bounds on the error probabilities or approximate expressions for these probabilities. In this section we derive some simple bounds and approximations which are useful in many problems of practical importance. The basic results, due to Chernoff [28], were extended initially by Shannon [23]. They have been further extended by Fano [24], Shannon, Gallager, and Berlekamp [25], and Gallager [26] and applied to a problem of interest to us by Jacobs [27]. Our approach is based on the last two references. Because the latter part of the development is heuristic in nature, the interested reader should consult the references given for more careful derivations. From the standpoint of use in later sections, we shall not use the results until Chapter II-3 (the results are also needed for some of the problems in Chapter 4).

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

The likelihood ratio test is

$$l(\mathbf{R}) \triangleq \ln \Lambda(\mathbf{R}) = \ln \left[\frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)}\right] \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{440}$$

The variable $l(\mathbf{R})$ is a random variable whose probability density depends on which hypothesis is true. In Fig. 2.36 we show a typical $p_{l|H_1}(L|H_1)$ and $p_{l|H_0}(L|H_0)$.

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

$$P_D = \int_\gamma^\infty p_{l|H_1}(L|H_1)\, dL, \tag{441}$$

$$P_F = \int_\gamma^\infty p_{l|H_0}(L|H_0)\, dL. \tag{442}$$

The difficulty is that it is often hard to find $p_{l|H_i}(L|H_i)$, and even if it can be found it is cumbersome. Typical of this complexity is Case 1A

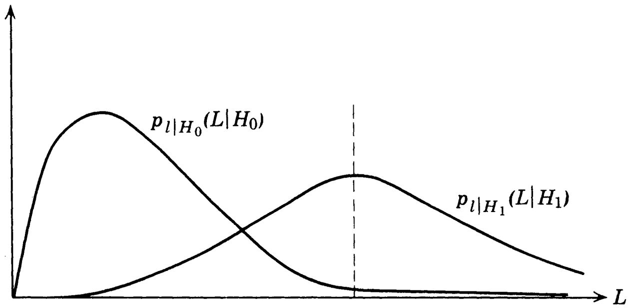


Fig. 2.36 Typical densities.

on p. 108, in which there are N Gaussian variables with equal variances making up the signal. To analyze a given system, the errors may be evaluated numerically. On the other hand, if we set out to synthesize a system, it is inefficient (if not impossible) to try successive systems and evaluate each numerically. Therefore we should like to find some simpler approximate expressions for the error probabilities.

In this section we derive some simple expressions that we shall use in the sequel. We first focus our attention on cases in which $l(\mathbf{R})$ is a sum of independent random variables. This suggests that its characteristic function may be useful, for it will be the product of the individual characteristic functions of the R_i. Similarly, the moment-generating function will be the product of individual moment-generating functions. Therefore an approximate expression based on one of these functions should be relatively easy to evaluate. The first part of our discussion develops bounds on the error probabilities in terms of the moment-generating function of $l(\mathbf{R})$.

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

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

$$\phi_{l|H_0}(s) \triangleq E(e^{sl}|H_0) = \int_{-\infty}^{\infty} e^{sL} p_{l|H_0}(L|H_0)\, dL, \tag{443}$$

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

$$\phi_{l|H_0}(s) \triangleq \exp\,[\mu(s)], \tag{444}$$

so that

$$\mu(s) = \ln \int_{-\infty}^{\infty} e^{sL} p_{l|H_0}(L|H_0)\, dL. \tag{445}$$

We may also express $\mu(s)$ in terms of $p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)$ and $p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)$. Because l is just a function of $\mathbf{r}$, we can write (443) as

$$\phi_{l|H_0}(s) = \int_{-\infty}^{\infty} e^{sl(\mathbf{R})} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\, d\mathbf{R}. \tag{446}$$

Then

$$\mu(s) = \ln \int_{-\infty}^{\infty} e^{sl(\mathbf{R})} p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\, d\mathbf{R}. \tag{447}$$

Using (440),

$$\mu(s) = \ln \int_{-\infty}^{\infty} \left(\frac{p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)}\right)^s p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)\, d\mathbf{R}, \tag{448}$$

or

$$\boxed{\mu(s) = \ln \int_{-\infty}^{\infty} [p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)]^s [p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)]^{1-s}\, d\mathbf{R}.} \tag{449}$$

The function $\mu(s)$ plays a central role in the succeeding discussion. It is now convenient to rewrite the error expressions in terms of a new random variable whose mean is in the vicinity of the threshold. The reason for this step is that we shall use the central limit theorem in the second part of our derivation. It is most effective near the mean of the random variable of interest. Consider the simple probability density shown in Fig. 2.37*a*. To get the new family of densities shown in Figs. 2.37*b*–*d* we multiply $p_x(X)$ by e^{sX} for various values of s (and normalize to obtain a unit area). We see that for $s > 0$ the mean is shifted to the right. For the moment we leave s as a parameter. We see that increasing s "tilts" the density more.

Denoting this new variable as x_s, we have

$$p_{x_s}(X) \triangleq \frac{e^{sX} p_{l|H_0}(X|H_0)}{\displaystyle\int_{-\infty}^{\infty} e^{sL} p_{l|H_0}(L|H_0)\, dL} = \frac{e^{sX} p_{l|H_0}(X|H_0)}{e^{\mu(s)}}. \tag{450}$$

Observe that we define x_s in terms of its density function, for that is what we are interested in. Equation 450 is a general definition. For the density shown in Fig. 2.37, the limits would be $(-A, A)$.

We now find the mean and variance of x_s:

$$E(x_s) = \int_{-\infty}^{\infty} X p_{x_s}(X)\, dX = \frac{\displaystyle\int_{-\infty}^{\infty} X e^{sX} p_{l|H_0}(X|H_0)\, dX}{\displaystyle\int_{-\infty}^{\infty} e^{sL} p_{l|H_0}(L|H_0)\, dL}. \tag{451}$$

Comparing (451) and (445), we see that

$$E(x_s) = \frac{d\mu(s)}{ds} \triangleq \dot{\mu}(s). \tag{452}$$

Similarly, we find

$$\operatorname{Var}(x_s) = \ddot{\mu}(s). \tag{453}$$

[Observe that (453) implies that $\mu(s)$ is convex.]

We now rewrite P_F in terms of this tilted variable x_s:

$$P_F = \int_\gamma^\infty p_{l|H_0}(L|H_0)\,dL = \int_\gamma^\infty e^{\mu(s)-sX} p_{x_s}(X)\,dX$$
$$= e^{\mu(s)} \int_\gamma^\infty e^{-sX} p_{x_s}(X)\,dX. \tag{454}$$

We can now find a simple upper bound on P_F. For values of $s \geq 0$,

$$e^{-sX} \leq e^{-s\gamma}, \qquad \text{for } X \geq \gamma. \tag{455}$$

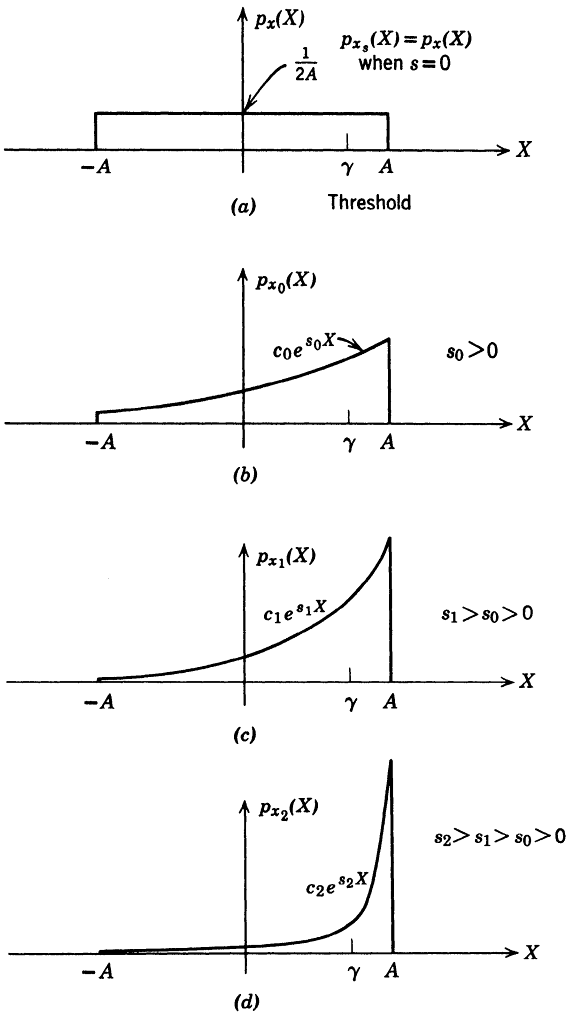


Fig. 2.37 Tilted probability densities.

Thus

$$P_F \le e^{\mu(s) - s\gamma} \int_{\gamma}^{\infty} p_{x_s}(X)\,dX, \qquad s \ge 0. \tag{456}$$

Clearly the integral is less than one. Thus

$$P_F \le e^{\mu(s) - s\gamma}, \qquad s \ge 0. \tag{457}$$

To get the best bound we minimize the right-hand side of (457) with respect to s. Differentiating the exponent and setting the result equal to zero, we obtain

$$\dot{\mu}(s) = \gamma. \tag{458}$$

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

$$\dot{\mu}(0) \le \gamma \le \dot{\mu}(\infty). \tag{459}$$

Because

$$\dot{\mu}(0) = E(l \mid H_0), \tag{460}$$

the left inequality implies that the threshold must be to the right of the mean of l on H_0. Assuming that (459) is valid, we have the desired result:

$$P_F \le \exp\,[\mu(s) - s\dot{\mu}(s)], \qquad s \ge 0, \tag{461}$$

where s satisfies (458). (We have assumed $\mu(s)$ exists for the desired s.)

Equation 461 is commonly referred to as the Chernoff bound [28]. Observe that s is chosen so that the mean of the tilted variable x_s is at the threshold.

The next step is find a bound on P_M, the probability of a miss:

$$P_M = \int_{-\infty}^{\gamma} p_{l \mid H_1}(X \mid H_1)\,dX, \tag{462}$$

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

Using an argument identical to that in (88) through (94), we see that

$$p_{l \mid H_1}(X \mid H_1) = e^{X} p_{l \mid H_0}(X \mid H_0). \tag{463}$$

Substituting (463) into the right side of (450), we have

$$p_{l \mid H_1}(X \mid H_1) = e^{\mu(s) + (1-s)X} p_{x_s}(X). \tag{464}$$

Substituting into (462),

$$P_M = e^{\mu(s)} \int_{-\infty}^{\gamma} e^{(1-s)X} p_{x_s}(X)\,dX. \tag{465}$$

For $s \le 1$

$$e^{(1-s)X} \le e^{(1-s)\gamma}, \qquad \text{for } X \le \gamma. \tag{466}$$

Thus

$$\begin{aligned} P_M &\le e^{\mu(s) + (1-s)\gamma} \int_{-\infty}^{\gamma} p_{x_s}(X)\,dX \\ &\le e^{\mu(s) + (1-s)\gamma}, \qquad s \le 1. \end{aligned} \tag{467}$$

Once again the bound is minimized for

$$\gamma = \dot{\mu}(s) \tag{468}$$

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

$$\dot{\mu}(1) = E(l|H_1), \tag{469}$$

we see that this requires the threshold to be to the left of the mean of l on H_1.

Combining (461) and (467), we have

$$\boxed{\begin{aligned} P_F &\leq \exp[\mu(s) - s\dot{\mu}(s)] \\ & \qquad\qquad\qquad\qquad 0 \leq s \leq 1 \\ P_M &\leq \exp[\mu(s) + (1-s)\dot{\mu}(s)] \end{aligned}} \tag{470}$$

and

$$\gamma = \dot{\mu}(s)$$

is the threshold that lies *between* the means of l on the two hypotheses. Confining s to [0, 1] is not too restrictive because if the threshold is not between the means the error probability will be large on one hypothesis (greater than one half if the median coincides with the mean). If we are modeling some physical system this would usually correspond to unacceptable performance and necessitate a system change.

As pointed out in [25], the exponents have a simple graphical interpretation. A typical $\mu(s)$ is shown in Fig. 2.38. We draw a tangent at the point at which $\dot{\mu}(s) = \gamma$. This tangent intersects vertical lines at $s = 0$ and $s = 1$. The value of the intercept at $s = 0$ is the exponent in the P_F bound. The value of the intercept at $s = 1$ is the exponent in the P_M bound.

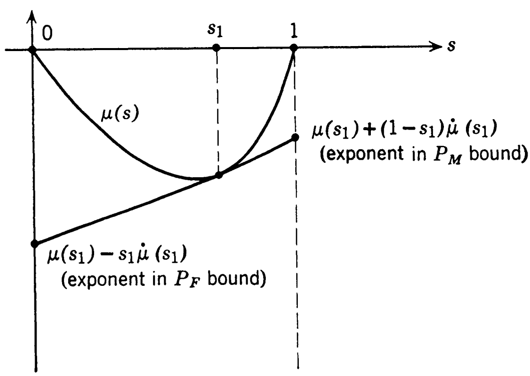


Fig. 2.38 Exponents in bounds.

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

The probability of error Pr (ϵ) is

$$\Pr(\epsilon) = \tfrac{1}{2}P_F + \tfrac{1}{2}P_M. \tag{471}$$

Substituting (456) and (467) into (471) and denoting the value s for which $\dot{\mu}(s) = 0$ as s_m, we have

$$\Pr(\epsilon) \leq \tfrac{1}{2}e^{\mu(s_m)} \int_0^{\infty} p_{x_s}(X)\, dX + \tfrac{1}{2}e^{\mu(s_m)} \int_{-\infty}^{0} p_{x_s}(X)\, dX, \tag{472}$$

or

$$\boxed{\Pr(\epsilon) \leq \tfrac{1}{2}e^{\mu(s_m)}.} \tag{473}$$

Up to this point we have considered arbitrary binary hypothesis tests. The bounds in (470) and (473) are always valid if $\mu(s)$ exists. In many cases of interest $l(\mathbf{R})$ consists of a sum of a large number of independent random variables, and we can obtain a simple approximate expression for P_F and P_M that provides a much closer estimate of their actual value than the above bounds. The exponent in this expression is the same, but the multiplicative factor will often be appreciably smaller than unity.

We start the derivation with the expression for P_F given in (454). Motivated by our result in the bound derivation (458), we choose s so that

$$\dot{\mu}(s) = \gamma.$$

Then (454) becomes

$$P_F = e^{\mu(s)} \int_{\dot{\mu}(s)}^{\infty} e^{-sX} p_{x_s}(X)\, dX. \tag{474}$$

This can be written as

$$P_F = e^{\mu(s) - s\dot{\mu}(s)} \int_{\dot{\mu}(s)}^{\infty} e^{+s[\dot{\mu}(s) - X]} p_{x_s}(X)\, dX. \tag{475}$$

The term outside is just the bound in (461). We now use a central limit theorem argument to evaluate the integral. First define a standardized variable:

$$y \triangleq \frac{x_s - E(x_s)}{(\mathrm{Var}\,[x_s])^{1/2}} = \frac{x_s - \dot{\mu}(s)}{\sqrt{\ddot{\mu}(s)}}. \tag{476}$$

Substituting (476) into (475), we have

$$P_F = e^{\mu(s) - s\dot{\mu}(s)} \int_0^{\infty} e^{-s\sqrt{\ddot{\mu}(s)}\, Y} p_y(Y)\, dY. \tag{477}$$

In many cases the probability density governing $\mathbf{r}$ is such that y approaches a Gaussian random variable as N (the number of components of $\mathbf{r}$) approaches infinity.† A simple case in which this is true is the case in which the r_i are independent, identically distributed random variables with finite means and variances. In such cases, y approaches a zero-mean Gaussian random variable with unit variance and the integral in (477) can be evaluated by substituting the limiting density.

$$\int_0^\infty e^{-s\sqrt{\ddot{\mu}(s)}\,Y} \frac{1}{\sqrt{2\pi}} e^{-(Y^2/2)}\, dY = e^{s^2\ddot{\mu}(s)/2} \operatorname{erfc}_* (s\sqrt{\ddot{\mu}(s)}). \tag{478}$$

Then

$$P_F \simeq \left\{\exp\left[\mu(s) - s\dot{\mu}(s) + \frac{s^2}{2}\ddot{\mu}(s)\right]\right\} \operatorname{erfc}_* [s\sqrt{\ddot{\mu}(s)}]. \tag{479}$$

The approximation arises because y is only approximately Gaussian for finite N. For values of $s\sqrt{\ddot{\mu}(s)} > 3$ we can approximate $\operatorname{erfc}_*(\cdot)$ by the upper bound in (71). Using this approximation,

$$\boxed{P_F \simeq \frac{1}{\sqrt{2\pi s^2 \ddot{\mu}(s)}} \exp\,[\mu(s) - s\dot{\mu}(s)], \qquad s \geq 0.} \tag{480}$$

It is easy to verify that the approximate expression in (480) can also be obtained by letting

$$p_y(Y) \simeq p_y(0) \simeq \frac{1}{\sqrt{2\pi}}. \tag{481}$$

Looking at Fig. 2.39, we see that this is valid when the exponential function decreases to a small value while $Y \ll 1$.

In exactly the same manner we obtain

$$P_M \simeq \left\{\exp\left[\mu(s) + (1 - s)\,\dot{\mu}(s) + \frac{(s-1)^2}{2}\ddot{\mu}(s)\right]\right\} \operatorname{erfc}_* [(1 - s)\sqrt{\ddot{\mu}(s)}]. \tag{482}$$

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

$$\boxed{P_M \simeq \frac{1}{\sqrt{2\pi(1-s)^2\ddot{\mu}(s)}} \exp\,[\mu(s) + (1 - s)\dot{\mu}(s)], \qquad s \leq 1.} \tag{483}$$

Observe that the exponents in (480) and (483) are identical to those obtained by using the Chernoff bound. The central limit theorem argument has provided a multiplicative factor that will be significant in many of the applications of interest to us.

† An excellent discussion is contained in Feller [33], pp. 517–520.

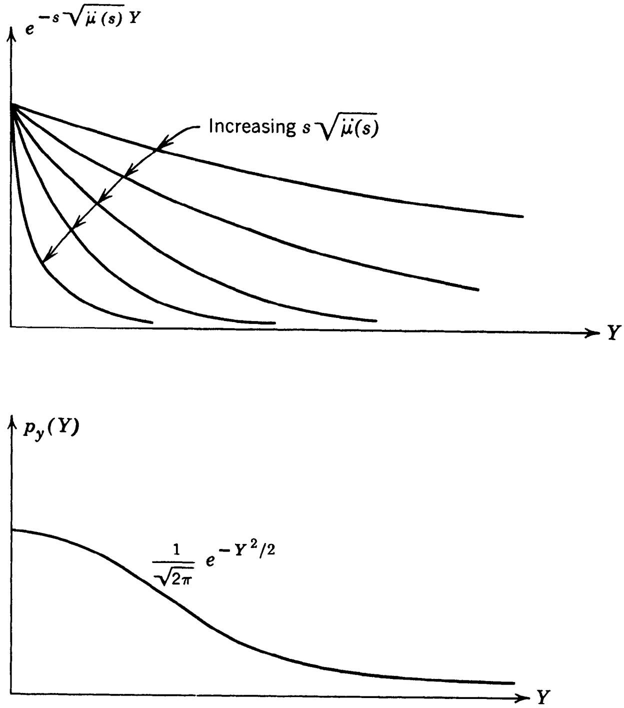


Fig. 2.39 Behavior of functions.

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

$$\begin{aligned}
\Pr(\epsilon) &= \tfrac{1}{2}P_F + \tfrac{1}{2}P_M \\
&= \tfrac{1}{2}\exp\left[\mu(s_m) + \frac{s_m^{\ 2}}{2}\,\ddot{\mu}(s_m)\right] \operatorname{erfc}_*\left[s_m\sqrt{\ddot{\mu}(s_m)}\right] \\
&\quad + \tfrac{1}{2}\exp\left[\mu(s_m) + \frac{(1-s_m)^2}{2}\,\ddot{\mu}(s_m)\right] \operatorname{erfc}_*\left[(1-s_m)\sqrt{\ddot{\mu}(s_m)}\right],
\end{aligned} \tag{484}$$

where s_m is defined in the statement preceding (472) [i.e., $\dot{\mu}(s_m) = 0 = \gamma$]. When both $s_m\sqrt{\ddot{\mu}(s_m)} > 3$ and $(1 - s_m)\sqrt{\ddot{\mu}(s_m)} > 3$, this reduces to

$$\boxed{\Pr(\epsilon) \simeq \frac{1}{[2(2\pi\ddot{\mu}(s_m))^{1/2}s_m(1-s_m)]}\exp\mu(s_m).} \tag{485}$$

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

Example 1. In this example we consider the simple Gaussian problem first introduced on p. 27:

$$p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(R_i - m)^2}{2\sigma^2}\right] \tag{486}$$

and

$$p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{R_i^2}{2\sigma^2}\right). \tag{487}$$

Then, using (449)

$$\mu(s) = \ln \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(R_i - m)^2 s + R_i^2(1 - s)}{2\sigma^2}\right] dR_i. \tag{488a}$$

Because all the integrals are identical,

$$\mu(s) = N \ln \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(R - m)^2 s + R^2(1 - s)}{2\sigma^2}\right] dR. \tag{488b}$$

Integrating we have

$$\mu(s) = Ns(s - 1)\frac{m^2}{2\sigma^2} \triangleq \frac{s(s - 1)d^2}{2}, \tag{489}$$

where d^2 was defined in the statement after (64). The curve is shown in Fig. 2.40:

$$\dot{\mu}(s) = \frac{(2s - 1)d^2}{2}. \tag{490}$$

Using the bounds in (470), we have

$$\left.\begin{aligned} P_F &\le \exp\left(\frac{-s^2 d^2}{2}\right) \\ P_M &\le \exp\left[-\frac{(1 - s)^2 d^2}{2}\right] \end{aligned}\right. \qquad 0 \le s \le 1. \tag{491}$$

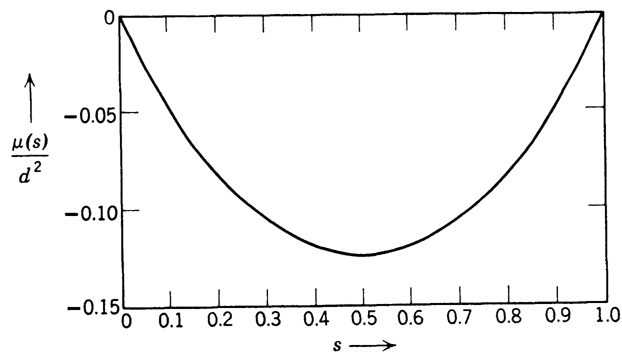


Fig. 2.40 $\mu(s)$ **for Gaussian variables with unequal means.**

Because $l(\mathbf{R})$ is the sum of Gaussian random variables, the expressions in (479) and (482) are exact. Evaluating $\ddot{\mu}(s)$, we obtain

$$\ddot{\mu}(s) = d^2. \tag{492}$$

Substituting into (479) and (482), we have

$$P_F = \operatorname{erfc}_* [s\sqrt{\ddot{\mu}(s)}] = \operatorname{erfc}_* (sd) \tag{493}$$

and

$$P_M = \operatorname{erfc}_* [(1 - s)\sqrt{\ddot{\mu}(s)}] = \operatorname{erfc}_* [(1 - s)d]. \tag{494}$$

These expressions are identical to (64) and (68) (let $s = (\ln \eta)/d^2 + \frac{1}{2}$).

An even simpler case is one in which the total probability of error is the criterion. Then we choose an s_m such as $\dot{\mu}(s_m) = 0$. From Fig. 2.40, we see that $s_m = \frac{1}{2}$. Using (484) and (485) we have

$$\Pr(\epsilon) = \operatorname{erfc}_* \left(\frac{d}{2}\right) \simeq \left(\frac{2}{\pi d^2}\right)^{1/2} \exp\left(-\frac{d^2}{8}\right), \tag{495}$$

where the approximation is very good for $d > 6$.

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

$$\mu(\tfrac{1}{2}) = \ln \int_{-\infty}^{\infty} [p_{\mathbf{r}|H_1}(\mathbf{R}|H_1)]^{1/2}[p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)]^{1/2}\, d\mathbf{R}. \tag{496}$$

The negative of this quantity is frequently referred to as the Bhattacharyya distance (e.g., [29]). It is important to note that it is the significant quantity only when $s_m = \frac{1}{2}$.

In our next example we look at a more interesting case.

Example 2. This example is Case 1A of the general Gaussian problem described on p. 108:

$$\begin{aligned} p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{R_i^2}{2\sigma_1^2}\right), \\ p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{R_i^2}{2\sigma_0^2}\right). \end{aligned} \tag{497}$$

Substituting (497) into (499) gives,

$$\mu(s) = N \ln \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi}\,\sigma_1^s \sigma_0^{1-s})} \exp\left[-\frac{sR^2}{2\sigma_1^2} - \frac{(1-s)R^2}{2\sigma_0^2}\right] dR \tag{498}$$

or

$$\mu(s) = \frac{N}{2} \ln \left[\frac{(\sigma_0^2)^s (\sigma_1^2)^{1-s}}{s\sigma_0^2 + (1 - s)\sigma_1^2}\right]. \tag{499}$$

A case that will be of interest in the sequel is

$$\begin{aligned} \sigma_1^2 &= \sigma_n^2 + \sigma_s^2, \\ \sigma_0^2 &= \sigma_n^2. \end{aligned} \tag{500}$$

Substituting (500) into (499) gives

$$\frac{\mu(s)}{N/2} = \left\{(1 - s)\ln\left(1 + \frac{\sigma_s^2}{\sigma_n^2}\right) - \ln\left[1 + (1 - s)\frac{\sigma_s^2}{\sigma_n^2}\right]\right\}. \tag{501}$$

This function is shown in Fig. 2.41.

$$\dot{\mu}(s) = \frac{N}{2}\left[-\ln\left(1 + \frac{\sigma_s^2}{\sigma_n^2}\right) + \frac{\sigma_s^2/\sigma_n^2}{1 + (1 - s)\sigma_s^2/\sigma_n^2}\right] \tag{502}$$

and

$$\ddot{\mu}(s) = \frac{N}{2}\left[\frac{\sigma_s^2/\sigma_n^2}{1 + (1 - s)(\sigma_s^2/\sigma_n^2)}\right]^2. \tag{503}$$

By substituting (501), (502), and (503) into (479) and (482) we can plot an approximate receiver operating characteristic. This can be compared with the exact ROC in Fig. 2.35*a* to estimate the accuracy of the approximation. In Fig. 2.42 we show the comparison for $N = 4$ and 8, and $\sigma_s^2/\sigma_n^2 = 1$. The lines connect the equal threshold points. We see that the approximation is good. For larger N the exact and approximate ROC are identical for all practical purposes.

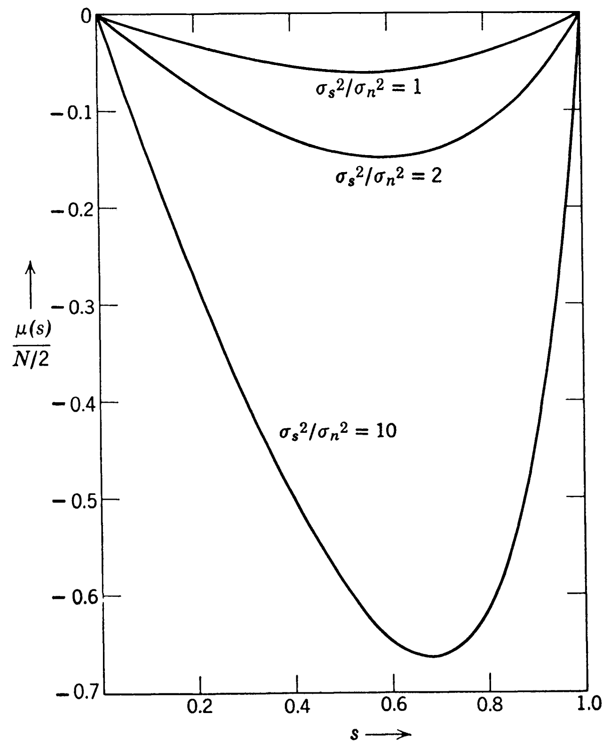


Fig. 2.41 **$\mu(s)$ for Gaussian variables with unequal variances.**

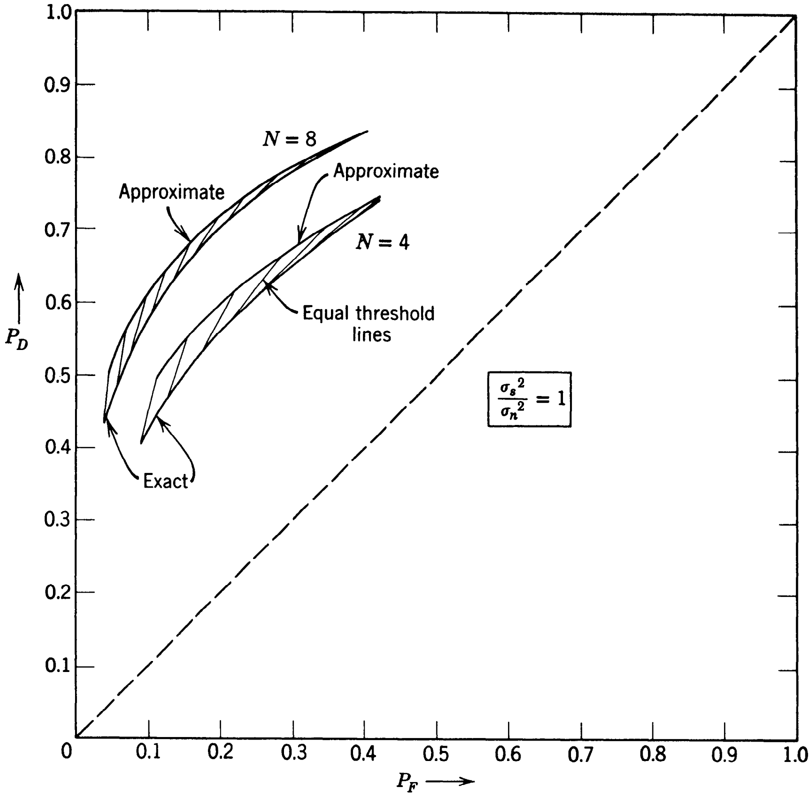


Fig. 2.42 Approximate receiver operating characteristics.

Example 3. In this example we consider first the simplified version of the symmetric hypothesis situation described in Case 2A (p. 115) in which $N = 2$.

$$p_{\mathbf{r}|H_1}(\mathbf{R}|H_1) = \frac{1}{(2\pi)^2\sigma_1^2\sigma_0^2} \exp\left(-\frac{R_1^2 + R_2^2}{2\sigma_1^2} - \frac{R_3^2 + R_4^2}{2\sigma_0^2}\right) \tag{504}$$

and

$$p_{\mathbf{r}|H_0}(\mathbf{R}|H_0) = \frac{1}{(2\pi)^2\sigma_1^2\sigma_0^2} \exp\left(-\frac{R_1^2 + R_2^2}{2\sigma_0^2} - \frac{R_3^2 + R_4^2}{2\sigma_1^2}\right), \tag{505}$$

where

$$\begin{aligned} \sigma_1^2 &= \sigma_s^2 + \sigma_n^2 \\ \sigma_0^2 &= \sigma_n^2. \end{aligned} \tag{506}$$

Then

$$\begin{aligned} \mu(s) &= s \ln \sigma_n^2 + (1 - s) \ln (\sigma_n^2 + \sigma_s^2) - \ln (\sigma_n^2 + \sigma_s^2 s) \\ &\quad + (1 - s) \ln \sigma_n^2 + s \ln (\sigma_n^2 + \sigma_s^2) - \ln [\sigma_n^2 + \sigma_s^2(1 - s)] \\ &= \ln\left(1 + \frac{\sigma_s^2}{\sigma_n^2}\right) - \ln\left[\left(1 + \frac{s\sigma_s^2}{\sigma_n^2}\right)\left(1 + \frac{(1 - s)\sigma_s^2}{\sigma_n^2}\right)\right]. \end{aligned} \tag{507}$$

The function $\mu(s)$ is plotted in Fig. 2.43a. The minimum is at $s = \frac{1}{2}$. This is the point of interest at which minimum Pr (ϵ) is the criterion.

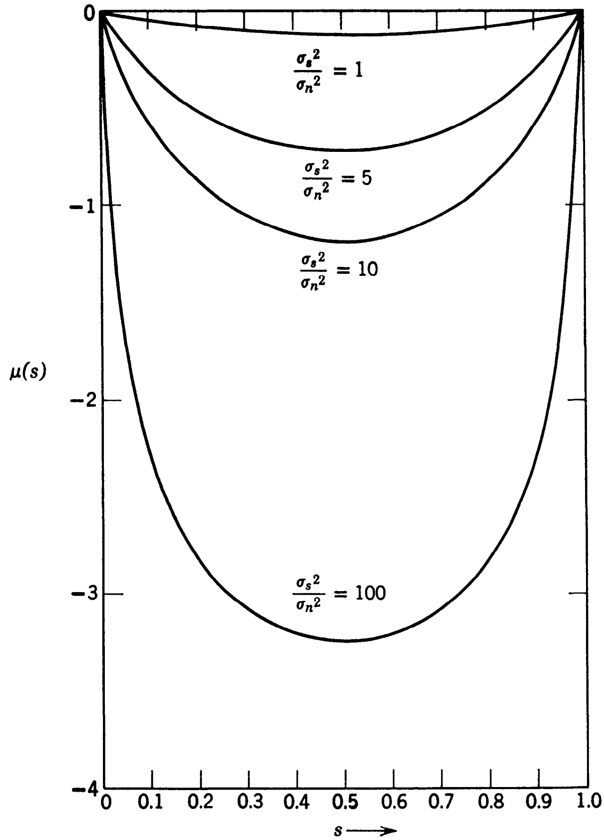


Fig. 2.43*a* **$\mu(s)$ for the binary symmetric hypothesis problem.**

Thus from (473), a *bound* on the error is,

$$\Pr(\epsilon) \leq \frac{1}{2}\,\frac{(1 + \sigma_s^2/\sigma_n^2)}{(1 + \sigma_s^2/2\sigma_n^2)^2}. \tag{508}$$

The bound in (508) is plotted in Fig. 2.43*b*.

Example 3A. An interesting extension of Example 3 is the problem in which

$$\mathbf{K}_s = \begin{bmatrix} \sigma_1^2 & & & & & & & \\ & \sigma_1^2 & & & & & 0 & \\ & & \sigma_2^2 & & & & & \\ & & & \sigma_2^2 & & & & \\ & & & & \sigma_3^2 & & & \\ & & & & & \ddots & & \\ & & 0 & & & & \sigma_{N/2}^2 & \\ & & & & & & & \sigma_{N/2}^2 \end{bmatrix}. \tag{509}$$

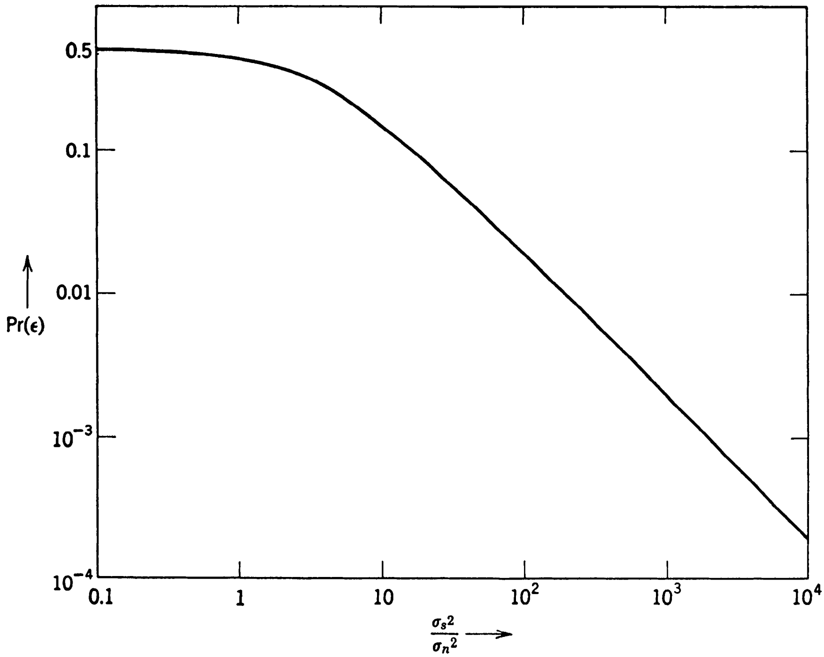


Fig. 2.43*b* Bound on the probability of error (Pr(ε)).

The r_i's are independent variables and their variances are pairwise equal. This is a special version of Case 2B on p. 115. We shall find later that it corresponds to a physical problem of appreciable interest.

Because of the independence, $\mu(s)$ is just the sum of the $\mu(s)$ for each pair, but each pair corresponds to the problem in Example 3. Therefore

$$\mu(s) = \sum_{i=1}^{N/2} \ln\left(1 + \frac{\sigma_{s_i}^2}{\sigma_n^2}\right) - \sum_{i=1}^{N/2} \ln\left\{\left(1 + s\frac{\sigma_{s_i}^2}{\sigma_n^2}\right)\left(1 + (1-s)\frac{\sigma_{s_i}^2}{\sigma_n^2}\right)\right\}. \tag{510}$$

Then

$$\dot{\mu}(s) = -\sum_{i=1}^{N/2} \left[\frac{\sigma_{s_i}^2}{\sigma_n^2 + s\sigma_{s_i}^2} - \frac{\sigma_{s_i}^2}{\sigma_n^2 + (1-s)\sigma_{s_i}^2}\right] \tag{511}$$

and

$$\ddot{\mu}(s) = \sum_{i=1}^{N/2} \left\{\frac{\sigma_{s_i}^4}{(\sigma_n^2 + s\sigma_{s_i}^2)^2} + \frac{\sigma_{s_i}^4}{[\sigma_n^2 + (1-s)\sigma_{s_i}^2]^2}\right\}. \tag{512}$$

For a minimum probability of error criterion it is obvious from (511) that $s_m = \frac{1}{2}$. Using (485), we have

$$\Pr(\epsilon) \simeq \left[\pi \sum_{i=1}^{N/2} \frac{\sigma_{s_i}^4}{(\sigma_n^2 + \frac{1}{2}\sigma_{s_i}^2)^2}\right]^{-1/2} \exp\left[\sum_{i=1}^{N/2} \ln\left(1 + \frac{\sigma_{s_i}^2}{\sigma_n^2}\right) - 2\sum_{i=1}^{N/2} \ln\left(1 + \frac{\sigma_{s_i}^2}{2\sigma_n^2}\right)\right] \tag{513}$$

or

$$\Pr(\epsilon) \simeq \left[\pi \sum_{i=1}^{N/2} \frac{\sigma_{s_i}^4}{(\sigma_n^2 + \frac{1}{2}\sigma_{s_i}^2)^2}\right]^{-1/2} \prod_{i=1}^{N/2} \frac{\left(1 + \frac{\sigma_{s_i}^2}{\sigma_n^2}\right)}{\left(1 + \frac{\sigma_{s_i}^2}{2\sigma_n^2}\right)^2}. \tag{514}$$

For the special case in which the variances are equal

$$\sigma_{s_i}^2 = \sigma_s^2 \tag{515}$$

and (514) reduces to

$$\Pr(\epsilon) \simeq \sqrt{\frac{2}{\pi N}} \frac{(1 + \sigma_s^2/\sigma_n^2)^{N/2}}{(\sigma_s^2/\sigma_n^2)(1 + \sigma_s^2/2\sigma_n^2)^{N-1}}. \tag{516}$$

Alternately, we can use the approximation given by (484). For this case it reduces to

$$\Pr(\epsilon) \simeq \left[\frac{1 + \sigma_s^2/\sigma_n^2}{(1 + \sigma_s^2/2\sigma_n^2)^2}\right]^{N/2} \exp\left[\frac{N}{8}\left(\frac{\sigma_s^2/\sigma_n^2}{1 + \sigma_s^2/2\sigma_n^2}\right)^2\right] \operatorname{erfc}_*\left[\left(\frac{N}{4}\right)^{1/2}\left(\frac{\sigma_s^2/\sigma_n^2}{1 + \sigma_s^2/2\sigma_n^2}\right)\right]. \tag{517}$$

In Fig. 2.44 we have plotted the approximate $\Pr(\epsilon)$ using (517) and exact $\Pr(\epsilon)$ which was given by (434). We see that the approximation is excellent.

The principal results of this section were the bounds on P_F and P_M given in (470) and (473) and the approximate error expressions given in (479), (480), (482), (483), (484), and (485). These expressions will enable us to find performance results for a number of cases of physical interest.

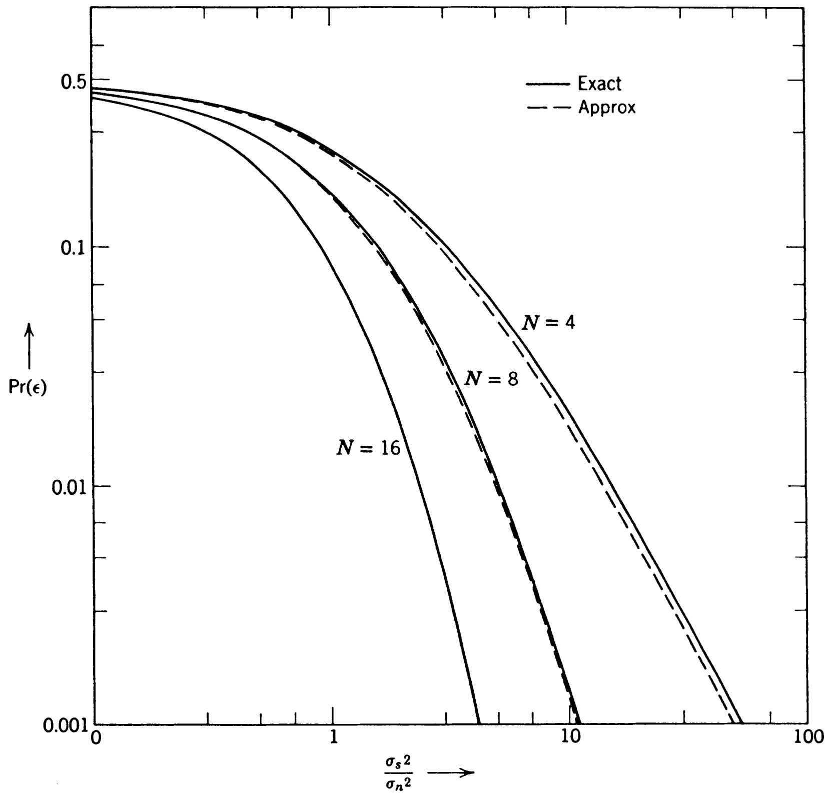


Fig. 2.44 Exact and approximate error expressions for the binary symmetric hypothesis case.

Results for some other cases are given in Yudkin [34] and Goblick [35] and the problems. In Chapter II-3 we shall study the detection of Gaussian signals in Gaussian noise. Suitable extensions of the above bounds and approximations will be used to evaluate the performance of the optimum processors.

2.8 SUMMARY

In this chapter we have derived the essential detection and estimation theory results that provide the basis for our work in the remainder of the book.

We began our discussion by considering the simple binary hypothesis testing problem. Using either a Bayes or a Neyman-Pearson criterion, we were led to a likelihood ratio test, whose performance was described by a receiver operating characteristic. Similarly, the M-hypothesis problem led to the construction of a set of likelihood ratios. This criterion-invariant reduction of the observation to a single number in the binary case or to $M - 1$ numbers in the M hypothesis case is the key to our ability to solve the detection problem when the observation is a waveform.

The development of the necessary estimation theory results followed a parallel path. Here, the fundamental quantity was a likelihood function. As we pointed out in Section 2.4, its construction is closely related to the construction of the likelihood ratio, a similarity that will enable us to solve many parallel problems by inspection. The composite hypothesis testing problem showed further how the two problems were related.

Our discussion through Section 2.5 was deliberately kept at a general level and for that reason forms a broad background of results applicable to many areas in addition to those emphasized in the remainder of the book. In Section 2.6 we directed our attention to the general Gaussian problem, a restriction that enabled us to obtain more specific results than were available in the general case. The waveform analog to this general Gaussian problem plays the central role in most of the succeeding work.

The results in the general Gaussian problem illustrated that although we can always find the optimum processor the exact performance may be difficult to calculate. This difficulty motivated our discussion of error bounds and approximations in Section 2.7. These approximations will lead us to useful results in several problem areas of practical importance.

2.9 PROBLEMS

The problems are divided into sections corresponding to the major sections in the chapter. For example, section P2.2 pertains to text material in Section 2.2. In sections in which it is appropriate the problems are divided into topical groups.

As pointed out in the Preface, solutions to individual problems are available on request.

P2.2 Binary Hypothesis Tests

Simple Binary Tests

Problem 2.2.1. Consider the following binary hypothesis testing problem:

$$H_1: r = s + n,$$
$$H_0: r = n,$$

where s and n are independent random variables.

$$p_s(S) = \begin{cases} ae^{-aS} & S \geq 0, \\ 0 & S < 0, \end{cases}$$

$$p_n(N) = \begin{cases} be^{-bN} & N \geq 0, \\ 0 & N < 0. \end{cases}$$

1. Prove that the likelihood ratio test reduces to

$$R \underset{H_0}{\overset{H_1}{\gtrless}} \gamma.$$

2. Find γ for the optimum Bayes test as a function of the costs and a priori probabilities.

3. Now assume that we need a Neyman-Pearson test. Find γ as a function of P_F, where

$$P_F \triangleq \Pr(\text{say } H_1 | H_0 \text{ is true}).$$

Problem 2.2.2. The two hypotheses are

$$H_1: p_r(R) = \frac{1}{2} \exp(-|R|)$$

$$H_0: p_r(R) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} R^2\right)$$

1. Find the likelihood ratio $\Lambda(R)$.
2. The test is

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

Compute the decision regions for various values of η.

Problem 2.2.3. The random variable x is $N(0, \sigma)$. It is passed through one of two nonlinear transformations.

$$H_1: y = x^2,$$
$$H_0: y = x^3.$$

Find the LRT.

Problem 2.2.4. The random variable x is $N(m, \sigma)$. It is passed through one of two nonlinear transformations.

$$H_1: y = e^x,$$
$$H_0: y = x^2.$$

Find the LRT.

Problem 2.2.5. Consider the following hypothesis-testing problem. There are K independent observations.

$$H_1: r_i \text{ is Gaussian, } N(0, \sigma_1), \qquad i = 1, 2, \ldots, K,$$
$$H_0: r_i \text{ is Gaussian, } N(0, \sigma_0), \qquad i = 1, 2, \ldots, K,$$

where $\sigma_0 < \sigma_1$.

1. Compute the likelihood ratio.
2. Assume that the threshold is η:

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

Show that a sufficient statistic is $l(\mathbf{R}) = \sum_{i=1}^{K} R_i^2$. Compute the threshold γ for the test

$$l(\mathbf{R}) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

in terms of η, σ_0, σ_1.

3. Define

$$P_F = \Pr\,(\text{choose } H_1 |\ H_0 \text{ is true}),$$
$$P_M = \Pr\,(\text{choose } H_0 |\ H_1 \text{ is true}).$$

Find an expression for P_F and P_M.

4. Plot the ROC for $K = 1$, $\sigma_1^2 = 2$, $\sigma_0^2 = 1$.
5. What is the threshold for the minimax criterion when $C_M = C_F$ and $C_{00} = C_{11} = 0$?

Problem 2.2.6. The observation r is defined in the following manner:

$$r = bm_1 + n : H_1,$$
$$r = n \qquad\quad : H_0,$$

where b and n are independent zero-mean Gaussian variables with variances σ_b^2 and σ_n^2, respectively

1. Find the LRT and draw a block diagram of the optimum processor.
2. Draw the ROC.
3. Assume that the two hypotheses are equally likely. Use the criterion of minimum probability of error. What is the $\Pr(\epsilon)$?

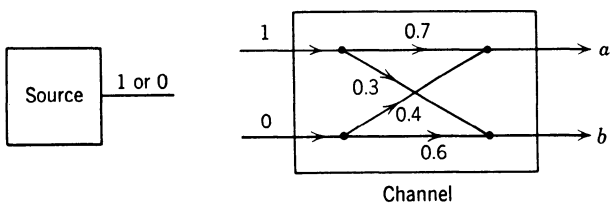
Problem 2.2.7. One of two possible sources supplies the inputs to the simple communication channel as shown in the figure.

Both sources put out either 1 or 0. The numbers on the line are the channel transition probabilities; that is,

$$\Pr(a \text{ out} \mid 1 \text{ in}) = 0.7.$$

The source characteristics are

$$\text{Source 1:}\quad \Pr(1) = 0.5,\quad \Pr(0) = 0.5;$$
$$\text{Source 2:}\quad \Pr(1) = 0.6,\quad \Pr(0) = 0.4.$$



To put the problem in familiar notation, define

(a) false alarm—say source 2 when source 1 is present;
(b) detection—say source 2 when source 2 is present.

1. Compute the ROC of a test that maximizes P_D subject to the constraint that $P_F = \alpha$.
2. Describe the test procedure in detail for $\alpha = 0.25$.

Problem 2.2.8. The probability densities on the two hypotheses are

$$p_{x|H_i}(X|H_i) = \frac{1}{\pi[1 + (X - a_i)^2]} \qquad -\infty < X < \infty : H_i, \qquad i = 0, 1.$$

where $a_0 = 0$ and $a_1 = 1$.

1. Find the LRT.
2. Plot the ROC.

Problem 2.2.9. Consider a simple coin tossing problem:

$$\begin{aligned} &H_1: \text{ heads are up,} && \Pr[H_1] \triangleq P_1, \\ &H_0: \text{ tails are up,} && \Pr[H_0] \triangleq P_0 < P_1. \end{aligned}$$

N independent tosses of the coin are made. Show that the number of observed heads, N_H, is a sufficient statistic for making a decision between the two hypotheses.

Problem 2.2.10. A sample function of a simple Poisson counting process $N(t)$ is observed over the interval T:

$$\begin{aligned} &\text{hypothesis } H_1: \text{ the mean rate is } k_1 : \Pr(H_1) = \tfrac{1}{2}, \\ &\text{hypothesis } H_0: \text{ the mean rate is } k_0 : \Pr(H_0) = \tfrac{1}{2}. \end{aligned}$$

1. Prove that the number of events in the interval T is a "sufficient statistic" to choose hypothesis H_0 or H_1.
2. Assuming equal costs for the possible errors, derive the appropriate likelihood ratio test and the threshold.
3. Find an expression for the probability of error.

Problem 2.2.11. Let

$$y = \sum_{i=0}^{n} x_i,$$

where the x_i are statistically independent random variables with a Gaussian density $N(0, \sigma)$. The number of variables in the sum is a random variable with a Poisson distribution:

$$\Pr(n = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, \ldots.$$

We want to decide between the two hypotheses,

$$\begin{aligned} H_1 &: n \leq 1, \\ H_0 &: n > 1. \end{aligned}$$

Write an expression for the LRT.

Problem 2.2.12. Randomized tests. Our basic model of the decision problem in the text (p. 24) did not permit randomized decision rules. We can incorporate them by assuming that at each point $\mathbf{R}$ in Z we say H_1 with probability $\phi(\mathbf{R})$ and say H_0 with probability $1 - \phi(\mathbf{R})$. The model in the text is equivalent to setting $\phi(\mathbf{R}) = 1$ for all $\mathbf{R}$ in Z_1 and $\phi(\mathbf{R}) = 0$ for all $\mathbf{R}$ in Z_0.

1. We consider the Bayes criterion first. Write the risk for the above decision model.
2. Prove that a LRT minimizes the risk and a randomized test is *never* necessary.

3. Prove that the risk is constant over the interior of any straight-line segment on an ROC. Because straight-line segments are generated by randomized tests, this is an alternate proof of the result in Part 2.

4. Consider the Neyman-Pearson criterion. Prove that the optimum test always consists of either

(i) an ordinary LRT with $P_F = \alpha$ or

(ii) a probabilistic mixture of *two* ordinary likelihood ratio tests constructed as follows: Test 1: $\Lambda(\mathbf{R}) \underset{}{\overset{H_1}{\geq}} \eta$ gives $P_F = \alpha^+$. Test 2: $\Lambda(\mathbf{R}) \overset{H_1}{>} \eta$ gives $P_F = \alpha^-$, where $[\alpha^-, \alpha^+]$ is the smallest interval containing α. $\phi(\mathbf{R})$ is 0 or 1 except for those $\mathbf{R}$ where $\phi(\mathbf{R}) = \eta$. (Find $\phi(\mathbf{R})$ for this set.)

Mathematical Properties

Problem 2.2.13. The random variable $\Lambda(\mathbf{R})$ is defined by (13) and has a different probability density on H_1 and H_0. Prove the following:

1. $E(\Lambda^n|H_1) = E(\Lambda^{n+1}|H_0)$,
2. $E(\Lambda|H_0) = 1$,
3. $E(\Lambda|H_1) - E(\Lambda|H_0) = \text{Var}\,(\Lambda|H_0)$.

Problem 2.2.14. Consider the random variable Λ. In (94) we showed that

$$p_{\Lambda|H_1}(X|H_1) = Xp_{\Lambda|H_0}(X|H_0).$$

1. Verify this relation by direct calculation of $p_{\Lambda|H_1}(\cdot)$ and $p_{\Lambda|H_0}(\cdot)$ for the densities in Example 1 [p. 27, (19) and (20)].

2. On page 37 we saw that the performance of the test in Example 1 was completely characterized by d^2. Show that

$$d^2 = \ln\,[1 + \text{Var}\,(\Lambda|H_0)].$$

Problem 2.2.15. The function $\text{erfc}_*\,(X)$ is defined in (66):

1. Integrate by parts to establish the bound

$$\frac{1}{\sqrt{2\pi}\,X}\left(1 - \frac{1}{X^2}\right)\exp\left(-\frac{X^2}{2}\right) < \text{erfc}_*\,(X) < \frac{1}{\sqrt{2\pi}\,X}\exp\left(-\frac{X^2}{2}\right), \qquad X > 0.$$

2. Generalize part 1 to obtain the asymptotic series

$$\text{erfc}_*\,(X) = \frac{1}{\sqrt{2\pi}\,X}\,e^{-X^2/2}\left[1 + \sum_{m=1}^{n-1}(-1)^m\,\frac{1\cdot 3\cdots(2m-1)}{X^{2m}} + R_n\right].$$

The remainder is less than the magnitude of the $n+1$ term and is the same sign. *Hint.* Show that the remainder is

$$R_n = \left[(-1)^{n+1}\,\frac{1\cdot 3\cdots(2n-1)}{X^{2n+2}}\right]\theta,$$

where

$$\theta = \int_0^\infty e^{-t}\left(1 + \frac{2t}{X^2}\right)^{-n-\frac{1}{2}}dt < 1.$$

3. Assume that $X = 3$. Calculate a simple bound on the *percentage* error when $\text{erfc}_*(3)$ is approximated by the first n terms in the asymptotic series. Evaluate this percentage error for $n = 2, 3, 4$ and compare the results. Repeat for $X = 5$.

Problem 2.2.16.

1. Prove

$$\operatorname{erfc}_*(X) < \frac{1}{2}\exp\left(-\frac{X^2}{2}\right), \qquad X > 0.$$

Hint. Show

$$[\operatorname{erfc}_*(X)]^2 = \Pr(x \geq X, y \geq X) < \Pr(x^2 + y^2 \geq 2X^2),$$

where x and y are independent zero-mean Gaussian variables with unit variance.

2. For what values of X is this bound better than (71)?

HIGHER DIMENSIONAL DECISION REGIONS

A simple binary test can always be reduced to a one-dimensional decision region. In many cases the results are easier to interpret in two or three dimensions. Some typical examples are illustrated in this section.

Problem 2.2.17.

$$H_1{:}p_{x_1,x_2|H_1}(X_1, X_2|H_1) = \frac{1}{4\pi\sigma_1\sigma_0}\left[\exp\left(-\frac{X_1^2}{2\sigma_1^2} - \frac{X_2^2}{2\sigma_0^2}\right) + \exp\left(-\frac{X_1^2}{2\sigma_0^2} - \frac{X_2^2}{2\sigma_1^2}\right)\right],$$
$$-\infty < X_1, X_2 < \infty.$$

$$H_0{:}p_{x_1,x_2|H_0}(X_1, X_2|H_0) = \frac{1}{2\pi\sigma_0^2}\exp\left(-\frac{X_1^2}{2\sigma_0^2} - \frac{X_2^2}{2\sigma_0^2}\right), \qquad -\infty < X_1, X_2 < \infty.$$

1. Find the LRT.
2. Write an exact expression for P_D and P_F. Upper and lower bound P_D and P_F by modifying the region of integration in the exact expression.

Problem 2.2.18. The joint probability density of the random variables x_i ($i = 1, 2, .., M$) on H_1 and H_0 is

$$p_{\mathbf{x}|H_1}(\mathbf{X}|H_1) = \sum_{k=1}^{M} p_k \frac{1}{(2\pi\sigma^2)^{M/2}}\exp\left[-\frac{(X_k - m)^2}{2\sigma^2}\right]\prod_{\substack{i=1\\ i \neq k}}^{M}\exp\left(-\frac{X_i^2}{2\sigma^2}\right),$$

where

$$\sum_{k=1}^{M} p_k = 1,$$

$$p_{\mathbf{x}|H_0}(\mathbf{X}|H_0) = \prod_{i=1}^{M}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{X_i^2}{2\sigma^2}\right) \qquad -\infty < X_i < \infty.$$

1. Find the LRT.
2. Draw the decision regions for various values of η in the X_1, X_2-plane for the special case in which $M = 2$ and $p_1 = p_2 = \frac{1}{2}$.
3. Find an upper and lower bound to P_F and P_D by modifying the regions of integration.

Problem 2.2.19. The probability density of r_i on the two hypotheses is

$$p_{r_i|H_k}(R_i|H_k) = \frac{1}{\sqrt{2\pi}\,\sigma_k}\exp\left[-\frac{(R_i - m_k)^2}{2\sigma_k^2}\right], \qquad \begin{aligned} i &= 1, 2, \ldots, N,\\ k &= 0, 1.\end{aligned}$$

The observations are independent.

1. Find the LRT. Express the test in terms of the following quantities:

$$l_\alpha = \sum_{i=1}^{N} R_i,$$

$$l_\beta = \sum_{i=1}^{N} R_i^2.$$

2. Draw the decision regions in the l_α, l_β-plane for the case in which

$$2m_0 = m_1 > 0,$$
$$2\sigma_1 = \sigma_0.$$

Problem 2.2.20 (continuation).

1. Consider the special case

$$m_0 = 0,$$
$$\sigma_0 = \sigma_1.$$

Draw the decision regions and compute the ROC.

2. Consider the special case

$$m_0 = m_1 = 0,$$
$$\sigma_1^2 = \sigma_s^2 + \sigma_n^2,$$
$$\sigma_0 = \sigma_n.$$

Draw the decision regions.

Problem 2.2.21. A shell is fired at one of two targets: under H_1 the point of aim has coordinates x_1, y_1, z_1; under H_0 it has coordinates x_0, y_0, z_0. The distance of the actual landing point from the point of aim is a zero-mean Gaussian variable, $N(0, \sigma)$, in each coordinate. The variables are independent. We wish to observe the point of impact and guess which hypothesis is true.

1. Formulate this as a hypothesis-testing problem and compute the likelihood ratio. What is the simplest sufficient statistic? Is the ROC in Fig. 2.9*a* applicable? If so, what value of d^2 do we use?

2. Now include the effect of time. Under H_k the desired explosion time is t_k ($k = 1, 2$). The distribution of the actual explosion time is

$$p_{\tau|H_k}(\tau) = \frac{1}{\sqrt{2\pi}\,\sigma_t} \exp\left(-\frac{(\tau - t_k)^2}{2\sigma_t^2}\right), \qquad \begin{matrix} -\infty < \tau < \infty, \\ k = 1, 2. \end{matrix}$$

Find the LRT and compute the ROC.

P2.3 M-Hypothesis Tests

Problem 2.3.1.

1. Verify that the M-hypothesis Bayes test always leads to a decision space whose dimension is less than or equal to $M - 1$.

2. Assume that the coordinates of the decision space are

$$\Lambda_k(\mathbf{R}) \triangleq \frac{p_{\mathbf{r}|H_k}(\mathbf{R}|H_k)}{p_{\mathbf{r}|H_0}(\mathbf{R}|H_0)}, \qquad k = 1, 2, \ldots, M - 1.$$

Verify that the decision boundaries are hyperplanes.

Problem 2.3.2. The formulation of the M-hypothesis problem in the text leads to an efficient decision space but loses some of the symmetry.

1. Starting with (98) prove that an equivalent form of the Bayes test is the following:

Compute

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

and choose the *smallest*.

2. Consider the special cost assignment

$$
\begin{aligned}
C_{ii} &= 0, \qquad i = 0, 1, 2, \ldots, M-1, \\
C_{ij} &= C, \qquad i \neq j,\ i, j = 0, 1, 2, \ldots, M-1.
\end{aligned}
$$

Show that an equivalent test is the following:
Compute

$$
\Pr(H_i|\mathbf{R}), \qquad i = 0, 1, 2, \ldots, M-1,
$$

and choose the *largest*.

Problem 2.3.3. The observed random variable is Gaussian on each of five hypotheses.

$$
p_{r|H_k}(R|H_k) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(R-m_k)^2}{2\sigma^2}\right), \qquad \begin{array}{l} -\infty < R < \infty, \\ k = 1, 2, \ldots, 5, \end{array}
$$

where

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

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

1. Draw the decision regions on the R-axis.
2. Compute the error probability.

Problem 2.3.4. The observed random variable r has a Gaussian density on the three hypotheses,

$$
p_{r|H_k}(R|H_k) = \frac{1}{\sqrt{2\pi}\,\sigma_k} \exp\left[-\frac{(R-m_k)^2}{2\sigma_k^2}\right], \qquad \begin{array}{l} -\infty < R < \infty \\ k = 1, 2, 3, \end{array}
$$

where the parameter values on the three hypotheses are,

$$
\begin{aligned}
H_1&: m_1 = 0, \quad \sigma_1 = \sigma_\alpha, \\
H_2&: m_2 = m, \quad \sigma_2 = \sigma_\alpha, \qquad (m > 0), \\
H_3&: m_3 = 0, \quad \sigma_3 = \sigma_\beta, \qquad (\sigma_\beta > \sigma_\alpha).
\end{aligned}
$$

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

1. Find the optimum Bayes test.
2. Draw the decision regions on the R-axis for the special case,

$$
\begin{aligned}
\sigma_\beta^2 &= 2\sigma_\alpha^2, \\
\sigma_\alpha &= m.
\end{aligned}
$$

3. Compute the $\Pr(\epsilon)$ for this special case.

Problem 2.3.5. The probability density of $\mathbf{r}$ on the three hypotheses is

$$
p_{r_1, r_2|H_k}(R_1, R_2|H_k) = (2\pi\sigma_{1k}\sigma_{2k})^{-1} \exp\left[-\frac{1}{2}\left(\frac{R_1^2}{\sigma_{1k}^2} + \frac{R_2^2}{\sigma_{2k}^2}\right)\right], \qquad \begin{array}{l} -\infty < R_1, R_2 < \infty, \\ k = 1, 2, 3, \end{array}
$$

where

$$
\begin{aligned}
&\sigma_{11}^2 = \sigma_{21}^2 = \sigma_n^2, \\
&\sigma_{12}^2 = \sigma_s^2 + \sigma_n^2, \qquad \sigma_{22}^2 = \sigma_n^2, \\
&\sigma_{13}^2 = \sigma_n^2, \qquad \sigma_{23}^2 = \sigma_s^2 + \sigma_n^2.
\end{aligned}
$$

The cost matrix is

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & \alpha \\ 1 & \alpha & 0 \end{bmatrix},$$

where $0 \leq \alpha < 1$ and $\Pr(H_2) = \Pr(H_3) \triangleq p$. Define $l_1 = R_1^2$ and $l_2 = R_2^2$.

1. Find the optimum test and indicate the decision regions in the l_1, l_2-plane.
2. Write an expression for the error probabilities. (Do not evaluate the integrals.)
3. Verify that for $\alpha = 0$ this problem reduces to 2.2.17.

Problem 2.3.6. On H_k the observation is a value of a Poisson random variable

$$\Pr(r = n) = \frac{k_m{}^n}{n!} e^{-k_m}, \qquad m = 1, 2, \dots, M,$$

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

1. Find the optimum test.
2. Find a simple expression for the boundaries of the decision regions and indicate how you would compute the $\Pr(\epsilon)$.

Problem 2.3.7. Assume that the received vector on each of the three hypotheses is

$$\begin{aligned} H_0 &: \mathbf{r} = \mathbf{m}_0 + \mathbf{n}, \\ H_1 &: \mathbf{r} = \mathbf{m}_1 + \mathbf{n}, \\ H_2 &: \mathbf{r} = \mathbf{m}_2 + \mathbf{n}, \end{aligned}$$

where

$$\mathbf{r} \triangleq \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}, \qquad \mathbf{m}_i \triangleq \begin{bmatrix} m_{i1} \\ m_{i2} \\ m_{i3} \end{bmatrix}, \qquad \mathbf{n} \triangleq \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}.$$

The $\mathbf{m}_i$ are known vectors, and the components of $\mathbf{n}$ are statistically independent, zero-mean Gaussian random variables with variance σ^2.

1. Using the results in the text, express the Bayes test in terms of two sufficient statistics.

$$l_1 = \sum_{i=1}^{3} c_i r_i,$$

$$l_2 = \sum_{i=1}^{3} d_i r_i.$$

Find explicit expressions for c_i and d_i. Is the solution unique?

2. Sketch the decision regions in the l_1, l_2-plane for the particular cost assignment,

$$\begin{aligned} C_{00} &= C_{11} = C_{22} = 0, \\ C_{12} &= C_{21} = C_{01} = C_{10} = \tfrac{1}{2}C_{02} = \tfrac{1}{2}C_{20} > 0. \end{aligned}$$

P2.4 Estimation

BAYES ESTIMATION

Problem 2.4.1. Let

$$r = ab + n,$$

where a, b, and independent zero-mean Gaussian variables with variances σ_a^2, σ_b^2, and σ_n^2.

1. What is $\hat{a}_{\text{map}}$?
2. Is this equivalent to simultaneously finding $\hat{a}_{\text{map}}$, $\hat{b}_{\text{map}}$?

3. Now consider the case in which

$$r = a + \sum_{i=1}^{k} b_i + n,$$

where the b_i are independent zero-mean Gaussian variables with variances $\sigma_{b_i}^2$.

(a) What is $\hat{a}_{map}$?
(b) Is this equivalent to simultaneously finding $\hat{a}_{map}$, $\hat{b}_{i,map}$?
(c) Explain intuitively why the answers to part 2 and part 3b are different.

Problem 2.4.2. The observed random variable is x. We want to estimate the parameter λ. The probability density of x as a function of λ is,

$$\begin{aligned} p_{x|\lambda}(X|\lambda) &= \lambda e^{-\lambda X}, \qquad X \geq 0, \lambda > 0, \\ &= 0, \qquad X < 0. \end{aligned}$$

The a priori density of λ depends on two parameters: n_*, l_*.

$$p_{\lambda|n_*,l_*}(\lambda|n_*, l_*) \triangleq \begin{cases} \dfrac{l_*^{n_*}}{\Gamma(n_*)} e^{-\lambda l_*}\lambda^{n_*-1}, & \lambda \geq 0, \\ 0, & \lambda < 0. \end{cases}$$

1. Find $E(\lambda)$ and Var (λ) *before* any observations are made.
2. Assume that one observation is made. Find $p_{\lambda|x}(\lambda|X)$. What interesting property does this density possess? Find $\hat{\lambda}_{ms}$ and $E[(\hat{\lambda}_{ms} - \lambda)^2]$.
3. Now assume that n independent observations are made. Denote these n observations by the vector $\mathbf{x}$. Verify that

$$p_{\lambda|\mathbf{x}}(\lambda|\mathbf{X}) \triangleq \begin{cases} \dfrac{(l')^{n'}}{\Gamma(n')} e^{-\lambda l'}\lambda^{n'-1}, & \lambda \geq 0, \\ 0, & \lambda < 0, \end{cases}$$

where

$$\begin{aligned} l' &= l + l_*, \\ n' &= n + n_*, \end{aligned}$$

and

$$l = \sum_{i=1}^{n} X_i.$$

Find $\hat{\lambda}_{ms}$ and $E[(\hat{\lambda}_{ms} - \lambda)^2]$.
4. Does $\hat{\lambda}_{map} = \hat{\lambda}_{ms}$?

Comment. Reproducing Densities. The reason that the preceding problem was simple was that the a priori and a posteriori densities had the same functional form. (Only the parameters changed.) In general,

$$p_{a|r}(A|R) = \frac{p_{r|a}(R|A)p_a(A)}{p_r(R)},$$

and we say that $p_a(A)$ is a *reproducing density* or a *conjugate prior* density [with respect to the transition density $p_{r|a}(R|A)$] if the a posteriori density is of the same form as $p_a(A)$. Because the choice of the a priori density is frequently somewhat arbitrary, it is convenient to choose a reproducing density in many cases. The next two problems illustrate other reproducing densities of interest.

Problem 2.4.3. Let

$$r = a + n,$$

where n is $N(0, \sigma_n)$. Then

$$p_{r|a}(R|A) = \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left[-\frac{(R - A)^2}{2\sigma_n^2}\right].$$

1. Verify that a conjugate priori density for a is $N\left(m_0, \frac{\sigma_n}{k_0}\right)$ by showing that

$$p_{a|r}(A|R) = N(m_1, \sigma_1),$$

where

$$m_1 = \frac{m_0 k_0^2 + R}{(1 + k_0^2)}$$

and

$$\sigma_1^2 = \frac{\sigma_n^2}{1 + k_0^2}.$$

2. Extend this result to N independent observations by verifying that

$$p_{a|\mathbf{r}}(A|\mathbf{R}) = N(m_N, \sigma_N),$$

where

$$m_N = \frac{m_0 k_0^2 + Nl}{N + k_0^2},$$

$$\sigma_N^2 = \frac{\sigma_n^2}{N + k_0^2},$$

and

$$l \triangleq \frac{1}{N} \sum_{i=1}^{N} R_i.$$

Observe that the a priori parameter k_0^2 can be interpreted as an equivalent number of observations (fractional observations are allowed).

Problem 2.4.4. Consider the observation process

$$p_{r|a}(R|A) = \frac{A^{1/2}}{(2\pi)^{1/2}} \exp\left[-\frac{A}{2}(R - m)^2\right],$$

where m is known and A is the parameter of interest (it is the reciprocal of the variance). We assume that N independent observations are available.

1. Verify that

$$p_a(A|k_1, k_2) = c(A^{\frac{k_1}{2}-1}) \exp(-\tfrac{1}{2}Ak_1k_2), \qquad A \geq 0, \quad k_1, k_2 > 0,$$

(c is a normalizing factor) is a conjugate prior density by showing that

$$p_{a|\mathbf{r}}(A|\mathbf{R}) = p_a(A|k_1', k_2'),$$

where

$$k_2' = \frac{1}{k_1'}(k_1k_2 + Nw),$$

$$k_1' = k_1 + N,$$

$$w = \frac{1}{N} \sum_{i=1}^{N} (R_i - m)^2.$$

Note that k_1, k_2 are simply the parameters in the a priori density which are chosen based on our a priori knowledge.

2. Find $\hat{a}_{ms}$.

Problem 2.4.5. We make K observations: $R_1, \ldots, R_K$, where

$$r_i = a + n_i.$$

The random variable a has a Gaussian density $N(0, \sigma_a)$. The n_i are independent Gaussian variables $N(0, \sigma_n)$.

1. Find the MMSE estimate $\hat{a}_{ms}$.
2. Find the MAP estimate $\hat{a}_{map}$.
3. Compute the mean-square error.
4. Consider an alternate procedure using the same r_i.

   (a) Estimate a after each observation using a MMSE criterion.

This gives a sequence of estimates $\hat{a}_1(R_1)$, $\hat{a}_2(R_1, R_2) \ldots \hat{a}_j(R_1, \ldots R_j) \ldots \hat{a}_K(R_1, \ldots, R_K)$. Denote the corresponding variances as $\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2$.

   (b) Express $\hat{a}_j$ as a function of $\hat{a}_{j-1}$, σ_{j-1}^2, and R_j.
   (c) Show that

$$\frac{1}{\sigma_j^2} = \frac{1}{\sigma_a^2} + \frac{j}{\sigma_n^2}.$$

Problem 2.4.6. [36]. In this problem we outline the proof of Property 2 on p. 61. The assumptions are the following:

(a) The cost function is a symmetric, nondecreasing function. Thus

$$C(X) = C(-X)$$

$$C(X_1) \geq C(X_2) \quad \text{for} \quad X_1 \geq X_2 \geq 0, \tag{P.1}$$

which implies

$$\frac{dC(X)}{dX} \geq 0 \quad \text{for} \quad X \geq 0. \tag{P.2}$$

(b) The a posteriori probability density is symmetric about its conditional mean and is nonincreasing.

(c)
$$\lim_{X \to \infty} C(X) p_{x|\mathbf{r}}(X|\mathbf{R}) = 0. \tag{P.3}$$

We use the same notation as in Property 1 on p. 61. Verify the following steps:

1. The conditional risk using the estimate $\hat{a}$ is

$$\mathcal{R}(\hat{a}|\mathbf{R}) = \int_{-\infty}^{\infty} C(Z) p_{z|\mathbf{r}}(Z + \hat{a} - \hat{a}_{ms}|\mathbf{R})\, dZ. \tag{P.4}$$

2. The difference in conditional risks is

$$\Delta\mathcal{R} = \mathcal{R}(\hat{a}|\mathbf{R}) - \mathcal{R}(\hat{a}_{ms}|\mathbf{R}) = \int_0^{\infty} C(Z)[p_{z|\mathbf{r}}(Z + \hat{a} - \hat{a}_{ms}|\mathbf{R}) p_{z|\mathbf{r}}(Z - \hat{a} + \hat{a}_{ms}|\mathbf{R}) - 2p_{z|\mathbf{r}}(Z|\mathbf{R})]\, dZ. \tag{P.5}$$

3. For $\hat{a} > \hat{a}_{ms}$ the integral of the terms in the bracket with respect to Z from 0 to Z_0 is

$$\int_0^{\hat{a} - \hat{a}_{ms}} [p_{z|\mathbf{r}}(Z_0 + Y|\mathbf{R}) - p_{z|\mathbf{r}}(Z_0 - Y|\mathbf{R})]\, dY \triangleq g(Z_0) \tag{P.6}$$

4. Integrate (P.5) by parts to obtain

$$\Delta\mathcal{R} = C(Z)g(Z)\Big|_0^{\infty} - \int_0^{\infty} \frac{dC(Z)}{dZ} g(Z)\, dZ, \qquad \hat{a} > \hat{a}_{ms}. \tag{P.7}$$

5. Show that the assumptions imply that the first term is zero and the second term is nonnegative.
6. Repeat Steps 3 to 5 with appropriate modifications for $\hat{a} < \hat{a}_{ms}$.

7. Observe that these steps prove that $\hat{a}_{ms}$ minimizes the Bayes risk under the above assumptions. Under what conditions will the Bayes estimate be unique?

NONRANDOM PARAMETER ESTIMATION

Problem 2.4.7. We make n statistically independent observations: $r_1, r_2, \ldots, r_n$, with mean m and variance σ^2. Define the sample variance as

$$V = \frac{1}{n}\sum_{j=1}^{n}\left(R_j - \sum_{i=1}^{n}\frac{R_i}{n}\right)^2.$$

Is it an unbiased estimator of the actual variance?

Problem 2.4.8. We want to estimate a in a binomial distribution by using n observations.

$$\Pr(r \text{ events}|a) = \binom{n}{r}a^r(1-a)^{n-r}, \qquad r = 0, 1, 2, \ldots, n.$$

1. Find the ML estimate of a and compute its variance.
2. Is it efficient?

Problem 2.4.9.

1. Does an efficient estimate of the standard deviation σ of a zero-mean Gaussian density exist?
2. Does an efficient estimate of the variance σ^2 of a zero-mean Gaussian density exist?

Problem 2.4.10 (continuation). The results of Problem 2.4.9 suggest the general question. Consider the problem of estimating some function of the parameter A, say, $f_1(A)$. The observed quantity is R and $p_{r|a}(R|A)$ is known. Assume that A is a nonrandom variable.

1. What are the conditions for an efficient estimate $\hat{f}_1(A)$ to exist?
2. What is the lower bound on the variance of the error of any unbiased estimate of $f_1(A)$?
3. Assume that an efficient estimate of $f_1(A)$ exists. When can an efficient estimate of some other function $f_2(A)$ exist?

Problem 2.4.11. The probability density of r, given A_1 and A_2 is:

$$p_{r|a_1,a_2}(R|A_1, A_2) = (2\pi A_2)^{-1/2}\exp\left[-\frac{(R-A_1)^2}{2A_2}\right];$$

that is, A_1 is the mean and A_2 is the variance.

1. Find the joint ML estimates of A_1 and A_2 by using n independent observations.
2. Are they biased?
3. Are they coupled?
4. Find the error covariance matrix.

Problem 2.4.12. We want to transmit two parameters, A_1 and A_2. In a simple attempt to achieve a secure communication system we construct two signals to be transmitted over separate channels.

$$s_1 = x_{11}A_1 + x_{12}A_2,$$
$$s_2 = x_{21}A_1 + x_{22}A_2,$$

where x_{ij}, $i, j = 1, 2$, are known. The received variables are

$$r_1 = s_1 + n_1,$$
$$r_2 = s_2 + n_2.$$

The additive noises are independent, identically distributed, zero-mean Gaussian random variables, $N(0, \sigma_n)$. The parameters A_1 and A_2 are nonrandom.

1. Are the ML estimates $\hat{a}_1$ and $\hat{a}_2$ unbiased?
2. Compute the variance of the ML estimates $\hat{a}_1$ and $\hat{a}_2$.
3. Are the ML estimates efficient? In other words, do they satisfy the Cramér-Rao bound with equality?

Problem 2.4.13. Let

$$y = \sum_{i=1}^{N} x_i,$$

where the x_i are independent, zero-mean Gaussian random variables with variance σ_x^2. We observe y. In parts 1 through 4 treat N as a continuous variable.

1. Find the maximum likelihood estimate of N.
2. Is $\hat{n}_{ml}$ unbiased?
3. What is the variance of $\hat{n}_{ml}$?
4. Is $\hat{n}_{ml}$ efficient?
5. Discuss qualitatively how you would modify part 1 to take into account that N is discrete.

Problem 2.4.14. We observe a value of the discrete random variable x.

$$\Pr(x = i|A) = \frac{A^i}{i!}e^{-A}, \qquad i = 0, 1, 2, \ldots,$$

where A is nonrandom.

1. What is the lower bound on the variance of any unbiased estimate, $\hat{a}(x)$?
2. Assuming n independent observations, find an $\hat{a}(\mathbf{x})$ that is efficient.

Problem 2.4.15. Consider the Cauchy distribution

$$p_{x|a}(X|A) = \{\pi[1 + (X - A)^2]\}^{-1}.$$

Assume that we make n independent observations in order to estimate A.

1. Use the Cramér-Rao inequality to show that the variance of any unbiased estimate of A has a variance greater than $2/n$.
2. Is the sample mean a consistent estimate?
3. We can show that the sample median is asymptotically normal, $N(A, \pi/\sqrt{4n})$. (See pp. 367–369 of Cramér [9].) What is the asymptotic efficiency of the sample median as an estimator?

Problem 2.4.16. Assume that

$$p_{r_1,r_2|\rho}(R_1, R_2|\rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}}\exp\left\{-\frac{(R_1^2 - 2\rho R_1 R_2 + R_2^2)}{2(1-\rho^2)}\right\}.$$

We want to estimate the correlation coefficient ρ by using n independent observations of (R_1, R_2).

1. Find the equation for the ML estimate $\hat{\rho}$.
2. Find a lower bound on the variance of any unbiased estimate of ρ.

Mathematical Properties

Problem 2.4.17. Consider the biased estimate $\hat{a}(\mathbf{R})$ of the *nonrandom* parameter A.

$$E(\hat{a}(\mathbf{R})) = A + B(A).$$

Show that

$$\operatorname{Var}[\hat{a}(\mathbf{R})] \geq \frac{(1 + dB(A)/dA)^2}{E\left\{\left[\dfrac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right]^2\right\}}.$$

This is the Cramér-Rao inequality for biased estimates. Note that it is a bound on the mean-square error.

Problem 2.4.18. Let $p_{\mathbf{r}|a}(\mathbf{R}|A)$ be the probability density of $\mathbf{r}$, given A. Let h be an arbitrary random variable that is independent of r defined so that $A + h$ ranges over all possible values of A. Assume that $p_{h_1}(H)$ and $p_{h_2}(H)$ are *two* arbitrary probability densities for h. Assuming that $\hat{a}(\mathbf{R})$ is unbiased, we have

$$\int [\hat{a}(\mathbf{R}) - (A + H)] p_{\mathbf{r}|a}(\mathbf{R}|A + H)\, d\mathbf{R} = 0.$$

Multiplying by $p_{h_i}(H)$ and integrating over H, we have

$$\int dH\, p_{h_i}(H) \int [\hat{a}(\mathbf{R}) - (A + H)] p_{\mathbf{r}|a}(\mathbf{R}|A + H)\, d\mathbf{R} = 0.$$

1. Show that

$$\operatorname{Var}[\hat{a}(R) - A] \geq \frac{[E_1(h) - E_2(h)]^2}{\displaystyle\int \left(\frac{\left(\int p_{\mathbf{r}|a}(\mathbf{R}|A + H)\,[p_{h_1}(H) - p_{h_2}(H)]\, dH\right)^2}{p_{\mathbf{r}|a}(\mathbf{R}|A)}\right) d\mathbf{R}}$$

for any $p_{h_1}(H)$ and $p_{h_2}(H)$. Observe that because this is true for all $p_{h_1}(H)$ and $p_{h_2}(H)$, we may write

$$\operatorname{Var}[\hat{a}(R) - A] \geq \sup_{p_{h_1}, p_{h_2}} \text{(right-hand side of above equation).}$$

Comment. Observe that this bound does not require any regularity conditions. Barankin [15] has shown that this is the greatest lower bound.

Problem 2.4.19 (continuation). We now derive two special cases.

1. First, let $p_{h_2}(H) = \delta(H)$. What is the resulting bound?
2. Second, let $p_{h_1}(H) = \delta(H - H_0)$, where $H_0 \neq 0$. Show that

$$\operatorname{Var}[\hat{a}(\mathbf{R}) - A] \geq \left(\inf_{H_0}\left\{\frac{1}{H_0^2}\left[\int \frac{p^2_{\mathbf{r}|a}(\mathbf{R}|A + H_0)}{p_{\mathbf{r}|a}(\mathbf{R}|A)}\, d\mathbf{R} - 1\right]\right\}\right)^{-1}.$$

The infimum being over all $H_0 \neq 0$ such that $p_{\mathbf{r}|a}(\mathbf{R}|A) = 0$ implies

$$p_{\mathbf{r}|a}(\mathbf{R}|A + H_0) = 0.$$

3. Show that the bound given in part 2 is always as good as the Cramér-Rao inequality when the latter applies.

Problem 2.4.20. Let

$$\mathbf{a} = \mathbf{L}\mathbf{b},$$

where $\mathbf{L}$ is a nonsingular matrix and $\mathbf{a}$ and $\mathbf{b}$ are vector random variables. Prove that

$$\hat{\mathbf{a}}_{\text{map}} = \mathbf{L}\hat{\mathbf{b}}_{\text{map}} \qquad \text{and} \qquad \hat{\mathbf{a}}_{ms} = \mathbf{L}\hat{\mathbf{b}}_{ms}.$$

Problem 2.4.21. An alternate way to derive the Cramér-Rao inequality is developed in this problem. First, construct the vector $\mathbf{z}$.

$$\mathbf{z} \triangleq \begin{bmatrix} \hat{a}(\mathbf{R}) - A \\ \hdashline \dfrac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} \end{bmatrix}.$$

1. Verify that for unbiased estimates $E(\mathbf{z}) = \mathbf{0}$.

2. Assuming that $E(\mathbf{z}) = \mathbf{0}$, the covariance matrix is

$$\mathbf{\Lambda}_{\mathbf{z}} = E(\mathbf{z}\mathbf{z}^T).$$

Using the fact that $\mathbf{\Lambda}_{\mathbf{z}}$ is nonnegative definite, derive the Cramér-Rao inequality. If the equality holds, what does this imply about $|\mathbf{\Lambda}_{\mathbf{z}}|$?

Problem 2.4.22. Repeat Problem 2.4.21 for the case in which a is a random variable. Define

$$\mathbf{z} = \begin{bmatrix} \hat{a}(\mathbf{R}) - a \\ \hline \dfrac{\partial \ln p_{\mathbf{r},a}(\mathbf{R}, A)}{\partial A} \end{bmatrix}$$

and proceed as before.

Problem 2.4.23. Bhattacharyya Bound. Whenever an efficient estimate does not exist, we can improve on the Cramér-Rao inequality. In this problem we develop a conceptually simple but algebraically tedious bound for unbiased estimates of nonrandom variables.

1. Define an $(N + 1)$-dimensional vector,

$$\mathbf{z} \triangleq \begin{bmatrix} \hat{a}(\mathbf{R}) - A \\ \hline \dfrac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \dfrac{\partial p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} \\ \hline \dfrac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \dfrac{\partial^2 p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2} \\ \hline \vdots \\ \hline \dfrac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \dfrac{\partial^N p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^N} \end{bmatrix}.$$

Verify that

$$\mathbf{\Lambda}_{\mathbf{z}} \triangleq E(\mathbf{z}\mathbf{z}^T) = \left[\begin{array}{c|c|c} \sigma_\epsilon^2 & 1 & \mathbf{0} \\ \hline 1 & \multicolumn{2}{c}{} \\ \cline{1-1} & \multicolumn{2}{c}{\tilde{\mathbf{J}}} \\ \mathbf{0} & \multicolumn{2}{c}{} \end{array}\right]$$

What are the elements in $\tilde{\mathbf{J}}$? Is $\mathbf{\Lambda}_{\mathbf{z}}$ nonnegative definite? Assume that $\tilde{\mathbf{J}}$ is positive definite. When is $\mathbf{\Lambda}_{\mathbf{z}}$ *not* positive definite?

2. Verify that the results in part 1 imply

$$\sigma_\epsilon^2 \geq \tilde{J}^{11}.$$

This is the Bhattacharyya bound. Under what conditions does the equality hold?

3. Verify that for $N = 1$ the Bhattacharyya bound reduces to Cramér-Rao inequality.

4. Does the Bhattacharyya bound always improve as N increases?

Comment. In part 2 the condition for equality is

$$\hat{a}(\mathbf{R}) - A = \sum_{i=1}^{N} c_i(A) \frac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \frac{\partial^i p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^i}.$$

This condition could be termed Nth-order efficiency but does not seem to occur in many problems of interest.

5. Frequently it is easier to work with

$$\frac{\partial^i \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^i}.$$

Rewrite the elements $\tilde{J}_{ij}$ in terms of expectations of combinations of these quantities for $N = 1$ and 2.

Problem 2.4.24 (continuation). Let $N = 2$ in the preceding problem.

1. Verify that

$$\sigma_\epsilon^2 \geq \frac{1}{\tilde{J}_{11}} + \frac{\tilde{J}_{12}^2}{\tilde{J}_{11}(\tilde{J}_{11}\tilde{J}_{22} - \tilde{J}_{12}^2)}.$$

The second term represents the improvement in the bound.

2. Consider the case in which $\mathbf{r}$ consists of M independent observations with identical densities and finite conditional means and variances. Denote the elements of $\tilde{\mathbf{J}}$ due to M observations as $\tilde{J}_{ij}(M)$. Show that $\tilde{J}_{11}(M) = M\tilde{J}_{11}(1)$. Derive similar relations for $\tilde{J}_{12}(M)$ and $\tilde{J}_{22}(M)$. Show that

$$\sigma_\epsilon^2 \geq \frac{1}{M\tilde{J}_{11}(1)} + \frac{\tilde{J}_{12}^2(1)}{2M^2\tilde{J}_{11}^4(1)} + o\left(\frac{1}{M^2}\right).$$

Problem 2.4.25. [11] Generalize the result in Problem 2.4.23 to the case in which we are estimating a function of A, say $f(A)$. Assume that the estimate is unbiased. Define

$$\mathbf{z} = \begin{bmatrix} \hat{a}(\mathbf{R}) - f(A) \\ \hline k_1 \dfrac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \dfrac{\partial p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} \\ \hline k_2 \dfrac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \dfrac{\partial^2 p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2} \\ \hline \vdots \\ \hline k_N \dfrac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \dfrac{\partial^N p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^N} \end{bmatrix}.$$

Let

$$y = [\hat{a}(\mathbf{R}) - f(A)] - \sum_{i=1}^{N} k_i \frac{1}{p_{\mathbf{r}|a}(\mathbf{R}|A)} \cdot \frac{\partial^i p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^i}.$$

1. Find an expression for $\xi_y = E[y^2]$. Minimize ξ_y by choosing the k_i appropriately.
2. Using these values of k_i, find a bound on $\mathrm{Var}[\hat{a}(\mathbf{R}) - f(A)]$.
3. Verify that the result in Problem 2.4.23 is obtained by letting $f(A) = A$ in (2).

Problem 2.4.26.

1. Generalize the result in Problem 2.4.23 to establish a bound on the mean-square error in estimating a random variable.
2. Verify that the matrix of concern is

$$\mathbf{\Lambda_z} = \left[\begin{array}{c|c|c} E(a_\epsilon^2) & 1 & 0 \\ \hline 1 & \multicolumn{2}{c}{\multirow{2}{*}{$\tilde{\mathbf{J}}_T$}} \\ \cline{1-1} \mathbf{0} & & \end{array}\right].$$

What are the elements in $\tilde{\mathbf{J}}_T$?

3. Find $\boldsymbol{\Lambda}_{\mathbf{z}}$ for the special case in which a is $N(0, \sigma_a)$.

Multiple Parameters

Problem 2.4.27. In (239) we defined the partial derivative matrix $\nabla_{\mathbf{x}}$.

$$\nabla_{\mathbf{x}} \triangleq \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \dfrac{\partial}{\partial x_2} \\ \vdots \\ \dfrac{\partial}{\partial x_n} \end{bmatrix}.$$

Verify the following properties.

1. The matrix $\mathbf{A}$ is $n \times 1$ and the matrix $\mathbf{B}$ is $n \times 1$. Show that
$$\nabla_{\mathbf{x}}(\mathbf{A}^T\mathbf{B}) = (\nabla_{\mathbf{x}}\mathbf{A}^T)\mathbf{B} + (\nabla_{\mathbf{x}}\mathbf{B}^T)\mathbf{A}.$$
2. If the $n \times 1$ matrix $\mathbf{B}$ is not a function of $\mathbf{x}$, show that
$$\nabla_{\mathbf{x}}(\mathbf{B}^T\mathbf{x}) = \mathbf{B}.$$
3. Let $\mathbf{C}$ be an $n \times m$ constant matrix,
$$\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{C}) = \mathbf{C}.$$
4. $\nabla_{\mathbf{x}}(\mathbf{x}^T) = \mathbf{I}$.

Problem 2.4.28. A problem that occurs frequently is the differentiation of a quadratic form.

$$Q = \mathbf{A}^T(\mathbf{x})\,\boldsymbol{\Lambda}\mathbf{A}(\mathbf{x}),$$

where $\mathbf{A}(\mathbf{x})$ is a $m \times 1$ matrix whose elements are a function of $\mathbf{x}$ and $\boldsymbol{\Lambda}$ is a symmetric nonnegative definite $m \times m$ matrix. Recall that this implies that we can write

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}.$$

1. Prove
$$\nabla_{\mathbf{x}}Q = 2(\nabla_{\mathbf{x}}\mathbf{A}^T(\mathbf{x}))\,\boldsymbol{\Lambda}\mathbf{A}(\mathbf{x})$$
2. For the special case
$$\mathbf{A}(\mathbf{x}) = \mathbf{B}\mathbf{x},$$
prove
$$\nabla_{\mathbf{x}}Q = 2\mathbf{B}^T\boldsymbol{\Lambda}\mathbf{B}\mathbf{x}.$$
3. For the special case
$$Q = \mathbf{x}^T\boldsymbol{\Lambda}\mathbf{x},$$
prove
$$\nabla_{\mathbf{x}}Q = 2\boldsymbol{\Lambda}\mathbf{x}.$$

Problem 2.4.29. Go through the details of the proof on p. 83 for arbitrary K.

Problem 2.4.30. As discussed in (284), we frequently estimate,

$$\mathbf{d} \triangleq \mathbf{g}_d(\mathbf{A}).$$

Assume the estimates are unbiased. Derive (286).

Problem 2.4.31. The cost function is a scalar-valued function of the vector $\mathbf{a}_\epsilon$, $C(\mathbf{a}_\epsilon)$. Assume that it is symmetric and convex,

1. $C(\mathbf{a}_\epsilon) = C(-\mathbf{a}_\epsilon)$,
2. $C(b\mathbf{x}_1 + (1-b)\mathbf{x}_2) \leq bC(\mathbf{x}_1) + (1-b)\,C(\mathbf{x}_2), \qquad 0 \leq b \leq 1.$

Assume that the a posteriori density is symmetric about its conditional mean. Prove that the conditional mean of **a** minimizes the Bayes risk.

Problem 2.4.32. Assume that we want to estimate K nonrandom parameters $A_1, A_2, \ldots, A_K$, denoted by **A**. The probability density $p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})$ is known. Consider the biased estimates $\hat{\mathbf{a}}(\mathbf{R})$ in which

$$\mathrm{B}(a_i) \triangleq \int [\hat{a}_i(\mathbf{R}) - A_i] p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})\, d\mathbf{R}.$$

1. Derive a bound on the mean-square error in estimating A_i.
2. The error correlation matrix is

$$\mathbf{R}_\epsilon \triangleq E[(\hat{\mathbf{a}}(\mathbf{R}) - \mathbf{A})(\hat{\mathbf{a}}^T(\mathbf{R}) - \mathbf{A}^T)]$$

Find a matrix $\mathbf{J}_B$ such that, $\mathbf{J}_B - \mathbf{R}_\epsilon^{-1}$ is nonnegative definite.

MISCELLANEOUS

Problem 2.4.33. Another method of estimating nonrandom parameters is called the method of moments (Pearson [37]). If there are k parameters to estimate, the first k sample moments are equated to the actual moments (which are functions of the parameters of interest). Solving these k equations gives the desired estimates. To illustrate this procedure consider the following example. Let

$$\begin{aligned} p_{x|\lambda}(X|\lambda) &= \frac{1}{\Gamma(\lambda)} X^{\lambda-1} e^{-X}, \qquad && X \geq 0,\\ &= 0, && X < 0, \end{aligned}$$

where λ is a positive parameter. We have n independent observations of x.

1. Find a lower bound on the variance of any unbiased estimate.
2. Denote the method of moments estimate as $\hat{\lambda}_{mm}$. Show

$$\hat{\lambda}_{mm} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

and compute $E(\hat{\lambda}_{mm})$ and $\mathrm{Var}\,(\hat{\lambda}_{mm})$.

Comment. In [9] the efficiency of $\hat{\lambda}_{mm}$ is computed. It is less than 1 and tends to zero as $n \to \infty$.

Problem 2.4.34. Assume that we have n independent observations from a Gaussian density $N(m, \sigma)$. Verify that the method of moments estimates of m and σ are identical to the maximum-likelihood estimates.

P2.5 Composite Hypotheses

Problem 2.5.1. Consider the following composite hypothesis testing problem,

$$H_0 : p_r(R) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{R^2}{2\sigma_0^2}\right),$$

where σ_0 is known,

$$H_1 : p_r(R) = \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{R^2}{2\sigma_1^2}\right),$$

where $\sigma_1 > \sigma_0$. Assume that we require $P_F = 10^{-2}$.

1. Construct an upper bound on the power function by assuming a perfect measurement scheme coupled with a likelihood ratio test.
2. Does a uniformly most powerful test exist?
3. If the answer to part 2 is negative, construct the power function of a generalized likelihood ratio test.

Problem 2.5.2. Consider the following composite hypothesis testing problem. *Two* statistically independent observations are received. Denote the observations as R_1 and R_2. Their probability densities on the two hypotheses are

$$H_0: p_{r_i}(R_i) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{R_i^2}{2\sigma_0^2}\right), \qquad i = 1, 2,$$

where σ_0 is known,

$$H_1: p_{r_i}(R_i) = \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{R_i^2}{2\sigma_1^2}\right), \qquad i = 1, 2,$$

where $\sigma_1 > \sigma_0$. Assume that we require a $P_F = \alpha$.

1. Construct an upper bound on the power function by assuming a perfect measurement scheme coupled with a likelihood ratio test.
2. Does a uniformly most powerful test exist?
3. If the answer to part 2 is negative, construct the power function of a generalized likelihood ratio test.

Problem 2.5.3. The observation consists of a set of values of the random variables, $r_1, r_2, \ldots, r_M$.

$$\begin{aligned} r_i &= s_i + n_i, \qquad & i &= 1, 2, \ldots, M, \qquad & &H_1, \\ r_i &= n_i, \qquad & i &= 1, 2, \ldots, M, \qquad & &H_0. \end{aligned}$$

The s_i and n_i are independent, identically distributed random variables with densities $N(0, \sigma_s)$ and $N(0, \sigma_n)$, respectively, where σ_n is known and σ_s is unknown.

1. Does a UMP test exist?
2. If the answer to part 1 is negative, find a generalized LRT.

Problem 2.5.4. The observation consists of a set of values of the random variables $r_1, r_2, \ldots, r_M$, which we denote by the vector $\mathbf{r}$. Under H_0 the r_i are statistically independent, with densities

$$p_{r_i}(R_i) = \frac{1}{\sqrt{2\pi\lambda_i^0}} \exp\left(-\frac{R_i^2}{2\lambda_i^0}\right)$$

in which the λ_i^0 are known. Under H_1 the r_i are statistically independent, with densities

$$p_{r_i}(R_i) = \frac{1}{\sqrt{2\pi\lambda_i^1}} \exp\left(-\frac{R_i^2}{2\lambda_i^1}\right)$$

in which $\lambda_i^1 > \lambda_i^0$ for all i. Repeat Problem 2.5.3.

Problem 2.5.5. Consider the following hypothesis testing problem. *Two* statistically independent observations are received. Denote the observations R_1 and R_2. The probability densities on the two hypotheses are

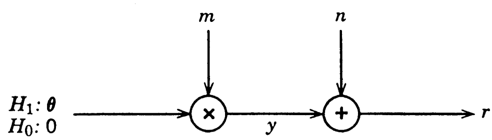
$$H_0: p_{r_i}(R_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{R_i^2}{2\sigma^2}\right), \qquad i = 1, 2,$$

$$H_1: p_{r_i}(R_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(R_i - m)^2}{2\sigma^2}\right] \qquad i = 1, 2,$$

where m can be *any* nonzero number. Assume that we require $P_F = \alpha$.

1. Construct an upper bound on the power function by assuming a perfect measurement scheme coupled with a likelihood ratio test.
2. Does a uniformly most powerful test exist?
3. If the answer to part 2 is negative, construct the power function of a generalized likelihood ratio test.

Problem 2.5.6. Consider the following hypothesis-testing problem.



Under H_1 a nonrandom variable $\theta\,(-\infty < \theta < \infty)$ is transmitted. It is multiplied by the random variable m. A noise n is added to the result to give r. Under H_0 nothing is transmitted, and the output is just n. Thus

$$H_1: r = m\theta + n,$$
$$H_0: r = n.$$

The random variables m and n are independent.

$$p_n(N) = \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left(-\frac{N^2}{2\sigma_n^2}\right),$$

$$p_m(M) = \tfrac{1}{2}\,\delta(M-1) + \tfrac{1}{2}\,\delta(M+1).$$

1. Does a uniformly most powerful test exist? If it does, describe the test and give an expression for its power function? If it does not, indicate why.
2. Do one of the following:
   (a) If a UMP test exists for this example, derive a necessary and sufficient condition on $p_m(M)$ for a UMP test to exist. (The rest of the model is unchanged.)
   (b) If a UMP test does not exist, derive a generalized likelihood ratio test and an expression for its power function.

Problem 2.5.7 (CFAR receivers.) We have N independent observations of the variable x. The probability density on H_k is

$$p_{x_i|H_k}(X|H_k) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{\frac{-(X_i - m_k)^2}{2\sigma^2}\right\} \qquad -\infty < X_i < \infty, \qquad \begin{aligned} i &= 1, 2, \ldots N, \\ H_k: k &= 0,1, \\ m_0 &= 0. \end{aligned}$$

The variance σ^2 is *unknown*. Define

$$l_1 = \sum_{i=1}^{N} x_i$$

$$l_2 = \sum_{i=1}^{N} x_i^2$$

(a) Consider the test

$$l_1^2 \underset{H_0}{\overset{H_1}{\gtrless}} \alpha l_2$$

Verify that the P_F of this test does *not* depend on σ^2. (*Hint.* Use formula in Problem 2.4.6.)

(b) Find α as a function of P_F.
(c) Is this a UMP test?
(d) Consider the particular case in which $N = 2$ and $m_1 = m$. Find P_D as a function of P_F and m/σ. Compare your result with Figure 2.9*b* and see how much the lack of knowledge about the variance σ^2 has decreased the system performance.

Comment. Receivers of this type are called CFAR (constant false alarm rate) receivers in the radar/sonar literature.

Problem 2.5.8 (continuation). An alternate approach to the preceding problem would be a generalized LRT.

1. Find the generalized LRT and write an expression for its performance for the case in which $N = 2$ and $m_1 = m$.
2. How would you decide which test to use?

Problem 2.5.9. Under H_0, x is a Poisson variable with a known intensity k_0.

$$\Pr(x = n) = \frac{k_0^{\,n}}{n!}e^{-k_0}, \qquad n = 0, 1, 2, \ldots.$$

Under H_1, x is a Poisson variable with an unknown intensity k_1, where $k_1 > k_0$.

1. Does a UMP test exist?
2. If a UMP test does not exist, assume that M independent observations of x are available and construct a generalized LRT.

Problem 2.5.10. How are the results to Problem 2.5.2 changed if we know that $\sigma_0 < \sigma_c$ and $\sigma_1 > \sigma_c$ where σ_c is known. Neither σ_0 or σ_1, however, is known. If a UMP test does not exist, what test procedure (other than a generalized LRT) would be logical?

P2.6 General Gaussian Problem

Detection

Problem 2.6.1. The M-hypothesis, general Gaussian problem is

$$p_{\mathbf{r}|H_i}(\mathbf{R}|H_i) = [(2\pi)^{N/2}|\mathbf{K}_i|^{1/2}]^{-1}\exp\left[-\tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_i^T)\,\mathbf{Q}_i(\mathbf{R} - \mathbf{m}_i)\right], \qquad i = 1, 2, \ldots, M.$$

1. Use the results of Problem 2.3.2 to find the Bayes test for this problem.
2. For the particular case in which the cost of a correct decision is zero and the cost of any wrong decision is equal show that the test reduces to the following:

Compute

$$l_i(\mathbf{R}) = \ln P_i - \tfrac{1}{2}\ln|\mathbf{K}_i| - \tfrac{1}{2}(\mathbf{R}^T - \mathbf{m}_i^T)\,\mathbf{Q}_i(\mathbf{R} - \mathbf{m}_i)$$

and choose the largest.

Problem 2.6.2 (continuation). Consider the special case in which all $\mathbf{K}_i = \sigma_n^2\mathbf{I}$ and the hypotheses are equally likely. Use the costs in Part 2 of Problem 2.6.1.

1. What determines the dimension of the decision space? Draw some typical decision spaces to illustrate the various alternatives.
2. Interpret the processor as a minimum-distance decision rule.

Problem 2.6.3. Consider the special case in which $\mathbf{m}_i = 0$, $i = 1, 2, \ldots, M$, and the hypotheses are equally likely. Use the costs in Part 2 of Problem 2.6.1.

1. Show that the test reduces to the following:

Compute

$$l_i(\mathbf{R}) = \mathbf{R}^T\mathbf{Q}_i\mathbf{R} + \ln|\mathbf{K}_i|$$

and choose the *smallest*.

2. Write an expression for the Pr (ϵ) in terms of $p_{l|H_k}(\mathbf{L}|H_k)$, where

$$l \triangleq \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_M \end{bmatrix}.$$

Problem 2.6.4. Let

$$q_{\mathbf{B}} \triangleq \mathbf{x}^T\mathbf{B}\mathbf{x},$$

where $\mathbf{x}$ is a Gaussian vector $N(\mathbf{0}, \mathbf{I})$ and $\mathbf{B}$ is a symmetric matrix.

1. Verify that the characteristic function of $q_{\mathbf{B}}$ is

$$M_{q_{\mathbf{B}}}(jv) \triangleq E(e^{jvq_{\mathbf{B}}}) = \prod_{i=1}^{N} (1 - 2jv\lambda_{\mathbf{B}i})^{-1/2},$$

where $\lambda_{\mathbf{B}i}$ are the eigenvalues of $\mathbf{B}$.

2. What is $p_{q\mathbf{B}}(Q)$ when the eigenvalues are equal?

3. What is the form of $p_{q\mathbf{B}}(Q)$ when N is even and the eigenvalues are pair-wise equal but otherwise distinct; that is,

$$\lambda_{2i-1} = \lambda_{2i}, \qquad i = 1, 2, \ldots, \frac{N}{2},$$
$$\lambda_{2i} \neq \lambda_{2j}, \qquad \text{all } i \neq j.$$

Problem 2.6.5.

1. Modify the result of the preceding problem to include the case in which $\mathbf{x}$ is a Gaussian vector $N(\mathbf{0}, \mathbf{\Lambda_x})$, where $\mathbf{\Lambda_x}$ is positive definite.

2. What is $M_{q_{\mathbf{\Lambda_x}^{-1}}}(jv)$? Does the result have any interesting features?

Problem 2.6.6. Consider the M-ary hypothesis-testing problem. *Each* observation is a three-dimensional vector.

$$\begin{aligned} H_0&: \mathbf{r} = \mathbf{m}_0 + \mathbf{n}, \\ H_1&: \mathbf{r} = \mathbf{m}_1 + \mathbf{n}, \\ H_2&: \mathbf{r} = \mathbf{m}_2 + \mathbf{n}, \\ H_3&: \mathbf{r} = \mathbf{m}_3 + \mathbf{n}, \\ \mathbf{m}_0 &= +A, 0, B, \\ \mathbf{m}_1 &= 0, +A, B, \\ \mathbf{m}_2 &= -A, 0, B, \\ \mathbf{m}_3 &= 0, -A, B. \end{aligned}$$

The components of the noise vector are independent, identically distributed Gaussian variables, $N(0, \sigma)$. We have K independent observations. Assume a minimum Pr(ϵ) criterion and equally-likely hypotheses. Sketch the decision region and compute the Pr(ϵ).

Problem 2.6.7. Consider the following detection problem. Under either hypothesis the observation is a *two*-dimensional vector $\mathbf{r}$.

Under H_1:

$$\mathbf{r} \triangleq \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \mathbf{x} + \mathbf{n}.$$

Under H_0:

$$\mathbf{r} \triangleq \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \mathbf{y} + \mathbf{n}.$$

The signal vectors $\mathbf{x}$ and $\mathbf{y}$ are known. The length of the signal vector is constrained to equal $\sqrt{E}$ under both hypotheses; that is,

$$x_1^2 + x_2^2 = E,$$
$$y_1^2 + y_2^2 = E.$$

The noises are *correlated* Gaussian variables.

$$p_{n_1 n_2}(N_1, N_2) = \frac{1}{2\pi\sigma^2(1-\rho^2)^{1/2}} \exp\left(-\frac{N_1^2 - 2\rho N_1 N_2 + N_2^2}{2\sigma^2(1-\rho^2)}\right).$$

1. Find a sufficient statistic for a likelihood ratio test. Call this statistic $l(\mathbf{R})$. We have already shown that the quantity

$$d^2 = \frac{[E(l|H_1) - E(l|H_0)]^2}{\mathrm{Var}\,(l|H_0)}$$

characterizes the performance of the test in a monotone fashion.

2. Choose $\mathbf{x}$ and $\mathbf{y}$ to maximize d^2. Does the answer depend on ρ?
3. Call the d^2 obtained by using the best $\mathbf{x}$ and $\mathbf{y}$, d_0^2. Calculate d_0^2 for $\rho = -1, 0$, and draw a rough sketch of d_0^2 as ρ varies from -1 through 0 to 1.
4. Explain why the performance curve in part 3 is intuitively correct.

ESTIMATION

Problem 2.6.8. The observation is an N-dimensional vector

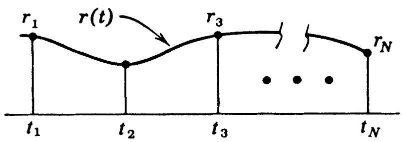
$$\mathbf{r} = \mathbf{a} + \mathbf{n},$$

where $\mathbf{a}$ is $N(\mathbf{0}, \mathbf{K_a})$, $\mathbf{n}$ is $N(\mathbf{0}, \mathbf{K_n})$, and $\mathbf{a}$ and $\mathbf{n}$ are statistically independent.

1. Find $\hat{\mathbf{a}}_{\mathrm{map}}$. *Hint.* Use the properties of $\nabla_{\mathbf{a}}$ developed in Problems 2.4.27 and 2.4.28.
2. Verify that $\hat{\mathbf{a}}_{\mathrm{map}}$ is efficient.
3. Compute the error correlation matrix

$$\mathbf{\Lambda}_\epsilon \triangleq E[(\hat{\mathbf{a}}_{\mathrm{ms}} - \mathbf{a})(\hat{\mathbf{a}}_{\mathrm{ms}} - \mathbf{a})^T].$$

Comment. Frequently this type of observation vector is obtained by sampling a random process $r(t)$ as shown below,



We denote the N samples by the vector $\mathbf{r}$. Using $\mathbf{r}$, we estimate the samples of $a(t)$ which are denoted by a_i. An error of interest is the sum of the squares of errors in estimating the a_i.

$$a_{\epsilon_i} = \hat{a}_i - a,$$

then

$$\xi_I \triangleq E\left[\sum_{i=1}^{N} (\hat{a}_i - a)^2\right] = E\left(\sum_{i=1}^{N} a_{\epsilon_i}{}^2\right) = E(\mathbf{a}_\epsilon{}^T \mathbf{a}_\epsilon) = \mathrm{Tr}\,(\mathbf{\Lambda}_\epsilon).$$

Problem 2.6.9 (continuation). Consider the special case

$$\mathbf{K}_n = \sigma_n^2 \mathbf{I}.$$

1. Verify that

$$\hat{\mathbf{a}}_{ms} = (\sigma_n^2\mathbf{I} + \mathbf{K}_a)^{-1}\mathbf{K}_a\mathbf{R}.$$

2. Now recall the detection problem described in Case 1 on p. 107. Verify that

$$l(\mathbf{R}) = \frac{1}{\sigma_n^2}\mathbf{R}^T\hat{\mathbf{a}}_{ms}.$$

Draw a block diagram of the processor. Observe that this is identical to the "unequal mean-equal covariance" case, except the mean $\mathbf{m}$ has been replaced by the mean-square estimate of the mean, $\hat{\mathbf{a}}_{ms}$.

3. What is the mean-square estimation error ξ_I?

Problem 2.6.10. Consider an alternate approach to Problem 2.6.8.

$$\mathbf{r} = \mathbf{a} + \mathbf{n},$$

where $\mathbf{a}$ is $N(\mathbf{0}, \mathbf{K_a})$ and $\mathbf{n}$ is $N(\mathbf{0}, \sigma_n^2\mathbf{I})$. Pass $\mathbf{r}$ through the matrix operation $\mathbf{W}$, which is defined in (369). The eigenvectors are those of $\mathbf{K}_a$.

$$\mathbf{r}' \triangleq \mathbf{W}\mathbf{r} = \mathbf{x} + \mathbf{n}'$$

1. Verify that $\mathbf{W}\mathbf{W}^T = \mathbf{I}$.
2. What are the statistics of $\mathbf{x}$ and $\mathbf{n}'$?
3. Find $\hat{\mathbf{x}}$. Verify that

$$\hat{x}_i = \frac{\lambda_i}{\lambda_i + \sigma_n^2} R_i',$$

where λ_i are the eigenvalues of $\mathbf{K_a}$.

4. Express $\hat{\mathbf{a}}$ in terms of a linear transformation of $\hat{\mathbf{x}}$. Draw a block diagram of the over-all estimator.
5. Prove

$$\xi_I \triangleq E[\mathbf{a}_\epsilon^T\mathbf{a}_\epsilon] = \sigma_n^2 \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i + \sigma_n^2}.$$

Problem 2.6.11 (***Nonlinear Estimation***). In the general Gaussian nonlinear estimation problem

$$\mathbf{r} = \mathbf{s}(\mathbf{A}) + \mathbf{n},$$

where $\mathbf{s}(\mathbf{A})$ is a nonlinear function of $\mathbf{A}$. The noise $\mathbf{n}$ is Gaussian $N(\mathbf{0}, \mathbf{K_n})$ and independent of $\mathbf{a}$.

1. Verify that

$$p_{\mathbf{r}|\mathbf{s}(\mathbf{A})}(\mathbf{R}|\mathbf{s}(\mathbf{A})) = [(2\pi)^{N/2}|\mathbf{K_n}|^{1/2}]^{-1} \exp\left[-\tfrac{1}{2}(\mathbf{R}^T - \mathbf{s}^T(\mathbf{A}))\mathbf{Q_n}(\mathbf{R} - \mathbf{s}(\mathbf{A}))\right].$$

2. Assume that $\mathbf{a}$ is a Gaussian vector $N(\mathbf{0}, \mathbf{K_a})$. Find an expression for $\ln p_{\mathbf{r},\mathbf{a}}(\mathbf{R}, \mathbf{A})$.
3. Using the properties of the derivative matrix $\nabla_{\mathbf{a}}$ derived in Problems 2.4.27 and 2.4.28, find the MAP equation.

Problem 2.6.12 (***Optimum Discrete Linear Filter***). Assume that we have a sequence of scalar observations $r_1, r_2, r_3, \ldots, r_K$, where $r_i = a_i + n_i$ and

$$E(a_i) = E(n_i) = 0,$$

$$E(\mathbf{r}\mathbf{r}^T) = \mathbf{\Lambda_r}, \qquad (N \times N),$$
$$E(\mathbf{r}a_i) = \mathbf{\Lambda}_{\mathbf{r}a_i}, \qquad (N \times 1).$$

We want to estimate a_K by using a realizable discrete linear filter. Thus

$$\hat{a}_K = \sum_{i=1}^{K} h_i R_i = \mathbf{h}^T\mathbf{R}.$$

Define the mean-square point estimation error as

$$\xi_P \triangleq E\{[\hat{a}_K(\mathbf{R}) - a_K]^2\}.$$

1. Use $\nabla_{\mathbf{h}}$ to find the discrete linear filter that minimizes ξ_P.
2. Find ξ_P for the optimum filter.
3. Consider the special case in which $\mathbf{a}$ and $\mathbf{n}$ are statistically independent. Find $\mathbf{h}$ and ξ_P.
4. How is $\hat{a}_K(\mathbf{R})$ for part 3 related to $\hat{\mathbf{a}}_{map}$ in Problem 2.6.8.

Note. No assumption about Gaussianness has been used.

Sequential Estimation

Problem 2.6.13. Frequently the observations are obtained in a time-sequence, $r_1, r_2, r_3, \ldots, r_N$. We want to estimate the k-dimensional parameter $\mathbf{a}$ in a sequential manner.

The ith observation is

$$r_i = \mathbf{C}\mathbf{a} + w_i, \qquad i = 1, 2, \ldots, N,$$

where $\mathbf{C}$ is a known $1 \times k$ matrix. The noises w_i are independent, identically distributed Gaussian variables $N(0, \sigma_n)$. The a priori knowledge is that $\mathbf{a}$ is Gaussian, $N(\mathbf{m}_0, \mathbf{\Lambda_a})$.

1. Find $p_{\mathbf{a}|r_1}(\mathbf{A}|R_1)$.
2. Find the minimum mean-square estimate $\hat{\mathbf{a}}_1$ and the error correlation matrix $\mathbf{\Lambda}_{\epsilon_1}$. Put your answer in the form

$$p_{\mathbf{a}|r_1}(\mathbf{A}|R_1) = c \exp\left[-\tfrac{1}{2}(\mathbf{A} - \hat{\mathbf{a}}_1)^T \mathbf{\Lambda}_{\epsilon_1}^{-1}(\mathbf{A} - \hat{\mathbf{a}}_1)\right],$$

where

$$\mathbf{\Lambda}_{\epsilon_1}^{-1} = \mathbf{\Lambda_a}^{-1} + \mathbf{C}^T \sigma_n^{-2} \mathbf{C}$$

and

$$\hat{\mathbf{a}}_1 = \mathbf{m}_0 + \frac{1}{\sigma_n^2} \mathbf{\Lambda}_{\epsilon_1} \mathbf{C}^T (R_1 - \mathbf{C}\mathbf{m}_0).$$

3. Draw a block diagram of the optimum processor.
4. Now proceed to the second observation R_2. What is the a priori density for this observation? Write the equations for $p_{\mathbf{a}|r_1,r_2}(\mathbf{A}|r_1, r_2)$, $\mathbf{\Lambda}_{\epsilon_2}^{-1}$, and $\hat{\mathbf{a}}_2$ in the same format as above.
5. Draw a block diagram of the sequential estimator and indicate exactly what must be stored at the end of each estimate.

Problem 2.6.14. Problem 2.6.13 can be generalized by allowing each observation to be an m-dimensional vector. The ith observation is

$$\mathbf{r}_i = \mathbf{C}\mathbf{a} + \mathbf{w}_i,$$

where $\mathbf{C}$ is a known $m \times k$ matrix. The noise vectors $\mathbf{w}_i$ are independent, identically distributed Gaussian vectors, $N(\mathbf{0}, \mathbf{\Lambda_w})$, where $\mathbf{\Lambda_w}$ is positive-definite.

Repeat Problem 2.6.13 for this model. Verify that

$$\hat{\mathbf{a}}_i = \hat{\mathbf{a}}_{i-1} + \mathbf{\Lambda}_{\epsilon_i} \mathbf{C}^T \mathbf{\Lambda_w}^{-1} (\mathbf{R}_i - \mathbf{C}\hat{\mathbf{a}}_{i-1})$$

and

$$\mathbf{\Lambda}_{\epsilon_i}^{-1} = \mathbf{\Lambda}_{\epsilon_{i-1}}^{-1} + \mathbf{C}^T \mathbf{\Lambda_w}^{-1} \mathbf{C}.$$

Draw a block diagram of the optimum processor.

Problem 2.6.15. **Discrete Kalman Filter.** Now consider the case in which the parameter $\mathbf{a}$ changes according to the equation

$$\mathbf{a}_{k+1} = \mathbf{\Phi a}_k + \mathbf{\Gamma u}_k, \qquad k = 1, 2, 3, \ldots,$$

where $\mathbf{a}_1$ is $N(\mathbf{m}_0, \mathbf{P}_0)$, $\mathbf{\Phi}$ is an $n \times n$ matrix (known), $\mathbf{\Gamma}$ is an $n \times p$ matrix (known), $\mathbf{u}_k$ is $N(\mathbf{0}, \mathbf{Q})$, and $\mathbf{u}_k$ is independent of $\mathbf{u}_j$ for $j \neq k$. The observation process is

$$\mathbf{r}_k = \mathbf{Ca}_k + \mathbf{w}_k, \qquad k = 1, 2, 3, \ldots,$$

where $\mathbf{C}$ is an $m \times n$ matrix, $\mathbf{w}_k$ is $N(\mathbf{0}, \mathbf{\Lambda_w})$ and the $\mathbf{w}_k$ are independent of each other and $\mathbf{u}_j$.

PART I. We first estimate $\mathbf{a}_1$, using a mean-square error criterion.

1. Write $p_{\mathbf{a}_1|\mathbf{r}_1}(\mathbf{A}_1|\mathbf{R}_1)$.
2. Use the $\nabla_{\mathbf{a}_1}$ operator to obtain $\hat{\mathbf{a}}_1$.
3. Verify that $\hat{\mathbf{a}}_1$ is efficient.
4. Use $\nabla_{\mathbf{a}_1}\{[\nabla_{\mathbf{a}_1}(\ln p_{\mathbf{a}_1|\mathbf{r}_1}(\mathbf{A}_1|\mathbf{R}_1))]^T\}$ to find the error covariance matrix $\mathbf{P}_1$,

where

$$\mathbf{P}_i \triangleq E[(\hat{\mathbf{a}}_i - \mathbf{a}_i)(\hat{\mathbf{a}}_i - \mathbf{a}_i)^T], \qquad i = 1, 2, \ldots.$$

Check.

$$\hat{\mathbf{a}}_1 = \mathbf{m}_0 + \mathbf{P}_1\mathbf{C}^T\mathbf{\Lambda_w}^{-1}[\mathbf{R} - \mathbf{Cm}_0]$$

and

$$\mathbf{P}_1^{-1} = \mathbf{P}_0^{-1} + \mathbf{C}^T\mathbf{\Lambda_w}^{-1}\mathbf{C}.$$

PART II. Now we estimate $\mathbf{a}_2$.

1. Verify that

$$p_{\mathbf{a}_2|\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}_2|\mathbf{R}_1, \mathbf{R}_2) = \frac{p_{\mathbf{r}_2|\mathbf{a}_2}(\mathbf{R}_2|\mathbf{A}_2)\, p_{\mathbf{a}_2|\mathbf{r}_1}(\mathbf{A}_2|\mathbf{R}_1)}{p_{\mathbf{r}_2|\mathbf{r}_1}(\mathbf{R}_2|\mathbf{R}_1)}.$$

2. Verify that $p_{\mathbf{a}_2|\mathbf{r}_1}(\mathbf{A}_2|\mathbf{R}_1)$ is $N(\mathbf{\Phi}\hat{\mathbf{a}}_1, \mathbf{M}_2)$, where

$$\mathbf{M}_2 \triangleq \mathbf{\Phi P}_1\mathbf{\Phi}^T + \mathbf{\Gamma Q \Gamma}^T.$$

3. Find $\hat{\mathbf{a}}_2$ and $\mathbf{P}_2$.

Check.

$$\hat{\mathbf{a}}_2 = \mathbf{\Phi}\hat{\mathbf{a}}_1 + \mathbf{P}_2\mathbf{C}^T\mathbf{\Lambda_w}^{-1}(\mathbf{R}_2 - \mathbf{C\Phi}\hat{\mathbf{a}}_1),$$
$$\mathbf{P}_2^{-1} = \mathbf{M}_2^{-1} + \mathbf{C}^T\mathbf{\Lambda_w}^{-1}\mathbf{C}.$$

4. Write

$$\mathbf{P}_2 = \mathbf{M}_2 - \mathbf{B}$$

and verify that $\mathbf{B}$ must equal

$$\mathbf{B} = \mathbf{M}_2\mathbf{C}^T(\mathbf{CM}_2\mathbf{C}^T + \mathbf{\Lambda_w})^{-1}\mathbf{CM}_2.$$

5. Verify that the answer to part 3 can be written as

$$\hat{\mathbf{a}}_2 = \mathbf{\Phi}\hat{\mathbf{a}}_1 + \mathbf{M}_2\mathbf{C}^T(\mathbf{CM}_2\mathbf{C}^T + \mathbf{\Lambda_w})^{-1}(\mathbf{R}_2 - \mathbf{C\Phi}\hat{\mathbf{a}}_1).$$

Compare the two forms with respect to ease of computation. What is the dimension of the matrix to be inverted?

PART III

1. Extend the results of Parts I and II to find an expression for $\hat{\mathbf{a}}_k$ and $\mathbf{P}_k$ in terms of $\hat{\mathbf{a}}_{k-1}$ and $\mathbf{M}_k$. The resulting equations are called the Kalman filter equations for discrete systems [38].
2. Draw a block diagram of the optimum processor.

PART IV. Verify that the Kalman filter reduces to the result in Problem 2.6.13 when $\mathbf{\Phi} = \mathbf{I}$ and $\mathbf{Q} = \mathbf{0}$.

Special Applications

A large number of problems in the areas of pattern recognition, learning systems, and system equalization are mathematically equivalent to the general Gaussian problem. We consider three simple problems (due to M. E. Austin) in this section. Other examples more complex in detail but not in concept are contained in the various references.

Problem 2.6.16. Pattern Recognition. A pattern recognition system is to be implemented for the classification of noisy samples taken from a set of M patterns. Each pattern may be represented by a set of parameters in which the mth pattern is characterized by the vector $\mathbf{s}_m$. In general, the $\mathbf{s}_m$ vectors are unknown. The samples to be classified are of the form

$$\mathbf{x} = \mathbf{s}_m + \mathbf{n},$$

where the $\mathbf{s}_m$ are assumed to be independent Gaussian random variables with mean $\bar{\mathbf{s}}_m$ and covariance $\mathbf{\Lambda}_m$ and $\mathbf{n}$ is assumed to be zero-mean Gaussian with covariance $\mathbf{\Lambda}_\mathbf{n}$ independent from sample to sample, and independent of $\mathbf{s}_m$.

1. In order to classify the patterns the recognition systems needs to know the pattern characteristics. We provide it with a "learning" sample:

$$\mathbf{x}_m = \mathbf{s}_m + \mathbf{n},$$

where the system knows that the mth pattern is present.

Show that if J learning samples, $\mathbf{x}_m^{(1)}, \mathbf{x}_m^{(2)}, \ldots, \mathbf{x}_m^{(J)}$, of the form $\mathbf{x}_m^{(j)} = \mathbf{s}_m + \mathbf{n}^{(j)}$ are available for each $m = 1, \ldots, M$, the pattern recognition system need store only the quantities

$$\mathbf{l}_m = \frac{1}{J} \sum_{j=1}^{J} \mathbf{x}_m{}^{(j)}$$

for use in classifying additional noisy samples; that is, show that the $\mathbf{l}_m$, $m = 1, \ldots M$' form a set of sufficient statistics extracted from the MJ learning samples.

2. What is the MAP estimate of $\mathbf{s}_m$? What is the covariance of this estimate as a function of J, the number of learning samples?

3. For the special case of two patterns ($M = 2$) characterized by unknown scalars s_1 and s_2, which have a priori densities $N(\bar{s}_1, \sigma)$ and $N(\bar{s}_2, \sigma)$, respectively, find the optimum decision rule for equiprobable patterns and observe that this approaches the decision rule of the "known patterns" classifier asymptotically with increasing number of learning samples J.

Problem 2.6.17. Intersymbol Interference. Data samples are to be transmitted over a known dispersive channel with an impulse response $h(t)$ in the presence of white Gaussian noise. The received waveform

$$r(t) = \sum_{k=-K}^{K} \xi_k h(t - kT) + n(t)$$

may be passed through a filter matched to the channel impulse response to give a set of numbers

$$a_j = \int r(t)\, h(t - jT)\, dt$$

for $j = 0, \pm 1, \pm 2, \ldots, \pm K$, which forms a set of sufficient statistics in the MAP estimation of the ξ_k. (This is proved in Chapter 4.) We denote the sampled channel autocorrelation function as

$$b_j = \int h(t)\, h(t - jT)\, dt$$

and the noise at the matched filter output as

$$n_j = \int n(t)\, h(t - jT)\, dt.$$

The problem then reduces to an estimation of the ξ_k, given a set of relations

$$a_j = \sum_{k=-K}^{K} \xi_k b_{j-k} + n_j \qquad \text{for } j, k = 0, \pm 1, \pm 2, \dots \pm K.$$

Using obvious notation, we may write these equations as

$$\mathbf{a} = \mathbf{B}\boldsymbol{\xi} + \mathbf{n}.$$

1. Show that if $n(t)$ has double-sided spectral height $\frac{1}{2}N_0$, that the noise vector $\mathbf{n}$ has a covariance matrix $\boldsymbol{\Lambda}_\mathbf{n} = \frac{1}{2}N_0\mathbf{B}$.
2. If the ξ_k are zero-mean Gaussian random variables with covariance matrix $\boldsymbol{\Lambda}_{\boldsymbol{\xi}}$ show that the MAP estimate of $\boldsymbol{\xi}$ is of the form $\boldsymbol{\xi} = \mathbf{Ga}$ and therefore that $\hat{\boldsymbol{\xi}}_0 = \mathbf{g}^T\mathbf{a}$. Find $\mathbf{g}$ and note that the estimate of ξ_0 can be obtained by passing the sufficient statistics into a tapped delay line with tap gains equal to the elements of $\mathbf{g}$. This cascading of a matched filter followed by a sampler and a transversal filter is a well-known equalization method employed to reduce intersymbol interference in digital communication via dispersive media.

Problem 2.6.18. Determine the MAP estimate of ξ_0 in Problem 2.6.17; assuming further that the ξ_k are independent and that the ξ_k are known (say through a "teacher" or infallible estimation process) for $k < 0$. Show then that the weighting of the sufficient statistics is of the form

$$\hat{\xi}_0 = \sum_{j>0} g_j a_j - \sum_{j<0} f_j \xi_j$$

and find g_j and f_j. This receiver may be interpreted as passing the sampled matched-filter output through a transversal filter with tap gains g_j and subtracting the output from a second transversal filter whose input is the sequence of ξ_k which estimates have been made. Of course, in implementation such a receiver would be self-taught by using its earlier estimates as correct in the above estimation equation.

Problem No. 2.6.19. Let

$$z = \mathbf{G}^T\mathbf{r}$$

and assume that z is $N(m_z, \sigma_z)$ for all finite $\mathbf{G}$.

1. What is $M_z(jv)$? Express your result in terms of m_z and σ_z.
2. Rewrite the result in (1) in terms of $\mathbf{G}$, $\mathbf{m}$, and $\boldsymbol{\Lambda}_\mathbf{r}$ [see (316)–(317) for definitions].
3. Observe that

$$M_z(ju) \triangleq E[e^{juz}] = E[e^{ju\mathbf{G}^T\mathbf{r}}]$$

and

$$M_\mathbf{r}(j\mathbf{v}) \triangleq E[e^{j\mathbf{v}^T\mathbf{r}}]$$

and therefore

$$M_z(ju) = M_\mathbf{r}(j\mathbf{v}) \quad \text{if} \quad \mathbf{G}u = \mathbf{v}.$$

Use these observations to verify (317).

Problem No. 2.6.20 (continuation).

(a) Assume that the $\mathbf{\Lambda_r}$ defined in (316) is positive definite. Verify that the expression for $p_{\mathbf{r}}(\mathbf{R})$ in (318) is correct. [*Hint.* Use the diagonalizing transformation $\mathbf{W}$ defined in (368).]

(b) How must (318) be modified if $\mathbf{\Lambda_r}$ is singular? What does this singularity imply about the components of $\mathbf{r}$?

P2.7 Performance Bounds and Approximations

Problem 2.7.1. Consider the binary test with N independent observations, r_i, where

$$p_{r_i|H_k} = N(m_k, \sigma_k), \qquad k = 0, 1, \quad i = 1, 2, \ldots, N.$$

Find $\mu(s)$.

Problem 2.7.2 (continuation). Consider the special case of Problem 2.7.1 in which

$$m_0 = 0,$$
$$\sigma_0^2 = \sigma_n^2,$$

and

$$\sigma_1^2 = \sigma_s^2 + \sigma_n^2.$$

1. Find $\mu(s)$, $\dot{\mu}(s)$, and $\ddot{\mu}(s)$.
2. Assuming equally likely hypotheses, find an upper bound on the minimum $\Pr(\epsilon)$.
3. With the assumption in part 2, find an approximate expression for the $\Pr(\epsilon)$ that is valid for large N.

Problem 2.7.3. A special case of the binary Gaussian problem with N observations is

$$p_{\mathbf{r}|H_k}(\mathbf{R}|H_k) = \frac{1}{(2\pi)^{N/2}|\mathbf{K}_k|^{1/2}} \exp\left(-\frac{\mathbf{R}^T\mathbf{K}_k^{-1}\mathbf{R}}{2}\right), \qquad k = 0, 1.$$

1. Find $\mu(s)$.
2. Express it in terms of the eigenvalues of the appropriate matrices.

Problem 2.7.4 (continuation). Consider the special case in which

$$\mathbf{K}_0 = \sigma_n^2\mathbf{I}$$

and

$$\mathbf{K}_1 = \mathbf{K}_s + \mathbf{K}_0.$$

Find $\mu(s)$, $\dot{\mu}(s)$, $\ddot{\mu}(s)$.

Problem 2.7.5 (alternate continuation of 2.7.3). Consider the special case in which $\mathbf{K}_1$ and $\mathbf{K}_0$ are partitioned into the 4 $N \times N$ matrices given by (422) and (423).

1. Find $\mu(s)$.
2. Assume that the hypotheses are equally likely and that the criterion is minimum $\Pr(\epsilon)$. Find a bound on the $\Pr(\epsilon)$.
3. Find an approximate expression for the $\Pr(\epsilon)$.

Problem 2.7.6. The general binary Gaussian problem for N observations is

$$p_{\mathbf{r}|H_k}(\mathbf{R}|H_k) = \frac{1}{(2\pi)^{N/2}|\mathbf{K}_k|^{1/2}} \exp\left[-\frac{(\mathbf{R}^T - \mathbf{m}_k^T)\mathbf{K}_k^{-1}(\mathbf{R} - \mathbf{m}_k)}{2}\right], \qquad k = 0, 1.$$

Find $\mu(s)$.

Problem 2.7.7. Consider Example 3A on p. 130. A bound on the $\Pr(\epsilon)$ is

$$\Pr(\epsilon) \leq \tfrac{1}{2}\left[\frac{(1 + \sigma_s^2/\sigma_n^2)}{(1 + \sigma_s^2/2\sigma_n^2)^2}\right]^{N/2}$$

1. Constrain $N\sigma_s^2/\sigma_n^2 = x$. Find the value of N that minimizes the bound.
2. Evaluate the approximate expression in (516) for this value of N.

Problem 2.7.8. We derived the Chernoff bound in (461) by using tilted densities. This approach prepared us for the central limit theorem argument in the second part of our discussion. If we are interested only in (461), a much simpler derivation is possible.

1. Consider a function of the random variable x which we denote as $f(x)$. Assume

$$f(x) \geq 0, \qquad \text{all } x,$$

$$f(x) \geq f(X_0) > 0, \qquad \text{all } x \geq X_0.$$

Prove

$$\Pr\,[x \geq X_0] \leq \frac{E[f(x)]}{f(X_0)}.$$

2. Now let

$$f(x) = e^{sx}, \qquad s \geq 0,$$

and

$$X_0 = \gamma.$$

Use the result in (1) to derive (457). What restrictions on γ are needed to obtain (461)?

Problem 2.7.9. The reason for using tilted densities and Chernoff bounds is that a straightforward application of the central limit theorem gives misleading results when the region of interest is on the tail of the density. A trivial example taken from [4-18] illustrates this point.

Consider a set of statistically independent random variables x_i which assumes values 0 and 1 with equal probability. We are interested in the probability

$$\Pr\left[y_N = \frac{1}{N}\sum_{i=1}^{N} x_i \geq 1\right] \triangleq \Pr\,[A_N].$$

(a) Define a standardized variable

$$z \triangleq \frac{y_N - \bar{y}_N}{\sigma_{y_N}}.$$

Use a central limit theorem argument to estimate $\Pr\,[A_N]$. Denote this estimate as $\widehat{\Pr}\,[A_N]$.

(b) Calculate $\Pr\,[A_N]$ exactly.

(c) Verify that the fractional error is,

$$\frac{\widehat{\Pr}\,[A_N]}{\Pr\,[A_N]} \propto e^{0.19N}$$

Observe that the fractional error grows exponentially with N.

(d) Estimate $\Pr\,[A_N]$ using the Chernoff bound of Problem 2.7.8. Denote this estimate as $\Pr_c\,[A_N]$. Compute $\dfrac{\Pr_c\,[A_N]}{\Pr\,[A_N]}$.

REFERENCES

[1] W. B. Davenport and W. L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.

[2] A. T. Bharucha-Reid, *Elements of the Theory of Markov Processes and their Applications*, McGraw-Hill, New York, 1960.

[3] M. Abramovitz and I. A. Stegun (ed.), *Handbook of Mathematical Functions*, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., June 1964.

[4] J. A. Greenwood and H. O. Hartley, *Guides to Tables in Mathematical Statistics*, Princeton University Press, Princeton, New Jersey, 1962.

[5] R. A. Fisher, "Theory of Statistical Estimation," *Proc. Cambridge Phil. Soc.*, **22**, 700 (1925).

[6] R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Phil. Trans. Roy. Soc., London*, **222**, 309 (1922).

[7] R. A. Fisher, "Two New Properties of Mathematical Likelihood," *Proc. Roy. Soc., London*, **144**, 285 (1934).

[8] R. A. Fisher, "The Logic of Inductive Inference," *J. Roy. Statist. Soc.*, **98**, 39 (1935).

[9] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New Jersey, 1946.

[10] S. S. Wilks, *Mathematical Statistics*, Wiley, New York, 1962.

[11] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 2, Inference and Relationship*, Hafner, New York, 1961.

[12] C. R. Rao, "Information and Accuracy Attainable in the Estimation of Statistical Parameters," *Bull. Calcutta Math. Soc.*, **37**, 81–91 (1945).

[13] A. Bhattacharyya, "On Some Analogues of the Amount of Information and their Use in Statistical Estimation," *Sankhya*, **8**, 1, 201, 315 (1946, 1947, 1948).

[14] H. L. Van Trees, "A Generalized Bhattacharyya Bound," Internal Memo., Detection & Estimation Theory Group, MIT, 1966.

[15] E. W. Barankin, "Locally Best Unbiased Estimates," *Ann. Math. Stat.*, **20**, 477 (1949).

[16] R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960.

[17] E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, 1959.

[18] F. B. Hildebrand, *Methods of Applied Mathematics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1952.

[19] R. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, Oliver & Boyd, Edinburgh, 1953.

[20] S. Sherman, "Non-mean-square Error Criteria," *IRE Trans. Information Theory*, **IT-4**, No. 3, 125–126 (1958).

[21] K. Pearson, *Tables of the Incomplete* Γ*-Function*, Cambridge University Press, Cambridge, 1934.

[22] J. N. Pierce, "Theoretical Diversity Improvement in Frequency-Shift Keying," *Proc. IRE*, **46**, 903–910 (May 1958).

[23] C. E. Shannon, "Seminar Notes for Seminar in Information Theory," MIT, 1956 (unpublished).

[24] R. M. Fano, *Transmission of Information*, MIT Press, Cambridge, Massachusetts, and Wiley, New York, 1961.

[25] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels: I," *Information and Control*, Vol. **10**, No. 1, 65–103 (1967).

[26] R. G. Gallager, "Lower Bounds on the Tails of Probability Distributions," MIT, RLE, QPR 77, 277–291 (April 1965).

[27] I. M. Jacobs, "Probability-of-Error Bounds for Binary Transmission on the Slow Fading Rician Channel," *IEEE Trans. Information Theory*, Vol. **IT-12**, No. 4, Oct, 1966.

[28] H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesi based on the Sum of Observations," *Annals Math. Stat.*, **23**, 493–507 (1962).

[29] A. Bhattacharyya, "On a Measure of Divergence Between Two Statistical Populations defined by their Probability Distributions," *Bull. Calcutta Math. Soc.*, **35**, No. 3, 99–110 (1943).

[30] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, Wiley, New York, 1950, 1957.

[31] D. Dugué, "Application des Proprietes de la Limite au Sens du Calcul des Probabilities a L'etude des Diverses Questions D'estimation," *Ecol. Poly.*, **3**, No. 4, 305–372 (1937).

[32] A. Sommerfeld, *An Introduction to the Geometry of N Dimensions*, Dutton, New York, 1929.

[33] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York, 1966.

[34] H. L. Yudkin, "An Error Bound for Gaussian Signals in Gaussian Noise," MIT, RLE, QPR 73, April 15, 1964.

[35] T. J. Goblick, "Study of An Orbiting Dipole Belt Communication System," Lincoln Laboratory, Technical Report 369, December 22, 1964.

[36] A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966.

[37] K. Pearson, "On the Systematic Fitting of Curves to Observations and Measurements," *Biometrika*, **1**, 265 (1902).

[38] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *J. Basic Eng.* (ASME Trans.) **82D**, 35–45 (1960).