CONTEMPORARY MATHEMATICS

1

Markov Random Fields and Their Applications

Ross Kindermann
J. Laurie Snell

American Mathematical Society
Providence, Rhode Island

Prepared with support from the Sloan Foundation grant S32-951 and the National Science Foundation LOCI grant SER77-02326.

2000 *Mathematics Subject Classification.* Primary 60K35.

Library of Congress Cataloging-in-Publication Data

Kindermann, Ross, 1950–
Markov random fields and their applications.
(Contemporary mathematics; v. 1)
Bibliography: p.
1. Random fields. 2. Markov processes. I. Snell, James Laurie, joint author. II. American Mathematical Society. III. Title. IV. Series: Contemporary mathematics (Providence, R.I.); v. 1.
QA274.45.K56 519.2 80-22764
ISBN 0-8218-5001-6

---

10 9 8 7 6 5 4 04 03 02 01 00

CONTENTS

PREFACE

Markov random fields is a new branch of probability theory that promises to be important both in the theory and application of probability. The existing literature on the subject is quite technical and often only understandable to the expert. This paper is an attempt to present the basic ideas of the subject and its application to a wider audience. We have relied on examples and computer graphics to convey the meaning of the results when they are too technical to prove. This graphical work was made possible by a grant to Dartmouth College from the Sloan Foundation. This work is also part of a project to produce modules on applications of probability under a grant from the National Science Foundation. The authors wish also to acknowledge the support of Kiewit Computation Center at Dartmouth College.

We were assisted by Robert Beck and William Myers in the writing of computer programs. We are very grateful to Marie Slack for a fine job of typing the material in a form suitable for a sometimes friendly and sometimes less-than-friendly computer.

The work of Ross Kindermann was supported by a Maude Hammond Fling Faculty Research Fellowship awarded through the University of Nebraska-Lincoln Research Council.

MARKOV RANDOM FIELDS AND THEIR APPLICATIONS

Ross Kindermann

J. Laurie Snell

1. THE ISING MODEL

In recent years, a new type of stochastic process, called a Markov random field, has been introduced in the theory of probability. The motivation for looking at such processes came originally from statistical physics, but it is clear that these processes form a natural generalization of Markov processes in which a time index is replaced by a space index.

The foundations of the theory of Markov random fields may be found in Preston (1974) or Spitzer (1971). The purpose of this work is to present this subject at a level which will make the material available to people outside of mathematics, as well as to discuss certain of its applications to other areas. It would seem that Markov fields should enjoy the same wide variety of applications that Markov chains have.

The concept of a Markov random field came from attempts to put into a general probabilistic setting a very specific model named after the German physicist Ernst Ising. Ising was a student of Lenz and wrote his doctoral thesis on a model now called the Ising model. He tried to explain, using this model, certain empirically observed facts about ferromagnetic materials. When Ising (1925) published a summary of his results, he stated that the model was suggested by Lenz. A paper written by Lenz (1920) gives a very sketchy idea of the model. For an

interesting historical discussion of the origins and development of the Ising model see Brush (1967).

The first formulation given by Ising is as follows: consider a sequence, $0,1,2,..,n$ of points on the line. At each point, or site, there is a small dipole or "spin" which at any given moment is in one of two positions, "up" or "down". It is customary now to indicate the spins in the form of a configuration as shown in Figure 1.

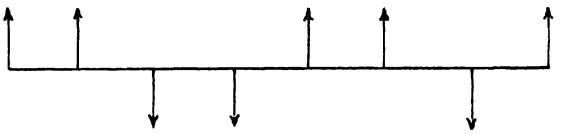


FIGURE 1

Following Ising, we are going to put a probability measure on the set of all possible configurations. Such a measure is called a random field. Using current probability notation we choose as sample space the space Ω of all sequences

$$\omega = (\omega_0, \omega_1, \ldots, \omega_n)$$

where $\omega_j = +$ or $-$ with "+" indicating a spin up and "-" a spin down. Then we can think of the spin σ_j as a function defined on Ω such that $\sigma_j(\omega) = 1$ if $\omega_j = +$ and -1 if $\omega_j = -$. Ising defined a probability measure on Ω as follows. To each configuration ω an energy $U(\omega)$ is assigned by

$$(1) \quad U(\omega) = -J\sum_{i,j} \sigma_i(\omega)\sigma_j(\omega) - mH\sum_i \sigma_i(\omega).$$

Here the first sum is taken over all pairs i,j of points which are one unit apart. (We count each pair only once.) The first term represents the energy caused by the interaction of the

spins. Ising made the simplifying assumption that only interactions between neighboring spins need be taken into account. The constant J is a property of the material being considered. The case $J > 0$ is called the attractive case. The reason for this is that the interaction tends to keep neighboring spins aligned the same. The case $J < 0$ is called the repulsive case since it tends to reinforce pairs in which the spins are of opposite orientation. The second term represents the effect of an external magnetic field of intensity H. The constant $m > 0$ is a property of the material. In the attractive case, the first term contributes minimum energy when all of the spins are lined up in the same direction. The second term contributes minimum energy when all the spins are in the same direction as the external field.

Ising then assigned probabilities to configurations ω proportional to

$$(2) \qquad e^{-\frac{1}{kT}U(\omega)}$$

where T is the temperature and k is a universal constant. The probability measure on Ω is thus given by

$$(3) \qquad P(\omega) = \frac{e^{-\frac{1}{kT}U(\omega)}}{Z}$$

where the normalizing constant Z, defined by

$$(4) \qquad Z = \sum_{\omega} e^{-\frac{1}{kT}U(\omega)},$$

is called the partition function.

A useful way to think of this measure is the following. Let us associate with each point i an energy U_i equal to

$$U_i(\omega) = -\frac{J}{2}\sum_{|j-i|=1}\sigma_i(\omega)\sigma_j(\omega) - mH\sigma_i(\omega) \ .$$

Then

$$P(\omega) = \frac{1}{Z}\prod_i e^{-\frac{1}{kT}U_i(\omega)} \ .$$

Thus the relative probability of a configuration is simply obtained by taking a product over all the points and using the energy at each point to determine the weight of that point.

We note that we could equally well have considered the spins located on a set of lattice points in 2 or more dimensions. A point i is then replaced by a point with two coordinates (i,j) where i and j are integers. A typical configuration in a two dimensional lattice is shown in Figure 2. The energy defined by (1) allows interaction between a point and its neighbors and the probability measure is defined in exactly the same way by (2), (3), and (4). Note that in two dimensions a point will normally have 4 neighbors unless it is on the boundary, when it will have 2 or 3.

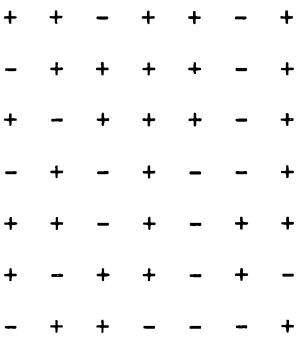


FIGURE 2

While Ising discussed only the magnetic interpretation, the same model has since been found applicable to a number of other physical and biological systems such as gases, binary alloys, and

cell structures. A sociologically oriented application has been suggested by Weidlich (1971). Here one considers a group of people, each of whom at a given moment is a "conservative" ("up") or a "liberal" ("down"). The energy (1) might better be called "tension". The first term in (1) is the tension caused by people interacting. The external field represents, for example, the current state of the government, liberal or conservative. Minimum tension (maximum boredom) occurs if all people agree and agree with the government. Of course in such an application we would want to drop the restriction to a regular lattice and specific neighbors but we shall see that this generalization can be made. Such applications will be discussed in more detail in the final section.

A probability measure of the form (3) defined by an energy function U is called a Gibbs measure. It is, of course, natural to ask why this particular measure should be interesting. There are two explanations. The first is in terms of entropy and probably led to the use of these measures in statistical mechanics.

For any probability measure $p(\omega)$ on a finite space Ω, the the entropy $S(p)$ is defined by

$$(5)\quad S(p) = -\sum_{\omega} p(\omega) \log p(\omega).$$

The entropy of a measure may be interpreted as the amount of uncertainty in the outcome. For example, if Ω has n points, the measure with greatest entropy is the measure which assigns all outcomes equal probability.

In the typical application of the Ising model, one is trying to assign a probability measure to a sample space Ω which represents outcomes which cannot be observed. In practice, only

very broad properties of the system can be observed. Assume that we could estimate at least the expected value of the energy, $E(U) = a$. Then the Gibbs measure defined by (3) is the measure which maximizes entropy among all measures which make the expected energy agree with our estimated value a. This will be discussed further in Section 3. Thus we have chosen the measure which has the greatest uncertainty, as measured by entropy, among all possible measures with given expected energy. A more detailed discussion of this approach may be found in Jaynes (1957).

A second and more important characterization, from the point of view of probability theory, is the following. We shall show in Section 2 that the Gibbs measure with nearest neighbor interaction has the following property in terms of conditional probability. Let N_j be the set of neighbors of a vertex j, that is, all points which are one unit away. Then

$$(6)\quad P(\sigma_j = a \mid \sigma_k,\ k \neq j) = P(\sigma_j = a \mid \sigma_k,\ k \text{ in } N_j).$$

Property (6) is in fact a Markov type property, i.e., the probability of a certain spin at the point j, given the spin values at all the other points of the lattice, is the same as the probability of that spin at point j, given only the spin values at the neighbors of the point j. A measure with property (6) is called a Markov random field. The probabilities expressed in (6) are called local characteristics. In the one-dimensional case, this is a generalization of a two-state Markov chain. The measure determined by such a Markov chain has property (6) when we interpret the time points as sites. Of course in two dimensions we cannot interpret the sites as time but we can consider measures defined by (3) or measures with property (6) as

generalizations of Markov processes to spatial situations. Not only does a Gibbs measure have property (6) but conversely, any strictly positive measure with property (6) can be represented as a Gibbs measure with a suitable choice of energy function. We shall in fact see that this is true if we replace a lattice by an arbitrary finite graph.

As an introduction to some of the new problems encountered in the study of Markov fields, we return to Ising's work and ask what he hoped to establish. Basically, he was interested in the case of no exterior field. It was thought that for sufficiently low temperatures, even if the spins were random to begin with, they would tend to move to a state of lower energy, i.e., mostly up or mostly down, forming a magnet. Thus, in equilibrium, if $n_+(\omega)$ is the number of up spins and $n_-(\omega)$ is the number of down spins, the total magnetization

$$M(\omega) = n_+(\omega) - n_-(\omega)$$

would be expected to have a distribution with two peaks as in Figure 3. This would result in "spontaneous magnetization", i.e., the spins would tend to be either mostly + or mostly - .

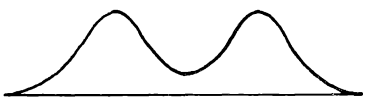


FIGURE 3

Ising remarked that this did not occur and then went on to consider the case of two dimensions. Here he made a mistake which held back the development of his model for many years. He argued that this magnetic effect would be even more noticeable if

in two dimensions he allowed different attractive forces J_1 and J_2 for the two possible directions. He in fact considered the case where the force J_2 in the vertical direction went to infinity and remarked that then all rows would be forced to conform to each other as in Figure 4.

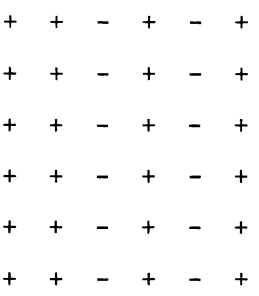


FIGURE 4

He then could apply his one-dimensional result and was again led to a unimodal distribution for the magnetization M. Thus he came to the conclusion that his model was too crude to explain magnetization. Ising was forced to leave Germany in 1936 and was cut off from the scientific community and unable to pursue his work. About the only immediate attention given to his paper was by Heisenberg (1928) who used the apparent failure as a reason to introduce a more complicated model. However, interest in the Ising model was revived by Bethe (1935) and others interested in applications such as formation of binary alloys. Peierls (1936) developed a method to show that in two or more dimensions the "spontaneous magnetization effect" could be seen to occur in the Ising model. His proof was not quite rigorous and careful proofs were given later by Griffiths (1964) and Dobrushin (1965) independently. However, the method introduced by Peierls is still a major tool in the study of the Ising model and so we shall discuss his technique in some detail. It is amusing that

by 1936 Peierls assumed that the Heisenberg model was the better one and remarked that the Ising model is therefore "now only of mathematical interest." While both models are still important, the Ising model continues to enjoy great success in a wide variety of applications.

We return now to the one dimensional case. The total magnetization with n spins is a random variable M_n and Ising was interested in the distribution of M_n. If we had in fact no interaction and zero exterior field, then all configurations would have the same probability and we would have independent trials with "+" occurring with probability 1/2 and "-" with probability 1/2. The law of large numbers would say that

$$P(|\frac{M_n}{n}| > d) \dashrightarrow 0$$

as n tends to infinity for any $d > 0$. Furthermore the central limit theorem would say that M_n, properly scaled, would be approximately normally distributed. Assume now that we do allow interaction but no exterior field. Then our Gibbs measure has the form

$$P(\omega) = \frac{1}{Z} e^{\frac{J}{kT} \sum_{i,j} \sigma_i(\omega)\sigma_j(\omega)} .$$

We can rewrite this as

$$P(\omega) = \frac{1}{Z} e^{\frac{J}{kT} (n_e(\omega) - n_o(\omega))}$$

where $n_e(\omega)$ is the number of even bonds (pairs of adjacent spins of the same sign) and $n_o(\omega)$ is the number of odd bonds (pairs of adjacent spins of opposite sign). If n_b is the total number of bonds we can rewrite this as

$$P(\omega) = \frac{1}{Z} e^{\frac{J}{kT} (n_b(\omega) - 2n_o(\omega))} .$$

Finally, since $n_b(\omega)$ is a constant we can write our probability measure in the form

$$P(\omega) = \frac{1}{Z} e^{-bn_o(\omega)}$$

where $b = \frac{2J}{kT}$ and Z is a new constant. Assume now that a probability measure on the space Ω of possible configurations of n spins is assigned in an apparently very different way. Specifically, assume that it is assigned as if the spins were generated by a two state Markov chain with transition matrix

$$P = \begin{array}{c} \\ - \\ + \end{array} \begin{array}{c} \begin{array}{cc} - & + \end{array} \\ \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \end{array} .$$

We assume the initial state is chosen at random. Then the probability $\bar{P}(\omega)$ for a specific sequence of spins, for example,

$$\omega = (+,-,-,+,-,\ldots,+) ,$$

is

$$\bar{P}(\omega) = \frac{1}{2} p^{n_e(\omega)} (1-p)^{n_o(\omega)}$$

$$= \frac{1}{2} p^{n_b - n_o(\omega)} (1-p)^{n_o(\omega)}$$

$$= \frac{1}{Z} \left(\frac{p}{1-p}\right)^{-n_o(\omega)}$$

with Z a constant. But now we see that our Gibbs measure and our Markov chain measure are exactly the same measures. We need only choose b so that $e^b = \frac{p}{1-p}$. Thus, since the law of large numbers and the central limit theorem are valid for these Markov

chains, Ising could not expect a double-humped curve, and presumably he had something similar in mind when he said that there would be a single maximum. The same kind of comparison with Markov chains, and hence the same conclusions, can be made if there is an exterior field.

We shall later see that the double-humped distribution that Ising desired can occur in two dimensions. First we shall illustrate the difference between one and two dimensions in terms of a simpler property of Markov chains. Consider the measure defined on a finite number of spins $\sigma_0, \sigma_1, \ldots, \sigma_N$ in one dimension. Since this measure may be viewed as a measure determined by a two state Markov chain, it follows from Markov chain theory that

$$\lim_{N\to\infty} P(\sigma_N = 1 | \sigma_0 = 1) = \lim_{N\to\infty} P(\sigma_N = 1 | \sigma_0 = -1) .$$

In terms of the magnet, this may be interpreted as saying that the value of a spin has little effect on the value of spins far away. This basic property of Markov chains need not hold for Markov random fields in two dimensions. We shall now use the argument of Griffiths and Peierls to make this more precise.

We begin by considering a lattice in which we fix all the points on the outside to have a "+" orientation. These points will be called boundary points. The remaining points, or interior points, interact and change just as in Ising's formulation. A typical configuration is shown in Figure 5.

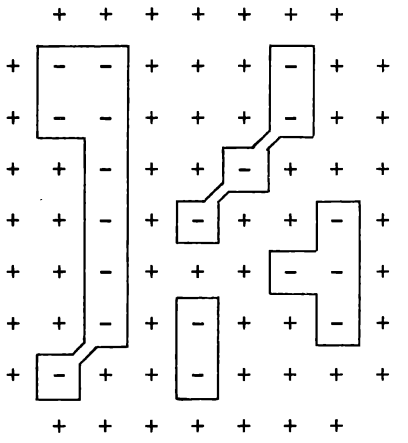


FIGURE 5

We have drawn lines separating spins of opposite signs, or odd bonds. We assign our probability measure by (3), and obtain

$$(7) \qquad P(\omega) = Z^{-1} e^{\frac{J}{kT} \sum_{i,j} \sigma_i(\omega)\sigma_j(\omega) + \frac{mH}{kT} \sum_i \sigma_i(\omega)} .$$

As we did in one dimension, we can rewrite this in terms of the number of odd bonds $n_o(\omega)$ and the magnetization $M_n(\omega)$, which is equal to the difference between the number of +'s and -'s in ω, as follows:

$$(8) \qquad P(\omega) = \frac{e^{-bn_o(\omega) + hM_n(\omega)}}{Z}$$

where $b = \frac{2J}{kT}$ and $h = \frac{mH}{kT}$. For zero external field, $h = 0$, the energy depends only on the number of odd bonds. But we notice that the number of odd bonds $n_o(\omega)$ is just the total length of all the lines in Figure 5. This was the essential device introduced by Peierls to study P.

In Figure 5 we fixed all boundary points to be "+" spins. We could equally well have fixed all the boundary points to have "-" spins. In either case we can ask for probabilities relating to the center points for very large lattices. We shall show that

if the temperature is low enough (b large enough), the probabilities near the center obtained for the two different boundaries differ by a positive amount, no matter how large the square is. This means that, in this case, the points far from O, the boundary points, have an influence on the center no matter how far away they are.

We shall look now at $P(\sigma_0 = -1)$ when we put all + spins on the boundary. Here σ_0 is a spin in the center of the box. For a configuration with $\sigma_0 = -1$, when we draw our borders as in Figure 5, this configuration will have a closed curve c that includes O. Let us call this curve a circuit and denote its length by L. All points inside this circuit have negative spins. Let $\bar{\omega}$ denote an arbitrary configuration which has the same circuit c around O. With such an $\bar{\omega}$ we associate a new configuration ω' which agrees with $\bar{\omega}$ except that all the - spins inside the circuit are changed to +. This has the effect of removing the circuit and decreases the number of odd bonds by L. That is,

$$n_o(\omega') = n_o(\bar{\omega}) - L.$$

Then the probability that the circuit c around O occurs is

$$(9)\quad P(\text{circuit } c) = \frac{\sum_{\bar{\omega}} e^{-bn_o(\bar{\omega})}}{\sum_{\omega} e^{-bn_o(\omega)}} .$$

If we limit the number of terms in the denominator of (9) we can only increase the fraction. Thus,

$$P(\text{circuit } c) \leq \frac{\sum_{\omega} e^{-bn_0(\bar{\omega})}}{\sum_{\omega'} e^{-bn_0(\omega')}}$$

$$= \frac{\sum_{\omega} e^{-bn_0(\bar{\omega})}}{\sum_{\omega} e^{-bn_0(\bar{\omega})} + bL}$$

$$= e^{-Lb}.$$

Since every configuration ω with $\sigma_0(\omega) = -1$ contains a circuit c,

$$P(\sigma_0(\omega) = -1) \leq \sum_{L=4,6,..} r(L)e^{-Lb}$$

where $r(L)$ is the number of configurations $\bar{\omega}$ with circuits of length L. Consider such a circuit. This circuit must have each point a distance no more than $\frac{L}{2}$ from the point O. There are at most L^2 ways to choose a starting point for this circuit. Having chosen a starting point there are at most 4 ways to choose the next point and 3 ways to choose each succeeding point moving around the circuit. Thus there are at most $4L^2 3^L$ circuits of length L containing O. That is

$$P(\sigma_0 = -1) \leq \sum_{L=4,6,..} 4L^2 3^L e^{-bL} \tag{10}$$

$$= \sum_{L=4,6,..} 4L^2 (3e^{-b})^L .$$

But since, for $|x| < 1$, it is true that $\sum_L L^2 x^L < \infty$, we can certainly make the right side of (10) less than 1/3 by choosing $b=b_0$ large enough. Of course by symmetry if we had started with all negative spins on the boundary points we would have been led to $P(\sigma_0 = 1) < 1/3$, and therefore $P(\sigma_0 = -1) > 2/3$. Thus the effect of the boundary is felt at the center no matter how far away it is, our first indication that

the model in two dimensions is really different from the model in one dimension, contrary to Ising's claim.

The very simple argument of Griffiths and Peierls showing the effect of the boundary has been very much improved in recent years. In fact it has been shown that in two dimensions there is a value b_c called the critical value such that, for $0 \leq b \leq b_c$, the boundary loses its effect and, for $b > b_c$ it does not. The critical value is about .88. The corresponding value for temperature is called the critical temperature.

Our limit laws for Markov chains are theorems that relate to repeating the experiment indefinitely. From just the knowledge of the initial probability vector and the transition matrix we can uniquely determine a probability measure on the infinite sequence of possible outcomes

$$\omega = (+,-,+,\ldots)$$

appropriate for continuing the process indefinitely. This measure is determined by the probabilities of the cylinder sets. These cylinder sets are sets consisting of points ω with the first n outcomes specified.

In two dimensions we would want, in a similar way, to determine a measure on the space of all infinite configurations ω. Such a configuration is now an assignment to each lattice point of a + or a -. We would expect this measure to be determined by the probabilities of cylinder sets where a cylinder set is a set of configurations with prescribed spins on a finite box. However, unlike the case of Markov chains, it is no longer clear how to consistently assign probabilities to boxes. With a Markov chain, time served as a direction and we could start at the initial point and build up the probability by our initial

probability vector and P. We can assign a probability measure by assigning a Gibbs measure with parameters b and h to boxes if we specify the values on the boundary of the box. We might expect to obtain a probability measure on the infinite configurations by taking limits of these measures.

Assume for example that we take a sequence of boxes with + spins on the boundary. Then for any point a in the box

$$P(\sigma_a = + | \sigma_k, k \neq a) = P(\sigma_a = + | \sigma_k, k \text{ in } N_a).$$

If the point a does not have any neighbors which are boundary points, then the right side of this equation will not depend upon the values fixed on the boundary. Thus, if we choose a sequence of such measures, they will all have the same local characteristics at the point a, and we can expect the limiting measure to have these same local characteristics. These characteristics are the same at each point and are determined by b and h. We now consider such a sequence. We show the first two boxes in Figure 6.

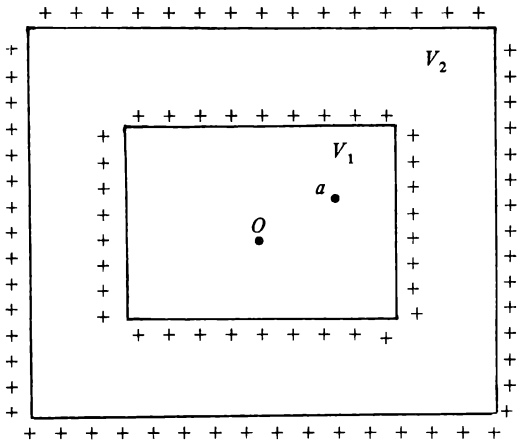


FIGURE 6

Since we are in the attractive case, the positive values on

the boundary will tend to make the spin at the center positive. The influence of the boundary of the box V_2 must be passed through the points of the boundary of V_1. Thus since V_1 has all boundary points +, it is quite intuitive that the probability of a + spin at 0 under the measure determined by V_2 is less than that under the measure determined by V_1. More generally, as $V_1, V_2, \ldots$ increase in size, the probabilities $P^{(n)}(\sigma_0=1)$ for V_n decrease to a limiting value $P_+(\sigma_0=1)$. This will be proved in Appendix 1. The same argument applies to $P^{(n)}(\sigma_a=1,\sigma_b=1)$. In this way we obtain a limiting probability for any events of the form

$$P_+(\sigma_a=1,\sigma_b=1,\ldots,\sigma_s=1)$$

These probabilities determine all probabilities relating to the state of a finite set of points. For example,

$$P_+(\sigma_a=1,\sigma_b=-1) = P_+(\sigma_a=1)-P_+(\sigma_a=1,\sigma_b=1).$$

Other probabilities are determined in a similar fashion.

Thus by the limiting procedure using only finite Gibbs measures we can construct probabilities of the form

$$P_+(\sigma_a=1,\sigma_b=1,\ldots,\sigma_s=1)$$

for any finite number of lattice points a,b,...,s. These probabilities in turn determine a unique measure on the set of all infinite configurations, and this measure will be shown to be stationary in Appendix 1. As we have observed, it will have local characteristics depending only on the energy parameters.

All of the previous discussion applies equally if we put -1's on the boundaries of the finite boxes. Again we get the same local characteristics, but we have seen that if the

temperature is low enough we will get a different $P_-(\sigma_0=1)$ for the probability that $\sigma_0=1$. Thus we will have, for sufficiently low temperatures, two measures which will both have the same local characteristics. The two measures obtained by + and - boundary conditions are called pure phases. If these two measures are different the process will be said to have a phase transition. It has been proven that for $b > 0$ and $h \neq 0$ these two measures are the same and represent the only possible measure with given local characteristics determined by b and h. Further for $h = 0$ and $0 \leq b \leq b_c$, where b_c is the critical value, there is also only one such measure. For $b > b_c$, $P_+ \neq P_-$, i.e., there are two pure phases. Finally, any measure with local characteristics determined by these parameters is a convex combination of these two pure phases. That is, if P is any measure with local characteristics the same as P_+ and P_-, there is a number t, where $0 \leq t \leq 1$, such that probabilities relating to P can be obtained from P_+ and P_- by

$$\begin{aligned} &P(\sigma_a=-1,\sigma_b=1,\ldots,\sigma_u=-1,\sigma_v=1) \\ &= tP_+(\sigma_a=-1,\sigma_b=1,\ldots,\sigma_u=-1,\sigma_v=1) \\ &+(1-t)P_-(\sigma_a=-1,\sigma_b=1,\ldots,\sigma_u=-1,\sigma_v=1). \end{aligned}$$

The fact that P_+ and P_- determine all Gibbs measures has only recently been proven by Aizenman (1979) and Higuchi (1979). This result does not hold in three dimensions.

In addition to the positive and negative boundary we shall discuss the free boundary and the periodic boundary. For the free boundary we treat the boundary points the same as any other point. They will of course have fewer neighbors. The periodic boundary will be defined in Section 3. For these boundaries the finite Gibbs measures on the boxes V_n approach a measure P

with $P = .5P_+ + .5P_-$. If the temperature is low enough so that $P_+ \neq P_-$, we can see for these boundaries that the behaviour Ising sought should occur. If we take a large area, the density of + spins will, with probability 1/2, be as predicted by P_+, i.e., greater than 1/2. With probability 1/2 the density will be as predicted by P_-, i.e., less than 1/2. It is this fact which suggests a double-humped distribution for the magnetization M_n when the magnet is a large two-dimensional set. Thus, below the critical temperature, it is reasonable to expect the system to have a majority of one type of spin, i.e., to be a magnet.

So far we have assumed that there is no exterior field. As we have remarked, for the case of exterior field $h \neq 0$, it will always be the case that $P_+ = P_-$ so that there will be only one measure P_h. However, there is another version of spontaneous magnetization which may be described as follows: for values of $b > b_c$ as h tends to 0 through positive values, the measures P_h approach the measure P_+. As h approaches 0 through negative values these measures approach the different measure P_-. This is often described by saying that if we impose a very small field below the critical temperature, we can make a positive magnet no matter how small the field is. We cannot do this above the critical temperature.

The measures P_+ and P_- have an important property called stationarity. This means that if we compute the probability for a specific configuration, this probability will be the same as for the same configuration shifted by a fixed amount. This property holds also for our Markov measure in one dimension when the chain is started in equilibrium. It might be expected that this should always be the case, but if $b < 0$, so that we are in the repulsive case, this need not be true, as we will now show.

A pair of lattice coordinates (i,j) is called even if i+j is an even number. It is called odd if i+j is odd. For a finite lattice we shall associate with each configuration ω a new configuration $\bar{\omega}$ as follows. To obtain $\bar{\omega}$ from ω we simply reverse the spins at the odd points. In Figure 7 we show the result of starting with a 5 x 5 configuration with positive boundary and reversing the spins at the odd points.

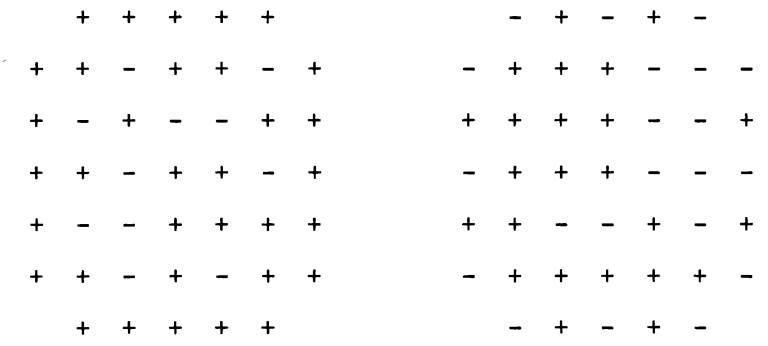


FIGURE 7

The important property of this transformation is that it changes even bonds into odd bonds and odd bonds into even bonds. Thus if $n_o(\omega)$ is the number of odd bonds in ω out of a total of $\bar{n}(\omega)$ bonds, then

$$n_o(\bar{\omega}) = \bar{n}(\omega) - n_o(\omega).$$

Now every $\bar{\omega}$ corresponds to exactly one ω. Thus if we put a probability measure on the space of $\bar{\omega}$'s we obtain a measure on the space Ω of all configurations. Assume that $h = 0$ and $\bar{b} < 0$. We define a Gibbs measure Q by

$$\begin{aligned} Q(\bar{\omega}) &= \frac{1}{Z_o} e^{-\bar{b}n_o(\bar{\omega})} \\ &= \frac{1}{Z_o} e^{-\bar{b}(\bar{n}(\omega)-n_o(\omega))} \\ &= \frac{1}{Z} e^{\bar{b}n_o(\omega)} . \end{aligned}$$

In the last step we have absorbed the constant term involving $\bar{n}(\omega)$ into the Z. We then put $b = -\bar{b}$ and define P by

$$(11) \quad P(\omega) = \frac{1}{Z} e^{-bn_o(\omega)} .$$

This means that $P(\omega)$ is a Gibbs measure with zero external field and an attractive interaction. We recall that $\bar{b}$ was negative so b is positive. We shall now use our information about P as an attractive measure to obtain information about Q. We know that for $b > b_c$ as the lattice size (n x n) increases, $P_+^{(n)}$ approaches P_+ and $P_-^{(n)}$ approaches P_-, where P_+ and P_- are measures on the space of all configurations and are not the same measures even though they have the same local characteristics. Let us denote by $Q_+^{(n)}$ the measure corresponding to $P_+^{(n)}$ under the transformation and $Q_-^{(n)}$ the measure obtained from $P_-^{(n)}$. Then $Q_+^{(n)}$ approaches a limiting measure Q_+ and $Q_-^{(n)}$ approaches a limiting measure Q_-, with Q_+ and Q_- two different measures. Also we know that

$$P_-(\omega_a=+) < P_+(\omega_a=+),$$

where ω_a represents the value given to the point a by the configuration ω. If a is an even site then

$$Q_+(\omega_a=+) = P_+(\omega_a=+)$$

but if a is an odd site then

$$Q_+(\omega_a=+) = P_-(\omega_a=+).$$

Thus if we define $\bar{b}_c = -b_c$, we see that for $\bar{b} < \bar{b}_c$ the measure Q_+ is not a stationary measure. By symmetry the same is true for the measure Q_-. For $0 \geq \bar{b} \geq \bar{b}_c$ the measures Q_+ and Q_- are the same measures since P_+ and P_- are the same for $0 \leq b$

$\leq b_c$.

If we choose $\bar{b} < \bar{b}_c$ and leave the boundary free, then from the fact that for $b > b_c$

$$P = \frac{1}{2} P_+ + \frac{1}{2} P_- ,$$

it follows that the distributions $Q^{(n)}$ will converge to

$$Q = \frac{1}{2} Q_+ + \frac{1}{2} Q_- .$$

By symmetry it is clear that Q is now a stationary measure. Recall that in the attractive case with free boundary and zero external field, the limiting measure P was unstable in the following sense: if we impose an external field h and let h go to zero through positive values, the measures P_h converge to P_+, but if we let h go to zero through negative values these measures converge to P_-. This characterizes spontaneous magnetization. The corresponding fact for the repulsive case becomes the following: if we impose a very small positive field on the even points and a very small negative field on the odd points the resulting measure will be near Q_+. If we impose a very small positive field on the odd points and an equal negative field on the even points the measure will be close to Q_-.

The results stated here for the two-dimensional Ising model represent the research of a large number of physicists and mathematicians. The value b_c or the critical temperature $T_c = 1/b_c$ was established by the pioneering work of Onsager (1944),(1949) and Yang (1952). They discussed phase transition in more analytic terms. Onsager considered the case where the horizontal interaction parameter J_1 and the vertical interaction parameter J_2 may be different. For any fixed J_1 and J_2 he showed that, below a critical temperature T_c which

depends on J_1 and J_2, phase transition occurs. If J_1 is held fixed and J_2 is allowed to tend to oo, T_c approaches O. Thus, in a sense, Ising was correct about this limiting case but it was not the place to look for a phase transition. The description in terms of the measures P_+ and P_- represent more recent work inspired by the approach of Dobrushin (1968a),(1968b),(1968c), and Lanford and Ruelle (1969). A proof that $P_+ \neq P_-$ for the case $J_1 \neq J_2$ for sufficiently high temperature may be found for example in Malysev (1979). We shall give a more detailed historical discussion in Section 3.

2. MARKOV FIELDS ON GRAPHS

So far we have dealt mostly with lattice points in the plane. Following Preston (1974) we shall now generalize this and carry out our discussion in the context of a graph $G = (T,E)$ where T is the finite set of vertices of the graph, and E is the set of edges of the graph. The following is a picture of a typical graph.

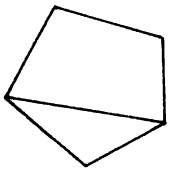


FIGURE 1

Two points will be called <u>neighbors</u> if there is an edge connecting them. The set of all points which are neighbors of a point t will be denoted by N_t.

We will assign to each point in the graph one of a finite set S of labels. Such an assignment will be called a <u>configuration</u>. We shall assign probability measures to the set Ω of all possible configurations ω. We denote by ω_t the value given to the point t by the configuration ω. We use ω_A to represent the configuration ω restricted to the subset A of T. We can think of ω_A as a configuration on the smaller graph restricting T to points of A. The <u>local characteristics</u> of a probability measure P defined on Ω are the conditional probabilities of the form

$$P(\omega_t | \omega_{T-t})$$

that is, the probability that the point t is assigned the value

ω_t, given the values at all other points of the graph. As in earlier sections, a probability measure will be said to define a Markov random field if the local characteristics depend only on the knowledge of the outcomes at neighboring points, i.e., if for every ω

$$P(\omega_t|\omega_{T-t}) = P(\omega_t|\omega_{N_t}).$$

This class of processes includes Markov processes but is also much more general as it applies to an arbitrary graph.

In previous sections we have discussed the relation between Gibbs measures and entropy. We next relate the Gibbs measure to the Markov property. Recall that a Markov process has the Markov property described as follows: the probability of an outcome at time n+1 given all previous outcomes depends only on the outcome at time n. These latter probabilities are called transition probabilities. To determine a specific Markov process we need a starting distribution and these transition probabilities. The Markov property stated above lacks symmetry in time, since it refers to the probability of of future outcomes given the past. However, a Markov process has a more symmetric property. This may be stated as follows: the probability that a particular outcome occurs at time n given all previous and all future outcomes depends only on the outcomes at times n-1 and n+1. The advantage of the more symmetric form for the Markov property is that it generalizes immediately to any graph. This generalization is the notion of Markov random field defined above.

We shall now examine the sense in which every Markov random field is a Gibbs distribution. We note that in the lattice example, $U(\omega)$ was the sum of energies which depend upon the

states of single points or of two points. Given an arbitrary graph we shall say that a set of points C is a clique if every pair of points in C are neighbors. We include the empty set as a clique. In the lattices we have considered, only the empty set, single points, and pairs of points are cliques. We could of course construct a lattice of triangles as in Figure 2.

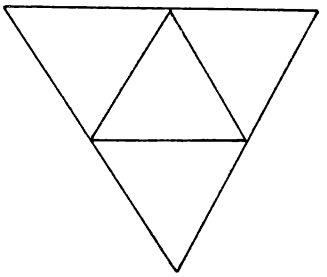


FIGURE 2

In this case, the empty set, single points, pairs of neighboring points, and triangles are cliques.

We now generalize the notion of energy. A potential V will be a way to assign a number $V_A(\omega)$ to every subconfiguration ω_A of a configuration ω. Given a potential we shall say that it defines an energy $U(\omega)$ on the set of all configurations ω by

$$U(\omega) = -\sum_A V_A(\omega)$$

where for fixed ω the sum is taken over all subsets A of T including the empty set.

The Gibbs measure induced by U is defined by

$$(1) \qquad P(\omega) = \frac{e^{-U(\omega)}}{Z} .$$

where

$$Z = \sum_\omega e^{-U(\omega)}$$

is called the partition function.

A potential V is called a nearest neighbor Gibbs potential if $V_A(\omega) = 0$ whenever A is not a clique. When we refer to a Gibbs measure we shall mean a measure induced by a nearest neighbor Gibbs potential. More general Gibbs potentials are studied but we shall not need these.

Note that this definition agrees with our definition of a nearest neighbor energy for a lattice. Points and pairs of adjacent points contribute to the energy, and these are the only cliques other than the empty set. If we had considered a lattice with triangles we would have had to allow an energy contribution for triangles in order to have the most general nearest neighbor potential. The inclusion of all cliques in assigning energies for Gibbs measures is necessary to establish the equivalence between Gibbs measures and Markov random fields. This result was proved first for lattices by Averintsev (1970) and extended to graphs by Preston (1974), Grimmett (1973), and Griffeath (1973).

We now proceed to prove that a nearest neighbor Gibbs measure determines a Markov random field.

Let $P(\omega)$ be a probability measure determined on Ω by a nearest neighbor Gibbs potential V, that is

$$(2) \qquad P(\omega) = \frac{1}{Z} e^{\sum_C V_C(\omega)}$$

where the sum is taken over all cliques on the graph G. Then

$$P(\omega_t | \omega_{T-t})$$

$$= \frac{P(\omega)}{\sum_{\omega'} P(\omega')}$$

where ω' is any configuration which agrees with ω at all

points except possibly t.

Thus

$$(3) \quad P(\omega_t | \omega_{T-t}) = \frac{e^{\sum_C V_C(\omega)}}{\sum_{\omega'} e^{\sum_C V_C(\omega')}} .$$

For any clique C that does not contain t, $V_C(\omega) = V_C(\omega')$. Thus terms that correspond to cliques that do not contain t cancel from both the numerator and denominator of (3) and therefore this probability depends only on the values at t and its neighbors. Thus P is a Markov random field.

The proof that a Markov random field determines a nearest neighbor Gibbs measure is much more involved and may be found in Preston (1974), Grimmett (1973) or in the contribution of Griffeath to Kemeny, Snell, and Knapp (1976). We note that there is not a unique potential function. However, there is a unique canonical potential which is singled out in the following manner: the states are renumbered $0,1,2,\ldots,r$ with 0 playing the role of a preferred state. The potential is then said to be a canonical potential if $V_C(\omega) = 0$ when ω assigns the value 0 to at least one site in C. It is then proved that there is a unique canonical potential for a given Markov random field. We shall now show how this canonical potential is described. Denote by ω^A the configuration which agrees with ω on A but assigns the value 0 at all sites outside of A. For the empty set $\emptyset$ we define $V_\emptyset(\omega) = 0$. For $A \neq \emptyset$ we define

$$(4) \quad V_A(\omega) = \sum_{B \subset A} (-1)^{|A-B|} \log P(\omega^B)$$

where $|A-B|$ is the number of elements in $A-B$. The potential $V_A(\omega)$ can be obtained also from the formula

$$(4') \quad V_A(\omega) = \sum_{B \subset A} (-1)^{|A-B|} \ln P(\omega_a^B \mid \omega_x^B \; x \in N_a)$$

where a is any point in A. Then the energy $U(\omega)$ is given by $U(\omega) = -\sum_C V_C(\omega)$. Using this energy

$$(5) \quad P(\omega) = \frac{e^{-U(\omega)}}{Z} .$$

Formula 4' shows that the measure $P(\omega)$ is determined by its local characteristics.

Let us now return for a moment to the Markov chain example discussed in Section 1. We start this with the equilibrium vector (1/2,1/2). Using (4) we shall actually construct the Gibbs potential for this Markov chain. Before we do this, let us note that it is often convenient to look at such a chain in two different ways. The first is where the states are 0 and 1 (as in coin tossing where we are looking at the number of heads). In this case the probability of a configuration is given by (using s_j to represent the outcome at time j)

$$P(\omega) = \frac{1}{Z} e^{v_1 \sum_{i,j} s_i s_j + v_0 \sum_{0<i<n} s_i + \bar{v}_0 \sum_{i=0,n} s_i} .$$

Here the first term in the exponent represents all cliques containing two points and the other terms represent the one-point cliques. On the other hand we may want to look at the states as +1 and -1 (spins) instead, in which case we will use σ_j for the value at point j instead of s_j. Then the probability of a configuration ω becomes

$$P(\omega) = \frac{1}{Z} e^{b_0 \sum_{i,j} \sigma_i \sigma_j + h \sum_{0<i<n} \sigma_i + \bar{h} \sum_{i=0,n} \sigma_i} .$$

The transformation

$$(6)\quad \sigma_i = 2s_i - 1$$

$$s_i = \frac{\sigma_i + 1}{2}$$

when substituted into the above formulas makes it possible to go between the two formulations easily enough. Using (6) we can compute that in terms of v_o, $\bar{v}_o$ and v_1,

$$(7)\quad b_o = \frac{v_1}{4},\ h = \frac{v_1+v_o}{2},\ \bar{h} = \frac{2\bar{v}_o+v_1}{4}$$

and in terms of b_o, h, and $\bar{h}$,

$$v_1 = 4b_o,\ v_o = 2h - 4b_o, \bar{v}_o = 2\bar{h} - 2b_o.$$

The following observation will be important for later work. Assume that we use 1 and -1 for the states and assign a measure using b_o, h and $\bar{h}$. Then h = 0 corresponds to the important special case of no exterior field. If we convert our measure to the measure using v_o, v_1 and $\bar{v}_o$, then in one dimension h = 0 corresponds to $v_o = -v_1$.

We now compute the canonical potential V for our Markov chain example.

We write the transition matrix in the form

$$P = \begin{pmatrix} s & 1-s \\ 1-t & t \end{pmatrix}$$

Let $w = (w_0, w_1)$ be the fixed vector for P, i.e., wp = w. Then

$$w_0 = \frac{1-t}{(1-s)+(1-t)}$$

$$w_1 = \frac{1-s}{(1-s)+(1-t)}$$

Let $\underline{0} = (000\ldots0)$ be the configuration with 0 at all sites. By (5)

$$Z = -\log P(\underline{0})$$

$$= -\log w_0 s^n .$$

Next,

$$P(1000\ldots000) = \frac{e^{\bar{v}_0}}{Z} .$$

Thus

$$\bar{v}_0 = \log P(100\ldots000) + \log Z$$

$$= \log w_1 (1-t) s^{n-1} - \log w_0 s^n$$

$$= \log \frac{(1-s)}{2-s-t}(1-t)s^{n-1} - \log \frac{(1-t)s^n}{2-s-t}$$

$$= \log \frac{1-s}{s}.$$

Continuing in this manner we can compute

$$(8)\quad v_0 = \log \frac{(1-s)(1-t)}{s^2}$$

and

$$(9)\quad v_1 = \log \frac{st}{(1-s)(1-t)} .$$

We see that it is easy to determine a Markov field if we have a nearest neighbor potential. If we have a Markov random field on a finite graph, then the local characteristics do uniquely determine this measure. In fact the canonical potential can be determined from these local characteristics by the formula (4'), and the canonical potential determines the measure. However, we cannot just choose any set of local characteristics

and be sure that there is a measure consistent with these characteristics. Certain consistency conditions have to be satisfied and these might be difficult to check.

The difficulty of choosing appropriate local characteristics would suggest that the concept of the potential function may be the key in studying Markov fields. We have seen that we can assign measures with the Markov field property to configurations on infinite graphs. In the Ising model we obtained these as limits of finite measures. We can define such Markov measures directly. The natural definition would require that

$$P(\omega_a=x|\omega_s,\ s\neq a) = P(\omega_a=x|\omega_s,\ s\in N_a)\ .$$

The first of these conditional probabilities is technically difficult to define. However, we can use the equivalent definition

$$P(\omega_a=x|\omega_s,\ s\in A) = P(\omega_a=x|\omega_s,\ s\in N_a)$$

for any finite set A which contains N_a and not the point a. As the Ising model shows, on infinite graphs there may be more than one measure with the same local characteristics. When this happens we can conclude that probabilities relating to a fixed finite set will be affected by the knowledge of the outcome arbitrarily far from this set. This, in turn, tells us that the same would be true for large finite models which the infinite models approximate.

It is of course important to know when such a "phase transition" (more than one measure with the same local characteristics) will not occur.

A condition for the absence of a phase transition has been given by Dobrushin (1968). He considers the more general problem

of random fields without assuming the Markov property. We shall give his result for the Markov case.

Assume the possible states at each point of the graph are from a finite set S. Define a distance between any two probability measures p and $\bar{p}$ on S by

$$d(p,\bar{p}) = \max_{A \subset S} |p(A) - \bar{p}(A)| .$$

This distance is also equal to

$$d(p,\bar{p}) = \frac{1}{2} \sum_{x \in S} |p(x) - \bar{p}(x)| .$$

If we specify the values ω_{N_a} of ω on the neighbors N_a of a, then the local characteristics give a measure $q_{\omega_{N_a}}$ on S. Define

$$R_{a,u} = \max_{\omega_{N_a}, \bar{\omega}_{N_a}} d(q_{\omega_{N_a}}, q_{\bar{\omega}_{N_a}})$$

where ω_{N_a} and $\bar{\omega}_{N_a}$ differ only at the point u in N_a. Then Dobrushin proved that there is a unique Markov random field with the same local characteristics as q, if there is a number $w < 1$ such that

$$\sum_u R_{a,u} < w$$

for all a.

3. FINITE LATTICES.

As we have indicated, some of the really surprising aspects of the Ising model show up as limiting results of finite lattices. In this section we shall give some exact results for finite lattices and indicate the limiting results. We shall consider square lattices in two dimensions. It will be convenient to describe an n x n lattice as the set of points of the form (i,j) with $1 \leq i \leq n$ and $1 \leq j \leq n$. If we add a boundary, the boundary points will be the points (0,j), (n+1,j) $1 \leq j \leq n$ and (i,0), (i, n+1) for $1 \leq i \leq n$. For example, in Figure 1 we show a typical configuration for a 4 x 4 lattice with boundary points fixed at +.

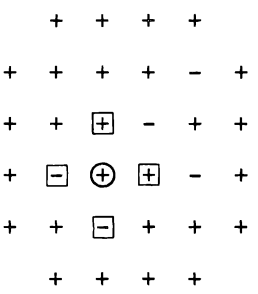


FIGURE 1

We have indicated a typical interior point and its neighbors.

One boundary condition that we shall examine is the free boundary. This is one of the boundaries that Ising studied in one and two dimensions. In this case we, in effect, have no boundary. The outer points of the square are simply assumed to have less than four neighbors. For example, again for the 4 x 4 case, a typical configuration with free boundary is shown in Figure 2.

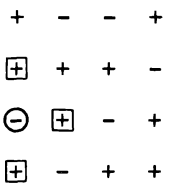


FIGURE 2

The point (1,2) has only three neighbors, and a corner point such as (4,4) has only two neighbors.

We shall now describe the other boundary considered by Ising, the periodic boundary. As with the free boundary, we do not fix any boundary points. Instead we make every point have four neighbors as follows: For an n x n lattice a point of the form (i,n), $1 < i < n$, has neighbors (i-1,n), (i,n-1), (i+1,n) and (i,1). Similarly a point of the form (n,j) with $1 < j < n$ has neighbors (n-1,j), (n,j-1), (n,j+1) and (1,j). A corner point, say (1,1), has neighbors (1,2), (2,1), (1,n) and (n,1). Thus an extreme row or column sees the opposite extreme row or column as neighbors. Figure 3 shows a 5 x 5 lattice.

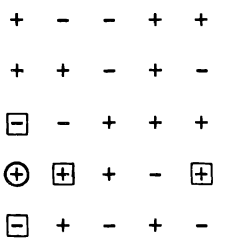


FIGURE 3

The points in squares are boundary points for the circled point.

The periodic boundary is important because, in a pioneering work, Onsager (1944) showed that in two dimensions using periodic boundary one can solve problems relating to finite lattices

analytically and then obtain limiting results from this. Properties of lattices with free boundary are very similar to those of lattices with periodic boundary, and in this section we shall find it convenient to use the free boundary instead of the periodic boundary. In later sections we shall prefer to use the periodic boundary.

We see from (7) of Section 1, taking $J = m = 1$, that if $b = \frac{1}{kT}$, the Gibbs measure with parameters b and H can be written in the form

$$P(\omega) = \frac{1}{Z_n} e^{b(\sum_{i,j} \sigma_i(\omega)\sigma_j(\omega) + HM_n(\omega))}$$

where the sum is over nearest neighbors. If we have fixed boundary values, we let i,j range over the boundary points, but $M_n(\omega)$ counts only the spins on the interior. The parameter b determines the strength of the interaction and H represents the strength of the exterior field. The partition function Z_n for the n x n lattice is defined by

$$Z_n = \sum_{\omega} e^{b(\sum_{i,j} \sigma_i(\omega)\sigma_j(\omega) + HM_n(\omega))} .$$

We consider first the finite lattice with a negative boundary, i.e., negative spins at all boundary points. Then M_n represents the magnetization on the interior of the lattice. If $b = 0$ (temperature infinity), then the spins are independent with each spin being + with probability 1/2 and - with probability 1/2. In this case, the classical limit laws apply to M_n. For example the law of large numbers states that the magnetization will, with probability one, approach 0. Letting $\sigma(M_n)$ denote the standard deviation of M_n, the central limit theorem states that the distribution of

$$\bar{M}_n = \frac{M_n}{\sigma(M_n)}$$

will approximate a bell-shaped normal curve for large enough n. This central limit theorem has been shown to hold for all temperatures above the critical temperature (Malysev 1975). As the temperature gets close to the critical temperature, however, the denominator goes to infinity, thus the normalization becomes quite severe. Later in this section, we will give physical interpretations of some of these quantities.

Assume now that $H = 0$ but $b > 0$. As we have stated earlier, if we fix the boundary points of a lattice to be all negative and let the size of the lattice tend to infinity we obtain a limiting measure, P_-, which is a "pure state" associated with the parameters b, 0. Similarly, if we put all + values on the boundary we obtain a possibly different "pure state" P_+ associated with the same parameters. We stated that $P_- = P_+$ if $b \leq b_c$, where $b_c \doteq .88$ is the critical value. If $b > b_c$, $P_- \neq P_+$ and we have two different pure phases. Each of these pure states acts very much like the measure corresponding to independent trials. For example there is a "law of large numbers" for P_+. This states that for the positive boundary

$$\frac{M_n}{n^2} \longrightarrow a^+(b),$$

as n tends to infinity. Here $a^+(b)$ is a constant depending only on b. Also for the measure P_-, corresponding to negative boundaries,

$$\frac{M_n}{n^2} \longrightarrow a^-(b),$$

as n tends to infinity, where $a^-(b)$ is a constant depending

on b. A proof of these assertions may be found in Georgii (1970).

The function $a^-(b)$ which originated in the work of Onsager (1949) and Yang (1952, is given by

$$(1) \quad a^-(b) = -\left(1 - \frac{1}{(\sinh b)^4}\right)^{1/8} \quad \text{if } b \geq b_c$$
$$= 0 \quad \text{if } b \leq b_c.$$

By symmetry the values of $a^+(b)$ are given by $-a^-(b)$.

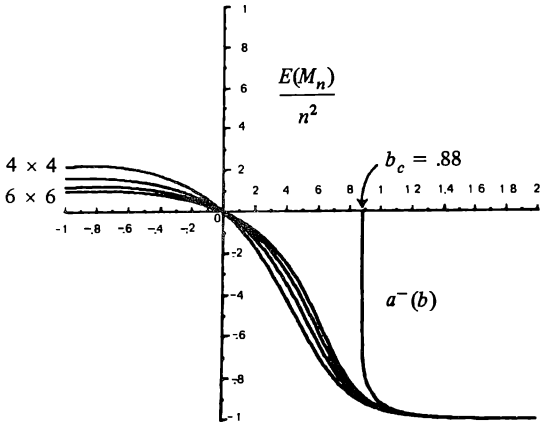


FIGURE 4

Expected magnetization per site as a function of b for lattice sizes $3 \times 3, 4 \times 4, 5 \times 5, 6 \times 6$ with negative boundary and zero external field.

In Figure 4 we have graphed the function (1) together with the expected value of $\frac{M_n}{n^2}$ with zero field and b between -1 and 2, for 3 x 3, 4 x 4, 5 x 5 and 6 x 6 lattices. We note that for negative values of b, the expected proportion of + spins is slightly larger than 0. This is caused by the repulsive effect of the boundary which has all - spins. For positive values of b we see that $E(\frac{M_n}{n^2})$ drops off quite fast as b increases. By the limiting results we know that the curves for these finite lattices should approach the limiting curve $a^-(b)$ as n tends to infinity. We see that they have a long way to

go! In physical applications it is reasonable to assume that the lattice size is of the order of $2^{20} \times 2^{20}$ so that it is probably true that by this time the approximation is reasonable.

For $H = 0$ and $b > b_c$, we have noted that the states P_+ and P_- are distinct. For these pure states, the random variables M_n satisfy the central limit theorem and law of large numbers. When $H \neq 0$ or $b < b_c$, and $P_+ = P_-$, the same theorems hold. A discussion of this theorem and further references may be found in Gallavotti and Jona-Lasinio (1975). In Figure 5 we have plotted the distribution of M_6 (the 6 x 6 case) for a sequence of b values from 0 to 1.

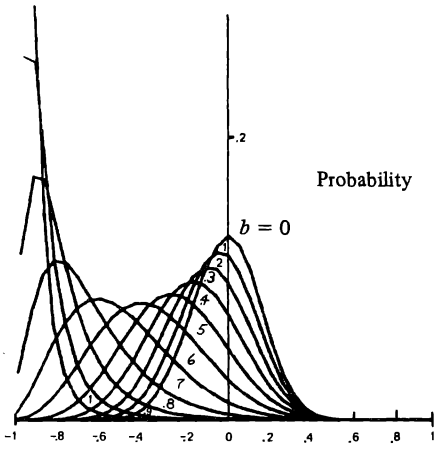


FIGURE 5

Distributions of the magnetization per site M_n/n^2 in the 6 × 6 lattice with negative boundary for b values between 0 and 1.

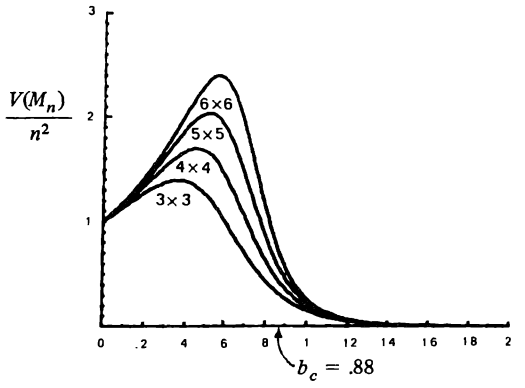


FIGURE 6

Graph of variance per site of the magnetization for free boundary, zero external field as a function of b for lattice sizes 3 × 3, 4 × 4, 5 × 5, 6 × 6.

We see that for $b = 0$ to .7 the curves are bell-shaped and then as b passes through .8 the distribution becomes quite concentrated at the configuration with all spins negative. The limit theorems for the pure states refer to $b \neq b_c$ being fixed and letting n tend to infinity. We see that even for the small lattice we have a bell-shaped distribution for $b < b_c$ (temperature above the critical temperature). Note that as b

decreases towards b_c, the distributions have increasing variances.

For the central limit theorem to hold, we would want the limit

$$\frac{V(M_n)}{n^2} \longrightarrow c(b)$$

to exist, where $V(M_n)$ denotes the variance of M_n. Then $c(b)$ would be used to normalize M_n. This limit exists, but $c(b)$ becomes infinite as b approaches b_c from either side. In Figure 6, we have plotted $\frac{V(M_n)}{n^2}$ for $H = 0$ as a function of b for $n=3,4,5,$ and 6. We see that the maximum values are increasing and the values of b for which they occur are moving toward b_c.

Let us return now to the case $H = 0$ but with a free boundary. This was the case that interested Ising. He was looking for a distribution for M_n which had two maximum values corresponding to the fact that the interactions alone should, for temperatures low enough, cause most of the spins to be in the same direction.

With $H = 0$, the basic probability measure can be written in the form

$$P(\omega) = \frac{1}{Z_n} e^{-bn_o(\omega)}$$

as in (8) of Section 1.

In this case, if we reverse all the spins for a configuration ω, we obtain a configuration $\bar{\omega}$ with the same number of odd bonds. This symmetry means that the distribution of M_n will have the property

$$P(M_n = i) = P(M_n = -i).$$

Recall that, for the attractive case with free boundary, the finite Gibbs measures converge to a limiting measure P as the size of the lattice increases indefinitely. If $b \leq b_c$ then $P_+ = P_-$, and P is equal to this common measure.

When $H \neq 0$ or when $H = 0$ and $b < b_c$, the magnetization M_n will again satisfy the law of large numbers and the central limit theorem. As for the negative boundary, the variance of M_n increases as b increases to b_c. In fact, for $H = 0$, above the critical temperature

$$\frac{\mathrm{Var}(M_n)}{n^2} \longrightarrow c(b)$$

independent of the boundary condition and $c(b) \longrightarrow \infty$ as b increases to b_c. Thus, above the critical temperature, we will have a single maximum for the magnetization, but large variation in the magnetization can be expected near the critical temperature. If $b > b_c$ then the measure P is given by

$$P = \frac{1}{2} P_+ + \frac{1}{2} P_-.$$

This may be interpreted as follows. For a large lattice, to compute $P(\omega)$ we can assume that we choose P_+ or P_- by the toss of a coin, and then if P_+ is obtained let $P(\omega) = P_+(\omega)$, and if P_- is obtained let $P(\omega) = P_-(\omega)$. For $b > b_c$ the law of large numbers will no longer be true. In fact we will have

$$\frac{M_n}{n^2} \longrightarrow W$$

where W is a random variable which takes on the value $a^+(b)$ with probability 1/2 and $a^-(b)$ with probability 1/2. Similarly the distribution of M_n/n^2 will approach a mixture of two distributions which are bell-shaped about the points $a^-(b)$ and

$a^+(b)$. Hence we can indeed expect, as Ising wished to have, a distribution with two maximum values for $b > b_c$. In Figure 7 we have plotted the distribution of $\frac{M_n}{n^2}$ for the 6 x 6 lattice with free boundary for values of b from 0 to 1.2. We see that as b passes through .8 the curves change from bell-shaped curves to curves with a double maximum, as predicted, even for this small value of n.

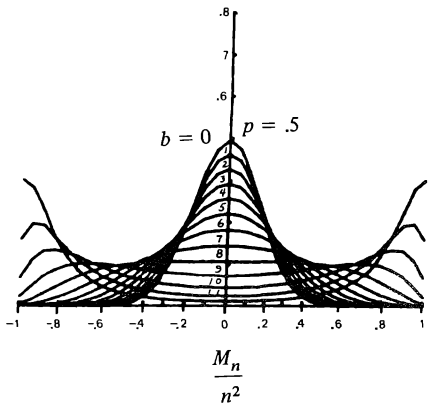


FIGURE 7

Distribution of the magnetization per site for M_n^2/n^2 a 6 × 6 lattice with free boundary, zero external field and attractive interaction.

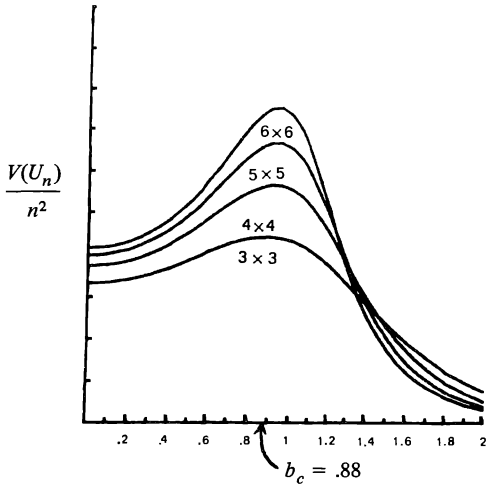


FIGURE 8

Variance per spin of the energy U_n as a function of b for 3 × 3, 4 × 4, 5 × 5 and 6 × 6 lattices with free boundary and no external field.

Of course for any finite lattice in either one or two dimensions we would expect that, as b tends to infinity, the distribution would have two maxima, one at each of the values corresponding to a configuration with all spins the same. The fundamental difference between one and two dimensions is that, for any fixed b, in one dimension the distribution will have a single maximum for large enough n, but in two dimensions the double maximum will persist for arbitrarily large boxes when $b > b_c$.

In the case of zero external field, the energy $U_n(\omega)$ is $-n_o(\omega)$. For b = 0, each bond will have an equal probability of

being even or odd and the bonds are independent. Hence the classical limit theorems would apply to U_n.

Assume now that $H = 0$ but $b > 0$. Then for $b < b_c$, the law of large numbers applies to $\frac{U_n}{n^2}$, and $\frac{U_n}{n^2} \longrightarrow U(b)$ where $U(b)$ is a constant depending only on b. For $b > b_c$, as with magnetization, the law of large numbers fails and $\frac{U_n}{n^2} \longrightarrow U$ where $U = U_+$ with probability 1/2 and U_- with probability 1/2. Here U_+ is the limit computed with the + boundary and U_- the limit computed with the - boundary. For $b < b_c$, the central limit theorem holds but we need to normalize with $\lim_{n\to\infty} \frac{V(U_n)}{n^2} = v(b)$. For $b \neq b_c$, this limit exists, but as was the case for the magnetization, $v(b)$ tends to infinity as $b \longrightarrow b_c$ from either side. In Figure 8 we have graphed $\frac{V(U_n)}{n^2}$ as a function of b for several lattice sizes. We note that even for these small lattice sizes, the maximum occurs near b_c.

The fact that $v_n(b) = \frac{V(U_n)}{n^2}$ approaches infinity as the lattice size approaches infinity suggests that, for $b = b_c$, a limit law for the distribution of this quantity would be of a different form than the central limit theorem. No limit theorem has been proven for $b = b_c$.

We turn now to the situation in which we have an exterior field. We consider only the attractive case. If $H \neq 0$ the measure P_+ and P_- are always the same measure P_H. Thus the presence of an exterior field is sufficient to cause the boundary influence to disappear as the size of the lattice increases indefinitely. This is in contrast to the case of zero field and $b > b_c$. The law of large numbers and the central limit theorem apply for the magnetization M_n using the limiting measure P_H.

To emphasize the limitation of looking only at finite lattices, we show in Figure 9 the distribution of $\frac{M_n}{n^2}$ for values

of b from 0 to 1.2 for h = .05. We note that, just as in the zero field case, we obtain two maxima for values of b sufficiently high. However, for any fixed b, as n increases, M_n normalized to have mean 0 and variance 1 will tend to a normal distribution. The expected value of the $\frac{M_n}{n^2}$ will tend to a positive value corresponding to the fact that the exterior field has caused positive magnetization.

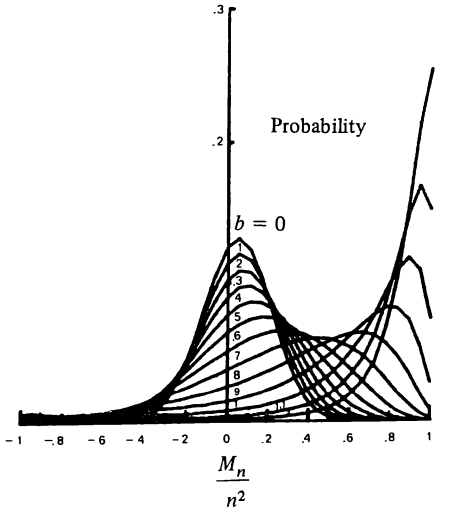


FIGURE 9

Distribution of the magnetization per site M_n/n^2 with a positive exterior field ($H = .05$) for the 6 × 6 lattice with free boundary.

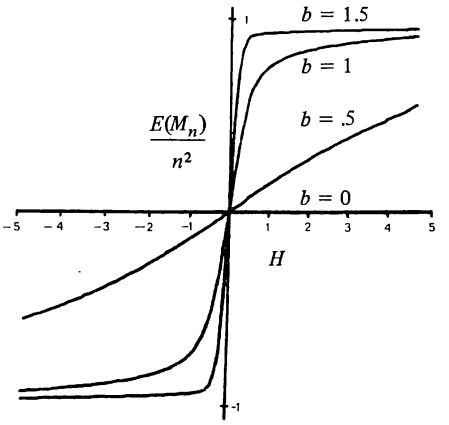


FIGURE 10

Expected value of magnetization per site as a function of exterior field H for $b = 0, .5, 1, 1.5$.

We now describe another very surprising limiting result. In Figure 10 we have graphed the expected value of $\frac{M_n}{n^2}$ in the 6 x 6 lattice for values of H ranging from -.5 to .5 for three values of b. We see that as b increases, the slope of the curves near H = 0 increases. Again, for fixed b and H, a limiting result has been proven by Martin-Löf (1973). This states that

$$E\left(\frac{M_n}{n^2}\right) \longrightarrow a(b,H)$$

where a(b,H) has a graph as indicated in Figure 11, if b is

fixed with $o \leq b \leq b_c$. However if $b > b_c$ the limiting curve has the form shown in Figure 12.

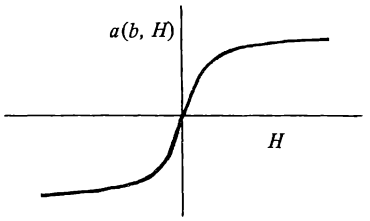


FIGURE 11

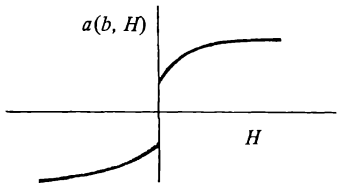


FIGURE 12

As indicated, if $b > b_c$ the limit, $a^-(b)$, of $E(\frac{M_n}{n^2})$ as H approaches O through negative field values is different than the limit $a^+(b)$ as H approaches O through positive values. The limit $a^-(b)$ has the values given by (1). This fact is interpreted as "spontaneous magnetization".

So far we have illustrated our basic descriptive quantities in terms of exact calculations. Basically, exact calculations are only possible for quite small lattices or for the periodic boundary conditions. To compute moments of our descriptive quantities for more general situations we need new methods. A technique that has been found quite useful in theoretical computations uses the function defined by

$$f_n(b,H) = -b^{-1} \ln Z_n .$$

The function $f_n(b,H)$ is called the free energy. We shall now show how the function $f_n(b,H)$ can be used to obtain the descriptive quantities that we have already discussed relating to the Gibbs measure. For example,

$$- \frac{d}{dH} f_n(b,H) = \frac{1}{bZ_n} \frac{dZ_n}{dH}$$

$$= \frac{1}{bZ_n} \sum_{\omega} bM_n(\omega) e^{b(\sum_{i,j} \sigma_i(\omega)\sigma_j(\omega) + HM_n(\omega))}$$

$$= E(M_n) .$$

Thus from the knowledge of $f_n(b,H)$ we can obtain the expected value of M_n. Similarly from the second derivative of f_n, we can obtain the variance of M_n. Specifically,

$$V(M_n) = - \frac{1}{b} \frac{d^2}{dH^2} f_n(b,H).$$

If we recall that the energy of the configuration ω is denoted by

$$U_n(\omega) = - \sum_{i,j} \sigma_i(\omega)\sigma_j(\omega) - HM_n(\omega) ,$$

then

$$E(U_n) = \frac{d}{db} (-\ln Z_n) = \frac{d}{db} (bf_n(b,H))$$

and

$$V(U_n) = \frac{d^2}{db^2} (-\ln Z_n) = - \frac{d^2}{db^2} (bf_n(b,H)) .$$

In the physical literature, <u>susceptibility</u> X_n is defined as the rate of change of the expected magnetization with respect to H. Thus

$$X_n = \frac{d}{dH} E(M_n)$$

$$= - \frac{d^2}{dH^2} f_n(b,H) = bV(M_n) .$$

Similarly, the variance of the energy is closely related to the thermodynamic quantity <u>specific heat</u>. Specific heat is defined as the rate of change of the energy with respect to T. For finite systems, we use the expected value of the energy. Thus, the specific heat C_n is

$$C_n = \frac{d}{dT} E(U_n) = \frac{d}{dT} \frac{d}{db} (bf_n(b,H))$$

$$= - kb^2 \frac{d^2}{db^2} (bf_n(b,H))$$

$$= - kb^2 V(U_n) .$$

Recall that $b = \frac{1}{kT}$. Next we note that these concepts tie up very nicely with the entropy as defined in Section 1. If $\bar{p}(\omega)$ is the Gibbs measure on the n x n lattice, then

$$\bar{p}(\omega) = \frac{1}{Z_n} e^{-bU_n(\omega)} .$$

Therefore

$$\ln \bar{p}(\omega) = -bU_n(\omega) - \ln Z_n$$

and

$$S(\bar{p}) = - \sum_{\omega} \bar{p}(\omega) \ln \bar{p}(\omega) = \sum_{\omega} \bar{p}(\omega) (bU_n(\omega) + \ln Z_n)$$

$$= bE(U_n) + \ln Z_n .$$

Thus $f_n(b,H) = E(U_n) - b^{-1}S(\bar{p})$. This shows that the entropy is closely related to the free energy. In fact, using the previous fact, it is not hard to show that given an energy function $U_n(\omega)$, the Gibbs measure $\bar{p}$ associated with this energy can be characterized as the measure which minimizes free energy where we define the free energy of the measure p by

$$f(p) = E_p(U_n) - b^{-1}S(p) .$$

Thus, by the previous formula, the Gibbs measure maximizes uncertainty as measured by entropy among all measures having the same expected value for the energy as $\bar{p}$.

Since in physical applications one is interested in very large systems, much of the attention in this literature centers

around the thermodynamic limit of these quantities defined as

$$f(b,H) = \lim_{n\to\infty} \frac{f_n(b,H)}{n^2}$$

$$M = \lim_{n\to\infty} \frac{E(M_n)}{n^2}$$

$$U = \lim_{n\to\infty} \frac{E(U_n)}{n^2}$$

$$X = \lim_{n\to\infty} \frac{X_n}{n^2} = b \lim_{n\to\infty} \frac{V(M_n)}{n^2}$$

$$C = \lim_{n\to\infty} \frac{C_n}{n^2} = kb^2 \lim_{n\to\infty} \frac{V(U_n)}{n^2} \quad .$$

These limits are functions of b and H though we have indicated this only for the case of the free energy. If $H \neq 0$, all these limits exist as finite limits and are related in the same manner as for the finite systems. Thus, $M = -\frac{d}{dH} f(b,H)$, $U = \frac{d}{db}(bf(b,H))$, $X = \frac{d}{dH} M = -\frac{d^2}{dH^2} f(b,H)$, and $C = \frac{d}{dT} U = -kb^2 \frac{d^2}{db^2}(bf(b,H))$.

We have seen that a "phase transition" occurs if $P_+ \neq P_-$. Historically a phase transition was first discussed in terms of the physical concepts of free energy, specific heat, susceptibility and spontaneous magnetization. Of course to the physicist, these are still the interesting quantities. A brief account of the history of these ideas will put our discussion in a better perspective.

In 1944 Onsager presented a calculation of the partition function for the case of a zero exterior field and periodic boundary which enabled him to obtain a closed form expression for the limiting value C of the specific heat per site $\frac{C_n}{n^2}$. This function has a singularity at $T = T_c$ or $b = b_c$. Its graph as a function of b has the form shown in Figure 13.

For finite lattices no such singularities occur, and indeed the the graphs of C_n/n^2 will look essentially the same as those for the variance of the number of odd bonds shown in Figure 6. Recall that these quantities are simply related. They have maximum values near b_c.

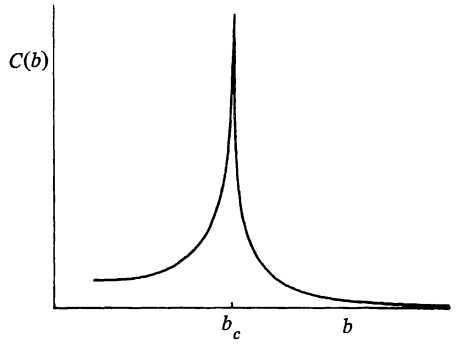


FIGURE 13

The spontaneous magnetization was defined by

$$M = \lim_{H\to 0+} \lim_{n\to\infty} E\left(\frac{M_n}{n^2}\right) = \lim_{H\to 0+} \lim_{n\to\infty} E(\sigma_0) \ .$$

A direct calculation of M would require calculation of the type carried out by Onsager for the case H>0. Neither Onsager nor anyone since has been able to carry out direct computation as Onsager did for $H \neq 0$. However, in a meeting at Cornell in 1948, Onsager wrote down his famous formula $M = \left(1 - \frac{1}{(\sinh b)^4}\right)^{1/8}$ for $b > b_c$. Onsager never published a derivation of this formula but it appeared in print in Onsager (1947). C. N. Yang (1952) gave a derivation of this formula using periodic type boundary conditions and letting H go to zero by setting $H = \frac{a}{n}$ for fixed a. Montroll, Potts, and Ward (1963) showed that the same formula for M can be obtained from zero field calculations alone. They show that for zero field and periodic boundary values the limits

$$E_p(\sigma_i) = \lim_{n\to\infty} E(\sigma_i)$$

and

$$E_p(\sigma_i\sigma_j) = \lim_{n\to\infty} E(\sigma_i\sigma_j)$$

exist. These values correspond to the infinite measure obtained by taking limits using the periodic boundary. They then showed that

$$\lim_{j\to\infty} E_p(\sigma_0\sigma_j) = (E_p(\sigma_0))^2 = M^2$$

where M is again given by Onsager's formula. They suggest that this must have been the way Onsager found the formula. It is not obvious that these two calculations should give the same result but we can now see this as follows. If for H=0 we calculate
$E_+(\sigma_i\sigma_j) = \lim_{n\to\infty} E(\sigma_i\sigma_j)$ using + boundary values and
$E_-(\sigma_i\sigma_j) = \lim_{n\to\infty} E(\sigma_i\sigma_j)$ using - boundary values, we would get the same result by symmetry. Since $P_p = \frac{1}{2}P_+ + \frac{1}{2}P_-$ we see that $E_p(\sigma_i\sigma_j)$ also has this common value. Now since P_+ is extreme for all b and H

$$\lim_{j\to\infty} E_+(\sigma_0\sigma_j) - E_+(\sigma_0)E_+(\sigma_j) = 0 .$$

Thus if we let H->0+ we should obtain, at 0 field, $\lim_{j\to\infty} E(\sigma_0\sigma_j) = M^2$ with $M = \lim_{H\to 0+} E_+(\sigma_0)$. The appropriate limits with + boundary conditions were carried out by Abraham and Martin Löf (1973) and they showed that the two methods do give the same results. Their proofs also show that T_c is characterized by the fact that $P_+ = P_-$ for $t > T_c$ and $P_+ \neq P_-$

for $t < T_c$.

The presence of the critical temperature shows up also in the concept of susceptibility. Recall that

$$X = b \lim_{n\to\infty} \frac{V(M_n)}{n^2}$$

$$= b \lim_{n\to\infty} \sum_{i,j} \frac{E(\sigma_i\sigma_j) - E(\sigma_i)E(\sigma_j)}{n^2}$$

$$= b \sum_j (E_p(\sigma_0\sigma_j) - E_p(\sigma_0)^2)$$

Thus the behavior of X at H=0 can be determined if we know the behavior of the correlation functions $E_p(\sigma_0\sigma_j)$. These have the following asymptotic properties. For T fixed, $E(\sigma_0\sigma_R)$ is asymptotically equal to $c''R^{-1/2}\, e^{-R/c}$ as $R \longrightarrow\infty$ if $T > T_c$, equal to $R^{-1/4}$ if $T = T_c$, and equal to $M^2 + c' \frac{e^{-2R/c}}{R^2}$ if $T < T_c$, where c, c', and c" are constants depending only on T. These results are due to Wu (1966) and Cheng and Wu (1967). Additional analytic results may be found in McCoy and Wu (1973). These results can be used to show that X gets very large near the critical temperature as $n \longrightarrow\infty$. In fact X is infinite for $T = T_c$ and increases without bound to either side of that point. An intuitive way of looking at this phenomenon is the following. As the temperature gets close to the critical temperature, the correlations between spins (determined by the constants above) become large for spins arbitrarily far away and since the susceptibility is a sum of these correlation functions, it also becomes very large. Since susceptibility is the amount of change produced in the magnetization by a small change in the field H, we see that near the critical temperature this change

will be large. For the behavior of the spin correlation near T_c, see Wu, McCoy, Tracy and Barouch (1976). For a survey of these results, see Tracy (1978).

It would seem natural that the finiteness of the susceptibility should imply the central limit theorem for the magnetization. Newman (1979) has shown that this is the case if in addition the FKG inequalities are valid. These inequalities hold valid for the Ising model with attractive potential (see Appendix 1).

Finally, we mention still one more way to determine the critical temperature in the two-dimensional Ising model. Consider a Gibbs measure on the infinite lattice. For a given configuration, we can look at sets of plus spins connected through bonds of the lattice. Maximal such sets are called clusters. Consider the event: "the origin belongs to an infinite cluster." Let p be its probability. Then if $p>0$, one says that percolation occurs. It was proven by Coniglio, Nappi, Peruggi, and Russo (1977) that for $H=0$ percolation will occur in P_+ if and only if $T < T_c$. Again it seems reasonable that the long range dependence which occurs below the critical temperature should come through the presence of large clusters. However, it is not believed that this same result holds in three dimensions. It is believed that percolation first takes place at a temperature $\overline{T}$ below T_c.

We next consider a quite different way of looking at Gibbs measures on finite lattices. The technique that we will now look at will further serve as an introduction to the next chapter on dynamic models.

The use of Gibbs measures in physical applications arose from attempts to give a physical description of a system of spins

in a magnet or molecules of a gas in an equilibrium state. This means that, while there are chance fluctuations, exterior conditions that effect the system have been held constant and enough time has passed so that the effect of special initial conditions has worn off. For example, if the magnet starts with random orientations for the spins and a positive exterior field is imposed, it might take some time for this field to have an effect and during this time we can expect the distribution to change from random orientation to one with a higher probability for positive orientation (the equilibrium distribution).

While it is generally believed that Gibbs measures are appropriate for the equilibrium distributions, the process by which this distribution is reached is less well understood. However, a method introduced by Glauber (1963) for a physical system in one dimension has been generalized to give a class of processes which have Gibbs measures as equilibrium distributions. We shall describe these processes now.

First, let us state some basic facts about Markov chains. Consider a finite Markov chain with state space S and transition matrix P. The Markov chain is called <u>regular</u> if it is possible to go between any two states i and j in some fixed number N of steps. For any regular Markov chain there is a unique probability measure w defined on S which is invariant in the sense that $wP = w$. For any starting distribution $w^{(0)}$ for the chain, the distribution after time n, $w^{(n)}$, will approach w. That is, w_s represents the long range prediction for finding the process in state s. The law of large numbers for regular chains states that the average number of times the process will be in state s in the long run will be given by

w_s. Some, but not all, regular chains have the special property of being reversible. This means that when the process is in equilibrium, or started with the fixed vector w,

$$P(X_{n+1} = j | X_n = i) = P(X_{n-1} = j | X_n = i).$$

That is, in equilibrium the process looks the same forward as it does backwards; hence the name reversible. A regular chain is reversible if and only if

$$(2) \quad w_s P(s,t) = w_t P(t,s)$$

for all s,t in S.

Further, if P is regular and w is any probability vector that makes (2) valid, P is reversible with limiting vector w.

Consider now a finite graph G on which a Gibbs measure

$$p(x) = \frac{1}{Z} e^{-U(x)}$$

is assigned. We assume that each component of x is a +1 or a -1 and use the terminology "spin up" or "spin down." We shall form a Markov chain with states the possible configurations x on G. We define the transition matrix R as follows. We choose an arbitrary strictly positive probability measure q(s) defined on G. If at any time we find the process in state x, we choose a site by q. Let $\bar{x}$ be the configuration obtained from x by reversing the spin at site s. We compute the energy difference $U(\bar{x})-U(x)$. If $U(\bar{x})-U(x) \leq 0$, the chain makes a transition from x to $\bar{x}$. If $U(\bar{x})-U(x) > 0$ it makes this transition with probability $e^{U(x)-U(\bar{x})}$. If it does not make a transition to $\bar{x}$ it remains in its original state. Thus the transition probabilities are given by

(a) If $U(\bar{x})-U(x) \leq 0$

$$P(x,\bar{x}) = q(s)$$

(b) If $U(\bar{x})-U(x) > 0$

$$P(x,\bar{x}) = q(s)e^{U(x)-U(\bar{x})} .$$

Note that, in this chain, only one spin can change during one step. Then P is a regular chain with equilibrium measure P(x). To see this we need only verify (2), i.e.,

$$\frac{e^{-U(x)}}{Z} P(x,\bar{x}) = \frac{e^{-U(\bar{x})}}{Z} P(\bar{x},x)$$

or

$$(3) \quad \frac{P(x,\bar{x})}{P(\bar{x},x)} = \frac{e^{-U(\bar{x})}}{e^{-U(x)}} = e^{U(x)-U(\bar{x})} .$$

Consider first the case $U(\bar{x})-U(x) \geq 0$. Then $U(x)-U(\bar{x}) \leq 0$. Thus

$$P(x,\bar{x}) = q(s)e^{U(x)-U(\bar{x})}$$
$$P(\bar{x},x) = q(s)$$

and (3) is satisfied. Similarly if $U(\bar{x})-U(x) \leq 0$ then $U(x)-U(\bar{x}) \geq 0$ and

$$P(\bar{x},x) = q(s)e^{U(\bar{x})-U(x)}$$
$$P(x,\bar{x}) = q(s) .$$

Thus, in both cases, (3) is satisfied, and p(x) is the equilibrium distribution for P.

We have given one choice for the transition matrix P which yields a reversible Markov chain with fixed vector a given Gibbs measure. This choice is by no means unique. A different choice will be given in the next section.

It is very difficult to compute the properties of a Gibbs

measure $p(x)$ on a lattice if the size of the lattice is larger than 6 x 6. The fact that we can obtain $p(x)$ as the limiting distribution has been used to find random configurations distributed according to $p(x)$ as follows: given the parameters and the boundary conditions that determine $p(x)$, we construct the Markov chain with limiting measure $p(x)$. We choose q to be the equiprobable measure. Then we run this Markov chain for a large number of trials and look at the configuration. This represents a sample chosen at random according to $p(x)$, assuming that our Markov chain has reached equilibrium. If a certain property is very likely to occur under $p(x)$, we can expect to see this property in our sample.

As our first example, we return to the familiar question asked by Ising. Can we expect that if the temperature is low enough a substance will become a magnet as a result of interactions only? To consider this, we would want to choose a Markov chain with no boundary effect and no exterior field but a number of sites relatively large. We simulate the Markov chain corresponding to a 50 x 50 lattice with periodic boundary condition, no external field. We start with a distribution with the orientations chosen independently to be + or - with equal probability. We have had the computer draw contour lines around groups with like spins. A convention is adopted which chooses $\begin{matrix} - & + \\ + & - \end{matrix}$ instead of $\begin{matrix} - & + \\ + & - \end{matrix}$ and $\begin{matrix} + & - \\ - & + \end{matrix}$ instead of $\begin{matrix} + & - \\ - & + \end{matrix}$. For random choice (no interaction) there are no large areas having the same spin orientation. In Figure 14 we show the result of 500,000 iterations for $T = .7T_c$. Recall that $T = 1/b$. We see that we have obtained a positive magnet. This was achieved without the use of any exterior field. In Figure 15 we show the result after 500,000 iterations when we choose $T = T_c$. We now have quite

irregular regions of a variety of sizes. In Figure 16 we show the result after 500,000 iterations when $T = 1.5T_c$.

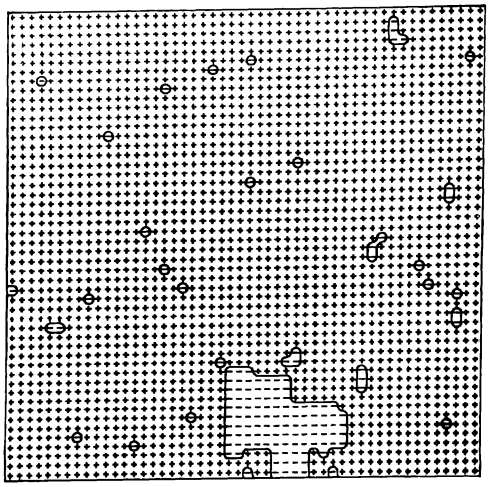


FIGURE 14

Simulation for $T = .7T_c$ with periodic boundary condition after 500,000 iterations.

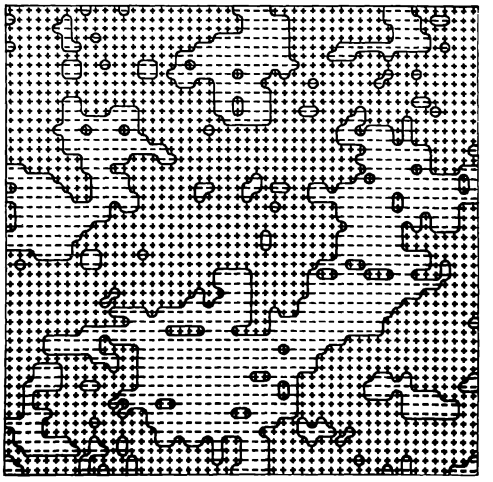


FIGURE 15

Simulation for $T = T_c$ with periodic boundary condition after 500,000 iterations.

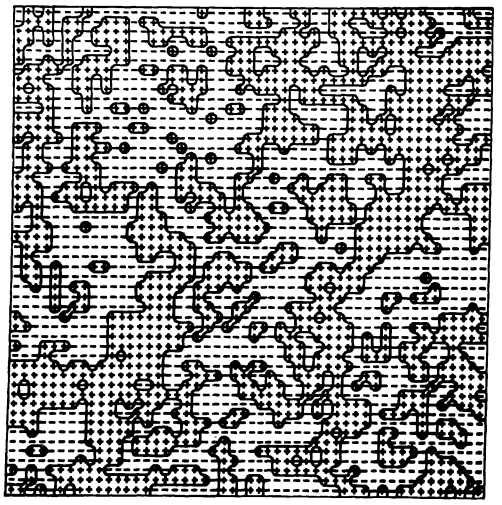


FIGURE 16

Simulation for $T = 1.5T_c$ with periodic boundary condition after 500,000 iterations.

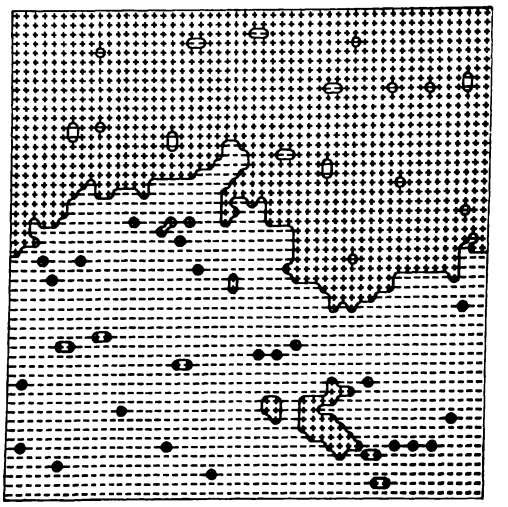


FIGURE 17

Simulation with $T = .9T_c$ and boundary condition, top half + and bottom half −, after 500,000 iterations.

Now we have a great deal of irregularity and the picture is similar to that one obtains by purely random choices, i.e. no interaction. Of course in each case we have considered only one sample. However, these do represent "typical" configurations for these three choices of temperature.

As a second example, consider the following question. Let us take a square lattice and fix the top half of the boundary to be + spins and the bottom half to be - spins. Then if we draw contour lines we will obtain one line which goes between the two boundary points which separate the + and - boundary spins. We call this a surface contour. All other contours will be closed regions. We choose $h = 0$ and $T = .9T_c$. In Figure 17 we show the result of the simulation after 500,000 iterations. We see that we have a rather long surface line and in the top half we have mostly positive spins and the bottom we have mostly negative spins as might be expected.

As usual, our finite results do not tell us what will happen if we keep b fixed and increase the size of the lattice indefinitely. It was proven by Messager and Miracle-Sole (1975) that when this is done we obtain a limiting measure which is the same as that obtained for the free boundary, namely the average of P_+ and P_-. This is rather surprising since it says that, in the limit, the probability of finding a + spin at, for example, one unit above the center is the same as at the center. Since Aizenman (1979) has shown that every Gibbs measure in two dimensions is a convex combination of P_+ and P_- we see that the separation of phases that we might expect can, in fact, never occur in two dimensions. The corresponding result in three dimensions is quite different. In this case as the cube increases indefinitely, a limiting measure P is obtained which,

for b sufficiently large, is not spatially homogeneous. More specifically Dobrushin (1973) has shown that there is a $\bar{b}$ such that for $b > \bar{b}$ the probability of a + spin is not the same at all sites. For example, the probability under P of finding a + spin at one unit above the center of the cube is now greater than at the center. A simplified proof of Dobrushin's result may be found in Van Beijeren (1975). Dobrushin only proved that the phases can be separated below a temperature $\bar{T} < T_c$. It is conjectured that this result will in fact only hold for T less than some $T_c' < T_c$.

It is also of some interest to observe quantities relating to a Markov process as it changes with time. In Figure 18 we give a graph which plots the proportion of + spins for a 6 x 6 lattice every 1000 steps for 100,000 iterations. We have assumed b = .88, i.e., around the critical value, no exterior field, and periodic boundary. We see that the fluctuations are very great. Indeed the graph shows that most of the time we would have almost all spins up or all spins down.

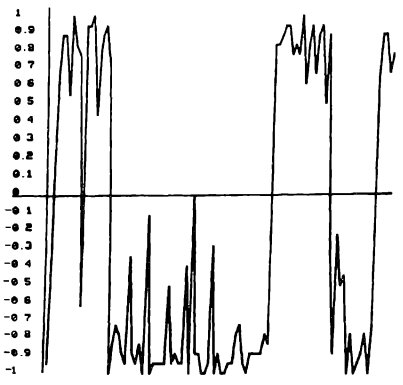


FIGURE 18

Magnetization every 1,000 steps for simulation with 500,000 steps 6 × 6 lattice with periodic boundary condition.

This stochastic Ising model has also been used to study quantities like susceptibility and specific heat. As we have seen, these quantities represent variances of the magnetization

and energy computed with respect to the Gibbs measure.

To see how this may be achieved, we first recall the law of large numbers for regular Markov chains. If f(i) is any function defined on the state space of such a chain, the law of large numbers states that, if $x^{(1)}, x^{(2)}, \ldots$ are the outcomes for the chain,

$$\frac{f(x^{(1)}) + f(x^{(2)}) + \ldots + f(x^{(n)})}{n} \longrightarrow E_x(f)$$

where $E_x(f) = \sum_i x(i)f(i)$ is the expected value of f computed using the equilibrium measure x. The reliability of this estimate will be determined by the limiting variance of the average.

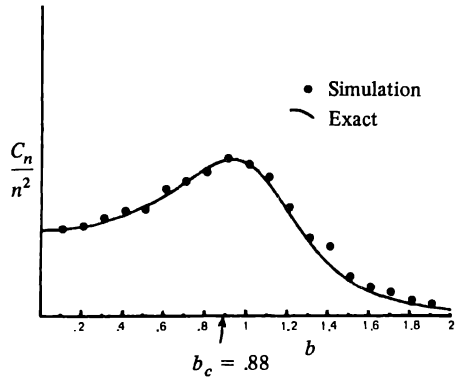


FIGURE 19

Specific heat per site for a 6 × 6 lattice with free boundary exact and by Monte-Carlo.

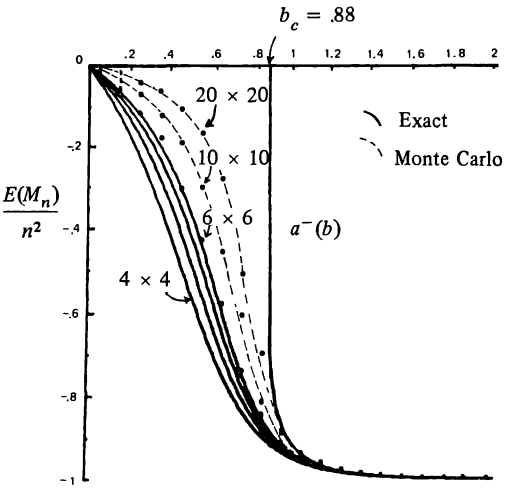


FIGURE 20

Expected magnetization per site as a function of b for lattice sizes 3 × 3, 4 × 4, 5 × 5, 6 × 6 (exact calculations) and 6 × 6, 10 × 10, 20 × 20 by Monte-Carlo estimates. Negative boundary and zero external field.

Consider the problem of determining the specific heat for an n x n lattice with free boundary. If $U_n(x)$ is the number of odd bonds in x, we want $E(U_n^2)$ computed with respect to the Gibbs measure. By the law of large numbers, we can estimate this by averaging this quantity over a large number of iterations of our stochastic Ising model. In Figure 19, we show the result of

doing this for b = 0 to 2 for a 6 x 6 lattice. We have plotted the specific heat per site. That is, $\frac{C_n}{n^2} = b^2 \frac{V(U_n)}{n^2}$. We can use our 6 x 6 exact calculation to compare with our simulation. We see the fit is quite good.

By choosing a - boundary we can, by simulation, extend our graphs for the magnetization shown in Figure 4. In Figure 20, we show the result of the 6 x 6, 10 x 10, and 20 x 20 together with our previous exact values. Again we see that the 6 x 6 fit is quite good, but the 20 x 20 lattice is still far from the limiting case.

4. DYNAMIC MODELS

We consider now a broad class of Markov processes which will include dynamic Ising models similar to those that we have already encountered.

We have, at each lattice point in R^n, an object which, at any moment, is in one of a finite number of states. The object may be an automaton, a voter, a neuron, or even a tree. The state of the voter might be "yes" or "no", the state of the neuron "firing" or "not firing", the state of the tree "diseased" or "not diseased". We shall often use the neutral term lamps. The Russian school initiated a systematic study of these processes. We shall, for simplicity, consider the case of only two possible states, off (0) and on (1).

We consider these processes both in discrete time and continuous time. In discrete time our model may be described as follows. Each lamp has a finite set N of neighbors (possibly including itself). At each time unit each of the lamps observes the state of its neighbors. We then toss a coin to decide whether the lamp will be on or off at the next time unit. The particular coin used depends upon the state of its neighbors, thus $2^{|N|}$ coins may be necessary. All the tosses occur simultaneously and are independent of each other. The whole system is invariant under translations. That is, if the neighbors of the lamp at a are N_a then the neighbors of the lamp at $a + b$ are $N_a + b$. For the lamp at a and $a + b$, we will use the same kind of coin if the states of their neighbors are the same.

One of the first classes of models studied consisted of models for neuron firing. We put a 1 if the neuron is firing and a 0 if not. For this model we assume that a threshold

number m is given. The neuron at a will fire next time if the number of its neighbors that are presently firing is $\geq$ m. In addition there is a probability θ that it will fire spontaneously independent of the state of its neighbors. A special case which has been studied extensively is the Stavskaya model (Stavskaya, Pyateskii-Shapiro (1968)). Here we are in one dimension and have a lamp at each integer positive or negative. The neighbors of the lamp at a are the lamps at a and $a + 1$.

The threshold value is $m = 2$. Thus the transition rules are as follows. If a and $a + 1$ are on, the lamp at a is necessarily on at the next time. If either of these lamps is off then the lamp will be on with probability θ at the next time. Thus $p(11) = 1$ and $p(00) = p(01) = p(10) = \theta$.

Assume that initially all the lamps are off. Let $p_n(\theta)$ be the probability that the lamp at the point 0 is on at time n. We shall show later in this section:

(A) $p_n(\theta)$ is increasing in n and hence has a limit $p(\theta)$,

(B) $p_n(\theta)$ increases with θ,

(C) if θ is sufficiently small $p(\theta) < 1$, and if it is sufficiently large $p(\theta) = 1$.

Thus there will be a critical value $\bar{\theta}$ such that for $\theta < \bar{\theta}$, $p(\theta) < 1$ and for $\theta > \bar{\theta}$, $p(\theta) = 1$. In the latter case, the probability that any finite set of lamps are all on will approach 1. If we start with any of the lamps on we would only make $p_n(\theta)$ greater than it is starting with all the lamps off. Thus if $p_n(\theta)$ approaches 1 starting with all the lamps off, the same will be true for any initial configuration. Therefore for

$\theta > \bar{\theta}$ the long range predictions are independent of the initial state. This is an ergodic type behaviour which we now describe a bit more carefully. A more detailed discussion of ergodicity for this model may be found in Shnirman (1968).

Let $x_a^{(n)}$ be the state of lamp a at time n. Assume that we choose the initial states by a measure $m^{(0)}$. When there are only two outcomes, 0 and 1, this measure will be determined by the measure of cylinder sets of the form

$$\{x_{a_1}^{(0)} = 1,\ x_{a_2}^{(0)} = 1,\ \ldots,\ x_{a_r}^{(0)} = 1\} .$$

These probabilities and our transition rules determine a measure $m^{(1)}$ for the outcomes at time 1. The measure $m^{(1)}$ is determined by its value on cylinder sets of the form

$$\{x_{a_1}^{(1)} = 1,\ x_{a_2}^{(1)} = 1,\ \ldots,\ x_{a_r}^{(1)} = 1\} .$$

From our assumptions, these probabilities depend only on the state of a finite set of lamps at time 0 (the lamps which are neighbors of lamps $a_1, a_2, \ldots, a_r$). Thus we need only use our cylinder measures at time 0 to calculate the measure of cylinder sets at time 1. We shall indicate this transformation by $m^{(1)} = m^{(0)}P$ where P is a transition operator. P acts as a transition matrix as in finite chains. Indeed, if we had only a finite number of lamps we could consider our process as a finite Markov chain with state at time n the configuration $(0, 1, 0, 1, \ldots, 1)$ giving the state of each lamp. Then P would be the transition matrix for this chain.

If m is a measure such that $m = mP$, we say that m is an invariant measure. For our processes there is always at least one such measure. If there is only one such measure m we say the process is ergodic. In this case, for any initial measure

$m^{(0)}$, $m^{(n)} = m^{(0)}P^n$ will converge to m in the sense that the measure of cylinder sets computed by $m^{(n)}$ converges to the measure of these sets as determined by m.

In the Stavskaya model, if we start with all lamps on, they will stay on for all time. Thus an invariant measure is the measure which assigns probability one to the sequence of all 1's. If $\theta > \bar{\theta}$ this is the only such measure. If $\theta < \bar{\theta}$, in addition to this measure, the limiting measure obtained by starting with all 0's will be invariant. Any convex combination of these measures will be invariant. For θ small enough it has been proven (see Vasilyev (1970)) that these are the only invariant measures.

We now briefly describe the way in which one proves the existence of the critical value $\bar{\theta}$ for the Stavskaya model. We shall first calculate the probability p_4 that the 0 lamp is off at time 4 when we start with all lamps off at time 0. We first note that this probability depends only on the lamps in the triangle in Figure 1. We introduce an auxiliary process as follows. At each point of our triangle in Figure 1, corresponding to time between 1 and 4, we assign a 1 with probability θ and a 0 with probability $1 - \theta$. We put a 0 at all points corresponding to time 0.

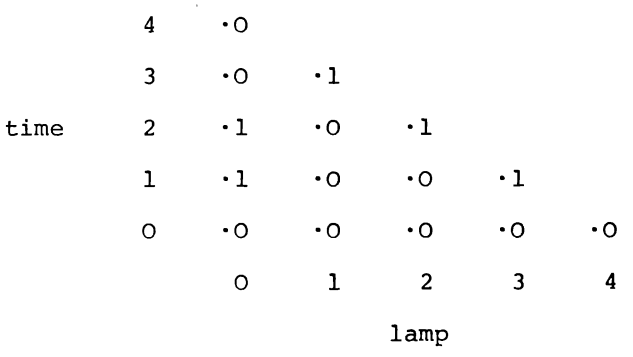


FIGURE 1

These outcomes will be used for the spontaneous firing when appropriate. Thus at any point where there is a 1 the lamp will go on. Of course, the lamp can also go on when the auxiliary process assigns a 0 because of the threshold effect. Now we claim that the 0 lamp will be off at time 4 if and only if, in the auxiliary process, there is a path of 0's going down from the top to level 1 (and hence, also to level 0). For a path we are allowed to move vertically down or diagonally down and to the right. To see this, we note that if the 0 lamp is off at time 4 we must have a 0 at the top in the auxiliary process. Then one of the two auxiliary outcomes at level 3 must be a 0, for if they were both 1, lamps 0 and 1 would be on at time 3 and lamp 0 would be on at time 4. Thus we have our path from level 4 to level 3. Now at level 3 one of the lamps 0 or 1 must be off or the 0 lamp would be on at time 4. Thus we can repeat the argument for this point and the two points below it to extend our path from level 3 to level 2 and then again from level 2 to level 1. Thus if the 0 lamp is off at time 4, there will be such a path. Conversely, assume that such a path exists. Then at level 1 the lamp on the path will be off since the auxiliary process is 0 and both lamps below it are off. At level 2, the lamp on the path must be off because the auxiliary process is 0 and one lamp below it is off. Continuing this way we see that the lamps must all be off along the path.

Thus the problem of determining the probability that the lamp 0 is off at time 4 and more generally at time n is reduced to a problem purely in terms of the auxiliary process. The problem is to find the probability that there will be a path of 0's from level 1 and, hence, also from the bottom of our

triangle to the top. This can be viewed as a "percolation problem" as follows. Let us "wet" all the sites at level 0. We will then let liquid flow vertically up or diagonally up and to the left if this leads to a 0 in the auxiliary process. If there is a 1, liquid cannot flow to this point. That is, a 1 "blocks" the flow of water into a point, and so it does not become wet. Thus in our example, the points which would be wet are as in Figure 2.

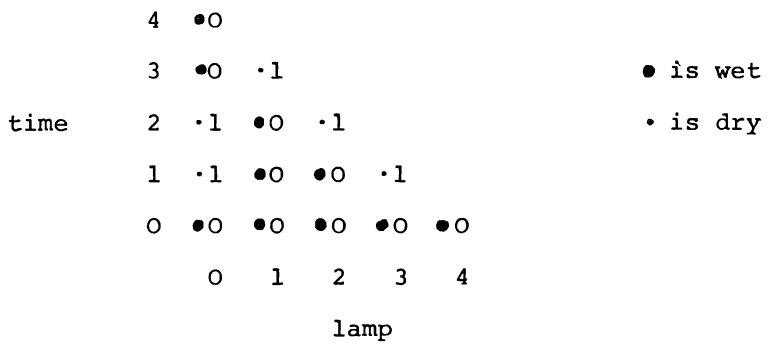


FIGURE 2

Recall that $p_n(\theta)$ is the probability that the lamp at 0 is on at time n when we start with all lamps off. In terms of our percolation process, $q_n(\theta) = 1-p_n(\theta)$ is the probability that water will reach the top when we wet all the bottom points. It is clear from the percolation interpretation that $q_n(\theta)$ decreases with n and increases with θ. We want to show (a) for θ large enough $q_n(\theta) \to 0$ as $n\to\infty$, and (b) for θ small enough $q_n(\theta) \to b(\theta) > 0$. In terms of the lamps, this means that for θ large enough, the probability that the lamp at 0 is on at time n tends to 1 and for θ small enough, this probability tends to a limit < 1.

First we prove (a). There are 2^n possible paths on which water can flow to the top. The probability that any one of these

paths is open is $(1-\theta)^n$. Thus the probability that there is at least one path open is $\leq 2^n(1-\theta)^n$. Therefore if $\theta > \frac{1}{2}$, this probability approaches 0 and (a) is established. More sophisticated arguments have established that the same is true for $\theta \geq .318$. (Gray, Wierman and Smythe (1979))

The proof of (b) uses a contour argument similar to the Griffiths-Peierls argument for the Ising model. We next sketch this proof.

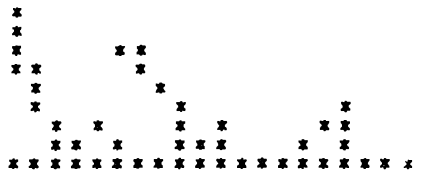


FIGURE 3

Simulation of the percolation process for the Stavskaya model with $m = 20$ and $\theta = 45$. The points $*$ are wetted.

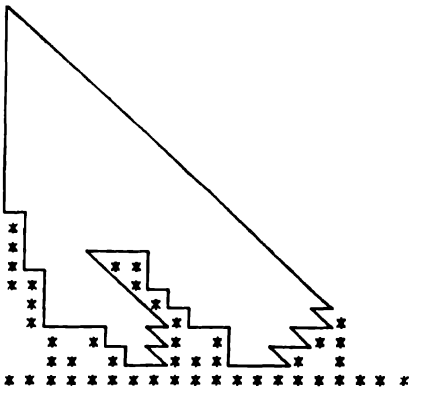


FIGURE 4

The blocking path for the simulation shown in Figure 3.

In Figure 3 we show the result of simulating our percolation process for $n = 20$ and $\theta = .45$. We note that water did not reach the top. Now when this happens we can draw a closed contour as follows. We add new points to our space time diagram located at the same levels but midway between the original points and to the left of the points corresponding to position 0. Our contour will be obtained by connecting certain of these points. We start at the point $(-.5, n+1)$ at the top of our triangle. We are allowed to move diagonally down and to the right, horizontally to the left, and vertically upwards. We want our contour to return to the starting point and have the property

that each horizontal segment passes through a point which blocks water. It can be shown that such a contour always exists. In Figure 4 we show the contour for our example in Figure 3. Now any such contour has the same number of unit line segments of each of the three types (else it would not get back to the starting point). Thus if the contour has length 3K the probability that this contour will occur is $\leq \theta^K$ (since it must pass through K blocked points). The number of these contours of length K is at most 3^{3K} (since at each move it must make one of at most three choices for direction). Thus the probability that water does not get to the top is bounded by

$$\sum_{K=1}^{\infty} 3^{3K} \theta^K = \frac{27\theta}{1-27\theta} \quad .$$

Therefore, if $\theta < \frac{1}{54}$ the probability that water does not reach the top will be less than $\bar{b} < 1$ for all n. This establishes (b).

We have seen that the Stavskaya model can be non-ergodic. The presence of a trap (all 1's) makes this a rather special kind of non-ergodicity. The question arises whether one can have non-ergodicity if none of the coin probabilities are zero or one. We shall see later that the dynamic Ising model will, for low enough temperature, be non-ergodic in two or more dimensions. It will be ergodic in one dimension for all temperatures. This is not surprising since we will define this model, as we did in the finite case, in such a way that the Gibbs measures will be the invariant measures. We know that in one dimension this measure is unique but not necessarily in two or more dimensions.

It would be natural to conjecture that lamp processes with

positive probabilities for each outcome should be ergodic in one dimension but not in two or higher dimensions. Toom (1974) has given examples in two dimensions that are not ergodic. Kurdyumov (see Gach, Kurdyumov and Levin (1978)) has announced a counter-example to the conjecture in one dimension but it is very complicated and the details have not been published.

A non-ergodic example in two dimensions is the following. We first apply to each point a deterministic transformation D defined as follows. The neighbors of (0,0) are the points (0,1), (1,0), (-1,0), (0,-1). The transformation D turns lamp (0,0) on if either both lamps (0,1) and (1,0) are on or if both (-1,0) and (0,-1) are on. It turns the lamp off if either both (0,1) and (-1,0) are off or both (1,0) and (0,-1) are off. If none of these cases happens, it leaves the lamp unchanged. After D is applied the result is changed (error occurs) with probability q, independent of the neighbors. In this scheme the configurations of all 0's and of all 1's are fixed points for the deterministic transformation. If the initial configuration is all 0's, Toom proved that, for small enough q, the probability that the lamp at (0,0) is on is less than $r < \frac{1}{2}$ for all n. By symmetry, if we start with all 1's the probability will be greater than $1-r > \frac{1}{2}$ and hence the process will not be ergodic.

A somewhat more natural model, called the voter model, is generally thought to behave in much the same way. In this model, the neighbors of a point are the point itself and the points a unit distance from it. The deterministic transformation is simply a majority vote. After this vote, the voter switches his vote with a small probability q independent of the position of his neighbors. Again the configurations of all 1's and all

O's are fixed points for the deterministic transformation. This example is believed to be non-ergodic in 2 dimensions and ergodic in one dimension.

Gach, Kurdyumov and Levin (1978) have suggested that simple counterexamples to ergodicity in one dimension might be obtained as follows. Find a deterministic transformation D with the following properties. First it has two fixed points, say the configurations of all O's and all 1's. Next require that D "wash out islands" in the following sense: if a configuration has only a finite number of 1's then after a finite number of iterations of D all the one's will disappear. Similarly, if there are only a finite number of O's , D should wash these out in a finite time. One example of such transformation is the following. If the lamp at a is on it looks at the lamps at a, a+1, and a+3 and takes a majority vote. If it is off it does the same but the vote is among the lamps a, a-1, and a-3. We shall call this the unsymmetric voter model. Gach, Kurdyumov, and Levin provide computer simulations to support the claim that this process with sufficiently small random error is non-ergodic.

The role of computer simulations in the question of ergodicity raises an important point. When we simulate our infinite process we must use a finite number of lamps. Thus we will have a finite Markov chain and if all coin probabilities are non-zero, the chain will certainly be ergodic. When the corresponding infinite model is non-ergodic, these finite models will take a very long time to reach equilibrium. Thus examples where the finite models take a very long time to reach equilibrium are natural candidates for non-ergodic infinite models. Since it has been suggested that the symmetric voter model is ergodic and the unsymmetric model not, it is interesting

to compare the time required to reach equilibrium in these two models. We shall examine this by looking at the proportion of 1's after time n when we start our finite chain with all 0's. By symmetry, in equilibrium, the distribution of this quantity will be symmetric about $\frac{1}{2}$. In Figure 5 we show the result of simulating these two models for q = .02. We have assumed 1000 voters and run the process 10,000 times. We note that the symmetric voter model has reached equilibrium but that this is not the case in the non-symmetric voter model.

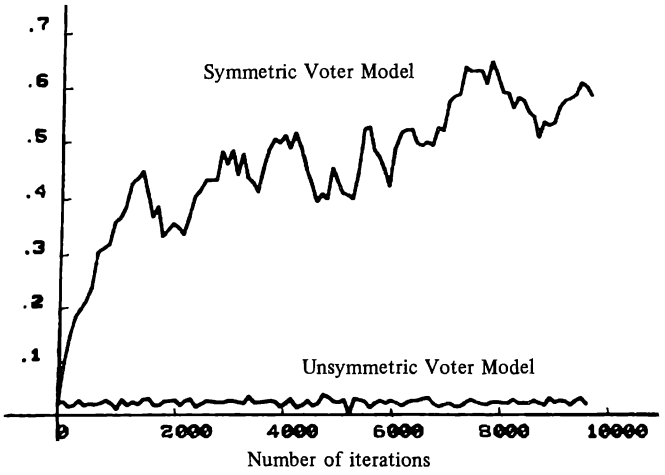


FIGURE 5

Proportion of ones after *n* iterations with 1000 voters starting with 2% ones.

We consider next continuous time models. We again have a lamp at each lattice point. Each lamp is, at any time t, either on or off. We shall denote by x a typical configuration which specifies the state of each lamp. We denote by x_a the state of lamp a in such a configuration. If we are looking at the state of the lamps at time t we shall indicate these quantities by x^t and x_a^t. We associate with each lamp a clock. This clock has an alarm set to go off at an exponential time with mean $1/c_a(x)$. Thus, loosely speaking, $c_a(x)dt$ is the

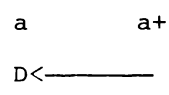
probability that the clock will ring in a short time dt when the configuration is x. We shall, as usual, assume there is a finite set of neighbors N_a associated with a. We assume that $c_a(x)$ depends only on x through the states of N_a. When the alarm at a goes off, this lamp switches its state and we have a new configuration $\bar{x}$. When this happens the clock at any point b which has a as a neighbor changes its rate to $c_b(\bar{x})$ to take into account the new configuration $\bar{x}$. If we have a finite number of lamps we can construct a continuous time Markov chain to represent this process. The states would be configurations x and the infinitesimal matrix $Q(x,y)$ would be determined by

$$Q(x,\bar{x}) = c_a(x) \quad \text{if } x=\bar{x} \text{ except at } a$$
$$= 0 \quad x \text{ and } \bar{x} \text{ differ at two or more sites}$$
$$Q(x,x) = -\sum_{\bar{x}\neq x} Q(x,\bar{x}) \ .$$

Then $P^t = e^{Qt}$ and we can analyze our process by standard Markov chain methods. We could obtain our infinite process as a limit of these Markov chains. Other methods have been developed to define the infinite processes directly. See Liggett (1977) for a review of these methods. It is possible in certain classes of examples to avoid some of the technical difficulties of defining these infinite processes by means of a graphical representation due to Harris (1978). A general description of this method and its application to a number of models may be found in Griffeath (1979). We next briefly describe this method.

With each point of our lattice we associate a finite collection of exponential clocks. With the ith clock we associate a labeled directed graph G_i. The points of the graph G_i are a finite set of lattice points. Directed means that we draw an arrow between certain of these points. Labeled means

that we label certain of these points with a D or a B. When the ith clock goes off we attach the graph G_i to our space time diagram at the time point corresponding to the times that the clock goes off. For example, assume that the graph G_i is on the points a,a+1 and has the form

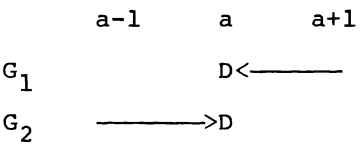


Then if the clock goes off at rate 1/2 we would attach an arrow with a D at the tip from a+1 to a at each of a sequence of time points over the point a generated by exponential random variables with mean 2.

When we have done this for all our points and clocks, we have a random graph with D's, B's and arrows, as in Figure 6. Such a graph is called a percolation structure. Note that these rules allow D's to come at either end of the arrows. In our percolation interpretation, we will wet a certain number of the points at time 0 and all the points (s,t_n) which have been labeled B (these are spontaneous springs). Water will flow up the lines. If it comes to a D, it is blocked from flowing up. If it comes to the foot of an arrow, it will flow along this arrow and then up the line at the tip of the arrow. We say that there is a path from (a,s) to (b,t) if there is a way to move from (a,s) to (b,t) going up the lines or horizontally along the arrows without passing vertically through a D. Then water will reach the point (a,t) if there is a path from a point (b,0) which was wet or a path from a point (b,s) which was labeled with a B. We shall define a Markov process x^t with state space consisting of configurations of 0's and 1's. The initial configuration x^0 will be the points which are wet

at time 0. Then $x_a^t = 1$ if water can reach a at time t and 0 otherwise.

We illustrate this first in terms of a voter model. In this model each voter has, at any time, one of two positions on a particular issue, "for" (denoted by 1) or "against" (denoted by 0). A voter's neighbors are the points one unit away. At exponential times the voter switches his position. The rate for the voter change is equal to the proportion of his neighbors who hold an opinion opposite to his. For this model, in one dimension we use the following percolation structure. With each point a, we have two clocks. Each clock goes off at exponential rate 1/2. The graphs associated with these points are on the set of points a-1,a,a+1. They are



Thus if the first clock goes off at time s, we put a D at point (a,s) and draw an arrow from (a+1,s) to (a,s). If the second clock goes off at time t, we put a D at (a,t) and draw an arrow from (a-1,t) to (a,t). This results in a graph as shown in Figure 6.

If there is a 1 at point a at time s, then the probability that this changes to a 0 in a short time dt is the probability that the next clock that rings is a clock associated with a 0 neighbor, and this is $\frac{1}{2}ndt$ where n is the number of 0 neighbors. Similarly, if the state at a at time s is 0, then this will change to a 1 in a short time dt if the next clock to go off is from a neighbor which has a 1. This happens with probability $\frac{1}{2}ndt$ if n neighbors are 1.

In Figure 7, we show the result of choosing an initial configuration and using the auxiliary graph of Figure 6 to determine the distribution at time t.

We wish to determine the distribution for the voters at time t. As in the discrete time case, it is sufficient to compute probabilities of the form

$$P(x^t_{a_1}=1,\ x^t_{a_2}=1,\ \ldots,\ x^t_{a_r}=1)$$

for a given initial configuration of 0's and 1's. In terms of the percolation process this is the probability that the random graph is such that for our initial configuration these points will be wet at time t. It is often easier to find these probabilities in terms of a dual graph obtained by reversing all of the arrows.

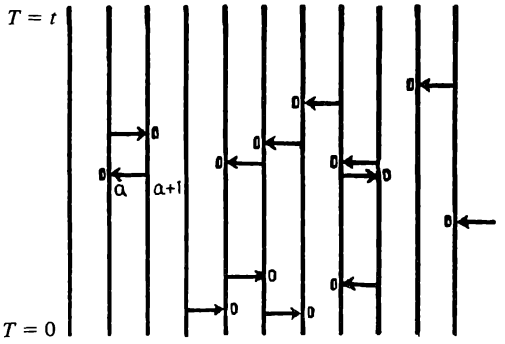


FIGURE 6

A typical percolation structure for the voter model.

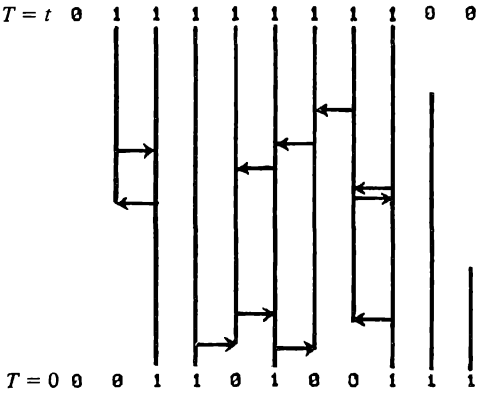


FIGURE 7

The percolation paths for the percolation structure of Figure 6.

The dual of our graph in Figure 6 is the graph in Figure 8. With the dual graph, we associate a new process which starts at the top with a configuration of 0's and 1's and moves down with the same rules as our original process moved up. Note that there

is a path from (b,s) to (a,t) in the original graph if and only if there is a path from (a,0) to (b,t-s) in the dual graph.

In Figure 9, we show the result of choosing an initial configuration and using the graph of Figure 8 to determine the distribution at time t.

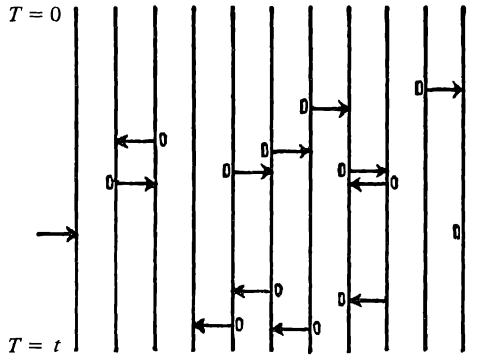


FIGURE 8

The percolation structure for the dual of the voter model simulation shown in Figure 6.

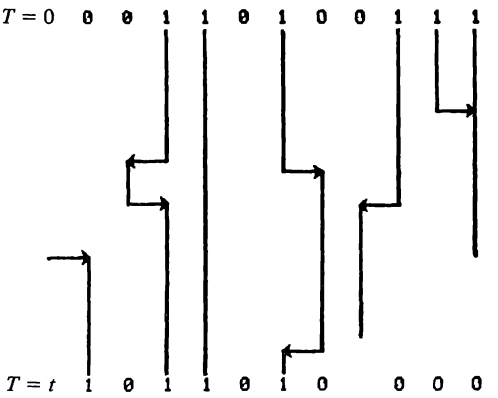


FIGURE 9

The percolation paths (top to bottom) resulting from percolation structure of Figure 8.

In our voter example, the dual process turns out to be easier to study than the original. Whenever we come to an arrow in the dual graph we move one unit to the left or to the right. We arrive at arrows at exponential rate 1 and then have probability $\frac{1}{2}$ of going to the right and probability $\frac{1}{2}$ of going to the left. Thus, paths in the dual graph from a point (a,t) can be viewed as paths of a continuous time random walk.

Consider now $P(x_a^t = 1)$. In our original process, this is the probability that there is a path which starts at a point (b,0) with a one and ends at the point (a,t). In the dual process, this is the probability that our random walk starting at (a,0) and moving down will, after time t, be at a point b

which was assigned a 1 in the initial configuration for the original process. While this probability is well defined for any t, the limit as $t \longrightarrow \infty$ may not exist.

As an example where this limit does exist, assume that initially all voters on the positive integers have opinion 1 and those at the negative integers and 0 have opinion 0. Then the only change that can occur is in the boundary of the 0's and 1's. The position of this boundary will again be a continuous time random walk. Thus the probability that any finite set of voters will be to the right of the border at time t and, hence, all have position 1 approaches $\frac{1}{2}$ as $t \longrightarrow \infty$. However, the boundary will be infinitely often to the right of these voters, and at these times they will all have opinion 0. It will also be infinitely often to the left, and at these times they will have position 1. Thus while we can expect after a long time to find consensus for any finite group, the position of the consensus will keep changing.

Clifford and Sudbury (1973) remark that if the initial configuration is such that the proportion of 1's in the interval from -N to N approaches p as $N \longrightarrow \infty$, then the limit of $P(x_{a_1}^t = 1, \ldots, x_{a_r}^t = 1)$ as $t \longrightarrow \infty$ will also exist and be equal to p.

Assume next that the initial configuration is chosen by assigning the positions at random with probability p for a 1. Let us consider the probability that $x_a^t = 1$. By the dual argument this is the probability that our random walk path, starting t units above the point a and moving down, ends at a point with a 1. But, wherever it ends, this point was assigned a 1 with probability p independent of any of the path constructions. Thus $P(x_a^t = 1) = p$.

Assume now that we look at two voters a and b and wish to find the probability that at time t both a and b have position 1. Then we look at the dual graph and trace their random paths back to time 0. These paths are not independent random walks since if they ever meet, they continue on as a single path representing a single random walk. In one and two dimensions the probability that two random walks meet approaches 1 as t tends to infinity. Thus the limiting probability that a and b are both 1 is the same as the probability that a's final position is a 1. Thus, again the probability in the long run that both a and b have position 1 is p. More generally, the probability that any finite number of voters all have position 1 tends to p.

This means that if our assignments are independent initially, the measure after time t for our configuration will approach a measure which assigns probability p to a configuration of all 1's and probability 1-p to a configuration of all 0's. Thus we have seen two initial configurations which lead to the same consensus prediction. The measure which assigns probability one to the configuration of all 1's is an invariant measure for the voter model. The same is true for the measure which assigns probability one to the configuration of all 0's. Holley and Liggett (1975) showed that in one and two dimensions any invariant measure must be a convex combination of these two measures. Thus any limiting measure for the voter model in one and two dimensions must be a mixture of these two pure measures, and, in general, we can expect consensus.

In three dimensions, the probability that two random walks meet tends, as $t \longrightarrow \infty$, to a limit less than one. Thus the

limiting consensus that we obtained in one and two dimensions will no longer hold for three dimensions. In this case Holley and Liggett showed that if the initial distribution for the voters is random, with probability p for 1, there will be a limiting measure $m^{(p)}$ which assigns a positive probability for every possible opinion for any finite number of voters. While the voter opinions are no longer perfectly correlated in the limit as they were in the one or two dimensional cases, they will have strong correlations. These correlations can be calculated since they involve the probability that two random walks will meet and this has been studied in some detail. Using results of Spitzer about random walks, Griffeath and Bramson (1979) showed that, in terms of the measure $m^{(p)}$,

$$|E(x_r \cdot x_0) - E(x_r)E(x_0)| \doteq \frac{c}{r}$$

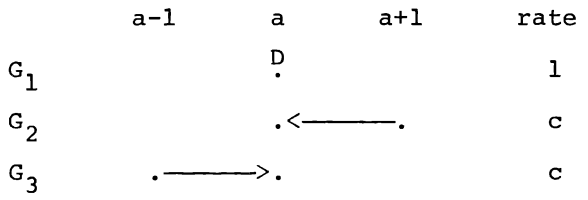
as $r \longrightarrow \infty$ where c is a known constant. This non-exponential decay of the correlations suggests that the classical central limit theorem should not hold. Indeed, Griffeath and Bramson showed that if S_n is the number of 1 voters in a cube centered at 0 with side n, then the distribution of

$$\frac{S_n - E(S_n)}{n^{5/2}}$$

tends to a normal distribution. They also proved a renormalization theorem exhibiting the dependence between large non-overlapping boxes.

A second model where this percolation structure applies is the process studied initially by Harris (1974). This is a model for the spread of a disease, say, in a forest. The states are, as usual, 0 and 1; 0 means a well tree and 1 a diseased tree. We shall illustrate this model in one dimension. Assume

that each tree a has two neighbors a-1 and a+1. A diseased tree recovers at a constant rate 1. A well tree becomes infected at a rate nc where n is the number of neighbors which are diseased. To form the percolation structure for this process, we associate with each point a three clocks. The graphs associated with these clocks and their rates are



The first clock goes off at rate 1. Thus if this clock goes off at time s we put a D at the point (a,s). The second clock has rate c. If this clock goes off at time t, we put an arrow from (a+1,t) to (a,t). The third clock also goes off at rate c. If this clock goes off at time u, we put an arrow from (a-1,u) to (a,u). This process is called a contact process.

Assume that we start with all 1's, in other words all trees infected. Let $p_t(c)$ be the probability that the 0 tree is infected after time t. In the percolation interpretation, this is the probability that water will get to 0 at time t. In Figure 10 we show a typical percolation structure for the contact process. Figure 11 shows a typical initial configuration and the paths which permit water to rise to the points at time t.

Assume that $s < t$. Then for water to get to 0 in time t it must get to level t-s and then in time s get to 0. Thus $p_t(c) \leq p_s(c)$ since $p_s(c)$ is the probability that 0 will be wet after time s if we start with all sites wet. Thus $p_t(c)$ has a limit $p(c)$ as t tends to infinity. Also, $p_t(c)$ will be increasing in c. Thus we again have the possibility for a

critical value $\bar{c}$ such that for $c < \bar{c}$, $p(c) = 0$ and for $c > \bar{c}$, $p(c) > 0$. The existence of this $\bar{c}$ was demonstrated by Harris (1974). We shall indicate a way to prove that $\bar{c} \geq 1$. The proof that $\bar{c} < \infty$ is much more difficult. The first proof due to Harris gave no easily computable upper bound. Holley and Liggett (1975) showed that $\bar{c} \leq 2$. For a more complete discussion of these and other estimates, see Griffeath (1979), (1980).

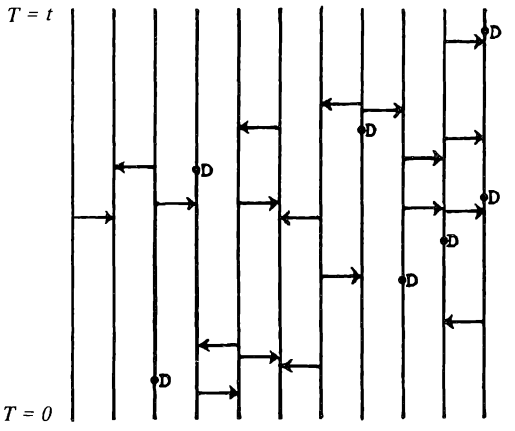


FIGURE 10

A typical percolation structure for the contact process.

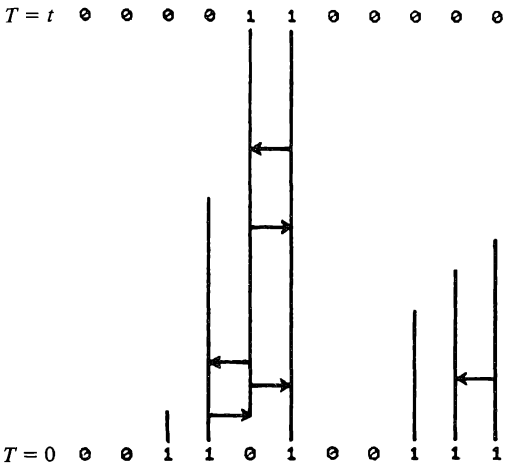


FIGURE 11

The percolation paths for the contact process with random initial configuration using the percolation structure of Figure 10.

The study of this model again makes use of the dual process. In the dual process if we start at (a,t) and move down the graph, when we come to an arrow, we can now continue down the path and also move horizontally. We shall regard this as a branching process thinking of a particle moving down the graph and splitting into two particles when it meets an arrow. If it meets a D, it dies. If two particles come together they continue together as a single particle. Thus the dual process is a continuous time branching process in which the particles coalesce when they come together.

Recall that $p_t(c)$ is the probability that the point 0 will be a 1 at time t when the contact process is started with all 1's. In the dual process this is the probability that a particle starting at 0 at time t will in time t reach a point that was initially a 1. But since we started with all 1's this is simply the probability that the dual process does not die out in time t. This process, in a short time dt, will die with probability dt or will split into two particles with probability cdt. Since even without coalescing such a process will die out if $c < 1$, we see that for c small enough the dual process will die out. In the original process this means that $p(c) = 0$ for such a c. To prove that $p(c) > 0$ for c sufficiently large, we would have to show that there is positive probability that the coalescing branching process grows indefinitely for c sufficiently large. This is of course true with coalescing but more difficult to prove. We have not drawn the dual graphs because the contact process is self-dual in the sense that when we form the dual process, we obtain a process of exactly the same probabilistic structure as the original. Thus the original process as illustrated in Figure 10 represents a coalescing branching process.

A variation in the voter model, the <u>biased voter model</u>, studied by Schwartz (1977), has recently been shown to have application to cell growth processes. For the biased voter model it is assumed that a voter switches opinions with rates

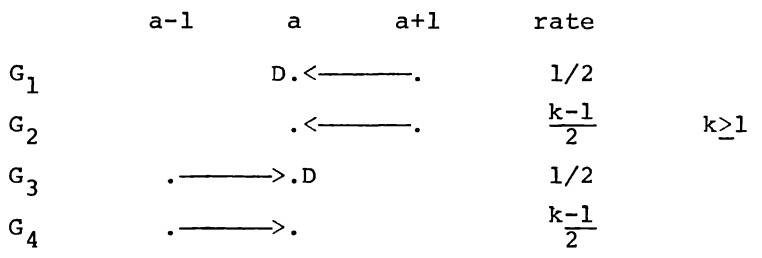
$c_a(x)$ = proportion of the neighbors that are 0 when $x_a=1$

$c_a(x) = k \cdot$ proportion of the neighbors that are 1 when $x_a=0$

where $k > 1$. Thus "for" voters have more influence than

"against" voters.

To construct this process by a percolation structure, we proceed as follows: With each point we associate four clocks. The graphs associated with these clocks are all on the set of points a-1, a, a+1. The graphs and the rates are:



If we start the dual process at a point a moving down from level t, then when it reaches an arrow with a D, it will act like our random walk. If it reaches an arrow without a D, it acts like a branching process. Both processes are coalescing since they act as only one particle if they come together. While this is a more complicated dual process, the study of the model has made heavy use of the dual process.

We consider next certain cell growth models related to these voter models.

Williams and Bjerknes (1972) and Richardson (1973) considered a number of models for cell growth. We can best describe these models in our setting as follows. For the Williams and Bjerknes model, at each lattice point in two dimensions, we have a 0 or a 1. A 1 stands for a bad (cancerous) cell and a 0 for a good cell. At exponential times, each cell splits into two cells. When the split occurs one cell remains at the same site and the other cell occupies a randomly chosen neighboring site and replaces the cell that was there. The rate at which a good cell splits is 1 and the rate at

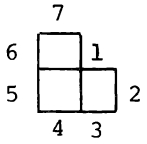
which a bad cell splits is $k > 1$. This is easily seen to be the same as the biased voter model.

Richardson (1973) studied several different models in which only the bad cells can split. He was interested in showing that under fairly general conditions there would be an asymptotic shape for the growth of the bad cells. He showed, in particular, that this is true for the Williams and Bjerknes model for the limiting case $k = \infty$. This model had already been considered by Eden (1961) who had conjectured a circular growth for the bad region. Richardson also showed that there is an asymptotic shape for a discrete time model defined as follows: at each time unit each of the good cells becomes a bad cell with probability p if there is at least one bad neighbor. If there is no such neighbor the cell does not change. He called this the G_p model.

For a class of models which includes these two models, Richardson proved that, if we start with a single cell, the cell growth will have an asymptotic shape in the following sense. The shape will be a circle determined by a distance $d(x,y)$, where a "circle" with center 0, using this distance, is the set of all lattice points x such that $d(x,0) \leq r$. Then r is the radius of the circle. One choice for this distance would be the Euclidean distance which would give familiar circles. A second choice might be the length of the shortest path from x to y passing through lattice points. With this distance, "circles" are diamonds. Richardson's results do not say what the distance is. They do say that in terms of the distance, the cell grows, as t increases, like a circle at a linear rate. More precisely, for any small $b > 0$, the probability that the growth includes a circle with a radius $t-tb$ and is included in a circle with radius $t+tb$ tends to 1 as $t \to \infty$. That is, if

A_t is the set of points that are 1 at time t, then the set of points $\frac{A_t}{t}$ will approach a circle using the distance d associated with the growth process being modeled.

The shape of the original Eden growth model is still not known. Much simulation has been carried out for this model. For this simulation, it is easier to think of the original Eden formulation. We start with a unit square and add a square at random to one of the available faces. We continue this procedure. For example, if after we have added two new squares and we have the configuration



then the next square will be put in position 1 with probability 2/7 and the other six positions with probability 1/6 each. In our simulation, we consider the squares centered at the latter points and put a dot at this point if the square is present in the growth. In Figure 12 we show a growth with 6000 squares.

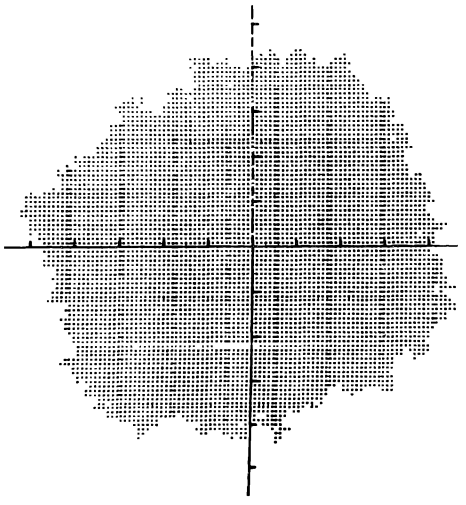


FIGURE 12

One simulation of the Eden growth model with 6000 points.

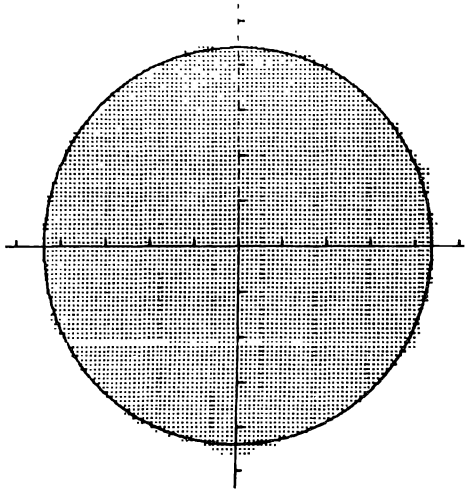


FIGURE 13

The points that occurred 20 or more times in 50 simulations of the Eden growth model, each simulation had 6000 points.

We see that the boundary of a single growth has a great deal of irregularity. To see if we can get a more regular region, we have repeated this simulation 50 times and plotted the points which occurred in 20 or more of these simulations. The results are shown in Figure 13. We have drawn a circle around the boundary. This would seem to support Eden's conjecture that the asymptotic shape is a circle.

The one growth model for which results about the shape have been proven is Richardson's Gp model. Richardson simulated the model for different values of p and conjectured that as p varies from 1 to 0, the shape changes from a diamond to a circle. Success with this model has come from identifying the growth with a percolation process, Cox (1979), and as a contact process, Durrett and Liggett (1979). Durrett and Liggett obtain a contact process by watching the growth along the 45 degree lines L_n consisting of the set of points (i,j) with $i+j = n$ for successive values of n. The growth can reach L_n in n generations but no further. A point (i,j) on L_n will be infected at time n, with probability p, if one or both of the points (i-1,j), (i,j+1) on L_{n-1} are infected at time n-1. Thus the growth along the lines L_n represents a Markov process which is like the contact process we have described earlier. There is a critical values p_c such that for $p > p_c$, this contact process does not die out, and for $p < p_c$ it does. In the former case, Holley and Liggett showed that the unit circle C_p contains a segment containing the point (1/2,1/2). For $p < p_c$, they showed that C_p does not intersect the line $x+y = 1$. The critical value p_c is not known but it follows from results of Gray, Wierman and Smythe (1979) that $p_c \leq .688$. We have also simulated this model. In Figure 14 we show the

results of a single simulation for the case p = .9 and the diamond shape is very evident. In Figure 15 we show the result of a single simulation for p = .1. Now we again have a quite irregular region. We have tried again our smoothing process for values of p = .1, .2, .3 and .6. In each case, we have plotted the points that occur 20 or more times in 50 simulations. In each simulation, the growth is allowed to reach 5000 points. The results are shown in Figure 16.

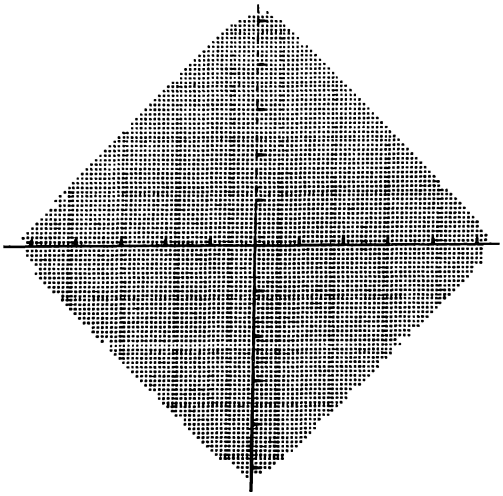


FIGURE 14

A single simulation of the *Gp* model with $p = .9$ and 5000 points.

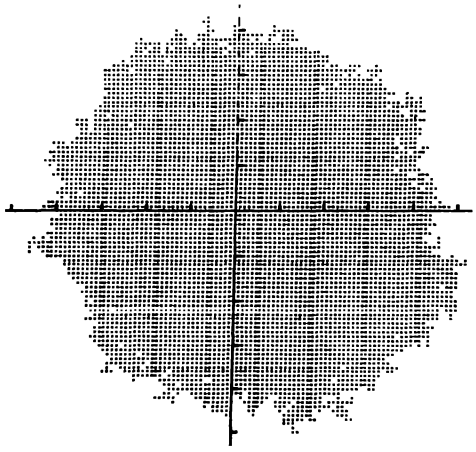
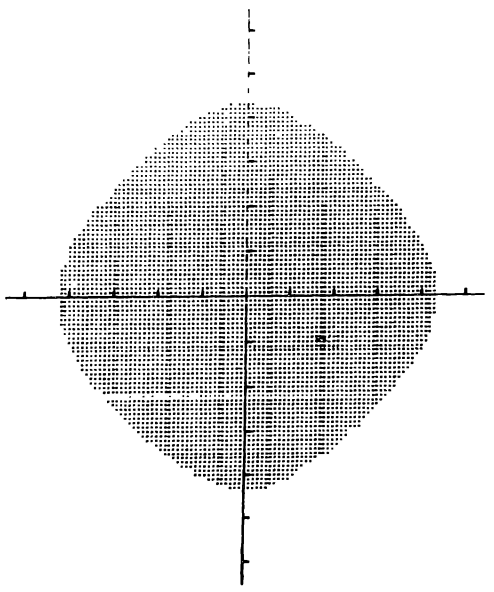


FIGURE 15

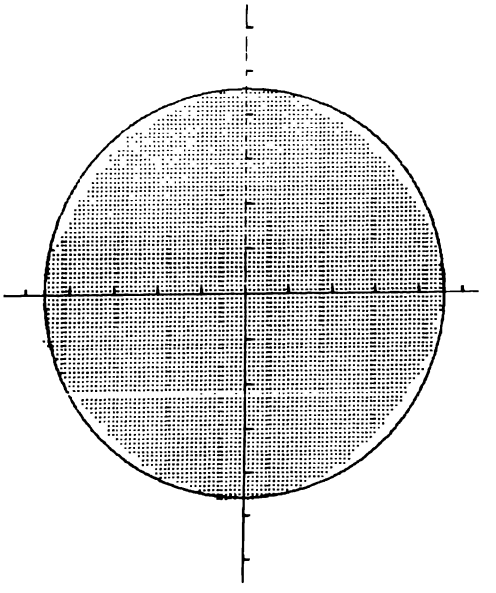
A single simulation of the *Gp* model with $p = .1$ and 5000 points.

For the lower values of p, we have tried to fit a circle. The results of Cox and Durrett and Liggett verify the first half of Richardson's conjecture and these simulations suggest that the second part may also be true. (The programs for the simulation for these growth models were written by Linda Gundal and Andrew M. Drexler.)

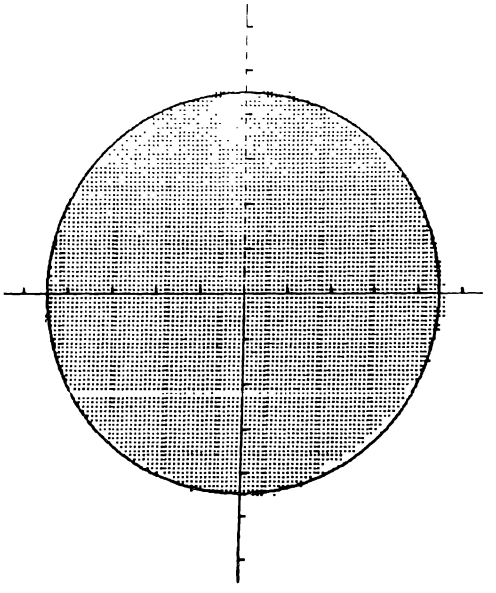
Using the fact that the Williams-Bjerknes model is the same as the biased voter model, Griffeath and Bramson (1979) have extended Richardson's results to the general Williams-Bjerknes model (the biased voter model). We have simulated a discrete



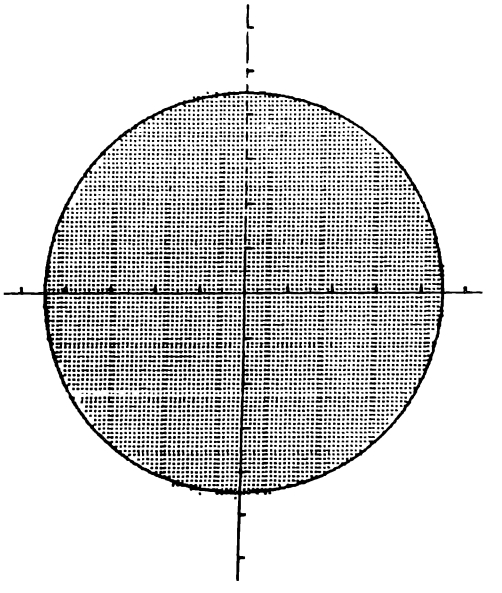
$p = .6$



$p = .3$



$p = .2$



$p = .1$

FIGURE 16

The points that occurred 20 or more times in 50 simulations of the *Gp* model for $p = .6, .3, .2, .1$. Each simulation had 5000 points.

time version of this model. Here, at each time unit, each good cell changes to a bad cell with probability $\frac{j}{4}$ where j is the number of neighbors of opposite type and each bad cell changes to a good cell with probability $k \cdot \frac{j}{4}$ with $k \leq 1$. In Figures 17,18 we show the results of simulating the discrete time version of the Williams and Bjerknes model for k = .7. As usual, the single simulation is irregular but a circle fits the smoothed simulation quite well.

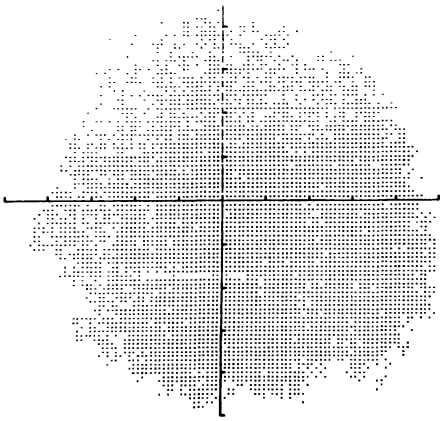


FIGURE 17

Discrete time Williams-Bjerknes model with $k = .7$ a good cell changes to a bad cell with probability $j/4$ and a bad cell changes to a good cell with probability $k(j/4)$ with j = number of neighbors of opposite type.

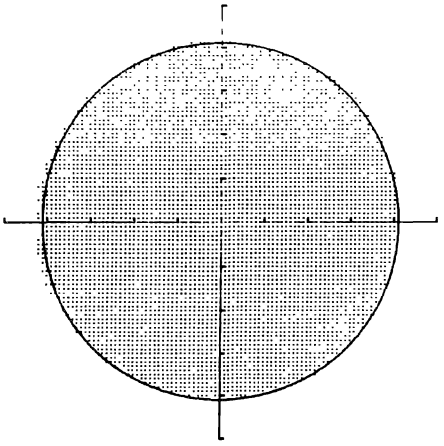


FIGURE 18

The points that occurred at least twenty times in 50 simulations of discrete time Williams-Bjerknes model with $k = .7$.

Discrete time models very similar to the above voter models have been studied by Sawyer (1976), (1979a,b) in connection with genetics models. Sawyer studied a model called the stepping stone model of Malécot (1948,1975) and Kimura and Weiss (1964) for migration in population genetics. The stepping stone model is a generalization of the well-known Fisher-Wright (Fisher 1930)(Wright 1931) model for generation fixation. This model, in its simplest form, may be described as follows. We have an urn with n balls each of which is white or red. For each generation, we choose n balls from the urn, choosing each time

with replacement. The balls chosen represent the gene types present at the next generation. The proportion $\frac{X_n}{n}$ of red balls at time n represents a finite state Markov chain with 0 and 1 as absorbing states. It is also a martingale so the process will eventually reach either 0 or 1 and the probability of reaching 1 is equal to the proportion of red balls initially in the urn. For the stepping stone model, we imagine one such urn situated on a finite number of points or colonies. In choosing the balls for the next generation, the ith colony again chooses n balls again with replacement, but now on each draw the ball is chosen from the urn at colony j with a probability p_{ij} to represent possible migration. The number of balls of each type in each of the urns again form a finite Markov chain and even if we allow more than two colors for the balls, eventually one color will win out. A simple example of the stepping stone model that is interesting to simulate is the following. Assume that, at each lattice point on an nxn square, we have a single colored ball with k different colors possible. For the next generation, each ball is replaced by a random choice from the five balls consisting of this ball and those at the four neighboring points. (Assume a periodic boundary.) Simulating this model on your home computer with color graphics makes an interesting set of patterns and, with a random initial choice of colors, takes a very long time before one color wins out.

In Sawyer's model there are a denumerable number of different species of an animal and these animals occupy a denumerable number of colonies. During each generation the population undergoes random mating, and all individuals of the new generation replace the old and independently migrate to other colonies. The motion of a representative group of N animals in

each colony is studied. We again illustrate this model in terms of a special case. We assume that colonies are located on the integer lattice points. We also assume that there is only one element in each colony. At each time unit, the animals migrate to one of the two neighboring points. We observe the species of one animal in each colony. Then this animal at time n+1 comes, with equal probability, from each of the two neighboring points. Thus the probability that it is of species j is equal to the proportion of animals of this species at neighboring points at time n. This corresponds to the voter model where the voter can have a denumerable number of positions on the issue at hand.

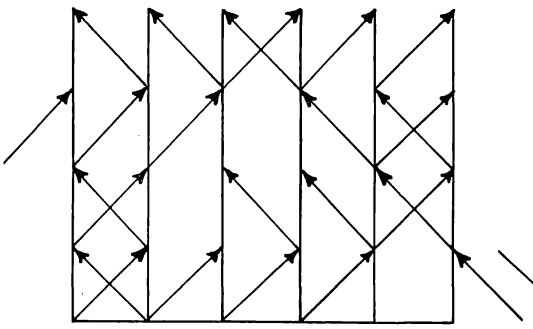


FIGURE 19

A typical percolation structure for stepping stone model (discrete time voter model)

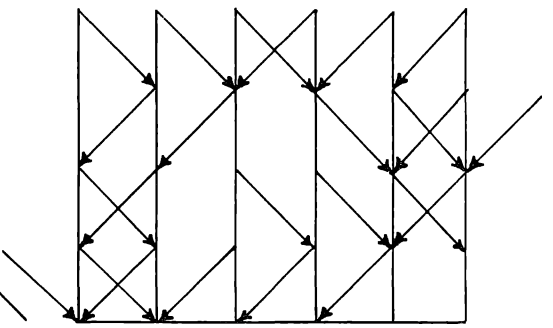


FIGURE 20

A percolation structure for the dual of the stepping stone model. (dual of Figure 19).

The graphical structure for this discrete time process can be described as follows: at each time n and for each point a, we draw an arrow either from (n-1,a) to (n,a) or from (n-1,a-1) to (n-1,a). These two possibilities occur with equal probability. A typical graph is shown in Figure 19. For our percolation interpretation, we can think of the point being wet initially with different colored fluid. The fluid then moves up

and along the arrows as before, coloring the points it passes through. The probability that a particular point a is of color j (species j) at time n is the probability that in our random graph there is a path from a point colored j at time 0 to the point (n,a). The dual structure is again obtained by reversing all the arrows. The dual of the graph in Figure 19 is shown in Figure 20.

The dual process again corresponds to coalescing random walk. The walks now can be identified by their colors. It follows, just as in the voter model, that if we start with any initial colors, the probability that a particular group of points will be all the same color tends to one as n tends to infinity. This result will be true in one and two dimensions but not in three dimensions. On the other hand, any one particular color will eventually die out. To see this, let X_n be the number of points of color j at time n. Then each such point colors 0, 1, or 2 points on the next generation with probabilities 1/4, 1/2, 1/4 respectively. Thus the expected number of particles that it colors is one. Thus X_n is a martingale. Since it is non-negative, it converges, with probability one, to a finite limit. Since it is integer valued, this limit must be 0. Thus, while any finite set of points will with high probability be the same color, this color will keep changing. A discussion of these and many other interesting properties of this model can be found in the papers of Sawyer.

As our next example of a dynamic process, we return to the dynamic Ising model. Recall that we used such a model to generate typical configurations of a Gibbs measure on a finite lattice. If $p(x) = \frac{1}{Z} e^{-bU(x)}$ was our Gibbs measure, we formed a reversible discrete time Markov chain with states the possible

configurations x. We needed only to require that the transition matrix $P(x,\bar{x})$ satisfy

$$p(x)P(x,\bar{x}) = p(\bar{x})P(\bar{x},x) .$$

For our simulations, we accomplished this by first choosing at random a spin to consider changing, letting $\bar{x}$ be the configuration x with this spin flipped. If $U(\bar{x}) \leq U(x)$, we flipped it. If $U(\bar{x}) > U(x)$, it was flipped with probability $e^{-b(U(\bar{x})-U(x))}$. A more symmetric choice would be to choose a spin at random and flip it resulting in $\bar{x}$ with probability

$$\frac{e^{-bU(\bar{x})}}{e^{-bU(\bar{x})} + e^{-bU(x)}} = \frac{1}{1 + e^{b(U(\bar{x})-U(x))}} .$$

Note that in both cases the probability that the chosen site is changed depends only on its state and that of its neighbors. Thus our Markov chain process differs from our lamp problems only in the fact that we choose only one point at a time to change. If we were to have them all attempt to change with these probabilities, we would find that $p(x)$ is no longer an invariant measure. The trouble is that we would be basing our transition probabilities on energy changes which would not be the true energy change if two neighboring points were to change. On the other hand, if we consider a continuous time process, we can achieve our lamp format. For these processes we need

$$p(x)Q(x,\bar{x}) = p(\bar{x})Q(\bar{x},x)$$

where Q is the infinitesimal matrix. We can achieve this if, following our first model, we assume that each spin will flip at a rate

$$c_a(x) = 1 \quad \text{if} \quad U(\bar{x}) \leq U(x)$$
$$= e^{-b(U(\bar{x})-U(x))} \quad \text{otherwise}$$

or following our second model if

$$c_a(x) = \frac{1}{1 + e^{b(U(\bar{x})-U(x))}} .$$

We need now only extend these finite continuous time chains to processes on infinite configurations. We can no longer use our graphical technique and so the construction of the infinite process is more difficult. There are a number of ways to proceed. The simplest to explain is that carried out by Dobrushin (1971). He obtained the infinite process as a limit of finite processes associated with a given potential U. He showed the following. Starting with any infinite configuration x, let A be a square centered at 0, and let x_A be x restricted to A. Consider a finite Markov chain determined by the potential U as described above. Here x restricted to the boundary of A gives boundary values. The boundary values remain fixed. Let $a_1, a_2, \ldots, a_r$ be points in A. Consider, for fixed t, the cylinder probabilities

$$P_{x_A}(x^t_{a_1} = +, x^t_{a_2} = +, \ldots, x^t_{a_r} = +) ,$$

where x_A is the initial configuration. Dobrushin showed that as A is increased to the whole space, these probabilities coverge to a limit

$$P_x(x^t_{a_1} = +, x^t_{a_2} = +, \ldots, x^t_{a_r} = +)$$

which depends only on the initial choice x and not on the choice of the boundary values on the boxes. This defines, for each t, a measure on the cylinder sets and hence on the set of all infinite configurations. This in turn determines a family

P_x^t of measures which represents the continuous time Markov chain associated with the given potential U.

While the construction of the Ising model is more difficult, the study of this model has been greatly enhanced by the understanding of the theory of Gibbs measure. In constructing P_+, we used the "attractive" nature of the Ising model when $b > 0$. In particular, we said that if we were trying to find the probability of a + at 0 for given boundary points, this probability is only increased by increasing the number of +'s on the boundary. The same reasoning suggests that in the dynamic model, if we consider the cylinder probability for all +'s at a finite number of sites

$$(1) \quad P_x(x^t_{a_1} = +, x^t_{a_2} = +, \ldots, x^t_{a_r} = +)$$

this probability should be a maximum when the x is the configuration of all +'s. Further, this probability should decrease to a limit as $t \longrightarrow \infty$. This would show that the measures after time t would converge to a limiting measure as $t \longrightarrow \infty$ if we start with x having all +'s. Similarly if we start with all - spins, we obtain a limiting measure. These measures will be Gibbs measures for U. Assume now that there is only one such Gibbs measure. Then these two measures will be the same. For any other initial configurations, the probability (1) will fall between those for "all +" and "all -" starting configurations. Hence, it will also converge to the same limit. Holley (1972) followed this approach to prove that if U determines only one infinite Gibbs measure, then the Ising model is ergodic. Dobrushin (1971) used a condition similar to that which he used to prove uniqueness of a Gibbs measure to show that if the attraction is weak enough (temperature above a certain

value $\bar{t}$), then the process is ergodic and the convergence to a limiting measure is exponential. The temperature $\bar{t}$ is well above the critical temperature. Whether or not this stronger result holds down to the critical temperature is not known. The method of duality exploited by Holley and Stroock (1976) leads to a similar result. This latter method can be discussed by percolation structures (see Griffeath (1979)).

While it is easy to check that every Gibbs measure associated with U is an invariant measure for the process, it is by no means obvious that these are the only invariant measures. An important tool in studying this problem has been the concept of free energy.

Consider any finite state continuous time Markov chain such that $m^{(t)} = mP^t \longrightarrow \bar{m}$ where $\bar{m}$ is the unique invariant measure. Then Renyi (1961) showed that one could prove convergence by the use of relative entropy with respect to $\bar{m}$ of the measure $m^{(t)}$ defined by

$$\sum_i m_i^{(t)} \ln \frac{m_i^{(t)}}{\bar{m}_i} .$$

In particular, he showed that this quantity was monotone in t and increasing to 0, indicating that the process is getting closer to the equilibrium distribution as time increases. Now if we apply this to our Ising chain, we have that

$$\sum_x m^{(t)}(x) \ln \frac{m^{(t)}(x)}{\frac{1}{Z} e^{-bU(x)}}$$

$$= \sum_x m^{(t)}(x) \ln m^{(t)}(x) + \sum_x m^{(t)}(x) bU(x) + \ln Z.$$

But the first two terms are the free energy determined by $m^{(t)}$, and we have indicated this is a minimum and equal to $-\ln Z$ when $m^{(t)}$ is the Gibbs measure. Thus, this shows that

our measure $m^{(t)}$ should converge to a Gibbs measure. Holley (1971) extended this result to infinite processes to show that if the initial measure is translation invariant, then the process will converge to a Gibbs measure. Holley and Stroock (1977) showed that in one and two dimensions, all invariant measures for the stochastic Ising model are Gibbs measures. Since all such measures are now known to be translation invariant for $b > 0$, this means that all invariant measures in one and two dimensions are translation invariant. In three dimensions, there are non-translation invariant Gibbs measures. Even the question of whether all invariant measures for the Ising model are Gibbs measures in three or more dimensions remains open.

What happens below the critical temperature? Here Holley (1974) has shown results which are associated with some aspects of phase transition. For example, suppose we have no external field and we start our Ising process with a Gibbs measure with temperature above the critical temperature. After the process has run for a time long enough to reach equilibrium, we lower the temperature to a temperature below the critical temperature. Then Holley proved that there will be a limiting measure and this limiting measure will be the measure $\frac{1}{2} P_+ + \frac{1}{2} P_-$. Next assume that we start with a temperature below the critical temperature and a positive exterior field. Then after the process has run a long time, we turn off the field. Then Holley showed that the process would converge to the distribution P^+. It will converge to P^- if we start with a negative external field and then turn it off. Holley and Stroock have established a number of other interesting results in the dynamic Ising model. A survey of their work can be found in Stroock (1978). Much of the work on dynamic models was inspired by the paper of Spitzer (1970). See Durrett (1980) for a more complete review.

5. THE TREE MODEL

The Ising model is by far the most important example of a Markov random field. However, as we have stated before it is very difficult to prove some of the basic results. An example which is mathematically much easier to understand is obtained when one takes the basic graph to be a tree rather than the set of all lattice points. This model is often called the Cayley tree model. Our discussion of the model will be based on the work of Preston (1973), Spitzer (1974), and Moore and Snell (1979). The model actually had its origins in the early work of Bethe (1935) as an approximation to the Ising model.

As the name suggests, the basic graph for the Cayley tree model can be shown by a tree diagram. An example is shown in Figure 1.

FIGURE 1

The root r is called the 0th level. From the root we have q branches (q = 2 in our example). The points at the ends of these branches are called the points of the 1st level. From each point of this level there are q new branches whose endpoints constitute the 2nd level, etc. We shall carry out the discussion for the case q = 2. The case q = 1 is the same as the one-dimensional Markov chain which we have already studied.

The basic results for $q \geq 2$ are essentially the same as for $q = 2$. As in the Ising model, we shall start with a finite graph (tree) and then go to the infinite graph (tree) as a limiting case. A configuration ω on a tree of n levels will be an assignment of a $+$ or a $-$ to each point. The nth level will constitute the boundary. We define the energy $U(\omega)$ just as in the Ising model. That is,

$$(1) \quad U(\omega) = -J \sum_{i,j} \sigma_i(\omega)\sigma_j(\omega) - mH\Sigma\sigma_i(\omega)$$

where, as usual, the sum in the first term is taken over nearest neighbors and counts each pair only once. The pairs including boundary points are counted. The second sum may be taken over only the interior points when the boundary is fixed. The resulting Gibbs measure is then defined by

$$P(\omega) = \frac{e^{-\frac{1}{kT}U(\omega)}}{Z} .$$

We shall again make the substitution $b = \frac{J}{kT}$ and $h = \frac{mH}{kT}$ and write our Gibbs measure as

$$P(\omega) = \frac{1}{Z} e^{-bn_0(\omega)+hM(\omega)}$$

where $n_0(\omega)$ is the number of odd bonds in ω and $M(\omega)$ is the magnetization.

When all boundary points are fixed as $+$ we have the positive boundary and when they are fixed as $-$ we have the negative boundary. The free boundary corresponds to treating the boundary points like all other points and, in this case, the second sum in (1) is over all points.

We begin by considering the first question for the Ising model: will the positive and negative boundaries give different

probabilities for a + at the root as the tree grows indefinitely?

In looking at the Cayley tree with n levels, we divide the partition function $Z^{(n)}$ into two sums

$$Z^{(n)} = Z_-^{(n)} + Z_+^{(n)}$$

where

$$Z_-^{(n)} = \sum_{\omega_r=-} e^{-bn_0(\omega)+hM(\omega)}$$

and

$$Z_+^{(n)} = \sum_{\omega_r=+} e^{-bn_0(\omega)+hM(\omega)} .$$

Thus $Z_-^{(n)}$ sums over all configurations which assign - to the root, and $Z_+^{(n)}$ sums over all configurations which assign +. We now compute the ratio of the probability of a - at the root to the probability of a + at the root. This ratio, of course, determines the probability of a + at the root since the sum of the two probabilities is 1. We define

$$u_n = \frac{Z_-^{(n)}}{Z_+^{(n)}} .$$

If we can find the limit of u_n as n tends to infinity, we will find the limiting ratio for the probability of a - to the probability of a + at the root.

We consider the possibilities for the first level of our simple tree with a + at the root. There are now only three essentially different possibilities as shown in Figure 2.

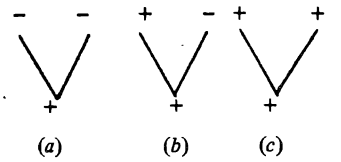


FIGURE 2

The cases (a) and (c) can occur in 1 way and (b) in 2 ways. Thus we can write

$$z_+^{(n)} = e^{-2b+h}(z_-^{(n-1)})^2 + 2e^{-b+h}\, z_-^{(n-1)} z_+^{(n-1)} + e^{h}(z_+^{(n-1)})^2$$

$$= e^{h}(e^{-b}\, z_-^{(n-1)} + z_+^{(n-1)})^2 .$$

Similarly,

$$z_-^{(n)} = e^{-h}(z_-^{(n-1)} + e^{-b}\, z_+^{(n-1)})^2 .$$

Thus

$$(3) \qquad \frac{P^{(n)}(\omega_r=-)}{P^{(n)}(\omega_r=+)} = \frac{z_-^{(n)}}{z_+^{(n)}}$$

$$= e^{-2h}\,\frac{(z_-^{(n-1)} + e^{-b}z_+^{(n-1)})^2}{(e^{-b}z_-^{(n-1)} + z_+^{(n-1)})^2}$$

Then from (3) we have

$$u_n = \frac{(u_{n-1} + e^{-b})^2}{e^{2h}(e^{-b}u_{n-1}+1)^2} .$$

The value u_1 may be obtained by considering a tree of height 1. With positive boundary

$$u_1 = \frac{P(\overset{+}{-}\ +)}{P(\overset{+}{+}\ +)} = e^{-2b-2h} .$$

Similarly, with negative boundary

$$u_1 = e^{2b-2h} .$$

We can then compute u_n for $n > 1$ by using the fact that

$$u_n = f(u_{n-1})$$

where

$$f(x) = \frac{(x+e^{-b})^2}{e^{2h}(e^{-b}x+1)^2} .$$

For the free boundary, we can start with a one point tree and obtain

$$u_0 = \frac{P(-)}{P(+)} = e^{-2h} .$$

We then ask if the sequence u_n a tends to limit u, and, if so, does this limit depend upon the value of u_1?

It is not hard to show by simple calculus arguments that in the attractive case $(b > 0)$, u_n has a limit. We note that if there is a u such that $u_n \longrightarrow u$, then we must have that $u = f(u)$, i.e., u is a fixed point of f. This follows from the fact that f is continuous and $u_n = f(u_{n-1})$. Again by standard calculus arguments, one can show that for $0 < b < \ln 9/2$, $f(x)$ has only one fixed point, and the values of u_n converge to this fixed point. If $b > \ln 9/2$, then there will be a interval of h values, including $h = 0$ such that $f(x) = x$ has three fixed points x_1, x_2, x_3.

In certain special cases, two of these points coincide. It is easy to show graphically (see Figure 3) that if $0 < u_0 < x_1$, u_n will be monotone increasing to x_1, the smallest fixed point.

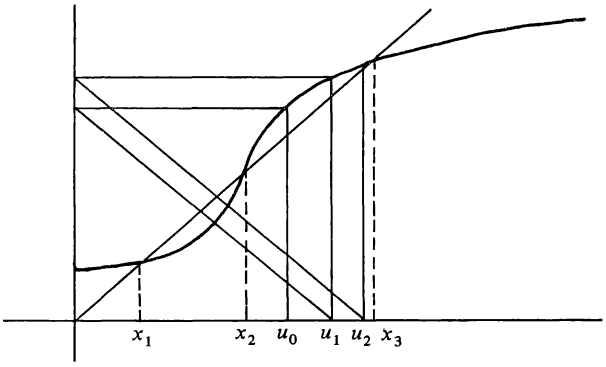


FIGURE 3

For values $x_1 < u_0 < x_2$, u_n will decrease monotonically to x_1. For $x_2 < u_0 < x_3$ they will increase monotonically to x_3. Finally if $x_3 < u_0$ the values of u_n decrease monotonically to x_3. The smallest fixed point gives the limiting probability ratio for the positive boundary. The largest fixed point gives the limiting probability ratio for the negative boundary. It would be nice if the middle value corresponded to the free boundary. This is not true in general, but it is true for the case of no external field. In this case $u_0 = 1$. We need only check that $u_0 = 1$ is a solution of $x = f(x)$ when $h = 0$, and this is easily seen to be true.

If, in the case of the free boundary, we impose a small positive exterior field, the ratio of the probability of a - to that of a + at the root will approach the lower fixed point. If we impose a small negative field it will approach the higher fixed point. These measures correspond to the positive and negative boundaries respectively. Thus we again have the effect of spontaneous magnetization as in the Ising model.

Note that, in contrast to the Ising model, it is not necessary to have the exterior field 0 to have a phase

transition, but, just as in the case of the Ising model, we obtain more than one limiting value for the probability of a + at the root if b is sufficiently large (temperature sufficiently low). For values of b between 0 and ln 9 there is a unique limit, corresponding to having no phase transition for temperatures sufficiently high.

The limits of the sequences u_n have been proven to exist. From (3) we see that this proves that $P^{(n)}(\omega_r = -)$ tends to a limit as n tends to infinity for a Cayley tree. From this point it would be natural to prove that there is a limiting probability for any specified configuration on a fixed finite subgraph. This would serve as the probability of a cylinder set for the infinite tree and these probabilities would serve as a basis for a measure on the set of all infinite configurations on the infinite tree. One can prove by limiting processes similar to that which we have employed that these limiting probabilities do exist and that the cylinder measures do not depend upon the location of the starting point. Such a measure is called by Spitzer homogeneous. This property plays the role of stationarity in the lattice models. Furthermore, these limiting measures (at most three) have a very simple form. They belong to the class of measures on the tree constructed as follows:

Choose a transition matrix

$$P = \begin{array}{c} \\ - \\ + \end{array} \begin{array}{c} \begin{array}{cc} - & + \end{array} \\ \begin{pmatrix} s & 1-s \\ 1-t & t \end{pmatrix} \end{array}$$

with fixed vector w. For a finite tree we define a measure as follows. We form a product starting at the root with w_r. Then for each branch going out the tree we multiply by the appropriate factor p_{ab}, for example:

$$P\left(\begin{array}{ccccccc} + & & + & & - & & + \\ & + & & & & - & \\ & & & - & & & \end{array}\right) = w_-(1-s)stts(1-s)$$

$$= w_-(1-s)^2s^2t^2.$$

The resulting measure is easily checked to be a Markov field. It is also easy to see that the measures obtained on the finite trees are consistent and form measures for cylinder sets, determining a measure on the infinite tree. This measure does not depend upon the particular choice of a point for the root and hence is homogeneous.

To identify the Markov chain that goes with a specific limiting measure, say P_+, it is only necessary to calculate the ratios of cylinder set probabilities

$$\frac{P_+\left(\begin{array}{ccc} + & & - \\ & + & \end{array}\right)}{P_+\left(\begin{array}{ccc} + & & + \\ & + & \end{array}\right)} \quad \text{and} \quad \frac{P_+\left(\begin{array}{ccc} + & & - \\ & - & \end{array}\right)}{P_+\left(\begin{array}{ccc} + & & + \\ & - & \end{array}\right)}.$$

The first ratio is $\frac{1-t}{t}$ and the second $\frac{s}{1-s}$. From these s and t, and hence P, are determined.

We shall call measures constructed by P Markov measures. The measures corresponding to P_+ and P_- represent pure phases in the sense that they cannot be written as a convex combination of other measures. The existence of these two measures occurs below the critical temperature T_c. When there is no external field the third Markov chain, which we shall call the middle chain, is obtained from a free boundary condition. In contrast to the Ising model in two dimensions, it is not a convex combination of P_+ and P_-. On the other hand, it also does not necessarily represent a pure phase. We will now consider some of the properties of these Markov measures. We follow here Moore

and Snell (1979).

We observe first that there is no long-range order, that is,

$$P(\omega_m = 1 | \omega_n = 1) \longrightarrow w_1$$

and

$$P(\omega_m = 1 | \omega_n = 0) \longrightarrow w_1 ,$$

as m goes further out on the tree from n. This follows from the fact that a Markov measure P, restricted to any path in the tree isomorphic to the integers, is an ordinary 2-state Markov chain with transition matrix P.

The limit theorems for the Markov measures are best studied as examples of multitype branching processes. We consider first the "law of large numbers" for a Markov measure. For a tree of height n there are $1+2+2^2+\ldots+2^n = 2^{n+1}-1$ points. Then the law of large numbers would state that the average number of + spins in n levels,

$$\frac{S^{(n)}(\omega)}{2^{n+1}} ,$$

converges to a constant. We shall show that this is true by using a well-known theorem from the theory of multitype branching processes. We shall only need to use a two-type branching process and so shall describe the result for this case.

We are given a particle which is one of two types, a or b. If it is of either type it can produce, for the next generation, particles of both types. We denote by $P_a(i,j)$ the probability that a particle of type a produces i particles of type a and j particles of type b. $P_b(i,j)$ is defined similarly, but starting with a particle of type b. The particles produced by the initial particle represent the first

generation. Each particle in the first generation independently produces particles according to the distributions $P_a(i,j)$ and $P_b(i,j)$. These particles form the second generation. This process continues through n generations. The behavior of the process depends very much on the expected number of each type of particle produced by a given type. These numbers are

$$m_{aa} = \sum_{i,j} i\, P_a(i,j)$$

$$m_{ab} = \sum_{i,j} j\, P_a(i,j)$$

$$m_{ba} = \sum_{i,j} i\, P_b(i,j)$$

$$m_{bb} = \sum_{i,j} j\, P_b(i,j).$$

Let M be the matrix

$$M = \begin{pmatrix} m_{aa} & m_{ab} \\ m_{ba} & m_{bb} \end{pmatrix}.$$

We assume that M is a strictly positive matrix. Then there is a largest positive eigenvalue e with eigenvectors u and v such that

$$uM = eu$$

$$Mv = ev.$$

Here u is a positive row vector and v is a positive column vector. Now let $X_a^{(n)}$ be the total number of particles of type a in the nth generation and $X_b^{(n)}$ be the total number of particles of type b in this generation. A basic theorem then states that if $e > 1$ the random vector

$$\frac{X^{(n)}}{e^n} = \left(\frac{X_a^{(n)}}{e^n}, \frac{X_b^{(n)}}{e^n} \right)$$

converges to a vector in the direction of the vector u. The sum

$$\frac{v_a X_a^{(n)} + v_b X_b^{(n)}}{e^n}$$

converges to a random variable with expected value uv. A proof of these results may be found in Harris (1963).

We shall now apply this theorem to our Markov measures. A particle will be a point in the tree. Thus each particle produces exactly two particles. The distribution $P_0(i,j)$ is

$$P_0(0,2) = P_{01}^2$$

$$P_0(1,1) = 2P_{00}P_{01}$$

$$P_0(2,0) = P_{00}^2.$$

Thus

$$m_{00} = 2P_{00}P_{01} + 2P_{00}^2 = 2P_{00}$$

$$m_{01} = 2P_{00}P_{01} + 2P_{01}^2 = 2P_{01}.$$

Similarly

$$m_{10} = 2P_{10}$$

$$m_{11} = 2P_{11}$$

and

$$M = 2\begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix}.$$

The largest eigenvalue for M is $e = 2$ with eigenvectors $w = (w_0, w_1)$ and $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Now $X_-^{(n)}$ will be the number of -'s on the nth level and $X_+^{(n)}$ the number of +'s on the nth level. The branching theorem then states that

$$X^{(n)} = \left(\frac{X_-^{(n)}}{2^n}, \frac{X_+^{(n)}}{2^n} \right)$$

converges to a vector in the direction $w = (w_0, w_1)$. Since

$$\frac{X_-^{(n)} + X_+^{(n)}}{2^n} = 1$$

this limiting vector must in fact be w.

This proves that for a Markov measure the proportion of +'s on the top of the tree approaches the component w_1 of the fixed vector w for the Markov chain matrix P that determines the measure. We are interested in the number S_n of +'s over the entire tree. The fact that if

$$\frac{X_+^{(n)}}{2^n}$$

converges, then

$$\frac{S_+^{(n)}}{2^{n+1}}$$

converges follows from standard results on summing series.

If we let $Z^{(n)} = X_+^{(n)} - X_-^{(n)} = M_n$, then M_n is the total magnetization at the top of the tree. In this case if r is the second largest eigenvalue for M, then if $r^2 < 2$, limit theorems of Kesten and Stigum (1966) imply that

$$\frac{M_n - E(M_n)}{2^{n/2}}$$

converges in distribution to a normal law. For $r^2 = 2$, we instead will have convergence of

$$\frac{M_n - E(M_n)}{(2n)^{n/2}}$$

to a normal law. In the case $r^2 > 2$,

$$\frac{M_n - E(M_n)}{r^n}$$

converges to a limiting random variable which depends upon the starting state and the transition matrix P. The phenomenon

observed here leads to the existence of a critical temperature T_c at which the central limit theorem breaks down (noting that the relationship between r and T is monotone). Computer calculations suggest that for $T < T_c$, M_n has limiting distribution which is bimodal.

If we analyze this Markov chain as we did when considering the Ising model we are led to a critical temperature $\bar{T}_c$ below which it is possible to have two different measures P_+ and P_-, with the same local characteristics. It is interesting to note that these two critical temperatures are not the same ($\bar{T}_c$ is slightly higher than T_c). The measure corresponding to the middle Markov chain is quite different from the other measures, for example, the case $r^2 > 2$ described above will only occur in dealing with this measure. Müller-Hartmann and Zittartz (1974) have in fact shown that for this measure there is a sequence of temperatures $T_c = T_2 < T_3 < T_4 < \ldots$ at which qualitative changes take place. See Moore and Snell (1979) for a more detailed treatment of this subject.

For the repulsive case $b < 0$, the equation $x = f(x)$ has only one fixed point. This means that any homogeneous boundary conditions will lead to the same Markov measure. Spitzer (1975) showed that if this Markov measure is determined by

$$P = \begin{pmatrix} s & 1-s \\ 1-t & t \end{pmatrix},$$

then providing that $s + t \geq 3/2$, there will be non-homogeneous measures with the same local characteristics determined by a pair of 2-state Markov chain matrices P and $\bar{P}$ with the property that

$$(4) \quad w_a P_{ab} = \bar{w}_b \bar{P}_{ba}$$

for any a, b. Call the points on levels 0,2,4,... even points and those on the levels 1,3,5,... odd points. The measure is then assigned as before but using P for transitions from even to odd points and $\bar{P}$ for transitions from odd to even points. Condition (4) is necessary to insure that the measure does not depend upon the starting point.

6. ADDITIONAL APPLICATIONS

It has been mentioned previously that we believe that Markov random fields should enjoy a wide range of applicability. We will now conclude by giving a few examples of various ways in which these processes have already been applied. One important point to keep in mind throughout the following discussion is the equivalence of Markov and Gibbs measures. Thus if we wish our measure on a lattice or graph to have the Markov property, we are required to choose a Gibbs type measure. The models we shall consider may be assumed to be on an arbitrary graph unless otherwise stated.

We will first consider an application to economics due to Hans Föllmer (1973). The problem he considers is one introduced by W. Hildenbrand (1971) in which the stabilization of an economy composed of independent agents is examined. The independence of the agents leads to the existence of a law of large numbers, thus making stabilization possible. Föllmer allows interaction between the agents thereby leading to a more natural situation. We shall now describe his model.

He considers a countably infinite set A of economic agents (as in physical examples the number of agents considered is often quite large and so we consider the infinite case to approximate such situations), each being in a state s specified by his preferences and resources. He then allows interaction between the different agents in the following manner.

First, the environment e of the economic agent a is a configuration on A-{a} which specifies the states of the other agents. The collection of local (microeconomic) characteristics of the form $r_a(s|e)$ can then be given as the conditional

probability that a is in the state s given the environment e. Then any probability measure Ω which possesses the local characteristics given by r is called a global phase of the economy.

A price is next defined as a vector $p = (p_1,\ldots,p_k)$. Then, based on some maximization scheme using the agent's preferences, a well-defined individual excess demand $d(\omega(a),p)$ is determined. The individual's excess demand is to be thought of as the difference between his demand and what he already has. The price p is said to stabilize the global phase Ω of the economy E if

$$(1)\quad \lim \frac{1}{|A_n|} \sum_{a\in A_n} d(\omega(a),p) = 0$$

whenever A_n is an increasing sequence of finite subsets of A which exhausts A. Equation (1) is interpreted as having the per capita excess demand going to 0. The questions then considered are

(a) Whether a given phase of the economy can be stabilized.

(b) Whether stabilization can be based on the microeconomic data, i.e., local characteristics and states of agents, irrespective of the phase which may prevail.

In order to consider these questions the Markov assumption, that is, $r_a(.|e) = r_a(.|e')$ whenever e and e' agree on N(a), is introduced.

What is needed here is an ergodic theorem stating that every pure phase on a general graph is ergodic, i.e., partial sums of the averages defined by (1) converge as n goes to infinity. Föllmer observes that in the case of the k-dimensional lattice

this is indeed the case, and thus any pure phase can be stabilized. The problem then is reduced to one of when there is a phase transition, which we know does not occur for $k = 1$ or for values of the interaction parameter less than b_c. Föllmer shows that for the Ising model, if there are two pure phases there is no price which makes (1) true. Since this is true even for the Ising model, we would also expect that it is not always possible to stabilize the economy in more general models.

Various types of applications of Markov fields are to be found in sociology, where they are used to describe polarization phenomena in society, as in Weidlich (1971). The simplest setup here is that people can take one of two basic stands on an issue, thus we could think of them as having one of two "spins", as we have remarked before, where there are parameters J_{ij} representing the effect of interaction on the people and there is an external field parameter h (e.g., degree to which one attitude is preferred by all individuals). In the most general formulation, the energy of a configuration would be given by

$$U(\omega) = -\sum_{i,j} J_{ij}\, \sigma_i(\omega)\sigma_j(\omega) - h \sum_i \sigma_i(\omega)$$

which would allow interactions of various strengths between different people. More generalized potentials such as these are discussed by Dobrushin (1970). However, it is frequently reasonable to make the assumption that $J_{ij} = J$ for all i,j in which case the probability of a configuration would be given by

$$P(\omega) = Z^{-1} e^{-\frac{1}{T} U(\omega)},$$

where U would now be defined as in Section 1. The system is allowed to change by a dynamic method as discussed in Section 4. The parameter T is a "climate parameter" describing whether the

society is of a more liberal or totalitarian nature. A totalitarian system would correspond to temperature that is below the critical value and thus would admit phase transition, producing the phenomenon that one attitude by chance happens to take hold and then takes over the whole society with only a few individuals left to oppose it. Similarly the liberal system in which thinking is much more independent, i.e., less "long range order", would correspond to high temperature, and the critical temperature, a point in between the two, would produce a highly polarized society with large clusters of people with opposing attitudes.

Various generalizations could be introduced into the system, for instance: a graph rather than lattice type structure, individual field parameters (prejudices) h_i, more than two possible attitudes on a subject, or a tradition effect taking into account past states as well. Results on such models could then be obtained by computer simulation. Note that the last generalization need no longer be Markovian.

Along similar lines, we shall next describe a few models of voting behavior. Here the parameters have analagous interpretations to those of the preceding discussion. The following three models have been proposed:

(a) Martin and May (1970). Each voter is initially assumed to have probability 1/2 of voting "yes" or "no", i.e., no external field. Then the relative probability that a voter votes "yes" is increased multiplicatively by the factor $e^{D/2}$ for each neighbor who votes "yes" and decreased by the factor $e^{-D/2}$ for each neighbor who votes "no".

(b) Smukler (1971). A dynamic model. Assume that we are given an interaction at time t. At time t+1 we choose a voter

at random. This voter will then be of the same opinion that the majority of his neighbors (including the voter himself) had at time t with probability $1 - x$, where $0 \leq x \leq 1$ is given. Note that as long as we adhere to the lattice interpretation there always is a majority.

(c) Kindermann (1973). A model in which at time $t+1$ each voter says "yes" with probability

$$\frac{m^a}{m^a+m^b}$$

where m is a constant and a is the number of his neighbors which said "yes" at time t and b is the number which said "no".

A modification of (a) has also been examined (Martin and May) in which all neighbors are allowed to affect a given voter's probability of voting "yes" but this turns out not to be a Markov field. The one-dimensional periodic boundary case has been examined for each of the above models with the following results: model (a) determines a nearest neighbor Gibbs measure and thus is a Markov field, model (b) is a case of a "birth and death" process and so is a time-reversible Markov process with limiting distribution of the number of "yes" votes being essentially binomial as $x \to \frac{1}{2}$ and essentially "all or none" as $x \to 0$, and model (c) is an example of a synchronous process, i.e., a process in which each voter makes his choice independent of what choices the others are making at that time. Most studies of these processes have been restricted by assuming that the decision for a site to change depends only on the values at its neighbors and that we want a reversible process. In this case Dawson (1974) has proven that the limiting measure, while it need

not be a Markov random field, will have the property that its local characteristics depend on sites at most two steps away. That is, for the limiting measure

$$P(\omega_a = r|\omega_b,\ b \in T\text{-}a)$$
$$= P(\omega_a = r|\omega_b,\ b \in N_a^{(2)})$$

where $N_a^{(2)}$ is the set of all points which can be reached from a by following one or two edges in the graph. Such a measure is called a 2-Markov field. Therefore the invariant measure for example (c) is a 2-Markov field. It is interesting to note that in model (c) the 2-Gibbs potential depends only on bonds between neighbors separated by one and not upon nearest neighbors.

Another sociological application of Markov random fields comes out of work done by Holland and Leinhardt (1977) and further developed by Wasserman (1977). They use probability measures and stochastic processes to study social networks, considering a directed graph G, where the vertices (a,b,...) represent individuals. The directed edges between the vertices specify relationships between the individuals, for instance, a—>b may mean "a influences b", "a admires b", "a knows b", etc. An example of such a graph is shown in Figure 1.

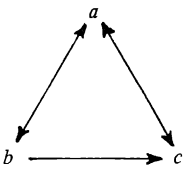


FIGURE 1

Each such graph corresponds to an adjacency matrix which, for our example, is

$$X = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}.$$

The problem of finding an appropriate probability measure for the set X of all possible graphs or adjacency matrices x is then considered. The probability measure

$$p_1(x) = (1/c)e^{v_1 n_m(x) + \Sigma_a v_a n_{a+}(x) + \Sigma_a \bar{v}_a n_{+a}(x)}$$

is proposed, where $n_m(x)$ is the number of mutual edges, $n_{a+}(x)$ is the number of edges coming out of point a, $n_{+a}(x)$ is the number of edges coming into a, c is a normalizing constant, and v_1, v_a, and $\bar{v}_a$ are parameters. This measure can be shown to specify a Markov random field (see Kindermann and Snell (1979)). Dynamic models in which the networks change in time have also been considered and can be analyzed using the voter model of Section 4.

People, however, are not the only creatures to which theories of this type can be applied. Callen and Shapero (1974) apply the Ising model to fish aligned in schools as they swim. Here the states can be two directions, e.g., north and south, in which the fish can swim. In this model the "temperature" parameter would affect the amount of order or disorder in the system, disorder corresponding to high temperature. Near the critical value response to a change in the "field" (for example, water turbulence) would be the greatest. It is amazing to note that if one actually watches fish swimming in schools, the whole school tends to change direction with amazing rapidity. Similar

phenomena have also been observed in the synchrony of fireflies as they flash on and off in swarms.

We have already, in Section 4, mentioned a number of cell growth models and a neuron firing model. Another biological model deals with diffusion across membranes. This model deals with the transport of neutral molecules between two baths separated by a symmetrical membrane. The membrane consists of units which may be in one of two states, one in which the unit is susceptible to reaction with the molecules, and the other in which it is not. There are also cooperative effects which arise from interactions between the units which introduce a dependence on the conformation of these units. Thus we see another case in which the idea of random fields can be used. Hill and Chen (1970) discuss some of the dynamic aspects of this model. Other references may also be found in this article.

We shall conclude our discussion of applications with two other models. The first, of which we have previously discussed a special case, is known as a contact process and might serve as a model for an epidemic. Here a site is either "infected" or "susceptible". If it is susceptible at time t, then it has probability $c_k dt + O(dt)$ of being infected at time $t + dt$ if k of its nearest neighbors were infected at time t, where the c_k's are constants. If a site is infected at time t, it has probability $ddt + O(dt)$ of being removed and replaced by a new susceptible element by time $t + dt$, where d is a constant. A reasonable application for such a model would be the spreading of insect pests or disease in a tree orchard, and the equilibrium measure would represent a distribution on what the final state of the orchard will be after the disease has had sufficient time to progress. A detailed study of processes of this kind has been

made by Harris (1974).

The second model is due to Holley and Liggett (1975) and is a bit more complicated. It is referred to as a proximity process. Here the set of sites I is countably infinite or finite and each site may take on the values 0 or 1. To each site i is assigned a collection of sets $\{N_{i,k}\}$, $k \geq 0$, where $N_{i,0} = \emptyset$, and a probability distribution $f_i(k)$, $k \geq 0$ on the nonnegative integers. Then the probability p_i of a 1 at time t+1 at site i is given by

$$p_i = \sum_{k \in D} f_i(k)$$

where the set D includes 0 as well as all integers k such that at time t the set $N_{i,k}$ contained at least one element which was a 1. The interpretation of such a model would be that the site considers various subsets of its environment and for each of these subsets which contains a 1, a certain quantity, $f_i(k)$, is added to its probability of being a 1 at time t+1. At time t+1 all sites then change to their new states, each making the choices independently of others, thus creating a synchronous process. Such a model might again depict a type of contagion process in which a member is more likely to catch the disease from certain members than from others.

Of course in order to use any of these models one would have to do some experimentation with the parameters to see which parameter values best describe reality. Often this may be rather difficult, depending on how easy it is to separate various features of the system. For instance, in the sociological context, it might be rather hard to separate the influences of government from interactive influences. Of course, the observation of general characteristics of the system will be a

great aid in limiting the range of parameters that needs to be considered. Once a workable model has been obtained, then the results of Sections 1-5 can be applied, or where they are not applicable, computer simulation (see Sections 3 and 4) can be employed to examine interesting properties of the system.

APPENDIX 1

In this appendix we shall discuss certain basic inequalities and show how they apply to prove that in the Ising model the measures P_+ and P_- are well defined and stationary.

We are going to use a dynamic model similar to the one introduced in Section 3. We shall, however, consider a slightly more general situation. We shall assume that we have a finite graph and configurations on the graph consist of the assignment of one of the outcomes $0,1,\ldots,n$ to each of the points. We also assume that on the space Ω of all possible configurations we have a Markov random field $P(\omega)$.

As in Section 3 we shall form a Markov chain in which the state space is Ω. We shall allow a transition from ω to $\bar{\omega}$ only if $\omega = \bar{\omega}$ or if they differ at only one site t and at this site $|\omega_t - \bar{\omega}_t| = 1$. When we pick a particular site we shall use $\omega =$ (i e) to represent the configuration with state i at the site chosen and configuration e at the other sites. We can then write the transition probabilities as R(i e,i+1 e),R(i e,i-1 e), or R(i e,i e) corresponding to the possible changes or no change at all. Recall that we must define R in such a way that

$$(1)\quad P(\omega)R(\omega,\bar{\omega}) = P(\bar{\omega})R(\bar{\omega},\omega)$$

whenever transition from ω to $\bar{\omega}$ is possible.

There are many ways to achieve this. One method is the following: Define

$$R(i\ e,i+1\ e) = c \qquad \text{for } i = 0,1,2,\ldots,n-1.$$

$$R(i\ e,i-1\ e) = c\,\frac{P(i-1\ e)}{P(i\ e)} \quad \text{for } i = 1,2,\ldots,n.$$

If neither transition takes place the state remains the

same. The constant c is chosen to make the sum of the transition probabilities from a given state less than or equal to 1. The resulting Markov chain is ergodic and condition (1) is satisfied since

$$P(i\ e)R(i\ e,i+1\ e) = c\ P(i\ e)$$
$$P(i+1\ e)R(i+1\ e,i\ e) = P(i+1\ e)\ c\ \frac{P(i\ e)}{P(i+1\ e)}\ .$$

We now use the Markov chain determined above to prove certain inequalities basic to the study of random fields. To do this, we shall need to form a new type of Markov chain called a coupled chain. We start with a space $\Omega = \{\omega\}$ of configurations on a graph. We say that $\omega \geq \bar{\omega}$ if $\omega_t \geq \bar{\omega}_t$ for all t. We form a new state space U consisting of the set of all pairs $(\omega,\bar{\omega})$ where ω and $\bar{\omega}$ are in Ω and $\omega \geq \bar{\omega}$. The elements of U will be the state space for the coupled chain. We are going to allow transitions between states only when they have a site at which ω and $\bar{\omega}$ have values which differ by at most one. We allow a transition by changing such a site according to the following transition probabilities where P_1 and P_2 are any two random fields:

$$R((i\ e,i\ \bar{e}),(i+1\ e,i+1\ \bar{e})) = c$$

$$R((i\ e,i\ \bar{e}),(i\ e,i-1\ \bar{e})) = c\ \frac{P_2(i-1\ \bar{e})}{P_2(i\ \bar{e})} - c\ \frac{P_1(i-1\ e)}{P_1(i\ e)}$$

$$R((i\ e,i\ \bar{e}),(i-1\ e,i-1\ \bar{e})) = c\ \frac{P_1(i-1\ e)}{P_1(i\ e)}$$

$$R((i\ e,i-1\ \bar{e}),(i\ e,\ i\ \bar{e})) = c$$

$$R((i\ e,i-1\ \bar{e}),(i-1\ e,i-1\ \bar{e})) = c\ \frac{P_1(i-1\ e)}{P_1(i\ e)}\ .$$

Note that we again put the site which is changed first. Note also that the environment of these sites may differ but we will have $e \geq \bar{e}$ since $\omega \geq \bar{\omega}$.

THEOREM (Holley (1974)). Let $P_1(\omega)$ and $P_2(\omega)$ be two random fields defined on Ω. Assume that

$$(2) \quad \frac{P_1(i\ e)}{P_1(i\text{-}1\ e)} \geq \frac{P_2(i\ \bar{e})}{P_2(i\text{-}1\ \bar{e})}$$

whenever $e \geq \bar{e}$. Then there is a measure P defined on the space U of all pairs $(\omega,\bar{\omega})$ with $\omega \geq \bar{\omega}$ such that

$$P_1(\omega) = \sum_{\bar{\omega}} P(\omega,\bar{\omega})$$

$$P_2(\omega) = \sum_{\bar{\omega}} P(\bar{\omega},\omega) \ .$$

Proof.

To prove the theorem, we form the above coupled Markov chain. Then this chain is an ergodic Markov chain and there is a measure $P(\omega,\bar{\omega})$ such that $P(\omega,\bar{\omega})$ represents the average time in the limit spent in the state $(\omega,\bar{\omega})$. Now watch the chain only when the first component changes. Then a transition from $\omega = (i\ e)$ to $(i+1\ e)$ can only occur when the original chain goes from a state of the form $(i\ e,\ i\ \bar{e})$ to $(i+1\ e, i+1\ \bar{e})$. This has probability c independent of $\bar{e}$. A transition from $\omega = (i\ e)$ to $(i\text{-}1\ e)$ can occur in one of two ways. The original process could be in a state $(i\ e,\ i\ \bar{e})$ and move to $(i\text{-}1\ e, i\text{-}1\ \bar{e})$ with probability $c\,\frac{P_1(i\text{-}1\ e)}{P_1(i\ e)}$ or it could be in a state $(i\ e, i\text{-}1\ \bar{e})$ and move to $(i\text{-}1\ e, i\text{-}1\ \bar{e})$ with the same probability. Thus watching only the first component we have a Markov chain with transition probabilities

$$R(i\ e, i+1\ e) = c$$

$$R(i\ e, i\text{-}1\ e) = c\,\frac{P_1(i\text{-}1\ e)}{P_1(i\ e)}$$

As we have seen above, this process has $P_1(\omega)$ as a limiting measure. Since $P_1(\omega)$ represents the proportion of time the first component is in state ω, $P_1(\omega) = \sum_{\bar{\omega}} P(\omega,\bar{\omega})$.

A similar argument shows that the original process watched only when the second component changes is a Markov chain with $P_2(\omega)$ as stationary measure. Hence $P_2(\omega) = \sum_{\bar{\omega}} P(\bar{\omega},\omega)$.

We now state two important corollaries of this theorem. If f is a function defined on a space $\Omega = \{\omega\}$ of configurations, we say that f is <u>increasing</u> if $f(\omega) \geq f(\bar{\omega})$ whenever $\omega \geq \bar{\omega}$.

<u>Corollary 1</u>. Let $P_1(\omega)$ and $P_2(\omega)$ be random fields satisfying (2). Let f be an increasing function defined on Ω. Then

$$E_1(f) \geq E_2(f).$$

Proof: Let R be constructed as in the theorem. Then

$$\begin{aligned} E_1(f) &= \sum_{\omega} f(\omega) P_1(\omega) \\ &= \sum_{\omega \geq \bar{\omega}} f(\omega) R(\omega,\bar{\omega}) \\ &\geq \sum_{\omega \geq \bar{\omega}} f(\bar{\omega}) R(\omega,\bar{\omega}) \\ &= \sum_{\bar{\omega}} f(\bar{\omega}) P_2(\bar{\omega}) \\ &= E_2(f). \end{aligned}$$

The following corollary is usually called the FKG inequalities and is due to Fortuin, Kasteleyn, and Ginibre (1971).

<u>Corollary 2</u>. Let P be a random field such that

$$\frac{P(i\ e)}{P(i\text{-}1\ e)} \geq \frac{P(i\ \bar{e})}{P(i\text{-}1\ \bar{e})}$$

when $e \geq \bar{e}$ for $i = 1,2,\ldots,n$. Then if f and g are two increasing functions on Ω,

$$E(fg) \geq E(f)\, E(g).$$

Proof: Let

$$P_1(\omega) = \frac{g(\omega)P(\omega)}{E(g)} .$$

Then if $e \geq \bar{e}$

$$\frac{P_1(i\ e)}{P_1(i\text{-}1\ e)} = \frac{g(i\ e)P(i\ e)}{g(i\text{-}1\ e)P(i\text{-}1\ e)} \geq \frac{P(i\ e)}{P(i\text{-}1\ e)} \geq \frac{P(i\ \bar{e})}{P(i\text{-}1\ \bar{e})} .$$

Thus by the Corollary 1 applied to P_1 and P

$$\sum_{\omega} \frac{f(\omega)g(\omega)P(\omega)}{E(g)} \geq \sum_{\omega} f(\omega)P(\omega)$$

or

$$E(fg) \geq E(f)\, E(g).$$

We shall now show how our inequalities can be used to prove a basic result in the theory of the Ising model. This result is the following: In the case of an attractive potential with positive boundary, $P_+^{(n)}$ converges to a limiting measure P_+. Recall that $P_+^{(n)}$ is the measure obtained from an $n \times n$ lattice with positive boundary.

Before proving this we shall prove a lemma which follows easily from the inequalities proven above.

Lemma 1. Let L be an $n \times n$ lattice. Let P and Q be two Gibbs measures assigned to L with the same local characteristics and an attractive potential. P is determined by fixing a subset R of the boundary to have value 1 and Q is

determined by fixing a subset S to have value 1, with $R \subset S$. Then for any subset A of L

$$P(\omega_a = 1,\ a \in A) \leq Q(\omega_a = 1,\ a \in A).$$

Proof: We note first that P and Q satisfy the conditions of Corollary 2.

In fact, if s is a site with k interior neighbors 1 and $\bar{k}$ boundary neighbors 1,

$$\frac{P(1,e)}{P(0,e)} = e^{v_0+v_1 k+v_1 \bar{k}}.$$

Using the measure Q this ratio will be the same or greater. It will be greater only if there is a boundary point of s in S but not in R thus increasing the value of $\bar{k}$. Recall that $v_1 \geq 0$. Let f be a set function defined on subsets of sites of L as follows:

$$f(C) = 1 \quad \text{if} \quad A \subset C$$
$$= 0 \quad \text{otherwise.}$$

Then f is an increasing function and hence by Corollary 1

$$E_P(f) \leq E_Q(f).$$

But

$$E_P(f) = P(\omega_a = 1,\ a \in A)$$
$$E_Q(f) = Q(\omega_a = 1,\ a \in A)$$

and thus the lemma is proven.

We now formulate our basic result. We have a sequence of $n \times n$ lattices centered at $0 = (0,0)$. Let $P^{(n)}$ be the sequence of measures determined by putting all 1's on the boundary (positive boundary) and all having the same local

characteristics determined by v_0, v_1, and with $v_1 \geq 0$.

Let A be a finite subset of lattice points. Then, for sufficiently large N, A will be a subset of L_n (n x n lattice) for $n \geq N$. For each $P^{(n)}$ the probabilities $P^{(n)}(\omega_A)$ are defined, where ω_A is a subconfiguration on A. Then we want to show that $P^{(n)}(\omega_A)$ tends to a limit $P_+(\omega_A)$. Specifying the outcomes on a finite subset A describes a cylinder set in the set of all infinite configurations, namely, all infinite ω which agree with ω_A. Then if the probabilities of the form $P_+(\omega_A)$ are consistent, they uniquely determine a probability measure P_+ on the infinite lattice. The consistency will follow from the fact that the $P^{(n)}$ are consistent. Our main job is to prove that $P^{(n)}(\omega_A)$ converges.

We note first that it is sufficient to prove this result for certain special ω_A, namely those which assign a 1 to each point of A. We denote these subconfigurations by 1_A. To see this consider the following two set functions f and g defined on subsets of lattice sites of L_n:

$$f(A) = P^{(n)}(\omega_A = 1_A, \omega_{L_n - A} = 0_{L_n - A})$$
$$g(A) = P^{(n)}(1_A) .$$

Then

$$g(A) = \sum_{A \subset B} f(B) .$$

Thus by the Moebius inversion formula

$$f(A) = \sum_{A \subset B} (-1)^{|B-A|} g(B) .$$

Thus the values of f(A) are determined by knowing g. But f(A) is simply the probability of a specific configuration where we specify that the 1's are on A and other sites are 0.

Thus if we can prove that $P^{(n)}(1_A)$ tends to a limit for all finite A we will have shown that $P^{(n)}(\omega_A)$ tends to a limit for finite A.

We shall prove in fact that $P^{(n)}(1_A)$ is monotone decreasing for fixed A as n tends to infinity. This will follow from the next lemma.

Lemma 2. Let L and M be two lattices which contain A and such that L is a sublattice of M. Let Q and P be the respective Gibbs measures with the same local characteristics obtained by assigning positive boundary values. Then

$$P(1_A) \leq Q(1_A) .$$

Proof:
$$\begin{aligned} P(1_A) &= \sum_{\omega_{M-L}} P(1_A, \omega_{M-L}) \\ &= \sum_{\omega_{M-L}} P(1_A|\omega_{M-L})P(\omega_{M-L}) \\ &= \sum_{\omega_{M-L}} P(1_A|\omega_B)P(\omega_{M-L}) \end{aligned}$$

where B is the boundary of L. But Q is the measure determined on L by putting 1 at each boundary point. Thus

$$P(1_A|\omega_B) \leq Q(1_A),$$

and so

$$\begin{aligned} P(1_A) &\leq \sum_{\omega_{M-L}} Q(1_A)P(\omega_{M-L}) \\ &= Q(1_A) \end{aligned}$$

as was to be proved.

Thus for any finite set A, $P^{(n)}(1_A)$ is monotone decreasing in n and tends to a limit $P_+(1_A)$. As remarked earlier this proves more generally that $P^{(n)}(\omega_A)$ tends to a limit $P_+(\omega_A)$. These limiting values represent the measure of cylinder sets for

the measure P_+. Had we put 0's on the boundaries we would have obtained the result that $P^{(n)}(1_A)$ is monotone increasing to a limit $P_-(1_A)$. This determines the measure P_-. The proof also shows that $P_- \leq P_+$. A phase transition occurs if $P_- \neq P_+$. It is clear from these constructions that both P_- and P_+ have the same local characteristics.

Next we shall prove that the limiting measure P obtained from lattices with positive boundaries is a stationary measure. For this, we need to prove that for any finite set A and lattice point $t = (t_1, t_2)$ then $P(A) = P(t+A)$ where $t + A$ are the lattice points obtained by adding t to points of A.

To prove this result we shall construct lattices alternately centered at 0 and t each containing the previous lattice (see Figure 1).

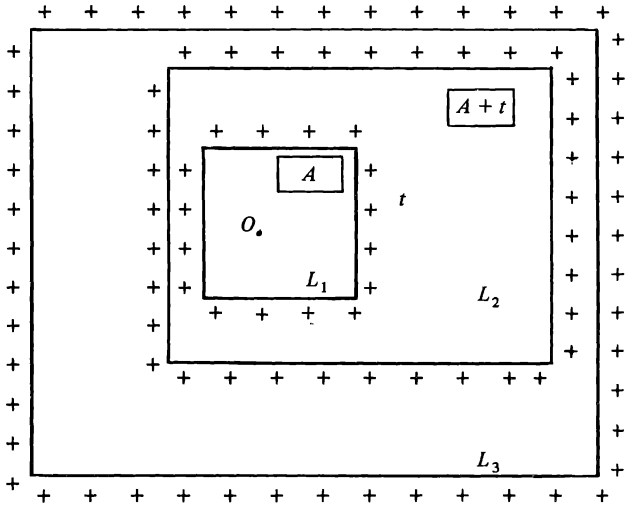


FIGURE 1

Here L_1 is an $n_1 \times n_1$ lattice centered at 0 which contains A, L_2 is an $n_2 \times n_2$ lattice centered at t large enough to include L_1 and $A + t$, and finally L_3 is an $n_3 \times n_3$ lattice centered at 0 containing L_2. Let P_1, P_2 and

P_3 be the Gibbs measures determined on L_1, L_2, L_3 with positive boundaries. Then by Lemma 2

$$P_1(l_A) \geq P_2(l_{A+t}) \geq P_3(l_A).$$

Forming a sequence of triples in this way and passing to the limit, we see that

$$P(l_A) \geq P(l_{A+t}) \geq P(l_A).$$

Hence $P(l_A) = P(l_{A+t})$ as was to be proven.

Abraham, D.B. and Martin-Löf, A. (1973): "The transfer matrix for a pure phase in the two-dimensional Ising model," Commun. Math. Phys., 32, 245-268.

Aczël, J., Forte, B., and Ng, C.T. (1974): "Why the Shannon and Hartley entropies are 'natural'," Advances in Applied Probability, 6, 131-146.

Aizenman, M. (1979): "Instability of Phase Coexistence and Translation Invariance in Two Dimensions," Physical Review Letters, Vol. 43, No. 6, 407-409.

Aizenman, M. (1980): "Translation Invariance and Instability of Phase Coexistence in the Two-Dimensional Ising System," Commun. Math. Phys. 73, 83-94.

Averintsev, M.B. (1970): "On a method of describing discrete parameter random fields," Problemy Peredachi Informatsii, 6, 100-109.

Bethe, H.A. (1935): "Statistical theory of superlattices," Proc. Roy. Soc., (London), A150, 552-575.

Bramson, M. and Griffeath, D. (1979): "Renormalizing the 3-dimensional voter model," Ann. Probability, 7, 418-432.

Bramson, M. and Griffeath, D. (1979): "On the Williams-Bjerknes tumour growth model I," to appear in Annals of Probability.

Bramson, M. and Griffeath, D.: "On the Williams-Bjerknes tumour growth model II," to appear in Cambridge Philosophical Society Mathematical Proceedings.

Brush, S. (1967): "History of the Lenz-Ising Model," Reviews of Modern Physics, 39, 883-893.

Callen, F. and Shapero, D. (1974): "A theory of social imitation," Physics Today, 27, 23-28.

Clifford, P. and Sudbury, A. (1973): "A model for spatial conflict," Biometrika, 60, 581-588.

Coniglio, A., Nappi, C.R., Peuiggi, F. and Russo, L. (1977): "Percolation points and critical point in the Ising Model," J. Physics A: Math. Gen., Vol. 10, No. 2, 205-218.

Cox, T. (1979): "The time constant of first-passage percolation on the square lattice," to appear.

Dawson, D. (1974): "Discrete Markov systems," Carleton Math. Lecture Notes No. 10, Carleton University, Ottawa.

Dobrushin, R.L. (1965): "Existence of a phase transition in two and three dimensional Ising models," Theor. Prob. Appl., 10, 193-213.

Dobrushin, R.L. (1968): "The description of a random field by means of conditional probabilities and conditions of its regularity," Theor. Prob. Appl., 13, 197-224.

Dobrushin, R.L. (1968): "The Gibbsian random fields for lattice systems with pairwise interactions," Functional Anal. Appl., 2, 292-301.

Dobrushin, R.L. (1968): "The problem of uniqueness of a Gibbsian random field and the problem of phase transitions," Functional Anal. Appl., 2, 302-312.

Dobrushin, R.L. (1970): "Prescribing a system of random variables by conditional distributions," Theor. Prob. Appl. 15, 458-486.

Dobrushin, R.L. (1971): "Markov processes with a large number of locally interacting components," Problems of Information Transmission, 7, 149-164 and 235-241.

Dobrushin, R.L. (1973): "Investigation of Gibbsian states for lattice systems," Theor. Prob. Appl., 18, 253-271.

Dobrushin, R.L., Kryukov, V.I., and Toom, A.L. (1978): Locally Interacting Systems and Their Application in Biolog Lecture Notes in Mathematics No. 653, Springer, New York.

Durrett, R. (1980): "An introduction to infinite particle systems," preprint.

Durrett, R. and Liggett, T.M. (1979): "The shape of the limit set in Richardson's growth model," to appear in Ann. Probability.

Eden, M. (1961): "A two-dimensional growth process," Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV, Berkeley, California.

Fisher, R.A. (1930): The Genetical Theory of Natural Selection, Oxford University Press.

Föllmer, H. (1974): "Random economies with many interacting systems," Journal of Math. Economics, 1, 51-62.

Fortuin, C.M., Kasteleyn, P.W., and Ginibre, J. (1971): "Correlation inequalities on some partially ordered sets," Comm. Math. Phys., 22, 89-103.

Gach, P., Kurdymov, G.L., and Levin, L.A. (1978): "One dimensional uniform arrays that wash out finite islands," Problemy Peredachi Informatsii, Vol. 14, No. 3, 92-96.

Gallavotti, G. and Jona-Lasinio, G. (1975): "Limit theorems for multidimensional Markov processes," Commun. Math. Phys., 41, 301-307.

Georgii, H.O. (1970): "Two remarks on external equilibrium states," Comm. Math. Phys., 32, 107-118.

Glauber, R.J. (1963): "Time dependent statistics of the Ising model," J. Math. Phys., 4, 294-307.

Gray, L., Wierman, J. and Smythe, R.T. (1979): "Lower bounds for the critical probability in percolation models with oriented bonds," preprint.

Griffeath, D. (1973): "Markov random fields," Unpublished manuscript.

Griffeath, D. (1975): "Ergodic theorems for graph interaction," Advances in Applied Probability, 7, 179-194.

Griffeath, D. (1976): "Introduction to random fields," Denumerable Markov Chains by Kemeny, Knapp, and Snell, (2nd edition), Springer-Verlag.

Griffeath, D. (1979): Additive and Cancellative Interacting Particle Systems, Lecture Notes in Mathematics Number 724, Springer, New York.

Griffeath, D. (1980): The Basic Contact Processes, to appear.

Griffiths, R.B. (1964): "Peierl's proof of spontaneous magnetization in two dimensional Ising ferromagnet," Phys. Rev., A136, 437-438.

Grimmett, G.R. (1973): "A theorem about random fields," Bull. London Math. Soc., 5, 81-84.

Harris, T.E. (1963): The Theory of Branching Processes, Springer.

Harris, T.E. (1974): "Contact interactions on a lattice," Ann. Probability, 2, 969-988.

Harris, T.E. (1978): "Additive set-valued Markov processes and percolation methods," Ann. Probability, 6, 355-378.

Heisenberg, W. (1928): "Zur theorie des Ferromagnetismus," Z. Physik, 49, 619-636.

Higuchi, Y. (1979): "On the absence of non-translationally invariant Gibbs states for the two-dimensional Ising model," Preprint No. 17, Universität Heidelberg.

Hildenbrand, W. (1971): "Random preferences and equilibrium analysis," Journal of Economic Theory, 3, 414-429.

Hill, T.L. and Chen, Yi-der (1970): "Cooperative effects in models of steady-state transport across membranes," I. Proc. Nat. Aca. Sci., 65, 1069-1076.

Holland, P. and Leinhardt, S. (1977): "A dynamic model for social networks," J. Math. Soc., 5, 5-20.

Holland, P. and Leinhardt, S. (1977): "Notes on the statistical analysis of social network data," preprint.

Holley, R. (1971): "Free energy in a Markovian model of a lattice spin system," Communications in Math. Physics, 23, 87-99.

Holley, R. (1972): "An ergodic theorem for interacting systems with attractive interactions," Z. Warscheinlichkeitstheorie verw. Geb., 24, 325-334.

Holley, R. (1974): "Recent results on the stochastic Ising model," Rocky Mt. J. of Math., 4, 479-496.

Holley, R. and Liggett, T.M. (1975): "Ergodic theorems for weakly interacting infinite systems and the voting model," Ann. Probability, 3, 643-663.

Holley, R. and Liggett, T.M. (1978): "The survival of contact processes," Ann. Probability, 6, 198-206.

Holley, R. and Stroock, D. (1976): "A martingale approach to infinite systems of interacting processes," Ann. Probability, 4, 195-228.

Holley, R. and Stroock, D. (1977): "In one and two dimensions, every stationary measure for a stochastic Ising model is a Gibbs state," Commun. Math. Phys., 55, 37-45.

Ising, E. (1925): "Beitrag sur Theorie des Ferromagnetismus," Zeit. fur Physik, 31, 253-258.

Jaynes, E.T. (1957): "Information theory and statistical mechanics," Phys. Rev., 106, 620-630.

Kesten, H. and Stigum, B.P. (1966): "Additional limit theorems for indecomposable multidimensional Galton-Watson processes," Ann. Math. Statist., 37, 1463-1481.

Kimura, M. and Weiss, G.H. (1964): "The stepping stone model of population structure and the decrease of genetic correlation with distance," Genetics, 49, 561-576.

Kindermann, R.P. (1973): "Random fields: theorems and examples," J. of Undergraduate Math., 5, 25-34.

Kindermann, R. and Snell, J.L. (1980): "On the relation between Markov random fields and social networks," J. Math. Soc., 7, No. 1, 1-13.

Lanford, O.E. and Ruelle, D. (1969): "Observables at infinity and states with short range correlations in statistical mechanics," Commun. Math. Phys., 13, 194-215.

Lebowitz, J.L. and Martin-Löf, A. (1972): "On the uniqueness of the equilibrium state for Ising spin systems," Comm. Math Phys., 25, 276-282.

Lenz, W. (1920): "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern," Zeit. Physik, 21, 613-615.

Liggett, T.M. (1977): "The stochastic evolution of infinite systems of interacting particles," Lecture Notes in Mathematics, 598, 187-248, Springer-Verlag, New York.

Malécot, G. (1948): The Mathematics of Heredity, English translation, W.H. Freeman, San Francisco, 1969.

Malécot, G. (1975): "Hetrozygosity and relationship in regularly subdivided populations," Theoret. Population Biology, 8, 212-241.

Malysev, V.A. (1975): "The central limit theorem for Gibbsian random fields," Soviet Math. Dokl., 16, 1141-1145.

Malysev, V.A. (1979): "Soliton sectors in continuous time lattice models," Functional Analysis and its Applications, 13, 23-32.

Martin, B.L. and May, R.M. (1970): "Non-linear models for voting behavior," Unpublished manuscript.

Martin-Löf, A. (1973): "Mixing properties, differentiability of the free energy and the central limit theorem for a pure phase in the Ising model at low temperature," Comm. Math. Phys., 32, 75-92.

McCoy, B.M. and Wu, T.T. (1973): The Two-Dimensional Ising Model, Harvard Press, Cambridge, Massachusetts.

Messager, A. and Miracle-Sole, S. (1975): "Equilibrium states of the two-dimensional Ising model in the two-phase region," Comm. Math. Phys., 40, 187-296.

Montroll, E.W., Potts, R.B. and Wards, J.C. (1963): "Correlations and Spontaneous Magnetization of the Two-Dimensional Ising Model," J. Math. Phys., 4, 308-322.

Moore, T. and Snell, J.L. (1979): "A branching process showing a phase transition," J. Appl. Prob., 16, 252-260.

Müller-Hartmann, E. and Zittartz, J. (1974): "New type of phase transition," Phys. Rev. letters, 33, 893-897.

Newman, C.N. (1979): "Normal fluctuations and the FKG inequalities," preprint.

Onsager, L. (1944): "Crystal statistics, I. A two-dimensional model with an order-disorder transition," Phys. Rev., 65, 117-149.

Onsager, L. (1949): discussion, Nuovo Cimento 6, Suppl., 261.

Peierls, R.E. (1936): "On Ising's ferromagnet model," Proc. Camb. Phil. Soc., 32, 477-481.

Preston, C.J. (1974): Gibbs States on Countable Sets, Cambridge University Press.

Rényi, A. (1961): "On measures of entropy and information," Proceedings of Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1, 547-561.

Rényi, A. (1970): Probability Theory, North Holland.

Richardson, D. (1973): "Random growth in a tessellation," Proc. Cambridge Phil. Soc., 74, 515-528.

Russo, L. (1979): "The infinite cluster method in the two-dimensional Ising model," Comm. Math. Phys. 67, 251-266.

Sawyer, S. (1979): "Results for the Stepping Stone Model for Migration in Population Genetics," Ann. Probability, 4, 699-728.

Sawyer, S. (1979): "A limit theorem for patch size in a selectively-neutral migration model," J. App. Prob. 16, 482-495.

Schwartz, D. (1977): "Applications of duality to a class of Markov processes," Ann. Probability, 5, 522-532.

Shnirman, N.G. (1968): "On the ergodicity of a certain Markov chain," Problemy Kibernet., 20, 115-124.

Smukler, J. (1971): "The ballot problem with random error," Soviet Mathematics, Amer. Math. Soc., 12, 275-279.

Spitzer, F. (1970): "Interaction of Markov processes, Adv. in Math., 5, 246-290.

Spitzer, F. (1971): "Random fields and interacting particle systems," M.A.A. Summer Seminar Notes, 1971.

Spitzer, F. (1974): "Introduction aux processus de Markov a parametres dans Z_2," Lecture Notes in Mathematics, 390, Springer-Verlag.

Spitzer, F. (1975): "Markov random fields on an infinite tree," Ann. Probability, 3, 387-398.

Stavskaya, O.N. and Pyatetskii-Shapiro, I.I. (1968): "Homogeneous networks of spontaneous active elements," Problemi Cibernetiki, 20, M. Nauka, 91-106.

Stroock, D. (1978): Lectures on Infinite Interacting Systems, Kyoto University Lectures in Math., No. 11.

Sullivan, W.G. (1975): Markov Processes for Random Fields, Communications Dublin Inst., Advanced Studies A, No. 23.

Thompson, C.J. (1972): Mathematical Statistical Mechanics, The Macmillan Co., New York.

Toom, A.L. (1974): "Nonergodic multidimensional systems of automata," Problemy Peredachi Informatsii, 10, 70-79.

Tracy, C.A. (1978): "Painlevé Transcendents and Scaling Functions of the Two-Dimensional Ising Model," Nonlinear Equations in Physics and Mathematics, 221-237, A.O. Barut (ed.), D. Reidel Pub. Co., Holland.

Van Beijeren, H. (1975): "Interface sharpness in the Ising system," Commun. Math. Phys., 40, 1-6.

Vasershtein, L.N. (1969): "Markov processes over denumerable products of spaces, describing large systems of automata," Problemy Peredachi Informatsii, 5, 64-72.

Vasilyev, N.B. (1970): "Correlation equations for the stationary measure of a Markov chain," Theor. Prob. Appl., 15, 521-525.

Wasserman, S. (1977): Stochastic Models for Directed Graphs, Ph.D. thesis, Department of Statistics, Harvard University.

Weidlich, W. (1971): "The statistical description of polarization phenomena in society," Br. J. Math Statist. Psychol., 251-266.

Williams, T. and Bjerknes, R. (1972): "Stochastic model for abnormal clone spread through epithelial basal layer," Nature, 236, 19-21.

Wright, S. (1931): "Evolution in Mendelean populations," Genetics, 16, 97-159.

Wu, T.T. (1966): "Theory of Toeplitz determinants and the spin correlations of the two-dimensional Ising model I," Phys. Rev., 149, 380.

Wu, T.T. and Cheng, H. (1967): "Theory of Toeplitz determinants and the spin correlations of the two-dimensional Ising model III," Phys. Rev., 164, 719.

Wu, T.T., McCoy, B.M., Tracy, C.A. and Barouch, E. (1976): "Spin-spin correlation functions for the two-dimensional Ising model: exact theory in the scaling region," Phys. Rev., B13, 316.

Yang, C.N. (1952): "The spontaneous magnetization of the two-dimensional Ising model," Phys. Rev., 85, 808-816.