13. In Problem 2(a), find MLE of $d(\theta) = \theta^2$ and its asymptotic distribution.

14. Let $X_1, X_2, \ldots, X_n$ be a random sample from some DF F on the real line. Suppose we observe $x_1, x_2, \ldots, x_n$ which are all different. Show that the MLE of F is F_n^*, the empirical DF of the sample.

15. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, 1)$. Suppose $\Theta = \{\mu \geq 0\}$. Find the MLE of μ.

16. Let $(X_1, X_2, \ldots, X_{k-1})$ have a multinomial distribution with parameters $n, p_1, \ldots, p_{k-1}$, $0 \leq p_1, p_2, \ldots, p_{k-1} \leq 1$, $\sum_1^{k-1} p_j \leq 1$, where n is known. Find the MLE of $(p_1, p_2, \ldots, p_{k-1})$.

17. Consider the one parameter exponential density introduced in Section 5.5 in its natural form with PDF

$$f_\theta(x) = \exp\{\eta T(x) + D(\eta) + S(x)\}.$$

(a) Show that the MGF of $T(X)$ is given by

$$M(t) = \exp\{D(\eta) - D(\eta + t)\}$$

for t in some neighborhood of the origin. Moreover, $E_\eta T(X) = -D'(\eta)$ and $\text{var}(T(X)) = -D''(\eta)$.

(b) If the equation $E_\eta T(X) = T(x)$ has a solution, it must be the unique MLE of η.

18. In Problem 1(b) show that the unique MLE of θ is consistent. Is it asymptotically normal?

8.8 BAYES AND MINIMAX ESTIMATION

In this section we consider the problem of point estimation in a decision-theoretic setting. We will consider here Bayes and minimax estimation.

Let $\{f_\theta : \theta \in \Theta\}$ be a family of PDFs (PMFs) and $X_1, X_2, \ldots, X_n$ be a sample from this distribution. Once the sample point $(x_1, x_2, \ldots, x_n)$ is observed, the statistician takes an *action* on the basis of these data. Let us denote by $\mathcal{A}$ the set of all *actions* or *decisions* open to the statistician.

Definition 1. A decision function δ is a statistic that takes values in $\mathcal{A}$, that is, δ is a Borel-measurable function that maps $\mathcal{R}_n$ into $\mathcal{A}$.

If $\mathbf{X} = \mathbf{x}$ is observed, the statistician takes action $\delta(\mathbf{X}) \in \mathcal{A}$.

Example 1. Let $\mathcal{A} = \{a_1, a_2\}$. Then any decision function δ partitions the space of values of $(X_1, \ldots, X_n)$, namely, $\mathcal{R}_n$, into a set C and its complement C^c, such that if $\mathbf{x} \in C$ we take action a_1, and if $\mathbf{x} \in C^c$ action a_2 is taken. This is the problem of testing hypotheses, which we will discuss in Chapter 9.

Example 2. Let $\mathcal{A} = \Theta$. In this case we face the problem of estimation.

Another element of decision theory is the specification of a *loss function*, which measures the loss incurred when we take a decision.

Definition 2. Let $\mathcal{A}$ be an arbitrary space of actions. A nonnegative function L that maps $\Theta \times \mathcal{A}$ into $\mathcal{R}$ is called a loss function.

The value $L(\theta, a)$ is the loss to the statistician if he takes action a when θ is the true parameter value. If we use the decision function $\delta(\mathbf{X})$ and loss function L and θ is the true parameter value, then the loss is the RV $L(\theta, \delta(\mathbf{X}))$. (As always, we will assume that L is a Borel-measurable function.)

Definition 3. Let $\mathcal{D}$ be a class of decision functions that map $\mathcal{R}_n$ into $\mathcal{A}$, and let L be a loss function on $\Theta \times \mathcal{A}$. The function R defined on $\Theta \times \mathcal{D}$ by

$$R(\theta, \delta) = E_\theta L(\theta, \delta(\mathbf{X})) \tag{1}$$

is known as the risk function associated with δ at θ.

Example 3. Let $\mathcal{A} = \Theta \subseteq \mathcal{R}$, $L(\theta, a) = |\theta - a|^2$. Then

$$R(\theta, \delta) = E_\theta L(\theta, \delta(X)) = E_\theta\{\delta(X) - \theta\}^2,$$

which is just the MSE. If we restrict attention to estimators that are unbiased, the risk is just the variance of the estimator.

The basic problem of decision theory is the following: Given a space of actions A, and a loss function $L(\theta, a)$, find a decision function δ in D such that the risk $R(\theta, \delta)$ is "minimum" in some sense for all $\theta \in \Theta$. We need first to specify some criterion for comparing the decision functions δ.

Definition 4. The principle of minimax is to choose $\delta^* \in \mathcal{D}$ so that

$$\max_\theta R(\theta, \delta^*) \le \max_\theta R(\theta, \delta) \tag{2}$$

for all δ in $\mathcal{D}$. Such a rule δ^*, if it exists, is called a minimax (decision) rule.

If the problem is one of estimation, that is, if $\mathcal{A} = \Theta$, we call δ^* satisfying (2) a *minimax estimator* of θ.

Example 4. Let $X \sim b(1, p)$, $p \in \Theta = \{\frac{1}{4}, \frac{1}{2}\}$, and $\mathcal{A} = \{a_1, a_2\}$. Let the loss function be defined as follows.

	a_1	a_2
$p_1 = \frac{1}{4}$	1	4
$p_2 = \frac{1}{2}$	3	2

The set of decision rules includes four functions: $\delta_1, \delta_2, \delta_3, \delta_4$, defined by $\delta_1(0) = \delta_1(1) = a_1$; $\delta_2(0) = a_1$, $\delta_2(1) = a_2$; $\delta_3(0) = a_2$, $\delta_3(1) = a_1$; and $\delta_4(0) = \delta_4(1) = a_2$. The risk function takes the following values

i	$R(p_1, \delta_i)$	$R(p_2, \delta_i)$	$\underset{p_1, p_2}{\text{Max}}\, R(p, \delta_i)$	$\underset{i}{\text{Min}}\, \underset{p_1, p_2}{\text{Max}}\, R(p, \delta_i)$
1	1	3	3	
2	$\frac{7}{4}$	$\frac{5}{2}$	$\frac{5}{2}$	$\frac{5}{2}$
3	$\frac{13}{4}$	$\frac{5}{2}$	$\frac{13}{4}$	
4	4	2	4	

Thus the minimax solution is $\delta_2(x) = a_1$ if $x = 0$ and $= a_2$ if $x = 1$.

The computation of minimax estimators is facilitated by the use of the *Bayes estimation method*. So far, we have considered θ as a fixed constant and $f_\theta(\mathbf{x})$ has represented the PDF (PMF) of the RV $\mathbf{X}$. In Bayesian estimation we treat θ as a random variable distributed according to PDF (PMF) $\pi(\theta)$ on Θ. Also, π is called the *a priori distribution*. Now $f(\mathbf{x} \mid \theta)$ represents the conditional probability density (or mass) function of RV $\mathbf{X}$, given that $\theta \in \Theta$ is held fixed. Since π is the distribution of θ, it follows that the joint density (PMF) of θ and $\mathbf{X}$ is given by

$$f(\mathbf{x}, \theta) = \pi(\theta) f(\mathbf{x} \mid \theta). \tag{3}$$

In this framework $R(\theta, \delta)$ is the conditional average loss, $E\{L(\theta, \delta(\mathbf{X})) \mid \theta\}$, given that θ is held fixed. (Note that we are using the same symbol to denote the RV θ and a value assumed by it.)

Definition 5. The Bayes risk of a decision function δ is defined by

$$R(\pi, \delta) = E_\pi R(\theta, \delta). \tag{4}$$

If θ is a continuous RV and $\mathbf{X}$ is of the continuous type, then

$$\begin{aligned} R(\pi, \delta) &= \int R(\theta, \delta) \pi(\theta)\, d\theta \\ &= \iint L(\theta, \delta(\mathbf{x})) f(\mathbf{x} \mid \theta) \pi(\theta)\, d\mathbf{x}\, d\theta \\ &= \iint L(\theta, \delta(\mathbf{x})) f(\mathbf{x}, \theta)\, d\mathbf{x}\, d\theta. \end{aligned} \tag{5}$$

If θ is discrete with PMF π and $\mathbf{X}$ is of the discrete type, then

$$R(\pi, \delta) = \sum_\theta \sum_{\mathbf{x}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}, \theta). \tag{6}$$

Similar expressions may be written in the other two cases.

Definition 6. A decision function δ^* is known as a Bayes rule (procedure) if it minimizes the Bayes risk, that is, if

$$R(\pi, \delta^*) = \inf_{\delta} R(\pi, \delta). \tag{7}$$

Definition 7. The conditional distribution of RV θ, given $\mathbf{X} = \mathbf{x}$, is called the a posteriori probability distribution of θ, given the sample.

Let the joint PDF (PMF) be expressed in the form

$$f(\mathbf{x}, \theta) = g(\mathbf{x}) h(\theta \mid \mathbf{x}), \tag{8}$$

where g denotes the joint marginal density (PMF) of $\mathbf{X}$. The a priori PDF (PMF) $\pi(\theta)$ gives the distribution of θ before the sample is taken, and the a posteriori PDF (PMF) $h(\theta \mid \mathbf{x})$ gives the distribution of θ after sampling. In terms of $h(\theta \mid \mathbf{x})$ we may write

$$R(\pi, \delta) = \int g(\mathbf{x}) \left\{ \int L(\theta, \delta(\mathbf{x})) h(\theta \mid \mathbf{x})\, d\theta \right\} d\mathbf{x} \tag{9}$$

or

$$R(\pi, \delta) = \sum_{\mathbf{x}} g(\mathbf{x}) \left\{ \sum_{\theta} L(\theta, \delta(\mathbf{x})) h(\theta \mid \mathbf{x}) \right\}, \tag{10}$$

depending on whether f and π are both continuous or both discrete. Similar expressions may be written if only one of f and π is discrete.

Theorem 1. Consider the problem of estimation of a parameter $\theta \in \Theta \subseteq \mathcal{R}$ with respect to the quadratic loss function $L(\theta, \delta) = (\theta - \delta)^2$. A Bayes solution is given by

$$\delta(\mathbf{x}) = E\{\theta \mid \mathbf{X} = \mathbf{x}\} \tag{11}$$

($\delta(x)$ defined by (11) is called the *Bayes estimator*).

Proof. In the continuous case, if π is the prior PDF of θ, then

$$R(\pi, \delta) = \int g(\mathbf{x}) \left\{ \int [\theta - \delta(\mathbf{x})]^2 h(\theta \mid \mathbf{x}) d\theta \right\} d\mathbf{x},$$

where g is the marginal PDF of $\mathbf{X}$, and h is the conditional PDF of θ, given $\mathbf{x}$. The Bayes rule is a function δ that minimizes $R(\pi, \delta)$. Minimization of $R(\pi, \delta)$ is the same as minimization of

$$\int [\theta - \delta(\mathbf{x})]^2 h(\theta \mid \mathbf{x})\, d\theta,$$

which is minimum if and only if

$$\delta(\mathbf{x}) = E\{\theta \mid \mathbf{x}\}.$$

The proof for the remaining cases is similar.