SIGNAL DETECTION THEORY

In remote sensing and communications, we are often required to decide whether a particular signal is present—or to distinguish among several possible signals—on the basis of noisy observations. For example, a radar transmits a known electromagnetic signal pulse into space and detects the presence of targets (e.g., airplanes or missiles) by the echoes which they reflect. In digital communications, a transmitter sends data symbols to a distant receiver by representing each symbol by a distinct signal. In automatic speech recognition, an electrical microphone signal is processed in order to extract a sequence of phonemes, the elementary sounds that make up spoken language. Similar problems arise in sonar, image processing, medical signal processing, automatic target recognition, radio astronomy, seismology, and many other applications.

In each of these applications, the signal received is typically distorted and corrupted by spurious interference, which complicates the task of deciding whether the desired signal is present. In particular, the received signal in radar and communication systems inevitably contains random fluctuations, or *noise,* due to the thermal motion of electrons in the receiver and ions in the atmosphere, atmospheric disturbances, electromagnetic clutter, and other sources. This noise component is inherently unpredictable; the best we can do is to describe it statistically in terms of probability distributions. In some situations, the desired signal may also be distorted in unpredictable ways—by unknown time delays, constructive and destructive interference of multiple signal reflections, and other channel impairments—and may be best modeled as a random process.

The task of the receiver is to decide whether the desired signal is present in an observation corrupted by random noise and other distortions. The mathematical framework for dealing with such problems comes from the field of *hypothesis testing,* a branch of the theory of statistical inference. In engineering circles, this is also called *detection theory* because of its early application to radar problems. Detection theory provides the basis for the design of receivers in communication and radar applications, algorithms for identifying edges and other features in images, algorithms for parsing an electrical speech signal into words, and many other applications.

In addition to deciding whether a signal is present, we often want to estimate real-valued parameters associated with the signal, such as amplitude, frequency, phase, or relative time delay. For example, once a target has been detected, a radar will typically attempt to determine its range by estimating the round-trip propagation delay of the pulse echo. Problems of this type are the province of *estimation theory,* a field of statistical inference closely related to detection theory. In essence, detection theory deals with the problem of deciding among a finite number of alternatives, whereas estimation theory seeks to approximate real-valued signal parameters.

This article provides an overview of the basic principles and selected results of signal detection theory. The subject of estimation theory is treated elsewhere in this volume (see ESTIMATION THEORY). A more complete and detailed treatment of both topics can be found in Refs. 1 and 2. In the next section, we introduce some fundamental concepts that underlie the design of optimal detection procedures. In succeeding sections, we apply these concepts to the problem of detecting signals in additive Gaussian noise. Finally, we close with a discussion of selected advanced topics in detection theory.

BASIC PRINCIPLES

Simple Hypotheses

The design of optimal receivers in radar and communications is based on principles from the theory of statistical hypothesis testing. The fundamental problem of hypothesis testing is to decide which of several possible statistical models best describes an observation Y. For simplicity, consider the problem of deciding between two models, "target present" or "target absent." Suppose the probability density function (pdf) of Y is given by $p_1(y)$ when the target is present and by $p_0(y)$ when the target is absent. The problem of deciding whether Y is best modeled by $p_0(y)$ or $p_1(y)$ can be expressed as a choice between two hypotheses:

$$\begin{aligned} H_0 &: \; Y \text{ has pdf } p_0(y) \qquad \text{(target absent)} \\ H_1 &: \; Y \text{ has pdf } p_1(y) \qquad \text{(target present)} \end{aligned} \tag{1}$$

where H_0 is often called the *null hypothesis* and H_1 is the *alternative hypothesis.* A *detector* (or *decision rule* or *hypothesis test*) is a procedure for deciding which hypothesis is true on the basis of the observation Y. More precisely, a detector is a function that assigns to each possible observation $Y = y$ a decision $d(y) = H_0$ or H_1. There are two possible ways for the detector to make an error: It may conclude that a target is present when there is none (a *false alarm*), or it may decide that no target is present when in fact there is one (a *miss*). The performance of a detector d can therefore be measured by two quantities, the *probability of false alarm* $P_F(d)$ and the *probability of a miss* $P_M(d)$. Ideally, we would like to make both error measures as small as possible; however, these are usually conflicting objectives in the sense that reducing one often increases the other. In order to determine which detector is best for a particular application, we must strike a balance between $P_F(d)$ and $P_M(d)$ which reflects the relative importance of these two types of errors.

Several methods can be used to weigh the relative importance of $P_F(d)$ and $P_M(d)$. If the prior probabilities of the hypotheses are known, say $\pi = \Pr\{H_0\} = 1 - \Pr\{H_1\}$, it is natural to seek a *minimum-probability-of-error detector*—that is, one that minimizes the average error probability:

$$\pi P_F(d) + (1-\pi)P_M(d)$$

Such detectors are appropriate in digital communication receivers where the hypotheses represent the possible transmitted data symbols and where the goal is to minimize the average number of errors that occur in a series of transmissions. More generally, when the two kinds of errors are not equally serious, we can assign a cost C_{ij} to choosing hypothesis H_i when H_j is actually true ($i, j = 0, 1$). A detector that minimizes the *average cost* (or *risk*) is called a *Bayes detector.* It sometimes happens that the prior probabilities of H_0 and H_1 are not known, in which case the Bayes and minimum-probability-of-error detectors cannot be applied. In this case, it often makes sense to choose a detector that minimizes the average cost for the worst prior probability—for example, one that minimizes

$$\max_{0\le\pi\le1} \pi P_F(d) + (1-\pi)P_M(d) = \max\{P_F(d), P_M(d)\}$$

The resulting detector is called a *minimax detector.* Finally, in other circumstances it may be difficult to assign costs or prior probabilities. In radar, for example, what is the prior probability of an incoming missile, and what numerical cost is incurred by failing to detect it? In situations like this, it seems inappropriate to weigh the relative importance of false alarms and misses in terms of numerical costs. An alternative approach is to seek a detector that makes the probability of a miss as small as possible for a given probability of false alarm:

$$\text{minimize } P_{\text{M}}(d) \quad \text{subject to} \quad P_{\text{F}}(d) \le \alpha$$

A detector of this type is called a *Neyman–Pearson detector.*

Remarkably, all of the optimal detectors mentioned above take the same general form. Each involves computing a *likelihood ratio* from the received observation

$$\Lambda(y) = \frac{p_1(y)}{p_0(y)} \tag{2}$$

and comparing it with a threshold τ. When Y is observed, the detectors choose H_1 if $\Lambda(Y) > \tau$ and choose H_0 if $\Lambda(Y) < \tau$. This detector can be expressed concisely as

$$\Lambda(Y) \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{3}$$

When $\Lambda(Y) = \tau$, minimax and Neyman–Pearson detectors may involve a random decision, such as choosing H_0 or H_1 based on the toss of a biased coin. The minimum-probability-of-error, Bayes, minimax, and Neyman–Pearson detectors mentioned earlier differ only in their choice of threshold τ and behavior on the boundary $\Lambda(Y) = \tau$.

Composite Hypotheses

Thus far we have assumed that the probability distribution of Y is known perfectly under both hypotheses. It is very common, however, for a signal to depend on parameters that are not known precisely at the detector. In radar, for example, since the distance to the target is not known at the outset, the radar pulse will experience an unknown propagation delay as it travels to the target and back. In digital communications, the phase of the carrier signal is often unknown to the receiver. In such situations, the hypothesis "target present" corresponds to a collection of possible probability distributions, rather than one. A hypothesis of this type is called a *composite hypothesis,* in contrast to a *simple hypothesis* in which Y is described by a single pdf.

Let θ denote an unknown parameter associated with the observation, and let $p_0(y|\theta)$ and $p_1(y|\theta)$ denote the conditional probability densities of Y given θ under H_0 and H_1, respectively. In some cases, it may be appropriate to model θ as a random variable with known probability densities $q_0(\theta)$ and $q_1(\theta)$ under hypothesis H_0 and H_1, respectively. In such cases, the composite hypothesis testing problem is equivalent to a simple hypothesis testing problem with probability densities

$$p_0(y) = \int_\theta p_0(y|\theta) q_0(\theta)\, d\theta, \qquad p_1(y) = \int_\theta p_1(y|\theta) q_1(\theta)\, d\theta \tag{4}$$

and the optimal detectors are again of the form shown in Eq. (3). If θ is a random variable with unknown probability densities under H_0 and H_1, we can follow a minimax-type approach and look for the detector that minimizes the worst-case average cost over all probability densities $q_0(\theta)$ and $q_1(\theta)$.

When θ cannot be modeled as a random variable, the situation is more complex. Occasionally, there exists a detector that is simultaneously optimal for all θ, in the sense that it minimizes $P_{\text{M}}(d|\theta)$ for each θ over all detectors with a given false-alarm probability, $\max_\theta P_{\text{F}}(d|\theta) \le \alpha$. A detector with this property is said to be *uniformly most powerful.* When a uniformly most powerful detector does not exist, it is natural to use an estimate $\hat{\theta}$ of the unknown parameter derived from the observation $Y = y$. The most commonly used estimates are *maximum likelihood estimates,* which are defined as the value of θ that maximizes the conditional probability density of the observation:

$$p_i(y|\hat{\theta}_i) = \max_\theta p_i(y|\theta), \qquad i = 0,\ 1$$

Substituting the maximum likelihood estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ into the likelihood ratio, we obtain the *generalized likelihood ratio* (GLR):

$$\Lambda_{\text{G}}(y) = \frac{\max_\theta p_1(y|\theta)}{\max_\theta p_0(y|\theta)}$$

Detectors based on the GLR take the same form as the likelihood ratio detector [Eq. (3)], with $\Lambda_{\text{G}}(y)$ substituted for $\Lambda(y)$.

Multiple Hypotheses and Observations

Each of the detectors described above extends in a straightforward way to a sequence of observations $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$. In this case, the hypotheses become

$$\begin{aligned} H_0 &: \ \boldsymbol{Y} \text{ has pdf } p_0(y_1, \ldots, y_n) \qquad \text{(target absent)} \\ H_1 &: \ \boldsymbol{Y} \text{ has pdf } p_1(y_1, \ldots, y_n) \qquad \text{(target present)} \end{aligned}$$

Again, the minimum-probability-of-error, Bayes, minimax, and Neyman–Pearson detectors are of the form shown in Eq. (3), where the likelihood ratio is

$$\Lambda(y_1, \ldots, y_n) = \frac{p_1(y_1, \ldots, y_n)}{p_0(y_1, \ldots, y_n)}$$

The generalized likelihood ratio detector also extends in an analogous way.

We have so far considered only detection problems involving two hypotheses. In some situations there may be more than two possible models for the observed data. For example, digital communication systems often use nonbinary signaling techniques in which one of M possible symbols is transmitted to the receiver in each unit of time. The receiver then has M hypotheses from which to choose, one corresponding to each possible transmitted symbol. The hypothesis-testing problem can then be expressed as

$$H_i : Y \text{ has pdf } p_i(y), \qquad i = 0, \ldots, M-1$$

In such situations, we are usually interested in finding a minimum-probability-of-error detector for some given prior probabilities $\pi_i = \Pr\{H_i\}$, $i = 0, \ldots, M-1$. The average probability of error for a detector d is given by

$$\sum_{i=0}^{M-1} \Pr\{d(Y) \ne H_i | H_i \text{ is true}\}\pi_i \tag{5}$$

This error probability is minimized by the *maximum a posteriori probability* (MAP) detector, which chooses the hypothesis that is most probable given the observation $Y = y$. Mathematically, the MAP detector takes the form

$$d(y) = H_i \text{ that maximizes } q(H_i|y) \tag{6}$$

where

$$q(H_i|y) = \frac{p_i(y)\pi_i}{p(y)}, \qquad p(y) = \sum_i p_i(y)\pi_i$$

is the conditional probability of hypothesis H_i given the observation $Y = y$. In digital communications, the possible transmitted symbols are often equally likely ($\pi_i = 1/M$, $i = 1, \ldots, M - 1$), in which case the MAP detector reduces to the *maximum likelihood* (ML) detector

$$d(y) = H_i, \qquad \text{where } i \text{ maximizes } p_i(y) \tag{7}$$

It is easy to check that the MAP and ML detectors reduce to likelihood ratio detectors when $M = 2$.

The results presented in this section form the basis for the design of optimal receivers for a wide variety of communications and remote sensing problems. In the following sections, we apply these results to the problem of detecting signals in noise. In the process, we obtain several of the most important and widely used receivers in communications and radar as particular instances of the likelihood ratio detector [Eq. (3)].

DETECTION OF KNOWN SIGNALS IN NOISE

We now consider the problem of detecting the presence or absence of a discrete-time signal observed in noise. A detector for this purpose is also called a *receiver* in the terminology of radar and communications. We assume for now that both the signal and the noise statistics are known precisely at the receiver, in which case the detection problem can be expressed as a choice between the simple hypotheses:

$$\begin{aligned} H_0: &\quad Y_i = N_i, & i = 1, \ldots, n \\ H_1: &\quad Y_i = s_i + N_i, & i = 1, \ldots, n \end{aligned}$$

where s_i, $i = 1, \ldots, n$ is a deterministic signal and N_i, $i = 1, \ldots, n$ is a *white Gaussian noise* sequence—that is, a sequence of independent and identically distributed (i.i.d.) Gaussian random variables with mean zero and variance $\sigma^2 > 0$. Thus, the probability densities of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ under both hypotheses are multivariate Gaussian where

$$p_1(\boldsymbol{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - s_i)^2\right\} \tag{8}$$

and $p_0(\boldsymbol{y})$ is given by the same formula, with the s_i's set to zero.

From the previous section, we know that each of the optimal detectors (minimum probability of error, Bayes, minimax, Neyman–Pearson) reduces to a likelihood ratio detector [Eq. (3)]. From Eq. (8), the likelihood ratio for this problem takes the form

$$\Lambda(\boldsymbol{y}) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(y_i - s_i)^2 - y_i^2]\right\}$$

It is easy to verify that $\Lambda(\boldsymbol{y})$ is a monotonically increasing function of the test statistic

$$\sum_{i=1}^{n} y_i s_i - \frac{1}{2}\sum_{i=1}^{n} s_i^2 \tag{9}$$

Thus, the likelihood ratio detector [Eq. (3)] can be expressed in the equivalent form

$$\sum_{i=1}^{n} Y_i S_i \underset{H_0}{\overset{H_1}{\gtrless}} \tau' \tag{10}$$

where the quadratic term in Eq. (9) has been merged with the threshold τ'. The optimal receiver thus consists of correlating the received sequence against the desired signal and comparing the result to a threshold. A receiver of this type is called a *correlation receiver.*

This receiver extends in a natural way to continuous-time detection. The correlation receiver extends in a natural way to continuous-time detection problems. A proof of this extension is nontrivial and requires generalizing the likelihood ratio to continuous-time observations (see Chapter 6 of Ref. 1 for details). Consider the signal detection problem

$$\begin{aligned} H_0: &\quad Y(t) = N(t), & 0 \le t < T \\ H_1: &\quad Y(t) = s(t) + N(t), & 0 \le t < T \end{aligned}$$

where $s(t)$ is a known deterministic signal and $N(t)$ is a *continuous-time white Gaussian noise* process with two-sided power spectral density $N_0/2$ (see KALMAN FILTERS AND OBSERVERS). The likelihood ratio is again a monotonically increasing function of a correlation statistic

$$\int_0^T y(t)s(t)\,dt - \frac{1}{2}\int_0^T s^2(t)\,dt \tag{11}$$

Merging the second term with the threshold, we again find that the likelihood ratio detector is a correlation receiver, which is illustrated in Fig. 1.

The correlation in Fig. 1 can also be expressed as a filtering operation:

$$\int_0^T y(t)s(t)\,dt = \int_{-\infty}^{\infty} h(T-t)y(t)\,dt$$

where $h(t) = s(T - t)$, $0 \le t \le T$. Here $h(t)$ can be regarded as the impulse response of a linear time-invariant filter. The frequency response of this filter is given by the Fourier transform of $h(t)$:

$$H(f) = S^*(f)e^{-2\pi jfT}$$

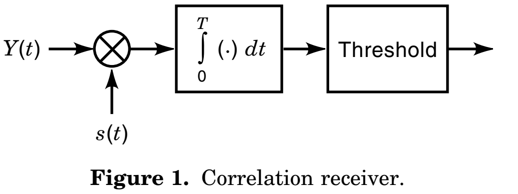

Figure 1. Correlation receiver.

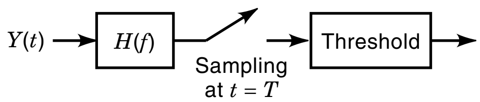

Figure 2. Matched filter receiver.

where $S^*(f)$ is the complex conjugate of the Fourier transform of $s(t)$. The correlation receiver can therefore be implemented in the form of a filter sampled at time $t = T$, as illustrated in Fig. 2.

Since the amplitude of the filter $H(f)$ matches the signal spectrum $S(f)$, this form of the detector is called a *matched-filter receiver.* The matched filter has the property that it maximizes the *signal-to-noise ratio* (the ratio of signal power to noise power) at the input to the threshold operation (see Ref. 2).

The receiver in Fig. 1 is optimal for deciding whether a known signal is present or absent in white Gaussian noise. Very similar correlation structures appear in receivers for deciding among several possible signals. For a detection problem involving M possible signals, the minimum-probability-of-error detector will compare the outputs of a bank of M correlation branches, one for each possible signal. For example, consider the problem of deciding between two equally likely ($\pi_0 = \pi_1 = \frac{1}{2}$) signals in white Gaussian noise:

$$\begin{aligned} H_0:\ & Y(t) = s_0(t) + N(t), \qquad 0 \le t < T \\ H_1:\ & Y(t) = s_1(t) + N(t), \qquad 0 \le t < T \end{aligned}$$

The receiver that minimizes the average probability of error [Eq. (5)] in this case is the maximum likelihood detector. When $Y(t) = y(t)$ is received, the ML detector chooses the hypothesis H_i such that $s(t) = s_i(t)$ maximizes the correlation statistic [Eq. (11)]. Thus, the optimal receiver consists of a correlation receiver with a branch for each possible transmitted signal, as illustrated in Fig. 3 [where E_i is the energy of $s_i(t)$].

As in the case of one signal, the correlation receiver in Fig. 3 can be implemented in the alternative form of a bank of matched filters, each sampled at $t = T$.

DETECTION OF SIGNALS WITH UNKNOWN PARAMETERS

In the preceding section, we assumed that the desired signal is known precisely at the receiver. However, this assumption

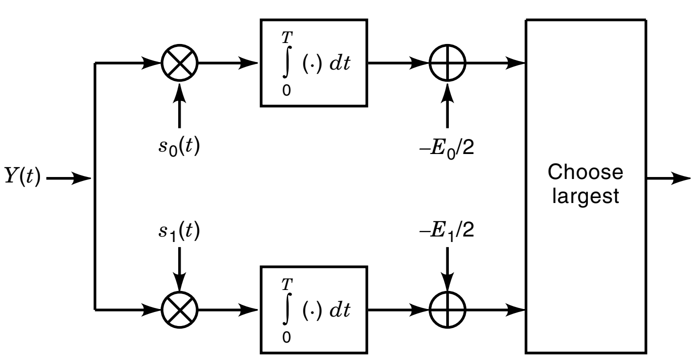

Figure 3. Correlation receiver for binary signals.

is often unrealistic in practice. Unknown path losses, Doppler shifts, and propagation delays can lead to uncertainty about the amplitude, phase, frequency, and delay of the signal. When signal parameters are unknown, the detection problem involves composite hypotheses. As discussed earlier, detection procedures for composite-hypothesis testing depend on whether the unknown parameter is modeled as random or deterministic. In this section, we consider only the example of an unknown random parameter.

Many radar and communication problems involve detection of a sinusoidal signal with an unknown phase. The phase is typically modeled as a random variable Θ, uniformly distributed on $[0, 2\pi)$. For example, consider the discrete-time binary detection problem:

$$\begin{aligned} H_0:\ & Y_i = N_i, \qquad i = 1, \cdots, n \\ H_1:\ & Y_i = A\cos(\omega i T_s + \Theta) + N_i, \qquad i = 1, \ldots, n \end{aligned}$$

where A is a known constant, T_s is the sampling interval, ω is a frequency such that $n\omega T_s$ is an integer multiple of 2π, and N_i is a discrete-time white Gaussian noise sequence which is independent of Θ. The likelihood ratio for this detection problem is given by Eqs. (3) and (4). Given $\Theta = \theta$, the conditional probability density of $\boldsymbol{Y}$ under H_1 is

$$p_1(\mathbf{y}|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - A\cos(\omega i T_s + \theta)]^2\right\}$$

and the unconditional pdf is given by

$$p_1(\mathbf{y}) = \frac{1}{2\pi}\int_0^{2\pi} p_1(\mathbf{y}|\theta)\, d\theta$$

After some manipulation, the likelihood ratio reduces to

$$\Lambda(\mathbf{y}) = \frac{p_1(\mathbf{y})}{p_0(\mathbf{y})} = \exp\left\{-\frac{A^2 n}{4\sigma^2}\right\} I_0\left(\frac{Aq}{\sigma^2}\right)$$

where

$$q^2 = \left[\sum_{i=1}^{n} y_i \cos(\omega i T_s)\right]^2 + \left[\sum_{i=1}^{n} y_i \sin(\omega i T_s)\right]^2 \qquad (12)$$

and

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp\{x\cos\theta\}\, d\theta$$

is a modified Bessel function of the first kind. Since $I_0(x)$ is symmetric in x and monotonically increasing for $x \ge 0$, the likelihood ratio is an increasing function of the quadrature statistic [Eq. (12)], and the likelihood ratio detector [Eq. (3)] can be expressed in the alternate form:

$$q^2 \underset{H_0}{\overset{H_1}{\gtrless}} \tau'$$

This detector is called a *quadrature receiver.* It consists of correlating the received signal with two phase-shifted versions of the desired signal, $\cos(\omega i T_s)$ and $\sin(\omega i T_s)$. The two correla-

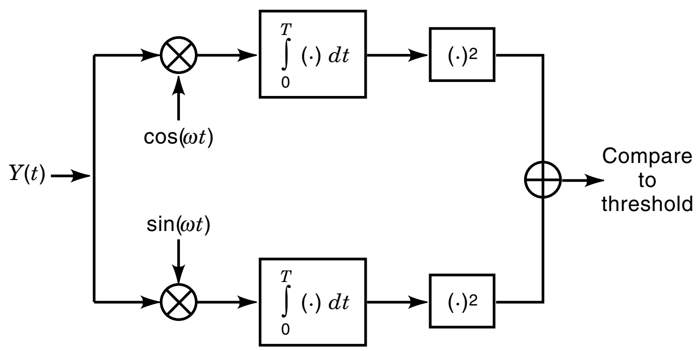

Figure 4. Quadrature receiver.

tions are then squared, summed, and compared to a threshold.

This detector extends in a straightforward way to the detection of continuous-time sinusoidal signals with random phase. Consider the detection problem

$$\begin{aligned} H_0 &: \quad Y(t) = N(t), && 0 \le t < T \\ H_1 &: \quad Y(t) = A\cos(\omega t + \Theta) + N(t), && 0 \le t < T \end{aligned}$$

where Θ is a random phase uniformly distributed on $[0, 2\pi)$, A is a constant, ω is an integer multiple of $2\pi/T$, and $N(t)$ is white Gaussian noise. The likelihood ratio detector reduces to a threshold test involving the quadrature statistic:

$$\left[\int_0^T y(t)\cos(\omega t)\,dt\right]^2 + \left[\int_0^T y(t)\sin(\omega t)\,dt\right]^2$$

The resulting continuous-time quadrature receiver is illustrated in Fig. 4.

DETECTION OF RANDOM SIGNALS

So far we have assumed the receiver knows the desired signal exactly, with the possible exception of specific parameters such as amplitude, phase, or frequency. However, sometimes the received signal may be so distorted by the channel that it must be modeled by a more complex type of random process. In certain situations, for example, the transmitted signal propagates to the receiver by many different paths due to signal reflection and scattering. In such cases, the received signal consists of many weak replicas of the original signal, called *multipath signals,* with different amplitudes and relative time delays. The superposition of these multipath signals can resemble a Gaussian random process statistically. Typical examples include channels that use ionospheric reflection or tropospheric scattering as a primary mode of propagation, and land mobile radio, where scattering and reflection by nearby ground structures can produce a similar effect.

In this section, we consider the problem of detecting signals that are described by random processes. We again begin by considering a discrete-time detection problem:

$$\begin{aligned} H_0 &: \quad Y_i = N_i, && i = 1, \ldots, n \\ H_1 &: \quad Y_i = S_i + N_i, && i = 1, \ldots, n \end{aligned}$$

where $\boldsymbol{S} = (S_1, \ldots, S_n)$ is a zero-mean Gaussian random sequence with known covariance $\boldsymbol{E}\{SS^{\mathrm{T}}\} = \Sigma$, N_i is discrete-time white Gaussian noise, $\boldsymbol{E}\{\cdot\}$ denotes the expectation, and T denotes transpose. Note that $\boldsymbol{Y}$ is a zero-mean Gaussian vector under hypotheses H_0 and H_1, with respective covariances $\sigma^2\boldsymbol{I}$ and $\sigma^2\boldsymbol{I} + \Sigma$. The likelihood ratio is then

$$\Lambda(\boldsymbol{y}) = |\boldsymbol{I} + \sigma^{-2}\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{y}^T(\sigma^2\boldsymbol{I} + \Sigma)^{-1}\boldsymbol{y} + \frac{1}{2}\boldsymbol{y}^T(\sigma^2\boldsymbol{I})^{-1}\boldsymbol{y}\right\}$$

Since this is a monotonically increasing function of the test statistic $\boldsymbol{y}^{\mathrm{T}}Q\boldsymbol{y}$, where

$$Q = \boldsymbol{I} - (\boldsymbol{I} + \sigma^{-2}\Sigma)^{-1} = \Sigma(\sigma^2\boldsymbol{I} + \Sigma)^{-1} \tag{13}$$

the likelihood ratio detector [Eq. (3)] can be expressed as

$$\boldsymbol{Y}^T Q\boldsymbol{Y} \underset{H_0}{\overset{H_1}{\gtrless}} \tau'$$

This detector is called a *quadratic receiver.* In the particular case when the desired signal is also a white noise process (i.e., $\Sigma = \alpha^2\boldsymbol{I}$), the quadratic receiver statistic $\boldsymbol{y}^{\mathrm{T}}Q\boldsymbol{y}$ is proportional to $\|\boldsymbol{y}\|^2$ and the likelihood ratio detector reduces to

$$\|\boldsymbol{Y}\|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \tau''$$

Since $\|\boldsymbol{y}\|^2$ is proportional to the average energy in the sequence $\boldsymbol{y}$, this detector is called an *energy detector* or *radiometer.*

In continuous time, the likelihood ratio detector takes a more complex but analogous form. Consider the problem of deciding among the hypotheses

$$\begin{aligned} H_0 &: \quad Y(t) = N(t), && 0 \le t < T \\ H_1 &: \quad Y(t) = S(t) + N(t), && 0 \le t < T \end{aligned}$$

where $S(t)$ is a zero-mean Gaussian noise process with autocovariance function

$$C(t, u) = E\{S(t)S(u)\}$$

and $N(t)$ is a white Gaussian noise process with one-sided power spectral density $N_0/2$. The likelihood ratio detector for this problem can also be expressed in terms of a quadratic statistic

$$\int_0^T \int_0^T Q(t, u)Y(t)Y(u)\,dt\,du$$

where $Q(t, u)$ is the solution to the integral equation

$$C(t, u) = \int_0^T Q(t, \xi)C(\xi, u)\,d\xi + \frac{N_0}{2}Q(t, u), \qquad 0 \le t, u < T$$

This equation is a continuous-time analog of Eq. (13), as can be seen by writing Eq. (13) in the alternative form $\Sigma = Q\Sigma + \sigma^2 Q$.

ADVANCED TOPICS

Detection in Colored Gaussian Noise

In the preceding sections, we assumed white Gaussian noise models for both the discrete and continuous-time detection problems. When the noise is Gaussian but not white, we can transform the detection problem into an equivalent problem involving white noise. For example, suppose we are interested in detecting a known signal $\boldsymbol{s} = (s_1, \ldots, s_n)$,

$$H_0: \quad \boldsymbol{Y} = \boldsymbol{N}$$
$$H_1: \quad \boldsymbol{Y} = \boldsymbol{s} + \boldsymbol{N}$$

where $\boldsymbol{N}$ is a zero-mean Gaussian noise vector with positive-definite covariance matrix Σ_N. Using the Cholesky decomposition (see p. 84 of Ref. 1), the noise covariance can be written in the form

$$\Sigma_N = CC^{\mathrm{T}}$$

where C is an $n \times n$ nonsingular lower-triangular matrix. Since C is invertible, it is intuitive that no information is lost by taking the observation to be $\boldsymbol{Y}' = C^{-1}\boldsymbol{Y}$ instead of $\boldsymbol{Y}$. The detection problem can then be expressed as

$$H_0: \quad \boldsymbol{Y}' = \boldsymbol{N}'$$
$$H_1: \quad \boldsymbol{Y}' = \boldsymbol{s}' + \boldsymbol{N}'$$

where $\boldsymbol{s}' = C^{-1}\boldsymbol{s}$ and $\boldsymbol{N}' = C^{-1}\boldsymbol{N}$. It is easy to verify that $\boldsymbol{N}'$ is a white Gaussian noise vector with covariance $\Sigma_{N'} = \boldsymbol{I}$; thus, the likelihood ratio detector is the correlation receiver [Eq. (10)]. Here, the overall approach is to *prewhiten* the original detection problem, by transforming it to an equivalent problem involving white noise. After prewhitening, the detection problem can be solved by the methods described in the previous sections.

A similar prewhitening procedure can be performed for continuous-time detection problems. Let $N(t)$ be a zero-mean colored Gaussian noise process with known autocovariance function $R(t, u) = E\{N(t)N(u)\}$. Under mild conditions on $R(t, u)$, there is a *whitening filter* $h(t, u)$ with the property that

$$N'(t) = \int_0^T h(t, u)N(u)\,du, \qquad 0 \le t < T$$

is a white Gaussian noise process with unit power spectral density. This filter can be used to transform a detection problem involving $N(t)$ into an equivalent problem involving white Gaussian noise.

Detection in Non-Gaussian Noise

We have thus far focused exclusively on Gaussian noise models. Gaussian processes can accurately model many important noise sources encountered in practice, such as electrical noise due to thermal agitation of electrons in the receiver electronics, radio emissions from the motion of ions in the atmosphere, and cosmic background radiation. However, other sources of noise are not well described by Gaussian distributions, such as impulsive noise due to atmospheric disturbances or radar clutter.

While the receivers presented in this article can be used in the presence of non-Gaussian noise, they are not optimal for this purpose and may perform poorly in comparison to the likelihood ratio detector [Eq. (3)] based on the actual non-Gaussian noise statistics. In contrast to the simple linear correlation operations that arise in Gaussian detection problems, optimal detectors for non-Gaussian noise often involve more complicated nonlinear operations. A thorough treatment of detection methods for i.i.d. non-Gaussian noise can be found Kassam (3). A recent survey of detection techniques for dependent non-Gaussian noise sequences is given in Poor and Thomas (4).

Nonparametric Detection

Throughout this article, we have assumed the receiver knows the probability density of the observation under each hypothesis, with the possible exception of a real-valued parameter θ. Under this assumption, the detection problem is a choice between composite hypotheses that each represents a collection of possible densities, say

$$\Omega_0 = \{p_0(y|\theta) : \theta \in \Theta_0\}, \qquad \Omega_1 = \{p_1(y|\theta) : \theta \in \Theta_1\}$$

This is called a *parametric model,* because the set of possible probability distributions under both hypotheses can be indexed by a finite number of real parameters.

In practice, however, precise models for the signal and the underlying noise statistics are frequently not available. In such cases, it is desirable to find detectors that perform well for a large class of possible probability densities. When the probability classes Ω_0 and Ω_1 are so broad that a parametric model cannot describe them, the model is said to be *nonparametric.* In general, nonparametric detection methods may be classified as robust or simply nonparametric depending on the breadth of the underlying probability classes Ω_0 and Ω_1.

In *robust detection,* the probability densities of the observation are known approximately under each hypothesis and the aim is to design detectors that perform well for small deviations from these densities. Usually, the probability classes Ω_0 and Ω_1 consist of small nonparametric neighborhoods of the nominal probability densities. One widely studied model for these neighborhoods is the *ϵ-contamination class*

$$\Omega_i = \{p(y) : p(y) = (1-\epsilon)p_i(y) + \epsilon\, h(y)\}, \qquad i = 0, 1$$

where $p_i(y)$ is the nominal probability density under hypothesis H_i, $0 \le \epsilon < 1$ is small enough so that Ω_0 and Ω_1 do not overlap, and $h(y)$ is an arbitrary probability density. In robust detection, the performance of a detector d is typically measured by worst-case performance over all probability densities in Ω_0 and Ω_1. Optimal detectors are those that yield the best worst-case performance. For ϵ-contamination models, the optimal robust detector consists of a likelihood ratio detector for the nominal probability densities that includes some type of soft-limiting operation. For example, a robust form of the correlation receiver (appropriate for small deviations from the Gaussian noise model) is obtained by replacing $Y_i s_i$ with $g(Y_i s_i)$ in Eq. (10), where g is a soft-limiter of the form

$$g(x) = \begin{cases} b & \text{if } x > b \\ x & \text{if } a < x < b \\ a & \text{if } x < a \end{cases}$$

An extensive survey of the robust detection literature prior to 1985 can be found in Kassam and Poor (5).

The term *nonparametric detection* is usually reserved for situations in which very little is known about the probability distribution of the underlying noise, except perhaps that it is symmetric and possesses a probability density. In such situations, the aim is to develop detectors that provide a guaranteed false-alarm probability over very wide classes of noise distributions. The simplest nonparametric detector is the *sign detector,* which counts the number of positive observations in a sequence and compares it to a threshold. It can be shown that this detector provides a constant false-alarm probability for detecting the presence or absence of a constant positive signal in any i.i.d. zero-median additive noise sequence. A discussion of further results in nonparametric detection may be found in Gibson and Melsa (6).

Sequential Detection

All of the discrete-time detection problems considered above involve a fixed number of observations. There are some situations, however, in which it may be advantageous to vary the number of observations. In radar systems, for example, a series of observations might correspond to repeated measurements of a weak target. Naturally, we want to detect the target as soon as possible—that is, using the fewest observations. Detection methods that permit a variable number of observations are the subject of *sequential detection.* Such methods are applicable whenever each observation carries a cost, and we want to minimize the overall average cost of making a reliable decision.

One of the most important techniques in sequential detection is a Neyman–Pearson-type test called the *sequential probability ratio test* (SPRT) (1). Suppose we want to decide between the two hypotheses

$$\begin{aligned} H_0 : &\quad Y_i \text{ is i.i.d. with pdf } p_0(y), \qquad i = 1, 2, \ldots \\ H_1 : &\quad Y_i \text{ is i.i.d. with pdf } p_1(y), \qquad i = 1, 2, \ldots \end{aligned}$$

using the smallest average number of observations necessary to achieve a probability of false alarm P_F and probability of miss P_M. The SPRT involves testing the accumulated data after each observation time $j = 1, 2, ...$. The test statistic at time j consists of the likelihood ratio of all observations up to that time, that is,

$$\Lambda_j(y_1, \ldots, y_j) = \frac{\prod_{i=1}^{j} p_1(y_i)}{\prod_{i=1}^{j} p_0(y_i)}$$

At time j, we calculate $\Lambda_j(Y_1, . . ., Y_j)$ and compare it to two thresholds, τ_0 and τ_1. If $\Lambda_j \geq \tau_1$ we decide in favor of H_1, if $\Lambda_j \leq \tau_0$ we decide in favor of H_1, otherwise we take another observation and repeat the test. The thresholds τ_0 and τ_1 are chosen to provide the desired false-alarm and miss probabilities. The SPRT minimizes the average number of observations under both H_0 and H_1, subject to constraints on P_F and P_M.

FURTHER READING

A more complete and detailed treatment of most of the topics covered in this article can be found in the books by Poor (1) and by Srinath, Rajasekaran, and Viswanathan (2). Further information on the applications of detection theory in communications and radar is contained in the books by Proakis (7) and by Nathanson (8).

Current research in signal detection and its applications is published in a wide variety of journals. Perhaps chief among these are the *IEEE Transactions on Information Theory, IEEE Transactions on Signal Processing,* and the *IEEE Transactions on Communications.*

BIBLIOGRAPHY

1. H. V. Poor, *An Introduction to Signal Detection and Estimation,* New York: Springer-Verlag, 1988.
2. S. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to Statistical Signal Processing with Applications,* Englewood Cliffs, NJ: Prentice-Hall, 1996.
3. S. A. Kassam, *Signal Detection in Non-Gaussian Noise,* New York: Springer-Verlag, 1987.
4. H. V. Poor and J. B. Thomas, Signal Detection in Dependent Non-Gaussian Noise, in H. V. Poor and J. B. Thomas, (eds.), *Advances in Statistical Signal Processing,* Vol. 2, *Signal Detection,* Greenwich, CT: JAI Press, 1993.
5. S. A. Kassam and H. V. Poor, Robust techniques for signal processing: A survey, *Proc. IEEE,* **73**: 433–481, 1985.
6. J. D. Gibson and J. L. Melsa, *Introduction to Nonparametric Detection with Applications,* New York: Academic Press, 1975.
7. J. G. Proakis, *Digital Communications,* 2nd ed., New York: McGraw-Hill, 1989.
8. F. E. Nathanson, *Radar Design Principles,* New York: McGraw-Hill, 1991.

BRIAN L. HUGHES
North Carolina State University