Chapter 9

Interval Estimation

"I fear," said Holmes, "that if the matter is beyond humanity it is certainly beyond me. Yet we must exhaust all natural explanations before we fall back upon such a theory as this."

Sherlock Holmes
The Adventure of the Devil's Foot

9.1 Introduction

In Chapter 7 we discussed point estimation of a parameter θ, where the inference is a guess of a single value as the value of θ. In this chapter we discuss interval estimation and, more generally, set estimation. The inference in a set estimation problem is the statement that "$\theta \in C$," where $C \subset \Theta$ and $C = C(\mathbf{x})$ is a set determined by the value of the data $\mathbf{X} = \mathbf{x}$ observed. If θ is real-valued, then we usually prefer the set estimate C to be an interval. Interval estimators will be the main topic of this chapter.

As in the previous two chapters, this chapter is divided into two parts, the first concerned with finding interval estimators and the second part concerned with evaluating the worth of the estimators. We begin with a formal definition of interval estimator, a definition as vague as the definition of point estimator.

Definition 9.1.1 An *interval estimate* of a real-valued parameter θ is any pair of functions, $L(x_1, \ldots, x_n)$ and $U(x_1, \ldots, x_n)$, of a sample that satisfy $L(\mathbf{x}) \le U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. If $\mathbf{X} = \mathbf{x}$ is observed, the inference $L(\mathbf{x}) \le \theta \le U(\mathbf{x})$ is made. The random interval $[L(\mathbf{X}), U(\mathbf{X})]$ is called an *interval estimator.*

We will use our previously defined conventions and write $[L(\mathbf{X}), U(\mathbf{X})]$ for an interval estimator of θ based on the random sample $\mathbf{X} = (X_1, \ldots, X_n)$ and $[L(\mathbf{x}), U(\mathbf{x})]$ for the realized value of the interval. Although in the majority of cases we will work with finite values for L and U, there is sometimes interest in *one-sided* interval estimates. For instance, if $L(\mathbf{x}) = -\infty$, then we have the one-sided interval $(-\infty, U(\mathbf{x})]$ and the assertion is that "$\theta \le U(\mathbf{x})$," with no mention of a lower bound. We could similarly take $U(\mathbf{x}) = \infty$ and have a one-sided interval $[L(\mathbf{x}), \infty)$.

Although the definition mentions a closed interval $[L(\mathbf{x}), U(\mathbf{x})]$, it will sometimes be more natural to use an open interval $(L(\mathbf{x}), U(\mathbf{x}))$ or even a half-open and half-closed interval, as in the previous paragraph. We will use whichever seems most appropriate for the particular problem at hand, although the preference will be for a closed interval.

Example 9.1.2 (Interval estimator) For a sample X_1, X_2, X_3, X_4 from a $\mathrm{n}(\mu, 1)$, an interval estimator of μ is $[\bar{X}-1, \bar{X}+1]$. This means that we will assert that μ is in this interval. ||

At this point, it is natural to inquire as to what is gained by using an interval estimator. Previously, we estimated μ with $\bar{X}$, and now we have the less precise estimator $[\bar{X}-1, \bar{X}+1]$. We surely must gain something! By giving up some precision in our estimate (or assertion about μ), we have gained some confidence, or assurance, that our assertion is correct.

Example 9.1.3 (Continuation of Example 9.1.2) When we estimate μ by $\bar{X}$, the probability that we are exactly correct, that is, $P(\bar{X}=\mu)$, is 0. However, with an interval estimator, we have a positive probability of being correct. The probability that μ is covered by the interval $[\bar{X}-1, \bar{X}+1]$ can be calculated as

$$\begin{aligned}
P(\mu \in [\bar{X}-1, \bar{X}+1]) &= P(\bar{X}-1 \le \mu \le \bar{X}+1) \\
&= P(-1 \le \bar{X}-\mu \le 1) \\
&= P\left(-2 \le \frac{\bar{X}-\mu}{\sqrt{1/4}} \le 2\right) \\
&= P(-2 \le Z \le 2) \qquad \left(\frac{\bar{X}-\mu}{\sqrt{1/4}} \text{ is standard normal}\right) \\
&= .9544.
\end{aligned}$$

Thus we have over a 95% chance of covering the unknown parameter with our interval estimator. Sacrificing some precision in our estimate, in moving from a point to an interval, has resulted in increased confidence that our assertion is correct. ||

The purpose of using an interval estimator rather than a point estimator is to have some guarantee of capturing the parameter of interest. The certainty of this guarantee is quantified in the following definitions.

Definition 9.1.4 For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, the *coverage probability* of $[L(\mathbf{X}), U(\mathbf{X})]$ is the probability that the random interval $[L(\mathbf{X}), U(\mathbf{X})]$ covers the true parameter, θ. In symbols, it is denoted by either $P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$ or $P(\theta \in [L(\mathbf{X}), U(\mathbf{X})] | \theta)$.

Definition 9.1.5 For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, the *confidence coefficient* of $[L(\mathbf{X}), U(\mathbf{X})]$ is the infimum of the coverage probabilities, $\inf_\theta P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$.

There are a number of things to be aware of in these definitions. One, it is important to keep in mind that the *interval* is the random quantity, not the parameter. Therefore, when we write probability statements such as $P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$, these probability statements refer to $\mathbf{X}$, not θ. In other words, think of $P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$, which might look like a statement about a random θ, as the algebraically equivalent $P_\theta(L(\mathbf{X}) \le \theta, U(\mathbf{X}) \ge \theta)$, a statement about a random $\mathbf{X}$.

Interval estimators, together with a measure of confidence (usually a confidence coefficient), are sometimes known as *confidence intervals*. We will often use this term interchangeably with *interval estimator*. Although we are mainly concerned with confidence *intervals*, we occasionally will work with more general sets. When working in general, and not being quite sure of the exact form of our sets, we will speak of confidence *sets*. A confidence set with confidence coefficient equal to some value, say $1-\alpha$, is simply called a $1-\alpha$ *confidence set*.

Another important point is concerned with coverage probabilities and confidence coefficients. Since we do not know the true value of θ, we can only guarantee a coverage probability equal to the infimum, the confidence coefficient. In some cases this does not matter because the coverage probability will be a constant function of θ. In other cases, however, the coverage probability can be a fairly variable function of θ.

Example 9.1.6 (Scale uniform interval estimator) Let $X_1, \ldots, X_n$ be a random sample from a uniform$(0, \theta)$ population and let $Y = \max\{X_1, \ldots, X_n\}$. We are interested in an interval estimator of θ. We consider two candidate estimators: $[aY, bY], 1 \le a < b$, and $[Y+c, Y+d]$, $0 \le c < d$, where a, b, c, and d are specified constants. (Note that θ is necessarily larger than y.) For the first interval we have

$$\begin{aligned} P_\theta(\theta \in [aY, bY]) &= P_\theta(aY \le \theta \le bY) \\ &= P_\theta\left(\frac{1}{b} \le \frac{Y}{\theta} \le \frac{1}{a}\right) \\ &= P_\theta\left(\frac{1}{b} \le T \le \frac{1}{a}\right). \qquad (T = Y/\theta) \end{aligned}$$

We previously saw (Example 7.3.13) that $f_Y(y) = ny^{n-1}/\theta^n, 0 \le y \le \theta$, so the pdf of T is $f_T(t) = nt^{n-1}$, $0 \le t \le 1$. We therefore have

$$P_\theta\left(\frac{1}{b} \le T \le \frac{1}{a}\right) = \int_{1/b}^{1/a} nt^{n-1}\, dt = \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n.$$

The coverage probability of the first interval is independent of the value of θ, and thus $(\frac{1}{a})^n - (\frac{1}{b})^n$ is the confidence coefficient of the interval.

For the other interval, for $\theta \ge d$ a similar calculation yields

$$\begin{aligned} P_\theta(\theta \in [Y+c, Y+d]) &= P_\theta(Y + c \le \theta \le Y + d) \\ &= P_\theta\left(1 - \frac{d}{\theta} \le T \le 1 - \frac{c}{\theta}\right) \qquad (T = Y/\theta) \\ &= \int_{1-d/\theta}^{1-c/\theta} nt^{n-1}\, dt \\ &= \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n. \end{aligned}$$

In this case, the coverage probability depends on θ. Furthermore, it is straightforward to calculate that for any constants c and d,

$$\lim_{\theta\to\infty}\left(1-\frac{c}{\theta}\right)^n-\left(1-\frac{d}{\theta}\right)^n=0,$$

showing that the confidence coefficient of this interval estimator is 0. ||

9.2 Methods of Finding Interval Estimators

We present four subsections of methods of finding estimators. This might seem to indicate that there are four different methods for finding interval estimators. This is really not so; in fact, operationally all of the methods presented in the next four subsections are the same, being based on the strategy of inverting a test statistic. The last subsection, dealing with Bayesian intervals, presents a different construction method.

9.2.1 Inverting a Test Statistic

There is a very strong correspondence between hypothesis testing and interval estimation. In fact, we can say in general that every confidence set corresponds to a test and vice versa. Consider the following example.

Example 9.2.1 (Inverting a normal test) Let $X_1,\dots,X_n$ be iid $\mathrm{n}(\mu,\sigma^2)$ and consider testing H_0: $\mu=\mu_0$ versus H_1: $\mu\neq\mu_0$. For a fixed α level, a reasonable test (in fact, the most powerful unbiased test) has rejection region $\{\mathbf{x}: |\bar{x}-\mu_0|>z_{\alpha/2}\sigma/\sqrt{n}\}$. Note that H_0 is accepted for sample points with $|\bar{x}-\mu_0|\le z_{\alpha/2}\sigma/\sqrt{n}$ or, equivalently,

$$\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu_0\le\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Since the test has size α, this means that $P(H_0 \text{ is rejected}|\mu=\mu_0)=\alpha$ or, stated in another way, $P(H_0 \text{ is accepted}|\mu=\mu_0)=1-\alpha$. Combining this with the above characterization of the acceptance region, we can write

$$P\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu_0\le\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\middle|\mu=\mu_0\right)=1-\alpha.$$

But this probability statement is true *for every* μ_0. Hence, the statement

$$P_\mu\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

is true. The interval $[\bar{x}-z_{\alpha/2}\sigma/\sqrt{n},\ \bar{x}+z_{\alpha/2}\sigma/\sqrt{n}]$, obtained by *inverting* the acceptance region of the level α test, is a $1-\alpha$ confidence interval. ||

We have illustrated the correspondence between confidence sets and tests. The acceptance region of the hypothesis test, the set in the *sample space* for which

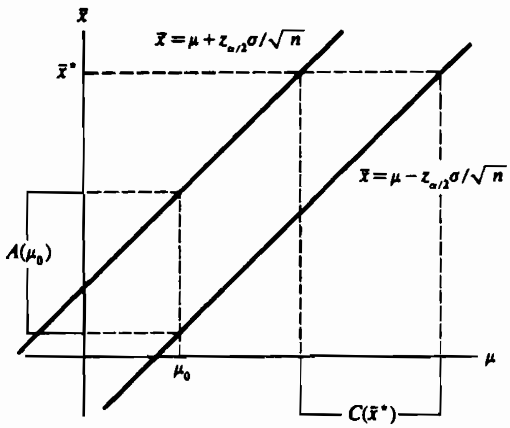

Figure 9.2.1. *Relationship between confidence intervals and acceptance regions for tests. The upper line is $\bar{x} = \mu + z_{\alpha/2}\sigma/\sqrt{n}$ and the lower line is $\bar{x} = \mu - z_{\alpha/2}\sigma/\sqrt{n}$.*

H_0: $\mu = \mu_0$ is accepted, is given by

$$A(\mu_0) = \left\{(x_1, \ldots, x_n) : \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\},$$

and the *confidence interval*, the set in the *parameter space* with plausible values of μ, is given by

$$C(x_1, \ldots, x_n) = \left\{\mu : \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}.$$

These sets are connected to each other by the tautology

$$(x_1, \ldots, x_n) \in A(\mu_0) \Leftrightarrow \mu_0 \in C(x_1, \ldots, x_n).$$

The correspondence between testing and interval estimation for the two-sided normal problem is illustrated in Figure 9.2.1. There it is, perhaps, more easily seen that both tests and intervals ask the same question, but from a slightly different perspective. Both procedures look for consistency between sample statistics and population parameters. The hypothesis test fixes the parameter and asks what sample values (the acceptance region) are consistent with that fixed value. The confidence set fixes the sample value and asks what parameter values (the confidence interval) make this sample value most plausible.

The correspondence between acceptance regions of tests and confidence sets holds in general. The next theorem gives a formal version of this correspondence.

Theorem 9.2.2 *For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test of $H_0: \theta = \theta_0$. For each $\mathbf{x} \in \mathcal{X}$, define a set $C(\mathbf{x})$ in the parameter space by*

$$C(\mathbf{x}) = \{\theta_0 : \mathbf{x} \in A(\theta_0)\}. \tag{9.2.1}$$

Then the random set $C(\mathbf{X})$ *is a* $1-\alpha$ *confidence set. Conversely, let* $C(\mathbf{X})$ *be a* $1-\alpha$ *confidence set. For any* $\theta_0 \in \Theta$, *define*

$$A(\theta_0) = \{\mathbf{x}: \theta_0 \in C(\mathbf{x})\}.$$

Then $A(\theta_0)$ *is the acceptance region of a level* α *test of* $H_0: \theta = \theta_0$.

Proof: For the first part, since $A(\theta_0)$ is the acceptance region of a level α test,

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) \leq \alpha \quad \text{and hence} \quad P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \geq 1-\alpha.$$

Since θ_0 is arbitrary, write θ instead of θ_0. The above inequality, together with (9.2.1), shows that the coverage probability of the set $C(\mathbf{X})$ is given by

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

showing that $C(\mathbf{X})$ is a $1-\alpha$ confidence set.

For the second part, the Type I Error probability for the test of $H_0: \theta = \theta_0$ with acceptance region $A(\theta_0)$ is

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) = P_{\theta_0}(\theta_0 \notin C(\mathbf{X})) \leq \alpha.$$

So this is a level α test. □

Although it is common to talk about inverting a test to obtain a confidence set, Theorem 9.2.2 makes it clear that we really have a family of tests, one for each value of $\theta_0 \in \Theta$, that we invert to obtain one confidence set.

The fact that tests can be inverted to obtain a confidence set and vice versa is theoretically interesting, but the really useful part of Theorem 9.2.2 is the first part. It is a relatively easy task to construct a level α acceptance region. The difficult task is constructing a confidence set. So the method of obtaining a confidence set by inverting an acceptance region is quite useful. All of the techniques we have for obtaining tests can immediately be applied to constructing confidence sets.

In Theorem 9.2.2, we stated only the null hypothesis $H_0: \theta = \theta_0$. All that is required of the acceptance region is

$$P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \geq 1-\alpha.$$

In practice, when constructing a confidence set by test inversion, we will also have in mind an alternative hypothesis such as $H_1: \theta \neq \theta_0$ or $H_1: \theta > \theta_0$. The alternative will dictate the form of $A(\theta_0)$ that is reasonable, and the form of $A(\theta_0)$ will determine the shape of $C(\mathbf{x})$. Note, however, that we carefully used the word *set* rather than *interval*. This is because there is no guarantee that the confidence set obtained by test inversion will be an interval. In most cases, however, one-sided tests give one-sided intervals, two-sided tests give two-sided intervals, strange-shaped acceptance regions give strange-shaped confidence sets. Later examples will exhibit this.

The properties of the inverted test also carry over (sometimes suitably modified) to the confidence set. For example, unbiased tests, when inverted, will produce unbiased confidence sets. Also, and more important, since we know that we can confine

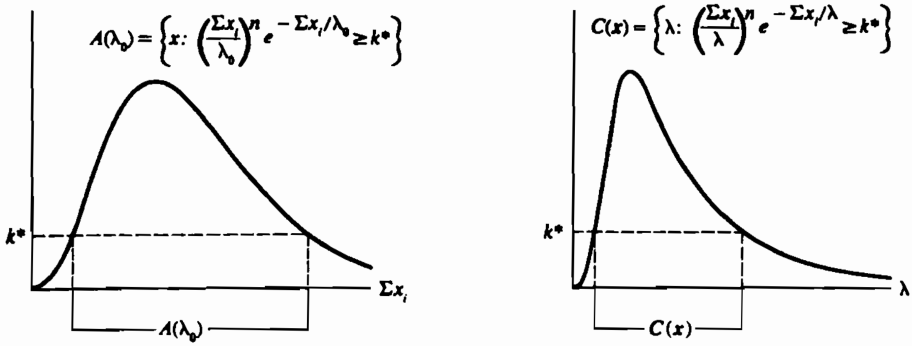

Figure 9.2.2. *Acceptance region and confidence interval for Example 9.2.3. The acceptance region is* $A(\lambda_0) = \left\{\mathbf{x} : \left(\sum_i x_i/\lambda_0\right)^n e^{-\sum_i x_i/\lambda_0} \geq k^*\right\}$ *and the confidence region is* $C(\mathbf{x}) = \left\{\lambda : \left(\sum_i x_i/\lambda\right)^n e^{-\sum_i x_i/\lambda} \geq k^*\right\}$.

attention to sufficient statistics when looking for a good test, it follows that we can confine attention to sufficient statistics when looking for good confidence sets.

The method of test inversion really is most helpful in situations where our intuition deserts us and we have no good idea as to what would constitute a reasonable set. We merely fall back on our all-purpose method for constructing a reasonable test.

Example 9.2.3 (Inverting an LRT) Suppose that we want a confidence interval for the mean, λ, of an exponential(λ) population. We can obtain such an interval by inverting a level α test of $H_0: \lambda = \lambda_0$ versus $H_1: \lambda \neq \lambda_0$.

If we take a random sample $X_1, \ldots, X_n$, the LRT statistic is given by

$$\frac{\frac{1}{\lambda_0^n} e^{-\Sigma x_i/\lambda_0}}{\sup_\lambda \frac{1}{\lambda^n} e^{-\Sigma x_i/\lambda}} = \frac{\frac{1}{\lambda_0^n} e^{-\Sigma x_i/\lambda_0}}{\frac{1}{(\sum x_i/n)^n} e^{-n}} = \left(\frac{\sum x_i}{n\lambda_0}\right)^n e^n e^{-\Sigma x_i/\lambda_0}.$$

For fixed λ_0, the acceptance region is given by

$$A(\lambda_0) = \left\{\mathbf{x} \colon \left(\frac{\sum x_i}{\lambda_0}\right)^n e^{-\Sigma x_i/\lambda_0} \geq k^*\right\}, \tag{9.2.2}$$

where k^* is a constant chosen to satisfy $P_{\lambda_0}(\mathbf{X} \in A(\lambda_0)) = 1 - \alpha$. (The constant e^n/n^n has been absorbed into k^*.) This is a set in the sample space as shown in Figure 9.2.2. Inverting this acceptance region gives the $1 - \alpha$ confidence set

$$C(\mathbf{x}) = \left\{\lambda \colon \left(\frac{\sum x_i}{\lambda}\right)^n e^{-\Sigma x_i/\lambda} \geq k^*\right\}.$$

This is an interval in the parameter space as shown in Figure 9.2.2.

The expression defining $C(\mathbf{x})$ depends on $\mathbf{x}$ only through $\sum x_i$. So the confidence interval can be expressed in the form

$$C(\textstyle\sum x_i) = \{\lambda \colon L(\sum x_i) \leq \lambda \leq U(\sum x_i)\}, \tag{9.2.3}$$

where L and U are functions determined by the constraints that the set (9.2.2) has probability $1-\alpha$ and

$$\left(\frac{\sum x_i}{L(\sum x_i)}\right)^n e^{-\Sigma x_i/L(\Sigma x_i)} = \left(\frac{\sum x_i}{U(\sum x_i)}\right)^n e^{-\Sigma x_i/U(\Sigma x_i)}. \tag{9.2.4}$$

If we set

$$\frac{\sum x_i}{L(\sum x_i)} = a \quad \text{and} \quad \frac{\sum x_i}{U(\sum x_i)} = b, \tag{9.2.5}$$

where $a > b$ are constants, then (9.2.4) becomes

$$a^n e^{-a} = b^n e^{-b}, \tag{9.2.6}$$

which yields easily to numerical solution. To work out some details, let $n = 2$ and note that $\sum X_i \sim \text{gamma}(2, \lambda)$ and $\sum X_i/\lambda \sim \text{gamma}(2,1)$. Hence, from (9.2.5), the confidence interval becomes $\{\lambda\colon \frac{1}{a}\sum x_i \le \lambda \le \frac{1}{b}\sum x_i\}$, where a and b satisfy

$$P_\lambda\left(\frac{1}{a}\sum X_i \le \lambda \le \frac{1}{b}\sum X_i\right) = P\left(b \le \frac{\sum X_i}{\lambda} \le a\right) = 1-\alpha$$

and, from (9.2.6), $a^2e^{-a} = b^2e^{-b}$. Then

$$\begin{aligned} P\left(b \le \frac{\sum X_i}{\lambda} \le a\right) &= \int_b^a te^{-t}\,dt \\ &= e^{-b}(b+1) - e^{-a}(a+1). \qquad \begin{pmatrix}\text{integration}\\ \text{by parts}\end{pmatrix} \end{aligned}$$

To get, for example, a 90% confidence interval, we must simultaneously satisfy the probability condition and the constraint. To three decimal places, we get $a = 5.480$, $b = .441$, with a confidence coefficient of .90006. Thus,

$$P_\lambda\left(\frac{1}{5.480}\sum X_i \le \lambda \le \frac{1}{.441}\sum X_i\right) = .90006. \qquad \|$$

The region obtained by inverting the LRT of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \ne \theta_0$ (Definition 8.2.1) is of the form

$$\text{accept } H_0 \text{ if } \frac{L(\theta_0|\mathbf{x})}{L(\hat\theta|\mathbf{x})} \le k(\theta_0),$$

with the resulting confidence region

$$\{\theta : L(\theta|\mathbf{x}) \ge k'(\mathbf{x},\theta)\}, \tag{9.2.7}$$

for some function k' that gives $1-\alpha$ confidence.

In some cases (such as the normal and the gamma distribution) the function k' will not depend on θ. In such cases the likelihood region has a particularly pleasing interpretation, consisting of those values of θ for which the likelihood is highest. We will also see such intervals arising from optimality considerations in both the frequentist (Theorem 9.3.2) and Bayesian (Corollary 9.3.10) realms.

The test inversion method is completely general in that we can invert any test and obtain a confidence set. In Example 9.2.3 we inverted LRTs, but we could have used a test constructed by any method. Also, note that the inversion of a two-sided test gave a two-sided interval. In the next examples, we invert one-sided tests to get one-sided intervals.

Example 9.2.4 (Normal one-sided confidence bound) Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{n}(\mu, \sigma^2)$ population. Consider constructing a $1-\alpha$ upper confidence bound for μ. That is, we want a confidence interval of the form $C(\mathbf{x}) = (-\infty, U(\mathbf{x})]$. To obtain such an interval using Theorem 9.2.2, we will invert one-sided tests of $H_0\colon \mu = \mu_0$ versus $H_1\colon \mu < \mu_0$. (Note that we use the specification of H_1 to determine the form of the confidence interval here. H_1 specifies "large" values of μ_0, so the confidence set will contain "small" values, values less than a bound. Thus, we will get an upper confidence bound.) The size α LRT of H_0 versus H_1 rejects H_0 if

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -t_{n-1,\alpha}$$

(similar to Example 8.2.6). Thus the acceptance region for this test is

$$A(\mu_0) = \left\{\mathbf{x} : \bar{x} \geq \mu_0 - t_{n-1,\alpha}\frac{s}{\sqrt{n}}\right\}$$

and $\mathbf{x} \in A(\mu_0) \Leftrightarrow \bar{x} + t_{n-1,\alpha}s/\sqrt{n} \geq \mu_0$. According to (9.2.1), we define

$$C(\mathbf{x}) = \{\mu_0 : \mathbf{x} \in A(\mu_0)\} = \left\{\mu_0 \colon \bar{x} + t_{n-1,\alpha}\frac{s}{\sqrt{n}} \geq \mu_0\right\}.$$

By Theorem 9.2.2, the random set $C(\mathbf{X}) = (-\infty, \bar{X} + t_{n-1,\alpha}S/\sqrt{n}]$ is a $1-\alpha$ confidence set for μ. We see that, indeed, it is the right form for an upper confidence bound. Inverting the one-sided test gave a one-sided confidence interval. ||

Example 9.2.5 (Binomial one-sided confidence bound) As a more difficult example of a one-sided confidence interval, consider putting a $1-\alpha$ lower confidence bound on p, the success probability from a sequence of Bernoulli trials. That is, we observe $X_1, \ldots, X_n$, where $X_i \sim \mathrm{Bernoulli}(p)$, and we want the interval to be of the form $(L(x_1, \ldots, x_n), 1]$, where $P_p(p \in (L(X_1, \ldots, X_n), 1]) \geq 1 - \alpha$. (The interval we obtain turns out to be open on the left, as will be seen.)

Since we want a one-sided interval that gives a lower confidence bound, we consider inverting the acceptance regions from tests of

$$H_0\colon \quad p = p_0 \qquad \text{versus} \qquad H_1\colon \quad p > p_0.$$

To simplify things, we know that we can base our test on $T = \sum_{i=1}^n X_i \sim$ binomial(n, p), since T is sufficient for p. (See the Miscellanea section.) Since the binomial distribution has monotone likelihood ratio (see Exercise 8.25), by the Karlin–Rubin Theorem (Theorem 8.3.17) the test that rejects H_0 if $T > k(p_0)$ is the UMP test of its size. For each p_0, we want to choose the constant $k(p_0)$ (it can be an integer) so that we have a level α test. We cannot get the size of the test to be exactly α, except for certain values of p_0, because of the discreteness of T. But we choose $k(p_0)$ so that the size of the test is as close to α as possible, without being larger. Thus, $k(p_0)$ is defined to be the integer between 0 and n that simultaneously satisfies the inequalities

$$\sum_{y=0}^{k(p_0)} \binom{n}{y} p_0^y (1-p_0)^{n-y} \geq 1-\alpha$$

(9.2.8)

$$\sum_{y=0}^{k(p_0)-1} \binom{n}{y} p_0^y (1-p_0)^{n-y} < 1-\alpha.$$

Because of the MLR property of the binomial, for any $k = 0, \ldots, n$, the quantity

$$f(p_0|k) = \sum_{y=0}^{k} \binom{n}{y} p_0^y (1-p_0)^{n-y}$$

is a decreasing function of p_0 (see Exercise 8.26). Of course, $f(0|0) = 1$, so $k(0) = 0$ and $f(p_0|0)$ remains above $1-\alpha$ for an interval of values. Then, at some point $f(p_0|0) = 1-\alpha$ and for values of p_0 greater than this value, $f(p_0|0) < 1-\alpha$. So, at this point, $k(p_0)$ increases to 1. This pattern continues. Thus, $k(p_0)$ is an integer-valued *step-function.* It is constant for a range of p_0; then it jumps to the next bigger integer. Since $k(p_0)$ is a nondecreasing function of p_0, this gives the lower confidence bound. (See Exercise 9.5 for an upper confidence bound.) Solving the inequalities in (9.2.8) for $k(p_0)$ gives both the acceptance region of the test and the confidence set.

For each p_0, the acceptance region is given by $A(p_0) = \{t : t \leq k(p_0)\}$, where $k(p_0)$ satisfies (9.2.8). For each value of t, the confidence set is $C(t) = \{p_0 : t \leq k(p_0)\}$. This set, in its present form, however, does not do us much practical good. Although it is formally correct and a $1-\alpha$ confidence set, it is defined implicitly in terms of p_0 and we want it to be defined explicitly in terms of p_0.

Since $k(p_0)$ is nondecreasing, for a given observation $T = t, k(p_0) < t$ for all p_0 less than or equal to some value, call it $k^{-1}(t)$. At $k^{-1}(t)$, $k(p_0)$ jumps up to equal t and $k(p_0) \geq t$ for all $p_0 > k^{-1}(t)$. (Note that at $p_0 = k^{-1}(t), f(p_0|t-1) = 1-\alpha$. So (9.2.8) is still satisfied by $k(p_0) = t-1$. Only for $p_0 > k^{-1}(t)$ is $k(p_0) \geq t$.) Thus, the confidence set is

$$C(t) = \{p_0 : t \leq k(p_0)\} = \{p_0 : p_0 > k^{-1}(t)\}, \tag{9.2.9}$$

and we have constructed a $1-\alpha$ lower confidence bound of the form $C(T) = (k^{-1}(T), 1]$.

The number $k^{-1}(t)$ can be defined as

$$k^{-1}(t) = \sup\left\{p : \sum_{y=0}^{t-1} \binom{n}{y} p^y (1-p)^{n-y} \geq 1-\alpha\right\}. \tag{9.2.10}$$

Realize that $k^{-1}(t)$ is not really an inverse of $k(p_0)$ because $k(p_0)$ is not a one-to-one function. However, the expressions in (9.2.8) and (9.2.10) give us well-defined quantities for k and k^{-1}.

The problem of binomial confidence bounds was first treated by Clopper and Pearson (1934), who obtained answers similar to these for the two-sided interval (see Exercise 9.21) and started a line of research that is still active today. See Miscellanea 9.5.2. ‖

9.2.2 Pivotal Quantities

The two confidence intervals that we saw in Example 9.1.6 differed in many respects. One important difference was that the coverage probability of the interval $\{aY, bY\}$ did not depend on the value of the parameter θ, while that of $\{Y+c, Y+d\}$ did. This happened because the coverage probability of $\{aY, bY\}$ could be expressed in terms of the quantity Y/θ, a random variable whose distribution does not depend on the parameter, a quantity known as a *pivotal quantity*, or *pivot*.

The use of pivotal quantities for confidence set construction, resulting in what has been called *pivotal inference*, is mainly due to Barnard (1949, 1980) but can be traced as far back as Fisher (1930), who used the term *inverse probability*. Closely related is D. A. S. Fraser's theory of *structural inference* (Fraser 1968, 1979). An interesting discussion of the strengths and weaknesses of these methods is given in Berger and Wolpert (1984).

Definition 9.2.6 A random variable $Q(\mathbf{X}, \theta) = Q(X_1, \ldots, X_n, \theta)$ is a *pivotal quantity* (or *pivot*) if the distribution of $Q(\mathbf{X}, \theta)$ is independent of all parameters. That is, if $\mathbf{X} \sim F(\mathbf{x}|\theta)$, then $Q(\mathbf{X}, \theta)$ has the same distribution for all values of θ.

The function $Q(\mathbf{x}, \theta)$ will usually explicitly contain both parameters and statistics, but for any set $\mathcal{A}$, $P_\theta(Q(\mathbf{X}, \theta) \in \mathcal{A})$ cannot depend on θ. The technique of constructing confidence sets from pivots relies on being able to find a pivot and a set $\mathcal{A}$ so that the set $\{\theta\colon Q(\mathbf{x}, \theta) \in \mathcal{A}\}$ is a set estimate of θ.

Example 9.2.7 (Location-scale pivots) In location and scale cases there are lots of pivotal quantities. We will show a few here; more will be found in Exercise 9.8. Let $X_1, \ldots, X_n$ be a random sample from the indicated pdfs, and let $\bar{X}$ and S be the sample mean and standard deviation. To prove that the quantities in Table 9.2.1 are pivots, we just have to show that their pdfs are independent of parameters (details in Exercise 9.9). Notice that, in particular, if $X_1, \ldots, X_n$ is a random sample from

Table 9.2.1. *Location-scale pivots*

Form of pdf	Type of pdf	Pivotal quantity
$f(x-\mu)$	Location	$\bar{X} - \mu$
$\frac{1}{\sigma} f(\frac{x}{\sigma})$	Scale	$\frac{\bar{X}}{\sigma}$
$\frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$	Location–scale	$\frac{\bar{X}-\mu}{S}$

a $n(\mu, \sigma^2)$ population, then the t statistic $(\bar{X} - \mu)/(S/\sqrt{n})$ is a pivot because the t distribution does not depend on the parameters μ and σ^2. ||

Of the intervals constructed in Section 9.2.1 using the test inversion method, some turned out to be based on pivots (Examples 9.2.3 and 9.2.4) and some did not (Example 9.2.5). There is no all-purpose strategy for finding pivots. However, we can be a little clever and not rely totally on guesswork. For example, it is a relatively easy task to find pivots for location or scale parameters. In general, *differences* are pivotal for location problems, while *ratios* (or products) are pivotal for scale problems.

Example 9.2.8 (Gamma pivot) Suppose that $X_1, \ldots, X_n$ are iid exponential(λ). Then $T = \sum X_i$ is a sufficient statistic for λ and $T \sim \text{gamma}(n, \lambda)$. In the gamma pdf t and λ appear together as t/λ and, in fact the gamma(n, λ) pdf $(\Gamma(n)\lambda^n)^{-1}t^{n-1}e^{-t/\lambda}$ is a scale family. Thus, if $Q(T, \lambda) = 2T/\lambda$, then

$$Q(T, \lambda) \sim \text{gamma}(n, \lambda(2/\lambda)) = \text{gamma}(n, 2),$$

which does not depend on λ. The quantity $Q(T, \lambda) = 2T/\lambda$ is a pivot with a gamma($n, 2$), or χ^2_{2n}, distribution. ||

We can sometimes look to the form of the pdf to see if a pivot exists. In the above example, the quantity t/λ appeared in the pdf and this turned out to be a pivot. In the normal pdf, the quantity $(\bar{x} - \mu)/\sigma$ appears and this quantity is also a pivot. In general, suppose the pdf of a statistic T, $f(t|\theta)$, can be expressed in the form

$$f(t|\theta) = g\left(Q(t, \theta)\right) \left| \frac{\partial}{\partial t} Q(t, \theta) \right| \tag{9.2.11}$$

for some function g and some monotone function Q (monotone in t for each θ). Then Theorem 2.1.5 can be used to show that $Q(T, \theta)$ is a pivot (see Exercise 9.10).

Once we have a pivot, how do we use it to construct a confidence set? That part is really quite simple. If $Q(\mathbf{X}, \theta)$ is a pivot, then for a specified value of α we can find numbers a and b, which do not depend on θ, to satisfy

$$P_\theta(a \leq Q(\mathbf{X}, \theta) \leq b) \geq 1 - \alpha.$$

Then, for each $\theta_0 \in \Theta$,

$$A(\theta_0) = \{\mathbf{x} \colon a \leq Q(\mathbf{x}, \theta_0) \leq b\} \tag{9.2.12}$$

is the acceptance region for a level α test of $H_0\colon \theta = \theta_0$. We will use the test inversion method to construct the confidence set, but we are using the pivot to specify the specific form of our acceptance regions. Using Theorem 9.2.2, we invert these tests to obtain

$$C(\mathbf{x}) = \{\theta_0 \colon a \leq Q(\mathbf{x}, \theta_0) \leq b\}, \tag{9.2.13}$$

and $C(\mathbf{X})$ is a $1 - \alpha$ confidence set for θ. If θ is a real-valued parameter and if, for each $\mathbf{x} \in \mathcal{X}$, $Q(\mathbf{x}, \theta)$ is a monotone function of θ, then $C(\mathbf{x})$ will be an interval. In fact, if

$Q(\mathbf{x},\theta)$ is an increasing function of θ, then $C(\mathbf{x})$ has the form $L(\mathbf{x},a) \le \theta \le U(\mathbf{x},b)$. If $Q(\mathbf{x},\theta)$ is a decreasing function of θ (which is typical), then $C(\mathbf{x})$ has the form $L(\mathbf{x},b) \le \theta \le U(\mathbf{x},a)$.

Example 9.2.9 (Continuation of Example 9.2.8) In Example 9.2.3 we obtained a confidence interval for the mean, λ, of the exponential(λ) pdf by inverting a level α LRT of $H_0\colon \lambda = \lambda_0$ versus $H_1: \lambda \ne \lambda_0$. Now we also see that if we have a sample $X_1,\dots,X_n$, we can define $T = \sum X_i$ and $Q(T,\lambda) = 2T/\lambda \sim \chi^2_{2n}$.

If we choose constants a and b to satisfy $P(a \le \chi^2_{2n} \le b) = 1-\alpha$, then

$$P_\lambda\left(a \le \frac{2T}{\lambda} \le b\right) = P_\lambda(a \le Q(T,\lambda) \le b) = P\big(a \le \chi^2_{2n} \le b\big) = 1-\alpha.$$

Inverting the set $A(\lambda) = \{t\colon a \le \frac{2t}{\lambda} \le b\}$ gives $C(t) = \{\lambda\colon \frac{2t}{b} \le \lambda \le \frac{2t}{a}\}$, which is a $1-\alpha$ confidence interval. (Notice that the lower endpoint depends on b and the upper endpoint depends on a, as mentioned above. $Q(t,\lambda) = 2t/\lambda$ is decreasing in λ.) For example, if $n = 10$, then consulting a table of χ^2 cutoffs shows that a 95% confidence interval is given by $\{\lambda\colon \frac{2T}{34.17} \le \lambda \le \frac{2T}{9.59}\}$. ||

For the location problem, even if the variance is unknown, construction and calculation of pivotal intervals are quite easy. In fact, we have used these ideas already but have not called them by any formal name.

Example 9.2.10 (Normal pivotal interval) It follows from Theorem 5.3.1 that if $X_1,\dots,X_n$ are iid $\mathrm{n}(\mu,\sigma^2)$, then $(\bar{X}-\mu)/(\sigma/\sqrt{n})$ is a pivot. If σ^2 is known, we can use this pivot to calculate a confidence interval for μ. For any constant a,

$$P\left(-a \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le a\right) = P(-a \le Z \le a),\ (Z \text{ is standard normal})$$

and (by now) familiar algebraic manipulations give us the confidence interval

$$\left\{\mu\colon \bar{x} - a\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + a\frac{\sigma}{\sqrt{n}}\right\}.$$

If σ^2 is unknown, we can use the location–scale pivot $\frac{\bar{X}-\mu}{S/\sqrt{n}}$. Since $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ has Student's t distribution,

$$P\left(-a \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le a\right) = P(-a \le T_{n-1} \le a).$$

Thus, for any given α, if we take $a = t_{n-1,\alpha/2}$, we find that a $1-\alpha$ confidence interval is given by

$$\left\{\mu\colon \bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right\}, \tag{9.2.14}$$

which is the classic $1-\alpha$ confidence interval for μ based on Student's t distribution.

Continuing with this case, suppose that we also want an interval estimate for σ. Because $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, $(n-1)S^2/\sigma^2$ is also a pivot. Thus, if we choose a and b to satisfy

$$P\left(a \le \frac{(n-1)S^2}{\sigma^2} \le b\right) = P\left(a \le \chi^2_{n-1} \le b\right) = 1-\alpha,$$

we can invert this set to obtain the $1-\alpha$ confidence interval

$$\left\{\sigma^2 \colon \frac{(n-1)s^2}{b} \le \sigma^2 \le \frac{(n-1)s^2}{a}\right\}$$

or, equivalently,

$$\left\{\sigma \colon \sqrt{\frac{(n-1)s^2}{b}} \le \sigma \le \sqrt{\frac{(n-1)s^2}{a}}\right\}.$$

One choice of a and b that will produce the required interval is $a = \chi^2_{n-1,1-\alpha/2}$ and $b = \chi^2_{n-1,\alpha/2}$. This choice splits the probability equally, putting $\alpha/2$ in each tail of the distribution. The χ^2_{n-1} distribution, however, is a skewed distribution and it is not immediately clear that an equal probability split is optimal for a skewed distribution. (It is not immediately clear that an equal probability split is optimal for a symmetric distribution, but our intuition makes this latter case more plausible.) In fact, for the chi squared distribution, the equal probability split is not optimal, as will be seen in Section 9.3. (See also Exercise 9.52.)

One final note for this problem. We now have constructed confidence intervals for μ and σ separately. It is entirely plausible that we would be interested in a confidence set for μ and σ *simultaneously*. The Bonferroni Inequality is an easy (and relatively good) method for accomplishing this. (See Exercise 9.14.) ‖

9.2.3 Pivoting the CDF

In previous section we saw that a pivot, Q, leads to a confidence set of the form (9.2.13), that is

$$C(\mathbf{x}) = \{\theta_0 \colon a \le Q(\mathbf{x}, \theta_0) \le b\}.$$

If, for every $\mathbf{x}$, the function $Q(\mathbf{x}, \theta)$ is a monotone function of θ, then the confidence set $C(\mathbf{x})$ is guaranteed to be an interval. The pivots that we have seen so far, which were mainly constructed using location and scale transformations, resulted in monotone Q functions and, hence, confidence intervals.

In this section we work with another pivot, one that is totally general and, with minor assumptions, will guarantee an interval.

If in doubt, or in a strange situation, we would recommend constructing a confidence set based on inverting an LRT, if possible. Such a set, although not guaranteed to be optimal, will never be very bad. However, in some cases such a tactic is too difficult, either analytically or computationally; inversion of the acceptance region can sometimes be quite a chore. If the method of this section can be applied, it is rather straightforward to implement and will usually produce a set that is reasonable.

To illustrate the type of trouble that could arise from the test inversion method, without extra conditions on the exact types of acceptance regions used, consider the following example, which illustrates one of the early methods of constructing confidence sets for a binomial success probability.

Example 9.2.11 (Shortest length binomial set) Sterne (1954) proposed the following method for constructing binomial confidence sets, a method that produces a set with the shortest length. Given α, for each value of p find the size α acceptance region composed of the most probable x values. That is, for each p, order the $x = 0, \ldots, n$ values from the most probable to the least probable and put values into the acceptance region $A(p)$ until it has probability $1-\alpha$. Then use (9.2.1) to invert these acceptance regions to get a $1-\alpha$ confidence set, which Sterne claimed had length optimality properties.

To see the unexpected problems with this seemingly reasonable construction, consider a small example. Let $X \sim$ binomial$(3, p)$ and use confidence coefficient $1-\alpha = .442$. Table 9.2.2 gives the acceptance regions obtained by the Sterne construction and the confidence sets derived by inverting this family of tests.

Surprisingly, the confidence set is not a confidence *interval.* This seemingly reasonable construction has led us to an unreasonable procedure. The blame is to be put on the pmf, as it does not behave as we expect. (See Exercise 9.18.) ||

We base our confidence interval construction for a parameter θ on a real-valued statistic T with cdf $F_T(t|\theta)$. (In practice we would usually take T to be a sufficient statistic for θ, but this is not necessary for the following theory to go through.) We will first assume that T is a continuous random variable. The situation where T is discrete is similar but has a few additional technical details to consider. We, therefore, state the discrete case in a separate theorem.

First of all, recall Theorem 2.1.10, the Probability Integral Transformation, which tells us that the random variable $F_T(T|\theta)$ is uniform$(0,1)$, a pivot. Thus, if $\alpha_1+\alpha_2 =$

Table 9.2.2. *Acceptance region and confidence set for Sterne's construction, $X \sim$ binomial$(3,p)$ and $1-\alpha = .442$*

p	Acceptance region $= A(p)$	x	Confidence set $= C(x)$
$[.000, .238]$	$\{0\}$		
$(.238, .305)$	$\{0,1\}$	0	$[.000, .305) \cup (.362, .366)$
$[.305, .362]$	$\{1\}$		
$(.362, .366)$	$\{0,1\}$	1	$(.238, .634]$
$[.366, .634]$	$\{1,2\}$		
$(.634, .638)$	$\{2,3\}$	2	$[.366, .762)$
$[.638, .695]$	$\{2\}$		
$(.695, .762)$	$\{2,3\}$	3	$(.634, .638) \cup (.695, 1.00]$
$[.762, 1.00]$	$\{3\}$		

α, an α-level acceptance region of the hypothesis $H_0 : \theta = \theta_0$ is (see Exercise 9.11)

$$\{t : \alpha_1 \le F_T(t|\theta_0) \le 1 - \alpha_2\},$$

with associated confidence set

$$\{\theta : \alpha_1 \le F_T(t|\theta) \le 1 - \alpha_2\}.$$

Now to guarantee that the confidence set is an interval, we need to have $F_T(t|\theta)$ to be monotone in θ. But we have seen this already, in the definitions of *stochastically increasing* and *stochastically decreasing.* (See the Miscellanea section of Chapter 8 and Exercise 8.26, or Exercises 3.41–3.43.) A family of cdfs $F(t|\theta)$ is *stochastically increasing in* θ (*stochastically decreasing in* θ) if, for each $t \in \mathcal{T}$, the sample space of $T, F(t|\theta)$ is a decreasing (increasing) function of θ. In what follows, we need only the fact that F is monotone, either increasing or decreasing. The more statistical concepts of stochastic increasing or decreasing merely serve as interpretational tools.

Theorem 9.2.12 (Pivoting a continuous cdf) *Let T be a statistic with continuous cdf $F_T(t|\theta)$. Let $\alpha_1 + \alpha_2 = \alpha$ with $0 < \alpha < 1$ be fixed values. Suppose that for each $t \in \mathcal{T}$, the functions $\theta_{\mathrm{L}}(t)$ and $\theta_{\mathrm{U}}(t)$ can be defined as follows.*

i. *If $F_T(t|\theta)$ is a decreasing function of θ for each t, define $\theta_{\mathrm{L}}(t)$ and $\theta_{\mathrm{U}}(t)$ by*

$$F_T(t|\theta_{\mathrm{U}}(t)) = \alpha_1, \quad F_T(t|\theta_{\mathrm{L}}(t)) = 1 - \alpha_2.$$

ii. *If $F_T(t|\theta)$ is an increasing function of θ for each t, define $\theta_{\mathrm{L}}(t)$ and $\theta_{\mathrm{U}}(t)$ by*

$$F_T(t|\theta_{\mathrm{U}}(t)) = 1 - \alpha_2, \quad F_T(t|\theta_{\mathrm{L}}(t)) = \alpha_1.$$

Then the random interval $[\theta_{\mathrm{L}}(T), \theta_{\mathrm{U}}(T)]$ is a $1 - \alpha$ confidence interval for θ.

Proof: We will prove only part (i). The proof of part (ii) is similar and is left as Exercise 9.19.

Assume that we have constructed the $1 - \alpha$ acceptance region

$$\{t : \alpha_1 \le F_T(t|\theta_0) \le 1 - \alpha_2\}.$$

Since $F_T(t|\theta)$ is a decreasing function of θ for each t and $1 - \alpha_2 > \alpha_1$, $\theta_{\mathrm{L}}(t) < \theta_{\mathrm{U}}(t)$, and the values $\theta_{\mathrm{L}}(t)$ and $\theta_{\mathrm{U}}(t)$ are unique. Also,

$$F_T(t|\theta) < \alpha_1 \Leftrightarrow \theta > \theta_{\mathrm{U}}(t),$$
$$F_T(t|\theta) > 1 - \alpha_2 \Leftrightarrow \theta < \theta_{\mathrm{L}}(t),$$

and hence $\{\theta : \alpha_1 \le F_T(t|\theta) \le 1 - \alpha_2\} = \{\theta : \theta_{\mathrm{L}}(T) \le \theta \le \theta_{\mathrm{U}}(T)\}$. □

We note that, in the absence of additional information, it is common to choose $\alpha_1 = \alpha_2 = \alpha/2$. Although this may not always be optimal (see Theorem 9.3.2), it is certainly a reasonable strategy in most situations. If a one-sided interval is desired, however, this can easily be achieved by choosing either α_1 or α_2 equal to 0.

The equations for the stochastically increasing case,

$$F_T(t|\theta_{\mathrm{U}}(t)) = \alpha_1, \quad F_T(t|\theta_{\mathrm{L}}(t)) = 1 - \alpha_2, \tag{9.2.15}$$

can also be expressed in terms of the pdf of the statistic T. The functions $\theta_{\mathrm{U}}(t)$ and $\theta_{\mathrm{L}}(t)$ can be defined to satisfy

$$\int_{-\infty}^{t} f_T(u|\theta_{\mathrm{U}}(t))\,du = \alpha_1 \quad \text{and} \quad \int_{t}^{\infty} f_T(u|\theta_{\mathrm{L}}(t))\,du = \alpha_2.$$

A similar set of equations holds for the stochastically decreasing case.

Example 9.2.13 (Location exponential interval) This method can be used to get a confidence interval for the location exponential pdf. (In Exercise 9.25 the answer here is compared to that obtained by likelihood and pivotal methods. See also Exercise 9.41.)

If $X_1, \ldots, X_n$ are iid with pdf $f(x|\mu) = e^{-(x-\mu)} I_{[\mu,\infty)}(x)$, then $Y = \min\{X_1, \ldots, X_n\}$ is sufficient for μ with pdf

$$f_Y(y|\mu) = n e^{-n(y-\mu)} I_{[\mu,\infty)}(y).$$

Fix α and define $\mu_{\mathrm{L}}(y)$ and $\mu_{\mathrm{U}}(y)$ to satisfy

$$\int_{\mu_{\mathrm{U}}(y)}^{y} n e^{-n(u-\mu_{\mathrm{U}}(y))}\,du = \frac{\alpha}{2}, \quad \int_{y}^{\infty} n e^{-n(u-\mu_{\mathrm{L}}(y))}\,du = \frac{\alpha}{2}.$$

These integrals can be evaluated to give the equations

$$1 - e^{-n(y-\mu_{\mathrm{U}}(y))} = \frac{\alpha}{2}, \quad e^{-n(y-\mu_{\mathrm{L}}(y))} = \frac{\alpha}{2},$$

which give us the solutions

$$\mu_{\mathrm{U}}(y) = y + \frac{1}{n}\log\left(1 - \frac{\alpha}{2}\right), \quad \mu_{\mathrm{L}}(y) = y + \frac{1}{n}\log\left(\frac{\alpha}{2}\right).$$

Hence, the random interval

$$C(Y) = \left\{\mu \colon Y + \frac{1}{n}\log\left(\frac{\alpha}{2}\right) \le \mu \le Y + \frac{1}{n}\log\left(1 - \frac{\alpha}{2}\right)\right\},$$

a $1-\alpha$ confidence interval for μ. ||

Note two things about the use of this method. First, the actual equations (9.2.15) need to be solved only for the value of the statistics actually observed. If $T = t_0$ is observed, then the realized confidence interval on θ will be $[\theta_{\mathrm{L}}(t_0), \theta_{\mathrm{U}}(t_0)]$. Thus, we need to solve only the two equations

$$\int_{-\infty}^{t_0} f_T(u|\theta_{\mathrm{U}}(t_0))\,du = \alpha_1 \quad \text{and} \quad \int_{t_0}^{\infty} f_T(u|\theta_{\mathrm{L}}(t_0))\,du = \alpha_2$$

for $\theta_{\mathrm{L}}(t_0)$ and $\theta_{\mathrm{U}}(t_0)$. Second, realize that even if these equations cannot be solved analytically, we really only need to solve them numerically since the proof that we have a $1-\alpha$ confidence interval did not require an analytic solution.

We now consider the discrete case.

Theorem 9.2.14 (Pivoting a discrete cdf) *Let T be a discrete statistic with cdf $F_T(t|\theta) = P(T \le t|\theta)$. Let $\alpha_1 + \alpha_2 = \alpha$ with $0 < \alpha < 1$ be fixed values. Suppose that for each $t \in \mathcal{T}$, $\theta_L(t)$ and $\theta_U(t)$ can be defined as follows.*

i. *If $F_T(t|\theta)$ is a decreasing function of θ for each t, define $\theta_L(t)$ and $\theta_U(t)$ by*

$$P(T \le t|\theta_U(t)) = \alpha_1, \quad P(T \ge t|\theta_L(t)) = \alpha_2.$$

ii. *If $F_T(t|\theta)$ is an increasing function of θ for each t, define $\theta_L(t)$ and $\theta_U(t)$ by*

$$P(T \ge t|\theta_U(t)) = \alpha_1, \quad P(T \le t|\theta_L(t)) = \alpha_2.$$

Then the random interval $[\theta_L(T), \theta_U(T)]$ is a $1 - \alpha$ confidence interval for θ.

Proof: We will only sketch the proof of part (i). The details, as well as the proof of part (ii), are left to Exercise 9.20.

First recall Exercise 2.10, where it was shown that $F_T(T|\theta)$ is stochastically greater than a uniform random variable, that is, $P_\theta(F_T(T|\theta) \le x) \le x$. Furthermore, this property is shared by $\bar{F}_T(T|\theta) = P(T \ge t|\theta)$, and this implies that the set

$$\{\theta : F_T(T|\theta) \le \alpha_1 \text{ and } \bar{F}_T(T|\theta) \le \alpha_2\}$$

is a $1 - \alpha$ confidence set.

The fact that $F_T(t|\theta)$ is a decreasing function of θ for each t implies that $\bar{F}(t|\theta)$ is a nondecreasing function of θ for each t. It therefore follows that

$$\theta > \theta_U(t) \Rightarrow F_T(t|\theta) < \tfrac{\alpha}{2},$$
$$\theta < \theta_L(t) \Rightarrow \bar{F}_T(t|\theta) < \tfrac{\alpha}{2},$$

and hence $\{\theta : F_T(T|\theta) \le \alpha_1 \text{ and } \bar{F}_T(T|\theta) \le \alpha_2\} = \{\theta : \theta_L(T) \le \theta \le \theta_U(T)\}$. □

We close this section with an example to illustrate the construction of Theorem 9.2.14. Notice that an alternative interval can be constructed by inverting an LRT (see Exercise 9.23).

Example 9.2.15 (Poisson interval estimator) Let $X_1, \ldots, X_n$ be a random sample from a Poisson population with parameter λ and define $Y = \sum X_i$. Y is sufficient for λ and $Y \sim \text{Poisson}(n\lambda)$. Applying the above method with $\alpha_1 = \alpha_2 = \alpha/2$, if $Y = y_0$ is observed, we are led to solve for λ in the equations

$$\sum_{k=0}^{y_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = \frac{\alpha}{2} \quad \text{and} \quad \sum_{k=y_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = \frac{\alpha}{2}. \tag{9.2.16}$$

Recall the identity, from Example 3.3.1, linking the Poisson and gamma families. Applying that identity to the sums in (9.2.16), we can write (remembering that y_0 is the *observed* value of Y)

$$\frac{\alpha}{2} = \sum_{k=0}^{y_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = P(Y \le y_0|\lambda) = P\left(\chi^2_{2(y_0+1)} > 2n\lambda\right),$$

where $\chi^2_{2(y_0+1)}$ is a chi squared random variable with $2(y_0+1)$ degrees of freedom. Thus, the solution to the above equation is to take

$$\lambda = \frac{1}{2n}\chi^2_{2(y_0+1),\alpha/2}.$$

Similarly, applying the identity to the other equation in (9.2.16) yields

$$\frac{\alpha}{2} = \sum_{k=y_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = P(Y \geq y_0|\lambda) = P\left(\chi^2_{2y_0} < 2n\lambda\right).$$

Doing some algebra, we obtain the $1-\alpha$ confidence interval for λ as

(9.2.17) $$\left\{\lambda\colon \frac{1}{2n}\chi^2_{2y_0,1-\alpha/2} \leq \lambda \leq \frac{1}{2n}\chi^2_{2(y_0+1),\alpha/2}\right\}.$$

(At $y_0 = 0$ we define $\chi^2_{0,1-\alpha/2} = 0$.)

These intervals were first derived by Garwood (1936). A graph of the coverage probabilities is given in Figure 9.2.5. Notice that the graph is quite jagged. The jumps occur at the endpoints of the different confidence intervals, where terms are added or subtracted from the sum that makes up the coverage probability. (See Exercise 9.24.)

For a numerical example, consider $n = 10$ and observe $y_0 = \sum x_i = 6$. A 90% confidence interval for λ is given by

$$\frac{1}{20}\chi^2_{12,.95} \leq \lambda \leq \frac{1}{20}\chi^2_{14,.05},$$

which is

$$.262 \leq \lambda \leq 1.184.$$

Similar derivations, involving the negative binomial and binomial distributions, are given in the exercises. ||

9.2.4 Bayesian Intervals

Thus far, when describing the interactions between the confidence interval and the parameter, we have carefully said that the interval *covers* the parameter, not that the parameter *is inside* the interval. This was done on purpose. We wanted to stress that the random quantity is the interval, not the parameter. Therefore, we tried to make the action verbs apply to the interval and not the parameter.

In Example 9.2.15 we saw that if $y_0 = \sum_{i=1}^{10} x_i = 6$, then a 90% confidence interval for λ is $.262 \leq \lambda \leq 1.184$. It is tempting to say (and many experimenters do) that "the probability is 90% that λ is in the interval [.262, 1.184]." Within classical statistics, however, such a statement is invalid since the parameter is assumed fixed. Formally, the interval [.262, 1.184] is one of the possible *realized values* of the random interval $[\frac{1}{2n}\chi^2_{2Y,.95}, \frac{1}{2n}\chi^2_{2(Y+1),.05}]$ and, since the parameter λ does not move, λ is in the *realized interval* [.262, 1.184] with probability either 0 or 1. When we say that the realized interval [.262, 1.184] has a 90% chance of coverage, we only mean that we know that 90% of the sample points of the random interval cover the true parameter.

In contrast, the Bayesian setup allows us to say that λ *is inside* [.262, 1.184] with some probability, not 0 or 1. This is because, under the Bayesian model, λ is a random variable with a probability distribution. All Bayesian claims of coverage are made with respect to the posterior distribution of the parameter.

To keep the distinction between Bayesian and classical sets clear, since the sets make quite different probability assessments, the Bayesian set estimates are referred to as *credible sets* rather than confidence sets.

Thus, if $\pi(\theta|\mathbf{x})$ is the posterior distribution of θ given $\mathbf{X} = \mathbf{x}$, then for any set $A \subset \Theta$, the credible probability of A is

$$P(\theta \in A|\mathbf{x}) = \int_A \pi(\theta|\mathbf{x})\, d\theta, \tag{9.2.18}$$

and A is a *credible set* for θ. If $\pi(\theta|\mathbf{x})$ is a pmf, we replace integrals with sums in the above expressions.

Notice that both the interpretation and construction of the Bayes credible set are more straightforward than those of a classical confidence set. However, remember that nothing comes free. The ease of construction and interpretation comes with additional assumptions. The Bayesian model requires more input than the classical model.

Example 9.2.16 (Poisson credible set) We now construct a credible set for the problem of Example 9.2.15. Let $X_1, \ldots, X_n$ be iid Poisson(λ) and assume that λ has a gamma prior pdf, $\lambda \sim \text{gamma}(a, b)$. The posterior pdf of λ (see Exercise 7.24) is

$$\pi(\lambda|\textstyle\sum X = \sum x) = \text{gamma}(a + \sum x, [n + (1/b)]^{-1}). \tag{9.2.19}$$

We can form a credible set for λ in many different ways, as any set A satisfying (9.2.18) will do. One simple way is to split the α equally between the upper and lower endpoints. From (9.2.19) it follows that $\frac{2(nb+1)}{b}\lambda \sim \chi^2_{2(a+\Sigma x_i)}$ (assuming that a is an integer), and thus a $1 - \alpha$ credible interval is

$$\left\{\lambda\colon \frac{b}{2(nb+1)}\chi^2_{2(\Sigma x+a),1-\alpha/2} \le \lambda \le \frac{b}{2(nb+1)}\chi^2_{2(\Sigma x+a),\alpha/2}\right\}. \tag{9.2.20}$$

If we take $a = b = 1$, the posterior distribution of λ given $\sum X = \sum x$ can then be expressed as $2(n+1)\lambda \sim \chi^2_{2(\Sigma x+1)}$. As in Example 9.2.15, assume $n = 10$ and $\sum x = 6$. Since $\chi^2_{14,.95} = 6.571$ and $\chi^2_{14,.05} = 23.685$, a 90% credible set for λ is given by [.299, 1.077].

The realized 90% credible set is different from the 90% confidence set obtained in Example 9.2.15, [.262, 1.184]. To better see the differences, look at Figure 9.2.3, which shows the 90% credible intervals and 90% confidence intervals for a range of x values. Notice that the credible set has somewhat shorter intervals, and the upper endpoints are closer to 0. This reflects the prior, which is pulling the intervals toward 0. ||

It is important not to confuse credible probability (the Bayes posterior probability) with coverage probability (the classical probability). The probabilities are very different entities, with different meanings and interpretations. Credible probability comes

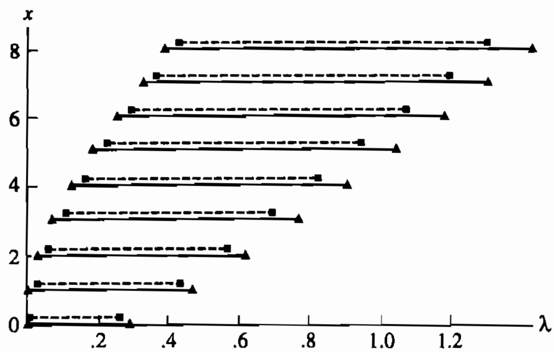

Figure 9.2.3. *The 90% credible intervals (dashed lines) and 90% confidence intervals (solid lines) from Example 9.2.16*

from the posterior distribution, which in turn gets its probability from the prior distribution. Thus, credible probability reflects the experimenter's subjective beliefs, as expressed in the prior distribution and updated with the data to the posterior distribution. A Bayesian assertion of 90% coverage means that the experimenter, upon combining prior knowledge with data, is 90% sure of coverage.

Coverage probability, on the other hand, reflects the uncertainty in the sampling procedure, getting its probability from the objective mechanism of repeated experimental trials. A classical assertion of 90% coverage means that in a long sequence of identical trials, 90% of the realized confidence sets will cover the true parameter.

Statisticians sometimes argue as to which is the better way to do statistics, classical or Bayesian. We do not want to argue or even defend one over another. In fact, we believe that there is no one best way to do statistics; some problems are best solved with classical statistics and some are best solved with Bayesian statistics. The important point to realize is that the solutions may be quite different. A Bayes solution is often not reasonable under classical evaluations and vice versa.

Example 9.2.17 (Poisson credible and coverage probabilities) The 90% confidence and credible sets of Example 9.2.16 maintain their respective probability guarantees, but how do they fare under the other criteria? First, lets look at the credible probability of the confidence set (9.2.17), which is given by

$$(9.2.21) \qquad P\left\{\frac{1}{2n}\chi^2_{2\Sigma x,1-\alpha/2} \le \lambda \le \frac{1}{2n}\chi^2_{2(\Sigma x+1),\alpha/2}\right\},$$

where λ has the distribution (9.2.19). Figure 9.2.4 shows the credible probability of the set (9.2.20), which is constant at $1-\alpha$, along with the credible probability of the confidence set (9.2.21).

This latter probability seems to be steadily decreasing, and we want to know if it remains above 0 for all values of Σx_i (for each fixed n). To do this, we evaluate the probability as $\Sigma x_i \to \infty$. Details are left to Exercise 9.30, but it is the case that, as

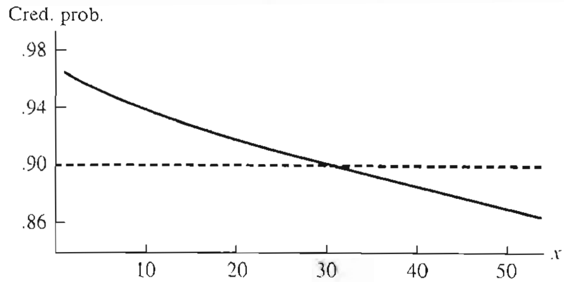

Figure 9.2.4. *Credible probabilities of the 90% credible intervals (dashed line) and 90% confidence intervals (solid line) from Example 9.2.16*

$\Sigma x_i \to \infty$, the probability (9.2.21) $\to 0$ unless $b = 1/n$. Thus, the confidence interval cannot maintain a nonzero credible probability.

The credible set (9.2.20) does not fare much better when evaluated as a confidence set. Figure 9.2.5 suggests that the coverage probability of the credible set is going to 0 as $\lambda \to \infty$. To evaluate the coverage probability, write

$$\lambda = \frac{\lambda}{\chi^2_{2Y}} \chi^2_{2Y},$$

where χ^2_{2Y} is a chi squared random variable with $2Y$ degrees of freedom, and $Y \sim$ Poisson $(n\lambda)$. Then, as $\lambda \to \infty$, $\lambda/\chi^2_{2Y} \to 1/(2n)$, and the coverage probability of (9.2.20) becomes

$$P\left(\frac{nb}{nb+1}\chi^2_{2(Y+a),1-\alpha/2} \le \chi^2_{2Y} \le \frac{nb}{nb+1}\chi^2_{2(Y+a),\alpha/2}\right). \tag{9.2.22}$$

That this probability goes to 0 as $\lambda \to \infty$ is established in Exercise 9.31. ||

The behavior exhibited in Example 9.2.17 is somewhat typical. Here is an example where the calculations can be done explicitly.

Example 9.2.18 (Coverage of a normal credible set) Let $X_1, \dots, X_n$ be iid $\mathrm{n}(\theta, \sigma^2)$, and let θ have the prior pdf $\mathrm{n}(\mu, \tau^2)$, where μ, σ, and τ are all known. In Example 7.2.16 we saw that

$$\pi(\theta|\bar{x}) \sim \mathrm{n}(\delta^{\mathrm{B}}(\bar{x}), \mathrm{Var}(\theta|\bar{x})),$$

where

$$\delta^{\mathrm{B}}(x) = \frac{\sigma^2}{\sigma^2 + n\tau^2}\mu + \frac{n\tau^2}{\sigma^2 + n\tau^2}\bar{x} \quad \text{and} \quad \mathrm{Var}(\theta|\bar{x}) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}.$$

It therefore follows that under the posterior distribution,

$$\frac{\theta - \delta^{\mathrm{B}}(\bar{x})}{\sqrt{\mathrm{Var}(\theta|\bar{x})}} \sim \mathrm{n}(0, 1),$$

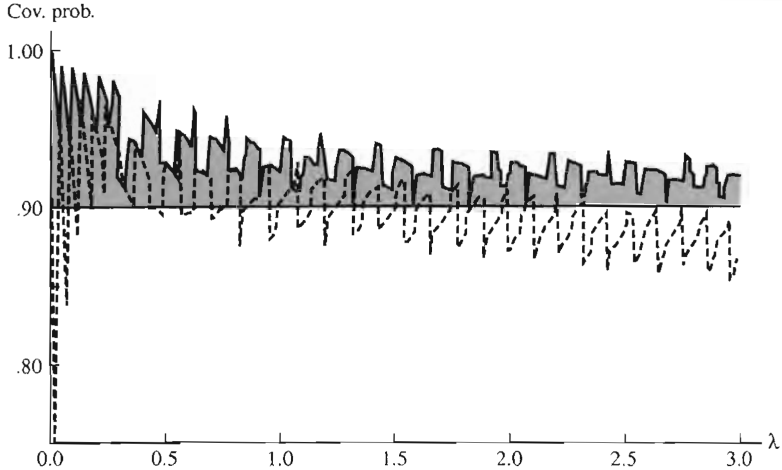

Figure 9.2.5. *Coverage probabilities of the 90% credible intervals (dashed lines) and 90% confidence intervals (solid lines) from Example 9.2.16*

and a $1-\alpha$ credible set for θ is given by

$$\delta^{\mathrm{B}}(\bar{x}) - z_{\alpha/2}\sqrt{\mathrm{Var}(\theta|\bar{x})} \le \theta \le \delta^{\mathrm{B}}(\bar{x}) + z_{\alpha/2}\sqrt{\mathrm{Var}(\theta|\bar{x})}. \tag{9.2.23}$$

We now calculate the coverage probability of the Bayesian region (9.2.23). Under the classical model $\bar{X}$ is the random variable, θ is fixed, and $\bar{X} \sim \mathrm{n}(\theta, \sigma^2/n)$. For ease of notation define $\gamma = \sigma^2/(n\tau^2)$, and from the definitions of $\delta^{\mathrm{B}}(\bar{X})$ and $\mathrm{Var}(\theta|\bar{X})$ and a little algebra, the coverage probability of (9.2.23) is

$$\begin{aligned}
&P_\theta\left(|\theta - \delta^{\mathrm{B}}(\bar{X})| \le z_{\alpha/2}\sqrt{\mathrm{Var}(\theta|\bar{X})}\right)\\
&\quad = P_\theta\left(\left|\theta - \left(\frac{\gamma}{1+\gamma}\mu + \frac{1}{1+\gamma}\bar{X}\right)\right| \le z_{\alpha/2}\sqrt{\frac{\sigma^2}{n(1+\gamma)}}\right)\\
&\quad = P_\theta\left(-\sqrt{1+\gamma}\,z_{\alpha/2} + \frac{\gamma(\theta-\mu)}{\sigma/\sqrt{n}} \le Z \le \sqrt{1+\gamma}\,z_{\alpha/2} + \frac{\gamma(\theta-\mu)}{\sigma/\sqrt{n}}\right),
\end{aligned}$$

where the last equality used the fact that $\sqrt{n}(\bar{X}-\theta)/\sigma = Z \sim \mathrm{n}(0,1)$.

Although we started with a $1-\alpha$ credible set, we do not have a $1-\alpha$ confidence set, as can be seen by considering the following parameter configuration. Fix $\theta \ne \mu$ and let $\tau = \sigma/\sqrt{n}$, so that $\gamma = 1$. Also, let $\sigma/\sqrt{n}$ be very small ($\to 0$). Then it is easy to see that the above probability goes to 0, since if $\theta > \mu$ the lower bound goes to infinity, and if $\theta < \mu$ the upper bound goes to $-\infty$. If $\theta = \mu$, however, the coverage probability is bounded away from 0.

On the other hand, the usual $1-\alpha$ confidence set for θ is $\{\theta\colon |\theta-\bar{x}| \le z_{\alpha/2}\sigma/\sqrt{n}\}$. The credible probability of this set (now $\theta \sim \pi(\theta|\bar{x})$) is given by

$$\begin{aligned}
&P_{\bar{x}}\left(|\theta-\bar{x}| \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\\
&\quad = P_{\bar{x}}\left(\left|[\theta-\delta^{\mathrm{B}}(\bar{x})] + [\delta^{\mathrm{B}}(\bar{x})-\bar{x}]\right| \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\\
&\quad = P_{\bar{x}}\left(-\sqrt{1+\gamma}z_{\alpha/2} + \frac{\gamma(\bar{x}-\mu)}{\sqrt{1+\gamma}\sigma/\sqrt{n}} \le Z \le \sqrt{1+\gamma}z_{\alpha/2} + \frac{\gamma(\bar{x}-\mu)}{\sqrt{1+\gamma}\sigma/\sqrt{n}}\right),
\end{aligned}$$

where the last equality used the fact that $(\theta-\delta^{\mathrm{B}}(\bar{x}))/\sqrt{\mathrm{Var}(\theta|\bar{x})} = Z \sim \mathrm{n}(0,1)$. Again, it is fairly easy to show that this probability is not bounded away from 0, showing that the confidence set is also not, in general, a credible set. Details are in Exercise 9.32. ||

9.3 Methods of Evaluating Interval Estimators

We now have seen many methods for deriving confidence sets and, in fact, we can derive different confidence sets for the same problem. In such situations we would, of course, want to choose a best one. Therefore, we now examine some methods and criteria for evaluating set estimators.

In set estimation two quantities vie against each other, size and coverage probability. Naturally, we want our set to have small size and large coverage probability, but such sets are usually difficult to construct. (Clearly, we can have a large coverage probability by increasing the size of our set. The interval $(-\infty,\infty)$ has coverage probability 1!) Before we can optimize a set with respect to size and coverage probability, we must decide how to measure these quantities.

The coverage probability of a confidence set will, except in special cases, be a function of the parameter, so there is not one value to consider but an infinite number of values. For the most part, however, we will measure coverage probability performance by the *confidence coefficient*, the infimum of the coverage probabilities. This is one way, but not the only available way of summarizing the coverage probability information. (For example, we could calculate an average coverage probability.)

When we speak of the *size* of a confidence set we will usually mean the *length* of the confidence set, if the set is an interval. If the set is not an interval, or if we are dealing with a multidimensional set, then length will usually become *volume*. (There are also cases where a size measure other than length is natural, especially if equivariance is a consideration. This topic is treated by Schervish 1995, Chapter 6, and Berger 1985, Chapter 6.)

9.3.1 Size and Coverage Probability

We now consider what appears to be a simple, constrained minimization problem. For a given, specified coverage probability find the confidence interval with the shortest length. We first consider an example.

Example 9.3.1 (Optimizing length) Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, where σ is known. From the method of Section 9.2.2 and the fact that

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is a pivot with a standard normal distribution, any a and b that satisfy

$$P(a \le Z \le b) = 1 - \alpha$$

will give the $1 - \alpha$ confidence interval

$$\left\{ \mu : \bar{x} - b\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} - a\frac{\sigma}{\sqrt{n}} \right\}.$$

Which choice of a and b is best? More formally, what choice of a and b will minimize the length of the confidence interval while maintaining $1 - \alpha$ coverage? Notice that the length of the confidence interval is equal to $(b - a)\sigma/\sqrt{n}$ but, since the factor $\sigma/\sqrt{n}$ is part of each interval length, it can be ignored and length comparisons can be based on the value of $b - a$. Thus, we want to find a pair of numbers a and b that satisfy $P(a \le Z \le b) = 1 - \alpha$ and minimize $b - a$.

In Example 9.2.1 we took $a = -z_{\alpha/2}$ and $b = z_{\alpha/2}$, but no mention was made of optimality. If we take $1 - \alpha = .90$, then any of the following pairs of numbers give 90% intervals:

Three 90% normal confidence intervals

a	b	Probability		$b - a$
-1.34	2.33	$P(Z < a) = .09$,	$P(Z > b) = .01$	3.67
-1.44	1.96	$P(Z < a) = .075$,	$P(Z > b) = .025$	3.40
-1.65	1.65	$P(Z < a) = .05$,	$P(Z > b) = .05$	3.30

This numerical study suggests that the choice $a = -1.65$ and $b = 1.65$ gives the best interval and, in fact, it does. In this case splitting the probability α equally is an optimal strategy. ||

The strategy of splitting α equally, which is optimal in the above case, is not always optimal. What makes the equal α split optimal in the above case is the fact that the height of the pdf is the same at $-z_{\alpha/2}$ and $z_{\alpha/2}$. We now prove a theorem that will demonstrate this fact, a theorem that is applicable in some generality, needing only the assumption that the pdf is unimodal. Recall the definition of unimodal: A pdf $f(x)$ is *unimodal* if there exists x^* such that $f(x)$ is nondecreasing for $x \le x^*$ and $f(x)$ is nonincreasing for $x \ge x^*$. (This is a rather weak requirement.)

Theorem 9.3.2 *Let $f(x)$ be a unimodal pdf. If the interval $[a, b]$ satisfies*

i. $\int_a^b f(x)\,dx = 1 - \alpha$,

ii. $f(a) = f(b) > 0$, *and*

iii. *$a \le x^* \le b$, where x^* is a mode of $f(x)$,*
then $[a, b]$ is the shortest among all intervals that satisfy (i).

Proof: Let $[a', b']$ be any interval with $b' - a' < b - a$. We will show that this implies $\int_{a'}^{b'} f(x)\,dx < 1 - \alpha$. The result will be proved only for $a' \le a$, the proof being similar if $a < a'$. Also, two cases need to be considered, $b' \le a$ and $b' > a$.

If $b' \le a$, then $a' \le b' \le a \le x^*$ and

$$\begin{aligned}
\int_{a'}^{b'} f(x)\,dx &\le f(b')(b' - a') && (x \le b' \le x^* \Rightarrow f(x) \le f(b')) \\
&\le f(a)(b' - a') && (b' \le a \le x^* \Rightarrow f(b') \le f(a)) \\
&< f(a)(b - a) && (b' - a' < b - a \text{ and } f(a) > 0) \\
&\le \int_a^b f(x)\,dx && \left(\begin{array}{c}\text{(ii), (iii), and unimodality} \\ \Rightarrow f(x) \ge f(a) \text{ for } a \le x \le b\end{array}\right) \\
&= 1 - \alpha, && \text{(i)}
\end{aligned}$$

completing the proof in the first case.

If $b' > a$, then $a' \le a < b' < b$ for, if b' were greater than or equal to b, then $b' - a'$ would be greater than or equal to $b - a$. In this case, we can write

$$\begin{aligned}
\int_{a'}^{b'} f(x)\,dx &= \int_a^b f(x)\,dx + \left[\int_{a'}^{a} f(x)\,dx - \int_{b'}^{b} f(x)\,dx\right] \\
&= (1 - \alpha) + \left[\int_{a'}^{a} f(x)\,dx - \int_{b'}^{b} f(x)\,dx\right],
\end{aligned}$$

and the theorem will be proved if we show that the expression in square brackets is negative. Now, using the unimodality of f, the ordering $a' \le a < b' < b$, and (ii), we have

$$\int_{a'}^{a} f(x)\,dx \le f(a)(a - a')$$

and

$$\int_{b'}^{b} f(x)\,dx \ge f(b)(b - b').$$

Thus,

$$\begin{aligned}
\int_{a'}^{a} f(x)\,dx - \int_{b'}^{b} f(x)\,dx &\le f(a)(a - a') - f(b)(b - b') \\
&= f(a)\left[(a - a') - (b - b')\right] \qquad (f(a) = f(b)) \\
&= f(a)\left[(b' - a') - (b - a)\right],
\end{aligned}$$

which is negative if $(b' - a') < (b - a)$ and $f(a) > 0$. $\square$

If we are willing to put more assumptions on f, for instance, that f is continuous, then we can simplify the proof of Theorem 9.3.2. (See Exercise 9.38.)

Recall the discussion after Example 9.2.3 about the form of likelihood regions, which we now see is an optimal construction by Theorem 9.3.2. A similar argument, given in Corollary 9.3.10, shows how this construction yields an optimal Bayesian region. Also, we can see now that the equal α split, which is optimal in Example 9.3.1, will be optimal for any symmetric unimodal pdf (see Exercise 9.39). Theorem 9.3.2 may even apply when the optimality criterion is somewhat different from minimum length.

Example 9.3.3 (Optimizing expected length) For normal intervals based on the pivot $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ we know that the shortest length $1-\alpha$ confidence interval of the form

$$\bar{x} - b\frac{s}{\sqrt{n}} \le \mu \le \bar{x} - a\frac{s}{\sqrt{n}}$$

has $a = -t_{n-1,\alpha/2}$ and $b = t_{n-1,\alpha/2}$. The interval length is a function of s, with general form

$$\text{Length}(s) = (b-a)\frac{s}{\sqrt{n}}.$$

It is easy to see that if we had considered the criterion of *expected length* and wanted to find a $1-\alpha$ interval to minimize

$$\mathrm{E}_\sigma(\text{Length}(S)) = (b-a)\frac{\mathrm{E}_\sigma S}{\sqrt{n}} = (b-a)c(n)\frac{\sigma}{\sqrt{n}},$$

then Theorem 9.3.2 applies and the choice $a = -t_{n-1,\alpha/2}$ and $b = t_{n-1,\alpha/2}$ again gives the optimal interval. (The quantity $c(n)$ is a constant dependent only on n. See Exercise 7.50.) ||

In some cases, especially when working outside of the location problem, we must be careful in the application of Theorem 9.3.2. In scale cases in particular, the theorem may not be directly applicable, but a variant may be.

Example 9.3.4 (Shortest pivotal interval) Suppose $X \sim \text{gamma}(k, \beta)$. The quantity $Y = X/\beta$ is a pivot, with $Y \sim \text{gamma}(k, 1)$, so we can get a confidence interval by finding constants a and b to satisfy

$$P(a \le Y \le b) = 1 - \alpha. \tag{9.3.1}$$

However, blind application of Theorem 9.3.2 will not give the shortest confidence interval. That is, choosing a and b to satisfy (9.3.1) and also $f_Y(a) = f_Y(b)$ is not optimal. This is because, based on (9.3.1), the interval on β is of the form

$$\left\{\beta : \frac{x}{b} \le \beta \le \frac{x}{a}\right\},$$

so the length of the interval is $(\frac{1}{a} - \frac{1}{b})x$; that is, it is proportional to $(1/a) - (1/b)$ and not to $b - a$.

Although Theorem 9.3.2 is not directly applicable here, a modified argument can solve this problem. Condition (a) in Theorem 9.3.2 defines b as a function of a, say $b(a)$. We must solve the following constrained minimization problem:

$$\textit{Minimize, with respect to } a\textit{:}\quad \tfrac{1}{a} - \tfrac{1}{b(a)}$$

$$\textit{subject to:}\quad \int_a^{b(a)} f_Y(y)\,dy = 1-\alpha.$$

Differentiating the first equation with respect to a and setting it equal to 0 yield the identity $db/da = b^2/a^2$. Substituting this in the derivative of the second equation, which must equal 0, gives $f(b)b^2 = f(a)a^2$ (see Exercise 9.42). Equations like these also arise in interval estimation of the variance of a normal distribution; see Example 9.2.10 and Exercise 9.52. Note that the above equations define not the shortest *overall* interval, but the shortest *pivotal* interval, that is, the shortest interval based on the pivot X/β. For a generalization of this result, involving the Neyman-Pearson Lemma, see Exercise 9.43. ||

9.3.2 Test-Related Optimality

Since there is a one-to-one correspondence between confidence sets and tests of hypotheses (Theorem 9.2.2), there is some correspondence between optimality of tests and optimality of confidence sets. Usually, test-related optimality properties of confidence sets do not directly relate to the size of the set but rather to the probability of the set covering false values.

The probability of covering false values, or the *probability of false coverage*, indirectly measures the size of a confidence set. Intuitively, smaller sets cover fewer values and, hence, are less likely to cover false values. Moreover, we will later see an equation that links size and probability of false coverage.

We first consider the general situation, where $\mathbf{X} \sim f(\mathbf{x}|\theta)$, and we construct a $1-\alpha$ confidence set for θ, $C(\mathbf{x})$, by inverting an acceptance region, $A(\theta)$. The probability of coverage of $C(\mathbf{x})$, that is, the probability of *true coverage*, is the function of θ given by $P_\theta(\theta \in C(\mathbf{X}))$. The probability of *false coverage* is the function of θ and θ' defined by

$$\begin{aligned}
&P_\theta\left(\theta' \in C(\mathbf{X})\right), \theta \neq \theta', &&\text{if } C(\mathbf{X}) = [L(\mathbf{X}), U(\mathbf{X})],\\
&P_\theta\left(\theta' \in C(\mathbf{X})\right), \theta' < \theta, &&\text{if } C(\mathbf{X}) = [L(\mathbf{X}), \infty),\\
&P_\theta\left(\theta' \in C(\mathbf{X})\right), \theta' > \theta, &&\text{if } C(\mathbf{X}) = (-\infty, U(\mathbf{X})],
\end{aligned} \tag{9.3.2}$$

the probability of covering θ' when θ is the true parameter.

It makes sense to define the probability of false coverage differently for one-sided and two-sided intervals. For example, if we have a lower confidence bound, we are asserting that θ is greater than a certain value and false coverage would occur only if we cover values of θ that are too small. A similar argument leads us to the definitions used for upper confidence bounds and two-sided bounds.

A $1-\alpha$ confidence set that minimizes the probability of false coverage over a class of $1-\alpha$ confidence sets is called a *uniformly most accurate* (UMA) confidence set. Thus, for example, we would consider looking for a UMA confidence set among sets of the form $[L(\mathbf{x}), \infty)$. UMA confidence sets are constructed by inverting the acceptance regions of UMP tests, as we will prove below. Unfortunately, although a UMA confidence set is a desirable set, it exists only in rather rare circumstances (as do UMP tests). In particular, since UMP tests are generally one-sided, so are UMA intervals. They make for elegant theory, however. In the next theorem we see that a UMP test of $H_0: \theta = \theta_0$ versus $H_1: \theta > \theta_0$ yields a UMA lower confidence bound.

Theorem 9.3.5 *Let $\mathbf{X} \sim f(\mathbf{x}|\theta)$, where θ is a real-valued parameter. For each $\theta_0 \in \Theta$, let $A^*(\theta_0)$ be the UMP level α acceptance region of a test of H_0: $\theta = \theta_0$ versus H_1: $\theta > \theta_0$. Let $C^*(\mathbf{x})$ be the $1-\alpha$ confidence set formed by inverting the UMP acceptance regions. Then for any other $1-\alpha$ confidence set C,*

$$P_\theta(\theta' \in C^*(\mathbf{X})) \leq P_\theta(\theta' \in C(\mathbf{X})) \text{ for all } \theta' < \theta.$$

Proof: Let θ' be any value less than θ. Let $A(\theta')$ be the acceptance region of the level α test of $H_0: \theta = \theta'$ obtained by inverting C. Since $A^*(\theta')$ is the UMP acceptance region for testing $H_0: \theta = \theta'$ versus $H_1: \theta > \theta'$, and since $\theta > \theta'$, we have

$$\begin{aligned} P_\theta(\theta' \in C^*(\mathbf{X})) &= P_\theta(\mathbf{X} \in A^*(\theta')) && \text{(invert the confidence set)} \\ &\leq P_\theta(\mathbf{X} \in A(\theta')) && \begin{pmatrix}\text{true for any } A \\ \text{since } A^* \text{ is UMP}\end{pmatrix} \\ &= P_\theta(\theta' \in C(\mathbf{X})). && \begin{pmatrix}\text{invert } A \text{ to} \\ \text{obtain } C\end{pmatrix} \end{aligned}$$

Notice that the above inequality is "$\leq$" because we are working with probabilities of acceptance regions. This is $1-$ power, so UMP tests will minimize these acceptance region probabilities. Therefore, we have established that for $\theta' < \theta$, the probability of false coverage is minimized by the interval obtained from inverting the UMP test. □

Recall our discussion in Section 9.2.1. The UMA confidence set in the above theorem is constructed by inverting the family of tests for the hypotheses

$$H_0: \quad \theta = \theta_0 \qquad \text{versus} \qquad H_1: \quad \theta > \theta_0,$$

where the form of the confidence set is governed by the alternative hypothesis. The above alternative hypotheses, which specify that θ_0 is less than a particular value, lead to *lower* confidence bounds; that is, if the sets are intervals, they are of the form $[L(\mathbf{X}), \infty)$.

Example 9.3.6 (UMA confidence bound) Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, where σ^2 is known. The interval

$$C(\bar{x}) = \left\{\mu : \mu \geq \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}\right\}$$

is a $1-\alpha$ UMA lower confidence bound since it can be obtained by inverting the UMP test of $H_0: \mu = \mu_0$ versus $H_1: \mu > \mu_0$.

The more common two-sided interval,

$$C(\bar{x}) = \left\{\mu : \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\},$$

is not UMA, since it is obtained by inverting the two-sided acceptance region from the test of $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$, hypotheses for which no UMP test exists. ||

In the testing problem, when considering two-sided tests, we found the property of *unbiasedness* to be both compelling and useful. In the confidence interval problem, similar ideas apply. When we deal with two-sided confidence intervals, it is reasonable to restrict consideration to unbiased confidence sets. Remember that an unbiased test is one in which the power in the alternative is always greater than the power in the null. Keep that in mind when reading the following definition.

Definition 9.3.7 A $1-\alpha$ confidence set $C(\mathbf{x})$ is *unbiased* if $P_\theta(\theta' \in C(\mathbf{X})) \le 1-\alpha$ for all $\theta \neq \theta'$.

Thus, for an unbiased confidence set, the probability of false coverage is never more than the minimum probability of true coverage. Unbiased confidence sets can be obtained by inverting unbiased tests. That is, if $A(\theta_0)$ is an unbiased level α acceptance region of a test of $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$ and $C(\mathbf{x})$ is the $1-\alpha$ confidence set formed by inverting the acceptance regions, then $C(\mathbf{x})$ is an unbiased $1-\alpha$ confidence set (see Exercise 9.46).

Example 9.3.8 (Continuation of Example 9.3.6) The two-sided normal interval

$$C(\bar{x}) = \left\{\mu : \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

is an unbiased interval. It can be obtained by inverting the unbiased test of $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ given in Example 8.3.20. Similarly, the interval (9.2.14) based on the t distribution is also an unbiased interval, since it also can be obtained by inverting a unbiased test (see Exercise 8.38). ||

Sets that minimize the probability of false coverage are also called *Neyman-shortest.* The fact that there is a length connotation to this name is somewhat justified by the following theorem, due to Pratt (1961).

Theorem 9.3.9 (Pratt) *Let X be a real-valued random variable with $X \sim f(x|\theta)$, where θ is a real-valued parameter. Let $C(x) = [L(x), U(x)]$ be a confidence interval for θ. If $L(x)$ and $U(x)$ are both increasing functions of x, then for any value θ^*,*

$$\mathrm{E}_{\theta^*}(\mathrm{Length}[C(\mathbf{X})]) = \int_{\theta \neq \theta^*} P_{\theta^*}(\theta \in C(\mathbf{X}))\, d\theta. \tag{9.3.3}$$

Theorem 9.3.9 says that the expected length of $C(x)$ is equal to a sum (integral) of the probabilities of false coverage, the integral being taken over all false values of the parameter.

Proof: From the definition of expected value we can write

$$
\begin{aligned}
\mathrm{E}_{\theta^*}(\text{Length}[C(X)]) &= \int_{\mathcal{X}} \text{Length}[C(x)] f(x|\theta^*)\, dx \\
&= \int_{\mathcal{X}} [U(x) - L(x)] f(x|\theta^*)\, dx && \text{(definition of length)} \\
&= \int_{\mathcal{X}} \left[\int_{L(x)}^{U(x)} d\theta \right] f(x|\theta^*)\, dx && \left(\begin{array}{c}\text{using } \theta \text{ as a} \\ \text{dummy variable}\end{array}\right) \\
&= \int_{\Theta} \left[\int_{U^{-1}(\theta)}^{L^{-1}(\theta)} f(x|\theta^*)\, dx \right] d\theta && \left(\begin{array}{c}\text{invert the order of} \\ \text{integration—see below}\end{array}\right) \\
&= \int_{\Theta} \left[P_{\theta^*}\left(U^{-1}(\theta) \le X \le L^{-1}(\theta)\right) \right] d\theta && \text{(definition)} \\
&= \int_{\Theta} \left[P_{\theta^*}(\theta \in C(X)) \right] d\theta && \left(\begin{array}{c}\text{invert the} \\ \text{acceptance region}\end{array}\right) \\
&= \int_{\theta \ne \theta^*} \left[P_{\theta^*}(\theta \in C(X)) \right] d\theta. && \left(\begin{array}{c}\text{one point does} \\ \text{not change value}\end{array}\right)
\end{aligned}
$$

The string of equalities establishes the identity and proves the theorem. The interchange of integrals is formally justified by Fubini's Theorem (Lehmann and Casella 1998, Section 1.2) but is easily seen to be justified as long as all of the integrands are finite. The inversion of the confidence interval is standard, where we use the relationship

$$\theta \in \{\theta \colon L(x) \le \theta \le U(x)\} \Leftrightarrow x \in \left\{x \colon U^{-1}(\theta) \le x \le L^{-1}(\theta)\right\},$$

which is valid because of the assumption that L and U are increasing. Note that the theorem could be modified to apply to an interval with decreasing endpoints. □

Theorem 9.3.9 shows that there is a formal relationship between the length of a confidence interval and its probability of false coverage. In the two-sided case, this implies that minimizing the probability of false coverage carries along some guarantee of length optimality. In the one-sided case, however, the analogy does not quite work. In that case, intervals that are set up to minimize the probability of false coverage are concerned with parameters in only a portion of the parameter space and length optimality may not obtain. Madansky (1962) has given an example of a $1 - \alpha$ UMA interval (one-sided) that can be beaten in the sense that another, shorter $1-\alpha$ interval can be constructed. (See Exercise 9.45.) Also, Maatta and Casella (1987) have shown that an interval obtained by inverting a UMP test can be suboptimal when measured against other reasonable criteria.

9.3.3 Bayesian Optimality

The goal of obtaining a smallest confidence set with a specified coverage probability can also be attained using Bayesian criteria. If we have a posterior distribution $\pi(\theta|\mathbf{x})$, the posterior distribution of θ given $\mathbf{X} = \mathbf{x}$, we would like to find the set $C(\mathbf{x})$ that satisfies

$$\text{(i)} \quad \int_{C(\mathbf{x})} \pi(\theta|\mathbf{x}) d\mathbf{x} = 1 - \alpha$$

$$\text{(ii)} \quad \text{Size } (C(\mathbf{x})) \leq \text{ Size } (C'(\mathbf{x}))$$

for any set $C'(\mathbf{x})$ satisfying $\int_{C'(\mathbf{x})} \pi(\theta|\mathbf{x}) d\mathbf{x} \geq 1 - \alpha$.

If we take our measure of size to be length, then we can apply Theorem 9.3.2 and obtain the following result.

Corollary 9.3.10 *If the posterior density $\pi(\theta|\mathbf{x})$ is unimodal, then for a given value of α, the shortest credible interval for θ is given by*

$$\{\theta : \pi(\theta|\mathbf{x}) \geq k\} \quad \textit{where} \quad \int_{\{\theta:\pi(\theta|\mathbf{x})\geq k\}} \pi(\theta|\mathbf{x}) d\theta = 1 - \alpha.$$

The credible set described in Corollary 9.3.10 is called a *highest posterior density* (HPD) region, as it consists of the values of the parameter for which the posterior density is highest. Notice the similarity in form between the HPD region and the likelihood region.

Example 9.3.11 (Poisson HPD region) In Example 9.2.16 we derived a $1 - \alpha$ credible set for a Poisson parameter. We now construct an HPD region. By Corollary 9.3.10, this region is given by $\{\lambda \colon \pi(\lambda|\sum x) \geq k\}$, where k is chosen so that

$$1 - \alpha = \int_{\{\lambda:\pi(\lambda|\Sigma x)\geq k\}} \pi(\lambda|\textstyle\sum x) \, d\lambda.$$

Recall that the posterior pdf of λ is gamma$(a + \sum x, [n + (1/b)]^{-1})$, so we need to find λ_L and λ_U such that

$$\pi(\lambda_L|\sum x) = \pi(\lambda_U|\sum x) \quad \text{and} \quad \int_{\lambda_L}^{\lambda_U} \pi(\lambda|\sum x) d\lambda = 1 - \alpha.$$

If we take $a = b = 1$ (as in Example 9.2.16), the posterior distribution of λ given $\sum X = \sum x$ can be expressed as $2(n+1)\lambda \sim \chi^2_{2(\Sigma x+1)}$ and, if $n = 10$ and $\sum x = 6$, the 90% HPD credible set for λ is given by $[.253, 1.005]$.

In Figure 9.3.1 we show three $1 - \alpha$ intervals for λ: the $1 - \alpha$ equal-tailed Bayes credible set of Example 9.2.16, the HPD region derived here, and the classical $1 - \alpha$ confidence set of Example 9.2.15. ‖

The shape of the HPD region is determined by the shape of the posterior distribution. In general, the HPD region is not symmetric about a Bayes point estimator but, like the likelihood region, is rather asymmetric. For the Poisson distribution this is clearly true, as the above example shows. Although it will not always happen, we can usually expect asymmetric HPD regions for scale parameter problems and symmetric HPD regions for location parameter problems.

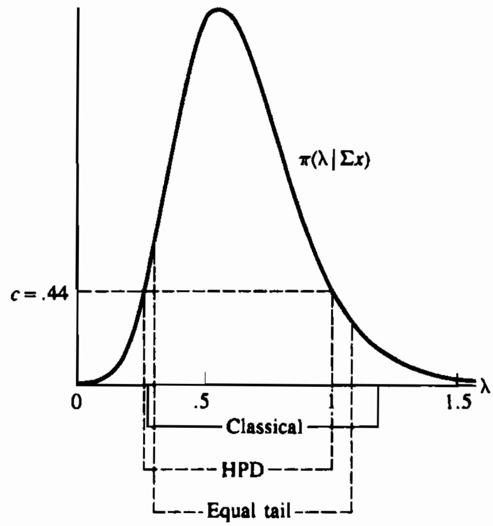

Figure 9.3.1. *Three interval estimators from Example 9.2.16*

Example 9.3.12 (Normal HPD region) The equal-tailed credible set derived in Example 9.2.18 is, in fact, an HPD region. Since the posterior distribution of θ is normal with mean δ^{B}, it follows that $\{\theta\colon \pi(\theta|\bar{x}) \geq k\} = \{\theta\colon \theta \in \delta^{B} \pm k'\}$ for some k' (see Exercise 9.40). So the HPD region is symmetric about the mean $\delta^{\mathrm{B}}(\bar{x})$. ||

9.3.4 Loss Function Optimality

In the previous two sections we looked at optimality of interval estimators by first requiring them to have a minimum coverage probability and then looking for the shortest interval. However, it is possible to put these requirements together in one loss function and use decision theory to search for an optimal estimator. In interval estimation, the action space $\mathcal{A}$ will consist of subsets of the parameter space Θ and, more formally, we might talk of "set estimation," since an optimal rule may not necessarily be an interval. However, practical considerations lead us to mainly consider set estimators that are intervals and, happily, many optimal procedures turn out to be intervals.

We use C (for confidence interval) to denote elements of $\mathcal{A}$, with the meaning of the action C being that the interval estimate "$\theta \in C$" is made. A decision rule $\delta(\mathbf{x})$ simply specifies, for each $\mathbf{x} \in \mathcal{X}$, which set $C \in \mathcal{A}$ will be used as an estimate of θ if $\mathbf{X} = \mathbf{x}$ is observed. Thus we will use the notation $C(\mathbf{x})$, as before.

The loss function in an interval estimation problem usually includes two quantities: a measure of whether the set estimate correctly includes the true value θ and a measure of the size of the set estimate. We will, for the most part, consider only sets C that are intervals, so a natural measure of size is Length(C) = length of C. To express the correctness measure, it is common to use

$$I_C(\theta) = \begin{cases} 1 & \theta \in C \\ 0 & \theta \notin C. \end{cases}$$

That is, $I_C(\theta) = 1$ if the estimate is correct and 0 otherwise. In fact, $I_C(\theta)$ is just the indicator function for the set C. But realize that C will be a random set determined by the value of the data $\mathbf{X}$.

The loss function should reflect the fact that a good estimate would have Length(C) small and $I_C(\theta)$ large. One such loss function is

$$L(\theta, C) = b\,\text{Length}(C) - I_C(\theta), \tag{9.3.4}$$

where b is a positive constant that reflects the relative weight that we want to give to the two criteria, a necessary consideration since the two quantities are very different. If there is more concern with correct estimates, then b should be small, while a large b should be used if there is more concern with interval length.

The risk function associated with (9.3.4) is particularly simple, given by

$$\begin{aligned} R(\theta, C) &= b\mathrm{E}_\theta\left[\text{Length}(C(\mathbf{X}))\right] - \mathrm{E}_\theta I_{C(\mathbf{X})}(\theta) \\ &= b\mathrm{E}_\theta\left[\text{Length}(C(\mathbf{X}))\right] - P_\theta(I_{C(\mathbf{X})}(\theta) = 1) \\ &= b\mathrm{E}_\theta\left[\text{Length}(C(\mathbf{X}))\right] - P_\theta(\theta \in C(\mathbf{X})). \end{aligned}$$

The risk has two components, the expected length of the interval and the coverage probability of the interval estimator. The risk reflects the fact that, simultaneously, we want the expected length to be small and the coverage probability to be high, just as in the previous sections. But now, instead of requiring a minimum coverage probability and then minimizing length, the trade-off between these two quantities is specified in the risk function. Perhaps a smaller coverage probability will be acceptable if it results in a greatly decreased length.

By varying the size of b in the loss (9.3.4), we can vary the relative importance of size and coverage probability of interval estimators, something that could not be done previously. As an example of the flexibility of the present setup, consider some limiting cases. If $b = 0$, then size does not matter, only coverage probability, so the interval estimator $C = (-\infty, \infty)$, which has coverage probability 1, is the best decision rule. Similarly, if $b = \infty$, then coverage probability does not matter, so point sets are optimal. Hence, an entire range of decision rules are possible candidates. In the next example, for a specified finite range of b, choosing a good rule amounts to using the risk function to decide the confidence coefficient while, if b is outside this range, the optimal decision rule is a point estimator.

Example 9.3.13 (Normal interval estimator) Let $X \sim \mathrm{n}(\mu, \sigma^2)$ and assume σ^2 is known. X would typically be a sample mean and σ^2 would have the form τ^2/n, where τ^2 is the known population variance and n is the sample size. For each $c \geq 0$, define an interval estimator for μ by $C(x) = [x - c\sigma, x + c\sigma]$. We will compare these estimators using the loss in (9.3.4). The length of an interval, Length$(C(x)) = 2c\sigma$, does not depend on x. Thus, the first term in the risk is $b(2c\sigma)$. The second term in the risk is

$$
\begin{aligned}
P_\mu(\mu \in C(X)) &= P_\mu(X - c\sigma \le \mu \le X + c\sigma) \\
&= P_\mu\left(-c \le \frac{X-\mu}{\sigma} \le c\right) \\
&= 2P(Z \le c) - 1,
\end{aligned}
$$

where $Z \sim \mathrm{n}(0,1)$. Thus, the risk function for an interval estimator in this class is

$$
R(\mu, C) = b(2c\sigma) - [2P(Z \le c) - 1]. \tag{9.3.5}
$$

The risk function is constant, as it does not depend on μ, and the best interval estimator in this class is the one corresponding to the value c that minimizes (9.3.5).

If $b\sigma > 1/\sqrt{2\pi}$, it can be shown that $R(\mu, C)$ is minimized at $c = 0$. That is, the length portion completely overwhelms the coverage probability portion of the loss, and the best *interval* estimator is the *point* estimator $C(x) = [x, x]$. But if $b\sigma \le 1/\sqrt{2\pi}$, the risk is minimized at $c = \sqrt{-2\log(b\sigma\sqrt{2\pi})}$. If we express c as $z_{\alpha/2}$ for some α, then the interval estimator that minimizes the risk is just the usual $1-\alpha$ confidence interval. (See Exercise 9.53 for details.) ||

The use of decision theory in interval estimation problems is not as widespread as in point estimation or hypothesis testing problems. One reason for this is the difficulty in choosing b in (9.3.4) (or in Example 9.3.13). We saw in the previous example that a choice that might seem reasonable could lead to unintuitive results, indicating that the loss in (9.3.4) may not be appropriate. Some who would use decision theoretic analysis for other problems still prefer to use only interval estimators with a fixed confidence coefficient $(1-\alpha)$. They then use the risk function to judge other qualities like the size of the set.

Another difficulty is in the restriction of the shape of the allowable sets in $\mathcal{A}$. Ideally, the loss and risk functions would be used to judge which shapes are best. But one can always add isolated points to an interval estimator and get an improvement in coverage probability with no loss penalty regarding size. In the previous example we could have used the estimator

$$
C(x) = [x - c\sigma, x + c\sigma] \cup \{\text{all integer values of } \mu\}.
$$

The "length" of these sets is the same as before, but now the coverage probability is 1 for all integer values of μ. Some more sophisticated measure of size must be used to avoid such anomalies. (Joshi 1969 addressed this problem by defining equivalence classes of estimators.)

9.4 Exercises

9.1 If $L(x)$ and $U(x)$ satisfy $P_\theta(L(X) \le \theta) = 1 - \alpha_1$ and $P_\theta(U(X) \ge \theta) = 1 - \alpha_2$, and $L(x) \le U(x)$ for all x, show that $P_\theta(L(X) \le \theta \le U(X)) = 1 - \alpha_1 - \alpha_2$.

9.2 Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, 1)$. A 95% confidence interval for θ is $\bar{x} \pm 1.96/\sqrt{n}$. Let p denote the probability that an additional independent observation, X_{n+1}, will fall in this interval. Is p greater than, less than, or equal to .95? Prove your answer.

9.3 The independent random variables $X_1, \ldots, X_n$ have the common distribution

$$P(X_i \leq x) = \begin{cases} 0 & \text{if } x \leq 0 \\ (x/\beta)^\alpha & \text{if } 0 < x < \beta \\ 1 & \text{if } x \geq \beta. \end{cases}$$

(a) In Exercise 7.10 the MLEs of α and β were found. If α is a known constant, α_0, find an upper confidence limit for β with confidence coefficient .95.

(b) Use the data of Exercise 7.10 to construct an interval estimate for β. Assume that α is known and equal to its MLE.

9.4 Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{n}(0, \sigma_X^2)$, and let $Y_1, \ldots, Y_m$ be a random sample from a $\mathrm{n}(0, \sigma_Y^2)$, independent of the Xs. Define $\lambda = \sigma_Y^2/\sigma_X^2$.

(a) Find the level α LRT of $H_0\colon \lambda = \lambda_0$ versus $H_1\colon \lambda \neq \lambda_0$.

(b) Express the rejection region of the LRT of part (a) in terms of an F random variable.

(c) Find a $1-\alpha$ confidence interval for λ.

9.5 In Example 9.2.5 a lower confidence bound was put on p, the success probability from a sequence of Bernoulli trials. This exercise will derive an upper confidence bound. That is, observing $X_1, \ldots, X_n$, where $X_i \sim \text{Bernoulli}(p)$, we want an interval of the form $[0, U(x_1, \ldots, x_n))$, where $P_p(p \in [0, U(X_1, \ldots, X_n))) \geq 1 - \alpha$.

(a) Show that inversion of the acceptance region of the test

$$H_0\colon p = p_0 \qquad \text{versus} \qquad H_1\colon p < p_0$$

will give a confidence interval of the desired confidence level and form.

(b) Find equations, similar to those given in (9.2.8), that can be used to construct the confidence interval.

9.6 (a) Derive a confidence interval for a binomial p by inverting the LRT of $H_0\colon p = p_0$ versus $H_1\colon p \neq p_0$.

(b) Show that the interval is a highest density region from $p^y(1-p)^{n-y}$ and is not equal to the interval in (10.4.4).

9.7 (a) Find the $1-\alpha$ confidence set for a that is obtained by inverting the LRT of $H_0: a = a_0$ versus $H_1: a \neq a_0$ based on a sample $X_1, \ldots, X_n$ from a $\mathrm{n}(\theta, a\theta)$ family, where θ is unknown.

(b) A similar question can be asked about the related family, the $\mathrm{n}(\theta, a\theta^2)$ family. If $X_1, \ldots, X_n$ are iid $\mathrm{n}(\theta, a\theta^2)$, where θ is unknown, find the $1-\alpha$ confidence set based on inverting the LRT of $H_0\colon a = a_0$ versus $H_1\colon a \neq a_0$.

9.8 Given a sample $X_1, \ldots, X_n$ from a pdf of the form $\frac{1}{\sigma} f((x-\theta)/\sigma)$, list at least five different pivotal quantities.

9.9 Show that each of the three quantities listed in Example 9.2.7 is a pivot.

9.10 (a) Suppose that T is a real-valued statistic. Suppose that $Q(t, \theta)$ is a monotone function of t for each value of $\theta \in \Theta$. Show that if the pdf of T, $f(t|\theta)$, can be expressed in the form (9.2.11) for some function g, then $Q(T, \theta)$ is a pivot.

(b) Show that (9.2.11) is satisfied by taking $g = 1$ and $Q(t, \theta) = F_\theta(t)$, the cdf of T. (This is the Probability Integral Transform.)

9.11 If T is a continuous random variable with cdf $F_T(t|\theta)$ and $\alpha_1 + \alpha_2 = \alpha$, show that an α level acceptance region of the hypothesis $H_0 : \theta = \theta_0$ is $\{t : \alpha_1 \le F_T(t|\theta_0) \le 1 - \alpha_2\}$, with associated confidence $1 - \alpha$ set $\{\theta : \alpha_1 \le F_T(t|\theta) \le 1 - \alpha_2\}$.

9.12 Find a pivotal quantity based on a random sample of size n from a $\mathrm{n}(\theta, \theta)$ population, where $\theta > 0$. Use the pivotal quantity to set up a $1 - \alpha$ confidence interval for θ.

9.13 Let X be a single observation from the beta$(\theta, 1)$ pdf.

(a) Let $Y = -(\log X)^{-1}$. Evaluate the confidence coefficient of the set $[y/2, y]$.
(b) Find a pivotal quantity and use it to set up a confidence interval having the same confidence coefficient as the interval in part (a).
(c) Compare the two confidence intervals.

9.14 Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, where both parameters are unknown. Simultaneous inference on both μ and σ can be made using the Bonferroni Inequality in a number of ways.

(a) Using the Bonferroni Inequality, combine the two confidence sets

$$\left\{\mu \colon \bar{x} - \frac{ks}{\sqrt{n}} \le \mu \le \bar{x} + \frac{ks}{\sqrt{n}}\right\} \quad \text{and} \quad \left\{\sigma^2 \colon \frac{(n-1)s^2}{b} \le \sigma^2 \le \frac{(n-1)s^2}{a}\right\}$$

into one confidence set for (μ, σ). Show how to choose a, b, and k to make the simultaneous set a $1 - \alpha$ confidence set.

(b) Using the Bonferroni Inequality, combine the two confidence sets

$$\left\{\mu \colon \bar{x} - \frac{k\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \frac{k\sigma}{\sqrt{n}}\right\} \quad \text{and} \quad \left\{\sigma^2 \colon \frac{(n-1)s^2}{b} \le \sigma^2 \le \frac{(n-1)s^2}{a}\right\}$$

into one confidence set for (μ, σ). Show how to choose a, b, and k to make the simultaneous set a $1 - \alpha$ confidence set.

(c) Compare the confidence sets in parts (a) and (b).

9.15 Solve for the roots of the quadratic equation that defines Fieller's confidence set for the ratio of normal means (see Miscellanea 9.5.3). Find conditions on the random variables for which

(a) the parabola opens upward (the confidence set is an interval).
(b) the parabola opens downward (the confidence set is the complement of an interval).
(c) the parabola has no real roots.

In each case, give an interpretation of the meaning of the confidence set. For example, what would you tell an experimenter if, for his data, the parabola had no real roots?

9.16 Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, \sigma^2)$, where σ^2 is known. For each of the following hypotheses, write out the acceptance region of a level α test and the $1 - \alpha$ confidence interval that results from inverting the test.

(a) $H_0 \colon \theta = \theta_0$ versus $H_1 \colon \theta \ne \theta_0$
(b) $H_0 \colon \theta \ge \theta_0$ versus $H_1 \colon \theta < \theta_0$
(c) $H_0 \colon \theta \le \theta_0$ versus $H_1 \colon \theta > \theta_0$

9.17 Find a $1 - \alpha$ confidence interval for θ, given $X_1, \ldots, X_n$ iid with pdf

(a) $f(x|\theta) = 1, \ \theta - \frac{1}{2} < x < \theta + \frac{1}{2}$.
(b) $f(x|\theta) = 2x/\theta^2, \ 0 < x < \theta, \ \theta > 0$.

9.18 In this exercise we will investigate some more properties of binomial confidence sets and the Sterne (1954) construction in particular. As in Example 9.2.11, we will again consider the binomial$(3, p)$ distribution.

(a) Draw, as a function of p, a graph of the four probability functions $P_p(X = x)$, $x = 0, \ldots, 3$. Identify the maxima of $P_p(X = 1)$ and $P_p(X = 2)$.
(b) Show that for small ϵ, $P_p(X = 0) > P_p(X = 2)$ for $p = \frac{1}{3} + \epsilon$.
(c) Show that the *most probable construction* is to blame for the difficulties with the Sterne sets by showing that the following acceptance regions can be inverted to obtain a $1 - \alpha = .442$ confidence *interval*.

p	Acceptance region
$[.000, .238]$	$\{0\}$
$(.238, .305)$	$\{0,\ 1\}$
$[.305, .362]$	$\{1\}$
$(.362, .634)$	$\{1,\ 2\}$
$[.634, .695]$	$\{2\}$
$(.695, .762)$	$\{2,\ 3\}$
$[.762, 1.00]$	$\{3\}$

(This is essentially Crow's 1956 modification of Sterne's construction; see Miscellanea 9.5.2.)

9.19 Prove part (b) of Theorem 9.2.12.

9.20 Some of the details of the proof of Theorem 9.2.14 need to be filled in, and the second part of the theorem needs to be proved.

(a) Show that if $F_T(T|\theta)$ is stochastically greater than or equal to a uniform random variable, then so is $\bar{F}_T(T|\theta)$. That is, if $P_\theta(F_T(T|\theta) \le x) \le x$ for every $x, 0 \le x \le 1$, then $P_\theta(\bar{F}_T(T|\theta) \le x) \le x$ for every $x, 0 \le x \le 1$.
(b) Show that for $\alpha_1 + \alpha_2 = \alpha$, the set $\{\theta : F_T(T|\theta) \le \alpha_1 \text{ and } \bar{F}_T(T|\theta) \le \alpha_2\}$ is a $1 - \alpha$ confidence set.
(c) If the cdf $F_T(t|\theta)$ is a decreasing function of θ for each t, show that the function $\bar{F}_T(t|\theta)$ defined by $\bar{F}_T(t|\theta) = P(T \ge t|\theta)$ is a nondecreasing function of θ for each t.
(d) Prove part (b) of Theorem 9.2.14.

9.21 In Example 9.2.15 it was shown that a confidence interval for a Poisson parameter can be expressed in terms of chi squared cutoff points. Use a similar technique to show that if $X \sim \text{binomial}(n, p)$, then a $1 - \alpha$ confidence interval for p is

$$\frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1),2x,\alpha/2}} \le p \le \frac{\frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}{1 + \frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}},$$

where $F_{\nu_1,\nu_2,\alpha}$ is the upper α cutoff from an F distribution with ν_1 and ν_2 degrees of freedom, and we make the endpoint adjustment that the lower endpoint is 0 if $x = 0$ and the upper endpoint is 1 if $x = n$. These are the Clopper and Pearson (1934) intervals.

(*Hint*: Recall the following identity from Exercise 2.40, which can be interpreted in the following way. If $X \sim \text{binomial}(n, \theta)$, then $P_\theta(X \ge x) = P(Y \le \theta)$, where $Y \sim \text{beta}(x, n - x + 1)$. Use the properties of the F and beta distributions from Chapter 5.)

9.22 If $X \sim \text{negative binomial}(r, p)$, use the relationship between the binomial and negative binomial to show that a $1 - \alpha$ confidence interval for p is given by

$$\frac{1}{1 + \frac{x+1}{r} F_{2(x+1),2r,\alpha/2}} \le p \le \frac{\frac{r}{x} F_{2r,2x,\alpha/2}}{1 + \frac{r}{x} F_{2r,2x,\alpha/2}},$$

with a suitable modification if $x = 0$.

9.23 (a) Let $X_1, \ldots, X_n$ be a random sample from a Poisson population with parameter λ and define $Y = \sum X_i$. In Example 9.2.15 a confidence interval for λ was found using the method of Section 9.2.3. Construct another interval for λ by inverting an LRT, and compare the intervals.

(b) The following data, the number of aphids per row in nine rows of a potato field, can be assumed to follow a Poisson distribution:

$$155, \quad 104, \quad 66, \quad 50, \quad 36, \quad 40, \quad 30, \quad 35, \quad 42.$$

Use these data to construct a 90% LRT confidence interval for the mean number of aphids per row. Also, construct an interval using the method of Example 9.2.15.

9.24 For $X \sim \text{Poisson}(\lambda)$, show that the coverage probability of the confidence interval $[L(X), U(X)]$ in Example 9.2.15 is given by

$$P_\lambda(\lambda \in [L(X), U(X)]) = \sum_{x=0}^{\infty} I_{[L(x),U(x)]}(\lambda) \frac{e^{-\lambda}\lambda^x}{x!}$$

and that we can define functions $x_l(\lambda)$ and $x_u(\lambda)$ so that

$$P_\lambda(\lambda \in [L(X), U(X)]) = \sum_{x=x_l(\lambda)}^{x_u(\lambda)} \frac{e^{-\lambda}\lambda^x}{x!}.$$

Hence, explain why the graph of the coverage probability of the Poisson intervals given in Figure 9.2.5 has jumps occurring at the endpoints of the different confidence intervals.

9.25 If $X_1, \ldots, X_n$ are iid with pdf $f(x|\mu) = e^{-(x-\mu)} I_{[\mu,\infty)}(x)$, then $Y = \min\{X_1, \ldots, X_n\}$ is sufficient for μ with pdf

$$f_Y(y|\mu) = ne^{-n(y-\mu)} I_{[\mu,\infty)}(y).$$

In Example 9.2.13 a $1 - \alpha$ confidence interval for μ was found using the method of Section 9.2.3. Compare that interval to $1 - \alpha$ intervals obtained by likelihood and pivotal methods.

9.26 Let $X_1, \ldots, X_n$ be iid observations from a beta$(\theta, 1)$ pdf and assume that θ has a gamma(r, λ) prior pdf. Find a $1 - \alpha$ Bayes credible set for θ.

9.27 (a) Let $X_1, \ldots, X_n$ be iid observations from an exponential(λ) pdf, where λ has the conjugate IG(a, b) prior, an inverted gamma with pdf

$$\pi(\lambda|a, b) = \frac{1}{\Gamma(a)b^a} \left(\frac{1}{\lambda}\right)^{a+1} e^{-1/(b\lambda)}, \quad 0 < \lambda < \infty.$$

Show how to find a $1 - \alpha$ Bayes HPD credible set for λ.

(b) Find a $1 - \alpha$ Bayes HPD credible set for σ^2, the variance of a normal distribution, based on the sample variance s^2 and using a conjugate IG(a, b) prior for σ^2.

(c) Starting with the interval from part (b), find the limiting $1-\alpha$ Bayes HPD credible set for σ^2 obtained as $a \to 0$ and $b \to \infty$.

9.28 Let $X_1, \ldots, X_n$ be iid n(θ, σ^2), where both θ and σ^2 are unknown, but there is only interest on inference about θ. Consider the prior pdf

$$\pi(\theta, \sigma^2|\mu, \tau^2, a, b) = \frac{1}{\sqrt{2\pi\tau^2\sigma^2}} e^{-(\theta-\mu)^2/(2\tau^2\sigma^2)} \frac{1}{\Gamma(a)b^a} \left(\frac{1}{\sigma^2}\right)^{a+1} e^{-1/(b\sigma^2)},$$

a n$(\mu, \tau^2\sigma^2)$ multiplied by an IG(a, b).

(a) Show that this prior is a conjugate prior for this problem.
(b) Find the posterior distribution of θ and use it to construct a $1-\alpha$ credible set for θ.
(c) The classical $1-\alpha$ confidence set for θ can be expressed as

$$\left\{\theta\colon |\theta-\bar{x}|^2 \le F_{1,n-1,\alpha}\frac{s^2}{n}\right\}.$$

Is there any (limiting) sequence of τ^2, a, and b that would allow this set to be approached by a Bayes set from part (b)?

9.29 Let $X_1,\dots,X_n$ are a sequence of n Bernoulli(p) trials.

(a) Calculate a $1-\alpha$ credible set for p using the conjugate beta(a,b) prior.
(b) Using the relationship between the beta and F distributions, write the credible set in a form that is comparable to the form of the intervals in Exercise 9.21. Compare the intervals.

9.30 Complete the credible probability calculation needed in Example 9.2.17.

(a) Assume that a is an integer, and show that $T=\frac{2(nb+1)}{b}\lambda\sim\chi^2_{2(a+\Sigma x)}$.
(b) Show that

$$\frac{\chi^2_\nu-\nu}{\sqrt{2\nu}}\to \mathrm{n}(0,1)$$

as $\nu\to\infty$. (Use moment generating functions. The limit is difficult to evaluate—take logs and then do a Taylor expansion. Alternatively, see Example A.0.8 in Appendix A.)
(c) Standardize the random variable T of part (a), and write the credible probability (9.2.21) in terms of this variable. Show that the standardized lower cutoff point $\to\infty$ as $\Sigma x_i\to\infty$, and hence the credible probability goes to 0.

9.31 Complete the coverage probability calculation needed in Example 9.2.17.

(a) If χ^2_{2Y} is a chi squared random variable with $Y\sim$Poisson(λ), show that $\mathrm{E}(\chi^2_{2Y})=2\lambda$, $\mathrm{Var}(\chi^2_{2Y})=8\lambda$, the mgf of χ^2_Y is given by $\exp\left(-\lambda+\frac{\lambda}{1-2t}\right)$, and

$$\frac{\chi^2_{2Y}-2\lambda}{\sqrt{8\lambda}}\to \mathrm{n}(0,1)$$

as $\lambda\to\infty$. (Use moment generating functions.)
(b) Now evaluate (9.2.22) as $\lambda\to\infty$ by first standardizing χ^2_{2Y}. Show that the standardized upper limit goes to $-\infty$ as $\lambda\to\infty$, and hence the coverage probability goes to 0.

9.32 In this exercise we will calculate the classical coverage probability of the HPD region in (9.2.23), that is, the coverage probability of the Bayes HPD region using the probability model $\bar{X}\sim\mathrm{n}(\theta,\sigma^2/n)$.

(a) Using the definitions given in Example 9.3.12, prove that

$$P_\theta\left(|\theta-\delta^{\mathrm{B}}(\bar{X})|\le z_{\alpha/2}\sqrt{\mathrm{Var}(\theta|\bar{X})}\right)$$
$$=P_\theta\left[-\sqrt{1+\gamma}z_{\alpha/2}+\frac{\gamma(\theta-\mu)}{\sigma/\sqrt{n}}\le Z\le\sqrt{1+\gamma}z_{\alpha/2}+\frac{\gamma(\theta-\mu)}{\sigma/\sqrt{n}}\right].$$

(b) Show that the above set, although a $1-\alpha$ credible set, is not a $1-\alpha$ confidence set. (Fix $\theta \neq \mu$, let $\tau = \sigma/\sqrt{n}$, so that $\gamma = 1$. Prove that as $\sigma^2/n \to 0$, the above probability goes to 0.)

(c) If $\theta = \mu$, however, prove that the coverage probability is bounded away from 0. Find the minimum and maximum of this coverage probability.

(d) Now we will look at the other side. The usual $1-\alpha$ confidence set for θ is $\{\theta : |\theta - \bar{x}| \leq z_{\alpha/2}\sigma/\sqrt{n}\}$. Show that the credible probability of this set is given by

$$P_{\bar{x}}\left(|\theta - \bar{x}| \leq z_{\alpha/2}\sigma/\sqrt{n}\right)$$

$$= P_{\bar{x}}\left[-\sqrt{1+\gamma}z_{\alpha/2} + \frac{\gamma(\bar{x}-\mu)}{\sqrt{1+\gamma}\sigma/\sqrt{n}} \leq Z \leq \sqrt{1+\gamma}z_{\alpha/2} + \frac{\gamma(\bar{x}-\mu)}{\sqrt{1+\gamma}\sigma/\sqrt{n}}\right]$$

and that this probability is not bounded away from 0. Hence, the $1-\alpha$ confidence set is not a $1-\alpha$ credible set.

9.33 Let $X \sim \mathrm{n}(\mu, 1)$ and consider the confidence interval

$$C_a(x) = \{\mu\colon \min\{0, (x-a)\} \leq \mu \leq \max\{0, (x+a)\}\}.$$

(a) For $a = 1.645$, prove that the coverage probability of $C_a(x)$ is exactly .95 for all μ, with the exception of $\mu = 0$, where the coverage probability is 1.

(b) Now consider the so-called noninformative prior $\pi(\mu) = 1$. Using this prior and again taking $a = 1.645$, show that the posterior credible probability of $C_a(x)$ is exactly .90 for $-1.645 \leq x \leq 1.645$ and increases to .95 as $|x| \to \infty$.

This type of interval arises in the problem of *bioequivalence*, where the objective is to decide if two treatments (different formulations of a drug, different delivery systems of a treatment) produce the same effect. The formulation of the problem results in "turning around" the roles of the null and alternative hypotheses (see Exercise 8.47), resulting in some interesting statistics. See Berger and Hsu (1996) for a review of bioequivalence and Brown, Casella, and Hwang (1995) for generalizations of the confidence set.

9.34 Suppose that $X_1, \ldots, X_n$ is a random sample from a $\mathrm{n}(\mu, \sigma^2)$ population.

(a) If σ^2 is known, find a minimum value for n to guarantee that a .95 confidence interval for μ will have length no more than $\sigma/4$.

(b) If σ^2 is unknown, find a minimum value for n to guarantee, with probability .90, that a .95 confidence interval for μ will have length no more than $\sigma/4$.

9.35 Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{n}(\mu, \sigma^2)$ population. Compare expected lengths of $1-\alpha$ confidence intervals for μ that are computed assuming

(a) σ^2 is known.

(b) σ^2 is unknown.

9.36 Let $X_1, \ldots, X_n$ be independent with pdfs $f_{X_i}(x|\theta) = e^{i\theta - x} I_{[i\theta, \infty)}(x)$. Prove that $T = \min_i(X_i/i)$ is a sufficient statistic for θ. Based on T, find the $1-\alpha$ confidence interval for θ of the form $[T+a, T+b]$ which is of minimum length.

9.37 Let $X_1, \ldots, X_n$ be iid uniform$(0, \theta)$. Let Y be the largest order statistic. Prove that Y/θ is a pivotal quantity and show that the interval

$$\left\{\theta\colon y \leq \theta \leq \frac{y}{\alpha^{1/n}}\right\}$$

is the shortest $1-\alpha$ pivotal interval.

9.38 If, in Theorem 9.3.2, we assume that f is continuous, then we can simplify the proof. For fixed c, consider the integral $\int_a^{a+c} f(x)dx$.

(a) Show that $\frac{d}{da}\int_a^{a+c} f(x)\,dx = f(a+c) - f(a)$.

(b) Prove that the unimodality of f implies that $\int_a^{a+c} f(x)\,dx$ is maximized when a satisfies $f(a+c) - f(a) = 0$.

(c) Suppose that, given α, we choose c^* and a^* to satisfy $\int_{a^*}^{a^*+c^*} f(x)\,dx = 1 - \alpha$ and $f(a^* + c^*) - f(a^*) = 0$. Prove that this is the shortest $1-\alpha$ interval.

9.39 Prove a special case of Theorem 9.3.2. Let $X \sim f(x)$, where f is a *symmetric unimodal* pdf. For a fixed value of $1-\alpha$, of all intervals $[a, b]$ that satisfy $\int_a^b f(x)\,dx = 1-\alpha$, the shortest is obtained by choosing a and b so that $\int_{-\infty}^{a} f(x)\,dx = \alpha/2$ and $\int_b^{\infty} f(x)\,dx = \alpha/2$.

9.40 Building on Exercise 9.39, show that if f is symmetric, the optimal interval is of the form $m \pm k$, where m is the mode of f and k is a constant. Hence, show that (a) symmetric likelihood functions produce likelihood regions that are symmetric about the MLE if k' does not depend on the parameter (see (9.2.7)), and (b) symmetric posterior densities produce HPD regions that are symmetric about the posterior mean.

9.41 (a) Prove the following, which is related to Theorem 9.3.2. Let $X \sim f(x)$, where f is a *strictly decreasing* pdf on $[0, \infty)$. For a fixed value of $1-\alpha$, of all intervals $[a, b]$ that satisfy $\int_a^b f(x)\,dx = 1-\alpha$, the shortest is obtained by choosing $a = 0$ and b so that $\int_0^b f(x)\,dx = 1-\alpha$.

(b) Use the result of part (a) to find the shortest $1-\alpha$ confidence interval in Example 9.2.13.

9.42 Referring to Example 9.3.4, to find the shortest pivotal interval for a gamma scale parameter, we had to solve a constrained minimization problem.

(a) Show that the solution is given by the a and b that satisfy $\int_a^b f_Y(y)\,dy = 1-\alpha$ and $f(b)b^2 = f(a)a^2$.

(b) With one observation from a gamma(k, β) pdf with known shape parameter k, find the shortest $1-\alpha$ (pivotal) confidence interval of the form $\{\beta : x/b \le \beta \le x/a\}$.

9.43 Juola (1993) makes the following observation. If we have a pivot $Q(X, \theta)$, a $1-\alpha$ confidence interval involves finding a and b so that $P(a < Q < b) = 1-\alpha$. Typically the length of the interval will be some function of a and b like $b-a$ or $1/b^2 - 1/a^2$. If Q has density f and the length can be expressed as $\int_a^b g(t)\,dt$ the shortest pivotal interval is the solution to

$$\min_{\{a,b\}} \int_a^b g(t)\,dt \text{ subject to } \int_a^b f(t)\,dt = 1-\alpha$$

or, more generally,

$$\min_{C} \int_C g(t)\,dt \text{ subject to } \int_C f(t)\,dt \ge 1-\alpha.$$

(a) Prove that the solution is $C = \{t : g(t) < \lambda f(t)\}$, where λ is chosen so that $\int_C f(t)\,dt = 1-\alpha$. (*Hint*: You can adapt the proof of Theorem 8.3.12, the Neyman-Pearson Lemma.)

(b) Apply the result in part (a) to get the shortest intervals in Exercises 9.37 and 9.42.

9.44 (a) Let $X_1, \ldots, X_n$ be iid Poisson(λ). Find a UMA $1-\alpha$ confidence interval based on inverting the UMP level α test of $H_0\colon \lambda = \lambda_0$ versus $H_1 : \lambda > \lambda_0$.

(b) Let $f(x|\theta)$ be the logistic$(\theta, 1)$ location pdf. Based on one observation, x, find the UMA one-sided $1-\alpha$ confidence interval of the form $\{\theta\colon \theta \le U(x)\}$.

9.45 Let $X_1, \ldots, X_n$ be iid exponential(λ).

(a) Find a UMP size α hypothesis test of $H_0\colon \lambda = \lambda_0$ versus $H_1 : \lambda < \lambda_0$.

(b) Find a UMA $1-\alpha$ confidence interval based on inverting the test in part (a). Show that the interval can be expressed as

$$C^*(x_1, \ldots, x_n) = \left\{\lambda\colon 0 \le \lambda \le \frac{2\sum x_i}{\chi^2_{2n,\alpha}}\right\}.$$

(c) Find the expected length of $C^*(x_1, \ldots, x_n)$.

(d) Madansky (1962) exhibited a $1-\alpha$ interval whose expected length is shorter than that of the UMA interval. In general, Madansky's interval is difficult to calculate, but in the following situation calculation is relatively simple. Let $1-\alpha = .3$ and $n = 120$. Madansky's interval is

$$C^{\mathrm{M}}(x_1, \ldots, x_n) = \left\{\lambda\colon 0 \le \lambda \le -\frac{x_{(1)}}{\log(.99)}\right\},$$

which is a 30% confidence interval. Use the fact that $\chi^2_{240,.7} = 251.046$ to show that the 30% UMA interval satisfies

$$\mathrm{E}\left[\mathrm{Length}\left(C^*(x_1, \ldots, x_n)\right)\right] = .956\lambda > \mathrm{E}\left[\mathrm{Length}\left(C^{\mathrm{M}}(x_1, \ldots, x_n)\right)\right] = .829\lambda.$$

9.46 Show that if $A(\theta_0)$ is an unbiased level α acceptance region of a test of $H_0\colon \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ and $C(\mathbf{x})$ is the $1-\alpha$ confidence set formed by inverting the acceptance regions, then $C(\mathbf{x})$ is an unbiased $1-\alpha$ confidence set.

9.47 Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{n}(\theta, \sigma^2)$ population, where σ^2 is known. Show that the usual one-sided $1-\alpha$ upper confidence bound $\{\theta\colon \theta \le \bar{x} + z_\alpha \sigma/\sqrt{n}\}$ is unbiased, and so is the corresponding lower confidence bound.

9.48 Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{n}(\theta, \sigma^2)$ population, where σ^2 is unknown.

(a) Show that the interval $\theta \le \bar{x} + t_{n-1,\alpha}\frac{s}{\sqrt{n}}$ can be derived by inverting the acceptance region of an LRT.

(b) Show that the corresponding two-sided interval in (9.2.14) can also derived by inverting the acceptance region of an LRT.

(c) Show that the intervals in parts (a) and (b) are unbiased intervals.

9.49 (*Cox's Paradox*) We are to test

$$H_0\colon \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta > \theta_0,$$

where θ is the mean of one of two normal distributions and θ_0 is a fixed but arbitrary value of θ. We observe the random variable X with distribution

$$X \sim \begin{cases} \mathrm{n}(\theta, 100) & \text{with probability } p \\ \mathrm{n}(\theta, 1) & \text{with probability } 1-p. \end{cases}$$

(a) Show that the test given by

$$\text{reject } H_0 \text{ if } X > \theta_0 + z_\alpha \sigma,$$

where $\sigma = 1$ or 10 depending on which population is sampled, is a level α test. Derive a $1 - \alpha$ confidence set by inverting the acceptance region of this test.

(b) Show that a more powerful level α test (for $\alpha > p$) is given by

$$\text{reject } H_0 \text{ if } X > \theta_0 + z_{(\alpha-p)/(1-p)} \text{ and } \sigma = 1; \text{ otherwise } \textit{always} \text{ reject } H_0.$$

Derive a $1 - \alpha$ confidence set by inverting the acceptance region of this test, and show that it is the empty set with positive probability. (Cox's Paradox states that classic optimal procedures sometimes ignore the information about conditional distributions and provide us with a procedure that, while optimal, is somehow unreasonable; see Cox 1958 or Cornfield 1969.)

9.50 Let $X \sim f(x|\theta)$, and suppose that the interval $\{\theta : a(X) \leq \theta \leq b(X)\}$ is a UMA confidence set for θ.

(a) Find a UMA confidence set for $1/\theta$. Note that if $a(x) < 0 < b(x)$, this set is $\{1/\theta : 1/b(x) \leq 1/\theta\} \cup \{1/\theta : 1/\theta \leq 1/a(x)\}$. Hence it is possible for the UMA confidence set to be neither an interval nor bounded.

(b) Show that, if h is a strictly increasing function, the set $\{h(\theta) : h(a(X)) \leq h(\theta) \leq h(b(X))\}$ is a UMA confidence set for $h(\theta)$. Can the condition on h be relaxed?

9.51 If $X_1, \ldots, X_n$ are iid from a location pdf $f(x - \theta)$, show that the confidence set

$$C(x_1, \ldots, x_n) = \{\theta : \bar{x} - k_1 \leq \theta \leq \bar{x} + k_2\},$$

where k_1 and k_2 are constants, has constant coverage probability. (*Hint:* The pdf of $\bar{X}$ is of the form $f_{\bar{X}}(\bar{x} - \theta)$.)

9.52 Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{n}(\mu, \sigma^2)$ population, where both μ and σ^2 are unknown. Each of the following methods of finding confidence intervals for σ^2 results in intervals of the form

$$\left\{\sigma^2 : \frac{(n-1)s^2}{b} \leq \sigma^2 \leq \frac{(n-1)s^2}{a}\right\},$$

but in each case a and b will satisfy different constraints. The intervals given in this exercise are derived by Tate and Klett (1959), who also tabulate some cutoff points.

Define $f_p(t)$ to be the pdf of a χ^2_p random variable with p degrees of freedom. In order to have a $1 - \alpha$ confidence interval, a and b must satisfy

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

but additional constraints are required to define a and b uniquely. Verify that each of the following constraints can be derived as stated.

(a) *The likelihood ratio interval:* The $1 - \alpha$ confidence interval obtained by inverting the LRT of $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma \neq \sigma_0$ is of the above form, where a and b also satisfy $f_{n+2}(a) = f_{n+2}(b)$.

(b) *The minimum length interval:* For intervals of the above form, the $1-\alpha$ confidence interval obtained by minimizing the interval length constrains a and b to satisfy $f_{n+3}(a) = f_{n+3}(b)$.

(c) *The shortest unbiased interval:* For intervals of the above form, the $1-\alpha$ confidence interval obtained by minimizing the probability of false coverage among all unbiased intervals constrains a and b to satisfy $f_{n+1}(a) = f_{n+1}(b)$. This interval can also be derived by minimizing the ratio of the endpoints.

(d) *The equal-tail interval:* For intervals of the above form, the $1-\alpha$ confidence interval obtained by requiring that the probability above and below the interval be equal constrains a and b to satisfy

$$\int_0^a f_{n-1}(t)\,dt = \frac{\alpha}{2}, \qquad \int_b^\infty f_{n-1}(t)\,dt = \frac{\alpha}{2}.$$

(This interval, although very common, is clearly nonoptimal no matter what length criterion is used.)

(e) For $\alpha = .1$ and $n = 3$, find the numerical values of a and b for each of the above cases. Compare the length of this intervals.

9.53 Let $X \sim \mathrm{n}(\mu, \sigma^2), \sigma^2$ known. For each $c \geq 0$, define an interval estimator for μ by $C(x) = [x - c\sigma, x + c\sigma]$ and consider the loss in (9.3.4).

(a) Show that the risk function, $R(\mu, C)$, is given by

$$R(\mu, C) = b(2c\sigma) - P(-c \leq Z \leq c).$$

(b) Using the Fundamental Theorem of Calculus, show that

$$\frac{d}{dc}R(\mu, C) = 2b\sigma - \frac{2}{\sqrt{2\pi}}e^{-c^2/2}$$

and, hence, the derivative is an increasing function of c for $c \geq 0$.

(c) Show that if $b\sigma > 1/\sqrt{2\pi}$, the derivative is positive for all $c \geq 0$ and, hence, $R(\mu, C)$ is minimized at $c = 0$. That is, the best interval estimator is the point estimator $C(x) = [x, x]$.

(d) Show that if $b\sigma \leq 1/\sqrt{2\pi}$, the c that minimizes the risk is $c = \sqrt{-2\log(b\sigma\sqrt{2\pi})}$. Hence, if b is chosen so that $c = z_{\alpha/2}$ for some α, then the interval estimator that minimizes the risk is just the usual $1-\alpha$ confidence interval.

9.54 Let $X \sim \mathrm{n}(\mu, \sigma^2)$, but now consider σ^2 unknown. For each $c \geq 0$, define an interval estimator for μ by $C(x) = [x - cs, x + cs]$, where s^2 is an estimator of σ^2 independent of X, $\nu S^2/\sigma^2 \sim \chi^2_\nu$ (for example, the usual sample variance). Consider a modification of the loss in (9.3.4),

$$L((\mu, \sigma), C) = \frac{b}{\sigma}\text{Length}(C) - I_C(\mu).$$

(a) Show that the risk function, $R((\mu, \sigma), C)$, is given by

$$R((\mu, \sigma), C) = b(2cM) - [2P(T \leq c) - 1],$$

where $T \sim t_\nu$ and $M = \mathrm{E}S/\sigma$.

(b) If $b \leq 1/\sqrt{2\pi}$, show that the c that minimizes the risk satisfies

$$b = \frac{1}{\sqrt{2\pi}}\left(\frac{\nu}{\nu + c^2}\right)^{(\nu+1)/2}.$$

(c) Reconcile this problem with the known σ^2 case. Show that as $\nu \to \infty$, the solution here converges to the solution in the known σ^2 problem. (Be careful of the rescaling done to the loss function.)

9.55 The decision theoretic approach to set estimation can be quite useful (see Exercise 9.56) but it can also give some unsettling results, showing the need for thoughtful implementation. Consider again the case of $X \sim \mathrm{n}(\mu, \sigma^2)$, σ^2 unknown, and suppose that we have an interval estimator for μ by $C(x) = [x - cs, x + cs]$, where s^2 is an estimator of σ^2 independent of $X, \nu S^2/\sigma^2 \sim \chi^2_\nu$. This is, of course, the usual t interval, one of the great statistical procedures that has stood the test of time. Consider the loss

$$L((\mu, \sigma), C) = b \text{ Length}(C) - I_C(\mu),$$

similar to that used in Exercise 9.54, but without scaling the length. Construct another procedure C' as

$$C' = \begin{cases} [x - cs, x + cs] & \text{if } s < K \\ \emptyset & \text{if } s \geq K, \end{cases}$$

where K is a positive constant. Notice that C' does *exactly the wrong thing.* When s^2 is big and there is a lot of uncertainty, we would want the interval to be wide. But C' is empty! Show that we can find a value of K so that

$$R((\mu, \sigma), C') \leq R((\mu, \sigma), C) \quad \text{for every } (\mu, \sigma)$$

with strict inequality for some (μ, σ).

9.56 Let $X \sim f(x|\theta)$ and suppose that we want to estimate θ with an interval estimator C using the loss in (9.3.4). If θ has the prior pdf $\pi(\theta)$, show that the Bayes rule is given by

$$C^\pi = \{\theta \colon \pi(\theta|x) \geq b\}.$$

(*Hint:* Write $\text{Length}(C) = \int_C 1\, d\theta$ and use the Neyman–Pearson Lemma.)

The following two problems relate to Miscellanea 9.5.4.

9.57 Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, where σ^2 is known. We know that a $1 - \alpha$ confidence interval for μ is $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

(a) Show that a $1 - \alpha$ prediction interval for X_{n+1} is $\bar{x} \pm z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}$.

(b) Show that a $1 - \alpha$ tolerance interval for $100p\%$ of the underlying population is given by $\bar{x} \pm z_{\alpha/2}\sigma\left(1 + \frac{1}{\sqrt{n}}\right)$.

(c) Find a $1 - \alpha$ prediction interval for X_{n+1} if σ^2 is unknown.

(If σ^2 is unknown, the $1 - \alpha$ tolerance interval is quite an involved calculation.)

9.58 Let $X_1, \ldots, X_n$ be iid observations from a population with median m. *Distribution-free* intervals can be based on the order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$ in the following way.

(a) Show that the one-sided intervals $(-\infty, x_{(n)}]$ and $[x_{(1)}, \infty)$ are each confidence intervals for m with confidence coefficient $1 - (1/2)^n$, and the confidence coefficient of the interval $[x_{(1)}, x_{(n)}]$ is $1 - 2(1/2)^n$.

(b) Show that the one-sided intervals of part (a) are prediction intervals with coefficient $n/(n+1)$ and the two-sided interval is a prediction interval with coefficient $(n-1)/(n+1)$.

(c) The intervals in part (a) can also be used as tolerance intervals for proportion p of the underlying population. Show that, when considered as tolerance intervals, the one-sided intervals have coefficient $1-p^n$ and the two-sided interval has coefficient $1-p^n-n(1-p)p^{n-1}$. Vardeman (1992) refers to this last calculation as a "nice exercise in order statistics."

9.5 Miscellanea

9.5.1 Confidence Procedures

Confidence sets and tests can be related formally by defining an entity called a *confidence procedure* (Joshi 1969). If $X \sim f(x|\theta)$, where $x \in \mathcal{X}$ and $\theta \in \Theta$, then a confidence procedure is a set in the space $\mathcal{X} \times \Theta$, the Cartesian product space. It is defined as

$$\{(x,\theta)\colon (x,\theta) \in \mathbf{C}\}$$

for a set $\mathbf{C} \in \mathcal{X} \times \Theta$.

From the confidence procedure we can define two slices, or sections, obtained by holding one of the variables constant. For fixed x, we define the *θ-section* or confidence set as

$$C(x) = \{\theta\colon (x,\theta) \in \mathbf{C}\}.$$

For fixed θ, we define the *x-section* or acceptance region as

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

Although this development necessitates working with the product space $\mathcal{X} \times \Theta$, which is one reason we do not use it here, it does provide a more straightforward way of seeing the relationship between tests and sets. Figure 9.2.1 illustrates this correspondence in the normal case.

9.5.2 Confidence Intervals in Discrete Distributions

The construction of optimal (or at least improved) confidence intervals for parameters from discrete distributions has a long history, as indicated in Example 9.2.11, where we looked at the Sterne (1954) modification to the intervals of Clopper and Pearson (1934). Of course, there are difficulties with the Sterne construction, but the basic idea is sound, and Crow (1956) and Blyth and Still (1983) modified Sterne's construction, with the latter producing the shortest set of exact intervals. Casella (1986) gave an algorithm to find a class of shortest binomial confidence intervals.

The history of Poisson interval research (which often includes other discrete distributions) is similar. The Garwood (1936) construction is exactly the Clopper-Pearson argument applied to the binomial, and Crow and Gardner (1959) improved the intervals. Casella and Robert (1989) found a class of shortest Poisson intervals.

Blyth (1986) produces very accurate approximate intervals for a binomial parameter, Leemis and Trivedi (1996) compare normal and Poisson approximations, and Agresti and Coull (1998) argue that requiring discrete intervals to maintain coverage above the nominal level may be too stringent. Blaker (2000) constructs improved intervals for binomial, Poisson, and other discrete distributions that have a nesting property: For $\alpha < \alpha'$, the $1-\alpha$ intervals contain the corresponding $1-\alpha'$ intervals.

9.5.3 Fieller's Theorem

Fieller's Theorem (Fieller 1954) is a clever argument to get an exact confidence set on a ratio of normal means.

Given a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a bivariate normal distribution with parameters $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, a confidence set on $\theta = \mu_Y/\mu_X$ can be formed in the following way. For $i = 1, \ldots, n$, define $Z_{\theta i} = Y_i - \theta X_i$ and $\bar{Z}_\theta = \bar{Y} - \theta \bar{X}$. It can be shown that $\bar{Z}_\theta$ is normal with mean 0 and variance

$$V_\theta = \frac{1}{n}\left(\sigma_Y^2 - 2\theta\rho\sigma_Y\sigma_X + \theta^2\sigma_X^2\right).$$

V_θ can be estimated with $\hat{V}_\theta$, given by

$$\begin{aligned}\hat{V}_\theta &= \frac{1}{n(n-1)}\sum_{i=1}^{n}(Z_{\theta i} - \bar{Z}_\theta)^2 \\ &= \frac{1}{n-1}\left(S_Y^2 - 2\theta S_{YX} + \theta^2 S_X^2\right),\end{aligned}$$

where

$$S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \quad S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad S_{YX} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X}).$$

Furthermore, it also can be shown that $\mathrm{E}\hat{V}_\theta = V_\theta$, $\hat{V}_\theta$ is independent of $\bar{Z}_\theta$, and $(n-1)\hat{V}_\theta/V_\theta \sim \chi^2_{n-1}$. Hence, $\bar{Z}_\theta/\sqrt{\hat{V}_\theta} \sim t_{n-1}$ and the set

$$\left\{\theta\colon \frac{\bar{z}_\theta^2}{\hat{v}_\theta} \le t^2_{n-1,\alpha/2}\right\}$$

defines a $1-\alpha$ confidence set for θ, the ratio of the means. This set defines a parabola in θ, and the roots of the parabola give the endpoints of the confidence set. Writing the set in terms of the original variables, we get

$$\left\{\theta\colon \left(\bar{x}^2 - \frac{t^2_{n-1,\alpha/2}}{n-1}S_x^2\right)\theta^2 - 2\left(\bar{x}\,\bar{y} - \frac{t^2_{n-1,\alpha/2}}{n-1}S_{yx}\right)\theta + \left(\bar{y}^2 - \frac{t^2_{n-1,\alpha/2}}{n-1}S_y^2\right) \le 0\right\}.$$

One interesting feature of this set is that, depending on the roots of the parabola, it can be an interval, the complement of an interval, or the entire real line (see

Exercise 9.15). Furthermore, to maintain $1-\alpha$ confidence, this interval must be infinite with positive probability. See Hwang (1995) for an alternative based on bootstrapping, and Tsao and Hwang (1998, 1999) for an alternative confidence approach.

9.5.4 *What About Other Intervals?*

Vardeman (1992) asks the question in the title of this Miscellanea, arguing that mainstream statistics should spend more time on intervals other than two-sided confidence intervals. In particular, he lists (a) one-sided intervals, (b) distribution-free intervals, (c) prediction intervals, and (d) tolerance intervals.

We have seen one-sided intervals, and distribution-free intervals are intervals whose probability guarantee holds with little (or no) assumption on the underlying cdf (see Exercise 9.58). The other two interval definitions, together with the usual confidence interval, provide use with a hierarchy of inferences, each more stringent than the previous.

If $X_1, X_2, \ldots, X_n$ are iid from a population with cdf $F(x|\theta)$, and $C(\mathbf{x}) = [l(\mathbf{x}), u(\mathbf{x})]$ is an interval, for a specified value $1-\alpha$ it is a

(i) *confidence* interval if $P_\theta[l(\mathbf{X}) \leq \theta \leq u(\mathbf{X})] \geq 1-\alpha$;

(ii) *prediction* interval if $P_\theta[l(\mathbf{X}) \leq X_{n+1} \leq u(\mathbf{X})] \geq 1-\alpha$;

(iii) *tolerance* interval if, for a specified value p, $P_\theta[F(u(\mathbf{X})|\theta) - F(l(\mathbf{X})|\theta) \geq p] \geq 1-\alpha$.

So a confidence interval covers a mean, a prediction interval covers a new random variable, and a tolerance interval covers a proportion of the population. Thus, each gives a different inference, with the appropriate one being dictated by the problem at hand.