Chapter 9

Confidence Intervals

9.1 Introduction

Definition 9.1. Let the data $Y_1, ..., Y_n$ have pdf or pmf $f(\boldsymbol{y}|\theta)$ with parameter space Θ and support $\mathcal{Y}$. Let $L_n(\boldsymbol{Y})$ and $U_n(\boldsymbol{Y})$ be statistics such that $L_n(\boldsymbol{y}) \le U_n(\boldsymbol{y})$, $\forall \boldsymbol{y} \in \mathcal{Y}$. Then $(L_n(\boldsymbol{y}), U_n(\boldsymbol{y}))$ is a 100 $(1-\alpha)$ % **confidence interval** (CI) for θ if

$$P_\theta(L_n(\boldsymbol{Y}) < \theta < U_n(\boldsymbol{Y})) = 1 - \alpha$$

for all $\theta \in \Theta$. The interval $(L_n(\boldsymbol{y}), U_n(\boldsymbol{y}))$ is a large sample 100 $(1-\alpha)$ % CI for θ if

$$P_\theta(L_n(\boldsymbol{Y}) < \theta < U_n(\boldsymbol{Y})) \to 1 - \alpha$$

for all $\theta \in \Theta$ as $n \to \infty$.

Definition 9.2. Let the data $Y_1, ..., Y_n$ have pdf or pmf $f(\boldsymbol{y}|\boldsymbol{\theta})$ with parameter space Θ and support $\mathcal{Y}$. The random variable $R(\boldsymbol{Y}|\boldsymbol{\theta})$ is a **pivot** or pivotal quantity if the distribution of $R(\boldsymbol{Y}|\boldsymbol{\theta})$ is independent $\boldsymbol{\theta}$. The quantity $R(\boldsymbol{Y}, \boldsymbol{\theta})$ is an **asymptotic pivot** or asymptotic pivotal quantity if the limiting distribution of $R(\boldsymbol{Y}, \boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$.

The first CI in Definition 9.1 is sometimes called an exact CI. The words "exact" and "large sample" are often omitted. In the following definition, the scaled asymptotic length is closely related to asymptotic relative efficiency of an estimator and high power of a test of hypotheses.

Definition 9.3. Let (L_n, U_n) be a 100 $(1-\alpha)$% CI or large sample CI for θ. If

$$n^{\delta}(U_n - L_n) \xrightarrow{P} A_{\alpha},$$

then A_{α} is the *scaled asymptotic length* of the CI. Typically $\delta = 0.5$ but superefficient CIs have $\delta = 1$. For fixed δ and fixed coverage $1-\alpha$, a CI with smaller A_{α} is "better" than a CI with larger A_{α}. If $A_{1,\alpha}$ and $A_{2,\alpha}$ are for two competing CIs with the same δ, then $(A_{2,\alpha}/A_{1,\alpha})^{1/\delta}$ is a measure of "asymptotic relative efficiency."

Definition 9.4. Suppose a nominal $100(1-\alpha)$% CI for θ has actual coverage $1-\delta$, so that $P_{\theta}(L_n(\boldsymbol{Y}) < \theta < U_n(\boldsymbol{Y})) = 1-\delta$ for all $\theta \in \Theta$. If $1-\delta > 1-\alpha$, then the CI is *conservative*. If $1-\delta < 1-\alpha$, then the CI is *liberal*. Conservative CIs are generally considered better than liberal CIs. Suppose a nominal $100(1-\alpha)$% large sample CI for θ has actual coverage $1-\delta_n$where $\delta_n \to \delta$ as $n \to \infty$ for all $\theta \in \Theta$. If $1-\delta > 1-\alpha$, then the CI is *asymptotically conservative*. If $1-\delta < 1-\alpha$, then the CI is *asymptotically liberal*. It is possible that $\delta \equiv \delta(\theta)$ depends on θ, and that the CI is (asymptotically) conservative or liberal for different values of θ, in that the (asymptotic) coverage is higher or lower than the nominal coverage, depending on θ.

Example 9.1. a) Let $Y_1, ..., Y_n$ be iid $N(\mu, \sigma^2)$ where $\sigma^2 > 0$. Then

$$R(\boldsymbol{Y}|\mu, \sigma^2) = \frac{\overline{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

is a pivot or pivotal quantity.

To use this pivot to find a CI for μ, let $t_{p,\alpha}$ be the α percentile of the t_p distribution. Hence $P(T \le t_{p,\alpha}) = \alpha$ if $T \sim t_p$. Using $t_{p,\alpha} = -t_{p,1-\alpha}$ for $0 < \alpha < 0.5$, note that

$$1 - \alpha = P(-t_{n-1,1-\alpha/2} \le \frac{\overline{Y} - \mu}{S/\sqrt{n}} \le t_{n-1,1-\alpha/2}) =$$

$$P(-t_{n-1,1-\alpha/2} \ \ S/\sqrt{n} \le \overline{Y} - \mu \le t_{n-1,1-\alpha/2} \ \ S/\sqrt{n}) =$$

$$P(-\overline{Y} - t_{n-1,1-\alpha/2} \ \ S/\sqrt{n} \le -\mu \le -\overline{Y} + t_{n-1,1-\alpha/2} \ \ S/\sqrt{n}) =$$

$$P(\overline{Y} - t_{n-1,1-\alpha/2} \ \ S/\sqrt{n} \le \mu \le \overline{Y} + t_{n-1,1-\alpha/2} \ \ S/\sqrt{n}).$$

Thus

$$\overline{Y} \pm t_{n-1,1-\alpha/2} \ \ S/\sqrt{n}$$

is a $100(1-\alpha)\%$ CI for μ.

b) If $Y_1, ..., Y_n$ are iid with $E(Y) = \mu$ and VAR$(Y) = \sigma^2 > 0$, then, by the CLT and Slutsky's Theorem,

$$R(\boldsymbol{Y}|\mu, \sigma^2) = \frac{\overline{Y} - \mu}{S/\sqrt{n}} = \frac{\sigma}{S} \ \ \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)$$

is an asymptotic pivot or asymptotic pivotal quantity.

To use this asymptotic pivot to find a large sample CI for μ, let z_α be the α percentile of the $N(0,1)$ distribution. Hence $P(Z \le z_\alpha) = \alpha$ if $Z \sim N(0,1)$. Using $z_\alpha = -z_{1-\alpha}$ for $0 < \alpha < 0.5$, note that for large n,

$$1 - \alpha \approx P(-z_{1-\alpha/2} \le \frac{\overline{Y} - \mu}{S/\sqrt{n}} \le z_{1-\alpha/2}) =$$

$$P(-z_{1-\alpha/2} \ \ S/\sqrt{n} \le \overline{Y} - \mu \le z_{1-\alpha/2} \ \ S/\sqrt{n}) =$$
$$P(-\overline{Y} - z_{1-\alpha/2} \ \ S/\sqrt{n} \le -\mu \le -\overline{Y} + z_{1-\alpha/2} \ \ S/\sqrt{n}) =$$
$$P(\overline{Y} - z_{1-\alpha/2} \ \ S/\sqrt{n} \le \mu \le \overline{Y} + z_{1-\alpha/2} \ \ S/\sqrt{n}).$$

Thus

$$\overline{Y} \pm z_{1-\alpha/2} \ \ S/\sqrt{n}$$

is a large sample $100(1-\alpha)\%$ CI for μ.

Since $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$ but $t_{n-1,1-\alpha/2} \to z_{1-\alpha/2}$ as $n \to \infty$,

$$\overline{Y} \pm t_{n-1,1-\alpha/2} \ \ S/\sqrt{n}$$

is also a large sample $100(1-\alpha)\%$ CI for μ. This t interval is the same as that in a), and is the most widely used confidence interval in statistics. Replacing $z_{1-\alpha/2}$ by $t_{n-1,1-\alpha/2}$ makes the CI longer and hence less likely to be liberal.

Large sample theory can be used to find a CI from the asymptotic pivot. Suppose that $\boldsymbol{Y} = (Y_1, ..., Y_n)$ and that $W_n \equiv W_n(\boldsymbol{Y})$ is an estimator of some parameter μ_W such that

$$\sqrt{n}(W_n - \mu_W) \xrightarrow{D} N(0, \sigma_W^2)$$

where σ_W^2/n is the asymptotic variance of the estimator W_n. The above notation means that if n is large, then for probability calculations

$$W_n - \mu_W \approx N(0, \sigma_W^2/n).$$

Suppose that S_W^2 is a consistent estimator of σ_W^2 so that the (asymptotic) *standard error* of W_n is $\text{SE}(W_n) = S_W/\sqrt{n}$. As in Example 9.1, let $P(Z \le z_\alpha) = \alpha$ if $Z \sim N(0,1)$. Then for large n

$$1 - \alpha \approx P(-z_{1-\alpha/2} \le \frac{W_n - \mu_W}{SE(W_n)} \le z_{1-\alpha/2}),$$

and an approximate or large sample $100(1-\alpha)\%$ CI for μ_W is given by

$$(W_n - z_{1-\alpha/2}SE(W_n), W_n + z_{1-\alpha/2}SE(W_n)). \tag{9.1}$$

Since

$$\frac{t_{p,1-\alpha/2}}{z_{1-\alpha/2}} \to 1$$

if $p \equiv p_n \to \infty$ as $n \to \infty$, another large sample $100(1-\alpha)\%$ CI for μ_W is

$$(W_n - t_{p,1-\alpha/2}SE(W_n), W_n + t_{p,1-\alpha/2}SE(W_n)). \tag{9.2}$$

The CI (9.2) often performs better than the CI (9.1) in small samples. The quantity $t_{p,1-\alpha/2}/z_{1-\alpha/2}$ can be regarded as a small sample correction factor. The CI (9.2) is longer than the CI (9.1). Hence the CI (9.2) more *conservative* than the CI (9.1).

Suppose that there are two independent samples $Y_1, ..., Y_n$ and $X_1, ..., X_m$ and that

$$\begin{pmatrix} \sqrt{n}(W_n(\boldsymbol{Y}) - \mu_W(Y)) \\ \sqrt{m}(W_m(\boldsymbol{X}) - \mu_W(X)) \end{pmatrix} \xrightarrow{D} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_W^2(Y) & 0 \\ 0 & \sigma_W^2(X) \end{pmatrix} \right).$$

Then

$$\begin{pmatrix} (W_n(\boldsymbol{Y}) - \mu_W(Y)) \\ (W_m(\boldsymbol{X}) - \mu_W(X)) \end{pmatrix} \approx N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_W^2(Y)/n & 0 \\ 0 & \sigma_W^2(X)/m \end{pmatrix} \right),$$

and

$$W_n(\boldsymbol{Y}) - W_m(\boldsymbol{X}) - (\mu_W(Y) - \mu_W(X)) \approx N(0, \frac{\sigma_W^2(Y)}{n} + \frac{\sigma_W^2(X)}{m}).$$

Hence $SE(W_n(\boldsymbol{Y}) - W_m(\boldsymbol{X})) =$

$$\sqrt{\frac{S^2_W(\boldsymbol{Y})}{n} + \frac{S^2_W(\boldsymbol{X})}{m}} = \sqrt{[SE(W_n(\boldsymbol{Y}))]^2 + [SE(W_m(\boldsymbol{X}))]^2},$$

and the large sample $100(1-\alpha)\%$ CI for $\mu_W(Y) - \mu_W(X)$ is given by

$$(W_n(\boldsymbol{Y}) - W_m(\boldsymbol{X})) \pm z_{1-\alpha/2} SE(W_n(\boldsymbol{Y}) - W_m(\boldsymbol{X})). \quad (9.3)$$

If p_n is the degrees of freedom used for a single sample procedure when the sample size is n, let $p = \min(p_n, p_m)$. Then another large sample $100(1-\alpha)\%$ CI for $\mu_W(Y) - \mu_W(X)$ is given by

$$(W_n(\boldsymbol{Y}) - W_m(\boldsymbol{X})) \pm t_{p,1-\alpha/2} SE(W_n(\boldsymbol{Y}) - W_m(\boldsymbol{X})). \quad (9.4)$$

These CIs are known as *Welch intervals.* See Welch (1937) and Yuen (1974).

Example 9.2. Consider the single sample procedures where $W_n = \overline{Y}_n$. Then $\mu_W = E(Y)$, $\sigma^2_W = \text{VAR}(Y)$, $S_W = S_n$, and $p = n-1$. Let t_p denote a random variable with a t distribution with p degrees of freedom and let the α percentile $t_{p,\alpha}$ satisfy $P(t_p \le t_{p,\alpha}) = \alpha$. Then the classical *t-interval* for $\mu \equiv E(Y)$ is

$$\overline{Y}_n \pm t_{n-1,1-\alpha/2} \frac{S_n}{\sqrt{n}}$$

and the *t-test statistic* for $Ho: \mu = \mu_o$ is

$$t_o = \frac{\overline{Y} - \mu_o}{S_n/\sqrt{n}}.$$

The right tailed p-value is given by $P(t_{n-1} > t_o)$.

Now suppose that there are two samples where $W_n(\boldsymbol{Y}) = \overline{Y}_n$ and $W_m(\boldsymbol{X}) = \overline{X}_m$. Then $\mu_W(Y) = E(Y) \equiv \mu_Y$, $\mu_W(X) = E(X) \equiv \mu_X$, $\sigma^2_W(Y) = \text{VAR}(Y) \equiv \sigma^2_Y$, $\sigma^2_W(X) = \text{VAR}(X) \equiv \sigma^2_X$, and $p_n = n-1$. Let $p = \min(n-1, m-1)$. Since

$$SE(W_n(\boldsymbol{Y}) - W_m(\boldsymbol{X})) = \sqrt{\frac{S^2_n(\boldsymbol{Y})}{n} + \frac{S^2_m(\boldsymbol{X})}{m}},$$

the *two sample t-interval* for $\mu_Y - \mu_X$

$$(\overline{Y}_n - \overline{X}_m) \pm t_{p,1-\alpha/2} \sqrt{\frac{S^2_n(\boldsymbol{Y})}{n} + \frac{S^2_m(\boldsymbol{X})}{m}}$$

and *two sample t-test statistic*

$$t_o = \frac{\overline{Y}_n - \overline{X}_m}{\sqrt{\frac{S_n^2(\boldsymbol{Y})}{n} + \frac{S_m^2(\boldsymbol{X})}{m}}}.$$

The right tailed p-value is given by $P(t_p > t_o)$. For sample means, values of the degrees of freedom that are more accurate than $p = \min(n-1, m-1)$ can be computed. See Moore (2007, p. 474).

The remainder of this section follows Olive (2008b, Section 2.4) closely. Let $\lfloor x \rfloor$ denote the "greatest integer function" (so $\lfloor 7.7 \rfloor = 7$). Let $\lceil x \rceil$ denote the smallest integer greater than or equal to x (so $\lceil 7.7 \rceil = 8$).

Example 9.3: inference with the sample median. Let $U_n = n - L_n$ where $L_n = \lfloor n/2 \rfloor - \lceil \sqrt{n/4}\ \rceil$ and use

$$SE(\text{MED}(n)) = 0.5(Y_{(U_n)} - Y_{(L_n+1)}).$$

Let $p = U_n - L_n - 1$. Then a large sample $100(1-\alpha)\%$ confidence interval for the population median MED(Y) is

$$\text{MED}(n) \pm t_{p,1-\alpha/2} SE(\text{MED}(n)). \tag{9.5}$$

Example 9.4: inference with the trimmed mean. The symmetrically trimmed mean or the δ *trimmed mean*

$$T_n = T_n(L_n, U_n) = \frac{1}{U_n - L_n} \sum_{i=L_n+1}^{U_n} Y_{(i)} \tag{9.6}$$

where $L_n = \lfloor n\delta \rfloor$ and $U_n = n - L_n$. If $\delta = 0.25$, say, then the δ trimmed mean is called the 25% trimmed mean.

The trimmed mean is estimating a truncated mean μ_T. Assume that Y has a probability density function $f_Y(y)$ that is continuous and positive on its support. Let y_δ be the number satisfying $P(Y \le y_\delta) = \delta$. Then

$$\mu_T = \frac{1}{1-2\delta} \int_{y_\delta}^{y_{1-\delta}} y f_Y(y) dy. \tag{9.7}$$

Notice that the 25% trimmed mean is estimating

$$\mu_T = \int_{y_{0.25}}^{y_{0.75}} 2y f_Y(y) dy.$$

To perform inference, find $d_1, ..., d_n$ where

$$d_i = \begin{cases} Y_{(L_n+1)}, & i \leq L_n \\ Y_{(i)}, & L_n + 1 \leq i \leq U_n \\ Y_{(U_n)}, & i \geq U_n + 1. \end{cases}$$

Then the Winsorized variance is the sample variance $S_n^2(d_1, ..., d_n)$ of $d_1, ..., d_n$, and the scaled Winsorized variance

$$V_{SW}(L_n, U_n) = \frac{S_n^2(d_1, ..., d_n)}{([U_n - L_n]/n)^2}. \tag{9.8}$$

The standard error of T_n is $SE(T_n) = \sqrt{V_{SW}(L_n, U_n)/n}$.

A large sample 100 $(1 - \alpha)\%$ confidence interval (CI) for μ_T is

$$T_n \pm t_{p,1-\frac{\alpha}{2}} SE(T_n) \tag{9.9}$$

where $P(t_p \leq t_{p,1-\frac{\alpha}{2}}) = 1 - \alpha/2$ if t_p is from a t distribution with $p = U_n - L_n - 1$ degrees of freedom. This interval is the classical t–interval when $\delta = 0$, but $\delta = 0.25$ gives a robust CI.

Example 9.5. Suppose the data below is from a symmetric distribution with mean μ. Find a 95% CI for μ.

6, 9, 9, 7, 8, 9, 9, 7

Solution. When computing small examples by hand, the steps are to sort the data from smallest to largest value, find n, L_n, U_n, $Y_{(L_n)}$, $Y_{(U_n)}$, p, MED(n) and SE(MED(n)). After finding $t_{p,1-\alpha/2}$, plug the relevant quantities into the formula for the CI. The sorted data are 6, 7, 7, 8, 9, 9, 9, 9. Thus MED(n) = $(8 + 9)/2 = 8.5$. Since $n = 8$, $L_n = \lfloor 4 \rfloor - \lceil \sqrt{2} \rceil = 4 - \lceil 1.414 \rceil = 4 - 2 = 2$ and $U_n = n - L_n = 8 - 2 = 6$. Hence $SE(\text{MED}(n)) = 0.5(Y_{(6)} - Y_{(3)}) = 0.5 * (9 - 7) = 1$. The degrees of freedom $p = U_n - L_n - 1 = 6 - 2 - 1 = 3$. The cutoff $t_{3,0.975} = 3.182$. Thus the 95% CI for MED(Y) is

$$\text{MED}(n) \pm t_{3,0.975} SE(\text{MED}(n))$$

$= 8.5 \pm 3.182(1) = (5.318, 11.682)$. The classical t–interval uses $\overline{Y} = (6+7+7+8+9+9+9+9)/8$ and $S_n^2 = (1/7)[(\sum_{i=1}^n Y_i^2) - 8(8^2)] = (1/7)[(522 - 8(64)] = 10/7 \approx 1.4286$, and $t_{7,0.975} \approx 2.365$. Hence the 95% CI for μ is $8 \pm 2.365(\sqrt{1.4286/8}) = (7.001, 8.999)$. Notice that the t-cutoff = 2.365 for the classical interval is less than the t-cutoff = 3.182 for the median interval and that $SE(\overline{Y}) < SE(\text{MED}(n))$.

Example 9.6. In the last example, what happens if the 6 becomes 66 and a 9 becomes 99?

Solution. Then the ordered data are 7, 7, 8, 9, 9, 9, 66, 99. Hence MED(n) = 9. Since L_n and U_n only depend on the sample size, they take the same values as in the previous example and $SE(\text{MED}(n)) = 0.5(Y_{(6)} - Y_{(3)}) = 0.5 * (9 - 8) = 0.5$. Hence the 95% CI for MED(Y) is MED(n) $\pm$ $t_{3,0.975}SE(\text{MED}(n)) = 9 \pm 3.182(0.5) = (7.409, 10.591)$. Notice that with discrete data, it is possible to drive $SE(\text{MED}(n))$ to 0 with a few outliers if n is small. The classical confidence interval $\overline{Y} \pm t_{7,0.975}S/\sqrt{n}$ blows up and is equal to $(-2.955, 56.455)$.

Example 9.7. The Buxton (1920) data contains 87 heights of men, but five of the men were recorded to be about 0.75 inches tall! The mean height is $\overline{Y} = 1598.862$ and the classical 95% CI is (1514.206, 1683.518). MED(n) = 1693.0 and the resistant 95% CI based on the median is (1678.517, 1707.483). The 25% trimmed mean $T_n = 1689.689$ with 95% CI (1672.096, 1707.282).

The heights for the five men were recorded under their head lengths, so the outliers can be corrected. Then $\overline{Y} = 1692.356$ and the classical 95% CI is (1678.595, 1706.118). Now MED(n) = 1694.0 and the 95% CI based on the median is (1678.403, 1709.597). The 25% trimmed mean $T_n = 1693.200$ with 95% CI (1676.259, 1710.141). Notice that when the outliers are corrected, the three intervals are very similar although the classical interval length is slightly shorter. Also notice that the outliers roughly shifted the median confidence interval by about 1 mm while the outliers greatly increased the length of the classical t–interval.

9.2 Some Examples

Example 9.8. Suppose that $Y_1, ..., Y_n$ are iid from a one parameter exponential family with parameter τ. Assume that $T_n = \sum_{i=1}^n t(Y_i)$ is a complete sufficient statistic. Then from Theorems 3.6 and 3.7, often $T_n \sim G(na, 2b\ \tau)$ where a and b are known positive constants. Then

$$\hat{\tau} = \frac{T_n}{2nab}$$

is the UMVUE and often the MLE of τ. Since $T_n/(b\ \tau) \sim G(na, 2)$, a

$100(1-\alpha)\%$ confidence interval for τ is

$$\left(\frac{T_n/b}{G(na,2,1-\alpha/2)}, \frac{T_n/b}{G(na,2,\alpha/2)}\right) \approx \left(\frac{T_n/b}{\chi^2_d(1-\alpha/2)}, \frac{T_n/b}{\chi^2_d(\alpha/2)}\right) \tag{9.10}$$

where $d = \lfloor 2na \rfloor$, $\lfloor x \rfloor$ is the greatest integer function (e.g. $\lfloor 7.7 \rfloor = \lfloor 7 \rfloor = 7$), $P[G \leq G(\nu,\lambda,\alpha)] = \alpha$ if $G \sim G(\nu,\lambda)$, and $P[X \leq \chi^2_d(\alpha)] = \alpha$ if X has a chi-square χ^2_d distribution with d degrees of freedom.

This confidence interval can be inverted to perform two tail tests of hypotheses. By Theorem 7.3, the uniformly most powerful (UMP) test of $H_o : \tau \leq \tau_o$ versus $H_A : \tau > \tau_o$ rejects H_o if and only if $T_n > k$ where $P[G > k] = \alpha$ when $G \sim G(na, 2b\ \tau_o)$. Hence

$$k = G(na, 2b\ \tau_o, 1-\alpha). \tag{9.11}$$

A good approximation to this test rejects H_o if and only if

$$T_n > b\ \tau_o \chi^2_d(1-\alpha)$$

where $d = \lfloor 2na \rfloor$.

Example 9.9. If Y is half normal HN(μ,σ) then the pdf of Y is

$$f(y) = \frac{2}{\sqrt{2\pi}\ \sigma} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$$

where $\sigma > 0$ and $y > \mu$ and μ is real. Since

$$f(y) = \frac{2}{\sqrt{2\pi}\ \sigma} I[y > \mu] \exp\left[\left(\frac{-1}{2\sigma^2}\right)(y-\mu)^2\right],$$

Y is a 1P–REF if μ is known.

Since $T_n = \sum (Y_i - \mu)^2 \sim G(n/2, 2\sigma^2)$, in Example 9.8 take $a = 1/2$, $b = 1$, $d = n$ and $\tau = \sigma^2$. Then a $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{T_n}{\chi^2_n(1-\alpha/2)}, \frac{T_n}{\chi^2_n(\alpha/2)}\right). \tag{9.12}$$

The UMP test of $H_o : \sigma^2 \leq \sigma^2_o$ versus $H_A : \sigma^2 > \sigma^2_o$ rejects H_o if and only if

$$T_n/\sigma^2_o > \chi^2_n(1-\alpha).$$

Now consider inference when both μ and σ are unknown. Then the family is no longer an exponential family since the support depends on μ. Let

$$D_n = \sum_{i=1}^n (Y_i - Y_{1:n})^2. \tag{9.13}$$

Pewsey (2002) showed that $(\hat{\mu}, \hat{\sigma}^2) = (Y_{1:n}, \frac{1}{n} D_n)$ is the MLE of (μ, σ^2), and that

$$\frac{Y_{1:n} - \mu}{\sigma \Phi^{-1}(\frac{1}{2} + \frac{1}{2n})} \xrightarrow{D} EXP(1).$$

Since $(\sqrt{\pi/2})/n$ is an approximation to $\Phi^{-1}(\frac{1}{2} + \frac{1}{2n})$ based on a first order Taylor series expansion such that

$$\frac{\Phi^{-1}(\frac{1}{2} + \frac{1}{2n})}{(\sqrt{\pi/2})/n} \to 1,$$

it follows that

$$\frac{n(Y_{1:n} - \mu)}{\sigma \sqrt{\frac{\pi}{2}}} \xrightarrow{D} EXP(1). \tag{9.14}$$

Using this fact, it can be shown that a large sample $100(1-\alpha)\%$ CI for μ is

$$(\hat{\mu} + \hat{\sigma} \log(\alpha)\ \Phi^{-1}(\frac{1}{2} + \frac{1}{2n})\ (1 + 13/n^2),\quad \hat{\mu}) \tag{9.15}$$

where the term $(1+13/n^2)$ is a small sample correction factor. See Abuhassan and Olive (2008).

Note that

$$D_n = \sum_{i=1}^n (Y_i - Y_{1:n})^2 = \sum_{i=1}^n (Y_i - \mu + \mu - Y_{1:n})^2 =$$

$$\sum_{i=1}^n (Y_i - \mu)^2 + n(\mu - Y_{1:n})^2 + 2(\mu - Y_{1:n}) \sum_{i=1}^n (Y_i - \mu).$$

Hence

$$D_n = T_n + \frac{1}{n}[n(Y_{1:n} - \mu)]^2 - 2[n(Y_{1:n} - \mu)] \frac{\sum_{i=1}^n (Y_i - \mu)}{n},$$

or

$$\frac{D_n}{\sigma^2} = \frac{T_n}{\sigma^2} + \frac{1}{n}\frac{1}{\sigma^2}[n(Y_{1:n} - \mu)]^2 - 2[\frac{n(Y_{1:n} - \mu)}{\sigma}]\frac{\sum_{i=1}^n (Y_i - \mu)}{n\sigma}. \tag{9.16}$$

Consider the three terms on the right hand side of (9.16). The middle term converges to 0 in distribution while the third term converges in distribution to a $-2EXP(1)$ or $-\chi^2_2$ distribution since $\sum_{i=1}^n (Y_i - \mu)/(\sigma n)$ is the sample mean of HN(0,1) random variables and $E(X) = \sqrt{2/\pi}$ when $X \sim HN(0,1)$.

Let $T_{n-p} = \sum_{i=1}^{n-p} (Y_i - \mu)^2$. Then

$$D_n = T_{n-p} + \sum_{i=n-p+1}^{n} (Y_i - \mu)^2 - V_n \tag{9.17}$$

where

$$\frac{V_n}{\sigma^2} \xrightarrow{D} \chi^2_2.$$

Hence

$$\frac{D_n}{T_{n-p}} \xrightarrow{D} 1$$

and D_n/σ^2 is asymptotically equivalent to a χ^2_{n-p} random variable where p is an arbitrary nonnegative integer. Pewsey (2002) used $p = 1$.

Thus when both μ and σ^2 are unknown, a large sample $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{D_n}{\chi^2_{n-1}(1-\alpha/2)}, \frac{D_n}{\chi^2_{n-1}(\alpha/2)} \right). \tag{9.18}$$

It can be shown that $\sqrt{n}$ CI length converges to $\sigma^2\sqrt{2}(z_{1-\alpha/2} - z_{\alpha/2})$ for CIs (9.12) and (9.18) while n length CI (9.15) converges to $-\sigma \log(\alpha)\sqrt{\pi/2}$.

When μ and σ^2 are unknown, an approximate α level test of $H_o : \sigma^2 \le \sigma^2_o$ versus $H_A : \sigma^2 > \sigma^2_o$ that rejects H_o if and only if

$$D_n/\sigma^2_o > \chi^2_{n-1}(1-\alpha) \tag{9.19}$$

has nearly as much power as the α level UMP test when μ is known if n is large.

Example 9.10. Following Mann, Schafer, and Singpurwalla (1974, p. 176), let $W_1, ..., W_n$ be iid $EXP(\theta, \lambda)$ random variables. Let

$$W_{1:n} = \min(W_1, ..., W_n).$$

Then the MLE

$$(\hat{\theta}, \hat{\lambda}) = \left(W_{1:n}, \frac{1}{n} \sum_{i=1}^{n} (W_i - W_{1:n}) \right) = (W_{1:n}, \overline{W} - W_{1:n}).$$

Let $D_n = n\hat{\lambda}$. For $n > 1$, a $100(1-\alpha)\%$ confidence interval (CI) for θ is

$$(W_{1:n} - \hat{\lambda}[(\alpha)^{-1/(n-1)} - 1], W_{1:n}) \tag{9.20}$$

while a $100(1-\alpha)\%$ CI for λ is

$$\left(\frac{2D_n}{\chi^2_{2(n-1),1-\alpha/2}}, \frac{2D_n}{\chi^2_{2(n-1),\alpha/2}} \right). \tag{9.21}$$

Let $T_n = \sum_{i=1}^n (W_i - \theta) = n(\overline{W} - \theta)$. If θ is known, then

$$\hat{\lambda}_\theta = \frac{\sum_{i=1}^n (W_i - \theta)}{n} = \overline{W} - \theta$$

is the UMVUE and MLE of λ, and a $100(1-\alpha)\%$ CI for λ is

$$\left(\frac{2T_n}{\chi^2_{2n,1-\alpha/2}}, \frac{2T_n}{\chi^2_{2n,\alpha/2}} \right). \tag{9.22}$$

Using $\chi^2_{n,\alpha}/\sqrt{n} \approx \sqrt{2} z_\alpha + \sqrt{n}$, it can be shown that $\sqrt{n}$ CI length converges to $\lambda(z_{1-\alpha/2} - z_{\alpha/2})$ for CIs (9.21) and (9.22) (in probability). It can be shown that n length CI (9.20) converges to $-\lambda \log(\alpha)$.

When a random variable is a simple transformation of a distribution that has an easily computed CI, the transformed random variable will often have an easily computed CI. Similarly the MLEs of the two distributions are often closely related. See the discussion above Example 5.11. The first 3 of the following 4 examples are from Abuhassan and Olive (2008).

Example 9.11. If Y has a Pareto distribution, $Y \sim \text{PAR}(\sigma, \lambda)$, then $W = \log(Y) \sim EXP(\theta = \log(\sigma), \lambda)$. If $\theta = \log(\sigma)$ so $\sigma = e^\theta$, then a

100 $(1-\alpha)\%$ CI for θ is (9.20). A 100 $(1-\alpha)\%$ CI for σ is obtained by exponentiating the endpoints of (9.20), and a 100 $(1-\alpha)\%$ CI for λ is (9.21). The fact that the Pareto distribution is a log-location-scale family ($W = \log(Y)$ is from a location–scale family) and hence has simple inference does not seem to be well known.

Example 9.12. If Y has a power distribution, $Y \sim POW(\lambda)$, then $W = -\log(Y)$ is $EXP(0, \lambda)$. A 100 $(1-\alpha)\%$ CI for λ is (9.22).

If Y has a two parameter power distribution, $Y \sim power(\tau, \lambda)$, then

$$F(y) = \left(\frac{y}{\tau}\right)^{1/\lambda}$$

for $0 < y \le \tau$. The pdf

$$f(y) = \frac{1}{\tau\lambda}\left(\frac{y}{\tau}\right)^{\frac{1}{\lambda}-1} I(0 < y \le \tau).$$

Then $W = -\log(Y) \sim EXP(-\log(\tau), \lambda)$. Thus (9.21) is an exact $100(1-\alpha)\%$ CI for λ, and (9.20) $= (L_n, U_n)$ is an exact $100(1-\alpha)\%$ CI for $-\log(\tau)$. Hence (e^{L_n}, e^{U_n}) is a $100(1-\alpha)\%$ CI for $1/\tau$, and (e^{-U_n}, e^{-L_n}) is a $100(1-\alpha)\%$ CI for τ.

Example 9.13. If Y has a truncated extreme value distribution, $Y \sim TEV(\lambda)$, then $W = e^Y - 1$ is $EXP(0, \lambda)$. A 100 $(1-\alpha)\%$ CI for λ is (9.22).

Example 9.14. If Y has a lognormal distribution, $Y \sim LN(\mu, \sigma^2)$, then $W_i = \log(Y_i) \sim N(\mu, \sigma^2)$. Thus a $(1-\alpha)100\%$ CI for μ when σ is unknown is

$$(\overline{W}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{S_W}{\sqrt{n}}, \overline{W}_n + t_{n-1,1-\frac{\alpha}{2}}\frac{S_W}{\sqrt{n}})$$

where

$$S_W = \frac{n}{n-1}\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(W_i - \overline{W})^2},$$

and $P(t \le t_{n-1,1-\frac{\alpha}{2}}) = 1 - \alpha/2$ when $t \sim t_{n-1}$.

Example 9.15. Let $X_1, ..., X_n$ be iid Poisson(θ) random variables. The classical large sample 100 $(1-\alpha)\%$ CI for θ is

$$\overline{X} \pm z_{1-\alpha/2}\sqrt{\overline{X}/n}$$

where $P(Z \le z_{1-\alpha/2}) = 1 - \alpha/2$ if $Z \sim N(0,1)$.

Following Byrne and Kabaila (2005), a modified large sample 100 $(1-\alpha)$% CI for θ is (L_n, U_n) where

$$L_n = \frac{1}{n}\left(\sum_{i=1}^n X_i - 0.5 + 0.5 z^2_{1-\alpha/2} - z_{1-\alpha/2}\sqrt{\sum_{i=1}^n X_i - 0.5 + 0.25 z^2_{1-\alpha/2}}\right)$$

and

$$U_n = \frac{1}{n}\left(\sum_{i=1}^n X_i + 0.5 + 0.5 z^2_{1-\alpha/2} + z_{1-\alpha/2}\sqrt{\sum_{i=1}^n X_i + 0.5 + 0.25 z^2_{1-\alpha/2}}\right).$$

Following Grosh (1989, p. 59, 197–200), let $W = \sum_{i=1}^n X_i$ and suppose that $W = w$ is observed. Let $P(T < \chi^2_d(\alpha)) = \alpha$ if $T \sim \chi^2_d$. Then an "exact" 100 $(1-\alpha)$% CI for θ is

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

for $w \ne 0$ and

$$\left(0, \frac{\chi^2_2(1-\alpha)}{2n}\right)$$

for $w = 0$.

The "exact" CI is conservative: the actual coverage $(1-\delta_n) \ge 1 - \alpha =$ the nominal coverage. This interval performs well if θ is very close to 0. See Problem 9.3.

Example 9.16. Let $Y_1, ..., Y_n$ be iid bin$(1, \rho)$. Let $\hat{\rho} = \sum_{i=1}^n Y_i/n =$ number of "successes"$/n$. The classical large sample 100 $(1-\alpha)$% CI for ρ is

$$\hat{\rho} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n}}$$

where $P(Z \le z_{1-\alpha/2}) = 1 - \alpha/2$ if $Z \sim N(0,1)$.

The Agresti Coull CI takes $\tilde{n} = n + z^2_{1-\alpha/2}$ and

$$\tilde{\rho} = \frac{n\hat{\rho} + 0.5 z^2_{1-\alpha/2}}{n + z^2_{1-\alpha/2}}.$$

(The method "adds" $0.5z^2_{1-\alpha/2}$ "0's and $0.5z^2_{1-\alpha/2}$ "1's" to the sample, so the "sample size" increases by $z^2_{1-\alpha/2}$.) Then the large sample 100 $(1-\alpha)$% Agresti Coull CI for ρ is

$$\tilde{\rho} \pm z_{1-\alpha/2}\sqrt{\frac{\tilde{\rho}(1-\tilde{\rho})}{\tilde{n}}}.$$

Now let $Y_1, ..., Y_n$ be independent bin(m_i, ρ) random variables, let $W = \sum_{i=1}^n Y_i \sim \text{bin}(\sum_{\text{i=1}}^{\text{n}} \text{m}_\text{i}, \rho)$ and let $n_w = \sum_{i=1}^n m_i$. Often $m_i \equiv 1$ and then $n_w = n$. Let $P(F_{d_1,d_2} \leq F_{d_1,d_2}(\alpha)) = \alpha$ where F_{d_1,d_2} has an F distribution with d_1 and d_2 degrees of freedom. Assume $W = w$ is observed. Then the Clopper Pearson "exact" 100 $(1-\alpha)$% CI for ρ is

$$\left(0, \frac{1}{1 \ + \ n_w \ F_{2n_w,2}(\alpha)}\right) \quad \text{for} \quad \text{w} = 0,$$

$$\left(\frac{n_w}{n_w \ + \ F_{2,2n_w}(1-\alpha)}, 1\right) \quad \text{for} \quad \text{w} = \text{n}_\text{w},$$

and (ρ_L, ρ_U) for $0 < w < n_w$ with

$$\rho_L = \frac{w}{w + (n_w - w + 1)F_{2(n_w-w+1),2w}(1-\alpha/2)}$$

and

$$\rho_U = \frac{w+1}{w + 1 + (n_w - w)F_{2(n_w-w),2(w+1)}(\alpha/2)}.$$

The "exact" CI is conservative: the actual coverage $(1-\delta_n) \geq 1-\alpha =$ the nominal coverage. This interval performs well if ρ is very close to 0 or 1. The classical interval should only be used if it agrees with the Agresti Coull interval. See Problem 9.2.

Example 9.17. Let $\hat{\rho}$ = number of "successes"$/n$. Consider a taking a simple random sample of size n from a finite population of known size N. Then the classical finite population large sample 100 $(1-\alpha)$% CI for ρ is

$$\hat{\rho} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n-1}\left(\frac{N-n}{N}\right)} = \hat{\rho} \pm z_{1-\alpha/2}SE(\hat{\rho}) \qquad (9.23)$$

where $P(Z \leq z_{1-\alpha/2}) = 1-\alpha/2$ if $Z \sim N(0,1)$.

Following DasGupta (2008, p. 121), suppose the number of successes Y has a hypergeometric $(C, N-C, n)$ where $p = C/N$. If $n/N \approx \lambda \in (0,1)$ where n and N are both large, then

$$\hat{\rho} \approx N\left(\rho, \frac{\rho(1-\rho)(1-\lambda)}{n}\right).$$

Hence CI (9.23) should be good if the above normal approximation is good.

Let $\tilde{n} = n + z^2_{1-\alpha/2}$ and

$$\tilde{\rho} = \frac{n\hat{\rho} + 0.5 z^2_{1-\alpha/2}}{n + z^2_{1-\alpha/2}}.$$

(Heuristically, the method adds $0.5z^2_{1-\alpha/2}$ "0's" and $0.5z^2_{1-\alpha/2}$ "1's" to the sample, so the "sample size" increases by $z^2_{1-\alpha/2}$.) Then a large sample 100 $(1-\alpha)$% Agresti Coull type (ACT) finite population CI for ρ is

$$\tilde{\rho} \pm z_{1-\alpha/2}\sqrt{\frac{\tilde{\rho}(1-\tilde{\rho})}{\tilde{n}}\left(\frac{N-n}{N}\right)} = \tilde{\rho} \pm z_{1-\alpha/2} SE(\tilde{\rho}). \tag{9.24}$$

Notice that a 95% CI uses $z_{1-\alpha/2} = 1.96 \approx 2$.

For data from a finite population, large sample theory gives useful approximations as N and $n \to \infty$ and $n/N \to 0$. Hence theory suggests that the ACT CI should have better coverage than the classical CI if the p is near 0 or 1, if the sample size n is moderate, and if n is small compared to the population size N. The coverage of the classical and ACT CIs should be very similar if n is large enough but small compared to N (which may only be possible if N is enormous). As n increases to N, $\hat{\rho}$ goes to p, SE$(\hat{\rho})$ goes to 0, and the classical CI may perform well. SE$(\tilde{\rho})$ also goes to 0, but $\tilde{\rho}$ is a biased estimator of ρ and the ACT CI will not perform well if n/N is too large.

Want an interval that gives good coverage even if ρ is near 0 or 1 or if n/N is large. A simple method is to combine the two intervals. Let (L_C, U_C) and (L_A, U_A) be the classical and ACT $100(1-\alpha)$% intervals. Let the modified $100(1-\alpha)$% interval be

$$(\max[0, \min(L_C, L_U)], \min(1, \max(U_C, U_A)]). \tag{9.25}$$

The modified interval seems to perform well. See Problem 9.4.

Example 9.18. If $Y_1, ..., Y_n$ are iid Weibull (ϕ, λ), then the MLE $(\hat{\phi}, \hat{\lambda})$ must be found before obtaining CIs. The likelihood

$$L(\phi, \lambda) = \frac{\phi^n}{\lambda^n} \prod_{i=1}^n y_i^{\phi-1} \ \exp\left[\frac{-1}{\lambda} \sum y_i^{\phi}\right],$$

and the log likelihood

$$\log(L(\phi, \lambda)) = n\log(\phi) - n\log(\lambda) + (\phi - 1)\sum_{i=1}^n \log(y_i) - \frac{1}{\lambda}\sum y_i^{\phi}.$$

Hence

$$\frac{\partial}{\partial\lambda}\log(L(\phi, \lambda)) = \frac{-n}{\lambda} + \frac{\sum y_i^{\phi}}{\lambda^2} \stackrel{set}{=} 0,$$

or $\sum y_i^{\phi} = n\lambda$, or

$$\hat{\lambda} = \frac{\sum y_i^{\hat{\phi}}}{n}.$$

Now

$$\frac{\partial}{\partial\phi}\log(L(\phi, \lambda)) = \frac{n}{\phi} + \sum_{i=1}^n \log(y_i) - \frac{1}{\lambda}\sum y_i^{\phi}\log(y_i) \stackrel{set}{=} 0,$$

so

$$n + \phi[\sum_{i=1}^n \log(y_i) - \frac{1}{\lambda}\sum y_i^{\phi}\log(y_i)] = 0,$$

or

$$\hat{\phi} = \frac{n}{\frac{1}{\hat{\lambda}}\sum y_i^{\hat{\phi}}\log(y_i) - \sum_{i=1}^n \log(y_i)}.$$

One way to find the MLE is to use iteration

$$\hat{\lambda}_k = \frac{\sum y_i^{\hat{\phi}_{k-1}}}{n}$$

and

$$\hat{\phi}_k = \frac{n}{\frac{1}{\hat{\lambda}_k}\sum y_i^{\hat{\phi}_{k-1}}\log(y_i) - \sum_{i=1}^n \log(y_i)}.$$

Since $W = \log(Y) \sim SEV(\theta = \log(\lambda^{1/\phi}), \sigma = 1/\phi)$, let

$$\hat{\sigma}_R = MAD(W_1, ..., W_n)/0.767049$$

and

$$\hat{\theta}_R = MED(W_1, ..., W_n) - \log(\log(2))\hat{\sigma}_R.$$

Then $\hat{\phi}_0 = 1/\hat{\sigma}_R$ and $\hat{\lambda}_0 = \exp(\hat{\theta}_R/\hat{\sigma}_R)$. The iteration might be run until both $|\hat{\phi}_k - \hat{\phi}_{k-1}| < 10^{-6}$ and $|\hat{\lambda}_k - \hat{\lambda}_{k-1}| < 10^{-6}$. Then take $(\hat{\phi}, \hat{\lambda}) = (\hat{\phi}_k, \hat{\lambda}_k)$.

By Example 8.13,

$$\sqrt{n}\left(\begin{pmatrix} \hat{\lambda} \\ \hat{\phi} \end{pmatrix} - \begin{pmatrix} \lambda \\ \phi \end{pmatrix} \right) \xrightarrow{D} N_2(\mathbf{0}, \mathbf{\Sigma})$$

where $\mathbf{\Sigma} =$

$$\begin{bmatrix} 1.109\lambda^2(1 + 0.4635\log(\lambda) + 0.5482(\log(\lambda))^2) & 0.257\phi\lambda + 0.608\lambda\phi\log(\lambda) \\ & \\ 0.257\phi\lambda + 0.608\lambda\phi\log(\lambda) & 0.608\phi^2 \end{bmatrix}.$$

Thus $1 - \alpha \approx P(-z_{1-\alpha/2}\sqrt{0.608}\hat{\phi} < \sqrt{n}(\hat{\phi} - \phi) < z_{1-\alpha/2}\sqrt{0.608}\hat{\phi})$ and a large sample $100(1-\alpha)\%$ CI for ϕ is

$$\hat{\phi} \pm z_{1-\alpha/2}\ \hat{\phi}\ \sqrt{0.608/n}. \tag{9.26}$$

Similarly, a large sample $100(1-\alpha)\%$ CI for λ is

$$\hat{\lambda} \pm \frac{z_{1-\alpha/2}}{\sqrt{n}}\ \sqrt{1.109\hat{\lambda}^2[1 + 0.4635\log(\hat{\lambda}) + 0.5824(\log(\hat{\lambda}))^2]}. \tag{9.27}$$

In simulations, for small n the number of iterations for the MLE to converge could be in the thousands, and the coverage of the large sample CIs is poor for $n < 50$. See Problem 9.7.

Iterating the likelihood equations until "convergence" to a point $\hat{\boldsymbol{\theta}}$ is called a fixed point algorithm. Such algorithms may not converge, so check that $\hat{\boldsymbol{\theta}}$ satisfies the likelihood equations. Other methods such as Newton's method may perform better.

Newton's method is used to solve $\boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$ for $\boldsymbol{\theta}$, where the solution is called $\hat{\boldsymbol{\theta}}$, and uses

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - [\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta}_k)}]^{-1}\boldsymbol{g}(\boldsymbol{\theta}_k) \tag{9.28}$$

where

$$\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} = \begin{bmatrix} \frac{\partial}{\partial\theta_1}g_1(\boldsymbol{\theta}) & \dots & \frac{\partial}{\partial\theta_p}g_1(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial\theta_1}g_p(\boldsymbol{\theta}) & \dots & \frac{\partial}{\partial\theta_p}g_p(\boldsymbol{\theta}) \end{bmatrix}.$$

If the MLE is the solution of the likelihood equations, then use $\boldsymbol{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), ..., g_p(\boldsymbol{\theta}))^T$ where

$$g_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_i} \log(L(\boldsymbol{\theta})).$$

Let $\boldsymbol{\theta}_0$ be an initial estimator, such as the method of moments estimator of $\boldsymbol{\theta}$. Let $\boldsymbol{D} = \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})}$. Then

$$D_{ij} = \frac{\partial}{\partial \theta_j} g_i(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(L(\boldsymbol{\theta})) = \sum_{k=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(x_k|\boldsymbol{\theta})),$$

and

$$\frac{1}{n} D_{ij} = \frac{1}{n} \sum_{k=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(X_k|\boldsymbol{\theta})) \stackrel{D}{\to} E\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(X|\boldsymbol{\theta}))\right].$$

Newton's method converges if the initial estimator is sufficiently close to $\boldsymbol{\theta}$, but may diverge otherwise. Hence $\sqrt{n}$ consistent initial estimators are recommended. Newton's method is also popular because if the partial derivative and integration operations can be interchanged, then

$$\frac{1}{n} \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} \stackrel{D}{\to} -\boldsymbol{I}(\boldsymbol{\theta}). \tag{9.29}$$

For example, the regularity conditions hold for a kP-REF by Proposition 8.20. Then a 100 $(1-\alpha)$% large sample CI for θ_i is

$$\hat{\theta}_i \pm z_{1-\alpha/2} \sqrt{-\boldsymbol{D}_{ii}^{-1}} \tag{9.30}$$

where

$$\boldsymbol{D}^{-1} = \left[\boldsymbol{D}_{\boldsymbol{g}(\hat{\boldsymbol{\theta}})}\right]^{-1}.$$

This result follows because

$$\sqrt{-\boldsymbol{D}_{ii}^{-1}} \approx \sqrt{[\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})]_{ii}/n}.$$

Example 9.19. Problem 9.8 simulates CIs for the Rayleigh (μ, σ) distribution of the form (9.30) although no check has been made on whether (9.29) holds for the Rayleigh distribution (which is not a 2P-REF).

$$L(\mu, \sigma) = \left(\prod \frac{y_i - \mu}{\sigma^2}\right) \exp\left[-\frac{1}{2\sigma^2} \sum (y_i - \mu)^2\right].$$

Notice that for fixed σ, $L(Y_{(1)}, \sigma) = 0$. Hence the MLE $\hat{\mu} < Y_{(1)}$. Now the log likelihood

$$\log(L(\mu, \sigma)) = \sum_{i=1}^n \log(y_i - \mu) - 2n\log(\sigma) - \frac{1}{2}\sum \frac{(y_i - \mu)^2}{\sigma^2}.$$

Hence $g_1(\mu, \sigma) =$

$$\frac{\partial}{\partial \mu}\log(L(\mu, \sigma)) = -\sum_{i=1}^n \frac{1}{y_i - \mu} + \frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \mu) \stackrel{set}{=} 0,$$

and $g_2(\mu, \sigma) =$

$$\frac{\partial}{\partial \sigma}\log(L(\mu, \sigma)) = \frac{-2n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (y_i - \mu)^2 \stackrel{set}{=} 0,$$

which has solution

$$\hat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^n (Y_i - \hat{\mu})^2. \tag{9.31}$$

To obtain initial estimators, let $\hat{\sigma}_M = \sqrt{S^2/0.429204}$ and $\hat{\mu}_M = \overline{Y} - 1.253314\hat{\sigma}_M$. These would be the method of moments estimators if S_M^2 was used instead of the sample variance S^2. Then use $\mu_0 = \min(\hat{\mu}_M, 2Y_{(1)} - \hat{\mu}_M)$ and $\sigma_0 = \sqrt{\sum(Y_i - \mu_0)^2/(2n)}$. Now $\boldsymbol{\theta} = (\mu, \sigma)^T$ and

$$\boldsymbol{D} \equiv \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} = \begin{bmatrix} \frac{\partial}{\partial \mu} g_1(\boldsymbol{\theta}) & \frac{\partial}{\partial \sigma} g_1(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \mu} g_2(\boldsymbol{\theta}) & \frac{\partial}{\partial \sigma} g_2(\boldsymbol{\theta}) \end{bmatrix} =$$

$$\begin{bmatrix} -\sum_{i=1}^n \frac{1}{(y_i - \mu)^2} - \frac{n}{\sigma^2} & -\frac{2}{\sigma^3}\sum_{i=1}^n (y_i - \mu) \\ -\frac{2}{\sigma^3}\sum_{i=1}^n (y_i - \mu) & \frac{2n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^n (y_i - \mu)^2 \end{bmatrix}.$$

So

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \begin{bmatrix} -\sum_{i=1}^n \frac{1}{(y_i - \mu_k)^2} - \frac{n}{\sigma_k^2} & -\frac{2}{\sigma_k^3}\sum_{i=1}^n (y_i - \mu_k) \\ -\frac{2}{\sigma_k^3}\sum_{i=1}^n (y_i - \mu_k) & \frac{2n}{\sigma_k^2} - \frac{3}{\sigma_k^4}\sum_{i=1}^n (y_i - \mu_k)^2 \end{bmatrix}^{-1} \boldsymbol{g}(\boldsymbol{\theta}_k)$$

where

$$\boldsymbol{g}(\boldsymbol{\theta}_k) = \begin{pmatrix} -\sum_{i=1}^n \frac{1}{(y_i-\mu_k)} + \frac{1}{\sigma_k^2}\sum_{i=1}^n (y_i - \mu_k) \\ \\ \frac{-2n}{\sigma_k} + \frac{1}{\sigma_k^3}\sum_{i=1}^n (y_i - \mu_k)^2 \end{pmatrix}.$$

This formula could be iterated for 100 steps resulting in $\boldsymbol{\theta}_{101} = (\mu_{101}, \sigma_{101})^T$. Then take $\hat{\mu} = \min(\mu_{101}, 2Y_{(1)} - \mu_{101})$ and

$$\hat{\sigma} = \sqrt{\frac{1}{2n}\sum_{i=1}^n (Y_i - \hat{\mu})^2}.$$

Then $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})^T$ and compute $\boldsymbol{D} \equiv \boldsymbol{D}_{\boldsymbol{g}(\hat{\boldsymbol{\theta}})}$. Then (assuming (9.29) holds) a 100 $(1-\alpha)$% large sample CI for μ is

$$\hat{\mu} \pm z_{1-\alpha/2}\sqrt{-\boldsymbol{D}_{11}^{-1}}$$

and a 100 $(1-\alpha)$% large sample CI for σ is

$$\hat{\sigma} \pm z_{1-\alpha/2}\sqrt{-\boldsymbol{D}_{22}^{-1}}.$$

Example 9.20. Assume that $Y_1, ..., Y_n$ are iid discrete uniform $(1, \eta)$ where η is an integer. For example, each Y_i could be drawn with replacement from a population of η tanks with serial numbers 1, 2, ..., η. The Y_i would be the serial number observed, and the goal would be to estimate the population size η = number of tanks. Then $P(Y_i = i) = 1/\eta$ for $i = 1, ..., \eta$. Then the cdf of Y is

$$F(y) = \sum_{i=1}^{\lfloor y \rfloor} \frac{1}{\eta} = \frac{\lfloor y \rfloor}{\eta}$$

for $1 \leq y \leq \eta$. Here $\lfloor y \rfloor$ is the greatest integer function, eg, $\lfloor 7.7 \rfloor = 7$.

Now let $Z_i = Y_i/\eta$ which has cdf

$$F_Z(t) = P(Z \leq t) = P(Y \leq t\eta) = \frac{\lfloor t\eta \rfloor}{\eta} \approx t$$

for $0 < t < 1$. Let $Z_{(n)} = Y_{(n)}/\eta = \max(Z_1, ..., Z_n)$. Then

$$F_{Z_{(n)}}(t) = P(\frac{Y_{(n)}}{\eta} \leq t) = \left(\frac{\lfloor t\eta \rfloor}{\eta}\right)^n$$

for $1/\eta < t < 1$.

Want c_n so that

$$P(c_n \le \frac{Y_{(n)}}{\eta} \le 1) = 1 - \alpha$$

for $0 < \alpha < 1$. So

$$1 - F_{Z_{(n)}}(c_n) = 1 - \alpha \;\text{ or }\; 1 - \left(\frac{\lfloor c_n \eta \rfloor}{\eta}\right)^n = 1 - \alpha$$

or

$$\frac{\lfloor c_n \eta \rfloor}{\eta} = \alpha^{1/n}.$$

The solution may not exist, but $c_n - 1/\eta \le \alpha^{1/n} \le c_n$. Take $c_n = \alpha^{1/n}$ then

$$[Y_{(n)}, \frac{Y_{(n)}}{\alpha^{1/n}})$$

is a CI for η that has coverage slightly less than $100(1-\alpha)\%$ for small n, but the coverage converges in probability to 1 as $n \to \infty$.

For small n the midpoint of the 95% CI might be a better estimator of η than $Y_{(n)}$. The left endpoint is closed since $Y_{(n)}$ is a consistent estimator of η. If the endpoint was open, coverage would go to 0 instead of 1. It can be shown that n (length CI) converges to $-\eta \log(\alpha)$ in probability. Hence n (length 95% CI) $\approx 3\eta$. Problem 9.9 provides simulations that suggest that the 95% CI coverage and length is close to the asymptotic values for $n \ge 10$.

Example 9.21. Assume that $Y_1, ..., Y_n$ are iid uniform $(0, \theta)$. Let $Z_i = Y_i/\theta \sim U(0,1)$ which has cdf $F_Z(t) = t$ for $0 < t < 1$. Let $Z_{(n)} = Y_{(n)}/\theta = \max(Z_1, ..., Z_n)$. Then

$$F_{Z_{(n)}}(t) = P(\frac{Y_{(n)}}{\theta} \le t) = t^n$$

for $0 < t < 1$.

Want c_n so that

$$P(c_n \le \frac{Y_{(n)}}{\theta} \le 1) = 1 - \alpha$$

for $0 < \alpha < 1$. So

$$1 - F_{Z_{(n)}}(c_n) = 1 - \alpha \;\text{ or }\; 1 - c_n^n = 1 - \alpha$$

or

$$c_n = \alpha^{1/n}.$$

Then

$$\left(Y_{(n)}, \ \frac{Y_{(n)}}{\alpha^{1/n}} \right)$$

is an exact $100(1-\alpha)\%$ CI for θ. It can be shown that n (length CI) converges to $-\theta \log(\alpha)$ in probability.

If $Y_1, ..., Y_n$ are iid $U(\theta_1, \theta_2)$ where θ_1 is known, then $Y_i - \theta_1$ are iid $U(0, \theta_2 - \theta_1)$ and

$$\left(Y_{(n)} - \theta_1, \frac{Y_{(n)} - \theta_1}{\alpha^{1/n}} \right)$$

is a $100(1-\alpha)\%$ CI for $\theta_2 - \theta_1$. Thus if θ_1 is known, then

$$\left(Y_{(n)}, \ \theta_1 (1 - \frac{1}{\alpha^{1/n}}) + \frac{Y_{(n)}}{\alpha^{1/n}} \right)$$

is a $100(1-\alpha)\%$ CI for θ_2.

Example 9.22. Assume $Y_1, ..., Y_n$ are iid with mean μ and variance σ^2. Bickel and Doksum (2007, p. 279) suggest that

$$W_n = n^{-1/2} \left[\frac{(n-1)S^2}{\sigma^2} - n \right]$$

can be used as an asymptotic pivot for σ^2 if $E(Y^4) < \infty$. Notice that $W_n =$

$$n^{-1/2} \left[\frac{\sum (Y_i - \mu)^2}{\sigma^2} \ - \ \frac{n(\overline{Y} - \mu)^2}{\sigma^2} \ - \ n \right] =$$

$$\sqrt{n} \left[\frac{\sum \left(\frac{Y_i - \mu}{\sigma} \right)^2}{n} \ - \ 1 \right] \ - \ \frac{1}{\sqrt{n}} \ n \ \left(\frac{\overline{Y} - \mu}{\sigma} \right)^2 = X_n - Z_n.$$

Since $\sqrt{n} Z_n \stackrel{D}{\rightarrow} \chi^2_1$, the term $Z_n \stackrel{D}{\rightarrow} 0$. Now $X_n = \sqrt{n}(\overline{U} - 1) \stackrel{D}{\rightarrow} N(0, \tau)$ by the CLT since $U_i = [(Y_i - \mu)/\sigma]^2$ has mean $E(U_i) = 1$ and variance

$$V(U_i) = \tau = E(U_i^2) - (E(U_i))^2 = \frac{E[(Y_i - \mu)^4]}{\sigma^4} - 1 = \kappa + 2$$

where κ is the kurtosis of Y_i. Thus $W_n \stackrel{D}{\rightarrow} N(0, \tau)$.

Hence

$$1-\alpha \approx P(-z_{1-\alpha/2} < \frac{W_n}{\sqrt{\tau}} < z_{1-\alpha/2}) = P(-z_{1-\alpha/2}\sqrt{\tau} < W_n < z_{1-\alpha/2}\sqrt{\tau})$$

$$= P(-z_{1-\alpha/2}\sqrt{n\tau} < \frac{(n-1)S^2}{\sigma^2} - n < z_{1-\alpha/2}\sqrt{n\tau})$$

$$= P(n - z_{1-\alpha/2}\sqrt{n\tau} < \frac{(n-1)S^2}{\sigma^2} < n + z_{1-\alpha/2}\sqrt{n\tau}).$$

Hence a large sample $100(1-\alpha)\%$ CI for σ^2 is

$$\left(\frac{(n-1)S^2}{n + z_{1-\alpha/2}\sqrt{n\hat{\tau}}}, \; \frac{(n-1)S^2}{n - z_{1-\alpha/2}\sqrt{n\hat{\tau}}} \right)$$

where

$$\hat{\tau} = \frac{\frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^4}{S^4} - 1.$$

Notice that this CI needs $n > z_{1-\alpha/2}\sqrt{n\hat{\tau}}$ for the right endpoint to be positive. It can be shown that $\sqrt{n}$ (length CI) converges to $2\sigma^2 z_{1-\alpha/2}\sqrt{\tau}$ in probability.

Problem 9.10 uses an asymptotically equivalent $100(1-\alpha)\%$ CI of the form

$$\left(\frac{(n-a)S^2}{n + t_{n-1,1-\alpha/2}\sqrt{n\hat{\tau}}}, \; \frac{(n+b)S^2}{n - t_{n-1,1-\alpha/2}\sqrt{n\hat{\tau}}} \right)$$

where a and b depend on $\hat{\tau}$. The goal was to make a 95% CI with good coverage for a wide variety of distributions (with 4th moments) for $n \geq 100$. The price is that the CI is too long for some of the distributions with small kurtosis. The $N(\mu, \sigma^2)$ distribution has $\tau = 2$, while the EXP(λ) distribution has $\sigma^2 = \lambda^2$ and $\tau = 8$. The quantity τ is small for the uniform distribution but large for the lognormal LN(0,1) distribution.

By the binomial theorem, if $E(Y^4)$ exists and $E(Y) = \mu$ then

$$E(Y-\mu)^4 = \sum_{j=0}^{4} \binom{4}{j} E[Y^j](-\mu)^{4-j} =$$

$$\mu^4 - 4\mu^3 E(Y) + 6\mu^2(V(Y) + [E(Y)]^2) - 4\mu E(Y^3) + E(Y^4).$$

This fact can be useful for computing

$$\tau = \frac{E[(Y_i-\mu)^4]}{\sigma^4} - 1 = \kappa + 2.$$

Example 9.23. Following DasGupta (2008, p. 402-404), consider the pooled t CI for $\mu_1 - \mu_2$. Let $X_1, ..., X_{n_1}$ be iid with mean μ_1 and variance σ_1^2. Let $Y_1, ..., Y_{n_2}$ be iid with mean μ_2 and variance σ_2^2. Assume that the two samples are independent and that $n_i \to \infty$ for $i = 1, 2$ in such a way that $\hat{\rho} = \frac{n_1}{n_1+n_2} \to \rho \in (0, 1)$. Let $\theta = \sigma_2^2/\sigma_1^2$, and let the pooled sample variance

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Then

$$\begin{pmatrix} \sqrt{n_1}(\overline{X} - \mu_1) \\ \sqrt{n_2}(\overline{Y} - \mu_2) \end{pmatrix} \xrightarrow{D} N_2(\mathbf{0}, \mathbf{\Sigma})$$

where $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2)$. Hence

$$\begin{pmatrix} \frac{1}{\sqrt{n_1}} & \frac{-1}{\sqrt{n_2}} \end{pmatrix} \begin{pmatrix} \sqrt{n_1}(\overline{X} - \mu_1) \\ \sqrt{n_2}(\overline{Y} - \mu_2) \end{pmatrix} = \overline{X} - \overline{Y} - (\mu_1 - \mu_2) \xrightarrow{D} N(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}).$$

So

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{D} N(0, 1).$$

Thus

$$\frac{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \; \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \xrightarrow{D} N(0, \tau^2)$$

where

$$\frac{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}{(\frac{1}{n_1} + \frac{1}{n_2})\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1+n_2}} = \frac{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}{\hat{\rho}\sigma_1^2 + (1 - \hat{\rho})\sigma_2^2} \; \frac{1/\sigma_1^2}{1/\sigma_2^2} \; \frac{n_1 n_2}{n_1 + n_2}$$

$$= \frac{\frac{1}{n_1} + \frac{\theta}{n_2}}{\hat{\rho} + (1 - \hat{\rho})\theta} \; \frac{n_1 n_2}{n_1 + n_2} \xrightarrow{D} \frac{1 - \rho + \rho\theta}{\rho + (1 - \rho)\theta} = \tau^2.$$

Now let $\hat{\theta} = S_2^2/S_1^2$ and

$$\hat{\tau}^2 = \frac{1 - \hat{\rho} + \hat{\rho}\ \hat{\theta}}{\hat{\rho} + (1 - \hat{\rho})\ \hat{\theta}}.$$

Notice that $\hat{\tau} = 1$ if $\hat{\rho} = 1/2$, and $\hat{\tau} = 1$ if $\hat{\theta} = 1$.

The usual large sample $(1-\alpha)100\%$ pooled t CI for $(\mu_1 - \mu_2)$ is

$$\overline{X} - \overline{Y} \ \pm \ t_{n_1+n_2-2,1-\alpha/2} \ \ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \tag{9.32}$$

The large sample theory says that this CI is valid if $\tau = 1$, and that

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\hat{\tau} \ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \xrightarrow{D} N(0,1).$$

Hence a large sample $(1-\alpha)100\%$ CI for $(\mu_1 - \mu_2)$ is

$$\overline{X} - \overline{Y} \ \pm \ z_{1-\alpha/2} \ \hat{\tau} \ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Then the large sample $(1-\alpha)100\%$ modified pooled t CI for $(\mu_1 - \mu_2)$ is

$$\overline{X} - \overline{Y} \ \pm \ t_{n_1+n_2-4,1-\alpha/2} \ \hat{\tau} \ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \tag{9.33}$$

The large sample $(1-\alpha)100\%$ Welch CI for $(\mu_1 - \mu_2)$ is

$$\overline{X} - \overline{Y} \ \pm \ t_{d,1-\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \tag{9.34}$$

where $d = max(1, [d_0])$, and

$$d_0 = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{1}{n_1-1}(\frac{S_1^2}{n_1})^2 + \frac{1}{n_2-1}(\frac{S_2^2}{n_2})^2}.$$

Suppose $n_1/(n_1 + n_2) \to \rho$. It can be shown that if the CI length is multiplied by $\sqrt{n_1}$, then the scaled length of the pooled t CI converges in probability to $2z_{1-\alpha/2}\sqrt{\frac{\rho}{1-\rho}\sigma_1^2 + \sigma_2^2}$ while the scaled lengths of the modified pooled t CI and Welch CI both converge in probability to $2z_{1-\alpha/2}\sqrt{\sigma_1^2 + \frac{\rho}{1-\rho}\sigma_2^2}$.

9.3 Bootstrap and Randomization Tests

Randomization tests and bootstrap tests and confidence intervals are resampling algorithms used to provide information about the sampling distribution of a statistic $T_n \equiv T_n(F) \equiv T_n(\boldsymbol{Y}_n)$ where $\boldsymbol{Y}_n = (Y_1, ..., Y_n)$ and the Y_i are iid from a distribution with cdf $F(y) = P(Y \leq y)$. Then T_n has a cdf $H_n(y) = P(T_n \leq y)$. If $F(y)$ is known, then m independent samples $\boldsymbol{Y}^*_{j,n} = (Y^*_{j,1}, ..., Y^*_{j,n})$ of size n could be generated, where the $Y^*_{j,k}$ are iid from a distribution with cdf F and $j = 1, ..., m$. Then the statistic T_n is computed for each sample, resulting in m statistics $T_{1,n}(F), ..., T_{m,n}(F)$ which are iid from a distribution with cdf $H_n(y)$. Equivalent notation $T_{i,n}(F) \equiv T^*_{i,n}(\boldsymbol{Y}^*_{i,n})$ is often used, where $i = 1, ..., m$.

If $W_1, ..., W_m$ are iid from a distribution with cdf F_W, then the empirical cdf F_m corresponding to F_W is given by

$$F_m(y) = \frac{1}{m} \sum_{i=1}^{m} I(W_i \leq y)$$

where the indicator $I(W_i \leq y) = 1$ if $W_i \leq y$ and $I(W_i \leq y) = 0$ if $W_i > y$. Fix m and y. Then $mF_m(y) \sim$ binomial $(m, F_W(y))$. Thus $E[F_m(y)] = F_W(y)$ and $V[F_m(y)] = F_W(y)[1 - F_W(y)]/m$. By the central limit theorem,

$$\sqrt{m}(F_m(y) - F_W(y)) \xrightarrow{D} N(0, F_W(y)[1 - F_W(y)]).$$

Thus $F_m(y) - F_W(y) = O_P(m^{-1/2})$, and F_m is a reasonable estimator of F_W if the number of samples m is large.

Let $W_i = T_{i,n}(F)$. Then $F_m \equiv \tilde{H}_{m,n}$ is an empirical cdf corresponding to H_n. Let $W_i = Y_i$ and $m = n$. Then F_n is the empirical cdf corresponding to F. Let $\boldsymbol{y}_n = (y_1, ..., y_n)^T$ be the observed data. Now F_n is the cdf of the population that consists of $y_1, ..., y_n$ where the probability of selecting y_i is $1/n$. Hence an iid sample of size d from F_n is obtained by drawing a sample of size d with replacement from $y_1, ..., y_n$. If $d = n$, let $\boldsymbol{Y}^*_{j,n} = (Y^*_{j,1}, ..., Y^*_{j,n})$ be an iid sample of size n from the empirical cdf F_n. Hence each $Y^*_{j,k}$ is one of the $y_1, ..., y_n$ where repetition is allowed. Take m independent samples from F_n and compute the statistic T_n for each sample, resulting in m statistics $T_{1,n}(F_n), ..., T_{m,n}(F_n)$ where $T_{i,n}(F_n) \equiv T^*_{i,n}(\boldsymbol{Y}^*_{i,n})$ for $i = 1, ..., m$. This type of sampling can be done even if F is unknown, and if $T_n(F_n) \approx T_n(F)$, then the empirical cdf based on the $T_{i,n}(F_n)$ may be a useful approximation for H_n.

For general resampling algorithms let $T^*_{i,n}(\boldsymbol{Y}^*_{i,n})$ be the statistic based on a randomly chosen sample $\boldsymbol{Y}^*_{i,n}$ used by the resampling algorithm. Let $H_{A,n}$ be the cdf of the $T^*_{i,n}$ based on all J_n possible samples, and let $H_{m,n}$ be the cdf of the $T^*_{i,n}$ based on m randomly chosen samples. Often theoretical results are given for $H_{A,n}$ but are not known for $H_{m,n}$. Let $G_{N,n}$ be a cdf based on a normal approximation for H_n. Central limit type theorems are used and $G_{N,n}$ is often first order accurate: $H_n(y) - G_{N,n}(y) = O_P(n^{-1/2})$. Approximations $G_{E,n}$ based on the Edgeworth expansion (which is not a cdf) and $H_{A,n}$ are sometimes second order accurate: $H_n(y) - H_{A,n}(y) = O_P(n^{-1})$. The following two examples follow DasGupta (2008, pp. 462, 469, 513).

Example 9.24. Let $Y_1, ..., Y_n$ be iid with cdf F. Then the *ordinary bootstrap distribution* of T_n is $H_{A,n}(y) = P_{F_n}(T_n(\boldsymbol{Y}^*_{i,n}) \le y)$ where $\boldsymbol{Y}^*_{i,n} = (Y^*_{i,1}, ..., Y^*_{i,n})$ is an iid sample of size n from the empirical cdf F_n obtained by selecting with replacement from $Y_1, ..., Y_n$. Here $T^*_{i,n}(\boldsymbol{Y}^*_{i,n}) = T_n(\boldsymbol{Y}^*_{i,n})$. Note that there are $J_n = n^n$ ordered samples and $n^n/n!$ unordered samples from F_n. The bootstrap distribution $H_{m,n}$ typically used in practice is based on m samples randomly selected with replacement. Both $H_{A,n}$ and $H_{m,n}$ are estimators of H_n, the cdf of T_n.

For example, suppose the data is 1, 2, 3, 4, 5, 6, 7. Then $n = 7$ and the sample median T_n is 4. Using *R*, we drew $m = 2$ bootstrap samples (samples of size n drawn with replacement from the original data) and computed the sample median $T^*_{1,n} = 3$ and $T^*_{2,n} = 4$.

```
b1 <- sample(1:7,replace=T)
b1
[1] 3 2 3 2 5 2 6
median(b1)
[1] 3
b2 <- sample(1:7,replace=T)
b2
[1] 3 5 3 4 3 5 7
median(b2)
[1] 4
```

Heuristically, suppose $T_n(F_n)$ is an unbiased estimator of θ. Let $T^*_{i,n} = T^*_{i,n}(\boldsymbol{Y}^*_{i,n})$. Then $T^*_{1,n}, ..., T^*_{m,n}$ each gives an unbiased estimator of θ. If m is large, then typical values of $T^*_{i,n}$ should provide information about θ. For example, the middle 95% of the T_i^* should be an approximate 95% *percentile method* CI for θ. Then reject $H_0: \theta = \theta_0$ if θ_0 is not in the CI. This bootstrap inference has two sources of error. First, n needs to be large enough so that $T_n(F_n) \approx T_n(F)$. Second, the $T_{i,n}^*$ are used to form an empirical cdf $H_{m,n}$ corresponding to $H_{A,n}$, so m needs to be large enough so that the empirical cdf $H_{m,n}$ is a good estimator of $H_{A,n}$.

Example 9.25. Let $X_1, ..., X_{k_1}$ be iid with pdf $f(y)$ while $Y_1, ..., Y_{k_2}$ are iid with pdf $f(y - \mu)$. Let $n = k_1 + k_2$ and consider testing $H_0 : \mu = 0$. Let $T_n \equiv T_{k_1,k_2}$ be the two sample t-statistic. Under H_0, the random variables in the combined sample $X_1, ..., X_{k_1}, Y_1, ..., Y_{k_2}$ are iid with pdf $f(y)$. Let $\boldsymbol{Z}_n$ be any permutation of $(X_1, ..., X_{k_1}, Y_1, ..., Y_{k_2})$ and compute $T_n(\boldsymbol{Z}_n)$ for each permutation. Then $H_{A,n}$ is the cdf based on all of the $T_n(\boldsymbol{Z}_n)$. H_0 is rejected if T_n is in the extreme tails of $H_{A,n}$. The number of ordered samples is $J_n = n!$ while the number of unordered samples is $\binom{n}{k_1}$. Such numbers get enormous quickly. Usually m randomly drawn permutations are selected with replacement, resulting in a cdf $H_{m,n}$ used to choose the appropriate cutoffs c_L and c_U.

For randomization tests that used a fixed number $m = B$ of permutations, calculations using binomial approximations suggest that $B = 999$ to 5000 will give a test similar to those based on using all permutations. See Efron and Tibshirani (1993, pp. 208-210). Jöckel (1986) shows, under regularity conditions, that the power of a randomization test is increasing and converges as $m \to \infty$. It is suggested that the tests have good power if $m = 999$, but the pvalue of such a test is bounded below by 0.001 since the pvalue = (1 + the number of the m test statistics at least as extreme as the observed statistic)/$(m+1)$. Buckland (1984) shows that the expected coverage of the nominal $100(1-\alpha)\%$ percentile method confidence interval is approximately correct, but the standard deviation of the coverage is proportional to $1/\sqrt{m}$. Hence the percentile method is a large sample confidence interval, in that the true coverage converges in probability to the nominal coverage, only if $m \to \infty$ as $n \to \infty$. These results are good reasons for using $m = \max(B, [n \log(n)])$ samples.

The key observation for theory is that $H_{m,n}$ is an empirical cdf. To see this claim, recall that $H_{A,n}(y) \equiv H_{A,n}(y|\boldsymbol{Y}_n)$ is a random cdf: it depends on the data $\boldsymbol{Y}_n$. Hence $H_{A,n}(y) \equiv H_{A,n}(y|\boldsymbol{y}_n)$ is the observed cdf based on the observed data. $H_{A,n}(y|\boldsymbol{y}_n)$ can be computed by finding $T_{i,n}^*(\boldsymbol{Y}_{i,n}^*)$ for all J_n possible samples $\boldsymbol{Y}^*_{i,n}$. If m samples are selected with replacement from all possible samples, then the samples are iid and $T^*_{1,n}, ..., T^*_{m,n}$ are iid with cdf $H_{A,n}(y|\boldsymbol{y}_n)$. Hence $F_m \equiv H_{m,n}$ is an empirical cdf corresponding to $F \equiv H_{A,n}(y|\boldsymbol{y}_n)$.

Thus empirical cdf theory can be applied to $H_{m,n}$. Fix n and y. Then $mH_{m,n}(y) \sim$ binomial $(m, H_{A,n}(y|\boldsymbol{y}_n))$. Thus $E[H_{m,n}(y)] = H_{A,n}(y|\boldsymbol{y}_n)$ and $V[H_{m,n}(y)] = H_{A,n}(y|\boldsymbol{y}_n)[1 - H_{A,n}(y|\boldsymbol{y}_n)]/m$. Also

$$\sqrt{m}(H_{m,n}(y) - H_{A,n}(y|\boldsymbol{y}_n)) \xrightarrow{D} N(0, H_{A,n}(y|\boldsymbol{y}_n)[1 - H_{A,n}(y|\boldsymbol{y}_n)]).$$

Thus $H_{m,n}(y) - H_{A,n}(y|\boldsymbol{y}_n) = O_P(m^{-1/2})$. Note that the probabilities and expectations depend on m and on the observed data $\boldsymbol{y}_n$.

This result suggests that if $H_{A,n}$ is a first order accurate estimator of H_n, then $H_{m,n}$ can not be a first order accurate estimator of H_n unless m is proportional to n. If $m = \max(1000, [n \log(n)])$, then $H_{m,n}$ is asymptotically equivalent to $H_{A,n}$ up to terms of order $n^{-1/2}$. If $m = \max(1000, [0.1n^2 \log(n)])$, then $H_{m,n}$ asymptotically equivalent to $H_{A,n}$ up to terms of order n^{-1}.

As an application, Efron and Tibshirani (1993, pp. 187, 275) state that percentile method for bootstrap confidence intervals is first order accurate and that the coefficient of variation of a bootstrap percentile is proportional to $\sqrt{\frac{1}{n} + \frac{1}{m}}$. If $m = 1000$, then the percentile bootstrap is not first order accurate. If $m = \max(1000, [n \log(n)])$, then the percentile bootstrap is first order accurate. Similarly, claims that a bootstrap method is second order accurate are false unless m is proportional to n^2. See a similar result in Robinson (1988).

Practical resampling algorithms often use $m = B = 1000$, 5000 or 10000. The choice of $m = 10000$ works well for small n and for simulation studies since the cutoffs based on $H_{m,n}$ will be close to those based on $H_{A,n}$ with high probability since $V[H_{10000,n}(y)] \leq 1/40000$. For the following theorem, also see Serfling (1981, pp. 59-61).

Theorem 9.1. Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ be iid $k \times 1$ random vectors from a distribution with cdf $F(\boldsymbol{y}) = P(Y_1 \leq y_1, ..., Y_k \leq y_k)$. Let

$$D_n = \sup_{\boldsymbol{y} \in \Re^k} |F_n(\boldsymbol{y}) - F(\boldsymbol{y})|.$$

a) Massart (1990) $k = 1$: $P(D_n > d) \leq 2\exp(-2nd^2)$ if $nd^2 \geq 0.5 \log(2)$.

b) Kiefer (1961) $k \geq 2 : P(D_n > d) \leq C \exp(-(2-\epsilon)nd^2)$ where $\epsilon > 0$ is fixed and the positive constant C depends on ϵ and k but not on F.

To use Theorem 9.1a, fix n (and suppressing the dependence on $\boldsymbol{y}_n$), take $F = H_{A,n}$ computed from the observed data and take $F_m = H_{m,n}$. Then

$$D_m = \sup_{y \in \Re} |H_{m,n}(y) - H_{A,n}(y)|.$$

Recalling that the probability is with respect to the observed data, consider the following choices of m.

i) If $m = 10000$, then $P(D_m > 0.01) \leq 2e^{-2} \approx 0.271$.

ii) If $m = \max(10000, [0.25n \log(n)])$, then for $n > 5000$
$P\left(D_m > \frac{1}{\sqrt{n}}\right) \leq 2\exp(-2[0.25n\log(n)]/n) \approx 2/\sqrt{n}$.

iii) If $m = \max(10000, [0.5n^2 \log(n)])$, then for $n > 70$
$P\left(D_m > \frac{1}{n}\right) \leq 2\exp(-2[0.5n^2\log(n)]/n^2) \approx 2/n$.

Two tail tests with nominal level α and confidence intervals with nominal coverage $1 - \alpha$ tend to use the lower and upper $\alpha/2$ percentiles from $H_{m,n}$. This procedure corresponds to an interval covering $100(1-\alpha)\%$ of the mass. The interval is short if the distribution corresponding to $H_{m,n}$ is approximately symmetric. Skewness or approximate symmetry can be checked by plotting the $T^*_{i,n}$. Shorter intervals can be found if the distribution is skewed by using the shorth(c) estimator where $c = \lceil m(1-\alpha) \rceil$ and $\lceil x \rceil$ is the smallest integer $\geq x$, e.g., $\lceil 7.7 \rceil = 8$. See Grübel (1988). That is, let $T^*_{(1)}, ..., T^*_{(m)}$ be the order statistics of the $T^*_{1,n}, ..., T^*_{m,n}$ computed by the resampling algorithm. Compute $T^*_{(c)} - T^*_{(1)}, T^*_{(c+1)} - T^*_{(2)}, ..., T^*_{(m)} - T^*_{(m-c+1)}$. Let $[T^*_{(s)}, T^*_{(s+c-1)}]$ correspond to the closed interval with the smallest distance. Then reject $H_0 : \theta = \theta_0$ if θ_0 is not in the interval.

Resampling methods can be used in courses on resampling methods, nonparametric statistics, and experimental design. In such courses it can be stated that it is well known that $H_{m,n}$ has good statistical properties (under regularity conditions) if $m \to \infty$ as $n \to \infty$, but algorithms tend to use $m = B$ between 999 and 10000. Such algorithms may perform well in simulations, but lead to tests with pvalue bounded away from 0, confidence intervals with coverage that fails to converge to the nominal coverage, and fail to take advantage of the theory derived for the impractical all subset algorithms. Since $H_{m,n}$ is the empirical cdf corresponding to the all subset algorithm cdf $H_{A,n}$, taking $m = \max(B, [n\log(n)])$ leads to a practical algorithm with good theoretical properties (under regularity conditions) that performs well in simulations.

Although theory for resampling algorithms given in Lehmann (1999, p. 425) and Sen and Singer (1993, p. 365) has the number of samples $m \to \infty$ as the sample size $n \to \infty$, much of the literature suggests using $m = B$ between 999 and 10000. This choice is often justified using simulations and binomial approximations. An exception is Shao (1989) where $n/m \to 0$ as $n \to \infty$. Let $[x]$ be the integer part of x, so $[7.7] = 7$. Then $m = [n^{1.01}]$ may give poor results for $n < 900$. To combine theory with empirical results, we suggest using $m = \max(B, [n\log(n)])$.

Theory for resampling algorithms such as first order accuracy of the bootstrap and the power of randomization tests is usually for the impractical algorithm that uses all J_n samples. Practical algorithms use B randomly drawn samples where B is chosen to give good performance when n is small. We suggest using $m = \max(B, [n\log(n)])$ randomly drawn samples results in a practical algorithm that is asymptotically equivalent to the impractical algorithm up to terms of order $n^{-1/2}$ while also having good small sample performance.

Example 9.26. Suppose F is the cdf of the $N(\mu, \sigma^2)$ distribution and $T_n(F) = \overline{Y}_n \sim N(\mu, \sigma^2/n)$. Suppose m independent samples $(Y^*_{j,1}, ..., Y^*_{j,n}) = \boldsymbol{Y}^*_{j,n}$ of size n are generated, where the $Y^*_{j,k}$ are iid $N(\mu, \sigma^2)$ and $j = 1, ..., m$. Then let the sample mean $T^*_{j,n} = \overline{Y}^*_{j,n} \sim N(\mu, \sigma^2/n)$ for $j = 1, ..., m$.

We want to examine, for a given m and n, how well do the sample quantiles $T^*_{(\lceil m\ \rho \rceil)} = \overline{Y}^*_{(\lceil m\ \rho \rceil),n}$ of the $\overline{Y}^*_{j,n}$ estimate the quantiles $\xi_{\rho,n}$ of the $N(\mu, \sigma^2/n)$ distribution and how well does $(T^*_{(\lceil m\ 0.025 \rceil)}, T^*_{(\lceil m\ 0.975 \rceil)})$ perform as a 95% CI for μ. Here $P(X \le \xi_{\rho,n}) = \rho$ if $X \sim N(\mu, \sigma^2/n)$. Note that $\xi_{\rho,n} = \mu + z_\rho \sigma/\sqrt{n}$ where $P(Z \le z_\rho) = \rho$ if $Z \sim N(0, 1)$.

Fix n and let f_n be the pdf of the $N(\mu, \sigma^2/n)$ distribution. By Theorem 8.27, as $m \to \infty$

$$\sqrt{m}(\overline{Y}^*_{(\lceil m\ \rho \rceil),n} - \xi_{\rho,n}) \xrightarrow{D} N(0, \tau_n^2)$$

where

$$\tau_n^2 \equiv \tau_n^2(\rho) = \frac{\rho(1-\rho)}{[f_n(\xi_\rho)]^2} = \frac{\rho(1-\rho)2\pi\sigma^2}{n\exp(-z_\rho^2)}.$$

Since the quantile $\xi_{\rho,n} = \mu + z_\rho \sigma/\sqrt{n}$, need m fairly large for the estimated quantile to be good. To see this claim, suppose we want m so that

$$P(\xi_{0.975,n} - 0.04\sigma/\sqrt{n} < \overline{Y}^*_{(\lceil m\ 0.975\rceil),n} < \xi_{0.975,n} - 0.04\sigma/\sqrt{n}) > 0.9.$$

(For $N(0,1)$ data, this would be similar to wanting the estimated 0.975 quantile to be between 1.92 and 2.00 with high probability.) Then $0.9 \approx$

$$P(\frac{-0.04\sigma\sqrt{m}}{\tau_n\sqrt{n}} < Z < \frac{0.04\sigma\sqrt{m}}{\tau_n\sqrt{n}}) \approx P(-0.01497\sqrt{m} < Z < 0.01497\sqrt{m}\ )$$

or

$$m \approx \left(\frac{z_{0.05}}{-0.01497}\right)^2 \approx 12076.$$

With $m = 1000$, the above probability is only about 0.36. To have the probability go to one, need $m \to \infty$ as $n \to \infty$.

Note that if $m = B = 1000$, say, then the sample quantile is not a consistent estimator of the population quantile $\xi_{\rho,n}$. Also, $(\overline{Y}^*_{(\lceil m\ \rho\rceil),n} - \xi_{\rho,n}) = O_P(n^{-\delta})$ needs $m \propto n^{2\delta}$ where $\delta = 1/2$ or 1 are the most interesting cases. For good simulation results, typically need m larger than a few hundred, eg $B = 1000$, for small n. Hence $m = \max(B, n\log(n))$ combines theory with good simulation results.

The CI length behaves fairly well for large n. For example, the 95% CI length will be close to $3.92/\sqrt{n}$ since roughly 95% of the $\overline{Y}^*_{j,n}$ are between $\mu - 1.96\sigma/\sqrt{n}$ and $\mu + 1.96\sigma/\sqrt{n}$. The coverage is conservative (higher than 95%) for moderate m. To see this, note that the 95% CI contains μ if $T^*_{(\lceil m\ 0.025\rceil)} < \mu$ and $T^*_{(\lceil m\ 0.975\rceil)} > \mu$. Let $W \sim$ binomial $(m, 0.5)$. Then

$$P(T^*_{(\lceil m\ 0.975\rceil)}) > \mu) \approx P(W > 0.025m) \approx P(Z > \frac{0.025m - 0.5m}{0.5\sqrt{m}}) =$$

$P(Z > -0.95\sqrt{m}) \to 1$ as $m \to \infty$. (Note that if $m = 1000$, then $T^*_{(\lceil m\ 0.975\rceil)} > \mu$ if 225 or more $\overline{Y}^*_{j,n} > \mu$ or if fewer than 975 $\overline{Y}^*_{j,n} < \mu$.)

Since F is not known, we can not sample from $T_n(F)$, but sampling from $T_n(F_n)$ can at least be roughly approximated using computer generated random numbers. The bootstrap replaces m samples from $T_n(F)$ by m samples from $T_n(F_n)$, that is, there is a single sample $Y_1, ..., Y_n$ of data. Take a sample of size n with replacement from $Y_1, ..., Y_n$ and compute the sample mean $\overline{Y}^*_{1,n}$. Repeat to obtain the bootstrap sample $\overline{Y}^*_{1,n}, ..., \overline{Y}^*_{m,n}$. Expect the bootstrap estimator of the quantile to perform less well than that based on samples from $T_n(F)$. So still need m large so that the estimated quantiles are near the population quantiles.

Simulated coverage for the bootstrap percentile 95% CI tends to be near 0.95 for moderate m, and we expect the length of the 95% CI to again be near $3.92/\sqrt{n}$. The bootstrap sample tends to be centered about the observed value of $\overline{Y}$. If there is a "bad sample" so that $\overline{Y}$ is in the left or right tail of the sampling distribution, say $\overline{Y} > \mu + 1.96\sigma/\sqrt{n}$ or $\overline{Y} < \mu - 1.96\sigma/\sqrt{n}$, then the coverage may be much less that 95%. But the probability of a "bad sample" is 0.05 for this example.

9.4 Complements

Guenther (1969) is a useful reference for confidence intervals. Agresti and Coull (1998), Brown, Cai and DasGupta (2001, 2002) and Pires and Amado (2008) discuss CIs for a binomial proportion. Agresti and Caffo (2000) discuss CIs for the difference of two binomial proportions $\rho_1 - \rho_2$ obtained from 2 independent samples. Barker (2002) and Byrne and Kabaila (2005) discuss CIs for Poisson (θ) data. Brown, Cai and DasGupta (2003) discuss CIs for several discrete exponential families. Abuhassan and Olive (2008) consider CIs for some transformed random variables. Also see Brownstein and Pensky (2008).

A comparison of CIs with other intervals (such as prediction intervals) is given in Vardeman (1992). Also see Hahn and Meeker (1991).

Newton's method is described, for example, in Peressini, Sullivan and Uhl (1988, p. 85).

9.5 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

Refer to Chapter 10 for the pdf or pmf of the distributions in the problems below.

9.1. (Aug. 2003 Qual): Suppose that $X_1, ..., X_n$ are iid with the Weibull distribution, that is the common pdf is

$$f(x) = \begin{cases} \frac{b}{a} x^{b-1} e^{-\frac{x^b}{a}} & 0 < x \\ 0 & \text{elsewhere} \end{cases}$$

where a is the unknown parameter, but $b(> 0)$ is assumed known.

a) Find a minimal sufficient statistic for a

b) Assume $n = 10$. Use the Chi-Square Table and the minimal sufficient statistic to find a 95% two sided confidence interval for a.

R/Splus Problems

Use the command *source("A:/sipack.txt")* **to download the functions**. See Section 11.1. Typing the name of the `sipack` function, eg *accisimf*, will display the code for the function. Use the `args` command, eg *args(accisimf)*, to display the needed arguments for the function.

9.2. Let $Y_1, ..., Y_n$ be iid binomial$(1, \rho)$ random variables.

From the website (www.math.siu.edu/olive/sipack.txt), enter the *R/Splus* function `bcisim` into *R/Splus*. This function simulates the 3 CIs (classical, Agresti Coull and exact) from Example 9.16, but changes the CI (L,U) to (max(0,L),min(1,U)) to get shorter lengths.

To run the function for $n = 10$ and $\rho \equiv p = 0.001$, enter the *R/Splus* command `bcisim(n=10,p=0.001)`. Make a table with header "n p ccov clen accov aclen ecov elen." Fill the table for $n = 10$ and $p = 0.001$, 0.01, 0.5, 0.99, 0.999 and then repeat for $n = 100$. The "cov" is the proportion of 500 runs where the CI contained p and the nominal coverage is 0.95. A coverage between 0.92 and 0.98 gives little evidence that the true coverage differs from the nominal coverage of 0.95. A coverage greater that 0.98 suggests that the CI is conservative while a coverage less than 0.92 suggests that the CI is liberal. Typically want the true coverage $\geq$ to the nominal coverage, so conservative intervals are better than liberal CIs. The "len" is the average scaled length of the CI and for large n should be near $2(1.96)\sqrt{p(1-p)}$.

From your table, is the classical estimator or the Agresti Coull CI better? When is the exact interval good? Explain briefly.

9.3. Let $X_1, ..., X_n$ be iid Poisson(θ) random variables.

From the website (www.math.siu.edu/olive/sipack.txt), enter the *R/Splus* function `poiscisim` into *R/Splus*. This function simulates the 3 CIs (classical, modified and exact) from Example 9.15. To run the function for $n = 100$

and $\theta = 5$, enter the *R/Splus* command `poiscisim(theta=5)`. Make a table with header "theta ccov clen mcov mlen ecov elen." Fill the table for $theta = 0.001$, 0.1, 1.0, and 5.

The "cov" is the proportion of 500 runs where the CI contained θ and the nominal coverage is 0.95. A coverage between 0.92 and 0.98 gives little evidence that the true coverage differs from the nominal coverage of 0.95. A coverage greater that 0.98 suggests that the CI is conservative while a coverage less than 0.92 suggests that the CI is liberal (too short). Typically want the true coverage $\geq$ to the nominal coverage, so conservative intervals are better than liberal CIs. The "len" is the average scaled length of the CI and for large $n\theta$ should be near $2(1.96)\sqrt{\theta}$ for the classical and modified CIs.

From your table, is the classical CI or the modified CI or the exact CI better? Explain briefly. (Warning: in a 1999 version of *R*, there was a bug for the Poisson random number generator for $\theta \geq 10$. The 2007 version of *R* seems to work.)

9.4. This problem simulates the CIs from Example 9.17.

a) Download the function `accisimf` into *R/Splus*.

b) The function will be used to compare the classical, ACT and modified 95% CIs when the population size $N = 500$ and p is close to 0.01. The function generates such a population, then selects 5000 independent simple random samples from the population. The 5000 CIs are made for both types of intervals, and the number of times the true population p is in the ith CI is counted. The simulated coverage is this count divided by 5000 (the number of CIs). The nominal coverage is 0.95. To run the function for $n = 50$ and $p \approx 0.01$, enter the command `accisimf(n=50,p=0.01)`. Make a table with header "n p ccov clen accov aclen mcov mlen." Fill the table for $n = 50$ and then repeat for $n = 100, 150, 200, 250, 300, 350, 400$ and 450. The "len" is $\sqrt{n}$ times the mean length from the 5000 runs. The "cov" is the proportion of 5000 runs where the CI contained p and the nominal coverage is 0.95. For 5000 runs, an observed coverage between 0.94 and 0.96 gives little evidence that the true coverage differs from the nominal coverage of 0.95. A coverage greater that 0.96 suggests that the CI is conservative while a coverage less than 0.94 suggests that the CI is liberal. Typically want the true coverage $\geq$ to the nominal coverage, so conservative intervals are better than liberal CIs. The "ccov" is for the classical CI, "accov" is for the Agresti Coull type (ACT) CI and "mcov" is for the modified interval. Given good coverage > 0.94, want short length.

c) First compare the classical and ACT intervals. From your table, for what values of n is the ACT CI better, for what values of n are the 3 intervals about the same, and for what values of n is the classical CI better?

d) Was the modified CI ever good?

9.5. This problem simulates the CIs from Example 9.10.

a) Download the function `expsim` into *R/Splus*.

The output from this function are the coverages scov, lcov and ccov of the CI for λ, θ and of λ if θ is known. The scaled average lengths of the CIs are also given. The lengths of the CIs for λ are multiplied by $\sqrt{n}$ while the length of the CI for θ is multiplied by n.

b) The 5000 CIs are made for 3 intervals, and the number of times the true population parameter λ or θ is in the ith CI is counted. The simulated coverage is this count divided by 5000 (the number of CIs). The nominal coverage is 0.95. To run the function for $n = 5$, $\theta = 0$ and $\lambda = 1$ enter the command `expsim(n=5)`. Make a table with header
"CI for λ CI for θ CI for λ, θ known."
Then make a second header "n cov slen cov slen cov slen" where "cov slen" is below each of the three CI headers. Fill the table for $n = 5$ and then repeat for $n = 10, 20, 50, 100$ and 1000. The "cov" is the proportion of 5000 runs where the CI contained λ or θ and the nominal coverage is 0.95. For 5000 runs, an observed coverage between 0.94 and 0.96 gives little evidence that the true coverage differs from the nominal coverage of 0.95. A coverage greater that 0.96 suggests that the CI is conservative while a coverage less than 0.94 suggests that the CI is liberal. As n gets large, the values of slen should get closer to 3.92, 2.9957 and 3.92.

9.6. This problem simulates the CIs from Example 9.9.

a) Download the function `hnsim` into *R/Splus*.

The output from this function are the coverages scov, lcov and ccov of the CI for σ^2, μ and of σ^2 if μ is known. The scaled average lengths of the CIs are also given. The lengths of the CIs for σ^2 are multiplied by $\sqrt{n}$ while the length of the CI for μ is multiplied by n.

b) The 5000 CIs are made for 3 intervals, and the number of times the true population parameter $\theta = \mu$ or σ^2 is in the ith CI is counted. The simulated coverage is this count divided by 5000 (the number of CIs). The

nominal coverage is 0.95. To run the function for $n = 5$, $\mu = 0$ and $\sigma^2 = 1$ enter the command `hnsim(n=5)`. Make a table with header
"CI for σ^2 CI for μ CI for σ^2, μ known."
Then make a second header "n cov slen cov slen cov slen" where "cov slen" is below each of the three CI headers. Fill the table for $n = 5$ and then repeat for $n = 10, 20, 50, 100$ and 1000. The "cov" is the proportion of 5000 runs where the CI contained θ and the nominal coverage is 0.95. For 5000 runs, an observed coverage between 0.94 and 0.96 gives little evidence that the true coverage differs from the nominal coverage of 0.95. A coverage greater that 0.96 suggests that the CI is conservative while a coverage less than 0.94 suggests that the CI is liberal. As n gets large, the values of slen should get closer to 5.5437, 3.7546 and 5.5437.

9.7. a) Download the function `wcisim` into *R/Splus*.

The output from this function includes the coverages `pcov` and `lcov` of the CIs for ϕ and λ if the simulated data $Y_1, ..., Y_n$ are iid Weibull (ϕ, λ). The scaled average lengths of the CIs are also given. The values `pconv` and `lconv` should be less than 10^{-5}. If this is not the case, increase `iter`. 100 samples of size $n = 100$ are used to create the 95% large sample CIs for ϕ and λ given in Example 9.18. If the sample size is large, then `sdphihat`, the sample standard deviation of the 100 values of the MLE $\hat{\phi}$, should be close to `phiasd` $= \phi\sqrt{.608}$. Similarly, `sdlamhat` should be close to the asymptotic standard deviation `lamasd` $= \sqrt{1.109\lambda^2(1 + 0.4635\log(\lambda) + 0.5282(\log(\lambda))^2)}$.

b) Type the command

```
wcisim(n = 100, phi = 1, lam = 1, iter = 100)
```

and record the coverages for the CIs for ϕ and λ.

c) Type the command

```
wcisim(n = 100, phi = 20, lam = 20, iter = 100)
```

and record the coverages for the CIs for ϕ and λ.

9.8. a) Download the function `raysim` into *R/Splus* to simulate the CI of Example 9.19.

b) Type the command

```
raysim(n = 100, mu = 20, sigma = 20, iter = 100)
```

and record the coverages for the CIs for μ and σ.

9.9. a) Download the function `ducisim` into *R/Splus* to simulate the CI of Example 9.20.

b) Type the command
`ducisim(n=10,nruns=1000,eta=1000)`.
Repeat for $n = 50, 100, 500$ and make a table with header
"n coverage n 95% CI length."
Fill in the table for $n = 10, 50, 100$ and 500.

c) Are the coverages close to or higher than 0.95 and is the scaled length close to $3\eta = 3000$?

9.10. a) Download the function `varcisim` into *R/Splus* to simulate a modified version of the CI of Example 9.22.

b) Type the command `varcisim(n = 100, nruns = 1000, type = 1)` to simulate the 95% CI for the variance for iid N(0,1) data. Is the coverage *vcov* close to or higher than 0.95? Is the scaled length *vlen* $= \sqrt{n}$ (CI length) $= 2(1.96)\sigma^2\sqrt{\tau} = 5.554\sigma^2$ close to 5.554?

c) Type the command `varcisim(n = 100, nruns = 1000, type = 2)` to simulate the 95% CI for the variance for iid EXP(1) data. Is the coverage *vcov* close to or higher than 0.95? Is the scaled length *vlen* $= \sqrt{n}$ (CI length) $= 2(1.96)\sigma^2\sqrt{\tau} = 2(1.96)\lambda^2\sqrt{8} = 11.087\lambda^2$ close to 11.087?

d) Type the command `varcisim(n = 100, nruns = 1000, type = 3)` to simulate the 95% CI for the variance for iid LN(0,1) data. Is the coverage *vcov* close to or higher than 0.95? Is the scaled length *vlen* long?

9.11. a) Download the function `pcisim` into *R/Splus* to simulate the three CIs of Example 9.23. The modified pooled t CI is almost the same as the Welch CI, but uses degrees of freedom $= n_1 + n_2 - 4$ instead of the more complicated formula for the Welch CI. The pooled t CI should have coverage that is too low if

$$\frac{\rho}{1-\rho}\sigma_1^2 + \sigma_2^2 \ < \ \sigma_1^2 + \frac{\rho}{1-\rho}\sigma_2^2.$$

b) Type the command `pcisim(n1=100,n2=200,var1=10,var2=1)` to simulate the CIs for $N(\mu_i, \sigma_i^2)$ data for $i = 1, 2$. The terms *pcov*, *mpcov* and *wcov* are the simulated coverages for the pooled, modified pooled and Welch 95% CIs. Record these quantities. Are they near 0.95?

9.12. (Aug. 2009 qual): Let $X_1, ..., X_n$ be a random sample from a uniform$(0, \theta)$ distribution. Let $Y = \max(X_1, X_2, ..., X_n)$.

a) Find the pdf of Y/θ.

b) To find a confidence interval for θ, can Y/θ be used as a pivot?

c) Find the shortest $(1 - \alpha)\%$ confidence interval for θ.