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 $L_n(\boldsymbol{y}) \leq U_n(\boldsymbol{y})$, $\forall \boldsymbol{y} \in \mathcal{Y}$. 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** if the limiting distribution of $R(\boldsymbol{Y}, \boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$.

Example 9.1. 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 pivotal quantity. If $Y_1, ..., Y_n$ are iid with $E(Y) = \mu$ and and $\text{VAR}(Y) = \sigma^2 > 0$, 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.

Large sample theory can be used to find 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 $\text{SE}(W_n) = S_W/\sqrt{n}$. Let z_α be the α percentile of the N(0,1) distribution. Hence $P(Z \le z_\alpha) = \alpha$ if $Z \sim N(0, 1)$. Then

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

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 interval (9.2) is longer than the interval (9.1). Hence the interval (9.2) more *conservative* than interval (9.1).

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

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

Then

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

and

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

Hence

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

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

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

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

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

These CI's 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*

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

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 (2004, p. 452).

9.2 Some Examples

Example 9.3. 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 Theorem 3.6, 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

$$\left(\frac{T_n/b}{G(na, 2, 1-\alpha/2)}, \frac{T_n/b}{G(na, 2, \alpha/2)}\right) \approx \left(\frac{T_n/b}{\chi^2_d(1-\alpha/2)}, \frac{T_n/b}{\chi^2_d(\alpha/2)}\right) \quad (9.5)$$

where $\lfloor x \rfloor$ is the greatest integer function (e.g. $\lfloor 7.7 \rfloor = \lfloor 7 \rfloor = 7$), $d = \lfloor 2na \rfloor$, $P[G \leq G(\nu, \beta, \alpha)] = \alpha$ if $G \sim G(\nu, \beta)$, 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 7.1, 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

$$k = G(na, 2b\ \tau_o, 1 - \alpha). \tag{9.6}$$

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 9.4. 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 9.3 take $a = 1/2$, $b = 1$, $d = n$ and $\tau = \sigma^2$. Then a $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{T_n}{\chi^2_n(1-\alpha/2)}, \frac{T_n}{\chi^2_n(\alpha/2)}\right). \tag{9.7}$$

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 $Y_{1:n} = \min(Y_1, ..., Y_n) = Y_{(1)}$ and let

$$W_n = \sum_{i=1}^n (Y_i - Y_{1:n})^2. \tag{9.8}$$

To show that the large sample inference using W_n in place of T_n is valid, some notation from large sample theory is useful. A sequence of random variables Z_n is *bounded in probability*, written $Z_n = O_P(1)$, if for every $\epsilon > 0$ there exist constants N_ϵ and D_ϵ such that $P(|Z_n| \leq D_\epsilon) > 1 - \epsilon$ for all $n > N_\epsilon$. For example, if Z_n converges in distribution to a random variable Z, written $Z_n \stackrel{D}{\rightarrow} Z$, then $Z_n = O_P(1)$. The notation $Z_n = O_P(a_n)$ means that the sequence $Z_n/a_n = O_P(1)$. See Definition 8.10.

From the theory of large deviations, $n(Y_{1:n} - \mu) \stackrel{D}{\rightarrow} Z$ for some random variable Z and thus $Y_{1:n} = O_P(1/n)$. By the central limit theorem, $T_n/n = O_P(1/\sqrt{n})$ and $\sum(Y_i - \mu)/n = O_P(1/\sqrt{n})$. Note that

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

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

Hence

$$W_n - T_n = \frac{1}{n}[n(Y_{1:n} - \mu)]^2 - 2[n(Y_{1:n} - \mu)] \frac{\sum(Y_i - \mu)}{n},$$

or

$$W_n - T_n = O_P(1/n) + O_P(1) O_P(1/\sqrt{n})$$

or

$$W_n - T_n = O_P(1/\sqrt{n}). \tag{9.9}$$

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

$$\left(\frac{W_n}{\chi^2_n(1-\alpha/2)}, \frac{W_n}{\chi^2_n(\alpha/2)} \right) \tag{9.10}$$

can be obtained by replacing T_n by W_n in Equation (9.7).

Similarly when μ and σ^2 are unknown, an approximate α level test of $H_o : \sigma^2 \leq \sigma_o^2$ versus $H_A : \sigma^2 > \sigma_o^2$ that rejects H_o if and only if

$$W_n > \sigma_o^2 \chi^2_n(1-\alpha) \tag{9.11}$$

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

Since $Y_{1:n} > \mu$, notice that $\sum(Y_i - \mu)^2 > \sum(Y_i - Y_{1:n})^2$. Pewsey (2002) gives a result similar to Equation (9.10); however, he concluded that $W_n \approx$

χ^2_{n-1} instead of χ^2_n and thus replaced n by $n-1$. His simulations showed that the confidence interval coverage was good for n as small as 20, suggesting replacing n by $n-1$ in Equations (9.10) and (9.11) is a useful technique when the sample size n is small.

9.3 Complements

Guenther (1969) is a useful reference for confidence intervals. Agresti and Coull (1998) and Brown, Cai and Das Gupta (2001) 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. Byrne and Kabaila (2005) discuss CIs for Poisson (θ) data.

A comparison of CIs with other intervals is given in Vardeman (1992).

9.4 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. Let $\hat{p}$ = number of "successes"$/n$. The classical large sample 100 $(1-\alpha)$% CI for p is $\hat{p} \pm z_{1-\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ 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

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

(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 p is $\tilde{p} \pm z_{1-\alpha/2}\sqrt{\dfrac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$.

From the website (www.math.siu.edu/olive/sipack.txt), enter the *R/Splus* function `accisim` into *R*. To run the function for $n = 10$ and $p = 0$, enter the *R/Splus* command `accisim(n=10,p=0)`. Make a table with header "n p ccov clen accov aclen." Fill the table for $n = 10$ and $p = 0, 0.01, 0.5, 0.99, 1$ 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 0.95 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 CI or the Agresti Coull CI better? Explain briefly.

9.2. 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 $P(Z \leq z_{1-\alpha/2}) = 1 - \alpha/2$ 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

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

and

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

From the website (www.math.siu.edu/olive/sipack.txt), enter the *R/Splus* function `poiscisim` into *R*. To run the function for $n = 100$ and $\theta = 10$, enter the *R/Splus* command `poiscisim(theta=10)`. Make a table with header "theta ccov clen mcov mlen." Fill the table for $theta = 0.001, 0.1, 1.0$, and 10. 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 0.95 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 $2(1.96)\sqrt{\theta}$.

From your table, is the classical CI or the modified CI better? Explain briefly.

9.3. (Aug. 2003 SIU 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.