CHAPTER 11

Confidence Estimation

11.1 INTRODUCTION

In many problems of statistical inference the experimenter is interested in constructing a family of sets that contain the true (unknown) parameter value with a specified (high) probability. If X, for example, represents the length of life of a piece of equipment, the experimenter is interested in a lower bound $\underline{\theta}$ for the mean θ of X. Since $\underline{\theta} = \underline{\theta}(X)$ will be a function of the observations, one cannot ensure with probability 1 that $\underline{\theta}(X) \le \theta$. All that one can do is to choose a number $1-\alpha$ that is close to 1 so that $P_\theta\{\underline{\theta}(X) \le \theta\} \ge 1-\alpha$ for all θ. Problems of this type are called problems of *confidence estimation*. In this chapter we restrict ourselves mostly to the case where $\Theta \subseteq \mathcal{R}$ and consider the problem of setting confidence limits for the parameter θ.

In Section 11.2 we introduce the basic ideas of confidence estimation. Section 11.3 deals with various methods of finding confidence intervals, while Section 11.4 deals with shortest-length confidence intervals. In Section 11.5 we study unbiased and equivariant confidence intervals.

11.2 SOME FUNDAMENTAL NOTIONS OF CONFIDENCE ESTIMATION

So far we have considered a random variable or some function of it as the basic observable quantity. Let X be an RV, and a, b be two given positive real numbers. Then

$$\begin{aligned} P\{a < X < b\} &= P\{a < X \text{ and } X < b\} \\ &= P\left\{\frac{bX}{a} > b \text{ and } X < b\right\} \\ &= P\left\{X < b < \frac{bX}{a}\right\}, \end{aligned}$$

and if we know the distribution of X and a, b, we can determine the probability $P\{a < X < b\}$. Consider the interval $I(X) = (X, bX/a)$. This is an interval with endpoints that are functions of the RV X, and hence it takes the value $(x, bx/a)$ when X takes the value x. In other words, $I(X)$ assumes the value $I(x)$ whenever X assumes the value x. Thus $I(X)$ is a random quantity and is an example of a *random interval*. Note that $I(X)$ includes the value b with a certain fixed probability. For example, if $b = 1, a = \frac{1}{2}$ and X is $U(0, 1)$, the interval $(X, 2X)$ includes point 1 with probability $\frac{1}{2}$. We note that $I(X)$ is a family of intervals with associated *coverage probability* $P(I(X) \ni 1) = \frac{1}{2}$. It has (random) length $l(I(X)) = 2X - X = X$. In general, the larger the length of the interval, the larger the coverage probability. Let us formalize these notions.

Definition 1. Let $\mathcal{P}_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}_k$, be the set of probability distributions of an RV $\mathbf{X}$. A family of subsets $S(\mathbf{x})$ of Θ, where $S(\mathbf{x})$ depends on the observation $\mathbf{x}$ but not on $\boldsymbol{\theta}$, is called a *family of random sets*. If, in particular, $\Theta \subseteq \mathcal{R}$ and $S(\mathbf{x})$ is an interval $(\underline{\theta}(\mathbf{x}), \overline{\theta}(\mathbf{x}))$, where $\underline{\theta}(\mathbf{x})$ and $\overline{\theta}(\mathbf{x})$ are functions of $\mathbf{x}$ alone (and not θ), we call $S(\mathbf{X})$ a random interval with $\underline{\theta}(\mathbf{X})$ and $\overline{\theta}(\mathbf{X})$ as lower and upper bounds, respectively. $\overline{\theta}(\mathbf{X})$ may be $-\infty$, and $\overline{\theta}(\mathbf{X})$ may be $+\infty$.

In a wide variety of inference problems, one is not interested in estimating the parameter or testing some hypothesis concerning it. Rather, one wishes to establish a lower or an upper bound, or both, for the real-valued parameter. For example, if X is the time to failure of a piece of equipment, one may be interested in a lower bound for the mean of X. If the RV X measures the toxicity of a drug, the concern is to find an upper bound for the mean. Similarly, if the RV X measures the nicotine content of a certain brand of cigarettes, one may be interested in determining an upper and a lower bound for the average nicotine content of these cigarettes.

In this chapter we are interested in the *problem of confidence estimation*, namely, that of finding a family of random sets $S(\mathbf{x})$ for a parameter $\boldsymbol{\theta}$ such that for a given $\alpha, 0 < \alpha < 1$ (usually small),

$$P_{\boldsymbol{\theta}}\{S(\mathbf{X}) \ni \boldsymbol{\theta}\} \geq 1 - \alpha \qquad \text{for all } \boldsymbol{\theta} \in \Theta. \tag{1}$$

We restrict our attention mainly to the case where $\theta \in \Theta \subseteq \mathcal{R}$.

Definition 2. Let $\theta \in \Theta \subseteq \mathcal{R}$ and $0 < \alpha < 1$. A function $\underline{\theta}(\mathbf{X})$ satisfying

$$P_\theta\{\underline{\theta}(\mathbf{X}) \leq \theta\} \geq 1 - \alpha \qquad \text{for all } \theta \tag{2}$$

is called a *lower confidence bound* for θ at confidence level $1 - \alpha$. The quantity

$$\inf_{\theta \in \Theta} P_\theta\{\underline{\theta}(\mathbf{X}) \leq \theta\} \tag{3}$$

is called the *confidence coefficient*.

Definition 3. A function $\underline{\theta}$ that minimizes

$$P_\theta\{\underline{\theta}(\mathbf{X}) \leq \theta'\} \qquad \text{for all } \theta' < \theta \tag{4}$$

subject to (2) is known as a *uniformly most accurate* (UMA) *lower confidence bound* for θ at confidence level $1-\alpha$.

Remark 1. Suppose that $\mathbf{X} \sim P_\theta$ and (2) holds. Then the smallest probability of *true coverage*, $P_\theta\{\underline{\theta}(\mathbf{X}) \le \theta) = P_\theta\{[\underline{\theta}(\mathbf{X}), \infty) \ni \theta\}$ is $1-\alpha$. The probability of *false* (or *incorrect) coverage* is $P_\theta\{[\underline{\theta}(\mathbf{X}), \infty) \ni \theta'\} = P_\theta\{\underline{\theta}(\mathbf{X}) \le \theta'\}$ for $\theta' < \theta$. According to Definition 3, among the class of all lower confidence bounds satisfying (2), a UMA lower confidence bound has the smallest probability of false coverage.

Similar definitions are given for an upper confidence bound for θ and a UMA upper confidence bound.

Definition 4. A family of subsets $S(\mathbf{x})$ of $\Theta \subseteq \mathcal{R}_k$ is said to constitute a family of confidence sets at confidence level $1-\alpha$ if

$$(5) \qquad P_{\boldsymbol{\theta}}\{S(\mathbf{X}) \ni \boldsymbol{\theta}\} \ge 1-\alpha \qquad \text{for all } \boldsymbol{\theta} \in \Theta,$$

that is, the random set $S(\mathbf{X})$ covers the true parameter value $\boldsymbol{\theta}$ with probability $\ge 1-\alpha$. A lower confidence bound corresponds to the special case where $k=1$ and

$$(6) \qquad S(\mathbf{x}) = \{\theta : \underline{\theta}(\mathbf{x}) \le \theta < \infty\};$$

and an upper confidence bound to the case where

$$(7) \qquad S(\mathbf{x}) = \{\theta : \overline{\theta}(\mathbf{x}) \ge \theta > -\infty\}.$$

If $S(\mathbf{x})$ is of the form

$$(8) \qquad S(\mathbf{x}) = (\underline{\theta}(\mathbf{x}), \overline{\theta}(\mathbf{x}))$$

we will call it a confidence interval at confidence level $1-\alpha$, provided that

$$(9) \qquad P_\theta\{\underline{\theta}(\mathbf{X}) < \theta < \overline{\theta}(\mathbf{X})\} \ge 1-\alpha \qquad \text{for all } \theta,$$

and the quantity

$$(10) \qquad \inf_\theta P_\theta\{\underline{\theta}(\mathbf{X}) < \theta < \overline{\theta}(\mathbf{X})\}$$

will be referred to as the confidence coefficient associated with the random interval.

Remark 2. We write $S(\mathbf{X}) \ni \boldsymbol{\theta}$ to indicate that $\mathbf{X}$, and hence $S(\mathbf{X})$, is random here and not $\boldsymbol{\theta}$, so the probability distribution referred to is that of $\mathbf{X}$.

Remark 3. When $\mathbf{X} = \mathbf{x}$ is the realization, the confidence interval (set) $S(\mathbf{x})$ is a fixed subset of $\mathcal{R}_k$. No probability is attached to $S(\mathbf{x})$ itself since neither $\boldsymbol{\theta}$ nor $S(\mathbf{x})$ has a probability distribution. In fact, either $S(\mathbf{x})$ covers $\boldsymbol{\theta}$ or it does not, and we will never know which since $\boldsymbol{\theta}$ is unknown. One can give a relative frequency interpretation. If $(1-\alpha)$-level confidence sets for $\boldsymbol{\theta}$ were computed a large number of times, a fraction (approximately) $1-\alpha$ of these would contain the true (but unknown) parameter value.

Definition 5. A family of $(1-\alpha)$-level confidence sets $\{S(\mathbf{x})\}$ is said to be a UMA family of confidence sets at level $1-\alpha$ if

$$P_{\boldsymbol{\theta}}\{S(\mathbf{X}) \text{ contains } \boldsymbol{\theta}'\} \leq P_{\boldsymbol{\theta}}\{S'(\mathbf{X}) \text{ contains } \boldsymbol{\theta}'\}$$

for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$ and any $(1-\alpha)$-level family of confidence sets $S'(\mathbf{X})$.

Example 1. Let $X_1, X_2, \ldots, X_n$ be iid RVs, $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Consider the interval $(\overline{X} - c_1, \overline{X} + c_2)$. In order for this to be a $(1-\alpha)$-level confidence interval, we must have

$$P\{\overline{X} - c_1 < \mu < \overline{X} + c_2\} \geq 1 - \alpha,$$

which is the same as

$$P\{\mu - c_2 < \overline{X} < \mu + c_1\} \geq 1 - \alpha.$$

Thus

$$P\left\{-\frac{c_2}{\sigma}\sqrt{n} < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < \frac{c_1}{\sigma}\sqrt{n}\right\} \geq 1 - \alpha.$$

Since $\sqrt{n}(\overline{X} - \mu)/\sigma \sim \mathcal{N}(0, 1)$, we can choose c_1 and c_2 to have equality, namely,

$$P\left\{-\frac{c_2}{\sigma}\sqrt{n} < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < \frac{c_1}{\sigma}\sqrt{n}\right\} = 1 - \alpha,$$

provided that σ is known. There are infinitely many such pairs of values (c_1, c_2). In particular, an intuitively reasonable choice is $c_1 = -c_2 = c$, say. In that case

$$\frac{c\sqrt{n}}{\sigma} = z_{\alpha/2},$$

and the confidence interval is $(\overline{X} - (\sigma/\sqrt{n})z_{\alpha/2}, \overline{X} + (\sigma/\sqrt{n})z_{\alpha/2})$. The length of this interval is $(2\sigma/\sqrt{n})z_{\alpha/2}$. Given σ and α, we can choose n to get a confidence interval of a fixed length.

If σ is not known, we have from

$$P\{-c_2 < \overline{X} - \mu < c_1\} \geq 1 - \alpha$$

that

$$P\left\{-\frac{c_2}{S}\sqrt{n} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < \frac{c_1}{S}\sqrt{n}\right\} \geq 1-\alpha,$$

and once again we can choose pairs of values (c_1, c_2) using a t-distribution with $n-1$ d.f. such that

$$P\left\{-\frac{c_2\sqrt{n}}{S} < \frac{\overline{X}-\mu}{S}\sqrt{n} < \frac{c_1\sqrt{n}}{S}\right\} = 1-\alpha.$$

In particular, if we take $c_1 = -c_2 = c$, say, then

$$c\frac{\sqrt{n}}{S} = t_{n-1,\alpha/2},$$

and $(\overline{X}-(S/\sqrt{n})t_{n-1,\alpha/2}, \overline{X}+(S/\sqrt{n})t_{n-1,\alpha/2})$, is a $(1-\alpha)$-level confidence interval for μ. The length of this interval is $(2S/\sqrt{n})t_{n-1,\alpha/2}$, which is no longer constant. Therefore, we cannot choose n to get a fixed-width confidence interval of level $1-\alpha$. Indeed, the length of this interval can be quite large if σ is large. Its expected length is

$$\frac{2}{\sqrt{n}}t_{n-1,\alpha/2}E_\sigma S = \frac{2}{\sqrt{n}}t_{n-1,\alpha/2}\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}\sigma,$$

which can be made as small as we please by choosing n large enough.

Example 2. In Example 1, suppose that we wish to find a confidence interval for σ^2 instead when μ is unknown. Consider the interval (c_1S^2, c_2S^2), $c_1, c_2 > 0$. We have

$$P\{c_1S^2 < \sigma^2 < c_2S^2\} \geq 1-\alpha,$$

so that

$$P\left\{c_2^{-1} < \frac{S^2}{\sigma^2} < c_1^{-1}\right\} \geq 1-\alpha.$$

Since $(n-1)S^2/\sigma^2$ is $\chi^2(n-1)$, we can choose pairs of values (c_1, c_2) from the tables of the chi-square distribution. In particular, we can choose c_1, c_2 so that

$$P\left\{\frac{S^2}{\sigma^2} \geq \frac{1}{c_1}\right\} = \frac{\alpha}{2} = P\left\{\frac{S^2}{\sigma^2} \leq \frac{1}{c_2}\right\}.$$

Then

$$\frac{n-1}{c_1} = \chi^2_{n-1,\alpha/2} \quad \text{and} \quad \frac{n-1}{c_2} = \chi^2_{n-1,1-\alpha/2}.$$

Thus

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

is a $(1-\alpha)$-level confidence interval for σ^2 whenever μ is unknown. If μ is known, then

$$\sum_1^n \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n).$$

Thus we can base the confidence interval on $\sum_1^n (X_i-\mu)^2$. Proceeding similarly, we get a $(1-\alpha)$-level confidence interval as

$$\left(\frac{\sum_1^n (X_i-\mu)^2}{\chi^2_{n,\alpha/2}}, \qquad \frac{\sum_1^n (X_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}\right).$$

Next suppose that both μ and σ^2 are unknown and that we want a confidence set for (μ, σ^2). We have from Boole's inequality

$$\begin{aligned}
&P\left\{\overline{X} - \frac{S}{\sqrt{n}} t_{n-1,\alpha_1/2} < \mu < \overline{X} + \frac{S}{\sqrt{n}} t_{n-1,\alpha_1/2}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha_2/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha_2/2}}\right\} \\
&\quad \geq 1 - P\left\{\overline{X} + \frac{S}{\sqrt{n}} t_{n-1,\alpha_1/2} \leq \mu \text{ or } \overline{X} - \frac{S}{\sqrt{n}} t_{n-1,\alpha_1/2} \geq \mu\right\} \\
&\qquad - P\left\{\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha_2/2}} \leq \sigma^2 \text{ or } \frac{(n-1)S^2}{\chi^2_{n-1,\alpha_2/2}} \geq \sigma^2\right\} \\
&\quad = 1 - \alpha_1 - \alpha_2,
\end{aligned}$$

so that the Cartesian product,

$$S(\mathbf{X}) = \left(\overline{X} - \frac{S}{\sqrt{n}} t_{n-1,\alpha_1/2}, \overline{X} + \frac{S}{\sqrt{n}} t_{n-1,\alpha_1/2}\right) \times \left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha_2/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha_2/2}}\right)$$

is a $(1-\alpha_1-\alpha_2)$-level confidence set for (μ, σ^2).

11.3 METHODS OF FINDING CONFIDENCE INTERVALS

We now consider some common methods of constructing confidence sets. The most common of these is the method of pivots.

Definition 1. Let $\mathbf{X} \sim P_{\boldsymbol{\theta}}$. A random variable $T(\mathbf{X}, \boldsymbol{\theta})$ is known as a *pivot* if the distribution of $T(\mathbf{X}, \boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$.

In many problems, especially in location and scale problems, pivots are easily found. For example, in sampling from $f(x-\theta)$, $X_{(n)}-\theta$ is a pivot and so is $\overline{X}-\theta$. In sampling from $(1/\sigma)f(x/\sigma)$, a scale family, $X_{(n)}/\sigma$ is a pivot and so is $X_{(1)}/\sigma$, and in sampling from $(1/\sigma)f((x-\theta)/\sigma)$, a location-scale family, $(\overline{X}-\theta)/S$, is a pivot, and so is $(X_{(2)}+X_{(1)}-2\theta)/S$.

If the DF F_θ of X_i is continuous, then $F_\theta(X_i) \sim U[0,1]$ and, in case of random sampling, we can take

$$T(\mathbf{X},\theta) = \prod_{i=1}^{n} F_\theta(X_i),$$

or

$$-\log T(\mathbf{X},\theta) = -\sum_{i=1}^{n} \log F_\theta(X_i)$$

as a pivot. Since $F_\theta(X_i) \sim U[0,1]$, $-\log F_\theta(X_i) \sim G(1,1)$ and $-\sum_{i=1}^{n}\log F_\theta(X_i) \sim G(n,1)$. It follows that $-\sum_{i=1}^{n}\log F_\theta(X_i)$ is a pivot.

The following result gives a simple sufficient condition for a pivot to yield a confidence interval for a real-valued parameter θ.

Theorem 1. Let $T(\mathbf{X},\theta)$ be a pivot such that for each θ, $T(\mathbf{X},\theta)$ is a statistic, and as a function of θ, T is either strictly increasing or decreasing at each $\mathbf{x} \in \mathcal{R}_n$. Let $\Lambda \subseteq \mathcal{R}$ be the range of T, and for every $\lambda \in \Lambda$ and $\mathbf{x} \in \mathcal{R}_n$, let the equation $\lambda = T(\mathbf{x},\theta)$ be solvable. Then one can construct a confidence interval for θ at any level.

Proof. Let $0<\alpha<1$. Then we can choose a pair of numbers $\lambda_1(\alpha)$ and $\lambda_2(\alpha)$ in Λ not necessarily unique such that

$$(1) \qquad P_\theta\{\lambda_1(\alpha) < T(\mathbf{X},\theta) < \lambda_2(\alpha)\} \geq 1-\alpha \qquad \text{for all } \theta.$$

Since the distribution of T is independent of θ, it is clear that λ_1 and λ_2 are independent of θ. Since, moreover, T is monotone in θ, we can solve the equations

$$(2) \qquad T(\mathbf{x},\theta) = \lambda_1(\alpha) \quad \text{and} \quad T(\mathbf{x},\theta) = \lambda_2(\alpha)$$

for every $\mathbf{x}$ uniquely for θ. We have

$$(3) \qquad P_\theta\{\underline{\theta}(\mathbf{X}) < \theta < \overline{\theta}(\mathbf{X})\} \geq 1-\alpha \qquad \text{for all } \theta,$$

where $\underline{\theta}(\mathbf{X}) < \overline{\theta}(\mathbf{X})$ are RVs. This completes the proof.

Remark 1. The condition that $\lambda = T(\mathbf{x}, \theta)$ be solvable will be satisfied if, for example, T is continuous and strictly increasing or decreasing as a function of θ in Θ.

Note that in the continuous case (that is, when the DF of T is continuous) we can find a confidence interval with equality on the right side of (1). In the discrete case, however, this is usually not possible.

Remark 2. Relation (1) is valid even when the assumption of monotonicity of T in the theorem is dropped. In that case, inversion of the inequalities may yield a set of intervals (random set) $S(\mathbf{X})$ in Θ instead of a confidence interval.

Remark 3. The argument used in Theorem 1 can be extended to cover the multiparameter case, and the method will determine a confidence set for all the parameters of a distribution.

Example 1. Let $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, where σ is unknown and we seek a $(1-\alpha)$-level confidence interval for μ. Let us choose

$$T(\mathbf{X}, \mu) = \frac{\overline{X} - \mu}{S}\sqrt{n},$$

where $\overline{X}, S^2$ are the usual sample statistics. The RV $T(\mathbf{X}, \mu)$ has Student's t-distribution with $n-1$ d.f., which is independent of μ and $T(\mathbf{X}, \mu)$, as a function of μ is monotone. We can clearly choose $\lambda_1(\alpha), \lambda_2(\alpha)$ (not necessarily uniquely) so that

$$P\{\lambda_1(\alpha) < T(\mathbf{X}, \mu) < \lambda_2(\alpha)\} = 1 - \alpha \qquad \text{for all } \mu.$$

Solving

$$\lambda_1(\alpha) = \frac{\overline{X} - \mu}{S}\sqrt{n},$$

we get

$$\underline{\mu}(\mathbf{X}) = \overline{X} - \frac{S}{\sqrt{n}}\lambda_2(\alpha), \qquad \overline{\mu}(\mathbf{X}) = \overline{X} - \frac{S}{\sqrt{n}}\lambda_1(\alpha),$$

and a $(1-\alpha)$-level confidence interval is

$$\left(\overline{X} - \frac{S}{\sqrt{n}}\lambda_2(\alpha),\ \overline{X} - \frac{S}{\sqrt{n}}\lambda_1(\alpha)\right).$$

In practice, one chooses $\lambda_2(\alpha) = -\lambda_1(\alpha) = t_{n-1,\alpha/2}$.

Example 2. Let $X_1, X_2, \dots, X_n$ be iid with common PDF

$$f_\theta(x) = \exp\{-(x-\theta)\}, \qquad x > \theta \quad \text{and} \quad 0 \text{ elsewhere.}$$

Then the joint PDF of $\mathbf{X}$ is

$$f(\mathbf{x}; \theta) = \exp\left(-\sum_{i=1}^{n} x_i + n\theta\right) I_{[x_{(1)}>\theta]}.$$

Clearly, $T(\mathbf{X}, \theta) = X_{(1)} - \theta$ is a pivot. We can choose $\lambda_1(\alpha)$, $\lambda_2(\alpha)$ such that

$$P_\theta\left\{\lambda_1(\alpha) < X_{(1)} - \theta < \lambda_2(\alpha)\right\} = 1 - \alpha \qquad \text{for all } \theta$$

which yields $(X_{(1)} - \lambda_2(\alpha), X_{(1)} - \lambda_1(\alpha))$ as a $(1-\alpha)$-level confidence interval for θ.

Remark 4. In Example 1 we chose $\lambda_2 = -\lambda_1$, whereas in Example 2 we did not indicate how to choose the pair (λ_1, λ_2) from an infinite set of solutions to $P_\theta\{\lambda_1(\alpha) < T(\mathbf{X}, \theta) < \lambda_2(\alpha)\} = 1 - \alpha$. One choice is the *equal-tails* confidence interval, which is arrived at by assigning probability $\alpha/2$ to each tail of the distribution of T. This means that we solve

$$\frac{\alpha}{2} = P_\theta\{T(\mathbf{X}, \theta) < \lambda_1\} = P\{T(\mathbf{X}, \theta) > \lambda_2\}.$$

In Example 1, symmetry of the distribution leads to the choice indicated. In Example 2, $Y = X_{(1)} - \theta$ has PDF

$$g(y) = n\exp(-ny) \qquad \text{for } y > 0$$

so we choose (λ_1, λ_2) from

$$P_\theta\left\{X_{(1)} - \theta < \lambda_1\right\} = \frac{\alpha}{2} = P_\theta\left\{X_{(1)} - \theta > \lambda_2\right\},$$

giving $\lambda_2(\alpha) = (1/n)\ln(\alpha/2)$, and $\lambda_1(\alpha) = -(1/n)\ln(1-\alpha/2)$. Yet another method is to choose λ_1, λ_2 in such a way that the resulting confidence interval has smallest length. We discuss this method in Section 11.4.

We next consider the method of *test inversion* and explore the relationship between a test of hypothesis for a parameter θ and confidence interval for θ. Consider the following example.

Example 3. Let $X_1, X_2, \dots, X_n$ be a sample from $\mathcal{N}(\mu, \sigma_0^2)$ where σ_0 is known. In Example 11.2.1 we showed that

$$\left(\overline{X} - \frac{1}{\sqrt{n}} z_{\alpha/2}\sigma_0, \ \overline{X} + \frac{1}{\sqrt{n}} z_{\alpha/2}\sigma_0\right)$$

is a $(1-\alpha)$-level confidence interval for μ. If we define a test φ that rejects a value of $\mu = \mu_0$ if and only if μ_0 lies outside this interval; that is, if and only if

$$\frac{\sqrt{n}\,|\overline{X} - \mu_0|}{\sigma_0} \geq z_{\alpha/2},$$

then

$$P_{\mu_0}\left\{\sqrt{n}\frac{|\overline{X} - \mu_0|}{\sigma_0} \geq z_{\alpha/2}\right\} = \alpha,$$

and the test φ is a size α test of $\mu = \mu_0$ against the alternatives $\mu \neq \mu_0$.

Conversely, a family of α-level tests for the hypothesis $\mu = \mu_0$ generates a family of confidence intervals for μ by simply taking, as the confidence interval for μ_0, the set of those μ for which one cannot reject $\mu = \mu_0$.

Similarly, we can generate a family of α-level tests from a $(1-\alpha)$-level lower (or upper) confidence bound. Suppose that we start with the $(1-\alpha)$-level lower confidence bound $\overline{X} - z_\alpha(\sigma_0/\sqrt{n})$ for μ. Then, by defining a test $\varphi(\mathbf{X})$ that rejects $\mu \leq \mu_0$ if and only if $\mu_0 < \overline{X} - z_\alpha(\sigma_0/\sqrt{n})$, we get an α-level test for a hypothesis of the form $\mu \leq \mu_0$.

Example 3 is a special case of the duality principle proved in Theorem 2 below. In the following we restrict attention to the case in which the rejection (acceptance) region of the test is the indicator function of a (Borel-measurable) set, that is, we consider only nonrandomized tests (and confidence intervals). For notational convenience we write $H_0(\theta_0)$ for the hypothesis $H_0\colon \theta = \theta_0$ and $H_1(\theta_0)$ for the alternative hypothesis, which may be one- or two-sided.

Theorem 2. Let $A(\theta_0)$, $\theta_0 \in \Theta$, denote the region of acceptance of an α-level test of $H_0(\theta_0)$. For each observation $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, let $S(\mathbf{x})$ denote the set

(4) $$S(\mathbf{x}) = \{\theta\colon \mathbf{x} \in A(\theta), \theta \in \Theta\}.$$

Then $S(\mathbf{x})$ is a family of confidence sets for θ at confidence level $1-\alpha$. If, moreover, $A(\theta_0)$ is UMP for the problem $(\alpha, H_0(\theta_0), H_1(\theta_0))$, then $S(\mathbf{X})$ minimizes

(5) $$P_\theta\{S(\mathbf{X}) \ni \theta'\} \qquad \text{for all } \theta \in H_1(\theta')$$

among all $(1-\alpha)$-level families of confidence sets. That is, $S(\mathbf{X})$ is UMA.

Proof. We have

(6) $$S(\mathbf{x}) \ni \theta \qquad \text{if and only } \mathbf{x} \in A(\theta),$$

so that

$$P_\theta\{S(\mathbf{X}) \ni \theta\} = P_\theta\{\mathbf{X} \in A(\theta)\} \geq 1-\alpha,$$

as asserted.

If $S^*(\mathbf{X})$ is any other family of $(1-\alpha)$-level confidence sets, let $A^*(\theta) = \{\mathbf{x}\colon S^*(\mathbf{x}) \ni \theta\}$. Then

$$P_\theta\{\mathbf{X} \in A^*(\theta)\} = P_\theta\{S^*(\mathbf{X}) \ni \theta\} \geq 1-\alpha;$$

and since $A(\theta_0)$ is UMP for $(\alpha, H_0(\theta_0), H_1(\theta_0))$, it follows that

$$P_\theta\{\mathbf{X} \in A^*(\theta_0)\} \geq P_\theta\{\mathbf{X} \in A(\theta_0)\} \qquad \text{for any } \theta \in H_1(\theta_0).$$

Hence

$$P_\theta\{S^*(\mathbf{X}) \ni \theta_0\} \geq P_\theta\{\mathbf{X} \in A(\theta_0)\} = P_\theta\{S(\mathbf{X}) \ni \theta_0\}$$

for all $\theta \in H_1(\theta_0)$. This completes the proof.

Example 4. Let $\mathbf{X}$ be an RV of the continuous type with one-parameter exponential PDF given by

$$f_\theta(\mathbf{x}) = \exp[Q(\theta)T(\mathbf{x}) + S'(\mathbf{x}) + D(\theta)],$$

where $Q(\theta)$ is a nondecreasing function of θ. Let $H_0\colon \theta = \theta_0$ and $H_1\colon \theta < \theta_0$. Then the acceptance region of a UMP size α test of H_0 is of the form

$$A(\theta_0) = \{\mathbf{x}\colon T(\mathbf{x}) > c(\theta_0)\}.$$

Since for $\theta \geq \theta'$,

$$P_{\theta'}\{T(\mathbf{X}) \leq c(\theta')\} = \alpha = P_\theta\{T(\mathbf{X}) \leq c(\theta)\} \leq P_{\theta'}\{T(\mathbf{X}) \leq c(\theta)\},$$

$c(\theta)$ may be chosen to be nondecreasing. (The last inequality follows because the power of the UMP test is at least α, the size.) We have

$$S(\mathbf{x}) = \{\theta\colon \mathbf{x} \in A(\theta)\},$$

so that $S(\mathbf{x})$ is of the form $(-\infty, c^{-1}(T(\mathbf{x})))$ or $(-\infty, c^{-1}(T(\mathbf{x}))]$, where c^{-1} is defined by

$$c^{-1}(T(\mathbf{x})) = \sup_\theta\{\theta\colon c(\theta) \leq T(\mathbf{x})\}.$$

In particular, if $X_1, X_2, \ldots, X_n$ is a sample from

$$f_\theta(x) = \begin{cases} \dfrac{1}{\theta} e^{-x/\theta}, & x > 0, \\ 0, & \text{otherwise}, \end{cases}$$

then $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$; and for testing $H_0\colon \theta = \theta_0$ against $H_1\colon \theta < \theta_0$, the UMP acceptance region is of the form

$$A(\theta_0) = \left\{ \mathbf{x}\colon \sum_{i=1}^{n} x_i \ge c(\theta_0) \right\},$$

where $c(\theta_0)$ is the unique solution of

$$\int_{c(\theta_0)/\theta_0}^{\infty} \frac{y^{n-1}}{(n-1)!} e^{-y}\, dy = 1 - \alpha, \qquad 0 < \alpha < 1.$$

The UMA family of $(1-\alpha)$-level confidence sets is of the form

$$S(\mathbf{x}) = \{\theta\colon \mathbf{x} \in A(\theta)\}.$$

In the case $n = 1$,

$$c(\theta_0) = \theta_0 \log\left(\frac{1}{1-\alpha}\right) \quad \text{and} \quad S(x) = \left[0, \frac{x}{-\log(1-\alpha)}\right].$$

Example 5. Let $X_1, X_2, \ldots, X_n$ be iid $U(0, \theta)$ RVs. In Problem 9.4.3 we asked the reader to show that the test

$$\phi(x) = \begin{cases} 1, & x_{(n)} > \theta_0 \quad \text{or} \quad x_{(n)} < \theta_0 \alpha^{1/n}, \\ 0, & \text{otherwise} \end{cases}$$

is UMP size α test of $\theta = \theta_0$ against $\theta \neq \theta_0$. Then

$$A(\theta_0) = \{\mathbf{x}\colon \theta_0 \alpha^{1/n} \le x_{(n)} \le \theta_0\}$$

and it follows that $[x_{(n)}, x_{(n)}\alpha^{-1/n}]$ is a $(1-\alpha)$-level UMA confidence interval for θ.

The third method we consider is based on Bayesian analysis, where we take into account any prior knowledge that the experimenter has about θ. This is reflected in the specification of the prior distribution $\pi(\theta)$ on Θ. Under this setup the claims of probability of coverage are based not on the distribution of $\mathbf{X}$ but on the conditional distribution of θ given $\mathbf{X} = \mathbf{x}$, the posterior distribution of θ.

Let Θ be the parameter set, and let the observable RV $\mathbf{X}$ have PDF (PMF) $f_\theta(\mathbf{x})$. Suppose that we consider θ as an RV with distribution $\pi(\theta)$ on Θ. Then $f_\theta(\mathbf{x})$ can be considered as the conditional PDF (PMF) of $\mathbf{X}$, given that the RV θ takes the value θ. Note that we are using the same symbol for the RV θ and the value that it assumes.

We can determine the joint distribution of $\mathbf{X}$ and θ, the marginal distribution of $\mathbf{X}$, and also the conditional distribution of θ, given $\mathbf{X} = \mathbf{x}$ as usual. Thus the joint distribution is given by

$$f(\mathbf{x}, \theta) = \pi(\theta) f_\theta(\mathbf{x}), \tag{7}$$

and the marginal distribution of $\mathbf{X}$ by

$$g(\mathbf{x}) = \begin{cases} \sum \pi(\theta) f_\theta(\mathbf{x}) & \text{if } \pi \text{ is a PMF,} \\ \int \pi(\theta) f_\theta(\mathbf{x})\, d\theta & \text{if } \pi \text{ is a PDF.} \end{cases} \tag{8}$$

The conditional distribution of θ, given that $\mathbf{x}$ is observed, is given by

$$h(\theta \mid \mathbf{x}) = \frac{\pi(\theta) f_\theta(\mathbf{x})}{g(\mathbf{x})}, \qquad g(\mathbf{x}) > 0. \tag{9}$$

Given $h(\theta \mid \mathbf{x})$, it is easy to find functions $l(\mathbf{x})$, $u(\mathbf{x})$ such that

$$P\{l(\mathbf{X}) < \theta < u(\mathbf{X})\} \geq 1 - \alpha,$$

where

$$P\{l(\mathbf{X}) < \theta < u(\mathbf{X}) \mid \mathbf{X} = \mathbf{x}\} = \begin{cases} \int_l^u h(\theta \mid \mathbf{x})\, d\theta, \\ \sum_l^u h(\theta \mid \mathbf{x}), \end{cases} \tag{10}$$

depending on whether h is a PDF or a PMF.

Definition 2. An interval $(l(\mathbf{x}), u(\mathbf{x}))$ that has probability at least $1-\alpha$ of including θ is called a $(1-\alpha)$-*level Bayes interval* for θ. Also, $l(\mathbf{x})$ and $u(\mathbf{x})$ are called the *lower and upper limits of the interval.*

One can similarly define one-sided Bayes intervals or $(1-\alpha)$-level lower and upper Bayes limits.

Remark 5. We note that under the Bayesian setup, we can speak of the probability that θ lies in the interval $(l(\mathbf{x}), u(\mathbf{x}))$ with probability $1-\alpha$ because l and u are computed based on the posterior distribution of θ given $\mathbf{x}$. To emphasize this distinction between Bayesian and classical analysis, some authors prefer the term *credible sets* for Bayesian confidence sets.

Example 6. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, 1)$, $\mu \in \mathcal{R}$, and let the a priori distribution of μ be $\mathcal{N}(0, 1)$. Then from Example 8.8.6 we know that $h(\mu \mid \mathbf{x})$ is

$$\mathcal{N}\left(\frac{\sum_1^n x_i}{n+1}, \frac{1}{n+1}\right).$$

Thus a $(1-\alpha)$-level Bayesian confidence interval is

$$\left(\frac{n\overline{x}}{n+1}-\frac{z_{\alpha/2}}{\sqrt{n+1}}, \frac{n\overline{x}}{n+1}+\frac{z_{\alpha/2}}{\sqrt{n+1}}\right).$$

A $(1-\alpha)$-level confidence interval for μ (treating μ as fixed) is a random interval with value

$$\left(\overline{x}-\frac{z_{\alpha/2}}{\sqrt{n}}, \overline{x}+\frac{z_{\alpha/2}}{\sqrt{n}}\right).$$

Thus the Bayesian interval is somewhat shorter in length. This is to be expected since we assumed more in the Bayesian case.

Example 7. Let $X_1, X_2, \ldots, X_n$ be iid $b(1, p)$ RVs, and let the prior distribution on $\Theta = (0, 1)$ be $U(0, 1)$. A simple computation shows that the posterior PDF of p, given x, is

$$h(p|x) = \begin{cases} \dfrac{p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}}{B\left(\sum_1^n x_i + 1, n - \sum_1^n x_i + 1\right)}, & 0 < p < 1 \\ 0, & \text{otherwise}, \end{cases}$$

Given a table of incomplete beta integrals and the observed value of $\sum_1^n x_i$, one can easily construct a Bayesian confidence interval for p.

Finally, we consider some large-sample methods of constructing confidence intervals. Suppose that $T(\mathbf{X}) \sim AN(\theta, v(\theta)/n)$. Then

$$\sqrt{n}\frac{T(X)-\theta}{\sqrt{v(\theta)}} \xrightarrow{L} Z,$$

where $Z \sim N(0, 1)$. Suppose further that there is a statistic $S(\mathbf{X})$ such that $S(\mathbf{X}) \xrightarrow{P} v(\theta)$. Then, by Slutsky's theorem,

$$\sqrt{n}\frac{T(\mathbf{X})-\theta}{\sqrt{S(\mathbf{X})}} \xrightarrow{L} Z$$

and we can obtain an (approximate) $(1-\alpha)$-level confidence interval for θ by inverting the inequality

$$\left|\sqrt{n}\frac{T(\mathbf{X})-\theta}{\sqrt{S(\mathbf{X})}}\right| \le z_{\alpha/2}.$$

Example 8. Let $X_1, X_2, \ldots, X_n$ be iid RVs with finite variance. Also, let $EX_i = \mu$ and $EX_i^2 = \sigma^2 + \mu^2$. From the CLT it follows that

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{L} Z,$$

where $Z \sim \mathcal{N}(0, 1)$. Suppose that we want a $(1-\alpha)$-level confidence interval for μ when σ is not known. Since $S \xrightarrow{P} \sigma$, for large n the quantity $[\sqrt{n}(\overline{X}-\mu)/S]$ is approximately normally distributed with mean 0 and variance 1. Hence, for large n, we can find constants c_1, c_2 such that

$$P\left\{c_1 < \frac{\overline{X}-\mu}{S}\sqrt{n} < c_2\right\} = 1-\alpha.$$

In particular, we can choose $-c_1 = c_2 = z_{\alpha/2}$ to give

$$\left(\overline{x} - \frac{s}{\sqrt{n}} z_{\alpha/2},\ \overline{x} + \frac{s}{\sqrt{n}} z_{\alpha/2}\right)$$

as an approximate $(1-\alpha)$-level confidence interval for μ.

Recall that if $\hat{\theta}$ is the MLE of θ and the conditions of Theorem 8.7.4 or 8.7.5 are satisfied (*caution:* see Remark 8.7.4), then

$$\frac{\sqrt{n}\,(\hat{\theta}-\theta)}{\sigma} \xrightarrow{L} \mathcal{N}(0, 1) \qquad \text{as } n \to \infty,$$

where

$$\sigma^2 = \left[E_\theta\left\{\frac{\partial \log f_\theta(X)}{\partial \theta}\right\}^2\right]^{-1} = \frac{1}{I(\theta)}.$$

Then we can invert the statement

$$P_\theta\left\{-z_{\alpha/2} < \frac{\hat{\theta}-\theta}{\sigma}\sqrt{n} < z_{\alpha/2}\right\} \geq 1-\alpha$$

to give an approximate $(1-\alpha)$-level confidence interval for θ.

Yet another possible procedure has universal applicability and hence can be used for large or small samples. Unfortunately, however, this procedure usually yields confidence intervals that are much too large in length. The method employs the well-known Chebychev inequality (see Section 3.4):

$$P\left\{|X - EX| < \varepsilon\sqrt{\operatorname{var}(X)}\right\} > 1 - \frac{1}{\varepsilon^2}.$$

If $\hat{\theta}$ is an estimate of θ (not necessarily unbiased) with finite variance $\sigma^2(\theta)$, then by Chebychev's inequality

$$P\left\{|\hat{\theta}-\theta| < \varepsilon\sqrt{E(\hat{\theta}-\theta)^2}\right\} > 1-\frac{1}{\varepsilon^2}.$$

It follows that

$$\left(\hat{\theta}-\varepsilon\sqrt{E(\hat{\theta}-\theta)^2}\,,\hat{\theta}+\varepsilon\sqrt{E(\hat{\theta}-\theta)^2}\right)$$

is a $[1-(1/\varepsilon^2)]$-level confidence interval for θ. Under some mild consistency conditions one can replace the normalizing constant $\sqrt{[E(\hat{\theta}-\theta)^2]}$, which will be some function $\lambda(\theta)$ of θ, by $\lambda(\hat{\theta})$.

Note that the estimator $\hat{\theta}$ need not have a limiting normal law.

Example 9. Let $X_1, X_2, \ldots, X_n$ be iid $b(1, p)$ RVs and it is required to find a confidence interval for p. We know that $E\overline{X} = p$, and

$$\operatorname{var}(\overline{X}) = \frac{\operatorname{var}(X)}{n} = \frac{p(1-p)}{n}.$$

It follows that

$$P\left\{|\overline{X}-p| < \varepsilon\sqrt{\frac{p(1-p)}{n}}\right\} > 1-\frac{1}{\varepsilon^2}.$$

Since $p(1-p) \le \frac{1}{4}$, we have

$$P\left\{\overline{X}-\frac{1}{2\sqrt{n}}\varepsilon < p < \overline{X}+\frac{1}{2\sqrt{n}}\varepsilon\right\} > 1-\frac{1}{\varepsilon^2}.$$

One can now choose ε and n or, if n is kept constant at a given number, ε to get the desired level.

Actually, the confidence interval obtained above can be improved somewhat. We note that

$$P\left\{|\overline{X}-p| < \varepsilon\sqrt{\frac{p(1-p)}{n}}\right\} > 1-\frac{1}{\varepsilon^2},$$

so that

$$P\left\{|\overline{X}-p|^2 < \frac{\varepsilon^2 p(1-p)}{n}\right\} > 1-\frac{1}{\varepsilon^2}.$$

Now

$$|\overline{X}-p|^2 < \frac{\varepsilon^2}{n}p(1-p)$$

if and only if

$$\left(1+\frac{\varepsilon^2}{n}\right)p^2-\left(2\overline{X}+\frac{\varepsilon^2}{n}\right)p+\overline{X}^2<0.$$

This last inequality holds if and only if p lies between the two roots of the quadratic equation

$$\left(1+\frac{\varepsilon^2}{n}\right)p^2-\left(2\overline{X}+\frac{\varepsilon^2}{n}\right)p+\overline{X}^2=0.$$

The two roots are

$$\begin{aligned}p_1&=\frac{2\overline{X}+(\varepsilon^2/n)-\sqrt{[2\overline{X}+(\varepsilon^2/n)]^2-4[1+(\varepsilon^2/n)]\overline{X}^2}}{2[1+(\varepsilon^2/n)]}\\&=\frac{\overline{X}}{1+(\varepsilon^2/n)}+\frac{(\varepsilon^2/n)-\sqrt{4(\varepsilon^2/n)\overline{X}(1-\overline{X})+(\varepsilon^4/n^2)}}{2[1+(\varepsilon^2/n)]}\end{aligned}$$

and

$$\begin{aligned}p_2&=\frac{2\overline{X}+(\varepsilon^2/n)+\sqrt{[2\overline{X}+(\varepsilon^2/n)]^2-4[1+(\varepsilon^2/n)]\overline{X}^2}}{2[1+(\varepsilon^2/n)]}\\&=\frac{\overline{X}}{1+(\varepsilon^2/n)}+\frac{(\varepsilon^2/n)+\sqrt{4(\varepsilon^2/n)\overline{X}(1-\overline{X})+(\varepsilon^4/n^2)}}{2[1+(\varepsilon^2/n)]}\end{aligned}$$

It follows that

$$P\{p_1<p<p_2\}>1-\frac{1}{\varepsilon^2}.$$

Note that when n is large,

$$p_1\approx\overline{X}-\varepsilon\sqrt{\frac{\overline{X}(1-\overline{X})}{n}},\qquad p_2\approx\overline{X}+\varepsilon\sqrt{\frac{\overline{X}(1-\overline{X})}{n}},$$

as one should expect in view of the fact that $\overline{X}\to p$ with probability 1 and $\sqrt{[\overline{X}(1-\overline{X})/n]}$ estimates $\sqrt{[p(1-p)/n]}$. Alternatively, we could have used the CLT (or large-sample property of the MLE) to arrive at the same result but with ε replaced by $z_{\alpha/2}$.

Example 10. Let $X_1, X_2, \dots, X_n$ be a sample from $U(0,\theta)$. We seek a confidence interval for the parameter θ. The estimator $\hat{\theta}=X_{(n)}$ is the MLE of θ, which is also sufficient for θ. From Example 5, $[X_{(n)}, \alpha^{-1/n}X_{(n)}]$ is a $(1-\alpha)$-level UMA confidence interval for θ.

Let us now apply the method of Chebychev's inequality to the same problem. We have

$$E_\theta X_{(n)} = \frac{n}{n+1}\theta$$

and

$$E_\theta (X_{(n)} - \theta)^2 = \theta^2 \frac{2}{(n+1)(n+2)}.$$

Thus

$$P\left\{\frac{|X_{(n)} - \theta|}{\theta}\sqrt{\frac{(n+1)(n+2)}{2}} < \varepsilon\right\} > 1 - \frac{1}{\varepsilon^2}.$$

Since $X_{(n)} \xrightarrow{P} \theta$, we replace θ by $X_{(n)}$ in the denominator, and for moderately large n,

$$P\left\{\frac{|X_{(n)} - \theta|}{X_{(n)}}\sqrt{\frac{(n+1)(n+2)}{2}} < \varepsilon\right\} > 1 - \frac{1}{\varepsilon^2}.$$

It follows that

$$\left(X_{(n)} - \varepsilon X_{(n)}\frac{\sqrt{2}}{\sqrt{(n+1)(n+2)}},\ X_{(n)} + \varepsilon X_{(n)}\frac{\sqrt{2}}{\sqrt{(n+1)(n+2)}}\right)$$

is a $1 - (1/\varepsilon^2)$ confidence interval for θ. Choosing $1 - (1/\varepsilon^2) = 1 - \alpha$, or $\varepsilon = 1/\sqrt{\alpha}$, and noting that $1/\sqrt{[(n+1)(n+2)]} \approx 1/n$ for large n, and the fact that with probability 1, $X_{(n)} \le \theta$, we can use the approximate confidence interval

$$\left(X_{(n)},\ X_{(n)}\left(1 + \frac{1}{n}\sqrt{\frac{2}{\alpha}}\right)\right)$$

for θ.

In the examples given above we see that for a given confidence interval $1 - \alpha$, a wide choice of confidence intervals is available. Clearly, the larger the interval, the better the chance of trapping a true parameter value. Thus the interval $(-\infty, +\infty)$, which ignores the data completely, will include the real-valued parameter θ with confidence level 1. However, the larger the confidence interval, the less meaningful it is. Therefore, for a given confidence level $1-\alpha$, it is desirable to choose the shortest possible confidence interval. Since the length $\overline{\theta} - \underline{\theta}$, in general, is a random variable, one can show that a confidence interval of level $1-\alpha$ with uniformly minimum length among all such intervals does not exist in most cases. The alternative, to minimize $E_\theta(\overline{\theta} - \underline{\theta})$, is also quite unsatisfactory. In the next section we consider the problem of finding shortest-length confidence interval based on a suitable statistic.

PROBLEMS 11.3

1. A sample of size 25 from a normal population with variance 81 produced a mean of 81.2. Find a 0.95 level confidence interval for the mean μ.

2. Let $\overline{X}$ be the mean of a random sample of size n from $\mathcal{N}(\mu, 16)$. Find the smallest sample size n such that $(\overline{X}-1, \overline{X}+1)$ is a 0.90 level confidence interval for μ.

3. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$, respectively. Find a confidence interval for $\mu_1 - \mu_2$ at confidence level $1 - \alpha$ when (a) σ is known, and (b) σ is unknown.

4. Two independent samples, each of size 7, from normal populations with common unknown variance σ^2 produced sample means 4.8 and 5.4 and sample variances 8.38 and 7.62, respectively. Find a 0.95-level confidence interval for $\mu_1 - \mu_2$, the difference between the means of samples 1 and 2.

5. In Problem 3, suppose that the first population has variance σ_1^2 and the second population has variance σ_2^2, where both σ_1^2, and σ_2^2 are known. Find a $(1 - \alpha)$-level confidence interval for $\mu_1 - \mu_2$. What happens if both σ_1^2 and σ_2^2 are unknown and unequal?

6. In Problem 5, find a confidence interval for the ratio σ_2^2/σ_1^2, both when μ_1, μ_2 are known and when μ_1, μ_2 are unknown. What happens if either μ_1 or μ_2 is unknown but the other is known?

7. Let $X_1, X_2, \ldots, X_n$ be a sample from a $G(1, \beta)$ distribution. Find a confidence interval for the parameter β with confidence level $1 - \alpha$.

8. (a) Use the large-sample properties of the MLE to construct a $(1 - \alpha)$-level confidence interval for the parameter θ in each of the following cases: (i) $X_1, X_2, \ldots, X_n$ is a sample from $G(1, 1/\theta)$, and (ii) $X_1, X_2, \ldots, X_n$ is a sample from $P(\theta)$.

   (b) In part (a), use Chebychev's inequality to do the same.

9. For a sample of size 1 from the population

$$f_\theta(x) = \frac{2}{\theta^2}(\theta - x), \qquad 0 < x < \theta,$$

   find a $(1 - \alpha)$-level confidence interval for θ.

10. Let $X_1, X_2, \ldots, X_n$ be a sample from the uniform distribution on N points. Find an upper $(1-\alpha)$-level confidence bound for N, based on $\max(X_1, X_2, \ldots, X_n)$.

11. In Example 10, find the smallest n such that the length of the $(1-\alpha)$-level confidence interval $(X_{(n)}, \alpha^{-1/n} X_{(n)}) < d$, provided it is known that $\theta \le a$, where a is a known constant.

12. Let X and Y be independent RVs with PDFs $\lambda e^{-\lambda x}$ $(x > 0)$ and $\mu e^{-\mu y}$ $(y > 0)$, respectively. Find a $(1-\alpha)$-level confidence region for (λ, μ) of the form $\{(\lambda, \mu)\colon \lambda X + \mu Y \le k\}$.

13. Let $X_1, X_2, \dots, X_n$ be a sample from $\mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Find a UMA $(1-\alpha)$-level upper confidence bound for μ.

14. Let $X_1, X_2, \dots, X_n$ be a sample from a Poisson distribution with unknown parameter λ. Assuming that λ is a value assumed by a $G(\alpha, \beta)$ RV, find a Bayesian confidence interval for λ.

15. Let $X_1, X_2, \dots, X_n$ be a sample from a geometric distribution with parameter θ. Assuming that θ has a priori PDF that is given by the density of a $B(\alpha, \beta)$ RV, find a Bayesian confidence interval for θ.

16. Let $X_1, X_2, \dots, X_n$ be a sample from $\mathcal{N}(\mu, 1)$, and suppose that the a priori PDF for μ is $U(-1, 1)$. Find a Bayesian confidence interval for μ.

11.4 SHORTEST-LENGTH CONFIDENCE INTERVALS

We have already remarked that we can increase the confidence level simply by taking a longer-length confidence interval. Indeed, the worthless interval $-\infty < \theta < \infty$, which simply says that θ is a point on the real line, has confidence level 1. In practice, one would like to set the level at a given fixed number $1-\alpha$ $(0 < \alpha < 1)$ and, if possible, construct an interval as short as possible among all confidence intervals with the same level. Such an interval is desirable since it is more informative. We have already remarked that shortest-length confidence intervals do not always exist. In this section we investigate the possibility of constructing shortest-length confidence intervals based on simple RVs. The discussion here is based on Guenther [34]. Theorem 11.3.1 is really the key to the following discussion.

Let $X_1, X_2, \dots, X_n$ be a sample from a PDF $f_\theta(x)$, and $T(X_1, X_2, \dots, X_n, \theta) = T_\theta$ be a pivot for θ. Also, let $\lambda_1 = \lambda_1(\alpha)$, $\lambda_2 = \lambda_2(\alpha)$ be chosen so that

$$P\{\lambda_1 < T_\theta < \lambda_2\} = 1 - \alpha, \tag{1}$$

and suppose that (1) can be rewritten as

$$P\{\underline{\theta}(\mathbf{X}) < \theta < \overline{\theta}(\mathbf{X})\} = 1 - \alpha. \tag{2}$$

For every T_θ, λ_1 and λ_2 can be chosen in many ways. We would like to choose λ_1 and λ_2 so that $\overline{\theta} - \underline{\theta}$ is minimum. Such an interval is a $(1-\alpha)$-level shortest-length confidence interval based on T_θ. It may be possible, however, to find another RV T_θ^* that may yield an even shorter interval. Therefore, we are not asserting that the procedure, if it succeeds, will lead to a $(1-\alpha)$-level confidence interval that has shortest length among all intervals of this level. For T_θ we use the simplest RV that is a function of a sufficient statistic and θ.

Remark 1. An alternative to minimizing the length of the confidence interval is to minimize the expected length $E_\theta\{\overline{\theta}(\mathbf{X}) - \underline{\theta}(\mathbf{X})\}$. Unfortunately, this also is quite unsatisfactory since, in general, there does not exist a member of the class of all $(1-\alpha)$-level confidence intervals that minimizes $E_\theta\{\overline{\theta}(\mathbf{X}) - \underline{\theta}(\mathbf{X})\}$ for all θ. The procedures applied in finding the shortest-length confidence interval based on a pivot are also applicable in finding an interval that minimizes the expected length. We remark here that the restriction to unbiased confidence intervals is natural if we wish to minimize $E_\theta[\overline{\theta}(\mathbf{X}) - \underline{\theta}(\mathbf{x})]$. See Section 11.5 for definitions and further details.

Example 1. Let $X_1, X_2, \ldots, X_n$ be sample from $\mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Then $\overline{X}$ is sufficient for μ and take

$$T_\mu(\mathbf{X}) = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}.$$

Then

$$1-\alpha = P\left\{a < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < b\right\} = P\left\{\overline{X} - b\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} - a\frac{\sigma}{\sqrt{n}}\right\}.$$

The length of this confidence interval is $(\sigma/\sqrt{n})(b-a)$. We wish to minimize $L = (\sigma/\sqrt{n})(b-a)$ such that

$$\Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,dx = \int_a^b \varphi(x)\,dx = 1-\alpha.$$

Here φ and Φ, respectively, are the PDF and DF of an $\mathcal{N}(0,1)$ RV. Thus

$$\frac{dL}{da} = \frac{\sigma}{\sqrt{n}}\left(\frac{db}{da} - 1\right)$$

and

$$\varphi(b)\frac{db}{da} - \varphi(a) = 0,$$

giving

$$\frac{dL}{da} = \frac{\sigma}{\sqrt{n}}\left[\frac{\varphi(a)}{\varphi(b)} - 1\right].$$

The minimum occurs when $\varphi(a) = \varphi(b)$, that is, when $a = b$ or $a = -b$. Since $a = b$ does not satisfy

$$\int_a^b \varphi(t)\,dt = 1 - \alpha,$$

we choose $a = -b$. The shortest confidence interval based on T_μ is therefore the equal-tails interval,

$$\left(\overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \quad \text{or} \quad \left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

The length of this interval is $2z_{\alpha/2}(\sigma/\sqrt{n})$. In this case we can plan our experiment to give a prescribed confidence level and a prescribed length for the interval. To have level $1 - \alpha$ and length $\le 2d$, we choose the smallest n such that

$$d \ge z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad n \ge z_{\alpha/2}^2\frac{\sigma^2}{d^2}.$$

This can also be interpreted as follows. If we estimate μ by $\overline{X}$, taking a sample of size $n \ge z_{\alpha/2}^2(\sigma^2/d^2)$, we are $100(1 - \alpha)$ percent confident that the error in our estimate is at most d.

Example 2. In Example 1, suppose that σ is unknown. In that case we use

$$T_\mu(\mathbf{X}) = \frac{\overline{X} - \mu}{S}\sqrt{n}$$

as a pivot. T_μ has Student's t-distribution with $n - 1$ d.f. Thus

$$1 - \alpha = P\left\{a < \frac{\overline{X} - \mu}{S}\sqrt{n} < b\right\} = P\left\{\overline{X} - b\frac{S}{\sqrt{n}} < \mu < \overline{X} - a\frac{S}{\sqrt{n}}\right\}.$$

We wish to minimize

$$L = (b - a)\frac{S}{\sqrt{n}}$$

subject to

$$\int_a^b f_{n-1}(t)\,dt = 1 - \alpha,$$

where $f_{n-1}(t)$ is the PDF of T_μ. We have

$$\frac{dL}{da} = \left(\frac{db}{da} - 1\right)\frac{S}{\sqrt{n}} \qquad \text{and} \qquad f_{n-1}(b)\frac{db}{da} - f_{n-1}(a) = 0,$$

giving

$$\frac{dL}{da} = \left[\frac{f_{n-1}(a)}{f_{n-1}(b)} - 1\right]\frac{S}{\sqrt{n}}.$$

It follows that the minimum occurs at $a = -b$ (the other solution, $a = b$, is not admissible). The shortest-length confidence interval based on T_μ is the equal-tails interval,

$$\left(\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}},\ \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right).$$

The length of this interval is $2t_{n-1,\alpha/2}(S/\sqrt{n})$, which, being random, may be arbitrarily large. Note that the same confidence interval minimizes the expected length of the interval, namely, $EL = (b-a)c_n(\sigma/\sqrt{n})$, where c_n is a constant determined from $ES = c_n\sigma$ and the minimum expected length is $2t_{n-1,\alpha/2}c_n(\sigma/\sqrt{n})$.

Example 3. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ RVs. Suppose that μ is known and we want a confidence interval for σ^2. The obvious choice for a pivot T_{σ^2} is given by

$$T_{\sigma^2}(\mathbf{x}) = \frac{\sum_1^n (X_i - \mu)^2}{\sigma^2},$$

which has a chi-square distribution with n d.f. Now

$$P\left\{a < \frac{\sum_1^n (X_i - \mu)^2}{\sigma^2} < b\right\} = 1 - \alpha,$$

so that

$$P\left\{\frac{\sum_1^n (X_i - \mu)^2}{b} < \sigma^2 < \frac{\sum_1^n (X_i - \mu^2)}{a}\right\} = 1 - \alpha.$$

We wish to minimize

$$L = \left(\frac{1}{a} - \frac{1}{b}\right)\sum_1^n (X_i - \mu)^2$$

subject to

$$\int_a^b f_n(t)\,dt = 1 - \alpha,$$

where f_n is the PDF of a chi-square RV with n d.f. We have

$$\frac{dL}{da} = \left(\frac{1}{a^2} - \frac{1}{b^2}\frac{db}{da}\right)\sum_1^n (X_i - \mu)^2$$

and

$$\frac{db}{da} = \frac{f_n(a)}{f_n(b)},$$

so that

$$\frac{dL}{da} = \left[\frac{1}{a^2} - \frac{1}{b^2}\frac{f_n(a)}{f_n(b)}\right]\sum_1^n (X_i - \mu)^2,$$

which vanishes if

$$\frac{1}{a^2} = \frac{1}{b^2}\frac{f_n(a)}{f_n(b)}.$$

Numerical results giving values of a and b to four significant places of decimals are available (see Tate and Klett [111]). In practice, the simpler equal-tails interval,

$$\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n,\alpha/2}}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n,1-\alpha/2}}\right),$$

may be used.

If μ is unknown, we use

$$T_{\sigma^2}(\mathbf{X}) = \frac{\sum_1^n (X_i - \overline{X})^2}{\sigma^2} = (n-1)\frac{S^2}{\sigma^2}$$

as a pivot. T_{σ^2} has a $\chi^2(n-1)$ distribution. Proceeding as above, we can show that the shortest-length confidence interval based on T_{σ^2} is $((n-1)(S^2/b), (n-1)(S^2/a))$; here a and b are a solution of

$$P\{a < \chi^2(n-1) < b\} = 1 - \alpha$$

and

$$a^2 f_{n-1}(a) = b^2 f_{n-1}(b),$$

where f_{n-1} is the PDF of a $\chi^2(n-1)$ RV. Numerical solutions due to Tate and Klett [111] may be used, but in practice, the simpler equal-tails confidence interval,

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

is employed.

Example 4. Let $X_1, X_2, \ldots, X_n$ be a sample from $U(0, \theta)$. Then $X_{(n)}$ is sufficient for θ with density

$$f_n(y) = n\frac{y^{n-1}}{\theta^n}, \qquad 0 < y < \theta.$$

The RV $T_\theta = X_{(n)}/\theta$ has PDF

$$h(t) = nt^{n-1}, \qquad 0 < t < 1.$$

Using T_θ as pivot, we see that the confidence interval is $(X_{(n)}/b, X_{(n)}/a)$ with length $L = X_{(n)}(1/a - 1/b)$. We minimize L subject to

$$\int_a^b nt^{n-1}\,dt = b^n - a^n = 1 - \alpha.$$

Now

$$(1-\alpha)^{1/n} < b \leq 1$$

and

$$\frac{dL}{db} = X_{(n)}\left(-\frac{1}{a^2}\frac{da}{db} + \frac{1}{b^2}\right) = X_{(n)}\left(\frac{a^{n+1} - b^{n+1}}{b^2 a^{n+1}}\right) < 0,$$

so that the minimum occurs at $b = 1$. The shortest interval is therefore $(X_{(n)}, X_{(n)}/\alpha^{1/n})$. Note that

$$EL = \left(\frac{1}{a} - \frac{1}{b}\right) EX_{(n)} = \frac{n\theta}{n+1}\left(\frac{1}{a} - \frac{1}{b}\right),$$

which is minimized subject to

$$b^n - a^n = 1 - \alpha,$$

where $b = 1$ and $a = \alpha^{1/n}$. The expected length of the interval that minimizes EL is $[(1/\alpha^{1/n}) - 1][n\theta/(n+1)]$, which is also the expected length of the shortest confidence interval based on $X_{(n)}$. Note that the length of the interval $(X_{(n)}, \alpha^{-1/n}X_{(n)})$ goes to 0 as $n \to \infty$.

For some results on asymptotically shortest-length confidence intervals, we refer the reader to Wilks [117, pp. 374–376].

PROBLEMS 11.4

1. Let $X_1, X_2, \dots, X_n$ be a sample from

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)} & \text{if } x > \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Find the shortest-length confidence interval for θ at level $1-\alpha$, based on a sufficient statistic for θ.

2. Let $X_1, X_2, \dots, X_n$ be a sample from $G(1, \theta)$. Find the shortest-length confidence interval for θ at level $1-\alpha$, based on a sufficient statistic for θ.

3. In Problem 11.3.9, how will you find the shortest-length confidence interval for θ at level $1-\alpha$ based on the statistic X/θ?

4. Let $T(\mathbf{X}, \theta)$ be a pivot of the form $T(\mathbf{X}, \theta) = T_1(\mathbf{X}) - \theta$. Show how one can construct a confidence interval for θ with fixed width d and maximum possible confidence coefficient. In particular, construct a confidence interval that has fixed width d and maximum possible confidence coefficient for the mean μ of a normal population with variance 1. Find the smallest size n for which this confidence interval has a confidence coefficient $\geq 1-\alpha$. Repeat the above in sampling from an exponential PDF

$$f_\mu(x) = e^{\mu - x} \quad \text{for } x > \mu \quad \text{and} \quad f_\mu(x) = 0 \quad \text{for } x \leq \mu.$$

(Desu [20])

5. Let $X_1, X_2, \dots, X_n$ be a random sample from

$$f_\theta(x) = \frac{1}{2\theta} \exp\left(\frac{-|x|}{\theta}\right), \qquad x \in \mathcal{R}, \quad \theta > 0.$$

Find the shortest-length $(1-\alpha)$-level confidence interval for θ, based on the sufficient statistic $\sum_{i=1}^n |X_i|$.

6. In Example 4, let $R = X_{(n)} - X_{(1)}$. Find a $(1-\alpha)$-level confidence interval for θ of the form $(R, R/c)$. Compare the expected length of this interval to the one computed in Example 4.

7. Let $X_1, X_2, \dots, X_n$ be a random sample from a Pareto PDF $f_\theta(x) = \theta/x^2$, $x > \theta$, and $= 0$ for $x \leq \theta$. Show that the shortest-length confidence interval for θ based on $X_{(1)}$ is $(X_{(1)}\alpha^{1/n}, X_{(1)})$. (Use $\theta/X_{(1)}$ as a pivot.)

8. Let $X_1, X_2, \dots, X_n$ be a sample from PDF $f_\theta(x) = 1/(\theta_2 - \theta_1)$, $\theta_1 \leq x \leq \theta_2$, $\theta_1 < \theta_2$ and $= 0$ otherwise. Let $R = X_{(n)} - X_{(1)}$. Using $R/(\theta_2 - \theta_1)$ as a pivot for estimating $\theta_2 - \theta_1$, show that the shortest-length confidence interval is of the form $(R, R/c)$, where c is determined from the level as a solution of $c^{n-1}[(n-1)c - n] + \alpha = 0$. (Ferentinos [24])

11.5 UNBIASED AND EQUIVARIANT CONFIDENCE INTERVALS

In Section 11.3 we studied test inversion as one of the methods of constructing confidence intervals. We showed that UMP tests lead to UMA confidence intervals. In Chapter 9 we saw that UMP tests generally do not exist. In such situations we either restrict consideration to smaller subclasses of tests by requiring that the test functions have some desirable properties, or we restrict the class of alternatives to those near the null parameter values. In this section we follow a similar approach in constructing confidence intervals.

Definition 1. A family $\{S(\mathbf{x})\}$ of confidence sets for a parameter θ is said to be *unbiased* at confidence level $1-\alpha$ if

$$P_\theta\{S(\mathbf{X}) \text{ contains } \theta\} \geq 1-\alpha \tag{1}$$

and

$$P_\theta\{S(\mathbf{X}) \text{ contains } \theta'\} \leq 1-\alpha \qquad \text{for all } \theta,\ \theta' \in \Theta, \quad \theta \neq \theta'. \tag{2}$$

If $S(\mathbf{X})$ is an interval satisfying (1) and (2), we call it a $(1-\alpha)$-level unbiased confidence interval. If a family of unbiased confidence sets at level $1-\alpha$ is UMA in the class of all $(1-\alpha)$-level unbiased confidence sets, we call it a UMA unbiased (UMAU) family of confidence sets at level $1-\alpha$. In other words, if $S^*(\mathbf{x})$ satisfies (1) and (2) and minimizes

$$P_\theta\{S(\mathbf{X}) \text{ contains } \theta'\} \qquad \text{for } \theta,\ \theta' \in \Theta, \quad \theta \neq \theta'$$

among all unbiased families of confidence sets $S(\mathbf{X})$ at level $1-\alpha$, then $S^*(\mathbf{X})$ is a UMAU family of confidence sets at level $1-\alpha$.

Remark 1. Definition 1 says that a family $S(\mathbf{X})$ of confidence sets for a parameter θ is unbiased at level $1-\alpha$ if the probability of true coverage is at least $1-\alpha$ and that of false coverage is at most $1-\alpha$. In other words, $S(\mathbf{X})$ traps a true parameter value more often than it does a false one.

Theorem 1. Let $A(\theta_0)$ be the acceptance region of a UMP unbiased size α test of $H_0(\theta_0)\colon \theta=\theta_0$ against $H_1(\theta_0)\colon \theta \neq \theta_0$ for each θ_0. Then $S(\mathbf{x}) = \{\theta : \mathbf{x} \in A(\theta)\}$ is a UMA unbiased family of confidence sets at level $1-\alpha$.

Proof. To see that $S(\mathbf{x})$ is unbiased, we note that since $A(\theta)$ is the acceptance region of an unbiased test,

$$P_\theta\{S(\mathbf{X}) \text{ contains } \theta'\} = P_\theta\{\mathbf{X} \in A(\theta')\} \leq 1-\alpha.$$

We next show that $S(\mathbf{X})$ is UMA. Let $S^*(\mathbf{x})$ be any other unbiased $(1-\alpha)$-level family of confidence sets, and write $A^*(\theta) = \{\mathbf{x}\colon S^*(\mathbf{x}) \text{ contains } \theta\}$. Then $P_\theta\{\mathbf{X} \in$

$A^*(\theta')\} = P_\theta\{S^*(\mathbf{X})$ contains $\theta'\} \le 1-\alpha$, and it follows that $A^*(\theta)$ is the acceptance region of an unbiased size α test. Hence

$$\begin{aligned} P_\theta\{S^*(\mathbf{X}) \text{ contains } \theta'\} &= P_\theta\{\mathbf{X} \in A^*(\theta')\} \\ &\ge P_\theta\{\mathbf{X} \in A(\theta')\} \\ &= P_\theta\{S(\mathbf{X}) \text{ contains } \theta'\}. \end{aligned}$$

The inequality follows since $A(\theta)$ is the acceptance region of a UMP unbiased test. This completes the proof.

Example 1. Let $X_1, X_2, \ldots, X_n$ be a sample from $\mathcal{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown. For testing $H_0\colon \mu = \mu_0$ against $H_1: \mu \ne \mu_0$, it is known (Ferguson [25, p. 232]) that the t-test

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \dfrac{|\sqrt{n}(\overline{x} - \mu_0)|}{s} > c, \\ 0, & \text{otherwise,} \end{cases}$$

where $\overline{x} = \sum x_i/n$ and $s^2 = (n-1)^{-1} \sum (x_i - \overline{x})^2$ is UMP unbiased. We choose c from the size requirement

$$\alpha = P_{\mu=\mu_0}\left\{\left|\frac{\sqrt{n}\,(\overline{X} - \mu_0)}{S}\right| > c\right\},$$

so that $c = t_{n-1,\alpha/2}$. Thus

$$A(\mu_0) = \left\{\mathbf{x}\colon \left|\frac{\sqrt{n}\,(\overline{x} - \mu_0)}{s}\right| \le t_{n-1,\alpha/2}\right\}$$

is the acceptance region of a UMP unbiased size α test of $H_0\colon \mu = \mu_0$ against $H_1: \mu \ne \mu_0$. By Theorem 1 it follows that

$$\begin{aligned} S(\mathbf{x}) &= \{\mu\colon \mathbf{x} \in A(\mu)\} \\ &= \left\{\overline{x} - \frac{s}{\sqrt{n}} t_{n-1,\alpha/2} \le \mu \le \overline{x} + \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}\right\} \end{aligned}$$

is a UMA unbiased family of confidence sets at level $1-\alpha$.

If the measure of precision of a confidence interval is its expected length, one is naturally led to a consideration of unbiased confidence intervals. Pratt [79] has shown that the expected length of a confidence interval is the average of false coverage probabilities.

Theorem 2. Let Θ be an interval on the real line and f_θ be the PDF of $\mathbf{X}$. Let $S(\mathbf{X})$ be a family of $(1-\alpha)$-level confidence intervals of finite length; that is, let

$S(\mathbf{X}) = (\underline{\theta}(\mathbf{X}), \overline{\theta}(\mathbf{X}))$, and suppose that $\overline{\theta}(\mathbf{X}) - \underline{\theta}(\mathbf{X})$ is (random) finite. Then

$$\int (\overline{\theta}(\mathbf{x}) - \underline{\theta}(\mathbf{x})) f_\theta(\mathbf{x})\, d\mathbf{x} = \int_{\theta' \neq \theta} P_\theta\{S(\mathbf{X}) \text{ contains } \theta'\}\, d\theta' \tag{3}$$

for all $\theta \in \Theta$.

Proof. We have

$$\overline{\theta} - \underline{\theta} = \int_{\underline{\theta}}^{\overline{\theta}} d\theta'.$$

Thus for all $\theta \in \Theta$,

$$\begin{aligned}
E_\theta\{\overline{\theta}(\mathbf{X}) - \underline{\theta}(\mathbf{X})\} &= E_\theta\left\{\int_{\underline{\theta}}^{\overline{\theta}} d\theta'\right\} \\
&= \int f_\theta(\mathbf{x}) \left(\int_{\underline{\theta}}^{\overline{\theta}} d\theta'\right) d\mathbf{x} \\
&= \int \left[\int_{\underline{\theta}}^{\overline{\theta}} f_\theta(\mathbf{x})\, d\mathbf{x}\right] d\theta' \\
&= \int P_\theta\{S(\mathbf{X}) \text{ contains } \theta'\}\, d\theta' \\
&= \int_{\theta' \neq \theta} P_\theta\{S(\mathbf{X}) \text{ contains } \theta'\}\, d\theta'.
\end{aligned}$$

Remark 2. If $S(\mathbf{X})$ is a family of UMAU $(1-\alpha)$-level confidence intervals, the expected length of $S(\mathbf{X})$ is minimal. This follows since the left-hand side of (3) is the expected length, if θ is the true value, of $S(\mathbf{X})$ and $P_\theta\{S(\mathbf{X})$ contains $\theta'\}$ is minimal [because $S(\mathbf{X})$ is UMAU], by Theorem 1, with respect to all families of $1-\alpha$ unbiased confidence intervals uniformly in $\theta(\theta \neq \theta')$.

Since a reasonably complete discussion of UMP unbiased tests (see Section 9.5) is beyond the scope of this book, the following procedure for determining unbiased confidence intervals is sometimes quite useful (see Guenther [35]). Let $X_1, X_2, \ldots, X_n$ be a sample from an absolutely continuous DF with PDF $f_\theta(x)$, and suppose that we seek an unbiased confidence interval for θ. Following the discussion in Section 11.4, suppose that

$$T(X_1, X_2, \ldots, X_n, \theta) = T(\mathbf{X}, \theta) = T_\theta$$

is a pivot, and suppose that the statement

$$P\{\lambda_1(\alpha) < T_\theta < \lambda_2(\alpha)\} = 1 - \alpha$$

can be converted to

$$P_\theta\{\underline{\theta}(\mathbf{X}) < \theta < \overline{\theta}(\mathbf{X})\} = 1 - \alpha.$$

For $(\underline{\theta}, \overline{\theta})$ to be unbiased, we must have

$$P(\theta, \theta') = P_\theta\{\underline{\theta}(\mathbf{X}) < \theta' < \overline{\theta}(\mathbf{X})\} = 1 - \alpha \qquad \text{if } \theta' = \theta \tag{4}$$

and

$$P(\theta, \theta') < 1 - \alpha \qquad \text{if } \theta' \neq \theta. \tag{5}$$

If $P(\theta, \theta')$ depends only on a function γ of θ, θ', we may write

$$P(\gamma) \begin{cases} = 1 - \alpha & \text{if } \theta' = \theta, \\ < 1 - \alpha & \text{if } \theta' \neq \theta, \end{cases} \tag{6}$$

and it follows that $P(\gamma)$ has a maximum at $\theta' = \theta$.

Example 2. Let $X_1, X_2, \dots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ RVs, and suppose that we desire an unbiased confidence interval for σ^2. Then

$$T(\mathbf{X}, \sigma^2) = \frac{(n-1)S^2}{\sigma^2} = T_\sigma$$

has a $\chi^2(n-1)$ distribution, and we have

$$P\left\{\lambda_1 < (n-1)\frac{S^2}{\sigma^2} < \lambda_2\right\} = 1 - \alpha,$$

so that

$$P\left\{(n-1)\frac{S^2}{\lambda_2} < \sigma^2 < (n-1)\frac{S^2}{\lambda_1}\right\} = 1 - \alpha.$$

Then

$$\begin{aligned} P(\sigma^2, \sigma'^2) &= P_{\sigma^2}\left\{(n-1)\frac{S^2}{\lambda_2} < \sigma'^2 < (n-1)\frac{S^2}{\lambda_1}\right\} \\ &= P\left\{\frac{T_\sigma}{\lambda_2} < \gamma < \frac{T_\sigma}{\lambda_1}\right\}, \end{aligned}$$

where $\gamma = \sigma'^2/\sigma^2$ and $T_\sigma \sim \chi^2(n-1)$. Thus

$$P(\gamma) = P\{\lambda_1\gamma < T_\sigma < \lambda_2\gamma\}.$$

Then

$$P(1) = 1-\alpha \quad \text{and} \quad P(\gamma) < 1-\alpha.$$

Thus we need λ_1, λ_2 such that

$$P(1) = 1-\alpha \tag{7}$$

and

$$\left.\frac{dP(\gamma)}{d\gamma}\right|_{\gamma=1} = \lambda_2 f_{n-1}(\lambda_2) - \lambda_1 f_{n-1}(\lambda_1) = 0, \tag{8}$$

where f_{n-1} is the PDF of T_σ. Equations (7) and (8) have been solved numerically for λ_1, λ_2 by several authors (see, for example, Tate and Klett [111]). Having obtained λ_1, λ_2 from (7) and (8), we have as the unbiased $(1-\alpha)$-level confidence interval

$$\left((n-1)\frac{S^2}{\lambda_2},\ (n-1)\frac{S^2}{\lambda_1}\right). \tag{9}$$

Note that in this case the shortest-length confidence interval (based on T_σ) derived in Example 11.4.3, the usual equal-tails confidence interval, and (9) are all different. The length of the confidence interval (9), however, can be considerably greater than that of the shortest interval of Example 11.4.3. For large n all three sets of intervals are approximately the same.

Finally, let us briefly investigate how invariance considerations apply to confidence estimation. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n) \sim f_\theta, \theta \in \Theta \subseteq \mathcal{R}$. Let $\mathcal{G}$ be a group of transformations on $\mathcal{X}$ that leaves $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$ invariant. Let $S(\mathbf{X})$ be a $(1-\alpha)$-level confidence set for θ.

Definition 2. Let $\mathcal{P}$ be invariant under $\mathcal{G}$, and let $S(\mathbf{x})$ be a confidence set for θ. Then S is *equivariant* under $\mathcal{G}$ if for every $\mathbf{x} \in \mathcal{X}, \theta \in \Theta$, and $g \in \mathcal{G}$,

$$S(\mathbf{x}) \in \theta \Leftrightarrow S(g(\mathbf{x})) \ni \bar{g}\theta. \tag{10}$$

Example 3. Let $X_1, X_2, \ldots, X_n$ be a sample from PDF

$$f_\theta(x) = \exp[-(x-\theta)], \qquad x > \theta$$

and $= 0$ if $x \le \theta$. Let $\mathcal{G} = \{\{a, 1\} : a \in \mathcal{R}\}$, where $\{a, 1\}\mathbf{x} = (x_1 + a, x_2 + a, \ldots, x_n + a)$ and $\mathcal{G}$ induces $\bar{\mathcal{G}} = \mathcal{G}$ on $\Theta = \mathcal{R}$. The family $\{f_\theta\}$ remains invariant under $\mathcal{G}$. Consider a confidence interval of the form

$$S(\mathbf{x}) = \{\theta : \overline{x} - c_1 \le \theta \le \overline{x} + c_2\}$$

where c_1, c_2 are constants. Then

$$S(\{a, 1\}\mathbf{x}) = \{\theta : \overline{x} + a - c_1 \le \theta \le \overline{x} + a - c_2\}.$$

Clearly,

$$\begin{aligned} S(\mathbf{x}) \ni \theta &\Longleftrightarrow \overline{x} + a - c_1 \le \theta + a \le \overline{x} + a - c_2 \\ &\Longleftrightarrow S(\{a, 1\}\mathbf{x}) \ni \overline{g}\theta \end{aligned}$$

and it follows that $S(\mathbf{x})$ is an equivariant confidence interval.

The most useful method of constructing invariant confidence intervals is test inversion. Inverting the acceptance region of invariant tests often leads to equivariant confidence intervals under certain conditions. Recall that a group $\mathcal{G}$ of transformations leaves a hypothesis-testing problem invariant if $\mathcal{G}$ leaves both Θ_0 and Θ_1 invariant. For each $H_0 : \theta = \theta_0$, $\theta_0 \in \Theta$ we have a different group of transformations, $\mathcal{G}_{\theta_0}$, which leaves the problem of testing $\theta = \theta_0$ invariant. The equivariant confidence interval, on the other hand, must be equivariant with respect to $\mathcal{G}$, which is a much larger group since $\mathcal{G} \supset \mathcal{G}_{\theta_0}$ for all θ_0. The relationship between an equivariant confidence set and invariant tests is more complicated when the family $\mathcal{P}$ has a nuisance parameter τ.

Under certain conditions there is a relationship between equivariant confidence sets and associated invariant tests. Rather than pursue this relationship, we refer the reader to Ferguson [27, p. 262]; it is generally easy to check that (10) holds for a given confidence interval S to show that S is invariant. The following example illustrates this point.

Example 4. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ RVs where both μ and σ^2 are unknown. In Example 9.5.3 we showed that the test

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_1^n (x_i - \overline{x})^2 \le \sigma_0^2 \chi^2_{n-1,1-\alpha} \\ 0 & \text{otherwise} \end{cases}$$

is UMP invariant, under translation group for testing $H_0 : \sigma^2 \ge \sigma_0^2$ against $H_1 : \sigma^2 < \sigma_0^2$. Then the acceptance region of ϕ is

$$A(\mathbf{x}) = \left\{ \mathbf{x} : \sum_1^n (x_i - \overline{x})^2 > \sigma_0^2 \chi^2_{n-1,1-\alpha} \right\}.$$

Clearly,

$$\mathbf{x} \in A(\mathbf{x}) \Longleftrightarrow \sigma_0^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha}}$$

and it follows that

$$S(\mathbf{x}) = \left\{\sigma^2 : \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha}}\right\}$$

is a $(1-\alpha)$-level confidence interval (upper confidence bound) for σ^2. We show that S is invariant with respect to the scale group. In fact,

$$S(\{0, c\}\mathbf{x}) = \left\{\sigma^2 : \sigma^2 < \frac{c^2(n-1)s^2}{\chi^2_{n-1,1-\alpha}}\right\}$$

and

$$\sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha}} \Longleftrightarrow S(\{0, c\}\mathbf{x}) \ni \bar{g}\sigma^2 = \{0, c\}\sigma^2$$

and it follows that $S(\mathbf{x})$ is an equivariant confidence interval for σ^2.

PROBLEMS 11.5

1. Let $X_1, X_2, \ldots, X_n$ be a sample from $U(0, \theta)$. Show that the unbiased confidence intervals for θ based on the pivot $\max X_i/\theta$, coincides with the shortest-length confidence interval based on the same pivot.

2. Let $X_1, X_2, \ldots, X_n$ be a sample from $\mathcal{G}(1, \theta)$. Find the unbiased confidence interval for θ based on the pivot $2\sum_{i=1}^n X_i/\theta$.

3. Let $X_1, X_2, \ldots, X_n$ be a sample from the PDF

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)} & \text{if } x > \theta \\ 0 & \text{otherwise.} \end{cases}$$

   Find the unbiased confidence interval based on the pivot $2n[\min X_i - \theta]$.

4. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ RVs where both μ and σ^2 are unknown. Using the pivot $T_{\mu,\sigma} = \sqrt{n}(\overline{X} - \mu)/S$, show that the shortest-length unbiased $(1-\alpha)$-level confidence interval for μ is the equal-tails interval $(\overline{X} - t_{n-1,\alpha/2}S/\sqrt{n}, \overline{X} + t_{n-1,\alpha/2}S/\sqrt{n})$.

5. Let $X_1, X_2, \ldots, X_n$ be iid with PDF $f_\theta(x) = \theta/x^2$, $x \geq \theta$, and $= 0$ otherwise. Find the shortest length $(1-\alpha)$-level unbiased confidence interval for θ based on the pivot $\theta/X_{(1)}$.

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from a location family $\mathcal{P} = \{f_\theta(x) = f(x-\theta); \theta \in \mathcal{R}\}$. Show that a confidence interval of the form $S(\mathbf{x}) = \{\theta : T(\mathbf{x}) - c_1 \leq \theta \leq T(\mathbf{x}) + c_2\}$, where $T(\mathbf{x})$ is an equivariant estimate under location group is an equivariant confidence interval.

7. Let $X_1, X_2, \ldots, X_n$ be iid RVs with common scale PDF $f_\sigma(x) = (1/\sigma)f(x/\sigma)$, $\sigma > 0$. Consider the scale group $\mathcal{G} = \{\{0, b\} : b > 0\}$. If $T(\mathbf{x})$ is an equivariant estimate of σ, show that a confidence interval of the form

$$S(\mathbf{x}) = \left\{\sigma : c_1 \leq \frac{T(\mathbf{x})}{\sigma} \leq c_2\right\}$$

is equivariant.

8. Let $X_1, X_2, \ldots, X_n$ be iid RVs with PDF $f_\theta(x) = \exp[-(x-\theta)]$, $x > \theta$ and $= 0$, otherwise. For testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$, consider the (UMP) test

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } x_{(1)} \geq \theta_0 - \dfrac{\ln \alpha}{n}, \\ 0, & \text{otherwise.} \end{cases}$$

Is the acceptance region of this α-level test an equivariant $(1-\alpha)$-level confidence interval (lower bound) for θ with respect to the location group?