

Chapter 3

DETECTION, DECISIONS, AND HYPOTHESIS TESTING

3.1 Introduction

Detection, decision making, and hypothesis testing are synonyms. The word *detection* refers to the effort to decide whether some phenomenon is present or not in a given situation. For example, a radar system attempts to detect whether or not a target is present; a quality control system attempts to detect whether a unit is defective; a medical test detects whether a given disease is present. The meaning has been extended in the communication field to detect which one, among a set of mutually exclusive alternatives, is correct. *Decision making* is, again, the process of deciding between a number of mutually exclusive alternatives. *Hypothesis testing* is the same, and here the mutually exclusive alternatives are called hypotheses. We use the word hypotheses for these alternatives in what follows, since the word conjures up the appropriate intuitive image.

These problems will be studied initially in a purely probabilistic setting. That is, there is a probability model within which each hypothesis is an event. These events are mutually exclusive and collectively exhaustive, i.e., the sample outcome of the experiment lies in one and only one of these events, which means that in each performance of the experiment, one and only one hypothesis is correct. Assume there are m hypotheses, numbered $0, 1, \dots, m-1$, and let H be a random variable whose sample value is the correct hypothesis i , $0 \leq i \leq m-1$ for that particular sample point. The probability of hypothesis i , $P_H(i)$, is denoted P_i and is usually referred to as the *a priori probability* of i . There is also a random vector (rv) \vec{Y} , called the observation vector. We observe a sample value \vec{y} of \vec{Y} , and on the basis of that observation, we want to make a decision between the possible hypotheses.

Before discussing how to make these decisions, it is important to understand when and why decisions must be made. As an example, suppose we conclude, on the basis of the observation, that hypothesis 0 is correct with probability $2/3$ and hypothesis 1 with probability $1/3$. Simply making a decision on hypothesis 0 and forgetting about the probabilities

seems to be throwing away much of the information that we have gathered. The problem is that sometimes choices must be made. In a communication system, the user wants to receive the message rather than a set of probabilities. In a control system, the controls must occasionally take action. Similarly managers must occasionally choose between courses of action, between products, and between people to hire. In a sense, it is by making decisions (and, in Chapter 4, by making estimates) that we return from the world of mathematical probability models to the world being modeled.

There are a number of possible criteria to use in making decisions, and initially, we assume that the criterion is to maximize the probability of choosing correctly. That is, when the experiment is performed, the resulting sample point maps into a sample value i for H and into a sample value \vec{y} for \vec{Y} . The decision maker observes \vec{y} (but not i) and maps \vec{y} into a decision $\hat{H}(\vec{y})$. The decision is correct if $\hat{H}(\vec{y}) = i$. In principal, maximizing the probability of choosing correctly is almost trivially simple. Given \vec{y} , we calculate $P_{H|\vec{Y}}(i | \vec{y})$ for each i , $0 \leq i \leq m - 1$. This is the probability that i is the correct hypothesis conditional on \vec{y} . Thus the rule for maximizing the probability of being correct is to choose $\hat{H}(\vec{y})$ to be that i for which $P_{H|\vec{Y}}(i | \vec{y})$ is maximized. This is denoted

$$\hat{H}(\vec{y}) = \arg \max_i [P_{H|\vec{Y}}(i | \vec{y})] \quad (\text{MAP rule}) \quad (3.1)$$

where $\arg \max_i$ means the argument i that maximizes the function. If the maximum is not unique, it makes no difference to the probability of being correct which maximizing i is chosen. To be explicit, we arbitrarily choose the largest maximizing i . The conditional probability $P_{H|\vec{Y}}(i | \vec{y})$ is called an *a posteriori probability*, and thus the decision rule in (3.1) is called the maximum a posteriori probability (MAP) rule.

When we want to distinguish between different decision rules, we denote the MAP decision rule in (3.1) as $\hat{H}_{MAP}(\vec{y})$. Since the MAP rule maximizes the probability of correct decision for each sample value \vec{y} , it also maximizes the probability of correct decision averaged over all \vec{y} . To see this analytically, let $\hat{H}_A(\vec{y})$ be an arbitrary decision rule. Since \hat{H} maximizes $P_{H|\vec{Y}}(i | \vec{y})$ over i ,

$$P_{H|\vec{Y}}(\hat{H}_{MAP}(\vec{y}) | \vec{y}) \geq P_{H|\vec{Y}}(\hat{H}_A(\vec{y}) | \vec{y}); \quad \text{for any rule A and all } \vec{y} \quad (3.2)$$

For simplicity of notation, we assume in what follows that the observation random vector \vec{Y} , conditional on each hypothesis, has a probability density, and thus that \vec{Y} has an unconditional probability density, $p_{\vec{Y}}$. Averaging (3.2) over observation vectors,

$$\int p_{\vec{Y}}(\vec{y}) P_{H|\vec{Y}}(\hat{H}_{MAP}(\vec{y}) | \vec{y}) d\vec{y} \geq \int p_{\vec{Y}}(\vec{y}) P_{H|\vec{Y}}(\hat{H}_A(\vec{y}) | \vec{y}) d\vec{y} \quad (3.3)$$

The quantity on the left is the probability of correct decision using \hat{H}_{MAP} , and that on the right is the probability of correct decision using \hat{H}_A . The above results are very simple, but also important and fundamental. We summarize them in the following theorem.

Theorem 3.1 The MAP rule, given in (3.1), maximizes the probability of correct decision for each observed sample value \vec{y} and also maximizes the overall probability of correct decision.

Before discussing the implications and use of the MAP rule, we review the assumptions that have been made. First, we assumed a probability experiment in which all probabilities are known, and in which, for each performance of the experiment, one and only one hypothesis is correct. This conforms very well to a communication model in which a transmitter sends one of a set of possible signals, and the receiver, given signal plus noise, makes a decision on the signal actually sent. It does not always conform well to a scientific experiment attempting to verify the existence of some new phenomenon; in such situations, there is often no sensible way to model a priori probabilities. In section 3.5, we find ways to avoid depending on a priori probabilities.

The next assumption was that maximizing the probability of correct decision is an appropriate decision criterion. In many situations, the cost of a wrong decision is highly asymmetric. For example, when testing for a treatable but deadly disease, making an error when the disease is present is far more costly than making an error when the disease is not present. In section 3.4, we adopt a minimum cost formulation which allows us to treat these asymmetric cases.

The next five sections are restricted to the case of binary hypotheses, ($m = 2$). This allows us to understand most of the important ideas but simplifies the notation considerably. In section 3.7, we again consider an arbitrary number of hypotheses.

3.2 Binary Detection with the MAP Criterion

We now continue our discussion of the MAP criterion. Assume a probability model in which the correct hypothesis H is a binary random variable with possible values 0 and 1 and with positive a priori probabilities P_0 and P_1 . Let \vec{Y} be a r.v. whose conditional probability density, $p_{\vec{Y}|H}(\vec{y} | i)$, is initially assumed to be finite and non-zero for all \vec{y} and for $i = 0, 1$.

The conditional densities $p_{\vec{Y}|H}(\vec{y} | i)$, $i = 0, 1$ are called *likelihoods* in the jargon of hypothesis testing. The marginal density of \vec{Y} is given by $p_{\vec{Y}}(\vec{y}) = P_0 p_{\vec{Y}|H}(\vec{y} | 0) + P_1 p_{\vec{Y}|H}(\vec{y} | 1)$. The a posteriori probability of H , for $i = 0$ or 1 , is given by

$$P_{H|\vec{Y}}(i | \vec{y}) = \frac{P_i p_{\vec{Y}|H}(\vec{y} | i)}{p_{\vec{Y}}(\vec{y})} \quad (3.4)$$

Writing out (3.1) explicitly for this case,

$$\frac{P_1 p_{\vec{Y}|H}(\vec{y} | 1)}{p_{\vec{Y}}(\vec{y})} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \frac{P_0 p_{\vec{Y}|H}(\vec{y} | 0)}{p_{\vec{Y}}(\vec{y})} \quad (3.5)$$

This "equation" indicates that the decision is 1 if the left side is greater than or equal to the right, and is 0 if the left side is less than the right. Choosing the decision to be 1 when

equality holds is arbitrary and does not affect the probability of being correct. Canceling $p_{\vec{Y}}(\vec{y})$ and rearranging,

$$\Lambda(\vec{y}) = \frac{p_{\vec{Y}|H}(\vec{y} | 1)}{p_{\vec{Y}|H}(\vec{y} | 0)} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \frac{P_0}{P_1} = \eta \quad (3.6)$$

$\Lambda(\vec{y}) = p_{\vec{Y}|H}(\vec{y} | 1)/p_{\vec{Y}|H}(\vec{y} | 0)$ is called the *likelihood ratio*, and is a function of \vec{y} . Similarly $\eta = P_0/P_1$ is called the *threshold*. The binary MAP rule (or MAP test, as it is often called) is then to compare the likelihood ratio with the threshold, and decide on hypothesis 0 if the threshold is not reached, and on hypothesis 1 otherwise. Note that if the a priori probability P_0 is increased, the threshold increases, and the set of \vec{y} for which hypothesis 0 is chosen increases; this corresponds to our intuition—the more certain we are initially that H is 0, the stronger the evidence required to make us change our minds. We shall find later, when we look at the minimum cost problem, that the only effect of minimizing over costs is to change the threshold η in (3.6).

An important special case of (3.6) is that in which $P_0 = P_1$. In this case $\eta = 1$, and the rule chooses $\hat{H}(\vec{y}) = 1$ for $p_{\vec{Y}|H}(\vec{y} | 1) \geq p_{\vec{Y}|H}(\vec{y} | 0)$ and chooses $\hat{H}(\vec{y}) = 0$ otherwise. This is called a *maximum likelihood (ML) rule or test*. The maximum likelihood test is often used when P_0 and P_1 are unknown, as is discussed in Section 3.5.

We now find the probability of error under each hypothesis, $\Pr(e | H=0)$ and $\Pr(e | H=1)$. From this we can also find the overall probability of error, $\Pr(e) = P_0 \Pr(e | H=0) + P_1 \Pr(e | H=1)$. In the radar field, $\Pr(e | H=0)$ is called the probability of false alarm, and $\Pr(e | H=1)$ is called the probability of a miss. Also $1 - \Pr(e | H=1)$ is called the probability of detection. In statistics, $\Pr(e | H=1)$ is called the probability of error of the second kind, and $\Pr(e | H=0)$ is the probability of error of the first kind.

Note that (3.6) partitions the space of observed sample values into 2 regions. $R_1 = \{\vec{y} : \Lambda(\vec{y}) \geq \eta\}$ is the region for which $\hat{H} = 1$ and $R_0 = \{\vec{y} : \Lambda(\vec{y}) < \eta\}$ is the region for which $\hat{H} = 0$. For $H=0$, an error occurs iff \vec{y} is in R_1 , and for $H=1$, an error occurs iff \vec{y} is in R_0 . Thus,

$$\Pr(e | H=0) = \int_{\vec{y} \in R_1} p_{\vec{Y}|H}(\vec{y} | 0) d\vec{y} \quad (3.7)$$

$$\Pr(e | H=1) = \int_{\vec{y} \in R_0} p_{\vec{Y}|H}(\vec{y} | 1) d\vec{y} \quad (3.8)$$

Another, frequently more useful, approach to finding the probability of error is to work directly with the likelihood ratio. Since $\Lambda(\vec{y})$ is a function of the observed sample value \vec{y} , we can define the likelihood ratio random variable $\Lambda(\vec{Y})$ in the usual way, i.e., for every sample point ω , $Y(\omega)$ is the corresponding sample value \vec{y} , and $\Lambda(\vec{Y})$ is then shorthand for $\Lambda(\vec{Y}(\omega))$. In the same way, $\hat{H}(\vec{Y})$ (or more briefly \hat{H}) is the decision random variable. In these terms, (3.6) states that

$$\hat{H} = 1 \quad \text{iff} \quad \Lambda(\vec{Y}) \geq \eta \quad (3.9)$$

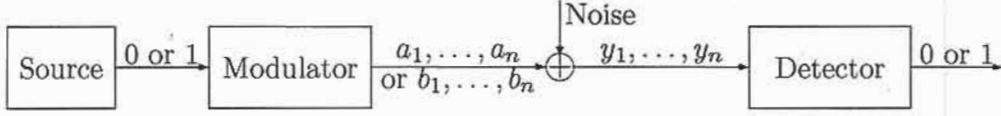


Figure 3.1: The source, in some given interval of time, attempts to transmit a binary digit, either 0 or 1, to the detector. The transmitter is here abstracted into a modulator which maps 0 into the signal vector $\vec{a} = (a_1, \dots, a_n)^T$ and maps 1 into the vector $\vec{b} = (b_1, \dots, b_n)^T$. A Gaussian noise vector, independent of the source output, is added to the modulated signal. The detector receives the signal plus noise. Based on this observation it makes a decision on the source output.

Thus,

$$\Pr(e | H=0) = \Pr(\hat{H}=1 | H=0) = \Pr(\Lambda(\vec{Y}) \geq \eta | H=0) \quad (3.10)$$

$$\Pr(e | H=1) = \Pr(\hat{H}=0 | H=1) = \Pr(\Lambda(\vec{Y}) < \eta | H=1) \quad (3.11)$$

A *sufficient statistic* is defined as a function of the observation vector \vec{y} from which the likelihood ratio can be calculated. For example, \vec{y} itself, $\Lambda(\vec{y})$, and any one to one function of $\Lambda(\vec{y})$ are sufficient statistics. $\Lambda(\vec{y})$ and functions of it are often simpler to work with than \vec{y} in calculating the probability of error, since they are one dimensional variables rather than vector variables. We have seen that the MAP rule (and, as we find later, essentially any sensible decision rule) can be specified in terms of the likelihood ratio. Thus, once a sufficient statistic has been calculated from the observed vector, the observed vector has no further value. For example, we see from (3.10) and (3.11) that the conditional error probabilities are determined simply from the conditional distribution functions of the likelihood ratio. We will often find that the log likelihood ratio, $\text{LLR}(\vec{Y}) = \ln[\Lambda(\vec{Y})]$ is even more convenient to work with than $\Lambda(\vec{Y})$. We next look at some widely used examples of binary MAP detection.

3.3 Binary Detection in Additive Gaussian Noise

We look at three progressively more complex examples here. Each can be visualized most easily in terms of the communication situation depicted in Figure 3.1, but each also applies more generally to many situations in which noisy measurements are taken to distinguish between two alternatives.

Example 3.1 First we look at the scalar version of Figure 3.1 where $n = 1$. That is, the correct hypothesis H is either 1 or 0; 1 is mapped into the real number b and 0 is mapped into the real number a . Let $Z \sim \mathcal{N}(0, \sigma^2)$ be the Gaussian noise rv, independent of H . The observation rv Y is $b + Z$ or $a + Z$, depending on whether $H = 1$ or 0. Thus, conditional on $H = 1$, $Y \sim \mathcal{N}(b, \sigma^2)$ and, conditional on $H = 0$, $Y \sim \mathcal{N}(a, \sigma^2)$.

$$p_{Y|H}(y | 1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y-b)^2}{(2\sigma^2)} \right]; \quad p_{Y|H}(y | 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y-a)^2}{(2\sigma^2)} \right]$$

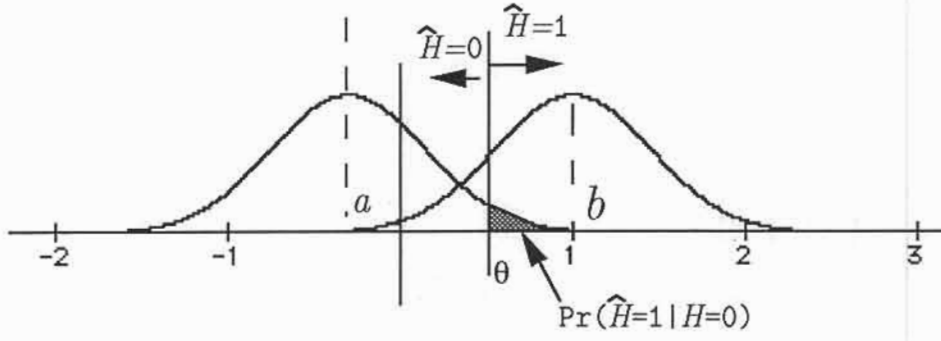


Figure 3.2: Binary hypothesis testing for signal plus noise

The likelihood ratio is the ratio of these likelihoods, and given by

$$\begin{aligned}\Lambda(y) &= \exp \left[\frac{(y-a)^2 - (y-b)^2}{2\sigma^2} \right] = \exp \left[\frac{2(b-a)y + (a^2 - b^2)}{2\sigma^2} \right] \\ &= \exp \left[\left(\frac{b-a}{\sigma^2} \right) \left(y - \frac{b+a}{2} \right) \right]\end{aligned}\quad (3.12)$$

Substituting this into (3.6), we have

$$\exp \left[\left(\frac{b-a}{\sigma^2} \right) \left(y - \frac{b+a}{2} \right) \right] \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \frac{P_0}{P_1} = \eta \quad (3.13)$$

This is further simplified by taking the logarithm, yielding

$$\text{LLR}(y) = \left[\left(\frac{b-a}{\sigma^2} \right) \left(y - \frac{b+a}{2} \right) \right] \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \ln(\eta) \quad (3.14)$$

Assuming that $b > a$, (3.14) can be rewritten as a threshold rule on y directly,

$$y \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \frac{\sigma^2 \ln(\eta)}{b-a} + \frac{b+a}{2} = \theta \quad (3.15)$$

This says that comparing $\Lambda(y)$ to a threshold η is equivalent to comparing y to a threshold $\theta = \sigma^2 \ln(\eta)/(b-a) + (b+a)/2$. In the maximum likelihood (ML) case ($P_1 = P_0$), the threshold η for Λ is 1 and the threshold θ for y is the midpoint between a and b (i.e., $\theta = (b+a)/2$). For the MAP case, If η is larger or smaller than 1, θ is respectively larger or smaller than $(b+a)/2$ (see Figure 3.2).

From (3.15), $\Pr(e | H=0) = \Pr(Y \geq \theta | H=0)$. Given $H = 0$, $Y \sim \mathcal{N}(a, \sigma^2)$, so, given $H = 0$, $(Y - a)/\sigma$ is a normalized Gaussian variable.

$$\Pr(Y \geq \theta | H=0) = \Pr\left(\frac{Y - a}{\sigma} \geq \frac{\theta - a}{\sigma} | H=0\right) = Q\left(\frac{\theta - a}{\sigma}\right) \quad (3.16)$$

where $Q(x)$ is defined as the complementary distribution function of a normalized Gaussian rv, i.e.,

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \quad (3.17)$$

$Q(x)$ cannot be evaluated in closed form, but is a standard function that will appear frequently. Replacing θ in (3.16) by its value in (3.15),

$$\Pr(e | H=0) = Q\left(\frac{\sigma \ln(\eta)}{b - a} + \frac{b - a}{2\sigma}\right) \quad (3.18)$$

We evaluate $\Pr(e | H=1) = \Pr(Y < \theta | H=1)$ in the same way. Given $H = 1$, Y is $\mathcal{N}(b, \sigma^2)$, so

$$\Pr(Y < \theta | H=0) = \Pr\left(\frac{Y - b}{\sigma} < \frac{\theta - b}{\sigma} | H=0\right) = 1 - Q\left(\frac{\theta - b}{\sigma}\right)$$

Using (3.15) for θ and noting that $Q(x) = 1 - Q(-x)$ for any x ,

$$\Pr(e | H=1) = Q\left(\frac{-\sigma \ln(\eta)}{b - a} + \frac{b - a}{2\sigma}\right) \quad (3.19)$$

Note that (3.18) and (3.19) are functions only of $(b-a)/\sigma$ and η . That is, only the difference between b and a is relevant, not the individual values, and it is only this difference relative to σ that is relevant. This should be intuitively clear from Figure 3.2. If we define $\gamma = (b-a)/\sigma$, then (3.18) and (3.19) simplify to

$$\Pr(e | H=0) = Q\left(\frac{\ln(\eta)}{\gamma} + \frac{\gamma}{2}\right) \quad \Pr(e | H=1) = Q\left(\frac{-\ln(\eta)}{\gamma} + \frac{\gamma}{2}\right) \quad (3.20)$$

For ML detection, $\eta = 1$, so this simplifies further to

$$\Pr(e | H=0) = \Pr(e | H=1) = \Pr(e) = Q(\gamma/2) \quad (3.21)$$

The quantity γ^2 here is a signal to noise ratio, i.e., the ratio of the signal difference energy to the noise energy. We shall find that the next two examples also reduce to this same form.

Example 3.2 Now we look at the vector version of Figure 3.1. If the source output is 0, ($H=0$), the modulator produces the real vector $\vec{a} = (a_1, \dots, a_n)^T$. If the source output is 1, the real vector $\vec{b} = (b_1, \dots, b_n)^T$ is produced. $\vec{Z} = (Z_1, \dots, Z_n)^T$ is a noise r \vec{v} assumed to be $\mathcal{N}(0, \sigma^2 I)$. That is, Z_1, \dots, Z_n are IID Gaussian rv's, also independent of H . The observation r \vec{v} \vec{Y} is $\vec{b} + \vec{Z}$ or $\vec{a} + \vec{Z}$, depending on whether $H = 1$ or 0. Thus, given $H=1$, $\vec{Y} \sim \mathcal{N}(\vec{b}, \sigma^2 I_n)$, so that

$$p_{\vec{Y}|H}(\vec{y} | 1) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \sum_{k=1}^n \frac{-(y_k - b_k)^2}{2\sigma^2} \quad (3.22)$$

Similarly, given $H = 0$, $\vec{Y} \sim \mathcal{N}(\vec{a}, \sigma^2 I)$, so that

$$p_{\vec{Y}|H}(\vec{y} | 0) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \sum_{k=1}^n \frac{-(y_k - a_k)^2}{2\sigma^2} \quad (3.23)$$

The likelihood ratio is then given by

$$\Lambda(\vec{y}) = \exp \sum_{k=1}^n \frac{(y_k - a_k)^2 - (y_k - b_k)^2}{2\sigma^2} \quad (3.24)$$

$$\begin{aligned} &= \exp \sum_{k=1}^n \frac{2(b_k - a_k)y_k + (a_k^2 - b_k^2)}{2\sigma^2} \\ &= \exp \left[\frac{(\vec{b} - \vec{a})^T \vec{y}}{\sigma^2} + \frac{\vec{a}^T \vec{a} - \vec{b}^T \vec{b}}{2\sigma^2} \right] \end{aligned} \quad (3.25)$$

Substituting this into (3.6) and taking the logarithm of both sides,

$$\text{LLR}(\vec{y}) = \frac{(\vec{b} - \vec{a})^T \vec{y}}{\sigma^2} + \frac{\vec{a}^T \vec{a} - \vec{b}^T \vec{b}}{2\sigma^2} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \ln \frac{P_0}{P_1} = \ln(\eta) \quad (3.26)$$

It can be seen that the test involves the observation \vec{y} only in terms of the inner product $(\vec{b} - \vec{a})^T \vec{y}$, so we can rewrite (3.26) in the form

$$(\vec{b} - \vec{a})^T \vec{y} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \sigma^2 \ln(\eta) + \frac{(\vec{b}^T \vec{b} - \vec{a}^T \vec{a})}{2} = \phi \quad (3.27)$$

These equations are interpreted in Figure 3.3. Contours of equal probability density for $p_{\vec{Y}|H}(\vec{y} | 0)$ are concentric spherical shells centered at \vec{a} , whereas contours of equal probability density for $p_{\vec{Y}|H}(\vec{y} | 1)$ are concentric spherical shells centered at \vec{b} . As can be seen from (3.24), the locus of points of constant likelihood ratio are points \vec{y} for which the squared distance to \vec{a} , less the squared distance to \vec{b} , is a constant. This set of points forms a straight line for the two dimensional case shown in Figure 3.3. In general, as seen analytically by (3.25), points of constant likelihood ratio are points for which $(\vec{b} - \vec{a})^T \vec{y}$ is constant, and this is the equation of an affine space.¹

We have seen from (3.27) that comparing $\Lambda(\vec{y})$ to the threshold η is equivalent to comparing $(\vec{b} - \vec{a})^T \vec{y}$ to the threshold ϕ . Thus the affine space $(\vec{b} - \vec{a})^T \vec{y} = \phi$ separates the observation space into two regions, where $\hat{H} = 1$ for $(\vec{b} - \vec{a})^T \vec{y} \geq \phi$ and $\hat{H} = 0$ otherwise.

¹In linear algebra, an $n - 1$ dimensional hyperplane in n dimensional space is by definition a linear space in its own right; such a space must include the origin, and be spanned by $n - 1$ vectors. The translation of a hyperplane away from the origin is called an *affine space*. For the case here, points \vec{y} for which $(\vec{b} - \vec{a})^T \vec{y} = 0$ form a hyperplane of points perpendicular to $(\vec{b} - \vec{a})$. The set of points for which $(\vec{b} - \vec{a})^T \vec{y} = c$, for some constant c , is thus an affine space.

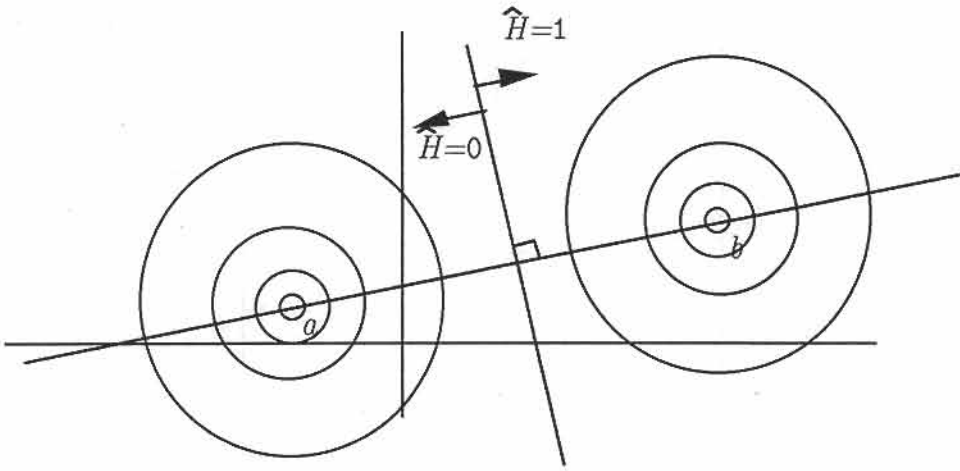


Figure 3.3: Decision regions for binary signals in additive IID Gaussian noise.

We also see from (3.26) that $\Lambda(\vec{y})$ can be calculated from $(\vec{b} - \vec{a})^T \vec{y}$, so that $(\vec{b} - \vec{a})^T \vec{y}$ is a sufficient statistic. This says that the threshold test for this problem is based on the value of a single number which is simply a linear combination of the observed variables. Note that each observation y_k is weighted by $(b_k - a_k)$ in forming the sufficient statistic. This makes sense intuitively, since if $b_k - a_k$ is very small, the observation y_k is mostly noise, whereas if $b_k - a_k$ is large, the observation gives a much better indication of which hypothesis is correct.

We can view $(\vec{b} - \vec{a})^T \vec{y}$ as the correlation² between $\vec{b} - \vec{a}$ and the observation \vec{y} . Thus a threshold detector, for this additive Gaussian noise case, is often called a *correlation detector* in communication theory. Often \vec{b} and \vec{a} are separately correlated with \vec{y} and the results compared; this is also called a correlation detector.

If we view \vec{y} as a discrete time sequence y_1, \dots, y_n , then we can also visualize performing this correlation function by convolving y_1, \dots, y_n with $(b_n - a_n), (b_{n-1} - a_{n-1}), \dots, (b_1 - a_1)$. This is the output, at the appropriate sampling time, of a digital filter with the impulse response $(b_n - a_n), \dots, (b_1 - a_1)$. A filter with this impulse response is said to be a *matched filter* to $(b_1 - a_1), \dots, (b_n - a_n)$. We will look at correlation detectors and matched filters again later when we consider detection of waveforms. The important point to note here, however, is that both the correlation detector and the matched filter simply compute the inner product $(\vec{b} - \vec{a})^T \vec{y}$.

Another way of viewing (3.27), and the most fundamental, is to view it in a different co-ordinate basis. That is, view the observation \vec{y} as a point in n dimensional space represented in a particular co-ordinate system. Consider a different orthonormal basis where one of the basis elements is $(\vec{b} - \vec{a}) / \|\vec{b} - \vec{a}\|$, where $\|\vec{b} - \vec{a}\|$ is the length of $\vec{b} - \vec{a}$,

$$\|\vec{b} - \vec{a}\| = \sqrt{(\vec{b} - \vec{a})^T (\vec{b} - \vec{a})}, \quad (3.28)$$

Thus $(\vec{b} - \vec{a}) / \|\vec{b} - \vec{a}\|$ is the vector $\vec{b} - \vec{a}$ normalized to unit length.

²For the moment, we ignore any similarity between this use of the word correlation as an inner product and the use of correlation as an expectation between random variables; we discuss this similarity later.

The two hypotheses can then only be distinguished by the component of the observation vector in this direction, i.e., by $(\vec{b} - \vec{a})^T \vec{y} / \|\vec{b} - \vec{a}\|$. This is what (3.27) says, but we now see that this is very intuitive geometrically. The measurements in orthogonal directions only measure noise. Because the noise is IID, the noise in these directions is independent of both the signal and the noise in the direction of interest, and thus can be ignored. This is sometimes called the theorem of irrelevance.

Note that $\vec{b}^T \vec{b} - \vec{a}^T \vec{a} = (\vec{b} - \vec{a})^T (\vec{b} + \vec{a})$. Substituting this in (3.26), we get

$$\text{LLR}(\vec{y}) = \frac{(\vec{b} - \vec{a})^T}{\sigma^2} \left(\vec{y} - \frac{\vec{b} + \vec{a}}{2} \right) \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \ln \frac{P_0}{P_1} = \ln(\eta) \quad (3.29)$$

This says that for ML detection, where $\ln \eta = 0$, the decision regions are separated by the affine space that forms the perpendicular bisector between \vec{a} and \vec{b} .

Finally, we use (3.29) to evaluate $\Pr(e | H=0)$. $E[\vec{Y} - (\vec{b} + \vec{a})/2 | H=0] = (\vec{a} - \vec{b})/2$, so

$$E[\text{LLR}(\vec{Y}) | H=0] = \frac{-(\vec{b} - \vec{a})^T (\vec{b} - \vec{a})}{2\sigma^2}$$

Defining γ as

$$\gamma = \frac{\|\vec{b} - \vec{a}\|}{\sigma} \quad (3.30)$$

this simplifies to $E[\text{LLR}(\vec{Y}) | H=0] = -\gamma^2/2$. Similarly, we see that $\text{VAR}[\text{LLR}(\vec{Y}) | H=0] = \gamma^2$. Thus, conditional on $H = 0$, $\text{LLR}(\vec{Y}) \sim \mathcal{N}(-\gamma^2/2, \gamma^2)$. The probability of error can then be found (see Exercise 3.1 as

$$\Pr(e | H=0) = \Pr[\text{LLR}(\vec{Y}) \geq \ln(\eta) | H=0] = Q\left(\frac{\ln(\eta)}{\gamma} + \frac{\gamma}{2}\right) \quad (3.31)$$

Analyzing $\text{LLR}(\vec{Y})$ conditional on $H = 1$ in the same way, we find that, conditional on $H = 1$, $\text{LLR}(\vec{Y}) \sim \mathcal{N}(\gamma^2/2, \gamma^2)$, and it follows that

$$\Pr(e | H=1) = Q\left(\frac{-\ln(\eta)}{\gamma} + \frac{\gamma}{2}\right) \quad (3.32)$$

Note that both error probabilities are functions only of $\gamma = \|\vec{b} - \vec{a}\|/\sigma$. This is not surprising in terms of our geometric interpretation. $\|\vec{b} - \vec{a}\|$ is the distance from \vec{a} to \vec{b} , and this is normalized by the standard deviation of the noise. That is, if we measure both $\|\vec{b} - \vec{a}\|$ and σ in some other units, the error probability cannot change. We can interpret $\|\vec{b} - \vec{a}\|^2$ as the energy in the difference between the signals. We can also interpret σ^2 as the energy per measurement of the noise. This says that what is relevant is not the number of different measurement values (i.e., n), but rather the total signal difference energy used over the set of measurements. With IID Gaussian noise, this signal difference energy can be split up in any way without affecting the error probability. This is why the signal to noise ratio is such an important parameter in digital communication.

Example 3.3 We consider Figure 3.1 again, but now we generalize the noise to be $\mathcal{N}(0, K_{\bar{Z}})$ where $K_{\bar{Z}}$ is non-singular. The likelihoods are then

$$p_{\bar{Y}|H}(\bar{y} | 1) = \frac{\exp \left[-\frac{1}{2}(\bar{y} - \bar{b})^T K_{\bar{Z}}^{-1}(\bar{y} - \bar{b}) \right]}{(2\pi)^{n/2} \sqrt{\det(K_{\bar{Z}})}} \quad (3.33)$$

$$p_{\bar{Y}|H}(\bar{y} | 0) = \frac{\exp \left[-\frac{1}{2}(\bar{y} - \bar{a})^T K_{\bar{Z}}^{-1}(\bar{y} - \bar{a}) \right]}{(2\pi)^{n/2} \sqrt{\det(K_{\bar{Z}})}} \quad (3.34)$$

The log likelihood ratio is

$$\text{LLR}(\bar{y}) = \frac{1}{2}(\bar{y} - \bar{a})^T K_{\bar{Z}}^{-1}(\bar{y} - \bar{a}) - \frac{1}{2}(\bar{y} - \bar{b})^T K_{\bar{Z}}^{-1}(\bar{y} - \bar{b}) \quad (3.35)$$

$$= (\bar{b} - \bar{a})^T K_{\bar{Z}}^{-1} \bar{y} + \frac{1}{2} \bar{a}^T K_{\bar{Z}}^{-1} \bar{a} - \frac{1}{2} \bar{b}^T K_{\bar{Z}}^{-1} \bar{b} \quad (3.36)$$

This can be rewritten as

$$\text{LLR}(\bar{y}) = (\bar{b} - \bar{a})^T K_{\bar{Z}}^{-1} \left[\bar{y} - \frac{\bar{b} + \bar{a}}{2} \right] \quad (3.37)$$

The quantity $(\bar{b} - \bar{a})^T K_{\bar{Z}}^{-1} \bar{y}$ is a sufficient statistic, and is simply a linear combination of the measurement variables y_1, \dots, y_n . Note that

$$\mathbb{E} [\text{LLR}(\bar{Y} | H=0)] = -(\bar{b} - \bar{a})^T K_{\bar{Z}}^{-1}(\bar{b} - \bar{a})/2$$

Defining γ as

$$\gamma = \sqrt{(\bar{b} - \bar{a})^T K_{\bar{Z}}^{-1}(\bar{b} - \bar{a})}, \quad (3.38)$$

we see that $\mathbb{E}[\text{LLR}(\bar{Y} | H=0)] = -\gamma^2/2$. Similarly,

$$\text{VAR} [\text{LLR}(\bar{Y} | H=0)] = \gamma^2$$

Then, as before, the conditional distribution of the log likelihood ratio is given by (see Exercise 3.2

$$\text{Given } H=0, \text{ LLR}(\bar{Y}) \sim \mathcal{N}(-\gamma^2/2, \gamma^2) \quad (3.39)$$

In the same way,

$$\text{Given } H=1, \text{ LLR}(\bar{Y}) \sim \mathcal{N}(\gamma^2/2, \gamma^2) \quad (3.40)$$

The probability of error is then

$$\Pr(e | H=0) = Q \left(\frac{\ln \eta}{\gamma} + \frac{\gamma}{2} \right); \quad \Pr(e | H=1) = Q \left(\frac{-\ln \eta}{\gamma} + \frac{\gamma}{2} \right) \quad (3.41)$$

Note that the previous two examples are special cases of this more general result. The following theorem summarizes this.

Theorem 3.2 Let the observed rv \bar{Y} be given by $\bar{Y} = \bar{a} + \bar{Z}$ under $H=0$ and by $\bar{Y} = \bar{b} + \bar{Z}$ under $H=1$ and let $Z \sim \mathcal{N}(0, K_{\bar{Z}})$ where $K_{\bar{Z}}$ is nonsingular and \bar{Z} is independent of H . Then the distribution of the conditional log likelihood ratio rv's are given by (3.39, 3.40) and the conditional error probabilities by (3.41).

3.4 Binary Detection with the Minimum Cost Criterion

An hypothesis test with a minimum cost criterion is often called a *Bayes test*. Bayes was an early statistician who advocated studying statistics within the context of a probability model (this is now called Bayesian statistics). In the next section, we discuss the Neyman-Pearson test, which is a good example of a test that does not depend on a priori probabilities.

Assume a probability model in which the correct hypothesis H is a binary random variable with possible values 0 and 1 and a priori probabilities P_0 and P_1 . Let \vec{Y} be the observation, with the likelihood functions $p_{\vec{Y}|H}(\vec{y} | 0)$ and $p_{\vec{Y}|H}(\vec{y} | 1)$. The a posteriori probability of H is given, as before, by

$$P_{H|\vec{Y}}(i | \vec{y}) = \frac{P_i p_{\vec{Y}|H}(\vec{y} | i)}{p_{\vec{Y}}(\vec{y})} \quad (3.42)$$

Finally, we assume that there are costs associated with each decision and hypothesis. Let C_{ij} be the cost of choosing i if j is the correct hypothesis. Given a sample value \vec{y} for the observation, the expected cost of decision i is

$$E[\text{Cost of } \hat{H}=i | \vec{Y}=\vec{y}] = C_{i0}P_{H|\vec{Y}}(0 | \vec{y}) + C_{i1}P_{H|\vec{Y}}(1 | \vec{y}) \quad (3.43)$$

The decision, \hat{H} , that minimizes the expected cost for the observed sample value \vec{y} , is then

$$\hat{H} = \arg \min_i [C_{i0}P_{H|\vec{Y}}(0 | \vec{y}) + C_{i1}P_{H|\vec{Y}}(1 | \vec{y})] \quad (3.44)$$

Writing this out explicitly,

$$C_{00}P_{H|\vec{Y}}(0 | \vec{y}) + C_{01}P_{H|\vec{Y}}(1 | \vec{y}) \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} C_{10}P_{H|\vec{Y}}(0 | \vec{y}) + C_{11}P_{H|\vec{Y}}(1 | \vec{y})$$

Rearranging terms,

$$(C_{01} - C_{11})P_{H|\vec{Y}}(1 | \vec{y}) \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} (C_{10} - C_{00})P_{H|\vec{Y}}(0 | \vec{y}) \quad (3.45)$$

Assuming that $C_{01} - C_{11} > 0$ (i. e., that the cost of an erroneous decision is greater than that of a correct decision), we can divide (3.45) by this term and by $P_{H|\vec{Y}}(0 | \vec{y})$ to get

$$\frac{P_{H|\vec{Y}}(1 | \vec{y})}{P_{H|\vec{Y}}(0 | \vec{y})} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})} \quad (3.46)$$

Substituting (3.42) into this, we get our final result,

$$\Lambda(\vec{y}) = \frac{p_{\vec{Y}|H}(\vec{y} | 1)}{p_{\vec{Y}|H}(\vec{y} | 0)} \underset{\hat{H}=0}{\overset{\hat{H}=1}{\geq}} \frac{(C_{10} - C_{00})P_0}{(C_{01} - C_{11})P_1} = \eta \quad (3.47)$$

Note that the decision rule in (3.47) is again a threshold test, i.e., the decision depends only on a comparison of the likelihood ratio $\Lambda(\vec{y})$ with the threshold η . The only difference from our previous results is that η is now different. Note also that the costs affect the threshold only through the differences $C_{01} - C_{11}$ and $C_{10} - C_{00}$. This is to be expected, since we can view C_{11} as a fixed cost that occurs whenever hypothesis 1 occurs, and view the difference $C_{01} - C_{11}$ as the additional cost if we choose $\hat{H} = 0$.

Since the decision rule is a threshold test, $\Pr(e | H=0)$ and $\Pr(e | H=1)$ can again be found as before (from (3.7) and (3.8) or from (3.10) and (3.11)). The expected cost then consists of the fixed costs above plus the added costs when errors are made. Thus

$$\begin{aligned} E[C] &= P_0 C_{00} + P_1 C_{11} + P_0 \Pr(e | H=0)[C_{10} - C_{00}] \\ &\quad + P_1 \Pr(e | H=1)[C_{01} - C_{11}] \end{aligned} \quad (3.48)$$

The following theorem summarizes this.

Theorem 3.3 Assuming that $C_{01} - C_{11} > 0$, the Bayes rule, given in (3.47), minimizes the expected cost conditional on each observed sample value \vec{y} and also minimizes the overall expected cost. The expected cost is given in (3.48), with $\Pr(e | H=0) = \Pr(\Lambda(\vec{Y}) \geq \eta | H=0)$ and $\Pr(e | H=1) = \Pr(\Lambda(\vec{Y}) < \eta | H=0)$.

So far, we have looked at the minimum cost (Bayes) rule, the MAP rule, and the maximum likelihood (ML) rule. All of them are threshold tests where the decision is based on whether the likelihood ratio is above or below a threshold. For all of them, (3.7) to (3.11) determine the probability of error conditional on $H = 0$ and $H = 1$, and these quantities determine the other quantities of interest, as in (3.48). The next section treats how these error probabilities change as a function of the threshold.

3.5 Effect of Threshold Variation; the Neyman-Pearson Test

In a binary threshold test, the likelihood ratio, $\Lambda(\vec{y})$, is calculated from the observation \vec{y} . $\Lambda(\vec{y})$ is then compared with the threshold η and \hat{H} is chosen to be 1 or 0, depending on whether $\Lambda(\vec{y}) \geq \eta$ or $\Lambda(\vec{y}) < \eta$. Let $L = \Lambda(\vec{Y})$ be the likelihood random variable. An error occurs if either $H = 0$ and $L \geq \eta$ or if $H = 1$ and $L < \eta$. Since we want to compare different values of η , and since the error event depends on η , we denote the error event as $e(\eta)$ in

this section. When we want to look at some arbitrary non-threshold test A , we denote the error event as $e(A)$. We then have

$$\Pr(e(\eta) | H=0) = \Pr[L \geq \eta | H=0]; \quad \Pr(e(\eta) | H=1) = \Pr[L < \eta | H=1] \quad (3.49)$$

Assuming that L has a finite probability density under each hypothesis,

$$\frac{d\Pr(e(\eta) | H=0)}{d\eta} = -p_{L|H}(\eta | 0); \quad \frac{d\Pr(e(\eta) | H=1)}{d\eta} = p_{L|H}(\eta | 1) \quad (3.50)$$

We now relate the conditional probability densities $p_{L|H}(\eta | 1)$ and $p_{L|H}(\eta | 0)$. Ignoring terms of order smaller than δ for small δ ,

$$p_{L|H}(\eta | i)\delta = \int_{\vec{y}: \eta \leq \Lambda(\vec{y}) < \eta + \delta} p_{\vec{Y}|H}(\vec{y} | i) d\vec{y} \quad (3.51)$$

Taking the ratio of these terms for $i = 1$ and $i = 0$,

$$\frac{p_{L|H}(\eta | 1)}{p_{L|H}(\eta | 0)} = \lim_{\delta \rightarrow 0} \frac{\int_{\vec{y}: \eta \leq \Lambda(\vec{y}) < \eta + \delta} p_{\vec{Y}|H}(\vec{y} | 1) d\vec{y}}{\int_{\vec{y}: \eta \leq \Lambda(\vec{y}) < \eta + \delta} p_{\vec{Y}|H}(\vec{y} | 0) d\vec{y}}$$

Over the range of these integrals, $\eta p_{\vec{Y}|H}(\vec{y} | 0) \leq p_{\vec{Y}|H}(\vec{y} | 1) < (\eta + \delta) p_{\vec{Y}|H}(\vec{y} | 0)$. Thus the ratio of the integrals is between η and $\eta + \delta$. Going to the limit as δ approaches 0,

$$p_{L|H}(\eta | 1) = \eta p_{L|H}(\eta | 0) \quad (3.52)$$

Substituting this in (3.50), we get

$$\frac{d\Pr(e(\eta) | H=1)}{d\Pr(e(\eta) | H=0)} = -\eta \quad (3.53)$$

With (3.53), we can plot $\Pr(e(\eta) | H=1)$ as a parametric function of $\Pr(e(\eta) | H=0)$; we call this the error curve.³ As η is increased from 0 to ∞ , $\Pr(e(\eta) | H=0)$ decreases from 1 to 0, and $\Pr(e(\eta) | H=1)$ increases from 0 to 1. For any η , the slope of the error curve at $\Pr(e(\eta) | H=0)$ is $-\eta$. This is illustrated in Figure 3.4.

Since the slope of the error curve increases as η decreases, and hence as $\Pr(e(\eta) | H=0)$ increases, we see that the error curve is convex.

We can interpret the error curve in terms of MAP detection. For a priori probabilities P_1 and P_0 , an arbitrary test A has the error probability

$$\begin{aligned} \Pr(e(A)) &= P_1 \Pr(e(A) | H=1) + P_0 \Pr(e(A) | H=0) \\ &= P_1 [\Pr(e(A) | H=1) + \eta \Pr(e(A) | H=0)]; \quad \eta = P_0/P_1 \end{aligned}$$

³In the radar field, one often plots $1 - \Pr(e(\eta) | H=1)$ as a function of $\Pr(e(\eta) | H=0)$. This is called the receiver operating characteristic (ROC). If one flips the error curve vertically around the point 1/2, the ROC results.

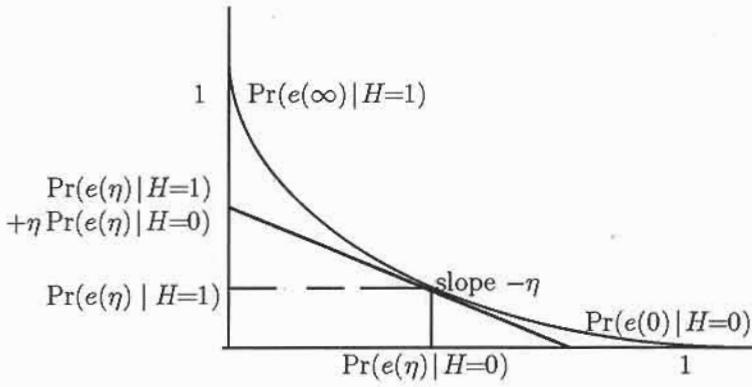


Figure 3.4: The error curve; $\Pr(e(\eta)|H=1)$ as a parametric function of $\Pr(e(\eta)|H=0)$

Here $\Pr(e(A) | H=i)$ is the probability of error when test A is used and the correct hypothesis is i . For arbitrary positive P_0, P_1 , the MAP test, i.e., the threshold test with $\eta = P_0/P_1$, minimizes $\Pr(e(A))$ over all tests, so for $\eta = P_0/P_1$ and for all A ,

$$P_1[\Pr(e(\eta) | H=1) + \eta \Pr(e(\eta) | H=0)] \leq P_1[\Pr(e(A) | H=1) + \eta \Pr(e(A) | H=0)] \quad (3.54)$$

Figure 3.4 interprets this equation. The vertical axis intercept of the tangent line is $\Pr(e(\eta) | H=1) + \eta \Pr(e(\eta) | H=0)$. Eq. (3.54) says that for any other test, the point $[\Pr(e(A) | H=0), \Pr(e(A) | H=1)]$ lies on or above that tangent line. Since this is true for all positive choices of a priori probabilities, and thus for all $\eta, 0 < \eta < \infty$, the point $[\Pr(e(A) | H=0), \Pr(e(A) | H=1)]$ lies on or above all the tangent lines, and thus lies on or above the curve of threshold tests. Thus the threshold tests are optimal in the sense of achieving the best tradeoffs between $\Pr(e(A) | H=1)$ and $\Pr(e(A) | H=0)$.

The *Neyman-Pearson* test is defined as the test A that minimizes $\Pr(e(A) | H=1)$ for a given maximum allowable value α for $\Pr(e(A) | H=0)$. This test is then a threshold test, choosing that value of η for which $\alpha = \Pr(e(\eta) | H=0)$. More explicitly, in terms of the error curve in Figure 3.4, we find the point α on the horizontal axis. The minimum value of $\Pr(e | H=1)$, for $\Pr(e | H=0) \leq \alpha$, is then the value of the error curve evaluated at α , and the threshold is the magnitude of the slope at that point.

The derivation of $\Pr(e(\eta) | H=1)$ as a function of $\Pr(e(\eta) | H=0)$ above depended on the assumption that the likelihood random variable L has a finite density everywhere. We next look at the more general case where the distribution function of L , conditional on $H=0$, contains discontinuities. This occurs both when \vec{Y} is discrete and also in cases like that shown in Figure 3.5. At a point of discontinuity, say η^* , we have $\Pr(L=\eta^* | H=0) > 0$. As in (3.51), we have $P_{L|H}(\eta^* | i) = \int_{\vec{y}: \eta^* = \Lambda(\vec{y})} p_{\vec{Y}|H}(\vec{y} | i) d\vec{y}$. Over the range of these integrals, $p_{\vec{Y}|H}(\vec{y} | 1) = \eta^* p_{\vec{Y}|H}(\vec{y} | 0)$. Thus, $\Pr(L=\eta^* | H=1) = \eta^* \Pr(L=\eta^* | H=0)$. The same result can be seen to hold when \vec{Y} does not have a density.

For a threshold test at η^* , the “don’t care” event, $L = \eta^*$, is significant, since it occurs with non-zero probability. Thus, as η goes from just below η^* to just above η^* , $\Pr(e(\eta) | H=0)$ jumps downward by $\Pr(L=\eta^* | H=0)$ and $\Pr(e(\eta) | H=1)$ jumps upward by $|\eta^*|$ times as much. This is illustrated in Figure 3.5d by a straight line portion, of slope $-\eta^*$, on the

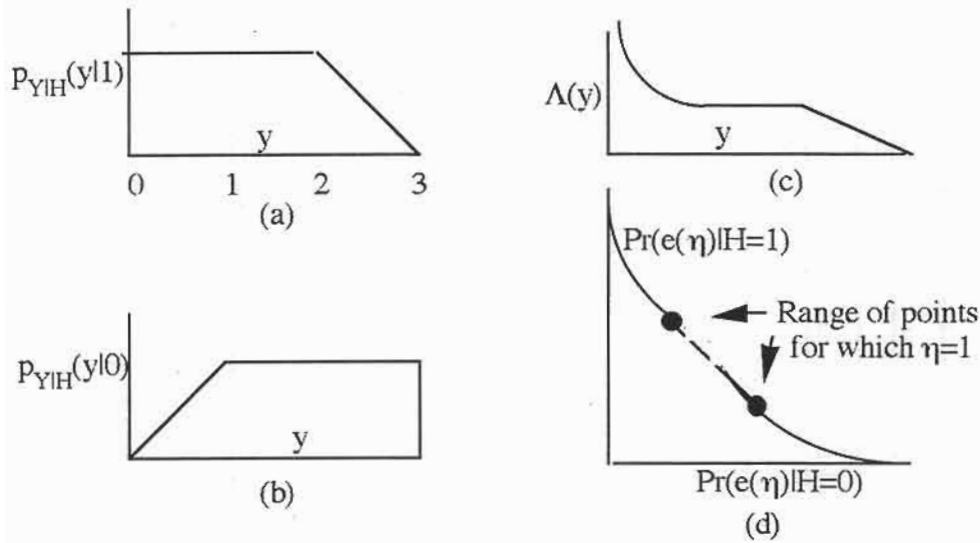


Figure 3.5: The two likelihood functions in (a) and (b) are equal for $1 \leq y \leq 2$, and the likelihood ratio is 1 over this region. For a threshold test with $\eta=1$, the region $1 \leq y \leq 2$ is a don't care region. As the fraction of those points mapped into \hat{H} increases, $\Pr(e | H=0)$ increases and $\Pr(e | H=1)$ decreases along the straight line indicated between the solid dots in part (d).

error curve. If the don't care cases with $L = \eta^*$ are all decided as $\hat{H} = 1$ (according to our convention as in (3.5)), the corresponding point on the error curve is at the lower right of the straight line portion. If the don't care cases are all decided as $\hat{H} = 0$, then the corresponding point is at the upper left of the straight line portion. As the fraction of those points decided as $\hat{H} = 1$ is increased from 0 to 1, we move from upper left to lower right on the straight line portion of the curve.

We have been referring to the event $L = \eta^*$, for the threshold test at η^* as a 'don't care case' because decisions in this case do not effect the overall error probability for the MAP rule with $P_0/P_1 = \eta^*$. We have now seen that if η^* is a point of discontinuity for L , then decisions in this case do effect the individual error probabilities $\Pr(e | H=i)$, $i = 0, 1$. To handle this systematically, we generalize our definition of a likelihood test to allow flexibility in handling the 'don't care cases'. In particular, define $\Pr(e(\eta, q) | H=0)$ as the probability of error, given $H=0$, for a threshold test at η for which, if $\Lambda(\bar{y}) = \eta$, the decision is $\hat{H}=1$ with probability q and $\hat{H}=0$ with probability $1 - q$. Then for an η^* at which L is discontinuous, $\Pr(e(\eta^*, q) | H=0)$ increases with q , and the corresponding $\Pr(e(\eta^*, q) | H=1)$ decreases with q , along the straight line portion of the error curve.

For a Neyman-Pearson test under the constraint $\Pr(e | H=0) \leq \alpha$, we find the point on the error curve where $\alpha = \Pr(e(\eta, q) | H=0)$. If the resulting η is at a point of discontinuity for L , then, when the event $L = \eta$ occurs, the decision $\hat{H} = 1$ is made with probability q and otherwise $\hat{H} = 0$. For situations such as that in Figure 3.5, this decision can be made according to the particular value of y . On the other hand, if Y is discrete, then some additional binary random variable X , with $P_X(1) = q$ is required to make the decision.

If Y is discrete, then $\Pr(e(\eta, q) | H=1)$, as a parametric function of $\Pr(e(\eta, q) | H=0)$, is piecewise linear. Each linear portion of the curve corresponds to a particular value η^* taken

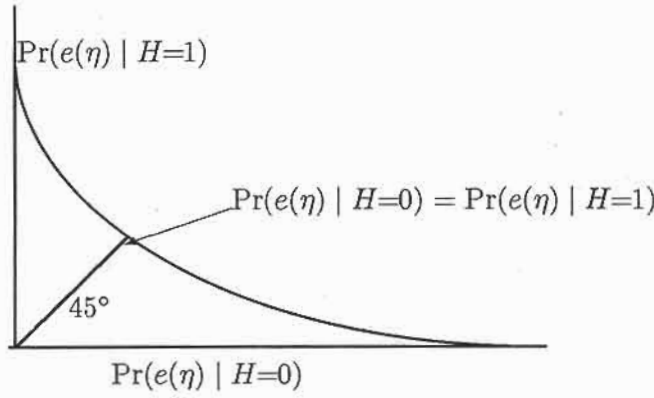


Figure 3.6: The minmax test. Increasing η from the point shown increases $\Pr(e | H=1, \eta)$, and decreasing η increases $\Pr(e | H=0, \eta)$.

on by the likelihood variable with non-zero probability. The point on that linear portion is then determined by q . The following theorem summarizes the above results

Theorem 3.4 The error curve is given by $\Pr(e(\eta, q) | H=1)$ as a parametric function in η and q of $\Pr(e(\eta, q) | H=0)$. The error curve is convex and has straight line segments of slope η^* at each η^* at which the distribution function of $L = \Lambda(\vec{Y})$ (conditional on $H=0$) has a discontinuity. For any test A , the point $(\Pr(e(A) | H=0), \Pr(e(A) | H=1))$ lies on or above the error curve.

There is one more interesting variation on the theme of threshold tests. If the a priori probabilities are unknown, we might want to minimize the maximum probability of error. That is, we visualize that we first choose a test, and then nature chooses H to maximize the probability of error. Our objective is to minimize the probability of error under the assumption that nature will then choose H to maximize $\Pr(e)$. The resulting test is called a minmax test. It can be seen geometrically from Figure 3.6 that the minmax test is the threshold test at the intersection of the error curve with a 45° line from the origin. If there is symmetry between $H = 0$ and $H = 1$, then the error curve will be symmetric around the 45° degree line, and the threshold will be at $\eta = 1$ (i. e., the ML test is also the minmax test).

3.6 Repeated Observations

Consider an experiment in which n identically distributed observations are made with the same hypothesis. That is, the observations $(Y_1, Y_2, \dots, Y_n) = \vec{Y}^T$ have the same conditional density $p_{\vec{Y}|H}(\vec{y} | H=i) = \prod_{k=1}^n p_{Y|H}(y_k | i)$ where $p_{Y|H}$ is the conditional density for each variable Y_1, \dots, Y_n . Example 3.2 above, with the additional constraint that $a_k = a_1$ and $b_k = b_1$ for $1 \leq k \leq n$ is an example of such repeated observations. This is also a reasonable model for many situations in which multiple noisy measurements are made of a given set

of alternatives. For the binary hypothesis case, the likelihood ratio is given by

$$\Lambda(\vec{y}) = \frac{\prod_{k=1}^n p_{Y|H}(y_k | 1)}{\prod_{k=1}^n p_{Y|H}(y_k | 0)} = \prod_{k=1}^n \frac{p_{Y|H}(y_k | 1)}{p_{Y|H}(y_k | 0)} \quad (3.55)$$

Taking the log of each side, we get the log likelihood ratio, $\text{LLR}(\vec{y})$,

$$\text{LLR}(\vec{y}) = \sum_{k=1}^n \ln \left[\frac{p_{Y|H}(y_k | 1)}{p_{Y|H}(y_k | 0)} \right] = \sum_{k=1}^n \text{LLR}(y_k) \quad (3.56)$$

where $\text{LLR}(y_k)$ is the log likelihood ratio of the k^{th} observation y_k ,

$$\text{LLR}(y_k) = \ln \left[\frac{p_{Y|H}(y_k | 1)}{p_{Y|H}(y_k | 0)} \right] \quad (3.57)$$

Thus, conditional on each hypothesis, $\text{LLR}(\vec{Y})$ is a sum of n IID random variables. For a threshold test with threshold η , define $\beta = \ln(\eta)$. Then

$$\Pr(e | H=1) = \Pr[\text{LLR}(\vec{Y}) < \beta | H=1] = \Pr \left[\sum_{k=1}^n \text{LLR}(Y_k) < \beta | H=1 \right] \quad (3.58)$$

$$\Pr(e | H=0) = \Pr[\text{LLR}(\vec{Y}) \geq \beta | H=0] = \Pr \left[\sum_{k=1}^n \text{LLR}(Y_k) \geq \beta | H=0 \right] \quad (3.59)$$

We could calculate these error probabilities for any given likelihood ratios and choice of n , but we are more interested in finding a simple upper bound to these quantities that shows what happens when n gets large. The appropriate tool here is the *Chernoff bound*, or *exponential bound*. Recall that the moment generating function (MGF) of a random variable U is $g_U(s) = \mathbb{E}[\exp(sU)]$. The Chernoff bound states that for any $s \geq 0$ such that $g_U(s)$ exists, and for any real number β ,

$$\Pr(U \geq \beta) \leq e^{-s\beta} g_U(s) \quad (3.60)$$

To derive this, assume that U has a density $p_U(u)$. Then

$$\begin{aligned} \Pr(U \geq \beta) &= \int_{u=\beta}^{\infty} p_U(u) du \leq \int_{u=\beta}^{\infty} e^{s(u-\beta)} p_U(u) du \\ &\leq \int_{-\infty}^{\infty} e^{s(u-\beta)} p_U(u) du = e^{-s\beta} g_U(s) \end{aligned}$$

where the first inequality follows because $\exp[s(u-\beta)] \geq 1$ for $u \geq \beta$, and the second follows because the integrand is non-negative for all u . Since this bound is valid for all $s \geq 0$, we have⁴

$$\Pr(U \geq \beta) \leq \min_{s \geq 0} e^{-s\beta} g_U(s) \quad (3.61)$$

⁴In a few very bizarre cases, this minimum has to be replaced with an infimum, but we won't have to worry about that here.

This bound can be rewritten in the following more convenient form:

$$\Pr(U \geq \beta) \leq \min_{s \geq 0} \exp[-s\beta + \ln(g_U(s))] \quad (3.62)$$

It can be shown that the second derivative of $-s\beta + \ln[g_U(s)]$ is non-negative over the values of s for which $g_U(s)$ exists. Thus we can minimize $-s\beta + \ln[g_U(s)]$ over s by setting the derivative with respect to s equal to 0. The minimum (over all s) thus occurs at $\beta = d[\ln(g_U(s))/ds]$. If $\beta \geq E[U]$, this minimum occurs for $s \geq 0$, so the s for which $\beta = d[\ln(g_U(s))/ds]$ minimizes (3.62). If $\beta < E[U]$, then (3.62) is minimized by $s=0$, leading to the trivial bound $\Pr(U \geq \beta) \leq 1$.

This bound is particularly convenient for sums of IID random variables. Let $U = V_1 + V_2 + \dots + V_n$, where V_1, \dots, V_n are IID. Then, as shown in Exercise 3.4, $g_U(s) = [g_V(s)]^n$. Thus

$$\Pr(U \geq \beta) \leq \min_{s \geq 0} \exp[-s\beta + n \ln(g_V(s))] \quad (3.63)$$

For the application at hand, let $g_0(s)$ be the MGF of $\text{LLR}(Y_k)$ under hypothesis 0, so

$$g_0(s) = \int_{-\infty}^{\infty} p_{Y|H}(y|0) \left[\frac{p_{Y|H}(y|1)}{p_{Y|H}(y|0)} \right]^s dy = \int_{-\infty}^{\infty} [p_{Y|H}(y|0)]^{1-s} [p_{Y|H}(y|1)]^s dy \quad (3.64)$$

Thus, for a threshold test with threshold η , $\Pr(e | H=0) \leq \exp[-s\eta + n \ln(g_0(s))]$ where $g_0(s)$ satisfies (3.64). By almost the same argument as used in deriving (3.60), there is an equivalent bound on $P(U \leq \beta)$. For any $r < 0$,

$$\Pr(U \leq \beta) \leq \exp[r\beta + \ln(g_U(r))] \quad (3.65)$$

If U is the sum of n IID rv's each with MGF $g(r)$, this becomes

$$\Pr(U \leq \beta) \leq \exp[r\beta + n \ln(g(r))] \quad (3.66)$$

For the application at hand, we let $g_1(r)$ be the MGF of $\text{LLR}(Y_k)$ under hypothesis 1, so

$$g_1(r) = \int_{-\infty}^{\infty} p_{Y|H}(y|1) \left[\frac{p_{Y|H}(y|1)}{p_{Y|H}(y|0)} \right]^r dy \quad (3.67)$$

$$= \int_{-\infty}^{\infty} [p_{Y|H}(y|0)]^{-r} [p_{Y|H}(y|1)]^{1+r} dy \quad (3.68)$$

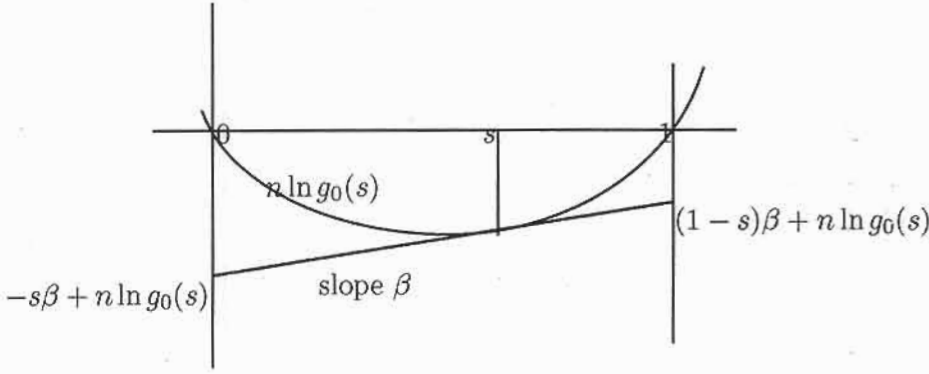
Note that $g_1(-1+s) = g_0(s)$, so we have the final answer:

$$\Pr(e | H=0) \leq \min_{0 \leq s \leq 1} \exp[-s\eta + n \ln(g_0(s))] \quad (3.69)$$

$$\Pr(e | H=1) \leq \min_{0 \leq s \leq 1} \exp[(1-s)\eta + n \ln(g_0(s))] \quad (3.70)$$

It can be seen from the figure that as n increases, the exponents increase approximately linearly with n . Eventually, the threshold η becomes insignificant. It can be shown that these bounds are asymptotically tight in the sense that for both $H = 1$ and $H = 0$,

$$\lim_{n \rightarrow \infty} \min_i \frac{\ln[\Pr(e | H=i)]}{n} = \min_{0 \leq s \leq 1} \ln[g_0(s)] \quad (3.71)$$

Figure 3.7: Exponents in Chernoff bound on $\Pr(e | H=0)$ and $\Pr(e | H=1)$

Exercise 3.12 shows that for example 3.2, with $a_k = a$ and $b_k = b$,

$$g_0(s) = \exp \left[-s(1-s)(b-a)^2/(2\sigma^2) \right], \quad (3.72)$$

so the bound becomes

$$\Pr(e | H=i) \leq \exp \left[-(b-a)^2/(8\sigma^2) \right] \quad (3.73)$$

As shown in Exercise 3.5, $Q(x)$ can be approximated closely for large x by $\frac{\exp(-x^2/2)}{\sqrt{2\pi}x}$ so this bound has the right exponent but ignores the term $1/[\sqrt{2\pi}x]$.

3.7 More than Two Hypotheses

Consider an hypothesis testing problem with m hypotheses. H is then a random variable with the possible values $0, 1, \dots, m-1$ and given a priori probabilities $P_i = P_H(i)$. Assume that C_{ij} is the cost of choosing i when $H = j$ and consider the Bayes test where \hat{H} is selected to minimize cost. As in (3.43), for an observed sample value \vec{y} of the observation $r\vec{V}\vec{Y}$,

$$E[\text{cost of } \hat{H}=i | \vec{Y}=\vec{y}] = \sum_{j=0}^{m-1} C_{ij} P_{H|\vec{Y}}(j | \vec{y}) \quad (3.74)$$

The minimum cost decision for the observed sample value \vec{y} is then

$$\hat{H} = \arg \min_i \sum_{j=0}^{m-1} C_{ij} P_{H|\vec{Y}}(j | \vec{y}) \quad (3.75)$$

As before, if the minimizing i is not unique, we follow the convention of choosing the largest minimizing i . Using Bayes' formula for $P_{H|\vec{Y}}(j | \vec{y})$ and canceling $p_{\vec{Y}}(\vec{y})$ since it is common to all terms in the minimization,

$$\hat{H} = \arg \min_i \sum_{j=0}^{m-1} C_{ij} P_j p_{\vec{Y}|H}(\vec{y} | j) \quad (3.76)$$

Finally, define the likelihood ratios, $\Lambda_i(\vec{y}) = p_{\vec{y}|H}(\vec{y} | i) / p_{\vec{y}|H}(\vec{y} | 0)$. With this definition, $\Lambda_0(\vec{y}) = 1$. The use of hypothesis 0 for normalization is arbitrary, but we assume it here. We then have

$$\hat{H} = \arg \min_i \left[C_{i0}P_0 + \sum_{j=1}^{m-1} C_{ij}P_j\Lambda_j(\vec{y}) \right] \quad (3.77)$$

Recall that for $m = 2$, the decision was simply a threshold test on the value of $\Lambda_1(\vec{y})$. Here the decision is based on the $m-1$ dimensional vector of likelihood ratios $(\Lambda_1(\vec{y}), \dots, \Lambda_{m-1}(\vec{y}))^T$. Let R_i be the region in this vector space mapped into a decision i . Then points in R_i are those points that satisfy

$$\left[C_{i0}P_0 + \sum_{j=1}^{m-1} C_{ij}P_j\Lambda_j(\vec{y}) \right] \leq \left[C_{k0}P_0 + \sum_{j=1}^{m-1} C_{kj}P_j\Lambda_j(\vec{y}) \right] /,; \quad \text{all } k < i$$

$$\left[C_{i0}P_0 + \sum_{j=1}^{m-1} C_{ij}P_j\Lambda_j(\vec{y}) \right] < \left[C_{k0}P_0 + \sum_{j=1}^{m-1} C_{kj}P_j\Lambda_j(\vec{y}) \right] ; \quad \text{all } k > i$$

This can be rewritten as

$$\left[(C_{i0} - C_{k0})P_0 + \sum_{j=1}^{m-1} (C_{ij} - C_{kj})P_j\Lambda_j(\vec{y}) \right] \leq 0; \quad \text{all } k < i \quad (3.78)$$

$$\left[(C_{i0} - C_{k0})P_0 + \sum_{j=1}^{m-1} (C_{ij} - C_{kj})P_j\Lambda_j(\vec{y}) \right] < 0; \quad \text{all } k > i \quad (3.79)$$

For each $k \neq i$, this inequality corresponds to a 'half space,' i.e., the set of points on one side of the affine space given by

$$\left[(C_{i0} - C_{k0})P_0 + \sum_{j=1}^{m-1} (C_{ij} - C_{kj})P_j\Lambda_j(\vec{y}) \right] = 0$$

Thus, R_i is the convex region bounded by this set of $m-1$ affine spaces. These affine spaces separate the various decision regions, but, for example, the affine space separating decision region 0 from region 1 depends on all the likelihood ratios, since all the likelihood ratios enter the decision (i.e., C_{12} might be much greater than C_{02} so that when Λ_2 is large, the decision is biased toward $\hat{H} = 0$). Not a great deal more can be said about these decision regions with this generality.

Many of the situations in which costs are important can be put in the form where $C_{ii} = 0$ for all i , and $C_{ij} = f_j$ for all $i \neq j$. That is, f_j is the cost of making an error when $H = j$, but it does not depend on which incorrect decision is made. In this case,

$$\sum_{j=0}^{m-1} C_{ij}P_j\Lambda_j(\vec{y}) = \sum_{j=0}^{m-1} f_jP_j\Lambda_j(\vec{y}) - f_iP_i\Lambda_i(\vec{y}) \quad (3.80)$$

Since the sum on the right hand side of (3.80) is common to all i , (3.77) becomes

$$\hat{H} = \arg \max_i f_i P_i \Lambda_i(\vec{y}) \quad (3.81)$$

This can be viewed as a set of binary threshold comparisons, i.e., for all $j, i, j > i$,

$$\frac{\Lambda_j(\vec{y})}{\Lambda_i(\vec{y})} = \frac{p_{\vec{y}|H}(\vec{y} | j)}{p_{\vec{y}|H}(\vec{y} | i)} \underset{\hat{H} \neq j}{\overset{\hat{H} \neq i}{\geq}} \frac{f_i P_i}{f_j P_j} = \eta_{ji} \quad (3.82)$$

It can be seen that, if $\Lambda_j(\vec{y}) > \Lambda_i(\vec{y})$ for all $i \neq j$, then (3.82) eliminates all hypotheses other than j , which then becomes the ML decision.

Example 3.4 Consider the same communication situation as in Figure 3.1, but assume the source produces one of m possible outputs, 0 to $m - 1$, and output i is mapped into $\vec{a}_i = (a_{i1}, a_{i2}, \dots, a_{in})$. Using the same analysis as in example 3.2, the analogous result to (3.27) is

$$(\vec{a}_j - \vec{a}_i)^T \vec{y} \underset{\hat{H} \neq j}{\overset{\hat{H} \neq i}{\geq}} \sigma^2 \ln(\eta_{ji}) + \frac{\vec{a}_j^T \vec{a}_j - \vec{a}_i^T \vec{a}_i}{2} \quad (3.83)$$

The geometric interpretation of this, in the space of observed vectors \vec{y} , is shown in Figure 3.8. The decision threshold between each pair of hypotheses is again an affine space perpendicular to the line joining the two signals. We also note that $(\vec{a}_j - \vec{a}_0)^T \vec{y}$ for $1 \leq j \leq m - 1$ is a sufficient statistic for this m -ary problem. Thus if the dimension n of the observed vectors \vec{y} is greater than $m - 1$, we can reduce the problem to $m - 1$ dimensions by transforming to a co-ordinate basis in which, for each i , $1 \leq i \leq m$, $\vec{a}_i - \vec{a}_0$ is a linear combination of $m - 1$ (or perhaps fewer) basis vectors. Using the theorem of irrelevance, the components of \vec{y} in all other directions can be ignored.

Even after the simplification of representing an additive Gaussian noise m -ary detection problem in the appropriate $m - 1$ or fewer dimensions, calculating the probability of error for each hypothesis can be messy. For example, in Figure 3.8, $\Pr(e | H=2)$ is the probability that the noise, added to \vec{a}_2 , carries the observation \vec{y} outside of the region where $\hat{H} = 2$.

This can be evaluated numerically using a two dimensional integral over the given constraint region. In typical problems of this type, however, the boundaries of the constraint region are several standard deviations away from \vec{a}_2 and it is often sufficient to provide a good upper bound to the error probability. The appropriate bound here is the *union bound*. That is, for any set of events, E_1, E_2, \dots, E_k ,

$$P \left(\bigcup_{j=1}^k E_j \right) \leq \sum_{j=1}^k P(E_j) \quad (3.84)$$

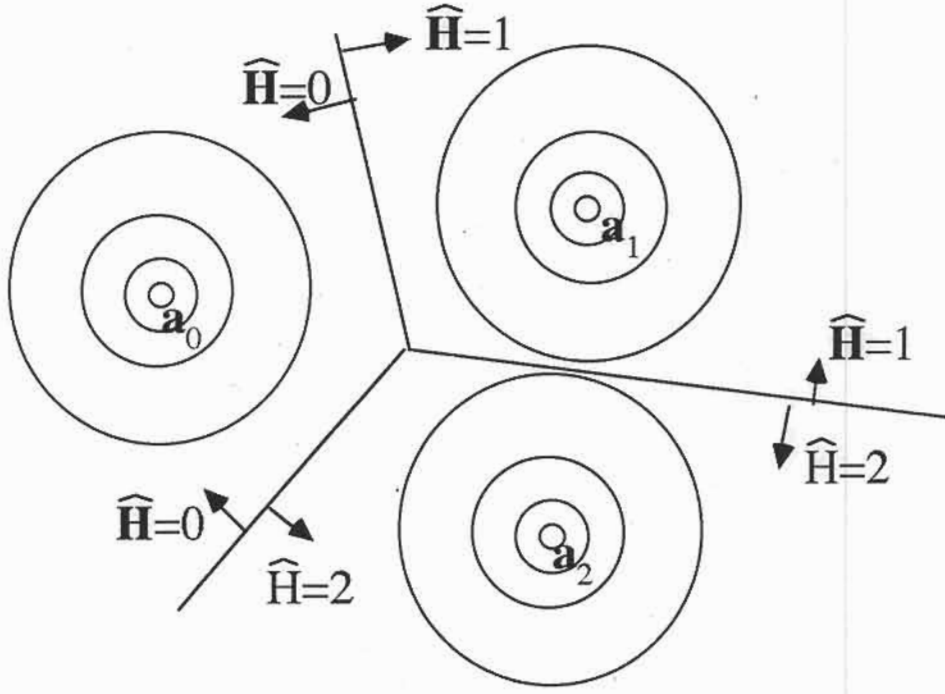


Figure 3.8: Decision regions for 3 signals in additive IID Gaussian noise.

For the problem at hand, the error event, conditional on $H = i$, is the union of the events that the individual binary thresholds are crossed. Thus, using (3.83),

$$\Pr(e | H=i) \leq \sum_{j \neq i} P \left((\vec{a}_j - \vec{a}_i)^T \vec{y} \geq \sigma^2 \ln(\eta_{ji}) + \frac{\vec{a}_j^T \vec{a}_j - \vec{a}_i^T \vec{a}_i}{2} \right) \quad (3.85)$$

Using (3.31) to evaluate the terms on the right hand side,

$$\Pr(e | H=i) \leq \sum_{j \neq i} Q \left(\frac{\sigma \ln(\eta_{ji})}{\|\vec{a}_j - \vec{a}_i\|} + \frac{\|\vec{a}_j - \vec{a}_i\|}{2\sigma} \right) \quad (3.86)$$

3.8 EXERCISES

- Exercise 3.1** a) Verify (3.31) from the fact that, conditional on $H = 0$, $\text{LLR}(\vec{Y}) \sim \mathcal{N}(-\gamma^2/2, \gamma^2)$.
b) Similarly, verify (3.32) from the fact that, conditional on $H = 1$, $\text{LLR}(\vec{Y}) \sim \mathcal{N}(\gamma^2/2, \gamma^2)$.
c) Define $U = (\vec{b} - \vec{a})^T \vec{Y}$. Find the mean and variance of U conditional on $H=0$. Use this, along with (3.27) to find $\Pr(e | H=0)$. Verify that your answer agrees with (3.31).
d) Note that U , as defined above, is a sufficient statistic. View the sample value u of U as the observation, and find the LLR of u .
e) Find $\Pr(e | H=0)$ and $\Pr(e | H=1)$ by applying (3.18) to this one dimensional problem.

Exercise 3.2 a) Let $U = (\vec{b} - \vec{a})^T K_{\vec{Z}}^{-1} \vec{Y}$. Find the conditional variance of U conditional on $H = 0$ and on $H = 1$.

b) Find $E[U | H=0]$ and $E[U | H=1]$.

c) Give the threshold test in terms of the sample value u of U , and evaluate $\Pr(e | H=0)$ and $\Pr(e | H=1)$ from this and part b). Show that your answer agrees with (3.41).

d) Explain what happens if $K_{\vec{Z}}$ is singular. Hint: you must look at two separate cases, depending on the vector $\vec{b} - \vec{a}$.

Exercise 3.3 Let \vec{Y} be the observation $r\vec{v}$ for a binary detection problem, let \vec{y} be the observed sample value. Let $v = f(\vec{y})$ be a sufficient statistic and let V be the corresponding random variable. Show that the likelihood ratio, $\Lambda(\vec{y})$, is equal to $p_{V|H}(f(\vec{y}) | 1) / p_{V|H}(f(\vec{y}) | 0)$. In other words, show that the likelihood ratio of a sufficient statistic is the same as the likelihood ratio of the original observation.

Exercise 3.4 Let $U = V_1 + \dots + V_n$ where V_1, \dots, V_n are IID rv's with the MGF $g_V(s)$. Show that $g_U(s) = [g_V(s)]^n$. Hint: You should be able to do this simply in a couple of lines.

Exercise 3.5 a) Consider example 3.3, and let $\vec{Z} = A\vec{W}$ where $\vec{W} \sim \mathcal{N}(0, I)$ is normalized IID Gaussian. The observation $r\vec{Y}$ is $\vec{a} + \vec{Z}$ given $H = 0$ and is $\vec{b} + \vec{Z}$ given $H = 1$. Suppose the observed sample value \vec{y} is transformed into $\vec{v} = A^{-1}\vec{y}$. Explain why \vec{v} is a sufficient statistic for this detection problem (and thus why MAP detection based on \vec{v} must yield the same decision as that based on \vec{y}).

b) Consider the detection problem where $\vec{V} = A^{-1}\vec{a} + \vec{W}$ given $H = 0$ and $A^{-1}\vec{b} + \vec{W}$ given $H = 1$. Find the log likelihood ratio $\text{LLR}(\vec{v})$ for a sample value \vec{v} of \vec{V} . Show that this is the same as the log likelihood ratio for a sample value $\vec{y} = A\vec{v}$ of \vec{Y} .

c) Find $\Pr(e | H=0)$ and $\Pr(e | H=1)$ for the detection problem in part b) by using the results of example 3.2. Show that your answer agrees with (3.41). Note: the methodology here is to transform the observed sample value to make the noise IID; this approach is often both useful and insightful, and we use it often in subsequent chapters.

Exercise 3.6 a) Calculate $g_0(s)$ as given in (3.64) for the decision problem in example 3.2 with $a_k = a$ and $b_k = b$, $1 \leq k \leq n$. Verify (3.72) and (3.73).

b) Upper bound $Q(x)$ by substituting $y = z - x$ for z as the variable of integration in the integral defining $Q(x)$ and then dropping the quadratic term in y . Explain why this results in a good approximation for large x (nothing very elaborate is expected here).

Exercise 3.7 Derive the union bound given in (3.84).

Exercise 3.8 Binary frequency shift keying (FSK) with incoherent reception can be modeled in terms of a 4 dimensional observation vector $\vec{Y} = (Y_1, Y_2, Y_3, Y_4)^T$. $\vec{Y} = \vec{X} + \vec{Z}$ where $\vec{Z} \sim \mathcal{N}(0, \sigma^2 I)$ and \vec{Z} is independent of \vec{X} . Under $H = 0$, $\vec{X} = (a \cos \phi, a \sin \phi, 0, 0)^T$, whereas under $H = 1$, $\vec{X} = (0, 0, a \cos \phi, a \sin \phi)^T$. The random variable ϕ is uniformly distributed between 0 and 2π and is independent of everything else. The a priori probabilities are $P_0 = P_1 = 1/2$.

a) Convince yourself from the circular symmetry of the situation that the ML receiver calculates the sample values v_0 and v_1 of $V_0 = Y_1^2 + Y_2^2$ and $V_1 = Y_3^2 + Y_4^2$ and chooses $\hat{H} = 0$ if $v_0 \geq v_1$ and chooses $\hat{H} = 1$ otherwise.

b) Find $\Pr(V_1 > v_1 \mid H=0)$ as a function of $v_1 > 0$.

c) Show that

$$p_{Y_1, Y_2 | H, \phi}(y_1, y_2 \mid 0, 0) = \frac{1}{2\pi\sigma^2} \exp \left[\frac{-y_1^2 - y_2^2 + 2y_1 a - a^2}{2\sigma^2} \right]$$

d) Show that

$$\Pr(V_1 > V_0 \mid H=0, \phi=0) = \int p_{Y_1, Y_2 | H, \phi}(y_1, y_2 \mid 0, 0) \Pr(V_1 > y_1^2 + y_2^2) dy_1 dy_2$$

Show that this is equal to $(1/2) \exp(-a^2/(4\sigma^2))$.

e) Explain why this is the probability of error (i.e., why the event $V_1 > V_0$ is independent of ϕ , and why $\Pr(e \mid H=0) = \Pr(e \mid H=1)$).

Exercise 3.9 Binary frequency shift keying (FSK) on a Rayleigh fading channel can be modeled in terms of a 4 dimensional observation vector $\vec{Y} = (Y_1, Y_2, Y_3, Y_4)^T$. $\vec{Y} = \vec{X} + \vec{Z}$ where $\vec{Z} \sim \mathcal{N}(0, \sigma^2 I)$ and \vec{Z} is independent of \vec{X} . Under $H = 0$, $\vec{X} = (X_1, X_2, 0, 0)^T$, whereas under $H = 1$, $\vec{X} = (0, 0, X_3, X_4)^T$. The random variables $X_i \sim \mathcal{N}(0, a^2)$ are IID. The a priori probabilities are $P_0 = P_1 = 1/2$.

a) Convince yourself from the circular symmetry of the situation that the ML receiver calculates sample values v_0 and v_1 for $V_0 = Y_1^2 + Y_2^2$ and $V_1 = Y_3^2 + Y_4^2$ and chooses $\hat{H} = 0$ if $v_0 \geq v_1$ and chooses $\hat{H} = 1$ otherwise.

b) Find $p_{V_0 | H}(v_0 \mid 0)$ and find $p_{V_1 | H}(v_1 \mid 0)$.

c) Let $U = V_0 - V_1$ and find $p_U(u \mid H=0)$.

d) Show that $\Pr(e \mid H=0) = [2 + a^2/\sigma^2]^{-1}$. Explain why this is also the unconditional probability of an incorrect decision.

Exercise 3.10 A disease has two strains, 0 and 1, which occur with a priori probabilities P_0 and P_1 respectively.

a) Initially, a rather noisy test was developed to test which strain is present for patients who are known to have one of the two strains. The output of the test is the sample value y_1 of a random variable Y_1 . Given strain 0 ($H=0$), $Y_1 = 5 + Z_1$, and given strain 1 ($H=1$), $Y_1 = 1 + Z_1$. The measurement noise Z_1 is independent of H and is Gaussian, $Z_1 \sim \mathcal{N}(0, \sigma^2)$. Give the MAP decision rule, i.e., determine the set of observations y_1 for which the decision is $\hat{H} = 1$. Give $\Pr(e | H=0)$ and $\Pr(e | H=1)$ in terms of the function $Q(x)$

b) A budding medical researcher determines that the test is making too many errors. A new measurement procedure is devised with two observation random variables Y_1 and Y_2 . Y_1 is the same as in part a). Y_2 , under hypothesis 0, is given by $Y_2 = 5 + Z_1 + Z_2$, and, under hypothesis 1, is given by $Y_2 = 1 + Z_1 + Z_2$. Assume that Z_2 is independent of both Z_1 and H , and that $Z_2 \sim \mathcal{N}(0, \sigma^2)$. Find the MAP decision rule for \hat{H} in terms of the joint observation (y_1, y_2) , and find $\Pr(e | H=0)$ and $\Pr(e | H=1)$. Hint: Find $p_{Y_2|Y_1, H}(y_2 | y_1, 0)$ and $p_{Y_2|Y_1, H}(y_2 | y_1, 1)$.

c) Explain in laymen's terms why the medical researcher should learn more about probability.

d) Now suppose that Z_2 , in part b), is uniformly distributed between 0 and 1 rather than being Gaussian. We are still given that Z_2 is independent of both Z_1 and H . Find the MAP decision rule for \hat{H} in terms of the joint observation (y_1, y_2) and find $\Pr(e | H=0)$ and $\Pr(e | H=1)$.

e) Finally, suppose that Z_1 is also uniformly distributed between 0 and 1. Again find the MAP decision rule and error probabilities.

Exercise 3.11 a) Consider a binary hypothesis testing problem, and denote the hypotheses as $H=1$ and $H=-1$. Let $\vec{a} = (a_1, a_2, \dots, a_n)^T$ be an arbitrary real n -vector and let the observation be a sample value \vec{y} of the random vector $\vec{Y} = H\vec{a} + \vec{Z}$ where $Z \sim \mathcal{N}(0, \sigma^2 I_n)$ and I_n is the n by n identity matrix. Assume that Z and H are independent. Find the maximum likelihood decision rule and find the probabilities of error $\Pr(e | H=0)$ and $\Pr(e | H=1)$ in terms of the function $Q(x)$.

b) Now suppose a third hypothesis, $H=0$, is added to the situation of part a). Again the observation random vector is $\vec{Y} = H\vec{a} + \vec{Z}$, but here H can take on values $-1, 0$, or $+1$. Find a one dimensional sufficient statistic for this problem (i.e., a one dimensional function of y from which the likelihood ratios

$$\Lambda_1(\vec{y}) = \frac{p_{Y|H}(\vec{y} | 1)}{p_{Y|H}(\vec{y} | 0)} \quad \text{and} \quad \Lambda_{-1}(\vec{y}) = \frac{p_{Y|H}(\vec{y} | -1)}{p_{Y|H}(\vec{y} | 0)}$$

can be calculated).

c) Find the maximum likelihood decision rule for the situation in part b) and find the probabilities of error, $\Pr(e | H=h)$ for $h = -1, 0, +1$.

d) Now suppose that Z_1, \dots, Z_n in part a) are IID and each is uniformly distributed over the interval -2 to $+2$. Also assume that $\vec{a} = (1, 1, \dots, 1)^T$. Find the maximum likelihood decision rule for this situation.

Exercise 3.12 Verify (3.72).