2 Bayesian Hypothesis Testing

In a wide range of applications, one must make decisions based on a set of observations. Examples include medical diagnosis, voice and face recognition, DNA sequence analysis, air traffic control, and digital communication. In general, the observations are noisy, incomplete, or otherwise imperfect, and thus the decisions produced will not always be correct. However, we would like to use a decision process that is as good as possible in an appropriate sense.

Addressing such problems is the aim of decision theory, and a natural framework for setting up such problems is in terms of a hypothesis test. In this framework, each of the possible scenarios corresponds to a hypothesis. When there are M hypotheses, we denote the set of possible hypotheses using $\mathcal{H} = \{H_0, H_1, \ldots, H_{M-1}\}$.[1] For each of the possible hypotheses, there is a different model for the observed data, and this is what we will exploit to distinguish among the hypotheses.

In our formulation, the observed collection of data is represented as a random vector y, which may be discrete- or continuous-valued. There are a variety of ways to model the hypotheses. In this section, we follow what is referred to as the *Bayesian* approach, and model the valid hypothesis as a (discrete-valued) random variable, and thus we denote it using H.

In a Bayesian hypothesis testing problem, the complete model therefore consists of the *a priori* probabilities

$$p_{\mathsf{H}}(H_m), \qquad m = 0, 1, \ldots, M-1,$$

together with a characterization of the observed data under each hypothesis, which takes the form of the conditional probability distributions[2]

$$p_{\mathsf{y}|\mathsf{H}}(\cdot|H_m), \qquad m = 0, 1, \ldots, M-1. \tag{1}$$

Of course, a complete characterization of our knowledge of the correct hypothesis based on our observations is the set of *a posteriori* probabilities

$$p_{\mathsf{H}|\mathsf{y}}(H_m|\mathbf{y}), \qquad m = 0, 1, \ldots, M-1. \tag{2}$$

The distribution of possible values of H is often referred to as our *belief* about the hypothesis. From this perspective, we can view the *a priori* probabilities as our prior belief, and view (2) as the revision of our belief based on having observed the

[1] Note that H_0 is sometimes referred to as the "null" hypothesis, particularly in asymmetric problems where it has special significance.

[2] As related terminology, the function $p_{\mathsf{y}|\mathsf{H}}(\mathbf{y}|\cdot)$, where $\mathbf{y}$ is the actual observed data, is referred to as the *likelihood function*.

data $\mathbf{y}$. The belief update is, of course, computed from the particular data $\mathbf{y}$ based on the model via Bayes' Rule:[3]

$$p_{H|\mathbf{y}}(H_m|\mathbf{y}) = \frac{p_{\mathbf{y}|H}(\mathbf{y}|H_m)\, p_H(H_m)}{\sum_{m'} p_{\mathbf{y}|H}(\mathbf{y}|H_{m'})\, p_H(H_{m'})}.$$

While the belief is a complete characterization of our knowledge of the true hypothesis, in applications one must often go further and make a decision (i.e., an intelligent guess) based on this information. To make a good decision we need some measure of goodness, appropriately chosen for the application of interest. In the sequel, we develop a framework for such decision-making, restricting our attention to the binary ($M = 2$) case to simplify the exposition.

2.1 Binary Hypothesis Testing

Specializing to the binary case, our model consists of two components. One is the set of prior probabilities

$$\begin{aligned} P_0 &= p_H(H_0) \\ P_1 &= p_H(H_1) = 1 - P_0. \end{aligned} \tag{3}$$

The second is the observation model, corresponding to the likelihood functions

$$\begin{aligned} H_0 &: \; p_{\mathbf{y}|H}(\mathbf{y}|H_0) \\ H_1 &: \; p_{\mathbf{y}|H}(\mathbf{y}|H_1). \end{aligned} \tag{4}$$

The development is essentially the same whether the observations are discrete or continuous. We arbitrarily use the continuous case in our development. The discrete case differs only in that integrals are replaced by summations.

We begin with a simple example to which we will return later.

Example 1. As a highly simplified scenario, suppose a single bit of information $m \in \{0, 1\}$ is encoded into a codeword s_m and sent over a communication channel, where s_0 and s_1 are both deterministic, known quantities. Let's further suppose that the channel is noisy; specifically, what is received is

$$y = s_m + w,$$

where w is a zero-mean Gaussian random variable with variance σ^2 and independent of H. From this information, we can readily construct the probability density for the

[3] In applications where further data is obtained, beliefs can be further revised, again using Bayes' Rule as for the computation. This updating is a simple form of what is referred to as *belief propagation*.

observation under each of the hypotheses, obtaining:

$$
\begin{aligned}
p_{y|H}(y|H_0) &= \mathtt{N}(y; s_0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-s_0)^2/(2\sigma^2)} \\
p_{y|H}(y|H_1) &= \mathtt{N}(y; s_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-s_1)^2/(2\sigma^2)}.
\end{aligned}
\tag{5}
$$

In addition, if 0's and 1's are equally likely to be transmitted we would set the *a priori* probabilities to

$$P_0 = P_1 = 1/2.$$

2.1.1 Optimum Decision Rules: The Likelihood Ratio Test

The solution to a hypothesis test is specified in terms of a *decision rule*. We focus for the time being on *deterministic* decision rules. Mathematically, such a decision rule is a function $\hat{H}(\cdot)$ that uniquely maps every possible observation $\mathbf{y} \in \mathcal{Y}$ to one of the two hypotheses, i.e., $\hat{H} : \mathcal{Y} \mapsto \mathcal{H}$, where $\mathcal{H} = \{H_0, H_1\}$. From this perspective, we see that choosing the function $\hat{H}(\cdot)$ is equivalent to partitioning the observation space $\mathcal{Y}$ into two disjoint "decision" regions, corresponding to the values of $\mathbf{y}$ for which each of the two possible decisions are made. Specifically, we use $\mathcal{Y}_m$ to denote those values of $\mathbf{y} \in \mathcal{Y}$ for which our rule decides H_m, i.e.,

$$
\begin{aligned}
\mathcal{Y}_0 &= \{\mathbf{y} \in \mathcal{Y} \,:\, \hat{H}(\mathbf{y}) = H_0\} \\
\mathcal{Y}_1 &= \{\mathbf{y} \in \mathcal{Y} \,:\, \hat{H}(\mathbf{y}) = H_1\}.
\end{aligned}
\tag{6}
$$

These regions are depicted schematically in Fig. 1.

Our goal, then, is to design this bi-valued function (equivalently the associated decision regions $\mathcal{Y}_0$ and $\mathcal{Y}_1$) in such a way that the best possible performance is obtained. In order to do this, we need to be able to quantify the notion of "best." This requires that we have a well-defined objective function corresponding to a suitable measure of goodness. In the Bayesian approach, we use an objective function taking the form of an expected cost function. Specifically, we use

$$\tilde{C}(H_j, H_i) \triangleq C_{ij} \tag{7}$$

to denote the "cost" of deciding that the hypothesis is $\hat{H} = H_i$ when the correct hypothesis is $H = H_j$. Then the optimum decision rule takes the form

$$\hat{H}(\cdot) = \arg\min_{f(\cdot)} \varphi(f) \tag{8}$$

where the average cost, which is referred to as the "Bayes risk," is

$$\varphi(f) = \mathbb{E}\left[\tilde{C}(H, f(\mathbf{y}))\right], \tag{9}$$

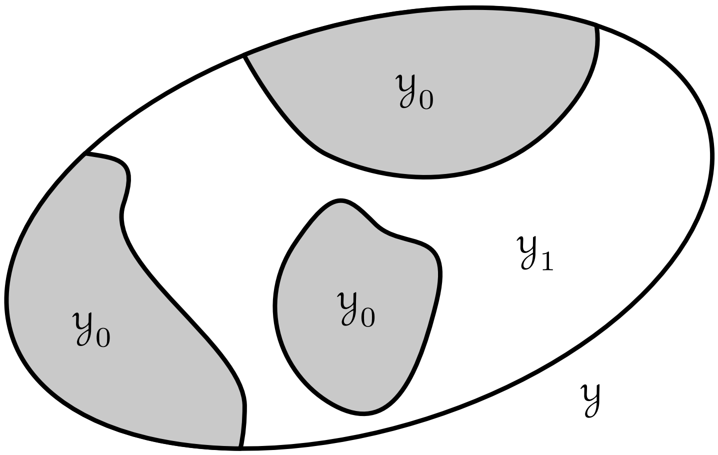

Figure 1: The regions $\mathcal{Y}_0$ and $\mathcal{Y}_1$ as defined in (6) corresponding to an example decision rule $\hat{H}(\cdot)$, where $\mathcal{Y}$ is the the observation alphabet.

and where the expectation in (9) is over both y and H, and where $f(\cdot)$ is a decision rule.

Generally, the application dictates an appropriate choice of the costs C_{ij}. For example, a symmetric cost function of the form $C_{ij} = 1 - \mathbb{1}_{i=j}$, i.e.,

$$\begin{aligned} C_{00} &= C_{11} = 0 \\ C_{01} &= C_{10} = 1, \end{aligned} \tag{10}$$

corresponds to seeking a decision rule that minimizes the probability of a decision error. However, there are many applications for which such symmetric cost functions are not well-matched. For example, in a medical diagnosis problem where H_0 denotes the hypotheses that the patient does not have a particular disease and H_1 that he does, we would typically want to select cost assignments such that $C_{01} \gg C_{10}$.

Definition 1. *A set of costs $\{C_{ij}\}$ is* valid *if the cost of a correct decision is lower than the cost of an incorrect decision, i.e., $C_{jj} < C_{ij}$ whenever $i \neq j$.*

Theorem 1. *Given* a priori *probabilities P_0, P_1, data $\mathbf{y}$, observation models $p_{\mathsf{y}|H}(\cdot|H_0)$, $p_{\mathsf{y}|H}(\cdot|H_1)$, and valid costs $C_{00}, C_{01}, C_{10}, C_{11}$, the optimum Bayes' decision rule takes the form:*

$$L(\mathbf{y}) \triangleq \frac{p_{\mathsf{y}|H}(\mathbf{y}|H_1)}{p_{\mathsf{y}|H}(\mathbf{y}|H_0)} \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} \triangleq \eta, \tag{11}$$

i.e., the decision is H_1 when $L(\mathbf{y}) > \eta$, the decision is H_0 when $L(\mathbf{y}) < \eta$, and the decision can be made arbitrarily when $L(\mathbf{y}) = \eta$.

Before establishing this result, we make a few remarks. First, the left-hand side of (11) is referred to as the *likelihood ratio*, and thus (11) is typically referred to as a *likelihood ratio test* (LRT). Note too that the likelihood ratio—which we denote using $L(\mathbf{y})$—is constructed from the observations model and the data. Meanwhile, the right-hand side of (11)—which we denote using η—is a precomputable threshold that is determined from the *a priori* probabilities and costs.

Proof. Consider an arbitrary but fixed decision rule $f(\cdot)$. In terms of this generic $f(\cdot)$, the Bayes risk can be expanded in the form

$$\begin{aligned} \varphi(f) &= \mathbb{E}\left[\tilde{C}(\mathsf{H}, f(\mathbf{y}))\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\tilde{C}(\mathsf{H}, f(\mathbf{y})) \,\Big|\, \mathsf{y} = \mathbf{y}\right]\right] \\ &= \int \tilde{\varphi}(f(\mathbf{y}), \mathbf{y})\, p_{\mathsf{y}}(\mathbf{y})\, \mathrm{d}\mathbf{y}, \end{aligned} \tag{12}$$

with

$$\tilde{\varphi}(H, \mathbf{y}) = \mathbb{E}\left[\tilde{C}(\mathsf{H}, H) \,\Big|\, \mathsf{y} = \mathbf{y}\right], \tag{13}$$

and where to obtain the second equality in (12) we have used iterated expectation.

Note from (12) that since $p_{\mathsf{y}}(\mathbf{y})$ is nonnegative, it is clear that we minimize φ if we minimize $\tilde{\varphi}(f(\mathbf{y}), \mathbf{y})$ *for each particular value of* $\mathbf{y}$. Hence, we can determine the optimum decision rule $\hat{H}(\cdot)$ on a point-by-point basis, i.e., $\hat{H}(\mathbf{y})$ for each $\mathbf{y}$.

Let's consider a particular (observation) point $\mathsf{y} = \mathbf{y}_*$. For this point, if we choose the assignment

$$\hat{H}(\mathbf{y}_*) = H_0,$$

then our conditional expectation (13) takes the value

$$\tilde{\varphi}(H_0, \mathbf{y}_*) = C_{00}\, p_{\mathsf{H}|\mathsf{y}}(H_0|\mathbf{y}_*) + C_{01}\, p_{\mathsf{H}|\mathsf{y}}(H_1|\mathbf{y}_*). \tag{14}$$

Alternatively, if we choose the assignment

$$\hat{H}(\mathbf{y}_*) = H_1,$$

then our conditional expectation (13) takes the value

$$\tilde{\varphi}(H_1, \mathbf{y}_*) = C_{10}\, p_{\mathsf{H}|\mathsf{y}}(H_0|\mathbf{y}_*) + C_{11}\, p_{\mathsf{H}|\mathsf{y}}(H_1|\mathbf{y}_*). \tag{15}$$

Hence, the optimum assignment for the value $\mathbf{y}_*$ is simply the choice corresponding to the smaller of (14) and (15). It is convenient to express this optimum decision rule using the following notation (now replacing our particular observation $\mathbf{y}_*$ with a generic observation $\mathbf{y}$):

$$\begin{aligned} C_{00}\, p_{H|\mathbf{y}}(H_0|\mathbf{y}) & \quad \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} \quad C_{10}\, p_{H|\mathbf{y}}(H_0|\mathbf{y}) \\ + C_{01}\, p_{H|\mathbf{y}}(H_1|\mathbf{y}) & \qquad\qquad\quad + C_{11}\, p_{H|\mathbf{y}}(H_1|\mathbf{y}). \end{aligned} \tag{16}$$

Note that when the two sides of (16) are equal, then either assignment is equally good—both have the same effect on the objective function (12).

A minor rearrangement of the terms in (16) results in

$$(C_{01} - C_{11}) p_{H|\mathbf{y}}(H_1|\mathbf{y}) \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} (C_{10} - C_{00}) p_{H|\mathbf{y}}(H_0|\mathbf{y}). \tag{17}$$

Since for any valid choice of costs the terms in parentheses in (17) are both positive, we can equivalently write (17) in the form[4]

$$\frac{p_{H|\mathbf{y}}(H_1|\mathbf{y})}{p_{H|\mathbf{y}}(H_0|\mathbf{y})} \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})}. \tag{18}$$

When we then substitute (19) into (18) and multiply both sides by P_0/P_1, we obtain the decision rule in its final form (11), directly in terms of the measurement densities.

As a final remark, observe that, not surprisingly, the optimum decision produced by (17) is a particular function of our beliefs, i.e., the *a posteriori* probabilities

$$p_{H|\mathbf{y}}(H_m|\mathbf{y}) = \frac{p_{\mathbf{y}|H}(\mathbf{y}|H_m)\, P_m}{p_{\mathbf{y}|H}(\mathbf{y}|H_0)\, P_0 + p_{\mathbf{y}|H}(\mathbf{y}|H_1)\, P_1}. \tag{19}$$

□

2.1.2 Properties of the Likelihood Ratio Test

Several observations lend insight into the optimum decision rule (11). First, note that the likelihood ratio $L(\cdot)$ is a scalar-valued function, i.e., $L : \mathcal{Y} \to \mathbb{R}$, regardless of the dimension or alphabet of the data. In fact, $L(\mathbf{y})$ is an example of what is referred to as a *sufficient statistic* for the problem: it summarizes everything we need to know about the observation vector in order to make a decision. Phrased differently, in terms of our ability to make the optimum decision (in the Bayesian sense in this case), knowledge of $L(\mathbf{y})$ is as good as knowledge of the full data vector $\mathbf{y}$ itself.

[4]Technically, we have to be careful about dividing by zero here if $p_{H|\mathbf{y}}(H_0|\mathbf{y}) = 0$. To simplify our exposition, however, as we discuss in Section 2.1.2, we will generally restrict our attention to the case where this does not happen.

We will develop the notion of a sufficient statistic more precisely and in greater generality in a subsequent section of the notes; however, at this point it suffices to make two observations with respect to our hypothesis testing problem. First, (11) tells us an explicit construction for a scalar sufficient statistic for the Bayesian binary hypothesis testing problem. Second, sufficient statistics are not unique. For example, any invertible function of $L(\mathbf{y})$ is also a sufficient statistic. In fact, for the purposes of implementation or analysis it is often more convenient to rewrite the likelihood ratio test in the form

$$L'(\mathbf{y}) = g(L(\mathbf{y})) \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} g(\eta), \tag{20}$$

where $g(\cdot)$ is some suitably chosen, monotonically increasing function. An important example is the case corresponding to $g(\cdot) = \ln(\cdot)$, which simplifies many tests involving densities with exponential factors, such as Gaussians.[5]

It is also important to emphasize that $\mathsf{L} = L(\mathsf{y})$ is a random variable—i.e., it takes on a different value in each experiment. As such, we will frequently be interested in its probability density function—or at least moments such as its mean and variance—under each of H_0 and H_1. Such densities can be derived using the usual method of events, and are often used in calculating performance of the decision rule.

It follows immediately from the definition in (11) that the likelihood ratio is a nonnegative quantity. Furthermore, depending on the problem, some values of $\mathbf{y}$ may lead to $L(\mathbf{y})$ being zero or infinite. In particular, the former occurs when $p_{\mathsf{y}|H}(\mathbf{y}|H_1) = 0$ but $p_{\mathsf{y}|H}(\mathbf{y}|H_0) > 0$, which is an indication that values in a neighborhood of $\mathbf{y}$ effectively cannot occur under H_1 but can under H_0. In this case, there will be values of $\mathbf{y}$ for which we'll effectively know with certainty that the correct hypothesis is H_0. When the likelihood ratio is infinite, corresponding a division-by-zero scenario, an analogous situation exists, but with the roles of H_0 and H_1 reversed. These cases where such perfect decisions are possible are referred to as *singular* decision scenarios. In some practical problems, these scenarios do in fact occur. However, in other cases they suggest a potential lack of robustness in the data modeling, i.e., that some source of inherent uncertainty may be missing from the model. In any event, to simplify our development for the remainder of the topic we will largely restrict our attention to the case where $0 < L(\mathbf{y}) < \infty$ for all $\mathbf{y}$.

While the likelihood ratio focuses the observed data into a single scalar for the purpose of making an optimum decision, the threshold η for the test plays a complementary role. In particular, from (11) we see that η focuses the relevant features of the cost function and *a priori* probabilities into a single scalar. Furthermore, this information is combined in a manner that is intuitively satisfying. For example, as (11) also reflects, an increase in P_0 means that H_0 is more likely, so that η is increased to appropriately bias the test toward deciding H_0 for any particular observation. Sim-

[5]We will discuss an important such family of distributions—exponential families—in detail in a subsequent section of the notes.

ilarly, an increase in C_{10} means that deciding H_1 when H_0 is true is more costly, so η is increased to appropriately bias the test toward deciding H_0 to offset this risk. Finally, note that adding a constant to the cost function (i.e., to all C_{ij}) has, as we would anticipate, no effect on the threshold. Hence, without loss of generality we may set at least one of the correct decision costs—i.e., C_{00} or C_{11}—to zero.

Finally, it is important to emphasize that the likelihood ratio test (11) indirectly determines the decision regions (6). In particular, we have

$$\begin{aligned} \mathcal{Y}_0 &= \{\mathbf{y} \in \mathcal{Y} : \hat{H}(\mathbf{y}) = H_0\} = \{\mathbf{y} \in \mathcal{Y} : L(\mathbf{y}) < \eta\} \\ \mathcal{Y}_1 &= \{\mathbf{y} \in \mathcal{Y} : \hat{H}(\mathbf{y}) = H_1\} = \{\mathbf{y} \in \mathcal{Y} : L(\mathbf{y}) > \eta\}. \end{aligned} \tag{21}$$

As Fig. 1 suggests, while a decision rule expressed in the measurement data space $\mathcal{Y}$ can be complicated,[6] (11) tells us that the observations can be transformed into a one-dimensional space defined via $L = L(\mathbf{y})$ where the decision regions have a particularly simple form: the decision $\hat{H}(L) = H_0$ is made whenever L lies to the left of some point η on the line, and $\hat{H}(L) = H_1$ whenever L lies to the right.

2.1.3 Maximum A Posteriori and Maximum Likelihood Decision Rules

An important cost assignment for many problems is that given by (10), which as we recall corresponds to a minimum probability-of-error criterion. Indeed, in this case, we have

$$\varphi(\hat{H}) = \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_0, \mathsf{H} = H_1\right) + \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_1, \mathsf{H} = H_0\right).$$

The corresponding decision rule in this case can be obtained as a special case of (11).

Corollary 1. *The minimum probability-of-error decision rule takes the form*

$$\hat{H}(\mathbf{y}) = \underset{H \in \{H_0, H_1\}}{\arg\max}\; p_{\mathsf{H}|\mathbf{y}}(H|\mathbf{y}). \tag{22}$$

The rule (22), in which one chooses the hypothesis for which our belief is largest, is referred to as the *maximum a posteriori* (MAP) decision rule.

Proof. Instead of specializing (11), we specialize the equivalent test (17), from which we obtain a form of the minimum probability-of-error test expressed in terms of the *a posteriori* probabilities for the problem, viz.,

$$p_{\mathsf{H}|\mathbf{y}}(H_1|\mathbf{y}) \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} p_{\mathsf{H}|\mathbf{y}}(H_0|\mathbf{y}). \tag{23}$$

From (23) we see that the desired decision rule can be expressed in the form (22) □

[6] Indeed, neither of the respective sets $\mathcal{Y}_0$ and $\mathcal{Y}_1$ are even connected in general.

Still further simplification is possible when the hypotheses are equally likely ($P_0 = P_1 = 1/2$). In this case, we have the following.

Corollary 2. *When the hypotheses are equally likely, the minimum probability of error decision rule takes the form*

$$\hat{H}(\mathbf{y}) = \underset{H \in \{H_0, H_1\}}{\arg\max}\ p_{\mathbf{y}|H}(\mathbf{y}|H). \tag{24}$$

The rule (24), which is referred to as the *maximum likelihood* (ML) decision rule, chooses the hypothesis for which the corresponding likelihood function is largest.

Proof. Specializing (11) we obtain

$$\frac{p_{\mathbf{y}|H}(\mathbf{y}|H_1)}{p_{\mathbf{y}|H}(\mathbf{y}|H_0)} \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} 1, \tag{25}$$

or, equivalently,

$$p_{\mathbf{y}|H}(\mathbf{y}|H_1) \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} p_{\mathbf{y}|H}(\mathbf{y}|H_0),$$

whence (24) □

Example 2. Continuing with Example 1, we obtain from (5) that the likelihood ratio test for this problem takes the form

$$L(y) = \frac{\dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-s_1)^2/(2\sigma^2)}}{\dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-s_0)^2/(2\sigma^2)}} \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} \eta. \tag{26}$$

As (26) suggests—and as is generally the case in Gaussian problems—the natural logarithm of the likelihood ratio is a more convenient sufficient statistic to work with in this example. In this case, taking logarithms of both sides of (26) yields

$$L'(y) = \frac{1}{2\sigma^2}\left[(y-s_0)^2 - (y-s_1)^2\right] \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} \ln\eta. \tag{27}$$

Expanding the quadratics and cancelling terms in (27) we obtain the test in its simplest form, which for $s_1 > s_0$ is given by

$$y \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} \frac{s_1+s_0}{2} + \frac{\sigma^2 \ln\eta}{s_1 - s_0} \triangleq \gamma. \tag{28}$$

In this form, the resulting error probability is easily obtained, and is naturally expressed in terms of Q-function notation.

We also remark that with a minimum probability-of-error criterion, if $P_0 = P_1$ then $\ln \eta = 0$ and we see immediately from (27) that the optimum test takes the form

$$|y - s_0| \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} |y - s_1|,$$

which corresponds to a "minimum-distance" decision rule, i.e.,

$$\hat{H}(y) = H_{\hat{m}}, \qquad \hat{m} = \underset{m \in \{0,1\}}{\arg\min} |y - s_m|.$$

This minimum-distance property turns out to hold in multidimensional Gaussian problems as well, and leads to convenient analysis in terms of Euclidean geometry.

Note too that in this problem the decisions regions on the y-axis have a particularly simple form; for example, for $s_1 > s_0$ we obtain

$$\begin{aligned} \mathcal{Y}_0 &= \{y \in \mathbb{R} : y < \gamma\} \\ \mathcal{Y}_1 &= \{y \in \mathbb{R} : y > \gamma\}. \end{aligned} \tag{29}$$

In other problems—even Gaussian ones—the decision regions can be more complicated, as our next example illustrates.

Example 3. Suppose that a zero-mean Gaussian random variable has one of two possible variances, σ_1^2 or σ_0^2, where $\sigma_1^2 > \sigma_0^2$. Let the costs and prior probabilities be arbitrary. Then the likelihood ratio test for this problem takes the form

$$L(y) = \frac{\dfrac{1}{\sqrt{2\pi\sigma_1^2}} e^{-y^2/(2\sigma_1^2)}}{\dfrac{1}{\sqrt{2\pi\sigma_0^2}} e^{-y^2/(2\sigma_0^2)}} \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} \eta.$$

In this problem, it is a straightforward exercise to show that the test simplifies to one of the form

$$|y| \underset{\hat{H}(y)=H_0}{\overset{\hat{H}(y)=H_1}{\gtrless}} \sqrt{2 \frac{\sigma_0^2 \sigma_1^2}{\sigma_1^2 - \sigma_0^2} \ln\left(\eta \frac{\sigma_1}{\sigma_0}\right)} \triangleq \gamma.$$

Hence, the decision region $\mathcal{Y}_1$ is the union of two disconnected regions in this case, i.e.,

$$\mathcal{Y}_1 = \{y \in \mathbb{R} : y > \gamma\} \cup \{y \in \mathbb{R} : y < -\gamma\}.$$