# Probability and Finance

WILEY SERIES IN PROBABILITY AND STATISTICS
*FINANCIAL ENGINEERING SECTION*

Established by WALTER A. SHEWHART and SAMUEL S. WILKS



A complete list of the titles in this series appears at the end of this volume.

# Probability and Finance

## It's Only a Game!

GLENN SHAFER

*Rutgers University*
*Newark, New Jersey*

VLADIMIR VOVK

*Royal Holloway, University of London*
*Egham, Surrey, England*

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This text is printed on acid-free paper. ♾



For ordering and customer service, call 1-800-CALL WILEY.

***Library of Congress Cataloging-in-Publication Data:***

Shafer, Glenn, 1946–
Probability and finance : it's only a game! / Glenn Shafer and Vladimir Vovk.
p. cm. — (Wiley series in probability and statistics. Financial engineering section)
Includes bibliographical references and index.
ISBN 0-471-40226-5 (acid-free paper)
1. Investments—Mathematics. 2. Statistical decision. 3. Financial engineering. I. Vovk, Vladimir, 1960-. II. Title. III. Series.

HG4515.3 .S534 2001
332'.01'1—dc21 2001024030

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

# Contents

# Preface

This book shows how probability can be based on game theory, and how this can free many uses of probability, especially in finance, from distracting and confusing assumptions about randomness.

The connection of probability with games is as old as probability itself, but the game-theoretic framework we present in this book is fresh and novel, and this has made the book exciting for us to write. We hope to have conveyed our sense of excitement and discovery to the reader. We have only begun to mine a very rich vein of ideas, and the purpose of the book is to put others in a position to join the effort. We have tried to communicate fully the power of the game-theoretic framework, but whenever a choice had to be made, we have chosen clarity and simplicity over completeness and generality. This is not a comprehensive treatise on a mature and finished mathematical theory, ready to be shelved for posterity. It is an invitation to participate.

Our names as authors are listed in alphabetical order. This is an imperfect way of symbolizing the nature of our collaboration, for the book synthesizes points of view that the two of us developed independently in the 1980s and the early 1990s. The main mathematical content of the book derives from a series of papers Vovk completed in the mid-1990s. The idea of organizing these papers into a book, with a full account of the historical and philosophical setting of the ideas, emerged from a pleasant and productive seminar hosted by Aalborg University in June 1995. We are very grateful to Steffen Lauritzen for organizing that seminar and for persuading Vovk that his papers should be put into book form, with an enthusiasm that subsequently helped Vovk persuade Shafer to participate in the project.

Shafer's work on the topics of the book dates back to the late 1970s, when his study of Bayes's argument for conditional probability [274] first led him to insist that protocols for the possible development of knowledge should be incorporated into the foundations of probability and conditional probability [275]. His recognition that such protocols are equally essential to objective and subjective interpretations of probability led to a series of articles in the early 1990s arguing for a foundation of probability that goes deeper than the established measure-theoretic foundation but serves a diversity of interpretations [276, 277, 278, 279, 281]. Later in the 1990s, Shafer used event trees to explore the representation of causality within probability theory [283, 284, 285].

Shafer's work on the book itself was facilitated by his appointment as a Visiting Professor in Vovk's department, the Department of Computer Science at Royal Holloway, University of London. Shafer and Vovk are grateful to Alex Gammerman, head of the department, for his hospitality and support of this project. Shafer's work on the book also benefited from sabbatical leaves from Rutgers University in 1996–1997 and 2000–2001. During the first of these leaves, he benefited from the hospitality of his colleagues in Paris: Bernadette Bouchon-Meunier and Jean-Yves Jaffray at the Laboratoire d'Informatique de l'Université de Paris 6, and Bertrand Munier at the École Normale Supérieure de Cachan. During the second leave, he benefited from support from the German Fulbright Commission and from the hospitality of his colleague Hans-Joachim Lenz at the Free University of Berlin. During the 1999–2000 and 2000–2001 academic years, his research on the topics of the book was also supported by grant SES-9819116 from the National Science Foundation.

Vovk's work on the topics of the book evolved out of his work, first as an undergraduate and then as a doctoral student, with Andrei Kolmogorov, on Kolmogorov's finitary version of von Mises's approach to probability (see [319]). Vovk took his first steps towards a game-theoretic approach in the late 1980s, with his work on the law of the iterated logarithm [320, 321]. He argued for basing probability theory on the hypothesis of the impossibility of a gambling system in a discussion paper for the Royal Statistical Society, published in 1993. His paper on the game-theoretic Poisson process appeared in *Test* in 1993. Another, on a game-theoretic version of Kolmogorov's law of large numbers, appeared in *Theory of Probability and Its Applications* in 1996. Other papers in the series that led to this book remain unpublished; they provided early proofs of game-theoretic versions of Lindeberg's central limit theorem [328], Bachelier's central limit theorem [325], and the Black-Scholes formula [327], as well as a finance-theoretic strong law of large numbers [326].

While working on the book, Vovk benefited from a fellowship at the Center for Advanced Studies in the Behavioral Sciences, from August 1995 to June 1996, and from a short fellowship at the Newton Institute, November 17–22, 1997. Both venues provided excellent conditions for work. His work on the book has also benefited from several grants from EPSRC (GR/L35812, GR/M14937, and GR/M16856) and from visits to Rutgers. The earliest stages of his work were generously supported by George Soros's International Science Foundation. He is grateful to all his colleagues in the Department of Computer Science at Royal Holloway for a stimulating research environment and to his former Principal, Norman Gowar, for administrative and moral support.

Because the ideas in the book have taken shape over several decades, we find it impossible to give a complete account of our relevant intellectual debts. We do wish to acknowledge, however, our very substantial debt to Phil Dawid. His work on what he calls the "prequential" framework for probability and statistics strongly influenced us both beginning in the 1980s. We have not retained his terminology, but his influence is pervasive. We also wish to acknowledge the influence of the many colleagues who have discussed aspects of the book's ideas with us while we have been at work on it. Shashi Murthy helped us a great deal, beginning at a very early stage, as we sought to situate our ideas with respect to the existing finance literature. Others who have been exceptionally helpful at later stages include Steve Allen, Nick Bingham, Bernard Bru, Kaiwen Chen, Neil A. Chris, Pierre Crépel, Joseph L. Doob, Didier Dubois, Adlai Fisher, Hans Föllmer, Peter R. Gillett, Jean-Yves Jaffray, Phan Giang, Yuri Kalnichkan, Jack L. King, Eberhard Knobloch, Gabor Laszlo, Tony Martin, Nell Irvin Painter, Oded Palmon, Jan von Plato, Richard B. Scherl, Teddy Seidenfeld, J. Laurie Snell, Steve Stigler, Vladimir V'yugin, Chris Watkins, and Robert E. Whaley.

GLENN SHAFER

*Rutgers University, New Jersey, USA*

VLADIMIR VOVK

*Royal Holloway, University of London, Surrey, UK*

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 1

# Introduction: Probability and Finance as a Game

We propose a framework for the theory and use of mathematical probability that rests more on game theory than on measure theory. This new framework merits attention on purely mathematical grounds, for it captures the basic intuitions of probability simply and effectively. It is also of philosophical and practical interest. It goes deeper into probability's conceptual roots than the established measure-theoretic framework, it is better adapted to many practical problems, and it clarifies the close relationship between probability theory and finance theory.

Jean Ville (1910–1988) as a student at the *École Normale Supérieure* in Paris. His study of martingales helped inspire our framework for probability.

From the viewpoint of game theory, our framework is very simple. Its most essential elements were already present in Jean Ville's 1939 book, *Étude critique de la notion de collectif*, which introduced martingales into probability theory. Following Ville, we consider only two players. They alternate moves, each is immediately informed of the other's moves, and one or the other wins. In such a game, one player has a winning strategy (§4.6), and so we do not need the subtle solution concepts now at the center of game theory in economics and the other social sciences.

Our framework is a straightforward but rigorous elaboration, with no extraneous mathematical or philosophical baggage, of two ideas that are fundamental to both probability and finance:

- **The Principle of Pricing by Dynamic Hedging.** When simple gambles can be combined over time to produce more complex gambles, prices for the simple gambles determine prices for the more complex gambles.

- **The Hypothesis of the Impossibility of a Gambling System.** Sometimes we hypothesize that no system for selecting gambles from those offered to us can both (1) be certain to avoid bankruptcy and (2) have a reasonable chance of making us rich.

The principle of pricing by dynamic hedging can be discerned in the letters of Blaise Pascal to Pierre de Fermat in 1654, at the very beginning of mathematical probability, and it re-emerged in the last third of the twentieth century as one of the central ideas of finance theory. The hypothesis of the impossibility of a gambling system also has a long history in probability theory, dating back at least to Cournot, and it is related to the efficient-markets hypothesis, which has been studied in finance theory since the 1970s. We show that in a rigorous game-theoretic framework, these two ideas provide an adequate mathematical and philosophical starting point for probability and its use in finance and many other fields. No additional apparatus such as measure theory is needed to get probability off the ground mathematically, and no additional assumptions or philosophical explanations are needed to put probability to use in the world around us.

Probability becomes game-theoretic as soon as we treat the expected values in a probability model as prices in a game. These prices may be offered to an imaginary player who stands outside the world and bets on what the world will do, or they may be offered to an investor whose participation in a market constitutes a bet on what the market will do. In both cases, we can learn a great deal by thinking in game-theoretic terms. Many of probability's theorems turn out to be theorems about the existence of winning strategies for the player who is betting on what the world or market will do. The theorems are simpler and clearer in this form, and when they are in this form, we are in a position to reduce the assumptions we make—the number of prices we assume are offered—down to the minimum needed for the theorems to hold. This parsimony is potentially very valuable in practical work, for it allows and encourages clarity about the assumptions we need and are willing to take seriously.

Defining a probability measure on a sample space means recommending a definite price for each uncertain payoff that can be defined on the sample space, a price at which one might buy or sell the payoff. Our framework requires much less than this. We may be given only a few prices, and some of them may be one-sided—certified only for selling, not for buying, or vice versa. From these given prices, using dynamic hedging, we may obtain two-sided prices for some additional payoffs, but only upper and lower prices for others.

The measure-theoretic framework for probability, definitively formulated by Andrei Kolmogorov in 1933, has been praised for its philosophical neutrality: it can guide our mathematical work with probabilities no matter what meaning we want to give to these probabilities. Any numbers that satisfy the axioms of measure may be called probabilities, and it is up to the user whether to interpret them as frequencies, degrees of belief, or something else. Our game-theoretic framework is equally open to diverse interpretations, and its greater conceptual depth enriches these interpretations. Interpretations and uses of probability differ not only in the source of prices but also in the role played by the hypothesis of the impossibility of a gambling system.

Our framework differs most strikingly from the measure-theoretic framework in its ability to model open processes—processes that are open to influences we cannot model even probabilistically. This openness can, we believe, enhance the usefulness of probability theory in domains where our ability to control and predict is substantial but very limited in comparison with the sweep of a deterministic model or a probability measure.

From a mathematical point of view, the first test of a framework for probability is how elegantly it allows us to formulate and prove the subject's principal theorems, especially the classical limit theorems: the law of large numbers, the law of the iterated logarithm, and the central limit theorem. In Part I, we show how our game-theoretic framework meets this test. We contend that it does so better than the measure-theoretic framework. Our game-theoretic proofs sometimes differ little from standard measure-theoretic proofs, but they are more transparent. Our game-theoretic limit theorems are more widely applicable than their measure-theoretic counterparts, because they allow reality's moves to be influenced by moves by other players, including experimenters, professionals, investors, and citizens. They are also mathematically more powerful; the measure-theoretic counterparts follow from them as easy corollaries. In the case of the central limit theorem, we also obtain an interesting one-sided generalization, applicable when we have only upper bounds on the variability of individual deviations.

In Part II, we explore the use of our framework in finance. We call Part II "Finance without Probability" for two reasons. First, the two ideas that we consider fundamental to probability—the principle of pricing by dynamic hedging and the hypothesis of the impossibility of a gambling system—are also native to finance theory, and the exploitation of them in their native form in finance theory does not require extrinsic stochastic modeling. Second, we contend that the extrinsic stochastic modeling that does sometimes seem to be needed in finance theory can often be advantageously replaced by the further use of markets to set prices. Extrinsic stochastic modeling can also be accommodated in our framework, however, and Part II includes a game-theoretic treatment of diffusion processes, the extrinsic stochastic models that are most often used in finance and are equally important in a variety of other fields.

In the remainder of this introduction, we elaborate our main ideas in a relatively informal way. We explain how dynamic hedging and the impossibility of a gambling system can be expressed in game-theoretic terms, and how this leads to game-theoretic formulations of the classical limit theorems. Then we discuss the diversity of ways in which game-theoretic probability can be used, and we summarize how our relentlessly game-theoretic point of view can strengthen the theory of finance.

## 1.1 A GAME WITH THE WORLD

At the center of our framework is a sequential game with two players. The game may have many—perhaps infinitely many—rounds of play. On each round, Player I bets on what will happen, and then Player II decides what will happen. Both players have perfect information; each knows about the other's moves as soon as they are made.

In order to make their roles easier to remember, we usually call our two players Skeptic and World. Skeptic is Player I; World is Player II. This terminology is inspired by the idea of testing a probabilistic theory. Skeptic, an imaginary scientist who does not interfere with what happens in the world, tests the theory by repeatedly gambling imaginary money at prices the theory offers. Each time, World decides what does happen and hence how Skeptic's imaginary capital changes. If this capital becomes too large, doubt is cast on the theory. Of course, not all uses of mathematical probability, even outside of finance, are scientific. Sometimes the prices tested by Skeptic express personal choices rather than a scientific theory, or even serve merely as a straw man. But the idea of testing a scientific theory serves us well as a guiding example.

In the case of finance, we sometimes substitute the names Investor and Market for Skeptic and World. Unlike Skeptic, Investor is a real player, risking real money. On each round of play, Investor decides what investments to hold, and Market decides how the prices of these investments change and hence how Investor's capital changes.

### Dynamic Hedging

The principle of pricing by dynamic hedging applies to both probability and finance, but the word "hedging" comes from finance. An investor hedges a risk by buying and selling at market prices, possibly over a period of time, in a way that balances the risk. In some cases, the risk can be eliminated entirely. If, for example, Investor has a financial commitment that depends on the prices of certain securities at some future time, then he may be able to cover the commitment exactly by investing shrewdly in the securities during the rounds of play leading up to that future time. If the initial

***Table 1.1*** Instead of the uninformative names Player I and Player II, we usually call our players Skeptic and World, because it is easy to remember that World decides while Skeptic only bets. In the case of finance, we often call the two players Investor and Market.

| | PROBABILITY | FINANCE |
|---|---|---|
| **Player I** bets on what will happen. | **Skeptic** bets against the probabilistic predictions of a scientific theory. | **Investor** bets by choosing a portfolio of investments. |
| **Player II** decides what happens. | **World** decides how the predictions come out. | **Market** decides how the price of each investment changes. |

capital required is $\$\alpha$, then we may say that Investor has a strategy for turning $\$\alpha$ into the needed future payoff. Assuming, for simplicity, that the interest rate is zero, we may also say that $\$\alpha$ is the game's price for the payoff. This is the principle of pricing by dynamic hedging. (We assume throughout this chapter and in most of the rest of the book that the interest rate is zero. This makes our explanations and mathematics simpler, with no real loss in generality, because the resulting theory extends readily to the case where the interest rate is not zero: see §12.1.)

As it applies to probability, the principle of pricing by dynamic hedging says simply that the prices offered to Skeptic on each round of play can be compounded to obtain prices for payoffs that depend on more than one of World's moves. The prices for each round may include probabilities for what World will do on that round, and the global prices may include probabilities for World's whole sequence of play. We usually assume that the prices for each round are given either at the beginning of the game or as the game is played, and prices for longer-term gambles are derived. But when the idea of a probability game is used to study the world, prices may sometimes be derived in the opposite direction. The principle of pricing by dynamic hedging then becomes merely a principle of coherence, which tells us how prices at different times should fit together.

We impose no general rules about how many gambles are offered to Skeptic on different rounds of the game. On some rounds, Skeptic may be offered gambles on every aspect of World's next move, while on other rounds, he may be offered no gambles at all. Thus our framework always allows us to model what science models and to leave unmodeled what science leaves unmodeled.

### The Fundamental Interpretative Hypothesis

In contrast to the principle of pricing by dynamic hedging, the hypothesis of the impossibility of a gambling system is optional in our framework. The hypothesis boils down, as we explain in §1.3, to the supposition that events with zero or low probability are unlikely to occur (or, more generally, that events with zero or low upper probability are unlikely to occur). This supposition is fundamental to many uses of probability, because it makes the game to which it is applied into a theory about the world. By adopting the hypothesis, we put ourselves in a position to test the prices in the game: if an event with zero or low probability does occur, then we can reject the game as a model of the world. But we do not always adopt the hypothesis. We do not always need it when the game is between Investor and Market, and we do not need it when we interpret probabilities subjectively, in the sense advocated by Bruno de Finetti. For de Finetti and his fellow neosubjectivists, a person's subjective prices are nothing more than that; they are merely prices that systematize the person's choices among risky options. See §1.4 and §2.6.

We have a shorter name for the hypothesis of the impossibility of a gambling system: we call it the *fundamental interpretative hypothesis* of probability. It is interpretative because it tells us what the prices and probabilities in the game to which it is applied mean in the world. It is not part of our mathematics. It stands outside the mathematics, serving as a bridge between the mathematics and the world.

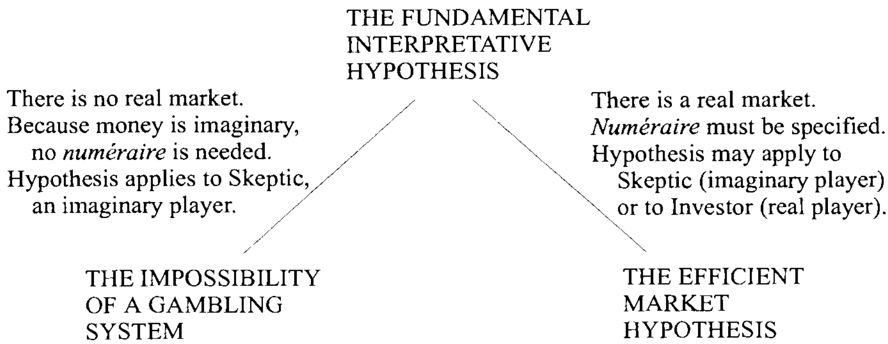



***Fig. 1.1*** The fundamental interpretative hypothesis in probability and finance.

When we are working in finance, where our game describes a real market, we use yet another name for our fundamental hypothesis: we call it the *efficient-market hypothesis*. The efficient-market hypothesis, as applied to a particular financial market, in which particular securities are bought and sold over time, says that an investor (perhaps a real investor named Investor, or perhaps an imaginary investor named Skeptic) cannot become rich trading in this market without risking bankruptcy. In order to make such a hypothesis precise, we must specify not only whether we are talking about Investor or Skeptic, but also the *numéraire*—the unit of measurement in which this player's capital is measured. We might measure this capital in nominal terms (making a monetary unit, such as a dollar or a ruble, the *numéraire*), we might measure it relative to the total value of the market (making some convenient fraction of this total value the *numéraire*), or we might measure it relative to a risk-free bond (which is then the *numéraire*), and so on. Thus the efficient-market hypothesis can take many forms. Whatever form it takes, it is subject to test, and it determines upper and lower probabilities that have empirical meaning.

Since about 1970, economists have debated an efficient-markets hypothesis, with *markets* in the plural. This hypothesis says that financial markets are efficient in general, in the sense that they have already eliminated opportunities for easy gain. As we explain in Part II (§9.4 and Chapter 15), our efficient-market hypothesis has the same rough rationale as the efficient-markets hypothesis and can often be tested in similar ways. But it is much more specific. It requires that we specify the particular securities that are to be included in the market, the exact rule for accumulating capital, and the *numéraire* for measuring this capital.

### Open Systems within the World

Our austere picture of a game between Skeptic and World can be filled out in a great variety of ways. One of the most important aspects of its potential lies in the possibility of dividing World into several players. For example, we might divide World into three players:

Experimenter, who decides what each round of play will be about.
Forecaster, who sets the prices.
Reality, who decides the outcomes.

This division reveals the open character of our framework. The principle of pricing by dynamic hedging requires Forecaster to give coherent prices, and the fundamental interpretative hypothesis requires Reality to respect these prices, but otherwise all three players representing World may be open to external information and influence. Experimenter may have wide latitude in deciding what experiments to perform. Forecaster may use information from outside the game to set prices. Reality may also be influenced by unpredictable outside forces, as long as she acts within the constraints imposed by Forecaster.

Many scientific models provide testable probabilistic predictions only subsequent to the determination of many unmodeled auxiliary factors. The presence of Experimenter in our framework allows us to handle these models very naturally. For example, the standard mathematical formalization of quantum mechanics in terms of Hilbert spaces, due to John von Neumann, fits readily into our framework. The scientist who decides what observables to measure is Experimenter, and quantum theory is Forecaster (§8.4).

Weather forecasting provides an example where information external to a model is used for prediction. Here Forecaster may be a person or a very complex computer program that escapes precise mathematical definition because it is constantly under development. In either case, Forecaster will use extensive external information—weather maps, past experience, etc. If Forecaster is required to announce every evening a probability for rain on the following day, then there is no need for Experimenter; the game has only three players, who move in this order:

Forecaster, Skeptic, Reality.

Forecaster announces odds for rain the next day, Skeptic decides whether to bet for or against rain and how much, and Reality decides whether it rains. The fundamental interpretative hypothesis, which says that Skeptic cannot get rich, can be tested by any strategy for betting at Forecaster's odds.

It is more difficult to make sense of the weather forecasting problem in the measure-theoretic framework. The obvious approach is to regard the forecaster's probabilities as conditional probabilities given what has happened so far. But because the forecaster is expected to learn from his experience in giving probability forecasts, and because he uses very complex and unpredictable external information, it makes no sense to interpret his forecasts as conditional probabilities in a probability distribution formulated at the outset. And the forecaster does not construct a probability distribution along the way; this would involve constructing probabilities for what will happen on the next day not only conditional on what has happened so far but also conditional on what might have happened so far.

In the 1980s, A. Philip Dawid proposed that the forecasting success of a probability distribution for a sequence of events should be evaluated using only the actual outcomes and the sequence of forecasts (conditional probabilities) to which these outcomes give rise, without reference to other aspects of the probability distribution. This is Dawid's *prequential principle* [82]. In our game-theoretic framework, the prequential principle is satisfied automatically, because the probability forecasts provided by Forecaster and the outcomes provided by Reality are all we have. So long as Forecaster does not adopt a strategy, no probability distribution is even defined.

The explicit openness of our framework makes it well suited to modeling systems that are open to external influence and information, in the spirit of the nonparametric, semiparametric, and martingale models of modern statistics and the even looser predictive methods developed in the study of machine learning. It also fits the open spirit of modern science, as emphasized by Karl Popper [250]. In the nineteenth century, many scientists subscribed to a deterministic philosophy inspired by Newtonian physics: at every moment, every future aspect of the world should be predictable by a superior intelligence who knows initial conditions and the laws of nature. In the twentieth century, determinism was strongly called into question by further advances in physics, especially in quantum mechanics, which now insists that some fundamental phenomena can be predicted only probabilistically. Probabilists sometimes imagine that this defeat allows a retreat to a probabilistic generalization of determinism: science should give us probabilities for everything that might happen in the future. In fact, however, science now describes only islands of order in an unruly universe. Modern scientific theories make precise probabilistic predictions only about some aspects of the world, and often only after experiments have been designed and prepared. The game-theoretic framework asks for no more.

### Skeptic and World Always Alternate Moves

Most of the mathematics in this book is developed for particular examples, and as we have just explained, many of these examples divide World into multiple players. It is important to notice that this division of World into multiple players does not invalidate the simple picture in which Skeptic and World alternate moves, with Skeptic betting on what World will do next, because we will continue to use this simple picture in our general discussions, in the next section and in later chapters.

One way of seeing that the simple picture is preserved is to imagine that Skeptic moves just before each of the players who constitute World, but that only the move just before Reality can result in a nonzero payoff for Skeptic. Another way, which we will find convenient when World is divided into Forecaster and Reality, is to add just one dummy move by Skeptic, at the beginning of the game, and then to group each of Forecaster's later moves with the preceding move by Reality, so that the order of play becomes

Skeptic, Forecaster, Skeptic, (Reality, Forecaster),
Skeptic, (Reality, Forecaster), . . . .

Either way, Skeptic alternates moves with World.

### The Science of Finance

Other players sometimes intrude into the game between Investor and Market. Finance is not merely practice; there is a theory of finance, and our study of it will sometimes require that we bring Forecaster and Skeptic into the game. This happens in several different ways. In Chapter 14, where we give a game-theoretic reading of the usual stochastic treatment of option pricing, Forecaster represents a probabilistic theory about the behavior of the market, and Skeptic tests this theory. In our study of the efficient-market hypothesis (Chapter 15), in contrast, the role of Forecaster is played by Opening Market, who sets the prices at which Investor, and perhaps also Skeptic, can buy securities. The role of Reality is then played by Closing Market, who decides how these investments come out.

In much of Part II, however, especially in Chapters 10–13, we study games that involve Investor and Market alone. These may be the most important market games that we study, because they allow conclusions based solely on the structure of the market, without appeal to any theory about the efficiency of the market or the stochastic behavior of prices.

## 1.2 THE PROTOCOL FOR A PROBABILITY GAME

Specifying a game fully means specifying the moves available to the players—we call this the *protocol* for the game—and the rule for determining the winner. Both of these elements can be varied in our game between Skeptic and World, leading to many different games, all of which we call *probability games*. The protocol determines the sample space and the prices (in general, upper and lower prices) for variables. The rule for determining the winner can be adapted to the particular theorem we want to prove or the particular problem where we want to use the framework. In this section we consider only the protocol.

The general theory sketched in this section applies to most of the games studied in this book, including those where Investor is substituted for Skeptic and Market for World. (The main exceptions are the games we use in Chapter 13 to price American options.) We will develop this general theory in more detail in Chapters 7 and 8.

### The Sample Space

The protocol for a probability game specifies the moves available to each player, Skeptic and World, on each round. This determines, in particular, the sequences of moves World may make. These sequences—the possible complete sequences of play by World—constitute the *sample space* for the game. We designate the sample space by $\Omega$, and we call its elements *paths*. The moves available to World may depend on moves he has previously made. But we assume that they do not depend on moves Skeptic has made. Skeptic's bets do not affect what is possible in the world, although World may consider them in deciding what to do next.

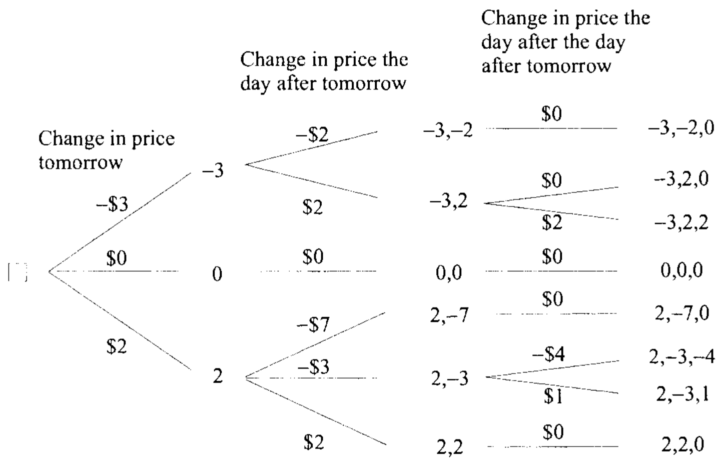



***Fig. 1.2*** An unrealistic sample space for changes in the price of a stock. The steps in the tree represent possible moves by World (in this case, the market). The nodes (situations) record the moves made by World so far. The initial situation is designated by □. The terminal nodes record complete sequences of play by World and hence can be identified with the paths that constitute the sample space. The example is unrealistic because in a real stock market there is a wide range of possible changes for a stock's price at each step, not just two or three.

We can represent the dependence of World's possible moves on his previous moves in terms of a tree whose paths form the sample space, as in Figure 1.2. Each node in the tree represents a *situation*, and the branches immediately to the right of a nonterminal situation represent the moves World may make in that situation. The initial situation is designated by □.

Figure 1.2 is finite: there are only finitely many paths, and every path terminates after a finite number of moves. We do not assume finiteness in general, but we do pay particular attention to the case where every path terminates; in this case we say the game is *terminating*. If there is a bound on the length of the paths, then we say the game has a *finite horizon*. If none of the paths terminate, we say the game has an *infinite horizon*.

In general, we think of a situation (a node in the tree) as the sequence of moves made by World so far, as explained in the caption of Figure 1.2. So in a terminating game, we may identify the terminal situation on each path with that path; both are the same sequence of moves by World.

In measure-theoretic probability theory, a real-valued function on the sample space is called a *random variable*. Avoiding the implication that we have defined a probability measure on the sample space, and also whatever other ideas the reader may associate with the word "random", we call such a function simply a *variable*. In the example of Figure 1.2, the variables include the prices for the stock for each of the next three days, the average of the three prices, the largest of the three prices,

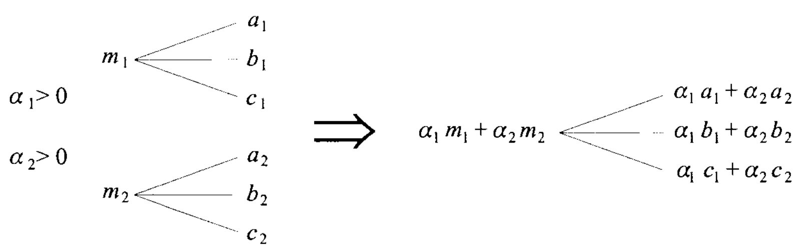



***Fig. 1.3*** Forming a nonnegative linear combination of two gambles. In the first gamble Skeptic pays $m_1$ in order to get $a_1$, $b_1$, or $c_1$ in return, depending on how things come out. In the second gamble, he pays $m_2$ in order to get $a_2$, $b_2$, or $c_2$ in return.

and so on. We also follow established terminology by calling a subset of the sample space an *event*.

## Moves and Strategies for Skeptic

To complete the protocol for a probability game, we must also specify the moves Skeptic may make in each situation. Each move for Skeptic is a gamble, defined by a price to be paid immediately and a payoff that depends on World's following move. The gambles among which Skeptic may choose may depend on the situation, but we always allow him to combine available gambles and to take any fraction or multiple of any available gamble. We also allow him to borrow money freely without paying interest. So he can take any nonnegative linear combination of any two available gambles, as indicated in Figure 1.3.

We call the protocol *symmetric* if Skeptic is allowed to take either side of any available gamble. This means that whenever he can buy the payoff $x$ at the price $m$, he can also sell $x$ at the price $m$. Selling $x$ for $m$ is the same as buying $-x$ for $-m$ (Figure 1.4). So a symmetric protocol is one in which the gambles available to Skeptic in each situation form a linear space; he may take any linear combination of the available gambles, whether or not the coefficients in the linear combination are nonnegative. If we neglect bid-ask spreads and transaction costs, then protocols based on market prices are symmetric, because one may buy as well as sell a security at its market price. Protocols corresponding to complete probability measures are also symmetric. But many of the protocols we will study in this book are asymmetric.

A *strategy* for Skeptic is a plan for how to gamble in each nonterminal situation he might encounter. His strategy together with his initial capital determine his capital in every situation, including terminal situations. Given a strategy $\mathcal{P}$ and a situation $t$, we write $\mathcal{K}^{\mathcal{P}}(t)$ for Skeptic's capital in $t$ if he starts with capital 0 and follows $\mathcal{P}$. In the terminating case, we may also speak of the capital a strategy produces at the end of the game. Because we identify each path with its terminal situation, we may write $\mathcal{K}^{\mathcal{P}}(\xi)$ for Skeptic's final capital when he follows $\mathcal{P}$ and World takes the path $\xi$.


***Fig. 1.4*** Taking the gamble on the left means paying $m$ and receiving $a$, $b$, or $c$ in return. Taking the other side means receiving $m$ and paying $a$, $b$, or $c$ in return—i.e., paying $-m$ and receiving $-a$, $-b$, or $-c$ in return. This is the same as taking the gamble on the right.

| | **Meaning** | **Net payoff** | **Simulated satisfactorily by $\mathcal{P}$ if** |
|---|---|---|---|
| Buy $x$ for $\alpha$ | Pay $\alpha$, get $x$ | $x - \alpha$ | $\mathcal{K}^{\mathcal{P}} \geq x - \alpha$ |
| Sell $x$ for $\alpha$ | Get $\alpha$, pay $x$ | $\alpha - x$ | $\mathcal{K}^{\mathcal{P}} \geq \alpha - x$ |

***Table 1.2*** How a strategy $\mathcal{P}$ in a probability game can simulate the purchase or sale of a variable $x$.

## Upper and Lower Prices

By adopting different strategies in a probability game, Skeptic can simulate the purchase and sale of variables. We can price variables by considering when this succeeds. In order to explain this idea as clearly as possible, we make the simplifying assumption that the game is terminating.

A strategy simulates a transaction satisfactorily for Skeptic if it produces at least as good a net payoff. Table 1.2 summarizes how this applies to buying and selling a variable $x$. As indicated there, $\mathcal{P}$ simulates buying $x$ for $\alpha$ satisfactorily if $\mathcal{K}^{\mathcal{P}} \geq x - \alpha$. This means that

$$\mathcal{K}^{\mathcal{P}}(\xi) \geq x(\xi) - \alpha$$

for every path $\xi$ in the sample space $\Omega$. When Skeptic has a strategy $\mathcal{P}$ satisfying $\mathcal{K}^{\mathcal{P}} \geq x - \alpha$, we say he *can buy $x$ for $\alpha$*. Similarly, when he has a strategy $\mathcal{P}$ satisfying $\mathcal{K}^{\mathcal{P}} \geq \alpha - x$, we say he *can sell $x$ for $\alpha$*. These are two sides of the same coin: selling $x$ for $\alpha$ is the same as buying $-x$ for $-\alpha$.

Given a variable $x$, we set

$$\overline{\mathbb{E}}\,x := \inf\left\{\alpha \mid \text{ there is some strategy } \mathcal{P} \text{ such that } \mathcal{K}^{\mathcal{P}} \geq x - \alpha\right\}. \tag{1.1}$$ [1]

We call $\overline{\mathbb{E}}\,x$ the *upper price* of $x$ or the *cost* of $x$; it is the lowest price at which Skeptic can buy $x$. (Because we have made no compactness assumptions about the protocol—and will make none in the sequel—the infimum in (1.1) may not be attained, and so strictly speaking we can only be sure that Skeptic can buy $x$ for

[1] We use := to mean "equal by definition"; the right-hand side of the equation is the definition of the left-hand side.

$\overline{\mathbb{E}}\,x + \epsilon$ for every $\epsilon > 0$. But it would be tedious to mention this constantly, and so we ask the reader to indulge the slight abuse of language involved in saying that Skeptic can buy $x$ for $\overline{\mathbb{E}}\,x$.)

Similarly, we set

$$\underline{\mathbb{E}}\,x := \sup\left\{\alpha \mid \text{ there is some strategy } \mathcal{P} \text{ such that } \mathcal{K}^{\mathcal{P}} \geq \alpha - x\right\}. \qquad (1.2)$$

We call $\underline{\mathbb{E}}\,x$ the *lower price* of $x$ or the *scrap value* of $x$; it is the highest price at which Skeptic can sell $x$.

It follows from (1.1) and (1.2), and also directly from the fact that selling $x$ for $\alpha$ is the same as buying $-x$ for $-\alpha$, that

$$\underline{\mathbb{E}}\,x = -\,\overline{\mathbb{E}}[-x]$$

for every variable $x$.

The idea of hedging provides another way of talking about upper and lower prices. If we have an obligation to pay something at the end of the game, then we hedge this obligation by trading in such a way as to cover the payment no matter what happens. So we say that the strategy $\mathcal{P}$ *hedges* the obligation $y$ if

$$\mathcal{K}^{\mathcal{P}}(\xi) \geq y(\xi) \qquad (1.3)$$

for every path $\xi$ in the sample space $\Omega$. Selling a variable $x$ for $\alpha$ results in a net obligation of $x - \alpha$ at the end of the game. So $\mathcal{P}$ hedges selling $x$ for $\alpha$ if $\mathcal{P}$ hedges $x - \alpha$, that is, if $\mathcal{P}$ simulates buying $x$ for $\alpha$. Similarly, $\mathcal{P}$ hedges buying $x$ for $\alpha$ if $\mathcal{P}$ simulates selling $x$ for $\alpha$. So $\overline{\mathbb{E}}\,x$ is the lowest price at which selling $x$ can be hedged, and $\underline{\mathbb{E}}\,x$ is the highest price at which buying it can be hedged, as indicated in Table 1.3.

These definitions implicitly place Skeptic at the beginning of the game, in the initial situation $\Box$. They can also be applied, however, to any other situation; we simply consider Skeptic's strategies for play from that situation onward. We write $\overline{\mathbb{E}}_t\,x$ and $\underline{\mathbb{E}}_t\,x$ for the upper and lower price, respectively, of the variable $x$ in the situation $t$.

***Table 1.3*** Upper and lower price described in terms of simulation and described in terms of hedging. Because hedging the sale of $x$ is the same as simulating the purchase of $x$, and vice versa, the two descriptions are equivalent.

| | **Name** | **Description in terms of the simulation of buying and selling** | **Description in terms of hedging** |
|---|---|---|---|
| $\overline{\mathbb{E}}\,x$ | Upper price of $x$ | Lowest price at which Skeptic can buy $x$ | Lowest selling price for $x$ Skeptic can hedge |
| $\underline{\mathbb{E}}\,x$ | Lower price of $x$ | Highest price at which Skeptic can sell $x$ | Highest buying price for $x$ Skeptic can hedge |

Upper and lower prices are interesting only if the gambles Skeptic is offered do not give him an opportunity to make money for certain. If this condition is satisfied in situation $t$, we say that the protocol is *coherent* in $t$. In this case,

$$\underline{\mathbb{E}}_t\, x \leq \overline{\mathbb{E}}_t\, x$$

for every variable $x$, and

$$\overline{\mathbb{E}}_t\, \mathbf{0} = \underline{\mathbb{E}}_t\, \mathbf{0} = 0,$$

where $\mathbf{0}$ denotes the variable whose value is 0 on every path in $\Omega$.

When $\underline{\mathbb{E}}_t\, x = \overline{\mathbb{E}}_t\, x$, we call their common value the *exact price* or simply the *price* for $x$ in $t$ and designate it by $\mathbb{E}_t\, x$. Such prices have the properties of expected values in measure-theoretic probability theory, but we avoid the term "expected value" in order to avoid suggesting that we have defined a probability measure on our sample space. We do, however, use the word "variance"; when $\mathbb{E}_t\, x$ exists, we set

$$\overline{\mathbb{V}}_t\, x := \overline{\mathbb{E}}_t(x - \mathbb{E}_t\, x)^2 \quad \text{and} \quad \underline{\mathbb{V}}_t\, x := \underline{\mathbb{E}}_t(x - \mathbb{E}_t\, x)^2,$$

and we call them, respectively, the *upper variance* of $x$ in $t$ and the *lower variance* of $x$ in $t$. If $\overline{\mathbb{V}}_t\, x$ and $\underline{\mathbb{V}}_t\, x$ are equal, we write $\mathbb{V}_t\, x$ for their common value; this is the (game-theoretic)*variance* of $x$ in $t$.

When the game is not terminating, definitions (1.1), (1.2), and (1.3) do not work, because $\mathcal{P}$ may fail to determine a final capital for Skeptic when World takes an infinite path; if there is no terminal situation on the path $\xi$, then $\mathcal{K}^{\mathcal{P}}(t)$ may or may not converge to a definite value as $t$ moves along $\xi$. Of the several ways to fix this, we prefer the simplest: we say that $\mathcal{P}$ hedges $y$ if on every path $\xi$ the capital $\mathcal{K}^{\mathcal{P}}(t)$ eventually reaches $y(\xi)$ and stays at or above it, and we similarly modify (1.1) and (1.2). We will study this definition in §8.3. On the whole, we make relatively little use of upper and lower price for nonterminating probability games, but as we explain in the next section, we do pay great attention to one special case, the case of probabilities exactly equal to zero or one.

## 1.3 THE FUNDAMENTAL INTERPRETATIVE HYPOTHESIS

The fundamental interpretative hypothesis asserts that no strategy for Skeptic can both (1) be certain to avoid bankruptcy and (2) have a reasonable chance of making Skeptic rich. Because it contains the undefined term "reasonable chance", this hypothesis is not a mathematical statement; it is neither an axiom nor a theorem. Rather it is an interpretative statement. It gives meaning in the world to the prices in the probability game. Once we have asserted that Skeptic does not have a reasonable chance of multiplying his initial capital substantially, we can identify other likely and unlikely events and calibrate just how likely or unlikely they are. An event is unlikely if its happening would give an opening for Skeptic to multiply his initial capital substantially, and it is the more unlikely the more substantial this multiplication is.

We use two distinct versions of the fundamental interpretative hypothesis, one *finitary* and one *infinitary*:

- **The Finitary Hypothesis.** No strategy for Skeptic can both (1) be certain to avoid bankruptcy and (2) have a reasonable chance of multiplying his initial capital by a large factor. (We usually use this version in terminating games.)
- **The Infinitary Hypothesis.** No strategy for Skeptic can both (1) be certain to avoid bankruptcy and (2) have a reasonable chance of making him infinitely rich. (We usually use this version in infinite-horizon games.)

Because our experience with the world is finite, the finitary hypothesis is of more practical use, but the infinitary hypothesis often permits clearer and more elegant mathematical statements. As we will show in Part I, the two forms lead to the two types of classical limit theorems. The finitary hypothesis leads to the weak limit theorems: the weak law of large numbers and the central limit theorem. The infinitary hypothesis leads to the strong limit theorems: the strong law of large numbers and the law of the iterated logarithm.

It is easy for World to satisfy the fundamental interpretative hypothesis in a probability game with a coherent protocol, for he can always move so that Skeptic does not make money. But becoming rich is not Skeptic's only goal in the games we study. In many of these games, Skeptic wins *either* if he becomes rich *or* if World's moves satisfy some other condition $E$. If Skeptic has a winning strategy in such a game, then the fundamental interpretative hypothesis authorizes us to conclude that $E$ will happen. In order to keep Skeptic from becoming rich, World must move so as to satisfy $E$.

## Low Probability and High Probability

In its finitary form, the fundamental interpretative hypothesis provides meaning to small upper probabilities and large lower probabilities.

We can define upper and lower probabilities formally as soon as we have the concepts of upper and lower price. As we mentioned earlier, an *event* is a subset of the sample space. Given an event $E$, we define its *indicator variable* $\mathbb{I}_E$ by

$$\mathbb{I}_E(\xi) := \begin{cases} 1 \text{ if } \xi \in E \\ 0 \text{ if } \xi \notin E. \end{cases}$$

Then we define its *upper probability* by

$$\overline{\mathbb{P}}E := \overline{\mathbb{E}}\,\mathbb{I}_E \tag{1.4}$$

and its *lower probability* by

$$\underline{\mathbb{P}}E := \underline{\mathbb{E}}\,\mathbb{I}_E\,. \tag{1.5}$$

Assuming the protocol is coherent, upper and lower probability obey

$$0 \leq \underline{\mathbb{P}}E \leq \overline{\mathbb{P}}E \leq 1 \tag{1.6}$$

and

$$\underline{\mathbb{P}}E = 1 - \overline{\mathbb{P}}E^{c}. \tag{1.7}$$

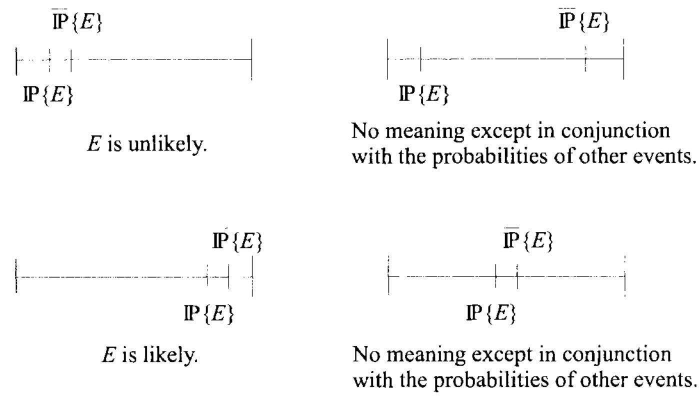



***Fig. 1.5*** Only extreme probabilities have meaning in isolation.

Here $E^c$ is the *complement* of $E$ in $\Omega$—the set of paths for World that are not in $E$, or the event that $E$ does not happen.

What meaning can be attached to $\overline{\mathbb{P}}\,E$ and $\underline{\mathbb{P}}\,E$? The fundamental interpretative hypothesis answers this question when the two numbers are very close to zero. Suppose, for example, that $\overline{\mathbb{P}}\,E = 0.001$. (In this case, $\underline{\mathbb{P}}\,E$ is also close to zero; by (1.6), it is between 0 and 0.001.) Then Skeptic can buy $\mathbb{I}_E$ for 0.001. Because $\mathbb{I}_E \geq 0$, the purchase does not open him to possible bankruptcy, and yet it results in a thousandfold increase in his investment if $E$ happens. The fundamental interpretative hypothesis says that this increase is unlikely and hence implies that $E$ is unlikely.

Similarly, we may say that $E$ is very likely to happen if $\underline{\mathbb{P}}\,E$ and hence also $\overline{\mathbb{P}}\,E$ are very close to one. Indeed, if $\underline{\mathbb{P}}\,E$ is close to one, then by (1.7), $\overline{\mathbb{P}}\,E^c$ is close to zero, and hence it is unlikely that $E^c$ will happen—that is, it is likely that $E$ will happen.

These interpretations are summarized in Figure 1.5. If $\overline{\mathbb{P}}\,E$ and $\underline{\mathbb{P}}\,E$ are neither both close to zero nor both close to one, as on the right in the figure, then they have little or no meaning in isolation. But if they are both close to zero, then we may say that $E$ has "low probability" and is unlikely to happen. And if they are both close to one, then we may say that $E$ has "high probability" and is likely to happen.

Strictly speaking, we should speak of the probability of $E$ only if $\overline{\mathbb{P}}\,E$ and $\underline{\mathbb{P}}\,E$ are exactly equal, for then their common value may be called the (game-theoretic) probability of $E$. But as the figure indicates, it is much more meaningful for the two values to both be close to zero or both be close to one than for them to be exactly equal.

The most important examples of low and high probability in this book occur in the two weak laws that we study in Chapters 6 and 7: the weak law of large numbers and the central limit theorem. The weak law of large numbers, in its simplest form, says that when Skeptic is offered even odds on each of a long sequence of events, the probability is high that the fraction of the events that happen will fall within a small interval around $1/2$: an interval that may be narrowed as the number of the events increases. The central limit theorem gives numerical estimates of this high probability. According to our definition of high probability, these theorems say something about Skeptic's opportunities to make money. The law of large numbers says that Skeptic has a winning strategy in a game that he wins if World either stays close to $1/2$ or allows Skeptic to multiply his stake substantially, and the central limit theorem calibrates the tradeoff between how far World can stray from $1/2$ and how much he can constrain Skeptic's wealth.

Middling probabilities, although they do not have meaning in isolation, can acquire collective meaning from the limit theorems. The law of large numbers tells us, for example, that many probabilities for successive events all equal to $1/2$ produce a very high probability that the relative frequency of the events will approximate $1/2$.

### Probability Zero and Probability One

As we have just emphasized, the finitary version of our fundamental hypothesis gives meaning to probabilities very close to zero or one. Skeptic is unlikely to become very rich, and therefore an event with a very low probability is unlikely to occur. The infinitary version sharpens this by giving meaning to probabilities exactly equal to zero or one. It is practically impossible for Skeptic to become infinitely rich, and therefore an event that makes this possible is practically certain not to occur.

Formally, we say that an event $E$ is *practically impossible* if Skeptic, beginning with some finite positive capital, has a strategy that guarantees that

- his capital does not become negative (he does not go bankrupt), and
- if $E$ happens, his capital increases without bound (he becomes infinitely rich).

We say that an event $E$ is *practically certain*, or that it happens *almost surely*, if its complement $E^c$ is practically impossible. It follows immediately from these definitions that a practically impossible event has upper probability (and hence also lower probability) zero, and that a practically certain event has lower probability (and hence also upper probability) one (§8.3).

The size of Skeptic's initial capital does not matter in the definitions of practical certainty and practical impossibility, provided it is positive. If the strategy $\mathcal{P}$ will do what is required when his initial capital is $a$, then the strategy $\frac{b}{a}\mathcal{P}$ will accomplish the same trick when his initial capital is $b$. Requiring that Skeptic's capital not become negative is equivalent to forbidding him to borrow money, because if he dared to gamble on borrowed money, World could force his capital to become negative. The real condition, however, is not that he never borrow but that his borrowing be bounded. Managing on initial capital $a$ together with borrowing limited to $b$ is the same as managing on initial capital $a + b$.

As we show in Chapters 3, 4, and 5, these definitions allow us to state and prove game-theoretic versions of the classical strong limit theorems—the strong law of large numbers and the law of the iterated logarithm. In its simplest form, the game-theoretic strong law of large numbers says that when Skeptic is offered even odds on each of an infinite sequence of events, the fraction of the events that happen will almost certainly converge to $1/2$. The law of the iterated logarithm gives the best possible bound on the rate of convergence.

## Beyond Frequencies

As we explain in some detail in Chapter 2, the law of large numbers, together with the empiricist philosophy of the time, led in the nineteenth and early twentieth centuries to a widespread conviction that the theory of probability should be founded on the concept of relative frequency. If many independent trials are made of an event with probability $p$, then the law of large numbers says that the event will happen $p$ of the time and fail $1 - p$ of the time. This is true whether we consider all the trials, or only every other trial, or only some other subsequence selected in advance. And this appears to be the principal empirical meaning of probability. So why not turn the theory around, as Richard von Mises proposed in the 1920s, and say that a probability is merely a relative frequency that is invariant under selection of subsequences?

As it turned out, von Mises was mistaken to emphasize frequency to the exclusion of other statistical regularities. The predictions about a sequence of events made by probability theory do not all follow from the invariant convergence of relative frequency. In the late 1930s, Jean Ville pointed out a very salient and decisive example: the predictions that the law of the iterated logarithm makes about the rate and oscillation of the convergence. Von Mises's theory has now been superseded by the theory of algorithmic complexity, which is concerned with the properties of sequences whose complexity makes them difficult to predict, and invariant relative frequency is only one of many such properties.

Frequency has also greatly receded in prominence within measure-theoretic probability. Where independent identically distributed random variables were once the central object of study, we now study stochastic processes in which the probabilities of events depend on preceding outcomes in complex ways. These models sometimes make predictions about frequencies, but instead of relating a frequency to a single probability, they may predict that a frequency will approximate the average of a sequence of probabilities. In general, emphasis has shifted from sums of independent random variables to martingales.

For some decades, it has been clear to mathematical probabilists that martingales are fundamental to their subject. Martingales remain, however, only an advanced topic in measure-theoretic probability theory. Our game-theoretic framework puts what is fundamental at the beginning. Martingales come at the beginning, because they are the capital processes for Skeptic. The fundamental interpretative hypothesis, applied to a particular nonnegative martingale, says that the world will behave in such a way that the martingale remains bounded. And the many predictions that follow include the convergence of relative frequencies.

## 1.4 THE MANY INTERPRETATIONS OF PROBABILITY

Contemporary philosophical discussions often divide probabilities into two broad classes:

- *objective probabilities*, which describe frequencies and other regularities in the world, and
- *subjective probabilities*, which describe a person's preferences, real or hypothetical, in risk taking.

Our game-theoretic framework accommodates both kinds of probabilities and enriches our understanding of them, while opening up other possibilities as well.

### Three Major Interpretations

From our point of view, it makes sense to distinguish three major ways of using the idea of a probability game, which differ in how prices are established and in the role of the fundamental interpretative hypothesis, as indicated in Table 1.4.

Games of statistical regularity express the objective conception of probability within our framework. In a game of statistical regularity, the gambles offered to Skeptic may derive from a scientific theory, from frequencies observed in the past, or from some relatively poorly understood forecasting method. Whatever the source, we adopt the fundamental interpretative hypothesis, and this makes statistical regularity the ultimate authority: the prices and probabilities determined by the gambles offered to Skeptic must be validated by experience. We expect events assigned small upper probabilities not to happen, and we expect prices to be reflected in average values.

Games of belief bring the neosubjectivist conception of probability into our framework. A game of belief may draw on scientific theories or statistical regularities to determine the gambles offered on individual rounds. But the presence of these gambles in the game derives from some individual's commitment to use them to rank and choose among risks. The individual does not adopt the fundamental interpretative

***Table 1.4*** Three classes of probability games.

| | **Authority for the Prices** | **Role of the Fundamental Interpretative Hypothesis** |
|---|---|---|
| **Games of Statistical Regularity** | Statistical regularities | Adopted |
| **Games of Belief** | Personal choices among risks | Not adopted |
| **Market Games** | Market for financial securities | Optional |

hypothesis, and so his prices cannot be falsified by what actually happens. The upper and lower prices and probabilities in the game are not the individual's hypotheses about what will happen; they merely indicate the risks he will take. A low probability does not mean the individual thinks an event will not happen; it merely means he is willing to bet heavily against it.

Market games are distinguished by the source of their prices: these prices are determined by supply and demand in some market. We may or may not adopt the hypothesis that the market is efficient. If we do adopt it, then we may test it or use it to draw various conclusions (see, e.g., the discussion of the Iowa Electronic Markets on p. 71). If we do not adopt it, even provisionally, then the game can still be useful as a framework for understanding the hedging of market risks.

Our understanding of objective and subjective probability in terms of probability games differs from the usual explanations of these concepts in its emphasis on sequential experience. Objective probability is often understood in terms of a population, whose members are not necessarily examined in sequence, and most expositions of subjective probability emphasize the coherence of one's belief about different events without regard to how those events might be arranged in time. But we do experience the world through time, and so the game-theoretic framework offers valuable insights for both the objective and the subjective conceptions. Objective probabilities can only be tested over time, and the idea of a probability game imposes itself whenever we want to understand the testing process. The experience anticipated by subjective probabilities must also be arrayed in time, and probability games are the natural framework in which to understand how subjective probabilities change as that experience unfolds.

### Looking at Interpretations in Two Dimensions

The uses and interpretations of probability are actually very diverse—so much so that we expect most readers to be uncomfortable with the standard dichotomy between objective and subjective probability and with the equally restrictive categories of Table 1.4. A more flexible categorization of the diverse possibilities for using the mathematical idea of a probability game can be developed by distinguishing uses along two dimensions: (1) the source of the prices, and (2) the attitude taken towards the fundamental interpretative hypothesis. This is illustrated in Figure 1.6.

We use quantum mechanics as an example of a scientific theory for which the fundamental interpretative hypothesis is well supported. From a measure-theoretic point of view, quantum mechanics is sometimes seen as anomalous, because of the influence exercised on its probabilistic predictions by the selection of measurements by observers, and because its various potential predictions, before a measurement is selected, do not find simple expression in terms of a single probability measure. From our game-theoretic point of view, however, these features are prototypical rather than anomalous. No scientific theory can support probabilistic predictions without protocols for the interface between the phenomenon being predicted and the various observers, controllers, and other external agents who work to bring and keep the phenomenon into relation with the theory.

| SOURCE OF THE PRICES | Irrelevant | Working Hypothesis | Believed | Well Supported |
|---|---|---|---|---|
| Scientific Theory | | | | Quantum mechanics |
| Observed Regularities | | Hypothesis testing | Statistical modeling and estimation | |
| Personal Choices | Neosubjective probability | Decision analysis | Weather forecasting | |
| Market | Hedging | Testing the EMH | Inference based on the EMH | |

STATUS OF THE FUNDAMENTAL INTERPRETATIVE HYPOTHESIS

***Fig. 1.6*** Some typical ways of using and interpreting a probability game, arrayed in two dimensions. (Here EMH is an acronym for the efficient-market hypothesis.)

Statistical modeling, testing, and estimation, as practiced across the natural and social sciences, is represented in Figure 1.6 in the row labeled "observed regularities". We speak of regularities rather than frequencies because the empirical information on which statistical models are based is usually too complex to be summarized by frequencies across identical or exchangeable circumstances.

As we have already noted, the fundamental interpretative hypothesis is irrelevant to the neosubjectivist conception of probability, because a person has no obligation to take any stance concerning whether his or her subjective probabilities and prices satisfy the hypothesis. On the other hand, an individual might conjecture that his or her probabilities and prices do satisfy the hypothesis, with confidence ranging from "working hypothesis" to "well supported". The probabilities used in decision analysis and weather forecasting can fall anywhere in this range. We must also consider another dimension, not indicated in the figure: With respect to whose knowledge is the fundamental interpretative hypothesis asserted? An individual might peers odds that he or she is not willing to offer to more knowledgeable observers.

Finally, the bottom row of Figure 1.6 lists some uses of probability games in finance, a topic to which we will turn shortly.

## Folk Stochasticism

In our listing of different ways probability theory can be used, we have not talked about using it to study stochastic mechanisms that generate phenomena in the world. Although quite popular, this way of talking is not encouraged by our framework.

What is a stochastic mechanism? What does it mean to suppose that a phenomenon, say the weather at a particular time and place, is generated by chance according to a particular probability measure? Scientists and statisticians who use probability theory often answer this question with a self-consciously outlandish metaphor: A demigod tosses a coin or draws from a deck of cards to decide what the weather will be, and our job is to discover the bias of the coin or the proportions of different types of cards in the deck (see, e.g., [23], p. 5).

In *Realism and the Aim of Science*, Karl Popper argued that objective probabilities should be understood as the *propensities* of certain physical systems to produce certain results. Research workers who speak of stochastic mechanisms sometimes appeal to the philosophical literature on propensities, but more often they simply assume that the measure-theoretic framework authorizes their way of talking. It authorizes us to use probability measures to model the world, and what can a probability measure model other than a stochastic mechanism—something like a roulette wheel that produces random results?

The idea of a probability game encourages a somewhat different understanding. Because the player who determines the outcome in a probability game does not necessarily do so by tossing a coin or drawing a card, we can get started without a complete probability measure, such as might be defined by a biased coin or a deck of cards. So we can accommodate the idea that the phenomenon we are modeling might have only limited regularities, which permit the pricing of only some of its uncertainties.

The metaphor in which the flow of events is determined by chance drives statisticians to hypothesize full probability measures for the phenomena they study and to make these measures yet more extensive and complicated whenever their details are contradicted by empirical data. In contrast, our metaphor, in which outcomes are determined arbitrarily within constraints imposed by certain prices, encourages a minimalist philosophy. We may put forward only prices we consider well justified, and we may react to empirical refutation by withdrawing some of these prices rather than adding more.

We do, however, use some of the language associated with the folk stochasticism we otherwise avoid. For example, we sometimes say that a phenomenon is governed by a probability measure or by some more restrained set of prices. This works in our framework, because government only sets limits or general directions; it does not determine all details. In our games, Reality is governed in this sense by the prices announced by Forecaster: these prices set boundaries that Reality must respect in order to avoid allowing Skeptic to become rich. In Chapter 14 we explain what it means for Reality to be governed in this sense by a stochastic differential equation.

## 1.5 GAME-THEORETIC PROBABILITY IN FINANCE

Our study of finance theory in Part II is a case study of our minimalist philosophy of probability modeling. Finance is a particularly promising field for such a case study, because it starts with a copious supply of prices—market prices for stocks, bonds, futures, and other financial securities—with which we may be able to do something without hypothesizing additional prices based on observed regularities or theory.

We explore two distinct paths. The path along which we spend the most time takes us into the pricing of options. Along the other path, we investigate the hypothesis that market prices are efficient, in the sense that an investor cannot become very rich relative to the market without risking bankruptcy. This hypothesis is widely used in the existing literature, but always in combination with stochastic assumptions. We show that these assumptions are not always needed. For example, we show that market efficiency alone can justify the advice to hold the market portfolio.

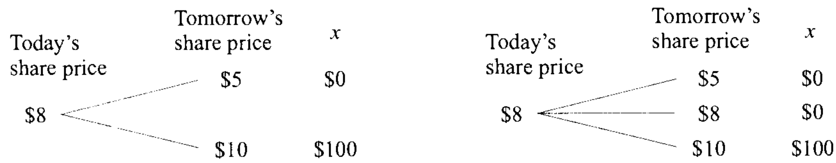
We conclude this introductory chapter with a brief preview of our approach to option pricing and with some comments about how our framework handles continuous time. A more extensive introduction to Part II is provided by Chapter 9.

### The Difficulty in Pricing Options

The worldwide market in derivative financial securities has grown explosively in recent years. The total nominal value of transactions in this market now exceeds the total value of the goods and services the world produces. Many of these transactions are in organized exchanges, where prices for standardized derivatives are determined by supply and demand. A larger volume of transactions, however, is in over-the-counter derivatives, purchased directly by individuals and corporations from investment banks and other financial intermediaries. These transactions typically involve hedging by both parties. The individual or corporation buys the derivative (a future payoff that depends, for example, on future stock or bond prices or on future interest or currency exchange rates) in order to hedge a risk arising in the course of their business. The issuer of the derivative, say an investment banker, buys and sells other financial instruments in order to hedge the risk acquired by selling the derivative. The cost of the banker's hedging determines the price of the derivative.

The bulk of the derivatives business is in futures, forwards, and swaps, whose payoffs depend linearly on the future market value of existing securities or currencies. These derivatives are usually hedged without considerations of probability [154]. But there is also a substantial market in options, whose payoffs depend nonlinearly on future prices. An option must be hedged dynamically, over the period leading up to its maturity, and according to established theory, the success of such hedging depends on stochastic assumptions. (See [128], p. xii, for some recent statistics on the total sales of different types of derivatives.)

For readers not yet familiar with options, the artificial examples in Figure 1.7 may be helpful. In both examples, we consider a stock that today sells for \$8 a share and tomorrow will either (1) go down in price to \$5, (2) go up in price to \$10, or (3) (in Example 2) stay unchanged in price. Suppose you want to purchase an option to buy 50 shares tomorrow at today's price of \$8. If you buy this option and the price goes up, you will buy the stock at \$8 and resell it at \$10, netting \$2 per share, or \$100. What price should you pay today for the option? What is the value today of a payoff $x$ that takes the value \$100 if the price of the stock goes up and the value \$0 otherwise? As explained in the caption to the figure, $x$ is worth \$60 in Example 1, for this price can be hedged exactly. In Example 2, however, no price for $x$ can be hedged exactly. The option in Example 1 can be priced because its payoff is actually a linear




**Example 1** **Example 2**

***Fig. 1.7*** The price of a share is now \$8. In Example 1, we assume that it will go up to \$10 or down to \$5 tomorrow. In Example 2, its price is also permitted to stay unchanged. In both cases, we are interested in the value today of the derivative $x$. In Example 1, $x$ has a definite value: $\mathbb{E}\,x = \$60$. This price for $x$ can be hedged exactly by buying 20 shares of the stock. If the stock down from \$8 to \$5, the loss of \$3 per share wipes out the \$60, but if it goes up to \$10, the gain of \$2 per share is just enough to provide the additional \$40 needed to provide $x$'s \$100 payoff. In Example 2, no price for $x$ can be hedged exactly. Instead we have $\overline{\mathbb{E}}\,x = \$60$ and $\underline{\mathbb{E}}\,x = \$0$. We should emphasize again that both examples are unrealistic. In a real financial market there is a whole range of possibilities—not just two or three possibilities—for how the price of a security can change over a single trading period.

function of the stock price. When there are only two possible values for a stock price, any function of that price is linear and hence can be hedged. In Example 2, where the stock price has three possible values, the payoff of the option is nonlinear. In real stock markets, there is a whole range of possible values for the price of a stock at some future time, and hence there are many nonlinear derivatives that cannot be priced by hedging in the stock itself without additional assumptions.

A range of possible values can be obtained by a sequence of binary branchings. This fact can be combined with the idea of dynamic hedging, as in Figure 1.8, to provide a misleadingly simple solution to our problem. The solution is so simple that we might be tempted to believe in some imaginary shadow market, speedier and more liquid than the real market, where changes in stock prices really are binary but produce the less restricted changes seen in the slower moving real market. Unfortunately, there is no traction in this idea, for we can hedge only in real markets. In real stock markets, many different price changes are possible over the time periods during which we hold stock, and so we can never hedge exactly. The best we can do is hedge in a way that works on average, counting on the errors to average out. This is why probabilistic assumptions are needed.

The probabilistic models most widely used for option pricing are usually formulated, for mathematical tractability, in continuous time. These models include the celebrated Black-Scholes model, as well as models that permit jumps. As it turns out, binomial trees, although unrealistic as models of the market, can serve as computationally useful approximations to these widely used (although perhaps equally unrealistic, alas) continuous-time models. This point, first demonstrated in the late 1970s [66, 67, 256], has made binomial trees a standard topic in textbooks on option pricing.

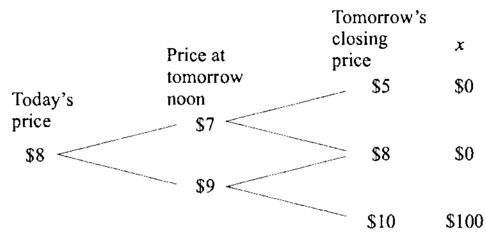



***Fig. 1.8*** In this example, $\mathbb{E}\,x = \$25$. To hedge this price, we first buy 25 shares of the stock today. We adjust this hedge at noon tomorrow, either by selling the 25 shares (if the price has gone down to \$7) or by buying another 25 shares (if the price has gone up to \$8).

### Making More Use of the Market

The most common probability model for option pricing in continuous time, the Black-Scholes model, assumes that the underlying stock price follows a geometric Brownian motion. Under this assumption, options can be priced by a formula—the Black-Scholes formula—that contains a parameter representing the volatility of the stock price; the value of this parameter is usually estimated from past fluctuations. The assumption of geometric Brownian motion can be interpreted from our thoroughly game-theoretic point of view (Chapter 14). But if we are willing to make more use of the market, we can instead eliminate it (Chapters 10–13). The simplest options on some stocks now trade in sufficient volume that their prices are determined by supply and demand rather than by the Black-Scholes formula. We propose to rely on this trend, by having the market itself price one type of option, with a range of maturity dates. If this traded option pays a smooth and strictly convex function of the stock price at maturity, then other derivatives can be priced using the Black-Scholes formula, provided that we reinterpret the parameter in the formula and determine its value from the price of the traded option. Instead of assuming that the prices of the stock and the traded option are governed by some stochastic model, we assume only certain limits on the fluctuation of these prices. Our market approach also extends to the Poisson model for jumps (§12.3).

### Probability Games in Continuous Time

Our discussion of option pricing in Part II involves an issue that is important both for our treatment of probability and for our treatment of finance: how can the game-theoretic framework accommodate continuous time? Measure theory's claim to serve as a foundation for probability has been based in part on its ability to deal with continuous time. In order to compete as a mathematical theory, our game-theoretic framework must also meet this challenge.

It is not immediately clear how to make sense of the idea of a game in which two players alternate moves continuously. A real number does not have an immediate predecessor or an immediate successor, and hence we cannot divide a continuum of time into points where Skeptic moves and immediately following points where World moves. Fortunately, we now have at our disposal a rigorous approach to continuous mathematics—nonstandard analysis—that does allow us to think of continuous time as being composed of discrete infinitesimal steps, each with an immediate predecessor and an immediate successor. First introduced by Abraham Robinson in the 1960s, long after the measure-theoretic framework for probability was established, nonstandard analysis is still unfamiliar and even intimidating for many applied mathematicians. But it provides a ready framework for putting our probability games into continuous time, with the great advantage that it allows a very clear understanding of how the infinite depends on the finite.

In Chapter 10, where we introduce our market approach to pricing options, we work in discrete time, just as real hedging does. Instead of obtaining an exact price for an option, we obtain upper and lower prices, both approximated by an expression similar to the familiar Black-Scholes formula. The accuracy of the approximation can be bounded in terms of the jaggedness of the market prices of the underlying security and the traded derivative. All this is very realistic but also unattractive and hard to follow because the approximations are crude, messy, and often arbitrary. In Chapter 11, we give a nonstandard version of the same theory. The nonstandard version, as it turns out, is simple and transparent. Moreover, the nonstandard version clearly says nothing that is not already in the discrete version, because it follows from the discrete version by the *transfer principle*, a general principle of nonstandard analysis that sometimes allows one to move between nonstandard and standard statements [136].

Some readers will see the need to appeal to nonstandard analysis as a shortcoming of our framework. There are unexpected benefits, however, in the clarity with which the transfer principle allows us to analyze the relation between discrete-time and continuous-time results. Although the discrete theory of Chapter 10 is very crude, its ability to calibrate the practical accuracy of our new purely game-theoretic Black-Scholes method goes well beyond what has been achieved by discrete-time analyses of the stochastic Black-Scholes method.

After introducing our approach to continuous time in Chapter 11, we use it to elaborate and extend our methods for option pricing (Chapters 12–13) and to give a general game-theoretic account of diffusion processes (Chapter 14), without working through corresponding discrete theory. This is appropriate, because the discrete theory will depend on the details of particular problems where the ideas are put to use. Discrete theory should be developed, however, in conjunction with any effort to put these ideas into practice. In our view, discrete theory should always be developed when continuous-time models are used, so that the accuracy of the continuous-time results can be studied quantitatively.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# *Part I*

# *Probability without Measure*

We turn now to consider what our game-theoretic framework does for probability. In Chapter 2, we show how it accommodates and extends the various viewpoints on the foundations and meaning of probability that have developed over the past three centuries. In Chapters 3–5, we formulate and prove game-theoretic versions of the classical strong limit theorems—the strong law of large numbers and the law of the iterated logarithm. In Chapters 6 and 7 we do the same for the classical weak limit theorems—the weak law of large numbers and the central limit theorem.

Our historical review in Chapter 2 ranges from Pascal and Fermat in the seventeenth century to von Mises, Kolmogorov, and de Finetti in the twentieth century. We emphasize the emergence of measure theory as a foundation for probability, Kolmogorov's explanation of how measure-theoretic probability can be related to the world, and the development of the hypothesis of the impossibility of a gambling system. This review reveals that our framework combines elements used in the past by a variety of authors, whose viewpoints are often seen as sharply divergent. The keys to this catholicity are (1) our sharp distinction between the idea of pricing by hedging and the hypothesis of the impossibility of a gambling strategy and (2) our flexibility with regard to the latter.

Using the simple case of bounded variables, Chapter 3 paints a reasonably full picture of how the game-theoretic framework handles strong laws. Chapter 4, which deals with Kolmogorov's strong law, and Chapter 5, which deals with the law of the iterated logarithm, confirm that the approach extends to more challenging examples.

Chapter 6 introduces our approach to the weak laws. In this chapter, we prove the central limit theorem for the simplest example: the fair coin. This simple setting permits a clear view of the martingale method of proof that we use repeatedly in later chapters, for central limit theorems and for option pricing. The method begins with a conjectured price as a function of time and an underlying process, verifies that the conjectured price satisfies a parabolic partial differential equation (the heat equation for the central limit theorem; most commonly the Black-Scholes equation for option pricing), expands the conjectured price in a Taylor's series, and uses the differential equation to eliminate some of the leading terms of the Taylor's series, so that the remaining leading terms reveal the price to be an approximate martingale as a function of time.

When we toss a fair coin, coding heads as 1 and tails as $-1$, we know exactly how far the result of each toss will fall from the average value of 0. We can relax this certainty about the magnitude of each deviation by demanding only a variance for each deviation—a price for its square—and still prove a central limit theorem. At the end of Chapter 6, we show that something can be done with even less: we can obtain interesting upper prices for the size of the average deviation beginning only with upper bounds for each deviation. This idea leads us into parabolic potential theory—the same mathematics we will use in Part II to price American options.

Chapter 6 is the most essential chapter in Part I for readers primarily interested in Part II. Those who seek a fuller understanding of the game-theoretic central limit theorem will also be interested in Chapter 7, where we formulate and prove a game-theoretic version of Lindeberg's central limit theorem. This theorem is more abstract than the main theorems of the preceding chapters; it is a statement about an arbitrary martingale in a symmetric probability game.

In Chapter 8, we step back and take a broader look at our framework. We verify that the game-theoretic results derived in Chapters 3 through 7 imply their measure-theoretic counterparts, we compare the game-theoretic and measure-theoretic frameworks as generalizations of coin tossing, and we review some general properties of game-theoretic probability.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# 2

# *The Game-Theoretic Framework in Historical Context*

The mathematical and philosophical foundations of probability have been debated ever since Blaise Pascal and Pierre de Fermat exchanged their letters in the seventeenth century. Different authors have shown how probability can be understood in terms of price, belief, frequency, measure, and algorithmic complexity. In this chapter, we review the historical development of these different ways of understanding probability, with a view to placing our own framework in historical context.

We begin with Pascal and Fermat's discussion of the problem of points and the subsequent development of their ideas into a theory unifying the belief and frequency aspects of probability. We then recount the emergence of measure theory and review Andrei Kolmogorov's definitive formulation of the measure-theoretic framework. After listing Kolmogorov's axioms and definitions, we compare his explanation of how they are to be used in practice with the earlier thinking of Antoine Augustin Cournot and the later attitudes of Joseph L. Doob and other mathematical probabilists. We then review Richard von Mises's collectives, Kolmogorov's algorithmic complexity, Jean Ville's martingales, A. Philip Dawid's prequential principle, and Bruno de Finetti's neosubjectivism. We also examine the history of the idea of the impossibility of a gambling system.

The authors whose work we review in this chapter often disagreed sharply with each other, but the game-theoretic framework borrows from them all. Our dual emphasis on the coherence of pricing and the hypothesis of the impossibility of a gambling system is in a tradition that goes back to Cournot, and our placement of the hypothesis of the impossibility of a gambling system outside the mathematics of probability, in an interpretative role, makes our viewpoint remarkably compatible with both subjective and objective interpretations of probability.

## 2.1 PROBABILITY BEFORE KOLMOGOROV

Our story about the foundations of probability before Kolmogorov begins with Pascal and Fermat, extends through the invention of probability by Jacob Bernoulli, and locates the beginnings of measure theory's dominance in the late nineteenth century.

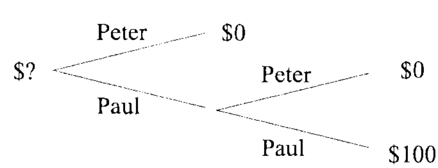
### The Precursors: Pascal and Fermat

The beginning of mathematical probability is often dated from the correspondence in the summer of 1654 between two French mathematicians, the Parisian Blaise Pascal and the Toulousian Pierre de Fermat. One of the problems they discussed was the problem of points—the problem of dividing the stakes when a game is cut short before any of the players has enough points to win. Figure 2.1 poses a very simple example of the problem. Here the stakes total \$100, and there are only two players, Peter and Paul, who lack one and two points, respectively. If play must be halted and the stakes divided, how much should Peter get and how much should Paul get?

Blaise Pascal (1623–1662), as imagined by the nineteenth-century artist Hippolyte Flandrin.

Pascal began with the principle that it is fair for two players to contend on equal terms. Equal terms means, *inter alia*, that the players make the same contribution to the total stakes. If Paul contributes $\$a$, his opponent Peter must also contribute $\$a$, and so the total stakes, which go to the winner, will be $\$2a$. Thus it is fair for Paul to pay $\$a$ for an

***Fig. 2.1*** Paul wins \$100 if he wins both of the next two rounds.



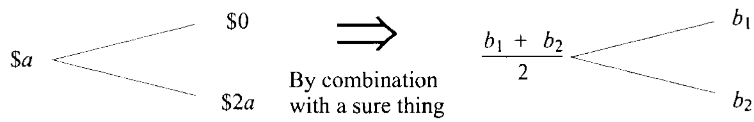



***Fig. 2.2*** Contending on equal terms.

opportunity to get $\$2a$ if he defeats Peter and \$0 if he loses to Peter, as indicated on the left of Figure 2.2.

Suppose now that we propose to Paul that he receive $b_2$ if he defeats Peter and $b_1$ if he loses to Peter, where $b_1 < b_2$. What price can we fairly charge Paul for this opportunity? Pascal answered this question by recognizing that the payoff can be reproduced by combining two payoffs whose fair price is already established:

- Paul gets $b_1$ no matter what happens. It is surely fair to charge Paul $b_1$ for this.
- In addition, Paul gets $b_2 - b_1$ if he defeats Peter and 0 if he loses to him. By the preceding argument, it is fair to charge Paul $(b_2 - b_1)/2$ for this.

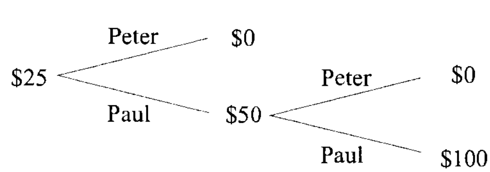
So the fair price to charge Paul is $b_1 + (b_2 - b_1)/2$, or $(b_1 + b_2)/2$, as indicated on the right of Figure 2.2.

Once the price $(b_1 + b_2)/2$ is established, we can solve the problem of points recursively. Figure 2.3 displays the solution for the example of Figure 2.1. If Paul wins the first round, he will be on even terms with Peter, and so his position will be worth \$50. So his present position, where he is contending for \$0 or \$50, is worth \$25.

In a pamphlet he published in 1657 to explain Pascal and Fermat's ideas, the Dutch mathematician Christiaan Huygens (1629–1695) slightly simplified Pascal's justification for the price $(b_1 + b_2)/2$. As Huygens pointed out, the players would be contending on equal terms if both paid this much, with the understanding that the winner would get $b_2$ and the loser would get $b_1$.

In his response to Pascal, Fermat proposed an alternative combinatorial method. Imagine, he said, that the players continue for fixed number of rounds even if one player gains enough points to win before the last round is played, and count the number of combinations that favor each player. Because each combination has an

***Fig. 2.3*** Paul's present position is worth \$25.



equal chance of happening, the stakes should be divided in this proportion. In our example, there are four combinations:

> Peter wins the first round; Peter wins the second.
> Peter wins the first round; Paul wins the second.
> Paul wins the first round; Peter wins the second.
> Paul wins the first round; Paul wins the second.

The first three win the game for Peter, and only the fourth wins the game for Paul. So Peter should get $75 and Paul should get $25.

Pascal acknowledged the validity of the combinatorial method, and his *Traité du triangle arithmétique*, which he shared with Fermat, was as concerned with combinations as with games of chance. Subsequent authors, including Jacob Bernoulli, Abraham De Moivre, Pierre de Montmort, and Nicholas Bernoulli, freely used both Fermat's method (the enumeration of combinations) and Pascal's method (recursive determination of fair price) to find odds in games of chance. In time, the combinatorial method came to be seen as more fundamental.

In retrospect, the combinatorial method can be seen as a precursor of measure-theoretic probability, just as Pascal's method can be seen as a precursor of our framework. Moreover, we share the eclectic attitude of the early authors: each approach has its advantages in different problems. Although we believe that our game-theoretic framework provides the best general understanding of probability, we have no desire to give up the mathematical tools that have developed within measure theory over the past century, and we believe that measure theory is more helpful than game theory for many problems and topics in probability. The basic ideas of ergodic theory, for example, seem to be essentially measure-theoretic.

The best account of the work of Pascal is provided by A. W. F. Edwards; see also [143, 280, 303]. English translations of the letters between Pascal and Fermat are provided in [79] and [290]; the originals can be found in their respective collected works. Huygens's pamphlet on games of chance was originally published in Latin and then in Dutch. The Dutch version, accompanied by a French translation, is in his collected works, along with his other writings on probability. The Latin version was reproduced in Jacob Bernoulli's *Ars Conjectandi*. Early English translations were published by Arbuthnot and Browne.

## Belief and Frequency

Pascal was not insensitive to the ambition of applying his ideas about games of chance to larger issues. He made a famous argument for wagering on the existence of God, and the *Port Royal Logic*, written by his friends at the Port Royal convent, argued for apportioning belief to the frequency with which events happen [8, 141]. But it was left to Jacob Bernoulli to clarify how price can serve as a foundation for numerical probability and how belief and frequency can be united in this concept.

As Bernoulli explained in his *Ars Conjectandi*, published posthumously in 1713, a numerical probability is a degree of certainty, which can be merited by an argument just as a share of the stakes in a game of chance can be merited by a position in that game. As a degree of certainty, probability is subjective. Any imperfect certainty must be subjective, for all things are known to God with complete certainty. On the other hand, a probability is not necessarily known to us. Sometimes we must make extensive observations in order to learn its magnitude, and in this respect it has an objective reality—a meaning that goes beyond any particular person's opinion. This uneasy balance between the objective and the subjective is encapsulated in Bernoulli's phrase "equally possible", which played an important role in his work and remained central to the foundations of probability for two centuries thereafter. We begin by counting the equally possible cases. The probability of an event is the proportion of these cases that favor it.

Bernoulli's greatest achievement was to represent within probability theory the idea that a probability can be learned from observation. Consider a game in which $r$ of the equally possible cases lead to success and $s$ to failure. Bernoulli's theorem says that by playing sufficiently many rounds of the game, we can be as certain as we like that the numbers of successes and failures we observe, say $y$ and $z$, will be in approximately the same ratio as $r$ and $s$. Christened the *law of large numbers* by Poisson, Bernoulli's theorem is now usually formulated as a comparison of probability and frequency. If we write $N$ for the total number of rounds played and $p$ for the probability of success, so that $N = y + z$, $p = r/(r + s)$, and the observed frequency of success is $y/N$, then the theorem says that for any $\epsilon > 0$ and $\delta > 0$,

$$\mathbb{P}\left\{\left|\frac{y}{N} - p\right| < \epsilon\right\} > 1 - \delta \tag{2.1}$$

when $N$ is sufficiently large, where $\mathbb{P}\,E$ denotes the probability of $E$. This justifies the use of $y/N$ as an estimate of $p$ when we do not know $r$ and $s$ and hence cannot calculate $p$ directly. Bernoulli derived an upper bound on how large the number of rounds $N$ needs to be in order to estimate $p$ with a given accuracy. He calculated, for example, that if $p = 0.6$, then 25,550 rounds is enough to make the odds at least a thousand to one that $y/N$ will fall between 0.58 and 0.62 ([15], p. 239, [3], p. 14).

In the 1730s, Abraham De Moivre (1667–1754) sharpened Bernoulli's theorem by estimating (rather than merely bounding) how large $N$ needs to be. As De Moivre showed,

$$\mathbb{P}\left\{\left|\frac{y}{N} - p\right| \leq c\sqrt{\frac{p(1-p)}{N}}\right\} \approx \frac{1}{\sqrt{2\pi}}\int_{-c}^{c} e^{-u^2/2}du \tag{2.2}$$

when $N$ is large. By evaluating the integral in (2.2), we find that only 6,700 rounds (not 25,550) are needed in order to estimate $p$ within 0.02 with thousand-to-one odds when $p$ is actually equal to 0.6.

Equal possibility, with all its internal tension between the subjective and the objective, remained the starting point for the French mathematician Pierre Simon Laplace (1749–1827). Laplace became the leading authority on mathematical probability by the 1770s, and his *Théorie analytique des probabilités* remained authoritative for over half a century after his death [44, 189, 191]. Laplace shared Bernoulli's subjectivism, not because he shared Bernoulli's belief in God but because he shared his own era's determinism: every event is objectively certain because it is determined by the past and the laws of physics. Yet Laplace also shared Bernoulli's interest in the empirical estimation of probabilities, and he created much of what we now recognize as mathematical statistics [295]. In particular, he generalized De Moivre's theorem by showing that if $x_1, \ldots, x_N$ are measurements resulting from $N$ independent trials of an experiment, with expected value $\mu$ and variance $\sigma^2$, then

$$\mathbb{P}\left\{\left|\frac{\sum_{n=1}^{N} x_n}{N} - \mu\right| \leq \frac{c\sigma}{\sqrt{N}}\right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-u^2/2} du. \tag{2.3}$$

De Moivre's theorem is the special case of (2.3) where $x_n$ is equal to either one (success) or zero (failure), so that $y = \sum_{n=1}^{N} x_n$. Laplace's proof is valid only for the case where $x_n$ has only a finite number of possible values, but the result is an instance of what we now call the central limit theorem for independent identically distributed random variables (see §7.4).

Equally possible cases provide the simplest instance of measure theory. In this sense, measure theory was already the mathematical foundation for probability in the eighteenth and nineteenth centuries. But as the nineteenth century wore on, this foundation became increasingly untenable. It faced two philosophical problems. First, the traditional problem of God's omniscience, now transformed into the problem of determinism. If God has foreseen what will happen, then how can a different outcome be possible? Or if the future can be predicted according to nature's laws from the present, then how can a different future be possible? Second, the increasing demands of empiricism, which now required one to question the very ideas of possibility and probability. If equal possibility can be discerned only by observing equal frequency, then what is probability but mere frequency?

Following Laplace's death, his French successors began to speak of different kinds of probability. Siméon-Denis Poisson (1781–1840) distinguished between *subjective* and *objective* probability. Antoine Augustin Cournot (1801–1877) spoke of *probability* and *chance*. Whereas Laplace had explained determinism by imagining a superior intelligence who could observe the present fully and exactly and hence predict the future precisely according to nature's laws, Cournot suggested that even an intelligence who observed everything and knew all the laws of nature would not be able to dispense with probability; chances are the probabilities that would be adopted by such an intelligence [58, 59, 222].

Although Cournot's understanding of objective probability is now attracting renewed attention, it found little sympathy in the nineteenth century, which was increasingly concerned with the empirical grounding of our knowledge. This advancing empiricism led to an equation between probability and frequency, especially in England, where it was articulated by John Venn (1834–1923) and John Stuart Mill (1806–1873). This equation could only discourage interest in mathematical probability, for it cast doubt on the basic theorems of probability, beginning with Bernoulli's theorem. If probability is nothing but frequency, then there can be nothing but confusion in a complicated proof that the two are probably approximately equal.

Serious mathematical work on probability slowed to a standstill with Laplace's death, partly because he was too great and too difficult a giant to see past, but also because his subjectivist defense of probability against determinism did not convince an empirical age. The reigning determinism of the nineteenth century, represented in France by Claude Bernard (1813–1878) in the biological sciences and Auguste Comte (1798–1857) in the physical sciences, was resolutely uninterested in probability. This disinterest persisted even as the project of basing social science on statistical fact became increasingly popular [142, 252]. Social statisticians, like frequentist philosophers, agreed that probability is mere frequency and had little interest in any mathematical or philosophical theory of probability.

A renewal of wide interest in probability came only near the end of the nineteenth century, as it began to appear that it could play a fundamental role in physics. Statistical physics began in the middle of the century, with the work of James Clerk Maxwell, Lord Kelvin, and Ludwig Boltzmann. At first, probability ideas played no greater role here than in social statistics; when physicists spoke of probabilities, they were thinking merely of frequencies over time. But as the subject developed, and especially as the philosophical conundrums of statistical thermodynamics emerged, the mathematics and philosophy of probability came to seem more and more relevant. People began to feel, as many still feel today, that statistical thermodynamics is confused about the relation between time averages and phase-state averages, and they began to hope that a clearer foundation for probability itself might bring clarity to this whole area of physics. When the German mathematician David Hilbert (1862–1943) listed the most important open problems in mathematics, at the second international congress of mathematicians at Paris in 1900, he named probability as one of the subdisciplines of physics whose axiomatic foundations needed investigation [139, 152]. Hilbert wanted a clearer and more rigorous treatment of what Jan von Plato [318] has called modern probability—probability for continuous quantities in continuous time.

The strongest mathematical work in probability in the late nineteenth century was that of the Russian Pafnutii L. Chebyshev (1821–1894) and his student Andrei A. Markov (1856–1922). By the early twentieth century, interest in mathematical probability had revived in France, with the work of Henri Poincaré (1854–1912), a major figure in mathematical physics and the philosophy of science, Émile Borel (1871–1956), a pioneer in measure theory, and Paul Lévy (1886–1971), who pioneered the study of dependent random variables. Other leaders included Francesco Cantelli (1887–1972) in Italy, Hugo Steinhaus (1875–1966) in Lviv, and Sergei Bernstein (1880–1968), who studied in Paris before starting his career in Kharkiv.

### The Advent of Measure Theory

From the modern measure-theoretic point of view, the relevance of integration to probability theory is obvious. A variable is a function on a sample space, and the expected value of the variable is the integral of that function with respect to a probability measure on the sample space. But this is hindsight. In the nineteenth century, probability theory remained firmly anchored in the traditional framework of equally possible cases, and continuous probability was considered admissible only when it could be seen as a well-behaved extension of that framework. The development of the theory of integration, in contrast, was driven by the desire to integrate more and more badly behaved functions [151, 246]. Moreover, the theory of integration was concerned with functions of real or complex numbers, whereas probability was more abstract. It was only after the turn of the century, when the theory of measure and integration had been developed for abstract spaces, that it could be seen as a potential foundation for probability theory.

Major milestones in the increasingly general and abstract understanding of measure and integration were set by Bernhard Riemann, Camille Jordan, Émile Borel, Henri Lebesgue, Johann Radon, and Maurice Fréchet. In his *Habilitationsschrift* of 1854, Riemann developed the theory of integration for continuous functions in the modern sense—mappings from real numbers to real numbers not necessarily given by formulas. The Riemann integral is obtained as the limit of discrete approximations in which the domain of the function is divided into small intervals, rectangles, or other elementary sets, and the integral of the function over each elementary set is approximated from the volume of the elementary set and typical values of the function on the set. In 1892, Jordan reformulated Riemann's theory of integration on the basis of a theory of measure in the domain of the function being integrated; the small sets in the domain used to approximate the integral, instead of being intervals or rectangles, might now be more complex sets, which could nevertheless be approximated by finite unions of intervals or rectangles. In 1898, Borel extended Jordan's measure theory by using sets approximated by countably infinite rather than merely finite unions of elementary sets. Borel was concerned with the theory of complex analytic functions rather than the theory of integration. But his work inspired Lebesgue's development, in 1901, of an entirely novel theory of integration. In Lebesgue's theory, the sets going into an approximation of an integral are formed by grouping points together on the basis of the value of the function rather than their contiguity in the domain, thus liberating integration from any assumption about the continuity of the function. Lebesgue was still concerned with integration of functions on Euclidean spaces, but because his ideas required only a measure (not a topology) on the domain, others, including Radon in 1913 and Fréchet in 1915, soon extended his theory of measure and integration to more abstract spaces.

In the 1890s, when the new measure theory was being developed, there was no felt need within probability for improved theories of integration. But the idea of measure zero was becoming more important for probability, and this made the new theories of measure and integration appear increasingly relevant. To understand the growing importance of measure zero in probability, we may begin with Henri Poincaré's recurrence theorem of 1890. This theorem says that an isolated mechanical system consisting of three bodies will eventually return arbitrarily close to its initial state, provided only that this initial state is not exceptional. In order to make precise how few exceptional states there are, Poincaré proposed, as a convention, that we take the relative probability of two regions of phase space to be proportional to their volumes. He then showed that within any region of finite volume, the states for which the recurrence does not hold are contained in subregions whose volume is arbitrarily small, and hence these states have zero probability. This is the sense in which they are exceptional ([247], p. 71). The idea that an event could be possible but have zero probability was familiar to mathematicians from geometric probability: when we choose a point at random from a line segment or from a region of a higher-dimensional space, the probability of any particular point being chosen is zero, even though the choice is quite possible. But according to von Plato (1994, p. 90), Poincaré's article on the three-body problem marks the first application of the idea of measure zero to a mechanical system. It was followed by many more.

Poincaré's assumption about the probabilities for the initial state of a mechanical system was, he said, a convention. He did not use the word carelessly. He is remembered for an influential philosophy of science that stressed the conventional nature of scientific laws. He considered even Newton's law and Euclid's axioms conventional inasmuch as they are deliberately chosen by scientists for their helpfulness in organizing our experience. Not all his contemporaries shared his broad conventionalism, but his fellow French mathematicians did tend to share his caution about the meaningfulness of continuous probability.

Borel shared Poincaré's cautious attitude towards continuous probability [164, 245, 270]. For Borel, a concrete meaning could be given to the idea of choosing at random from a finite set and perhaps even from a denumerable set, but because of the nonconstructive character of the totality of real numbers, no meaning could be given to choosing at random a real number from an interval or a point from a region of the plane. But in 1909, Borel published a celebrated article that brought continuous measure theory into the very heart of probability. In this article, Borel investigated the behavior of the infinite sequence of successes and failures that would result were a game repeated indefinitely. Consider the simple case where the probability of success on each round is $1/2$ and the rounds are independent, and write $y_n$ for the number of successes in the first $n$ rounds. Borel claimed that with probability one there will be a number $N$ such that

Émile Borel (1871–1956), minister of the French navy from 1925 to 1940.

$$\left| \frac{y_n}{n} - \frac{1}{2} \right| \leq \frac{\ln(n/2)}{\sqrt{2n}} \tag{2.4}$$

for all $n$ larger than $N$. Since $\lim_{n\to\infty} \ln(n/2)/\sqrt{2n} = 0$, this implies that

$$\lim_{n\to\infty} \frac{y_n}{n} = \frac{1}{2} \tag{2.5}$$

with probability one. The latter statement, generalized from the case where the probability on each round is $1/2$ to the case of an arbitrary probability $p$, is now called *Borel's strong law of large numbers.* Bernoulli's theorem has been demoted; it is now merely a *weak* law of large numbers.

For the sake of historical accuracy, we hasten to add that Borel did not express his claim in the form (2.4). Instead, he asserted that there are sequences $\lambda_n$ such that $\lim_{n\to\infty} \lambda_n/\sqrt{n} = 0$ and the number of successes in $2n$ trials is eventually bounded between

$$n - \lambda_n\sqrt{n} \quad \text{and} \quad n + \lambda_n\sqrt{n}. \tag{2.6}$$

To fix ideas, he took $\lambda_n$ to be $\ln n$ (Borel 1909, p. 259). When we replace $\lambda_n$ by $\ln n$ and then $2n$ by $n$ in (2.6), we obtain bounds equivalent to (2.4). There were gaps in Borel's proof [104]; a complete proof was first given by Georg Faber (1877–1966) in 1910. Moreover, Borel did not investigate how slowly $\lambda_n$ may grow and hence how tightly the frequency of successes can be bounded; this question was settled by Khinchin's law of the iterated logarithm, which appeared in 1924 (see Chapter 5).

Borel insisted on a constructive and ultimately finitary interpretation of his results. As he explained, they mean merely that as $n$ increases, one may safely give indefinitely great odds that (2.4) will continue to hold no matter how much longer one continues to play. Yes, the relative frequency of successes converges to one-half with probability one. But this is only a manner of speaking. There is no infinite sequence that may converge or diverge, and "probability one" does not mean certainty. "Probability one" means that we may give indefinitely great odds as the number of rounds increases; "converges to 1/2" means that these odds are for the relative frequency remaining in some ever tightening bounds around 1/2.

But as Borel also understood and explained, his results can be expressed in terms of geometric probability. If we encode success as 1 and failure as 0, then the infinite sequence of 1s and 0s is the binary expansion of a real number from the closed interval $[0, 1]$, and the assumption that the rounds are independent with probability $1/2$ for 1 on each round corresponds to the uniform probability measure on this interval—the measure that gives each subinterval probability equal to its length. So if we imagine drawing a real number $x$ at random from the interval $[0, 1]$, and we write $y_n$ for the number of 1s in the first $n$ digits of $x$'s binary expansion, then Borel's strong law says that (2.5) holds with probability one. For Borel, this interpretation was mathematically instructive but conceptually inadmissible, simply because the idea of choosing a real number at random from $[0, 1]$ was inadmissible.

An important example of the impact of Borel's thinking is provided by Paul Lévy's *Calcul des probabilités*, published in 1925. In an appendix to the book, Lévy explains how Borel's example generalizes to a method for defining probabilities in an abstract space, such as a space of functions: partition the space into a denumerable collection of subsets, assign these subsets probabilities that add to one, then partition each of them in the same way, and so on indefinitely. In general, this does not produce probabilities for all interesting events. One cannot, for example, speak of the probability that a random function on $[0, 1]$ will never take the value $1/2$ (Lévy 1937, pp. 24–25). This limitation, for Borel and Lévy, was not an embarrassment; it was the result of the need for probabilities to have empirical meaning.

In the end, however, Borel's scruples proved less influential than the elegance of his mathematics. His colleagues were surprised and fascinated by his results—surprised because it seemed paradoxical that the binary expansions of nearly all real numbers should have equal numbers of zeros and ones in the limit; fascinated because it now seemed that measure theory was fundamental to probability. Author after author declared that Borel had reduced the study of infinite sequences of events to measure theory ([318], pp. 57–60). The way was paved for the free-wheeling equation of probability and measure that found its full expression in the work of Kolmogorov.

## 2.2 KOLMOGOROV'S MEASURE-THEORETIC FRAMEWORK

Andrei Kolmogorov (1903–1987), in his dacha at Komarovka.

Although measure and integration occupied center stage in probability theory after 1909, and the idea that measure could provide an axiomatic basis for probability was in the air in the 1920s, Hilbert's call for an axiomatization of probability was answered in a fully satisfying way only by Kolmogorov's famous monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung*, published in 1933. As Maurice Fréchet said in his opening address to the colloquium on probability theory at the University of Geneva in 1937 ([130], p. 54), it is not enough to have all the ideas in mind and recall them here and there; one must bring them together explicitly, make sure that the whole is sufficient, and take responsibility for saying that nothing further is needed in order to construct the theory. This is what Kolmogorov did.

From a purely mathematical point of view, Kolmogorov's framework was extremely simple—six axioms and a few definitions. Among mathematicians, its simplicity, clarity, and power made it the easy victor in the spirited debate on the foundations of probability that took place in the 1930s. Its champions included Maurice Fréchet (1878–1973) in France, Harald Cramér (1893–1985) in Sweden, and William Feller (1906–1970) and Joseph L. Doob (born 1910) in the United States. After World War II, it was the accepted wisdom. Exposited in countless textbooks, it is now familiar to all who study probability beyond the most elementary level.

In addition to Kolmogorov's axioms and definitions, we review his philosophical account of probability—his explanation of how these axioms and definitions can be used in practice. This explanation was as attuned to Kolmogorov's time as his mathematics was, but it seemed less important to the mathematical probabilists who followed him. For us, it is very important. It is the key to understanding how our fundamental interpretative hypothesis relates both to Kolmogorov's axioms and to the traditions that preceded him.

### Kolmogorov's Axioms and Definitions

Although Kolmogorov's axioms and definitions are now very familiar, they are also very concise, and we will take the time to repeat them. We do so in the order and form in which he gave them, except that we deviate slightly from his notation and terminology, in favor of usages that will be more familiar to most readers and will recur later in this book.

Because he wants to relate his ideas to the traditional framework of equally possible cases, Kolmogorov begins his exposition with the five of his six axioms that are relevant to finite spaces. These axioms concern a set $\Omega$, which we call the *sample space*, and a set $\mathcal{F}$ of subsets of $\Omega$, which we call *events*. With this notation, the five axioms run as follows:

1. $\mathcal{F}$ is a *field* of sets. (This means that whenever $\mathcal{F}$ contains $E$ and $F$ it also contains the union $E \cup F$, the intersection $E \cap F$, and the difference $E \setminus F$.)

2. $\mathcal{F}$ contains the set $\Omega$. (Together with Axiom 1, this says that $\mathcal{F}$ is an *algebra* of sets. When $\mathcal{F}$ is also closed under countably infinite intersections and unions, it is called a *$\sigma$-algebra*.)

3. To each set $E$ in $\mathcal{F}$ is assigned a nonnegative real number $\mathbb{P}\, E$. This number is called the *probability* of the event $E$.

4. $\mathbb{P}\, \Omega$ equals 1.

5. If $E$ and $F$ are disjoint ($E \cap F = \emptyset$, where $\emptyset$ is the empty set), then $\mathbb{P}[E \cup F] = \mathbb{P}\, E + \mathbb{P}\, F$.

With these axioms in place, Kolmogorov introduces the ideas of conditional probability, experiment, and independence.

- If $E$ and $F$ are events and $\mathbb{P}\, E > 0$, then the *conditional probability* of the event $F$ under the condition $E$, $\mathbb{P}[F \mid E]$, is given by

$$\mathbb{P}[F \mid E] := \frac{\mathbb{P}[E \cap F]}{\mathbb{P}\, E}. \tag{2.7}$$

- An *experiment* is a decomposition of $\Omega$—that is, a collection $E_1, \ldots, E_r$ of disjoint events whose union is $\Omega$.

- A collection of $n$ experiments, say $\mathcal{E}^i = \{E_1^i, \ldots, E_{r_i}^i\}$ for $i = 1, \ldots, n$, are *mutually independent* if

$$\mathbb{P}[F^1 \cap \cdots \cap F^n] = \mathbb{P}\, F^1 \cdots \mathbb{P}\, F^n$$

whenever $F^i$ is an element of $\mathcal{E}^i$ for $i = 1, \dots, n$.

These axioms and definitions are satisfied in the traditional theory, in which the equally possible cases constitute, in effect, the sample space. The novelty of (2.7) is that only the unconditional probabilities $\mathbb{P}\,E$ are primitive. In earlier treatments, both conditional and unconditional probabilities were either defined in terms of equally possible cases or else taken as primitive, and (2.7) was a theorem rather than a definition. For historical accuracy, we should note that in 1933 Kolmogorov actually wrote $P_A(B)$ for the probability of $B$ given $A$. This notation was widely used at the time; it goes back at least to an article by Felix Hausdorff in 1901. Kolmogorov switched to the now standard $P(B \mid A)$ in the second Russian edition of his monograph, published in 1974.

Finally, going beyond classical theory, Kolmogorov adds a sixth axiom, which is restrictive only when the sample space is infinite:

6. For a decreasing sequence of events $E_1 \supseteq E_2 \supseteq \cdots$ for which $\bigcap_{n=1}^{\infty} E_n = \emptyset$, $\lim_{n \to \infty} \mathbb{P}\,E_n = 0$.

A function $\mathbb{P}$ satisfying axioms 4, 5, and 6 is called a *probability measure* on the algebra $\mathcal{F}$. Any probability measure on an algebra $\mathcal{F}$ can be extended to a probability measure on the smallest $\sigma$-algebra containing $\mathcal{F}$, and the extension will be countably additive:

$$\mathbb{P} \bigcup_{n=1}^{\infty} E_n = \sum_{n=1}^{\infty} \mathbb{P}\,E_n$$

for every sequence $E_1, E_2, \dots$ of disjoint events in the $\sigma$-algebra (here we still write $\mathbb{P}$ for the extension).

So Kolmogorov now assumes that $\mathcal{F}$ is a $\sigma$-algebra. He then introduces the concepts of random variable, equivalence of random variables, expected value, and conditional expectation.

- A *random variable* is a function $x$ on $\Omega$ that is measurable with respect to the $\sigma$-algebra $\mathcal{F}$. (This means that for each interval $I$ of real numbers, the subset $\{\xi \mid x(\xi) \in I\}$ of $\Omega$ is in $\mathcal{F}$. When this is the case, we can define a probability measure on the real numbers by setting the probability for each interval $I$ equal to $\mathbb{P}\{\xi \mid x(\xi) \in I\}$ and extending to the smallest $\sigma$-algebra containing all intervals. This probability measure on the real numbers is called the *probability distribution* for the random variable.[1]

- Two random variables $x$ and $y$ are *equivalent* if the probability of the event $x \neq y$ is zero.

[1] When one speaks of a probability measure on the real numbers, it is always assumed that the $\sigma$-algebra is either the Borel $\sigma$-algebra or the Lebesgue $\sigma$-algebra. The Borel $\sigma$-algebra, whose elements are called Borel sets, is the smallest $\sigma$-algebra containing the intervals. The Lebesgue $\sigma$-algebra, whose elements are called Lebesgue sets, is the smallest $\sigma$-algebra containing both the Borel sets and any set contained in a Borel set that has measure zero with respect to the uniform measure. In practice, any probability measure on the real numbers is called a "probability distribution"; the requirement that it be constructed in the way specified is dropped.

- The *expected value* of a random variable $x$, $\mathbb{E}\,x$, is its integral with respect to the measure $\mathbb{P}$. If the integral does not exist, then we say that $x$ does not have an expected value. If $x$ does have an expected value, then we write $\mathbb{V}\,x$ for $\mathbb{E}[(x - \mathbb{E}\,x)^2]$, the *variance* of $x$.

- A *version of the conditional expectation* of a random variable $y$ given another random variable $x$ is any random variable $z$ that (a) is a function of $x$ and (b) has the same integral with respect to $\mathbb{P}$ as $y$ over every event defined in terms of $x$. (According to a theorem of measure theory now called the *Radon-Nikodym theorem*, at least one such random variable $z$ exists, and any two are equivalent. We may write $\mathbb{E}[y \mid x]$ for any one of them,[2] and we may write $\mathbb{E}[y \mid x = c]$ for its value when $x = c$.)

Conditional expectation generalizes conditional probability as defined by (2.7); if $y$ takes the value 1 on an event $B$ and 0 elsewhere, and the event $x = c$ has positive probability, then all versions of $\mathbb{E}[y \mid x]$ agree that $\mathbb{E}[y \mid x = c] = \mathbb{P}[B \mid x = c]$ (and this equality serves as the definition of conditional probability in general).

As a first step towards demonstrating that these axioms and definitions are a sufficient mathematical foundation for continuous probability, Kolmogorov showed that in order to define a joint probability distribution for an indexed set of random variables, $x_t$, where the index $t$ ranges over a set of arbitrary cardinality, it suffices to give mutually consistent joint probability distributions for all finite collections $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$. He also showed, in the *Grundbegriffe* and related articles, that the framework is adequate for fundamental results such as the strong law of large numbers.

Of course, the reduction of probability to measure did not immediately persuade everyone. Borel, who lived until 1956, and Lévy, who lived until 1971, always felt that the mathematical theory should incorporate the passage from the finite to the continuous that makes abstract probabilities meaningful and makes conditional probability primitive [36, 201, 202]. De Finetti argued for decades against countable additivity (Axiom 6) and the demotion of conditional probability [92]. Along with de Finetti, Alfréd Rényi carried the battle for conditional probability into the 1960s [257, 258]. There were mathematical as well as philosophical reasons to hesitate about the new framework. In practice, the conditional expectations that characterize a stochastic process are given directly, not derived from unconditional expectations, and we want them to have properties that are not guaranteed by Kolmogorov's axioms. These axioms do not even guarantee, for example, that the conditional probabilities $\mathbb{P}[B \mid x = c]$, for fixed $c$, form a probability measure. It was some years after the *Grundbegriffe* that topological conditions under which such regular conditional probabilities exist were formulated [262].

[2]Because our game-theoretic framework does not neglect what happens only on a set of measure zero, we avoid taking such liberties in this book. When we use a symbol for a variable in a protocol for a game, it always refers to a single particular variable, never to a class of variables every two members of which agree only almost surely.

### How Kolmogorov Related His Axioms to the World

Kolmogorov was sympathetic with the frequentist interpretation of probability developed and championed by Richard von Mises. As we will explain in the next section, probability was not primitive for von Mises; it was derived from the idea of a *collective*—a sequence of actual outcomes in which particular outcomes have limiting frequencies that remain invariant when we shift attention to subsequences defined without foreknowledge. Kolmogorov did not incorporate von Mises's philosophical ideas into his mathematics; instead he axiomatized the mathematics of probability directly. But he regarded his axioms as axioms for a frequentist concept of probability, and he tried to explain why.

His explanation is brief enough to quote in full. We use the translation by Nathan Morrison ([178], 1950, pp. 3–4) except that we continue to deviate modestly from Kolmogorov's notation.

> We apply the theory of probability to the actual world of experiments in the following manner:
>
> 1. There is assumed a complex of conditions, $\mathcal{C}$, which allows of any number of repetitions.
> 2. We study a definite set of events which could take place as a result of the establishment of the conditions $\mathcal{C}$. In individual cases where the conditions are realized, the events occur, generally, in different ways. Let $\Omega$ be the set of all possible variants $\omega_1, \omega_2, \ldots$ of the outcome of the given events. Some of these variants might in general not occur. We include in set $\Omega$ all the variants we regard *a priori* as possible.
> 3. If the variant of the events which has actually occurred upon realization of conditions $\mathcal{C}$ belongs to the set $E$ (defined in any way), then we say that the event $E$ has taken place.
> 4. Under certain conditions, which we shall not discuss here, we may assume that to an event $E$ which may or may not occur under conditions $\mathcal{C}$, is assigned a real number $\mathbb{P}\,E$ which has the following characteristics:
>    (a) One can be practically certain that if the complex of conditions $\mathcal{C}$ is repeated a large number of times, $n$, then if $y$ be the number of occurrences of event $E$, the ratio $y/n$ will differ very slightly from $\mathbb{P}\,E$.
>    (b) If $\mathbb{P}\,E$ is very small, one can be practically certain that when conditions $\mathcal{C}$ are realized only once, the event $E$ would not occur at all.

For us, the most striking part of this passage is its final point, 4(b), which is closely related to our fundamental interpretative hypothesis. Point 4(b) is the only grounding Kolmogorov gives to the meaning of probabilities when an experiment is conducted only once. So it would appear that for him, as for us (see §1.3), only probabilities very close to zero or one have any meaning in isolation. A probability close to zero means that the event almost certainly will not occur, and a probability close to one means that the event almost certainly will occur. But in the absence of repetition, a middling probability such as $1/3$ or $1/2$ has no particular meaning at all.

A. A. Cournot (1801–1877). The principle that events with zero probability cannot happen is Cournot's bridge from probability theory to the physical world.

In order to understand fully the significance of Kolmogorov's points 4(a) and 4(b), we need to recall their antecedents. The basic idea goes back to Jacob Bernoulli, who first argued that an event with very high probability is practically (or "morally") certain and then concluded that frequency is practically certain to approximate probability ([15], pp. 217, 226). Cournot brought Bernoulli's argument into the empiricist age by recasting practical certainty as physical certainty. It may be mathematically possible for a heavy cone to stand in equilibrium on its point, but such an event is physically impossible, Cournot argued, because of its very small or even zero probability. Similarly, it is physically impossible for the frequency of an event in a sufficiently long sequence of trials to differ substantially from the event's probability ([58], pp. 57, 106). This has been called *Cournot's bridge* from probability theory to the physical world ([5], p. 106; [222], p. 175).

Cournot's bridge was alive and well in the early twentieth century. It was expressed most clearly by Paul Lévy. In his 1925 book, Lévy explained that probability is founded on two principles: the principle of equally likely cases and the principle of the very unlikely event. The first principle underlies the mathematics of probability, but it produces only a subjective theory. The second principle, which says that a very unlikely event will not happen, provides the bridge from this subjective theory to the world of experience. It is only this bridge that allows us to treat a mathematical probability as a physical quantity, which we can measure by repeated observations just as objectively as we measure length or weight ([196], pp. 21–22, 29–30, 34). The importance of the principle of the very unlikely event to practice was also emphasized by the English statistician Ronald A. Fisher [125].

Kolmogorov's formulation of 4(a) and 4(b) echoes this tradition, but he does not acknowledge that 4(b) renders 4(a) superfluous as an interpretive assumption. Once 4(b) is accepted, 4(a) is a theorem, and the frequency aspect of probability, as important as it is, seems derivative rather than fundamental. Passing over this point in silence was Kolmogorov's way, perhaps, of expressing his own preference for an understanding of the relation between probability theory and the world that begins with the idea of frequency rather than the idea of practical impossibility.

## The Expulsion of Philosophy from Probability

Although Kolmogorov himself did not see his framework as a substitute for what von Mises had tried to do, many mathematicians in the late 1930s did see Kolmogorov's and von Mises's approaches as rival mathematical foundations for the frequentist conception of probability. This is explicit in Maurice Fréchet's contribution to the Geneva colloquium in 1937 and in Joseph L. Doob's debate with von Mises in Hanover, New Hampshire, in 1940 [98, 313, 315]. In these discussions, Fréchet and Doob point to the law of large numbers as one aspect of the correspondence between Kolmogorov's mathematics and what happens in the world, but neither of them mention any need for a principle such as Kolmogorov's 4(b) or the similar principles that had seemed so important to Cournot and Lévy.

For Doob, we need merely to "find a translation of the results of the formal calculus which makes them relevant to empirical practice" ([98], p. 208). Doob's solution was to set up a correspondence between subsets of $\Omega$ and events in the world, and to say that the measure assigned to a subset of $\Omega$ in the formal calculus corresponds to the probability of the event, which is found by the practitioner using "a judicious mixture of experiments with reason founded on theory and experience" ([98], p. 209). Whereas Lévy offered the principle of practical impossibility as a way of giving meaning to probability, Doob preferred to leave the meaning, numerical values, and validity of probabilities entirely up to the practitioner.

Joseph L. Doob, president of the American Mathematical Society, 1963–1964.

In the last analysis, Kolmogorov's axioms triumphed and endured because they made probability legitimate as mathematics, freeing mathematical probabilists from the need to worry about interpretation. Before Kolmogorov's *Grundbegriffe*, most mathematicians were suspicious of probability, because it seemed to be based on assumptions that could not be stated clearly within mathematics. After the *Grundbegriffe*, probability was mathematics, with fully respectable axioms and definitions, in the set-theoretic tradition championed by Hilbert. The intuitions remained, but now they provided only inspiration, not impediments to rigor. Other mathematicians could use Kolmogorov's framework whether or not they shared his frequentism. And once it was accepted by mathematicians, the framework also was accepted by most other scholars interested in probability, regardless of their philosophical leanings. Thorough neosubjectivists still share de Fintetti's [14, 186] reservations, but even many subjectivists now take Kolmogorov as their starting point [259].

## 2.3 REALIZED RANDOMNESS

Richard von Mises (1883–1957), in Berlin in 1930.

Game theory is a mathematical account of potentiality; it analyzes what players can do, not what they have done or will do. Hence the probabilities in our framework are measures of potentiality. But our thinking owes much to a twentieth-century line of work that treated probabilities as aspects of a concrete reality. This work was begun between the two world wars by Richard von Mises, who sought to explain what it means for a realized sequence to be random. Vigorously debated in the 1930s, von Mises's ideas inspired Jean Ville's work on martingales, an important ingredient in our framework. Once the adequacy of Kolmogorov's axioms as a foundation for mathematical work in probability became clear, most mathematicians lost interest in trying to define randomness mathematically, but the project was revived in the 1960s by Kolmogorov and others, who developed it into a theory of algorithmic complexity.

### Von Mises's Collectives

Von Mises, who served in the Austro-Hungarian army in World War I, was educated in Vienna and taught in Strasbourg, Dresden, and Berlin before fleeing to Turkey when the Nazis came to power in 1933. In 1939, he emigrated to the United States, becoming a professor of mathematics at Harvard [127]. He explained his concept of a collective in an article in 1919 and then in a book, *Wahrscheinlichkeit, Statistik, und Wahrheit*, first published in 1928. His fundamental intuition was that a probability is an invariant limiting frequency—a frequency that (a) is obtained in the limit as one looks at longer and longer initial segments of an infinite sequence and (b) remains the same when one replaces the sequence with a subsequence.

A *collective* (*Kollektiv* in German), according to von Mises, is an actual sequence for which such invariant limiting frequencies exist. Consider, for example, an infinite sequence of 0s and 1s that begins like this:

$$0\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,0\,0\,1\,1\,1\,1\,1\,0\ldots.$$

As we move through the sequence, the cumulative relative frequency of 1s changes:

$$\frac{0}{1}\;\frac{0}{2}\;\frac{1}{3}\;\frac{1}{4}\;\frac{1}{5}\;\frac{2}{6}\;\frac{3}{7}\;\frac{3}{8}\;\frac{3}{9}\;\frac{4}{10}\;\frac{4}{11}\;\frac{5}{12}\;\frac{5}{13}\;\frac{5}{14}\;\frac{6}{15}\;\frac{7}{16}\;\frac{8}{17}\;\frac{9}{18}\;\frac{10}{19}\;\frac{10}{20}\ldots.$$

Von Mises's definition requires that this cumulative relative frequency converge to a constant, say $p$. And it must converge to the same constant $p$ in any subsequence, such as the subsequence consisting of every other entry:

$$0\,1\,0\,1\,0\,0\,0\,1\,1\,1\ldots$$

or the subsequence consisting of entries preceded by a 1:

$$0\,1\,0\,0\,1\,1\,1\,1\,0\ldots.$$

Thus a gambler who bets on 1s, say, cannot improve his performance by selecting only some trials on which to bet. In order to qualify as a collective, the sequence must be too irregular to permit any such strategy to succeed.

In von Mises's philosophy, the collective comes first, and then probability. The collective can consist of zeros and ones, as in our example, or its entries can be drawn from any other sample space. In any case, an event can have a probability only if it is situated in a collective, in which case its probability is that collective's limiting frequency for the event. And the rules of probability derive from the properties of collectives. Probabilities add, for example, because frequencies add.

Von Mises presented his ideas in an engaging, confident, but nonrigorous way. He did not explain how subsequences are to be selected, and so he was not in a position to consider the mathematical problem of proving that collectives—infinite sequences for which all permitted subsequences have the same frequencies—exist. But these challenges were taken up by others.

The first authors to clarify how subsequences can be selected were Arthur H. Copeland, Sr., Karl Popper, and Hans Reichenbach [57, 249, 254]. Their ideas about how to select a subsequence from $x_1, x_2, \ldots$ all boil down to fixing a natural number $r$ and a sequence $a_1, \ldots, a_r$ of elements of the sample space and then including $x_n$ in the subsequence if it is immediately preceded in $x_1, x_2, \ldots$ by $a_1, \ldots, a_r$. They suggested that $x_1, x_2, \ldots$ be considered a collective if the same limiting frequencies are obtained in every infinite subsequence selected in this way, but this was unacceptable to von Mises, because it may allow some sequences defined by mathematical laws to qualify as collectives. For von Mises, a collective had to be unpredictable.

In 1937, Abraham Wald (1902–1950) gave a formulation that von Mises did accept. Wald proposed that we permit any rule for selecting a subsequence that decides whether to include $x_n$ on the basis of the preceding outcomes $x_1, \ldots, x_{n-1}$. But he demanded that such a rule be expressed in a formal logic, and he pointed out that only countably many rules can be expressed in a logic with a finite alphabet. If the sample space is finite, then it follows easily from standard probability theory that sequences with the same limiting frequencies for a given countable number of subsequences do exist, and similar results can also be obtained for infinite sample spaces if only well-behaved subsets are considered. In 1940, shortly after the publication of Wald's ideas, the American logician Alonzo Church (1903–1995) used the concept of effective computability to support Wald's argument that countably many rules would be enough to account for every implementable method of selection. Church had introduced the mathematical notion of effective computability in 1936, and its adequacy for representing the empirical notion of effective calculation had been supported by Alan Turing's demonstration that it is equivalent to computability by a universal computing machine.

Church's contribution marked the end of the intense discussion of collectives that flourished in the 1930s. The concept of a collective was consistent, but for most mathematicians, it was not very useful. In spite of von Mises's efforts, especially in [311], the derivation of the rules of probability from the concept of a collective remained messy, and the idea of an infinite sequence seemed to contribute nothing to actual uses of probability. Kolmogorov's axioms were a much clearer and more convenient starting point. (For a more detailed but still concise account of the work on collectives in the 1930s, see Martin-Löf 1969. See also [187, 269].)

### Ville's Strengthening of the Concept of a Collective

A important contribution to the theory of collectives, which had relatively little impact at the time but has turned out to be of fundamental importance, was made by the young French mathematician Jean Ville (1910–1988). As Ville showed, probability theory requires even more irregularity from a sequence than is required by the Mises-Wald definition of a collective [307, 308, 306].

Ville's favorite example concerns the oscillation of a cumulative relative frequency as it converges. We expect the frequency to oscillate above and below its limiting value, and this expectation became part of probability theory when Khinchin proved the law of the iterated logarithm (see p. 99). But as Ville showed, the Mises-Wald definition does not imply this oscillation. No matter what countable set of subsequence selection rules are adopted, one can construct a Mises-Wald collective of 0s and 1s in which the frequency of 1s approaches its limiting value from above.

Ville dramatized the unacceptability of a collective of 0s and 1s in which the limiting frequency of 1s is approached from above by demonstrating that it is susceptible to a gambling system that varies its bets. Suppose, to fix ideas, that the limiting frequency is $1/2$. Then there is a system for placing even-money bets, possibly using knowledge of the previous outcomes, that will avoid bankruptcy no matter what happens but will increase one's capital without bound if $y_n/n$ always stays above $1/2$.

Considered as tests of the randomness of sequences of 0s and 1s, gambling systems are more powerful, as a class, than subsequence selection rules. Given any subsequence selection rule, there is a system for placing even-money bets that avoids bankruptcy no matter what happens and produces an infinite return if the frequency of 1s in the subsequence does not converge to $1/2$ (see §3.1). In other words, there is a gambling system that detects any deviation from randomness that a given subsequence selection rule detects. And as Ville's example shows, there are also gambling systems that detect deviations no subsequence selection rule detects.

In fact, Ville's method for testing for randomness is universal, in the sense that it can detect any deviation from randomness by a sequence of 0s and 1s. A deviation from randomness, by definition, is the failure of a property $E$ that has probability one under classical probability theory (or under measure-theoretic probability theory; there is no difference in the simple case of independent trials of an experiment with only two outcomes). For any such property $E$, as Ville showed, there is a system for betting at even odds that avoids bankruptcy no matter what happens and produces an infinite return if $E$ does not happen (see §8.1). The oscillation required by the law of the iterated logarithm is merely one example of a property $E$ that has probability one.

Ville used von Mises's own words to buttress his view that testing for randomness using a gambling system is just as valid as testing for randomness using a subsequence selection rule. In explaining the requirement that the limiting frequency in a collective be invariant across subsequences, von Mises had argued that the futility of even the most complicated systems for beating the odds in games of chance is part of "the experimental basis of our definition of probability" (p. 25 of von Mises 1957). This experience teaches that "one cannot improve one's chances by a systematic choice of trials" (p. 11 of von Mises 1931, our translation). As Ville noted, it also teaches that one cannot improve one's chances by varying how and how much one bets. Gambling systems usually do vary their bets depending on preceding outcomes, and this feature is mentioned in discussions of their futility in nineteenth-century books on probability (Lacroix 1822, pp. 123–124; Cournot 1843, §62; Bertrand 1889, §113).

So why not use Ville's notion of a gambling system, rather than von Mises's weaker notion of subsequence selection, to define the concept of a collective? As Ville noted, for any $p$ between 0 and 1 (we need not take $p$ equal to $1/2$) and any countable set of systems for gambling at odds $p : 1-p$ without risking bankruptcy, a sequence of 0s and 1s exists such that the capital from each of the gambling systems remains bounded. (The capital process for each gambling system remains bounded with probability one, the intersection of countably many events with probability one also has probability one, and a set that has probability one cannot be empty.) By including appropriate gambling systems in the countable set, we can make sure that $p$ is the limiting frequency of 1s in the sequence and in any given countable number of subsequences, thus accomplishing all that von Mises and Wald accomplished. And we can also make sure that the sequence satisfies any given countable set of further properties that have probability one under the probability measure that makes the outcomes independent and assigns probability $p$ to each coming out 1.

Wald welcomed this strengthening of the idea of a collective. As he noted, the number of gambling systems expressible in a formal language, like the number of subsequence selection rules expressible in a formal language, is necessarily countable ([91], pp. 15–16). Von Mises, however, never conceded Ville's point ([312], p. 66). Apparently von Mises's appeal to the impossibility of a gambling system had only been rhetorical. For him, frequency, not the impossibility of a gambling system, always remained the irreducible empirical core of the idea of probability.

For his own part, Ville retained a critical stance towards collectives, refusing even to embrace his own strengthening of the concept. He agreed with his adviser, Fréchet, that probability should be founded on axioms (perhaps Kolmogorov's axioms or perhaps a similar system that treats conditional probabilities as primitive), not on collectives. And he saw paradox lurking in Wald's idea that one can enumerate all definable subsequence selection rules or gambling systems. He knew how to construct a collective $x_S$ for a countable set $S$ of selection rules, and he thought that the 1s in $x_S$ would effectively define a new selection rule not in $S$, producing a contradiction ([306], p. 134). In 1940, Church made it clear that this paradox is avoided: the effectively computable selection rules are enumerable but not effectively so. But this was not enough to revive interest in collectives. Twenty-five years passed before the topic was seriously revisited, by Per Martin-Löf 1969 and Claus Peter Schnorr 1971.

### Kolmogorov Complexity

As a frequentist, Kolmogorov always felt that his axioms needed something like von Mises's theory in order to connect with frequencies in the world. But he agreed with the critics that the infinitary character of von Mises's collectives made them useless for connecting with our finite experience. For many decades, he thought it was impossible to express his finitary frequentism in mathematics. But in the early 1960s, he realized that finite sequences and selection rules for them can be studied using Turing's idea of a universal computing device, which permits an objective measurement of complexity and simplicity for finite mathematical objects: given a universal computing device, one mathematical object can be considered simpler than another if there is a shorter program for generating it.

Kolmogorov first considered applying the notion of algorithmic simplicity to finite sequences in the same way Wald and Church had applied effective computability to von Mises's infinite sequences. In an article entitled "On tables of random numbers", published in 1963, Kolmogorov pointed out that only a few of the many subsequence selection rules that might be applied to a very long but finite sequence of 0s and 1s can be algorithmically simple. He suggested that we call the sequence $p$-random if the frequency of 1s is close to $p$ in all sufficiently long subsequences selected by these easily described rules. But he soon turned to a simpler idea, not so purely frequentist but too natural to resist. Why not consider directly the complexity of a finite sequence of zeros and ones? If the sequence's algorithmic complexity is close to maximum (the length of the sequence), it can be regarded as random. Per Martin-Löf developed this idea when he visited Moscow in 1963–1964. In "The definition of random sequences", a celebrated paper published in 1966, Martin-Löf showed that there exists a universal statistical test for detecting sequences that are nonrandom in Kolmogorov's sense. He also extended the idea of a universal statistical test to infinite sequences. In 1971 Schnorr showed that Martin-Löf's notion of randomness for infinite sequences is also obtained if one imposes tests of Ville's type; this result was also independently obtained by Leonid Levin. There are yet other variants on the idea; for reviews of the work in this area, see [183, 204, 269].

As for Kolmogorov himself, he continued to be concerned with finite sequences, and he eventually proposed an approach to statistical modeling that bypasses probability, using algorithmic complexity directly to define statistical models [331]. The field of study now called Kolmogorov or algorithmic complexity encompasses various ideas for using complexity for statistical work, including ideas on inductive inference championed by Ray J. Solomonoff, who studied the universal measurement of complexity independently of Kolmogorov [294]. An excellent overview of the field is provided by Li and Vitányi 1997.

The success of Kolmogorov complexity has vindicated von Mises's ambition to analyze the idea of randomness, and Kolmogorov complexity has contributed to the development of our framework. Saying that Skeptic cannot become infinitely rich is the same as saying that Reality's moves will pass any computable test based on Forecaster's prices—that is, they will be random in the sense of Schnorr. But we are concerned with potential unpredictability rather than concrete realized randomness. Moreover, we do not even attempt to measure unpredictability quantitatively in this book. We do believe, however, that the future development of our framework will involve more complicated games, in which different Forecasters compete and are compared quantitatively. This will require a greater use of ideas from Kolmogorov complexity.

## 2.4 WHAT IS A MARTINGALE?

In this section, we discuss the idea of a martingale: its origins in gambling, its introduction into probability theory by Jean Ville, its role in measure-theoretic probability, and its somewhat different role in the game-theoretic framework.

### Martingales before Jean Ville

A martingale, in English and French, is a system by which a gambler tries to become rich. The word appears in the dictionary of the French Academy as early as 1762, where it is said to refer to the particular system where the gambler doubles his bet until he finally wins. If the gambler starts by betting $\alpha$, and he finally wins on the $n$th bet, then his profit on that bet, $2^n\alpha$, will exceed his total loss,

$$\alpha + 2\alpha + \cdots + 2^{n-1}\alpha,$$

by $\alpha$. The difficulty is that a modest losing streak will exhaust the gambler's capital or bring him up against the house's limit on the size of a bet. As Sylvestre François Lacroix (1765–1843) wrote in his textbook on probability in 1822,

> Gamblers with little capital cannot follow a scheme like this, which is called making the martingale, without hurting or ruining themselves, and they are often forced to give up, losing everything they have laid out. As for the rich, they can make much better use of their capital ...

(p. 124, our translation). But such schemes are constantly reinvented by mathematically inclined novice gamblers. We can find manuals for gamblers who want to follow martingales (e.g., [76]), and the theme of the martingale appears frequently in literature (see the works by Cocteau, Picard, and Thackeray in our list of references).

In an apparently older sense, the word martingale refers to the strap of a horse's harness that passes from the noseband through the forelegs and keeps the horse from raising his head. In this meaning, at least, the word came into French and English from Arabic through Spanish, although many authorities believe it was also influenced by *martegalo*, the word in Provençal for a female inhabitant of the Mediterranean port of Martigues. Some French dictionaries quote Rabelais, in 1491, ridiculing shoes that fastened in the back, *à la martingale*. Was he comparing them to a horse's harness, referring to the style of dress in Martigues, or already evoking the wild abandon of the gambler going for all or nothing?

## Martingales in Jean Ville's *Étude critique de la notion de collectif*

Martingales are wild systems of play, but the distinction between wild and tame is not sharp. So in mathematics, we might call any strategy for placing bets a martingale. This is what Ville did, with no sense of originality. But he then gave the word a twist. In the case of coin tossing, which he emphasized, strategies are in a one-to-one correspondence with their capital processes once we have fixed the initial capital. So Ville called the capital processes martingales.

Formally, Ville was considering a sequence of play with outcomes $x_1, x_2, \ldots$ and a gambler who begins with initial capital equal to 1. A strategy for the gambler is a rule that chooses, for each initial subsequence $x_1, \ldots, x_{n-1}$, a gamble on $x_n$ (see §1.2). The strategy's capital process is the real-valued function $\mathcal{L}$ that assigns to each sequence $x_1, \ldots, x_n$ the capital $\mathcal{L}(x_1, \ldots, x_n)$ the gambler will have right after his gamble on $x_n$ is settled. If we assume (as Ville did but as we do not do in this book) that the permitted gambles define a probability distribution for $x_1, x_2, \ldots$, then we can speak of the expected value of $\mathcal{L}(x_1, \ldots, x_n)$, and we have

$$\mathbb{E}[\mathcal{L}(x_1, \ldots, x_n) \mid x_1, \ldots, x_{n-1}] = \mathcal{L}(x_1, \ldots, x_{n-1}). \tag{2.8}$$

Conversely, any process $\mathcal{L}$ starting at 1 and satisfying (2.8) can be obtained from some strategy. So Ville took (2.8) and $\mathcal{L}(\Box) = 1$ as his mathematical definition of a *martingale* (Ville 1939, p. 99).

As we have already noted, Ville demonstrated that for any set $E$ of $x_1, x_2, \ldots$ that has probability one, there exists a nonnegative martingale that diverges to infinity if $E$ fails. He demonstrated, in fact, that this is an equivalence: an event $E$ has probability one if and only if there exists a nonnegative martingale that diverges to infinity if the event fails. Moreover, he came very close to showing that nonzero probabilities can also be interpreted in terms of martingales: for any event $E$,

$$\mathbb{P}E = \inf \left\{ \mathcal{L}_0 \,\middle|\, \liminf_{n \to \infty} \mathcal{L}_n \geq \mathbb{I}_E \right\}, \tag{2.9}$$

where $\mathcal{L}$ ranges over all martingales, $\mathcal{L}_0$ is the initial value of $\mathcal{L}$, and $\mathcal{L}$ is its value at time $n$. We can put (2.9) in words by saying that the probability of an event is the smallest possible initial value for a nonnegative martingale that eventually reaches or exceeds 1 if the event happens. (We give proofs of Ville's results in §8.5.)

Ville also explored other uses of the concept of a martingale in probability theory. He used it to prove several limit theorems, and he even ventured to extend his results, along with his definition of martingale, to continuous time. He did not, however, pursue the study of martingales further after his book was published. (We provide a brief biography of Ville in §8.6.)

## Game-Theoretic Martingales

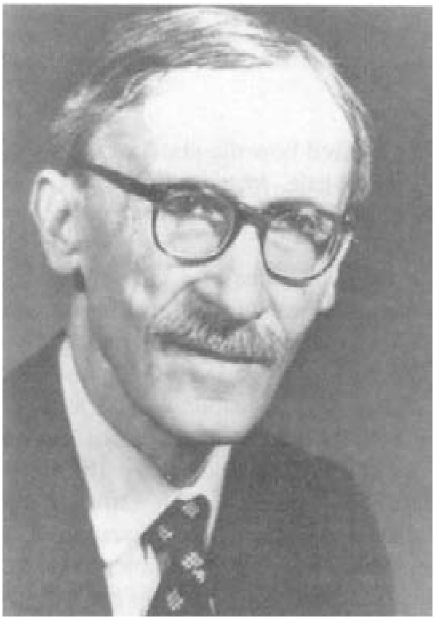
It is scarcely an exaggeration to say that Ville showed how the classical theory of probability can be founded on the notion of a martingale. Martingales can serve as our starting point, because probability can be defined in terms of martingales, using Equation (2.9).

Our game-theoretic framework simply develops this idea in a thorough-going way. Instead of beginning with classical probabilities, as Ville did, we begin with opportunities for betting, which define possible capital processes. Then we define probability in terms of these capital processes, in the way suggested by Ville's results. We say that an event happens almost surely (that is, with probability one) if there is a nonnegative capital process that diverges to infinity if the event does not happen (p. 17), and we define upper probability in the spirit of Equation (2.9): the upper probability of an event is the smallest possible initial value for a nonnegative capital process that eventually reaches or exceeds 1 if the event happens (see Equation (1.4) on p. 15). Because Ville began with classical probabilities, it was possible to recover these probabilities by Equation (2.9). We begin instead with arbitrary betting opportunities, which may or may not define probabilities; in general they define only upper probabilities. But this is the only essential difference between our viewpoint and Ville's, which was already relatively game-theoretic.

When the protocol for our probability game is symmetric (Skeptic can always buy tickets at the same price at which he is allowed to sell them), we even follow Ville by using the word "martingale": we call a capital process for Skeptic in a symmetric probability protocol a (game-theoretic) *martingale*. This is more general than Ville's notion of a martingale, because even in a symmetric probability protocol there is no requirement that Skeptic be offered enough gambles to define a probability measure (see p. 67 and p. 188). But it preserves Ville's basic intuitions about martingales. The value of a game-theoretic martingale $\mathcal{L}$ in the initial situation $\Box$ in a symmetric probability protocol is the price at which Skeptic can buy or sell $\mathcal{L}$'s future payoffs. This is because there is a strategy $\mathcal{P}$ that produces these payoffs starting with $\mathcal{L}(\Box)$. If Skeptic follows $\mathcal{P}$ starting with zero instead of $\mathcal{L}(\Box)$, he gets $\mathcal{L} - \mathcal{L}(\Box)$; this is buying $\mathcal{L}$ for $\mathcal{L}(\Box)$. If he follows the opposite strategy, $-\mathcal{P}$, starting with zero, he gets $-\mathcal{L} + \mathcal{L}(\Box)$; this is selling $\mathcal{L}$ for $\mathcal{L}(\Box)$.

A nonnegative game-theoretic martingale in a symmetric probability protocol is the same thing as a capital process for a strategy that does not risk bankruptcy, and our fundamental interpretative hypothesis can be expressed by saying that any given nonnegative game-theoretic martingale will remain bounded. Equivalently, a game-theoretic martingale that is bounded below will not become infinitely large. Because $-\mathcal{L}$ is a game-theoretic martingale whenever $\mathcal{L}$ is, we can also say that a game-theoretic martingale that is bounded above will not become infinitely negative.

## Measure-Theoretic Martingales



Paul Lévy (1886–1971). His beloved teacher Jacques Hadamard disapproved of his work in probability, but Bourbaki ranked him as one of the leading French mathematicians between the two world wars [202, 270].

In the 1940s and 1950s, Joseph L. Doob adapted Ville's concept of a martingale to the measure-theoretic framework.

Doob's thinking was influenced by work on dependent random variables that goes back to Andrei A. Markov in the nineteenth century and Sergei Bernstein in the early twentieth century. Markov was interested in extending the central limit theorem from independent random variables to dependent random variables. Bernstein, a Russian mathematician who completed a doctorate in analysis in Paris in 1904, began working on this topic in about 1917, and in the 1920s, he published proofs of the central limit theorem for sequences of random variables $x_1, x_2, \ldots$ obeying various conditions on the mean and variance of $x_n$ given $x_1, \ldots, x_{n-1}$ [17, 18].

Around 1930, following Bernstein's lead, Paul Lévy took up Markov's problem. In 1935, Lévy began to emphasize what he called *condition $C$*: the mean of $x_n$ given $x_1, \ldots, x_{n-1}$ should be zero. If we write $\mathcal{S}(x_1, \ldots, x_n)$ for the sum of $x_1, \ldots, x_n$, then we can express Lévy's condition $C$ in the form

$$\mathbb{E}[\mathcal{S}(x_1, \ldots, x_n) \mid x_1, \ldots, x_{n-1}] = \mathcal{S}(x_1, x_2, \ldots, x_{n-1}),$$

which looks like Ville's definition of a martingale.

Ville had not investigated the central limit theorem for dependent variables, but Doob, who reviewed Ville's book right after its publication [96], immediately saw the connection. He also thought it more natural to write

$$\mathbb{E}[\mathcal{S}_n \mid \mathcal{S}_1, \ldots, \mathcal{S}_{n-1}] = \mathcal{S}_{n-1}, \tag{2.10}$$

where $\mathcal{S}_n = \mathcal{S}(x_1, \ldots, x_n)$. This is a property of a sequence of random variables $\mathcal{S}_1, \mathcal{S}_2, \ldots$, which can be studied in its own right. One can put matters more abstractly: writing $\mathcal{F}_n$ for the $\sigma$-algebra generated by $\mathcal{S}_1, \ldots, \mathcal{S}_n$, one can rewrite (2.10) as

$$\mathbb{E}[\mathcal{S}_n \mid \mathcal{F}_{n-1}] = \mathcal{S}_{n-1}. \tag{2.11}$$

It is also interesting to relax the condition that $\mathcal{F}_n$ be the $\sigma$-algebra generated by $\mathcal{S}_1, \ldots, \mathcal{S}_n$ to the condition that these random variables be measurable with respect to $\mathcal{F}_n$ and that the $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$ for $n = 1, 2, \ldots$ (see §8.1).

In an article he published in 1940, in which he called (2.11) *condition* $\mathcal{E}$, Doob began to explore this condition in the measure-theoretic framework, in both discrete and continuous time. As he explained later, $\mathcal{E}$ did not stand for expected value; it was merely the second letter in the alphabet after $\mathcal{C}$. It would have been too self-centered to use $\mathcal{D}$ [293]. Ten years later, when working on his immensely influential *Stochastic Processes*, which appeared in 1953, Doob returned to condition $\mathcal{E}$ and began calling sequences satisfying it martingales [99, 100]. This is now the accepted meaning of the word in measure-theoretic probability.

We use "martingale" in both its game-theoretic and measure-theoretic senses in this book, appending the adjective "game-theoretic" or "measure-theoretic" when this is necessary in order to avoid confusion.

## 2.5 THE IMPOSSIBILITY OF A GAMBLING SYSTEM

Because the hypothesis of the impossibility of a gambling system is one of the two pillars of our framework, our historical review would not be complete without a search for its antecedents. Where, in the historical development of probability theory, did the hypothesis appear, and how did it develop?

### Different Ways of Understanding the Futility of Gambling Systems

When was the impossibility of a gambling system first articulated? This is a tricky question because, as Maurice Fréchet pointed out, the impossibility can be understood in many different ways ([130], p. 40).

Within classical or measure-theoretic probability theory, we might make the following points in order to elaborate the assertion that a gambling system cannot succeed in a fair game:

1. If $x_1, x_2, \ldots$ are independent and identically distributed random variables, then the sequence of random variables obtained by applying a subsequence selection rule to $x_1, x_2, \ldots$ has the same joint probability distribution.

2. If at the beginning of each round of a game, the expected gain for each gamble available on that round is zero, then the initial expected value of one's total gain is zero, no matter what strategy one adopts.

3. A player's gain in a fair game, being a random variable with expected value zero, cannot be positive with positive probability without also being negative with positive probability.

4. If the player follows a system that does not risk bankruptcy, then the odds are at least $K - 1$ to one against his multiplying his stake by $K$.

Points 2 and 3, which assume constraints on the size of bets, were made by Cournot 1843 and Bertrand 1889 in their discussions of the futility of gambling systems. Doob 1936 advanced Point 1 as an expression of the impossibility of a gambling system within measure-theoretic probability. Ville, as we have seen, put the spotlight on Point 4. In 1965, Lester E. Dubins and Leonard J. Savage christened Point 2 the *conservation of fairness* ([105], §1.3).

Jean Ville put the emphasis on Point 4, which he expressed by means of what we now call *Doob's inequality*: if a nonnegative martingale $\mathcal{L}$ has the initial value 1, and $\lambda > 0$, then

$$\mathbb{P}\left\{\sup_n \mathcal{L}(x_1, \ldots, x_n) \geq \lambda\right\} \leq \frac{1}{\lambda} \tag{2.12}$$

(Ville 1939, p. 100). This generalizes *Markov's inequality*, which says that a non-negative random variable $x$ satisfies

$$\mathbb{P}\{x \geq \lambda\} \leq \frac{\mathbb{E}\, x}{\lambda}$$

(Markov 1900, §13). As we have noted, Ville also proved an infinitary version of (2.12): if a strategy does not risk bankruptcy, then the probability that it makes you infinitely rich is zero. And for the classical fair coin (where $x_n$ is 0 or 1 with probability 1/2, no matter how $x_1, \ldots, x_{n-1}$ come out), he proved a converse: if a certain property of the sequence $x_1, x_2, \ldots$ has probability zero, then there is a strategy that makes you infinitely rich whenever that property holds (see §8.1). Thus there is a full equivalence, in the canonical classical case of the fair coin, between probability zero and Skeptic's becoming infinitely rich. This equivalence was significant for Ville's approach to collectives, because it gave intuitive grounds, conceptually prior to probability theory, for requiring a collective to avoid any property that is given probability zero by probability theory. It is equally significant for us, because it makes clear that "unless Skeptic becomes infinitely rich" derives as much legitimacy from classical probability theory as "except on a set of measure zero".

Points 1–4 are valid, with appropriate qualifications and changes in terminology, within our game-theoretic framework. But in whatever framework they are understood (classical, measure-theoretic, or game-theoretic), they are only statements about probabilities. Our *hypothesis* of the impossibility of a gambling strategy, in contrast, is not a statement about probabilities. It is a hypothesis that relates a certain game to the world, which we can state before we compute probabilities for events from the prices in the game. When adopted for particular prices for a particular phenomenon, this hypothesis asserts directly our practical certainty that a gambler cannot get rich at the prices offered in the game without risking bankruptcy. This assertion stands outside the mathematics of probability, serving as a bridge between it and the world. It is our version of Cournot's bridge.

Our understanding of the impossibility of a gambling system, as something prior to the computation of probabilities, is relatively novel. It departs from the thinking of Cournot, Lévy, and Kolmogorov, for they formulated principles for interpreting zero or small probabilities, not hypotheses expressed without reference to probabilities. Von Mises anticipated our way of thinking, for he did say that the impossibility of a gambling system is something more primitive than probability. But his allegiance to this idea, as we have seen, was half-hearted; in the end he refused to acknowledge that the impossibility of a gambling system is more fundamental than invariant frequency.

And Ville was less radical, in some respects, than von Mises; in spite of his discovery of the fundamental properties of martingales, he insisted on classical or measure-theoretic probabilities as his starting point. So our understanding of the impossibility of a gambling system as a primitive hypothesis can be seen as the distinguishing novelty in our game-theoretic framework.

### The Prequential Principle

In order to appreciate fully the independence of our fundamental hypothesis from classical or measure-theoretic probability theory, we should recognize that the prices whose testing is enabled by the hypothesis may fall short of defining a full probability distribution for a process, *even if the prices for each step of the process define a full probability distribution for what happens on that step.* This aspect of our framework finds an important antecedent in A. Philip Dawid's work on probability forecasting.

We can begin our discussion of probability forecasting by recalling an exchange between Émile Borel and Hans Reichenbach. Reichenbach argued that even the probability of an isolated event, such as Julius Caesar having visited Britain, can be given a frequency interpretation. When a historian asserts that Caesar visited Britain, he should set his probability equal to the frequency with which historians are correct in cases with similar evidence. Reichenbach acknowledged that there might be some question about how to choose the class of similar cases for which to calculate the frequency, but he counseled choosing the most restricted class, consisting of the most similar cases (Reichenbach 1937, p. 314). In his *Valeur pratique et philosophie des probabilités*, Borel agreed with Reichenbach that it makes sense to talk about the probability of an isolated event but disputed the viability of a purely frequentist interpretation. In Borel's view, the class of entirely similar cases—past and future—is usually too small to permit us to speak of frequencies. In order to assess the probability of an isolated event, we must take details into account by subjectively adjusting general frequencies (Borel 1939, p. 86).

Ville took up the issue at the very end of his own book (Ville 1939, p. 139). In order to deal with isolated events, Ville contended, we must generalize von Mises's notion of a homogeneous sequence of events by considering an observer O who makes probability judgments about a sequence $E_1, E_2, \ldots$ of events:

$$\mathbb{P} E_1 = p_1, \ \ \mathbb{P} E_2 = p_2, \ \ \ldots .$$

Each judgment $\mathbb{P} E_i = p_i$ can be influenced by previous outcomes. The problem, Ville said, is to find "an objective method of verifying the value of O's judgments".

The problem was taken up in 1950 by Glenn W. Brier, who was engaged in training weather forecasters. For Brier, $E_n$ might be the event that it rains on day $n$. Brier's work led to a substantial literature [84, 265]. Most of this literature lies beyond the scope of this book, because it is concerned with nonlinear methods of scoring the forecaster (an extension of our methods to this case is given in [330]). It is instructive for us, however, to note the mismatch between the forecasting story and the measure-theoretic framework.

How can we put the forecaster's $p_n$ into the measure-theoretic framework? The obvious way is to think of each one as the forecaster's conditional probability given all his information at the time he announces it. If forecasts are given for four days, we have four such conditional probabilities:

$$\mathbb{P}[E_1 \mid \mathcal{I}_0], \quad \mathbb{P}[E_2 \mid \mathcal{I}_1], \quad \mathbb{P}[E_3 \mid \mathcal{I}_2], \text{ and } \mathbb{P}[E_4 \mid \mathcal{I}_3], \tag{2.13}$$

where $\mathcal{I}_n$ represents all the forecaster's relevant information (current weather maps, previous outcomes, success of previous forecasts, etc.) when he makes the forecast for day $n+1$. But the forecaster has not defined a probability measure for which these are conditional probabilities. He has not even specified a sample space, which needs to include, in addition to the $2^4$ possible outcomes for the sequence $E_1, E_2, E_3, E_4$, some number of alternative outcomes for the information sequence $\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3$.

We can certainly define a sample space and a probability measure on it that will have the conditional probabilities given in (2.13). In fact, there are many ways of doing this. But the forecaster need not do it, and whether he does it is irrelevant to the evaluation of his success. This is the point of the *prequential principle*, which Dawid first proposed in 1984. This principle says that the evaluation of a probability forecaster should depend only on his actual probability forecasts and the corresponding outcomes. The additional information contained in a probability measure that has these probability forecasts as conditional probabilities should not enter into the evaluation, even if the probability measure was actually constructed and the probability forecasts were derived from it.

The word *prequential* is Dawid's invention; as he explained, it "combines *pro*bability forecasting with *sequential pre*diction". In 1985, Dawid applied the principle to infinite sequences of probability forecasts. In subsequent articles, including [85, 272], he used it to propose new statistical methods. Unfortunately, these methods do not quite comply with the prequential principle; they still depend on some mild assumptions about the measure from which one's probability forecasts are derived as conditional probabilities (see Dawid and Vovk 1997).

Within the game-theoretic framework, we can evaluate probability forecasts in a way that complies strictly with the prequential principle. We do not begin with a probability measure. We begin only with a sequential game. Forecaster moves by giving a probability for rain, based on whatever information he has and wants to use, in whatever way he wants to use it. Skeptic bets at the odds defined by this probability, and Reality decides whether it rains. Any strategy for Skeptic that does not risk bankruptcy casts doubt on Forecaster's bona fides if it multiplies Skeptic's initial capital by an unreasonably large factor. But in the end, we will have have only $n$ probabilities $p_1, \ldots, p_n$, not a full probability measure for the process $E_1, \ldots, E_n$.

We will return to the idea of probability forecasting in §7.3, where we consider one aspect of the evaluation of probability forecasters (calibration), and at greater length in Chapter 8, where we analyze a probability measure for a process as a strategy for our player Forecaster in a probability protocol, and where we discuss scientifically important examples of open probability protocols, in which a scientific theory provides probability forecasts without defining such a strategy.

## 2.6 NEOSUBJECTIVISM

Bruno de Finetti (1906–1985) at a reception in Rome, around 1958.

We have already mentioned the neosubjectivist interpretation of probability associated with Bruno de Finetti. De Finetti participated in the foundational debates of the 1930s, and in subsequent decades his reasoned but uncompromising subjectivism emerged as the leading alternative to the vague stochasticism associated with the measure-theoretic framework.

Our most important point of agreement with de Finetti, perhaps, is our emphasis on price over probability. Like de Finetti and Pascal, we begin with prices and derive probabilities. We also agree with de Finetti that Kolmogorov's sixth axiom should be regarded as optional. Because our experience is finite, infinitary statements are less fundamental than finitary statements. We do allow Skeptic to combine a countably infinite set of strategies (see, e.g., Lemma 3.2, p. 68), but this is much less problematic than Kolmogorov's sixth axiom, because it does not appeal to infinite experience: it is merely a way of obtaining a new strategy which can then be used directly, not as a combination of infinitely many components. Of course, we do sometimes take advantage of the mathematical clarity permitted by the use of infinity, just as de Finetti did. Chapters 3–5, for example, are devoted to infinite-horizon games.

De Finetti also insisted that probability is primarily a choice made by an individual. Our framework allows this role of probability to be made absolutely clear: in many cases, a probability is a move made by Forecaster. We also speak, however, of probabilities that are derived from the structure of the game, as in (1.4) and (1.5).

The most important point where we differ with de Finetti is in our emphasis on upper and lower price. De Finetti was not persuaded that there is anything to be gained by generalizing price to possibly unequal upper and lower prices ([93], Vol. 2, Appendix 19.3). For him, use of probability required that one assign to each uncertain quantity $x$ a price $\mathbb{E}\,x$ such that one is indifferent between $x$ and $\mathbb{E}\,x$. Even here, however, de Finetti has had his influence. Our analysis of upper and lower price finds antecedents in the work of other authors who were directly inspired by de Finetti: Peter M. Williams (1976) and Peter Walley (1991).

We also differ with de Finetti with respect to the sequential nature of experience, which we emphasize more than he did, and Bayes's theorem, which we emphasize less. Bayes's theorem can be developed in our framework, but we do not develop it in this book. Advances in Bayesian statistics in recent decades have vindicated de Finetti's insistence on its usefulness, but other probabilistic learning methods are also useful, and we are particularly interested in those that exploit directly the idea of a sequential game.

## 2.7 CONCLUSION

Let us summarize briefly the most important sources and antecedents for the different aspects of our game-theoretic framework.

1. Pricing by dynamic hedging goes back to Pascal, and our emphasis on the role of price in probability is also influenced by de Finetti.

2. Our understanding of how probability theory can be bridged with the world has roots in the thinking of Bernoulli and Cournot, and especially in Lévy's articulation of the principle of practical impossibility.

3. Our understanding of the hypothesis of the impossibility of a gambling strategy has roots in the work of von Mises and Ville.

4. Our emphasis on testing only using what actually happens is inspired by Dawid's prequential principle.

As we will explain in Chapter 9, our understanding of dynamic hedging also owes a great deal to work in finance theory, from Bachelier to Merton. The reader may be able to identify other antecedents of our ideas. The depth, diversity, and number of these antecedents demonstrate the richness of the game-theoretic framework and helps explain its ability to capture so many of our intuitions about probability.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 3

# *The Bounded Strong Law of Large Numbers*

In this chapter, we formulate and prove the simplest forms of the game-theoretic strong law of large numbers.

In its very simplest form, the strong law of large numbers is concerned with a sequence of events, each with the same two possible outcomes, which we may call heads and tails. The law says that as one proceeds in the sequence, the proportion of heads converges to one-half almost surely. In symbols:

$$\lim_{n\to\infty} \frac{y_n}{n} = \frac{1}{2} \tag{3.1}$$

almost surely, where $y_n$ is the number of heads among the first $n$ events. A framework for mathematical probability must provide a precise mathematical context for this statement, including a mathematical definition of the term *almost surely*.

In the measure-theoretic framework, the mathematical context is provided by adopting a certain probability measure for the infinite sequence of events: heads has probability one-half each time, and the events are independent. The term *almost surely* means except on a set of measure zero. This makes our claim about convergence into the precise statement known as Borel's strong law: the sequences of outcomes for which the convergence to one-half fails have measure zero under the specified probability measure.

The game between Skeptic and Reality that we study in this chapter makes the claim about convergence precise in a different way. No probability measure is given, but before each event, Skeptic is allowed to bet as much as he wants on heads or on tails, at even odds. The meaning of *almost surely* is this: an event happens almost surely if Skeptic has a strategy for betting that does not risk bankruptcy and allows him to become infinitely rich if the event does not happen. This is all we need: the statement that the proportion of heads converges to one-half almost surely is now a precise statement in game theory, which we prove in this chapter.

The preceding paragraph does not mention our fundamental interpretative hypothesis, which says that Skeptic cannot become infinitely rich without risking bankruptcy. As we explained in Chapter 1, this hypothesis stands outside the mathematical theory. It has nothing to do with the mathematical meaning of the strong law of large numbers or with its proof. It comes into play only when we use the theorem. If we adopt the hypothesis for a particular sequence of events, then the statement "the proportion of these events that happen will converge to one-half almost surely" acquires a useful meaning: we think the convergence will happen. If we do not adopt the hypothesis for the particular sequence, then the statement does not have this meaning.

The game-theoretic formulation is more constructive than the measure-theoretic formulation. We construct a computable strategy for Skeptic that is sure to keep him from becoming bankrupt and allows him to become infinitely rich if the convergence of the proportion of heads to one-half fails. Moreover, our formulation is categorical—it is a statement about all sequences, not merely about sequences outside a set of measure zero. Every sequence either results in an infinite gain for Skeptic or else has a proportion of heads converging to one-half.

The game-theoretic formulation has a further unfamiliar and very interesting facet. In the folk picture of stochastic reality, outcomes are determined independently of how any observer bets. In the game between Skeptic and Reality, in contrast, Reality is allowed to take Skeptic's bets into account in deciding on outcomes. Yet this does not prevent Skeptic from constructing a winning strategy. No matter how diabolically Reality behaves, she cannot violate the required convergence without yielding an infinite gain to Skeptic.

We do not propose to replace stochastic reality with a rational, diabolical reality. We propose, rather, to eliminate altogether from the general theory of probability any particular assumption about how outcomes are determined. It is consistent with our framework to suppose that Reality is nothing more than the actual outcomes of the events—that Reality has no strategy, that there is no sense in the question of how she would have made the second event come out had the first event come out differently or had Skeptic bet differently. By lingering over this supposition, we underline the concreteness of the strong law of large numbers; it concerns only our beliefs about a single sequence of actual outcomes. But it is equally consistent with our framework to imagine an active, strategic Reality. This diversity of possible suppositions hints at the breadth of possible applications of probability, a breadth not yet, perhaps, fully explored.

We formalize our game of heads and tails in §3.1. Then, in §3.2, we generalize it to a game in which Reality decides on values for a bounded sequence of centered variables $x_1, x_2, \ldots$. The strong law for this game says that the average of the first $N$ of the $x_n$ will converge to zero as $N$ increases. In order to explain this in the measure-theoretic framework, we postulate a complete probability distribution for all the variables (this amounts to a specification of prices for all measurable functions of the variables), and then we conclude that the convergence will occur except on a set of measure zero, provided the conditional expectation (given the information available before time $n$) of each $x_n$ is zero. Our game-theoretic formulation dispenses not only with the use of measure zero but also with the complete probability distribution. We assume only that each $x_n$ is offered to Skeptic at the price of zero just before it is announced by Reality. Skeptic may buy the $x_n$ in any positive or negative amounts, but nonlinear functions of the $x_n$ need not be priced. This formulation is more widely—or at least more honestly—applicable than the measure-theoretic formulation.

In §3.3, we generalize to the case where the successive variables have prices not necessarily equal to zero. In this case, the strong law says that the average difference between the variables and their prices converges to zero almost surely. The mathematical content of this generalization is slight (we continue to assume a uniform bound for the variables and their prices), but the generalization is philosophically interesting, because we must now discuss how the prices are set, a question that is very important for meaning and application. The diversity of interpretations of probability can be attributed to the variety of ways in which prices can be set.

In §3.4, we briefly discuss the generalization from two-sided prices—prices at which Skeptic is allowed both to buy and sell—to one-sided prices, at which he is allowed only to buy or only to sell. If Skeptic is only allowed, for example, to buy at given prices, and not to sell, then our belief that he cannot become infinitely rich implies only that the long-term average difference between the variables and their prices will almost surely not exceed zero.

In an appendix, §3.5, we comment on the computability of the strategies we construct and on the desirability of detailed investigation of their computational properties.

The main results of this chapter are special cases of more general results we establish in Chapter 4, where we allow Reality's moves, the variables $x_n$, to be unbounded. For most readers, however, this chapter will be a better introduction to the basic ideas of the game-theoretic framework than Chapter 4, because it presents these ideas without the additional complications that arise in the unbounded case.

## 3.1 THE FAIR-COIN GAME

Now we formalize our game of heads and tails. We call it the *fair-coin game*, but not too much meaning should be read into this name. The outcomes need not be determined by tossing a coin, and even if they are, there is no need for the coin to have any property that might be called fairness. All that is required is that Skeptic be allowed to bet at even odds on heads or on tails, as he pleases.

Skeptic begins with some initial capital, say \$1. He bets by buying some number, say $M$, of tickets in favor of heads; $M$ may be any real number—positive, zero, or negative. Each ticket, which sells for \$0, pays the bearer \$1 if Reality chooses heads, and requires the bearer to forfeit \$1 if Reality chooses tails. So buying $M$ tickets means agreeing to pay $\$M$ if Reality chooses tails in order to gain $\$M$ if Reality chooses heads; if $M$ is negative, then this is really a bet in favor of tails. If we code tails as $-1$ and heads as $1$, then the protocol for the game can be described as follows:

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots$:
  Skeptic announces $M_n \in \mathbb{R}$.
  Reality announces $x_n \in \{-1, 1\}$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.

The quantity $\mathcal{K}_n$ is Skeptic's capital just after the bet on the $n$th toss is settled.

Skeptic wins the game if (1) his capital $\mathcal{K}_n$ is never negative, and (2) either

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0 \tag{3.2}$$

or

$$\lim_{n\to\infty} \mathcal{K}_n = \infty \tag{3.3}$$

holds. Otherwise Reality wins.

Equation (3.2) says that the proportion of heads in the first $n$ tosses converges to one-half. It is equivalent to (3.1), because $\sum_{i=1}^{n} x_i = 2y_n - n$. We use (3.2) instead of (3.1) only because it is suitable for the more general bounded forecasting game that we consider in the next section.

The rule for determining the winner, given by (3.2) and (3.2), completes our specification of the *fair-coin game*. It is a two-person, zero-sum, perfect-information game. Zero-sum because Skeptic wins if and only if Reality loses. (Since a win is conventionally scored as a 1 and a loss as $-1$, the two players' scores sum to zero.) Perfect-information because each player knows all the previous moves when he makes his own next move. Here is the strong law of large numbers for this game:

**Proposition 3.1** *Skeptic has a winning strategy in the fair-coin game.*

This means the convergence (3.2) occurs almost surely, in the game-theoretic sense of this term explained on p. 17 and p. 61. Skeptic has a strategy that forces Reality to arrange the convergence if she is to keep him from becoming infinitely rich. To the extent that we believe that Skeptic cannot become infinitely rich, we should also believe that the convergence will happen. If we adopt the fundamental interpretative hypothesis, then we may simply assert that the convergence will occur.

Some readers might prefer to allow Skeptic to borrow money. Skeptic does not need any such concession, however; he has a winning strategy even if he is not allowed to borrow. Moreover, allowing Skeptic to borrow would not really change the picture so long as there were a limit to his borrowing; allowing him to borrow up to $\$\beta$ would have the same effect on our reasoning as changing his initial capital from $\$1$ to $\$(1 + \beta)$, and so long as his initial capital is positive, its value makes no difference in our reasoning.

We will prove a generalization of Proposition 3.1 in §3.2. Our proof is constructive; we spell out Skeptic's strategy explicitly. The strategy can be described roughly as follows: If Skeptic establishes an account for betting on heads, and if at each step he bets a fixed proportion $\epsilon$ of the money then in the account on heads, then Reality can keep the account from getting indefinitely large only by eventually holding the average $\frac{1}{n}\sum_{i=1}^{n} x_i$ at or below $\epsilon$. So Skeptic can force Reality to hold the average at zero or less by splitting a portion of his capital into an infinite number of accounts for betting on heads, including accounts with $\epsilon$ that come arbitrarily close to zero. By also setting up similar accounts for betting on tails, Skeptic can force Reality to make the average converge exactly to zero.

The rule for determining the winner in our game will seem simpler if we break out the two players' *collateral duties* so as to emphasize the main question, whether (3.2) holds. The collateral duty of Skeptic is to make sure that his capital $\mathcal{K}_n$ is never negative. The collateral duty of Reality is to make sure that $\mathcal{K}_n$ does not tend to infinity. If a player fails to perform his or her collateral duties, he or she loses the game. (More precisely, the first player to fail loses. If Skeptic and Reality both fail, then Skeptic loses and Reality wins, because Skeptic's failure happens at some particular trial, while Reality's failure happens later, at the end of the infinite sequence of trials.) If both players perform their collateral duties, Skeptic wins if and only if (3.2) is satisfied.

Equation (3.2) is a particular event—a particular property of Reality's moves. We can define a whole gamut of analogous games with the same protocol and the same collateral duties but with other events in the place of (3.2). In this more general context, we say that a strategy *forces* an event $E$ if it is a winning strategy for Skeptic in the game in which $E$ replaces (3.2) as Skeptic's main goal. We say that Skeptic *can force* $E$ if he has a strategy that forces $E$—that is, if $E$ happens almost surely. As we will see in later chapters, Skeptic can force many events.

## 3.2 FORECASTING A BOUNDED VARIABLE

Suppose now that instead of being required to choose heads or tails (1 or $-1$) on each trial, Reality is allowed to choose any real number $x$ between $-1$ and 1. This number becomes the payoff (positive, negative, or zero) in dollars for a ticket Skeptic can buy for \$0 before the trial. Skeptic is again allowed to buy any number $M$ of such tickets; when he buys a positive number, he is betting Reality will choose $x$ positive; when he buys a negative number, he is betting Reality will choose $x$ negative.

Our new game generalizes the fair-coin game only in that Reality chooses from the closed interval $[-1, 1]$ rather than from the two-element set $\{-1, 1\}$:

BOUNDED FORECASTING GAME WITH FORECASTS SET TO ZERO
**Players:** Skeptic, Reality
**Protocol:**

$\mathcal{K}_0 = 1$.
FOR $n = 1, 2, \ldots$:
    Skeptic announces $M_n \in \mathbb{R}$.
    Reality announces $x_n \in [-1, 1]$.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.

**Winner:** Skeptic wins if $\mathcal{K}_n$ is never negative and either (3.2) or (3.3) holds. Otherwise Reality wins.

Why do we call this a *forecasting* game? Who is forecasting what? The answer is that we have forecast Reality's move $x_n$, and the forecast is zero. This forecast has an economic meaning: Skeptic can buy $x_n$ for zero. In §3.3, we generalize by allowing forecasts different from zero, made in the course of the game. These forecasts will also serve as prices for Skeptic.

Because the generalization from the fair-coin game to the bounded forecasting game with forecasts at zero involves only an enlargement of Reality's move space, the reformulation in terms of collateral duties and the concept of forcing developed in the preceding section apply here as well. We refrain from repeating the definitions.

The game-theoretic strong law of large numbers says that Skeptic can always win this game; if Reality is committed to avoiding making him infinitely rich, then he can force her to make the average of the $x_1, x_2, \ldots$ converge to zero.

**Proposition 3.2** *Skeptic has a winning strategy in the bounded forecasting game with forecasts set to zero.*

Because Proposition 3.2 generalizes Proposition 3.1, our proof of it will establish Proposition 3.1 as well.

The proof of Proposition 3.2 will be facilitated by some additional terminology and notation.

As we explained in § 1.2, a complete sequence of moves by World is called a *path*, and the set of all paths is called the *sample space* and designated by $\Omega$. In the game at hand, the bounded forecasting game with forecasts set to zero, World consists of the single player Reality, Reality's moves always come from the interval $[-1, 1]$, and the game continues indefinitely. So $\Omega$ is the infinite Cartesian product $[-1, 1]^\infty$. Each path is an infinite sequence $x_1, x_2, \ldots$ of numbers in $[-1, 1]$.

As we said in § 1.2, any function on the sample space is a *variable*. Here this means that any function of the $x_1, x_2, \ldots$ is a variable. In particular, the $x_n$ themselves are variables.

A *situation* is a finite sequence of moves by Reality. For example, $x_1 x_2$ is the situation after Reality has chosen $x_1$ as her first move and $x_2$ as her second move. We write $\Omega^\diamond$ for the set of all situations. In the game at hand, $\Omega^\diamond$ is the set of all finite sequences of numbers from $[-1, 1]$, including the sequence of length zero, the *initial situation*, which we designate by $\Box$.

We say that the situation $s$ *precedes* the situation $t$ if $t$, as a sequence, contains $s$ as an initial segment—say $s = x_1 x_2 \ldots x_m$ and $t = x_1 x_2 \ldots x_m \ldots x_n$. We write $s \sqsubseteq t$ when $s$ precedes $t$. If $s$ is a situation and $x \in [-1, 1]$, we write $sx$ for the situation obtained by concatenating $s$ with $x$; thus if $s = x_1 \ldots x_n$, then $sx = x_1 \ldots x_n x$. If $s$ and $t$ are situations and neither precedes the other, then we say they are *divergent*. We write $|s|$ for the length of $s$; thus $|x_1 x_2 \ldots x_n| = n$. If $\xi$ is a path for Reality, say $\xi = x_1 x_2 \ldots$, we write $\xi^n$ for the situation $x_1 x_2 \ldots x_n$. We say that $s$ *begins* $\xi$ whenever $s$ is a situation, $\xi$ is a path, and $s = \xi^n$ for some $n$.

We call a real-valued function on $\Omega^\diamond$ a *process*. Any process $\mathcal{P}$ can be interpreted as a *strategy* for Skeptic; for each situation $s$, we interpret $\mathcal{P}(s)$ as the number of tickets Skeptic is to buy in situation $s$. This definition of strategy puts no constraints on Skeptic. In particular, his initial capital does not constrain him; he is allowed to borrow money indefinitely. In the games in this chapter, however, a strategy that may require borrowing money cannot be a winning strategy for Skeptic. In these games, Skeptic loses if his capital becomes negative. If he adopts a strategy that would result in a negative capital in any situation, Reality can defeat him by choosing a path that goes through that situation.

In our game, Skeptic begins with the initial capital 1, but we can also consider the capital process that would result from his beginning with any given capital $\mathcal{K}_0$, positive, negative, or zero, and following a particular strategy $\mathcal{P}$. As in §1.2, we write $\mathcal{K}^{\mathcal{P}}$ for his capital process when he begins with zero: $\mathcal{K}^{\mathcal{P}}(\Box) = 0$ and

$$\mathcal{K}^{\mathcal{P}}(x_1 x_2 \ldots x_n) := \mathcal{K}^{\mathcal{P}}(x_1 x_2 \ldots x_{n-1}) + \mathcal{P}(x_1 x_2 \ldots x_{n-1}) x_n. \tag{3.4}$$

When he uses the $\mathcal{P}$ with any other initial capital $\alpha$, his capital follows the process $\alpha + \mathcal{K}^{\mathcal{P}}$. We call a process a *martingale* if it is of the form $\alpha + \mathcal{K}^{\mathcal{P}}$—that is, if it is the capital process for some strategy and some initial capital.[1]

The capital processes that begin with zero form a linear space, for $\beta\mathcal{K}^{\mathcal{P}} = \mathcal{K}^{\beta\mathcal{P}}$ and $\mathcal{K}^{\mathcal{P}_1} + \mathcal{K}^{\mathcal{P}_2} = \mathcal{K}^{\mathcal{P}_1+\mathcal{P}_2}$. It follows that the set of all capital processes (the set of all martingales) is also a linear space.

If $\alpha_1$ and $\alpha_2$ are nonnegative numbers that add to one, and $\mathcal{P}_1$ and $\mathcal{P}_2$ are strategies, then the martingale that results from using the strategy $\alpha_1\mathcal{P}_1 + \alpha_2\mathcal{P}_2$ starting with capital 1 is given by the same convex combination of the martingales that result from using the respective strategies starting with capital 1:

$$1 + \mathcal{K}^{\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2} = \alpha_1(1 + \mathcal{K}^{\mathcal{P}_1}) + \alpha_2(1 + \mathcal{K}^{\mathcal{P}_2}). \tag{3.5}$$

We can implement the convex combination in (3.5) by dividing the initial capital 1 between two accounts, putting $\alpha_1$ in one and $\alpha_2$ in the other, and then applying the strategy $\alpha_k\mathcal{P}_k$ (which is simply $\mathcal{P}_k$ scaled down to the initial capital $\alpha_k$) to the $k$th account.

We will also find occasion to form infinite convex combinations of strategies. If $\mathcal{P}_1, \mathcal{P}_2, \ldots$ are strategies, $\alpha_1, \alpha_2, \ldots$ are nonnegative real numbers adding to one, and the sum $\sum_{k=1}^{\infty} \alpha_k \mathcal{P}_k$ converges, then the sum $\sum_{k=1}^{\infty} \alpha_k \mathcal{K}^{\mathcal{P}_k}$ will also converge (by induction on (3.4)), and $1 + \sum_{k=1}^{\infty} \alpha_k \mathcal{K}^{\mathcal{P}_k}$ will be the martingale the strategy $\sum_{k=1}^{\infty} \alpha_k \mathcal{P}_k$ produces when it starts with initial capital 1. The strategy $\sum_{k=1}^{\infty} \alpha_k \mathcal{P}_k$ starting with 1 is implemented by dividing the initial capital of 1 among a countably infinite number of accounts, with $\alpha_k$ in the $k$th account, and applying $\alpha_k\mathcal{P}_k$ to the $k$th account.

Recall that an *event* is a subset of the sample space. We say that a strategy $\mathcal{P}$ for Skeptic *forces* an event $E$ if

$$\mathcal{K}^{\mathcal{P}}(t) \geq -1 \tag{3.6}$$

for every $t$ in $\Omega^{\diamond}$ and

$$\lim_{n\to\infty} \mathcal{K}^{\mathcal{P}}(\xi^n) = \infty \tag{3.7}$$

[1] As we explained in §2.4 (p. 53), we use the word "martingale" in this way only in symmetric probability protocols. The protocol we are studying now is symmetric: Skeptic can buy $x_n$-tickets in negative as well as in positive amounts.

for every path $\xi$ not in $E$. This agrees with the definition given in §3.1; condition (3.6) says that Skeptic does not risk bankruptcy using the strategy starting with the capital 1, no matter what Reality does. We say that Skeptic *can force* $E$ if he has a strategy that forces $E$; this is the same as saying that there exists a nonnegative martingale starting at 1 that becomes infinite on every path not in $E$.

We say that $\mathcal{P}$ *weakly forces* $E$ if (3.6) holds and every path $\xi$ not in $E$ satisfies

$$\sup_n \mathcal{K}^{\mathcal{P}}(\xi^n) = \infty. \tag{3.8}$$

By these definitions, any strategy $\mathcal{P}$ for which (3.6) holds weakly forces $\sup_n \mathcal{K}^{\mathcal{P}}(\xi^n) < \infty$. We say that Skeptic *can weakly force* $E$ if he has a strategy that weakly forces $E$; this is the same as saying that there exists a nonnegative martingale starting at 1 that is unbounded on every path not in $E$.

The following lemma shows that the concepts of forcing and weak forcing are nearly equivalent.

**Lemma 3.1** *If Skeptic can weakly force $E$, then he can force $E$.*

*Proof* Suppose $\mathcal{P}$ is a strategy that weakly forces $E$. For any $C > 0$, define a new strategy $\mathcal{P}^{(C)}$ by

$$\mathcal{P}^{(C)}(s) := \begin{cases} \mathcal{P}(s) \text{ if } \mathcal{K}^{\mathcal{P}}(t) < C \text{ for all } t \sqsubseteq s \\ 0 \text{ otherwise.} \end{cases}$$

This strategy mimics $\mathcal{P}$ except that it quits betting as soon as Skeptic's capital reaches $C$. Define a strategy $\mathcal{Q}$ by

$$\mathcal{Q} := \sum_{k=1}^{\infty} 2^{-k} \mathcal{P}^{(2^k)}.$$

Then $\lim_{n\to\infty} \mathcal{K}^{\mathcal{Q}}(\xi^n) = \infty$ for every $\xi$ for which $\sup_n \mathcal{K}^{\mathcal{P}}(\xi^n) = \infty$. Since $\mathcal{K}^{\mathcal{P}} \geq -1$, $\mathcal{K}^{\mathcal{Q}} \geq -1$. Since $\sup_n \mathcal{K}^{\mathcal{P}}(\xi^n) = \infty$ for every $\xi$ not in $E$, $\lim_{n\to\infty} \mathcal{K}^{\mathcal{Q}}(\xi^n) = \infty$ for every $\xi$ not in $E$. So $\mathcal{Q}$ forces $E$. ∎

Proving Proposition 3.2 means showing Skeptic can force (3.2), and according to Lemma 3.1, it suffices to show he can weakly force (3.2). The next two lemmas will make this easy.

**Lemma 3.2** *If Skeptic can weakly force each of a sequence $E_1, E_2, \ldots$ of events, then he can weakly force $\bigcap_{k=1}^{\infty} E_k$.*

*Proof* Let $\mathcal{P}_k$ be a strategy that weakly forces $E_k$. The capital process $1+\mathcal{K}^{\mathcal{P}_k}$ is nonnegative, and in our game this implies that it can at most double on each step:

$$1 + \mathcal{K}^{\mathcal{P}_k}(x_1 \ldots x_n) \leq 2^n.$$

Since $|\mathcal{P}_k| \leq 1 + \mathcal{K}^{\mathcal{P}_k}$ (see (3.4)), we can also say that

$$|\mathcal{P}_k(x_1 \ldots x_n)| \leq 2^n$$

for all $k$, which implies that a strategy $\mathcal{Q}$ can be defined by

$$\mathcal{Q} := \sum_{k=1}^{\infty} 2^{-k} \mathcal{P}_k.$$

Since $\mathcal{P}_k$ weakly forces $E_k$, $\mathcal{Q}$ also weakly forces $E_k$. So $\mathcal{Q}$ weakly forces $\bigcap_{k=1}^{\infty} E_k$. ∎

**Lemma 3.3** *Suppose* $\epsilon > 0$. *Then Skeptic can weakly force*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i \le \epsilon \tag{3.9}$$

*and*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i \ge -\epsilon. \tag{3.10}$$

*Proof* We may suppose that $\epsilon < 1/2$. The game specifies that the initial capital is 1. Let $\mathcal{P}$ be the strategy that always buys $\epsilon\alpha$ tickets, where $\alpha$ is the current capital. Since Reality's move $x$ is never less than $-1$, this strategy loses at most the fraction $\epsilon$ of the current capital, and hence the capital process $1 + \mathcal{K}^{\mathcal{P}}$ is nonnegative. It is given by $1 + \mathcal{K}^{\mathcal{P}}(\Box) = 1$ and

$$1 + \mathcal{K}^{\mathcal{P}}(x_1 \ldots x_n) = (1 + \mathcal{K}^{\mathcal{P}}(x_1 \ldots x_{n-1}))(1 + \epsilon x_n) = \prod_{i=1}^{n} (1 + \epsilon x_i)\,.$$

Let $\xi = x_1 x_2 \ldots$ be a path such that $\sup_n \mathcal{K}^{\mathcal{P}}(x_1 \ldots x_n) < \infty$. Then there exists a constant $C_\xi > 0$ such that

$$\prod_{i=1}^{n} (1 + \epsilon x_i) \le C_\xi$$

for all $n$. This implies that

$$\sum_{i=1}^{n} \ln (1 + \epsilon x_i) \le D_\xi$$

for all $n$ for some $D_\xi$. Since $\ln(1+t) \ge t - t^2$ whenever $t \ge -\frac{1}{2}$, $\xi$ also satisfies

$$\epsilon \sum_{i=1}^{n} x_i - \epsilon^2 \sum_{i=1}^{n} x_i^2 \le D_\xi,$$

$$\epsilon \sum_{i=1}^{n} x_i - \epsilon^2 n \le D_\xi,$$

$$\epsilon \sum_{i=1}^{n} x_i \le D_\xi + \epsilon^2 n,$$

or

$$\frac{1}{n} \sum_{i=1}^{n} x_i \le \frac{D_\xi}{\epsilon n} + \epsilon$$

for all $n$ and hence satisfies (3.9). Thus $\mathcal{P}$ weakly forces (3.9). The same argument, with $-\epsilon$ in place of $\epsilon$, establishes that Skeptic can weakly force (3.10). ∎

In order to complete the proof that Skeptic can weakly force (3.2), we now simply consider the events (3.9) and (3.10) for $\epsilon = 2^{-k}$, where $k$ ranges over all natural numbers; this defines a countable number of events Skeptic can weakly force, and their intersection, which he can also weakly force (by Lemma 3.2), is (3.2).

## 3.3 WHO SETS THE PRICES?

We have formulated the bounded forecasting game in the simplest possible way. The variables $x_1, x_2, \ldots$ are all between $-1$ and 1, and they all have the same price: zero. But our proof applies equally well when each variable has a different price, say $x_n$ has the price $m_n$, provided only that both $x_n$ and $m_n$ are uniformly bounded. (In fact, it is enough that the net payoffs of the tickets, the differences $x_n - m_n$, be uniformly bounded.) Of course, we must then replace (3.2) by

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) = 0. \tag{3.11}$$

The price $m_n$ can be chosen in whatever manner we please; we require only that it be announced before Skeptic places his bet $M_n$.

As we explained in §1.1, the idea that prices can be set freely can be expressed within our framework by introducing a third player, Forecaster, who sets them. The game then takes the following form:

BOUNDED FORECASTING GAME
**Parameter:** $C > 0$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
  Forecaster announces $m_n \in [-C, C]$.
  Skeptic announces $M_n \in \mathbb{R}$.
  Reality announces $x_n \in [-C, C]$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n)$.

**Winner:** Skeptic wins if $\mathcal{K}_n$ is never negative and either (3.11) or (3.3) holds. Otherwise Reality wins.

Since Forecaster can always choose the $m_n$ to be zero, and since the bound $C$ can be 1, this game generalizes the game of the preceding section. And Proposition 3.2 generalizes as well:

**Proposition 3.3** *Skeptic has a winning strategy in the bounded forecasting game.*

The proof of Proposition 3.2 generalizes immediately to a proof of Proposition 3.3. Alternatively, any winning strategy in the game with zero prices can be adapted in an obvious way to produce a winning strategy in the game with arbitrary prices.

We can recover Proposition 3.2 from Proposition 3.3 by setting $C$ equal to 1 and requiring Forecaster to set each $m_n$ equal to zero. Imposing this constraint on Forecaster changes the game (because his move is entirely determined, he is no longer really in the game!), but this change obviously does not impair the validity of the proposition. A strategy for Skeptic that wins when his opponents, Forecaster and Reality, have complete freedom of action will obviously still win when they are constrained, partially or completely.

We introduce Forecaster, with his complete freedom of action, in order to emphasize that how the $m_n$ are selected is quite immaterial to the reasoning by which we establish the strong law of large numbers. No matter how prices are set, Skeptic has a winning strategy. This is not to say that how prices are set is unimportant to users of the law of large numbers. On the contrary, the practical significance of the law depends both on how the prices $m_n$ are determined and on how the outcomes $x_n$ are determined. And they can be determined in a variety of ways. In physics, the $m_n$ are furnished by theory, while the $x_n$ are furnished by reality. In finance, both $m_n$ and $x_n$ are determined by a market; in many cases, $m_n$ is the price of a stock at the beginning of day $n$, and $x_n$ is its price at the end of day $n$.

It is possible for the price $m_n$ to be set by a market even though $x_n$ is determined outside the market. This happens, for example, in the Iowa Electronic Markets, which generate prices for events such as the outcomes of elections. Suppose the Iowa Electronic Markets continues to organize trading in contracts for the outcomes of U.S. presidential elections into the indefinite future: each November 1 before such an election, it determines a price for a contract that pays \$1 if the Democratic candidate wins. Then the game-theoretic strong law of large numbers tells us that either we can become infinitely rich without risking more than \$1 or else the market is calibrated, in the sense that the long-term average price of the contract approaches the long-term relative frequency with which the Democratic candidates win. Of course, this is a very idealized statement; we are assuming that the system that pits Democrats against Republicans will go on forever, that money is infinitely divisible, and that we can neglect transaction costs and bid-ask spreads. But in Chapter 6 we will prove a finitary game-theoretic law of large numbers, and the other idealizing assumptions can also be relaxed.

When we say that $m_n$ is the price for $x_n$ when Skeptic makes his move $M_n$, our manner of speaking is consistent with the general definition of price given in Chapter 1 (p. 14). Because it is determined by the path $x_1, x_2, \ldots$, we are entitled to call $x_n$ a variable, and because Skeptic can buy it exactly for $m_n$, we have $\overline{\mathbb{E}}_t\, x_n = \underline{\mathbb{E}}_t\, x_n = m_n$ in the situation $t$ where Forecaster has just announced $m_n$.

In measure-theoretic probability, the strong law for a sequence $x_1, x_2, \ldots$ of variables is formulated beginning with the assumption that the variables have a joint probability distribution, and the price $m_n$ is the conditional expected value of $x_n$ given $x_1, \ldots, x_{n-1}$; the conclusion is that (3.11) holds almost surely. We will leave for Chapter 8 the formal derivation of this measure-theoretic result from Proposition 3.3, but it is intuitively obvious that our game-theoretic formulation is more powerful, in the sense that it arrives at the same conclusion (Equation (3.11) holds almost surely) with fewer assumptions. Postulating a joint probability distribution for $x_1, x_2, \ldots$ amounts to assuming that for all $n$, every measurable function of $x_n$ (and even of $x_n, x_{n+1}, \ldots$) is priced conditional on the outcomes $x_1, \ldots, x_{n-1}$. But the game-theoretic formulation assumes only that a price for $x_n$ itself is given in light of $x_1, \ldots, x_{n-1}$. In the simplest case, where each $x_n$ has only two possible values, heads or tails, there is no difference between pricing the $x_n$ and pricing all measurable functions of it. But when each $x_n$ can be chosen from a large range of values, the difference is immense, and consequently the game-theoretic result is much more powerful than the measure-theoretic result.

As we pointed out earlier, Proposition 3.3 continues to be true when the bounded forecasting game is modified by a restriction, partial or complete, on the freedom of action of Skeptic's opponents. When we constrain Forecaster by setting the $m_n$ equal to a common value $m$ at the beginning of the game, we obtain a game-theoretic generalization of the measure-theoretic strong law for the case where the $x_1, x_2, \ldots$ are independent random variables with a common mean $m$. If we then constrain Reality to choose $x_n$ from the set $\{0, 1\}$, we obtain the game-theoretic result corresponding to the measure-theoretic strong law for a possibly biased coin; if 1 represents heads and 0 represents tails, then $m$ corresponds to the probability of heads on each toss, and (3.11) says that $\sum_{i=1}^{n} x_i$, the proportion of heads in the first $n$ tosses, converges to $m$. If $m = 1/2$, then we are back to the fair-coin game with which we began the chapter, except that we are using 0 rather than $-1$ to represent heads.

## 3.4 ASYMMETRIC BOUNDED FORECASTING GAMES

Our bounded strong law of large numbers, Equation (3.11), can be decomposed into two parts:

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) \leq 0, \tag{3.12}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) \geq 0. \tag{3.13}$$

Moreover, the proof of Lemma 3.3 makes it clear that these two parts depend on different assumptions. If Skeptic can buy tickets at the prices $m_n$, then either (3.12) will hold or else he can become infinitely rich. If Skeptic can sell tickets at the prices $m_n$, then either (3.13) will hold or else he can become infinitely rich.

We can express this point more formally by adapting the game of the preceding section as follows:

BOUNDED UPPER FORECASTING GAME

**Parameter:** $C > 0$

**Players:** Forecaster, Skeptic, Reality

**Protocol:**

$\mathcal{K}_0 := 1$.

FOR $n = 1, 2, \ldots$:

Forecaster announces $m_n \in [-C, C]$.

Skeptic announces $M_n \geq 0$.

Reality announces $x_n \in [-C, C]$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n)$.

**Winner:** Skeptic wins if and only if $\mathcal{K}_n$ is never negative and either (3.12) or (3.3) holds. Otherwise Reality wins.

This is an asymmetric protocol, in the sense explained on p. 11. Using only the first half of Lemma 3.3, we obtain our usual result:

**Proposition 3.4** *Skeptic has a winning strategy in the bounded upper forecasting game.*

The hypothesis of the impossibility of a gambling system, as applied to this game, says that the prices $m_n$ are high enough that Skeptic cannot get infinitely rich by buying tickets. If we adopt this hypothesis, then we may conclude that (3.12) will hold for these prices and, a fortiori, for any higher prices.

We can similarly define a bounded lower forecasting game, in which (1) Skeptic selects a nonpositive rather than a nonnegative real number, and (2) Skeptic wins if his capital remains nonnegative and either (3.13) or (3.3) holds. Again Skeptic will have a winning strategy.

## 3.5 APPENDIX: THE COMPUTATION OF STRATEGIES

The scope of this book is limited to showing how the game-theoretic framework can handle traditional questions in probability theory (Part I) and finance theory (Part II). But our results raise many new questions, especially questions involving computation. All our theoretical results are based on the explicit construction of strategies, and it should be both interesting and useful to study the computational properties of these constructions.

All the strategies we construct are in fact computable. For example, the construction in §3.2 is obviously computable, and hence we can strengthen Proposition 3.3 to the following:

**Proposition 3.5** *Skeptic has a computable winning strategy in the bounded forecasting game.*

This strengthening is relevant to points we have already made. For example, our argument on p. 71 concerning the Iowa Electronic Markets obviously requires that Skeptic's strategy be computable.

Proposition 3.5 is mathematically trivial, but it suggests many nontrivial questions. For example, fixing a computational model (such as the one-head and one-tape Turing machine), we can ask questions such as this:

> Does there exist a winning strategy for Skeptic in the fair-coin game such that the move at step $n$ can be computed in time $O(n^c)$, for some $c$? If yes, what is the infimum of such $c$?

Similar questions, which may be of practical interest when one undertakes to implement the game-theoretic approach, can be asked about other computational resources, such as the required memory.

In a different direction, we can ask about the rate at which Skeptic can increase his capital if the sequence of outcomes produced by Reality in the bounded forecasting game does not satisfy (3.11). For example, the construction in §3.2 shows that the following is true.

**Proposition 3.6** *Skeptic has a computable winning strategy in the bounded forecasting game with the condition (3.3) that his capital tends to infinity replaced by the condition*

$$\limsup_{n\to\infty} \frac{\log \mathcal{K}_n}{n} > 0$$

*that his capital increases exponentially fast.*

It might be interesting to study the trade-off (if any) between the computational efficiency of a strategy and the rate at which its capital tends to infinity.

Questions similar to that answered in Proposition 3.6 have been asked in algorithmic probability theory: see Schnorr (1970, 1971) in connection with the strong law of large numbers and Vovk (1987) in connection with the law of the iterated logarithm and the recurrence property.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 4

# *Kolmogorov's Strong Law of Large Numbers*

Andrei Kolmogorov (1903–1987)

In 1930, Kolmogorov proved a measure-theoretic strong law for possibly unbounded variables. In this chapter, we reformulate this strong law and its proof in game-theoretic terms, just as we reformulated the bounded strong law in the preceding chapter.

In our game for Kolmogorov's strong law, Reality announces a sequence $x_1, x_2, \ldots$ of real numbers. There is no bound on how large Reality can make $x_n$, but just before she announces $x_n$, Forecaster announces $m_n$ and $v_n$ and makes two offers to Skeptic:

- Skeptic can buy $x_n$ for $m_n$, and
- Skeptic can buy $(x_n - m_n)^2$ for $v_n$.

At this point, the numbers $m_n$ and $v_n$ are the game-theoretic price and variance, respectively, for $x_n$. Our game-theoretic version of Kolmogorov's strong law says that Skeptic can force Reality to satisfy

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) = 0 \tag{4.1}$$

if

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty. \tag{4.2}$$

Condition (4.2) is necessary as well as sufficient; as we will see, Skeptic can force (4.1) if and only if Forecaster makes his $v_n$ satisfy (4.2).

We use "Skeptic can force $E$" in the same way here as in Chapter 3. It has the same meaning as "$E$ happens almost surely": Skeptic has a strategy that makes him infinitely rich without risk of bankruptcy if $E$ does not happen. As always, the existence of such a strategy acquires practical meaning in a particular instance where the game is played only if we adopt the fundamental interpretative hypothesis for that instance, in which case the practical meaning is that we expect $E$ to happen. But in this chapter we consider only the mathematical question of the existence of such a strategy, leaving aside issues of practical interpretation.

We put no restrictions on how Forecaster chooses $m_n$ and $v_n$ in the course of the game. Therefore (4.2), a condition on the $v_n$, cannot be verified at the beginning of the game. But our claim that (4.2) is necessary and sufficient for Skeptic to be able to force (4.1) can be made precise as follows:

- Skeptic has a strategy that always keeps his capital nonnegative and guarantees that if Forecaster obeys (4.2), then either Reality obeys (4.1) or Skeptic becomes infinitely rich.

- The condition that Forecaster obey (4.2) cannot be relaxed. In fact, Reality has a strategy that guarantees that if Forecaster violates (4.2) and Skeptic's capital is always nonnegative, then Skeptic does not become infinitely rich and (4.1) fails.

We prove these assertions in §4.2 and §4.3, respectively. The two proofs differ somewhat in their logical status. We construct an explicit strategy for Skeptic, using ideas from the usual proof of Kolmogorov's strong law. But we give only a very nonconstructive proof of the existence of the strategy for Reality, a proof that combines Kolmogorov's counterexample to (4.1) when (4.2) fails with Martin's theorem, which says that quasi-Borel games are determinate. Kolmogorov's example yields a randomized strategy for Reality that wins with probability one. This implies that Reality can sometimes win and hence that Skeptic does not have a strategy that always wins. Martin's theorem tells us that there is necessarily a winning strategy for one of the two players, and so we conclude that Reality has one.

In §4.4, we relax the assumption that Skeptic can both buy and sell $x_n$ at the price $m_n$ to the assumption that he can only buy at that price or only sell at that price. The result is the same as we found in the bounded case: one-sided prices produce a one-sided bound on the oscillation of the average deviation rather than convergence. In §4.5, we show that the game-theoretic version of Kolmogorov's strong law implies analogous strong laws for any symmetric probability protocol and for many asymmetric probability protocols. In an appendix, §4.6, we review the statement of Martin's theorem.

Our game-theoretic version of Kolmogorov's strong law is more powerful than any measure-theoretic version, because it does not assume that the variables $x_1, x_2, \ldots$ are governed by a probability distribution. Kolmogorov assumed that the variables are independent with means $m_1, m_2, \ldots$ and variances $v_1, v_2, \ldots$, respectively and concluded that (4.1) holds except on a set of measure zero if (4.2) holds. He also proved a converse: given a sequence of nonnegative numbers $v_1, v_2, \ldots$ violating (4.2) and any sequence of numbers $m_1, m_2, \ldots$, we can construct a probability distribution for a sequence $x_1, x_2, \ldots$ of independent random variables with means $m_1, m_2, \ldots$ and variances $v_1, v_2, \ldots$ such that (4.1) is violated except on a set of measure zero. The modern measure-theoretic version of his theorem relaxes the assumption of independence in the first statement to the assumption that $m_n$ and $v_n$ are the mean and variance of $x_n$ conditional on all the information available before time $n$, including knowledge of $x_1, x_2, \ldots, x_{n-1}$ (a precise statement is given on p. 171). The game-theoretic result established in this chapter (Statement 1 of Theorem 4.1) assumes still less. It assumes only the existence of prices $m_1, m_2, \ldots$ and $v_1, v_2, \ldots$, not the existence of an entire probability measure for $x_1, x_2, \ldots$. Because it assumes less, it implies the measure-theoretic result (see Corollary 8.1 on p. 171).

## 4.1 TWO STATEMENTS OF KOLMOGOROV'S STRONG LAW

We now state our game-theoretic version of Kolmogorov's strong law precisely. We do this in two slightly different ways. First, we consider a game that Skeptic wins if (4.1) happens; here the game-theoretic strong law says that Skeptic can win if Forecaster obeys (4.2), while Reality can win if Forecaster violates (4.2). Then we formulate two different games, one that Skeptic can win playing against Reality and Forecaster together, and one that Reality can win playing against Skeptic and Forecaster together.

### The Unbounded Game

Our game-theoretic version of Kolmogorov's strong law says that Skeptic has a strategy that wins the following game if Forecaster makes his $v_n$ satisfy (4.2).

UNBOUNDED FORECASTING GAME
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $m_n \in \mathbb{R}$ and $v_n \geq 0$.
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \geq 0$.
Reality announces $x_n \in \mathbb{R}$.
$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n) + V_n((x_n - m_n)^2 - v_n). \tag{4.3}$$
**Winner:** Skeptic wins if $\mathcal{K}_n$ is never negative and either $\mathcal{K}_n$ tends to infinity or else (4.1) holds. Otherwise Reality wins.

This resembles the bounded forecasting game we studied in §3.3. Forecaster again has the task of forecasting Reality's move $x_n$. But now he does so with two numbers instead of one:

- $m_n$ is his forecast of $x_n$, and
- $v_n$ is his forecast of $(x_n - m_n)^2$, the squared error in his forecast of $x_n$.

Skeptic may use both forecasts as prices. He may buy any number (positive, negative, or zero) of $x_n$-tickets and any nonnegative number of $(x_n - m_n)^2$-tickets. We write

- $M_n$ for the number of $x_n$-tickets he buys, and
- $V_n$ for the number of $(x_n - m_n)^2$-tickets he buys.

Skeptic's total net gain is the sum of his net gains from the two kinds of tickets:

- $M_n(x_n - m_n)$ from the $x_n$-tickets, and
- $V_n((x_n - m_n)^2 - v_n)$ from the $(x_n - m_n)^2$-tickets.

So (4.3) is his capital after Reality announces $x_n$.

Here is one way of saying what Skeptic and Reality can each achieve in the unbounded forecasting game.

**Proposition 4.1**

1. *Skeptic has a strategy that assures that he wins if (4.2) holds.*
2. *Reality has a strategy that assures that she wins if (4.2) fails.*

Statement 1 of Proposition 4.1 generalizes Proposition 3.2, which asserts that Skeptic has a winning strategy in the bounded forecasting game. This becomes clear if we consider the following constraints on Reality and Forecaster:

- Reality must choose $x_n$ from $[-C, C]$.
- Forecaster must choose $m_n$ from $[-C, C]$ and must set each $v_n$ equal to $10C^2$.

Constraining Reality and Forecaster in this way can only make it easier for Skeptic to win the game; so Statement 1 will remain true. But now Skeptic will never buy $(x_n - m_n)^2$-tickets, because he would necessarily lose money on them. So we can remove them from the description of the game, reducing it to the bounded forecasting game. And since Equation (4.2) is satisfied when the $v_n$ are all equal to $10C^2$, we can drop it as a condition in Statement 1, which thereby becomes identical with Proposition 3.2.

The rule for determining the winner of our unbounded forecasting game can be reformulated in terms of collateral duties, in the same way as in the bounded forecasting game. Keeping $\mathcal{K}_n$ nonnegative is Skeptic's collateral duty, and keeping $\mathcal{K}_n$ from tending to infinity is Reality's collateral duty. A player loses as soon as he or she fails to perform his or her collateral duties. If both players perform their collateral duties, then Skeptic wins if (4.1) holds, and Reality wins if it fails.

### The Unbounded Protocol

By dropping the particular goal (4.1) from our unbounded forecasting game, we obtain this protocol:

UNBOUNDED FORECASTING
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $m_n \in \mathbb{R}$ and $v_n \geq 0$.
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \geq 0$.
Reality announces $x_n \in \mathbb{R}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n) + V_n((x_n - m_n)^2 - v_n)$.
**Collateral Duties:** Skeptic must keep $\mathcal{K}_n$ nonnegative. Reality must keep $\mathcal{K}_n$ from tending to infinity.

We can make this protocol back into a game by providing Skeptic with some other goal—some other event $E$. The complement $E^c$ of $E$ then becomes Reality's goal. Skeptic wins if he performs his collateral duty and either $E$ happens or Reality fails to perform her collateral duty. Reality wins if either (1) Skeptic fails to perform his collateral duty or (2) Reality performs her collateral duty and her goal $E^c$ happens.

What exactly is an event $E$ for this protocol? In general, as we learned in §1.2, an event is a subset of the sample space and so consists of paths formed by possible sequences of moves by World. Here World consists of Reality together with Forecaster, and a path of their moves is a sequence $m_1 v_1 x_1 m_2 v_2 x_2 \ldots$, with the $v_n$ all nonnegative. So an event $E$ is any set of sequences of this form. If we take $E$ to be the set of sequences $m_1 v_1 x_1 m_2 v_2 x_2 \ldots$ that satisfy (4.1), then we are back to the unbounded forecasting game with which we began.

We can speak of both Skeptic and Reality forcing events. A strategy for Skeptic *forces* an event $E$ if it is a winning strategy for Skeptic when $E$ is Skeptic's goal. A strategy for Reality *forces* $E$ if it is a winning strategy for Reality when $E$ is her goal. We say that a player, Skeptic or Reality, *can force* $E$ whenever he or she has a strategy that forces $E$. Notice that if one of the two players can force $E$, then the other cannot force $E^c$.

With these definitions, Proposition 4.1 can be strengthened to the following:

**Theorem 4.1**

1. *Skeptic can force*

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \Longrightarrow \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) = 0.^{1} \tag{4.4}$$

[1] Here we use $\Longrightarrow$ for material implication: if $E$ and $F$ are events, then $E \Longrightarrow F$ is the event that $F$ holds if $E$ holds. Since these events are sets (of sequences of moves by Forecaster and Reality), we can write $(E \Longrightarrow F) = E^c \cup F$.

2. *Reality can force*

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} = \infty \Longrightarrow \left( \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) \textit{ does not converge to } 0 \right). \tag{4.5}$$

Statement 1 of Theorem 4.1 is completely equivalent to Statement 1 of Proposition 4.1. But Statement 2 of Theorem 4.1 is stronger than Statement 2 of Proposition 4.1. The strategy for Reality mentioned in Statement 2 of Proposition 4.1 need not accomplish anything when (4.2) holds, but a strategy for Reality that validates Statement 2 of Theorem 4.1 must satisfy Reality's collateral duty whether (4.2) holds or not.

We can imagine that Forecaster is doing the bidding of Reality when we interpret Statement 1 of Theorem 4.1; since Skeptic can force (4.4) no matter what the other players do, he can do so even if Reality is allowed to set the prices as well as determine the outcomes. Similarly, we can think of Forecaster as doing the bidding of Skeptic when we interpret Statement 2; Reality can force (4.5) even if Skeptic is allowed to set the prices.

Statement 2 of Theorem 4.1 is equivalent, in light of Martin's Theorem, to the statement that Skeptic does not have a winning strategy in the game in which Reality has (4.5) as her goal, even if he is allowed to set the prices. In other words, he cannot force

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} = \infty \;\;\&\;\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) = 0 \tag{4.6}$$

even if he is allowed to set the prices. We can restate Theorem 4.1 as a whole by saying that Skeptic can force (4.4) even if Reality sets the prices but cannot force (4.6) even if he sets the prices himself.

The unbounded forecasting protocol, as we have formulated it, requires both $v_n$ and $V_n$ to be nonnegative. No generality is lost by requiring that $v_n$ be nonnegative; it is a forecast of the nonnegative quantity $(x_n - m_n)^2$, and if it is negative Skeptic can make as much money as he wants. The requirement that $V_n$ be nonnegative requires more explanation; it means that Skeptic is allowed only to buy $(x_n - m_n)^2$-tickets, not to sell them. He can bet that $v_n$ is too low a price, but not that it is too high a price. This means that $v_n$ is only the upper price for $(x_n - m_n)^2$—only the upper variance for $x_n$. In fact, Theorem 4.1 remains true if we change the protocol by allowing Skeptic to make $V_n$ negative, so that $v_n$ becomes the game-theoretic variance for $x_n$ rather than merely the game-theoretic upper variance. We have chosen to state the theorem in the form in which the $v_n$ are merely upper variances because this broadens its applicability to include practical problems where it is much more reasonable to set upper bounds on squared errors than to set two-sided prices for them.

We will prove both forms of the theorem. Our proof of Statement 1, in §4.2, assumes that Skeptic must make $V_n$ nonnegative, but it also applies when this restriction is lifted, because increasing Skeptic's freedom cannot make it harder for him to have a winning strategy. Similarly, our proof of Statement 2, in §4.3, assumes that Skeptic's choice of the $V_n$ is not restricted, and it then applies to the other case as well, because limiting Skeptic's freedom cannot make it harder for Reality to have a winning strategy.

## 4.2 SKEPTIC'S STRATEGY

We now prove Statement 1 of Theorem 4.1 by constructing the required strategy for Skeptic. Our proof is similar to the usual proofs of Kolmogorov's strong law; it uses a form of Doob's martingale convergence theorem.

For simplicity, we suppose that Forecaster is required to set all the $m_n$ to zero. This involves no loss of generality, because any strategy for Skeptic that forces

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0 \tag{4.7}$$

when Forecaster is required to set the $m_n$ equal to zero can easily be adapted to yield a strategy that forces (4.4) when Forecaster is free to make the $m_n$ nonzero.

A *path* is now a sequence of numbers $v_1 x_1 v_2 x_2 \ldots$ with the $v_n$ nonnegative. As usual, an *event* is a set of paths, and a *variable* is a function on the paths. A *situation* is a finite sequence of numbers $v_1 x_1 \ldots v_n x_n$ with all $v_n$ nonnegative, and a *process* is a function on the situations. If $\xi$ designates a path $v_1 x_1 v_2 x_2 \ldots$ and $n$ is a nonnegative integer, then $\xi^n$ designates the situation $v_1 x_1 \ldots v_n x_n$. We extend situations freely by appending additional numbers or sequences of numbers; whenever $a$ and $b$ designate numbers or finite sequences of numbers, $ab$ designates the result of concatenating the two with $b$ on the right.

Following the usual practice in probability theory, we sometimes identify a process $\mathcal{S}$ with the sequence $\mathcal{S}_0, \mathcal{S}_1, \ldots$ of variables defined by

$$\mathcal{S}_n(\xi) := \mathcal{S}(\xi^n).$$

The first of these variables, $\mathcal{S}_0$, is always constant:

$$\mathcal{S}_0(\xi) = \mathcal{S}(\xi^0) = \mathcal{S}(\Box)$$

for all $\xi$.

A process $\mathcal{A}$ is *predictable* if for every integer $n$ greater than or equal to one, $\mathcal{A}_n(v_1 x_1 v_2 x_2 \ldots)$ does not depend on $x_n$. When we are dealing with a predictable process $\mathcal{A}$, we will sometimes abbreviate $\mathcal{A}(v_1 x_1 \ldots v_n x_n)$ to $\mathcal{A}(v_1 x_1 \ldots v_n)$. A strategy for Skeptic consists of two predictable processes, a process $\mathcal{M}$ that specifies the number $\mathcal{M}(v_1 x_1 \ldots v_n)$ of $x_n$-tickets Skeptic buys, and a process $\mathcal{V}$ that specifies the number $\mathcal{V}(v_1 x_1 \ldots v_n)$ of $x_n^2$-tickets he buys. A pair $(\mathcal{M}, \mathcal{V})$ of predictable processes qualifies as a strategy for Skeptic provided only that $\mathcal{V}$ is nonnegative.[2]

[2]Representing a strategy for Skeptic in this way, as a pair of predictable processes $(\mathcal{M}, \mathcal{V})$, is awkward in one respect: the values $\mathcal{M}(\Box)$ and $\mathcal{V}(\Box)$ are superfluous; we can specify them arbitrarily since they do not specify moves for Skeptic. Notice also that this way of representing a strategy differs from the approach we used in the simpler setting of §3.2, where there was no Forecaster. There the value $\mathcal{P}(s)$ of a strategy $\mathcal{P}$ was the move by Skeptic that combined with Reality's move to produce Skeptic's capital in the situation following $s$. Now $(\mathcal{M}(s), \mathcal{V}(s))$ is the move by Skeptic that combines with Reality's move (and Forecaster's preceding move) to produce Skeptic's capital in the situation $s$ itself.

Skeptic's capital process is determined by his initial stake and strategy according to Equation (4.3). If the initial stake is 0 and the strategy is $\mathcal{P} = (\mathcal{M}, \mathcal{V})$, then the capital process is $\mathcal{K}^{\mathcal{P}}$, where

$$\mathcal{K}^{\mathcal{P}}(\Box) := 0$$

and

$$\mathcal{K}^{\mathcal{P}}(v_1 x_1 \dots v_n x_n) := \mathcal{K}^{\mathcal{P}}(v_1 x_1 \dots v_{n-1} x_{n-1}) + \mathcal{M}(v_1 x_1 \dots v_n) x_n + \mathcal{V}(v_1 x_1 \dots v_n)(x_n^2 - v_n).$$

If the initial stake is $\alpha$ instead of 0, then the capital process is $\alpha + \mathcal{K}^{\mathcal{P}}$.

In the bounded forecasting game, we called capital processes martingales (see p. 67). Here we do not call them martingales, because they lack an important property that is often associated with the word "martingale": the fact that $\mathcal{S}$ is a capital process does not imply that $-\mathcal{S}$ is also a capital process. This is because the protocol is asymmetric: Skeptic cannot make $V_n$ negative. In §4.5, where we derive Kolmogorov's strong law for an abstract symmetric protocol, we will again call capital processes "martingales".

**Lemma 4.1** *The capital processes form a convex cone: if $\mathcal{S}^1, \dots, \mathcal{S}^K$ are capital processes and $c_1 \geq 0, \dots, c_K \geq 0$, then $c_1\mathcal{S}^1 + \cdots + c_K\mathcal{S}^K$ is a capital process.*

*Proof* If $\mathcal{S}^k$ is the capital process resulting from the initial stake $\alpha_k$ and the strategy $(\mathcal{M}^k, \mathcal{V}^k)$, $k = 1, \dots, K$, then $c_1\mathcal{S}^1 + \cdots + c_K\mathcal{S}^K$ is the capital process resulting from the initial stake $c_1\alpha_1 + \cdots + c_K\alpha_K$ and the strategy

$$(c_1\mathcal{M}^1 + \cdots + c_K\mathcal{M}^K, c_1\mathcal{V}^1 + \cdots + c_K\mathcal{V}^K).$$ ∎

A *supermartingale* is a process of the form $\mathcal{S} - \mathcal{B}$, where $\mathcal{S}$ is a capital process and $\mathcal{B}$ is an increasing process. (A process $\mathcal{B}$ is *increasing* if $\mathcal{B}_n(\xi) \leq \mathcal{B}_{n+1}(\xi)$ for all $n$ and all $\xi$. A process that is identically equal to zero qualifies as increasing, and hence a capital process itself qualifies as a supermartingale.) We can think of $\mathcal{S} - \mathcal{B}$ as the capital for Skeptic were he to follow the strategy that produces $\mathcal{S}$ but were also to throw money away on each round; $\mathcal{B}$ records the cumulative amount of money thrown away. If we write $\mathcal{T}$ for $\mathcal{S} - \mathcal{B}$, then the fact that $\mathcal{B}$ is increasing can be expressed by the inequality

$$\mathcal{T}_{n+1}(\xi) - \mathcal{T}_n(\xi) \leq \mathcal{S}_{n+1}(\xi) - \mathcal{S}_n(\xi),$$

which holds for all $n$ and all $\xi$; the increments for $\mathcal{T}$ are smaller than the increments for $\mathcal{S}$ because of the money thrown away. We call $\mathcal{S}$ a *bounding capital process* for the supermartingale $\mathcal{T}$, and we call a strategy $(\mathcal{M}, \mathcal{V})$ that generates $\mathcal{S}$ (with the help of some initial stake) a *bounding strategy.*

If a nonnegative supermartingale $\mathcal{T}$ tends to infinity on a path, then any bounding capital process with initial stake $\mathcal{T}(\Box)$ is also nonnegative and also tends to infinity on that path. So we can substitute nonnegative supermartingales for nonnegative capital processes in the definition of "almost surely": An event $E$ happens almost surely if and only if there exists a nonnegative supermartingale that tends to infinity on every path in $E^c$. We also say that any such nonnegative supermartingale is a *witness* of the almost certain happening of $E$.

**Lemma 4.2** *The supermartingales form a convex cone: if $\mathcal{T}^1, \ldots, \mathcal{T}^K$ are supermartingales and $c_1 \geq 0, \ldots, c_K \geq 0$, then $c_1\mathcal{T}^1 + \cdots + c_K\mathcal{T}^K$ is a supermartingale.*

*Proof* If $\mathcal{T}^k = \mathcal{S}^k - \mathcal{B}^k$, where the $\mathcal{S}^k$ are martingales and the $\mathcal{B}^k$ are increasing, then

$$c_1\mathcal{T}^1 + \cdots + c_K\mathcal{T}^K = (c_1\mathcal{S}^1 + \cdots + c_K\mathcal{S}^K) - (c_1\mathcal{B}^1 + \cdots + c_K\mathcal{B}^K). \qquad \blacksquare$$

Given a supermartingale $\mathcal{T}$ we define a predictable process $\|\mathcal{T}\|$ by

$$\begin{aligned} \|\mathcal{T}\|(sv) := \inf\{\|(M,V)\| :\ & V \text{ is nonnegative, and} \\ \mathcal{T}(svx) - \mathcal{T}(s) \leq Mx + V(x^2 - v) & \text{ for all } x \in \mathbb{R}\} \end{aligned} \tag{4.8}$$

for every situation $s$ and every nonnegative number $v$. Here $\|\cdot\|$ is the usual norm for two-dimensional Euclidean space: $\|(M,V)\| = \sqrt{M^2 + V^2}$.[3] We say that a bounding strategy $(\mathcal{M}, \mathcal{V})$ for $\mathcal{T}$ is *minimal* if $\|\mathcal{M}(s), \mathcal{V}(s)\| = \|\mathcal{T}\|(s)$ for all $s \neq \Box$.

**Lemma 4.3** *Every supermartingale has a minimal bounding strategy.*

*Proof* It suffices to show that the infimum over $M$ and nonnegative $V$ in (4.8) is attained for fixed $s$ and $v$. Let $\mu(x)$ denote the closed subset of $\mathbb{R}^2$ consisting of all $(M, V)$, with $V$ nonnegative, that satisfy the inequality in (4.8) for $x$. Since $\mathcal{T}$ has at least one bounding strategy, the intersection of the $\mu(x)$ over all $x$ is nonempty. As the intersection of closed sets, it is closed. So it has an element closest to the origin. $\blacksquare$

**Lemma 4.4** *Suppose $\mathcal{T}^1$, $\mathcal{T}^2$,... is a sequence of nonnegative supermartingales and $c_1$, $c_2$,... is a sequence of nonnegative numbers such that*

$$\sum_{k=1}^{\infty} c_k \mathcal{T}_0^k < \infty \tag{4.9}$$

*and*

$$\sum_{k=1}^{\infty} c_k \|\mathcal{T}^k\|(s) < \infty \tag{4.10}$$

*for every situation $s$. Set*

$$\mathcal{T} := \sum_{k=1}^{\infty} c_k \mathcal{T}^k.$$

*Then $\mathcal{T}$ is a nonnegative supermartingale and*

$$\|\mathcal{T}\| \leq \sum_{k=1}^{\infty} c_k \|\mathcal{T}^k\|. \tag{4.11}$$

[3] The choice of this norm is arbitrary; many other norms would work just as well. We may also specify $\|\mathcal{T}\|$'s value at $\Box$ arbitrarily; say $\|\mathcal{T}\|(\Box) := 0$.

*Proof* Choose a minimal bounding strategy $(\mathcal{M}^k, \mathcal{V}^k)$ for each $\mathcal{T}^k$ and set

$$\mathcal{M} := \sum_{k=1}^{\infty} c_k \mathcal{M}^k; \ \mathcal{V} := \sum_{k=1}^{\infty} c_k \mathcal{V}^k.$$

(It is clear from (4.10) that the series defining $\mathcal{M}$ and $\mathcal{V}$ converge to finite numbers at every situation.)

Let us prove that $\mathcal{T}_n < \infty$ for every $n$; since $\mathcal{T}_0 < \infty$ by (4.9), it is sufficient to prove that $\mathcal{T}_{n+1} < \infty$ assuming $\mathcal{T}_n < \infty$, where $n = 1, 2, \ldots$. Summing the inequality

$$\mathcal{T}^k(svx) - \mathcal{T}^k(s) \leq \mathcal{M}^k(s)x + \mathcal{V}^k(s)(x^2 - v)$$

(cf. (4.8)) multiplied by $c_k$ over $k$, we obtain

$$\mathcal{T}(svx) - \mathcal{T}(s) \leq \mathcal{M}(sv)x + \mathcal{V}(sv)(x^2 - v), \tag{4.12}$$

which shows that $\mathcal{T}(svx)$ is finite as soon as $\mathcal{T}(s)$ is finite.

Equation (4.12) also shows that $(\mathcal{M}, \mathcal{V})$ bounds $\mathcal{T}$, and so equation (4.11) follows from

$$\begin{aligned} \|\mathcal{T}\|(sv) &\leq \|(\mathcal{M}(sv), \mathcal{V}(sv))\| = \left\| \sum_{k=1}^{\infty} c_k (\mathcal{M}^k(sv), \mathcal{V}^k(sv)) \right\| \\ &\leq \sum_{k=1}^{\infty} c_k \left\| (\mathcal{M}^k(sv), \mathcal{V}^k(sv)) \right\| = \sum_{k=1}^{\infty} c_k \|\mathcal{T}^k\|(sv). \end{aligned}$$

∎

Now our argument will follow Liptser's proof of Kolmogorov's strong law [287]. The next step is to establish a version of Doob's convergence theorem for our game-theoretic nonnegative supermartingales. Doob's theorem asserts that a nonnegative measure-theoretic supermartingale converges almost surely in the sense of measure theory—that is, except on a set of measure zero. Our version asserts that a nonnegative game-theoretic supermartingale converges almost surely in our sense—that is, there is another nonnegative game-theoretic supermartingale that tends to infinity on all paths where the convergence fails.

**Lemma 4.5** *If $\mathcal{T}$ is a nonnegative supermartingale, then $\mathcal{T}_n$ converges*[4] *almost surely.*

*Proof [Doob]* We will prove the lemma by constructing a witnessing nonnegative supermartingale—a nonnegative supermartingale $\mathcal{T}^*$ that tends to infinity on every path where $\mathcal{T}$ does not converge.

Choose a minimal bounding strategy $\mathcal{P} = (\mathcal{M}, \mathcal{V})$ for $\mathcal{T}$. Let $a$, $b$ be positive rational numbers such that $a < b$. Set $\tau_0 := 0$ and, for $k = 1, 2, \ldots$, set

$$\sigma_k := \min\{i > \tau_{k-1} : \mathcal{T}_i > b\}, \ \tau_k := \min\{i > \sigma_k : \mathcal{T}_i < a\}.$$

Let $\mathcal{P}^{a,b}$ be the strategy given by

$$\mathcal{P}_i^{a,b} := \begin{cases} \mathcal{P}_i & \text{if } \exists k \colon \tau_{k-1} < i \leq \sigma_k \\ (0,0) & \text{otherwise,} \end{cases}$$

[4] We say a sequence *converges* when it has a finite limit.

and let $\mathcal{T}^{a,b}$ be the nonnegative martingale $\mathcal{T}(\Box) + \mathcal{K}^{\mathcal{P}^{a,b}}$. It is easy to see that

$$\mathcal{T}_0^{a,b} = \mathcal{T}_0, \ \|\mathcal{T}^{a,b}\| \leq \|\mathcal{T}\|,$$

and always

$$\left.\begin{array}{l}\liminf_{n\to\infty} \mathcal{T}_n < a \\ \limsup_{n\to\infty} \mathcal{T}_n > b\end{array}\right\} \Longrightarrow \lim_{n\to\infty} \mathcal{T}_n^{a,b} = \infty. \tag{4.13}$$

Arrange all such pairs $(a, b)$ in a sequence $(a_1, b_1), (a_2, b_2), \ldots$ and put

$$\mathcal{T}^* := \frac{1}{2}\mathcal{T} + \sum_{k=1}^{\infty} 2^{-k-1}\mathcal{T}^{a_k,b_k}. \tag{4.14}$$

By Lemma 4.4, $\mathcal{T}^*$ is a nonnegative supermartingale with

$$\|\mathcal{T}^*\| \leq \frac{1}{2}\|\mathcal{T}\| + \sum_{k=1}^{\infty} 2^{-k-1}\|\mathcal{T}^{a_k,b_k}\| \leq \|\mathcal{T}\|.$$

We need to show that if $\mathcal{T}_n$ does not have a finite limit, then $\mathcal{T}_n^*$ tends to infinity. But (4.14) implies that if $\mathcal{T}_n$ tends to infinity, then $\mathcal{T}_n^*$ does as well. And $\mathcal{T}_n$ having no limit at all is equivalent to

$$\exists (a, b) \in \mathbb{Q}^2 : \liminf_{n\to\infty} \mathcal{T}_n < a \ \ \& \ \ \limsup_{n\to\infty} \mathcal{T}_n > b,$$

and by (4.13) and (4.14), this implies that $\mathcal{T}_n^*$ tends to infinity. ∎

Almost certain convergence also holds for processes that can grow faster than nonnegative supermartingales but in a limited and predictable way. To explain this, we need another definition. A *semimartingale* is a process that can be written in the form $\mathcal{U} = \mathcal{T} + \mathcal{A}$, where $\mathcal{T}$ is a supermartingale and $\mathcal{A}$ is an increasing predictable process; the process $\mathcal{A}$ is called a *compensator* for $\mathcal{U}$.

**Lemma 4.6** *If $\mathcal{U}$ is a nonnegative semimartingale with $\mathcal{A}$ as a compensator, then*

$$(\mathcal{A}_\infty \textit{ is finite}) \Longrightarrow (\mathcal{U}_n \textit{ converges}) \tag{4.15}$$

*almost surely. (Here $\mathcal{A}_\infty$ is the limit, finite or infinite, of the increasing process $\mathcal{A}_n$.)*

*Proof* Set $\mathcal{T} := \mathcal{U} - \mathcal{A}$; $\mathcal{T}$ is a supermartingale. For $C = 1, 2, \ldots$, define the nonnegative supermartingales $\mathcal{T}^C$ by the requirements $\mathcal{T}_0^C = C$ and

$$\Delta\mathcal{T}_n^C = \begin{cases} \Delta\mathcal{T}_n & \text{if } \mathcal{A}_n \leq C \\ 0 & \text{otherwise,} \end{cases} \qquad n = 1, 2, \ldots \tag{4.16}$$

($\Delta a_n$ stands for $a_n - a_{n-1}$). Since $\|\mathcal{T}^C\| \leq \|\mathcal{T}\|$, Lemma 4.4 implies that

$$\mathcal{R} := \sum_{C=1}^{\infty} 2^{-C}(\mathcal{T}^C)^*, \tag{4.17}$$

where $^*$ is the transformation (4.14), witnesses that (4.15) holds almost surely. ∎

**Lemma 4.7** *Suppose $\mathcal{S}$ is a supermartingale and $\mathcal{S}^2$ is a semimartingale with compensator $\mathcal{A}$. Then*

$$(\mathcal{A}_\infty \textit{ is finite}) \Longrightarrow (\mathcal{S}_n \textit{ converges})$$

*almost surely.*

*Proof* If $\mathcal{A}$ is a compensator of $\mathcal{S}^2$, then $\mathcal{A}$ is a compensator of $(\mathcal{S}+1)^2$ as well. By Lemma 4.6, $\mathcal{S}_n^2$ and $(\mathcal{S}_n+1)^2$ converge when $\mathcal{A}_\infty < \infty$, almost surely. It remains to note that

$$\mathcal{S}_n = \frac{1}{2}\Big((\mathcal{S}_n+1)^2 - \mathcal{S}_n^2 - 1\Big).$$ ∎

To conclude our proof that Skeptic has a winning strategy, consider the capital process

$$\mathcal{S}_n(v_1x_1v_2x_2\ldots) := \sum_{i=1}^n \frac{x_i}{i}$$

and the increasing predictable process

$$\mathcal{A}_n(v_1x_1v_2x_2\ldots) := \sum_{i=1}^n \frac{v_i}{i^2}.$$

The difference

$$\mathcal{S}_n^2 - \mathcal{A}_n = \left(\sum_{i=1}^n \frac{x_i}{i}\right)^2 - \sum_{i=1}^n \frac{v_i}{i^2}$$

$$= \sum_{i=1}^n 2\left(\sum_{j=1}^{i-1}\frac{x_j}{j}\right)\frac{x_i}{i} + \sum_{i=1}^n \frac{x_i^2 - v_i}{i^2}$$

is itself a capital process, and hence $\mathcal{S}^2$ is a semimartingale with $\mathcal{A}$ as its compensator. Applying Lemma 4.7, we deduce that

$$\sum_{n=1}^\infty \frac{v_n}{n^2} < \infty \Longrightarrow \left(\sum_{n=1}^\infty \frac{x_n}{n} \text{ exists and is finite}\right)$$

almost surely. It remains only to apply Kronecker's lemma (see, e.g., [287]), which implies that

$$\left(\sum_{n=1}^\infty \frac{x_n}{n} \text{ exists and is finite}\right) \Longrightarrow \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i = 0.$$

This completes the proof.

Implicit in this proof is the construction of a nonnegative supermartingale that tends to infinity on all paths outside the event (4.4). The following proposition specifies this supermartingale explicitly.

**Proposition 4.2** *The following nonnegative supermartingale tends to infinity on all paths outside the event (4.4):*

$$\sum_{C=1}^{\infty} 2^{-C} \left(\mathcal{S}^C\right)^* + \sum_{C=1}^{\infty} 2^{-C} \left(\mathcal{T}^C\right)^*, \tag{4.18}$$

*where $\mathcal{S}$ and $\mathcal{T}$ are the supermartingales defined by*

$$\mathcal{S}_n := \left(\sum_{i=1}^{n} \frac{x_i}{i}\right)^2, \quad \mathcal{T}_n := \left(1 + \sum_{i=1}^{n} \frac{x_i}{i}\right)^2,$$

*the operation $\mathcal{T} \mapsto \mathcal{T}^C$ is defined by (4.16) with*

$$\mathcal{A}_n := \sum_{i=1}^{n} \frac{v_i}{i^2},$$

*and $^*$ is the transformation (4.14).*

*Proof* If (4.4) is violated, then

$$\sum_{i=1}^{\infty} \frac{v_i}{i^2} < \infty$$

and

$$\frac{1}{n} \sum_{i=1}^{n} x_i \text{ does not converge to } 0. \tag{4.19}$$

By Kronecker's lemma, Equation (4.19) implies

$$\sum_{i=1}^{n} \frac{x_i}{i} \text{ does not converge.} \tag{4.20}$$

As seen from the proof of Lemma 4.7, Equation (4.20) implies the disjunction of the events

$$\left(\sum_{i=1}^{n} \frac{x_i}{i}\right)^2 \text{ does not converge}$$

and

$$\left(1 + \sum_{i=1}^{n} \frac{x_i}{i}\right)^2 \text{ does not converge.}$$

Now Equation (4.17) implies that the supermartingale (4.18) witnesses that (4.4) holds almost surely. ∎

## 4.3 REALITY'S STRATEGY

We turn now to Statement 2 of Theorem 4.1: Reality has a strategy that forces (4.5). We will prove this nonconstructively, by first showing that she has a randomized strategy that wins with probability one. This implies that Skeptic and Forecaster do not have a winning strategy and hence, by Martin's theorem, the other player, Reality, does have a winning strategy. Martin's theorem applies because the event Skeptic and Forecaster are trying to make happen is quasi-Borel.

As in the preceding section, we assume without loss of generality that Forecaster always sets $m_n$ equal to zero. A strategy that always wins for Reality in this case can easily be adapted to a strategy that always win for Reality in the general case; we simply replace each move $x_n$ the strategy recommends by $x_n + m_n$.

The randomized strategy for Reality we use was devised by Kolmogorov (1930): if $v_n < n^2$,

$$x_n := \begin{pmatrix} n \\ -n \\ 0 \end{pmatrix} \text{ with probability } \begin{pmatrix} v_n/(2n^2) \\ v_n/(2n^2) \\ 1 - v_n/n^2 \end{pmatrix},$$

respectively; if $v_n \geq n^2$,

$$x_n := \begin{pmatrix} \sqrt{v_n} \\ -\sqrt{v_n} \end{pmatrix} \text{ with probability } \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}.$$

Suppose Reality plays this strategy while Skeptic plays some (nonrandomized) strategy with a nonnegative capital process $\mathcal{K}_n$. Equation (4.3) implies that $\mathcal{K}_n$ is a measure-theoretic martingale, and hence it tends to $\infty$ with measure-theoretic probability 0. In order to win against Reality, Skeptic and Forecaster must make

$$(\mathcal{K}_n \text{ tends to infinity}) \text{ or } \left( \sum_{n=1}^{\infty} \frac{v_n}{n^2} = \infty \;\&\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0 \right)$$

happen always. Since they can make $\mathcal{K}_n$ tend to infinity only with probability zero, they must make

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} = \infty \;\&\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0$$

happen with probability one. But the Borel-Cantelli-Lévy lemma (see, e.g., [287]) shows that with probability one $\sum_n v_n/n^2 = \infty$ (or, equivalently, $\sum_n \min(v_n/n^2, 1/2) = \infty$) implies that $|x_n| \geq n$ for infinitely many $n$ and hence that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0$ fails. So Skeptic and Forecaster lose with probability one.

There is one hole in our argument. Doob's inequality and the Borel-Cantelli-Lévy lemma require that the objects to which they are applied, $\mathcal{K}_n$, $v_n$, and $x_n$ in this case, be measurable. It is not quite obvious that they are, because these objects depend on Skeptic's and Forecaster's strategy, which is not assumed to be measurable. We can plug this hole simply by placing the argument in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ so simple that the required measurability cannot be avoided. We set $\Omega := \{-1, 0, 1\}^{\infty}$,

$$x_n(\omega) := \begin{cases} 0 & \text{if } \omega_n = 0 \\ \pm n & \text{if } \omega_n = \pm 1 \text{ and } v_n < n^2 \\ \pm\sqrt{v_n} & \text{if } \omega_n = \pm 1 \text{ and } v_n \geq n^2, \end{cases}$$

where $\omega = \omega_1\omega_2 \ldots \in \Omega$, $\mathcal{F}$ is the usual $\sigma$-algebra on $\Omega$, and

$$\mathbb{P}(\omega_n = 1 \mid \mathcal{F}_{n-1}) := \mathbb{P}(\omega_n = -1 \mid \mathcal{F}_{n-1}) := \begin{cases} v_n/(2n^2) & \text{if } v_n < n^2 \\ 1/2 & \text{otherwise,} \end{cases}$$

where $\mathcal{F}_n$ is the $\sigma$-algebra generated by $\omega_1, \ldots, \omega_n$. Reality's moves $x_n$ are $\mathcal{F}_n$-measurable, and the other players' moves, $v_n$, $M_n$, and $V_n$, are $\mathcal{F}_{n-1}$-measurable; therefore, $\mathcal{K}_n$ is $\mathcal{F}_n$-measurable.

## 4.4 THE UNBOUNDED UPPER FORECASTING PROTOCOL

In the preceding chapter (§3.4), we showed that the bounded strong law can be generalized to a one-sided strong law when Skeptic is allowed only to buy, not to sell, $x_n$-tickets. The unbounded strong law generalizes in the same way.

For the generalization, we modify the unbounded forecasting protocol (p. 79) by adding the requirement that $M_n \geq 0$. We call this the *unbounded upper forecasting protocol*; it allows Skeptic to buy only nonnegative amounts of $x_n$-tickets and also only nonnegative amounts of $(x_n - m_n)^2$-tickets.

**Proposition 4.3** *Skeptic can force*

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \Longrightarrow \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) \leq 0$$

*in the unbounded upper forecasting protocol.*

*Proof* As usual, we assume without loss of generality that Forecaster is required to set $m_n = 0$ for all $n$. We then need to show that Skeptic has a strategy that forces

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \Longrightarrow \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i \leq 0. \tag{4.21}$$

Suppose for the moment that Skeptic is allowed to buy negative as well as nonnegative numbers of $x_n$-tickets. We can then suppose with no loss of generality that Reality always chooses a nonnegative value for $x_n$ whenever Skeptic buys a negative number of $x_n$-tickets. By doing so, Reality both decreases Skeptic's capital and makes violating (4.21) easier, and so any strategy for Skeptic that defeats Reality when she plays under this constraint will defeat her regardless. We also know that Skeptic has a winning strategy—the one we constructed for him in §4.2. It forces (4.21) because it forces (4.7). This strategy may sometimes recommend buying a negative number of $x_n$-tickets. But since Reality will make $x_n$ nonnegative in these cases, Skeptic will do better by changing any such negative values to zero, and the resulting strategy, which will consequently still force (4.21), does qualify as a strategy in the present game. ∎

## 4.5 A MARTINGALE STRONG LAW

Theorem 4.1 is set in a very specific probability protocol: the unbounded forecasting protocol. But as we show in this section, it implies an abstract strong law of large numbers, which applies to martingales in any symmetric probability protocol and to supermartingales in a broad class of asymmetric probability protocols.

Recall that a *probability protocol* involves two players, Skeptic and World. On each round, World moves after Skeptic, and their two moves together determine a gain for Skeptic (see §1.2 and §8.3). We sometimes divide World into two players:

- Reality, who announces (when World moves, after Skeptic's $n$th move) the aspects of World's $n$th move that are relevant to Skeptic's $n$th payoff, and
- Forecaster, who announces (just before Skeptic's $n$th move, because the Skeptic has no need for the information earlier) the aspects of World's first $n-1$ moves that are relevant to Skeptic's $n$th payoff.

(See §1.1.) If we then make the simplifying assumption that Skeptic's gain function and the three players' move spaces do not change from round to round, we can describe the protocol as follows:

CONSTANT MOVE SPACES AND GAIN FUNCTION
**Parameters:** $\mathbf{F}$, $\mathbf{S}$, $\mathbf{R}$, $\lambda$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $\mathbf{f}_n \in \mathbf{F}$.
Skeptic announces $\mathbf{s}_n \in \mathbf{S}$.
Reality announces $\mathbf{r}_n \in \mathbf{R}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + \lambda(\mathbf{f}_n, \mathbf{s}_n, \mathbf{r}_n)$.
**Collateral Duties:** Skeptic must keep $\mathcal{K}_n$ nonnegative. Reality must keep $\mathcal{K}_n$ from tending to infinity.

We assume that the gain function and move spaces are constant only to make the notation manageable; our results (Propositions 4.4 and 4.5) and their proofs can be extended, merely by elaborating the notation, to the general case where the gain function and move spaces may depend on the situation (the previous moves of Forecaster and Reality).

With the notation used here, a *situation* is a finite sequence $\mathbf{f}_1\mathbf{r}_1 \ldots \mathbf{f}_n\mathbf{r}_n$. A *path* is an infinite sequence, $\mathbf{f}_1\mathbf{r}_1\mathbf{f}_2\mathbf{r}_2 \ldots$. An *event*, as usual, is a set of paths. Fixing an event $E$ as Skeptic's *goal* makes the probability protocol into a probability game: Skeptic wins if the goal happens and he performs his collateral duty. As usual, we say that a strategy for Skeptic *forces* $E$ in the protocol if it is a winning strategy in this game, and we say that $E$ happens *almost surely* if Skeptic has a strategy that forces $E$. The concepts of *supermartingale*, *predictable process*, *semimartingale*, and *compensator* apply here as well.

We begin with the symmetric case. Recall that a probability protocol is symmetric if Skeptic can take either side of any gamble (p. 11). Here this means that $\mathbf{S}$ is a linear space and $\lambda(\mathbf{f}, \mathbf{s}, \mathbf{r})$ is linear in $\mathbf{s}$. In this case, the capital processes for Skeptic also form a linear space (any initial capital is allowed). In particular, if $\mathcal{S}$ is a capital process for Skeptic, then so is $-\mathcal{S}$. We call the capital processes in a symmetric probability protocol *martingales* (cf. p. 67 and p. 82).

In a symmetric probability protocol, the compensator of the square of a martingale, when it exists, has an interpretation that generalizes the idea of quadratic variation in measure-theoretic probability. We say that an increasing predictable process $\mathcal{A}$ is a *quadratic supervariation* for a martingale $\mathcal{S}$ if there exists a martingale $\mathcal{K}$ such that

$$(\mathcal{S}(s\mathbf{fr}) - \mathcal{S}(s))^2 - (\mathcal{A}(s\mathbf{f}) - \mathcal{A}(s)) \leq \mathcal{K}(s\mathbf{fr}) - \mathcal{K}(s) \tag{4.22}$$

for all situations $s$ and all $\mathbf{f} \in \mathbf{F}$ and $\mathbf{r} \in \mathbf{R}$. The idea behind the name "quadratic supervariation" is that the right-hand side of (4.22), as the increment of a martingale, is expected to be about zero on average, so that the increment of $\mathcal{A}$ is roughly an upper bound, on average, on the square of the increment of $\mathcal{S}$. The relation between the idea of compensator and that of quadratic supervariation is made precise by the following lemma.

**Lemma 4.8** *If $\mathcal{S}$ is a martingale in a symmetric probability protocol, then the following conditions are equivalent.*

1. *$\mathcal{A}$ is a quadratic supervariation for $\mathcal{S}$.*
2. *$\mathcal{S}^2$ is a semimartingale with $\mathcal{A}$ as a compensator.*

*Proof* For the moment, let us simplify the notation in (4.22) so that it reads

$$(\mathcal{S}_+ - \mathcal{S})^2 - (\mathcal{A}_+ - \mathcal{A}) \leq \mathcal{K}_+ - \mathcal{K}. \tag{4.23}$$

Multiplying out the square, we find that this is equivalent to

$$(\mathcal{S}_+^2 - \mathcal{S}^2) - (\mathcal{A}_+ - \mathcal{A}) \leq (\mathcal{K}_+ - \mathcal{K}) + 2\mathcal{S}(\mathcal{S}_+ - \mathcal{S}). \tag{4.24}$$

The right-hand side of (4.24) is itself the increment of a capital process, which we may designate by $\mathcal{K}^*$. Indeed, if $\mathcal{P}_\mathcal{K}$ is a strategy that produces $\mathcal{K}$ and $\mathcal{P}_\mathcal{S}$ is a strategy that produces $\mathcal{S}$, then $\mathcal{K}^*$ is produced by the strategy $\mathcal{P}$, where

$$\mathcal{P}(s) = \mathcal{P}_\mathcal{K}(s) + 2\mathcal{S}(s^-)\mathcal{P}_\mathcal{S}(s) \tag{4.25}$$

for every noninitial situation $s$ ($s^-$ being $s$'s *parent*—the situation preceding $s$). Since (4.23) implies (4.24) we may conclude that condition 1 implies condition 2. To derive condition 1 from condition 2, we rewrite (4.24) and (4.23) as

$$(\mathcal{S}_+^2 - \mathcal{S}^2) - (\mathcal{A}_+ - \mathcal{A}) \leq \mathcal{K}_+^* - \mathcal{K}^* \tag{4.26}$$

and

$$(\mathcal{S}_+ - \mathcal{S})^2 - (\mathcal{A}_+ - \mathcal{A}) \leq (\mathcal{K}_+^* - \mathcal{K}^*) - 2\mathcal{S}(\mathcal{S}_+ - \mathcal{S}), \tag{4.27}$$

respectively, and we similarly use the fact that (4.26) implies (4.27). ∎

Here is our strong law of large numbers for martingales, which is similar in form to the measure-theoretic strong law for martingales (Corollary 8.1 on p. 171):

**Proposition 4.4** *If $\mathcal{S}$ is a martingale in a symmetric probability protocol, and $\mathcal{A}$ is a quadratic supervariation for $\mathcal{S}$, then Skeptic can force*

$$\sum_{n=1}^{\infty} \frac{\Delta \mathcal{A}_n}{n^2} < \infty \Longrightarrow \lim_{n\to\infty} \frac{\mathcal{S}_n}{n} = 0, \tag{4.28}$$

*where $\Delta \mathcal{A}_n := \mathcal{A}_n - \mathcal{A}_{n-1}$.*

*Proof* Let $\mathcal{P}_1$ be a strategy with capital process $\mathcal{S}$ (i.e., $\mathcal{S} = \mathcal{K}^{\mathcal{P}_1}$), and let $\mathcal{P}_2$ be a strategy whose capital process satisfies (4.22), so that all increments obey

$$\Delta \mathcal{S} = \Delta \mathcal{K}^{\mathcal{P}_1}, \tag{4.29}$$

and

$$(\Delta \mathcal{S})^2 - \Delta \mathcal{A} \leq \Delta \mathcal{K}^{\mathcal{P}_2}. \tag{4.30}$$

We will show how Skeptic can force (4.28) by combining $\mathcal{P}_1$ and $\mathcal{P}_2$ with information obtained from a winning strategy $(\mathcal{M}, \mathcal{V})$ in the unbounded forecasting protocol with zero prices for $x$-tickets.

In order to use the information in $(\mathcal{M}, \mathcal{V})$, Skeptic simulates the unbounded forecasting protocol as he plays the symmetric probability protocol, pretending that Forecaster and Reality make moves in the unbounded forecasting protocol that he computes from their moves in the symmetric probability protocol and computing his own moves in the symmetric probability protocol from what happens in the unbounded forecasting protocol. His procedure is spelled out in Table 4.1, which shows one round in the two protocols. At the beginning of the round, the Skeptic is in situation $s_{\text{act}}$ in the symmetric probability protocol, the protocol he is actually playing, and in situation $s_{\text{sim}}$ in the unbounded forecasting protocol, the protocol he is simulating. He then observes Forecaster's move $\mathbf{f}$ in the symmetric probability protocol. He calculates the resulting increment $\Delta \mathcal{A}$, and he interprets this as Forecaster's move $v$ (his upper

***Table 4.1*** As he plays the symmetric probability protocol, Skeptic simulates the unbounded forecasting protocol with zero prices for $x$-tickets.

| | **Symmetric Probability Protocol** | | **Unbounded Forecasting Protocol** |
|---|---|---|---|
| Situation | $s_{\text{act}}$ (actual) | | $s_{\text{sim}}$ (simulated) |
| Forecaster's move | $\mathbf{f}$ | $\rightarrow$ | $v := \Delta\mathcal{A} = \mathcal{A}(s_{\text{act}}\mathbf{f}) - \mathcal{A}(s_{\text{act}})$ $\downarrow$ |
| Skeptic's move | $\mathbf{s} := M\mathcal{P}_1(s_{\text{act}}\mathbf{f}) + V\mathcal{P}_2(s_{\text{act}}\mathbf{f})$ $\downarrow$ | $\leftarrow$ | $M := \mathcal{M}(s_{\text{sim}}v)$ $V := \mathcal{V}(s_{\text{sim}}v)$ |
| Reality's move | $\mathbf{r}$ | $\rightarrow$ | $x := \Delta\mathcal{S} = \mathcal{S}(s_{\text{act}}\mathbf{fr}) - \mathcal{S}(s_{\text{act}})$ |

price for $x^2$-tickets) in the simulated unbounded forecasting protocol. Using $v$, he calculates the number $M$ of $x$-tickets and the number $V$ of $x^2$-tickets to buy in the unbounded forecasting protocol from his winning strategy $(\mathcal{M}, \mathcal{V})$. He then calculates $M\mathcal{P}_1(s_{\text{act}}\mathbf{f}) + V\mathcal{P}_2(s_{\text{act}}\mathbf{f})$ and uses this as his own move in the symmetric probability protocol. After waiting for Reality to complete the round in symmetric probability protocol by making her move $\mathbf{r}$, he completes the round in the unbounded forecasting protocol by pretending that Reality's move there is the same as the increment of $\mathcal{S}$ in the symmetric probability protocol.

In the unbounded forecasting protocol, Skeptic's gain on this round is

$$Mx + V(x^2 - v) = M\Delta\mathcal{S} + V((\Delta\mathcal{S})^2 - \Delta\mathcal{A}). \tag{4.31}$$

But according to (4.29) and (4.30),

$$M\Delta\mathcal{S} + V((\Delta\mathcal{S})^2 - \Delta\mathcal{A}) \leq M\Delta\mathcal{K}^{\mathcal{P}_1} + V\Delta\mathcal{K}^{\mathcal{P}_2}, \tag{4.32}$$

and since the gain function is linear in Skeptic's move, $M\Delta\mathcal{K}^{\mathcal{P}_1} + V\Delta\mathcal{K}^{\mathcal{P}_2}$ is Skeptic's gain in the symmetric probability protocol. So Skeptic makes at least as much money in the symmetric probability protocol as in the unbounded forecasting protocol. Since $v = \Delta\mathcal{A}$ and $x = \Delta\mathcal{S}$, (4.4) happens in the unbounded forecasting protocol if and only if (4.28) happens in the symmetric probability protocol. So the fact that $(\mathcal{M}, \mathcal{V})$ is a winning strategy in the unbounded forecasting protocol implies that the strategy explained in Table 4.1 is a winning strategy in the symmetric probability protocol. ∎

We are also interested in probability protocols that are not symmetric (after all, the unbounded forecasting protocol itself is not symmetric). When we do not assume symmetry, however, we must assume some other form of regularity on Skeptic's move space and gain function. The following proposition is an example of the kind of result we can obtain.

**Proposition 4.5** *Suppose Skeptic's move space is a closed convex cone in a Banach space (i.e., a complete normed linear space), and suppose the gain function $\lambda$ is linear and continuous in Skeptic's move.*[5] *Suppose further that $\mathcal{S}$ is a supermartingale and $\mathcal{A}$ is a compensator for $\mathcal{S}^2$. Then Skeptic can force (4.28).*

*Proof* This proposition can be proven by mimicking the proof of the strong law for the unbounded forecasting protocol. It is clear that Lemmas 4.1 and 4.2 continue to hold under our new assumptions. We can replace (4.8) with

$$\|\mathcal{T}\|(s\mathbf{f}) := \inf\Big\{\|\mathbf{s}\| : \mathcal{T}(s\mathbf{a}\mathbf{r}) - \mathcal{T}(s) \leq \lambda(\mathbf{f}, \mathbf{s}, \mathbf{r}) \text{ for all } \mathbf{r} \in \mathbf{R}\Big\}$$

and define minimal bounding strategies essentially as we did before, but, unfortunately, Lemma 4.3 ceases to be true, since it makes use of the compactness of closed balls in Euclidean spaces. However, the following relaxation of Lemma 4.3 is obviously true:

[5]Since we are not assuming that Skeptic's move space is linear, the condition that $\lambda(\mathbf{f}, \mathbf{s}, \mathbf{r})$ be linear in $\mathbf{s}$ must be understood carefully. For us it means that (a) if $\mathbf{s}$ is in $\mathbf{S}$ and $c \geq 0$, then $\lambda(\mathbf{f}, c\mathbf{s}, \mathbf{r}) = c\lambda(\mathbf{f}, \mathbf{s}, \mathbf{r})$ (notice that $c\mathbf{s}$ is in $\mathbf{S}$ automatically), and (b) if $\mathbf{s}_1$ and $\mathbf{s}_2$ are both in $\mathbf{S}$, then $\lambda(\mathbf{f}, \mathbf{s}_1 + \mathbf{s}_2, \mathbf{r}) = \lambda(\mathbf{f}, \mathbf{s}_1, \mathbf{r}) + \lambda(\mathbf{f}, \mathbf{s}_2, \mathbf{r})$ (notice that $\mathbf{s}_1 + \mathbf{s}_2$ is in $\mathbf{S}$ automatically).

**Lemma 4.9** *Let $\epsilon$ be a positive constant; we say that a bounding strategy $\mathcal{B}$ for $\mathcal{T}$ is $\epsilon$-*minimal *if $\|\mathcal{B}(s)\| \leq (1+\epsilon)\|\mathcal{T}\|(s)$ for all $s \neq \Box$. Every supermartingale has an $\epsilon$-minimal bounding strategy.*

It is clear that Lemma 4.4 continues to be true: in its proof we can replace "minimal bounding strategy" with "$\epsilon$-minimal bounding strategy" and let $\epsilon \to 0$; the convergence of series will follow from the completeness of the Banach space. The proof proceeds further as before; we leave the details to the reader. ∎

The hypothesis of Proposition 4.5 is incomparable with the hypothesis of linearity used by Proposition 4.4. (A cone is not necessarily a linear space, and a linear space is not necessarily Banach.) But either proposition can serve as the starting point for deriving the strong law of large numbers for securities markets that we study in §15.1.

## 4.6 APPENDIX: MARTIN'S THEOREM

Tony Martin, the UCLA mathematician who proved the determinacy of Borel and quasi-Borel games, in 1990 at his home in Santa Monica, California.

Martin's theorem, the only mathematically advanced result from game theory used in this book, says that quasi-Borel games are determinate. In this appendix, we explain the meaning of this statement. We also comment briefly on the significance of Martin's theorem for topics that go far beyond the concerns of this book: the general theory of games and the foundations of mathematics.

Consider a perfect-information game with two players, Player I and Player II, who alternate moves. Such a game can be specified mathematically by specifying (1) the game tree and (2) the set of paths through the game tree that produce a win for Player I. In 1913, Ernst Zermelo (1871–1953) proved an almost obvious fact: if the game has a finite horizon, then one of the players has a winning strategy. Figure 4.1 illustrates why this theorem is almost obvious: as soon as we see that the winner is determined in every penultimate situation in the game tree, it becomes obvious that the theorem can be proven by backward induction.

In the general case, where the game tree may contain arbitrarily long or even infinite paths, it is not at all obvious that one of the players has a winning strategy. Martin's theorem tells us, however, that this is the case unless the subset of paths on which Player I wins is extremely pathological.

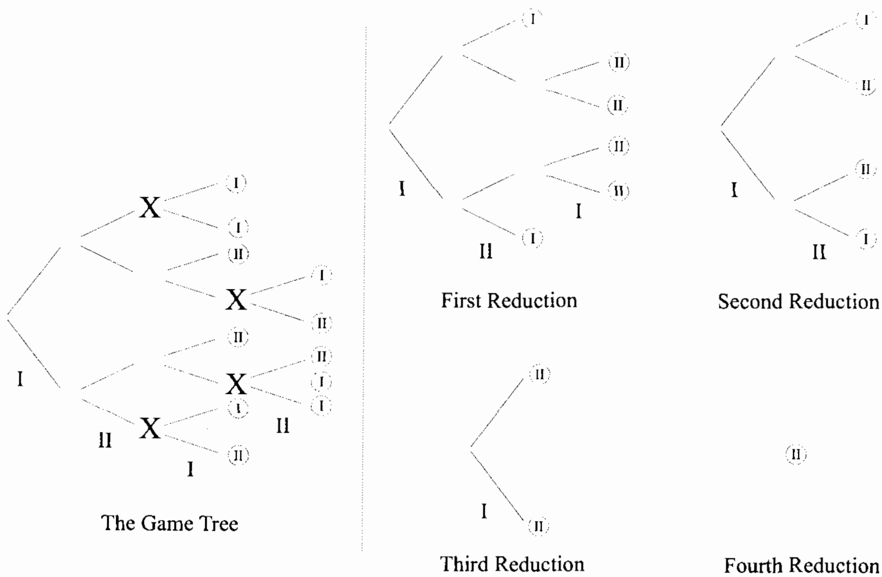



***Fig. 4.1*** In the game tree on the left, Player I moves first, and the game ends after three or at most four moves. The winner is indicated at the end of each path; Player I wins on six paths, and Player II wins on the other five. But Player II has a winning strategy. To see this, we progressively prune away terminal moves. In the first reduction, we prune the moves following the four situations marked with an X, replacing each X with the winner, on the assumption that the player who has the move will choose a winning move if he has one. Three more reductions bring us back to the initial situation, marked with Player II as the winner.

### Precise Statement

In the games between Skeptic and World studied in this book, we use trees that describe only the moves made by World. We are now concerned, however, with a game tree in the classic sense—a tree that describes the moves of both players. In order to avoid any confusion on this point, we use a different notation; we write $\Gamma$ instead of $\Omega$ for the set of possible paths, and we write $\mathbf{t}$ instead of $\Omega^\diamond$ for the set of situations. We suppose, for simplicity, that the game tree has an infinite horizon: All the paths are infinite. This does not limit the generality of the discussion, because (as Fermat first noted) we can always imagine that play continues even though the winner has already been determined.

The literature on perfect-information games in mathematical logic, on which this appendix draws, formalizes the infinite-horizon notion of a game tree starting with the set $\mathbf{t}$ of situations: a set of finite sequences $\mathbf{t}$ is a *game tree* if

1. $t \in \mathbf{t}$ and $s \sqsubseteq t$ ($t$ extends $s$) imply that $s \in \mathbf{t}$, and

2. $s \in \mathbf{t}$ implies that there is a $t \in \mathbf{t}$ such that $s \sqsubset t$ ($t$ strictly extends $s$).

A *path* in a game tree $\mathbf{t}$ is an infinite sequence $a_1 a_2 \dots$ such that $a_1 \dots a_n \in \mathbf{t}$ for all $n = 1, 2, \dots$. We write $\Gamma$ for the set of all paths. Player I makes the odd-numbered moves, and Player II makes the even-numbered ones:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| I | $a_1$ | | $a_3$ | | $a_5$ | | | $\dots$ |
| II | | $a_2$ | | $a_4$ | | $\dots$ | | |

Each situation $a_1 \dots a_n$, $n = 1, 2, \dots$, belongs to $\mathbf{t}$.

For $E \subseteq \Gamma$ we denote by $\mathcal{G}(E)$ the game on $\mathbf{t}$ with the following winning conditions: I wins a play $a_1 a_2 \dots$ of $\mathcal{G}(E)$ if $a_1 a_2 \dots \in E$; otherwise, II wins. The game $\mathcal{G}(E)$ is *determined* if either I or II has a winning strategy. The notion of a strategy is based on the assumption that each player has perfect information: he always knows the other player's moves so far.

If $t \in \mathbf{t}$, then $\Gamma(t) \subseteq \Gamma$ is defined to be the set of all infinite extensions of $t$. We give $\Gamma$ a topology by taking the $\Gamma(t)$, $t \in \mathbf{t}$, as basic open sets. Each of the $\Gamma(t)$ is both closed and open.

The *quasi-Borel* subsets of $\Gamma$ form the smallest class of subsets of $\Gamma$ that contains all open sets and is closed under (1) complementation, (2) finite or countable union, and (3) open-separated union. We will never use property (3), but for completeness we give the definition of open-separated union: $E \subseteq \Gamma$ is the *open-separated union* of a family $\{F_j : j \in J\}$ of subsets of $\Gamma$ if (1) $E = \cup_{j \in J} F_j$ and (2) there are disjoint open sets $D_j$, $j \in J$, such that $F_j \subseteq D_j$ for each $j \in J$. This is a larger class than the Borel subsets: every Borel subset is a quasi-Borel subset.

The following result was proven by Donald A. Martin in 1990 ([220], Corollary 1):

**Martin's Theorem** *If $E \subseteq \Gamma$ is quasi-Borel, then $\mathcal{G}(E)$ is determined.*

The winning sets for all games considered in this book are quasi-Borel. For example, let us check that the winning condition

$$\left(\mathcal{K}_n < 0 \text{ for some } n\right) \text{ or } \left( (\mathcal{K}_n \text{ does not tend to } \infty) \text{ and } \left( \sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \text{ or } \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) \text{ does not tend to } 0 \right) \right)$$

for Reality in the game of forcing (4.5) is quasi-Borel. Since the class of quasi-Borel events is an algebra, it suffices to prove that the four events

$$\mathcal{K}_n < 0 \text{ for some } n, \qquad \mathcal{K}_n \text{ does not tend to } \infty,$$

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty, \qquad \text{and} \qquad \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) \text{ does not tend to } 0$$

are quasi-Borel. All these events can be expressed by means of events depending only on a finite number of the players' moves (such events are open and so quasi-Borel) using countable unions and intersections and therefore are quasi-Borel.

Martin reviews the history of his theorem in [219]. As he explains, it was the culmination of a lengthy investigation by a number of mathematicians following Zermelo. Polish mathematicians proved the theorem for additional special cases, but a general result came in 1953, when David Gale and F. M. Stewart proved the determinacy of open games (games $\mathcal{G}(E)$ for which $E$ is an open, or $\Sigma_1^0$, set) on trees with at most countably many situations. Philip Wolfe extended this result to $\Sigma_2^0$ games in 1955, Morton Davis extended it to $\Sigma_3^0$ games in 1964, and Jeff Paris extended it to $\Sigma_4^0$ games in 1972. Martin extended it to Borel games in 1975, then, in 1990, he established the general result that we cite: all quasi-Borel games are determined, even if the game tree is uncountable. The definition of the classes $\Sigma_m^0$, as well as further information about the determinacy of Borel and projective sets, can be found in Kechris [168]. In our games, the goal $E$ is usually very simple and hence very low in the $\Sigma_m^0$ hierarchy, but Martin's 1990 result relieves us of any need to examine $E$'s complexity closely; it is enough that $E$ should be quasi-Borel.

Perfect-information games occupy only a small corner of the general theory of games. In most of the two-person games studied in the social sciences and in operations research, the players do not alternate moves while observing each other. But in 1998, Martin showed that his 1990 theorem implies determinacy for a very wide class of Blackwell games. In a *Blackwell game*, Players I and II move simultaneously and play randomized strategies. Such a game can be defined by payoff functions rather than by winning sets, and then determinacy means the existence of a value for the game. Von Neumann's minimax theorem asserts the determinacy of one-move Blackwell games. Martin's result, which generalizes this to arbitrary Blackwell games with Borel measurable payoff functions, suggests that perfect-information games play a fundamental role in the general theory of games.

In 1962 Mycielski and Steinhaus proposed a statement similar to Martin's theorem as an axiom for mathematics in general. Their *axiom of determinacy* AD (see [233]) asserts that every perfect-information game $\mathcal{G}(E)$ on the countable tree $\mathbb{N}^*$ (the set of all finite sequences of natural numbers) is determined. In his 1998 article, Martin showed that this implies the analogous statement for Blackwell games (see also [304]). The axiom AD contradicts the axiom of choice, which is assumed in Martin's proof, but in combination with the *axiom of dependent choice* DC it is a strong competitor with the axiom of choice. The mathematics based on AD + DC is more regular in some ways than standard mathematics, which uses the axiom of choice. For example, the following surprising results can be proven assuming AD + DC:

**Mycielski-Świerczkowski [234]:** Any set of real numbers is Lebesgue measurable.

**Davis [81]:** Any uncountable set of real numbers contains a perfect subset.

The second result solves the continuum hypothesis, since every perfect (i.e., nonempty, closed, and without isolated points) subset of the real line has the cardinality of the continuum.

Of course, we do not need a new foundation for mathematics in this book, and accepting AD instead of AC would actually be awkward for our Chapters 11–14. According to a folk theorem (whose proof can be found in [165], p. 52), non-principal ultrafilters on $\mathbb{N}$ do not exist under AD, and this would eliminate the construction of nonstandard analysis described in §11.5 and used throughout those chapters.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# 5

# *The Law of the Iterated Logarithm*

Aleksandr Khinchin (1894–1959)

The strong law of large numbers says that under certain circumstances an average converges almost surely. The law of the iterated logarithm concerns the rate and oscillation of this convergence. It was first formulated and proven by Aleksandr Yakovlevich Khinchin in work published in 1924. Khinchin considered only coin tossing, but general measure-theoretic versions of the law were established by Andrei N. Kolmogorov (1929) and William F. Stout (1970). In this chapter, we state and prove a very general game-theoretic version of the law. As we will see in Chapter 8 (p. 172), this game-theoretic version implies Kolmogorov's and Stout's measure-theoretic results.

Khinchin's law is already very interesting in the simplest case, where the coin being tossed is fair. In this case, the strong law of large numbers says that $y_n/n$ converges almost surely to $1/2$, where $y_n$ is the number of heads in the first $n$ tosses. Khinchin's law adds that it almost surely oscillates as it converges, with maximal deviations on both sides asymptotically close to $\sqrt{\ln\ln n}/\sqrt{2n}$. More precisely,

$$\limsup_{n\to\infty} \frac{\frac{y_n}{n} - \frac{1}{2}}{\sqrt{\frac{\ln\ln n}{2n}}} = 1 \tag{5.1}$$

and

$$\liminf_{n\to\infty} \frac{\frac{y_n}{n} - \frac{1}{2}}{\sqrt{\frac{\ln \ln n}{2n}}} = -1 \tag{5.2}$$

both hold almost surely. We can interpret "almost surely" either measure-theoretically or game-theoretically.

Our general game-theoretic result is set in a variant of the unbounded forecasting protocol, the protocol that we used in the preceding chapter for our game-theoretic version of Kolmogorov's strong law. In the unbounded forecasting protocol, Forecaster gives a price $m_n$ for Reality's move $x_n$ and an upper price $v_n$ for the squared deviation $(x_n - m_n)^2$. We show that

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln \ln A_n}} = 1, \tag{5.3}$$

with $A_n := \sum_{i=1}^n v_i$, holds almost surely in the game-theoretic sense provided that the rules of the protocol are adjusted so that $A_n$ must tend to infinity and the $x_n - m_n$ must stay within bounds that do not grow too fast. Equation (5.3) generalizes (5.1). We can also generalize (5.2), obtaining

$$\liminf_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln \ln A_n}} = -1. \tag{5.4}$$

But because of symmetry between $x_n$ and $-x_n$ in the unbounded forecasting protocol, (5.3) and (5.4) are equivalent. So we discuss only (5.3).

Our game-theoretic study of the law of the iterated logarithm leads to insights not brought out in standard measure-theoretic expositions. These insights begin with our noticing that (5.3) combines two assertions:

- $\sqrt{2A_n \ln \ln A_n}$ is almost surely *valid* as an asymptotic bound on the cumulative sum $|\sum_{i=1}^n (x_i - m_i)|$. For any positive number $\epsilon$,

$$\frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln \ln A_n}} \tag{5.5}$$

  will eventually stay between $-(1+\epsilon)$ and $1+\epsilon$ almost surely.

- The bound is also almost surely *sharp*. For any positive number $\epsilon$, the ratio (5.5) will almost surely come within $\epsilon$ of both $-1$ and $1$ infinitely often.

Sharpness is more delicate than validity, but measure-theoretic authors have made it a point of honor to prove both under the same conditions on the $x_n - m_n$ and their variances. In the game-theoretic framework, a different attitude is natural, for Skeptic can force validity under weaker conditions than sharpness. He can force validity provided only that Reality keeps her values for $|x_n - m_n|$ bounded relative to Forecaster's values for $A_n$. But in order also to force sharpness, the bounds $A_n$ must themselves be sharp (i.e., the $v_n$, which Forecaster announces to Skeptic in advance of his move, must be prices rather than merely upper prices), and Forecaster must even give some extra advance warning of particularly large values of $|x_n - m_n|$.

The iterated-logarithm bound is valid but not sharp for many of the protocols in this book, including the bounded forecasting protocol of Chapter 3 and the unbounded forecasting protocol of Chapter 4. We obtain sharpness only after changing the unbounded forecasting rules in Skeptic's favor, and these changes will not fit all applications. In our study of securities markets in Chapter 15, we will see an example where we might hope for the weak limits on Reality's moves that make the bound valid but not for the advance warnings about her moves that make it sharp.

In §5.1, we introduce our protocols and state our theorems concerning validity and sharpness. Then we turn to the proofs: §5.2 deals with validity and §5.3 with sharpness. In §5.4, we consider the case of an abstract symmetric probability protocol. There are two appendixes: §5.5 discusses the history of the law of the iterated logarithm, and §5.6 explains its finitary interpretation.

## 5.1 UNBOUNDED FORECASTING PROTOCOLS

Our main theorems are set in two general protocols: the unbounded forecasting protocol (familiar from the preceding chapter) and a variant, the predictably unbounded forecasting protocol. We also consider some simpler protocols.

### The Unbounded and Predictably Unbounded Protocols

Here again is the unbounded forecasting protocol, which we used in the preceding chapter for our game-theoretic version of Kolmogorov's strong law.

UNBOUNDED FORECASTING
**Players:** Forecaster, Skeptic, Reality
**Protocol:**

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $m_n \in \mathbb{R}$ and $v_n \geq 0$.
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \geq 0$.
Reality announces $x_n \in \mathbb{R}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n) + V_n((x_n - m_n)^2 - v_n)$.

**Forcing by Skeptic:** A strategy for Skeptic (or for Skeptic and Forecaster) forces an event $E$ if it assures that $\mathcal{K}_n$ is never negative and either $\mathcal{K}_n$ tends to infinity or else $E$ happens, regardless of the moves of the other player(s).
**Forcing by Reality:** A strategy for Reality (or for Reality and Forecaster) forces an event $E$ if it assures that, provided $\mathcal{K}_n$ is never negative, $\mathcal{K}_n$ remains bounded and $E$ happens, regardless of the moves of the other player(s).

An unbounded forecasting game is obtained from this protocol by choosing a player or coalition (Skeptic, Reality, or a coalition of one of them with Forecaster) and a goal $E$. The player or coalition wins by forcing the goal. In Chapter 4, we showed that Skeptic can force (4.4) and Reality can force (4.5). In this chapter, we consider goals related to the iterated-logarithm bound.

As we shall prove, the iterated-logarithm bound is valid but not sharp in the unbounded forecasting protocol. In order to have sharpness as well as validity, we must slightly limit Reality's freedom of action and slightly enlarge Skeptic's freedom of action, as follows:

PREDICTABLY UNBOUNDED FORECASTING
**Protocol:**

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \dots$:
Forecaster announces $m_n \in \mathbb{R}$, $c_n \geq 0$, and $v_n \geq 0$.
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Reality announces $x_n \in \mathbb{R}$ such that $|x_n - m_n| \leq c_n$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n) + V_n((x_n - m_n)^2 - v_n)$.

Here we have given Skeptic two advantages: (1) He is given, just before his choice of $M_n$ and $V_n$, a bound $c_n$ on how far Reality's subsequent choice of $x_n$ will be from $m_n$. (2) He is allowed to choose $V_n$ negative.

We shall prove the following two theorems. The first says that the iterated-logarithm bound is both valid and sharp in the predictably unbounded protocol. The second says that it remains valid in our original unbounded protocol. In both theorems, $A_n := \sum_{i=1}^n v_i$.

**Theorem 5.1** *In the predictably unbounded forecasting protocol, Skeptic can force*

$$\left(A_n \to \infty \ \&\ c_n = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln\ln A_n}} = 1.$$

**Theorem 5.2** *In the unbounded forecasting protocol, Skeptic can force*

$$\left(A_n \to \infty \ \&\ |x_n - m_n| = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln\ln A_n}} \leq 1. \tag{5.6}$$

The statement of Theorem 5.2 remains true, of course, if we increase Skeptic's freedom by allowing negative values of $V_n$.

In the next section, §5.2, we prove the validity of the iterated-logarithm bound: Skeptic can force (5.6) in the unbounded protocol and hence also force

$$\left(A_n \to \infty \ \&\ c_n = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln\ln A_n}} \leq 1$$

in the predictably unbounded protocol (where the rules are more favorable to him). This establishes Theorem 5.2 and half of Theorem 5.1—the half we obtain by replacing $= 1$ by $\leq 1$. In the section following, §5.3, we prove sharpness in the predictably unbounded protocol, thus establishing the other half of Theorem 5.1—the half we obtain by replacing $= 1$ by $\geq 1$.

The following proposition confirms that the iterated-logarithm bound is not sharp for the unbounded forecasting protocol.

**Proposition 5.1** *In the unbounded forecasting protocol, Forecaster and Reality can force*

$$A_n = n \; \& \; x_n = m_n = 0$$

*for all $n$, even if we increase Skeptic's freedom by allowing negative $V_n$.*

*Proof* Forecaster always announces $m_n = 0$ and $v_n = 1$. If Skeptic chooses a negative number for $V_n$, Reality busts him by choosing $x_n$ with sign opposite to that of $M_n$ (any sign, if $M_n = 0$) and so large in absolute value that Skeptic's capital,

$$\mathcal{K}_{n-1} + M_n x_n + V_n(x_n^2 - 1),$$

becomes negative. If Skeptic chooses $V_n$ nonnegative, Reality chooses $x_n = 0$, producing a zero change or decrease in Skeptic's capital. So $A_n = n$ and $m_n = 0$ will always be satisfied, Skeptic's capital will never increase, and either Skeptic will lose by going into debt or $x_n = 0$ will always be satisfied. ∎

## Simplifications

If we are only interested in the validity of the iterated-logarithm bound, we can discard the requirement that Forecaster issue a variance $v_n$ at every step, as in the following simplified protocol:

SIMPLIFIED PREDICTABLY UNBOUNDED FORECASTING

**Protocol:**

$\mathcal{K}_0 := 1$.

FOR $n = 1, 2, \ldots$:

Forecaster announces $m_n \in \mathbb{R}$ and $c_n \geq 0$.

Skeptic announces $M_n \in \mathbb{R}$.

Reality announces $x_n \in [-c_n, c_n]$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n)$.

**Corollary 5.1** *In the simplified predictably unbounded forecasting protocol, Skeptic can force*

$$\left(A_n \to \infty \;\&\; c_n = o\left(\sqrt{\frac{A_n}{\ln \ln A_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln \ln A_n}} \leq 1,$$

*where $A_n$ is defined to be $\sum_{i=1}^n c_i^2$.*

*Proof* Without loss of generality we can assume that $m_n \in [-c_n, c_n]$ (otherwise the protocol will not be coherent), but notice that we cannot just set $m_n = 0$ (as we do in the proofs of Theorems 5.1 and 5.2 below).

In view of Theorem 5.2, it suffices to show that Skeptic can simulate buying the variance tickets with payoff $(x_n - m_n)^2$ and price $c_n^2$ with the existing tickets. The lowest price for which such a simulation is possible is

$$\sup_{m_n} \inf_{b_n} \sup_{x_n \in [-c_n, c_n]} \left( (x_n - m_n)^2 + b_n(x_n - m_n) \right) \tag{5.7}$$

(recall that $x_n - m_n$ is the net payoff of the available tickets, which can be bought in any number to hedge against too large payoffs $(x_n - m_n)^2$). It is easy to see that indeed (5.7) does not exceed $c_n^2$; dropping the subindex $n$, we can transform (5.7) as follows:

$$\begin{aligned} &\sup_m \inf_b \sup_{x \in [-c,c]} \left( (x-m)^2 + b(x-m) \right) \\ &= \sup_m \inf_b \max\left( (-c-m)^2 + b(-c-m), (c-m)^2 + b(c-m) \right) \\ &= \sup_m \left( c^2 - m^2 \right) = c^2 \end{aligned}$$

(the optimal value of $b$ in this chain is $2m$, the one that equalizes the two expressions after the max sign). ∎

If we now require that Forecaster set each $c_n$ to some constant $C$, given in advance, then this protocol reduces, essentially, to the bounded forecasting protocol we studied in §3.3. The conclusion then reduces to

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2C^2 n \ln \ln n}} \le 1. \tag{5.8}$$

Forecaster and Reality are still free, however, to make the differences $x_n - m_n$ all equal to zero, and when they do so Skeptic certainly will not become infinitely rich. So Skeptic cannot force the bound to be sharp as well as valid even in the bounded forecasting protocol.

## 5.2 THE VALIDITY OF THE ITERATED-LOGARITHM BOUND

In this section we prove Theorem 5.2. This proof is an extension of the proof in Dawid and Vovk 1997 (which, in its turn, was adapted from Vovk 1988), but its idea goes back to Ville 1939.

We are working in the unbounded forecasting protocol, where we assume, without loss of generality, that $m_n = 0$ for all $n$.

### The Main Idea of the Proof

Our task is to construct a strategy for Skeptic that will make his capital unbounded if the sum $\sum_{i=1}^n x_i$ persists in having extremely large values so that the iterated-logarithm bound is violated. (Once we have a strategy with unbounded capital, we can construct one with capital that tends to infinity using the trick that we learned in Chapter 3.) This goal would certainly be achieved if we could always multiply

Skeptic's capital on the $n$th step by, say, $\exp \kappa x_n$ for some positive $\kappa$, for this would produce the capital

$$\exp\left(\kappa \sum_{i=1}^{n} x_i\right) \tag{5.9}$$

after $n$ steps, which is unbounded if $\sum_{i=1}^{n} x_i$ is unbounded, no matter how small the positive constant $\kappa$ is. Of course, (5.9) is too much to hope for, but we can come close to multiplying the capital on the $n$th step by $\exp\left(\kappa x_n - \kappa^2 v_n/2\right)$, which is nearly the same when $\kappa$ is very small; this will produce the capital

$$\exp\left(\kappa \sum_{i=1}^{n} x_i - \frac{1}{2}\kappa^2 \sum_{i=1}^{n} v_i\right) = \exp\left(\kappa \sum_{i=1}^{n} x_i - \frac{1}{2}\kappa^2 A_n\right). \tag{5.10}$$

In order for this to work, we need to use a number of values of $\kappa$, chosen to take maximal advantage of successively larger values of the deviation $\sum_{i=1}^{n} x_i$. The large deviations we want to take advantage of are those just where $\sum_{i=1}^{n} x_i$ slightly exceeds

$$(2A_n \ln\ln A_n)^{1/2},$$

since this is the maximal value allowed by the law of the iterated logarithm. The optimal value of $\kappa$ for taking advantage of such a deviation is

$$\kappa(n) = \frac{\sum_{i=1}^{n} x_i}{A_n} \approx \sqrt{\frac{2\ln\ln A_n}{A_n}}$$

(cf. (5.21); to see why $\kappa$ should be given by this expression notice that it provides the maximum of the exponent in (5.10)). The next step is to approximate $A_n$, and this produces (5.14) with $k$ defined by (5.19). The problem is that even $k$ is unknown in advance; the solution is to take a mixture with different $k$ with weights $w(k)$ shrinking (as $k \to \infty$) as slowly as possible. We use $w(k) = k^{-1-\delta}$ (see (5.15)).

In this argument the value

$$\kappa \approx \sqrt{\frac{2\ln\ln A_n}{A_n}}$$

is obtained from the statement of the law of the iterated logarithm. A less *ad hoc* way to obtain it is to minimize the sum of the first two addends in the right-hand side of (5.24) with respect to $\kappa$.

## The Proof

Fix temporarily a number $\delta \in (0,1)$. (The intuition behind our reasoning will rely on $\delta$ being very small.) For every $\kappa \in (0,1)$, Skeptic has a strategy $\mathcal{P}^{(\kappa)}$ in the unbounded forecasting protocol (with initial capital \$1) such that, for all $i$, his capital $\mathcal{K}_i^{(\kappa)}$ satisfies

$$\mathcal{K}_{i+1}^{(\kappa)} = \mathcal{K}_i^{(\kappa)} \frac{1 + \kappa x_i + (1+\delta)\kappa^2 x_i^2/2}{1 + (1+\delta)\kappa^2 v_i/2} \tag{5.11}$$

when he is using that strategy (indeed, (5.11) is equivalent to

$$\mathcal{K}_{i+1}^{(\kappa)} = \mathcal{K}_i^{(\kappa)} + \frac{\mathcal{K}_i^{(\kappa)}}{1 + (1+\delta)\kappa^2 v_i/2} \left(\kappa x_i + (1+\delta)(\kappa^2/2)(x_i^2 - v_i)\right),$$

so $\mathcal{P}^{(\kappa)}$ recommends buying

$$\frac{\mathcal{K}_i^{(\kappa)}\kappa}{1 + (1+\delta)\kappa^2 v_i/2} \tag{5.12}$$

$x_i$-tickets and

$$\frac{\mathcal{K}_i^{(\kappa)}(1+\delta)\kappa^2/2}{1 + (1+\delta)\kappa^2 v_i/2} \tag{5.13}$$

$x_i^2$-tickets at every step $i$). Notice that (5.11) is always nonnegative. Put

$$\kappa(k) := \sqrt{2(1+\delta)^{-k}\ln k}, \tag{5.14}$$

$$\mathcal{P} := \sum_{k=K}^{\infty} k^{-1-\delta}\mathcal{P}^{(\kappa(k))}, \tag{5.15}$$

where $K = K(\delta)$ is chosen so that $\kappa(k) \in (0,1)$ for $k \geq K$. (Later we shall impose additional conditions on how large $K$ should be.) Let $\mathcal{K}_i > 0$ be the capital process for $\mathcal{P}$ with the initial capital

$$\mathcal{K}_0 := \sum_{k=K}^{\infty} k^{-1-\delta}. \tag{5.16}$$

Skeptic can weakly force $\sup_n \mathcal{K}_n < \infty$ and hence also

$$\sup_n \mathcal{K}_n^{\kappa(k)} = O\left(k^{1+\delta}\right). \tag{5.17}$$

Consider a play where both (5.17) and

$$A_n \to \infty \ \&\ |x_n| = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right) \tag{5.18}$$

hold. For such a play and $n$ sufficiently large, we set

$$k(n) := \lfloor \log_{1+\delta} A_n \rfloor \tag{5.19}$$

and obtain

$$\ln \mathcal{K}_n^{(\kappa[n])} \leq (1+\delta)\ln k(n) + O(1) \leq (1+\delta)\ln\ln A_n + O(1),$$

where $\kappa[n] := \kappa(k(n))$. By (5.11), we further obtain

$$\ln \prod_{i=1}^{n} \frac{1 + \kappa[n]x_i + (1+\delta)\kappa^2[n]x_i^2/2}{1 + (1+\delta)\kappa^2[n]v_i/2} \leq (1+\delta)\ln\ln A_n + O(1),$$

that is,

$$\begin{aligned}&\sum_{i=1}^{n} \ln\left(1+\kappa[n]x_i+(1+\delta)\kappa^2[n]x_i^2/2\right)\\ \le &\sum_{i=1}^{n} \ln\left(1+(1+\delta)\kappa^2[n]v_i/2\right)+(1+\delta)\ln\ln A_n+O(1).\end{aligned} \tag{5.20}$$

For $n$ large enough,

$$\frac{1}{1+\delta}\sqrt{\frac{2\ln\ln A_n}{A_n}} \le \kappa[n] \le (1+\delta)\sqrt{\frac{2\ln\ln A_n}{A_n}}; \tag{5.21}$$

in conjunction with (5.18) this gives

$$\sup_{i\le n}\kappa[n]|x_i| \to 0 \ (n\to\infty). \tag{5.22}$$

For $t$ small enough in absolute value we have

$$\ln\left(1+t+(1+\delta)t^2/2\right) \ge t, \tag{5.23}$$

so (5.20) and (5.22) imply

$$\kappa[n]\sum_{i=1}^{n} x_i \le \sum_{i=1}^{n} \ln\left(1+(1+\delta)\kappa^2[n]v_i/2\right)+(1+\delta)\ln\ln A_n+O(1).$$

Since always $\ln(1+t)\le t$, we further obtain

$$\sum_{i=1}^{n} x_i \le (1+\delta)\kappa[n]A_n/2+\frac{1+\delta}{\kappa[n]}\ln\ln A_n+O\left(\frac{1}{\kappa[n]}\right). \tag{5.24}$$

This and (5.21) give

$$\sum_{i=1}^{n} x_i \le (1+\delta)^2\sqrt{2A_n\ln\ln A_n}+O\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right). \tag{5.25}$$

We have shown that for any $\delta\in(0,1)$ the strategy $\mathcal{P}=\mathcal{P}(\delta)$ given by (5.11) and (5.15), with initial capital (5.16), weakly forces (5.25) or the failure of (5.18). Setting

$$\mathcal{P} := \sum_{j=1}^{\infty} 2^{-j}\mathcal{P}(1/j), \tag{5.26}$$

we obtain a strategy that weakly forces (5.6) starting with initial capital

$$\sum_{j=1}^{\infty} 2^{-j}\mathcal{K}_0(1/j). \tag{5.27}$$

To get rid of "weakly" in "weakly forces", use the device of Lemma 3.2 (p. 68).

It remains to consider the questions of convergence. First we prove that (5.27) is finite. It is easy:

$$\sum_{j=1}^{\infty} 2^{-j}\mathcal{K}_0(1/j) = \sum_{j=1}^{\infty} 2^{-j} \sum_{k=K(1/j)}^{\infty} k^{-1-1/j} \le \sum_{j=1}^{\infty} 2^{-j} = 1, \qquad (5.28)$$

provided that $K(1/j)$ are so big that

$$\sum_{k=K(1/j)}^{\infty} k^{-1-1/j} \le 1, \ \forall j.$$

So it remains to prove the convergence of series (5.15) and (5.26) (getting rid of "weakly" is trivial convergence-wise). Here we use the fact that all our strategies recommend *nonnegative* amounts of $x_n$-tickets and $x_n^2$-tickets. This means that it suffices to prove that the strategy $\mathcal{P}$ recommends a *finite* number of $x_n$-tickets and $x_n^2$-tickets in each situation.

First we note that $\delta \in (0,1)$, $\kappa \in (0,1)$, and (5.11) imply that $\mathcal{K}_i^{(\kappa(k,\delta))}$ (now we explicitly indicate $\delta$ in our notation for $\kappa = \sqrt{2(1+\delta)^{-k} \ln k}$ ) as function of $k$ and $\delta$ is bounded in every fixed situation. Since $\mathcal{P}^{(\kappa(k,\delta))}$ recommends buying (5.12) $x_i$-tickets and (5.13) $x_i^2$-tickets, we are only required to prove that

$$\sum_{j=1}^{\infty} 2^{-j} \sum_{k=K(1/j)}^{\infty} k^{-1-1/j}\kappa(k,1/j) < \infty \qquad (5.29)$$

and

$$\sum_{j=1}^{\infty} 2^{-j} \sum_{k=K(1/j)}^{\infty} k^{-1-1/j}\kappa^2(k,1/j) < \infty. \qquad (5.30)$$

Both (5.29) and (5.30) follow from (5.28).

## 5.3 THE SHARPNESS OF THE ITERATED-LOGARITHM BOUND

Again we assume without loss of generality that $m_n = 0$ for all $n$. Our proof will be traditional, modeled on Kolmogorov (1929); see also Petrov [242, 243].

### Reduction to a Large-Deviation Inequality

Here is the general idea of the proof. Suppose the iterated-logarithm bound is not sharp on a particular sequence $x_1, x_2, \ldots$ output by Reality; $\sum_{i=1}^n x_i$ is too small, at least for large $n$. Divide the whole sequence $x_1 x_2 \ldots$ into pieces of rapidly increasing cumulative variance. For a typical piece $x_n x_{n+1} \ldots x_m$, the sum $x_n + \cdots + x_m$ will also be too small. We will prove a large-deviation inequality that enables Skeptic to increase his capital slightly for such a piece $x_n x_{n+1} \ldots x_m$. The cumulative effect of these slight increases will bring Skeptic's capital from \$1 to infinity.

We define a *stopping time* to be a set of disjoint situations. A stopping time $\tau$ is identified with the extended (this means that a value of $\infty$ is allowed) nonnegative variable that maps every path $\xi$ into $n$, where $n$ is the length of $\xi$'s prefix $\xi^n$ which belongs to $\tau$ if such a prefix exists and $n = \infty$ if such a prefix does not exist.

Here is the large-deviation inequality we need.

**Lemma 5.1** *For arbitrarily small $\epsilon > 0$, for sufficiently small positive $\delta < \delta(\epsilon)$, and for sufficiently large $C > C(\epsilon, \delta)$, the following holds: Consider the unbounded forecasting protocol in which Reality is additionally required to ensure that*

$$|x_n| \le \delta \sqrt{\frac{C}{\ln \ln C}}, \ \forall n. \tag{5.31}$$

*Define a stopping time $\tau$ by*

$$\tau := \min\{n \mid A_n \ge C\}$$

*(so $\tau = \infty$ if $A_n$ never reaches $C$). Then there exists a positive martingale $\mathcal{L}$ such that*

$$\mathcal{L}(\Box) = 1 \qquad \textit{and} \qquad \mathcal{L}(v_1 x_1 \dots v_n x_n) \ge 1 + \frac{1}{\ln C}$$

*for any situation $v_1 x_1 \dots v_n x_n \in \tau$ such that*

$$\sum_{i=1}^{n} x_i \le (1 - \epsilon)\sqrt{2C \ln \ln C}. \tag{5.32}$$

We will write $\mathcal{P}(\epsilon, \delta, C)$ for the strategy for Skeptic whose existence is asserted in this lemma. The lemma could, of course. be stated in terms of upper probability.

Let us show that Lemma 5.1 does imply the sharpness of the iterated-logarithm bound in the predictably unbounded forecasting protocol, ignoring issues of convergence, which we will consider after we prove the lemma. In other words, let us prove that Skeptic can force

$$\left(A_n \to \infty \ \& \ c_n = o\left(\sqrt{\frac{A_n}{\ln \ln A_n}}\right)\right) \implies \left(\limsup_{n \to \infty} \frac{\sum_{i=1}^n x_i}{\sqrt{2A_n \ln \ln A_n}} \ge 1 - 2\epsilon\right) \tag{5.33}$$

for a fixed constant $\epsilon > 0$. Combining the strategies corresponding to $\epsilon_m$ with $\epsilon_m \to 0$ as $m \to \infty$, we shall obtain the strategy forcing the sharpness of the iterated-logarithm bound.

Take $D$ sufficiently large (we will be precise later) and $\delta$ sufficiently small in the sense of Lemma 5.1. Skeptic's strategy for forcing (5.33) can be described as follows (the variable $k$ is the ordinal number of the current piece of the path, or the *cycle* of the loop):

> Start with initial capital $\mathcal{K} := \$1$.
> FOR $k = 1, 2, \dots$:
>   $C := D^k$.
>   Apply strategy $\mathcal{K}\mathcal{P}(\epsilon, \delta, C)$ until $A_n \ge C$ or $c_n > \delta\sqrt{C/\ln \ln C}$.
>   Clear memory, except for the current capital $\mathcal{K}$. (5.34)

(Command (5.34) means that Skeptic should forget all that he has seen so far and act as if the protocol has just started with the initial capital $\mathcal{K}$. In the preceding command, "until" is understood inclusively for $A_n$ but exclusively for $c_n$: the strategy $\mathcal{K}\mathcal{P}(\epsilon, \delta, C)$ is still applied at the step $n$ when $A_n \ge C$ for the first time but is not applied when $c_n > \delta\sqrt{C/\ln \ln C}$ for the first time.) Suppose Skeptic plays this strategy, the iterated-logarithm bound is valid on the path chosen by Forecaster and Reality, and this path satisfies the antecedent but does not satisfy the consequent of (5.33). In this case, $k$ will grow indefinitely, the condition $c_n \le \delta\sqrt{C/\ln \ln C}$ will be always satisfied from some $k$ on, and, from some $k$ on, inequality (5.32) will be satisfied, with $x_1, \dots, x_n$ Reality's moves during the $k$th cycle of the loop in the description of Skeptic's strategy. (Recall that we are assuming that the iterated-logarithm bound is valid on the path chosen by Forecaster and Reality and that $D$ is sufficiently large; these conditions allow us to make the influence of the previous pieces negligible.) Therefore, from some $k$ on, Skeptic's capital will be multiplied by at least

$$1+\frac{1}{\ln C}=1+\frac{1}{k\ln D}$$

at cycle $k$. It remains to notice that

$$\prod_k\left(1+\frac{1}{k\ln D}\right)=\infty.$$

## Proof of the Large-Deviation Inequality (Lemma 5.1)

It suffices to construct a martingale $\mathcal{T}$ that is always less than $\ln C+1$ and satisfies

$$\mathcal{T}(\square)=1 \qquad \text{and} \qquad \mathcal{T}(v_1x_1\dots v_nx_n)\le 0$$

when (5.32) holds. (We can then obtain the required $\mathcal{L}$ by putting $\mathcal{L}:=1+\frac{1-\mathcal{T}}{\ln C}$.) We shall do more: we shall find a martingale $\mathcal{L}$ that never exceeds $\frac{\ln C}{2}$ and satisfies

$$\mathcal{L}(\square)\ge\frac{1}{2} \qquad \text{and} \qquad \mathcal{L}(v_1x_1\dots v_nx_n)\le 0$$

when (5.32) holds.

Our construction is fairly complicated, so we first explain the basic idea. The crucial point, as in the proof of the validity of the iterated-logarithm bound, is that the process

$$\exp\left(\kappa\mathcal{S}_n-\frac{\kappa^2}{2}\mathcal{A}_n\right),$$

where

$$\mathcal{S}_n:=\sum_{i=1}^n x_i,\ \mathcal{A}_n:=\sum_{i=1}^n v_i,$$

is nearly a martingale for small $\kappa$ ($\mathcal{A}_n$ is the same as $A_n$ but is thought of as a process rather than as a number). The initial value of this approximate martingale is 1, and its value in a situation in $\tau$ will be close to

$$\exp\left(\kappa\mathcal{S}-\frac{\kappa^2}{2}C\right), \tag{5.35}$$

and hence will depend, approximately, only on the value of $\mathcal{S}$ at this situation. In this proof we shall only be interested in martingales and approximate martingales whose value in a situation in $\tau$ is at least approximately a function of the value of $\mathcal{S}$ at this situation; this function will be called the *payoff* of the martingale under consideration.

Our goal is to find an approximate martingale whose payoff is bounded above by the function denoted by the thick line in Figure 5.1. As a very crude first approximation to this line, we use the payoff (5.35) of a suitable approximate martingale, which we call the *basic payoff*; the basic payoff is shown as the curved line in Figure 5.1. It is clear from Figure 5.1 that it suffices to remove regions A, B, and C from the basic payoff. We shall do this by subtracting several payoffs of the form (5.35).

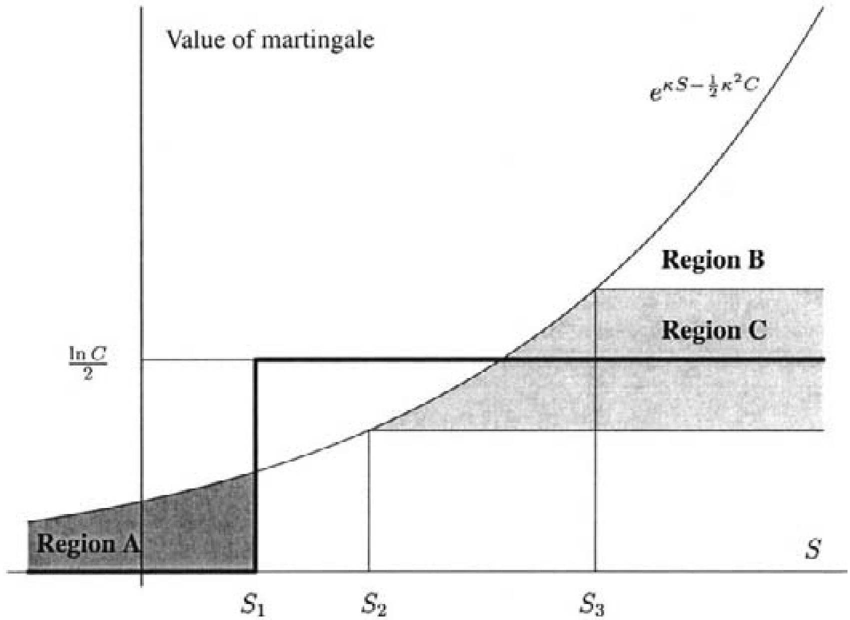



***Fig. 5.1*** Proving the large-deviation inequality by reducing an initial payoff. Here $S_1 = (1-\epsilon)\sqrt{2C \ln\ln C}$, $S_2 = (1+\epsilon^*)^2(1-\epsilon)\sqrt{2C \ln\ln C}$, and $S_3 = 2\sqrt{2C \ln\ln C}$.

To implement our plan, we shall need the following simple auxiliary result: Suppose we are given some value $S$ of $\mathcal{S}$. Then the value of $\kappa$ for which (5.35) attains its maximum is

$$\kappa = S/C, \tag{5.36}$$

and the corresponding value of the payoff (5.35) is

$$\exp\left(\frac{S^2}{2C}\right). \tag{5.37}$$

In this proof, we call this value of $\kappa$ *optimal.*

Let $\epsilon^* > 0$ be a constant, small even compared to $\epsilon$. Our basic payoff will be optimal with respect to the value

$$\mathcal{S} = (1+\epsilon^*)(1-\epsilon)\sqrt{2C \ln\ln C}$$

and will therefore correspond to

$$\kappa = (1+\epsilon^*)(1-\epsilon)\sqrt{\frac{2\ln\ln C}{C}}. \tag{5.38}$$

Let us check that the bottom of the redundant region C lies below the thick line. For this $\kappa$ and for

$$\mathcal{S} = S_2 = (1+\epsilon^*)^2(1-\epsilon)\sqrt{2C \ln\ln C},$$

we obtain the following value for the basic payoff:

$$\exp\left(\kappa S_2 - \frac{\kappa^2}{2}C\right) = \exp\left((1+\epsilon^*)^3(1-\epsilon)^2 2\ln\ln C - (1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C\right)$$
$$= \exp\left((1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C + 2\epsilon^*(1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C\right)$$
$$\le \exp\left((1-\epsilon)\ln\ln C\right) = (\ln C)^{1-\epsilon},$$

which is indeed smaller than $\ln C/2$. Later, in Lemma 5.2, we shall see that we can replace our approximate martingale with a genuine martingale without losing more than a factor $(\ln C)^{O(\delta)}$, and this will not spoil the picture.

Now we remove the regions A, B, and C with martingales having small initial values.

We remove region A using the payoff (5.35) optimal with respect to the rightmost $S$ for this region,

$$S = S_1 = (1-\epsilon)\sqrt{2C\ln\ln C},$$

which corresponds to

$$\kappa = (1-\epsilon)\sqrt{\frac{2\ln\ln C}{C}}. \tag{5.39}$$

We multiply this payoff by a suitable positive weight and subtract it from the basic payoff. Let $\kappa_0$ be the basic value of $\kappa$ given by (5.38). Since the ratio

$$\frac{\exp\left(\kappa S - \frac{\kappa^2}{2}C\right)}{\exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C\right)} \propto e^{(\kappa-\kappa_0)S} \tag{5.40}$$

decreases as $S$ increases, the weight with which we should take the payoff corresponding to (5.39) is

$$\frac{\exp\left(\kappa_0(1-\epsilon)\sqrt{2C\ln\ln C} - \frac{\kappa_0^2}{2}C\right)}{\exp\left(\kappa(1-\epsilon)\sqrt{2C\ln\ln C} - \frac{\kappa^2}{2}C\right)}$$
$$= \frac{\exp\left((1+\epsilon^*)(1-\epsilon)^2 2\ln\ln C - (1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C\right)}{\exp\left((1-\epsilon)^2 2\ln\ln C - (1-\epsilon)^2\ln\ln C\right)}$$
$$= \frac{\exp\left((1+\epsilon^*)(1-\epsilon^*)(1-\epsilon)^2\ln\ln C\right)}{\exp\left((1-\epsilon)^2\ln\ln C\right)}$$
$$= \exp\left(-(\epsilon^*)^2(1-\epsilon)^2\ln\ln C\right) = (\ln C)^{-(\epsilon^*)^2(1-\epsilon)^2} \ll 1.$$

The region $B \cup C$ is much more difficult than A; this is why we split it into B and C. Dealing with B is relatively easy, similar to dealing with A. We remove B using the payoff (5.35) optimal with respect to the leftmost $S$ for this region,

$$S = S_3 = 2\sqrt{2C\ln\ln C},$$

which corresponds to

$$\kappa = 2\sqrt{\frac{2\ln\ln C}{C}}. \tag{5.41}$$

Again, we multiply this optimal payoff by a suitable positive weight and subtract it from the basic payoff. Recall that $\kappa_0$ is the basic value of $\kappa$ given by (5.38). Since the ratio (5.40) now increases as $S$ increases, the weight with which we should take the payoff corresponding to

(5.41) is

$$\frac{\exp\left(\kappa_0 2\sqrt{2C\ln\ln C} - \frac{\kappa_0^2}{2}C\right)}{\exp\left(\kappa 2\sqrt{2C\ln\ln C} - \frac{\kappa^2}{2}C\right)}$$

$$= \frac{\exp\left((1+\epsilon^*)(1-\epsilon)4\ln\ln C - (1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C\right)}{\exp\left(8\ln\ln C - 4\ln\ln C\right)}$$

$$\approx \frac{\exp\left(3\ln\ln C\right)}{\exp\left(4\ln\ln C\right)} = (\ln C)^{-1} \ll 1.$$

Here $\approx$ means approximate equality taking into account the smallness of $\epsilon$ and $\epsilon^*$.

Removing region C is most difficult and will be accomplished by subtracting not just one but continuously many payoffs. To remove the narrow strip of C between the horizontal lines at heights

$$\exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C\right) \quad \text{and} \quad \exp\left(\kappa_0(S+dS) - \frac{\kappa_0^2}{2}C\right),$$

we use the payoff (5.35) optimal at $S$, corresponding to $\kappa = S/C$ and taking the value $e^{\frac{S^2}{2C}}$ at $S$; see (5.36) and (5.37). Since the width of this strip is

$$\exp\left(\kappa_0(S+dS) - \frac{\kappa_0^2}{2}C\right) - \exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C\right)$$

$$= \exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C\right)\left(e^{\kappa_0 dS} - 1\right) \sim \exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C\right)\kappa_0 dS,$$

this payoff should be taken with the weight

$$\frac{\exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C\right)\kappa_0 dS}{\exp\left(\frac{S^2}{2C}\right)} = \kappa_0 dS \exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C - \frac{S^2}{2C}\right),$$

and so the total weight for the whole of C is

$$\int_{(1+\epsilon^*)^2(1-\epsilon)\sqrt{2C\ln\ln C}}^{2\sqrt{2C\ln\ln C}} \kappa_0 \exp\left(\kappa_0 S - \frac{\kappa_0^2}{2}C - \frac{S^2}{2C}\right) dS. \tag{5.42}$$

The expression under the integral sign attains its maximum in $S$ when

$$S = \kappa_0 C = (1+\epsilon^*)(1-\epsilon)\sqrt{2C\ln\ln C},$$

which is to the left of the region of integration in (5.42). Since the integrand decreases over the region of integration, we can bound (5.42) from above by

$$2\sqrt{2C\ln\ln C}\kappa_0 \exp\left(\kappa_0(1+\epsilon^*)^2(1-\epsilon)\sqrt{2C\ln\ln C} - \frac{\kappa_0^2}{2}C \right.$$
$$\left. - \frac{((1+\epsilon^*)^2(1-\epsilon)\sqrt{2C\ln\ln C})^2}{2C}\right)$$

$$= 4(1+\epsilon^*)(1-\epsilon)\ln\ln C \exp\Bigl(2(1+\epsilon^*)^3(1-\epsilon)^2\ln\ln C$$

$$-(1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C - (1+\epsilon^*)^4(1-\epsilon)^2\ln\ln C\Bigr)$$

$$= 4(1+\epsilon^*)(1-\epsilon)\ln\ln C$$

$$\exp\Bigl((1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C\bigl(2(1+\epsilon^*)-1-(1+\epsilon^*)^2\bigr)\Bigr)$$

$$= 4(1+\epsilon^*)(1-\epsilon)\ln\ln C \exp\Bigl((1+\epsilon^*)^2(1-\epsilon)^2\ln\ln C\bigl(-(\epsilon^*)^2\bigr)\Bigr)$$

$$= 4(1+\epsilon^*)(1-\epsilon)\frac{\ln\ln C}{(\ln C)^{(\epsilon^*)^2(1+\epsilon^*)^2(1-\epsilon)^2}} \ll 1.$$

We have now removed all three redundant regions shown in Figure 5.1.

The next lemma shows that our approximate martingales can be replaced by genuine martingales.

**Lemma 5.2** *Consider the protocol of Lemma 5.1. Let*

$$(1-\epsilon)\sqrt{\frac{2\ln\ln C}{C}} \le \kappa \le 2\sqrt{\frac{2\ln\ln C}{C}}. \tag{5.43}$$

*There exists a positive martingale $\mathcal{L}$ that starts with* \$1 *and satisfies*

$$(\ln C)^{-8\delta} \le \frac{\mathcal{L}(s)}{\exp\left(\kappa S(s) - \frac{\kappa^2}{2}C\right)} \tag{5.44}$$

*for every situation $s \in \tau$. There exists another positive martingale $\mathcal{L}$ that starts from* \$1 *and satisfies*

$$\frac{\mathcal{L}(s)}{\exp\left(\kappa S(s) - \frac{\kappa^2}{2}C\right)} \le (\ln C)^{8\delta} \tag{5.45}$$

*for every situation $s \in \tau$.*

*Proof* First we prove (5.44). As in the proof of validity, we consider the martingale $\mathcal{L}$ satisfying $\mathcal{L}(\Box) = 1$ and

$$\mathcal{L}_{i+1} = \mathcal{L}_i \frac{1+\kappa x_i + (1+\delta)\kappa^2 x_i^2/2}{1+(1+\delta)\kappa^2 v_i/2} \tag{5.46}$$

for $i = 1, 2, \ldots$ (cf. (5.11) on p. 105). We need to prove

$$(\ln C)^{-8\delta} \le \frac{\prod_{i=1}^{\tau} \frac{1+\kappa x_i+(1+\delta)\kappa^2 x_i^2/2}{1+(1+\delta)\kappa^2 v_i/2}}{\exp\left(\kappa S(s) - \frac{\kappa^2}{2}C\right)}. \tag{5.47}$$

Using the inequalities

$$|t| \le 2\sqrt{2}\delta \Longrightarrow 1 + t + (1+\delta)\frac{t^2}{2} \ge e^t \tag{5.48}$$

(this elaboration of (5.23) is true for $\delta$ small enough) and

$$|\kappa x_i| \le 2\sqrt{\frac{2\ln\ln C}{C}}\delta\sqrt{\frac{C}{\ln\ln C}} = 2\sqrt{2}\delta \tag{5.49}$$

(cf. (5.43) and (5.31)), we reduce the task of proving (5.47) to proving

$$(\ln C)^{-8\delta} \le \frac{\prod_{i=1}^{\tau}\frac{e^{\kappa x_i}}{1+(1+\delta)\kappa^2 v_i/2}}{\exp\left(\kappa\mathcal{S}(s) - \frac{\kappa^2}{2}C\right)},$$

or

$$\prod_{i=1}^{\tau}\left(1 + (1+\delta)\kappa^2 v_i/2\right) \le (\ln C)^{8\delta}\exp\left(\frac{\kappa^2}{2}C\right),$$

which is equivalent to

$$\sum_{i=1}^{\tau}\ln\left(1 + (1+\delta)\kappa^2 v_i/2\right) \le 8\delta\ln\ln C + \frac{\kappa^2}{2}C.$$

Using the inequality $\ln(1+t) \le t$, we can see that it suffices to prove

$$\sum_{i=1}^{\tau}(1+\delta)\kappa^2 v_i/2 \le 8\delta\ln\ln C + \frac{\kappa^2}{2}C. \tag{5.50}$$

We can assume, without loss of generality, that

$$v_i \le \delta^2\frac{C}{\ln\ln C}, \ \forall i \tag{5.51}$$

(see (5.31)). Using the definition of $\tau$, we can therefore reduce the inequality (5.50) to

$$(1+\delta)\frac{\kappa^2}{2}\left(C + \delta^2\frac{C}{\ln\ln C}\right) \le 8\delta\ln\ln C + \frac{\kappa^2}{2}C,$$

or

$$\delta\frac{\kappa^2}{2}C + (1+\delta)\frac{\kappa^2}{2}\delta^2\frac{C}{\ln\ln C} \le 8\delta\ln\ln C;$$

recalling (5.43), we further reduce it to

$$4\ln\ln C + 4(1+\delta)\delta \le 8\ln\ln C,$$

which is indeed true for small $\delta$ and large $C$.

Now let us prove (5.45). Here we consider the martingale $\mathcal{L}$ satisfying $\mathcal{L}(\Box) = 1$ and

$$\mathcal{L}_{i+1} = \mathcal{L}_i\frac{1 + \kappa x_i + (1-\delta)\kappa^2 x_i^2/2}{1 + (1-\delta)\kappa^2 v_i/2} \tag{5.52}$$

for $i = 1, 2, \ldots$. (Notice that it is positive for small $\delta$.) We need to prove

$$\frac{\prod_{i=1}^{\tau} \frac{1+\kappa x_i+(1-\delta)\kappa^2 x_i^2/2}{1+(1-\delta)\kappa^2 v_i/2}}{\exp\left(\kappa \mathcal{S}(s) - \frac{\kappa^2}{2} C\right)} \le (\ln C)^{8\delta}. \tag{5.53}$$

The inequalities (5.49) and

$$|t| \le 2\sqrt{2}\delta \Longrightarrow 1 + t + (1-\delta)\frac{t^2}{2} \le e^t$$

(this is analogous to (5.48)) reduce (5.53) to

$$\frac{\prod_{i=1}^{\tau} \frac{e^{\kappa x_i}}{1+(1-\delta)\kappa^2 v_i/2}}{\exp\left(\kappa \mathcal{S}(s) - \frac{\kappa^2}{2} C\right)} \le (\ln C)^{8\delta},$$

or

$$\exp\left(\frac{\kappa^2}{2} C\right) \le (\ln C)^{8\delta} \prod_{i=1}^{\tau} \left(1 + (1-\delta)\frac{\kappa^2}{2} v_i\right).$$

Taking logarithms of both sides, we obtain

$$\frac{\kappa^2}{2} C \le 8\delta \ln\ln C + \sum_{i=1}^{\tau} \ln\left(1 + (1-\delta)\frac{\kappa^2}{2} v_i\right).$$

Since $\ln(1+t) \ge t - t^2/2$ for all nonnegative $t$, it suffices to prove

$$\frac{\kappa^2}{2} C \le 8\delta \ln\ln C + \sum_{i=1}^{\tau} \left((1-\delta)\frac{\kappa^2}{2} v_i - (1-\delta)^2 \frac{\kappa^4}{8} v_i^2\right). \tag{5.54}$$

Since

$$C \le \sum_{i=1}^{\tau} v_i \le \frac{3}{2} C \tag{5.55}$$

for large $C$ and small $\delta$ (see (5.51) and remember the definition of $\tau$), (5.54) reduces to

$$(1-\delta)^2 \frac{\kappa^4}{8} \sum_{i=1}^{\tau} v_i^2 \le 8\delta \ln\ln C - \delta \frac{\kappa^2}{2}\left(\frac{3}{2} C\right);$$

by (5.43), this further reduces to

$$8(\ln\ln C/C)^2 \sum_{i=1}^{\tau} v_i^2 \le 8\delta \ln\ln C - 6\delta \ln\ln C,$$

or

$$\sum_{i=1}^{\tau} v_i^2 \le \frac{\delta C^2}{4 \ln\ln C}. \tag{5.56}$$

According to (5.51) and (5.55),

$$\sum_{i=1}^{\tau} v_i^2 \le \left(\max_i v_i\right) \sum_{i=1}^{\tau} v_i \le \delta^2 \frac{C}{\ln\ln C} \frac{3}{2} C = \frac{3\delta^2 C^2}{2 \ln\ln C},$$

which is stronger than (5.56) for small $\delta$. ∎

Our proof of the large-deviation lemma (Lemma 5.1) is now complete except for convergence issues (considered in the next paragraph) and the following small point: we were required to construct a martingale $\mathcal{L}$ that never exceeds $\ln C/2$ in the situations in and before the stopping time $\tau$, but so far we have only shown that our $\mathcal{L}$ does not exceed $\ln C/2$ in the situations in $\tau$. Let us show that $\mathcal{L}$ does not exceed $\ln C/2$ in the situations before $\tau$ as well. (This would be obvious were $\tau$ a *complete* stopping time—one that intersects every path.) Suppose Skeptic plays the strategy we have constructed (the one whose capital process is the martingale $\mathcal{L}$ guaranteed not to exceed $\ln C/2$ inside $\tau$). Suppose that in some situation $s$ before $\tau$ we have $\mathcal{L}(s) > \ln C/2$. We shall arrive at a contradiction considering the following strategy for Forecaster and Reality. After $s$, Forecaster always sets

$$v_n := \delta^2 \frac{C}{\ln \ln C},$$

and Reality always sets

$$x_n := \begin{cases} \delta\sqrt{\frac{C}{\ln \ln C}} & \text{if } M_n \geq 0 \\ -\delta\sqrt{\frac{C}{\ln \ln C}} & \text{if } M_n < 0. \end{cases}$$

This ensures that Skeptic's capital $\mathcal{L}_n$ will never decrease after $s$ (the $x_n^2$-tickets will produce zero loss since their price always equals their payoff, and the $x_n$-tickets will produce a profit unless $M_n = 0$) and that $\mathcal{A}_n$ will eventually exceed $C$ (in other words, $\tau$ will be reached).

## Convergence Check

Let us fix a cycle $k$ of Skeptic's strategy and a step $n$ (counting from the beginning of the game) in this cycle; we now show that the sums of the amounts of $x_n$- and $x_n^2$-tickets required by the constituent strategies (defined by (5.46) and (5.52)) converges; our argument will also show that the sum of the initial capitals required by the constituent strategies converges. According to (5.49), (5.46), and (5.52), $\mathcal{L}_n$ at most doubles at every step; therefore, $\mathcal{L}_n \leq 2^n$, and every constituent strategy involves amounts of $x_n$-tickets bounded (in absolute value) by $\kappa 2^n$ and amounts of $x_n^2$ tickets bounded by $\kappa^2 2^n$. Since $\kappa$ satisfies (5.43), we can assume $\kappa < 1$, and so these amounts are bounded by some constant (namely, $2^n$) which depends only on $n$.

Now we will list all constituent strategies and their weights:

- the one corresponding to the basic payoff is taken with the weight 1;
- the one removing region A is taken with weight $\ll 1$;
- the one removing region B is taken with weight $\ll 1$;
- the family removing region C has cumulative weight (5.42), which we have shown is $\ll 1$.

This shows that the combined strategy for Lemma 5.1 also requires amounts of tickets bounded by a constant depending only on $n$.

The final step in the proof of the theorem is to combine the different $\epsilon$ in (5.33). It suffices to take the strategy corresponding to $\epsilon_m > 0$ with weight $p_m > 0$, where $\epsilon_m$ is any sequence convergent to 0 and $\sum p_m$ is any convergent series.

## 5.4 A MARTINGALE LAW OF THE ITERATED LOGARITHM

Recall that in a symmetric probability protocol (a probability protocol in which Skeptic can take either side of any gamble), we call the capital processes for Skeptic *martingales* (p. 67 and p. 82). Recall also the game-theoretic definition of *quadratic supervariation* given on p. 91: a quadratic supervariation for a martingale $\mathcal{S}$ is an increasing predictable process $\mathcal{A}$ such that the inequality (4.22) holds for some martingale $\mathcal{K}$. If the inequality holds with equality, we will call $\mathcal{A}$ a (game-theoretic) *quadratic variation* for $\mathcal{S}$.

The following propositions follow from Theorems 5.1 and 5.2, by the method we used to prove Proposition 4.4 (p. 92):

**Proposition 5.2** *Suppose $\mathcal{A}$ is a quadratic supervariation of a martingale $\mathcal{S}$ in a symmetric probability protocol. Then Skeptic can force*

$$\left(\mathcal{A}_n \to \infty \ \& \ |\Delta\mathcal{S}_n| = o\left(\sqrt{\frac{\mathcal{A}_n}{\ln\ln\mathcal{A}_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{|\mathcal{S}_n|}{\sqrt{2\mathcal{A}_n \ln\ln\mathcal{A}_n}} \leq 1,$$

*where $\Delta\mathcal{S}_n := \mathcal{S}_n - \mathcal{S}_{n-1}$.*

**Proposition 5.3** *Suppose $\mathcal{A}$ is a quadratic variation of a martingale $\mathcal{S}$ in a symmetric probability protocol. And suppose $c_1, c_2, \ldots$ is a predictable process such that $|\Delta\mathcal{S}_n| \leq c_n$ for all $n$. Then Skeptic can force*

$$\left(\mathcal{A}_n \to \infty \ \& \ c_n = o\left(\sqrt{\frac{\mathcal{A}_n}{\ln\ln\mathcal{A}_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\mathcal{S}_n}{\sqrt{2\mathcal{A}_n \ln\ln\mathcal{A}_n}} = 1.$$

In Chapter 15, we will apply Proposition 5.2 to securities markets.

Similar results hold when Skeptic's move space is an arbitrary closed convex cone in a Banach space, as in Proposition 4.5 on p. 93.

## 5.5 APPENDIX: HISTORICAL COMMENTS

This appendix surveys the work leading up to Khinchin's law of the iterated logarithm and the subsequent development of the law within the measure-theoretic framework.

As we mentioned in Chapter 2, Émile Borel's original proof of his strong law for coin tossing, published in 1909, already gave an upper bound on the rate of convergence. This bound was far from sharp; it was of order $\sqrt{n}\ln n$ rather than $\sqrt{n}\ln\ln n$. Borel made it clear, however, that he was not trying to give the tightest possible bound, and his method of proof—approximating the probability of his bound being violated for each $n$ and then concluding that this would happen for only finitely many $n$ because the sum of the probabilities converges—readily yields tighter bounds.

In a two-page account of Borel's strong law in the chapter on measure theory at the end of his first book on set theory (1914), Felix Hausdorff (1868–1942) gave a different (and more complete) proof than Borel but announced a bound on the rate of convergence, $n^{1/2+\epsilon}$ for any $\epsilon > 0$, that is actually weaker than Borel's bound. Perhaps because of the shortcomings of Borel's proof, Hausdorff's contribution has been seen as an important step on the way from Borel to Khinchin. Feller ([120], p. 196) states that Khinchin's law gave an answer to a problem "treated in a series of papers initiated by" Hausdorff and Hardy and Littlewood. Chow and Teicher ([51], p. 368) state that Kolmogorov's version of the law "was the culmination of a series of strides by mathematicians of the caliber of Hausdorff, Hardy, Littlewood, and Khintchine". Other authors [134, 188] make similar statements.

The English mathematicians Godfrey H. Hardy and John E. Littlewood did touch on the rate of convergence in Borel's law in a lengthy paper on number theory in 1914 ([145], p. 190). Their method was essentially the same as Borel's, but they found the tightest asymptotic bound the method will yield: with probability one, the deviation of $\sum_{i=1}^n x_i$ from zero is, in their words, "not of order exceeding $\sqrt{n \ln n}$". They also made a start, as neither Borel nor Hausdorff had done, on investigating how large a deviation can be counted on; they showed that with probability one, "the deviation, in both directions, is sometimes of order exceeding $\sqrt{n}$".

The big step after Borel was really Khinchin's alone. Borel, Hausdorff, and Hardy and Littlewood had all followed the same strategy: they estimated the probability of a large deviation for $\sum_{i=1}^n x_i$ separately for each $n$ and then summed these probabilities. This is crude, because the deviations are highly dependent for adjacent $n$. Khinchin estimated the probability of at least one deviation within increasingly wide ranges of $n$. This much more difficult calculation, which Khinchin made in 1922 and published in 1924, yielded an asymptotic bound on the rate convergence in Borel's strong law that was sharp as well as valid. Like his predecessors, Khinchin considered independent tosses of a coin with any probability $p$ for heads; he showed that the number $y_n$ of heads will almost surely satisfy

$$\limsup_{n \to \infty} \frac{y_n - np}{\sqrt{2pqn \ln \ln n}} = 1.$$

In 1929, Kolmogorov generalized Khinchin's law from coin tossing to independent random variables $x_1, x_2, \ldots$ with means $m_1, m_2, \ldots$ and variances $v_1, v_2, \ldots$; he showed that (5.3), with $A_n := \sum_{i=1}^n v_i$, holds almost surely in the measure-theoretic sense if $A_n$ tends to infinity and the $x_n - m_n$ stay within bounds that do not grow too fast. Stout 1970 generalized Kolmogorov's result to the martingale setting, where the $m_n$ and $v_n$ are conditional means and variances relative to a filtration to which $x_1, x_2, \ldots$ is adapted. In Chapter 8 (p. 172), we deduce Stout's theorem from our game-theoretic result.

In 1941, Hartman and Wintner obtained a law of the iterated logarithm for independent identically distributed random variables. Although they used Kolmogorov's theorem as the starting point of their proof, their result does not follow simply from Kolmogorov's theorem or from Stout's. Nor does it have a simple game-theoretic counterpart. A game could no doubt be constructed on the basis of their proof, but it would price so many functions of the $x_n$ that it would hold little interest. In general, we consider a game-theoretic generalization of a measure-theoretic result interesting only if it manages to reduce significantly the number of prices that are assumed.

## 5.6 APPENDIX: KOLMOGOROV'S FINITARY INTERPRETATION

Infinity is thoroughly embedded in the usual statement of the law of the iterated logarithm. It is possible, however, to formulate the law in finitary terms, thereby making clearer its relevance to practical problems. In this appendix, we review the finitary formulation given by Kolmogorov in 1929.

We consider only the fair coin. Here the law can be written

$$\limsup_{n\to\infty} \frac{y_n - \frac{n}{2}}{\sqrt{\frac{n}{2}\ln\ln n}} = 1 \tag{5.57}$$

(cf. (5.1)). According to Kolmogorov, this equation says two things:

1. For arbitrary positive numbers $\eta$ and $\delta$ one can find a natural number $n$ such that for any $p$ the probability that at least one of the inequalities

   $$\frac{y_k - \frac{k}{2}}{\sqrt{\frac{k}{2}\ln\ln k}} > 1 + \delta \; (k = n, n+1, \ldots, n+p)$$

   is satisfied is less than $\eta$.

2. For arbitrary positive numbers $\eta$ and $\delta$, and any natural number $m$, one can find a natural number $q$ such that the probability of simultaneous satisfaction of all the inequalities

   $$\frac{y_k - \frac{k}{2}}{\sqrt{\frac{k}{2}\ln\ln k}} < 1 - \delta \; (k = m, m+1, \ldots, m+q)$$

   is less than $\eta$.

For practical applications, we need to go further, explicitly telling how large $n$ and $q$ need to be in terms of the given parameters. In Point 1, for example, we need an explicit function $n(\eta, \delta)$ that outputs the required $n$ from $\eta$ and $\delta$.

The probabilities referred to in Points 1 and 2 can be interpreted game-theoretically. (In fact, as we explain in §8.2, the measure-theoretic and game-theoretic frameworks are essentially equivalent for coin tossing.) Further development of finitary versions of the game-theoretic strong limit theorems is outside the scope of this book, but some work in this direction is provided in [322].

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 6 The Weak Laws

We now turn from the strong laws, which have occupied us for the last three chapters, to the weak laws—the weak law of large numbers and the central limit theorem. Though less elegant than the strong laws, the weak laws are a step closer to practice, because they do not use the concept of infinity.

The most elementary weak laws are Bernoulli's theorem (the weak law of large numbers for coin tossing) and De Moivre's theorem (the central limit theorem for coin tossing). In this chapter, we formulate and prove game-theoretic versions of these theorems, for the simplest case, where the coin is fair. We use the same game as in §3.1, except that there are only $N$ tosses: before each toss, Skeptic is allowed to bet at even odds on whether Reality will produce a head or a tail. Our game-theoretic version of Bernoulli's theorem says that

$$\overline{\underline{\mathbb{P}}}\left\{\left|\frac{y}{N}-\frac{1}{2}\right|<\epsilon\right\}>1-\delta \tag{6.1}$$

for fixed $\epsilon > 0$ and $\delta > 0$ when $N$ is sufficiently large, where $y$ is the number of heads. Our game-theoretic version of De Moivre's theorem says that

$$\overline{\underline{\mathbb{P}}}\left\{a<\frac{y-N/2}{\sqrt{N/4}}<b\right\}\approx\frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}dz \tag{6.2}$$

for fixed $a$ and $b$ when $N$ is sufficiently large. The symbol $\overline{\underline{\mathbb{P}}}$ represents both upper and lower probability; (6.1) and (6.2) are each to be read as the conjunction of two statements, one with $\overline{\mathbb{P}}$ and the other with $\underline{\mathbb{P}}$. These two game-theoretic theorems are exactly the same as the classical theorems for the fair coin, except that we have replaced $\mathbb{P}$ with $\overline{\underline{\mathbb{P}}}$; cf. (2.1) and (2.2).

Our proof of the game-theoretic version of De Moivre's theorem is the part of this chapter that is most important for the rest of the book. It uses the method of proof first set out by Lindeberg in his 1920 and 1922 articles on the central limit theorem, and it makes two fundamental aspects of this method explicit: (1) the method's game-theoretic character, and (2) its relation to the heat equation. This is only the first of several times that we use Lindeberg's method. We use it in the next chapter to prove Lindeberg's more general central limit theorem, and we use it again in Part II to price European options in the Bachelier and Black-Scholes models. The applications of Lindeberg's method in these later chapters are self-contained, but the exposition in this chapter is much more transparent.

In order to explain the game-theoretic character of De Moivre's theorem, it is convenient to write $x_n$ for a variable that is equal to $1/\sqrt{N}$ when the $n$th toss is a head and $-1/\sqrt{N}$ when it is a tail. We then have

$$\sum_{n=1}^{N} x_n = y\frac{1}{\sqrt{N}} + (N-y)\left(-\frac{1}{\sqrt{N}}\right) = \frac{y - N/2}{\sqrt{N/4}},$$

and we can rewrite (6.2) as

$$\overline{\mathbb{P}}\left\{a < \sum_{n=1}^{N} x_n < b\right\} \approx \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}dz. \tag{6.3}$$

We actually prove a more general approximation:

$$\overline{\mathbb{E}}\,U\left(\sum_{n=1}^{N} x_n\right) \approx \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} U(z)e^{-z^2/2}dz, \tag{6.4}$$

where $U$ is any well-behaved function. Equation (6.4) reduces to (6.3) when $U$ is the indicator function for the interval $[a, b]$. The generalization from (6.3) to (6.4) is of little interest in typical applications of De Moivre's theorem, which rely on the fundamental interpretative hypothesis, because this hypothesis gives an intuitive meaning only to probabilities close to one, all of which can be calculated using (6.3). But it is mathematically more convenient to work with (6.4).

Equation (6.4) gives an approximate price at the beginning of the game for the payoff $U(\sum_{n=1}^{N} x_n)$ at the end of the game. We can also give an approximate price at intermediate points; the approximate price just after the $n$th toss is

$$\overline{U}\left(\sum_{i=1}^{n} x_i, (N-n)/N\right), \tag{6.5}$$

where

$$\overline{U}(s, D) := \int_{-\infty}^{\infty} U(s+z)\,\mathcal{N}_{0,D}(dz).$$

(Here we write $\mathcal{N}_{\mu,\sigma^2}$ for the Gaussian distribution with mean $\mu$ and variance $\sigma^2$; see §6.4.) At the beginning of the game, when $n = 0$, the expression (6.5) reduces to $\overline{U}(0,1)$, which is equal to the right-hand side of (6.4), as required. At the end, when $n = N$, it correctly gives the payoff itself, $U(\sum_{n=1}^{N} x_n)$, because the Gaussian distribution $\mathcal{N}_{0,D}$ approaches a probability distribution assigning probability one to 0 as $D$ approaches 0. Because (6.5) gives the correct price at the end of the game, we can show that it gives an approximately correct price earlier by showing that it is approximately a martingale (a capital process for Skeptic). We do this by expanding its increments in a Taylor's series and eliminating some of the terms using the fact that $\overline{U}(s,D)$ satisfies

$$\frac{\partial \overline{U}}{\partial D} = \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}, \tag{6.6}$$

the heat equation. In fact, $\overline{U}(s,D)$ is the solution of the Cauchy problem for the heat equation with the initial condition given by $U$. This means that it satisfies the equation for all $s$ and all $D > 0$, subject to the condition that it approach $U(s)$ as $D$ approaches 0. When we think of $D$ as time, $\overline{U}(s,D)$ describes the propagation of heat from an initial temperature distribution $U(s)$ at time 0; see Figure 6.1 on p. 130.

Bernoulli's theorem generalizes readily from coin tossing to bounded forecasting, where Reality can choose each $x_n$ freely from a bounded interval of real numbers. The generalization of De Moivre's theorem to bounded forecasting is more interesting. In general, we do not obtain an approximate price for the payoff $U(\sum_{n=1}^{N} x_n)$ in the bounded forecasting game; instead we obtain upper and lower prices that may be quite different. Upper prices, as it turns out, are given by a function $\overline{U}(s,D)$ that satisfies

$$\overline{U}(s,D) \geq U(s).$$

This function describes heat propagation with a heat source at each point $s$ that switches on when the temperature at $s$ threatens to drop below $U(s)$, as in Figure 6.2 on p. 135; only at times $D$ when $\overline{U}(s,D) > U(s)$ is the evolution of the temperature governed by the heat equation. As we will see, some understanding of $\overline{U}(s,D)$ can be obtained using elementary results from the parabolic potential theory developed by Joseph L. Doob in his *Classical Potential Theory and Its Probabilistic Counterpart* (1984).

We call the version of De Moivre's theorem that we obtain for bounded forecasting a *one-sided* central limit theorem. It gives us high lower probabilities for Reality's average move $\sum_{n=1}^{N} x_n$ ending up within certain distances of zero, but unlike De Moivre's theorem, it does not give us any similar reassurance about this average move ending up even a small distance away from zero. This is a natural result of the difference between bounded forecasting and coin tossing. In the bounded forecasting protocol, Skeptic cannot force Reality away from zero, for Skeptic gains nothing when Reality sets all her $x_n$ exactly equal to zero. In coin tossing, in contrast, Reality must set each $x_n$ equal to $1/\sqrt{N}$ or $-1/\sqrt{N}$, and if she makes the average too close to zero, she gives Skeptic an opening to get rich by betting that the average will move towards zero whenever it is a slight distance away.

We prove the game-theoretic version of Bernoulli's theorem in §6.1, the game-theoretic version of De Moivre's theorem in §6.2, and the one-sided central limit theorem for bounded forecasting in §6.3.

## 6.1 BERNOULLI'S THEOREM

Jacob Bernoulli (1654–1705). His theorem first appeared in his posthumously published book *Ars conjectandi*.

In this section, we prove two game-theoretic versions of Bernoulli's theorem. The first does not explicitly use upper and lower probability; it merely asserts that Skeptic can multiply his capital in the fair-coin game by a certain factor if the proportion of heads, $y/N$, does not approximate one-half closely enough. The second does use upper and lower probability; it says that

$$\underline{\mathbb{P}}\left\{\left|\frac{y}{N}-\frac{1}{2}\right|<\epsilon\right\}>1-\delta$$

holds for given $\epsilon$ and $\delta$ when $N$ is large enough. The second version follows from the first, because the factor by which Skeptic can multiply his stake determines the lower probability for the event $|\frac{y}{N}-\frac{1}{2}|<\epsilon$.

### Bernoulli's Theorem without Probability

To facilitate the generalization from coin tossing to bounded forecasting, we code Reality's move $x_n$ as 1 for heads and $-1$ for tails. We then define a process $\mathcal{S}$ by

$$\mathcal{S}_0 := 0 \quad \text{and} \quad \mathcal{S}_n := \sum_{i=1}^{n} x_i \text{ for } n = 1, \ldots, N.$$

With this coding, the condition that $y/N$ be close to one-half becomes the condition that $\mathcal{S}_N/N$ be close to zero.

We also find it convenient to start Skeptic's capital at a small positive number $\alpha$ and to challenge him to multiply it by the large factor $\alpha^{-1}$. This produces the following coherent symmetric probability game.

THE FINITE-HORIZON FAIR-COIN GAME
**Parameters:** $N$, $\epsilon > 0, \alpha > 0$
**Players:** Reality, Skeptic
**Protocol:**
$\mathcal{K}_0 := \alpha$.
FOR $n = 1, \ldots, N$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{-1, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.
**Winning:** Skeptic wins if $\mathcal{K}_n$ is never negative and either $\mathcal{K}_N \geq 1$ or $|\mathcal{S}_N/N| < \epsilon$.

This is a finitary version of the fair-coin game of Chapter 3. The number of rounds of play is finite, and instead of trying to make his capital infinite, Skeptic only tries to multiply it by a large factor $\alpha^{-1}$. And instead of trying to make the average of Reality's moves converge exactly to zero, Skeptic only tries to keep it close to zero.

Bernoulli's theorem says that Skeptic has a winning strategy if $N$ is large enough. Our proof uses the following lemma.

**Lemma 6.1** *Set*

$$\mathcal{L}_n := \frac{\mathcal{S}_n^2 + N - n}{N} \text{ for } n = 0, 1, \ldots, N. \tag{6.7}$$

*Then $\mathcal{L}_n$ is a nonnegative martingale, with $\mathcal{L}_0 = 1$.*

*Proof* Because $\mathcal{S}_n^2 - \mathcal{S}_{n-1}^2 = 2\mathcal{S}_{n-1}x_n + x_n^2$, the increment of $\mathcal{S}_n^2 - n$ is

$$(\mathcal{S}_n^2 - n) - (\mathcal{S}_{n-1}^2 - (n-1)) = 2\mathcal{S}_{n-1}x_n + (x_n^2 - 1). \tag{6.8}$$

Since $x_n^2 = 1$, $\mathcal{S}_n^2 - n$ is a martingale; it is obtained by starting with capital 0 and then buying $2\mathcal{S}_{n-1}$ tickets on the $n$th round. The process $\mathcal{L}_n$ is therefore also a martingale. Moreover, it is obvious from (6.7) that $\mathcal{L}_0 = 1$ and $\mathcal{L}_n \geq 0$ for $n = 1, \ldots, N$. ∎

**Proposition 6.1 (Bernoulli's Theorem)** *Skeptic has a winning strategy in the finite-horizon fair-coin game if $N \geq 1/(\alpha\epsilon^2)$.*

*Proof* Suppose Skeptic starts with $\alpha$ and plays $\alpha\mathcal{P}$, where $\mathcal{P}$ is a strategy that produces the martingale $\mathcal{L}_n$ when he starts with 1. His capital at the end of the game is then $\alpha\mathcal{S}_N^2/N$, and if this is 1 or more, then he wins. Otherwise $\alpha\mathcal{S}_N^2/N < 1$. Multiplying this by $1/(\alpha\epsilon^2) \leq N$, we obtain $|\mathcal{S}_N/N| < \epsilon$; Skeptic again wins. ∎

We can easily generalize this proof to a bounded forecasting game in which the two-element set $\{-1, 1\}$ is replaced by the interval of real numbers $[-1, 1]$. In this game, $\mathcal{S}_n - n$ and $\mathcal{L}_n$ are merely nonnegative supermartingales, not necessarily nonnegative martingales. As we learned in §4.2, a *supermartingale* is a process for Skeptic's capital obtained when Skeptic is allowed to give away money on each round. To see that $\mathcal{S}_n - n$ is a supermartingale in the bounded forecasting protocol, we use the fact that $x_n^2 \leq 1$. By (6.8), Skeptic gets $\mathcal{S}_n^2 - n$ when he buys $2\mathcal{S}_{n-1}$ tickets and gives away $1 - x_n^2$ on the $n$th round.

We can generalize further, as we did for the strong law in Chapters 3 and 4. But let us instead turn to see how the weak law for bounded forecasting can be re-expressed in terms of upper and lower probability.

## Bernoulli's Theorem with Upper or Lower Probability

We now drop the condition for winning, and merely consider the upper and lower probabilities determined by the fair-coin protocol:

**Protocol:**

$\mathcal{K}_0 := 1$.

FOR $n = 1, \ldots, N$:

Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{-1, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.

Recall that the upper probability of an event $E$ measures the degree to which a strategy for betting on $E$ can multiply one's capital without risk of bankruptcy:

$$\overline{\mathbb{P}}\, E = \inf\{\alpha \mid \text{there is a strategy that begins with } \alpha \text{ and ends up with at least 1 if } E \text{ happens and at least 0 otherwise}\}.$$

(This formula follows from (1.1) and (1.4), on pp. 12 and 15, respectively.) Substituting $\{|\mathcal{S}/N| \geq \epsilon\}$ for $E$, we obtain

$$\overline{\mathbb{P}}\{|\mathcal{S}/N| \geq \epsilon\} = \inf\{\alpha \mid \text{there is a strategy that begins with } \alpha \text{ and ends up with at least 1 if } |\mathcal{S}/N| \geq \epsilon \text{ and at least 0 if } |\mathcal{S}/N| < \epsilon\}.$$

According to Proposition 6.1, Skeptic has the desired strategy if $N \geq 1/(\alpha\epsilon^2)$, or $\alpha \geq 1/(N\epsilon^2)$. So

$$\overline{\mathbb{P}}\left\{\left|\frac{\mathcal{S}_N}{N}\right| \geq \epsilon\right\} \leq \frac{1}{N\epsilon^2}.$$

Because lower probability is never greater than upper probability in a coherent protocol, we can also write

$$\underline{\mathbb{P}}\left\{\left|\frac{\mathcal{S}_N}{N}\right| \geq \epsilon\right\} \leq \frac{1}{N\epsilon^2}.$$

Equivalently,

$$\underline{\mathbb{P}}\left\{\left|\frac{\mathcal{S}_N}{N}\right| < \epsilon\right\} \geq 1 - \frac{1}{N\epsilon^2}. \tag{6.9}$$

If we want $\mathcal{S}_N/N$ to be within $\epsilon$ of zero with lower probability $1-\delta$, then it suffices, according to (6.9), to make the number of tosses at least $1/(\epsilon^2\delta)$. Actually, a much smaller number of tosses will suffice. Bernoulli himself gave a much better upper bound, but it was still very conservative. For a sharp bound, we need De Moivre's theorem, which gives approximate probabilities for given deviations of $\mathcal{S}_N/N$ from zero.

## 6.2 DE MOIVRE'S THEOREM

We outlined our game-theoretic version of De Moivre's theorem for the fair coin in the introduction to this chapter. As we indicated there, it is convenient for this theorem to code Reality's move $x_n$ as $1/\sqrt{N}$ for heads and $-1/\sqrt{N}$ for tails. We then define a process $\mathcal{S}$ by $\mathcal{S}_0 := 0$ and $\mathcal{S}_n := \sum_{i=1}^n x_i$ for $n = 1, \ldots, N$. Thus $\mathcal{S}_n$ is the difference between the number of heads and the number of tails in the first $n$ rounds, divided by $\sqrt{N}$. De Moivre's theorem says that when $N$ is sufficiently large,

$\mathcal{S}_N$ is approximately standard Gaussian at the beginning of the game, inasmuch as $\int_{-\infty}^{\infty} U(z)\,\mathcal{N}_{0,1}(dz)$ is an approximate price for $U(\mathcal{S}_N)$ when $U$ is a well-behaved function.

Fixing the number $N$ of tosses and a positive initial capital $\mathcal{K}_0$ for Skeptic, we have the following protocol:

THE DE MOIVRE FAIR-COIN PROTOCOL
**Parameters:** $N, \mathcal{K}_0 > 0$
**Players:** Reality, Skeptic
**Protocol:**
FOR $n = 1, \ldots, N$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{-N^{-1/2}, N^{-1/2}\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.

We say that a function is *smooth* if it is infinitely differentiable.

**Proposition 6.2 (De Moivre's Theorem)** *Let $U$ be a smooth function constant outside a finite interval. Then for $N$ sufficiently large, the initial upper and lower prices of $U(\mathcal{S}_N)$ are both arbitrarily close to $\int_{-\infty}^{\infty} U(z)\,\mathcal{N}_{0,1}(dz)$.*

The upper and lower prices will not coincide exactly for any value of $N$, but they will both approach $\int_{-\infty}^{\infty} U(z)\,\mathcal{N}_{0,1}(dz)$ as $N$ grows, and so we may say that the game assigns $U(\mathcal{S}_N)$ this price in the limit.

Because many functions can be approximated by smooth functions constant outside a finite interval, De Moivre's theorem can easily be extended to larger classes of functions. We leave these details for the next chapter, because we want to avoid any complication that might obscure the simplicity of the proof. But we should note that the theorem does extend to the case where $U$ is the indicator function for an event $E$ that consists, say, of a finite union of intervals. Hence we can use De Moivre's theorem to compute upper and lower probabilities for $\mathcal{S}_N$ and thus for $y$, the number of heads in the $N$ tosses, using the relation $\mathcal{S}_N = (2y-N)/\sqrt{N}$. Table 6.1 gives a couple of examples. As indicated there, De Moivre's theorem not only says that $y/N$ is likely to be close to $1/2$; it also says that it is unlikely to be too close.

Abraham De Moivre (1667–1754), from a portrait dated 1736, when he was about 70. He published his theorem for coin tossing in 1738.

### The Heat Equation

Our proof of De Moivre's theorem uses the idea of smoothing a function $U$ by calculating the expected result of adding to its argument a Gaussian perturbation with mean zero and variance $D$. This transforms $U$ into the function $\overline{U}(\cdot, D)$, where

$$\overline{U}(s,D) := \int_{-\infty}^{\infty} U(s+z)\,\mathcal{N}_{0,D}(dz) = \int_{-\infty}^{\infty} U(z)\,\mathcal{N}_{s,D}(dz). \tag{6.10}$$

The function $\overline{U}(s,D)$ plays an important role in many branches of science, because, as we mentioned in the introduction to the chapter, it satisfies the heat equation,

$$\frac{\partial \overline{U}}{\partial D} = \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}, \tag{6.11}$$

for all $s \in \mathbb{R}$ and all $D > 0$. This can be verified informally using Taylor's approximation: for a small positive constant $dD$,

$$\begin{aligned}
\frac{\partial \overline{U}}{\partial D}(s,D)dD &\approx \overline{U}(s,D+dD) - \overline{U}(s,D) \\
&= \int_{-\infty}^{\infty} U(s+z)\,\mathcal{N}_{0,D+dD}(dz) - \overline{U}(s,D) \\
&= \int_{-\infty}^{\infty} \overline{U}(s+z,D)\,\mathcal{N}_{0,dD}(dz) - \overline{U}(s,D) \\
\approx \int_{-\infty}^{\infty} \left(\overline{U}(s,D) + \frac{\partial \overline{U}}{\partial s}(s,D)z + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(s,D)z^2\right) &\mathcal{N}_{0,dD}(dz) - \overline{U}(s,D) \\
&= \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(s,D)dD.
\end{aligned}$$

A rigorous proof requires some regularity conditions on $U$. For example, if $U$ is a Borel function such that

$$\lim_{s\to\infty} |U(s)|e^{-\epsilon s^2} = 0 \tag{6.12}$$

for any $\epsilon > 0$ (this holds if $U(s)$ grows at most polynomially fast), then a proof can be obtained by differentiating $\overline{U}(s,D)$, in the form

$$\frac{1}{\sqrt{2\pi D}} \int_{-\infty}^{\infty} U(z) e^{-(z-s)^2/(2D)} dz, \tag{6.13}$$

***Table 6.1*** Some high probabilities, from De Moivre's theorem, concerning the number $y$ of heads in $N$ trials.

| |
|---|
| $\overline{\mathbb{P}}\left\{\left\lvert\frac{y}{N} - \frac{1}{2}\right\rvert \geq \frac{0.02}{\sqrt{N}}\right\} = \overline{\mathbb{P}}\{\lvert S_N\rvert \geq 0.01\} = 0.992$ |
| $\overline{\mathbb{P}}\left\{\left\lvert\frac{y}{N} - \frac{1}{2}\right\rvert \leq \frac{6}{\sqrt{N}}\right\} = \overline{\mathbb{P}}\{\lvert S_N\rvert \leq 3\} = 0.997$ |

under the integral sign; this is authorized by Leibniz's differentiation rule for integrals [49, 38]. Other proofs are given in standard references on partial differential equations, for example, [49]. It is easy to see that (6.10) converges to $U(s)$ as $D \to 0$ and $S \to s$ for any real number $s$, provided that $U$ is continuous and (6.12) holds. This means that $\overline{U}$ solves the heat equation for the initial conditions given by $U$.

The heat equation was first successfully studied in the early nineteenth century by Joseph Fourier (1768–1830), who provided solutions in terms of trigonometric series for the case where initial conditions $U(s)$ are given on a finite interval in $[s_1, s_2]$, and boundary conditions $\overline{U}(s_1, D)$ and $\overline{U}(s_2, D)$ are also provided. Laplace, in 1809, was the first to provide the solution (6.10) for the case where $U(s)$ is given for all $s$ and hence no boundary conditions are needed [138].

Joseph B. Fourier (1768–1830), from an engraving made by Jules Boilly in 1873.

Fourier and other earlier writers were concerned with the propagation of heat. In this context, $D$ represents time, and the coefficient $1/2$ is replaced by an arbitrary constant. The value $\overline{U}(s, D)$ is the temperature at point $s$ at time $D$, and the equation says that the increase in the temperature at a point is proportional to how much warmer adjacent points are. Figure 6.1 shows the resulting propagation of heat for one particular initial distribution $\overline{U}(s, 0)$. Because the equation is also used to describe the diffusion of a substance in a fluid, it is often called the *diffusion equation*. Our application is no exception to the rule that the variable whose first derivative appears in the equation represents time; superficially, $D$ represents a variance, but this variance decreases proportionally with time.

## The Idea of the Proof

The proposition we are undertaking to prove, De Moivre's theorem, says that $\overline{U}(0, 1)$, where $\overline{U}$ is given by (6.10), is an approximate price at the beginning of the fair-coin game for the payoff $U(\mathcal{S}_N)$ at the end of the game. Our proof exhibits a strategy for Skeptic that begins with the capital $\overline{U}(0, 1)$ and ends up approximately with $U(\mathcal{S}_N)$. Indeed, the capital process for this strategy is approximately

$$\overline{U}(0,1),\ \overline{U}\left(\mathcal{S}_1, \frac{N-1}{N}\right), \dots, \overline{U}\left(\mathcal{S}_{N-1}, \frac{1}{N}\right), \overline{U}\left(\mathcal{S}_N, 0\right). \tag{6.14}$$

In order to find the strategy, we use a Taylor's series to study the increments in $\overline{U}$. If we write $\Delta\overline{U}$ for the increment on the $n$th round,

$$\Delta\overline{U} := \overline{U}\left(\mathcal{S}_n, \frac{N-n}{N}\right) - \overline{U}\left(\mathcal{S}_{n-1}, \frac{N-n+1}{N}\right),$$

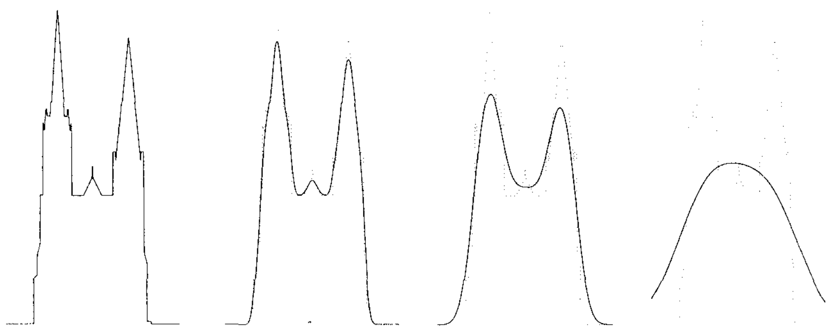



***Fig. 6.1*** Heat propagation according to the equation $\partial\overline{U}/\partial D = \frac{1}{2}\partial^2\overline{U}/\partial s^2$. Part **a** shows an initial temperature distribution $U(s)$, or $\overline{U}(s,0)$. Parts **b**, **c**, and **d** show $\overline{U}(s,D)$ as a function of $s$ at times $D = 1, 10, 100$, respectively.

then

$$\begin{aligned}\Delta\overline{U} \approx & \frac{\partial\overline{U}}{\partial s}\Delta s + \frac{\partial\overline{U}}{\partial D}\Delta D + \frac{1}{2}\frac{\partial^2\overline{U}}{\partial s^2}(\Delta s)^2 \\ & + \frac{\partial^2\overline{U}}{\partial D\partial s}\Delta s\Delta D + \frac{1}{2}\frac{\partial^2\overline{U}}{\partial D^2}(\Delta D)^2.\end{aligned} \tag{6.15}$$

Here $\Delta s = x_n$, while $\Delta D = -1/N$; the derivatives are evaluated at (or near) the point $(S_n, (N-n)/N)$. Thus $\Delta s$ is of order $N^{-1/2}$, while $(\Delta s)^2$ and $\Delta D$ are of order $N^{-1}$. We can neglect higher order terms, including the terms in $(\Delta D)^2$ and $\Delta s\Delta D$ shown in (6.15), but we cannot neglect $(\Delta s)^2$ and $\Delta D$. Although they are only of order $N^{-1}$, each moves in a consistent direction (the one up and other down), so that in $N$ steps they add to something of order 1. (The $\Delta s$, to the extent that Reality varies them, may be both positive and negative and hence may also add only to something of this order.) But according to the heat equation, $\Delta D$ and $(\Delta s)^2$ have the same coefficient. So (6.15) reduces to

$$\Delta\overline{U} \approx \frac{\partial\overline{U}}{\partial s}\Delta s + \frac{\partial\overline{U}}{\partial D}\left((\Delta s)^2 + \Delta D\right),$$

or

$$\Delta\overline{U} \approx c_1 x_n + c_2\left(x_n^2 - \frac{1}{N}\right). \tag{6.16}$$

Since $x_n$ is always equal to $N^{-1/2}$ or $-N^{-1/2}$, this further reduces to

$$\Delta\overline{U} \approx c_1 x_n. \tag{6.17}$$

So Skeptic's strategy is simple: on each round, he buys $c_1$ tickets.

The step from (6.16) to (6.17) relies on the fact that Reality is restricted to two moves in coin tossing; $x_n$ must be $N^{-1/2}$ or $-N^{-1/2}$. In more general protocols, such as bounded forecasting or the Lindeberg protocol that we will study in the next chapter, the second term in (6.16) can be dealt with only if Skeptic is allowed to buy $x_n^2$-tickets as well as $x_n$-tickets. In this case, the price of these tickets will take the place of $1/N$, and Skeptic's strategy will be to buy $c_1$ $x_n$-tickets and $c_2$ $x_n^2$-tickets.

This method of proof can be used not only to verify the price given in Proposition 6.2 but also to discover it. If there is a game-theoretic price for the payoff $U(\mathcal{S}_N)$, then it must begin a martingale that ends with $U(\mathcal{S}_N)$. A conjecture that the martingale's value at the $n$th step depends only on the value of $\mathcal{S}_n$ and $n$ would lead to the Taylor's series (6.15), from which it is clear that the desired increment can be obtained from tickets available to Skeptic only if the function expressing the current value of the martingale satisfies the heat equation; we would then solve the heat equation to obtain (6.10).

The method also works in the measure-theoretic framework. In this case, we understand *martingale* in the measure-theoretic sense, as a sequence of random variables in which each step is a zero-mean perturbation. Since the mean of $x_n$ will be zero and the mean of $x_n^2$ will be $1/N$ in the measure-theoretic case, (6.16) makes it clear that (6.14) is an approximate martingale in this sense as well.

This method of proof for the central limit theorem originated with Lindeberg, who used it in articles published in 1920 and 1922. Lindeberg's argument was taken up by Paul Lévy, in his *Calcul des Probabilités* in 1925 (§51), and it has been subsequently repeated in a number of textbooks, including Breiman 1968 (pp. 167–170) and Billingsley 1968 (pp. 42–44). Neither Lindeberg nor these subsequent authors made explicit reference to the heat equation, which tends to disappear from view when one develops a rigorous treatment for the error terms in the approximation (6.15). The underlying gambling ideas are also usually left implicit.

## Details of the Proof

Making our proof rigorous is mainly a matter of dealing carefully with the remainder terms in the Taylor expansion.

As a first step, notice that it suffices to show that for any positive number $\epsilon$,

$$\overline{\mathbb{E}}\,U(\mathcal{S}_N) \leq \int U(z)\,\mathcal{N}_{0,1}(dz) + \epsilon \tag{6.18}$$

when $N$ is sufficiently large. For when we substitute $-U$ for $U$ in (6.18), we obtain

$$\int U(z)\,\mathcal{N}_{0,1}(dz) - \epsilon \leq \underline{\mathbb{E}}\,U(\mathcal{S}_N), \tag{6.19}$$

and (6.18) and (6.19), together with the fact that the lower price cannot exceed the upper price, give the result we want.

In order to establish (6.18), we need to construct a martingale $\mathcal{V}$ that starts at $\int U(z)\,\mathcal{N}_{0,1}(dz)$ and ends up with $\mathcal{V}_N$ satisfying $U(\mathcal{S}_N) \leq \mathcal{V}_N + \epsilon$. In other words, we need a strategy $\mathcal{P}$ whose capital process $\mathcal{K}^{\mathcal{P}}$ satisfies $U(\mathcal{S}_N) \leq \int U(z)\,\mathcal{N}_{0,1}(dz) + \mathcal{K}_N^{\mathcal{P}} + \epsilon$. The strategy is simple. On round $n+1$, Skeptic knows $x_1, \ldots, x_n$. So he buys $(\partial\overline{U}/\partial s)(S_n, D_n)$ tickets, where

$$S_n := \mathcal{S}_n = \sum_{i=1}^{n} x_i, \; D_n := 1 - \frac{n}{N},$$

and $\overline{U}(s, D)$ is defined, for $s \in \mathbb{R}$ and $D \geq 0$, by (6.10).

We have[1]

$$\begin{aligned} d\overline{U}(S_n, D_n) &= \frac{\partial \overline{U}}{\partial s}(S_n, D_n) dS_n + \frac{\partial \overline{U}}{\partial D}(S_n, D_n) dD_n \\ &+ \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(S_n', D_n')(dS_n)^2 + \frac{\partial^2 \overline{U}}{\partial s \partial D}(S_n', D_n') dS_n dD_n \\ &+ \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial D^2}(S_n', D_n')(dD_n)^2, \end{aligned} \tag{6.20}$$

for $n = 0, \ldots, N-1$, where $(S_n', D_n')$ is a point strictly between $(S_n, D_n)$ and $(S_{n+1}, D_{n+1})$. Applying Taylor's formula to $\partial^2\overline{U}/\partial s^2$, we find

$$\frac{\partial^2 \overline{U}}{\partial s^2}(S_n', D_n') = \frac{\partial^2 \overline{U}}{\partial s^2}(S_n, D_n) + \frac{\partial^3 \overline{U}}{\partial s^3}(S_n'', D_n'') dS_n' + \frac{\partial^3 \overline{U}}{\partial D \partial s^2}(S_n'', D_n'') dD_n',$$

where $(S_n'', D_n'')$ is a point strictly between $(S_n, D_n)$ and $(S_n', D_n')$, and $dS_n'$ and $dD_n'$ satisfy $|dS_n'| \leq |dS_n|$, $|dD_n'| \leq |dD_n|$. Plugging this equation and the heat equation, (6.11), into (6.20), we obtain

$$\begin{aligned} d\overline{U}(S_n, D_n) &= \frac{\partial \overline{U}}{\partial s}(S_n, D_n) dS_n + \frac{\partial \overline{U}}{\partial D}(S_n, D_n)\left((dS_n)^2 + dD_n\right) \\ &+ \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial s^3}(S_n'', D_n'') dS_n' (dS_n)^2 + \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial D \partial s^2}(S_n'', D_n'') dD_n' (dS_n)^2 \\ &+ \frac{\partial^2 \overline{U}}{\partial s \partial D}(S_n', D_n') dS_n dD_n + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial D^2}(S_n', D_n')(dD_n)^2. \end{aligned} \tag{6.21}$$

The first term on the right-hand side is the increment of our martingale $\mathcal{V}$: the gain from buying $\frac{\partial \overline{U}}{\partial s}(S_n, D_n)$ tickets on round $n+1$. So we only need to show that the other terms are negligible when $N$ is sufficiently large.

The second term is identically zero in the current context (where $(dS_n)^2 = x_{n+1}^2 = 1/N = -dD_n$), and so we are concerned only with the last four terms. All the partial derivatives involved in those terms are bounded: the heat equation implies

$$\begin{aligned} \frac{\partial^3 \overline{U}}{\partial D \partial s^2} = \frac{\partial^3 \overline{U}}{\partial s^2 \partial D} = \frac{1}{2}\frac{\partial^4 \overline{U}}{\partial s^4}, \quad \frac{\partial^2 \overline{U}}{\partial s \partial D} = \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial s^3}, \\ \frac{\partial^2 \overline{U}}{\partial D^2} = \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial D \partial s^2} = \frac{1}{4}\frac{\partial^4 \overline{U}}{\partial s^4}, \end{aligned} \tag{6.22}$$

and $\partial^3\overline{U}/\partial s^3$ and $\partial^4\overline{U}/\partial s^4$, being averages of $U^{(3)}$ and $U^{(4)}$, are bounded. So the four terms will have at most the order of magnitude $O(N^{-3/2})$, and their total cumulative contribution be at most $O(N^{-1/2})$.

[1]For any sequence $A_n$, $\Delta A_n := A_n - A_{n-1}$ and $dA_n := A_{n+1} - A_n$. We use both in this book, switching from one to the other depending on which makes for the simpler expressions in the particular context.

All the central limit theorems and the option pricing formulas studied in this book are based on variants of Equation (6.21).

## 6.3 A ONE-SIDED CENTRAL LIMIT THEOREM

We now turn to the case where Reality has more than two alternatives on each round of play. For simplicity, we assume that Reality may freely choose her move $x_n$ from the entire closed interval $[-N^{-1/2}, N^{-1/2}]$. This degree of freedom is not essential, however. All the results of this section continue to hold if, for example, we assume instead that she chooses from the three-element set $\{-N^{-1/2}, 0, N^{-1/2}\}$.

FINITE-HORIZON BOUNDED FORECASTING PROTOCOL
**Parameters:** $N, \mathcal{K}_0 > 0$
**Players:** Reality, Skeptic
**Protocol:**
FOR $n = 1, \ldots, N$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in [-N^{-1/2}, N^{-1/2}]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.

As we showed in §6.1, Bernoulli's theorem generalizes readily to this protocol. De Moivre's theorem does not. This is because De Moivre's theorem is two-sided: as we noted in Table 6.1, it tells us not only that Skeptic can force Reality to bring her average move $\sum_{n=1}^{N} x_n$ close to zero but also that he can force her not to bring it too close. In this protocol (and also in the protocol where Reality chooses from the three-element set $\{-N^{-1/2}, 0, N^{-1/2}\}$), Skeptic cannot force Reality to stay away from zero. She can set all her $x_n$ exactly equal to zero with no penalty.

The best we can do is generalize De Moivre's theorem to a one-sided central limit theorem. We set $\mathcal{S}_0 := 0$ and $\mathcal{S}_n := \sum_{i=1}^{n} x_i$ for $n = 1, \ldots, N$, as in the preceding section. Using parabolic potential theory, we will define a function $\overline{U}(s, D)$, $s \in \mathbb{R}$ and $D > 0$, that approaches $U(s)$ as $D \to 0$, satisfies

$$\overline{U}(s, D) \geq U(s)$$

for all $s$ and $D$, and satisfies the heat equation,

$$\frac{\partial \overline{U}}{\partial D} = \frac{1}{2} \frac{\partial^2 \overline{U}}{\partial s^2},$$

for all $s$ and $D$ such that $\overline{U}(s, D) > U(s)$. And then we will prove the following one-sided theorem:

**One-Sided Central Limit Theorem (Informal Statement)** *If the payoff function $U$ satisfies appropriate regularity conditions, then* $\lim_{n\to\infty} \overline{\mathbb{E}}\, U(\mathcal{S}_N) = \overline{U}(0, 1)$.

The regularity conditions we use are given in Theorem 6.1 in the next subsection.

The value of $\overline{U}(0,1)$ can be found numerically (just as we find the value of the Gaussian integral numerically in the case of De Moivre's theorem), and because $\underline{\mathbb{E}}\, U(\mathcal{S}_N) = -\overline{\mathbb{E}}[-U(\mathcal{S}_N)]$, this allows us to find both upper and lower prices for $U(\mathcal{S}_N)$. As Table 6.2 illustrates, the upper and lower prices may be quite different.

Joseph L. Doob, receiving the National Medal of Science from President Carter in 1979.

We can gain insight into the behavior of the function $\overline{U}$ from its interpretation in terms of heat propagation. In this interpretation, we start at time $D = 0$ with a distribution of heat over the real line described by the function $U(s)$, and at every point $s$, there is a thermostat that switches on whenever this is necessary to keep the temperature from falling below $U(s)$. Whenever the temperature at $s$ rises above $U(s)$, this local thermostat switches off and the evolution of the temperature is then governed by the heat equation. The function $\overline{U}(s, D)$, which is defined for all $s \in \mathbb{R}$ and all $D \geq 0$, gives the resulting temperature at $s$ at time $D$. Figure 6.2 depicts the temperature distribution $s \mapsto \overline{U}(s, D)$ for several values of $D$, starting with the same initial temperature distribution as in Figure 6.1 (p. 130), which depicts heat propagation with no heat sources or sinks.

For each $D \geq 0$, set $U_D(s) := \overline{U}(s, D)$. It is clear that if $U_{D_1} \geq U_{D_2}$, then $U_{D_1+\delta} \geq U_{D_2+\delta}$ for $\delta > 0$; given the physics of heat propagation, and given that the behavior of the local thermostats is fixed, more heat everywhere at the beginning of a time period of length $\delta$ can only produce more heat everywhere at the end of the time period. We must certainly have $\overline{U}_\delta \geq \overline{U}_0$, because the thermostats will not allow any decrease from the initial distribution $\overline{U}_0$. So by induction, the temperature is always increasing in $D$:

$$\frac{\partial \overline{U}}{\partial D}(s, D) \geq 0 \tag{6.23}$$

***Table 6.2*** Some upper and lower probabilities from the one-sided central limit for the sum $\mathcal{S}_N$ in the finite-horizon bounded forecasting protocol. This table should be compared with Table 6.1. For any $\alpha > 0$, $\underline{\mathbb{P}}\{|\mathcal{S}_N| \geq \alpha\} = 0$, and $\overline{\mathbb{P}}\{|\mathcal{S}_N| \leq \alpha\} = 1$. But $\underline{\mathbb{P}}\{|\mathcal{S}_N| \leq \alpha\}$ and $\overline{\mathbb{P}}\{|\mathcal{S}_N| \geq \alpha\}$ must be found numerically.

| **Upper probabilities** | **Lower probabilities** |
|---|---|
| $\overline{\mathbb{P}}\{|\mathcal{S}_N| \geq 0.01\} = 1.000$ | $\underline{\mathbb{P}}\{|\mathcal{S}_N| \geq 0.01\} = 0$ |
| $\overline{\mathbb{P}}\{|\mathcal{S}_N| \leq 3\} = 1$ | $\underline{\mathbb{P}}\{|\mathcal{S}_N| \leq 3\} = 0.995$ |

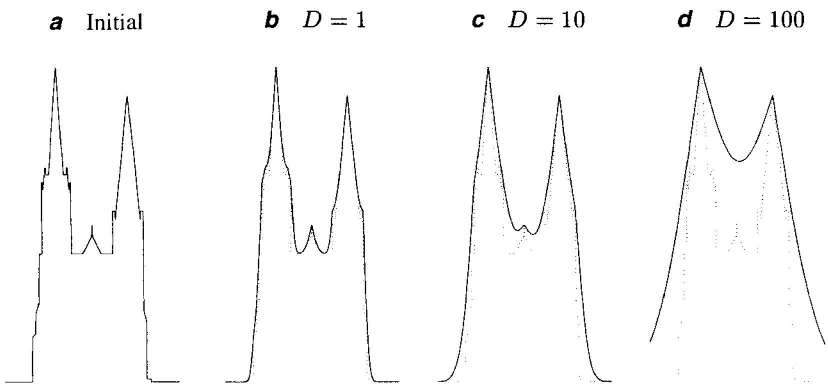



***Fig. 6.2*** Heat propagation with thermostats set to keep the temperature at $s$ from falling below $U(s)$. Part **a** shows the same initial temperature distribution as in Figure 6.1. Parts **b**, **c**, and **d** show $\overline{U}(s, D)$ as a function of $s$ at times $D = 1, 10, 100$, respectively.

for all $s \in \mathbb{R}$ and all $D \geq 0$. For each $D$, $U_D$ divides the infinite rod $-\infty < s < \infty$ into two parts: the part where the thermostats are active (the *exercise region*) and the part where they are not (the *continuation region*) (cf. the discussion of exercise and continuation regions for American options in [107], 8E and 8F). In the continuation region $U_D$ is convex (by (6.23) and the heat equation (6.11)); in the exercise region $U_D$ coincides with $U$. The heat propagation, together with the influx of heat from the local thermostats, gradually increases the temperature everywhere, usually by making it increasingly convex.

Another very valuable source of intuition for understanding the function $\overline{U}$, described in detail by Doob (1984), is its interpretation in terms of Brownian motion. In order to avoid the confusion that might arise from working with too many metaphors at once, we leave this interpretation for an appendix, §6.5.

Our one-sided central limit theorem says that $\lim_{n\to\infty} \overline{\mathbb{E}}\, U(\mathcal{S}_N) = \overline{U}(0,1)$. Before setting out the parabolic potential theory needed for a precise definition of $\overline{U}(0,1)$ and a proof of the theorem, we can give a heuristic proof of part of it, the inequality

$$\limsup_{N\to\infty} \overline{\mathbb{E}}\, U(\mathcal{S}_N) \leq \overline{U}(0,1).$$

This inequality says that starting with $\overline{U}(0,1)$ at the beginning of the game, Skeptic can obtain at least $U(\mathcal{S}_N)$, approximately, at the end. Reasoning as in our heuristic proof of De Moivre's theorem (see (6.15)–(6.17)), but with our current definition of $\overline{U}$, we obtain, instead of (6.16), that $\Delta\overline{U}$ does not exceed, approximately,

$$\frac{\partial \overline{U}}{\partial s} x_n + \frac{\partial \overline{U}}{\partial D}\left(x_n^2 - \frac{1}{N}\right) \leq \frac{\partial \overline{U}}{\partial s} x_n;$$

the inequality follows from (6.23) and the restriction of $x_n$ to $\left[-N^{-1/2}, N^{-1/2}\right]$. By buying $\partial\overline{U}/\partial s$ tickets, Skeptic can ensure that his capital at the end of round $n$ is close to or exceeds $\overline{U}(\mathcal{S}_n, 1 - n/N)$. When $n = N$ this will be close to or more than $U(\mathcal{S}_N)$, as required.

## Rigorous Treatment

Potential theory dates from the eighteenth century [70]. It is mainly concerned with elliptic differential equations, but in recent decades it has been extended to parabolic differential equations as well [55]. The first systematic exposition of parabolic potential theory was provided by Doob 1984 [103], who demonstrated remarkable parallels between classical potential theory, parabolic potential theory, and parts of probability theory. We use only a small part of Doob's theory. For a more elementary treatment of some of the key notions we use, see [49], Chapter 16.

This section also requires some elementary notions from general topology that we do not use elsewhere. We write $\overline{A}$ for the closure of a set $A$ and $\partial A$ for its boundary.

We begin with the notion of a *parabolic measure*. Doob ([103], §1.XV.10) shows how parabolic measures can be defined for a wide class of domains, but we will be mainly concerned with rectangles of the form

$$B(s, D, \delta) := (s - \delta, s + \delta) \times (D - \delta^2, D), \tag{6.24}$$

where $(s, D) \in \mathbb{R}^2$ and $\delta > 0$ (see Figure 6.3). The height of the rectangle is the square of its width; intuitively, this is because space scales as the square root of time in heat propagation, as revealed by the exponent $-(z - s)^2/(2D)$ in (6.13). We divide the boundary of $B(s, D, \delta)$ into four parts:

- the *lower boundary* $(s - \delta, s + \delta) \times \{D - \delta^2\}$,
- the *upper boundary* $(s - \delta, s + \delta) \times \{D\}$,
- the *lateral boundary* $\{s - \delta, s + \delta\} \times (D - \delta^2, D)$,
- and the *corners* $\{s - \delta, s + \delta\} \times \{D - \delta^2, D\}$.

There is a parabolic measure for every point in the interior of the rectangle, and all these measures are concentrated on the lower and lateral boundaries of the rectangle. The parabolic measure for $(s', D')$ is denoted by $\mu_{B(s,D,\delta)}(s', D')$, and it is actually concentrated on the lower boundary and the parts of the lateral boundary below $D'$. The purpose of this measure is to tell what temperature at $s'$ will result at time $D'$ for any given initial conditions at time $D - \delta^2$ and boundary conditions at $s \pm \delta$ (assuming no heat sources or sinks). If the initial and boundary conditions are expressed by a function $u$ on the lower and lateral boundaries of the rectangle (the temperature distribution at time $D - \delta^2$ on the lower boundary, and the boundary conditions on the lateral boundaries), then the temperature at $s'$ at time $D'$ is the expected value of $u$ with respect to the parabolic measure:

$$\int u \, d\mu_{B(s,D,\delta)}(s', D').$$

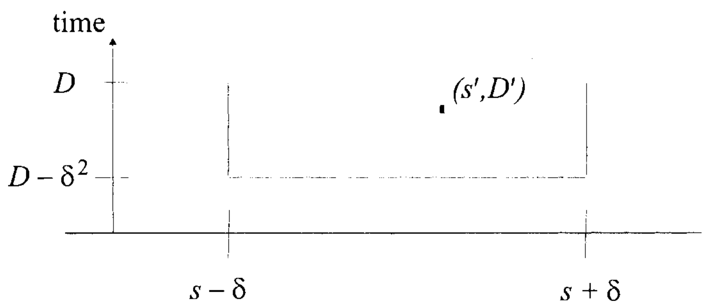



***Fig. 6.3*** A rectangle in space and time. For each point $(s', D')$ inside the rectangle (or on the upper boundary), there is a probability measure, the *parabolic measure*, which is concentrated on the lower and lateral boundaries and gives $s'$'s temperature at time $D'$ as the expected value of a function $u$ that is defined on the lower and lateral boundaries. The values of $u$ on the lower boundary are initial conditions (the temperature distribution at time $D - \delta^2$), and the values of $u$ on the lateral boundary are boundary conditions (temperatures that are imposed at the points $s - \delta$ and $s + \delta$).

The corners have too low a dimension to influence the temperature at $(s', D')$, and the upper boundary is situated later in time.

It is easy to extend the definition of the parabolic measure $\mu_{B(s,D,\delta)}(s', D')$ to the case where $(s', D')$ lies on the upper boundary of $B(s, D, \delta)$. For example, we can define it as $\mu_A(s', D')$, where $A$ is a rectangle obtained from $B(s, D, \delta)$ by raising its upper boundary. We are mainly interested in the case $(s', D') = (s, D)$, and we abbreviate $\mu_{B(s,D,\delta)}(s, D)$ to $\mu_{B(s,D,\delta)}$.

A function $u(s, D)$, $u : \mathbb{R} \times (0, \infty) \to (-\infty, \infty]$, is called *superparabolic* if

- for any $(s, D)$ and $\delta > 0$ such that $\delta < \sqrt{D}$,

$$u(s, D) \geq \int u d\mu_{B(s,D,\delta)}$$

  (in particular, the integral is required to exist),

- it is lower semicontinuous (this means that the set $\{u > c\}$ is open for any $c \in \mathbb{R}$), and

- it is finite on a dense subset of $\mathbb{R} \times (0, \infty)$.

This definition is analogous to the classical definition of a superharmonic function. The first condition says that only heat sources are allowed, no heat sinks: the temperature $u(s, D)$ should be $\int u d\mu_{B(s,D,\delta)}$, the temperature at $s$ that would result at time $D$ from unfettered propagation from nearby points, or else higher, because of heat sources. The second condition says that if a point $(s, D)$ is warm (perhaps because it was heated by some heat source), the points next to it should also be warm (for the same reason). The third condition is not really relevant for us: we consider only finite (in this chapter, bounded) superparabolic functions.

A function $u$ is *subparabolic* if $-u$ is superparabolic (only heat sinks are allowed). A function defined on some open domain is *parabolic* if it (1) is continuously differentiable, (2) is twice continuously differentiable with respect to the space variable $s$, and (3) satisfies the standard heat equation. This is equivalent to the function being both superparabolic and subparabolic ([103], §1.XV.12).

The least superparabolic majorant of a function $u : \mathbb{R} \times (0, \infty) \to (-\infty, \infty]$, if it exists, is denoted by $\mathrm{LM}\, u$. Now we can state our one-sided central limit theorem:

**Theorem 6.1** *Suppose the payoff function $U$ is continuous and constant outside a finite interval. Then* $\mathrm{LM}\, u$, *where* $u(s, D) := U(s)$, *exists and*

$$\lim_{N\to\infty} \overline{\mathbb{E}}\, U(\mathcal{S}_N) = (\mathrm{LM}\, u)(0, 1).$$

For the proof, we need a few more concepts from parabolic potential theory and general topology:

- If $u$ is an integrable[2] function on the lower and lateral boundary of $B = B(s, D, \delta)$, the function $\mathrm{PI}(B, u)$ on $B$ given by

$$\mathrm{PI}(B, u)(s', D') := \int u\, d\mu_B(s', D')$$

is well defined and is called the *parabolic Poisson integral* of $u$. The formula extends to the case where $(s', D')$ is on the upper boundary of $B$. (See [103], §1.XV.12.)

- If $u$ is a Borel function on $\mathbb{R} \times (0, \infty)$, the closure of $B = B(s, D, \delta)$ lies in $\mathbb{R} \times (0, \infty)$, and the restriction of $u$ to the lower and lateral boundaries of $B$ is integrable, define $\tau_B u$ by (1) $\tau_B u := \mathrm{PI}(B, u)$ on $B$ and its upper boundary (this ensures the lower semicontinuity of $\tau_B u$ when $u$ is superparabolic), and (2) $\tau_B u := u$ on the rest of $\mathbb{R} \times (0, \infty)$. If a function $u$ defined on $\mathbb{R} \times (0, \infty)$ is superparabolic and the closure of $B = B(s, D, \delta)$ lies in $\mathbb{R} \times (0, \infty)$, then $\tau_B u \le u$, $\tau_B u$ is superparabolic, and $\tau_B u$ is parabolic in $B$. (See [103], §1.XV.14.)

- If $u$ is a function defined on $\mathbb{R} \times (0, \infty)$, we let $\tilde{u}$ stand for the greatest lower semicontinuous minorant (also called the *lower semicontinuous smoothing*)

$$\tilde{u}(\xi) := \min\left(u(\xi), \liminf_{\eta\to\xi} u(\eta)\right), \quad \xi \in \mathbb{R} \times (0, \infty),$$

of $u$.

[2]By "integrable" we always mean integrable with respect to the Lebesgue measure.

We also need the following results about superparabolic functions ([103], §1.XV.12 and §1.XV.14):

**Smooth superparabolic functions.** If $u : \mathbb{R} \times (0, \infty) \to (-\infty, \infty)$ is infinitely differentiable, then it is superparabolic if and only if

$$\frac{\partial \overline{U}}{\partial D} \geq \frac{1}{2} \frac{\partial^2 \overline{U}}{\partial s^2} \tag{6.25}$$

holds.

**Approximation theorem.** Every superparabolic function $u$ is the limit, on each compact subset of $\mathbb{R} \times (0, \infty)$, of an increasing sequence $u_n$ of infinitely differentiable superparabolic functions. The proof of this theorem is constructive: we take $A_{1/n}u$ as $u_n$, where the averaging operator $A_\delta$, $\delta > 0$, is defined by

$$A_\delta u(s, D) := \int_0^1 \phi(r) \int u d\mu_{B(s,D,r\delta)} dr$$

and $\phi(r)$ is a weight function, $\int_0^1 \phi(r) dr = 1$. Here the averaging over $B(s, D, \delta)$ is combined with further averaging over $\delta$. The function $A_\delta u$ will be infinitely differentiable for an appropriate choice of $\phi$. Notice that $A_\delta u(s, D)$ will only be defined for $D > \delta^2$. Because the smoothing operators $A_\delta$ do not change parabolic functions at points not too close to the boundary, the construction shows that parabolic functions are infinitely differentiable.

**Convergence theorem.** (This is part of what Doob [103] calls the parabolic fundamental convergence theorem.) Set $u := \inf_{\alpha \in I} u_\alpha$, where $\{u_\alpha \mid \alpha \in I\}$ is a family of superparabolic functions locally uniformly bounded below. Then the lower semicontinuous smoothing $\tilde{u}$ is superparabolic. (This theorem will allow us to establish the existence of LM $f$ in the case that interests us.)

The last result we need to prove Theorem 6.1 is the continuity of LM $u$ under the conditions of the theorem.

**Lemma 6.2** *Under the conditions of Theorem 6.1 (or, more generally, if $U$ is uniformly continuous and bounded),* $(\text{LM}\, u)(s, D)$ *exists and is increasing in $D$ and continuous.*

*Proof* Consider the infimum $f$ of the set of all superparabolic majorants of $u$ (this set is nonempty since $U$ is bounded). Since $f \geq u$ and $u$ is lower semicontinuous, $\tilde{f} \geq u$. According to the convergence theorem, $\tilde{f}$ is superparabolic and, therefore, is the least superparabolic majorant LM $u$ of $u$; in particular, $\tilde{f} = f$.

The fact that LM $u$ is nondecreasing in $D$ is intuitively clear (cf. (6.23)) and easy to verify formally: indeed, to see that $(\text{LM}\, u)(s, D_1) \leq (\text{LM}\, u)(s, D_2)$ when $D_1 < D_2$, notice that $(\text{LM}\, u)(s, D_1) \leq (\text{LM}\, \tilde{u})(s, D_2) \leq (\text{LM}\, u)(s, D_2)$, where $\tilde{u}$ is defined as $u$ in the region $D > D_2 - D_1$ and as $\inf u$ in $D \leq D_2 - D_1$. (A cleaner argument not requiring boundedness but involving parabolic potential theory for domains different from $\mathbb{R} \times (0, \infty)$

is: $(\mathrm{LM}\, u)(s, D_1) = (\mathrm{LM}\, \tilde{u})(s, D_2) \le (\mathrm{LM}\, u)(s, D_2)$, where $\tilde{u}$ is the restriction of $u$ to $\mathbb{R} \times (D_2 - D_1, \infty)$.)

Fix some *modulus of continuity* for $U$, that is, a function $f : (0, \infty) \to (0, \infty)$ such that $\lim_{\delta \downarrow 0} f(\delta) = 0$ and, for any $\delta$, $s_1$, and $s_2$,

$$|s_1 - s_2| < \delta \Longrightarrow |U(s_1) - U(s_2)| \le f(\delta).$$

It is easy to see that, for every $D > 0$, the function $(\mathrm{LM}\, u)(s, D)$ is uniformly continuous in $s$ with the same modulus of continuity $f$: indeed, if $|s_1 - s_2| < \delta$,

$$(\mathrm{LM}\, u)(s_1, D) = (\mathrm{LM}\, \tilde{u})(s_2, D) \le (\mathrm{LM}\, u)(s_2, D) + f(\delta),$$

where $\tilde{u}(s) := u(s + s_1 - s_2)$. (A similar "shifting" argument, but along the time rather space axis, was used in the above proof that $\mathrm{LM}\, u$ increases in $D$.)

It is now easy to see that the continuity of $\mathrm{LM}\, u$ will be implied by its continuity in the variable $D$. Since $(\mathrm{LM}\, u)(s, D)$ is continuous in $D$ from below (being lower semicontinuous and increasing in $D$), we only need to prove

$$\lim_{\delta \downarrow 0} (\mathrm{LM}\, u)(s, D + \delta) = (\mathrm{LM}\, u)(s, D) \tag{6.26}$$

for every point $(s, D)$ in the domain of $\mathrm{LM}\, u$. Assuming (6.26) is violated for a point $(s, D)$, that is,

$$G := \lim_{\delta \downarrow 0} (\mathrm{LM}\, u)(s, D + \delta) > (\mathrm{LM}\, u)(s, D),$$

let us choose positive numbers $\epsilon$ and $\delta$ such that

$$G > (\mathrm{LM}\, u)(s, D) + 2\epsilon, \; f(\delta) < \epsilon.$$

Setting $A := (\mathrm{LM}\, u)(s, D) + \epsilon$, we obtain

$$(\mathrm{LM}\, u)(s, D') > A + \epsilon > A > (\mathrm{LM}\, u)(s', D)$$

for all $s' \in (s - \delta, s + \delta)$ and $D' > D$. We can see that on the lower boundary of $B(s, D + \delta^2, \delta)$ the temperature is less than $A$, on the lateral boundary it is bounded by $\sup U$, and strictly above $(s, D)$ it jumps to at least $A + \epsilon$. This contradicts the continuity of heat propagation. ∎

Now we are ready to prove Theorem 6.1. We will deal separately with the positive part,

$$\limsup_{N \to \infty} \overline{\mathbb{E}}\, U(\mathcal{S}_N) \le (\mathrm{LM}\, u)(0, 1),$$

and the negative part,

$$\liminf_{N \to \infty} \overline{\mathbb{E}}\, U(\mathcal{S}_N) \ge (\mathrm{LM}\, u)(0, 1). \tag{6.27}$$

*Proof of the positive part of Theorem 6.1* Let $f := \mathrm{LM}\, u$; it exists by Lemma 6.2.

Remember that $D_n := 1 - n/N$. Fix a large constant $C > 0$ and a small constant $\delta \in (0, 1)$; how large and how small will become clear later. First we show that Skeptic, starting with $f(0, 1)$, can attain capital close to $f(\mathcal{S}_n, \delta)$ (or more) when $(\mathcal{S}_n, D_n)$, $n = 0, \ldots, N$, hits the set

$$\{(s, D) \mid |s| \ge C \text{ or } D \le \delta\} \tag{6.28}$$

(after that Skeptic will stop playing or will play just to prevent big changes in $\mathcal{S}_n$: see the next two paragraphs). According to the approximation theorem, there is a smooth superparabolic function $\overline{U} \leq f$ which is arbitrarily close to $f$ on $[-C, C] \times [\delta, 1]$. The construction in the proof of the approximation theorem (see above) shows that $\overline{U}$ is, like $f$, nondecreasing in $D$. As in the proof of the central limit theorem, using (6.25), we will obtain (6.21) with $=$ replaced by $\leq$ (where $S_n = \mathcal{S}_n$ and $D_n = 1 - n/N$). Since the second addend in (6.21) is nonpositive (it is the product of a nonnegative and a nonpositive factor) and all addends after the second add, over $n = 0, \ldots, N-1$, to $O(N^{-1/2})$, Skeptic will be able to super-replicate, by buying $\partial\overline{U}/\partial s$ tickets, $\overline{U}(S_n, D_n)$ with arbitrary accuracy before the set (6.28) is hit.

Let us first assume that only the border $D = \delta$ of (6.28) is hit. We have proved that Skeptic can attain capital close to $f(\mathcal{S}_n, \delta)$ starting with $f(0, 1)$. Since $f(s, D) \geq U(s)$ for all $s$ and $D$, it remains to prove that Skeptic can ensure that $U(\mathcal{S}_n)$ will not move significantly after $D_n$ reaches $\delta$ unless his capital is multiplied substantially; since $U$ is bounded, this will allow Skeptic to hedge against moves of $\mathcal{S}_n$ exceeding some small constant $\epsilon$ with a small increase in the required initial capital. Since $U$ is continuous and constant outside a finite interval (and, therefore, uniformly continuous), it is sufficient for Skeptic to prevent big changes in $\mathcal{S}_n$. Below we prove a version of the (one-sided) weak law of large numbers which is sufficient for this purpose[3]: setting $a := N^{-1/2}$ and $\tilde{N} := \lceil \delta N \rceil$, we can see that Skeptic has a winning strategy in the game of Lemma 6.3 when $\delta$ is small enough (as compared to $\epsilon^2$).

Now let us see what happens if the border $|s| = C$ of (6.28) is (or both $|s| = C$ and $D = \delta$ are) hit. According to Lemma 6.3 (with $a := N^{-1/2}$, $\tilde{N} := N$, and $\epsilon = C$), the probability of this event is small when $C$ is large, and Skeptic will be able to simulate a low-cost insurance policy with payoff exceeding $\sup U - \inf U$ when $|s| = C$ is hit (after it is hit, Skeptic can stop any trading). ∎

The following lemma is a variation on the result that Skeptic has a winning strategy in the finite-horizon bounded forecasting game with forecasts set at 0 (this is the game we obtain when we generalize the fair-coin game by replacing $\{-1, 1\}$ with $[-1, 1]$; see p. 125). We recast the game as follows:

**Parameters:** $\mathcal{K}_0 > 0, \tilde{N}, a > 0, K > 0$
**Players:** Reality, Skeptic
**Protocol:**
FOR $n = 1, \ldots, \tilde{N}$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in [-a, a]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n$.
**Winning:** Skeptic wins if $\mathcal{K}_n$ is never negative and
either $\mathcal{K}_{\tilde{N}} \geq K\mathcal{K}_0$ or $\left|\sum_{n=1}^{\tilde{N}} x_n\right| < \epsilon$.

[3]Lemma 6.2, which asserts that LM $u$ is continuous, provides a different means of coping with irregularities arising when $D_n$ approaches 0: we could use the trick (as we do in the proof of Theorem 13.1 on p. 331) of starting with $(\mathrm{LM}\, u)(0, 1 + \delta)$ (for a small $\delta > 0$) rather than with $(\mathrm{LM}\, u)(0, 1)$, instead of using a one-sided weak law of large numbers. However, we would still need a one-sided weak law of large numbers to hedge the possibility of hitting the lines $|s| = C$.

**Lemma 6.3** *Skeptic has a winning strategy in this game if* $Ka^2\tilde{N} < \epsilon^2$.

This lemma can be obtained as a simple corollary from Proposition 6.1; alternatively, consider the nonnegative supermartingale $\mathcal{L}_n := \mathcal{K}_0(\mathcal{S}_n^2 + a^2\tilde{N} - a^2 n)/(a^2\tilde{N})$.

*Proof of the negative part of Theorem 6.1* Here we prove (6.27). Since LM $u$ is lower semicontinuous, the set $\mathcal{D}$ where it is different from the continuous $u$ is open. We already know that $(s, D)$ is in $\mathcal{D}$ as soon as $(s, D')$ is in $\mathcal{D}$ for some $D' < D$. A standard trick shows that LM $u$ is parabolic inside $\mathcal{D}$: indeed, replacing LM $u$ inside a small interval $B \subseteq \mathcal{D}$ of the form (6.24) by the Poisson integral PI$(B, \text{LM } u)$ (more accurately, applying the $\tau_B$ operation), we obtain a function which is simultaneously a superparabolic majorant of $u$ and a minorant of LM $u$; since LM $u$ is the least superparabolic majorant, it coincides with PI$(B, \text{LM } u)$ inside $B$ and so is parabolic inside $B$. Varying $B$ we obtain that LM $u$ is parabolic in $\mathcal{D}$.

Suppose that, contrary to what we are proving, Skeptic can ensure, for a large enough $N$, the final capital of at least $U(\mathcal{S}_N)$ starting with $(\text{LM } u)(0, 1) - 2\epsilon$, for some constant $\epsilon > 0$. Let $\mathcal{D}^*$ be the following open subset of $\mathcal{D}$:

$$\mathcal{D}^* := \{(s, D) \mid \text{LM } u - u > \epsilon\} .$$

Notice that LM $u$, as a parabolic function in $\mathcal{D}$, is infinitely differentiable inside $\mathcal{D}$; therefore, all its partial derivatives are bounded in any compact subset of $\overline{\mathcal{D}^*} \subseteq \mathcal{D}$ (this inclusion follows from the continuity of LM $u$: see Lemma 6.2). Without loss of generality we assume that $U$ is zero outside a finite interval. Let $C$ be a large constant; in particular: $U(s) = 0$ when $|s| \geq C$; $(\text{LM } u)(s, D)$ is uniformly negligible for $D \in (0, 1)$ and $|s| \geq C$. (The latter property for large $C$ can be seen from, e.g., the fact that $(s, D) \mapsto e^{-s+D/2}$ and $(s, D) \mapsto e^{s+D/2}$ are parabolic functions in $\mathbb{R} \times (0, \infty)$.) If Skeptic starts with $(\text{LM } u)(0, 1) - 2\epsilon$, it is clear from (6.21) (where $\overline{U}$ is LM $u$) that, by choosing $dS_n = \pm N^{-1/2}$ with a suitable sign, Reality will be able to keep Skeptic's capital below $(\text{LM } u)(\mathcal{S}_n, D_n) - \epsilon$ until (and including, since all partial derivatives are bounded) the first time $(\mathcal{S}_n, D_n)$ is outside $\{(s, D) \in \mathcal{D}^* | s \in (-C, C)\}$ (here and below we are assuming that $N$ is large enough); after that time, by choosing $dS_i = 0$, $i \geq n$, Reality will lock Skeptic's capital below $(\text{LM } u)(\mathcal{S}_n, D_n) - \epsilon$, which is

- less than $U(\mathcal{S}_n) = U(\mathcal{S}_N)$ if $|\mathcal{S}_n| < C$ (by the definition of $\mathcal{D}^*$);
- negative if $|\mathcal{S}_n| \geq C$ and, therefore, again less than $U(\mathcal{S}_n) = U(\mathcal{S}_N) = 0$.

Therefore, Reality will always be able to prevent Skeptic from getting $U(\mathcal{S}_N)$ or more in the end. ∎

### Indicator functions

The case of indicator functions, although it is the most important case in applications where we adopt the fundamental interpretative hypothesis, is not covered directly by Theorem 6.1, because an indicator function is not continuous. But if the event $E$ is reasonably simple (if it is a finite union of intervals, say), then the indicator function $\mathbb{I}_E$ can be approximated arbitrarily closely by continuous functions, both from above and below, and so there is no difficulty in proving an analogue of Theorem 6.1.

Suppose $u = \mathbb{I}_E$ is the indicator function of $E$, a finite union of closed intervals; let us see what LM $u$ looks like in this case. It is clear that $(\text{LM } u)(s, D) = 1$ if $s \in E$; therefore we assume that $s$ is outside $E$. Let $a$ be the largest number in $E$

such that $a < s$ and $b$ be the smallest number in $E$ such that $b > s$. The function $\overline{U}(s, D)$, for all $s \in [a, b]$, is the solution of the heat equation (6.11) with the initial condition $\overline{U}(s, 0) = 0$ and the boundary conditions $\overline{U}(a, D) = 1$ and $\overline{U}(b, D) = 1$. It is clear that the solution is convex in $s \in [a, b]$. The explicit form of the solution can be found in standard texts about partial differential equations (see, e.g., [49], Chapter 6), but we will never need it.

### Convex functions

Theorem 6.1 gives us no help when $U$ is an unbounded convex function. Dealing with unbounded convex initial conditions will become important for us in Part II, because put and call options have unbounded convex payoffs; see Chapter 13. But this problem is actually relatively easy, because when the initial temperature distribution $U$ is convex, the thermostats will never switch on. This suggests that under suitable regularity conditions,

$$\lim_{N \to \infty} \overline{\mathbb{E}}\, U(\mathcal{S}_N) = \int U(z) d\mathcal{N}_{0,1}(dz). \tag{6.29}$$

The proof of De Moivre's theorem shows that (6.29) holds provided that $U$, in addition to being convex, is smooth and has its third and fourth derivatives bounded in absolute value. (Indeed, $\partial^2 \overline{U}/\partial s^2$, being an average of $U^{(2)}$, is nonnegative; the heat equation shows that $\partial \overline{U}/\partial D$ is also nonnegative, and so the second term on the right-hand side of (6.21) is nonpositive. This proves the positive part; the negative part follows directly from De Moivre's theorem if the condition that $U$ is constant outside a finite interval is relaxed to the one really used: $|U^{(3)}|$ and $|U^{(4)}|$ are bounded.) Such functions can approximate a wide class of functions, including the payoffs of standard put and call options.

## 6.4 APPENDIX: THE GAUSSIAN DISTRIBUTION

A random variable $x$ with probabilities

$$\mathbb{P}\{a < x < b\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz \tag{6.30}$$

has the standard *normal* or standard *Gaussian* distribution. We use the latter name in this book. We write $\mathcal{N}_{0,1}$ for the probability distribution

$$\mathcal{N}_{0,1}([a, b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz, \tag{6.31}$$

and we may write (6.30) in the form

$$\mathbb{P}\{a < x < b\} = \int_a^b \mathcal{N}_{0,1}(dz).$$

More generally, we write $\mathcal{N}_{\mu,\sigma^2}$ for the Gaussian distribution with mean $\mu$ and variance $\sigma^2$. If $x$ is standard Gaussian, then $\mu + \sigma x$ has this distribution.

Further information about the Gaussian distribution is provided by any textbook on probability, and most give tables of (6.31); see, for example, [120], §VII.1.

## 6.5 APPENDIX: STOCHASTIC PARABOLIC POTENTIAL THEORY

In §6.3, we relied on intuitions about heat propagation to guide our thinking about the heat equation. The stochastic interpretation of the heat equation in terms of Brownian motion provides another important source of intuition. Leaving aside the needed regularity conditions on $U$, consider the least superparabolic majorant LM $u$ of $u(s, D) := U(s)$. The standard stochastic representation of $\overline{U} = \text{LM}\, u$ is

$$\overline{U}(s, D) = \sup_{\tau \in \mathcal{T}[0,D]} \mathbb{E}\, U\left(s + W(\tau)\right), \tag{6.32}$$

where $W$ is standard Brownian motion, and $\mathcal{T}[0, D]$ is the set of all stopping times in $[0, D]$. The reader familiar with the standard approach to pricing American options (see also §13.4 below) will notice close similarity between the one-sided central limit theorem in the form $\overline{\mathbb{E}}\, U(\mathcal{S}_N) \to \overline{U}(0,1)$, where $\overline{U}(0,1)$ is defined by (6.32), and the usual formulas for the value of an American option in terms of the Snell envelope ([107], §8E). The one-sided central limit theorem is actually of the same form (in both the stochastic and physical representations) as the pricing formula for American options in the Bachelier model of Chapter 11.

The parabolic measure $\mu_{B(s,D,\delta)}(s', D')$ gives the probabilities for where the space-time Brownian motion $(s' + W(t), D' - t)$, starting from $(s', D')$, hits the boundary of $B(s, D, \delta)$ ([103], §2.VII.12; a similar assertion is also true for sets much more general than $B(s, D, \delta)$: see [103], Theorem 3.II.2(b), parabolic context). This stochastic interpretation of $\mu_{B(s,D,\delta)}$ makes (6.32) almost obvious. It also shows that superparabolicity of a function $u$ is a variant of the supermartingale property (at least in a local sense) for the composition $u(W(t), D - t)$.

To help the reader appreciate the usefulness of the stochastic interpretation, we list several cases where stochastic arguments shed new light on what we did in §6.3:

- The representation (6.32) provides a heuristic argument in favor of the negative half of the one-sided central limit theorem,

$$\liminf_{N \to \infty} \overline{\mathbb{E}}\, U\left(\mathcal{S}_N\right) \geq \sup_{\tau \in \mathcal{T}[0,1]} \mathbb{E}\, U\left(W(\tau)\right) \tag{6.33}$$

  (as before, even when Reality is restricted to choosing $x_n$ from the three-element set $\{-N^{-1/2}, 0, N^{-1/2}\}$). Fix some stopping time $\tau$ (which can be arbitrarily close to attaining the supremum in (6.33)). Let Reality choose $x_n = N^{-1/2}$ or $x_n = -N^{-1/2}$ with equal probabilities $1/2$, independently for different $n$, before stopping time $N\tau$; after $N\tau$, Reality always chooses $x_n = 0$. De Moivre's theorem in its functional form (see, e.g., [24]) asserts that the normalized sequence $\mathcal{S}_n$, $n = 1, \ldots, N$, is well approximated by the standard Wiener process ($N$ is assumed sufficiently large). Since $\mathcal{K}_n$ is a martingale, its starting value cannot be much smaller than the right-hand side of (6.33) if its final value is at least $U(\mathcal{S}_N)$.

- In our proof of the negative half of Theorem 6.1 we used the fact that for a large constant $C$, $(\mathrm{LM}\, u)(s, D)$ is negligible for $D \in (0, 1]$ and $|s| \geq C$. Instead of the argument given on p. 142, one can use (6.32).

- In the subsection about convex functions (p. 143) we proved that for convex $U$ the game-theoretic price of $U(\mathcal{S}_N)$ is given by the usual Gaussian integral. For the stochastic interpretation of this observation, consider the representation (6.32). According to Jensen's inequality, the convexity of $U$ implies

$$\overline{U}(\mathcal{S}_n, D_n) = \sup_{\tau \in \mathcal{T}[0, D_n]} \mathbb{E}\, U\left(\mathcal{S}_n + W(\tau)\right) = \mathbb{E}\, U\left(\mathcal{S}_n + W(D_n)\right).$$

- In the following subsection we considered the case where $U$ was the indicator of some set $E$, which we assumed to be the union of a finite set of closed intervals. If $U(s) = 1$, then the stochastic definition (6.32) of $\overline{U}(s, D)$ reduces to 1; if $U(s) = 0$, it reduces to the probability that a standard one-dimensional Wiener process starting from $s$ will hit $E$ in time $D$ or less. At $(s, D)$ the optimal stopping time $\tau$ in (6.32) is "now" if $U(s) = 1$ and the hitting time for $E$ if $U(s) = 0$.

In general, the stochastic interpretation of differential equations has become a very powerful mathematical tool ([166], Chapter 4, or [107], Appendix E)—so much so that some probabilists now regard potential theory, both classical and parabolic, as a special topic in the theory of Markov processes.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 7

# *Lindeberg's Theorem*

Jarl Waldemar Lindeberg (1876–1932)

In a famous article published in 1922, Jarl Waldemar Lindeberg gave a condition under which the central limit theorem holds for independent random variables with finite variances. Lindeberg's condition, as it turns out, is roughly necessary as well as sufficient, and so his theorem is now regarded as the definitive statement of the classical central limit theorem. It is now usually proven in a general measure-theoretic context, where it applies to measure-theoretic martingales rather than to independent random variables.

In this chapter, we generalize Lindeberg's theorem to martingales in certain probability protocols. In this context, Lindeberg's condition requires that extreme values for the increments of a martingale have small game-theoretic upper probability, and the theorem concludes that the price of any well-behaved function of the final value of the martingale is given by the integral of the function with respect to the standard Gaussian distribution.

This game-theoretic version of Lindeberg's theorem generalizes the game-theoretic version of De Moivre's theorem that we proved in the preceding chapter, and its proof has the same basic structure. To establish the approximate validity of a Gaussian price for a given function of the final value of a given martingale, we construct an approximate martingale that begins with the proposed price and ends with the value of the function. To show that the constructed process is indeed an approximate martingale, we relate its increments to the increments of the given martingale by expanding it in a Taylor's series and then using the heat equation to combine some of the leading terms. As we will see in Part II, this same strategy of proof also underlies the pricing of options in finance. As we remarked in the preceding chapter, it was also more or less the strategy of Lindeberg's own proof.

Our game-theoretic version of De Moivre's theorem was concerned with a concrete protocol—a protocol for betting on $N$ successive tosses of a fair coin. Our game-theoretic version of Lindeberg's theorem, in contrast, is abstract; it is concerned with a martingale in any *Lindeberg protocol*—any coherent, terminating, symmetric probability protocol. We introduced such protocols in Chapter 1, but our discussion there was informal, and in subsequent chapters we have almost always worked with specific examples. So we begin this chapter with a brief but formal study of Lindeberg protocols (§7.1). We then state and prove the theorem (§7.2).

Although Lindeberg's theorem is very powerful, its abstractness can disguise its relevance to specific examples. In §7.3, we try to remedy this by describing several specific protocols in which the theorem gives interesting results.

In an appendix, §7.4, we review the history of the central limit theorem. The measure-theoretic version of Lindeberg's theorem is stated precisely in Chapter 8, where we show that it follows from the game-theoretic version of this chapter (Corollary 8.4).

## 7.1 LINDEBERG PROTOCOLS

As we have just said, a Lindeberg protocol is a coherent terminating symmetric probability protocol. In this section, we repeat carefully the definitions of these terms that were given in Chapter 1, and we extend our notation and terminology slightly. Further abstract theory about probability protocols, which is not needed for Lindeberg's theorem but may interest some readers, is presented in the next chapter.

### Terminating Protocols

A probability protocol is defined by specifying (1) the moves each player is allowed to make, which may depend on the preceding moves made by World, and (2) Skeptic's gain on each round as a function of the moves made so far by World and the move made on that round by Skeptic.

In the terminating case, which we are now considering, World makes only a finite sequence of moves, no matter what happens. The possibilities for this finite sequence constitute the *sample space* $\Omega$ for the protocol. The sample space $\Omega$ can be any nonempty set of finite nonempty sequences that form the leaves of a tree. This means that no proper initial subsequence of a sequence in $\Omega$ is also in $\Omega$. We call an element of $\Omega$ a *path*.

We call an initial segment of a path a *situation*. We include among the situations the empty sequence, which we call the *initial situation* and designate by $\Box$. A path is itself a situation—a *final* situation. We write $\Omega^\diamond$ for the set consisting of all situations. Thus $\Omega$ is a subset of $\Omega^\diamond$.

In each situation $t$, we set

$$\mathbf{W}_t := \{\mathbf{w} \mid t\mathbf{w} \in \Omega^\diamond\},$$

where $t\mathbf{w}$ is the sequence obtained by placing the single move $\mathbf{w}$ at the end of the sequence of moves $t$. This is the set of moves available to World in $t$; we call it *World's move space* in $t$. An important assumption is being made here: the moves available to World may depend only on his own previous moves, not on the previous moves by Skeptic.

To visualize the sample space, we picture it as a tree rooted at an initial node $\Box$, as in Figure 1.2. The nodes in the tree represent the situations. The branches to the right of a situation $t$ form the set $\mathbf{W}_t$. A path starts with $\Box$ and ends at a final situation.

We adopt the following terminology and notation, mostly familiar from earlier chapters.

- When the situation $t$ is an initial segment of the situation $u$, we say that $t$ *precedes* $u$ and that $u$ *follows* $t$, and we write $t \sqsubseteq u$. (The initial situation $\Box$ precedes every situation.)
- When $t \sqsubseteq u$ and $t \neq u$, we say that $t$ *strictly precedes* $u$ and that $u$ *strictly follows* $t$, and we write $t \sqsubset u$.
- A situation $u$ is a *child* of a situation $t$ (and $t$ is the *parent* of $u$) if $t \sqsubset u$ and there is no third situation between them—no situation $v$ such that $t \sqsubset v \sqsubset u$. Every noninitial situation $u$ has a unique parent.
- When $\xi$ is a path and $t \sqsubseteq \xi$, we say that $t$ *is on* $\xi$ and that $\xi$ *goes through* $t$.
- We call a function on $\Omega^\diamond$ a *process*. We call a partial process (a function on a subset of $\Omega^\diamond$) a *t-process* if its domain of definition includes all situations that follow $t$. (Thus a process is the same as a $\Box$-process.)
- We call a function on $\Omega$ a *variable*. We call a partial variable (a function on a subset of $\Omega$) a *t-variable* if its domain of definition includes all paths that go through $t$.

We obtain a variable by considering the possible final values of a process—its values in $\Omega$. We write $\mathcal{S}_\Omega$ for the variable thus obtained from a process $\mathcal{S}$:

$$\mathcal{S}_\Omega(\xi) := \mathcal{S}(\xi)$$

for each path $\xi \in \Omega$. When all the sequences in $\Omega$ have the same length $N$, we write $\mathcal{S}_N$ for $\mathcal{S}_\Omega$. We sometimes omit the subscript $\Omega$ or $N$ if the context makes clear that we are treating $\mathcal{S}$ as a variable rather than as a process.

We designate by $\mathbf{S}_t$ the set of moves available to Skeptic in a given nonfinal situation $t$, and we call $\mathbf{S}_t$ *Skeptic's move space* in $t$. We assume (the assumption of symmetry is included here) that $\mathbf{S}_t$ is a linear space: if $\mathbf{s}_1$ and $\mathbf{s}_2$ are in $\mathbf{S}_t$ and $a_1$ and $a_2$ are real numbers, then $a_1\mathbf{s}_1 + a_2\mathbf{s}_2$ is in $\mathbf{S}_t$.

We also specify for each nonfinal situation $t$ a real-valued function $\lambda_t$ on $\mathbf{S}_t \times \mathbf{W}_t$, which we call Skeptic's *gain function*. The quantity $\lambda_t(\mathbf{s}, \mathbf{w})$ is the increase in Skeptic's capital in situation $t$ if he moves $\mathbf{s}$ and then World moves $\mathbf{w}$. We assume that $\lambda_t$ is linear in $\mathbf{s}$.

A *strategy* for Skeptic is a partial process defined on the nonfinal situations. In the nonfinal situation $t$, the strategy $\mathcal{P}$ directs Skeptic to make the move $\mathcal{P}(t)$. We write $\mathcal{K}^{\mathcal{P}}$ for Skeptic's capital process when he starts with capital 0 and uses the strategy $\mathcal{P}$; this process is defined inductively by $\mathcal{K}^{\mathcal{P}}(\Box) = 0$ and

$$\mathcal{K}^{\mathcal{P}}(t\mathbf{w}) = \mathcal{K}^{\mathcal{P}}(t) + \lambda_t(\mathcal{P}(t), \mathbf{w}). \tag{7.1}$$

Although $\mathcal{P}$ may be defined only for nonfinal situations, $\mathcal{K}^{\mathcal{P}}$ is defined for all situations.

When Skeptic begins with capital $\alpha$ in $\Box$ and follows the strategy $\mathcal{P}$, his capital process is $\alpha + \mathcal{K}^{\mathcal{P}}$. Under the assumptions we have made (including symmetry), we call a process a *martingale* if it is equal to $\alpha + \mathcal{K}^{\mathcal{P}}$ for some real number $\alpha$ and some strategy $\mathcal{P}$—that is, if it is Skeptic's capital process for some initial capital (which may be negative) and some strategy (see p. 53). Because the strategies form a linear space, the martingales also form a linear space.

## Beginning in a Situation

In addition to strategies and capital processes that begin in the initial situation $\Box$, we can talk about strategies and capital processes that begin in some later situation $s$. A partial process that is defined in the nonfinal situations following $s$ can be interpreted as telling Skeptic how to move in $s$ and afterwards; this is an *$s$-strategy*. When Skeptic begins in $s$ with capital 0 and follows a particular *$s$-strategy* $\mathcal{P}$, his capital process from $s$ onward is the $s$-process $\mathcal{K}^{\mathcal{P}}$ defined by $\mathcal{K}^{\mathcal{P}}(s) = 0$ and (7.1). When he begins with $\alpha$ instead of 0 in $s$, it is $\alpha + \mathcal{K}^{\mathcal{P}}$. We call any $s$-process of this form an *$s$-martingale*.

## Coherence

We call a protocol *coherent* in $\Box$ if World has a strategy that guarantees that Skeptic ends the game with no cumulative gain. This is equivalent (see Appendix 4.6) to the requirement that Skeptic have no strategy that guarantees a cumulative gain. In other words, for every strategy $\mathcal{P}$ for Skeptic, there is a path $\xi$ such that $\mathcal{K}^{\mathcal{P}}(\xi) \leq 0$.

The notion of coherence can, of course, be applied beginning in any situation $t$. We call a protocol *coherent in $t$* if World has a strategy for play beginning in $t$ that guarantees that Skeptic makes no cumulative gain from $t$ onward. This implies that for every $\mathbf{s} \in \mathbf{S}_t$, World has a move $\mathbf{w} \in \mathbf{W}_t$ such that $\lambda_t(\mathbf{s}, \mathbf{w}) \leq 0$; otherwise

Skeptic could achieve a cumulative gain by moving s in $t$ and then preserve it by moving 0 thereafter.

We call a protocol *everywhere coherent* if it is coherent in every situation $t$. This is equivalent to the requirement that for every situation $t$ and every $\mathbf{s} \in \mathbf{S}_t$, World has a move $\mathbf{w} \in \mathbf{W}_t$ such that $\lambda_t(\mathbf{s}, \mathbf{w}) \leq 0$.

## Game-Theoretic Price and Probability

Now we study upper and lower prices and probabilities. We define them just as in Chapter 1, but with slightly more formality.

Given a nonfinal situation $t$ and a $t$-variable $x$, we define $\overline{\mathbb{E}}_t\, x$, the *upper price* of $x$ in $t$, and $\underline{\mathbb{E}}_t\, x$, the *lower price* of $x$ in $t$, by

$$\overline{\mathbb{E}}_t\, x := \inf \left\{ \alpha \mid \text{there is some } t\text{-strategy } \mathcal{P} \text{ such that } \mathcal{K}_\Omega^\mathcal{P} \geq x - \alpha \right\} \tag{7.2}$$

and

$$\underline{\mathbb{E}}_t(x) := \sup \left\{ \alpha \mid \text{there is some } t\text{-strategy } \mathcal{P} \text{ such that } \mathcal{K}_\Omega^\mathcal{P} \geq \alpha - x \right\}. \tag{7.3}$$

Here $\mathcal{K}^\mathcal{P}$ is defined as in (7.1). The inequalities in (7.2) and (7.3) must be interpreted in light of the fact that $x$ and $\mathcal{K}_\Omega^\mathcal{P}$ are only $t$-variables. In (7.2), for example, $\mathcal{K}_\Omega^\mathcal{P} \geq x - \alpha$ means that $\mathcal{K}_\Omega^\mathcal{P}(\xi) \geq x(\xi) - \alpha$ for every path $\xi$ that goes through $t$.

It follows from (7.2) and (7.3) that

$$\underline{\mathbb{E}}_t\, x = -\overline{\mathbb{E}}_t[-x]. \tag{7.4}$$

From (7.2) and (7.4), we obtain the following proposition.

**Proposition 7.1** *If $\mathcal{S}$ is a t-martingale, then $\overline{\mathbb{E}}_t\, \mathcal{S}_\Omega \leq \mathcal{S}(t)$ and $\underline{\mathbb{E}}_t\, \mathcal{S}_\Omega \geq \mathcal{S}(t)$.*

When $\overline{\mathbb{E}}_t\, x$ and $\underline{\mathbb{E}}_t\, x$ are equal, we write $\mathbb{E}_t\, x$ for their common value and call $\mathbb{E}_t\, x$ the (game-theoretic) *price* of $x$ in $t$. When $\mathbb{E}_t\, x$ exists, we say that $x$ is *exactly priced* in $t$.

So far we have not used coherence. When we do, we obtain the following proposition.

**Proposition 7.2** *Suppose the protocol is coherent in t.*

1. *If $x$ is a t-variable, then $\underline{\mathbb{E}}_t\, x \leq \overline{\mathbb{E}}_t\, x$.*

2. *If $\mathcal{S}$ is a t-martingale, then $\mathbb{E}_t\, \mathcal{S}_\Omega = \mathcal{S}(t)$.*

3. *If $r$ is a real number and $\mathbf{r}$ designates the variable that is equal to $r$ on every path, then $\mathbb{E}_t\, \mathbf{r} = r$.*

*Proof* If $\underline{\mathbb{E}}_t\, x \leq \overline{\mathbb{E}}_t\, x$ did not hold, then there would exist constants $\alpha_1$ and $\alpha_2$ satisfying $\overline{\mathbb{E}}_t\, x < \alpha_1 < \alpha_2 < \underline{\mathbb{E}}_t\, x$ and $t$-strategies $\mathcal{P}_1$ and $\mathcal{P}_2$ satisfying $\mathcal{K}_\Omega^{\mathcal{P}_1} \geq x - \alpha_1$ and $\mathcal{K}_\Omega^{\mathcal{P}_2} \geq \alpha_2 - x$. The $t$-strategy $\mathcal{P}_1 + \mathcal{P}_2$ would then satisfy $\mathcal{K}_\Omega^{\mathcal{P}_1+\mathcal{P}_2} > 0$, contradicting coherence in

$t$. Statement 2 follows from Statement 1 and Proposition 7.1. Statement 3 then follows from the fact that $\mathbf{r}$ is a martingale. ∎

The *upper and lower probabilities* of an event $E$ in the situation $t$ are the upper and lower prices of the indicator function $\mathbb{I}_E$ in those situations:

$$\overline{\mathbb{P}}_t\, E := \overline{\mathbb{E}}_t\, \mathbb{I}_E \quad \text{and} \quad \underline{\mathbb{P}}_t\, E := \underline{\mathbb{E}}_t\, \mathbb{I}_E.$$

As we explained in Chapter 1, the fundamental hypothesis of probability makes $\overline{\mathbb{P}}_t\, E$ interesting when it is close to zero (this means $E$ is very unlikely) and $\underline{\mathbb{P}}_t\, E$ interesting when it is close to one (this means $E$ is nearly certain).

In measure-theoretic probability, we are accustomed to making the notion of a probability very close to zero vivid by imagining a sequence of events whose probability converges to zero. We say, for example, that a sequence of variables $x_1, x_2, \ldots$, in probability spaces $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$, $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2), \ldots$, respectively, *converges in probability* to a real number $c$ if for any positive real numbers $\epsilon$ and $\delta$ there is an integer $N$ such that $\mathbb{P}_n\{|x_n - c| > \delta\} < \epsilon$ for all $n$ greater than $N$. We can make an analogous definition for probability protocols: a sequence of variables $x_1, x_2, \ldots$ in a sequence of probability protocols *converges in probability* to a real number $c$ if for any positive real numbers $\epsilon$ and $\delta$ there is an integer $N$ such that $\overline{\mathbb{P}}_{\square_n}\{|x_n - c| > \delta\} < \epsilon$ for all $n$ greater than $N$, where $\square_n$ denotes the initial situation in the $n$th protocol.

When $x$ is exactly priced in $t$, we defineits *upper variance* $\overline{\mathbb{V}}_t\, x$ and its *lower variance* $\underline{\mathbb{V}}_t\, x$ by

$$\overline{\mathbb{V}}_t\, x := \overline{\mathbb{E}}_t(x - \mathbb{E}_t(x))^2 \quad \text{and} \quad \underline{\mathbb{V}}_t\, x := \underline{\mathbb{E}}_t(x - \mathbb{E}_t(x))^2.$$

When $\overline{\mathbb{V}}_t\, x$ and $\underline{\mathbb{V}}_t\, x$ are equal, we write $\mathbb{V}_t(x)$ for their common value, the (game-theoretic) *variance* of $x$ in $t$.

As in the preceding chapter, we use the symbol $\overline{\underline{\mathbb{E}}}$ when $\overline{\mathbb{E}}_t\, x$ and $\underline{\mathbb{E}}_t\, x$ are not necessarily equal but we want to make the same assertion about both of them. For example, $\overline{\underline{\mathbb{E}}}_t\, x < b$ means that both $\overline{\mathbb{E}}_t\, x$ and $\underline{\mathbb{E}}_t\, x$ are less than $b$. Similarly, we use $\overline{\underline{\mathbb{P}}}$ to make simultaneous statements about an event's upper and lower price.

## The Gains of a Martingale

Given a martingale $\mathcal{S}$ and a nonfinal situation $t$, we define a process $d_t\mathcal{S}$ by

$$d_t\mathcal{S}(u) := \begin{cases} \mathcal{S}(t_u^+) - \mathcal{S}(t) & \text{if } t \sqsubset u \\ 0 & \text{otherwise,} \end{cases} \tag{7.5}$$

where $t_u^+$ is the next situation towards $u$ after $t$—i.e., $t_u^+ := t\mathbf{w}$, where $\mathbf{w}$ is the next entry in the sequence $u$ after the end of the initial subsequence $t$ (see Figure 7.1). We call $d_t\mathcal{S}$ the *gain* of $\mathcal{S}$ in $t$. The gain $d_t\mathcal{S}$ is itself a martingale; it is the capital process for Skeptic when he moves in $t$ following the strategy that produces $\mathcal{S}$ but moves 0 everywhere else. We can recover a martingale by adding up its gains:

$$\mathcal{S} = \mathcal{S}(\square) + \sum_{t \text{ nonfinal}} d_t\mathcal{S}$$

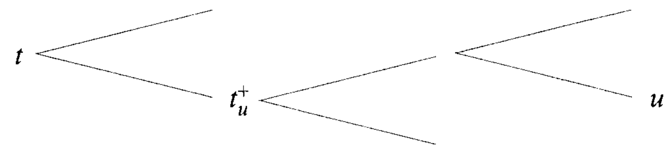



***Fig. 7.1*** When the situation $u$ strictly follows the situation $t$, we write $t_u^+$ for the first situation on the way to $u$ from $t$.

(this sum may be even uncountable, but in every situation only finitely many addends are nonzero). Because the martingale $d_t\mathcal{S}$ changes value only in $t$, all the information in it is also contained in the variable $(d_t\mathcal{S})_\Omega$. When there is no danger of confusion, we use $d_t\mathcal{S}$, without the subscript, to designate the variable as well as the martingale. We say, for example, that $\mathbb{E}_u\, d_t\mathcal{S} = d_t\mathcal{S}(u)$; in particular, $\mathbb{E}_t\, d_t\mathcal{S} = 0$.

We may get some idea of what to expect from a martingale by looking at the upper and lower variances of its gains. In particular, we may want to look at the accumulated upper and lower variances, $\overline{\mathrm{Var}}_\mathcal{S}$ and $\underline{\mathrm{Var}}_\mathcal{S}$, defined by

$$\overline{\mathrm{Var}}_\mathcal{S}(\xi) := \sum_{t \sqsubset \xi} \overline{\mathbb{V}}_t\, d_t\mathcal{S}$$

and

$$\underline{\mathrm{Var}}_\mathcal{S}(\xi) := \sum_{t \sqsubset \xi} \underline{\mathbb{V}}_t\, d_t\mathcal{S}.$$

As we will see in Example 5 in §7.3, the sample space $\Omega$ is sometimes designed so that these variables are likely to reach a certain level. This may allow us to use Lindeberg's theorem to test whether the final size of $\mathcal{S}$ is consistent with the claim that it is a martingale.

## 7.2 STATEMENT AND PROOF OF THE THEOREM

Consider a martingale $\mathcal{S}$ in a Lindeberg protocol. For every positive number $\delta$, we define a variable $L_{\mathcal{S},\delta}$:

$$L_{\mathcal{S},\delta}(\xi) = \sum_{t \sqsubset \xi} \overline{\mathbb{E}}_t[(d_t\mathcal{S})^2 \mathbb{I}_{|d_t\mathcal{S}| \geq \delta}].$$

These variables measure the degree to which exceptionally large increments are found along the path $\xi$, and Lindeberg's condition requires them to be small in probability.

Our game-theoretic version of Lindeberg's theorem can be stated as follows:

**Theorem 7.1** *Suppose $U$ is a bounded continuous function. Then the initial upper and lower prices of $U(\mathcal{S}_\Omega)$ are both arbitrarily close to $\int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz)$ for every* *martingale $\mathcal{S}$ in every Lindeberg protocol for which the initial lower probability of the conjunction*

$$1-\delta \leq \underline{\mathrm{Var}}_{\mathcal{S}} \leq \overline{\mathrm{Var}}_{\mathcal{S}} \leq 1+\delta \text{ and } L_{\mathcal{S},\delta} \leq \delta$$

*is sufficiently close to one for a positive number $\delta$ sufficiently close to* 0.

The condition that $L_{\mathcal{S},\delta} \leq \delta$ have lower probability close to one is our version of Lindeberg's condition. The condition that the lower probability of $1-\delta \leq \underline{\mathrm{Var}}_{\mathcal{S}} \leq \overline{\mathrm{Var}}_{\mathcal{S}} \leq 1+\delta$ be close to one makes precise the idea that successive gains have approximate game-theoretic variances that probably approximately sum to 1. The theorem remains valid if we replace $1 \pm \delta$ in this condition by $\sigma^2 \pm \delta$, where $\sigma^2$ is any positive real number; in this case $\mathcal{N}_{0,1}$ becomes $\mathcal{N}_{0,\sigma^2}$.

If we fix an interval $[a,b]$ and choose the continuous function $U$ so that it approximates the (discontinuous) indicator function for $[a,b]$, then the conclusion of the theorem, that $\overline{\mathbb{E}}_{\square}\, U(\mathcal{S}_{\Omega})$ and $\underline{\mathbb{E}}_{\square}\, U(\mathcal{S}_{\Omega})$ are both close to $\int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz)$, becomes

$$\overline{\underline{\mathbb{P}}}_{\square}\{a < \mathcal{S}_{\Omega} < b\} \approx \int_a^b \mathcal{N}_{0,1}(dz).$$

More colloquially: $\mathcal{S}_{\Omega}$ has approximately a standard Gaussian distribution at the beginning of the game.

The meaning of the terms "sufficiently close" and "arbitrarily close" can be spelled out as follows: For every bounded continuous function $U$ and every $\epsilon > 0$, there exist $\delta > 0$ and $\zeta > 0$ such that if $\mathcal{S}$ is a martingale in a coherent terminating probability protocol and $\mathcal{S}$ satisfies

$$\underline{\mathbb{P}}_{\square}\left\{1-\delta \leq \underline{\mathrm{Var}}_{\mathcal{S}} \leq \overline{\mathrm{Var}}_{\mathcal{S}} \leq 1+\delta \ \&\ L_{\mathcal{S},\delta} \leq \delta\right\} > 1-\zeta,$$

then

$$\left|\overline{\underline{\mathbb{E}}}_{\square}\, U(\mathcal{S}_{\Omega}) - \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz)\right| < \epsilon.$$

The same $\delta$ and $\zeta$ work for every martingale in every Lindeberg protocol.

Alternatively, we can restate the theorem in terms of convergence in probability. Suppose $U$ is a bounded continuous function and $\mathcal{S}^1, \mathcal{S}^2, \ldots$ is a sequence of martingales in a sequence of coherent terminating probability protocols. Write $\square_n$ for the initial situation and $\Omega_n$ for the sample space in the $n$th protocol. And suppose these two conditions are satisfied:

1. $\overline{\mathrm{Var}}_{\mathcal{S}^n}$ and $\underline{\mathrm{Var}}_{\mathcal{S}^n}$ converge in probability to 1.

2. For every $\delta$ greater than zero, $L_{\mathcal{S}^n,\delta}$ converges in probability to zero.

Then $\overline{\underline{\mathbb{E}}}_{\square_n}\, U(\mathcal{S}^n_{\Omega_n})$ converges to $\int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz)$.

Readers who wish to compare our formulation of Lindeberg's theorem with the classical formulation can find the latter reviewed in §7.4.

*Proof of Lindeberg's theorem.* Our proof will follow the same lines as the proof of De Moivre's theorem in the preceding chapter, with one important difference: we move the center of Taylor's expansion.

We begin by noticing that it suffices to show that for every $\epsilon > 0$ the hypotheses of the theorem imply

$$\overline{\mathbb{E}}_{\square}\, U(\mathcal{S}) \le \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) + \epsilon. \tag{7.6}$$

First we assume that $U$ is a smooth function that is constant outside a compact interval. We introduce the function

$$g(h) := \sup_x \left| U(x+h) - U(x) - U'(x)h - \frac{1}{2}U''(x)h^2 \right|$$

and note that

$$\left| (U(x+h) - U(x)) - \left( U'(x)h + \frac{1}{2}U''(x)h^2 \right) \right| \le g(h) \tag{7.7}$$

and, for some constant $K$,

$$g(h) \le K|h|^3,\ g(h) \le Kh^2,\ \forall h.$$

We shall suppose that $K > 2$.

Fix arbitrarily small $\delta \in (0,1)$. Let $\zeta \in (0,1)$ be so small that

$$\left| \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1+\zeta}(dz) - \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) \right| \le \delta \tag{7.8}$$

and, for all $v \in \mathbb{R}$ and all $\Delta \in (0, 2\zeta]$,

$$\left| \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{v,\Delta}(dz) - U(v) \right| \le \delta. \tag{7.9}$$

Let us check that such $\zeta$ indeed exists. The existence of $\zeta$ satisfying (7.9) follows from $U$ being uniformly continuous and bounded—see, for example, [121], Lemma VII.1.1. Substituting the function $v \mapsto \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{v,1}(dz)$ for $U$, $\zeta$ for $\Delta$, and 0 for $v$ in (7.9), we obtain (7.8).

We assume that the lower probability of the conjunction of the inequalities

$$1 - \zeta \le \sum_{s \sqsubset \xi} \underline{\mathbb{E}}_s (d_s\mathcal{S})^2 \le \sum_{s \sqsubset \xi} \overline{\mathbb{E}}_s (d_s\mathcal{S})^2 \le 1 + \zeta, \tag{7.10}$$

$$\sum_{s \sqsubset \xi} \overline{\mathbb{E}}_s [(d_s\mathcal{S})^2 \mathbb{I}_{|d_s\mathcal{S}| \ge \delta^2}] \le \delta^4 \tag{7.11}$$

(which is "sufficiently small" according to the statement of the theorem) is at least $1 - \delta$. Inequality (7.11) implies

$$\sum_{s \sqsubset \xi} \overline{\mathbb{E}}_s [(d_s\mathcal{S})^2 \mathbb{I}_{|d_s\mathcal{S}| \ge \delta}] \le \delta \tag{7.12}$$

and, for all $s \sqsubset \xi$,

$$\overline{\mathbb{E}}_s (d_s\mathcal{S})^2 \le 2\delta^4. \tag{7.13}$$

We shall show that for some martingale $\mathcal{V}$ that is bounded below by a constant independent of $\delta$, we have

$$\mathcal{V}(\square) \le \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1+\zeta}(dz) + C(\delta + \zeta), \tag{7.14}$$

where $C$ is a constant (depending only on $U$), and, for all $\xi$ satisfying (7.10) and (7.11),

$$\mathcal{V}(\xi) \geq U(\mathcal{S}(\xi)). \tag{7.15}$$

Since $\xi$ satisfies (7.10) and (7.11) with lower probability at least $1-\delta$, our goal will be achieved (remember (7.8), that $\delta$ can be taken arbitrarily small, and that $\mathcal{V}$ is bounded below).

Put, for each situation $s$,

$$S(s) := \mathcal{S}(s), \quad D(s) := 1 + \zeta - \sum_{t \sqsubset s} A(t)$$

(these are analogues of $S_n$ and $D_n$ in (6.21)), where

$$A(t) := \frac{\overline{\mathbb{E}}_t(d_t\mathcal{S})^2 + \underline{\mathbb{E}}_t(d_t\mathcal{S})^2}{2}.$$

Let $\overline{U}$ be defined as in Chapter 6 (see (6.10) on p. 128), except that we put $\overline{U}(s, D) := U(s)$ when $D \leq 0$. We construct $\mathcal{V}$ so that it approximates, or at least is not too far below, $\overline{U}(S(s), D(s))$.

Define variables $S_n$ and $D_n$ by

$$S_n(\xi) := S(\xi^n) \qquad \text{and} \qquad D_n(\xi) := D(\xi^n),$$

for $n = 0, 1, \ldots$. Instead of using (6.21) directly, we will use a similar expansion, but around the point $(S_n, D_{n+1})$ rather than $(S_n, D_n)$. It should be clear why this makes sense in our current context: we are interested in the connection between $\overline{U}(S_n, D_n)$ and $\overline{U}(S_{n+1}, D_{n+1})$, the point $(S_n, D_{n+1})$ is the middle ground between $(S_n, D_n)$ and $(S_{n+1}, D_{n+1})$, and Skeptic knows $S_n$ and $D_{n+1}$ when he makes his own move in round $n+1$. Therefore, we replace (6.21) with

$$\begin{aligned} d\overline{U}(S_n, D_n) \leq \frac{\partial \overline{U}}{\partial D}(S_n, D_{n+1})\left(dD_n + (dS_n)^2\right) + \frac{\partial \overline{U}}{\partial S}(S_n, D_{n+1}) dS_n \\ - \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial D^2}(S_n, D_n')(dD_n)^2 + g(dS_n) \end{aligned} \tag{7.16}$$

($D_n'$ lies between $D_n$ and $D_{n+1}$), which can be obtained by subtracting

$$\overline{U}(S_n, D_n) - \overline{U}(S_n, D_{n+1}) = \frac{\partial \overline{U}}{\partial D}(S_n, D_{n+1})(-dD_n) + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial D^2}(S_n, D_n')(dD_n)^2$$

from

$$\begin{aligned} &\overline{U}(S_{n+1}, D_{n+1}) - \overline{U}(S_n, D_{n+1}) \\ \leq &\frac{\partial \overline{U}}{\partial D}(S_n, D_{n+1})(dS_n)^2 + \frac{\partial \overline{U}}{\partial S}(S_n, D_{n+1}) dS_n + g(dS_n). \end{aligned}$$

The last inequality follows from (7.7): first we average (7.7) with respect to $x$ distributed as $\mathcal{N}_{S_n, D_{n+1}}$ (we can do this by Leibniz's differentiation rule for integrals), and then we replace $(1/2)\partial^2\overline{U}/\partial S^2$ by $\partial\overline{U}/\partial D$.

Let us first concentrate on the first two terms on the right-hand side of (7.16); the rest will be negligible. We established in Chapter 6 (see (6.22)) that all partial derivatives on the right-hand side of (7.16) are bounded by some constant $c$. It is clear that there exists a martingale $\mathcal{V}^{(1)}$ starting from

$$\delta + c\zeta \tag{7.17}$$

such that $\mathcal{V}^{(1)}(\xi)$ exceeds the sum of the first two terms summed along $\xi$, for every $\xi$ satisfying (7.10). Indeed, the second term on the right-hand side of (7.16) can be exactly replicated by Skeptic, and $(dD_n + (dS_n)^2)$ in the first term can be replicated with the accuracy equal to the half-difference between the upper and lower variance plus some allowance (which should be taken exponentially decreasing with $n$) for the potential nonattainability of the supremum or infimum in the definition of upper and lower variances. The $\zeta$ in (7.17) arises because of the discrepancies between the lower and upper variances, and we took $\delta$ to be an upper bound on the error due to the allowances involved in super-replicating the first term.

For the third term, we have

$$\sum_n (dD_n)^2 \le \left(\sum_n |dD_n|\right)\left(\max_n |dD_n|\right), \tag{7.18}$$

which shows that it is negligible (cf. (7.10) and (7.13)). For the last term, we split the corresponding sum as follows:

$$\sum_n g(dS_n) \le K\left(\sum_{n:|dS_n|\ge\delta} (dS_n)^2 + \sum_{n:|dS_n|<\delta} |dS_n|^3\right).$$

The first sum on the right-hand side, according to Lindeberg's condition (see (7.12)), can be super-replicated by a martingale $\mathcal{V}^{(2)}$ starting from $2\delta$. The second sum can be bounded as

$$\sum_{n:|dS_n|<\delta} |dS_n|^3 \le \delta \sum_{n:|dS_n|<\delta} (dS_n)^2 \le \delta \sum_n (dS_n)^2$$

and so can be, according to (7.10), super-replicated by a martingale $\mathcal{V}^{(3)}$ starting from $2\delta$.

Setting

$$\mathcal{V} := \overline{U}(S(\square), D(\square)) + \mathcal{V}^{(1)} + K\mathcal{V}^{(2)} + K\mathcal{V}^{(3)} + \delta \tag{7.19}$$

will ensure that (7.14) and (7.15) are satisfied (the $\delta$ in (7.19) corresponds to the $\delta$ in (7.9)).

Let us now show that the sum $\mathcal{V}^{(1)} + K\mathcal{V}^{(2)} + K\mathcal{V}^{(3)}$ can be chosen bounded below. We will assume that the martingales $\mathcal{V}^{(1)}$, $\mathcal{V}^{(2)}$, $\mathcal{V}^{(3)}$ are stopped as soon as it becomes clear that one of the relations (7.10) or (7.11) will be violated; more accurately, Skeptic starts playing 0 from the first situation $\xi^N$ where

$$\sum_{s\sqsubseteq\xi^N} \overline{\mathbb{E}}_s (d_s S)^2 > 1 + \zeta,$$

$$\sum_{s\sqsubseteq\xi^N} \overline{\mathbb{E}}_s (d_s S)^2 - \sum_{s\sqsubseteq\xi^N} \underline{\mathbb{E}}_s (d_s S)^2 > 2\zeta, \tag{7.20}$$

or

$$\sum_{s\sqsubseteq\xi^N} \overline{\mathbb{E}}_s [(d_s S)^2 \mathbb{I}_{|d_s S|\ge\delta^2}] > \delta^4.$$

According to (7.16), in no situation will the sum $\mathcal{V}^{(1)} + K\mathcal{V}^{(2)} + K\mathcal{V}^{(3)}$ be smaller than the increment (from the root to that situation) in $\overline{U}$ plus a multiple of the following values:

- the accumulated error (bounded by $2\zeta$; see (7.20)) due to using $-dD_n$ in place of the corresponding upper or lower variance plus the allowances (as discussed above);
- $\sum_{n<N} (dD_n)^2$, which is negligible by (7.18).

Hence we established that $\mathcal{V}$ is bounded below, which completes the proof under the assumption that $U$ is smooth and constant outside a finite interval.

**Lemma 7.1** *For any $\epsilon > 0$ and bounded continuous $U: \mathbb{R} \to \mathbb{R}$, there exists a constant outside a finite interval and smooth function $\tilde{U}$ such that $U \leq \tilde{U}$ and*

$$\int_{-\infty}^{\infty} \tilde{U}(z)\ \mathcal{N}_{0,1}(dz) \leq \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) + \epsilon. \tag{7.21}$$

*Proof* There are an increasing finite sequence $t_1, \ldots, t_k$ of real numbers and a function $g: \mathbb{R} \to \mathbb{R}$ which is constant on each of the intervals

$$(-\infty, t_1], (t_1, t_2], \ldots, (t_{k-1}, t_k], (t_k, \infty)$$

and satisfies $g(-\infty) = g(\infty)$ and

$$U \leq g, \int_{-\infty}^{\infty} g(z)\ \mathcal{N}_{0,1}(dz) \leq \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) + \frac{\epsilon}{2}.$$

It is well known that there exists a smooth function $\phi: \mathbb{R} \to [0,1]$ such that $\phi(t) = 1$ for $t \leq 0$ and $\phi(t) = 0$ for $t \geq 1$ (see, e.g., (7.1) of [24]). Choose $\delta > 0$ which does not exceed any of the values $(t_2 - t_1)/2, \ldots, (t_k - t_{k-1})/2$. Using the function $\phi$, one can easily construct a smooth function $\tilde{U}$ such that:

- $g \leq \tilde{U}$,
- $\tilde{U}(t) = g(t)$ for each $t \notin \bigcup_{i=1}^{k} (t_i - \delta, t_i + \delta)$;
- $\tilde{U}(t) \in [g(t_i-), g(t_i+)]$ when $t \in (t_i - \delta, t_i + \delta)$.

Choosing sufficiently small $\delta$, we ensure

$$\int_{-\infty}^{\infty} \tilde{U}(z)\ \mathcal{N}_{0,1}(dz) \leq \int_{-\infty}^{\infty} g(z)\ \mathcal{N}_{0,1}(dz) + \frac{\epsilon}{2}.$$

∎

Now we can easily finish the proof. We are required to establish (7.6) for any bounded continuous $U$. By Lemma 7.1, there exists a smooth and constant outside a finite interval function $\tilde{U}$ such that $U \leq \tilde{U}$ and (7.21) holds. Since (7.6) holds for $\tilde{U}$, we have:

$$\begin{aligned} &\overline{\mathbb{E}}\, U(\mathcal{S}) \leq \overline{\mathbb{E}}\, \tilde{U}(\mathcal{S}) \\ &\leq \int_{-\infty}^{\infty} \tilde{U}(z)\ \mathcal{N}_{0,1}(dz) + \epsilon \leq \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) + 2\epsilon. \end{aligned}$$

Since $\epsilon$ can be arbitrarily small, this completes the proof. ∎

## 7.3 EXAMPLES OF THE THEOREM

We now we look at what Lindeberg's theorem says for some particular martingales in some particular Lindeberg protocols.

### Example 1. Bounded Variables with Identical Variances

This is a finite-horizon version of the bounded forecasting protocol we studied in Chapter 3. As always, the parameters are fixed and known to the players before the game begins.

BOUNDED FORECASTING WITH CONSTANT VARIANCE
**Parameters:** $N, A \geq 1, \mathcal{K}_0 > 0$
**Players:** Skeptic, Reality
**Protocol:**
FOR $n = 1, \ldots, N$:
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Reality announces $x_n \in [-AN^{-1/2}, AN^{-1/2}]$.
$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n + V_n(x_n^2 - 1/N). \tag{7.22}$$

Reality may always set $x_n$ equal to $N^{-1/2}$ in absolute value, with sign opposite that of $M_n$, thereby making Skeptic's gain in (7.22) nonpositive. So the protocol is everywhere coherent.

In every situation up to and including $x_1 x_2 \ldots x_{n-1}$, the variable $x_n$ has exact price zero (by setting all his $M_i$ and $V_i$ equal to zero except for $M_n$, which he sets equal to 1, Skeptic gets $x_n$ with an investment of zero) and exact game-theoretic variance $1/N$ (by setting them all equal to zero except for $V_n$, which he sets equal to 1, he gets $x_n^2$ with an investment of $1/N$).

The martingale $\mathcal{S}$ that interests us is the sum of Reality's moves; we set $\mathcal{S}(\Box) := 0$ and $\mathcal{S}(x_1 x_2 \ldots x_n) := \sum_{i=1}^n x_i$. We verify as follows that $\mathcal{S}$ satisfies the hypotheses of Lindeberg's theorem.

- The $x_n$ are the gains of our martingale: $d_{x_1 x_2 \ldots x_{n-1}} \mathcal{S} = x_n$. So
$$\mathbb{V}_{x_1 x_2 \ldots x_{n-1}} d_{x_1 x_2 \ldots x_{n-1}} \mathcal{S} = 1/N$$
for $n = 1, \ldots, N$, and so $\overline{\text{Var}}_{\mathcal{S}}(\xi) = \underline{\text{Var}}_{\mathcal{S}}(\xi) = 1$ for all $\xi$.
- Reality is constrained to keep $x_n^2 \leq A^2/N$. So the inequality $x_n^2 \geq \delta$ is ruled out and $L_{\mathcal{S},\delta}$ is identically zero when $N > A^2/\delta$.

We conclude that $\mathcal{S}_N$ is approximately standard Gaussian when $A$ is fixed and $N$ is made large enough.

For simplicity, we have arranged for the total game-theoretic variance to be one in this example, but we may rescale the protocol to make it any positive number $\sigma^2$, and Lindeberg's theorem will then easily give the conclusion that the sum of Reality's moves will be approximately Gaussian with mean zero and variance $\sigma^2$. We can also allow Forecaster to enter the game and give $x_n$ a possibly nonzero price $m_n$ just before Skeptic moves; as in Chapter 3, this makes no difference. Reality still has the same freedom in choosing the difference $x_n - m_n$, which is all that matters. With these two modifications, the protocol reads as follows:

**Parameters:** $N, A \geq 1, \sigma^2 \geq 0, \mathcal{K}_0 > 0$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
FOR $n = 1, \ldots, N$:
Forecaster announces $m_n \in \mathbb{R}$.
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Reality announces $x_n$ such that $(x_n - m_n) \in [-A\sigma N^{-1/2}, A\sigma N^{-1/2}]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n) + V_n(x_n^2 - \sigma^2/N)$.

The conclusion is that $\mathcal{S}$, the sum of the $x_n - m_n$, is approximately Gaussian with mean zero and variance $\sigma^2$ when $A$ and $\sigma^2$ are fixed and $N$ is made large enough.

It is natural to compare Lindeberg's theorem for this example with the classical central limit theorem for independent identically distributed random variables. We have no probability distribution, but our variables $x_n - m_n$ are identically and independently priced in the sense that they all have the same price (zero) and their squares all have the same price ($\sigma^2/N$) independently of how the preceding ones have come out. On the other hand, we have the additional condition that the variables be uniformly bounded.

The uniform bound and the constant variance are one way of making precise the condition, always needed in one form or another for a central limit theorem, that none of the individual variables contributes too substantially to the total variance.

### Example 2. Bounded Variables with Fixed Variances

Instead of assuming that the $x_n$ all have the same game-theoretic variance, we may assume simply that their game-theoretic variances are fixed and known to all players at the beginning of the game.

BOUNDED FORECASTING WITH FIXED VARIANCES
**Parameters:** $N, A \geq 1, \sigma^2 \geq 0$, nonnegative $v_1, v_2, \ldots, v_N$ adding to $\sigma^2$, $\mathcal{K}_0 > 0$
**Players:** Skeptic, Reality
**Protocol:**
FOR $n = 1, \ldots, N$:
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Reality announces $x_n \in [-A\sqrt{v_n}, A\sqrt{v_n}]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n + V_n(x_n^2 - v_n)$.

If we set $v_n$ equal to $\sigma^2/N$, this reduces to the preceding example.

Our reasoning for this example differs little from that for the preceding example. To see that the protocol is everywhere coherent, we notice that Reality may set $x_n$ equal to $\sqrt{v_n}$ in absolute value, with sign opposite that of $M_n$, thereby making Skeptic's gain nonpositive. And by the same argument as in the preceding example, $x_n$ has exact price zero and exact game-theoretic variance $v_n$ in every situation up to and including $x_1 x_2 \ldots x_{n-1}$.

Again taking $\mathcal{S}$ to be the sum of Reality's moves, we have $\overline{\mathrm{Var}}_{\mathcal{S}}(\xi) = \underline{\mathrm{Var}}_{\mathcal{S}}(\xi) = \sigma^2$ for all $\xi$.

Now fix a positive constant $B$, and let us assume that

$$v_n \leq B\sigma^2/N. \tag{7.23}$$

(Because the $v_n$ must add to $\sigma^2$, $B$ cannot be less than 1.) Because Reality is constrained to keep $x_n^2 \leq A^2 v_n$, this means that $x_n^2 \leq B\sigma^2 A^2/N$. So the inequality $x_n^2 \geq \delta$ is ruled out and $L_{\mathcal{S},\delta}$ is identically zero when $N > B\sigma^2 A^2/\delta$. We conclude by Lindeberg's theorem that $\mathcal{S}_N$ is approximately Gaussian with mean zero and variance $\sigma^2$ when the $v_n$ satisfy (7.23) and $N$ is sufficiently large.

### Example 3. Bounded Variables with Bounded Variances

In this example, we suppose that the game-theoretic variance $v_n$, instead of being fixed in advance of the game, is determined by Forecaster on the $n$th round of play, just before Skeptic moves.

BOUNDED FORECASTING WITH BOUNDED VARIANCES
**Parameters:** $A \geq 1, B \geq 1, \sigma^2 > 0, \mathcal{K}_0 > 0$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
FOR $n = 1, \ldots, N$:
Respecting the constraint $\sum_{i=1}^N v_i = \sigma^2$,
Forecaster announces $v_n \in [0, B\sigma^2/N]$.
Skeptic announces $M_n \in \mathbb{R}$ and a number $V_n \in \mathbb{R}$.
Reality announces $x_n \in [-A\sqrt{v_n}, A\sqrt{v_n}]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n + V_n(x_n^2 - v_n)$.

This protocol is obviously coherent: Forecaster and Reality can prevent Skeptic from increasing his capital by setting all the $v_n$ equal to $\sigma^2/N$ and setting $x_n$ equal in absolute value to $\sigma/\sqrt{N}$, with sign opposite to $M_n$.

In order to fit the example into our formulation of Lindeberg's theorem, we must treat Forecaster's move $v_n$ as part of Reality's move on the preceding round of play, the $(n-1)$st round. This means that we have a 0th round of play in which Reality makes a move $x_0$ and Forecaster makes her first move, $v_1$. The move $x_0$, although irrelevant to Skeptic's capital, becomes part of the record of play, along with the moves by Forecaster. So the situation just before Reality's move $x_n$ is now $x_0 v_1 x_1 v_2 x_2 v_3 \ldots x_{n-1} v_n$ instead of $x_1 x_2 \ldots x_{n-1}$. By our usual argument, we find that the prices of $x_n$ and $x_n^2$ in this situation are 0 and $v_n$, respectively.

As in the preceding examples, we study the martingale $\mathcal{S}_n = \sum_{i=1}^n x_i$. So once again $\overline{\text{Var}}_{\mathcal{S}}(\xi) = \underline{\text{Var}}_{\mathcal{S}}(\xi) = \sigma^2$ for all $\xi$. Forecaster is constrained to keep $v_n \leq B\sigma^2/N$, and Reality is then constrained to keep $x_n^2 \leq A^2 v_n$, and so $x_n^2 \leq B\sigma^2 A^2/N$. So again the inequality $x_n^2 \geq \delta$ is ruled out and $L_{\mathcal{S},\delta}$ is identically zero when or $N > B\sigma^2 A^2/\delta$. We again conclude by Lindeberg's theorem that $\mathcal{S}_N$ has approximately a normal distribution with mean zero and variance $\sigma^2$ when $N$ is large enough—this time large enough relative to $B\sigma^2 A^2$.

## Example 4. Lyapunov's Condition

Before Lindeberg's 1922 article, the weakest conditions under which the central limit theorem was known to hold had been given by Lyapunov. Lyapunov's condition was that the absolute third moments of the variables be bounded. As Lindeberg showed, Lyapunov's theorem followed easily from his own. This is also clear in the game-theoretic context.

In order to express Lyapunov's theorem game-theoretically, we require Forecaster to give upper prices for the $|x_n|^3$. At the beginning of the game, we fix and make known to both players a constant $\epsilon > 0$.

LYAPUNOV'S PROTOCOL
**Parameter:** $N, \epsilon > 0, \mathcal{K}_0 > 0$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
FOR $n = 1, \ldots, N$:
Forecaster announces $v_n \geq 0$ and $w_n \geq 0$.
Skeptic announces $M_n \in \mathbb{R}$, $V_n \in \mathbb{R}$, and $W_n \in \mathbb{R}$.
Reality announces $x_n \in \mathbb{R}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n + V_n(x_n^2 - v_n) + W_n(|x_n|^3 - w_n)$.
**Additional Constraints on Forecaster:** $\sum_{n=1}^N v_n = 1$, $\sum_{n=1}^N w_n \leq \epsilon$.

**Corollary 7.1 (Lyapunov's Theorem)** *Suppose $U$ is a bounded continuous function. If $\epsilon$ is sufficiently small and $N$ is sufficiently large, then the initial upper and lower prices of $U\left(\sum_{n=1}^N x_n\right)$ will both be arbitrarily close to $\int U(z)\,\mathcal{N}_{0,1}(dz)$.*

To derive this as a corollary of Lindeberg's Theorem, notice that

$$|x_n|^2\,\mathbb{I}_{x_n^2 \geq \delta} \leq \delta^{-1/2}|x_n|^3.$$

This, together with $\sum_{n=1}^N w_n \leq \epsilon$, makes it clear that we can satisfy Lindeberg's condition with any $\delta$ taking $\epsilon$ small enough. This is the standard proof of Lyapunov's theorem; see, for example, [287].

This game-theoretic version of Lyapunov's theorem can be strengthened by requiring Skeptic to make $W_n$ nonnegative, but because this produces an asymmetric protocol, the result does not follow from our version of Lindeberg's theorem.

## Example 5. Calibrating the Weather Forecaster

Forecaster gives a probability $p$ that it will rain, Skeptic bets for or against rain at the odds defined by $p$, and then Reality decides whether it rains or not. As we explained in §1.1, this is the prototypical probability game.

Lindeberg's theorem can be used to assess one particular aspect of the Forecaster's $p$s: their *calibration*. Forecaster is said to be calibrated if it rains as often as he leads us to expect. It should rain about 80% of the days for which $p$ is 0.8, about 70% of the days for which $p$ is 0.7, and so on. During a period when $p$ varies, the frequency with which it rains should approximate the average of the $p$s. So in order to judge whether Forecaster is doing as well as he ought, we might monitor him for $N$ days and calculate the difference between the his average probability and the frequency with which it rains. But how should we choose $N$? And how large should the difference be before we reproach Forecaster for his lack of calibration?

We cannot choose $N$ in advance, because we can learn something about Forecaster's calibration only on days when forecasting is difficult. When rain is clearly impossible, Forecaster's $p$ will be 0; when rain is clearly inevitable, his $p$ will be 1. If all the days we monitor are like this, we will learn nothing about how well Forecaster does when forecasting is difficult. We need to wait, however long it takes, until some hard cases come along.

Phil Dawid, in the foreground, with Glenn Shafer on Lake Huron, during an excursion from the Fields Institute, October 1999.

Reality's move on day $n$, $x_n$, is coded as 1 for rain and $-1$ for no rain, then (as we will verify shortly) the game-theoretic variance for $x_n$ is $p_n(1 - p_n)$, where $p_n$ is Forecaster's $p$ for day $n$. This is zero when $p_n$ is 0 or 1. The most telling tests of Forecaster's calibration will come when the game-theoretic variance is relatively large, say when $p_n$ is between 0.2 and 0.8. We do not know in advance how often Forecaster will give forecasts with large $p_n(1 - p_n)$, so we may want to continue our monitoring until the sum of the $p_n(1 - p_n)$ reaches some fixed level that we consider adequate. This leads us to the following protocol.

BINARY PROBABILITY FORECASTING
**Parameters:** $C > 0, \mathcal{K}_0 > 0$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
FOR $n = 1, 2, \ldots$:
Forecaster announces $p_n \in [0, 1]$.
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n)$.
STOP if $\sum_{i=1}^{n} p_i(1 - p_i) \geq C$.

We require that Forecaster eventually satisfy $\sum_{i=1}^{n} p_i(1-p_i) \geq C$ so the game will stop. We assume that $C$ is very large, and since the increments of the sum are always between zero and .25, it will stop at a value relatively quite close to $C$.

It is easy to check that this is a Lindeberg protocol. Forecaster and Reality can keep Skeptic from making money by setting $p_n = 1/2$, $x_n := 0$ when $M_n \geq 0$, and $x_n := 1$ when $M_n < 0$.

In the situation just before Reality announces $x_n$, it has price $p_n$ and game-theoretic variance $p_n(1-p_n)$, just as we expected. It has price $p_n$ because Skeptic has a strategy for getting exactly $x_n$ beginning with capital $p_n$: he sets $M_n$ equal to 1 and his later $M_i$ equal to zero. It has variance $p_n(1-p_n)$ because Skeptic has a strategy for getting exactly $(x_n - p_n)^2$ starting with capital $p_n(1-p_n)$: he sets $M_n$ equal to $1 - 2p_n$ and his later $M_i$ equal to zero. (This returns $(1-2p_n)(x_n - p_n)$, which is equal, because $x_n^2 = x_n$, to $(x_n - p_n)^2 - p_n(1-p_n)$.)

Our test statistic will be $\mathcal{S}_N$, where $\mathcal{S}$ is the martingale defined by

$$\mathcal{S}_n = \frac{1}{\sqrt{C}} \sum_{i=1}^{n} (x_i - p_i), \quad n = 0, 1, \ldots.$$

We have

$$d_{x_1 x_2 \ldots x_{n-1}} \mathcal{S} = \frac{1}{\sqrt{C}} (x_n - p_n).$$

So

$$\mathbb{V}_{x_1 x_2 \ldots x_{n-1}} d_{x_1 x_2 \ldots x_{n-1}} \mathcal{S} = \frac{1}{C} p_n (1 - p_n),$$

and hence

$$\overline{\text{Var}}_{\mathcal{S}}(\xi) = \underline{\text{Var}}_{\mathcal{S}}(\xi) = \frac{1}{C} \sum_{i=1}^{n} p_i(1-p_i),$$

which will be quite close to 1 at the stopping time no matter what path $\xi$ Reality takes. The squared gain $\frac{1}{C}(x_i - p_i)^2$ never exceeds $1/C$. So, provided $C$ is big enough, the inequality $\frac{1}{C}(x_i - p_i)^2 \geq \delta$ is ruled out and $L_{\mathcal{S},\delta}$ is identically zero. So the conditions of Lindeberg's theorem are met.

We conclude that $\mathcal{S}_N$ should have approximately a standard Gaussian distribution. So we will be entitled to question the calibration of Forecaster if $|\mathcal{S}_N|$ exceeds 2 or 3. For further discussion, see Dawid [84, 85] and Dawid and Vovk [86].

## 7.4 APPENDIX: THE CLASSICAL CENTRAL LIMIT THEOREM

As we noted in §2.1, the central limit theorem grew out of Jacob Bernoulli and Abraham De Moivre's investigations of the speed with which the frequency of an event approaches its probability in repeated independent trials, such as the toss of a coin. When a coin is tossed $N$ times, with probability $p$ of heads each time, the probabilities for the number of heads, $y$, are

$$\mathbb{P}\{y = k\} = \frac{N!}{k!(N-k)!} p^k q^{N-k},$$

where $q = 1 - p$. This is the *binomial distribution*. Bernoulli and De Moivre proved their respective theorems by studying the binomial distribution directly [3]. De Moivre's theorem can be written in the form

$$\mathbb{P}\left\{a < \frac{y - Np}{\sqrt{Npq}} < b\right\} \approx \int_a^b \mathcal{N}_{0,1}(dz), \tag{7.24}$$

where $\mathcal{N}_{0,1}$ is the standard Gaussian distribution. De Moivre did not have the idea of a continuous probability distribution, but subsequent work by Laplace and Gauss led to the understanding that (7.24) is an approximation of the binomial, a discrete distribution, by the standard Gaussian, a continuous distribution.

Laplace generalized De Moivre's result to independent identically distributed random variables $x_1, \ldots, x_N$: if $N$ is large enough, then the normalized sum

$$\frac{\sum_{n=1}^{N} x_n - N\mu}{\sqrt{N\sigma^2}},$$

where $\mu$ and $\sigma^2$ are the common mean and variance, will be approximately standard Gaussian (Stigler 1986). Laplace's proof, based on a discrete version of what probabilists now call the characteristic function and other mathematicians now call the Fourier transform, really applied only to the very simple case where the $x_n$ are bounded and take only values that are integer multiples of some small number. The task of perfecting the proof and relaxing its hypotheses was taken up by the Russian school of probability theorists—Pafnutii Chebyshev (1821–1894) and his students Andrei Markov (1856–1922) and Aleksandr Lyapunov (1857–1918). Chebyshev and Markov proved the central limit theorem under conditions weaker than Laplace's using moments, and Lyapunov later proved the same theorems using characteristic functions.

Pierre Simon Laplace (1749–1827), the most illustrious scientist of the golden age of French science.

The nineteenth-century authors did not use the name "central limit theorem". Apparently it was George Pólya (1887–1985), writing in German in 1920, who first spoke of the *zentralen Grenzwertsatz* of probability theory. Since *Grenzwertsatz*, the German equivalent of *limit theorem*, is a single word, Pólya clearly meant that the theorem, not the limit, is central. But some authors in English and French have used the name to underline the fact that the theorem concerns the behavior in the limit of central rather than extreme values of a sum of random variables ([192], p. 79).

Jarl Waldemar Lindeberg (1876–1932), a mathematician at the University of Helsinki, first advanced his method of proof of the central limit theorem in 1920, in an article published in a Finnish journal. He was unaware of Lyapunov's work when he published this article, but he subsequently realized that he could use his method with weaker conditions and derive Lyapunov's theorem as a corollary; he did this in his 1922 article. His method of proof was immediately taken up by Paul Lévy, who used it, together with the characteristic-function method, in his 1925 book.

Lindeberg expressed his result as a limit theorem, applicable to an infinite sequence of independent random variables $x_1, x_2, \ldots$ with means $\mu_1, \mu_2, \ldots$ and variances $\sigma_1^2, \sigma_2^2 \ldots$. He showed that

$$\lim_{N\to\infty} \mathbb{P}\left\{a < \frac{\sum_{n=1}^N (x_n - \mu_n)}{\sqrt{\sum_{n=1}^N \sigma_n^2}} < b\right\} = \int_a^b \mathcal{N}_{0,1}(dz) \tag{7.25}$$

if

$$\lim_{N\to\infty} \sum_{i=1}^N \mathbb{E}\left[\left(\frac{x_i - \mu_i}{\sqrt{\sum_{n=1}^N \sigma_n^2}}\right)^2 \mathbb{I}_{\frac{|x_i-\mu_i|}{\sqrt{\sum_{n=1}^N \sigma_n^2}} \geq \epsilon}\right] = 0 \tag{7.26}$$

for every positive number $\epsilon$. The condition (7.26) says that none of the $|x_n - \mu_n|$ are likely to be very large relative to $\sqrt{\sum_{n=1}^N \sigma_n^2}$, the standard deviation of their sum; each makes an individually negligible contribution to the sum. Lindeberg's result is definitive for independent random variables because, as Paul Lévy and William Feller both showed in 1935 [122, 198, 192], (7.26) is roughly necessary as well as sufficient for (7.25). The statement is only roughly true because some of the variables in the sum can be themselves exactly Gaussian, and such summands can have any variances whatsoever ([287], III.5; [356], Theorem 5.2.5). We can make a precise statement by considering the condition

$$\lim_{N\to\infty} \max_{i\leq N} \left(\frac{\sigma_i^2}{\sum_{n=1}^N \sigma_n^2}\right) = 0, \tag{7.27}$$

which is implied by (7.26). It turns out that (7.26) holds if and only if both (7.27) and (7.25) hold. So when (7.27) is assumed, we can say that Lindeberg's condition, (7.26), is necessary and sufficient for the central limit theorem, (7.25), for independent random variables.

Beginning with further work by Lévy in the 1930s, work on the central limit theorem shifted to the martingale case, where Lindeberg's condition is also applicable [144, 206]. Lévy also initiated, in 1937, work on versions of the central limit theorem that do not even require the existence of the variances $\sigma_1^2, \sigma_2^2 \ldots$. These versions justify a very general qualitative statement: a sum of independent variables, appropriately centered and normalized, will be nearly Gaussian if and only if each variable has a dispersion small relative to the dispersion of the sum or is itself nearly Gaussian ([192], Theorem 2; [356], Theorem 5.2.5).

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 8

# *The Generality of Probability Games*

In preceding chapters, we demonstrated the value of the game-theoretic framework by example. We looked at probability's classical limit theorems one by one, showing in each case how well the game-theoretic framework captures the theorem's classical intuition and generalizes its scope. In this chapter, we study the effectiveness of the game-theoretic framework in more abstract ways.

We begin, in §8.1, by showing that the game-theoretic framework is strictly more powerful, in at least one important respect, than the measure-theoretic framework. The game-theoretic limit theorems that we proved in the preceding chapters actually imply the corresponding measure-theoretic versions of these theorems (Kolmogorov's strong law, the law of the iterated logarithm, and Lindeberg's theorem). This is because the game-theoretic limit theorems all assert the existence of certain martingales, which are constructed from other martingales in a computable and hence Borel measurable way. These constructions can also be applied to martingales in the measure-theoretic framework, and this yields the corresponding measure-theoretic limit theorems. But there is no easy way to go in the opposite direction, from the measure-theoretic to the game-theoretic limit theorems.

In §8.2, we look at how the game-theoretic framework deals with the historical and conceptual kernel of probability theory: coin tossing. Remarkably, the game-theoretic and measure-theoretic treatments of coin tossing have much in common: the same sample space, the same martingales, and the same probabilities. So we can say that the two frameworks represent two different ways of generalizing a common classical kernel.

In §8.3, we return to the abstract study of game-theoretic price and probability that we began in §1.2 and §1.3 and continued in §7.1. One point of this section is to reassure those readers who think of Kolmogorov's axioms and definitions as the essence of probability theory that similar and equally abstract properties can be formulated and derived in the game-theoretic framework. These include the rule of iterated expectation and linearity (or sublinearity or superlinearity) of price.

In §8.4, we underline the practical importance of the generality of the game-theoretic framework by looking at two scientifically important probability models that are easily understood in terms of probability protocols but do not fit into measure theory: quantum mechanics and the failure model in medical statistics. These models are probabilistic but open; they model processes that are open to external and unmodeled influences. Probabilities are determined as such a process proceeds, but not by the model alone; they are also influenced by factors outside the process. In the end, the model and these outside influences have determined a sequence of probabilities (or even probability distributions) for the successive steps, but no probability measure for the process as a whole is ever determined.

In an appendix, §8.5, we state and prove Ville's theorem, which relates probabilities in a filtered probability space to martingales. This theorem is used in §8.1 and §8.2. In a second appendix, §8.6, we provide some biographical information about Jean Ville.

## 8.1 DERIVING THE MEASURE-THEORETIC LIMIT THEOREMS

In this section we derive the three most important measure-theoretic limit theorems—Kolmogorov's strong law of large numbers, the law of the iterated logarithm, and Lindeberg's theorem—from the game-theoretic versions that we proved in Chapters 4, 5, and 7, respectively.

The game-theoretic strong laws assert that Skeptic can force certain events in the unbounded forecasting game. Kolmogorov's strong law says that Skeptic can force the convergence of a certain average deviation to zero. The law of the iterated logarithm asserts that he can force the oscillation during the convergence to have certain properties. Here we restate these two game-theoretic results so that they also assert the Borel measurability of the strategies that enable Skeptic to force these events. The measure-theoretic versions of the two strong laws then follow easily.

Our derivation of the measure-theoretic version of Lindeberg's theorem from the game-theoretic version is similar; we begin by restating the game-theoretic version so that it asserts the measurability of what Skeptic does, and then we derive the measure-theoretic version from this restatement. The argument is more complicated than in the case of the strong laws, however, because the game-theoretic version of Lindeberg's theorem begins with a game-theoretic martingale $\mathcal{S}$ that need not be measurable and constructs a strategy for Skeptic from $\mathcal{S}$ and from martingales witnessing certain conditions on $\mathcal{S}$. Here it is the transformation, not the strategy itself, that is computable and therefore Borel measurable.

In the rest of this section we emphasize measurability rather than computability, but this should not be misunderstood. In our view, measurability has no foundational role within the game-theoretic framework per se. We introduce it here only to make the relation between the game-theoretic and measure-theoretic frameworks as clear as possible for readers already trained in measure theory. From the game-theoretic point of view, the important point about the strategies constructed in Chapters 4 and 5 and the martingale transformation constructed in Chapter 7 is that they are computable. Computability is obviously important for applications, and it raises many interesting questions for further research, as we noted in §3.5. Computable functions and transformations are necessarily Borel measurable, but measurability by itself does not accomplish anything within the game-theoretic framework.

## Filtrations and Measure-Theoretic Martingales

Before we look at the limit theorems, we must first extend the review of measure-theoretic terminology and notation that we began in §2.2.

A *measurable space* is a nonempty set $\Omega$ together with a $\sigma$-algebra $\mathcal{F}$ of subsets of $\Omega$. A subset of $\Omega$ is *measurable* if it is in $\mathcal{F}$. Elements of $\mathcal{F}$ are also called *events*. A real-valued function $x$ on $\Omega$ is *measurable* if $\{\xi \mid x(\xi) \leq a\}$ is measurable for every real number $a$.

A *filtration* in a measurable space $(\Omega, \mathcal{F})$ is a sequence $\{\mathcal{F}_n\}_{n=0}^{\infty}$ of successively larger $\sigma$-algebras all contained in $\mathcal{F}$: $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$. We write $\mathcal{F}_\infty$ for the smallest $\sigma$-algebra containing all the $\mathcal{F}_n$.

A *filtered measurable space* is a measurable space with a filtration. A sequence $x_1, x_2, \dots$ of functions on a filtered measurable space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty})$ is *adapted* if $x_n$ is measurable with respect to $\mathcal{F}_n$ for $n = 1, 2, \dots$; it is *predictable* if $x_n$ is measurable with respect to $\mathcal{F}_{n-1}$ for $n = 1, 2, \dots$. A sequence $\mathcal{V}_0, \mathcal{V}_1, \dots$ of functions is a *process* if $\mathcal{V}_n$ is measurable with respect to $\mathcal{F}_n$ for $n = 0, 1, \dots$. If $\mathcal{V}_1, \mathcal{V}_2, \dots$ is a predictable sequence, we say the process is *predictable*. A function $\tau$ is a *stopping time* if (1) $\tau(\omega)$ is a natural number for all $\omega \in \Omega$ and (2) $\{\omega \mid \tau(\omega) = n\} \in \mathcal{F}_n$ for $n = 1, 2, \dots$.

A *probability space* is a measurable space $(\Omega, \mathcal{F})$ together with a probability measure $\mathbb{P}$ on $\mathcal{F}$. A *filtered probability space* is a probability space with a filtration. It is customary to assume that the $\sigma$-algebra $\mathcal{F}_0$ in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ contains all of $\mathbb{P}$'s null events—that is, every $E \in \Omega$ such that $\mathbb{P}\,E = 0$, but we will sometimes find it convenient to assume instead that $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

A measurable function on a probability space is called a *random variable*. Given a random variable $x$ whose expected value exists and a $\sigma$-algebra $\mathcal{G}$ contained in $\mathcal{F}$, there exists at least one random variable $y$ that is measurable with respect to $\mathcal{G}$ and satisfies $\mathbb{E}[x\,\mathbb{I}_E] = \mathbb{E}[y\,\mathbb{I}_E]$ for every $E$ in $\mathcal{G}$. Any such random variable $y$ is called a *version* of the *conditional expectation* of $x$ with respect to $\mathcal{G}$. Any two versions of the same conditional expectation are equal except on a set of measure zero.

It is customary to write $\mathbb{E}[x \mid \mathcal{G}] = y$ to indicate that $y$ is a version of the conditional expectation of $x$ with respect to $\mathcal{G}$. More generally, $\mathbb{E}[x \mid \mathcal{G}]$ can represent an arbitrary version of the conditional expectation in an equation or inequality that is stated to hold almost surely. This does not entail any ambiguity, because the equation or inequality will hold almost surely for one version if and only if it holds almost surely for a different one.[1] Conditional variance is handled in the same way; one writes $\mathbb{V}[x \mid \mathcal{G}]$ for $\mathbb{E}[(x - \mathbb{E}[x \mid \mathcal{G}])^2 \mid \mathcal{G}]$.

A process $\mathcal{L}_0, \mathcal{L}_1, \ldots$ in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ is a *measure-theoretic martingale* if

$$\mathbb{E}[\mathcal{L}_n \mid \mathcal{F}_{n-1}] = \mathcal{L}_{n-1} \tag{8.1}$$

for $n = 1, 2, \ldots$. If $\mathcal{F}_0 = \{\emptyset, \Omega\}$, then the initial variable $\mathcal{L}_0$ will necessarily be a constant. Equation (8.1) says that $\mathcal{L}_{n-1}$ is a version of the conditional expectation of $\mathcal{L}_n$ given $\mathcal{F}_{n-1}$.

### The Game-Theoretic Strong Laws with Measurability

Our game-theoretic versions of Kolmogorov's strong law and the law of the iterated logarithm involve no conditions of measurability. But as we noted at the beginning of this section, the strategies and capital processes that we constructed to prove these theorems are measurable in the sense required. In order to make the theorems comparable to measure-theoretic theorems, we now restate them in a way that makes this measurability explicit.

Consider the unbounded forecasting protocol (p. 79), which we used for Kolmogorov's strong law of large numbers. In this protocol, Forecaster makes moves $m_n \in \mathbb{R}$ and $v_n \geq 0$, then Skeptic makes moves $M_n \in \mathbb{R}$ and $V_n \geq 0$, and then Reality makes a move $x_n \in \mathbb{R}$. A strategy for Skeptic is therefore a pair of processes $\mathcal{M}_1, \mathcal{M}_2, \ldots$ and $\mathcal{V}_1, \mathcal{V}_2, \ldots$ such that for each $n$, $\mathcal{M}_n$ and $\mathcal{V}_n$ are both functions of $m_1, v_1, x_1, \ldots, m_{n-1}, v_{n-1}, x_{n-1}, m_n, v_n$. Mathematically, $\mathcal{M}_n$ and $\mathcal{V}_n$ can be regarded as real-valued functions of $3n - 1$ real variables. If these functions are all Borel measurable, then we say that the strategy is *Borel measurable.*

If $E$ is an event in the protocol, and Skeptic has a Borel measurable winning strategy in the game in which his goal is $E$, then we say that Skeptic *can Borel force $E$.* The strategy we constructed for Skeptic in §4.2, being the result of simple arithmetic and limiting processes, obviously is Borel measurable, and hence we may strengthen Statement 1 of Theorem 4.1 to the following:

**Proposition 8.1** *Skeptic can Borel force*

$$\sum_{n=1}^{\infty} \frac{v_n}{n^2} < \infty \Longrightarrow \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) = 0 \tag{8.2}$$

*in the unbounded forecasting protocol.*

We can similarly restate Theorems 5.1 and 5.2, which together form our game-theoretic version of the law of the iterated logarithm:

[1] We hasten to remind the reader that this practice does not carry over to the game-theoretic framework. In the game-theoretic framework, a symbol for a variable always refers to a single particular variable, never to a class of variables every two members of which agree only almost surely.

**Proposition 8.2** *Skeptic can Borel force*

$$\left(A_n \to \infty \;\&\; c_n = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln\ln A_n}} = 1$$

*in the predictably unbounded forecasting protocol, where* $A_n := \sum_{i=1}^n v_i$.

**Proposition 8.3** *Skeptic can Borel force*

$$\left(A_n \to \infty \;\&\; |x_n - m_n| = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^n (x_i - m_i)}{\sqrt{2A_n \ln\ln A_n}} \le 1$$

*in the unbounded forecasting protocol, where, again,* $A_n := \sum_{i=1}^n v_i$.

In general, the capital process for Skeptic resulting from an initial capital and a particular strategy in the unbounded or predictably unbounded forecasting protocol can be thought of as a sequence of variables $\mathcal{L}_0, \mathcal{L}_1, \ldots$. Here $\mathcal{L}_0$ is the initial capital and

$$\begin{aligned} \mathcal{L}_n(m_1 v_1 x_1 \ldots m_n v_n x_n) &:= \mathcal{L}_{n-1}(m_1 v_1 x_1 \ldots m_{n-1} v_{n-1} x_{n-1}) \\ &\quad + \mathcal{M}_n(m_1 v_1 x_1 \ldots m_n v_n)(x_n - m_n) \\ &\quad + \mathcal{V}_n(m_1 v_1 x_1 \ldots m_n v_n)((x_n - m_n)^2 - v_n) \end{aligned} \tag{8.3}$$

is the capital at the end of the $n$th round of play. The function $\mathcal{L}_n$ specified by (8.3) is a very simple—certainly measurable—function of the moves by Forecaster and Reality. So if Skeptic's strategy $(\mathcal{M}_1, \mathcal{V}_1), (\mathcal{M}_2, \mathcal{V}_2), \ldots$ is Borel measurable, and if all the $m_n$, $v_n$, and $x_n$ are taken to be measurable functions on some measurable space $(\Omega, \mathcal{F})$,[2] then $\mathcal{L}_0, \mathcal{L}_1, \ldots$ will also become measurable functions on $(\Omega, \mathcal{F})$. In fact, if $\{\mathcal{F}_n\}_{n=0}^\infty$ is a filtration in $(\Omega, \mathcal{F})$, and the functions $m_n$, $v_n$, and $x_n$ are measurable with respect to $\mathcal{F}_n$ for each $n$, then $\mathcal{L}_n$ will also be measurable with respect to $\mathcal{F}_n$ for each $n$.[3]

## Deriving the Measure-Theoretic Strong Laws

The following corollary of Proposition 8.1 is the measure-theoretic form of Kolmogorov's strong law.

[2] Here $(\Omega, \mathcal{F})$ is an arbitrary measurable space. We are not using the symbol $\Omega$ to designate the sample space for the probability protocol, as we usually do.

[3] We can make this mathematical observation more vivid by imagining that Forecaster and Reality choose a particular element $\omega$ of $\Omega$ at the beginning of the game, behind Skeptic's back, and then simply make the moves specified by the measurable functions: $m_1(\omega), v_1(\omega), m_2(\omega), v_2(\omega), \ldots$ for Forecaster and $x_1(\omega), x_2(\omega), \ldots$ for Reality.

**Corollary 8.1** *If $x_1, x_2, \ldots$ is an adapted sequence of random variables in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$, then*

$$\sum_{n=1}^{\infty} \frac{\mathbb{V}[x_n \mid \mathcal{F}_{n-1}]}{n^2} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - \mathbb{E}[x_i \mid \mathcal{F}_{i-1}]) = 0 \qquad (8.4)$$

*almost surely.*[4]

*Proof* According to Proposition 8.1, the player Skeptic in the unbounded forecasting game has a Borel measurable strategy $(\mathcal{M}_1, \mathcal{V}_1), (\mathcal{M}_2, \mathcal{V}_2), \ldots$ for which the capital process $\mathcal{L}_0, \mathcal{L}_1, \ldots$, given by (8.3), is nonnegative no matter how Forecaster and Reality move and diverges to infinity if (8.2) fails. Fix versions of the conditional expectations $\mathbb{E}(x_n \mid \mathcal{F}_{n-1})$ and then of the conditional variances $\mathbb{V}(x_n \mid \mathcal{F}_{n-1})$ (making sure the latter are nonnegative), and substitute them for $m_n$ and $v_n$ in (8.3). Similarly, substitute the random variable $x_n$ for the move $x_n$ in (8.3). As we explained in our comments following (8.3), the resulting function $\mathcal{L}_n$ on $\Omega$ is measurable with respect to $\mathcal{F}_n$. Similarly, $\mathcal{M}_n$ and $\mathcal{V}_n$ become functions on $\Omega$ measurable with respect to $\mathcal{F}_{n-1}$, and we can rewrite (8.3) in the form

$$\mathcal{L}_n := \mathcal{L}_{n-1} + \mathcal{M}_n(x_n - \mathbb{E}[x_n \mid \mathcal{F}_{n-1}]) + \mathcal{V}_n((x_n - \mathbb{E}[x_n \mid \mathcal{F}_{n-1}])^2 - \mathbb{V}[x_n \mid \mathcal{F}_{n-1}]).$$

This implies (8.1). So $\mathcal{L}_0, \mathcal{L}_1, \ldots$ is a measure-theoretic martingale. It is nonnegative and diverges to infinity if (8.4) fails. But by Doob's martingale convergence theorem ([287], Theorem VII.4.1), a nonnegative measure-theoretic martingale diverges to infinity with probability zero. (This is one part of what we call Ville's theorem; see Proposition 8.14 on p. 196.) So (8.4) happens almost surely. ∎

Similarly, Propositions 8.2 and 8.3 have the following corollaries, which together form a slightly stronger-than-standard version of the measure-theoretic law of the iterated logarithm.

**Corollary 8.2** *If $x_n$ is an adapted sequence of random variables and $c_n$ a predictable sequence of random variables in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$, and*

$$|x_n - \mathbb{E}[x_n \mid \mathcal{F}_{n-1}]| \le c_n$$

*for all n, then*

$$\left(A_n \to \infty \;\&\; c_n = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^{n}(x_i - \mathbb{E}[x_i \mid \mathcal{F}_{i-1}])}{\sqrt{2A_n \ln\ln A_n}} = 1$$

*almost surely, where $A_n := \sum_{i=1}^{n} \mathbb{V}[x_i \mid \mathcal{F}_{i-1}]$.*

[4]The conclusion is true even if some or all of the $x_n$ fail to have (finite) conditional expected values or variances on a set of $\omega \in \Omega$ of a positive measure. At such $\omega$ the conditional variance $\mathbb{V}[x_n \mid \mathcal{F}_{n-1}]$ does not exist or is infinite, and so the left-hand side of the implication (8.4) fails and thus the implication itself happens (remember that $(A \Longrightarrow B) = A^c \cup B$).

**Corollary 8.3** *If $x_n$ is an adapted sequence of random variables in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$, then*

$$\left(A_n \to \infty \;\&\; |x_n - \mathbb{E}[x_n \mid \mathcal{F}_{n-1}]| = o\left(\sqrt{\frac{A_n}{\ln\ln A_n}}\right)\right)$$
$$\Longrightarrow \limsup_{n\to\infty} \frac{\sum_{i=1}^{n}(x_i - \mathbb{E}[x_i \mid \mathcal{F}_{i-1}])}{\sqrt{2A_n \ln\ln A_n}} \leq 1$$

*almost surely, where, again, $A_n := \sum_{i=1}^{n} \mathbb{V}[x_i \mid \mathcal{F}_{i-1}]$.*

The first of these, Corollary 8.2, is the sharpness part of the law of the iterated logarithm as proven by William F. Stout in 1970 (see also his 1973 book). The second, Corollary 8.3, strengthens the validity part of Stout's theorem.

### The Game-Theoretic Form of Lindeberg's Theorem with Measurability

We now explain how to add measurability to the conclusion of Theorem 7.1, our game-theoretic version of Lindeberg's theorem.

Recall that a probability protocol is a *Lindeberg protocol* if it is symmetric, coherent, and terminating. In a terminating protocol, a process $\mathcal{L}$ eventually has a final value, which we designate by $\mathcal{L}_\Omega$. We say that a game-theoretic martingale $\mathcal{L}$ in a Lindeberg protocol *witnesses* the relation $\overline{\mathbb{E}}_\Box\, x < \alpha$, where $x$ is a variable and $\alpha$ is a number, if $\mathcal{L}(\Box) < \alpha$ and $\mathcal{L}_\Omega(\omega) \geq x$ for all $\omega \in \Omega$; and we say that $\mathcal{L}$ *witnesses* the relation $\overline{\mathbb{P}}_\Box\, E < \alpha$, where $E$ is an event, if it witnesses $\overline{\mathbb{E}}_\Box\, \mathbb{I}_E < \alpha$.

Theorem 7.1 considers certain conditions on a game-theoretic martingale $\mathcal{S}$ in a Lindeberg protocol. Under these conditions, the Gaussian integral $\int U(z)\, \mathcal{N}_{0,1}(dz)$ is an approximate initial price for the payoff $U(\mathcal{S}_\Omega)$, where $U$ is a bounded continuous function. Our proof, in §7.2, constructed a witnessing game-theoretic martingale. Our restatement will point out that the witnessing martingale is obtained from $\mathcal{S}$ and other game-theoretic martingales in the problem (game-theoretic martingales that witness the conditions on $\mathcal{S}$) by a measurable martingale transformation.

In general, we can construct a new martingale $\mathcal{L}$ from martingales $\mathcal{S}^1, \ldots, \mathcal{S}^K$ by selecting an initial value for $\mathcal{L}$ and setting each increment of $\mathcal{L}$ equal to a linear combination of the corresponding increments of the $\mathcal{S}^k$, say

$$\Delta\mathcal{L}_n = c_n^1 \Delta\mathcal{S}_n^1 + c_n^2 \Delta\mathcal{S}_n^2 + \cdots + c_n^K \Delta\mathcal{S}_n^K. \tag{8.5}$$

The $c_n^k$ may depend on the previous values of the $\mathcal{S}^k$ and also on the current and previous values of other processes, provided the additional processes are predictable. Each such process has $n$ previous values (say $\mathcal{V}_0 \ldots, \mathcal{V}_{n-1}$) in addition to its current value (say $\mathcal{V}_n$). So we can formulate the following definitions:

- A $(K, K')$-*martingale transformation* $\mathbf{C}$ consists of functions $\mathbf{c}_n^k$, for $k = 1, \ldots, K$ and $n = 1, 2, \ldots$, where $\mathbf{c}_n^k$ is a real-valued function of $nK + (n+1)K'$ real variables.

- A $(K, K')$-martingale transformation $\mathbf{C}$ is *measurable* if all its $\mathbf{c}_n^k$ are Borel measurable.
- A process $\mathcal{L}$ is a *martingale transform* of martingales $\mathcal{S}^1, \ldots, \mathcal{S}^K$ given auxiliary predictable processes $\mathcal{R}^1, \ldots, \mathcal{R}^{K'}$ if there exists a $(K, K')$-martingale transformation $\mathbf{C}$ such that (8.5) is satisfied, with the coefficient $c_n^k$ given by applying $\mathbf{c}_n^k$ to the previous values of $\mathcal{S}^1, \ldots, \mathcal{S}^K$ and current and previous values of $\mathcal{R}^1, \ldots, \mathcal{R}^{K'}$.

These definitions can be used both for game-theoretic martingales (in a given symmetric probability protocol, where the predictability of $\mathcal{R}^k$ means that $\mathcal{R}_n^k$ is a function of World's first $n-1$ moves) and for measure-theoretic martingales (in a given filtered probability space, where the predictability of $\mathcal{R}^k$ means that $\mathcal{R}_n^k$ is measurable with respect to $\mathcal{F}_{n-1}$). In a symmetric probability protocol, a martingale transform $\mathcal{L}$ is always a game-theoretic martingale, because if $\mathcal{S}^1, \ldots, \mathcal{S}^K$ are game-theoretic martingales, then (8.5) is enough to ensure that $\mathcal{L}$ will be as well. In a filtered probability space, on the other hand, we need a bit more. In order to be sure that $\mathcal{L}_n$ is $\mathcal{F}_n$-measurable, we need to know that the martingale transformation $\mathbf{C}$ is measurable.

In the preceding paragraph, a process $\mathcal{V}$ is represented as a sequence of variables, $\mathcal{V}_0, \mathcal{V}_1, \ldots$. This permits us to write $\Delta\mathcal{V}_n$ for the difference $\mathcal{V}_n - \mathcal{V}_{n-1}$ and $d\mathcal{V}_n$ for the difference $\mathcal{V}_{n+1} - \mathcal{V}_n$. In the game-theoretic framework, especially in the case of a terminating protocol, we often use instead a notation that refers to a specific situation $t$; we write $\mathcal{V}(t)$ for $\mathcal{V}$'s value in $t$, and we write $\Delta_t\mathcal{V}$ (or $d_t\mathcal{V}$) for the increment of $\mathcal{V}$ immediately before (or after) $t$ (see p. 152). We may relate the two notations by writing $t_n$ for the $n$th situation (counting $\Box$ as the 0th situation); then

$$\mathcal{V}_n = \mathcal{V}(t_n), \quad \Delta\mathcal{V}_n = \Delta_{t_n}\mathcal{V}, \quad \text{and} \quad d\mathcal{V}_n = d_{t_n}\mathcal{V}. \tag{8.6}$$

The identity of the $n$th situation depends, of course, on the path $\omega$. If we write $t_n(\omega)$ to make this explicit, then (8.6) becomes

$$\mathcal{V}_n(\omega) = \mathcal{V}(t_n(\omega)), \quad \Delta\mathcal{V}_n(\omega) = \Delta_{t_n(\omega)}\mathcal{V}, \quad \text{and} \quad d\mathcal{V}_n(\omega) = d_{t_n(\omega)}\mathcal{V}.$$

This makes it clear that $\mathcal{V}_n$ and $\Delta\mathcal{V}_n$ only have a meaning for a path $\omega$ on which there are at least $n$ situations after $\Box$, and $d\mathcal{V}_n$ only has a meaning for a path $\omega$ on which there are at least $n+1$ situations after $\Box$.

Consider now a game-theoretic martingale $\mathcal{S}$ in a Lindeberg protocol. If

$$\underline{\mathbb{E}}_t(d_t\mathcal{S})^2 = \overline{\mathbb{E}}_t(d_t\mathcal{S})^2$$

for every nonterminal situation $t$, then we call $\mathcal{S}$ *scaled* (or *0-scaled*). If $\mathcal{S}$ is scaled, we define a process $\mathcal{A}$, the game-theoretic *quadratic variation* of $\mathcal{S}$, by setting

$$\mathcal{A}(t) := \sum_{s \sqsubset t} \mathbb{E}_s(d_s\mathcal{S})^2 \tag{8.7}$$

for every situation $t$ (so that $\mathcal{A}(\Box) = 0$), and we define a martingale $\mathcal{B}$ by setting $\mathcal{B}(\Box) := 0$ and

$$d_t\mathcal{B} := (d_t\mathcal{S})^2 - \mathbb{E}_t(d_t\mathcal{S})^2 \tag{8.8}$$

for every nonterminal situation $t$. The current value of $\mathcal{B}$ is always determined by current and past values of $\mathcal{S}$ and $\mathcal{A}$.

If $\delta > 0$ and

$$\underline{\mathbb{E}}_t \left[(d_t\mathcal{S})^2 \mathbb{I}_{|d_t\mathcal{S}| \geq \delta}\right] = \overline{\mathbb{E}}_t \left[(d_t\mathcal{S})^2 \mathbb{I}_{|d_t\mathcal{S}| \geq \delta}\right]$$

for every nonterminal situation $t$, then we call $\mathcal{S}$ *$\delta$-scaled.* If $\mathcal{S}$ is $\delta$-scaled, we define a process $\mathcal{C}$ by setting

$$\mathcal{C}(t) := \sum_{s \sqsubset t} \mathbb{E}_s \left[(d_s\mathcal{S})^2 \mathbb{I}_{|d_s\mathcal{S}| \geq \delta}\right] \tag{8.9}$$

for every situation $t$ (so $\mathcal{C}(\Box) = 0$), and we define a martingale $\mathcal{U}$ by setting $\mathcal{U}(\Box) := 0$ and

$$d_t\mathcal{U} := (d_t\mathcal{S})^2 \mathbb{I}_{|d_t\mathcal{S}| \geq \delta} - \mathbb{E}_t \left[(d_t\mathcal{S})^2 \mathbb{I}_{|d_t\mathcal{S}| \geq \delta}\right] \tag{8.10}$$

for every nonterminal situation $t$. The current value of $\mathcal{U}$ is always determined by current and past values of $\mathcal{S}$ and $\mathcal{C}$.

The proof of Theorem 7.1 in §7.2 involved the construction of a measurable martingale transformation. We now restate the theorem to make this explicit. For the sake of clarity, we simplify a bit as we do so; we consider only scaled and $\delta$-scaled martingales.

**Proposition 8.4** *Suppose $U$ is a bounded continuous function, and suppose $\epsilon > 0$. Then there exists $\delta > 0$ and a measurable $(4,2)$-martingale transformation* **C** *such that if $\mathcal{S}$ is a scaled and $\delta$-scaled game-theoretic martingale in a Lindeberg protocol and*

$$\overline{\mathbb{P}}_{\Box} \{|\mathcal{A}_\Omega - 1| > \delta \text{ or } \mathcal{C}_\Omega > \delta\} < \delta, \tag{8.11}$$

*then*

$$\overline{\mathbb{E}}_{\Box} [U(\mathcal{S}_\Omega)] < \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) + \epsilon \tag{8.12}$$

*is witnessed by the game-theoretic martingale obtained by applying* **C** *to $\mathcal{S}$, $\mathcal{B}$, $\mathcal{U}$, and $\mathcal{W}$ given $\mathcal{A}$ and $\mathcal{C}$, where $\mathcal{W}$ is any game-theoretic martingale witnessing (8.11), and $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{U}$ are defined by (8.7), (8.8), (8.9), and (8.10), respectively.*

The restriction to game-theoretic martingales that are scaled and $\delta$-scaled makes Proposition 8.4 less general than Theorem 7.1. This restriction is not necessary, but it simplifies the statement of the proposition considerably, and the result is still sufficient for the derivation of the theorem's measure-theoretic counterpart.

When the conditions of Proposition 8.4 are satisfied by $\mathcal{S}$, they are also satisfied by $-\mathcal{S}$, and when the proposition is applied to $-\mathcal{S}$ and to the bounded continuous function $-U(-s)$, it yields a game-theoretic martingale witnessing

$$\underline{\mathbb{E}}_{\Box} [U(\mathcal{S}_\Omega)] > \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) - \epsilon.$$

So the final conclusion is the same as in Theorem 7.1: The initial upper and lower prices of $U(\mathcal{S}_\Omega)$ are both close to $\int U(z)\ \mathcal{N}_{0,1}(dz)$.

### Deriving the Measure-Theoretic Form of Lindeberg's Theorem

We can now derive the measure-theoretic form of Lindeberg's theorem (see, e.g., [206], Theorem 5.5.9) as a corollary of Proposition 8.4.

We begin by associating processes $\mathcal{A}$ and $\mathcal{C}$ with any measure-theoretic martingale $\mathcal{S}$ in a filtered probability space. For $n = 0, 1, \ldots$, we set

$$\mathcal{A}_n := \sum_{i=1}^{n} \mathbb{E}[(\Delta \mathcal{S}_i)^2 \mid \mathcal{F}_{i-1}]$$

and

$$\mathcal{C}_n := \sum_{i=1}^{n} \mathbb{E}[(\Delta \mathcal{S}_i)^2 \mathbb{I}_{|\Delta \mathcal{S}_i| \geq \delta} \mid \mathcal{F}_{i-1}].$$

The process $\mathcal{C}$ depends, of course, on the choice of the positive constant $\delta$. The process $\mathcal{A}$ is the measure-theoretic *quadratic variation* of $\mathcal{S}$.

**Corollary 8.4** *Suppose $U$ is a bounded continuous function, and suppose $\epsilon > 0$. Then there exists $\delta > 0$ such that if $\mathcal{S}$ is a measure-theoretic martingale and $\tau$ is a stopping time in a filtered probability space, and*

$$\mathbb{P}\{|\mathcal{A}_\tau - 1| > \delta \text{ or } \mathcal{C}_\tau > \delta\} < \delta, \tag{8.13}$$

*then*

$$\mathbb{E}[U(\mathcal{S}_\tau)] < \int_{-\infty}^{\infty} U(z)\ \mathcal{N}_{0,1}(dz) + \epsilon. \tag{8.14}$$

*Proof* Consider the number $\delta > 0$ and the measurable martingale transformation $\mathbf{C}$ given for $\epsilon$ and $U$ by Proposition 8.4. Fix a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$, a stopping time $\tau$, and a measure-theoretic martingale $\mathcal{S}$ satisfying (8.13).

By Ville's theorem (Proposition 8.13), there exists a measure-theoretic martingale $\mathcal{W}$ that witnesses (8.13)—that is, for which $\mathcal{W}_0 < \delta$ and $\mathcal{W}_\tau$ exceeds the indicator function for the event $\{|\mathcal{A}_\tau - 1| > \delta \text{ or } \mathcal{C}_\tau > \delta\}$. Fix $\mathcal{W}$. Define measure-theoretic martingales $\mathcal{B}$ and $\mathcal{U}$ by setting $\mathcal{B}_0$ and $\mathcal{U}_0$ equal to zero and setting

$$\Delta \mathcal{B}_n := (\Delta \mathcal{S}_n)^2 - \mathbb{E}[(\Delta \mathcal{S}_n)^2 \mid \mathcal{F}_{n-1}]$$

and

$$\Delta \mathcal{U}_n := (\Delta \mathcal{S}_n)^2 \mathbb{I}_{|\Delta \mathcal{S}_n| \geq \delta} - \mathbb{E}[(\Delta \mathcal{S}_n)^2 \mathbb{I}_{|\Delta \mathcal{S}_n| \geq \delta} \mid \mathcal{F}_{n-1}]$$

for $n = 1, 2, \ldots$. Let $\mathcal{L}$ be the measure-theoretic martingale obtained by applying $\mathbf{C}$ to $\mathcal{S}$, $\mathcal{B}$, $\mathcal{U}$, and $\mathcal{W}$ given $\mathcal{A}$ and $\mathcal{C}$.

We now define a Lindeberg protocol from the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$, the stopping time $\tau$, and the four processes $\mathcal{S}$, $\mathcal{W}$, $\mathcal{A}$, and $\mathcal{C}$. First we set

$$\mathbf{w}_n(\omega) := (\mathcal{S}_n(\omega), \mathcal{W}_n(\omega), \mathcal{A}_n(\omega), \mathcal{C}_n(\omega), \mathbb{I}_{\tau(\omega)=n}) \tag{8.15}$$

for each $\omega \in \Omega$. Then we set

$$\Omega' := \{\mathbf{w}_1(\omega), \mathbf{w}_2(\omega), \ldots, \mathbf{w}_{\tau(\omega)}(\omega) \mid \omega \in \Omega\};$$

this is the sample space of the Lindeberg protocol. (The flag $\mathbb{I}_{\tau(\omega)=n}$ in (8.15) ensures that no proper initial subsequence of a sequence in $\Omega'$ is also in $\Omega'$; see p. 149.) We can interpret the four processes $\mathcal{S}$, $\mathcal{W}$, $\mathcal{A}$, and $\mathcal{C}$ as processes in this game-theoretic sample space, because their $n$th values are always identified by World's $n$th move $\mathbf{w}_n$. The same is true for $\mathcal{B}$ and $\mathcal{U}$, because their current values are always identified by the current and previous values of $\mathcal{S}$, $\mathcal{A}$, and $\mathcal{C}$. Then we define Skeptic's move space and gain function in each nonterminal situation $t$ by allowing him to buy arbitrary multiples of the increments $d_t\mathcal{S}$, $d_t\mathcal{W}$, $d_t\mathcal{B}$, and $d_t\mathcal{U}$ for 0.

Because Skeptic can buy their increments for 0, the four processes $\mathcal{S}$, $\mathcal{W}$, $\mathcal{B}$, and $\mathcal{U}$ are game-theoretic martingales in the Lindeberg protocol. So the martingale transform $\mathcal{L}$ is as well. By Proposition 8.4, $\mathcal{L}$ witnesses (8.12) in the Lindeberg protocol. This means that $\mathcal{L}_0 < \int U(z)\ \mathcal{N}_{0,1}(dz) + \epsilon$ and $\mathcal{L}_\tau \geq U(\mathcal{S}_\tau)$. Because $\mathcal{L}$ is a measure-theoretic martingale in the filtered probability space, this implies (8.14) by Ville's theorem. ∎

By applying Corollary 8.4 to $-\mathcal{S}$ and $-U(-s)$, we can establish further that $\mathbb{E}\left[U(\mathcal{S}_\tau)\right]$ exceeds $\int U(z)\ \mathcal{N}_{0,1}(dz) - \epsilon$. So altogether we have the usual conclusion: $\mathbb{E}\left[U(\mathcal{S}_\tau)\right]$ is close to $\int U(z)\ \mathcal{N}_{0,1}(dz)$.

## 8.2 COIN TOSSING

In this section, we compare the measure-theoretic and game-theoretic frameworks for coin tossing. Superficially, the game-theoretic picture looks very different from measure theory even in this simple case. The outcome of each toss is a move by Reality. Probabilities enter as moves by Forecaster. But once we fix a strategy for Forecaster, we obtain a probability measure for Reality's moves, and then the picture looks more like measure theory. Indeed, as we will verify, we get the same martingales and probabilities whether we think game-theoretically or measure-theoretically.

The two generalizations go in different directions, but in some cases the game-theoretic interpretation of a probability measure can be extended beyond coin tossing, and this can have some practical advantages.

### Binary Probability Forecasting

Consider the following coherent probability protocol, in which Reality always makes a binary choice, and Forecaster always gives a probability for what she will do.

BINARY PROBABILITY FORECASTING
**Parameter:** $\mathcal{K}_0 \in \mathbb{R}$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
FOR $n = 1, 2, \ldots$:
Forecaster announces $p_n \in (0, 1)$.
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{0, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n)$.

This is identical with the binary forecasting protocol we studied in Chapter 7 (p. 162), except that it does not terminate.

Forecaster has two opponents, Skeptic and Reality. We are interested in strategies for Forecaster that ignore previous moves by Skeptic, taking into account only previous moves by Reality. Such a strategy—we may call it a *neutral forecasting strategy*—can be specified by a family of numbers

$$\{\mathbf{p}_t\}_{t \in \{0,1\}^*}, \tag{8.16}$$

where $\{0,1\}^*$ denotes the set of all finite sequences of 0s and 1s, including the empty sequence $\Box$, and $0 < \mathbf{p}_t < 1$. The strategy directs Forecaster to set $p_1$ equal to $\mathbf{p}_\Box$ and to set $p_n$ equal to $\mathbf{p}_{x_1 \ldots x_{n-1}}$ if Reality's first $n-1$ moves are $x_1 \ldots x_{n-1}$.

If we require Forecaster to use a particular neutral forecasting strategy, then he has no decisions left to make, and we can remove him from the description of the protocol. We thereby obtain the following coherent probability protocol involving Skeptic and Reality only.

COIN TOSSING
**Parameter:** $\mathcal{K}_0 \in \mathbb{R}$, neutral forecasting strategy $\{\mathbf{p}_t\}_{t \in \{0,1\}^*}$
**Players:** Skeptic, Reality
**Protocol:**
FOR $n = 1, 2, \ldots$:
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{0,1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - \mathbf{p}_{x_1 \ldots x_{n-1}})$.

If the $\mathbf{p}_t$ are all equal to $1/2$, this reduces, essentially, to the fair-coin protocol that we studied in Chapter 3 (p. 64).

### For Coin-Tossing, Forecaster's Strategy = Probability Distribution

As the reader may have noticed, a neutral forecasting strategy for the binary forecasting protocol amounts to the same thing as a positive probability distribution for an infinite sequence of 0s and 1s.

Recall that a *probability distribution* for an infinite sequence of 0s and 1s is a probability measure on the filtered measurable space $(\{0,1\}^\infty, \mathcal{C}, \{\mathcal{C}_n\}_{n=0}^\infty)$, where $\mathcal{C}_n$ consists of cylinder sets corresponding to subsets of $\{0,1\}^n$, and $\mathcal{C} = \mathcal{C}_\infty$. Such a probability distribution $\mathbb{P}$ is completely determined once we give, for every finite sequence $x_1 \ldots x_n$ of 0s and 1s, the probability for the cylinder set corresponding to the element $x_1 \ldots x_n$ of $\{0,1\}^n$. We may write $[x_1 \ldots x_n]$ for this cylinder set and $\mathbb{P}[x_1 \ldots x_n]$ for its probability. If

$$\mathbb{P}[x_1 \ldots x_n] > 0$$

for every finite sequence $x_1 \ldots x_n$ of 0s and 1s, then we say that the probability distribution $\mathbb{P}$ is *positive*.

There is a one-to-one correspondence between neutral forecasting strategies and positive probability distributions for an infinite sequence of 0s and 1s:

- Given $\mathbb{P}$, we can define $\{\mathbf{p}_t\}_{t\in\{0,1\}^*}$ by setting the $\mathbf{p}_t$ equal to $\mathbb{P}$'s conditional probabilities:

$$\mathbf{p}_{x_1\ldots x_{n-1}} := \frac{\mathbb{P}[x_1 \ldots x_{n-1}1]}{\mathbb{P}[x_1 \ldots x_{n-1}]}.$$

- Given $\{\mathbf{p}_t\}_{t\in\{0,1\}^*}$, we can define probability distributions $\{\mathbf{P}_t\}_{t\in\{0,1\}^*}$ in $\{0,1\}$ by

$$\mathbf{P}_t\{x\} := \begin{cases} \mathbf{p}_t & \text{if } x = 1 \\ 1 - \mathbf{p}_t & \text{if } x = 0, \end{cases} \tag{8.17}$$

  and then we can define $\mathbb{P}$ by setting

$$\mathbb{P}[x_1 \ldots x_n] := \mathbf{P}_{\Box}\{x_1\}\mathbf{P}_{x_1}\{x_2\} \cdots \mathbf{P}_{x_1\ldots x_{n-1}}\{x_n\} \tag{8.18}$$

  for every finite sequence $x_1 \ldots x_n$ of 0s and 1s.

## For Coin Tossing, Game-Theoretic = Measure-Theoretic

We now come to our main point. For coin tossing, we get the same martingales and probabilities whether we think game-theoretically or measure-theoretically.

**Proposition 8.5** *Consider a coin-tossing protocol with parameter $\{\mathbf{p}_t\}_{t\in\{0,1\}^*}$ and the filtered probability space $(\mathbb{P}, \{0,1\}^\infty, \mathcal{C}, \{\mathcal{C}_n\}_{n=0}^\infty)$ defined by (8.17) and (8.18).*

1. *Every game-theoretic martingale in the coin-tossing protocol is a measure-theoretic martingale in the probability space.*

2. *Every measure-theoretic martingale in the probability space is a game-theoretic martingale in the coin-tossing protocol.*

3. *Every event in the probability space has a game-theoretic probability in the protocol, which is equal to the event's probability in the probability space.*

*Proof* We begin by observing that the protocol and the probability space have the same processes. In either case, a process is a real-valued function $\mathcal{L}$ on $\{0,1\}^*$. It can also be described as a sequence $\mathcal{L}_0, \mathcal{L}_1, \ldots$, where $\mathcal{L}_n$ is a function of $n$ binary variables $x_1, \ldots, x_n$. Because of the simplicity of the probability space, there are no measurability problems; all functions of $x_1 \ldots x_n$ are measurable.

To prove Statement 1, we begin by observing that $\mathbf{p}_{x_1\ldots x_{n-1}}$, considered as a function of $x_1, \ldots, x_{n-1}$, is the unique version of the conditional expectation $\mathbb{E}[x_n \mid x_1, \ldots, x_{n-1}]$. The condition for a process $\mathcal{L}$ to qualify as a game-theoretic martingale is that

$$\mathcal{L}(x_1, \ldots, x_n) = \mathcal{L}(x_1, \ldots, x_{n-1}) + \mathcal{M}(x_1, \ldots, x_{n-1})(x_n - \mathbf{p}_{x_1\ldots x_{n-1}}) \tag{8.19}$$

for some process $\mathcal{M}$. Taking the conditional expectation of both sides with respect to $x_1, \ldots, x_{n-1}$ and substituting $\mathbb{E}[x_n \mid x_1, \ldots, x_{n-1}]$ for $\mathbf{p}_{x_1\ldots x_{n-1}}$, we obtain

$$\mathbb{E}[\mathcal{L}(x_1, \ldots, x_n) \mid x_1, \ldots, x_{n-1}] = \mathcal{L}(x_1, \ldots, x_{n-1}), \tag{8.20}$$

the condition for $\mathcal{L}$ to qualify as a measure-theoretic martingale.

To prove Statement 2, we use the fact that $x_n$ can only take the values 0 and 1 to write

$$\mathcal{L}(x_1, \ldots, x_n) = \mathcal{L}(x_1, \ldots, x_{n-1}, 1)x_n + \mathcal{L}(x_1, \ldots, x_{n-1}, 0)(1 - x_n)$$

for any process $\mathcal{L}$. This implies

$$\begin{aligned} &\mathcal{L}(x_1, \ldots, x_n) - \mathbb{E}[\mathcal{L}(x_1, \ldots, x_n) \mid x_1, \ldots, x_{n-1}] \\ &= \mathcal{M}(x_1, \ldots, x_{n-1})(x_n - \mathbb{E}[x_n \mid x_1, \ldots, x_{n-1}]) \end{aligned} \tag{8.21}$$

almost surely, where $\mathcal{M}$ is the process defined by

$$\mathcal{M}(x_1, \ldots, x_{n-1}) := \mathcal{L}_n(x_1, \ldots, x_{n-1}, 1) - \mathcal{L}_n(x_1, \ldots, x_{n-1}, 0).$$

Because conditional expectations are unique, we can substitute $\mathbf{p}_{x_1 \ldots x_{n-1}}$ for $\mathbb{E}[x_n \mid x_1, \ldots, x_{n-1}]$ in (8.21). If $\mathcal{L}$ is a measure-theoretic martingale, then we can also substitute $\mathcal{L}(x_1, \ldots, x_{n-1})$ for $\mathbb{E}[\mathcal{L}(x_1, \ldots, x_n) \mid x_1, \ldots, x_{n-1}]$, obtaining (8.19).

Now suppose $E$ is an event in the probability space—that is, $E \in \mathcal{C}$. Then by Ville's theorem for positive probabilities, Proposition 8.13,

$$\mathbb{P} E = \inf \left\{ \mathcal{L}_0 \,\middle|\, \liminf_{n \to \infty} \mathcal{L}_n \geq \mathbb{I}_E \right\}, \tag{8.22}$$

where $\mathcal{L}$ ranges over measure-theoretic martingales. The initial game-theoretic upper probability of $E$, $\overline{\mathbb{P}}_\square E$, is given by exactly the same formula, with $\mathcal{L}$ ranging over game-theoretic martingales. Because the measure-theoretic and game-theoretic martingales are the same, we can immediately conclude that $\overline{\mathbb{P}}_\square E = \mathbb{P} E$. Similarly, $\overline{\mathbb{P}}_\square E^c = \mathbb{P} E^c$, or $\mathbb{P} E = \underline{\mathbb{P}}_\square E$. ∎

## Generalizing from Coin Tossing

It is instructive to think about the game-theoretic and measure-theoretic frameworks as two different generalizations of coin tossing. The game-theoretic generalization enlarges the freedom of both Reality and Forecaster. Reality may have more than a binary choice, and Forecaster may say much or little about what Reality will do. The measure-theoretic generalization, on the other hand, keeps Reality subject to a probability measure. This measure is analogous to our Forecaster, but in general it is more ambitious; it gives a price for every possible aspect of what Reality will do.

To what extent can the measure-theoretic generalization of coin tossing be understood inside our game-theoretic generalization?

Certainly we can generalize probability forecasting beyond the binary case. We might, for example, ask Forecaster to give a whole probability measure for an outcome (such as a real number) to be announced by Reality, thus obtaining this generalization:

PROBABILITY FORECASTING
**Parameters:** Measurable space $(\Omega_\bullet, \mathcal{F})$, $\mathcal{K}_0 \in \mathbb{R}$
**Players:** Forecaster, Skeptic, Reality
**Protocol:**
FOR $n = 1, 2, \ldots$:
Forecaster announces a probability measure $p_n$ on $(\Omega_\bullet, \mathcal{F})$.

Skeptic announces a measurable function $f_n$ on $\Omega_\bullet$ such that $\int_{\Omega_\bullet} f_n dp_n$, the expected value of $f_n$ with respect to $p_n$, exists.
Reality announces $x_n \in \Omega_\bullet$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(x_n) - \int_{\Omega_\bullet} f_n dp_n$.

This is a coherent probability protocol, with a very complex sample space. The paths are sequences $p_1, x_1, p_2, x_2, \ldots$, where the $x_n$ are elements of $\Omega_\bullet$ and the $p_n$ are probability measures on $\Omega_\bullet$.

A neutral forecasting strategy for this protocol is a family of probability distributions $\{\mathbf{P}_t\}_{t\in\Omega_\bullet^*}$, where $\Omega_\bullet^*$ denotes the set of all finite sequences of elements of $\Omega_\bullet$. Imposing such a strategy on Forecaster produces this generalization of the coin-tossing protocol:

GENERALIZED COIN TOSSING
**Parameters:** $(\Omega_\bullet, \mathcal{F})$, $\mathcal{K}_0 \in \mathbb{R}$, neutral forecasting strategy $\{\mathbf{P}_t\}_{t\in\Omega_\bullet^*}$
**Players:** Skeptic, Reality
**Protocol:**
FOR $n = 1, 2, \ldots$:
Skeptic announces a measurable function $f_n$ on $\Omega_\bullet$ such that $\mathbf{E}_{x_1\ldots x_{n-1}} f_n$, the expected value of $f_n(x_n)$ with respect to $\mathbf{P}_{x_1\ldots x_{n-1}}$, exists.
Reality announces $x_n \in \Omega_\bullet$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(x_n) - \mathbf{E}_{x_1\ldots x_{n-1}} f_n$.

Here the paths are merely sequences $x_1, x_2, \ldots$, and the sample space is the Cartesian product $\Omega_\bullet^\infty$. A simple special case is where $\Omega_\bullet$ is equal to $\{0, 1\}$; in this case the generalized coin-tossing protocol is the same as the coin-tossing protocol except that some of the probabilities $\mathbf{p}_t$ may be equal zero or one.

Intuitively, the $\mathbf{P}_{x_1\ldots x_{n-1}}$ in the generalized coin-tossing protocol are conditional probabilities, and so this protocol can be thought of as a game-theoretic representation of probability models, such as Markov chains, that are defined by means of successive conditional distributions.

If the neutral forecasting strategy is measurable (i.e., $\mathbf{P}_{x_1\ldots x_{n-1}}(A)$ is a measurable function of $x_1, \ldots, x_{n-1}$ for all $n$ and all measurable $A \subseteq \Omega_\bullet$), then it determines a probability measure on $\Omega_\bullet^\infty$ by Ionescu-Tulcea's theorem ([287], §II.9). We can then prove the following generalization of Statement 3 of Proposition 8.5.

**Proposition 8.6** *If $\mathbb{P}$ is the probability measure determined by Ionescu-Tulcea's theorem for the neutral forecasting strategy $\{\mathbf{P}_t\}_{t\in\Omega_\bullet^*}$, then every measurable subset $E$ in $\Omega_\bullet^\infty$ has a game-theoretic probability in the corresponding generalized coin-tossing protocol, and this game-theoretic probability is equal to $\mathbb{P}(E)$.*

Here we cannot assert that the game-theoretic and measure-theoretic martingales are exactly the same. However, a measure-theoretic martingale can be always made into a game-theoretic martingale by changes on a set of measure zero. Proposition 8.6 can be proven using this fact together with the ideas in Proposition 8.5—Ville's theorem and the definition of game-theoretic upper probability in terms of game-theoretic martingales.

The measure-theoretic framework *starts* with a probability distribution $\mathbb{P}$ in $\Omega_\bullet^\infty$, and then the question is whether there exists a neutral forecasting strategy that gives rise to $\mathbb{P}$. Standard results on the existence of regular conditional probabilities ([287], Theorem II.7.5) imply that the answer is "yes" provided $\Omega_\bullet$ is a Borel space (in particular, if $\Omega_\bullet = \mathbb{R}$).

What advantage is there in the game-theoretic representation of models already understood within the measure-theoretic framework? The main advantage may be in interpretation. The explicit interpretation of conditional probabilities as forecasts may be less mysterious and more helpful in applications than the notion that $x_1, x_2, \ldots$ are generated by the probability measure $\mathbb{P}$. The protocol discourages us from saying that $\mathbb{P}$ generates $x_1, x_2, \ldots$; they are obviously chosen by Reality. On the other hand, we can give a clear meaning to the statement that $\mathbb{P}$ *governs* $x_1, x_2, \ldots$; this can be interpreted as an assertion of the fundamental interpretative hypothesis. Reality will avoid letting Skeptic become too rich, it says, and so will avoid any path $x_1, x_2, \ldots$ on which any particular nonnegative martingale for $\mathbb{P}$ becomes too large. Forecaster does not force Reality's hand, but he has somehow found and adopted forecasts that Reality respects—forecasts that guide or govern her choices without determining them.

## 8.3 GAME-THEORETIC PRICE AND PROBABILITY

In this section, we return to the abstract study of game-theoretic price and probability that we began in §1.2 and §1.3 and continued in §7.1. We organize this study by distinguishing three types of protocols for our sequential perfect-information game between Skeptic and World. They involve increasingly strong assumptions about Skeptic's move space:

- **Gambling Protocols.** We call the protocol a gambling protocol if (1) Skeptic's moves do not affect the moves later available to World, and (2) the moves by Skeptic and World in each situation determine an immediate monetary payoff for Skeptic. This is enough to permit us to define upper and lower prices.

- **Probability Protocols.** Here we add some minimal assumptions on Skeptic's move space and gain function, corresponding to the assumption that he can combine tickets and can buy any *positive* fraction or multiple of a ticket. These are the assumptions we made in §1.2. They are enough to permit us to define useful concepts of upper and lower probability.

- **Symmetric Probability Protocols.** Here we assume further that Skeptic can buy negative as well as positive amounts of tickets, so that his move space and gain function are linear in the usual sense. This is the one of the assumptions we made in Chapter 7.

In this section, in contrast, with §7.1, we do not assume that play terminates. Protocols of all three types may allow play to continue indefinitely.

It is not our intention in this section to formulate a framework so general that it accommodates all the ways game theory might be used as a foundation for probability. Game theory is an extraordinarily flexible tool, and we expect that others will find yet other ways to relate it to probability. Even in this book, in Chapter 13, we use games more general than the ones considered here.

## Gambling Protocols

We begin, as always, with the *sample space* $\Omega$, which consists of all sequences of moves World is allowed to make. The space $\Omega$ may be any set of sequences with the property that no proper initial subsequence of a sequence in $\Omega$ is also in $\Omega$. Some or all of the sequences in $\Omega$ may be infinite.

We adopt most of the terminology and notation for sample spaces that we have used in preceding chapters. An element of $\Omega$ is a *path*. A finite initial segment of a path is a *situation*. We write $|\xi|$ for the length of the path $\xi$; $|\xi|$ may be infinite. If $|\xi| \geq n$, we write $\xi^n$ for $\xi$'s initial segment of length $n$. The *initial* (empty) situation is denoted by $\Box$, and the set of all situations is denoted by $\Omega^\diamond$. World's *move space* in situation $t$ is denoted by $\mathbf{W}_t$:

$$\mathbf{W}_t := \{\mathbf{w} \mid t\mathbf{w} \in \Omega^\diamond\},$$

where $t\mathbf{w}$ is the sequence obtained by placing the single move $\mathbf{w}$ after the sequence of moves $t$. We define *precede, follow, strictly precede, strictly follow, child, parent, process, t-process, event, variable*, and *t-variable* as in §7.1.

If the situation $t$ is an initial segment of the path $\xi$, then $\xi$ *goes through* $t$, and $t$ *is on* $\xi$. Given a real-valued $t$-process $\mathcal{U}$, we define a $t$-variable $\liminf \mathcal{U}$ by

$$\liminf \mathcal{U}(\xi) := \begin{cases} \liminf_{n\to\infty} \mathcal{U}(\xi^n) & \text{if } \xi \text{ is infinite} \\ \mathcal{U}(u) & \text{if } u \text{ is the final situation on } \xi \end{cases}$$

for all $\xi$ that go through $t$. In words: lim inf is defined in the usual way if the path is infinite but is equal to the value in the final situation if the path is finite. We define lim sup and lim similarly.

A set $U$ of situations is a *cut* of a situation $t$ if (1) all the situations in $U$ follow $t$ and (2) there is exactly one situation from $U$ on each path through $t$, as in Figure 8.1.

For each nonfinal situation $t$, we specify Skeptic's *move space* in $t$, a nonempty set $\mathbf{S}_t$, and Skeptic's *gain function* in $t$, a real-valued function $\lambda_t$ on $\mathbf{S}_t \times \mathbf{W}_t$. As usual, $\lambda_t(\mathrm{s}, \mathbf{w})$ is the gain for Skeptic when he moves s and World then moves $\mathbf{w}$; notice that because of the subindex $t$ Skeptic's gain may depend not only on World's current move but also on his previous moves. For the moment, we make no further assumptions about $\mathbf{S}_t$ and $\lambda_t$.

As usual, a *strategy* is a strategy for Skeptic—a partial process $\mathcal{P}$ defined on all nonfinal situations and satisfying $\mathcal{P}(t) \in \mathbf{S}_t$. When Skeptic starts with capital 0 and follows $\mathcal{P}$, his *capital process* is $\mathcal{K}^{\mathcal{P}}$, where $\mathcal{K}^{\mathcal{P}}(\Box) = 0$ and

$$\mathcal{K}^{\mathcal{P}}(t\mathbf{w}) = \mathcal{K}^{\mathcal{P}}(t) + \lambda_t(\mathcal{P}(t), \mathbf{w}). \tag{8.23}$$

When he starts with capital $\alpha$ and follows $\mathcal{P}$, his capital process is $\alpha + \mathcal{K}^{\mathcal{P}}$. A process $\mathcal{T}$ is a *supermartingale* if there is a strategy $\mathcal{P}$ such that

$$\mathcal{T}(t\mathbf{w}) \leq \mathcal{T}(t) + \lambda_t(\mathcal{P}(t), \mathbf{w}) \tag{8.24}$$

for every nonterminal situation $t$ and every $\mathbf{w} \in \mathbf{W}_t$. Skeptic's capital will follow $\mathcal{T}$ if he starts with $\mathcal{T}(\Box)$, follows $\mathcal{P}$, but gives away $\mathcal{T}(t) + \lambda_t(\mathcal{P}(t), \mathbf{w}) - \mathcal{T}(t\mathbf{w})$ upon arrival in $t\mathbf{w}$. Thus the supermartingales are the processes that Skeptic's capital can follow if he is allowed to give money away. As in §7.1, we also consider $t$-strategies, which begin in the situation $t$. They give rise to $t$-capital processes and $t$-supermartingales.

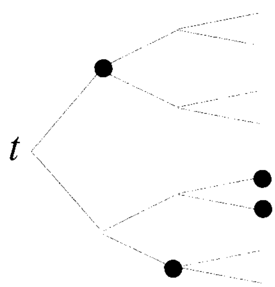



***Fig. 8.1*** The situations marked with dots form a cut of $t$.

Given a variable $x$ and a situation $t$, we define the *upper price* of $x$ in $t$ by

$$\overline{\mathbb{E}}_t\, x := \inf\{\mathcal{S}(t) \mid \mathcal{S} \text{ is a } t\text{-capital process, and } \liminf \mathcal{S} \geq x\} \tag{8.25}$$

(relations such as $A \geq B$ are understood as satisfied for all $\xi$ in the domain of both $A$ and $B$; therefore, $\liminf \mathcal{S} \geq x$ means that $\liminf \mathcal{S}(\xi) \geq x(\xi)$ for all $\xi$ passing through $x$). This is equivalent to the definition we stated verbally at the end of §1.2: $\overline{\mathbb{E}}\, x$ is the infimum of all $\alpha$ such that there is a capital process starting at $\alpha$ that eventually reaches $x$ and does not go back below it. Similarly, we define the *lower price* of $x$ in $t$ by

$$\underline{\mathbb{E}}_t\, x := \sup\{-\mathcal{S}(t) \mid \mathcal{S} \text{ is a } t\text{-capital process, and } \liminf \mathcal{S} \geq -x\}. \tag{8.26}$$

When the protocol is terminating, Equations (8.25) and (8.26) reduce to the definitions for terminating protocols given in §1.2 and repeated in §7.1. They imply, of course, that

$$\underline{\mathbb{E}}_t\, x = -\overline{\mathbb{E}}_t[-x] \tag{8.27}$$

for every situation $t$ and every $t$-variable $x$. As usual, we write $\mathbb{E}_t\, x$ for the common value of $\overline{\mathbb{E}}_t\, x$ and $\underline{\mathbb{E}}_t\, x$ when they are equal; this is the *price* of $x$ in $t$.

Suppose $x$ is a $t$-variable and $U$ is a cut of $t$. Then we can define a $t$-variable $\overline{\mathbb{E}}_U\, x$ by

$$(\overline{\mathbb{E}}_U\, x)(\xi) := \overline{\mathbb{E}}_u\, x,$$

where $u$ is the unique situation on $\xi$ in $U$.

**Proposition 8.7** *Suppose $x$ is a $t$-variable and $U$ is a cut of $t$. Then*

$$\overline{\mathbb{E}}_t\, x = \overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x] \tag{8.28}$$

*and*

$$\underline{\mathbb{E}}_t\, x = \underline{\mathbb{E}}_t[\underline{\mathbb{E}}_U\, x]. \tag{8.29}$$

*Proof* Choose $\epsilon > 0$, and choose a $t$-strategy $\mathcal{P}$ that produces at least $\overline{\mathbb{E}}_U\, x$ when it starts in $t$ with $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x] + \epsilon$. And for each $u \in U$, choose a $u$-strategy $\mathcal{P}_u$ that produces at least $x$ when it starts in $u$ with $\overline{\mathbb{E}}_u\, x + \epsilon$. Combining these strategies in the obvious way (starting in $t$, play $\mathcal{P}$ and then after you go through a situation $u$ in $U$, play $\mathcal{P}_u$), we obtain a $t$-strategy that produces at least $x$ when started with $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x] + 2\epsilon$. So $\overline{\mathbb{E}}_t\, x \leq \overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x] + 2\epsilon$. This shows that $\overline{\mathbb{E}}_t\, x \leq \overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x]$.

Now choose a $t$-strategy $\mathcal{P}$ that produces at least $x$ when it starts in $t$ with $\overline{\mathbb{E}}_t\, x + \epsilon$. For each $u$ in $U$, this strategy produces at least $x$ when it starts in $u$ with capital $\mathcal{K}^{\mathcal{P}}(u)$. So

$$\overline{\mathbb{E}}_u\, x \leq \mathcal{K}^{\mathcal{P}}(u) \tag{8.30}$$

for each $u$ in $U$. Let us write $\mathcal{K}^{\mathcal{P}}_U$ for the $t$-variable whose value for $\xi$ is $\mathcal{K}^{\mathcal{P}}(u)$ when $\xi$ goes through $u \in U$. Then (8.30) says that $\overline{\mathbb{E}}_U\, x \leq \mathcal{K}^{\mathcal{P}}_U$, and it follows that $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x] \leq \overline{\mathbb{E}}_t\, \mathcal{K}^{\mathcal{P}}_U$. On the other hand, we have a strategy (play $\mathcal{P}$ until $U$ and then stop) that produces $\mathcal{K}^{\mathcal{P}}_U$ starting in $t$ with $\overline{\mathbb{E}}_t\, x + \epsilon$. So $\overline{\mathbb{E}}_t\, \mathcal{K}^{\mathcal{P}}_U \leq \overline{\mathbb{E}}_t\, x + \epsilon$. So $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x] \leq \overline{\mathbb{E}}_t\, x + \epsilon$. This shows that $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_U\, x] \leq \overline{\mathbb{E}}_t\, x$.

Equation (8.29) follows from (8.28) by (8.27). ∎

In probability theory, the equation $\mathbb{E}_t\, x = \mathbb{E}_t[\mathbb{E}_U\, x]$ is sometimes called the *law of iterated expectation*. Its appearance at this stage of our study, where we have made only the minimal assumptions needed to define upper and lower price, shows how purely game-theoretic it is.

### Probability Protocols

We call a gambling protocol a *probability protocol* when it satisfies the following assumptions:

1. Each $\mathbf{S}_t$ is a convex cone in a linear space. In other words, if $\mathbf{s}_1$ and $\mathbf{s}_2$ are in $\mathbf{S}_t$ and $a_1$ and $a_2$ are nonnegative numbers, then $a_1\mathbf{s}_1 + a_2\mathbf{s}_2$ is in $\mathbf{S}_t$.

2. Each $\lambda_t$ has the following linearity property: if $\mathbf{s}_1$ and $\mathbf{s}_2$ are in $\mathbf{S}_t$ and $a_1$ and $a_2$ are nonnegative numbers, then $\lambda_t(a_1\mathbf{s}_1 + a_2\mathbf{s}_2, \mathbf{w}) = a_1\lambda_t(\mathbf{s}_1, \mathbf{w}) + a_2\lambda_t(\mathbf{s}_2, \mathbf{w})$ for every $\mathbf{w} \in \mathbf{W}_t$.

These two assumptions were stated informally in §1.2 and adopted in all the games we have studied so far: Skeptic can combine freely all the tickets offered to him, and he can buy any positive fraction or multiple of a ticket. He is also always free to choose the move 0, so that his capital does not change.

In a probability protocol, the capital processes form a convex cone in the space of all processes. The supermartingales do as well.

The following proposition lists some properties of upper price in a probability protocol.

**Proposition 8.8** *Suppose $t$ is a nonfinal situation. Then the upper prices have the following properties:*

1. *If $x$ is a t-variable, then* $\overline{\mathbb{E}}_t\, x \leq \sup\{x(\xi) \mid \xi \in \Omega, \xi$ *goes through* $t\}$.

2. *If $x_1$ and $x_2$ are t-variables, then* $\overline{\mathbb{E}}_t[x_1 + x_2] \leq \overline{\mathbb{E}}_t\, x_1 + \overline{\mathbb{E}}_t\, x_2$.

3. *If $x$ is a t-variable and $\alpha > 0$, then* $\overline{\mathbb{E}}_t[\alpha x] = \alpha\, \overline{\mathbb{E}}_t\, x$.

4. *If $x$ is a t-variable and $\alpha \in \mathbb{R}$, then* $\overline{\mathbb{E}}_t[x + \alpha] = \overline{\mathbb{E}}_t\, x + \alpha$.

5. *If $x_1$ and $x_2$ are t-variables, and $x_1 \leq x_2$, then* $\overline{\mathbb{E}}_t\, x_1 \leq \overline{\mathbb{E}}_t\, x_2$.

These properties follow directly from the definition of upper price and our assumptions about $\mathbf{S}_t$ and $\lambda_t$. We leave it to the reader to verify these properties and to formulate the corresponding properties for lower price.

The first three properties in Proposition 8.8 were taken as the defining properties of upper probability by Peter Walley (1991, p. 65); see also Peter M. Williams (1976). Walley and Williams were inspired by Bruno de Finetti, who had emphasized similar properties for probability (see p. 188). These authors also considered the relation between unconditional and conditional prices, but they were not working in a dynamic framework and so did not formulate Proposition 8.7.

As in §7.1, we call a probability protocol *coherent* in $t$ if World has a $t$-strategy that guarantees that Skeptic ends the game with no cumulative gain from $t$ onward. We call it simply *coherent* if it is coherent in every situation $t$. Coherence in $t$ implies that for every $\mathbf{s} \in \mathbf{S}_t$, World has a move $\mathbf{w} \in \mathbf{W}_t$ such that $\lambda_t(\mathbf{s}, \mathbf{w}) \leq 0$. Coherence is equivalent to the requirement that for every situation $t$ and every $\mathbf{s} \in \mathbf{S}_t$, World has a move $\mathbf{w} \in \mathbf{W}_t$ such that $\lambda_t(\mathbf{s}, \mathbf{w}) \leq 0$.

**Proposition 8.9** *Suppose $t$ is a nonfinal situation in a coherent probability protocol.*

1. *If $x$ is a t-variable, then* $\underline{\mathbb{E}}_t\, x \leq \overline{\mathbb{E}}_t\, x$.

2. *If $a$ is a real number and* $\mathbf{a}$ *designates the variable that is equal to $a$ on every path, then* $\mathbb{E}_t\, \mathbf{a} = a$.

When the protocol is coherent, we can also strengthen Property 3 in Proposition 8.8 by relaxing $\alpha > 0$ to $\alpha \geq 0$.

We define upper and lower probabilities for an event $E$ in a situation $t$ as usual: $\overline{\mathbb{P}}_t\, E := \overline{\mathbb{E}}_t\, \mathbb{I}_E$ and $\underline{\mathbb{P}}_t\, E := \underline{\mathbb{E}}_t\, \mathbb{I}_E$, where $\mathbb{I}_E$ is the indicator variable for $E$.

**Proposition 8.10** *Suppose $t$ is a nonfinal situation in a coherent probability protocol. Then*

1. $\overline{\mathbb{P}}_t\, E = 1 - \underline{\mathbb{P}}_t\, E^c$,

2. $\overline{\mathbb{P}}_t[E_1 \cup E_2] \leq \overline{\mathbb{P}}_t\, E_1 + \overline{\mathbb{P}}_t\, E_2$, *and*

3. $\underline{\mathbb{P}}_t[E_1 \cap E_2] \geq \underline{\mathbb{P}}_t\, E_1 + \underline{\mathbb{P}}_t\, E_2 - 1$.

When the fundamental interpretative hypothesis is adopted, so that an upper probability close to zero means that an event is very unlikely and a lower probability close to one means that it is very likely, Statements 2 and 3 of this proposition acquire substantive interpretations. Statement 2 says that the union of two very unlikely events is very unlikely, and Statement 3 says that the intersection of two very likely events is very likely.

**Proposition 8.11** *Suppose t is a nonfinal situation in a coherent probability protocol. Then*

$$\overline{\mathbb{P}}_t\, E = \inf\{\mathcal{S}(t) \mid \mathcal{S} \text{ is a nonnegative } t\text{-capital process, and } \liminf \mathcal{S} \geq 1 \text{ on } E\}. \tag{8.31}$$

*Proof* The condition that $\liminf \mathcal{S}(\xi) \geq 0$ for all $\xi$, together with coherence, implies that $\mathcal{S}$ is nonnegative, for if $\mathcal{S}(t) < 0$, then coherence implies that there is a path going through $t$ along which $\mathcal{S}$ remains negative. So (8.31) follows directly from (8.25). ∎

Another way of expressing (8.31) is to say that $\overline{\mathbb{P}}_t\, E$ is the infimum of all nonnegative numbers $\alpha$ such that Skeptic has a strategy that starts with capital 1 and reaches $1/\alpha$ on $E$ without risking bankruptcy.

Recall the definition of forcing in Chapter 3 (p. 65): a strategy *forces* $E$ if it has a nonnegative capital process $\mathcal{S}$ satisfying $\lim_{n\to\infty} \mathcal{S}(\xi^n) = \infty$ for all $\xi$ in the complement $E^c$. When Skeptic has a strategy that forces $E$, we say that Skeptic *can force* $E$, and that $E$ happens *almost surely*. Comparing the definition of forcing with (8.31), we obtain the following proposition:

**Proposition 8.12** *If Skeptic can force* $E$*, then* $\overline{\mathbb{P}}\, E^c = 0$.

Intuitively, the converse should also be true. If there are strategies that do not risk bankruptcy and multiply one's capital by an arbitrarily large factor when $E$ does not happen, we should be able to obtain a strategy that makes the multiplication infinite by forming an infinite convex combination, as in Lemma 3.1 (p. 68). In the present abstract context, however, some regularity conditions would be needed in order to assure the existence of the convex combination. We will not investigate such conditions here. (But see the discussion in connection with Proposition 4.5 on p. 93.)

## Symmetric Probability Protocols

We call a probability protocol *symmetric* if whenever $\mathbf{s} \in \mathbf{S}_t$, we also have $-\mathbf{s} \in \mathbf{S}_t$ and $\lambda_t(-\mathbf{s}, \mathbf{w}) = -\lambda_t(\mathbf{s}, \mathbf{w})$ for all $\mathbf{w} \in \mathbf{W}_t$. When we describe the moves available to Skeptic in terms of tickets that are offered to him, this means that Skeptic can buy negative as well as positive amounts of any ticket—he can sell at the same price at which he can buy.

Given the assumptions already satisfied by $\mathbf{S}_t$ and $\lambda_t$ in a probability space (p. 185), symmetry means that $\mathbf{S}_t$ is a linear space and $\lambda_t(\mathbf{s}, \mathbf{w})$ is linear in $\mathbf{s}$. It follows that the strategies and capital processes available to Skeptic also form linear spaces. In particular, if $\mathcal{P}$ is a strategy for Skeptic, then $-\mathcal{P}$ is also a strategy for Skeptic, and by (8.23),

$$\mathcal{K}^{-\mathcal{P}} = -\mathcal{K}^{\mathcal{P}}. \tag{8.32}$$

In a symmetric probability protocol, we call a capital process (no matter whether it starts with 0 or some other initial capital—positive or negative) a *martingale*.

In an asymmetric probability protocol, Skeptic may have moves that amount to throwing money away. There may, for example, be a move $\mathbf{s} \in \mathbf{S}_t$ such that $\lambda_t(\mathbf{s}, \mathbf{w}) = -1$ for all $\mathbf{w} \in \mathbf{W}_t$: if Skeptic moves $\mathbf{s}$, he loses 1 no matter how World moves. This cannot happen in a coherent symmetric probability protocol. If there were a move by Skeptic that necessarily loses money, then there would be one that necessarily makes money, contrary to coherence.

In a symmetric probability protocol, we can use (8.32) to rewrite (8.26) as

$$\underline{\mathbb{E}}_t\, x = \sup\{\mathcal{S}(t) \mid \mathcal{S} \text{ is a } t\text{-capital process, and } \limsup \mathcal{S} \leq x\}. \tag{8.33}$$

This makes lower price analogous to upper price in a new way. Equation (8.25) says that the upper price of $x$ is the smallest initial capital that will enable Skeptic to make at least as much as $x$; Equation (8.33) says that the lower price of $x$ is the greatest initial capital with which Skeptic can manage to make no more than $x$.

As we have already seen in the preceding chapters, the probability protocols associated most closely with measure-theoretic probability are symmetric. For example, our formulation of Lindeberg's theorem in Chapter 7 assumed symmetry. The probability protocols considered in finance theory are also symmetric; investors can buy at the same prices at which they can sell. Only occasionally do finance theorists relax this condition in order to study the effect of a market's bid-ask spread (see, e.g., [162]). But from our point of view, symmetry is not a crucial condition for a probability game. There are many interesting asymmetric probability games, and the most important general properties of upper and lower price hold for asymmetric as well as symmetric protocols.

## Adding Tickets

If we abstract away from the details of Skeptic's move spaces and gain functions in a probability protocol and consider only the set of capital processes, then it is natural to consider how the upper and lower prices and probabilities change as this set is changed. If the probability protocol is coherent, then enlarging the set of capital processes corresponds to making more tickets available to Skeptic, at least in some situations, and it is obvious that this can only increase the lower prices and decrease the upper prices. It is natural to ask whether a sufficient enlargement will bring the lower and upper prices together, producing exact prices for all variables and exact probabilities for all events.

If the coherent probability protocol is symmetric and terminating, then the answer is yes. In the enlargement, the exact prices in each situation will satisfy de Finetti's rules for probability ([93], English edition, Vol. 1, p. 74):

1. If $x$ is a $t$-variable, then

$$\inf\{x(\xi) \mid \xi \text{ goes through } t\} \leq \mathbb{E}_t\, x \leq \sup\{x(\xi) \mid \xi \text{ goes through } t\}.$$

2. If $x_1$ and $x_2$ are $t$-variables, then $\mathbb{E}_t(x_1 + x_2) = \mathbb{E}_t\, x_1 + \mathbb{E}_t\, x_2$.

3. If $x$ is a $t$-variable and $\alpha \in \mathbb{R}$, then $\mathbb{E}_t(\alpha x) = \alpha \, \mathbb{E}_t \, x$.

And the upper and lower prices in the protocol will be the suprema and infima of the exact prices obtained by different enlargements ([336], Chapter 3). Similar results may be possible for nonterminating symmetric probability protocols. Readers who see exact probabilities as fundamental may be interested in pursuing this line of thought further, because it provides a way of understanding upper and lower prices in terms of exact prices.

## 8.4 OPEN SCIENTIFIC PROTOCOLS

In this section, we further illustrate the generality of probability games by explaining how they accommodate two widely used scientific models that use probabilities but fall outside the measure-theoretic framework: quantum mechanics and Cox's regression model. Both models go beyond the measure-theoretic framework because they are open in the sense we explained in §1.1: their forecasts are influenced by events in the world that lie outside the model.

### Quantum Mechanics

Quantum mechanics is nearly as old as measure-theoretic probability, and probabilists and statisticians have extended the measure-theoretic framework to accommodate it in many different ways [213, 231]. Here we do not try to survey or evaluate this work; we merely sketch how one version of von Neumann's axioms for quantum mechanics fits into the game-theoretic framework.

This version of von Neumann's axioms is adapted from Jammer 1974, Chapter 1 (see also Chapter III of von Neumann 1955).

**Axiom 8.1** *To every physical system corresponds a separable complex Hilbert space $\mathcal{H}$ whose unit vectors completely describe the state of the system.*

**Axiom 8.2** *To every observable (such as position, momentum, energy, or angular momentum) corresponds a unique self-adjoint operator $A$ acting in $\mathcal{H}$.*

Let us call an observable $A$ *simple* if there exist a complete orthonormal system of states $\psi_1, \psi_2, \ldots$ (*eigenfunctions*) and a sequence $a_1, a_2, \ldots$ of distinct real numbers (the corresponding *eigenvalues*) such that $A\phi = \sum_i a_i \lambda_i \psi_i$ for each state $\phi$, where the $\lambda_i$ are the coefficients in the expansion $\phi = \sum_i \lambda_i \psi_i$. We consider only simple observables.

**Axiom 8.3** *Let $A$ be an observable with eigenfunctions $\psi_1, \psi_2, \ldots$ and corresponding eigenvalues $\alpha_1, \alpha_2, \ldots$. For a system in a state $\phi = \sum_i \lambda_i \psi_i$, the probability that the result of a measurement of the observable $A$ is $\alpha_i$ equals $|\lambda_i|^2$ (and so results other than $\alpha_1, \alpha_2, \ldots$ are impossible).*

**Axiom 8.4** *The time development of the state $\phi$ is determined by the Schrödinger equation*

$$\frac{\partial \phi}{\partial t} = \frac{1}{i\hbar} H\phi,$$

*where $H$ is an observable called the* Hamiltonian *of the system.*

**Axiom 8.5** *Let $A$ be an observable with eigenfunctions $\psi_1, \psi_2, \ldots$ and corresponding eigenvalues $\alpha_1, \alpha_2, \ldots$. If a measurement of $A$ gives $\alpha_i$, then the state of the system immediately after the measurement is $\psi_i$.*

Axiom 8.5 is a condition on the observer's measuring instruments. In Pauli's terminology, the measuring instruments are required to be "ideal and of the first kind" ([12], §8.1). For simplicity we only consider a single physical system (otherwise an extra axiom, such as Postulate IV in [32], should be added).

Writing $\mathcal{P}(\mathbb{N})$ for the set of all probability measures on the natural numbers $\mathbb{N}$, we write the protocol for our game as follows:

GAME OF QUANTUM MECHANICS
**Parameters:** $\mathcal{K}_0 > 0$ (Skeptic's initial capital), $t_0$ (initial time)
**Players:** Observer, Quantum Mechanics, Skeptic, Reality
**Protocol:**

FOR $n = 1, 2, \ldots$:

Observer announces an observable $A_n$ and a time $t_n > t_{n-1}$.
Quantum Mechanics announces $a_n : \mathbb{N} \to \mathbb{R}$ and $p_n \in \mathcal{P}(\mathbb{N})$.
Skeptic announces a function $f_n : \mathbb{N} \to \mathbb{R}$ such that $\int f_n dp_n$ exists.
At time $t_n$, Reality announces the measurement $a_n(i_n)$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(i_n) - \int f_n dp_n$. (8.34)

In words, the protocol is:

- Observer decides what he is going to measure ($A_1$) and when he is going to measure it ($t_1$).
- Quantum Mechanics decides on the possible results $a_1(i)$, $i = 1, 2, \ldots$, of the measurement, together with their probabilities $p_1(i)$.
- Skeptic bets on what integer $i$ Reality will choose, at odds given by the probabilities $p_1(i)$.
- At time $t_1$, Reality chooses the measurement result $a_1(i_1)$ (which is equivalent to choosing the index $i_n$, since our observables are simple). Her choice is governed by the probabilities $p_1(i)$ only in the sense that she is expected to avoid allowing Skeptic to become rich by betting at the corresponding odds.
- Skeptic's capital is updated from $\mathcal{K}_0$ to $\mathcal{K}_1$ according to his bets $f_n$; (8.34) means that Skeptic buys $f_n$ for $\int f_n dp_n$, Quantum Mechanics's price for $f$.
- Everything is repeated: Observer decides what ($A_2$) and when ($t_2$) he is going to measure, etc.

Quantum Mechanics follows a fixed strategy in this game. This strategy uses an auxiliary variable $\phi_n \in \mathcal{H}$, the state of the system; an initial state $\phi_0 \in \mathcal{H}$ (a vector in $\mathcal{H}$ of unit length) is given at the outset. Quantum Mechanics calculates her move $(a_n, p_n)$ as follows:

John von Neumann (1903–1957). Born in Budapest under the Austro-Hungarian empire, he became the most trusted adviser of the United States military during the early years of the cold war. This photograph appears to have been taken in the 1930s.

- Let $A_n$'s eigenvalues and eigenfunctions be $a_{n,i}$ and $\psi_{n,i}$ $(i = 1, 2, \ldots)$, respectively.
- Quantum Mechanics solves $\frac{\partial \phi}{\partial t} = \frac{1}{i\hbar} H\phi$ with the initial condition $\phi(t_{n-1}) = \phi_{n-1}$ and sets $\phi_n^- := \phi(t_n)$ (this is the state of the system immediately before measuring $A_n$).
- Quantum Mechanics expands $\phi_n^-$ as a linear combination of $A_n$'s eigenfunctions: $\phi_n^- = \sum_i \lambda_{n,i}\psi_{n,i}$.
- Quantum Mechanics announces the function $a_n : i \mapsto a_{n,i}$ and the probability measure $p_n : i \mapsto |\lambda_{n,i}|^2$.
- Once Reality chooses $i_n$, the new state of the system, $\phi_n$, becomes $\psi_{n,i_n}$.

Skeptic's goal is to falsify Quantum Mechanics or to make some point (such as (8.35)); he succeeds in falsifying Quantum Mechanics if $\mathcal{K}_n$ is never negative and tends to infinity. As usual, we say that Skeptic *can force* an event $E$ (such as (8.35)) if he has a strategy that falsifies Quantum Mechanics outside $E$.

Our limit theorems can be applied to this game in many different ways. As an example, we give a corollary of the strong law of large numbers for the bounded forecasting game.

**Corollary 8.5** *Suppose Observer repeatedly measures observables bounded by some constant $C$. Skeptic can force the event*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left( a_n(i_n) - \int a_n dp_n \right) = 0. \tag{8.35}$$

## Cox's Regression Model

To illustrate how open statistical models can fit into the game-theoretic framework, we look at failure models in general and Cox's regression model in particular. This model was introduced by David R. Cox (1972); see also [63] and [65], Chapter 7.

We start with a general protocol for testing a failure model; Cox's regression model will be a particular strategy for one of the players, Model.

TESTING A FAILURE MODEL

**Parameters:** $\mathcal{K}_0 > 0$ (Skeptic's initial capital), $K \in \mathbb{N}$ (number of covariates)

**Players:** Reality, Skeptic, Model

**Protocol:**

FOR $n = 1, 2, \ldots$:

Reality announces group $B_n$ and $z_n^i \in \mathbb{R}^K$ (*covariates*) for every $i \in B_n$.

Model announces $p_n \in \mathcal{P}(B_n)$.

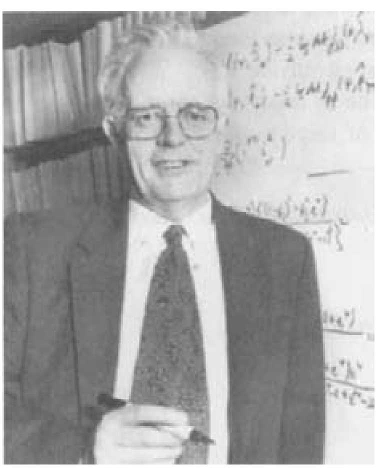
Skeptic announces a function $f_n : B_n \to \mathbb{R}$.

Reality announces $b_n \in B_n$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + f_n(b_n) - \int_{B_n} f_n dp_n$.

Reality stops when she pleases.

The group $B$ can be any finite non-empty set; we write $\mathcal{P}(B)$ for the set of all probability measures on $B$. This protocol describes the process of observation of a group of individuals, from which members may be deleted (say, by failure or censoring) or even added. Typically there is some continuity in the composition of the group, but this is not assumed. Every individual in the group is characterized by a set of $K$ covariates, such as blood pressure, sex, etc., which may be useful in estimating the individual's chances of failure. At the moment when the $n$th failure occurs, the experimenters record the composition of the group just prior to failure ($B_n$) and the covariates for each individual $i$ who was then in the group ($z_n^i$), including the covariates for the individual who failed. With this information (but without knowing which individual failed), Model provides a probability distribution $p_n$ for $b_n$, and Skeptic, who has the same information, is allowed to buy any payoff $f_n$ for the price $\int f_n dp_n$. Then the identity of the individual who failed, $b_n$, is announced. We ignore the possibility that two individuals might fail simultaneously.



Sir David R. Cox (born 1924), at Oxford in 1990.

Cox's regression model is the following strategy for Model: for some constant vector $\theta \in \mathbb{R}^K$ (unknown in advance),

$$p_n\{i\} := \frac{e^{z_n^i \cdot \theta}}{\sum_{j \in B_n} e^{z_n^j \cdot \theta}}, \quad i \in B_n.$$

In [63] Cox defined the *partial log-likelihood* for a data sequence

$$(B_1, z_1, b_1), \ldots, (B_N, z_N, b_N),$$

where $N$ is the duration of the game (we assume that Reality eventually stops), as the function $L$ of $\theta$ defined by

$$e^{L(\theta)} = \prod_{n=1}^{N} \frac{e^{z_n^{b_n} \cdot \theta}}{\sum_{i \in B_n} e^{z_n^i \cdot \theta}}.$$

Because we assume no stochastic mechanism, the adjective "partial" is not helpful. So we call $L$ simply the *log-likelihood*.

Let us assume, for simplicity, that $K = 1$. First we give a simple test of the null hypothesis $\theta = 0$ (the individuals fail at random) against the alternative $\theta > 0$, assuming that $B_1$ (the initial group) and $N$ (the duration of the game) are known in advance, that $B_{n+1} = B_n \setminus \{b_n\}$ for all $n$ (individuals are only removed when they fail, and none are added), and that the $z_n^i$ do not depend on $n$ and are known in advance ($z_n^i = z^i$). In this case, $N \leq |B_1|$. This test is a game-theoretic formulation of the test given in Cox and Oakes [65], §7.4.

For any sequence $i_1, \dots, i_N$ of different elements of $B_1$, set

$$T(i_1, \dots, i_N) := \sum_{n=1}^{N} \left( z^{i_n} - \frac{1}{|R_n|} \sum_{i \in R_n} z^i \right), \tag{8.36}$$

where $R_n := B_1 \setminus \{i_1, \dots, i_{n-1}\}$. The intuition behind $T(i_1, \dots, i_N)$ is that if $i_1, \dots, i_N$ were the individuals who failed in the order of failing ($(b_1, \dots, b_N) = (i_1, \dots, i_N)$), then $T(i_1, \dots, i_N)$ would measure how much the covariate for the failed individuals exceeded the average covariate for the individuals who were at risk at the moment of each failure. Under the null hypothesis $\theta = 0$, we expect $T$ to be neither excessively large nor excessively small, whereas under the alternative $\theta > 0$ we expect $T$ to be large (individuals with large $z$ will tend to fail). Under our simplifying assumptions, the failure model protocol is essentially a special case of the probability forecasting protocol on p. 181, and so, by Proposition 8.6, we obtain probabilities for $T$ that are both game-theoretic and measure-theoretic: For any threshold $u \in \mathbb{R}$,

$$\mathbb{P}\{T(b_1, \dots, b_N) \geq u\} = \frac{\#\{(i_1, \dots, i_N) \mid T(i_1, \dots, i_N) \geq u\}}{\#\{(i_1, \dots, i_N)\}},$$

where $(i_1, \dots, i_N)$ ranges over all sequences of length $N$ of distinct elements of $B_1$. We can use this probability distribution for $T$ to test the hypothesis $\theta = 0$. We might choose some conventional significance level $\alpha$ (such as 1%), define $u_\alpha$ as the smallest $u$ for which $\overline{\mathbb{P}}\{T(b_1, \dots, b_N) > u\} \leq \alpha$, and reject the hypothesis (at level $\alpha$) if $T(b_1, \dots, b_N) > u_\alpha$ for the observed Reality's moves. Alternatively, p-values can be calculated, exactly as in the standard measure-theoretic statistics (see, e.g., [64]).

We can also apply the law of the iterated logarithm for the bounded forecasting game (see (5.8) on p. 104):

**Corollary 8.6** *Suppose $K = 1$, Reality always chooses nonempty groups $B_n$ and covariates $z_n^i$ bounded by some constant $C$ in absolute value, she never stops, and Model follows Cox's strategy with $\theta = 0$. Then*

$$\limsup_{N \to \infty} \frac{T_N}{\sqrt{2C^2 N \ln \ln N}} \leq 1,$$

*where (cf. (8.36))*

$$T_N := \sum_{n=1}^{N} \left( z_n^{b_n} - \frac{1}{|B_n|} \sum_{i \in B_n} z^i \right).$$

Other limit theorems studied in this book can also be applied to Cox's model, but these do not answer the asymptotic questions about the model that are most interesting to statisticians. These questions have been widely explored within the measure-theoretic framework and also deserve exploration in the game-theoretic framework. For example (assuming $K = 1$):

- Let $U(\theta)$ (the *efficient score*) be the derivative of the log-likelihood $L(\theta)$ and $-T(\theta)$ be the second derivative of $L(\theta)$ (for very intuitive explicit expressions, see [62], p. 191). One can test the null hypothesis $\theta = \theta_0$ if $U(\theta_0)$ is asymptotically Gaussian with zero mean and variance $T(\theta_0)$—more precisely, if the random variable $X := U(\theta_0)/\sqrt{T(\theta_0)}$ is close to standard Gaussian for large $N$. (In particular, this provides another way of testing $\theta = 0$; cf. [62], §2.) An open problem is to prove some variant of this in the game-theoretic framework. One might, for instance, prove that if Model plays Cox's strategy with $\theta = \theta_0$ and the path is required to satisfy some natural conditions, then the upper probability of the event $X \geq C$ is close to $\int_C^\infty \mathcal{N}_{0,1}$.

- It is often assumed that $Y := (\hat{\theta} - \theta_0)\sqrt{T(\hat{\theta})}$, where $\hat{\theta} := \arg\max_\theta L(\theta)$ is the maximum likelihood estimate, is approximately standard Gaussian under $\theta = \theta_0$ (cf. (11) in [63]); this assumption can be used for finding confidence limits for $\theta$. Again an open problem is to prove some version of this in the game-theoretic framework.

## 8.5 APPENDIX: VILLE'S THEOREM

Ville's theorem interprets probabilities in an arbitrary filtered probability space in terms of measure-theoretic martingales. Although quite elementary, the theorem is not usually discussed in expositions of the measure-theoretic framework, because it has no particular role to play there. The whole point of the measure-theoretic framework is to make probability measures, which give probabilities directly, one's starting point, and so there is no reason to dwell on the point that martingales could be taken as the starting point instead. The theorem is clearly important, however, for understanding the relationship of the measure-theoretic framework with the game-theoretic framework, where capital processes (martingales in the symmetric case) are the starting point.

Although we speak of it as one theorem, Ville's theorem consists of two claims, one for positive probabilities and one for zero probabilities. We stated the two claims roughly and informally in Chapter 2 (p. 52):

- The probability of an event is the smallest possible initial value for a nonnegative measure-theoretic martingale that reaches or exceeds one if the event happens.

- An event has probability zero if and only if there is a nonnegative martingale that diverges to infinity if the event happens.

They are made precise by Propositions 8.13 and 8.14, respectively.

**Lemma 8.1** *Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ is a filtered probability space, and suppose $E \in \mathcal{F}_\infty$. If $\mathbb{P} E < \epsilon$ for some $\epsilon > 0$, then there is a nonnegative measure-theoretic martingale $\mathcal{L}_0, \mathcal{L}_1, \ldots$ such that*

$$\mathbb{E}\,\mathcal{L}_0 < \epsilon \;\&\; \liminf_{n\to\infty} \mathcal{L}_n \geq 1 \text{ on } E.$$

*Proof* The union of all $\mathcal{F}_i$, $i = 0, 1, \ldots$, is an algebra (i.e., is closed under the usual Boolean operations). The distribution $\mathbb{P}$ can be extended from this algebra to the $\sigma$-algebra $\mathcal{F}_\infty$ uniquely by Carathéodory's construction (see, e.g., [25]). This construction shows that there is a sequence $E_0 \in \mathcal{F}_0$, $E_1 \in \mathcal{F}_1$,... of disjoint events such that $E \subseteq \cup_i E_i$ and $\sum_i \mathbb{P} E_i < \epsilon$. For $i = 1, 2, \ldots$, define a measure-theoretic martingale $\mathcal{L}_0^i, \mathcal{L}_1^i, \ldots$ by taking $\mathcal{L}_n^i$ to be $\mathbb{I}_{E_i}$ if $n \geq i$ and to be an arbitrary version of $\mathbb{E}[\mathbb{I}_{E_i} | \mathcal{F}_n]$ otherwise. (Since $\mathbb{I}_{E_i}$ is a version of $\mathbb{E}[\mathbb{I}_{E_i} | \mathcal{F}_n]$ when $n \geq i$, this means that $\mathcal{L}_n^i$ is always a version of $\mathbb{E}[\mathbb{I}_{E_i} | \mathcal{F}_n]$.) Finally, set

$$\mathcal{L}_n := \sum_{i=1}^{\infty} \mathcal{L}_n^i.$$

This sequence of nonnegative extended (meaning that they can, *a priori*, take infinite values) random variables satisfies

$$\mathbb{E}\,\mathcal{L}_0 = \sum_i \mathbb{E}\,\mathcal{L}_0^i = \sum_i \mathbb{P} E_i < \epsilon;$$

it is easy to see that, since all $\mathcal{L}_n^i$ are nonnegative measure-theoretic martingales, $\mathcal{L}_n$ is an almost surely finite nonnegative measure-theoretic martingale. Every $\omega \in E$ belongs to some $E_i$ and therefore satisfies

$$\liminf_n \mathcal{L}_n(\omega) \geq \liminf_n \mathcal{L}_n^i(\omega) = \mathbb{I}_{E_i}(\omega) = 1. \qquad \blacksquare$$

**Proposition 8.13** *Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ is a filtered probability space, and suppose $E \in \mathcal{F}_\infty$. Then*

$$\mathbb{P} E = \inf\left\{\mathbb{E}\,\mathcal{L}_0 \,\middle|\, \liminf_{n\to\infty} \mathcal{L}_n \geq \mathbb{I}_E\right\} = \inf\left\{\mathbb{E}\,\mathcal{L}_0 \,\middle|\, \sup_n \mathcal{L}_n \geq \mathbb{I}_E\right\}, \tag{8.37}$$

*where $\mathcal{L}$ ranges over measure-theoretic martingales.*

*Proof* It is clear that

$$\inf\left\{\mathbb{E}\,\mathcal{L}_0 \,\middle|\, \liminf_{n\to\infty} \mathcal{L}_n \geq \mathbb{I}_E\right\} \geq \inf\left\{\mathbb{E}\,\mathcal{L}_0 \,\middle|\, \sup_n \mathcal{L}_n \geq \mathbb{I}_E\right\},$$

so it suffices to prove

$$\mathbb{P} E \leq \inf\left\{\mathbb{E}\,\mathcal{L}_0 \,\middle|\, \sup_n \mathcal{L}_n \geq \mathbb{I}_E\right\} \tag{8.38}$$

and

$$\inf\left\{\mathbb{E}\,\mathcal{L}_0 \,\middle|\, \liminf_{n\to\infty} \mathcal{L}_n \geq \mathbb{I}_E\right\} \leq \mathbb{P} E. \tag{8.39}$$

But (8.38) follows from Doob's inequality ([287], Theorem VII.3.1), and (8.39) follows from Lemma 8.1. ∎

We are most interested in Proposition 8.13 in the case where $\mathcal{F}_0 = \{\emptyset, \Omega\}$. In this case, $\mathcal{L}_0$ is a constant for every measure-theoretic martingale $\mathcal{L}$, and so we can write

$$\mathbb{P}\{E\} = \inf\left\{\mathcal{L}_0 \,\middle|\, \liminf_{n\to\infty} \mathcal{L}_n \geq \mathbb{I}_E\right\}.$$

Informally, this says that $\mathbb{P}\,E$ is the smallest possible initial value for a measure-theoretic martingale that eventually equals or exceeds 1 if $E$ happens and zero otherwise. By (8.1), if a measure-theoretic martingale is eventually nonnegative no matter what happens it can be made always nonnegative by changes on a set of measure zero, and so this can also be put in the form we gave earlier: $\mathbb{P}\,E$ is the smallest possible initial value for a nonnegative measure-theoretic martingale that eventually equals or exceeds 1 if $E$ happens.

Proposition 8.13 applies in particular to the case where the probability is zero; it tells us that $\mathbb{P}\,E = 0$ if and only if there is a nonnegative measure-theoretic martingale that multiplies its initial value by an arbitrarily large factor if $E$ happens. Using a limiting argument, we could further conclude that $\mathbb{P}\,E = 0$ if and only if there is a nonnegative measure-theoretic martingale that becomes infinite if $E$ happens. It is simpler, however, to prove this directly, as follows.

**Proposition 8.14** *Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ is a filtered probability space, and suppose $E \in \mathcal{F}$.*

1. *If a nonnegative martingale diverges to infinity when $E$ happens, then $\mathbb{P}\,E = 0$.*

2. *If $E \in \mathcal{F}_\infty$ and $\mathbb{P}\,E = 0$, then then there is a nonnegative measure-theoretic martingale that diverges to infinity if $E$ happens.*

*Proof* To prove Statement 1, consider a nonnegative measure-theoretic martingale $\mathcal{L}_n$, $n = 0, 1, \ldots$. For any positive integer $j$, $\mathcal{L}_n \mathbb{I}_{\mathcal{L}_0 \leq j}$ is also a nonnegative measure-theoretic martingale and satisfies, by Doob's inequality,

$$\mathbb{P}\left\{\sup_n \mathcal{L}_n \mathbb{I}_{\mathcal{L}_0 \leq j} \geq C\right\} \leq \frac{j}{C},$$

for any constant $C > 0$. So

$$\mathbb{P}\left\{\sup_n \mathcal{L}_n = \infty \text{ and } \mathcal{L}_0 \leq j\right\} = \mathbb{P}\left\{\sup_n \mathcal{L}_n \mathbb{I}_{\mathcal{L}_0 \leq j} = \infty\right\} = 0.$$

Since the union of $\{\mathcal{L}_0 \leq j\}$ over all positive integers $j$ is the certain event, it follows that $\mathbb{P}\{\sup_n \mathcal{L}_n = \infty\} = 0$.

To prove Statement 2, notice that for each positive integer $k$ there exists, by Lemma 8.1, a nonnegative measure-theoretic martingale $\mathcal{L}_0^k, \mathcal{L}_1^k, \ldots$ such that $\mathbb{E}(\mathcal{L}_0^k) < 2^{-k}$ and $\liminf_n \mathcal{L}_n^k \geq 1$ on $E$. The sequence $\mathcal{L}_0, \mathcal{L}_1, \ldots$ defined by

$$\mathcal{L}_i := \sum_{k=1}^{\infty} \mathcal{L}_i^k,$$

since it satisfies $\mathbb{E}\,\mathcal{L}_0 < 1$, is finite almost surely and hence defines a nonnegative measure-theoretic martingale that is divergent on $E$. ∎

We used Statement 1 in the derivation of the measure-theoretic strong laws in §8.1; as we noted there, it follows immediately from Doob's convergence theorem. Statement 2 is trivially true if we assume that $\mathcal{F}_0$ already includes all of $\mathbb{P}$'s null sets, because then we can simply set $\mathcal{L}_n := n\,\mathbb{I}_E$, where $\mathbb{I}_E$ is the indicator of $E$ [262]. This is one reason Ville's theorem does not get much attention in standard measure-theoretic expositions. But our proof establishes that the statement holds no matter how we choose $\mathcal{F}_0$—it is true, for example, when $\mathcal{F}_0 = \{\Omega, \emptyset\}$. If $\mathcal{F} = \mathcal{F}_\infty$, then the two statements together reduce to our informal statement: an event $E$ has probability zero if and only if there is a measure-theoretic martingale that tends to infinity whenever $E$ happens.

## 8.6 APPENDIX: A BRIEF BIOGRAPHY OF JEAN VILLE

Because Jean Ville's work is so important to our framework, we provide here some details about the sources of his ideas and his subsequent career. Most of this information has been provided by Pierre Crépel of the University of Lyons, who interviewed Ville in 1984 [69].

Jean Ville (1910–1988)

In France in the 1930s, as Ville told Crépel in 1984, probability was an honorable pastime for those who had already distinguished themselves in pure mathematics. It was not an appropriate subject for an aspiring mathematician's dissertation. Ville was a promising young mathematician, and Maurice Fréchet, his adviser, had put him to work on a problem in Hilbert spaces. But Ville was intrigued by the foundations of probability, and after making no progress on his assigned topic, he went to Berlin for the academic year 1933–1934, in the hope of learning more about von Mises's ideas, which Fréchet had found interesting but flawed. He found no one interested in the topic in Berlin; Hitler was dictator, and von Mises had fled. The following year, Ville went to Vienna on an Asconti-Visconti fellowship, and there he found a much more stimulating environment, especially in Karl Menger's seminar, where there was keen interest in the foundations of mathematics and economics. There, during 1934–1935, he met Constantin Carathéodory, Abraham Wald, and Kurt Gödel, and he made his discoveries about collectives and martingales. In 1936, after he returned to Paris, he published two reports on the work in the *Comptes rendus* [307, 308].

What led Ville to his insights about martingales? In 1984, he recalled knowing in his youth a relative of the woman who later became his wife, one M. Parcot, who apparently made a modest living gambling on lotteries and roulette after making laborious secret calculations. Ville had suspected that Parcot's secret was related to Laplace's idea for taking advantage of the bias in a coin without knowing whether heads or tails is favored [190], and it was this old idea that led him to consider how to make money if one knew that the frequency of heads converged to one-half from above or below, without knowing which.

After writing up his work on collectives as a dissertation, Ville turned, as Émile Borel's research assistant, to other topics. His most striking success was a new proof of John von Neumann's minimax theorem for games; he showed that von Neumann's appeal to a fixed-point theorem could be replaced with a simpler argument relying on convexity [305, 194]. Together with Wolfgang Doeblin, a brilliant young colleague who had come with his family from Germany to Paris to escape the Nazis, he organized a seminar on probability theory, which was soon taken over by Borel.

Fréchet had reluctantly agreed that the work on collectives could be Ville's dissertation, and he presented Ville's results in detail in his address in Geneva in 1937 (or at least in the published version, [130]), as part of his argument for Kolmogorov's framework and against von Mises's collectives. But still hoping that Ville would do some real mathematics (game theory also did not count), he continually delayed the defense of the dissertation. In desperation, Ville took a job in a lycée in Nantes in 1938. The defense was finally held, after Borel's intervention, in March of 1939.

Ville's book, *Étude critique de la notion de collectif*, appeared in Borel's series of monographs on probability soon after the defense. Except for an added introductory chapter, it is identical with the dissertation. The book received far less attention than it deserved, in part because of onslaught of the war. Aside from Fréchet, Borel, Lévy, and Doeblin, there were few mathematicians in Paris interested in probability theory. Doeblin died as a soldier in 1940 (shooting himself when his unit was about to be captured by the Germans). Lévy paid no attention. According to Ville's recollection in 1984, Lévy actually refused to read his dissertation because it was printed by an Italian press: "You had your thesis printed by the fascists." "I didn't have any money." "I won't read it." Lévy's reluctance to read the work of other mathematicians was notorious, but like Doeblin, he had other things to worry about. He survived the German occupation of Paris only by hiding, with false papers, in the home of his non-Jewish son-in-law [270]. It was only ten years later, when Doob put the idea of a martingale in measure-theoretic form (see p. 55), that it attracted wide interest. Ville himself did not participate in this measure-theoretic development. He never met Doob or Feller.

Ville's subsequent work was distributed across information theory, operations research, statistics, computing, and economics. He worked at the University of Poitiers starting in 1941, then at the University of Lyons starting in 1943. Passed over for a professorship at Lyons in 1947, he held several academic and industrial jobs, finally becoming professor of econometrics at the University of Paris in 1958.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# *Part II*

# *Finance without Probability*

As the branch of economics concerned with time and uncertainty, finance theory has always been infused with probability. It was a probabilist, Louis Bachelier, who first developed a mathematical model for stock-market prices. Bachelier's work, which began with his dissertation in 1900, was not noticed by economists until the late 1950s, but theoretical work on finance within economics, which might be dated from the work on portfolio selection by Markowitz and Tobin in the early 1950s, has always been equally probabilistic.

In this part of the book, we demonstrate that finance theory can profit from the purely game-theoretic understanding of probability developed in Part I. We concentrate on two central topics: the pricing of options and the efficient market hypothesis. In both cases, as we show, a purely game-theoretic understanding, without stochastic assumptions, is both possible and illuminating.

Since the 1970s, the theory and practice of the pricing of options and other derivatives has been dominated by the Black-Scholes equation, which is based on a game-theoretic argument. The price of a derivative based on a particular stock is supposed to be the initial capital needed to reproduce the payoff of the derivative by continuously shifting capital between the stock and a risk-free bond as their relative prices change in the market. But a stochastic assumption is used in the proof that this works. One assumes that the price of the stock follows a geometric Brownian motion, and one uses the estimated variance of this geometric Brownian motion in the Black-Scholes formula for the price of the derivative. The assumption that prices follow a geometric Brownian motion is empirically shaky, and our deconstruction of the idea of stochasticity in Part I can only serve to make it less persuasive. How can we do without it?

We offer a substantive proposal that may be practical for stocks for which there is a large market in derivatives. Instead of using the estimated variance of the price of a stock to price all derivatives in the stock, we propose that one particular derivative, which pays at maturity some strictly convex function of the current stock price, be itself priced by supply and demand. The market price of this derivative can be used to determine the theoretical price of an imagined derivative that pays a continuous dividend equal to the squared relative change in the stock price, and this theoretical price can replace the estimated or predicted variance of the stock price in the Black-Scholes formula. Implementation of this proposal would require that traders shift into the new derivative from derivatives they now prefer, usually call and put options. Our theory suggests, however, that the costs of this shift may be repaid by a substantial increase in the stability and reliability of pricing.

Our game-theoretic approach to option pricing is explained in rigorous detail in Chapters 10–13. Chapters 10 and 11 lay out the basic argument, in discrete time and continuous time, respectively. In addition to the Black-Scholes model, these two chapters also treat the Bachelier model, which assumes that prices follow an ordinary Brownian motion. Because it permits negative stock prices, the Bachelier model is of little use in practice, but because it is slightly simpler, it provides a good starting point for our exposition. Later chapters demonstrate the flexibility of our approach; we can handle interest rates and jumps (Chapter 12) and American options (Chapter 13).

In Chapter 14, we show how diffusion processes can be handled within our continuous-time game-theoretic framework. In addition to making absolutely clear the strength of the assumption that prices follow a diffusion process, this may provide a starting point for a better understanding of other uses of diffusion processes.

The idea of efficient markets, which we study in Chapter 15, provides another example of the clarity that can be gained by eliminating stochasticity in favor of a game-theoretic understanding of probability. The existing literature on efficient markets begins with the simple proposition that opportunities for substantial profit-making in large capital markets are quickly eliminated by competition. This conceptual simplicity disappears when, in order to test efficiency empirically, stochastic assumptions are introduced. Our game-theoretic viewpoint allows us to recover the simplicity, for it shows us that the stochasticity is unnecessary. All that is needed for statistical testing is a variant of the fundamental interpretive assumption of probability—we assume that an investor is highly unlikely to do substantially better than the market. Since this is essentially identical with the intuitive hypothesis of market efficiency, it is fair to say that our approach to testing market efficiency requires no addition of probability ideas at all.

The next chapter, Chapter 9, provides a more detailed introduction to Part II. Because we hope that the audience for this book will include probabilists and others who have not previously studied finance theory, the tone of this chapter is relatively elementary. For details about established theory, however, we often refer the reader to the excellent introductions already available in book form. For practical information on option pricing, we most often cite Hull (2000). For mathematical details on stochastic option pricing, we usually cite Karatzas and Shreve (1991). In addition, readers may wish to consult standard textbooks and more popular accounts such as Bernstein (1992) and Malkiel (1996).

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 9

# *Game-Theoretic Probability in Finance*

In this introductory chapter, we sketch our game-theoretic approach to some basic topics of finance theory. We concentrate on how the Black-Scholes method for pricing and hedging a European option can be made purely game-theoretic, but begin with an examination of the apparently random behavior of stock-market prices, and we conclude with a discussion of informational efficiency.

The usual derivation of the Black-Scholes formula for the price of an option relies on the assumption that the market price $S(t)$ of the underlying security $\mathcal{S}$ follows a diffusion process. As we explain in the first two sections of this chapter, this stochastic assumption is used in two crucial ways:

**Taming the Market** The diffusion model limits the wildness of fluctuations in $S(t)$. This is the celebrated $\sqrt{dt}$ effect: the change in $S(t)$ over an increment of time of positive length $dt$ has the order of magnitude $(dt)^{1/2}$. This is wild enough, because $(dt)^{1/2}$ is much larger than $dt$ when $dt$ is small, but one can imagine much wilder fluctuations—say fluctuations of order $(dt)^{1/3}$.

**Averaging Market Changes** The diffusion model authorizes the use of the law of large numbers on a relatively fine time scale. The model says that relative changes in $S(t)$ over nonoverlapping time intervals are independent, no matter how small the intervals, and so by breaking a small interval of time $[t_1, t_2]$ into many even smaller intervals, we can use the law of large numbers to replace certain effects by their theoretical mean values.

Our purely game-theoretic approach does not need the diffusion model for either of these purposes.

A limit on the wildness of price changes can be expressed straightforwardly in our game-theoretic framework: as a constraint on the market, it can be listed among the rules of the game between Investor and Market. Market simply is not allowed to move too wildly. In §9.1, we discuss how such a constraint can be expressed. In a realistic discrete-time framework, it can be expressed in terms of the variation spectrum of the price series $S(t)$. In a theoretical continuous-time framework, where clearer and more elegant theorems can be proven, it can be expressed in terms of the variation exponent or the Hölder exponent of $S(t)$.

The notion that the market is constrained in how it can change prices should, of course, be taken with a grain of salt. The market can do what it wants. But no theory is possible without some regularity assumptions, and the game-theoretic framework can be credited for making clear the nature of the assumptions that the Black-Scholes argument requires.

As we explain in §9.2, the use of the law of large numbers on a fine time scale is more problematic. The problem is that the hedging of options must actually be implemented on a relatively coarse time scale. Transaction costs limit the frequency with which it is practical or desirable to trade in $S$, and the discreteness of actual trades limits the fineness of the scale at which the price process $S(t)$ is even well defined. In practice, the interval $dt$ between adjustments in one's holdings of $S$ is more likely to be a day than a millisecond, and this makes the appeal to the law of large numbers in the Black-Scholes argument appear pollyannaish, for this appeal requires the total change in $S(t)$ to remain negligible during enough $dt$s for the law of large numbers to do its work.

In our judgment, the appeal to the law of large numbers is the weak point of the Black-Scholes method and may be partly responsible for the substantial departures from the Black-Scholes formula often observed when option prices are determined by supply and demand. In any case, the appeal is unpersuasive in the game-theoretic framework, and in order to eliminate it, we need a significant change in our understanding of how options should be priced and hedged.

When time is measured in days, acceptance of the Black-Scholes use of the law of large numbers amounts, roughly speaking, to the assumption that a derivative security that pays daily dividends equal to $(dS(t)/S(t))^2$ should decrease in price linearly: its price at time $t$ should be $\sigma^2(T-t)$, where $\sigma^2$ is the variance rate of the process $S(t)$ and $T$ is the date the commitment to pay the dividends expires. We propose to drop this assumption and have the market actually price this dividend-paying derivative, the *variance derivative*, as we shall call it. As we show in §9.3, this produces a purely game-theoretic version of the Black-Scholes formula, in which the current market price of the variance derivative replaces the statistical estimate of $\sigma^2(T-t)$ that appears in the standard Black-Scholes formula. A derivative that pays dividends may not be very manageable, and in §12.2 we explain how to replace it with a more practical instrument. But the variance derivative is well suited to this chapter's explanation of our fundamental idea: cure the shortcomings of the Black-Scholes method and make it purely game-theoretic by asking the market to price $S$'s volatility as well as $S$ itself.

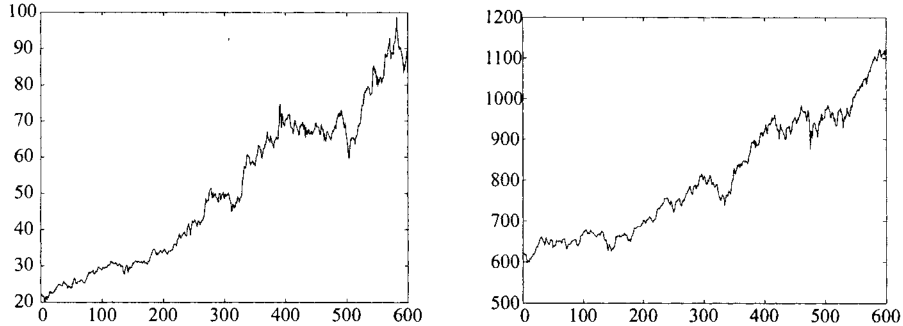



***Fig. 9.1*** The graph on the left shows daily closing prices, in dollars, for shares of Microsoft for 600 working days starting January 1, 1996. The graph on the right shows daily closing values of the S&P 500 index for the same 600 days.

After discussing stochastic and game-theoretical models for the behavior of stock-market prices and their application to option pricing, we move on, in §9.4, to a more theoretical topic: the efficient-market hypothesis. This section serves as an introduction to Chapter 15.

In an appendix, §9.5, we discuss various ways that the prices given by the Black-Scholes formula are adjusted in practice to bring them into line with the forces of supply and demand. In some cases these adjustments can be thought of as responses to the market's pricing of the volatility of $S(t)$. So our proposal is not radical with respect to practice.

In a second appendix, §9.6, we provide additional information on stochastic option pricing, including a derivation of the stochastic differential equation for the logarithm of a diffusion process, a statement of Itô's lemma, and a sketch of the general theory of risk-neutral valuation.

## 9.1 THE BEHAVIOR OF STOCK-MARKET PRICES

The erratic and apparently random character of changes in stock-market prices has been recognized ever since the publication of Louis Bachelier's doctoral dissertation, *Théorie de la Speculation*, in 1900 [9, 60]. To appreciate just how erratic the behavior of these prices is, it suffices to glance at a few time-series plots, as in Figure 9.1.

Bachelier proposed that the price of a stock moves like what is now called *Brownian motion*. This means that changes in the price over nonoverlapping intervals of time are independent and Gaussian, with the variance of each price change proportional to the length of time involved. Prominent among the several arguments Bachelie gave for each change being Gaussian was the claim that it is sum of many smaller changes, resulting from many independent influences; the idea that the Gaussian distribution appears whenever many small independent influences are at play was already widely accepted in 1900 ([3], Chapter 6).

A important shortcoming of Bachelier's model, which he himself recognized, is that it allows the price to become negative. The price of a share of a corporation cannot be negative, because the liability of the owner of the share is limited to what he pays to buy it. But this difficulty is easily eliminated: we may assume that the logarithm of the share price, rather than the price itself, follows a Brownian motion. In this case, we say that the price follows a *geometric Brownian motion.*

Norbert Wiener (1894–1964) at MIT in the 1920s. He was the first to put Brownian motion into the measure-theoretic framework.

In this section, we study how these stochastic models constrain the jaggedness of the path followed by the prices. In the next two chapters, we will use what we learn here to express these constraints directly in game-theoretic terms—as constraints on Market's moves in the game between Investor and Market. As we will see, this has many advantages. One advantage is that we can study the constraints in a realistically finite and discrete setting, instead of relying only on asymptotic theory obtained by making the length of time between successive price measurements or portfolio adjustments infinitely small. A second advantage is that we cannot avoid acknowledging the contingency of the constraints. At best, they are expectations based on the past behavior of Market, or perhaps on our understanding of the strategic interaction among the many players who comprise Market, and Market may well decide to violate them if Investor, perhaps together with some of these other players, puts enough money on the table.

Our main tool for describing constraints on Market's moves in a discrete-time game between Market and Investor is the variation spectrum. We define the variation spectrum in this section, and we explain how it behaves for the usual stochastic models and for typical price series such as those in Figure 9.1. We also discuss the variation and Hölder exponents. These exponents can be defined only asymptotically, but once we understand how they are related to the variation spectrum, which is meaningful in the discrete context, we will be able to relate continuous-time theory based on them to discrete games between Investor and Market.

### Brownian Motion

Five years after Bachelier published his dissertation, which was concerned with the motion of stock prices, the physicists Albert Einstein (1879–1955) and Marian von Smoluchowski (1872–1917) proposed the same model for the motion of particles suspended in a liquid (Einstein 1905, Smoluchowski 1906; see also [45, 110, 318]).

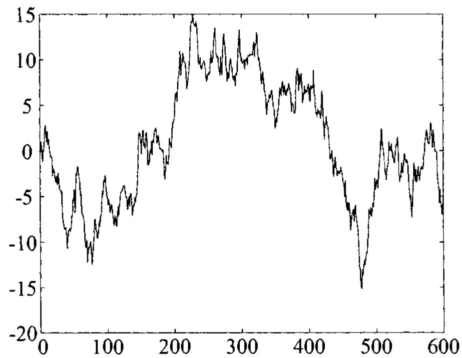



***Fig. 9.2*** A realization of the Wiener process. Each of the 600 steps in this graph are independent, with mean zero and variance one.

Because experimental study of the motion of such particles had been initiated by the naturalist Robert Brown in 1828, and because Einstein's and Smoluchowski's predictions were promptly verified experimentally, the stochastic process that Bachelier, Einstein, and Smoluchowski studied became known as *Brownian motion*.

Intuitively, Brownian motion is a continuous limit of a random walk. But as Norbert Wiener showed in the early 1920s, it can be described directly in terms of a probability measure over a space of continuous paths [102, 346]. As Wiener showed, it is legitimate to talk about a random real-valued continuous function $W$ on $[0, \infty)$ such that

- $W(0) = 0$,
- for each $t > 0$, $W(t)$ is Gaussian with mean zero and variance $t$, and
- if the intervals $[t_1, t_2]$ and $[u_1, u_2]$ do not overlap, then the random variables $W(t_2) - W(t_1)$ and $W(u_2) - W(u_1)$ are independent.

We now call such a random function a *Wiener process*. It is nowhere differentiable, and its jaggedness makes it similar to observed price series like those in Figure 9.1. If $dt$ is a small positive number, then the increment $W(t + dt) - W(t)$ is Gaussian with mean zero and variance $dt$. This means in particular that its order of magnitude is $(dt)^{1/2}$; this is the $\sqrt{dt}$ effect.

One realization of the Wiener process (one continuous path chosen at random according to the probabilities given by the Wiener measure) is shown in Figure 9.2. Actually, this figure shows a random walk with 600 steps up or down, each Gaussian with mean zero and variance one. In theory, this is not the same thing as a realization of the Wiener process, which has the same jagged appearance no matter how fine the scale at which it is viewed. But this finer structure would not be visible at the scale of the figure.

### Diffusion Processes

In practice, we may want to measure Brownian motion on an arbitrary scale and allow the mean to change. So given a Wiener process $W$, we call any process $S$ of the form

$$S(t) = \mu t + \sigma W(t), \tag{9.1}$$

where $\mu \in \mathbb{R}$ and $\sigma \geq 0$, a *Brownian motion*. The random variable $S(t)$ then has mean $\mu t$ and variance $\sigma^2 t$. We call $\mu$ the *drift* of the process, we call $\sigma$ its *volatility*, and we call $\sigma^2$ its *variance*.

Equation (9.1) implies that any positive real number $dt$, we may write

$$dS(t) = \mu dt + \sigma dW(t), \tag{9.2}$$

where, as usual, $dS(t) := S(t+dt) - S(t)$ and $dW(t) := W(t+dt) - W(t)$. When $dt$ is interpreted as a infinitely small number rather than an ordinary real number, this is called a *stochastic differential equation* and is given a rigorous meaning in terms of a corresponding stochastic integral equation (see §9.6).

We obtain a wider class of stochastic processes when we allow the drift and volatility in the stochastic differential equation to depend on $S$ and $t$:

$$dS(t) = \mu(S(t), t)dt + \sigma(S(t), t)dW(t). \tag{9.3}$$

Stochastic processes that satisfy stochastic differential equations of this form are called *diffusion processes*, because of the connection with the diffusion equation (also known as the heat equation; see p. 128) and because the random walk represented by $W$ diffuses the probabilities for the position of the path as time goes on.

A diffusion process $S$ that satisfies (9.3) has the Markov property: the probabilities for what $S$ does next depend only on the current state, $S(t)$. We can generalize further, allowing the drift and volatility to depend on the whole preceding path of $S$ rather than merely the current value $S(t)$. This produces a wider class of processes, which are often called *Itô processes* in honor of the Japanese mathematician Kiyosi Itô. But in this book, we consider only the Markov case, (9.3).

Itô developed a calculus for the study of stochastic differential equations, the *stochastic calculus*. The centerpiece of this calculus is Itô's lemma, which allows us write down, from knowledge of the stochastic differential equation governing an Itô process $S$, the stochastic differential equation governing the Itô process $U(S)$, where $U$ is a well-behaved function. We state Itô's lemma in §9.5, and we prove a game-theoretic variant in §14.2. But in the body of this chapter we avoid Itô's lemma in favor of more direct heuristic arguments, whose robustness is more easily analyzed when we relax the rather strong assumptions that define the Wiener process.

The measure-theoretic account of continuous-time stochastic processes is essentially asymptotic: it shows us only what is left in the limit as the time steps of the underlying random walk (represented by $W$) are made shorter and shorter. Although it makes for beauty and clarity, the asymptotic character of the theory can create difficulties when we want to gauge the value and validity of applications to phenomena such as finance, which are discrete on a relatively macroscopic level. One of our goals in this book is to develop less asymptotic tools, whose relevance to discrete phenomena can be understood more directly and more clearly.

## Osborne's Log-Gaussian Model (Geometric Brownian Motion)

Although Bachelier continued publishing on probability and finance through the 1930s, his fellow probabilists gave his work on probability little prominence and ignored his work on finance [60]. Consequently, the British and American statisticians and economists who began examining stock-market prices empirically and theoretically in the middle of the twentieth century had to rediscover for themselves the relevance of the Wiener process. The principal contribution was made in 1959 by the American astrophysicist M. F. Maury Osborne, who was the first to publish a detailed study of the hypothesis that $S(t)$ follows a geometric Brownian motion. This has been called Osborne's *log-Gaussian model*, because it says that the logarithm of the price $S(t)$, rather than $S(t)$ itself, follows a Brownian motion [237, 337].

If $\ln S(t)$ follows a Brownian motion, then we may write

$$d\ln S(t) = \mu_0 dt + \sigma_0 dW(t).$$

It follows (see p. 231) that the relative increments $dS(t)/S(t)$ satisfy a stochastic differential equation of the same form:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t). \tag{9.4}$$

So $S$ itself is also a diffusion process:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t). \tag{9.5}$$

The stochastic differential equation (9.5) will be the starting point for our review in the next section of the Black-Scholes model for option pricing.

Figure 9.3 shows two realized paths of the random motion defined by (9.5), with parameters chosen to match the apparent trend and volatility of the Microsoft prices shown in Figure 9.1. The parameters are the same for the two paths. Both start at 22.4, the initial price of Microsoft in Figure 9.1, and for both the theoretical drift $\mu$ is 0.0024 (the average daily change in the logarithm of Microsoft's price), and the theoretical volatility $\sigma$ is 0.0197, (the square root of the average daily squared change in the logarithm of Microsoft's price). The fact that $\sigma$ is multiplied by $S(t)$ in (9.5) is visible in these graphs: the magnitude of the fluctuations tends to be larger when $S(t)$ is larger. This same tendency towards larger fluctuations when the price is higher is also evident in Figure 9.1. On the other hand, the two paths show very different overall trends; the first ends up about 50% higher than Microsoft does at the end of 600 days, while the second ends up only about $1/3$ as high. This illustrates one aspect of the relatively weak role played by the drift $\mu$ in the diffusion model; it can easily be dominated by the random drift produced by the failure of the $dW(t)$ to average exactly to zero.

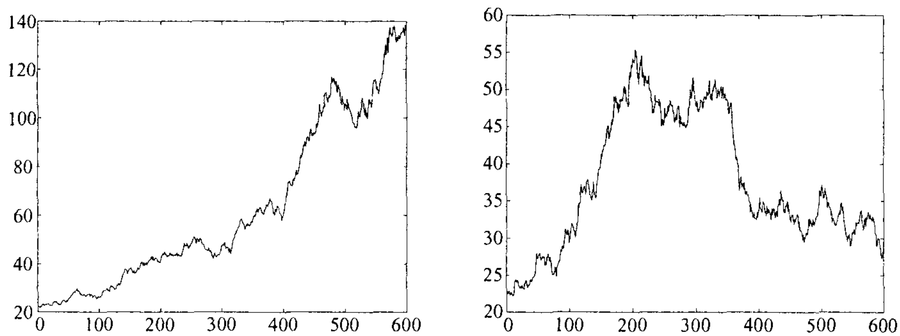
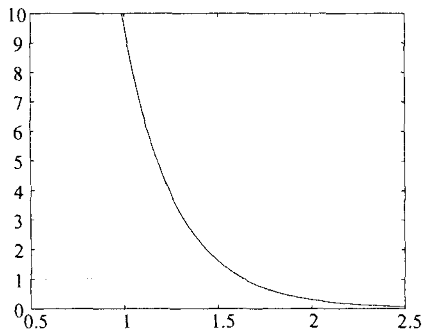



***Fig. 9.3*** Two realizations of the same geometric Brownian motion, whose parameters are chosen to match the apparent trend and volatility in the Microsoft price series in Figure 9.1: $S(0) = 22.4375$, $\mu = 0.0024$, and $\sigma = 0.0197$.

Benoit Mandelbrot (born 1924) in 1983. Since the 1960s, Mandelbrot has championed alternatives to the log-Gaussian model.

In the four decades since Osborne's formulation of the log-Gaussian model, abundant evidence has been found for deviations from it: dependencies, skewness, non-Gaussian kurtosis, and heteroskedasticity [212, 216, 208]. Some authors, most notably Benoit Mandelbrot, have argued for alternative stochastic models. But most researchers in finance have not found these alternatives attractive or useful. They have continued to take the log-Gaussian model as their starting point, making ad hoc adjustments and elaborations as necessary when empirical deviations become too troublesome.

We agree with Mandelbrot that the log-Gaussian model is a doubtful starting point. But we do not propose replacing it with a different stochastic model. Instead, we propose dropping the assumption that the market is governed by a stochastic model. We now turn to a tool that can help us do this, the variation spectrum.



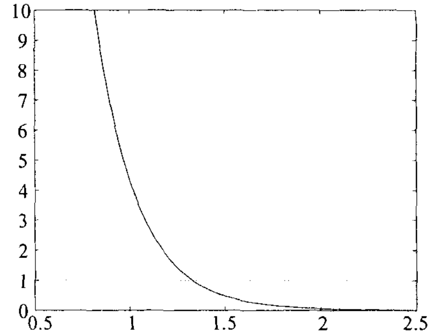




***Fig. 9.4*** Variation spectra for the two price series in Figure 9.1—Microsoft on the left and the S&P 500 on the right. In both cases, we have rescaled the data in Figure 9.1 by taking the median value as our unit. The median price for Microsoft over the 600 working days in Figure 9.1 is approximately $48.91; the median value of the S&P 500 index is approximately 775. This unit determines the vertical scale in the picture of the variation spectrum. As an example, consider $\mathbf{var}_{600}(1)$, which is the sum of the absolute values of the daily changes. In the case of Microsoft, $\mathbf{var}_{600}(1)$ is approximately 9.5, meaning that the total of the absolute sizes of the daily changes over 600 days is 9.5 times the median price. In the case of the S&P 500, $\mathbf{var}_{600}(1)$ is approximately 4.3, meaning that the total of the absolute sizes of the daily changes is 4.3 times the median value of the index.

## The Variation Spectrum

Consider a continuous function $S(t)$ on the finite interval $[0, T]$. Choose an integer $N$, set

$$x_n = S\left(n\frac{T}{N}\right) - S\left((n-1)\frac{T}{N}\right)$$

for $n = 1, \ldots, N$, and set

$$\mathbf{var}_{S,N}(p) := \sum_{n=1}^{N} |x_n|^p \tag{9.6}$$

for all $p > 0$. We call $\mathbf{var}_{S,N}(p)$ the *p-variation* of $S$, and we call the function $\mathbf{var}_{S,N}$ the *variation spectrum* for $S$. We abbreviate $\mathbf{var}_{S,N}$ to $\mathbf{var}_S$ or $\mathbf{var}_N$ or even to $\mathbf{var}$ in contexts that fix the missing parameters. If we set $dt := T/N$, then we can also write (9.6) in the form

$$\mathbf{var}_{S,N}(p) := \sum_{n=0}^{N-1} |dS(ndt)|^p, \tag{9.7}$$

where, as usual, $dS(t) := dS(t + dt) - dS(t)$.

Figure 9.4 shows the variation spectra for the Microsoft and S&P 500 data shown in Figure 9.1. These spectra display some typical features: the $p$-variation has large values for $p$ less than 1 and small values for $p$ greater than 2. This is to be expected; if $|x_n| < 1$ for all $n$, then $\mathbf{var}_{S,N}(p)$ decreases continuously in $p$, with $\lim_{p\to 0} \mathbf{var}_{S,N}(0) = N$ and $\lim_{p\to\infty} \mathbf{var}_{S,N}(p) = 0$. The fall from large to small values is gradual, however, and the apparent position of the transition depends strongly on the unit of measurement for the original series. This dependence is illustrated by Table 9.1, which gives $\mathbf{var}_{600}(2)$ for the Microsoft and S&P 500 data measured with several different units.

In many of the games between Investor and Market that we will study, the ability of Investor to hedge the sale of an option depends on Market choosing his moves $x_1, \ldots, x_N$ so that $\mathbf{var}_N(2)$ is small—or at least so that $\mathbf{var}_N(2+\epsilon)$ is small for some small positive $\epsilon$. In Chapter 10, for example, we prove that under certain conditions, which include $\mathbf{var}_N(2+\epsilon) \leq \delta$, the game-theoretic upper and lower prices of an option is approximated in discrete time by our Black-Scholes formula with an error no greater than $K\delta$ for some constant $K$. Here it makes sense that $\mathbf{var}_N$ should depend on the unit of measurement for the price of the underlying security, for the error in pricing the option is also measured using some arbitrary unit, and the constant $K$ depends on how these two units are related. But the devil is in the details. It is not enough to say that the Black-Scholes formula will be accurate if $\mathbf{var}_N(2+\epsilon)$ is small. Just how small a bound we can put on $\mathbf{var}_N(2+\epsilon)$ will determine whether the formula will give a reasonable approximation or be an asymptotic irrelevance.

## The Variation and Hölder Exponents

In addition to a practical theory, in which we wade through messy calculations to obtain bounds on the accuracy of hedging in terms of bounds on 2-variation, we do also want an asymptotic theory, in which we clear away the clutter of practical detail in order to see the big picture more clearly. This book is far from practical, and asymptotic theory will occupy most of the remaining chapters. In this asymptotic theory, we will emphasize the value of $p$ at which the $p$-variation drops from large (asymptotically infinite) to small (asymptotically zero). In practice, as we have just seen, this value is scarcely sharply defined. But an asymptotic theory will assume,

***Table 9.1*** Dependence of $\mathbf{var}_{600}(2)$ and $\mathbf{var}_{600}(2.5)$ on the choice of a unit for the Microsoft price or the S&P 500 index. The $p$-variations are given to three significant figures or three decimal places, whichever is less.

| Microsoft | | | S&P 500 | | |
|---|---|---|---|---|---|
| unit | $\mathbf{var}(2)$ | $\mathbf{var}(2.5)$ | unit | $\mathbf{var}(2)$ | $\mathbf{var}(2.5)$ |
| dollar | 742 | 1,100 | index point | 37,700 | 160,000 |
| median ($48.91) | 0.310 | 0.066 | median (775) | 0.063 | 0.010 |
| initial ($22.44) | 1.473 | 0.463 | initial (621) | 0.098 | 0.017 |
| final ($86.06) | 0.100 | 0.016 | final (1124) | 0.030 | 0.004 |

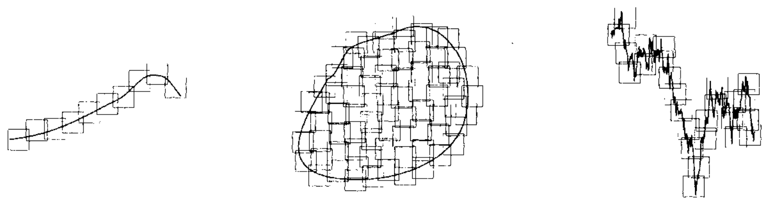


***Fig. 9.5*** Mandelbrot's concept of box dimension ([216], pp. 172–173). Intuitively, the box dimension of an object in the plane is the power to which we should raise $1/dt$ to get, to the right order of magnitude, the number of $dt \times dt$ boxes required to cover it. In the case of an object of area $A$, about $A/(dt)^2$ boxes are required, so the box dimension is 2. In the case of a smooth curve of length $T$, $T/dt$ boxes are required, so the box dimension is 1. In the case of the graph of a function on $[0, T]$ with Hölder exponent $H$, we must cover a vertical distance $(dt)^H$ above the typical increment $dt$ on the horizontal axis, which requires $(dt)^H/dt$ boxes. So the order of magnitude of the number of boxes needed for all $T/dt$ increments is $(dt)^{H-2}$; the box dimension is $2 - H$.

one way or another, that it is defined. It is called the *variation exponent*, and its inverse is called the *Hölder exponent*, in honor of Ludwig Otto Hölder (1859–1937).

We brush shoulders with the idea of a Hölder exponent whenever we talk informally about the order of magnitude of small numbers. What is the order of magnitude of $x_1, \ldots, x_N$ relative to $1/N$? If $N$ is large and the $|x_n|$ have the same order of magnitude as $(1/N)^H$ on average, where $0 < H \leq 1$, then the order of magnitude of the $p$-variation will be

$$N\left(\frac{1}{N}\right)^{Hp} = N^{1-Hp}. \tag{9.8}$$

This will be close to zero for $p > 1/H + \epsilon$ and large for $p < 1/H - \epsilon$, where $\epsilon$ is some small positive number. Intuitively, $1/H$ is the variation exponent, and $H$ is the Hölder exponent. In other words, the Hölder exponent of $S(t)$ is the number $H$ such that $dS(t)$ has the order of magnitude $(dt)^H$.

Following Mandelbrot ([215], Part II, §2.3; [216], p. 160), we also introduce a name for $2 - H$; this is the *box dimension*. Mandelbrot's rationale for this name is sketched in Figure 9.5. Although this rationale is only heuristic, the concept gives visual meaning to the Hölder exponent. In general, the box dimension of the graph of a time series falls between 1 and 2, because the jaggedness of the graph falls somewhere between that that of a line or a smooth curve and that of a graph so dense that it almost completely fills a two-dimensional area.

As Figure 9.5 suggests, an ordinary continuous function should have box dimension 1, which means that its Hölder exponent and variation exponent should also be 1. For a wide class of stochastic processes with independent increments, including all diffusion processes, we expect the Hölder exponent of a path to be $1/2$; this expresses the idea that the increment $dS$ has the order of magnitude $(dt)^{1/2}$. Functions that fall between an ordinary function and a diffusion process in their jaggedness may be called *substochastic*. These benchmarks are summarized in Table 9.2.

For a greater variety of values for the Hölder exponent, we may consider the fractional Brownian motions. The *fractional Brownian motion* with index $h \in (0,1)$ is a stochastic process $B_h$ such that $B_h(0) = 0$, values $B_h(t)$ for $t > 0$ are jointly Gaussian, and the variance of an increment $B_h(t) - B_h(s)$, where $0 < s < t$, is $(t-s)^{2h}$. If $h = 0.5$, then the fractional Brownian motion reduces to the usual Brownian motion, but for other values of $h$ even nonoverlapping increments are correlated. We can assume that the sample paths are continuous. The Hölder exponent should approximate the index $h$. Figure 9.6 shows some sample paths for different $h$. See [13, 48, 214, 216].

We use such cautious turns of phrase ("the Hölder exponent is supposed to be such and such") because there is no canonical way to define the Hölder exponent precisely. In practice, we cannot determine a Hölder exponent for an arbitrary continuous function $S(t)$ on $[0,t]$ with any meaningful precision unless we chop the interval into an absolutely huge number of increments, and even then there will be an unattractive dependence on just how we do the chopping. In order to formulate a theoretical definition, we cannot merely look at the behavior in the limit of a division of $[0,T]$ into $N$ equal parts; in general, we must pass to infinity much faster than this, and exactly how we do it matters. In the nonstochastic and nonstandard framework that we use in Chapter 11, this arbitrariness is expressed as the choice of an arbitrary infinite positive integer $N$.

The practical implication of the relative indeterminacy of the Hölder exponent is clear. Unlike many other theoretical continuous-time concepts, the Hölder exponent is not something that we should try to approximate in a discrete reality. It has a less direct meaning. When a statement in continuous-time theory mentions a Hölder exponent $H$, we can use that statement in a discrete reality only after translating it into a statement about the $p$-variation for values of $p$ at a decent distance from $1/H$. For example, the statement that Market's path will have Hölder exponent $1/2$ or less must be translated into a condition about the relative smallness of the $p$-variation for $p = 2 + \epsilon$, where $\epsilon > 0$ is not excessively small.

***Table 9.2*** Typical values for our three related measures of the jaggedness of a continuous real-valued function $S$. The figures given for a substochastic function are typical of those reported for price series by Mandelbrot.

| | Hölder exponent $H$ | variation exponent vex | box dimension dim |
|---|---|---|---|
| definition | | $\text{vex} := 1/H$ | $\dim := 2 - H$ |
| range | $0 \le H \le 1$ | $1 \le \text{vex} \le \infty$ | $1 \le \dim \le 2$ |
| ordinary function | 1 | 1 | 1 |
| substochastic | 0.57 | 1.75 | 1.43 |
| diffusion process | 0.5 | 2 | 1.5 |

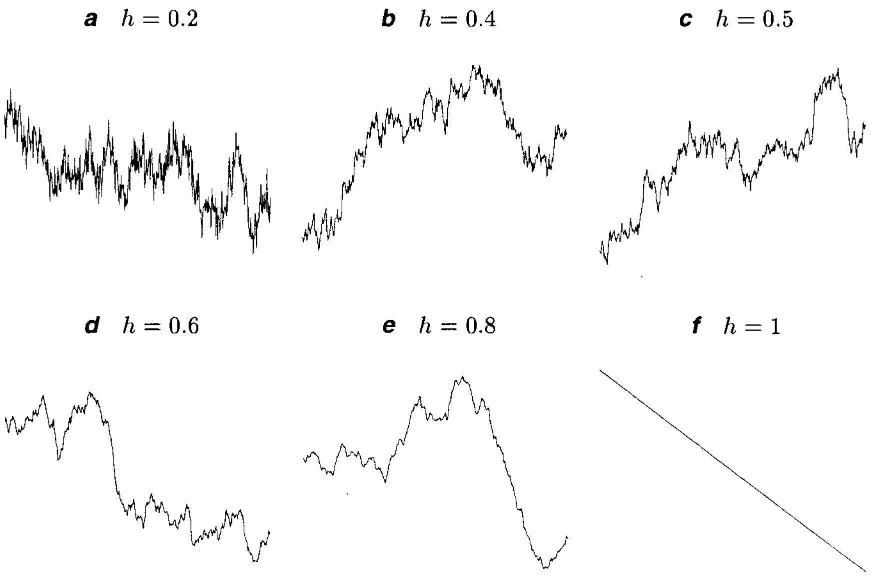



***Fig. 9.6*** Sample paths for fractional Brownian motions with different values for the index $h$. The Hölder exponent of a sample path is supposed to approximate the index $h$. When $h = 0.5$, the fractional Brownian motion reduces to ordinary Brownian motion. When $h = 1$, it reduces to a straight line.

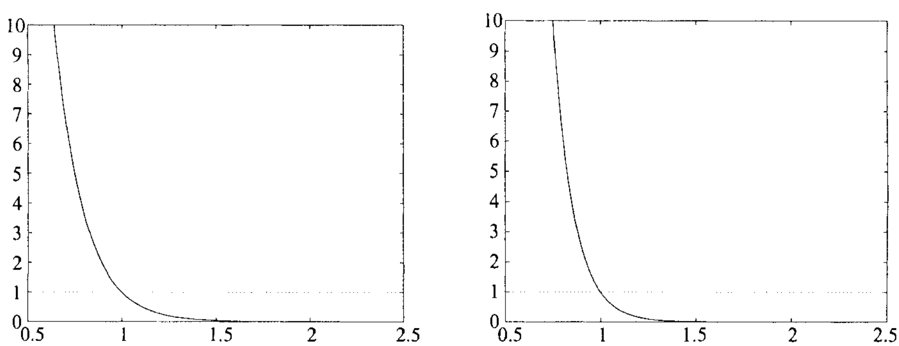



***Fig. 9.7*** Graphs of variation spectra for the straight line $S(t) = t$ on $0 \leq t \leq 1$. On the left is the variation spectrum based on a division of the interval $[0, 1]$ into 600 steps. On the right is the variation spectrum based on a division into 10,000 steps. These graphs fall from values greater than one to values less than one when we cross $p = 1$ on the horizontal axis, but the fall is less than abrupt.

In order to be fully persuasive on this point, we should perhaps look at some further examples of the indeterminacy of the Hölder exponent even when $N$ is huge. Consider the best behaved function there is: the linear function $S(t) = t$ for $0 \leq t \leq 1$. We split the interval $[0, 1]$ into $N$ parts, and (9.7) becomes

$$\mathbf{var}_{S,N}(p) := \sum_{n=0}^{N-1} \left(\frac{1}{N}\right)^p = N^{1-p}.$$

Figure 9.7 graphs this function for $N = 600$ and $N = 10,000$. As $N$ tends to infinity, $N^{1-p}$ tends to zero for $p > 1$ and to infinity for $p < 1$, confirming that the Hölder exponent for $S$ is 1. But how large does $N$ have to be in order for the variation spectrum to identify this value to within, say, 10%? For example, how large does $N$ have to be in order for

$$\mathbf{var}_{S,N}(p) \geq 10 \text{ for } p \leq 0.9 \qquad \text{and} \qquad \mathbf{var}_{S,N}(p) \leq 0.1 \text{ for } p \geq 1.1 \tag{9.9}$$

to hold? This is not hard to answer: the conditions $N^{1-0.9} \geq 10$ and $N^{1-1.1} \leq 0.1$ are both equivalent to $N \geq 10^{10}$. And even when $N$ is in the billions, there are still arbitrary choices to be made. If we had chosen our linear function on $[0, 1]$ to be $S(t) = 100t$ instead of $S(t) = t$, $\mathbf{var}_N(1)$ would not be 1, as shown in Figure 9.7 and taken for granted by our question about (9.9). Instead, it would be 100.

The story is similar for other continuous functions, including sample paths for diffusion processes: we do not get even the appearance of a sharply defined value for the Hölder exponent unless $N$ is in the billions. Figure 9.8 illustrates the point with the variation spectra for two realizations of the Wiener process, one with $N = 600$ and one with $N = 10,000$. In this case, the Hölder exponent is 0.5: the $p$-variation tends to infinity with $N$ for $p$ less than 2 and to zero for $p$ greater than 2. The scale in which the Wiener process is measured has the convenient feature that $\mathbf{var}_N(p) = 2$ for all $N$, but the drop at this point is no more abrupt than for the linear function. We should note that Figure 9.8 is not affected by sampling error; different realizations give curves that cannot be distinguished visually.

The $\mathbf{var}_{600}$ for Microsoft and the $\mathbf{var}_{600}$ for the S&P 500, both of which we displayed in Figure 9.4, are similar to the $\mathbf{var}_{600}$ for the Wiener process in Figure 9.8. But in the Microsoft and S&P 500 cases, we must use empirical scales for the data (in Figure 9.1, our unit was the observed median in each case), and so we cannot create the illusion that we can identify 2 as the crossover point from large to small $p$-variation.

## Why Do Stock-Market Prices Move Like Brownian Motion?

As we have explained, we will use the variation spectrum and the variation exponent in game-theoretic versions of Black-Scholes pricing that do not rely on the assumption that stock-market prices are stochastic. In these versions, the assumption that the stock price $S(t)$ follows a geometric Brownian motion with theoretical volatility $\sigma$ is replaced by the assumption that the market prices a derivative security that

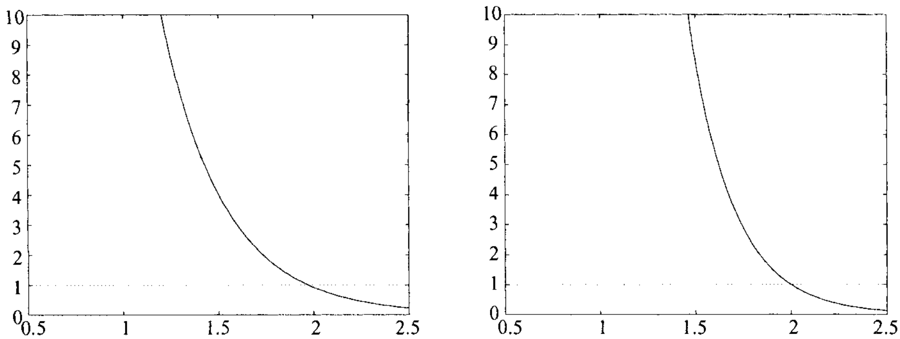



***Fig. 9.8*** The variation spectrum for a sample path of the Wiener process, for 600 and 10,000 sample points.

pays dividends that add up to the actual future (relative) variance of $S(t)$. We then assume bounds on the $p$-variations of $S(t)$ and the price $D(t)$ of the traded derivative. In discrete time, we assume upper bounds on $\mathbf{var}_S(2+\epsilon)$ and $\mathbf{var}_D(2-\epsilon)$ (Proposition 10.3, p. 249). In our continuous-time idealization, bounds on the wildness of $S(t)$ are not even needed if we can assume that the payoff of the option we want to price is always nonnegative and that $S$ does not become worthless; in this case it is enough that the variation exponent of the variance security $\mathcal{D}$ be greater than 2 (or, equivalently, that its Hölder exponent be less than 1/2) (Theorem 11.2, p. 280).

The movement of stock-market prices does look roughly like Brownian motion. But in our theory this is a consequence, not an assumption, of our market games for pricing options. This is made clear in our continuous-time theory, where we prove (Proposition 11.1, p. 281) that Market can avoid allowing Investor to become infinitely rich only by choosing his $dS(t)$ so that $S(t)$ has variation exponent exactly equal to 2. This is a way of expressing the truism that it is the operation of the market that makes price changes look like a random walk. The game-theoretic framework clears away the sediment of stochasticism that has obscured this truism, and in its discrete-time form it allows us to explore how far the effect must go in order for Black-Scholes pricing to work.

## 9.2 THE STOCHASTIC BLACK-SCHOLES FORMULA

The *Black-Scholes formula* was published in 1973, in two celebrated articles, one by Fischer Black and Myron Scholes, the other by Robert C. Merton [30, 229, 28, 16]. This formula, which prices a wide class of derivatives, has facilitated an explosive growth in the markets for options and more complex derivatives. In addition to filling the need for objective pricing (a starting point for valuation by creditors and auditors), it gives those who write (originate) options guidance on how to hedge the risks they are taking, and it permits a great variety of adjustments that can bring the ultimate pricing into line with supply and demand. It was recognized by a Nobel Prize in 1997.

As we explained in the introduction to this chapter, the Black-Scholes formula relies on the assumption that the price of the underlying stock follows a geometric Brownian motion. In this section, we review the derivation, informally but with careful attention to how the stochastic assumption is used. This will set the stage for our presentation, in the next section, of our purely game-theoretic Black-Scholes formula.

### European Options

Recall that a derivative (or, more properly, a derivative security) is a contract whose payoff depends on the future movement of the prices of one or more commodities, securities, or currencies. The derivative's payoff may depend on these prices in a complicated way, but the easiest derivatives to study are the *European options*, whose payoffs depend only on the price of a single security at a fixed date of maturity. A European option $\mathcal{U}$ on an underlying security $\mathcal{S}$ is characterized by its maturity date, say $T$, and its payoff function, say $U$. Its payoff at time $T$ is, by definition,

$$\mathcal{U}(T) := U(\mathcal{S}(T)). \qquad (9.10)$$

The problem is to price $\mathcal{U}$ at a time $t$ before $T$. What price should a bank charge at time 0, say, for a contract that requires it to pay (9.10) at time $T$?

Fischer Black (1938–1995) in 1975. Because of his early death, Black did not share in the 1997 Nobel prize for economics, which was awarded to Myron Scholes and Robert C. Merton.

The most familiar European option is the *European call*, which allows the holder to buy a security at a price set in advance. The holder of a European call on $\mathcal{S}$ with *strike price* $c$ and maturity $T$ will exercise it only if $\mathcal{S}(T)$ exceeds $c$, and he can then immediately realize the profit $\mathcal{S}(T) - c$ by reselling the security. So the payoff function $U$ of the call is

$$U(S) := \begin{cases} S - c & \text{if } S > c \\ 0 & \text{if } S \leq c. \end{cases}$$

In practice, the bank selling the option and the customer buying it usually do not bother with the underlying security; the bank simply agrees to pay the customer $U(\mathcal{S}(T))$.

At first glance, one might expect that buyers of call options would be motivated by the belief that the price of the stock $S$ will go up. This is often true in practice, especially for buyers who cannot afford to buy the stock outright. But buying the stock is a more straightforward and efficient way to bet on a price increase. The buyer of a call option will generally be charged more for his potential payoff than the interest on the money needed to buy the stock, because he is not risking the loss a buyer of the stock would incur if its price goes down instead of up. If he can afford to buy the stock but buys the stock option instead, the fact that he is paying to avoid this risk reveals that he is not really so confident that the price will go up. He may be counting only on a big change in the price. If it goes up, he will make a lot of money. If it goes down, he will lose only the relatively small price he pays for the option.

Another important European option is the European put, which entitles its holder to sell stock at a given price at a given time. It can be analyzed in the same way as we have analyzed the European call. European options are less popular, however, than American options, which can be exercised at a time of the holder's choosing. We analyze American options in Chapter 13.

### The Market Game

In the 1950s and early 1960s, when American economists first attacked the problem of pricing derivatives, they considered a number of factors that might influence the price an investor would be willing to pay, including the investor's attitude toward risk and the prospects for the underlying security [56, 114, 264]. But the formula derived by Black and Scholes involved, to their own surprise, only the time to maturity $T$, the option's payoff function $U$, the current price $S(t)$ of the underlying security $S$, and the volatility of $S$'s price. Moreover, as they came to realize in discussions with Merton, the formula requires very little economic theory for its justification. If the price of $S$ follows a geometric Brownian motion, then the price the formula gives for $\mathcal{U}$ is its game-theoretic price, in the sense of Part I, in a game between Investor and Market in which Investor is allowed to continuously adjust the amount of the stock $S$ he holds. The protocol for this game looks like this:

THE BLACK-SCHOLES PROTOCOL
**Parameters:** $T > 0$ and $N \in \mathbb{N}$; $dt := T/N$
**Players:** Investor, Market
**Protocol:**

$\mathcal{I}(0) := 0$.
Market announces $S(0) > 0$.
FOR $t = 0, dt, 2dt, \dots, T - dt$:
  Investor announces $\delta(t) \in \mathbb{R}$.
  Market announces $dS(t) \in \mathbb{R}$.
  $S(t + dt) := S(t) + dS(t)$.
  $\mathcal{I}(t + dt) := \mathcal{I}(t) + \delta(t)dS(t)$. (9.11)

Investor's move $\delta(t)$ is the number of shares of the stock he holds during the period $t$ to $t + dt$, Market's move $dS(t)$ is the change in the price per share over this period, and hence $\delta(t)dS(t)$ is Investor's gain (or loss). We write $\mathcal{I}$ for Investor's capital process.

Investor starts with zero capital, but at each step he can borrow money to buy stock or borrow stock to sell, in amounts as large as he wants. So his move $\delta(t)$ may be any number, positive, zero, or negative. Market may also choose his move positive, zero, or negative, but he cannot allow $S(t)$ to become negative. If $S(t)$ ever becomes zero, the firm issuing the stock is bankrupt, and $S(t)$ must remain zero.

For simplicity, we assume that the interest rate is zero. Thus we do not need to adjust Investor's gain, $\delta(t)dS(t)$, to account for his payment or receipt of interest. We also assume that transaction costs are zero; Investor does not incur any fees when he buys or sells stock in order to change the number of shares he is holding from $\delta(t)$ to $\delta(t + dt)$.

### The Stochastic Assumption

To the protocol we have just described, Black and Scholes added the assumption that $S(t)$ follows a geometric Brownian motion. In other words, Market's moves must obey the stochastic differential equation for geometric Brownian motion,

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t). \tag{9.12}$$

But the derivation of the Black-Scholes formula does not actually use the full force of this assumption. What it does use can be boiled down to three subsidiary assumptions:

1. **The $\sqrt{dt}$ effect.** The variation exponent of the $S(t)$ is 2. In other words, the order of magnitude of the $dS(t)$ is $(dt)^{1/2}$. This is a constraint on the wildness of Market's moves. They cannot take too jagged a path.

2. **Variance proportional to price.** The expected value of $(dS(t))^2$ just before Market makes the move $dS(t)$ is approximately $\sigma^2 S^2(t)dt$.

3. **Authorization to use of the law of large numbers.** In a calculation where the squared increments $(dS(t))^2$ are added, they can be replaced by their expected values, $\sigma^2 S^2(t)dt$. This follows from the assumption that the $dW(t)$ are independent, together with the law of large numbers, which is applicable if the time increment $dt$ is sufficiently small.

In our judgment, the most troublesome of these three assumptions is the third. The first assumption, as we explained in the preceding section, can be re-expressed in game-theoretic terms. Adjustments (more or less convincing and more or less cumbersome) can be made to correct for deviations from the second. But the third is risky, simply because the number of terms being averaged may fail to be large enough to justify the use of the law of large numbers. The new Black-Scholes method that we introduce in the next section is motivated mainly by our dissatisfaction with this risky use of the law of large numbers.

Assumption 1, that the $dS(t)$ have order of magnitude $(dt)^{1/2}$, follows, of course, from (9.12) and the fact that the $dW(t)$, as increments in a Wiener process, have this order of magnitude. The first term on the right-hand side of (9.12), $\mu S(t)dt$, can be neglected, because $dt$ is much smaller than $(dt)^{1/2}$.

Assumption 2 follows similarly when we square both sides of (9.12):

$$(dS(t))^2 = S^2(t)\left(\mu^2(dt)^2 + 2\mu\sigma dt dW(t) + \sigma^2(dW(t))^2\right). \qquad (9.13)$$

Because $dW(t)$ is of order $(dt)^{1/2}$, $(dW(t))^2$ is of order $dt$ and dominates the other two terms, and hence the approximate expected value of $(dS(t))^2$ is obtained by dropping them and replacing $(dW(t))^2$ by its expected value, $dt$.

In order to explain Assumption 3 carefully, we first note that as the square of a Gaussian variable with mean zero and variance $dt$, $(dW(t))^2$ has mean $dt$ and variance $2(dt)^2$. (The Gaussian assumption is not crucial here; the fact that the coefficient of $(dt)^2$ in the variance is exactly 2 depends on it, but this coefficient is of no importance.) This can also be expressed by writing

$$(dW(t))^2 = dt + z, \qquad (9.14)$$

where the $z$ has mean zero and variance $2(dt)^2$. In words: $(dW(t))^2$ is equal to $dt$ plus a fluctuation of order $dt$. Summing (9.14) over all $N$ increments

$$dW(0), dW(dt), dW(2dt), \ldots, dW(T - dt),$$

we obtain

$$\sum_{n=0}^{N-1}(dW(ndt))^2 = T + \sum_{n=0}^{N-1} z_n.$$

Because $\sum_{n=0}^{N-1} z_n$ has a total variance of only $2Tdt$, we may neglect it and say that the $(dW(t))^2$ add to the total time $T$; the $z_n$ cancel each other out. More generally, if the squared increments $(dW(t))^2$ are added only after being multiplied by slowly varying coefficients, such as $\sigma^2(S(t))^2$, we can still expect the $z_n$ to cancel each other out, and so we can simply replace each $(dW(t))^2$ in the sum with $dt$. Here it is crucial that the time step $dt$ be sufficiently small; before there is any substantial change in the coefficient $S^2(t)$, there must be enough time increments to average the effects of the $z_n$ to zero.

### The Derivation

Now to our problem. We want to find $\mathcal{U}(t)$, the price at time $t$ of the European option $\mathcal{U}$ that pays $U(S(T))$ at its maturity $T$. We begin by optimistically supposing that there is such a price and that it depends only on $t$ and on the current price of the stock, $S(t)$. This means that there is a function of two variables, $\overline{U}(s,t)$, such that $\mathcal{U}(t) = \overline{U}(S(t), t)$. In order to find $\overline{U}$, we investigate the behavior of its increments by means of a Taylor's series.

Considering only terms of order $(dt)^{1/2}$ or $dt$ (i.e., omitting terms of order $(dt)^{3/2}$ and higher, which are much smaller), we obtain

$$d\overline{U}(S(t),t) \approx \frac{\partial \overline{U}}{\partial s} dS(t) + \frac{\partial \overline{U}}{\partial t} dt + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(dS(t))^2. \tag{9.15}$$

Here there is one term of order $(dt)^{1/2}$—the term in $dS(t)$. There are two terms of order $dt$—the term in $dt$ and the term in $(dS(t))^2$. The terms of order $dt$ must be included because their coefficients are always positive and hence their cumulative effect (there are $T/dt$ of them) will be nonnegligible. Individually, the $dS(t)$ are much larger, but because they oscillate between positive and negative values while their coefficient varies slowly, their total effect may be comparable to that of the $dt$ terms. (We made this same argument in our heuristic proof of De Moivre's theorem in §6.2.)

Substituting the right-hand side of (9.13) for $(dS(t))^2$ in (9.15) and again retaining only terms of order $(dt)^{1/2}$ and $dt$, we obtain

$$d\overline{U}(S(t),t) \approx \frac{\partial \overline{U}}{\partial s} dS(t) + \frac{\partial \overline{U}}{\partial t} dt + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 \overline{U}}{\partial s^2}(dW(t))^2. \tag{9.16}$$

We still have one term of order $(dt)^{1/2}$ and two terms of order $dt$.

Because its coefficient in (9.16) varies slowly, we replace $(dW(t))^2$ with $dt$, obtaining

$$d\overline{U}(S(t),t) \approx \frac{\partial \overline{U}}{\partial s} dS(t) + \left(\frac{\partial \overline{U}}{\partial t} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 \overline{U}}{\partial s^2}\right) dt. \tag{9.17}$$

This is the risky use of the law of large numbers. It is valid only if the coefficient $S^2(t)\partial^2\overline{U}/\partial s^2$ holds steady during enough $dt$ for the $(dW(t))^2$ to average out. Notice that we simplified in our preliminary discussion of this point. The variability in $(dW(t))^2$'s coefficient comes from $\partial^2\overline{U}/\partial s^2$ in addition to $S^2(t)$.

Now we look again at the Black-Scholes protocol. According to (9.11),

$$d\mathcal{I}(t) = \delta(t) dS(t),$$

where $\delta(t)$ is the amount of stock Investor holds from $t$ to $t + dt$. Comparing this with (9.17), we see that we can achieve our goal by setting

$$\delta(t) := \frac{\partial \overline{U}}{\partial s}(S(t),t) \tag{9.18}$$

if we are lucky enough to have

$$\frac{\partial \overline{U}}{\partial t}(S(t),t) + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 \overline{U}}{\partial s^2}(S(t),t) = 0$$

for all $t$, no matter what value $S(t)$ happens to take.

Our problem is thus reduced to a purely mathematical one: we need to find a function $\overline{U}(s,t)$, for $0 < t < T$ and $0 < s < \infty$, that satisfies the partial differential equation

$$\frac{\partial \overline{U}}{\partial t} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 \overline{U}}{\partial s^2} = 0 \tag{9.19}$$

(this is the *Black-Scholes equation*) and the final condition

$$\overline{U}(s,t) \to U(s) \quad (t \to T).$$

As it turns out (see Chapter 11), there is a solution,

$$\overline{U}(s,t) = \int_{-\infty}^{\infty} U\left(se^z\right) \mathcal{N}_{-\sigma^2(T-t)/2,\,\sigma^2(T-t)}(dz). \tag{9.20}$$

As the reader will have noted, (9.19) differs only slightly from the heat equation, which we used in our proof of De Moivre's theorem, and (9.20) is similar to the solution of that equation. Both equations are parabolic equations, a class of partial differential equations that have been thoroughly studied and are easily solved numerically; see §6.3 and [68, 131, 352].

So an approximate price at time $t$ for the European option $\mathcal{U}$ with maturity $T$ and payoff $U(S(T))$ is given by

$$\mathcal{U}(t) = \int_{-\infty}^{\infty} U\left(S(t)e^z\right) \mathcal{N}_{-\sigma^2(T-t)/2,\,\sigma^2(T-t)}(dz). \tag{9.21}$$

This is the *Black-Scholes formula* for an arbitrary European option. (The formula is more often stated in a form that applies only to calls and puts.) We can replicate the option $\mathcal{U}$ at the price (9.21) by holding a continuously adjusted amount of the underlying security $\mathcal{S}$, the amount $\delta(t)$ held at time $t$ being given by (9.18). Financial institutions that write options often do just this; it is called *delta-hedging*.

Only one of the parameters in (9.12), the volatility $\sigma$, plays a role in the derivation we have just outlined. The other parameter, the drift $\mu$, does not appear in the Black-Scholes equation or in the Black-Scholes formula.

Most expositions simplify the argument by using Itô's lemma (p. 232). We have avoided this simplification, because Itô's lemma itself is based on an asymptotic application of the law of large numbers, and so using it would obscure just where such asymptotic approximation comes into play. As we have explained, we are uncomfortable with the application of the law of large numbers that takes us from (9.16) to (9.17), because in practice the length of time $dt$ may be equal to a day or longer, and it may be unreasonable to expect $S^2\partial^2\overline{U}/\partial s^2$ to hold steady for a large number of days.

## 9.3 A PURELY GAME-THEORETIC BLACK-SCHOLES FORMULA

We now turn to our game-theoretic version of the Black-Scholes formula. We have already explained the main ideas: (1) Market is asked to price both $\mathcal{S}$ and a derivative security $\mathcal{D}$ that pays dividends $(dS(t)/S(t))^2$, and (2) constraints on the wildness of price changes are adopted directly as constraints on Market's moves. We now explain informally how these ideas produce a Black-Scholes formula. The argument is made rigorous in the next two chapters, in discrete time in Chapter 10 and in continuous time in Chapter 11.

### Another Look at the Stochastic Derivation

Consider again the derivation of the stochastic Black-Scholes formula. It begins with a Taylor's series:

$$d\overline{U}(S(t),t) \approx \frac{\partial \overline{U}}{\partial s} dS(t) + \frac{\partial \overline{U}}{\partial t} dt + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(dS(t))^2. \tag{9.22}$$

The right-hand side of this approximation is the increment in the capital process of an investor who holds shares of two securities during the period from $t$ to $t + dt$:

- $\partial\overline{U}/\partial s$ shares of $S$, and
- $-\sigma^{-2}\partial\overline{U}/\partial t$ shares of a security $\mathcal{D}$ whose price per share at time $t$ is $\sigma^2(T-t)$ (the remaining variance of $S$), and which pays a continuous dividend per share amounting, over the period from $t$ to $t + dt$, to

  $$-\frac{\sigma^2}{\partial\overline{U}/\partial t}\frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(dS(t))^2. \tag{9.23}$$

  The second term on the right-hand side of (9.22) is the capital gain from holding the $-\sigma^{-2}\partial\overline{U}/\partial t$ shares of $\mathcal{D}$, and the third term is the total dividend.

The Black-Scholes equation tells us to choose the function $\overline{U}$ so that the dividend per share, (9.23), reduces to

$$\left(\frac{dS(t)}{S(t)}\right)^2,$$

and the increment in the capital process, (9.22), becomes

$$d\overline{U}(S(t),t) \approx \frac{\partial \overline{U}}{\partial s} dS(t) + \frac{\partial \overline{U}}{\partial t} dt - \frac{\partial \overline{U}}{\partial t}\frac{(dS(t))^2}{\sigma^2 S^2(t)}. \tag{9.24}$$

Only at this point do we need the assumption that $S(t)$ follows a geometric Brownian motion. It tells us that $(dS(t))^2 \approx \sigma^2 S^2(t)dt$, so that (9.24) reduces to

$$d\overline{U}(S(t),t) \approx \frac{\partial \overline{U}}{\partial s} dS(t) + \frac{\partial \overline{U}}{\partial t} dt - \frac{\partial \overline{U}}{\partial t} dt,$$

which will be easier to interpret if we write it in the form

$$d\overline{U}(S(t),t) \approx \frac{\partial \overline{U}}{\partial s} dS(t) - \sigma^{-2}\frac{\partial \overline{U}}{\partial t}(-\sigma^2 dt) - \sigma^{-2}\frac{\partial \overline{U}}{\partial t}(\sigma^2 dt).$$

The capital gain on each share of $\mathcal{D}$, $-\sigma^2 dt$, is cancelled by the dividend, $\sigma^2 dt$. So there is no point in holding $-\sigma^{-2}\partial\overline{U}/\partial t$ shares, or any number of shares, of $\mathcal{D}$. The increment in the capital process is simply

$$d\overline{U}(S(t), t) \approx \frac{\partial \overline{U}}{\partial s} dS(t),$$

which we achieve just by holding $\partial\overline{U}/\partial s$ shares of $\mathcal{S}$.

This way of organizing the Black-Scholes argument points the way to the elimination of the stochastic assumption. We can do without the assumption if the market really does price a security $\mathcal{D}$ whose dividend accounts for the $(dS(t))^2$ term in the Taylor's series.

### The Purely Game-Theoretic Derivation

Assume now that between 0 and $T$, Investor trades in two securities: (1) a security $\mathcal{S}$ that pays no dividends and (2) a security $\mathcal{D}$, each share of which periodically pays the dividend $(dS(t)/S(t))^2$. This produces the following protocol:

THE NEW BLACK-SCHOLES PROTOCOL
**Parameters:** $T > 0$ and $N \in \mathbb{N}$; $dt := T/N$
**Players:** Investor, Market
**Protocol:**
Market announces $S(0) > 0$ and $D(0) > 0$.
$\mathcal{I}(0) := 0$.
FOR $t = 0, dt, 2dt, \ldots, T - dt$:
Investor announces $\delta(t) \in \mathbb{R}$ and $\lambda(t) \in \mathbb{R}$.
Market announces $dS(t) \in \mathbb{R}$ and $dD(t) \in \mathbb{R}$.
$S(t + dt) := S(t) + dS(t)$.
$D(t + dt) := D(t) + dD(t)$.

$$\mathcal{I}(t + dt) := \mathcal{I}(t) + \delta(t)dS(t) + \lambda(t)\left(dD(t) + (dS(t)/S(t))^2\right). \quad (9.25)$$

**Additional Constraints on Market:** (1) $D(t) > 0$ for $0 < t < T$ and $D(T) = 0$, (2) $S(t) \geq 0$ for all $t$, and (3) the wildness of Market's moves is constrained.

Once $D$ pays its last dividend, at time $T$, it is worthless: $D(T) = 0$. So Market is constrained to make his $dD(t)$ add to $-D(0)$. We also assume, as we did in the previous section, that the interest rate is zero. We do not spell out the constraints on the wildness of Market's moves, which will take different forms in the different versions of game-theoretic Black-Scholes pricing that we will study in the next two chapters. Here we simply assume that these constraints are sufficient to justify our usual approximation by a Taylor's series.

Consider a European option $\mathcal{U}$ with maturity date $T$ and payoff function $U$. We begin by optimistically assuming that the price of $\mathcal{U}$ before $T$ is given in terms of the current prices of $\mathcal{D}$ and $\mathcal{S}$ by

$$\mathcal{U}(t) = \overline{U}(S(t), D(t)),$$

where the function $\overline{U}(s, D)$ satisfies the initial condition

$$\overline{U}(s, 0) = U(s). \tag{9.26}$$

We approximate the increment in $\mathcal{U}$'s price from $t$ to $t + dt$ by a Taylor's series:

$$d\overline{U}(S(t), D(t)) \approx \frac{\partial \overline{U}}{\partial s} dS(t) + \frac{\partial \overline{U}}{\partial D} dD(t) + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(dS(t))^2. \tag{9.27}$$

We assume that the rules of the game constrain Market's moves $dS(t)$ and $dD(t)$ so that higher order terms in the Taylor expansion are negligible.

Comparing Equations (9.25) and (9.27), we see that we need

$$\delta(t) = \frac{\partial \overline{U}}{\partial s}, \qquad \lambda(t) = \frac{\partial \overline{U}}{\partial D}, \qquad \text{and} \qquad \frac{\lambda(t)}{S^2(t)} = \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}.$$

The two equations involving $\lambda(t)$ require that the function $\overline{U}$ satisfy the partial differential equation

$$-\frac{\partial \overline{U}}{\partial D} + \frac{1}{2} s^2 \frac{\partial^2 \overline{U}}{\partial s^2} = 0$$

for all $s$ and all $D > 0$. This is the Black-Scholes equation, adapted to the market in which both $\mathcal{S}$ and $\mathcal{D}$ are traded. Its solution, with the initial condition (9.26), is

$$\overline{U}(s, D) = \int_{-\infty}^{\infty} U\left(se^z\right) \mathcal{N}_{-D/2,D}(dz).$$

This is the Black-Scholes formula for this market.

To summarize, the price for the European option $\mathcal{U}$ in a market where both the underlying security $\mathcal{S}$ and a volatility security $\mathcal{D}$ with dividend $(dS(t)/S(t))^2$ are traded is

$$\mathcal{U}(t) = \int_{-\infty}^{\infty} U\left(S(t)e^z\right) \mathcal{N}_{-D(t)/2,D(t)}(dz). \tag{9.28}$$

To hedge this price, we hold a continuously changing portfolio, containing

$$\frac{\partial \overline{U}}{\partial s}(S(t), D(t)) \text{ shares of } \mathcal{S}$$

and

$$\frac{\partial \overline{U}}{\partial D}(S(t), D(t)) \text{ shares of } \mathcal{D}$$

at time $t$.

By the argument of the preceding subsection, the derivative $\mathcal{D}$ is redundant if $S(t)$ follows a geometric Brownian motion. In this case, $\mathcal{D}$'s dividends are independent nonnegative random variables with expected value $\sigma^2 dt$. By the law of large numbers, the remaining dividends at time $t$ will add to almost exactly $\sigma^2(T - t)$, and hence this will be the market price, known in advance.

In the following chapters, we refine and elaborate the derivation of (9.28) in various ways. In Chapter 10, we derive (9.28) as an approximate price in a discrete-time game in which Market is constrained to keep $\mathbf{var}_S(2+\epsilon)$ and $\mathbf{var}_D(2-\epsilon)$ small. In Chapter 11, we derive it as an exact price in a continuous-time game in which Market is constrained to make $\operatorname{vex} D < 2$; it turns out that in this game Market is further obliged to make $\operatorname{vex} S = 2$ in order to avoid allowing Investor to become infinitely rich. As we have already noted, this gives some insight into why stock-market prices resemble diffusion processes as much as they do: the game itself pushes them in this direction. In Chapter 12, we extend the argument to the case of a known interest rate and show that we can replace the dividend-paying security $\mathcal{D}$ with a derivative that pays at time $T$ a strictly convex function of $S(T)$.

### Other Choices for the Dividend-Paying Security

The core idea of the preceding argument is to have the market price by supply and demand a derivative security $\mathcal{D}$ that pays a continuous dividend locally proportional to $\mathcal{S}$'s incremental variance, $(dS(t))^2$. We chose for $\mathcal{D}$'s dividend to be $(dS(t))^2/S^2(t)$, but this is not the only possible choice. If we take $\mathcal{D}$'s dividend to be

$$\frac{(dS(t))^2}{g(S(t))}, \tag{9.29}$$

then we obtain the partial differential equation

$$-\frac{\partial \overline{U}}{\partial D} + \frac{1}{2} g(s) \frac{\partial^2 \overline{U}}{\partial s^2} = 0$$

for the price $\overline{U}$, and there are many different functions $g(s)$ for which this equation has solutions. The choice $g(s) := s^2$, which we have just studied and will study further in the next two chapters, leads to the Black-Scholes formula. The choice $g(s) := 1$, which we will also study in the next two chapters, leads to Bachelier's formula. Bachelier's formula makes sense only if $S(t)$ can be negative, which is impossible for the price of stock in a limited-liability corporation. Powers of $s$ intermediate between 0 and 2 (as in the Constant Elasticity of Variance model of Cox and Ross 1976) also have this defect, but there are many choices for $g(s)$ that avoid it; it is sufficient that $g(s)$ go to 0 fast enough as $s$ goes to 0 ([107], p. 294).

In general, the game in which Investor can buy a derivative that pays the dividend (9.29) has as its stochastic counterpart the diffusion model

$$dS(t) = \mu(S(t), t)dt + \sqrt{g(S(t))}dW(t). \tag{9.30}$$

As we explain in an appendix (p. 231), the stochastic theory of risk-neutral valuation, which generalizes the Black-Scholes theory, tells us that if $S(t)$ follows the diffusion model (9.30), then all well-behaved derivatives are exactly priced, and the dividend (9.29), will total to exactly $T - t$ over the period from $t$ to $T$. So this diffusion model makes it redundant to trade in a derivative that pays the dividend (9.29), just as geometric Brownian motion makes it redundant to trade in a derivative that pays the dividend $(dS(t))^2/S^2(t)$.

Since the Taylor's series on which our reasoning is based is only an approximation, the local proportionality of the dividend to $(dS(t))^2$ does not need to be exact, and this suggests another possibility: $(dS(t))^2$ might be smoothed to limit the dependence on extreme values and the susceptibility of the market to manipulation by major investors. But as a practical matter, it seems more promising to take the more conventional approach we have already discussed: we ask the market to price a derivative that pays a strictly convex function of $S(T)$ at $T$, and we calculate from its price an implied price for our theoretical derivative $\mathcal{D}$ (§12.2).

## 9.4 INFORMATIONAL EFFICIENCY

The hypothesis that capital markets are informationally efficient emerged from efforts in the 1960s to give an economic explanation for the apparent randomness of prices in markets for stocks and markets for commodity futures, and it is formulated in the context of a stochastic assumption. According to this stochastic assumption, each price $p$ in such a market is based on a probability distribution for the ultimate value $x$ of the contract being priced—the discounted value of the future stream of dividends in the case of a stock, the value at delivery of the commodity in the case of a futures contract. Neglecting interest, transactions costs, and so on, the assumption is that $p$ is the expected value of $x$ conditional on certain current information. What information? Different answers are possible. The hypothesis of informational efficiency says that $p$ is the expected value of $x$ conditional on *all* information available to investors, including all the information in past prices, so that an investor cannot expect, on average, to profit from buying at current prices and selling later.

Our rejection of stochasticity obviously undercuts this whole discussion. If there is no probability distribution for $x$, then there is no point to arguing about how the market uses such a probability distribution. But as we pointed out in §1.1, our game-theoretic framework permits a much simpler interpretation of the hypothesis of informational efficiency: it is simply the hypothesis of the impossibility of a gambling strategy in a game where the imaginary player Skeptic is allowed to buy and sell securities at current prices. It says that Skeptic does not, in this setting, have a strategy that will make him very rich without risk of bankruptcy. No assumptions of stochasticity are made, and yet there are many ways of testing the hypothesis: any strategy that does not risk bankruptcy can be the basis for such a test. As we will see in Chapter 15, under certain conditions there are strategies that allow Skeptic to become rich without risk of bankruptcy if returns do not average to zero in the long-run. So tests of the stochastic hypothesis of market efficiency that check whether returns do approximately average to zero can be made into tests of our hypothesis of market efficiency.

In addition to allowing us to test market efficiency, this understanding of the market efficiency also opens the possibility of using game-theoretic probability in various contexts where established finance theory uses stochastic ideas. We explore a couple of examples in Chapter 15: the trade-off between risk and return, and the measurement of value at risk.

### Why Should Prices Be Stochastic?

Why should prices in markets for stocks and commodity futures be stochastic? In 1972, Paul Samuelson summarized the answer that comes to an economist's mind as follows ([264], p. 17):

> Expected future price must be closely equal to present price, or else present price will be different from what it is. If there were a bargain, which all could recognize, that fact would be "discounted" in advance and acted upon, thereby raising or lowering present price until the expected discrepancy with the future price were sensibly zero. It is true that people in the marketplace differ in their guesses about the future: and that is a principal reason why there are transactions in which one man is buying and another is selling. But at all times there is said to be as many bulls as bears, and in some versions there is held to be a wisdom in the resultant of the mob that transcends any of its members and perhaps transcends that of any outside jury of scientific observers. The opinions of those who make up the whole market are not given equal weights: those who are richer, more confident, perhaps more volatile, command greater voting power; but since better-informed, more-perceptive speculators tend to be more successful, and since the unsuccessful tend both to lose their wealth and voting potential and also to lose their interest and participation, the verdict of the marketplace as recorded in the record of auction prices is alleged to be as accurate ex ante and ex post as one can hope for and may perhaps be regarded as more accurate and trustworthy than would be the opinions formed by governmental planning agencies.

Samuelson did not represent this argument as his own opinion, and his tone suggests some misgivings. But he did represent it as "a faithful reproduction of similar ideas to be found repeatedly in the literature of economics and of practical finance". He cited a collection of articles edited by Cootner [56], which included a translation of Louis Bachelier's dissertation.

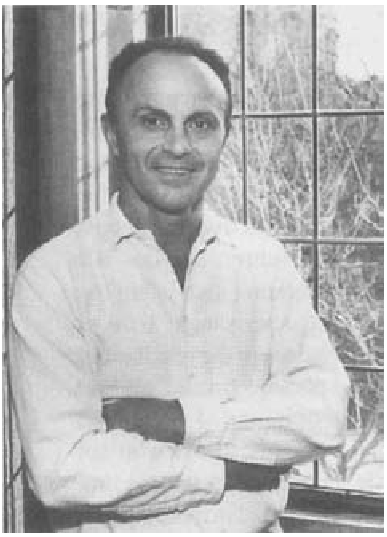
As Samuelson had observed in 1965, the assumption that the current price of a stock (or a futures contract) is the expected value of its price at some future time has a simple consequence: the successive prices of the stock will form a martingale [263]. This means that if $p_t$ is the price of a stock at time $t$, then

$$p_t = \mathbb{E}_t(p_{t+1}),$$

or

$$\mathbb{E}_t(p_{t+1} - p_t) = 0, \tag{9.31}$$

where $\mathbb{E}_t$ represents the expected value conditional on information available at time $t$. Before Samuelson's observation, economists had been investigating the hypothesis that prices follow a random walk—that is, have statistically independent increments [116]. The increments of an arbitrary martingale have only the weaker property (9.31); each has expected value zero just before it is determined. Subsequent to Samuelson's calling attention to the martingale property, economists shifted from testing for a random walk to testing (9.31), and they began saying that they are testing market efficiency.



Eugene Fama (born 1939) in 1999. His work on efficient markets has helped make him the most frequently cited professor of finance.

Tests of the stochastic efficiency of markets have spawned an immense literature, chronicled in successive reviews by Eugene Fama [117, 118, 119]. Many authors contend that the empirical results in this literature confirm that financial markets generally are efficient; as Fama put it in 1998, "the expected value of abnormal returns is zero, but chance generates deviations from zero (anomalies) in both directions" ([119], p. 284). Other authors see deviations from efficiency everywhere [288] and conclude that stock-market prices are the result of "indifferent thinking by millions of people" ([286], p. 203) that can hardly identify correct probabilities for what will happen in the future. Yet other authors have suggested that the financial markets can be considered efficient even if they do not conform exactly to a stochastic model or eliminate entirely the possibility for abnormal returns [48, 140, 207, 208].

The diversity of interpretation of the empirical results can be explained in part by the fact, acknowledged by everyone in the debate, that the efficient-markets hypothesis cannot really be tested by itself. By itself, it says only that prices are expected values with respect to some stochastic model. An effective test requires that we specify the stochastic model, substantially if not completely, and then we will be testing not merely the efficient-markets hypothesis but also specific model. This is the *joint hypothesis problem* ([48], p. 24; [118], pp. 1575–1576).

### Game-Theoretic Efficiency

Our game-theoretic efficient-market hypothesis is in the spirit of Samuelson's argument but draws a weaker conclusion. We do not suppose that there is some mysteriously correct probability distribution for future prices, and therefore we reject the words with which Samuelson's argument begins: "expected future price". But we accept the notion that an efficient market is one in which bargains have already been discounted in advance and acted upon. We hypothesize that our Skeptic cannot become rich without risking bankruptcy because any bargains providing Skeptic security against large loss would have already been snapped up, so much so that prices would have adjusted to eliminate them. By the principles of §1.3 and §8.3, this is enough to determine game-theoretic upper and lower probabilities for other events in the market being considered.

The purely game-theoretic approach obviously avoids the joint-hypothesis problem. We do not assume a stochastic model, and so we do not need to specify one in order to test our efficient-market hypothesis. We must specify, however, just what market we are talking about. Are we asserting that Skeptic cannot get rich without risking bankruptcy by trading in stocks on the New York Stock Exchange? By trading in options on the Chicago Board Options Exchange? Or by trading just in stocks in the S&P 500 index? These are all well-defined markets, and the hypothesis that Skeptic cannot get rich is a different hypothesis for each of one of them, requiring different tests and perhaps leading to different practical conclusions. Ours is an efficient-market hypothesis, not an efficient-markets hypothesis.

We must also specify a unit of measurement for Skeptic's gains—a *numéraire*. We may hypothesize that Skeptic cannot get rich relative to the total value of the market (if this is well-defined for the particular market we are considering). Or we may hypothesize that he cannot get rich in terms of some monetary unit, such as the dollar or the yen. Or we may hypothesize that he cannot get rich relative to the value of a risk-free bond. And so on. These are all different hypotheses, subject to different tests and possibly having different implications concerning what we should expect in the future.

## 9.5 APPENDIX: TWEAKING THE BLACK-SCHOLES MODEL

In practice, the Black-Scholes formula is only a starting point for pricing an option. There are a host of practical problems and many ways of adjusting the formula [27].

The first problem is that of estimating the volatility $\sigma$ of the price process $S(t)$. We can use the *historic volatility*—the standard deviation of $dS(t)$ in the historical record. Or, if an option on $S$ is traded, we can find the *implied volatility*—the value of $\sigma$ that makes the Black-Scholes formula agree with the market price for the option. Unfortunately, these methods do not always give well-defined answers. The historic volatility may vary from one period to another, and the implied volatility may vary with the type of option, the strike price, and the maturity date.

Writers of derivatives keep an eye on apparent changes in the volatility of $S(t)$ and often combine different financial instruments to hedge against such changes. The derivation of the Black-Scholes formula tells us how to hedge against changes in the share price itself; we hold $\delta(t)$ shares of $S$ at time $t$, where $\delta(t)$ is the derivative at time $t$ of the Black-Scholes price with respect to the share price. To hedge against changes in $\sigma^2$, one combines different securities so that the the derivative of the total price with respect to $\sigma^2$ is zero. This is called *vega-hedging*. Though it is widely used, it does not have a theoretical justification like that for delta-hedging.

Econometricians have developed elaborations of the Black-Scholes model that provide a theoretical basis for variations in historic and implied volatility. These models can allow changes in volatility to depend on time, on current and previous values of the security's price, and even on current and previous values of the volatility itself, while still permitting hedging by strategies that trade only in the security. Among econometricians, the most popular of these models is probably the generalized autoregressive conditional heteroskedasticity (GARCH) model. It is sufficiently developed to use in practice ([48], Chapter 12), but it has not been widely perceived as useful by practitioners [1].

One reason for the lack of interest in GARCH models among practitioners is the very fact that they recommend hedging by trading only in the underlying security. When there is enough trading in a security's derivatives to justify attention to variation in implied volatility, practitioners tend to have more confidence in the market prices of these traded derivatives than in any theory derived from them, and so they want to include the traded derivatives in their hedging strategy. This is done in practice. Typically, the practitioner inverts the Black-Scholes equation using prices for a single type of option, say a call option, obtaining the implied volatility as a function of the strike price and maturity. This volatility surface is usually unstable in time and has no more theoretical justification than vega-hedging. But it can be used to price more exotic and thinly traded options in a way that is consistent with the market pricing of the call. One then tries to replicate the exotic option by trading in the call, so as to reduce the risk taken in writing the exotic to the more familiar risk in the call ([154], §18.8; [351], Chapter 22).

Theoretically, hedging using traded derivatives can be justified by models in which the volatility is influenced by an independent source of randomness. This is *stochastic volatility* in the full sense of the term. We might assume, for example, that the volatility $\sigma$ in (9.4) follows a diffusion process independent of the Wiener process $W(t)$ in that formula. Stochastic volatility models are often able to predict future option prices better than the Black-Scholes model, but their use by practitioners has been limited by their complexity ([1]; [351], Chapter 23).

A preference for relying on the market, carried to its logical conclusion, should result in market pricing for derivatives such as our variance derivative $\mathcal{D}$ or the more conventional $R(S(T))$, where $R$ is a strictly convex function, that we propose in §12.2. But this may be slow in coming; options are bought and sold in organized exchanges because there is a demand for them, not because they are interesting to professionals. In the early 1990s, the Chicago Board Options Exchange commissioned Robert E. Whaley of Duke University to design a market volatility index to serve as the underlying for volatility options that it was contemplating introducing. The resulting CBOE Market Volatility Index (ticker symbol "VIX") is based on implied volatilities of S&P 100 index options [339, 340] and has been disseminated on a real-time basis since 1993. Although the CBOE has yet to introduce exchange-traded volatility derivatives, there is now a substantial over-the-counter market in equity volatility and variance swaps, which pay out observed volatility or variance during the period up to maturity [52, 225].

## 9.6 APPENDIX: ON THE STOCHASTIC THEORY

Kiyosi Itô (born 1915), at his desk in the national statistics bureau of Japan at the age of 27.

In this appendix, we provide some additional information on stochastic option pricing and stochastic differential equations, aimed at readers new to these topics who would like a fuller picture at the heuristic level of this chapter. First, we fill some holes in our discussion of stochastic differential equations: we explain why (9.4) represents a geometric Brownian motion, and we state Itô's lemma. Then we discuss what appears from the measure-theoretic point of view to be the general theory of option pricing: the theory of risk-neutral valuation.

### The Stochastic Differential Equation for Geometric Brownian Motion

We used the stochastic differential equation

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t) \tag{9.32}$$

as the starting point for our study of geometric Brownian motion in §9.1, instead of the differential equation

$$d \ln S(t) = \mu_0 dt + \sigma_0 dW(t), \tag{9.33}$$

which also expresses the assumption that the logarithm of $S(t)$ is a Brownian motion. To see that (9.32) implies (9.33), we may use Taylor's expansion of the logarithm to obtain

$$\begin{aligned} d \ln S(t) = \ln(S(t) + dS(t)) - \ln S(t) &= \ln\left(1 + \frac{dS(t)}{S(t)}\right) \\ \approx \frac{dS(t)}{S(t)} - \frac{1}{2}\left(\frac{dS(t)}{S(t)}\right)^2&. \end{aligned} \tag{9.34}$$

If (9.32) holds, then by (9.13) and (9.14), $(dS(t)/S(t))^2$ differs from $\sigma^2 dt$ by at most a random term of order $dt$ that makes no difference in the interpretation of the stochastic differential equation. So (9.34) becomes

$$d \ln S(t) = \mu dt + \sigma dW(t) - \frac{1}{2}\sigma^2 dt. \tag{9.35}$$

This makes clear how $\mu$ and $\sigma$ are related to the coefficients $\mu_0$ and $\sigma_0$ in (9.33): $\sigma = \sigma_0$ and $\mu = \mu_0 + \sigma_0^2/2$.

Some readers may be puzzled that approximations can produce the equals sign in (9.35). In general, however, equals signs in stochastic differential equations are meaningful only in the limit as $dt$ is made smaller. For this reason, the equations are often translated into corresponding stochastic integral equations, in which equality has its usual measure-theoretic meaning: equal except on a set of measure zero. Higher-order terms are irrelevant to the integrals.

In our game-theoretic framework, we do not necessarily have a probability measure, and hence equations that relate increments such as $dt$ and $dS(t)$ must be given a pathwise meaning. In the discrete-time case (Chapter 10), this meaning is familiar: the $dt$ and the other increments are real numbers. We handle the continuous-time case (Chapter 11) by supposing that they are infinitesimal.

## Statement of Itô's Lemma

In terms of stochastic differential equations, Itô's lemma says that if

$$dS(t) = \mu dt + \sigma dW(t)$$

and $f(s,t)$ is a well-behaved function, then

$$df(S(t),t) = \left(\mu \frac{\partial f}{\partial s}(S(t),t) + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial s^2}(S(t),t) + \frac{\partial f}{\partial t}(S(t),t)\right) dt \\ + \sigma \frac{\partial f}{\partial s}(S(t),t) dW(t).$$

(The drift $\mu$ and volatility $\sigma$ may depend on $S(t)$ and $t$, and perhaps even on other information available at time $t$.) The equivalent statement in terms of stochastic integral equations is that if

$$S(t) = S(0) + \int_0^t \mu dt + \int_0^t \sigma dW,$$

then

$$f(S(t),t) = f(S(0),0) + \int_0^t \left(\mu \frac{\partial f}{\partial s} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial s^2} + \frac{\partial f}{\partial t}\right) dt + \int_0^t \sigma \frac{\partial f}{\partial s} dW,$$

where the stochastic integral (the integral with respect to $dW$) is understood in the sense by Itô (see Itô 1944 or, e.g., [26], §5.5).

The equivalence of (9.32) and (9.33) follows directly from Itô's lemma, and the derivation of the stochastic Black-Scholes formula is simplified by it. We prove a game-theoretic variant of the lemma in §14.2.

## Risk-Neutral Valuation

Although it is central to option pricing in practice, the Black-Scholes model appears from the measure-theoretic point of view as merely one of many stochastic models that can be used to help price options, all of which fall under a general theory of continuous-time hedging, or *risk-neutral valuation*. (The name refers to the game-theoretic character of the price; although it does depend on a stochastic assumption, it does not depend on any assumption about anyone's attitude towards risk.)

Because the theory of risk-neutral valuation is part of the theory of continuous-time stochastic processes, a rigorous mathematical understanding of its main results would require a detour through several pages of definitions. (Most significantly, the definition of the capital process resulting from a strategy involves stochastic integration.) But the results are quite simple and can be stated clearly as soon as it is stipulated that the measure-theoretic concepts of process, strategy, martingale, etc., which we have studied in discrete time, can be extended to continuous time. We start with a process $S(t)$, which represents as usual the price of a security $\mathcal{S}$. Technically, this is a collection of random variables $\{S(t)\}_{t\geq0}$ in a probability space with a probability measure $\mathbb{P}$. For simplicity, we consider only $t$ from 0 to $T$. We say that the price process $S(t)$ is *arbitrage-free* if there is no well-behaved strategy for trading in $\mathcal{S}$, starting with capital zero, whose capital at time $T$ is nonnegative with probability one and positive with positive probability (we say "well-behaved" to rule out wild strategies, such as those that make unbounded bets). Then we can state the main results of the theory as follows:

**Result 1** The price process $S(t)$ is arbitrage-free if and only if there exists at least one probability measure equivalent to $\mathbb{P}$ under which $S$ is a martingale [75, 113]. (Recall that two probability measures on the same space are *equivalent* if they give positive probabilities to the same events.)

**Result 2** If there is only one such measure, say $\mathbb{Q}$, then every well-behaved European option is priced, and its price is its expected value under $\mathbb{Q}$ [184, 147].

Although the generality of these results makes them mathematically attractive, their main use in practice seems to be in the well-known examples we discuss in this book—binomial trees (discussed in §1.5), diffusion processes (discussed in this chapter), and Poisson jump processes (discussed in §12.3).

The content of the general risk-neutral results is most easily understood in a completely discrete and finite setting, where we are concerned only with a sequence of prices, say $S_0, \ldots, S_N$, for $\mathcal{S}$, and there are always only a finite number of possibilities for the change $S_{n+1} - S_n$. This puts us into a finite tree, as in the examples we considered in §1.5. We may assume that $\mathbb{P}$ gives the positive probability to all branches. In this case, the requirement that the price process be arbitrage-free is the same as the requirement that the game be coherent: in each situation $t$, the possible changes $dS$ are never all positive or all negative. This is necessary and sufficient for the existence of positive probabilities for the branches from $t$ that make the current price of $\mathcal{S}$ the expected value of its next price. This defines an equivalent martingale measure, which is unique only if the branching is always binary. This is also the condition, as we saw in §1.5, for options to be exactly priced, and their price is then their expected value under the equivalent martingale measure. Since a unique equivalent martingale measure exists in a finite tree only if it is binary, we would expect to obtain a unique equivalent martingale measure in continuous-time model only if it can be conveniently approximated by a binary tree. So it is not surprising that continuous-time risk-neutral pricing has been used mainly for the Gaussian and Poisson cases.

In the case of the diffusion model, the change from the true probability measure $\mathbb{P}$ to the equivalent martingale measure $\mathbb{Q}$ amounts to dropping the term in the stochastic differential equation that represents the drift, after which we can find option prices as expected values under $\mathbb{Q}$. We can easily reorganize the derivation in a way that makes this clear—and also applies to any diffusion model, not just geometric Brownian motion. We begin with the general stochastic differential equation for the diffusion model

$$dS(t) = \mu(S(t), t)dt + \sigma(S(t), t)dW(t).$$

As usual, we write $\overline{U}(s, t)$ for the price of the derivative at time $t$ when $S(t) = s$. Our problem is to find the function $\overline{U}$ starting with knowledge of its values when $t = T$. The problem will be solved by backwards induction if we can find the function $\overline{U}(s, t)$ of $s$ for fixed $t$ from knowledge of the function $\overline{U}(s, t + dt)$. Suppose we do know $\overline{U}(s, t + dt)$ as a function of $s$, and designate it by $U$:

$$U(s) := \overline{U}(s, t + dt).$$

Write $S$ for the price of the security at time $t$ and $S + dS$ for the price at time $t + dt$. Investor can complete his hedging of the derivative if his capital at time $t + dt$ is $U(S + dS)$. But

$$U(S + dS) \approx U(S) + U'(S)dS + \frac{1}{2}U''(S)(dS)^2. \tag{9.36}$$

(Here we use $\approx$ to represent equality up to $o(dt)$. As usual, we ignore terms of order $o(dt)$, because their cumulative effect, when summed over the $O(1/dt)$ time increments, is negligible.) He can approximately achieve the capital (9.36) at time $t + dt$ if at time $t$ he has

$$U(S) + \frac{1}{2}U''(S)\sigma^2 dt \tag{9.37}$$

(we omit the arguments $t$ and $S$ of $\sigma$ and $\mu$), because:

- he can replicate the term $U'(S)dS$ by buying $U'(S)$ shares of $S$, and
- as we showed in §9.2 using the law of large numbers and the approximation $(dS)^2 \approx \sigma^2(dW(t))^2$, the difference between the third term of (9.36) and the second term of (9.37),

  $$\frac{1}{2}U''(S)(dS)^2 - \frac{1}{2}U''(S)\sigma^2 dt,$$

  will eventually cancel out when we sum over all $t$.

So

$$\overline{U}(S, t) \approx U(S) + \frac{1}{2}U''(S)\sigma^2 dt.$$

This is just a variant of the Black-Scholes equation, (9.19), generalized to the case where $\sigma$ may depend on $S$ and $t$, but it is written in a form that shows that $\overline{U}(S,t)$ is the expected value at time $t$ of $U(S+dS)$ under the assumption that $dS$ is distributed as $\sigma dW(t)$. To find the price of an option, calculate the expected value ignoring the drift.

The generalization to the general diffusion model carries over, in large measure, to our game-theoretic approach. We explained at the end of §9.3 how it generalizes to the case where $\mu$ and $\sigma$ may depend on $S(t)$. We cannot allow direct dependence on the time $t$, but we can allow dependence on the current value $D(t)$ of the new security $\mathcal{D}$. Then instead of (9.37) we obtain

$$U(S) + \frac{1}{2}U''(S)g(S,D)dD,$$

$S$ and $D$ being the current prices of $\mathcal{S}$ and $\mathcal{D}$; no law of large numbers is needed.

To conclude this appendix, we will briefly describe an alternative approach to risk-neutral valuation based on Girsanov's theorem ([26], p. 159). We have seen that, when valuing options, we should ignore the drift of $\mathcal{S}$. The idea of the approach based on Girsanov's theorem is to change the probability distribution governing the generation of $dS(t)$ so that its expectation becomes zero; then the price of the option will be given by the expected value with respect to the modified distribution.

Let us fix some $t$ and suppose $S = S(t)$ is known. For simplicity let us assume that $dW(t)$ can only take a finite range of values, $dW_k$, $k = 1, \ldots, K$; the probability of value $dW_k$ will be denoted $p_k$. The values of the increment $dS(t)$ corresponding to different possible values $dW_k$ are

$$dS_k = \mu dt + \sigma dW_k$$

(as before, we drop the arguments of $\mu$ and $\sigma$). The new probabilities will have the form

$$\tilde{p}_k = \alpha p_k e^{-\beta dW_k}$$

for some constants $\alpha$ and $\beta$. This is a natural form, analogous to (5.9) used in the proof of the law of the iterated logarithm. The constant of proportionality $\alpha$ is determined by the requirement that the $\tilde{p}_k$ sum to 1, and $\beta$ will be chosen later so that the average drift $\sum_k \tilde{p}_k dS_k$ is zero. If we again write $U$ for the function $s \mapsto \overline{U}(s, t+dt)$, the target function for time $t+dt$, we get

$$\begin{aligned}\sum_k \tilde{p}_k U(S+dS_k) &\approx \sum_k \tilde{p}_k \left(U(S) + U'(S)dS_k + \frac{1}{2}U''(S)(dS_k)^2\right) \\ &\approx \sum_k \tilde{p}_k \left(U(S) + \frac{1}{2}U''(S)\sigma^2(dW_k)^2\right) \approx U(S) + \frac{1}{2}U''(S)\sigma^2 dt,\end{aligned}$$

which coincides with (9.37). (The final approximation relies on the fact that the average of the $(dW_k)^2$ with respect to $\tilde{p}_k$ coincides, to our usual accuracy $o(dt)$, with their average with respect to $p_k$.)

To have $\sum_k \tilde{p}_k = 1$, we need $\alpha = e^{-(\beta^2/2)dt}$:

$$\sum_k \tilde{p}_k \approx \sum_k \alpha p_k \left(1 - \beta dW_k + \frac{\beta^2 (dW_k)^2}{2}\right) = \alpha \left(1 + \frac{\beta^2 dt}{2}\right) \approx \alpha e^{(\beta^2/2)dt}.$$

Let us see now that we can indeed choose $\beta$ so that the mean drift becomes zero:

$$\sum_k \tilde{p}_k dS_k \approx \sum_k p_k \left(1 - \beta dW_k + \frac{\beta^2 (dW_k)^2}{2}\right) (\mu dt + \sigma dW_k) \approx \mu dt - \beta \sigma dt;$$

therefore, it suffices to put $\beta = \mu/\sigma$ (this ratio is sometimes called the *market price of risk*).

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# 10

# Games for Pricing Options in Discrete Time

Myron Scholes (born 1941), at a press conference at Stanford in October 1997, after the announcement that he and Robert C. Merton had been awarded the 1997 Nobel Prize for economics.

In the preceding chapter, we explained informally how an investor can hedge European options on a security $\mathcal{S}$ if the market prices both $\mathcal{S}$ and a dividend-paying derivative security $\mathcal{D}$. The argument was entirely game-theoretic. We required that the prices $S(t)$ and $D(t)$ of $\mathcal{S}$ and $\mathcal{D}$ not fluctuate too wildly, but we made no stochastic assumptions.

This chapter makes the argument mathematically precise in a realistic discrete-time setting. We prove that European options can be approximately hedged in discrete time at a price given by the Black-Scholes formula, with the market price $D(t)$ in the place of $(T - t)\sigma^2$, provided that $S(t)$ is always positive, $\mathcal{D}$ pays $(dS(t)/S(t))^2$ as a dividend, and both $S(t)$ and $D(t)$ obey certain constraints on their $p$-variations.

For historical and mathematical reasons, we preface this Black-Scholes result with an analogous result for Bachelier's formula. In the Bachelier case, $S(t)$ is not required to remain positive, and $\mathcal{D}$ pays the simpler dividend $(dS(t))^2$. Although a model that allows $S(t)$ to be negative is of little substantive interest, the mathematical simplicity of the Bachelier model allows us to see clearly the close relation between option pricing and the central limit theorem. Bachelier's formula is another central limit theorem.

We derive Bachelier's formula under two different sets of conditions. In the first derivation, in §10.1, we use conditions similar to those that we studied in Part I. These conditions are not realistic for finance, but they are similar to the martingale argument we used to prove De Moivre's theorem in Chapter 6. In the second derivation, in §10.2, we use conditions on the $p$-variations of $S$ and $D$.

We also use two different sets of conditions to derive the Black-Scholes formula. In §10.3, we use nearly the same conditions as we use for Bachelier's formula in §10.2. Then, in §10.5, we change the condition on $S(t)$ slightly; instead of constraining the $p$-variation of its absolute fluctuations, we constrain the $p$-variation of its relative (percentage) fluctuations. This approach gives simpler bounds.

Our discrete-time results are messy, involving many error terms and arbitrary decisions about how approximations will be made. In the next chapter, we will clear away the mess by passing to a limit in which the intervals between successive hedges are infinitesimal, thus obtaining elegant continuous-time counterparts for the main results of this chapter (see Table 10.1). But the details cleared away by continuous-time models must be brought back in practice. In finance, certainly, the possibility of making discrete-time calculations is fundamental. Traders usually balance their portfolios daily, and the gap between infinitely often and daily is so great that the meaningfulness of continuous-time results is in doubt unless we can demonstrate that they can be approximated in realistically coarse discrete time.

In the latter part of this chapter, in §10.4 and §10.6, we look at some stock-price data to see whether the error bounds calculated by our discrete methods are tight enough to be useful. Here we encounter two important difficulties:

- Because our dividend-paying derivative $\mathcal{D}$ is not now traded for any $\mathcal{S}$, we must speculate about the $p$-variations for $\mathcal{D}$ that will be observed if it is traded. In §10.4, we use an assumption that is surely overly optimistic: we suppose that the $p$-variations are the same as would be obtained if the market knew $\mathcal{D}$'s future dividend stream (the fluctuations in $S(t)$) in advance. In §10.6, we use also prices for $\mathcal{D}$ implied by market prices for calls and puts.

***Table 10.1*** The correspondence between the discrete-time results of this chapter and the continuous-time results of the next chapter.

| | Discrete Time | Continuous Time |
|---|---|---|
| **Bachelier**<br>absolute fluctuations in $S(t)$ | Proposition 10.2<br>p. 244 | Theorem 11.1<br>p. 278 |
| **Black-Scholes**<br>absolute fluctuations in $S(t)$ | Proposition 10.3<br>p. 249 | Theorem 11.2<br>p. 280 |
| **Black-Scholes**<br>relative fluctuations in $S(t)$ | Proposition 10.4<br>p. 260 | Theorem 11.3<br>p. 282 |

- In general, a discrete-time calculation is likely to give useful results only if it is tailored to one's particular circumstances—more tailored than any general theorem, even a messy discrete-time theorem, can possibly be. In the case of option pricing, the calculation needs to be tailored not only to the particular stock (different conditions on the fluctuations may be reasonable for different stocks) but also to the particular option and even its practical purpose, which may affect how exactly it needs to be replicated. Because we are discussing option pricing in general, not just a particular option with a particular purpose, we must settle for conclusions that involve relatively arbitrary assumptions about the smoothness of the option's payoff function.

In spite of these difficulties, it is clear from our calculations that the hedging recommended by our theory can be practical. We obtain promising results in §10.4, where we apply §10.3's method using our very optimistic assumption about $\mathcal{D}$'s $p$-variation. We obtain even more promising results in §10.6, where we apply §10.5's method to the same data and also to data for which we have prices for $\mathcal{D}$ implied by market prices for calls and puts.

These results compare very favorably with existing work on discrete-time Black-Scholes hedging, which has remained very asymptotic and therefore unconvincing. In fact, the results are so promising that we feel serious consideration should be given to implementing trading in our dividend-paying derivative or some more conventional derivative that would simulate it; see the discussion in §12.2.

## 10.1 BACHELIER'S CENTRAL LIMIT THEOREM

In this section, we derive a new central limit theorem, which we call *Bachelier's central limit theorem*, because it gives Bachelier's formula for option prices.

Bachelier's central limit theorem differs from the central limit theorems we studied in Part I in the way it handles the variances of Reality's moves. In De Moivre's and Lindeberg's central limit theorems, a variance for Reality's move $x_n$ is known before Skeptic decides how many tickets to buy. In Bachelier's central limit theorem, Skeptic has less information before he decides how many tickets to buy: he has only a total variance for all of Reality's remaining moves, $x_n, x_{n+1}, \ldots, x_N$.

As a reminder of how a Lindeberg protocol handles variance, we reproduce here, in simplified form, one of the Lindeberg protocols we discussed in §7.3.

A LINDEBERG PROTOCOL (Example 3 from Chapter 7, p. 161)
**Parameters:** $N, \mathcal{K}_0 > 0, C \geq 1, \sigma^2 > 0$
**Players:** World (Forecaster + Reality), Skeptic
**Protocol:**

FOR $n = 1, \ldots, N$:

Forecaster announces $v_n \in \mathbb{R}$.

Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.

Reality announces $x_n \in \mathbb{R}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n + V_n(x_n^2 - v_n)$.

**Constraints on World:** Forecaster must make $\sum_{i=1}^{n} v_i < \sigma^2$ for $n = 1, \ldots, N-1$, $\sum_{n=1}^{N} v_n = \sigma^2$, and $|v_n| \leq CN^{-1}$ for all $n$. Reality must make $|x_n| \leq CN^{-1/2}$ for all $n$.

In this Lindeberg protocol, Skeptic is told the total variance $\sigma^2$ at the beginning of the game. At the beginning of the $n$th round, he already knows $v_1, \ldots, v_{n-1}$, and hence he knows the total variance of Reality's remaining moves, $\sum_{i=n}^{N} v_i$. Before he makes his moves $M_n$ and $V_n$, he is also told $v_n$, and so he knows how the remaining total variance $\sum_{i=n}^{N} v_i$ is split between the current round and later rounds.

The protocol we now consider differs from this Lindeberg protocol in only one particular: Forecaster does not announce the variance $v_n$ until *after* Skeptic and Reality move:

PROTOCOL FOR BACHELIER'S CENTRAL LIMIT THEOREM
**Parameters:** $N, \mathcal{K}_0 > 0, C \geq 1, \sigma^2 > 0$
**Players:** World (Forecaster + Reality), Skeptic
**Protocol:**
FOR $n = 1, \ldots, N$:
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Reality announces $x_n \in \mathbb{R}$.
Forecaster announces $v_n \in \mathbb{R}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n x_n + V_n(x_n^2 - v_n)$.
**Constraints on World:** Forecaster must make $\sum_{n=1}^{N} v_n = \sigma^2$, $\sum_{i=1}^{n} v_i < \sigma^2$ for $n < N$, and $|v_n| \leq CN^{-1}$ for all $n$. Reality must make $|x_n| \leq CN^{-1/2}$ for all $n$.

This is a symmetric probability protocol, and it is obviously coherent in the initial situation: Forecaster and Reality can ensure that Skeptic never makes money by choosing $v_n := 1/N$ and

$$x_n := \begin{cases} N^{-1/2} & \text{if } M_n < 0 \\ -N^{-1/2} & \text{if } M_n \geq 0. \end{cases}$$

As it turns out, we can still prove the central limit theorem for this protocol: for any well-behaved function $U$ and large $N$, $\int U(z)\ \mathcal{N}_{0,\sigma^2}(dz)$ is an approximate price for $U\left(\sum_{n=1}^{N} x_n\right)$.

It is convenient to write the protocol in the terms of the cumulative sums of World's moves:

$$S_0 := 0 \quad \text{and} \quad S_n := \sum_{i=1}^{n} x_i \text{ for } n = 1, \ldots, N$$

and

$$D_n := \sum_{i=n+1}^{N} v_i \text{ for } n = 0, \ldots, N-1 \quad \text{and} \quad D_N = 0.$$

The moves $x_n$ and $v_n$ can then be represented as increments:

$$\Delta S_n := S_n - S_{n-1} = x_n \quad \text{and} \quad \Delta D_n := D_n - D_{n-1} = -v_n.$$

The quantity $D_0$ is the same as the total variance, $\sigma^2$, and instead of calling it a parameter, we may have Forecaster announce it at the beginning of the game. This puts the protocol in the following form:

BACHELIER'S PROTOCOL IN TERMS OF $D$ AND $S$
**Parameters:** $N, \mathcal{K}_0 > 0, C \geq 1$
**Players:** World (Forecaster + Reality), Skeptic
**Protocol:**
Forecaster announces $D_0 > 0$.
FOR $n = 1, \ldots, N$:
Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Reality announces $S_n \in \mathbb{R}$.
Forecaster announces $D_n \geq 0$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n \Delta S_n + V_n((\Delta S_n)^2 + \Delta D_n)$.
**Additional Constraints on World:** Forecaster must make $D_n > 0$ for $n < N$, $D_N = 0$, and $|\Delta D_n| \leq CN^{-1}$ for all $n$. Reality must make $|\Delta S_n| \leq CN^{-1/2}$ for all $n$.

This protocol is coherent both in the initial situation, before Forecaster announces $D_0$, and in the situation just after Forecaster announces $D_0$, where it reduces to the version in which $\sigma^2$ is a parameter. The central limit theorem for this protocol is a statement about prices in the situation just after Forecaster announces $D_0$: it says that $\int U(z)\, \mathcal{N}_{0,D_0}(dz)$ is then an approximate price for $U(S_N)$. Lindeberg's theorem does not include this particular central limit theorem, and our proof of Lindeberg's theorem cannot be adapted to cover it, because it was essential to that proof that Skeptic knew $D_{n+1}$ when choosing his move on round $n + 1$. We used this when we moved the center of Taylor's expansion of $\overline{U}$ forward from $(S_n, D_n)$ (its position in Equation (6.21), p. 132, in our proof of De Moivre's theorem) to $(S_n, D_{n+1})$ (Equation (7.16), p. 156). In our new protocol, where Skeptic does not know $D_{n+1}$ when making his moves $M_{n+1}$ and $V_{n+1}$, we must return to the expansion around $(S_n, D_n)$ that we used for De Moivre's theorem.

Before proving the central limit theorem for Bachelier's protocol, let us restate it to emphasize its financial interpretation, where $S_n$ and $D_n$ are the prices of the securities $\mathcal{S}$ and $\mathcal{D}$, respectively, and Skeptic and World are renamed Investor and Market, respectively. Market determines the price of both securities. Instead of assuming that $S_0 = 0$, we now permit Market the same freedom in setting $S_0$ as he has in setting $D_0$.

BACHELIER'S PROTOCOL WITH MARKET TERMINOLOGY
**Parameters:** $N, \mathcal{I}_0 > 0, C \geq 1$
**Players:** Market, Investor
**Protocol:**
Market announces $S_0 \in \mathbb{R}$ and $D_0 > 0$.
FOR $n = 1, \ldots, N$:
Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.

Market announces $S_n \in \mathbb{R}$ and $D_n \geq 0$.
$\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n((\Delta S_n)^2 + \Delta D_n)$. (10.1)
**Additional Constraints on Market:** Market must make $|\Delta D_n| \leq CN^{-1}$ and $|\Delta S_n| \leq CN^{-1/2}$ for $n = 1, \ldots, N$ and must ensure that $D_N = 0$ and $D_n > 0$ for $n < N$.

Now $M_n$ and $V_n$ are the number of shares of $\mathcal{S}$ and $\mathcal{D}$, respectively, that Investor holds during the $n$th period. According to (10.1), his net gain is

$$\Delta \mathcal{I}_n = M_n \Delta S_n + V_n \Delta D_n + V_n (\Delta S_n)^2. \tag{10.2}$$

The first two terms of this expression are his capital gains from holding the $M_n$ shares of $\mathcal{S}$ and the $V_n$ shares of $\mathcal{D}$. The third term amounts to a dividend of $(\Delta S_n)^2$ for each share of $\mathcal{D}$ that he holds. After the end of the $N$ periods, the security $\mathcal{D}$ will pay no more dividends and is worthless: $D_N = 0$. The other constraints limit how wildly the prices of the two securities can fluctuate in a single period.

The price $D_n$ is a signal from Market to Investor concerning how much the price of $\mathcal{S}$ will fluctuate during the remaining rounds of play. There is no requirement, however, that $D_n$ should decrease monotonically from its initial value of $D_0$ to its final value of 0. Market may change his mind about the likely remaining fluctuation, either because of his own and Investor's moves so far or because of information coming from outside the game.

Now we state and prove Bachelier's central limit theorem .

**Proposition 10.1** *Suppose $U$ is a bounded continuous function. Then in Bachelier's protocol with market terminology, the upper and lower prices of $U(S_N)$ in the situation where $S_0$ and $D_0$ have just been announced are both arbitrarily close to $\int U(z)\, \mathcal{N}_{S_0, D_0}(dz)$ for $N$ sufficiently large.*

*Proof* By the symmetry of the protocol, it suffices to prove that the difference between the upper price of $U(S_N)$ and $\int U(z)\, \mathcal{N}_{0,D_0}(dz)$ goes to zero as $N$ increases without bound.

First assume that $U$ is a smooth function constant outside a finite interval, so that its third and fourth derivatives are bounded: for some constant $c$ and all $s \in \mathbb{R}$, $|U^{(3)}(s)| \leq c$ and $|U^{(4)}(s)| \leq c$. As usual, we set $\overline{U}(s, D) := \int U(z)\, \mathcal{N}_{s,D}(dz)$ for $s \in \mathbb{R}$ and $D \geq 0$.

As in our proof of De Moivre's theorem in Chapter 6, we construct a strategy that makes Investor's capital at the end of round $n$ approximately equal to $\overline{U}(S_n, D_n)$ when he starts at $\overline{U}(S_0, D_0)$. From (6.21) (p. 132) we see that Investor should set

$$M_{n+1} = \frac{\partial \overline{U}}{\partial s}(S_n, D_n) \qquad \text{and} \qquad V_{n+1} = \frac{\partial \overline{U}}{\partial D}(S_n, D_n). \tag{10.3}$$

In this case, the first two terms of (6.21) give the increment of Investor's capital (cf. (10.2)), and we can rewrite (6.21) as

$$\begin{aligned} d\overline{U}(S_n, D_n) &= d\mathcal{I}_n + \frac{1}{2} \frac{\partial^3 \overline{U}}{\partial s^3}(S_n'', D_n'') dS_n' (dS_n)^2 + \frac{1}{2} \frac{\partial^3 \overline{U}}{\partial D \partial s^2}(S_n'', D_n'') dD_n' (dS_n)^2 \\ &+ \frac{\partial^2 \overline{U}}{\partial D \partial s}(S_n', D_n') dD_n dS_n + \frac{1}{2} \frac{\partial^2 \overline{U}}{\partial D^2}(S_n', D_n') (dD_n)^2. \end{aligned} \tag{10.4}$$

From (6.22), the fact that $\partial^3\overline{U}/\partial s^3$ and $\partial^4\overline{U}/\partial s^4$ as averages of $U^{(3)}$ and $U^{(4)}$ cannot exceed $c$ in absolute value, (10.4), and the constraints on Market's moves, we obtain:

$$\begin{aligned}|d\overline{U}(S_n, D_n) - d\mathcal{I}_n| &\le c\left(|dS_n|^3 + |dD_n||dS_n|^2 + |dD_n||dS_n| + |dD_n|^2\right)\\ &\le c\left(C^{1.5}N^{-1.5} + C^3N^{-2} + C^2N^{-1.5} + C^2N^{-2}\right) = O\left(N^{-3/2}\right).\end{aligned}$$

Summing over $n = 0, \ldots, N-1$, we obtain

$$\left|\left(\overline{U}(S_N, D_N) - \overline{U}(S_0, D_0)\right) - (\mathcal{I}_N - \mathcal{I}_0)\right| = O\left(N^{-1/2}\right).$$

This completes the proof under the assumption that $U$ is a smooth function constant outside a finite interval. To drop that assumption it suffices to apply Lemma 7.1 (p. 158). ▮

Proposition 10.1 is not really useful, because the constraints $|\Delta D_n| \le CN^{-1}$ and $|\Delta S_n| \le CN^{-1/2}$ are unrealistically strong, and because the proposition does not provide an explicit bound on the accuracy of the approximation. While still remaining within the unrealistic Bachelier setting, we now take a step towards more useful bounds.

## 10.2 BACHELIER PRICING IN DISCRETE TIME

In this section we derive an explicit bound on the accuracy of Bachelier's central limit theorem, using a version of Bachelier's protocol that constrains the variation spectra of the prices of $\mathcal{S}$ and $\mathcal{D}$. This derivation will serve as a model for the more applicable but slightly more complicated derivation, in the next section, of a discrete-time error bound for Black-Scholes.

The Taylor's expansion that we studied in §9.3 should have a negligible remainder, as required by the argument there, if the variation exponents of the realized paths for $S(t)$ and $D(t)$ satisfy

$$\operatorname{vex} S \le 2 \qquad \text{and} \qquad \operatorname{vex} D < 2. \tag{10.5}$$

As we learned in §9.1, the practical (discrete-time) meaning of the variation exponent having a certain value $\alpha$ is that $\mathbf{var}(p)$ will be small for $p$ greater than $\alpha$. So the meaning of (10.5) is that for small $\epsilon > 0$, $\mathbf{var}_S(2+\epsilon)$ and $\mathbf{var}_D(2-\epsilon)$ will be small. So even for a small $\delta > 0$, there should exist an $\epsilon \in (0,1)$ such that

$$\mathbf{var}_S(2+\epsilon) < \delta \qquad \text{and} \qquad \mathbf{var}_D(2-\epsilon) < \delta. \tag{10.6}$$

This is asymptotically weaker than the condition that we used in the preceding section—the condition that

$$|\Delta D_n| \le CN^{-1} \qquad \text{and} \qquad |\Delta S_n| \le CN^{-1/2}. \tag{10.7}$$

Indeed, (10.7) implies that

$$\mathbf{var}_S(2.5) = \sum_{n=1}^{N} |\Delta S_n|^{2.5} \le NC^{2.5}N^{-1.25} = C^{2.5}N^{-0.25}$$

and

$$\mathbf{var}_D(1.5) = \sum_{n=1}^{N} |\Delta D_n|^{1.5} \leq NC^{1.5}N^{-1.5} = C^{1.5}N^{-0.5},$$

so that (10.6) holds with $\epsilon = 0.5$ when $N$ is sufficiently large.

The condition that an $\epsilon$ satisfying (10.6) exists can be written compactly as

$$\inf_{\epsilon \in (0,1)} \max\left(\mathbf{var}_S(2+\epsilon),\, \mathbf{var}_D(2-\epsilon)\right) < \delta. \tag{10.8}$$

This is the constraint we impose on Market in our new version of Bachelier's protocol:

BACHELIER'S PROTOCOL WITH CONSTRAINED VARIATION
**Parameters:** $N, \mathcal{I}_0, \delta \in (0,1)$
**Players:** Market, Investor
**Protocol:**
Market announces $S_0 \in \mathbb{R}$ and $D_0 > 0$.
FOR $n = 1, \ldots, N$:
Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Market announces $S_n \in \mathbb{R}$ and $D_n \geq 0$.
$\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n((\Delta S_n)^2 + \Delta D_n)$.
**Additional Constraints on Market:** Market must set $D_N = 0$, $D_n > 0$ for $n < N$, and must make $S_0, \ldots, S_N$ and $D_0, \ldots, D_N$ satisfy (10.8).

We now state and prove our central limit theorem for this protocol. Because we are now concerned with pricing options rather than with statistical applications of the central limit theorem, we change our regularity conditions on the function $U$ that we want to price. In the central limit theorems of Part I and also in Proposition 10.1, we assumed that $U$ is bounded and continuous, a condition that is not even satisfied by European calls and puts. Now, and in subsequent theorems concerning option pricing, we assume instead that $U$ is Lipschitzian. Recall that a function $U\colon \mathbb{R} \to \mathbb{R}$ is *Lipschitzian* with coefficient $c$ if $|U(x) - U(y)| \leq c|x - y|$ for all $x$ and $y$ in $\mathbb{R}$.

**Proposition 10.2** *Suppose $U\colon \mathbb{R} \to \mathbb{R}$ is Lipschitzian with coefficient $c$. Then in Bachelier's protocol with constrained variation, the variable $U(S_N)$ is priced by*

$$\int U(z)\, \mathcal{N}_{S_0,D_0}(dz) \tag{10.9}$$

*with accuracy*

$$6c\delta^{1/4} \tag{10.10}$$

*in the situation where $S_0$ and $D_0$ have just been determined.*

For standard call and put options, we have $c = 1$, and (10.10) becomes $6\delta^{1/4}$.

*Proof of Proposition 10.2* We will establish the proposition by showing that (10.9) prices $U(S_N)$ with accuracy (10.10) provided (10.6) is satisfied, using a strategy for Investor that does not depend on $\epsilon$. (Equation (10.13), which describes the strategy, does not depend on $\delta$

either, but later we will replace $U$ with a smoothing, and the degree of smoothing will depend on $\delta$; see (10.32).)

First we state two well-known inequalities, involving nonnegative sequences $X_n$ and $Y_n$ and nonnegative real numbers $p$ and $q$, that we will use frequently. The first, Hölder's inequality, says that if $1/p + 1/q = 1$, then

$$\sum_{n=1}^{N} X_n Y_n \leq \left( \sum_{n=1}^{N} X_n^p \right)^{1/p} \left( \sum_{n=1}^{N} Y_n^q \right)^{1/q} . \tag{10.11}$$

The other, sometimes called Jensen's inequality, says that if $p \leq q$, then

$$\left( \sum_{n=1}^{N} X_n^p \right)^{1/p} \geq \left( \sum_{n=1}^{N} X_n^q \right)^{1/q} . \tag{10.12}$$

For proofs, see, for example, [11, 146]. These two inequalities imply that if $1/p + 1/q \geq 1$, then (10.11) holds. In particular,

$$\sum_{n=1}^{N} X_n Y_n \leq \left( \sum_{n=1}^{N} X_n^{2+\epsilon} \right)^{1/(2+\epsilon)} \left( \sum_{n=1}^{N} Y_n^{2-\epsilon} \right)^{1/(2-\epsilon)}$$

First assume that the derivatives $\|U^{(3)}\|$ and $\|U^{(4)}\|$ exist and their norms

$$\|U^{(3)}\| := \sup_s |U^{(3)}(s)|, \quad \|U^{(4)}\| := \sup_s |U^{(4)}(s)|$$

are bounded by positive constants $c_3$ and $c_4$, respectively. (We encountered a similar condition in §6.2: see the argument after (6.22), p. 132.) Taking as Investor's strategy

$$\Bigl( \underbrace{\frac{\partial \overline{U}}{\partial s}(S_n, D_n)}_{\text{shares of } \mathcal{S}}, \underbrace{\frac{\partial \overline{U}}{\partial D}(S_n, D_n)}_{\text{shares of } \mathcal{D}} \Bigr), \tag{10.13}$$

as in (10.3), and summing over $n = 0, \ldots, N-1$ in (6.21) (p. 132), we can see that

$$\begin{aligned} &\left| \left( \overline{U}(S_N, D_N) - \overline{U}(S_0, D_0) \right) - (\mathcal{I}_N - \mathcal{I}_0) \right| \\ &\leq \frac{1}{2} \sup \left| \frac{\partial^3 \overline{U}}{\partial s^3} \right| \mathbf{var}_S(3) + \frac{1}{2} \sup \left| \frac{\partial^3 \overline{U}}{\partial D \partial s^2} \right| \sum_n |dD_n||dS_n|^2 \\ &\quad + \sup \left| \frac{\partial^2 \overline{U}}{\partial D \partial s} \right| \sum_n |dD_n||dS_n| + \frac{1}{2} \sup \left| \frac{\partial^2 \overline{U}}{\partial D^2} \right| \mathbf{var}_D(2), \end{aligned} \tag{10.14}$$

all suprema being over the convex hull of $\{(S_n, D_n) \mid n = 0, \ldots, N\}$. Remember that $\overline{U}$ is defined by (6.10), p. 128.

Since, for $n = 3, 4$,

$$\left| \frac{\partial^n \overline{U}}{\partial s^n} \right| = \left| \int_{\mathbb{R}} U^{(n)}(s+z)\, \mathcal{N}_{0,D}(dz) \right| \leq \|U^{(n)}\| \int_{\mathbb{R}} \mathcal{N}_{0,D}(dz) = \|U^{(n)}\| \tag{10.15}$$

(we may differentiate under the integral sign by Leibniz's differentiation rule for integrals), we can bound from above all coefficients of (10.14) using (6.22) on p. 132:

$$\left|\frac{\partial^3 \overline{U}}{\partial s^3}\right| \le c_3, \quad \left|\frac{\partial^3 \overline{U}}{\partial D \partial s^2}\right| = \frac{1}{2}\left|\frac{\partial^4 \overline{U}}{\partial s^4}\right| \le \frac{1}{2}c_4,$$
$$\left|\frac{\partial^2 \overline{U}}{\partial D \partial s}\right| = \frac{1}{2}\left|\frac{\partial^3 \overline{U}}{\partial s^3}\right| \le \frac{1}{2}c_3, \quad \left|\frac{\partial^2 \overline{U}}{\partial D^2}\right| = \frac{1}{4}\left|\frac{\partial^4 \overline{U}}{\partial s^4}\right| \le \frac{1}{4}c_4.$$

Therefore, (10.14) gives the following accuracy for (10.9) pricing the final payoff $U(S_N)$:

$$\frac{c_3}{2}\,\mathbf{var}_S(3) + \frac{c_4}{4}\sum_n |dD_n||dS_n|^2 + \frac{c_3}{2}\sum_n |dD_n||dS_n| + \frac{c_4}{8}\,\mathbf{var}_D(2). \tag{10.16}$$

Using Hölder's inequality and the variant

$$a \le b \Longrightarrow \sum_n z_n^b \le \left(\sum_n z_n^a\right)^{b/a}$$

($a$, $b$, and all $z_n$ are positive numbers) of (10.12), we obtain, with $p := 2 + \epsilon$ and $q := 2 - \epsilon$,

$$\sum_n |dS_n|^3 \le \left(\sum_n |dS_n|^p\right)^{3/p} = (\mathbf{var}_S(p))^{3/p} \le \delta^{3/p} \le \delta, \tag{10.17}$$

$$\begin{aligned}\sum_n |dD_n||dS_n|^2 &\le \left(\sum_n |dS_n|^p\right)^{2/p}\left(\sum_n |dD_n|^{\frac{p}{p-2}}\right)^{\frac{p-2}{p}}\\ &\le \left(\sum_n |dS_n|^p\right)^{2/p}\left(\sum_n |dD_n|^q\right)^{\frac{p}{(p-2)q}\frac{p-2}{p}}\\ &= (\mathbf{var}_S(p))^{2/p}(\mathbf{var}_D(q))^{1/q} \le \delta^{2/p+1/q} \le \delta^{1+1/p} \le \delta\end{aligned} \tag{10.18}$$

(here we used the inequality $p/(p-2) \ge q$, which is easily checked),

$$\begin{aligned}\sum_n |dD_n||dS_n| &\le \left(\sum_n |dS_n|^p\right)^{1/p}\left(\sum_n |dD_n|^q\right)^{1/q}\\ &= (\mathbf{var}_S(p))^{1/p}(\mathbf{var}_D(q))^{1/q} \le \delta^{1/p}\delta^{1/q} \le \delta,\end{aligned} \tag{10.19}$$

$$\sum_n |dD_n|^2 \le \left(\sum_n |dD_n|^q\right)^{2/q} = (\mathbf{var}_D(q))^{2/q} \le \delta^{2/q} \le \delta, \tag{10.20}$$

and so the accuracy (10.16) can be bounded from above by

$$c_3\delta + \frac{3}{8}c_4\delta. \tag{10.21}$$

Now we drop the assumption that $\|U^{(3)}\|$ and $\|U^{(4)}\|$ are bounded, assuming only, as stated in the proposition, that $U$ is Lipschitzian with coefficient $c$. As usual, we introduce a new function $V$, a smoothed version of $U$, that does have bounded $\|V^{(3)}\|$ and $\|V^{(4)}\|$. But because we have replaced our usual condition on $U$, that it is bounded and continuous, by the condition that it is Lipschitzian, we cannot now rely on the method in Lemma 7.1 (p. 158) to construct $V$.

Let $\sigma > 0$ be some (small) parameter; put

$$V(s) := \int_{\mathbb{R}} U(s+z)\, \mathcal{N}_{0,\sigma^2}(dz); \tag{10.22}$$

this is the smoothed version of $U$ we are going to use. We know how to attain $V(S_N)$ starting with $\int V d\mathcal{N}_{S_0,D_0}$ with accuracy (10.21), where $c_3 = \|V^{(3)}\|$ and $c_4 = \|V^{(4)}\|$ (we will see momentarily that these derivatives exist and are bounded). So our plan is:

- Show that $U(S_N)$ is close to $V(S_N)$.
- Show that $\int U d\mathcal{N}_{S_0,D_0}$ is close to $\int V d\mathcal{N}_{S_0,D_0}$.
- Bound $c_3$ and $c_4$ from above.

From time to time we will make use of the formula

$$\int_0^\infty y^{2n+1} e^{-y^2/2} dy = 2^n \int_0^\infty x^n e^{-x} dx = 2^n \Gamma(n+1) = 2^n n!. \tag{10.23}$$

First let us check that $U(s)$ is close to $V(s)$, using the fact that $U$ is Lipschitzian with coefficient $c$:

$$\begin{aligned} |V(s) - U(s)| &= \left| \int_{\mathbb{R}} U(s+z) - U(s)\, \mathcal{N}_{0,\sigma^2}(dz) \right| \\ \le \int_{\mathbb{R}} |U(s+z) - U(s)|\, &\mathcal{N}_{0,\sigma^2}(dz) \le c \int_{\mathbb{R}} |z|\, \mathcal{N}_{0,\sigma^2}(dz) \\ &= c\sigma \int_{\mathbb{R}} |z|\, \mathcal{N}_{0,1}(dz) = \sqrt{2/\pi}\, c\sigma \end{aligned} \tag{10.24}$$

(the last equality follows from (10.23)). We can also find

$$\begin{aligned} &\left| \int_{\mathbb{R}} V(S_0+z) - U(S_0+z)\, \mathcal{N}_{0,D_0}(dz) \right| \\ &\le \int_{\mathbb{R}} |V(S_0+z) - U(S_0+z)|\, \mathcal{N}_{0,D_0}(dz) \le \sqrt{2/\pi}\, c\sigma \end{aligned} \tag{10.25}$$

(the last inequality follows from (10.24)).

Now we find upper bounds for all derivatives $V^{(n)}$. We start by proving that, for $n = 0, 1, \ldots$,

$$V^{(n)}(s) = \frac{1}{\sqrt{2\pi}\sigma^{n+1}} \int_{\mathbb{R}} e^{-(x-s)^2/(2\sigma^2)} H_n\left(\frac{x-s}{\sigma}\right) U(x) dx, \tag{10.26}$$

where $H_n$ are Hermite's polynomials (see, e.g., [287], Example II.11.1). Equality (10.26) can be proven by induction in $n$. Indeed, for $n = 0$ it is obvious (recall that $H_0(x) = 1$), and assuming (10.26) we can obtain, by differentiating (10.26), the analogous equality with $n+1$ in place of $n$:

$$\begin{aligned} &V^{(n+1)}(s) \\ &= \frac{1}{\sqrt{2\pi}\sigma^{n+1}} \int_{\mathbb{R}} e^{-(x-s)^2/(2\sigma^2)} \left( \frac{x-s}{\sigma^2} H_n\left(\frac{x-s}{\sigma}\right) - \frac{1}{\sigma} H_n'\left(\frac{x-s}{\sigma}\right) \right) U(x) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma^{n+2}} \int_{\mathbb{R}} e^{-(x-s)^2/(2\sigma^2)} H_{n+1}\left(\frac{x-s}{\sigma}\right) U(x) dx \end{aligned}$$

(we used the easily checked recurrence for the Hermite polynomials, $H_{n+1}(x) = xH_n(x) - H_n'(x)$). This completes the proof of (10.26).

Assuming, without loss of generality, $U(s) = 0$, we obtain for $n = 1, 2, \ldots$:

$$\begin{aligned}|V^{(n)}(s)| &\le \frac{1}{\sqrt{2\pi}\sigma^{n+1}} \int_{\mathbb{R}} e^{-(x-s)^2/(2\sigma^2)} \left| H_n\left(\frac{x-s}{\sigma}\right)\right| c|x-s|\,dx \\ &= \frac{c}{\sqrt{2\pi}\sigma^{n-1}} \int_{\mathbb{R}} e^{-y^2/2} |H_n(y)||y|\,dy.\end{aligned} \tag{10.27}$$

So we have the following upper bounds for $\|V^{(3)}\|$ and $\|V^{(4)}\|$:

$$\|V^{(3)}\| \le \frac{c}{\sqrt{2\pi}\sigma^2} \int_{\mathbb{R}} e^{-y^2/2} |y^3 - 3y||y|\,dy \le \frac{c}{\sqrt{2\pi}\sigma^2} \int_{\mathbb{R}} e^{-y^2/2} (y^4 + 3y^2)\,dy = \frac{6c}{\sigma^2} \tag{10.28}$$

(the second and fourth moments of the Gaussian distribution are 1 and 3) and

$$\begin{aligned}\|V^{(4)}\| &\le \frac{c}{\sqrt{2\pi}\sigma^3} \int_{\mathbb{R}} e^{-y^2/2} |y^4 - 6y^2 + 3||y|\,dy \\ &\le \frac{2c}{\sqrt{2\pi}\sigma^3} \int_0^\infty e^{-y^2/2} (y^5 + 6y^3 + 3y)\,dy = \frac{2c}{\sqrt{2\pi}\sigma^3}(8 + 12 + 3) = \frac{23\sqrt{2}c}{\sqrt{\pi}\sigma^3};\end{aligned} \tag{10.29}$$

in (10.29) we used formula (10.23).

Plugging (10.28) and (10.29) into (10.21) and taking into account (10.24) and (10.25), we obtain the accuracy

$$2\sqrt{2/\pi}c\sigma + \frac{6c}{\sigma^2}\delta + \frac{3}{8}\frac{23\sqrt{2}c}{\sqrt{\pi}\sigma^3}\delta \tag{10.30}$$

of reproducing the payoff $U(S_N)$ when starting with (10.9).

Finally, we need to choose some $\sigma$ in expression (10.30). This expression is of the form

$$A\sigma + B\sigma^{-3} + C\sigma^{-2}; \tag{10.31}$$

we will choose $\sigma$ so as to minimize the sum of the first two terms. This gives

$$\sigma = (3B/A)^{1/4}, \tag{10.32}$$

and so we can bound (10.31) from above by

$$\begin{aligned}&A(3B/A)^{1/4} + B(3B/A)^{-3/4} + C(3B/A)^{-1/2} \\ &= (3^{1/4} + 3^{-3/4})A^{3/4}B^{1/4} + 3^{-1/2}A^{1/2}B^{-1/2}C.\end{aligned}$$

Therefore, we obtain the following upper bound on (10.30):

$$\begin{aligned}&(3^{1/4} + 3^{-3/4})\left(2\sqrt{2/\pi}c\right)^{3/4}\left(\frac{3}{8}\frac{23\sqrt{2}c}{\sqrt{\pi}}\delta\right)^{1/4} \\ &\quad + 3^{-1/2}\left(2\sqrt{2/\pi}c\right)^{1/2}\left(\frac{3}{8}\frac{23\sqrt{2}c}{\sqrt{\pi}}\delta\right)^{-1/2} 6c\delta \\ &= (3^{1/2} + 3^{-1/2})2^{1/2}\pi^{-1/2}23^{1/4}c\delta^{1/4} + 8 \times 23^{-1/2}c\delta^{1/2} \\ &\le \left((3^{1/2} + 3^{-1/2})2^{1/2}\pi^{-1/2}23^{1/4} + 8 \times 23^{-1/2}\right)c\delta^{1/4} \le 5.71c\delta^{1/4},\end{aligned}$$

which is slightly better than promised in (10.10). ∎

## 10.3 BLACK-SCHOLES PRICING IN DISCRETE TIME

We now derive the Black-Scholes formula in a discrete-time protocol similar to the one we just used for Bachelier's formula. This protocol differs from the discrete-time Bachelier protocol in two essential particulars: (1) Market is required to keep the price of the underlying security $\mathcal{S}$ positive, and (2) the derivative security $\mathcal{D}$ now pays the relative change in $S(t)$, rather than the absolute change. We also bound the fluctuations in the prices of $\mathcal{S}$ and $\mathcal{D}$.

BLACK-SCHOLES PROTOCOL WITH CONSTRAINED VARIATION
**Parameters:** $N, \mathcal{I}_0 > 0, \delta \in (0,1), C > 0$
**Players:** Market, Investor
**Protocol:**
Market announces $S_0 > 0$ and $D_0 > 0$.
FOR $n = 1, \dots, N$:
Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Market announces $S_n > 0$ and $D_n \geq 0$.
$\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n((\Delta S_n / S_{n-1})^2 + \Delta D_n)$.
**Additional Constraints on Market:** Market's moves must satisfy $0 < S_n < C$ for $n = 1, \dots, N, 0 < D_n < C$ for $n = 1, \dots, N-1$, $D_N = 0$, and

$$\inf_{\epsilon \in (0,1)} \max (\mathbf{var}_S(2+\epsilon),\ \mathbf{var}_D(2-\epsilon)) < \delta. \tag{10.33}$$

This protocol is coherent: Market can prevent Investor from making any gains by announcing

$$D_n = \frac{N-n}{N} D_0$$

and

$$\frac{\Delta S_n}{S_{n-1}} = \pm \sqrt{\frac{D_0}{N}},$$

with + or − chosen as needed.

**Proposition 10.3** *Suppose $U \colon (0, \infty) \to \mathbb{R}$ is Lipschitzian with coefficient $c$. Then in the Black-Scholes protocol with constrained variation, the price of $U(S_N)$ is*

$$\int U(S_0 e^z)\ \mathcal{N}_{-D_0/2,\, D_0}(dz) \tag{10.34}$$

*with accuracy*

$$8ce^{5C}\delta^{1/4} \tag{10.35}$$

*in the situation where $S_0$ and $D_0$ have just been announced.*

For standard calls and puts, we have $c = 1$, and so (10.35) becomes $10e^{5C}\delta^{1/4}$. This is a relatively crude bound (see §10.4). We can eliminate the assumption $S_n < C$ from the protocol, but then we will obtain an even cruder bound.

*Proof of Proposition 10.3* This proof is modeled on the proof of Proposition 10.2. We assume that

$$\mathbf{var}_S(2+\epsilon) \le \delta, \quad \mathbf{var}_D(2-\epsilon) \le \delta, \tag{10.36}$$

and we prove that (10.34) prices $U(S_N)$ with accuracy (10.35).

First we assume that the second through fourth derivatives exist and their norms $\|U^{(2)}\|$, $\|U^{(3)}\|$, and $\|U^{(4)}\|$ are bounded by positive constants $c_2$, $c_3$, and $c_4$, respectively.

Put, for $D \ge 0$ and $s > 0$,

$$\overline{U}(s,D) := \int_{\mathbb{R}} U(se^z)\ \mathcal{N}_{-D/2,\,D}(dz). \tag{10.37}$$

It is clear that $\overline{U}$ is continuous and satisfies the initial condition $\overline{U}(s,0) = U(s)$. It can be checked by direct differentiation that for $D > 0$,

$$\frac{\partial \overline{U}}{\partial D} = \frac{1}{2}s^2\frac{\partial^2 \overline{U}}{\partial s^2}. \tag{10.38}$$

In fact, (10.37) is the only solution of (10.38) with this initial condition that satisfies the polynomial growth condition (see [107], Appendix E).

Our trading strategy is exactly the same as in the proof of Proposition 10.2, (10.13) (only $\overline{U}$ is defined differently). Since the heat equation (6.6) is replaced by (10.38), we now have

$$\begin{aligned} d\overline{U}(S_n,D_n) &= \frac{\partial \overline{U}}{\partial s}(S_n,D_n)dS_n + \frac{\partial \overline{U}}{\partial D}(S_n,D_n)\left(dD_n + \left(\frac{dS_n}{S_n}\right)^2\right) \\ &+ \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial s^3}(S_n'',D_n'')dS_n'(dS_n)^2 + \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial D\partial s^2}(S_n'',D_n'')dD_n'(dS_n)^2 \\ &+ \frac{\partial^2 \overline{U}}{\partial D\partial s}(S_n',D_n')dD_n dS_n + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial D^2}(S_n',D_n')(dD_n)^2 \end{aligned} \tag{10.39}$$

instead of (6.21) (p. 132). Summing over $n$, we obtain

$$\begin{aligned} &\left|(\overline{U}(S_N,D_N) - \overline{U}(S_0,D_0)) - (\mathcal{I}_N - \mathcal{I}_0)\right| \\ &\le \frac{1}{2}\sup\left|\frac{\partial^3 \overline{U}}{\partial s^3}\right|\mathbf{var}_S(3) + \frac{1}{2}\sup\left|\frac{\partial^3 \overline{U}}{\partial D\partial s^2}\right|\sum_n |dD_n||dS_n|^2 \\ &+ \sup\left|\frac{\partial^2 \overline{U}}{\partial D\partial s}\right|\sum_n |dD_n||dS_n| + \frac{1}{2}\sup\left|\frac{\partial^2 \overline{U}}{\partial D^2}\right|\mathbf{var}_D(2), \end{aligned} \tag{10.40}$$

with all suprema over the convex hull of $\{(S_n,D_n) \mid 0 \le n \le N\}$. (This is the same as (10.14).)

Now we prepare to bound the suprema in (10.40) from above. From (10.38),

$$\begin{aligned} \frac{\partial^3 \overline{U}}{\partial D\partial s^2} &= \frac{\partial^3 \overline{U}}{\partial s^2\partial D} = \frac{1}{2}\frac{\partial^2}{\partial s^2}\left(s^2\frac{\partial^2 \overline{U}}{\partial s^2}\right) = \frac{\partial^2 \overline{U}}{\partial s^2} + 2s\frac{\partial^3 \overline{U}}{\partial s^3} + \frac{1}{2}s^2\frac{\partial^4 \overline{U}}{\partial s^4}, \\ \frac{\partial^2 \overline{U}}{\partial D\partial s} &= \frac{1}{2}\frac{\partial}{\partial s}\left(s^2\frac{\partial^2 \overline{U}}{\partial s^2}\right) = s\frac{\partial^2 \overline{U}}{\partial s^2} + \frac{1}{2}s^2\frac{\partial^3 \overline{U}}{\partial s^3}, \\ \frac{\partial^2 \overline{U}}{\partial D^2} &= \frac{1}{2}\frac{\partial}{\partial D}\left(s^2\frac{\partial^2 \overline{U}}{\partial s^2}\right) = \frac{1}{2}s^2\frac{\partial^3 \overline{U}}{\partial D\partial s^2}. \end{aligned} \tag{10.41}$$

To bound the partial derivatives $\partial^n \overline{U}/\partial s^n$, $n = 1, 2, \ldots$, from above in absolute value, we first note that $\mathbb{E}\, e^{x\xi} = e^{x^2/2}$ for $x \in \mathbb{R}$ and a standard Gaussian variable $\xi$ (this is a restatement of the well-known fact that $e^{-x^2/2}$ is the characteristic function of $\xi$). Therefore, applying Leibniz's differentiation rule, we obtain

$$\begin{aligned}\left|\frac{\partial^n \overline{U}}{\partial s^n}\right| &= \left|\int_{\mathbb{R}} U^{(n)}(se^z)e^{nz}\ \mathcal{N}_{-D/2,\,D}(dz)\right| \le \|U^{(n)}\| \int_{\mathbb{R}} e^{nz}\ \mathcal{N}_{-D/2,\,D}(dz) \\ &= \|U^{(n)}\|\,\mathbb{E}\,e^{n(-D/2+\sqrt{D}\xi)} = \|U^{(n)}\|e^{-nD/2}e^{n^2D/2} = \|U^{(n)}\|e^{n(n-1)D/2}.\end{aligned} \tag{10.42}$$

Now we are ready to bound the suprema in (10.40). From (10.41) and (10.42) we obtain

$$\left|\frac{\partial^3 \overline{U}}{\partial s^3}\right| \le c_3 e^{3C}, \tag{10.43}$$

$$\left|\frac{\partial^3 \overline{U}}{\partial D \partial s^2}\right| \le \left|\frac{\partial^2 \overline{U}}{\partial s^2}\right| + 2C\left|\frac{\partial^3 \overline{U}}{\partial s^3}\right| + \frac{1}{2}C^2\left|\frac{\partial^4 \overline{U}}{\partial s^4}\right| \le c_2e^C + 2Cc_3e^{3C} + \frac{1}{2}C^2c_4e^{6C}, \tag{10.44}$$

$$\left|\frac{\partial^2 \overline{U}}{\partial D \partial s}\right| \le C\left|\frac{\partial^2 \overline{U}}{\partial s^2}\right| + \frac{1}{2}C^2\left|\frac{\partial^3 \overline{U}}{\partial s^3}\right| \le Cc_2e^C + \frac{1}{2}C^2c_3e^{3C}, \tag{10.45}$$

$$\left|\frac{\partial^2 \overline{U}}{\partial D^2}\right| \le \frac{1}{2}C^2\left|\frac{\partial^3 \overline{U}}{\partial D \partial s^2}\right| \le \frac{1}{2}C^2c_2e^C + C^3c_3e^{3C} + \frac{1}{4}C^4c_4e^{6C}. \tag{10.46}$$

Combining Equations (10.43)–(10.46) and (10.17)–(10.20) with Equation (10.40), we obtain the following bound for the accuracy of pricing $U$:

$$\delta\left(\left(\frac{C^2}{4} + C + \frac{1}{2}\right)e^C c_2 + \left(\frac{C^3}{2} + \frac{C^2}{2} + C + \frac{1}{2}\right)e^{3C}c_3 + \left(\frac{C^4}{8} + \frac{C^2}{4}\right)e^{6C}c_4\right). \tag{10.47}$$

Now we remove the restriction on $\|U^{(2)}\|$–$\|U^{(4)}\|$. As usual, we introduce a new function $V$ by Equation (10.22) and apply the bound (10.47) to $V$. Equation (10.24) still holds, but (10.25) has to be modified (remember that $c$ is the constant from the definition of $U$ being Lipschitzian):

$$\begin{aligned}|\overline{V}(S_0, D_0) - \overline{U}(S_0, D_0)| &= \left|\int_{\mathbb{R}} V(S_0e^z) - U(S_0e^z)\ \mathcal{N}_{-D_0/2,\,D_0}(dz)\right| \\ &\le \int_{\mathbb{R}} |V(S_0e^z) - U(S_0e^z)|\ \mathcal{N}_{-D_0/2,\,D_0}(dz) \le \sqrt{2/\pi}\,c\sigma\end{aligned} \tag{10.48}$$

(the last inequality follows from (10.24)); we can see that the inequality between the extreme terms in (10.25) still holds.

Using inequality (10.27), we obtain (analogously to (10.28) and (10.29)):

$$\begin{aligned}\|V^{(2)}\| &\le \frac{c}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} e^{-y^2/2}|y^2 - 1||y|\,dy \\ &\le \frac{2c}{\sqrt{2\pi}\sigma}\int_0^\infty e^{-y^2/2}(y^3 + y)\,dy = \frac{2c}{\sqrt{2\pi}\sigma}(2+1) = \frac{3\sqrt{2}c}{\sqrt{\pi}\sigma}.\end{aligned} \tag{10.49}$$

Combining (10.49), (10.28), (10.29), the inequalities (true for $C > 0$)

$$\frac{C^2}{4} + C + \frac{1}{2} \le 0.5e^{[5\frac{2}{3}]C}, \quad \frac{C^3}{2} + \frac{C^2}{2} + C + \frac{1}{2} \le 0.5e^{[3\frac{2}{3}]C},$$
$$\frac{C^4}{8} + \frac{C^2}{4} \le 3.14e^{\frac{2}{3}C}$$

(we let $[k\frac{m}{n}]$ stand for the mixed fraction $k + m/n$), and the inequalities (10.24) and (10.48) relating $V$ and $U$, we obtain from (10.47), interpreted as the accuracy for pricing the derivative $V(S_N)$, the accuracy

$$2\sqrt{2/\pi}\, c\sigma + \delta c e^{[6\frac{2}{3}]C}\left(0.5\frac{3\sqrt{2}}{\sqrt{\pi}\sigma} + 0.5\frac{6}{\sigma^2} + 3.14\frac{23\sqrt{2}}{\sqrt{\pi}\sigma^3}\right) \tag{10.50}$$

for pricing $U(S_N)$. Similarly to our treatment of (10.31), we notice that (10.50) is an expression of the form

$$A\sigma + B\sigma^{-3} + C\sigma^{-2} + D\sigma^{-1};$$

substituting the same value (10.32) for $\sigma$ as before, we obtain

$$(3^{1/4} + 3^{-3/4})A^{3/4}B^{1/4} + 3^{-1/2}A^{1/2}B^{-1/2}C + 3^{-1/4}A^{1/4}B^{-1/4}D,$$

which for (10.50) becomes

$$\begin{aligned}
&(3^{1/4} + 3^{-3/4})\left(2\sqrt{2/\pi}c\right)^{3/4}\left(\delta c e^{[6\frac{2}{3}]C} 3.14\frac{23\sqrt{2}}{\sqrt{\pi}}\right)^{1/4}\\
&+ 3^{-1/2}\left(2\sqrt{2/\pi}c\right)^{1/2}\left(\delta c e^{[6\frac{2}{3}]C} 3.14\frac{23\sqrt{2}}{\sqrt{\pi}}\right)^{-1/2}\delta c e^{[6\frac{2}{3}]C} 0.5 \times 6\\
&+ 3^{-1/4}\left(2\sqrt{2/\pi}c\right)^{1/4}\left(\delta c e^{[6\frac{2}{3}]C} 3.14\frac{23\sqrt{2}}{\sqrt{\pi}}\right)^{-1/4}\delta c e^{[6\frac{2}{3}]C} 0.5\frac{3\sqrt{2}}{\sqrt{\pi}}\\
\le\ & \delta^{1/4} c e^{5C}\Big((3^{1/4} + 3^{-3/4})2^{5/4}\pi^{-1/2}3.14^{1/4}23^{1/4} + 2^{1/2}3^{1/2}3.14^{-1/2}23^{-1/2}\\
&+ 2^{-1/4}3^{3/4}\pi^{-1/2}3.14^{-1/4}23^{-1/4}\Big) \le 7.53 c e^{5C}\delta^{1/4};
\end{aligned}$$

this proves the proposition. ∎

## 10.4 HEDGING ERROR IN DISCRETE TIME

In this section, we look at $p$-variations for the American stock price data that we studied in the preceding chapter (prices for Microsoft stock and values for the S&P 500 index), with a view to assessing whether the methods used in our proof of Proposition 10.3 can be used in practice for hedging options. The results are encouraging. In §10.5, we will obtain even better results using the slightly different approach of Proposition 10.4.

### The Variation Spectrum for $S$

We will use the first 512 days of the 600 days of data for Microsoft and the S&P 500 that we displayed in Figure 9.1 (p. 203). We work with 512 days because this is a power of two close to a typical maturity for long-term options such as the Long-Term Equity Anticipation Securities (LEAPS) introduced by the Chicago Board Options Exchange in 1990 and now traded at all four United States options exchanges ([163], p. 56).

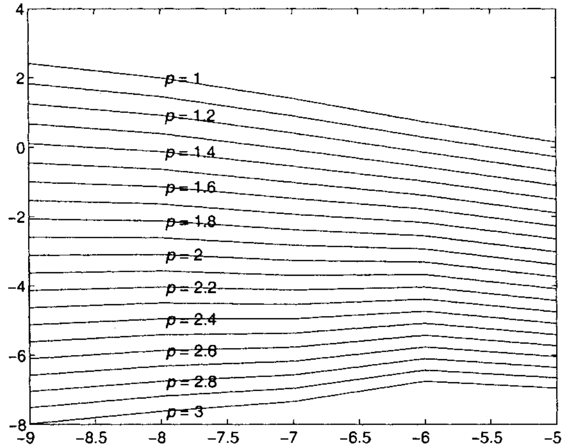



***Fig. 10.1*** Plots of log $\mathbf{var}(p)$ against log $dt$ for different values of $p$ for the Microsoft stock price for the 512 working days starting January 1, 1996. We use 512 days as the unit for measuring time, and we use the final (day 512) price as the unit for measuring price. The logarithms are base 2. The time step $dt$ runs from one day ($1/512$, or $-9$ on the logarithmic scale) to 16 days ($16/512$, or $-5$ on the logarithmic scale).

We use daily closing prices, as we did in §9.1. Using one day as our time step corresponds to the usual banking practice of updating hedging portfolios daily. Fortunately, the exact choice of the time step does not make too much difference. The approximation

$$\log \mathbf{var}(p) \approx (Hp - 1) \log dt \tag{10.51}$$

seems to hold within wide bounds, so multiplying $dt$ by a moderate factor multiplies each $p$-variation by a similar factor, and this does not change whether a particular $p$-variation is close to zero. The approximation (10.51) follows directly from the heuristic relation (9.8), assuming that (1) the time horizon $T$ is our unit for measuring time, and (2) the $dS_t$ have the order of magnitude $(dt)^H$. Figures 10.1–10.4 confirm that the approximation is qualitatively accurate for our Microsoft and S&P 500 data when $dt$ ranges from 1 to 16 days. Figures 10.1 and 10.2 show that $\log \mathbf{var}(p)$ is approximately linear in $\log dt$, while Figures 10.3 and 10.4 show that it is approximately linear in $p$. The important point is that whether we take $dt$ to be 1, 2, 4, 8, or 16 days does not make much difference in the smallness of the $p$-variation for $p$ greater than 1. This is indicated by the flatness of the lines in Figures 10.1 and 10.2 and by their approximate agreement for $p > 2$ in Figures 10.3 and 10.4.

Mandelbrot and his collaborators have reported similar results for a variety of price series. In [124], Fisher, Mandelbrot, and Calvet report that the variation spectrum for the USD/DM exchange rate is largely the same for $dt$ from about two hours to about one hundred days. (They study the variation spectrum for the logarithm of the exchange rate, but theoretically, at least, this should make no difference.) A diagram

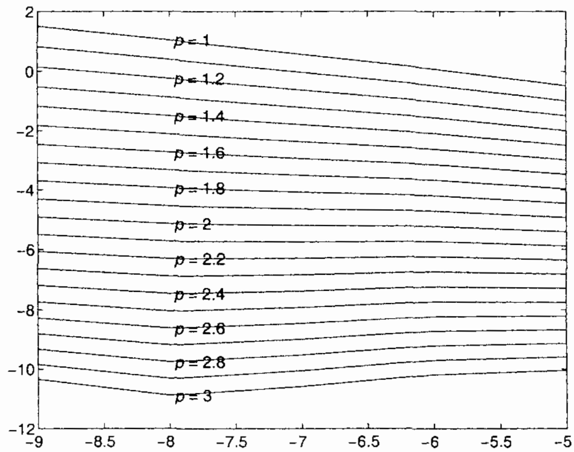



***Fig. 10.2*** Plots of log $\mathbf{var}(p)$ against log $dt$ for different values of $p$ for the S&P 500 index for the 512 working days starting January 1, 1996.

in [216], p. 176, shows that the dependence of log $\mathbf{var}(p)$ on log $dt$ is indeed close to being linear. This linear dependence breaks down for $dt$ below 2 hours (at about 1.4 hours according to [46]); this is attributed to market frictions such as bid-ask spreads and discontinuous trading [124]. The results depend on a filtering that expands time during typically active periods of the week, but there is no strong dependence on the choice of filter [124]. Calvet and Fisher report similar findings for five major US stocks and an equity index [46]. These results are reassuring, although they are not strictly applicable to our question, because they are based on analyses in which intercepts of straight lines (relevant for us but irrelevant to the multifractal model the authors are testing) have been discarded.

In order for the bound in Proposition 10.3 to be useful, it is necessary that $S$'s 3-variation, at least, should be small. So it is reassuring that the daily 3-variation (the value of the curve marked $p = 3$ at the point $-9$ on the horizontal axis in Figures 10.1 and 10.2) is quite small, approximately $2^{-8}$ in the case of Microsoft and $2^{-10}$ in the case of the S&P 500.

## The Variation Spectrum for $\mathcal{D}$

Because the security $\mathcal{D}$ is not now traded, we can only speculate about its variation spectrum. We obtain an interesting benchmark, however, by looking at what $\mathcal{D}$ would be worth if all the future prices of $S$ were known:

$$D_n^{\circ} = \sum_{i=n}^{511} \left( \frac{S_{i+1} - S_i}{S_i} \right)^2 ,$$

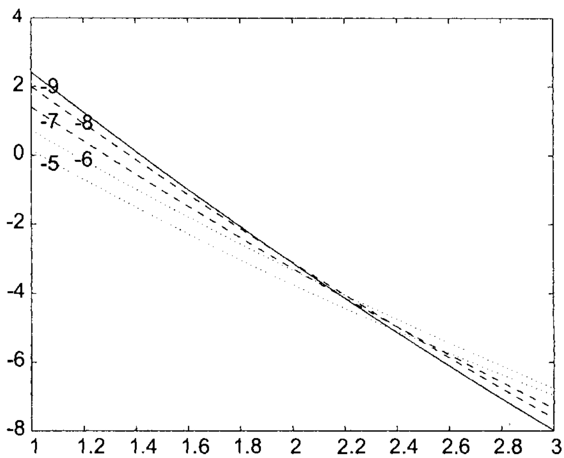



***Fig. 10.3*** Plots of log $\mathbf{var}(p)$ against $p$ for different time steps for the Microsoft data. The lines are labeled by the logarithm of the time step, $-9$ to $-5$.

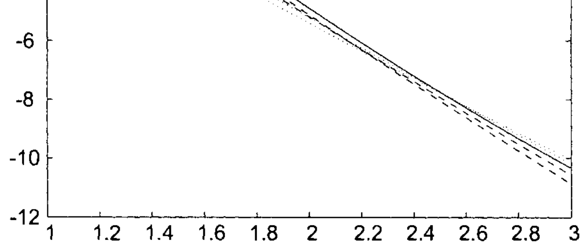



***Fig. 10.4*** Plots of log $\mathbf{var}(p)$ against $p$ for different time steps for the S&P 500 data.

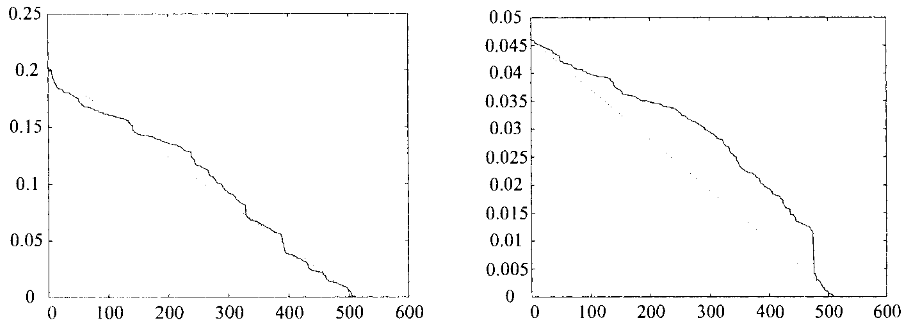



***Fig. 10.5*** The oracular prices $D_n^{\circ}$ for $n = 0, \ldots, 512$, for the Microsoft stock price on the left and the S&P 500 index on the right. The straight line in each plot shows, for reference, the linear decrease in the value of $\mathcal{D}$ predicted by the diffusion model.

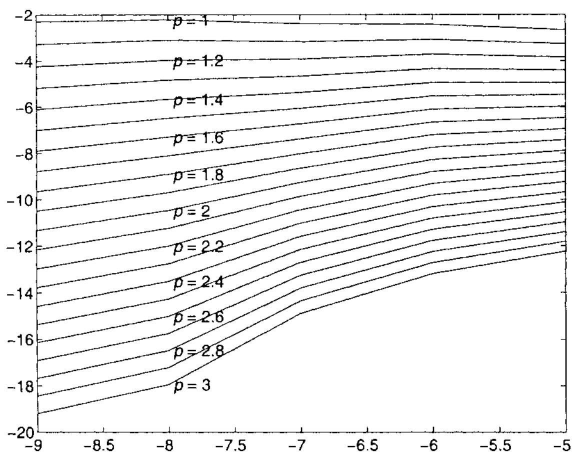



***Fig. 10.6*** The loglog plot of $\mathbf{var}(p)$ against the time step for the oracular prices $D_n^{\circ}$ for Microsoft.

for $n = 0, \ldots, 511$, and $D_{512}^{\circ} = 0$. We call $D_0^{\circ}, \ldots, D_{512}^{\circ}$ *oracular prices*; these are the prices that an oracle would set. Figure 10.5 shows the oracular prices for our two data sets, and Figures 10.6 and 10.7 show their variation spectra.

In order for the bound in Proposition 10.3 to be useful, it is necessary that $\mathcal{D}$'s 2-variation be small. Figures 10.6 and 10.7 show daily 2-variations of approximately $2^{-11}$ for Microsoft and $2^{-15}$ for the S&P 500. Market prices might show greater volatility than this (cf. §10.6), but the 2-variations could be several orders of magnitude greater and still be very small. The diffusion model would produce computed

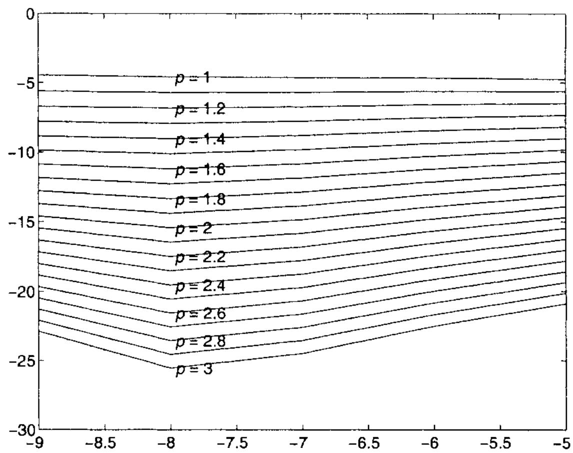



***Fig. 10.7*** The loglog plot of $\mathbf{var}(p)$ against the time step for oracular prices for the S&P 500 index.

values of approximately $2^{-13}$ for Microsoft and $2^{-18}$ for the S&P 500 when the time step is one day, but this is surely too optimistic.

## Is the Hedging Error in Proposition 10.3 Acceptable?

The estimates of the variation spectra for $S$ and $D$ that we have just developed show that Proposition 10.3 itself is too crude to give useful error bounds. But it comes close enough to doing so to give us reason to hope that its method can be useful when tailored to particular options.

Recall that the error bound for Proposition 10.3 reduces to $8e^{5C}\delta^{1/4}$ for a call or put. In the case of Microsoft, we see from the values of $\log \mathbf{var}(p)$ given in Figures 10.1 and 10.6 for $\log dt = -9$ (daily trading) that the smallest $\delta$ satisfying (10.33) is about $2^{-6}$, approximately the 2.6-variation for $S$ and the 1.4-variation for $D$ ($\epsilon = 0.6$). Optimistically setting $C = 1$, we find that a call or put can be hedged with accuracy

$$8e^{5C}\delta^{1/4} \approx 420.$$

This is of no practical use, because the payoff of the call or put will be less than one. (Our unit of measurement is the final price of Microsoft, and the payoff of a call or a put will usually be less than the stock's price.) But it demonstrates that we are not hopelessly stuck at infinity.

A re-examination of the proof of Proposition 10.3 shows several opportunities to tighten the estimate and obtain a meaningful bound. Most importantly, the right-hand side of the inequality (10.40) can be used directly, and it can give a very meaningful result. The first step where a significant loss of accuracy occurs in the proof is where *cross-variations*

$$\sum_n |dS_n|^a |dD_n|^b$$

(the second and third terms on the right-hand side of (10.40) are of this form) are bounded in terms of the variations for $S$ and $D$ (cf. (10.19) on p. 246). This loss of accuracy is not too serious; in the case of Microsoft, for example, if we take $\epsilon = 0.6$ (i.e., $p = 2+\epsilon = 2.6$ and $q = 2-\epsilon = 1.4$), then we obtain from (10.19) the accuracy

$$\begin{aligned}\sum_n |dS_n||dD_n| &\leq (\mathbf{var}_S(p))^{1/p} (\mathbf{var}_D(q))^{1/q} \\ &\approx (2^{-6})^{1/2.6}(2^{-6})^{1/1.4} \approx 0.01\end{aligned}$$

for the main cross-variation term. (We may alternatively take $p = 3$ and $q = 1.5$: the essential condition in the proof is $1/p + 1/q \geq 1$, not $p = 2+\epsilon$ and $q = 2-\epsilon$. This gives 0.006 as the approximate upper bound.) The other cross-variation will be much smaller. The really big losses of accuracy occur when we estimate the coefficients in front of variations and cross-variations in (10.40) and when we smooth (10.22). By using suitable estimation procedures, these losses can be avoided if the coefficients (which are determined by partial derivatives of $\overline{U}$) are all moderate in magnitude; for example, if the coefficients are all bounded in absolute value by 1, then we arrive at an overall bound on the hedging error less than 0.04.

The prospects are therefore good for practical error bounds. A realistic assessment will require, however, both a better handle on the behavior of $\mathcal{D}$ (see §10.6) and consideration of transaction costs, which are steadily decreasing because of new technologies but never zero.

### The Accuracy of Stochastic Hedging

There is an extensive literature on discrete hedging under the stochastic Black-Scholes model, going back to work by Black and Scholes themselves in the early 1970s [29]. This literature is concerned, however, with error that would be eliminated entirely by hedging in our variance derivative $\mathcal{D}$. It does not deal with the $O((dt)^{3/2})$ error studied in this section—an error that is smaller than the error we eliminate but still, as we have seen, nonnegligible.

The seminal work on discrete stochastic hedging is by Boyle and Emanuel (1980), who studied expected error under geometric Brownian motion when Black-Scholes delta-hedging is implemented only discretely. Subsequent work moved in directions not central to our concerns. Schäl (1994) derived the optimal (minimum expected squared error) strategy for hedging in discrete time under geometric Brownian motion (this differs slightly from the Black-Scholes delta, which is error-free and therefore optimal only for continuous hedging). Schweizer (1995) generalized this idea to other stochastic processes, and Mercurio and Vorst (1996) considered the effect of transaction costs (see also Wilmott 1998).

Stochastic error analysis probably cannot approach the specificity and concreteness of our error analysis. The only nonasymptotic error bounds it can provide are probabilistic and are therefore questionable as soon as we acknowledge doubts about the diffusion model. The central problem in relying on the diffusion model is the implicit appeal to the law of large numbers to deal with the unhedged squared relative returns $(dS(t)/S(t))^2$. If this is invalid, there is a potentially catastrophic error, which we avoid by variance hedging. Because we deal with $(dS(t)/S(t))^2$ by hedging rather than with stochastic theory, our discrete theory is completely model-free. Our error bounds involve no probabilities at all.

## 10.5 BLACK-SCHOLES WITH RELATIVE VARIATIONS FOR $S$

Given a sequence $S_1, S_2, \ldots$ of positive numbers, we can define a function of $p$ based on their relative increments:

$$\mathbf{var}_S^{\mathrm{rel}}(p) := \sum_n \left| \frac{dS_n}{S_n} \right|^p, \tag{10.52}$$

where $dS_n := S_{n+1} - S_n$. We call this function of $p$ the *relative variation sprectrum*. It is dimensionless and invariant with respect to the unit in which $S_n$ is measured.

The dimensionlessness of the relative variation spectrum contrasts, of course, with the behavior of the absolute variation spectrum,

$$\mathbf{var}_S(p) = \sum_n |dS_n|^p .$$

As we saw in §9.1, to our repeated discomfort, $\mathbf{var}_S(p)$ is measured in the $p$th power of whatever monetary unit we use for $S$.

Proposition 10.3 (p. 249) can be adapted directly to the relative variation spectrum: replace the first inequality in (10.36) with

$$\mathbf{var}_S^{\mathrm{rel}}(2+\epsilon) := \sum_t \left| \frac{S(t+dt) - S(t)}{S(t)} \right|^{2+\epsilon} \leq \delta,$$

and replace (10.35) with

$$8ce^{5C}(\delta C^3)^{1/4}. \tag{10.53}$$

The extra factor $C^3$ in (10.53) arises from the inequality $\mathbf{var}_S(2+\epsilon) \leq \|S\|_\infty^{2+\epsilon}\, \mathbf{var}_S^{\mathrm{rel}}(2+\epsilon) \leq C^3\, \mathbf{var}_S^{\mathrm{rel}}(2+\epsilon)$.

It is more natural, however, to extract a new Black-Scholes formula from the proof of Proposition 10.3 rather than from its statement. To this end, consider the protocol obtained from the Black-Scholes protocol with constrained variation (p. 249) by replacing (10.33) with

$$\inf_{\epsilon \in (0,1)} \max\left(\mathbf{var}_S^{\mathrm{rel}}(2+\epsilon),\, \mathbf{var}_D(2-\epsilon)\right) < \delta. \tag{10.54}$$

We may call this the *relative Black-Scholes protocol with constrained variation.*

A function $U : (0, \infty) \to \mathbb{R}$ is *log-Lipschitzian* (with coefficient $c$) if the function $S \in \mathbb{R} \mapsto U(e^S)$ is Lipschitzian (with coefficient $c$).

**Proposition 10.4** *Suppose $U\colon (0,\infty) \to \mathbb{R}$ is log-Lipschitzian with coefficient $c$ and suppose $\delta \in (0,1)$. Then in the relative Black-Scholes protocol with constrained variation, the price of $U(S_N)$ is*

$$\int U(S_0 e^z)\ \mathcal{N}_{-D_0/2,\,D_0}(dz)$$

*with accuracy*

$$40c\delta^{1/4} \tag{10.55}$$

*in the situation where $S_0$ and $D_0$ have just been announced.*

Put options are log-Lipschitzian, whereas call options are not. So the condition imposed on $U$ in Proposition 10.4 is quite strong. A strong condition is needed because Investor must replicate the final payoff $U(S(T))$ with small absolute error, whereas we expect the oscillations $dS(t)$ to grow as $S(t)$ grows.

*Proof of Proposition 10.4* First we assume that the norms

$$c_m := \sup_{s\in(0,\infty)} |s^m U^{(m)}(s)| \tag{10.56}$$

are finite for $m = 2, 3, 4$.

Analogously to (10.40) (but using relative rather than absolute increments of $S$), we obtain

$$\begin{aligned}
&\left|\left(\overline{U}(S_N, D_N) - \overline{U}(S_0, D_0)\right) - (\mathcal{I}_N - \mathcal{I}_0)\right| \\
&\le \frac{1}{2}\sup\left|s^3\frac{\partial^3\overline{U}}{\partial s^3}\right| \operatorname{var}_S^{\mathrm{rel}}(3) + \frac{1}{2}\sup\left|s^2\frac{\partial^3\overline{U}}{\partial D\partial s^2}\right| \sum_n |dD_n| \left|\frac{dS_n}{S_n}\right|^2 \\
&\quad + \sup\left|s\frac{\partial^2\overline{U}}{\partial D\partial s}\right| \sum_n |dD_n| \left|\frac{dS_n}{S_n}\right| + \frac{1}{2}\sup\left|\frac{\partial^2\overline{U}}{\partial D^2}\right| \operatorname{var}_D(2),
\end{aligned} \tag{10.57}$$

with all suprema again over the convex hull of $\{(S_n, D_n) \mid 0 \le n \le N\}$. These suprema can be bounded from above using (10.41) and the following analog of (10.42):

$$\left|s^n\frac{\partial^n\overline{U}}{\partial s^n}\right| = \left|\int_{\mathbb{R}} (se^z)^n U^{(n)}(se^z)\ \mathcal{N}_{-D/2,\,D}(dz)\right| \le c_n.$$

In place of (10.43)–(10.46), we obtain

$$\left|s^3\frac{\partial^3\overline{U}}{\partial s^3}\right| \le c_3, \qquad \left|s^2\frac{\partial^3\overline{U}}{\partial D\partial s^2}\right| \le c_2 + 2c_3 + \frac{1}{2}c_4, \tag{10.58}$$

$$\left|s\frac{\partial^2\overline{U}}{\partial D\partial s}\right| \le c_2 + \frac{1}{2}c_3, \qquad \left|\frac{\partial^2\overline{U}}{\partial D^2}\right| \le \frac{1}{2}c_2 + c_3 + \frac{1}{4}c_4. \tag{10.59}$$

Combining (10.58)–(10.59), the relative versions of (10.17)–(10.20), and (10.57), we obtain a simpler bound for the accuracy of pricing $U$:

$$\delta\left(1.75c_2 + 2.5c_3 + 0.375c_4\right). \tag{10.60}$$

To remove the restriction on the existence and finiteness of $c_2$–$c_4$, we define the $\sigma$-smoothing $V$ of $U$ in three steps:

- represent $U$ in the log-picture: $f(S) := U(e^S)$, where $S \in \mathbb{R}$;
- smooth: for some $\sigma^2 > 0$, $g(S) := \int f(S+z)\,\mathcal{N}_{0,\sigma^2}(dz)$;
- leave the log picture: $V(s) := g(\ln s)$, $s$ ranging over $(0,\infty)$.

Remember that $f$ is $c$-Lipschitzian. Applying (10.28), (10.29), (10.49), and

$$\|V^{(1)}\| \le \frac{c}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2}|y|^2\,dy = c$$

to $g$ rather than $V$, we obtain for $V$'s norms (in the sense of (10.56)):

$$\left|s^2 V^{(2)}\right| = \left|s^2 \frac{d^2}{ds^2} g(\ln s)\right| = \left|g''(\ln s) - g'(\ln s)\right| \le \frac{3\sqrt{2}c}{\sqrt{\pi}\sigma} + c,$$

$$\left|s^3 V^{(3)}\right| = \left|s^3 \frac{d^3}{ds^3} g(\ln s)\right| = \left|g^{(3)}(\ln s) - 3g''(\ln s) + 2g'(\ln s)\right|$$
$$\le \frac{6c}{\sigma^2} + 3\frac{3\sqrt{2}c}{\sqrt{\pi}\sigma} + 2c,$$

$$\left|s^4 V^{(4)}\right| = \left|s^4 \frac{d^4}{ds^4} g(\ln s)\right| = \left|g^{(4)}(\ln s) - 6g^{(3)}(\ln s) + 11g''(\ln s) - 6g'(\ln s)\right|$$
$$\le \frac{23\sqrt{2}c}{\sqrt{\pi}\sigma^3} + 6\frac{6c}{\sigma^2} + 11\frac{3\sqrt{2}c}{\sqrt{\pi}\sigma} + 6c.$$

Applying (10.60) to $V$ and taking into account the accuracy (10.24), (10.48) of $V$ approximating $U$, we obtain the accuracy

$$2\sqrt{2/\pi}c\sigma + \delta\,(1.75c_2 + 2.5c_3 + 0.375c_4)$$
$$\le 2\sqrt{2/\pi}c\sigma + \delta\left(0.375\frac{23\sqrt{2}c}{\sqrt{\pi}\sigma^3} + 4.75\frac{6c}{\sigma^2} + 13.375\frac{3\sqrt{2}c}{\sqrt{\pi}\sigma} + 9c\right)$$

for capital of Investor (who is playing the strategy computed from $V$) approximating the target payoff $U$. As in (10.50), we obtain the accuracy

$$A\sigma + B\sigma^{-3} + C\sigma^{-2} + D\sigma^{-1} + E$$
$$= (3^{1/4} + 3^{-3/4})A^{3/4}B^{1/4} + 3^{-1/2}A^{1/2}B^{-1/2}C + 3^{-1/4}A^{1/4}B^{-1/4}D + E$$
$$= (3^{1/4} + 3^{-3/4})\left(2\sqrt{2/\pi}c\right)^{3/4}\left(0.375\frac{23\sqrt{2}}{\sqrt{\pi}}c\delta\right)^{1/4}$$
$$+3^{-1/2}\left(2\sqrt{2/\pi}c\right)^{1/2}\left(0.375\frac{23\sqrt{2}}{\sqrt{\pi}}c\delta\right)^{-1/2} 4.75 \times 6c\delta$$
$$+3^{-1/4}\left(2\sqrt{2/\pi}c\right)^{1/4}\left(0.375\frac{23\sqrt{2}}{\sqrt{\pi}}c\delta\right)^{-1/4} 13.375\frac{3\sqrt{2}}{\sqrt{\pi}}c\delta + 9c\delta$$
$$\le \delta^{1/4}c\Big((3^{1/4} + 3^{-3/4})2^{5/4}\pi^{-1/2}0.375^{1/4}23^{1/4}$$

$$+2^{3/2}3^{1/2}0.375^{-1/2}23^{-1/2}4.75+2^{3/4}3^{3/4}\pi^{-1/2}0.375^{-1/4}23^{-1/4}13.375+9\Big)$$

$$\leq 37.84c\delta^{1/4}.$$

This completes the proof. ∎

Although we have singled out the relative variation spectrum for further study in this book, there are other reasonable ways of defining a variation spectrum when $S(t)$ is required to be positive. For example, we can set

$$\sum_n \left|\frac{dS_n}{S_n^\alpha}\right|^p$$

for some $\alpha \in (0,1)$. Or we can use the logarithm to obtain an alternative dimensionless quantity:

$$\sum_n \left|\ln \frac{S_{n+1}}{S_n}\right|^p.$$

These alternatives also merit serious study.

## 10.6 HEDGING ERROR WITH RELATIVE VARIATIONS FOR $S$

As we will see in this section, we can obtain much better error bounds on our Black-Scholes hedging error using relative variation rather than absolute variation for $S$ (although comparison is impeded by the fact that these error bounds depend on different assumptions about the payoff function). We verify this very quickly for the American data that we used in §10.4. Then we look at another data set, for which we have a more convincing way of estimating the prices for $\mathcal{D}$.

### The American Data, using Absolute Variations for Oracular $D$ and Relative Variations for $S$

Figures 10.8 and 10.9 show the relative variation spectra for the Microsoft and S&P 500 data we studied earlier. Because these relative variation spectra are roughly similar to the absolute variation spectra shown in Figures 10.1 and 10.2, this data should satisfy the relative Black-Scholes protocol with approximately the same $\delta$ as we used in §10.4 when we studied it under the absolute Black-Scholes protocol. Hence we can compare the error bound $40c\delta^{1/4}$ given by Proposition 10.4 for the relative protocol with the bound $8ce^{5C}\delta^{1/4}$ given by Proposition 10.3 for the absolute protocol. Because 40 is much less than $8e^{5C}$ for reasonable values of $C$, the bound based on relative variation should be much better.

For Microsoft, for example, we can take $\delta \approx 2^{-6}$ in (10.60), especially if (10.54) is replaced by what is actually used in the proof,

$$\inf_{p,q\in(0,1):1/p+1/q=1} \max\left(\mathbf{var}_S^{\mathrm{rel}}(p),\ \mathbf{var}_D(q)\right) < \delta.$$

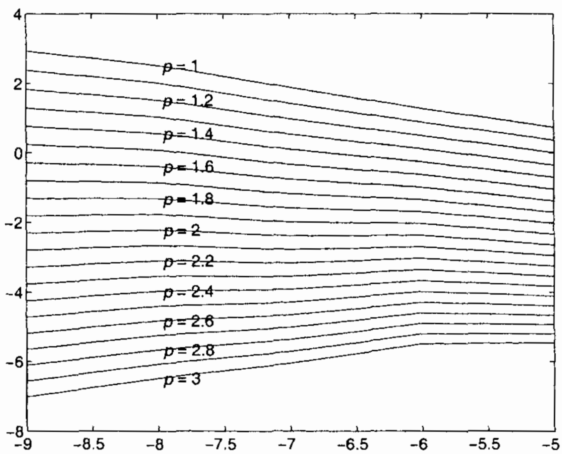



***Fig. 10.8*** The loglog plot of $\mathbf{var}^{\mathrm{rel}}(p)$ against the time step for Microsoft.

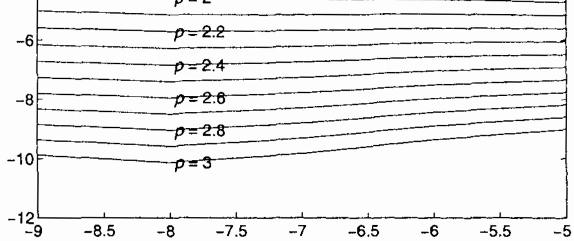



***Fig. 10.9*** The loglog plot of $\mathbf{var}^{\mathrm{rel}}(p)$ against the time step for the S&P 500.

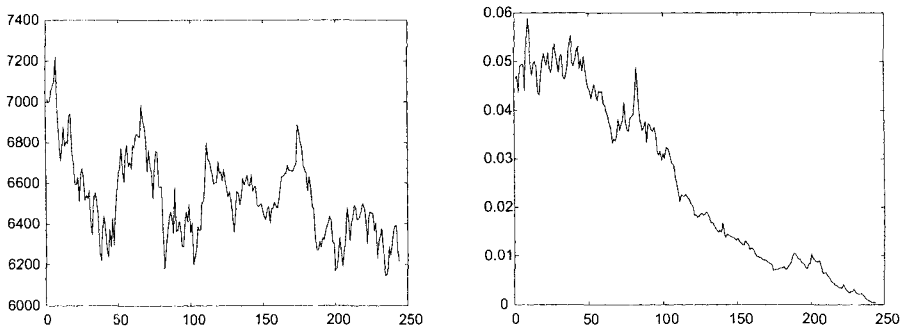



***Fig. 10.10*** The graph on the left shows values of the FTSE 100 index for 244 trading days ending in December 2000. The graph on the right shows the index's implied $D$ for this period, calculated from at-the-money implied volatility.

This gives a reasonable accuracy if $c_2$ through $c_4$ are not too large. Even (10.55), the hedging accuracy declared in the statement of the proposition, gives approximately 14 when $\delta = 2^{-6}$ and $c = 1$; this is still impractical but much better than the 420 we had in the previous section.

## The British Data, using Absolute Variations for $D$ Implied by the Market Prices of Calls and Puts and Relative Variations for $S$

We now look at recent data for the FTSE 100 index and for the stock prices of two British companies, Vodafone and BP (British Petroleum) Amoco. In April 2001 these two companies (ticker symbols VOD and BP) were by far the two largest in Europe by market capitalization, each accounting for approximately 5% of the FTSE Eurotop 100 index. Vodafone's price was quickly dropping and was very volatile (and, also important for us, even "second order" volatile, with implied volatility undergoing rapid changes), whereas BP Amoco shares were relatively stable.

Our data, obtained from LIFFE, the London International Financial Futures and Options Exchange [205], also gives implied volatility from at-the-money call and put options on the index and the two stocks. In the case of the FTSE 100, we use data on European calls and puts (ticker symbol ESX) with maturity in December 2000 and a life span of 244 days. (We removed 28 April 2000 from the data sets; on that day ESX migrated from pit trading to electronic trading, and all volatilities for that day are listed as 0.) In the case of Vodafone and BP, we use data on American calls and puts with maturity in October 2000 and a life span of 187 days. For all the three cases, the LIFFE data gives the at-the-money implied volatility for the underlying as a function of time—that is, the volatility $\sigma(t)$ implied by market prices for calls and puts with strike prices at the current value for the underlying. Because LIFFE

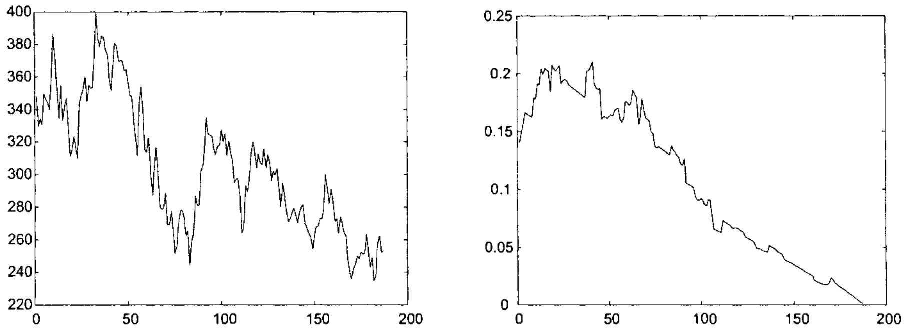



***Fig. 10.11*** The graph on the left shows stock prices for Vodafone for 187 trading days ending in October 2000. The graph on the right shows the stock's implied $D$ for this period, calculated from at-the-money implied volatility.

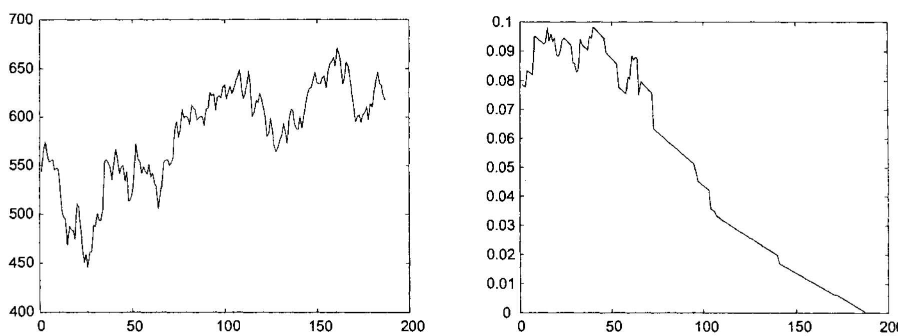



***Fig. 10.12*** The graph on the left shows stock prices for BP Amoco for 187 trading days ending in October 2000. The graph on the right shows the stock's implied $D$ for this period, calculated from at-the-money implied volatility.

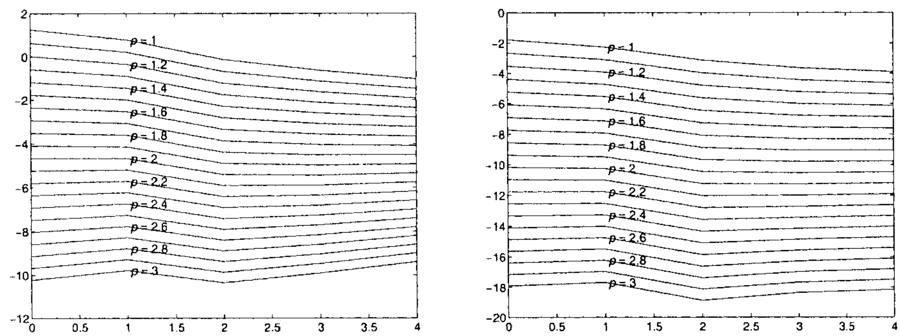



***Fig. 10.13*** The variation spectra for $S$ (on the left) and $D$ (on the right) for FTSE 100. The horizontal axis shows the time step, from 1 to 16 days, in logarithms to the base 2. This and the remaining figures in this section are all loglog plots with logarithms to base 2.

assumes put/call parity, the option's class (call or put) is not important. The unit for $t$ is one year, which we equate with 250 trading days.

From the implied volatilities given by LIFFE, we can estimate an implied price path for $\mathcal{D}$:

$$D(t) = (T - t)\sigma^2(t),$$

where the time $t$ is measured from the time the option started trading, and $T$ is the option's life span. (There would be difficulties, however, in replicating $\mathcal{D}$ using the at-the-money calls and puts; see §12.2.) We can then compute the variation spectrum of the path. The price processes for the underlyings and the implied $D$ are shown in Figures 10.10–10.12.

Figures 10.13–10.15 show the variation plots for $S$ and implied $D$ for the three underlyings. In our further discussion, we will use the information for a time step of 1 day (the left side of the bounding rectangle). When the time step is 16 days (the right side), there are too few observations (just 11 in the case of Vodafone and BP Amoco shares) for the computations to be reliable. Because our time spans are no longer powers of 2 (as they were in our data for Microsoft and the S&P 500), we now label the horizontal axis with the log of the time period in days. We will be interested, in each of the three cases, in the 3-variation for the underlying ($p = 3$ on the left) and the 2-variation for the implied $D$ ($p = 2$ on the right).

Figures 10.16–10.18 report all four quantities occurring in Equation (10.57) and contributing to the accuracy of hedging: $\mathbf{var}_S^{\mathrm{rel}}(3)$, $\mathbf{var}_D(2)$ (these two quantities are shown in the previous figures as well), and the cross-variations

$$\mathrm{cov}(1,1) := \sum_n \left|\frac{dS_n}{S_n}\right| |dD_n|$$

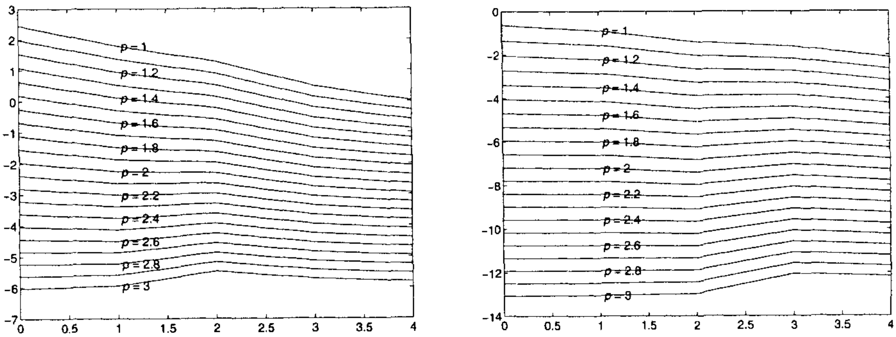



***Fig. 10.14*** The variation spectra for $S$ (on the left) and $D$ (on the right) for Vodafone.

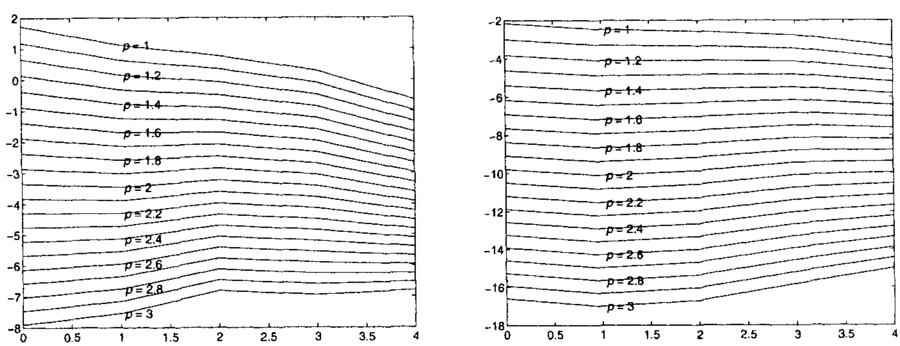



***Fig. 10.15*** The variation spectra for $S$ (on the left) and $D$ (on the right) for BP Amoco.

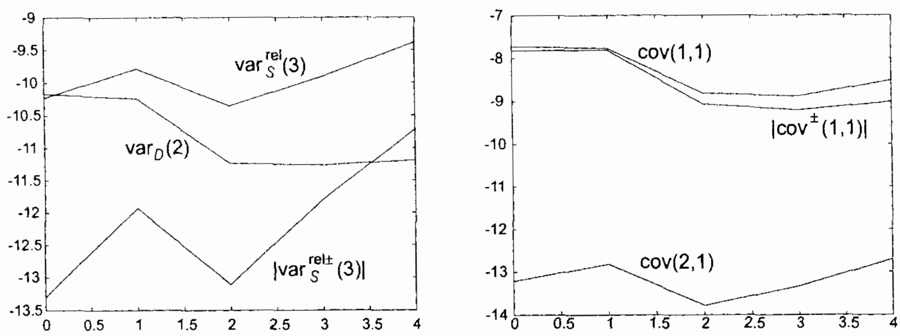



***Fig. 10.16*** Variations and cross-variations for the FTSE 100.

and

$$\operatorname{cov}(2,1) := \sum_n \left| \frac{dS_n}{S_n} \right|^2 |dD_n|.$$

The values given by $\mathbf{var}_S^{\mathrm{rel}}(3)$ and $\operatorname{cov}(1,1)$ may be too pessimistic, inasmuch as the corresponding hedging errors can cancel each other out. So we also report the values

$$\left|\mathbf{var}_S^{\mathrm{rel}\pm}(3)\right| := \left|\sum_n \left(\frac{dS_n}{S_n}\right)^3\right|$$

and

$$\left|\operatorname{cov}^{\pm}(1,1)\right| := \left|\sum_n \frac{dS_n}{S_n} dD_n\right|.$$

To assess the hedging accuracy corresponding to these pictures, assume (somewhat arbitrarily) that $c_2$–$c_4$ in (10.56) do not exceed 1. Then the hedging accuracy as given by the right-hand side of (10.57) is bounded by

$$\frac{1}{2}\,\mathbf{var}_S^{\mathrm{rel}}(3) + \frac{7}{4}\operatorname{cov}(2,1) + \frac{3}{2}\operatorname{cov}(1,1) + \frac{7}{8}\,\mathbf{var}_D(2)$$

(we have used (10.58)–(10.59) here). Substituting the numbers obtained as described above from the LIFFE data, we obtain the following bounds (with accuracy 0.001): 0.008 for FTSE 100, 0.052 for Vodafone, and 0.011 for BP Amoco. The largest contribution to these totals is invariably from $\operatorname{cov}(1,1)$ (0.036 in the case of Vodafone); if we are lucky, the addends in the second sum in the right-hand side of (10.57) (corresponding to $\operatorname{cov}(1,1)$) will cancel each other out; see the $|\operatorname{cov}^{\pm}(1,1)|$ lines in the pictures.

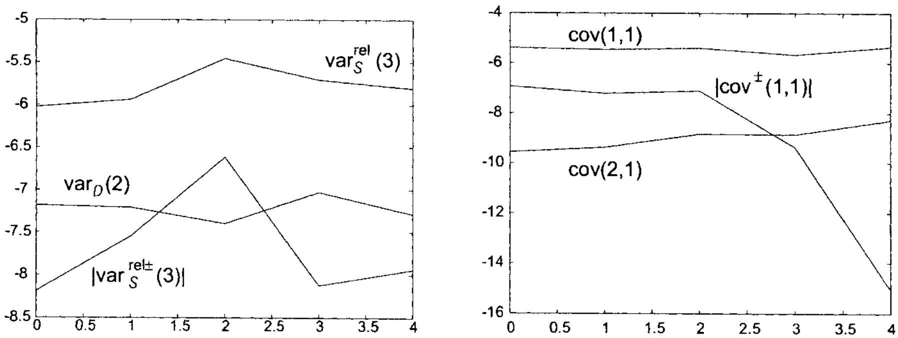



***Fig. 10.17*** Variations and cross-variations for Vodafone.

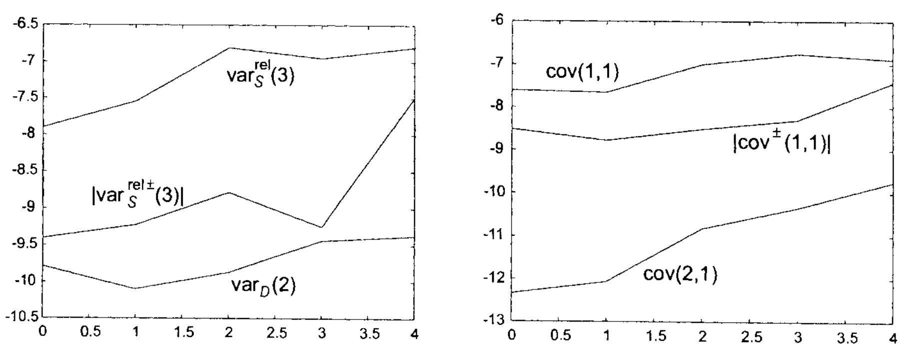



***Fig. 10.18*** Variations and cross-variations for BP Amoco.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 11

# *Games for Pricing Options in Continuous Time*

Abraham Robinson (1918–1974), the inventor of nonstandard analysis, photographed in 1971 on his appointment as Sterling Professor, Yale University.

In the preceding chapter, we demonstrated that a Black-Scholes game for option pricing is feasible in a realistic discrete-time setting. In our view, this justifies a theoretical exploration, which we will now undertake, of game-theoretic option pricing in continuous time. Our continuous-time theory will neglect details of implementation but will give a clearer and broader perspective.

One of the most important purposes of this chapter is to introduce our way of handling a continuous market game. In order to do so as clearly and simply as possible, we limit our scope: we consider only the simplest forms of Bachelier and Black-Scholes option pricing. We leave for later chapters issues that complicate the picture, such as interest rates, price processes with jumps, the pricing of American options, and so on.

We use nonstandard analysis. Although it is still novel for many applied mathematicians, nonstandard analysis is well adapted to game theory, because, as we will show, it allows us to import into continuous time the fundamental picture in which two players alternate moves. Our time interval is a continuum—all real numbers between 0 and $T$ are included, where $T$ is a positive real number. But we divide the interval $[0,T]$ into an infinitely large number $N$ of steps of equal infinitesimal length $dt$; $dt := T/N$. The time horizon $T$ and the infinitely large positive integer $N$ will be fixed throughout the rest of this book. We designate the resulting set of time points by $\mathbb{T}$:

$$\mathbb{T} := \{ndt \mid 0 \leq n \leq N\}.$$

The infinitely large number $N$ is the number of rounds of play in our game between Investor and Market, and the infinitesimal number $dt$ is the amount of time each round takes. After an initial move by Market at time 0, the first round takes place between time 0 and time $dt$, the second between time $dt$ and time $2dt$, and so on. The last round takes place between time $(N-1)dt$ and time $Ndt = T$. If $1 < n < N-1$, then the $n$th round, which takes place between time $(n-1)dt$ and time $ndt$, has an immediately preceding round and an immediately following round, even if $n$ is already infinitely large.

We use only simplest notions of nonstandard analysis, such as *infinitely large* and *infinitesimal* (infinitely close to zero). The reader will be able to follow our reasoning at an intuitive level if he or she reads these terms simply as "very large" and "very small". But they have an exact meaning in nonstandard analysis, and our reasoning is rigorous when these exact meanings are used. Explanations sufficient to allow the reader to understand the limited parts of nonstandard analysis we use are provided in an appendix, §11.5.

Nonstandard analysis studies the *hyperreal numbers*, which consist of (1) the ordinary real numbers, which are called *standard*, together with (2) *nonstandard numbers*. A nonstandard number is either infinite (positive or negative) or else differs from a standard number by an infinitesimal amount. The set $\mathbb{T}$ defined above is a subset of the hyperreal numbers; it includes all the real numbers in the interval $[0,T]$, together with many (but not all) of the nonstandard numbers that lie infinitely close to a real number in $[0,T]$.

Our claim that the continuous theory of this chapter is justified by the discrete-time theory of the preceding chapter is true quite literally. As we mentioned in §1.5, the *transfer principle* of nonstandard analysis allows some nonstandard theorems to be deduced directly from corresponding standard theorems ([136], Chapter 4). Aside from a few details, such is the situation here; the principal continuous-time results in this chapter follow from the corresponding discrete-time results of the preceding chapter:

- Our continuous-time Bachelier formula, Theorem 11.1 (p. 278), follows by the transfer principle from Proposition 10.2 (p. 244).

- We did not state a discrete-time result corresponding exactly to our main statement of the Black-Scholes formula, Theorem 11.2 (p. 280), but application of the transfer principle to Proposition 10.3 (p. 249) under the assumption that $U$ is smooth with bounded $U^{(2)}$–$U^{(4)}$ produces a continuous-time result that makes Theorem 11.2 nearly obvious.

- Our continuous-time relative Black-Scholes formula, Theorem 11.3 (p. 282) follows by the transfer principle from Proposition 10.4 (p. 260).

(In fact, these implications are very simple and will be explained in §11.5: see p. 286.) Thus our continuous-time theory is literally a simplification of our discrete-time theory, in which important but messy practical details are cleared away. The absence of these details makes the proofs of the continuous-time theorems much simpler than the proofs of the corresponding discrete-time results. In order to make this clear, we prove directly all the results in this chapter.

We begin our exposition, here as in Chapter 9, by defining the variation spectrum and the variation and Hölder exponents (§11.1). Then we derive our game-theoretic versions of Bachelier's formula (§11.2) and the Black-Scholes formula (§11.3). In a final section, §11.4, we consider an important implication of our continuous-time game-theoretic treatment of option pricing: in a market where the variance of the price of a security $S$ is traded, Investor can force Market to make her $S(t)$ have Hölder exponent $1/2$ (i.e., Investor can become infinitely rich if Market does not do so). This is the $\sqrt{dt}$ effect, which is assumed at the outset by the diffusion model.

In addition to the appendix in which we provide a brief introduction to nonstandard analysis (§11.5), we also include a second appendix, §11.6. in which we show that our definition of the Hölder exponent (which differs from the definitions usually used in stochastic theory) gives the expected result in the diffusion model: a random function that obeys the diffusion stochastic differential equation does indeed have Hölder exponent $1/2$.

## 11.1 THE VARIATION SPECTRUM

In this section, we define the variation spectrum, the variation exponent, and the Hölder exponent for nonstandard continuous functions. In addition to the variation spectrum with absolute differences, which we introduced in §9.1 and emphasized in the preceding chapter, we also study the relative variation spectrum, which we introduced in §10.5.

### Continuity for Nonstandard Functions

A real-valued function of real variables (such as $x \mapsto x^2$ or $x, y \mapsto x + y$) automatically extends to a hyperreal-valued function of hyperreal variables (p. 284). This type of hyperreal function is called *standard*, although it may assign nonstandard values to nonstandard numbers (e.g., if $x$ and $y$ are nonstandard, then $x + y$ may be as well). Other hyperreal-valued functions are called *nonstandard*; they may assign nonstandard values even to standard numbers.

Given a function $f$ on $\mathbb{T}$, we write $df(t)$ for $f(t+dt) - f(t)$ whenever $t \in \mathbb{T} \setminus \{T\}$, and we call $f$ *continuous* if $\sup_{t \in \mathbb{T} \setminus \{T\}} |df(t)|$ is infinitesimal. In the case where $f$ is standard, this condition is equivalent to $f$ being continuous in the ordinary sense when it is considered as a real-valued function on the interval $[0, T]$ (see p. 286).

## The Variation Spectrum

Given a continuous, possibly nonstandard, function $f$ on $\mathbb{T}$, and a real number $p \in [1, \infty)$, we set

$$\mathbf{var}_f(p) := \sum_{t \in \mathbb{T} \setminus \{T\}} |df(t)|^p. \tag{11.1}$$

We call the number $\mathbf{var}_f(p)$, which may be nonstandard, the *p-variation* of $f$. We call the function $\mathbf{var}_f$ the *variation spectrum.*

**Lemma 11.1** *Suppose $f$ is a continuous, possibly nonstandard, function, on $\mathbb{T}$. Then there exists a unique real number*[1] $\operatorname{vex} f \in [1, \infty]$ *such that*

- $\mathbf{var}_f(p)$ *is infinitely large when* $1 \leq p < \operatorname{vex} f$, *and*
- $\mathbf{var}_f(p)$ *is infinitesimal when* $p > \operatorname{vex} f$.

*Proof* If $f$ is constant, then $\mathbf{var}_f(p) = 0$ for all $p \in [1, \infty)$. This means that $\operatorname{vex} f$ exists and is equal to 1.

Suppose now that $f$ is not constant. Then it suffices to consider real numbers $p_1$ and $p_2$ satisfying $1 \leq p_1 < p_2$ and to show that the ratio

$$\sum_t |df(t)|^{p_2} \Big/ \sum_t |df(t)|^{p_1}, \tag{11.2}$$

where $t$ ranges over $\mathbb{T} \setminus \{T\}$, is infinitesimal. (Because $f$ is not constant, the denominator is positive.) To do this, we show that (11.2) is less than $\epsilon$ for an arbitrary positive real number $\epsilon$. Let $\epsilon_1 > 0$ be so small that $\epsilon_1^{p_2 - p_1} < \epsilon$. Since $f$ is continuous, $|df(t)| \leq \epsilon_1$ for all $t$. So we have:

$$\sum_t |df(t)|^{p_2} = \sum_t |df(t)|^{p_2 - p_1} |df(t)|^{p_1} \leq \epsilon_1^{p_2 - p_1} \sum_t |df(t)|^{p_1} < \epsilon \sum_t |df(t)|^{p_1}. \quad \blacksquare$$

We call $\operatorname{vex} f$ the *variation exponent* of $f$, and we call $1/\operatorname{vex} f$ the *Hölder exponent.* We write $H(f)$ for the Hölder exponent. We defined $\operatorname{vex} f$ and $H(f)$ in this same way in §9.1, but there these definitions were merely heuristic. By dint of fixing a particular infinitely large integer $N$, we have ensured that $\operatorname{vex} f$ and $H(f)$ are well defined in the present idealized continuous-time context.

If $\operatorname{vex} f = 1$, then we say that $f$ has *bounded variation.* It is clear that $\operatorname{vex} f = 1$ when $f$ is bounded and monotonic, for then $\mathbf{var}_f(1) = \sum_t |df(t)| = |f(T) - f(0)|$, which is finite and not infinitesimal. We obtain the same conclusion when $f$ is bounded and $[0, 1]$ can be divided into a finite number of intervals where $f$ is monotonic. This justifies the loose statement, which we made in §9.1, that ordinary well-behaved functions have variation exponent 1.

We call $f$ *stochastic* when $\operatorname{vex} f = 2$, *substochastic* when $\operatorname{vex} f < 2$, and *superstochastic* when $\operatorname{vex} f > 2$. As we will confirm in §11.6, we can expect the path of a diffusion process to be stochastic. A substochastic function is less jagged than the path of a diffusion process; a superstochastic one is more jagged.

[1]Sometimes we include $\infty$ and $-\infty$ in the real numbers.

### The Relative Variation Spectrum

We turn now to the relative variation spectrum, which we studied for the discrete case in §10.5 (p. 259). As we will see, it normally leads, in the present idealized setting, to the same variation and Hölder exponents as the absolute variation spectrum.

We call a possibly nonstandard function $f$ on $\mathbb{T}$ *positive* if $f(t) > 0$ for all $t \in \mathbb{T}$. (This does not forbid infinitesimal values for $f(t)$.) We call a positive, possibly nonstandard, function $f$ *relatively continuous* if $\sup_{t \in \mathbb{T} \setminus \{T\}} df(t)/f(t)$ is infinitesimal. If $f$ is relatively continuous, we set

$$\mathbf{var}_f^{\mathrm{rel}}(p) := \sum_t \left| \frac{df(t)}{f(t)} \right|^p .$$

As in the discrete case, we call $\mathbf{var}_f^{\mathrm{rel}}(p)$ the *relative p-variation* of $f$, and we call the function $\mathbf{var}_f^{\mathrm{rel}}$ the *relative variation spectrum.*

**Lemma 11.2** *If $f$ is a relatively continuous positive, possibly nonstandard, function on $\mathbb{T}$, then there exists a unique real number* $\operatorname{vex}^{\mathrm{rel}} f \in [1, \infty]$ *such that*

- $\mathbf{var}_f^{\mathrm{rel}}(p)$ *is infinitely large when* $1 \leq p < \operatorname{vex}^{\mathrm{rel}} f$,
- $\mathbf{var}_f^{\mathrm{rel}}(p)$ *is infinitesimal when* $p > \operatorname{vex}^{\mathrm{rel}} f$.

The proof is analogous to that of Lemma 11.1. We call $\operatorname{vex}^{\mathrm{rel}} f$ the *relative variation exponent* of $f$.

A possibly nonstandard function $f$ on $\mathbb{T}$ is *strictly positive* if there exists a real number $\epsilon > 0$ such that $f(t) > \epsilon$ for all $t \in \mathbb{T}$. It is *bounded* if there exists a real number $C < \infty$ such that $\sup_t |f(t)| < C$.

**Lemma 11.3** *If a strictly positive and bounded nonstandard function $f$ is relatively continuous, then* $\operatorname{vex}^{\mathrm{rel}} f$ *coincides with its absolute counterpart:* $\operatorname{vex}^{\mathrm{rel}} f = \operatorname{vex} f$.

This lemma follows from the coincidence of $\mathbf{var}_f(p)$ and $\mathbf{var}_f^{\mathrm{rel}}(p)$ to within a constant factor.

## 11.2 BACHELIER PRICING IN CONTINUOUS TIME

As in the preceding chapter, we begin, for simplicity, with the Bachelier game, even though it has less practical importance because it allows negative stock prices.

In the Bachelier game, two securities are available to Investor: a security $\mathcal{S}$, which pays no dividends, and a security $\mathcal{D}$, which pays a regular dividend equal to the square of the most recent change in the price of $\mathcal{S}$. Just as in the discrete case, Market sets prices for $\mathcal{S}$ and $\mathcal{D}$ at $N+1$ successive points in time:

$$\text{prices } S_0, \ldots, S_N \text{ for } \mathcal{S}, \text{ and prices } D_0, \ldots, D_N \text{ for } \mathcal{D},$$

and the dividend paid by $\mathcal{D}$ at point $n$ is $(\Delta S_n)^2$, where $\Delta S_n := S_n - S_{n-1}$. But now $N$ is infinitely large, and the sequences $S_0, \ldots, S_N$ and $D_0, \ldots, D_N$ define functions $S$ and $D$, respectively, on our time interval $\mathbb{T}$:

$$S(ndt) := S_n \quad \text{and} \quad D(ndt) := D_n$$

for $n = 0, \ldots, N$.

The protocol for the Bachelier game can be written exactly as in the preceding chapter (p. 244):

Market announces $S_0 \in \mathbb{R}$ and $D_0 > 0$.
FOR $n = 1, 2, \ldots, N$:
  Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
  Market announces $S_n \in \mathbb{R}$ and $D_n \geq 0$.
  $\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n((\Delta S_n)^2 + \Delta D_n)$. (11.3)

We can also write (11.3) as

$$\mathcal{I}_{n+1} := \mathcal{I}_n + M_{n+1}(S_{n+1} - S_n) + V_{n+1}((S_{n+1} - S_n)^2 + (D_{n+1} - D_n))$$

or

$$d\mathcal{I}_n := M_{n+1} dS_n + V_{n+1}((dS_n)^2 + dD_n) \tag{11.4}$$

for $n = 0, \ldots, N-1$. Here, as always, $df_n := f_{n+1} - f_n$.

Investor decides on his moves $M_n$ and $V_n$ in the situation

$$S_0 D_0 \ldots S_{n-1} D_{n-1}. \tag{11.5}$$

So a strategy for Investor is a pair of functions, say $\mathcal{M}$ and $\mathcal{V}$, each of which maps each situation of the form (11.5), for $n = 1, \ldots, N$, to a real number. When Investor uses the strategy $(\mathcal{M}, \mathcal{V})$, and Market's moves are recorded by the functions $S$ and $D$, (11.4) becomes

$$d\mathcal{I}_n := \mathcal{M}(S_0 D_0 \ldots S_n D_n) dS_n + \mathcal{V}(S_0 D_0 \ldots S_n D_n)((dS_n)^2 + dD_n). \tag{11.6}$$

Let us write $\mathcal{I}^{\mathcal{M},\mathcal{V}}(S, D)$ for Investor's total change in capital over the course of the game when he follows the strategy $(\mathcal{M}, \mathcal{V})$ and Market plays $(S, D)$. This is obtained by summing the increments (11.6):

$$\begin{aligned} \mathcal{I}^{\mathcal{M},\mathcal{V}}(S, D) := \sum_{n=0}^{N-1} \Big( & \mathcal{M}(S_0 D_0 \ldots S_n D_n) dS_n \\ & + \mathcal{V}(S_0 D_0 \ldots S_n D_n)((dS_n)^2 + dD_n) \Big). \end{aligned} \tag{11.7}$$

If Investor starts with initial capital $\alpha$ at time 0, then his final capital at time $T$ will be $\alpha + \mathcal{I}^{\mathcal{M},\mathcal{V}}(S, D)$.

If Investor's decision on how many units of each security to hold during a time period depends only on the current prices of the two securities, then $\mathcal{M}$ and $\mathcal{V}$ are merely functions of two variables (the strategies we construct for Investor will in fact have this Markovian property), and (11.7) reduces to

$$\mathcal{I}^{\mathcal{M},\mathcal{V}}(S,D) := \sum_{n=0}^{N-1} \Big( \mathcal{M}(S_n, D_n) dS_n + \mathcal{V}(S_n, D_n)((dS_n)^2 + dD_n) \Big).$$

This can also be written as

$$\mathcal{I}^{\mathcal{M},\mathcal{V}}(S,D) := \sum_{t \in \mathbb{T}\setminus\{T\}} \mathcal{M}(S(t), D(t)) dS(t) + \sum_{t \in \mathbb{T}\setminus\{T\}} \mathcal{V}(S(t), D(t)) \left((dS(t))^2 + dD(t)\right).$$

In discrete time, we say that $\alpha$ is an approximate price for a European option $\mathcal{U}$ if Investor has a strategy that produces approximately the same capital starting with $\alpha$. In the discrete-time Bachelier and Black-Scholes protocols, the approximation improves with the number of rounds played: for every positive real number $\epsilon$, there is a number of rounds $N$ that allows the price to be approximated to within $\epsilon$, provided that Market's moves obey the rules of the game. In our present setting, the number of rounds is infinite, and so we should be able to make the discrepancy between $\mathcal{U}$'s payoff and Investor's final capital $\alpha + \mathcal{I}^{\mathcal{M},\mathcal{V}}(S,D)$ smaller than every positive real number $\epsilon$. This is equivalent to saying that the difference should be infinitesimal.

So here is how we define the price of a European option $\mathcal{U}$ in continuous time. First we complete the definition of the game by fixing constraints on Market's paths $S$ and $D$ under which our protocol is coherent. Then we select the payoff function $U$ for $\mathcal{U}$. Finally, we consider a real number $\alpha$. We say that $\alpha$ is the *price* of $\mathcal{U}$ at time 0 (just after Market has announced his initial moves $S_0$ and $D_0$) if for any $\epsilon > 0$ Investor has a strategy $(\mathcal{M}, \mathcal{V})$ such that

$$\left|\alpha + \mathcal{I}^{\mathcal{M},\mathcal{V}}(S,D) - U(S(T))\right| < \epsilon \tag{11.8}$$

for all paths $S$ and $D$ permitted by the rules of the game. If a price exists, it is unique. Indeed, if (11.8) holds and Investor also has a strategy $(\mathcal{M}', \mathcal{V}')$ such that

$$\left|\alpha' + \mathcal{I}^{\mathcal{M}',\mathcal{V}'}(S,D) - U(S(T))\right| < \epsilon,$$

then

$$\left|\left(\mathcal{I}^{\mathcal{M}-\mathcal{M}',\mathcal{V}-\mathcal{V}'}(S,D)\right) - (\alpha - \alpha')\right| < 2\epsilon.$$

This means that Investor can obtain capital arbitrarily close to $\alpha - \alpha'$ by playing $(\mathcal{M} - \mathcal{M}', \mathcal{V} - \mathcal{V}')$ starting with zero, and coherence then implies that $\alpha = \alpha'$.

We should note the limited role played by nonstandard objects in this machinery. The payoff $U$ is a real valued function of a real variable, and the price $\alpha$ is a real number. The moves $M_n$ and $V_n$ made by Investor are also always real numbers.

Only the time interval $\mathbb{T}$, the paths $S$ and $D$ taken by Market, and the resulting capital process $\mathcal{I}^{\mathcal{M},\mathcal{V}}(S,D)$ are nonstandard.

There are several different sets of constraints for Market's moves that will produce the usual Gaussian prices for European options in the Bachelier game. Here we assume that $S$ is not superstochastic and that $D$ is substochastic: $\operatorname{vex} S \leq 2$ and $\operatorname{vex} D < 2$. In terms of the Hölder exponents: $H(S) \geq 1/2$ and $H(D) > 1/2$. A weaker set of conditions that also works is $H(S) > 1/3$, $H(D) > 1/2$, and $H(S) + H(D) > 1$.

BACHELIER'S PROTOCOL IN CONTINUOUS TIME

**Players:** Investor, Market

**Protocol:**

$\mathcal{I}_0 := 0$.

Market announces $S_0 \in \mathbb{R}$ and $D_0 > 0$.

FOR $n = 1, 2, \ldots, N$:

Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.

Market announces $S_n \in \mathbb{R}$ and $D_n \geq 0$.

$\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n((\Delta S_n)^2 + \Delta D_n)$.

**Additional Constraints on Market:** Market must ensure that $S$ and $D$ are continuous and satisfy $\operatorname{vex} S \leq 2$ and $\operatorname{vex} D < 2$ and that $S_0$, $D_0$ are neither infinitely large nor infinitesimal. Moreover, he must make $D_n > 0$ for $n = 1, \ldots, N-1$ and $D_N = 0$.

The coherence of this protocol is verified as usual: Market can keep Investor from making money by setting, say, $D_0 := T$ and then always setting $\Delta D_n = -dt$ and $\Delta S_n = \pm\sqrt{dt}$, with the sign opposite that of $M_n$. In this case $\operatorname{vex} S = 2$ and $\operatorname{vex} D = 1$. (The path $D$ decreases linearly from $D_0$ to 0.)

**Theorem 11.1** *Let $U\colon \mathbb{R} \to \mathbb{R}$ be Lipschitzian. Then in Bachelier's protocol in continuous time, the price at time 0 (right after $S(0)$ and $D(0)$ are announced) for the European option $U(S(T))$ is*

$$\int_{\mathbb{R}} U(S(0) + z)\ \mathcal{N}_{0,D(0)}(dz).$$

*Proof* This proof is modeled on the proof of Proposition 10.2 (p. 244). The ingredients are all familiar. First we assume that $U$ is smooth with bounded $U^{(3)}$ and $U^{(4)}$, define $\overline{U}(s,D)$ by (6.10), and use the strategy $(\mathcal{M},\mathcal{V})$ given by (10.13): $\mathcal{M}(s,D) := \partial\overline{U}/\partial s$ and $\mathcal{V}(s,D) := \partial\overline{U}/\partial D$. We need to show that

$$\overline{U}(S(0), D(0)) + \mathcal{I}^{\mathcal{M},\mathcal{V}}(S,D) \approx \overline{U}(S(T), D(T)). \tag{11.9}$$

Letting $t$ run over $\mathbb{T} \setminus \{T\}$, we rewrite (10.14) as

$$\begin{aligned}
&\left|(\overline{U}(S(T), D(T)) - \overline{U}(S(0), D(0))) - \mathcal{I}^{\mathcal{M},\mathcal{V}}(S, D)\right| \\
&\le \frac{1}{2} \sup \left|\frac{\partial^3 \overline{U}}{\partial s^3}\right| \mathbf{var}_S(3) + \frac{1}{2} \sup \left|\frac{\partial^3 \overline{U}}{\partial D \partial s^2}\right| \sum_t |dD(t)||dS(t)|^2 \\
&+ \sup \left|\frac{\partial^2 \overline{U}}{\partial D \partial s}\right| \sum_t |dD(t)||dS(t)| + \frac{1}{2} \sup \left|\frac{\partial^2 \overline{U}}{\partial D^2}\right| \mathbf{var}_D(2),
\end{aligned} \tag{11.10}$$

all suprema being taken over the convex hull of $\{(S(t), D(t)) \mid 0 < t < T\}$ (or, in this proof, even the whole of $\mathbb{R} \times (0, \infty)$).

We are required to show that the right-hand side of (11.10) is infinitesimal. Let $p := 1 + \epsilon$ and $q := 1 - \epsilon$ for some $\epsilon \in (0, 1)$ such that $\operatorname{vex} D < q$; we know that $\operatorname{vex} S < p$. Remember that we can use Hölder's inequality (10.11), since $1/p + 1/q \ge 1$. First we will show that

$$\sum_t |dD(t)||dS(t)|^2 \quad \text{and} \quad \sum_t |dD(t)||dS(t)|$$

are infinitesimal. This is easy: Hölder's inequality gives

$$\sum_t |dD(t)||dS(t)| \le \left(\sum_t |dD(t)|^q\right)^{1/q} \left(\sum_t |dS(t)|^p\right)^{1/p} \approx 0; \tag{11.11}$$

and the uniform continuity of $S$ on $[0, T]$ gives $\sup_t |dS(t)| \le 1$ and, thus,

$$\sum_t |dD(t)||dS(t)|^2 \le \sum_t |dD(t)||dS(t)| \approx 0. \tag{11.12}$$

In the preceding chapter (cf. (6.22) and (10.15)) we saw that all suprema in (11.10) are finite. Therefore, the right-hand side of (11.10) is infinitesimal, which completes the proof for the case of smooth $U$ with bounded $U^{(3)}$ and $U^{(4)}$.

The assumption that $U$ is smooth with bounded $U^{(3)}$ and $U^{(4)}$ is dropped in the same way as in the proof of Proposition 10.2: for a small $\sigma > 0$ define $V$ by (10.22), apply (11.9) to $V$, and notice that $U(S(T))$ is close to $V(S(T))$ (by (10.24)) and $\int U d\mathcal{N}_{S(0),D(0)}$ is close to $\int V d\mathcal{N}_{S(0),D(0)}$ (by (10.25)). This proves the theorem. ∎

## 11.3 BLACK-SCHOLES PRICING IN CONTINUOUS TIME

We turn now to the Black-Scholes protocol. This differs from the Bachelier protocol primarily in that we require $S_n$ to be positive and have $\mathcal{D}$ pay $(\Delta S_n / S_{n-1})^2$ instead of $(\Delta S_n)^2$ as a dividend. We also relax the condition $\operatorname{vex} S \le 2$.

THE BLACK-SCHOLES PROTOCOL IN CONTINUOUS TIME
**Players:** Investor, Market
**Protocol:**
$\mathcal{I}_0 := 0$.
Market announces $S_0 > 0$ and $D_0 > 0$.
FOR $n = 1, 2, \dots, N$:

Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Market announces $S_n > 0$ and $D_n \geq 0$.
$\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n((\Delta S_n / S_{n-1})^2 + \Delta D_n)$.

**Additional Constraints on Market:** Market must ensure that $S$ is continuous, $\inf_n S_n$ is positive and not infinitesimal, and $\sup_n S_n$ is finite. He must also ensure that $D$ is continuous, $D_n > 0$ for $n = 1, \ldots, N-1$, $D_N = 0$, $\sup_n D_n$ is finite, and $\operatorname{vex} D < 2$.

**Theorem 11.2** *Let $U\colon (0,\infty) \to \mathbb{R}$ be Lipschitzian and bounded below. Then in the Black-Scholes protocol in continuous time, the price at time 0 (right after $S(0)$ and $D(0)$ are announced) for the European option $U(S(T))$ is*

$$\int_{\mathbb{R}} U(S(0)e^z)\, \mathcal{N}_{-D(0)/2,\, D(0)}(dz). \tag{11.13}$$

*Proof* This proof is modeled on the proofs of Theorem 11.1 and Proposition 10.3. First we assume that $\operatorname{vex} S \leq 2$ and that $U$ is a smooth function such that the derivatives $U^{(1)}$–$U^{(4)}$ are bounded.

Our goal is to find a strategy $(\mathcal{M}, \mathcal{V})$ such that

$$\int_{\mathbb{R}} U(S(0)e^z)\, \mathcal{N}_{-D(0)/2,\, D(0)}(dz) + \mathcal{I}^{\mathcal{M},\mathcal{V}}(S, D) \approx U(S(T)). \tag{11.14}$$

Defining $\overline{U}(S, D)$ by (10.37), we can rewrite (11.14) as (11.9). The strategy $(\mathcal{M}, \mathcal{V})$ is the same as before, (10.13).

In our present continuous-time notation we can rewrite (10.40) as

$$\begin{aligned}
&\left|\left(\overline{U}(S(T), D(T)) - \overline{U}(S(0), D(0))\right) - \mathcal{I}^{\mathcal{M},\mathcal{V}}(S, D)\right| \\
&\leq \frac{1}{2}\sup\left|\frac{\partial^3 \overline{U}}{\partial s^3}\right| \mathbf{var}_S(3) + \frac{1}{2}\sup\left|\frac{\partial^3 \overline{U}}{\partial D \partial s^2}\right| \sum_t |dD(t)||dS(t)|^2 \\
&\quad + \sup\left|\frac{\partial^2 \overline{U}}{\partial D \partial s}\right| \sum_t |dD(t)||dS(t)| + \frac{1}{2}\sup\left|\frac{\partial^2 \overline{U}}{\partial D^2}\right| \mathbf{var}_D(2).
\end{aligned}$$

We need to show that the right-hand side of this inequality is infinitesimal. In view of (11.11) and (11.12), it is sufficient to prove that all suprema in it are finite; this was done in the preceding chapter (see (10.41) and (10.42)). This completes the proof for the case of smooth $U$ with bounded $U^{(1)}$–$U^{(4)}$.

Now we drop the assumption that $U$ is smooth with bounded $U^{(1)}$–$U^{(4)}$. Taking a small $\sigma > 0$, define $V$ by (10.22). Combining (10.24), (10.48), and the boundedness of $\|V^{(2)}\|$–$\|V^{(4)}\|$ (Equations (10.49), (10.28), (10.29)) shows that (11.14) can be achieved for any Lipschitzian $U$.

It remains to remove the assumption $\operatorname{vex} S \leq 2$. Since $U$ is bounded below, we can assume that Investor's capital in the above construction never drops below some known constant (as soon as this constant is reached, Investor can start choosing zero moves, since it means that Market violated some of its obligations). But spending an arbitrarily small amount $\epsilon > 0$ on $\mathcal{D}$ will make sure that Investor's dividends

$$\sum_t \left(\frac{dS(t)}{S(t)}\right)^2 \geq \frac{\mathbf{var}_S(2)}{(\inf_t S(t))^2}$$

from holding $\mathcal{D}$ will be infinitely large when $\operatorname{vex} S > 2$ and so will more than compensate any losses incurred by his main hedging strategy. ∎

Notice the difference between Theorems 11.1 and 11.2: in the former we require that $\operatorname{vex} S \leq 2$, whereas in the latter this requirement is replaced by the requirement that $U$ should be bounded below and $\inf_t S(t)$ should not be infinitesimal. In principle, either of these requirements can be used in both theorems, but the condition that $U$ should be bounded below, or even nonnegative, is more natural in the Black-Scholes model: share prices being nonnegative makes nonnegative options natural. In practice, options are usually nonnegative.

The assumption $\operatorname{vex} D < 2$ can be regarded as a limit on the amount of new information that arrives concerning the future rates of return $dS(t)/S(t)$.

There already exists some literature on option pricing without probabilities. For example, Bick and Willinger (1994, Proposition 1) prove a path-wise variant of the Black-Scholes formula. However, Bick and Willinger do not go very far beyond the diffusion model: they essentially assume that

$$(S(t+dt) - S(t))^2 \approx \sigma^2 S^2(t)dt,$$

where $\sigma$ is a constant. In §14.5 we discuss some other work where stochastic processes are treated path by path.

## 11.4 THE GAME-THEORETIC SOURCE OF THE $\sqrt{dt}$ EFFECT

As we mentioned at the beginning of the chapter, Investor can multiply his capital substantially in our continuous Black-Scholes protocol unless Market makes $\operatorname{vex} S = 2$ (half of this result has just been used at the end of the proof of Theorem 11.2).

If $\operatorname{vex} S > 2$, it is possible to become infinitely rich buying our derivative security $\mathcal{D}$, and if $\operatorname{vex} S < 2$, it is possible to become infinitely rich shorting $\mathcal{D}$.

**Proposition 11.1** *For any $\epsilon > 0$ (arbitrarily small) there exists a strategy which, starting from $\epsilon$ at the moment when $S(0)$ and $D(0)$ are announced, never goes to debt and earns more than 1 if*

$$S \text{ is continuous}, \ \sup_{0 \leq t \leq T} S(t) < \infty, \ \operatorname{st}\left(\inf_{0 \leq t \leq T} S(t)\right) > 0 \qquad (11.15)$$

*(where $\operatorname{st}(a) > 0$ means that the hyperreal $a$ is positive and non-infinitesimal) and*

$$\operatorname{vex} S \neq 2.$$

*Proof* We assume that conditions (11.15) hold, and we show how to get rich when $\operatorname{vex} S < 2$ or $\operatorname{vex} S > 2$.

First we show how to get rich when $\operatorname{vex} S > 2$ (we have actually done in the proof of Theorem 11.2). Since $\operatorname{vex} S > 2$, $\mathbf{var}_S(2)$ is infinitely large, whereas $D(0)$ is finite. Buying \$$\epsilon$ worth of $\mathcal{D}$, we will get an infinitely large amount,

$$\epsilon \sum_t \left(\frac{dS(t)}{S(t)}\right)^2 \geq \epsilon \left(\sup_t S(t)\right)^{-2} \mathbf{var}_S(2),$$

in dividends, however small $\epsilon > 0$ is.

Now let us see how to hedge against $\operatorname{vex} S < 2$. In this case, $\mathbf{var}_S(2)$ is infinitesimal. Short selling \$1 worth of $\mathcal{D}$, we will get \$1 outright, whereas we will owe the infinitesimal amount

$$\sum_t \left(\frac{dS(t)}{S(t)}\right)^2 \leq \left(\inf_t S(t)\right)^{-2} \mathbf{var}_S(2)$$

of dividends that can be covered with an arbitrarily small initial investment $\epsilon$.

We can see that, sacrificing $\epsilon$, we can ensure that we get at least 1 when $\operatorname{vex} S$ is different from 2. This proves the proposition. ∎

It is instructive to compare this result with the following measure-theoretic results:

- Fractional Brownian motion with $h \neq 0.5$ is not a semimartingale (Rogers 1997; for $h > 0.5$, see also Example 4.9.2 of Liptser and Shiryaev 1986). Therefore (see, e.g., [206], Theorem 4.5.5), it cannot be turned into a martingale by an equivalent measure change. This shows that the traditional equivalent martingale measure approach to contingent claim valuation does not apply for the fractional Black-Scholes model.

- The geometric fractional Brownian motion with parameter $h \neq 0.5$ allows for arbitrage opportunities (this "almost follows" from the previous item), as shown by Rogers 1997 and, in a nonstandard framework for $h > 0.5$, Cutland, Kopp, and Willinger 1995. Remembering that the variation exponent of the fractional Brownian motion with parameter $h$ is $1/h$, we see that this corresponds to the substochastic and superstochastic cases, $\operatorname{vex} S \neq 2$.

### Black-Scholes with the Relative Variation Spectrum

Therefore, Theorem 11.2 will remain true if we replace $\operatorname{vex} S$ with $\operatorname{vex}^{\mathrm{rel}} S$ and impose the requirement that $S$ should be strictly positive and bounded. Proposition 10.4 on p. 260, however, implies the following simplified version:

**Theorem 11.3** *Suppose $U: \mathbb{R} \to \mathbb{R}$ is log-Lipschitzian and bounded below. Then the price for the European option $U(S(T))$ at the time when $S(0)$ and $D(0)$ have just been determined is*

$$\int_{\mathbb{R}} U(S(0)e^z)\, \mathcal{N}_{-D(0)/2,\, D(0)}(dz),$$

*provided that $S$ is positive and relatively continuous, $D$ is continuous, and* $\operatorname{vex} D < 2$.

The proof immediately follows from Proposition 10.4 (p. 260) combined with the following analog of Proposition 11.1.

**Proposition 11.2** *For any $\epsilon > 0$ (arbitrarily small) there exists a strategy which, starting from $\epsilon$ at the moment when $S(0)$ and $D(0)$ are announced, never goes to debt and earns more than* 1 *provided that $S$ is positive and relatively continuous and*

$$\operatorname{vex}^{\mathrm{rel}} S \neq 2.$$

The proof is similar to (but simpler than) the proof of Proposition 11.1.

## 11.5 APPENDIX: ELEMENTS OF NONSTANDARD ANALYSIS

Nonstandard analysis was invented by Abraham Robinson (1918–1974) in 1960 [78]. The first edition of his book on the subject appeared in 1966 [260].

Robinson's invention was based on insights into the logical foundations of mathematics. As he realized, the usual axioms for arithmetic do not rule out the existence of infinite integers and therefore can be extended by assuming that infinite integers do exist. There are different ways of using nonstandard analysis. Some emphasize logical concepts [235, 236]. But most applications, including those in probability [209, 6] and option pricing [72, 73, 74], rely on the representation of hyperreals within classical mathematics by means of the ultrapower construction ([136], Chapter 3), and we follow this tradition.

The nonstandard ideas that we use do not go beyond what is explained in this appendix, but readers who want a fuller account may consult Davis (1977), Hoskins (1990), and, especially, Goldblatt (1998).

Most often, nonstandard analysis is used to prove theorems that can be formulated in standard analysis. Here, in contrast, nonstandard analysis is essential to the statement of the theorems we prove. We use it in drawing as well as in studying our idealized picture of a continuous-time game. The meaning of this picture is to be found directly in the corresponding discrete-time game, not in some intermediate picture, involving continuous time but formulated in standard analysis. We do not know how to represent a continuous-time game in standard analysis, and it might even be impossible.

### The Ultrapower Construction of the Hyperreals

An *ultrafilter* in the set $\mathbb{N}$ of natural numbers (i.e., positive integers) is a family $\mathcal{U}$ of subsets of $\mathbb{N}$ such that

1. $\mathbb{N} \in \mathcal{U}$ and $\emptyset \notin \mathcal{U}$,
2. if $A \in \mathcal{U}$ and $A \subseteq B \subseteq \mathbb{N}$, then $B \in \mathcal{U}$,
3. if $A \in \mathcal{U}$ and $B \in \mathcal{U}$, then $A \cap B \in \mathcal{U}$, and
4. if $A \subseteq \mathbb{N}$, then either $A \in \mathcal{U}$ or $\mathbb{N} \setminus A \in \mathcal{U}$.

(The first three properties define a *filter*.) An ultrafilter $\mathcal{U}$ is *nontrivial* if it does not contain a set consisting of a single integer; this implies that all the sets in $\mathcal{U}$ are infinite. It follows from the axiom of choice that a nontrivial ultrafilter exists. We fix a nontrivial ultrafilter $\mathcal{U}$.

We say that a property of natural numbers holds for *most* natural numbers (or for most $k$, as we will say for brevity) if the set of natural numbers for which it holds is in $\mathcal{U}$; Condition 2 justifies this usage. It follows from Condition 4 in the definition of ultrafilter that for any property $A$, either $A$ holds for most $k$ or else the negation of $A$ holds for most $k$. It follows from Conditions 1 and 3 that $A$ and its negation cannot both hold for most $k$.

A *hyperreal number* $a$ is a sequence $[a^{(1)}a^{(2)}\dots]$ of real numbers. Sometimes we abbreviate $[a^{(1)}a^{(2)}\dots]$ to $[a^{(k)}]$. Operations (addition, multiplication, etc.) over hyperreals are defined term by term. For example,

$$\left[a^{(1)}a^{(2)}\dots\right]+\left[b^{(1)}b^{(2)}\dots\right]:=\left[\left(a^{(1)}+b^{(1)}\right)\left(a^{(2)}+b^{(2)}\right)\dots\right].$$

Relations (equals, greater than, etc.) are extended to the hyperreals by voting. For example, $[a^{(1)}a^{(2)}\dots]\le[b^{(1)}b^{(2)}\dots]$ if $a^{(k)}\le b^{(k)}$ for most $k$. For all $a,b\in{}^*\mathbb{R}$ one and only one of the following three possibilities holds: $a<b$, $a=b$, or $a>b$.

Perhaps we should dwell for a moment on the fact that a hyperreal number $a=[a^{(1)}a^{(2)}\dots]$ is always below, equal to, or above another hyperreal number $b=[b^{(1)}b^{(2)}\dots]$: $a<b$, $a=b$, or $a>b$. Obviously some of the $a^{(k)}$ can be above $b^{(k)}$, some equal to $b^{(k)}$, and some below $b^{(k)}$. But the set of $k$ satisfying one these three conditions is in $\mathcal{U}$ and outvotes the other two.

We do not distinguish hyperreals $a$ and $b$ such that $a=b$. Technically, this means that a hyperreal is an equivalence class of sequences rather than an individual sequence: $[a^{(1)}a^{(2)}\dots]$ is the equivalence class containing $a^{(1)}a^{(2)}\dots$.

For each $A\subseteq\mathbb{R}$ we denote by ${}^*A$ the set of all hyperreals $[a^{(k)}]$ with $a^{(k)}\in A$ for all $k$. We embed $A$ into ${}^*A$ by identifying each $a\in A$ with $[a,a,\dots]\in{}^*A$.

We say that $a\in{}^*\mathbb{R}$ is *infinitesimal* if $|a|<\epsilon$ for each real $\epsilon>0$. The only real number that qualifies as an infinitesimal by this definition is 0. We say that $a\in{}^*\mathbb{R}$ is *infinitely large* if $a>C$ for each positive integer $C$, and we say that $a\in{}^*\mathbb{R}$ is *finite* if $a<C$ for some positive integer $C$.

We write $a\approx b$ when $a-b$ is infinitesimal. For every hyperreal number $a\in{}^*\mathbb{R}$ there exists a unique standard number $\mathrm{st}(a)$ (its *standard part*) such that $a\approx b$.

The representation of the hyperreals as equivalence classes of sequences with respect to a nontrivial ultrafilter is constructive only in a relative sense, because the proof that a nontrivial ultrafilter exists is nonconstructive; no one knows how to exhibit one. However, the representation provides an intuition that helps us think about hyperreals. For example, an infinite positive integer is represented by a sequence of positive integers that increases without bound, such as $[1,2,4,\dots]$, and the faster it grows the larger it is.

Whereas there are canonical ways to construct rational numbers from integers and real numbers from rational numbers, the construction of hyperreals from reals depends on the arbitrary choice of a nontrivial ultrafilter. This has been one source of dissatisfaction with nonstandard analysis. However, the continuum hypothesis implies that the choice of the ultrafilter is irrelevant: if the continuum hypothesis is adopted as an axiom, then all systems of hyperreal numbers are isomorphic as ordered fields ([136], p. 33).

### Games and Strategies

The simplest notions of nonstandard analysis introduced in the previous subsection suffice to formalize our informal exposition in the bulk of the chapter. In this subsection we will explain how this formalization is done.

At the beginning of this chapter we fixed a positive real number $T$ and an infinitely large positive integer $N$; let

$$N = [N^{(k)}] = [N^{(1)}, N^{(2)}, \ldots].$$

For each natural number $k$, set

$$\mathbb{T}^{(k)} := \{nT/N^{(k)} \mid n = 0, 1, \ldots, N^{(k)}\}.$$

To each $k$ corresponds a "finitary framework" (which we will call the $k$-finitary framework), where the time interval is the finite set $\mathbb{T}^{(k)}$ rather than the infinite set $\mathbb{T}$. The "limit" (formally, ultraproduct) of these finitary frameworks will be the infinitary framework based on $\mathbb{T}$; as in the previous subsection, this "limit" is defined as follows:

- An object in the infinitary framework, such as strategy, should be defined as a family of finitary objects: for every $k$, an object in the $k$-finitary framework should be defined (cf. the definition of hyperreals in the previous subsection).

- Functionals defined on finitary objects are extended to infinitary objects termwise, analogously to the previous subsection. (By "functionals" we mean functions of objects of complex nature, such as paths or strategies.)

- Relations (in particular, properties) are defined by voting (again as in the previous subsection).

(In nonstandard analysis such limiting infinitary structures are called hyperfinite.)

This defines the procedure of "translation" of the informal statements into the formal nonstandard framework. Let us first consider some random examples of infinitary objects. The Bachelier game is a family of finitary games indexed by $k = 1, 2, \ldots$; finitary game $k$ is obtained from the protocol given in §11.2 by replacing $N$ with $N^{(k)}$. A *strategy* in the infinitary game is actually a set of strategies indexed by $k$ in the finitary games; it should specify, for any $k$ and $n = 0, \ldots, N^{(k)} - 1$, two functions, $\mathcal{M}^{(k)}(S_0, D_0, \ldots, S_n, D_n)$ and $\mathcal{V}^{(k)}(S_0, D_0, \ldots, S_n, D_n)$. A *path* (such as $S$ and $D$ in the theorems) in the infinitary game is a sequence, indexed by $k$, of finitary functions (i.e., functions defined on $\mathbb{T}^{(k)}$); for example, in the case of $S$ every finitary function $S^{(k)}$ is just declared by Market as the values $S_n^{(k)}$ in the $k$th protocol.

Now let us consider examples of infinitary functionals. Sum (11.1) is interpreted term by term: if $N = [N^{(k)}]$, then (11.1) is the hyperreal

$$\left( \sum_{n=0}^{N^{(k)}-1} \left| f^{(k)}\left(\frac{(n+1)T}{N^{(k)}}\right) - f^{(k)}\left(\frac{nT}{N^{(k)}}\right) \right|^c \right)_{k=1}^{\infty}.$$

Analogously, the sum in (11.7) is also interpreted term by term; namely, as the hyperreal

$$\left(\sum_{n=0}^{N^{(k)}-1}\left(\mathcal{M}^{(k)}(S_0,D_0,\ldots,S_n,D_n)\,(S_{n+1}-S_n)\right.\right.$$
$$\left.\left.+\,\mathcal{V}^{(k)}(S_0,D_0,\ldots,S_n,D_n)\left(D_{n+1}-D_n+\left(\frac{S_{n+1}-S_n}{S_n}\right)^2\right)\right)\right)_{k=1}^{\infty},$$

where, as usual, $S_n = S(nT/N^{(k)})$ and $D_n = D(nT/N^{(k)})$. The functionals in the definition of continuous and relatively continuous functions are of course understood term-wise, as the positive hyperreals

$$\sup_{t\in\mathbb{T}^{(k)}\setminus\{T\}} |f^{(k)}(t+dt) - f^{(k)}(t)|,\ k = 1,2,\ldots,$$

and

$$\sup_{t\in\mathbb{T}^{(k)}\setminus\{T\}} \left|\frac{f^{(k)}(t+dt) - f^{(k)}(t)}{f^{(k)}(t)}\right|,\ k = 1,2,\ldots,$$

respectively (of course, in the finitary frameworks sup can be replaced by max). A function is *continuous* or *relatively continuous*, respectively, if this positive hyperreal is infinitesimal.

And finally, examples of relations (actually, properties). The most important examples are, of course, Theorems 11.1 and 11.2 themselves. Theorem 11.1, for example, asserts that, for every $\epsilon$, there exists a strategy in the infinitary game replicating the European option's payoff with accuracy $\epsilon$. In the proof we actually constructed such a strategy: it is obvious how our construction works for every finitary game $k$; this family of finitary strategies indexed by $k$ determines the desired infinitary strategy (it will replicate the European option with accuracy $\epsilon$ for most $k$).

At this point it is very easy to demonstrate that the finitary propositions 10.2 and 10.3 of the previous chapter imply the theorems 11.1 and 11.2 of this chapter; this is a very special case of the general transfer principle briefly discussed above. Let us consider, for concreteness, the Bachelier formula. Suppose a play of the game satisfies $\operatorname{vex} S \le 2$ and $\operatorname{vex} D < 2$. For any real $\delta > 0$, (10.8) will be satisfied for $S = S^{(k)}$ and $D = D^{(k)}$ and most $k$ (take any $0 < \epsilon < 2 - \operatorname{vex} D$). Proposition 10.2 asserts the existence of a $k$-finitary strategy that reproduces the European option with accuracy $6c\delta^{1/4}$, which can be made arbitrarily small.

The above construction of the "infinitary structure" on $\mathbb{T}$ from the finitary structures on $\mathbb{T}^{(k)}$ is an informal version of the ultraproduct (see, e.g., [111]). The use of ultraproducts goes back to work by Gödel and Skolem in the 1930s. The first systematic study and the transfer principle for ultraproducts is due to the Polish (born in Lviv) mathematician Jerzy Łoś, in a paper published in 1955, before the invention of nonstandard analysis in 1960. Only this precursor of nonstandard analysis is really used in this book.

## 11.6 APPENDIX: ON THE DIFFUSION MODEL

From an intuitive point of view, it is obvious that the variation exponent, as we have defined it in this chapter, is identical with the 2-variation as it is usually defined in the theory of stochastic processes, and so it follows that paths of diffusion processes will have a variation exponent of two almost surely. Rigorous verification of this intuition requires some work, however, because of differences between the measure-theoretic and game-theoretic frameworks and between standard and nonstandard definitions.

In the first two subsections, we relate the definitions of $p$-variation and variation exponent that we used in this chapter to definitions used in the standard diffusion model—that is, the diffusion model as it is defined using standard rather than nonstandard analysis. In the third subsection, we relate our definitions to definitions that have been used for the nonstandard diffusion model.

### The $\sqrt{dt}$ Effect in the Standard Diffusion Model

The first three propositions of this subsection confirm that a path of the diffusion model will almost surely have a variation exponent equal to two. This is intuitively obvious because of the similarity between our definition of the 2-variation $\mathbf{var}_S(2)$ and the usual definition of optional quadratic variation $[S,S](T)$. But the proofs are not quite trivial, partly because of the mismatch between our nonstandard treatment of the game-theoretic approach and this subsection's standard treatment of the measure-theoretic approach. The final subsection of the appendix covers the same ground in a cleaner but more demanding way, by making the comparison between our nonstandard game-theoretic approach and a nonstandard version of the measure-theoretic approach.

**Proposition 11.3** *The path* $W\colon [0,T] \to \mathbb{R}$ *of a standard Wiener process satisfies* $\operatorname{vex} W = 2$ *almost surely. Moreover,* $\mathbf{var}_W(2) \approx T$ *almost surely.*

*Proof* Let $\epsilon > 0$ be arbitrarily small. For each $N = 1, 2, \ldots$, we have

$$\mathbb{P}\left\{\sum_{n=0}^{N-1}(W((n+1)T/N) - W(nT/N))^2 \geq T + \epsilon\right\} \leq e^{-Nc(\epsilon)}, \tag{11.16}$$

$$\mathbb{P}\left\{\sum_{n=0}^{N-1}(W((n+1)T/N) - W(nT/N))^2 \leq T - \epsilon\right\} \leq e^{-Nc(\epsilon)}, \tag{11.17}$$

where $c(\epsilon)$ is a positive constant. These inequalities follow from

$$\mathbb{E}(W((n+1)T/N) - W(nT/N))^2 = T/N$$

and the standard large-deviation results (such as [287], §IV.5, (11) and (12)). Combining (11.16) and (11.17) with the Borel-Cantelli lemma [287], we obtain that, with probability 1,

$$\left|\sum_{n=0}^{N-1}(W((n+1)T/N) - W(nT/N))^2 - T\right| \geq \epsilon$$

only for finitely many $N$; therefore, $|\,\mathbf{var}_W(2) - T| < \epsilon$ almost surely. Since $\epsilon$ can be taken arbitrarily small, $\mathbf{var}_W(2) \approx T$ almost surely. ∎

Proposition 11.3 is a nonstandard version of a result in Paul Lévy 1937.

The following lemma is obvious now:

**Proposition 11.4** *The path* $S\colon [0,T] \to \mathbb{R}$ *of the diffusion process governed by (9.2) satisfies* $\operatorname{vex} S = 2$ *almost surely.*

Now we extend this result to the (standard) Black-Scholes model.

**Proposition 11.5** *The path* $S\colon [0,T] \to \mathbb{R}$ *of the diffusion process governed by (9.4) satisfies* $\operatorname{vex} S = 2$ *almost surely.*

*Proof* Since

$$S(t) = S(0)e^{(\mu-\sigma^2/2)t+\sigma W(t)} \tag{11.18}$$

for a standard Wiener process $W$, we have:

$$dS(t) = S(t)e^{(\mu-\sigma^2/2)dt+\sigma dW(t)} - S(t)$$
$$= S(t)e^{\theta(t)((\mu-\sigma^2/2)dt+\sigma dW(t))}((\mu - \sigma^2/2)dt + \sigma dW(t)) \asymp (\mu - \sigma^2/2)dt + \sigma dW(t),$$

where $\theta(t) \in (0,1)$ and $a(t) \asymp b(t)$ means that $|a(t)| \le C|b(t)|$ and $|b(t)| \le C|a(t)|$ for some positive constant $C$ (maybe, dependent on the path $S(t)$). Therefore,

$$\mathbf{var}_S(2) \asymp \sum_t ((\mu - \sigma^2/2)dt + \sigma dW(t))^2$$
$$= (\mu - \sigma^2/2)^2 \sum_t (dt)^2 + 2(\mu - \sigma^2/2)\sigma \sum_t dtdW(t) + \sigma^2 \sum_t (dW(t))^2$$
$$\approx \sigma^2\, \mathbf{var}_W(2),$$

$t$ ranging over $\mathbb{T} \setminus \{T\}$, and Lemma 11.3 on p. 287 shows that, with probability one, $\mathbf{var}_S(2)$ is neither infinitesimal nor infinitely large. ∎

Generalizing Propositions 11.4 and 11.5, it is natural to expect that almost all paths of a regular continuous semimartingale $S$ will satisfy $\operatorname{vex} S \in \{1,2\}$ ($\operatorname{vex} S = 1$ corresponding to a trivial martingale component) and

$$\mathbf{var}_S(2) = [S,S](T) = \langle S,S\rangle(T)$$

(cf. [158], Theorems I.4.47 and I.4.52, and Lepingle's result, discussed on p. 290).

The next two propositions assert the uselessness of the derivative security $\mathcal{D}$ in the Bachelier and Black-Scholes models, respectively.

**Proposition 11.6 (Bachelier derivative)** *Let the derivative security* $\mathcal{D}$ *pay a dividend of* $(dS(t))^2$ *at the end of each interval* $[t, t + dt]$, $t \in \mathbb{T} \setminus \{T\}$, *where* $S(t)$ *is governed by (9.2). With probability 1, the total amount of dividends paid by* $\mathcal{D}$ *during every time interval* $[t_1, t_2]$, $0 \le t_1 < t_2 \le T$, *is infinitely close to* $\sigma^2(t_2 - t_1)$.

*Proof* First we consider fixed $t_1$ and $t_2$. We can assume that the dividend paid by $\mathcal{D}$ at the end of each $[t, t + dt]$ is $(\sigma dW(t))^2$, where $W(t)$ is a standard Wiener process. Analogously to (11.16) and (11.17), we have, for any $\epsilon > 0$ and from some $N$ on,

$$\mathbb{P}\left\{\sum_{n=\lceil Nt_1/T\rceil-1}^{\lfloor Nt_2/T\rfloor-1}(\sigma W((n+1)T/N)-\sigma W(nT/N))^2 \geq \sigma^2(t_2-t_1)+\epsilon\right\} \leq e^{-Nc(\epsilon)},$$

$$\mathbb{P}\left\{\sum_{n=\lceil Nt_1/T\rceil-1}^{\lfloor Nt_2/T\rfloor-1}(\sigma W((n+1)T/N)-\sigma W(nT/N))^2 \leq \sigma^2(t_2-t_1)-\epsilon\right\} \leq e^{-Nc(\epsilon)}.$$

Again applying the Borel-Cantelli lemma, we obtain that the total amount of dividends paid during $[t_1, t_2]$ is $\sigma^2(t_2 - t_1)$ to within $\epsilon$, almost surely; since $\epsilon$ can be arbitrarily small, it is infinitely close to $\sigma^2(t_2 - t_1)$, almost surely.

Now it is clear that for almost all paths $S(t)$ the total amount of dividends paid by $\mathcal{D}$ during $[t_1, t_2]$ is infinitely close to $\sigma^2(t_2 - t_1)$ for all rational $t_1, t_2 \in [0, T]$, $t_1 < t_2$. Consider a path $S(t)$ satisfying this property; let $t_1, t_2 \in [0, T]$, $t_1 < t_2$, be not necessarily rational. Since $t_1$ and $t_2$ can be arbitrarily accurately approximated from below and from above by rational numbers, the total amount of dividends paid by $\mathcal{D}$ during $[t_1, t_2]$ is infinitely close to $\sigma^2(t_2 - t_1)$. ∎

**Proposition 11.7 (Black-Scholes derivative)** *Let the derivative security $\mathcal{D}$ pay a dividend of $(dS(t)/S(t))^2$ at the end of each interval $[t, t + dt]$, $t \in \mathbb{T} \setminus \{T\}$, where $S(t)$ is governed by (9.4). With probability 1, the total amount of dividends paid by $\mathcal{D}$ during every time interval $[t_1, t_2]$, $0 \leq t_1 < t_2 \leq T$, is infinitely close to $\sigma^2(t_2 - t_1)$.*

*Proof* First we consider fixed $t_1$ and $t_2$. It is easy to see that we can assume that the dividend paid by $\mathcal{D}$ at the end of each $[t, t + dt]$ is $(\sigma dW(t))^2$ rather than $(dS(t)/S(t))^2$, where $W(t)$ is the standard Wiener process satisfying (11.18). This puts us in the same position as in the proof of Proposition 11.6, and we can use the same argument. ∎

Propositions 11.6 and 11.7 immediately imply that in the case where the derivative security $\mathcal{D}$ is traded in an arbitrage-free market its price $D(t)$ (as usual in probability theory, we assume that $D(t)$ is continuous from the right) satisfies $D(t) = \sigma^2(T - t)$ almost surely and, therefore, $\operatorname{vex} D = 1$ almost surely.

## Strong $p$-variation

We now turn to an alternative approach to the variation exponent that is popular in the measure-theoretic literature. Let $f$ be a (standard) real-valued function defined on the interval $[0, T]$. Its *strong p-variation*, for $p > 0$, is

$$\overline{\mathbf{var}}_f(p) := \sup_{\kappa}\sum_{i=1}^{n}|f(t_i) - f(t_{i-1})|^p,$$

where $n$ ranges over all positive integers and $\kappa$ over all subdivisions $0 = t_0 < t_1 < \cdots < t_n = T$ of the interval $[0, T]$. For any function $f$ there exists a unique number $\overline{\operatorname{vex}}\, f$ such that $\overline{\mathbf{var}}_f(p)$ is finite when $p > \overline{\operatorname{vex}}\, f$ and infinite when $p < \overline{\operatorname{vex}}\, f$. This is true for all functions $f$, although $\overline{\operatorname{vex}}\, f$ can be finite only for functions that are *regulated*—that is, have left and right limits everywhere. We are mainly interested in the case where $f$ is continuous and not constant.

**Lemma 11.4** *If* $f : [0,T] \to \mathbb{R}$ *is continuous and not constant,* $\operatorname{vex} f \leq \overline{\operatorname{vex}}\, f$.

*Proof* For any $p \geq 1$ and any continuous $f : [0,T] \to \mathbb{R}$, $\mathbf{var}_f(p) = \infty$ implies $\overline{\mathbf{var}}_f(p) = \infty$; it is also obvious that $\overline{\mathbf{var}}_f(p) = \infty$ when $p < 1$ since $f$ is not a constant. ∎

Lepingle 1976 showed that $\overline{\operatorname{vex}}\, f \leq 2$ for almost all paths $f$ of a semimartingale; therefore, $\operatorname{vex} f \leq 2$. This shows that the processes with $\operatorname{vex} f > 2$ are beyond the reach of the general theory of stochastic processes, which is based on the notion of a semimartingale.

The analog for the result of the previous subsection that $\operatorname{vex} W = 2$ for almost all paths of the Brownian motion is that $\overline{\operatorname{vex}}\, W = 2$ for almost all paths of the Brownian motion (Lévy 1940). For the boundary value $p = 2$, Lévy proved that $\overline{\mathbf{var}}_W(p) = \infty$ almost surely; moreover,

$$\lim_{\delta \to 0} \sup_{\kappa \in K_\delta} \sum_{i=1}^{n_\kappa} \left(W(t_i) - W(t_{i-1})\right)^2$$

is infinite almost surely, where $K_\delta$ is the set of all finite partitions $\kappa = \{0 = t_0 < \cdots < t_{n_\kappa} = T\}$ whose mesh is less than $\delta$: $\max |t_i - t_{i-1}| < \delta$. This result found a beautiful continuation in Taylor (1972):

$$\lim_{\delta \to 0} \sup_{\kappa \in K_\delta} \sum_{i=1}^{n_\kappa} \psi\left(|W(t_i) - W(t_{i-1})|\right) = 1 \quad \text{almost surely}$$

for

$$\psi(t) = \frac{t^2}{2 \ln \ln \frac{1}{t}}.$$

Recall that a *fractional Brownian motion* with index $h \in (0,1)$ is a continuous zero-mean Gaussian process $B_h(t)$ whose increments $B_h(t) - B_h(s)$, where $0 \leq s \leq t$, have variance $(t-s)^{2h}$. As shown by Kawada and Kôno [167], $\overline{\operatorname{vex}}\, B_h \leq 1/h$ almost surely, with equality when $h \leq 0.5$. This implies that $\operatorname{vex} B_h \leq 1/h$ almost surely. Comparing this result with our requirement (implicit in Theorem 11.2) $\operatorname{vex} S \leq 2$, we see that fractional Brownian motions with $h < 0.5$ may be too irregular even for our methods for the Bachelier and Black-Scholes formulas. See also [132, 167].

### The $\sqrt{dt}$ Effect in the Nonstandard Diffusion Model

The comparison of the game-theoretic and measure-theoretic approaches to the variation exponent becomes transparent if we do it entirely in terms of nonstandard analysis.

Before we can proceed, we need to discuss the key notion of *internal* objects in nonstandard analysis. Intuitively, internal sets, functions, measures, and so on, can be defined in terms of sequences of standard objects of the same nature; this makes them tame, unlike the unmanageable *external* objects (i.e., objects that are not internal). For example, an internal subset $Y$ of a standard set $Z$ can be defined to be a sequence $[Y^{(k)}]$ of subsets $Y^{(k)} \subseteq Z$; a nonstandard object $[z^{(k)}] \in {}^*Z$ is in $Y$ if and only if $z^{(k)} \in Y^{(k)}$ for most $k$. Similarly, an internal function $f$ on $Z$ is a sequence $[f^{(k)}]$ of standard functions $f^{(k)} : Z \to \mathbb{R}$; internal functions are applied to elements of ${}^*Z$ term by term. We can also define internal objects for nonstandard domains such as $\mathbb{T}$; for example, an internal function $f$ on $\mathbb{T}$ is a sequence $[f^{(k)}]$ of functions $f^{(k)} : \mathbb{T}^{(k)} \to \mathbb{R}$.

The nonstandard treatment of Brownian motion (Anderson 1976) uses an important property of internal sets called *countable saturation*. This property is the key element in constructing the Loeb measure on the set of internal functions on $\mathbb{T}$; the next theorem will be applied to internal subsets of $\mathbb{T}$, but it is true in great generality and we denote by $X$ the set whose internal subsets are considered.

**Saturation Theorem** *The following property of countable saturation holds: the intersection of a decreasing sequence*

$$A_1 \supseteq A_2 \supseteq \cdots$$

*of nonempty internal sets in $X$ is always nonempty.*

In this theorem, the relation $[A^{(k)}] \supseteq [B^{(k)}]$ means, of course, that $A^{(k)} \supseteq B^{(k)}$ for most of $k$. For its proof (just a simple diagonalization argument) see, for example, Goldblatt [136], p. 138.

In 1975 Peter Loeb introduced what is now known as the Loeb measure. Let $\nu$ be a finitely additive nonstandard (which just means "taking values in ${}^*\mathbb{R}^+$", where $\mathbb{R}^+$ is the set of nonnegative reals) measure on an algebra of internal subsets of $X$. Let $\operatorname{st}\nu$ be the composition of st and $\nu$ (i.e., $(\operatorname{st}\nu)(A)$ is defined as $\operatorname{st}(\nu(A))$); because of the saturation theorem, $\operatorname{st}\nu$ (as well as $\nu$) will automatically be $\sigma$-additive (and, in general, taking values in $[0, \infty]$ rather than $[0, \infty)$). This allows one to apply Carathéodory's extension theorem (see, e.g., [347] or [25]) to define a $\sigma$-additive extension $L_\nu$ of $\operatorname{st}\nu$ to the smallest $\sigma$-algebra containing the initial algebra on which $\nu$ is defined (such an extension is unique if $\nu$ only takes finite values; this will be the case in our applications). This extension $L_\nu$ is called the Loeb measure on $X$.

The Loeb measure was used by Anderson (1976) to define a measure, analogous to the Wiener measure, on the set of internal functions on $\mathbb{T}$. First we introduce some specific *internal measure* $\nu = [\nu^{(k)}]$ on the internal functions on $\mathbb{T}$: for any $k$, $\nu^{(k)}$ is the uniform distribution on the set

$$\left\{ f_{j_0, j_1, \ldots, j_{N^{(k)}}} \mid j_0, j_1, \ldots, j_{N^{(k)}} \in \{-1, 1\} \right\} \tag{11.19}$$

of all functions defined by

$$f_{j_1, \ldots, j_{N^{(k)}}}(t) := \sqrt{\frac{T}{N^{(k)}}} \sum_{n=0}^{\lfloor N^{(k)} t/T \rfloor} j_n;$$

for any internal set $A = [A^{(k)}]$, where each $A^{(k)}$ is a subset of (11.19), $\nu(A)$ is defined to be the hyperreal $[\nu^{(k)}(A^{(k)})]$. It is obvious that $\nu$ is finitely additive. Let $L_\nu$ be Carathéodory's extension of $\operatorname{st}\nu$ (Loeb measure). The image $P$ of $L_\nu$ under the mapping $f \mapsto \operatorname{st} f$ is a measure (actually a probability distribution) on the set of all functions of the type $[0, T] \to [-\infty, \infty]$.

This is Anderson's construction of the standard (i.e., variance $T$) Brownian motion on $[0, T]$. He proves the following results about his construction ([6], Theorems 26 and 27, Corollary 28):

- $P$-almost all functions $f : [0, T] \to [-\infty, \infty]$ are continuous and finite.
- The restriction of $P$ to the continuous functions $f : [0, T] \to \mathbb{R}$ (equipped with the usual $\sigma$-algebra) is the standard Wiener measure.

Propositions 11.3–11.7 are trivial under this construction. Under Cutland's variant [71], where the jumps $\pm 1$ are replaced by Gaussian jumps, they are less trivial but still can be proven very easily.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 12

# *The Generality of Game-Theoretic Pricing*

Robert C. Merton (born 1944). He was the first to understand and exploit fully the game-theoretic nature of stochastic Black-Scholes pricing.

In this chapter, we demonstrate the scope and flexibility of our purely game-theoretic approach to option pricing by extending it in several directions. We show how to handle interest rates. We show that the dividend-paying derivative that we have used for Black-Scholes pricing can be replaced with a more conventional derivative. And we show how to deal with the case where there can be jumps in the price process for the underlying security.

Interest rates are in the picture from the outset in most expositions of stochastic Black-Scholes pricing—including the original exposition by Black and Scholes in 1973. And aside from a minor adjustment in the dividend paid by our security $\mathcal{D}$, there is no conceptual difference between the way they enter the stochastic picture and the way they enter our picture. We have chosen to leave them aside until now only in order to make our exposition as simple and clear as possible. In §12.1 we correct the omission.

In §12.2, we show that the dividend-paying security $\mathcal{D}$ we use in game-theoretic Black-Scholes pricing can be replaced with a European option $\mathcal{R}$, which pays off only at its maturity $T$. Many different European options might play the role of $\mathcal{R}$: the main requirement is that $\mathcal{R}$'s payoff at $T$ be a strictly convex function of the price at $T$ of the underlying security $\mathcal{S}$. The simplest choice is to have $\mathcal{R}$ pay the square of $\mathcal{S}$'s price at $T$. In any case, we can infer from the current price of $\mathcal{R}$ an implied price for our dividend-paying security $\mathcal{D}$, and this suffices for purely game-theoretic pricing of other derivatives with maturity $T$. Unfortunately, calls and puts, which now constitute the bulk of options that are traded, are not strictly convex and can hardly serve our purpose. This suggests a market reform that might be both feasible and useful: the market might be more liquid and efficient if calls and puts were replaced by a strictly convex derivative.

We deal with jumps in §12.3. The established stochastic theory uses the Poisson process to model jumps, just as it uses geometric Brownian motion to model continuous price processes. Our game-theoretic analog involves marketing an instrument that pays dividends when market crashes occur—an insurance policy similar to the weather derivatives used to insure against extreme storms. Our treatment of jumps, like the usual stochastic treatment, is in continuous time. This demonstrates the generality of our mathematical framework for continuous time, but it also means, from a practical point of view, that our results are only preliminary. Putting them into practice will require discrete-time analyses along the lines sketched in Chapter 10 for the game-theoretic Bachelier and Black-Scholes protocols.

In an appendix, §12.4, we review the theory of stable and infinitely divisible probability distributions, which is related to the stochastic pricing of discontinuous processes and contains ideas that could lead to additional game-theoretic methods.

## 12.1 THE BLACK-SCHOLES FORMULA WITH INTEREST

In this section, we explain how interest rates can be taken into account in game-theoretic Black-Scholes pricing. We rely on an argument that has also been used in stochastic option-pricing theory:

- our reasoning ignoring interest rates is correct if all prices are measured relative to a risk-free bond, and

- formulas in which prices are expressed in dollars, say, can be derived from the formulas in which the prices are expressed relative to the risk-free bond.

One point does arise, however, that does not arise in the stochastic theory: we need to adjust the dividend paid by $\mathcal{D}$ at the end of each period to take account of the bond's change in value during the period.

### What Difference Does Interest Make?

How are interest rates relevant to the game-theoretic hedging problem? Indeed, how have we used the assumption that the interest rate is zero?

When we look at our protocols for Black-Scholes pricing, we see that the need to consider interest enters only when we consider the rule for updating Investor's capital. Consider our protocol for purely game-theoretic Black-Scholes pricing in discrete time, on p. 249 of §10.3. The rule for updating Investor's capital that appears there can be written in the form

$$\begin{aligned}\mathcal{I}_n := (\mathcal{I}_{n-1} - M_n S_{n-1} - V_n D_{n-1})\\ + (M_n S_n + V_n D_n) + V_n(\Delta S_n / S_{n-1})^2.\end{aligned} \tag{12.1}$$

The first term on the right-hand side, $\mathcal{I}_{n-1} - M_n S_{n-1} - V_n D_{n-1}$, is Investor's net cash position during the $n$th round of trading. If interest is paid on money during the $n$th round at nonzero rate $r_n$, then this term should be multiplied by $1 + r_n$, in order to account for the interest that Investor pays (if his net cash position is negative) or receives (if it is positive). The second term, $M_n S_n + V_n D_n$, is the net cash Investor receives or pays when he liquidates his position at the end of the round; this is not affected by the interest rate, except insofar as the Market might take it into account in setting the prices $S_n$ and $D_n$. The final term, $V_n(\Delta S_n/S_{n-1})^2$, represents Investor's dividends from holding $V_n$ shares of the security $\mathcal{D}$. This is also unaffected by the interest rate, but as we will see shortly, we need to change the dividend $(\Delta S_n/S_{n-1})^2$ in order to replicate the option $\mathcal{U}$.

## Using a Risk-Free Bond as *Numéraire*

Our first step is to recognize that all our reasoning concerning the game-theoretic Black-Scholes protocol, in both the discrete and continuous cases, is valid if all prices are measured relative to a risk-free bond $\mathcal{B}$ which is priced by the market and whose changes in value define the interest rate.

Let us assume that a unit of $\mathcal{B}$ pays \$1 at maturity (the end of round $N$) and costs $B_n$ at the end of round $n$. Let us also assume that the interest rate for Investor is defined by the changes in $\mathcal{B}$'s value:

$$r_n := \frac{B_n - B_{n-1}}{B_{n-1}} = \frac{\Delta B_n}{B_{n-1}}.$$

Investor begins with capital zero. We assume that he pays interest at the rate $r_n$ if he needs to borrow money during period $n$ and that he draws interest at this rate on any spare cash that he does not use during period $n$. In theory (i.e., neglecting transaction costs), he can do this by investing all spare cash in $\mathcal{B}$ and borrowing any money that he needs by going short in $\mathcal{B}$. We can then say that his strategy involves trading in three securities: the underlying security $\mathcal{S}$, the derivative security $\mathcal{D}$, and the risk-free bond $\mathcal{B}$.

When we make the role of $\mathcal{B}$ explicit in the protocol for discrete Black-Scholes pricing on p. 249, we obtain the following:

A DISCRETE BLACK-SCHOLES PROTOCOL WITH A RISK-FREE BOND
**Parameters:** $N, \mathcal{I}_0 > 0, \delta \in (0,1), C > 0$

**Players:** Market, Investor
**Protocol:**
Market announces $B_0 > 0$, $S_0 > 0$, and $D_0 > 0$.
FOR $n = 1, \ldots, N$:
Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Market announces $B_n > 0$, $S_n \in \mathbb{R}$, and $D_n \geq 0$.

$$\mathcal{I}_n := (\mathcal{I}_{n-1} - M_n S_{n-1} - V_n D_{n-1})\tfrac{B_n}{B_{n-1}} + (M_n S_n + V_n D_n) + V_n \tfrac{B_{n-1}^2}{B_n}\left(\tfrac{\Delta S_n}{S_{n-1}} - \tfrac{\Delta B_n}{B_{n-1}}\right)^2. \tag{12.2}$$

**Additional Constraints on Market:** Market must satisfy the constraints on $S_n$ and $D_n$ stated in the protocol on p. 249 and must also satisfy $B_N = 1$.

The first line of the rule for updating Investor's capital $\mathcal{I}_n$ is different from the first line in (12.1) because of the interest paid; we now multiply Investor's net cash position by $B_n/B_{n-1}$, or $1 + r_n$, to reflect the payment of interest. The third line is different because we are changing the dividend paid by the security $\mathcal{D}$. We now have $\mathcal{D}$ pay the dividend

$$\frac{B_{n-1}^2}{B_n}\left(\frac{\Delta S_n}{S_{n-1}} - \frac{\Delta B_n}{B_{n-1}}\right)^2 \tag{12.3}$$

at the end of period $n$, instead of $(\Delta S_n/S_{n-1})^2$. Typically (12.3) will be approximately equal to $B_n(\Delta S_n/S_{n-1} - r_n)^2$.

Let us now re-express (12.2), the rule for updating Investor's capital, with the capital and prices measured relative to $B_n$, the current price of the bond. We use a dagger for quantities measured in this *numéraire*:

$$\mathcal{I}_i^\dagger = \frac{\mathcal{I}_i}{B_i}, S_i^\dagger = \frac{S_i}{B_i}, \text{ and } D_i^\dagger = \frac{D_i}{B_i}.$$

Substituting $B_i\mathcal{I}_i^\dagger$ for $\mathcal{I}_i$, $B_i S_i^\dagger$ for $S_i$ and $B_i D_i^\dagger$ for $D_i$ in (12.2) and simplifying, we obtain

$$\mathcal{I}_n^\dagger = (\mathcal{I}_{n-1}^\dagger - M_n S_{n-1}^\dagger - V_n D_{n-1}^\dagger) + (M_n S_n^\dagger + V_n D_n^\dagger) + V_n(\Delta S_n^\dagger / S_{n-1}^\dagger)^2,$$

which is exactly the same as (12.1) except for the presence of the daggers. In other words, our new protocol reduces to Chapter 10's protocol for Black-Scholes when we measure prices and capital relative to the bond. The dividend (12.3) is chosen, of course, to make this happen.

Because $B_N = 1$, the shift from pricing in dollars to pricing relative to the risk-free bond makes no difference in how we describe a European option $\mathcal{U}$ with maturity at the end of period $N$; its payoff function $U$ gives its price relative to $\mathcal{B}$, which is the same as its price in dollars, and this price can be described as $U(S(T))$ or $U(S^\dagger(T))$; the two are the same. So Theorem 10.3, applied to the protocol in which prices are measured relative to $\mathcal{B}$, tells us that the initial price of $\mathcal{U}$ relative to $\mathcal{B}$ is approximately

$$\mathcal{U}_0^\dagger = \int_{\mathbb{R}} U\left(S_0^\dagger e^z\right) \mathcal{N}_{-D_0^\dagger/2, D_0^\dagger}(dz).$$

To obtain the initial price of $\mathcal{U}$ in dollars, $\mathcal{U}_0$, we multiply this expression by $B_0$:

$$\mathcal{U}_0 = B_0 \int_{\mathbb{R}} U\left(S_0^{\dagger} e^z\right) \mathcal{N}_{-D_0^{\dagger}/2,\, D_0^{\dagger}}(dz),$$

or

$$\mathcal{U}_0 = B_0 \int_{\mathbb{R}} U\left(\frac{S_0}{B_0} e^z\right) \mathcal{N}_{-D_0/(2B_0),\, D_0/B_0}(dz). \tag{12.4}$$

## Dollar Pricing in Discrete Time

We can simplify the expression (12.4) slightly by rescaling the derivative security $\mathcal{D}$. Indeed, if we divide the dividend (12.3) by $B_0$, then Market will divide the price of $\mathcal{D}$ at all times by $B_0$, thereby eliminating the need to introduce this divisor in the mean and variance of the Gaussian distribution in (12.4). So we have the following result, which we state informally:

**Black-Scholes Formula with Interest; Discrete Time** *Suppose Market prices a security $\mathcal{S}$, a bond $\mathcal{B}$ that pays \$1 at maturity $N$, and a derivative security $\mathcal{D}$ that pays the dividend*

$$\$\frac{B_{n-1}^2}{B_0 B_n}\left(\frac{\Delta S_n}{S_{n-1}} - \frac{\Delta B_n}{B_{n-1}}\right)^2 \tag{12.5}$$

*at the end of period $n$ for $n = 1, \ldots, N$. And suppose $\mathcal{U}$ is a European option that pays $\$U(S_N)$ at the end of period $N$. Then after Market has announced the prices $\$S_0$, $\$B_0$, and $\$D_0$, Investor can approximately replicate $\mathcal{U}$ starting with the initial capital*

$$\$B_0 \int_{\mathbb{R}} U\left(\frac{S_0}{B_0} e^z\right) \mathcal{N}_{-D_0/2,\, D_0}(dz).$$

## Dollar Pricing in Continuous Time

When we pass to continuous time, $t = (n-1)dt$, the expression (12.5) becomes

$$\frac{B^2(t)}{B(0)B(t+dt)}\left(\frac{dS(t)}{S(t)} - r(t)dt\right)^2 \approx \frac{B(t)}{B(0)}\left(\frac{dS(t)}{S(t)} - r(t)dt\right)^2,$$

where $r(t)dt$ corresponds to the discrete-time $r_n$ (since $dt$ is now infinitesimal, we prefer to deal with the "annualized" rate of return $r(t)$). So our informal result for continuous time reads as follows:

**Black-Scholes Formula with Interest; Continuous Time** *Suppose Market prices a security $\mathcal{S}$, a bond $\mathcal{B}$ that pays \$1 at maturity $T$, and a derivative security $\mathcal{D}$ that pays the continuous dividend*

$$\$\frac{B(t)}{B(0)}\left(\frac{dS(t)}{S(t)} - r(t)dt\right)^2.$$

*And suppose $\mathcal{U}$ is a European option that pays $\$U(S(T))$ at maturity $T$. Then after Market has announced the prices $\$S(0)$, $\$B(0)$, and $\$D(0)$, Investor can approximately replicate $\mathcal{U}$ starting with the initial capital*

$$\$B(0) \int_{\mathbb{R}} U\left(\frac{S(0)}{B(0)} e^{z}\right) \mathcal{N}_{-D(0)/2,\, D(0)}(dz).$$

When the interest rate is constant, say equal to $r$, $B(t) = e^{-r(T-t)}$. In this case $\mathcal{D}$'s continuous dividend is

$$\$e^{rt}\left(\frac{dS(t)}{S(t)} - r dt\right)^2,$$

and the Black-Scholes price for $\mathcal{U}$ at time 0 is

$$\$e^{-rT} \int_{\mathbb{R}} U\left(S(0)e^{z}\right) \mathcal{N}_{rT-D(0)/2,\, D(0)}(dz).$$

## 12.2 BETTER INSTRUMENTS FOR BLACK-SCHOLES

Although the dividend-paying derivative $\mathcal{D}$ plays a fundamental theoretical role in game-theoretic Black-Scholes pricing, it has potentially fatal practical disadvantages. In this section, we show how it can be replaced in practice with a European option $\mathcal{R}$. The replacement proceeds in a way that is familiar from stochastic option-pricing theory. Assuming that the market prices $\mathcal{R}$, we find the price for $\mathcal{D}$ implied by $\mathcal{R}$'s price. Then we price any other option $\mathcal{U}$ as if this price for $\mathcal{D}$ had been determined directly by the market, and we hedge the price of $\mathcal{U}$ by trading in $\mathcal{S}$ and $\mathcal{R}$.

By pricing $\mathcal{R}$ with a range of different maturities, the market would indirectly determine prices for $\mathcal{D}$ for a range of maturities, thereby allowing us to price all European options. In the current situation, where puts and calls are the options most often directly priced from the market, similar efforts are often made to price other options from these puts and calls, but these efforts lack the theoretical grounding of our proposal, and they are also less practical, because a doubly indexed array of puts and calls is required for effective pricing; we must have prices for puts and calls not only for a range of maturities but also for a range of strike prices. Our proposal, if implemented, would concentrate market activity on a single instrument with a range of maturities, thereby producing more reliable prices for this instrument and a more efficient and liquid market overall.

### Difficulties with the Dividend-Paying Derivative

Before explaining our replacement for the dividend-paying derivative $\mathcal{D}$, let us pause to note the nature of the difficulties that would be involved in trading $\mathcal{D}$.

The derivative $\mathcal{D}$ is supposed to pay the dividend $(dS(t)/S(t))^2$ at the end of the period from $t$ to $t + dt$. This raises several questions. Most importantly:

1. What should $dt$ be? One day? One hour? One week?

2. How is the price of $\mathcal{S}$ at the end of each period defined? By bids/offers or by actual trades?

Possible answers to the second question raise further questions about possible biases and, most importantly, about the potential for manipulation by brokers and large traders. If $S(t)$ is defined in terms of actual trades, the dividends paid by $\mathcal{D}$ might be excessively high, because sell orders are executed at a price lower (by the margin) than buy orders. And anyone with discretion over the timing of transactions might affect the dividend $(dS(t)/S(t))^2$ by pushing a transaction to one side or the other of the time boundary $d + dt$.

In addition to these issues, there are also practical problems in crediting dividends for units of $\mathcal{D}$ that change hands rapidly. It is conceivable that all these problems can be resolved, but the results of this section show that they do not really need to be addressed.

### Black-Scholes with the Square

Consider a European option $\mathcal{R}$ with payoff function $R : \mathbb{R} \to \mathbb{R}$; $\mathcal{R}$ pays $R(S(T))$ at maturity $T$. If our variance derivative $\mathcal{D}$ were traded by the market, then the price $\mathcal{R}$ at time $t$ would be given by our game-theoretic Black-Scholes formula:

$$\mathcal{R}(t) = \int_{\mathbb{R}} R(S(t)e^z)\ \mathcal{N}_{-D(t)/2,\,D(t)}(dz). \tag{12.6}$$

If $\mathcal{D}$ is not actually priced by the market, but $\mathcal{R}$ is, and if the function $R$ is well-enough behaved that we can solve (12.6) for $D(t)$, then we call the solution $D(t)$ the *implied remaining variance* of $\mathcal{S}$ at time $t$. We will emphasize the example where $\mathcal{R}$ is the square of $\mathcal{S}$: $R(s) := s^2$. In this case, $\mathcal{R}$'s price is

$$\mathcal{R}(t) = \int_{\mathbb{R}} (S(t)e^z)^2\ \mathcal{N}_{-D(t)/2,\,D(t)}(dz) = S^2(t)e^{D(t)}, \tag{12.7}$$

and this can be solved analytically; the implied remaining volatility at time $t$ is

$$D(t) = \ln \frac{\mathcal{R}(t)}{S^2(t)}. \tag{12.8}$$

As we will show, $D(t)$ can then be used to price another European option $\mathcal{U}$ with payoff function $U$ using our usual game-theoretic Black-Scholes formula:

$$\mathcal{U}(t) = \int_{\mathbb{R}} U(S(t)e^z)\ \mathcal{N}_{-D(t)/2,\,D(t)}(dz). \tag{12.9}$$

This works under the usual constraints on the paths $S(t)$ and $D(t)$.

This price for $\mathcal{U}$ can be hedged by trading in $\mathcal{S}$ and $\mathcal{R}$. From time $t$ to $t + dt$, we hold

$$\frac{\partial \overline{U}}{\partial s}(S(t), D(t)) - \frac{2}{S(t)}\frac{\partial \overline{U}}{\partial D}(S(t), D(t)) \tag{12.10}$$

units of $\mathcal{S}$ and

$$\frac{1}{\mathcal{R}(t)}\frac{\partial \overline{U}}{\partial D}(S(t), D(t)) \tag{12.11}$$

units of $\mathcal{R}$, where, as usual,

$$\overline{U}(s, D) := \int_{\mathbb{R}} U(se^z)\ \mathcal{N}_{-D/2,\,D}(dz).$$

This recipe for pricing and hedging a security $\mathcal{U}$ is similar to but distinct from recipes that have emerged from models for stochastic volatility; see p. 229.

The following protocol formalizes the assumption that Investor can trade in both $\mathcal{S}$ and $\mathcal{R}$:

THE BLACK-SCHOLES PROTOCOL FOR THE SQUARE
**Players:** Investor, Market
**Protocol:**

$\mathcal{I}_0 := 0$.
Market announces $S_0 > 0$ and $\mathcal{R}_0 > 0$.
FOR $n = 1, 2, \ldots, N$:
  Investor announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
  Market announces $S_n > 0$ and $\mathcal{R}_n \geq 0$.
  $\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n \Delta \mathcal{R}_n$.

**Additional Constraints on Market:** Market must ensure that $S$ is continuous, $\inf_n S_n$ is positive and not infinitesimal, and $\sup_n S_n$ is finite. He must also ensure that the process $D_n$ defined by $D_n = \ln(\mathcal{R}_n / S_n^2)$ is continuous and satisfies $D_n > 0$ for $n = 0, 1, \ldots, N-1$, $D_N = 0$, $\sup_n D_n$ is finite, and $\operatorname{vex} D < 2$.

**Theorem 12.1** *Suppose $U\colon \mathbb{R} \to \mathbb{R}$ is a smooth European option with all derivatives (including $U$ itself) bounded. Then the price for $U(S(T))$ in the Black-Scholes protocol for the square just after $S(0)$ and $\mathcal{R}(0)$ are announced is*

$$\int_{\mathbb{R}} U(S(0)e^z)\ \mathcal{N}_{-D(0)/2,\,D(0)}(dz), \tag{12.12}$$

*where $D(0) := \ln(\mathcal{R}(0)/S^2(0))$.*

*Proof (Sketch Only)* It suffices to show that if Investor starts with initial capital (12.12) and follows the strategy given by (12.10) and (12.11), then his capital at time $t$ will be infinitely close to (12.9). Using Taylor's formula and the fact that $\overline{U}$ satisfies the Black-Scholes equation, we obtain a simplified version of (6.20):

$$\begin{aligned} d\overline{U}(S(t), D(t)) &\approx \frac{\partial \overline{U}}{\partial s}(S(t), D(t))dS(t) + \frac{\partial \overline{U}}{\partial D}(S(t), D(t))dD(t) \\ &\quad + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial s^2}(S(t), D(t))(dS(t))^2 \\ &= \frac{\partial \overline{U}}{\partial s}(S(t), D(t))dS(t) + \frac{\partial \overline{U}}{\partial D}(S(t), D(t))\left(dD(t) + \left(\frac{dS(t)}{S(t)}\right)^2\right) \end{aligned} \tag{12.13}$$

(where $\approx$ stands for equality to within $O((dt)^{3/2})$). For $\mathcal{R}$, we similarly obtain

$$d\overline{R}(S(t), D(t)) \approx \frac{\partial \overline{R}}{\partial s}(S(t), D(t))dS(t) + \frac{\partial \overline{R}}{\partial D}(S(t), D(t)) \left( dD(t) + \left( \frac{dS(t)}{S(t)} \right)^2 \right);$$

solving for $dD(t) + (dS(t)/S(t))^2$,

$$\begin{aligned} &dD(t) + \left( \frac{dS(t)}{S(t)} \right)^2 \\ &\approx \frac{1}{(\partial \overline{R}/\partial D)(S(t), D(t))} d\mathcal{R}(t) - \frac{(\partial \overline{R}/\partial s)(S(t), D(t))}{(\partial \overline{R}/\partial D)(S(t), D(t))} dS(t). \end{aligned} \tag{12.14}$$

Substituting (12.14) into (12.13) and summing over $t$, we find that the strategy of holding

$$\frac{\partial \overline{U}}{\partial s}(S(t), D(t)) - \frac{(\partial \overline{U}/\partial D)(S(t), D(t))\ (\partial \overline{R}/\partial s)(S(t), D(t))}{(\partial \overline{R}/\partial D)(S(t), D(t))} \tag{12.15}$$

units of $\mathcal{S}$ and

$$\frac{(\partial \overline{U}/\partial D)(S(t), D(t))}{(\partial \overline{R}/\partial D)(S(t), D(t))} \tag{12.16}$$

units of $\mathcal{R}$ works if the approximations implicit in (12.13) are not too crude. Here $\overline{R}$ is defined as usual:

$$\overline{R}(s, D) := \int_{\mathbb{R}} R(se^z)\ \mathcal{N}_{-D/2,\, D}(dz).$$

In the case $R(s) = s^2$ the strategy (12.15)–(12.16) reduces to (12.10)–(12.11).

It is clear that the accumulated hedging error has the order of magnitude

$$\begin{aligned} &\mathbf{var}_S(3) + \sum_{t \in \mathbb{T} \setminus \{T\}} |dD(t)||dS(t)| + \mathbf{var}_D(2) \\ &\leq \mathbf{var}_S(3) + (\mathbf{var}_D(p))^{1/p} (\mathbf{var}_S(q))^{1/q} + \mathbf{var}_D(2), \end{aligned}$$

where $p, q \in (1, \infty)$ are chosen to satisfy $\operatorname{vex} D < p$, $\operatorname{vex} S < q$ and $1/p + 1/q = 1$. So the assumptions $\operatorname{vex} D < 2$ and $\operatorname{vex} S = 2$ (as we know, the latter can be added without loss of generality) imply that the accumulated error will be infinitesimal. ∎

The theorem is easily generalized to typical payoff functions, such as those for calls and puts, which can be arbitrarily well approximated with smooth functions with bounded derivatives.

Implementation of our proposal should be based, of course, on a careful analysis of the error in discrete time, along the lines of our analysis of the Bachelier and Black-Scholes protocols in Chapter 10.

### Alternatives to the Square

As we have indicated, the square can be replaced by a different strictly convex function $R$, and in some cases this may be desirable. This subsection consists of two parts: first we discuss possible implementations of the square and then we discuss more substantive alternatives.

The simplest alternative is to set $R(s) := (s-a)^2$ for some real number $a$. Having the market price $(S(T)-a)^2$ is obviously equivalent to having it price $S^2(T)$, because

$$(S(T)-a)^2 = S^2(T) - 2aS(T) + a^2,$$

and the market is already pricing $S(T)$: its price for $S(T)$ at time $t$ is $S(t)$. But at a practical level, trading in $(S(T)-a)^2$ for a value $a$ close to $S$'s current price may be more sensible than trading in $S^2(T)$, if only because $(S(T)-a)^2$ is easier to think about. Further "normalization" of the square can be achieved by defining the final payoff to be, say,

$$\left(\frac{S(T)-S(0)}{S(0)}\right)^2. \tag{12.17}$$

Notice that Equation (12.17) has exactly the same form as the dividends for $\mathcal{D}$; the difference is that for $\mathcal{D}$ the dividends are paid very often and for the square, as represented by (12.17), there is only one dividend, at the maturity.

Other convex (but not linear) functions provide more substantive alternatives; their advantage over the square might be a slower growth (exponential in the case of the square; see (12.7)) of $\mathcal{R}(t)$ as $D(t)$ increases. In what follows we will exclude the trivial case of a linear payoff function from consideration. Convexity is sufficient for

$$\overline{R}(s,D) := \int_{\mathbb{R}} R(se^z)\, \mathcal{N}_{-D/2,\,D}(dz) \tag{12.18}$$

to be a strictly increasing function of $D$ for fixed $s$; since this function will typically be continuous, (12.6) can be solved to obtain $D(t)$. In Chapter 10 we studied the volatility path $D(t)$ implied by call and put options. Whereas we believe that encouraging results obtained there show the feasibility of our approach to option pricing, calls and puts, with their piecewise linear payoffs, might be too awkward analytically to be used effectively for hedging in place of the square (for concreteness we are discussing a call):

- Formally, $\overline{R}(s,D)$ (as defined in (12.18)) is a convex function of $s$, but in reality it will be almost flat unless the call is nearly at-the-money (especially for small $D$). Therefore, to simulate the dividend-paying security $\mathcal{D}$ (see (12.14)) we will need a large amount of the call, which may make the transaction costs involved too heavy.

- Our hedging strategies require that the function $\overline{R}(s,D)$ should be smooth and that its derivatives should be sufficiently well-behaved; this becomes a problem in the case of a call when $D$ becomes small: in the limit as $D \to 0$, the payoff is not smooth at all.

Ideally, the payoff function $R$ should be smooth and strictly convex (recall that a smooth function $R$ is *strictly convex* if $R''(s) > 0$ for all $s > 0$).

## 12.3 GAMES FOR PRICE PROCESSES WITH JUMPS

In this section, we consider protocols in which the price of the underlying security $\mathcal{S}$ is allowed to jump. We show that our methods can handle jumps if the market itself prices securities that pay dividends when jumps of different sizes occur.

A fundamental tool for dealing with jumps is the *Poisson distribution*, which plays a role analogous to the role played by the Gaussian distribution for continuous processes. This distribution depends on a nonnegative parameter $D$, equal to its mean. We write $\mathcal{P}_D$ for the Poisson distribution with parameter $D$; it assigns probability

$$\mathcal{P}_D\{z\} := \frac{D^z}{z!}e^{-D} \tag{12.19}$$

to each $z \in \mathbb{Z}^+$. Here $\mathbb{Z}^+$ designates the set of all nonnegative integers, and $0^0$ and $0!$ are understood to equal 1.

We begin by studying a continuous-time process $\mathcal{S}$ that consists purely of jumps, always of the same size. Such counting processes occur in many domains, but for concreteness, we imagine that $\mathcal{S}$ counts the number of hurricanes. As we show, a payoff that depends on the number of hurricanes between time 0 and time $T$ can be priced by its expected value under (12.19) if $\mathcal{D}$ is the market price for a dividend-paying derivative that pays a dollar every time a hurricane occurs.

We then turn to the case of two simultaneous counting processes, for both of which the market prices a dividend-paying security. In this case, a payoff that depends on both counting processes can be priced using two Poisson distributions, with means equal to the market prices for the two dividend-paying derivatives. The distributions are used as if they were independent, but the source of the independence is not stochastic; rather it lies in the fact that an investor can choose how many units of the one derivative he holds independently of how many units of the other he holds (combined with the requirement that the processes do not jump simultaneously).

Finally, we turn to the problem of pricing an option on a security whose price can change continuously and can also jump. Combining what we know about the Bachelier process with what we have learned about simultaneous counting processes, we propose having the market price separate dividend-paying securities for continuous volatility and jumps. This approach can be applied to Black-Scholes pricing with jumps, but for simplicity we only briefly discuss Bachelier pricing with jumps.

### Weather Derivatives

Suppose $S(t)$ is the number of hurricanes in the time interval $[0, t]$, and suppose that as time proceeds, $S(t)$ increases in distinct jumps, always by 1; if two hurricanes occur simultaneously, we count them as one hurricane. And suppose we want to price a European option $U(S(T))$. Such an option is an example of a *weather derivative*. Such derivatives are actually traded; they are used to hedge insurance policies against weather damage. (See, e.g., the web site of the Met Office, the British weather service.)

We can price $U(S(T))$ if the market prices a security $\mathcal{D}$ that pays the dividend

$$\begin{cases} 0 \text{ if } dS(t) = 0 \\ 1 \text{ if } dS(t) = 1 \end{cases}$$

during each infinitesimal interval $[t, t + dt]$.

In our protocol, Investor plays against Reality, who determines the hurricanes, and Market, who sets the price for $\mathcal{D}$. Investor is Player I and Reality and Market together are Player II (see §1.1). As always, prices are defined by Player I's strategies and their capital processes.

THE POISSON PROTOCOL
**Players:** Investor, Market, Reality
**Protocol:**

$\mathcal{I}_0 := 0$.
$S_0 := 0$.
Market announces $D_0 > 0$.
FOR $n = 1, 2, \ldots, N$:
  Investor announces $V_n \in \mathbb{R}$.
  Reality announces $x_n \in \{0, 1\}$.
  Market announces $D_n \geq 0$.
  $S_n := S_{n-1} + x_n$.
  $\mathcal{I}_n := \mathcal{I}_{n-1} + V_n(x_n + \Delta D_n)$. (12.20)

**Additional Constraints on Reality and Market:** Reality must make $S_N < \infty$. Market must make the path $D$ is continuous, with $D_n > 0$ for $n = 1, \ldots, N-1$, $D_N = 0$, $\sup_n D_n < \infty$, and $\operatorname{vex} D < 2$.

This protocol codes a hurricane as 1 and its absence as 0. Equation (12.20), as usual, takes into account both the dividend $x_n$ and capital gain $D_n - D_{n-1}$ for each of the $V_n$ shares of $\mathcal{D}$.

**Proposition 12.1** *Suppose the function $U: \mathbb{Z}^+ \to \mathbb{R}^+$ satisfies the growth condition*

$$\exists C > 0: \ U(s) = O(C^s). \tag{12.21}$$

*Then the price for a European option $U(S(T))$ at the time when $D(0)$ has just been announced in the Poisson protocol is*

$$\int_{\mathbb{Z}^+} U(z) \ \mathcal{P}_{D(0)}(dz).$$

*Proof* We will find a strategy $\mathcal{V}$ such that

$$\int_{\mathbb{Z}^+} U(z) \ \mathcal{P}_{D(0)}(dz) + \mathcal{I}^{\mathcal{V}}(S, D) \approx U(S(T)). \tag{12.22}$$

Set

$$\overline{U}(s, D) := \int_{\mathbb{Z}^+} U(s + z) \ \mathcal{P}_D(dz)$$

for $s = 0, 1, \ldots$ and $D \geq 0$. Equation (12.22) can be rewritten as

$$\overline{U}(S(0), D(0)) + \mathcal{I}^{\mathcal{V}}(S, D) \approx \overline{U}(S(T), D(T)). \tag{12.23}$$

It can be checked by direct differentiation that $\overline{U}$ satisfies, for $D > 0$,

$$\frac{\partial \overline{U}}{\partial D}(s, D) = \overline{U}(s+1, D) - \overline{U}(s, D). \tag{12.24}$$

(According to Leibniz's differentiation rule for integrals, (12.21), which implies $U(s) = o(\epsilon^s s!)$, $\forall \epsilon > 0$, is a sufficient condition for the differentiation to be valid.)

Our strategy is simply

$$\mathcal{V} := \frac{\partial \overline{U}}{\partial D}$$

(exactly as in the Black-Scholes case, except that no shares of $S$ are bought). Using Taylor's formula, we find for $t \in \mathbb{T} \setminus \{T\}$: if $dS(t) = 0$,

$$d\overline{U}(S(t), D(t)) = \frac{\partial \overline{U}}{\partial D}(S(t), D(t))dD(t) + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial D^2}(S(t), D(t) + \theta dD(t))(dD(t))^2 \tag{12.25}$$

(as usual, $\theta$ stand for different numbers in $(0,1)$), and if $dS(t) = 1$,

$$\begin{aligned} d\overline{U}(S(t), D(t)) &= \left(\overline{U}(S(t)+1, D(t)) - \overline{U}(S(t), D(t))\right) \\ &+ \left(\overline{U}(S(t)+1, D(t) + dD(t)) - \overline{U}(S(t)+1, D(t))\right) \\ &= \frac{\partial \overline{U}}{\partial D}(S(t), D(t)) + \frac{\partial \overline{U}}{\partial D}(S(t)+1, D(t) + \theta dD(t))dD(t) \end{aligned} \tag{12.26}$$

(equation (12.24) was used here). Comparing (12.25) with

$$d\mathcal{I}_n = \frac{\partial \overline{U}}{\partial D}(S(t), D(t))dD(t)$$

(valid when $dS(t) = 0$) and (12.26) with

$$d\mathcal{I}_n = \frac{\partial \overline{U}}{\partial D}(S(t), D(t))(1 + dD(t))$$

(valid when $dS(t) = 1$), we see that the difference between the left-hand and right-hand sides of (12.23) does not exceed

$$\begin{aligned} &\frac{1}{2}\sum_t \left|\frac{\partial^2 \overline{U}}{\partial D^2}(S(t), D(t) + \theta dD(t))(dD(t))^2\right| + \sum_{t:dS(t)=1} \left|\frac{\partial \overline{U}}{\partial D}(S(t), D(t))dD(t)\right| \\ &\quad + \sum_{t:dS(t)=1} \left|\frac{\partial \overline{U}}{\partial D}(S(t)+1, D(t) + \theta dD(t))dD(t)\right| \\ &\quad \leq \frac{1}{2}c_2 \sum_t (dD(t))^2 + 2S(T)c_1 \max_t |dD(t)| \approx 0, \end{aligned}$$

where $c_1$ is an upper bound on $\left|(\partial \overline{U}/\partial D)(s, D)\right|$ and $c_2$ is an upper bound on $\left|(\partial^2 \overline{U}/\partial D^2)(s, D)\right|$, $s$ ranging over $[0, S(T)+1] \cap \mathbb{Z}^+$ and $D$ over $[0, \sup_t D(t)]$. An apparent gap in this argument is that $c_1$ and $c_2$ may not necessarily be finite because of the possibility of erratic behavior of $(\partial \overline{U}/\partial D)$ and $(\partial^2 \overline{U}/\partial D^2)$ as $D \to 0$. However, it can be seen that condition (12.21) (in conjunction with the Lebesgue dominated convergence theorem) implies $\overline{U}(s, D) \to U(s)$ as $D \to 0$; therefore, (12.24) implies that the limit

$$\lim_{D\to 0} \frac{\partial \overline{U}}{\partial D}(s, D) = U(s+1) - U(s)$$

exists for every $s$, and the easy corollary

$$\begin{aligned} \frac{\partial^2 \overline{U}}{\partial D^2}(s, D) &= \frac{\partial \overline{U}}{\partial D}(s+1, D) - \frac{\partial \overline{U}}{\partial D}(s, D) \\ = \Big(\overline{U}(s+2, D) - \overline{U}(s+1, D)\Big) &- \Big(\overline{U}(s+1, D) - \overline{U}(s, D)\Big) \\ &= \overline{U}(s+2, D) - 2\overline{U}(s+1, D) + \overline{U}(s, D) \end{aligned}$$

of (12.24) implies that the limit

$$\lim_{D\to 0} \frac{\partial^2 \overline{U}}{\partial D^2}(s, D) = U(s+2) - 2U(s+1) + U(s)$$

exists for every $s$. ∎

As the proof shows, the growth condition (12.21) can be relaxed to $U(s) = (o(s))^s$ or, more formally, $\lim_{s\to\infty}(U(s))^{1/s}/s = 0$.

Proposition 12.1 is analogous to a result in Dambis (1965) and Dubins and Schwarz (1965), which says that a point process with continuous compensator can be transformed into a Poisson process by a change of time. (Similarly, our game-theoretic Bachelier formula is analogous to the measure-theoretic result in Meyer (1971) and Papangelou (1972), that a continuous martingale can be transformed into a Wiener process by a change of time.)

### Multivalued Processes

The preceding theory can be generalized to price options that depend on any number of counting processes. For simplicity, we consider the case of two counting processes; the extension to more than two is straightforward though notationally awkward.

Let $S^{(1)}(t)$ and $S^{(2)}(t)$ be the number of hurricanes and hailstorms, respectively, in the time interval $[0, t]$. We assume that only one storm can occur at a time: $dS^{(1)}(t)$ and $dS^{(2)}(t)$ cannot exceed 1, and $dS^{(1)}(t)$ and $dS^{(2)}(t)$ cannot be 1 simultaneously.

In order to price a European option $U(S^{(1)}(T), S^{(2)}(T))$, where $U : (\mathbb{Z}^+)^2 \to \mathbb{R}^+$, we will need the Market to price two dividend-paying derivatives, $D^{(1)}$ and $D^{(2)}$. The derivative $D^{(k)}$, for $k = 1, 2$, pays the dividend

$$\begin{cases} 0 \text{ if } dS^{(k)}(t) = 0 \\ 1 \text{ if } dS^{(k)}(t) = 1 \end{cases}$$

during the infinitesimal interval $[t, t + dt]$.

POISSON PROTOCOL FOR TWO PROCESSES
**Players:** Investor, Market, Reality

**Protocol:**

$\mathcal{I}_0 := 0.$
$S_0^{(1)} := 0, S_0^{(2)} := 0.$
Market announces $D_0^{(1)} > 0$ and $D_0^{(2)} > 0$.
FOR $n = 1, 2, \ldots, N$:
  Investor announces $V_n^{(1)} \in \mathbb{R}$ and $V_n^{(2)} \in \mathbb{R}$.
  Reality announces $x_n^{(1)}, x_n^{(2)} \in \{0, 1\}$ with $x_n^{(1)} x_n^{(2)} = 0$.
  Market announces $D_n^{(1)} \geq 0$ and $D_n^{(2)} \geq 0$.
  $S_n^{(1)} := S_{n-1}^{(1)} + x_n^{(1)}$.
  $S_n^{(2)} := S_{n-1}^{(2)} + x_n^{(2)}$.
  $\mathcal{I}_n := \mathcal{I}_{n-1} + V_n^{(1)} \left( x_n^{(1)} + \Delta D_n^{(1)} \right) + V_n^{(2)} \left( x_n^{(2)} + \Delta D_n^{(2)} \right)$.

**Additional Constraints on Reality and Market:** Reality must make $S_N^{(k)} < \infty$. Market must make the path $D^{(k)}$ is continuous, with $D_n^{(k)} > 0$ for $n = 1, \ldots, N-1$, $D_N^{(k)} = 0$, $\sup_n D_n^{(k)} < \infty$, and $\operatorname{vex} D^{(k)} < 2$.

**Theorem 12.2** *Suppose the function $U\colon (\mathbb{Z}^+)^2 \to \mathbb{R}^+$ satisfies the growth condition*

$$\exists C > 0: \; U(s^{(1)}, s^{(2)}) = O\left( C^{s^{(1)} + s^{(2)}} \right). \tag{12.27}$$

*The price for a European option $U(S^{(1)}(T), S^{(2)}(T))$ at the time when $D^{(1)}(0)$ and $D^{(2)}(0)$ have just been announced in the Poisson protocol for two processes is*

$$\int U(z_1, z_2) \; \mathcal{P}_{D^{(1)}(0)}(dz_1) \, \mathcal{P}_{D^{(2)}(0)}(dz_2).$$

*Proof* Our goal is to find strategies $\mathcal{V}^{(1)}$ and $\mathcal{V}^{(2)}$ such that

$$\int U(z_1, z_2) \; \mathcal{P}_{D^{(1)}(0)}(dz_1) \, \mathcal{P}_{D^{(2)}(0)}(dz_2) + \mathcal{I}^{\mathcal{V}^{(1)}, \mathcal{V}^{(2)}} \left( S^{(1)}, S^{(2)}, D^{(1)}, D^{(2)} \right) \approx U(S^{(1)}(T), S^{(2)}(T)). \tag{12.28}$$

For nonnegative $s^{(1)}$, $s^{(2)}$, $D^{(1)}$, and $D^{(2)}$, set

$$\overline{U} \left( s^{(1)}, s^{(2)}, D^{(1)}, D^{(2)} \right) := \int U(s^{(1)} + z_1, s^{(2)} + z_2) \; \mathcal{P}_{D^{(1)}}(dz_1) \, \mathcal{P}_{D^{(2)}}(dz_2)$$

and rewrite the equality in (12.28) as

$$\overline{U} \left( S^{(1)}(0), S^{(2)}(0), D^{(1)}(0), D^{(2)}(0) \right) + \mathcal{I}^{\mathcal{V}^{(1)}, \mathcal{V}^{(2)}} \left( S^{(1)}, S^{(2)}, D^{(1)}, D^{(2)} \right) \approx \overline{U} \left( S^{(1)}(T), S^{(2)}(T), D^{(1)}(T), D^{(2)}(T) \right). \tag{12.29}$$

Direct differentiation and application of Leibniz's differentiation rule for integrals shows that our growth condition (12.27) on $U$ implies

$$\begin{aligned}
&\frac{\partial \overline{U}}{\partial D^{(1)}}\left(s^{(1)}, s^{(2)}, D^{(1)}, D^{(2)}\right) \\
&= \overline{U}\left(s^{(1)}+1, s^{(2)}, D^{(1)}, D^{(2)}\right) - \overline{U}\left(s^{(1)}, s^{(2)}, D^{(1)}, D^{(2)}\right), \\
&\frac{\partial \overline{U}}{\partial D^{(2)}}\left(s^{(1)}, s^{(2)}, D^{(1)}, D^{(2)}\right) \\
&= \overline{U}\left(s^{(1)}, s^{(2)}+1, D^{(1)}, D^{(2)}\right) - \overline{U}\left(s^{(1)}, s^{(2)}, D^{(1)}, D^{(2)}\right).
\end{aligned} \tag{12.30}$$

Our strategy is to buy

$$\mathcal{V}^{(1)} := \frac{\partial \overline{U}}{\partial D^{(1)}} \quad \text{and} \quad \mathcal{V}^{(2)} := \frac{\partial \overline{U}}{\partial D^{(2)}}$$

shares of $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(2)}$, respectively. If $dS^{(1)}(t) = dS^{(2)}(t) = 0$, Taylor's formula gives

$$\begin{aligned}
&d\overline{U}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right) \\
&= \frac{\partial \overline{U}}{\partial D^{(1)}}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right) dD^{(1)}(t) \\
&+ \frac{\partial \overline{U}}{\partial D^{(2)}}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right) dD^{(2)}(t) \\
&+ \frac{1}{2}\frac{\partial^2 \overline{U}}{(\partial D^{(1)})^2}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t)+\theta dD^{(1)}(t), D^{(2)}(t)+\theta dD^{(2)}(t)\right)\left(dD^{(1)}(t)\right)^2 \\
&+ \frac{1}{2}\frac{\partial^2 \overline{U}}{(\partial D^{(2)})^2}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t)+\theta dD^{(1)}(t), D^{(2)}(t)+\theta dD^{(2)}(t)\right)\left(dD^{(2)}(t)\right)^2 \\
&+ \frac{\partial^2 \overline{U}}{\partial D^{(1)}\partial D^{(2)}}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t)+\theta dD^{(1)}(t), D^{(2)}(t)+\theta dD^{(2)}(t)\right) \\
&\qquad dD^{(1)}(t)dD^{(2)}(t).
\end{aligned}$$

If $dS^{(1)}(t) = 1$, we get

$$\begin{aligned}
&d\overline{U}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right) \\
&= \left(\overline{U}\left(S^{(1)}(t)+1, S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right) - \overline{U}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right)\right) \\
&+ \left(\overline{U}\left(S^{(1)}(t)+1, S^{(2)}(t), D^{(1)}(t)+dD^{(1)}(t), D^{(2)}(t)+dD^{(2)}(t)\right)\right. \\
&\qquad \left. - \overline{U}\left(S^{(1)}(t)+1, S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right)\right) \\
&= \frac{\partial \overline{U}}{\partial D^{(1)}}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right) \\
&+ \frac{\partial \overline{U}}{\partial D^{(1)}}\left(S^{(1)}(t)+1, S^{(2)}(t), D^{(1)}(t)+\theta dD^{(1)}(t), D^{(2)}(t)+\theta dD^{(2)}(t)\right) dD^{(1)}(t)
\end{aligned}$$

$$+ \frac{\partial \overline{U}}{\partial D^{(2)}} \left( S^{(1)}(t) + 1, S^{(2)}(t), D^{(1)}(t) + \theta dD^{(1)}(t), D^{(2)}(t) + \theta dD^{(2)}(t) \right) dD^{(2)}(t).$$

If $dS^{(2)}(t) = 1$, we get

$$d\overline{U}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right)$$

$$= \left( \overline{U}\left(S^{(1)}(t), S^{(2)}(t) + 1, D^{(1)}(t), D^{(2)}(t)\right) - \overline{U}\left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right) \right)$$

$$+ \left( \overline{U}\left(S^{(1)}(t), S^{(2)}(t) + 1, D^{(1)}(t) + dD^{(1)}(t), D^{(2)}(t) + dD^{(2)}(t)\right) \right.$$

$$\left. - \overline{U}\left(S^{(1)}(t), S^{(2)}(t) + 1, D^{(1)}(t), D^{(2)}(t)\right) \right)$$

$$= \frac{\partial \overline{U}}{\partial D^{(2)}} \left(S^{(1)}(t), S^{(2)}(t), D^{(1)}(t), D^{(2)}(t)\right)$$

$$+ \frac{\partial \overline{U}}{\partial D^{(1)}} \left( S^{(1)}(t), S^{(2)}(t) + 1, D^{(1)}(t) + \theta dD^{(1)}(t), D^{(2)}(t) + \theta dD^{(2)}(t) \right) dD^{(1)}(t)$$

$$+ \frac{\partial \overline{U}}{\partial D^{(2)}} \left( S^{(1)}(t), S^{(2)}(t) + 1, D^{(1)}(t) + \theta dD^{(1)}(t), D^{(2)}(t) + \theta dD^{(2)}(t) \right) dD^{(2)}(t).$$

We used (12.30) in the last two inequalities.

We can see that the difference between the left-hand and right-hand sides of (12.29) does not exceed

$$\frac{1}{2} c_{1,1} \sum_t (dD^{(1)}(t))^2 + \frac{1}{2} c_{2,2} \sum_t (dD^{(2)}(t))^2 + c_{1,2} \sum_t \left| dD^{(1)}(t) dD^{(2)}(t) \right|$$
$$+ 2\left(S^{(1)}(T) + S^{(2)}(T)\right) c_1 \max_t |dD^{(1)}(t)| + 2\left(S^{(1)}(T) + S^{(2)}(T)\right) c_2 \max_t |dD^{(2)}(t)|$$
$$\approx 0,$$

where $c_k$ is an upper bound (over the relevant area) on $\left|\partial \overline{U} / \partial D^{(k)}\right|$ and $c_{k,l}$ is an upper bound on $\left|\partial^2 \overline{U} / \left(\partial D^{(k)} \partial D^{(l)}\right)\right|$. We have used the facts that the cross-product is infinitesimal,

$$\sum_t \left| dD^{(1)}(t) dD^{(2)}(t) \right| \le \left( \sum_t \left( dD^{(1)}(t) \right)^2 \right)^{1/2} \left( \sum_t \left( dD^{(2)}(t) \right)^2 \right)^{1/2} \approx 0,$$

and that $c_k$ and $c_{k,l}$ are finite (this can be proven as in Proposition 12.1). ∎

Consider the process $S(t)$ defined by

$$S(t) := S^{(1)}(t) - S^{(2)}(t).$$

We could think of this as a security price that can change only by jumps, a jump always being 1 or $-1$. A European option on such a security is priced by Theorem 12.2; its price is

$$\int U(z_1 - z_2)\, \mathcal{P}_{D^{(1)}(0)}(dz_1)\, \mathcal{P}_{D^{(2)}(0)}(dz_2).$$

This simple model is in the spirit of the model of Cox, Ross, and Rubinstein 1979, except that here the timing of the jumps is not known in advance. Other differences between the two models, such as the positivity of prices in the Cox, Ross, and Rubinstein model, are easily eliminated.

## Putting Jumps in Price Processes

The ideas of the preceding subsections can be combined to price options on share prices that can change continuously and can also jump.

In its simplest form, at least, this combination requires that market participants be able to identify when a jump occurs, and it limits the different possible sizes for jumps. For each size that is permitted, there must be a dividend-paying security. It may be possible to extend the method so as to relax these assumptions.

For simplicity, let us assume that there are only two kinds of external events: positive ones, which cause the price to jump by 1, and negative ones, which cause it to jump by $-1$. We write $S^{(0)}$, $S^{(-1)}$, and $S^{(1)}$ for the three components of $S$: the diffusion component $S^{(0)}$, the cumulative effect $S^{(1)}$ of jumps of 1, and the cumulative effect $S^{(-1)}$ of jumps of $-1$. For each component, we ask the market to price dividend-paying derivatives: $\mathcal{D}^{(0)}$, responsible for the diffusion component $S^{(0)}$ of $S$, $\mathcal{D}^{(-1)}$, responsible for negative jumps, and $\mathcal{D}^{(1)}$, responsible for positive jumps.

This gives the following protocol:

BACHELIER'S PROTOCOL WITH JUMPS
**Players:** Investor, Market
**Protocol:**

$\mathcal{I}_0 := 0$.
$S_0^{(-1)} := 0$, $S_0^{(1)} := 0$.
Market announces $S_0^{(0)} \in \mathbb{R}$, $D_0^{(0)} > 0$, $D_0^{(-1)} > 0$, and $D_0^{(1)} > 0$.
FOR $n = 1, 2, \dots, N$:
Investor announces $M_n \in \mathbb{R}$, $V_n \in \mathbb{R}$, $V_n^{(-1)} \in \mathbb{R}$, and $V_n^{(1)} \in \mathbb{R}$.
Market announces $x_n^{(0)} \in \mathbb{R}$.
Market announces $x_n^{(-1)} \in \{-1, 0\}$ and $x_n^{(1)} \in \{0, 1\}$ with $x_n^{(-1)} x_n^{(1)} = 0$.
Market announces $D_n^{(0)} \geq 0$, $D_n^{(-1)} \geq 0$, and $D_n^{(1)} \geq 0$.
$S_n^{(0)} := S_{n-1}^{(0)} + x_n^{(0)}$.
$S_n^{(-1)} := S_{n-1}^{(-1)} + x_n^{(-1)}$.
$S_n^{(1)} := S_{n-1}^{(1)} + x_n^{(1)}$.
$S_n := S_n^{(0)} + S_n^{(-1)} + S_n^{(1)}$.
$$\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n \left( \Delta D_n^{(0)} + \left( \Delta S_n^{(0)} \right)^2 \right) + V_n^{(-1)} \left( x_n^{(-1)} + \Delta D_n^{(-1)} \right) + V_n^{(1)} \left( x_n^{(1)} + \Delta D_n^{(1)} \right).$$

**Additional Constraints on Market:** For $k = -1, 0, 1$, Market must ensure that $S_N^{(k)} < \infty$, the path $S^{(0)}$ is continuous, $\inf_n S_n^{(0)}$ is positive and not infinitesimal,

$\sup_n S_n^{(0)} < \infty$, the path $D^{(k)}$ is continuous, with $D_n^{(k)} > 0$ for $n = 1, \ldots, N-1$, $D_N^{(k)} = 0$, $\sup_n D_n^{(k)} < \infty$, and $\operatorname{vex} D^{(k)} < 2$.

This protocol does lead to a game-theoretic price for a well-behaved option on $\mathcal{S}$, which can be found from the appropriate partial differential equation by numerical methods. Because there does not seem to be a simple formula for the price, we do not state this conclusion as a formal proposition.

## 12.4 APPENDIX: THE STABLE AND INFINITELY DIVISIBLE LAWS

In this appendix, we provide some standard definitions, results, and references concerning stable and infinitely divisible probability distributions (or laws, as one often says).

### Stable Distributions

The study of the stable distributions goes back to the work of Augustin Louis Cauchy (1789–1857) in 1853. The topic was taken up by Paul Lévy in 1925. (For an excellent historical account see Gnedenko [134] .) Different authors define stability in slightly different ways, but we can use the definition due to Khinchin: a probability distribution $P$ is *stable* if, for any two independent random variables $x$ and $y$ distributed as $P$ and any two real numbers $a$ and $b$, the sum $ax + by$ has the same distribution as $cx + d$ for some real numbers $c$ and $d$. The stable distributions arise in the theory of sums of independent random variables: a distribution can be the limit of the distributions of the random variables

$$y_n := \frac{1}{a_n} \sum_{i=1}^{n} x_i - b_n,$$

where $x_i$ are independent and identically distributed, if and only if it is stable; see Gnedenko and Kolmogorov [135], Theorem 35.2.

Recall that the *characteristic function* for a random variable $X$ is the function $\psi(u)$, in general complex-valued, defined by $\psi(u) := \mathbb{E}(e^{iuX})$. Khinchin and Lévy (1936) showed that for a typical stable law,

$$\log \psi(u) = i\gamma u - c|u|^{\alpha} \left(1 + i\beta(\operatorname{sign} u) \tan \frac{\alpha\pi}{2}\right), \qquad (12.31)$$

where $\gamma$ is a location parameter, $c \geq 0$ is a scale parameter, $\beta \in [-1, 1]$ is an index of skewness, and $\alpha$, usually of most interest, is the *characteristic exponent*; recall that $\operatorname{sign} u$ is defined to be 1 if $u$ is positive, $-1$ if $u$ is negative, and 0 if $u$ is 0. The characteristic exponent must lie in the range $0 < \alpha \leq 2$. Equation (12.31) is actually valid only when $\alpha \neq 1$; when $\alpha = 1$, $\tan \frac{\alpha\pi}{2}$ is not defined and has to be replaced by $\frac{2}{\pi} \log |u|$. Except for the Gaussian distribution ($\alpha = 2$) and a couple of other special cases, there is no simple expression for the density of a stable law ([121], Chapter XVII), but (12.31) shows how small the set of stable laws is: the stable types (for any probability distribution $P$ its type is defined to be the distributions of the random variables $aX + b$, $a$ and $b$ ranging over the reals) depend on just two parameters, $\alpha$ and $\beta$.

Cauchy discovered that the characteristic functions of symmetric stable laws are all of the simple form

$$\log \psi(u) = -c|u|^{\alpha}. \tag{12.32}$$

Important special cases, besides the Gaussian distribution ($\alpha = 2$), are the Cauchy distribution ($\alpha = 1$; actually discovered by Poisson) and the Holtzmark distributions ($\alpha = 0.5, 1.5$). The latter describe the gravitational field created by randomly placed stars; $\alpha = 0.5$ when the stars are distributed in a one-dimensional space, and $\alpha = 1.5$ in the three-dimensional case. In the one-dimensional case the Holtzmark distribution gives way to distribution (12.32) if the usual gravitational law of inverse squares is replaced by the law of inverse $1/\alpha$th power, for any $0 < \alpha < 2$ (the Gaussian case, $\alpha = 2$, is excluded). In all three cases, $\alpha = 0.5$, $\alpha = 1$, and $\alpha = 1.5$, closed-form expressions for densities are known. For details, see [121, 188].

The characteristic exponent characterizes the behavior of the tails of the distribution—that is, the probabilities for extreme values. For a nondegenerate random variable $X$ following a stable law with a characteristic exponent $\alpha$ less than 2, the variance does not exist; in fact, the moment $\mathbb{E}(|X|^k)$ exists when and only when $k < \alpha$ (Gnedenko and Kolmogorov 1954).

## Infinitely Divisible Distributions

The stable distributions are included in a larger class of distributions that are called *infinitely divisible*. A probability distribution is infinitely divisible if for any positive integer $k$ it can be represented as the law of a sum $\sum_{n=1}^{k} x_n$ of independent identically distributed random variables $x_n$. The study of infinitely divisible distributions began in 1929 with de Finetti [89, 88, 87, 90], and was continued by Kolmogorov (1932) and Lévy (1934). For a typical infinitely divisible law,

$$\log \psi(u) = ibu + \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} \left(e^{iut} - 1 - iut\,\mathbb{I}_{|t|\leq 1}\right) \Pi(dt), \tag{12.33}$$

where $\Pi$ is a measure on $\mathbb{R} \setminus \{0\}$ with

$$\int_{[-1,1]} t^2 \Pi(dt) < \infty \quad \text{and} \quad \Pi((-\infty, -1) \cup (1, \infty)) < \infty,$$

$b$ is a real number (a location parameter), and $\sigma^2$ is a nonnegative number (the variance of the Gaussian component). Because $\Pi$ , which is called the *Lévy measure*, can be chosen freely, the class of infinitely divisible laws is infinite dimensional, in contrast with the class of stable laws, which, aside from the location and scale parameters, has only the two parameters $\alpha$ and $\beta$.

Formula (12.33) is *Lévy's formula* or the *Lévy-Khinchin formula*; it was proven by Lévy (1934) and Khinchin (1937). Different authors, including Lévy and Khinchin, write it in slightly different ways.

The role of infinitely divisible distributions as limits of sums of independent random variables is clarified by the following fundamental result [171, 242, 243]:

**Khinchin's Theorem** *Suppose the random variables*

$$\begin{matrix} X_{11}, X_{12}, \ldots, X_{1k_1} \\ X_{21}, X_{22}, \ldots, X_{2k_2} \\ \vdots \end{matrix}$$

*are independent within rows. Suppose $k_n \to \infty$ as $n \to \infty$ and the condition of infinite smallness*

$$\max_{1 \le k \le k_n} \mathbb{P}\{X_{nk} \ge \epsilon\} \to 0 \quad (n \to \infty)$$

*holds for every fixed $\epsilon > 0$. If the distributions of the sum $\sum_{k=1}^{k_n} X_{nk}$ converge to some probability distribution, that probability distribution is infinitely divisible.*

Boris Gnedenko (1912–1995) established necessary and sufficient conditions under which $\sum_{k=1}^{k_n} X_{nk}$ converges to a given infinitely divisible law [135, 242, 243]. There are a number of results here waiting to be extended to the game-theoretic framework.

It remains to explain how (12.31) is a special case of (12.33). As we know, $\gamma$ and $c$ in (12.31) are the location and scale parameters. In the case $\alpha = 2$, the Lévy measure $\Pi$ is zero and $\sigma^2 = c$; in all other cases $\sigma^2 = 0$. When $0 < \alpha < 2$, the Lévy measure $\Pi$ satisfies

$$\Pi((-\infty, -t)) = c_1 t^{-\alpha} \ \& \ \Pi((t, \infty)) = c_2 t^{-\alpha}, \quad \forall t > 0,$$

for some constants $c_1 \ge 0$ and $c_2 \ge 0$. The index of skewness $\beta$ can be defined as

$$\beta = \frac{c_1 - c_2}{c_1 + c_2};$$

arbitrarily if $c_1 + c_2 = 0$ (see [135]).

## Lévy Processes

A càdlàg stochastic process $X$ with time parameter ranging from zero to infinity is a *Lévy process* if its increments $X(t+h) - X(t)$ and independent and stationary (the distribution of $X(t+h) - X(t)$ depends only on $h$ but not on $t$). This implies that the increments have infinitely divisible distributions. In fact, there is a one-to-one correspondence between the Lévy processes and infinitely divisible distributions; for every infinitely divisible distribution $P$, there is a Lévy process $X$ such that $X(1)$ is distributed as $P$. An excellent book on Lévy processes is Bertoin [20]; for general processes with independent increments, see Skorohod [289]. For a brief and readable review of Gaussian and Lévy processes, see §I.4 of [262].

Suppose $X$ is a Lévy process such that the characteristic function $\psi$ of $X(1)$ satisfies Equation (12.33). Then $\Pi$ describes the jumps of $X$. Heuristically, if $A$

is an open interval of the real line not containing 0, then $\Pi(A) < \infty$, and for any infinitesimal interval $[t, t + dt)$ the number

$$\eta(t, A) = \# \{s \in [t, t + dt) \mid X(s) - X(s-) \in A\}$$

of jumps in that interval belonging to $A$ has expected value $\Pi(A)dt$; for disjoint intervals $[t_1 + dt)$ and $[t_2 + dt)$, the random variables $\eta(t_1, A)$ and $\eta(t_2, A)$ are independent [20].

Equation (12.33) has the following interpretation in the language of Lévy processes. Every Lévy process can be represented as the sum of four independent components:

1. The deterministic function $X_1(t) = bt$; the characteristic function $\psi_1$ of $X_1(1)$ satisfies
$$\log \psi_1(u) = ibu$$
(i.e., corresponds to the first addend in (12.33)).

2. A scaled Brownian motion $X_2(t) = \sigma W(t)$, where $W$ is a standard Wiener process. The characteristic function $\psi_2$ of $X_2(1)$ satisfies
$$\log \psi_2(u) = \frac{1}{2}\sigma^2 u^2$$
(corresponds to the second addend in (12.33)).

3. A pure jump process $X_3$ with all jumps greater than 1 in absolute value. This is a compound Poisson process (the latter is defined similarly to the usual Poisson process, but with arbitrary independent identically distributed jumps). Its jumps are described by the restriction of the Lévy measure $\Pi(dt)$ to $(-\infty, -1) \cup (1, \infty)$. The characteristic function $\psi_3$ of $X_3(1)$ satisfies
$$\log \psi_3(u) = \int_{\mathbb{R}} \left(e^{iut} - 1\right) \mathbb{I}_{|t|>1} \, \Pi(dt).$$

4. A pure jump martingale $X_4$ whose jumps never exceed one and are described by the restriction of the Lévy measure $\Pi(dt)$ to $[-1, 1]$; the characteristic function $\psi_4$ of $X_4(1)$ satisfies
$$\log \psi_4(u) = \int_{\mathbb{R}} \left(e^{iut} - 1 - iut\right) \mathbb{I}_{|t|\leq 1} \, \Pi(dt).$$
(In principle, a Lévy process can have more than one representation in the form (12.33), but there will always be a representation under which $X_4$ is a martingale.)

For a proof, again see Bertoin [20]. Similar decompositions are known for general processes with independent increments (see Skorohod [289]).

Similar decomposition results have recently been developed for semimartingales. For any semimartingale, it is possible to define a "triplet of predictable characteristics", analogous to the deterministic term $bt$, the Wiener term $\sigma W(t)$, and the Lévy measure $\Pi$; in the following informal discussion we will call $b$ the *drift term* and $\sigma^2$ the *diffusion term*. This representation has a clear game-theoretic interpretation: at every step Forecaster announces $b$, $\sigma^2$, and $\Pi$, Skeptic is allowed to buy corresponding tickets, and Reality chooses the increment in $X(t)$, which influences the amount received by Skeptic.

It should be rewarding to review this literature from our game-theoretic point of view, seeking to simplify results by reducing Forecaster's moves to what is actually needed. For some results, for example, it suffices for Forecaster to give only mean value and variance of a jump (defined to be 0 if there is no jump) instead of announcing a complete Lévy measure $\Pi$. We have already demonstrated this point with Theorems 4.1–5.2, where Forecaster was only required to price a few tickets associated with the next jump instead of giving the full distribution $\Pi$. Those theorems were set in discrete time, but they fit into the general framework as a special case where $X(t)$ changes only when $t$ is integer, $b$ and $\sigma^2$ are always zero, and the changes in $X(t)$ are governed by $\Pi$. The continuous-time protocols in this part of the book can also be seen as part of this general approach. In §14.1, we will consider continuous-time protocols in which Forecaster announces only a diffusion term (denoted $v_n$), omitting any drift terms. The continuous-time protocols we considered in Chapter 11 go even further, requiring only a remaining cumulative variance instead of an instantaneous variance $v_n$. The Poisson protocol of §12.3 corresponds, of course, to the case where the Lévy measure $\Pi(t)$ at step $t$ is concentrated at $\{1\}$, but in this protocol we also ask for the remaining cumulative variance. The final protocol of §12.3, which combines Bachelier's protocol with jumps, can be interpreted very roughly as an extension of Bachelier's protocol from Brownian motion to Lévy processes whose Lévy measure is concentrated on a finite set.

We have not tried to create a general game-theoretic framework involving all three characteristics for semimartingales, because we do not yet have enough interesting special cases to justify such a general construction. Instead, we have been opportunistic, developing only the mathematics needed for specific problems. When such a framework is developed, however, we expect it to be more flexible than the standard approach, with fewer technical conditions. The measure-theoretic general theory of stochastic processes is notorious for its highly technical nature; as Paul Meyer once said, "...il faut...un cours de six mois sur les définitions. Que peut on y faire?" (quoted by Williams in the first edition of [262], 1979). The game-theoretic framework should support a more accessible theory.

### Stable Lévy Processes

Under any reasonable formalization, a stochastic process obtained by summing independent identically distributed stable infinitesimal increments will be a stable Lévy process, that is, a Lévy process with stable increments. In this case, if $\alpha \neq 2$, the

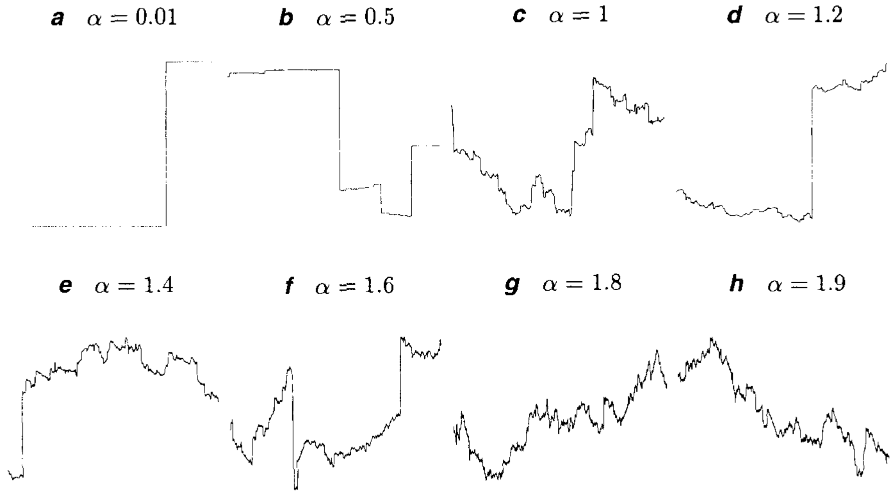



***Fig. 12.1*** Simulated price paths with increments $dS(t)$ that are stable with different values for the parameter $\alpha$.

diffusion component is absent and the Lévy measure $\Pi$ is given by

$$\Pi((-\infty, -t)) = c_1 t^{-\alpha}, \quad \Pi((t, \infty)) = c_2 t^{-\alpha}$$

($t > 0$) for some nonnegative constants $c_1$ and $c_2$. The Lévy measure is absolutely continuous with the density $c_1'(-t)^{-1-\alpha}$ on $(-\infty, 0)$ and $c_2' t^{-1-\alpha}$ on $(0, \infty)$. Its interpretation in terms of jumps makes computer modeling of stable processes quite easy.

Figure 12.1 shows some paths of stable Lévy processes with different $\alpha$. In contrast with the Wiener process (Figure 9.2), they do appear discontinuous. This is what we expect when $\alpha \in (0, 2)$. Lévy showed that the only continuous Lévy processes are diffusion processes ([262], Theorem 28.12), as we would expect from the interpretation we have given for the Lévy measure $\Pi$.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 13
# *Games for American Options*

In this chapter, we extend our purely game-theoretic approach to option pricing from European options to American options.

An American option differs from a European option in that it can be exercised early. If a European option on a security $\mathcal{S}$ pays $U(S(T))$, where $S(T)$ is the price of $\mathcal{S}$ at time $T$, then the corresponding American option pays $U(S(t))$, where $t$ is the time when the holder of the option decides to exercise it. This flexibility is attractive to investors, because it allows them to use information coming from outside the market. Consider, for example, an American call option, which allows its holder to buy a share of $\mathcal{S}$ at the strike price $\alpha$. If the current price of $\mathcal{S}$ is greater than $\alpha$, then the immediate exercise of the option would produce a net gain. If the holder of the option learns something that convinces him that the price of $\mathcal{S}$ will soon fall and stay low at least until after $T$, then he will want to exercise the option immediately.

Because of their flexibility, the pricing of American options raises some new issues. In order to establish that the option is worth no more than \$$c$ at time 0, it is not enough to show that we can reproduce a certain payoff at time $T$ starting with \$$c$. We must also show that we can reproduce the option's flexibility starting with \$$c$. This leads to a splitting of the concept of upper price. In general, we may have three prices for an American option $H$:

$$\underline{\mathbb{E}}\, H \le \overline{\mathbb{E}}\, H \le \overline{\overline{\mathbb{E}}} H. \qquad (13.1)$$

Roughly speaking, the *lower price* $\underline{\mathbb{E}}\, H$ is the largest amount of money the holder of the option can produce for certain (using the option and also trading directly in the market), the *weak upper price* $\overline{\mathbb{E}}\, H$ is the least initial capital that will enable a trader without the option to produce any payoff variable that a trader with the option can produce with no additional capital, and the *strong upper price* $\overline{\overline{\mathbb{E}}}H$ is the least initial capital that will enable a trader without the option to reproduce all the flexibility of the trader with the option.

In this chapter, we show that an American option is priced—that is, the prices in (13.1) are all equal—under certain assumptions similar to those we used for pricing European options: the dividend-paying security $\mathcal{D}$ is traded, Market obeys certain conditions on the paths $S$ and $D$, and the payoff function $U$ is well-behaved. To show this, and to find the price, we use the ideas from parabolic potential theory that we introduced in §6.3.

As an exercise in game theory, this chapter uses a more general framework than the rest of the book. Elsewhere, neither our particular games nor our abstract theory ever goes beyond protocols in which the capital of our protagonist, Skeptic or Investor, increases by increments that depend only on his current move. Here we consider more general protocols, in which Investor's move on one round can influence what he can achieve by his move on another round.

We introduce these more general protocols, which accommodate American options and many other types of options as well, in §13.1. In §13.2, we show how to construct a two-player game for comparing two of these protocols. In this game, each player follows his own protocol, except that he also observes the other player's moves. This introduces an element of competition; we can see which protocol allows its player to do better. If one of the protocols merely gives its player cash to invest in the market, then we are comparing the other protocol to a given amount of cash. This leads to the weak and strong notions of price, which we formalize in §13.3.

After all this careful groundwork, we finally show how to price American options in §13.4. That section concludes with a few examples.

## 13.1 MARKET PROTOCOLS

The protocols we studied in earlier chapters, in which the capital of Player I (Skeptic or Investor) increases by increments he influences only through his current move, are *gambling protocols* (see §8.3). We will call a protocol in which Player I's final capital is determined in a more complicated way a *market protocol*.

We begin the chapter by introducing some general notation for market protocols. Then we discuss the two examples of market protocols that interest us the most: the investor who begins owning an American option, and the investor who begins holding a certain amount of cash, say $\$c$. We assume that both investors are entitled to trade freely in the market, going short in securities if they want. Generalizing from these two examples while retaining the feature that the investor can trade freely in the market in addition to exploiting his initial endowment, we obtain the concept of a *financial instrument*. Both an American option and an endowment of $\$c$ are financial instruments. Other types of options are also financial instruments. The notation that we lay out for financial instruments in general in this section will help us define the concepts of weak and strong price in the next section.

### The Abstract Market Protocol

Here is our market protocol in its most general form. It is a discrete-time, perfect-information protocol. It generalizes the gambling protocol but retains the feature that the moves available to Player II (Market) can be affected only by his own earlier moves, not by the moves of Player I (Investor).

MARKET PROTOCOL
**Parameters:** $N$, set $\mathbf{S}$, set $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$, $\Lambda: \Omega \times \mathbf{S}^N \to \mathbb{R}$
**Players:** Investor, Market
**Protocol:**
  Market announces $\mathbf{w}_0 \in \mathbf{W}$.
  FOR $n = 1, 2, \ldots, N$:
    Investor announces $\mathbf{s}_n \in \mathbf{S}$.
    Market announces $\mathbf{w}_n \in \mathbf{W}$.
  $I := \Lambda(\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_N, \mathbf{s}_1, \ldots, \mathbf{s}_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_N) \in \Omega$.

The quantity $I$ is Investor's capital at the end of the game.

As usual, we call $\Omega$ the *sample space*, we call an element of $\Omega$ a *path*, and we call a function on $\Omega$ a *variable*. We will call an initial subsequence of a path a *market situation*, and we write $\Omega^\diamond$ for the set of all market situations.

We assume that Investor has the same move space $\mathbf{S}$ on every round only to keep the notation simple. Because a move by Investor has no meaning aside from its effect on the final capital $I$ (and this effect may depend on the round and on other moves by both players), the effective move space on each round is really defined by $\mathbf{S}$ together with the function $\Lambda$, and this may be quite different from round to round.

The notion of a market protocol is obviously extremely general. Investor may begin the game with various kinds of property, including cash, shares of stock, and options. Investor's moves may include all sorts of decisions he takes with respect to this endowment (including decisions about whether to exercise an American option), provided only that they have a well-defined effect on the final capital $I$.

This generality changes the role played by coherence. We called a gambling protocol coherent if World had a strategy guaranteeing that Skeptic ended the game with nonpositive capital. This would be inappropriate here; there is nothing incoherent about Investor starting with cash or with a promissory note guaranteeing some positive capital in the future. But we will encounter an appropriate concept of coherence for the market protocols that interest us (p. 322).

A *strategy* for Investor is a function $\mathcal{P}$ that assigns an element of $\mathbf{S}$ to every market situation. Given a strategy $\mathcal{P}$, we set

$$I^{\mathcal{P}}(\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_N) :=$$
$$\Lambda(\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N, \mathcal{P}(\mathbf{w}_0), \mathcal{P}(\mathbf{w}_0, \mathbf{w}_1), \ldots, \mathcal{P}(\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{N-1})).$$

This defines a function $I^{\mathcal{P}}$ on the sample space, which we call the *payoff variable* for the strategy $\mathcal{P}$.

### The American Option

Let us assume that an investor who holds an American option on $\mathcal{S}$ also has the opportunity to trade in $\mathcal{S}$ and whatever auxiliary securities (such as our dividend-paying security $\mathcal{D}$ or our strictly convex derivative $\mathcal{R}$) are in the market. This means that he makes two moves on each round:

- a market move—how much of each traded security to hold on that round, and
- a move intrinsic to the American option—whether to hold or exercise it.

Leaving open exactly what auxiliary securities are traded, we use the following notation for the market move:

- The move space for the market move is a linear space $\mathbf{S}$. (If Investor can trade in $\mathcal{S}$ and $\mathcal{D}$, then $\mathbf{S}$ is $\mathbb{R}^2$, and an element $(M_n, V_n)$ of $\mathbf{S}$ is interpreted as holding $M_n$ shares of $\mathcal{S}$ and $V_n$ shares of $\mathcal{D}$.)
- The payoff from a market move on the $n$th round, say $\mathbf{s}_n \in \mathbf{S}$, is given by a function $\Lambda_S : \mathbf{S} \times \Omega^{\diamond} \to \mathbb{R}$ that is linear in $\mathbf{s} \in \mathbf{S}$. (In the case where Investor is trading in $\mathcal{S}$ and $\mathcal{D}$ and Market moves by announcing new prices for $\mathcal{S}$ and $\mathcal{D}$, the payoff in the market situation $s_n$ just after announcing $S_n$ and $D_n$ is $\Lambda_S(M_n, V_n, s_n) = M_n \Delta S_n + V_n((\Delta S_n / S_{n-1})^2 + \Delta D_n)$.)

And we use the following notation for the intrinsic move:

- Given $(h_1, \ldots, h_N) \in \{\text{Exercise}, \text{Hold}\}^N$, we write $\mathbf{1stE}(h_1, \ldots, h_N)$ for the first $n$ for which $h_n = \text{Exercise}$; if $h_n = \text{Hold}$ for all $n$, then we set $\mathbf{1stE}(h_1, \ldots, h_N) = N$.
- We write $\underline{U}(\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_\tau)$ for the payoff of the option when it is exercised on round $\tau$. (Normally this is $U(S_\tau)$, where $U$ is the payoff function for the option.)

With this notation, the market protocol for the American option reads as follows:

MARKET PROTOCOL FOR AN AMERICAN OPTION
**Market Parameters:** $N, \mathbf{W}, \Omega \subseteq \mathbf{W}^{N+1}, \mathbf{S}, \Lambda_S : \mathbf{S} \times \mathbf{W}^* \to \mathbb{R}$
**Instrument Parameter:** $U : \mathbf{W}^* \to \mathbb{R}$
**Players:** Investor, Market
**Protocol:**

$\mathcal{I}_0 := 0$.
Market announces $\mathbf{w}_0 \in \mathbf{W}$.
FOR $n = 1, 2, \ldots, N$:
  Investor announces $h_n \in \{\text{Exercise}, \text{Hold}\}$ and $\mathbf{s}_n \in \mathbf{S}$.
  Market announces $\mathbf{w}_n \in \mathbf{W}$.
  $\mathcal{I}_n = \mathcal{I}_{n-1} + \Lambda_S(\mathbf{s}_n, \mathbf{w}_0, \ldots, \mathbf{w}_n)$.
$I := \mathcal{I}_N + \underline{U}(\mathbf{w}_0, \ldots, \mathbf{w}_\tau)$, where $\tau = \mathbf{1stE}(h_1, \ldots, h_N)$.

**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \ldots, \mathbf{w}_N) \in \Omega$.

Investor's total capital $I$ at the end of the game is obtained by adding his capital from trading in the market, $\mathcal{I}_N$, to his payoff from the option.

## Cash

If a holder of $\$c$ is also allowed to trade in the same securities as the holder of the American option, then his market protocol can be formalized as follows:

MARKET PROTOCOL FOR $\$c$
**Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$, $\mathbf{S}$, $\Lambda_S: \mathbf{S} \times \mathbf{W}^* \to \mathbb{R}$, $c$
**Players:** Cash, Market
**Protocol:**
 $\mathcal{I}_0 = 0$.
 Market announces $\mathbf{w}_0 \in \mathbf{W}$.
 FOR $n = 1, 2, \ldots, N$:
  Cash announces $\mathbf{s}_n \in \mathbf{S}$.
  Market announces $\mathbf{w}_n \in \mathbf{W}$.
  $\mathcal{I}_n = \mathcal{I}_{n-1} + \Lambda_S(\mathbf{s}_n, \mathbf{w}_0, \ldots, \mathbf{w}_n)$.
 $I := \mathcal{I}_N + c$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \ldots, \mathbf{w}_N) \in \Omega$.

As usual, we assume that $\mathbf{S}$ is a linear space and that $\Lambda_S$ is linear in its first argument.

## The Concept of a Financial Instrument

In order to have a general framework for comparing cash and the American option, we now introduce the concept of a *financial instrument*. A financial instrument $H$, in the sense in which we shall use the term in this chapter, is defined by the parameters $\mathbf{H}$ and $\Lambda_H: \Omega \times \mathbf{H}^N \to \mathbb{R}$ in the following protocol. A holder of a financial instrument has the right to play the role of Investor in the protocol with these parameters.

MARKET PROTOCOL FOR FINANCIAL INSTRUMENT $H$
**Market Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$, $\mathbf{S}$, $\Lambda_S: \mathbf{S} \times \mathbf{W}^* \to \mathbb{R}$
**Instrument Parameters:** $\mathbf{H}$, $\Lambda_H: \Omega \times \mathbf{H}^N \to \mathbb{R}$
**Players:** Investor, Market
**Protocol:**
 $\mathcal{I}_0 := 0$.
 Market announces $\mathbf{w}_0 \in \mathbf{W}$.
 FOR $n = 1, 2, \ldots, N$:
  Investor announces $h_n \in \mathbf{H}$ and $\mathbf{s}_n \in \mathbf{S}$.
  Market announces $\mathbf{w}_n \in \mathbf{W}$.
  $\mathcal{I}_n = \mathcal{I}_{n-1} + \Lambda_S(\mathbf{s}_n, \mathbf{w}_0, \ldots, \mathbf{w}_n)$.
 $I := \mathcal{I}_N + \Lambda_H(\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_N, h_1, \ldots, h_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \ldots, \mathbf{w}_N) \in \Omega$.

We again assume that $\mathbf{S}$ is a linear space and that $\Lambda_S$ is linear in its first argument. We call $h_n$ Investor's *intrinsic move*; it is his action, if any, with respect to the instrument. We call $\mathbf{s}_n$ his *market move*; it tells how many traded securities (such as shares of $\mathcal{S}$ and its traded derivatives) he holds on round $n$.

The market for securities is defined by its sample space $\Omega$ and the function $\Lambda_S$ with which it rewards the positions taken by investors in traded securities, while the instrument is defined by its move space $\mathbf{H}$ and its payoff $\Lambda_H$.

A strategy for Investor can be decomposed into a *market strategy* $\mathcal{P}$, which tells him how to choose his market moves, and an *intrinsic strategy* $\mathcal{H}$, which tells him how to choose his intrinsic moves. These strategies each determine a payoff variable; the payoff $\mathcal{I}^{\mathcal{P}}$ from the market strategy is given by

$$\mathcal{I}^{\mathcal{P}}(\mathbf{w}_0, \dots, \mathbf{w}_N) := \sum_{n=1}^{N} \Lambda_S(\mathcal{P}(\mathbf{w}_0, \dots, \mathbf{w}_{n-1}), \mathbf{w}_0, \dots, \mathbf{w}_n)$$

and the payoff $\mathcal{I}^{\mathcal{H}}$ from the intrinsic strategy is given by

$$\mathcal{I}^{\mathcal{H}}(\mathbf{w}_0, \dots, \mathbf{w}_N) := \Lambda_H(\mathbf{w}_0, \dots, \mathbf{w}_N, \mathcal{H}(\mathbf{w}_0), \dots, \mathcal{H}(\mathbf{w}_0, \dots, \mathbf{w}_{N-1})).$$

The total payoff is $\mathcal{I}^{\mathcal{H}} + \mathcal{I}^{\mathcal{P}}$.

If Market can guarantee that Investor's gain $\mathcal{I}_N$ from trading in the market is nonpositive, then we say that *the market is coherent*.

### Passive Instruments

We say that a financial instrument $H$ is *passive* if Investor cannot do anything to affect its payoff. Cash and European options are passive; American options are not. We can formalize this by saying that the move space $\mathbf{H}$ has only one element, a move called "do nothing" that has no effect on the function $\Lambda_H$. For a cleaner protocol, we can simply delete the intrinsic moves:

MARKET PROTOCOL FOR A PASSIVE INSTRUMENTS $H$
**Market Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$, $\mathbf{S}$, $\Lambda_S : \mathbf{S} \times \mathbf{W}^* \to \mathbb{R}$
**Instrument Parameters:** $\Lambda_H : \Omega \to \mathbb{R}$
**Players:** Investor, Market
**Protocol:**
    $\mathcal{I}_0 := 0$.
    Market announces $\mathbf{w}_0 \in \mathbf{W}$.
    FOR $n = 1, 2, \dots, N$:
        Investor announces $\mathbf{s}_n \in \mathbf{S}$.
        Market announces $\mathbf{w}_n \in \mathbf{W}$.
        $\mathcal{I}_n = \mathcal{I}_{n-1} + \Lambda_S(\mathbf{s}_n, \mathbf{w}_0, \dots, \mathbf{w}_n)$.
    $I := \mathcal{I}_N + \Lambda_H(\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \dots, \mathbf{w}_N) \in \Omega$.

Again, $\mathbf{S}$ is a linear space, and $\Lambda_S$ is linear in its first argument. Here Investor only makes market moves, and so a strategy for him is merely a market strategy.

### Other Options

Although we are mainly concerned with American options in this chapter, our concept of a market protocol is broad enough to accommodate almost all types of options, including the following types of non-European options.

**Asian option.** An "average price call" on a security $S$ with maturity $N$ and strike price $c$ pays at $N$ the amount by which the average price of $S$ over the period $\{1, \ldots, N\}$ exceeds $c$.

**Make-your-mind-up option.** This is a variant of the American option; the option can be exercised early, but the holder must give notice a certain number of days in advance that he will exercise it on a given day.

**Bermudan option.** This is another variant on the American option, in which exercise is only allowed on days in a specified subset of the $N$ trading days.

**Shout option.** This type of option combines the advantages of European and American options by allowing the holder, on one occasion of his choosing, to lock in the gain he would obtain by the immediate exercise of an American option, without losing the possibility of doing even better by waiting until maturity. A simple shout call option on $S$ with strike price $c$ and maturity $N$, for example, pays

$$(S_\tau - c) + \max(S_N - S_\tau, 0),$$

where $\tau$ is round where the holder shouts. Only one shout is allowed, on a round $n$ where $S_n - c > 0$.

**Chooser option.** Here the holder of the option has the right to choose, at some specified time $n$, whether it will be a call or a put option, which can then be exercised only at maturity.

**Passport option.** Also called a *perfect trader option*, this option permits the holder to trade in the underlying security $S$, under some restrictions (for example, he may be allowed to hold only between $-C$ and $C$ shares of $S$ on any round), without risking a net loss; the option pays the profit from the trading in $S$ if it is positive, zero otherwise ([351], p. 243).

Asian options are passive; the other types in this list are active. For reviews of different types of options and other financial instruments, see [154, 351].

## 13.2 COMPARING FINANCIAL INSTRUMENTS

In this section, we define weak and strong ways of comparing two financial instruments. The weak comparison takes into account only the possible payoff variables that each instrument can produce, while the strong comparison also takes into account the extent to which the instrument's holder can change goals in the course of the game.

Two different financial instruments in the same market, say Instrument $A$ and Instrument $B$, can be represented by market protocols with the same moves for Market but different moves and payoffs for Investor:

MARKET PROTOCOL FOR INSTRUMENT $A$
**Market Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$
**Instrument Parameters:** $\mathbf{A}$, $\Lambda_A: \Omega \times \mathbf{A}^N \to \mathbb{R}$
**Players:** Investor A, Market
**Protocol:**
Market announces $\mathbf{w}_0 \in \mathbf{W}$.
FOR $n = 1, 2, \ldots, N$:
Investor A announces $\mathbf{a}_n \in \mathbf{A}$.
Market announces $\mathbf{w}_n \in \mathbf{W}$.
$I := \Lambda_A(\mathbf{w}_0, \ldots, \mathbf{w}_N, \mathbf{a}_1, \ldots, \mathbf{a}_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \ldots, \mathbf{w}_N) \in \Omega$.

MARKET PROTOCOL FOR INSTRUMENT $B$
**Market Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$
**Instrument Parameters:** $\mathbf{B}$, $\Lambda_B: \Omega \times \mathbf{B}^N \to \mathbb{R}$
**Players:** Investor B, Market
**Protocol:**
Market announces $\mathbf{w}_0 \in \mathbf{W}$.
FOR $n = 1, 2, \ldots, N$:
Investor B announces $\mathbf{b}_n \in \mathbf{B}$.
Market announces $\mathbf{w}_n \in \mathbf{W}$.
$I := \Lambda_B(\mathbf{w}_0, \ldots, \mathbf{w}_N, \mathbf{b}_1, \ldots, \mathbf{b}_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \ldots, \mathbf{w}_N) \in \Omega$.

These two protocols are simplified versions of the market protocol for financial instrument $H$ in the previous section; the moves $\mathbf{a}_n \in \mathbf{A}$ and $\mathbf{b}_n \in \mathbf{B}$ now include both the intrinsic moves with respect to the instruments $A$ and $B$, respectively, and the market moves.

We can compare what the two instruments can achieve by comparing the payoff variables that the investors can produce. We say that Instrument $A$ *weakly super-replicates* Instrument $B$ if for every strategy $\mathcal{Q}$ for Investor B, there is a strategy $\mathcal{P}$ for Investor A such that $I^{\mathcal{P}} \geq I^{\mathcal{Q}}$. The relation of weak super-replication is obviously transitive.

We call this kind of super-replication *weak* because it does not capture everything that investors consider when they compare financial instruments. If the investor intends to follow a strategy that he fixes at the beginning of the game, then it is appropriate for him to consider only the payoff variables that an instrument makes available. But a strategy determines each move based only on the preceding moves by the market, and investors often use other information—information from outside the market—when deciding on their moves. If an investor intends to use outside information that he acquires after the game begins, then he will be interested not only in what payoff variables are available at the beginning of the game but also in the flexibility he has to change from one payoff variable to another in the course of the game.

One way to take this flexibility into account in the comparison of two instruments is to play them against each other in what we call a *super-replication game*:

GAME FOR SUPER-REPLICATING INSTRUMENT $B$ WITH INSTRUMENT $A$
**Market Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$
**Parameters for Instrument** $A$**:** $\mathbf{A}$, $\Lambda_A: \Omega \times \mathbf{A}^N \to \mathbb{R}$
**Parameters for Instrument** $B$**:** $\mathbf{B}$, $\Lambda_B: \Omega \times \mathbf{B}^N \to \mathbb{R}$
**Players:** Investor A, (Investor B + Market)
**Protocol:**
Market announces $\mathbf{w}_0 \in \mathbf{W}$.
FOR $n = 1, 2, \ldots, N$:
Investor B announces $\mathbf{b}_n \in \mathbf{B}$.
Investor A announces $\mathbf{a}_n \in \mathbf{A}$.
Market announces $\mathbf{w}_n \in \mathbf{W}$.
$I_A := \Lambda_A(\mathbf{w}_0, \ldots, \mathbf{w}_N, \mathbf{a}_1, \ldots, \mathbf{a}_N)$.
$I_B := \Lambda_B(\mathbf{w}_0, \ldots, \mathbf{w}_N, \mathbf{b}_1, \ldots, \mathbf{b}_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \ldots, \mathbf{w}_N) \in \Omega$.
**Winner:** Investor A wins if $I_A \geq I_B$. Otherwise Investor B and Market win.

We allow Investor A to move after Investor B, with knowledge of how Investor B has moved, because we are asking whether Instrument $A$ gives at least as much flexibility as Instrument $B$. We are asking whether Investor A, using Instrument $A$, can always match or improve on what Investor B, using Instrument $B$, has achieved. If we were interested instead in which investor had the best outside information, then we might look at a game in which each must make his $n$th move without knowing the other's $n$th move.

We say that Instrument $A$ *strongly super-replicates* Instrument $B$ if Investor A has a winning strategy in the game of super-replicating Instrument $B$ with Instrument $A$. This relation is obviously transitive.

It is also obvious that strong super-replication implies weak super-replication. The following example, in which two investors are trying to predict $\mathbf{w}_2$, shows that weak super-replication does not imply strong super-replication.

EXAMPLE OF GAME FOR SUPER-REPLICATING $B$ WITH $A$
**Players:** Investor A, Investor B, Market
**Protocol:**
On round 1:
Investor A announces $\mathbf{a}_1: \{0,1\} \to \{0,1\}$.
Market announces $\mathbf{w}_1 \in \{0,1\}$.
On round 2:
Investor B announces $\mathbf{b}_2 \in \{0,1\}$.
Market announces $\mathbf{w}_2 \in \{0,1\}$.

$$I_A := \begin{cases} 1 & \text{if } \mathbf{a}_1(\mathbf{w}_1) = \mathbf{w}_2 \\ 0 & \text{otherwise.} \end{cases}$$
$$I_B := \begin{cases} 1 & \text{if } \mathbf{b}_2 = \mathbf{w}_2 \\ 0 & \text{otherwise.} \end{cases}$$

**Winner:** Investor A wins if $I_A \geq I_B$.

In principle, both investors move on both rounds. But $\mathbf{b}_1$ does not affect $I_B$, and $\mathbf{a}_2$ does not affect $I_A$, and so we have simplified the protocol by omitting them. To see that Instrument $A$ weakly super-replicates Instrument $B$, notice that a strategy $\mathcal{P}$ for Investor B produces the payoff variable

$$I_B^{\mathcal{P}} = \begin{cases} 1 & \text{if } \mathcal{P}(\mathbf{w}_1) = \mathbf{w}_2 \\ 0 & \text{otherwise,} \end{cases}$$

which Investor A can match merely by choosing his first move $\mathbf{a}_1$ to agree with $\mathcal{P}$: $\mathbf{a}_1(0) = \mathcal{P}(0)$ and $\mathbf{a}_1(1) = \mathcal{P}(1)$. On the other hand, Investor A does not have a winning strategy in the super-replication game. After Investor A has made his first move $\mathbf{a}_1$ and Market has made his first move $\mathbf{w}_1$, Investor B can always make $\mathbf{b}_2$ different from $\mathbf{a}_1(\mathbf{w}_1)$, and Market can then make $\mathbf{w}_2$ equal to $\mathbf{b}_2$, resulting in a final capital of 0 for Investor A and 1 for Investor B. So Instrument $A$ clearly does not strongly super-replicate Instrument $B$.

If $\mathbf{B}$ has just one element ("do nothing"), then there is no difference between strongly and weakly super-replicating Instrument $B$; because it is known in advance that the holder of $B$ will not do anything, it makes no difference whether the holder of Instrument $A$ gets to watch what he is doing. A similar conclusion can be drawn about passive instruments (whose owners are free to choose the market moves): there is no difference between strongly and weakly super-replicating a passive instrument with another instrument. But to see that this is true, we need to look a little closer at the relevant super-replication game.

GAME FOR SUPER-REPLICATING PASSIVE $H$ WITH $G$

**Market Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$, $\mathbf{S}$, $\Lambda_S: \mathbf{S} \times \mathbf{W}^* \to \mathbb{R}$

**Parameters for Instrument** $G$**:** $\mathbf{G}$, $\Lambda_G: \Omega \times \mathbf{G}^N \to \mathbb{R}$

**Parameters for Instrument** $H$**:** $\Lambda_H: \Omega \to \mathbb{R}$

**Players:** Investor G, (Investor H + Market)

**Protocol:**

$\mathcal{I}_0^G := 0$.

$\mathcal{I}_0^H := 0$.

Market announces $\mathbf{w}_0 \in \mathbf{W}$.

FOR $n = 1, 2, \ldots, N$:

Investor H announces $\mathbf{s}_n^H \in \mathbf{S}$.

Investor G announces $g_n \in \mathbf{G}$ and $\mathbf{s}_n^G \in \mathbf{S}$.

Market announces $\mathbf{w}_n \in \mathbf{W}$.

$\mathcal{I}_n^G = \mathcal{I}_{n-1}^G + \Lambda_S(\mathbf{s}_n^G, \mathbf{w}_0, \ldots, \mathbf{w}_n)$.

$\mathcal{I}_n^H = \mathcal{I}_{n-1}^H + \Lambda_S(\mathbf{s}_n^H, \mathbf{w}_0, \ldots, \mathbf{w}_n)$.

$I_G := \mathcal{I}_N^G + \Lambda_G(\mathbf{w}_0, \dots, \mathbf{w}_N, g_1, \dots, g_N)$.
$I_H := \mathcal{I}_N^H + \Lambda_H(\mathbf{w}_0, \dots, \mathbf{w}_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \dots, \mathbf{w}_N) \in \Omega$.
**Winner:** Investor G wins if $I_G \geq I_H$. Otherwise Investor H and Market win.

As usual, we assume that $\mathbf{S}$ is a linear space and that $\Lambda_S$ is linear in its first argument.

Notice that Investor G can exactly match whatever Investor H accomplishes by trading in the market: he simply makes $\mathbf{s}_n^G$ the same as $\mathbf{s}_n^H$ on every round $n$. Because of the linearity of market trading, he can first do this and then make whatever other market moves he wants. So Investor H's ability to trade in the market makes no difference to whether Investor G has a winning strategy: Investor G has a winning strategy in this game if and only if he has a winning strategy in the following simpler game, where Investor H does not trade:

SIMPLIFIED GAME FOR SUPER-REPLICATING PASSIVE $H$ WITH $G$
**Market Parameters:** $N$, $\mathbf{W}$, $\Omega \subseteq \mathbf{W}^{N+1}$, $\mathbf{S}$, $\Lambda_S: \mathbf{S} \times \mathbf{W}^* \to \mathbb{R}$
**Parameters for Instrument $G$:** $\mathbf{G}$, $\Lambda_G: \Omega \times \mathbf{G}^N \to \mathbb{R}$
**Parameters for Instrument $H$:** $\Lambda_H: \Omega \times \mathbf{B}^N \to \mathbb{R}$
**Players:** Investor G, Market
**Protocol:**
$\mathcal{I}_0^G := 0$.
Market announces $\mathbf{w}_0 \in \mathbf{W}$.
FOR $n = 1, 2, \dots, N$:
Investor G announces $g_n \in \mathbf{G}$ and $\mathbf{s}_n^G \in \mathbf{S}$.
Market announces $\mathbf{w}_n \in \mathbf{W}$.
$\mathcal{I}_n^G = \mathcal{I}_{n-1}^G + \Lambda_S(\mathbf{s}_n^G, \mathbf{w}_0, \dots, \mathbf{w}_n)$.
$I_G := \mathcal{I}_N^G + \Lambda_G(\mathbf{w}_0, \dots, \mathbf{w}_N, g_1, \dots, g_N)$.
$I_H := \Lambda_H(\mathbf{w}_0, \dots, \mathbf{w}_N)$.
**Additional Constraint on Market:** Market must make $(\mathbf{w}_0, \dots, \mathbf{w}_N) \in \Omega$.
**Winner:** Investor G wins if $I_G \geq I_H$. Otherwise Market wins.

With this protocol in view, we easily obtain the following proposition.

**Proposition 13.1** *Suppose G and H are instruments, H is passive, and G weakly super-replicates H. Then G strongly super-replicates H.*

*Proof* The hypothesis says that for every strategy $\mathcal{P}$ for Investor H, there is a market strategy $\mathcal{Q}$ and an intrinsic strategy $\mathcal{G}$ for Investor G such that $\mathcal{I}^{\mathcal{G}} + \mathcal{I}^{\mathcal{Q}} \geq \mathcal{I}^{\mathcal{P}} + \Lambda_H$. Because market strategies form a linear space and their payoff variables combine linearly, this can be written as $\mathcal{I}^{\mathcal{G}} + \mathcal{I}^{\mathcal{Q}-\mathcal{P}} \geq \Lambda_H$. In fact, we can omit mention of $\mathcal{P}$: the condition is simply that Investor G have a market strategy $\mathcal{Q}$ and an intrinsic strategy $\mathcal{G}$ such that $\mathcal{I}^{\mathcal{G}} + \mathcal{I}^{\mathcal{Q}} \geq \Lambda_H$. This is obviously equivalent to having a winning strategy in the simplified game for super-replicating $H$. ∎

## 13.3 WEAK AND STRONG NOTIONS OF PRICE

In this section we formulate weak and strong notions of price, corresponding to the weak and strong notions of super-replication, for financial instruments.

Intuitively, pricing a financial instrument $H$ is a matter of comparing it with cash. If an investor can always achieve as much with the instrument as he can with $\$c$, then the instrument is worth at least $c$; if he can always achieve as much with $\$c$ as he can with the instrument, then the instrument is worth no more than $c$.

Because $\$c$ is a passive instrument, there is no difference between weakly and strongly super-replicating it with another instrument. So we may define $H$'s lower price by

$$\underline{\mathbb{E}}\,H := \sup\{c \in \mathbb{R} \mid H \text{ weakly super-replicates } \$c\} = \sup\{c \in \mathbb{R} \mid H \text{ strongly super-replicates } \$c\}.$$

On the other hand, there may be a difference between weakly and strongly super-replicating $H$, even with cash, and so we must consider strong and weak versions of upper price. The weak upper price is

$$\overline{\mathbb{E}}\,H := \inf\{c \in \mathbb{R} \mid \$c \text{ weakly super-replicates } H\},$$

and the strong upper price is

$$\overline{\overline{\mathbb{E}}}H := \inf\{c \in \mathbb{R} \mid \$c \text{ strongly super-replicates } H\}.$$

Because strong super-replication implies weak super-replication, $\overline{\mathbb{E}}\,H \le \overline{\overline{\mathbb{E}}}H$.

Using the coherence of the market, we see that lower price cannot exceed weak upper price:

**Proposition 13.2** $\underline{\mathbb{E}}\,H \le \overline{\mathbb{E}}\,H$.

*Proof* If $\underline{\mathbb{E}}\,H > \overline{\mathbb{E}}\,H$, then there exist real numbers $c$ and $d$ such that $c < d$, $\$c$ weakly super-replicates $H$, and $H$ weakly super-replicates $\$d$. This implies that $\$c$ weakly super-replicates $\$d$. So given a strategy $\mathcal{P}$ for trading in the market, there is a strategy $\mathcal{Q}$ for trading in the market such that $\mathcal{I}^{\mathcal{Q}} + c \ge \mathcal{I}^{\mathcal{P}} + d$. This implies $\mathcal{I}^{\mathcal{Q}-\mathcal{P}} \ge d - c$, contradicting the coherence of the market. ∎

So we have the relation (13.1) mentioned in the introduction to this chapter:

$$\underline{\mathbb{E}}\,H \le \overline{\mathbb{E}}\,H \le \overline{\overline{\mathbb{E}}}H.$$

As usual, we speak simply of price when upper and lower prices coincide:

- When $\underline{\mathbb{E}}\,H$ and $\overline{\mathbb{E}}\,H$ coincide, their common value is the *weak price* for $H$.
- When $\underline{\mathbb{E}}\,H$ and $\overline{\overline{\mathbb{E}}}H$ coincide, their common value is the *strong price* for $H$.

When the strong price exists, the weak price also exists and has the same value.

We will be mainly interested in $\underline{\mathbb{E}}_0 H$, $\overline{\mathbb{E}}_0 H$, and $\overline{\overline{\mathbb{E}}}_0 H$, which are defined in exactly the same way as $\underline{\mathbb{E}}\, H$, $\overline{\mathbb{E}}\, H$, and $\overline{\overline{\mathbb{E}}} H$ but starting in the situation 0 where $\mathbf{w}_0$ has just been announced.

If $H$ is a European option or some other passive instrument, super-replicating it strongly is the same as super-replicating it weakly, and therefore there will be no difference between its weak or strong upper price. Moreover, as the following proposition makes clear, our new concepts of upper and lower price agree, in the case of European options, with the concepts of upper and lower price that we used in preceding chapters.

**Proposition 13.3** *If $\mathcal{U}$ is a European option with payoff function $U$, then $\underline{\mathbb{E}}_0 \mathcal{U} = \underline{\mathbb{E}}_0(U(S_N))$ and $\overline{\mathbb{E}}_0 \mathcal{U} = \overline{\mathbb{E}}_0(U(S_N))$, 0 being the situation where $\mathbf{w}_0$ has just been announced.*

*Proof* Reasoning as in the proof of Proposition 13.2, we see that \$$c$ weakly replicates $H$ if and only if there is a market strategy $\mathcal{P}$ such that $c + \mathcal{I}^{\mathcal{P}}(\mathbf{w}_0, \ldots, \mathbf{w}_N) \geq U(S_N)$ for all $\mathbf{w}_1, \ldots, \mathbf{w}_N$. This is same as saying that Investor in the protocol for trading in $\mathcal{S}$ alone can super-replicate $U$ if he starts with capital $c$. So $\underline{\mathbb{E}}_0 \mathcal{U} = \underline{\mathbb{E}}_0(U(S_N))$.

Similarly, $H$ weakly replicates \$$c$ if and only if there is a market strategy $\mathcal{P}$ such that $U(S_N) + \mathcal{I}^{\mathcal{P}}(\mathbf{w}_0, \ldots, \mathbf{w}_N) \geq c$ for all $\mathbf{w}_1, \ldots, \mathbf{w}_N$. This is same as saying that Investor in the protocol for trading in $\mathcal{S}$ alone can super-replicate $-U$ if he starts with capital $-c$. So $\overline{\mathbb{E}}_0 \mathcal{U} = \overline{\mathbb{E}}_0(U(S_N))$. ∎

## 13.4 PRICING AN AMERICAN OPTION

We now consider the problem of pricing an American option on a security $\mathcal{S}$ with payoff function $U$, assuming that the dividend-paying security $\mathcal{D}$ is traded and that Market makes the paths $S$ and $D$ satisfy requirements similar to those that we used for pricing European options in Chapter 11.

We begin by writing down the games for super-replicating the American option with \$$c$ and vice versa. We use the same discrete-time notation as in the preceding sections, but now $N$ should be interpreted as an infinitely large number and the arithmetic should be understood in terms of nonstandard analysis.

Because \$$c$ is passive, we can use the simplified protocol on p. 327 for super-replicating \$$c$ with the American option. Because Investor can never be sure to make money trading in the market, his strategy for being sure to have \$$c$ must accomplish this by the time $\tau$ he exercises the option, and so we can simplify the notation by ending the game at that point. So here is our game:

GAME FOR SUPER-REPLICATING \$$c$ WITH THE AMERICAN OPTION
**Market Parameters:** $N, C > 0$
**Parameter for the American Option:** $U\colon (0, \infty) \to \mathbb{R}$
**Parameter for \$$c$:** $c$
**Players:** Investor, Market
**Protocol:**

$\mathcal{I}_0 := 0$.
Market announces $S_0 > 0$ and $D_0 > 0$.
FOR $n = 1, 2, \ldots, N$:
Investor announces $h_n \in \{\text{Exercise}, \text{Hold}\}$, $M_n \in \mathbb{R}$, and $V_n \in \mathbb{R}$.
Market announces $S_n > 0$ and $D_n \geq 0$.
$\mathcal{I}_n := \mathcal{I}_{n-1} + M_n \Delta S_n + V_n \left( (\Delta S_n / S_{n-1})^2 + \Delta D_n \right)$.
$\tau := n$.
EXIT if $h_n = \text{Exercise}$.
**Additional Constraints on Market:** Market must ensure that $S$ is continuous, $1/C < S_n < C$ for all $n$, and $\operatorname{vex} S \leq 2$. He must also ensure that $D$ is continuous, $0 < D_n < C$ for $n = 1, \ldots, N-1$, $D_N = 0$, and $\operatorname{vex} D < 2$.
**Winner:** Investor if $\mathcal{I}_\tau + U(S_\tau) \geq c$.

In the game for super-replicating the American option with $\$c$, we cannot simplify as much, but we can, for the same reasons as in the first game, end the game after the round where the option is exercised and eliminate the market trading by the player who moves first, the holder of the option in this case. This gives the following somewhat simplified game.

GAME FOR SUPER-REPLICATING THE AMERICAN OPTION WITH $\$c$
**Market Parameters:** $N$, $C > 0$
**Parameter for $\$c$:** $c$
**Parameter for the American Option:** $U\colon (0, \infty) \to \mathbb{R}$
**Players:** Cash, (Investor + Market)
**Protocol:**
$\mathcal{I}_0^C := 0$.
Market announces $S_0 > 0$ and $D_0 > 0$.
FOR $n = 1, 2, \ldots, N$:
Investor announces $h_n \in \{\text{Exercise}, \text{Hold}\}$.
Cash announces numbers $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Market announces $S_n > 0$ and $D_n \geq 0$.
$\mathcal{I}_n^C := \mathcal{I}_{n-1}^C + M_n \Delta S_n + V_n \left( (\Delta S_n / S_{n-1})^2 + \Delta D_n \right)$.
$\tau := n$.
EXIT if $h_n = \text{Exercise}$.
**Additional Constraints on Market:** Market must ensure that $S$ is continuous, $1/C < S_n < C$ for all $n$, and $\operatorname{vex} S \leq 2$. He must also ensure that $D$ is continuous, $0 < D_n < C$ for $n = 1, \ldots, N-1$, $D_N = 0$, and $\operatorname{vex} D < 2$.
**Winner:** Cash if $\mathcal{I}_\tau^C + c \geq U(S_\tau)$.

In order to show that $\$c$ is the strong price of the American option, we must show that there is a winning strategy in both games—a winning strategy for Investor in the first game and for Cash in the second game.

In order to identify the price $\$c$ for which such strategies exist, we need notions of parabolic potential theory similar to those used in Chapter 6 but for a nonuniform medium. (The Black-Scholes equation is the heat equation corresponding to a medium with density proportional to $s^{-2}$.)

There are several possible approaches. We could develop parabolic potential theory for general parabolic differential equations, but this would be time-consuming: most standard references (such as Doob [103] and Constantinescu and Cornea [55]) only discuss the simplest heat equation. Alternatively, we could exploit the stochastic interpretation, replacing Brownian motion with a stochastic process defined by the stochastic differential equation

$$dS(t) = S(t)dW(t). \tag{13.2}$$

The simplest approach, however, is to reduce the Black-Scholes equation to the heat equation: a smooth function $u(s, D)$ satisfies the heat equation

$$\frac{\partial u}{\partial D} = \frac{1}{2}\frac{\partial^2 u}{\partial s^2}$$

if and only if the function

$$\tilde{u}(s, D) = u\left(\ln s - \frac{D}{2}, D\right)$$

satisfies the Black-Scholes equation

$$\frac{\partial \tilde{u}}{\partial D} = \frac{1}{2}s^2\frac{\partial^2 \tilde{u}}{\partial s^2}$$

(see (13.4) below). We might remark that this change of variables has a stochastic interpretation. When $W(t)$ is a Brownian motion, $S(t) := e^{W(t)-t/2}$ satisfies, by Itô's formula, the stochastic differential equation (13.2). Since $D$ plays the role of remaining volatility, $D = T - t$, the transformation $s \mapsto \ln s - D/2$ will transform a solution to (13.2) into a Brownian motion.

We say that a function $u(s, D)$, $(s, D) \in (0, \infty)^2$, is *supermarket* if the function $(S, D) \in \mathbb{R} \times (0, \infty) \mapsto u\left(e^{S+D/2}, D\right)$ is superparabolic (as defined in Chapter 6). For any function $u : (0, \infty)^2 \to \mathbb{R}$, we define LSM $u$ to be the Least SuperMarket majorant of $u$, if it exists.

Now we can state our theorem for pricing American options, which is analogous to Theorem 11.2 (p. 280) for pricing European options.

**Theorem 13.1** *Suppose the payoff function $U$: $(0, \infty) \to \mathbb{R}$ is continuous and grows only at most polynomially fast as $s \to 0$ and $s \to \infty$:*

$$\exists k \in \{1, 2, \ldots\} : \lim_{s\to\infty} |U(s)|s^{-k} = 0 \ \&\ \lim_{s\to 0} |U(s)|s^k = 0. \tag{13.3}$$

*Define $u : (0, \infty)^2 \to \mathbb{R}$ by $u(s, D) := U(s)$. Then* LSM $u$ *exists, and in the situation where $S(0)$ and $D(0)$ have just been announced,*

$$(\text{LSM}\, u)(S(0), D(0))$$

*is the strong price for the American option with payoff $U$ and maturity $T$, under the constraints on Market given in our super-replication protocols.*

*Proof* This proof is modeled on the proof of the one-sided central limit theorem in Chapter 6.

Along with every function $f(s,D)$, $f:(0,\infty)^2 \to \mathbb{R}$, we will also consider the function $\tilde{f}(S,D) := f(s,D)$ obtained from $f$ by the change of variable $s = e^{S+D/2}$. As we have already mentioned, when smooth, $f(s,D)$ satisfies the Black-Scholes equation if and only if $\tilde{f}(S,D)$ satisfies the standard heat equation. To see why this is true and for future reference, we give the connections between partial derivatives of $f(s,D)$, $(s,D) \in (0,\infty)^2$, and $\tilde{f}(S,D)$:

$$\frac{\partial \tilde{f}}{\partial D} = \frac{\partial f}{\partial D} + \frac{1}{2}s\frac{\partial f}{\partial s}, \quad \frac{\partial^2 \tilde{f}}{\partial S^2} = s^2\frac{\partial^2 f}{\partial s^2} + s\frac{\partial f}{\partial s}. \tag{13.4}$$

As before, we will also continue to use tilde for the opposite operation: if $f(S,D)$ is a function of the type $\mathbb{R} \times (0,\infty) \to \mathbb{R}$, rather then $(0,\infty)^2 \to \mathbb{R}$, $\tilde{f}:(0,\infty)^2 \to \mathbb{R}$ is defined by the equation $\tilde{f}(s,D) = f(S,D)$, where $S = \ln s - D/2$. To follow the proof of the theorem, it is helpful to split the change of variable $(s,D) \mapsto (S,D)$ into two steps: first the taking of the logarithm $(s,D) \mapsto (S',D)$, where $S' := \ln s$, and then the shearing $(S',D) \mapsto (S' - D/2, D)$.

We say that a function $u$ is *submarket* if $-u$ is supermarket; $u$ is *market* if it is both supermarket and submarket. We know from Chapter 6 that being both superparabolic and subparabolic is equivalent to satisfying the standard heat equation for locally bounded functions; therefore, locally bounded market functions satisfy the Black-Scholes equation.

The function $\tilde{u}(S,D)$ has the least superparabolic majorant $\mathrm{LM}\,\tilde{u}$; it is clear that $\mathrm{LSM}\,u$ will exist and coincide with $\widetilde{\mathrm{LM}\,\tilde{u}}$.

Let $c$ be a large constant (in particular, $c \gg C$; other requirements will be added later). We can (and will) assume that $U = 0$ outside the interval $1/c \le s \le c$; indeed:

- $S(t)$ will never reach that interval.
- Changing $U$ outside that interval will not change $(\mathrm{LSM}\,u)(S_0, D_0)$ much since:
  - for any constant $\gamma \in \mathbb{R}$, the function

    $$s^\gamma \exp\left(\frac{\gamma^2 - \gamma}{2}D\right)$$

    is market (this class of functions is the image under the tilde-transformation of the familiar class

    $$(S,D) \mapsto \exp\left(\gamma S + \frac{\gamma^2}{2}D\right)$$

    of functions parabolic in $\mathbb{R} \times (0,\infty)$, and even in $\mathbb{R}^2$; see, e.g., [103], 1.XV.2, Example (a));
  - we can take $\gamma = \pm k$ where $k$ is a number whose existence is asserted in (13.3).

Since $\tilde{u}(S,D)$ is constant along the lines $S + D/2 = \text{const}$, the function $\mathrm{LM}\,\tilde{u}$ increases[1] along those lines (the positive direction along such a line being the one in which $D$ increases). The formal argument (the same "shifting" argument as in Lemma 6.2 on p. 139) is:

$$\mathrm{LM}\,\tilde{u}(S,D) \le \mathrm{LM}\,\tilde{u}\,(S - \delta/2, D + \delta)$$

[1] "Increases" is used in the wide sense in this proof.

when $\delta > 0$ because

$$\mathrm{LM}\,\tilde{u}(S,D) \le f\,(S-\delta/2, D+\delta) \le \mathrm{LM}\,\tilde{u}\,(S-\delta/2, D+\delta)\,,$$

where $f := \mathrm{LM}\,u^*$ and $u^*(S,D)$ is defined as $\tilde{u}(S,D)$ if $D > \delta$ and as $\inf \tilde{u} = \inf u$ otherwise. Therefore, the function LSM $u$ increases in $D$.

First we prove that $(\mathrm{LSM}\,u)(S_0, D_0)+\epsilon$, for any $\epsilon > 0$, strongly super-replicates the American option $U$. We start with proving that, for an arbitrarily small $\delta > 0$, Cash can maintain his capital at any round $n$ above $(\mathrm{LSM}\,u)(S_n, D_n+\delta)$ if starting with $(\mathrm{LSM}\,u)(S_0, D_0+\delta)+\epsilon/2$. By the approximation theorem (stated on p. 139), there is a smooth supermarket function $\overline{U} \le \mathrm{LSM}\,u$ defined in the region

$$(s, D) \in (1/c, c) \times (\delta/2, C + 2\delta)$$

and arbitrarily close to LSM $u$. The construction in the proof of the approximation theorem gives $\overline{U}$ increasing in $D$; the property (6.25) (see p. 139) of smooth superparabolic functions and (13.4) imply that $\overline{U}$, like any other smooth supermarket function, satisfies the "Black-Scholes inequality"

$$\frac{\partial \overline{U}}{\partial D} \ge \frac{1}{2} s^2 \frac{\partial^2 \overline{U}}{\partial s^2}.$$

Therefore, we will have

$$\begin{aligned} d\overline{U}(S_n, D_n) \le{}& \frac{\partial \overline{U}}{\partial s}(S_n, D_n+\delta)dS_n + \frac{\partial \overline{U}}{\partial D}(S_n, D_n+\delta)\left(dD_n + \left(\frac{dS_n}{S_n}\right)^2\right) \\ &+ \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial s^3}(S''_n, D''_n+\delta)dS'_n(dS_n)^2 + \frac{1}{2}\frac{\partial^3 \overline{U}}{\partial D \partial s^2}(S''_n, D''_n+\delta)dD'_n(dS_n)^2 \\ &+ \frac{\partial^2 \overline{U}}{\partial D \partial s}(S'_n, D'_n+\delta)dD_n dS_n + \frac{1}{2}\frac{\partial^2 \overline{U}}{\partial D^2}(S'_n, D'_n+\delta)(dD_n)^2 \end{aligned} \tag{13.5}$$

in place of (10.39) (p. 250). Since the sum of the addends after the second on the right-hand side of (13.5) is negligible, Cash will be able to maintain his capital above LSM $u(S_n, D_n+\delta)$ until $D_n+\delta$ reaches $\delta$ at time $n = N$. (Notice that, since $\overline{U}$ is increasing in $D$, Cash only needs to hold long positions in security $\mathcal{D}$ when using this hedging strategy. Similarly, Investor's strategy constructed below will require short positions in $\mathcal{D}$.) Since $(\mathrm{LSM}\,u)(S_n, D_n+\delta) \ge u(S_n, D_n+\delta) = U(S_n)$, we will prove that $(\mathrm{LSM}\,u)(S_0, D_0)+\epsilon$ strongly super-replicates the American option if we show that the starting capital $(\mathrm{LSM}\,u)(S_0, D_0+\delta)+\epsilon/2$ does not exceed $(\mathrm{LSM}\,u)(S_0, D_0)+\epsilon$ for sufficiently small $\delta > 0$. This immediately follows from the continuity of $(\mathrm{LSM}\,u)(s, D)$ in $D$, which, in its turn, is an easy implication of the continuity of LM $\tilde{u}$ (see Lemma 6.2 on p. 139).

We have proved the "positive" part of the theorem; but before we can proceed, we will need several definitions. Let $\tilde{\mathcal{D}}$ be the set where LSM $u$ is different from $u$; since LSM $u \ge u$, LSM $u$ is lower semicontinuous and $u$ is continuous, $\mathcal{D}$ is open. The argument of Theorem 6.1 (p. 138) applied to $\widetilde{\mathrm{LSM}\,u} = \mathrm{LM}\,\tilde{u}$ shows that LSM $u$ is market (i.e., satisfies the Black-Scholes equation) inside $\mathcal{D}$.

Now we can prove that $c = (\mathrm{LSM}\,u)(S_0, D_0) - \epsilon$, where $\epsilon$ is a positive constant, is strongly super-replicated by the American option. First we consider the log-picture, $(s, D) \mapsto (S, D) := (\ln s - D/2, D)$. Let $\mathcal{D}$ be the set where LM $\tilde{u} > \tilde{u}$ (i.e., $\mathcal{D}$ is $\tilde{\mathcal{D}}$ in the log-picture). Analogously to the proof of Theorem 6.1 on p. 138, define

$$\mathcal{D}^* := \{(s, D) \mid \mathrm{LM}\,\tilde{u} - \tilde{u} > \epsilon/2\}\,.$$

Our goal is to prove that in the game for super-replicating \$$c$ with the American option, Investor can, starting with 0, end up with at least LSM $u(S_0, D_0) - \epsilon$ after exercising his American option. Investor's market strategy before $(S_n, D_n)$ hits $\partial\tilde{\mathcal{D}}^*$ will be to replicate the first two addends on the right-hand side of (10.39) with the opposite sign, where $\overline{U}$ is LSM $u$. Therefore, at the time of hitting $\partial\tilde{\mathcal{D}}^*$ Investor's capital will be close to $-U(S_n) +$ LSM $u(S_0, D_0)$; exercising the American option, he will procure a final capital close to LSM $u(S_0, D_0)$. ∎

## Examples

Suppose all the conditions of both Theorems 11.2 and 13.1, including the regularity conditions on the payoff function $U$, are satisfied. The solution to the Black-Scholes equation will be

$$\overline{u}(s, D) = \int_{\mathbb{R}} U(se^z)\ \mathcal{N}_{-D/2,\,D}(dz)$$

(cf. (11.13) on p. 280). If $U$ is convex, Jensen's inequality tells us that $\overline{u}(s, D) \geq U(sc)$, where $c$ is the mean value of $e^z$ under the distribution $\mathcal{N}_{-D/2,\,D}(dz)$. A simple calculation shows that $c = 1$. This result,

$$\overline{u}(s, D) \geq U(s),$$

shows that $\overline{u} =$ LSM $u$ ($u$ is defined in Theorem 13.1) and that it is optimal to never exercise convex American options, such as plain put and call options. It should be remembered, however, that we are assuming zero interest rates and no dividends for $S$. If either of these assumptions is violated, early exercise may be optimal; see, e.g., [154].

The case of binary options (see [351]) is not formally covered by Theorem 13.1, because their payoff functions are discontinuous. But it is easy to see that in the case of an American binary call,

$$U(s) = \begin{cases} 1 \text{ if } s > \alpha \\ 0 \text{ otherwise} \end{cases}$$

(where $\alpha$ is the strike price), the fair price is given (assuming $D(0) < \alpha$) by the value at the point $(S(0), D(0))$ of the solution $u(s, D)$ to the Black-Scholes equation in the region $(s, D) \in (0, \alpha) \times (0, \infty)$ with the initial condition $u(s, 0) = 0$ and the boundary condition $u(\alpha, D) = 1$. The analogous assertion for binary puts is also true.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk
Copyright © 2001 John Wiley & Sons, Inc.
ISBN: 0-471-40226-5

# 14

# *Games for Diffusion Processes*

So far in our study of game-theoretic option pricing, we have moved as far as possible away from the usual stochastic theory. Instead of assuming that the price of the underlying security $\mathcal{S}$ is governed by a stochastic differential equation, we have put much weaker constraints on the price process $S(t)$, and we have required that the market should price a derivative $\mathcal{D}$ that anticipates the realized variance of $\mathcal{S}$.

In this chapter, we translate the stochastic Black-Scholes theory more directly into our game-theoretic framework. It should be no surprise that such a translation is possible, for as we learned in §8.2, any stochastic process can be embedded in a probability game. But it is instructive to work out the details. We obtain new insights not only into the Black-Scholes theory, but into diffusion processes in general.

As we learned in Chapter 9, the stochastic Black-Scholes theory has both game-theoretic and stochastic elements. On the game-theoretic side, we have Investor playing a hedging strategy: hold $(\partial \overline{U}/\partial s)(S(t), t)$ shares of $\mathcal{S}$ at time $t$. On the stochastic side, Market somehow generates the increments $dS(t)$ according to the stochastic differential equation

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t). \qquad (14.1)$$

The idea of this chapter is to interpret (14.1) game-theoretically, so that the entire setup becomes game-theoretic. As we learned in §8.2, we make a sequential stochastic model game-theoretic by interpreting the probabilities given by the model as a fixed strategy for Forecaster in a game involving Forecaster, Skeptic, and Reality; the probabilities are Forecaster's forecasts for what Reality will do, which are binding on Reality in the sense that Reality can violate them on average only if she is willing to allow Skeptic to become rich. So the game-theoretic interpretation of (14.1) is

- $\mu dt$ is the price Forecaster sets for $dS(t)/S(t)$, and
- $\sigma dt$ is the price Forecaster sets for $(dS(t)/S(t))^2$.

With this interpretation, we call (14.1) a *game-theoretic differential equation.*

The game-theoretic interpretation of what was a stochastic element adds additional players to the game. In our discussion in Chapter 9 (see the protocol on p. 217) we had only Investor, who was hedging, and Market, who was behaving stochastically. Now we have added Skeptic. We need not add Reality, because Market is already there to choose the $dS(t)$. And we need not count Forecaster as a player, because he is required to follow the strategy defined by (14.1), which is known to the other players and therefore becomes part of the protocol of the game. So the game now has three players: Investor, Skeptic, and Market. Both Skeptic and Investor are playing against Market. Investor's goal is to reproduce the derivative $\mathcal{U}$ starting with the Black-Scholes price by trading in $\mathcal{S}$ alone. Skeptic's goal is to become rich. Skeptic and Investor win if one of them succeeds. We will show that as a team they do have a winning strategy. Thus Investor can succeed in his hedging almost surely—i.e., unless Skeptic becomes infinitely rich.

To summarize, this chapter's game-theoretic reinterpretation of the stochastic Black-Scholes theory contrasts with the game-theoretic Black-Scholes theory of Chapters 10 and 11 in two major respects:

1. In the Black-Scholes game of Chapters 10 and 11, Investor trades in both $\mathcal{S}$ and $\mathcal{D}$. In the game of this chapter, he trades only in $\mathcal{S}$.
2. In the Black-Scholes game of Chapters 10 and 11, Investor's hedging always succeeds. In the game of this chapter, it succeeds only almost surely—that is, only if Skeptic does not become rich.

One instructive sidelight of the Black-Scholes argument that we study in this chapter is the way the drift $\mu$ drops out. In the stochastic argument, the drift is usually irrelevant to what we want to do, but we are accustomed to thinking that it needs to be present to complete the stochastic model. In the game-theoretic interpretation, it becomes clear that we can dispense with it altogether. We do not need Forecaster to price $dS$ (this is what he would do with $\mu$); we only need for him to price $(dS)^2$ (this is what he does with $\sigma$). We represent this possibility symbolically by writing

$$\frac{dS(t)}{S(t)} = \text{no price for drift} + \sigma dW(t). \tag{14.2}$$

Whereas (14.1) is a game-theoretic differential equation for drift and volatility, (14.2) is a game-theoretic differential equation for volatility only. It means that Skeptic can buy $(dS(t)/S(t))^2$ for $\sigma dt$ but cannot buy the relative increment $dS(t)/S(t)$ at any price. It should not be confused with (14.1) with $\mu = 0$, which says that Skeptic can buy $(dS(t)/S(t))^2$ for $\sigma dt$ and $dS(t)/S(t)$ for 0.

The game-theoretic interpretation of diffusion processes may be useful not only in option pricing but also in other fields where diffusion processes are used. In general, it clarifies the meaning of the assumption that a phenomenon follows a diffusion process, and it expresses the assumption in a way that may often point to weaker or different assumptions.

We begin this chapter with a general discussion of protocols for diffusion processes (§14.1) and a game-theoretic version of Itô's lemma (§14.2). Then we use these ideas to obtain our game-theoretic reinterpretation of the stochastic Black-Scholes formula (§14.3). In appendixes (§14.4 and §14.5), we comment on the nonstandard mathematics of this chapter and on some related stochastic literature.

## 14.1 GAME-THEORETIC DIFFUSION PROCESSES

In this section, we lay out some protocols in which a game-theoretic differential equation can serve as a strategy for Forecaster. The players in these protocols are Forecaster, Skeptic, and Reality. Our Black-Scholes game in §14.3 will add Investor and substitute Market for Reality.

We begin with protocols in which Forecaster prices both drift and volatility, and then we consider a simpler protocol where he prices only volatility.

As usual, our games are played over the time period $[0, T]$, during which a huge (nominally infinite) number $N$ of rounds are played; each round occupies a tiny (nominally infinitesimal) time period of length $dt := T/N$. As in our protocol for the strong law of large numbers in §4.1, we define a process $S_n$ in terms of increments $x_n$.

### Pricing Drift and Volatility

We begin with two protocols in which Forecaster can use a game-theoretic differential equation for drift and volatility.

The first protocol is very similar to the protocol that we used to prove the strong law of large numbers in §4.1.

DIFFUSION PROTOCOL 0
**Players:** Forecaster, Skeptic, Reality
**Protocol:**

$\mathcal{K}_0 := 1$.
$S_0 := 0, T_0 := 0, A_0 := 0$.
FOR $n = 1, 2, \ldots, N$:
Forecaster announces $m_n \in \mathbb{R}$ and $v_n \geq 0$.
Skeptic announces numbers $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.
Reality announces number $x_n \in \mathbb{R}$.
$S_n := S_{n-1} + x_n$.
$T_n := T_{n-1} + m_n$.
$A_n := A_{n-1} + v_n$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n) + V_n((x_n - m_n)^2 - v_n)$.

This protocol includes definitions of the *trend* $T_n$ and the *quadratic variation* $A_n$ of the process $S_n$.

In the second protocol, Forecaster provides forecasts for $x_n$ and $x_n^2$.

DIFFUSION PROTOCOL 1

**Players:** Forecaster, Skeptic, Reality

**Protocol:**

$\mathcal{K}_0 := 1$.

$S_0 := 0, T_0 := 0, A_0 := 0$.

FOR $n = 1, 2, \ldots, N$:

Forecaster announces $m_n \in \mathbb{R}$ and $v_n \geq 0$.

Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.

Reality announces $x_n \in \mathbb{R}$.

$S_n := S_{n-1} + x_n$.

$T_n := T_{n-1} + m_n$.

$A_n := A_{n-1} + v_n$.

$$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n) + V_n(x_n^2 - v_n). \tag{14.3}$$

These two protocols differ only in the role of $v_n$. In Diffusion Protocol 0, $v_n$ is the price of $(x_n - m_n)^2$—that is, the variance of $x_n$. In Diffusion Protocol 1, it is simply the price of $x_n^2$. Because

$$x_n^2 = ((x_n - m_n) + m_n)^2 = (x_n - m_n)^2 + 2m_n(x_n - m_n) + m_n^2,$$

and because there is a ticket with payoff $x_n - m_n$, the real difference between the two protocols is in the term $m_n^2$. Because $m_n$ typically has the order of magnitude $O(dt)$, summing $m_n^2$ with bounded coefficients will produce something also of order $O(dt)$. So the two protocols are very close.

In either of the two protocols, if Forecaster always sets

$$m_n := \mu(S_{n-1}, ndt)dt \quad \text{and} \quad v_n := \sigma(S_{n-1}, ndt)dt, \tag{14.4}$$

where the functions $\mu : \mathbb{R}^2 \to \mathbb{R}$ and $\sigma : \mathbb{R}^2 \to [0, \infty)$ are known in advance to all the players, then we say that Forecaster *follows* the game-theoretic differential equation

$$dS(t) = \mu(S(t), t)dt + \sigma(S(t), t)dW(t). \tag{14.5}$$

If we then adopt the fundamental interpretative hypothesis, then we may say that Reality is *governed by* this equation (see p. 182).

### Pricing Volatility Alone

Consider now the following simplified protocol:

DIFFUSION PROTOCOL 2

**Players:** Forecaster, Skeptic, Reality

**Protocol:**

$\mathcal{K}_0 := 1.$
$S_0 := 0$, $A_0 := 0$.
FOR $n = 1, 2, \ldots, N$:
  Forecaster announces $v_n \geq 0$.
  Skeptic announces $V_n \in \mathbb{R}$.
  Reality announces $x_n \in \mathbb{R}$.
  $S_n := S_{n-1} + x_n$.
  $A_n := A_{n-1} + v_n$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + V_n(x_n^2 - v_n)$.

In this protocol, Skeptic is not offered $x_n$-tickets at any price, and Forecaster does not say anything that can be interpreted as an assumption about trend.

If Forecaster always sets

$$v_n := \sigma(S_{n-1}, ndt)dt,$$

where $\sigma : \mathbb{R}^2 \to [0, \infty)$ is known in advance to all the players, then we say that Forecaster follows the game-theoretic differential equation

$$dS(t) = \text{no price for drift} + \sigma(S(t), t)dW(t).$$

This is a game-theoretic differential equation for volatility alone.

## Quadratic Variation and 2-Variation

There are two distinct concepts of quadratic variation in our protocols: $\mathbf{var}_S(2)$ and $A_N$. The following proposition relates the two. We do not make any assumption here about the strategy followed by Forecaster.

**Proposition 14.1** *Suppose Reality is required, in one of our diffusion protocols, to make $S(t)$ continuous. Then almost surely, $\mathbf{var}_S(2) < \infty$ if and only if $A_N < \infty$.*

*Proof* Because $S$ is continuous, all the $x_n$ are infinitesimal and hence less than 1. Suppose $\mathbf{var}_S(2) < \infty$. For $k = 1, 2, \ldots$, define a martingale $\mathcal{T}^{(k)}$ inductively by

$$\mathcal{T}_n^{(k)} = 2^k + \sum_{i=1}^{n} (v_i - x_i^2)$$

for all values of $n$ starting with $n = 1$ and switching, when and if the right-hand side gets down to the value 1, to $\mathcal{T}_n^{(k)} = \mathcal{T}_{n-1}^{(k)}$. Then the martingale

$$\sum_{k=1}^{\infty} 2^{-2k} \mathcal{T}_n^{(k)} \tag{14.6}$$

exists (because the strategies corresponding to $\mathcal{T}_n^{(k)}$ take only two values, 0 or $-1$) and goes from 1 to an infinitely large amount when $\sum_i x_i^2$ is finite but $\sum_i v_i$ is infinitely large.

The construction is similar in the case where we start with the assumption that $A(T) < \infty$. In this case, we define

$$\mathcal{T}_n^{(k)} = 2^k + \sum_{i=1}^{n}(x_i^2 - v_i)$$

before the right-hand side reaches 1 and $\mathcal{T}_n^{(k)} = \mathcal{T}_{n-1}^{(k)}$ afterwards. The martingale (14.6) again exists and goes from 1 to an infinitely large amount when $\sum_i v_i$ is finite but $\sum_i x_i^2$ is infinite. ∎

In measure-theoretic probability [112, 158, 206], the process $A(t)$ corresponds to the *predictable quadratic variation* of the price process $S(t)$, whereas $\mathbf{var}_S(2)$ corresponds to the final value of the *optional quadratic variation* of $S(t)$. For continuous semimartingales, the two coincide almost surely ([158], Theorem I.4.52). Here we draw only the weaker conclusion that that the finiteness of $\mathbf{var}_S(2)$ and of $A(T)$ imply each other almost surely, but they actually nearly coincide with high lower probability; this can be shown in the same way that we show that (14.11) is small in absolute value in the proof of Proposition 14.2 below.

## 14.2 ITÔ'S LEMMA

In this section, we formulate and prove Itô's lemma for our diffusion protocols. We do not assume that Forecaster follows any particular strategy, but we do impose constraints on Forecaster's and Reality's moves.

Let $f$ and $g$ be real valued functions on $\mathbb{T}$. We define the *integral* (or *stochastic integral*) $\int_0^T f dg$ as follows:

$$\int_0^T f(t)dg(t) := \sum_{n=1}^{N} f((n-1)dt)\big(g(ndt) - g((n-1)dt)\big)$$

(we will not need other limits of integration). This definition uses $f((n-1)dt)$ rather than $f(ndt)$ because $f((n-1)dt)$ is known before Skeptic's move at trial $n$. This is analogous to the standard requirement of predictability for the integrand in measure-theoretic stochastic integration, where it is usually assumed that $f$ is a path of a locally bounded predictable process and $g$ is a path of a semimartingale [112, 158].

Let $A$ and $B$ be two variables (i.e., functions of Forecaster and Reality's moves). We will say that $A$ and $B$ coincide almost surely, $A \overset{a.s.}{=} B$, under some set of conditions $\gamma$ on Forecaster and Reality's moves in a diffusion protocol if, for any (arbitrarily small) $\epsilon > 0$ and any (arbitrarily large) $C > 0$, Skeptic has a strategy in the protocol guaranteeing that $\min_{0 \le t \le T} \mathcal{K}(t) \ge 0$ and either

$$|A - B| \le \epsilon$$

or

$$\mathcal{K}(T) \ge C,$$

when $\gamma$ is added to the protocol as a constraint on Forecaster's and Reality's moves.

**Proposition 14.2 (Itô's Lemma)** *In Diffusion Protocol 2,*

$$f(S(T),T) - f(S(0),0) \overset{a.s.}{=} \int_0^T \frac{\partial f}{\partial t}(S(t),t)dt + \int_0^T \frac{\partial f}{\partial s}(S(t),t)dS(t) + \frac{1}{2}\int_0^T \frac{\partial^2 f}{\partial s^2}(S(t),t)dA(t), \tag{14.7}$$

*provided $f(s,t)$ is a smooth function and Forecaster and Reality are required to satisfy*

- $\mathbf{var}_S(2) < \infty$ *or* $A(T) < \infty$,
- $\sup_{0\le t\le T}|S(t)| < \infty$, *and*
- $A(t)$ *and* $S(t)$ *are continuous.*

*Proof* By Proposition 14.1, the two conditions $\mathbf{var}_S(2) < \infty$ and $A(T) < \infty$ imply each other almost surely, and hence we assume that both $\mathbf{var}_S(2) < \infty$ and $A(T) < \infty$.

Let $A =^{\epsilon} B$ (in words: $A$ and $B$ are $\epsilon$-close) stand for $|A - B| \le \epsilon$. Expanding the definitions, we can rewrite (14.7) as: $\inf_n \mathcal{K}_n \ge 0$ and either

$$\begin{aligned} &f(S_N,T) - f(S_0,0) \\ &=^{\epsilon} \sum_{n=1}^N \frac{\partial f}{\partial t}(S_{n-1},(n-1)dt)dt + \sum_{n=1}^N \frac{\partial f}{\partial s}(S_{n-1},(n-1)dt)x_n \\ &\quad + \frac{1}{2}\sum_{n=1}^N \frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)v_n \end{aligned} \tag{14.8}$$

or $\mathcal{K}_N \ge C$. Here, as usual, $dt = T/N$.

Using Taylor's formula, we find ($\theta_i$ are numbers between 0 and 1):

$$\begin{aligned} f(S_N,T) - f(S_0,0) &= \sum_{n=1}^N \Big( f(S_n,ndt) - f(S_{n-1},(n-1)dt) \Big) \\ = \sum_{n=1}^N \Big( f(S_n,ndt) &- f(S_n,(n-1)dt) + f(S_n,(n-1)dt) - f(S_{n-1},(n-1)dt) \Big) \\ &= \sum_{n=1}^N \Big( \frac{\partial f}{\partial t}(S_n,(n-\theta_n)dt)dt + \frac{\partial f}{\partial s}(S_{n-1},(n-1)dt)x_n \\ &\quad + \frac{1}{2}\frac{\partial^2 f}{\partial s^2}(S_{n-1}+\theta_{N+n}(S_n-S_{n-1}),(n-1)dt)x_n^2 \Big) \\ &= \sum_{n=1}^N \frac{\partial f}{\partial t}(S_n,(n-\theta_n)dt)dt + \sum_{n=1}^N \frac{\partial f}{\partial s}(S_{n-1},(n-1)dt)x_n \\ &\quad + \frac{1}{2}\sum_{n=1}^N \frac{\partial^2 f}{\partial s^2}(S_{n-1}+\theta_{N+n}(S_n-S_{n-1}),(n-1)dt)x_n^2. \end{aligned}$$

Comparing this to (14.8), we can see that it is sufficient to prove that

$$\sum_{n=1}^{N}\left(\frac{\partial f}{\partial t}(S_n,(n-\theta_n)dt)-\frac{\partial f}{\partial t}(S_{n-1},(n-1)dt)\right)dt \tag{14.9}$$

and

$$\sum_{n=1}^{N}\left(\frac{\partial^2 f}{\partial s^2}(S_{n-1}+\theta_{N+n}(S_n-S_{n-1}),(n-1)dt)-\frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)\right)x_n^2 \tag{14.10}$$

are infinitesimal and that

$$\sum_{n=1}^{N}\frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)(x_n^2-v_n) \tag{14.11}$$

is less than $\epsilon$ in absolute value unless some nonnegative martingale starts at 1 and ends at $C$.

The assertion that (14.9) is infinitesimal follows from the fact that, uniformly in $n$,

$$\left|\frac{\partial f}{\partial t}(S_n,(n-\theta_n)dt)-\frac{\partial f}{\partial t}(S_{n-1},(n-1)dt)\right|\le c(|x_n|+dt)\approx 0,$$

where $c$ is $\sup_{0\le t\le T}\max\left(\partial^2 f/\partial t^2,\partial^2 f/(\partial s\partial t)\right)$ (recall that the function $f=f(s,t)$ is smooth and so bounded with all its derivatives in the region $t\in[0,T]$ and $|s|\le\sup_t|S(t)|$). To see that (14.10) is infinitesimal, notice that

$$\frac{\partial^2 f}{\partial s^2}(S_{n-1}+\theta_{N+n}(S_n-S_{n-1}),(n-1)dt)\approx\frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)$$

uniformly in $n$, and therefore, for an arbitrarily small $\delta>0$,

$$\left|\sum_{n=1}^{N}\left(\frac{\partial^2 f}{\partial s^2}(S_{n-1}+\theta_{N+n}(S_n-S_{n-1}),(n-1)dt)-\frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)\right)x_n^2\right|\le\delta c,$$

where $c$ is an upper bound on $\mathbf{var}_S(2)$.

To complete the proof, we will prove that (14.11) squared, that is,

$$\left(\sum_{n=1}^{N}\frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)(x_n^2-v_n)\right)^2,$$

is less than $\epsilon^2$ unless some nonnegative martingale increases its initial value $C$-fold; our proof will again be similar to the proof of the weak law of large numbers in Chapter 6. Let us consider the nonnegative submartingale

$$\mathcal{S}_n:=\left(\sum_{i=1}^{n}\frac{\partial^2 f}{\partial s^2}(S_{i-1},(i-1)dt)(x_i^2-v_i)\right)^2.$$

Since

$$\mathcal{S}_n - \mathcal{S}_{n-1} = 2\left(\sum_{i=1}^{n-1} \frac{\partial^2 f}{\partial s^2}(S_{i-1},(i-1)dt)(x_i^2 - v_i)\right)\frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)(x_n^2 - v_n) + \left(\frac{\partial^2 f}{\partial s^2}(S_{n-1},(n-1)dt)(x_n^2 - v_n)\right)^2,$$

its "compensator" is

$$\mathcal{A}_n := \sum_{i=1}^{n}\left(\frac{\partial^2 f}{\partial s^2}(S_{i-1},(i-1)dt)(x_i^2 - v_i)\right)^2.$$

(This is not a usual compensator, because it is not predictable.) Writing $c$ for an upper bound on $\partial^2 f/\partial s^2$, we see that

$$\mathcal{A}_N \le c^2 \sum_{n=1}^{N}(x_n^2 - v_n)^2 \le 2c^2 \sum_{n=1}^{N} x_n^4 + 2c^2 \sum_{n=1}^{N} v_n^2$$
$$\le 2c^2 \left(\sum_{n=1}^{N} x_n^2\right)\left(\max_{n=1}^{N} x_n^2\right) + 2c^2 \left(\sum_{n=1}^{N} v_n\right)\left(\max_{n=1}^{N} v_n\right),$$

and this is infinitesimal because of the continuity of $S(t)$ and $A(t)$ and the finiteness of $\mathbf{var}_S(2)$ and $A(T)$. Therefore, the martingale

$$\mathcal{T}_n := \epsilon^2/C + \mathcal{S}_n - \mathcal{A}_n \tag{14.12}$$

is nonnegative. If its final value $\mathcal{T}_N$ exceeds $\epsilon^2$, it will have made $\epsilon^2$ out of $\epsilon^2/C$. ∎

Proposition 14.2 extends easily to Diffusion Protocols 0 and 1. Skeptic can use the strategy constructed for Diffusion Protocol 2 and always take $M_n = 0$; in Diffusion Protocol 1 $m_n$ will be completely irrelevant, and in Diffusion Protocol 0 the role of $x_n$ will be played by the difference $x_n - m_n$. In the case of Diffusion Protocol 0, however, continuity and boundedness (and finiteness of the 2-variation if $A(T) < \infty$ is omitted) should be imposed on the difference $S - T$ rather than on $S$.

The assumption that $\sup_t |S(t)|$ is finite is essential for Proposition 14.2. We cannot deduce it from the finiteness of $A(T)$ since no assumptions about the drift of $S(t)$ are made (the usual idea of bounding a compensator of $S^2(t)$, as used in the proof of the weak law of large numbers in §6.1, will not work since compensators will change). To verify that the proposition fails when the condition $\sup_t |S(t)| < \infty$ is dropped, set $f(s,t) := e^s$. Arguing indirectly, suppose the proposition does remain true, suppose $v_n = dt$ for all $n$, and suppose Reality randomly chooses $x_n = 0$ or $x_n = \sqrt{2dt}$. Skeptic's capital $\mathcal{K}_n$ will be a martingale; therefore, by Doob's inequality $\mathcal{K}_{N-1}$ will be at most 100 with probability at least 99%. Fix such a typical play through step $N-1$. Consider the difference $D_0$ between the left-hand side and right-hand side of (14.7) when $x_N = 0$ and the difference $D_+$ between them when $x_N = \sqrt{2dt}$. The difference $D_+ - D_0$ will be at least

$$\left(e^{S_{N-1}+\sqrt{2dt}} - e^{S_{N-1}}\sqrt{2dt}\right) - \left(e^{S_{N-1}}\right) = e^{S_{N-1}}\left(e^{\sqrt{2dt}} - 1 - \sqrt{2dt}\right) \ge e^{S_{N-1}}dt,$$

which, with probability close to 1, is infinitely large, comparable to $e^{\sqrt{N}}/N$. Since in both cases (whether $x_N$ is 0 or $\sqrt{2dt}$) $\mathcal{K}_N$ will not exceed 200, we arrive at a contradiction.

## 14.3 GAME-THEORETIC BLACK-SCHOLES DIFFUSION

We now expand our diffusion protocols by adding another player, Investor, who tries to replicate a European option $U(S(T))$ at the cost given by the Black-Scholes formula. As we explained in the introduction, if Forecaster follows the appropriate game-theoretic differential equation, Investor and Skeptic can assure that Investor achieves his goal almost surely—that is, unless Skeptic becomes infinitely rich.

Adding Investor and his capital process to Diffusion Protocol 1, changing the name of Reality to Market, and requiring $S_n > 0$, we obtain this protocol:

BLACK-SCHOLES DIFFUSION PROTOCOL 1

**Parameters:** $\mu, \sigma > 0, c$

**Players:** Forecaster, Skeptic, Investor, Market

**Protocol:**

$\mathcal{K}_0 := 1$.

$\mathcal{I}_0 := c$.

Market chooses $S_0 > 0$.

FOR $n = 1, 2, \ldots, N$:

Forecaster announces $m_n \in \mathbb{R}$ and $v_n \geq 0$.

Skeptic announces $M_n \in \mathbb{R}$ and $V_n \in \mathbb{R}$.

Investor announces $X_n \in \mathbb{R}$.

Market announces $S_n > 0$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(\Delta S_n - m_n) + V_n((\Delta S_n)^2 - v_n)$.

$\mathcal{I}_n := \mathcal{I}_{n-1} + X_n \Delta S_n$.

**Additional Constraints:** Forecaster must follow (14.1), considered as a game-theoretic differential equation; this means that he must set

$$m_n := \mu S_{n-1} dt \quad \text{and} \quad v_n := \sigma^2 S_{n-1}^2 dt.$$

Market must assure that $S$ is continuous and $\max |S| < \infty$.

Here $\mathcal{K}_n$ is Skeptic's capital process, and $\mathcal{I}_n$ is Investor's capital process.

Suppose $\gamma$ is a property that may or may not be satisfied by the path $S(t)$ chosen by Market. We say that $c \in \mathbb{R}$ is the *almost sure price given* $\gamma$ for a European option $U(S(T))$ if, for any (arbitrarily small) $\epsilon > 0$ and for any (arbitrarily large) $C > 0$, there exist strategies for Investor and Skeptic in the preceding protocol that guarantee that $\inf_t \mathcal{K}(t) \geq 0$ and either $\mathcal{K}_N \geq C$ or

$$\gamma(S) \Rightarrow |\mathcal{I}_N - U(S(T))| \leq \epsilon$$

if Market is constrained to satisfy $\gamma$.

The following is the analog of Theorem 11.2 (p. 280) closest to the usual measure-theoretic Black-Scholes formula:

**Proposition 14.3** *Suppose $U$ is a Lipschitzian function. Then when $S(0) > 0$ has just been announced in Black-Scholes Diffusion Protocol 1, the almost sure price for $U(S(T))$ is*

$$\int_{\mathbb{R}} U(S(0)e^z)\, \mathcal{N}_{-T\sigma^2/2, T\sigma^2}(dz).$$

This is almost a direct translation of the measure-theoretic Black-Scholes formula.

We now simplify in the spirit of Diffusion Protocol 2, omitting the drift altogether.

BLACK-SCHOLES DIFFUSION PROTOCOL 2

**Parameters:** $\mu, \sigma > 0, c$

**Players:** Forecaster, Skeptic, Investor, Market

**Protocol:**

$\mathcal{K}_0 := 1.$
$\mathcal{I}_0 := c.$
Market chooses $S_0 > 0$.
FOR $n = 1, 2, \ldots, N$:
  Forecaster announces $v_n \geq 0$.
  Skeptic announces $V_n \in \mathbb{R}$.
  Investor announces $X_n \in \mathbb{R}$.
  Market announces $S_n > 0$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + V_n((\Delta S_n)^2 - v_n)$.
  $\mathcal{I}_n := \mathcal{I}_{n-1} + X_n \Delta S_n$.

**Additional Constraints:** Forecaster must follow (14.2), considered as a game-theoretic differential equation; this means that he must set

$$v_n := \sigma^2 S_{n-1}^2 dt.$$

Market must assure that $S$ is continuous and $\max |S| < \infty$.

**Theorem 14.1** *Suppose $U$ is a Lipschitzian function. Then when $S(0) > 0$ has just been announced in Black-Scholes Diffusion Protocol 2, the almost sure price for $U(S(T))$ is*

$$\int_{\mathbb{R}} U(S(0)e^z)\, \mathcal{N}_{-T\sigma^2/2, T\sigma^2}(dz).$$

*Proof* The proof is similar to that of Theorem 11.2 in Chapter 11; the most essential difference is that now we take $D = D(t) := \sigma^2(T - t)$. Assuming that $U$ is smooth (we reduce the case where $U$ is merely Lipschitzian to the case where it is smooth as in Chapter 11) and putting $f(s,t) := \overline{U}(s, \sigma^2(T-t))$, we find from the Black-Scholes equation (10.38):

$$\frac{\partial f}{\partial t} = \frac{\partial \overline{U}}{\partial D}\frac{\partial D}{\partial t} = \frac{1}{2}s^2\frac{\partial^2 \overline{U}}{\partial s^2}(-\sigma^2) = -\frac{1}{2}\sigma^2 s^2 \frac{\partial^2 f}{\partial s^2},$$

so from Itô's lemma (Proposition 14.2 on p. 341) we obtain, setting $x_n := \Delta S_n$,

$$
\begin{aligned}
& f(S_N, T) - f(S_0, 0) \\
&\overset{\text{a.s.}}{=} \sum_{n=1}^{N} \frac{\partial f}{\partial t}(S_{n-1}, (n-1)dt)dt + \sum_{n=1}^{N} \frac{\partial f}{\partial s}(S_{n-1}, (n-1)dt)x_n \\
&\quad + \frac{1}{2}\sum_{n=1}^{N} \frac{\partial^2 f}{\partial s^2}(S_{n-1}, (n-1)dt)v_n \\
&= -\frac{1}{2}\sum_{n=1}^{N} \sigma^2 S_{n-1}^2 \frac{\partial^2 f}{\partial s^2}(S_{n-1}, (n-1)dt)dt + \sum_{n=1}^{N} \frac{\partial f}{\partial s}(S_{n-1}, (n-1)dt)x_n \\
&\quad + \frac{1}{2}\sum_{n=1}^{N} \frac{\partial^2 f}{\partial s^2}(S_{n-1}, (n-1)dt)\sigma^2 S_{n-1}^2 dt \\
&= \sum_{n=1}^{N} \frac{\partial f}{\partial s}(S_{n-1}, (n-1)dt)x_n;
\end{aligned}
$$

since the last sum is exactly the gain of Investor's strategy which recommends buying

$$\frac{\partial f}{\partial s}(S_{n-1}, (n-1)dt)$$

shares on round $n$, this completes the proof. ∎

Proposition 14.3 follows immediately from Theorem 14.1, because Skeptic can reduce Black-Scholes Diffusion Protocol 1 to Black-Scholes Diffusion Protocol 2 by setting $M_n$ to 0 for all $n$.

## 14.4 APPENDIX: THE NONSTANDARD INTERPRETATION

In this appendix we give a few guidelines concerning how our formulation and proof of Itô's lemma (p. 341) is to be understood in the nonstandard framework spelled out in §11.5.

As usual, we consider a sequence of protocols indexed by $k = 1, 2, \ldots$, the $k$th protocol having $N_k$ rounds of play. The condition $\mathbf{var}_S(2) < \infty$ was explained in Chapter 11. The condition $A(T) < \infty$ is interpreted analogously: the sequence $A^{(k)}(T)$ of quadratic variances corresponding to $k = 1, 2, \ldots$ (with a fixed nontrivial ultrafilter on the set of $k$s) is a finite nonstandard number. In a similar way $\sup_t |S(t)| < \infty$ means that for every $k$ we consider $M^{(k)} := \sup_t |S^{(k)}(t)|$, and the sequence $M^{(k)}$ is a finite nonstandard number; $S(t)$ being continuous means that the sequence $d^{(k)} := \sup_{n=1,\ldots,N^{(k)}} \left|x_n^{(k)}\right|$ is an infinitesimal nonstandard number. The condition $\inf_{0 \le t \le T} \mathcal{K}(t) \ge 0$ means that the capital process is nonnegative for most $k$. The martingale (14.12) is also nonnegative for most $k$. The continuity of $A(t)$ and $S(t)$ implies that the martingale (14.6) is nonnegative for most $k$ (namely, for the $k$ such that $\sup_n |x_n^{(k)}| \le 1$ and $\sup_n v_n^{(k)} \le 1$).

## 14.5 APPENDIX: RELATED STOCHASTIC THEORY

In this appendix, we relate the ideas of this chapter to some of the existing stochastic literature. We comment on stochastic differential equations, risk-neutral valuation, and attempts to study the paths of stochastic processes outside of measure theory.

### Stochastic Differential Equations

Measure-theoretic probability theory distinguishes between strong and weak solutions for stochastic differential equations. In the case of weak solutions, we do not really begin with a Wiener process $W(t)$ in (14.5): we ask only for a process with infinitesimal increments with mean $\mu(S(t),t)dt$ and variance $\sigma^2(S(t),t)dt$. This chapter's game-theoretic interpretation of stochastic differential equations is in this spirit. For strong solutions, on the other hand, we actually begin with a Wiener process $W(t)$, and we use (14.5) as a recipe for constructing a process $S(t)$ from $W(t)$. The recipe does not always work; for example, when we try to follow (14.5) starting at $t = 0$, $S(t)$ may explode to infinity in finite time. It is not trivial to identify conditions under which the recipe does work, and these conditions are of course stronger than those required for weak solutions. But the main difference is that strong solutions require much stronger conditions on the drift term $\mu(S,t)$, which does not even appear in our game-theoretic treatment of the Black-Scholes formula. We might construct strong game-theoretic solutions by means of an auxiliary protocol similar to Diffusion Protocol 0 but with Forecaster always choosing $m_n = 0$ and $v_n = dt$; Reality's path $W(t)$ in this protocol would have the properties of Brownian motion, and we could then call the process $S(t)$ defined by (14.5) a strong solution.

Stroock and Varadhan's representation of the problem of finding a weak solution to a stochastic differential equation as a martingale problem is especially close to our game-theoretic approach; see [299, 300] and [166], §5.4. Stroock and Varadhan's idea, loosely speaking, is to look for a probability measure on the set of all continuous functions $y : [0,\infty) \to \mathbb{R}$ such that for every function $f$ that is twice continuously differentiable,

$$df(y(t)) - \mu(y(t),t)f'(y(t))dt - \frac{1}{2}\sigma^2(y(t),t)f''(y(t))dt \tag{14.13}$$

is a martingale difference (at least in a local sense). It is sufficient to require this condition for $f(y) = y$ and $f(y) = y^2$; see, e.g., [166], Proposition 5.4.6. For $f(y) = y$, (14.13) reduces to requiring that

$$dy(t) - \mu(y(t),t)dt \tag{14.14}$$

should be a martingale difference; in game-theoretic terms, Skeptic should be able to buy or replicate the $x_n$-tickets responsible for the addend $M_n(x_n - m_n)$ in (14.3), where $m_n$ is given by (14.4). For $f(y) = y^2$, (14.13) reduces to requiring that

$$\begin{aligned} &d\left(y^2(t)\right) - 2\mu(y(t),t)y(t)dt - \sigma^2(y(t),t)dt \\ &= 2y(t)[dy(t) - \mu(y(t),t)dt] + [(dy(t))^2 - \sigma^2(y(t),t)dt] \end{aligned}$$

be a martingale difference; since (14.14) is a martingale difference, this is equivalent to

$$dy(t))^2 - \sigma^2(y(t), t)dt$$

being a martingale difference. In the game-theoretic terms, this says Skeptic should be able to buy or replicate the $x_n^2$-tickets responsible for the addend $V_n(x_n^2 - v_n)$ in (14.3), where $v_n$ is given by (14.4).

A large part of the theory of stochastic differential equations is concerned with the existence and uniqueness of solutions. We have not attempted here even to define precisely what existence and uniqueness would mean in the game-theoretic approach, but the meanings are clear at an intuitive level. Existence of a solution to a game-theoretic differential equation means that the protocol is coherent when the equation is interpreted as a constraint on Forecaster, so that the lower price of a variable never exceeds its upper price. Uniqueness means that the lower and upper price coincide for variables that are well behaved. This corresponds to uniqueness in the sense of probability law in measure-theoretic probability ([166], §5.3.A).

### Risk-Neutral Valuation

The stochastic principle of risk-neutral valuation [147], discussed in detail in §9.6, can be summarized as follows:

1. Modify the probability measure governing the underlying security $S$ so that under the modification (usually called the *equivalent martingale measure* in the probabilistic literature and the *risk-neutral* or *risk-adjusted* probability in the financial literature) the price process $S(t)$ (properly discounted if our assumption of zero interest rates is dropped) is a martingale.

2. Find the price of the given contingent claim by computing the payoff's expectation under the risk-neutral probability.

The modification mentioned in Step 1 can be effected by removing the drift term from the stochastic differential equation governing $S(t)$ (more generally, equating the drift to the interest rate).

The key difference between the stochastic and game-theoretic approaches is that whereas the stochastic approach essentially replaces a strategy for Forecaster that gives arbitrary values for the drift by one that always gives 0 for the drift, the game-theoretic approach does not start with any price for the drift but still obtains the same probabilities (now game-theoretic).

### Stochastic Processes without Measure

Because it does not assume a measure as its starting point, our game-theoretic interpretation of stochastic processes deals directly with paths. A number of other authors, while not taking a game-theoretic approach, have also considered how paths of stochastic processes might be analyzed directly. Here we briefly review some examples.

Föllmer (1981) describes a pathwise approach to Itô's integral. For some functions $f : [0,T] \to \mathbb{R}$ and sequences $(\tau_n)$ of finite partitions of $[0,T]$ whose mesh tends to 0, Föllmer defines the notion of quadratic variation of $f$ relative to $(\tau_n)$ and uses this definition to prove a pathwise version of Itô's lemma, where stochastic integral is defined as the limit of Riemann–Stieltjes sums along $(\tau_n)$. The dependence on $(\tau_n)$ is, of course, unattractive. Another approach to pathwise stochastic integration is described in Willinger and Taqqu (1989) and applied to finance in Willinger and Taqqu (1991).

The integral $\int f dg$ has been studied for processes $g$ that are not semimartingales. In the special case where $f$ and $g$ are continuous and $1/\overline{\text{vex}}\, f + 1/\overline{\text{vex}}\, g > 1$, this integral can be defined pathwise as the Riemann-Stieltjes integral (Young 1936). Bertoin (1989) proves a variant of Itô's lemma for Young's integral. Pathwise stochastic differential and integral equations involving a driving process with trajectories $g$ satisfying $\overline{\text{vex}}\, g < 2$ are considered in recent papers by Mikosch and Norvaiša (2000) and Klingenhöfer and Zähle (1999).

Föllmer's work was the starting point for the work by Bick and Willinger (1994) on dynamic hedging without probabilities, which we mentioned in Chapter 11 (p. 281). They give conditions on a path under which an option can be priced by the Black-Scholes formula, without any assumption about other paths that are not realized. These conditions are much closer, however, to the standard stochastic conditions than our game-theoretic conditions are. For their Proposition 1, they assume that the quadratic variation of the log price path $\ln S(t)$ is linear; in their notation, $[Y,Y]_t = Y^0 + \sigma^2 t$, where $S(t) = \exp Y(t)$ and $[Y,Y]_t$ is the pathwise quadratic variation. They also consider (their Proposition 3) a more general case where $d[S,S]_t = \beta^2(S(t),t)$ for a continuous function $\beta$; but in this case $\beta$ has to be known in advance for successful hedging to be possible.

Bick and Willinger's work was continued by Britten-Jones and Neuberger (1996). In their framework, a trading strategy is a function of the current price of the underlying security and its future variance. The future variance, in our notation, is $(\ln S(t+dt) - \ln S(t))^2 + (\ln S(t+2dt) - \ln S(t+dt))^2 + \cdots$ (they do not assume, however, a constant time step $dt$ between the trades). Their main result is an upper bound on the cost of replicating a European call option provided the jumps $|\ln S(t+dt) - \ln S(t)|$ never exceed some constant. They do not clarify how the investor can know the future variances.

The case where the existing literature manages to price options without probability most convincingly remains, of course, the elementary case we discussed in §1.5, where there are only two possible changes in the price of the underlying security at each step.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# 15

# *The Game-Theoretic Efficient-Market Hypothesis*

In this final chapter, we explore tests and uses of our game-theoretic efficient-market hypothesis. As we explained in §9.4, this hypothesis must always be considered relative to a particular financial market, defined by particular opportunities to buy and sell, and relative to a particular *numéraire* with respect to which prices and capital are measured. The hypothesis says that our imaginary player Skeptic cannot become rich in the particular market relative to the particular *numéraire* without risking bankruptcy. This hypothesis can be tested in a multitude of ways, and each test specifies an event that we can count on if the hypothesis is true. In fact, the hypothesis determines game-theoretic upper and lower probabilities for everything that might happen in the market, and a test is defined by every event that has high lower probability.

Paul Samuelson (born 1915) launched the literature on efficient markets with his suggestion, in 1965, that the price of a stock should follow a martingale.

In the first three sections of this chapter, we study tests of the efficient-market hypothesis corresponding to limit theorems that we studied in Part I. In §15.1 and §15.2, we consider tests that lead to a strong law of large numbers and a law of the iterated logarithm for financial markets. The strong law of large numbers says that returns average to zero in the long run, while the law of the iterated logarithm describes how the average return oscillates as it converges to zero. In §15.3, we prove two weak laws for financial markets, a simple weak law of large numbers and a one-sided central limit theorem.

In the last two sections, we illustrate how upper and lower probabilities determined by the efficient-market hypothesis can make substantive contributions to finance theory. First, in §15.4, we derive some interesting upper and lower probability judgments concerning the relation between risk and return in a market where the same securities are traded over many periods. Then, in §15.5, we discuss how a market that includes options on an index can provide nontrivial upper and lower probabilities for the growth in the index and for a trader's value at risk.

This chapter, in contrast with the four preceding chapters, uses no nonstandard analysis.

## 15.1 A STRONG LAW OF LARGE NUMBERS FOR A SECURITIES MARKET

In this section, we study a strong law of large numbers for financial markets. This strong law says that if an imaginary investor (Skeptic) is unable to become infinitely rich, then the daily gains of an actual investor (Investor) must average to zero.

In general, Skeptic represents a strategy for investing in a particular market. We call Skeptic imaginary in order to caution that there is always a degree of idealization involved in supposing that such a strategy can be implemented without transaction costs or other frictions, and of course the idea that one's capital might become infinitely large is already very much an idealization. Strategies vary a great deal in how closely they can actually be implemented, even within a single well-defined market. In some cases, it might be possible for an actual investor to play the role of Skeptic, to a reasonable approximation. In other cases, Skeptic must be seen as representing the limit of what might be achieved by actual investors.

We assume, in this section, that the market we are considering is well defined; we have specified just what securities are traded each day. We also assume that the total number of all outstanding shares of all these securities is well defined, so that the total value of the market at current prices is well defined. We suppose that investments and capital are measured relative to this total market value. To make this concrete, let us say that a *market unit* is $10^{-10}$ of the total value of the market at any particular time. This will be our *numéraire*; when we say that an investor invests $h$ in a particular security, we will mean that he invests $h$ market units, not $\$h$ or $h$ of any other monetary unit. If the dollar price of the security changes in proportion to the total value of the market, then we will say that the investor's gain from his investment in the security is 0, but if the price of the security increases by 5% relative to the total value of the market, we will say that he has gained $0.05h$; his investment is now worth $1.05h$ market units.

We derive our strong law of large numbers for a securities markets from the game-theoretic martingale strong law of large numbers that we proved in §4.5. It itself, however, is still very general. As we will see, it still implies the measure-theoretic martingale strong law of large numbers.

We begin the section by formulating a protocol in which a fixed number $K$ of securities is traded each day. We prove our strong law of large numbers for this protocol, discuss its applications, and show how it implies a martingale strong law for discrete probability spaces. We then turn to a generalization in which the total value of the market is distributed across an abstract measurable space of securities; our strong law of large numbers in this more general protocol directly implies the measure-theoretic martingale strong law.

## The Securities Market Game

Formally, our strong law for financial markets is a corollary of the strong law for martingales that we studied in §4.5. It involves a game between Skeptic and World, where World is divided into three players:

- Opening Market, who sets the opening prices for each of the market's securities each day,
- Investor, who decides, after hearing the opening prices, on a portfolio of securities to hold for the day, and
- Closing Market, who decides how the price of each security changes and hence how Skeptic's and Investor's investments turn out each day.

Substantively, Investor represents an individual or institutional investor (perhaps a mutual fund) or an investment strategy that we want to test. In the game against Skeptic, however, Opening Market and Investor together play the role that was played by Forecaster in Part I.

Here is our protocol:

SECURITIES MARKET PROTOCOL
**Parameters:** $\mathcal{K}_0 > 0$, natural number $K > 1$
**Players:** Opening Market, Investor, Skeptic, Closing Market
**Protocol:**
FOR $n = 1, 2, \ldots$:
Opening Market selects $m_n \in [0,1]^K$ such that $\sum_{k=1}^K m_n^k = 1$.
Investor selects $g_n \in \mathbb{R}^K$.
Skeptic selects $h_n \in \mathbb{R}^K$.
Closing Market selects $x_n \in [-1, \infty)^K$ such that $m_n \cdot x_n = 0$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + h_n \cdot x_n$.

Here $\cdot$ represents the dot product: if $m$ and $x$ are two vectors of the same length, say $m = (m^1, \ldots, m^K)$ and $x = (x^1, \ldots, x^K)$, then $m \cdot x = \sum_{k=1}^K m^k x^k$.

The protocol assumes that $K$ securities are traded each day in our market. The vectors $m_n$, $x_n$, $g_n$, and $h_n$ are interpreted as follows:

- The vector $m_n = (m_n^1, \ldots, m_n^K)$ specifies how the total value of the market is distributed among the $K$ securities at the beginning of day $n$. In other words, $m_n^k$ is the total value of all the outstanding units of security $k$, as a fraction of the total value of all outstanding units of all $K$ securities. Notice that $m_n^k$ is less than one. If we measure the total value of all the outstanding units of security $k$ in market units, we will get a much larger number, namely $10^{10}m_n^k$.
- The vector $x_n = (x_n^1, \ldots, x_n^K)$ gives the rates of return relative to the market for the $K$ securities on day $n$; $x_n^k = -1$ means security $k$ becomes worthless, $x_n^k = 0$ means its share of the market does not change, and $x_n^k = 0.07$ means its price relative to the market goes up $7\%$.
- The vector $g_n = (g_n^1, \ldots, g_n^K)$ is the portfolio Investor holds during day $n$. Its entries are measured in market units: holding $g_n$ means holding $g_n^k$ market units worth of security $k$, for $k = 1, \ldots, K$. We permit short positions: $g_n^k$ may be negative. For simplicity in writing the protocol, we have assumed that $g_n^k$ can be any real number, even exceeding the total value in market units of security $k$. In practice, we can assume some restrictions on the size of $g_n^k$, but such restrictions will make no difference to our theory; our conclusion will be that Skeptic can achieve a certain goal playing against Investor and Market, and this conclusion will not be affected by any change in the protocol that restricts the moves of Investor.
- Similarly, $h_n$ is the portfolio Skeptic holds during day $n$. In the case of Skeptic, we do want to allow arbitrarily large values of $h_n^k$; as his capital tends to infinity, he may make larger and larger investments. But this is part of the imaginary aspect of Skeptic; as his capital passes from large to astronomical, we must suppose that he is dealing in imaginary money.

The market share of the $k$th security is $m_n^k$ at the beginning of the day and $m_n^k(1+x_n^k)$ at the end of the day, and both must add to one:

$$\sum_{k=1}^{K} m_n^k = 1 \quad \text{and} \quad \sum_{k=1}^{K} m_n^k(1 + x_n^k) = 1.$$

The second of these constraints is expressed in the protocol by the requirement that $m_n \cdot x_n = 0$.

Skeptic's capital $\mathcal{K}_n$ is measured in market units; it changes on the $n$th day by the change of value, in market units, of the portfolio $h_n$ that Skeptic holds that day. We assume that all spare capital is invested in the market, and so the change in Skeptic's capital on day $n$ has the following components:

- the capital invested in security $k$ changes as

$$h_n^k \mapsto h_n^k(1 + x_n^k); \tag{15.1}$$

- the spare capital (which can be negative) is invested in the market, and so does not change,

$$\mathcal{K}_{n-1} - \sum_k h_n^k \mapsto \mathcal{K}_{n-1} - \sum_k h_n^k. \tag{15.2}$$

Summing (15.1) over $k$ and adding (15.2) shows that on day $n$ the capital changes from $\mathcal{K}_{n-1}$ to $\mathcal{K}_{n-1} + h_n \cdot x_n$, which is the expression given in the protocol.

The protocol is obviously a symmetric probability protocol. In fact, it satisfies the conditions of §4.5. The players Opening Market and Investor together constitute Forecaster, and Closing Market is Reality. Skeptic's move space, $\mathbb{R}^K$, is a linear space, and his gain function, $h_n \cdot x_n$, is linear in his move $h_n$. The protocol is coherent, because Closing Market can keep Skeptic from making money by setting all his $x_n$ equal to zero.

The vector $m_n$, which gives the securities' shares of the total value of the market, can also be used to distribute an investment of one market unit across the $K$ securities: we invest the fraction $m_n^k$ of a market unit in security $k$. When $m_n$ is used in this way, we call it the *market portfolio.*

There is a certain indeterminacy in our representation of portfolios by means of vectors. In general, two portfolios $v$ and $v + cm_n$, where $c$ is a constant and $m_n$ is the market portfolio for day $n$, produce the same net gain relative to the market on day $n$:

$$(v + cm_n) \cdot x_n = v \cdot x_n + c(m_n \cdot x_n) = v \cdot x_n.$$

We call two such portfolios *equivalent* for day $n$. One way to eliminate this indeterminacy is to represent portfolios only with vectors that indicate no net use of capital, all purchases of securities being offset by short selling in other securities. We call such vectors *arbitrage portfolios.* Formally, an arbitrage portfolio is a vector $v$ with the property that $\sum_k v^k = 0$. For each portfolio $v$, there is a unique arbitrage portfolio that is equivalent to it for day $n$: $v - cm_n$, where $c = \sum_k v^k$.

We may assume, without loss of generality, that Skeptic is always required to choose an arbitrage portfolio. This makes no difference in what Skeptic can achieve in the game, and because it leaves Skeptic with a linear move space, it does not affect the symmetry of the protocol. We will not, however, assume that Investor necessarily chooses an arbitrage portfolio.

## Interpretations

The simplest and most concrete interpretation of the securities market protocol is obtained by supposing that the same $K$ securities are included in the protocol on each of the $N$ days and that the opening price of each security each morning is its closing price from the preceding evening: $m_{n+1}^k = m_n^k(1 + x_n^k)$. In this case, Skeptic could be an actual investor, except for the consideration of transaction costs and until his capital becomes too large relative to the entire market. We can specify arbitrarily the $K$ securities that we want to study for $N$ days; they need not include all the securities traded in any existing exchange.

The formulation also permits us, however, to change the $K$ securities in the protocol, perhaps in the same way as stock exchanges and market indexes update their listings: we throw out securities that have declined relative to the total market, and we bring in newly valuable securities. If such shifts change the total value of the market[1] in dollars (or whatever monetary unit is relevant), Skeptic becomes more imaginary, however; the formula $\mathcal{K}_n := \mathcal{K}_{n-1} + h_n \cdot x_n$ assumes that his capital in market units at the end of day $n-1$, $\mathcal{K}_{n-1}$, is also his capital in market units on the morning of day $n$, and if the total dollar value of the market has increased overnight because of new listings, then we must give Skeptic enough dollars for this to remain true.

If we want to change the securities from day to day, then we can even change the total number $K$ from day to day. This will become clear when we generalize our results to the case where the total value of the market is distributed across an arbitrary measurable space of securities.

In fact, the protocol is very broad, with many possible interpretations. As we will see shortly, it even accommodates a general sequential gambling scenario (essentially what we called generalized coin tossing in Chapter 8), and so in a certain sense it contains the theory of probability.

## The Finance-Theoretic Strong Law of Large Numbers

The following proposition, a corollary of Proposition 4.4 (p. 92), is our finance-theoretic strong law.

**Proposition 15.1** *Skeptic can force the event*

$$\sum_{n=1}^{\infty} n^{-2} \sum_{k=1}^{K} \frac{(g_n^k)^2}{m_n^k} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g_i \cdot x_i = 0. \tag{15.3}$$

To make this proposition stronger, we resolve the possible uncertainty $0/0$ in the fraction on the left-hand side to 0. Notice that the proposition remains true if "can force" is replaced by "can Borel force".

*Proof* If the $g_n$ are arbitrage portfolios, then it suffices to apply Proposition 4.4 to the martingale

$$\mathcal{S}_n := \sum_{i=1}^{n} g_i \cdot x_i, \tag{15.4}$$

using the fact, established in the next lemma, that

$$\mathcal{A}_n := \sum_{i=1}^{n} \sum_{k=1}^{K} \frac{(g_i^k)^2}{m_i^k} \tag{15.5}$$

is a quadratic supervariation for $\mathcal{S}_n$.

[1]Such a change will usually be negligible in the idealized case of an index always containing $K$ companies with the highest market capitalization.

When the $g_n$ are arbitrage portfolios, the substitution of equivalent portfolios $g_n + c_n m_n$ yields

$$\sum_{k=1}^{K} \frac{(g_n^k + c_n m_n^k)^2}{m_n^k} = \sum_{k=1}^{K} \frac{(g_n^k)^2}{m_n^k} + c_n^2 \geq \sum_{k=1}^{K} \frac{(g_n^k)^2}{m_n^k}.$$

So the sum on the left-hand side of the implication (15.3) can only be increased by substituting equivalent portfolios, while the sum on the right-hand side is not changed. So if the implication holds when the $g_n$ are arbitrage portfolios, it holds in general. ∎

**Lemma 15.1** *If the $g_n$ are arbitrage portfolios, then (15.5) is a quadratic supervariation of (15.4).*

*Proof* We must show that Skeptic has a strategy that will ensure that

$$\Delta(\mathcal{S}^2)_n - \Delta\mathcal{A}_n \leq h_n \cdot x_n,$$

or

$$(g_n \cdot x_n)^2 + 2\mathcal{S}_{n-1}(g_n \cdot x_n) - \sum_k \frac{(g_n^k)^2}{m_n^k} \leq h_n \cdot x_n$$

for $n \geq 1$. This is equivalent to his having a strategy that ensures

$$(g_n \cdot x_n)^2 - \sum_k \frac{(g_n^k)^2}{m_n^k} \leq h_n \cdot x_n.$$

Thus our task is to find a move $h$ for Skeptic, following moves $m$ by Opening Market and $g$ by Investor, such that

$$(g \cdot x)^2 - h \cdot x \leq \sum_k \frac{(g^k)^2}{m^k} \tag{15.6}$$

will hold no matter what move $x$ is chosen by Closing Market.

The case $\exists k : m^k = 0 \;\&\; g^k > 0$ is trivial, so we assume $\forall k : m^k = 0 \Rightarrow g^k = 0$. Excluding the $k$ for which $m^k = 0$ from consideration (for such $k$ Skeptic can set $h^k := 0$), we further assume that $m^k > 0$ for all $k$.

The move $x \in \mathbb{R}^K$ must satisfy

$$m \cdot x = 0 \;\&\; \forall k : x^k \geq -1. \tag{15.7}$$

The left-hand side of (15.6) is a convex function of $x$, so it suffices for Skeptic to ensure that (15.6) holds for the vertices of the polyhedron (15.7). These vertices are $x_j$, $j = 1, \dots, K$, with

$$x_j^k = \begin{cases} 1/m^j - 1 & \text{if } k = j \\ -1 & \text{otherwise.} \end{cases}$$

The inequality (15.6) is satisfied for the vertex $x_j$ if

$$\left(\frac{g^j}{m^j}\right)^2 - \frac{h^j}{m^j} + \sum_k h^k \leq \sum_k \frac{(g^k)^2}{m^k}.$$

(Here we have used the assumption that $g$ is an arbitrage portfolio.) This is satisfied if we set

$$h^j := \frac{(g^j)^2}{m^j} - m^j \sum_k \frac{(g^k)^2}{m^k},$$

thus defining an arbitrage portfolio $h$ for Skeptic. ∎

Proposition 15.1 says that Skeptic can force $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g_i \cdot x_i = 0$ on paths where Investor's moves are not too unusual—not too greatly and too consistently deviant from the market portfolio. To illustrate this, consider the case where Investor chooses a particular security, say the $k$th one, and always invests one market unit in this security. In other words, he always chooses the portfolio $g$ given by

$$g^j := \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise.} \end{cases}$$

Substituting this portfolio for all the $g_n$ in (15.3), we find that Skeptic can force the event

$$\sum_{n=1}^{\infty} \frac{1}{n^2 m_n^k} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i^k = 0.$$

For example, Skeptic can force $\lim_{n\to\infty}(1/n)\sum_{i=1}^{n} x_i^k = 0$ on paths where $m_n^k$ is bounded below by $\epsilon \ln^2 n/n$ for some $\epsilon > 0$.

## Is the Finance-Theoretic Strong Law Satisfied Empirically?

Many studies of the stochastic efficient-markets hypothesis test the hypothesis using average returns for well-defined investment strategies or for actual investors, such as mutual funds. Provided that the larger market in which the strategies or investors operate is well defined, Proposition 15.1 tells us these tests can be regarded, at least to a first approximation, as tests of the game-theoretic efficient-market hypothesis for these particular markets with a *numéraire* proportional to the total value of the market.

It is beyond the scope of this book to review these studies in detail in an attempt to judge which game-theoretic efficient-market hypotheses stand up to these tests. There is a widely held consensus, however, that funds that track market indexes such as the S&P 500 in the United States or the MSCI in Europe can do so successfully and usually outperform, often dramatically, other funds that follow different strategies [212]. The evidence on which this consensus is based may also serve as evidence that the markets defined by these indexes do satisfy our game-theoretic efficient-market hypothesis.

Of course, Proposition 15.1 does not merely say that an investor should be unable to beat the market; it also says that the investor should do as well as the market. Thus the below-market returns often achieved by individual investors [10] and funds [160, 211, 50, 2] must be taken as evidence against the game-theoretic efficient-market hypothesis except in cases where it can be attributed to violations of the securities market protocol (transactions costs or trading outside the index that is being tested, perhaps in emerging markets or small value stocks) or violations of the left-hand side of (15.3) (excessively wild trading).

### Horse Races

Although it may seem rather specialized, our strong law for a securities market actually contains a martingale strong law for elementary (discrete) probability theory as a special case. This becomes clear if we think about how the securities market game can be used to model a sequence of horse races [61, 169].

Suppose we conduct a horse race every day, always with $K$ horses. At the beginning of day $n$, Opening Market posts the current odds for the day's race: $m_n^k$ is the probability that the $k$th horse will win. The probabilities $m_n^1, \ldots, m_n^K$ add to one. Investor can bet on horse $k$ at odds $m_n^k : (1 - m_n^k)$. Putting \$1 on horse $k$ produces

$$\begin{cases} 1/m_n^k - 1 & \text{if horse } k \text{ wins the race} \\ -1 & \text{if horse } k \text{ loses the race.} \end{cases}$$

So if Investor puts $g_n^k$ on horse $k$, for $k = 1, 2, \ldots, K$, his total gain will be $g_n \cdot x_n$, where $x_n$ is the vector given by

$$x_n^k = \begin{cases} 1/m_n^k - 1 & \text{if } k = \text{winner} \\ -1 & \text{otherwise.} \end{cases} \tag{15.8}$$

The vector $x_n$ satisfies the constraint $m_n \cdot x_n = 0$. So the game played by Investor in the horse race is a special case of our securities market game, special only inasmuch as the moves by Closing Market, $x_n$, must be of the special form (15.8).

From the viewpoint of probability theory, the sequence of winners can be thought of as coming from a probability space $\{1, \ldots, K\}^\infty$ in which $m_n^k$ is the conditional probability that the $n$th element of the sequence will be $k$ given the first $n-1$ choices. Suppose $\xi_n = (\xi_n^1, \ldots, \xi_n^K)$ is a sequence of martingale differences in this space. This means simply that $\sum_{k=1}^K \xi_n^k m_n^k = 0$ for all $n$—that is, the vectors $\xi_n m_n$ are arbitrage portfolios. Proposition 15.1 with $g_n^k = \xi_n^k m_n^k$ says that

$$\sum_{n=1}^{\infty} n^{-2} \sum_{k=1}^{K} (\xi_n^k)^2 m_n^k < \infty \Longrightarrow \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \xi_i = 0 \tag{15.9}$$

holds almost surely ($\xi_i$ on the right-hand side is used, following the usual practice in probability theory, to mean $\xi_i^{k_i}$, where $k_i$ is the $i$th race's winner). This is the strong law for martingales for elementary probability theory; the left-hand side of (15.9) is the usual condition on the variances.

If $K = 2$, Opening Market always gives the two horses the same chance of winning ($m_n^1 = m_n^2 = 1/2$), and Investor always puts $1/2$ on the second horse ($g_n^1 = -1/2$ and $g_n^2 = 1/2$, so that $\xi_n^1 = -1$ and $\xi_n^2 = 1$), then the convergence on the left-hand side of (15.9) is guaranteed, and the martingale strong law reduces to Borel's strong law.

### Markets with Infinitely Many Securities

Real markets can only contain finitely many securities, but there is no obstacle to a generalization of our mathematical results in the direction of measure theory, so that the total value of the market is distributed, on each day, across an arbitrary measurable space of securities.

For this measure-theoretic generalization, it is convenient to rewrite the protocol using the following quantities:

- $G_n^k := g_n^k / m_n^k$ instead of $g_n^k$ (this means that we measure an investment in security $k$ in units proportional to the total market value of this security),
- $H_n^k := h_n^k / m_n^k$ instead of $h_n^k$, and
- $X_n^k := (1 + x_n^k) m_n^k$ instead of $x_n^k$ (this is security $k$'s share of the total value of the market at the end of day $n$).

Skeptic's gain from his portfolio $H_n$ will be

$$h_n \cdot x_n = \sum_k h_n^k x_n^k = \sum_k (H_n^k m_n^k)(X_n^k / m_n^k - 1) = \sum_k H_n^k X_n^k - \sum_k H_n^k m_n^k.$$

Passing from this discrete notation to the corresponding but more general measure-theoretic notation, we obtain the following protocol.

ABSTRACT SECURITIES MARKET PROTOCOL
**Parameters:** Measurable space $(\Omega_\bullet, \mathcal{F})$, Skeptic's initial capital $\mathcal{K}_0 > 0$
**Players:** Opening Market, Investor, Skeptic, Closing Market
**Protocol:**
FOR $n = 1, 2, \ldots$:
Opening Market selects $m_n \in \mathcal{P}(\Omega_\bullet)$.
Investor selects measurable $G_n : \Omega_\bullet \to \mathbb{R}$.
Skeptic selects measurable $H_n : \Omega_\bullet \to \mathbb{R}$.
Closing Market selects $X_n \in \mathcal{P}(\Omega_\bullet)$.

$$\mathcal{K}_n := \begin{cases} \mathcal{K}_{n-1} + \int H_n d(X_n - m_n) & \\ \qquad \text{if } \int H_n dX_n \text{ and } \int H_n dm_n \text{ are defined} \\ -\infty & \text{if } \int H_n dm_n \text{ is undefined} \\ \infty & \text{otherwise.} \end{cases}$$

Recall that $\mathcal{P}(B)$ stands for the set of all probability measures on the measurable space $B$. We have divided the job of ensuring that the integral $\int H_n d(X_n - m_n)$ is defined between Skeptic and Closing Market in a natural way: Skeptic is required to ensure that $\int H_n dm_n$ is defined (i.e., $\int |H_n| dm_n < \infty$); if he complies, Closing Market is required to ensure that $\int H_n dX_n$ is defined (i.e., $\int |H_n| dX_n < \infty$).

The analog of Lemma 15.1 (p. 357) is:

**Lemma 15.2** *The process*

$$\mathcal{A}_n := \sum_{i=1}^{n} \int G_i^2 dm_i \tag{15.10}$$

*is a quadratic supervariation of the martingale*

$$\mathcal{S}_n := \sum_{i=1}^{n} \int G_i d(X_i - m_i).$$

*Proof* We are required to prove that, for some $H_n$,

$$\left(\int G_n d(X_n - m_n)\right)^2 \leq \int G_n^2 dm_n + \int H_n d(X_n - m_n).$$

If $\int G_n^2 dm_n = \infty$, this is trivially true, so we will assume that $\int G_n^2 dm_n$ is finite. We simplify the notation by dropping the subscript $n$; our goal is to find $H$ for which

$$\left(\int Gd(X - m)\right)^2 - \int G^2 dm - \int Hd(X - m) \leq 0. \tag{15.11}$$

The left-hand side is a convex function of $X$; therefore, by Jensen's inequality (see the next paragraph), it is sufficient to establish this inequality for $X$ that are concentrated at one point $\omega \in \Omega$:

$$\left(G(\omega) - \int Gdm\right)^2 - \int G^2 dm - \left(H(\omega) - \int Hdm\right) \leq 0. \tag{15.12}$$

To show that such an $H$ exists, it is enough to prove that the $m$-expected value of

$$\left(G(\omega) - \int Gdm\right)^2 - \int G^2 dm \tag{15.13}$$

is nonpositive, which is obvious: the variance never exceeds the second moment. We will take (15.13) as $H$; notice that $\int Hdm$ will be defined.

In conclusion, we will discuss the application of Jensen's inequality above. If the function on the left-hand side of (15.11) is $f(X)$ and $[\omega]$ is the probability distribution concentrated at $\omega$, we have

$$f(X) = f\left(\int [\omega] dX\right) \leq \int f([\omega]) dX \leq 0$$

when $f([\omega]) \leq 0$ for all $\omega$. The problem with this argument is that the domain of $f$ is the set $\mathcal{P}(\Omega_\bullet)$ of all probability distributions in $\Omega_\bullet$, whereas the most usual form of Jensen's inequality is only applicable to convex functions $f : \mathbb{R} \to \mathbb{R}$. However, the result we need can easily be derived from the usual proof of Jensen's inequality (as in, e.g., [287]): assuming (15.12),

$$\begin{aligned} 0 \geq f([\omega]) &= \left(\int Gd((X - m) + ([\omega] - X))\right)^2 - \int G^2 dm - \int Hd([\omega] - m) \\ &= \left(\int Gd(X - m)\right)^2 + 2\int Gd(X - m) \int Gd([\omega] - X) + \left(\int Gd([\omega] - X)\right)^2 \\ &\qquad - \int G^2 dm - \int Hd([\omega] - m) \\ &\geq \left(\int Gd(X - m)\right)^2 + 2\int Gd(X - m) \int Gd([\omega] - X) \\ &\qquad - \int G^2 dm - \int Hd([\omega] - m) \end{aligned}$$

$$= \left(\int Gd(X - m)\right)^2 + 2\int Gd(X - m)\left(G(\omega) - \int GdX\right)$$
$$- \int G^2 dm - \left(H(\omega) - \int Hdm\right);$$

integrating the last term of this chain over the probability distribution $X(d\omega)$ gives (15.11). We have been assuming that $\int GdX$ and $\int HdX$ are defined; the latter, which is Closing Market's responsibility, implies the former (and even $\int G^2 dX < \infty$). ▮

This lemma immediately implies the following strong law for the abstract securities market protocol.

**Proposition 15.2 (Fuzzy Strong Law of Large Numbers)** *Skeptic can force the event*

$$\sum_{n=1}^{\infty} n^{-2} \int G_n^2 dm_n < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int G_i d(X_i - m_i) = 0. \qquad (15.14)$$

We call it fuzzy because the outcome of each round of play is another probability distribution on $\Omega_\bullet$ rather than a crisply defined point in $\Omega_\bullet$. As usual, the forcing strategy for Skeptic can be chosen measurable.

### Deducing the Martingale Strong Law of Large Numbers

As we mentioned at the beginning of the section, our strong law for the abstract securities market game (the fuzzy strong law we have just proved, with Skeptic's forcing strategy understood to be measurable) implies the usual measure-theoretic martingale strong law.

Suppose, indeed, that $x_1, x_2, \ldots$ is a measure-theoretic martingale difference with respect to a sequence of $\sigma$-algebras $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$; let $(\Omega, \mathcal{F}, P)$ be the underlying probability space. To apply our fuzzy strong law, set $\Omega_\bullet := \mathbb{R}$ and assume that Opening Market, Investor, and Closing Market play the following strategies (deterministic for Investor and stochastic, i.e., dependent on $\omega \in \Omega$, for the Markets): $m_n$ is always a regular conditional distribution of $x_n$ with respect to $\mathcal{F}_{n-1}$ evaluated at $\omega$ (here we rely on the result in measure-theoretic probability that regular probability distributions do exist; see, for example, [287]), $G_n$ is the identity function ($G_n(t) = t$ for all $t \in \Omega_\bullet$), and $X_n$ is concentrated at one point, $x_n(\omega)$. The implication (15.14) then reduces to

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\left(x_n^2 \mid \mathcal{F}_{n-1}\right)}{n^2} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0. \qquad (15.15)$$

Since Skeptic's strategy that forces (15.15) is measurable, it gives a supermartingale (see Proposition 4.2 on p. 86); therefore, our fuzzy strong law implies that (15.15) holds almost surely. This is the usual measure-theoretic martingale strong law.

## 15.2 A LAW OF THE ITERATED LOGARITHM FOR A SECURITIES MARKET

We now turn to the law of the iterated logarithm for our securities market. This law explains how Skeptic can control the oscillation of Investor's average gain relative to the market as it converges to zero. We derive the law first for our discrete securities market ($K$ securities) and then for our abstract securities market (a measurable space of securities).

Here is the law for the discrete market.

**Proposition 15.3** *In the securities market protocol, Skeptic can force*

$$\left(\mathcal{A}_n \to \infty \;\&\; |g_n \cdot x_n| = o\left(\sqrt{\frac{\mathcal{A}_n}{\ln\ln \mathcal{A}_n}}\right)\right) \Longrightarrow \limsup_{n\to\infty} \frac{|\sum_{i=1}^n g_i \cdot x_i|}{\sqrt{2\mathcal{A}_n \ln\ln \mathcal{A}_n}} \le 1. \tag{15.16}$$

Because $\mathcal{A}_n$, as given by (15.5), is a quadratic supervariation of $\sum_{i=1}^n g_i \cdot x_i$ (Lemma 15.1), this proposition is just a special case of Proposition 5.2 (p. 118).

The conditions on the left-hand side of the implication (15.16) can be interpreted as follows:

- The condition $\mathcal{A}_n \to \infty$ says that Investor does not play too timidly. It is implied, for example, by $\sum_{i=1}^\infty \|g_i\|^2 = \infty$.
- The condition $|g_n \cdot x_n| = o(\sqrt{\mathcal{A}_n / \ln\ln \mathcal{A}_n})$, which limits the growth of $|g_n \cdot x_n|$, requires, in general, that $g_n$ and $x_n$ do not get too far from zero too fast—that is, neither Investor nor Market play too wildly. It is implied by $|g_n \cdot x_n| = O(1)$ (assuming $\mathcal{A}_n \to \infty$).

As an illustration of the law, consider the case where Investor always puts one market unit on security $k$. This means that his gain on the day $n$ is always $x_n^k$. The arbitrage portfolio representing his bet on day $n$, $g_n$, is given by $g_n^k = 1 - m_n^k$ and $g_n^j = -m_n^j$ for $j \neq k$. We find

$$\mathcal{A}_n = \sum_{i=1}^n \left(\frac{(1-m_i^k)^2}{m_i^k} + \sum_{j\neq k} \frac{(-m_i^j)^2}{m_i^j}\right) = \sum_{i=1}^n \left(\frac{(1-m_i^k)^2}{m_i^k} + (1-m_i^k)\right)$$
$$= \sum_{i=1}^n \frac{1-m_i^k}{m_i^k}.$$

This will tend to infinity if the market share of the single security $k$ never becomes too large—say if $m_n^k \le 1-\epsilon$ for some $\epsilon > 0$ and all $n$. If we assume that $|x_n^k| < 1$ for all $n$, then Proposition 15.3 implies that Skeptic can force the event

$$\limsup_{n\to\infty} \frac{|\sum_{i=1}^n x_i^k|}{\sqrt{2\sum_{i=1}^n \frac{1-m_i^k}{m_i^k} \ln\ln \sum_{i=1}^n \frac{1-m_i^k}{m_i^k}}} \le 1. \tag{15.17}$$

If we assume further that $m_n^k \geq \epsilon$, so that the total value of the $k$th security is never overwhelmingly dominated by that of the other securities, then the cumulative gain grows not faster than $\sqrt{n \ln \ln n}$,

$$\left| \sum_{i=1}^{n} x_i^k \right| = O\left(\sqrt{n \ln \ln n}\right), \quad n \to \infty.$$

We are not aware of empirical studies that have tested the efficient-market hypothesis by looking for violations of the limits given by Equations (15.16) or (15.17). But this could be done, using some of the longer series of data now available, provided we restate Proposition 15.3 in finitary terms; see §5.6 and [322].

Lemma 15.2 implies the following abstract game-theoretic law of the iterated logarithm:

**Proposition 15.4 (Fuzzy Law of the Iterated Logarithm)** *Skeptic can force the event*

$$\left( \mathcal{A}_n \to \infty \;\&\; \left| \int G_n d(X_n - m_n) \right| = o\left( \sqrt{\frac{\mathcal{A}_n}{\ln \ln \mathcal{A}_n}} \right) \right)$$
$$\Longrightarrow \limsup_{n \to \infty} \frac{\left| \sum_{i=1}^{n} \int G_i d(X_i - m_i) \right|}{\sqrt{2\mathcal{A}_n \ln \ln \mathcal{A}_n}} \leq 1,$$

*where $\mathcal{A}_n$ is defined by (15.10).*

## 15.3 WEAK LAWS FOR A SECURITIES MARKET

We turn now to weak finance-theoretic laws, first a weak law of large numbers and then a one-sided central limit theorem.

### A Weak Law of Large Numbers for a Securities Market

Here is a simple form of the weak law of large numbers for a securities market.

**Proposition 15.5** *Let $C > 0$ be a (large) constant, $\epsilon > 0$ be a (small) constant, and $\mathcal{S}_n$ and $\mathcal{A}_n$ be defined by (15.4) and (15.5). Suppose the players in the securities market protocol play a fixed number $N$ of trials. Then*

$$\overline{\mathbb{P}}\left\{ \mathcal{A}_N \leq C \;\&\; \left| \frac{\mathcal{S}_N}{N} \right| \geq \epsilon \right\} \leq \frac{C}{\epsilon^2 N^2}. \tag{15.18}$$

*Proof* Adapting the proof of Proposition 6.1 on p. 125, we consider the nonnegative supermartingale

$$\mathcal{L}_n := \begin{cases} \frac{\mathcal{S}_n^2 + C - \mathcal{A}_n}{C} & \text{if } n \leq \tau \\ \mathcal{L}_\tau & \text{otherwise} \end{cases}$$

with $\mathcal{L}_0 = 1$, where $\tau$ is the last trial $n$ with $\mathcal{A}_n \leq C$:

$$\tau := \max\{n \mid \mathcal{A}_n \leq C\}. \tag{15.19}$$

On the event $\mathcal{A}_N \le C$ & $|\mathcal{S}_N|/N \ge \epsilon$, we have $\mathcal{L}_N \ge N^2\epsilon^2/C$. ∎

The inequality (15.18) suggests that the efficient-market hypothesis should be rejected when $\mathcal{S}_N/N$, Investor's average return relative to the market, is not small in absolute value and our other conditions are satisfied. But the presence of $C$ makes the inequality difficult to interpret more specifically, because we are told nothing about the relation between $C$ and $N$. (In the coin-tossing case, in contrast, we can take $C = N$.) So it is convenient to replace (15.18) with

$$\overline{\mathbb{P}}\left\{|\mathcal{S}_\tau| \ge \sqrt{C/\delta}\right\} \le \delta, \qquad (15.20)$$

where $\delta > 0$ is some conventional threshold, such as 1%; now $N$ becomes irrelevant: the game can last infinitely long. It is possible that $\tau$ in (15.18) will be infinite, and so $\mathcal{S}_\tau$ may not be defined; (15.20) says that the upper probability that it is defined and satisfies $|\mathcal{S}_\tau| \ge \sqrt{C/\delta}$ does not exceed $\delta$. We will use this convention throughout. Equation (15.20) can be proven in the same way as (15.18): on the event $|\mathcal{S}_\tau| \ge \sqrt{C/\delta}$ we have $\mathcal{L}_\tau \ge 1/\delta$.

The efficient-market hypothesis can be tested by looking at whether mutual funds or investment strategies do better than Investor is supposed to do in our game. It might be tempting to look at a large number of mutual funds and see what proportion of them achieve $|\mathcal{S}_\tau| \ge \sqrt{C/\delta}$, but this would not tell us anything, because different funds or strategies can be arbitrarily correlated. A more interesting approach is to consider a single fund or strategy over successive periods. The following strong law points to how this can be done:

**Proposition 15.6** *Let $C$ and $\delta$ be two positive constants, and let $\mathcal{S}_n$ and $\mathcal{A}_n$ be defined by (15.4) and (15.5). Let $\tau_0 = 0$ and $\tau_i$, $i = 1, 2, \ldots$, be the last time that $\mathcal{A}_{\tau_i} - \mathcal{A}_{\tau_{i-1}} \le C$ (this is an inductive definition). Then*

$$\limsup_{I\to\infty} \frac{\#\left\{i = 1, \ldots, I \mid |\mathcal{S}_{\tau_i} - \mathcal{S}_{\tau_{i-1}}| \ge \sqrt{C/\delta}\right\}}{I} \le \delta.$$

*Proof* It suffices to combine (15.20) with the one-sided strong law of large numbers in Chapter 3 (Proposition 3.4 on p. 73). ∎

This is a strong law, but a corresponding weak law holds for finitely many rounds.

### A One-Sided Central Limit Theorem for a Securities Market

We now establish a one-sided finance-theoretic central limit theorem, generalizing the one-sided central limit theorem of §6.3.

For any constant $C$ let $\tau_C$ be the moment when $\mathcal{A}_n$ (see (15.5) on p. 356) reaches the value $C$,

$$\tau_C := \min\{n \mid \mathcal{A}_n \ge C\}$$

(we could define $\tau_C$ in the same way as $\tau$ in (15.19), but now we follow what we did in Chapter 7, Example 5; both definitions are good for our purpose), and let $\mathcal{X}_C$ be a normalized version of (15.4),

$$\mathcal{X}_C := \frac{1}{\sqrt{C}} \sum_{i=1}^{\tau_C} g_i \cdot x_i.$$

**Proposition 15.7** *Consider the securities market protocol with the additional restriction that Investor and Closing Market are required to choose $g_n^k$ and $x_n^k$ bounded in absolute value by $c\sqrt{m_n^k}$ and $c$, respectively, where $c$ is a given constant. Let $U$ be the indicator function of a closed set $E \subseteq \mathbb{R}$ that does not contain 0. Let $\tilde{U}(s, D)$ be the solution to the heat equation $\partial\tilde{U}(s, D)/\partial D = (1/2)\partial^2\tilde{U}(s, D)/\partial s^2$ with the initial condition $\tilde{U}(s, 0) = 0$, $\forall s \notin E$, and the boundary condition $\tilde{U}(s, D) = 1$, $\forall s \in E, D > 0$. Then as $C \to \infty$,*

$$\overline{\mathbb{P}}[\mathcal{X}_C \in E] \to \tilde{U}(0, 1).$$

*Proof* Without loss of generality we can assume that $E = (-\infty, a] \cup [b, \infty)$, $E = (-\infty, a]$, or $E = [b, \infty)$, for some $a < 0 < b$. (We can define $a := \sup(E \cap (-\infty, 0))$ and $b := \inf(E \cap (0, \infty))$.) There exist smooth functions $V \leq \mathbb{I}_E$ and $W \geq \mathbb{I}_E$ that are constant outside a finite interval such that $\tilde{V}(0, 1)$ and $\tilde{W}(0, 1)$ are arbitrarily close to $\tilde{U}(0, 1)$ ($\tilde{V}$ and $\tilde{W}$ are defined as LM $v$ and LM $w$, respectively, where $v(s, D) := V(s)$ and $w(s, D) := W(s)$). It is sufficient to prove that, for any fixed $\epsilon > 0$ and large enough $C$, $\overline{\mathbb{E}}[W(\mathcal{X}_C)] \leq \tilde{W}(0, 1) + \epsilon$ and $\overline{\mathbb{E}}[V(\mathcal{X}_C)] \geq \tilde{V}(0, 1) - \epsilon$. The proofs will be analogous to the proof of the one-sided central limit theorem in Chapter 6. Set

$$S_n := \frac{1}{\sqrt{C}} \sum_{i=1}^{n} g_i \cdot x_i, \quad D_n := \frac{C - \mathcal{A}_n}{C}. \tag{15.21}$$

*Proof of* $\overline{\mathbb{E}}[W(\mathcal{X}_C)] \leq \tilde{W}(0, 1) + \epsilon$. As before, choose large $C > 0$ and small $\delta \in (0, 1)$; our first goal is to show that Skeptic, starting with $\tilde{W}(0, 1) + \epsilon$, can attain capital at least $\tilde{W}(S_n, \delta) + \epsilon/2$ when $(S_n, D_n)$ hits the set (6.28) (p. 140). Find a smooth superparabolic $\overline{U} \leq \tilde{W}$ sufficiently close to $\tilde{W}$ on $[-C, C] \times [\delta, 1]$. As usual, the proof is based on (6.21) (p. 132); the most important difference from the situation of Chapter 6 is that the second addend in (6.21), instead of being zero, is bounded from above by the return of some portfolio; this follows from $\mathcal{A}_n$ being a quadratic supervariation and the fact that $\partial\overline{U}/\partial D$ is never negative (see Lemma 6.2 on p. 139).

Let us prove that the cumulative effect of the last four addends in (6.21) is negligible. First notice that the conditions $|g_n^k| \leq c\sqrt{m_n^k}$ and $|x_n^k| \leq c$ imply that $|dD_n| = O(C^{-1})$ and $|dS_n| = O(C^{-1/2})$. The last addend is easiest: $\sum_n |dD_n| \leq 1 + O(C^{-1})$, and so $\sum_n (dD_n)^2 = O(C^{-1})$. The third and fourth addends are negligible because $|dS'_n|$ and $|dD'_n|$ are negligible and $\sum_n (dS_n)^2$ is moderately large with lower probability close to 1 (this last point follows from the game-theoretic Markov inequality and the fact that, according to Lemma 15.1 on p. 357, the upper expectation of $\sum_n (dS_n)^2$ is bounded above by $C^{-1}(C + O(1)) = O(1)$). For the penultimate addend, we apply our usual argument for the weak law of large numbers. Lemma 15.1 implies that the upper expectation of

$$\left( \sum_{n=1}^{\tau_C} \frac{\partial^2 \overline{U}}{\partial s \partial D}(S'_n, D'_n) dS_n dD_n \right)^2$$

is $O(C^{-2})$; therefore, by the game-theoretic Markov inequality, the sum over $n$ of the penultimate addend does not exceed $AC^{-1}$ in absolute value with lower probability arbitrarily close to 1 (provided $A$ is sufficiently large).

Similar arguments show that: (1) with high lower probability, $S_n$ will not change much after $(S_n, D_n)$ hits the boundary $D = \delta$ of (6.28); (2) with high lower probability the boundary $|s| = C$ will not be hit. ∎

*Proof of* $\overline{\mathbb{E}}[V(\mathcal{X}_C)] \geq \tilde{V}(0,1) - \epsilon$. This is simple, since in the current game Skeptic's opponents have much more freedom than in the one-sided game of Chapter 6. The proof of the negative part of Theorem 6.1 (with $u$ defined by $u(s, D) := V(s)$, $N$ replaced by $C$, and other obvious changes) works in the current setting if we assume the following horse-race scenario:

- Opening Market's move is always $m_n = (0.5, 0.5, 0, 0, \ldots)$ (only the first two stocks will matter);
- Investor's move is always $g_n = (-0.5, 0.5, 0, 0, \ldots)$;
- at the beginning of the game Closing Market (who plays the role of Chapter 6's Reality) chooses $x_n = (\pm 1, \mp 1, 0, 0, \ldots)$ with a suitable sign; and at the end of the game he chooses $x_n = (0, 0, \ldots)$.

As before, $S_n$ and $D_n$ should be defined by (15.21). ∎

This proposition can be combined with a one-sided law of large numbers, in the spirit of Proposition 15.6.

## 15.4 RISK VS. RETURN

An important aspect of stochastic asset pricing theories is the relation between risk and return: riskier securities tend to bring higher returns. In this section, we study to what extent this relation follows from our game-theoretic efficient-market hypothesis.

For this study, we leave aside the very special protocol of the preceding sections, which assumes a well-defined total value for the market in which Skeptic is trading and takes the *numéraire* proportional to this total value. We now use a finite-horizon protocol, in which the *numéraire* can be chosen arbitrarily. This protocol tracks the capital of Investor as well as that of Skeptic and does not allow Investor to go short in securities. The requirement that Investor cannot go short might appear restrictive, but the protocol produces interesting results already in the case where $K = 1$ and $g_n = 1$ for all $n$ (Investor always holds one unit worth of the only security included in the protocol).

We write $\mathcal{I}_n$ for Investor's capital, dispose of Opening Market, and rename Closing Market to Market.

FINITE-HORIZON SECURITIES MARKET PROTOCOL

**Parameters:** $N, \mathcal{K}_0 > 0, \mathcal{I}_0 > 0$

**Players:** Investor, Skeptic, Market

**Protocol:**

FOR $n = 1, 2, \ldots, N$:

Investor selects $g_n \in [0, \infty)^K$.

Skeptic selects $h_n \in [0,\infty)^K$.
Market selects $x_n \in [-1,\infty)^K$.
$\mathcal{I}_n := \mathcal{I}_{n-1} + g_n \cdot x_n$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + h_n \cdot x_n$.

We set

$$r_n := \frac{g_n \cdot x_n}{\mathcal{I}_{n-1}},$$

and we call it Investor's *return* on his capital on day $n$. We also set

$$\mu := \frac{1}{N}\sum_{n=1}^{N} r_n$$

and

$$\sigma^2 := \frac{1}{N}\sum_{n=1}^{N}(r_n - \mu)^2, \quad \sigma_0^2 := \frac{1}{N}\sum_{n=1}^{N} r_n^2;$$

$\mu$ and $\sigma^2$ are the empirical mean and variance, respectively, of Investor's returns, and $\sigma_0^2$ is the uncentered empirical variance usually used in the finance literature. We will call $\sigma_0$ the *empirical volatility* and use it in Propositions 15.8 and 15.9 below; to state these propositions in terms of empirical variance, replace $\sigma_0^2$ by $\sigma^2 + \mu^2$ (but the difference $\mu^2$ between $\sigma_0^2$ and $\sigma^2$ is usually negligible: see, e.g., the figures for Microsoft and IBM below).

**Proposition 15.8** *For any* $\alpha > 0$,

$$\underline{\mathbb{P}}\left\{\mu < \frac{\sigma_0^2}{2} - \frac{\ln\alpha}{N} + \left[\frac{1}{3N}\sum_{n:r_n<0}\left|\frac{r_n}{1+r_n}\right|^3\right]\right\} \geq 1-\alpha. \tag{15.22}$$

*Proof* The proof is based on an elementary idea: the daily logarithmic returns, $\ln(1+r_n)$, are more useful than the $r_n$ themselves because they are additive and because $\ln(1+r_n) \approx r_n - r_n^2/2$ will be adversely influenced by high volatility.

Suppose at every trial $n$ Skeptic takes his portfolio $h_n = g_n\mathcal{K}_0/\mathcal{I}_0$ to be the same as Investor's portfolio $g_n$ but scaled to his initial capital. Then $\mathcal{K}_n = (1+r_n)\mathcal{K}_{n-1}$. So the lower probability of the event

$$\prod_{n=1}^{N}(1+r_n) < \frac{1}{\alpha}$$

is at least $1-\alpha$. It remains to take the log of both sides and apply the formulas (following from Taylor's expansion)

$$\ln(1+r_n) \geq r_n - \frac{1}{2}r_n^2, \quad r_n \geq 0,$$

$$\ln(1+r_n) \geq r_n - \frac{1}{2}r_n^2 + \frac{1}{3}\left(\frac{r_n}{1+r_n}\right)^3, \quad r_n < 0.$$ ∎

In typical applications (where $|r_n| \ll 1$) the quantity in the square brackets in (15.22) is small. The significance level $\alpha$ (typically something like 1%) is a number such that events of upper probability $\alpha$ or less are considered unlikely. Assuming that $|r_n| \ll 1$ and that $(-\ln\alpha)/N$ is of even lower order of magnitude, Equation (15.22) says that

$$\mu \leq \frac{\sigma_0^2}{2} + \text{small amount}$$

(cf. (9.35) on p. 231). This is applicable to the case when there is just one security in Investor's portfolio. So we can say that the efficient-market hypothesis implies a high lower probability that the average of the returns from a particular security will be less than half their squared volatility.

This result is much more concrete than typical results from stochastic finance theory. It is about the mean and variance of the actual returns, not the mean and variance for some stochastic mechanism that is hypothesized to generate these returns.

In order to verify that Proposition 15.8 is useful in practice, consider the returns from Microsoft stock for the period from March 13, 1986 (the first day on which data on Yahoo is available) until September 21, 2000, with S&P 500 as the *numéraire*. Here $N = 3,671$, and the empirical mean and volatility of the daily returns, to three significant digits, are

$$\mu = 0.00141, \text{ or } 0.141\%; \qquad \text{and} \qquad \sigma_0 = 0.0254, \text{ or } 2.54\%.$$

Taking $\alpha = 1\%$, the inequality

$$\mu < \frac{\sigma_0^2}{2} - \frac{\ln\alpha}{N} + \left[\frac{1}{3N}\sum_{n:r_n<0}\left|\frac{r_n}{1+r_n}\right|^3\right] \tag{15.23}$$

within the lower probability assertion in (15.22) becomes, numerically,

$$0.141\% < 0.0322\% + 0.125\% + 0.000512\%$$

(we use $\%$ to mean multiplication by 0.01), or

$$0.141\% < 0.158\%. \tag{15.24}$$

Thus the bound that is asserted to hold with lower probability of at least 99% did hold for Microsoft in this period, even though the mean return $\mu$ (0.141%) substantially exceeded $\sigma^2/2$ (0.0322%). Of course, we can scarcely expect the bound (15.23) to be as tight as (15.24) in general, because $\mu$ will seldom be so large. The similar calculation for IBM, for the 9,748 days from the beginning of 1962 (the earliest data available on Yahoo) until 21 September 2000 yields $0.008 < 0.059$ (to accuracy 0.001), a bound that is much looser, although the two sides are still of the same order of magnitude.

The bound (15.23) can be violated of course. We see this when we put $\alpha$ equal to 10% for the Microsoft data. In this case, the term $-\ln\alpha/N$ in (15.23) becomes 0.063% instead of 0.125%, and (15.23) becomes $0.141\% < 0.095\%$. The violation is hardly surprising, because Microsoft's performance between 1986 and 2000 was in the top tenth of anyone's range of expectations.

It should not be supposed that $\mu$ can approach the limit $\sigma_0^2/2$ asymptotically as the number of periods $N$ increases. On the contrary, as $N$ increases the limit on the size of $\mu$ at any fixed significance level $\alpha$ tends to zero (assuming $r_n = O(1)$), and the increase in the limit associated with increasing $\sigma_0^2$ becomes more and more negligible. This is clear from the following proposition, which appears asymptotically much stronger than Proposition 15.8.

**Proposition 15.9** *Under conditions of Proposition 15.8, for any* $N \geq 3$ *and significance level* $\alpha > 0$,

$$\underline{\mathbb{P}}\left\{\mu < \frac{1}{\sqrt{N}}\left(\frac{\sigma_0^2}{2} + \ln\frac{1}{\alpha} + \frac{1}{\sqrt{N}}\right)\right\} \geq 1 - \alpha.$$

*Proof* From the proof of the previous proposition we know that the lower probability of the event

$$\prod_{n=1}^{N}(1 + \kappa r_n) < \frac{1}{\alpha}$$

is at least $1 - \alpha$, where $0 < \kappa < 1$. (See our analogous use of $\kappa$ in the proof of the validity of the law of the iterated logarithm in Chapter 5.) Taking logarithms of both sides and again applying Taylor's expansion, we obtain

$$\sum_{n=1}^{N} \kappa r_n - \frac{1}{2}\sum_{n=1}^{N} \kappa^2 r_n^2 + \frac{1}{3}\sum_{n:r_n<0}\left(\frac{\kappa r_n}{1 + \kappa r_n}\right)^3 < \ln\frac{1}{\alpha}.$$

Setting $\kappa := N^{-1/2}$, we further obtain

$$\mu\sqrt{N} - \frac{1}{2}\sigma_0^2 + N^{-3/2}\frac{1}{3}\sum_{n:r_n<0}\left(\frac{r_n}{1 + N^{-1/2}r_n}\right)^3 < \ln\frac{1}{\alpha}. \tag{15.25}$$

Because $r_n \geq -1$ and $N \geq 3$, we see that

$$\sum_{n:r_n<0}\left(\frac{r_n}{1 + N^{-1/2}r_n}\right)^3 \geq -\frac{N}{1 - N^{-1/2}} \geq -3N.$$

So we can transform (15.25) to

$$\mu\sqrt{N} - \frac{\sigma_0^2}{2} - N^{-1/2} < \ln\frac{1}{\alpha}.$$

∎

The asymptotic power of Proposition 15.9 is irrelevant, however, for values of $N$ that we are likely to encounter in applications, because its error term involving the significance level $\alpha$ (namely, $-\ln\alpha/\sqrt{N}$), which is much larger than the corresponding term in Proposition 15.8 ($-\ln\alpha/N$), is usually too large to impose any interesting limit on $\mu$. Indeed, in our Microsoft example, where $N = 3,671$ and $\alpha = 1\%$, we obtain $-\ln\alpha/\sqrt{N} = 7.60\%$, hardly interesting as part of an upper bound on average daily returns.

Because of our insistence that Skeptic cannot risk bankruptcy, he cannot take short positions, and this means that we can prove Propositions 15.8 and 15.9 only under the assumption of no short selling. It also means that we cannot prove sharpness results for the law of the iterated logarithm and the central limit theorem. One way to remove this restriction on Skeptic is to assume that securities' returns do not exceed some specified positive constant $C$ [333]. Another, more flexible, possibility is to allow Skeptic to buy a new kind of tickets which will allow him to hedge against large returns for securities he is short in; these tickets might be derivatives, such as out-of-the-money call options, which would insure against abnormal behavior by the market, or merely theoretical tickets expressing our beliefs about the market.

## 15.5 OTHER FORMS OF THE EFFICIENT-MARKET HYPOTHESIS

In this concluding section, we briefly discuss two more possibilities for choosing the *numéraire* for our efficient-market hypothesis: (1) a given market index might be used as the *numéraire* for a larger market game in which Skeptic is allowed to trade in options as well as in the securities that make up the index, and (2) a monetary unit such as the dollar might be used when evaluating a complex range of market opportunities.

### When Skeptic Can Trade in Options

If the securities that are traded in a market include options, then we cannot take the total value of all outstanding units of all traded securities in the market as a *numéraire*, because a market order for an option may be filled by someone writing a new option, and hence the number of outstanding options contracts (this is called the *open interest* in the option and is included in daily market reports) can change from minute to minute. But we can consider the efficient-market hypothesis relative to some other *numéraire*.

We have already encountered one very important example: our game-theoretic Black-Scholes protocol, where the *numéraire* is a risk-free bond. Skeptic is absent from this protocol; the only players are Market and Investor, who is allowed to trade in the bond, in a stock $\mathcal{S}$, and in a derivative depending on $\mathcal{S}$ (the variance derivative $\mathcal{D}$ or the strictly convex option $\mathcal{R}$). But we can formulate an efficient-market hypothesis for Investor: we can suppose that he cannot get infinitely rich without risking bankruptcy. As we showed in §11.4, this game-theoretic efficient-market hypothesis implies that Market will produce a price path for $\mathcal{S}$ with variation exponent equal to 2, the typical value for a diffusion process. This example is important because it demonstrates dramatically how specific stochastic properties can emerge from a market game, without any assumption or conclusion that market prices are fully stochastic.

Another interesting implication of the efficient-market hypothesis in the Black-Scholes protocol with the risk-free bond as *numéraire* is discussed by Bodie (1995), who shows that as the time horizon grows it becomes more and more expensive to ensure, using the Black-Scholes formula, against a stock $\mathcal{S}$ earning less than the risk-free rate of interest. In our purely game-theoretic Black-Scholes protocol, with our usual simplifying assumption that the interest rate is zero (and the bond price is constant), we may write

$$\mathbb{P}\{-D_T/2 + \xi\sqrt{D_T} \geq 0\} = \mathbb{P}\{\xi \geq \sqrt{D_T}/2\}$$

for the price at time 0 of an option that pays 1 if the stock $\mathcal{S}$ does as well as the bond between 0 and $T$, where $D_T$ is the price at time 0 of our dividend-paying security $\mathcal{D}$ with maturity $T$ and $\xi$ is a standard Gaussian random variable. A simple arbitrage argument shows that $D_T$ increases with the time horizon $T$, and we can reasonably assume that $D_T \to \infty$ as $T \to \infty$. So the game-theoretic probability that $\mathcal{S}$ does as well as the bond tends to 0 as the time horizon $T$ increases. For $D_T = 100$, for example, the game-theoretic probability that $\mathcal{S}$ does as well as the bond is less than $10^{-6}$. So the efficient-market hypothesis with the bond as *numéraire* says that the $\mathcal{S}$ will not do as well as the bond. This does not contradict the fact that $\mathcal{S}$'s current price is the current game-theoretic price for what it will be worth at the time $T$ for which $D_T = 100$. Rather, it says that if we believe the efficient-market hypothesis for this market with the bond as *numéraire*, then buying the stock to hold for the long-run is like buying a lottery ticket: we are paying for a very large payoff that will almost certainly not happen.

Another interesting example is provided by the existence of options on market indexes $I$ (such as the European option on FTSE 100 discussed in Chapter 10). Consider the efficient-market hypothesis that takes $I$ as the *numéraire* but allows Skeptic to trade in the options on $I$ as well as in the stocks in $I$ (and hence, effectively, in $I$ itself). This hypothesis may allow us to draw conclusions about the rate of growth of $I$. For simplicity, suppose that a European call option on $I$, with maturity time $T$ and strike price $c$, is traded. If the option is far out of the money (i.e., the strike price $c$ is much greater than the current value $I(0)$ of the index), its price $p$ will be very low: $p \ll I(0)$. Suppose we subscribe to the efficient-market hypothesis to the extent that we do not consider it possible to beat the market by a factor of $1/\delta$ ($\delta$ might be 1%). Assume that the option is so far out of the money that $p < \delta I(0)$. Then holding one option to maturity will multiply Skeptic's capital by $(I(T) - c)^+/p$, where $I(T)$ is the value of the index at time $T$. Because the *numéraire* increases $I(T)/I(0)$-fold, the efficient-market hypothesis implies

$$\frac{(I(T) - c)/p}{I(T)/I(0)} \leq \frac{1}{\delta},$$

which is equivalent, under the assumption $p < \delta I(0)$, to

$$I(T) \leq \frac{c}{1 - \frac{p}{\delta I(0)}}.$$

This bound may be too loose to be useful, but it demonstrates the possibility of nontrivial implications for the performance of the market as a whole.

### Value at Risk

It is currently popular in financial practice to summarize the uncertainties concerning the investment portfolio of an individual or institution in terms of the portfolio's *value at risk* (VaR) at a certain significance level $\alpha$. This is defined in terms of a stipulated stochastic model $\mathbb{P}$ for the behavior of the prices of the securities in the portfolio. Writing $S(0)$ and $S(T)$ for the initial and final values of the portfolio, and assuming that its holder does not intend to make changes during the period $[0, T]$,

$$\mathrm{VaR} = \inf\left\{V \mid \mathbb{P}\{S(T) \geq S(0) - V\} \geq 1 - \alpha\right\}.$$

The significance level $\alpha$ is a small number, such as 5% or 1%, measuring the largest probability of disaster that we are willing to tolerate. The value at risk is the smallest bound on the portfolio's loss that we can endorse with confidence $1 - \alpha$.

Suppose that we adopt the efficient-market hypothesis with the dollar as *numéraire* for a market that includes all the investments of a particular individual or firm. This gives us upper and lower probabilities for all events concerning these investments, and hence it allows us to define the value at risk in an analogous way, using game-theoretic probability instead of stochastic probability:

$$\inf\left\{V \mid \underline{\mathbb{P}}\{S(T) \geq S(0) - V\} \geq 1 - \alpha\right\},$$

where the lower probability is evaluated in the situation in which $S(0)$ and other relevant financial data for time 0 (such as the prices $P(0)$ and $D(0)$ below) have just become known.

Suppose there is only one security traded in the market for which we have adopted the efficient-market hypothesis, and assume for simplicity that the portfolio consists of one share of this security. The VaR of this portfolio equals its current price; everything can be lost. There exist lower probabilities that are nontrivial (different from 0), such as

$$\underline{\mathbb{P}}\{S(T) \leq cS(0)\} = 1 - \frac{1}{c}$$

for $c > 1$, but all events of the form $\{S(T) \geq a\}$, where $a > 0$ may depend on the information available at time 0, have lower probability zero.

Now add another security to our market, a European put option with maturity $T$ (or an American put option with maturity at least $T$) and strike price $c$; suppose its initial price is $P(0)$. In this case the value at risk is at most

$$S(0) - c + \frac{P(0)}{\alpha} \tag{15.26}$$

(in some extreme cases this value can be negative). This is because the event that $S(0) - S(T)$ exceeds (15.26) can be written as

$$c - S(T) > \frac{P(0)}{\alpha},$$

and the upper probability of the last event is at most $\alpha$, as evidenced by the strategy of buying the put and holding it to maturity.

An extreme case is where instead of the put the dividend-paying derivative $\mathcal{D}$ introduced in Chapter 11 is added to the market. It easily follows from Theorem 11.2 (p. 280) that VaR in this case is

$$S(0)\left(1-\exp\left(-\frac{D(0)}{2}-k_\alpha^*\sqrt{D(0)}\right)\right),$$

where $k_\alpha^*$ is the upper $\alpha$-quantile of the standard Gaussian distribution. This expression is obtained by solving the equation

$$\int_{z:S(0)e^z<S(0)-V} \mathcal{N}_{-D(0)/2,\,D(0)}(dz)=\alpha \tag{15.27}$$

in $V$. The left-hand side of (15.27) is a special case of (11.13) for $U$ the indicator function of the set $(-\infty, S(0)-V)$. According to Theorem 11.2, if (15.27) holds and the final value $S(T)$ drops below $S(0)-V$, some trading strategy will have made 1 out of $\alpha$ without any risk of debt. (Formally, $U$ is required to be a Lipschitzian function, but it is clear that the indicator function can be approximated by a Lipschitzian function to any accuracy.) In this extreme case, our game-theoretic VaR is simply the usual stochastic VaR using the risk-neutral probability as the true probability.

It is not feasible or reasonable to calculate value at risk from options markets at the present time. The range of maturities for marketed derivatives is too limited, and the efficient-market hypothesis for markets that include derivatives has not been sufficiently tested. But if some of this book's ideas for improving option markets are implemented, game-theoretic value at risk may become feasible in the future.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# *References*

The following list includes references cited in the text by number and also those cited by author. Whenever possible, we provide first names (or middle names in some cases) rather than merely initials. If the first name appeared in the publication cited, we reproduce it as it appeared, even if this results in more than one form of the same person's name. (William Feller was Willy Feller in 1935.) In most cases where only initials were given in the publication but a first or middle name is known to us, we replace the initial with the name without comment. But if the name might be spelled or transliterated in more than one way, we put our addition in parentheses: A(leksandr) Khinchine.


1. Peter A. Abken and Saikat Nandi. Options and volatility. *Economic Review*, 81:21–35, December 1996. Federal Reserve Bank of Atlanta. A survey of competing proposals, deterministic and stochastic, for improving the performance of the Black-Scholes model by modeling changes in volatility.

2. Carl Ackermann, Richard McEnally, and David Ravenscraft. The performance of hedge funds: Risk, return, and incentives. *Journal of Finance*, 54:833–874, 1999.

3. William J. Adams. *The Life and Times of the Central Limit Theorem.* Kaedmon, New York, 1974. A pleasantly written history of the central limit theorem from De Moivre to Lyapunov.

4. Yutaka Aida and Yoshiro Sigeta. *NHK Special: Monetary Revolution 2—The Standard-Bearers of Financial Engineering*. NHK Publishing Company, Tokyo, 1999. NHK is an acronym for Nippon Housou Kyoukai, or the Japan Broadcasting Corporation. The book is based on a television program.

5. Oskar Anderson. *Probleme der statistischen Methodenlehre in den Sozialwissenschaften*. Physica, Würzburg, 1954.

6. Robert M. Anderson. A non-standard representation for Brownian motion and Itô integration. *Israel Journal of Mathematics*, 25:15–46, 1976.

7. John Arbuthnot. *Of the Laws of Chance.* B. Motte and C. Bathurst, London, fourth edition, 1738. A loose translation of Christiaan Huygens's *De ratiociniis in ludo aleae.* The first edition appeared in 1692.

8. Antoine Arnauld and Pierre Nicole. *La logique, ou l'art de penser.* Vrin, Paris, 1993. This critical edition, by Pierre Clair and François Girbal, was first issued in 1981. It is based on the 1683 edition of *l'Art de penser*, often called the *Port Royal Logic.* The first edition appeared in 1662, the year of Pascal's death.

9. Louis Bachelier. Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure, 3e série*, 17:21–86, 1900. This was Bachelier's doctoral dissertation. It was reprinted in facsimile in 1995 by Éditions Jacques Gabay, Paris. An English translation, by A. James Boness, appears on pp. 17–78 of [56].

10. Brad M. Barber and Terrance Odean. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, 55:773–806, 2000.

11. Edwin F. Beckenbach and Richard Bellman. *Inequalities.* Springer, Berlin, 1965.

12. Enrico G. Beltrametti and Gianni Cassinelli. *The Logic of Quantum Mechanics.* Addison-Wesley, Reading, MA, 1981.

13. Jan Beran. *Statistics for Long-Memory Processes.* Chapman and Hall, New York, 1994.

14. José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory.* Wiley, Chichester, England, 1994.

15. Jacob Bernoulli. *Ars Conjectandi.* Thurnisius, Basel, 1713. This pathbreaking work appeared eight years after Bernoulli's death. Bound together with it were two other works by Bernoulli, a treatise in Latin on infinite series, and a study in French of odds in court tennis (*Lettre à un Amy, sur les Parties du Jeu de Paume*). A facsimile reprinting of the original Latin text of *Ars Conjectandi* is marketed by Éditions Jacques Gabay. A German translation (*Wahrscheinlichkeitsrechnung von Jakob Bernoulli. Anmerkungen von R. Haussner.* Ostwald's Klassiker, Nr. 107–108, Engelmann, Leipzig) appeared in 1899. No English edition is available.

16. Peter L. Bernstein. *Capital Ideas: The Improbable Origins of Modern Wall Street.* Free Press, New York, 1992. A lively account of the development and intertwining of financial practice and theory, beginning with the stories of Louis Bachelier in Paris and Charles Dow and Eddie Jones on Wall Street a century ago and concluding with the adventures of the Wall Street professors of the 1970s and 1980s. Bernstein, a financial analyst and founding editor of the *Journal of Portfolio Management* is personally acquainted with many of the individuals who play starring roles in his story.

17. Serge Bernstein. Sur le théorème limite du calcul des probabilités. *Mathematische Annalen*, 85:237–241, 1922.

18. Serge Bernstein. Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, 97:1–59, 1927.

19. Jean Bertoin. Sur une intégrale pour les processus à $\alpha$-variation bornée. *Annals of Probability*, 17:1521–1535, 1989.

20. Jean Bertoin. *Lévy Processes.* Cambridge University Press, Cambridge, 1996.

21. Joseph Bertrand. *Calcul des Probabilités*. Gauthier-Villars, Paris, 1889. Some copies of the first edition are dated 1888. A second edition appeared in 1907 and was republished by Chelsea, New York, in 1972.

22. Avi Bick and Walter Willinger. Dynamic spanning without probabilities. *Stochastic Processes and their Applications*, 50:349–374, 1994.

23. Patrick Billingsley. *Ergodic Theory and Information*. Wiley, New York, 1965.

24. Patrick Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

25. Patrick Billingsley. *Probability and Measure*. Wiley, New York, third edition, 1995. Previous editions appeared in 1979 and 1986.

26. Nicholas H. Bingham and Rüdiger Kiesel. *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*. Springer-Verlag, London, 1998. A textbook on stochastic processes and option pricing. The authors present both topics first for discrete and then for continuous time.

27. Fischer Black. How to use the holes in Black-Scholes. *Journal of Applied Corporate Finance*, 1(4):67–73, Winter 1989. Reprinted in [174].

28. Fischer Black. How we came up with the option formula. *Journal of Portfolio Management*, 15:4–8, Winter 1989.

29. Fischer Black and Myron Scholes. The valuation of option contracts and a test of market efficiency. *Journal of Finance*, 27:399–417, 1972.

30. Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973. This is the article in which the Black-Scholes equation and the Black-Scholes formula were first derived. It is concerned as much with the pricing of warrants and other corporate liabilities as with options and derivatives.

31. Zvi Bodie. On the risk of stocks in the long run. *Financial Analysts Journal*, pp. 18–22, May–June 1995.

32. Arno Bohm. *Quantum Mechanics: Foundations and Applications*. Springer, New York, second edition, 1986.

33. Émile Borel. *Leçons sur la théorie des fonctions*. Gauthier-Villars, Paris, 1898.

34. Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–270, 1909. Reprinted in [37], Volume 2, pp. 1055–1079.

35. Émile Borel. *Applications aux jeux de hasard*. Gauthier-Villars, Paris, 1938. The title page lists Borel as the author but indicates that the entire book was actually written by Jean Ville. Ville is listed as the author, however, of an important note at the end of the book, [305].

36. Émile Borel. *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, Paris, 1939.

37. Émile Borel. *Œuvres de Émile Borel*. Éditions du Centre National de la Recherche Scientifique, Paris, 1972. Four volumes.

38. Nicolas Bourbaki. *Eléments de mathématique, Livre IV, Fonctions d'une variable réelle (théorie élémentaire)*. Hermann, Paris, second edition, 1958.

39. Phelim P. Boyle and David Emanuel. Discretely adjusted option hedges. *Journal of Financial Economics*, 8:259–282, 1980.

40. Leo Breiman. *Probability*. Addison-Wesley, Reading, MA, 1968. Second edition: SIAM, Philadelphia, 1992.

41. Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

42. Mark Britten-Jones and Anthony Neuberger. Arbitrage pricing with incomplete markets. *Applied Mathematical Finance*, 3:347–363, 1996.

43. W. Browne. *Christiani Hugenii Libellus de Ratiociniis in Ludo Aleae. Or, the Value of all Chances in Games of Fortune; Cards, Dice, Wagers, Lotteries, etc. Mathematically Demonstrated.* Woodward, London, 1714. A translation of Christiaan Huygens's *De ratiociniis in ludo aleae*.

44. Bernard Bru. Postface to the 1986 edition of Laplace's "Essai philosophique sur les probabilités". In *[190]*, pp. 245–303. Christian Bourgois, Paris, 1986.

45. Stephen G. Brush. A history of random sequences. I: Brownian movement from Brown to Perrin. *Archive for History of Exact Sciences*, 5:1–36, 1968.

46. Laurent Calvet and Adlai Fisher. A multifractal model of asset returns, November 1999. Available from `http://www.stern.nyu.edu/~afisher`, this paper simplifies and elaborates results on the multifractal model of asset prices. Empirical calculations of the partition function are extended to equities, which were not covered by [124].

47. Laurent Calvet, Adlai Fisher, and Benoit B. Mandelbrot. Large deviations and the distribution of price changes. Cowles Foundation Discussion Paper No. 1166, September 1997. Yale University. Paper available from the SSRN database `http://www.ssrn.com`.

48. John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 1997.

49. John Rozier Cannon. *The One-Dimensional Heat Equation*. Addison-Wesley, Reading, MA, 1984.

50. Mark M. Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52:57–82, 1997.

51. Yuan S. Chow and Henry Teicher. *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York, third edition, 1997. One of many excellent textbooks on measure-theoretic probability. Its first edition appeared in 1978.

52. Neil Chriss and William Morokoff. Volatility and variance swaps. *Risk*, 12, October 1999.

53. Alonzo Church. On the concept of a random sequence. *Bulletin of the American Mathematical Society*, 46:130–135, 1940.

54. Jean Cocteau. *Œuvres complètes de Jean Cocteau*. Marguerat, Lausanne, Switzerland, 1946–1951. The lovely poem entitled "Martingale", originally published in Cocteau's *Opéra* (1925–1927), is on p. 136 of Volume IV.

55. Corneliu Constantinescu and Aurel Cornea. *Potential Theory of Harmonic Spaces*. Springer, Berlin, 1972.

56. Paul H. Cootner, editor. *The Random Character of Stock Market Prices*. MIT Press, Cambridge, MA, 1964.

57. Arthur H. Copeland Sr. Admissible numbers in the theory of probability. *American Journal of Mathematics*, 50:535–552, 1928.

58. Antoine Augustin Cournot. *Exposition de la théorie des chances et des probabilités*. Hachette, Paris, 1843. Reprinted in 1984 as Volume I (B. Bru, editor) of [59].

59. Antoine Augustin Cournot. *Œuvres complètes*. Vrin, Paris, 1973–1984. 10 volumes.

60. Jean-Michel Courtault et al. Louis Bachelier: On the centenary of Théorie de la Speculation. *Mathematical Finance*, 10:339–353, 2000.

61. Thomas M. Cover and Erik Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42:348–363, 1996.

62. David R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society. Series B*, 34:187–220, 1972.

63. David R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

64. David R. Cox and David V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1974.

65. David R. Cox and David Oakes. *Analysis of Survival Data*. Chapman and Hall, London, 1984.

66. John C. Cox and Stephen A. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, 1976.

67. John C. Cox, Stephen A. Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7:229–263, 1979.

68. John Crank. *The Mathematics of Diffusion*. Oxford University Press, Oxford, second edition, 1975.

69. Pierre Crépel. Private communication, November 2000. Professor Crépel has shared the following documents with us: (1) a letter from Crépel to Jean Ville dated 22 August 1984, (2) Crépel's notes from his interview of Ville on 27 August 1984, in Ville's home at Langon, France, (3) a letter from Crépel to Ville, dated 21 January 1985, (4) a letter from Ville to Crépel, dated February 2, 1985, (5) two documents from the Fréchet Archives of the probability department at the University of Paris VI: a curriculum vitae for Ville from around 1956 and a 15-page summary by Ville of his scientific work, dated May 1955, and (6) Crépel's 65-page paper on the history of martingales, *Quelques matériaux pour l'histoire de la théorie des martingales (1920–1940)*, which he presented on 5 November 1984 to the *Seminaires de Probabilités* of the University of Rennes.

70. J. J. Cross. Potential theory. In Grattan-Guinness [137], pp. 475–488 (Volume 1).

71. Nigel J. Cutland. Infinitesimals in action. *Journal of the London Mathematical Society*, 35:202–216, 1987.

72. Nigel J. Cutland, P. Ekkehard Kopp, and Walter Willinger. A nonstandard approach to option pricing. *Mathematical Finance*, 1:1–38, 1991.

73. Nigel J. Cutland, P. Ekkehard Kopp, and Walter Willinger. From discrete to continuous financial models: New convergence results for option pricing. *Mathematical Finance*, 3:101–123, 1993.

74. Nigel J. Cutland, P. Ekkehard Kopp, and Walter Willinger. Stock price returns and the Joseph effect: A fractional version of the Black-Scholes model. In E. Bolthausen, M. Dozzi, and F. Russo, editors, *Seminar on Stochastic Analysis, Random Fields and Applications*, pp. 327–351. Birkhäuser, Basel, 1995.

75. Robert C. Dalang, Andrew Morton, and Walter Willinger. Equivalent martingale measures and no-arbitrage in stochastic securities market model. *Stochastics and Stochastics Reports*, 29:185–201, 1990.

76. G. D'Albigny. *Les martingales modernes. Nouveau recueil du joueur de Roulette et de Trente-et-Quarante, contenant les meilleurs systèmes.* Librairie des Mathurins, Paris, 1902.

77. K. E. Dambis. On the decomposition of continuous submartingales. *Theory of Probability and Its Applications*, 10:401–410, 1965.

78. Joseph Warren Dauben. *Abraham Robinson: The Creation of Nonstandard Analysis. A Personal and Mathematical Odyssey.* Princeton University Press, Princeton, NJ, 1995.

79. Florence N. David. *Games, Gods, and Gambling.* Griffin, London, 1962.

80. Martin Davis. *Applied Nonstandard Analysis.* Wiley, New York, 1977.

81. Morton Davis. Infinite games with perfect information. In M. Dresher, L. S. Shapley, and A. W. Tucker, editors, *Advances in Game Theory*, pp. 85–101. Princeton University Press, Princeton, NJ, 1964.

82. A. Philip Dawid. Statistical theory: The prequential approach (with discussion). *Journal of the Royal Statistical Society. Series A*, 147:278–292, 1984.

83. A. Philip Dawid. Calibration-based empirical probability (with discussion). *Annals of Statistics*, 13:1251–1285, 1985.

84. A. Philip Dawid. Probability forecasting. In S. Kotz, N. L. Johnson, and C. B. Read, editors, *Encyclopedia of Statistical Sciences*, volume 7, pp. 210–218. Wiley, New York, 1986.

85. A. Philip Dawid. Prequential data analysis. In Malay Ghosh and Pramod K. Pathak, editors, *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, IMS Lecture Notes–Monographs Series 17, pp. 113–126. Institute of Mathematical Statistics, Hayward, California, 1992.

86. A. Philip Dawid and Vladimir G. Vovk. Prequential probability: Principles and properties. *Bernoulli*, 5:125–162, 1999. This paper is the first to state formally a game-theoretic law of the iterated logarithm.

87. Bruno de Finetti. Integrazione delle funzioni a incremento aleatorio. *Atti della Reale Accademia Nazionale dei Lincei. Classe di scienze fisiche, matematiche, e naturali. Rendiconti Serie VI*, 10:548–553, 1929.

88. Bruno de Finetti. Sulla possibilità di valori eccezionali per una legge di incrementi aleatori. *Atti della Reale Accademia Nazionale dei Lincei. Classe di scienze fisiche, matematiche, e naturali. Rendiconti Serie VI*, 10:325–329, 1929.

89. Bruno de Finetti. Sulle funzione a incremento aleatorio. *Atti della Reale Accademia Nazionale dei Lincei. Classe di scienze fisiche, matematiche, e naturali. Rendiconti Serie VI*, 10:163–168, 1929.

90. Bruno de Finetti. Le funzioni caratteristiche di legge istantanea. *Atti della Reale Accademia Nazionale dei Lincei. Classe di scienze fisiche, matematiche, e naturali. Rendiconti Serie VI*, 12:278–282, 1930.

91. Bruno de Finetti. *Compte rendu critique du colloque de Genève sur la théorie des probabilités.* Number 766 in *Actualités Scientifiques et Industrielles.* Hermann, Paris, 1939. This is the eighth fascicle of [338].

92. Bruno de Finetti. Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Triestini dell'Università di Trieste*, 19:29–81, 1949. An English translation by Giandomenico Majone appears as "On the axiomatization of probability theory", Chapter 5 of [94].

93. Bruno de Finetti. *Teoria Delle Probabilità*. Einaudi, Turin, 1970. An English translation, by Antonio Machi and Adrian Smith, was published as *Theory of Probability* by Wiley (London, England) in two volumes in 1974 and 1975.

94. Bruno de Finetti. *Probability, Induction and Statistics: The Art of Guessing*. Wiley, London, 1972. Ten articles in English, all previously published in English, French, or Italian in the period from 1949 to 1967.

95. Joseph L. Doob. Note on probability. *Annals of Mathematics*, 37:363–367, 1936. This article demonstrates, within the measure-theoretic framework, that the probabilities for the outcomes of an infinite sequence of independent, identically distributed random variables are not changed by a gambling system—in this case, a system for selecting a subsequence.

96. Joseph L. Doob. Review of *Étude critique de la notion de collectif*. *Bulletin of the American Mathematical Society*, 45:824, 1939.

97. Joseph L. Doob. Regularity properties of certain families of chance variables. *Transactions of the American Mathematical Society*, 47:455–486, 1940.

98. Joseph L. Doob. Probability as measure. *Annals of Mathematical Statistics*, 12:206–214, 1941. This article originated as a paper for a meeting of the Institute of Mathematical Statistics in Hanover, NH, in September 1940. It was published together with an article by von Mises [313] and comments by Doob and von Mises on each other's articles [315].

99. Joseph L. Doob. Application of the theory of martingales. In *Le Calcul des Probabilités et ses Applications*, Colloques Internationaux, pp. 23–27. Centre National de la Recherche Scientifique, Paris, 1949.

100. Joseph L. Doob. Continuous parameter martingales. In Jerzy Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 269–277. University of California Press, Berkeley, 1951.

101. Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953. Classical monograph that popularized martingales.

102. Joseph L. Doob. Wiener's work in probability theory. *Bulletin of the American Mathematical Society*, 72(1, Part II):69–72, 1966. This is a special obituary issue for Wiener, and its pages are numbered separately from the rest of Volume 72.

103. Joseph L. Doob. *Classical Potential Theory and Its Probabilistic Counterpart*. Springer, New York, 1984.

104. Joseph L. Doob. The development of rigor in mathematical probability, 1900–1950. In Pier [245], pp. 157–170.

105. Lester E. Dubins and Leonard J. Savage. *How to Gamble if You Must: Inequalities for Stochastic Processes*. McGraw-Hill, New York, 1965. A second edition, with the title and subtitle interchanged, was published in 1976 by Dover, New York.

106. Lester E. Dubins and Gideon Schwarz. On continuous martingales. *Proceedings of the National Academy of Sciences of the USA*, 53:913–916, 1965.

107. Darrell Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, NJ, second edition, 1996. In addition to a relatively rigorous treatment of continuous-time arbitrage pricing, this book includes discussions of equilibrium pricing and optimal portfolio selection. The first edition appeared in 1992.

108. A. W. F. Edwards. *Pascal's Arithmetical Triangle*. Charles Griffin & Company Limited and Oxford University Press, London and New York, 1987.

109. Albert Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17:549–560, 1905.

110. Albert Einstein. *Investigations on the Theory of the Brownian Movement*. Dover, New York, 1956. First published by Dutton, New York, in 1926, this is a translation, by A. D. Cowper, of *Untersuchungen über die Theorie der Brownschen Bewegung*. It includes [109] together with Einstein's other papers on Brownian motion.

111. P. Eklof. Ultraproducts for algebraists. In J. Barwise, editor, *Handbook of Mathematical Logic*, pp. 105–137. North Holland, Amsterdam, 1977.

112. Robert J. Elliott. *Stochastic Analysis and Applications*. Springer, New York, 1982.

113. Robert J. Elliott and P. Ekkehard Kopp. *Mathematics of Financial Markets*. Springer, New York, 1999.

114. Edwin J. Elton and Martin J. Gruber, editors. *Security Evaluation and Portfolio Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1972.

115. Georg Faber. über stetige Funktionen. *Mathematische Annalen*, 69:372–443, 1910.

116. Eugene F. Fama. The behavior of stock-market prices. *Journal of Business*, 38:34–105, 1965.

117. Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.

118. Eugene F. Fama. Efficient capital markets, II. *Journal of Finance*, 46:1575–1617, 1991.

119. Eugene F. Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49:283–306, 1998.

120. William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, New York, third edition, 1968. Previous editions of this volume appeared in 1950 and 1957.

121. William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, second edition, 1971. The first edition of this volume appeared in 1966.

122. Willy Feller. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 40:521–559, 1935.

123. Pierre de Fermat. *Œuvres*. Gauthier-Villars, Paris, 1891–1912. Edited by Paul Tannery and Charles Henry in four volumes, together with a supplement that appeared in 1922. Fermat's correspondence, including the surviving letters to Pascal on the problem of points and the problem of the gambler's ruin, are in the second volume, which appeared in 1894.

124. Adlai Fisher, Laurent Calvet, and Benoit B. Mandelbrot. Multifractality of Deutschmark / US Dollar exchange rates. Cowles Foundation Discussion Paper No. 1166, September

1997. Yale University. Available from the SSRN database `http://www.ssrn.com`. Among other things, the paper studies the variation spectrum for the DM/USD market.

125. Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925. The thirteenth edition appeared in 1958.

126. Hans Föllmer. Calcul d'Itô sans probabilités. In J. Azéma and M. Yor, editors, *Séminaire de Probabilités XV*, volume 850 of *Lecture Notes in Mathematics*, pp. 143–150. Springer, Berlin, 1981.

127. Hans Föllmer and Uwe Küchler. Richard von Mises. In Heinrich G. W. Begehr, Helmut Koch, Jürg Kramer, Norbert Schappacher, and Ernst-Jochen Theile, editors, *Mathematics in Berlin*, pp. 111–116. Birkhäuser Verlag, Berlin, 1998.

128. Jack Clark Francis, William W. Toy, and J. Gregg Whittaker, editors. *The Handbook of Equity Derivatives*. Wiley, New York, revised edition, 2000.

129. Maurice Fréchet. Sur l'intégrale d'une fonctionelle étendue à un ensemble abstrait. *Bulletin de la Société mathématique de France*, 43:248–265, 1915.

130. Maurice Fréchet. Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. In Rolin Wavre, editor, *Les fondements du calcul des probabilités*, number 735 in *Actualités Scientifiques et Industrielles*, pp. 22–55. Hermann, Paris, 1938. Fréchet reviews von Mises's concept of a collective and the clarifications due to Reichenbach, Popper, Copeland, Wald, and Ville. He concludes that the theory of the collective is not successful as a foundation for probability and argues that Kolmogorov's axioms provide a sufficient formal foundation. *Les fondements du calcul des probabilités* is the second fascicle of [338].

131. Avner Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, NJ, 1964.

132. Bert Fristedt and S. James Taylor. Strong variation for the sample functions of a stable process. *Duke Mathematical Journal*, 40:259–278, 1973.

133. David Gale and F. M. Stewart. Infinite games with perfect information. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, II*, pp. 245–266. Princeton University Press, Princeton, NJ, 1953.

134. Boris V. Gnedenko. *Theory of Probability*. Gordon and Breach, St. Leonards, Australia, 1997. This is an English translation, with revisions by the author, of the sixth Russian edition of Курс теории вероятностей, published in 1988 by Nauka, Moscow. The translator is Igor A. Ushakov. The first Russian edition of the book appeared in 1950.

135. Boris V. Gnedenko and Andrei N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley, Cambridge, MA, 1954. The original Russian edition: Предельные распределения для сумм независимых случайных величин, 1949. Translated and annotated by K. L. Chung. With an appendix by Joseph L. Doob.

136. Robert Goldblatt. *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis*. Springer, New York, 1998. An excellent exposition based on ultrafilters. Contains a chapter on the Loeb measure.

137. Ivor Grattan-Guinness, editor. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*. Routledge, London, 1994. two volumes.

138. Ivor Grattan-Guinness. Heat diffusion. In *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences* [137], pp. 1165–1170 (Volume 2).

139. Jeremy J. Gray. *The Hilbert Challenge*. Oxford University Press, Oxford, 2000.

140. Sanford Grossman and Joseph Stiglitz. On the impossibility of informationally efficient markets. *American Economic Review*, 70:393–408, 1980.

141. Ian Hacking. *The Emergence of Probability*. Cambridge University Press, New York, 1975.

142. Ian Hacking. *The Taming of Chance*. Cambridge University Press, New York, 1990.

143. Anders Hald. *A History of Probability and Statistics and their Applications before 1750*. Wiley, New York, 1990.

144. Peter Hall and Christopher C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980.

145. Godfrey H. Hardy and John E. Littlewood. Some problems of Diophantine approximation. *Acta Mathematica*, 37:155–239, 1914.

146. Godfrey H. Hardy, John E. Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, Cambridge, second edition, 1967. The first edition appeared in 1934.

147. J. Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11:215–260, 1981. This article is credited with the first rigorous and general proof that the absence of arbitrage and the ability to hedge depend only on the existence and uniqueness of an equivalent martingale measure.

148. Philip Hartman and Avi Wintner. On the law of the iterated logarithm. *American Journal of Mathematics*, 63:169–176, 1941.

149. Felix Hausdorff. Beiträge zur Wahrscheinlichkeitsrechnung. *Sitzungsberichte der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 53:152–178, 1901.

150. Felix Hausdorff. *Grundzüge der Mengenlehre*. Verlag Von Veit & Comp., Leipzig, 1914.

151. Thomas Hawkins. *Lebesgue's Theory of Integration: Its Origins and Development*. Chelsea, New York, second edition, 1974. The first edition was published in 1970 by the University of Wisconsin Press, Madison.

152. David Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 8:437–479, 1902. Hilbert's famous address to the International Congress of Mathematicians in Paris in 1900, in which he listed twenty-three open problems central to mathematics. Translated from the German by Mary W. Newson.

153. Roland F. Hoskins. *Standard and Nonstandard Analysis: Fundamental Theory, Techniques, and Applications*. Ellis Horwood, New York, 1990. Chapter 4 gives the construction of $^*\mathbb{R}$ via ultrafilters. Later chapters contain applications in mathematical analysis including convergence, continuity, differentiation, and integration.

154. John C. Hull. *Options, Futures, and Other Derivatives*. Prentice-Hall, Upper Saddle River, NJ, fourth edition, 2000. A popular introduction to the study of derivatives at the MBA level. It discusses both the theory and practice of derivatives and includes numerical procedures for implementing the most common pricing algorithms.

155. Christiaan Huygens. *Œuvres complètes de Christiaan Huygens*. Nijhoff, La Haye, 1888–1950. There are twenty-two volumes. Volume XIV, which contains Huygens's work on probability, appeared in 1920.

156. Kiyosi Itô. Stochastic integral. *Proceedings of the Imperial Academy of Tokyo*, 20:519–524, 1944. Reprinted in [157], pp. 85–90.

157. Kiyosi Itô. *Selected Papers*. Springer, New York, 1987. Edited by Daniel W. Stroock and S. R. Sriniva Varadhan.

158. Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 1987.

159. Max Jammer. *The Philosophy of Quantum Mechanics: The Interpretations of Quantum Mechanics in Historical Perspective*. Wiley, New York, 1974.

160. Michael C. Jensen. Risk, the pricing of capital assets, and evaluation of investment portfolios. *Journal of Business*, 42:167–247, 1969.

161. Camille Jordan. Remarques sur les intégrales définies. *Journal de mathématiques pures et appliquées (4)*, 8:69–99, 1892.

162. E. Jouini and H. Kallal. Martingales and arbitrage in securities markets with transactions costs. *Journal of Economic Theory*, 66:178–197, 1995.

163. Joan Junkus. U. S. options exchanges. In Francis et al. [128], pp. 55–76.

164. Jean-Pierre Kahane. Des séries de Taylor au mouvement brownien, avec un aperçu sur le retour. In Pier [245], pp. 415–429.

165. Vladimir G. Kanovei. Аксиома выбора и аксиома детерминированности *(The axiom of choice and the axiom of determinacy)*. Nauka, Moscow, 1984.

166. Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, second edition, 1991.

167. Takayuki Kawada and Norio Kôno. On the variation of Gaussian processes. In G. Maruyama and Yurii V. Prokhorov, editors, *Proceedings of the Second Japan-USSR Symposium on Probability Theory*, volume 330 of *Lecture Notes in Mathematics*, pp. 176–192. Springer, Berlin, 1973.

168. Alexander S. Kechris. *Classical Descriptive Set Theory*. Springer, New York, 1995.

169. John L. Kelley, Jr. A new interpretation of information rate. *Bell System Technical Journal*, 35:917–926, 1956.

170. A(leksandr) Khinchine. Über einen Satz der Wahrscheinlichkeitsrechnung. *Fundamenta Mathematicae*, VI:9–20, 1924. The date "December 1922" is given at the end of the article.

171. A(leksandr) Khinchine. Zur Theorie der unbeschränkt teilbaren Vorteilungsgesetze. *Matematicheskii Sbornik*, 2:79–119, 1937.

172. A(leksandr) Khinchine and Paul Lévy. Sur les lois stables. *Comptes rendus des Séances de l'Académie des Sciences*, 202:374–376, 1936.

173. Frank Klingenhöfer and Martina Zähle. Ordinary differential equations with fractal noise. *Proceedings of the American Mathematical Society*, 127:1021–1028, 1999.

174. Robert W. Kolb, editor. *Practical Readings in Financial Derivatives*. Blackwell, Malden, MA, 1998.

175. Andrei N. Kolmogorov. Über das Gesetz der iterierten Logarithmus. *Mathematische Annalen*, 101:126–135, 1929. Appears in English in [181], pp. 32–42, as "On the law of the iterated logarithm".

176. Andrei N. Kolmogorov. Sur la loi forte des grands nombres. *Comptes rendus des Séances de l'Académie des Sciences*, 191:910–912, 1930. Appears in English in [181], pp. 60–61, as "On the strong law of large numbers".

177. Andrei N. Kolmogorov. Sulla forma generale di un processo stocastico omogeneo (Un probleme di Bruno de Finetti). *Atti della Reale Accademia Nazionale dei Lincei. Classe di scienze fisiche, matematiche, e naturali. Rendiconti. Serie VI*, 15:805–808 and 866–869, 1932. The second part of the article (pp. 866–869) is entitled "Ancora sulla forma generale di un processo omogeneo". The article appears in English in [181], pp. 121–127, as "On the general form of a homogeneous stochastic process".

178. Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956. A Russian translation, by G. M. Bavli, appeared under the title Основные понятия теории вероятностей (Nauka, Moscow) in 1936, with a second edition, slightly expanded by Kolmogorov with the assistance of Albert N. Shiryaev, in 1974.

179. Andrei N. Kolmogorov. On tables of random numbers. *Sankhya, The Indian Journal of Statistics. Series A*, 25:369–376, 1963. Republished in [182], pp. 176–183. The two English versions both contain potentially misleading errors, but in different places.

180. Andrei N. Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume I: Mathematics and Mechanics*. Kluwer, Dordrecht, 1991. Translated by V. M. Volosov from the Russian original, Избранные труды. Математика и механика, edited by Vladimir M. Tikhomirov and published in 1985 by Nauka, Moscow. The volume includes commentary on some of the articles. Some of the articles were originally published in languages other than Russian, but the translations of these articles in this volume appear to have been made from the Russian, without reference to the originals or other translations. Consequently, even the titles are sometimes different.

181. Andrei N. Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume II: Probability Theory and Mathematical Statistics*. Kluwer, Dordrecht, 1992. Translated by G. Lindquist from the Russian original, Теория вероятностей и математическая статистика, edited by Albert N. Shiryaev and published in 1986 by Nauka, Moscow. Our comments about [180] also apply to this volume.

182. Andrei N. Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume III: Information Theory and the Theory of Algorithms*. Kluwer, Dordrecht, 1993. Translated by A. B. Sossinsky from the Russian original, Теория информации и теория алгоритмов, edited by Albert N. Shiryaev and published in 1987 by Nauka, Moscow. Our comments about [180] also apply to this volume.

183. Andrei N. Kolmogorov and Vladimir A. Uspenskii. Algorithms and randomness. *Theory of Probability and Its Applications*, 32:389–412, 1987.

184. David M. Kreps. Multiperiod securities and the efficient allocation of risk: A comment on the Black-Scholes option pricing model. In John J. McCall, editor, *The Economics of Information and Uncertainty*, pp. 203–232. Chicago University Press, Chicago, 1982.

185. Sylvestre François Lacroix. *Traité élémentaire du calcul des probabilités*. Bachelier, Paris, 2nd edition, 1822.

186. Frank Lad. *Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction*. Wiley, New York, 1996.

187. Michiel van Lambalgen. *Random Sequences.* PhD thesis, University of Amsterdam, 1987.

188. John W. Lamperti. *Probability.* Wiley, New York, second edition, 1996. The first edition appeared in 1966.

189. Pierre Simon de Laplace. *Théorie analytique des probabilités.* Courcier, Paris, first edition, 1812. This monumental work had later editions in 1814 and 1820. The third edition was reprinted in Volume 7 of Laplace's *Œuvres complètes.*

190. Pierre Simon de Laplace. *Essai philosophique sur les probabilités.* Courcier, Paris, first edition, 1814. The fifth and definitive edition was published in 1825. A modern edition, edited by Bernard Bru, was published by Christian Bourgois, Paris, in 1986.

191. Pierre Simon de Laplace. *Philosophical Essay on Probabilities.* Springer, New York, 1994. English translation of [190]; translated by Andrew I. Dale.

192. Lucien Le Cam. The central limit theorem around 1935 (with comments by Hale F. Trotter, Joseph L. Doob, and David Pollard). *Statistical Science*, 1:78–96, 1986.

193. Henri Lebesgue. Sur une généralisation de l'intégrale définie. *Comptes rendus des Séances de l'Académie des Sciences*, 132:1025–1028, 1901.

194. Robert J. Leonard. Creating a context for game theory. In E. Roy Weintraub, editor, *Toward a History of Game Theory*, pp. 29–76. Duke University Press, Durham, NC, 1992.

195. Dominique Lepingle. La variation d'ordre $p$ des semi-martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36:295–316, 1976.

196. Paul Lévy. *Calcul de probabilités.* Gauthier-Villars, Paris, 1925. This pathbreaking book includes the first theory of stable distributions. The first few chapters provide an eclectic understanding of the subjective and objective sides of probability. An appendix discusses the laws of probability in abstract spaces.

197. Paul Lévy. Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Annali della R. Scuola Normale Superiore di Pisa. Series 2.*, 3:337–366, 1934. See also the addendum in 4:217–218, 1935. The article and addendum are reprinted in [203], Volume IV, pp. 9–40.

198. Paul Lévy. Propriétés asymptotiques des sommes de variables indépendantes ou enchainées. *Journal des mathématiques pures et appliquées. Series 9.*, 14(4):347–402, 1935. This article was omitted from [203], the six-volume collection of Lévy's works.

199. Paul Lévy. *Théorie de l'addition des variables aléatoires.* Gauthier-Villars, Paris, 1937. Second edition: 1954.

200. Paul Lévy. Le mouvement brownien plan. *American Journal of Mathematics*, 62:487–550, 1940. Reprinted in [203], Volume V, pp. 10–74.

201. Paul Lévy. Les fondements du calcul des probabilités. *Revue de métaphysique et de morale*, 59(2):164–179, 1954. Reprinted in [203], Volume VI, pp. 95–110.

202. Paul Lévy. *Quelques aspects de la pensée d'un mathématicien.* Blanchard, Paris, 1970.

203. Paul Lévy. *Œuvres de Paul Lévy.* Gauthier-Villars, Paris, 1973–1980. In six volumes. Edited by Daniel Dugué.

204. Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer, New York, second edition, 1997. The first edition appeared in 1993.

205. LIFFEdata historical CD-Rom: Equity products end of day data. LIFFE Data Services, Basingstoke, UK, March 2001. Website http://www.liffedata.com, e-mail sales@liffedata.com.

206. Robert S. Liptser and Albert N. Shiryaev. *Theory of Martingales.* Kluwer, Dordrecht, 1989. The Russian original, Теория мартингалов, was published by Nauka, Moscow, in 1986.

207. Andrew W. Lo, editor. *Market Efficiency: Stock Market Behaviour in Theory and Practice.* Elgar, Cheltenham, United Kingdom, 1997. Two volumes. A comprehensive and extremely expensive collection of articles on the efficient market hypothesis over three decades.

208. Andrew W. Lo and A. Craig MacKinlay, editors. *A Non-Random Walk Down Wall Street.* Princeton, NJ, Princeton University Press, 1999.

209. Peter A. Loeb. Conversion from nonstandard to standard measure spaces and applications in probability theory. *Transactions of the American Mathematical Society*, 211:113–122, 1975.

210. Jerzy Łoś. Quelques remarques, théoremes et problèmes sur les classes définissables d'algèbres. In *Mathematical Interpretation of Formal Systems*, pp. 98–113. North Holland, Amsterdam, 1955.

211. Burton G. Malkiel. Returns from investing in equity mutual funds 1971 to 1991. *Journal of Finance*, 50:549–572, 1995.

212. Burton G. Malkiel. *A Random Walk Down Wall Street.* Norton, New York, 1996.

213. James D. Malley and John Hornstein. Quantum statistical inference. *Statistical Science*, 8:433–457, 1993.

214. Benoit B. Mandelbrot. Une classe de processus stochastiques homothétiques à soi; application á la loi climatologique de H. E. Hurst. *Comptes rendus des Séances de l'Académie des Sciences*, 260:3274–3277, 1965.

215. Benoit B. Mandelbrot. Self-affine fractal sets. In L. Pietronero and E. Tosatti, editors, *Fractals in Physics*, pp. 3–28. North-Holland, Amsterdam, 1986.

216. Benoit B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk.* Springer, New York, 1997. This volume includes contributions Mandelbrot made to finance theory in the 1960s, with retrospective comments, together with reports on new developments, including work with Fisher and Calvet on the multifractal model of asset returns; see also [47], [124], and [217].

217. Benoit B. Mandelbrot, Adlai Fisher, and Laurent Calvet. A Multifractal Model of Asset Returns. Cowles Foundation Discussion Paper No. 1164, September 1997. Yale University. Available from the SSRN database http://www.ssrn.com.

218. Andrei A. Markov. Исчисление вероятностей. Типография Императорской Академии Наук, Saint Petersburg, 1900. The second edition, which appeared in 1908, was translated into German by Heinrich Liebmann: *Wahrscheinlichkeitsrechnung*, Teubner, Leipzig, 1912.

219. Donald A. Martin. Borel determinacy. *Annals of Mathematics*, 102:363–371, 1975.

220. Donald A. Martin. An extension of Borel determinacy. *Annals of Pure and Applied Logic*, 49:279–293, 1990.

221. Donald A. Martin. The determinacy of blackwell games. *The Journal of Symbolic Logic*, 63:1565–1581, 1998.

222. Thierry Martin. *Probabilités et critique philosophique selon Cournot.* Vrin, Paris, 1996.

223. Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

224. Per Martin-Löf. The literature on von Mises' Kollectivs revisited. *Theoria*, 35:12–37, 1969.

225. Nina Mehta. Equity vol swaps grow up. *Derivatives Strategy*, 4:11–12, July 1999.

226. D. H. Mellor. *The Matter of Chance.* Cambridge University Press, London, 1971.

227. F. Mentré. *Cournot et la renaissance du probabilisme au XIXe siècle.* Marcel Rivière, Paris, 1908.

228. Fabio Mercurio and Ton C. F. Vorst. Option pricing with hedging at fixed trading dates. *Applied Mathematical Finance*, 3:135–158, 1996.

229. Robert C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4:141–183, 1973.

230. Paul-André Meyer. Démonstration simplifiée d'un théorème de Knight. In Paul-André Meyer, editor, *Séminaire de Probabilités V*, volume 191 of *Lecture Notes in Mathematics*, pp. 191–195. Springer, Berlin, 1971.

231. Paul-André Meyer. Quantum probability as seen by a classical probabilist. In L. Accardi and Christopher C. Heyde, editors, *Probability Towards 2000*, volume 128 of *Lecture Notes in Statistics*, pp. 235–248. Springer, New York, 1998.

232. Thomas Mikosch and Rimas Norvaiša. Stochastic integral equations without probability. *Bernoulli*, 6:401–434, 2000.

233. Jan Mycielski. On the axiom of determinateness. *Fundamenta Mathematicae*, 53:204–224, 1964.

234. Jan Mycielski and S. Świerczkowski. On the Lebesgue measurability and the axiom of determinateness. *Fundamenta Mathematicae*, 54:67–71, 1964.

235. Edward Nelson. Internal set theory: A new approach to nonstandard analysis. *Bulletin of the American Mathematical Society*, 83:1165–1198, 1977.

236. Edward Nelson. *Radically Elementary Probability Theory.* Princeton University Press, Princeton, NJ, 1987.

237. M. F. Maury Osborne. Brownian motion in the stock market. *Operations Research*, 7:145–173, 1959.

238. Fredos Papangelou. Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, 165:483–506, 1972.

239. Jeff B. Paris. $ZF \vdash \Sigma_4^0$ determinateness. *Journal of Symbolic Logic*, 37:661–667, 1972.

240. Blaise Pascal. *Pascal: Œuvres complètes.* Desclée de Brouwer, Paris, 1970. Edited by J. Mesnard, four volumes. The third volume contains Pascal's letters to Fermat on the problem of points and the problem of the gambler's ruin.

241. Blaise Pascal. *Pascal: Œuvres complètes.* Gallimard, Paris, 1998–2000. Edited by Michel Le Guern, two volumes. The first volume, which appeared in 1998, contains the

letters to Fermat on the problem of points and the gambler's ruin. The second volume appeared in 2000.

242. Valentin V. Petrov. *Sums of Independent Random Variables.* Springer, Berlin, 1975.

243. Valentin V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables.* Oxford University Press, Oxford, 1995.

244. Louis B. Picard. *Œuvres de L. B. Picard.* Barba, Paris, 1821. The play entitled *Les provinciaux à Paris* is in Volume 3. On p. 401, Jérôme takes in a gambling house as he surveys the vices of Paris: "*L'un vient de perdre son dernier écu, et il va mettre sa boîte d'or en gage pour suivre sa martingale*".

245. Jean-Paul Pier, editor. *Development of Mathematics 1900–1950.* Birkhäuser, Basel, 1994.

246. Jean-Paul Pier. Intégration et mesure 1900–1950. In *Development of Mathematics 1900–1950* [245], pp. 517–564.

247. Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270, 1890.

248. Georg Pólya. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Mathematische Zeitschrift*, 8:170–181, 1920. Görgy Pólya was born in Budapest in 1887 and died in Palo Alto in 1985. After he immigrated to the United States, Pólya used the English spelling of his first name, George. But this article used the German spelling. The article contributes only indirectly to the central limit theorem; it is concerned with conditions under which the convergence of a sequence of probability distributions is implied by the convergence of their moments. It appears to be responsible, however, for the name "central limit theorem".

249. Karl R. Popper. *Logik der Forschung.* Springer, Vienna, 1935. An English translation, *The Logic of Scientific Discovery*, was published by Hutchinson, London, in 1959.

250. Karl R. Popper. *The Open Universe.* Hutchinson, London, 1982.

251. Karl R. Popper. *Realism and the Aim of Science.* Hutchinson, London, 1983.

252. Theodore Porter. *The Rise of Statistical Thinking, 1820–1900.* Princeton University Press, Princeton, NJ, 1986.

253. Johann Radon. Theorie und Anwendungen der absolut additiven Mengenfunctionen. *Sitzungsberichte der kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse*, $122_{IIa}$:1295–1438, 1913.

254. Hans Reichenbach. Axiomatik der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 34:568–619, 1932.

255. Hans Reichenbach. Les fondements logiques des probabilités. *Annales de l'Institut Henri Poincaré*, 7:267–348, 1937.

256. Richard J. Rendleman and Brit J. Bartter. Two-state option pricing. *Journal of Finance*, 34:1093–1110, 1979.

257. Alfréd Rényi. *Foundations of Probability.* Holden-Day, San Francisco, 1970.

258. Alfréd Rényi. *Probability Theory.* North-Holland, Amsterdam, 1970.

259. Christian P. Robert. *The Bayesian Choice.* Springer, New York, 1994.

260. Abraham Robinson. *Non-Standard Analysis.* North-Holland, Amsterdam, second edition, 1974. First edition: 1966.

261. L. Chris G. Rogers. Arbitrage with fractional Brownian motion. *Mathematical Finance*, 7:95–105, 1997. As the paper says, the fact that the fractional Brownian motion is a semimartingale only in the Brownian case has been known to its author since 1989.

262. L. Chris G. Rogers and David Williams. *Diffusions, Markov Processes, and Martingales*, volume 1: Foundations. Wiley, Chichester, second edition, 1994. Reissued by Cambridge University Press in *Cambridge Mathematical Library*, 2000.

263. Paul A. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–50, Spring 1965.

264. Paul A. Samuelson. Mathematics of speculative price. *SIAM Review*, 15:1–42, 1973.

265. Leonard J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66:783–801, 1971.

266. Manfred Schäl. On quadratic cost criterion for option hedging. *Mathematics of Operations Research*, 19:121–131, 1994.

267. Claus Peter Schnorr. Klassifikation der zufallsgesetze nach komplexität und ordnung. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 16:1–21, 1970.

268. Claus Peter Schnorr. A unified approach to the definition of random sequences. *Math. Systems Theory*, 5:246–258, 1971.

269. Claus Peter Schnorr. *Zufälligkeit und Wahrscheinlichkeit: Eine algorithmische Bergründung der Wahrscheinlichkeitstheorie*, volume 218 of *Lecture Notes in Mathematics*. Springer, Berlin, 1971.

270. Laurent Schwartz. Quelques réflexions et souvenirs sur Paul Lévy. In M. Metivier, J. Neveu, and S. R. Sriniva Varadhan, editors, *Colloque Paul Lévy sur les Processus Stochastiques (22–26 juin 1987, École Polytechnique, Palaiseau)*, number 157–158 in Astérisque, pp. 13–28. Société Mathématique de France, Paris, 1988.

271. Martin Schweizer. Variance-optimal hedging in discrete time. *Mathematics of Operations Research*, 20:1–32, 1995.

272. Françoise Seillier-Moiseiwitsch and A. Philip Dawid. On testing the validity of sequential probability forecasts. *Journal of the American Statistical Association*, 88:355–359, 1993.

273. Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

274. Glenn Shafer. Bayes's two arguments for the rule of conditioning. *Annals of Statistics*, 10:1075–1089, 1982. An examination of the role of time in Thomas Bayes's theory of probability.

275. Glenn Shafer. Conditional probability. *International Statistical Review*, 53:261–277, 1985. This article argues that conditional probability is meaningful only in the context of a protocol (or event tree) that spells out the possibilities for future observations.

276. Glenn Shafer. The unity and diversity of probability. *Statistical Science*, 5:435–462, 1990. Argues that a broad conception of the foundation and applications of probability is needed in order for statistics to flourish as a profession. The article is accompanied by comments by a number of distinguished statisticians.

277. Glenn Shafer. The unity of probability. In George von Furstenberg, editor, *Acting Under Uncertainty: Multidisciplinary Conceptions*, pp. 95–126. Kluwer, Boston, 1990. Argues that knowledge of the long run, value, and warranted belief are entangled in the

picture underlying mathematical probability to the degree that no single one of these ideas suffices as a conceptual foundation.

278. Glenn Shafer. What is probability? In David C. Hoaglin and David S. Moore, editors, *Perspectives on Contemporary Statistics*, MAA Notes 21, pp. 93–105. Mathematical Association of America, 1992. A philosophical and historical introduction to the competing interpretations of probability.

279. Glenn Shafer. Can the various meanings of probability be reconciled? In Gideon Keren and Charles Lewis, editors, *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, pp. 165–196. Lawrence Erlbaum, Hillsdale, NJ, 1993. Develops the elementary theory of probability and price in trees and shows how it can be used to reconcile competing philosophies of probability.

280. Glenn Shafer. The early development of mathematical probability. In Grattan-Guinness [137], pp. 1293–1302. Volume 2. This article surveys the development of the mathematics and philosophy of probability from Pascal to Laplace.

281. Glenn Shafer. The subjective aspect of probability. In George Wright and Peter Ayton, editors, *Subjective Probability*, pp. 53–73. Wiley, New York, 1994. A philosophical examination of the use of probability models as benchmarks or standards to which to compare claims of predictive efficacy. The conclusion is that the probabilities in such models satisfy neither Bayesian nor frequentist claims about the general meaning of probability.

282. Glenn Shafer. *The Art of Causal Conjecture*. MIT Press, Cambridge, MA, 1996.

283. Glenn Shafer. The significance of Jacob Bernoulli's *Ars Conjectandi* for the philosophy of probability today. *Journal of Econometrics*, 75:15–32, 1996.

284. Glenn Shafer. Mathematical foundations for probability and causality. In Frederick Hoffman, editor, *Mathematical Aspects of Artificial Intelligence*, volume 55 of *Symposia in Applied Mathematics*, pp. 207–270. American Mathematical Society, Providence, RI, 1998. Develops axiomatically the concepts of an event space, which synthesizes the concepts of Boolean algebra and event tree. In an event space, one situation may follow another in time (as in an event tree) or merely be more detailed than another (as in a Boolean algebra).

285. Glenn Shafer, Peter R. Gillett, and Richard B. Scherl. Probabilistic logic. *Annals of Mathematics and Artificial Intelligence*, 2000. Provides a more concise axiomatization for event trees, which were first axiomatized in [284].

286. Robert J. Shiller. *Irrational Exuberance*. Princeton University Press, Princeton, NJ, 2000.

287. Albert N. Shiryaev. *Probability*. Springer, New York, second edition, 1996. The first English edition appeared in 1984. Both editions where translated by R. P. Boas from the Russian originals, Вероятность, published by Nauka in Moscow in 1980 and 1989.

288. Andrei Shleifer. *Inefficient Markets*. Oxford University Press, Oxford, 2000.

289. Anatolii V. Skorohod. *Random Processes with Independent Increments*. Kluwer, Dordrecht, 1991. Russian original: Случайные процессы с независимыми приращениями, Nauka, Moscow, 1986 (second edition).

290. David E. Smith. *A Source Book in Mathematics*. McGraw-Hill, New York, 1929.

291. Marian von Smoluchowski. Zarys kinetycznej teoriji ruchów browna i roztworów metnych. *Rozprawy i Sprawozdania z Posiedzeń Wydzialu Matematyczno-Przyrodniczego*

*Akademii Umiejetności*, 3:257–282, 1906. A German translation appeared in the *Annalen der Physik*, 21, pp. 756–780 in 1906.

292. J. Laurie Snell. Gambling, probability, and martingales. *Mathematical Intelligencer*, 4(3):118–124, 1982.

293. J. Laurie Snell. A conversation with Joe Doob. *Statistical Science*, 12:301–311, 1997.

294. Ray J. Solomonoff. A formal theory of inductive inference. parts I and II. *Information and Control*, 7:1–22 and 224–254, 1964.

295. Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press, Cambridge, MA, 1986.

296. Stephen M. Stigler. Laplace's 1774 memoir on inverse probability. *Statistical Science*, 1:359–378, 1986.

297. William F. Stout. A martingale analogue of Kolmogorov's law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 15:279–290, 1970.

298. William F. Stout. *Almost Sure Convergence.* Wiley, New York, 1973.

299. Daniel W. Stroock and S. R. Sriniva Varadhan. Diffusion processes with continuous coefficients, I and II. *Communications in Pure and Applied Mathematics*, 22:345–400 and 479–530, 1969.

300. Daniel W. Stroock and S. R. Sriniva Varadhan. *Multidimensional Diffusion Processes.* Springer, Berlin, 1979.

301. S. James Taylor. Exact asymptotic estimates of Brownian path variation. *Duke Mathematical Journal*, 39:219–241, 1972.

302. William M. Thackeray. *The Newcomes: Memoirs of a Most Respectable Family.* Bradbury and Evans, London, 1853–1855. Monsieur de Florac admonishes Clive, "You have not played as yet? Do not do so; above all avoid a martingale, if you do. Play ought not to be an affair of calculation, but of inspiration." (Volume I, Chapter 28, p. 259.) The novel appeared in monthly installments starting in 1853 and concluding in 1855.

303. Isaac Todhunter. *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.* Macmillan, London, 1865. Extraordinarily dry, Todhunter's *History* was the authoritative history of early probability theory until Hald's history appeared in 1990. It was reprinted by Chelsea, New York, in 1949 and 1965.

304. Marco R. Vervoort. Blackwell games. In Thomas Ferguson, Lloyd Shapley, and James MacQueen, editors, *Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell*, volume 30 of *IMS Lecture Notes-Monograph Series*, pp. 369–390. Institute of Mathematical Statistics, 1996.

305. Jean Ville. Sur la théorie des jeux ou intervient l'habileté des joueurs. In *Application aux jeux de hasard* [35], pp. 105–113.

306. Jean Ville. *Étude critique de la notion de collectif.* Gauthier-Villars, Paris, 1939. This differs from Ville's dissertation, which was defended in March 1939, only in that a one-page introduction was replaced by a 17-page introductory chapter.

307. Jean-André Ville. Sur les suites indifférentes. *Comptes rendus des Séances de l'Académie des Sciences*, 202:1393–1394, 1936. Session of April 27, 1936.

308. Jean-André Ville. Sur la notion de collectif. *Comptes rendus des Séances de l'Académie des Sciences*, 203:26–27, 1936. Session of July 6, 1936.

309. Richard von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919.

310. Richard von Mises. *Wahrscheinlichkeitsrechnung, Statistik und Wahrheit*. Verlag von Julius Springer, Wien, 1928. The second edition appeared in 1936 and the third in 1951. A posthumous fourth edition, edited by his wife Hilda Geiringer, appeared in 1972. English editions, under the title *Probability, Statistics and Truth*, appeared in 1939 and 1957.

311. Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. F. Deuticke, Leipzig und Wien, 1931.

312. Richard von Mises. Quelques remarques sur les fondements du calcul des probabilités. In Rolin Wavre, editor, *Les fondements du calcul des probabilités*, number 735 in *Actualités Scientifiques et Industrielles*, pp. 57–66. Hermann, Paris, 1938. In this article, von Mises responds to the criticisms aired by Fréchet and others at the 1937 Geneva colloquium. *Les fondements du calcul des probabilités* is the second fascicle of [338].

313. Richard von Mises. On the foundations of probability and statistics. *Annals of Mathematical Statistics*, 12:191–205, 1941. This article originated as an address to the Institute of Mathematical Statistics in Hanover, NH, in September 1940. It was published together with an article by Doob [98] and comments by von Mises and Doob on each other's articles [315].

314. Richard von Mises. *Selected Papers of Richard von Mises. Volume One: Geometry, Mechanics, Analysis*. American Mathematical Society, Providence, RI, 1963. Selected and edited by Ph. Frank, S. Goldstein, M. Kac, W. Prager, G. Szegö, and G. Birkhoff, Chairman.

315. Richard von Mises and Joseph L. Doob. Discussion of papers on probability theory. *Annals of Mathematical Statistics*, 12:215–217, 1941. Discussion of [313] and [98].

316. John von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton, NJ, 1955. First German edition: *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin, 1932.

317. John von Neumann and Oscar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, third edition, 1953. Previous editions appeared in 1944 and 1947.

318. Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*. Cambridge University Press, Cambridge, 1994. A study of the emergence of measure-theoretic probability in the first third of the 20th century. It emphasizes statistical physics and includes a detailed look at the foundational work of Kolmogorov and de Finetti.

319. Vladimir G. Vovk. On the concept of the Bernoulli property. *Russian Mathematical Surveys*, 41:247–248, 1986. Studies connections between Kolmogorov's finitary variant of von Mises's notion of collective and a more traditional measure-theoretic variant.

320. Vladimir G. Vovk. The law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. *Theory of Probability and Its Applications*, 32:413–425, 1987. This paper gives an algorithmic proof of the law of the iterated logarithm which was adapted to a game-theoretic framework in Dawid and Vovk [86].

321. Vladimir G. Vovk. Kolmogorov-Stout law of the iterated logarithm. *Mathematical Notes*, 44:502–507, 1988. A more general version of the proof given in [320].

322. Vladimir G. Vovk. Finitary prequential probability: Asymptotic results. Technical report, Institute of New Technologies, Moscow, 1991. Game-theoretic finitary versions of strong limit theorems (not in this book). The two examples of §8.4.

323. Vladimir G. Vovk. Forecasting point and continuous processes: Prequential analysis. *Test*, 2:189–217, 1993. Versions of Theorem 11.1 (p. 278) and Proposition 12.1 (p. 304).

324. Vladimir G. Vovk. A logic of probability, with applications to the foundations of statistics (with discussion). *Journal of the Royal Statistical Society. Series B*, 55:317–351, 1993. This paper argues for basing the theory and applications of probability on the principle of impossibility of a gambling system. It proves a variant of Ville's finitary theorem (p. 195) and a prequential (essentially game-theoretic) central limit theorem.

325. Vladimir G. Vovk. Central limit theorem without probability, June 1995. Bachelier's central limit theorem (Proposition 10.1 on p. 242).

326. Vladimir G. Vovk. Game-theoretic versions of Kolmogorov's strong law of large numbers, June 1995. Predictive and finance-theoretic strong law of large numbers (Proposition 4.1 on p. 78 and Proposition 15.1 on p. 356).

327. Vladimir G. Vovk. Pricing European options without probability, June 1995. A version of the game-theoretic Black-Scholes formula (Theorem 11.2 on p. 280).

328. Vladimir G. Vovk. A purely martingale version of Lindeberg's central limit theorem, June 1995. A variant of Theorem 7.1 on p. 153.

329. Vladimir G. Vovk. A strictly martingale version of Kolmogorov's strong law of large numbers. *Theory of Probability and Its Applications*, 41:605–608, 1996. A variant of Proposition 4.4 on p. 92.

330. Vladimir G. Vovk. Probability theory for the Brier game. In *Proceedings of the Eighth International Workshop on Algorithmic Learning Theory*, 1997. Accepted for publication in *Theoretical Computer Science*. Limit theorems of probability theory for nonlinear protocols (not in this book).

331. Vladimir G. Vovk. Kolmogorov's complexity conception of probability. Technical Report CLRC-TR-00-01, Computer Learning Research Centre, Royal Holloway, University of London, January 2000. This report describes one of the sources of the game-theoretic approach, Kolmogorov's algorithmic and finitary approach to the foundations of probability. To appear in *Proceedings of the 1999 Conference on Statistics: Philosophy, Recent History and Relations to Science, Roskilde, Denmark*. Synthese Library Series, Kluwer Academic Publishers. Editors Vincent F. Hendricks, Stig Andur Pedersen, and Klaus Frovin Jørgensen.

332. Vladimir G. Vovk. Black-Scholes formula without stochastic assumptions. Technical Report CLRC-TR-00-02, Computer Learning Research Centre, Royal Holloway, University of London, March 2000. Also in *Models for Credit Risk*, pp. 149–154. UNICOM Seminars, London, May 2000. Nonstochastic Black-Scholes using squares for hedging (§12.2).

333. Vladimir G. Vovk and Chris J. H. C. Watkins. Universal portfolio selection. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, pp. 12–23, 1998.

334. Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, 8:38–72, 1937. This journal, or series of publications, reported results from Karl Menger's famous Vienna

Colloquium. Participants included von Neumann, Morgenstern, and Wald. The eighth volume was the last in the series, because the colloquium ended with the Nazi invasion of Austria in 1938. In 1939, Menger started a second series in English (*Reports of a mathematical colloquium*), at the University of Notre Dame.

335. Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes. In Rolin Wavre, editor, *Les fondements du calcul des probabilités*, number 735 in *Actualités Scientifiques et Industrielles*, pp. 79–99. Hermann, Paris, 1938. This is an abridged version of [334]. *Les fondements du calcul des probabilités* is the second fascicle of [338].

336. Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

337. Christian Walter. Une histoire du concept d'efficience sur les marchés financiers. *Annales: Histoire, Sciences Sociales*, number 4:873–905, July-August 1996.

338. Rolin Wavre. *Colloque consacré à la théorie des probabilités*. Hermann, Paris, 1938–1939. This celebrated colloquium was held in October 1937 at the University of Geneva, as part of a series (*Conférences internationales des sciences mathématiques*) that began in 1933. The colloquium was organized by Rolin Wavre and chaired by Maurice Fréchet. Other participants included Cramér, Dœblin, Feller, de Finetti, Heisenberg, Hopf, Lévy, Neyman, Pòlya, Steinhaus, and Wald, and communications were received from Bernstein, Cantelli, Glivenko, Jordan, Kolmogorov, von Mises, and Slutsky. The proceedings of the colloquium were published by Hermann in eight fascicles of 50 to 100 pages each, in their series *Actualités Scientifiques et Industrielles*. The first seven fascicles appeared in 1938 as numbers 734 through 740 of this series; the eighth appeared in 1939 as number 766. The second fascicle, entitled *Les fondements du calcul des probabilités*, includes contributions by Feller, Fréchet, von Mises, and Wald. The eighth fascicle consists of de Finetti's summary of the colloquium.

339. Robert E. Whaley. Derivatives on market volatility: Hedging tools long overdue. *Journal of Derivatives*, 1:71–84, 1993.

340. Robert E. Whaley. The investor fear gauge. *Journal of Derivatives*, 26:12–17, 2000.

341. Norbert Wiener. The mean of a functional of arbitrary elements. *Annals of Mathematics*, 22:66, 1920.

342. Norbert Wiener. The average of an analytical functional. *Proceedings of the National Academy of Science*, 7:253, 1921.

343. Norbert Wiener. The average of an analytical functional and the Brownian movement. *Proceedings of the National Academy of Science*, 7:294, 1921.

344. Norbert Wiener. Differential space. *Journal of Mathematics and Physics*, 2:131, 1923.

345. Norbert Wiener. The average value of a functional. *Proceedings of the London Mathematical Society*, 22:454, 1924.

346. Norbert Wiener. *Collected Works*. MIT Press, Cambridge, MA, 1976–1985. Four volumes. Edited by P. Pisani. Volume 1 includes Wiener's original papers on Brownian motion [341, 342, 343, 344, 345], with a commentary by Kiyosi Itô.

347. David Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.

348. Peter M. Williams. Indeterminate probabilities. In M. Przełęcki, K. Szaniawski, and R. Wojcícki, editors, *Formal Methods in the Methodology of Empirical Sciences*, pp. 229–246. Ossolineum & Reidel, Wrocław, 1976.

349. Walter Willinger and Murad S. Taqqu. Pathwise stochastic integration and application to the theory of continuous trading. *Stochastic Processes and their Applications*, 32:253–280, 1989.

350. Walter Willinger and Murad S. Taqqu. Towards a convergence theory for continuous stochastic securities market models. *Mathematical Finance*, 1:55–100, 1991.

351. Paul Wilmott. *Derivatives: The Theory and Practice of Financial Engineering*. Wiley, Chichester, 1998. An engaging and leisurely (739 pages) introduction to the pricing of derivatives.

352. Paul Wilmott, Jeff Dewynne, and Sam Howison. *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press, Oxford, 1993. A readable book aimed at the applied mathematician.

353. Philip Wolfe. The strict determinateness of certain infinite games. *Pacific Journal of Mathematics*, 5:841–847, 1955.

354. L. C. Young. An inequality of the Hölder type, connected with Stiltjes integration. *Acta Math.*, 67:251–282, 1936.

355. Ernst Zermelo. Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In E. W. Hobson and A. E. H. Love, editors, *Proceedings of the Fifth International Congress of Mathematicians*, volume II, pp. 501–504. Cambridge University Press, Cambridge, 1913.

356. Vladimir M. Zolotarev. Современная теория суммирования независимых случайных величин *(The modern theory of summation of independent random variables)*. Nauka, Moscow, 1986.

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# *Notation*

$A := B$: $A$ equals $B$ by definition, 12
$E \Longrightarrow F$: material implication, 79
$x \cdot y$: dot product, 353
log: logarithm to the base 2, 253

$\emptyset$: empty set, 40
$\mathbb{R}$: the real numbers, 64
$\mathbb{Q}$: the rational numbers, 85
$\mathbb{N}$: the positive integers, 190, 283
$\mathbb{Z}^+$: the nonnegative integers, 303

$\Delta A_n := A_n - A_{n-1}$: previous increment, 129
$dA_n := A_{n+1} - A_n$: next increment, 132

$\Omega$: game-theoretic sample space, 9, 148, 183
$\xi$: path, 11
$\xi^n$: initial segment of $\xi$, 66
$|\xi|$: length of path $\xi$, 183
$E$: event, 15
$\mathbb{I}_E$: indicator variable for $E$, 15
$E^c$: complement of $E$, 16
$\Omega^\diamond$: set of all situations, 66, 149, 183
$\Box$: initial situation, 10, 66, 149, 183
$t$: situation, 11
$|t|$: the length of $t$, 66

$s \sqsubseteq t$: $s$ precedes $t$, 66, 149
$s \sqsubset t$: $s$ strictly precedes $t$, 149
$sx$: concatenation of $s$ with $x$, 66
$t_u^+$: next situation towards $u$, 152

$m_n$: Forecaster's move, 70
$v_n$: Forecaster's variance move, 77
$\mathbf{f}_n$: Forecaster's abstract move, 90
$\mathbf{F}$: Forecaster's move space, 90
$M_n$: Skeptic's move, 64
$V_n$: Skeptic's variance move, 77
$\mathbf{s}_n$: Skeptic's abstract move, 90
$\mathbf{S}$: Skeptic's move space, 90
$\mathbf{S}_t$: Skeptic's move space in $t$, 183
$x_n$: Reality's move, 64
$\mathbf{r}_n$: Reality's abstract move, 90
$\mathbf{R}$: Reality's move space, 90
$\mathbf{W}_t$: World's move space in $t$, 149, 183

$\lambda$: Skeptic's gain function, 91
$\lambda_t$: Skeptic's gain function in $t$, 150, 183
$\mathcal{K}_n$: Skeptic's capital, 64
$\mathcal{P}$: strategy for Skeptic, 11, 66, 183
$\mathcal{K}^\mathcal{P}$: Skeptic's capital process, 11, 67, 150, 184

$\mathcal{A}$: quadratic supervariation, 91

$\mathbf{0}$: variable identically equal to 0, 14
$\mathcal{S}_{\Omega}$: variable from process $\mathcal{S}$, 149
$\mathcal{K}^{\mathcal{P}}(\xi)$: Skeptic's final capital, 11

$\overline{\mathbb{E}}\,x$: upper price, 12
$\underline{\mathbb{E}}\,x$: lower price, 13
$\overline{\mathbb{E}}_t\,x$: upper price in $t$, 13, 151, 184
$\underline{\mathbb{E}}_t\,x$: lower price in $t$, 13, 151, 184
$\mathbb{E}_t\,x$: game-theoretic price in $t$, 14, 151, 184
$\overline{\underline{\mathbb{E}}}_t\,x$: upper and lower prices in $t$, 152
$\overline{\mathbb{V}}\,x$: upper variance, 14, 152
$\underline{\mathbb{V}}\,x$: lower variance, 14, 152
$\mathbb{V}\,x$: game-theoretic variance, 14
$\overline{\mathbb{P}}\,E$: upper probability, 15
$\underline{\mathbb{P}}\,E$: lower probability, 15
$\overline{\underline{\mathbb{P}}}\,E$: upper and lower probability, 121
$\overline{\mathrm{Var}}_{\mathcal{S}}$: accumulated upper variance, 153
$\underline{\mathrm{Var}}_{\mathcal{S}}$: accumulated lower variance, 153
$d_t\mathcal{S}$: gain of $\mathcal{S}$ in $t$, 152

$\mathbb{P}\,E$: classical probability, 33

$\Omega$: measure-theoretic sample space, 40
$\mathcal{F}$: $\sigma$-algebra, 40
$\mathcal{P}(B)$: set of all probability measures on $B$, 190, 192, 360
$\mathbb{P}\,E$: measure-theoretic probability, 40
$\mathbb{P}[F \mid E]$: measure-theoretic conditional probability, 40
$x$: random variable, 41
$\mathbb{E}\,x$: measure-theoretic expected value, 42
$\mathbb{V}\,x$: measure-theoretic variance, 42
$\mathbb{E}[x \mid \mathcal{G}]$: measure-theoretic conditional expectation, 169
$\mathbb{E}[y|x]$: measure-theoretic conditional expectation, 42
$\mathbb{V}[x \mid \mathcal{G}]$: measure-theoretic conditional variance, 170
$\mathbb{E}_t\,x$: measure-theoretic expected value conditional on information at time $t$, 227
$\psi$: characteristic function, 311
$\mathcal{N}_{\mu,\sigma^2}$: Gaussian distribution, 122
$\overline{U}$: Gaussian smoothing of $U$, 123
$\mathcal{P}_D$: Poisson distribution, 303
$W$: Wiener process, 205
$B_h$: fractional Brownian motion, 212

$\Pi$: Lévy measure, 312
$\mu$: drift of diffusion process, 206
$\sigma$: volatility of diffusion process, 206
$\mathcal{T}[0, D]$: set of all stopping times, 144

$\overline{A}$: the closure of a set $A$, 136
$\partial A$: the boundary of a set $A$, 136

$B(s, D, \delta)$: space/time rectangle, 136
$\mathrm{LM}\,u$: the least superparabolic majorant of $u$, 138
$\mathrm{LSM}\,u$: the least supermarket majorant of $u$, 331
PI: parabolic Poisson integral, 138
$\tau_B u$, 138
$\tilde{u}$: lower semicontinuous smoothing, 138

$\mathbf{var}_{S,N}$: variation spectrum for $S$, 209
$\mathbf{var}_N$: same as $\mathbf{var}_{S,N}$, 209
$\mathbf{var}_S$: same as $\mathbf{var}_{S,N}$, 209
$\mathbf{var}_S^{\mathrm{rel}}$: relative variation spectrum, 259
$\overline{\mathbf{var}}_S$: strong variation spectrum, 289
$H(S)$: Hölder exponent of $S$, 212
$\dim S$: box dimension of $S$, 212
$\mathrm{vex}\,S$: variation exponent of $S$, 212
$\overline{\mathrm{vex}}\,S$: strong variation exponent, 289

$\mathcal{S}$: underlying security, 201
$\mathcal{I}$: Investor's capital process, 218, 344
$I$: Investor's final capital, 319
$\mathcal{U}$: European option, 216
$\mathcal{D}$: dividend-paying derivative, 222
$\mathcal{B}$: risk-free bond, 295
$\mathcal{R}$: strictly convex option, 298

$\overline{\overline{\mathbb{E}}}H$: strong upper price of option, 318
$\overline{\mathbb{E}}\,H$: upper price of option, 317
$\underline{\mathbb{E}}\,H$: lower price of option, 317

*Probability and Finance: It's Only a Game!* Glenn Shafer, Vladimir Vovk

ISBN: 0-471-40226-5

# *Index*