Detection, Estimation, and Modulation Theory, Part III: Radar–Sonar Signal Processing and Gaussian Signals in Noise. Harry L. Van Trees Copyright © 2001 John Wiley & Sons, Inc. ISBNs: 0-471-10793-X (Paperback); 0-471-22109-0 (Electronic)

# 6

# Estimation of the Parameters of a Random Process

The next topic of interest is the estimation of the parameters of a Gaussian process. We study this problem in Chapters 6 and 7. Before developing a quantitative model of the problem, we discuss several physical situations in which parameter estimation problems arise.

The first example arises whenever we model a physical phenomenon using random processes. In many cases, the processes are characterized by a mean-value function, covariance function, or spectrum. We then analyze the model assuming that these functions are known. Frequently we must observe a sample function of the process and estimate the process characteristics from this observation. The measurement problems can be divided into two categories. In the first, we try to estimate an entire function, such as the power density spectrum of stationary processes. In the second, we parameterize the function and try to estimate the parameters; for example, we assume that the spectrum has the form

$$S(\omega) = \frac{P}{\omega^2 + k^2},\tag{1}$$

and try to estimate P and k. In many cases, this second category will fit into the parameter estimation model of this section. An adequate discussion of the first category would take us too far afield. Some of the issues are discussed in the problems. Books that discuss this problem include [1]-[3].

The second example arises in such areas as spectroscopy, radio astronomy, and passive sonar classification. The source generates a narrow-band random process whose center frequency characterizes the source. Thus, the first step in the classification problem is to estimate the center frequency of the signal process.

#### 168 6.1 Parameter Estimation Model

The third example arises in the underground nuclear blast detection problem. An important parameter in deciding whether the event was an earthquake or bomb is the depth of the source. At the station, we receive seismic waves whose angle of arrival depend on the depth of the source.

The common feature in all these examples is that in the parameters of interest are imbedded in the process characteristics. In other words, the mapping from the parameter to the signal is random. In this chapter and the next, we develop techniques for solving this type of problem.

In Chapter 6 we develop the basic results. The quantitative model of the problem is given in Section 6.1. In Section 6.2 we derive the likelihood function, the maximum likelihood equations, and the maximum aposteriori probability equations. In Section 6.3 we develop procedures for analyzing the performance.

In our study of detection theory we saw that there were special categories of problems for which we could obtain complete solutions. In Chapter 7 we study four such special categories of problems. In Section 7.1 we consider the stationary-process, long-observation-time case. The examples in this section deal with estimating the amplitude of a known covariance function. Several issues arise that cannot be adequately resolved without developing new techniques, and so we digress and develop the needed expressions. This section is important because it illustrates how to bridge the gap between the general theory of Chapter 6 and the complete solution to an actual problem. In Sections 7.2, 7.3, and 7.4 we consider processes with a finite state representation, separable kernel processes, and lowenergy-coherence problems, respectively. In Sections 7.5 and 7.6 we extend the results to include multiple parameter estimation and summarize the important results of our estimation theory discussion.

Two observations are useful before we begin our quantitative discussion.

1. The discussion is a logical extension of our parameter estimation work in Chapters I-2 and I-4. We strongly suggest that the reader review Section I-2.4 (pages 52–86), Sections I-4.2.2—I-4.2.4 (pages 271–287), and Section I-4.3.5 (pages 307–309) before beginning this section.

2. Parameter estimation problems frequently require a fair amount of calculation to get to the final result. The casual reader can skim over this detail but should be aware of the issues that are involved.

#### 6.1 PARAMETER ESTIMATION MODEL

The model of the parameter estimation problem can be described easily. The received waveform r(t) consists of the sum of signal waveform and a noise waveform,

$$r(t) = s(t, A) + w(t), \qquad T_i \le t \le T_f.$$
 (2)

The waveform s(t, A) is a sample function from a random process whose characteristics depend on the parameter A, which we want to estimate.

To emphasize the nature of the model, assume that A is fixed and w(t) is identically zero. Then, each time the experiment is conducted, the signal waveform s(t, A) will be different because it is a sample function of a random process. By contrast, in the parameter estimation problems of Chapter I-4, the mapping from the parameter to the signal waveform is deterministic.

We assume that the signal process is a conditionally Gaussian process.

**Definition.** A random process s(t, A) is conditionally Gaussian if, given any value of A is the allowable parameter range  $\chi_a$ , s(t, A) is a Gaussian process.

A conditionally Gaussian process is completely characterized by a conditional mean-value function

$$E[s(t, A) \mid A] \triangleq m(t, A), \qquad T_i \le t \le T_f \tag{3}$$

and a conditional covariance function

$$E[(s(t, A) - m(t, A))(s(u, A) - m(u, A)) \mid A] \triangleq K_s(t, u : A),$$
  
$$T_i \le t, u \le T_f. \quad (4)$$

The noise process is a zero-mean, white Gaussian noise process with spectral height  $N_0/2$  and is statistically independent of the signal process. Thus r(t) is also a conditionally Gaussian process,

$$E[r(t) \mid A] = E[s(t, A) \mid A] = m(t, A), \qquad T_i \le t \le T_f, \tag{5}$$

and

$$E\{[r(t) - m(t, A)][r(u) - m(u, A)] \mid A\} \triangleq K_r(t, u: A)$$

$$= K_{s}(t, u: A) + \frac{N_{0}}{2} \delta(t - u), \qquad T_{i} \le t, u \le T_{f}.$$
 (6)

Observe that any colored noise component in r(t) can be included in s(t, A). We assume that m(t, A),  $K_s(t, u: A)$ , and  $N_0/2$  are known.

The parameter A will be modeled in two different ways. In the first, we assume that A is a nonrandom parameter that lies in some range  $\chi_a$ , and we use maximum likelihood estimation procedures. In the second, we assume that A is the value of a random variable with a known probability

density  $p_a(A)$ . For random parameters we can use Bayes estimates with various cost functions. We shall confine our discussion to MAP estimates.

These assumptions specify our model of the parameter estimation problem. We now develop an estimation procedure.

#### 6.2 ESTIMATOR STRUCTURE

Our approach to the estimation problem is analogous to the one taken in Chapters I-2 and I-4. We first find the likelihood function  $\Lambda(A)$ . Then, if A is a nonrandom parameter and we want an ML estimate, we find the value of A for which  $\Lambda(A)$  is a maximum. If A is the value of a random variable and we desire an MAP estimate, we construct the function

$$f(A) \triangleq \ln \Lambda(A) + \ln p_a(A), \tag{7}$$

and find that value of A where it is a maximum. The only new issue is the actual construction of  $\Lambda(A)$  and the processing needed to find the maximum. In this section we address these issues.

#### 6.2.1 Derivation of the Likelihood Function

The derivation of the likelihood function is similar to that of the likelihood ratio in Chapter 2, and so we can proceed quickly. The first step is to find a series expansion for r(t). We then find the conditional probability density of the coefficients (given A) and use this to find an appropriate likelihood function. The procedure is simplified if we choose the coordinate system so that the coefficients are conditionally statistically independent. This means that we must choose a coordinate system that is conditionally dependent on A. The coefficients are

$$r_i(A) \stackrel{\Delta}{=} \int_{T_i}^{T_f} r(t)\phi_i(t;A) dt.$$
(8)

The  $r_i(A)$  are Gaussian random variables whose mean and variance are functions of A.

$$E[r_i(A) \mid A] = E\left\{\int_{T_i}^{T_f} r(t)\phi_i(t;A) dt \mid A\right\} = \int_{T_i}^{T_f} m(t,A)\phi_i(t;A) dt \triangleq m_i(A).$$
(9)

We choose the  $\phi_i(t; A)$  so that

$$E[(r_i(A) - m_i(A))(r_j(A) - m_j(A)) | A] = \lambda_i(A)\delta_{ij}.$$
 (10)

From our earlier work, we know that to achieve this conditional independence the  $\phi_i(t; A)$  must be the eigenfunctions of the integral equation

$$\lambda_i(A)\phi_i(t:A) = \int_{T_i}^{T_f} K_s(t, u:A)\phi_i(u:A) \, du, \qquad T_i \le t \le T_f. \tag{11}$$

Because the covariance function depends on the parameter A, the eigenfunctions, eigenvalues, or both will depend on A. If  $K_s(t, u: A)$  is positive definite, the eigenfunctions form a complete set. If  $K_s(t, u: A)$  is only non-negative-definite, we augment the set of eigenfunctions to make it complete.

Since the resulting set is complete, we can expand the mean-value function m(t, A) and the received waveform r(t) in a series expansion. These series are

$$m(t, A) = \sum_{i=1}^{\infty} m_i(A)\phi_i(t; A), \qquad T_i \le t \le T_f$$
(12)

and

$$r(t) = \lim_{K \to \infty} \sum_{i=1}^{K} [r_i(A) - m_i(A)]\phi_i(t, A) + m(t, A), \qquad T_i \le t \le T_f.$$
(13)

We denote the first K coefficients by the vector **R**. The probability density of **r** given the value of A is

$$p_{\mathbf{r}|a}(\mathbf{R} \mid A) = \left(\prod_{i=1}^{K} \frac{1}{\sqrt{2\pi} \left(N_0/2 + \lambda_i(A)\right)^{1/2}}\right) \exp\left[-\frac{1}{2} \sum_{i=1}^{K} \left(\frac{[R_i - m_i(A)]^2}{N_0/2 + \lambda_i(A)}\right)\right].$$
(14)

Just as in the known signal case (Section I-4.2.3), it is convenient to define a likelihood function  $\Lambda_K(A)$ , which is obtained from  $p_{r|a}(\mathbf{R} \mid A)$  by dividing by some function that does not depend on A (see page I-274). As before, we divide by

$$\prod_{i=1}^{K} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{1}{2} \frac{R_i^2}{N_0/2}\right).$$
(15)

Dividing (14) by (15), taking the logarithm of the result, and letting  $K \rightarrow \infty$ , we have

$$\ln \Lambda(A) = \frac{1}{N_0} \sum_{i=1}^{\infty} \frac{\lambda_i(A)}{\lambda_i(A) + N_0/2} R_i^2 + \sum_{i=1}^{\infty} \frac{1}{\lambda_i(A) + N_0/2} m_i(A) R_i - \frac{1}{2} \sum_{i=1}^{\infty} \ln \left( 1 + \frac{2\lambda_i(A)}{N_0} \right) - \frac{1}{2} \sum_{i=1}^{\infty} \frac{m_i^2(A)}{\lambda_i(A) + N_0/2}.$$
(16)

Comparing (16) with the limit of (2.19) as  $K \to \infty$  in our detection theory discussion, we see that there is a one-to-one correspondence. Thus, all of the closed-form expressions in the detection theory section will have obvious analogs in the estimation problem. By proceeding in a manner identical with that in Chapter 2, we can obtain four terms corresponding to those in (2.31)-(2.34).

The first term can be written as

$$l_{R}(A) = \frac{1}{N_{0}} \iint_{T_{i}}^{T_{f}} r(t)h(t, u:A)r(u) dt du,$$
(17)

where h(t, u: A) satisfies the integral equation

$$\frac{N_0}{2}h(t, u:A) + \int_{T_i}^{T_f} h(t, z:A) K_s(z, u:A) dz = K_s(t, u:A),$$
$$T_i \le t, u \le T_f. \quad (18)$$

We see that h(t, u: A) is the optimum unrealizable filter for the problem in which we observe

$$r(t) = s(t, A) + w(t), \qquad T_i \le t \le T_f,$$
(19)

and we want to make the MMSE error estimate of s(t, A) under the assumption that A is known. As in the detection problem, we shall frequently use the inverse kernel  $Q_r(t, u: A)$ , which can be written as

$$Q_r(t, u:A) = \frac{2}{N_0} [\delta(t-u) - h(t, u:A)], \qquad T_i < t, u < T_f.$$
(20)

The second term in (16) can be written as

$$l_D(A) = \iint_{T_i}^{T_f} r(t) Q_r(t, u; A) m(u, A) dt du.$$
(21)

Recall that the subscript D denotes deterministic and is used because  $l_D(A)$  is analogous to the receiver output in the known signal problem. Alternatively,

$$l_D(A) = \int_{T_i}^{T_f} r(t)g(t, A) \, dt,$$
(22)

where g(t, A) is defined as

$$g(t, A) = \int_{T_i}^{T_f} Q_r(t, u; A) m(u, A) \, du.$$
(23)

We can also specify g(t, A) implicitly by the equation

$$m(t, A) = \int_{T_i}^{T_f} K_r(t, u; A) g(u, A) \, du, \qquad T_i \le t \le T_f.$$
(24)

The function g(t, A) is familiar from the problem of estimating the parameters of a known signal in colored noise.

The remaining terms in (16) are the bias terms. The first is

$$l_B^{[1]}(A) \stackrel{\Delta}{=} -\frac{1}{2} \sum_{i=1}^{\infty} \ln\left(1 + \frac{2\lambda_i(A)}{N_0}\right) = -\frac{1}{N_0} \int_{T_i}^{T_f} \xi_P(t;A) \, dt, \quad (25)$$

where  $\xi_P(t: A)$  is the realizable mean-square filtering error for the filtering problem in (19). As in the detection case, we can also evaluate the second term in  $l_B(A)$  by means of the Fredholm determinant [see (2.74)]. The second bias term is

$$l_{B}^{[2]}(A) = -\frac{1}{2} \iint_{T_{i}}^{T_{f}} m(t, A)Q_{r}(t, u:A)m(u, A) dt du$$
  
$$= -\frac{1}{2} \int_{T_{i}}^{T_{f}} m(t, A)g(t, A) dt.$$
 (26)

Notice that the integral in  $l_B^{[2]}(A)$  is just  $d^2(A)$  for the problem of detecting a known signal m(t, A) in colored noise. The likelihood function is

$$\ln \Lambda(A) = l_R(A) + l_D(A) + l_B^{[1]}(A) + l_B^{[2]}(A),$$
(27)

where the component terms are defined in (17), (22), (25), and (26).

We can now use  $\ln \Lambda(A)$  to find  $\hat{a}_{map}(r(t))$  or  $\hat{a}_{ml}(r(t))$ . The procedure is conceptually straightforward. To find  $\hat{a}_{ml}$ , we construct  $\ln \Lambda(A)$  as a function of A and find the value of A where it is a maximum. To find  $\hat{a}_{map}$  we construct the function

$$f(A) \triangleq \ln \Lambda(A) + \ln p_a(A) = l_R(A) + l_D(A) + l_B(A) + \ln p_a(A)$$
(28)

and find the value of A where it is a maximum.

Even though the procedure is well defined, the actual implementation is difficult. A receiver structure analogous to that in the PFM problem (Fig. I-4.31) of Section I-4.2.3 is usually needed.

To illustrate this, we consider the case of the maximum likelihood estimation of a parameter A. We assume that it lies in the interval  $[A_{\alpha}, A_{\beta}]$ . In addition, we assume that the mean m(t, A) is zero. We divide the

parameter range into intervals of length  $\Delta$ . The center points of these intervals are

$$A_1 = A_{\alpha} + \frac{\Delta}{2},$$

$$A_2 = A_{\alpha} + \frac{3\Delta}{2},$$
(29)

and so forth. There are M intervals. We then construct  $\ln \Lambda(A_i)$ ,  $i = 1, \ldots, M$ , by using the parallel processing shown in Fig. 6.1. Several observations are worthwhile:

1. In general we have to solve a different integral equation to find the filter in each path. Thus the estimation problem has the same degree of complexity as an *M*-ary detection problem in the sense that we must build *M*-parallel processors.

2. The bias terms are usually functions of A and cannot be neglected.

3. In analyzing the performance, we must consider both global and local errors.

4. We have to consider the effect of the grid size  $\Delta$ . There is a trade-off between accuracy and complexity.

Before leaving our discussion of the estimator structure, we digress briefly and derive two alternative forms for  $l_R(A)$ . Repeating (17),

$$l_R(A) = \frac{1}{N_0} \iint_{T_i}^{T_f} r(t) h(t, u: A) r(u) \, dt \, du.$$
(17)

This corresponds to Canonical Realization No. 1 in the detection problem. To obtain Canonical Realization No. 3, we define  $h^{[1/2]}(t, u; A)$  implicitly,

$$h(t, u:A) = \int_{T_i}^{T_f} h^{[1/2]}(z, t:A) h^{[1/2]}(z, u:A) dz, \qquad T_i \le t, u \le T_f. \quad (30)$$

Then

$$l_R(A) = \frac{1}{N_0} \int_{T_i}^{T_f} dz \left[ \int_{T_i}^{T_f} h^{[1/2]}(z, t; A) r(t) dt \right]^2.$$
(31)

This can be implemented by a filter-squarer-integrator for any particular A.

To obtain Canonical Realization No. 4, we go through an argument parallel to that on pages 19–21. The result is

$$l_R(A) = \frac{1}{N_0} \int_{T_i}^{T_f} [2r(t)\hat{s}_r(t;A) - \hat{s}_r^2(t;A)] dt, \qquad (32)$$



Fig. 6.1 Generation of  $\ln \Lambda(A)$ .

where

$$\hat{s}_{r}(t:A) = \int_{T_{i}}^{t} h_{r}(t, u:A) r(u) \, du.$$
(33)

The filter  $h_r(t, u: A)$  satisfies the equation

$$\frac{N_0}{2}h_r(t, u:A) + \int_{T_i}^t h_r(t, z:A)K_s(z, u:A) dz = K_s(t, u:A),$$
$$T_i \le u \le t. \quad (34)$$

For the zero-mean case, the function  $\hat{s}_r(t:A)$  is the realizable MMSE estimate of s(t:A), assuming that A is given. We encounter examples of these realizations in subsequent sections. Many of the same issues that we encountered in the detection problem will also arise in the estimation problem.

Before considering some specific cases, we derive the maximum likelihood (ML) equations and the maximum a-posteriori probability (MAP) equations.

## 6.2.2 Maximum Likelihood and Maximum A-Posteriori Probability Equations

If the maximum of  $\ln \Lambda(A)$  is interior to  $\chi_a$  and  $\ln \Lambda(A)$  has a continuous first derivative, the ML equations specify a *necessary* condition on  $\hat{a}_{ml}$ .

The ML equation follows easily by differentiating (27) and setting the results equal to zero. Taking the partial derivative of  $\ln \Lambda(A)$  with respect to A, we have

$$\frac{\partial \ln \Lambda(A)}{\partial A} = \frac{\partial l_R(A)}{\partial A} + \frac{\partial l_D(A)}{\partial A} + \frac{\partial l_B(A)}{\partial A}.$$
(35)

To evaluate the first term, we differentiate the function in (17). The result is

$$\frac{\partial l_R(A)}{\partial A} = \frac{1}{N_0} \iint_{T_i}^{T_f} r(t) \frac{\partial h(t, u:A)}{\partial A} r(u) dt du$$
$$= -\frac{1}{2} \iint_{T_i}^{T_f} r(t) \frac{\partial Q_r(t, u:A)}{\partial A} r(u) dt du.$$
(36)

Notice that to find  $\partial h(t, u: A)/\partial A$  we must solve (17) as a function of A and then differentiate it. To evaluate the second term, we differentiate (21) and use (20) to obtain

$$\frac{\partial l_D(A)}{\partial A} = -\frac{2}{N_0} \iint_{T_i}^{T_f} r(t) \frac{\partial h(t, u:A)}{\partial A} m(u, A) dt du + \iint_{T_i}^{T_f} r(t) Q_r(t, u:A) \frac{\partial m(u, A)}{\partial A} dt du.$$
(37)

Finally, from (25) and (26),

$$\frac{\partial l_B(A)}{\partial A} = \frac{1}{N_0} \iint_{T_i}^{T_f} m(t, A) \frac{\partial h(t, u:A)}{\partial A} m(u, A) dt du$$
$$- \iint_{T_i}^{T_f} m(t, A) Q_r(t, u:A) \frac{\partial m(u, A)}{\partial A} dt du$$
$$- \frac{1}{N_0} \int_{T_i}^{T_f} \frac{\partial \xi_{Ps}(t, A)}{\partial A} dt.$$
(38)

Two alternative forms of the last term in (38) are

$$-\frac{1}{N_0} \int_{T_i}^{T_f} \frac{\partial \xi_{Ps}(t,A)}{\partial A} dt = \frac{1}{2} \iint_{T_i}^{T_f} K_r(t,u:A) \frac{\partial Q_r(t,u:A)}{\partial A} dt du$$
$$= -\frac{1}{2} \iint_{T_i}^{T_f} \frac{\partial K_r(t,u:A)}{\partial A} Q_r(t,u:A) dt du \quad (39)$$

(see Problem 6.2.1). Collecting terms, we have

$$\frac{\partial \ln \Lambda(A)}{\partial A} = \frac{1}{2} \iint_{T_i}^{T_f} K_r(t, u; A) \frac{\partial Q_r(t, u; A)}{\partial A} dt du + \iint_{T_i}^{T_f} \frac{\partial m(t, A)}{\partial A} Q_r(t, u; A) [r(u) - m(u, A)] dt du - \frac{1}{2} \iint_{T_i}^{T_f} [r(t) - m(t, A)] \frac{\partial Q_r(t, u; A)}{\partial A} \times [r(u) - m(u, A)] dt du.$$
(40)

If we assume that the derivative exists at the maximum of  $\ln \Lambda(A)$  and that the maximum is interior to the range, then a necessary condition on the maximum likelihood estimate is obtained by equating the right side of (40) to zero. To find the MAP equation, we add  $(\partial \ln p_a(A)/\partial A)$  to (40) and equate the result to zero.

The likelihood equation obtained from (40) is usually difficult to solve. The reason is that even if the parameter appears linearly in the signal covariance function, it may not appear linearly in the inverse kernel. Thus, the necessary condition is somewhat less useful in the random signal case than it is in the known signal case.

#### 6.3 PERFORMANCE ANALYSIS

The performance analysis is similar to that in the nonlinear estimation problems in Chapters I-2 and I-4. We divide the errors into local and global errors. The variance of the local errors can be obtained from a power series approach or by a generalized Cramér-Rao bound. The global behavior can be analyzed by an extension of the analysis on pages I-279– I-284. In this section we derive the generalized Cramér-Rao bound and discuss methods of calculating it.

#### 6.3.1 A Lower Bound on the Variance

We assume that A is a nonrandom variable that we want to estimate. We desire a lower bound on the variance of any unbiased estimate of A.

#### 178 6.3 Performance Analysis

The derivation on pages I-66–I-68 extends easily to this case. The result is that for any unbiased estimate  $\hat{a}(r(t))$  of the nonrandom variable A, the variance satisfies the inequality

$$\operatorname{Var}\left[\hat{a}(r(t)) - A\right] \ge -\left\{ E\left[\frac{\partial^2 \ln \Lambda(A)}{\partial A^2}\right] \right\}^{-1}$$
(41)

with equality if and only if

$$\frac{\partial \ln \Lambda(A)}{\partial A} = \{\hat{a}(r(t)) - A\}k(A).$$
(42)

To evaluate the bound we differentiate (40). The result is

$$\frac{\partial^2 \ln \Lambda(A)}{\partial A^2} = \frac{1}{2} \iint_{T_i}^{T_f} K_r(t, u:A) \frac{\partial^2 Q_r(t, u:A)}{\partial A^2} dt du + \frac{1}{2} \iint_{T_i}^{T_f} \frac{\partial K_r(t, u:A)}{\partial A} \cdot \frac{\partial Q_r(t, u:A)}{\partial A} dt du - \iint_{T_i}^{T_f} \frac{\partial m(t, A)}{\partial A} Q_r(t, u:A) \frac{\partial m(u, A)}{\partial A} dt du - \frac{1}{2} \iint_{T_i}^{T_f} [r(t) - m(t, A)] \frac{\partial^2 Q_r(t, u:A)}{\partial A^2} [r(u) - m(u, A)] dt du + (terms whose expectations are zero).$$
(43)

When we take the expectation of the last integral, we find that it cancels the first term in (43). Thus any *unbiased* estimate of A will have a variance satisfying the bound

$$\operatorname{Var}\left[\hat{a}(r(t)) - A\right] \geq \left( \iint_{T_{i}}^{T_{f}} \frac{\partial m(t, A)}{\partial A} Q_{r}(t, u:A) \frac{\partial m(u, A)}{\partial A} dt du - \frac{1}{2} \iint_{T_{i}}^{T_{f}} \frac{\partial K_{r}(t, u:A)}{\partial A} \frac{\partial Q_{r}(t, u:A)}{\partial A} dt du \right)^{-1}$$
(44)

For notational convenience in the subsequent discussion, we denote the first term in the braces by  $J^{(1)}(A)$ , the second by  $J^{(2)}(A)$ , and the sum by J(A).

$$J^{(1)}(A) \stackrel{\Delta}{=} \iint_{T_i}^{T_f} \frac{\partial m(t, A)}{\partial A} Q_r(t, u; A) \frac{\partial m(u, A)}{\partial A} dt du$$
(45)

and

$$J^{(2)}(A) \triangleq -\frac{1}{2} \iint_{T_i}^{T_f} dt \, du \, \frac{\partial K_r(t, u:A)}{\partial A} \frac{\partial Q_r(t, u:A)}{\partial A} \,.$$
(46)

Several observations are useful:

1. The terms in the bound depend on A. Thus, as we have seen before, the variance depends on the actual value of the nonrandom parameter.

2. The bound assumes that the estimate is unbiased. If the estimate is biased, a different bound must be used. (See Problem 6.3.1.)

3. The first term is familiar in the context of detection of known signals in colored noise. Specifically, it is exactly the value of  $d^2$  for the simple binary detection problem in which we transmit  $\partial m(t, A)/\partial A$  and the additive colored noise has a covariance function  $K_r(t, u: A)$ . Thus, the techniques we have developed for evaluating  $d^2$  are applicable here.

We now consider efficient procedures for evaluating  $J^{(2)}(A)$ .

#### 6.3.2 Calculation of $J^{(2)}(A)^{\dagger}$

The  $J^{(2)}(A)$  term arises because the covariance function of the process depends on A. It is a term we have not encountered previously, and so we develop two convenient procedures for evaluating it. The first technique relates it to the Bhattacharyya distance (recall the discussion on pages 71-72), and the second expresses it in terms of eigenfunctions and eigenvalues.

The techniques developed in this section are applicable to arbitrary observation intervals and processes that are not necessarily stationary. In Section 7.1, we shall consider the stationary-process, long-observation-time case and develop a simple expression for  $J^{(2)}(A)$ .

† This section may be omitted on the first reading

#### 180 6.3 Performance Analysis

**Relation to Bhattacharyya Distance.** In this section we relate J(A) to the Bhattacharyya distance. We first work with  $r_K(t)$  and the vector **R** and then let  $K \to \infty$  in the final answer. We define a function

$$\mu(\frac{1}{2}, A_1, A) \stackrel{\Delta}{=} \ln \int_{-\infty}^{\infty} p_{\mathbf{r}|a}^{\frac{1}{2}}(\mathbf{R}|A_1) p_{\mathbf{r}|a}^{\frac{1}{2}}(\mathbf{R}|A) d\mathbf{R}.$$
(47)

This is simply  $\mu(\frac{1}{2})$  for the general binary detection problem in which

$$p_{\mathbf{r}|H_1}(\mathbf{R} \mid H_1) = p_{\mathbf{r}|a}(\mathbf{R} \mid A_1)$$
(48)

and

$$p_{\mathbf{r}|H_0}(\mathbf{R} \mid H_0) = p_{\mathbf{r}|a}(\mathbf{R} \mid A).$$
(49)

The Bhattacharyya distance is just

$$B(A_1, A) = -\mu(\frac{1}{2}, A_1, A).$$
(50)

Using (50) and (47) leads to

$$e^{-B(A_1,A)} = \int_{-\infty}^{\infty} p_{\mathbf{r}|a}^{\frac{1}{2}}(\mathbf{R}|A_1) p_{\mathbf{r}|a}^{\frac{1}{2}}(\mathbf{R}|A) d\mathbf{R}.$$
 (51)

We are interested in the case in which

$$\Delta A \stackrel{\Delta}{=} A_1 - A \tag{52}$$

is small, and so we expand both sides of (51) in a series. Expanding the left side in a Taylor series in  $A_1$  about the point  $A_1 = A$  gives

$$e^{-B(A_{1},A)} = e^{-B(A,A)} - \left[\frac{\partial B(A_{1},A)}{\partial A_{1}}e^{-B(A_{1},A)}\right]_{A_{1}=A} \Delta A + \frac{1}{2} \left[\left(\frac{\partial^{2}B(A_{1},A)}{\partial A_{1}^{2}} - \left(\frac{\partial B(A_{1},A)}{\partial A_{1}}\right)^{2}\right)e^{-B(A_{1},A)}\right]_{A_{1}=A} (\Delta A)^{2} + \cdots \right]$$
(53)

From (47), it follows easily that

$$\frac{\partial B(A_1, A)}{\partial A_1}\Big|_{A_1=A} = 0$$
(54)

(see Problem 6.3.2). Thus, (53) reduces to

$$e^{-B(A_1,A)} = 1 + \left(\frac{\partial^2 B(A_1,A)}{\partial A_1^2}\Big|_{A_1=A}\right) \frac{(\Delta A)^2}{2} + \cdots$$
 (55)

To expand the right side of (51), we use a Taylor series for the first term in the integrand and then integrate term by term. The result is

$$\int_{-\infty}^{\infty} p_{\mathbf{r}\mid a}^{\frac{1}{2}}(\mathbf{R}\mid A_1) p_{\mathbf{r}\mid a}^{\frac{1}{2}}(\mathbf{R}\mid A) dR \simeq 1 - \frac{(\Delta A)^2}{8} \int_{-\infty}^{\infty} \frac{\left(\frac{\partial p_{\mathbf{r}\mid a}(\mathbf{R}\mid A)}{p_{\mathbf{r}\mid a}(\mathbf{R}\mid A)}\right)^2}{p_{\mathbf{r}\mid a}(\mathbf{R}\mid A)} d\mathbf{R}.$$
(56)

The integral on the right side of (56) can be written as

$$\int_{-\infty}^{\infty} \left(\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R} \mid A)}{\partial A}\right)^{2} p_{\mathbf{r}|a}(\mathbf{R} \mid A) d\mathbf{R} = E \left\{\frac{\partial^{2} \ln p_{\mathbf{r}|a}(\mathbf{R} \mid A)}{\partial A^{2}}\right\}.$$
 (57)

The term on the right side of (57) is just the negative of J(A). Substituting (55) and (56) into (51) and equating the coefficients of  $(\Delta A)^2$ , we have

$$J(A) = 4\left(\frac{\partial^2 B(A_1, A)}{\partial^2 A_1}\Big|_{A_1=A}\right).$$
(58)

Notice that the expression in (58) includes both  $J^{(1)}(A)$  and  $J^{(2)}(A)$ . To calculate  $J^{(2)}(A)$ , we assume that the process is zero-mean and use the formula for  $\mu(s)$  given in (3.60),

$$B(A_{1}, A) = -\frac{1}{N_{0}} \int_{T_{i}}^{T_{f}} dt \left\{ \frac{1}{2} \xi_{P} \left( t \mid s(\cdot, A_{1}), \frac{N_{0}}{2} \right) + \frac{1}{2} \xi_{P} \left( t \mid s(\cdot, A), \frac{N_{0}}{2} \right) - \xi_{P} \left( t \mid \sqrt{\frac{1}{2}} s(\cdot, A_{1}) + \sqrt{\frac{1}{2}} s(\cdot, A), \frac{N_{0}}{2} \right) \right\}.$$
 (59)

In the last term we have a composite process of the type discussed in (3.63). We emphasize that s(t, A) and  $s(t, A_1)$  are statistically independent components in this composite process. Differentiating twice and substituting into (58) gives the desired result.

$$J^{(2)}(A) = \frac{4}{N_0} \left\{ \frac{\partial^2}{\partial A_1^2} \int_{T_i}^{T_f} \xi_P \left( t \mid \sqrt{\frac{1}{2}} \, s(\cdot, A_1) + \sqrt{\frac{1}{2}} \, s(\cdot, A), \frac{N_0}{2} \right) dt - \frac{1}{2} \frac{\partial^2}{\partial A_1^2} \int_{T_i}^{T_f} \xi_P \left( t \mid s(\cdot, A_1), \frac{N_0}{2} \right) dt \right\}_{A_1 = A}.$$
 (60)

It is worthwhile pointing out that in many cases it will be easier to evaluate  $J^{(2)}(A)$  by using the Fredholm determinant (e.g., Section 2.2.3).

**Eigenvalue Approach.** In this section we derive an expression for  $J^{(2)}(A)$  in terms of the eigenvalues and eigenfunctions of  $K_r(t, u: A)$ . From (20) it is clear that we could also write  $J^{(2)}(A)$  as

$$J^{(2)}(A) = \frac{1}{N_0} \iint_{T_i}^{T_f} dt \, du \, \frac{\partial K_r(t, u:A)}{\partial A} \frac{\partial h(t, u:A)}{\partial A} \,. \tag{61}$$

This expression still requires finding h(t, u : A), the optimum unrealizable filter for all t and u in  $[T_i, T_f]$ . In order to express  $J^{(2)}(A)$  in terms of eigenvalues and eigenfunctions, we first write h(t, u : A) as the series

$$h(t, u:A) = \sum_{i=1}^{\infty} \left( \frac{\lambda_i(A)}{N_0/2 + \lambda_i(A)} \right) \phi_i(t:A) \phi_i(u:A).$$
(62)

#### 182 6.3 Performance Analysis

Differentiating  $K_r(t, u : A) h(t, u : A)$  and using the results in (62), we obtain

$$J^{(2)}(A) = \frac{1}{2} \sum_{i=1}^{\infty} \left( \frac{[\partial \lambda_i(A)]/\partial A}{\lambda_i(A) + N_0/2} \right)^2 + \frac{2}{N_0} \sum_{i=1}^{\infty} \left( \frac{\lambda_i^2(A)b_i(A)}{\lambda_i(A) + N_0/2} \right) \\ - \frac{2}{N_0} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{\lambda_i(A)\lambda_j(A)}{\lambda_j(A) + N_0/2} a_{ij}(A), \quad (63)$$

where

$$b_i(A) \stackrel{\Delta}{=} \int_{T_i}^{T_f} \left(\frac{\partial \phi_i(t, A)}{\partial A}\right)^2 dt \tag{64}$$

and

$$a_{ij}(A) \stackrel{\Delta}{=} \int_{T_i}^{T_f} \frac{\partial \phi_i(t:A)}{\partial A} \phi_j(t:A) \, dt.$$
(65)

The expression in (63) is not particularly useful in the most cases. A special case of interest in which it is useful is the one in which the eigenfunctions do not depend on A. A common example of this case is when A corresponds to the amplitude of the covariance function. Then the last two terms in (63) are zero and

$$J^{(2)}(A) = \frac{1}{2} \sum_{i=1}^{\infty} \left( \frac{[\partial \lambda_i(A)]/\partial A}{\lambda_i(A) + N_0/2} \right)^2.$$
 (66)

The form in (66) is reasonably easy to evaluate in many problems.

A simple example illustrates the use of (66).

Example. The received waveform is

$$r(t) = s(t, A) + w(t), \qquad 0 \le t \le T.$$
(67)

The signal is a sample function of a Wiener process. It is a Gaussian process with statistics

$$E[s(t, A)] = 0, \quad t \ge 0$$
 (68)

and

$$s(0, A) = 0,$$
 (69)

$$K_s(t, u: A) = A \min(t, u), \quad 0 \le t, u.$$
 (70)

This process was first introduced on page I-195. The additive noise w(t) is a sample function from a statistically independent white Gaussian process with spectral height  $N_0/2$ . We want to estimate the nonrandom parameter A.

In Problem 7.2.1, we shall derive the optimum receiver for this problem. In the present example, we simply evaluate the expression in (66). From page I-196, the eigenvalues are

$$\lambda_i(A) = \frac{AT^2}{(i-\frac{1}{2})^2 \pi^2}, \qquad i = 1, 2, \dots,$$
(71)

and the eigenfunctions do not depend on A. Differentiating (71) gives

$$\frac{\partial \lambda_i(A)}{\partial A} = \frac{T^2}{(i-\frac{1}{2})^2 \pi^2} \,. \tag{72}$$

Using (71) and (72) in (66) gives

$$J(A) = J^{(2)}(A) = \frac{1}{2A^2} \sum_{i=1}^{\infty} \left( \frac{1}{1 + [N_0/2AT^2](i-1)^2 \pi^2} \right)^2.$$
(73)

The bound on the normalized variance of any unbiased estimate is

$$\frac{\operatorname{Var}\left[\hat{a} - A\right]}{A^2} \ge \frac{2}{\sum_{i=1}^{\infty} \left[(1 + \left[N_0/2AT^2\right](i-1)^2\pi^2\right)^2\right]^{-1}}.$$
(74)

The sum can be expressed in terms of polygamma functions whose values are tabulated (e.g., [4, page 265]). In Chapter 7 we shall see that for large values of  $2AT^2/N_0$ , the ML estimate is essentially unbiased and its variance approaches this bound. For small values of  $2AT^2/N_0$ , the bias is an important issue. We discuss the bias problem in detail in Section 7.1.

The final topic of interest is the performance when we estimate a random variable. In the next section we derive a lower bound on the minimum mean-square error.

#### 6.3.3 Lower Bound on the Mean-Square Error

To derive the bound on the mean-square error we go through a similar procedure (e.g., page I-72). Since the derivation is straightforward, we leave it as an exercise. The result is

$$E[(\hat{a}(R) - a)^2] \ge \left\{ E_a[J^{(1)}(A)] + E_a[J^{(2)}(A)] - E_a\left[\frac{\partial^2 \ln p_a(A)}{\partial A^2}\right] \right\}^{-1}.$$
 (75)

The expressions for  $J^{(1)}(A)$  and  $J^{(2)}(A)$  are given in (45) and (46). This bound holds under weak conditions analogous to those given on page I-72. Two observations are useful:

1. Since a is a random variable, there is no issue of bias. The bound is on the mean-square error, *not* the variance.

2. There is an expectation over  $p_a(A)$  in each term on the right side of (75). Thus the bound is not a function of the actual value of A. In most cases it is difficult to perform this integration over A.

Most of our examples in the text will deal with nonrandom variables. The extension of any particular example to the random-variable case is straightforward.

## 6.3.4 Improved Performance Bounds

Our discussion of performance has concentrated on generalizations of the Cramer-Rao bounds. In many problems when the processes are stationary, one can show that the variance of the ML estimate approaches the bound as the observation time increases (e.g., [5]). On the other hand, as we have seen before, there are a number of problems in which the bound does not give an accurate indication of the actual performance.

One procedure for obtaining a better estimate is suggested by the structure in Fig. 6.1. We consider the problem as an M-ary detection problem, find the error probability, and translate this into a global estimation error. This technique was introduced for the problem of estimating deterministic signal parameters by Woodward [6] and Kotel-nikov [7]. It was subsequently modified and extended [8]–[13]. We discussed the approach on pages I-278–I-284. The extension to the random signal parameter case is conceptually straightforward but usually difficult to carry out. In Problem 7.1.23, we go through the procedure for a particular estimation problem.

A second procedure for evaluating the performance is to use the Barankin bound [14]. This technique has been applied to the deterministic signal parameter problem [15]–[17]. Some progress has been made in the random signal problem by Baggeroer [18]. Once again, the basic ideas are straightforward but the actual calculations are difficult.

In Chapter 7, we study some particular estimation problems. At that point, we consider the performance question again in more detail. We may now summarize the results of this chapter.

#### 6.4 SUMMARY

In this chapter we have developed the basic results needed to study the parameter estimation problem. The formal derivation of the likelihood function was a straightforward extension of our earlier detection results. The resulting likelihood function is

$$\ln \Lambda(A) = \frac{1}{N_0} \iint_{T_i}^{T_f} r(t)h(t, u:A)r(u) \, dt \, du + \int_{T_i}^{T_f} r(t)g(t, A) \, dt$$
$$- \frac{1}{N_0} \int_{T_i}^{T_f} \xi_P(t:A) \, dt - \frac{1}{2} \int_{T_i}^{T_f} m(t, A)g(t, A) \, dt, \quad (76)$$

where the various functions are defined in (18), (23), and (25). To find  $\hat{a}_{ml}$  we plot  $\ln \Lambda(A)$  as a function of A and find the point where it is a maximum.

The next step was to find the performance of the estimator. A lower bound on the variance of any unbiased estimate was given in (44).

At this point in our discussion we have derived several general results. The next, and more important, step is to see how we can use these results actually to solve a particular estimation problem. We study this question in detail in Chapter 7.

#### 6.5 PROBLEMS

This problem section is brief because of the introductory nature of the chapter. Section 7.7 contains a number of interesting estimation problems.

#### P.6.2 Estimator Structure

**Problem 6.2.1.** Verify the result in (39). (*Hint*: use the original definition of  $Q_r(t, u: A)$  and an eigenfunction expansion of the various terms.)

Problem 6.2.2. Consider the vector version of the model in (2). The received waveform is

$$\mathbf{r}(t) = \mathbf{s}(t, A) + \mathbf{w}(t), \qquad T_i \le t \le T_f.$$

The signal process s(t, A) is a vector, conditionally Gaussian process with conditional mean-value function  $\mathbf{m}(t, A)$  and conditional covariance function matrix  $\mathbf{K}_{\mathbf{r}}(t, u: A)$ . The additive white Gaussian noise has a spectral matrix  $(N_0/2)\mathbf{I}$ .

1. Find an expression for  $\ln \Lambda(A)$ .

2. Find an expression for  $l_R(A)$  in terms of Canonical Realizations No. 1, 3, 4, and 4S.

3. Derive the vector version of the bound in (44).

**Problem 6.2.3.** In Section 6.1, we indicated that if a colored noise component was present it could be included in s(t, A). In this problem we indicate the colored noise explicitly as

$$r(t) = s_s(t, A) + n_c(t) + w(t), \qquad T_i \le t \le T_f.$$

The processes are zero-mean Gaussian processes with covariance functions  $K_{ss}(t, u)$ ,  $K_c(t, u)$ , and  $(N_0/2)\delta(t - u)$ , respectively.

1. Modify (16), (17), and (25) to include the effect of the colored noise explicitly.

2. Can any of the above expressions be simplified because of the explicit inclusion of the white noise?

**Problem 6.2.4.** The model in Problem 6.2.3 is analogous to a class  $B_w$  detection problem. Consider the model

$$r(t) = s(t, A) + n_c(t), \qquad T_i \le t \le T_f,$$

where  $n_c(t)$  does not contain a white component.

1. Derive an expression for  $\ln \Lambda(A)$ .

2. Derive a lower bound on the variance of any unbiased estimate analogous to (44). (*Hint*: Review Section 3.5.)

Problem 6.2.5. Assume that

$$r(t) = s(t, A) + w(t), \qquad T_i \le t \le T_f,$$

186 References

with probability p, and that

$$r(t) = w(t), \qquad T_i \le t \le T_f,$$

with probability (1 - p).

- 1. Derive an expression for  $\ln \Lambda(A)$ .
- 2. Check your answer for the degenerate cases when p = 0 and p = 1.

#### P.6.3 Performance

Problem 6.3.1. Assume that

$$E[(\hat{a} - A)] = B(A).$$
 (P.1)

Derive a lower bound on the variance of any estimate satisfying (P.1).

**Problem 6.3.2.** Use the definition of  $B(A_1, A)$  in (47) and (50) to verify that (54) is valid.

Problem 6.3.3. Carry out the details of the derivation of (75).

#### REFERENCES

- G. M. Jenkins and D. G. Watts, Spectral Analysis and Its Applications, Holden-Day, San Francisco, 1968.
- [2] R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra from the Point of View of Communications Engineering, Dover Press, New York, 1958.
- [3] J. S. Bendat and A. G. Piersol, Measurement and Analysis of Random Data, Wiley, New York, 1966.
- [4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Press, New York, 1965.
- [5] U. Grenander, "Stochastic Processes and Statistical Inference," Arkiv Matematik 1, 195–277 (April 1950).
- [6] P. M. Woodward, Probability and Information Theory, with Application to Radar, Pergamon Press, London, 1953.
- [7] V. A. Kotelnikov, The Theory of Optimum Noise Immunity, McGraw-Hill, New York, 1959.
- [8] S. Darlington, "Demodulation of Wideband, Low Power FM Signals, Bell Syst. Tech. J. 43, No. 1, Pt. 2, 339–374 (1964).
- [9] H. Akima, "Theoretical Studies on Signal-to-Noise Characteristics of an FM System," IEEE Trans. Space Electronics Tel. SET-9, 101-108 (1963).
- [10] J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering, Wiley, New York, 1965.
- [11] L. A. Wainstein and V. D. Zubakov, Extraction of Signals from Noise, Prentice-Hall, Englewood Cliffs, N.J., 1962.
- [12] J. Ziv and M. Zakai, "Some Lower Bounds on Signal Parameter Estimation," IEEE Trans. Information Theory IT-15, No. 3, 386-391 (May 1969).
- [13] M. Zakai and J. Ziv, "On the Threshold Effect in Radar Range Estimation," IEEE Trans. Information Theory IT-15, No. 1, 167–170 (Jan. 1969).
- [14] E. W. Barankin, "Locally Best Unbiased Estimates," Ann. Math. Statist. 20, 477-501 (1949).
- [15] P. Swerling, "Parameter Estimation for Waveforms in Additive Gaussian Noise," J. SIAM 7, 154–166 (1959).

- [16] P. Swerling, "Parameter Estimation Accuracy Formulas," IEEE Trans. Information Theory IT-10, 302–313 (Oct. 1964).
- [17] R. J. McAulay and L. P. Seidman, "A Useful Form of the Barankin Lower Bound and Its Application to PPM Threshold Analysis," IEEE Trans. Information Theory IT-15, No. 2, 273–279 (March 1969).
- [18] A. B. Baggeroer, "Barankin Bound on the Variance of Estimates of the Parameters of a Gaussian Random Process," Massachusetts Institute of Technology, RLE Quart. Prog. Report No. 92, 324–333, January 15, 1969.