* Hogg, R.V., McKean, J.W. & Craig, A.T. (2005). Introduction to Mathematical Stats. (6th edition)

Indice:

Example 6.1.6. In Example 6.1.1, we discussed the mle of the probability of success θ for a random sample $X_1, X_2, \ldots, X_n$ from the Bernoulli distribution with pmf

$$p(x) = \begin{cases} \theta^x(1-\theta)^{1-x} & x = 0, 1 \\ 0 & \text{elsewhere,} \end{cases}$$

where $0 \le \theta \le 1$. Recall that the mle is $\overline{X}$, the proportion of sample successes. Now suppose that we know in advance that, instead of $0 \le \theta \le 1$, θ is restricted by the inequalities $0 \le \theta \le 1/3$. If the observations were such that $\overline{x} > 1/3$, then $\overline{x}$ would not be a satisfactory estimate. Since $\frac{\partial l(\theta)}{\partial \theta} > 0$, provided $\theta < \overline{x}$, under the restriction $0 \le \theta \le 1/3$, we can maximize $l(\theta)$ by taking $\widehat{\theta} = \min\{\overline{x}, \frac{1}{3}\}$. ■

The following is an appealing property of maximum likelihood estimates.

Theorem 6.1.2. *Let $X_1, \ldots, X_n$ be iid with the pdf $f(x;\theta), \theta \in \Omega$. For a specified function g, let $\eta = g(\theta)$ be a parameter of interest. Suppose $\widehat{\theta}$ is the mle of θ. Then $g(\widehat{\theta})$ is the mle of $\eta = g(\theta)$.*

Proof: First suppose g is a one-to-one function. The likelihood of interest is $L(g(\theta))$, but because g is one-to-one,

$$\max L(g(\theta)) = \max_{\eta = g(\theta)} L(\eta) = \max_{\eta} L(g^{-1}(\eta)).$$

But the maximum occurs when $g^{-1}(\eta) = \widehat{\theta}$; i.e., take $\widehat{\eta} = g(\widehat{\theta})$.

Suppose g is not one-to-one. For each η in the range of g, define the set (preimage),

$$g^{-1}(\eta) = \{\theta : g(\theta) = \eta\}.$$

The maximum occurs at $\widehat{\theta}$ and the domain of g is Ω which covers $\widehat{\theta}$. Hence, $\widehat{\theta}$ is in one of these preimages and, in fact, it can only be in one preimage. Hence to maximize $L(\eta)$, choose $\widehat{\eta}$ so that $g^{-1}(\widehat{\eta})$ is that unique preimage containing $\widehat{\theta}$. Then $\widehat{\eta} = g(\widehat{\theta})$. ■

In Example 6.1.5, it might be of interest to estimate $\mathrm{Var}(X) = \theta^2/12$. Hence by Theorem 6.1.2, the mle is $\max\{X_i\}^2/12$. Next, consider Example 6.1.1, where $X_1, \ldots, X_n$ are iid Bernoulli random variables with probability of success p. As shown in the example, $\widehat{p} = \overline{X}$ is the mle of p. Recall that in the large sample confidence interval for p, (5.4.8), an estimate of $\sqrt{p(1-p)}$ is required. By Theorem 6.1.2, the mle of this quantity is $\sqrt{\widehat{p}(1-\widehat{p})}$.

We close this section by showing that maximum likelihood estimators, under regularity conditions, are consistent estimators. Recall that $\mathbf{X}' = (X_1, \ldots, X_n)$.

Theorem 6.1.3. *Assume that $X_1, \ldots, X_n$ satisfy the regularity conditions (R0) - (R2), where θ_0 is the true parameter, and further that $f(x;\theta)$ is differentiable with respect to θ in Ω. Then the likelihood equation,*

$$\frac{\partial}{\partial \theta} L(\theta) = 0$$

* Introduction to Mathematical Statistics (6th edition)
Hogg, R.V., McKean, J.W. & Craig, A.T. (2005). Pearson

or equivalently

$$\frac{\partial}{\partial \theta} l(\theta) = 0$$

has a solution $\widehat{\theta}_n$ such that $\widehat{\theta}_n \xrightarrow{P} \theta_0$.

Proof: Because θ_0 is an interior point in Ω, $(\theta_0 - a, \theta_0 + a) \subset \Omega$, for some $a > 0$. Define S_n to be the event

$$S_n = \{\mathbf{X} : l(\theta_0; \mathbf{X}) > l(\theta_0 - a; \mathbf{X})\} \cap \{\mathbf{X} : l(\theta_0; \mathbf{X}) > l(\theta_0 + a; \mathbf{X})\}.$$

By Theorem 6.1.1, $P(S_n) \to 1$. So we can restrict attention to the event S_n. But on S_n, $l(\theta)$ has a local maximum say, $\widehat{\theta}_n$ such that $\theta_0 - a < \widehat{\theta}_n < \theta_0 + a$ and $l'(\widehat{\theta}_n) = 0$. That is,

$$S_n \subset \left\{\mathbf{X} : |\widehat{\theta}_n(\mathbf{X}) - \theta_0| < a\right\} \cap \left\{\mathbf{X} : l'(\widehat{\theta}_n(\mathbf{X})) = 0\right\}.$$

Therefore,

$$1 = \lim_{n\to\infty} P(S_n) \le \varliminf_{n\to\infty} P\left[\left\{\mathbf{X} : |\widehat{\theta}_n(\mathbf{X}) - \theta_0| < a\right\} \cap \left\{\mathbf{X} : l'(\widehat{\theta}_n(\mathbf{X})) = 0\right\}\right] \le 1;$$

see Remark 4.3.3 for discussion on $\varliminf$. It follows that for the sequence of solutions $\widehat{\theta}_n$, $P[|\widehat{\theta}_n - \theta_0| < a] \to 1$.

The only contentious point in the proof is that the sequence of solutions might depend on a. But we can always choose a solution "closest" to θ_0 in the following way. For each n, the set of all solutions in the interval is bounded, hence the infimum over solutions closest to θ_0 exists. ■

Note that this theorem is vague in that it discusses solutions of the equation. If, however, we know that the mle is the unique solution of the equation $l'(\theta) = 0$, then it is consistent. We state this as a corollary:

Corollary 6.1.1. *Assume that $X_1, \ldots, X_n$ satisfy the regularity conditions (R0) - (R2), where θ_0 is the true parameter, and that $f(x;\theta)$ is differentiable with respect to θ in Ω. Suppose the likelihood equation has the unique solution $\widehat{\theta}_n$. Then $\widehat{\theta}_n$ is a consistent estimator of θ_0.*

EXERCISES

* **6.1.1.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\theta, \sigma^2)$ distribution, $-\infty < \theta < \infty$ with σ^2 known. Determine the mle of θ.

* **6.1.2.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a $\Gamma(\alpha = 3, \beta = \theta)$ distribution, $0 < \theta < \infty$. Determine the mle of θ.

* **6.1.3.** Let $X_1, X_2, \ldots, X_n$ represent a random sample from each of the distributions having the following pdfs or pmfs:

(a) $f(x;\theta) = \theta^x e^{-\theta}/x!$, $x = 0, 1, 2, \ldots$, $0 \le \theta < \infty$, zero elsewhere, where $f(0;0) = 1$.

(b) $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $0 < \theta < \infty$, zero elsewhere.

(c) $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere.

(d) $f(x;\theta) = e^{-(x-\theta)}$, $\theta \le x < \infty$, $-\infty < \theta < \infty$, zero elsewhere.

In each case find the mle $\hat{\theta}$ of θ.

6.1.4. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a distribution with pdf $f(x;\theta) = 1$, $\theta - \frac{1}{2} \le x \le \theta + \frac{1}{2}$, $-\infty < \theta < \infty$, zero elsewhere. Show that every statistic $u(X_1, X_2, \ldots, X_n)$ such that

$$Y_n - \tfrac{1}{2} \le u(X_1, X_2, \ldots, X_n) \le Y_1 + \tfrac{1}{2}$$

is a mle of θ. In particular, $(4Y_1 + 2Y_n + 1)/6$, $(Y_1 + Y_n)/2$, and $(2Y_1 + 4Y_n - 1)/6$ are three such statistics. Thus uniqueness is not, in general, a property of a mle.

6.1.5. Suppose $X_1, \ldots, X_n$ are iid with pdf $f(x;\theta) = 2x/\theta^2$, $0 < x \le \theta$, zero elsewhere, find:

(a) The mle $\hat{\theta}$ for θ.

(b) The constant c so that $E(c\hat{\theta}) = \theta$.

(c) The mle for the median of the distribution.

6.1.6. Suppose $X_1, X_2, \ldots, X_n$ are iid with pdf $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, zero elsewhere. Find the mle of $P(X \le 2)$.

6.1.7. Let the table

x	0	1	2	3	4	5
Frequency	6	10	14	13	6	1

represent a summary of a sample of size 50 from a binomial distribution having $n = 5$. Find the mle of $P(X \ge 3)$.

6.1.8. Let X_1, X_2, X_3, X_4, X_5 be a random sample from a Cauchy distribution with median θ, that is, with pdf.

$$f(x;\theta) = \frac{1}{\pi}\frac{1}{1 + (x-\theta)^2}, \quad -\infty < x < \infty,$$

where $-\infty < \theta < \infty$. If $x_1 = -1.94$, $x_2 = 0.59$, $x_3 = -5.98$, $x_4 = -0.08$, and $x_5 = -0.77$, find by numerical methods the mle of θ.

6.1.9. Let the table

x	0	1	2	3	4	5
Frequency	7	14	12	13	6	3

represent a summary of a random sample of size 50 from a Poisson distribution. Find the maximum likelihood estimate of $P(X = 2)$.

6.1.10. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p. If p is restricted so that we know that $\frac{1}{2} \le p \le 1$, find the mle of this parameter.

6.1.11. Let $X_1, X_2, \ldots, X_n$ be random sample from a $N(\theta, \sigma^2)$ distribution, where σ^2 is fixed but $-\infty < \theta < \infty$.

(a) Show that the mle of θ is $\overline{X}$.

(b) If θ is restricted by $0 \le \theta < \infty$, show that the mle of θ is $\widehat{\theta} = \max\{0, \overline{X}\}$.

6.1.12. Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with $0 < \theta \le 2$. Show that the mle of θ is $\widehat{\theta} = \min\{\overline{X}, 2\}$.

6.1.13. Let $X_1, X_2, \ldots, X_n$ ba random sample from a distribution with one of two pdfs. If $\theta = 1$, then $f(x;\theta = 1) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $-\infty < x < \infty$. If $\theta = 2$, then $f(x;\theta = 2) = 1/[\pi(1 + x^2)]$, $-\infty < x < \infty$. Find the mle of θ.

6.2 Rao-Cramér Lower Bound and Efficiency

In this section we establish a remarkable inequality called the **Rao-Cramér** lower bound which gives a lower bound on the variance of any unbiased estimate. We then show that, under regularity conditions, the variances of the maximum likelihood estimates achieve this lower bound asymptotically.

As in the last section, let X be a random variable with pdf $f(x;\theta)$, $\theta \in \Omega$, where the parameter space Ω is an open interval. In addition to the regularity conditions (6.1.1) of Section 6.1, for the following derivations, we require two more regularity conditions given by

Assumptions 6.2.1. (Additional Regularity Conditions).

(R3): *The pdf $f(x;\theta)$ is twice differentiable as a function of θ.*

(R4): *The integral $\int f(x;\theta)\,dx$ can be differentiated twice under the integral sign as a function of θ.*

Note that conditions (R1)-(R4) mean that the parameter θ does not appear in the endpoints of the interval in which $f(x;\theta) > 0$ and that we can interchange integration and differentiation with respect to θ. Our derivation is for the continuous case but the discrete case can be handled in a similar manner. We begin with the identity

$$1 = \int_{-\infty}^{\infty} f(x;\theta)\,dx.$$

Taking the derivative with respect to θ results in,

$$0 = \int_{-\infty}^{\infty} \frac{\partial f(x;\theta)}{\partial \theta}\,dx.$$

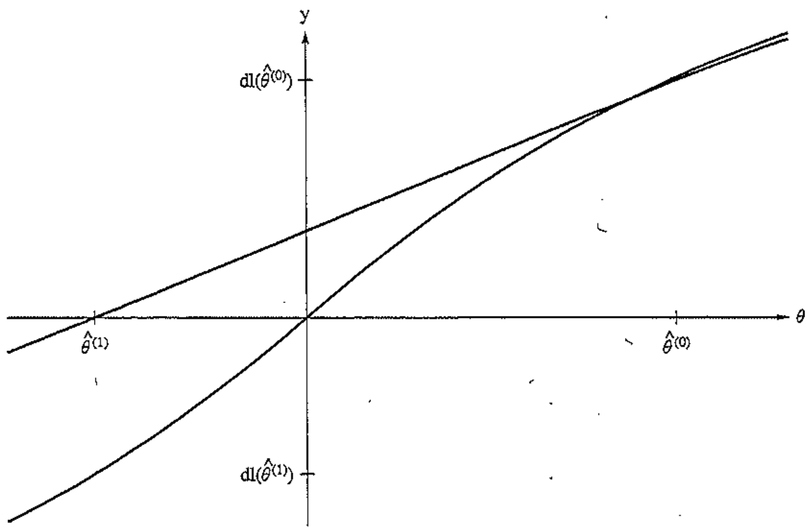

Figure 6.2.1: Beginning with starting value $\widehat{\theta}^{(0)}$, the one step estimate is $\widehat{\theta}^{(1)}$ which is the intersection of the tangent line to the curve $l'(\theta)$ at $\widehat{\theta}^{(0)}$ and the horizontal axis. In the figure, $dl(\theta) = l'(\theta)$.

EXERCISES

6.2.1. Prove that $\overline{X}$, the mean of a random sample of size n from a distribution that is $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$, is, for every known $\sigma^2 > 0$, an efficient estimator of θ.

6.2.2. Given $f(x;\theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere, with $\theta > 0$, formally compute the reciprocal of

$$nE\left\{\left[\frac{\partial \log f(X:\theta)}{\partial\theta}\right]^2\right\}.$$

Compare this with the variance of $(n+1)Y_n/n$, where Y_n is the largest observation of a random sample of size n from this distribution. Comment.

6.2.3. Given the pdf

$$f(x;\theta) = \frac{1}{\pi[1+(x-\theta)^2]}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty.$$

Show that the Rao-Cramér lower bound is $2/n$, where n is the size of a random sample from this Cauchy distribution. What is the asymptotic distribution of $\sqrt{n}(\widehat{\theta}-\theta)$, if $\widehat{\theta}$ is the mle of θ?

6.2.4. Consider Example 6.2.2, where we discussed the location model.

(a) Write the location model when e_i has the logistic pdf given in expression (5.2.8).

(b) Using expression (6.2.8), show that the information, $I(\theta) = 1/3$, for the model in Part (a). *Hint:* In the integral of expression (6.2.8), use the substitution $u = (1+e^{-z})^{-1}$. Then $du = f(z)dz$, where $f(z)$ is the pdf (5.2.8).

6.2.5. Using the same location model as in Part (a) Exercise 6.2.4, obtain the ARE of the sample median to mle of the model.
Hint: The mle of θ for this model is discussed in Example 6.2.7. Furthermore as shown in Theorem 10.2.3 of Chapter 10, Q_2 is asymptotically normal with asymptotic mean θ and asymptotic variance $1/(4f^2(0)n)$.

6.2.6. Consider a location model (Example 6.2.2) when the error pdf is the contaminated normal (3.4.14) with ϵ proportion of contamination and with σ_c^2 as the variance of the contaminated part. Show that the ARE of the sample median to the sample mean is given by

$$e(Q_2, \overline{X}) = \frac{2[1+\epsilon(\sigma_c^2-1)][1-\epsilon+(\epsilon/\sigma_c)]^2}{\pi}. \tag{6.2.34}$$

Use the hint in Exercise 6.2.5 for the median.

(a) If $\sigma_c^2 = 9$, use (6.2.34) to fill in the following table:

ϵ	0	0.05	0.10	0.15
$e(Q_2, \overline{X})$				

(b) Notice from the table that the sample median becomes the "better" estimator when ϵ increases from 0.10 to 0.15. Determine the value for ϵ where this occurs (this involves a third degree polynomial in ϵ, so one way of obtaining the root is to use the Newton algorithm discussed around expression (6.2.32)).

6.2.7. Let X have a gamma distribution with $\alpha = 4$ and $\beta = \theta > 0$.

(a) Find the Fisher information $I(\theta)$.

(b) If $X_1, X_2, \ldots, X_n$ is a random sample from this distribution, show that the mle of θ is an efficient estimator of θ.

(c) What is the asymptotic distribution of $\sqrt{n}(\widehat{\theta}-\theta)$?

6.2.8. Let X be $N(0,\theta)$, $0 < \theta < \infty$.

(a) Find the Fisher information $I(\theta)$.

(b) If $X_1, X_2, \ldots, X_n$ is a random sample from this distribution, show that the mle of θ is an efficient estimator of θ.

(c) What is the asymptotic distribution of $\sqrt{n}(\widehat{\theta}-\theta)$?

* **6.2.9.** If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with pdf

$$f(x;\theta) = \begin{cases} \frac{3\theta^3}{(x+\theta)^4} & 0 < x < \infty, 0 < \theta < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Show that $Y = 2\overline{X}$ is an unbiased estimator of θ and determine its efficiency.

* **6.2.10.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(0,\theta)$ distribution. We want to estimate the standard deviation $\sqrt{\theta}$. Find the constant c so that $Y = c\sum_{i=1}^{n}|X_i|$ is an unbiased estimator of $\sqrt{\theta}$ and determine its efficiency.

* **6.2.11.** Let $\overline{X}$ be the mean of a random sample of size n from a $N(\theta,\sigma^2)$ distribution, $-\infty < \theta < \infty, \sigma^2 > 0$. Assume that σ^2 is known. Show that $\overline{X}^2 - \frac{\sigma^2}{n}$ is an unbiased estimator of θ^2 and find its efficiency.

* **6.2.12.** Recall that $\widehat{\theta} = -n/\sum_{i=1}^{n}\log X_i$ is the mle of θ for a Beta$(\theta,1)$ distribution. Also, $W = -\sum_{i=1}^{n}\log X_i$ has the gamma distribution $\Gamma(n, 1/\theta)$.

(a) Show that $2\theta W$ has a $\chi^2(2n)$ distribution.

(b) Using Part (a), find c_1 and c_2 so that

$$P\left(c_1 < \frac{2\theta n}{\widehat{\theta}} < c_2\right) = 1-\alpha,$$

for $0 < \alpha < 1$. Next, obtain a $(1-\alpha)100\%$ confidence interval for θ.

(c) Let $n = 10$ and compare the length of this interval with the length of the interval found in Example 6.2.6.

6.2.13. By using expressions (6.2.21) and (6.2.22) obtain the result for the one-step estimate discussed at the end of this section.

* **6.2.14.** Let S^2 be the sample variance of a random sample of size $n > 1$ from $N(\mu,\theta)$, $0 < \theta < \infty$, where μ is known. We know $E(S^2) = \theta$.

(a) What is the efficiency of S^2?

(b) Under these conditions, what is the mle $\widehat{\theta}$ of θ?

(c) What is the asymptotic distribution of $\sqrt{n}(\widehat{\theta}-\theta)$?

6.3 Maximum Likelihood Tests

The last section presented an inference for pointwise estimation and confidence intervals based on likelihood theory. In this section, we present a corresponding inference for testing hypotheses.

As in the last section, let $X_1, \ldots, X_n$ be iid with pdf $f(x;\theta)$ for $\theta \in \Omega$. In this section, θ is a scalar but in Sections 6.4 and 6.5 extensions to the vector valued case will be discussed. Consider the two-sided hypotheses

$$H_0: \theta = \theta_0 \text{ versus } H_1: \theta \neq \theta_0, \tag{6.3.1}$$

where θ_0 is a specified value.

Recall that the likelihood function and its log are given by:

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta)$$
$$l(\theta) = \sum_{i=1}^{n}\log f(X_i;\theta).$$

Let $\widehat{\theta}$ denote the maximum likelihood estimate of θ.

To motivate the test, consider Theorem 6.1.1 which says that if θ_0 is the true value of θ then, asymptotically, $L(\theta_0)$ is the maximum value of $L(\theta)$. Consider the ratio of two likelihood functions, namely

$$\Lambda = \frac{L(\theta_0)}{L(\widehat{\theta})}. \tag{6.3.2}$$

Note that $\Lambda \leq 1$, but if H_0 is true Λ should be large (close to 1); while, if H_1 is true, Λ should be smaller. For a specified significance level α, this leads to the intuitive decision rule,

$$\text{Reject } H_0 \text{ in favor of } H_1 \text{ if } \Lambda \leq c, \tag{6.3.3}$$

where c is such that $\alpha = P_{\theta_0}[\Lambda \leq c]$. This test is called the **likelihood ratio test.** Theorem 6.3.1 derives the asymptotic distribution of Λ under H_0, but first we will look at two examples.

Example 6.3.1 (Likelihood Ratio Test for the Exponential Distribution). Suppose $X_1, \ldots, X_n$ are iid with pdf $f(x;\theta) = \theta^{-1}\exp\{-x/\theta\}$, for $x, \theta > 0$. Let the hypotheses be given by (6.3.1). The likelihood function simplifies to

$$L(\theta) = \theta^{-n}\exp\{-(n/\theta)\overline{X}\}.$$

From Example 6.1.2, the mle of θ is $\overline{X}$. After some simplification, the likelihood ratio test statistic simplifies to

$$\Lambda = e^n\left(\frac{\overline{X}}{\theta_0}\right)^n \exp\{-n\overline{X}/\theta_0\}. \tag{6.3.4}$$

Example 6.3.4 (Likelihood Tests for the Laplace Location Model). Consider the location model

$$X_i = \theta + e_i, \quad i = 1, \ldots, n,$$

where $-\infty < \theta < \infty$ and the random errors e_is are iid each having the Laplace pdf, (6.2.9). Technically, the Laplace distribution does not satisfy all of the regularity conditions (R0) - (R5) but the results below can be derived rigorously; see, for example, Hettmansperger and McKean (1998). Consider testing the hypotheses

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0,$$

where θ_0 is specified. Here $\Omega = (-\infty, \infty)$ and $\omega = \{\theta_0\}$. By Example 6.1.3, we know that the mle of θ under Ω is $Q_2 = \text{med}\{X, \ldots, X_n\}$, the sample median. It follows that

$$L(\widehat{\Omega}) = 2^{-n} \exp\{-\sum_{i=1}^{n} |x_i - Q_2|\},$$

while

$$L(\widehat{\omega}) = 2^{-n} \exp\{-\sum_{i=1}^{n} |x_i - \theta_0|\}.$$

Hence, the negative of twice the log of the likelihood ratio test statistic is

$$-2 \log \Lambda = 2 \left[\sum_{i=1}^{n} |x_i - \theta_0| - \sum_{i=1}^{n} |x_i - Q_2| \right]. \tag{6.3.24}$$

Thus the size α asymptotic likelihood ratio test for H_0 versus H_1 rejects H_0 in favor of H_1 if

$$2 \left[\sum_{i=1}^{n} |x_i - \theta_0| - \sum_{i=1}^{n} |x_i - Q_2| \right] \geq \chi^2_\alpha(1).$$

By (6.2.10), the Fisher information for this model is $I(\theta) = 1$. Thus the Wald type test statistic simplifies to

$$\chi^2_W = [\sqrt{n}(Q_2 - \theta_0)]^2.$$

For the scores test, we have

$$\frac{\partial \log f(x_i - \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left[\log \frac{1}{2} - |x_i - \theta| \right] = \text{sgn}(x_i - \theta).$$

Hence, the score vector for this model is $\mathbf{S}(\theta) = (\text{sgn}(X_1 - \theta), \ldots, \text{sgn}(X_n - \theta))'$. From the above discussion, (see equation (6.3.17)), the scores test statistic can be written as

$$\chi^2_R = (S^*)^2/n,$$

where

$$S^* = \sum_{i=1}^{n} \text{sgn}(X_i - \theta_0).$$

As Exercise 6.3.4 shows, under H_0, S^* is a linear function of a random variable with a $b(n, 1/2)$ distribution. ■

Which of the three tests should we use? Based on the above discussion, all three tests are asymptotically equivalent under the null hypothesis. Similar to the concept of asymptotic relative efficiency (ARE) we can derive an equivalent concept of efficiency for tests; see Chapter 10 and more advanced books such as Hettmansperger and McKean (1998). However, all three tests have the same asymptotic efficiency. Hence, asymptotic theory offers little help in separating the tests. There have been finite sample comparisons in the literature; but these studies have not selected any of these as a "best" test overall; see Chapter 7 of Lehmann (1999) for more discussion.

EXERCISES

6.3.1. Consider the decision rule (6.3.5) derived in Example 6.3.1. Obtain the distribution of the test statistic under a general alternative and use it to obtain the power function of the test. If computational facilities are available, sketch this power curve for the case when $\theta_0 = 1$, $n = 10$, and $\alpha = 0.05$.

6.3.2. Show that the test with decision rule (6.3.6) is like that of Example 5.6.1 except here σ^2 is known.

6.3.3. Consider the decision rule (6.3.6) derived in Example 6.3.2. Obtain an equivalent test statistic which has a standard normal distribution under H_0. Next obtain the distribution of this test statistic under a general alternative and use it to obtain the power function of the test. If computational facilities are available, sketch this power curve for the case when $\theta_0 = 0$, $n = 10$, $\sigma^2 = 1$, and $\alpha = 0.05$.

6.3.4. Consider Example 6.3.4.

(a) Show that we can write $S^* = 2T - n$ where $T = \#\{X_i > \theta_0\}$.

(b) Show that the scores test for this model is equivalent to rejecting H_0 if $T < c_1$ or $T > c_2$.

(c) Show that under H_0, T has the binomial distribution $b(n, 1/2)$; hence, determine c_1 and c_2 so the test has size α.

(d) Determine the power function for the test based on T as a function of θ.

6.3.5. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu_0, \sigma^2 = \theta)$ distribution, where $0 < \theta < \infty$ and μ_0 is known. Show that the likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ can be based upon the statistic $W = \sum_{i=1}^{n}(X_i - \mu_0)^2/\theta_0$. Determine the null distribution of W and give, explicitly, the rejection rule for a level α test.

6.3.6. For the test described in Exercise 6.3.5, obtain the distribution of the test statistic under general alternatives. If computational facilities are available, sketch this power curve for the case when $\theta_0 = 1$, $n = 10$, $\mu = 0$, and $\alpha = 0.05$.

6.3.7. Using the results of Example 6.2.4, find an exact size α test for the hypotheses (6.3.21).

* **6.3.8.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean $\theta > 0$.

  (a) Show that the likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is based upon the statistic $Y = \sum_{i=1}^n X_i$. Obtain the null distribution of Y.

  (b) For $\theta_0 = 2$ and $n = 5$, find the significance level of the test that rejects H_0 if $Y \leq 4$ or $Y \geq 17$.

* **6.3.9.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli $b(1, \theta)$ distribution, where $0 < \theta < 1$.

  (a) Show that the likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is based upon the statistic $Y = \sum_{i=1}^n X_i$. Obtain the null distribution of Y.

  (b) For $n = 100$ and $\theta_0 = 1/2$, find c_1 so that the test rejects H_0 when $Y \leq c_1$ or $Y \geq c_2 = 100 - c_1$ has the approximate significance level of $\alpha = 0.05$. *Hint:* Use the Central Limit Theorem.

* **6.3.10.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a $\Gamma(\alpha = 3, \beta = \theta)$ distribution, where $0 < \theta < \infty$.

  (a) Show that the likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is based upon the statistic $W = \sum_{i=1}^n X_i$. Obtain the null distribution of $2W/\theta_0$.

  (b) For $\theta_0 = 3$ and $n = 5$, find c_1 and c_2 so that the test that rejects H_0 when $W \leq c_1$ or $W \geq c_2$ has significance level 0.05.

* **6.3.11.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta \exp\left\{-|x|^\theta\right\}/2\Gamma(1/\theta)$, $-\infty < x < \infty$, where $\theta > 0$. Suppose $\Omega = \{\theta : \theta = 1, 2\}$. Consider the hypotheses $H_0 : \theta = 2$ (a normal distribution) versus $H_1 : \theta = 1$ (a double exponential distribution). Show that the likelihood ratio test can be based on the statistic $W = \sum_{i=1}^n (X_i^2 - |X_i|)$.

* **6.3.12.** Let $X_1, X_2, \ldots, X_n$ be a random sample from the beta distribution with $\alpha = \beta = \theta$ and $\Omega = \{\theta : \theta = 1, 2\}$. Show that the likelihood ratio test statistic Λ for testing $H_0 : \theta = 1$ versus $H_1 : \theta = 2$ is a function of the statistic $W = \sum_{i=1}^n \log X_i + \sum_{i=1}^n \log(1 - X_i)$.

6.3.13. Consider a location model

$$X_i = \theta + e_i, \quad i = 1, \ldots, n, \tag{6.3.25}$$

where $e_1, e_2, \ldots, e_n$ are iid with pdf $f(z)$. There is a nice geometric interpretation for estimating θ. Let $\mathbf{X} = (X_1, \ldots, X_n)'$ and $\mathbf{e} = (e_1, \ldots, e_n)'$ be the vectors of observations and random error, respectively, and let $\boldsymbol{\mu} = \theta\mathbf{1}$ where $\mathbf{1}$ is a vector with all components equal to one. Let V be the subspace of vectors of the form $\boldsymbol{\mu}$; i.e, $V = \{\mathbf{v} : \mathbf{v} = a\mathbf{1}, \text{ for some } a \in R\}$. Then in vector notation we can write the model as

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{e}, \quad \boldsymbol{\mu} \in V. \tag{6.3.26}$$

Then we can summarize the model by saying, "Except for the random error vector $\mathbf{e}$, $\mathbf{X}$ would reside in V." Hence, it makes sense intuitively to estimate $\boldsymbol{\mu}$ by a vector in V which is "closest" to $\mathbf{X}$. That is, given a norm $\|\cdot\|$ in R^n choose

$$\widehat{\boldsymbol{\mu}} = \text{Argmin}\|\mathbf{X} - \mathbf{v}\|, \quad \mathbf{v} \in V. \tag{6.3.27}$$

(a) If the error pdf is the Laplace, (6.2.9), show that the minimization in (6.3.27) is equivalent to maximizing the likelihood, when the norm is the l_1 norm given by

$$\|\mathbf{v}\|_1 = \sum_{i=1}^n |v_i|. \tag{6.3.28}$$

(b) If the error pdf is the $N(0,1)$, show that the minimization in (6.3.27) is equivalent to maximizing the likelihood, when the norm is given by the square of the l_2 norm

$$\|\mathbf{v}\|_2^2 = \sum_{i=1}^n v_i^2. \tag{6.3.29}$$

6.3.14. Continuing with the last exercise, besides estimation there is also a nice geometric interpretation for testing. For the model (6.3.26), consider the hypotheses

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0, \tag{6.3.30}$$

where θ_0 is specified. Given a norm $\|\cdot\|$ on R^n, denote by $d(\mathbf{X}, V)$ the distance between $\mathbf{X}$ and the subspace V; i.e., $d(\mathbf{X}, V) = \|\mathbf{X} - \widehat{\boldsymbol{\mu}}\|$, where $\widehat{\boldsymbol{\mu}}$ is defined in equation (6.3.27). If H_0 is true, then $\widehat{\boldsymbol{\mu}}$ should be close to $\boldsymbol{\mu} = \theta_0\mathbf{1}$ and, hence, $\|\mathbf{X} - \theta_0\mathbf{1}\|$ should be close to $d(\mathbf{X}, V)$. Denote the difference by

$$RD = \|\mathbf{X} - \theta_0\mathbf{1}\| - \|\mathbf{X} - \widehat{\boldsymbol{\mu}}\|. \tag{6.3.31}$$

Small values of RD indicate that the null hypothesis is true while large values indicate H_1. So our rejection rule when using RD is

$$\text{Reject } H_0 \text{ in favor of } H_1, \text{ if } RD > c. \tag{6.3.32}$$

(a) If the error pdf is the Laplace, (6.1.6), show that expression (6.3.31) is equivalent to the likelihood ratio test, when the norm is given by (6.3.28).

(b) If the error pdf is the $N(0,1)$, show that expression (6.3.31) is equivalent to the likelihood ratio test when the norm is given by the square of the l_2 norm, (6.3.29).

6.3.15. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pmf $p(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, where $0 < \theta < 1$. We wish to test $H_0 : \theta = 1/3$ versus $H_1 : \theta \neq 1/3$.

(a) Find Λ and $-2\log\Lambda$.

(b) Determine the Wald-type test.

where Z_1 and Z_2 are iid $N(0,1)$ random variables. Assume for $i = 1,2$ that, as $n \to \infty$, $n_i/n \to \lambda_i$, where $0 < \lambda_i < 1$ and $\lambda_1 + \lambda_2 = 1$. As Exercise 6.5.10 shows

$$\sqrt{n}[(\widehat{p}_1 - \widehat{p}_2) - (p_1 - p_2)] \xrightarrow{D} N\left(0, \frac{1}{\lambda_1}p_1(1-p_1) + \frac{1}{\lambda_2}p_2(1-p_2)\right). \tag{6.5.23}$$

It follows that the random variable

$$Z = \frac{(\widehat{p}_1 - \widehat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \tag{6.5.24}$$

has an approximate $N(0,1)$ distribution. Under H_0, $p_1 - p_2 = 0$. We could use Z as a test statistic, provided we replace the parameters $p_1(1-p_1)$ and $p_2(1-p_2)$ in its denominator with a consistent estimate. Recall that $\widehat{p}_i \to p_i$, $i = 1,2$, in probability. Thus under H_0, the statistic

$$Z^* = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}} \tag{6.5.25}$$

has an approximate $N(0,1)$ distribution. Hence, an approximate level α test is to reject H_0, if $|z^*| \geq z_{\alpha/2}$. Another consistent estimator of the denominator is discussed in Exercise 6.5.11. ■

EXERCISES

6.5.1. In Example 6.5.1 let $n = 10$, and let the experimental value of the random variables yield $\overline{x} = 0.6$ and $\sum_1^{10}(x_i - \overline{x})^2 = 3.6$. If the test derived in that example is used, do we accept or reject $H_0: \theta_1 = 0$ at the 5 percent significance level?

6.5.2. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution $N(\theta_1, \theta_2)$. Show that the likelihood ratio principle for testing $H_0: \theta_2 = \theta_2'$ specified, and θ_1 unspecified, against $H_1: \theta_2 \neq \theta_2'$, θ_1 unspecified, leads to a test that rejects when $\sum_1^n (x_i - \overline{x})^2 \leq c_1$ or $\sum_1^n (x_i - \overline{x})^2 \geq c_2$, where $c_1 < c_2$ are selected appropriately.

6.5.3. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be independent random samples from the distributions $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$, respectively.

(a) Show that the likelihood ratio for testing $H_0: \theta_1 = \theta_2$, $\theta_3 = \theta_4$ against all alternatives is given by

$$\frac{\left[\sum_1^n (x_i - \overline{x})^2/n\right]^{n/2}\left[\sum_1^m (y_i - \overline{y})^2/m\right]^{m/2}}{\left\{\left[\sum_1^n (x_i - u)^2 + \sum_1^m (y_i - u)^2\right] \Big/ (m+n)\right\}^{(n+m)/2}},$$

where $u = (n\overline{x} + m\overline{y})/(n+m)$.

(b) Show that the likelihood ratio test for testing $H_0: \theta_3 = \theta_4$, θ_1 and θ_2 unspecified, against $H_1: \theta_3 \neq \theta_4$, θ_1 and θ_2 unspecified, can be based on the random variable

$$F = \frac{\sum_1^n (X_i - \overline{X})^2/(n-1)}{\sum_1^m (Y_i - \overline{Y})^2/(m-1)}.$$

6.5.4. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be independent random samples from the two normal distributions $N(0, \theta_1)$ and $N(0, \theta_2)$.

(a) Find the likelihood ratio Λ for testing the composite hypothesis $H_0: \theta_1 = \theta_2$ against the composite alternative $H_1: \theta_1 \neq \theta_2$.

(b) This Λ is a function of what F-statistic that would actually be used in this test?

6.5.5. Let X and Y be two independent random variables with respective pdfs

$$f(x;\theta_i) = \begin{cases} \left(\frac{1}{\theta_i}\right) e^{-x/\theta_i} & 0 < x < \infty, 0 < \theta_i < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

for $i = 1,2$. To test $H_0: \theta_1 = \theta_2$ against $H_1: \theta_1 \neq \theta_2$, two independent samples of sizes n_1 and n_2, respectively, were taken from these distributions. Find the likelihood ratio Λ and show that Λ can be written as a function of a statistic having an F-distribution, under H_0.

6.5.6. Consider the two uniform distributions with respective pdfs

$$f(x;\theta_i) = \begin{cases} \frac{1}{2\theta_i} & -\theta_i < x < \theta_i, -\infty < \theta_i < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

for $i = 1,2$. The null hypothesis is $H_0: \theta_1 = \theta_2$ while the alternative is $H_1: \theta_1 \neq \theta_2$. Let $X_1 < X_2 < \cdots < X_{n_1}$ and $Y_1 < Y_2 < \cdots < Y_{n_2}$ be the order statistics of two independent random samples from the respective distributions. Using the likelihood ratio Λ, find the statistic used to test H_0 against H_1. Find the distribution of $-2\log\Lambda$ when H_0 is true. Note that in this nonregular case the number of degrees of freedom is two times the difference of the dimension of Ω and ω.

6.5.7. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random sample from a bivariate normal distribution with $\mu_1, \mu_2, \sigma_1^2 = \sigma_2^2 = \sigma^2, \rho = \frac{1}{2}$, where μ_1, μ_2, and $\sigma^2 > 0$ are unknown real numbers. Find the likelihood ratio Λ for testing $H_0: \mu_1 = \mu_2 = 0$, σ^2 unknown against all alternatives. The likelihood ratio Λ is a function of what statistic that has a well-known distribution?

6.5.8. Let n independent trials of an experiment be such that $x_1, x_2, \ldots, x_k$ are the respective numbers of times that the experiment ends in the mutually exclusive and exhaustive events $C_1, C_2, \ldots, C_k$. If $p_i = P(C_i)$ is constant throughout the n trials, then the probability of that particular sequence of trials is $L = p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$.

(a) Recalling that $p_1 + p_2 + \cdots + p_k = 1$, show that the likelihood ratio for testing $H_0 : p_i = p_{i0} > 0$, $i = 1, 2, \ldots, k$, against all alternatives is given by

$$\Lambda = \prod_{i=1}^{k} \left(\frac{(p_{i0})^{x_i}}{(x_i/n)^{x_i}} \right).$$

(b) Show that

$$-2 \log \Lambda = \sum_{i=1}^{k} \frac{x_i (x_i - np_{0i})^2}{(np_i')^2},$$

where p_i' is between p_{0i} and x_i/n.
Hint: Expand $\log p_{i0}$ in a Taylor's series with the remainder in the term involving $(p_{i0} - x_i/n)^2$.

(c) For large n, argue that $x_i/(np_i')^2$ is approximated by $1/(np_{i0})$ and hence

$$-2 \log \Lambda \approx \sum_{i=1}^{k} \frac{(x_i - np_{0i})^2}{np_{0i}}, \quad \text{when } H_0 \text{ is true.}$$

Theorem 6.5.1 says that the right-hand member of this last equation defines a statistic that has an approximate chi-square distribution with $k-1$ degrees of freedom. Note that

$$\text{dimension of } \Omega - \text{dimension of } \omega = (k-1) - 0 = k - 1.$$

6.5.9. Finish the derivation of the LRT found in Example 6.5.3. Simplify as much as possible.

6.5.10. Show that expression (6.5.23) of Example 6.5.3 is true.

6.5.11. As discussed in Example 6.5.3, Z, (6.5.25), can be used as a test statistic provided we have a consistent estimator of $p_1(1-p_1)$ and $p_2(1-p_2)$ when H_0 is true. In the example, we discussed an estimator which is consistent under both H_0 and H_1. Under H_0, though, $p_1(1-p_1) = p_2(1-p_2) = p(1-p)$, where $p = p_1 = p_2$. Show that the statistic (6.5.22) is a consistent estimator of p, under H_0. Thus determine another test of H_0.

* **6.5.12.** A machine shop that manufactures toggle levers has both a day and a night shift. A toggle lever is defective if a standard nut cannot be screwed onto the threads. Let p_1 and p_2 be the proportion of defective levers among those manufactured by the day and night shifts, respectively. We shall test the null hypothesis, $H_0 : p_1 = p_2$, against a two-sided alternative hypothesis based on two random samples, each of 1000 levers taken from the production of the respective shifts. Use the test statistic Z^* given in Example 6.5.3.

(a) Sketch a standard normal pdf illustrating the critical region having $\alpha = 0.05$.

(b) If $y_1 = 37$ and $y_2 = 53$ defectives were observed for the day and night shifts, respectively, calculate the value of the test statistic and the approximate p-value (note that this is a two-sided test). Locate the calculated test statistic on your figure in Part (a) and state your conclusion. Obtain the approximate p-value of the test.

* **6.5.13.** For the situation given in Part (b) of Exercise 6.5.12, calculate the tests defined in Exercises 6.5.9 and 6.5.11. Obtain the approximate p-values of all three tests. Discuss the results.

6.6 The EM Algorithm

In practice, we are often in the situation where part of the data is missing. For example, we may be observing lifetimes of mechanical parts which have been put on test and some of these parts are still functioning when the statistical analysis is carried out. In this section, we introduce the EM Algorithm which frequently can be used in these situations to obtain maximum likelihood estimates. Our presentation is brief. For further information, the interested reader can consult the literature in this area including the monograph by McLachlan and Krishnan (1997). Although, for convenience, we will write in terms of continuous random variables; the theory in this section holds for the discrete case as well.

Suppose we consider a sample of n items, where n_1 of the items are observed while $n_2 = n - n_1$ items are not observable. Denote the observed items by $\mathbf{X}' = (X_1, X_2, \ldots, X_{n_1})$ and the unobserved items by $\mathbf{Z}' = (Z_1, Z_2, \ldots, Z_{n_2})$. Assume that the X_is are iid with pdf $f(x|\theta)$, where $\theta \in \Omega$. Assume that Z_js and the X_is are mutually independent. The conditional notation will prove useful here. Let $g(\mathbf{x}|\theta)$ denote the joint pdf of $\mathbf{X}$. Let $h(\mathbf{x}, \mathbf{z}|\theta)$ denote the joint pdf of the observed and the unobserved items. Let $k(\mathbf{z}|\theta, \mathbf{x})$ denote the conditional pdf of the missing data given the observed data. By the definition of a conditional pdf, we have the identity

$$k(\mathbf{z}|\theta, \mathbf{x}) = \frac{h(\mathbf{x}, \mathbf{z}|\theta)}{g(\mathbf{x}|\theta)}. \tag{6.6.1}$$

The **observed likelihood** function is $L(\theta|\mathbf{x}) = g(\mathbf{x}|\theta)$. The **complete likelihood** function is defined by

$$L^c(\theta|\mathbf{x}, \mathbf{z}) = h(\mathbf{x}, \mathbf{z}|\theta). \tag{6.6.2}$$

Our goal is maximize the likelihood function $L(\theta|\mathbf{x})$ by using the complete likelihood $L^c(\theta|\mathbf{x}, \mathbf{z})$ in this process.

Using (6.6.1), we derive the following basic identity for an arbitrary but fixed discussion. Replacing w_i by γ_i in expression (6.6.19), the M step of the algorithm is to maximize

$$Q(\theta|\theta_0,\mathbf{x}) = \sum_{i=1}^{n}[(1-\gamma_i)\log f_1(x_i) + \gamma_i \log f_2(x_i)]. \quad (6.6.21)$$

This maximization is easy to obtain by taking partial derivatives of $Q(\theta|\theta_0,\mathbf{x})$ with respect to the parameters. For example,

$$\frac{\partial Q}{\partial \mu_1} = \sum_{i=1}^{n}(1-\gamma_i)(-1/2\sigma_1^2)(-2)(x_i-\mu_1).$$

Setting this to 0 and solving for μ_1 yields the estimate of μ_1. The estimates of the other mean and the variances can be obtained similarly. These estimates are:

$$\begin{aligned}
\widehat{\mu}_1 &= \frac{\sum_{i=1}^{n}(1-\gamma_i)x_i}{\sum_{i=1}^{n}(1-\gamma_i)}, \\
\widehat{\sigma}_1^2 &= \frac{\sum_{i=1}^{n}(1-\gamma_i)(x_i-\widehat{\mu}_1)^2}{\sum_{i=1}^{n}(1-\gamma_i)}, \\
\widehat{\mu}_2 &= \frac{\sum_{i=1}^{n}\gamma_i x_i}{\sum_{i=1}^{n}\gamma_i}, \\
\widehat{\sigma}_2^2 &= \frac{\sum_{i=1}^{n}\gamma_i(x_i-\widehat{\mu}_2)^2}{\sum_{i=1}^{n}\gamma_i},
\end{aligned} \quad (6.6.22)$$

Since γ_i is an estimate of $P[W_i = 1|\theta_0,\mathbf{x}]$, the average $n^{-1}\sum_{i=1}^{n}\gamma_i$ is an estimate of $\pi = P[W_i = 1]$. This average is our estimate of $\widehat{\pi}$.

EXERCISES

6.6.1. Rao (page 368, 1973) considers a problem in the estimation of linkages in genetics. McLachlan and Krishnan (1997) also discuss this problem and we present their model. For our purposes it can be described as a multinomial model with the four categories C_1, C_2, C_3 and C_4. For a sample of size n, let $\mathbf{X} = (X_1, X_2, X_3, X_4)'$ denote the observed frequencies of the four categories. Hence, $n = \sum_{i=1}^{4} X_i$. The probability model is

C_1	C_2	C_3	C_4
$\frac{1}{2}+\frac{1}{4}\theta$	$\frac{1}{4}-\frac{1}{4}\theta$	$\frac{1}{4}-\frac{1}{4}\theta$	$\frac{1}{4}\theta$

where the parameter θ satisfies $0 \le \theta \le 1$. In this exercise, we obtain the mle of θ.

(a) Show that likelihood function is given by

$$L(\theta|\mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!}\left[\frac{1}{2}+\frac{1}{4}\theta\right]^{x_1}\left[\frac{1}{4}-\frac{1}{4}\theta\right]^{x_2+x_3}\left[\frac{1}{4}\theta\right]^{x_4}. \quad (6.6.23)$$

(b) Show that the log of the likelihood function can be expressed as a constant (not involving parameters) plus the term

$$x_1\log[2+\theta] + [x_2+x_3]\log[1-\theta] + x_4\log\theta.$$

(c) Obtain the partial of the last expression, set the result to 0, and solve for the mle. (This will result in a quadratic equation which has one positive and one negative root.)

6.6.2. In this exercise, we set up an EM Algorithm to determine the mle for the situation described in Exercise 6.6.1. Split category C_1 into the two subcategories C_{11} and C_{12} with probabilities 1/2 and $\theta/4$, respectively. Let Z_{11} and Z_{12} denote the respective "frequencies." Then $X_1 = Z_{11} + Z_{12}$. Of course, we cannot observe Z_{11} and Z_{12}. Let $\mathbf{Z} = (Z_{11}, Z_{12})'$.

(a) Obtain the complete likelihood $L^c(\theta|\mathbf{x},\mathbf{z})$.

(b) Using the last result and (6.6.23), show that the conditional pmf $k(\mathbf{z}|\theta,\mathbf{x})$ is binomial with parameters x_1 and probability of success $\theta/(2+\theta)$.

(c) Obtain the E step of the EM Algorithm given an initial estimate $\widehat{\theta}^{(0)}$ of θ. That is, obtain

$$Q(\theta|\widehat{\theta}^{(0)},\mathbf{x}) = E_{\widehat{\theta}^{(0)}}[\log L^c(\theta|\mathbf{x},\mathbf{Z})|\widehat{\theta}^{(0)},\mathbf{x}].$$

Recall that this expectation is taken using the conditional pmf $k(\mathbf{z}|\widehat{\theta}^{(0)},\mathbf{x})$. Keep in mind the next step; i.e., we need only terms that involve θ.

(d) For the M step of the EM Algorithm, solve the equation $\partial Q(\theta|\widehat{\theta}^{(0)},\mathbf{x})/\partial\theta = 0$. Show that the solution is

$$\widehat{\theta}^{(1)} = \frac{x_1\widehat{\theta}^{(0)} + 2x_4 + x_4\widehat{\theta}^{(0)}}{n\widehat{\theta}^{(0)} + 2(x_2+x_3+x_4)}. \quad (6.6.24)$$

6.6.3. For the setup of Exercise 6.6.2, show that the following estimator of θ is unbiased.

$$\widetilde{\theta} = n^{-1}(X_1 - X_2 - X_3 + X_4). \quad (6.6.25)$$

6.6.4. Rao (page 368, 1973) presents data for the situation described in Exercise 6.6.1. The observed frequencies are: $\mathbf{x} = (125, 18, 20, 34)'$.

(a) Using computational packages, (eg, R or S-PLUS), with (6.6.25) as the initial estimate, write a program that obtains the stepwise EM estimates $\widehat{\theta}^{(k)}$.

(b) Using the data from Rao, compute the EM estimate of θ with your program. List the sequence of EM estimates, $\{\widehat{\theta}^k\}$, that you obtained. Did your sequence of estimates converge?

(c) Show that the mle using the likelihood approach in Exercise 6.6.1 is the positive root of the equation: $197\theta^2 - 15\theta - 68 = 0$. Compare it with your EM solution. They should be the same within roundoff error.

6.6.5. Suppose $X_1, X_2, \ldots, X_{n_1}$ are a random sample from a $N(\theta, 1)$ distribution. Suppose $Z_1, Z_2, \ldots, Z_{n_2}$ are missing observations. Show that the first step EM estimate is

$$\widehat{\theta}^{(1)} = \frac{n_1\overline{x} + n_2\widehat{\theta}^{(0)}}{n},$$

where $\widehat{\theta}^{(0)}$ is an initial estimate of θ and $n = n_1 + n_2$. Note that if $\widehat{\theta}^{(0)} = \overline{x}$, then $\widehat{\theta}^{(k)} = \overline{x}$ for all k.

6.6.6. Consider the situation described in Example 6.6.1. But suppose we have left censoring. That is, if $Z_1, Z_2, \ldots, Z_{n_2}$ are the censored items then all we know is that each $Z_j < a$. Obtain the EM Algorithm estimate of θ.

6.6.7. Suppose the following data follow the model of Example 6.6.1.

2.01	0.74	0.68	1.50+	1.47	1.50+	1.50+	1.52
0.07	-0.04	-0.21	0.05	-0.09	0.67	0.14	

where the superscript + denotes that the observation was censored at 1.50. Write a computer program to obtain the EM Algorithm estimate of θ.

6.6.8. The following data are observations of the random variable $X = (1-W)Y_1 + WY_2$, where W has a Bernoulli distribution with probability of success 0.70; Y_1 has a $N(100, 20^2)$ distribution; Y_2 has a $N(120, 25^2)$ distribution; W and Y_1 are independent; and W and Y_2 are independent.

119.0	96.0	146.2	138.6	143.4	98.2	124.5
114.1	136.2	136.4	184.8	79.8	151.9	114.2
145.7	95.9	97.3	136.4	109.2	103.2	

Program the EM Algorithm for this mixing problem as discussed at the end of the section. Use a dotplot to obtain initial estimates of the parameters. Compute the estimates. How close are they to the true parameters?

Chapter 7

Sufficiency

7.1 Measures of Quality of Estimators

In Chapter 6 we presented procedures for finding point estimates, interval estimates, and tests of statistical hypotheses based on likelihood theory. In this and the next chapter, we present some optimal point estimates and tests for certain situations. We first consider point estimation.

In this chapter, as in Chapter 6, we find it convenient to use the letter f to denote a pmf as well as a pdf. It will be clear from the context whether we are discussing the distributions of discrete or continuous random variables.

Suppose $f(x;\theta)$ for $\theta \in \Omega$ is the pdf (pmf) of a continuous (discrete) random variable X. Consider a point estimator $Y_n = u(X_1, \ldots, X_n)$ based on a sample $X_1, \ldots, X_n$. In Chapter 4, we discussed several properties of point estimators. Recall that Y_n is a consistent estimator (Definition 4.2.2) of θ if Y_n converges to θ in probability; i.e., Y_n is close to θ for large sample sizes. This is definitely a desirable property of a point estimator. Under suitable conditions, Theorem 6.1.3 shows that the maximum likelihood estimator is consistent. Another property was unbiasedness, (Definition 4.1.1), which says that Y_n is an unbiased estimator of θ if $E(Y_n) = \theta$. Recall that maximum likelihood estimators may not be unbiased; although, generally they are asymptotically unbiased, (see Theorem 6.2.2).

If two estimators of θ are unbiased, it would seem that we would choose the one with the smaller variance. This would be especially true if they were both approximately normal because by (5.4.3) the one with the smaller variance would tend to produce shorter asymptotic confidence intervals for θ. This leads to the following definition:

Definition 7.1.1. *For a given positive integer n, $Y = u(X_1, X_2, \ldots, X_n)$ will be called a* **minimum variance unbiased estimator**, *(MVUE), of the parameter θ, if Y is unbiased, that is, $E(Y) = \theta$, and if the variance of Y is less than or equal to the variance of every other unbiased estimator of θ.*

Example 7.1.1. As an illustration, let $X_1, X_2, \ldots, X_9$ denote a random sample from a distribution that is $N(\theta, \sigma^2)$, where $-\infty < \theta < \infty$. Because the statistic

There are two additional observations about decision rules and loss functions that should be made at this point. First, since Y is a statistic, the decision rule $\delta(Y)$ is also a statistic, and we could have started directly with a decision rule based on the observations in a random sample, say $\delta_1(X_1, X_2, \ldots, X_n)$. The risk function is then given by

$$\begin{aligned} R(\theta, \delta_1) &= E\{\mathcal{L}[\theta, \delta_1(X_1, \ldots, X_n)]\} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{L}[\theta, \delta_1(x_1, \ldots, x_n)] f(x_1;\theta) \cdots f(x_n;\theta)\, dx_1 \cdots dx_n \end{aligned}$$

if the random sample arises from a continuous-type distribution. We did not do this because, as you will see in this chapter, it is rather easy to find a good statistic, say Y, upon which to base all of the statistical inferences associated with a particular model. Thus we thought it more appropriate to start with a statistic that would be familiar, like the mle $Y = \overline{X}$ in Example 7.1.2. The second decision rule of that example could be written $\delta_2(X_1, X_2, \ldots, X_n) = 0$, a constant no matter what values of $X_1, X_2, \ldots, X_n$ are observed.

The second observation is that we have only used one loss function, namely the *squared-error loss function* $\mathcal{L}(\theta, \delta) = (\theta - \delta)^2$. The *absolute-error loss function* $\mathcal{L}(\theta, \delta) = |\theta - \delta|$ is another popular one. The loss function defined by

$$\begin{aligned} \mathcal{L}(\theta,\delta) &= 0, \quad |\theta - \delta| \le a, \\ &= b, \quad |\theta - \delta| > a, \end{aligned}$$

where a and b are positive constants, is sometimes referred to as the *goal post loss function*. The reason for this terminology is that football fans recognize that it is similar to kicking a field goal: There is no loss (actually a three-point gain) if within a units of the middle but b units of loss (zero points awarded) if outside that restriction. In addition, loss functions can be asymmetric as well as symmetric as the three previous ones have been. That is, for example, it might be more costly to underestimate the value of θ than to overestimate it. (Many of us think about this type of loss function when estimating the time it takes us to reach an airport to catch a plane.) Some of these loss functions are considered when studying Bayesian estimates in Chapter 11.

Let us close this section with an interesting illustration that raises a question leading to the likelihood principle which many statisticians believe is a quality characteristic that estimators should enjoy. Suppose that two statisticians, A and B, observe 10 independent trials of a random experiment ending in success or failure. Let the probability of success on each trial be θ, where $0 < \theta < 1$. Let us say that each statistician observes one success in these 10 trials. Suppose however, that A had decided to take $n = 10$ such observations in advance and found only one success while B had decided to take as many observations as needed to get the first success, which happened on the $10th$ trial. The model of A is that Y is $b(n = 10, \theta)$ and $y = 1$ is observed. On the other hand, B is considering the random variable Z that has a geometric pdf $g(z) = (1-\theta)^{z-1}\theta,\ z = 1, 2, 3, \ldots$, and $z = 10$ is observed. In either case, the relative frequency of success is

$$\frac{y}{n} = \frac{1}{z} = \frac{1}{10},$$

which could be used as an estimate of θ.

Let us observe, however, that one of the corresponding estimators, Y/n and $1/Z$, is biased. We have

$$E\left(\frac{Y}{10}\right) = \frac{1}{10}E(Y) = \frac{1}{10}(10\theta) = \theta$$

while

$$\begin{aligned} E\left(\frac{1}{Z}\right) &= \sum_{z=1}^{\infty} \frac{1}{z}(1-\theta)^{z-1}\theta \\ &= \theta + \tfrac{1}{2}(1-\theta)\theta + \tfrac{1}{3}(1-\theta)^2\theta + \cdots > \theta. \end{aligned}$$

That is, $1/Z$ is a biased estimator while $Y/10$ is unbiased. Thus A is using an unbiased estimator while B is not. Should we adjust B's estimator so that it, too, is unbiased?

It is interesting to note that if we maximize the two respective likelihood functions, namely

$$L_1(\theta) = \binom{10}{y}\theta^y(1-\theta)^{10-y}$$

and

$$L_2(\theta) = (1-\theta)^{z-1}\theta,$$

with $n = 10$, $y = 1$, and $z = 10$, we get exactly the same answer, $\hat{\theta} = \frac{1}{10}$. This must be the case, because in each situation we are maximizing $(1-\theta)^9\theta$. Many statisticians believe that this is the way it should be and accordingly adopt the *likelihood principle:*

Suppose two different sets of data from possibly two different random experiments lead to respective likelihood ratios, $L_1(\theta)$ and $L_2(\theta)$, that are proportional to each other. These two data sets provide the same information about the parameter θ and a statistician should obtain the same estimate of θ from either.

In our special illustration, we note that $L_1(\theta) \propto L_2(\theta)$, and the likelihood principle states that statisticians A and B should make the same inference. Thus believers in the likelihood principle would not adjust the second estimator to make it unbiased.

EXERCISES

* **7.1.1.** Show that the mean $\overline{X}$ of a random sample of size n from a distribution having pdf $f(x;\theta) = (1/\theta)e^{-(x/\theta)},\ 0 < x < \infty,\ 0 < \theta < \infty$, zero elsewhere, is an unbiased estimator of θ and has variance θ^2/n.

* **7.1.2.** Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution with mean zero and variance θ, $0 < \theta < \infty$. Show that $\sum_1^n X_i^2/n$ is an unbiased estimator of θ and has variance $2\theta^2/n$.

∗ **7.1.3.** Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size 3 from the uniform distribution having pdf $f(x;\theta) = 1/\theta$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere. Show that $4Y_1$, $2Y_2$, and $\frac{4}{3}Y_3$ are all unbiased estimators of θ. Find the variance of each of these unbiased estimators.

∗ **7.1.4.** Let Y_1 and Y_2 be two independent unbiased estimators of θ. Assume that the variance of Y_1 is twice the variance of Y_2. Find the constants k_1 and k_2 so that $k_1Y_1 + k_2Y_2$ is an unbiased estimator with smallest possible variance for such a linear combination.

7.1.5. In Example 7.1.2 of this section, take $\mathcal{L}[\theta, \delta(y)] = |\theta - \delta(y)|$. Show that $R(\theta, \delta_1) = \frac{1}{5}\sqrt{2/\pi}$ and $R(\theta, \delta_2) = |\theta|$. Of these two decision functions δ_1 and δ_2, which yields the smaller maximum risk?

7.1.6. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution with parameter θ, $0 < \theta < \infty$. Let $Y = \sum_1^n X_i$ and let $\mathcal{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. If we restrict our considerations to decision functions of the form $\delta(y) = b + y/n$, where b does not depend on y, show that $R(\theta, \delta) = b^2 + \theta/n$. What decision function of this form yields a uniformly smaller risk than every other decision function of this form? With this solution, say δ, and $0 < \theta < \infty$, determine $\max_\theta R(\theta, \delta)$ if it exists.

7.1.7. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\mu, \theta)$, $0 < \theta < \infty$, where μ is unknown. Let $Y = \sum_1^n (X_i - \overline{X})^2/n = V$ and let $\mathcal{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. If we consider decision functions of the form $\delta(y) = by$, where b does not depend upon y, show that $R(\theta, \delta) = (\theta^2/n^2)[(n^2-1)b^2 - 2n(n-1)b + n^2]$. Show that $b = n/(n+1)$ yields a minimum risk decision functions of this form. Note that $nY/(n+1)$ is not an unbiased estimator of θ. With $\delta(y) = ny/(n+1)$ and $0 < \theta < \infty$, determine $\max_\theta R(\theta, \delta)$ if it exists.

7.1.8. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $b(1, \theta)$, $0 \le \theta \le 1$. Let $Y = \sum_1^n X_i$ and let $\mathcal{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. Consider decision functions of the form $\delta(y) = by$, where b does not depend upon y. Prove that $R(\theta, \delta) = b^2 n\theta(1-\theta) + (bn-1)^2\theta^2$. Show that

$$\max_\theta R(\theta, \delta) = \frac{b^4n^2}{4[b^2n - (bn - -1)^2]},$$

provided that the value b is such that $b^2n \ge 2(bn-1)^2$. Prove that $b = 1/n$ does not maximize $\max_\theta R(\theta, \delta)$.

∗ **7.1.9.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean $\theta > 0$.

(a) Statistician A observes the sample to be the values $x_1, x_2, \ldots, x_n$ with sum $y = \sum x_i$. Find the mle of θ.

(b) Statistician B loses the sample values $x_1, x_2, \ldots, x_n$ but remembers the sum y_1 and the fact that the sample arose from a Poisson distribution. Thus B decides to create some fake observations which he calls $z_1, z_2, \ldots, z_n$ (as he knows they will probably not equal the original x-values) as follows. He notes that the conditional probability of independent Poisson random variables $Z_1, Z_2, \ldots, Z_n$ being equal to $z_1, z_2, \ldots, z_n$, given $\sum z_i = y_1$ is

$$\frac{\frac{\theta^{z_1}e^{-\theta}}{z_1!}\frac{\theta^{z_2}e^{-\theta}}{z_2!}\cdots\frac{\theta^{z_n}e^{-\theta}}{z_n!}}{\frac{(n\theta)^{y_1}e^{n\theta}}{y_1!}} = \frac{y_1!}{z_1!z_2!\cdots z_n!}\left(\frac{1}{n}\right)^{z_1}\left(\frac{1}{n}\right)^{z_2}\cdots\left(\frac{1}{n}\right)^{z_n}$$

since $Y_1 = \sum Z_i$ has a Poisson distribution with mean $n\theta$. The latter distribution is multinomial with y_1 independent trials, each terminating in one of n mutually exclusive and exhaustive ways, each of which has the same probability $1/n$. Accordingly, B runs such a multinomial experiment y_1 independent trials and obtains $z_1, z_2, \ldots, z_n$. Find the likelihood function using these z-values. Is it proportional to that of statistician A?
Hint: Here the likelihood function is the product of this conditional pdf and the pdf of $Y_1 = \sum Z_i$.

7.2 A Sufficient Statistic for a Parameter

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a distribution that has pdf $f(x;\theta)$, $\theta \in \Omega$. In Chapters 4 and 6 we constructed statistics to make statistical inferences as illustrated by point and interval estimation and tests of statistical hypotheses. We note that a statistic, for example, $Y = u(X_1, X_2, \ldots, X_n)$, is a form of data reduction. To illustrate, instead of listing all of the individual observations $X_1, X_2, \ldots, X_n$, we might prefer to give only the sample mean $\overline{X}$ or the sample variance S^2. Thus statisticians look for ways of reducing a set of data so that these data can be more easily understood without losing the meaning associated with the entire set of observations.

It is interesting to note that a statistic $Y = u(X_1, X_2, \ldots, X_n)$ really partitions the sample space of $X_1, X_2, \ldots, X_n$. For illustration, suppose we say that the sample was observed and $\overline{x} = 8.32$. There are many points in the sample space which have that same mean of 8.32, and we can consider them as belonging to the set $\{(x_1, x_2, \ldots, x_n) : \overline{x} = 8.32\}$. As a matter of fact, all points on the hyperplane

$$x_1 + x_2 + \cdots + x_n = (8.32)n$$

yield the mean of $\overline{x} = 8.32$, so this hyperplane is the set. However, there are many values that $\overline{X}$ can take and thus there are many such sets. So, in this sense, the sample mean $\overline{X}$, or any statistic $Y = u(X_1, X_2, \ldots, X_n)$, partitions the sample space into a collection of sets.

Often in the study of statistics the parameter θ of the model is unknown; thus we need to make some statistical inference about it. In this section we consider a statistic denoted by $Y_1 = u_1(X_1, X_2, \ldots, X_n)$, which we call a *sufficient statistic* because

$$2\sum_{i=1}^{n}(x_i-\overline{x})(\overline{x}-\theta)=2(\overline{x}-\theta)\sum_{i=1}^{n}(x_i-\overline{x})=0.$$

Thus the joint pdf of $X_1, X_2, \ldots, X_n$ may be written

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n}(x_i-\theta)^2/2\sigma^2\right]$$

$$=\{\exp[-n(\overline{x}-\theta)^2/2\sigma^2]\}\left\{\frac{\exp\left[-\sum_{i=1}^{n}(x_i-\overline{x})^2/2\sigma^2\right]}{(\sigma\sqrt{2\pi})^n}\right\}$$

Because the first factor of the right-hand member of this equation depends upon $x_1, x_2, \ldots, x_n$ only through $\overline{x}$, and the second factor does not depend upon θ, the factorization theorem implies that the mean $\overline{X}$ of the sample is, for any particular value of σ^2, a sufficient statistic for θ, the mean of the normal distribution. ■

We could have used the definition in the preceding example because we know that $\overline{X}$ is $N(\theta, \sigma^2/n)$. Let us now consider an example in which the use of the definition is inappropriate.

Example 7.2.5. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with pdf

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1} & 0<x<1\\ 0 & \text{elsewhere,}\end{cases}$$

where $0<\theta$. Using the factorization theorem, we will show that the product $u_1(X_1, X_2, \ldots, X_n)=\prod_{i=1}^{n}X_i$ is a sufficient statistic for θ. The joint pdf of $X_1, X_2, \ldots, X_n$ is

$$\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}=\left[\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta}\right]\left(\frac{1}{\prod_{i=1}^{n}x_i}\right),$$

where $0<x_i<1$, $i=1,2,\ldots,n$. In the factorization theorem let

$$k_1[u_1(x_1,x_2,\ldots,x_n);\theta]=\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta}$$

and

$$k_2(x_1,x_2,\ldots,x_n)=\frac{1}{\prod_{i=1}^{n}x_i}.$$

Since $k_2(x_1,x_2,\ldots,x_n)$ does not depend upon θ, the product $\prod_{i=1}^{n}X_i$ is a sufficient statistic for θ. ■

There is a tendency for some readers to apply incorrectly the factorization theorem in those instances in which the domain of positive probability density depends upon the parameter θ. This is due to the fact that they do not give proper consideration to the domain of the function $k_2(x_1, x_2, \ldots, x_n)$. This will be illustrated in the next example.

Example 7.2.6. In Example 7.2.3 with $f(x;\theta)=e^{-(x-\theta)}I_{(\theta,\infty)}(x)$, it was found that the first order statistic Y_1 is a sufficient statistic for θ. To illustrate our point about not considering the domain of the function, take $n=3$ and note that

$$e^{-(x_1-\theta)}e^{-(x_2-\theta)}e^{-(x_3-\theta)}=[e^{-3\max x_i+3\theta}][e^{-x_1-x_2-x_3+3\max x_i}]$$

or a similar expression. Certainly, in the latter formula, there is no θ in the second factor and it might be assumed that $Y_3=\max X_i$ is a sufficient statistic for θ. Of course, this is incorrect because we should have written the joint pdf of X_1, X_2, X_3 as

$$\prod_{i=1}^{3}[e^{-(x_i-\theta)}I_{(\theta,\infty)}(x_i)]=[e^{3\theta}I_{(\theta,\infty)}(\min x_i)]\left[\exp\left\{-\sum_{i=1}^{3}x_i\right\}\right]$$

because $I_{(\theta,\infty)}(\min x_i)=I_{(\theta,\infty)}(x_1)I_{(\theta,\infty)}(x_2)I_{(\theta,\infty)}(x_3)$. A similar statement cannot be made with $\max x_i$. Thus $Y_1=\min X_i$ is the sufficient statistic for θ, not $Y_3=\max X_i$. ■

EXERCISES

7.2.1. Let $X_1, X_2, \ldots, X_n$ be iid $N(0,\theta)$, $0<\theta<\infty$. Show that $\sum_1^n X_i^2$ is a sufficient statistic for θ.

7.2.2. Prove that the sum of the observations of a random sample of size n from a Poisson distribution having parameter θ, $0<\theta<\infty$, is a sufficient statistic for θ.

7.2.3. Show that the nth order statistic of a random sample of size n from the uniform distribution having pdf $f(x;\theta)=1/\theta$, $0<x<\theta$, $0<\theta<\infty$, zero elsewhere, is a sufficient statistic for θ. Generalize this result by considering the pdf $f(x;\theta)=Q(\theta)M(x)$, $0<x<\theta$, $0<\theta<\infty$, zero elsewhere. Here, of course,

$$\int_0^\theta M(x)\,dx=\frac{1}{Q(\theta)}.$$

7.2.4. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a geometric distribution that has pmf $f(x;\theta)=(1-\theta)^x\theta$, $x=0,1,2,\ldots$, $0<\theta<1$, zero elsewhere. Show that $\sum_1^n X_i$ is a sufficient statistic for θ.

7.2.5. Show that the sum of the observations of a random sample of size n from a gamma distribution that has pdf $f(x;\theta)=(1/\theta)e^{-x/\theta}$, $0<x<\infty$, $0<\theta<\infty$, zero elsewhere, is a sufficient statistic for θ.

* **7.2.6.** Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a beta distribution with parameters $\alpha = \theta$ and $\beta = 2$. Show that the product $X_1 X_2 \cdots X_n$ is a sufficient statistic for θ.

* **7.2.7.** Show that the product of the sample observations is a sufficient statistic for $\theta > 0$ if the random sample is taken from a gamma distribution with parameters $\alpha = \theta$ and $\beta = 6$.

* **7.2.8.** What is the sufficient statistic for θ if the sample arises from a beta distribution in which $\alpha = \beta = \theta > 0$?

* **7.2.9.** We consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with pdf $f(x;\theta) = (1/\theta)\exp(-x/\theta)$, $0 < x < \infty$, zero elsewhere, where $0 < \theta$. Possibly, in a life testing situation, however, we only observe the first r order statistics $Y_1 < Y_2 < \cdots < Y_r$.

  (a) Record the joint pdf of these order statistics and denote it by $L(\theta)$.

  (b) Under these conditions, find the mle, $\hat{\theta}$, by maximizing $L(\theta)$.

  (c) Find the mgf and pdf of $\hat{\theta}$.

  (d) With a slight extension of the definition of sufficiency, is $\hat{\theta}$ a sufficient statistic?

7.3 Properties of a Sufficient Statistic

Suppose $X_1, X_2, \ldots, X_n$ is a random sample on a random variable with pdf or pmf $f(x;\theta)$ where $\theta \in \Omega$. In this section we discuss how sufficiency is used to determine MVUEs. First note that a sufficient estimate is not unique in any sense. For if $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is a sufficient statistic and $Y_2 = g(Y_1)$ where $g(x)$ is a one-to-one function is a statistic then

$$\begin{aligned} f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta) &= k_1[u_1(y_1);\theta]k_2(x_1, x_2, \ldots, x_n) \\ &= k_1[u_1(g^{-1}(y_2));\theta]k_2(x_1, x_2, \ldots, x_n); \end{aligned}$$

hence, by the factorization theorem Y_2 is also sufficient. However, as the theorem below shows, sufficiency can lead to a best point estimate.

We first refer back to Theorem 2.3.1 of Section 2.3: If X_1 and X_2 are random variables such that the variance of X_2 exists, then

$$E[X_2] = E[E(X_2|X_1)]$$

and

$$\text{var}(X_2) \geq \text{var}[E(X_2|X_1)].$$

For the adaptation in context of sufficient statistics, we let the sufficient statistic Y_1 be X_1 and Y_2, an unbiased statistic of θ, be X_2. Thus, with $E(Y_2|y_1) = \varphi(y_1)$, we have

$$\theta = E(Y_2) = E[\varphi(Y_1)]$$

and

$$\text{var}(Y_2) \geq \text{var}[\varphi(Y_1)].$$

That is, through this conditioning, the function $\varphi(Y_1)$ of the sufficient statistic Y_1 is an unbiased estimator of θ having a smaller variance than that of the unbiased estimator Y_2. We summarize this discussion more formally in the following theorem, which can be attributed to Rao and Blackwell.

Theorem 7.3.1 (Rao-Blackwell). *Let $X_1, X_2, \ldots, X_n$, n a fixed positive integer, denote a random sample from a distribution (continuous or discrete) that has pdf or pmf $f(x;\theta)$, $\theta \in \Omega$. Let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for θ, and let $Y_2 = u_2(X_1, X_2, \ldots, X_n)$, not a function of Y_1 alone, be an unbiased estimator of θ. Then $E(Y_2|y_1) = \varphi(y_1)$ defines a statistic $\varphi(Y_1)$. This statistic $\varphi(Y_1)$ is a function of the sufficient statistic for θ; it is an unbiased estimator of θ; and its variance is less than that of Y_2.*

This theorem tells us that in our search for an MVUE of a parameter, we may, if a sufficient statistic for the parameter exists, restrict that search to functions of the sufficient statistic. For if we begin with an unbiased estimator Y_2 alone, then we can always improve on this by computing $E(Y_2|y_1) = \varphi(y_1)$ so that $\varphi(Y_1)$ is an unbiased estimator with smaller variance than that of Y_2.

After Theorem 7.3.1 many students believe that it is necessary to find first some unbiased estimator Y_2 in their search for $\varphi(Y_1)$, an unbiased estimator of θ based upon the sufficient statistic Y_1. This is not the case at all, and Theorem 7.3.1 simply convinces us that we can restrict our search for a best estimator to functions of Y_1. Furthermore, there is a connection between sufficient statistics and maximum likelihood estimates as the following theorem shows,

Theorem 7.3.2. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has pdf or pmf $f(x;\theta)$, $\theta \in \Omega$. If a sufficient statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ for θ exists and if a maximum likelihood estimator $\hat{\theta}$ of θ also exists uniquely, then $\hat{\theta}$ is a function of $Y_1 = u_1(X_1, X_2, \ldots, X_n)$.*

Proof. Let $f_{Y_1}(y_1;\theta)$ be the pdf or pmf of Y_1. Then by the definition of sufficiency, the likelihood function

$$\begin{aligned} L(\theta; x_1, x_2, \ldots, x_n) &= f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta) \\ &= f_{Y_1}[u_1(x_1, x_2, \ldots, x_n);\theta]H(x_1, x_2, \ldots, x_n), \end{aligned}$$

where $H(x_1, x_2, \ldots, x_n)$ does not depend upon θ. Thus L and f_{Y_1}, as functions of θ, are maximized simultaneously. Since there is one and only one value of θ

zero elsewhere. Thus

$$\begin{aligned} E\left(\frac{Y_1}{3}\middle| y_3\right) &= E\left(\frac{Y_1 - Y_3}{3}\middle| y_3\right) + E\left(\frac{Y_3}{3}\middle| y_3\right) \\ &= \left(\frac{1}{3}\right)\int_{y_3}^{\infty}\left(\frac{1}{\theta}\right)^2 (y_1 - y_3)^2 e^{-(y_1-y_3)/\theta}\, dy_1 + \frac{y_3}{3} \\ &= \left(\frac{1}{3}\right)\frac{\Gamma(3)\theta^3}{\theta^2} + \frac{y_3}{3} = \frac{2\theta}{3} + \frac{y_3}{3} = \Upsilon(y_3). \end{aligned}$$

Of course, $E[\Upsilon(Y_3)] = \theta$ and $\text{var}[\Upsilon(Y_3)] \le \text{var}(Y_1/3)$, but $\Upsilon(Y_3)$ is not a statistic as it involves θ and cannot be used as an estimator of θ. This illustrates the preceding remark. ■

EXERCISES

★ **7.3.1.** In each of the Exercises 7.2.1, 7.2.2, 7.2.3, and 7.2.4, show that the mle of θ is a function of the sufficient statistic for θ.

★ **7.3.2.** Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics of a random sample of size 5 from the uniform distribution having pdf $f(x;\theta) = 1/\theta$, $0 < x < \theta$, $0 < \theta < \infty$, zero elsewhere. Show that $2Y_3$ is an unbiased estimator of θ. Determine the joint pdf of Y_3 and the sufficient statistic Y_5 for θ. Find the conditional expectation $E(2Y_3|y_5) = \varphi(y_5)$. Compare the variances of $2Y_3$ and $\varphi(Y_5)$.

★ **7.3.3.** If X_1, X_2 is a random sample of size 2 from a distribution having pdf $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere, find the joint pdf of the sufficient statistic $Y_1 = X_1 + X_2$ for θ and $Y_2 = X_2$. Show that Y_2 is an unbiased estimator of θ with variance θ^2. Find $E(Y_2|y_1) = \varphi(y_1)$ and the variance of $\varphi(Y_1)$.

★ **7.3.4.** Let $f(x,y) = (2/\theta^2)e^{-(x+y)/\theta}$, $0 < x < y < \infty$, zero elsewhere, be the joint pdf of the random variables X and Y.

(a) Show that the mean and the variance of Y are, respectively, $3\theta/2$ and $5\theta^2/4$.

(b) Show that $E(Y|x) = x + \theta$. In accordance with the theory, the expected value of $X + \theta$ is that of Y, namely, $3\theta/2$, and the variance of $X + \theta$ is less than that of Y. Show that the variance of $X + \theta$ is in fact $\theta^2/4$.

★ **7.3.5.** In each of Exercises 7.2.1, 7.2.2, and 7.2.3, compute the expected value of the given sufficient statistic and, in each case, determine an unbiased estimator of θ that is a function of that sufficient statistic alone.

★ **7.3.6.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean θ. Find the conditional expectation $E\left(X_1 + 2X_2 + 3X_3 \middle| \sum_1^n X_i\right)$.

7.4 Completeness and Uniqueness

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution that has pmf

$$f(x;\theta) = \begin{cases} \frac{\theta^x e^{-\theta}}{x!} & x = 0,1,2,\ldots;\ 0 < \theta \\ 0 & \text{elsewhere.} \end{cases}$$

From Exercise 7.2.2, we know that $Y_1 = \sum_{i=1}^n X_i$ is a sufficient statistic for θ and its pmf is

$$g_1(y_1;\theta) = \begin{cases} \frac{(n\theta)^{y_1} e^{-n\theta}}{y_1!} & y_1 = 0,1,2,\ldots \\ 0 & \text{elsewhere.} \end{cases}$$

Let us consider the family $\{g_1(y_1;\theta) : 0 < \theta\}$ of probability mass functions. Suppose that the function $u(Y_1)$ of Y_1 is such that $E[u(Y_1)] = 0$ for every $\theta > 0$. We shall show that this requires $u(y_1)$ to be zero at every point $y_1 = 0, 1, 2, \ldots$. That is, $E[u(Y_1)] = 0$ for $0 < \theta$ requires

$$0 = u(0) = u(1) = u(2) = u(3) = \cdots.$$

We have for all $\theta > 0$ that

$$\begin{aligned} 0 = E[u(Y_1)] &= \sum_{y_1=0}^{\infty} u(y_1)\frac{(n\theta)^{y_1}e^{-n\theta}}{y_1!} \\ &= e^{-n\theta}\left[u(0) + u(1)\frac{n\theta}{1!} + u(2)\frac{(n\theta)^2}{2!} + \cdots\right]. \end{aligned}$$

Since $e^{-n\theta}$ does not equal zero, we have shown that

$$0 = u(0) + [nu(1)]\theta + \left[\frac{n^2u(2)}{2}\right]\theta^2 + \cdots.$$

However, if such an infinite (power) series converges to zero for all $\theta > 0$, then each of the coefficients must equal zero. That is,

$$u(0) = 0, \quad nu(1) = 0, \quad \frac{n^2u(2)}{2} = 0, \ldots$$

and thus $0 = u(0) = u(1) = u(2) = \cdots$, as we wanted to show. Of course, the condition $E[u(Y_1)] = 0$ for all $\theta > 0$ does not place any restriction on $u(y_1)$ when y_1 is not a nonnegative integer. So we see that, in this illustration, $E[u(Y_1)] = 0$ for all $\theta > 0$ requires that $u(y_1)$ equals zero except on a set of points that has probability zero for each pmf $g_1(y_1;\theta)$, $0 < \theta$. From the following definition we observe that the family $\{g_1(y_1;\theta) : 0 < \theta\}$ is complete.

Definition 7.4.1. *Let the random variable Z of either the continuous type or the discrete type have a pdf or pmf that is one member of the family $\{h(z;\theta) : \theta \in \Omega\}$. If the condition $E[u(Z)] = 0$, for every $\theta \in \Omega$, requires that $u(z)$ be zero except on a set of points that has probability zero for each $h(z;\theta)$, $\theta \in \Omega$, then the family $\{h(z;\theta) : \theta \in \Omega\}$ is called a* **complete family** *of probability density or mass functions.*

Remark 7.4.1. In Section 1.8 it was noted that the existence of $E[u(X)]$ implies that the integral (or sum) converges absolutely. This absolute convergence was tacitly assumed in our definition of completeness and it is needed to prove that certain families of probability density functions are complete. ■

In order to show that certain families of probability density functions of the continuous type are complete, we must appeal to the same type of theorem in analysis that we used when we claimed that the moment generating function uniquely determines a distribution. This is illustrated in the next example.

Example 7.4.1. Consider the family of pdfs $\{h(z;\theta) : 0 < \theta < \infty\}$. Suppose Z has a pdf in this family given by

$$h(z;\theta) = \begin{cases} \frac{1}{\theta}e^{-z/\theta} & 0 < z < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Let us say that $E[u(Z)] = 0$ for every $\theta > 0$. That is,

$$\frac{1}{\theta}\int_0^\infty u(z)e^{-z/\theta}\,dz = 0, \quad \theta > 0.$$

Readers acquainted with the theory of transformations will recognize the integral in the left-hand member as being essentially the Laplace transform of $u(z)$. In that theory we learn that the only function $u(z)$ transforming to a function of θ which is identically equal to zero is $u(z) = 0$, except (in our terminology) on a set of points that has probability zero for each $h(z;\theta)$, $0 < \theta$. That is, the family $\{h(z;\theta) : 0 < \theta < \infty\}$ is complete. ■

Let the parameter θ in the pdf or pmf $f(x;\theta)$, $\theta \in \Omega$, have a sufficient statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$, where $X_1, X_2, \ldots, X_n$ is a random sample from this distribution. Let the pdf or pmf of Y_1 be $f_{Y_1}(y_1;\theta)$, $\theta \in \Omega$. It has been seen that, if there is any unbiased estimator Y_2 (not a function of Y_1 alone) of θ, then there is at least one function of Y_1 that is an unbiased estimator of θ, and our search for a best estimator of θ may be restricted to functions of Y_1. Suppose it has been verified that a certain function $\varphi(Y_1)$, not a function of θ, is such that $E[\varphi(Y_1)] = \theta$ for all values of θ, $\theta \in \Omega$. Let $\psi(Y_1)$ be another function of the sufficient statistic Y_1 alone, so that we also have $E[\psi(Y_1)] = \theta$ for all value of θ, $\theta \in \Omega$. Hence,

$$E[\varphi(Y_1) - \psi(Y_1)] = 0, \quad \theta \in \Omega.$$

If the family $\{f_{Y_1}(y_1;\theta) : \theta \in \Omega\}$ is complete, the function of $\varphi(y_1) - \psi(y_1) = 0$, except on a set of points that has probability zero. That is, for every other unbiased estimator $\psi(Y_1)$ of θ, we have

$$\varphi(y_1) = \psi(y_1)$$

except possibly at certain special points. Thus, in this sense [namely $\varphi(y_1) = \psi(y_1)$, except on a set of points with probability zero], $\varphi(Y_1)$ is the unique function of Y_1, which is an unbiased estimator of θ. In accordance with the Rao-Blackwell theorem, $\varphi(Y_1)$ has a smaller variance than every other unbiased estimator of θ. That is, the statistic $\varphi(Y_1)$ is the MVUE of θ. This fact is stated in the following theorem of Lehmann and Scheffé.

Theorem 7.4.1 (Lehmann and Scheffé). *Let $X_1, X_2, \ldots, X_n$, n a fixed positive integer, denote a random sample from a distribution that has pdf or pmf $f(x;\theta)$, $\theta \in \Omega$, let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for θ, and let the family $\{f_{Y_1}(y_1;\theta) : \theta \in \Omega\}$ be complete. If there is a function of Y_1 that is an unbiased estimator of θ, then this function of Y_1 is the unique MVUE of θ. Here "unique" is used in the sense described in the preceding paragraph.*

The statement that Y_1 is a sufficient statistic for a parameter θ, $\theta \in \Omega$, and that the family $\{f_{Y_1}(y_1;\theta) : \theta \in \Omega\}$ of probability density functions is complete is lengthy and somewhat awkward. We shall adopt the less descriptive, but more convenient, terminology that Y_1 is a *complete sufficient statistic* for θ. In the next section we study a fairly large class of probability density functions for which a complete sufficient statistic Y_1 for θ can be determined by inspection.

EXERCISES

* **7.4.1.** If $az^2 + bz + c = 0$ for more than two values of z, then $a = b = c = 0$. Use this result to show that the family $\{b(2,\theta) : 0 < \theta < 1\}$ is complete.

* **7.4.2.** Show that each of the following families is not complete by finding at least one nonzero function $u(x)$ such that $E[u(X)] = 0$, for all $\theta > 0$.

(a)

$$f(x;\theta) = \begin{cases} \frac{1}{2\theta} & -\theta < x < \theta \quad \text{where } 0 < \theta < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

(b) $N(0,\theta)$, where $0 < \theta < \infty$.

* **7.4.3.** Let $X_1, X_2, \ldots, X_n$ represent a random sample from the discrete distribution having the pmf

$$f(x;\theta) = \begin{cases} \theta^x(1-\theta)^{1-x} & x = 0, 1, \ 0 < \theta < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Show that $Y_1 = \sum_1^n X_i$ is a complete sufficient statistic for θ. Find the unique function of Y_1 that is the MVUE of θ.
Hint: Display $E[u(Y_1)] = 0$, show that the constant term $u(0)$ is equal to zero, divide both members of the equation by $\theta \neq 0$, and repeat the argument.

* **7.4.4.** Consider the family of probability density functions $\{h(z;\theta) : \theta \in \Omega\}$, where $h(z;\theta) = 1/\theta$, $0 < z < \theta$, zero elsewhere.

(a) Show that the family is complete provided that $\Omega = \{\theta : 0 < \theta < \infty\}$.
Hint: For convenience, assume that $u(z)$ is continuous and note that the derivative of $E[u(Z)]$ with respect to θ is equal to zero also.

(b) Show that this family is not complete if $\Omega = \{\theta : 1 < \theta < \infty\}$.
Hint: Concentrate on the interval $0 < z < 1$ and find a nonzero function $u(z)$ on that interval such that $E[u(Z)] = 0$ for all $\theta > 1$.

7.4.5. Show that the first order statistic Y_1 of a random sample of size n from the distribution having pdf $f(x;\theta) = e^{-(x-\theta)}$, $\theta < x < \infty$, $-\infty < \theta < \infty$, zero elsewhere, is a complete sufficient statistic for θ. Find the unique function of this statistic which is the MVUE of θ.

* **7.4.6.** Let a random sample of size n be taken from a distribution of the discrete type with pmf $f(x;\theta) = 1/\theta$, $x = 1, 2, \ldots, \theta$, zero elsewhere, where θ is an unknown positive integer.

(a) Show that the largest observation, say Y, of the sample is a complete sufficient statistic for θ.

(b) Prove that

$$[Y^{n+1} - (Y-1)^{n+1}]/[Y^n - (Y-1)^n]$$

is the unique MVUE of θ.

* **7.4.7.** Let X have the pdf $f_X(x;\theta) = 1/(2\theta)$, for $-\theta < x < \theta$, zero elsewhere, where $\theta > 0$.

(a) Is the statistic $Y = |X|$ a sufficient statistic for θ? Why?

(b) Let $f_Y(y;\theta)$ be the pdf of Y. Is the family $\{f_Y(y;\theta) : \theta > 0\}$ complete? Why?

* **7.4.8.** Let X have the pmf $p(x;\theta) = \frac{1}{2}\binom{n}{|x|}\theta^{|x|}(1-\theta)^{n-|x|}$, for $x = \pm 1, \pm 2, \ldots, \pm n$, $p(0,\theta) = (1-\theta)^n$, and zero elsewhere, where $0 < \theta < 1$.

(a) Show that this family $\{p(x;\theta) : 0 < \theta < 1\}$ is not complete.

(b) Let $Y = |X|$. Show that Y is a complete and sufficient statistic for θ.

* **7.4.9.** Let $X_1, \ldots, X_n$ be iid with pdf $f(x;\theta) = 1/(3\theta)$, $-\theta < x < 2\theta$, zero elsewhere, where $\theta > 0$.

(a) Find the mle $\widehat{\theta}$ of θ.

(b) Is $\widehat{\theta}$ a sufficient statistic for θ? Why?

(c) Is $(n+1)\widehat{\theta}/n$ the unique MVUE of θ? Why?

* **7.4.10.** Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution with pdf $f(x;\theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere. The statistic Y_n is a complete sufficient statistic for θ and it has pdf

$$g(y_n;\theta) = \frac{ny_n^{n-1}}{\theta^n}, \quad 0 < y_n < \theta,$$

and zero elsewhere.

(a) Find the distribution function $H_n(z;\theta)$ of $Z = n(\theta - Y_n)$.

(b) Find the $\lim_{n\to\infty} H_n(z;\theta)$ and thus the limiting distribution of Z.

7.5 The Exponential Class of Distributions

In this section we discuss an important class of distributions, called the *exponential class*. As we will show, this class possesses complete and sufficient statistics which are readily determined from the distribution.

Consider a family $\{f(x;\theta) : \theta \in \Omega\}$ of probability density or mass functions, where Ω is the interval set $\Omega = \{\theta : \gamma < \theta < \delta\}$, where γ and δ are known constants (they may be $\pm\infty$), and where

$$f(x;\theta) = \begin{cases} \exp[p(\theta)K(x) + S(x) + q(\theta)] & x \in \mathcal{S} \\ 0 & \text{elsewhere}, \end{cases} \tag{7.5.1}$$

where $\mathcal{S}$ is the support of X. In this section we will be concerned with a particular class of the family called the regular exponential class.

Definition 7.5.1 (Regular Exponential Class). *A pdf of the form (7.5.1) is said to be a member of the* **regular exponential class** *of probability density or mass functions if*

1. *$\mathcal{S}$, the support of X, does not depend upon θ,*
2. *$p(\theta)$ is a nontrivial continuous function of $\theta \in \Omega$,*
3. *Finally,*
   (a) *if X is a continuous random variable then each of $K'(x) \not\equiv 0$ and $S(x)$ is a continuous function of $x \in \mathcal{S}$,*
   (b) *if X is a discrete random variable then $K(x)$ is a nontrivial function of $x \in \mathcal{S}$.*

For example, each member of the family $\{f(x;\theta) : 0 < \theta < \infty\}$, where $f(x;\theta)$ is $N(0,\theta)$, represents a regular case of the exponential class of the continuous type because

$$\begin{aligned} f(x;\theta) &= \frac{1}{\sqrt{2\pi\theta}}e^{-x^2/2\theta} \\ &= \exp\left(-\frac{1}{2\theta}x^2 - \log\sqrt{2\pi\theta}\right), \quad -\infty < x < \infty. \end{aligned}$$

On the other hand, consider the uniform density function given by

$$f(x;\theta) = \begin{cases} \exp\{-\log\theta\} & x \in (0,\theta) \\ 0 & \text{elsewhere}. \end{cases}$$

This can be written in the form (7.5.1) but the support is the interval $(0,\theta)$ which depends on θ. Hence, the uniform family is not a regular exponential family.

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that represents a regular case of the exponential class. The joint pdf or pmf of $X_1, X_2, \ldots, X_n$ is

$$\exp\left[p(\theta)\sum_1^n K(x_i) + \sum_1^n S(x_i) + nq(\theta)\right].$$

Here σ^2 is any fixed positive number. This is a regular case of the exponential class with

$$p(\theta) = \frac{\theta}{\sigma^2}, \quad K(x) = x,$$
$$S(x) = -\frac{x^2}{2\sigma^2} - \log\sqrt{2\pi\sigma^2}, \quad q(\theta) = -\frac{\theta^2}{2\sigma^2}.$$

Accordingly, $Y_1 = X_1 + X_2 + \cdots + X_n = n\overline{X}$ is a complete sufficient statistic for the mean θ of a normal distribution for every fixed value of the variance σ^2. Since $E(Y_1) = n\theta$, then $\varphi(Y_1) = Y_1/n = \overline{X}$ is the only function of Y_1 that is an unbiased estimator of θ; and being a function of the sufficient statistic Y_1, it has a minimum variance. That is, $\overline{X}$ is the unique MVUE of θ. Incidentally, since Y_1 is a one-to-one function of $\overline{X}$, $\overline{X}$ itself is also a complete sufficient statistic for θ. ■

Example 7.5.3 (Example 7.5.1, continued). Reconsider the discussion concerning the Poisson distribution with parameter θ found in Example 7.5.1. Based on this discussion the statistic $Y_1 = \sum_{i=1}^n X_i$ was sufficient. It follows from Theorem 7.5.2 that its family of distributions is complete. Since $E(Y_1) = n\theta$, it follows that $\overline{X} = n^{-1}Y_1$ is the unique MVUE of θ. ■

EXERCISES

* **7.5.1.** Write the pdf

$$f(x;\theta) = \frac{1}{6\theta^4}x^3e^{-x/\theta}, \quad 0 < x < \infty, \quad 0 < \theta < \infty,$$

zero elsewhere, in the exponential form. If $X_1, X_2, \ldots, X_n$ is a random sample from this distribution, find a complete sufficient statistic Y_1 for θ and the unique function $\varphi(Y_1)$ of this statistic that is the MVUE of θ. Is $\varphi(Y_1)$ itself a complete sufficient statistic?

* **7.5.2.** Let $X_1, X_2, \ldots, X_n$ denote a random sample of size $n > 1$ from a distribution with pdf $f(x;\theta) = \theta e^{-\theta x}$, $0 < x < \infty$, zero elsewhere, and $\theta > 0$. Then $\sum_1^n X_i$ is a sufficient statistic for θ. Prove that $(n-1)/Y$ is the MVUE of θ.

* **7.5.3.** Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution with pdf $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, and $\theta > 0$.

(a) Show that the *geometric mean* $(X_1X_2\cdots X_n)^{1/n}$ of the sample is a complete sufficient statistic for θ.

(b) Find the maximum likelihood estimator of θ, and observe that it is a function of this geometric mean.

* **7.5.4.** Let $\overline{X}$ denote the mean of the random sample $X_1, X_2, \ldots, X_n$ from a gamma-type distribution with parameters $\alpha > 0$ and $\beta = \theta > 0$. Compute $E[X_1|\overline{x}]$.
Hint: Can you find directly a function $\psi(\overline{X})$ of $\overline{X}$ such that $E[\psi(\overline{X})] = \theta$? Is $E(X_1|\overline{x}) = \psi(\overline{x})$? Why?

7.5.5. Let X be a random variable with pdf of a regular case of the exponential class. Show that $E[K(X)] = -q'(\theta)/p'(\theta)$, provided these derivatives exist, by differentiating both members of the equality

$$\int_a^b \exp[p(\theta)K(x) + S(x) + q(\theta)]\,dx = 1$$

with respect to θ. By a second differentiation, find the variance of $K(X)$.

7.5.6. Given that $f(x;\theta) = \exp[\theta K(x) + S(x) + q(\theta)]$, $a < x < b$, $\gamma < \theta < \delta$, represents a regular case of the exponential class, show that the moment-generating function $M(t)$ of $Y = K(X)$ is $M(t) = \exp[q(\theta) - q(\theta + t)]$, $\gamma < \theta + t < \delta$.

7.5.7. In the preceding exercise, given that $E(Y) = E[K(X)] = \theta$, prove that Y is $N(\theta, 1)$.
Hint: Consider $M'(0) = \theta$ and solve the resulting differential equation.

7.5.8. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution that has a pdf which is a regular case of the exponential class, show that the pdf of $Y_1 = \sum_1^n K(X_i)$ is of the form $f_{Y_1}(y_1;\theta) = R(y_1)\exp[p(\theta)y_1 + nq(\theta)]$.
Hint: Let $Y_2 = X_2, \ldots, Y_n = X_n$ be $n-1$ auxiliary random variables. Find the joint pdf of $Y_1, Y_2, \ldots, Y_n$ and then the marginal pdf of Y_1.

* **7.5.9.** Let Y denote the median and let $\overline{X}$ denote the mean of a random sample of size $n = 2k+1$ from a distribution that is $N(\mu, \sigma^2)$. Compute $E(Y|\overline{X} = \overline{x})$.
Hint: See Exercise 7.5.4.

* **7.5.10.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta^2 x e^{-\theta x}$, $0 < x < \infty$, where $\theta > 0$.

(a) Argue that $Y = \sum_1^n X_i$ is a complete sufficient statistic for θ.

(b) Compute $E(1/Y)$ and find the function of Y which is the unique MVUE of θ.

* **7.5.11.** Let $X_1, X_2, \ldots, X_n$, $n > 2$, be a random sample from the binomial distribution $b(1, \theta)$.

(a) Show that $Y_1 = X_1 + X_2 + \cdots + X_n$ is a complete sufficient statistic for θ.

(b) Find the function $\varphi(Y_1)$ which is the MVUE of θ.

(c) Let $Y_2 = (X_1 + X_2)/2$ and compute $E(Y_2)$.

(d) Determine $E(Y_2|Y_1 = y_1)$.

* **7.5.12.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta e^{-\theta x}$, $0 < x < \infty$, zero elsewhere where $0 < \theta$.

(a) What is the complete sufficient statistic, for example Y, for θ?

(b) What function of Y is an unbiased estimator of θ?

★ **7.5.13.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta^x(1-\theta)$, $x = 0, 1, 2, \ldots$, zero elsewhere, where $0 \le \theta \le 1$.

(a) Find the mle, $\hat{\theta}$, of θ.

(b) Show that $\sum_1^n X_i$ is a complete sufficient statistic for θ.

(c) Determine the MVUE of θ.

7.6 Functions of a Parameter

Up to this point we have sought an MVUE of a parameter θ. Not always, however, are we interested in θ but rather in a function of θ. There are several techniques we can use to the find the MVUE. One is by inspection of the expected value of a sufficient statistic. This is how we found the MVUEs in Examples 7.5.2 and 7.5.3 of the last section. In this section and its exercises we offer more examples of the inspection technique. The second technique is based on the conditional expectation of an unbiased estimate given a sufficient statistic. The second example illustrates this technique.

Recall in Chapter 5, under regularity conditions we obtained the asymptotic distribution theory for maximum likelihood estimators (mles). This allows certain asymptotic inferences (confidence intervals and tests) for these estimators. Such a simple theory is not available for MVUEs. As Theorem 7.3.2 shows, though, sometimes we can determine the relationship between the mle and the MVUE. In these situations, we can often obtain the asymptotic distribution for the MVUE based on the asymptotic distribution of the mle. We illustrate this for some of the following examples.

Example 7.6.1. Let $X_1, X_2, \ldots, X_n$ denote the observations of a random sample of size $n > 1$ from a distribution that is $b(1,\theta)$, $0 < \theta < 1$. We know that if $Y = \sum_1^n X_i$, then Y/n is the unique minimum variance unbiased estimator of θ. Now suppose we want to estimate the variance of Y/n which is $\theta(1-\theta)/n$. Let $\delta = \theta(1-\theta)$. Because Y is a sufficient statistic for θ, it is known that we can restrict our search to functions of Y. The maximum likelihood estimate of δ which is given by $\tilde{\delta} = (Y/n)(1 - Y/n)$ is a function of the sufficient statistic and seems to be a reasonable starting point. The expectation of this statistic is given by

$$E[\tilde{\delta}] = E\left[\frac{Y}{n}\left(1 - \frac{Y}{n}\right)\right] = \frac{1}{n}E(Y) - \frac{1}{n^2}E(Y^2).$$

Now $E(Y) = n\theta$ and $E(Y^2) = n\theta(1-\theta) + n^2\theta^2$. Hence

$$E\left[\frac{Y}{n}\left(1 - \frac{Y}{n}\right)\right] = (n-1)\frac{\theta(1-\theta)}{n}.$$

If we multiply both members of this equation by $n/(n-1)$, we find that the statistic $\hat{\delta} = (n/(n-1))(Y/n)(1-Y/n) = (n/(n-1))\tilde{\delta}$ is the unique MVUE of δ. Hence, the MVUE of δ/n, the variance of Y/n, is $\hat{\delta}/n$.

It is interesting to compare the mle $\tilde{\delta}$ with $\hat{\delta}$. Recall from Chapter 5 that the mle $\tilde{\delta}$ is a consistent estimate of δ and and that $\sqrt{n}(\tilde{\delta}-\delta)$ is asymptotically normal. Because,

$$\hat{\delta} - \tilde{\delta} = \tilde{\delta}\frac{1}{n-1} \xrightarrow{P} \delta \cdot 0 = 0,$$

it follows that $\hat{\delta}$ is also a consistent estimator of δ. Further,

$$\sqrt{n}(\hat{\delta} - \delta) - \sqrt{n}(\tilde{\delta} - \delta) = \frac{\sqrt{n}}{n-1}\tilde{\delta} \xrightarrow{P} 0. \qquad (7.6.1)$$

Hence, $\sqrt{n}(\hat{\delta}-\delta)$ has the same asymptotic distribution as $\sqrt{n}(\tilde{\delta}-\delta)$. Using the Δ-method, Theorem 4.3.9, we can obtain the asymptotic distribution of $\sqrt{n}(\tilde{\delta}-\delta)$. Let $g(\theta) = \theta(1-\theta)$. Then $g'(\theta) = 1 - 2\theta$. Hence, by Theorem 4.3.9, the asymptotic distribution of $\sqrt{n}(\tilde{\delta}-\delta)$, and (7.6.1), we have the asymptotic distribution

$$\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{D} N(0, \theta(1-\theta)(1-2\theta)^2),$$

provided $\theta \ne 1/2$; see Exercise 7.6.10 for the case $\theta = 1/2$. ■

A somewhat different, but also very important problem in point estimation is considered in the next example. In the example the distribution of a random variable X is described by a pdf $f(x;\theta)$ that depends upon $\theta \in \Omega$. The problem is to estimate the fractional part of the probability for this distribution which is at, or to the left of, a fixed point c. Thus we seek an MVUE of $F(c;\theta)$, where $F(x;\theta)$ is the cdf of X.

Example 7.6.2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n > 1$ from a distribution that is $N(\theta, 1)$. Suppose that we wish to find an MVUE of the function of θ defined by

$$P(X \le c) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}\, dx = \Phi(c - \theta),$$

where c is a fixed constant. There are many unbiased estimators of $\Phi(c-\theta)$. We first exhibit one of these, say $u(X_1)$, a function of X_1 alone. We shall then compute the conditional expectation, $E[u(X_1)|\overline{X} = \overline{x}] = \varphi(\overline{x})$, of this unbiased statistic, given the sufficient statistic $\overline{X}$, the mean of the sample. In accordance with the theorems of Rao-Blackwell and Lehmann-Scheffé, $\varphi(\overline{X})$ is the unique MVUE of $\Phi(c-\theta)$.

Consider the function $u(x_1)$, where

$$u(x_1) = \begin{cases} 1 & x_1 \le c \\ 0 & x_1 > c. \end{cases}$$

The expected value of the random variable $u(X_1)$ is given by

$$E[u(X_1)] = 1 \cdot P[X_1 - \theta \le c - \theta] = \Phi(c - \theta).$$

That is, $u(X_1)$ is an unbiased estimator of $\Phi(c-\theta)$.

We shall next discuss the joint distribution of X_1 and $\overline{X}$ and the conditional distribution of X_1, given $\overline{X} = \overline{x}$. This conditional distribution enables us to compute the conditional expectation $E[u(X_1)|\overline{X} = \overline{x}] = \varphi(\overline{x})$. In accordance with Exercise 7.6.6 the joint distribution of X_1 and $\overline{X}$ is bivariate normal with mean vector (θ, θ), variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 1/n$, and correlation coefficient $\rho = 1/\sqrt{n}$. Thus the conditional pdf of X_1, given $\overline{X} = \overline{x}$, is normal with linear conditional mean

$$\theta + \frac{\rho\sigma_1}{\sigma_2}(\overline{x} - \theta) = \overline{x}$$

and with variance

$$\sigma_1^2(1-\rho^2) = \frac{n-1}{n}.$$

The conditional expectation of $u(X_1)$, given $\overline{X} = \overline{x}$, is then

$$\begin{aligned}\varphi(\overline{x}) &= \int_{-\infty}^{\infty} u(x_1)\sqrt{\frac{n}{n-1}}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{n(x_1-\overline{x})^2}{2(n-1)}\right]dx_1 \\ &= \int_{-\infty}^{c}\sqrt{\frac{n}{n-1}}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{n(x_1-\overline{x})^2}{2(n-1)}\right]dx_1.\end{aligned}$$

The change of variable $z = \sqrt{n}(x_1 - \overline{x})/\sqrt{n-1}$ enables us to write this conditional expectation as

$$\varphi(\overline{x}) = \int_{-\infty}^{c'} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = \Phi(c') = \Phi\left[\frac{\sqrt{n}(c-\overline{x})}{\sqrt{n-1}}\right],$$

where $c' = \sqrt{n}(c-\overline{x})/\sqrt{n-1}$. Thus the unique MVUE of $\Phi(c-\theta)$ is, for every fixed constant c, given by $\varphi(\overline{X}) = \Phi[\sqrt{n}(c-\overline{X})/\sqrt{n-1}]$.

In this example the mle of $\Phi(c-\theta)$ is $\Phi(c-\overline{X})$. These two estimators are close because $\sqrt{n/(n-1)} \to 1$, as $n \to \infty$. ■

Remark 7.6.1. We should like to draw the attention of the reader to a rather important fact. This has to do with the adoption of a *principle*, such as the principle of unbiasedness and minimum variance. A principle is not a theorem; and seldom does a principle yield satisfactory results in all cases. So far, this principle has provided quite satisfactory results. To see that this is not always the case, let X have a Poisson distribution with parameter θ, $0 < \theta < \infty$. We may look upon X as a random sample of size 1 from this distribution. Thus X is a complete sufficient statistic for θ. We seek the estimator of $e^{-2\theta}$ that is unbiased and has minimum variance. Consider $Y = (-1)^X$. We have

$$E(Y) = E[(-1)^X] = \sum_{x=0}^{\infty}\frac{(-\theta)^x e^{-\theta}}{x!} = e^{-2\theta}.$$

Accordingly, $(-1)^X$ is the MVUE of $e^{-2\theta}$. Here this estimator leaves much to be desired. We are endeavoring to elicit some information about the number $e^{-2\theta}$, where $0 < e^{-2\theta} < 1$; yet our point estimate is either -1 or $+1$, each of which is a very poor estimate of a number between zero and 1. We do not wish to leave the reader with the impression that an MVUE is *bad*. That is not the case at all. We merely wish to point out that if one tries hard enough, one can find instances where such a statistic is *not good*. Incidentally, the maximum likelihood estimator of $e^{-2\theta}$ is, in the case where the sample size equals one, e^{-2X}, which is probably a much better estimator in practice than is the unbiased estimator $(-1)^X$. ■

EXERCISES

⁎ **7.6.1.** Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\theta, 1)$, $-\infty < \theta < \infty$. Find the MVUE of θ^2.
Hint: First determine $E(\overline{X}^2)$.

⁎ **7.6.2.** Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(0, \theta)$. Then $Y = \sum X_i^2$ is a complete sufficient statistic for θ. Find the MVUE of θ^2.

7.6.3. In the notation of Example 7.6.2 of this section, does $P(-c \le X \le c)$ have an MVUE ? Here $c > 0$.

⁎ **7.6.4.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter $\theta > 0$.

(a) Find the MVUE of $P(X \le 1) = (1+\theta)e^{-\theta}$. *Hint:* Let $u(x_1) = 1$, $x_1 \le 1$, zero elsewhere, and find $E[u(X_1)|Y = y]$, where $Y = \sum_1^n X_i$.

(b) Express the MVUE as a function of the mle.

(c) Determine the asymptotic distribution of the mle.

⁎ **7.6.5.** Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution with parameter $\theta > 0$. From the Remark of this section, we know that $E[(-1)^{X_1}] = e^{-2\theta}$.

(a) Show that $E[(-1)^{X_1}|Y_1 = y_1] = (1 - 2/n)^{y_1}$, where $Y_1 = X_1 + X_2 + \cdots + X_n$.
Hint: First show that the conditional pdf of $X_1, X_2, \ldots, X_{n-1}$, given $Y_1 = y_1$, is multinomial, and hence that of X_1 given $Y_1 = y_1$ is $b(y_1, 1/n)$.

(b) Show that the mle of $e^{-2\theta}$ is $e^{-2\overline{X}}$.

(c) Since $y_1 = n\overline{x}$, show that $(1 - 2/n)^{y_1}$ is approximately equal to $e^{-2\overline{x}}$ when n is large.

⁎ **7.6.6.** As in Example 7.6.2, let $X_1, X_2, \ldots, X_n$ be a random sample of size $n > 1$ from a distribution that is $N(\theta, 1)$. Show that the joint distribution of X_1 and $\overline{X}$ is bivariate normal with mean vector (θ, θ), variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 1/n$, and correlation coefficient $\rho = 1/\sqrt{n}$.

* **7.6.7.** Let a random sample of size n be taken from a distribution that has the pdf $f(x;\theta) = (1/\theta)\exp(-x/\theta)I_{(0,\infty)}(x)$. Find the mle and the MVUE of $P(X \le 2)$.

* **7.6.8.** Let $X_1, X_2, \ldots, X_n$ be a random sample with the common pdf $f(x) = \theta^{-1}e^{-x/\theta}$, for $x > 0$, zero elsewhere; that is, $f(x)$ is a $\Gamma(1,\theta)$ pdf.

  (a) Show that the statistic $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ is a complete and sufficient statistic for θ.

  (b) Determine the MVUE of θ.

  (c) Determine the mle of θ.

  (d) Often, though, this pdf is written as $f(x) = \tau e^{-\tau x}$, for $x > 0$, zero elsewhere. Thus $\tau = 1/\theta$. Use Theorem 6.1.2 to determine the mle of τ.

  (e) Show that the statistic $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ is a complete and sufficient statistic for τ. Show that $(n-1)/(n\overline{X})$ is the MVUE of $\tau = 1/\theta$. Hence, as usual the reciprocal of the mle of θ is the mle of $1/\theta$, but, in this situation, the reciprocal of the MVUE of θ is not the MVUE of $1/\theta$.

  (f) Compute the variances of each of the unbiased estimators in Parts (b) and (e).

* **7.6.9.** Consider the situation of the last exercise, but suppose we have the following two independent random samples: (1). $X_1, X_2, \ldots, X_n$ is a random sample with the common pdf $f_X(x) = \theta^{-1}e^{-x/\theta}$, for $x > 0$, zero elsewhere, and (2). $Y_1, Y_2, \ldots, Y_n$ is a random sample with common pdf $f_Y(y) = \tau e^{-\tau y}$, for $y > 0$, zero elsewhere. Assume that $\tau = 1/\theta$.
The last exercise suggests that, for some constant c, $Z = c\overline{X}/\overline{Y}$ might be an unbiased estimator of θ^2. Find this constant c and the variance of Z. *Hint:* Show that $\overline{X}/(\theta^2\overline{Y})$ has an F-distribution.

7.6.10. Obtain the asymptotic distribution of the MVUE in Example 7.6.1 for the case $\theta = 1/2$.

7.7 The Case of Several Parameters

In many of the interesting problems we encounter, the pdf or pmf may not depend upon a single parameter θ, but perhaps upon two (or more) parameters. In general, our parameter space Ω will be a subset of R^p, but in many of our examples p will be two.

Definition 7.7.1. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has pdf or pmf $f(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Omega \subset R^p$. Let $\mathcal{S}$ denote the support of X. Let $\mathbf{Y}$ be an m-dimensional random vector of statistics $\mathbf{Y} = (Y_1, \ldots, Y_m)'$, where $Y_i = u_i(X_1, X_2, \ldots, X_n)$, for $i = 1, \ldots, m$. Denote the pdf or pmf of $\mathbf{Y}$ by $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$ for $\mathbf{y} \in R^m$. The random vector of statistics $\mathbf{Y}$ is* **jointly sufficient** *for $\boldsymbol{\theta}$ if and only if*

$$\frac{\prod_{i=1}^n f(x_i;\boldsymbol{\theta})}{f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})} = H(x_1, x_2, \ldots, x_n), \quad \text{for all } x_i \in \mathcal{S},$$

where $H(x_1, x_2, \ldots, x_n)$ does not depend upon $\boldsymbol{\theta}$.

In general $m \neq p$, i.e., the number of sufficient statistics does not have to be the same as the number of parameters, but in most of our examples this will be the case.

As may be anticipated, the factorization theorem can be extended. In our notation it can be stated in the following manner. The vector of statistics $\mathbf{Y}$ is jointly sufficient for the parameter $\boldsymbol{\theta} \in \Omega$ if and only if we can find two nonnegative functions k_1 and k_2 such that

$$\prod_{i=1}^n f(x_i;\boldsymbol{\theta}) = k_1(\mathbf{y};\boldsymbol{\theta})k_2(x_1, \ldots, x_n), \quad \text{for all } x_i \in \mathcal{S}, \tag{7.7.1}$$

where the function $k_2(x_1, x_2, \ldots, x_n)$ does not depend upon $\boldsymbol{\theta}$.

Example 7.7.1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having pdf

$$f(x;\theta_1,\theta_2) = \begin{cases} \frac{1}{2\theta_2} & \theta_1 - \theta_2 < x < \theta_1 + \theta_2 \\ 0 & \text{elsewhere,} \end{cases}$$

where $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics. The joint pdf of Y_1 and Y_n is given by

$$f_{Y_1,Y_2}(y_1,y_n;\theta_1,\theta_2) = \frac{n(n-1)}{(2\theta_2)^n}(y_n - y_1)^{n-2}, \quad \theta_1 - \theta_2 < y_1 < y_n < \theta_1 + \theta_2,$$

and equals zero elsewhere. Accordingly, the joint pdf of $X_1, X_2, \ldots, X_n$ can be written, for all points in its support (all x_i such that $\theta_1 - \theta_2 < x_i < \theta_1 + \theta_2$),

$$\left(\frac{1}{2\theta_2}\right)^n = \frac{n(n-1)[\max(x_i) - \min(x_i)]^{n-2}}{(2\theta_2)^n}\left(\frac{1}{n(n-1)[\max(x_i) - \min(x_i)]^{n-2}}\right).$$

Since $\min(x_i) \le x_j \le \max(x_i)$, $j = 1, 2, \ldots, n$, the last factor does not depend upon the parameters. Either the definition or the factorization theorem assures us that Y_1 and Y_n are joint sufficient statistics for θ_1 and θ_2. ■

The concept of a complete family of probability density functions is generalized as follows: Let

$$\{f(v_1, v_2, \ldots, v_k;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Omega\}$$

denote a family of pdfs of k random variables $V_1, V_2, \ldots, V_k$ that depends upon the p-dimensional vector of parameters $\boldsymbol{\theta} \in \Omega$. Let $u(v_1, v_2, \ldots, v_k)$ be a function of $v_1, v_2, \ldots, v_k$ (but not a function of any or all of the parameters). If

$$E[u(V_1, V_2, \ldots, V_k)] = 0$$

For our last example, we consider a case where the set of parameters is the cdf.

Example 7.7.5. Let $X_1, X_2, \ldots, X_n$ be a random sample having the common continuous cdf $F(x)$. Let $Y_1 < Y_2 < \cdots < Y_n$ denote the corresponding order statistics. Note that given $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$, the conditional distribution of $X_1, X_2, \ldots, X_n$ is discrete with probability $\frac{1}{n!}$ on each of the $n!$ permutations of the vector $(y_1, y_2, \ldots, y_n)$, (because $F(x)$ is continuous we can assume that each of the values $y_1, y_2, \ldots, y_n$ are distinct). That is, the conditional distribution does not depend on $F(x)$. Hence, by the definition of sufficiency the order statistics are sufficient for $F(x)$. Furthermore, while the proof is beyond the scope of this book, it can be shown that the order statistics are also complete; see page 72 of Lehmann and Casella (1998).

Let $T = T(x_1, x_2, \ldots, x_n)$ be any statistic which is *symmetric in its arguments*; i.e., $T(x_1, x_2, \ldots, x_n) = T(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ for any permutation $(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ of $(x_1, x_2, \ldots, x_n)$. Then T is a function of the order statistics. This is useful in determining MVUEs for this situation; see Exercises 7.7.12 and 7.7.13. ■

EXERCISES

* **7.7.1.** Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size 3 from the distribution with pdf.

$$f(x;\theta_1,\theta_2) = \begin{cases} \frac{1}{\theta_2}\exp\left(-\frac{x-\theta_1}{\theta_2}\right) & \theta_1 < x < \infty, \ -\infty < \theta_1 < \infty, \ 0 < \theta_2 < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Find the joint pdf of $Z_1 = Y_1, Z_2 = Y_2$, and $Z_3 = Y_1 + Y_2 + Y_3$. The corresponding transformation maps the space $\{(y_1, y_2, y_3) : \theta_1 < y_1 < y_2 < y_3 < \infty\}$ onto the space

$$\{(z_1, z_2, z_3) : \theta_1 < z_1 < z_2 < (z_3 - z_1)/2 < \infty\}.$$

Show that Z_1 and Z_3 are joint sufficient statistics for θ_1 and θ_2.

7.7.2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution that has a pdf of form (7.7.2) of this section. Show that $Y_1 = \sum_{i=1}^{n} K_1(X_i), \ldots, Y_m = \sum_{i=1}^{m} K_m(X_i)$ have a joint pdf of form (7.7.4) of this section.

* **7.7.3.** Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ denote a random sample of size n from a bivariate normal distribution with means μ_1 and μ_2, positive variances σ_1^2 and σ_2^2, and correlation coefficient ρ. Show that $\sum_1^n X_i$, $\sum_1^n Y_i$, $\sum_1^n X_i^2$, $\sum_1^n Y_i^2$, and $\sum_1^n X_iY_i$ are joint complete sufficient statistics for the five parameters. Are $\overline{X} = \sum_1^n X_i/n$, $\overline{Y} = \sum_1^n Y_i/n$, $S_1^2 = \sum_1^n (X_i - \overline{X})^2/(n-1)$, $S_2^2 = \sum_1^n (Y_i - \overline{Y})^2/(n-1)$, and $\sum_1^n (X_i - \overline{X})(Y_i - \overline{Y})/(n-1)S_1S_2$ also joint complete sufficient statistics for these parameters?

7.7.4. Let the pdf $f(x;\theta_1,\theta_2)$ be of the form

$$\exp[p_1(\theta_1,\theta_2)K_1(x) + p_2(\theta_1,\theta_2)K_2(x) + S(x) + q_1(\theta_1,\theta_2)], \quad a < x < b,$$

zero elsewhere. Let $K_1'(x) = cK_2'(x)$. Show that $f(x;\theta_1,\theta_2)$ can be written in the form

$$\exp[p_1(\theta_1,\theta_2)K(x) + S(x) + q_1(\theta_1,\theta_2)], \quad a < x < b,$$

zero elsewhere. This is the reason why it is required that no one $K_j'(x)$ be a linear homogeneous function of the others, that is, so that the number of sufficient statistics equals the number of parameters.

7.7.5. In Example 7.7.2, find the MVUE of the standard deviation $\sqrt{\theta_2}$

* **7.7.6.** Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution with pdf $f(x;\theta_1,\theta_2) = 1/(2\theta_2)$, $\theta_1 - \theta_2 < x < \theta_1 + \theta_2$, where $-\infty < \theta_1 < \infty$ and $\theta_2 > 0$, and the pdf is equal to zero elsewhere.

(a) Show that $Y_1 = \min(X_i)$ and $Y_n = \max(X_i)$, the joint sufficient statistics for θ_1 and θ_2, are complete.

(b) Find the MVUEs of θ_1 and θ_2.

* **7.7.7.** Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta_1, \theta_2)$.

(a) If the constant b is defined by the equation $P(X \le b) = 0.90$, find the mle and the MVUE of b.

(b) If c is a given constant, find the mle and the MVUE of $P(X \le c)$.

7.7.8. In the notation of Example 7.7.3, show that the mle of p_jp_l is $n^{-2}Y_jY_l$.

7.7.9. Refer to Example 7.7.4 on sufficiency for the multivariate normal model.

(a) Determine the MVUE of the covariance parameters σ_{ij}.

(b) Let $h = \sum_{i=1}^{k} a_i\mu_i$, where $a_1, \ldots, a_k$ are specified constants. Find the MVUE for h.

* **7.7.10.** In a personal communication, LeRoy Folks noted that the inverse Gaussian pdf

$$f(x;\theta_1,\theta_2) = \left(\frac{\theta_2}{2\pi x^3}\right)^{1/2} \exp\left[\frac{-\theta_2(x-\theta_1)^2}{2\theta_1^2 x}\right], \quad 0 < x < \infty, \qquad (7.7.9)$$

where $\theta_1 > 0$ and $\theta_2 > 0$ is often used to model lifetimes. Find the complete sufficient statistics for (θ_1, θ_2), if $X_1, X_2, \ldots, X_n$ is a random sample from the distribution having this pdf.

that nowhere was this needed or assumed. The pdf or pmf may depend upon any finite number of parameters. What is essential is that the hypothesis H_0 and the alternative hypothesis H_1 be simple, namely that they completely specify the distributions. With this in mind, we see that the simple hypotheses H_0 and H_1 do not need to be hypotheses about the parameters of a distribution, nor, as a matter of fact, do the random variables $X_1, X_2, \ldots, X_n$ need to be independent. That is, if H_0 is the simple hypothesis that the joint pdf or pmf is $g(x_1, x_2, \ldots, x_n)$, and if H_1 is the alternative simple hypothesis that the joint pdf or pmf is $h(x_1, x_2, \ldots, x_n)$, then C is a best critical region of size α for testing H_0 against H_1 if, for $k > 0$:

1. $\dfrac{g(x_1, x_2, \ldots, x_n)}{h(x_1, x_2, \ldots, x_n)} \le k$ for $(x_1, x_2, \ldots, x_n) \in C$.

2. $\dfrac{g(x_1, x_2, \ldots, x_n)}{h(x_1, x_2, \ldots, x_n)} \ge k$ for $(x_1, x_2, \ldots, x_n) \in C^c$.

3. $\alpha = P_{H_0}[(X_1, X_2, \ldots, X_n) \in C]$.

An illustrative example follows.

Example 8.1.3. Let $X_1, \ldots, X_n$ denote a random sample from a distribution which has a pmf $f(x)$ that is positive on and only on the nonnegative integers. It is desired to test the simple hypothesis

$$H_0 : f(x) = \begin{cases} \frac{e^{-1}}{x!} & x = 0, 1, 2, \ldots \\ 0 & \text{elsewhere,} \end{cases}$$

against the alternative simple hypothesis

$$H_1 : f(x) = \begin{cases} (\frac{1}{2})^{x+1} & x = 0, 1, 2, \ldots \\ 0 & \text{elsewhere.} \end{cases}$$

Here

$$\frac{g(x_1, \ldots, x_n)}{h(x_1, \ldots, x_n)} = \frac{e^{-n}/(x_1!x_2!\cdots x_n!)}{(\frac{1}{2})^n(\frac{1}{2})^{x_1+x_2+\cdots+x_n}} = \frac{(2e^{-1})^n 2^{\sum x_i}}{\prod_1^n (x_i!)}.$$

If $k > 0$, the set of points $(x_1, x_2, \ldots, x_n)$ such that

$$\left(\sum_1^n x_i\right)\log 2 - \log\left[\prod_1^n (x_i!)\right] \le \log k - n\log(2e^{-1}) = c$$

is a best critical region C. Consider the case of $k = 1$ and $n = 1$. The preceding inequality may be written $2^{x_1}/x_1! \le e/2$. This inequality is satisfied by all points

in the set $C = \{x_1 : x_1 = 0, 3, 4, 5, \ldots\}$. Thus the power of the test when H_0 is true is

$$P_{H_0}(X_1 \in C) = 1 - P_{H_0}(X_1 = 1, 2) = 0.448,$$

approximately, in accordance with Table I of Appendix B; i.e., the level of this test is 0.448. The power of the test when H_1 is true is given by

$$P_{H_1}(X_1 \in C) = 1 - P_{H_1}(X_1 = 1, 2) = 1 - (\tfrac{1}{4} + \tfrac{1}{8}) = 0.625. \quad ■$$

Note that these results are consistent with Corollary 8.1.1.

Remark 8.1.2. In the notation of this section, say C is a critical region such that

$$\alpha = \int_C L(\theta') \quad \text{and} \quad \beta = \int_{C^c} L(\theta''),$$

where α and β equal the respective probabilities of the Type I and Type II errors associated with C. Let d_1 and d_2 be two given positive constants. Consider a certain linear function of α and β, namely

$$d_1\int_C L(\theta') + d_2\int_{C^c} L(\theta'') = d_1\int_C L(\theta') + d_2\left[1 - \int_C L(\theta'')\right] = d_2 + \int_C [d_1L(\theta') - d_2L(\theta'')].$$

If we wished to minimize this expression, we would select C to be the set of all $(x_1, x_2, \ldots, x_n)$ such that

$$d_1L(\theta') - d_2L(\theta'') < 0$$

or, equivalently,

$$\frac{L(\theta')}{L(\theta'')} < \frac{d_2}{d_1}, \quad \text{for all } (x_1, x_2, \ldots, x_n) \in C,$$

which according to the Neyman-Pearson theorem provides a best critical region with $k = d_2/d_1$. That is, this critical region C is one that minimizes $d_1\alpha + d_2\beta$. There could be others, including points on which $L(\theta')/L(\theta'') = d_2/d_1$, but these would still be best critical regions according to the Neyman-Pearson theorem. ■

EXERCISES

* **8.1.1.** In Example 8.1.2 of this section, let the simple hypotheses read $H_0 : \theta = \theta' = 0$ and $H_1 : \theta = \theta'' = -1$. Show that the best test of H_0 against H_1 may be carried out by use of the statistic $\overline{X}$, and that if $n = 25$ and $\alpha = 0.05$, the power of the test is 0.999+ when H_1 is true.

* **8.1.2.** Let the random variable X have the pdf $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, zero elsewhere. Consider the simple hypothesis $H_0 : \theta = \theta' = 2$ and the alternative hypothesis $H_1 : \theta = \theta'' = 4$. Let X_1, X_2 denote a random sample of size 2 from this distribution. Show that the best test of H_0 against H_1 may be carried out by use of the statistic $X_1 + X_2$.

8.1.3. Repeat Exercise 8.1.2 when $H_1 : \theta = \theta'' = 6$. Generalize this for every $\theta'' > 2$.

8.1.4. Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size 10 from a normal distribution $N(0, \sigma^2)$. Find a best critical region of size $\alpha = 0.05$ for testing $H_0 : \sigma^2 = 1$ against $H_1 : \sigma^2 = 2$. Is this a best critical region of size 0.05 for testing $H_0 : \sigma^2 = 1$ against $H_1 : \sigma^2 = 4$? Against $H_1 : \sigma^2 = \sigma_1^2 > 1$?

8.1.5. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution having pdf of the form $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, show that a best critical region for testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$ is $C = \left\{(x_1, x_2, \ldots, x_n) : c \le \prod_{i=1}^{n} x_i\right\}$.

8.1.6. Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a distribution that is $N(\theta_1, \theta_2)$. Find a best test of the simple hypothesis $H_0 : \theta_1 = \theta_1' = 0,\ \theta_2 = \theta_2' = 1$ against the alternative simple hypothesis $H_1 : \theta_1 = \theta_1'' = 1,\ \theta_2 = \theta_2'' = 4$.

8.1.7. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution $N(\theta, 100)$. Show that $C = \left\{(x_1, x_2, \ldots, x_n) : c \le \overline{x} = \sum_1^n x_i/n\right\}$ is a best critical region for testing $H_0 : \theta = 75$ against $H_1 : \theta = 78$. Find n and c so that

$$P_{H_0}[(X_1, X_2, \ldots, X_n) \in C] = P_{H_0}(\overline{X} \ge c) = 0.05$$

and

$$P_{H_1}[(X_1, X_2, \ldots, X_n) \in C] = P_{H_1}(\overline{X} \ge c) = 0.90,$$

approximately.

8.1.8. If $X_1, X_2, \ldots, X_n$ is a random sample from a beta distribution with parameters $\alpha = \beta = \theta > 0$, find a best critical region for testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$.

8.1.9. Let $X_1, X_2, \ldots, X_n$ be iid with pmf $f(x;p) = p^x(1-p)^{1-x}$, $x = 0, 1$, zero elsewhere. Show that $C = \left\{(x_1, \ldots, x_n) : \sum_1^n x_i \le c\right\}$ is a best critical region for testing $H_0 : p = \frac{1}{2}$ against $H_1 : p = \frac{1}{3}$. Use the Central Limit Theorem to find n and c so that approximately $P_{H_0}\left(\sum_1^n X_i \le c\right) = 0.10$ and $P_{H_1}\left(\sum_1^n X_i \le c\right) = 0.80$.

8.1.10. Let $X_1, X_2, \ldots, X_{10}$ denote a random sample of size 10 from a Poisson distribution with mean θ. Show that the critical region C defined by $\sum_1^{10} x_i \ge 3$ is a best critical region for testing $H_0 : \theta = 0.1$ against $H_1 : \theta = 0.5$. Determine, for this test, the significance level α and the power at $\theta = 0.5$.

8.2 Uniformly Most Powerful Tests

This section will take up the problem of a test of a simple hypothesis H_0 against an alternative composite hypothesis H_1. We begin with an example.

Example 8.2.1. Consider the pdf

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}e^{-x/\theta} & 0 < x < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

of Exercises 8.1.2 and 8.1.3. It is desired to test the simple hypothesis $H_0 : \theta = 2$ against the alternative composite hypothesis $H_1 : \theta > 2$. Thus $\Omega = \{\theta : \theta \ge 2\}$. A random sample, X_1, X_2, of size $n = 2$ will be used, and the critical region is $C = \{(x_1, x_2) : 9.5 \le x_1 + x_2 < \infty\}$. It was shown in the example cited that the significance level of the test is approximately 0.05 and the power of the test when $\theta = 4$ is approximately 0.31. The power function $\gamma(\theta)$ of the test for all $\theta \ge 2$ will now be obtained. We have

$$\begin{aligned} \gamma(\theta) &= 1 - \int_0^{9.5}\int_0^{9.5-x_2} \frac{1}{\theta^2}\exp\left(-\frac{x_1+x_2}{\theta}\right) dx_1 dx_2 \\ &= \left(\frac{\theta + 9.5}{\theta}\right) e^{-9.5/\theta}, \quad 2 \le \theta. \end{aligned}$$

For example, $\gamma(2) = 0.05$, $\gamma(4) = 0.31$, and $\gamma(9.5) = 2/e \doteq 0.74$. It is shown (Exercise 8.1.3) that the set $C = \{(x_1, x_2) : 9.5 \le x_1 + x_2 < \infty\}$ is a best critical region of size 0.05 for testing the simple hypothesis $H_0 : \theta = 2$ against each simple hypothesis in the composite hypothesis $H_1 : \theta > 2$. ∎

The preceding example affords an illustration of a test of a simple hypothesis H_0 that is a best test of H_0 against every simple hypothesis in the alternative composite hypothesis H_1. We now define a critical region when it exists, which is a best critical region for testing a simple hypothesis H_0 against an alternative composite hypothesis H_1. It seems desirable that this critical region should be a best critical region for testing H_0 against each simple hypothesis in H_1. That is, the power function of the test that corresponds to this critical region should be at least as great as the power function of any other test with the same significance level for every simple hypothesis in H_1.

Definition 8.2.1. *The critical region C is a* **uniformly most powerful** (UMP) **critical region** *of size α for testing the simple hypothesis H_0 against an alternative composite hypothesis H_1 if the set C is a best critical region of size α for testing H_0 against each simple hypothesis in H_1. A test defined by this critical region C is called a* **uniformly most powerful** (UMP) **test**, *with significance level α, for testing the simple hypothesis H_0 against the alternative composite hypothesis H_1.*

As will be seen presently, uniformly most powerful tests do not always exist. However, when they do exist, the Neyman-Pearson theorem provides a technique for finding them. Some illustrative examples are given here.

EXERCISES

*8.2.1. Let X have the pmf $f(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, zero elsewhere. We test the simple hypothesis $H_0 : \theta = \frac{1}{4}$ against the alternative composite hypothesis $H_1 : \theta < \frac{1}{4}$ by taking a random sample of size 10 and rejecting $H_0 : \theta = \frac{1}{4}$ if and only if the observed values $x_1, x_2, \ldots, x_{10}$ of the sample observations are such that $\sum_1^{10} x_i \le 1$. Find the power function $\gamma(\theta)$, $0 < \theta \le \frac{1}{4}$, of this test.

*8.2.2. Let X have a pdf of the form $f(x;\theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere. Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size 4 from this distribution. Let the observed value of Y_4 be y_4. We reject $H_0 : \theta = 1$ and accept $H_1 : \theta \ne 1$ if either $y_4 \le \frac{1}{2}$ or $y_4 > 1$. Find the power function $\gamma(\theta)$, $0 < \theta$, of the test.

*8.2.3. Consider a normal distribution of the form $N(\theta, 4)$. The simple hypothesis $H_0 : \theta = 0$ is rejected, and the alternative composite hypothesis $H_1 : \theta > 0$ is accepted if and only if the observed mean $\overline{x}$ of a random sample of size 25 is greater than or equal to $\frac{3}{5}$. Find the power function $\gamma(\theta)$, $0 \le \theta$, of this test.

*8.2.4. Consider the distributions $N(\mu_1, 400)$ and $N(\mu_2, 225)$. Let $\theta = \mu_1 - \mu_2$. Let $\overline{x}$ and $\overline{y}$ denote the observed means of two independent random samples, each of size n, from these two distributions. We reject $H_0 : \theta = 0$ and accept $H_1 : \theta > 0$ if and only if $\overline{x} - \overline{y} \ge c$. If $\gamma(\theta)$ is the power function of this test, find n and c so that $\gamma(0) = 0.05$ and $\gamma(10) = 0.90$, approximately.

8.2.5. If in Example 8.2.2 of this section $H_0 : \theta = \theta'$, where θ' is a fixed positive number, and $H_1 : \theta < \theta'$, show that the set $\left\{(x_1, x_2, \ldots, x_n) : \sum_1^n x_i^2 \le c\right\}$ is a uniformly most powerful critical region for testing H_0 against H_1.

8.2.6. If, in Example 8.2.2 of this section, $H_0 : \theta = \theta'$, where θ' is a fixed positive number, and $H_1 : \theta \ne \theta'$, show that there is no uniformly most powerful test for testing H_0 against H_1.

*8.2.7. Let $X_1, X_2, \ldots, X_{25}$ denote a random sample of size 25 from a normal distribution $N(\theta, 100)$. Find a uniformly most powerful critical region of size $\alpha = 0.10$ for testing $H_0 : \theta = 75$ against $H_1 : \theta > 75$.

*8.2.8. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution $N(\theta, 16)$. Find the sample size n and a uniformly most powerful test of $H_0 : \theta = 25$ against $H_1 : \theta < 25$ with power function $\gamma(\theta)$ so that approximately $\gamma(25) = 0.10$ and $\gamma(23) = 0.90$.

*8.2.9. Consider a distribution having a pmf of the form $f(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, zero elsewhere. Let $H_0 : \theta = \frac{1}{20}$ and $H_1 : \theta > \frac{1}{20}$. Use the central limit theorem to determine the sample size n of a random sample so that a uniformly most powerful test of H_0 against H_1 has a power function $\gamma(\theta)$, with approximately $\gamma(\frac{1}{20}) = 0.05$ and $\gamma(\frac{1}{10}) = 0.90$.

8.2.10. Illustrative Example 8.2.1 of this section dealt with a random sample of size $n = 2$ from a gamma distribution with $\alpha = 1$, $\beta = \theta$. Thus the mgf of the distribution is $(1 - \theta t)^{-1}$, $t < 1/\theta$, $\theta \ge 2$. Let $Z = X_1 + X_2$. Show that Z has a gamma distribution with $\alpha = 2$, $\beta = \theta$. Express the power function $\gamma(\theta)$ of Example 8.2.1 in terms of a single integral. Generalize this for a random sample of size n.

*8.2.11. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, where $\theta > 0$. Find a sufficient statistic for θ and show that a uniformly most powerful test of $H_0 : \theta = 6$ against $H_1 : \theta < 6$ is based on this statistic.

*8.2.12. Let X have the pdf $f(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, zero elsewhere. We test $H_0 : \theta = \frac{1}{2}$ against $H_1 : \theta < \frac{1}{2}$ by taking a random sample $X_1, X_2, \ldots, X_5$ of size $n = 5$ and rejecting H_0 if $Y = \sum_1^n X_i$ is observed to be less than or equal to a constant c.

(a) Show that this is a uniformly most powerful test.

(b) Find the significance level when $c = 1$.

(c) Find the significance level when $c = 0$.

(d) By using a *randomized test*, as discussed in Example 5.6.4, modify the tests given in Parts (b) and (c) to find a test with significance level $\alpha = \frac{2}{32}$.

*8.2.13. Let $X_1, \ldots, X_n$ denote a random sample from a gamma-type distribution with $\alpha = 2$ and $\beta = \theta$. Let $H_0 : \theta = 1$ and $H_1 : \theta > 1$.

(a) Show that there exists a uniformly most powerful test for H_0 against H_1, determine the statistic Y upon which the test may be based, and indicate the nature of the best critical region.

(b) Find the pdf of the statistic Y in Part (a). If we want a significance level of 0.05, write an equation which can be used to determine the critical region. Let $\gamma(\theta)$, $\theta \ge 1$, be the power function of the test. Express the power function as an integral.

8.3 Likelihood Ratio Tests

In the first section of this chapter we presented most powerful tests for simple versus simple hypotheses. In the second section, we extended this theory to uniformly most powerful tests for essentially one-sided alternative hypotheses and families of distributions which have monotone likelihood ratio. What about the general case? That is, suppose the random variable X has pdf or pmf $f(x;\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a vector of parameters in Ω. Let $\omega \subset \Omega$ and consider the hypotheses

$$H_0 : \boldsymbol{\theta} \in \omega \text{ versus } H_1 : \boldsymbol{\theta} \in \Omega \cap \omega^c. \tag{8.3.1}$$

8.3.3. Verify Equations (8.3.4) of Example 8.3.1 of this section.

8.3.4. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ follow the location model

$$\begin{aligned} X_i &= \theta_1 + Z_i, \quad i = 1, \ldots, n \\ Y_i &= \theta_2 + Z_{n+i}, \quad i = 1, \ldots, m, \end{aligned} \tag{8.3.9}$$

where $Z_1, \ldots, Z_{n+m}$ are iid random variables with common pdf $f(z)$. Assume that $E(Z_i) = 0$ and $\text{Var}(Z_i) = \theta_3 < \infty$.

(a) Show that $E(X_i) = \theta_1$, $E(Y_i) = \theta_2$, and $\text{Var}(X_i) = \text{Var}(Y_i) = \theta_3$.

(b) Consider the hypotheses of Example 8.3.1; i.e,

$$H_0 : \theta_1 = \theta_2 \text{ versus } H_1 : \theta_1 \neq \theta_2.$$

Show that under H_0, the test statistic T given in expression (8.3.5) has a limiting $N(0, 1)$ distribution.

(c) Using Part (b), determine the corresponding large sample test (decision rule) of H_0 versus H_1. (This shows that the test in Example 8.3.1 is asymptotically correct.)

8.3.5. Show that the likelihood ratio principle leads to the same test when testing a simple hypothesis H_0 against an alternative simple hypothesis H_1, as that given by the Neyman-Pearson theorem. Note that there are only two points in Ω.

* **8.3.6.** Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(\theta, 1)$. Show that the likelihood ratio principle for testing $H_0 : \theta = \theta'$, where θ' is specified, against $H_1 : \theta \neq \theta'$ leads to the inequality $|\overline{x} - \theta'| \geq c$.

(a) Is this a uniformly most powerful test of H_0 against H_1?

(b) Is this a uniformly most powerful unbiased test of H_0 against H_1?

* **8.3.7.** Let $X_1, X_2, \ldots, X_n$ be iid $N(\theta_1, \theta_2)$. Show that the likelihood ratio principle for testing $H_0 : \theta_2 = \theta_2'$ specified, and θ_1 unspecified, against $H_1 : \theta_2 \neq \theta_2'$, θ_1 unspecified, leads to a test that rejects when $\sum_1^n (x_i - \overline{x})^2 \leq c_1$ or $\sum_1^n (x_i - \overline{x})^2 \geq c_2$, where $c_1 < c_2$ are selected appropriately.

* **8.3.8.** Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be independent random samples from the distributions $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$, respectively.

(a) Show that the likelihood ratio for testing $H_0 : \theta_1 = \theta_2$, $\theta_3 = \theta_4$ against all alternatives is given by

$$\frac{\left[\sum_1^n (x_i - \overline{x})^2 / n\right]^{n/2} \left[\sum_1^m (y_i - \overline{y})^2 / m\right]^{m/2}}{\left\{\left[\sum_1^n (x_i - u)^2 + \sum_1^m (y_i - u)^2\right] \Big/ (m+n)\right\}^{(n+m)/2}},$$

where $u = (n\overline{x} + m\overline{y})/(n+m)$.

(b) Show that the likelihood ratio for testing $H_0 : \theta_3 = \theta_4$ with θ_1 and θ_4 unspecified can be based on the test statistic F given in expression (8.3.8).

* **8.3.9.** Let $Y_1 < Y_2 < \cdots < Y_5$ be the order statistics of a random sample of size $n = 5$ from a distribution with pdf $f(x; \theta) = \frac{1}{2} e^{-|x-\theta|}$, $-\infty < x < \infty$, for all real θ. Find the likelihood ratio test Λ for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

* **8.3.10.** A random sample $X_1, X_2, \ldots, X_n$ arises from a distribution given by

$$H_0 : f(x; \theta) = \frac{1}{\theta}, \quad 0 < x < \theta, \quad \text{zero elsewhere},$$

or

$$H_1 : f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty, \quad \text{zero elsewhere}.$$

Determine the likelihood ratio (Λ) test associated with the test of H_0 against H_1.

* **8.3.11.** Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with pdf $f(x; \theta) = \theta(1-x)^{\theta-1}$, $0 < x < 1$, zero elsewhere, where $\theta > 0$.

(a) Find the form of the uniformly most powerful test of $H_0 : \theta = 1$ against $H_1 : \theta > 1$.

(b) What is the likelihood ratio Λ for testing $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$?

* **8.3.12.** Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from two normal distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively, where σ^2 is the common but unknown variance.

(a) Find the likelihood ratio Λ for testing $H_0 : \mu_1 = \mu_2 = 0$ against all alternatives.

(b) Rewrite Λ so that it is a function of a statistic Z which has a well-known distribution.

(c) Give the distribution of Z under both null and alternative hypotheses.

* **8.3.13.** Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random sample from a bivariate normal distribution with $\mu_1, \mu_2, \sigma_1^2 = \sigma_2^2 = \sigma^2, \rho = \frac{1}{2}$, where μ_1, μ_2, and $\sigma^2 > 0$ are unknown real numbers. Find the likelihood ratio Λ for testing $H_0 : \mu_1 = \mu_2 = 0$, σ^2 unknown against all alternatives. The likelihood ratio Λ is a function of what statistic that has a well-known distribution?

8.3.14. Let X be a random variable with pdf $f_X(x) = (2b_X)^{-1} \exp\{-|x|/b_X\}$, for $-\infty < x < \infty$ and $b_X > 0$. First, show that the variance of X is $\sigma_X^2 = 2b_X^2$.

Now let Y, independent of X, have pdf $f_Y(y) = (2b_Y)^{-1} \exp\{-|y|/b_Y\}$, for $-\infty < x < \infty$ and $b_Y > 0$. Consider the hypotheses

$$H_0 : \sigma_X^2 = \sigma_Y^2 \text{ versus } H_1 : \sigma_X^2 > \sigma_Y^2.$$

* **7.7.11.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\theta_1, \theta_2)$ distribution.

(a) Show that $E[(X_1 - \theta_1)^4] = 3\theta_2^2$.

(b) Find the MVUE of $3\theta_2^2$.

7.7.12. Let $X_1, \ldots, X_n$ be a random sample from a distribution of the continuous type with cdf $F(x)$. Suppose the mean, $\mu = E(X_1)$, exists. Using Example 7.7.5, show that the sample mean, $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$ is the MVUE of μ.

7.7.13. Let $X_1, \ldots, X_n$ be a random sample from a distribution of the continuous type with cdf $F(x)$. Let $\theta = P(X_1 \le a) = F(a)$, where a is known. Show that the proportion $n^{-1}\#\{X_i \le a\}$ is the MVUE of θ.

7.8 Minimal Sufficiency and Ancillary Statistics

In the study of statistics, it is clear that we want to reduce the data contained in the entire sample as much as possible without losing relevant information about the important characteristics of the underlying distribution. That is, a large collection of numbers in the sample is not as meaningful as a few good summary statistics of those data. Sufficient statistics, if they exist, are valuable because we know that the statisticians with those summary measures have as much information as the statistician with the entire sample. Sometimes, however, there are several sets of joint sufficient statistics, and thus we would like to find the simplest one of these sets. For illustration, in a sense, the observations $X_1, X_2, \ldots, X_n$, $n > 2$, of a random sample from $N(\theta_1, \theta_2)$ could be thought of as joint sufficient statistics for θ_1 and θ_2. We know, however, that we can use $\overline{X}$ and S^2 as joint sufficient statistics for those parameters, which is a great simplification over using $X_1, X_2, \ldots, X_n$, particularly if n is large.

In most instances in this chapter, we have been able to find a single sufficient statistic for one parameter or two joint sufficient statistics for two parameters. Possibly the most complicated cases considered so far are given in Example 7.7.3, in which we find $k+k(k+1)/2$ joint sufficient statistics for $k+k(k+1)/2$ parameters; or the multivariate normal distribution given in Example 7.7.4; or the use the order statistics of a random sample for some completely unknown distribution of the continuous type as in Example 7.7.5.

What we would like to do is to change from one set of joint sufficient statistics to another, always reducing the number of statistics involved until we cannot go any further without losing the sufficiency of the resulting statistics. Those statistics that are there at the end of this reduction are called *minimal sufficient statistics*. These are sufficient for the parameters and are functions of every other set of sufficient statistics for those same parameters. Often, if there are k parameters, we can find k joint sufficient statistics that are minimal. In particular, if there is one parameter, we can often find a single sufficient statistic which is minimal. Most of the earlier examples that we have considered illustrate this point, but this is not always the case as shown by the following example.

Example 7.8.1. Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution over the interval $(\theta - 1, \theta + 1)$ having pdf

$$f(x;\theta) = (\tfrac{1}{2})I_{(\theta-1,\theta+1)}(x), \quad \text{where} -\infty < \theta < \infty.$$

The joint pdf of $X_1, X_2, \ldots, X_n$ equals the product of $(\frac{1}{2})^n$ and certain indicator functions, namely

$$(\tfrac{1}{2})^n \prod_{i=1}^{n} I_{(\theta-1,\theta+1)}(x_i) = (\tfrac{1}{2})^n \{I_{(\theta-1,\theta+1)}[\min(x_i)]\}\{I_{(\theta-1,\theta+1)}[\max(x_i)]\},$$

because $\theta - 1 < \min(x_i) \le x_j \le \max(x_i) < \theta + 1$, $j = 1, 2, \ldots, n$. Thus the order statistics $Y_1 = \min(X_i)$ and $Y_n = \max(X_i)$ are the sufficient statistics for θ. These two statistics actually are minimal for this one parameter, as we cannot reduce the number of them to less than two and still have sufficiency. ■

There is an observation that helps us observe that almost all the sufficient statistics that we have studied thus far are minimal. We have noted that the mle $\hat{\theta}$ of θ is a function of one or more sufficient statistics, when the latter exists. Suppose that this mle $\hat{\theta}$ is also sufficient. Since this sufficient statistic $\hat{\theta}$ is a function of the other sufficient statistics, Theorem 7.3.2, it must be minimal. For example, we have

1. The mle $\hat{\theta} = \overline{X}$ of θ in $N(\theta, \sigma^2)$, σ^2 known, is a minimal sufficient statistic for θ.

2. The mle $\hat{\theta} = \overline{X}$ of θ in a Poisson distribution with mean θ is a minimal sufficient statistic for θ.

3. The mle $\hat{\theta} = Y_n = \max(X_i)$ of θ in the uniform distribution over $(0, \theta)$ is a minimal sufficient statistic for θ.

4. The maximum likelihood estimators $\hat{\theta}_1 = \overline{X}$ and $\hat{\theta}_2 = S^2$ of θ_1 and θ_2 in $N(\theta_1, \theta_2)$ are joint minimal sufficient statistics for θ_1 and θ_2.

From these examples we see that the minimal sufficient statistics do not need to be unique, for any one-to-one transformation of them also provides minimal sufficient statistics. The linkage between minimal sufficient statistics and the mle, however, does not hold in many interesting instances. We illustrate this in the next two examples.

Example 7.8.2. Consider the model given in Example 7.8.1. There we noted that $Y_1 = \min(X_i)$ and $Y_n = \max(X_i)$ are joint sufficient statistics. Also, we have

$$\theta - 1 < Y_1 < Y_n < \theta + 1$$

or, equivalently,

$$Y_n - 1 < \theta < Y_1 + 1.$$

Hence, to maximize the likelihood function so that it equals $(\frac{1}{2})^n$, θ can be any value between $Y_n - 1$ and $Y_1 + 1$. For example, many statisticians take the mle to be the mean of these two end points, namely

$$\hat{\theta} = \frac{Y_n - 1 + Y_1 + 1}{2} = \frac{Y_1 + Y_n}{2},$$

for all $c > 0$. Then

$$Z = u(X_1, X_2, \dots, X_n) = u(\theta W_1, \theta W_2, \dots, \theta W_n) = u(W_1, W_2, \dots, W_n).$$

Since neither the joint pdf of $W_1, W_2, \dots, W_n$ nor Z contain θ, the distribution of Z must not depend upon θ. We say that Z is a **scale-invariant statistic**.

The following are some examples of scale-invariant statistics: $X_1/(X_1 + X_2)$, $X_1^2/\sum_1^n X_i^2$, $\min(X_i)/\max(X_i)$, and so on. ■

Example 7.8.6 (Location and Scale Invariant Statistics). Finally, consider a random sample $X_1, X_2, \dots, X_n$ which follows a location and scale model as in Example 7.7.5. That is,

$$X_i = \theta_1 + \theta_2 W_i, \quad i = 1, \dots, n, \tag{7.8.5}$$

where W_i are iid with the common pdf $f(t)$ which is free of θ_1 and θ_2. In this case the pdf of X_i is $\theta_2^{-1} f((x - \theta_1)/\theta_2)$. Consider the statistic $Z = u(X_1, X_2, \dots, X_n)$ where

$$u(cx_1 + d, \dots, cx_n + d) = u(x_1, \dots, x_n).$$

Then

$$Z = u(X_1, \dots, X_n) = u(\theta_1 + \theta_2 W_1, \dots, \theta_1 + \theta_2 W_n) = u(W_1, \dots, W_n).$$

Since neither the joint pdf of $W_1, \dots, W_n$ nor Z contains θ_1 and θ_2, the distribution of Z must not depend upon θ_1 nor θ_2. Statistics such as $Z = u(X_1, X_2, \dots, X_n)$ are called **location and scale invariant statistics**. The following are four examples of such statistics:

(a) $T_1 = [\max(X_i) - \min(X_i)]/S$,

(b) $T_2 = \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2/S^2$,

(c) $T_3 = (X_i - \overline{X})/S$,

(d) $T_4 = |X_i - X_j|/S,\ ; i \neq j$.

Thus these location invariant, scale invariant, and location and scale invariant statistics provide good illustrations, with the appropriate model for the pdf, of ancillary statistics. Since an ancillary statistic and a complete (minimal) sufficient statistic are such opposites, we might believe that there is, in some sense, no relationship between the two. This is true and in the next section we show that they are independent statistics.

EXERCISES

7.8.1. Let $X_1, X_2, \dots, X_n$ be a random sample from each of the following distributions involving the parameter θ. In each case find the mle of θ and show that it is a sufficient statistic for θ and hence a minimal sufficient statistic.

(a) $b(1, \theta)$, where $0 \le \theta \le 1$.

(b) Poisson with mean $\theta > 0$.

(c) Gamma with $\alpha = 3$ and $\beta = \theta > 0$.

(d) $N(\theta, 1)$, where $-\infty < \theta < \infty$.

(e) $N(0, \theta)$, where $0 < \theta < \infty$.

7.8.2. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from the uniform distribution over the closed interval $[-\theta, \theta]$ having pdf $f(x; \theta) = (1/2\theta) I_{[-\theta,\theta]}(x)$.

(a) Show that Y_1 and Y_n are joint sufficient statistics for θ.

(b) Argue that the mle of θ is $\hat{\theta} = \max(-Y_1, Y_n)$.

(c) Demonstrate that the mle $\hat{\theta}$ is a sufficient statistic for θ and thus is a minimal sufficient statistic for θ.

7.8.3. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution with pdf

$$f(x; \theta_1, \theta_2) = \left(\frac{1}{\theta_2}\right) e^{-(x-\theta_1)/\theta_2} I_{(\theta_1, \infty)}(x),$$

where $-\infty < \theta_1 < \infty$ and $0 < \theta_2 < \infty$. Find the joint minimal sufficient statistics for θ_1 and θ_2.

7.8.4. With random samples from each of the distributions given in Exercises 7.8.1(d), 7.8.2, and 7.8.3, define at least two ancillary statistics that are different from the examples given in the text. These examples illustrate, respectively, location invariant, scale invariant, and location and scale invariant statistics.

7.9 Sufficiency, Completeness and Independence

We have noted that if we have a sufficient statistic Y_1 for a parameter θ, $\theta \in \Omega$, then $h(z|y_1)$, the conditional pdf of another statistic Z, given $Y_1 = y_1$, does not depend upon θ. If, moreover, Y_1 and Z are independent, the pdf $g_2(z)$ of Z is such that $g_2(z) = h(z|y_1)$, and hence $g_2(z)$ must not depend upon θ either. So the independence of a statistic Z and the sufficient statistic Y_1 for a parameter θ means that the distribution of Z does not depend upon $\theta \in \Omega$. That is, Z is an ancillary statistic.

It is interesting to investigate a converse of that property. Suppose that the distribution of an ancillary statistic Z does not depend upon θ; then, are Z and the sufficient statistic Y_1 for θ independent? To begin our search for the answer, we know that the joint pdf of Y_1 and Z is $g_1(y_1; \theta) h(z|y_1)$, where $g_1(y_1; \theta)$ and $h(z|y_1)$

Example 7.9.4. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\theta_1, \theta_2)$, $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. In Example 7.7.2 it was proved that the mean $\overline{X}$ and the variance S^2 of the sample are joint complete sufficient statistics for θ_1 and θ_2. Consider the statistic

$$Z = \frac{\sum_1^{n-1}(X_{i+1} - X_i)^2}{\sum_1^n (X_i - \overline{X})^2} = u(X_1, X_2, \ldots, X_n),$$

which satisfies the property that $u(cx_1 + d, \ldots, cx_n + d) = u(x_1, \ldots, x_n)$. That is, the ancillary statistic Z is independent of both $\overline{X}$ and S^2. ■

In this section we have given several examples in which the complete sufficient statistics are independent of ancillary statistics. Thus in those cases, the ancillary statistics provide no information about the parameters. However, if the sufficient statistics are not complete, the ancillary statistics could provide some information as the following example demonstrates.

Example 7.9.5. We refer back to Examples 7.8.1 and 7.8.2. There the first and nth order statistics, Y_1 and Y_n, were minimal sufficient statistics for θ, where the sample arose from an underlying distribution having pdf $(\frac{1}{2})I_{(\theta-1,\theta+1)}(x)$. Often $T_1 = (Y_1 + Y_n)/2$ is used as an estimator of θ as it is a function of those sufficient statistics which is unbiased. Let us find a relationship between T_1 and the ancillary statistic $T_2 = Y_n - Y_1$.

The joint pdf of Y_1 and Y_n is

$$g(y_1, y_n; \theta) = n(n-1)(y_n - y_1)^{n-2}/2^n, \quad \theta - 1 < y_1 < y_n < \theta + 1,$$

zero elsewhere. Accordingly, the joint pdf of T_1 and T_2 is, since the absolute value of the Jacobian equals 1,

$$h(t_1, t_2; \theta) = n(n-1)t_2^{n-2}/2^n, \quad \theta - 1 + \frac{t_2}{2} < t_1 < \theta + 1 - \frac{t_2}{2}, \quad 0 < t_2 < 2,$$

zero elsewhere. Thus the pdf of T_2 is

$$h_2(t_2; \theta) = n(n-1)t_2^{n-2}(2 - t_2)/2^n, \quad 0 < t_2 < 2,$$

zero elsewhere, which of course is free of θ as T_2 is an ancillary statistic. Thus the conditional pdf of T_1, given $T_2 = t_2$, is

$$h_{1|2}(t_1|t_2; \theta) = \frac{1}{2 - t_2}, \quad \theta - 1 + \frac{t_2}{2} < t_1 < \theta + 1 - \frac{t_2}{2}, \quad 0 < t_2 < 2,$$

zero elsewhere. Note that this is uniform on the interval $(\theta - 1 + t_2/2, \theta + 1 - t_2/2)$; so the conditional mean and variance of T_1 are, respectively,

$$E(T_1|t_2) = \theta \quad \text{and} \quad \text{var}(T_1|t_2) = \frac{(2 - t_2)^2}{12}.$$

Given $T_2 = t_2$, we know something about the conditional variance of T_1. In particular, if that observed value of T_2 is large (close to 2), that variance is small and we can place more reliance on the estimator T_1. On the other hand, a small value of t_2 means that we have less confidence in T_1 as an estimator of θ. It is extremely interesting to note that this conditional variance does not depend upon the sample size n but only on the given value of $T_2 = t_2$. As the sample size increases, T_2 tends to becomes larger and, in those cases, T_1 has smaller conditional variance. ■

While Example 7.9.5 is a special one demonstrating mathematically that an ancillary statistic can provide some help in point estimation, this does actually happen in practice, too. For illustration, we know that if the sample size is large enough, then

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Of course, if the sample arises from a normal distribution, $\overline{X}$ and S are independent and T has a t-distribution with $n - 1$ degrees of freedom. Even if the sample arises from a symmetric distribution, $\overline{X}$ and S are uncorrelated and T has an approximate t-distribution and certainly an approximate standard normal distribution with sample sizes around 30 or 40. On the other hand, if the sample arises from a highly skewed distribution (say to the right), then $\overline{X}$ and S are highly correlated and the probability $P(-1.96 < T < 1.96)$ is not necessarily close to 0.95 unless the sample size is extremely large (certainly much greater than 30). Intuitively, one can understand why this correlation exists if the underlying distribution is highly skewed to the right. While S has a distribution free of μ (and hence is an ancillary), a large value of S implies a large value of $\overline{X}$, since the underlying pdf is like the one depicted in Figure 7.9.1. Of course, a small value of $\overline{X}$ (say less than the mode) requires a relatively small value of S. This means that unless n is extremely large, it is risky to say that

$$\overline{x} - \frac{1.96s}{\sqrt{n}}, \quad \overline{x} + \frac{1.96s}{\sqrt{n}}$$

provides an approximate 95 percent confidence interval with data from a very skewed distribution. As a matter of fact, the authors have seen situations in which this confidence coefficient is closer to 80 percent, rather than 95 percent, with sample sizes of 30 to 40.

EXERCISES

7.9.1. Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size $n = 4$ from a distribution having pdf $f(x; \theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere, where $0 < \theta < \infty$. Argue that the complete sufficient statistic Y_4 for θ is independent of each of the statistics Y_1/Y_4 and $(Y_1 + Y_2)/(Y_3 + Y_4)$.
Hint: Show that the pdf is of the form $(1/\theta)f(x/\theta)$, where $f(w) = 1$, $0 < w < 1$, zero elsewhere.

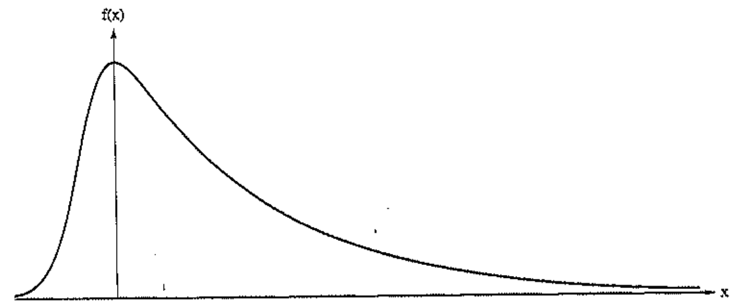

Figure 7.9.1: Graph of a Right Skewed Distribution; see also, Exercise 7.9.14

7.9.2. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$, distribution. Show that the distribution of $Z = Y_n - \overline{X}$ does not depend upon θ. Thus $\overline{Y} = \sum_1^n Y_i/n$, a complete sufficient statistic for θ is independent of Z.

7.9.3. Let $X_1, X_2, \ldots, X_n$ be iid with the distribution $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$. Prove that a necessary and sufficient condition that the statistics $Z = \sum_1^n a_i X_i$ and $Y = \sum_1^n X_i$, a complete sufficient statistic for θ, are independent is that $\sum_1^n a_i = 0$.

7.9.4. Let X and Y be random variables such that $E(X^k)$ and $E(Y^k) \neq 0$ exist for $k = 1, 2, 3, \ldots$. If the ratio X/Y and its denominator Y are independent, prove that $E[(X/Y)^k] = E(X^k)/E(Y^k), k = 1, 2, 3, \ldots$.
Hint: Write $E(X^k) = E[Y^k(X/Y)^k]$.

7.9.5. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution that has pdf $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere. Show that the ratio $R = nY_1 \Big/ \sum_1^n Y_i$ and its denominator (a complete sufficient statistic for θ) are independent. Use the result of the preceding exercise to determine $E(R^k)$, $k = 1, 2, 3, \ldots$.

7.9.6. Let $X_1, X_2, \ldots, X_5$ be iid with pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Show that $(X_1 + X_2)/(X_1 + X_2 + \cdots + X_5)$ and its denominator are independent.
Hint: The pdf $f(x)$ is a member of $\{f(x;\theta) : 0 < \theta < \infty\}$, where $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, zero elsewhere.

7.9.7. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from the normal distribution $N(\theta_1, \theta_2)$, $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. Show that the joint complete sufficient statistics $\overline{X} = \overline{Y}$ and S^2 for θ_1 and θ_2 are independent of each of $(Y_n - \overline{Y})/S$ and $(Y_n - Y_1)/S$.

7.9.8. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a distribution with the pdf

$$f(x;\theta_1,\theta_2) = \frac{1}{\theta_2}\exp\left(-\frac{x-\theta_1}{\theta_2}\right),$$

$\theta_1 < x < \infty$, zero elsewhere, where $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. Show that the joint complete sufficient statistics Y_1 and $\overline{X} = \overline{Y}$ for for the parameters θ_1 and θ_2 are independent of $(Y_2 - Y_1) \Big/ \sum_1^n (Y_i - Y_1)$.

7.9.9. Let $X_1, X_2, \ldots, X_5$ be a random sample of size $n = 5$ from the normal distribution $N(0, \theta)$.

(a) Argue that the ratio $R = (X_1^2 + X_2^2)/(X_1^2 + \cdots + X_5^2)$ and its denominator $(X_1^2 + \cdots + X_5^2)$ are independent.

(b) Does $5R/2$ have an F-distribution with 2 and 5 degrees of freedom? Explain your answer.

(c) Compute $E(R)$ using Exercise 7.9.4.

7.9.10. Referring to Example 7.9.5 of this section, determine c so that

$$P(-c < T_1 - \theta < c | T_2 = t_2) = 0.95.$$

Use this result to find a 95 percent confidence interval for θ, given $T_2 = t_2$; and note how its length is smaller when the range of t_2 is larger.

7.9.11. Show that $Y = |X|$ is a complete sufficient statistic for $\theta > 0$, where X has the pdf $f_X(x;\theta) = 1/(2\theta)$, for $-\theta < x < \theta$, zero elsewhere. Show that $Y = |X|$ and $Z = \text{sgn}(X)$ are independent.

7.9.12. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a $N(\theta, \sigma^2)$ distribution, where σ^2 is fixed but arbitrary. Then $\overline{Y} = \overline{X}$ is a complete sufficient statistic for θ. Consider another estimator T of θ, such as $T = (Y_i + Y_{n+1-i})/2$, for $i = 1, 2, \ldots, [n/2]$ or T could be any weighted average of these latter statistics.

(a) Argue that $T - \overline{X}$ and $\overline{X}$ are independent random variables.

(b) Show that $\text{Var}(T) = \text{Var}(\overline{X}) + \text{Var}(T - \overline{X})$.

(c) Since we know the $\text{Var}(\overline{X}) = \sigma^2/n$, it might be more efficient to estimate the $\text{Var}(T)$ by estimating the $\text{Var}(T - \overline{X})$ by Monte Carlo methods rather than doing that with $\text{Var}(T)$ directly, because $\text{Var}(T) \geq \text{Var}(T - \overline{X})$. This is often called the *Monte Carlo Swindle.*

7.9.13. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with pdf $f(x;\theta) = (1/2)\theta^3 x^2 e^{-\theta x}$, $0 < x < \infty$, zero elsewhere, where $0 < \theta < \infty$:

(a) Find the mle, $\hat{\theta}$, of θ. Is $\hat{\theta}$ unbiased?

Hint: Find the pdf of $Y = \sum_1^n X_i$ and then compute $E(\hat{\theta})$.

(b) Argue that Y is a complete sufficient statistic for θ.

(c) Find the MVUE of θ.

(d) Show that X_1/Y and Y are independent.

(e) What is the distribution of X_1/Y?

7.9.14. The pdf depicted in Figure 7.9.1 is given by

$$f_{m_2}(x) = e^x(1 + m_2^{-1}e^x)^{-(m_2+1)}, \quad -\infty < x < \infty, \tag{7.9.2}$$

where $m_2 > 0$, (the pdf graphed is for $m_2 = 0.1$). This is a member of a large family of pdfs, log F-family, which are useful in survival (lifetime) analysis; see Chapter 3 of Hettmansperger and McKean (1998).

(a) Let W be a random variable with pdf (7.9.2). Show that $W = \log Y$, where Y has an F-distribution with 2 and $2m_2$ degrees of freedom.

(b) Show that the pdf becomes the logistic (6.1.8) if $m_2 = 1$.

(c) Consider the location model where

$$X_i = \theta + W_i \quad i = 1, \ldots, n,$$

where $W_1, \ldots, W_n$ are iid with pdf (7.9.2). Similar to the logistic location model, the order statistics are minimal sufficient for this model. Show, similar to Example 6.1.4, that the mle of θ exists.

Chapter 8

Optimal Tests of Hypotheses

8.1 Most Powerful Tests

In Section 5.5 we introduced the concept of hypotheses testing and followed it with the introduction of likelihood ratio tests in Chapter 6. In this chapter we discuss certain best tests.

We are interested in a random variable X which has pdf or pmf $f(x;\theta)$ where $\theta \in \Omega$. We assume that $\theta \in \omega_0$ or $\theta \in \omega_1$ where ω_0 and ω_1 are subsets of Ω and $\omega_0 \cup \omega_1 = \Omega$. We label the hypotheses as

$$H_0: \theta \in \omega_0 \text{ versus } H_1: \theta \in \omega_1. \tag{8.1.1}$$

The hypothesis H_0 is referred to as the **null hypothesis** while H_1 is referred to as the **alternative hypothesis**. The test of H_0 versus H_1 is based on a sample $X_1, \ldots, X_n$ from the distribution of X. In this chapter we will often use the vector $\mathbf{X}' = (X_1, \ldots, X_n)$ to denote the random sample and $\mathbf{x}' = (x_1, \ldots, x_n)$ to denote the values of the sample. Let $\mathcal{S}$ denote the support of the random sample $\mathbf{X}' = (X_1, \ldots, X_n)$.

A test of H_0 versus H_1 is based on a subset C of $\mathcal{S}$. This set C is called the **critical region** and its corresponding decision rule is:

$$\begin{array}{ll} \text{Reject } H_0, \text{ (Accept } H_1), & \text{if } \mathbf{X} \in C \\ \text{Retain } H_0, \text{ (Reject } H_1), & \text{if } \mathbf{X} \in C^c. \end{array} \tag{8.1.2}$$

Note that a test is defined by its critical region. Conversely a critical region defines a test.

Recall that the 2×2 decision table, Table 5.5.1, summarizes the results of the hypothesis test in terms of the true state of nature. Besides the correct decisions, two errors can occur. A **Type I** error occurs if H_0 is rejected when it is true while a **Type II** error occurs if H_0 is accepted when H_1 is true. The **size** or **significance level** of the test is the probability of a Type I error, i.e.,

$$\alpha = \max_{\theta \in \omega_0} P_\theta(\mathbf{X} \in C). \tag{8.1.3}$$

(c) What is Rao's score statistic?

6.3.16. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean $\theta > 0$. Test $H_0: \theta = 2$ against $H_1: \theta \neq 2$ using

(a) $-2 \log \Lambda$.

(b) a Wald-type statistic.

(c) Rao's score statistic.

6.3.17. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $\Gamma(\alpha, \beta)$-distribution where α is known and $\beta > 0$. Determine the likelihood ratio test for $H_0: \beta = \beta_0$ against $H_1: \beta \neq \beta_0$.

6.3.18. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a uniform distribution on $(0, \theta)$, where $\theta > 0$.

(a) Show that Λ for testing $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$ is $\Lambda = (Y_n/\theta_0)^n$, $Y_n \leq \theta_0$, and $\Lambda = 0$, if $Y_n > \theta_0$.

(b) When H_0 is true, show that $-2 \log \Lambda$ has an exact $\chi^2(2)$ distribution, not $\chi^2(1)$. Note that the regularity conditions are not satisfied.

6.4 Multiparameter Case: Estimation

In this section we discuss the case where $\boldsymbol{\theta}$ is a vector of p parameters. There are analogs to the theorems in the previous sections in which θ is a scalar and we present their results, but for the most part, without proofs. The interested reader can find additional information in more advanced books; see, for instance, Lehmann and Casella (1998) and Rao (1973).

Let $X_1, \ldots, X_n$ be iid with common pdf $f(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Omega \subset R^p$. As before, the likelihood function and its log are given by

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n f(x_i; \boldsymbol{\theta}) \\ l(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta}), \end{aligned} \qquad (6.4.1)$$

for $\boldsymbol{\theta} \in \Omega$. The theory requires additional regularity conditions which are listed in the Appendix A, (A.1.1). In keeping with our number scheme in the last two sections, we have labeled these (R6)-(R9). In this section of Chapter 6, when we say under regularity conditions we mean all of the conditions of (6.1.1), (6.2.1), (6.2.2), and (A.1.1) which are relevant to the argument. The discrete case follows in the same way as the continuous case, so in general we will state material in terms of the continuous case.

Note that the proof of Theorem 6.1.1 did not depend on whether the parameter is a scalar or a vector. Therefore, with probability going to one, $L(\boldsymbol{\theta})$ is maximized

at the true value of $\boldsymbol{\theta}$. Hence, as an estimate of $\boldsymbol{\theta}$ we will consider the value which maximizes $L(\boldsymbol{\theta})$ or equivalently solves the vector equation $(\partial/\partial\boldsymbol{\theta})l(\boldsymbol{\theta}) = \mathbf{0}$. If it exists this value will be called the **maximum likelihood estimator** (mle) and we will denote it by $\widehat{\boldsymbol{\theta}}$. Often we are interested in a function of $\boldsymbol{\theta}$, say, the parameter $\eta = g(\boldsymbol{\theta})$. Because the second part of the proof of Theorem 6.1.2 remains true for $\boldsymbol{\theta}$ as a vector, $\widehat{\eta} = g(\widehat{\boldsymbol{\theta}})$ is the mle of η.

Example 6.4.1 (Maximum Likelihood Estimates of the Normal pdf). Suppose $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$. In this case, $\boldsymbol{\theta} = (\mu, \sigma^2)'$ and Ω is the product space $(-\infty, \infty) \times (0, \infty)$. The log of the likelihood simplifies to

$$l(\mu, \sigma^2) = -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \qquad (6.4.2)$$

Taking partial derivatives of (6.4.2) with respect to μ and σ and setting them to 0, we get the simultaneous equations,

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0. \end{aligned}$$

Solving these equations, we obtain $\widehat{\mu} = \overline{X}$ and $\widehat{\sigma} = \sqrt{(1/n)\sum_{i=1}^n (X_i - \overline{X})^2}$ as solutions. A check of the second partials shows that these maximize $l(\mu, \sigma^2)$, so these are the mles. Also, by Theorem 6.1.2, $(1/n)\sum_{i=1}^n (X_i - \overline{X})^2$ is the mle of σ^2. We know from our discussion in Section 5.4, that these are consistent estimates of μ and σ^2, respectively; that $\widehat{\mu}$ is an unbiased estimate of μ and that $\widehat{\sigma^2}$ is a biased estimate of σ^2 whose bias vanishes as $n \to \infty$. ■

Example 6.4.2 (General Laplace pdf). Let $X_1, X_2, \ldots, X_n$ be a random sample from the Laplace pdf $f_X(x) = (2b)^{-1} \exp\{-|x - a|/b\}$, $-\infty < x < \infty$, where the parameters (a, b) are in the space $\Omega = \{(a, b) : -\infty < a < \infty, b > 0\}$. Recall in the last sections we looked at the special case where $b = 1$. As we now show, the mle of a is the sample median, regardless of the value of b. The log of the likelihood function is,

$$l(a, b) = -n \log 2 - n \log b - \sum_{i=1}^n \left| \frac{x_i - a}{b} \right|.$$

The partial of $l(a, b)$ with respect to a is

$$\frac{\partial l(a, b)}{\partial a} = \frac{1}{b} \sum_{i=1}^n \operatorname{sgn}\left\{ \frac{x_i - a}{b} \right\} = \frac{1}{b} \sum_{i=1}^n \operatorname{sgn}\{x_i - a\},$$

where the second equality follows because $b > 0$. Setting this partial to 0, we obtain the mle of a to be $Q_2 = \operatorname{med}\{X_1, X_2, \ldots, X_n\}$, just as in Example 6.1.3. Hence, the

In Example 8.3.1, in testing the equality of the means of two normal distributions, it was assumed that the unknown variances of the distributions were equal. Let us now consider the problem of testing the equality of these two unknown variances.

Example 8.3.2. We are given the independent random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ from the distributions, which are $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$, respectively. We have

$$\Omega = \{(\theta_1, \theta_2, \theta_3, \theta_4) : -\infty < \theta_1, \theta_2 < \infty, 0 < \theta_3, \theta_4 < \infty\}.$$

The hypothesis $H_0 : \theta_3 = \theta_4$, unspecified, with θ_1 and θ_2 also unspecified, is to be tested against all alternatives. Then

$$\omega = \{(\theta_1, \theta_2, \theta_3, \theta_4) : -\infty < \theta_1, \theta_2 < \infty, 0 < \theta_3 = \theta_4 < \infty\}.$$

It is easy to show (see Exercise 8.3.8) that the statistic defined by $\Lambda = L(\hat{\omega})/L(\hat{\Omega})$ is a function of the statistic

$$F = \frac{\sum_1^n (X_i - \overline{X})^2/(n-1)}{\sum_1^m (Y_i - \overline{Y})^2/(m-1)}. \tag{8.3.8}$$

If $\theta_3 = \theta_4$, this statistic F has an F-distribution with $n-1$ and $m-1$ degrees of freedom. The hypothesis that $(\theta_1, \theta_2, \theta_3, \theta_4) \in \omega$ is rejected if the computed $F \leq c_1$ or if the computed $F \geq c_2$. The constants c_1 and c_2 are usually selected so that, if $\theta_3 = \theta_4$,

$$P(F \leq c_1) = P(F \geq c_2) = \frac{\alpha_1}{2}$$

where α_1 is the desired significance level of this test. ■

Example 8.3.3. Let the independent random variables X and Y have distributions that are $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$. In Example 8.3.1 we derived the likelihood ratio test statistic T of the hypothesis $\theta_1 = \theta_2$ when $\theta_3 = \theta_4$, while in Example 8.3.2 we obtained the likelihood ratio test statistic F of the hypothesis $\theta_3 = \theta_4$. The hypothesis that $\theta_1 = \theta_2$ is rejected if the computed $|T| \geq c$, where the constant c is selected so that $\alpha_2 = P(|T| \geq c; \theta_1 = \theta_2, \theta_3 = \theta_4)$ is the assigned significance level of the test. We shall show that, if $\theta_3 = \theta_4$, the likelihood ratio test statistics for equality of variances and equality of means, respectively F and T, are independent. Among other things, this means that if these two tests based on F and T, respectively, are performed sequentially with significance levels α_1 and α_2, the probability of accepting both these hypotheses, when they are true, is $(1-\alpha_1)(1-\alpha_2)$. Thus the significance level of this joint test is $\alpha = 1 - (1-\alpha_1)(1-\alpha_2)$.

Independence of F and T, when $\theta_3 = \theta_4$, can be established by an appeal to sufficiency and completeness. The three statistics $\overline{X}$, $\overline{Y}$, and $\sum_1^n (X_i - \overline{X})^2 + \sum_1^n (Y_i - \overline{Y})^2$ are joint complete sufficient statistics for the three parameters θ_1, θ_2, and $\theta_3 = \theta_4$. Obviously, the distribution of F does not depend upon θ_1, θ_2, or $\theta_3 = \theta_4$, and hence F is independent of the three joint complete sufficient statistics. However, T is a function of these three joint complete sufficient statistics alone, and, accordingly, T is independent of F. It is important to note that these two statistics are independent whether $\theta_1 = \theta_2$ or $\theta_1 \neq \theta_2$. This permits us to calculate probabilities other than the significance level of the test. For example, if $\theta_3 = \theta_4$ and $\theta_1 \neq \theta_2$, then

$$P(c_1 < F < c_2, \ |T| \geq c) = P(c_1 < F < c_2)P(|T| \geq c).$$

The second factor in the right-hand member is evaluated by using the probabilities of a noncentral t-distribution. Of course, if $\theta_3 = \theta_4$ and the difference $\theta_1 - \theta_2$ is large, we would want the preceding probability to be close to 1 because the event $\{c_1 < F < c_2, \ |T| \geq c\}$ leads to a correct decision, namely, accept $\theta_3 = \theta_4$ and reject $\theta_1 = \theta_2$. ■

Remark 8.3.2. We caution the reader on this last test for the equality of two variances. In Remark 8.3.1, we discussed that the one- and two-sample t-tests for means are asymptotically correct. The two sample variance test of the last example is not; see, for example, page 126 of Hettmansperger and McKean (1998). If the underlying distributions are not normal then the F-critical values may be far from valid critical values, (unlike the t-critical values for the means tests as discussed in Remark 8.3.1). In a large simulation study Conover, Johnson and Johnson (1981) showed that instead of having the nominal size of $\alpha = 0.05$, the F-test for variances using the F-critical values could have significance levels as high as 0.80, in certain nonnormal situations. Thus, the two sample F-test for variances does not possess robustness of validity. It should only be used in situations where the assumption of normality can be justified. See Exercise 8.3.14 for an illustrative data set. ■

In the above examples, we were able to determine the null distribution of the test statistic. This is often impossible in practice. As discussed in Chapter 6, though, minus twice the log of the likelihood ratio test statistic is asymptotically χ^2 under H_0. Hence, we can obtain an approximate test in most situations.

EXERCISES

8.3.1. In Example 8.3.1, suppose $n = m = 8$, $\overline{x} = 75.2$, $\overline{y} = 78.6$, $\sum_1^8 (x_i - \overline{x})^2 = 71.2$, $\sum_1^8 (y_i - \overline{y})^2 = 54.8$. If we use the test derived in that example, do we accept or reject $H_0 : \theta_1 = \theta_2$ at the 5 percent significance level? Obtain the p-value, see Remark (5.6.1), of the test.

8.3.2. Verify Equations (8.3.3) of Example 8.3.1 of this section.

and σ^2 must be estimated, but in this discussion, we assume that they are known. From theory we know that the probability is 0.997 that $\bar{x}$ is between

$$\text{LCL} = \mu - \frac{3\sigma}{\sqrt{n}} \quad \text{and} \quad \text{UCL} = \mu + \frac{3\sigma}{\sqrt{n}}.$$

These two values are called the lower (LCL) and upper (UCL) control limits, respectively. Samples like these are taken periodically resulting in a sequence of means, say $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$. These are usually plotted; and if they are between the LCL and UCL, we say that the process is *in control*. If one falls outside the limits, this would suggest that the mean μ has shifted, and the process would be investigated.

It was recognized by some that there could be a shift in the mean, say from μ to $\mu + (\sigma/\sqrt{n})$; and it would still be difficult to detect that shift with a single sample mean for now the probability of a single $\bar{x}$ exceeding UCL is only about 0.023. This means that we would need about $1/0.023 \approx 43$ samples, each of size n, on the average before detecting such a shift. This seems too long; so statisticians recognized that they should be cumulating experience as the sequence $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$ is observed in order to help them detect the shift sooner. It is the practice to compute the standardized variable $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$; thus we state the problem in these terms and provide the solution given by a sequential probability ratio test.

Here Z is $N(\theta, 1)$, and we wish to test $H_0: \theta = 0$ against $H_1: \theta = 1$ using the sequence of iid random variables $Z_1, Z_2, \ldots, Z_m, \ldots$. We use m rather than n, as the latter is the size of the samples taken periodically. We have

$$\frac{L(0,m)}{L(1,m)} = \frac{\exp\left[-\sum z_i^2/2\right]}{\exp\left[-\sum (z_i-1)^2/2\right]} = \exp\left[-\sum_{i=1}^{n}(z_i - 0.5)\right].$$

Thus

$$k_0 < \exp\left[-\sum_{i=1}^{n}(z_i - 0.5)\right] < k_1$$

can be written as

$$h = -\log k_0 > \sum_{i=1}^{m}(z_i - 0.5) > -\log k_1 = -h.$$

It is true that $-\log k_0 = \log k_1$ when $\alpha_a = \beta_a$. Often, $h = -\log k_0$ is taken to be about 4 or 5, suggesting that $\alpha_a = \beta_a$ is small, like 0.01. As $\sum(z_i - 0.5)$ is cumulating the sum of $z_i - 0.5$, $i = 1, 2, 3, \ldots$, these procedures are often called CUSUMS. If the CUSUM $= \sum(z_i - 0.5)$ exceeds h, we would investigate the process, as it seems that the mean has shifted upward. If this shift is to $\theta = 1$, the theory associated with these procedures shows that we need only eight or nine samples on the average, rather than 43, to detect this shift. For more information about these methods, the reader is referred to one of the many books on quality improvement through statistical methods. What we would like to emphasize here is that through sequential methods (not only the sequential probability ratio test), we should take advantage of all past experience that we can gather in making inferences.

EXERCISES

8.4.1. Let X be $N(0, \theta)$ and, in the notation of this section, let $\theta' = 4$, $\theta'' = 9$, $\alpha_a = 0.05$, and $\beta_a = 0.10$. Show that the sequential probability ratio test can be based upon the statistic $\sum_1^n X_i^2$. Determine $c_0(n)$ and $c_1(n)$.

8.4.2. Let X have a Poisson distribution with mean θ. Find the sequential probability ratio test for testing $H_0: \theta = 0.02$ against $H_1: \theta = 0.07$. Show that this test can be based upon the statistic $\sum_1^n X_i$. If $\alpha_a = 0.20$ and $\beta_a = 0.10$, find $c_0(n)$ and $c_1(n)$.

8.4.3. Let the independent random variables Y and Z be $N(\mu_1, 1)$ and $N(\mu_2, 1)$, respectively. Let $\theta = \mu_1 - \mu_2$. Let us observe independent observations from each distribution, say $Y_1, Y_2, \ldots$ and $Z_1, Z_2, \ldots$. To test sequentially the hypothesis $H_0: \theta = 0$ against $H_1: \theta = \frac{1}{2}$, use the sequence $X_i = Y_i - Z_i$, $i = 1, 2, \ldots$. If $\alpha_a = \beta_a = 0.05$, show that the test can be based upon $\overline{X} = \overline{Y} - \overline{Z}$. Find $c_0(n)$ and $c_1(n)$.

8.4.4. Suppose that a manufacturing process makes about 3 percent defective items, which is considered satisfactory for this particular product. The managers would like to decrease this to about 1 percent and clearly want to guard against a substantial increase, say to 5 percent. To monitor the process, periodically $n = 100$ items are taken and the number X of defectives counted. Assume that X is $b(n = 100, p = \theta)$. Based on a sequence $X_1, X_2, \ldots, X_m, \ldots$, determine a sequential probability ratio test that tests $H_0: \theta = 0.01$ against $H_1: \theta = 0.05$. (Note that $\theta = 0.03$, the present level, is in between these two values.) Write this test in the form

$$h_0 > \sum_{i=1}^{m}(x_i - nd) > h_1$$

and determine d, h_0, and h_1 if $\alpha_a = \beta_a = 0.02$.

8.4.5. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere.

(a) Find a complete sufficient statistic for θ.

(b) If $\alpha_a = \beta_a = \frac{1}{10}$, find the sequential probability ratio test of $H_0: \theta = 2$ against $H_1: \theta = 3$.

8.5 Minimax and Classification Procedures

We have considered several procedures which may be used in problems of point estimation. Among these were decision function procedures (in particular, minimax decisions). In this section, we apply minimax procedures to the problem of testing a or, for brevity,

$$ax + by \le c. \tag{8.5.3}$$

That is, if this linear function of x and y in the left-hand member of inequality (8.5.3) is less than or equal to a constant, we classify (x, y) as coming from the bivariate normal distribution with means μ_1'' and μ_2''. Otherwise, we classify (x, y) as arising from the bivariate normal distribution with means μ_1' and μ_2'. Of course, if the prior probabilities can be assigned as discussed in Remark 8.5.1 then k and thus c can be found easily; see Exercise 8.5.3. ■

Once the rule for classification is established, the statistician might be interested in the two probabilities of misclassifications using that rule. The first of these two is associated with the classification of (x, y) as arising from the distribution indexed by θ'' if, in fact, it comes from that index by θ'. The second misclassification is similar, but with the interchange of θ' and θ''. In the preceding example, the probabilities of these respective misclassifications are

$$P(aX + bY \le c; \mu_1', \mu_2') \quad \text{and} \quad P(aX + bY > c; \mu_1'', \mu_2'').$$

The distribution of $Z = aX + bY$ follows easily from Theorem 3.5.1, it follows that the distribution of $Z = aX + bY$ is given by,

$$N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + 2ab\rho\sigma_1\sigma_2 + b^2\sigma_2^2).$$

With this information, it is easy to compute the probabilities of misclassifications; see Exercise 8.5.3.

One final remark must be made with respect to the use of the important classification rule established in Example 8.5.2. In most instances the parameter values μ_1', μ_2' and μ_1'', μ_2'' as well as σ_1^2, σ_2^2, and ρ are unknown. In such cases the statistician has usually observed a random sample (frequently called a *training sample*) from each of the two distributions. Let us say the samples have sizes n' and n'', respectively, with sample characteristics,

$$\overline{x}', \overline{y}', (s_x')^2, (s_y')^2, r' \quad \text{and} \quad \overline{x}'', \overline{y}'', (s_x'')^2, (s_y'')^2, r'';$$

see Section 9.7 for the definiton of r. If in inequality (8.5.3) the parameters $\mu_1', \mu_2', \mu_1'', \mu_2'', \sigma_1^2, \sigma_2^2$, and $\rho\sigma_1\sigma_2$ are replaced by the unbiased estimates

$$\overline{x}', \overline{y}', \overline{x}'', \overline{y}'', \frac{(n'-1)(s_x')^2 + (n''-1)(s_x'')^2}{n'+n''-2}, \frac{(n'-1)(s_y')^2 + (n''-1)(s_y'')^2}{n'+n''-2},$$
$$\frac{(n'-1)r's_x's_y' + (n''-1)r''s_x''s_y''}{n'+n''-2},$$

the resulting expression in the left-hand member is frequently called Fisher's **linear discriminant function**. Since those parameters have been estimated, the distribution theory associated with $aX + bY$ does provide an approximation.

Although we have considered only bivariate distributions in this section, the results can easily be extended to multivariate normal distributions using the results of Section 3.5; see also, Chapter 6 of Seber (1984).

EXERCISES

8.5.1. Let $X_1, X_2, \ldots, X_{20}$ be a random sample of size 20 from a distribution which is $N(\theta, 5)$. Let $L(\theta)$ represent the joint pdf of $X_1, X_2, \ldots, X_{20}$. The problem is to test $H_0: \theta = 1$ against $H_1: \theta = 0$. Thus $\Omega = \{\theta : \theta = 0, 1\}$.

(a) Show that $L(1)/L(0) \le k$ is equivalent to $\overline{x} \le c$.

(b) Find c so that the significance level is $\alpha = 0.05$. Compute the power of this test if H_1 is true.

(c) If the loss function is such that $\mathcal{L}(1,1) = \mathcal{L}(0,0) = 0$ and $\mathcal{L}(1,0) = \mathcal{L}(0,1) > 0$, find the minimax test. Evaluate the power function of this test at the points $\theta = 1$ and $\theta = 0$.

8.5.2. Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size 10 from a Poisson distribution with parameter θ. Let $L(\theta)$ be the joint pdf of $X_1, X_2, \ldots, X_{10}$. The problem is to test $H_0: \theta = \frac{1}{2}$ against $H_1: \theta = 1$.

(a) Show that $L(\frac{1}{2})/L(1) \le k$ is equivalent to $y = \sum_1^n x_i \ge c$.

(b) In order to make $\alpha = 0.05$, show that H_0 is rejected if $y > 9$ and, if $y = 9$, reject H_0 with probability $\frac{1}{2}$ (using some auxiliary random experiment).

(c) If the loss function is such that $\mathcal{L}(\frac{1}{2}, \frac{1}{2}) = \mathcal{L}(1,1) = 0$ and $\mathcal{L}(\frac{1}{2}, 1) = 1$ and $\mathcal{L}(1, \frac{1}{2}) = 2$ show that the minimax procedure is to reject H_0 if $y > 6$ and, if $y = 6$, reject H_0 with probability 0.08 (using some auxiliary random experiment).

8.5.3. In Example 8.5.2 let $\mu_1' = \mu_2' = 0$, $\mu_1'' = \mu_2'' = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, and $\rho = \frac{1}{2}$.

(a) Find the distribution of the linear function $aX + bY$.

(b) With $k = 1$, compute $P(aX + bY \le c; \mu_1' = \mu_2' = 0)$ and $P(aX + bY > c; \mu_1'' = \mu_2'' = 1)$.

8.5.4. Determine Newton's algorithm to find the solution of equation (8.5.2). If software is available, write a program which performs your algorithm and then show that the solution is $c = 76.8$. If software is not available, solve (8.5.2) by "trial and error."

8.5.5. Let X and Y have the joint pdf.

$$f(x, y; \theta_1, \theta_2) = \frac{1}{\theta_1\theta_2} \exp\left(-\frac{x}{\theta_1} - \frac{y}{\theta_2}\right), \quad 0 < x < \infty, \quad 0 < y < \infty,$$

zero elsewhere, where $0 < \theta_1$, $0 < \theta_2$. An observation (x, y) arises from the joint distribution with parameters equal to either $(\theta_1' = 1, \theta_2' = 5)$ or $(\theta_1'' = 3, \theta_2'' = 2)$. Determine the form of the classification rule.

8.5.6. Let X and Y have a joint bivariate normal distribution. An observation (x, y) arises from the joint distribution with parameters equal to either

$$\mu_1' = \mu_2' = 0, \ (\sigma_1^2)' = (\sigma_2^2)' = 1, \ \rho' = \tfrac{1}{2}$$

or

$$\mu_1'' = \mu_2'' = 1, \ (\sigma_1^2)'' = 4, \ (\sigma_2^2)'' = 9, \ \rho'' = \tfrac{1}{2}.$$

Show that the classification rule involves a second-degree polynomial in x and y.

8.5.7. Let $\mathbf{W}' = (W_1, W_2)$ be an observation from one of two bivariate normal distributions, I and II, each with $\mu_1 = \mu_2 = 0$ but with the respective variance-covariance matrices

$$\mathbf{V}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \quad \text{and} \quad \mathbf{V}_2 = \begin{pmatrix} 3 & 0 \\ 0 & 12 \end{pmatrix}.$$

How would you classify $\mathbf{W}$ into I or II?

Chapter 9

Inferences about Normal Models

9.1 Quadratic Forms

A homogeneous polynomial of degree 2 in n variables is called a **quadratic** form in those variables. If both the variables and the coefficients are real, the form is called a **real quadratic** form. Only real quadratic forms will be considered in this book. To illustrate, the form $X_1^2 + X_1X_2 + X_2^2$ is a quadratic form in the two variables X_1 and X_2; the form $X_1^2 + X_2^2 + X_3^2 - 2X_1X_2$ is a quadratic form in the three variables X_1, X_2, and X_3; but the form $(X_1 - 1)^2 + (X_2 - 2)^2 = X_1^2 + X_2^2 - 2X_1 - 4X_2 + 5$ is not quadratic form in X_1 and X_2, although it is a quadratic form in the variables $X_1 - 1$ and $X_2 - 2$.

Let $\overline{X}$ and S^2 denote, respectively, the mean and the variance of a random sample $X_1, X_2, \ldots, X_n$ from an arbitrary distribution. Thus

$$\begin{aligned}(n-1)S^2 &= \sum_1^n (X_i - \overline{X})^2 = \sum_1^n \left(X_i - \frac{X_1 + X_2 + \cdots + X_n}{n}\right)^2 \\ &= \frac{n-1}{n}(X_1^2 + X_2^2 + \cdots + X_n^2) \\ &\quad - \frac{2}{n}(X_1X_2 + \cdots + X_1X_n + \cdots + X_{n-1}X_n)\end{aligned}$$

is a quadratic form in the n variables $X_1, X_2, \ldots, X_n$. If the sample arises from a distribution that is $N(\mu, \sigma^2)$, we know that the random variable $(n-1)S^2/\sigma^2$ is $\chi^2(n-1)$ regardless of the value of μ. This fact proved useful in our search for a confidence interval for σ^2 when μ is unknown.

It has been seen that tests of certain statistical hypotheses require a statistic that is a quadratic form. For instance, Example 8.2.2 made use of the statistic $\sum_1^n X_i^2$, which is a quadratic form in the variables $X_1, X_2, \ldots, X_n$. Later in this