Saleem A. Kassam

Signal Detection in Non-Gaussian Noise

Springer-Verlag

世界图书出版公司

Saleem A. Kassam

Signal Detection in Non-Gaussian Noise

Consulting Editor: John B. Thomas

With 38 Illustrations

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo
世界图书出版公司
北京・广州・上海・西安

A Dowden &
Culver Book

Saleem A. Kassam
Department of Electrical Engineering
University of Pennsylvania
Philadelphia, PA 19104
USA

Library of Congress Cataloging-in-Publication Data
Kassam, Saleem A.
Signal detection in non-Gaussian noise
(Springer texts in electrical engineering)
Bibliography p.
Includes index
1. Signal theory (Telecommunication) 2. Signal
detection. I. Title. II. Series.
TK5102.5.K357 1987 621.38'0436 87-32197

Reprinted by World Publishing Corporation, Beijing, 1992
for distribution and sale in The People's Republic of China only
ISBN 7-5062-1312-5

ISBN 0-387-96680-3 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-96680-3 Springer-Verlag Berlin Heidelberg New York

PREFACE

This book contains a unified treatment of a class of problems of signal detection theory. This is the detection of signals in additive noise which is not required to have Gaussian probability density functions in its statistical description. For the most part the material developed here can be classified as belonging to the general body of results of *parametric* theory. Thus the probability density functions of the observations are assumed to be known, at least to within a finite number of unknown parameters in a known functional form. Of course the focus is on noise which is not Gaussian; results for Gaussian noise in the problems treated here become special cases. The contents also form a bridge between the classical results of signal detection in Gaussian noise and those of *nonparametric* and *robust* signal detection, which are not considered in this book.

Three canonical problems of signal detection in additive noise are covered here. These allow between them formulation of a range of specific detection problems arising in applications such as radar and sonar, binary signaling, and pattern recognition and classification. The simplest to state and perhaps the most widely studied of all is the problem of detecting a *completely known deterministic signal* in noise. Also considered here is the detection of a *random non-deterministic signal* in noise. Both of these situations may arise for observation processes of the low-pass type and also for processes of the band-pass type. Spanning the gap between the known and the random signal detection problems is that of detection of a deterministic signal with random parameters in noise. The important special case of this treated here is the detection of *phase-incoherent narrowband signals* in narrowband noise.

There are some specific assumptions that we proceed under throughout this book. One of these is that ultimately all the data which our detectors operate on are *discrete sequences* of observation components, as opposed to being continuous-time waveforms. This is a reasonable assumption in modern implementations of signal detection schemes. To be able to treat non-Gaussian noise with any degree of success and obtain explicit, canonical, and useful results, a more stringent assumption is needed. This is the *independence* of the discrete-time additive noise components in the observation processes. There do exist many situations under which this assumption is at least a good approximation.

With the same objective of obtaining explicit canonical results of practical appeal, this book concentrates on *locally optimum* and *asymptotically optimum* detection schemes. These criteria are appropriate in detection of weak signals (the low signal-to-noise-ratio case), for which the use of optimum detectors is particularly meaningful and necessary to extract the most in detection performance.

Most of the development given here has not been given detailed exposition in any other book covering signal detection theory and applications, and many of the results have appeared relatively recently in technical journals. In presenting this material it is assumed only that the reader has had some exposure to the elements of statistical inference and of signal detection in Gaussian noise. Some of the basic statistics background needed to appreciate the rest of the development is reviewed in Chapter 1. This book should be suitable for use in a first graduate course on signal detection, to supplement the classical material on signal detection in Gaussian noise. Chapters 2-4 may be used to provide a fairly complete introduction to the known signal detection problem. Chapters 5 and 6 are on the detection of narrowband known and phase-incoherent signals, respectively, and Chapter 7 is on random signal detection. A more advanced course on signal detection may also be based on this book, with supplementary material on nonparametric and robust detection if desired. This book should also be useful as a reference to those active in research, as well as to those interested in the application of signal detection theory to problems arising in practice.

The completion of this book has been made possible through the understanding and help of many individuals. My family has been most patient and supportive. My graduate students have been very stimulating and helpful. Prashant Gandhi has been invaluable in getting many of the figures ready. For the excellent typing of the drafts and the final composition, I am grateful to Drucilla Spanner and to Diane Griffiths. Finally, I would like to acknowledge the research support I have received from the Air Force Office of Scientific Research and the Office of Naval Research, which eventually got me interested in writing this book.

Saleem A. Kassam
University of Pennsylvania
Philadelphia, PA 19104

TABLE OF CONTENTS

Chapter 1

ELEMENTS OF STATISTICAL HYPOTHESIS TESTING

1.1 Introduction

The signal processing problem which is the object of our study in this book is that of detecting the presence of a signal in noisy observations. Signal detection is a function that has to be implemented in a variety of applications, the more obvious ones being in radar, sonar, and communications. By viewing signal detection problems as problems of binary hypothesis testing in statistical inference, we get a convenient mathematical framework within which we can treat in a unified way the analysis and synthesis of signal detectors for different specific situations. The theory and results in mathematical statistics pertaining to binary hypothesis-testing problems are therefore of central importance to us in this book. In this first chapter we review some of these basic statistics concepts. In addition, we will find in this chapter some further results of statistical hypothesis testing with which the reader may not be as familiar, but which will be of use to us in later chapters.

We begin in Section 1.2 with a brief account of the basic concepts and definitions of hypothesis-testing theory, which leads to a discussion of most powerful tests and the Neyman-Pearson lemma in Section 1.3. In Section 1.4 this important result is generalized to yield the structures of locally optimum tests, which we will make use of throughout the rest of this book. Section 1.5 reviews briefly the Bayesian approach to construction of tests for hypotheses. We shall not be using the Bayesian framework very much except in Chapters 2 and 5, where we shall develop locally optimum Bayes' detectors for detection of known signals in additive noise.

In the last section of this chapter we will introduce a measure which we will make use of quite extensively in comparing the performances of different detectors for various signal detection problems in the following chapters. While we will give a more general discussion of the *asymptotic relative efficiency* and the *efficacy* in Section 1.6, these measures will be introduced and discussed in detail for the specific problem of detection of a known signal in additive noise in Section 2.4 of Chapter 2. Readers may find it beneficial to postpone study of Section 1.6 until after Chapter 2 has been read; they may then better appreciate the applicability of the ideas and results of this section.

1.2 Basic Concepts of Hypothesis Testing

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be a random vector of observations with joint probability density function (pdf) $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$, where θ is a parameter of the density function. Any specific realization $\mathbf{x} = (x_1, x_2, ..., x_n)$ of $\mathbf{X}$ will be a point in $\mathbb{R}^n$, where $\mathbb{R}$ is the set of all real numbers. In binary hypothesis-testing problems we have to decide between one of two hypotheses, which we will label as H and K, about the pdf $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$, given an observation vector in $\mathbb{R}^n$. Let Θ be the set of all possible values of θ; we usually identify H with one subset Θ_H of θ values and K with a disjoint subset Θ_K, so that $\Theta = \Theta_H \cup \Theta_K$. This may be expressed formally as

$$H: \mathbf{X} \text{ has pdf } f_{\mathbf{X}}(\mathbf{x} \mid \theta) \text{ with } \theta \in \Theta_H \tag{1-1}$$

$$K: \mathbf{X} \text{ has pdf } f_{\mathbf{X}}(\mathbf{x} \mid \theta) \text{ with } \theta \in \Theta_K \tag{1-2}$$

If Θ_H and Θ_K are made up of single elements, say θ_H and θ_K, respectively, we say that the hypotheses are *simple*; otherwise, the hypotheses are *composite*. If Θ can be viewed as a subset of $\mathbb{R}^p$ for a finite integer p, the pdf $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$ is completely specified by the finite number p of real components of θ, and we say that our hypotheses are *parametric*.

A *test* for the hypothesis H against K may be specified as a partition of the sample space $S = \mathbb{R}^n$ of observations into disjoint subsets S_H and S_K, so that $\mathbf{x}$ falling in S_H leads to acceptance of H, with K accepted otherwise. This may also be expressed by a *test function* $\delta(\mathbf{x})$ which is defined to have value $\delta(\mathbf{x}) = 1$ for $\mathbf{x} \in S_K$ and value $\delta(\mathbf{x}) = 0$ for $\mathbf{x} \in S_H$. The value of the test function is defined to be the probability with which the hypothesis K, the *alternative hypothesis*, is accepted. The hypothesis H is called the *null hypothesis*.

More generally, the test function can be allowed to take on probability values in the closed interval $[0,1]$. A test based on a test function taking on values inside $[0,1]$ is called a *randomized test*.

The *power function* $p(\theta \mid \delta)$ of a test based on a test function δ is defined for $\theta \in \Theta_H \cup \Theta_K$ as

$$\begin{aligned} p(\theta \mid \delta) &= E\{\delta(\mathbf{X}) \mid \theta\} \\ &= \int_{\mathbb{R}^n} \delta(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x} \mid \theta)\, d\mathbf{x} \end{aligned} \tag{1-3}$$

Thus it is the probability with which the test will accept the alternative hypothesis K for any particular parameter value θ. When

θ is in Θ_H the value of $p(\theta \mid \delta)$ gives the probability of an error, that of accepting K when H is correct. This is called a *type I error*, and depends on the particular value of θ in Θ_H. The *size* of a test is the quantity

$$\alpha = \sup_{\theta \in \Theta_H} p(\theta \mid \delta) \tag{1-4}$$

which may be considered as being the best upper bound on the type I error probability of the test.

In signal detection the null hypothesis is often a noise-only hypothesis, and the alternative hypothesis expresses the presence of a signal in the observations. For a detector D implementing a test function $\delta(\mathbf{x})$ the power function evaluated for any θ in Θ gives a *probability of detection* of the signal. Thus in later chapters we will use the notation $p_d(\theta \mid D)$ for the power function of a detector D, and in discussing the probability of detection at a particular value of the parameter θ in Θ_K (or for a simple alternative hypothesis K) we will use for it the notation p_d. The size of a detector is often called its *false-alarm probability*. This usage is encountered specifically when the noise-only null hypothesis is simple, and the notation for this probability is p_f.

1.3 Most Powerful Tests and the Neyman-Pearson Lemma

Given a problem of binary hypothesis testing such as defined by (1-1) and (1-2), the question arises as to how one may define and then construct an optimum test. Ideally, one would like to have a test for which the power function $p(\theta \mid \delta)$ has values close to zero for θ in Θ_H, and has values close to unity for θ in Θ_K. These are, however, conflicting requirements. We can instead impose the condition that the size α of any acceptable test be no larger than some reasonable level α_0, and subject to this condition look for a test for which $p(\theta \mid \delta)$, evaluated at a particular value θ_K of θ in Θ_K, has its largest possible value. Such a test is *most powerful* at level α_0 in testing H against the simple alternative $\theta = \theta_K$ in Θ_K; its test function $\delta^*(\mathbf{x})$ satisfies

$$\sup_{\theta \in \Theta_H} p(\theta \mid \delta^*) \leq \alpha_0 \tag{1-5}$$

and

$$p(\theta_K \mid \delta^*) \geq p(\theta_K \mid \delta) \tag{1-6}$$

for all other test functions $\delta(\mathbf{x})$ of size less than or equal to α_0. In most cases of interest a most powerful level α_0 test satisfies (1-5) with equality, so that its size is $\alpha = \alpha_0$.

For a simple null hypothesis H when $\theta = \theta_H$ is the only parameter value in Θ_H, the condition (1-5) becomes $p(\theta_H \mid \delta^*) \leq \alpha_0$ or $p_f \leq \alpha_0$, subject to which p_d at $\theta = \theta_K$ is maximized. For this problem of testing a simple H against a simple K, a fundamental result of Neyman and Pearson (called the Neyman-Pearson lemma) gives the structure of the most powerful test. We state the result here as a theorem:

Theorem 1: Let $\delta(\mathbf{x})$ be a test function of the form

$$\delta(\mathbf{x}) = \begin{cases} 1 & , \quad f_{\mathbf{X}}(\mathbf{x} \mid \theta_K) > t f_{\mathbf{X}}(\mathbf{x} \mid \theta_H) \\ r(\mathbf{x}) & , \quad f_{\mathbf{X}}(\mathbf{x} \mid \theta_K) = t f_{\mathbf{X}}(\mathbf{x} \mid \theta_H) \\ 0 & , \quad f_{\mathbf{X}}(\mathbf{x} \mid \theta_K) < t f_{\mathbf{X}}(\mathbf{x} \mid \theta_H) \end{cases} \tag{1-7}$$

for some constant $t \geq 0$ and some function $r(\mathbf{x})$ taking on values in [0,1]. Then the resulting test is most powerful at level equal to its size for H: $\theta = \theta_H$ versus K: $\theta = \theta_K$.

In addition to the above sufficient condition for a most powerful test it can be shown that conversely, if a test is known to be most powerful at level equal to its size, then its test function must be of the form (1-7) except perhaps on a set of $\mathbf{x}$ values of probability measure zero. Additionally, we may always require $r(\mathbf{x})$ in (1-7) to be a constant r in [0,1]. Finally, we note that we are always guaranteed the existence of such a test for H versus K, of given size α [Lehmann, 1959, Ch. 3].

From the above result we see that generally the structure of a most powerful test may be described as one comparing a *likelihood ratio* to a constant *threshold*,

$$\frac{f_{\mathbf{X}}(\mathbf{x} \mid \theta_K)}{f_{\mathbf{X}}(\mathbf{x} \mid \theta_H)} > t \tag{1-8}$$

in deciding if the alternative K is to be accepted. If the likelihood ratio on the left-hand side of (1-8) equals the threshold value t, the alternative K may be accepted with some probability r (the randomization probability). The constants t and r may be evaluated to obtain a desired size α using knowledge of the distribution function of the likelihood ratio under H.

When the alternative hypothesis K is composite we may look for a test which is *uniformly most powerful* (UMP) in testing H against K, that is, one which is most powerful for H against each $\theta = \theta_K$ in Θ_K. While UMP tests can be found in some cases, notably in many situations involving Gaussian noise in signal detection, such tests do not exist for many other problems of interest. One option in such situations is to place further restrictions on the class of acceptable or admissible tests in defining a most powerful test; for example, a requirement of *unbiasedness* or of *invariance* may be imposed [Lehmann, 1959, Ch. 4-6]. As an alternative, other performance criteria based on the power function may be employed. We will consider one such criterion, leading to *locally optimum* or *locally most powerful* tests for composite alternatives, in the next section. One approach to obtaining reasonable tests for composite hypotheses is to use maximum-likelihood estimates $\hat{\theta}_H$ and $\hat{\theta}_K$ of the parameter θ, obtained under the constraints $\theta \in \Theta_H$ and $\theta \in \Theta_K$, respectively, in place of θ_H and θ_K in (1-8). The resulting test is called a *generalized likelihood ratio* test or simply a *likelihood ratio* test (see, for example, [Bickel and Doksum, 1977, Ch. 6]).

1.4 Local Optimality and the Generalized Neyman-Pearson Lemma

Let us now consider the approach to construction of tests for composite alternative hypotheses which we will use almost exclusively in the rest of our development on signal detection in non-Gaussian noise. In this approach attention is concentrated on alternatives $\theta = \theta_K$ in Θ_K which are close, in the sense of a metric or distance, to the null-hypothesis parameter value $\theta = \theta_H$. Specifically, let θ be a real-valued parameter with value $\theta = \theta_0$ defining the simple null hypothesis and let $\theta > \theta_0$ define the composite alternative hypothesis. Consider the class of all tests based on test functions $\delta(\mathbf{x})$ of a particular desired size α for $\theta = \theta_0$ against $\theta > \theta_0$, and assume that the power functions $p(\theta \mid \delta)$ of these tests are continuous and also continuously differentiable at $\theta = \theta_0$. Then if we are interested primarily in performance for alternatives which are close to the null hypothesis, we can use as a measure of performance the slope of the power function at $\theta = \theta_0$, that is,

$$\begin{aligned} p'(\theta_0 \mid \delta) &= p'(\theta \mid \delta)\big|_{\theta=\theta_0} \\ &= \frac{d}{d\theta}\, p(\theta \mid \delta)\Big|_{\theta=\theta_0} \end{aligned} \tag{1-9}$$

From among our class of tests of size α, the test based on $\delta^*(\mathbf{x})$ which uniquely maximizes $p'(\theta_0 \mid \delta)$ has a power function satisfying

$$p(\theta \mid \delta^*) \geq p(\theta \mid \delta), \ \theta_0 < \theta < \theta_{max} \tag{1-10}$$

for some $\theta_{max} > \theta_0$. Such a test is called a *locally* most powerful or *locally optimum* (LO) test for $\theta = \theta_0$ against $\theta > \theta_0$. It is clearly of interest in situations such as the weak-signal case in signal detection, when the alternative-hypothesis parameter values of primary concern are those which define pdf's $f_X(x \mid \theta)$ close to the null-hypothesis noise-only pdf $f_X(x \mid \theta_H)$.

The following generalization of the Neyman-Pearson fundamental result of Theorem 1 can be used to obtain the structure of an LO test:

Theorem 2: Let $g(x)$ and $h_1(x), h_2(x), \ldots, h_m(x)$ be real-valued and integrable functions defined on R^n. Let an integrable function $\delta(x)$ on R^n have the characteristics

$$\delta(x) = \begin{cases} 1 \quad , & g(x) > \sum_{i=1}^{m} t_i h_i(x) \\ r(x) \quad , & g(x) = \sum_{i=1}^{m} t_i h_i(x) \\ 0 \quad , & g(x) < \sum_{i=1}^{m} t_i h_i(x) \end{cases} \tag{1-11}$$

for a set of constants $t_i \geq 0$, $i = 1,2,\ldots,m$, and where $0 \leq r(x) \leq 1$. Define, for $i = 1,2,\ldots,m$, the quantities

$$\alpha_i = \int_{R^n} \delta(x)\, h_i(x)\, dx \tag{1-12}$$

Then from within the class of all test functions satisfying the m constraints (1-12), the function $\delta(x)$ defined by (1-11) maximizes $\int_{R^n} \delta(x) g(x)\, dx$.

A more complete version of the above theorem, and its proof, may be found in [Lehmann, 1959, Ch. 3]; Ferguson [1967, Ch. 5] also discusses the use of this result.

To use the above result in finding an LO test for $\theta = \theta_0$ against $\theta > \theta_0$ defining Θ_H and Θ_K in (1-1) and (1-2), respectively, let us write (1-9) explicitly as

$$\begin{aligned} p'(\theta_0 \mid \delta) &= \frac{d}{d\theta} \int_{R^n} \delta(x) f_X(x \mid \theta)\, dx \Bigg|_{\theta = \theta_0} \\ &= \int_{R^n} \delta(x) \frac{d}{d\theta} f_X(x \mid \theta) \Bigg|_{\theta = \theta_0} dx \end{aligned} \tag{1-13}$$

assuming that our pdf's are such as to allow the interchange of the order in which limits and integration operations are performed. Taking $m = 1$ and identifying $h_1(\mathbf{x})$ with $f_{\mathbf{X}}(\mathbf{x} \mid \theta_0)$ and $g(\mathbf{x})$ with $\frac{d}{d\theta} f_{\mathbf{X}}(\mathbf{x} \mid \theta)\Big|_{\theta=\theta_0}$ in Theorem 2, we are led to the locally optimum test which accepts the alternative $K: \theta > \theta_0$ when

$$\frac{\frac{d}{d\theta} f_{\mathbf{X}}(\mathbf{x} \mid \theta)\Big|_{\theta=\theta_0}}{f_{\mathbf{X}}(\mathbf{x} \mid \theta_0)} > t \tag{1-14}$$

where t is the test threshold value which results in a size-α test satisfying

$$E\{\delta(\mathbf{X}) \mid H: \theta = \theta_0\} = \alpha \tag{1-15}$$

The test of (1-14) may also be expressed as one accepting the alternative when

$$\frac{d}{d\theta} \ln\{f_{\mathbf{X}}(\mathbf{x} \mid \theta)\}\Big|_{\theta=\theta_0} > t \tag{1-16}$$

Theorem 2 may also be used to obtain tests maximizing the *second* derivative $p''(\theta_0 \mid \delta)$ at $\theta = \theta_0$. This would be appropriate to attempt if it so happens that $p'(\theta_0 \mid \delta) = 0$ for all size-α tests for a given problem. The condition $p'(\theta_0 \mid \delta) = 0$ will occur if $\frac{d}{d\theta} f_{\mathbf{X}}(\mathbf{x} \mid \theta)\Big|_{\theta=\theta_0}$ is zero, assuming the requisite regularity conditions mentioned above. In this case Theorem 2 can be applied to obtain the locally optimum test accepting the alternative hypothesis $K: \theta > \theta_0$ when

$$\frac{\frac{d^2}{d\theta^2} f_{\mathbf{X}}(\mathbf{x} \mid \theta)\Big|_{\theta=\theta_0}}{f_{\mathbf{X}}(\mathbf{x} \mid \theta_0)} > t \tag{1-17}$$

One type of problem for which Theorem 2 is useful in characterizing locally optimum tests is that of testing $\theta = \theta_0$ against the *two-sided* alternative hypothesis $\theta \neq \theta_0$. We have previously mentioned that one can impose the condition of *unbiasedness* on the allowable tests for a problem. Unbiasedness of a size-α test for the hypotheses H and K of (1-1) and (1-2) means that the test satisfies

$$p(\theta \mid \delta) \leq \alpha, \text{ all } \theta \in \Theta_H \tag{1-18}$$

$$p(\theta \mid \delta) \geq \alpha, \text{ all } \theta \in \Theta_K \tag{1-19}$$

so that the detection probability for any $\theta_K \in \Theta_K$ is never less than the size α. For the two-sided alternative hypothesis $\theta \neq \theta_0$, suppose the pdf's $f_X(\mathbf{x} \mid \theta)$ are sufficiently regular so that the power functions of all tests are twice continuously differentiable at $\theta = \theta_0$. Then it follows that for any *unbiased* size-α test we will have $p(\theta_0 \mid \delta) = \alpha$ and $p'(\theta_0 \mid \delta) = 0$. Thus, the test function of a locally optimum unbiased test can be characterized by using these two constraints and maximizing $p''(\theta_0 \mid \delta)$ in Theorem 2. Another interpretation of the above approach for the two-sided alternative hypothesis is that the quantity $\omega = (\theta - \theta_0)^2$ may then be used as a measure of the distance of any alternative hypothesis from the null hypoth·sis $\theta = \theta_0$. We have

$$\frac{d}{d\omega} p(\theta \mid \delta)\bigg|_{\omega = 0} = \frac{1}{2(\theta - \theta_0)} \frac{d}{d\theta} p(\theta \mid \delta)\bigg|_{\theta = \theta_0}.$$

$$= \frac{1}{2} p''(\theta_0 \mid \delta) \tag{1-20}$$

if $p'(\theta_0 \mid \delta)$ is zero, for sufficiently regular pdf's $f_X(\mathbf{x} \mid \theta)$. Thus if $p'(\theta_0 \mid \delta)$ is zero for a class of size-α tests, then maximization of $p''(\theta_0 \mid \delta)$ leads to a test which is locally optimum within that class.

1.5 Bayes Tests

In general statistical decision theory which can treat estimation and hypothesis-testing problems within a single framework, there are four fundamental entities. These are (a) the observation space, which in our case is $\mathbb{R}^n$; (b) the set Θ of values of θ which parameterizes the possible distributions of the observations; (c) the set of all actions a which may be taken, the action space A; and (d) the loss function $l(\theta, a)$, a real-valued function measuring the loss suffered due to an action a when $\theta \in \Theta$ is the parameter value. In a binary hypothesis-testing problem the action space A will have only two possible actions, a_H and a_K, which, respectively, represent acceptance of the hypotheses H and K; and as a reasonable choice of loss function we can take

$$l(\theta, a) = c_{LJ}, \quad \theta \in \Theta_J, \text{ and } a = a_L \tag{1-21}$$

where $J,L = H$ or K, the c_{LJ} are non-negative, and $c_{JJ} = 0$. What is sought is a decision rule $d(\mathbf{x})$ taking on values in A, which specifies the action to be taken when an observation $\mathbf{x}$ has been made. More generally we can permit randomized decisions $\delta(\mathbf{x})$ which for each $\mathbf{x}$ specify a probability distribution over A.

The performance of any decision rule $d(\mathbf{x})$ can be characterized by the average loss that is incurred in using it; this is the *risk function*

$$R(\theta,d) = E\{l(\theta,d(\mathbf{x}) \mid \theta\}$$
$$= \int_{R^n} l(\theta,d(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x} \mid \theta)\, d\mathbf{x} \tag{1-22}$$

The risk function for any given decision rule is nonetheless a function of θ, so that a comparison of the performances of different decision rules over a set of values of θ is not quite straightforward. A single real number serving as a figure of merit is assigned to a decision rule in Bayesian decision theory; to do this there is assumed to be available information leading to an *a priori* characterization of a probability distribution over Θ. We will denote the corresponding pdf as $\pi(\theta)$, and obtain the *Bayes risk* for a given prior density $\pi(\theta)$ and a decision rule $d(\mathbf{x})$ as

$$r(\pi,d) = E\{R(\theta,d)\}$$
$$= \int_{\Theta} R(\theta,d)\pi(\theta)\, d\theta \tag{1-23}$$

In the binary hypothesis-testing problem of deciding between Θ_H and Θ_K for θ in $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$ [Equations (1-1) and (1-2)], the prior pdf $\pi(\theta)$ may be obtained as

$$\pi(\theta) = \pi_H\, \pi(\theta \mid H) + \pi_K\, \pi(\theta \mid K) \tag{1-24}$$

where π_H and π_K are the respective *a priori* probabilities that H and K are true ($\pi_H + \pi_K = 1$), and $\pi(\theta \mid H)$ and $\pi(\theta \mid K)$ are the conditional *a priori* pdf's over Θ, conditioned, respectively, on H and K being true. For the loss function of (1-21) this gives

$$r(\pi,d) = \int_{\Theta}\int_{R^n} l(\theta,d(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x} \mid \theta)\pi(\theta)\, d\mathbf{x}\, d\theta$$
$$= \int_{R^n}\int_{\Theta} l(\theta,d(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x} \mid \theta)\left\{\sum_{J=H,K} \pi_J\, \pi(\theta \mid J)\right\} d\theta\, d\mathbf{x} \tag{1-25}$$