1 Introduction and motivation to detection and estimation

1.1 Introduction

The second half of the twentieth century experienced an explosive growth in information technology, including data transmission, processing, and computation. This trend will continue at an even faster pace in the twenty-first century. Radios and televisions started in the 1920s and 1940s respectively, and involved transmission from a single transmitter to multiple receivers using AM and FM modulations. Baseband analog telephony, starting in the 1900s, was originally suited only for local area person-to-person communication. It became possible to have long-distance communication after using cascades of regeneration repeaters based on digital PCM modulation. Various digital modulations with and without coding, across microwave, satellite, and optical fiber links, allowed the explosive transmissions of data around the world starting in the 1950s–1960s. The emergence of Ethernet, local area net, and, finally, the World Wide Web in the 1980s–1990s allowed almost unlimited communication from any computer to another computer. In the first decade of the twenty-first century, by using wireless communication technology, we have achieved cellular telephony and instant/personal data services for humans, and ubiquitous data collection and transmission using ad hoc and sensor networks. By using cable, optical fibers, and direct satellite communications, real-time on-demand wideband data services in offices and homes are feasible.

Detection and estimation theories presented in this book constitute some of the most basic statistical and optimization methodologies used in communication/telecommunication, signal processing, and radar theory and systems. The purpose of this book is to introduce these basic concepts and their applications to readers with only basic junior/senior year linear system and probability knowledge. The modest probability prerequisites are summarized in Section 2.1. Other necessary random processes needed to understand the material in the rest of this book are also presented succinctly in Sections 2.2–2.4.

The author (KY) has taught a first-year detection and estimation graduate course at UCLA for many years. Given the university's location in southern California, our students have very diverse backgrounds. There are students who have been working for some years in various local aerospace, communications, and signal processing industries, and may have already encountered in their work various concepts in detection and estimation. They may already have quite good intuitions on many of the issues encountered in this course and may be highly motivated to learn these concepts in greater depth. On the other

hand, most students (domestic and foreign) in this course have just finished a BS degree in engineering or applied science with little or no prior engineering experience and have not encountered real-life practical information processing systems and technologies. Many of these students may feel the topics covered in the course to be just some applied statistical problems and have no understanding about why one would want to tackle such problems.

The detection problem is one of deciding at the receiver which bit of information, which for simplicity at this point can be assumed to be binary, having either a "one" or a "zero," was sent by the transmitter. In the absence of noise/disturbance, the decision can be made with no error. However, in the presence of noise, we want to maximize the probability of making a correct decision and minimize the probability of making an incorrect decision. It turns out the solution of this statistical problem is based on statistical hypothesis theory already formulated in statistics in the 1930s by Fisher [1] and Neyman–Pearson [2]. However, it was only during World War II, in the analysis and design of optimum radar and sonar systems, that a statistical approach to these problems was formulated by Woodward [3]. Of course, we also want to consider decisions for multiple hypotheses under more general conditions.

The parameter estimation problem is one of determining the value of a parameter in a communication system. For example, in a modern cellular telephony system, the base station as well as a hand-held mobile phone need to estimate the power of the received signal in order to control the power of the transmitted signal. Parameter estimation can be performed using many methods. The simplest one is based on the mean-square estimation criterion, which had its origin in the 1940s by Kolmogorov [4] and Wiener [5]. The related least-square-error criterion estimation method was formulated by Gauss [6] and Laplace [7] in the nineteenth century. Indeed, Galileo even formulated the least-absolute-error criterion estimation in the seventeenth century [8]. All of these estimation methods of Galileo, Gauss, and Laplace were motivated by practical astronomical tracking of various heavenly bodies.

The purpose of this course is to teach some basic statistical and associated optimization methods mainly directed toward the analysis, design, and implementation of modern communication systems. In network jargon, the topics we encounter in this book all belong to the physical layer problems. These methods are equally useful for the study of modern control, system identification, signal/image/speech processing, radar systems, mechanical systems, economic systems, and biological systems. In this course, we will encounter the issue of the modeling of a system of interest. In simple problems, this modeling may be sort of obvious. In complicated problems, the proper modeling of the problem may be its most challenging aspect. Once a model has been formulated, then we can consider the appropriate mathematical tools for the correct and computationally efficient solution of the modeled problem. Often, among many possible solutions, we may seek the theoretically optimized solution based on some analytically tractable criterion. After obtaining this optimum solution, we may consider some solutions that are only slightly sub-optimum but are practical from an engineering point of view (e.g., ease of implementation; low cost of implementation; etc.). If there is not a known optimum solution, how can we determine how good some ad hoc as well as practical solutions are?

In Chapter 1, our purpose is to provide some very simple motivational examples illustrating the concepts of: statistical decision of two possible hypotheses; correlation receiver; relationship of the receiver's detector signal-to-noise ratio (SNR) to the transmitter power, and deterministic and statistical estimations. These examples will relate various simply posed hypothetical problems and human-made and physical phenomena to their possible solutions based on statistical methods. In turn, these methods are useful for characterizing, modeling, analyzing, and designing various engineering problems discussed in the rest of the book.

We will use the notation of denoting a deterministic scalar variable or parameter by a lowercase letter such as z. A deterministic column vector of dimension $M \times 1$ will be denoted by a bold lowercase letter such as $\mathbf{z}$. A scalar random variable (r.v.) will be denoted by an uppercase letter like Z, while a vector random vector will be denoted by a bold uppercase letter like $\mathbf{Z}$. The realizations of the r.v. Z and the random vector $\mathbf{Z}$, being non-random (i.e., deterministic), will be denoted by their corresponding lowercase letters of z and $\mathbf{z}$ respectively. We will use the abbreviation for the "left-hand side" by "l.h.s." and the "right-hand side" by "r.h.s" of an equation. We will also use the abbreviation of "with respect to" by "w.r.t."

At the end of each chapter, for each section, we provide some casual historical background information and references to other relevant journals and books. An asterisk following the section title in a chapter indicates that section may be of interest only to some serious readers. Materials in those sections will not be needed for following chapters. Similarly, an asterisk following an example indicates these materials are provided for the serious readers.

1.2 A simple binary decision problem

In the first motivational example, we consider a simple and intuitively obvious example illustrating the concept of maximum-likelihood (ML) decision. It turns out the ML criterion is the basis of many modern detection, decoding, and estimation procedures. Consider two boxes denoted as Box 0 and Box 1. Each box contains ten objects colored either red (R) or black (B). Suppose we know the "prior distributions" of the objects in the two boxes as shown in (1.1) and in Fig. 1.1.

$$\text{Box 0} \begin{cases} 2 \text{ reds} \\ 8 \text{ blacks} \end{cases}; \quad \text{Box 1} \begin{cases} 8 \text{ reds} \\ 2 \text{ blacks} \end{cases}. \tag{1.1}$$

Furthermore, we define the random variable (r.v.) X by

$$X = \begin{cases} -1, & \text{for a red object,} \\ 1, & \text{for a black object.} \end{cases} \tag{1.2}$$

From these prior probabilities, the conditional probabilities $p_0(x) = p(x|\text{Box 0})$ and $p_1(x) = p(x|\text{Box 1})$ of the two boxes are described in Fig. 1.1. For an observed value of x (taking either 1 or -1), these conditional probabilities are called "likelihood" functions. Now, suppose we randomly (with equal probability) pick one object from one

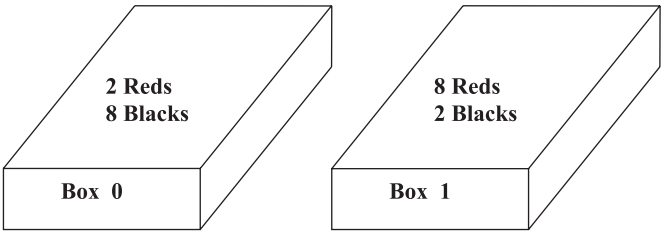

Figure 1.1 Known prior distributions of red and black objects in two boxes.

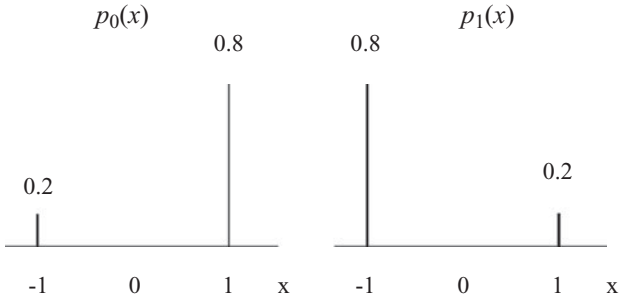

Figure 1.2 Conditional probabilities of $p_0(x)$ and $p_1(x)$.

of the boxes (whose identity is not known to us) and obtain a red object with the r.v. X taking the value of $x = -1$. Then by comparing the likelihood functions, we notice

$$p(x = -1|\text{Box } 0) = 0.2 < 0.8 = p(x = -1|\text{Box } 1). \tag{1.3}$$

After having observed a red object with $x = -1$, we declare Box 1 as the "most likely" box that the observed object was selected from. Thus, (1.3) illustrates the use of the ML decision rule. By looking at Fig. 1.1, it is "intuitively obvious" if a selected object is red, then statistically that object is more likely to come from Box 1 than from Box 0. A reasonably alert ten-year-old child might have come to that decision. On the other hand, if we observed a black object with $x = 1$, then by comparing their likelihood functions, we have

$$p(x = 1|\text{Box } 0) = 0.8 > 0.2 = p(x = 1|\text{Box } 1). \tag{1.4}$$

Thus, in this case with $x = 1$, we declare Box 0 as the "most likely" box that the observed object was selected from, again using the ML decision rule. In either case, we can form the likelihood ratio (LR) function

$$\frac{p_1(x)}{p_0(x)} \begin{cases} > 1 \Rightarrow \text{Declare Box 1,} \\ < 1 \Rightarrow \text{Declare Box 0.} \end{cases} \tag{1.5}$$

Then (1.3) and (1.4) are special cases of (1.5), where the LR function $p_1(x)/p_0(x)$ is compared to the threshold constant of 1. Specifically, for $x = -1$, $p_1(x = -1)/p_0(x = -1) > 1$, we decide for Box 1. For $x = 1$, $p_1(x = 1)/p_0(x = 1) < 1$, we decide for

Box 0. The decision procedure based on (1.5) is called the LR test. It turns out that we will repeatedly use the concepts of ML and LR tests in detection theory in Chapter 3.

Even though the decision based on the ML criterion (or equivalently the LR test) for the above binary decision problem is statistically reasonable, that does not mean the decision can not result in errors. Consider the evaluation of the probability of a decision error, given we have observed a red object with $x = -1$, is denoted by $P_{e|x=-1}$ in (1.6a):

$$\begin{aligned} P_{e|x=-1} &= \text{Prob(Decision error}|x = -1) && \text{(1.6a)} \\ &= \text{Prob(Decide Box 0}|x = -1) && \text{(1.6b)} \\ &= \text{Prob(Object came from Box 0}|x = -1) && \text{(1.6c)} \\ &= 2/10 = 1/5 = 0.2. && \text{(1.6d)} \end{aligned}$$

For an observed $x = -1$ the ML criterion makes the decision for Box 1. Thus, the "Decision error" in (1.6a) is equivalent to "Decide Box 0" in (1.6b). But the "Object came from Box 0 given $x = -1$" in (1.6c) means it is a red object from Box 0. The probability of a red object from Box 0 has the relative frequency of 2 over 10, which yields (1.6d). Similarly, the probability of a decision error, given we have observed a black object with $x = 1$, is denoted by $P_{e|x=1}$ in (1.7a):

$$\begin{aligned} P_{e|x=1} &= \text{Prob(Decision error}|x = 1) && \text{(1.7a)} \\ &= \text{Prob(Decide Box 1}|x = 1) && \text{(1.7b)} \\ &= \text{Prob(Object came from Box 1}|x = -1) && \text{(1.7c)} \\ &= 2/10 = 1/5 = 0.2. && \text{(1.7d)} \end{aligned}$$

Due to the symmetry of the number of red versus black objects in Box 0 with the number of black versus red objects in Box 1, it is not surprising that $P_{e|x=-1} = P_{e|x=1}$.

Next, suppose we want the average probability of a decision error P_e. Since the total number of red objects equals the total number of black objects, then $\text{P}(x = -1) = \text{P}(x = 1) = 1/2 = 0.5$. Thus,

$$P_e = P_{e|x=-1} \times 0.5 + P_{e|x=1} \times 0.5 = 1/5 = 0.2. \quad \text{(1.8)}$$

Equation (1.8) shows on the average the above ML decision rule results in an error 20% of the time. However, suppose we are told there are a total of ten black objects and ten red objects in both boxes, but are not given the prior probability information of (1.1) and Fig. 1.1. Then we can not use the ML decision criterion and the LR test. Given there are a total of ten black objects and ten red objects in both boxes, then regardless of whether the observed object is red or black, we should declare either Box 0 or Box 1 half of the time (by flipping a fair coin and deciding Box 0 if the coin shows a "Head" and Box 1 if the coin shows a "Tail"). In this random equi-probable decision rule, on the average an error occurs 50% of the time. This shows that by knowing the additional conditional probability information of (1.1) and using the ML decision criterion (or the LR test), we can achieve on average a smaller probability of error. This simple example shows the usefulness of statistical decision theory.

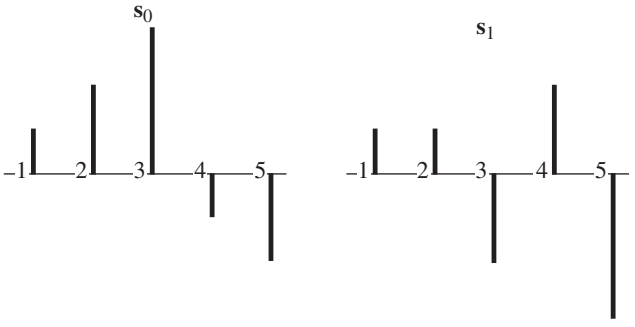

Figure 1.3 Two digital data vectors $\mathbf{s}_0$ and $\mathbf{s}_1$ of length five.

When the two boxes are originally chosen with equal probability in selecting an object, we saw the simple intuitive solution and the LR test solution are identical. However, if the two boxes are not chosen with equal probability, then after an observation the proper decision of the box is less intuitively obvious. As will be shown again in Chapter 3, for arbitrary probability distribution of the two boxes (i.e., hypotheses), the LR test can still be used to perform a statistically optimum decision.

1.3 A simple correlation receiver

In the second motivational example, consider two known digital signal data of length five denoted by the column vectors

$$\mathbf{s}_0 = [1,\ 2,\ 3,\ -1,\ -2]^T,\ \mathbf{s}_1 = [1,\ 1,\ -2,\ 2,\ -3]^T \tag{1.9}$$

as shown in Fig. 1.3. Define the *correlation* of the $\mathbf{s}_i$ vector with the $\mathbf{s}_j$ vector by

$$\mathbf{s}_i^T \mathbf{s}_j = \sum_{k=1}^{5} s_i(k) s_j(k),\ i, j = 0,\ 1. \tag{1.10}$$

Since $\mathbf{s}_i$ is a 5×1 column vector, and $\mathbf{s}_i^T$ is a 1×5 row vector, thus the correlation of the $\mathbf{s}_i$ vector with the $\mathbf{s}_j$ vector, $\mathbf{s}_i^T \mathbf{s}_j$, is a $1 \times 5 \times 5 \times 1 =$ one-dimensional scalar number. In particular, if $i = j$, then with each component of $\mathbf{s}_i$ having the unit of a volt, $\mathbf{s}_i^T \mathbf{s}_i = \|\mathbf{s}_i\|^2$ can be considered as the power of the $\mathbf{s}_i$ vector summed over four equally spaced time intervals. Thus, $\|\mathbf{s}_i\|^2$ represents the energy of the vector $\mathbf{s}_i$. By direct evaluation, we show both vectors have the same energy of

$$\begin{aligned} \|\mathbf{s}_0\|^2 &= 1^2 + 2^2 + 3^2 + (-1)^2 + (-2)^2 = 19 \\ &= 1^2 + 1^2 + (-2)^2 + (-2)^2 + (-3)^2 = \|\mathbf{s}_1\|^2. \end{aligned} \tag{1.11}$$

However, the correlation of $\mathbf{s}_1$ with $\mathbf{s}_0$ by direct evaluation is given by

$$\begin{aligned} \mathbf{s}_1^T \mathbf{s}_0 &= (1 \times 1) + (1 \times 2) + (-2 \times 3) + (-2 \times (-1)) + (-3 \times (-2)) \\ &= 1. \end{aligned} \tag{1.12}$$

Furthermore, suppose the observed vector $\mathbf{x}$ is either $\mathbf{s}_0$ or $\mathbf{s}_1$. That is, we have the model relating the observed data vector $\mathbf{x}$ to the signal vectors $\mathbf{s}_i$ given by

$$\mathbf{x} = \begin{cases} \mathbf{s}_0, \text{ under hypothesis } H_0 \\ \mathbf{s}_1, \text{ under hypothesis } H_1, \end{cases} \tag{1.13}$$

under the two hypotheses of H_0 and H_1. A simple decision rule to determine which vector or hypothesis is valid is to consider the correlation of $\mathbf{s}_1$ with the observed vector $\mathbf{x}$ given by

$$\gamma = \mathbf{s}_1^T \mathbf{x} = \begin{cases} \mathbf{s}_1^T \mathbf{s}_0 = 1 \Leftrightarrow \quad \mathbf{x} = \mathbf{s}_0 \Rightarrow \text{Declare hypothesis } H_0 \\ \mathbf{s}_1^T \mathbf{s}_1 = 19 \Leftrightarrow \mathbf{x} = \mathbf{s}_1 \Rightarrow \text{Declare hypothesis } H_1. \end{cases} \tag{1.14}$$

From (1.14), we note if the correlation value is low (i.e., 1), then we know the observed $\mathbf{x} = \mathbf{s}_0$ (or hypothesis H_0 is true), while if the correlation value is high (i.e., 19), then the observed $\mathbf{x} = \mathbf{s}_1$ (or hypothesis H_1 is true). Of course, in practice, the more realistic additive noise (AN) observation model replacing (1.13) may be generalized to

$$\mathbf{X} = \begin{cases} \mathbf{s}_0 + \mathbf{N}, \text{ under hypothesis } H_0 \\ \mathbf{s}_1 + \mathbf{N}, \text{ under hypothesis } H_1, \end{cases} \tag{1.15}$$

where $\mathbf{N}$ denotes the observation noise vector with some statistical properties. The introduction of an AN model permits the modeling of realistic physical communication or measurement channels with noises. Fortunately, an AN channel still allows the use of many statistical methods for analysis and synthesis of the receiver.

As we will show in Chapter 2, if $\mathbf{N}$ is a white Gaussian noise vector, then for the model of (1.15), the decision procedure based on the correlation method of (1.14) is still optimum, except the threshold values of 1 under hypothesis H_0 and 19 under hypothesis H_1 need to be modified under different statistical criteria (with details to be given in Chapter 3). We hope in the statistical analysis of complex systems, as the noise approaches zero, the operations of the noise-free systems may give us some intuitive hint on the optimum general solutions. The fact that the optimum binary receiver decision rule of (1.14) based on the correlation of $\mathbf{s}_1$ with the received vector $\mathbf{x}$ in the noise-free model of (1.13) is still optimum for the white Gaussian AN model of (1.15) is both interesting and satisfying.

1.4 Importance of SNR and geometry of the signal vectors in detection theory

One of the most important measures of the "goodness" of a waveform, whether in the analog domain (with continuous-time values and continuous-amplitude values), or in the digital data domain (after sampling in time and quantization in amplitude) as used in most communication, radar, and signal processing systems, is its signal-to-noise ratio (SNR) value. In order to make this concept explicit, consider the correlation receiver of Section 1.3 with the additive noise channel model of (1.15). In order to show quantitative results with explicit SNR values, we also need to impose an explicit statistical property on the noise vector $\mathbf{N}$ and also assume the two $n \times 1$ signal column vectors $\mathbf{s}_0$ and $\mathbf{s}_1$ have

equal energy $\|\mathbf{s}_0\|^2 = \|\mathbf{s}_1\|^2 = E$. The simplest (and fortunately still quite justifiable) assumption of the white Gaussian noise (WGN) property of $\mathbf{N} = [N_1, N_2, \ldots, N_n]^T$ being a column vector of dimension $n \times 1$ having zero mean and a covariance matrix of

$$\mathbf{\Lambda} = \mathrm{E}\{\mathbf{N}\mathbf{N}^T\} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n, \tag{1.16}$$

where σ^2 is the variance of each noise component N_k along the diagonal of $\mathbf{\Lambda}$, and having zero values in its non-diagonal values. Then the SNR of the signal $\mathbf{s}_0$ or $\mathbf{s}_1$ to the noise $\mathbf{N}$ is defined as the ratio of the energy E of $\mathbf{s}_0$ or $\mathbf{s}_1$ to the trace of $\mathbf{\Lambda}$ of (1.16) defined by

$$\mathrm{trace}(\mathbf{\Lambda}) = \sum_{k=1}^{n} \Lambda_{kk} = n\sigma^2. \tag{1.17}$$

Thus,

$$\mathrm{SNR} = \frac{E}{\mathrm{trace}(\mathbf{\Lambda})} = \frac{E}{n\sigma^2}. \tag{1.18}$$

Due to its possible large dynamic range, SNR is often expressed in the logarithmic form in units of dB defined by

$$\mathrm{SNR(dB)} = 10\log_{10}(\mathrm{SNR}) = 10\log_{10}(E/(n\sigma^2)). \tag{1.19}$$

Now, consider the binary detection problem of Section 1.3, where the two 5×1 signal column vectors $\mathbf{s}_0$ and $\mathbf{s}_1$ defined by (1.9) have equal energies of $\|\mathbf{s}_0\|^2 = \|\mathbf{s}_1\|^2 = E = 19$, and the AN channel of (1.15) has a WGN vector $\mathbf{N}$ of zero mean and covariance matrix given by (1.16). Thus, its SNR $= 19/(5\sigma^2)$.

Denote the correlation of $\mathbf{s}_1$ with $\mathbf{X}$ by $\Gamma = \mathbf{s}_1^T\mathbf{X}$. Under the noise-free channel model of (1.12), $\Gamma = \mathbf{s}_1^T\mathbf{x} = \mathbf{s}_1^T\mathbf{s}_0 = 1$ for hypothesis H_0 and $\Gamma = \mathbf{s}_1^T\mathbf{x} = \mathbf{s}_1^T\mathbf{s}_1 = 19$ for hypothesis H_1. But under the present AN channel of model (1.15), $\Gamma = \mathbf{s}_1^T\mathbf{X}$ under the two hypotheses yields two r.v.'s given by

$$\Gamma = \mathbf{s}_1^T\mathbf{X} = \begin{cases} \mathbf{s}_1^T(\mathbf{s}_0 + \mathbf{N}) \\ \mathbf{s}_1^T(\mathbf{s}_1 + \mathbf{N}) \end{cases} = \begin{cases} \mu_0 + \mathbf{s}_1^T\mathbf{N}, \text{ under hypothesis } H_0 \\ \mu_1 + \mathbf{s}_1^T\mathbf{N}, \text{ under hypothesis } H_1 \end{cases}, \tag{1.20}$$

where $\mu_0 = 1$ and $\mu_1 = 19$. Denote the correlation of $\mathbf{s}_1$ with $\mathbf{N}$ by $\tilde{\Gamma} = \mathbf{s}_1^T\mathbf{N}$, which is a Gaussian r.v. consisting of a sum of Gaussian r.v.'s of $\{N_1, \ldots, N_n\}$. Since all the r.v.'s of $\{N_1, \ldots, N_n\}$ have zero means, then $\tilde{\Gamma}$ has a zero mean. Then the variance of $\tilde{\Gamma}$ is given by

$$\sigma_{\tilde{\Gamma}}^2 = \mathrm{E}\{(\mathbf{s}_1^T\mathbf{N})(\mathbf{s}_1^T\mathbf{N})^T\} = \mathrm{E}\{\mathbf{s}_1^T\mathbf{N}\mathbf{N}^T\mathbf{s}_1\} = \sigma^2\mathbf{s}_1^T\mathbf{s}_1 = 19\sigma^2. \tag{1.21}$$

Using (1.20) and (1.21), we note $\Gamma = \mathbf{s}_1^T\mathbf{X}$ is a Gaussian r.v. of mean 1 and variance $19\sigma^2$ under hypothesis H_0, while it is a Gaussian r.v. of mean 19 and also variance $19\sigma^2$ under hypothesis H_1. The Gaussian pdfs of Γ under the two hypotheses are plotted in Fig. 1.4. Thus, for the additive white Gaussian noise (AWGN) channel, instead of using the decision rule of (1.14) for the noise-free case, under the assumption that both

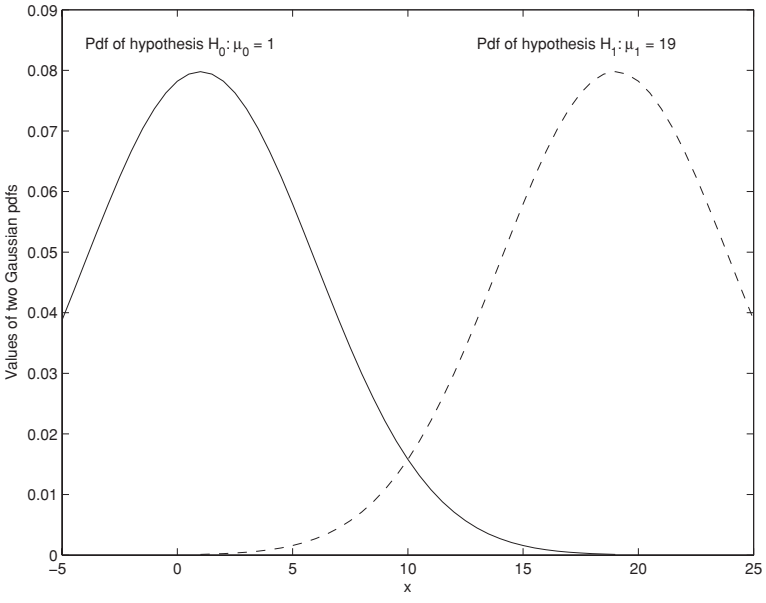

Figure 1.4 Plots of Gaussian pdfs under H_0 of $\mu_0 = 1$ and under H_1 of $\mu_1 = 19$.

hypotheses occur equi-probably and are of equal importance, then we should set the decision threshold at 10, which is the average of 1 and 19. Thus, the new decision rule is given by

$$\Gamma = \mathbf{s}_1^T \mathbf{X} \begin{cases} < 10 \Rightarrow \text{Declare hypothesis } H_0 \\ > 10 \Rightarrow \text{Declare hypothesis } H_1 \end{cases}. \tag{1.22}$$

From (1.20) and Fig. 1.4, under hypothesis H_0, the decision rule in (1.22) states if the noise is such that $\mathbf{s}_1^T \mathbf{X} = (1 + \mathbf{s}_1^T \mathbf{N}) < 10$, then there is no decision error. But if the noise drives $\mathbf{s}_1^T \mathbf{X} = (1 + \mathbf{s}_1^T \mathbf{N}) > 10$, then the probability of an error is given by the area of the solid Gaussian pdf curve under H_0 to the right of the threshold value of 10.

Thus, the probability of an error under H_0, $P_{e|H_0}$, is given by

$$\begin{aligned} P_{e|H_0} &= \int_{10}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\Gamma} e^{-\frac{(\gamma-1)^2}{2\sigma_\Gamma^2}} d\gamma = \int_{\frac{10-1}{\sigma_\Gamma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt \\ &= \int_{\frac{9}{\sqrt{19}\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt = \int_{\frac{9\sqrt{5\text{SNR}}}{19}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt \\ &= \text{Q}\left(\left((9/19)\sqrt{5}\right)\sqrt{\text{SNR}}\right) = \text{Q}\left(1.059\sqrt{\text{SNR}}\right). \end{aligned} \tag{1.23}$$

The second integral on the r.h.s. of (1.23) follows from the first integral by the change of variable $t = (\gamma - 1)/\sigma_\Gamma$, the third integral follows from the second integral by using the variance $\sigma_\Gamma^2 = 19\sigma^2$, the fourth integral follows from the third integral by using $\text{SNR} = 19/(5\sigma^2)$. Finally, the last expression of (1.23) follows from the definition of a complementary Gaussian distribution function of the zero mean and unit variance

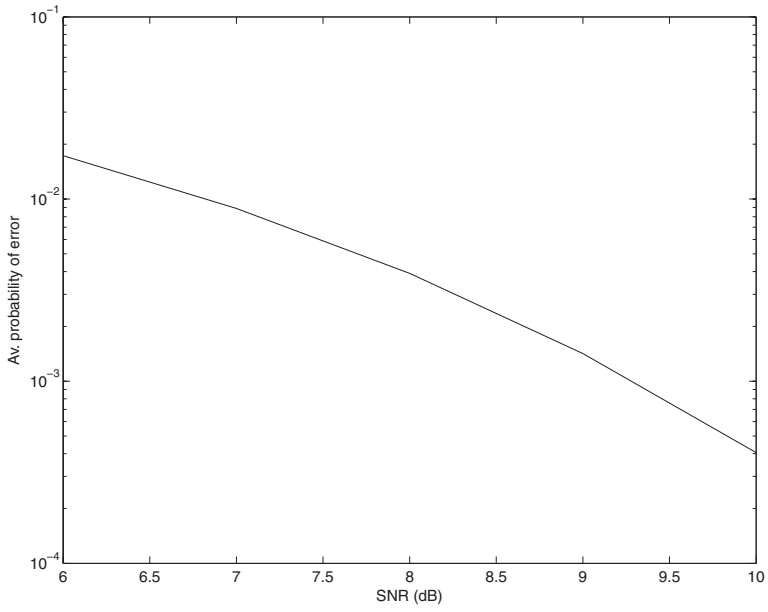

Figure 1.5 Plot of the average probability of error P_e vs. SNR(dB) for detection of $\mathbf{s}_0$ and $\mathbf{s}_1$ of (1.9) in WGN.

Gaussian r.v. defined by

$$\mathrm{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt. \tag{1.24}$$

By symmetry, under hypothesis H_1, the probability of an error is given by the area of the dashed Gaussian pdf curve under H_1 to the left of the threshold value of 10. Thus, the probability of an error under H_1, $P_{e|H_1}$, is given by

$$\begin{aligned} P_{e|H_1} &= \int_{-\infty}^{10} \frac{1}{\sqrt{2\pi}\sigma_\Gamma} e^{-\frac{(\gamma-19)^2}{2\sigma_r}} d\gamma = \int_{-\infty}^{\frac{10-19}{\sigma\Gamma}} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt \\ &= \int_{-\infty}^{\frac{-9}{\sqrt{19}\sigma}} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt = \int_{-\infty}^{\frac{-9\sqrt{5\mathrm{SNR}}}{19}} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt \\ &= \int_{\frac{9\sqrt{5\mathrm{SNR}}}{19}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt = \mathrm{Q}\left(1.059\sqrt{\mathrm{SNR}}\right). \end{aligned} \tag{1.25}$$

If hypothesis H_0 and hypothesis H_1 are equi-probable, then the average probability of an error P_e is given by

$$P_e = 1/2 \cdot P_{e|H_0} + 1/2 \cdot P_{e|H_1} = \mathrm{Q}\left(1.059\sqrt{\mathrm{SNR}}\right). \tag{1.26}$$

A plot of the average probability of error P_e of (1.25) is given in Fig. 1.5. Since the Q(·) function in (1.25) is a monotonically decreasing function, the average probability of error decreases as SNR increases as seen in Fig. 1.5.