3. PARAMETRIC INTERVAL ESTIMATION

In Section 2, we focussed on estimating particular numerical characteristics of a population (denoted as either parameters, in a parametric setting, or as functionals of an underlying distribution, in the non-parametric setting). In particular, we derived and discussed estimation methods and their properties. Inherent in this discussion was an effort to assess which estimates were "good" and which less so. This notion was generally encapsulated in an assessment of the "accuracy" of our chosen estimate (often measured using calculations of either bias or variance). Our previous focus was on using these measures of accuracy to choose between competing procedures. However, once we have settled on an estimation procedure, the accuracy of this chosen method is still an important issue and ought to somehow be reflected in any statements we make regarding our estimation of the characteristics of interest in our population of study. This attachment of an assessment of accuracy, or of a level of *confidence*, to an estimation procedure leads to the development of *interval estimates* which are the subject of the current sections of these notes. Generally speaking, while a point estimate is a single "best guess" at the value of the population characteristic of interest, an interval estimate provides a range of "plausible" values, in which we have a certain level of confidence (i.e., to which we can attach a certain level of credence for our belief that the true value of the population characteristic under study is contained within our given "plausible range"). We will begin to formalise these notions in the following sections; however, we will not be able to do so to the same extent as we did for point estimation, since such a level of investigation requires detailed mathematics beyond the scope of these notes (we will return to the issues in Section 4, though, where we will discuss the relationship between interval estimates and hypothesis tests). Before proceeding, we again note that the majority of our focus will be on parametric probability models. As in Section 2, though, we will briefly discuss some non-parametric interval estimation procedures which are not as heavily dependent on an appropriate choice of a probabilistic model.

3.1. Confidence Intervals and Regions

Point estimation procedures were formally outlined and discussed in Section 2. There, we saw that point estimates will almost never yield an estimate which is "correct" since, by their nature, they must be based on observations obtained through a random sampling mechanism. The notion of error in estimates was the fundamental guide used to describe various properties of estimation procedures and thereby choose between procedures on rigorous and quantifiable grounds. In this section, we take our discussion of the likely size of errors in estimation and attempt to incorporate it into the estimation procedure itself. This notion will certainly be a familiar one from common experience where estimated values are often presented with a "plus or minus" value which is designed to provide a rough idea of the likely discrepancy between the estimate and its target value. More formally, we recall the standard confidence interval procedures developed for estimates of population means and proportions which are the stalwart of any introductory unit in statistics. As such, we note that the most common sort of interval estimate will take the form of a point estimate plus or minus some measure of its *standard deviation* or *standard error*. [NOTE: the term *standard error* was originally used to represent the standard deviation of a sample average, however, in modern parlance it is essentially synonymous with *standard deviation*, the fundamental feature being the measure of the likely size of fluctuations in (random) values produced by an estimator or statistic]. We note, however, that the notion of interval estimators is more than just the extension of a single number estimate to a range of possible values. The key statistical aspect is the attachment of a level of confidence to these intervals. Moreover, we note that even the idea that the range of values must be an interval is not always required. In particular, if the characteristic or parameter of interest

is vector-valued, then we require confidence regions of appropriate dimension. In this section, we focus on a formal definition of a confidence interval or region and discuss the general principles associated with them.

3.1.1. Definitions: We shall start from the framework of the standard parametric model, wherein we assume that we have a random sample, $X_1, \ldots, X_n$, of observations from a population characterised by a distribution having a density, $f_X(x;\theta)$, belonging to a parameterised family indexed by the value $\theta \in \Theta$. Furthermore, we shall again assume that we are interested in estimating the value $\tau = \tau(\theta)$, and we shall further assume that τ is scalar (though, θ may be vector-valued). In this situation, we define a confidence interval as follows:

Definition 3.1: Any pair of statistics, $T_l = t_l(X_1, \ldots, X_n)$ and $T_u = t_u(X_1, \ldots, X_n)$ which satisfy $T_l \leq T_u$ [i.e., $Pr_\theta(T_l \leq T_u) = 1$] and

$$Pr_\theta\{T_l \leq \tau(\theta) \leq T_u\} = 1 - \alpha, \qquad \forall \theta \in \Theta,$$

is called a $100(1-\alpha)\%$ confidence interval. Also, $1-\alpha$ is usually referred to as the *confidence coefficient*, while T_l and T_u are referred to as the *lower* and *upper confidence limits*, respectively. More generally, if the parameter of interest, $\tau = \tau(\theta)$, takes values in the set $\mathcal{T}$ and $B_{X_1,\ldots,X_n} \subseteq \mathcal{T}$ is a (random) set such that

$$Pr_\theta\{\tau(\theta) \in B_{X_1,\ldots,X_n}\} = 1 - \alpha, \qquad \forall \theta \in \Theta,$$

then $B_{X_1,\ldots,X_n}$ is termed a $100(1-\alpha)\%$ *confidence region* for $\tau = \tau(\theta)$ [NOTE: This last definition is actually applicable to the case where τ is vector-valued].

In examining this definition, we must recall that a statistics is a function of the observed data only (i.e., it cannot depend on the value of the parameter θ). Moreover, we note that the main probability statement requires that the probability that the random values T_l and T_u surround $\tau(\theta)$ be $1-\alpha$ regardless of the parameter value θ (note that in general this probability would certainly depend on θ, and thus, a confidence interval cannot be composed of any two statistics). Finally, we must remind ourselves that, despite the way in which the probability statement is written, it is T_l and T_u which are the random quantities here. As such, the probability statement is to be interpretted as saying that in repeated samples from the population, the random interval $(T_l, T_u) = \{t_l(X_1, \ldots, X_n), t_u(X_1, \ldots, X_n)\}$ will enclose the true value $\tau(\theta)$ in $1-\alpha$ proportion of these samples [as opposed to the tempting but incorrect interpretation that $\tau(\theta)$ has $1-\alpha$ probability of falling within the specific observed values $t_l = t_l(x_1, \ldots, x_n)$ and $t_u = t_u(x_1, \ldots, x_n)$ derived from a specific realisation of the random sample, $x_1, \ldots, x_n$.] However, this notion of randomness in the parameter was seen in point estimation to lead to Bayesian estimation, and we will develop the same Bayesian notions of confidence regions in Section 3.2.3.

Once we have a formal definition of a confidence interval, we can make some important observations. First, if we suppose that the parameter of interest $\tau = \tau(\theta)$ takes values in the set $\mathcal{T} \subseteq [a, b]$ for some values a (possibly negatively infinite) and b (possibly infinite), then setting $T_l(X_1, \ldots, X_n) = a$ and finding an appropriate statistic $T_u(X_1, \ldots, X_n)$ to satisfy Definition 3.1 leads to a *lower one-sided confidence interval* for $\tau(\theta)$ and $T_u(X_1, \ldots, X_n)$ is termed an *upper one-sided confidence limit*. Similarly, setting $T_u(X_1, \ldots, X_n) = b$ and finding an appropriate statistic $T_l(X_1, \ldots, X_n)$ to satisfy Definition 3.1 leads to an *upper one-sided confidence interval* for $\tau(\theta)$ and $T_l(X_1, \ldots, X_n)$ is termed a *lower one-sided confidence limit*. Moreover, once we have found a confidence interval for $\tau(\theta)$, we have in principle found a confidence interval for any parameter $\gamma = \gamma(\theta) = g\{\tau(\theta)\}$ defined by a monotone (and thus invertible) function $g(\cdot)$. To see this, we note

that if $g(\cdot)$ is increasing and (T_l, T_u) is a $100(1-\alpha)\%$ confidence interval for $\tau(\theta)$, then we have:

$$1-\alpha = Pr_\theta\{T_l \leq \tau(\theta) \leq T_u\} = Pr_\theta\{g(T_l) \leq \gamma(\theta) \leq g(T_u)\}.$$

In other words, $\{g(T_l), g(T_u)\}$ is a $100(1-\alpha)\%$ confidence interval for $\gamma = \gamma(\theta)$. Similarly, if $g(\cdot)$ is decreasing than a nearly identical argument shows that $\{g(T_u), g(T_l)\}$ is a $100(1-\alpha)\%$ confidence interval for $\gamma = \gamma(\theta)$.

Unfortunately, finding appropriate statistics T_l and T_u is often extremely difficult. Moreover, as was the case for point estimation, there may in general be several competing choices for confidence intervals for any particular parameter of interest and we would like to be able to assess the quality of these competing procedures and thereby select among them. As noted in the introduction to this section, however, the full completion of such a task is beyond the scope of these notes, and we satisfy ourselves here with some basic descriptions of methods for deriving intervals and a brief discussion of their properties.

3.1.2. Pivotal Quantities: The critical feature of Definition 3.1 which makes finding confidence intervals so difficult is the requirement that the probability statement associated with the statistics T_l and T_u hold true for *any* value of the parameter θ. In general, of course, the distribution of a statistic (and thus any probability statement about that statistic) will depend on the parameter θ, since statistics are simply functions of the random observations $X_1, \ldots, X_n$. The search for confidence intervals, then, can often be seen as the search for quantities whose distribution does *not* depend on the parameter θ. In this vein, we define a particular such quantity as follows:

Definition 3.2: Let $X_1, \ldots, X_n$ be a random sample from a distribution having density function $f_X(x;\theta)$ for some value of the parameter $\theta \in \Theta$. Further, for a scalar parameter of interest $\tau = \tau(\theta)$, let $Q = q(\tau; X_1, \ldots, X_n)$ be a function of τ as well as $X_1, \ldots, X_n$ such that $q(\cdot; X_1, \ldots, X_n)$ is monotone in its first argument. That is, for any given realisations of the sample, $x_1, \ldots, x_n$, the function $h_{x_1,\ldots,x_n}(t) = q(t; x_1, \ldots, x_n)$ is either strictly increasing or strictly decreasing in t. If the distribution of the random quantity Q does not depend on θ, then Q is termed a *pivotal quantity* or sometimes just a *pivot*.

Note that a pivot is not a statistic, since it depends on the value $\tau = \tau(\theta)$ as well as on the observed data. However, the usefulness of pivots in constructing confidence intervals should be clear from the fact that their distributions are free of any dependence on θ, and therefore probability statements regarding them will be true for all $\theta \in \Theta$. Finally, we note that the requirement that a pivotal quantity be monotone in τ is necessary to ensure invertibility in this argument, and we shall see how this technical aspect comes into play in the following section where we demonstrate how pivots can be used to construct confidence intervals. Before heading to this discussion, however, we close this section by noting some of the common pivots which should be familiar from any introductory unit in statistics.

Example 3.1: Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then, the quantity

$$Q_1 = q_1(\mu; X_1, \ldots, X_n) = \frac{\sqrt{n}(\overline{X} - \mu)}{S},$$

where $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ and $S = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$ is well known to have a Student's t-distribution with $n-1$ degrees of freedom. Moreover, for any given values of the data, $x_1, \ldots, x_n$ (which in turn determine the observed values of $\overline{x}$ and s) we see that $q_1(\mu; x_1, \ldots, x_n) = s^{-1}\sqrt{n}\overline{x} - s^{-1}\sqrt{n}\mu$ is clearly a decreasing function in μ. As such, Q_1 is a pivotal quantity for the parameter of interest $\tau(\mu, \sigma^2) = \mu$. Furthermore, it is then easily seen that $Q_2 = -Q_1 = S^{-1}\sqrt{n}(\mu - \overline{X})$ is also a pivot.

Alternatively, the quantity

$$Q_3 = q_3(\sigma^2; X_1, \ldots, X_n) = \frac{(n-1)S^2}{\sigma^2},$$

is known to have a chi-squared distribution with $n-1$ degrees of freedom. Since, for any given values of $x_1, \ldots, x_n$, $q_3(\sigma^2; x_1, \ldots, x_n)$ is clearly strictly decreasing in σ^2 we see that Q_3 is a pivotal quantity for the parameter of interest $\tau(\mu, \sigma^2) = \sigma^2$. Finally, since the square-root function is itself monotone, the quantity $Q_4 = \sqrt{Q_3} = \sqrt{n-1}S/\sigma$ is seen to have a distribution which does not depend on θ (since Q_3 itself has this property) and is decreasing in σ (NOTE: the distribution of Q_4 is sometimes referred to as the *chi* distribution, or the *generalised Rayleigh* distribution). As such, it is a pivot for $\tau(\mu, \sigma^2) = \sigma$.

In closing, we note that the fact that the quantities Q_1 and Q_3 have a Student's t-distribution and a chi-squared distribution, respectively, is easily demonstrated using the (perhaps surprising) fact that, for normally distributed data, the sample average and the sample variance are independent of each other (the demonstration of which is left as an exercise).

3.2. Construction Methods

In this section, we discuss the construction of confidence intervals. In particular, we start by attempting to construct "exact" confidence intervals; that is, intervals which satisfy Definition 3.1 exactly. Of course, such intervals are extremely difficult to construct, and thus in the remaining sections we discuss "asymptotic" confidence intervals, which do not necessarily satisfy Definition 3.1 exactly for any fixed sized sample, but will satisfy the definition more and more closely as the sample size, n, tends towards infinity. In addition, we discuss the Bayesian framework for confidence regions (often termed credibility regions in this context) as well as some non-parametric approaches.

3.2.1. Pivots and Statistics: We start by employing the concept of pivotal quantities outlined in Definition 3.2. Suppose that $X_1, \ldots, X_n$ are a random sample from a distribution having density function $f_X(x; \theta)$ and that a pivotal quantity $Q = q(\tau; X_1, \ldots, X_n)$ exists for the (scalar) parameter of interest $\tau = \tau(\theta)$. From the definition of a pivotal quantity, we know that the function $h_{X_1,\ldots,X_n}(t) = q(t; X_1, \ldots, X_n)$ is monotone in t, and thus admits a monotone inverse function; that is, there exists a function $h^{-1}_{X_1,\ldots,X_n}(t)$ such that $h^{-1}_{X_1,\ldots,X_n}\{h_{X_1,\ldots,X_n}(t)\} = t$ and $h^{-1}_{X_1,\ldots,X_n}(t)$ is itself a monotone function of t. Moreover, we know that Q has a distribution which does not depend on θ. Suppose we denote the *CDF* of Q by $F_Q(x)$, and let the quantities $q_{l,\alpha}$ and $q_{u,\alpha}$ be such that $F_Q(q_{u,\alpha}) - F_Q(q_{l,\alpha}) = Pr(q_{l,\alpha} \leq Q \leq q_{u,\alpha}) = 1 - \alpha$. In this case, it is clear that $q_{l,\alpha}$ and $q_{u,\alpha}$ do not depend on θ, and so we see that (for all $\theta \in \Theta$):

$$\begin{aligned} 1 - \alpha &= Pr(q_{l,\alpha} \leq Q \leq q_{u,\alpha}) \\ &= Pr_\theta\{q_{l,\alpha} \leq q(\tau; X_1, \ldots, X_n) \leq q_{u,\alpha}\} \\ &= Pr_\theta\{h^{-1}_{X_1,\ldots,X_n}(q_{l,\alpha}) \leq \tau \leq h^{-1}_{X_1,\ldots,X_n}(q_{u,\alpha})\} \end{aligned}$$

if $h_{X_1,\ldots,X_n}(t)$ (and therefore its inverse as well) is an increasing function. If $h_{X_1,\ldots,X_n}(t)$ had been decreasing (and therefore its inverse as well), the final equality in the preceding argument would have read $Pr_\theta\{h^{-1}_{X_1,\ldots,X_n}(q_{u,\alpha}) \leq \tau \leq h^{-1}_{X_1,\ldots,X_n}(q_{l,\alpha})\}$, since the order of the inequality signs are reversed by the decreasing nature of the inverse function $h^{-1}_{X_1,\ldots,X_n}(t)$. In either case, we have used our pivotal quantity to arrive at an exact confidence interval for τ. Before proceeding to some examples, we note that it will generally be the case that more than one combination of values $q_{l,\alpha}$ and $q_{u,\alpha}$ exist which satisfy $F_Q(q_{u,\alpha}) - F_Q(q_{l,\alpha}) = 1 - \alpha$. Moreover, there may be more than

one possible pivotal quantity for any given parameter of interest, $\tau = \tau(\theta)$. As such, the pivotal method of finding confidence intervals described here admits several possible results. Of course, choosing between the possible intervals is an important issue (in just the same way as selecting among various estimators was for point estimation), and we will briefly discuss the issues involved in making such choices (though we will not be able to give a very detailed discussion, given the complex mathematics required to do so).

Example 3.1*(cont'd)*: For normally distributed data, $X_1, \ldots, X_n$, we have seen that

$$Q_1 = q_1(\mu; X_1, \ldots, X_n) = \frac{\sqrt{n}(\overline{X} - \mu)}{S} = -\frac{\sqrt{n}}{S}\mu + \frac{\sqrt{n}\overline{X}}{S} = h_{X_1,\ldots,X_n}(\mu),$$

is a pivotal quantity having a Student's t-distribution with $n-1$ degrees of freedom. In this case, the appropriate inverse function is

$$h^{-1}_{X_1,\ldots,X_n}(t) = -\frac{St}{\sqrt{n}} + \overline{X},$$

since clearly

$$\begin{aligned} h^{-1}_{X_1,\ldots,X_n}\{h_{X_1,\ldots,X_n}(\mu)\} &= -\frac{Sh_{X_1,\ldots,X_n}(\mu)}{\sqrt{n}} + \overline{X} \\ &= -\frac{\sqrt{n}(\overline{X} - \mu)}{\sqrt{n}} + \overline{X} \\ &= (\mu - \overline{X}) + \overline{X} \\ &= \mu. \end{aligned}$$

Thus, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\begin{aligned} \{h^{-1}_{X_1,\ldots,X_n}(q_{u,\alpha}), h^{-1}_{X_1,\ldots,X_n}(q_{l,\alpha})\} &= \left(-\frac{Sq_{u,\alpha}}{\sqrt{n}} + \overline{X}, -\frac{Sq_{l,\alpha}}{\sqrt{n}} + \overline{X} \right) \\ &= \left(\overline{X} - q_{u,\alpha}\frac{S}{\sqrt{n}}, \overline{X} - q_{l,\alpha}\frac{S}{\sqrt{n}} \right), \end{aligned}$$

where $q_{l,\alpha}$ and $q_{u,\alpha}$ are any two values satisfying $Pr(q_{l,\alpha} \leq Q_1 \leq q_{u,\alpha}) = 1 - \alpha$. In other words, $q_{l,\alpha}$ and $q_{u,\alpha}$ must be appropriate quantiles of the Student's t-distribution with $n-1$ degrees of freedom. Indeed, if we let $t_{(n-1)}(\gamma)$ be the γ-quantile of the Student's t-distribution with $n-1$ degrees of freedom, so that $Pr\{Q_1 \leq t_{(n-1)}(\gamma)\} = \gamma$, then we can choose $q_{l,\alpha} = t_{(n-1)}(\alpha - \beta)$ and $q_{u,\alpha} = t_{(n-1)}(1-\beta)$ for any $0 \leq \beta \leq \alpha$. To illustrate where the term "pivot" derives, we note that the interval construction can be viewed from the following argument:

$$\begin{aligned} 1 - \alpha &= Pr(q_{l,\alpha} \leq Q_1 \leq q_{u,\alpha}) \\ &= Pr\left\{ q_{l,\alpha} \leq \frac{\sqrt{n}(\overline{X} - \mu)}{S} \leq q_{u,\alpha} \right\} \\ &= Pr\left(q_{l,\alpha}\frac{S}{\sqrt{n}} \leq \overline{X} - \mu \leq q_{u,\alpha}\frac{S}{\sqrt{n}} \right) \\ &= Pr\left(q_{l,\alpha}\frac{S}{\sqrt{n}} - \overline{X} \leq -\mu \leq q_{u,\alpha}\frac{S}{\sqrt{n}} - \overline{X} \right) \\ &= Pr\left(\overline{X} - q_{u,\alpha}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} - q_{l,\alpha}\frac{S}{\sqrt{n}} \right), \end{aligned}$$

where we have "pivoted" the inequality within the probability statement from one focussed on Q_1 to one apparently focussed on $\tau(\mu, \sigma^2) = \mu$, the parameter of interest [NOTE: the

actual algebra of the "pivoting" is precisely the operation determined by the inverse function $h^{-1}_{X_1,\ldots,X_n}(t)$.] Note that, if we choose $\beta = 0$, then $q_{u,\alpha} = t_{(n-1)}(1) = \infty$ and we arrive at a lower one-sided confidence interval, while if we let $\beta = \alpha$, then $q_{l,\alpha} = t_{(n-1)}(0) = -\infty$ and we arrive at an upper one-sided confidence interval. More usually, we would choose $\beta = \alpha/2$, which, when combined with the fact that $t_{(n-1)}(\alpha/2) = -t_{(n-1)}(1-\alpha/2)$ due to the symmetry of the Student's t-distributions, would yield the familiar interval: $\overline{X} \pm t_{(n-1)}(\alpha/2)\frac{S}{\sqrt{n}}$. Such an interval is often termed *equal-tailed*, since $Pr(Q_1 < q_{l,\alpha}) = Pr(Q_1 > q_{u,\alpha})$ in such cases. It turns out that for this case, this familiar interval is also the shortest possible interval among all choices of β (a fact which again relies on the symmetry of the Student's t-distributions), adding evidence to support its widespread common usage (in fact, it can be shown to be the shortest among all possible confidence intervals, provided the data truly is normally distributed, though such a demonstration is beyond the scope of these notes).

Alternatively, we also saw that the quantity

$$Q_3 = q_3(\sigma^2; X_1, \ldots, X_n) = \frac{(n-1)S^2}{\sigma^2} = h_{X_1,\ldots,X_n}(\sigma^2),$$

was a pivotal quantity having a chi-squared distribution with $n-1$ degrees of freedom. Either by applying the appropriate inverse function, which in this case is given by:

$$h^{-1}_{X_1,\ldots,X_n}(t) = \frac{(n-1)S^2}{t},$$

or else by following a similar pivoting argument to that described previously for the case of intervals for μ, we see that a $100(1-\alpha)\%$ confidence interval for σ^2 can be constructed from the relationship:

$$\begin{aligned}
1-\alpha &= Pr(q_{l,\alpha} \leq Q_3 \leq q_{u,\alpha}) \\
&= Pr\left(q_{l,\alpha} \leq \frac{(n-1)S^2}{\sigma^2} \leq q_{u,\alpha}\right) \\
&= Pr\left(\frac{q_{l,\alpha}}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{q_{u,\alpha}}{(n-1)S^2}\right) \\
&= Pr\left(\frac{(n-1)S^2}{q_{u,\alpha}} \leq \sigma^2 \leq \frac{(n-1)S^2}{q_{l,\alpha}}\right),
\end{aligned}$$

where $q_{l,\alpha}$ and $q_{u,\alpha}$ are now appropriate quantiles of the chi-squared distribution with $n-1$ degrees of freedom. In particular, if we denote the γ-quantile of the chi-squared distribution with $n-1$ degrees of freedom by $\chi^2_{(n-1)}(\gamma)$, so that $Pr\{Q_3 \leq \chi^2_{(n-1)}(\gamma)\} = \gamma$, then we can set $q_{l,\alpha} = \chi^2_{(n-1)}(\alpha-\beta)$ and $q_{u,\alpha} = \chi^2_{(n-1)}(1-\beta)$ for any value $0 \leq \beta \leq \alpha$ and arrive at the $100(1-\alpha)\%$ confidence interval for σ^2:

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

As before, if we set $\beta = 0$, so that $q_{u,\alpha} = \chi^2_{(n-1)}(1) = \infty$, we arrive at the lower one-sided confidence interval, while if we set $\beta = \alpha$, so that $q_{l,\alpha} = \chi^2_{(n-1)}(0) = 0$, we arrive at the upper one-sided confidence interval. If we set $\beta = \alpha/2$, then we arrive at the equal-tailed interval. Of course, we might wish to choose β so that we arrive at the interval with shortest length. Unfortunately, in this case, constructing such an interval requires the solution to a difficult (though by no means impossible) calculus problem with no closed-form solution, and thus requiring a numerical or computer-based solution. While the actual value of β which leads

to the shortest interval is difficult to find, the lack of symmetry in the chi-squared distribution ensures that whatever the appropriate value is, it is not equal to $\alpha/2$. In other words, in this case, the shortest length interval is not the same as the equal-tailed interval.

Finally, we note that the "pivoting" argument here can obviously be taken one step further to show that:

$$1-\alpha = Pr\left(\frac{(n-1)S^2}{q_{u,\alpha}} \leq \sigma^2 \leq \frac{(n-1)S^2}{q_{l,\alpha}}\right) = Pr\left(\frac{\sqrt{n-1}S}{\sqrt{q_{u,\alpha}}} \leq \sigma \leq \frac{\sqrt{n-1}S}{\sqrt{q_{l,\alpha}}}\right),$$

to arrive at a confidence interval for σ. Of course, exactly the same interval would result from using the pivot $Q_4 = \sqrt{Q_3}$, which follows from the observation that the distribution of Q_4 is precisely the distribution of the square-root of a chi-squared distribution (which is termed the chi distribution for obvious reasons, as we noted earlier), and thus the quantiles of the distribution of Q_4 are precisely the square-root of the quantiles of the distribution of Q_3, the chi-squared distribution with $n-1$ degrees of freedom.

Confidence intervals constructed from pivots are extremely simple and intuitively appealing. Indeed, all of the standard confidence intervals discussed in introductory statistics units (i.e., intervals for population proportions or population averages, or differences thereof, and even those for regression coefficients) are derived from pivotal quantities under the assumption of normal data. However, outside of these familiar cases, it can be extremely difficult to find a pivotal quantity. In this vein, we note that if X_i has a distribution with continuous *CDF* $F_X(x;\theta)$, then the random quantity $Q_i = -\ln\{F_X(X_i;\theta)\}$ has an exponential distribution with unit mean parameter (the demonstration of which is left as an exercise). Therefore, we see that the quantity

$$Q = q(\theta; X_1, \ldots, X_n) = \sum_{i=1}^{n} Q_i = -\sum_{i=1}^{n} \ln\{F_X(X_i;\theta)\} = -\ln\left\{\prod_{i=1}^{n} F_X(X_i;\theta)\right\},$$

has a Gamma distribution with shape parameter n and unit scale parameter. In other words, Q has a distribution which does not depend on θ. Thus, if θ is a scalar and is the parameter of interest, Q is a candidate for a pivot.

Unfortunately, it is usually not the case that the function $q(\cdot; X_1, \ldots, X_n)$ is monotonic in its first argument, and thus Q turns out not to be an appropriate pivotal quantity. We now note, however, that the requirement that pivotal quantities be monotonic in the parameter of interest is not strictly required if we are willing to allow confidence regions which are not intervals. In particular, if $Q = q(\tau; X_1, \ldots, X_n)$ is a quantity whose distribution does not depend on θ, τ is a parameter of interest which takes values in the set $\mathcal{T}$, and $q_{l,\alpha}$ and $q_{u,\alpha}$ are chosen so that $Pr(q_{l,\alpha} \leq Q \leq q_{u,\alpha}) = 1-\alpha$, then the set $B_{X_1,\ldots,X_n} = \{\tau \in \mathcal{T} : q_{l,\alpha} \leq q(\tau; X_1, \ldots, X_n) \leq q_{u,\alpha}\}$ satisfies $Pr_\theta\{\tau(\theta) \in B_{X_1,\ldots,X_n}\}$ for all $\theta \in \Theta$, and is thus a $100(1-\alpha)\%$ confidence region for $\tau = \tau(\theta)$. Of course, $B_{X_1,\ldots,X_n}$ will only be an interval if $q(\tau; X_1, \ldots, X_n)$ is monotone in τ. Moreover, if $q(\tau; X_1, \ldots, X_n)$ is not monotone it can often be extremely difficult to explicitly determine the set $B_{X_1,\ldots,X_n}$. Nonetheless, this is a useful notion, and we return to it in a modified form later.

We note that the pivotal quantity method of confidence interval construction, while intuitive in origin, does not draw directly on our previous discussions of point estimation. It turns out that in the case that $\tau(\theta) = \theta$ (i.e., the parameter of interest is the same as the indexing parameter θ, which thus must be assumed to be a scalar), another approach to constructing exact confidence intervals (i.e., intervals which satisfy Definition 3.1) can be based on any point estimator $T = t(X_1, \ldots, X_n)$, provided we can readily calculate the quantiles of its distribution. In particular, suppose that T has *CDF* $F_T(t;\theta)$, and define the quantities $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ as the solutions to the equations:

$$F_T\{h_{l,\beta}(\theta);\theta\} = \alpha - \beta \qquad \text{and} \qquad F_T\{h_{u,\beta}(\theta);\theta\} = 1-\beta,$$

respectively, for any specified $0 \leq \beta \leq \alpha$. [NOTE: Technically, these equations do not uniquely define the quantities $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ if T has a discrete or mixed distribution, as it is possible for a range of values to satisfy the equations; however, we can overcome this problem by simply defining $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ to be the smallest such solution, and so we will ignore this technicality.] Next, define the sets

$$A_{\theta;\beta} = \{x_1, \ldots, x_n \in I\!R^n : \; h_{l,\beta}(\theta) \leq t(x_1, \ldots, x_n) \leq h_{u,\beta}(\theta)\}$$
$$B_{x_1,\ldots,x_n;\beta} = \{\theta \in \Theta : \; h_{l,\beta}(\theta) \leq t(x_1, \ldots, x_n) \leq h_{u,\beta}(\theta)\}.$$

Clearly, these sets are closely related, the set $A_{\theta;\beta}$ being the set of all possible samples for which the value of our chosen estimator lies between $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ for a given value of θ, while $B_{x_1,\ldots,x_n;\beta}$ is the set of all possible θ values for which the value of our chosen estimator calculated on a specific sample, $x_1, \ldots, x_n$, lies between $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$. In particular, note that the first set is a collection of samples for a fixed θ, while the second is a collection of parameter values for a fixed set of observed data values. Now, from the definition of $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$, it is clear that

$$\begin{aligned} Pr_\theta\{(X_1, \ldots, X_n) \in A_{\theta,\beta}\} &= Pr_\theta\{h_{l,\beta}(\theta) \leq t(X_1, \ldots, X_n) \leq h_{u,\beta}(\theta)\} \\ &= F_T\{h_{u,\beta}(\theta); \theta\} - F_T\{h_{l,\beta}(\theta); \theta\} = (1-\beta) - (\alpha - \beta) = 1 - \alpha. \end{aligned}$$

Next, we note that the event $\{(X_1, \ldots, X_n) \in A_{\theta,\beta}\}$ is clearly equivalent to the event $\{\theta \in B_{X_1,\ldots,X_n;\beta}\}$ (where we must take care to recognise that this second event deals with a fixed value of θ and a *random set*). As such, we see that $Pr_\theta\{\theta \in B_{X_1,\ldots,X_n;\beta}\} = Pr_\theta\{(X_1, \ldots, X_n) \in A_{\theta,\beta}\} = 1 - \alpha$. In other words, $B_{x_1,\ldots,x_n;\beta}$ is a $1 - \alpha$ confidence region for θ (note that we refer to this as a confidence region here, since there is no guarantee that $B_{x_1,\ldots,x_n;\beta}$ will be an interval). Of course, actually determining the set $B_{x_1,\ldots,x_n;\beta}$ can be difficult. However, the determination of $B_{x_1,\ldots,x_n;\beta}$, for this method can be easily described graphically. Suppose, for simplicity, that the functions $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ for some specified value of β are monotonically increasing and can be depicted as in Figure 3.1 below [NOTE: as long as $\alpha < 1$, it is clear that $h_{l,\beta}(\theta) < h_{u,\beta}(\theta)$, since $F_T\{h_{l,\beta}(\theta)\} = \beta < 1 - \alpha + \beta = F_T\{h_{u,\beta}(\theta)\}$ and $F_T(t)$, being a *CDF*, must be an increasing function.]

Figure 3.1

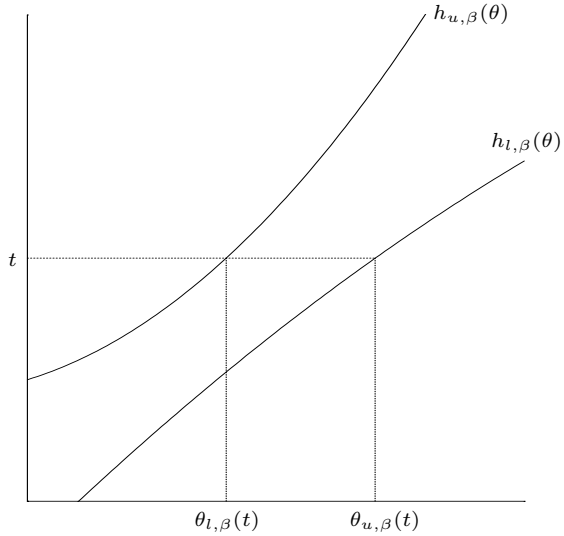


Figure 3.2

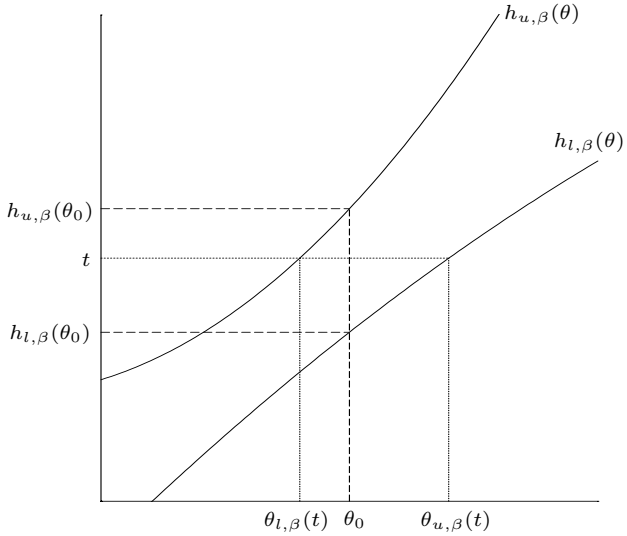


Figure 3.1 also shows the value of our chosen estimator, $t = t(x_1, \ldots, x_n)$, on the vertical axis, as well as the points $\theta_{l,\beta}(t)$ and $\theta_{u,\beta}(t)$ on the horizontal axis corresponding to the relationships:

$$h_{u,\beta}\{\theta_{l,\beta}(t)\} = t; \qquad h_{l,\beta}\{\theta_{u,\beta}(t)\} = t.$$

A little thought shows that $B_{x_1,\ldots,x_n;\beta}$ in this case is precisely the interval $\{\theta_{l,\beta}(t), \theta_{u,\beta}(t)\}$. To see why this is so, we examine Figure 3.2, which denotes the (unknown) "true" parameter as θ_0 as well as the corresponding values of $h_{l,\beta}(\theta_0)$ and $h_{u,\beta}(\theta_0)$. The value of $t = t(x_1, \ldots, x_n)$ shown falls within the two values $h_{l,\beta}(\theta_0)$ and $h_{u,\beta}(\theta_0)$, and we see that this corresponds to θ_0 falling in the interval $\{\theta_{l,\beta}(t), \theta_{u,\beta}(t)\}$. Indeed, further investigation easily shows that if $t = t(x_1, \ldots, x_n)$ falls below $h_{l,\beta}(\theta_0)$, then θ_0 will lie above the corresponding interval $\{\theta_{l,\beta}(t), \theta_{u,\beta}(t)\}$, while if $t = t(x_1, \ldots, x_n)$ falls above $h_{u,\beta}(\theta_0)$, then θ_0 will lie below the corresponding interval $\{\theta_{l,\beta}(t), \theta_{u,\beta}(t)\}$. In other words, $t = t(x_1, \ldots, x_n)$ will lie within $h_{l,\beta}(\theta_0)$ and $h_{u,\beta}(\theta_0)$ [which corresponds to $(x_1, \ldots, x_n) \in A_{\theta_0;\beta}$] if and only if θ_0 lies between $\theta_{l,\beta}\{t(x_1, \ldots, x_n)$ and $\theta_{u,\beta}\{t(x_1, \ldots, x_n)\}$ [which corresponds to $\theta_0 \in B_{x_1,\ldots,x_n;\beta}$].

So, we have constructed a confidence interval for θ based on an estimator T (which may be any estimator for which we are able to determine the *CDF* function explicitly). Of course, the region determined from Figures 3.1 and 3.2 was an interval due to the fact that the functions $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ were assumed to be monotonically increasing. If this had not been the case, then we could still construct our confidence region as $B_{X_1,\ldots,X_n;\beta}$, which corresponds to those values of θ for which $T = t(X_1, \ldots, X_n)$ lies between $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$, but this will not necessarily be an interval any more [you should draw your own version of Figure 3.1 with non-monotonic $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ to see how this works]. Furthermore, just as was the case for the pivotal method, we note that we have a choice of the value β. As before, choosing $\beta = 0$ and $\beta = \alpha$ will lead to lower and upper one-sided intervals, respectively, provided $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ are monotonic. Similarly, the case where $\beta = \alpha/2$ will be "equal-tailed", provided $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ are monotonic. Of course, when $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ are not monotonic, we are not guaranteed that the region $B_{X_1,\ldots,X_n;\beta}$ is even an interval, and thus discussions of one-sided intervals and equal-tailed intervals are not relevant.

Example 3.2: Suppose that $X_1, \ldots, X_n$ are a random sample from a population characterised by a distribution having *CDF* $F_X(x;\theta) = x\theta^{-1}I_{(0\le x\le\theta)}$ for some value $\theta > 0$. In other words, the X_i's are uniformly distributed on the interval $(0, \theta)$. We have seen for this model that $T = \max(X_1, \ldots, X_n)$ is the *MLE* (as well as a sufficient statistic) for the parameter θ. It is also easily shown that

$$\begin{aligned} F_T(T;\theta) = Pr_\theta(T \le t) &= Pr_\theta\{\max(X_1, \ldots, X_n) \le t\} \\ &= Pr_\theta(X_1 \le t, \ldots, X_n \le t) \\ &= Pr_\theta(X_1 \le t) \cdots Pr_\theta(X_n \le t) \\ &= \{F_X(t;\theta)\}^n \\ &= \frac{t^n}{\theta^n} I_{(0\le t\le\theta)}. \end{aligned}$$

So, for a given β, we see that:

$$\begin{aligned} F_T\{h_{l,\beta}(\theta);\theta\} = \alpha - \beta \qquad &\Longrightarrow \qquad \frac{\{h_{l,\beta}(\theta)\}^n}{\theta^n} I_{\{0\le h_{l,\beta}(\theta)\le\theta\}} = \alpha - \beta \\ &\Longrightarrow \qquad h_{l,\beta}(\theta) = \theta(\alpha - \beta)^{1/n}, \end{aligned}$$

(which is clearly less than θ, since $\alpha - \beta < 1$, and thus satisfies the requirement of the indicator function). An identical calculation shows that $h_{u,\beta}(\theta) = \theta(1 - \beta)^{1/n}$. Note that these functions are clearly monotonically increasing in θ (and indeed they are linearly so). Figures 3.3 and 3.4 show plots of these two functions versus θ for two different choices of β (the β used in Figure 3.4 is smaller than that used in Figure 3.3):

Figure 3.3

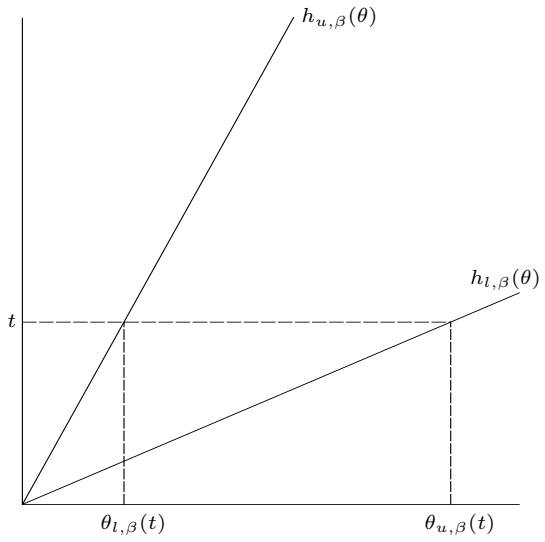
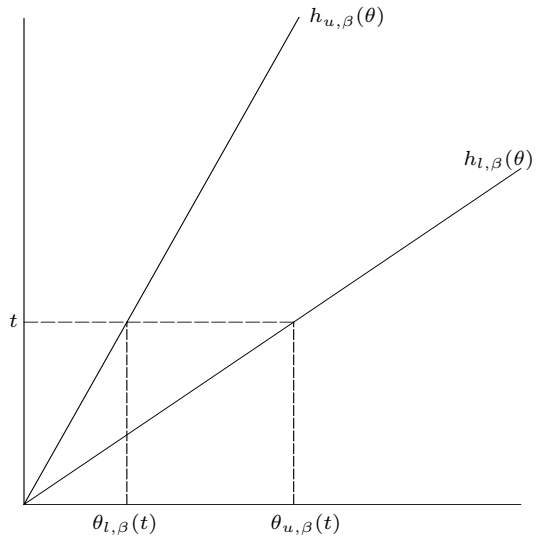


Figure 3.4



Further, Figures 3.3 and 3.4 show the construction of a confidence interval for θ based on the observed value $t = \max(x_1, \ldots, x_n)$ of the estimator T. Since $h_{l,\beta}(\theta)$ and $h_{u,\beta}(\theta)$ are monotonically increasing in this case, we can construct our $100(1-\alpha)\%$ confidence interval as $\{\theta_{l,\beta}(T), \theta_{u,\beta}(T)\}$ where:

$$h_{u,\beta}(\theta_{l,\beta}) = T \qquad \Longrightarrow \qquad \theta_{l,\beta}(1-\beta)^{1/n} = T \qquad \Longrightarrow \qquad \theta_{l,\beta} = T(1-\beta)^{-1/n}$$

and

$$h_{l,\beta}(\theta_{u,\beta}) = T \qquad \Longrightarrow \qquad \theta_{u,\beta}(\alpha-\beta)^{1/n} = T \qquad \Longrightarrow \qquad \theta_{u,\beta} = T(\alpha-\beta)^{-1/n}.$$

Now, we note that the length of the confidence interval is $T\{(\alpha-\beta)^{-1/n} - (1-\beta)^{-1/n}\}$, which is minimised when $\beta = 0$ (a fact which is left as an exercise); that is, the shortest interval in this case is the lower one-sided interval (note that Figure 3.4 foreshadows this result, since we see that a smaller β leads to a shorter interval). Also, note that the lower one-sided interval in this case has a lower endpoint of $T = \max(X_1, \ldots, X_n)$ (as opposed to the more common values of 0 or $-\infty$), since this is clearly the smallest possible value that θ can assume (since the set of possible values, or *support*, of the distribution of the X_i's is determined by θ).

We end this example by noting that, in this case, it turns out that the quantity $Q = T\theta^{-1}$ is a pivot, since

$$F_Q(q) = Pr(T\theta^{-1} \leq q) = Pr(T \leq q\theta) = \frac{(q\theta)^n}{\theta^n} I_{(0 \leq q\theta \leq \theta)} = q^n I_{(0 \leq q \leq 1)}.$$

As such, the confidence intervals derived here can also be arrived at directly using the pivotal method based on Q (the fact that this method arrives at the same intervals is left as an exercise). This approach is certainly easier to implement, however, it required us to first determine Q and show it was a pivot (admittedly not a very difficult exercise in this case).

In closing, we note that, in principal we can always construct an interval based on a given estimator $T = t(X_1, \ldots, X_n)$ using the method discussed here. However, in practice, this requires us to be able to calculate $F_T(t; \theta)$, which will generally be extremely complicated. As such, we will likely need computer-based solutions to find the required *CDF* and this makes this method somewhat cumbersome and unattractive. Moreover, we now have the issue of which estimation method to use as our starting point (of course, one option would be to search for an estimator which made

finding the required *CDF* an easy task, but this is a rather poor statistical justification). Finally, we note that this approach is only appropriate when the parameter of interest is θ [though, it can be modified to handle a general parameter $\tau = \tau(\theta)$ provided the chosen estimator T has a distribution which only depends on the parameter τ, as opposed to θ; but, such estimators are not always readily available]. So, we see that, unless we are in a rather standard situation where a pivotal quantity is readily available, we cannot really hope to find straightforward exact confidence intervals. As such, we turn our attention to finding "approximate" confidence intervals, which are simply intervals which satisfy Definition 3.1 asymptotically (i.e., as the sample size, n, tends to infinity).

3.2.2. Asymptotic Methods: In the preceding section, we discussed "exact" confidence intervals based on either pivotal quantities or on statistics with known distributions. In general, though, while pivots have simple distributions they are extremely difficult to determine outside of the standard scenarios dealt with in introductory units. On the other hand, estimators can usually be determined reasonably easily (though perhaps not optimally) for any given parameter of interest, but in general their exact distributions are extremely complicated except in a few special cases. As such, both of the "exact" confidence interval methods of the previous section are of somewhat limited use in more advanced settings. As an alternative, it is often easier to pursue confidence interval procedures which are approximate, rather than exact. In other words, we want to find simple statistics whose distributions can be approximated (usually in an asymptotic sense, so that the distribution in question is approximated by a limiting distribution appropriate when the sample size, n, is extremely large). In this way, we overcome the need to find "exact" pivots as well as the need to determine complicated "exact" distributions for estimators. Of course, the consequence of such an endeavour is that the confidence intervals so constructed are only asymptotically accurate. That is, they do not satisfy Definition 3.1, but rather satisfy the weaker condition that

$$\lim_{n\to\infty} Pr_\theta\{T_{l,n} \leq \tau(\theta) \leq T_{u,n}\} = 1 - \alpha, \qquad \forall\theta \in \Theta,$$

where $T_{l,n} = t_{l,n}(X_1, \ldots, X_n)$ and $T_{u,n} = t_{u,n}(X_1, \ldots, X_n)$ are confidence limits based on a specific sample of size n. The idea here is best illustrated by the standard approach to finding confidence intervals for population means. In such a setting, we know that the quantity $Z = \sqrt{n}(\overline{X} - \mu)/S$ is pivotal in the case of normally distributed X_i's. However, if the X_i's have some other distribution, then Z is no longer "exactly" Student's t-distributed, and indeed is no longer pivotal at all. Fortunately, the Central Limit Theorem provides us with a way of extending the result. Using this result, we know that the distribution of Z is "approximately" normal, indeed we know that it is asymptotically standard normal as outlined in Definition 2.5. Using this fact, we see that the quantity Z is approximately pivotal for large values of n, regardless of the original distribution of the X_i's (provided that distribution is not too unusual; that is, provided its variance is not infinite). As such, we can employ the pivotal method to Z using its limiting distribution, and thus arrive at an approximate confidence interval for the population mean.

Formally, suppose that we have a sequence of random quantities $Q_1 = q_1(\tau; X_1, \ldots, X_n), Q_2 = q_2(\tau; X_1, \ldots, X_n), \ldots$ with *CDF*s $F_1(q;\theta)$, $F_2(q;\theta)$, ..., such that $\lim_{n\to\infty} F_n(q;\theta) = F(q)$ for some known *CDF* $F(q)$ which does not depend on the parameter θ, and is not the *CDF* of a point mass distribution [i.e., is not of the form $F(q) = I_{(q\geq c)}$ for a constant c.] Note also, that we only require $F(q)$ to be known to us exactly, while the $F_i(q;\theta)$'s need only be defined in principle. Such a sequence is often termed *asymptotically pivotal*. Based on this idea, we can define an approximate $100(1-\alpha)\%$ confidence interval $B_{X_1,\ldots,X_n;\beta} = \{\tau \in \mathcal{T} : q_{l,\alpha} \leq q_n(\tau; X_1, \ldots, X_n) \leq q_{u,\alpha}\}$, where $q_{l,\alpha} = F^{-1}(\alpha - \beta)$, $q_{u,\alpha} = F^{-1}(1-\beta)$, β is any value between 0 and α, and $\mathcal{T}$ is the set of possible values for $\tau = \tau(\theta)$.

Of course, we might at this stage ask what has been gained, since we have merely replaced the task of finding a pivot with the task of finding an asymptotic pivot. However, it turns out that for most parametric probability models, an easy asymptotic pivot exists. In particular, if θ is a scalar, $\hat{\theta}$ is the *MLE* of θ and $\hat{\tau} = \tau(\hat{\theta})$ is the *MLE* for the parameter of interest $\tau = \tau(\theta)$, then the quantity $Q_n = (\hat{\tau} - \tau)/\sqrt{\sigma^2(\hat{\tau})}$ is approximately standard normally distributed with

$$\sigma^2(\hat{\tau}) = \{\tau'(\hat{\theta})\}^2\{I(\hat{\theta})\}^{-1},$$

where $I(\theta)$ is the expected Fisher information for the chosen probability model; that is, $I(\theta) = -E\{l''(\theta)\}$, where $l''(\theta)$ is the second derivative of the log-likelihood function $l(\theta)$; see Section 2.4.1. [NOTE: In the case that θ is vector-valued, the variance must be defined in terms of the expected Fisher information matrix, $I(\theta)$, as $\sigma^2\{\tau(\hat{\theta})\}$ where

$$\sigma^2\{\tau(\theta)\} = \{(\nabla\tau)^T I^{-1}(\theta)\nabla\tau\}^{-1},$$

and $\nabla\tau = \left\{\frac{\partial}{\partial\theta_1}\tau(\theta), \ldots, \frac{\partial}{\partial\theta_k}\tau(\theta)\right\}^T$ is the gradient vector (written as a column) of $\tau(\theta)$ and $I^{-1}(\theta)$ is the matrix inverse of the expected Fisher information matrix; see Remark *(i)* at the end of Section 2.4.1.]

To see why this is so, suppose that we have a sample, $X_1, \ldots, X_n$, from a distribution having density function $f_X(x;\theta)$ indexed by a scalar parameter $\theta \in \Theta$, so that the log-likelihood for θ based on the sample is given by $l(\theta) = \sum_{i=1}^n \ln\{f_X(X_i;\theta)\} = \sum_{i=1}^n l_i(\theta)$, where we have defined $l_i(\theta) = \ln\{f_X(X_i;\theta)\}$ to be the log-likelihood for θ based on the single observation X_i. Now, we note that the maximum likelihood estimate for θ is the solution to the equation $l'(\theta) = 0$ (provided, of course, that this equation has a solution within Θ, which we will assume throughout the remainder of this section). In other words, the maximum likelihood estimate, $\hat{\theta}$, satisfies $U(\hat{\theta}) = 0$, where the $U(\theta) = l'(\theta)/\sqrt{n}$ is called the *score statistic* or *score function*. Moreover, we see that $U(\theta) = n^{-1/2}\sum_{i=1}^n l_i'(\theta) = \sqrt{n}\overline{l'}$, where $\overline{l'} = n^{-1}\sum_{i=1}^n l_i'(\theta)$ and $l_i'(\theta) = \frac{\partial}{\partial\theta}\ln\{f_X(X_i;\theta)\} = \frac{\partial}{\partial\theta}f_X(X_i;\theta)/f_X(X_i;\theta)$. Now, since

$$E_\theta\{l_i'(\theta)\} = \int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial\theta}f_X(x;\theta)}{f_X(x;\theta)} f_X(x;\theta)dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}f_X(x;\theta)dx = \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f_X(x;\theta)dx = \frac{\partial}{\partial\theta}(1) = 0,$$

and $Var_\theta\{l_i'(\theta)\} = E_\theta[\{l_i'(\theta)\}^2] = -E_\theta\{l_i''(\theta)\} = i(\theta)$ (a fact which is left as an exercise; see Section 2.4.1), we see that $U(\theta)/\sqrt{i(\theta)} = \sqrt{n}(\overline{l'} - 0)/\sqrt{i(\theta)}$ is a standardised average and the Central Limit Theorem guarantees that $U(\theta)/\sqrt{i(\theta)}$ is asymptotically standard normal.

Next, we note that a Taylor expansion shows [NOTE: the details of the Taylor series approximations require mathematics which are beyond the scope of these notes and are thus omitted here; however, they are based on the fact that *MLE*s can be shown to be consistent, so that as the sample size increases, the discrepancy between an *MLE*, $\hat{\theta}$, and its target value, θ, goes to zero, meaning that $(\hat{\theta} - \theta)^2$ is essentially negligible]:

$$\begin{aligned} 0 = U(\hat{\theta}) &\approx U(\theta) + U'(\theta)(\hat{\theta} - \theta) \approx U(\theta) + E_\theta\{U'(\theta)\}(\hat{\theta} - \theta) = U(\theta) - \sqrt{n}i(\theta)(\hat{\theta} - \theta) \\ &\approx U(\theta) - \sqrt{n}i(\hat{\theta})(\hat{\theta} - \theta), \end{aligned}$$

where we have used the fact that $i(\theta) = -E_\theta\{l_i''(\theta)\}$ (see Section 2.4.1) which implies

$$E_\theta\{U'(\theta)\} = \frac{1}{\sqrt{n}}E_\theta\left\{\sum_{i=1}^n l_i''(\theta)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n E_\theta\{l_i''(\theta)\} = -\sqrt{n}i(\theta),$$

and the fact (again using Taylor expansion) that $i(\theta) \approx i(\hat{\theta}) + i'(\hat{\theta})(\theta - \hat{\theta}) = i(\hat{\theta}) - i'(\hat{\theta})(\hat{\theta} - \theta)$ which implies that

$$i(\theta)(\hat{\theta} - \theta) \approx i(\hat{\theta})(\hat{\theta} - \theta) - i'(\hat{\theta})(\hat{\theta} - \theta)^2 \approx i(\hat{\theta})(\hat{\theta} - \theta)$$

[again based on the fact that $\hat{\theta}$ is consistent, which means that $(\hat{\theta} - \theta)^2$ is essentially negligible].

Now, using these results, we see that

$$\sqrt{ni(\hat{\theta})}(\hat{\theta} - \theta) \approx \sqrt{ni(\theta)}(\hat{\theta} - \theta) \approx \frac{U(\theta)}{\sqrt{i(\theta)}} \quad \sim \quad N(0, 1).$$

Again using Taylor expansion, we have:

$$\tau(\theta) \approx \tau(\hat{\theta}) + \tau'(\hat{\theta})(\theta - \hat{\theta}) \approx \hat{\tau} - \tau'(\hat{\theta})(\hat{\theta} - \theta),$$

which implies that $\hat{\tau} - \tau(\theta) \approx \tau'(\hat{\theta})(\hat{\theta} - \theta)$, so that:

$$Q_n = q_n(\tau; X_1, \ldots, X_n) = \frac{\hat{\tau} - \tau(\theta)}{\sqrt{\sigma^2(\hat{\tau})\}} = \frac{\sqrt{ni(\hat{\theta})}\{\hat{\tau} - \tau(\theta)\}}{\tau'(\hat{\theta})} \approx \sqrt{ni(\hat{\theta})}(\hat{\theta} - \theta) \quad \sim \quad N(0, 1).$$

Formally, then, an approximate $100(1 - \alpha)\%$ confidence interval for a scalar parameter $\tau = \tau(\theta)$ can be constructed from the *MLE*, $\hat{\tau} = \tau(\hat{\theta})$, as the set

$$\begin{aligned}
B_{X_1,\ldots,X_n;\beta} &= \{\tau \in \mathcal{T} : \Phi^{-1}(\alpha - \beta) \leq q_n(\tau; X_1, \ldots, X_n) \leq \Phi^{-1}(1 - \beta)\} \\
&= \left\{\tau \in \mathcal{T} : \Phi^{-1}(\alpha - \beta) \leq \frac{\sqrt{ni(\hat{\theta})}\{\hat{\tau} - \tau(\theta)\}}{\tau'(\hat{\theta})} \leq \Phi^{-1}(1 - \beta)\right\} \\
&= \left\{\tau \in \mathcal{T} : \hat{\tau} - \Phi^{-1}(1 - \beta)\frac{\tau'(\hat{\theta})}{\sqrt{ni(\hat{\theta})}} \leq \tau \leq \hat{\tau} - \Phi^{-1}(\alpha - \beta)\frac{\tau'(\hat{\theta})}{\sqrt{ni(\hat{\theta})}}\right\} \\
&= \{\tau \in \mathcal{T} : \hat{\tau} - \Phi^{-1}(1 - \beta)\sqrt{\sigma^2(\hat{\tau})} \leq \tau \leq \hat{\tau} - \Phi^{-1}(\alpha - \beta)\sqrt{\sigma^2(\hat{\tau})}\} \\
&= \{\tau \in \mathcal{T} : \hat{\tau} - \Phi^{-1}(1 - \beta)\sqrt{\sigma^2(\hat{\tau})} \leq \tau \leq \hat{\tau} + \Phi^{-1}(1 - \alpha + \beta)\sqrt{\sigma^2(\hat{\tau})}\},
\end{aligned}$$

where we have used the symmetry of the standard normal distribution to conclude that $\Phi^{-1}(\alpha - \beta) = -\Phi^{-1}(1 - \alpha + \beta)$. As always, we have a choice of β. When $\beta = 0$ or $\beta = \alpha$ we arrive at lower and upper one-sided intervals, respectively. If $\beta = \alpha/2$ we arrive at an equal-tailed interval, which in this case is symmetric (due to the symmetry of the normal distribution) and has the shortest length among all choices for β. Note that this latter interval also has the standard form associated with confidence intervals; namely, an estimator plus or minus some number, $\Phi^{-1}(1 - \alpha/2)$, of standard deviations.

Example 3.3: Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with parameter θ, so that $f_X(x; \theta) = \theta e^{-\theta x}$, for $\theta > 0$. Suppose that we are interested in estimating $\tau(\theta) = \theta^{-1}$ (which is the expectation of the exponential distribution). In this case, it is straightforward to show that the log-likelihood function is given by:

$$l(\theta) = \sum_{i=1}^{n} \ln\{f_X(X_i; \theta)\} = n \ln(\theta) - \theta \sum_{i=1}^{n} X_i = n \ln(\theta) - n\theta\overline{X}.$$

Taking derivatives yields:

$$l'(\theta) = \frac{n}{\theta} - n\overline{X}; \qquad l''(\theta) = -\frac{n}{\theta^2}.$$

Therefore, we see that $\hat{\theta} = 1/\overline{X}$ which implies that $\hat{\tau} = \overline{X}$, and we see that $I(\theta) = n\theta^{-2}$. Furthermore, $\tau'(\theta) = -\theta^{-2}$, which implies that

$$\sigma^2(\hat{\tau}) = (\hat{\theta}^{-4})(n\hat{\theta}^{-2})^{-1} = n^{-1}\hat{\theta}^{-2} = n^{-1}(\overline{X})^2.$$

As such, we see that an equal-tailed $100(1-\alpha)\%$ approximate confidence interval for τ is given by:

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

In closing, we note that since $\hat{\tau} = \overline{X}$ in this case, the Central Limit Theorem can also be used to arrive at the standard approximate confidence interval for τ of:

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

where S is the usual sample standard deviation. If the data truly are exponentially distributed, than this latter interval will not be as accurate as the former as we shall discuss briefly in Section 3.3. (Of course, the latter interval will be valid even in the case that the data are not truly exponentially distributed.)

In the preceding example, we constructed a confidence interval for the parameter $\tau = \tau(\theta) = \theta^{-1}$. Suppose, however, that we are now interested in the parameter $\theta = \tau^{-1}$ itself. One simple way to arrive at a confidence interval for θ would be to appropriately transform the endpoints of the interval for τ. In other words, if (l, u) was the interval for τ, then we could use the interval (u^{-1}, l^{-1}) for θ. Alternatively, we could construct an interval directly for θ, based on $\hat{\theta} = (\overline{X})^{-1}$ and $\sigma^2(\hat{\theta}) = \{ni(\theta)\}^{-1}$. For the exponential distribution of Example 3.3, this leads to the two possible intervals:

$$\left[\left\{\overline{X} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\overline{X}}{\sqrt{n}}\right\}^{-1}, \left\{\overline{X} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\overline{X}}{\sqrt{n}}\right\}^{-1}\right]$$

and

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

Clearly, these two intervals are not the same. This raises an important issue regarding the current asymptotic approach to confidence interval construction; namely, it is not "functionally equivariant". In other words, the confidence region we arrive at for a new parameter τ is not simply the appropriate transformation of the original values in the confidence region for θ (recall that this was an important property of *MLE*s themselves; see Theorem 2.1 in Section 2.1.2). Of course, it can be shown that the above two intervals are approximately equal; however, the question arises as to which one is better. This is a complicated question, and we will not be able to answer it here; however, we do point out that the answer is dependent upon the circumstances of the distributions involved. The idea here is that the normal approximation to the distribution of $\sqrt{n}\{\tau(\hat{\theta}) - \tau(\theta)\}$ will be better (i.e., more closely normal) for some functions $\tau(\theta)$ than for others, and it is often difficult to know which functions of θ are the "most normal".

Given the lack of functional equivariance of the current approximate procedure, it seems sensible to see if we can find a simple alternative which does possess the equivariance property. It turns out that such a procedure is possible, as long as the functional transformation in question is invertible. A transformation of θ which is invertible is often referred to as a *reparameterisation*,

and procedures which are equivariant with respect to invertible transformations are termed *parameterisation equivariant*. We now give a brief outline of an asymptotic likelihood-based confidence interval procedure which is parameterisation equivariant. Specifically, this means that if we find a confidence region, C, for θ based on this new procedure and transform all of its values [which we sometimes denote as $\tau(C) = \{\tau(\theta) : \theta \in C\}$] then we will arrive at the same confidence region as if we had applied our new procedure to the parameter τ directly.

To define our new procedure, we first suppose that θ is the (scalar) parameter of interest. Again, Taylor series expansions show that:

$$l(\theta) \approx l(\hat{\theta}) + l'(\hat{\theta})(\theta - \hat{\theta}) + \tfrac{1}{2}l''(\hat{\theta})(\theta - \hat{\theta})^2 \approx l(\hat{\theta}) - \tfrac{1}{2}ni(\theta)(\hat{\theta} - \theta),$$

where we have used the fact that $l'(\hat{\theta}) = 0$ [again, provided the *MLE* solves the score equations; if the *MLE* lies on the boundary of the parameter space, other more complicated mathematical procedures are required]. Now, the preceding approximation shows that

$$2\{l(\hat{\theta}) - l(\theta)\} \approx ni(\theta)(\hat{\theta} - \theta)^2 = \left\{\sqrt{ni(\theta)}(\hat{\theta} - \theta)\right\}^2,$$

and we know that $\sqrt{ni(\theta)}(\hat{\theta} - \theta)$ has an asymptotic standard normal distribution. Therefore, $2\{l(\hat{\theta}) - l(\theta)\}$ will have an asymptotic chi-squared distribution with one degree of freedom. As such, we can construct an alternative approximate $100(1-\alpha)\%$ confidence interval for θ as:

$$C = \{\theta \in \Theta : 2\{l(\hat{\theta}) - l(\theta)\} \leq \chi^2_{(1)}(1-\alpha)\},$$

where $\chi^2_{(1)}(1-\alpha)$ is the $(1-\alpha)$-quantile of the chi-squared distribution with one degree of freedom.

Next, we note that if we are interested in some other parameter $\tau = \tau(\theta)$ then we must construct the confidence interval based on the induced likelihood function $M(\tau; x_1, \ldots, x_n)$ defined in Theorem 2.1 of Section 2.1.2. In general, recall that the induced likelihood is $M(\tau) = \sup_{\theta:\, \tau(\theta)=\tau} L(\theta; x_1, \ldots, x_n)$ and thus an asymptotic $100(1-\alpha)\%$ confidence interval for τ can then be construct as:

$$C_\tau = \{\tau \in \mathcal{T} : 2\{m(\hat{\tau}) - m(\tau)\} \leq \chi^2_{(1)}(1-\alpha)\},$$

where $\hat{\tau} = \tau(\hat{\theta})$ (since *MLE*s are functionally equivariant) and $m(\tau) = \ln\{M(\tau)\}$ is the induced log-likelihood. Now, if we assume that the transformation $\tau(\theta)$ is invertible, so that there exists an inverse function $\theta^{-1}(\tau)$ such that $\theta^{-1}\{\tau(\theta)\} = \theta$, then the set $\{\theta : \tau(\theta) = \tau\}$ has only the single element $\theta^{-1}(\tau)$ and thus the supremum operation is not required. Therefore, we see that $M(\tau) = L\{\theta^{-1}(\tau)\}$, implying $m(\tau) = l\{\theta^{-1}(\tau)\}$, and thus the set C_τ becomes

$$C_\tau = \{\tau \in \mathcal{T} : 2[l(\hat{\theta}) - l\{\theta^{-1}(\tau)\}] \leq \chi^2_{(1)}(1-\alpha)\},$$

since $m(\hat{\tau}) = l\{\theta^{-1}(\hat{\tau})\} = l[\theta^{-1}\{\tau(\hat{\theta})\}] = l(\hat{\theta})$. From this definition, it clearly follows that if $\tau \in C_\tau$ then $\theta = \theta^{-1}(\tau) \in C$. In other words, for every τ value in C_τ there is a corresponding θ value in C, which is precisely the definition of the set $\tau(C) = \{\tau(\theta) : \theta \in C\}$. In other words, $C_\tau = \tau(C)$, which means that the confidence interval defined by C does possess the property of parameterisation equivariance, since we can now see that C_τ is just an appropriate transformation of C.

Example 3.3 *(cont'd)*: We saw that for a sample from an exponential distribution with parameter θ, the log-likelihood function was:

$$l(\theta) = \sum_{i=1}^{n} \ln\{f_X(X_i; \theta)\} = n\ln(\theta) - \theta\sum_{i=1}^{n} X_i = n\ln(\theta) - n\theta\overline{X},$$

and the *MLE* of θ was $\hat{\theta} = (\overline{X})^{-1}$. Thus, $l(\hat{\theta}) = -n\ln(\overline{X}) - n$, and we see that an approximate $100(1-\alpha)\%$ confidence interval for θ is given by:

$$C = \{\theta > 0 : \ -2n\ln(\theta\overline{X}) + 2n(\theta\overline{X} - 1) \leq \chi^2_{(1)}(1-\alpha)\}.$$

Unfortunately, the relationship $-2n\ln(\theta\overline{X}) + 2n(\theta\overline{X} - 1) \leq \chi^2_{(1)}(1-\alpha)$ cannot be solved for θ explicitly and requires computer assistance to determine the endpoints of the confidence interval, $C = (\hat{\theta}_{l,C}, \hat{\theta}_{u,C})$. However, the interval can be investigated graphically by plotting the function $\Lambda(\theta) = -2n\ln(\theta\overline{X}) + 2n(\theta\overline{X} - 1)$ and examining the horizontal cut-off line at $\chi^2_{(1)}(1-\alpha)$, which is shown in Figure 3.5 below:

Figure 3.5

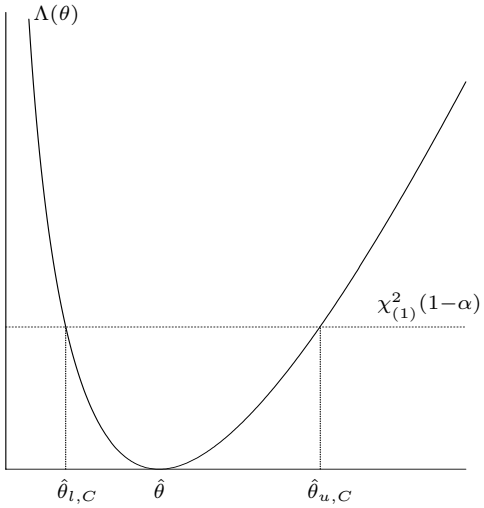
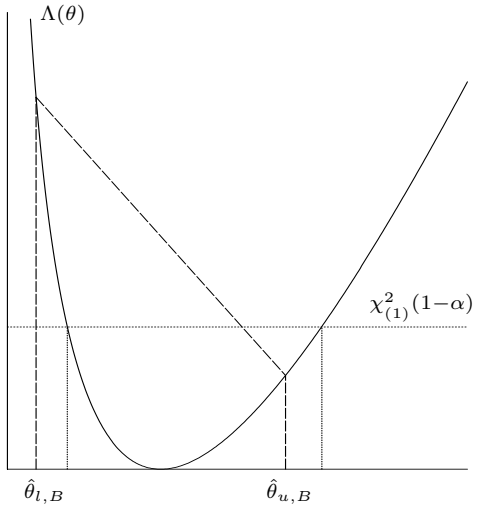


Figure 3.6



For comparison, the interval based on the normal approximation to the distribution of $\hat{\theta} = (\overline{X})^{-1}$, which we saw was:

$$(\hat{\theta}_{l,B}, \hat{\theta}_{u,B}) = \left\{(\overline{X})^{-1} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{1}{\sqrt{n\overline{X}}}, (\overline{X})^{-1} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{1}{\sqrt{n\overline{X}}}\right\},$$

is shown in Figure 3.6. Clearly, the two intervals differ (though, the difference here is not necessarily indicative of the usual difference between these two approaches). Moreover, we saw that the normal-based approach was not functionally equivariant. However, if we change our focus to estimation of $\tau = \tau(\theta) = \theta^{-1}$, we see that $\hat{\tau} = \overline{X}$ and the induced log-likelihood is given by:

$$m(\tau; x_1, \ldots, x_n) = l\{\theta^{-1}(\tau)\} = l(\tau^{-1}) = -n\ln(\tau) - n\frac{\overline{X}}{\tau},$$

[NOTE: clearly the relationship $\tau = \theta^{-1}$ is invertible] which leads to the approximate confidence interval for τ:

$$C_\tau = \{\tau > 0 : \ -2n\ln(\overline{X}\tau^{-1}) + 2n(\overline{X}\tau^{-1} - 1) \leq \chi^2_{(1)}(1-\alpha)\}.$$

It is easily seen that $\tau(\theta) \in C_\tau$ if and only if $\theta = \tau^{-1} \in C$. In other words, C_τ is just the set C transformed appropriately, which means that if $C = (\hat{\theta}_{l,C}, \hat{\theta}_{u,C})$ then $C_\tau = (\hat{\tau}_{l,C_\tau}, \hat{\tau}_{u,C_\tau}) = \tau\{(\hat{\theta}_{l,C}, \hat{\theta}_{u,C})\} = (\hat{\theta}^{-1}_{u,C}, \hat{\theta}^{-1}_{l,C})$, so that the intervals based on the asymptotic chi-squared pivot, $2\{l(\hat{\theta}) - l(\theta)\}$ have the desired parameterisation equivariance property.

Of course, as we saw in the preceding example, the intervals based on the asymptotic chi-squared pivot are not as easily determined explicitly as the intervals based on the asymptotic normal

pivots, $\sqrt{n}(\hat{\tau} - \tau)$ [NOTE: they may not even be intervals if the log-likelihood function is not unimodal]. In general, they will require computer-assisted solutions. Nonetheless, they possess the intuitively appealing parameterisation equivariance property, and can be considered preferable on those grounds. Moreover, we can see from Figures 3.5 and 3.6, that the intervals based on the asymptotic chi-squared pivot contain the values of θ which have the highest likelihoods [i.e., likelihoods as close as possible to $l(\hat{\theta})$, the maximum possible likelihood]. In other words, if $\theta_1 \in C$ and $l(\theta_2) > l(\theta_1)$ then $\theta_2 \in C$, so that all the elements of C have higher likelihoods than those elements of C^c, the complement of C. Clearly, the same cannot be said for the intervals based on the asymptotic normal pivots, as can be seen in Figure 3.6. Of course, if the sample size, n, is sufficiently large, the difference between these two approaches becomes negligible.

3.2.3. Bayesian Intervals: The discussion at the end of the preceding section noted that intervals based on the asymptotic chi-squared pivot, $2\{\hat{l(\theta)} - l(\theta)\}$, contain those elements of Θ with the largest likelihoods. This notion should remind us somewhat of the ideas underlying Bayesian point estimation, where we determined a posterior distribution for θ, which encapsulated the "likelihood" of any θ value based on the information contained in the data and the prior distribution. Combining these ideas leads us to the notions of Bayesian *credibility* and *highest posterior density* regions.

Formally, suppose that we have observed a sample $X_1, \ldots, X_n$ from a distribution with density function $f_X(x;\theta)$ and joint likelihood function $L(\theta; X_1, \ldots, X_n)$ [which is just the product of the density function evaluated at each of the data values, X_i]. Further, suppose that we have some prior belief about the likely value of θ, which we encapsulate in a prior distribution, $\pi(\theta)$. We saw in Section 2.5, that the posterior distribution, which incorporated our prior beliefs about θ and the information contained in the sample, was determined as:

$$\pi(\theta|X_1, \ldots, X_n) = \frac{L(\theta; X_1, \ldots, X_n)\pi(\theta)}{\int_\Theta L(t; X_1, \ldots, X_n)\pi(t)dt}.$$

Using this posterior distribution, we can define a $100(1-\alpha)\%$ *Bayesian credibility region* for θ as any set C for which

$$Pr_{\pi(\theta|x_1,\ldots,x_n)}(C) = \int_C \pi(\theta|x_1, \ldots, x_n)d\theta = 1 - \alpha.$$

Obviously, there will be many possible choices of sets C which satisfy this definition (just as there were many possible choices for confidence intervals based on the selection of a β value). There are, however, three common choices. The most common choice is to select C as the interval (θ_l, θ_u) where the endpoints of the interval satisfy

$$\int_{-\infty}^{\theta_l} \pi(\theta|x_1, \ldots, x_n)d\theta = \alpha/2; \qquad \int_{\theta_u}^{\infty} \pi(\theta|x_1, \ldots, x_n)d\theta = \alpha/2,$$

leading to an *equal-tailed* Bayesian interval estimate. Alternatively, we might choose θ_l and θ_u so as to minimise the length of the interval, $\theta_u - \theta_l$. Finally, we might choose C to be a so-called *highest posterior density (HPD)* Bayesian credibility region. To construct *HPD* regions, we define C to be that set with posterior probability $1 - \alpha$ which satisfies the criterion:

$$\theta_1 \in C \quad \text{and} \quad \pi(\theta_2|x_1, \ldots, x_n) > \pi(\theta_1|x_1, \ldots, x_n) \qquad \Longrightarrow \qquad \theta_2 \in C.$$

In other words, C contains the values of θ which have the highest posterior density values, so that we can determine *HPD* regions as the set

$$C = \{\theta \in \Theta : \pi(\theta|x_1, \ldots, x_n) \geq c_\alpha\},$$

where c_α is determined so that $Pr_{\pi(\theta|x_1,\ldots,x_n)}(C) = 1-\alpha$. Of course, this sort of region can be difficult to determine explicitly in practice; however, noting the similarity of its definition to that of the confidence interval based on the asymptotic chi-squared pivot, $2\{l(\hat{\theta}) - l(\theta)\}$, we can see that a graphical determination of these intervals is possible by simply plotting the posterior density and then drawing a horizontal line at height c_α, thus determining the *HPD* region as those value of θ for which the posterior density value falls above the horizontal cut-off line. We note that *HPD* regions need not be intervals. However, if the *HPD* region in a particular case is an interval (which it must be if the posterior density can be shown to be unimodal), then it can be shown that the *HPD* region is equivalent to the Bayesian interval estimate with shortest length.

Example 3.4: Suppose that $X_1, \ldots, X_n$ are a sample from a normal distribution with mean θ and unit variance. Further, suppose the prior distribution on θ is also normal, but with a (known) mean of δ and unit variance. It can be easily shown (and the demonstration is left as an exercise) that the posterior distribution of θ is normal with mean $\frac{n}{n+1}\overline{X} + \frac{1}{n+1}\delta$ and variance $\frac{1}{n+1}$. As such, we can find an equal-tailed $100(1-\alpha)\%$ Bayesian interval estimate by selecting two values θ_l and θ_u which satisfy:

$$1 - \Phi\left(\frac{\theta_u - \frac{n}{n+1}\overline{X} - \frac{1}{n+1}\delta}{\sqrt{\frac{1}{n+1}}}\right) = \Phi\left(\frac{\theta_l - \frac{n}{n+1}\overline{X} - \frac{1}{n+1}\delta}{\sqrt{\frac{1}{n+1}}}\right) = \frac{\alpha}{2}.$$

Since the normal distribution is symmetric, the equal-tailed interval will be the same as the shortest possible interval here, and it can easily be seen to be:

$$\theta_l = \frac{n\overline{X}+\delta}{n+1} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n+1}}; \quad \theta_u = \frac{n\overline{X}+\delta}{n+1} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n+1}}.$$

Moreover, since the normal distribution is unimodal, this interval will also by the *HPD* region.

In closing, we recall that usual confidence interval for the mean of a normal distribution with known unit variance is given by:

$$\overline{X} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{1}{\sqrt{n}}.$$

If we rewrite the Bayesian interval determined here as:

$$\frac{1}{n+1}\left(\delta + \sum_{i=1}^{n} X_i\right) \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{1}{\sqrt{n+1}},$$

we see that it has the same general form as the usual confidence interval, except that it appears to be based on an "additional observation" which is simply δ, the mean of the prior distribution. In this way, we can see directly how the Bayesian interval incorporates the additional information contained in the prior distribution into the interval estimate of θ for this example (of course, such direct insights will not often be possible in more complex problems).

Finally, we close this brief section by noting that Bayesian credibility regions can easily be constructed for other parameters of interest, $\tau = \tau(\theta)$, by simply suitably transforming the posterior distribution of θ to arrive at the posterior distribution for $\tau(\theta)$ using the change of variable formulae for densities [of course, this procedure can be complicated if the function $\tau(\theta)$ is complex]. Alternatively, if $\tau(\theta)$ is an invertible function, we can simply appropriately transform the Bayesian credibility region for θ to arrive at the set:

$$C_\tau = \{\tau : \tau = \tau(\theta) \text{ for some } \theta \in C\} = \tau(C),$$

that is, C_τ is the set of τ values which correspond to some θ value in the set C [NOTE: the notion here is precisely the same as was discussed at the end of Section 3.2.2 regarding the parameterisation invariance of confidence intervals based on the asymptotic chi-squared pivot].

3.2.4. Bootstrap Intervals: As with parametric point estimation, the interval estimation methods discussed so far are heavily dependent on the chosen probability model. In this section, we briefly introduce and examine some methods of confidence interval construction which do not rely on the explicit specification of a probability model. There are, of course, many such procedures, but we focus here on methods based on the idea of the bootstrap, introduced in Section 2.6.3.

Recall that the bootstrap was originally introduced in Section 2.6.3 as a method of assessing the bias and variability of an estimator, $\hat{\theta}$. In particular, the estimated standard deviation of any estimator $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ was derived by constructing some large number, B, of re-samples (with replacement) from the observed values of the sample, $X_1, \ldots, X_n$. The estimator was then applied to each of these re-samples, $\{X^\star_{1,1}, \ldots, X^\star_{n,1}\}, \ldots, \{X^\star_{1,b}, \ldots, X^\star_{n,b}\}, \ldots, \{X^\star_{1,B}, \ldots, X^\star_{n,B}\}$, to construct the bootstrap replicates:

$$\hat{\theta}^\star_b = \hat{\theta}(X^\star_{1,1}, \ldots, X^\star_{n,1}), \quad b = 1, \ldots, B.$$

The estimated standard deviation of the estimator $\hat{\theta}$ was then calculated as the sample standard deviation of the $\hat{\theta}^\star_b$'s:

$$\hat{\sigma}_B(\hat{\theta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^\star_b - \frac{1}{B}\sum_{c=1}^{B}\hat{\theta}^\star_c\right)}.$$

Now, we have seen from our investigation of asymptotic confidence intervals that the *MLE* is approximately normally distributed. It can further be said that most standard estimators are asymptotically normal and asymptotically unbiased. As such, the only requirement for constructing an approximate confidence interval is an estimate of the standard deviation of the chosen estimator. In the case of asymptotic likelihood-based intervals for the *MLE* we used the standard deviation calculated from the expected Fisher information, and this calculation was dependent (quite heavily) on the choice of parametric probability model. If we accept the asymptotic normality of our estimator $\hat{\theta}$, but are not willing to specify a probability model, then the bootstrap can provide us with an estimate of the standard deviation of $\hat{\theta}$ for use in constructing the approximate confidence interval

$$\left\{\hat{\theta} - \Phi^{-1}(1-\beta)\hat{\sigma}_B(\hat{\theta}), \hat{\theta} + \Phi^{-1}(1-\alpha+\beta)\hat{\sigma}_B(\hat{\theta})\right\},$$

where we have the usual choice of the value β (i.e., $\beta = 0$ and $\beta = \alpha$ lead to one-sided intervals, while $\beta = \alpha/2$ leads to an equal-tailed interval). Note that if $\hat{\theta}$ is an average [i.e., has the form $\hat{\theta} = n^{-1}\sum_{i=1}^{n} g(X_i)$ for some function $g(\cdot)$], then we can employ the Central Limit Theorem to arrive at a confidence interval [based on the sample standard deviation of the $g(X_i)$ values] without relying on the specification of a probability model. In this sense, the bootstrap interval provided above is a natural extension to the Central Limit Theorem for estimators which are not in the form of an average.

Just as for the Central Limit Theorem intervals, the accuracy of the normal approximation in bootstrap-based intervals can be poor in small samples. To combat this, we might employ Student's t-distribution quantiles, $t_{(n-1)}(1-\beta)$ and $t_{(n-1)}(1-\alpha+\beta)$ rather than the normal quantiles. Of course, we actually have a better alternative in this case. As we saw in Section 2.6.3, the values $\hat{\theta}^\star_b$ not only provide us with estimates of the bias and standard deviation of our estimator, $\hat{\theta}$, but also of its entire distribution. As such, rather than using either the normal

or Student's t-distribution quantiles, we could use the quantiles of the "bootstrap distribution" determined by a histogram of the $\hat{\theta}_b^\star$ values. Specifically, then, we can calculate the *bootstrap percentile method interval* as $C_B = \left\{\hat{\theta}^\star_{[B(\alpha-\beta)]}, \hat{\theta}^\star_{[B(1-\beta)]}\right\}$, where $\hat{\theta}^\star_{[B\gamma]}$ is the $B\gamma^{\text{th}}$ (rounded to the nearest integer) value in the (ascending) ordered collection of $\hat{\theta}_b^\star$'s. In other words, $\hat{\theta}^\star_{[B\gamma]}$ is the γ-quantile of the observed distribution of the $\hat{\theta}_b^\star$ values. To see that this interval is reasonable, we note that the histogram of the $\hat{\theta}_b^\star$ values can be seen as an approximation (which would be exact if B were infinite) to the conditional distribution of $\hat{\theta}^\star$ given the observed data $X_1, \ldots, X_n$ or, more compactly, given their empirical distribution $\hat{F}$. In other words, the quantiles of the bootstrap histogram are approximations to the quantities $Pr(\hat{\theta}^\star \leq c|\hat{F})$, where $\hat{\theta}^\star$ is the random variable representing the generation of a bootstrap replicate based on a re-sample from the data $X_1, \ldots, X_n$ (or equivalently from its empirical *CDF* $\hat{F}$). In this notation, then, the bootstrap percentile method interval is constructed as $(\hat{\theta}^\star_{B,l}, \hat{\theta}^\star_{B,u})$ where these endpoints satisfy:

$$Pr_{\hat{F}}(\hat{\theta}^\star \leq \hat{\theta}^\star_{B,l}) = \alpha - \beta; \qquad Pr_{\hat{F}}(\hat{\theta}^\star \leq \hat{\theta}^\star_{B,u}) = 1 - \beta.$$

In other words, $\hat{\theta}^\star_{B,l}$ and $\hat{\theta}^\star_{B,u}$ are chosen to be the appropriate quantiles of the conditional distribution of $\hat{\theta}^\star$ under $\hat{F}$, and these values are then approximated from the bootstrap histogram; $\hat{\theta}^\star_{B,l} \approx \hat{\theta}^\star_{[B(\alpha-\beta)]}$ and $\hat{\theta}^\star_{B,u} \approx \hat{\theta}^\star_{[B(1-\beta)]}$. In this way, we see that the interval C_B essentially follows the standard bootstrap paradigm by constructing an interval for θ by finding a corresponding interval for $\hat{\theta}$. Specifically, we see that a direct confidence interval for θ would consist of endpoints $\hat{\theta}_l$ and $\hat{\theta}_u$ satisfying:

$$Pr_F(\theta \leq \hat{\theta}_l) = \alpha - \beta; \qquad Pr_F(\theta \leq \hat{\theta}_u) = 1 - \beta.$$

The bootstrap paradigm replaces population quantities by sample quantities (e.g., F by $\hat{F}$ and θ by $\hat{\theta}$) and sample quantities by re-sample quantities (e.g., $\hat{\theta}_l$ by $\hat{\theta}^\star_{B,l}$ and $\hat{\theta}_u$ by $\hat{\theta}^\star_{B,u}$), and doing so yields the bootstrap percentile interval here, with the slight modification that we actually replace θ by $\hat{\theta}^\star$.

In fact, this "minor modification" in the application of the bootstrap paradigm leads us to consider various other bootstrap intervals in which the paradigm is followed more closely. In particular, instead of bootstrapping the quantity $\hat{\theta}$ itself (i.e., using the percentiles of the simulated distribution of $\hat{\theta}^\star$ under $\hat{F}$), we might instead choose to bootstrap some other quantity $Q(F, \hat{F})$ and use its simulated quantiles to construct an interval. The simplest example of such an approach is to consider a quantity which we believe is (approximately) pivotal; for example, $Q = Q(F, \hat{F}) = \frac{\theta(\hat{F}) - \theta(F)}{\hat{\sigma}(F)}$, where $\hat{\sigma}(F)$ is some reasonable estimator of $\sigma(F) = \sqrt{Var_F\{\theta(\hat{F})\}}$. If the quantity Q were exactly pivotal, we could construct a confidence interval for θ by first defining $q_{l,\beta}$ and $q_{u,\beta}$ to be the $\alpha-\beta$ and $1-\beta$ quantiles, respectively, of the distribution of Q [so that $Pr\{Q \leq q_{l,\beta}\} = \alpha - \beta$ and $Pr\{Q \leq q_{u,\beta}\} = 1 - \beta$ for any distribution] and then noting:

$$1 - \alpha = Pr_F\{q_{l,\beta} \leq Q(F, \hat{F}) \leq q_{u,\beta}\} = Pr_{\hat{F}}\{q_{l,\beta} \leq Q(\hat{F}, \hat{F}^\star) \leq q_{u,\beta}\}.$$

Using this result, we can see that $q_{l,\beta}$ and $q_{u,\beta}$ can be directly estimated by generating $\hat{Q}_b = Q(\hat{F}, \hat{F}_b^\star)$ values (by resampling from $\hat{F}$ to produce $\hat{F}_b$'s) and then approximating $q_{l,\beta}$ by $\hat{q}_{l,\beta} = \hat{Q}_{[B(\alpha-\beta)]}$ and $q_{u,\beta}$ by $\hat{q}_{u,\beta} = \hat{Q}_{[B(1-\beta)]}$. In the case of $Q = \frac{\theta(\hat{F}) - \theta(F)}{\hat{\sigma}(F)}$, we see that this leads to a confidence interval as follows:

1. Using B resamples, calculate $\hat{Q}_b = \frac{\hat{\theta}_b^\star - \hat{\theta}}{\hat{\sigma}(\hat{F})}$ and then define the estimates $\hat{q}_{l,\beta} = \hat{Q}_{[B(\alpha-\beta)]}$ and $\hat{q}_{u,\beta} = \hat{Q}_{[B(\alpha-\beta)]}$.

2. Construct the confidence interval using a "pivoting" argument:

$$\begin{aligned} 1-\alpha &= Pr_F\{q_{l,\beta} \leq Q(F,\hat{F}) \leq q_{u,\beta}\} \\ &\approx Pr_F\{\hat{q}_{l,\beta} \leq \frac{\theta(\hat{F})-\theta(F)}{\hat{\sigma}(F)} \leq \hat{q}_{u,\beta}\} \\ &= Pr_F\{\hat{\theta}-\hat{q}_{u,\beta}\hat{\sigma}(F) \leq \theta \leq \hat{\theta}-\hat{q}_{l,\beta}\hat{\sigma}(F)\}, \end{aligned}$$

   so that the confidence interval takes the form $\{\hat{\theta}-\hat{q}_{u,\beta}\hat{\sigma}(F), \hat{\theta}-\hat{q}_{l,\beta}\hat{\sigma}(F)\}$. In other words, the interval is precisely of the standard form, but instead of normal or t-distribution quantiles, we use quantiles based on the bootstrap distribution of Q.

Of course, this interval relies on the fact that the quantity Q is (at least approximately) pivotal, and thus the best intervals of this kind will be those based on Q's which are as close to pivotal as possible. However, we can now see that, even if Q is not pivotal, the idea behind this approach follows the bootstrap paradigm nicely; that is, we want two points, q_l and q_u, such that $Pr_F\{q_l \leq Q(F,\hat{F}) \leq q_u\} = 1-\alpha$ and we approximate these two points by $\hat{q}_l$ and $\hat{q}_u$ which satisfy $Pr_{\hat{F}}\{\hat{q}_l \leq Q(\hat{F},\hat{F}^\star) \leq \hat{q}_u\} = 1-\alpha$, which is just our original equation with all population quantities [i.e., F, θ, q_l and q_u] replaced by sample quantities [i.e., $\hat{F}$, $\hat{\theta}$, $\hat{q}_l$ and $\hat{q}_u$] and all sample quantities [i.e., $\hat{F}$] replaced by resample quantities [i.e., $\hat{F}^\star$]. Once we have these two points, $\hat{q}_l$ and $\hat{q}_u$, we substitute back into original equation to get:

$$1-\alpha \approx Pr_F\{\hat{q}_l \leq Q(F,\hat{F}) \leq \hat{q}_u\}$$

and then "pivot" based on the specific form of the quantity Q to arrive at a confidence interval for θ. Before moving to an example, we note that the interval described above based on the quantity $Q = \frac{\hat{\theta}-\theta}{\hat{\sigma}(F)}$ uses the upper end of the simulated distribution to construct the lower endpoint of the confidence interval and vice versa. This is in direct contrast to the percentile interval described originally, and mimics the usual pivotal constructions we saw in the original section on classical confidence interval construction. A natural question arises as to which approach is more sensible. Unfortunately, there is no easy answer to this question as both approaches have their benefits and drawbacks. We shall briefly discuss these issues further in the following sections; however, an example can serve to illustrate some of the key points.

Example 3.4: Recall Example 2.13, which considered a sample of LSAT scores and GPAs for the entering classes at 15 U.S. Law Schools. There we calculated the correlation coefficient between the measurements as $\hat{\rho} = 0.7764$. Further, based on $B = 10000$ re-samples, we approximated the variance of this estimator as $\widehat{Var}_B(\hat{\rho}) = 0.0180$. As such, we could construct an approximate (equal-tailed) 95% confidence interval for the true correlation coefficient, ρ, as:

$$0.7764 \pm 1.96\sqrt{0.0180} = (0.5134,\ 1.039).$$

Note that this interval extends beyond the allowable range for a correlation coefficient. Intervals for which this can happen are not *range-respecting*. We will briefly mention this property in Section 3.3. However, here we point out that the bootstrap percentile method is range-respecting. Using the histogram shown in Example 2.13, we can construct an approximate (equal-tailed) 95% bootstrap percentile interval for ρ as:

$$(\hat{\rho}^\star_{[250]}, \hat{\rho}^\star_{[9750]}) = (0.4640,\ 0.9612),$$

where $\hat{\rho}^\star_{[b]}$ represents the b^{th} order statistic of the re-sampled $\hat{\rho}^\star$'s (i.e., the b^{th} smallest value). Note the difference between this interval and the previous interval. In particular, we see that

this interval remains within the allowable range for correlation coefficients. Moreover, we see that this interval is not symmetric around the point estimate $\hat{\rho} = 0.7764$, which is clear from the skewness of the bootstrap histogram.

Alternatively, suppose we use the method described above based on a quantity $Q = \frac{\hat{\rho}-\rho}{\hat{\sigma}(\hat{\rho})}$, and we set $\hat{\sigma}^2(\hat{\rho}) = \widehat{Var}_J(\hat{\rho})$, the Jackknife estimate of variance of $\hat{\rho}$. In this case, the $\hat{Q}_b$ values are calculated by resampling from the original data, and then calculating $\hat{\rho}^\star$ and $\widehat{Var}_J(\hat{\rho}^\star)$ using the resampled observations. For example, in Example 2.13, we saw a particular resample determined by:

$$\begin{gathered}
X_1^\star = X_7 = (555, 3.00),\ X_2^\star = X_{15} = (594, 2.96),\ X_3^\star = X_{14} = (572, 2.88), \\
X_4^\star = X_3 = (558, 2.81),\ X_5^\star = X_7 = (555, 3.00),\ X_6^\star = X_{14} = (572, 2.88), \\
X_7^\star = X_7 = (555, 3.00),\ X_8^\star = X_7 = (555, 3.00),\ X_9^\star = X_{12} = (575, 2.74), \\
X_{10}^\star = X_3 = (558, 2.81),\ X_{11}^\star = X_6 = (580, 3.07),\ X_{12}^\star = X_6 = (580, 3.07), \\
X_{13}^\star = X_1 = (576, 3.39),\ X_{14}^\star = X_{10} = (605, 3.13),\ X_{15}^\star = X_{12} = (575, 2.74)
\end{gathered}$$

led to $\hat{\rho}^\star = 0.2585$, and this resample now also yields $\hat{\sigma}^2(\hat{\rho}^\star) = \widehat{Var}_J(\hat{\rho}^\star) = 0.0313$ by simply applying the Jackknife variance procedure to the collection $\{X_1^\star, \ldots, X_{15}^\star\}$. Implementing this approach using $B = 10000$ resamples yielded $\hat{Q}_{[250]} = -1.4412$ and $\hat{Q}_{[9750]} = 8.4305$. Lastly, recall that the Jackknife estimate of variance for $\hat{\rho}$ calculated in Example 2.13 was seen to be $\widehat{Var}_J(\hat{\rho}) = 0.0203$. Therefore, the bootstrap confidence interval based on this approach (which, by the way, is widely referred to as the *bootstrap-t* confidence interval, for obvious reasons) is:

$$(0.7764 - 8.4305\sqrt{0.0203}, 0.7764 + 1.4412\sqrt{0.0203}) = (-0.4248, 0.9817)$$

Note the rather dramatic difference between this interval and the percentile method interval. One possible explanation is to recall that we have already noted that the Jackknife estimate of variance is notoriously poor in many circumstances. Perhaps using the bootstrap variance estimate would be better, of course this starts to become very computationally expensive, requiring B^2 resamples in general. [Of course, there is no requirement that the internal B used for the variance calculation be the same as the external B used for the quantile estimation; indeed, if we use $B = 1000$ resamples to estimate the bootstrap variance within each of our original $B = 10000$ resamples, the resulting confidence interval is (-0.1162, 0.9695), though this interval took nearly two days of computing time to construct!]. Finally, we note that the actual interval produced by the bootstrap methods will vary slightly with each implementation, since they will be based on a different random collection of B re-samples. As long as B is reasonably large, however, this variation will be minimal.

In closing, we note that in Example 3.4, even if we were willing to assume a bivariate normal probability model for our data, finding pivots or the exact distribution of the sample correlation coefficient in order to construct exact confidence intervals is extremely difficult (though the distribution of the sample correlation coefficient does exist in closed form and can be found in various advanced texts). Alternatively, we might try an asymptotic likelihood-based approach, but this requires us to work with the general bivariate normal distribution which has a parameter vector of dimension 5 (the two means, two variances and the correlation). In other words, we would have to find the appropriate variance estimate by inverting the expected Fisher information matrix, which is a rather involved and tedious exercise in algebra and calculus. As such, we note that the bootstrap confidence intervals are not only more *robust* (i.e., they do not require assumptions about

the form of the distribution of the observations) but are more easily computed (in the sense that no complex analytical calculations must be performed, though there will certainly be substantial computer-based calculations required). In this sense, the bootstrap intervals are superior to the previous methods we have discussed. Of course, if we truly believe in some parametric assumptions regarding the distribution of our data, then the parametric methods will be "optimal" in some sense (e.g., the intervals will be shorter and more accurate).

3.3. Properties of Confidence Intervals

In this section, we briefly define and discuss several of the most common properties of confidence intervals and regions. We note that many of these have been noted throughout the previous sections in conjunction with the actual construction procedures themselves, and we will here only give general definitions and cursory discussions regarding how each of these procedures can be expected to behave. In particular, we note that these properties are commonly used to compare various construction procedures to hopefully find a "best" construction method. As you would expect, however, it is rarely the case that we can actual find an interval estimation procedure which is *uniformly* best. Indeed, even if we can find a procedure which is uniformly best with respect to one of the properties, it will almost never be the case that it is the best procedure with respect to some other property.

3.3.1. Length and Symmetry: As we have discussed throughout, nearly all of our interval construction procedures have a built in parameter β, which determines the degree of asymmetry in the "tail" behaviour of the intervals. Specifically, our choice of β determines the chance that the interval will "miss on the high side"; that is, the chance that the lower endpoint of the interval will be too large, so that the true value of the parameter of interest is outside and below the calculated interval.

It is quite common practice to choose equal-tailed intervals by choosing $\beta = \alpha/2$, and we have seen that such a practice will lead to the shortest interval (at least among all choices of β; though, it may be possible, of course, to choose a completely new construction method which produces shorter interval estimates) in situations where the underlying probability model is "symmetric" in some sense. The idea behind equal-tailed intervals is essentially the belief that errors in one direction are just as bad as errors of an equal size in the opposite direction. Quite often, this will be the case. Of course, if it is not, then equal-tailed procedures are not as compelling. Moreover, if we think of the length of an interval as a measure of how accurate (and therefore how ultimately useful) it is, then choosing β to minimise the length of the constructed interval is a reasonable approach. As noted, this is accomplished directly by choosing equal-tailed intervals in symmetric situations. Unfortunately, outside this situation, finding the appropriate β to minimise the length of the constructed interval can be a rather complicated (and even intractable) calculus problem.

In addition, we can consider choosing between different construction procedures based on considerations of length. For instance, the likelihood-based intervals constructed from the chi-squared pivot [often termed the *likelihood ratio statistic (LRS)*] will generally be shorter than the likelihood-based intervals derived using the normal approximation to the distribution of the *MLE*. Similarly, we noted that Bayesian *HPD*s are the shortest among all Bayesian credibility intervals. Finally, we note that the length of intervals constructed from inverting the distribution of a statistic T will depend on the variance of the statistic used. Therefore, we might prefer to construct intervals from UMVU estimators if possible.

3.3.2. Coverage Accuracy: For intervals which satisfy Definition 3.1 exactly (e.g., those based on pivots or on the exact distribution of some estimator), the probability that the constructed

interval will correctly contain the true parameter value is precisely equal to the *nominal confidence coefficient* $1-\alpha$. However, for the approximate, asymptotic intervals (e.g., bootstrap intervals or intervals based on the asymptotic distribution of the *MLE*), the question of the degree to which their *true* probability of correctly containing the proper parameter value (also termed the *coverage probability*) is close to the *nominal* probability, $1-\alpha$, arises. The degree to which the actual coverage probability of an interval construction method matches the nominal value, $1-\alpha$, is termed the *coverage accuracy* of the interval procedure.

It can be shown (though the mathematics required is rather advanced) that all the approximate intervals discussed here have a coverage probability which differs from the nominal value by an amount which is inversely proportional to the sample size, n. In particular, if C_n is an asymptotic interval based on a sample of size n, then for the approximate, asymptotic intervals described in the preceding sections (i.e., likelihood-based intervals and bootstrap intervals),

$$\lim_{n\to\infty} n[Pr_\theta\{\tau(\theta) \in C_n\} - (1-\alpha)] = c,$$

for some constant c (which will generally depend on the parameter value θ, or more generally on the true distribution F in non-parametric settings, as well as on the confidence level α). In other words, $Pr_\theta\{\tau(\theta) \in C_n\} \approx (1-\alpha) + \frac{c}{n}$. As such, one way to choose between competing asymptotic methods is to choose the one with the best coverage accuracy (i.e., the smallest value of c). Unfortunately, since the constant c depends on unknown parameters of the underlying distribution, such comparisons can often be quite difficult to perform in practice. Moreover, since the constant depends on the parameter of interest, there may be no uniformly best method for all values of θ (in just the same way that there is usually no uniformly best point estimator with respect to *MSE*).

Finally, we note that it is often possible to improve the coverage accuracy of confidence intervals based on the *LRS* by employing a technique known as *Bartlett correction*. If we define $A(\theta) = E_\theta[2\{l(\hat{\theta}) - l(\theta)\}]$ and let c_α be the $(1-\alpha)$-quantile of the chi-squared distribution with one degree of freedom, then the interval (for a scalar parameter θ):

$$C_{BC} = \{\theta \in \Theta : 2\{A(\hat{\theta})\}^{-1}\{l(\hat{\theta}) - l(\theta)\} \leq c_\alpha\}$$

will generally have a coverage error of $\frac{c}{n^2}$ for some constant c. In other words, $Pr_\theta\{\tau(\theta) \in C_{BC}\} \approx (1-\alpha) + \frac{c}{n}$ for all $\theta \in \Theta$. This result can also be translated to the case that $\tau(\theta)$ is the (scalar) parameter of interest using the induced likelihood, $M(\tau)$; see Theorem 2.1 of Section 2.1.2. Unfortunately, it is usually rather difficult to calculate $A(\theta)$ explicitly. However, this result does indicate that confidence regions based on the *LRS* have some claim to being the best construction method, and indeed, this claim can be substantiated to some degree with more advanced theory (provided, of course, that we are willing to accept the validity of the chosen parametric probability model on which the likelihood ratio statistic is based).

3.3.3. Parameterisation Equivariance and Range Respecting Intervals: We close this section of the notes with some very brief comments regarding two very desirable properties for confidence intervals to possess.

First, as discussed earlier, we would like our interval construction procedures to transform appropriately if we change our focus from $\tau = \tau(\theta)$ to $\gamma = \gamma(\tau) = \gamma\{\tau(\theta)\}$ for a reparameterisation $\gamma(\cdot)$ [recall that $\gamma(\cdot)$ is reparameterisation if it is an invertible transformation; that is, it is a strictly monotone function]. In particular, if $C_\tau = (l, u)$ is a confidence interval for $\tau = \tau(\theta)$ based on a particular construction method, then we would like the interval for $\gamma = \gamma\{\tau(\theta)\}$ based on this

same construction method to satisfy $C_\gamma = \gamma(C_\tau) = \{\gamma(l), \gamma(u)\}$ [provided that $\gamma(\cdot)$ is an increasing function; otherwise, if it is decreasing we need $C_\gamma = \{\gamma(u), \gamma(l)\}$.] Intervals which have this property are called *parameterisation equivariant*. We note that the bootstrap percentile intervals, the intervals based on the asymptotic chi-squared approximation of the distribution of the likelihood ratio statistic as well as the Bayesian intervals are all parameterisation equivariant. However, intervals based on asymptotic normal approximations generally are not parameterisation equivariant. In particular, the likelihood-based intervals derived from the asymptotic normal approximation to the distribution of the *MLE* or bootstrap intervals based on the bootstrap estimate of variance, $\hat{\sigma}^2_B$, are not parameterisation equivariant.

Finally, we would like our intervals to contain only values in the allowable parameter space for the parameter of interest; that is, if C is a confidence interval for $\tau = \tau(\theta) \in \mathcal{T}$, we would like $C \subseteq \mathcal{T}$. Intervals which satisfy this requirement are referred to as *range-respecting*. We note that Bayesian intervals and the bootstrap percentile interval are range-respecting, as are the intervals based on the asymptotic chi-squared approximation to the distribution of the likelihood ratio statistic. However, as with parameterisation equivariance, intervals based on normal approximations are generally not range respecting nor, in general, are the bootstrap intervals based on simulating the distribution of the (approximate) pivotal quantity $Q(F, \hat{F})$, unless we modify their construction to allow for a potential truncation. For example, if $\mathcal{T} = [a, b]$ and we construct a confidence interval C which is not contained within the range $[a, b]$, then we can modify our interval as $C' = C \cap [a, b]$ to ensure that we have an interval which is range respecting. Of course, if such truncations are required to make an interval construction procedure range respecting, we should consider whether an alternative procedure which is range respecting without the need for such truncations is possible.