Chapter 6

Parameter Estimation

6.1 INTRODUCTION

In Chapter 5, we considered the problem of detection theory, where the receiver receives a noisy version of a signal and decides which hypothesis is true among the $M$ possible hypotheses. In the binary case, the receiver had to decide between the null hypothesis $H_0$ and the alternate hypothesis $H_1$.

In this chapter, we assume that the receiver has made a decision in favor of the true hypothesis, but some parameter associated with the signal may not be known. The goal is to estimate those parameters in an optimum fashion based on a finite number of samples of the signal.

Let $Y_1, Y_2, \ldots, Y_K$ be $K$ independent and identically distributed samples of a random variable $Y$, with some density function depending on an unknown parameter $\theta$. Let $y_1, y_2, \ldots, y_K$ be the corresponding values of samples $Y_1, Y_2, \ldots, Y_K$ and $g(Y_1, Y_2, \ldots, Y_K)$, a function (a statistic) of the samples used to estimate the parameter $\theta$. We call $\hat{\theta}(Y_1, Y_2, \ldots, Y_K) = \theta$ (6.1)

the estimator of $\theta$. The value that the statistic assumes is called the estimate of $\theta$ and is equal to $\hat{\theta} = g(y_1, y_2, \ldots, y_K)$. In order to avoid any confusion between a random variable and its value, it should be noted that $\hat{\theta}$, the estimate of $\theta$, is actually $g(Y_1, Y_2, \ldots, Y_K)$. Consequently, when we speak of the mean of $\hat{\theta}$, $E[\hat{\theta}]$, we are actually referring to $E[g(Y_1, Y_2, \ldots, Y_K)]$.

The parameter to be estimated may be random or nonrandom. The estimation of random parameters is known as the Bayes’ estimation, while the estimation of nonrandom parameters is referred to as the maximum likelihood estimation (MLE).
In Section 6.2, we present the maximum likelihood estimator, then we use this estimator to compute the likelihood ratio test. This is called the generalized likelihood ratio test. In Section 6.4, we present the criteria for a “good” estimator. When the parameter to be estimated is a random variable, we use the Bayes’ estimation. Specifically, we study the minimum mean-square estimation, the minimum mean absolute value of error estimation, and the maximum a posteriori estimation. The Cramer-Rao lower bound on the estimator is presented in Section 6.6. Then, we generalize the above concepts to multiple parameter estimation. Based on the fact that sometimes it is not possible to determine the optimum mean-square estimate, even if it exists, we present the best linear unbiased estimator, which is a suboptimum solution, and discuss the conditions under which it becomes optimum. In Section 6.9, we present the least-square estimation, which is different than the above-mentioned methods, in the sense that it is not based on an unbiased estimator with minimum variance, but rather on minimizing the squared difference between the observed data and the signal data. We conclude the chapter with a brief section on recursive least-square estimation for real-time applications.

6.2 MAXIMUM LIKELIHOOD ESTIMATION

As mentioned in the previous function, the procedure commonly used to estimate nonrandom parameters is the maximum likelihood (ML) estimation. Let \( Y_1, Y_2, \ldots, Y_K \) be \( K \) observations of the random variable \( Y \), with sample values \( y_1, y_2, \ldots, y_K \). These random variables are independent and identically distributed. Let \( f_{Y|\theta}(y|\theta) \) denote the conditional density function of the random variable \( Y \). Note that the density function of \( Y \) depends on the parameter \( \theta \), \( \Theta \subseteq \theta \), which needs to be estimated. The likelihood function, \( L(\theta) \), is

\[
L(\theta) = f_{Y_1,\ldots,Y_K|\theta}(y_1, y_2, \ldots, y_K | \theta) = f_{Y|\theta}(y | \theta) = \prod_{k=1}^{K} f_{Y_k|\theta}(y_k | \theta) \tag{6.2}
\]

The value \( \hat{\theta} \) that maximizes the likelihood function is called the maximum likelihood estimator of \( \theta \). In order to maximize the likelihood function, standard techniques of calculus may be used. Because the logarithmic function \( \ln x \) is a monotonically increasing function of \( x \), as was shown in Chapter 5, maximizing \( L(\theta) \) is equivalent to maximizing \( \ln L(\theta) \). Hence, it can be shown that a necessary but not sufficient condition to obtain the ML estimate \( \hat{\theta} \) is to solve the likelihood equation.

\[
\frac{\partial}{\partial \theta} \ln f_{Y|\theta}(y | \theta) = 0 \tag{6.3}
\]
**Invariance Property.** Let $L(\theta)$ be the likelihood function of $\theta$ and $g(\theta)$ be a one-to-one function of $\theta$; that is, if $g(\theta_1) = g(\theta_2) \Leftrightarrow \theta_1 = \theta_2$. If $\hat{\theta}$ is an MLE of $\theta$, then $g(\hat{\theta})$ is an MLE of $g(\theta)$.

**Example 6.1**

In Example 5.2, the received signal under hypotheses $H_1$ and $H_0$ was

$$
H_1 : Y_k = m + N_k, \; k = 1, 2, ..., K \\
H_0 : Y_k = N_k, \; k = 1, 2, ..., K
$$

(a) Assuming the constant $m$ is not known, obtain the ML estimate $\hat{m}_{ml}$ of the mean.

(b) Suppose now that the mean $m$ is known, but the variance $\sigma^2$ is unknown. Obtain the MLE of $\theta = \sigma^2$.

**Solution**

Detection theory (Chapter 5) was used to determine which of the two hypotheses was true. In this chapter of estimation theory, we assume that $H_1$ is true. However, a parameter is not known and needs to be estimated using MLE.

(a) The parameter $\hat{\theta}$ to be determined in this example is $\hat{m}_{ml}$, where the mean $m \in M$. Since the samples are independent and identically distributed, the likelihood function, using (6.2), is

$$
f_{Y|M}(y|m) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y_k - m)^2}{2\sigma^2}\right] = \frac{1}{(2\pi)^{K/2} \sigma^K} \exp\left[-\frac{\sum_{k=1}^{K} (y_k - m)^2}{2\sigma^2}\right]
$$

Taking the logarithm on both sides, we obtain

$$
\ln f_{Y|M}(y|m) = \ln\left[\frac{1}{(2\pi)^{K/2} \sigma^K}\right] - \frac{\sum_{k=1}^{K} (y_k - m)^2}{2\sigma^2}
$$

The ML estimate is obtained by solving the likelihood equation, as shown in (6.3). Hence,
or \( m = (1/k) \sum_{k=1}^{K} y_k \). Thus, the ML estimator is \( \hat{m}_{ml} = (1/k) \sum_{k=1}^{K} y_k \).

(b) The likelihood function is

\[
L(\sigma^2) = \frac{1}{(2\pi)^{\frac{K}{2}} \sigma^K} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - m)^2\right]
\]

Taking the logarithm, we obtain

\[
\ln L(\sigma^2) = -\frac{K}{2} \ln 2\pi - K \ln \sigma - \frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - m)^2
\]

Observe that maximizing \( \ln L(\sigma^2) \) with respect to \( \sigma^2 \) is equivalent to minimizing

\[
g(\sigma^2) = K \ln \sigma + \frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - m)^2
\]

Using the invariance property, it is easier to differentiate \( g(\sigma^2) \) with respect to \( \sigma \) to obtain \( \hat{\sigma}_{ml} \) the MLE of \( \sigma \), instead of \( \hat{\sigma}_{ml}^2 \) the MLE of \( \sigma^2 \). Hence,

\[
\frac{dg(\sigma^2)}{d\sigma} = \frac{K}{\sigma} - \frac{1}{\sigma^3} \sum_{k=1}^{K} (y_k - m)^2 = 0 \quad \text{or} \quad \hat{\sigma} = \sqrt{\frac{1}{K} \sum_{k=1}^{K} (y_k - m)^2}
\]

Consequently, the MLE of \( \sigma^2 \) is \( \hat{\sigma}_{ml}^2 = (1/K) \sum_{k=1}^{K} (y_k - m)^2 \).

### 6.3 Generalized Likelihood Ratio Test

In Example 5.9, we solved the hypothesis testing problem where the alternative hypothesis was composite. The parameter \( m \) under hypothesis \( H_1 \) was unknown, although it was known that \( m \) was either positive or negative. When \( m \) was positive only (negative only), a UMP test existed and the decision rule was
Parameter Estimation

\[
H_1: \quad \sigma^2 > \frac{m}{\ln \eta + \frac{m}{2}} = \gamma_1
\]
\[
H_0
\]
for positive \(m\), and

\[
H_0: \quad \sigma^2 > \frac{m}{\ln \eta + \frac{m}{2}} = \gamma_2
\]
\[
H_1
\]
for negative \(m\). Since the test designed for positive \(m\) was not the same as the test designed for negative \(m\), we concluded that a UMP test did not exist for all possible values of \(m\); that is, positive and negative. This requires that different tests be used. One approach is to use the concepts developed in Section 6.2. That is, we use the required data to estimate \(\theta\), as though hypothesis \(H_1\) is true. Then, we use these estimates in the likelihood ratio test as if they are the correct values. There are many ways to estimate \(\theta\), as will be shown in this chapter. If the estimates used are the maximum likelihood estimates, then the result is called the generalized likelihood ratio test and is given by

\[
\Lambda_g(y) = \frac{\max_{\theta_1} \int y \theta_1(y | \theta_1)}{\max_{\theta_0} \int y \theta_0(y | \theta_0)} \quad H_1
\]
\[
\Lambda_g(y) < \eta \quad H_0
\]

(6.4)

\(\theta_1\) and \(\theta_0\) are the unknown parameters to be estimated under hypotheses \(H_1\) and \(H_0\), respectively.

Example 6.2
Consider the problem of Example 5.9, where \(m\) is an unknown parameter. Obtain the generalized likelihood ratio test and compare it to the optimum Neyman-Pearson test.

Solution
Since the \(K\) observations are independent, the conditional density functions under both hypotheses \(H_1\) and \(H_0\) are
\[ H_0 : f_{Y|M,H_0}(y|m,H_0) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{y_k^2}{2\sigma^2}\right) \]

\[ H_1 : f_{Y|M,H_1}(y|m,H_1) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(y_k - m)^2}{2\sigma^2}\right) \]

where \( m \) is an unknown parameter. Since hypothesis \( H_0 \) does not contain \( m \) (\( H_0 \) is simple), the estimation procedure is applicable to hypothesis \( H_1 \) only. From the likelihood equation given by (6.3), the ML estimate of \( m \) under \( H_1 \) is given by

\[
\frac{\partial}{\partial m} \ln f_{Y|M,H_1}(y|m,H_1) = 0
\]

Substituting for \( f_{Y|M,H_1}(y|m,H_1) \) in the above equation, we have

\[
\frac{\partial}{\partial m} \left[ -\frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - \hat{m})^2 \right] = 0 \quad \text{or} \quad \hat{m} = \frac{1}{K} \sum_{k=1}^{K} y_k
\]

The details are given in Example 6.1. The likelihood ratio test becomes

\[
\Lambda_n(y) = \frac{\prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2} (y_k - \hat{m})^2\right)_{H_1}}{\prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{y_k^2}{2\sigma^2}\right)}_{H_0} > \ln \eta
\]

Substituting for the obtained value of \( \hat{m} \) in the above expression, and simplifying after taking the logarithm, the test becomes

\[
\frac{1}{2\sigma^2 K} \left( \sum_{k=1}^{K} y_k \right)^2_{H_1} > \ln \eta
\]

\[
\text{Since } \left( \sum_{k=1}^{K} y_k \right)^2 \text{ is nonnegative, the decision will always be } H_1 \text{ if } \eta \text{ is less than one ( } \ln \eta \text{ negative) or } \eta \text{ is set equal to one. Consequently, } \eta \text{ can always be chosen greater than or equal to one. Thus, an equivalent test is}
\]
Parameter Estimation

\[
\left( \frac{1}{\sqrt{K}} \sum_{k=1}^{K} y_k \right)^2 \begin{cases} 
\text{if } H_1 & > 2\sigma^2 \ln \eta = \gamma_1^2 \\
\text{if } H_0 & < \end{cases}
\]

where \( \gamma_1 \geq 0 \). Equivalently, we can use the test

\[
|Z| = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} y_k \begin{cases} 
\text{if } H_1 & > \gamma_1 \\
\text{if } H_0 & \end{cases}
\]

The decision regions are shown in Figure 6.1.

Given the desired probability of false alarm, the value of \( \gamma_1 \) can be determined. Before we can get an expression for \( P_F \), the probability of false alarm, we need to determine the density function of \( Z \). Since

\[
Z = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} y_k
\]

the mean and variance of \( Y \) under hypothesis \( H_0 \) are zero and \( \sigma^2 \), respectively. All the observations are Gaussian and statistically independent. Thus, the density function of \( Z = \sum_{k=1}^{K} Y_k \) is Gaussian with mean zero and variance \( K\sigma^2 \).

Consequently, \( Z \) is Gaussian with mean zero and variance \( \sigma^2 \). That is,

\[
f_{Z\mid H_0}(Z \mid H_0) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{z^2}{2\sigma^2}\right)
\]

The probability of false alarm, from Figure 6.2, is

Figure 6.1 Decision regions of the generalized likelihood ratio test.
We observe that we are able to determine the value $\gamma_1$ from the derived probability of false alarm without any knowledge of $m$. However, the probability of detection cannot be determined without $m$, but can be evaluated with $m$ as a parameter. Under hypothesis $H_1$, $Z_1 = \sum_{k=1}^{K} Y_k$ is Gaussian with mean $Km$ and variance $K\sigma^2$. Hence, the density function of $Z$ is Gaussian with mean $\sqrt{K}m$ and variance $\sigma^2$. That is,

$$f_{Z|H_1}(z \mid H_1) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(z-\sqrt{K}m)^2}{2\sigma^2}\right]$$

The probability of detection for a given value of $m$, from Figure 6.3, is

$$P_D = P(\text{decide } H_1 \mid H_1 \text{ true})$$

$$= \int_{-\infty}^{\gamma_1} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(z-\sqrt{K}m)^2}{2\sigma^2}\right] dz + \int_{\gamma_1}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(z-\sqrt{K}m)^2}{2\sigma^2}\right] dz$$

$$= 1 - Q\left(\frac{\gamma_1 - \sqrt{K}m}{\sigma}\right) + Q\left(\frac{\gamma_1 + \sqrt{K}m}{\sigma}\right) = Q\left(\frac{\gamma_1 + \sqrt{K}m}{\sigma}\right) + Q\left(\frac{\gamma_1 - \sqrt{K}m}{\sigma}\right)$$
In Figure 3.31 of [1], it is shown that the generalized likelihood ratio test performs nearly as well as the Neyman-Pearson test.

### 6.4 SOME CRITERIA FOR GOOD ESTIMATORS

Since the estimator $\hat{\theta}$ is a random variable and may assume more than one value, some characteristics of a “good” estimate need to be determined.

**Unbiased Estimate** We say $\hat{\theta}$ is an unbiased estimator for $\theta$ if

$$E[\hat{\theta}] = \theta \quad \text{for all } \theta$$

(6.5)

**Bias of Estimator** Let

$$E[\hat{\theta}] = \theta + b(\theta)$$

(6.6)

1. If $b(\theta)$ does not depend on $\theta$, $b(\theta) = b$, we say that the estimator $\hat{\theta}$ has a known bias. That is, $(\hat{\theta} - \theta)$ is an unbiased estimate.

2. When $b(\theta) \neq b$, an unbiased estimate cannot be obtained, since $\theta$ is unknown. In this case, we say that the estimator has an unknown bias.

When the parameter $\theta$ to be estimated satisfies (6.5) and is not random (i.e., there is no a priori probability distribution for $\theta$), it is sometimes referred to as absolutely unbiased.
The fact that the estimator is unbiased, which means that the average value of the estimate is close to the true value, does not necessarily guarantee that the estimator is “good.” This is easily seen by the conditional density function of the estimator shown in Figure 6.4. We observe that even though the estimate is unbiased, sizable errors are likely to occur, since the variance of the estimate is large. However, if the variance is small, the variability of the estimator about its expected value is also small. Consequently, the variability of the estimator is close to the true value, since the estimate is unbiased, which is a desired feature. Hence, we say that the second measure of quality of the estimate is to have a small variance.

Unbiased Minimum Variance \( \hat{\theta} \) is a minimum variance and unbiased (MVU) estimate of \( \theta \) if, for all estimates \( \theta' \) such that \( E[\theta'] = \theta \), we have \( \text{var}[\hat{\theta}] \leq \text{var}[\theta'] \) for all \( \theta' \). That is, \( \hat{\theta} \) has the smallest variance among all unbiased estimates of \( \theta \).

Consistent Estimate \( \hat{\theta} \) is a consistent estimate of the parameter \( \theta \), based on \( K \) observed samples, if

\[
\lim_{K \to \infty} P \left( \left| \hat{\theta} - \theta \right| > \varepsilon \right) = 0 \quad \text{for all } \varepsilon > 0 \tag{6.7}
\]

where \( P(\cdot) \) denotes probability.

Applying the above definition to verify the consistency of an estimate is not simple. The following theorem is used instead.

Theorem. Let \( \hat{\theta} \) be an unbiased estimator of \( \theta \) based on \( K \) observed samples. If

\[
\lim_{K \to \infty} E[\hat{\theta}] = \theta \tag{6.8}
\]

Figure 6.4 Density function of the unbiased estimator \( \hat{\theta} \).
and if

$$\lim_{K \to \infty} \text{var} E[\hat{\theta}] = 0$$  \hspace{1cm} (6.9)$$

then \( \hat{\theta} \) is a consistent estimator of \( \theta \).

**Example 6.3**

(a) Verify if the estimator \( \hat{m}_{ml} \) of Example 6.1 is an unbiased estimate of \( m \).

(b) Is the estimator \( \hat{\sigma}_{ml}^2 \) unbiased?

**Solution**

(a) The estimator \( \hat{m}_{ml} \) is unbiased if \( E[\hat{m}_{ml}] = m \). After substitution, we obtain

$$E[\hat{m}_{ml}] = E\left[ \frac{1}{K} \sum_{k=1}^{K} Y_k \right] = \frac{1}{K} E\left[ \sum_{k=1}^{K} Y_k \right] = \frac{1}{K} Km = m$$

Hence, \( \hat{m}_{ml} \) is unbiased.

(b) The estimator \( \hat{\sigma}_{ml}^2 \) is unbiased if \( E[\hat{\sigma}_{ml}^2] = \sigma^2 \). That is,

$$E\left[ \frac{1}{K} \sum_{k=1}^{K} (Y_k - m)^2 \right] = \frac{1}{K} E\left[ Km^2 + \sum_{k=1}^{K} Y_k^2 - 2m \sum_{k=1}^{K} Y_k \right] = \sigma^2$$

Hence, \( \hat{\sigma}_{ml}^2 \) is unbiased.

**6.5 BAYES’ ESTIMATION**

In the Bayes’ estimation, we assign a cost \( C(\theta, \hat{\theta}) \) to all pairs \( (\theta, \hat{\theta}) \). The cost is a nonnegative real value function of the two random variables \( \theta \) and \( \hat{\theta} \). As in the Bayes’ detection, the risk function is defined to be the average value of the cost; that is,

$$\mathcal{R} = E[C(\theta, \hat{\theta})]$$  \hspace{1cm} (6.10)
The goal is to minimize the risk function in order to obtain \( \hat{\theta} \), which is the optimum estimate. In many problems, only the error \( \tilde{\theta} \) between the estimate and the true value is of interest; that is,

\[
\tilde{\theta} = \theta - \hat{\theta}
\]  

(6.11)

Consequently, we will only consider costs which are a function of the error. Three cases will be studied, and their corresponding sketches are shown in Figure 6.5.

1. Squared error

\[
C(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2
\]  

(6.12)

2. Absolute value of error

\[
C(\theta, \hat{\theta}) = |\theta - \hat{\theta}|
\]  

(6.13)

3. Uniform cost function

\[
C(\theta, \hat{\theta}) = \begin{cases} 
1, & |\theta - \hat{\theta}| \geq \frac{\Delta}{2} \\
0, & |\theta - \hat{\theta}| < \frac{\Delta}{2}
\end{cases}
\]  

(6.14)

The unknown parameter is assumed to be a continuous random variable with density function \( f_{\theta}(\theta) \). The risk function can then be expressed as

Figure 6.5 Cost functions: (a) squared error, (b) absolute value of error, and (c) uniform.
Parameter Estimation

\[ R = E[C(\theta, \hat{\theta})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\theta, \hat{\theta}) f_{\theta,Y}(\theta, y) d\theta dy \] \hspace{1cm} (6.15)

Note that we take the cost average over all possible values of \( \theta \) and \( Y \), where \( Y \) is the vector \([Y_1 \ Y_2 \ \ldots \ Y_K]^T\). We now find the estimator for the three cost functions considered.

### 6.5.1 Minimum Mean-Square Error Estimate

The estimator that minimizes the risk function for the cost given in (6.12) is referred to as a minimum mean-square estimate (MMSE). The corresponding risk function is denoted by \( R_{ms} \). We have

\[ R_{ms} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f_{\theta,Y}(\theta, y) d\theta dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f_{\theta,Y}(\theta, y) d\theta dy \] \hspace{1cm} (6.16)

Using (1.91), the risk function can be rewritten as

\[ R_{ms} = \int_{-\infty}^{\infty} dy f_Y(y) \left[ \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f_{\theta|Y}(\theta | y) d\theta \right] \] \hspace{1cm} (6.17)

Since the density function \( f_Y(y) \) is nonnegative, minimizing \( R_{ms} \) is equivalent to minimizing the expression in brackets of the above equation. Hence, taking the derivative with respect to \( \hat{\theta} \) and setting it equal to zero, we have

\[ \frac{d}{d\hat{\theta}} \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f_{\theta|Y}(\theta | y) d\theta = 0 \] \hspace{1cm} (6.18)

Using Leibniz’s rule given in (1.38), we obtain

\[ \hat{\theta}_{ms} = \int_{-\infty}^{\infty} \theta f_{\theta|Y}(\theta | y) d\theta = E[\theta | y] \] \hspace{1cm} (6.19)

That is, the minimum mean-square estimate \( \hat{\theta}_{ms} \) represents the conditional mean of \( \theta \) given \( Y \). It can easily be shown that the second derivative with respect to \( \hat{\theta}_{ms} \) is positive-definite, which corresponds to a unique minimum of \( R_{ms} \), and is given by
The conditional variance of $\theta$ given $Y$ is
\[
\text{var}[\theta | y] = \int_{-\infty}^{\infty} \{\theta - E[\theta | y]\}^2 f_{\theta | Y}(\theta | y) d\theta
\]  
(6.21)

Hence, $\mathcal{R}_m$ is just the conditional variance of $\theta$ given $Y$, averaged over all possible values of $Y$. This estimation procedure using the squared error criterion is sometimes referred to as a minimum variance (MV) of error estimation.

### 6.5.2 Minimum Mean Absolute Value of Error Estimate

In this case, the cost function is given by (6.13), and the risk is
\[
\mathcal{R}_{abs} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{\theta - \hat{\theta}\}^2 f_{\theta | Y}(\theta | y) d\theta dy
\]
(6.22)

Using the same arguments as in Section 6.5.1, the risk can be minimized by minimizing the integral in brackets, which is given by
\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{\theta - \hat{\theta}\}^2 f_{\theta | Y}(\theta | y) d\theta dy
\]
(6.23)

Differentiating (6.23) with respect to $\hat{\theta}$, and setting the result equal to zero, we obtain
\[
\int_{-\infty}^{\infty} f_{\theta | Y}(\theta | y) d\theta = \hat{\theta}_{abs}
\]
(6.24)

That is, the estimate $\hat{\theta}_{abs}$ is just the median of the conditional density function $f_{\theta | Y}(\theta | y)$. This estimate is also known as the minimum mean absolute value of error (MAVE) estimate, and thus $\hat{\theta}_{abs} = \hat{\theta}_{mave}$.
6.5.3 Maximum A Posteriori Estimate

For the uniform cost function given by (6.14), the Bayes’ risk becomes

\[
\mathcal{R}_{\text{unf}} = \int_{-\infty}^{\infty} f_Y(y) \left[ \int_{-\infty}^{\hat{\theta} + \frac{\Delta}{2}} f_{\theta|Y}(\theta|y) d\theta + \int_{\hat{\theta} - \frac{\Delta}{2}}^{\infty} f_{\theta|Y}(\theta|y) d\theta \right] dy
\]

or

\[
\mathcal{R}_{\text{unf}} = \int_{-\infty}^{\infty} f_Y(y) \left[ 1 - \int_{\hat{\theta} - \frac{\Delta}{2}}^{\hat{\theta} + \frac{\Delta}{2}} f_{\theta|Y}(\theta|y) d\theta \right] dy
\]

(6.25)

where

\[
\int_{\hat{\theta} - \frac{\Delta}{2}}^{\hat{\theta} + \frac{\Delta}{2}} f_{\theta|Y}(\theta|y) d\theta = P\left[ \hat{\theta} - \frac{\Delta}{2} \leq \Theta \leq \hat{\theta} + \frac{\Delta}{2} | y \right]
\]

(6.26)

\(P[\cdot]\) denotes probability. Hence, the risk \(\mathcal{R}_{\text{unf}}\) is minimized by maximizing (6.26). Note that in maximizing (6.26) (minimizing \(\mathcal{R}_{\text{unf}}\)), we are searching for the estimate \(\hat{\theta}\), which minimizes \(f_{\theta|Y}(\theta|y)\). This is called the maximum a posteriori estimate (MAP), \(\hat{\theta}_{\text{map}}\), which is defined as

\[
\frac{\partial f_{\theta|Y}(\theta|y)}{\partial \theta} \bigg|_{\theta = \hat{\theta}_{\text{map}}} = 0
\]

(6.27)

Using the logarithm, which is a monotonically increasing function, (6.27) becomes

\[
\frac{\partial \ln f_{\theta|Y}(\theta|y)}{\partial \theta} = 0
\]

(6.28)

Equation (6.28) is called the MAP equation. This is a necessary but not sufficient condition, since \(f_{\theta|Y}(\theta|y)\) may have several local maxima. Using the Bayes’ rule

\[
f_{\theta|Y}(\theta|y) = \frac{f_{Y\theta}(y|\theta) f_{\theta}(\theta)}{f_Y(y)}
\]

(6.29)
and the fact that

$$\ln f_{\omega,Y}(\theta | y) = \ln f_{Y|\omega}(y | \theta) + \ln f_{\omega}(\theta) - \ln f_Y(y)$$  \hspace{1cm} (6.30)$$

then the MAP equation may be rewritten as

$$\frac{\partial \ln f_{\omega,Y}(\theta | y)}{\partial \theta} = \frac{\partial \ln f_{Y|\omega}(y | \theta)}{\partial \theta} + \frac{\partial \ln f_{\omega}(\theta)}{\partial \theta} = 0$$  \hspace{1cm} (6.31)$$

We always assume that \( \Delta \) is sufficiently small, so that the estimate \( \hat{\theta}_{\text{map}} \) is given by the MAP equation. That is, the cost function shown in Figure 6.5 may be defined as

$$C(\hat{\theta}, \theta) = 1 - \delta(\theta, \hat{\theta})$$  \hspace{1cm} (6.32)$$

**Example 6.4**

Consider the problem where the observed samples are

$$Y_k = M + N_k, \quad k = 1, 2, \ldots, K$$

\( M \) and \( N_k \) are statistically independent Gaussian random variables with zero mean and variance \( \sigma^2 \). Find \( \hat{m}_{m}, \hat{m}_{\text{map}}, \) and \( \hat{m}_{\text{move}} \).

**Solution**

From (6.19), the estimate \( \hat{m}_{m} \) is the conditional mean of \( m \) given \( Y \). The density function \( f_{M|Y}(m | y) \) is expressed as

$$f_{M|Y}(m | y) = \frac{f_{Y|M}(y | m) f_{M}(m)}{f_Y(y)}$$

where

$$f_{M}(m) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{m^2}{2\sigma^2}\right), \quad f_{Y|M}(y | m) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(y_k - m)^2}{2\sigma^2}\right)$$

and the marginal density function \( f_Y(y) \) is
Parameter Estimation

\[ f_Y(y) = \int_{-\infty}^{\infty} f_{M,Y}(m, y) dm = \int_{-\infty}^{\infty} f_{M|Y}(m \mid y) f_M(m) dm \]

Note that \( f_{M|Y}(m \mid y) \) is a function of \( m \), but that \( f_Y(y) \) is a constant with \( y \) as a parameter needed to maintain the area under the conditional density function equal to one. That is,

\[ f_{M|Y}(m \mid y) = \frac{1}{(\sqrt{2\pi}\sigma)^K} \exp\left\{ -\frac{1}{2\sigma^2} \left[ \sum_{k=1}^{K} (y_k - m)^2 + m^2 \right] \right\} \]

Expanding the exponent, we have

\[ \sum_{k=1}^{K} (y_k^2 - 2y_km + m^2) + m^2 = m^2 (K+1) - 2m \sum_{k=1}^{K} y_k + \sum_{k=1}^{K} y_k^2 \]

\[ = (K+1) m^2 - \frac{2m}{K+1} \sum_{k=1}^{K} y_k + \sum_{k=1}^{K} y_k^2 \]

\[ = (K+1) \left(m - \frac{1}{K+1} \sum_{k=1}^{K} y_k\right)^2 - \left(\frac{1}{K+1} \sum_{k=1}^{K} y_k\right)^2 + \sum_{k=1}^{K} y_k^2 \]

The last two terms in the exponent do not involve \( m \), and can be absorbed in the multiplicative constant to obtain

\[ f_{M|Y}(m \mid y) = c(y) \exp\left\{ -\frac{1}{2\sigma_m^2} \left( m - \frac{1}{K+1} \sum_{k=1}^{K} y_k \right)^2 \right\} \]

where \( \sigma_m = \sigma/\sqrt{K+1} \). By inspection, the conditional mean is

\[ \hat{m}_m = E[M \mid y] = \frac{1}{K+1} \sum_{k=1}^{K} y_k \]

According to (6.20), \( \mathfrak{R}_{mr} \) is given by

\[ \mathfrak{R}_{mr} = \int_{-\infty}^{\infty} \text{var}[M \mid y] f_Y(y) dy \]
Hence, since $\int_{-\infty}^{\infty} f_Y(y) dy = 1$, then $\mathcal{Y}_{ms} = \sigma_m^2 \int_{-\infty}^{\infty} f_Y(y) dy = \sigma_m^2$.

The MAP estimate is obtained using (6.28) and (6.29). Taking the logarithm of $f_{MY}(m \mid y)$, we have

$$\ln f_{MY}(m \mid y) = \ln c(y) - \frac{1}{\sigma_m^2} \left( m - \frac{1}{K+1} \sum_{k=1}^{K} y_k \right)^2$$

Therefore,

$$\frac{\partial \ln f_{MY}(m \mid y)}{\partial m} = - \frac{1}{\sigma_m^2} \left( m - \frac{1}{K+1} \sum_{k=1}^{K} y_k \right) = 0$$

$$\Rightarrow \hat{m}_{map} = \frac{1}{K+1} \sum_{k=1}^{K} y_k$$

That is, $\hat{m}_{map} = \hat{m}_{ms}$. We could have obtained this result directly by inspection, since we have shown that $f_{MY}(m \mid y)$ is Gaussian. Consequently, the maximum of $f_{MY}(m \mid y)$ occurs at its mean value.

Using the fact that the Gaussian density function is symmetric, and that $\hat{m}_{movem}$ is the median of the conditional density function $f_{MY}(m \mid y)$, we conclude

$$\hat{m}_{movem} = \hat{m}_{ms} = \hat{m}_{map} = \frac{1}{K+1} \sum_{k=1}^{K} y_k$$

From (6.31), if $\theta$ is assumed to be random with $f_\theta(\theta) = 0$ for $-\infty < \theta < \infty$, then the ML estimate can then be considered to be a special case of the MAP estimate. Such a density function for $\theta$ connotes zero a priori information about $\theta$. Furthermore, the MAP estimate of a Gaussian distributed parameter is equivalent to the ML estimate as the variance increases; that is, the distribution of the parameter to be estimated tends to be uniform. In general, for a symmetric distribution centered at the maximum, as shown Figure 6.6(a), the mean, mode, and median are identical. If the distribution of the parameter is uniform, then the MAP, the MMSE, and the MAVE estimates are identical. In Figure 6.6(b), we illustrate the different estimates when the density function is not symmetric. Recall that the median is the value of $y$ for which $P(Y \leq y) = P(Y \geq y) = 1/2$, while the mode is the value that has the greatest probability of occurring.
Example 6.5

Find $\hat{x}_{\text{ms}}$, the minimum mean-square, and $\hat{x}_{\text{map}}$, the maximum a posteriori estimators, of $X$ from the observation $Y = X + N$. $X$ and $N$ are random variables with density functions

$$f_X(x) = \frac{1}{2} \delta(x) + \frac{1}{2} \delta(x - 1) \quad \text{and} \quad f_N(n) = \begin{cases} \frac{1}{2} e^n, & n \leq 0 \\ \frac{1}{2}, & n \geq 0 \end{cases}$$

Solution

The estimate $\hat{x}_{\text{map}}$ maximizes the density function $f_{XY}(x \mid y)$. Since the conditional probability density function is $f_{Y \mid X}(y \mid X) = (1/2)e^{-|y-x|}$, the probability density function of $Y$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y \mid X}(y \mid x)f_X(x)dx = \frac{1}{4} \int_{-\infty}^{\infty} e^{-|y-x|}[\delta(x) + \delta(x - 1)]dx$$

$$= \frac{1}{4} \begin{cases} e^y + e^{-y-1}, & y < 0 \\ 1, & 0 \leq y < 1 \\ e^{-y} + e^{y+1}, & y \geq 1 \end{cases}$$
The a posteriori density function is, from (6.29), given by

\[
f_{X|Y}(x | y) = \frac{f_{Y|X}(y | x)f_{X}(x)}{f_{Y}(y)} = e^{-|x-y|}[\delta(x) + \delta(x-1)]
\]

where \( f_{X}(x) = e^{-|x|} \). The above expression is zero except when \( x = 0 \) and \( x = 1 \). The above expression is maximized when \(|n-x|\) is minimized. Since \( x \) can take only two values, but must be close to \( n \), we have

\[
\hat{x}_{\text{map}} = \begin{cases} 
1 & \text{for } n \geq 1/2 \\
0 & \text{for } n < 1/2 
\end{cases}
\]

The mean-square error estimate is the mean of the a posteriori density function as given by (6.19). Hence,

\[
\hat{x}_{\text{ms}} = \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx = \int_{-\infty}^{\infty} x e^{-|x-y|}[\delta(x) + \delta(x-1)] e^{-|x|} e^{-|y|} dx
\]

Since \( \int_{-\infty}^{\infty} \delta(t-t_0)g(t)dt = g(t_0) \), the mean-square estimate is

\[
\hat{x}_{\text{ms}} = \frac{e^{-|n-y|}}{e^{-|n|} + e^{-|y|}}
\]

and we see that \( \hat{x}_{\text{map}} \) is not identical to \( \hat{x}_{\text{ms}} \).

### 6.6 CRAMER-RAO INEQUALITY

From the MAP equation of (6.31), if we set the density function of \( \theta \) to zero, for all \( \theta \), we obtain the likelihood equation of (6.3). That is, the ML estimate can be considered as a special case of the MAP estimate. In this case, to check whether the estimate is “good,” we need to compute its bias and error variance and determine its consistency. It may be very difficult to obtain an expression for the error variance. In this case, the “goodness” of the estimator is studied in terms of a lower bound on the error variance. This bound is known as the Cramer-Rao
bound. The Cramer-Rao bound of a constant parameter is given by the following theorem.

**Theorem.** Let the vector $Y = [Y_1, Y_2, \ldots, Y_K]^T$ represent $K$ observations, and $\hat{\theta}$ be the unbiased estimator of $\theta$. Then

$$\text{var}[(\hat{\theta} - \theta) | \theta] \geq \frac{1}{E \left[ \left( \frac{\partial \ln f_{Y|\theta} (y | \theta)}{\partial \theta} \right)^2 \right]} \quad (6.33)$$

where

$$E \left[ \left( \frac{\partial \ln f_{Y|\theta} (y | \theta)}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial^2 \ln f_{Y|\theta} (y | \theta)}{\partial \theta^2} \right] \quad (6.34)$$

**Proof.** For an unbiased estimator $\hat{\theta}$, we have

$$E[\hat{\theta} | \theta] = 0 \quad (6.35)$$

Therefore,

$$E[(\hat{\theta} - \theta) | \theta] = \int_{-\infty}^{\infty} (\hat{\theta} - \theta) f_{Y|\theta} (y | \theta) dy = 0 \quad (6.36)$$

Differentiating (6.36) with respect to $\theta$, we obtain

$$\int_{-\infty}^{\infty} (\hat{\theta} - \theta) \frac{\partial f_{Y|\theta} (y | \theta)}{\partial \theta} dy = \int_{-\infty}^{\infty} f_{Y|\theta} (y | \theta) dy = 0 \quad (6.37)$$

The second integral is equal to one. Using the fact that

$$\frac{\partial \ln g(x)}{\partial x} = \frac{1}{g(x)} \frac{\partial g(x)}{\partial x} \quad (6.38)$$

where $g(x)$ is a function of $x$, we can express $\frac{\partial f_{Y|\theta} (y | \theta)}{\partial \theta}$ as
\[
\frac{\partial f_{Y|\Theta}(y|\theta)}{\partial \theta} = f_{Y|\Theta}(y|0) \frac{\partial \ln f_{Y|\Theta}(y|\theta)}{\partial \theta}
\]  
(6.39)

Substituting (6.39) into (6.37), we obtain
\[
\int_{-\infty}^{\infty} (\hat{\theta} - \theta) f_{Y|\Theta}(y|0) \frac{\partial \ln f_{Y|\Theta}(y|\theta)}{\partial \theta} dy = 1
\]  
(6.40)

The Schwarz inequality states that
\[
\left[ \int_{-\infty}^{\infty} x^2(t) dt \right] \left[ \int_{-\infty}^{\infty} y^2(t) dt \right] \geq \left[ \int_{-\infty}^{\infty} x(t)y(t) dt \right]^2
\]  
(6.41)

where \(x(t)\) and \(y(t)\) are two functions of \(t\). Equality holds if and only if \(y(t) = c x(t)\), with \(c\) a constant. Rewriting (6.39) in order to use the Schwarz inequality, we have
\[
\int_{-\infty}^{\infty} \frac{\partial \ln f_{Y|\Theta}(y|\theta)}{\partial \theta} \sqrt{f_{Y|\Theta}(y|\theta)} \left[ (\hat{\theta} - \theta) \sqrt{f_{Y|\Theta}(y|0)} \right] dy = 1
\]  
(6.42)

or
\[
\left[ \int_{-\infty}^{\infty} (\hat{\theta} - \theta)^2 \sqrt{f_{Y|\Theta}(y|0)} dy \right] \left\{ \int_{-\infty}^{\infty} \frac{\partial \ln f_{Y|\Theta}(y|\theta)}{\partial \theta} \right\}^2 f_{Y|\Theta}(y|\theta) dy \geq 1
\]  
(6.43)

The first integral between brackets is actually \(\text{var}[(\hat{\theta} - \theta)|\theta]\). Hence, the inequality becomes
\[
\text{var}[(\hat{\theta} - \theta)|\theta] \geq \frac{1}{\mathbb{E}\left\{ \left( \frac{\partial \ln f_{Y|\Theta}(y|\theta)}{\partial \theta} \right)^2 \right\}}
\]  
(6.44)

which proves (6.33).

We now prove (6.34), which says that the Cramer-Rao bound can be expressed in a different form. We know that
Parameter Estimation

\[ \int_{-\infty}^{\infty} f_{Y|\theta}(y|\theta) \, dy = 1 \quad (6.45) \]

Differentiating both sides of the equation with respect to \( \theta \) results in

\[ \int_{-\infty}^{\infty} \frac{\partial f_{Y|\theta}(y|\theta)}{\partial \theta} \, dy = 0 \quad (6.46) \]

Rewriting (6.46) and using (6.38), we have

\[ \int_{-\infty}^{\infty} \frac{\partial \ln f_{Y|\theta}(y|\theta)}{\partial \theta} f_{Y|\theta}(y|\theta) \, dy = 0 \quad (6.47) \]

Differentiating again with respect to \( \theta \), we obtain

\[ \int_{-\infty}^{\infty} \frac{\partial^2 \ln f_{Y|\theta}(y|\theta)}{\partial \theta^2} f_{Y|\theta}(y|\theta) \, dy + \int_{-\infty}^{\infty} \frac{\partial \ln f_{Y|\theta}(y|\theta)}{\partial \theta} \frac{\partial f_{Y|\theta}(y|\theta)}{\partial \theta} = 0 \quad (6.48) \]

Substituting (6.47) for the second term of the second integral of (6.48), and rearranging terms yields

\[ E \left[ \frac{\partial^2 \ln f_{Y|\theta}(y|\theta)}{\partial \theta^2} \right] = -E \left[ \left( \frac{\partial \ln f_{Y|\theta}(y|\theta)}{\partial \theta} \right)^2 \right] \quad (6.49) \]

which is the same as (6.34), and the proof of the theorem is complete.

An important observation about (6.43) is that equality holds if and only if

\[ \frac{\partial \ln f_{Y|\theta}(y|\theta)}{\partial \theta} = c(\theta)(\hat{\theta} - \theta) \quad (6.50) \]

Any unbiased estimator that satisfies the equality in the Cramer-Rao inequality of (6.33) is said to be an efficient estimator.

If an efficient estimator exists, it can easily be shown that it equals the ML estimate. The ML equation is given by

\[ \left. \frac{\partial \ln f_{Y|\theta}(y|\theta)}{\partial \theta} \right|_{\theta = \hat{\theta}} = 0 \quad (6.51) \]
Using (6.50), provided that an efficient estimate exists, we have

\[
\frac{\partial \ln f_{Y \theta}(y \mid \theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}_{ML}} = c(0)(\hat{\theta} - 0) \bigg|_{\theta = \hat{\theta}_{ML}} \tag{6.52}
\]

which equals zero when \( \hat{\theta} = \hat{\theta}_{ML} \).

**Example 6.6**

Consider \( K \) observations, such that

\[ Y_k = m + N_k, \quad k = 1, 2, \ldots, K \]

where \( m \) is unknown and \( N_k \)s are statistically independent zero mean Gaussian random variables with unknown variance \( \sigma^2 \).

(a) Find the estimates \( \hat{m} \) and \( \hat{\sigma}^2 \) for \( m \) and \( \sigma^2 \), respectively.

(b) Is \( \hat{m} \) an efficient estimator?

(c) Find the conditional variance of the error \( \text{var}((\hat{m} - m) \mid m) \).

**Solution**

(a) Using (6.2), we can determine \( \hat{m} \) and \( \hat{\sigma}^2 \) simultaneously. The conditional density function of \( Y \) given \( m \) and \( \sigma^2 \) is

\[
f_Y(y \mid m, \sigma^2) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{(y_k - m)^2}{2\sigma^2} \right]
\]

Taking the logarithm, we have

\[
\ln f_Y(y \mid m, \sigma^2) = -\frac{K}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - m)^2
\]

We take the derivative of the above equation with respect to \( m \) and \( \sigma^2 \) to obtain two equations in two unknowns. That is,

\[
\frac{\partial \ln f_Y(y \mid m, \sigma^2)}{\partial m} = 2 \sum_{k=1}^{K} \frac{y_k - m}{2\sigma^2} = 0
\]
Parameter Estimation

and

\[ \frac{\partial \ln f_y (y \mid m, \sigma^2)}{\partial \sigma^2} = -\frac{K}{2\sigma^2} + \sum_{k=1}^{K} \frac{(y_k - m)^2}{2\sigma^4} = 0 \]

Solving for \( \hat{m}_{ml} \) and \( \hat{\sigma}_{ml}^2 \) simultaneously, we obtain

\[ \hat{m}_{ml} = \frac{1}{K} \sum_{k=1}^{K} y_k \]

and

\[ \hat{\sigma}_{ml}^2 = \frac{1}{K} \sum_{k=1}^{K} \left( y_k - \frac{1}{K} \sum_{k=1}^{K} y_k \right)^2 = \frac{1}{K} \sum_{k=1}^{K} (y_k - \hat{m}_{ml})^2 \]

(b) \( \hat{m}_{ml} \) is an unbiased estimator since

\[ E[\hat{m}_{ml}] = \frac{1}{K} E\left[ \sum_{k=1}^{K} y_k \right] = m \]

To check if the estimator is efficient, we use (6.50) to obtain

\[ \frac{\partial^2 \ln f_y (y \mid m, \sigma^2)}{\partial m^2} = \sum_{k=1}^{K} \frac{y_k - m}{\sigma^2} = \frac{K}{\sigma^2} \left( \frac{1}{K} \sum_{k=1}^{K} y_k - \hat{m} \right) \]

where \( c(m) = K / \sigma^2 \) and \( \hat{m} = (1 / K) \sum_{k=1}^{K} y_k = \hat{m}_{ml} \). Hence, the estimator is efficient.

(c) To determine the conditional variance of error, we use (6.33) and (6.34). Taking the derivative of the likelihood equation with respect to \( m \), we obtain

\[ \frac{\partial^2 \ln f_y (y \mid m, \sigma^2)}{\partial m^2} = -\frac{K}{\sigma^2} \]

Hence,
\[ \text{var}(\hat{m} - m) = -\frac{1}{E\left[\frac{\partial^2 \ln f_Y(y|m,\sigma^2)}{\partial m^2}\right]} = \frac{\sigma^2}{K} \]

**Cramer-Rao Inequality for a Random Parameter**

We suppose that $\theta$ is a random parameter, such that the joint density function $f_{Y\theta}(y|\theta)$ of the observation vector $Y$ and the parameter $\theta$ are known. Then,

\[ \text{var}(\hat{\theta} - \theta) \geq \frac{1}{E\left[\frac{\partial^2 \ln f_{Y\theta}(y,\theta)}{\partial \theta^2}\right]} \quad (6.53) \]

where

\[ E\left[\frac{\partial}{\partial \theta} \ln f_{Y\theta}(y,\theta)\right]^2 = -E\left[\frac{\partial^2}{\partial \theta^2} \ln f_{Y\theta}(y,\theta)\right] \quad (6.54) \]

Equality of (6.53) holds if and only if

\[ \frac{\partial}{\partial \theta} \ln f_{Y\theta}(y,\theta) = c(\hat{\theta} - \theta) \quad (6.55) \]

where $c$ is independent of $Y$ and $\theta$. Furthermore, the lower bound of (6.53) is achieved with equality if and if $f_{Y\theta}(\theta | y)$ is Gaussian.

It also can be shown that if the lower bound on the nonrandom parameter of (6.34) is denoted $J$ and if the lower bound on the random parameter of (6.54) is denoted $L$, then

\[ L = J - E\left[\frac{\partial^2 \ln f_{\theta}(\theta)}{\partial \theta^2}\right] \quad (6.56) \]

Next, we present the generalization of the Cramer-Rao bound for a vector parameter on multiple parameter estimation for both random and nonrandom parameters.
6.7 MULTIPLE PARAMETER ESTIMATION

In many radar and communication applications, it may be necessary to examine several parameters simultaneously. For example, in a radar application, a problem may be to estimate the range and velocity of a target; while in a communication application, the problem may be to estimate the amplitude, arrival time, and a carrier frequency of a received signal. Therefore, we can now extend the parameter estimation concepts to multiple parameters. The vector to be estimated may be random (in this case we use the Bayes’ estimation) or nonrandom (in this case we use the maximum likelihood estimation).

6.7.1 θ Nonrandom

In this case, the vector θ is

$$θ = [θ_1, θ_2, \ldots, θ_K]^T$$  \hspace{1cm} (6.57)

Then, (6.3) becomes the following set of simultaneous likelihood equations

$$\frac{∂}{∂θ_1} \ln f_{Y|θ}(y_1, y_2, \ldots, y_K \mid θ_1, θ_2, \ldots, θ_K) = 0$$

$$\frac{∂}{∂θ_2} \ln f_{Y|θ}(y_1, y_2, \ldots, y_K \mid θ_1, θ_2, \ldots, θ_K) = 0$$

$$\vdots$$

$$\frac{∂}{∂θ_K} \ln f_{Y|θ}(y_1, y_2, \ldots, y_K \mid θ_1, θ_2, \ldots, θ_K) = 0$$  \hspace{1cm} (6.58)

In order to write (6.58) in a more compact form, we define the partial derivative column vector by

$$∇_θ = \left[ \frac{∂}{∂θ_1}, \frac{∂}{∂θ_2}, \ldots, \frac{∂}{∂θ_K} \right]^T$$  \hspace{1cm} (6.59)

This operation is generally applied to row vectors only. That is, if

$$X^T = [X_1, X_2, \ldots, X_n]$$

then
The ML equation is then

$$\nabla_{\theta} \left[ \ln f_{Y|\theta}(y | \theta) \right] \bigg|_{\theta = \hat{\theta}_{\omega}(y)} = 0$$

(6.60)

We saw in Section 6.4 that a measure of quality of the estimate is the bias. The conditional mean of the estimate given by (6.6) becomes

$$E[\hat{\theta}(y) | \theta] = \theta + b(\theta)$$

(6.61)

If the bias vector $b(\theta) = 0$, that is, each component of the bias vector is zero for any $\theta$, then the estimate is said to be unbiased. We note that

$$b(\theta) = E[ (\hat{\theta}(y) - \theta) | \theta ] = E[ \tilde{\theta}(y) ] = E[ \hat{\theta}(y) ] - \theta$$

(6.62)

A second measure of quality of the estimate is the conditional variance of the error. For multiple parameters, the corresponding conditional covariance matrix of the error is

$$\tilde{C} = E[ (\tilde{\theta} - \tilde{\theta}_{\theta}) (\tilde{\theta}^T - \tilde{\theta}_{\theta}^T ) | \theta ]$$

(6.63)

where $\tilde{\theta}_{\theta}$ is the bias vector given by

$$\tilde{\theta}_{\theta} = E[ \tilde{\theta}(y) | \theta ] = b(\theta)$$

(6.64)

Note that $\tilde{C}$ is a $K \times K$ matrix. The $i$th element is

$$\tilde{C}_{ij} = E[ (\tilde{\theta}_i - \theta_{\theta i}) (\tilde{\theta}_j - \theta_{\theta j}) | \theta ]$$

(6.65)

while the $i$th diagonal element is the conditional variance given by
Parameter Estimation

\[ \text{var}[\hat{\theta}_i] = \tilde{C}_{ii} = \text{var}[\hat{\theta}_i | \theta] = \text{var}[(\hat{\theta}_i(y) - \theta_i) | \theta] \tag{6.66} \]

**Cramer-Rao Bound**

The extension of the Cramer-Rao bound is given by the following theorem.

**Theorem.** If \( \hat{\theta} \) is any absolutely unbiased estimator of \( \theta \) based on the observation vector \( Y \), then the covariance of the error in the estimator is bounded by the inverse, assuming it exists, of the Fisher information matrix \( J \).

\[ E[(\theta - \hat{\theta})(\theta - \hat{\theta})^T | \theta] \geq J^{-1} \tag{6.67} \]

where

\[ J = E\left[ \left( \frac{\partial}{\partial \theta} \ln f_{Y|\theta}(Y | \theta) \right)^T \left( \frac{\partial}{\partial \theta} \ln f_{Y|\theta}(Y | \theta) \right) \right] = -E\left[ \frac{\partial^2}{\partial \theta^2} \ln f_{Y|\theta}(Y | \theta) \right] \tag{6.68} \]

\( J^{-1} \) is the inverse matrix of the Fisher information matrix. Equality holds only if

\[ \left( \frac{\partial}{\partial \theta} \ln f_{Y|\theta}(y | \theta) \right)^T = \epsilon(\theta) [\theta - \hat{\theta}] \tag{6.69} \]

The derivatives \( \frac{\partial f_{Y|\theta}(y | \theta)}{\partial \theta} \) and \( \frac{\partial^2 f_{Y|\theta}(y | \theta)}{\partial \theta^2} \) are assumed to exist and to be absolutely integrable. The Fisher information matrix is defined as

\[ J = E\left[ \nabla_\theta \left[ \ln f_{Y|\theta}(y | \theta) \right] \right] \left( \nabla_\theta \left[ \ln f_{Y|\theta}(y | \theta) \right] \right)^T | \theta \tag{6.70} \]

which can also be rewritten as

\[ J = -E\left[ \nabla_\theta \left[ \ln f_{Y|\theta}(y | \theta) \right] \right] \left( \nabla_\theta \left[ \ln f_{Y|\theta}(y | \theta) \right] \right)^T | \theta \tag{6.71} \]

For simplicity, we give the conditional variance on the error \( \tilde{\theta}_i = \hat{\theta}_i - \theta_i, i = 1, 2, \ldots, K \), which is bounded by the inequality
\[ \sigma_{\theta_i}^2 = \text{var}\left[ \hat{\theta}_i \mid \theta \right] = \text{var}\left[ (\hat{\theta}_i(y) - 0) \mid \theta \right] \geq J^{ii} \quad (6.72) \]

\( J^{ii} \) is the \( i \)th diagonal element in the \( K \times K \) square matrix \( J^{-1} \). The \( ij \)th element of \( J \) in (6.70) is given by
\[
J_{ij} = \mathbb{E} \left[ \frac{\partial \ln f_{Y \mid \Theta}(y \mid \theta)}{\partial \theta_i} \frac{\partial \ln f_{Y \mid \Theta}(y \mid \theta)}{\partial \theta_j} \mid \theta \right] \quad (6.73)
\]

whereas the \( ji \)th element of (6.71) is given by
\[
J_{ji} = -\mathbb{E} \left[ \frac{\partial^2 \ln f_{Y \mid \Theta}(y \mid \theta)}{\partial \theta_i \partial \theta_j} \mid \theta \right] \quad (6.74)
\]

**Proof.** One way to prove the above theorem without resorting to excessive matrix operation is the following. Since the estimations are unbiased (the expected value of each estimator is the true value), we can write
\[
\mathbb{E}[\hat{\theta}_i(y) \mid \theta] = \int_{-\infty}^{\infty} \hat{\theta}_i(y) f_{Y \mid \Theta}(y \mid \theta) dy = \theta_i \quad (6.75)
\]
or
\[
\int_{-\infty}^{\infty} [\hat{\theta}_i(y) - \theta_i] f_{Y \mid \Theta}(y \mid \theta) dy = 0 \quad (6.76)
\]

Differentiating both sides of (6.76) with respect to \( \theta_j \), we have
\[
\int_{-\infty}^{\infty} \hat{\theta}_i(y) \frac{\partial \ln f_{Y \mid \Theta}(y \mid \theta)}{\partial \theta_j} dy = \frac{\partial \theta_i}{\partial \theta_j} \quad (6.77)
\]

Using (6.38) for the integral, and the fact that \( \partial \theta_i / \partial \theta_j \) is the Kronecker \( \delta_{ij} \) (unity for \( i = j \), and zero otherwise), (6.77) can be rewritten as
\[
\int_{-\infty}^{\infty} \hat{\theta}_i(y) f_{Y \mid \Theta}(y \mid \theta) \frac{\partial \ln f_{Y \mid \Theta}(y \mid \theta)}{\partial \theta_j} dy = \delta_{ij} \quad (6.78)
\]
Consider the case when $j = 1$, and define the $K+1$ dimensional vector $X$ ($K$ is the number of parameters to be estimated) as

$$
X = \begin{bmatrix}
\hat{\theta}_1 - \theta_1 \\
\frac{\partial \ln f_{Y\theta}(y \mid \theta)}{\partial \theta_1} \\
\frac{\partial \ln f_{Y\theta}(y \mid \theta)}{\partial \theta_2} \\
\vdots \\
\frac{\partial \ln f_{Y\theta}(y \mid \theta)}{\partial \theta_K}
\end{bmatrix}
$$

(6.79)

Note that the mean values of the components of $X$ are all zero. The first term is zero because the estimate is unbiased, while the other terms are zero in light of (6.35), which can be written as

$$
\int_{-\infty}^{\infty} \frac{\partial \ln f_{Y\theta}(y \mid \theta)}{\partial \theta} f_{Y\theta}(y \mid \theta) dy = E\left[\frac{\partial \ln f_{Y\theta}(y \mid \theta)}{\partial \theta}\right] = 0
$$

(6.80)

The covariance matrix of $X$ is then

$$
C_{XX} = E[XX^T] = \begin{bmatrix}
\sigma^2_{\hat{\theta}_1} & 1 & 0 & \cdots & 0 \\
1 & J_{11} & J_{12} & \cdots & J_{1K} \\
0 & J_{21} & J_{22} & \cdots & J_{2K} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & J_{K1} & J_{K2} & \cdots & J_{KK}
\end{bmatrix}
$$

(6.81)

or in partitioned form,

$$
C_{XX} = \begin{bmatrix}
\sigma^2_{\hat{\theta}_1} & 1 & 0 & \cdots & 0 \\
1 & J \\
0 & \vdots \\
0 & \vdots \\
0 & 0
\end{bmatrix}
$$

(6.82)
Since the covariance matrix is nonnegative definite, and consequently its determinant is nonnegative definite, the determinant of (6.81) is given by

$$\det(C_{xx}) = \sigma_0^2 |J| - \begin{vmatrix} J_{12} & \cdots & J_{1K} \\ \vdots & \ddots & \vdots \\ J_{K2} & \cdots & J_{KK} \end{vmatrix}$$

From (4.30), we observe that (6.83) can be written in terms of the cofactor $J_{11}$. Hence,

$$\det(C_{xx}) = \sigma_0^2 |J| - \text{cofactor} J_{11} \geq 0 \quad (6.84)$$

Assuming that the Fisher matrix $J$ is nonsingular, we have

$$C_{xx} = E[XX^T] = \sigma_0^2 |J| - \text{cofactor} J_{11} \geq 0 \quad (6.85)$$

or

$$\sigma_0^2 \geq \frac{\text{cofactor} J_{11}}{|J|} = J^{\mu} \quad (6.86)$$

which is the desired result given in (6.72).

### 6.7.2 Random Vector

In the Bayes’ estimation, we minimize the cost function $C(\theta, \hat{\theta}(y))$. Consider now the extension of the mean-square error criterion and the MAP criterion for multiple parameters estimation.

**Mean-Square Estimation**

In this case, the cost function is the sum of the squares of the error samples given by
Parameter Estimation

\[ C(\hat{\theta}(y)) = c(\hat{\theta}(y) - \theta) = \sum_{i=1}^{K} [\hat{\theta}_i(y) - \theta_i(y)]^2 = \sum_{i=1}^{K} \hat{\theta}_i(y)^2 = \bar{\theta}^T \bar{\theta}(y) \]  

(6.87)

The risk is

\[ \mathcal{R}_{ms} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\theta, \phi}(y, \theta, \phi) d\theta d\phi \]  

(6.88)

Substituting (6.87) in (6.88) and using the Bayes' rule, the risk becomes

\[ \mathcal{R}_{ms} = \int_{-\infty}^{\infty} f_{\theta}(y) d\theta \int_{-\infty}^{\infty} \left\{ \sum_{i=1}^{K} [\hat{\theta}_i(y) - \theta_i(y)]^2 \right\} f_{\theta|Y}(\theta|y) d\theta \]  

(6.89)

As before, minimizing the risk is equivalent to minimizing the expression in the brackets of (6.89). Each term between the brackets is positive, and thus the minimization is done term-by-term. From (6.19), the \( i \)th term \( \hat{\theta}_i(y) \) is minimized for

\[ \hat{\theta}_{ms}(y) = \int_{-\infty}^{\infty} f_{\theta|Y}(\theta|y) d\theta \]  

(6.90)

In vector form, the MMSE is given by

\[ \hat{\theta}_{ms} = E[\theta|y] = \int_{-\infty}^{\infty} f_{\theta|Y}(\theta|y) d\theta \]  

(6.91)

It can be shown that the mean-square estimation commutes over a \textit{linear transformation} to yield

\[ \hat{\phi}_{ms}(y) = D \hat{\theta}_{ms}(y) \]  

(6.92)

where \( D \) is an \( L \times K \) matrix.

\textit{MAP Estimation}

From (6.28), the MAP estimate \( \hat{\theta}_{map} \) is obtained by minimizing \( f_{\theta|Y}(\theta|y) \).

Generalizing the result to the estimation of multiple parameters estimation, we obtain the following set of MAP equations:
Using (6.59), the MAP equation can be written in a single vector to be

$$\nabla_\theta [\ln f_{\theta|y}(\theta | y)]\big|_{\hat{\theta} = \hat{\theta}_{\text{MAP}}(y)} = 0, \quad i = 1, 2, \ldots, K$$

(6.94)

\textit{Cramer-Rao Bound}

The covariance matrix of the error of any unbiased estimator $\hat{\theta}$ of $\theta$ is bounded below by the inverse of the Fisher information matrix, $L$, and is given by

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \geq \frac{1}{L}$$

(6.95)

where

$$L = -E \left[ \frac{\partial^2}{\partial \theta^2} \ln f_{Y|\theta}(y, \theta) \right]$$

(6.96)

Note that the equality holds if and only if

$$\left[ \frac{\partial}{\partial \theta} \ln f_{Y|\theta}(y, \theta) \right]^T = c(\hat{\theta} - \theta)$$

(6.97)

where $c$ is independent of $\theta$ and $Y$. If the conditional density function $f_{Y|\theta}(y | \theta)$ is Gaussian, the lower bound of (6.95) is achieved with equality.

The information matrix $L$ can be written in terms of $J$ as

$$L = J - E \left[ \frac{\partial^2}{\partial \theta^2} \ln f_{\theta}(\theta) \right]$$

(6.98)

6.8 \textbf{BEST LINEAR UNBIASED ESTIMATOR}

In many practical problems, it may be not possible to determine the MMSE estimators of a random or an unknown parameter, even if it exists. For example, we do not know the probability density function of the data, but we know the first-
order and second-order moments of it. In this case, the methods developed in estimating the parameters and determining the Cramer-Rao lower bound cannot be applied. However, we still would like to obtain a reasonable (suboptimum) or “best” estimator, in the sense that it is unbiased and has a minimum variance, usually called MVU estimator. To do so, we limit the estimator to be a linear function of the data, and thus it becomes possible to obtain an explicit expression for the best linear unbiased estimator (BLUE).

We first give the one parameter linear minimum mean-square estimation to present the fundamental concepts, and then generalize them to multiple parameters.

### 6.8.1 One Parameter Linear Mean-Square Estimation

The linear minimum-square estimate of a random parameter $\theta$ is given by

$$\hat{\theta}_{lms} = aY + b \quad (6.99)$$

The corresponding risk function is

$$\mathcal{R}_{lms} = E[C(\theta, \hat{\theta})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f_{\theta, Y}(\theta, y) d\theta dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - ay - b)^2 f_{\theta, Y}(\theta, y) d\theta dy \quad (6.100)$$

Following the same procedure as we did in Section 6.5.1, we observe that minimizing the risk involves finding the constants $a$ and $b$, so that $\mathcal{R}_{lms}$ is minimum. Hence, taking the derivatives of $\mathcal{R}_{lms}$ with respect to $a$ and $b$ and setting them equal to zero, we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - ay - b) f_{\theta, Y}(\theta, y) d\theta dy = 0 \quad (6.101)$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - ay - b) f_{\theta, Y}(\theta, y) d\theta dy = 0 \quad (6.102)$$

Using (1.45) and (1.108), (6.101) and (6.102) can be rewritten as

$$aE[Y^2] + bE[Y] = E[\theta Y]$$

$$aE[Y^2] + bE[Y] = E[\theta Y] \quad (6.103)$$
and
\[ aE[Y] + b = E[\theta] \] (6.104)

We have two equations in two unknowns. Solving for \(a\) and \(b\), we obtain

and

Knowing that the correlation coefficient \(\rho_{\theta Y}\) is given by
\[ \rho_{\theta Y} = \frac{E[(\hat{\theta} - m_\theta)(Y - m_y)]}{\sigma_\theta \sigma_y} \] (6.107)

with \(m_\theta = E[\theta]\), \(m_y = E[Y]\), \(\sigma_\theta = \sqrt{E[(\theta - m_\theta)^2]}\), and \(\sigma_y = \sqrt{E[(Y - m_y)^2]}\). Then,
\[ a = \rho_{\theta Y} \frac{\sigma_\theta}{\sigma_y} \] (6.108)

and
\[ b = m_\theta - \rho_{\theta Y} m_y \frac{\sigma_\theta}{\sigma_y} \] (6.109)

The optimal cost function can be obtained to be
\[ \mathcal{R}_{\text{inv}} = \sigma_\theta^2 (1 - \rho_{\theta Y}^2) \] (6.110)

It can be shown that if the joint density function \(f_{Y,\theta}(y, \theta)\) is Gaussian, then the conditional mean \(E[\theta | y]\) is linear in the observation data, and thus the minimum mean-square estimate is linear. In addition, we usually assume for
convenience that the parameter \( \theta \) and the observation \( Y \) have zero means. In this case, \( \hat{\theta}_{\text{blue}} \) is unbiased, and is given by

\[
\hat{\theta}_{\text{blue}} = C_\theta C_{\theta Y}^{-1} Y
\]  

(6.111)

where \( C_{\theta Y} = E[\theta Y] \) and \( C_{\theta Y}^{-1} = 1 / E[Y^2] \). We now can generalize the result of (6.111) for multiple parameter estimation.

### 6.8.2 \( \theta \) Random Vector

If now \( \theta \) is a random vector parameter and \( \theta \) and \( Y \) are assumed to have zero means, then it can be shown that the BLUE that minimizes the mean-square error (variance minimum) is given by

\[
\hat{\theta}_{\text{blue}} = C_{\theta Y} C_{YY}^{-1} Y
\]  

(6.112)

and the mean-square error is

\[
E[(\theta - \hat{\theta}_{\text{blue}})(\theta - \hat{\theta}_{\text{blue}})^T] = C_{\theta \theta} - C_{\theta Y} C_{YY}^{-1} C_{Y \theta}
\]  

(6.113)

\( C_{YY} \) is the covariance matrix of the observation vector \( Y \), \( C_{\theta Y}^{-1} \) is its inverse, and \( C_{\theta Y} \) is the cross-covariance matrix between \( Y \) and \( \theta \). Note that the mean and covariance of the data are unknown, and the means of \( Y \) and \( \theta \) are assumed to be zero, and thus the linear mean-square estimator is unbiased.

**Proof.** We now give a derivation of the result given in (6.112). Since \( \hat{\theta} \) is restricted to be a linear estimator for \( Y \), that is a linear function of the data, then \( \hat{\theta} \) can be written as

\[
\hat{\theta} = DY
\]  

(6.114)

The problem is to select the matrix \( D \) so that the mean-square given by (6.113) is minimized. Equation (6.113) is called the matrix-valued squared error loss function. Substituting (6.114) into (6.113), we have

\[
E[(\theta - \hat{\theta})(\theta - \hat{\theta})^T] = E[(\theta - DY)(\theta - DY)^T]
\]  

(6.115)

Using the fact that
then, (6.115) becomes

\[
E[ (\theta - \hat{\theta})(\theta - \hat{\theta})^T ] = \text{tr} E[ (\theta - \hat{\theta})(\theta - \hat{\theta})^T ] \tag{6.116}
\]

Note that

\[
(\mathbf{D} - \mathbf{C}_{0\mathbf{Y}}) \mathbf{C}_{\mathbf{Y}Y}^{-1} (\mathbf{D} - \mathbf{C}_{0\mathbf{Y}}) \mathbf{C}_{\mathbf{Y}Y}^{-1} = D_{\mathbf{C}_{\mathbf{Y}Y}}^{-1} D_{\mathbf{C}_{\mathbf{Y}Y}}^{-1} - D_{\mathbf{C}_{\mathbf{Y}Y}}^{-1} + \mathbf{C}_{\mathbf{C}_{\mathbf{Y}Y}}^{-1} \mathbf{C}_{\mathbf{Y}Y} \mathbf{C}_{\theta} \tag{6.117}
\]

Using (6.118), we can write

\[
E[ (\theta - \hat{\theta})(\theta - \hat{\theta})^T ] = \text{tr} \left[ \mathbf{C}_{\mathbf{Y}Y} + (\mathbf{D} - \mathbf{C}_{0\mathbf{Y}}) \mathbf{C}_{\mathbf{Y}Y}^{-1} (\mathbf{D} - \mathbf{C}_{0\mathbf{Y}}) \mathbf{C}_{\mathbf{Y}Y}^{-1} - \mathbf{C}_{0\mathbf{Y}} \mathbf{C}_{\mathbf{Y}Y} \mathbf{C}_{\theta} \right] \tag{6.119}
\]

We observe that the gain matrix \( \mathbf{D} \) appears only in the second term on the righthand side of (6.119). Thus, each diagonal element in the matrix \( E[ (\theta - \hat{\theta})(\theta - \hat{\theta})^T ] \) is minimized when \( \mathbf{D} \) is given by

\[
\mathbf{D} = \mathbf{C}_{0\mathbf{Y}} \mathbf{C}_{\mathbf{Y}Y}^{-1} \tag{6.120}
\]

Substituting (6.120) in (6.114), we have

\[
\hat{\theta} = \hat{\theta}_{\text{blue}} = \mathbf{C}_{0\mathbf{Y}} \mathbf{C}_{\mathbf{Y}Y}^{-1} \mathbf{Y} \tag{6.121}
\]

and the proof is complete.

Note that if \( \mathbf{Y} \) and \( \theta \) are not zero mean, such that \( E[\mathbf{Y}] = \mathbf{m}_Y \) and \( E[\theta] = \mathbf{m}_\theta \), then

\[
\hat{\theta}_{\text{ms}} = \mathbf{A}\mathbf{Y} + \mathbf{b} \tag{6.122}
\]

where the matrix \( \mathbf{A} \) and the vector \( \mathbf{b} \) are given by

\[
\mathbf{A} = \left( E[\mathbf{Y}\mathbf{Y}^T] - E[\mathbf{Y}]E[\mathbf{Y}^T] \right)^{-1} \left[ E[\theta\mathbf{Y}^T] - E[\theta]E[\mathbf{Y}^T] \right] = \mathbf{C}_{\mathbf{Y}Y}^{-1} \mathbf{C}_{0\mathbf{Y}} \tag{6.123}
\]
and

\[ b = E[\theta] - AE[Y] \]  \hspace{1cm} (6.124)

By direct substitution, we obtain

\[ \hat{\theta}_{blue} = m_\theta + C_{\theta Y} C_{YY}^{-1} (Y - m_Y) \]  \hspace{1cm} (6.125)

The BLUE given in (6.121) has several properties of interest:

\[ E[\hat{\theta}_{blue} Y^T] = C_{\theta Y} \]  \hspace{1cm} (6.126)

\[ E[\hat{\theta}_{blue} \hat{\theta}_{blue}^T] = C_{\theta Y} C_{YY}^{-1} C_{\theta} = C_{\theta_{blue} \theta_{blue}} \]  \hspace{1cm} (6.127)

\[ E[(\theta - \hat{\theta}_{blue})(\theta - \hat{\theta}_{blue})^T] = C_{\theta} - C_{\theta_{blue} \theta_{blue}} \]  \hspace{1cm} (6.128)

\[ E[(\theta - \hat{\theta}_{blue})Y^T] = 0 \]  \hspace{1cm} (6.129)

\[ E[(\theta - \hat{\theta}_{blue})\hat{\theta}_{blue}^T] = 0 \]  \hspace{1cm} (6.130)

We observe that property (6.129) means that the error in the estimate is orthogonal to the data \( Y \), while property (6.130) means that the error in the estimate is orthogonal to the estimator \( \hat{\theta}_{blue} \). This concept of orthogonality is an important result, which will be developed and used extensively in the next chapter on filtering.

### 6.8.3 BLUE in White Gaussian Noise

Consider the general problem of estimating a random vector with \( N \) parameters (denoted as the \( N \)-dimensional vectors \( \theta \)), to be estimated from \( K \) observations (denoted as the \( K \)-dimensional vector \( Y \)), in white Gaussian noise. The parameters \( \theta \) and measurements \( Y \) are assumed to be related by the so-called linear model

\[ Y = H \theta + N \]  \hspace{1cm} (6.131)

\( H \) is a \( K \times N \) known mapping matrix, \( Y \) is the \( K \times 1 \) observed random vector, \( \theta \) is an \( N \times 1 \) random vector to be estimated, and \( N \) is a \( K \times 1 \) vector representing errors in the measurement (noise). Assuming that \( \theta \) and \( N \) have zero means, then \( Y \) has zero mean. The covariance matrix of \( Y \) is
Signal Detection and Estimation

\[ C_{YY} = [(H \theta + N)(H \theta + N)^T] = HC_{\theta\theta}H^T + HC_{\theta N} + C_{\theta N}H^T + C_{NN} \]  \hspace{1cm} (6.132)

while the cross-covariance matrix of \( Y \) and \( \theta \) is

\[ C_{Y \theta} = HC_{\theta\theta} + C_{\theta N} \]  \hspace{1cm} (6.133)

Substituting (6.132) and (6.133) in (6.121), we obtain the BLUE estimate of \( \theta \) to be

\[ \hat{\theta}_{\text{blue}} = \left[ C_{\theta\theta}H^T + C_{\theta N} \right] \left[ HC_{\theta\theta}H^T + HC_{\theta N} + C_{\theta N}H^T + C_{NN} \right]^{-1} Y \]  \hspace{1cm} (6.134)

with error covariance matrix

\[ C_{\hat{\theta}\hat{\theta}} = C_{\theta\theta} - \left( C_{\theta\theta}H^T + C_{\theta N} \right) \left( HC_{\theta\theta}H^T + HC_{\theta N} + C_{\theta N}H^T + C_{NN} \right)^{-1} \left( HC_{\theta\theta} + C_{\theta N} \right) \]  \hspace{1cm} (6.135)

When \( \theta \) and \( N \) are uncorrelated, which is the usual assumed case, \( C_{\theta N} = 0 \), and the BLUE of \( \theta \) reduces to

\[ \hat{\theta} = C_{\theta\theta}H^T (HC_{\theta\theta}H^T + C_{NN})^{-1} Y \]  \hspace{1cm} (6.136)

while the error matrix becomes

\[ C_{\hat{\theta}\hat{\theta}} = C_{\theta\theta} - C_{\theta\theta}H^T (HC_{\theta\theta}H^T + C_{NN})^{-1} HC_{\theta\theta} \]  \hspace{1cm} (6.137)

Using the matrix inversion lemma given in Chapter 4, and after some matrix operation, we have

\[ \tilde{\theta}_{\text{blue}} = C_{\theta\theta}H^T C_{NN}^{-1} Y \]  \hspace{1cm} (6.138a)

where

\[ C_{\tilde{\theta}\tilde{\theta}} = (C_{\theta\theta} + H^T C_{NN}^{-1} H)^{-1} \]  \hspace{1cm} (6.138b)

If no a priori information about \( \theta \) is available, and thus if \( C_{\theta\theta}^{-1} \) is assumed zero, the BLUE of \( \hat{\theta} \) is given by
Note that in these results, we only assumed that $\theta$ is a random parameter. Consider now the problem of estimating the unknown vector $\theta$, but which is constrained to be a linear function of the data (measurements).

**The Estimator as a Linear Function of Data**

In this case, we require

$$\hat{\theta} = \sum_{k=1}^{K} a_{ik} Y_k + b_k, \quad i = 1, 2, \ldots, M$$

(6.140)

or, in matrix form

$$\hat{\theta} = AY + b$$

(6.141)

where $A$ is a $M \times K$ matrix, and $Y$ and $b$ are $K \times 1$ vectors. In order for $\hat{\theta}$ to be unbiased, we must have

$$E[\hat{\theta} \mid 0] = 0$$

(6.142)

Hence,

$$E[AY + b \mid 0] = AE[Y \mid 0] + b = AE[H \theta + N \mid 0] + b = AH \theta + b = 0$$

(6.143)

only if

$$AH = I$$

(6.144a)

and

$$b = 0$$

(6.144b)

The BLUE estimate is then given by

$$\hat{\theta} = (H^T C_{NN}^{-1} H)^{-1} H^T C_{NN}^{-1} Y$$

(6.145)

Therefore, with the noise Gaussian in the linear model, we can state the following result given by the Gauss-Markov theorem.
Gauss-Markov Theorem. If the data is of the general linear model form

$$Y = H \theta + N$$

(6.146)

where $H$ is a known $K \times M$ matrix, $\theta$ is an $M \times 1$ vector of parameters to be estimated, and $N$ is a $K \times 1$ noise vector with mean zero and covariance matrix $C_{NN}$, then the BLUE of $\theta$ that minimizes the mean-square error is

$$\hat{\theta} = (H^T C_{NN}^{-1} H)^{-1} H^T C_{NN}^{-1} Y$$

(6.147)

with error covariance matrix

$$C_{\hat{\theta}\hat{\theta}} = E[(\theta - \hat{\theta}_{blue})(\theta - \hat{\theta}_{blue})^T | \theta] = (H^T C_{NN}^{-1} H)^{-1}$$

(6.148)

The minimum variance of $\hat{\theta}_k$ is then

$$\text{var}(\hat{\theta}_k) = [(H^T C_{NN}^{-1} H)^{-1}]_{kk}$$

(6.149)

Example 6.7

Consider the problem of Example 6.2 where

$$Y_k = A + N_k, \quad k = 1, 2, \ldots, K$$

where $N_k$ is a zero mean white noise. Find the BLUE of $M$ if:

(a) The variance of $N_k$, $k = 1, 2, \ldots, K$ is $\sigma^2$.

(b) The noise components are correlated with variance $\sigma_k^2$, $k = 1, 2, \ldots, K$.

Solution

(a) The estimator is constrained to be a linear function of the data. Let

$$\hat{A}_k = \sum_{k=1}^{K} A_{jk} Y_k, \quad j = 1, 2, \ldots, M$$

where the $A_{jk}$s are the weighting coefficients to be determined. From (6.147), the BLUE is given by

$$\hat{A} = (H^T C_{NN}^{-1} H)^{-1} H^T C_{NN}^{-1} Y$$
where

\[ E[Y_k] = E[H_k A_k + N_k] = H_k E[A_k] \]

Since \( A_k \) must be unbiased, then \( E[A_k] = A_k \), \( H_k = 1 \), and thus \( H = 1 \). Substituting, we have

\[ \hat{A} = \left( I^T \frac{1}{\sigma^2} I \right)^{-1} \left( I^T \frac{1}{\sigma^2} y \right) = (\sigma^2 K)^{-1} \left( \sigma^2 \sum_{k=1}^{K} y_k \right) = \frac{1}{K} \sum_{k=1}^{K} y_k \]

Hence, we observe that the BLUE is the sample mean independently of the probability density function of the data, while the minimum variance is

\[ \text{var}(\hat{A}) = \frac{1}{(H^T C_{NN}^{-1} H)} = \frac{1}{1^T \frac{1}{\sigma^2} I 1} = \frac{\sigma^2}{K} \]

(b) In this case, the variance matrix is

\[ C_{NN} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_K^2 \end{bmatrix} \]

After substitution, the BLUE is

\[ \hat{A} = \frac{\sum_{k=1}^{K} \frac{1}{\sigma_k^2} y_k}{\sum_{k=0}^{K} \frac{1}{\sigma_k^2}} \]

while the minimum variance is

\[ \text{var}(\hat{A}) = \frac{1}{\sum_{k=0}^{K} \frac{1}{\sigma_k^2}} \]
6.9 LEAST-SQUARE ESTIMATION

In studying parameter estimation in the previous sections, our criteria were to find a “good” estimator that was unbiased and had minimum variance. In the least-square estimation, the criterion is only to minimize the squared difference between the given data (signal plus noise) and the assumed signal data.

Suppose we want to estimate \( M \) parameters, denoting the \( M \)-dimensional vector \( \theta \), from the \( K \) measurements, denoting the \( K \)-dimensional vector \( Y \) with \( K \geq M \). The relation between the parameters \( \theta \) and the observed data \( Y \) is given by the linear model

\[
Y = H\theta + N
\]

where \( H \) is a known \((K \times M)\) matrix, and \( N \) is the unknown \((K \times 1)\) error vector that occurs in the measurement of \( \theta \).

The least-square estimator (LSE) of \( \theta \) chooses the values that make \( X = H\theta \) closest to the observed data \( Y \). Hence, we minimize

\[
J(\theta) = \sum_{k=1}^{K} (Y_k - X_k)^2 = (Y - H\theta)^T (Y - H\theta) = YY^T - Y^T H \theta - \theta^T H^T Y + \theta^T H^T H \theta
\]

\[
= YY^T - 2Y^T H \theta + \theta^T H^T H \theta
\]

Note that \( Y^T H \theta \) is a scalar. Taking the first-order partial derivative of the cost function \( J(\theta) \) with respect to \( \theta \) (i.e., the gradient) and setting it equal to zero, we obtain the set of linear equations

\[
\frac{\partial J(\theta)}{\partial \theta} = -2HY + 2H^T H \theta = 0
\]

and the LSE is found to be

\[
\hat{\theta}_{ls} = (H^T H)^{-1} H^T Y
\]

Note that the second-order partial derivative is

\[
\frac{\partial^2 J(\theta)}{\partial \theta^2} = H^T H
\]
This matrix is positive-definite as long as $H$ is assumed to be of full rank to guarantee the inversion of $H^T H$. Thus, the solution (6.153) is unique and minimizes $J(\theta)$. The equations

$$H^T H \hat{\theta} = H^T Y$$

(6.155)

to be solved for $\hat{\theta} = \hat{\theta}_{ls}$ are referred to as the normal equations.

We observe that the error in the estimator $\hat{\theta}_{ls}$ is a linear function of the measurement errors $N$, since

$$\tilde{\theta}_{ls} \triangleq \theta - \hat{\theta}_{ls} = \theta - \left( H^T H \right)^{-1} H^T Y = \theta - \left( H^T H \right)^{-1} H^T [H \theta + N]$$

$$= \theta - \left( H^T H \right)^{-1} H^T H \theta - \left( H^T H \right)^{-1} H^T N$$

$$= -\left( H^T H \right)^{-1} H^T N$$

(6.156)

The minimum least-square $J_{min}$ can be shown, after some matrix operation, to be

$$J_{min} = J(\hat{\theta}_{ls}) = (Y - H \hat{\theta})^T (Y - H \hat{\theta}) = Y^T - Y^T H (H^T H)^{-1} H^T Y$$

$$= Y^T (Y - H \hat{\theta})$$

(6.157)

**Generalization of the Least-Square Problem**

The least-square cost function can be generalized by introducing a $K \times K$ positive definite weighting matrix $W$ to yield

$$J(\theta) = (Y - H \theta)^T W (Y - H \theta)$$

(6.158)

The elements of the weighting can be chosen to emphasize specific values of the data that are more reliable for the estimate $\hat{\theta}$.

The general form of the least-square estimator can be shown to be

$$\hat{\theta} = (H^T WH)^{-1} H^T W Y$$

(6.159)

while its minimum least-square error is

$$J_{min} = Y^T [W - WH (H^T WH)^{-1} H^T W] Y$$

(6.160)

The error covariance matrix becomes
\[ C_{NN} = (H^T W H)^{-1} H^T W R W H (H^T W H)^{-1} \]  
\[ (6.161) \]
where \( R_{NN} \) is a known positive-definite covariance matrix given by
\[
R_{NN} = E[N N^T] \tag{6.162}
\]
since \( E[N] = 0 \) (i.e., \( R_{NN} = C_{NN} \)).

If the measurement errors \( N \) are uncorrelated and have identical variance \( \sigma^2 \), then \( R = \sigma^2 I \); and if \( W = \sigma^2 I \), then (6.159) reduces to (6.153). That is, a constant scaling has no effect on the estimate.

It can also be shown that the least-square estimator and the linear minimum mean-square estimator are identical when the weighting matrix \( W \) is chosen as
\[
W = R^{-1} \tag{6.163}
\]
that is, the inverse of the measurement noise covariance matrix.

**Example 6.8**

Consider again the problem of Example 6.5 with \( y_k = A + N_k, k = 1, 2, \ldots, K \).

From (6.153), the least-square estimate is \( \hat{A} = (H^T H)^{-1} H^T Y \). \( H \) is the \((K \times 1)\) column matrix denoted \( 1^T = [1 \ 1 \ \ldots \ 1] \). Hence,
\[
\hat{A} = (1^T 1)^{-1} 1^T Y = \frac{1}{K} \sum_{k=1}^{K} y_k
\]
which is the sample mean. Observe that for this simple operation, instead of applying a derived result, we could have started by writing the least-square cost function \( J(A) = \sum_{k=1}^{K} (y_k - A)^2 \), then differentiating \( J(A) \) with respect to \( A \), setting the result equal to zero, and solving for \( \hat{A} = \hat{A}_0 \).

**Example 6.9**

Suppose that three measurements of signal \( s_k = \theta \exp(k/2) \), where \( \theta \) is the parameter to be estimated, are given by \( y_1 = 1.5 \), \( y_2 = 3 \), and \( y_3 = 5 \). Find the least-square estimate of \( \theta \).
Parameter Estimation

Solution

The data can be put in the form, \( Y = H\theta + N \) given by (6.150). Substituting for the values of \( k \), we have

\[
\begin{align*}
1.5 &= 1.648\theta + N_1 \\
3 &= 2.718\theta + N_2 \\
5 &= 4.820\theta + N_3
\end{align*}
\]

where \( y = [1.5 \ 3 \ 5]^T \) is a realization of \( Y \), \( H = [1.648 \ 2.718 \ 4.482]^T \), and \( N = [N_1 \ N_2 \ N_3] \) a realization of \( N \). The least-square estimate is given by

\[
\hat{\theta}_{ls} = (H^TH)^{-1}H^Ty
\]

where \( H^TH = \sum_{k=1}^{3} H_k^2 = 30.192 \), and \( H^Ty = \sum_{k=1}^{3} H_k Y_k = 30.036 \). Hence,

\[
\hat{\theta}_{ls} = (H^TH)^{-1}H^Ty = \frac{\sum_{k=1}^{3} H_k Y_k}{\sum_{k=1}^{3} H_k^2} = 0.995
\]

6.10 RECURSIVE LEAST-SQUARE ESTIMATOR

In real time estimation problems (filtering), it is necessary to write the estimator \( \hat{\theta} \) in a recursive form for efficiency. For example, consider a situation where an estimate \( \hat{\theta} \) is determined based on some data \( Y_K \). If new data \( Y_{K+1} \) is to be processed after having determined an estimate based on the data \( Y_K \), it is best to use the old solution along with the new data to determine the new least-square estimator. It is clear that discarding the estimate based on the data \( Y_K \) and restarting the computation for a solution is inefficient. This procedure of determining the least-square estimate from an estimate based on \( Y_K \) and the new data \( Y_{K+1} \) is referred to as sequential least-square estimation, or more commonly recursive least-square (RLS) estimation.

Consider the problem of estimating \( \theta \) from the data vectors \( Z_M \) given by the linear model
\[ Z_M = H_M \theta + U_M \]  \hspace{1cm} (6.164a)

where

\[ Z_M = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_M \end{bmatrix}^T \]  \hspace{1cm} (6.164b)

is an \((MK+1)\) collection of vectors \(Y_1, Y_2, \ldots, Y_M\), since each vector \(Y_k, k = 1, 2, \ldots, M\), is a \((K+1)\) vector,

\[ U_M = \begin{bmatrix} N_1 & N_2 & \cdots & N_M \end{bmatrix}^T \]  \hspace{1cm} (6.164c)

is an \((MK+1)\) error vector, and

\[ H_M = \begin{bmatrix} h_1 & h_2 & \cdots & h_M \end{bmatrix}^T \]  \hspace{1cm} (6.164d)

is an \((MK \times n)\) mapping matrix relating \(Z_M\) to the \((n \times 1)\) parameter vector \(\theta\) to be estimated.

It can be shown that the RLS estimator is given by

\[ \hat{\theta}_M = \hat{\theta}_{M-1} + V_M [U_M - H_M \hat{\theta}_{M-1}] \]  \hspace{1cm} (6.165)

where

\[ V_M = C_{UU} H_M^T R_{MM}^{-1} \]  \hspace{1cm} (6.166)

\(C\) is the error covariance matrix given by

\[ C_{UU} = E[U_M U_M^T] = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1M} \\ R_{21} & R_{22} & \cdots & R_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1M}^T & R_{2M}^T & \cdots & R_{MM}^T \end{bmatrix} \]  \hspace{1cm} (6.167)

and

\[ E[N_i N_j^T] = R_i \delta_{ij} \]  \hspace{1cm} (6.168)
The covariance matrix of the individual noise vector \( N \) is \( R_d \hat{=} R_r \). Equation (6.170) indicates that the estimator \( \hat{\theta}_M \) based on \( Z_M \) is formed as a linear combination of \( \hat{\theta}_{M-1} \) and a correction term \( V_M[U_M - H_M \hat{\theta}_{M-1}] \).

If \( \theta \) were a random variable, it can be shown that the generalization of the recursive least-square estimation leads to the Kalman filter [3]. In the next chapter on filtering, we present an introduction to Kalman filtering.

6.11 SUMMARY

In this chapter, we have developed the concept of parameter estimation. We used the maximum likelihood estimation to estimate nonrandom parameters. We first obtained the likelihood function in terms of the parameters to be estimated. Then, we maximized the likelihood function to obtain the estimator, which resulted from solving the likelihood equation. We linked this chapter to the previous one by presenting the generalized likelihood ratio test in Section 6.3. In the generalized likelihood ratio test, we used the maximum likelihood estimate of the unknown parameter in the composite hypothesis as its true value and then performed the likelihood ratio test. This was an alternative to the case where UMP tests did not exist. Measuring criteria for the estimator, such as bias and consistency, were presented to determine the quality of the estimator.

When the parameter to be estimated was a random variable, we used Bayes’ estimation. In Bayes’ estimation, we minimized the risk, which is a function of error between the estimate and the true value. Three cases were considered; the squared error, the absolute value error, and the uniform cost function. It was shown that the minimum mean-square error represents the conditional mean of the parameter (associated with the observation random variable) to be estimated. The resulting minimum risk was the conditional variance. In the absolute value error case, the estimate turned out to be the median of the conditional density function of the parameter to be estimated, given the observation random variable.

For the uniform Bayes’ cost, the estimator was actually the solution of the MAP equation. In comparing the ML estimate and MAP estimate, it was observed that the ML estimate was a special case of the MAP estimate and is obtained by setting to zero the density function of the parameter to be estimated in the MAP equation. In order to measure the “goodness” of the estimator, the Cramer-Rao bound was given as an alternate way to measure the error variance, since an expression for the error variance was difficult to obtain. The above results were generalized to multiple parameter estimation in Section 6.7.

Then, we presented linear mean-square estimation for situations where it may have been difficult to find the MMSE, even if existed. We defined the BLUE in the sense that the mean-square value is minimized. We verified that for a joint Gaussian density function of the observation and the parameter to be estimated, the linear mean-square estimator is the optimum MMSE. An introduction to least-square estimation was presented. We noted that least-square estimation was not...
based on the criteria of the unbiased and minimum variance estimator, but rather on minimizing the squared difference between the given data and the assumed signal data. We concluded the chapter with a brief section on recursive least-square estimation.

**PROBLEMS**

6.1 Let $Y_1, Y_2, \ldots, Y_K$ be the observed random variables, such that

$$Y_k = a + b x_k + Z_k, \quad k = 1, 2, \ldots, K$$

The constants $x_k, k = 1, 2, \ldots, K$, are known, while the constants $a$ and $b$ are not known. The random variables $Z_k, k = 1, 2, \ldots, K$, are statistically independent, each with zero mean and variance $\sigma^2$ known. Obtain the ML estimate of $(a, b)$.

6.2 Let $Y$ be a Gaussian random variable with mean zero and variance $\sigma^2$.

(a) Obtain the ML estimates of $\sigma$ and $\sigma^2$.

(b) Are the estimates efficient?

6.3 Let $Y_1$ and $Y_2$ be two statistically independent Gaussian random variables, such that $E[Y_1] = m, E[Y_2] = 3m$, and $\text{var}[Y_1] = \text{var}[Y_2] = 1$; $m$ is unknown.

(a) Obtain the ML estimates of $m$.

(b) If the estimator of $m$ is of the form $a_1 Y_1 + a_2 Y_2$, determine $a_1$ and $a_2$, so that the estimator is unbiased.

6.4 The observation sample of the envelope of a received signal is given by the following exponential distribution

$$f_{Y_k}(y_k) = \frac{1}{\theta} \exp\left(-\frac{y_k}{\theta}\right), \quad k = 1, 2, \ldots, K$$

$\theta$ is an unknown parameter and the observations are statistically independent.

(a) Obtain the ML estimate of $\theta$.

(b) Is the estimator unbiased?

(c) Determine the lower bound on the estimator.

(d) Is the estimator consistent?
6.5 Let the observation $Y$ satisfy the binomial law, such that the density function of $Y$ is

$$f_Y(y) = \binom{n}{k} p^k (1-p)^{n-k}$$

(a) Find an unbiased estimate for $p$.
(b) Is the estimate consistent?

6.6 Obtain the ML estimates of the mean $m$ and variance $\sigma^2$ for the independent observations $Y_1, Y_2, \ldots, Y_K$, such that

$$f_{Y_k}(y_k) = \frac{1}{\sqrt{2\pi \sigma}} \exp\left[-\frac{(y_k - m)^2}{2\sigma^2}\right], \quad k = 1, 2, \ldots, K$$

6.7 Let $x$ be an unknown deterministic parameter that can have any value in the interval $[-1, 1]$. Suppose we take two observations of $x$ with independent samples of zero-mean Gaussian noise, and with variance $\sigma^2$ superimposed on each of the observations.
(a) Obtain the ML estimate of $x$.
(b) Is $\hat{x}_{ml}$ unbiased?

6.8 Let $Y_1, Y_2, \ldots, Y_K$ be $K$ independent observed random variables, each having a Poisson distribution given by

$$f_{Y_k|\Theta}(y_k | \Theta) = e^{-\Theta} \frac{\Theta^{y_k}}{y_k!}, \quad y_k \geq 0, \quad k = 1, 2, \ldots, K.$$  

The parameter $\Theta$ is unknown.
(a) Obtain the ML estimate of $\Theta$.
(b) Verify that the estimator is unbiased and determine the lower bound.

6.9 Let $Y_1, Y_2, \ldots, Y_K$ be $K$ independent and identically distributed observations. The observations are uniformly distributed between $-\theta$ and $+\theta$, where $\theta$ is an unknown parameter to be estimated.
(a) Obtain the MLE of $\theta$.
(b) How is the estimator unbiased?
6.10 Let \( Y_1, Y_2, \ldots, Y_K \) be \( K \) independent variables with \( P(Y_k = 1) = p \) and \( P(Y_k = 0) = 1 - p \), where \( p, \ 0 \leq p < 1 \) is unknown.

(a) Obtain the ML estimate.
(b) Determine the lower bound on the variance of the estimator, assuming that the estimator is unbiased.

6.11 Find \( \hat{x}_{ms} \), the minimum mean-square error, and \( \hat{x}_{map} \), the maximum a posteriori estimators, of \( X \) from the observations \( Y = X + N \).

\( X \) and \( N \) are random variables with density functions

\[
f_X(x) = \frac{1}{2} [\delta(x-1) + \delta(x+1)] \quad \text{and} \quad f_N(n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n^2}{2\sigma^2}\right)
\]

6.12 The conditional density function of the observed random variable \( Y \) given a random parameter \( X \) is given by

\[
f_{Y|X}(y|x) = \begin{cases} xe^{-\alpha y}, & y \geq 0 \text{ and } x > 0 \\ 0, & y < 0 \end{cases}
\]

The a priori probability density function of \( X \) is

\[
f_X(x) = \begin{cases} \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x}, & x \geq 0 \\
0, & x < 0 \end{cases}
\]

where \( \alpha \) is a parameter, \( r \) is a positive integer, and \( \Gamma(r) \) is the gamma function.

(a) Obtain the a priori mean and variance of \( X \).
(b) For \( Y \) given,
   1. Obtain the minimum mean-square error estimate of \( X \).
   2. What is the variance of this estimate?
(c) Suppose we take \( K \) independent observations of \( Y_k, k = 1, 2, \ldots, K \), such that

\[
f_{Y_k|X}(y_k|x) = \begin{cases} xe^{-\alpha y_k}, & y_k \geq 0 \text{ and } x > 0 \\
0, & y_k < 0 \end{cases}
\]
1. Determine the minimum mean-square error estimate of $X$.
2. What is the variance of this estimate?
(d) Verify if the MAP estimate equals the MMSE estimate.

6.13 Consider the problem where the observation is given by $Y = \ln X + N$, where $X$ is the parameter to be estimated. $X$ is uniformly distributed over the interval $[0, 1]$, and $N$ has an exponential distribution given by

$$f_N(n) = \begin{cases} e^{-n}, & n \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Obtain
(a) The mean-square estimate, $\hat{x}_{ms}$.
(b) The MAP estimate, $\hat{x}_{map}$.
(c) The MAVE estimate, $\hat{x}_{mave}$.

6.14 The observation $Y$ is given by $Y = X + N$, where $X$ and $N$ are two random variables. $N$ is normal with mean one and variance $\sigma^2$, and $X$ is uniformly distributed over the interval $[0, 2]$. Determine the MAP estimate of the parameter $X$.

6.15 Show that the mean-square estimation $\hat{\theta}_{ms} = E[\theta \mid y]$ commutes over a linear transformation.

6.16 Suppose that the joint density function of the observation $Y$ and the parameter $\theta$ is Gaussian. The means $m_y$ and $m_\theta$ are assumed to be zero. $\theta$ can then be expressed as a linear form of the data. Determine an expression for the conditional density $f_{\theta \mid y}(\theta \mid y)$.

6.17 Consider the problem of estimating a parameter $\theta$ from one observation $Y$. Then, $Y = \theta + N$, where $\theta$ and the noise $N$ are statistically independent with

$$f_\theta(\theta) = \begin{cases} 1, & 0 \leq \theta \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad f_N(n) = \begin{cases} \frac{n}{2}, & 0 \leq n \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

Determine $\hat{\theta}_{blue}$, the best linear unbiased estimate of $\theta$. 

---

**Parameter Estimation** 397
References


Selected Bibliography


