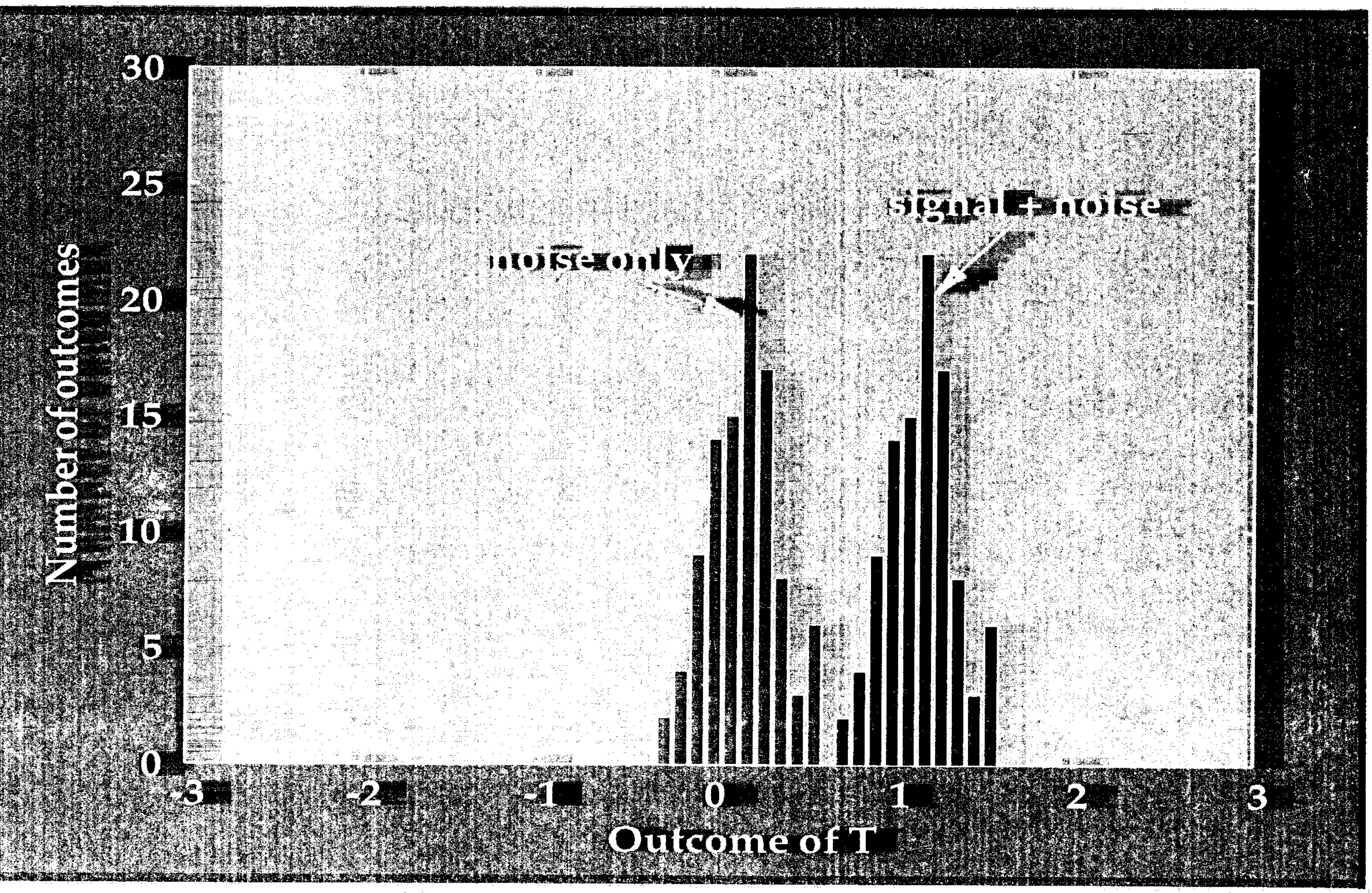
VOLUME II

FUNDAMENTALS OF STATISTICAL SIGNAL PROCESSING

DETECTION THEORY



STEVEN M. KAY

PRENTICE HALL SIGNAL PROCESSING SERIES
ALAN V. OPPENHEIM, SERIES EDITOR

Fundamentals of Statistical Signal Processing

Volume II

Detection Theory

Steven M. Kay

University of Rhode Island

Prentice Hall PTR

Upper Saddle River, New Jersey 07458

http://www.phptr.com

[redacted]

Fundamentals of Statistical Signal Processing

Volume II

Detection Theory

Steven M. Kay

University of Rhode Island

Prentice Hall PTR

Upper Saddle River, New Jersey 07458

http://www.phptr.com

Contents

Preface

This text is the second volume of a series of books addressing statistical signal processing. The first volume, *Fundamentals of Statistical Signal Processing: Estimation Theory*, was published in 1993 by Prentice-Hall, Inc. Henceforth, it will be referred to as [Kay-I 1993]. This second volume, entitled *Fundamentals of Statistical Signal Processing: Detection Theory*, is the application of statistical hypothesis testing to the detection of signals in noise. The series has been written to provide the reader with a broad introduction to the theory and application of statistical signal processing.

Hypothesis testing is a subject that is standard fare in the many books available dealing with statistics. These books range from the highly theoretical expositions written by statisticians to the more practical treatments contributed by the many users of applied statistics. This text is an attempt to strike a balance between these two extremes. The particular audience we have in mind is the community involved in the design and implementation of signal processing algorithms. As such, the primary focus is on obtaining optimal detection algorithms that may be implemented on a digital computer. The data sets are therefore assumed to be samples of a continuous-time waveform or a sequence of data points. The choice of topics reflects what we believe to be the important approaches to obtaining an optimal detector and analyzing its performance. As a consequence, some of the deeper theoretical issues have been omitted with references given instead.

It is the author's opinion that the best way to assimilate the material on detection theory is by exposure to and working with good examples. Consequently, there are numerous examples that illustrate the theory and others that apply the theory to actual detection problems of current interest. We have made extensive use of the MATLAB® scientific programming language (Version 4.2b)[1] for all computer-generated results. In some cases, actual MATLAB programs have been listed where a program was deemed to be of sufficient utility to the reader. Additionally, an abundance of homework problems has been included. They range from simple applications of the theory to extensions of the basic concepts. A solutions manual is available from the author. To aid the reader, summary sections have been provided at the beginning of each chapter. Also, an overview of all the principal detection approaches and the rationale for choosing a particular method can be found in

[1]MATLAB is a registered trademark of The MathWorks, Inc.

Chapter 11. Detection based on simple hypothesis testing is described in Chapters 3–5, while that based on composite hypothesis testing (to accommodate unknown parameters) is the subject of Chapters 6–9. Other chapters address detection in nonGaussian noise (Chapter 10), detection of model changes (Chapter 12), and extensions for complex/vector data useful in array processing (Chapter 13).

This book is an outgrowth of a one-semester graduate level course on detection theory given at the University of Rhode Island. It includes somewhat more material than can actually be covered in one semester. We typically cover most of Chapters 1–10, leaving the subjects of model change detection and complex data/vector data extensions to the student. It is also possible to combine the subjects of estimation and detection into a single semester course by a judicious choice of material from Volumes I and II. The necessary background that has been assumed is an exposure to the basic theory of digital signal processing, probability and random processes, and linear and matrix algebra. This book can also be used for self-study and so should be useful to the practicing engineer as well as the student.

The author would like to acknowledge the contributions of the many people who over the years have provided stimulating discussions of research problems, opportunities to apply the results of that research, and support for conducting research. Thanks are due to my colleagues L. Jackson, R. Kumaresan, L. Pakula, and P. Swaszek of the University of Rhode Island, and L. Scharf of the University of Colorado. Exposure to practical problems, leading to new research directions, has been provided by H. Woodsum of Sonetech, Bedford, New Hampshire, and by D. Mook and S. Lang of Sanders, a Lockheed-Martin Co., Nashua, New Hampshire. The opportunity to apply detection theory to sonar and the research support of J. Kelly of the Naval Undersea Warfare Center, J. Salisbury, formerly of the Naval Undersea Warfare Center, and D. Sheldon of the Naval Undersea Warfare Center, Newport, Rhode Island are also greatly appreciated. Thanks are due to J. Sjogren of the Air Force Office of Scientific Research, whose support has allowed the author to investigate the field of statistical signal processing. A debt of gratitude is owed to all my current and former graduate students. They have contributed to the final manuscript through many hours of pedagogical and research discussions as well as by their specific comments and questions. In particular, P. Djurić of the State University of New York proofread much of the manuscript, and S. Talwalkar of Motorola, Plantation, Florida proofread parts of the manuscript and helped with the finer points of MATLAB.

Steven M. Kay
University of Rhode Island
Kingston, RI 02881
Email: kay@ele.uri.edu

Chapter 1

Introduction

1.1 Detection Theory in Signal Processing

Modern detection theory is fundamental to the design of electronic signal processing systems for decision making and information extraction. These systems include

1. Radar
2. Communications
3. Speech
4. Sonar
5. Image processing
6. Biomedicine
7. Control
8. Seismology,

and all share the common goal of being able to decide when an event of interest occurs and then to determine more information about that event. The latter task, information extraction, is the subject of the first volume [Kay 1993]. The former problem, that of decision making, is the subject of this book and is broadly termed *detection theory*. Other names associated with it are *hypothesis testing* and *decision theory*. To illustrate the problem of detection as applied to signal processing, we briefly describe the first three of these systems.

In radar we are interested in determining the presence or absence of an approaching aircraft [Skolnik 1980]. To accomplish this task we transmit an electromagnetic pulse, which if reflected by a large moving object, will indicate the presence of an aircraft. If an aircraft is present, the received waveform will consist of the reflected pulse (at some time later) and noise due to ambient radiation and the receiver electronics. If an aircraft is not present, then only noise will be present. It is the function of the signal processor to decide whether the received waveform consists of noise only (no aircraft) or an echo in noise (aircraft present). As an example, in Figure 1.1a we have depicted a radar and in Figure 1.1b a typical received waveform for the two possible scenarios. When an echo is present, we see that the character of the received waveform is somewhat different, although possibly not by much. This is because the received echo is attenuated due to propagation loss and possibly distorted due to the interaction of multiple reflections. Of course, if the aircraft is detected, then it is of interest to determine its bearing, range, speed, etc. Hence, detection is the first task of the signal processing system while the second task is information extraction. Estimation theory provides the foundation for the second task and has already been described in Volume I [Kay-I 1993]. The optimal detector for the radar problem is the Neyman-Pearson detector, which is described in Chapter 4. A more practical detector which accommodates signal uncertainties, however, is discussed in Chapter 7.

A second application is in the design of a digital communication system. An example is the binary phase shift keyed (BPSK) system as shown in Figure 1.2a used to communicate the output of a digital data source that emits a "0" or "1" [Proakis 1989]. The data bit is first modulated, then transmitted, and at the receiver, demodulated and then detected. The modulator converts a 0 to the waveform $s_0(t) = \cos 2\pi F_0 t$ and a 1 to $s_1(t) = \cos(2\pi F_0 t + \pi) = -\cos 2\pi F_0 t$ to allow transmission through a bandpass channel whose center frequency is F_0 Hz (such as a microwave link). The phase of the sinusoid indicates whether a 0 or 1 has been sent. In this problem, the function of the detector is to decide between the two possibilities, as in the radar problem, although now, we always have a signal present – the question is *which* signal. Typical received waveforms are shown in Figure 1.2b. Since the sinusoidal carrier has been extracted by the demodulator, all that remains at the detector input is the baseband signal, either a positive or negative pulse. This signal is usually distorted due to limited channel bandwidth and is also corrupted by additive channel noise. The solution to this problem is given in Chapter 4.

Another application is in speech recognition where we wish to determine which word was spoken from among a group of possible words [Rabiner and Juang 1993]. A simple example is to discern among the digits "0", "1", ..., "9". To recognize a spoken digit using a digital computer we would need to *match* the spoken digit with some stored digit. For example, the waveforms for the spoken digits 0 and 1 are shown in Figure 1.3. They have been repeated three times by the same speaker. Note that the waveform changes slightly for each utterance of the same word. We may think of this change as "noise," although it is actually the natural variability of speech. Given an utterance, we wish to decide if it is a 0 or 1. More generally, we would need to decide among the ten possible digits. Such a problem is a generalization of that for radar and for digital communications in which only one of two possible choices need be made. The solution to this problem is discussed in Chapter 4.

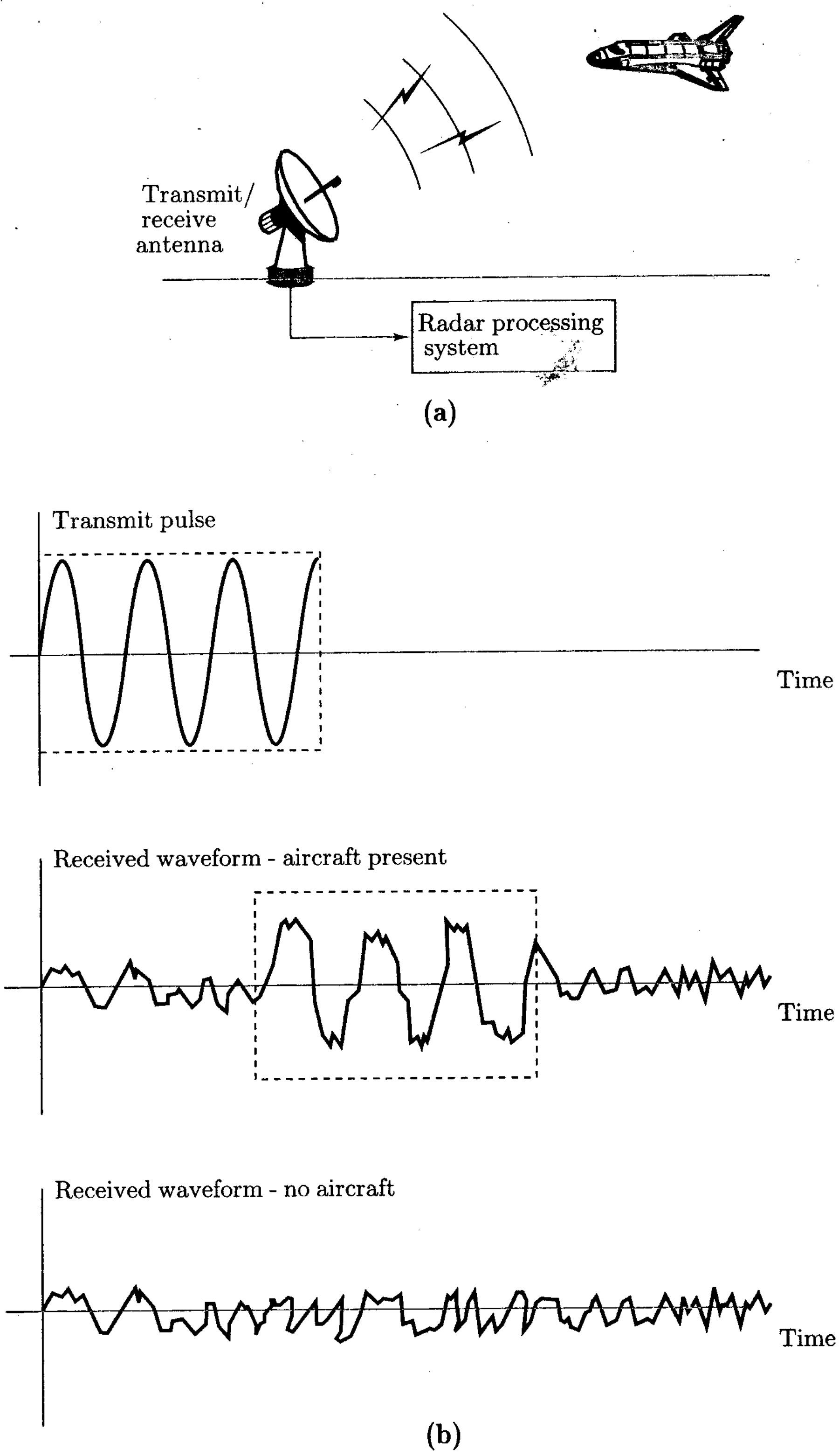


Figure 1.1. Radar system (a) Radar (b) Radar waveforms.

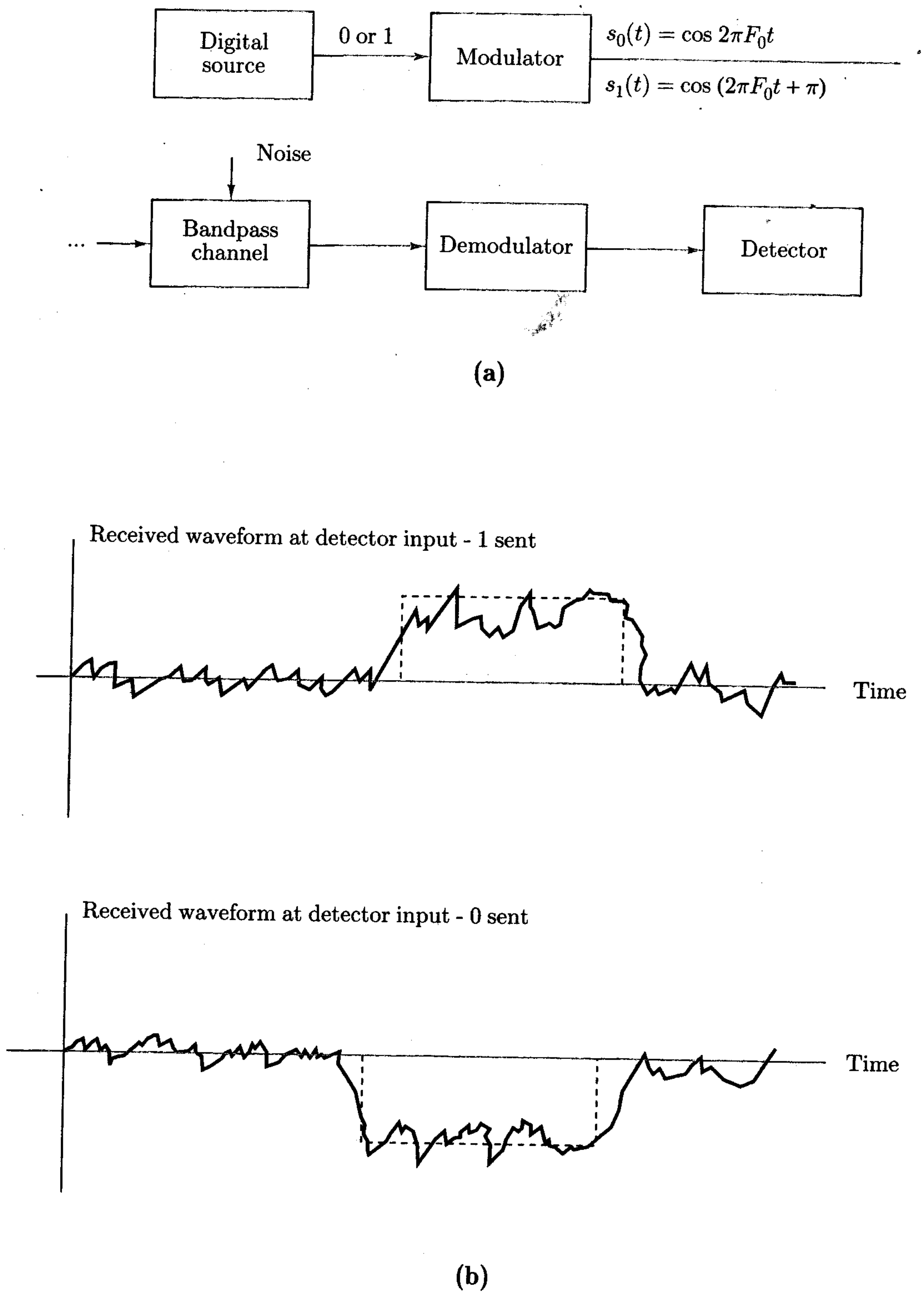
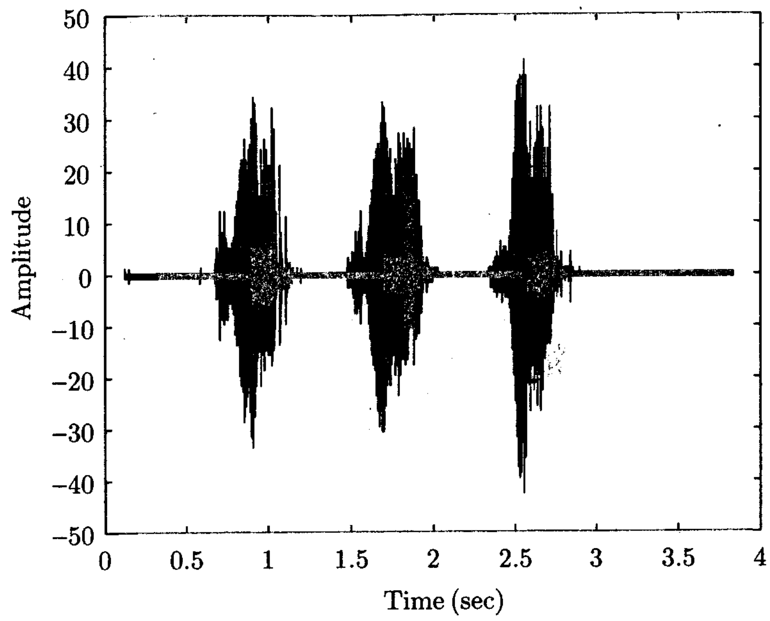
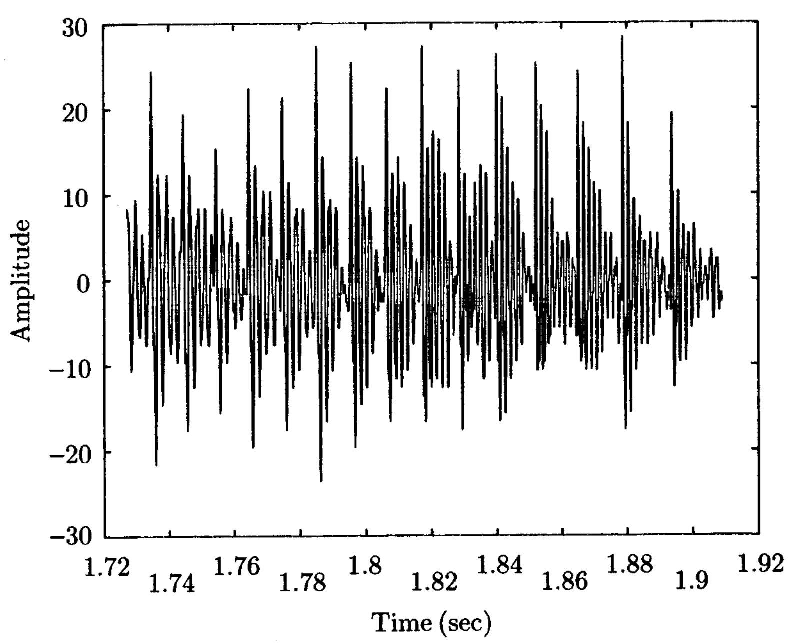


Figure 1.2. Binary phase shift keyed digital communication system (a) Basic system (b) BPSK baseband waveforms.

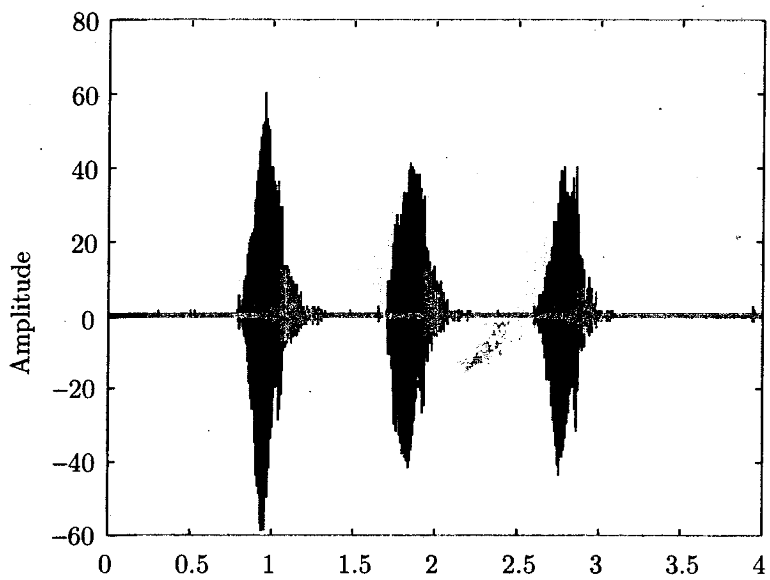


(a)

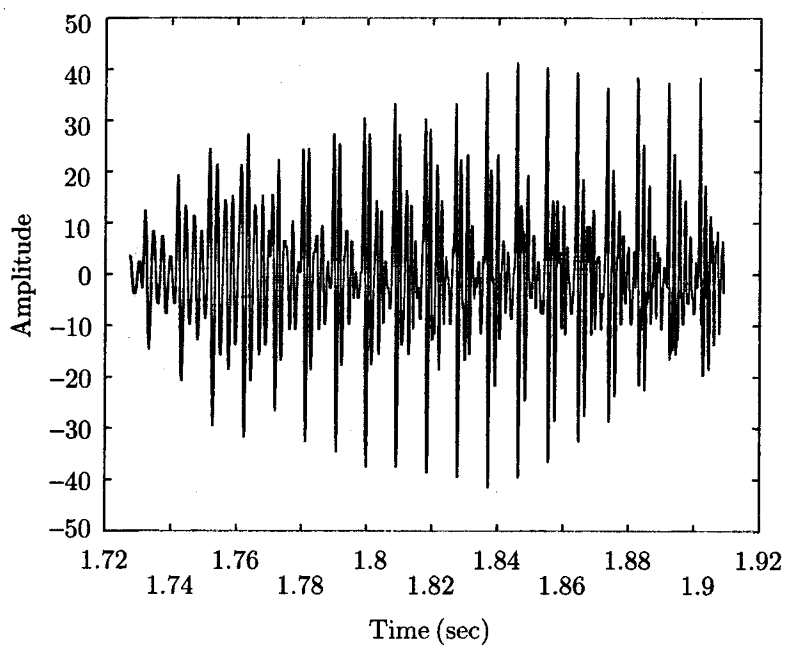


(b)

Figure 1.3. Speech waveforms for digits "zero" and "one" (a) "Zero" spoken three times (b) "Zero"–portion of utterance.



(c)



(d)

Figure 1.3. Continued (c) "One" spoken three times (d) "One"–portion of utterance.

In all of these systems, we are faced with the problem of making a decision based on a continuous-time waveform. Modern-day signal processing systems utilize digital computers to sample the continuous-time waveform and store the samples. As a result, we have the equivalent problem of making a decision based on a *discrete-time* waveform or *data set.* Mathematically, we assume the N-point data set $\{x[0], x[1], \ldots, x[N-1]\}$ is available. To arrive at a decision we first form a function of the data or $T(x[0], x[1], \ldots, x[N-1])$ and then make a decision based on its value. Determining the function T and mapping it into a decision is the central problem addressed in *detection theory.* Although electrical engineers at one time designed systems based on analog signals and analog circuits, the future trend is based on discrete-time signals or sequences and digital circuitry. With this transition the detection problem has evolved into one of making a decision based on the observation of a *time series*, which is just a discrete-time process. Therefore, our problem has now evolved into decision-making based on data, which is the subject of *statistical hypothesis testing.* All the theory and techniques developed are now at our disposal [Kendall and Stuart 1976–1979].

Before concluding our discussion of application areas, we complete the previous list.

4. Sonar - detect the presence of an enemy submarine [Knight, Pridham, and Kay 1981, Burdic 1984]

5. Image processing - detect the presence of an aircraft using infrared surveillance [Chan, Langan, and Staver 1990]

6. Biomedicine - detect the presence of a cardiac arrythmia [Gustafson et al. 1978]

7. Control - detect the occurence of an abrupt change in a system to be controlled [Willsky and Jones 1976]

8. Seismology - detect the presence of an underground oil deposit [Justice 1985]

Finally, the multitude of applications stemming from analysis of data from physical phenomena, economics, medical testing, etc., should also be mentioned [Ives 1981, Taylor 1986, Ellenberg et al. 1992].

1.2 The Detection Problem

The simplest detection problem is to decide whether a signal is present, which, as always, is embedded in noise, or if only noise is present. An example of this problem is the detection of an aircraft based on a radar return. Since we wish to decide between two possible hypotheses, signal and noise present versus noise only present, we term this the *binary hypothesis testing problem.* Our goal is to use the received data as efficiently as possible in making our decision and hopefully to be correct most of the time. A somewhat more general form of the binary hypothesis test was encountered in the communication problem. There our interest was in deciding which of two possible signals was transmitted. Our two hypotheses in this case consist of a sinusoid with phase 0° embedded in noise versus a sinusoid with phase 180° embedded in noise.

It also frequently occurs that we wish to decide among more than two hypotheses. In the speech recognition example, our goal was to determine which digit among the ten possible ones was spoken. Such a problem is referred to as the *multiple hypothesis testing problem.* Because we are essentially attempting to determine the speech pattern or to classify the spoken digit as one of a set of possible digits, it is also referred to as the *pattern recognition or classification* problem [Fukunaga 1990].

All these problems are characterized by the need to decide among two or more possible hypotheses based on an observed data set. As always, the data are inherently random in nature, with speech patterns and noise as examples, so that a statistical approach is necessary. In the next section we model the detection problem in a form that allows us to apply the theory of *statistical hypothesis testing* [Lehmann 1959].

1.3 The Mathematical Detection Problem

By way of introduction we consider the detection of a DC level of amplitude $A = 1$ embedded in white Gaussian noise $w[n]$ with variance σ^2. To simplify the discussion we assume that only one sample is available on which to base the decision. Hence, we wish to decide between the hypotheses $x[0] = w[0]$ (noise only) and $x[0] = 1 + w[0]$ (signal in noise). Since the noise is assumed to be zero mean, we might decide that a signal is present if

$$x[0] > \frac{1}{2} \tag{1.1}$$

and noise only is present if

$$x[0] < \frac{1}{2} \tag{1.2}$$

since $E(x[0]) = 0$ if noise only is present and $E(x[0]) = 1$ if a signal is present in noise. (The decision for $x[0] = 1/2$ can be made arbitrarily since the probability of this event is zero. We will henceforth omit this case.) Clearly, we will be in error whenever a signal is present and $w[0] < -1/2$ or whenever noise only is present and $w[0] > 1/2$ (see also Problem 1.1). Hence, we cannot expect to make a correct decision all the time. Hopefully, we will decide correctly most of the time. A better understanding can be obtained by considering what would happen if we repeated the experiment a number of times. This is to say that we observe $x[0]$ for 100 realizations of $w[0]$ when a signal is present and when it is not. Then, some typical results are shown in Figure 1.4a for $\sigma^2 = 0.05$. The "o"'s denote the outcomes when no signal is present and the "x"'s when a signal is present. Clearly, according to (1.1), (1.2) we may make an incorrect decision but only rarely. However, if $\sigma^2 = 0.5$, then our chances of making an error increase dramatically as shown in Figure 1.4b.

Of course, this is due to the increasing spread of the realizations of $w[0]$ as σ^2 increases. Specifically, the probability density function (PDF) of the noise is

$$p(w[0]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} w^2[0]\right). \tag{1.3}$$

This is illustrated in Figure 1.5 in which histograms of the data shown in Figure 1.4 have been plotted. The dashed plot is for noise only and the solid plot is for a signal in noise. The performance of any detector will depend upon how different the PDFs of $x[0]$ are under each hypothesis. For the same example we plot the PDFs as given by (1.3) in Figure 1.6 for $\sigma^2 = 0.05$ and $\sigma^2 = 0.5$. When noise only is present, they are

$$p(x[0]) = \begin{cases} \frac{1}{\sqrt{0.1\pi}} \exp\left(-10x^2[0]\right) & \sigma^2 = 0.05 \\ \frac{1}{\sqrt{\pi}} \exp\left(-x^2[0]\right) & \sigma^2 = 0.5 \end{cases}$$

and when a signal is embedded in noise, the PDFs are

$$p(x[0]) = \begin{cases} \frac{1}{\sqrt{0.1\pi}} \exp\left(-10(x[0]-1)^2\right) & \sigma^2 = 0.05 \\ \frac{1}{\sqrt{\pi}} \exp\left(-(x[0]-1)^2\right) & \sigma^2 = 0.5. \end{cases}$$

We will see later that the detection performance improves as the "distance" between the PDFs increases or as A^2/σ^2 (the signal-to-noise ratio (SNR)) increases. This example illustrates the basic result that the detection performance depends on the *discrimination* between the two hypotheses or equivalently between the two PDFs (see also Problems 1.2 and 1.3).

More formally, we model the previous detection problem as one of choosing between $\mathcal{H}_0$, which is termed the noise-only hypothesis, and $\mathcal{H}_1$, which is the signal-present hypothesis, or symbolically

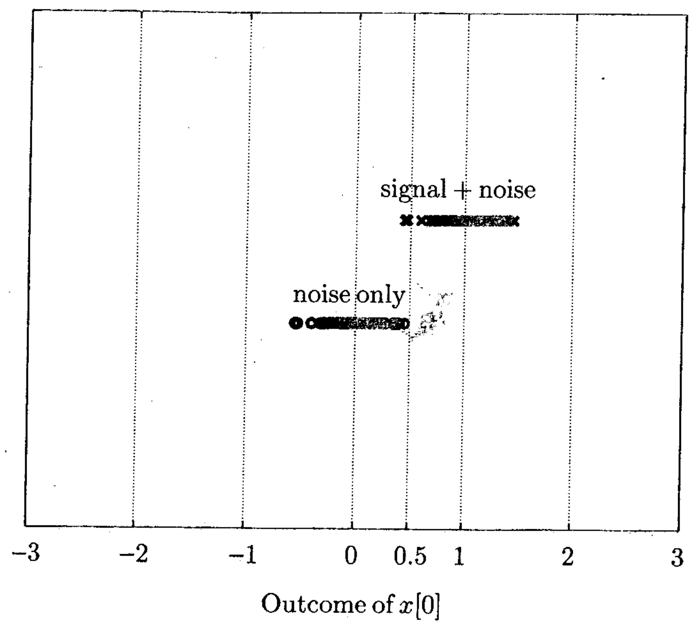
$$\begin{aligned} \mathcal{H}_0 &: x[0] = w[0] \\ \mathcal{H}_1 &: x[0] = 1 + w[0]. \end{aligned} \tag{1.4}$$

The PDFs under each hypothesis are denoted by $p(x[0];\mathcal{H}_0)$ and $p(x[0];\mathcal{H}_1)$, which for this example are

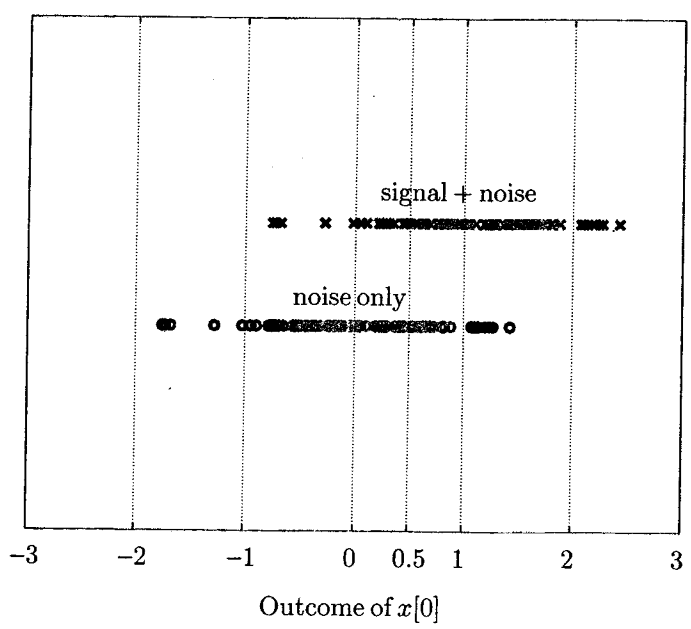
$$\begin{aligned} p(x[0];\mathcal{H}_0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} x^2[0]\right) \\ p(x[0];\mathcal{H}_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x[0]-1)^2\right). \end{aligned} \tag{1.5}$$

Note that in deciding between $\mathcal{H}_0$ and $\mathcal{H}_1$, we are essentially asking whether $x[0]$ has been generated according to the PDF $p(x[0];\mathcal{H}_0)$ or the PDF $p(x[0];\mathcal{H}_1)$. Alternatively, if we consider the *family* of PDFs

$$p(x[0];A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x[0]-A)^2\right) \tag{1.6}$$



(a)



(b)

Figure 1.4. Realizations of $x[0]$ for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

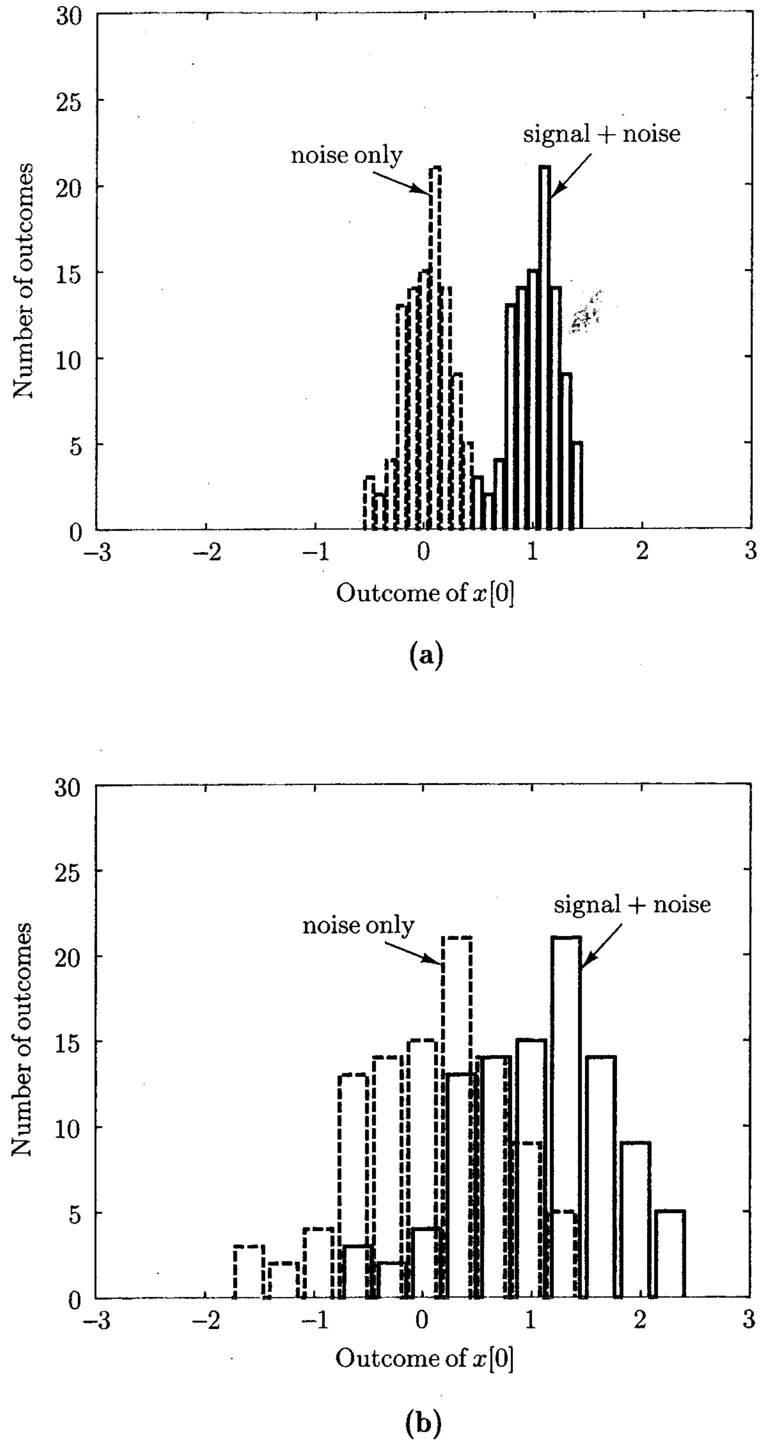


Figure 1.5. Histograms of $x[0]$ for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

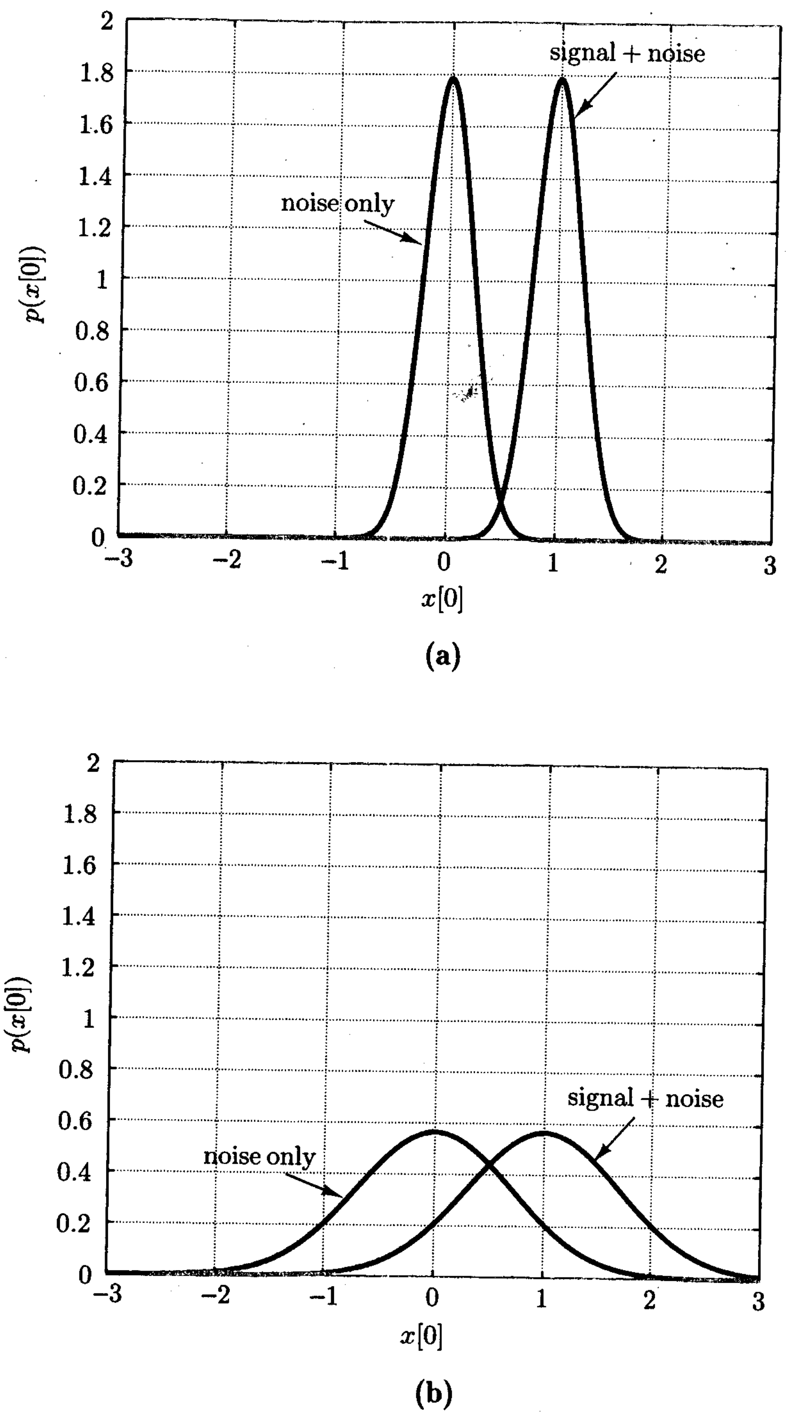


Figure 1.6. PDFs of $x[0]$ for signal present and signal absent (a) $\sigma^2 = 0.05$ (b) $\sigma^2 = 0.5$.

which is *parameterized* by A, then we obtain $p(x[0]; \mathcal{H}_0)$ if $A = 0$, and $p(x[0]; \mathcal{H}_1)$ if $A = 1$. We may, therefore, view the detection problem as a *parameter test.* Given the observation $x[0]$, whose PDF is given by (1.6), we wish to test if $A = 0$ or $A = 1$

or symbolically

$$\begin{array}{l} \mathcal{H}_0 : A = 0 \\ \mathcal{H}_1 : A = 1. \end{array} \tag{1.7}$$

This is termed a parameter test of the PDF, a viewpoint that will be useful later on.

At times it is convenient to assign *prior* probabilities to the possible occurrences of $\mathcal{H}_0$ and $\mathcal{H}_1$. For example, in an on-off keyed (OOK) digital communication system we transmit a "0" by sending no pulse and a "1" by sending a pulse with amplitude $A = 1$. Hence, the corresponding hypothesis test is given by (1.7). In an actual OOK system we will transmit a steady stream of data bits. Since the data bits, 0 or 1, are equally likely to be generated by the source (in the long run), we would expect $\mathcal{H}_0$ to be true half the time and $\mathcal{H}_1$ the other half. It makes sense then to regard the hypotheses as random events with probability 1/2. When we do so, our notation for the PDFs will be $p(x[0]|\mathcal{H}_0)$ and $p(x[0]|\mathcal{H}_1)$, in keeping with the standard notation of a conditional PDF. For this example, we have then that

$$\begin{aligned} p(x[0]|\mathcal{H}_0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2[0]\right) \\ p(x[0]|\mathcal{H}_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x[0]-1)^2\right) \end{aligned}$$

which should be contrasted with (1.5) (see also Problem 1.4). This distinction is analogous the the classical versus Bayesian approach to parameter estimation [Kay-I 1993].

1.4 Hierarchy of Detection Problems

The detection problems we will address will proceed from the simplest to the more difficult. The degree of difficulty is directly related to our knowledge of the signal and noise characteristics in terms of their PDFs. The ideal case occurs when we have exact knowledge of the PDFs. This is explored in Chapters 4 and 5. Then, at least in theory, we can obtain an optimal detector. When the PDFs are not completely known, the determination of a good (but possibly not optimal) detector is much more difficult. This case is discussed in Chapters 7–9. Another consideration in designing detectors is the mathematical tractability of the PDF. The Gaussian PDF is particularly convenient from a theoretical and practical viewpoint and will be the assumption most often made. In Chapter 10, however, the Gaussian PDF is replaced by the more general nonGaussian PDF. A summary of detection problems and where they are discussed is shown in Table 1.1. Along with the increasing difficulty of determining a detector, we will see that the detection *performance* decreases as we have less specific knowledge of the signal and noise characteristics.

Signal ↓ \ Noise →	Gaussian Known PDF	Gaussian Unknown PDF	NonGaussian Known PDF	NonGaussian Unknown PDF
Deterministic Known	4	9	10	*
Deterministic Unknown	7	9	10	*
Random Known PDF	5	9	*	*
Random Unknown PDF	8	*	*	*

* Not discussed (beyond scope of text)

Table 1.1. Hierarchy of detection problems and chapters where discussed.

1.5 Role of Asymptotics

In practice we are principally interested in detecting signals that are *weak* or signals whose SNR is small. If this were not the case, then there would be little need to bother with detection theory in that the signal would not be "buried" in the noise. This is in contrast to the estimation problem in which we usually desire a highly accurate estimate. For accurate estimation we are required to control the SNR so that it is high enough to meet some requirement. It followed then that in estimation problems an asymptotic or high SNR assumption was sometimes useful. In detection problems, however, we are generally faced with a low SNR signal so that our success depends on the data record length. As an illustration, assume we wish to detect the same DC level as before but we will do so by taking multiple measurements. Our data then consist of $x[n] = w[n]$ for $n = 0, 1, \ldots, N-1$ under $\mathcal{H}_0$ and $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$ under $\mathcal{H}_1$, or more formally we have the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $w[n]$ is WGN with variance σ^2. A reasonable approach might be to average the samples and compare the value obtained to a threshold γ so that we would decide $\mathcal{H}_1$ if

$$T = \frac{1}{N} \sum_{n=0}^{N-1} x[n] > \gamma.$$

(Note that (1.1) is just a special case when $N = 1$ and $\gamma = 1/2$.) Intuitively, we expect that as N increases, the detection performance should also increase. To justify our intuition we have plotted a histogram of T for $N = 1$ and $N = 10$ using $\sigma^2 = 0.5$ in Figure 1.7. The experiment has been repeated 100 times so that 100 outcomes of T have been generated. It is seen that, as expected, the overlap between the histograms (which are estimates of the PDFs to within a scale factor) is less as N increases. To quantify this we use a measure that increases with the differences of the means or $E(T;\mathcal{H}_1)-E(T;\mathcal{H}_0)$ and which increases as the variance of each PDF decreases. Noting that $\mathrm{var}(T;\mathcal{H}_0) = \mathrm{var}(T;\mathcal{H}_1)$ (see Problem 1.6) we have the measure, termed the *deflection coefficient*,

$$d^2 = \frac{(E(T;\mathcal{H}_1) - E(T;\mathcal{H}_0))^2}{\mathrm{var}(T;\mathcal{H}_0)}.$$

It can be shown (see Chapter 4) that the detection performance increases with increasing d^2. For the problem at hand it is easily shown that (see Problem 1.6)

$$\begin{aligned} E(T;\mathcal{H}_0) &= 0 \\ E(T;\mathcal{H}_1) &= A \\ \mathrm{var}(T;\mathcal{H}_0) &= \sigma^2/N \end{aligned}$$

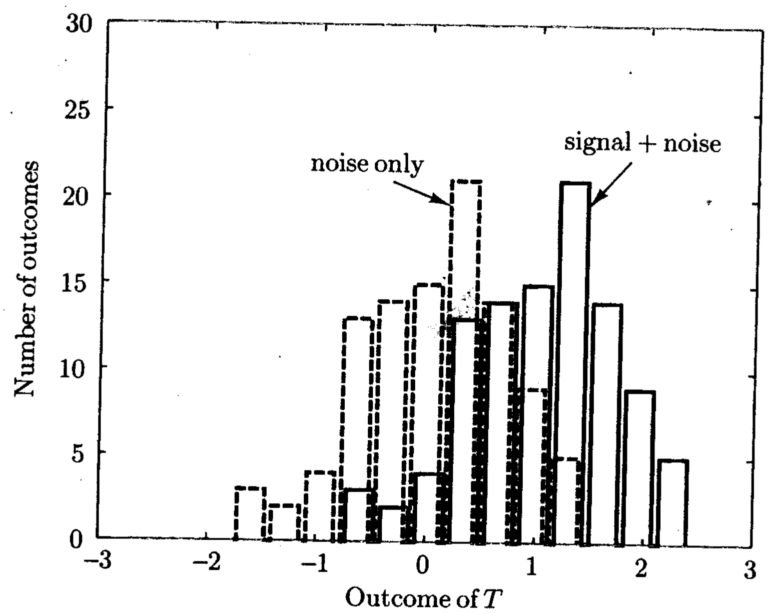
so that

$$d^2 = \frac{A^2}{\sigma^2/N} = \frac{NA^2}{\sigma^2}. \tag{1.8}$$

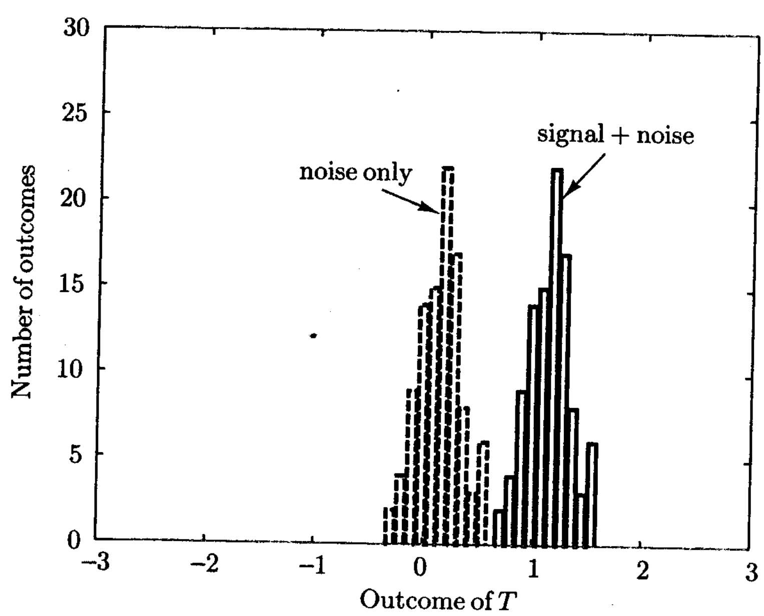
Hence, as intuited, the detection performance improves as the SNR A^2/σ^2 increases and/or *the data record length N increases.* For weak signals, for which A^2/σ^2 is small, we require N to be large for good detection performance. This has the effect of reducing the noise via averaging since the variance of T, which is due to noise, is σ^2/N. As a result, in detection theory asymptotic analysis (as $N \to \infty$) proves to be appropriate and quite useful. It allows us to derive detectors more easily and also to analyze their performance. For example, if $w[n]$ consisted of independent and identically distributed samples of *nonGaussian* noise, then T would not have a Gaussian PDF. However, as $N \to \infty$, we could invoke the central limit theorem to justify a Gaussian approximation. To determine the detection performance we would need only to obtain the first two moments of T.

1.6 Some Notes to the Reader

Our philosophy in presenting a theory of detection is to provide the user with the main ideas necessary for determining optimal detectors, where possible, and good detectors otherwise. We have included results that we deem to be most useful in practice, omitting some important theoretical issues. The latter can be found in many books on statistical theory, which have been written from a more theoretical viewpoint [Cox and Hinkley 1974, Lehmann 1959, Kendall and Stuart 1976–1979, Rao 1973]. Other books on detection theory that are similar to this one and



(a)



(b)

Figure 1.7. Histograms of T for signal present and signal absent (a) $N = 1$ (b) $N = 10$.

which the reader may wish to consult are [Van Trees 1968–1971, Helstrom 1995, McDonough and Whalen 1995]. In Chapter 11 we provide a "road map" for determining a good detector as well as a summary of the various methods and their properties. The reader may wish to read this chapter first to obtain an overview.

This text is part of a two-volume series on statistical signal processing. The first volume, *Fundamentals of Statistical Signal Processing: Estimation Theory*, is referenced as [Kay-I 1993]. For a full appreciation the reader is assumed to be familiar with Volume I, although we have tried to minimize its impact on this text. When this was not possible, references to Volume I by chapter or page number have been made.

We have also tried to maximize insight by including many examples and minimizing long mathematical expositions, although much of the tedious algebra and proofs have been included as appendices. The DC level in noise described earlier will serve as a standard example in introducing almost all of the detection approaches. It is hoped that in doing so the reader will be able to develop his or her own intuition by building upon previously assimilated concepts.

The mathematical notation for all common symbols is summarized in Appendix 2. The distinction between a continuous-time waveform and a discrete-time waveform or sequence is made through the symbolism $x(t)$ for continuous-time and $x[n]$ for discrete-time. Plots of $x[n]$, however, appear continuous in time, the points having been connected by straight lines for easier viewing. All vectors and matrices are *boldface* with all vectors being *column* vectors. All other symbolism is defined within the context of the discussion.

References

Burdic, W.S., *Underwater Acoustic System Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1984.

Chan, D.S.K, D.A. Langan, S.A. Staver, "Spatial Processing Techniques for the Detection of Small Targets in IR Clutter," *SPIE, Signal and Data Processing of Small Targets*, Vol. 1305, pp. 53–62, 1990.

Cox, D.R., D.V. Hinkley, *Theoretical Statistics*, Chapman and Hall, New York, 1974.

Ellenberg, S.S., D.M. Finklestein, D.A. Schoenfeld, "Statistical Issues Arising in AIDS Clinical Trials," *J. Am. Statist. Asssoc.*, Vol. 87, pp. 562–569, 1992.

Fukunaga, K., *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, New York, 1990.

Gustafson, D.E., A.S. Willsky, J.Y. Wang, M.C. Lancaster, J.H. Triebwasser, "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis. Part II: Identification of Transient Rhythms," *IEEE Trans. Biomed. Eng.*, Vol. BME-25, pp. 353–361, 1978.

Helstrom, C.W., *Elements of Signal Detection and Estimation*, Prentice-Hall, Englewood Cliffs, N.J., 1995.

Ives, R.B., "The Applications of Advanced Pattern Recognition Techniques for the Discrimination Between Earthquakes and Nuclear Detonations," *Pattern Recognition*, Vol. 14, pp. 155–161, 1981.

Justice, J.H. ,"Array Processing in Exploration Seismology," in *Array Signal Processing*, S. Haykin, Ed., Prentice-Hall, Englewood Cliffs, N.J., 1985.

Kay, S.M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1993.

Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics*, Vols. 1–3, Macmillan, New York, 1976–1979.

Knight, W.S., R.G. Pridham, S.M. Kay, "Digital Signal Processing for Sonar," *Proc. IEEE*, Vol. 69, pp. 1451–1506, Nov. 1981.

Lehmann, E.L., *Testing Statistical Hypotheses*, J. Wiley, New York, 1959.

McDonough, R.N., A.T. Whalen,*Detection of Signals in Noise*, 2nd Ed., Academic Press, 1995.

Proakis, J.G., *Digital Communications*, 2nd Ed., McGraw-Hill, New York, 1989.

Rabiner, L., B-H Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, N.J., 1993.

Rao, C.R., *Linear Statistical Inference and its Applications*, J. Wiley, New York, 1973.

Skolnik, M.I., *Introduction to Radar Systems*, McGraw-Hill, New York, 1980.

Taylor, S., *Modeling Financial Time Series*, J. Wiley, New York, 1986.

Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Vols. I–III, J. Wiley, New York, 1968–1971.

Willsky, A.S., H.L. Jones, "A Generalized Likelihood Ratio Approach to Detection and Estimation of Jumps in Linear Systems," *IEEE Trans. Automat. Contr.*, Vol. AC-21, pp. 108–112, 1976.

Problems

1.1 Consider the detection problem

$$\begin{aligned}\mathcal{H}_0 &: x[0] = w[0] \\ \mathcal{H}_1 &: x[0] = 1 + w[0]\end{aligned}$$

where $w[0]$ is a zero mean Gaussian random variable with variance σ^2. If the detector decides $\mathcal{H}_1$ if $x[0] > 1/2$, find the probability of making a wrong decision when $\mathcal{H}_0$ is true. To do so, determine the probability of deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true or $P_0 = \Pr\{x[0] > 1/2; \mathcal{H}_0\}$. For this to be 10^{-3} what must σ^2 be?

1.2 Consider the detection problem

$$\begin{aligned}\mathcal{H}_0 &: x[0] = w[0] \\ \mathcal{H}_1 &: x[0] = 1 + w[0]\end{aligned}$$

where $w[0]$ is a uniformly distributed random variable on the interval $[-a, a]$ for $a > 0$. Discuss the performance of the detector that decides $\mathcal{H}_1$ if $x[0] > 1/2$ as a increases.

1.3 We observe the datum $x[0]$ where $x[0]$ is a Gaussian random variable with mean A and variance $\sigma^2 = 1$. We wish to test if $A = A_0$ or $A = -A_0$, where $A_0 > 0$. Propose a test and discuss its performance as a function of A_0.

1.4 In Problem 1.1 now assume that the probability of $\mathcal{H}_0$ being true is 1/2. If we decide $\mathcal{H}_1$ is $x[0] > 1/2$, find the *total* probability of error or

$$P_e = \Pr\{x[0] > 1/2|\mathcal{H}_0\} \Pr\{\mathcal{H}_0\} + \Pr\{x[0] < 1/2|\mathcal{H}_1\} \Pr\{\mathcal{H}_1\}.$$

Plot P_e versus σ^2 and explain your result.

1.5 In Problem 1.1 now assume that we have two samples on which to base our decision. We decide that a signal is present if

$$T = \frac{1}{2}(x[0] + x[1]) > \frac{1}{2}.$$

Determine all the values of $x[0]$ and $x[1]$ that will result in a decision that a signal is present. Plot these values in a plane. Also, plot the point $[E(x[0]), E(x[1])]^T$ assuming $\mathcal{H}_0$ is true and then assuming $\mathcal{H}_1$ is true. Comment on the results.

1.6 Verify (1.8) by determining the means and variances of T.

1.7 Using (1.8) determine the number of samples required if the deflection coefficient is to be $d^2 = 100$ for adequate detection performance and the SNR is -20 dB.

Chapter 2

Summary of Important PDFs

2.1 Introduction

Evaluation of the performance of a detector depends upon the ability to determine the probability density function of a function of the data samples, either analytically or numerically. When this is not possible, we must resort to Monte Carlo computer simulation techniques. Familiarity with common probability density functions and their properties is essential to the success of the performance evaluation. In this chapter we provide the background material that will be called upon throughout the text. Our discussion is cursory at best due to space limitations. For further details the reader is referred to [Abramowitz and Stegun 1970, Kendall and Stuart 1976–1979, Johnson, Kotz, and Balakrishnan 1995], as well as to the specific references given within the chapter.

2.2 Fundamental Probability Density Functions and Properties

2.2.1 Gaussian (Normal)

The *Gaussian* probability density function (PDF) (also referred to as the *normal* PDF) for a scalar random variable x is defined as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \qquad -\infty < x < \infty \tag{2.1}$$

where μ is the mean and σ^2 is the variance of x. It is denoted by $\mathcal{N}(\mu, \sigma^2)$ and we say that $x \sim \mathcal{N}(\mu, \sigma^2)$, where "$\sim$" means "is distributed according to." If $\mu = 0$, its moments are

$$E(x^n) = \begin{cases} 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n & n \text{ even} \\ 0 & n \text{ odd.} \end{cases} \tag{2.2}$$

Otherwise, we use

$$E[(x+\mu)^n] = \sum_{k=0}^{n} \binom{n}{k} E(x^k)\mu^{n-k}$$

where $E(x^k)$ is given by (2.2). The cumulative distribution function (CDF) for $\mu = 0$ and $\sigma^2 = 1$, for which the PDF is termed a *standard normal* PDF, is defined as

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt.$$

A more convenient description, which is termed the *right-tail probability* and is the probability of exceeding a given value, is defined as $Q(x) = 1 - \Phi(x)$, where

$$Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt. \tag{2.3}$$

The function $Q(x)$ is also referred to as the *complementary cumulative distribution function.* This cannot be evaluated in closed-form. Its value is shown in Figure 2.1 on linear and log scales. For computational purposes we use the MATLAB program Q.m listed in Appendix 2C. An approximation that is sometimes useful is (see Problem 2.2)

$$Q(x) \approx \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{1}{2}x^2\right). \tag{2.4}$$

It is shown in Figure 2.2 along with the exact value of $Q(x)$. The approximation is quite accurate for $x > 4$. At times it is important to determine if a random variable has a PDF that is approximately Gaussian. An examination of its right-tail probability reveals whether this is so. By plotting $Q(x)$ on *normal probability paper*, the curve becomes a straight line as shown in Figure 2.3. The normal probability paper can be constructed as described in Appendix 2B. Also contained there is the MATLAB program plotprob.m for plotting the right-tail probability on normal probability paper. As an example, the right-tail probability of the nonGaussian PDF

$$p(x) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi \cdot 2}} \exp\left(-\frac{1}{2\cdot 2}x^2\right)$$

which is a Gaussian mixture PDF, is plotted on normal probability paper in Figure 2.4. The functional form for the right-tail probability is easily shown to be $Q(x)/2 + Q(x/\sqrt{2})/2$. The function $Q(x)$ is shown as a dashed straight line for comparison.

If it is known that a probability is given by $P = Q(\gamma)$, then we can determine γ for a given P. Symbolically, we have $\gamma = Q^{-1}(P)$, where Q^{-1} is the inverse function. The latter must exist since $Q(x)$ is strictly monotonically decreasing. The MATLAB program Qinv.m listed in Appendix 2C computes the function Q^{-1} and can be used to determine γ numerically.

The *multivariate Gaussian* PDF of an $n \times 1$ random vector $\mathbf{x}$ is defined as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{2.5}$$

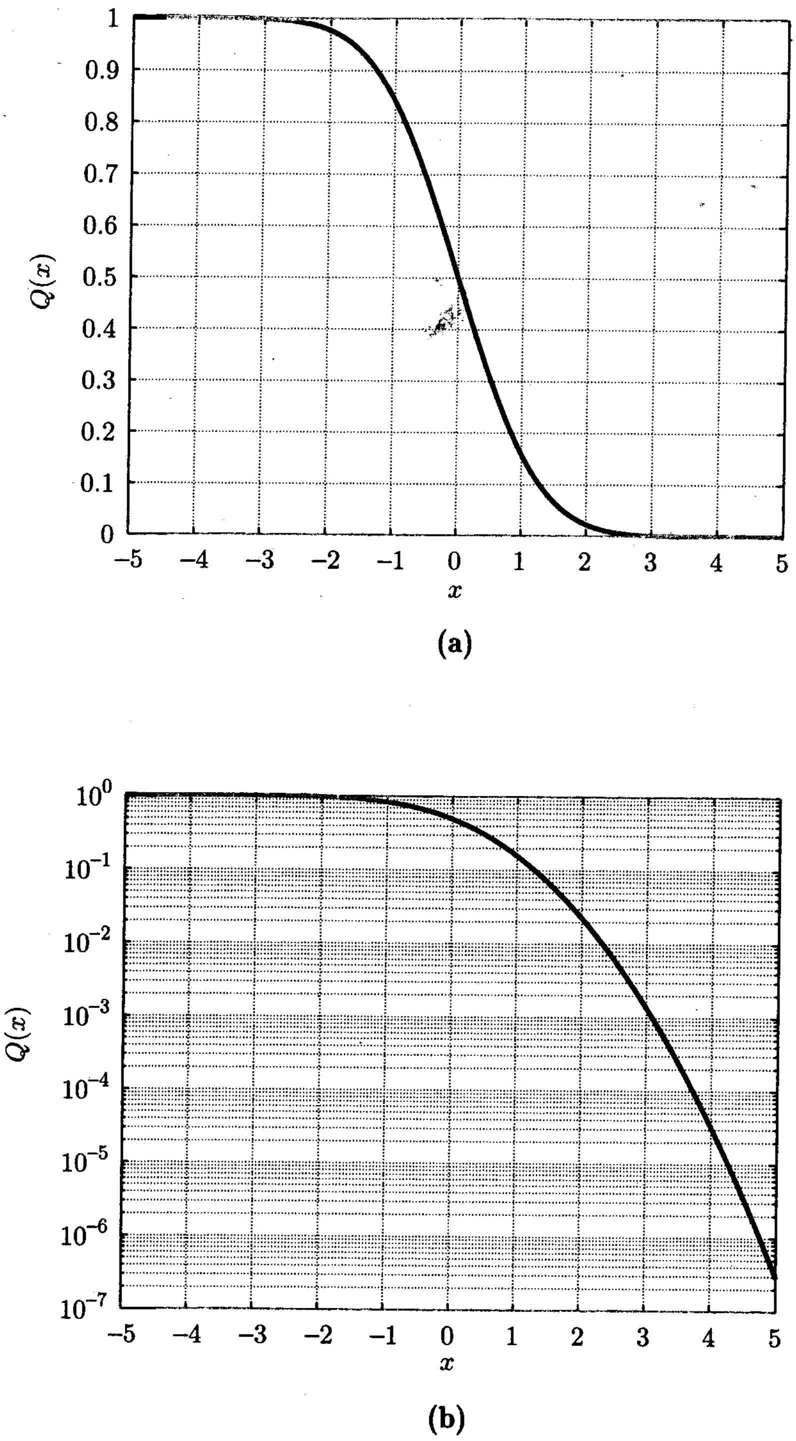


Figure 2.1. Right-tail probability for standard normal PDF (a) Linear vertical axis (b) Logarithmic vertical axis.

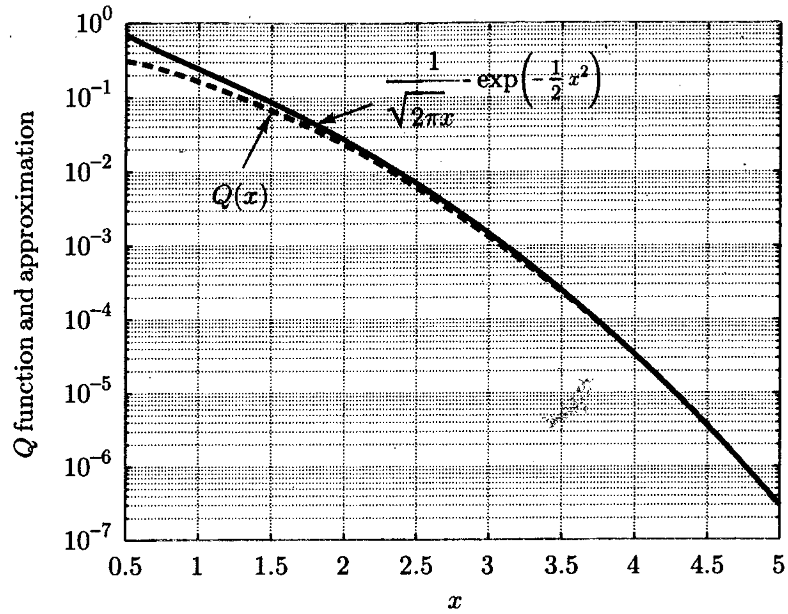


Figure 2.2. Approximation to Q function.

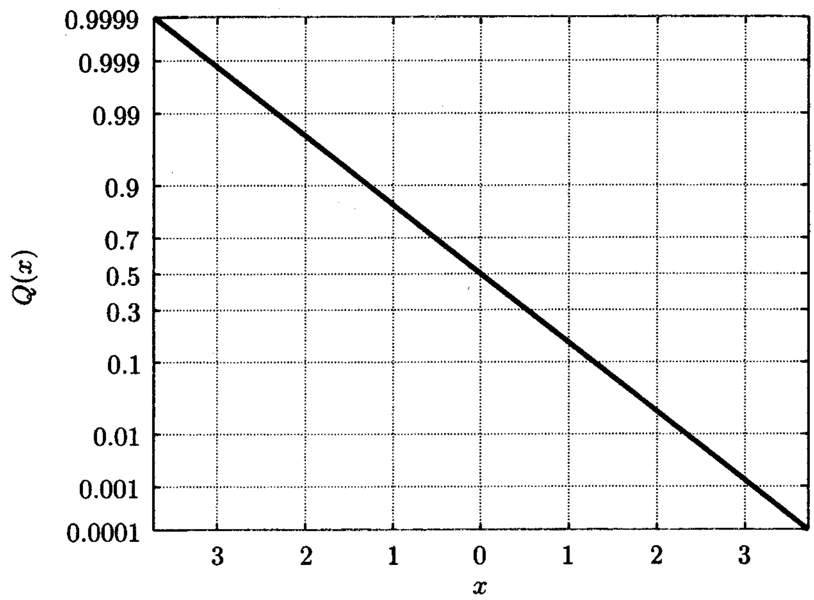


Figure 2.3. Q function plotted on normal probability paper.

where $\boldsymbol{\mu}$ is the mean vector and $\mathbf{C}$ is the covariance matrix and is denoted by $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$. It is assumed that $\mathbf{C}$ is positive definite and hence $\mathbf{C}^{-1}$ exists. The mean vector is defined as

$$[\boldsymbol{\mu}]_i = E(x_i) \qquad i = 1, 2, \ldots, n$$

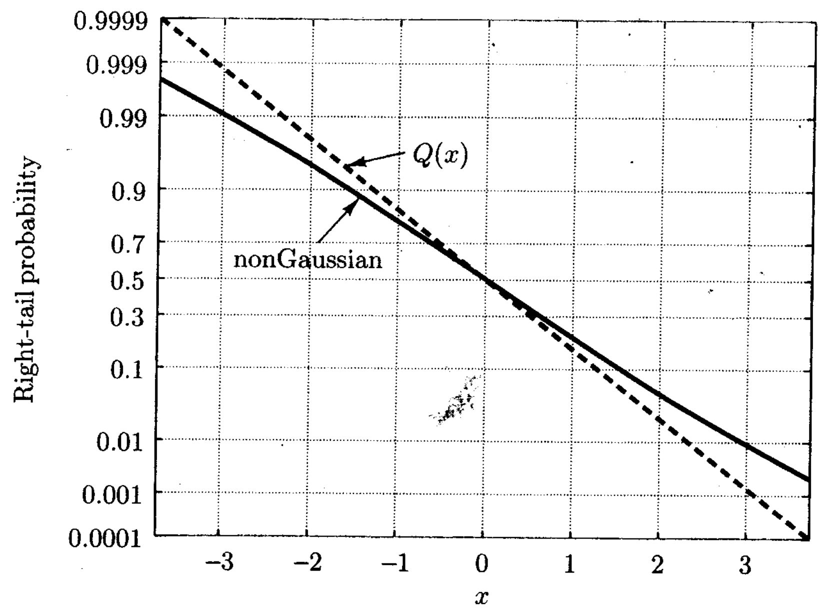


Figure 2.4. Right-tail probability for nongaussian PDF on normal probability paper.

and the covariance matrix as

$$[\mathbf{C}]_{ij} = E[(x_i - E(x_i))(x_j - E(x_j))] \qquad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n$$

or in more compact form as

$$\mathbf{C} = E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T].$$

If $\boldsymbol{\mu} = \mathbf{0}$, then all odd-order joint moments are zero. Even-order moments are found as combinations of second-order moments. In particular, a useful result for $\boldsymbol{\mu} = \mathbf{0}$ is

$$E(x_i x_j x_k x_l) = E(x_i x_j)E(x_k x_l) + E(x_i x_k)E(x_j x_l) + E(x_i x_l)E(x_j x_k). \tag{2.6}$$

2.2.2 Chi-Squared (Central)

A *chi-squared* PDF with ν degrees of freedom is defined as

$$p(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \exp(-\frac{1}{2}x) & x > 0 \\ 0 & x < 0 \end{cases} \tag{2.7}$$

and is denoted by χ^2_ν. The degrees of freedom ν is assumed to be an integer with $\nu \geq 1$. The function $\Gamma(u)$ is the Gamma function, which is defined as

$$\Gamma(u) = \int_0^\infty t^{u-1} \exp(-t)dt.$$

The relations $\Gamma(u) = (u-1)\Gamma(u-1)$ for any u, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and $\Gamma(n) = (n-1)!$ for n an integer can be used to evaluate it. Some examples of the PDF are given in Figure 2.5. It becomes Gaussian as ν becomes large. Note that for $\nu = 1$ the PDF is infinite at $x = 0$.

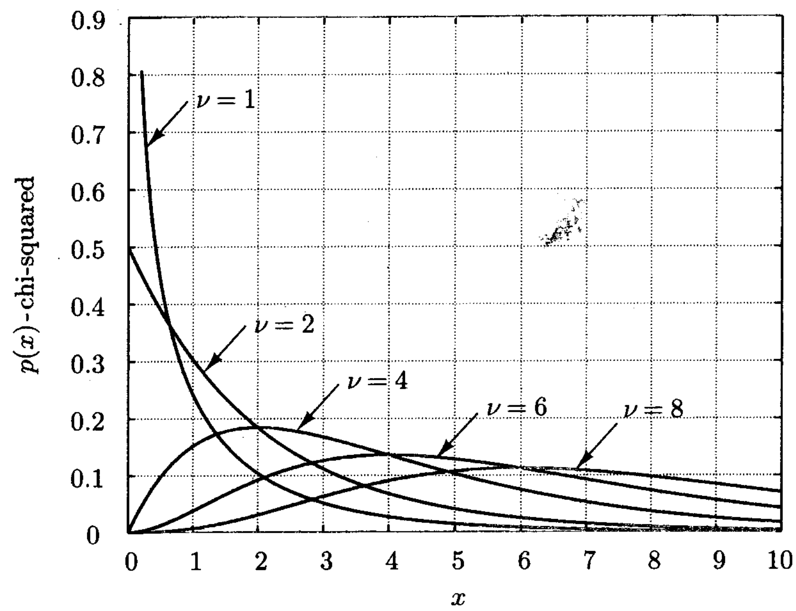


Figure 2.5. PDF for chi-squared random variable.

The chi-squared PDF arises as the PDF of x where $x = \sum_{i=1}^{\nu} x_i^2$ if $x_i \sim \mathcal{N}(0,1)$ and the x_i's are independent and identically distributed (IID). By the latter we mean that each x_i is independent of the others and each x_i has the same PDF (identically distributed). The mean and variance are

$$E(x) = \nu \tag{2.8}$$

$$\operatorname{var}(x) = 2\nu. \tag{2.9}$$

A specific case of interest occurs when $\nu = 2$ so that

$$p(x) = \begin{cases} \frac{1}{2}\exp(-\frac{1}{2}x) & x > 0 \\ 0 & x < 0 \end{cases}$$

and is referred to as an *exponential* PDF (see Figure 2.5).

The right-tail probability for a χ_ν^2 random variable is defined as

$$Q_{\chi_\nu^2}(x) = \int_x^\infty p(t)dt \qquad x > 0$$

and can be shown to be [Abramowitz and Stegun 1970] for ν even

$$Q_{\chi_\nu^2}(x) = \exp\left(-\frac{1}{2}x\right) \sum_{k=0}^{\frac{\nu}{2}-1} \frac{\left(\frac{x}{2}\right)^k}{k!} \qquad \nu \geq 2 \tag{2.10}$$

and for ν odd

$$Q_{\chi^2_\nu}(x) = \begin{cases} 2Q(\sqrt{x}) & \nu = 1 \\ 2Q(\sqrt{x}) + \dfrac{\exp(-\frac{1}{2}x)}{\sqrt{\pi}} \displaystyle\sum_{k=1}^{\frac{\nu-1}{2}} \dfrac{(k-1)!(2x)^{k-\frac{1}{2}}}{(2k-1)!} & \nu \geq 3. \end{cases} \tag{2.11}$$

The MATLAB program Qchipr2.m listed in Appendix 2D can be used to numerically evaluate $Q_{\chi^2_\nu}(x)$.

2.2.3 Chi-Squared (Noncentral)

A generalization of the χ^2_ν PDF arises as a result of summing the squares of IID Gaussian random variables with *nonzero means.* Specifically, if $x = \sum_{i=1}^{\nu} x_i^2$, where the x_i's are independent and $x_i \sim \mathcal{N}(\mu_i, 1)$, then x has a *noncentral chi-squared* PDF with ν degrees of freedom and *noncentrality parameter* $\lambda = \sum_{i=1}^{\nu} \mu_i^2$. The PDF is quite complicated and must be expressed either in integral or infinite series form. As an integral it is

$$p(x) = \begin{cases} \frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{\nu-2}{4}} \exp\left[-\frac{1}{2}(x+\lambda)\right] I_{\frac{\nu}{2}-1}(\sqrt{\lambda x}) & x > 0 \\ 0 & x < 0 \end{cases} \tag{2.12}$$

where $I_r(u)$ is the modified Bessel function of the first kind and order r. It is defined as

$$I_r(u) = \frac{(\frac{1}{2}u)^r}{\sqrt{\pi}\Gamma(r+\frac{1}{2})} \int_0^\pi \exp(u\cos\theta)\sin^{2r}\theta\, d\theta \tag{2.13}$$

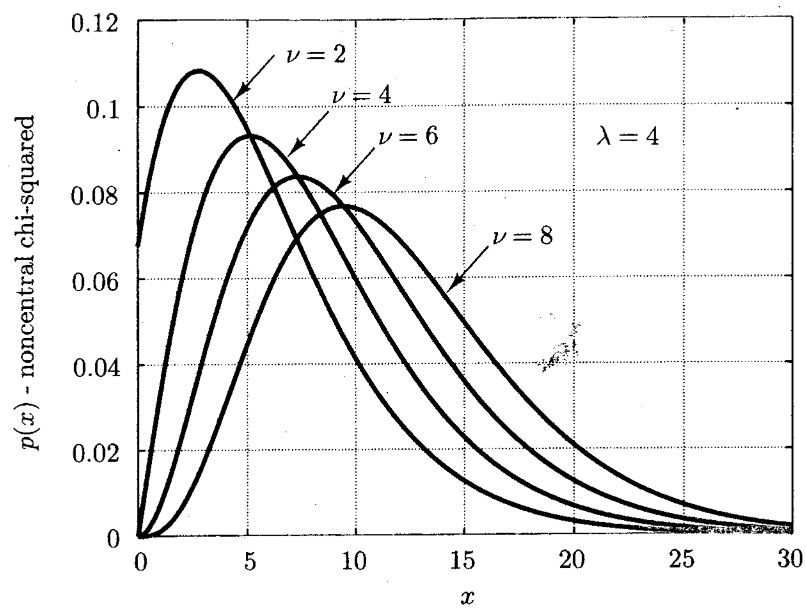
and has the series representation

$$I_r(u) = \sum_{k=0}^{\infty} \frac{(\frac{1}{2}u)^{2k+r}}{k!\,\Gamma(r+k+1)}. \tag{2.14}$$

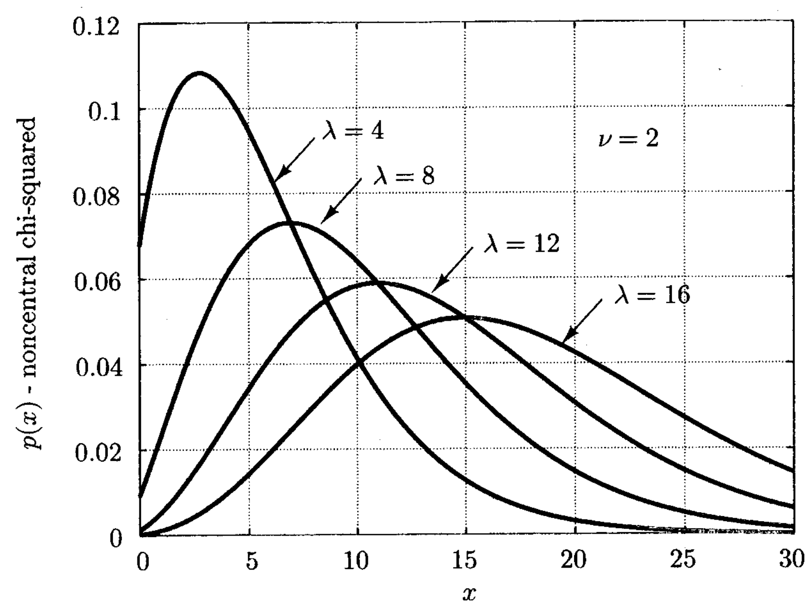
Some examples of the PDF are given in Figure 2.6. Note that the PDF becomes Gaussian as ν becomes large. Using the series expansion of (2.14) the PDF can also be expressed in infinite series form as

$$p(x) = \frac{x^{\frac{\nu}{2}-1}\exp[-\frac{1}{2}(x+\lambda)]}{2^{\frac{\nu}{2}}} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda x}{4}\right)^k}{k!\,\Gamma(\frac{\nu}{2}+k)}. \tag{2.15}$$

Note that for $\lambda = 0$, the noncentral chi-squared PDF reduces to the chi-squared PDF. The noncentral chi-squared PDF with ν degrees of freedom and noncentrality parameter λ is denoted by $\chi'^2_\nu(\lambda)$.



(a)



(b)

Figure 2.6. PDF for noncentral chi-squared random variable (a) Varying degrees of freedom (b) Varying noncentrality parameter.

The mean and variance are

$$\begin{aligned} E(x) &= \nu + \lambda \\ \text{var}(x) &= 2\nu + 4\lambda. \end{aligned} \tag{2.16}$$

We will denote the right-tail probability as

$$Q_{\chi_\nu'^2(\lambda)}(x) = \int_x^\infty p(t)dt \qquad x > 0$$

Its value can be numerically determined by the MATLAB program Qchipr2.m given in Appendix 2D.

2.2.4 F (Central)

The F PDF arises as the ratio of two independent χ^2 random variables. Specifically, if

$$x = \frac{x_1/\nu_1}{x_2/\nu_2}$$

where $x_1 \sim \chi^2_{\nu_1}$, $x_2 \sim \chi^2_{\nu_2}$, and x_1 and x_2 are independent, then x has the F PDF. It is given by

$$p(x) = \begin{cases} \dfrac{\left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{B\left(\dfrac{\nu_1}{2}, \dfrac{\nu_2}{2}\right)} \dfrac{x^{\frac{\nu_1}{2}-1}}{\left(1 + \dfrac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}} & x > 0 \\ 0 & x < 0 \end{cases} \tag{2.17}$$

where $B(u, v)$ is the Beta function, which can be related to the Gamma function as

$$B(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}.$$

The PDF is denoted by F_{ν_1,ν_2} as an F PDF with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom. Some examples of the PDF are given in Figure 2.7. The right-tail probability is denoted by $Q_{F_{\nu_1,\nu_2}}(x)$ and must be evaluated numerically [Abramowitz and Stegun 1970]. The mean and variance are

$$\begin{aligned} E(x) &= \frac{\nu_2}{\nu_2 - 2} & \nu_2 > 2 \\ \text{var}(x) &= \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} & \nu_2 > 4. \end{aligned} \tag{2.18}$$

Note that as $\nu_2 \to \infty$, $x \to x_1/\nu_1 \sim \chi^2_{\nu_1}/\nu_1$ since $x_2/\nu_2 \to 1$ (see Problem 2.3).

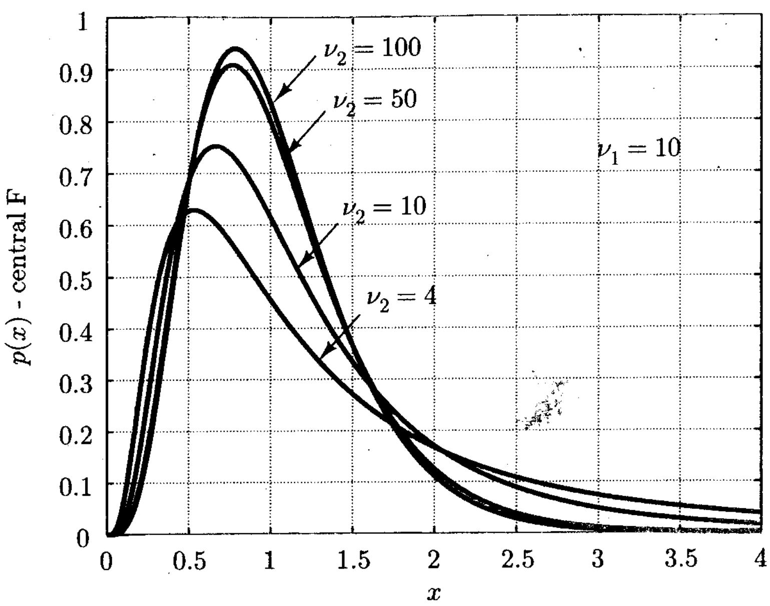


Figure 2.7. PDF for F random variable.

2.2.5 F (Noncentral)

The *noncentral* F PDF results from the ratio of a noncentral χ^2 random variable to a central χ^2 random variable. Specifically, if

$$x = \frac{x_1/\nu_1}{x_2/\nu_2}$$

where $x_1 \sim \chi'^2_{\nu_1}(\lambda)$ and $x_2 \sim \chi^2_{\nu_2}$ and x_1, x_2 are independent, then x has the noncentral F PDF. It is denoted by $F'_{\nu_1,\nu_2}(\lambda)$ as a noncentral F PDF with ν_1 numerator and ν_2 denominator degrees of freedom and noncentrality parameter λ. Its PDF in infinite series form is

$$\begin{aligned} p(x) &= \exp\left(-\frac{\lambda}{2}\right) \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^k}{k!} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}\nu_1+k}}{B\left(\frac{\nu_1+2k}{2}, \frac{\nu_2}{2}\right)} \\ &\quad \cdot x^{\frac{\nu_1}{2}+k-1} \left(1 + \frac{\nu_1}{\nu_2}x\right)^{-\frac{1}{2}(\nu_1+\nu_2)-k}. \qquad (2.19) \end{aligned}$$

Some examples of the PDF can be found in [Johnson and Kotz 1995]. For $\lambda = 0$ this reduces to the central F PDF (let $k = 0$ in (2.19)). Its mean and variance are

$$E(x) = \frac{\nu_2(\nu_1 + \lambda)}{\nu_1(\nu_2 - 2)} \qquad \nu_2 > 2$$

$$\text{var}(x) = 2\left(\frac{\nu_2}{\nu_1}\right)^2 \frac{(\nu_1+\lambda)^2+(\nu_1+2\lambda)(\nu_2-2)}{(\nu_2-2)^2(\nu_2-4)} \quad \nu_2 > 4. \tag{2.20}$$

The right-tail probability is denoted by $Q_{F'_{\nu_1,\nu_2}(\lambda)}(x)$ and requires numerical evaluation [Patnaik 1949]. Also, note that as $\nu_2 \to \infty$, $F'_{\nu_1,\nu_2}(\lambda) \to \chi'^2_{\nu_1}(\lambda)$ (see Problem 2.3).

2.2.6 Rayleigh

The *Rayleigh* PDF is obtained as the PDF of $x = \sqrt{x_1^2 + x_2^2}$, where $x_1 \sim \mathcal{N}(0, \sigma^2)$, $x_2 \sim \mathcal{N}(0, \sigma^2)$, and x_1, x_2 are independent. Its PDF is

$$p(x) = \begin{cases} \dfrac{x}{\sigma^2}\exp\left(-\dfrac{1}{2\sigma^2}x^2\right) & x > 0 \\ 0 & x < 0. \end{cases} \tag{2.21}$$

It is shown in Figure 2.8 for $\sigma^2 = 1$. The mean and variance are

$$\begin{aligned} E(x) &= \sqrt{\frac{\pi\sigma^2}{2}} \\ \text{var}(x) &= \left(2 - \frac{\pi}{2}\right)\sigma^2. \end{aligned} \tag{2.22}$$

The right-tail probability is easily found as

$$\int_x^\infty p(t)dt = \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{2.23}$$

The Rayleigh PDF is related to the χ_2^2 PDF since if x is a Rayleigh random variable, then $x = \sqrt{\sigma^2 y}$, where $y \sim \chi_2^2$. As a result, the right-tail probabilities can be related as

$$\begin{aligned} \Pr\{x > \sqrt{\gamma'}\} &= \Pr\left\{x/\sqrt{\sigma^2} > \sqrt{\gamma'/\sigma^2}\right\} \\ &= \Pr\left\{\sqrt{y} > \sqrt{\gamma'/\sigma^2}\right\} \\ &= \Pr\{y > \gamma'/\sigma^2\} \\ &= Q_{\chi_2^2}\left(\frac{\gamma'}{\sigma^2}\right) \end{aligned}$$

or

$$\Pr\{x > \gamma\} = Q_{\chi_2^2}\left(\frac{\gamma^2}{\sigma^2}\right)$$

which produces (2.23) since $Q_{\chi_2^2}(x) = \exp(-x/2)$.

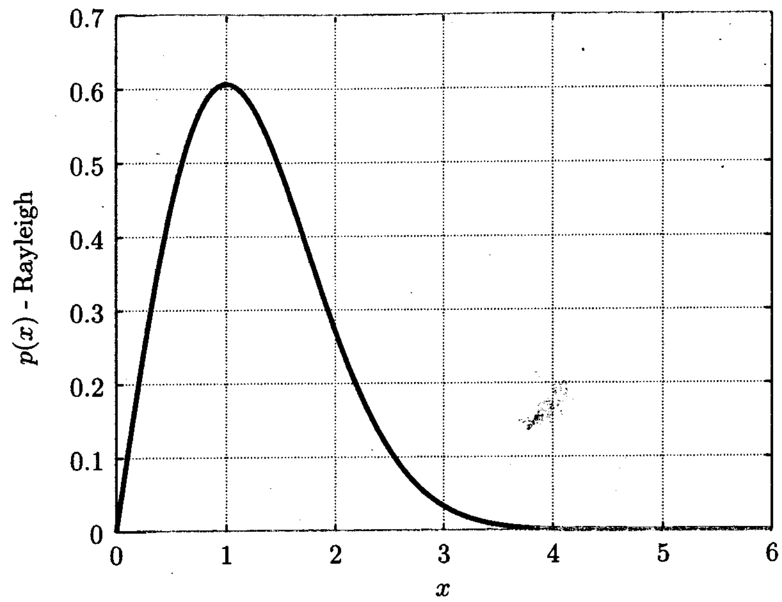


Figure 2.8. PDF for Rayleigh random variable ($\sigma^2 = 1$).

2.2.7 Rician

The *Rician* PDF is obtained as the PDF of $x = \sqrt{x_1^2 + x_2^2}$, where $x_1 \sim \mathcal{N}(\mu_1, \sigma^2)$, $x_2 \sim \mathcal{N}(\mu_2, \sigma^2)$, and x_1, x_2 are independent. Its PDF is (see Problem 7.19)

$$p(x) = \begin{cases} \dfrac{x}{\sigma^2} \exp\left[-\dfrac{1}{2\sigma^2}(x^2 + \alpha^2)\right] I_0\left(\dfrac{\alpha x}{\sigma^2}\right) & x > 0 \\ 0 & x < 0 \end{cases} \tag{2.24}$$

where $\alpha^2 = \mu_1^2 + \mu_2^2$ and $I_0(u)$ is given by (2.13) with $r = 0$ or

$$\begin{aligned} I_0(u) &= \frac{1}{\pi} \int_0^{\pi} \exp(u \cos\theta) d\theta \\ &= \int_0^{2\pi} \exp(u\cos\theta) \frac{d\theta}{2\pi}. \end{aligned} \tag{2.25}$$

Some examples are given in Figure 2.9 for $\sigma^2 = 1$. For $\alpha^2 = 0$ it reduces to the Rayleigh PDF. Its moments are expressable in terms of confluent hypergeometric functions, which can be found in [Rice 1948, McDonough and Whalen 1995]. The right-tail probability can be shown to be related to that of the noncentral χ^2 random variable and must be evaluated numerically (see also Problem 7.20). To do so we proceed as follows:

$$\Pr\{x > \sqrt{\gamma'}\} = \Pr\left\{\sqrt{\frac{x_1^2 + x_2^2}{\sigma^2}} > \sqrt{\frac{\gamma'}{\sigma^2}}\right\}$$

Chapter 3

Statistical Decision Theory I

3.1 Introduction

In this chapter we lay the basic statistical groundwork for the design of detectors of signals in noise. The approaches follow directly from the theory of hypothesis testing. In particular, we address the *simple hypothesis* testing problem in which the PDF for each assumed hypothesis is completely known. A much more complicated problem arises when the PDF has unknown parameters. We defer that discussion until Chapter 6. The primary approaches to simple hypothesis testing are the *classical* approach based on the Neyman-Pearson theorem and the *Bayesian* approach based on minimization of the Bayes risk. In many ways these approaches are analogous to the classical and Bayesian methods of statistical estimation theory. The particular method employed depends upon our willingness to incorporate prior knowledge about the probabilities of occurrence of the various hypotheses. The choice of an appropriate approach is therefore dictated by the problem at hand. Sonar and radar systems typically use the Neyman-Pearson criterion while communications and pattern recognition systems employ the Bayes risk.

3.2 Summary

The detector that maximizes the probability of detection for a given probability of false alarm is the likelihood ratio test of (3.3) as specified by the Neyman-Pearson theorem. The threshold is found from the false alarm constraint. For the mean-shifted Gauss-Gauss hypothesis testing problem, the detection performance is summarized by (3.10) and is monotonic with the deflection coefficient of (3.9). The performance of a detector can also be displayed using the receiver operating characteristics as discussed in Section 3.4. In some detection problems a subset of the data may be discarded as being irrelevant to a decision. This is discussed in Section 3.5, with the condition for irrelevant data being (3.11). To minimize the probability of decision error as given by (3.12), we should employ the detector of (3.13), where the threshold is now determined by the prior probabilities of the hypotheses. For equal prior probabilities of the hypotheses, the detector becomes the maximum likelihood detector of (3.14). More generally, the optimal decision rule that minimizes the probability of error is given by the maximum a posteriori probability detector of (3.16). A generalization of the minimum probability of error criterion is the Bayes risk as discussed in Section 3.7 with the detector given by (3.18). For multiple hypothesis testing the decision rule for minimizing the Bayes risk is given by (3.21). Specializing the result to the minimum probability of error criterion leads to the maximum a posteriori probability detector of (3.22) and the maximum likelihood detector of (3.24) for multiple hypothesis testing.

3.3 Neyman-Pearson Theorem

In discussing the Neyman-Pearson (NP) approach to signal detection we will center our discussion around a simple example of hypothesis testing. Assume that we observe a realization of a random variable whose PDF is either $\mathcal{N}(0,1)$ or $\mathcal{N}(1,1)$. The notation $\mathcal{N}(\mu,\sigma^2)$ denotes a Gaussian PDF with mean μ and variance σ^2. We must therefore determine if $\mu = 0$ or $\mu = 1$ based on a single observation $x[0]$. Each possible value of μ can be thought of as a hypothesis so that our problem is to choose among two competing hypotheses. These are summarized as follows:

$$\begin{aligned} \mathcal{H}_0 &: \mu = 0 \\ \mathcal{H}_1 &: \mu = 1 \end{aligned} \tag{3.1}$$

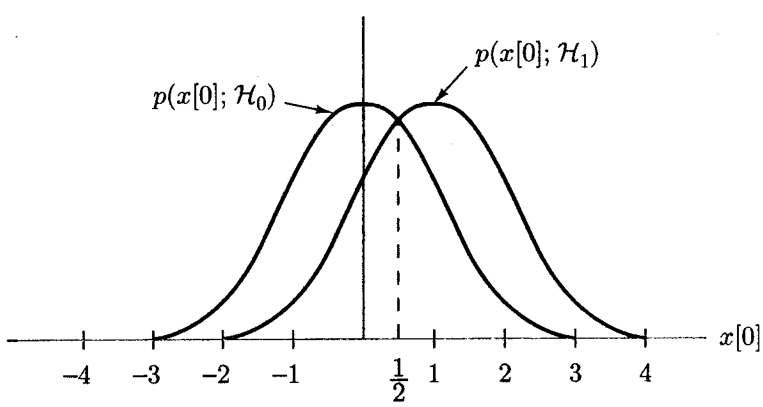


Figure 3.1. PDFs for hypothesis testing problem.

where $\mathcal{H}_0$ is referred to as the *null hypothesis* and $\mathcal{H}_1$ as the *alternative hypothesis*. This problem is known as a *binary* hypothesis test since we must choose between *two* hypotheses. The PDFs under each hypothesis are shown in Figure 3.1, with the difference in means causing the PDF under $\mathcal{H}_1$ to be shifted to the right. On the basis of a single sample it is difficult to determine which PDF generated it.

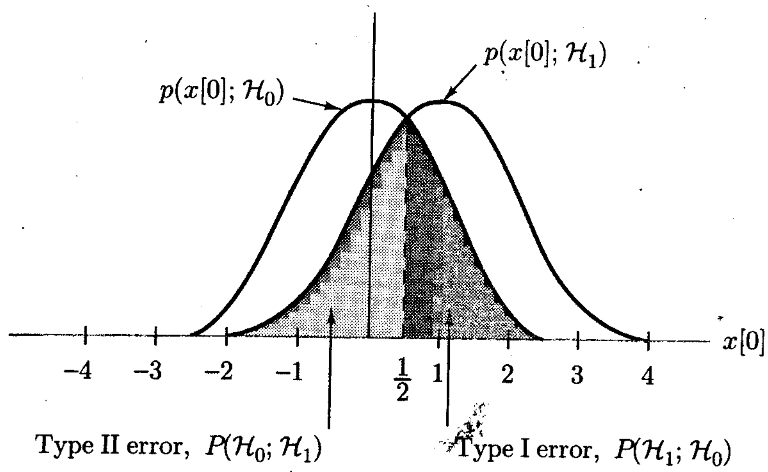


Figure 3.2. Possible hypothesis testing errors and their probabilities.

However, a reasonable approach might be to decide $\mathcal{H}_1$ if $x[0] > 1/2$. This is because if $x[0] > 1/2$, the observed sample is more *likely* if $\mathcal{H}_1$ is true. Or if $x[0] > 1/2$, we have from Figure 3.1 that $p(x[0];\mathcal{H}_1) > p(x[0];\mathcal{H}_0)$. Our detector then compares the observed datum value with 1/2, the latter being called the *threshold.* Note that with this scheme we can make two types of errors. If we decide $\mathcal{H}_1$ but $\mathcal{H}_0$ is true, we make a *Type I error.* On the other hand, if we decide $\mathcal{H}_0$ but $\mathcal{H}_1$ is true, we make a *Type II error.* These errors are illustrated in Figure 3.2. The notation $P(\mathcal{H}_i;\mathcal{H}_j)$ indicates the probability of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true. For example, $P(\mathcal{H}_1;\mathcal{H}_0) = \Pr\{x[0] > 1/2;\mathcal{H}_0\}$ and is shown as the darker area. These two errors are unavoidable to some extent but may be traded off against each other. To do so we need only change the threshold as shown in Figure 3.3. Clearly, the Type I error probability ($P(\mathcal{H}_1;\mathcal{H}_0)$) is decreased at the expense of increasing the Type II error probability ($P(\mathcal{H}_0;\mathcal{H}_1)$). *It is not possible to reduce both error probabilities simultaneously.* A typical approach then in designing an optimal detector is to hold one error probability fixed while minimizing the other. We choose to constrain $P(\mathcal{H}_1;\mathcal{H}_0)$ to a fixed value, say α. If we view the problem of (3.1) as an attempt to distinguish between the hypotheses

$$\begin{aligned}\mathcal{H}_0 &: x[0] = w[0]\\ \mathcal{H}_1 &: x[0] = s[0] + w[0]\end{aligned}$$

where $s[0] = 1$ and $w[0] \sim \mathcal{N}(0,1)$, then we have the signal detection problem. Deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true can be thought of as a false alarm. As a result, $P(\mathcal{H}_1;\mathcal{H}_0)$ is referred to as the *probability of false alarm* and is denoted by P_{FA}. Usually this is a small value, say 10^{-8}, in keeping with the disastrous effects that may ensue. For example, if we falsely say an enemy aircraft is present, we may initiate an attack. To design the optimal detector we then seek to minimize the other error $P(\mathcal{H}_0;\mathcal{H}_1)$ or equivalently to maximize $1 - P(\mathcal{H}_0;\mathcal{H}_1)$. The latter is just

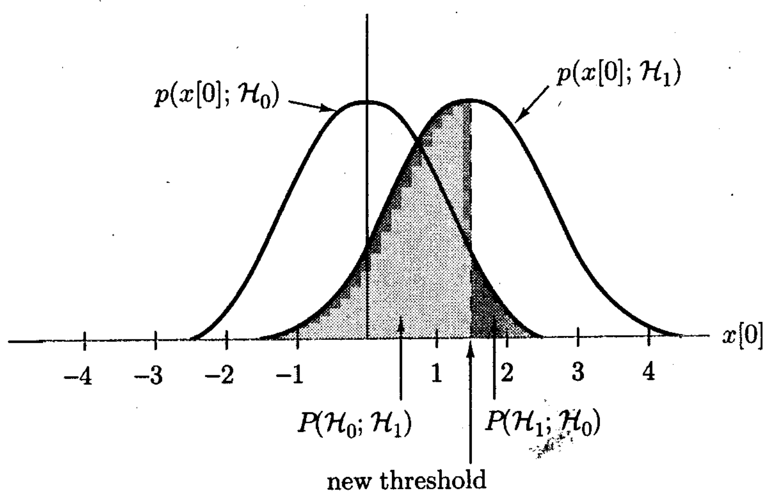


Figure 3.3. Trading off errors by adjusting threshold.

$P(\mathcal{H}_1;\mathcal{H}_1)$ and in keeping with the signal detection problem is called the *probability of detection.* It is denoted by P_D. This setup is termed the *Neyman-Pearson* (NP) approach to hypothesis testing or to signal detection. In summary, we wish to *maximize* $P_D = P(\mathcal{H}_1;\mathcal{H}_1)$ subject to the constraint $P_{FA} = P(\mathcal{H}_1;\mathcal{H}_0) = \alpha$.

Returning to the previous example we can constrain P_{FA} by choosing the threshold γ since

$$\begin{aligned} P_{FA} &= P(\mathcal{H}_1;\mathcal{H}_0) \\ &= \Pr\{x[0] > \gamma;\mathcal{H}_0\} \\ &= \int_\gamma^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \\ &= Q(\gamma). \end{aligned}$$

As an example, if $P_{FA} = 10^{-3}$, we have $\gamma = 3$. We therefore decide $\mathcal{H}_1$ if $x[0] > 3$. Furthermore, with this choice we have

$$\begin{aligned} P_D &= P(\mathcal{H}_1;\mathcal{H}_1) \\ &= \Pr\{x[0] > \gamma;\mathcal{H}_1\} \\ &= \int_\gamma^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt \\ &= Q(\gamma - 1) = Q(2) = 0.023. \end{aligned}$$

The question arises as to whether $P_D = 0.023$ is the maximum P_D for this problem. Our choice of the detector that decides $\mathcal{H}_1$ if $x[0] > \gamma$ was just a guess. Might there be a better approach?

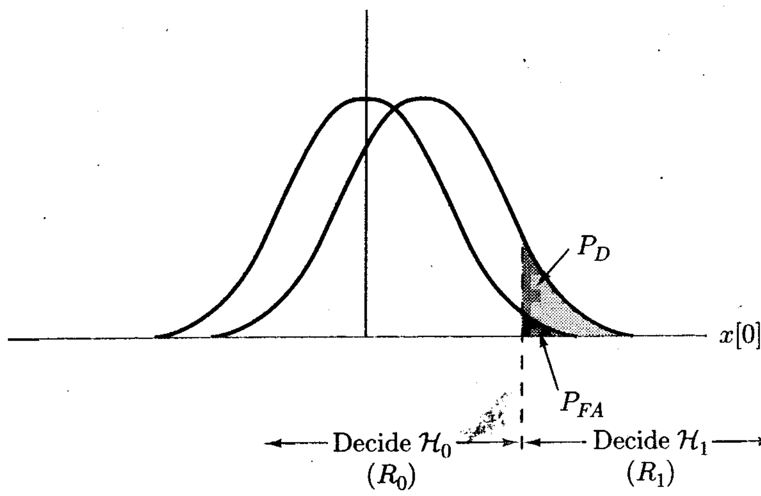


Figure 3.4. Decision regions and probabilities.

Before answering this question we first describe the operation of a detector in more general terms. The goal of a detector is to decide either $\mathcal{H}_0$ or $\mathcal{H}_1$ based on an observed set of data $\{x[0], x[1], \ldots, x[N-1]\}$. This is a mapping from each possible data set value into a decision. For the previous example the *decision regions* are shown in Figure 3.4. A detector then may be thought of as a mapping from the data values into a decision. In particular, let R_1 be the set of values in R^N that map into the decision $\mathcal{H}_1$ or

$$R_1 = \{\mathbf{x} : \text{decide } \mathcal{H}_1 \text{ or reject } \mathcal{H}_0\}.$$

This region is termed the *critical region* in statistics. The set of points in R^N that map into the decision $\mathcal{H}_0$ is the complement set of R_1 or $R_0 = \{\mathbf{x} : \text{decide } \mathcal{H}_0 \text{ or reject } \mathcal{H}_1\}$. Clearly, $R_0 \cup R_1 = R^N$ since R_0 and R_1 *partition* the data space. For the previous example the critical region was $x[0] > 3$. The P_{FA} constraint then becomes

$$P_{FA} = \int_{R_1} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x} = \alpha. \tag{3.2}$$

In statistics, α is termed the *significance level* or *size* of the test. Now there are many sets R_1 that satisfy (3.2) (see Problem 3.2). Our goal is to choose the one that maximizes

$$P_D = \int_{R_1} p(\mathbf{x}; \mathcal{H}_1) d\mathbf{x}.$$

In statistics, P_D is called the *power* of the test and the critical region that attains the maximum power is the *best critical region*. See Table 3.1 for a summary of the statistical terminology and the engineering equivalents.

The NP theorem tells us how to choose R_1 if we are given $p(\mathbf{x}; \mathcal{H}_0)$, $p(\mathbf{x}; \mathcal{H}_1)$, and α.

Statisticians	Engineers
Test statistic ($T(\mathbf{x})$) and threshold (γ)	Detector
Null hypothesis ($\mathcal{H}_0$)	Noise only hypothesis
Alternative hypothesis ($\mathcal{H}_1$)	Signal + noise hypothesis
Critical region	Signal present decision region
Type I error (decide $\mathcal{H}_1$ when $\mathcal{H}_0$ true)	False alarm (FA)
Type II error (decide $\mathcal{H}_0$ when $\mathcal{H}_1$ true)	Miss (M)
Level of significance or size of test (α)	Probability of false alarm (P_{FA})
Probability of Type II error (β)	Probability of miss (P_M)
Power of test ($1-\beta$)	Probability of detection (P_D)

Table 3.1. Cross-Reference of Statistical Terms for Binary Hypothesis Testing

Theorem 3.1 (Neyman-Pearson) *To maximize P_D for a given $P_{FA}=\alpha$ decide $\mathcal{H}_1$ if*

$$L(\mathbf{x}) = \frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma \tag{3.3}$$

where the threshold γ is found from

$$P_{FA} = \int_{\{\mathbf{x}:L(\mathbf{x})>\gamma\}} p(\mathbf{x};\mathcal{H}_0)d\mathbf{x} = \alpha.$$

The proof is given in Appendix 3A. The function $L(\mathbf{x})$ is termed the *likelihood ratio* since it indicates for each value of $\mathbf{x}$ the likelihood of $\mathcal{H}_1$ versus the likelihood of $\mathcal{H}_0$. The entire test of (3.3) is called the *likelihood ratio test* (LRT). We next illustrate the NP test with some examples.

Example 3.1 - Introductory Example (continued)

For the hypothesis test of (3.1) we can easily find the NP test. Assume that we require $P_{FA}=10^{-3}$. Then, from (3.3) we decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} = \frac{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x[0]-1)^2\right]}{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}x^2[0]\right]} > \gamma$$

or

$$\exp\left[-\frac{1}{2}(x^2[0]-2x[0]+1-x^2[0])\right] > \gamma$$

or finally

$$\exp\left(x[0]-\frac{1}{2}\right) > \gamma. \tag{3.4}$$

At this point we could determine γ from the false alarm constraint

$$P_{FA} = \Pr\left\{\exp\left(x[0] - \frac{1}{2}\right) > \gamma; \mathcal{H}_0\right\} = 10^{-3}.$$

This would require us to find the PDF of $\exp(x[0]-1/2)$. A much simpler approach is to note that the inequality of (3.4) is not changed if we take logarithms of both sides. This is because the logarithm is a monotonically increasing function (see Problem 3.3). Alternatively, since $\gamma > 0$, we can let $\gamma = \exp(\beta)$ so that we decide $\mathcal{H}_1$ if

$$\exp\left(x[0] - \frac{1}{2}\right) > \exp(\beta)$$

or

$$x[0] > \beta + \frac{1}{2} = \ln\gamma + \frac{1}{2}.$$

Letting $\gamma' = \ln\gamma + 1/2$ we decide $\mathcal{H}_1$ if $x[0] > \gamma'$. To explicitly find γ' (or equivalently γ) we use the P_{FA} constraint

$$\begin{aligned} P_{FA} = \Pr\{x[0] > \gamma'; \mathcal{H}_0\} &= 10^{-3} \\ \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt &= 10^{-3} \end{aligned}$$

so that $\gamma' = 3$. The NP test is to decide $\mathcal{H}_1$ if $x[0] > 3$. Thus, the detector of the previous example is indeed optimum in the NP sense in that it maximizes P_D. As before we find P_D as follows

$$\begin{aligned} P_D &= \Pr\{x[0] > 3; \mathcal{H}_1\} \\ &= \int_3^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt = 0.023. \end{aligned}$$

Note that the detection performance is poor. Although we have satisfied our false alarm constraint, we will only detect the signal a small fraction of the time. To improve the detection performance we can increase P_{FA}, employing the usual tradeoff. For example, if $P_{FA} = 0.5$, then the threshold is found from

$$0.5 = \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt$$

as $\gamma' = 0$. Then

$$\begin{aligned} P_D &= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt \end{aligned}$$

$$\begin{aligned} &= Q\left(\frac{0-1}{1}\right) = Q(-1) \\ &= 1 - Q(1) = 0.84. \end{aligned}$$

(Recall that if $x \sim \mathcal{N}(\mu, \sigma^2)$, the right-tail probability for a threshold γ' is $Q((\gamma' - \mu)/\sigma)$. See Chapter 2.) By changing the threshold we can trade off P_{FA} and P_D. This point is discussed further in the next section.

$\diamond$

Example 3.2 - DC Level in WGN

Now consider the more general signal detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where the signal is $s[n] = A$ for $A > 0$ and $w[n]$ is WGN with variance σ^2. The previous example is just a special case where $A = 1$, $N = 1$, and $\sigma^2 = 1$. Also, note that the current problem is actually a test of the mean of a multivariate Gaussian PDF. This is because under $\mathcal{H}_0$, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ while under $\mathcal{H}_1$, $\mathbf{x} \sim \mathcal{N}(A\mathbf{1}, \sigma^2\mathbf{I})$, where $\mathbf{1}$ is the vector of all ones. Hence, we have equivalently

$$\begin{array}{l} \mathcal{H}_0 : \boldsymbol{\mu} = \mathbf{0} \\ \mathcal{H}_1 : \boldsymbol{\mu} = A\mathbf{1}. \end{array}$$

We will often use this *parameter test of the PDF* interpretation in describing a signal detection problem. Now the NP detector decides $\mathcal{H}_1$ if

$$\frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > \gamma.$$

Taking the logarithm of both sides results in

$$-\frac{1}{2\sigma^2}\left(-2A\sum_{n=0}^{N-1}x[n] + NA^2\right) > \ln\gamma$$

which simplifies to

$$\frac{A}{\sigma^2}\sum_{n=0}^{N-1}x[n] > \ln\gamma + \frac{NA^2}{2\sigma^2}.$$

Since $A > 0$, we have finally

$$\frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\ln\gamma + \frac{A}{2} = \gamma'. \tag{3.5}$$

The NP detector compares the *sample mean* $\bar{x} = (1/N)\sum_{n=0}^{N-1} x[n]$ to a threshold γ'. This is intuitively reasonable since $\bar{x}$ may be thought of as an estimate of A. If the estimate is large and positive, then the signal is probably present. How large the estimate must be before we are willing to declare that a signal is present depends upon our concern that noise only may cause a large estimate. To avoid this possibility we adjust γ' to control P_{FA}, with larger threshold values reducing P_{FA} (as well as P_D).

To determine the detection performance we first note that the test statistic $T(\mathbf{x}) = (1/N)\sum_{n=0}^{N-1} x[n]$ is Gaussian under each hypothesis. The means and variances are

$$\begin{aligned} E(T(\mathbf{x});\mathcal{H}_0) &= E\left(\frac{1}{N}\sum_{n=0}^{N-1} w[n]\right) \\ &= \frac{1}{N}\sum_{n=0}^{N-1} E(w[n]) \\ &= 0. \end{aligned}$$

Similarly, $E(T(\mathbf{x});\mathcal{H}_1) = A$ and

$$\begin{aligned} \operatorname{var}(T(\mathbf{x});\mathcal{H}_0) &= \operatorname{var}\left(\frac{1}{N}\sum_{n=0}^{N-1} w[n]\right) \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1} \operatorname{var}(w[n]) \\ &= \frac{\sigma^2}{N}. \end{aligned}$$

Similarly, $\operatorname{var}(T(\mathbf{x});\mathcal{H}_1) = \sigma^2/N$ where we have noted that the noise samples are uncorrelated. Thus,

$$T(\mathbf{x}) \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(A, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_1. \end{cases}$$

We have then

$$\begin{aligned} P_{FA} &= \Pr\{T(\mathbf{x}) > \gamma';\mathcal{H}_0\} \\ &= Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right) \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} P_D &= \Pr\{T(\mathbf{x}) > \gamma'; \mathcal{H}_1\} \\ &= Q\left(\frac{\gamma' - A}{\sqrt{\sigma^2/N}}\right). \end{aligned} \tag{3.7}$$

We can relate P_D to P_{FA} more directly by noting that the Q function is monotonically decreasing since $1 - Q$ is a CDF, which is monotonically increasing. Thus, Q has an inverse that we denote as Q^{-1}. As a result, the threshold is found from (3.6) as

$$\gamma' = \sqrt{\frac{\sigma^2}{N}} Q^{-1}(P_{FA})$$

and

$$\begin{aligned} P_D &= Q\left(\frac{\sqrt{\sigma^2/N}Q^{-1}(P_{FA}) - A}{\sqrt{\sigma^2/N}}\right) \\ &= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{NA^2}{\sigma^2}}\right). \end{aligned} \tag{3.8}$$

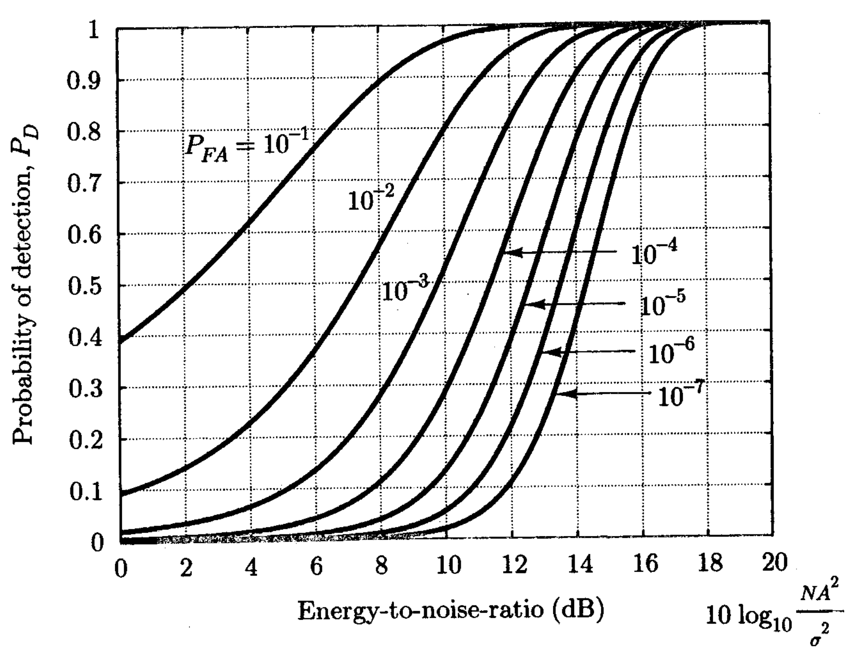


Figure 3.5. Detection performance for DC level in WGN.

It is seen that for a given P_{FA} the detection performance increases monotonically with NA^2/σ^2, which is the *signal energy-to-noise ratio* (ENR). An alternative interpretation is explored in Problem 3.5. The detection performance is shown in Figure 3.5 for various values of P_{FA}. It is sometimes convenient to display the detection curves on normal probability paper (see Chapter 2). This has the effect of straightening the curves when plotted versus $\sqrt{\text{ENR}}$ as shown in Figure 3.6. The advantage is an easier reading of the required ENR for a given P_D, especially for P_D's near one. The disadvantage is that the abscissa values are not in decibels (dB), which is customary in engineering. We will usually employ the former approach.

$\diamond$

The previous example illustrates a particularly useful hypothesis testing problem called the *mean-shifted Gauss-Gauss problem*. We observe the value of a test statistic T and decide $\mathcal{H}_1$ if $T > \gamma'$ and $\mathcal{H}_0$ otherwise. The PDF of T is assumed to be

$$T \sim \begin{cases} \mathcal{N}(\mu_0, \sigma^2) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mu_1, \sigma^2) & \text{under } \mathcal{H}_1 \end{cases}$$

where $\mu_1 > \mu_0$. Hence, we wish to decide between the two hypotheses that differ by a shift in the mean of T. In the previous example $T = \bar{x}$. For this type of detector

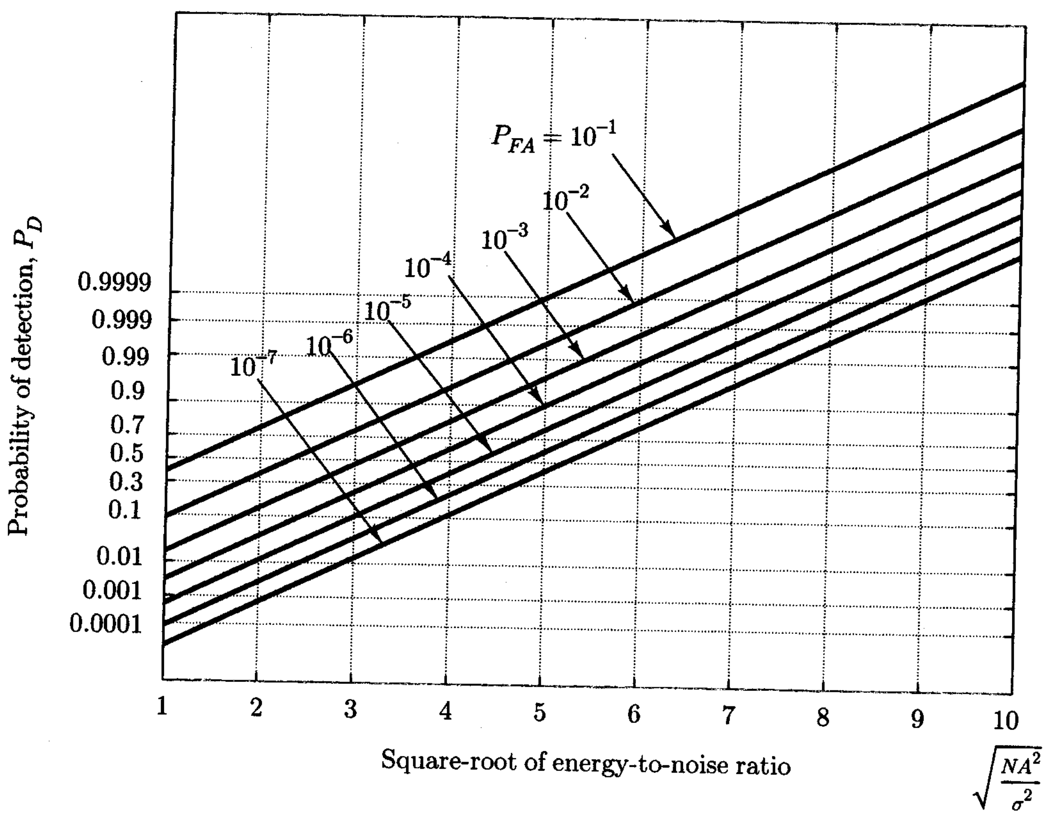


Figure 3.6. Detection performance for DC level in WGN-normal probability paper.

the detection performance is totally characterized by the *deflection coefficient* d^2. This is defined as

$$\begin{aligned} d^2 &= \frac{(E(T;\mathcal{H}_1) - E(T;\mathcal{H}_0))^2}{\mathrm{var}(T;\mathcal{H}_0)} \qquad (3.9) \\ &= \frac{(\mu_1 - \mu_0)^2}{\sigma^2}. \end{aligned}$$

In the case when $\mu_0 = 0$, $d^2 = \mu_1^2/\sigma^2$ may be interpreted as a signal-to-noise ratio (SNR). To verify the dependence of detection performance on d^2 we have that

$$\begin{aligned} P_{FA} &= \Pr\{T > \gamma'; \mathcal{H}_0\} \\ &= Q\left(\frac{\gamma' - \mu_0}{\sigma}\right) \\ P_D &= \Pr\{T > \gamma'; \mathcal{H}_1\} \\ &= Q\left(\frac{\gamma' - \mu_1}{\sigma}\right) \\ &= Q\left(\frac{\mu_0 + \sigma Q^{-1}(P_{FA}) - \mu_1}{\sigma}\right) \\ &= Q\left(Q^{-1}(P_{FA}) - \left(\frac{\mu_1 - \mu_0}{\sigma}\right)\right) \end{aligned}$$

and using (3.9) we have

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right) \qquad (3.10)$$

since $\mu_1 > \mu_0$. *The detection performance is therefore monotonic with the deflection coefficient.* We end this section with another example.

Example 3.3 - Change in Variance

This hypothesis testing example illustrates that a change in the variance of a Gaussian statistic can be used to distinguish between two hypotheses. We observe $x[n]$ for $n = 0, 1, \ldots, N-1$, where the $x[n]$'s are independent and identically distributed (IID). The latter qualification means that the first-order PDF for each $x[n]$ is the same. Assume that $x[n] \sim \mathcal{N}(0, \sigma_0^2)$ under $\mathcal{H}_0$ and $x[n] \sim \mathcal{N}(0, \sigma_1^2)$ under $\mathcal{H}_1$, where $\sigma_1^2 > \sigma_0^2$. Then the NP test is to decide $\mathcal{H}_1$ if

$$\frac{\dfrac{1}{(2\pi\sigma_1^2)^{\frac{N}{2}}} \exp\left(-\dfrac{1}{2\sigma_1^2}\displaystyle\sum_{n=0}^{N-1} x^2[n]\right)}{\dfrac{1}{(2\pi\sigma_0^2)^{\frac{N}{2}}} \exp\left(-\dfrac{1}{2\sigma_0^2}\displaystyle\sum_{n=0}^{N-1} x^2[n]\right)} > \gamma.$$

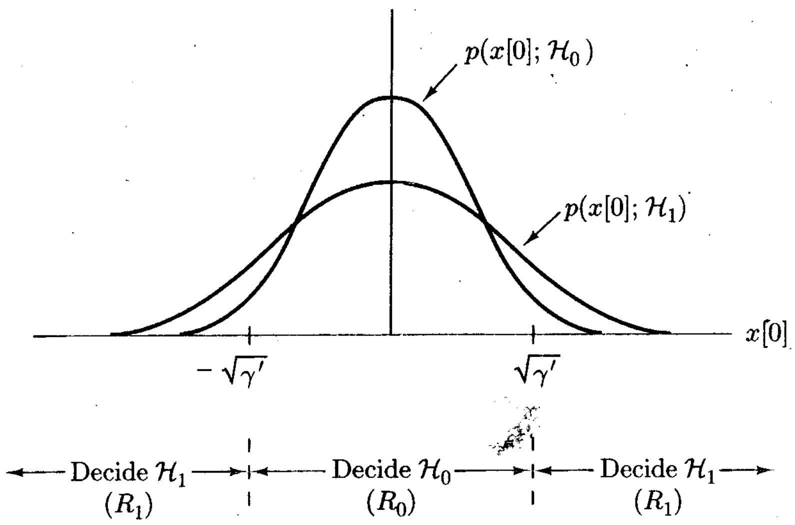


Figure 3.7. Decision regions for change in variance hypothesis test.

Taking logarithms of both sides we have

$$-\frac{1}{2}\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)\sum_{n=0}^{N-1} x^2[n] > \ln\gamma + \frac{N}{2}\ln\frac{\sigma_1^2}{\sigma_0^2}.$$

Since $\sigma_1^2 > \sigma_0^2$, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] > \gamma'$$

where

$$\gamma' = \frac{\frac{2}{N}\ln\gamma + \ln\frac{\sigma_1^2}{\sigma_0^2}}{\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}}.$$

The test statistic is just an estimate of the variance. We decide $\mathcal{H}_1$ if the power in the observed samples is large enough. In particular, if $N = 1$ we have a detector that decides $\mathcal{H}_1$ if $x^2[0] > \gamma'$ or equivalently if $|x[0]| > \sqrt{\gamma'}$. The decision regions are shown in Figure 3.7 and are seen to be plausible. The performance of this detector is examined in Problem 3.9 for $N = 2$ and in more generality in Chapter 5, where we discuss the energy detector.

◇

Note that for the DC level in WGN and the change in variance examples we distinguish between two hypotheses whose PDFs have different parameter values. We do so by estimating the parameter and comparing the estimated value to a threshold. This is not merely a coincidence but is due to the presence of a sufficient statistic [Kay-I 1993, Chapter 5]. In particular, assume that we observe

$\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$ from a PDF that is parameterized by θ. The PDF is denoted by $p(\mathbf{x}; \theta)$. (In the DC level in WGN example $\theta = A$.) We wish to test for the value of θ as

$$\begin{array}{l} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta = \theta_1. \end{array}$$

If a sufficient statistic exists for θ, then by the Neyman-Fisher factorization theorem [Kay-I 1993, Chapter 5] we can express the PDF as

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta) h(\mathbf{x})$$

where $T(\mathbf{x})$ is a sufficient statistic for θ. The NP test, which is

$$\frac{p(\mathbf{x}; \theta_1)}{p(\mathbf{x}; \theta_0)} > \gamma$$

then becomes

$$\frac{g(T(\mathbf{x}), \theta_1)}{g(T(\mathbf{x}), \theta_0)} > \gamma.$$

Clearly, the test will depend on the data only through $T(\mathbf{x})$. In the DC level in WGN example it can be shown (see Problem 3.10) that the sufficient statistic is $T(\mathbf{x}) = (1/N) \sum_{n=0}^{N-1} x[n]$ while in the change in variance example $T(\mathbf{x}) = (1/N) \sum_{n=0}^{N-1} x^2[n]$ is a sufficient statistic. In essence the sufficient statistic summarizes all the relevant information in the data about θ that is needed to make a decision. (See also Problem 3.11.) Furthermore, if $T(\mathbf{x})$ is an unbiased estimator of θ, then the detector will be based on an *estimate* of the unknown parameter. Unfortunately, sufficient statistics do not always exist, as our final example illustrates.

Example 3.4 - DC Level in NonGaussian Noise

Assume that under $\mathcal{H}_0$ we observe N IID samples $x[n] = w[n]$ for $n = 0, 1, \ldots, N-1$ from the noise PDF $p(w[n])$ while under $\mathcal{H}_1$ we observe $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$. Thus, under $\mathcal{H}_0$ we have

$$p(\mathbf{x}; \mathcal{H}_0) = \prod_{n=0}^{N-1} p(x[n])$$

and under $\mathcal{H}_1$ we have

$$p(\mathbf{x}; \mathcal{H}_1) = \prod_{n=0}^{N-1} p(x[n] - A).$$

The NP detector decides $\mathcal{H}_1$ if

$$\frac{\prod_{n=0}^{N-1} p(x[n] - A)}{\prod_{n=0}^{N-1} p(x[n])} > \gamma.$$

If the PDF of the noise is a *Gaussian mixture*

$$p(w[n]) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}w^2[n]\right) + \frac{1}{2}\frac{1}{\sqrt{4\pi}}\exp\left(-\frac{1}{4}w^2[n]\right)$$

then the detector becomes

$$\frac{\prod_{n=0}^{N-1}\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x[n]-A)^2\right) + \frac{1}{2}\frac{1}{\sqrt{4\pi}}\exp\left(-\frac{1}{4}(x[n]-A)^2\right)}{\prod_{n=0}^{N-1}\frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2[n]\right) + \frac{1}{2}\frac{1}{\sqrt{4\pi}}\exp\left(-\frac{1}{4}x^2[n]\right)} > \gamma.$$

No further simplification is possible due to the lack of a sufficient statistic for A. We will explore the nonGaussian detection problem further in Chapter 10.

◇

3.4 Receiver Operating Characteristics

An alternative way of summarizing the detection performance of a NP detector is to plot P_D versus P_{FA}. As an example, for the DC level in WGN we have from (3.6), (3.7), and (3.8)

$$\begin{aligned} P_{FA} &= Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right) \\ P_D &= Q\left(\frac{\gamma'-A}{\sqrt{\sigma^2/N}}\right) \end{aligned}$$

and

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$

where $d^2 = NA^2/\sigma^2$. The latter is shown in Figure 3.8 for $d^2 = 1$. Each point on the curve corresponds to a value of (P_{FA}, P_D) for a *given* threshold γ'. By adjusting γ' any point on the curve may be obtained. As expected as γ' increases, P_{FA} decreases but so does P_D and vice-versa. This type of performance summary is called the *receiver operating characteristic* (ROC). The ROC should always be above the 45° line (shown dashed in Figure 3.8). This is because the 45° ROC can be attained by a detector that bases its decision on flipping a coin, ignoring all the data. Consider the detector that decides $\mathcal{H}_1$ if a flipped coin comes up a head, where $\Pr\{\text{head}\} = p$. For a tail outcome we decide $\mathcal{H}_0$. Then,

$$\begin{aligned} P_{FA} &= \Pr\{\text{head}; \mathcal{H}_0\} \\ P_D &= \Pr\{\text{head}; \mathcal{H}_1\}. \end{aligned}$$

which becomes

$$\frac{\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right] \prod_{n=N}^{2N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x^2[n]\right]}{\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x[n]^2\right] \prod_{n=N}^{2N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x^2[n]\right]} > \gamma$$

or finally

$$\frac{p(\mathbf{x}_1; \mathcal{H}_1)}{p(\mathbf{x}_1; \mathcal{H}_0)} > \gamma$$

so that $\mathbf{x}_2$ is irrelevant to the detection problem. Thus, in practice, *for detection of signals in WGN we can limit the observation interval to the signal interval.* If, however, the noise is correlated, then for best performance we should also include noise samples from outside the signal interval in our detector.

The preceding discussion can be generalized using the NP theorem. The likelihood ratio is

$$\begin{aligned} L(\mathbf{x}_1, \mathbf{x}_2) &= \frac{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_1)}{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_0)} \\ &= \frac{p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_1)p(\mathbf{x}_1; \mathcal{H}_1)}{p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_0)p(\mathbf{x}_1; \mathcal{H}_0)}. \end{aligned}$$

It follows that if

$$p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_1) = p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_0) \tag{3.11}$$

then $L(\mathbf{x}_1, \mathbf{x}_2) = L(\mathbf{x}_1)$ and $\mathbf{x}_2$ is irrelevant to the detection problem. A special case occurs when $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent under either hypothesis and the PDF of $\mathbf{x}_2$ does not depend on the hypothesis. Then, (3.11) holds since $p(\mathbf{x}_2; \mathcal{H}_1) = p(\mathbf{x}_2; \mathcal{H}_0)$. The DC level in WGN with extra noise samples is an example. See also Problems 3.14 and 3.15.

3.6 Minimum Probability of Error

In some detection problems one can reasonably assign probabilities to the various hypotheses. In doing so, we express a prior belief in the likelihood of the hypotheses. An example is in digital communications in which the transmission of a "0" or "1" is equally likely. Then, it is reasonable to assign equal probabilities to $\mathcal{H}_0$ ("0" sent) and $\mathcal{H}_1$ ("1" sent). We say that $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$, where $P(\mathcal{H}_0)$, $P(\mathcal{H}_1)$ are the *prior probabilities* of the respective hypotheses. In other applications, such as sonar or radar, this is not possible. If one is attempting to detect an enemy submarine, then the likelihood of its appearance can usually not be determined. This type of approach, where we assign prior probabilities, is the Bayesian approach to hypothesis testing. It is completely analogous to the Bayesian philosophy of estimation theory in which a prior PDF is assigned to an unknown parameter.

With the Bayesian paradigm we can define a *probability of error* P_e as

$$\begin{aligned} P_e &= \Pr\{\text{decide } \mathcal{H}_0, \mathcal{H}_1 \text{ true}\} + \Pr\{\text{decide } \mathcal{H}_1, \mathcal{H}_0 \text{ true}\} \\ &= P(\mathcal{H}_0|\mathcal{H}_1)P(\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)P(\mathcal{H}_0) \end{aligned} \tag{3.12}$$

where $P(\mathcal{H}_i|\mathcal{H}_j)$ is the *conditional* probability that indicates the probability of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true. Note the slight distinction between $P(\mathcal{H}_i;\mathcal{H}_j)$ of the NP approach and $P(\mathcal{H}_i|\mathcal{H}_j)$ of the Bayesian approach. The former is the probability of deciding $\mathcal{H}_i$ if $\mathcal{H}_j$ *is true* with no probabilistic meaning assigned to the likelihood that $\mathcal{H}_j$ is true. The latter assumes that the outcome of a probabilistic experiment is observed to be $\mathcal{H}_j$ and that the probability of deciding $\mathcal{H}_i$ is *conditioned* on that outcome. Using the P_e criterion, the two errors are weighted appropriately to yield an overall error measure. Our goal will be to design a detector that minimizes P_e.

The derivation for the minimum P_e detector as a special case of the more general Bayesian detector is given in Appendix 3B. It is shown there that we should decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} = \gamma. \tag{3.13}$$

Similar to the NP test we compare the *conditional* likelihood ratio to a threshold. Here, however, the threshold is determined by the prior probabilities. If, as is commonly the case, the prior probabilities are equal, we decide $\mathcal{H}_1$ if

$$p(\mathbf{x}|\mathcal{H}_1) > p(\mathbf{x}|\mathcal{H}_0). \tag{3.14}$$

Equivalently, we choose the hypothesis with the larger conditional likelihood or the one that maximizes $p(\mathbf{x}|\mathcal{H}_i)$ for $i = 0, 1$. This is called the *maximum likelihood* (ML) detector. (Actually, we should term this the maximum *conditional* likelihood. We defer to common usage in not doing so.) An example follows.

Example 3.5 - DC Level in WGN - Minimum P_e Criterion

We have the detection problem

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

where $A > 0$ and $w[n]$ is WGN with variance σ^2. If this is a digital communication problem where we transmit either $s_0[n] = 0$ or $s_1[n] = A$ (called an on-off keyed (OOK) communication system), it is reasonable to assume $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$. The receiver that minimizes P_e is given by (3.13) with $\gamma = 1$. Hence, we decide $\mathcal{H}_1$

if

$$\frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > 1.$$

Taking logarithms yields

$$-\frac{1}{2\sigma^2}\left(-2A\sum_{n=0}^{N-1}x[n] + NA^2\right) > 0$$

or we decide $\mathcal{H}_1$ if $\bar{x} > A/2$. This is the same form of the detector as for the NP criterion except for the threshold (and, of course, the performance). To determine P_e we use (3.12) and note that

$$\bar{x} \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_0 \\ \mathcal{N}(A, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_1. \end{cases}$$

Thus

$$\begin{aligned} P_e &= \frac{1}{2}\left[P(\mathcal{H}_0|\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)\right] \\ &= \frac{1}{2}\left[\Pr\{\bar{x} < A/2|\mathcal{H}_1\} + \Pr\{\bar{x} > A/2|\mathcal{H}_0\}\right] \\ &= \frac{1}{2}\left[\left(1 - Q\left(\frac{A/2 - A}{\sqrt{\sigma^2/N}}\right)\right) + Q\left(\frac{A/2}{\sqrt{\sigma^2/N}}\right)\right] \end{aligned}$$

and since $Q(-x) = 1 - Q(x)$, we have finally

$$P_e = Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right). \tag{3.15}$$

The probability of error decreases monotonically with NA^2/σ^2, which is, of course, the deflection coefficient.

◇

Another form of the minimum P_e detector follows directly from (3.13). We decide $\mathcal{H}_1$ if

$$p(\mathbf{x}|\mathcal{H}_1)P(\mathcal{H}_1) > p(\mathbf{x}|\mathcal{H}_0)P(\mathcal{H}_0).$$

But from Bayes rule we have that

$$P(\mathcal{H}_i|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_i)P(\mathcal{H}_i)}{p(\mathbf{x})}$$

where the denominator $p(\mathbf{x})$ does not depend on the true hypothesis. In fact, $p(\mathbf{x})$ is just a normalizing factor that can be written as

$$p(\mathbf{x}) = p(\mathbf{x}|\mathcal{H}_0)P(\mathcal{H}_0) + p(\mathbf{x}|\mathcal{H}_1)P(\mathcal{H}_1).$$

As a result we decide $\mathcal{H}_1$ if

$$P(\mathcal{H}_1|\mathbf{x}) > P(\mathcal{H}_0|\mathbf{x}) \tag{3.16}$$

or we choose the hypothesis whose a posteriori (after the data are observed) probability is maximum. This detector, which minimizes P_e for any prior probability, is termed the *maximum a posteriori probability* (MAP) detector. Of course, for equal prior probabilities, the MAP detector reduces to the ML detector. The decision regions for the DC level in WGN with $N = 1$, $A = 1$, $\sigma^2 = 1$ are shown in Figure 3.10 for various prior probabilities (see Problem 3.16).

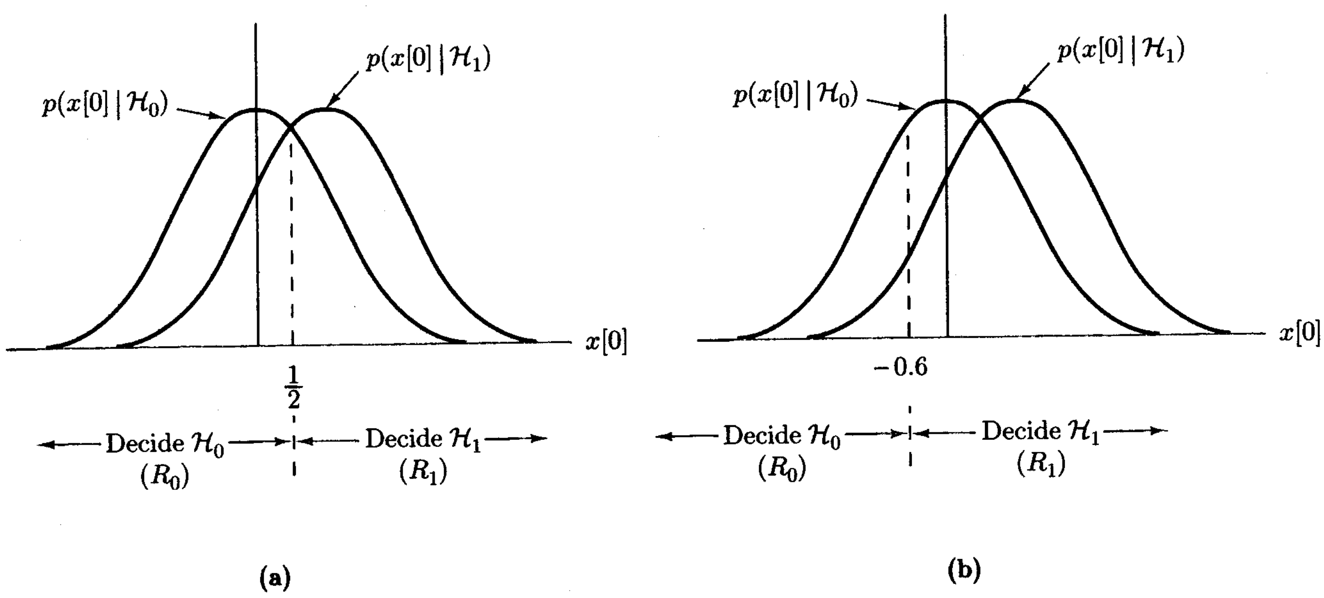


Figure 3.10. Effect of prior probability on decision regions (a) MAP detector with $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$ (b) MAP detector with $P(\mathcal{H}_0) = 1/4$, $P(\mathcal{H}_1) = 3/4$.

3.7 Bayes Risk

A generalization of the minimum P_e criterion assigns costs to each type of error. Suppose that we wish to design a system to automatically inspect a machine part. The result of the inspection is either to use the part in a product if it is deemed satisfactory or else to discard it. We could set up the hypothesis test

$$\begin{aligned}\mathcal{H}_0 &: \text{part is defective}\\ \mathcal{H}_1 &: \text{part is satisfactory}\end{aligned}$$

and assign costs to the errors. Let C_{ij} be the cost if we decide $\mathcal{H}_i$ but $\mathcal{H}_j$ is true. For example, we would probably want $C_{10} > C_{01}$. If we decide the part is satisfactory but it proves to be defective, the entire product may be defective and we incur a large cost (C_{10}). If, however, we decide that the part is defective when it is not, we incur the smaller cost of the part only (C_{01}). Once costs have been assigned, the decision rule is based on minimizing the expected cost or *Bayes risk* $\mathcal{R}$ defined as

$$\mathcal{R} = E(C) = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij}P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j). \tag{3.17}$$

Usually, if no error is made, we do not assign a cost so that $C_{00} = C_{11} = 0$. However, for convenience we will retain the more general form. Also, note that if $C_{00} = C_{11} = 0$, $C_{10} = C_{01} = 1$, then $\mathcal{R} = P_e$.

Under the reasonable assumption that $C_{10} > C_{00}$, $C_{01} > C_{11}$, the detector that minimizes the Bayes risk is to decide $\mathcal{H}_1$ if

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{(C_{10} - C_{00})P(\mathcal{H}_0)}{(C_{01} - C_{11})P(\mathcal{H}_1)} = \gamma. \tag{3.18}$$

See Appendix 3B for the proof. Once again, the conditional likelihood ratio is compared to a threshold.

3.8 Multiple Hypothesis Testing

We now consider the case where we wish to distinguish between M hypotheses, where $M > 2$. Such a problem arises quite frequently in communications, in which one of M signals must be detected. Also, pattern recognition systems make extensive use of the results in distinguishing between different patterns. In addition to signal detection, this problem also goes by the name of *classification* or *discrimination*. Although an NP criterion can be formulated for the M-ary hypothesis test, it seems to seldom be used in practice. The interested reader should consult [Lehmann 1959] for further details. More commonly the minimum P_e criterion or its generalization, the Bayes risk, is employed. We now consider the latter.

Assume that we now wish to decide among the M possible hypotheses $\{\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_{M-1}\}$. The cost assigned to the decision to choose $\mathcal{H}_i$ when $\mathcal{H}_j$ is true is denoted by C_{ij}. The expected cost or Bayes risk becomes

$$\mathcal{R} = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij}P(\mathcal{H}_i|\mathcal{H}_j)P(\mathcal{H}_j). \tag{3.19}$$

For the particular assignment

$$C_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \tag{3.20}$$

we have that $\mathcal{R} = P_e$. The decision rule that minimizes $\mathcal{R}$ is derived in Appendix 3C. There it is shown that we should choose the hypothesis that *minimizes*

$$C_i(\mathbf{x}) = \sum_{j=0}^{M-1} C_{ij} P(\mathcal{H}_j|\mathbf{x}) \tag{3.21}$$

over $i = 0, 1, \ldots, M-1$. To determine the decision rule that minimizes P_e we use (3.20). Then

$$\begin{aligned} C_i(\mathbf{x}) &= \sum_{\substack{j=0 \\ j\neq i}}^{M-1} P(\mathcal{H}_j|\mathbf{x}) \\ &= \sum_{j=0}^{M-1} P(\mathcal{H}_j|\mathbf{x}) - P(\mathcal{H}_i|\mathbf{x}). \end{aligned}$$

Since the first term is independent of i, $C_i(\mathbf{x})$ is minimized by *maximizing* $P(\mathcal{H}_i|\mathbf{x})$. Thus, the minimum P_e decision rule is to decide $\mathcal{H}_k$ if

$$P(\mathcal{H}_k|\mathbf{x}) > P(\mathcal{H}_i|\mathbf{x}) \qquad i \neq k. \tag{3.22}$$

As in the binary case we seek to maximize the a posteriori probability. This is the M-ary maximum a posteriori probability (MAP) decision rule. If, however, the prior probabilities are equal, then

$$\begin{aligned} P(\mathcal{H}_i|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{H}_i)P(\mathcal{H}_i)}{p(\mathbf{x})} \\ &= \frac{p(\mathbf{x}|\mathcal{H}_i)\frac{1}{M}}{p(\mathbf{x})} \end{aligned} \tag{3.23}$$

and to maximize $P(\mathcal{H}_i|\mathbf{x})$ we need only maximize $p(\mathbf{x}|\mathcal{H}_i)$. Hence, for equal prior probabilities we decide $\mathcal{H}_k$ if

$$p(\mathbf{x}|\mathcal{H}_k) > p(\mathbf{x}|\mathcal{H}_i) \qquad i \neq k. \tag{3.24}$$

This is the M-ary maximum likelihood (ML) decision rule.

Finally, observe from (3.23) that to maximize $P(\mathcal{H}_i|\mathbf{x})$ we can equivalently maximize $p(\mathbf{x}|\mathcal{H}_i)P(\mathcal{H}_i)$ since $p(\mathbf{x})$ does not depend on i. Equivalently then, the MAP rule maximizes

$$\ln p(\mathbf{x}|\mathcal{H}_i) + \ln P(\mathcal{H}_i).$$

An example follows.

Example 3.6 - Multiple DC Levels in WGN

Assume that we have the three hypotheses

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = -A + w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_1 : x[n] = w[n] & n = 0,1,\ldots,N-1 \\ \mathcal{H}_2 : x[n] = A + w[n] & n = 0,1,\ldots,N-1 \end{array}$$

where $A > 0$ and $w[n]$ is WGN with variance σ^2. Furthermore, if the prior probabilities are equal or $P(\mathcal{H}_0) = P(\mathcal{H}_1) = P(\mathcal{H}_2) = 1/3$, then the ML decision rule applies. Consider first the simple case of $N = 1$. We then have the PDFs shown in Figure 3.11. By symmetry it is clear from (3.24) that to minimize P_e we should decide $\mathcal{H}_0$ if $x[0] < -A/2$, $\mathcal{H}_1$ if $-A/2 < x[0] < A/2$, and $\mathcal{H}_2$ if $x[0] > A/2$. For multiple samples ($N > 1$) we cannot just plot the multivariate PDFs and observe the regions over which each one yields the maximum. Instead we need to derive a test statistic. The conditional PDF is

$$p(\mathbf{x}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A_i)^2\right]$$

where $A_0 = -A$, $A_1 = 0$, $A_2 = A$. To maximize $p(\mathbf{x}|\mathcal{H}_i)$ we can equivalently minimize

$$D_i^2 = \sum_{n=0}^{N-1}(x[n]-A_i)^2.$$

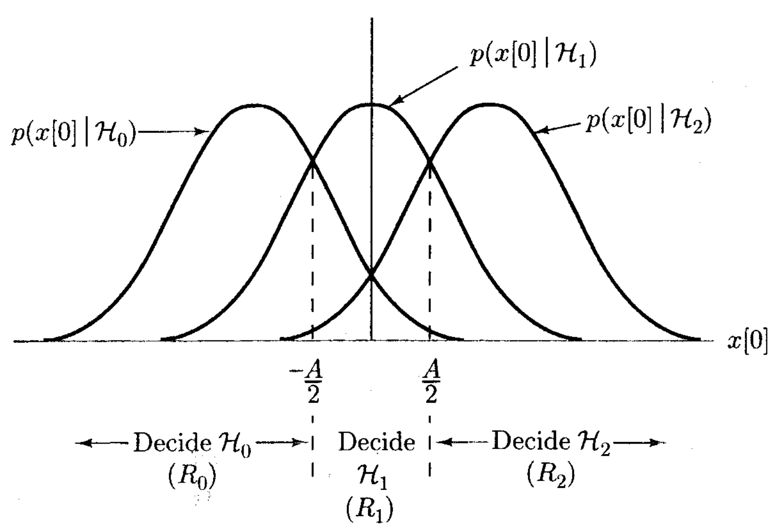


Figure 3.11. Decision regions for multiple DC levels in WGN ($N = 1$).

Chapter 4

Deterministic Signals

4.1 Introduction

The problem described in this chapter is the detection of a known signal in Gaussian noise. This is perhaps the simplest detection problem encountered because the resultant hypothesis test is a simple versus simple hypothesis. As discussed in Chapter 3, the optimal test is well known. If the probability of detection is to be maximized subject to a constant probability of false alarm, then the Neyman-Pearson criterion is employed, while to minimize the average cost, the Bayesian risk criterion is used. Furthermore, the resulting test statistic is a linear function of the data due to the Gaussian noise assumption, and therefore the performance of the detector is easily determined. The detector that evolves from these assumptions is termed the *matched filter*. It has found extensive use in applications in which the signal is under the designer's control so that the known signal assumption is valid. The salient example is in coherent communication systems.

4.2 Summary

The Neyman-Pearson detector for a known signal in WGN is the replica-correlator as given by (4.3). Alternatively, the matched filter implementation is given by (4.5) for use in (4.4). The detection performance is summarized in (4.14). When the Gaussian noise is not white, the Neyman-Pearson detector is given by (4.16) and its performance by (4.18). For the binary communication problem the optimal receiver is the minimum distance receiver given by (4.20). Its probability of error is determined by (4.25). In the M-ary communication case the optimal receiver is again a minimum distance receiver as summarized by (4.26). Its probability of error is found for orthogonal signals as (4.28). The general classical linear model is described in Section 4.6. For the case of known model parameters the Neyman-Pearson detector is given by (4.29) and is shown to be a special case of the detector for a known signal in colored noise. In Section 4.7 we apply the theory to multiple orthogonal signaling for communications, which illustrates the concept of channel capacity, and to pattern recognition for images.

4.3 Matched Filters

4.3.1 Development of Detector

We begin our discussion of optimal detection approaches by considering the problem of detecting a *known deterministic* signal in white Gaussian noise (WGN). The Neyman-Pearson (NP) criterion will be used, but as discussed in Chapter 3, the resulting test statistic will be identical to that obtained using the Bayesian risk criterion. Only the threshold and detection performance will differ. The detection problem is to distinguish between the hypotheses

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = s[n] + w[n] & n = 0, 1, \ldots, N-1 \end{array} \tag{4.1}$$

where the signal $s[n]$ is assumed known and $w[n]$ is WGN with variance σ^2. WGN is defined as a zero mean Gaussian noise process with autocorrelation function (ACF)

$$\begin{aligned} r_{ww}[k] &= E(w[n]w[n+k]) \\ &= \sigma^2\delta[k] \end{aligned}$$

where $\delta[k]$ is the discrete-time delta function ($\delta[k] = 1$ if $k = 0$ and $\delta[k] = 0$ for $k \neq 0$). Such a model can be derived from a continuous-time setup as described in [Kay-I 1993, pg. 54].

The NP detector decides $\mathcal{H}_1$ if the likelihood ratio exceeds a threshold or

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma \tag{4.2}$$

where $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$. Since

$$\begin{aligned} p(\mathbf{x}; \mathcal{H}_1) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n])^2\right] \\ p(\mathbf{x}; \mathcal{H}_0) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right] \end{aligned}$$

we have

$$L(\mathbf{x}) = \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}(x[n]-s[n])^2 - \sum_{n=0}^{N-1}x^2[n]\right)\right] > \gamma.$$

Taking the logarithm (a monotonically increasing transformation) of both sides does not change the inequality (see Section 3.3) so that

$$l(\mathbf{x}) = \ln L(\mathbf{x}) = -\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}(x[n]-s[n])^2 - \sum_{n=0}^{N-1}x^2[n]\right) > \ln\gamma.$$

We decide $\mathcal{H}_1$ if

$$\frac{1}{\sigma^2}\sum_{n=0}^{N-1}x[n]s[n] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}s^2[n] > \ln\gamma.$$

Since $s[n]$ is known (and thus not a function of the data), we may incorporate the energy term into the threshold to yield

$$T(\mathbf{x}) = \sum_{n=0}^{N-1}x[n]s[n] > \sigma^2\ln\gamma + \frac{1}{2}\sum_{n=0}^{N-1}s^2[n].$$

Calling the right-hand-side of the inequality a new threshold γ', we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1}x[n]s[n] > \gamma'. \tag{4.3}$$

This is our NP detector and as expected consists of a test statistic $T(\mathbf{x})$ (a function of the data) and a threshold γ', which is chosen to satisfy $P_{FA} = \alpha$ for a given α. We now determine the test statistic for some simple examples.

Example 4.1 - DC Level in WGN

Assume that $s[n] = A$ for some known level A, where $A > 0$. Then from (4.3), $T(\mathbf{x}) = A\sum_{n=0}^{N-1}x[n]$. An equivalent detector divides $T(\mathbf{x})$ by NA to decide $\mathcal{H}_1$ if

$$T'(\mathbf{x}) = \frac{1}{NA}T(\mathbf{x}) = \frac{1}{N}\sum_{n=0}^{N-1}x[n] = \bar{x} > \gamma''$$

where $\gamma'' = \gamma'/NA$. But this is just the sample mean detector discussed in Chapter 3. Its performance has also been described there. Note that if $A < 0$, the inequality reverses and we decide $\mathcal{H}_1$ if $\bar{x} < \gamma''$.

$\diamond$

Example 4.2 - Damped Exponential in WGN

Now let $s[n] = r^n$ for $0 < r < 1$. From (4.3) the test statistic is

$$T(\mathbf{x}) = \sum_{n=0}^{N-1}x[n]r^n$$

and is seen to weight the earlier samples more heavily than the later ones. This is due to the fact that the signal is decaying as n increases while the noise power remains constant. The signal-to-noise ratio (SNR) for the nth sample is $s^2[n]/\sigma^2 = r^{2n}/\sigma^2$, which decreases with n. The detection performance can easily be determined as described in Section 4.3.2.

◇

In general, the test statistic of (4.3) weights the data samples according to the value of the signal. More weight is reserved for the larger signal samples. Even negative signal samples are weighted in the same manner because by multiplying $x[n]$ by $s[n]$, negative samples yield a positive contribution to the sum. The detector of (4.3) is referred to as a *correlator* or *replica-correlator* since we correlate the received data with a replica of the signal. The detector is shown in Figure 4.1a. An alternative interpretation of the test statistic is based on relating the correlation process to the effect of a finite impulse response (FIR) filter on the data. Specifically, if $x[n]$ is the input to an FIR filter with impulse response $h[n]$, where $h[n]$ is nonzero

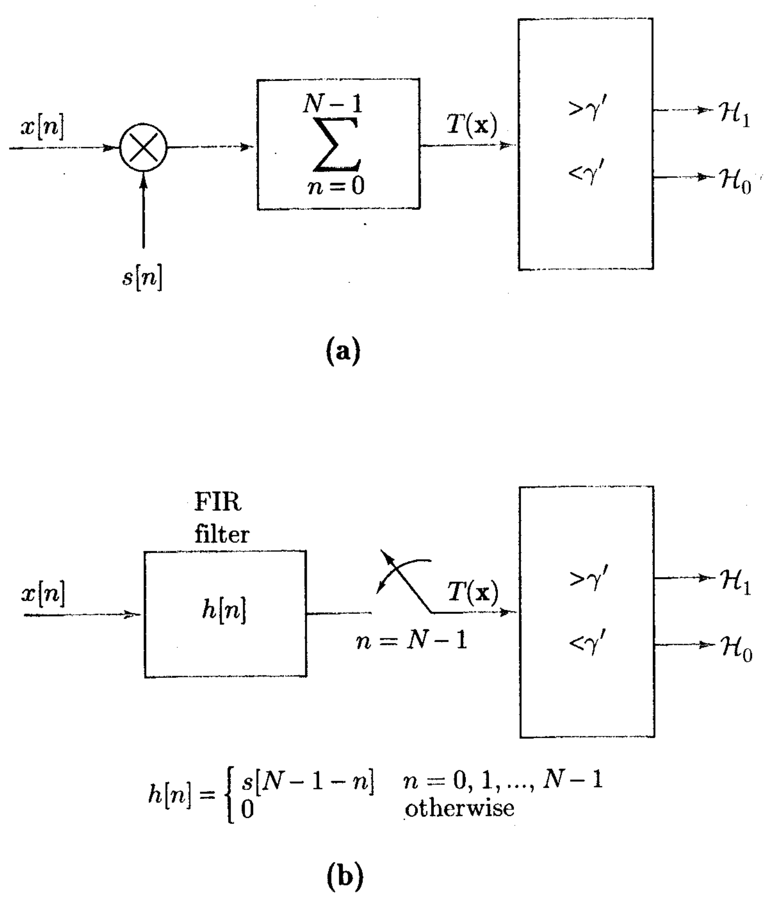


Figure 4.1. Neyman-Pearson detector for deterministic signal in white Gaussian noise (a) Replica-correlator (b) Matched filter.

for $n = 0, 1, \ldots, N-1$, then the output at time n is

$$y[n] = \sum_{k=0}^{n} h[n-k]x[k] \tag{4.4}$$

for $n \geq 0$. (For $n < 0$ the output is zero since we assume $x[n]$ is also nonzero only over the interval $[0, N-1]$.) If we let the impulse response be a "flipped around" version of the signal or

$$h[n] = s[N-1-n] \qquad n = 0, 1, \ldots, N-1 \tag{4.5}$$

then

$$y[n] = \sum_{k=0}^{n} s[N-1-(n-k)]x[k].$$

Now the output of the filter at time $n = N-1$ is

$$y[N-1] = \sum_{k=0}^{N-1} s[k]x[k]$$

which with a change of summation variable is identical to the replica-correlator. This implementation of the NP detector is shown in Figure 4.1b and is known as a *matched filter*. The filter impulse response is *matched* to the signal. Figure 4.2 shows some examples. The matched filter impulse response is obtained by flipping $s[n]$ about $n = 0$ and shifting to the right by $N-1$ samples.

Although the test statistic is obtained by sampling the matched filter output at time $n = N-1$, it is instructive to view the entire output of the matched filter. For the DC level signal shown in Figure 4.2b, the signal output is given via the convolution sum in Figure 4.3. Note that the signal output attains a maximum at the sampling time. This is true in general (see Problem 4.2). When noise is present, the maximum may be perturbed but it should be intuitively clear that the best detection performance will be obtained by sampling at $n = N-1$. If, however, the signal does not begin at $n = 0$, but we assume that it does and use the corresponding matched filter, poor detection performance may be obtained. An example is given in Problem 4.3. Thus, for signals with *unknown arrival times*, we cannot use the matched filter in its present form. In Chapter 7 we will see how to modify it to circumvent this problem.

The matched filter may also be viewed in the frequency domain. From (4.4) we have

$$y[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)X(f)\exp(j2\pi fn)df \tag{4.6}$$

where $H(f), X(f)$ are the discrete-time Fourier transforms of $h[n]$ and $x[n]$, respectively. But from (4.5), $H(f) = \mathcal{F}\{s[N-1-n]\}$, where $\mathcal{F}$ denotes the discrete-time Fourier transform. This may be shown to be

$$H(f) = S^*(f)\exp[-j2\pi f(N-1)] \tag{4.7}$$

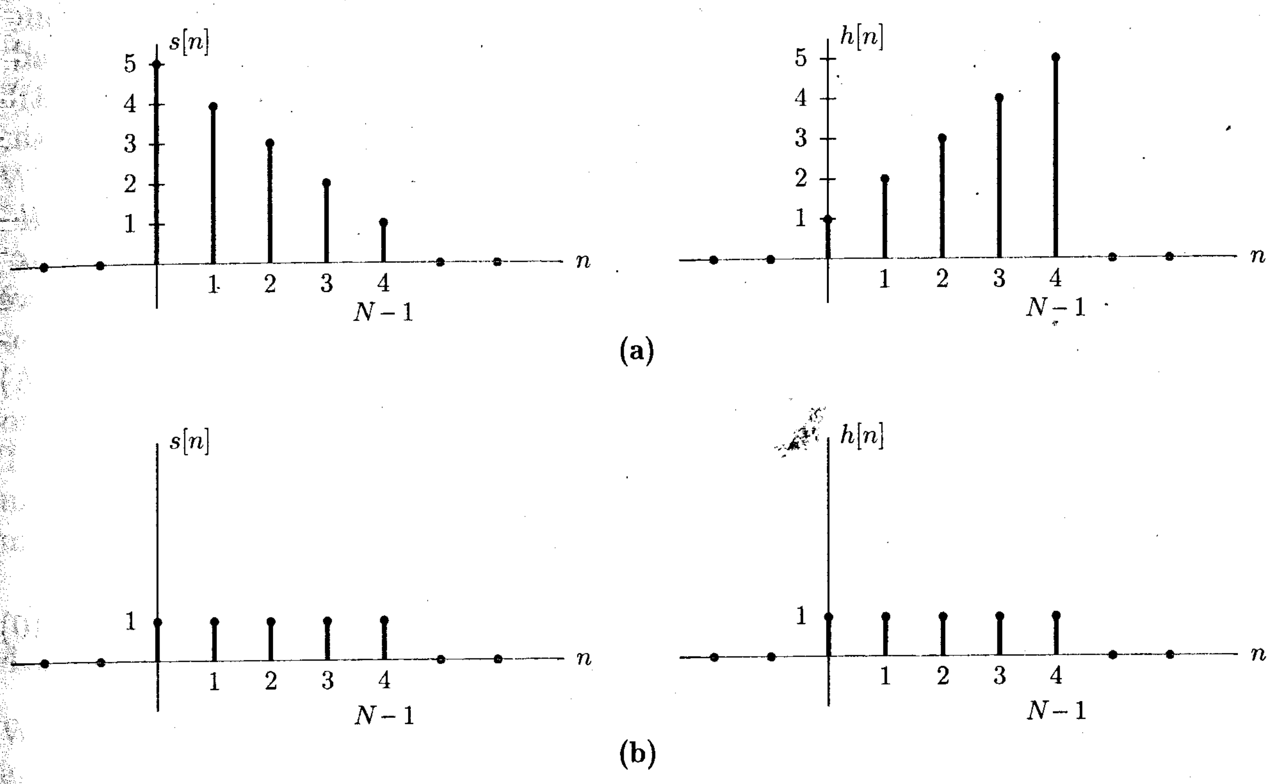


Figure 4.2. Examples of matched filter impulse response ($N = 5$).

so that (4.6) becomes

$$y[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} S^*(f)X(f)\exp[j2\pi f(n-(N-1))]df. \tag{4.8}$$

Sampling the output at $n = N - 1$ produces

$$y[N-1] = \int_{-\frac{1}{2}}^{\frac{1}{2}} S^*(f)X(f)df \tag{4.9}$$

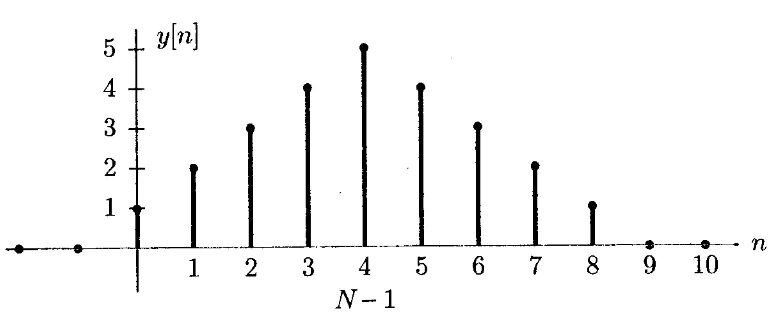


Figure 4.3. Matched filter signal output for DC level signal input ($N = 5$).

which could also have been obtained from the correlator implementation by using Parseval's theorem. Note from (4.7) that the matched filter emphasizes the bands where there is more signal power. Also, from (4.9) (or equivalently from (4.3)), when the noise is absent or $X(f) = S(f)$, the matched filter output is just the signal *energy*.

Another property of the matched filter is that *it maximizes the SNR at the output of an FIR filter.* In other words, we consider all detectors of the form of Figure 4.1b but with an arbitrary $h[n]$ over $[0, N-1]$ and zero otherwise. If we define the output SNR as

$$\begin{aligned}\eta &= \frac{E^2(y[N-1];\mathcal{H}_1)}{\operatorname{var}(y[N-1];\mathcal{H}_1)}\\ &= \frac{\left(\sum_{k=0}^{N-1} h[N-1-k]s[k]\right)^2}{E\left[\left(\sum_{k=0}^{N-1} h[N-1-k]w[k]\right)^2\right]}\end{aligned} \tag{4.10}$$

then the matched filter of (4.5) maximizes (4.10). To show this let $\mathbf{s} = [s[0]\, s[1] \ldots s[N-1]]^T$, $\mathbf{h} = [h[N-1]\, h[N-2] \ldots h[0]]^T$ and $\mathbf{w} = [w[0]\, w[1] \ldots w[N-1]]^T$. Then

$$\begin{aligned}\eta &= \frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{E\left[(\mathbf{h}^T\mathbf{w})^2\right]} = \frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{\mathbf{h}^T E(\mathbf{w}\mathbf{w}^T)\mathbf{h}}\\ &= \frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{\mathbf{h}^T\sigma^2\mathbf{I}\mathbf{h}} = \frac{1}{\sigma^2}\frac{\left(\mathbf{h}^T\mathbf{s}\right)^2}{\mathbf{h}^T\mathbf{h}}.\end{aligned}$$

By the Cauchy-Schwarz inequality (see Appendix 1)

$$(\mathbf{h}^T\mathbf{s})^2 \leq (\mathbf{h}^T\mathbf{h})(\mathbf{s}^T\mathbf{s})$$

with equality if and only if $\mathbf{h} = c\mathbf{s}$, where c is any constant. Hence

$$\eta \leq \frac{1}{\sigma^2}\mathbf{s}^T\mathbf{s}$$

with equality if and only if $\mathbf{h} = c\mathbf{s}$. The maximum output SNR is attained for (letting $c = 1$)

$$h[N-1-n] = s[n] \qquad n = 0, 1, \ldots, N-1$$

or equivalently

$$h[n] = s[N-1-n] \qquad n = 0, 1, \ldots, N-1$$

which is the matched filter. Note that the maximum SNR is $\eta_{\max} = \mathbf{s}^T\mathbf{s}/\sigma^2 = \mathcal{E}/\sigma^2$, where $\mathcal{E}$ is the signal energy. One would expect that the performance of the matched filter detector would increase monotonically with $\eta_{\max}$. This will be seen to be the case in the next section. For the detection problem of a known deterministic signal in WGN, the NP criterion and the maximum SNR criterion both lead to the matched filter detector. Since we know that the NP criterion produces an optimal detector, the maximum SNR criterion also does so for these model assumptions. However, for nonGaussian noise, as an example, the matched filter is not optimal in the NP sense but may still be said to maximize the SNR at the output of a *linear* FIR filter. (Actually, the matched filter more generally maximizes the SNR at the output of *any* linear filter, i.e., even ones with an infinite impulse response (see Problem 4.4).) This is because in the nonGaussian noise case the NP detector is not linear. For moderate deviations of the noise PDF from Gaussian, however, the matched filter may still be a good approximation. See Chapter 10 for a further discussion.

4.3.2 Performance of Matched Filter

We now determine the detection performance. Specifically, we will derive P_D for a given P_{FA}. Using the replica-correlator form we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]s[n] > \gamma'.$$

Under either hypothesis $x[n]$ is Gaussian and since $T(\mathbf{x})$ is a linear combination of Gaussian random variables, $T(\mathbf{x})$ is also Gaussian. Let $E(T;\mathcal{H}_i)$ and $\mathrm{var}(T;\mathcal{H}_i)$ denote the expected value and variance of $T(\mathbf{x})$ under $\mathcal{H}_i$. Then

$$\begin{aligned}
E(T;\mathcal{H}_0) &= E\left(\sum_{n=0}^{N-1} w[n]s[n]\right) = 0 \\
E(T;\mathcal{H}_1) &= E\left(\sum_{n=0}^{N-1} (s[n]+w[n])s[n]\right) = \mathcal{E} \\
\mathrm{var}(T;\mathcal{H}_0) &= \mathrm{var}\left(\sum_{n=0}^{N-1} w[n]s[n]\right) \\
&= \sum_{n=0}^{N-1} \mathrm{var}(w[n])s^2[n] \\
&= \sigma^2 \sum_{n=0}^{N-1} s^2[n] = \sigma^2\mathcal{E}
\end{aligned}$$

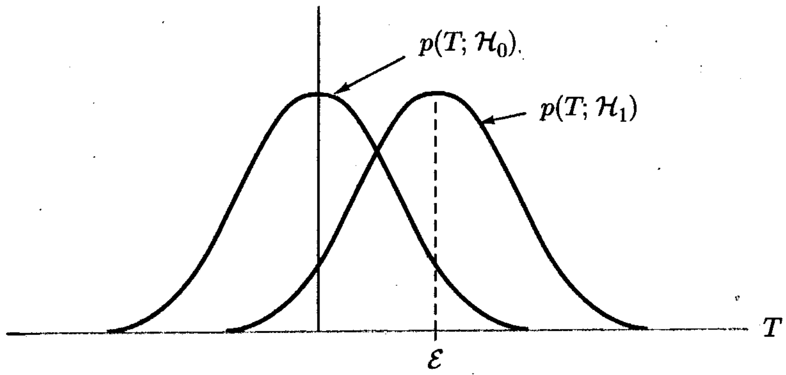


Figure 4.4. PDFs of matched filter test statistic.

where we have used the fact that the $w[n]$'s are uncorrelated. Similarly, $\mathrm{var}(T;\mathcal{H}_1) = \sigma^2\mathcal{E}$. Thus,

$$T \sim \begin{cases} \mathcal{N}(0,\sigma^2\mathcal{E}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mathcal{E},\sigma^2\mathcal{E}) & \text{under } \mathcal{H}_1 \end{cases} \tag{4.11}$$

which is shown in Figure 4.4. Note that the scaled test statistic $T' = T/\sqrt{\sigma^2\mathcal{E}}$ has the PDF

$$T' \sim \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\sqrt{\mathcal{E}/\sigma^2},1) & \text{under } \mathcal{H}_1. \end{cases}$$

It follows then that the detection performance must increase with $\sqrt{\mathcal{E}/\sigma^2}$ or with $\mathcal{E}/\sigma^2$ since as $\mathcal{E}/\sigma^2$ increases the PDFs retain the same shape but move further apart. To verify this we have from (4.11)

$$\begin{aligned} P_{FA} &= \Pr\{T > \gamma'; \mathcal{H}_0\} \\ &= Q\left(\frac{\gamma'}{\sqrt{\sigma^2\mathcal{E}}}\right) \end{aligned} \tag{4.12}$$

$$\begin{aligned} P_D &= \Pr\{T > \gamma'; \mathcal{H}_1\} \\ &= Q\left(\frac{\gamma'-\mathcal{E}}{\sqrt{\sigma^2\mathcal{E}}}\right) \end{aligned} \tag{4.13}$$

where

$$\begin{aligned} Q(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \\ &= 1 - \Phi(x) \end{aligned}$$

and $\Phi(x)$ is the CDF for a $\mathcal{N}(0,1)$ random variable. Now, since a CDF is a monotonically increasing function, $Q(x)$ must be monotonically decreasing and so has an inverse function $Q^{-1}(\cdot)$. (See Chapter 2 for an evaluation of $Q(\cdot)$ and $Q^{-1}(\cdot)$.) This allows us to write (4.12) as

$$\gamma' = \sqrt{\sigma^2 \mathcal{E}} Q^{-1}(P_{FA}).$$

Substituting in (4.13) produces

$$\begin{aligned} P_D &= Q\left(\frac{\sqrt{\sigma^2\mathcal{E}}Q^{-1}(P_{FA})}{\sqrt{\sigma^2\mathcal{E}}} - \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right) \\ &= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{\mathcal{E}}{\sigma^2}}\right). \end{aligned} \tag{4.14}$$

As asserted, as $\eta = \mathcal{E}/\sigma^2$ increases, the argument of $Q(\cdot)$ decreases, and P_D increases. The detection performance is summarized in Figure 4.5. The key parameter is the SNR at the matched filter output, which is the energy-to-noise ratio (ENR) or $\mathcal{E}/\sigma^2$. These detection curves are identical to those in Figure 3.5. The only difference is that for the DC level in WGN, the energy is NA^2 so that $\eta = NA^2/\sigma^2$. It is apparent from the curves that to improve the detection performance we can always increase P_{FA} and/or increase the ENR. The ENR is increased by increasing the signal energy, either by increasing the signal level (A) or signal duration (N) in the case of a DC level signal. *The shape of the signal does not affect the detection performance.* For example, the two signals shown in Figure 4.6 both yield the

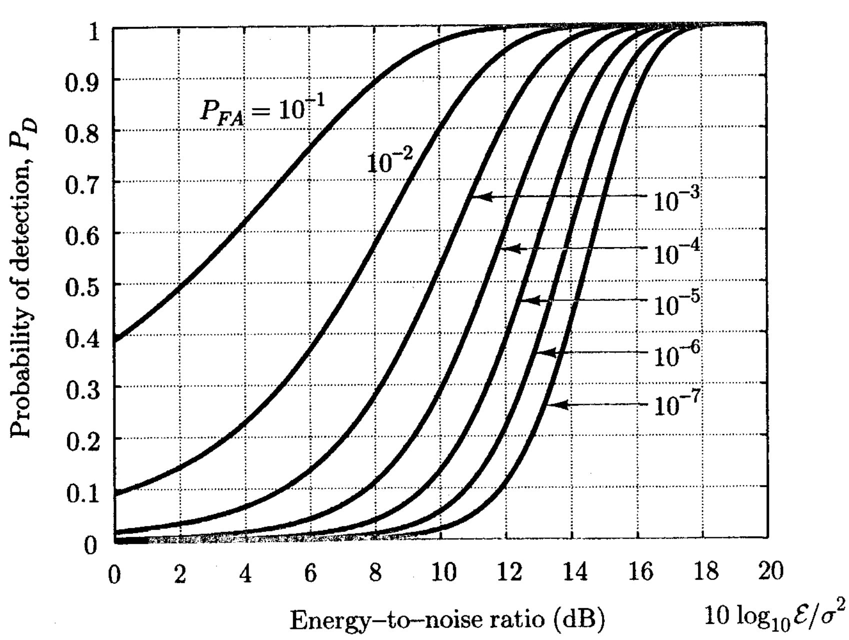


Figure 4.5. Detection performance of matched filter.

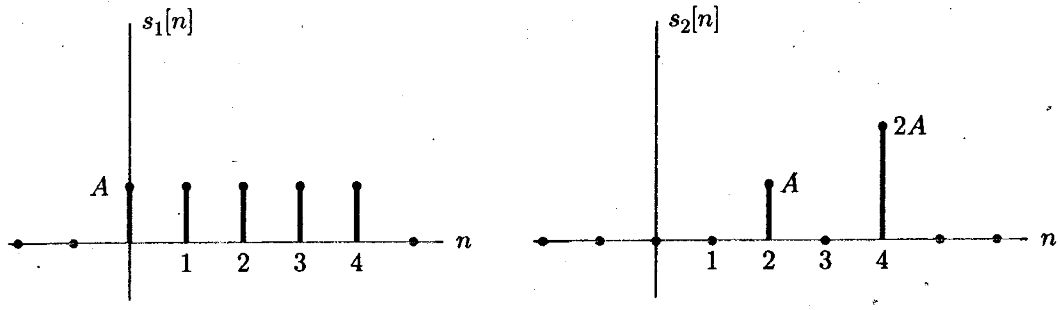


Figure 4.6. Signals yielding same detection performance.

same detection performance since their energies are equal. As we will see shortly, the signal shape does become an important design consideration when the noise is *colored.*

The use of a matched filter for signal detection leads to the concept of *processing gain.* Processing gain can be viewed as the advantage of making a decision based on a test statistic, which is an optimal combination of the data, as opposed to a decision based on a direct examination of the data. It is defined as the SNR of the optimal test statistic, i.e, the *processed data,* divided by the SNR of a single data sample, i.e., the unprocessed data. As an example, consider a DC level in WGN. If one were to attempt to detect the signal based on only a single sample, then the performance would be given by (4.14) with

$$\eta_{\text{in}} = \frac{\mathcal{E}}{\sigma^2} = \frac{A^2}{\sigma^2}.$$

By using a matched filter to process N samples the performance improves to

$$\eta_{\text{out}} = \frac{\mathcal{E}}{\sigma^2} = \frac{NA^2}{\sigma^2}.$$

The improvement in SNR or $\eta_{\text{out}}/\eta_{\text{in}}$ is termed the processing gain (PG) or

$$\begin{aligned} PG &= 10\log_{10}\frac{\eta_{\text{out}}}{\eta_{\text{in}}} \\ &= 10\log_{10} N \qquad \text{dB}. \end{aligned}$$

For example, from Figure 4.5 the ENR required to attain $P_D = 0.5$ for a $P_{FA} = 10^{-3}$ is about 10 dB. If we required our detector to have $P_D = 0.95$, the ENR must be increased by 4 dB. This is attainable if we increase N by a factor of 2.5, since then the PG will increase by 4 dB. Processing gain considerations are important in designing sonar/radar systems. See also Problem 4.9 for a further discussion.

Lastly, the matched filter test statistic has the PDF of the mean-shifted Gauss-Gauss problem ($\mathcal{N}(\mu_1, \sigma^2)$ versus $\mathcal{N}(\mu_2, \sigma^2)$). Thus, the deflection coefficient characterizes the detection performance as described in Chapter 3. Recall that the definition of this coefficient is

$$d^2 = \frac{(E(T;\mathcal{H}_1) - E(T;\mathcal{H}_0))^2}{\text{var}(T;\mathcal{H}_0)}. \tag{4.15}$$

From (4.11) this is $d^2 = \mathcal{E}/\sigma^2$ or just the SNR at the matched filter output. Not surprising then, the performance as given by (4.14) increases monotonically with d^2. It should be emphasized that only for the mean-shifted Gauss-Gauss problem does the deflection coefficient completely characterize the detection performance. However, it is often used for *approximate* detection performance evaluation for other detection problems.

4.4 Generalized Matched Filters

The matched filter is an optimal detector for a known signal in WGN. In many situations, however, the noise is more accurately modeled as *correlated* noise. Thus, we now assume that $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C}$ is the covariance matrix. If the noise is modeled as wide sense stationary (WSS), then $\mathbf{C}$ has the special form of a symmetric Toeplitz matrix (see Appendix 1). This is because for a zero mean WSS random process

$$[\mathbf{C}]_{mn} = \text{cov}(w[m], w[n]) = E(w[m]w[n]) = r_{ww}[m-n].$$

Hence, the elements along any NW to SE diagonal of $\mathbf{C}$ are the same. For nonstationary noise $\mathbf{C}$ will be an arbitrary covariance matrix.

For WGN we assumed that the received data samples were observed over the signal interval $[0, N-1]$. (It is assumed that the signal is zero outside the interval $[0, N-1]$.) The data samples outside this interval are irrelevant since the noise outside the interval is independent of the noise inside the interval and so can be discarded. Hence, in the presence of WGN there is no loss in detection performance by assuming the observation interval to be $[0, N-1]$ (see also Problem 4.4). However, for a signal embedded in correlated noise this is not the case. Improved performance may be obtained by incorporating data samples outside the signal interval into the detector. Problem 4.5 discusses a somewhat contrived example where perfect detection is possible by doing so. The reader may also wish to refer to Problem 4.14 in which the optimal detector is described for an infinite observation interval. In the discussion to follow we assume that the observed data samples are $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$. In practice, larger data intervals appear to be seldom used due to the nonstationarity of the noise samples. However, the reader should bear in mind that improved detection performance may be possible in some situations.

To determine the NP detector we again determine the likelihood ratio test (LRT) with

$$p(\mathbf{x};\mathcal{H}_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{s})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{s})\right]$$

$$p(\mathbf{x}; \mathcal{H}_0) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right]$$

where we have noted that under $\mathcal{H}_0$, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and under $\mathcal{H}_1$, $\mathbf{x} \sim \mathcal{N}(\mathbf{s}, \mathbf{C})$. In the WGN case $\mathbf{C} = \sigma^2\mathbf{I}$ and the PDFs reduce to the ones in Section 4.3.1. Now, we decide $\mathcal{H}_1$ if

$$l(\mathbf{x}) = \ln \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \ln \gamma.$$

But

$$\begin{aligned} l(\mathbf{x}) &= -\frac{1}{2}\left[(\mathbf{x}-\mathbf{s})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{s}) - \mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right] \\ &= -\frac{1}{2}\left[\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} + \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} - \mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right] \\ &= \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} - \frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} \end{aligned}$$

or by incorporating the non-data-dependent term into the threshold we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} > \gamma'. \tag{4.16}$$

Note that for WGN $\mathbf{C} = \sigma^2\mathbf{I}$ the detector reduces to

$$\frac{\mathbf{x}^T\mathbf{s}}{\sigma^2} > \gamma'$$

or

$$\mathbf{x}^T\mathbf{s} = \sum_{n=0}^{N-1} x[n]s[n] > \gamma''$$

as before. The detector of (4.16) is referred to as a *generalized replica-correlator* or *matched filter*. It may be viewed as a replica-correlator where the replica is the modified signal $\mathbf{s}' = \mathbf{C}^{-1}\mathbf{s}$. Then

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{C}^{-1}\mathbf{s} = \mathbf{x}^T\mathbf{s}'$$

so that we correlate with the modified signal. An example follows.

Example 4.3 - Uncorrelated Noise with Unequal Variances

If $w[n] \sim \mathcal{N}(0, \sigma_n^2)$ and the $w[n]$'s are uncorrelated, then $\mathbf{C} = \mathrm{diag}(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{N-1}^2)$ and $\mathbf{C}^{-1} = \mathrm{diag}(1/\sigma_0^2, 1/\sigma_1^2, \ldots, 1/\sigma_{N-1}^2)$. Hence, from (4.16) we decide $\mathcal{H}_1$ if

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} \frac{x[n]s[n]}{\sigma_n^2} > \gamma'.$$