2

Classical Detection Theory

2.1 INTRODUCTION

In this chapter, we develop in detail the basic ideas of classical detection theory. The first step is to define the various terms.

The basic components of a simple decision theory problem are shown in Figure 2.1.

The first is a *source* that generates an output. In the simplest case, this output is one of two choices. We refer to them as hypotheses and label them H_0 and H_1 in the two-choice case. More generally, the output might be one of M hypotheses, which we label $H_0, H_1, \ldots, H_{M-1}$. Some typical source mechanisms are the following:

1. A digital communication system transmits information by sending ones and zeros. When "one" is sent, we call it H_1, and when "zero" is sent, we call it H_0.
2. In a radar system, we look at a particular range and azimuth and try to decide whether a target is present; H_1 corresponds to the presence of a target and H_0 corresponds to no target.
3. In a medical diagnosis problem, we examine an electrocardiogram. Here H_1 could correspond to the patient having had a heart attack and H_0 to the absence of one.

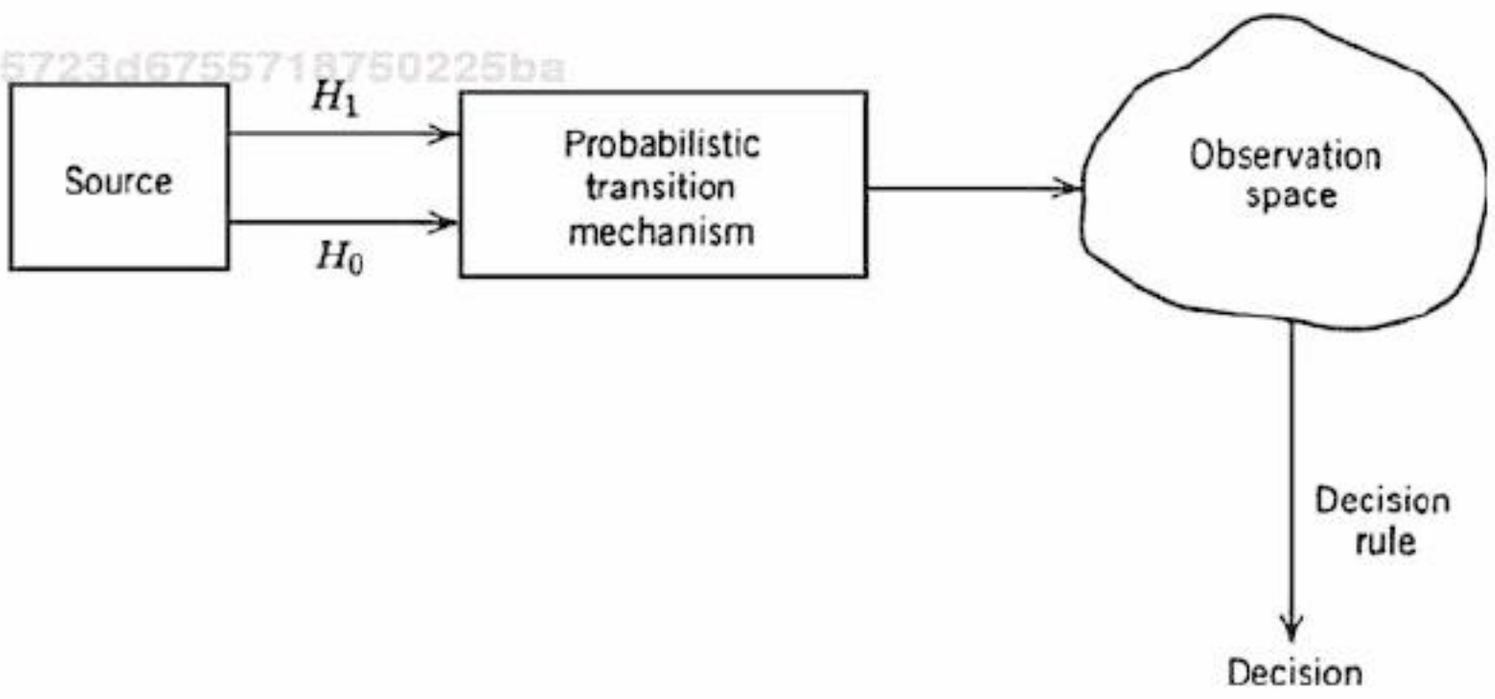
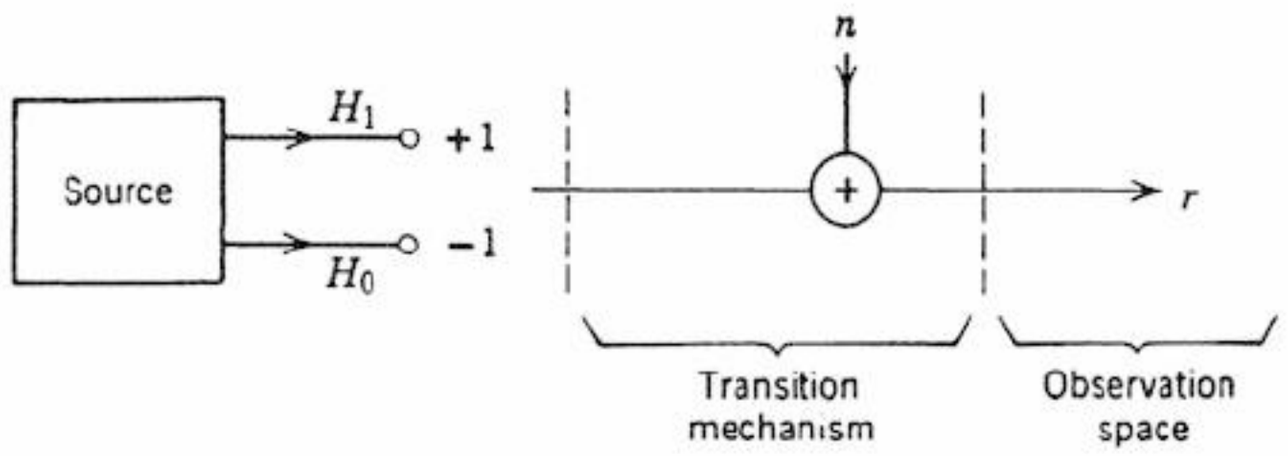
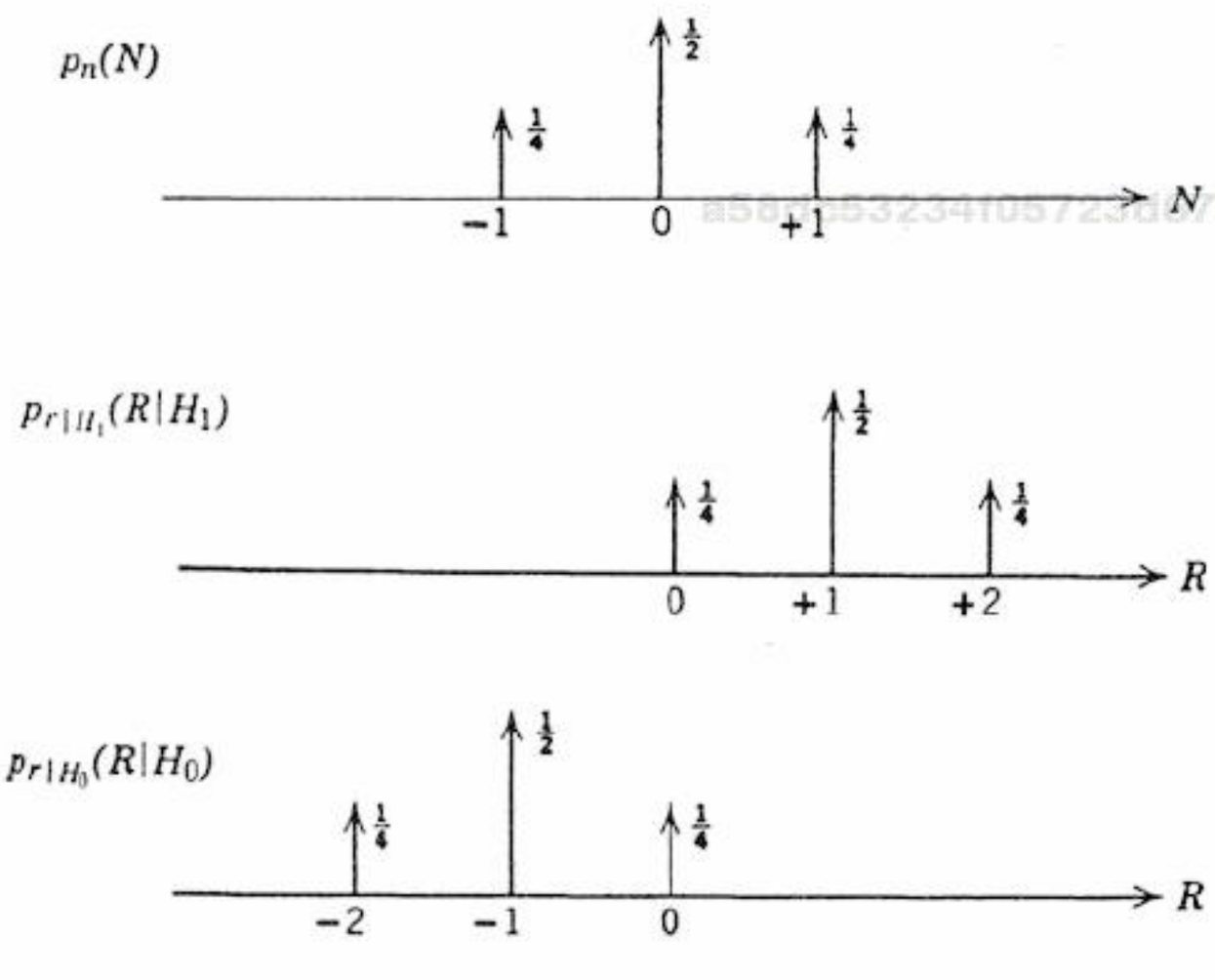


Figure 2.1: Components of a decision theory problem.

Detection, Estimation, and Modulation Theory, Second Edition. Harry L. Van Trees, Kristine L. Bell with Zhi Tian.



(*a*)



(*b*)

Figure 2.2: A simple decision problem: (*a*) model; (*b*) probability densities.

4. In a speaker classification problem we know the speaker is German, British, or American and either male or female. There are six possible hypotheses.

In the cases of interest to us, we do not know which hypothesis is true.

The second component of the problem is a probabilistic transition mechanism; the third is an observation space. The transition mechanism can be viewed as a device that knows which hypothesis is true. Based on this knowledge, it generates a point in the observation space according to some probability law. A simple example to illustrate these ideas is given in Figure 2.2. When H_1 is true, the source generates $+1$. When H_0 is true, the source generates -1. An independent discrete random variable n whose probability density is shown in Figure 2.2*b* is added to the source output. The sum of the source output and n is the observed variable r.

Under the two hypotheses, we have

$$\begin{aligned} H_1 &: r = 1 + n, \\ H_0 &: r = -1 + n. \end{aligned} \tag{2.1}$$

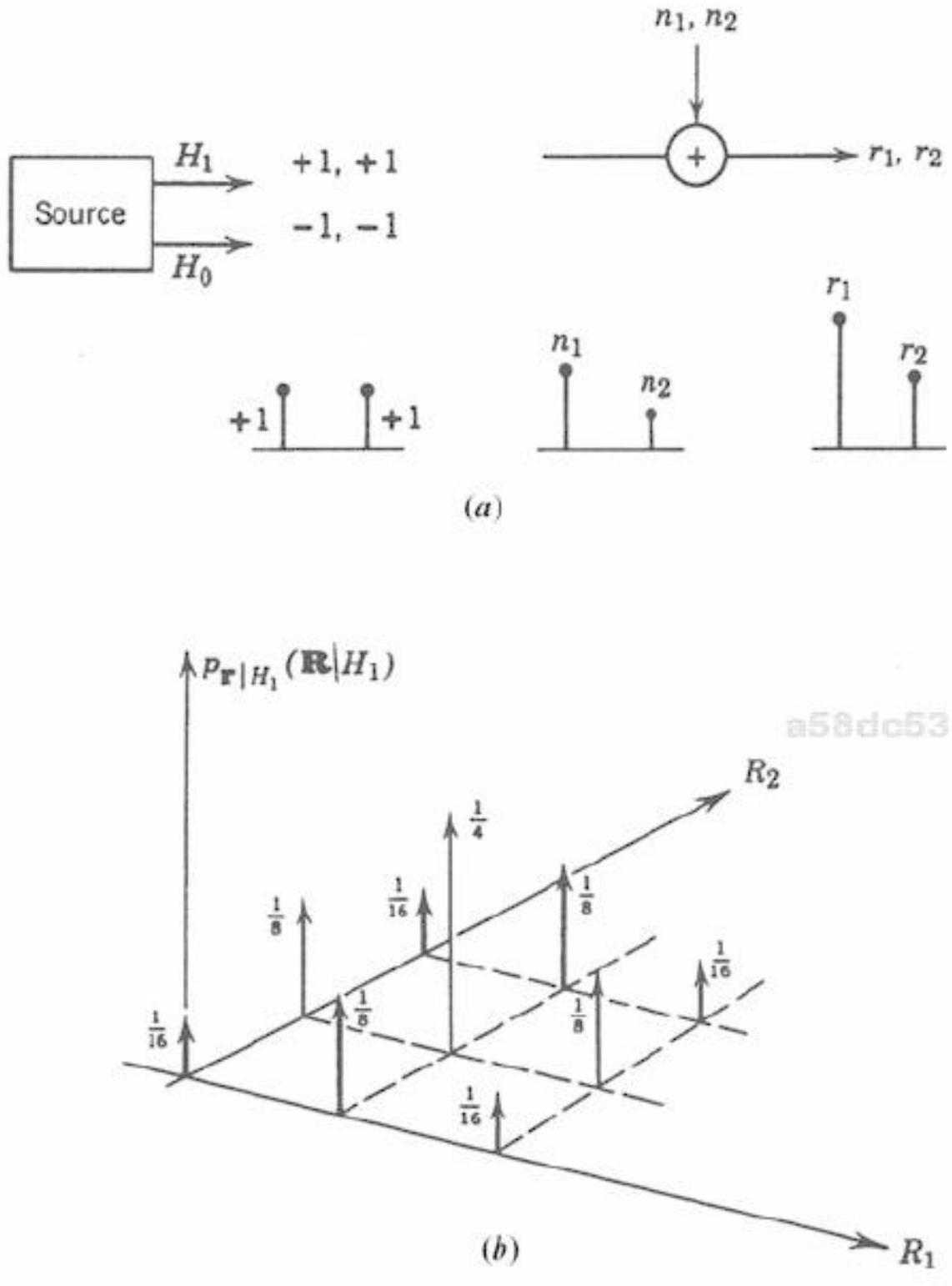


Figure 2.3: A two-dimensional problem: (*a*) model; (*b*) probability density.

The probability densities of r on the two hypotheses are shown in Figure 2.2*b*. The observation space is one dimensional, for any output can be plotted on a line.

A related example is shown in Figure 2.3*a* in which the source generates two numbers in sequence. A random variable n_1 is added to the first number and an independent random variable n_2 is added to the second.

Thus,

$$\begin{aligned} H_1 &: r_1 = 1 + n_1, \\ &\quad r_2 = 1 + n_2, \\ H_0 &: r_1 = -1 + n_1, \\ &\quad r_2 = -1 + n_2. \end{aligned} \tag{2.2}$$

The joint probability density of r_1 and r_2 when H_1 is true is shown in Figure 2.3*b*. The observation space is two dimensional and any observation can be represented as a point in a plane.

In this chapter, we confine our discussion to problems in which the observation space is finite dimensional. In other words, the observations consist of a set of N numbers and can be represented as a point in an N-dimensional space. This is the class of problem that

statisticians have treated for many years. For this reason, we refer to it as the *classical* decision problem.

The fourth component of the detection problem is a decision rule. After observing the outcome in the observation space we shall guess which hypothesis was true, and to accomplish this we develop a decision rule that assigns each point to one of the hypotheses. Suitable choices for decision rules will depend on several factors that we discuss in detail later. Our study will demonstrate how these four components fit together to form the total decision (or hypothesis testing) problem.

Organization. This chapter is organized in the following sections. In Section 2.2, we study the binary hypothesis testing problem. Then in Section 2.3, we extend the results to the case of M hypotheses.

In many cases of practical importance, we can develop the "optimum" decision rule according to certain criteria but cannot evaluate how well the test will work. In Section 2.4, we develop bounds and approximate expressions for the performance that will be necessary for some of the later chapters. In Section 2.5, we discuss Monte Carlo simulation and introduce a simulation technique using importance sampling. Finally, in Section 2.6, we summarize our results and indicate some of the topics that we have omitted.

2.2 SIMPLE BINARY HYPOTHESIS TESTS

As a starting point, we consider the decision problem in which each of two source outputs corresponds to a hypothesis. Each hypothesis maps into a point in the observation space. We assume that the observation space corresponds to a set of N observations: $r_1, r_2, r_3, \ldots, r_N$. Thus, each set can be thought of as a point in an N-dimensional space and can be denoted by a vector $\mathbf{r}$:

$$\mathbf{r} \triangleq \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}. \tag{2.3}$$

The probabilistic transition mechanism generates points in accord with the two known conditional probability densities $p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)$ and $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$. The object is to use this information to develop a suitable decision rule. To do this, we must look at various criteria for making decisions.

2.2.1 Decision Criteria

In the binary hypothesis problem, we know that either H_0 or H_1 is true. We shall confine our discussion to decision rules that are required to make a choice. (An alternative procedure would be to allow decision rules with three outputs: (a) H_0 true, (b) H_1 true, (c) do not know.) Thus, each time the experiment is conducted one of four things can happen:

1. H_0 true; choose H_0.
2. H_0 true; choose H_1.

3. H_1 true; choose H_1.
4. H_1 true; choose H_0.

The first and third alternatives correspond to correct choices. The second and fourth alternatives correspond to errors. The purpose of a decision criterion is to attach some relative importance to the four possible courses of action. It might be expected that the method of processing the received data ($\mathbf{r}$) would depend on the decision criterion we select. In this section, we show that for the two criteria of most interest, the Bayes and the Neyman–Pearson, the operations on $\mathbf{r}$ are identical.

Bayes Criterion. A Bayes test is based on two assumptions. The first is that the source outputs are governed by probability assignments, which are denoted by P_1 and P_0, respectively, and called the *a priori* probabilities. These probabilities represent the observer's information about the source before the experiment is conducted. The second assumption is that a cost is assigned to each possible course of action. We denote the cost for the four courses of action as C_{00}, C_{10}, C_{11}, C_{01}, respectively. The first subscript indicates the hypothesis chosen and the second, the hypothesis that is true. Each time the experiment is conducted, a certain cost will be incurred. We should like to design our decision rule so that *on the average* the cost will be as small as possible. To do this, we first write an expression for the expected value of the cost. We see that there are two probabilities that we must average over; the *a priori* probability and the probability that a particular course of action will be taken. Denoting the expected value of the cost as the risk $\mathcal{R}$, we have

$$\begin{aligned}\mathcal{R} = {} & C_{00}P_0 \Pr(\text{say } H_0|H_0 \text{ is true}) \\ & + C_{10}P_0 \Pr(\text{say } H_1|H_0 \text{ is true}) \\ & + C_{11}P_1 \Pr(\text{say } H_1|H_1 \text{ is true}) \\ & + C_{01}P_1 \Pr(\text{say } H_0|H_1 \text{ is true}).\end{aligned} \tag{2.4}$$

Because we have assumed that the decision rule must say either H_1 or H_0, we can view it as a rule for dividing the total observation space Z into two parts, Z_0 and Z_1, as shown in Figure 2.4. Whenever an observation falls in Z_0 we say H_0, and whenever an observation falls in Z_1 we say H_1.

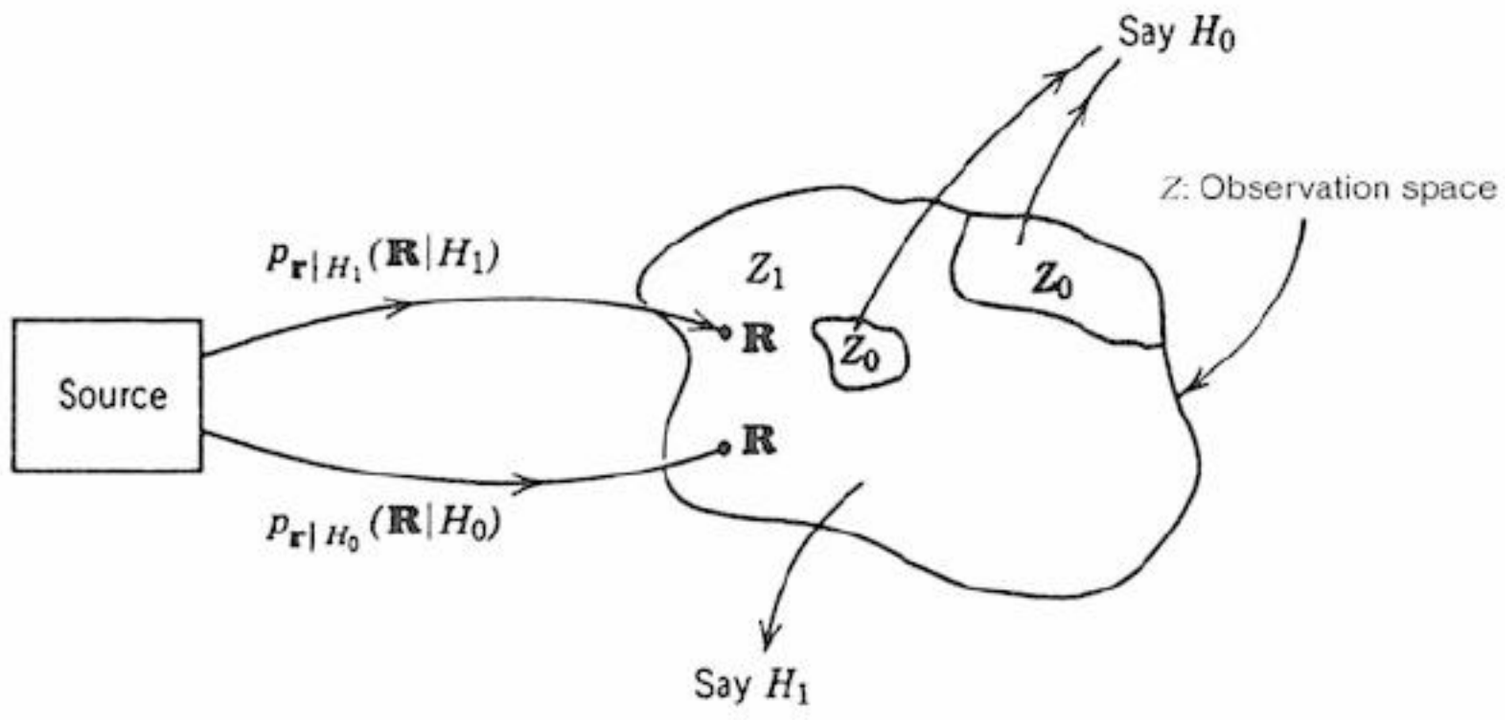


Figure 2.4: Decision regions.

We can now write the expression for the risk in terms of the transition probabilities and the decision regions:

$$\begin{aligned}\mathcal{R} = {} & C_{00}P_0 \int_{Z_0} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R} \\ & + C_{10}P_0 \int_{Z_1} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R} \\ & + C_{11}P_1 \int_{Z_1} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R} \\ & + C_{01}P_1 \int_{Z_0} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R}. \end{aligned} \tag{2.5}$$

For an N-dimensional observation space, the integrals in (2.5) are N-fold integrals. We shall assume throughout our work that the cost of a wrong decision is higher than the cost of a correct decision. In other words,

$$\begin{aligned} C_{10} &> C_{00}, \\ C_{01} &> C_{11}. \end{aligned} \tag{2.6}$$

Now, to find the Bayes test we must choose the decision regions Z_0 and Z_1 in such a manner that the risk will be minimized. Because we require that a decision be made, this means that we must assign each point $\mathbf{R}$ in the observation space Z to Z_0 or Z_1.

Thus,

$$Z = Z_0 + Z_1 \triangleq Z_0 \cup Z_1. \tag{2.7}$$

Rewriting (2.5), we have

$$\begin{aligned}\mathcal{R} = {} & P_0C_{00}\int_{Z_0} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R} + P_0C_{10}\int_{Z-Z_0} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R} \\ & + P_1C_{01}\int_{Z_0} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R} + P_1C_{11}\int_{Z-Z_0} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R}. \end{aligned} \tag{2.8}$$

Observing that

$$\int_Z p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R} = \int_Z p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R} = 1, \tag{2.9}$$

(2.8) reduces to

$$\mathcal{R} = P_0C_{10} + P_1C_{11} + \int_{Z_0} \Big[\big(P_1(C_{01} - C_{11})p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) \big) - \big(P_0(C_{10} - C_{00})p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) \big) \Big] d\mathbf{R}. \tag{2.10}$$

The first two terms represent the fixed cost. The integral represents the cost controlled by those points $\mathbf{R}$ that we assign to Z_0. The assumption in (2.6) implies that the two terms

inside the brackets are positive. Therefore, all values of $\mathbf{R}$ where the second term is larger than the first should be included in Z_0 because they contribute a negative amount to the integral. Similarly, all values of $\mathbf{R}$ where the first term is larger than the second should be excluded from Z_0 (assigned to Z_1) because they would contribute a positive amount to the integral. Values of $\mathbf{R}$ where the two terms are equal have no effect on the cost and may be assigned arbitrarily. We shall assume that these points are assigned to H_1 and ignore them in our subsequent discussion. Thus, the decision regions are defined by the statement: If

$$P_1(C_{01} - C_{11})p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) \geqslant P_0(C_{10} - C_{00})p_{\mathbf{r}|H_0}(\mathbf{R}|H_0), \tag{2.11}$$

assign $\mathbf{R}$ to Z_1 and consequently say that H_1 is true. Otherwise assign $\mathbf{R}$ to Z_0 and say H_0 is true.

Alternatively, we may write

$$\frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}. \tag{2.12}$$

The quantity on the left-hand side is called the *likelihood ratio* and denoted by $\Lambda(\mathbf{R})$

$$\boxed{\Lambda(\mathbf{R}) \triangleq \frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)}.} \tag{2.13}$$

Because it is the ratio of two functions of a random variable, it is a random variable. We see that regardless of the dimensionality of $\mathbf{R}$, $\Lambda(\mathbf{R})$ is a one-dimensional variable.

The quantity on the right-hand side of (2.12) is the threshold of the test and is denoted by η:

$$\eta \triangleq \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}. \tag{2.14}$$

Thus, Bayes criterion leads us to a *likelihood ratio test* (LRT)

$$\Lambda(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{2.15}$$

We see that all the data processing is involved in computing $\Lambda(\mathbf{R})$ and is not affected by *a priori* probabilities or cost assignments. This invariance of the data processing is of considerable practical importance. Frequently, the costs and *a priori* probabilities are merely educated guesses. The result in (2.15) enables us to build the entire processor and leave η as a variable threshold to accommodate changes in our estimates of *a priori* probabilities and costs.

Because the natural logarithm is a monotonic function, and both sides of (2.15) are positive, an equivalent test is

$$\ln \Lambda(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta. \tag{2.16}$$

Two forms of a processor to implement a likelihood ratio test are shown in Figure 2.5. Before proceeding to other criteria, we consider three simple examples.

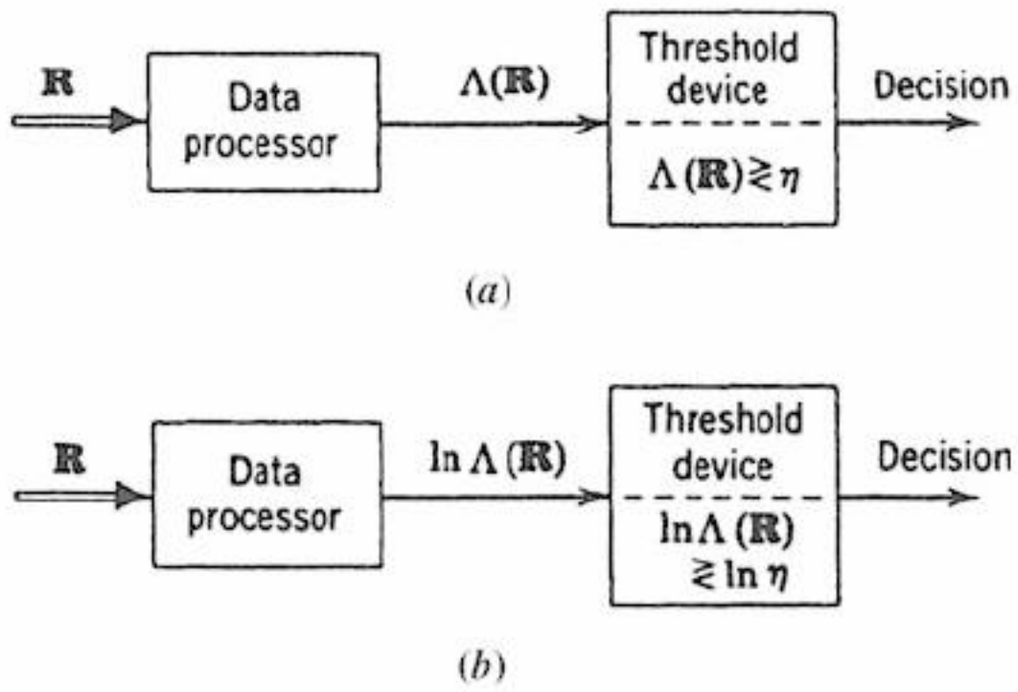


Figure 2.5: Likelihood ratio processors.

Example 2.1. We assume that under H_1 the source output is a constant voltage m. Under H_0, the source output is zero. Before observation, the voltage is corrupted by an additive noise. We sample the output waveform each second and obtain N samples. Each noise sample is a zero-mean Gaussian random variable n with variance σ^2. The noise samples at various instants are independent random variables and are independent of the source output. Looking at Figure 2.6, we see that the observations under the two hypotheses are

$$\begin{aligned} H_1 &: r_i = m + n_i, \quad i = 1, 2, \ldots, N, \\ H_0 &: r_i = n_i, \quad i = 1, 2, \ldots, N, \end{aligned} \tag{2.17}$$

and

$$p_{n_i}(X) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{X^2}{2\sigma^2}\right), \tag{2.18}$$

because the noise samples are Gaussian.

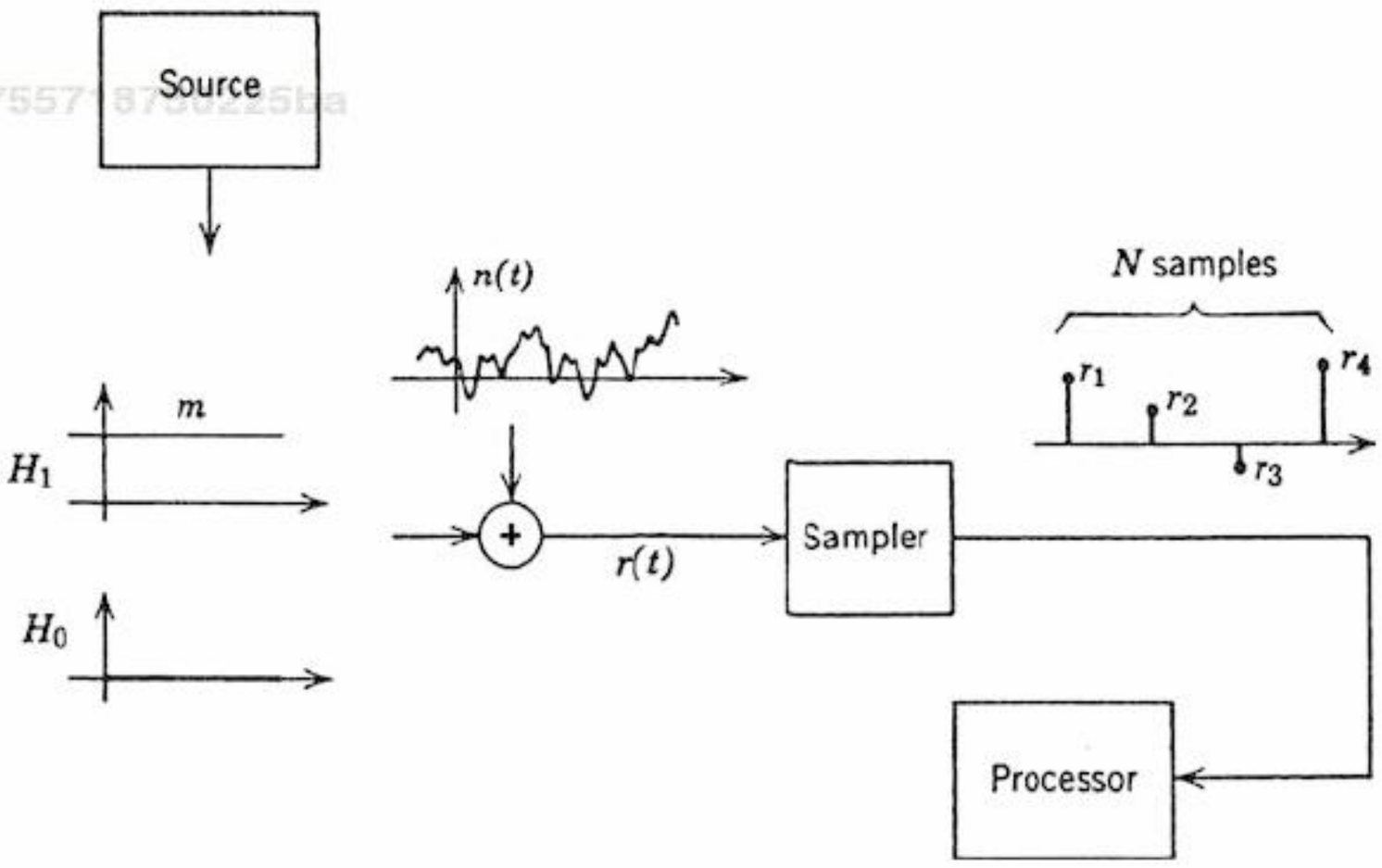


Figure 2.6: Model for Example 2.1.

The probability density of r_i under each hypothesis follows easily:

$$p_{r_i|H_1}(R_i|H_1) = p_{n_i}(R_i - m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(R_i - m)^2}{2\sigma^2} \right) \tag{2.19}$$

and

$$p_{r_i|H_0}(R_i|H_0) = p_{n_i}(R_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{R_i^2}{2\sigma^2} \right). \tag{2.20}$$

Because the n_i are statistically independent, the joint probability density of the r_i (or, equivalently, of the vector $\mathbf{r}$) is simply the product of the individual probability densities. Thus,

$$p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(R_i - m)^2}{2\sigma^2} \right) \tag{2.21}$$

and

$$p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{R_i^2}{2\sigma^2} \right). \tag{2.22}$$

Substituting into (2.13), we have

$$\Lambda(\mathbf{R}) = \frac{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(R_i - m)^2}{2\sigma^2} \right)}{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{R_i^2}{2\sigma^2} \right)}. \tag{2.23}$$

After canceling common terms and taking the logarithm, we have

$$\ln \Lambda(\mathbf{R}) = \frac{m}{\sigma^2} \sum_{i=1}^{N} R_i - \frac{Nm^2}{2\sigma^2}. \tag{2.24}$$

Thus, the likelihood ratio test is

$$\frac{m}{\sigma^2} \sum_{i=1}^{N} R_i - \frac{Nm^2}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta \tag{2.25}$$

or, equivalently,

$$\sum_{i=1}^{N} R_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma^2}{m} \ln \eta + \frac{Nm}{2} \triangleq \gamma. \tag{2.26}$$

We see that the processor simply adds the observations and compares them with a threshold. ■

In this example, the only way the data appear in the likelihood ratio test is in a sum. This is an example of a *sufficient statistic*, which we denote by $l(\mathbf{R})$ (or simply l when the argument is obvious). It is just a function of the received data, which has the property that $\Lambda(\mathbf{R})$ can be written as a function of l. In other words, when making a decision, knowing the value of the sufficient statistic is just as good as knowing $\mathbf{R}$. In Example 2.1, l is a linear function of the R_i. A case in which this is not true is illustrated in Example 2.2.

Example 2.2. Several different physical situations lead to the mathematical model of interest in this example. The observations consist of a set of N values: $r_1, r_2, r_3, \ldots, r_N$. Under both hypotheses,

the r_i are independent, identically distributed, zero-mean Gaussian random variables. Under H_1, each r_i has a variance σ_1^2. Under H_0, each r_i has a variance σ_0^2. Because the variables are independent, the joint density is simply the product of the individual densities. Therefore

$$p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{R_i^2}{2\sigma_1^2}\right) \tag{2.27}$$

and

$$p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{R_i^2}{2\sigma_0^2}\right). \tag{2.28}$$

Substituting (2.27) and (2.28) into (2.13) and taking the logarithm, we have

$$\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum_{i=1}^{N} R_i^2 + N \ln \frac{\sigma_0}{\sigma_1} \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta. \tag{2.29}$$

In this case, the sufficient statistic is the sum of the squares of the observations

$$l(\mathbf{R}) = \sum_{i=1}^{N} R_i^2, \tag{2.30}$$

and an equivalent test for $\sigma_1^2 > \sigma_0^2$ is

$$l(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\left(\ln \eta - N \ln \frac{\sigma_0}{\sigma_1}\right) \triangleq \gamma. \tag{2.31}$$

For $\sigma_1^2 < \sigma_0^2$, the inequality is reversed because we are multiplying by a negative number:

$$l(\mathbf{R}) \underset{H_1}{\overset{H_0}{\gtrless}} \frac{2\sigma_0^2\sigma_1^2}{\sigma_0^2 - \sigma_1^2}\left(N \ln \frac{\sigma_0}{\sigma_1} - \ln \eta\right) \triangleq \gamma'; \quad \sigma_1^2 < \sigma_0^2. \tag{2.32}$$

■

These two examples have emphasized Gaussian variables. In the next example, we consider a different type of distribution.

Example 2.3. The Poisson distribution of events is encountered frequently as a model of shot noise and other diverse phenomena (e.g., [DR58c] or [BR60]). Each time the experiment is conducted, a certain number of events occur. Our observation is just this number that ranges from 0 to ∞ and obeys a Poisson distribution on both hypotheses; that is,

$$\Pr(n \text{ events}) = \frac{(m_i)^n}{n!} e^{-m_i}, \quad n = 0, 1, 2, \ldots; i = 0, 1. \tag{2.33}$$

where m_i is the parameter that specifies the average number of events:

$$E(n) = m_i. \tag{2.34}$$

It is this parameter m_i that is different in the two hypotheses. Rewriting (2.33) to emphasize this point, we have for the two Poisson distributions

$$H_1 : \Pr(n \text{ events}) = \frac{(m_1)^n}{n!} e^{-m_1}, \quad n = 0, 1, 2, \ldots, \tag{2.35}$$

$$H_0 : \Pr(n \text{ events}) = \frac{(m_0)^n}{n!} e^{-m_0}, \quad n = 0, 1, 2, \ldots. \tag{2.36}$$

Then, the likelihood ratio test is

$$\Lambda(n) = \left(\frac{m_1}{m_0}\right)^n e^{-(m_1 - m_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.37}$$

or, equivalently,

$$n \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\ln \eta + m_1 - m_0}{\ln m_1 - \ln m_0}, \quad \text{if } m_1 > m_0,$$
$$n \underset{H_1}{\overset{H_0}{\gtrless}} \frac{\ln \eta + m_1 - m_0}{\ln m_1 - \ln m_0}, \quad \text{if } m_0 > m_1. \tag{2.38}$$

This example illustrates how the likelihood ratio test that we originally wrote in terms of probability densities can be simply adapted to accommodate observations that are discrete random variables. ■

Independent Distribution (ID) Model. In Examples 2.1 and 2.2, the components of the vector $\mathbf{r}$ are statistically independent and have identical probability distributions. This is referred to as an IID (independent identical distribution) model. In Examples 2.1 and 2.2, the probability densities are Gaussian but that is not necessary for an ID or IID model.

The likelihood ratio is given in (2.13). If the components are statistically independent (but not necessarily identically distributed), then

$$\Lambda(\mathbf{R}) = \frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} = \frac{\prod_{i=1}^{N} p_{r_i|H_1}(R_i|H_1)}{\prod_{i=1}^{N} p_{r_i|H_0}(R_i|H_0)}$$
$$\triangleq \prod_{i=1}^{N} \Lambda_i(R_i), \tag{2.39}$$

and the ID log-likelihood ratio is

$$\ln \Lambda(\mathbf{R}) = \sum_{i=1}^{N} \ln \Lambda_i(R_i). \tag{2.40}$$

If the components have identical probability densities,

$$p_{r_i|H_j}(R_i|H_j) = p_{r|H_j}(R_i|H_j) \quad j = 0, 1, \tag{2.41}$$

then (2.40) reduces to the IID log-likelihood ratio,

$$\ln \Lambda(\mathbf{R}) = \sum_{i=1}^{N} \ln \Lambda(R_i), \tag{2.42}$$

where

$$\ln \Lambda(R_i) \triangleq \ln \frac{p_{r|H_1}(R_i|H_1)}{p_{r|H_0}(R_i|H_0)} \tag{2.43}$$

is the marginal log-likelihood ratio. The log-likelihood ratio processor for the IID case is shown in Figure 2.7.

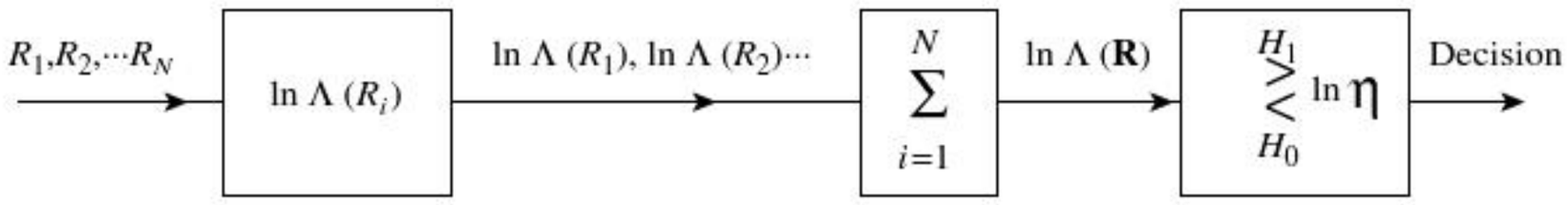


Figure 2.7: Likelihood ratio test for the IID model.

In Example 2.1, from (2.24)

$$\ln \Lambda(R_i) = \frac{m R_i}{\sigma^2} - \frac{m^2}{2\sigma^2}. \tag{2.44}$$

In Example 2.2, from (2.29)

$$\ln \Lambda(R_i) = \frac{1}{2}\left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2 \sigma_1^2}\right) R_i^2 - \frac{1}{2} \ln\left(\frac{\sigma_1^2}{\sigma_0^2}\right). \tag{2.45}$$

We now consider an IID model where $p_{\mathbf{r}|H_j}(\mathbf{r}|H_j)$; $j = 0, 1$ are non-Gaussian.

Example 2.4. The observations under the two hypotheses are

$$\begin{aligned} H_1 &: r_i = -\frac{m}{2} + n_i, \quad i = 1, 2, \ldots, N, \\ H_0 &: r_i = \frac{m}{2} + n_i, \qquad i = 1, 2, \ldots, N. \end{aligned} \tag{2.46}$$

The noise n_i is modeled as a zero-mean Generalized Gaussian random variable. The n_i are statistically independent random variables.

The Generalized Gaussian random variable family is defined as

$$p_n(X) = c \exp\left(-\left|\frac{X}{b}\right|^{\alpha}\right), \quad -\infty < X < \infty, \tag{2.47}$$

where α denotes a particular density in the family, b defines the spread of the density, and c normalizes the density. For $\alpha = 2$, we have a Gaussian random variable. For $\alpha = 1$, we have a Laplacian random variable.

The parameter b and constant c depend on the variance of the density,

$$b = \sigma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}, \tag{2.48}$$

$$c = \frac{\alpha}{2b\Gamma(1/\alpha)}, \tag{2.49}$$

where $\Gamma(\cdot)$ is the Gamma function.[1]

The probability densities for $\alpha = 1, 2$, and 3, and $\sigma = 1$ are shown in Figure 2.8. We see that as α decreases the densities decay more slowly.[2]

[1]The Gamma function is defined as $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$. It can be computed in Matlab using the function `gamma`. See Appendix A.

[2]Later, when we study clutter, we will see that these are referred to as heavy-tailed distributions.

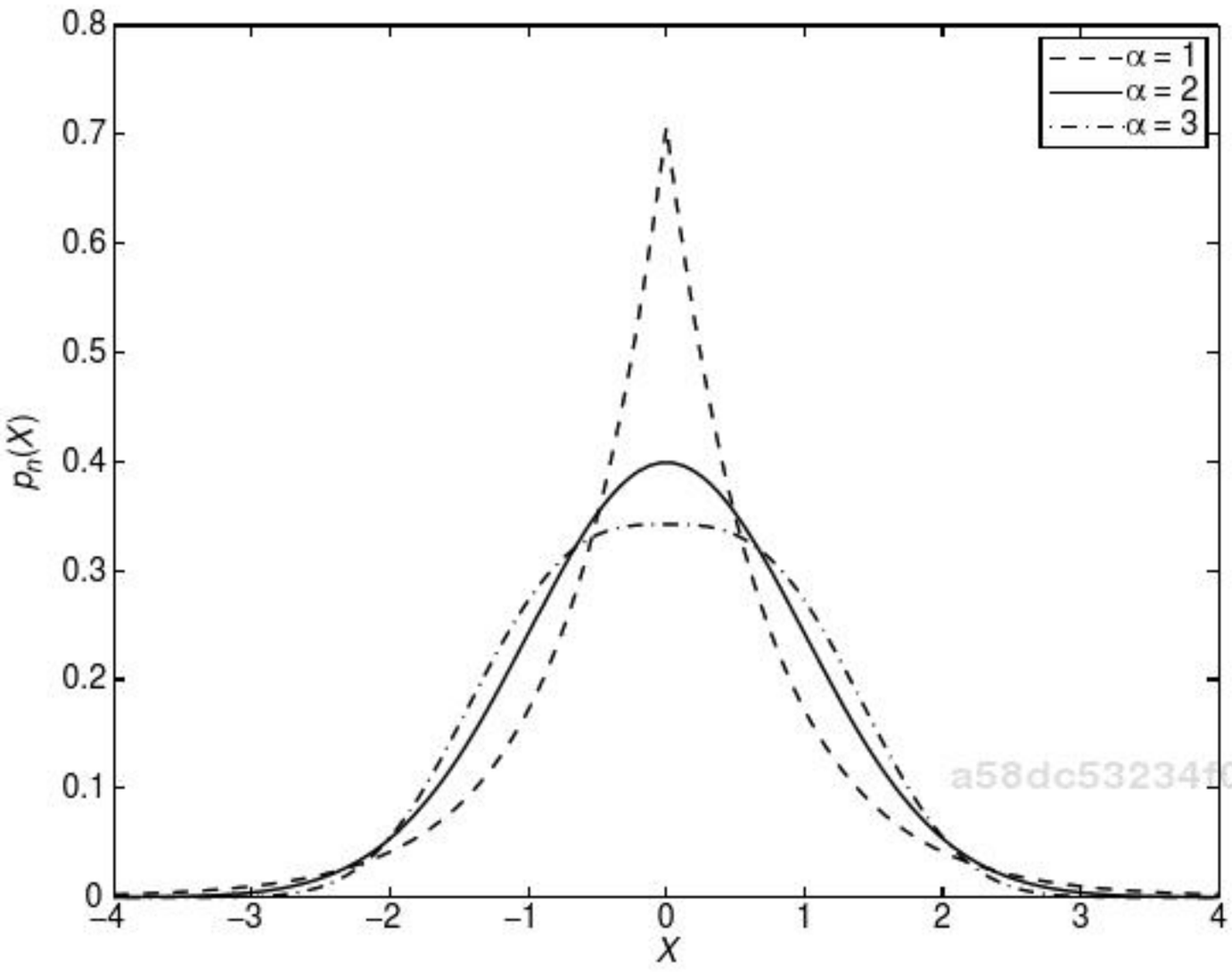


Figure 2.8: Probability densities; Generalized Gaussian random variable, $\alpha = 1$, 2, and 3.

The probability densities of the r_i on the two hypotheses are

$$p_{r|H_0}(R_i|H_0) = c\ \exp\left(-\left|\frac{R_i + m/2}{b}\right|^{\alpha}\right), \quad i = 1, 2, \ldots, N, \tag{2.50}$$

$$p_{r|H_1}(R_i|H_1) = c\ \exp\left(-\left|\frac{R_i - m/2}{b}\right|^{\alpha}\right), \quad i = 1, 2, \ldots, N. \tag{2.51}$$

The marginal log-likelihood ratio is found by substituting (2.50) and (2.51) in (2.43),

$$\begin{aligned}\ln \Lambda(R_i) &= \ln\left(\frac{c\ \exp\left(-\left|\frac{R_i - m/2}{b}\right|^{\alpha}\right)}{c\ \exp\left(-\left|\frac{R_i + m/2}{b}\right|^{\alpha}\right)}\right) \\ &= \frac{1}{b^{\alpha}}\left(\left|R_i + \frac{m}{2}\right|^{\alpha} - \left|R_i - \frac{m}{2}\right|^{\alpha}\right).\end{aligned} \tag{2.52}$$

In Figure 2.9, we plot $\ln \Lambda(R_i)$ for $\alpha = 1, 2,$ and 3. As α increases, the weighting of large values of R_i increases. ■

The LRT processor in Figure 2.7 is an important result because we can always find $\ln \Lambda(R_i)$ as defined in (2.43). In addition, since $\ln \Lambda(R_i)$ is a single variable-to-single variable mapping, we can always find the probability density of the output of the first box. This step will be important when we analyze the performance of the LRT.

We consider the IID model because it occurs frequently in practice. For the ID case when the components are statistically independent but have different probability densities, we obtain the more general model that is given in (2.39) and (2.40).

We now return to our general discussion of Bayes tests. There are several special kinds of Bayes test that are frequently used and that should be mentioned explicitly.

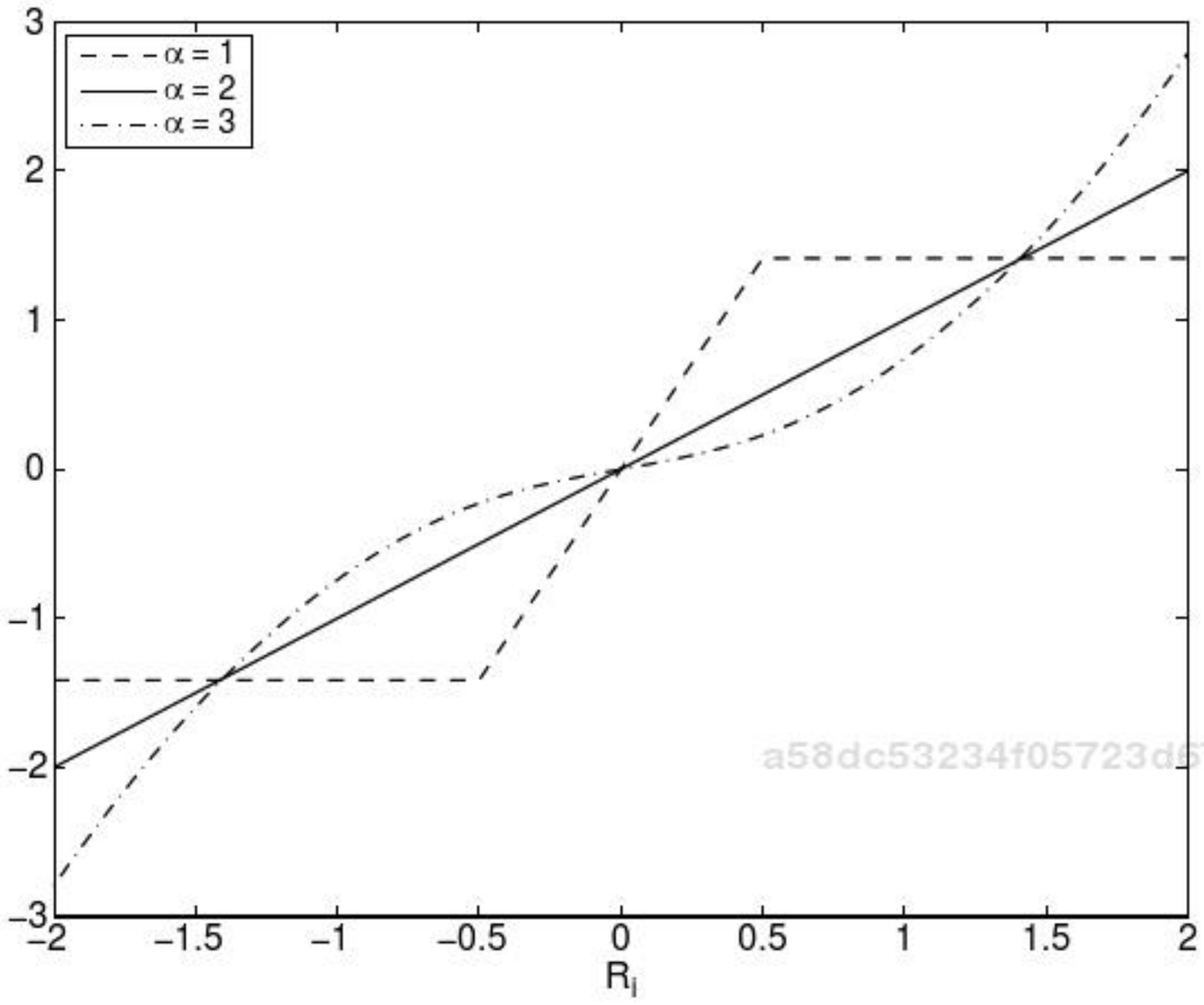


Figure 2.9: $\ln \Lambda(R_i)$; $\alpha = 1$, 2, and 3.

If we assume that C_{00} and C_{11} are zero and $C_{01} = C_{10} = 1$, the expression for the risk in (2.8) reduces to

$$\mathcal{R} = P_0 \int_{Z_1} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R} + P_1 \int_{Z_0} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R}. \tag{2.53}$$

We see that (2.53) is just the total probability of making an error. Therefore, for this cost assignment the Bayes test is minimizing the total probability of error. The test is

$$\ln \Lambda(\mathbf{R}) \underset{H_1}{\overset{H_0}{\gtrless}} \ln \frac{P_0}{P_1} = \ln P_0 - \ln(1 - P_0). \tag{2.54}$$

These processors are commonly referred to as minimum probability of error receivers. When the two hypotheses are equally likely, the threshold is zero. This assumption is normally true in digital communications systems.

We can also write the likelihood ratio test as

$$P_1 p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} P_0 p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) \tag{2.55}$$

or

$$\frac{P_1 p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}}(\mathbf{R})} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0 p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)}{p_{\mathbf{r}}(\mathbf{R})}. \tag{2.56}$$

The terms in (2.56) are the *a posteriori* probabilities of the two hypotheses,

$$\Pr(H_1|\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \Pr(H_0|\mathbf{R}). \tag{2.57}$$

Therefore, a minimum probability of error test is computing *a posteriori* probabilities of the two hypotheses and choosing the largest. It is frequently referred to as a maximum *a posteriori* probability (MAP) test.

A second special case of interest arises when the *a priori* probabilities are unknown. To investigate this case, we look at (2.8) again. We observe that once the decision regions Z_0 and Z_1 are chosen, the values of the integrals are determined. We denote these values in the following manner:

$$P_F = \int_{Z_1} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R},$$

$$P_D = \int_{Z_1} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R},$$

$$P_M = \int_{Z_0} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R} = 1 - P_D. \tag{2.58}$$

We see that these quantities are *conditional probabilities*. The subscripts are mnemonic and chosen from the radar problem in which hypothesis H_1 corresponds to the presence of a target and hypothesis H_0 corresponds to its absence. P_F is the probability of a false alarm (i.e., we say the target is present when it is not); P_D is the probability of detection (i.e., we say the target is present when it is); P_M is the probability of a miss (we say the target is absent when it is present). Although we are interested in a much larger class of problems than this notation implies, we shall use it for convenience.

For any choice of decision regions, the risk expression in (2.8) can be written in the notation of (2.58):

$$\mathcal{R} = P_0C_{10} + P_1C_{11} + P_1(C_{01} - C_{11})P_M - P_0(C_{10} - C_{00})(1 - P_F). \tag{2.59}$$

Because

$$P_0 = 1 - P_1, \tag{2.60}$$

(2.59) becomes

$$\mathcal{R}(P_1) = C_{00}(1 - P_F) + C_{10}P_F + P_1\big[(C_{11} - C_{00}) + (C_{01} - C_{11})P_M - (C_{10} - C_{00})P_F\big]. \tag{2.61}$$

Now, if all the costs and *a priori* probabilities are known, we can find a Bayes test. In Figure 2.10*a*, we plot the Bayes risk, $\mathcal{R}_B(P_1)$, as a function of P_1. Observe that as P_1 changes the decision regions for the Bayes test change and therefore P_F and P_M change.

Now consider the situation in which a certain P_1 (say $P_1 = P_1^*$) is assumed and the corresponding Bayes test designed. We now fix the threshold and assume that P_1 is allowed to change. We denote the risk for this fixed threshold test as $\mathcal{R}_F(P_1^*, P_1)$. Because the threshold is fixed, P_F and P_M are fixed, and (2.61) is just a straight line. Because it is a Bayes test for $P_1 = P_1^*$, it touches the $\mathcal{R}_B(P_1)$ curve at that point. Looking at (2.14), we

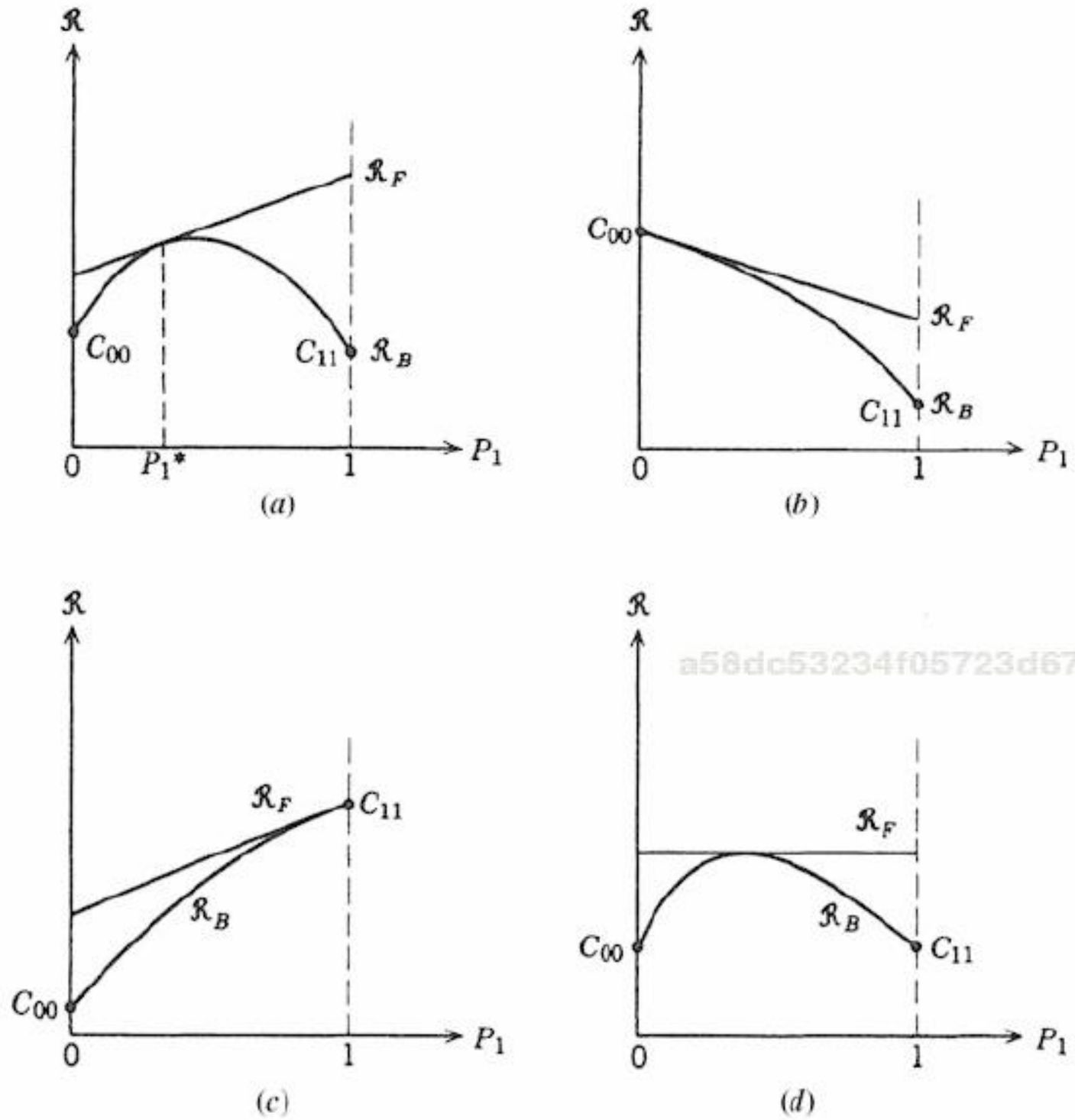


Figure 2.10: Risk curves: (*a*) fixed risk versus typical Bayes risk; (*b*) maximum value of $\mathcal{R}_1$ at $P_1 = 0$; (*c*) maximum value of $\mathcal{R}_1$ at $P_1 = 1$; (*d*) maximum value of $\mathcal{R}_1$ interior to [0, 1].

see that the threshold changes continuously with P_1. Therefore, whenever $P_1 \neq P_1^*$, the threshold in the Bayes test will be different. Because the Bayes test minimizes the risk,

$$\mathcal{R}_F(P_1^*, P_1) \geqslant \mathcal{R}_B(P_1). \tag{2.62}$$

If $\Lambda(\mathbf{R})$ is a continuous random variable with a probability distribution function that is strictly monotonic, then changing η always changes the risk. $\mathcal{R}_B(P_1)$ is strictly concave downward and the inequality in (2.62) is strict. This case, which is one of the particular interest to us, is illustrated in Figure 2.10*a*. We see that $\mathcal{R}_F(P_1^*, P_1)$ is tangent to $\mathcal{R}_B(P_1)$ at $P_1 = P_1^*$. These curves demonstrate the effect of incorrect knowledge of the *a priori* probabilities.

An interesting problem is encountered if we assume that the *a priori* probabilities are chosen to make our performance as bad as possible. In other words, P_1 is chosen to maximize our risk $\mathcal{R}_F(P_1^*, P_1)$. Three possible examples are given in Figures 2.10*b*, 2.10*c*, and 2.10*d*. In Figure 2.10*b*, the maximum of $\mathcal{R}_B(P_1)$ occurs at $P_1 = 0$. To minimize the maximum risk, we use a Bayes test designed assuming $P_1 = 0$. In Figure 2.10*c*, the maximum of $\mathcal{R}_B(P_1)$ occurs at $P_1 = 1$. To minimize the maximum risk, we use a Bayes test designed assuming $P_1 = 1$. In Figure 2.10*d*, the maximum occurs inside the interval [0, 1], and we

choose $\mathcal{R}_F$ to be the horizontal line. This implies that the coefficient of P_1 in (2.61) must be zero:

$$(C_{11} - C_{00}) + (C_{01} - C_{11})P_M - (C_{10} - C_{00})P_F = 0. \tag{2.63}$$

A Bayes test designed to minimize the maximum possible risk is called a *minimax test*. Equation (2.63) is referred to as the minimax equation and is useful whenever the maximum of $\mathcal{R}_B(P_1)$ is interior to the interval.

A special cost assignment that is frequently logical is

$$C_{00} = C_{11} = 0. \tag{2.64}$$

(This guarantees the maximum is interior.)

Denoting,

$$\begin{aligned} C_{01} &= C_M, \\ C_{10} &= C_F, \end{aligned} \tag{2.65}$$

the risk is,

$$\begin{aligned} \mathcal{R} &= C_F P_F + P_1(C_M P_M - C_F P_F) \\ &= P_0 C_F P_F + P_1 C_M P_M, \end{aligned} \tag{2.66}$$

and the minimax equation is

$$C_M P_M = C_F P_F. \tag{2.67}$$

Before continuing our discussion of likelihood ratio tests we shall discuss a second criterion and prove that it also leads to a likelihood ratio test.

Neyman–Pearson Tests. In many physical situations, it is difficult to assign realistic costs or *a priori* probabilities. A simple procedure to bypass this difficulty is to work with the conditional probabilities P_F and P_D. In general, we should like to make P_F as small as possible and P_D as large as possible. For most problems of practical importance, these are conflicting objectives. An obvious criterion is to constrain one of the probabilities and maximize (or minimize) the other. A specific statement of this criterion is the following:

Neyman–Pearson Criterion. Constrain $P_F = \alpha' \leqslant \alpha$ and design a test to maximize P_D (or minimize P_M) under this constraint.

The solution is obtained easily by using Lagrange multipliers. We construct the function f,

$$f = P_M + \lambda(P_F - \alpha'), \tag{2.68}$$

or

$$f = \int_{Z_0} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)d\mathbf{R} + \lambda\left(\int_{Z_1} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)d\mathbf{R} - \alpha'\right). \tag{2.69}$$

Clearly, if $P_F = \alpha'$, then minimizing f minimizes P_M. We rewrite (2.69) as

$$f = \lambda(1-\alpha') + \int_{Z_0} \left[p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) - p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\right] d\mathbf{R}. \tag{2.70}$$

Now, observe that for any positive value of λ an LRT will minimize f. (A negative value of λ gives an LRT with the inequalities reversed.)

This follows directly, because to minimize f we assign a point $\mathbf{R}$ to Z_0 only when the term in the bracket is negative. This is equivalent to the test

$$\frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} < \lambda, \quad \text{assign point to } Z_0 \text{ or say } H_0. \tag{2.71}$$

The quantity on the left is just the likelihood ratio. Thus, f is minimized by the likelihood ratio test

$$\Lambda(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{2.72}$$

To satisfy the constraint, we choose λ so that $P_F = \alpha$. If we denote the density of Λ when H_0 is true as $p_{\Lambda|H_0}(\Lambda|H_0)$, then we require

$$P_F = \int_{\lambda}^{\infty} p_{\Lambda|H_0}(\Lambda|H_0) d\Lambda = \alpha'. \tag{2.73}$$

Solving (2.73) for λ gives the threshold. The value of λ given by (2.73) will be nonnegative because $p_{\Lambda|H_0}(\Lambda|H_0)$ is zero for negative values of λ. Observe that decreasing λ is equivalent to increasing Z_1, the region where we say H_1. Thus, P_D increases as λ decreases. Therefore, we decrease λ until we obtain the largest possible $\alpha' \leqslant \alpha$. In most cases of interest to us, P_F is a continuous function of λ and we have $P_F = \alpha$. We shall assume this continuity in all subsequent discussions. Under this assumption, the Neyman–Pearson criterion leads to a likelihood ratio test. Later, we shall see the effect of the continuity assumption not being valid.

Summary. In this section, we have developed two ideas of fundamental importance in hypothesis testing. The first result is the demonstration that for a Bayes or a Neyman–Pearson criterion the optimum test consists of processing the observation $\mathbf{R}$ to find the likelihood ratio $\Lambda(\mathbf{R})$ and then comparing $\Lambda(\mathbf{R})$ to a threshold in order to make a decision. Thus, regardless of the dimensionality of the observation space, the decision space is one dimensional.

The second idea is that of a sufficient statistic $l(\mathbf{R})$. The idea of a sufficient statistic originated when we constructed the likelihood ratio and saw that it depended explicitly only on $l(\mathbf{R})$. If we actually construct $\Lambda(\mathbf{R})$, and then recognize $l(\mathbf{R})$, the notion of a sufficient statistic is perhaps of secondary value. A more important case is when we can recognize $l(\mathbf{R})$ directly. An easy way to do this is to examine the geometric interpretation of a sufficient statistic. We considered the observations $r_1, r_2, \ldots, r_N$ as a point $\mathbf{r}$ in an N-dimensional space, and one way to describe this point is to use these coordinates. When we choose a sufficient statistic, we are simply describing the point in a coordinate system that is more useful for the decision problem. We denote the first coordinate in this system by l, the

sufficient statistic, and the remaining $N-1$ coordinates which will not affect our decision by the $(N-1)$-dimensional vector $\boldsymbol{y}$. Thus,

$$\Lambda(\mathbf{R}) = \Lambda(L, \mathbf{Y}) = \frac{p_{l,\mathbf{y}|H_1}(L, \mathbf{Y}|H_1)}{p_{l,\mathbf{y}|H_0}(L, \mathbf{Y}|H_0)}. \tag{2.74}$$

Now, the expression on the right can be written as

$$\Lambda(L, \mathbf{Y}) = \frac{p_{l|H_1}(L|H_1)\, p_{\mathbf{y}\|l,H_1}(\mathbf{Y}|L, H_1)}{p_{l|H_0}(L|H_0)\, p_{\mathbf{y}\|l,H_0}(\mathbf{Y}|L, H_0)}. \tag{2.75}$$

If l is a sufficient statistic, then $\Lambda(\mathbf{R})$ must reduce to $\Lambda(L)$. This implies that the second terms in the numerator and denominator must be equal. In other words,

$$p_{\mathbf{y}\|l,H_0}(\mathbf{Y}|L, H_0) = p_{\mathbf{y}\|l,H_1}(\mathbf{Y}|L, H_1) \tag{2.76}$$

because the density of $\boldsymbol{y}$ cannot depend on which hypothesis is true. We see that choosing a sufficient statistic simply amounts to picking a coordinate system in which one coordinate contains all the information necessary to making a decision. The other coordinates contain no information and can be disregarded for the purpose of making a decision.

In Example 2.1, the new coordinate system could be obtained by a simple rotation. For example, when $N = 2$,

$$\begin{aligned} L &= \frac{1}{\sqrt{2}}(R_1 + R_2), \\ Y &= \frac{1}{\sqrt{2}}(R_1 - R_2). \end{aligned} \tag{2.77}$$

In Example 2.2, the new coordinate system corresponded to changing to polar coordinates. For $N = 2$,

$$\begin{aligned} L &= R_1^2 + R_2^2, \\ Y &= \tan^{-1}\frac{R_2}{R_1}. \end{aligned} \tag{2.78}$$

Notice that the vector $\boldsymbol{y}$ can be chosen in order to make the demonstration of the condition in (2.76) as simple as possible. The only requirement is that the pair $(l, \boldsymbol{y})$ must describe any point in the observation space. We should also observe that the condition

$$p_{\mathbf{y}|H_1}(\mathbf{Y}|H_1) = p_{\mathbf{y}|H_0}(\mathbf{Y}|H_0), \tag{2.79}$$

does not imply (2.76) unless l and $\boldsymbol{y}$ are independent under H_1 and H_0. Frequently, we will choose $\boldsymbol{y}$ to obtain this independence and then use (2.79) to verify that l is a sufficient statistic.

In the examples considered so far, the sufficient statistic has been a scalar quantity, but it may be multidimensional in general (see Problem 2.2.19).

2.2.2 Performance: Receiver Operating Characteristic

To complete our discussion of the simple binary problem, we must evaluate the performance of the likelihood ratio test. For a Neyman–Pearson test, the values of P_F and P_D completely

specify the test performance. Looking at (2.59), we see that the Bayes risk $\mathcal{R}_B$ follows easily if P_F and P_D are known. Thus, we can concentrate our efforts on calculating P_F and P_D.

We begin by considering Example 2.1 in Section 2.2.1.

Example 2.5 (continuation of Example 2.1). From (2.25), we see that an equivalent test is

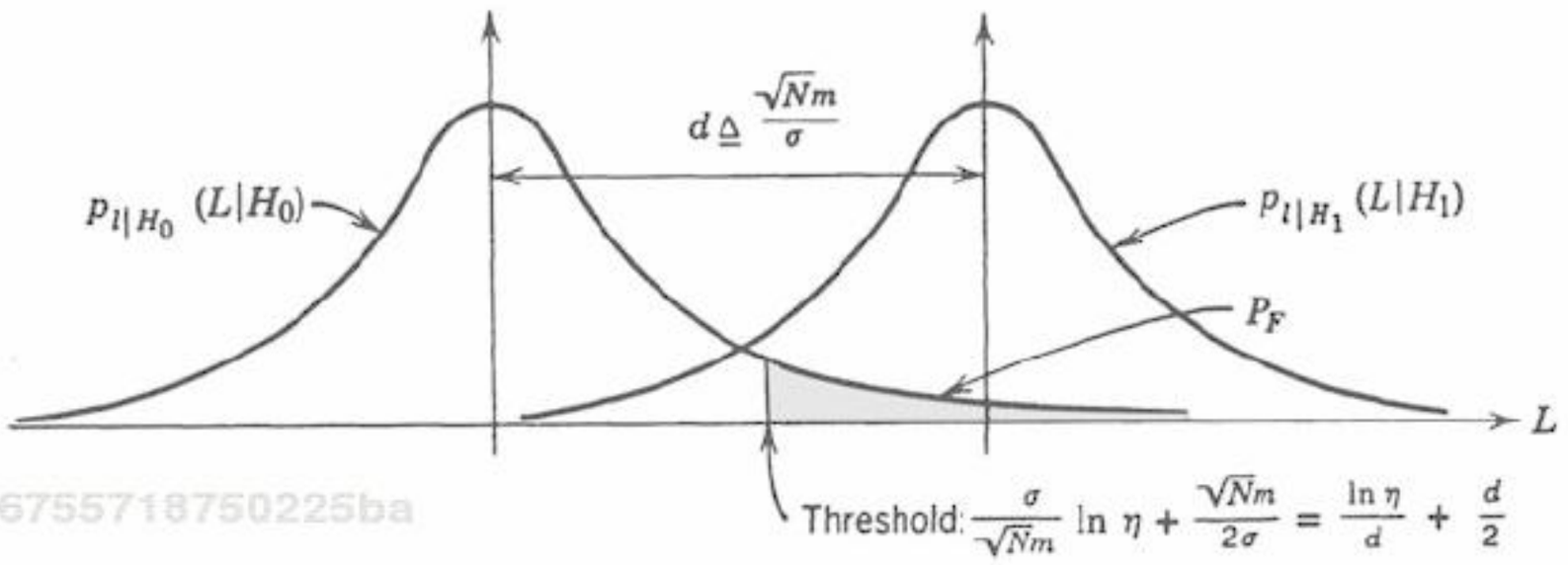
$$l(\mathbf{R}) = \frac{1}{\sqrt{N}\sigma} \sum_{i=1}^{N} R_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma}{\sqrt{N}m} \ln \eta + \frac{\sqrt{N}m}{2\sigma}. \tag{2.80}$$

We have multiplied (2.25) by $\sigma/\sqrt{N}m$ to normalize the next calculation. Under H_0, l is obtained by adding N independent zero-mean Gaussian variables with variance σ^2 and then dividing by $\sqrt{N}\sigma$. Therefore l is $N(0, 1)$.

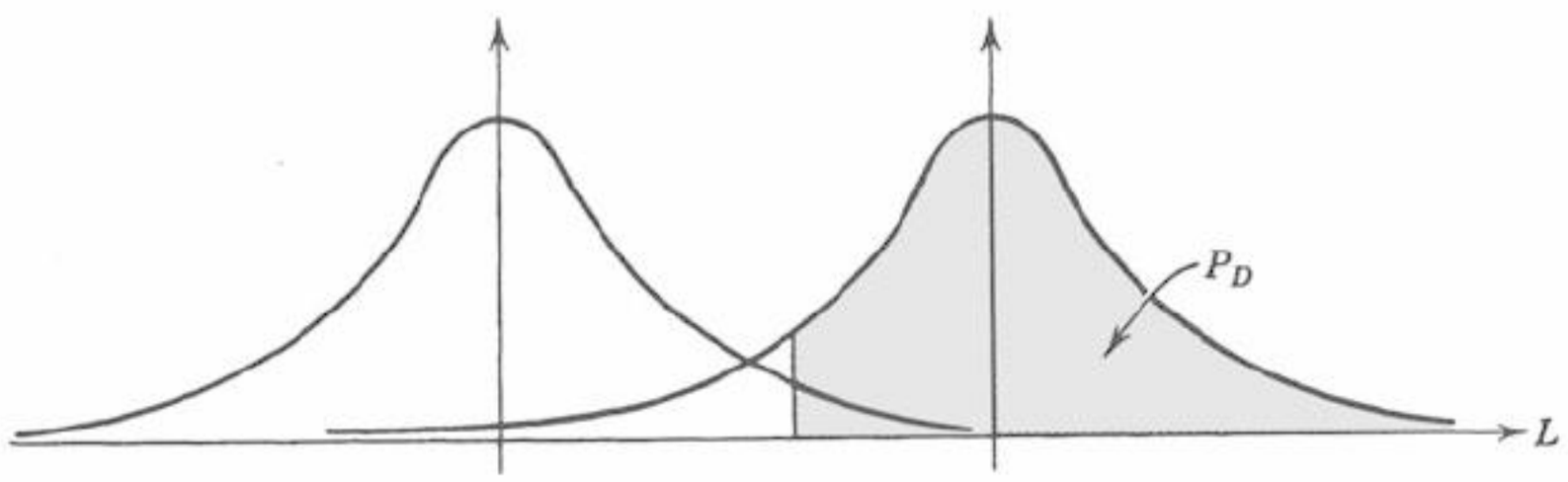
Under H_1, l is $N\left(\sqrt{N}m/\sigma, 1\right)$. The probability densities on the two hypotheses are sketched in Figure 2.11*a*. The threshold is also shown. Now, P_F is simply the integral of $p_{l|H_0}(L|H_0)$ to the right of the threshold.

Thus,

$$P_F = \int_{(\ln \eta)/d + d/2}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx, \tag{2.81}$$



(*a*)



(*b*)

Figure 2.11: Error probabilities: (*a*) P_F calculation; (*b*) P_D calculation.

where

$$d \triangleq \frac{\sqrt{N}m}{\sigma}, \tag{2.82}$$

is the distance between the means of the two densities. The integral in (2.81) is called the complementary error function and is tabulated in many references (e.g., [AS64] or [GH62]). We use a modified version of the standard definition that we denote as[3]

$$\text{erfc}_*(X) \triangleq \int_X^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx. \tag{2.83}$$

In this notation,

$$P_F = \text{erfc}_*\left(\frac{\ln\eta}{d} + \frac{d}{2}\right). \tag{2.84}$$

Similarly, P_D is the integral of $p_{l|H_1}(L|H_1)$ to the right of the threshold, as shown in Figure 2.11*b*:

$$\begin{aligned} P_D &= \int_{(\ln\eta)/d+d/2}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-d)^2}{2}\right) dx \\ &= \int_{(\ln\eta)/d-d/2}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \triangleq \text{erfc}_*\left(\frac{\ln\eta}{d} - \frac{d}{2}\right). \end{aligned} \tag{2.85}$$

In Figure 2.12*a*, we have plotted P_D versus P_F for various values of d with η as the varying parameter. For $\eta = 0$, $\ln\eta = -\infty$, and the processor always guesses H_1. Thus, $P_F = 1$ and $P_D = 1$. As η increases, P_F and P_D decrease. When $\eta = \infty$, the processor always guesses H_0 and $P_F = P_D = 0$.

As we would expect from Figure 2.11, the performance increases monotonically with d. In Figure 2.12*b*, we have replotted the results to give P_D versus d with P_F as a parameter on the curves. For a particular d, we can obtain any point on the curve by choosing η appropriately $(0, \infty)$.

The result in Figure 2.12*a* is referred to as the receiver operating characteristic (ROC). It completely describes the performance of the test as a function of the parameter of interest.

A special case that will be important when we look at communication systems is the case in which we want to minimize the total probability of error

$$\Pr(\epsilon) \triangleq P_0 P_F + P_1 P_M. \tag{2.86}$$

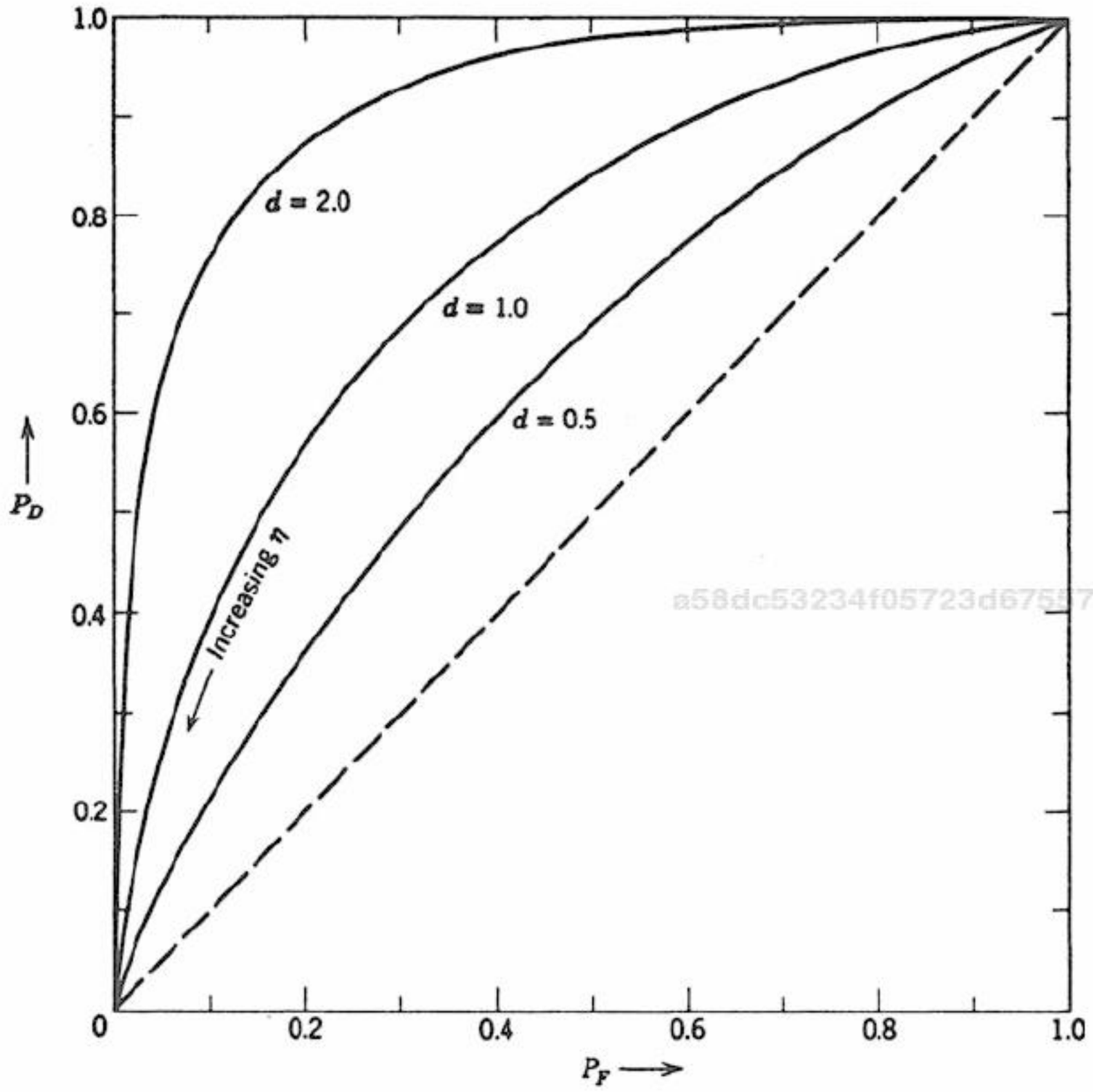
The threshold for this criterion was given in (2.54). For the special case in which $P_0 = P_1$ the threshold η equals one and

$$\Pr(\epsilon) = \frac{1}{2}(P_F + P_M). \tag{2.87}$$

Using (2.84) and (2.85) in (2.87), we have

$$\Pr(\epsilon) = \int_{d/2}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \text{erfc}_*\left(\frac{d}{2}\right). \tag{2.88}$$

[3]The standard complementary error function is defined as $\text{erfc}(X) = (2/\sqrt{\pi})\int_X^\infty e^{-x^2} dx$. Our modified version can be computed using the `normcdf` function in the Matlab Statistics Toolbox. See Appendix A.



(*a*)

Figure 2.12: (*a*) Receiver operating characteristic: Gaussian variables with unequal means.

It is obvious from (2.88) that we could also obtain the $\Pr(\epsilon)$ from the ROC. However, if this is the only threshold setting of interest, it is generally easier to calculate the $\Pr(\epsilon)$ directly. ■

Before calculating the performance of the other two examples, it is worthwhile to point out two simple bounds on $\text{erfc}_*(X)$. They will enable us to discuss its approximate behavior analytically. For $X > 0$,

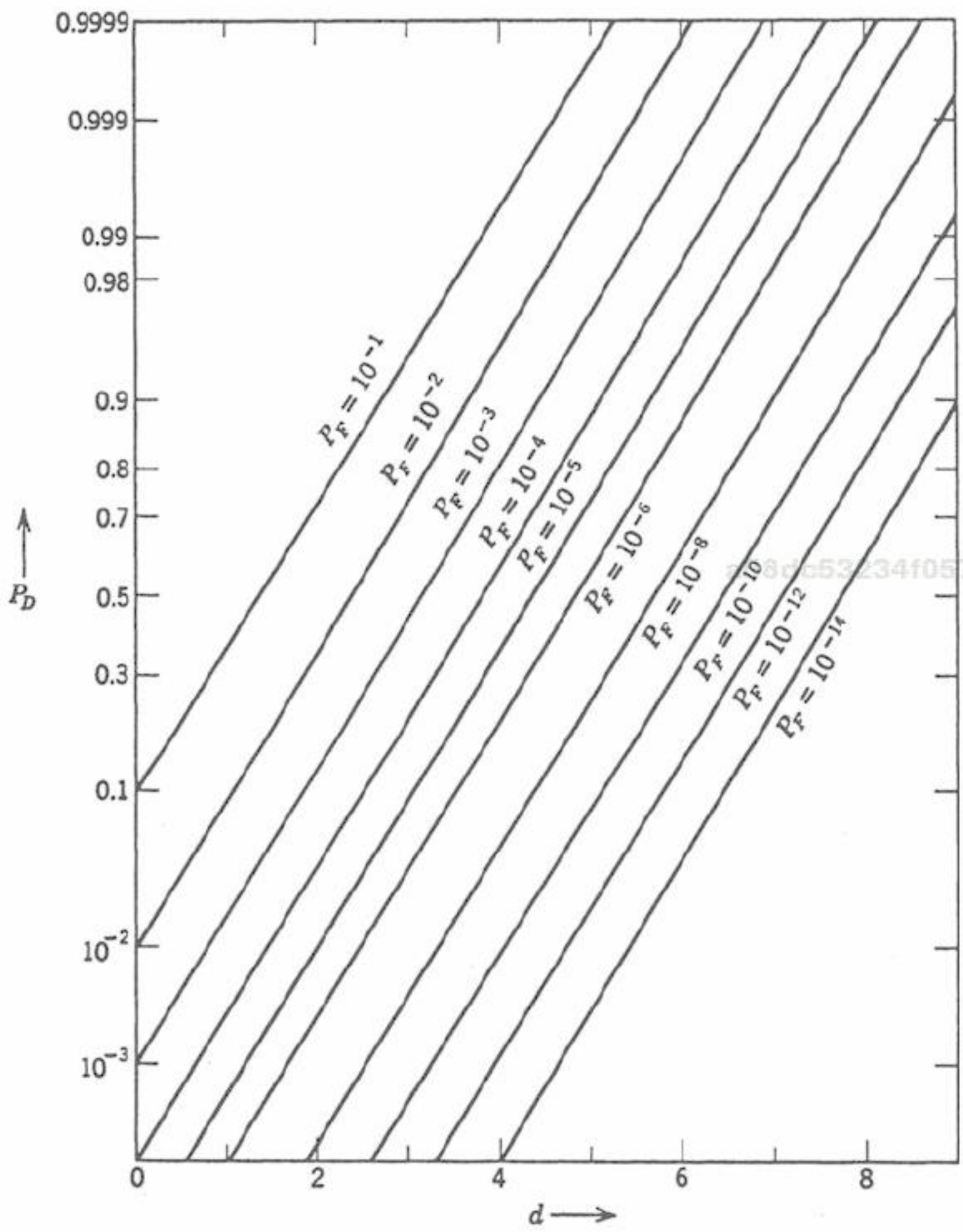
$$\frac{1}{\sqrt{2\pi}X}\left(1-\frac{1}{X^2}\right)\exp\left(-\frac{X^2}{2}\right) < \text{erfc}_*(X) < \frac{1}{\sqrt{2\pi}X}\exp\left(-\frac{X^2}{2}\right). \tag{2.89}$$

This can be derived by integrating by parts (see Problem 2.2.15 or Feller [Fel57]). A second bound is

$$\text{erfc}_*(X) < \frac{1}{2}\exp\left(-\frac{X^2}{2}\right), \quad x > 0, \tag{2.90}$$

which can also be derived easily (see Problem 2.2.16). The four curves are plotted in Figure 2.13. We note that $\text{erfc}_*(X)$ decreases exponentially.

The receiver operating characteristics for the other two examples are also of interest.



(*b*)

Figure 2.12: (*b*) Detection probability versus d.

Example 2.6 (continuation of Example 2.2). In this case, the test is

$$l(\mathbf{R}) = \sum_{i=1}^{N} R_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\left(\ln \eta - N \ln \frac{\sigma_0}{\sigma_1}\right) = \gamma, \quad \sigma_1 > \sigma_0. \tag{2.91}$$

A particularly simple case appearing frequently in practice is $N = 2$. Under H_0 the r_i are independent zero-mean Gaussian variables with variances equal to σ_0^2:

$$P_F = \Pr(l \geqslant \gamma | H_0) = \Pr(r_1^2 + r_2^2 \geqslant \gamma | H_0). \tag{2.92}$$

To evaluate the expression on the right, we change to polar coordinates:

$$\begin{aligned} r_1 &= z\cos\theta, \quad z = \sqrt{r_1^2 + r_2^2} \\ r_1 &= z\sin\theta, \quad \theta = \tan^{-1}\frac{r_2}{r_1}. \end{aligned} \tag{2.93}$$

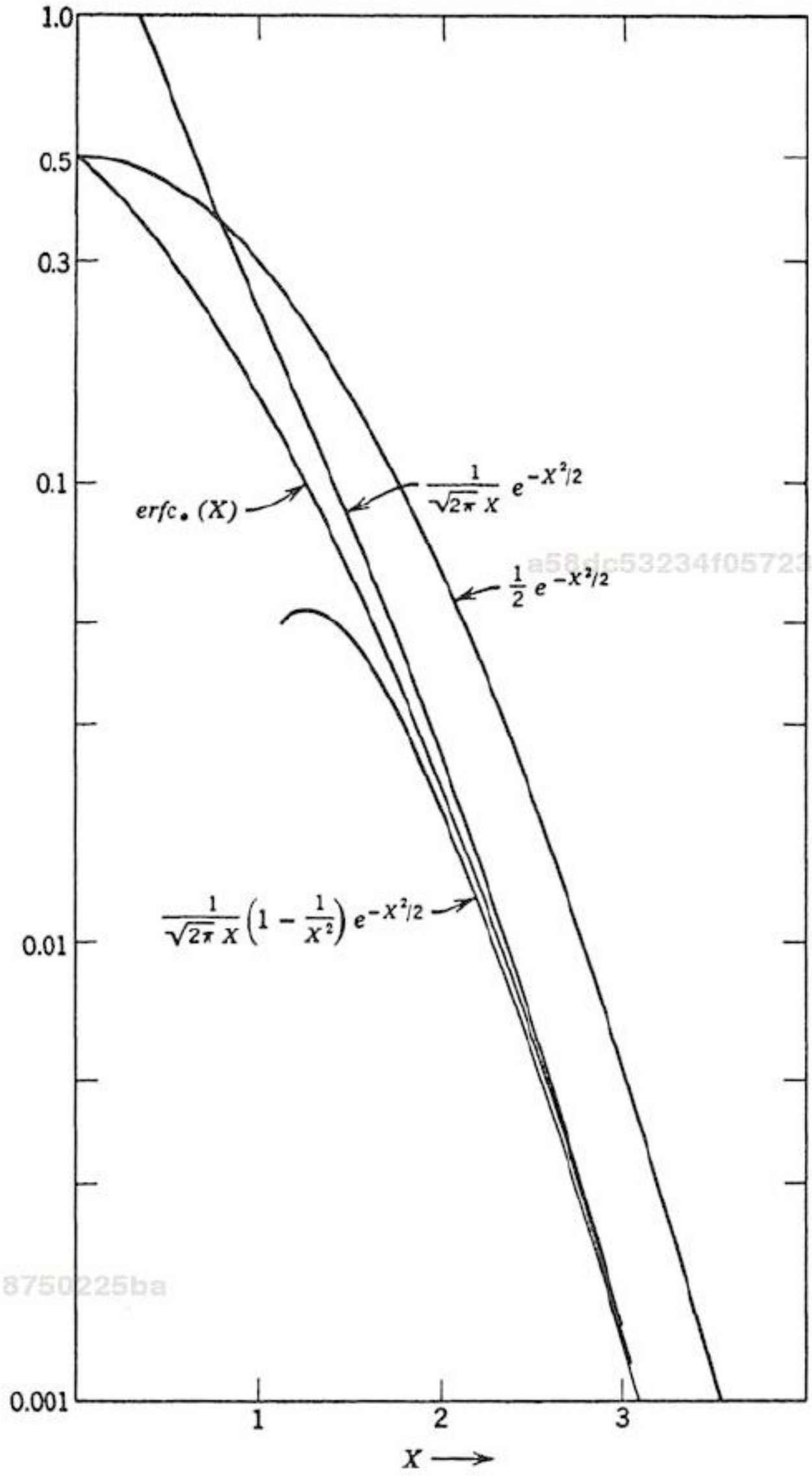


Figure 2.13: Plot of $\text{erfc}_*(X)$ and related functions.

Then,

$$\Pr(z^2 \geqslant \gamma | H_0) = \int_0^{2\pi} d\theta \int_{\sqrt{\gamma}}^{\infty} Z \frac{1}{2\pi\sigma_0^2} \exp\left(-\frac{Z^2}{2\sigma_0^2}\right) dZ. \tag{2.94}$$

Integrating with respect to θ, we have

$$P_F = \int_{\sqrt{\gamma}}^{\infty} Z \frac{1}{2\sigma_0^2} \exp\left(-\frac{Z^2}{2\sigma_0^2}\right) dZ. \tag{2.95}$$

We observe that l, the sufficient statistic, equals z^2. Changing variables, we have

$$P_F = \int_{\gamma}^{\infty} \frac{1}{2\sigma_0^2} \exp\left(-\frac{L}{2\sigma_0^2}\right) dL = \exp\left(-\frac{\gamma}{2\sigma_0^2}\right). \tag{2.96}$$

(Note that the probability density of the sufficient statistic is exponential.)

Similarly,

$$P_D = \exp\left(-\frac{\gamma}{2\sigma_1^2}\right). \tag{2.97}$$

To construct the ROC, we can combine (2.96) and (2.97) to eliminate the threshold γ. This gives

$$P_D = (P_F)^{\sigma_0^2/\sigma_1^2}. \tag{2.98}$$

In terms of logarithms,

$$\ln P_D = \frac{\sigma_0^2}{\sigma_1^2} \ln P_F. \tag{2.99}$$

As expected, the performance improves monotonically as the ratio σ_1^2/σ_0^2 increases.

We now calculate the performance for arbitrary N. On both hypotheses, $l(\mathbf{R})$ is the sum of the squares of N Gaussian variables. The difference in the hypotheses is in the variance of the Gaussian variables.

To find $p_{l|H_0}(L|H_0)$, we observe that the characteristic function (CF) of each R_i^2 is

$$\begin{aligned} M_{R_i^2|H_0}(jv) &\triangleq E\left\{ e^{jvR_i^2} \middle| H_0 \right\} = \int_{-\infty}^{\infty} e^{jvR_i^2} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-R_i^2/2\sigma_0^2} dR_i \\ &= \left(1 - 2jv\sigma_0^2\right)^{-1/2}. \end{aligned} \tag{2.100}$$

Because of the independence of the variables, $M_{l|H_0}(jv)$ can be written as a product. Therefore,

$$M_{l|H_0}(jv) = \left(1 - 2jv\sigma_0^2\right)^{-N/2}. \tag{2.101}$$

Taking the inverse transform, we obtain $p_{l|H_0}(L|H_0)$:

$$p_{l|H_0}(L|H_0) = \frac{L^{N/2-1} e^{-L/2\sigma_0^2}}{2^{N/2}\sigma_0^N \Gamma\left(\frac{N}{2}\right)} \quad L \geqslant 0, \tag{2.102}$$

which is familiar as the Gamma probability density function with shape parameter $a = N/2$ and scale parameter $b = 2\sigma_0^2$. The properties of the Gamma density are available in many sources (e.g., [JKB94]) and are summarized in Appendix A. Plots of the probability density are shown in Figure 2.14. When $\sigma_0^2 = 1$, the Gamma density in (2.102) is identical to a Chi-squared density with N degrees of freedom. For $N = 2$, it is easy to check that it is the simple Exponential density in (2.96).

Similarly, for H_1 we have,

$$p_{l|H_1}(L|H_1) = \frac{L^{N/2-1} e^{-L/2\sigma_1^2}}{2^{N/2}\sigma_1^N \Gamma\left(\frac{N}{2}\right)} \quad L \geqslant 0. \tag{2.103}$$

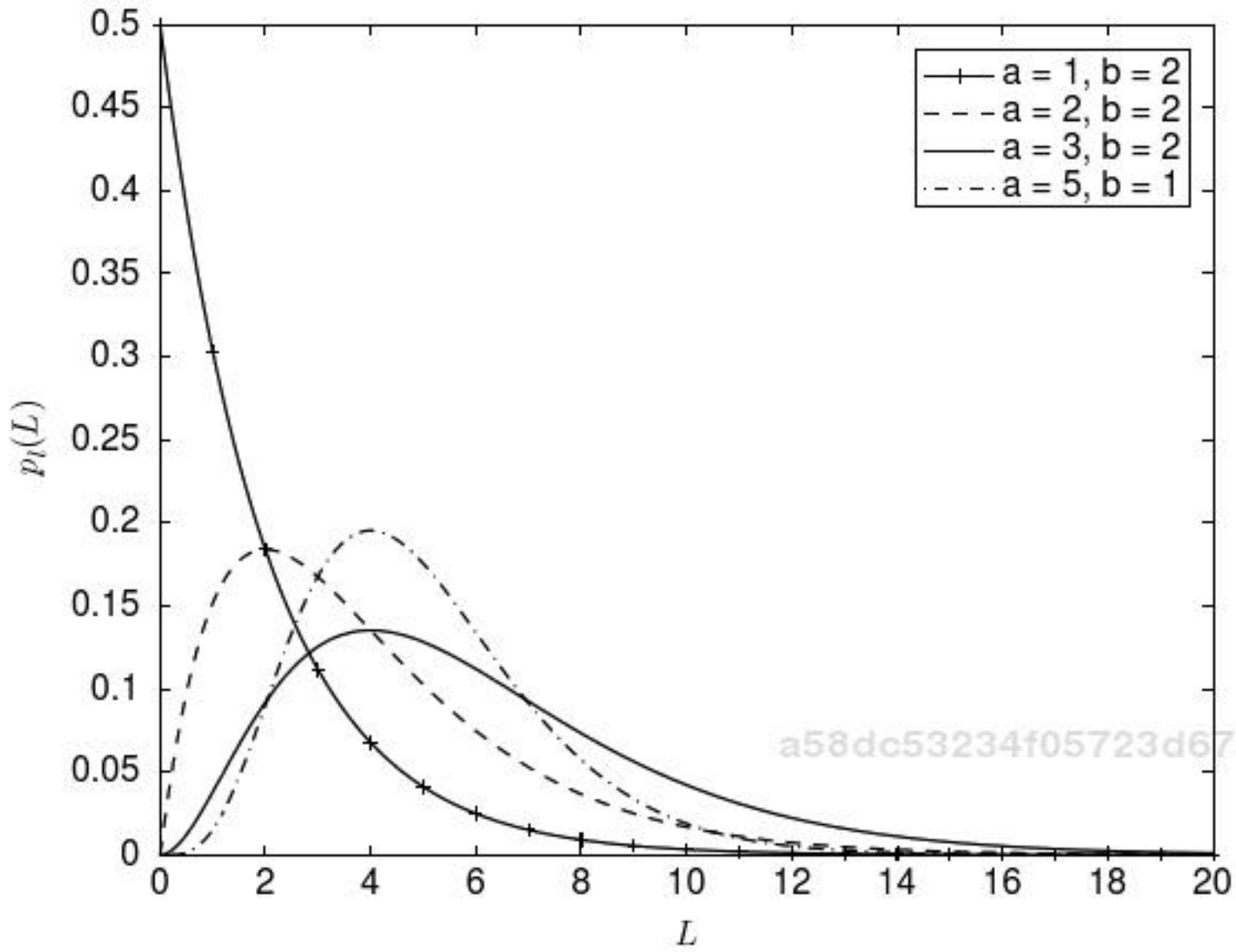


Figure 2.14: Gamma probability density for various a and b.

The expressions for P_D and P_F are,

$$P_D = \int_{\gamma}^{\infty} \left[2^{N/2} \sigma_1^N \, \Gamma\left(\frac{N}{2}\right) \right]^{-1} L^{N/2-1} e^{-L/2\sigma_1^2} \, dL, \tag{2.104}$$

and

$$P_F = \int_{\gamma}^{\infty} \left[2^{N/2} \sigma_0^N \, \Gamma\left(\frac{N}{2}\right) \right]^{-1} L^{N/2-1} e^{-L/2\sigma_0^2} \, dL. \tag{2.105}$$

Making the substitution,

$$X_1 = \frac{L}{2\sigma_1^2}, \tag{2.106}$$

in (2.104), and

$$X_0 = \frac{L}{2\sigma_0^2}, \tag{2.107}$$

in (2.105), the integrals are incomplete Gamma functions, where the normalized incomplete Gamma function is defined as[4]

$$\Gamma_a(x) \triangleq \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} \, dt. \tag{2.108}$$

[4]The incomplete Gamma function is tabulated in several references (e.g., [FY53] or [AS64]) and can be computed in Matlab using the `gammainc` function. See Appendix A.

Therefore,

$$P_D = 1 - \Gamma_{N/2}\left(\frac{\gamma'}{2\sigma_1^2}\right), \tag{2.109}$$

and

$$P_F = 1 - \Gamma_{N/2}\left(\frac{\gamma'}{2\sigma_0^2}\right). \tag{2.110}$$

We see that, for $N = 2$, P_F and P_D reduce to (2.96) and (2.97).

Consider the physical situation in which there is "noise" only on H_0,

$$\sigma_0^2 = \sigma_n^2, \tag{2.111}$$

and "signal" plus "noise" on H_1,

$$\sigma_1^2 = \sigma_s^2 + \sigma_n^2. \tag{2.112}$$

In Figure 2.15, we have plotted the receiver operating characteristic for some representative values of N and σ_s^2/σ_n^2.

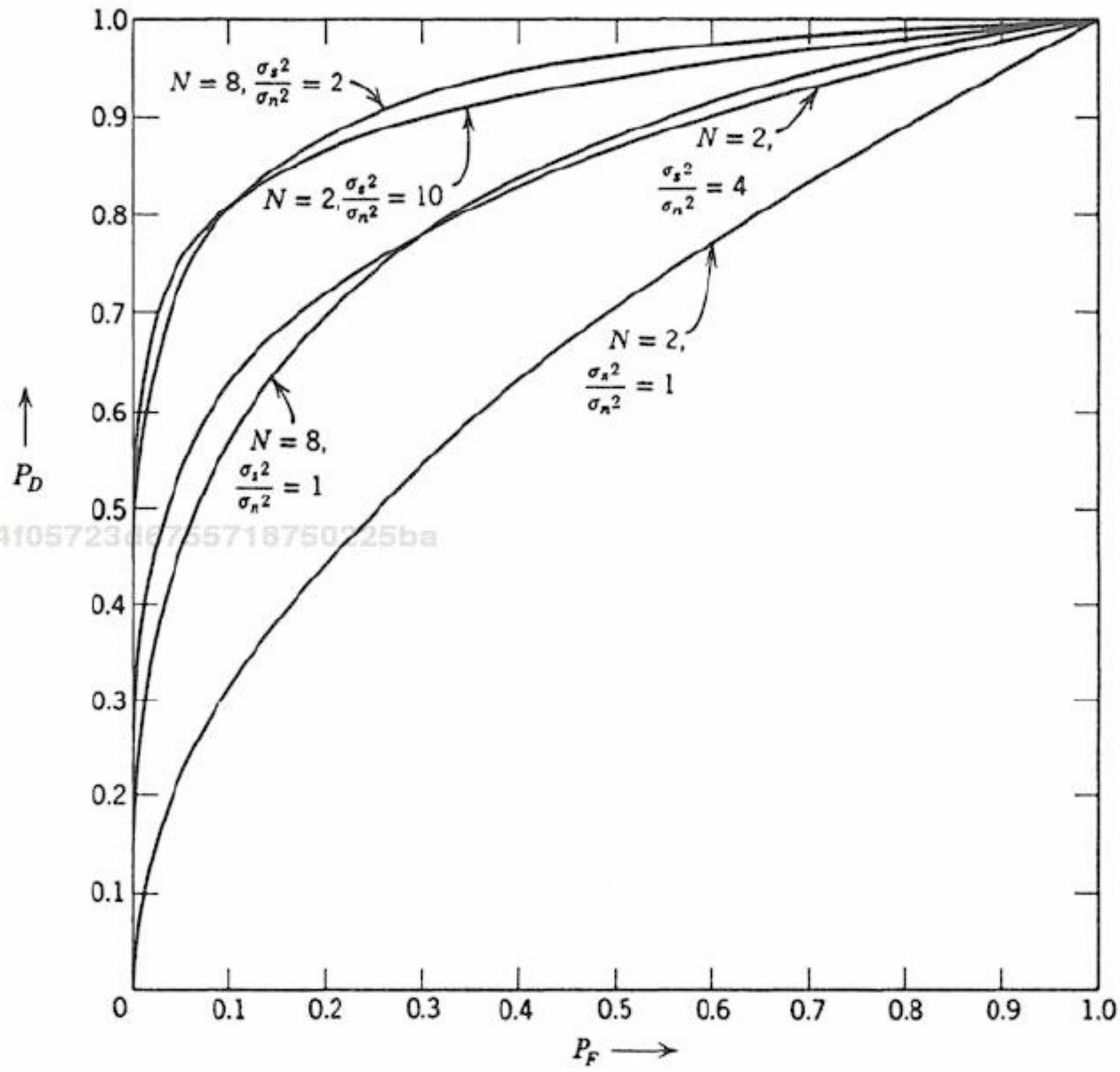


Figure 2.15: Receiver operating characteristic: Gaussian variables with identical means and unequal variances on the two hypotheses.

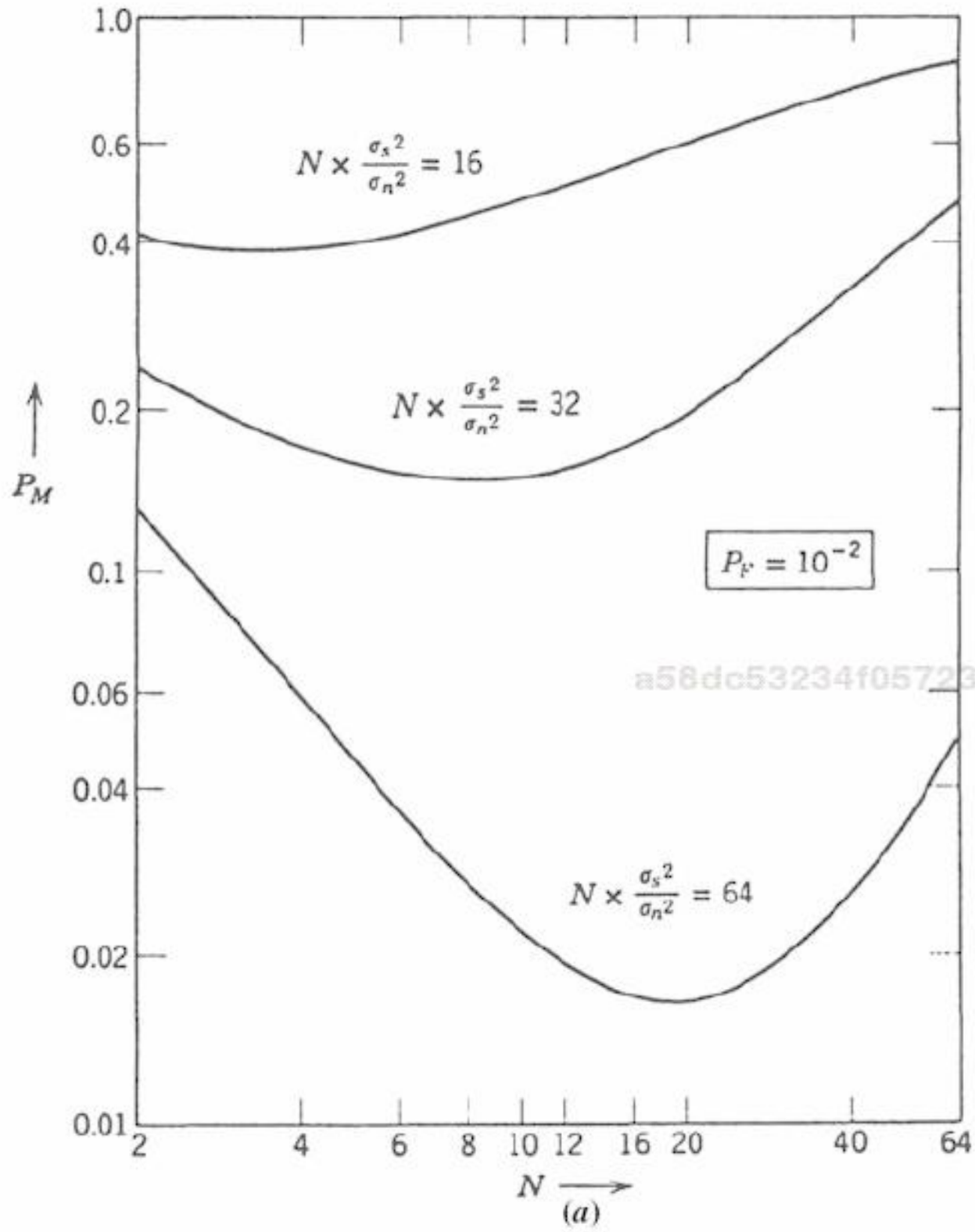


Figure 2.16: (*a*) P_M **as a function of** $N[P_F = 10^{-2}]$.

Two particularly interesting curves are those for $N = 8, \sigma_s^2/\sigma_n^2 = 1$, and $N = 2, \sigma_s^2/\sigma_n^2 = 4$. In both cases, the product $N\sigma_s^2/\sigma_n^2 = 8$. We see that when the desired P_F is greater than 0.3, P_D is higher if the available "signal strength" is divided into more components. This suggests that for each P_F and $N\sigma_s^2/\sigma_n^2$ there should be an optimum N. In Chapter 7, we shall see that this problem corresponds to optimum diversity in communication systems and the optimum energy per pulse in radars. In Figures 2.16*a* and *b*, we have plotted P_M as a function of N for $P_F = 10^{-2}$ and 10^{-4}, respectively, and various $N\sigma_s^2/\sigma_n^2$ products. We discuss the physical implications of these results in Chapters 7 and 10. ■

The two Poisson distributions are the third example.

Example 2.7 (continuation of Example 2.3). From (2.38), the likelihood ratio test is

$$n \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\ln \eta + m_1 - m_0}{\ln m_1 - \ln m_0} = \gamma, \quad m_1 > m_0. \tag{2.113}$$

Because n takes on only integer values, it is more convenient to rewrite (2.113) as

$$n \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_I, \quad \gamma_1 = 1, 2, \ldots. \tag{2.114}$$

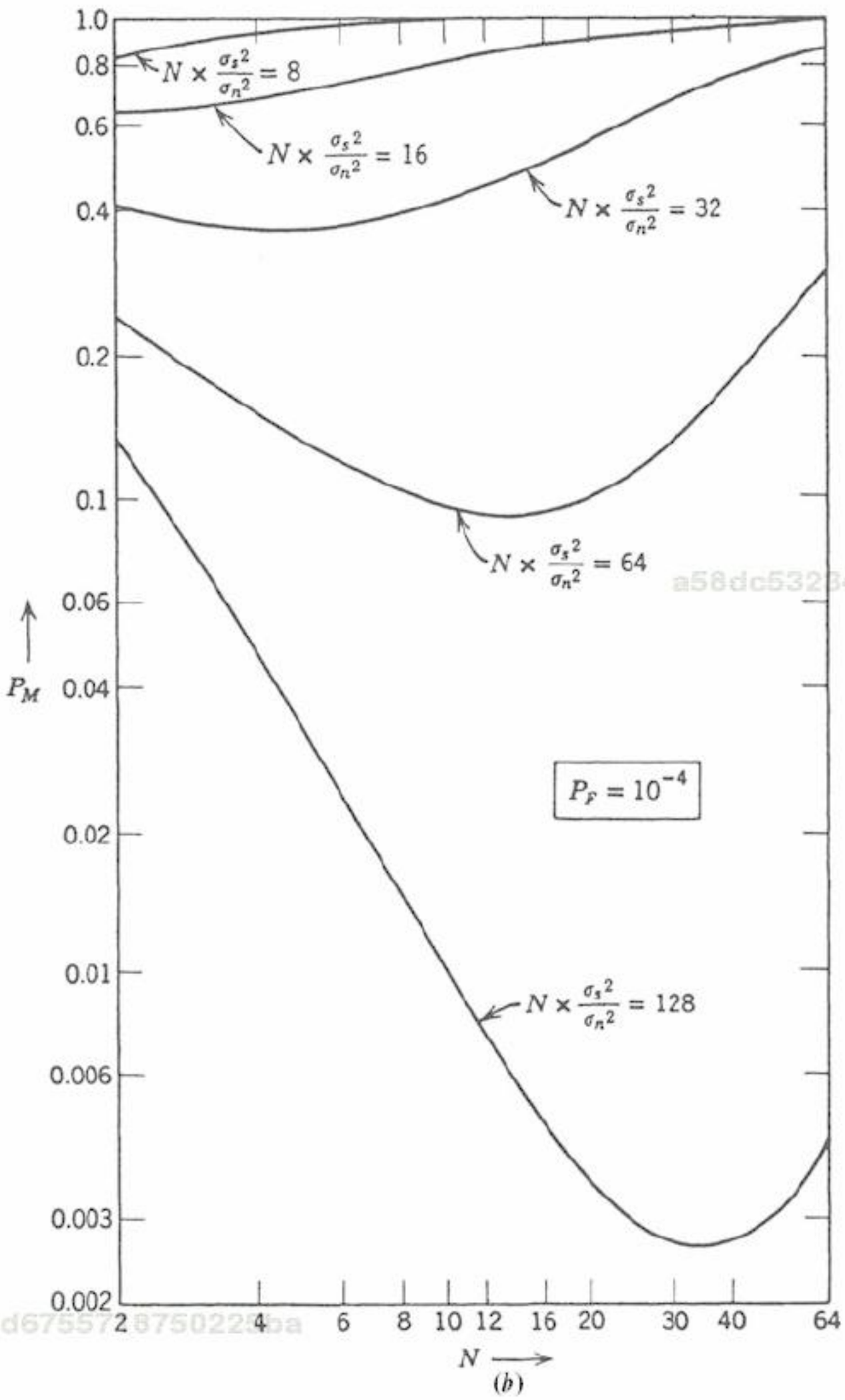


Figure 2.16: (*b*) P_M as a function of $N[P_F = 10^{-4}]$.

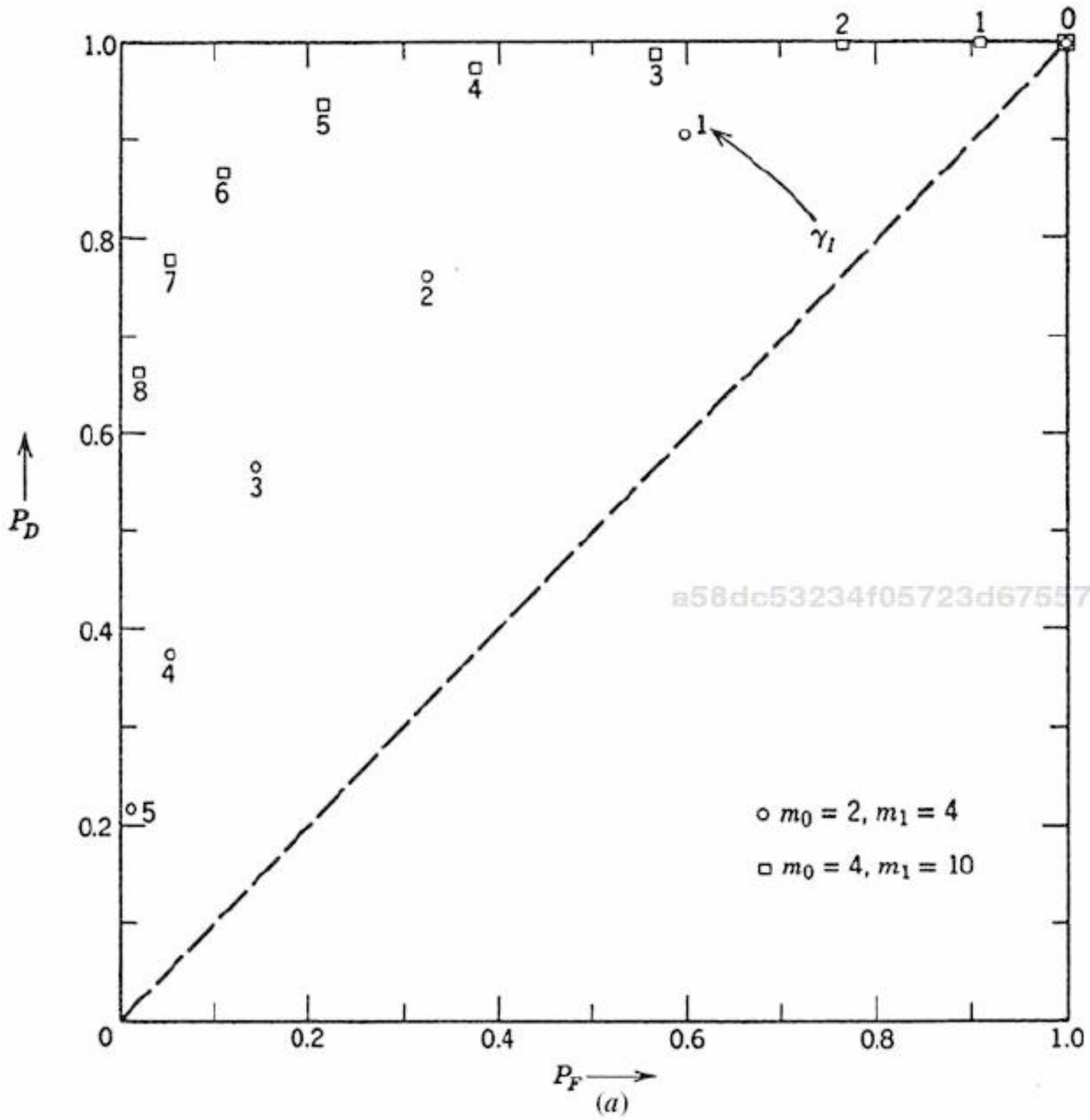
where γ_I takes on only integer values. Using (2.35),

$$P_D = 1 - e^{-m_1} \sum_{n=0}^{\gamma_I - 1} \frac{(m_1)^n}{n!}, \quad \gamma_I = 0, 1, 2, \ldots, \tag{2.115}$$

and from (2.36)

$$P_F = 1 - e^{-m_0} \sum_{n=0}^{\gamma_I - 1} \frac{(m_0)^n}{n!}, \quad \gamma_I = 0, 1, 2, \ldots. \tag{2.116}$$

The resulting ROC is plotted in Figure 2.17*a* for some representative values of m_0 and m_1.



(*a*)

Figure 2.17: (*a*) Receiver operating characteristic, Poisson problem.

We see that it consists of a series of points and that P_F goes from 1 to $1 - e^{-m_0}$ when the threshold is changed from 0 to 1. Now suppose we wanted P_F to have an intermediate value, say $1 - \frac{1}{2}e^{-m_0}$. To achieve this performance, we proceed in the following manner. Denoting the LRT with $\gamma_I = 0$ as LRT No. 0 and the LRT with $\gamma_I = 1$ as LRT No. 1, we have the following table:

LRT	γ_I	P_F	P_D
0	0	1	1
1	1	$1 - e^{-m_0}$	$1 - e^{-m_1}$

To get the desired value of P_F, we use LRT No. 0 with probability $\frac{1}{2}$ and LRT No. 1 with probability $\frac{1}{2}$. The test is

$$\begin{aligned} &\text{If } n = 0, \quad \text{say } H_1 \text{ with probability } \tfrac{1}{2}, \\ &\qquad\qquad \text{say } H_0 \text{ with probability } \tfrac{1}{2}, \\ &\text{If } n \geqslant 1, \quad \text{say } H_1. \end{aligned}$$

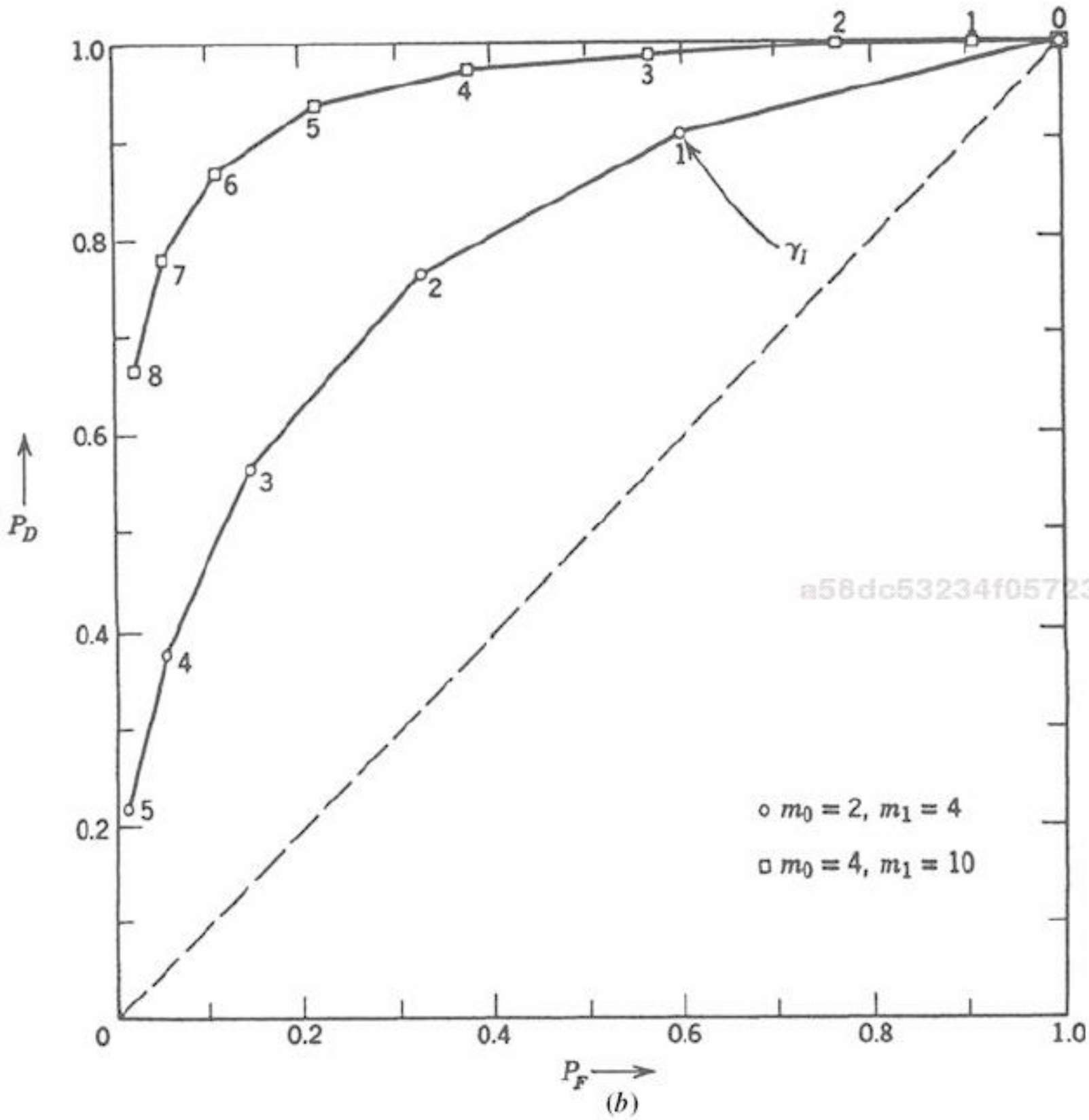


Figure 2.17: (*b*) Receiver operating characteristic with randomized decision rule.

This procedure, in which we mix two likelihood ratio tests in some probabilistic manner, is called a *randomized decision rule*. The resulting P_D is simply a weighted combination of detection probabilities for the two tests.

$$P_D = 0.5(1) + 0.5(1 - e^{-m_1}) = 1 - 0.5e^{-m_1}. \tag{2.117}$$

We see that the ROC for randomized tests consists of straight lines which connect the points in Figure 2.17*a*, as shown in Figure 2.17*b*. The reason that we encounter a randomized test is that the observed random variables are discrete. Therefore, $\Lambda(\mathbf{R})$ is a discrete random variable and, using an ordinary likelihood ratio test, only certain values of P_F are possible. ■

Looking at the expression for P_F in (2.73) and denoting the threshold by η, we have

$$P_F(\eta) = \int_{\eta}^{\infty} p_{\Lambda|H_0}(X|H_0)\,dX. \tag{2.118}$$

If $P_F(\eta)$ is a continuous function of η, we can achieve a desired value from 0 to 1 by a suitable choice of η and a randomized test will never be needed. This is the only case of interest to us in the sequel (see Problem 2.2.12).

In each of the first three examples, we were able to compute the probability density of the sufficient statistic on H_0 and H_1. However, in Example 2.4, this is difficult to do for arbitrary α. For observations that are IID, a Matlab-based approach provides a good alternative. We discuss this case briefly. From (2.39) and (2.42),

$$\Lambda(\mathbf{R}) = \frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} = \prod_{i=1}^{N} \frac{p_{r|H_1}(R_i|H_1)}{p_{r|H_0}(R_i|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{2.119}$$

$$l(\mathbf{R}) = \ln \Lambda(\mathbf{R}) = \sum_{i=1}^{N} \ln \Lambda(R_i), \tag{2.120}$$

where

$$\ln \Lambda(R_i) \triangleq \ln \frac{p_{r|H_1}(R_i|H_1)}{p_{r|H_0}(R_i|H_0)}. \tag{2.121}$$

To evaluate performance, we need to compute

$$P_D = \int_{\ln \eta}^{\infty} p_{l|H_1}(L|H_1)dL \tag{2.122}$$

and

$$P_F = \int_{\ln \eta}^{\infty} p_{l|H_0}(L|H_0)dL. \tag{2.123}$$

In many cases, we can compute the probability density of $\ln \Lambda(R_i)$ on H_0 and H_1, but an analytic formula for the density of $l(\mathbf{R})$ is hard to find. However, we know that it is an $(N-1)$-fold convolution of identical densities. This is straightforward to carry out using Matlab. We then perform numerical integration to obtain P_D and P_F.

Example 2.8 (continuation of Example 2.4). In Figure 2.18, we plot the ROC for $m = 1$, $\sigma = 0.5$, $N = 10$, and $\alpha = 1.1$, 1.5, and 2.0. For $\alpha = 2$, the density is Gaussian, so the result is the same as Example 2.5. ■

With these examples as a background, we now derive a few general properties of receiver operating characteristics. We confine our discussion to continuous likelihood ratio tests.

Two properties of all ROC's follow immediately from this example.

Property 1. All continuous likelihood ratio tests have ROC's that are concave downward. If they were not, a randomized test would be better. This would contradict our proof that a LRT is optimum (see Problem 2.2.12).

Property 2. All continuous likelihood ratio tests have ROC's that are above the $P_D = P_F$ line. This is just a special case of Property 1 because the points ($P_F = 0$, $P_D = 0$) and ($P_F = 1$, $P_D = 1$) are contained on all ROC's.

Property 3. The slope of a curve in a ROC at a particular point is equal to the value of the threshold η required to achieve the P_D and P_F of that point.

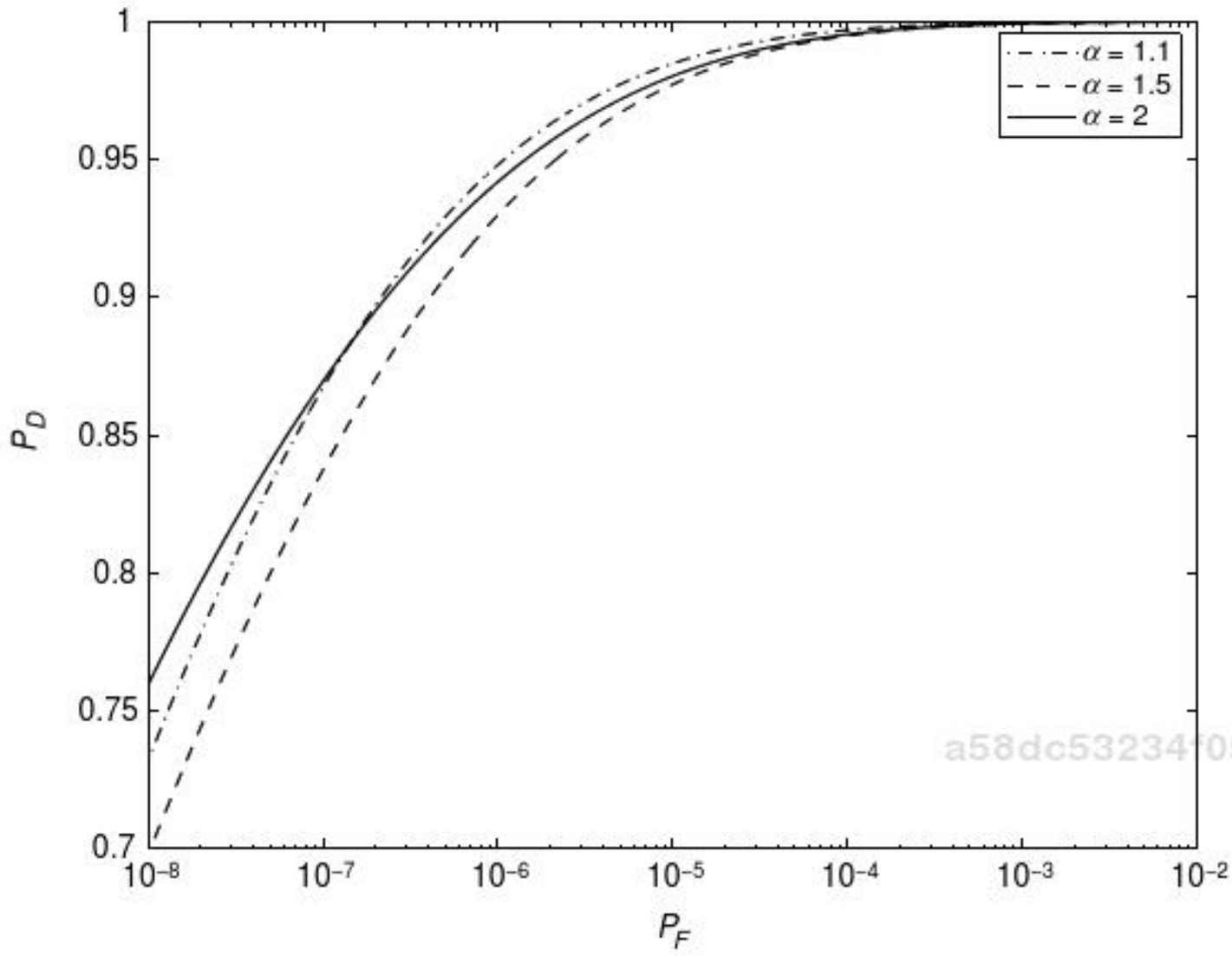


Figure 2.18: P_D versus P_F for Generalized Gaussian probability density: $m = 1$, $\sigma = 0.5$, $N = 10$; $\alpha = 1.1, 1.5, 2.0$.

Proof.

$$P_D = \int_{\eta}^{\infty} p_{\Lambda|H_1}(\Lambda|H_1)d\Lambda,$$

$$P_F = \int_{\eta}^{\infty} p_{\Lambda|H_0}(\Lambda|H_0)d\Lambda. \tag{2.124}$$

Differentiating both expressions with respect to η and writing the results as a quotient, we have

$$\frac{dP_D/d_\eta}{dP_F/d_\eta} = \frac{-p_{\Lambda|H_1}(\Lambda|H_1)}{-p_{\Lambda|H_0}(\Lambda|H_0)} = \frac{dP_D}{dP_F}. \tag{2.125}$$

We now show that

$$\frac{p_{\Lambda|H_1}(\Lambda|H_1)}{p_{\Lambda|H_0}(\Lambda|H_0)} = \eta. \tag{2.126}$$

Let

$$\Omega(\eta) \triangleq \{\mathbf{R}|\Lambda(\mathbf{R}) \geqslant \eta\} = \left[\mathbf{R}\left|\frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} \geqslant \eta\right.\right]. \tag{2.127}$$

Then,

$$P_D(\eta) \triangleq \Pr\{\Lambda(\mathbf{R}) \geqslant \eta | H_1\} = \int_{\Omega(\eta)} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) d\mathbf{R}$$
$$= \int_{\Omega(\eta)} \Lambda(\mathbf{R}) p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) d\mathbf{R}, \tag{2.128}$$

where the last equality follows from the definition of the likelihood ratio. Using the definition of $\Omega(\eta)$, we can rewrite the last integral

$$P_D(\eta) = \int_{\Omega(\eta)} \Lambda(\mathbf{R}) p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) d\mathbf{R} = \int_{\eta}^{\infty} X p_{\Lambda|H_0}(X|H_0) dX. \tag{2.129}$$

Differentiating (2.129) with respect to η, we obtain

$$\frac{dP_D(\eta)}{d\eta} = -\eta p_{\Lambda|H_0}(X|H_0). \tag{2.130}$$

Equating the expression for $dP_D(\eta)/d\eta$ in the numerator of (2.125) to the right side of (2.130) gives the desired result.

We see that this result is consistent with Example 2.5. In Figure 2.12*a*, the curves for nonzero d have zero slope at $P_F = P_D = 1$ ($\eta = 0$) and infinite slope at $P_F = P_D = 0$ ($\eta = \infty$).

Property 4. Whenever the maximum value of the Bayes risk is interior to the interval (0, 1) on the P_1 axis, the minimax operating point is the intersection of the line

$$(C_{11} - C_{00}) + (C_{01} - C_{11})(1 - P_D) - (C_{10} - C_{00})P_F = 0 \tag{2.131}$$

and the appropriate curve of the ROC (see (2.63)). In Figure 2.19, we show the special case defined by (2.67),

$$C_F P_F = C_M P_M = C_M(1 - P_D), \tag{2.132}$$

superimposed on the ROC of Example 2.5. We see that it starts at the point $P_F = 0$, $P_D = 1$, and intersects the $P_F = 1$ line at

$$P_F = 1 - \frac{C_F}{C_M}. \tag{2.133}$$

This completes our discussion of the binary hypothesis testing problem. Several key ideas should be re-emphasized:

1. Using either a Bayes criterion or a Neyman–Pearson criterion, we find that the optimum test is a likelihood ratio test. Thus, regardless of the dimensionality of the observation space, the test consists of comparing a scalar variable $\Lambda(\mathbf{R})$ with a threshold. (We assume $P_F(\eta)$ is continuous.)
2. In many cases, construction of the LRT can be simplified if we can identify a sufficient statistic. Geometrically, this statistic is just that coordinate in a suitable coordinate

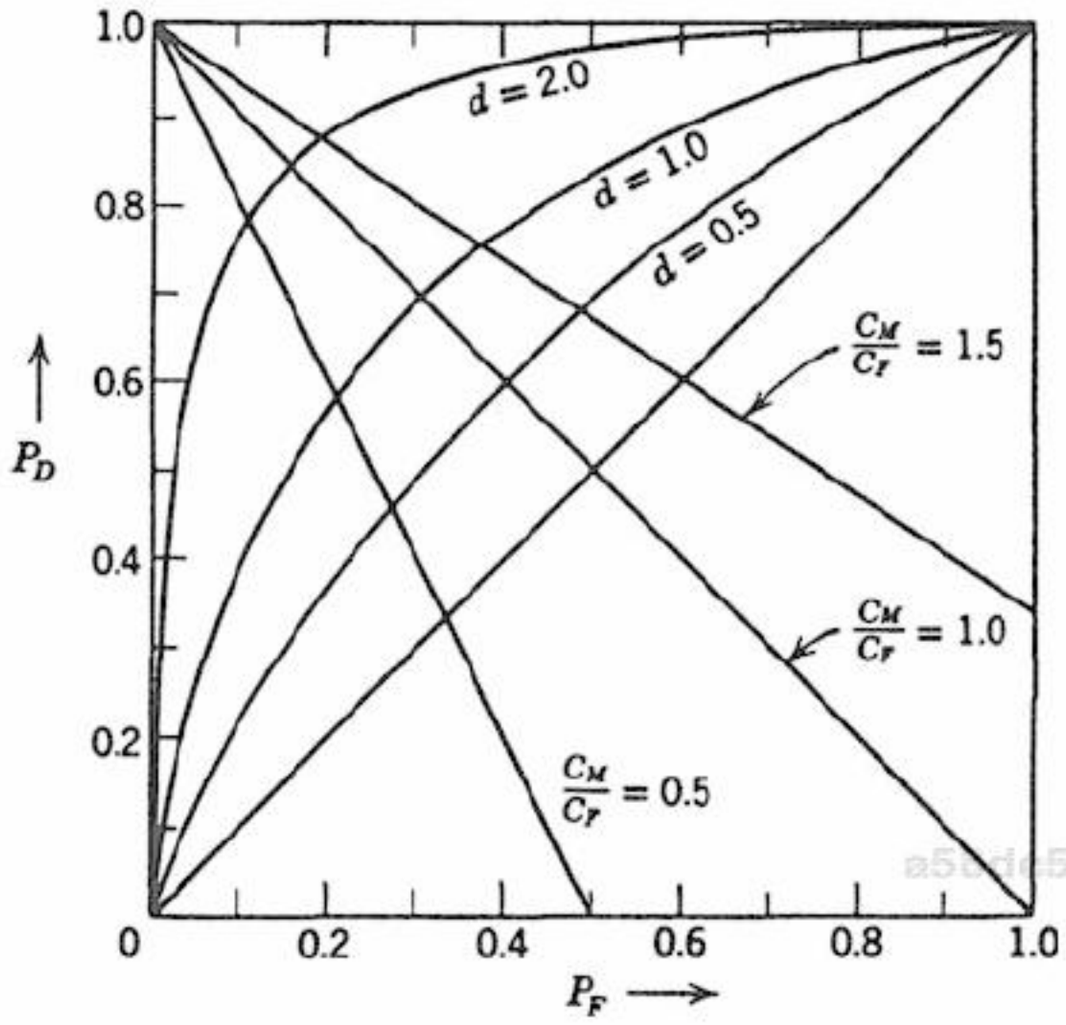


Figure 2.19: Determination of minimax operating point.

system that describes the observation space that contains *all* the information necessary to make a decision.

3. A complete description of the LRT performance was obtained by plotting the conditional probabilities P_D and P_F as the threshold η was varied. The resulting ROC could be used to calculate the Bayes risk for any set of costs. In many cases only one value of the threshold is of interest and a complete ROC is not necessary.

A number of interesting binary tests are developed in the problems.

2.3 *M* HYPOTHESES

The next case of interest is one in which we must choose one of M hypotheses. In the simple binary hypothesis test, there were two source outputs, each of which corresponded to a single hypothesis. In the simple M-ary test, there are M source outputs, each of which corresponds to one of M hypotheses. As before, we assume that we are forced to make a decision. Thus, there are M^2 alternatives that may occur each time the experiment is conducted. The Bayes criterion assigns a cost to each of these alternatives, assumes a set of *a priori* probabilities $P_0, \ldots, P_{M-1}$, and minimizes the risk. The generalization of the Neyman–Pearson criterion to M hypotheses is also possible. Because it is not widely used in practice, we shall discuss only the Bayes criterion in the text.

Bayes Criterion. To find a Bayes test, we denote that cost of each course of action as C_{ij}. The first subscript signifies that the ith hypothesis is chosen. The second subscript signifies that the jth hypothesis is true. We denote the region of the observation space in which we choose H_i as Z_i and the *a priori* probabilities are P_i. The model is shown in Figure 2.20. The expression for the risk is

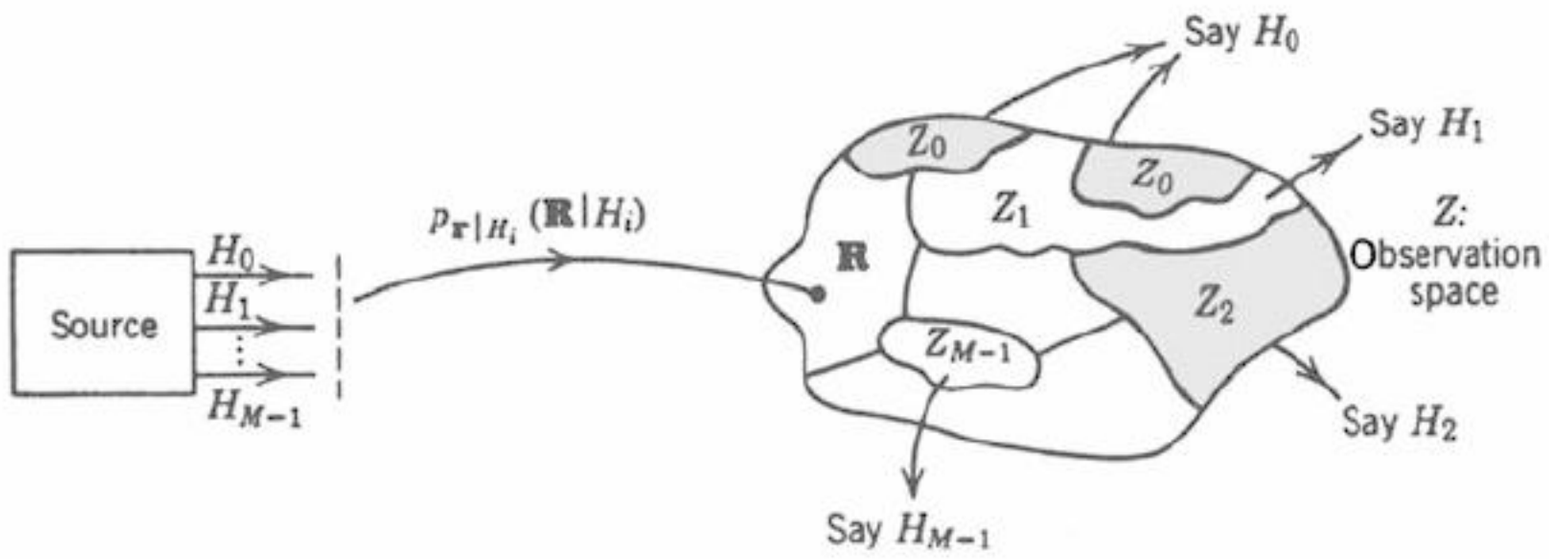


Figure 2.20: *M* hypothesis problem.

$$\mathcal{R} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_j C_{ij} \int_{Z_i} p_{\mathbf{r}|H_j}(\mathbf{R}|H_j)\, d\mathbf{R}. \tag{2.134}$$

To find the optimum Bayes test, we simply vary the Z_i to minimize $\mathcal{R}$. This is a straightforward extension of the technique used in the binary case.

Noting that $Z_0 = Z - Z_1 - Z_2$, because the regions are disjoint, we obtain

$$\begin{aligned}
\mathcal{R} = {} & P_0 C_{00} \int_{Z-Z_1-Z_2} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) d\mathbf{R} + P_0 C_{10} \int_{Z_1} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) d\mathbf{R} \\
& + P_0 C_{20} \int_{Z_2} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) d\mathbf{R} + P_1 C_{11} \int_{Z-Z_0-Z_2} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) d\mathbf{R} \\
& + P_1 C_{01} \int_{Z_0} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) d\mathbf{R} + P_1 C_{21} \int_{Z_2} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) d\mathbf{R} \\
& + P_2 C_{22} \int_{Z-Z_0-Z_1} p_{\mathbf{r}|H_2}(\mathbf{R}|H_2) d\mathbf{R} + P_2 C_{02} \int_{Z_0} p_{\mathbf{r}|H_2}(\mathbf{R}|H_2) d\mathbf{R} \\
& + P_2 C_{12} \int_{Z_1} p_{\mathbf{r}|H_2}(\mathbf{R}|H_2) d\mathbf{R}.
\end{aligned} \tag{2.135}$$

This reduces to

$$\begin{aligned}
\mathcal{R} = {} & P_0 C_{00} + P_1 C_{11} + P_2 C_{22} \\
& + \int_{Z_0} \left[P_2(C_{02} - C_{22}) p_{\mathbf{r}|H_2}(\mathbf{R}|H_2) + P_1(C_{01} - C_{11}) p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) \right] d\mathbf{R} \\
& + \int_{Z_1} \left[P_0(C_{10} - C_{00}) p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) + P_2(C_{12} - C_{22}) p_{\mathbf{r}|H_2}(\mathbf{R}|H_2) \right] d\mathbf{R} \\
& + \int_{Z_2} \left[P_0(C_{20} - C_{00}) p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) + P_1(C_{21} - C_{11}) p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) \right] d\mathbf{R}.
\end{aligned} \tag{2.136}$$

As before, the first three terms represent the fixed cost and the integrals represent the variable cost that depends on our choice of Z_0, Z_1, Z_2. Clearly, we assign each $\mathbf{R}$ to the region in which the value of the integrand is the smallest. Labeling these integrands $I_0(\mathbf{R})$, $I_1(\mathbf{R})$, and $I_2(\mathbf{R})$, we have the following rule:

$$\begin{aligned}
&\text{if } I_0(\mathbf{R}) < I_1(\mathbf{R}) \text{ and } I_2(\mathbf{R}), \text{ choose } H_0,\\
&\text{if } I_1(\mathbf{R}) < I_0(\mathbf{R}) \text{ and } I_2(\mathbf{R}), \text{ choose } H_1,\\
&\text{if } I_2(\mathbf{R}) < I_0(\mathbf{R}) \text{ and } I_1(\mathbf{R}), \text{ choose } H_2.
\end{aligned} \tag{2.137}$$

We can write terms in terms of likelihood ratios by defining

$$\begin{aligned}
\Lambda_1(\mathbf{R}) &\triangleq \frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)},\\
\Lambda_2(\mathbf{R}) &\triangleq \frac{p_{\mathbf{r}|H_2}(\mathbf{R}|H_2)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)}.
\end{aligned} \tag{2.138}$$

Using (2.138) in (2.136) and (2.137), we have

$$P_1(C_{01}-C_{11})\Lambda_1(\mathbf{R}) \underset{H_0 \text{ or } H_2}{\overset{H_1 \text{ or } H_2}{\gtrless}} P_0(C_{10}-C_{00}) + P_2(C_{12}-C_{02})\Lambda_2(\mathbf{R}), \tag{2.139}$$

$$P_2(C_{02}-C_{22})\Lambda_2(\mathbf{R}) \underset{H_0 \text{ or } H_1}{\overset{H_2 \text{ or } H_1}{\gtrless}} P_0(C_{20}-C_{00}) + P_1(C_{21}-C_{01})\Lambda_1(\mathbf{R}), \tag{2.140}$$

$$P_2(C_{12}-C_{22})\Lambda_2(\mathbf{R}) \underset{H_1 \text{ or } H_0}{\overset{H_2 \text{ or } H_0}{\gtrless}} P_0(C_{20}-C_{10}) + P_1(C_{21}-C_{11})\Lambda_1(\mathbf{R}). \tag{2.141}$$

We see that the decision rules correspond to three lines in the (Λ_1, Λ_2) plane. It is easy to verify that these lines intersect at a common point and therefore uniquely define three decision regions, as shown in Figure 2.21. The decision space is two dimensional for the three-hypothesis problem. It is easy to verify that M hypotheses *always* lead to a decision space that has, at most, $(M-1)$ dimensions.

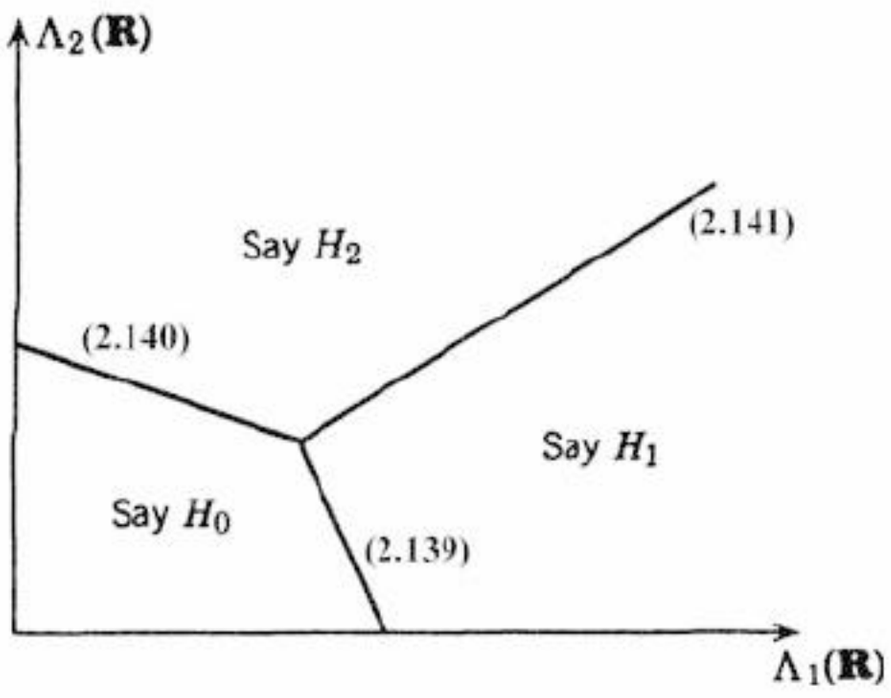


Figure 2.21: Decision space.

Several special cases will be useful in our later work. The first is defined by the assumptions

$$C_{00} = C_{11} = C_{22} = 0,$$
$$C_{ij} = 1, \quad i \neq j. \tag{2.142}$$

These equations indicate that any error is of equal importance. Looking at (2.134), we see that this corresponds to minimizing the total probability of error.

Substituting into (2.139)–(2.141), we have

$$P_1\Lambda_1(\mathbf{R}) \underset{H_0 \text{ or } H_2}{\overset{H_1 \text{ or } H_2}{\gtrless}} P_0,$$
$$P_2\Lambda_2(\mathbf{R}) \underset{H_0 \text{ or } H_1}{\overset{H_2 \text{ or } H_1}{\gtrless}} P_0, \tag{2.143}$$
$$P_2\Lambda_2(\mathbf{R}) \underset{H_1 \text{ or } H_0}{\overset{H_2 \text{ or } H_0}{\gtrless}} P_1\Lambda_1(\mathbf{R}).$$

The decision regions in the (Λ_1, Λ_2) plane are shown in Figure 2.22*a*. In this particular case, the transition to the $(\ln \Lambda_1, \ln \Lambda_2)$ plane is straightforward (Figure 2.22*b*). The equations are

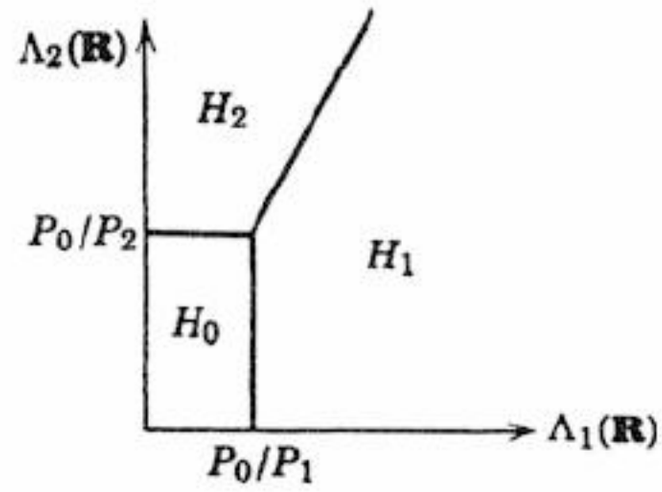


Figure 2.22: Decision spaces.

$$\ln \Lambda_1(\mathbf{R}) \underset{H_0 \text{ or } H_2}{\overset{H_1 \text{ or } H_2}{\gtrless}} \ln \frac{P_0}{P_1},$$
$$\ln \Lambda_2(\mathbf{R}) \underset{H_0 \text{ or } H_1}{\overset{H_2 \text{ or } H_1}{\gtrless}} \ln \frac{P_0}{P_2}, \tag{2.144}$$
$$\ln \Lambda_2(\mathbf{R}) \underset{H_1 \text{ or } H_0}{\overset{H_2 \text{ or } H_0}{\gtrless}} \ln \Lambda_1(\mathbf{R}) + \ln \frac{P_1}{P_2}.$$

The expressions in (2.143) and (2.144) are adequate, but they obscure an important interpretation of the processor. The desired interpretation is obtained by a little manipulation.

Substituting (2.138) into (2.139)–(2.141) and multiplying both sides by $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$, we have

$$P_1 p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) \underset{H_0 \text{ or } H_2}{\overset{H_1 \text{ or } H_2}{\gtrless}} P_0 p_{\mathbf{r}|H_0}(\mathbf{R}|H_0),$$
$$P_2 p_{\mathbf{r}|H_2}(\mathbf{R}|H_2) \underset{H_0 \text{ or } H_1}{\overset{H_2 \text{ or } H_1}{\gtrless}} P_0 p_{\mathbf{r}|H_0}(\mathbf{R}|H_0), \tag{2.145}$$
$$P_2 p_{\mathbf{r}|H_2}(\mathbf{R}|H_2) \underset{H_1 \text{ or } H_0}{\overset{H_2 \text{ or } H_0}{\gtrless}} P_1 p_{\mathbf{r}|H_1}(\mathbf{R}|H_1).$$

Looking at (2.145), we see that an equivalent test is to compute *a posteriori* probabilities $\Pr(H_0|\mathbf{R})$, $\Pr(H_1|\mathbf{R})$, and $\Pr(H_2|\mathbf{R})$ and choose the largest. (Simply divide both sides of each equation by $p_{\mathbf{r}}(\mathbf{R})$ and examine the resulting test.) For this reason, the processor for the minimum probability of error criterion is frequently referred to as a *maximum a posterior probability computer*.

The next two topics deal with degenerate tests. Both results will be useful in later applications. A case of interest is a degenerate one in which we combine H_1 and H_2. Then,

$$C_{12} = C_{21} = 0, \tag{2.146}$$

and, for simplicity, we can let

$$C_{01} = C_{10} = C_{20} = C_{02} \tag{2.147}$$

and

$$C_{00} = C_{11} = C_{22} = 0. \tag{2.148}$$

Then (2.139) and (2.140) both reduce to

$$P_1 \Lambda_1(\mathbf{R}) + P_2 \Lambda_2(\mathbf{R}) \underset{H_0}{\overset{H_1 \text{ or } H_2}{\gtrless}} P_0 \tag{2.149}$$

and (2.141) becomes an identity.

The decision regions are shown in Figure 2.23. Because we have eliminated all of the cost effect of a decision between H_1 and H_2, we have reduced it to a binary problem.

We next consider the dummy hypothesis technique. A simple example illustrates the idea. The actual problem has two hypotheses, H_1 and H_2, but occasionally we can simplify

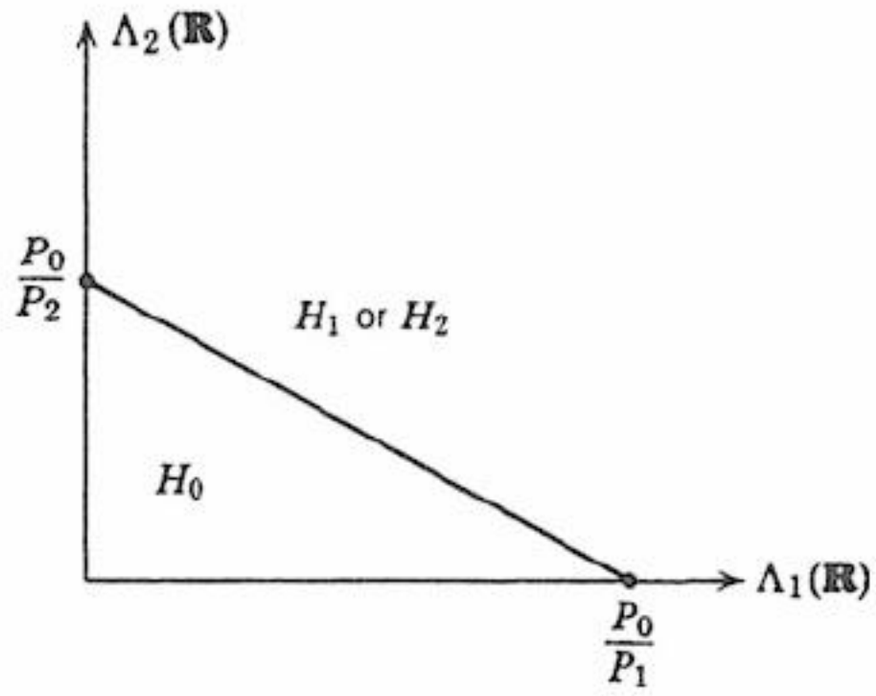


Figure 2.23: Decision spaces.

the calculations by introducing a dummy hypothesis H_0 that occurs with zero probability. We let

$$P_0 = 0, \quad P_1 + P_2 = 1,$$
$$C_{12} = C_{02}, C_{21} = C_{01}. \tag{2.150}$$

Substituting these values into (2.139)–(2.141), we find that (2.139) and (2.140) imply that we always choose H_1 or H_2 and the test reduces to

$$P_2(C_{12} - C_{22})\Lambda_2(\mathbf{R}) \underset{H_1}{\overset{H_2}{\gtrless}} P_1(C_{21} - C_{11})\Lambda_1(\mathbf{R}). \tag{2.151}$$

Looking at (2.12) and recalling the definition of $\Lambda_1(\mathbf{R})$ and $\Lambda_2(\mathbf{R})$, we see that this result is exactly what we would expect. [Just divide both sides of (2.12) by $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$.] On the surface this technique seems absurd, but it will turn out to be useful when the ratio

$$\frac{p_{\mathbf{r}|H_2}(\mathbf{R}|H_2)}{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}$$

is difficult to work with and the ratios $\Lambda_1(\mathbf{R})$ and $\Lambda_2(\mathbf{R})$ can be made simple by a proper choice of $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$.

The formulation of the M hypothesis problem in (2.134)–(2.141) leads to an efficient decision space but loses some of the symmetry. The optimum Bayes test can be written in a different manner by defining a function

$$\beta_i(\mathbf{R}) = \sum_{j=0}^{M-1} C_{ij} \Pr(H_j|\mathbf{R}), \quad i = 0, 1, \ldots, M-1. \tag{2.152}$$

Using Bayes rule, we can rewrite (2.134) as

$$\begin{aligned}\mathcal{R} &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} P_j\, C_{ij} \int_{Z_i} p_{\mathbf{r}|H_j}(\mathbf{R}|H_j)\, d\mathbf{R} \\ &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij} \int_{Z_i} p_{\mathbf{r}|H_j}(\mathbf{R}|H_j)\, P_j\, d\mathbf{R} \\ &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij} \int_{Z_i} \Pr(H_j|\mathbf{R})\, p_{\mathbf{r}}(\mathbf{R})\, d\mathbf{R} \\ &= \sum_{i=0}^{M-1} \int_{Z_i} \left(\sum_{j=0}^{M-1} C_{ij}\, \Pr(H_j|\mathbf{R}) \right) p_{\mathbf{r}}(\mathbf{R})\, d\mathbf{R}. \end{aligned} \tag{2.153}$$

Substituting (2.152) into (2.153)

$$\mathcal{R} = \sum_{i=0}^{M-1} \int_{Z_i} \beta_i(\mathbf{R})\, p_{\mathbf{r}}(\mathbf{R})\, d\mathbf{R}. \tag{2.154}$$

Evaluating (2.154) gives,

$$\mathcal{R} = \int_{Z_0} \beta_0(\mathbf{R})\, p_{\mathbf{r}}(\mathbf{R})\, d\mathbf{R} + \int_{Z_1} \beta_1(\mathbf{R})\, p_{\mathbf{r}}(\mathbf{R})\, d\mathbf{R} + \cdots + \int_{Z_{M-1}} \beta_{M-1}(\mathbf{R})\, p_{\mathbf{r}}(\mathbf{R})\, d\mathbf{R}. \tag{2.155}$$

Each particular $\mathbf{R}$ will be included in only one integral. We want to assign it to the region Z_i where it will make the smallest contribution to $\mathcal{R}$. Clearly this is done by choosing the smallest $\beta_i(\mathbf{R})$ and assigning $\mathbf{R}$ to that region. Thus, the optimum Bayes test is, compute

$$\boxed{\beta_i(\mathbf{R}) = \sum_{j=0}^{M-1} C_{ij}\, \Pr(H_j|\mathbf{R}), \quad i = 0, 1, \ldots, M-1} \tag{2.156}$$

and choose the smallest.

For a minimum probability of error test, we consider the costs,

$$\begin{aligned} C_{ii} &= 0, \quad i = 0, 1, \ldots, M-1, \\ C_{ij} &= C, \quad i \neq j; i, j = 0, 1, \ldots, M-1. \end{aligned} \tag{2.157}$$

Substituting into (2.152),

$$\beta_i(\mathbf{R}) = \sum_{j=0}^{M-1} C_{ij}\, \Pr(H_j|\mathbf{R}) = C \sum_{\substack{j=0 \\ j \neq i}}^{M-1} \Pr(H_j|\mathbf{R}), \tag{2.158}$$

or

$$\beta_i(\mathbf{R}) = C\big[1 - \Pr(H_i|\mathbf{R})\big]. \tag{2.159}$$

From (2.159), it is clear that choosing the largest $\Pr(H_i|\mathbf{R})$ is equivalent to choosing the smallest $\beta_i(\mathbf{R})$. This is the *maximum a posteriori* test that we previously encountered in (2.145) for $M = 3$. Thus, the optimum Bayes test is, compute

$$l_j(\mathbf{R}) = \Pr(H_j|\mathbf{R}), \quad j = 0, 1, \ldots, M-1, \tag{2.160}$$

and choose the largest.

A special case that occurs frequently in practice is equal *a priori* probabilities.

$$P_j = \frac{1}{M}, \quad j = 0, 1, \ldots, M-1. \tag{2.161}$$

Then, since

$$\begin{aligned}\Pr(H_j|\mathbf{R}) &= \frac{P_j\, p_{\mathbf{r}|H_j}(\mathbf{R}|H_j)}{p_{\mathbf{r}}(\mathbf{R})} \\ &= \left(\frac{1}{M\, p_{\mathbf{r}}(\mathbf{R})}\right) p_{\mathbf{r}|H_i}(\mathbf{R}|H_j), \quad j = 0, 1, \ldots, M-1,\end{aligned} \tag{2.162}$$

we can compute

$$l_j(\mathbf{R}) = p_{\mathbf{r}|H_j}(\mathbf{R}|H_j), \quad j = 0, 1, \ldots, M-1, \tag{2.163}$$

and choose the largest. Thus the maximum a posteriori test involves computing M sufficient statistics.

We now consider several examples.

Example 2.9. The observed random variable is Gaussian on each of five hypotheses.

$$p_{r|H_j}(R|H_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(R-m_j)^2}{2\sigma^2}\right), \quad -\infty < R < \infty;\ j = 1, 2, \ldots, 5, \tag{2.164}$$

where

$$\begin{aligned} m_1 &= -2m, \\ m_2 &= -m, \\ m_3 &= 0, \\ m_4 &= m, \\ m_5 &= 2m. \end{aligned} \tag{2.165}$$

The hypotheses are equally likely and the criterion is minimum $\Pr(\epsilon)$.

In this case, $M = 5$ and the decision space is one dimensional. From (2.160), we know that to minimize the $\Pr(\epsilon)$, we choose the H_j with the largest *a posteriori* probability.

$$\Pr(H_j|R) = \Pr(R|H_j)\,\frac{P_j}{p_r(R)}. \tag{2.166}$$

Since the H_j are equally likely, this is equivalent to choosing the H_j for which,

$$p_{r|H_j}(R|H_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(R-m_j)^2}{2\sigma^2}\right), \quad j = 1, 2, \ldots, 5, \tag{2.167}$$

is the largest. This, in turn, is equivalent to choosing the H_j for which

$$l_j(R) = |R - m_j|, \quad j = 1, 2, \ldots, 5, \tag{2.168}$$

is the smallest. The decision space and the boundaries are shown in Figure 2.24.

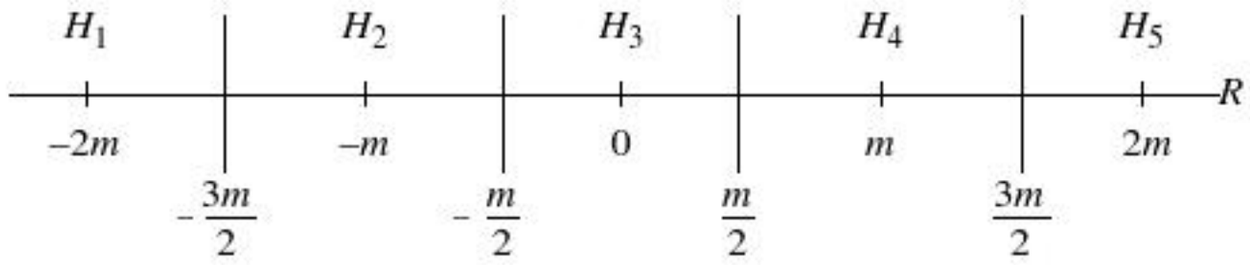


Figure 2.24: Decision space.

The probability of error is

$$\Pr(\epsilon) = \frac{1}{5}\Big[\Pr(\epsilon|H_1) + \cdots + \Pr(\epsilon|H_5)\Big]. \tag{2.169}$$

By comparing Figure 2.24 to Figure 2.11, we have

$$\Pr(\epsilon|H_j) = 2\,\mathrm{erfc}_*\left(\frac{m}{2\sigma}\right), \quad j = 2, 3, 4, \tag{2.170}$$

and

$$\Pr(\epsilon|H_j) = \mathrm{erfc}_*\left(\frac{m}{2\sigma}\right), \quad j = 1, 5, \tag{2.171}$$

so

$$\Pr(\epsilon) = \frac{8}{5}\,\mathrm{erfc}_*\left(\frac{m}{2\sigma}\right). \tag{2.172}$$

■

Example 2.10. Consider the case where $M = 4$ and $N = 2$

$$\begin{aligned} r_1 &= m_{1j} + n_1, \quad j = 0, 1, 2, 3 \\ r_2 &= m_{2j} + n_2, \quad j = 0, 1, 2, 3. \end{aligned} \tag{2.173}$$

The n_1 and n_2 are statistically independent, zero-mean Gaussian random variables with variance σ^2. The m_{1j} and m_{2j} form a two-dimensional vector whose components are

$$\begin{aligned} \mathbf{m}_0 &= \begin{bmatrix} m & 0 \end{bmatrix}^T, \\ \mathbf{m}_1 &= \begin{bmatrix} 0 & m \end{bmatrix}^T, \\ \mathbf{m}_2 &= \begin{bmatrix} -m & 0 \end{bmatrix}^T, \\ \mathbf{m}_3 &= \begin{bmatrix} 0 & m \end{bmatrix}^T. \end{aligned} \tag{2.174}$$

The hypotheses are equally likely and the criteria is minimum probability of error.
The joint probability density of r_1 and r_2 on H_j is

$$p_{r_1,r_2|H_j}(R_1, R_2|H_j) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\Big[(R_1 - m_{1j})^2 + (R_2 - m_{2j})^2\Big]\right). \tag{2.175}$$

From (2.163), we take the logarithm of (2.175) and choose the largest. Dropping terms that do not depend on the hypotheses gives

$$l_j(R_1, R_2) = -\Big[(R_1 - m_{1j})^2 + (R_2 - m_{2j})^2\Big], \quad j = 0, 1, 2, 3. \tag{2.176}$$

Choosing the largest in (2.176) is equivalent to choosing the smallest of

$$D_j = \Big[(R_1 - m_{1j})^2 + (R_2 - m_{2j})^2\Big], \quad j = 0, 1, 2, 3, \tag{2.177}$$

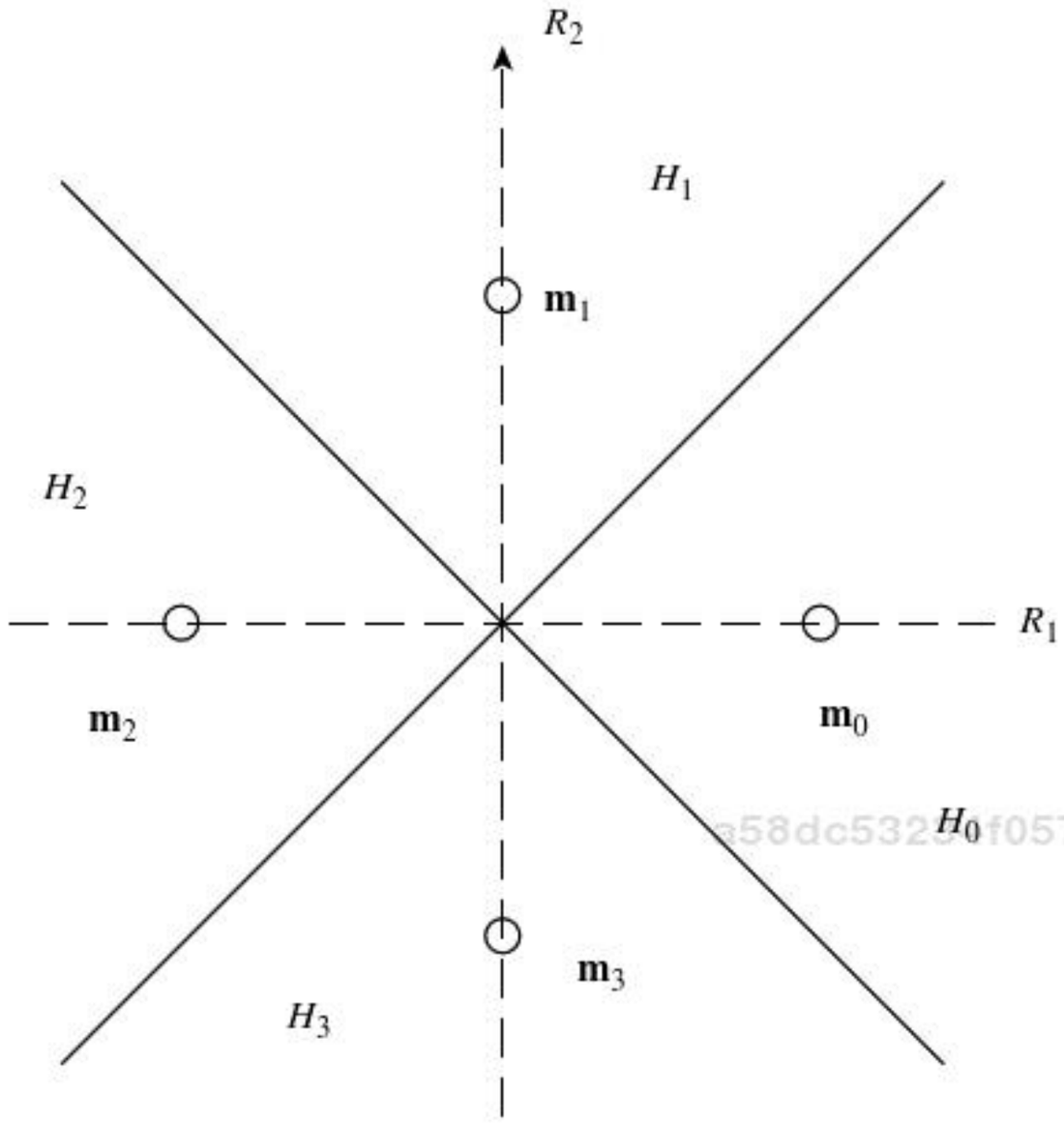


Figure 2.25: Decision space.

which is just the distance between the observations (R_1, R_2) and $\mathbf{m}_j$ in the two-dimensional space shown in Figure 2.25. The result is an example of a *minimum distance decision* rule. We choose the hypothesis corresponding to the mean vector that the observation is closest to.

Note that the minimum distance test applies to any set of $\mathbf{m}_j$. We can also write

$$D_j = -2(R_1 m_{1j} + R_2 m_{2j}) + R_1^2 + R_2^2 + m_{1j}^2 + m_{2j}^2, \quad j = 0, 1, 2, 3, \tag{2.178}$$

and observe that for the $\mathbf{m}_i$ in (2.174) only the first term depends on the hypotheses. Thus, we can compute

$$\mathbf{R}^T \mathbf{m}_j = R_1 m_{1j} + R_2 m_{2j}, \quad j = 0, 1, 2, 3, \tag{2.179}$$

and choose the largest. This is a *correlation* test.

To compute the $\Pr(\epsilon)$, we observe that

$$\Pr(\epsilon) = \frac{1}{4} \sum_{j=0}^{3} \Pr(\epsilon | H_j) \tag{2.180}$$

and that, from the symmetry of the decision space, all of the $\Pr(\epsilon|H_j)$ are identical.

We also observe that the answer would be invariant to a 45° rotation of the signal set because the noise is circularly symmetric.

Thus, the problem of interest reduces to the simple diagram shown in Figure 2.26. The $\Pr(\epsilon)$ is simply the probability that $\mathbf{r}$ lies outside the first quadrant when H_1 is true.

Now r_1 and r_2 are independent Gaussian variables with identical means and variances:

$$\begin{aligned} E(r_1|H_1) &= E(r_2|H_1) = m/\sqrt{2}, \\ \mathrm{Var}(r_1|H_1) &= \mathrm{Var}(r_2|H_1) = \sigma^2. \end{aligned} \tag{2.181}$$

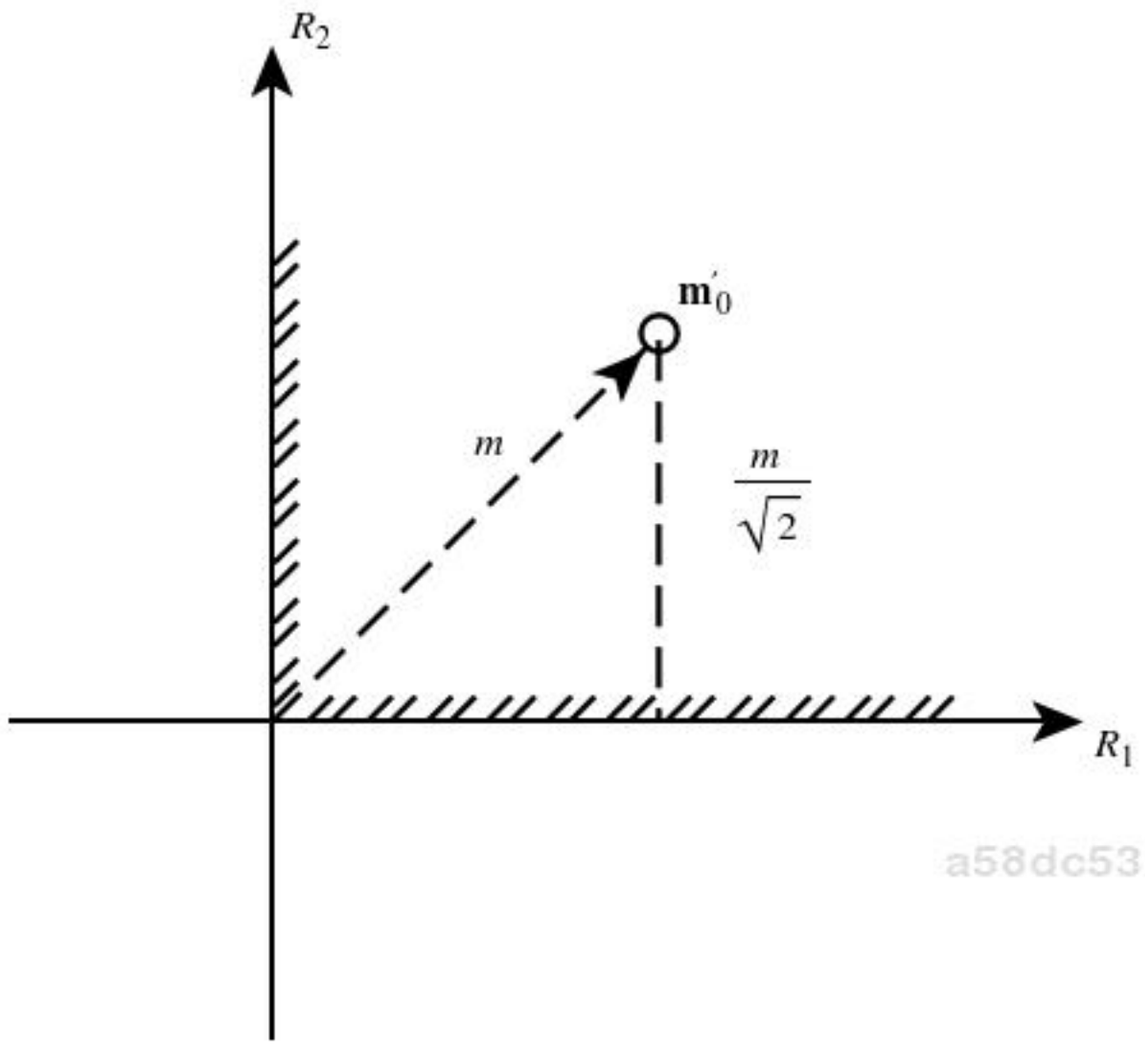


Figure 2.26: Rotation of signal.

The $\Pr(\epsilon)$ can be obtained by integrating $p_{r_1,r_2|H_1}(R_1, R_2|H_1)$ over the area outside the first quadrant. Equivalently, $\Pr(\epsilon)$ is the integral over the first quadrant subtracted from unity,

$$\Pr(\epsilon) = 1 - \left[\int_0^\infty (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(R_1 - m/\sqrt{2})^2}{2\sigma^2}\right) dR_1\right]\left[\int_0^\infty (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(R_2 - m/\sqrt{2})^2}{2\sigma^2}\right) dR_2\right]. \tag{2.182}$$

Changing variables, we have

$$\Pr(\epsilon) = 1 - \left(\int_{-m/\sqrt{2}\sigma}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx\right)^2 = 1 - \left[\operatorname{erfc}_*\left(-\frac{m}{\sqrt{2}\sigma}\right)\right]^2, \tag{2.183}$$

which is the desired result. ■

Example 2.11. In this example, we consider a non-Gaussian density with M hypotheses. We consider the Generalized Gaussian probability density introduced in Example 2.4. We assume that α is the same on all hypotheses. The mean is zero on all hypotheses but the variances σ_j^2 are different. The probability densities are

$$p_{r_i|H_j}(R_i|H_j) = c_j \exp\left(-\left|\frac{R_i}{b_j}\right|^\alpha\right), \quad j = 0, 1, \ldots, M-1, \tag{2.184}$$

where b_j and c_j were defined in (2.48) and (2.49). To simplify notation, we define

$$\beta(\alpha) = \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} = \frac{b_j}{\sigma_j} \tag{2.185}$$

and

$$\gamma(\alpha) = \frac{1}{2\beta(\alpha)\Gamma(1+1/\alpha)} = c_j\sigma_j. \tag{2.186}$$

Then, (2.184) can be rewritten as

$$p_{r_i|H_j}(R_i|H_j) = \frac{1}{\sigma_j}\,\gamma(\alpha)\,\exp\left(-\left|\frac{R_i}{\sigma_j\beta(\alpha)}\right|^{\alpha}\right) \tag{2.187}$$

and

$$p_{\mathbf{r}|H_j}(\mathbf{R}|H_j) = \prod_{i=1}^{N} p_{r_i|H_j}(R_i|H_j). \tag{2.188}$$

We assume that the hypotheses are equally likely, so we choose the H_j for which $p_{\mathbf{r}|H_j}(\mathbf{R}|H_j)$ is the largest.

$$\begin{aligned}\ln p_{\mathbf{r}|H_j}(\mathbf{R}|H_j) &= N\ln\gamma(\alpha) - N\ln\sigma_j - \sum_{i=1}^{N}\left|\frac{R_i}{\sigma_j\beta(\alpha)}\right|^{\alpha} \\ &= N\ln\gamma(\alpha) - N\ln\sigma_j - \frac{1}{\left[\sigma_j\beta(\alpha)\right]^{\alpha}}\sum_{i=1}^{N}|R_i|^{\alpha}.\end{aligned} \tag{2.189}$$

We can see that a sufficient statistic is

$$l_N(\mathbf{R}) = \sum_{i=1}^{N}|R_i|^{\alpha}. \tag{2.190}$$

We define

$$\bar{l}_N(\mathbf{R}) = \frac{1}{N}\sum_{i=1}^{N}|R_i|^{\alpha} \tag{2.191}$$

and choose the largest of

$$l_j(\mathbf{R}) = -\ln\sigma_j - \frac{1}{\left[\sigma_j\beta(\alpha)\right]^{\alpha}}\bar{l}_N(\mathbf{R}), \quad j = 0, 1, \ldots, M-1. \tag{2.192}$$

In Figure 2.27, we plot $l_j(\mathbf{R})$ versus $\bar{l}_N(\mathbf{R})$ for $M = 4$ hypotheses with $\alpha = 1$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 5$, and $\sigma_4 = 10$. We see that there are four distinct regions where each of the $l_j(\mathbf{R})$ is largest. The performance can be evaluated, but it is tedious. ■

In this section, we have developed the basic results needed for the M hypothesis problem. Several important points should be emphasized.

1. The dimension of the decision space is no more than $M-1$. The boundaries of the decision regions are hyperplanes in the $(\Lambda_1, \ldots, \Lambda_{M-1})$ plane.
2. The optimum Bayes test is straightforward to find. We compute

$$\beta_i(\mathbf{R}) = \sum_{j=0}^{M-1} C_{ij}\,\Pr(H_j|\mathbf{R}), \quad i = 0, 1, \ldots, M-1, \tag{2.193}$$

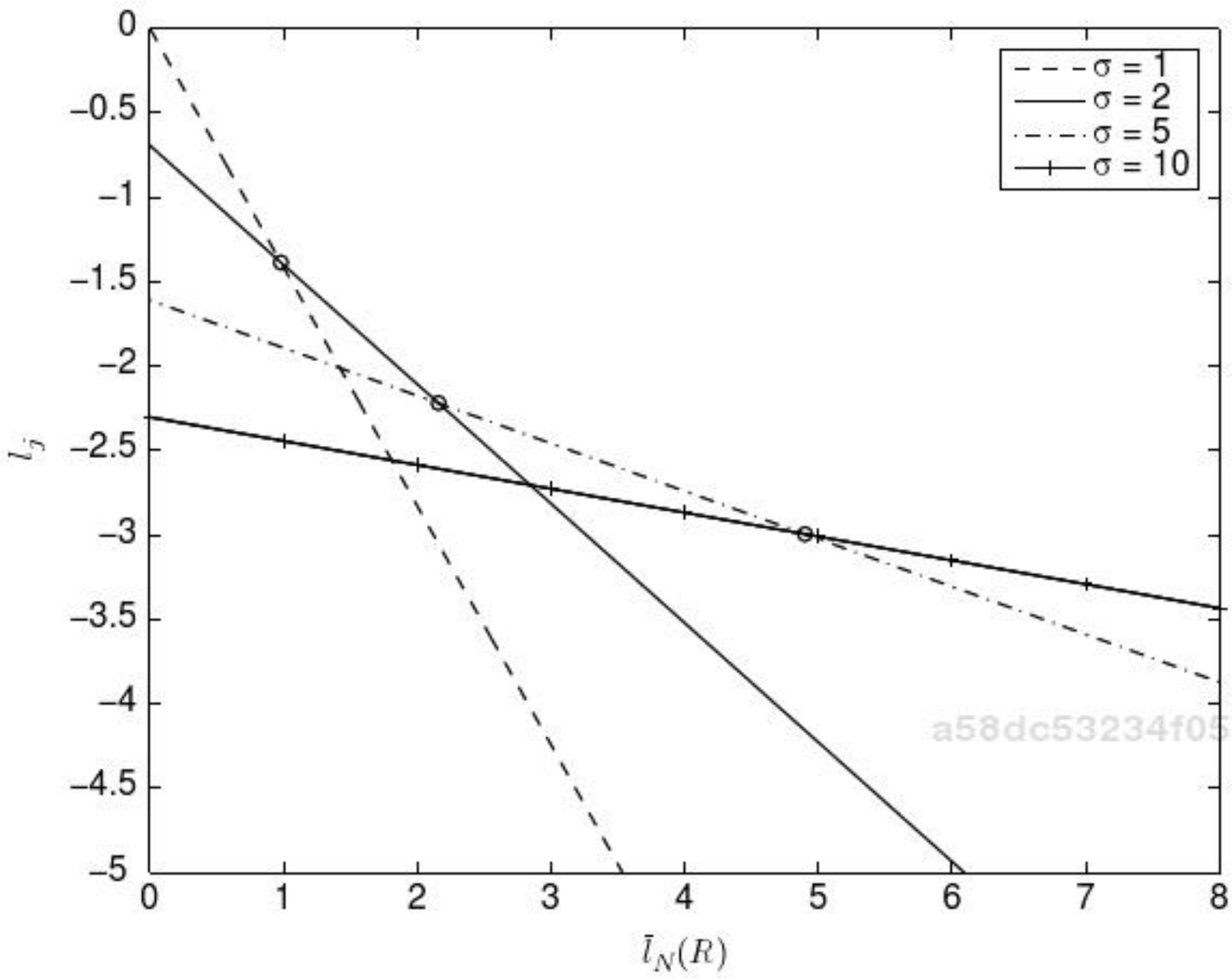


Figure 2.27: l_j **versus** $\bar{l}_N(\mathbf{R})$**.**

or

$$\beta_i'(\mathbf{R}) = \sum_{j=0}^{M-1} C_{ij}\, P_j\, p_{\mathbf{r}|H_j}(\mathbf{R}|H_j), \quad i = 0, 1, \ldots, M-1, \tag{2.194}$$

and choose the smallest. We shall find however, when we consider specific examples, that the error probabilities are frequently difficult to compute.

3. A particular test of importance is the minimum total probability of error test. Here we compute the *a posteriori* probability of each hypothesis $\Pr(H_i|\mathbf{R})$ and choose the largest.

These points will be appreciated more fully as we proceed through various applications.

2.4 PERFORMANCE BOUNDS AND APPROXIMATIONS

Up to this point, we have dealt primarily with problems in which we could derive the structure of the optimum receiver and obtain relatively simple expressions for the receiver operating characteristic or the error probability.

In many cases of interest, the optimum test can be derived but an exact analytic performance calculation is difficult or impossible. For these cases, we must resort to bounds on the error probabilities or approximate expressions for these probabilities. In this section, we derive some simple bounds and approximations that are useful in many problems of practical importance. The basic results, due to Chernoff [Che62], were extended initially by Shannon [Sha56]. They have been further extended by Fano [Fan61], Shannon, Gallager, and Berlekamp [SGB67], and Gallager [Gal65] and applied to a problem of interest to us by Jacobs [Jac66]. Our approach is based on the last two references. Because the latter part

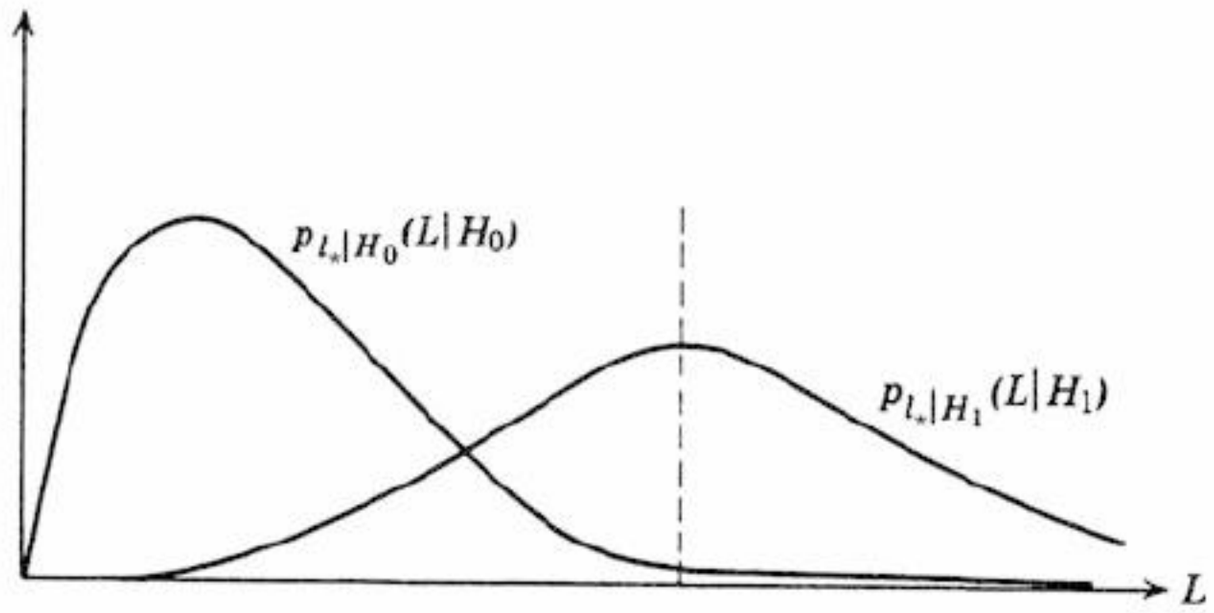


Figure 2.28: Typical densities.

of the development is heuristic in nature, the interested reader should consult the references given for more careful derivations.

The problem of interest is the general binary hypothesis test outlined in Section 2.2. From our results in that section we know that it will reduce to a likelihood ratio test. We begin our discussion at this point.

The likelihood ratio test is

$$l_*(\mathbf{R}) \triangleq \ln \Lambda(\mathbf{R}) = \ln \left[\frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} \right] \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta \triangleq \gamma_*. \tag{2.195}$$

We use the notation $l_*(\mathbf{R})$ to denote the sufficient statistic that is equal to the log-likelihood ratio and γ_* to denote the threshold for this statistic. The variable $l_*(\mathbf{R})$ is a random variable whose probability density depends on which hypothesis is true. In Figure 2.28, we show a typical $p_{l_*|H_1}(L|H_1)$ and $p_{l_*|H_0}(L|H_0)$.

If the two densities are known, then P_F and P_D are given by

$$P_D(\gamma_*) = \int_{\gamma_*}^{\infty} p_{l_*|H_1}(L|H_1)dL, \tag{2.196}$$

$$P_F(\gamma_*) = \int_{\gamma_*}^{\infty} p_{l_*|H_0}(L|H_0)dL, \tag{2.197}$$

where we have used the notation $P_D(\gamma_*)$ and $P_F(\gamma_*)$ to emphasize the dependence of these probabilities on the value of the threshold γ_*. The difficulty is that it is often hard to find $p_{l_*|H_j}(L|H_j)$; $j = 0, 1$. A common case in practice occurs when the components of $\mathbf{R}$ are statistically independent but are non-Gaussian. In a few cases, analytic expressions for $p_{l_*|H_i}(L|H_j)$; $j = 0, 1$ can be obtained, but in most cases, an $(N-1)$-fold convolution is required. On the other hand, if we set out to synthesize a system, it is inefficient (if not impossible) to try successive systems and evaluate each numerically. Therefore, we should like to find some simpler approximate expressions for the error probabilities.

When we discuss simulation in Section 2.5, we will find that the approximate expressions we derive in this section are the key to efficient simulation.

In this section, we derive some simple expressions that we shall use in the sequel. We first focus our attention on cases in which $l_*(\mathbf{R})$ is a sum of independent random variables. This suggests that its characteristic function may be useful, for it will be the product of the

individual characteristic functions of the R_i. Similarly, the moment-generating function will be the product of the individual moment-generating functions. Therefore, an approximate expression based on one of these functions should be relatively easy to evaluate. The first part of our discussion develops bounds on the error probabilities in terms of the moment-generating function of $l_*(\mathbf{R})$.

In the second part, we consider the case in which $l_*(\mathbf{R})$ is sum of a *large* number of independent random variables. By the use of the central limit theorem, we improve on the results obtained in the first part of the discussion.

We begin by deriving a simple upper bound on $P_F(\gamma_*)$ in terms of the moment-generating function. The moment-generating function of $l_*(\mathbf{R})$ on hypothesis H_0 is

$$\phi_{l_*|H_0}(s) \triangleq E\left(e^{sl_*}|H_0\right) = \int_{-\infty}^{\infty} e^{sL} p_{l_*|H_0}(L|H_0)\, dL, \tag{2.198}$$

where s is a real variable. (The range of s corresponds to those values for which the integral exists.) We shall see shortly that it is more useful to write

$$\phi_{l_*|H_0}(s) \triangleq \exp[\mu(s)], \tag{2.199}$$

so that

$$\mu(s) = \ln \int_{-\infty}^{\infty} e^{sL} p_{l_*|H_0}(L|H_0)\, dL. \tag{2.200}$$

We may also express $\mu(s)$ in terms of $p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)$ and $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$. Because l_* is just a function of $\mathbf{r}$, we can write (2.198) as

$$\phi_{l_*|H_0}(s) = \int_{-\infty}^{\infty} e^{sl_*(\mathbf{R})} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\, d\mathbf{R}. \tag{2.201}$$

Then,

$$\mu(s) = \ln \int_{-\infty}^{\infty} e^{sl_*(\mathbf{R})} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\, d\mathbf{R}. \tag{2.202}$$

Using (2.195),

$$\mu(s) = \ln \int_{-\infty}^{\infty} \left(\frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)}\right)^s p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\, d\mathbf{R}, \tag{2.203}$$

or

$$\boxed{\mu(s) = \ln \int_{-\infty}^{\infty} \left[p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)\right]^s \left[p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\right]^{1-s} d\mathbf{R}.} \tag{2.204}$$

The function $\mu(s)$ plays a central role in the succeeding discussion. It is now convenient to rewrite the error expressions in terms of a new random variable whose mean is in the vicinity of the threshold. The reason for this step is that we shall use the central limit theorem in the

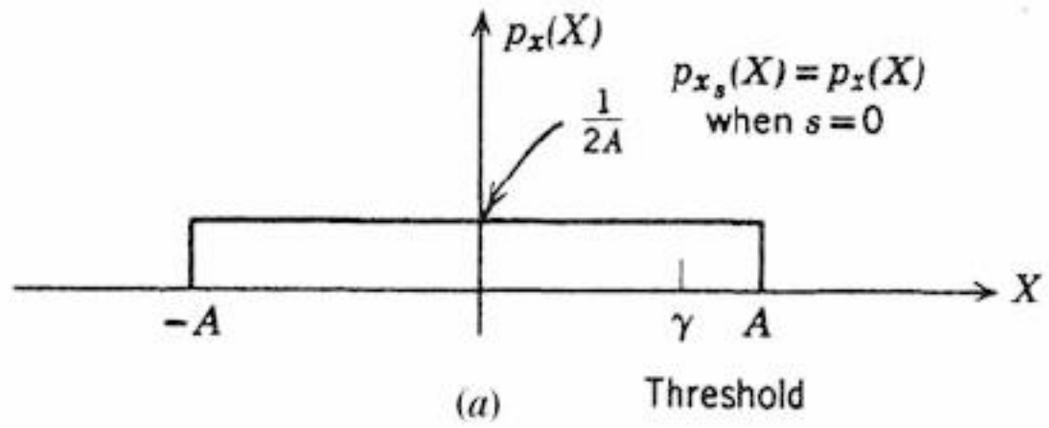


Figure 2.29: Tilted probability densities.

second part of our derivation. This is most effective near the mean of the random variable of interest. Consider the simple probability density shown in Figure 2.29*a*. To get the new family of densities shown in Figure 2.29, we multiply $p_x(X)$ by e^{sX} for various values of s (and normalize to obtain a unit area). We see that for $s > 0$ the mean is shifted to the right. For the moment, we leave s as a parameter. We see that increasing s "tilts" the density more.

Denoting this new variable as x_s, we have

$$p_{x_s}(X) \triangleq \frac{e^{sX} p_{l_*|H_0}(X|H_0)}{\int_{-\infty}^{\infty} e^{sL} p_{l_*|H_0}(L|H_0)\, dL} = \frac{e^{sX} p_{l_*|H_0}(X|H_0)}{e^{\mu(s)}} = e^{sX-\mu(s)}\, p_{l_*|H_0}(X|H_0). \quad (2.205)$$

Observe that we define x_s in terms of its density function, for that is what we are interested in. Equation (2.205) is a general definition. For the density shown in Figure 2.29, the limits would be $(-A, A)$.

We now find the mean and variance of x_s:

$$E(x_s) = \int_{-\infty}^{\infty} X p_{x_s}(X) dX = \frac{\int_{-\infty}^{\infty} X e^{sX} p_{l_*|H_0}(X|H_0)\, dX}{\int_{-\infty}^{\infty} e^{sL} p_{l_*|H_0}(L|H_0)\, dL}. \tag{2.206}$$

Comparing (2.206) and (2.200), we see that

$$E(x_s) = \frac{d\mu(s)}{ds} \triangleq \dot{\mu}(s). \tag{2.207}$$

Similarity, we find

$$\text{Var}(x_s) = \ddot{\mu}(s). \tag{2.208}$$

[Observe that (2.208) implies that $\mu(s)$ is convex.]

We now rewrite $P_F(\gamma_*)$ in terms of this tilted variable x_s:

$$P_F(\gamma_*) = \int_{\gamma_*}^{\infty} p_{l_*|H_0}(L|H_0)\, dL = \int_{\gamma_*}^{\infty} e^{\mu(s)-sX} p_{x_s}(X)\, dX$$

$$= e^{\mu(s)} \int_{\gamma_*}^{\infty} e^{-sX} p_{x_s}(X)\, dX. \tag{2.209}$$

We can now find a simple upper bound on $P_F(\gamma_*)$. For values of $s \geqslant 0$,

$$e^{-sX} \leqslant e^{-s\gamma_*}, \quad \text{for } X \geqslant \gamma_*. \tag{2.210}$$

Thus,

$$P_F(\gamma_*) \leqslant e^{\mu(s)-s\gamma_*} \int_{\gamma_*}^{\infty} p_{x_s}(X)\, dX, \quad s \geqslant 0. \tag{2.211}$$

Clearly the integral is less than one. Thus,

$$P_F(\gamma_*) \leqslant e^{\mu(s)-s\gamma_*}, \quad s \geqslant 0. \tag{2.212}$$

To get the best bound, we minimize the right-hand side of (2.212) with respect to s. Differentiating the exponent and setting the result equal to zero, we obtain

$$\dot{\mu}(s) = \gamma_*. \tag{2.213}$$

Because $\ddot{\mu}(s)$ is nonnegative, a solution will exist if

$$\dot{\mu}(0) \leqslant \gamma_* \leqslant \dot{\mu}(\infty). \tag{2.214}$$

Because

$$\dot{\mu}(0) = E(l_*|H_0), \tag{2.215}$$

the left inequality implies that the threshold must be to the right of the meant of l_* on H_0. If $\gamma_* < \dot{\mu}(0)$, the value of s that minimizes the bound in (2.212) will be $s_* = 0$, which gives the bound $P_F(\gamma_*) \leqslant 1$.

Let s_* denote the value of s that is the solution to (2.213),

$$s_* : \dot{\mu}(s_*) = \gamma_*. \tag{2.216}$$

The resulting bound is:

$$P_F(\gamma_*) \leqslant \begin{cases} e^{\mu(s_*)-s_*\dot{\mu}(s_*)}, & \gamma_* \geqslant \dot{\mu}(0) \\ 1 & \gamma_* < \dot{\mu}(0). \end{cases} \tag{2.217}$$

(Note that we have assumed $\mu(s)$ exists for the desired s_*.)

Equation (2.217) is commonly referred to as the Chernoff bound [Che62]. Comparing (2.216) and (2.207), we observe that s_* is chosen so that the mean of the tilted variable x_s is at the threshold γ_*.

The next step is to find a bound on $P_M(\gamma_*)$, the probability of a miss:

$$P_M(\gamma_*) = \int_{-\infty}^{\gamma_*} p_{l_*|H_1}(X|H_1)\,dX, \tag{2.218}$$

which we want to express in terms of the tilted variable x_s.

Using an argument identical to that in (2.124) through (2.130), we see that

$$p_{l_*|H_1}(X|H_1) = e^X p_{l_*|H_0}(X|H_0). \tag{2.219}$$

Substituting (2.219) into the right side of (2.205), we have

$$p_{l_*|H_1}(X|H_1) = e^{\mu(s)+(1-s)X} p_{x_s}(X). \tag{2.220}$$

Substituting into (2.218),

$$P_M(\gamma_*) = e^{\mu(s)} \int_{-\infty}^{\gamma_*} e^{(1-s)X} p_{x_s}(X)\,dX. \tag{2.221}$$

For $s \leqslant 1$,

$$e^{(1-s)X} \leqslant e^{(1-s)\gamma_*}, \quad \text{for } X \leqslant \gamma_*. \tag{2.222}$$

Thus,

$$\begin{aligned} P_M(\gamma_*) &\leqslant e^{\mu(s)+(1-s)\gamma_*} \int_{-\infty}^{\gamma_*} p_{x_s}(X)\,dX \\ &\leqslant e^{\mu(s)+(1-s)\gamma_*}, \quad s \leqslant 1. \end{aligned} \tag{2.223}$$

Once again the bound is minimized for

$$\gamma_* = \dot{\mu}(s) \tag{2.224}$$

if a solution exists for $s \leqslant 1$. Observing that

$$\dot{\mu}(1) = E(l_*|H_1), \tag{2.225}$$

we see that this requires that threshold to be to the left of the mean of l_* on H_1. If $\gamma_* > \dot{\mu}(1)$, then the value of s that minimizes the bound in (2.223) will be $s_* = 1$, which gives the bound $P_M(\gamma_*) \leqslant 1$.

Combining (2.217) and (2.223), we have

$$\begin{aligned} P_F(\gamma_*) &\leqslant \exp[\mu(s_*) - s_*\dot{\mu}(s_*)], & s_* \geqslant 0 \Rightarrow \gamma_* \geqslant \dot{\mu}(0), \\ P_M(\gamma_*) &\leqslant \exp[\mu(s_*) + (1 - s_*)\dot{\mu}(s_*)], & s_* \leqslant 1 \Rightarrow \gamma_* \leqslant \dot{\mu}(1), \end{aligned} \tag{2.226}$$

where

$$s_* : \dot{\mu}(s_*) = \gamma_*. \tag{2.227}$$

Confining $s_* \geqslant 0$ (and therefore $\gamma_* \geqslant \dot{\mu}(0)$) for the $P_F(\gamma_*)$ bound and $s_* \leqslant 1$ (therefore $\gamma_* \leqslant \dot{\mu}(1)$) for the $P_M(\gamma_*)$ bound is not too restrictive because if the threshold is not between the means then one of the error probabilities will be large (greater than one half if the median coincides with the mean).

As pointed out in [SGB67], the exponents have a simple graphical interpretation. A typical $\mu(s)$ is shown in Figure 2.30. We draw a tangent at the point at which $\dot{\mu}(s) = \gamma_*$. This tangent intersects vertical lines at $s = 0$ and $s = 1$. The value of the intercept at $s = 0$ is the exponent in the $P_F(\gamma_*)$ bound. The value of the intercept at $s = 1$ is the exponent in the $P_M(\gamma_*)$ bound.

For the special case in which the hypotheses are equally likely and the error costs are equal we know that $\gamma_* = 0$. Therefore to minimize the bound we choose that value of s where $\dot{\mu}(s) = 0$.

The probability of error $\Pr(\epsilon)$ is

$$\Pr(\epsilon) = \frac{1}{2}P_F(\gamma_* = 0) + \frac{1}{2}P_M(\gamma_* = 0). \tag{2.228}$$

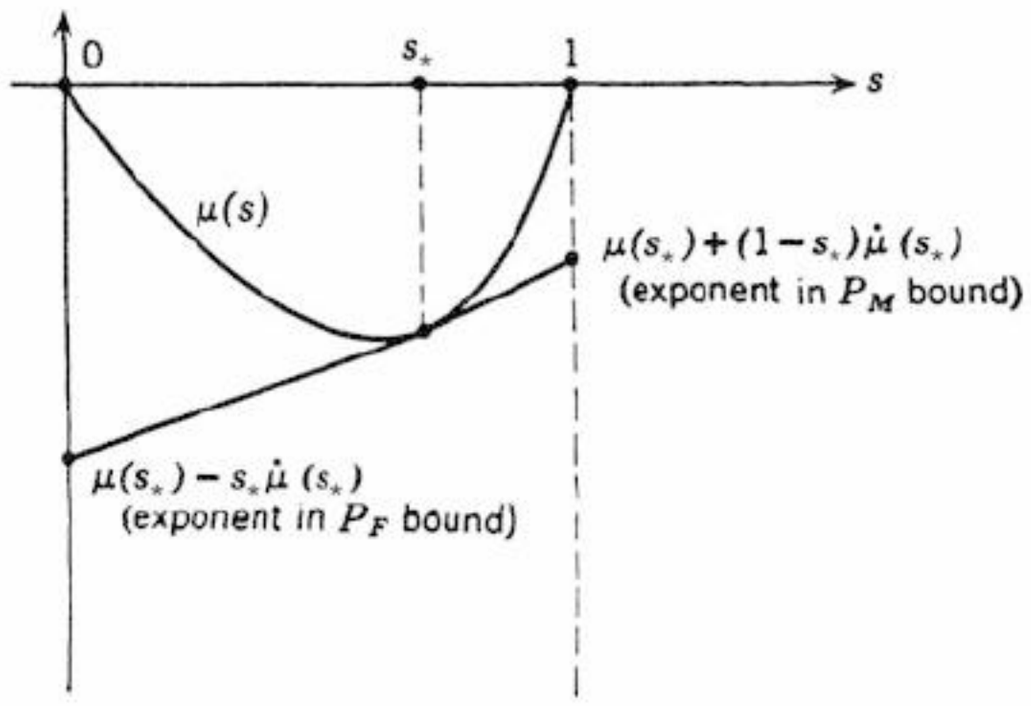


Figure 2.30: Exponents in bounds.

Substituting (2.211) and (2.223) into (2.228) and denoting the value s for which $\dot{\mu}(s) = 0$ as s_m, we have

$$\Pr(\epsilon) \leqslant \frac{1}{2} e^{\mu(s_m)} \int_0^\infty p_{x_s}(X)\, dX + \frac{1}{2} e^{\mu(s_m)} \int_{-\infty}^0 p_{x_s}(X)\, dX, \tag{2.229}$$

or

$$\boxed{\Pr(\epsilon) \leqslant \frac{1}{2} e^{\mu(s_m)}.} \tag{2.230}$$

where

$$s_m : \dot{\mu}(s_m) = 0. \tag{2.231}$$

Up to this point we have considered arbitrary binary hypothesis tests. The bounds in (2.226) and (2.230) are always valid if $\mu(s)$ exists. In many cases of interest, $l_*(\mathbf{R})$ consists of a sum of a large number of independent random variables, and we can obtain a simple approximate expression for $P_F(\gamma_*)$ and $P_M(\gamma_*)$ that provides a much closer estimate of their actual value than the above bounds. The exponent in this expression is the same, but the multiplicative factor will often be appreciably smaller than unity.

We start the derivation with the expression for $P_F(\gamma_*)$ given in (2.209). Motivated by our result in (2.213) in the bound derivation, we choose s_* so that

$$\dot{\mu}(s_*) = \gamma_*. \tag{2.232}$$

Then, (2.209) becomes

$$P_F(\gamma_*) = e^{\mu(s_*)} \int_{\dot{\mu}(s_*)}^\infty e^{-s_* X} p_{x_s}(X)\, dX. \tag{2.233}$$

This can be written as

$$P_F(\gamma_*) = e^{\mu(s_*) - s_* \dot{\mu}(s_*)} \int_{\dot{\mu}(s_*)}^\infty e^{+s_*[\dot{\mu}(s_*) - X]} p_{x_s}(X)\, dX. \tag{2.234}$$

The term outside is just the bound in (2.217). We now use a central limit theorem argument to evaluate the integral. First, define a standardized variable:

$$y \triangleq \frac{x_s - E(x_s)}{[\mathrm{Var}(x_s)]^{1/2}} = \frac{x_s - \dot{\mu}(s_*)}{\sqrt{\ddot{\mu}(s_*)}}. \tag{2.235}$$

Substituting (2.235) into (2.234), we have

$$P_F(\gamma_*) = e^{\mu(s_*) - s_* \dot{\mu}(s_*)} \int_0^\infty e^{-s_* \sqrt{\ddot{\mu}(s_*)} Y} p_y(Y)\, dY. \tag{2.236}$$

In many cases, the probability density governing $\mathbf{r}$ is such that y approaches a Gaussian random variable as N (the number of components of $\mathbf{r}$) approaches infinity.[5] A simple case in which this is true is the case in which the r_i are independent, identically distributed random variables with finite means and variances. In such cases, y approaches a zero-mean Gaussian random variable with unit variance and the integral in (2.236) can be evaluated by substituting the limiting density,

$$\int_0^\infty e^{-s_* \sqrt{\ddot{\mu}(s_*)} Y} \frac{1}{\sqrt{2\pi}} e^{-(Y^2/2)}\, dY = e^{s_*^2 \ddot{\mu}(s_*)/2} \operatorname{erfc}_* \left[s_* \sqrt{\ddot{\mu}(s_*)} \right]. \tag{2.237}$$

Then,

$$\boxed{P_F(\gamma_*) \approx \left\{ \exp\left[\mu(s_*) - s_* \dot{\mu}(s_*) + \frac{s_*^2}{2} \ddot{\mu}(s_*) \right] \right\} \operatorname{erfc}_* \left[s_* \sqrt{\ddot{\mu}(s_*)} \right]}. \tag{2.238}$$

The approximation arises because y is only approximately Gaussian for finite N. For values of $s_* \sqrt{\ddot{\mu}(s_*)} > 3$, we can approximate $\operatorname{erfc}_*(\cdot)$ by the upper bound in (2.89). Using this approximation,

$$P_F(\gamma_*) \approx \frac{1}{\sqrt{2\pi s_*^2 \ddot{\mu}(s_*)}} \exp\left[\mu(s_*) - s_* \dot{\mu}(s_*) \right], \quad s_* > 0. \tag{2.239}$$

It is easy to verify that the approximate expression in (2.239) can also be obtained by letting

$$p_y(Y) \approx p_y(0) \approx \frac{1}{\sqrt{2\pi}}. \tag{2.240}$$

Looking at Figure 2.31, we see that this is valid when the exponential function decreases to a small value while $Y \ll 1$.

In exactly the same manner, we obtain

$$\boxed{P_M(\gamma_*) \approx \left\{ \exp\left[\mu(s_*) + (1 - s_*) \dot{\mu}(s_*) + \frac{(s_* - 1)^2}{2} \ddot{\mu}(s_*) \right] \right\} \operatorname{erfc}_* \left[(1 - s_*) \sqrt{\ddot{\mu}(s_*)} \right]}. \tag{2.241}$$

For $(1 - s_*)\sqrt{\ddot{\mu}(s_*)} > 3$, this reduces to

$$P_M(\gamma_*) \approx \frac{1}{\sqrt{2\pi (1 - s_*)^2 \ddot{\mu}(s_*)}} \exp\left[\mu(s_*) + (1 - s_*) \dot{\mu}(s_*) \right], \quad s_* < 1. \tag{2.242}$$

Observe that the exponents in (2.239) and (2.242) are identical to those obtained by using the Chernoff bound. The central limit theorem argument has provided a multiplicative factor that will be significant in many of the applications of interest to us. In practice, we will normally use (2.238) and (2.241) in our numerical evaluations.

[5] An excellent discussion in contained in Feller [Fel66], pp. 517–520.

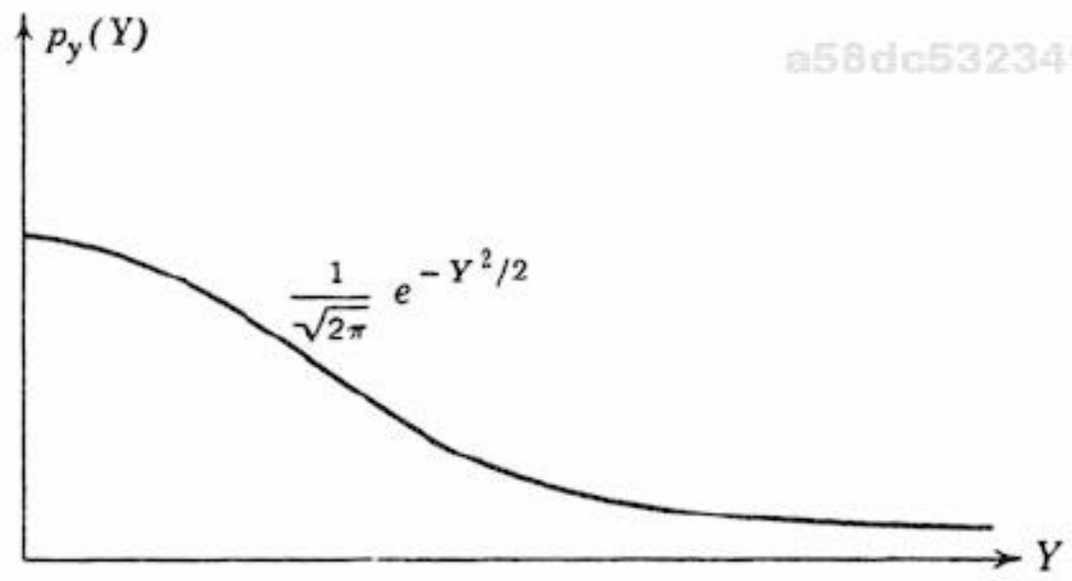


Figure 2.31: Behavior of functions.

For the case in which $\Pr(\epsilon)$ is the criterion and the hypotheses are equally likely we have

$$\begin{aligned}\Pr(\epsilon) &= \frac{1}{2}P_F(0) + \frac{1}{2}P_M(0)\\ &= \frac{1}{2}\exp\left[\mu(s_m) + \frac{s_m^2}{2}\dot{\mu}(s_m)\right]\ \operatorname{erfc}_*\left[s_m\sqrt{\ddot{\mu}(s_m)}\right]\\ &\quad + \frac{1}{2}\exp\left[\mu(s_m) + \frac{(1-s_m)^2}{2}\dot{\mu}(s_m)\right]\ \operatorname{erfc}_*\left[(1-s_m)\sqrt{\ddot{\mu}(s_m)}\right],\end{aligned} \tag{2.243}$$

where s_m is defined in (2.231) [i.e., $\dot{\mu}(s_m) = 0 = \gamma_*$]. When both $s_m\sqrt{\ddot{\mu}(s_m)} > 3$ and $(1-s_m)\sqrt{\ddot{\mu}(s_m)} > 3$, this reduces to

$$\boxed{\Pr(\epsilon) \approx \frac{1}{2\left[2\pi\,\ddot{\mu}(s_m)\right]^{1/2} s_m(1-s_m)}\exp\mu(s_m).} \tag{2.244}$$

Independent Observations. The function $\mu(s)$ defined in (2.204) and its derivatives $\dot{\mu}(s)$ and $\ddot{\mu}(s)$ are the key quantities in the bounds and approximations derived in this section. We now show that their expressions can be simplified in the case where the observation vector $\mathbf{r}$ consists of N statistically independent components. In many cases, the probability

densities of the observation components are identical, but that is not necessary for the current derivation. For independent observations, we have

$$p_{\mathbf{r}|H_j}(\mathbf{R}|H_j) = \prod_{i=1}^{N} p_{r_i|H_j}(R_i|H_j) \quad j = 0, 1. \tag{2.245}$$

From (2.204),

$$\mu(s) = \ln \int_{-\infty}^{\infty} \left[p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)\right]^s \left[p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\right]^{1-s} d\mathbf{R}. \tag{2.246}$$

Using (2.245) in (2.246), for the ID model, we have

$$\begin{aligned}\mu(s) &= \sum_{i=1}^{N} \ln \int \left[p_{r_i|H_1}(R_i|H_1)\right]^s \left[p_{r_i|H_0}(R_i|H_0)\right]^{1-s} dR_i \\ &= \sum_{i=1}^{N} \mu_i(s),\end{aligned} \tag{2.247}$$

where

$$\mu_i(s) = \ln \int \left[p_{r_i|H_1}(R_i|H_1)\right]^s \left[p_{r_i|H_0}(R_i|H_0)\right]^{1-s} dR_i. \tag{2.248}$$

Taking derivatives with respect to s yields

$$\dot{\mu}(s) = \sum_{i=1}^{N} \dot{\mu}_i(s), \tag{2.249}$$

$$\ddot{\mu}(s) = \sum_{i=1}^{N} \ddot{\mu}_i(s). \tag{2.250}$$

If the components of $\mathbf{r}$ are IID, then $\mu_i(s)$ is the same for all $i = 1, \ldots, N$, and we have

$$\mu(s) = N\mu_i(s), \tag{2.251}$$

$$\dot{\mu}(s) = N\dot{\mu}_i(s), \tag{2.252}$$

$$\ddot{\mu}(s) = N\ddot{\mu}_i(s). \tag{2.253}$$

We now consider several examples to illustrate the application of these ideas. The first is one in which the exact performance is known. We go through the bounds and approximations to illustrate the manipulations involved.

Example 2.12 (continuation of Examples 2.1 and 2.5). In this example, we consider the simple Gaussian problem first introduced in Example 2.1:

$$p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(R_i - m)^2}{2\sigma^2}\right] \tag{2.254}$$

and

$$p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{R_i^2}{2\sigma^2}\right). \tag{2.255}$$

Then, using (2.246)

$$\mu(s) = \ln \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(R_i - m)^2 s + R_i^2(1-s)}{2\sigma^2}\right] dR_i. \tag{2.256}$$

Because the observations are IID, all the integrals are identical, and

$$\mu(s) = N\mu_i(s) = N \ln \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(R-m)^2 s + R^2(1-s)}{2\sigma^2}\right] dR. \tag{2.257}$$

Integrating we have

$$\mu(s) = Ns(s-1)\frac{m^2}{2\sigma^2} = \frac{s(s-1)d^2}{2}, \tag{2.258}$$

where d was defined in (2.82). The curve is shown in Figure 2.32.

Taking the derivative with respect to s gives:

$$\dot{\mu}(s) = \frac{(2s-1)d^2}{2}. \tag{2.259}$$

Evaluating $\ddot{\mu}(s)$, we obtain

$$\ddot{\mu}(s) = d^2. \tag{2.260}$$

Using (2.227), the value of s_* is found from:

$$s_* = \frac{\gamma_*}{d^2} + \frac{1}{2}. \tag{2.261}$$

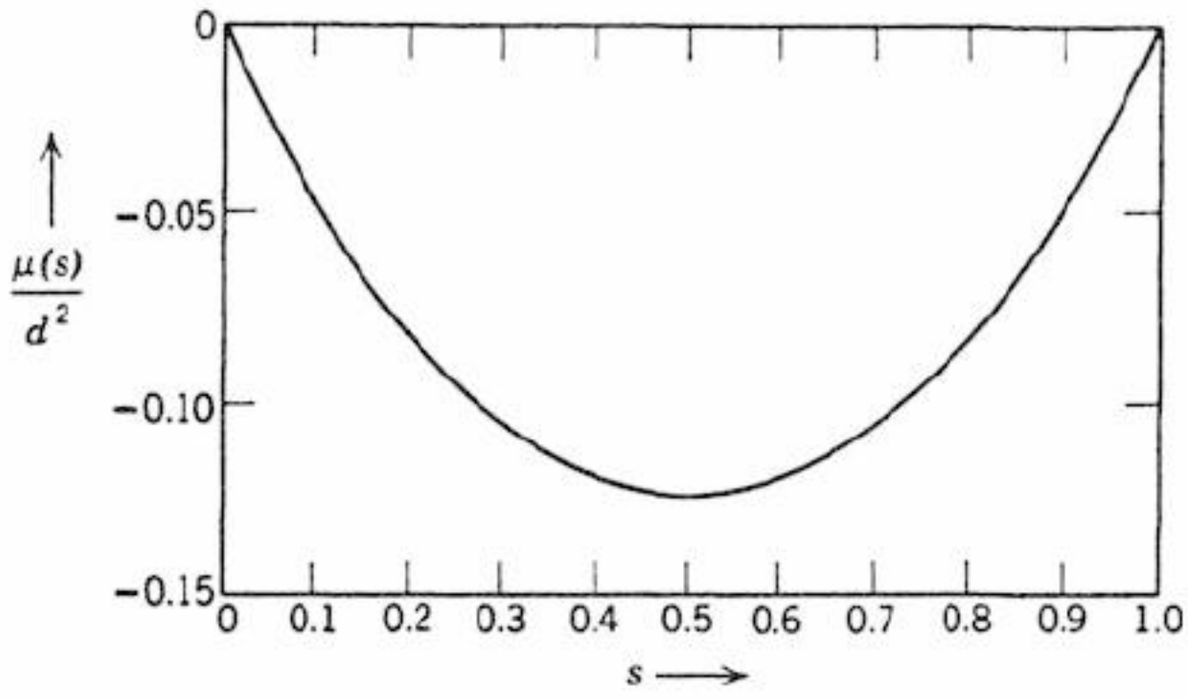


Figure 2.32: $\mu(s)$ for Gaussian variables with unequal means.

Using the bounds in (2.226), we have

$$P_F(\gamma_*) \leqslant \exp\left(\frac{-s_*^2 d^2}{2}\right); \quad s_* \geqslant 0 \Longrightarrow \gamma_* \geqslant -\frac{d^2}{2},$$
$$P_M(\gamma_*) \leqslant \exp\left[-\frac{(1-s_*)^2 d^2}{2}\right]; \quad s_* \leqslant 1 \Longrightarrow \gamma_* \leqslant \frac{d^2}{2}. \tag{2.262}$$

Because $l(\mathbf{R})$ is the sum of Gaussian random variables, the expressions in (2.238) and (2.241) are exact. Substituting (2.258)–(2.261) into (2.238) and (2.241), we have

$$P_F(\gamma_*) = \text{erfc}_*\left[s_*\sqrt{\ddot{\mu}(s_*)}\right] = \text{erfc}_*(s_* d) = \text{erfc}_*\left(\frac{\gamma_*}{d} + \frac{d}{2}\right) \tag{2.263}$$

and

$$P_M(\gamma_*) = \text{erfc}_*\left[(1-s_*)\sqrt{\ddot{\mu}(s_*)}\right] = \text{erfc}_*[(1-s_*)d] = \text{erfc}_*\left(-\frac{\gamma_*}{d} + \frac{d}{2}\right). \tag{2.264}$$

These expressions are identical to (2.84) and (2.85).

An even simpler case is one in which the total probability of error is the criterion. Then, we choose s_m such that $\dot{\mu}(s_m) = 0$. From Figure 2.32, we see that $s_m = \frac{1}{2}$. Using (2.243) and (2.244), we have

$$\Pr(\epsilon) = \text{erfc}_*\left(\frac{d}{2}\right) \approx \left(\frac{2}{\pi d^2}\right)^{1/2} \exp\left(-\frac{d^2}{8}\right). \tag{2.265}$$

The first expression is exact and identical to (2.88) and the second approximate expression is very good for $d > 6$.

This example is a special case of the binary symmetric hypothesis problem in which $\mu(s)$ is symmetric about $\frac{1}{2}$. When this is true and the criterion is minimum $\Pr(\epsilon)$, then $\mu\left(\frac{1}{2}\right)$ is the important quantity,

$$\mu\left(\frac{1}{2}\right) = \ln \int_{-\infty}^{\infty} \left[p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)\right]^{1/2} \left[p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\right]^{1/2} d\mathbf{R}. \tag{2.266}$$

The negative of this quantity is frequently referred to as the Bhattacharyya distance (e.g., [Bha43]). It is important to note that it is the significant quantity only when $s_m = \frac{1}{2}$. ■

The next example also considers a case in which we have exact expressions for $P_D(\gamma_*)$ and $P_F(\gamma_*)$. However, it serves as useful lead into the case where the σ_j^2; $j = 0, 1$ are different on each component (i.e., we have σ_{ij}^2; $j = 0, 1; i = 1, 2, \ldots, N$) that we will encounter in Chapter 3.

Example 2.13 (continuation of Examples 2.2 and 2.6). From (2.27) and (2.28),

$$p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{R_i^2}{2\sigma_1^2}\right),$$
$$p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{R_i^2}{2\sigma_0^2}\right). \tag{2.267}$$

Substituting (2.267) into (2.246) and using the IID property gives,

$$\mu(s) = N\mu_i(s) = N \ln \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1^s \sigma_0^{1-s}} \exp\left[-\frac{sR^2}{2\sigma_1^2} - \frac{(1-s)R^2}{2\sigma_0^2}\right] dR \tag{2.268}$$

or

$$\mu(s) = \frac{N}{2} \ln\left[\frac{(\sigma_0^2)^s(\sigma_1^2)^{1-s}}{s\sigma_0^2 + (1-s)\sigma_1^2}\right]. \tag{2.269}$$

A case that will be of interest in the sequel is

$$\begin{aligned} \sigma_1^2 &= \sigma_n^2 + \sigma_s^2, \\ \sigma_0^2 &= \sigma_n^2. \end{aligned} \tag{2.270}$$

Substituting (2.270) into (2.269) gives

$$\frac{\mu(s)}{N/2} = \left\{(1-s)\ln\left(1 + \frac{\sigma_s^2}{\sigma_n^2}\right) - \ln\left[1 + (1-s)\frac{\sigma_s^2}{\sigma_n^2}\right]\right\}. \tag{2.271}$$

This function is shown in Figure 2.33. Taking derivatives,

$$\dot{\mu}(s) = \frac{N}{2}\left[-\ln\left(1 + \frac{\sigma_s^2}{\sigma_n^2}\right) + \frac{\sigma_s^2/\sigma_n^2}{1 + (1-s)\sigma_s^2/\sigma_n^2}\right] \tag{2.272}$$

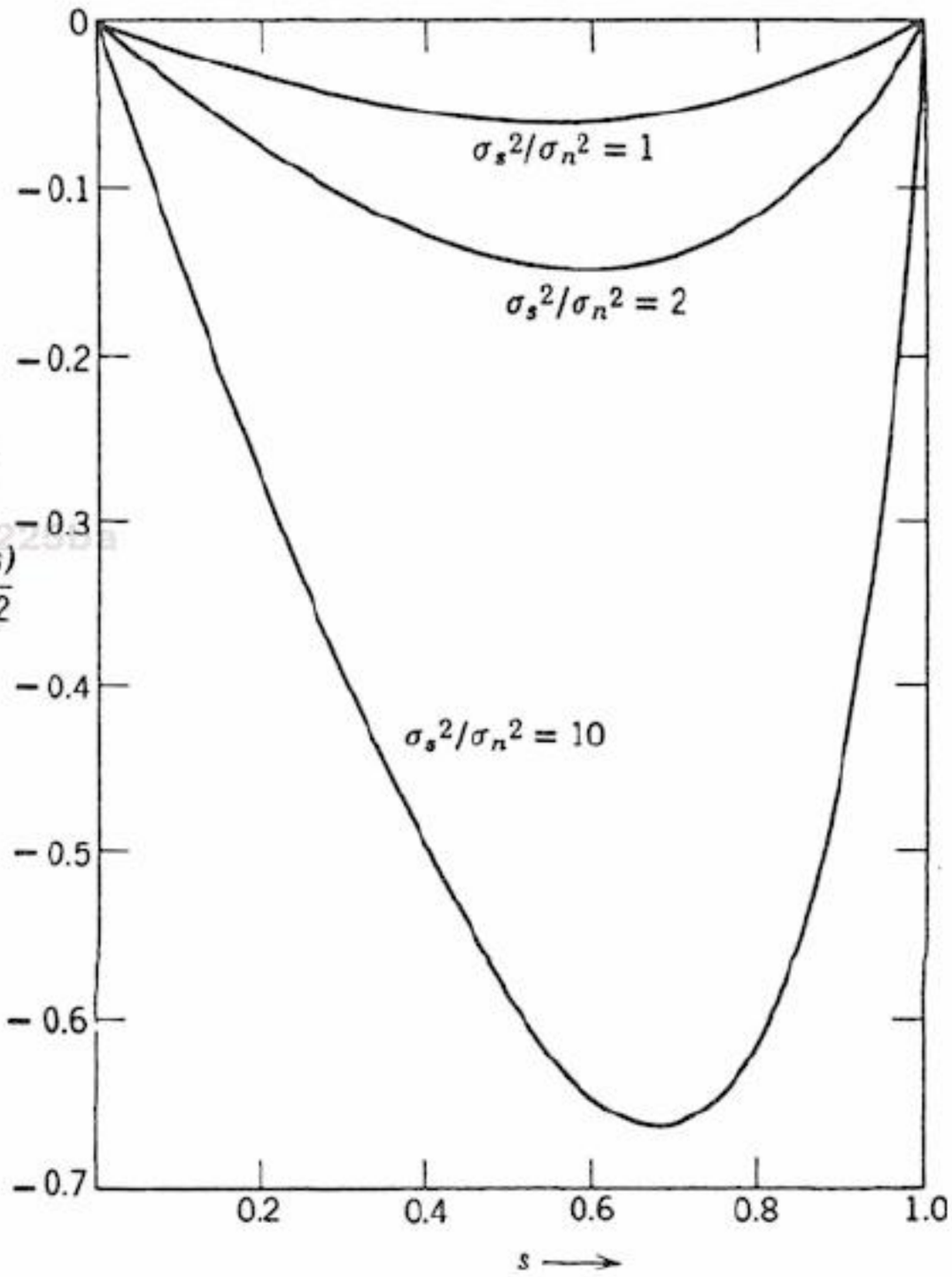


Figure 2.33: $\mu(s)$ for Gaussian variables with unequal variances.

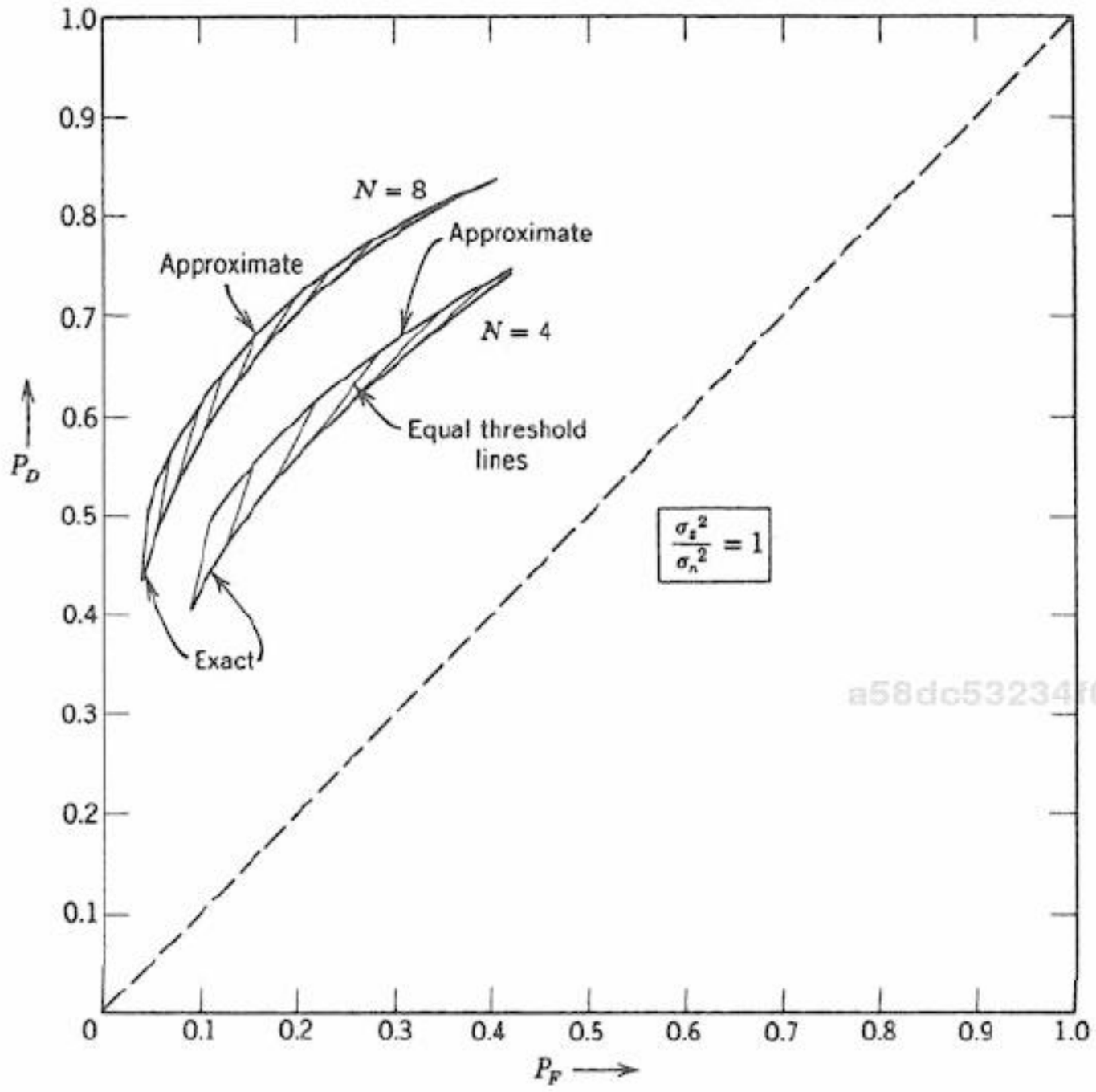


Figure 2.34: Approximate receiver operating characteristics.

and

$$\ddot{\mu}(s) = \frac{N}{2}\left[\frac{\sigma_s^2/\sigma_n^2}{1+(1-s)\sigma_s^2/\sigma_n^2}\right]^2. \tag{2.273}$$

By substituting (2.271), (2.272), and (2.273) into (2.238) and (2.241), we can plot an approximate receiver operating characteristic. This can be compared with the exact ROC in Figure 2.15 to estimate the accuracy of the approximation. In Figure 2.34, we show the comparison for $N = 4$ and 8, and $\sigma_s^2/\sigma_n^2 = 1$. The lines connect the equal threshold points. We see that the approximation is good. For larger N, the exact and approximate ROC are identical for all practical purposes. ■

Examples 2.12 and 2.13 allowed us to compare the P_D and P_F approximations to exact expressions. In the next example analytic solutions are not available for P_D and P_F.

Example 2.14. The observations on the two hypotheses are statistically independent samples from Weibull probability densities with different parameters,

$$p_{r_i|H_j}(R_i|H_j) = \frac{\alpha_j}{b_j}\left(\frac{R_i}{b_j}\right)^{\alpha_j - 1} e^{-(R_i/b_j)^{\alpha_j}}, \quad i = 1, 2, \ldots, N; \; j = 0, 1. \tag{2.274}$$

When we study radar applications later in the text, we will find that the Weibull probability density is a good model for clutter in many applications. The parameter $\alpha_j > 0$ controls the shape and the parameter $b_j > 0$ controls the scale. The mean is

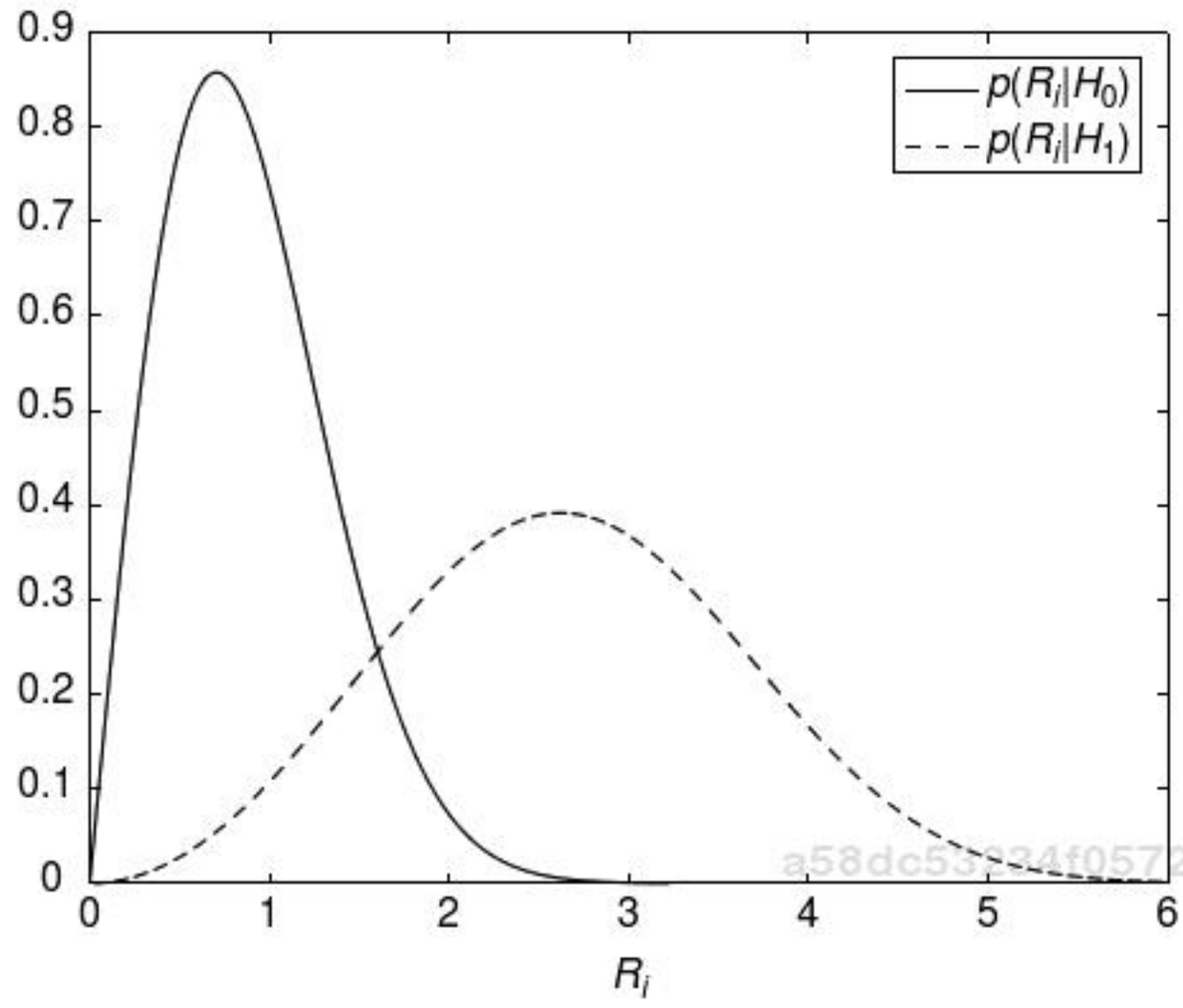


Figure 2.35: Weibull densities on H_0 and H_1, $\alpha_0 = 2$, $b_0 = 1$, $\alpha_1 = 3$, $b_1 = 3$.

$$E(r_i|H_j) = b_j\,\Gamma\left(1+\frac{1}{\alpha_j}\right) \tag{2.275}$$

and the variance is

$$\mathrm{Var}(r_i|H_j) = b_j^2\left[\Gamma\left(1+\frac{2}{\alpha_j}\right) - \Gamma^2\left(1+\frac{1}{\alpha_j}\right)\right]. \tag{2.276}$$

We consider the following specific values: $\alpha_0 = 2, b_0 = 1, \alpha_1 = 3, b_1 = 3$. The resulting probability densities on the two hypotheses are shown in Figure 2.35.

The functions $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$ are computed numerically. We use (2.274) to obtain $\mu_i(s)$ and then differentiate numerically to obtain $\dot{\mu}_i(s)$, and $\ddot{\mu}_i(s)$. The results are shown in Figure 2.36. We obtain $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$ from (2.251)–(2.253) by multiplying by N. In order to construct an approximate ROC using (2.238) and (2.241), we let s_* vary from 0 to 1. The $\dot{\mu}_i(s)$ curve in Figure 2.36 specifies the threshold $\gamma_* = N\dot{\mu}_i(s_*)$. The results are shown in Figure 2.37. As expected, the performance improves as N increases. In the problems, we will investigate the behavior for various α_j and b_j. We do not have an analytic result to verify the approximation, but in Section 2.5 we will discuss techniques for simulating the LRT to validate the results. ■

Summary. The principal results of this section were the bounds on P_F, P_M, and $\Pr(\epsilon)$ given in (2.226) and (2.230), and the approximate error expressions given in (2.238), (2.239), (2.241), (2.242), (2.243), and (2.244). These expressions will enable us to find performance results for a number of cases of physical interest.

The first two examples considered Gaussian densities on both hypotheses and analytic results were available. The results play a much more important role when we have statistically independent observations and the probability densities are non-Gaussian. Then,

$$\mu(s) = \sum_{i=1}^{N} \mu_i(s)$$

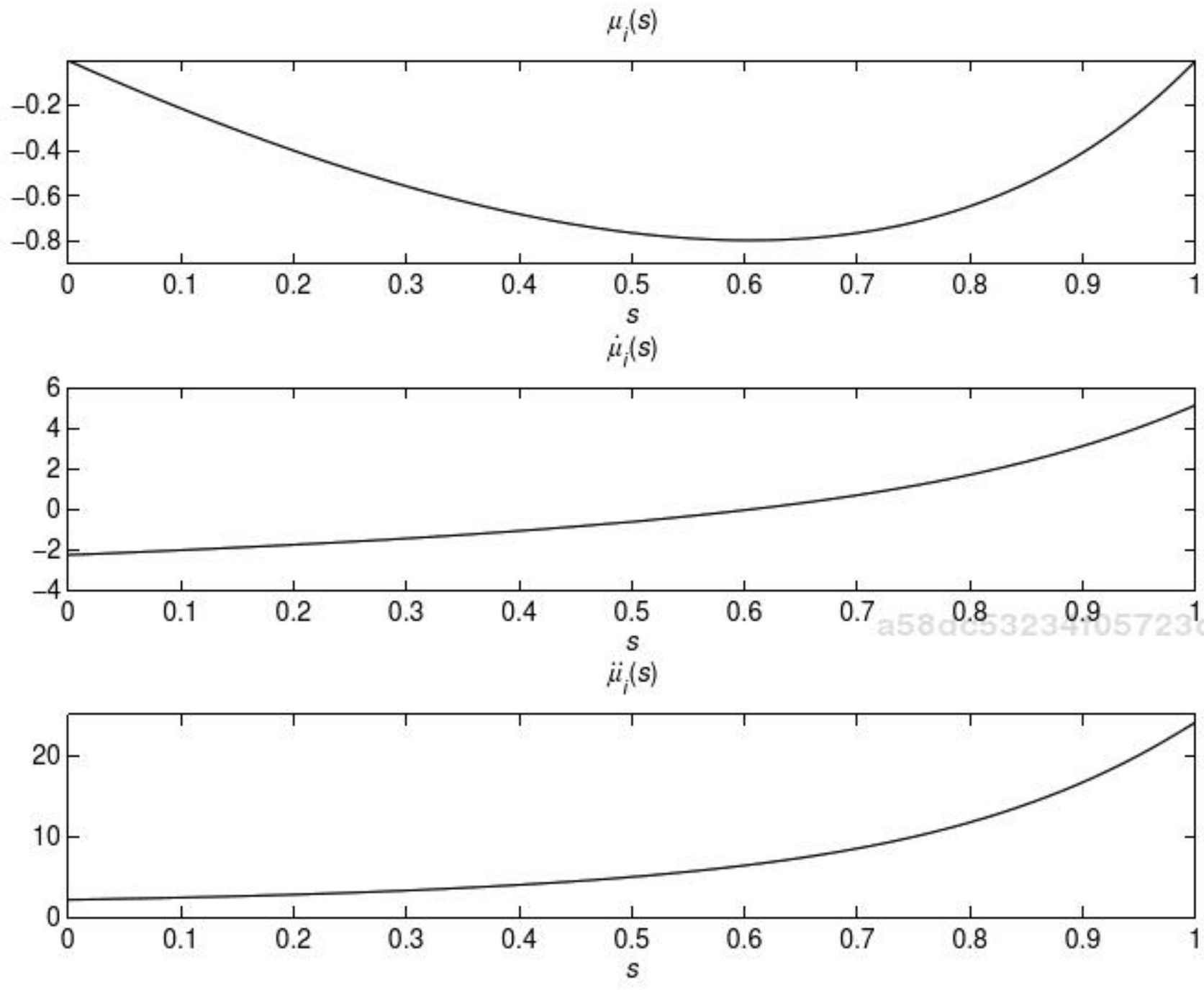


Figure 2.36: Weibull densities; $\mu_i(s)$, $\dot{\mu}_i(s)$, and $\ddot{\mu}_i(s)$.

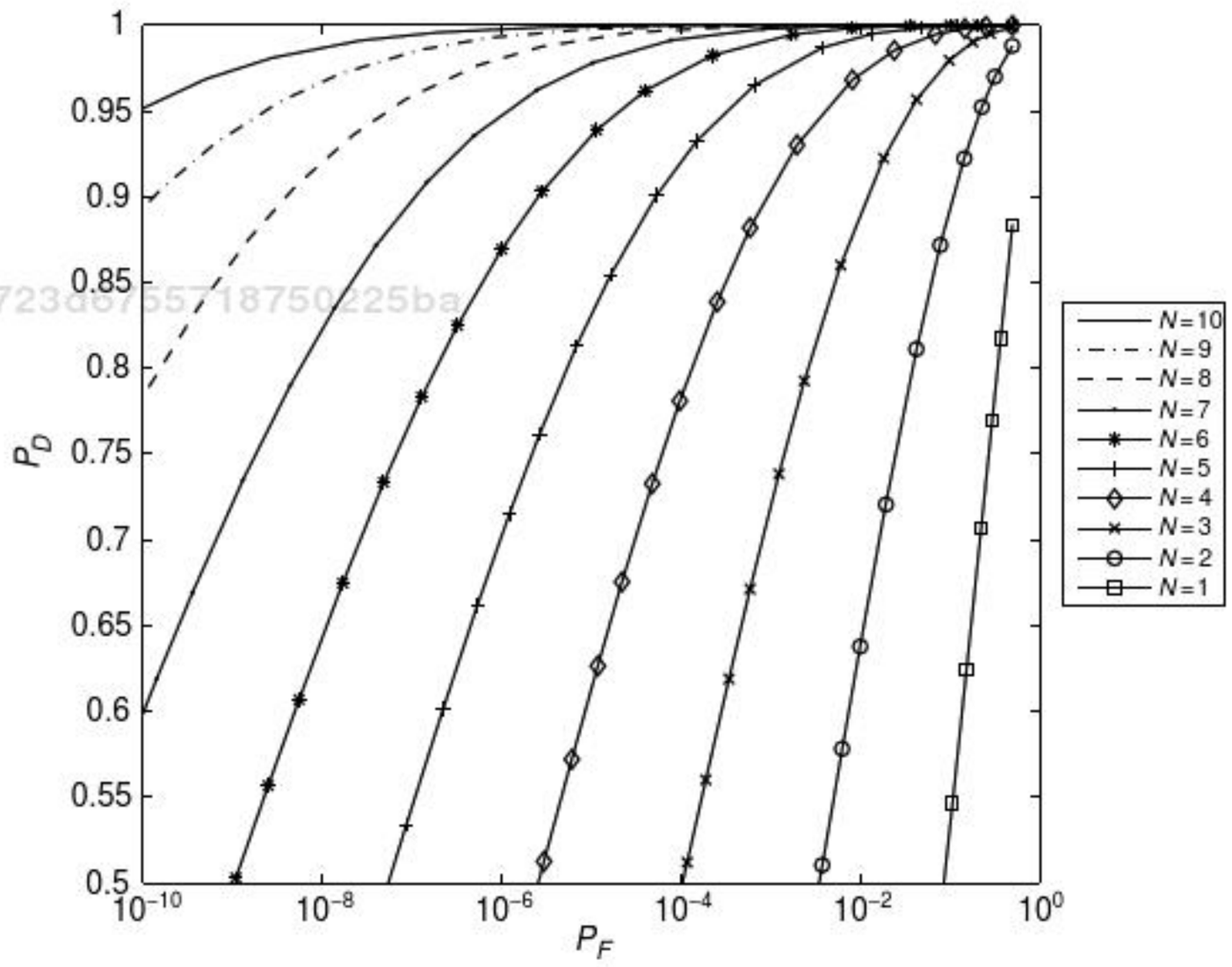


Figure 2.37: Receiver operating characteristic; Weibull densities, $N = 1, 2, \ldots, 10$.

and we can find $\mu_i(s)$ by a numerical one-dimensional integration of (2.248) if an analytic result is not available. We can then construct an approximate ROC as in Example 2.14. Thus, for an ID, non-Gaussian model, the results in Section 2.2 specify the optimum test and the results in this section allow us to bound the performance and construct approximate ROCs.

In the next section, we will discuss Monte Carlo simulation techniques and will find that the $\mu(s)$ function plays the central role in designing optimum simulations.

2.5 MONTE CARLO SIMULATION

In many of the applications of interest, it is necessary to simulate the detection algorithm in order to evaluate the performance. In Section 2.5.1, we give a brief introduction to Monte Carlo (MC) simulation. A key issue is the number of trials needed to have a desired level of confidence in the result. In most systems of interest, the desired P_F is very small (e.g., $P_F \leqslant 10^{-6}$ is frequently required). In these cases, the number of trials required to obtain a reasonable confidence level is prohibitively large.

In Section 2.4, we introduced the Chernoff bound and various extensions in order to obtain bounds on the performance and approximate expressions for the performance. The key idea was that, by defining a $\mu(s)$ function, we could tilt the relevant probability densities so that the mass of the tilted density was near the threshold. We can apply the same idea to the simulation problem. The resulting technique is called "importance sampling" in the literature. We develop the key results in Section 2.5.2 and apply them to several of the examples introduced earlier in the chapter.

In Section 2.5.3, we summarize our results.

2.5.1 Monte Carlo Simulation Techniques

The log-likelihood ratio test consists of comparing the log-likelihood ratio to a threshold. From (2.195),

$$l_*(\mathbf{R}) \triangleq \Lambda(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_*. \tag{2.277}$$

From (2.75), the log-likelihood ratio and the corresponding error probabilities can equivalently be expressed in terms of any sufficient statistic $\mathbf{x}$,

$$l_*(\mathbf{X}) \triangleq \ln \frac{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}. \tag{2.278}$$

The expressions for $P_F(\gamma_*)$ and $P_M(\gamma_*)$ in terms of $\mathbf{x}$ are

$$P_F(\gamma_*) = \Pr\left(l_*(\mathbf{X}) \geqslant \gamma_* | H_0\right) = \int_{l_*(\mathbf{X}) \geqslant \gamma_*} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\, d\mathbf{X} \tag{2.279}$$

and

$$P_M(\gamma_*) = 1 - P_D(\gamma_*) = \Pr\left(l_*(\mathbf{X}) < \gamma_* | H_1\right) = \int\limits_{l_*(\mathbf{X})<\gamma_*} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\, d\mathbf{X}. \tag{2.280}$$

At one extreme, the sufficient statistic may be the scalar log-likelihood ratio itself or a scalar function of the observations that is related to the log-likelihood ratio in a straightforward (invertible) manner. At the other extreme, it may be the N-dimensional observation vector $\mathbf{r}$. In other cases, $\mathbf{x}$ may be multidimensional with dimension less than N. The latter two cases are important when it is difficult to find $p_{l_*|H_0}(L|H_0)$ and $p_{l_*|H_1}(L|H_1)$. As in the previous sections, we will find it useful to work with a variety of sufficient statistics.

For this discussion, we will make use of the indicator function $\mathbb{I}(\cdot)$, which is equal to one when its argument is true and zero otherwise. The indicator function allows us to express probabilities as expected values. For example,

$$\mathbb{I}(l_*(\mathbf{X}) \geqslant \gamma_*) = \begin{cases} 1 & l_*(\mathbf{X}) \geqslant \gamma_* \\ 0 & l_*(\mathbf{X}) < \gamma_*. \end{cases} \tag{2.281}$$

With this definition, $P_F(\gamma_*)$ in (2.279) may be written as

$$P_F(\gamma_*) = \int \mathbb{I}(l_*(\mathbf{X}) \geqslant \gamma_*)\, p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\, d\mathbf{X} = E_0\left[\mathbb{I}(l_*(\mathbf{X}) \geqslant \gamma_*)\right], \tag{2.282}$$

where $E_0[\cdot]$ denotes expectation on H_0. Similarly,

$$P_M(\gamma_*) = E_1\left[\mathbb{I}(l_*(\mathbf{X}) < \gamma_*)\right], \tag{2.283}$$

where $E_1[\cdot]$ denotes expectation on H_1.

The procedure for estimating P_F is to simulate a random sample of the sufficient statistic $\mathbf{x}$ and estimate the expected value in (2.282) by computing the sample mean of the indicator function. To generate the random sample, we conduct a set of K_F independent trials. On each trial, we generate a realization of $\mathbf{x}$ from the probability density $p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)$. Our choice of sufficient statistic $\mathbf{x}$ will often by driven by the ease in which we can generate random samples from the specified distribution. There are a variety of techniques for generating random samples that are discussed in [DeV86] and [BFS87].

If the sufficient statistic is the scalar x and its cumulative distribution function (CDF) $P_x(X)$ is known, then we can use the inversion method. In this method, we generate a sample of a random variable y with a uniform probability density

$$p_y(Y) = \begin{cases} 1 & 0 \leqslant Y \leqslant 1 \\ 0 & \text{elsewhere} \end{cases} \tag{2.284}$$

and transform Y to X by the inverse of the CDF,

$$X = P_x^{-1}(Y). \tag{2.285}$$

Then,

$$Y = P_x(X). \tag{2.286}$$

To verify that X is a sample from $p_x(X)$, and the Jacobian of the transformation in (2.285) is

$$J = \frac{dY}{dX} = \frac{dP_x(X)}{dX} = p_x(X) \tag{2.287}$$

and

$$p_x(X) = p_y\big(P_x(X)\big)|J| = 1 \cdot p_x(X). \tag{2.288}$$

Thus, we can generate $U(0, 1)$ random samples and transform them with $P_x^{-1}(Y)$.

If the inverse of the CDF is not known explicitly, then we may use other methods such as numerical inversion or the rejection method [DeV86, BFS87].[6]

We denote the simulated sufficient statistic generated on the kth trial as $\mathbf{X}_k$ and the value of the log-likelihood ratio as $l_*(\mathbf{X}_k; H_0)$. On each trial, we compute $l_*(\mathbf{X}_k; H_0)$ and count the number of times that it exceeds the threshold γ_*. We denote this count by the random variable n_{γ_*}, which is defined as[7]

$$n_{\gamma_*} = \sum_{k=1}^{K_F} \mathbb{I}\left(l_*(\mathbf{X}_k; H_0) \geqslant \gamma_*\right). \tag{2.289}$$

It is a Binomial(K_F, P_F) random variable with probability mass function,

$$\Pr(n_{\gamma_*} = n) = \binom{K_F}{n} P_F^n (1 - P_F)^{K_F - n} \quad n = 0, 1, \ldots, K_F. \tag{2.290}$$

Given n_{γ_*}, we estimate $P_F(\gamma_*)$ as[8]

$$\hat{P}_F(\gamma_*) = \frac{n_{\gamma_*}}{K_F} = \frac{1}{K_F} \sum_{k=1}^{K_F} \mathbb{I}\left(l_*(\mathbf{X}_k; H_0) \geqslant \gamma_*\right). \tag{2.291}$$

The expectation of the estimate is

$$E\left[\hat{P}_F\right] = \frac{1}{K_F} E\left[n_{\gamma_*}\right] = \frac{K_F P_F}{K_F} = P_F, \tag{2.292}$$

which is an unbiased estimate. The variance of the estimate is

$$\mathrm{Var}(\hat{P}_F) = \frac{1}{K_F^2} \mathrm{Var}(n_{\gamma_*}) = \frac{K_F P_F (1 - P_F)}{K_F^2} = \frac{P_F(1 - P_F)}{K_F}. \tag{2.293}$$

The variance of the estimate decreases as the number of trials increases and we would like to determine how many trials we need to get a good estimate. We do this in terms of a confidence interval $\hat{P}_F \epsilon [(1 - \alpha)P_F, (1 + \alpha)P_F]$, where

[6]The Matlab Statistics Toolbox has random number generators for many standard distributions, including all of the distributions listed in Appendix A except for the Generalized Gaussian. The inversion or rejection methods could be used for that case.

[7]Our development is similar to a number of references (e.g., [Ech91] or [Sri02]).

[8]In Chapter 4, when we study estimation theory, we will see that $\hat{P}_F$ is the maximum likelihood estimate of P_F.

$$\Pr\left(\frac{|\hat{P}_F - P_F|}{P_F} \leqslant \alpha\right) = p_c, \tag{2.294}$$

where p_c is the confidence probability (e.g., 0.95) and α is the percent deviation tolerance (e.g., 0.1). Then, (2.294) can be written as

$$\Pr\left(|\hat{P}_F - P_F| \leqslant \alpha P_F\right) = p_c. \tag{2.295}$$

For large K_F, the probability density of $\hat{P}_F$ approaches a Gaussian density with mean and variance given by (2.292) and (2.293), that is,

$$\hat{P}_F \sim N\left(P_F, \frac{P_F(1-P_F)}{K_F}\right). \tag{2.296}$$

For a Gaussian random variable $z \sim N(\mu_z, \sigma_z^2)$, the probability that z has a value within c standard deviations of its mean is

$$\Pr\left(|z - \mu_z| \leqslant c\,\sigma_z\right) = p_c, \tag{2.297}$$

where c is given in Table 2.1 for several values of p_c. Using the Gaussian approximation for $\hat{P}_F$ from (2.296) in (2.297), we have

$$\Pr\left(|\hat{P}_F - P_F| \leqslant c\sqrt{\frac{P_F(1-P_F)}{K_F}}\right) = p_c. \tag{2.298}$$

Comparing this expression to (2.295), we can achieve the desired confidence level if

$$\alpha P_F = c\sqrt{\frac{P_F(1-P_F)}{K_F}}. \tag{2.299}$$

This occurs when

$$K_F = \frac{c^2}{\alpha^2}\frac{1-P_F}{P_F}. \tag{2.300}$$

The preceding analysis is valid when P_F is small (less than 0.5). In some cases, we may want to simulate a scenario where $P_F > 0.5$. In this case, we want the tolerance to be a fraction of $1 - P_F$. To cover both cases, we express the tolerance as a fraction of the smaller of P_F and $1 - P_F$, that is,

$$\Pr\left(|\hat{P}_F - P_F| \leqslant \alpha \min\left(P_F, 1-P_F\right)\right) = p_c. \tag{2.301}$$

Table 2.1: Confidence interval probabilities from the Gaussian distribution

p_c	0.900	0.950	0.954	0.990	0.997
c	1.645	1.960	2.000	2.576	3.000

Then, (2.299) becomes

$$\alpha \min(P_F, 1 - P_F) = c\sqrt{\frac{P_F(1 - P_F)}{K_F}}. \tag{2.302}$$

and the more general formula for the required number of trials is:

$$K_F = \frac{c^2}{\alpha^2} \frac{P_F(1 - P_F)}{\min(P_F, 1 - P_F)^2}. \tag{2.303}$$

For example, if $p_c = 0.954$, then $c = 2$. When $\alpha = 0.1$ and P_F is small, we have

$$K_F \approx \frac{4}{0.01} \frac{1}{P_F} = \frac{400}{P_F}. \tag{2.304}$$

Since false alarms occur with probability P_F, we would expect to observe one false alarm in about every $1/P_F$ trials. The result above tells us that we need to observe at least 400 false alarms to get a good estimate of P_F, which requires at least $400/P_F$ trials. If we want a higher confidence probability or lower tolerance factor, even more trials would be required.

To simulate an entire ROC curve, we need to estimate both P_F and P_D (or P_M) at various thresholds. To estimate P_M, we conduct a set of K_M independent trials in which we generate a realization of $\boldsymbol{x}$ from the probability density $p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)$ and count the number of times that $l_*(\mathbf{X}_k; H_1)$ falls below the threshold γ_*. Following a similar argument, the required number of trials is:

$$K_M = \frac{c^2}{\alpha^2} \frac{P_M(1 - P_M)}{\min(P_M, 1 - P_M)^2}, \tag{2.305}$$

and the estimate of $P_M(\gamma_*)$ is:

$$\hat{P}_M(\gamma_*) = \frac{1}{K_M} \sum_{k=1}^{K_M} \mathbb{I}(l_*(\mathbf{X}_k; H_1) < \gamma_*). \tag{2.306}$$

We consider the following example to demonstrate the results. It is a continuation of Examples 2.1 and 2.5. Because we have analytic results, we probably would not simulate this model in practice, but it is useful to introduce the simulation procedure.

Example 2.15 (continuation of Examples 2.1, 2.5, and 2.12). From (2.19) and (2.20),

$$\begin{aligned} H_1 &: r_i \sim N(m, \sigma^2) \quad i = 1, 2, \ldots N \\ H_0 &: r_i \sim N(0, \sigma^2) \quad i = 1, 2, \ldots N, \end{aligned} \tag{2.307}$$

and from (2.84) and (2.85),

$$P_F(\gamma_*) = \text{erfc}_*\left(\frac{\gamma_*}{d} + \frac{d}{2}\right), \tag{2.308}$$

$$P_M(\gamma_*) = 1 - \text{erfc}_*\left(\frac{\gamma_*}{d} - \frac{d}{2}\right), \tag{2.309}$$

where

$$d = \frac{\sqrt{N}m}{\sigma}. \tag{2.310}$$

Table 2.2: Simulation values for $d = \sqrt{40}$, $p_c = 0.954$, and $\alpha = 0.1$

P_F	Z_F	γ_*	P_M	K_F	K_M
10^{-2}	2.33	−5.29	3.2×10^{-5}	4×10^4	1.3×10^7
10^{-4}	3.72	3.52	0.0046	4×10^6	8.7×10^4
10^{-6}	4.75	10.06	0.0581	4×10^8	6488
10^{-8}	5.61	15.49	0.2381	4×10^{10}	1281
10^{-10}	6.36	20.23	0.5147	4×10^{12}	425
10^{-12}	7.03	24.49	0.7611	4×10^{14}	1275

We define Z_F as the argument of $\text{erfc}_*(\cdot)$ that corresponds to the value of P_F that we want to simulate.[9] We rewrite (2.308) as

$$P_F = \text{erfc}_*(Z_F). \tag{2.311}$$

Therefore,

$$Z_F = \frac{\gamma_*}{d} + \frac{d}{2} \tag{2.312}$$

and we calculate the threshold from

$$\gamma_* = Z_F d - \frac{d^2}{2} \tag{2.313}$$

and then calculate $P_M(\gamma_*)$ from (2.309). The required number of trials is obtained from (2.303) and (2.305). A summary is given in Table 2.2 for $d = \sqrt{40}$, $p_c = 0.954$ ($c = 2$), $\alpha = 0.1$, and a variety of ROC points. We observe that as P_F decreases, K_F increases correspondingly. At the same time, P_M is increasing and K_M is decreasing until P_M becomes greater than 0.5. Then $1 - P_M < P_M$, and K_M begins increasing.

To implement the simulation, we need to pick a sufficient statistic to simulate. We could choose the statistic to be the original observations, $\mathbf{x} = \mathbf{r}$. Then for each trial, we would generate N IID observations from the densities in (2.307). Alternatively, we could choose the statistic to be the log-likelihood ratio, $\mathbf{x} = l_*$, which is given in (2.24). Then, from (2.307) and (2.310),

$$H_1 : l_* \sim N\left(\frac{d}{2}, d^2\right) \tag{2.314}$$

$$H_0 : l_* \sim N\left(-\frac{d}{2}, d^2\right), \tag{2.315}$$

and for each trial, we would generate a scalar random variable from these densities.

To simulate the ROC curves in Figure 2.12*b*, for each point on the curve, we would determine the threshold γ_* from P_F and d, and calculate the required K_F and K_M. We would generate K_F trials from $p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)$ and compute $\hat{P}_F(\gamma_*)$ using (2.291), then generate K_M trials from $p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)$ and compute $\hat{P}_M(\gamma_*)$ using (2.306).

To simulate the curves in 2.12*a*, we could follow the same procedure. However, in this case only the threshold varies along each curve while the densities remain the same, and we can use a simpler procedure. We first determine the maximum number of trials needed over all points on the curve and denote these by $\overline{K}_F$ and $\overline{K}_M$. We then generate $\overline{K}_F$ trials from $p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)$ and compute $\hat{P}_F(\gamma_*)$ using (2.291) by varying γ_*, and do the same using $\overline{K}_M$ trials from $p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)$ to calculate $\hat{P}_M(\gamma_*)$. This

[9]For a specified P_F, Z_F can be computed in Matlab using the `norminv` function. See Appendix A.

method has the advantage that only one set of data is generated. The worst-case accuracy will be as specified by α and p_c; however, many of the points will have significantly better accuracy since the number of trials generated will be larger than required. ■

In the previous example, the simulation method was straightforward, however the number of trials was very large for the small values of P_F considered. Clearly, we need to find a better technique in order to reduce the number of trials required to achieve the desired accuracy and confidence level.

2.5.2 Importance Sampling

Our development of importance sampling is based on our discussion of tilted densities, $\mu(s)$, and the Chernoff bound in Section 2.4.

We introduced a new random variable x_s whose probability density was related to $p_{l|H_0}(L|H_0)$ by (2.205)

$$p_{x_s}(X) = e^{sX-\mu(s)}\, p_{l|H_0}(X|H_0), \tag{2.316}$$

where $s \geqslant 0$. In Figure 2.29, we saw that as s increased, more of the density was moved to the right of the threshold. The tightest bound on P_F was obtained by choosing s so that the mean of the tilted variable x_s was at the threshold.

In Section 2.5.1, we found that a key issue in MC simulation was that if P_F (or P_M) was very small, then the number of trials was prohibitively large. However, if we could run the simulation using the tilted density, then the probability of exceeding the threshold would be on the order of 0.5. (It would be exactly 0.5 if the median and mean were equal.) If we could relate this probability to the desired P_F (or P_M), then presumably the required number of trials would be significantly smaller.

There is an extensive literature on tilted densities and importance sampling. The earliest use of tilted densities appears in a paper by Esscher [Ess32]. His results are still widely used in the financial community and are referred to as the Esscher transformation (e.g., [The84]). Siegmund [Sie76] applied the technique to sequential detection. The first application to communications was by Cottrell et al. [CFM83]. The technique started to be referred to as *large deviation theory* and a number of applications, books, and journal articles began to appear.

Importance sampling was discussed in the early work of Kahn and Marshall [KM53] and the book by Hammersley and Handscomb [HH64]. The early work developed various techniques to modify the probability density for simulation purposes but did not utilize tilted densities. Papers by Sadowsky and Bucklew [SB90] and Sadowsky [Sad93] showed the optimality of tilted densities and the technique became more widely used.

The paper by Smith et al. [SSG97] provides an excellent review of the history and status of the area (circa 1997). It contains an extensive list of references. The book by Srinivasan [Sri02] provides a good development of importance sampling at a mathematical level similar to our discussion. Other books include Bucklew [Buc90] and Jeruchim et al. [JBS92].

Our objective in this section is to introduce the reader to the application of tilted densities (large deviation theory) to a specific set of problems:

1. We restrict our attention to the problem of estimating $\Pr(l_*(\mathbf{X}) \geqslant \gamma_*)$ and $\Pr(l_*(\mathbf{X}) < \gamma_*)$ by simulation, where $l_*(\mathbf{X})$ is the log-likelihood ratio in (2.278). This enables us to find $\hat{P}_F(\gamma_*)$ and $\hat{P}_M(\gamma_*)$.
2. We restrict our development to finding the "optimum" tilted density to use in the simulation. We do not discuss other importance sampling techniques that involve other types of biasing densities that may be simpler to implement (but may not perform as well).
3. We assume that the components of $\mathbf{r}$ are statistically independent, but not necessarily IID.

The approach should provide the necessary background so that students and practicing engineers can effectively simulate the algorithms that we develop in the text and problems.

The initial part of our discussion follows Srinivasan's book on importance sampling [Sri02]. However, after establishing some preliminary results we can proceed directly to the tilted densities of Section 2.4.

2.5.2.1 Simulation of P_F

First, consider P_F. We introduce a biasing probability density $p_{\mathbf{x}|0^*}(\mathbf{X})$[10] that is related to $p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)$ in a manner yet to be determined, and write $P_F(\gamma_*)$ as:

$$\begin{aligned} P_F(\gamma_*) &= \int \mathbb{I}\left(l_*(\mathbf{X}) \geqslant \gamma_*\right) p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\, d\mathbf{X} \\ &= \int \mathbb{I}\left(l_*(\mathbf{X}) \geqslant \gamma_*\right) \frac{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}{p_{\mathbf{x}|0^*}(\mathbf{X})}\, p_{\mathbf{x}|0^*}(\mathbf{X})\, d\mathbf{X}. \end{aligned} \tag{2.317}$$

Defining,

$$W_0(\mathbf{X}) \triangleq \frac{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}{p_{\mathbf{x}|0^*}(\mathbf{X})}, \tag{2.318}$$

we can write

$$\begin{aligned} P_F(\gamma_*) &= \int \mathbb{I}\left(l_*(\mathbf{X}) \geqslant \gamma_*\right) W_0(\mathbf{X})\, p_{\mathbf{x}|0^*}(\mathbf{X})\, d\mathbf{X} \\ &= E_{0^*}\left[\mathbb{I}\left(l_*(\mathbf{X}) \geqslant \gamma_*\right) W_0(\mathbf{X})\right]. \end{aligned} \tag{2.319}$$

Note that in general, that for any function $f(\mathbf{X})$,

$$E_0\left[f(\mathbf{X})\right] = E_{0^*}\left[f(\mathbf{X}) W_0(\mathbf{X})\right]. \tag{2.320}$$

We will simulate the test, choosing $\mathbf{X}_k$ from the biasing density $p_{\mathbf{x}|0^*}(\mathbf{X})$. Then,

$$\hat{P}_F(\gamma_*) = \frac{1}{K_F} \sum_{k=1}^{K_F} \mathbb{I}\left(l_*(\mathbf{X}_k; H_{0^*}) \geqslant \gamma_*\right) W_0(\mathbf{X}_k; H_{0^*}). \tag{2.321}$$

[10] We used $p_{x_r}(X)$ in Section 2.4. We need different notation because $\mathbf{X}$ may be a vector and, even if it is a scalar, the biasing density may be different.

The expectation of the estimate is

$$E_{0^*}\left[\hat{P}_F\right] = \frac{K_F}{K_F} E_{0^*}\left[\mathbb{I}\left(l_*(\mathbf{x}) \geqslant \gamma_*\right) W_0(\mathbf{x})\right] = E_0\left[\mathbb{I}\left(l_*(\mathbf{x}) \geqslant \gamma_*\right)\right] = P_F, \tag{2.322}$$

and the variance is

$$\mathrm{Var}_{0^*}(\hat{P}_F) = \frac{K_F}{K_F^2} \mathrm{Var}_{0^*}\left[\mathbb{I}\left(l_*(\mathbf{x}) \geqslant \gamma_*\right) W_0(\mathbf{x})\right] = \frac{1}{K_F}\left(I_F - P_F^2\right), \tag{2.323}$$

where

$$I_F(\gamma_*) \triangleq E_{0^*}\left[\mathbb{I}^2\left(l_*(\mathbf{x}) \geqslant \gamma_*\right) W_0^2(\mathbf{x})\right]. \tag{2.324}$$

Noting that $\mathbb{I}^2(\cdot) = \mathbb{I}(\cdot)$, and using the property in (2.320), we may also write $I_F(\gamma_*)$ as:

$$I_F(\gamma_*) = E_0\left[\mathbb{I}\left(l_*(\mathbf{x}) \geqslant \gamma_*\right) W_0(\mathbf{x})\right]. \tag{2.325}$$

Comparing (2.323) to (2.293), if $I_F < P_F$, then the variance of the importance sampling estimate is less than the variance of the conventional estimate for the same number of trials. Therefore, the number of trials can be reduced to get the same level of accuracy in the estimate. For the importance sampling estimate, (2.302) becomes

$$\alpha \min\left(P_F, 1 - P_F\right) = c\sqrt{\frac{\left(I_F - P_F^2\right)}{K_F}}, \tag{2.326}$$

and the required number of trials is:

$$K_{F,IS} = \frac{c^2}{\alpha^2} \frac{\left(I_F - P_F^2\right)}{\min\left(P_F, 1 - P_F\right)^2}. \tag{2.327}$$

We would like to find the biasing probability density $p_{\mathbf{x}|0^*}(\mathbf{X})$ that minimizes I_F. This is difficult, so instead we find an upper bound on I_F similar to the Chernoff bound and find the biasing density that minimizes the bound. We observe that for $s \geqslant 0$,

$$\mathbb{I}\left(l_*(\mathbf{X}) \geqslant \gamma_*\right) \leqslant e^{s\left[l_*(\mathbf{X}) - \gamma_*\right]} \quad s \geqslant 0. \tag{2.328}$$

This is illustrated in Figure 2.38.

Using (2.328) in (2.324),

$$I_F(\gamma_*) \leqslant E_{0^*}\left[e^{2s\left[l_*(\mathbf{x}) - \gamma_*\right]} W_0^2(\mathbf{x})\right] \triangleq \overline{I}_F(\gamma_*). \tag{2.329}$$

We now apply Jensen's inequality, which states that for any nonnegative function $f(\mathbf{X})$,

$$E\left[f^2(\mathbf{X})\right] \geqslant E^2\left[f(\mathbf{X})\right], \tag{2.330}$$

with equality *iff* $f(\mathbf{X}) = E\left[f(\mathbf{X})\right]$. For the quantities in (2.329), $\overline{I}_F(\gamma_*)$ will be minimized *iff*

$$\begin{aligned} e^{s\left[l_*(\mathbf{X}) - \gamma_*\right]} W_0(\mathbf{X}) &= E_{0^*}\left[e^{s\left[l_*(\mathbf{X}) - \gamma_*\right]} W_0(\mathbf{X})\right] \\ &= E_0\left[e^{s\left[l_*(\mathbf{X}) - \gamma_*\right]}\right] \\ &= e^{\mu(s) - s\gamma_*}, \end{aligned} \tag{2.331}$$

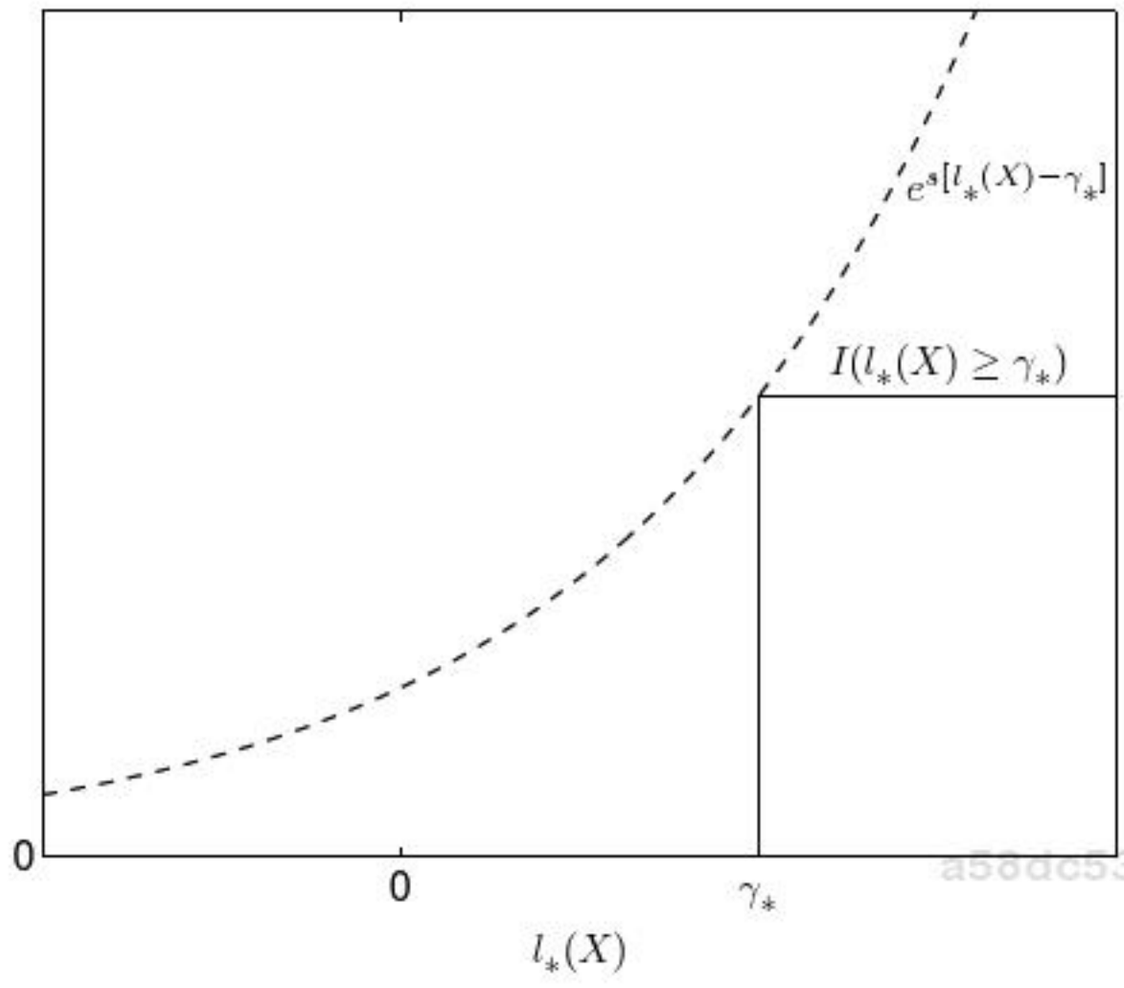


Figure 2.38: Upper bound on $\mathbb{I}(l_*(\mathbf{X}) \geqslant \gamma_*)$.

where $\mu(s)$ was defined in (2.200) in Section 2.4, and can be expressed in terms of any sufficient statistic as:

$$\mu(s) \triangleq \ln E_0\left[e^{sl_*(\mathbf{X})}\right] = \ln \int \left[p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\right]^s \left[p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\right]^{1-s} d\mathbf{X}. \tag{2.332}$$

Using (2.318) in (2.331) gives

$$e^{s[l_*(\mathbf{X})-\gamma_*]} \frac{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}{p_{\mathbf{x}|0^*}(\mathbf{X})} = e^{\mu(s)-s\gamma_*} \tag{2.333}$$

or

$$p_{\mathbf{x}|0^*}(\mathbf{X}; s) = e^{sl_*(\mathbf{X})-\mu(s)} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0). \tag{2.334}$$

Not surprisingly, the biasing density in (2.334) is the same as the tilted density in (2.205).

Substituting (2.334) into (2.318) gives

$$\boxed{W_0(\mathbf{X}; s) = e^{-sl_*(\mathbf{X})+\mu(s)}.} \tag{2.335}$$

These results are essentially the same as the results in [Sri02 equations (2.18) and (2.19)]. However, there are two differences that will be important in our later development. The results developed in [Sri02] are for a generic scalar statistic being compared to a threshold, while we are specifically considering the log-likelihood ratio expressed as a function of a multidimensional sufficient statistic $\mathbf{X}$,

$$l_*(\mathbf{X}) = \ln \frac{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)} \tag{2.336}$$

so that

$$e^{sl_*(\mathbf{X})} = \left(\frac{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}\right)^s. \tag{2.337}$$

In our case, $\mu(s)$ is defined as in (2.332) and the tilted density has the form

$$\boxed{p_{\mathbf{x}|0^*}(\mathbf{X};s) = e^{-\mu(s)} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)^{1-s} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)^s.} \tag{2.338}$$

The tilted density is specified in terms of the sufficient statistic $\mathbf{x}$, rather than the log-likelihood ratio l_*, and it can be found directly from the original densities $p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)$ and $p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)$. This is a key result, because in many, if not most, of the cases when we need to use this in practice, analytical expressions for $p_{l_*|H_0}(L|H_0)$ and $p_{l_*|H_1}(L|H_1)$ are not available. In addition, we will find that the optimum tilted density for estimating P_M is identical to $p_{\mathbf{x}|0^*}(\mathbf{X};s)$. In many cases we use $\mathbf{x} = \mathbf{r}$, corresponding to the original observations, and we tilt the N-dimensional densities $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$ and $p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)$. In practice, the case of most interest will be when the r_i are statistically independent (not necessarily identical) and we tilt the marginal probability densities. Thus, this more general approach enables us to find the tilted densities for a large number of useful applications. A final note is that although $W_0(\mathbf{X};s)$ is defined in terms of $p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)$ and $p_{\mathbf{x}|0^*}(\mathbf{X};s)$, the expression in (2.335) can be evaluated without specifying these densities.

From (2.330) and (2.331), the optimized upper bound is

$$\boxed{\overline{I}_F(\gamma_*;s) \triangleq e^{2[\mu(s)-s\gamma_*]} \quad s \geqslant 0.} \tag{2.339}$$

Comparing (2.339) to (2.212), we see that $\overline{I}_F(\gamma_*;s)$ is the square of the Chernoff bound. As with the Chernoff bound, we can get the tightest bound by minimizing it with respect to s. The optimum s is the same as for the Chernoff bound and is given by

$$s_* : \dot{\mu}(s_*) = \gamma_*, \tag{2.340}$$

as long as $\gamma_* \geqslant \dot{\mu}(0)$. If $\gamma_* < \dot{\mu}(0)$, the optimum value is $s_* = 0$, which gives the bound $\overline{I}_F(\gamma_*;0) = 1$.

The upper bound $\overline{I}_F(\gamma_*;s)$ was useful for showing that the optimal biasing density has the form of the tilted density given in (2.338) and for finding the value of s_* that minimizes the estimation variance. However, the bound is too weak to be used in (2.327) to determine the required number of trials. For this calculation, we would prefer to compute I_F itself. Substituting (2.335) in (2.325) gives an expression for I_F when the biasing density is the tilted density in (2.334),

$$\boxed{I_F(\gamma_*;s) \triangleq e^{\mu(s)} \int \mathbb{I}\big(l_*(\mathbf{X}) \geqslant \gamma_*\big) e^{-sl_*(\mathbf{X})} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\, d\mathbf{X}.} \tag{2.341}$$

Note that when $s = 0$, $I_F(\gamma_*;0) = P_F(\gamma_*)$ because $p_{\mathbf{x}|0^*}(\mathbf{X};0) = p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)$, and we do not get any advantage from importance sampling.

In problems where we can find an analytical expression for P_F, we will generally also be able to evaluate (2.341) to find an analytical expression for I_F. In cases where the evaluation is intractable, we can approximate I_F using the technique developed in Section 2.4 for approximating P_F. If we let $x = l_*$ and note that $p_{x|0^*}(X;s) = p_{x_s}(X)$, then substituting (2.335) into (2.324) and using the optimum s_* from (2.340) gives the following expression,

$$I_F(\gamma_*;s_*) = \int_{\dot{\mu}(s_*)}^{\infty} e^{2[\mu(s_*)-s_*X]} p_{x_s}(X)\, dX. \tag{2.342}$$

In Section 2.4, we started with a similar expression for P_F in (2.233) and derived the approximate expression in (2.238). Noting that (2.342) and (2.233) differ only in a factor of two in the exponent, we may follow the same argument and obtain an approximate expression for I_F,

$$\boxed{I_F(\gamma_*; s_*) \approx \left\{\exp\left[2\mu(s_*) - 2s_*\dot{\mu}(s_*) + 2s_*^2\ddot{\mu}(s_*)\right]\right\} \operatorname{erfc}_*\left[2s_*\sqrt{\ddot{\mu}(s_*)}\right].} \tag{2.343}$$

2.5.2.2 Simulation of P_M

The next step is to find the optimum biasing density to estimate P_M. The arguments are identical, so we omit some of the intermediate equations. We begin from

$$\begin{aligned} P_M(\gamma_*) &= \int \mathbb{I}(l_*(\mathbf{X}) < \gamma_*)\, p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\, d\mathbf{X} \\ &= \int \mathbb{I}(l_*(\mathbf{X}) < \gamma_*)\, W_1(\mathbf{X})\, p_{\mathbf{x}|1*}(\mathbf{X})\, d\mathbf{X}, \end{aligned} \tag{2.344}$$

where

$$W_1(\mathbf{X}) \triangleq \frac{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}{p_{\mathbf{x}|1*}(\mathbf{X})}. \tag{2.345}$$

The expression in (2.344) can be written as

$$P_M(\gamma_*) = E_{1*}\left[\mathbb{I}(l_*(\mathbf{X}) < \gamma_*)\, W_1(\mathbf{X})\right]. \tag{2.346}$$

We define

$$I_M(\gamma_*) \triangleq E_{1*}\left[\mathbb{I}^2(l_*(\mathbf{X}) < \gamma_*)\, W_1^2(\mathbf{X})\right] \tag{2.347}$$

$$= E_1\left[\mathbb{I}(l_*(\mathbf{X}) < \gamma_*)\, W_1(\mathbf{X})\right]. \tag{2.348}$$

Then,

$$\operatorname{Var}_{1*}\left(\hat{P}_M\right) = \frac{1}{K_M}\left(I_M - P_M^2\right) \tag{2.349}$$

and

$$K_{M,IS} = \frac{c^2}{\alpha^2}\frac{\left(I_M - P_M^2\right)}{\min\left(P_M, 1 - P_M\right)^2}. \tag{2.350}$$

We can upper bound $\mathbb{I}(l_*(\mathbf{X}) < \gamma_*)$ as shown in Figure 2.39,

$$\mathbb{I}(l_*(\mathbf{X}) < \gamma_*) \leqslant e^{-t[l_*(\mathbf{X}) - \gamma_*]} \quad t \geqslant 0. \tag{2.351}$$

Then,

$$I_M(\gamma_*) \leqslant E_{1*}\left[e^{-2t[l_*(\mathbf{x}) - \gamma_*]} W_1^2(\mathbf{x})\right] \triangleq \overline{I}_M(\gamma_*), \tag{2.352}$$

which is minimized when

$$e^{-t[l_*(\mathbf{X}) - \gamma_*]} W_1(\mathbf{X}) = E_1\left[e^{-t[l_*(\mathbf{X}) - \gamma_*]}\right]. \tag{2.353}$$

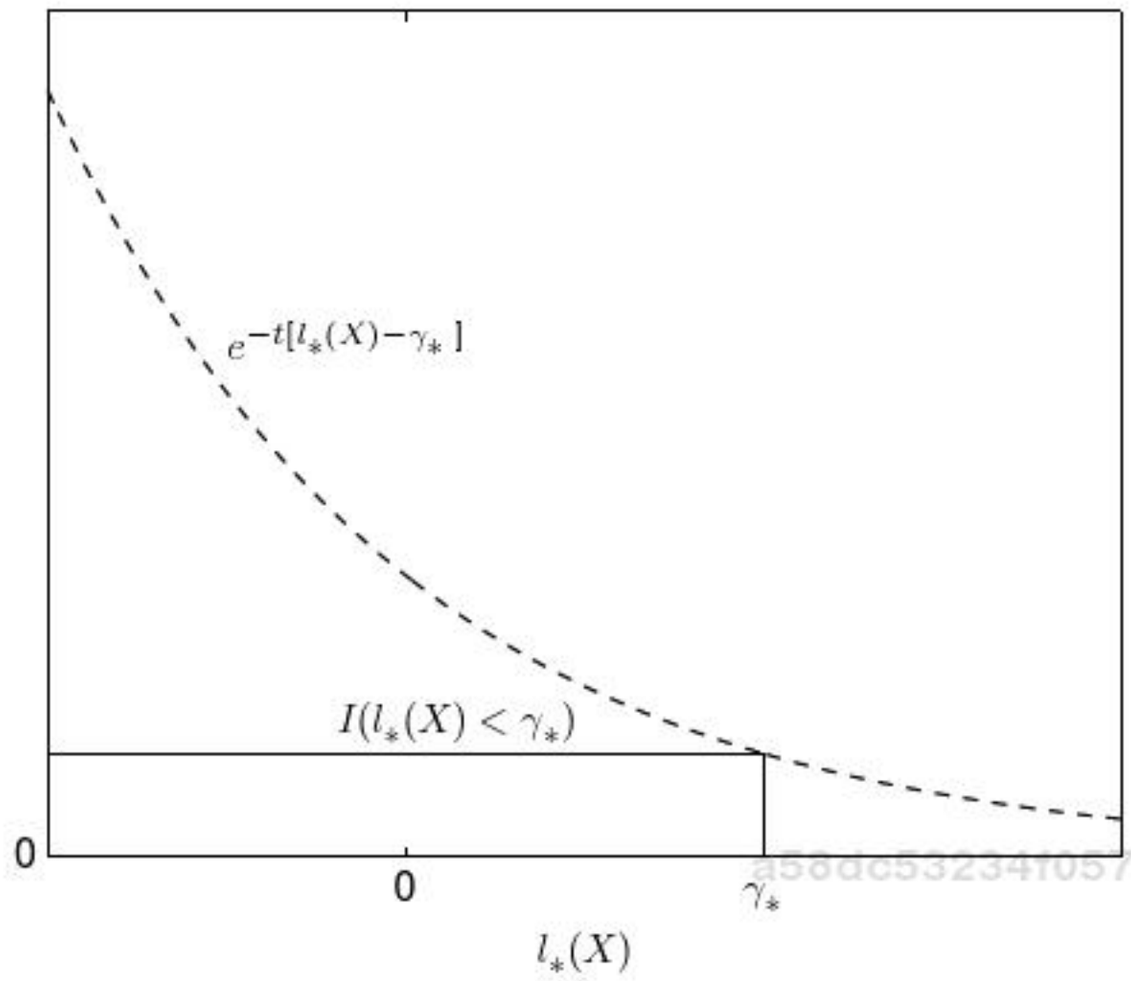


Figure 2.39: Upper bound on $\mathbb{I}(l_*(\mathbf{X}) < \gamma_*)$.

Now note that

$$e^{-tl_*(\mathbf{X})} = \left(\frac{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}\right)^{-t} = \left(\frac{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}\right)^{t}, \tag{2.354}$$

so

$$E_1\left[e^{-tl_*(\mathbf{X})}\right] = \int p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)^t p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)^{1-t}\, d\mathbf{X}$$
$$= e^{\mu(1-t)}. \tag{2.355}$$

From (2.353), $\overline{I}_M(\gamma_*)$ is minimized when

$$e^{-t[l_*(\mathbf{X})-\gamma_*]}\,\frac{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}{p_{\mathbf{x}|l^*}(\mathbf{X})} = e^{\mu(1-t)+t\gamma_*} \tag{2.356}$$

or

$$p_{\mathbf{x}|l^*}(\mathbf{X};t) = e^{-tl_*(\mathbf{X})-\mu(1-t)} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1). \tag{2.357}$$

Substituting (2.357) into (2.345), we obtain

$$W_1(\mathbf{X};t) = e^{tl_*(\mathbf{X})+\mu(1-t)}. \tag{2.358}$$

The optimized upper bound is

$$\overline{I}_M(\gamma_*;t) = e^{2[\mu(1-t)+t\gamma_*]} \quad t \geqslant 0. \tag{2.359}$$

Minimizing $\overline{I}_M(\gamma_*;t)$ with respect to t, we obtain

$$t_* : \dot{\mu}(1-t_*) = \gamma_*, \tag{2.360}$$

for $\gamma_* \leqslant \dot{\mu}(1)$. If $\gamma_* > \dot{\mu}(1)$, the optimum value is $t_* = 0$, which gives the bound $\overline{I}_M(\gamma_*;0) = 1$.

Substituting (2.358) in (2.348) gives the following expression for I_M,

$$I_M(\gamma_*;t) = e^{\mu(1-t)} \int \mathbb{I}(l_*(\mathbf{X}) < \gamma_*)\, e^{tl_*(\mathbf{X})} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\, d\mathbf{X}. \tag{2.361}$$

In this case when $t = 0$, $I_M(\gamma_*;0) = P_M(\gamma_*)$, and $p_{\mathbf{x}|1*}(\mathbf{X};0) = p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)$, and there is no importance sampling gain.

If we let

$$s = 1 - t \quad s \leqslant 1, \tag{2.362}$$

and use the expression for $l_*(\mathbf{X})$ in (2.336), the tilted density in (2.357) becomes

$$\boxed{p_{\mathbf{x}|1*}(\mathbf{X};s) = e^{-\mu(s)} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)^{1-s} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)^s} \tag{2.363}$$

which is the same as the tilted density in (2.338). We use the following notation for the optimum tilted density for estimating both P_F and P_M,

$$\boxed{p_{\mathbf{x}|s_*}(\mathbf{X}) = e^{-\mu(s_*)} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)^{1-s_*} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)^{s_*}.} \tag{2.364}$$

In terms of s, (2.358)–(2.361) become

$$\boxed{W_1(\mathbf{X};s) = e^{(1-s)l_*(\mathbf{X})+\mu(s)}.} \tag{2.365}$$

$$\overline{I}_M(\gamma_*;s) = e^{2[\mu(s)+(1-s)\gamma_*]} \quad s \leqslant 1. \tag{2.366}$$

$$s_* : \dot{\mu}(s_*) = \gamma_*, \tag{2.367}$$

$$I_M(\gamma_*;s) = e^{\mu(s)} \int \mathbb{I}(l_*(\mathbf{X}) < \gamma_*)\, e^{(1-s)l_*(\mathbf{X})} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\, d\mathbf{X}. \tag{2.368}$$

From (2.367) and (2.340), we see that the optimum tilting specified by s_* is the same for both P_M and P_F.

If we let $x = l_*$ and note that $p_{x|1*}(X;s) = p_{x_s}(X)$, then substituting (2.365) into (2.347) and using the optimum s_* from (2.367) gives the following expression,

$$I_M(\gamma_*;s_*) = \int_{-\infty}^{\dot{\mu}(s_*)} e^{2[\mu(s_*)+(1-s_*)X]} p_{x_s}(X)\, dX, \tag{2.369}$$

from which we can derive the approximate expression for $I_M(\gamma_*;s_*)$:

$$\boxed{I_M(\gamma_*;s_*) \approx \left\{\exp\left[2\mu(s_*)+2(1-s_*)\dot{\mu}(s_*)+2(1-s_*)^2\ddot{\mu}(s_*)\right]\right\} \operatorname{erfc}_*\left[2(1-s_*)\sqrt{\ddot{\mu}(s_*)}\right].} \tag{2.370}$$

2.5.2.3 Independent Observations

The optimum tilted density in terms of the original observations is found by letting $\mathbf{X} = \mathbf{R}$ in (2.364),

$$p_{\mathbf{r}|s_*}(\mathbf{R}) = e^{-\mu(s_*)} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)^{1-s_*} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)^{s_*}. \tag{2.371}$$

If the observations are independent, we substitute (2.245) and (2.247) in (2.371) to obtain

$$p_{\mathbf{r}|s_*}(\mathbf{R}) = \exp\left(-\sum_{i=1}^{N} \mu_i(s)\right)\left(\prod_{i=1}^{N} p_{r_i|H_0}(R_i|H_0)\right)^{1-s_*}\left(\prod_{i=1}^{N} p_{r_i|H_1}(R_i|H_1)\right)^{s_*}$$

$$= \prod_{i=1}^{N}\left(e^{-\mu_i(s_*)} p_{r_i|H_0}(R_i|H_0)^{1-s_*} p_{r_i|H_1}(R_i|H_1)^{s_*}\right) \tag{2.372}$$

that can be written as

$$p_{\mathbf{r}|s_*}(\mathbf{R}) = \prod_{i=1}^{N} p_{r_i|s_*}(R_i) \tag{2.373}$$

where the tilted marginal probability density is defined as

$$p_{r_i|s_*}(R_i) \triangleq e^{-\mu_i(s_*)} p_{r_i|H_0}(R_i|H_0)^{1-s_*} p_{r_i|H_1}(R_i|H_1)^{s_*}. \tag{2.374}$$

This is a key result for the cases in which it is difficult to compute $p_{l_*|H_0}(L|H_0)$ and $p_{l_*|H_1}(L|H_1)$. Most applications with non-Gaussian observations fall into this category. We can use the tilted marginal probability densities to simulate the r_i.

2.5.2.4 Simulation of the ROC

To summarize our results, we have the following procedure for simulating a (P_F, P_D) point on the ROC curve:

1. Compute $\mu(s)$ from (2.332),

$$\mu(s) = \ln \int \left[p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\right]^s \left[p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\right]^{1-s} d\mathbf{X}. \tag{2.375}$$

For an ID model, we can compute $\mu_i(s)$ from (2.248) and $\mu(s)$ from (2.247),

$$\mu_i(s) = \ln \int \left[p_{r_i|H_1}(R_i|H_1)\right]^s \left[p_{r_i|H_0}(R_i|H_0)\right]^{1-s} dR_i \tag{2.376}$$

$$\mu(s) = \sum_{i=1}^{N} \mu_i(s). \tag{2.377}$$

Differentiate $\mu(s)$ to obtain $\dot{\mu}(s)$ and $\ddot{\mu}(s)$. Note that $\ddot{\mu}(s)$ is only needed if we are computing approximate expressions in step 3.

2. Find s_* using

$$s_* = \begin{cases} 0 & \gamma_* < \dot{\mu}(0) \\ s_* : \dot{\mu}(s_*) = \gamma_* & \dot{\mu}(0) \leqslant \gamma_* \leqslant \dot{\mu}(1) \\ 1 & \gamma_* > \dot{\mu}(1) \end{cases} \tag{2.378}$$

3. Find $P_F(\gamma_*)$, $P_M(\gamma_*)$, $I_F(\gamma_*;s)$, and $I_M(\gamma_*;s)$. Exact analytical expressions can be obtained from (2.317), (2.344), (2.341), and (2.368), when they can be evaluated.

$$P_F(\gamma_*) = \int \mathbb{I}\big(l_*(\mathbf{X}) \geqslant \gamma_*\big)\, p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\, d\mathbf{X}, \tag{2.379}$$

$$P_M(\gamma_*) = \int \mathbb{I}\,(l_*(\mathbf{X}) < \gamma_*)\, p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\, d\mathbf{X}, \tag{2.380}$$

$$I_F(\gamma_*;s) = e^{\mu(s)} \int \mathbb{I}\big(l_*(\mathbf{X}) \geqslant \gamma_*\big) e^{-sl_*(\mathbf{X})} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)\, d\mathbf{X}, \tag{2.381}$$

$$I_M(\gamma_*;s) = e^{\mu(s)} \int \mathbb{I}\,(l_*(\mathbf{X}) < \gamma_*)\, e^{(1-s)l_*(\mathbf{X})} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)\, d\mathbf{X}. \tag{2.382}$$

Alternatively, use the approximate expressions in (2.238), (2.241), (2.343), and (2.370),

$$P_F(\gamma_*) \approx \left\{\exp\left[\mu(s_*) - s_*\dot{\mu}(s_*) + \frac{s_*^2}{2}\ddot{\mu}(s_*)\right]\right\} \times \operatorname{erfc}_*\left[s_*\sqrt{\ddot{\mu}(s_*)}\right], \tag{2.383}$$

$$P_M(\gamma_*) \approx \left\{\exp\left[\mu(s_*) + (1-s_*)\dot{\mu}(s_*) + \frac{(s_*-1)^2}{2}\ddot{\mu}(s_*)\right]\right\} \times \operatorname{erfc}_*\left[(1-s_*)\sqrt{\ddot{\mu}(s_*)}\right], \tag{2.384}$$

$$I_F(\gamma_*;s_*) \approx \left\{\exp\left[2\mu(s_*) - 2s_*\dot{\mu}(s_*) + 2s_*^2\ddot{\mu}(s_*)\right]\right\} \operatorname{erfc}_*\left[2s_*\sqrt{\ddot{\mu}(s_*)}\right], \tag{2.385}$$

$$I_M(\gamma_*;s_*) \approx \left\{\exp\left[2\mu(s_*) + 2(1-s_*)\dot{\mu}(s_*) + 2(1-s_*)^2\ddot{\mu}(s_*)\right]\right\} \times \operatorname{erfc}_*\left[2(1-s_*)\sqrt{\ddot{\mu}(s_*)}\right]. \tag{2.386}$$

Note that we can only use the approximate expressions for thresholds that satisfy $\dot{\mu}(0) \leqslant \gamma_* \leqslant \dot{\mu}(1)$, while the exact expressions are valid for any γ_*.

4. Pick a sufficient statistic $\mathbf{x}$ to be used in the simulation and evaluate (2.278) to find an expression for $l_*(\mathbf{X})$,

$$l_*(\mathbf{X}) = \ln \frac{p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)}{p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)}. \tag{2.387}$$

5. Find the optimum tilted density from (2.364),

$$p_{\mathbf{x}|s_*}(\mathbf{X}) = e^{-\mu(s_*)} p_{\mathbf{x}|H_0}(\mathbf{X}|H_0)^{1-s_*} p_{\mathbf{x}|H_1}(\mathbf{X}|H_1)^{s_*}. \tag{2.388}$$

For an ID model, the optimum tilted marginal densities are found from (2.374),

$$p_{r_i|s_*}(R_i) = e^{-\mu_i(s_*)} p_{r_i|H_0}(R_i|H_0)^{1-s_*} p_{r_i|H_1}(R_i|H_1)^{s_*}. \tag{2.389}$$

6. Find the weighting functions from (2.335) and (2.365),

$$W_0(\mathbf{X}; s_*) = e^{-s_* l_*(\mathbf{X}) + \mu(s_*)}, \tag{2.390}$$

$$W_1(\mathbf{X}; s_*) = e^{(1-s_*) l_*(\mathbf{X}) + \mu(s_*)}. \tag{2.391}$$

7. Specify the confidence interval parameters α and c and compute $K_{F,IS}$ and $K_{M,IS}$ from using (2.327) and (2.350),

$$K_{F,IS} = \frac{c^2}{\alpha^2} \frac{I_F(\gamma_*; s_*) - P_F(\gamma_*)^2}{\min\left[P_F(\gamma_*), 1 - P_F(\gamma_*)\right]^2} \tag{2.392}$$

$$K_{M,IS} = \frac{c^2}{\alpha^2} \frac{I_M(\gamma_*; s_*) - P_M(\gamma_*)^2}{\min\left[P_M(\gamma_*), 1 - P_M(\gamma_*)\right]^2}. \tag{2.393}$$

Select the larger value for the simulation,

$$K_{IS} = \max\left(K_{F,IS}, K_{M,IS}\right). \tag{2.394}$$

8. Generate K_{IS} independent realizations of $\mathbf{x}$ from $p_{\mathbf{x}|s_*}(\mathbf{X})$ and compute the estimates

$$\hat{P}_F(\gamma_*) = \frac{1}{K_{IS}} \sum_{k=1}^{K_{IS}} \mathbb{I}\left(l_*(\mathbf{X}_k; s_*) \geqslant \gamma_*\right) W_0(\mathbf{X}_k; s_*) \tag{2.395}$$

$$\hat{P}_M(\gamma_*) = \frac{1}{K_{IS}} \sum_{k=1}^{K_{IS}} \mathbb{I}\left(l_*(\mathbf{X}_k; s_*) < \gamma_*\right) W_1(\mathbf{X}_k; s_*). \tag{2.396}$$

An estimate of $P_D(\gamma_*)$ can be found from

$$\hat{P}_D(\gamma_*) = 1 - \hat{P}_M(\gamma_*) = \frac{1}{K_{IS}} \sum_{k=1}^{K_{IS}} \mathbb{I}\left(l_*(\mathbf{X}_k; s_*) \geqslant \gamma_*\right) W_1(\mathbf{X}_k; s_*). \tag{2.397}$$

It is important to note that $\mu(s)$, $P_F(\gamma_*)$, $P_M(\gamma_*)$, $I_F(\gamma_*; s)$, $I_M(\gamma_*; s)$, $W_0(\mathbf{X}; s)$, and $W_1(\mathbf{X}; s)$ are all defined in terms of a sufficient statistic $\mathbf{X}$. When computing these quantities, we do not need to use the same statistic, and in fact we will usually find it convenient to use different statistics.

2.5.2.5 Examples

In this section, we apply importance sampling using the optimum tilted density. The first two examples are continuations of Examples 2.1 and 2.2. Because we have analytic results, we probably would not simulate these models in practice, but they are useful to introduce the simulation procedure. The third example illustrates the case where we can find an analytic expression for $\mu(s)$, but cannot find $p_{l_*|H_0}(L|H_0)$ and $p_{l_*|H_1}(L|H_1)$ and cannot evaluate P_F and P_M. We use importance sampling with the approximate expressions and develop an iterative importance sampling scheme in Section 2.5.2.6 to solve this case.

Example 2.16 (continuation of Examples 2.1, 2.5, 2.12, and 2.15). For this model, $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$ were computed in Example 2.12. Recall from (2.257)–(2.259) that

$$\mu_i(s) = \frac{s(s-1)d^2}{2N}, \tag{2.398}$$

$$\mu(s) = \frac{s(s-1)d^2}{2}, \tag{2.399}$$

$$\dot{\mu}(s) = \frac{(2s-1)d^2}{2}. \tag{2.400}$$

To find s_*, we solve $\dot{\mu}(s) = \gamma_*$. The result is

$$s_* = \begin{cases} 0 & \gamma_* < -\dfrac{d^2}{2} \\ \dfrac{\gamma_*}{d^2} + \dfrac{1}{2} & -\dfrac{d^2}{2} \leqslant \gamma_* \leqslant \dfrac{d^2}{2} \\ 1 & \gamma_* > \dfrac{d^2}{2}. \end{cases} \tag{2.401}$$

$P_F(\gamma_*)$ and $P_M(\gamma_*)$ were found in Examples 2.5 and 2.15 and are given in (2.308) and (2.309).

We next compute $I_F(\gamma_*; s)$ using (2.381). It is convenient to use

$$x_0 = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} r_i, \tag{2.402}$$

which is $N(0, 1)$ on H_0. Then, from (2.24)

$$l_*(X_0) = dX_0 - \frac{d^2}{2} \tag{2.403}$$

and $l_*(X_0) \geqslant \gamma_*$ when $X_0 \geqslant Z_F$, where Z_F was defined in (2.312). Thus,

$$I_F(\gamma_*; s) = e^{\frac{s(s-1)d^2}{2}} \int_{Z_F}^{\infty} e^{-sdX_0 + \frac{sd^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{X_0^2}{2}} dX_0, \tag{2.404}$$

which reduces to

$$I_F(\gamma_*; s) = e^{s^2d^2}\,\mathrm{erfc}_*(Z_F + sd) = e^{(sd)^2}\,\mathrm{erfc}_*\left(\frac{\gamma_*}{d} + \frac{d}{2} + sd\right). \tag{2.405}$$

To calculate $I_M(\gamma_*; s)$, we proceed in a similar manner. We define

$$x_1 = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} r_i - d, \tag{2.406}$$

which is $N(0, 1)$ on H_1. In this case,

$$l_*(X_1) = dX_1 + \frac{d^2}{2} \tag{2.407}$$

and $l_*(X_1) < \gamma_*$ when $X_1 < Z_F - d$. Evaluating (2.382), we obtain

$$I_M(\gamma_*; s) = e^{[(s-1)d]^2}\left[1 - \mathrm{erfc}_*\left(\frac{\gamma_*}{d} - \frac{d}{2} + (s-1)d\right)\right]. \tag{2.408}$$

Table 2.3: Importance sampling simulation values for $d = 6$, $p_c = 0.954$, and $\alpha = 0.1$

P_F	γ_*	s_*	P_M	I_F	I_M	$K_{F,IS}$	$K_{M,IS}$
10^{-2}	−5.29	0.37	3.2×10^{-5}	3.7×10^{-4}	5.6×10^{-9}	1069	1804
10^{-4}	3.52	0.59	0.0046	5.2×10^{-8}	8.3×10^{-5}	1676	1185
10^{-6}	10.06	0.75	0.0581	6.4×10^{-12}	0.0099	2155	774
10^{-8}	15.49	0.89	0.2381	7.4×10^{-16}	0.1281	2562	504
10^{-10}	20.23	1	0.5147	8.3×10^{-20}	0.5147	2934	425
10^{-12}	24.49	1	0.7611	1.2×10^{-23}	0.7611	4539	1275

For the simulation, we choose our sufficient statistic to be the original data $\mathbf{x} = \mathbf{r}$, which are IID. From (2.24),

$$l_*(\mathbf{R}) = \left(\frac{1}{\sqrt{N}\sigma} \sum_{i=1}^{N} R_i \right) d - \frac{d^2}{2}. \tag{2.409}$$

The tilted marginal densities are given by (2.389). Substituting (2.19), (2.20), and (2.398) into (2.389) and simplifying gives

$$p_{r_i|s_*}(R_i) \sim N\left(s_* m, \sigma^2\right). \tag{2.410}$$

We see that the tilted density corresponds to a Gaussian density with the same variance but with a mean value of $s_* m$ that varies between 0 and m. When $s_* = 0$, $p_{r_i|s_*}(R_i) = p_{r_i|H_0}(R_i|H_0)$, and when $s_* = 1$, $p_{r_i|s_*}(R_i) = p_{r_i|H_1}(R_i|H_1)$.

From (2.390) and (2.391), the weighting functions are

$$W_0(\mathbf{R}; s_*) = e^{-s_* l_*(\mathbf{R})} e^{\mu(s_*)} = e^{-s_* l_*(\mathbf{R})} e^{s_*(s_*-1)\frac{d^2}{2}}, \tag{2.411}$$

$$W_1(\mathbf{R}; s_*) = e^{(1-s_*) l_*(\mathbf{R})} e^{\mu(s_*)} = e^{(1-s_*) l_*(\mathbf{R})} e^{s_*(s_*-1)\frac{d^2}{2}}. \tag{2.412}$$

We consider the parameter values used in Table 2.2: $d = \sqrt{40}$, $c = 2$, $\alpha = 0.1$, and a variety of ROC points. For $d = \sqrt{40}$, $\dot{\mu}(0) = -20$, and $\dot{\mu}(1) = 20$. We compute s_*, $I_F(\gamma_*; s_*)$, $I_M(\gamma_*; s_*)$, $K_{F,IS}$, and $K_{M,IS}$. The results are summarized in Table 2.3. We see that in all cases $I_F < P_F$ and $K_{F,IS}$ is significantly smaller than the corresponding K_F in Table 2.2. In the last two rows of the table, we have $\gamma_* > \dot{\mu}(1)$ and the optimum $s_* = 1$. For these cases, $I_M = P_M$ and $K_{M,IS} = K_M$. For the other rows, $I_M < P_M$ and $K_{M,IS} < K_M$. We see that importance sampling has dramatically reduced the required number of trials when P_F or P_M is very small.

We run the simulation for $P_F = 10^{-6}$, choosing $N = 40$, $m = 1$, and $\sigma^2 = 1$. From Table 2.3, $P_M = 0.0581$ and $P_D = 0.9419$, and the required number of trials is $K_{IS} = \max(2155, 774) = 2155$. The confidence intervals are:

$$\hat{P}_F \in [0.9, 1.1] \times 10^{-6},$$

$$\hat{P}_D \in [0.9384, 0.9454].$$

We generate the required number of trials using the tilted marginal density[11] and compute $\hat{P}_F$ and $\hat{P}_D$ from (2.395) and (2.397). (Without importance sampling, we would have had to run two separate

[11] The random samples are generated using the Matlab function `randn`.

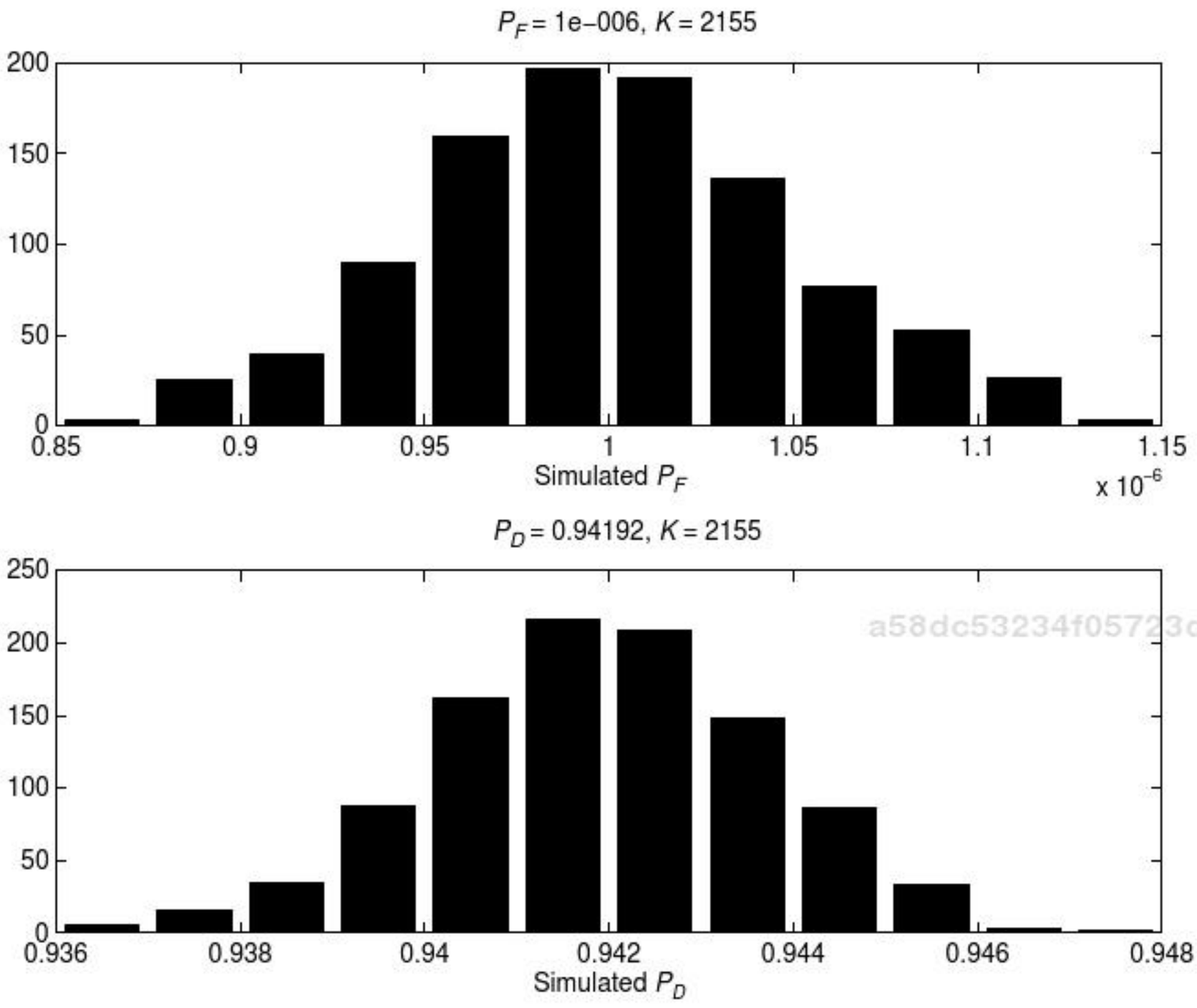


Figure 2.40: Histograms of $\hat{P}_F$ and $\hat{P}_D$.

simulations to estimate $\hat{P}_F$ and $\hat{P}_D$ requiring $K_F = 4 \times 10^8$ and $K_M = 6488$ trials, respectively.) We repeat the simulation 1000 times. Histograms of $\hat{P}_F$ and $\hat{P}_D$ are shown in Figure 2.40. We see that the results are consistent with the confidence interval specifications.

We next consider the ROC curves plotted in Figure 2.12*b*. We run the simulation for various values of d and P_F and plot the simulation results on top of the analytic results in Figure 2.41. We see that there is excellent agreement between the simulation and the analytic results. ■

Example 2.17 (continuation of Examples 2.2, 2.6, and 2.13). In this example, the r_i are IID zero-mean Gaussian random variables with different variances on the two hypotheses:[12]

$$p_{r_i|H_j}(R_i|H_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{1}{2}\frac{R_i^2}{\sigma_j^2}\right\}, \quad i = 1, 2, \ldots, N \text{ and } j = 0, 1. \tag{2.413}$$

We assume $\sigma_1^2 > \sigma_0^2$.

We computed P_F and P_D in Example 2.6. From (2.109) and (2.110),[13]

$$P_F(\gamma) = 1 - \Gamma_{N/2}\left(\frac{\gamma}{2\sigma_0^2}\right), \tag{2.414}$$

[12]This example also has an analytic solution for P_D and P_F. However, it serves as an introduction to the problem whose σ_j^2 is different on each observation and an analytic solution is not available.

[13]For a specified P_F or P_M, γ can be computed in Matlab using the `gaminv` function. See Appendix A.

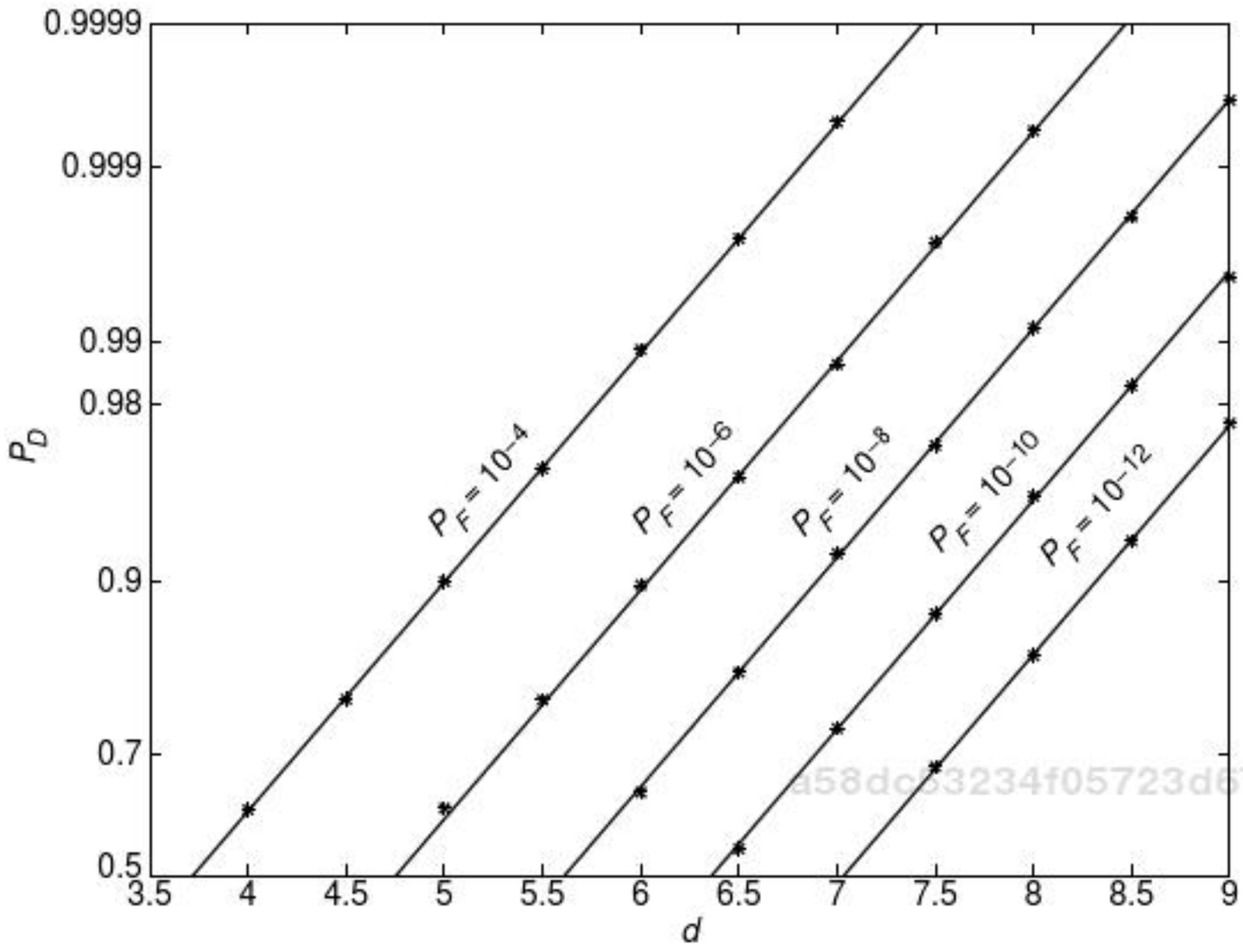


Figure 2.41: Simulated P_D versus d for various P_F using optimum tilted density.

$$P_M(\gamma) = \Gamma_{N/2}\left(\frac{\gamma}{2\sigma_1^2}\right), \tag{2.415}$$

where γ was defined in terms of $\gamma_* = \ln\eta$ in (2.91),

$$\gamma = \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\left[\gamma_* - \frac{N}{2}\ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right)\right]. \tag{2.416}$$

For notational simplicity, it will be easier to work with γ instead of γ_*. We can easily convert the results back to γ_* by rearranging (2.416) to obtain,

$$\gamma_* = \frac{\gamma\left(\sigma_1^2 - \sigma_0^2\right)}{2\sigma_0^2\sigma_1^2} + \frac{N}{2}\ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right). \tag{2.417}$$

We computed $\mu_i(s)$ and $\mu(s)$ in Example 2.13. Recall from (2.268) and (2.269),

$$\mu(s) = N\mu_i(s) = \frac{N}{2}\ln\left[\frac{\left(\sigma_0^2\right)^s\left(\sigma_1^2\right)^{1-s}}{s\sigma_0^2 + (1-s)\sigma_1^2}\right]. \tag{2.418}$$

Differentiating gives

$$\dot{\mu}(s) = \frac{N}{2}\left[\ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right) + \frac{\sigma_1^2 - \sigma_0^2}{s\sigma_0^2 + (1-s)\sigma_1^2}\right]. \tag{2.419}$$

To find s_* we solve $\dot{\mu}(s) = \gamma_*$ for $\dot{\mu}(0) \leqslant \gamma_* \leqslant \dot{\mu}(1)$ and then use (2.417) to express the results in terms of γ. The result is

$$s_* = \begin{cases} 0 & \gamma < N\sigma_0^2 \\ \dfrac{\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\left(1 - \dfrac{N\sigma_0^2}{\gamma}\right) & N\sigma_0^2 \leqslant \gamma \leqslant N\sigma_1^2 \\ 1 & \gamma > N\sigma_1^2. \end{cases} \tag{2.420}$$

To find $I_F(\gamma; s)$, we define

$$x = \sum_{i=1}^{N} R_i^2 \tag{2.421}$$

and evaluate (2.381) using an analysis similar to Example 2.6 to obtain[14]

$$I_F(\gamma; s) = \left[\frac{\sigma_1^2}{s\sigma_0^2 + (1-s)\sigma_1^2}\right]^{\frac{N}{2}} \left[\frac{\sigma_1^2}{-s\sigma_0^2 + (1+s)\sigma_1^2}\right]^{\frac{N}{2}} \left(1 - \Gamma_{N/2}\left[\gamma \frac{-s\sigma_0^2 + (1+s)\sigma_1^2}{2\sigma_0^2\sigma_1^2}\right]\right). \tag{2.422}$$

Similarly,

$$I_M(\gamma; s) = \left[\frac{\sigma_0^2}{s\sigma_0^2 + (1-s)\sigma_1^2}\right]^{\frac{N}{2}} \left[\frac{\sigma_0^2}{(2-s)\sigma_0^2 + (s-1)\sigma_1^2}\right]^{\frac{N}{2}} \Gamma_{N/2}\left[\gamma \frac{(2-s)\sigma_0^2 + (s-1)\sigma_1^2}{2\sigma_0^2\sigma_1^2}\right]. \tag{2.423}$$

For the simulation, we choose our sufficient statistic to be the original data $\mathbf{x} = \mathbf{r}$, which are IID. From (2.29), the log-likelihood ratio is

$$l_*(\mathbf{R}) = \frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum_{i=1}^{N} R_i^2 + N \ln \frac{\sigma_0}{\sigma_1} = \left(\frac{\sigma_1^2 - \sigma_0^2}{2\sigma_1^2\sigma_0^2}\right)\sum_{i=1}^{N} R_i^2 + \frac{N}{2}\ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right). \tag{2.424}$$

To find the tilted marginal densities, we use (2.413) and (2.418) in (2.389) to obtain

$$p_{r_i|s_*}(R_i) = \left[\frac{\left(\sigma_0^2\right)^{s_*}\left(\sigma_1^2\right)^{1-s_*}}{s_*\sigma_0^2 + (1-s_*)\sigma_1^2}\right]^{-1/2} \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\, e^{-R_i^2/2\sigma_0^2}\right)^{1-s_*} \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\, e^{-R_i^2/2\sigma_1^2}\right)^{s_*} \tag{2.425}$$

which reduces to

$$p_{r_i|s_*}(R_i) \sim N\left(0, \sigma_{s_*}^2\right) \tag{2.426}$$

where

$$\sigma_{s_*}^2 \triangleq \frac{\sigma_0^2\sigma_1^2}{s_*\sigma_0^2 + (1-s_*)\sigma_1^2}. \tag{2.427}$$

We see that the tilted marginal density is a zero-mean Gaussian density with variance $\sigma_{s_*}^2$. When $s_* = 0$, $\sigma_{s_*}^2 = \sigma_0^2$, and when $s_* = 1$, $\sigma_{s_*}^2 = \sigma_1^2$.

The weighting functions are

$$W_0(\mathbf{R}; s_*) = e^{-s_* l_*(\mathbf{R})} e^{\mu(s_*)} = e^{-s_* l_*(\mathbf{R})} \left[\frac{\left(\sigma_0^2\right)^{s_*}\left(\sigma_1^2\right)^{1-s_*}}{s_*\sigma_0^2 + (1-s_*)\sigma_1^2}\right]^{\frac{N}{2}}, \tag{2.428}$$

$$W_1(\mathbf{R}; s_*) = e^{(1-s_*) l_*(\mathbf{R})} e^{\mu(s_*)} = e^{(1-s_*) l_*(\mathbf{R})} \left[\frac{\left(\sigma_0^2\right)^{s_*}\left(\sigma_1^2\right)^{1-s_*}}{s_*\sigma_0^2 + (1-s_*)\sigma_1^2}\right]^{\frac{N}{2}}. \tag{2.429}$$

We consider the following parameter values: $N = 8$, $\sigma_0^2 = 1$, $\sigma_1^2 = 21$, $p_c = 0.954$ ($c = 2$), and $\alpha = 0.1$, and a variety of ROC points. We compute the quantities of interest and summarize the results in Table 2.4. In this example, $I_F < P_F$ for all the ROC points and $K_{F,IS} < K_F$. Also, $I_M < P_M$ and

[14]The analysis is straightforward but tedious.

Table 2.4: Importance sampling simulation values for $N = 8$, $\sigma_0^2 = 1$, $\sigma_1^2 = 21$, $p_c = 0.954$, and $\alpha = 0.1$

P_F	γ	s_*	P_M	I_F	I_M	K_F	$K_{F,IS}$	K_M	$K_{M,IS}$
10^{-4}	31.8	0.79	0.0076	8.8×10^{-8}	0.0001	4×10^6	3129	5.3×10^4	573
10^{-6}	42.7	0.85	0.0200	1.2×10^{-11}	0.0009	4×10^8	4495	2.0×10^4	536
10^{-8}	53.2	0.89	0.0398	1.6×10^{-15}	0.0036	4×10^{10}	5820	9658	503
10^{-10}	63.4	0.92	0.0668	1.9×10^{-19}	0.0097	4×10^{12}	7120	5585	472
10^{-12}	73.5	0.94	0.1007	2.2×10^{-23}	0.0214	4×10^{14}	8401	3573	444

$K_{M,IS} < K_M$. Again, importance sampling has significantly reduced the required number of trials when P_F and P_M are very small.

Next we let $\sigma_1^2 = \sigma_s^2 + \sigma_n^2$ and study the behavior as a function of σ_s^2/σ_n^2. We run the simulation for various values of σ_s^2/σ_n^2 and P_F by generating the required number of trials using the tilted marginal density and computing $\hat{P}_F$ and $\hat{P}_D$ using (2.395) and (2.397). In Figure 2.42, we plot the analytic results and simulation results for P_D versus σ_s^2/σ_n^2 for various P_F. Once again, the agreement is excellent. ■

In Examples 2.16 and 2.17, we had analytic expressions for P_F and P_D, so we could specify the desired P_F and solve for threshold γ_*. In this case, we do not really need to simulate the likelihood ratio test. A more realistic case is when we do not have an analytic expression relating P_F and γ_*. We consider this case in the next example.

Example 2.18. The observations on H_0 and H_1 are statistically independent draws from a Beta probability density,

$$p_{r_i|H_j}(R_i|H_j) = \frac{1}{B(a_j, b_j)} R_i^{a_j - 1}(1 - R_i)^{b_j - 1}, \quad 0 \leqslant r_i \leqslant 1; i = 1, 2, \ldots, N; j = 0, 1, \qquad (2.430)$$

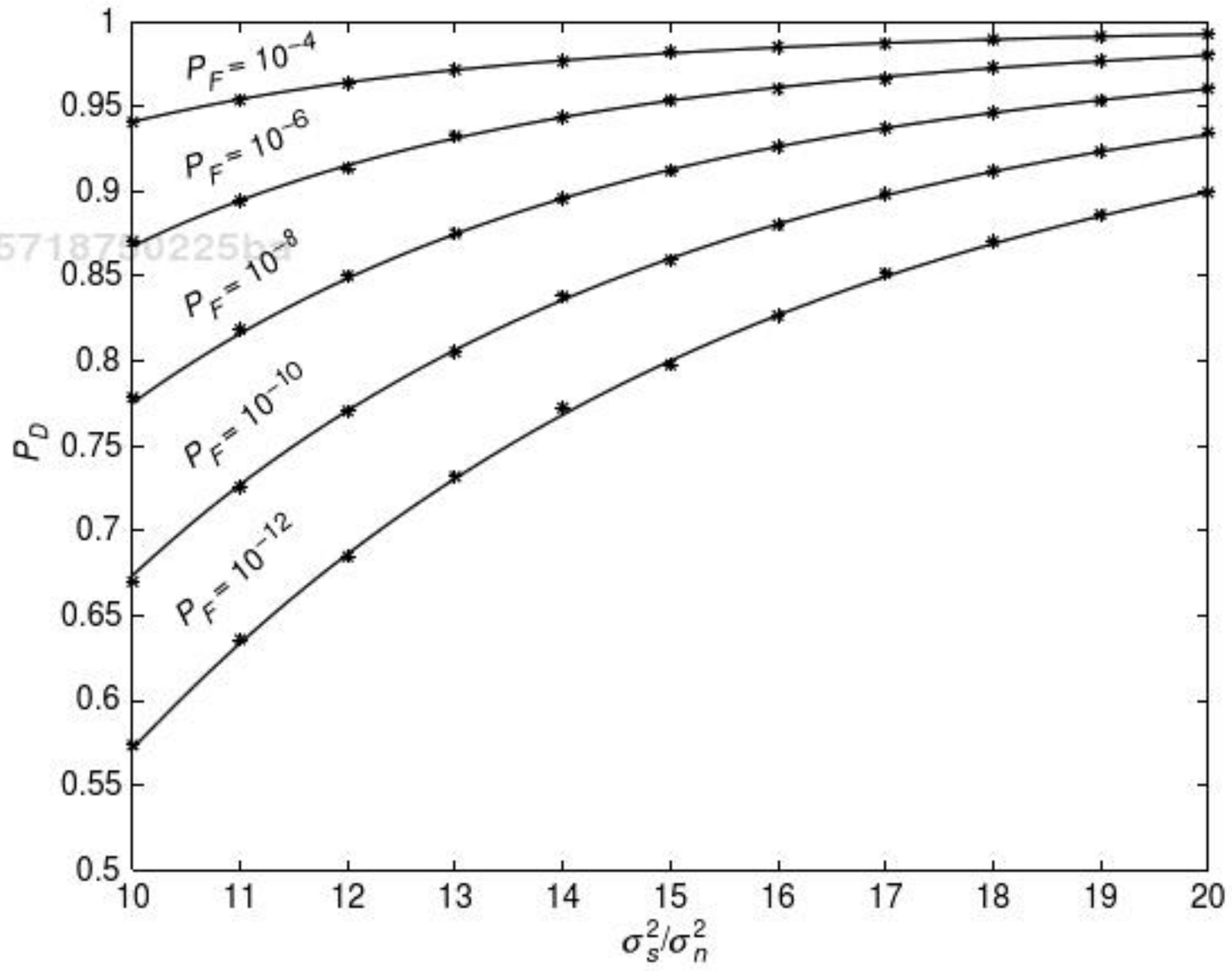


Figure 2.42: Simulated P_D versus σ_s^2/σ_n^2 for various P_F using optimum tilted density.

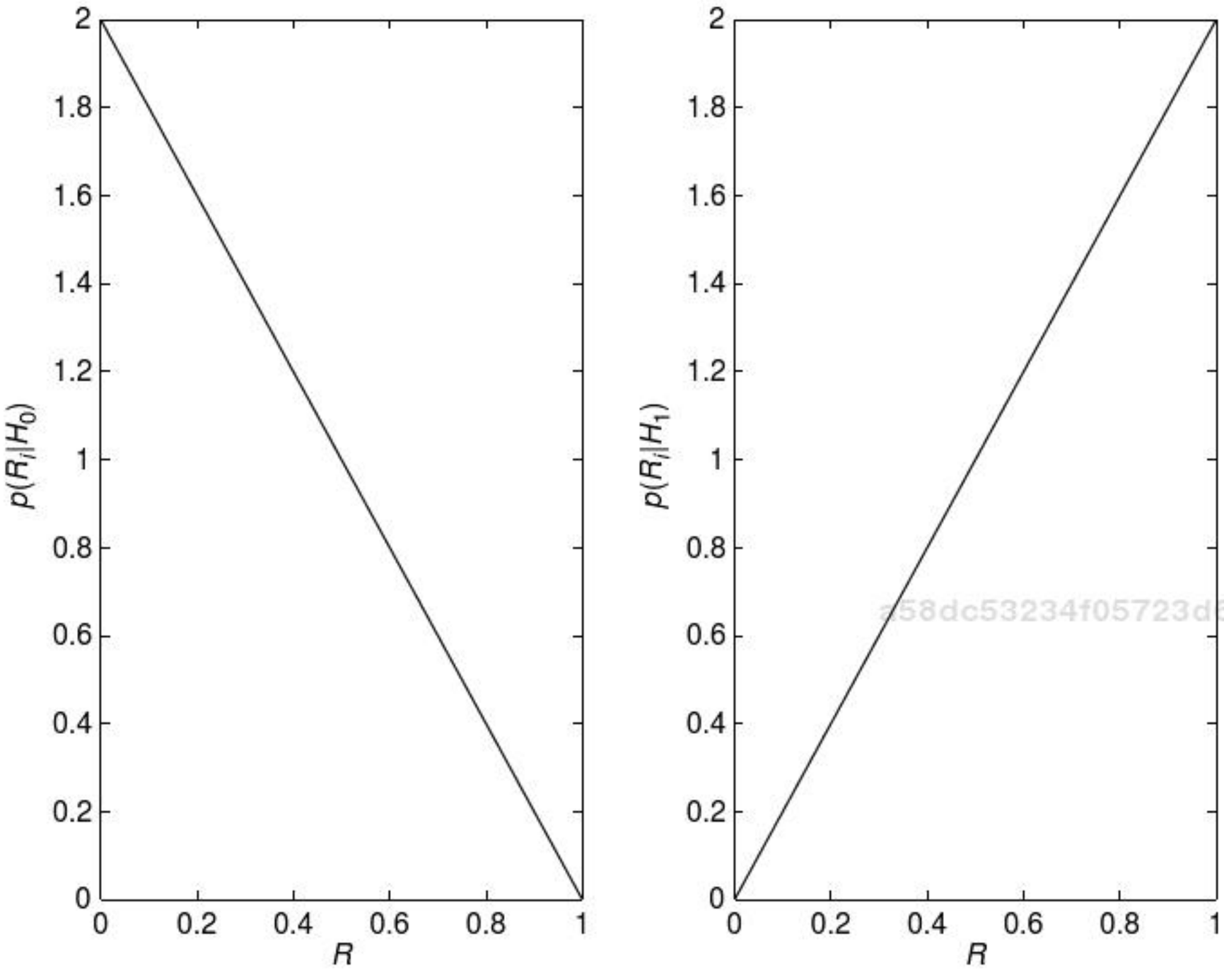


Figure 2.43: Probability densities; $p_{r_i|H_0}(R_i|H_0)$ and $p_{r_i|H_1}(R_i|H_1)$.

where $B(a, b)$ is the Beta function:[15]

$$B(a, b) \triangleq \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{2.431}$$

On H_0 we have $a_0 = 1, b_0 = 2$ and on H_1 we have $a_1 = 2, b_1 = 1$, which gives the following densities

$$\begin{aligned} H_0 &: p_{r_i|H_0}(R_i|H_0) = 2(1-R_i) \quad 0 \leqslant R_i \leqslant 1; i = 1, 2, \ldots, N, \\ H_1 &: p_{r_i|H_1}(R_i|H_1) = 2R_i \qquad\quad 0 \leqslant R_i \leqslant 1; i = 1, 2, \ldots, N. \end{aligned} \tag{2.432}$$

The densities on the two hypotheses are shown in Figure 2.43.

The log-likelihood ratio is

$$l_*(\mathbf{R}) = \ln \prod_{i=1}^{N} \frac{R_i}{1-R_i} = \sum_{i=1}^{N} \ln \frac{R_i}{1-R_i}. \tag{2.433}$$

We do not know how to evaluate $p_{l_*|H_0}(L|H_0)$ and $p_{l_*|H_1}(L|H_1)$, therefore we do not have analytical expressions for P_F and P_M. However we can find $\mu_i(s)$. Substituting (2.432) into (2.376), we have

$$e^{\mu_i(s)} = \int_0^1 p_{r_i|H_1}(R_i|H_1)^s p_{r_i|H_0}(R_i|H_0)^{1-s}\,dR_i = 2\int_0^1 R_i^s(1-R_i)^{1-s}dR_i. \tag{2.434}$$

[15]The Beta function can be computed in Matlab using the `beta` function. See Appendix A.

The integral can be evaluated using (2.431), which gives

$$e^{\mu_i(s)} = \frac{2\Gamma(s+1)\Gamma(2-s)}{\Gamma(3)} = \Gamma(s+1)\Gamma(2-s), \tag{2.435}$$

therefore,

$$\mu_i(s) = \ln\Gamma(s+1) + \ln\Gamma(2-s). \tag{2.436}$$

Differentiating yields

$$\dot{\mu}_i(s) = \frac{\Gamma'(s+1)}{\Gamma(s+1)} - \frac{\Gamma'(2-s)}{\Gamma(2-s)} = \psi_0(s+1) - \psi_0(2-s), \tag{2.437}$$

$$\ddot{\mu}_i(s) = \psi_0'(s+1) - \psi_0'(2-s) = \psi_1(s+1) + \psi_1(2-s), \tag{2.438}$$

where $\psi_0(\cdot)$ and $\psi_1(\cdot)$ are the Digamma and Trigamma functions defined as[16]

$$\psi_0(z) = \frac{\Gamma'(z)}{\Gamma(z)}, \tag{2.439}$$

$$\psi_1(z) = = \frac{d}{dz}\psi_0(z). \tag{2.440}$$

The samples are IID, so

$$\mu(s) = N\mu_i(s) = N\ln\Gamma(s+1) + N\ln\Gamma(2-s), \tag{2.441}$$

$$\dot{\mu}(s) = N\dot{\mu}_i(s) = N\psi_0(s+1) - N\psi_0(2-s), \tag{2.442}$$

$$\ddot{\mu}(s) = N\ddot{\mu}_i(s) = N\psi_1(s+1) + N\psi_1(2-s). \tag{2.443}$$

We want to run a simulation to estimate P_F and P_D and compare the results to the approximate expressions. We start by choosing a set of ROC points to simulate by choosing some values for s_* that satisfy $0 \leqslant s_* \leqslant 1$. We then find the corresponding threshold using

$$\gamma_* = \dot{\mu}(s_*) = N\psi_0(s_*+1) - N\psi_0(2-s_*). \tag{2.444}$$

We compute $P_F(\gamma_*)$, $P_M(\gamma_*)$, $I_F(\gamma_*; s_*)$, and $I_M(\gamma_*; s_*)$ using (2.383)–(2.386).

For the simulation, we choose $\mathbf{x} = \mathbf{r}$. The log-likelihood ratio is given in (2.433) and the tilted marginal density is found by substituting (2.432) into (2.389). The result is

$$\begin{aligned} p_{r_i|s_*}(R_i) &= \frac{1}{\Gamma(s_*+1)\Gamma(2-s_*)} 2^{s_*} R_i^{s_*} 2^{1-s_*}(1-R_i)^{1-s_*}, \quad 0 \leqslant R_i \leqslant 1, \\ &= \frac{1}{B(s_*+1, 2-s_*)} R_i^{s_*}(1-R_i)^{1-s_*}, \qquad 0 \leqslant R_i \leqslant 1, \end{aligned} \tag{2.445}$$

which is a Beta$(s_*+1, 2-s_*)$ probability density. In Figure 2.44, we show the tilted density for $s_* = 0.7$. We see that, although it is still a Beta probability density, it has a significantly different shape than the original densities.

The weighting functions are evaluated by substituting (2.433) and (2.441) into (2.390) and (2.391). We specify $\alpha = 0.1$ and $c = 3$ and calculate $K_{F,IS}$, $K_{M,IS}$, and K_{IS} using (2.392)–(2.394). We generate the random samples from the tilted marginal density[17] and compute the estimates $\hat{P}_F(\gamma_*)$ and $\hat{P}_D(\gamma_*)$ using (2.395) and (2.397). We plot the approximate ROC curve and the simulation estimates in Figure 2.45 for $N = 20$. The results show that the approximation is quite accurate for this example. ■

[16]The Digamma and Trigamma functions can be computed in Matlab using the `psi` function. See Appendix A.

[17]The random samples are generated using the Matlab function `betarnd`.

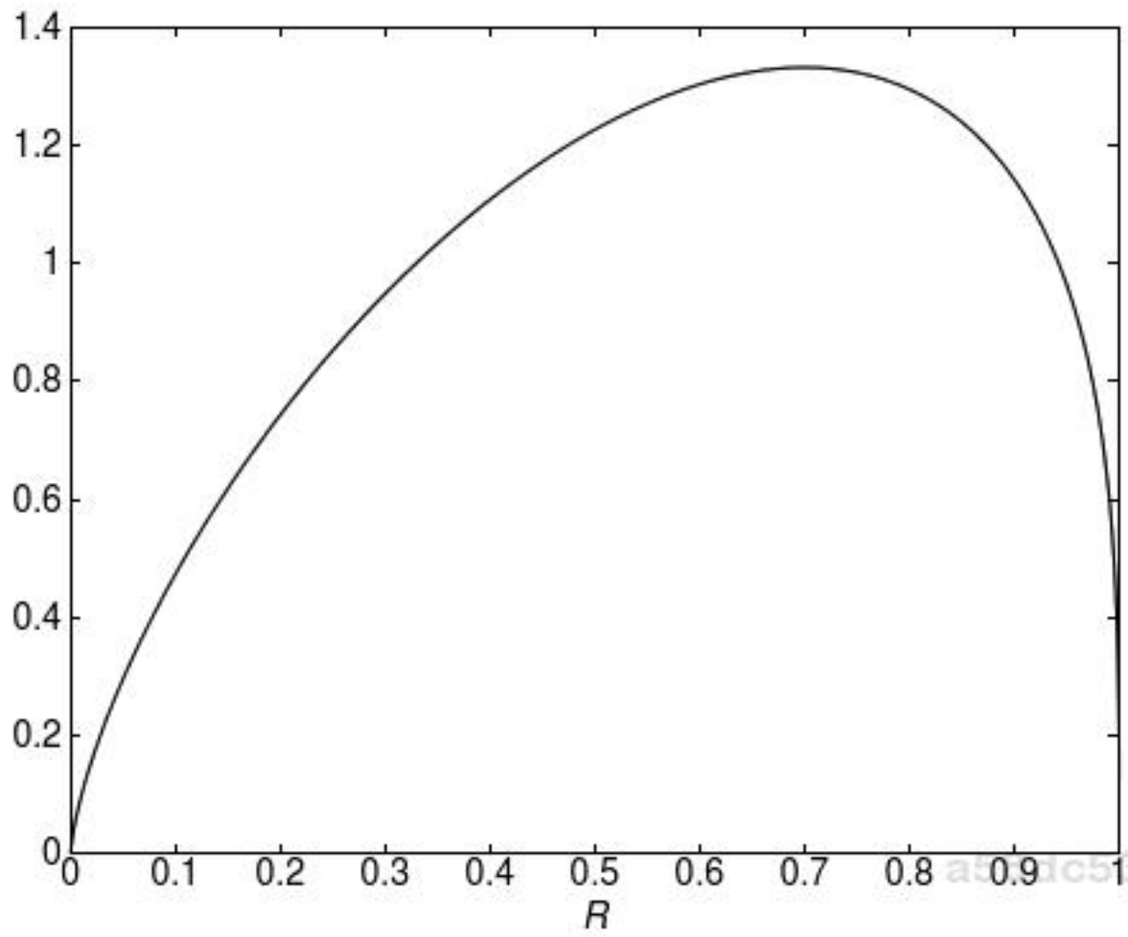


Figure 2.44: Tilted probability density for $s_* = 0.7$.

These three examples provide interesting results for the optimum "tilted" density. In Example 2.16, the original Gaussian density is translated so that its mean is $s_* m$ and its variance is unchanged. Example 2.17, the variance of the original Gaussian density is modified. This is equivalent to scaling the density. In neither case is the density actually tilted as shown in Figure 2.29. In Example 2.18, the parameters of the Beta density are modified so that another Beta density is obtained. The operation is not a scaling or shifting operation and "tilting" is a more appropriate description.

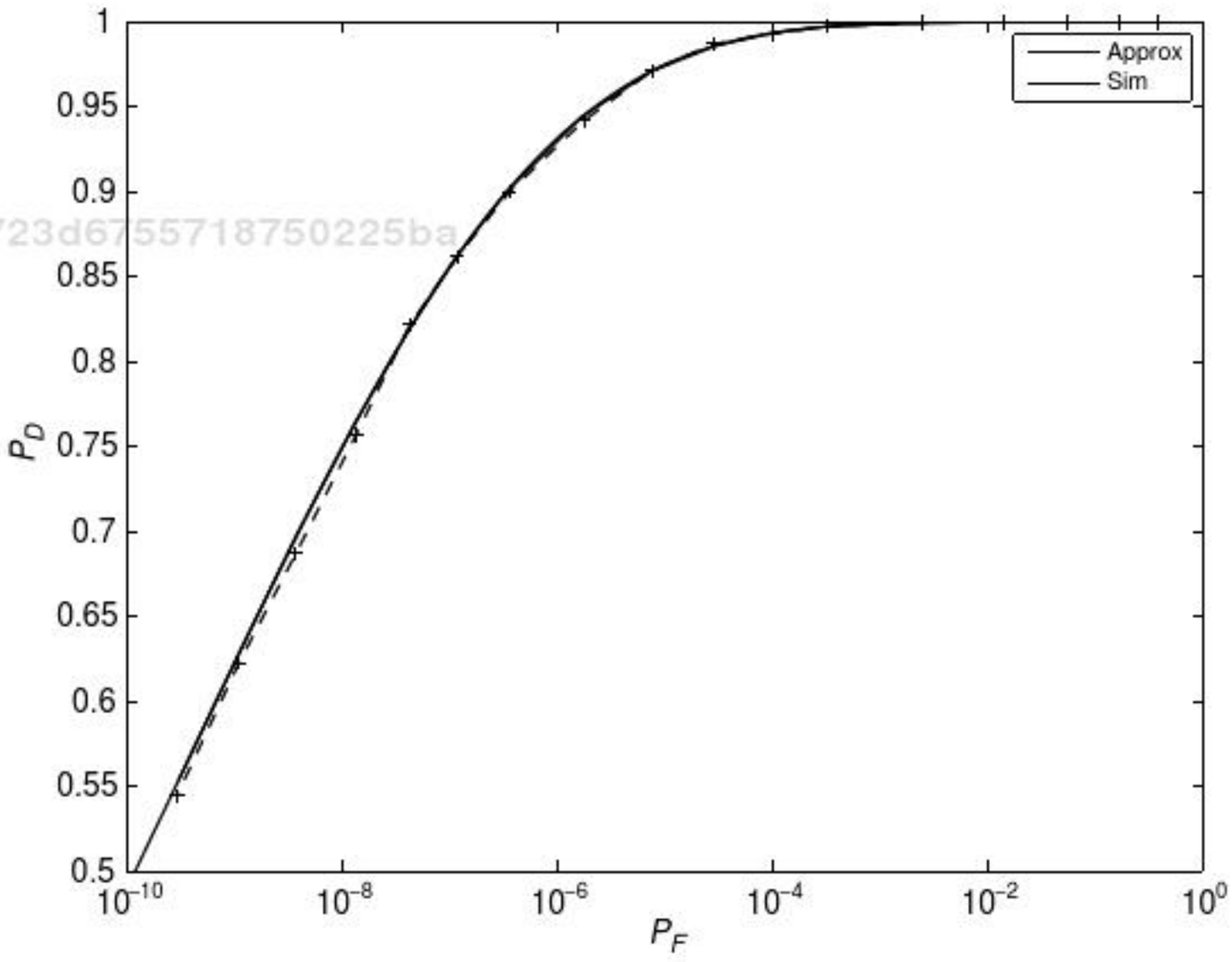


Figure 2.45: Approximate and simulated P_D versus P_F for $N = 20$.

Srinivasan [Sri02] and other references (e.g., [Mit81]) consider an approach that restricts $p_{x|0*}(X)$ to be a scaled transformation,

$$p_{x|0*}(X) = \frac{1}{a}\, p_{x|H_0}\left(\frac{X}{a}\right) \tag{2.446}$$

or a translated density

$$p_{x|0*}(X) = p_{x|H_0}(X - b) \tag{2.447}$$

and chooses a or b to minimize the variance. The disadvantage is that it is not clear which form is appropriate for a given problem. In addition, in many cases, the optimum tilted density will not be the same form as the original density.

Example 2.18 illustrated how simulation can be used to validate theoretical results. To make the comparison, we needed a representative set of points on the two ROC curves, but the specific values were not important. It turned out that the simulated P_F and P_D were quite close to the approximate values, however this is not always the case. If we are trying to simulate a particular value of P_F, such as when producing curves similar to those in Figures 2.41 and 2.42, the threshold γ_* corresponding to the approximate $P_F(\gamma_*)$ may produce an estimate $\hat{P}_F$ that is not within the specified confidence interval. We develop an iterative technique to handle this situation in the next section.

2.5.2.6 Iterative Importance Sampling

In this section, we consider the model in which we do not have an exact analytic expression relating P_F and γ_* and want to simulate a particular value of P_F, which we denote as P_F^*. We develop an iterative algorithm for solving this problem.[18]

We assume that we can find $\mu(s)$, either analytically or numerically, and can then find $\dot{\mu}(s)$ and $\ddot{\mu}(s)$. We use (2.383) to specify an approximate expression for P_F as a function of s,

$$P_F(s) \approx \left\{\exp\left[\mu(s) - s\dot{\mu}(s) + \frac{s^2}{2}\ddot{\mu}(s)\right]\right\} \operatorname{erfc}_*\left[s\sqrt{\ddot{\mu}(s)}\right]. \tag{2.448}$$

We solve (2.448) to find s (usually numerically) for the desired P_F^*, that is,

$$s : P_F(s) = P_F^*. \tag{2.449}$$

We then find the corresponding threshold using

$$\gamma = \dot{\mu}(s). \tag{2.450}$$

Note that we can only use this technique for values of P_F^* where $0 \leqslant s \leqslant 1$ or equivalently $\dot{\mu}(0) \leqslant \gamma \leqslant \dot{\mu}(1)$. We run the simulation and compute the estimate $\hat{P}_F$ using (2.395). If the estimate falls within our confidence interval, that is,

$$(1-\alpha)P_F^* \leqslant \hat{P}_F \leqslant (1+\alpha)P_F^* \tag{2.451}$$

then we accept the simulation results. If not, we adjust s and γ iteratively.

[18] Srinivasan [Sri02] (p. 55) also discusses this problem but our approach is significantly different. [SSG97] also discusses various adaptive techniques and gives references.

Our iterative procedure is a gradient approach based on a Taylor series expansion of $\ln P_F(s)$,

$$\ln P_F(s) \approx \ln P_F(s_0) + (s - s_0) \left. \frac{d \ln P_F(s)}{ds} \right|_{s=s_0}. \tag{2.452}$$

To evaluate the derivative in (2.452) we use the Chernoff bound in (2.217), which we denote as $\overline{P}_F(s)$,

$$\overline{P}_F(s) \triangleq e^{\mu(s) - s\dot{\mu}(s)}. \tag{2.453}$$

Although the Chernoff bound can be weak compared to the approximation $P_F(s)$, they have similar derivatives, that is,

$$\frac{d \ln \overline{P}_F(s)}{ds} \approx \frac{d \ln P_F(s)}{ds}. \tag{2.454}$$

Differentiating the log of (2.453) gives

$$\frac{d \ln \overline{P}_F(s)}{ds} = \frac{d}{ds}\left(\mu(s) - s\dot{\mu}(s)\right) = -s\ddot{\mu}(s). \tag{2.455}$$

Let $s^{(n)}$ and $\hat{P}_F^{(n)}$ denote the values of s and $\hat{P}_F$ at the nth iteration. In (2.452), we set $s_0 = s^{(n)}$, $s = s^{(n+1)}$, $P_F(s_0) = \hat{P}_F^{(n)}$, and $P_F(s) = P_F^*$. Then s can be updated as follows

$$s^{(n+1)} = s^{(n)} + \frac{\ln P_F^* - \ln \hat{P}_F^{(n)}}{s^{(n)}\ddot{\mu}\left(s^{(n)}\right)}, \tag{2.456}$$

and $\gamma^{(n+1)}$ is given by

$$\gamma^{(n+1)} = \dot{\mu}\left(s^{(n+1)}\right). \tag{2.457}$$

We iterate until convergence in $\hat{P}_F$ as specified in (2.451).

Note that, at each step in the iteration, the tilt in the marginal density changes per (2.456). Also note that we have used several approximations, so there is no guarantee that the algorithm will converge.

Example 2.19 (continuation of Example 2.18). We consider the same model as in Example 2.18. We specify $\alpha = 0.1$, $c = 2$, $N = 20$, and $P_F^* = 10^{-8}$. In Figure 2.46, we show the iteration for one trial. The simulated values are slightly less than the approximate values and it converges in three iterations. ■

Examples 2.18 and 2.19 are representative of the class of problems where we can find an analytic expression for $\mu(s)$ but we do not know how to find $p_{l_*|H_0}(L|H_0)$ and $p_{l_*|H_1}(L|H_1)$, and therefore cannot evaluate P_F and P_M. The approximate expressions used in Example 2.18 and the iterative algorithm used in Example 2.19 appear to provide efficient techniques for simulating this class of problems.

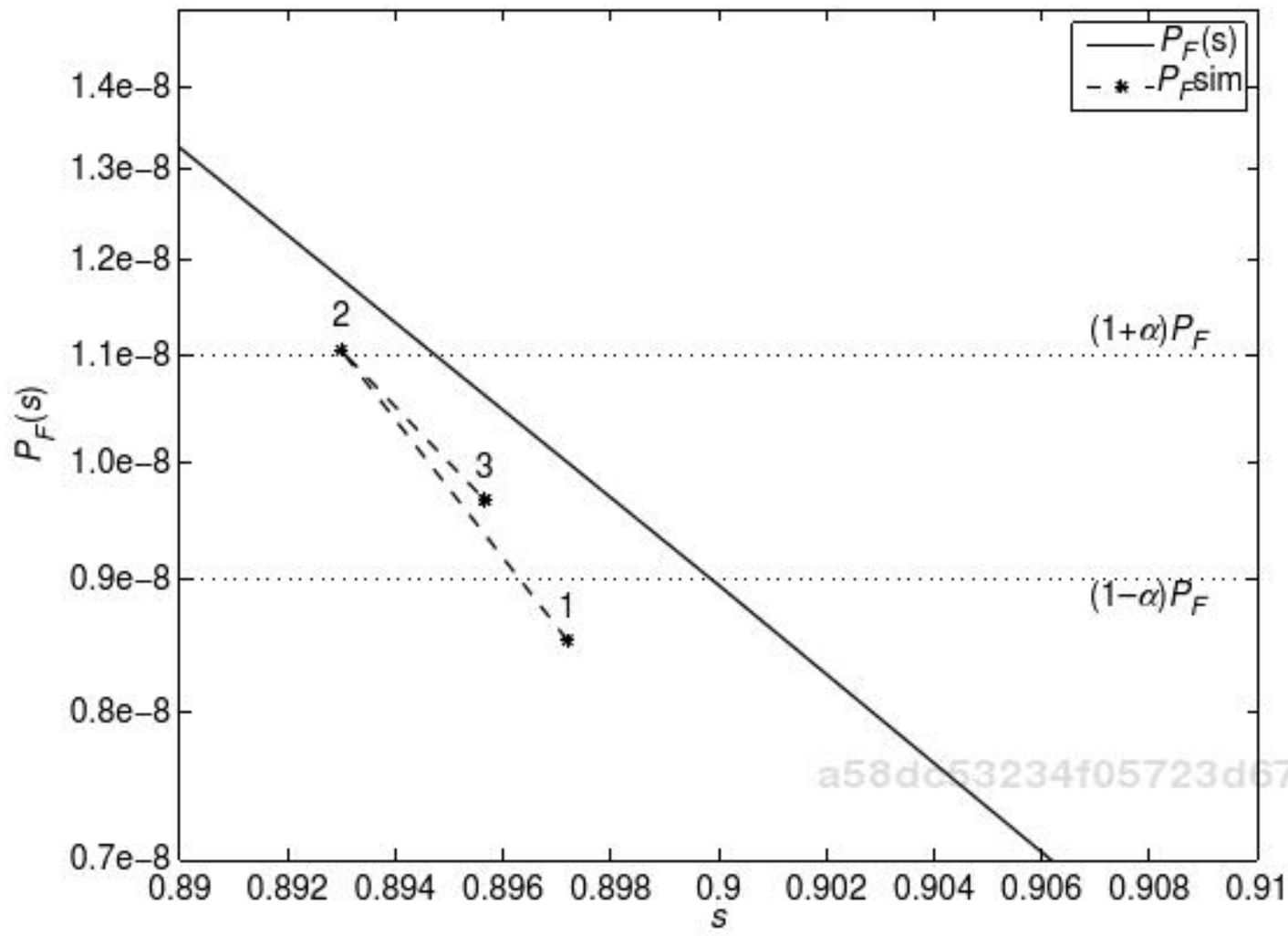


Figure 2.46: Convergence of iterative algorithm.

Example 2.14 in Section 2.4 is representative of the class of problems where we have to find $\mu(s)$ numerically. The approximate and iterative algorithms are also applicable to this class of problems (see Problem 2.5.6).

There are a number of references on adaptive importance sampling techniques. Srinivasan [Sri02] and Smith et al. [SSG97] have a discussion of various techniques and a list of references.

2.5.3 Summary

In this section, we have developed an approach to importance sampling that utilizes tilted densities to specify the probability density to be used for simulation. The technique is also referred to as exponential twisting or large deviation theory in the literature. Although we did not include the derivation, one can show that this approach is asymptotically efficient as $N \to \infty$ or $P_F \to 0$ (e.g., [SB90]).

By focusing on the log-likelihood ratio (which is optimal for our model), we were able to find $\mu(s)$ using the probability densities of a sufficient statistic $\mathbf{X}$ on H_0 and H_1. This enabled us to solve problems where $p_{l_*|H_0}(L|H_0)$ and $p_{l_*|H_1}(L|H_1)$ were difficult to find. This result was exceedingly useful when the components of $\mathbf{R}$ were statistically independent and we could tilt the marginal probability densities. This approach allowed us to achieve the required accuracy and confidence intervals with K_{IS} values that were typically lower by a factors of up to 10^{10} compared with classical Monte Carlo techniques.

We did not consider the simulation of suboptimal tests or tests that simply compared an arbitrary statistic to a threshold. All of the tilting ideas carry over, but the $\mu(s)$ relation in (2.375) no longer applies.

Our goal in this section was to provide an introduction to importance sampling that focused on the use of tilted densities. It provides adequate background to simulate most of the detection problems that we will encounter in the text.

2.6 SUMMARY

In this chapter, we have derived the essential detection theory results that provide the basis for much of our work in the remainder of the book.

We began our discussion in Section 2.2 by considering the simple binary hypothesis testing problem. There were several key results:

1. Using either a Bayes criterion or a Neyman–Pearson criterion, we find that the optimum test is a likelihood ratio test,

$$\Lambda(\mathbf{R}) = \frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

   Thus, regardless of the dimensionality of the observation space, the test consists of comparing a scalar variable $\Lambda(\mathbf{R})$ with a threshold.
2. In many cases, construction of the LRT can be simplified if we can identify a sufficient statistic. Geometrically, this statistic is just that coordinate in a suitable coordinate system that describes the observation space that contains *all* the information necessary to make a decision (see (2.74)–(2.76)).
3. A complete description of the LRT performance was obtained by plotting the conditional probabilities P_D and P_F as the threshold η was varied. The resulting ROC could be used to calculate the Bayes risk for any set of costs. In many cases, only one value of the threshold is of interest and a complete ROC is not necessary.

In Section 2.3, we introduced the M hypotheses problem. The key results were

1. The dimension of the decision space is no more than $M-1$. The boundaries of the decision regions are hyperplanes in the $(\Lambda_1, \ldots, \Lambda_{M-1})$ plane.
2. The optimum test is straightforward to find. From (2.156), we compute

$$\beta_i(\mathbf{R}) = \sum_{j=0}^{M-1} C_{ij} \Pr(H_j|\mathbf{R}), \quad i = 0, 1, \ldots, M-1,$$

   and choose the smallest. We shall find however, when we consider specific examples that the error probabilities are frequently difficult to compute.
3. A particular test of importance is the minimum total probability of error test. Here we compute the *a posteriori* probability of each hypothesis $\Pr(H_i|\mathbf{R})$ and choose the largest.

In Sections 2.2 and 2.3, we dealt primarily with problems in which we could derive the structure of the optimum test and obtain relatively simple analytic expressions for the receiver operating characteristic or the error probability. In Section 2.4, we developed bounds and approximate expressions for the error probabilities for the large group of problems where an exact solution is difficult. The key function in these results was the logarithm of the moment generating function of the likelihood ratio. From (2.204)

$$\mu(s) = \ln \int_{-\infty}^{\infty} \left[p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)\right]^s \left[p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\right]^{1-s} d\mathbf{R}.$$

The function $\mu(s)$ plays a central role in all of the bounds and approximate expressions that are derived in Section 2.4. It is straightforward to calculate when the components of $\mathbf{r}$ on the two hypotheses are statistically independent. Then,

$$\mu_i(s) = \ln \int_{-\infty}^{\infty} \left[p_{r_i|H_1}(R_i|H_1)\right]^s \left[p_{r_i|H_0}(R_i|H_0)\right]^{1-s} dR_i, \quad i = 1, \ldots, N,$$

and

$$\mu(s) = \sum_{i=1}^{N} \mu_i(s).$$

We have introduced $\mu(s)$ early in the text because of the central role it plays in the analysis of non-Gaussian models.

In many applications of interest, it is necessary to simulate the detection algorithm in order to evaluate the performance. In Section 2.5, we gave a brief introduction to Monte Carlo simulation. A key issue is the number of trials needed to have a desired level of confidence in the result. In most systems of interest, the desired P_F is very small (e.g., $P_F \leqslant 10^{-6}$ is frequently required). In these cases, the number of trials required to obtain a reasonable confidence level is prohibitively large. We introduced a technique called "importance sampling" that provided a dramatic reduction in the number of trials. The key function in our approach was $\mu(s)$ developed in Section 2.4.

In this chapter, we confined our discussion to the decision problem in which the transition probabilities $p_{\mathbf{r}|H_j}(\mathbf{R}|H_j)$ were known. This is referred to as *simple* hypothesis testing. In many applications, $p_{\mathbf{r}|H_j,\boldsymbol{\theta}}(\mathbf{R}|H_j, \boldsymbol{\theta})$ depends on an unknown vector parameter $\boldsymbol{\theta}$ that may be random or nonrandom. This is referred to as *composite* hypothesis testing and we study it in Chapter 4.

In this chapter, we have developed many of the key results in detection theory. All of our discussion dealt with arbitrary probability densities. A large number of important signal processing applications in communications, radar and sonar can be modeled assuming that $p_{\mathbf{r}|H_j}(\mathbf{R}|H_j)$, $j = 0, \ldots, M-1$ is a multivariate Gaussian probability density. In Chapter 3, we consider this model in detail.

2.7 PROBLEMS

The problems are divided into sections corresponding to the major sections in the chapter. For example, section P2.2 pertains to text material in Section 2.2. In sections in which it is appropriate the problems are divided into topical groups.

P2.2 Simple Binary Hypothesis Tests

SIMPLE BINARY TESTS

Problem 2.2.1. Consider the following binary hypothesis testing problem:

$$H_1 : r = s + n,$$
$$H_0 : r = n,$$

where s and n are independent random variables:

$$p_s(S) = \begin{cases} ae^{-aS} & S \geqslant 0 \\ 0 & S < 0 \end{cases}$$

and

$$p_n(N) = \begin{cases} be^{-bN} & N \geqslant 0 \\ 0 & N < 0. \end{cases}$$

1. Prove that the likelihood ratio test reduces to

$$R \underset{H_0}{\overset{H_1}{\gtrless}} \gamma.$$

2. Find γ for the optimum Bayes test as a function of the costs and *a priori* probabilities.
3. Now assume that we need a Neyman–Pearson test. Find γ as a function of P_F, where

$$P_F \triangleq \Pr(\text{say } H_1 | H_0 \text{ is true}).$$

Problem 2.2.2. The two hypotheses are

$$H_1 : p_r(R) = \frac{1}{2} \exp(-|R|),$$
$$H_0 : p_r(R) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} R^2\right).$$

(This is an example of the Generalized Gaussian density with $\alpha_0 = 1, b_0 = 1$, and $\alpha_1 = 2, b_1 = \sqrt{2}$.)

1. Find the likelihood ratio $\Lambda(R)$.
2. The test is

$$\Lambda(R) \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

Compute the decision regions for various values of η.

Problem 2.2.3. The random variable x is $N(0, \sigma^2)$. It is passed through one of two nonlinear transformations.

$$H_1 : y = x^2$$
$$H_0 : y = x^3.$$

Find the LRT.

Problem 2.2.4. The random variable x is $N(m, \sigma^2)$. It is passed through one of two nonlinear transformations.

$$H_1 : y = e^x,$$
$$H_0 : y = x^2.$$

Find the LRT.

Problem 2.2.5. Consider the following hypothesis testing problem. There are K independent observations:

$$H_1 : r_i \text{ is Gaussian}, N(0, \sigma_1^2) \quad i = 1, 2, \ldots, K$$
$$H_0 : r_i \text{ is Gaussian}, N(0, \sigma_0^2) \quad i = 1, 2, \ldots, K,$$

where $\sigma_0^2 < \sigma_1^2$.

1. Compute the likelihood ratio.
2. Assume that the threshold is η:

$$\Lambda(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

Show that a sufficient statistic is $l(\mathbf{R}) = \sum_{i=1}^{K} R_i^2$. Compute the threshold γ for the test

$$l(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

in terms of η, σ_0^2, and σ_1^2.

3. Define

$$P_F = \Pr(\text{choose } H_1 | H_0 \text{ is true}),$$
$$P_M = \Pr(\text{choose } H_0 | H_1 \text{ is true}).$$

Find an expression for P_F and P_M.

4. Plot the ROC for $K = 1$, $\sigma_1^2 = 2$, and $\sigma_0^2 = 1$.
5. What is the threshold for the minimax criterion when $C_M = C_F$ and $C_{00} = C_{11} = 0$?

Problem 2.2.6. The observation r is defined in the following manner:

$$H_1 : r = bm_1 + n,$$
$$H_0 : r = n,$$

where b and n are independent zero-mean Gaussian variables with variances σ_b^2 and σ_n^2, respectively.

1. Find the LRT and draw a block diagram of the optimum processor.
2. Draw the ROC.
3. Assume that the two hypotheses are equally likely. Use the criterion of minimum probability of error. What is the $\Pr(\epsilon)$?

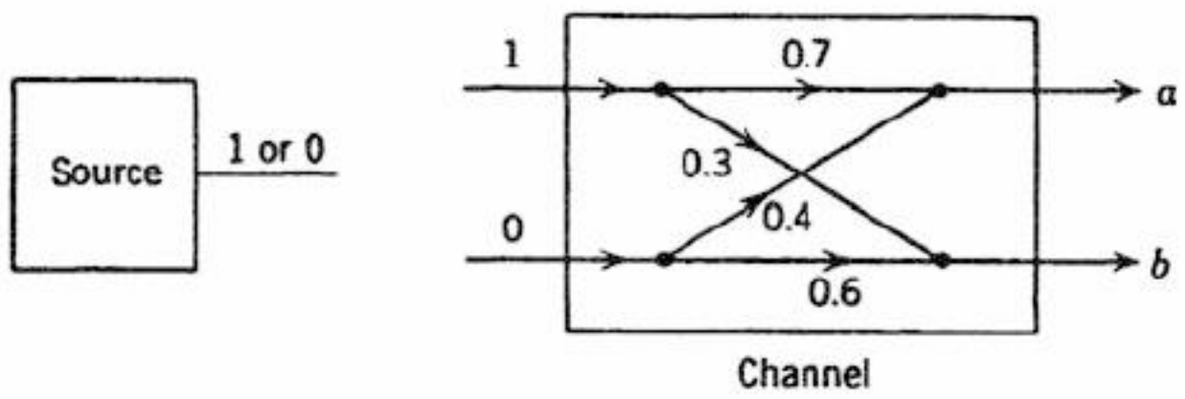


Figure P2.1: Binary communication channel.

Problem 2.2.7. One of two possible sources supplies the inputs to the simple communication channel as shown in Figure P2.1.

Both sources put out either 1 or 0. The numbers on the line are the channel transition probabilities; that is,

$$\Pr(a \text{ out}|\ 1 \text{ in}) = 0.7.$$

The source characteristics are

$$\begin{aligned} &\text{Source 1:} \quad \Pr(1) = 0.5 \quad \Pr(0) = 0.5 \\ &\text{Source 2:} \quad \Pr(1) = 0.6 \quad \Pr(0) = 0.4. \end{aligned}$$

To put the problem in familiar notation, define

(a) False alarm—say source 2 when source 1 is present;

(b) Detection—say source 2 when source 2 is present.

1. Compute the ROC of a test that maximizes P_D subject to the constraint that $P_F = \alpha$.
2. Describe the test procedure in detail for $\alpha = 0.25$.

Problem 2.2.8. The probability densities on the two hypotheses are

$$H_i : p_{x|H_i}(X|H_i) = \frac{1}{\pi\left[1 + (X - a_i)^2\right]} \quad i = 0, 1; -\infty < X < \infty.$$

where $a_0 = 0$ and $a_1 = 1$.

1. Find the LRT.
2. Plot the ROC.

Problem 2.2.9. Consider a simple coin tossing problem:

$$\begin{aligned} &H_1 : \text{heads are up: } \Pr[H_1] \triangleq P_1, \\ &H_0 : \text{tails are up: } \quad \Pr[H_0] \triangleq P_0 < P_1. \end{aligned}$$

N independent tosses of the coin are made. Show that the number of observed heads N_H is a sufficient statistic for making a decision between the two hypotheses.

Problem 2.2.10. A sample function of a simple Poisson counting process $N(t)$ is observed over the interval T:

$$H_1 : \text{The mean rate is } k_1 : \Pr(H_1) = \frac{1}{2},$$
$$H_0 : \text{The mean rate is } k_0 : \Pr(H_0) = \frac{1}{2}.$$

1. Prove that the number of events in the interval T is a "sufficient statistic" to choose hypothesis H_0 or H_1.
2. Assuming equal costs for the possible errors, derive the appropriate likelihood ratio test and the threshold.
3. Find an expression for the probability of error.

Problem 2.2.11. Let

$$y = \sum_{i=0}^{n} x_i,$$

where the x_i are statistically independent random variables with a Gaussian density $N(0, \sigma^2)$. The number of variables in the sum is a random variable with a Poisson distribution:

$$\Pr(n = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots.$$

We want to decide between the two hypotheses,

$$H_1 : n \leqslant 1$$
$$H_0 : n > 1.$$

Write an expression for the LRT.

Problem 2.2.12. Randomized tests. Our basic model of the decision problem in the text did not permit randomized decision rules. We can incorporate them by assuming that at each point $\mathbf{R}$ in Z we say H_1 with probability $\phi(\mathbf{R})$ and say H_0 with probability $1 - \phi(\mathbf{R})$. The model in the text is equivalent to setting $\phi(\mathbf{R}) = 1$ for all $\mathbf{R}$ in Z_1 and $\phi(\mathbf{R}) = 0$ for all $\mathbf{R}$ in Z_0.

1. We consider the Bayes criterion first. Write the risk for the above decision model.
2. Prove that an LRT minimizes the risk and a randomized test is *never* necessary.
3. Prove that the risk is constant over the interior of any straight-line segment on an ROC. Because straight-line segments are generated by randomized tests, this is an alternate proof of the result in Part 2.
4. Consider the Neyman–Pearson criterion. Prove that the optimum test always consists of either
   (*i*) an ordinary LRT with $P_F = \alpha$ or
   (*ii*) a probabilistic mixture of *two* ordinary likelihood ratio tests constructed as follows: Test 1: $\Lambda(\mathbf{R}) \overset{H_1}{\geqslant} \eta$ gives $P_F = \alpha^+$. Test 2: $\Lambda(\mathbf{R}) \overset{H_1}{>} \eta$ gives $P_F = \alpha^-$, where $[\alpha^-, \alpha^+]$ is the smallest interval containing α. $\phi(\mathbf{R})$ is 0 or 1 except for those $\mathbf{R}$ where $\phi(\mathbf{R}) = \eta$. (Find $\phi(\mathbf{R})$ for this set.)

MATHEMATICAL PROPERTIES

Problem 2.2.13. The random variable $\Lambda(\mathbf{R})$ is defined by (2.13) and has a different probability density on H_1 and H_0. Prove the following:

1. $E(\Lambda^n|H_1) = E(\Lambda^{n+1}|H_0)$.
2. $E(\Lambda|H_0) = 1$.
3. $E(\Lambda|H_1) - E(\Lambda|H_0) = \text{Var}(\Lambda|H_0)$.

Problem 2.2.14. Consider the random variable Λ. In (2.128)–(2.129), we showed that

$$p_{\Lambda|H_1}(X|H_1) = Xp_{\Lambda|H_0}(X|H_0).$$

1. Verify this relation by direct calculation of $p_{\Lambda|H_1}(\cdot)$ and $p_{\Lambda|H_0}(\cdot)$ for the densities in Examples 2.1 and 2.5.
2. We saw that the performance of the test in Example 2.5 was completely characterized by d^2. Show that

$$d^2 = \ln[1 + \text{Var}(\Lambda|H_0)].$$

Problem 2.2.15. The function $\text{erfc}_*(X)$ is defined in (2.83).

1. Integrate by parts to establish the bound

$$\frac{1}{\sqrt{2\pi}X}\left(1 - \frac{1}{X^2}\right)\exp\left(-\frac{X^2}{2}\right) < \text{erfc}_*(X) < \frac{1}{\sqrt{2\pi}X}\exp\left(-\frac{X^2}{2}\right) \quad X > 0.$$

2. Generalize part 1 to obtain the asymptotic series

$$\text{erfc}_*(X) = \frac{1}{\sqrt{2\pi}X}e^{-X^2/2}\left[1 + \sum_{m=1}^{n-1}(-1)^m\frac{1\cdot 3\cdots(2m-1)}{X^{2m}} + R_n\right].$$

The remainder is less than the magnitude of the $n+1$ term and is the same sign. *Hint*: Show that the remainder is

$$R_n = \left[(-1)^{n+1}\frac{1\cdot 3\cdots(2n-1)}{X^{2n+2}}\right]\theta,$$

where

$$\theta = \int_0^\infty e^{-t}\left(1 + \frac{2t}{X^2}\right)^{-n-\frac{1}{2}}dt < 1.$$

3. Assume that $X = 3$. Calculate a simple bound on the *percentage* error when $\text{erfc}_*(3)$ is approximated by the first n terms in the asymptotic series. Evaluate this percentage error for $n = 2, 3, 4$ and compare the results. Repeat for $X = 5$.

Problem 2.2.16.

1. Prove

$$\text{erfc}_*(X) < \frac{1}{2}\exp\left(-\frac{X^2}{2}\right) \quad X > 0.$$

Hint: Show

$$[\text{erfc}_*(X)]^2 = \Pr(x \geqslant X, y \geqslant Y) < \Pr\left(x^2 + y^2 \geqslant 2X^2\right),$$

where x and y are independent zero-mean Gaussian variables with unit variance.

2. For what values of X is this bound better than (2.89)?

Higher Dimensional Decision Regions

A simple binary test can always be reduced to a one-dimensional decision region. In many cases, the results are easier to interpret in two or three dimensions. Some typical examples are illustrated in this section.

Problem 2.2.17. The joint probability density of the random variables x_1 and x_2 on H_1 and H_0 is

$$H_1 : p_{x_1,x_2|H_1}(X_1, X_2|H_1) = \frac{1}{4\pi\sigma_1\sigma_0}\left[\exp\left(-\frac{X_1^2}{2\sigma_1^2} - \frac{X_2^2}{2\sigma_0^2}\right) + \exp\left(-\frac{X_1^2}{2\sigma_0^2} - \frac{X_2^2}{2\sigma_1^2}\right)\right],$$

$$H_0 : p_{x_1,x_2|H_0}(X_1, X_2|H_0) = \frac{1}{2\pi\sigma_0^2}\exp\left(-\frac{X_1^2}{2\sigma_0^2} - \frac{X_2^2}{2\sigma_0^2}\right),$$

where $-\infty < X_1, X_2 < \infty$.

1. Find the LRT.
2. Write an exact expression for P_D and P_F. Upper and lower bound P_D and P_F by modifying the region of integration in the exact expression.

Problem 2.2.18. The joint probability density of the random variables $x_i; i = 1, 2, \ldots, M$ on H_1 and H_0 is

$$H_1 : p_{\mathbf{x}|H_1}(\mathbf{X}|H_1) = \sum_{k=1}^{M} p_k \frac{1}{(2\pi\sigma^2)^{M/2}}\exp\left[-\frac{(X_k - m)^2}{2\sigma^2}\right]\prod_{i\neq k}^{M}\exp\left(-\frac{X_i^2}{2\sigma^2}\right),$$

$$H_0 : p_{\mathbf{x}|H_0}(\mathbf{X}|H_0) = \prod_{i=1}^{M}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{X_i^2}{2\sigma^2}\right),$$

where $-\infty < X_i < \infty$, and

$$\sum_{k=1}^{M} p_k = 1.$$

1. Find the LRT.
2. Draw the decision regions for various values of η in the X_1, X_2-plane for the special case in which $M = 2$ and $p_1 = p_2 = \frac{1}{2}$.
3. Find an upper and lower bound to P_F and P_D by modifying the regions of integration.

Problem 2.2.19. The probability density of r_i on the two hypotheses is

$$p_{r_i|H_k}(R_i|H_k) = \frac{1}{\sqrt{2\pi}\sigma_k}\exp\left[-\frac{(R_i - m_k)^2}{2\sigma_k^2}\right] \qquad \begin{matrix} i = 1, 2, \ldots, N \\ k = 0, 1. \end{matrix}$$

The observations are independent.

1. Find the LRT. Express the test in terms of the following quantities:

$$l_\alpha = \sum_{i=1}^{N} R_i,$$

$$l_\beta = \sum_{i=1}^{N} R_i^2.$$

2. Draw the decision regions in the l_α, l_β-plane for the case in which

$$2m_0 = m_1 > 0$$

$$2\sigma_1 = \sigma_0.$$

Problem 2.2.20 (continuation of Problem 2.2.19).

1. Consider the special case

$$m_0 = 0,$$

$$\sigma_0 = \sigma_1.$$

   Draw the decision regions and compute the ROC.
2. Consider the special case

$$m_0 = m_1 = 0,$$

$$\sigma_1^2 = \sigma_s^2 + \sigma_n^2,$$

$$\sigma_0 = \sigma_n.$$

Draw the decision regions.

Problem 2.2.21. A shell is fired at one of two targets: Under H_1 the point of aim has coordinates x_1, y_1, z_1; under H_0 it has coordinates x_0, y_0, z_0. The distance of the actual landing point from the point of aim is a zero-mean Gaussian variable, $N\left(0, \sigma^2\right)$, in each coordinate. The variables are independent. We wish to observe the point of impact and guess which hypothesis is true.

1. Formulate this as a hypothesis testing problem and compute the likelihood ratio. What is the simplest sufficient statistic? Is the ROC in Figure 2.12*a* applicable? If so, what value of d^2 do we use?
2. Now include the effect of time. Under H_k the desired explosion time is $t_k; k = 1, 2$. The distribution of the actual explosion time is

$$p_{\tau|H_k}(\tau) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{(\tau - t_k)^2}{2\sigma_t^2}\right) \quad \begin{array}{l} -\infty < \tau < \infty \\ k = 1, 2. \end{array}$$

   Find the LRT and compute the ROC.

Problem 2.2.22. Consider the model in Examples 2.2 and 2.6.

1. Plot P_M versus N for $P_F = 10^{-6}$.
2. Define N_{opt} as the value of N that gives the minimum P_M for a given

$$\text{SNR}_T \triangleq N\sigma_s^2/\sigma_n^2.$$

   Plot N_{opt} versus SNR_T for $P_F = 10^{-2}, 10^{-4}$, and 10^{-6}.
3. Plot $P_M(N_{\text{opt}})$ versus SNR_T for $P_F = 10^{-2}, 10^{-4}$, and 10^{-6}.

IID MODEL

The following problems assume the IID model in (2.39)–(2.43) and Figure 2.7.

Problem 2.2.23. Consider the Generalized Gaussian model in Example 2.4. Plot $\ln \Lambda(R_i)$ for $\alpha = 1, 1.1, 1.2, \ldots, 1.9$.

Problem 2.2.24. [Kay98] The observations on the two hypotheses are

$$\begin{aligned} H_1 &: r_i = m + n_i \quad i = 1, \ldots, N \\ H_0 &: r_i = n_i \qquad\quad i = 1, \ldots, N, \end{aligned}$$

where $p_{n_i}(N_i)$ is Cauchy

$$p_{n_i}(N_i) = \frac{1}{\pi(1 + N_i^2)} \quad -\infty < N_i < \infty.$$

Plot $\ln \Lambda(R_i)$.

Problem 2.2.25. The observations on the two hypotheses are

$$\begin{aligned} H_1 &: r_i = m + n_i, \quad i = 1, \ldots, N, \\ H_0 &: r_i = n_i, \qquad\quad i = 1, \ldots, N, \end{aligned}$$

where n_i is a sample from a Gaussian mixture density

$$p_{n_i}(N_i) = \alpha \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\frac{N_i^2}{\sigma_1^2}\right) + (1-\alpha)\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\frac{N_i^2}{\sigma_2^2}\right)$$

for $0 \leqslant \alpha \leqslant 1$.

(*a*) Assume $m = 1, \sigma_1^2 = 1$, and three values of $\sigma_2^2 : 2, 10$, and 20. Plot $\ln \Lambda(R_i)$ for $\alpha = 0, 0.1, \ldots, 1.0$.

(*b*) Fix $\alpha = 0.5, m = 1, \sigma_1^2 = 1$. Plot $\ln \Lambda(R_i)$ for various σ_2^2/σ_1^2.

Problem 2.2.26. The observations on the two hypotheses are:

$$\begin{aligned} H_1 &: r_i = m + n_i, \quad i = 1, \ldots, N, \\ H_0 &: r_i = n_i \qquad\quad i = 1, \ldots, N. \end{aligned}$$

The noise n_i is a sum of two statistically independent noise terms

$$n_i = w_i + x_i, \quad i = 1, \ldots, N,$$

where w_i has a Weibull density, defined in (2.274),

$$p_{w_i}(W_i) = \frac{\alpha}{b}\left(\frac{W_i}{b}\right)^{\alpha-1} e^{-(W_i/b)^\alpha}, \quad W_i \geqslant 0; i = 1, 2, \ldots, N,$$

and x_i has a Rayleigh density

$$p_{x_i}(X_i) = \frac{X_i}{\sigma_x^2} \exp\left(-\frac{X_i^2}{2\sigma_x^2}\right), \quad X_i \geqslant 0; i = 1, 2, \ldots, N.$$

The n_i are statistically independent.

Notice that the Weibull density corresponds to the Rayleigh density when $\alpha = 2$ and $b = \sqrt{2}\sigma_x$. The variance of w_i is given by (2.276). We will encounter this model later when we sample the output of a bandpass square-law detector.

(*a*) Assume $\alpha = b = 3$. The resulting density is shown in Figure 2.35. Assume $m/\sigma_x^2 = 10$. Find $p_{n_i}(N_i)$ by numerically convolving $p_{r_i}(R_i)$ and $p_{x_i}(X_i)$ for three variance ratios, $\frac{\sigma_w^2}{\sigma_x^2} = 0.1, 1.0, 10.0$.

(*b*) Plot $\ln \Lambda(R_i)$ for the values in part (a).

Problem 2.2.27. Consider the Generalized Gaussian model in Example 2.4. Plot $\ln \Lambda(R_i)$ for $\alpha < 1$. Explain your results.

Problem 2.2.28. Extend the results in (2.42)–(2.43) where the observations are vectors $\mathbf{R}_i$ that are IID. Explain your results.

P2.3 *M* Hypotheses

Problem 2.3.1.

1. Verify that the M hypothesis Bayes test always leads to a decision space whose dimension is less than or equal to $M - 1$.
2. Assume that the coordinates of the decision space are

$$\Lambda_k(\mathbf{R}) \triangleq \frac{p_{\mathbf{r}|H_k}(\mathbf{R}|H_k)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)} \quad k = 1, 2, \ldots, M-1.$$

Verify that the decision boundaries are hyperplanes.

Problem 2.3.2. The observed random variable r has a Gaussian density on the three hypotheses,

$$p_{r|H_k}(R|H_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{(R-m_k)^2}{2{\sigma_k}^2}\right] \quad \begin{array}{l} -\infty < R < \infty \\ k = 1, 2, 3, \end{array}$$

where the parameter values on the three hypotheses are,

$$\begin{aligned} H_1 &: m_1 = 0, \quad \sigma_1^2 = \sigma_a^2, \\ H_2 &: m_2 = m, \quad \sigma_2^2 = \sigma_a^2 \quad (m > 0), \\ H_3 &: m_3 = 0, \quad \sigma_3^2 = \sigma_\beta^2 \quad (\sigma_\beta^2 > \sigma_a^2). \end{aligned}$$

The three hypotheses are equally likely and the criterion is minimum $\Pr(\epsilon)$.

1. Find the optimum Bayes test.
2. Draw the decision regions on the R-axis for the special case,

$$\begin{aligned} \sigma_\beta^2 &= 2\sigma_a^2, \\ \sigma_a^2 &= m^2. \end{aligned}$$

3. Compute the $\Pr(\epsilon)$ for this special case.

Problem 2.3.3. The probability density of $\mathbf{r}$ on the three hypotheses is

$$p_{r_1,r_2|H_k}(R_1, R_2|H_k) = (2\pi\sigma_{1k}\sigma_{2k})^{-1} \exp\left[-\frac{1}{2}\left(\frac{R_1^2}{\sigma_{1k}^2} + \frac{R_2^2}{\sigma_{2k}^2}\right)\right] \quad \begin{matrix} -\infty < R_1, R_2 < \infty \\ k = 1, 2, 3, \end{matrix}$$

where

$$\begin{aligned} \sigma_{11}^2 &= \sigma_n^2, & \sigma_{21}^2 &= \sigma_n^2, \\ \sigma_{12}^2 &= \sigma_s^2 + \sigma_n^2, & \sigma_{22}^2 &= \sigma_n^2, \\ \sigma_{13}^2 &= \sigma_n^2, & \sigma_{23}^2 &= \sigma_s^2 + \sigma_n^2. \end{aligned}$$

The cost matrix is

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & \alpha \\ 1 & \alpha & 0 \end{bmatrix},$$

where $0 \leqslant \alpha < 1$ and $\Pr(H_2) = \Pr(H_3) \triangleq p$. Define $l_1 = R_1^2$ and $l_2 = R_2^2$.

1. Find the optimum test and indicate the decision regions in the l_1, l_2-plane.
2. Write an expression for the error probabilities. (Do not evaluate the integrals.)
3. Verify that for $\alpha = 0$ this problem reduces to Problem 2.2.17.

Problem 2.3.4. On H_k the observation is the value of a Poisson random variable

$$\Pr(r = n) = \frac{k_m^n}{n!} e^{-k_m}, \quad m = 1, 2, \ldots, M,$$

where $k_m = mk$. The hypotheses are equally likely and the criterion is minimum $\Pr(\epsilon)$.

1. Find the optimum test.
2. Find a simple expression for the boundaries of the decision regions and indicate how you would compute the $\Pr(\epsilon)$.

Problem 2.3.5. Assume that the received vector on each of the three hypotheses is

$$\begin{aligned} H_0 &: \mathbf{r} = \mathbf{m}_0 + \mathbf{n}, \\ H_1 &: \mathbf{r} = \mathbf{m}_1 + \mathbf{n}, \\ H_2 &: \mathbf{r} = \mathbf{m}_2 + \mathbf{n}, \end{aligned}$$

where

$$\mathbf{r} \triangleq \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \quad \mathbf{m}_i \triangleq \begin{bmatrix} m_{i1} \\ m_{i2} \\ m_{i3} \end{bmatrix} \quad \mathbf{n} \triangleq \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}.$$

The $\mathbf{m}_i$ are known vectors, and the components of $\mathbf{n}$ are statistically independent, zero-mean Gaussian random variables with variance σ^2.

1. Using the results in the text, express the Bayes test in terms of two sufficient statistics:

$$l_1 = \sum_{i=1}^{3} c_i r_i,$$

$$l_2 = \sum_{i=1}^{3} d_i r_i.$$

   Find explicit expressions for c_i and d_i. Is the solution unique?
2. Sketch the decision regions in the l_1, l_2-plane for the particular cost assignment:

$$C_{00} = C_{11} = C_{22} = 0,$$

$$C_{12} = C_{21} = C_{01} = C_{10} = \frac{1}{2}C_{02} = \frac{1}{2}C_{20} > 0.$$

P2.4 Performance Bounds and Approximations

Problem 2.4.1. Consider the binary test with N independent observations r_i, where

$$p_{r_i|H_k}(R_i|H_k) = N\left(m_k, \sigma_k^2\right) \qquad \begin{matrix} k = 0, 1 \\ i = 1, 2, \ldots, N. \end{matrix}$$

Find $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$.

Problem 2.4.2 (continuation of Problem 2.4.1). Consider the special case in which

$$m_0 = 0,$$

$$\sigma_0^2 = \sigma_n^2,$$

and

$$\sigma_1^2 = \sigma_s^2 + \sigma_n^2.$$

1. Find $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$.
2. Assuming equally likely hypotheses, find an upper bound on the minimum $\Pr(\epsilon)$.
3. With the assumption in part 2, find an approximate expression for the $\Pr(\epsilon)$ that is valid for large N.

Problem 2.4.3. We derived the Chernoff bound in (2.217) by using tilted densities. This approach prepared us for the central limit theorem argument in the second part of our discussion. If we are interested only in (2.217), a much simpler derivation is possible.

1. Consider a function of the random variable x that we denote as $f(x)$. Assume

$$f(x) \geqslant 0, \qquad \text{all } x$$

$$f(x) \geqslant f(X_0) > 0, \qquad \text{all } x \geqslant X_0.$$

Prove

$$\Pr[x \geqslant X_0] \leqslant \frac{E[f(x)]}{f(X_0)}.$$

2. Now let

$$f(x) = e^{sx}, \quad s \geqslant 0,$$

and

$$X_0 = \gamma.$$

Use the result in part 1 to derive (2.212). What restrictions on γ are needed to obtain (2.217)?

Problem 2.4.4. The reason for using tilted densities and Chernoff bounds is that a straightforward application of the central limit theorem gives misleading results when the region of interest is on the tail of the density. A trivial example taken from [WJ65] illustrates this point.

Consider a set of statistically independent random variables x_i that assume values 0 and 1 with equal probability. We are interested in the probability

$$\Pr\left[y_N = \frac{1}{N}\sum_{i=1}^{N} x_i \geqslant 1\right] \triangleq \Pr[A_N].$$

(*a*) Define a standardized variable

$$z \triangleq \frac{y_N - \bar{y}_N}{\sigma_{yN}}.$$

Use a central limit theorem argument to estimate $\Pr[A_N]$. Denote this estimate as $\widehat{\Pr}[A_N]$.

(*b*) Calculate $\Pr[A_N]$ exactly.

(*c*) Verify that the fractional error is

$$\frac{\widehat{\Pr}[A_N]}{\Pr[A_N]} \propto e^{0.19N}$$

Observe that the fractional error grows exponentially with N.

(*d*) Estimate $\Pr[A_N]$ using the Chernoff bound of Problem 2.4.6. Denote this estimate as $\Pr_c[A_N]$. Compute $\frac{\Pr_c[A_N]}{\Pr[A_N]}$.

Problem 2.4.5.

(*a*) Find $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$ for the model in Problem 2.2.1.

(*b*) Plot an approximate ROC.

Problem 2.4.6.

(*a*) Find $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$ for the model in Problem 2.2.23.

(*b*) Plot an approximate ROC.

Problem 2.4.7.

(*a*) Find $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$ for the model in Problem 2.2.24.

(*b*) Plot an approximate ROC.

P2.5 Monte Carlo Simulation

Problem 2.5.1. This problem is a continuation of Example 2.16. The observations on the two hypotheses are:

$$H_1 : r_i \sim N\left(m_{1_i}, \sigma^2\right), \quad i = 1, 2, \ldots, N,$$
$$H_0 : r_i \sim N\left(m_{0_i}, \sigma^2\right), \quad i = 1, 2, \ldots, N,$$

where

$$m_{1_i} = m(1 + 0.2i),$$
$$m_{0_i} = m(1 - 0.2i).$$

Simulate the optimum detector for $N = 40, m = 1$, and $\sigma^2 = 1$ and compare your results to the analytic solution.

Problem 2.5.2. This problem is a continuation of Example 2.16. The observations on the two hypotheses are:

$$H_1 : r_i \sim N\left(m, \sigma_i^2\right), \quad i = 1, 2, \ldots, N,$$
$$H_0 : r_i \sim N\left(0, \sigma_i^2\right), \quad i = 1, 2, \ldots, N,$$

where

$$\sigma_i^2 = \sigma^2(1 + 0.05i) \quad i = 1, 2, \ldots, N.$$

Simulate the optimum detector for $N = 40, m = 1$, and $\sigma^2 = 1$ and compare your results to the analytic solution.

Problem 2.5.3. This problem is a continuation of Example 2.17. The observations on the two hypotheses are:

$$H_1 : r_i \sim N\left(0, \sigma_{s_i}^2 + \sigma_n^2\right), \quad i = 1, 2, \ldots, N,$$
$$H_0 : r_i \sim N\left(0, \sigma_n^2\right), \quad i = 1, 2, \ldots, N,$$

where

$$\sigma_{s_i}^2 = \sigma_s^2(1 + 0.05i) \quad i = 1, 2, \ldots, N$$

Simulate the optimum detector and generate a curve similar to Figure 2.42.

Problem 2.5.4. This problem is a continuation of Examples 2.16 and 2.17. The observations on the two hypotheses are:

$$H_1 : r_i \sim N\left(m, \sigma_s^2 + \sigma_n^2\right), \quad i = 1, 2, \ldots, N,$$
$$H_0 : r_i \sim N\left(0, \sigma_s^2\right), \quad i = 1, 2, \ldots, N.$$

(*a*) Find the LRT.

(*b*) Simulate the optimum detector for $N = 40$ and $m = 1$. Plot a figure similar to Figure 2.42 versus σ_s^2/σ_n^2.

Problem 2.5.5. This problem is a continuation of Example 2.14.

(*a*) Find the LRT for the model in Example 2.14.

(*b*) Simulate the optimum detector and compare your results to the approximate ROC computed in Example 2.14.

Problem 2.5.6. This problem is a continuation of Example 2.18.

(*a*) Generalize the results in (2.436)–(2.438) to other values of a_j and b_j; $j = 0, 1$.

(*b*) Plot the results corresponding to Figures 2.43–2.45.

Problem 2.5.7. Simulate the model developed in Problems 2.2.23 and 2.4.6.

Problem 2.5.8. Simulate the model developed in Problems 2.2.24 and 2.4.7.