Lecture 22: Asymptotic confidence sets

Comparison of asymptotic confidence sets

Intuitively, if two asymptotic confidence sets are constructed using two different estimators, $\widehat{\theta}_{1n}$ and $\widehat{\theta}_{2n}$, and if $\widehat{\theta}_{1n}$ is asymptotically more efficient than $\widehat{\theta}_{2n}$ (§4.5.1), then the confidence set based on $\widehat{\theta}_{1n}$ should be better than the one based on $\widehat{\theta}_{2n}$ in some sense.

Proposition 7.4

Let

$$C_j(X) = \{\theta : \|\widehat{V}_{jn}^{-1/2}(\widehat{\theta}_{jn} - \theta)\|^2 \le \chi^2_{k,\alpha}\}, \quad j = 1,2,$$

be the confidence sets based on $\widehat{\theta}_{jn}$ satisfying

$$V_{jn}^{-1/2}(\widehat{\theta}_{jn} - \theta) \to_d N_k(0, I_k),$$

where $\widehat{V}_{jn}$ is consistent for V_{jn}, $j = 1,2$.
If $\mathrm{Det}(V_{1n}) < \mathrm{Det}(V_{2n})$ for sufficiently large n, where $\mathrm{Det}(A)$ is the determinant of A, then

$$P\big(\mathrm{vol}(C_1(X)) < \mathrm{vol}(C_2(X))\big) \to 1.$$

Proof

The result follows from the consistency of $\widehat{V}_{jn}$ and the fact that the volume of the ellipsoid $C_j(X)$ is equal to

$$\mathrm{vol}(C_j(X)) = \frac{\pi^{k/2}(\chi^2_{k,\alpha})^{k/2}[\mathrm{Det}(\widehat{V}_{jn})]^{1/2}}{\Gamma(1+k/2)}.$$

Asymptotic efficiency

If $\widehat{\theta}_{1n}$ is asymptotically more efficient than $\widehat{\theta}_{2n}$ (§4.5.1), then $\mathrm{Det}(V_{1n}) \leq \mathrm{Det}(V_{2n})$.

Hence, Proposition 7.4 indicates that a more efficient estimator of θ results in a better confidence set in terms of volume.

If $\widehat{\theta}_n$ is asymptotically efficient (optimal in the sense of having the smallest asymptotic covariance matrix; see Definition 4.4), then the corresponding confidence set $C(X)$ is asymptotically optimal (in terms of volume) among the confidence sets of the same form as $C(X)$.

Parametric likelihoods

In parametric problems, it is shown in §4.5 that MLE's or RLE's are asymptotically efficient.
Thus, we study more closely the asymptotic confidence sets based on MLE's and RLE's or, more generally, based on likelihoods.
Consider the case where $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ is a parametric family dominated by a σ-finite measure, where $\Theta \subset \mathscr{R}^k$.
Consider $\theta = (\vartheta, \varphi)$ and confidence sets for ϑ with dimension r.
Let $\ell(\theta)$ be the likelihood function based on the observation $X = x$.
The acceptance region of the LR test defined in §6.4.1 with $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$ is

$$A(\vartheta_0) = \{x : \ell(\vartheta_0, \widehat{\varphi}_{\vartheta_0}) \geq e^{-c_\alpha/2}\ell(\widehat{\theta})\},$$

where $\ell(\widehat{\theta}) = \sup_{\theta \in \Theta} \ell(\theta)$, $\ell(\vartheta, \widehat{\varphi}_\vartheta) = \sup_\varphi \ell(\vartheta, \varphi)$, and c_α is a constant related to the significance level α.
Under the conditions of Theorem 6.5, if c_α is chosen to be $\chi^2_{r,\alpha}$, the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r, then

$$C(X) = \{\vartheta : \ell(\vartheta, \widehat{\varphi}_\vartheta) \geq e^{-c_\alpha/2}\ell(\widehat{\theta})\}$$

is a $1-\alpha$ asymptotically correct confidence set.
Note that this confidence set and the one given by

$$C(X) = \{\theta : \|\widehat{V}_n^{-1/2}(\widehat{\theta}_n - \theta)\|^2 \le \chi^2_{k,\alpha}\}$$

are generally different.
In many cases $-\ell(\vartheta,\varphi)$ is a convex function of ϑ and, therefore, $C(X)$ based on LR tests is a bounded set in $\mathscr{R}^k$; in particular, $C(X)$ is a bounded interval when $k=1$.
In §6.4.2 we discussed two asymptotic tests closely related to the LR test: Wald's test and Rao's score test.
When $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$, Wald's test has acceptance region

$$A(\vartheta_0) = \{x : (\widehat{\vartheta} - \vartheta_0)^\tau \{C^\tau [I_n(\widehat{\theta})]^{-1} C\}^{-1} (\widehat{\vartheta} - \vartheta_0) \le \chi^2_{r,\alpha}\},$$

where $\widehat{\theta} = (\widehat{\vartheta}, \widehat{\varphi})$ is an MLE or RLE of $\theta = (\vartheta, \varphi)$, $I_n(\theta)$ is the Fisher information matrix based on X, $C^\tau = (\ I_r\ 0\ )$, and 0 is an $r \times (k-r)$ matrix of 0's.
By Theorem 4.17, the confidence set obtained by inverting $A(\vartheta)$ is

$$C(X) = \{\theta : \|\widehat{V}_n^{-1/2}(\widehat{\vartheta} - \vartheta)\|^2 \le \chi^2_{k,\alpha}\}$$

with $\widehat{V}_n = C^\tau[I_n(\widehat{\theta})]^{-1}C$.
When $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$, Rao's score test has acceptance region

$$A(\vartheta_0) = \{x : [s_n(\vartheta_0, \widehat{\varphi}_{\vartheta_0})]^\tau [I_n(\vartheta_0, \widehat{\varphi}_{\vartheta_0})]^{-1} s_n(\vartheta_0, \widehat{\varphi}_{\vartheta_0}) \le \chi^2_{r,\alpha}\},$$

where $s_n(\theta) = \partial \log \ell(\theta)/\partial\theta$.
The confidence set obtained by inverting $A(\vartheta)$ is also $1-\alpha$ asymptotically correct.

Example 7.23

Let $X_1, ..., X_n$ be i.i.d. binary random variables with $p = P(X_i = 1)$.
Since confidence sets for p with a given confidence coefficient are usually randomized (§7.2.3), asymptotically correct confidence sets may be considered when n is large.
The likelihood ratio for testing $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ is

$$\lambda(Y) = p_0^Y (1-p_0)^{n-Y} / \widehat{p}^Y (1-\widehat{p})^{n-Y},$$

where $Y = \sum_{i=1}^n X_i$ and $\widehat{p} = Y/n$ is the MLE of p.

The confidence set based on LR tests is equal to

$$C_1(X) = \{p : p^Y (1-p)^{n-Y} \geq e^{-c_\alpha/2} \widehat{p}^Y (1-\widehat{p})^{n-Y}\}.$$

When $0 < Y < n$, $-p^Y(1-p)^{n-Y}$ is strictly convex and equals 0 if $p = 0$ or 1 and, hence, $C_1(X) = [\underline{p}, \overline{p}]$ with $0 < \underline{p} < \overline{p} < 1$.
When $Y = 0$, $(1-p)^n$ is strictly decreasing and, therefore, $C_1(X) = (0, \overline{p}]$ with $0 < \overline{p} < 1$.
Similarly, when $Y = n$, $C_1(X) = [\underline{p}, 1)$ with $0 < \underline{p} < 1$.

The confidence set obtained by inverting acceptance regions of Wald's tests is simply

$$C_2(X) = [\widehat{p} - z_{1-\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}, \widehat{p} + z_{1-\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}],$$

since $I_n(p) = n/[p(1-p)]$ and $(\chi^2_{1,\alpha})^{1/2} = z_{1-\alpha/2}$, the $(1-\alpha/2)$th quantile of $N(0,1)$.
Note that

$$s_n(p) = \frac{Y}{p} - \frac{n-Y}{1-p} = \frac{Y-pn}{p(1-p)}$$

and
$$[s_n(p)]^2[I_n(p)]^{-1} = \frac{(Y-pn)^2}{p^2(1-p)^2}\frac{p(1-p)}{n} = \frac{n(\widehat{p}-p)^2}{p(1-p)}.$$
Hence, the confidence set obtained by inverting acceptance regions of Rao's score tests is
$$C_3(X) = \{p : n(\widehat{p}-p)^2 \le p(1-p)\chi^2_{1,\alpha}\}.$$
It can be shown (exercise) that $C_3(X) = [p_-, p_+]$ with
$$p_\pm = \frac{2Y + \chi^2_{1,\alpha} \pm \sqrt{\chi^2_{1,\alpha}[4n\widehat{p}(1-\widehat{p}) + \chi^2_{1,\alpha}]}}{2(n+\chi^2_{1,\alpha})}.$$

Example 7.24

Let $X_1,...,X_n$ be i.i.d. from $N(\mu,\varphi)$ with unknown $\theta = (\mu,\varphi)$.
Consider the problem of constructing a $1-\alpha$ asymptotically correct confidence set for θ.
The log-likelihood function is
$$\log \ell(\theta) = -\frac{1}{2\varphi}\sum_{i=1}^n (X_i-\mu)^2 - \frac{n}{2}\log\varphi - \frac{n}{2}\log(2\pi).$$

Since $(\bar{X}, \widehat{\varphi})$ is the MLE of θ, where $\widehat{\varphi} = (n-1)S^2/n$, the confidence set based on LR tests is

$$C_1(X) = \left\{ \theta : \frac{1}{\varphi} \sum_{i=1}^{n} (X_i - \mu)^2 + n \log \varphi \leq \chi^2_{2,\alpha} + n + n \log \widehat{\varphi} \right\}.$$

Note that

$$s_n(\theta) = \left(\frac{n(\bar{X} - \mu)}{\varphi}, \frac{1}{2\varphi^2} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{n}{2\varphi} \right) \qquad I_n(\theta) = \begin{pmatrix} \frac{n}{\varphi} & 0 \\ 0 & \frac{n}{2\varphi^2} \end{pmatrix}.$$

Hence, the confidence set based on Wald's tests is

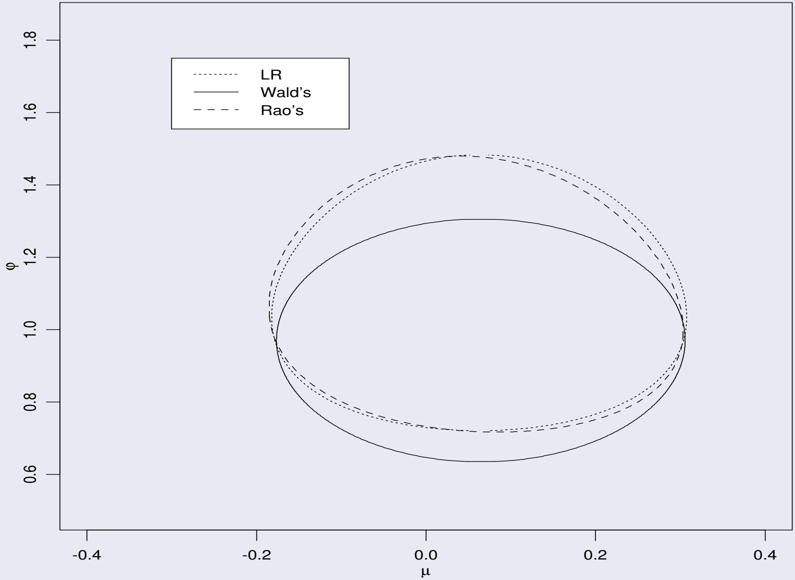
$$C_2(X) = \left\{ \theta : \frac{(\bar{X} - \mu)^2}{\widehat{\varphi}} + \frac{(\widehat{\varphi} - \varphi)^2}{2\widehat{\varphi}^2} \leq \frac{\chi^2_{2,\alpha}}{n} \right\},$$

which is an ellipsoid in $\mathscr{R}^2$, and the confidence set based on Rao's score tests is

$$C_3(X) = \left\{ \theta : \frac{(\bar{X} - \mu)^2}{\varphi} + \frac{1}{2} \left[\frac{1}{n\varphi} \sum_{i=1}^{n} (X_i - \mu)^2 - 1 \right]^2 \leq \frac{\chi^2_{2,\alpha}}{n} \right\}.$$

In general, $C_j(X)$, $j = 1, 2, 3$, are different.

Figure 7.2. Confidence sets obtained by inverting LR, Wald's, and Rao's score tests in Example 7.24



Example 7.24 (continued)

Consider now the construction of a confidence set for μ.

The confidence set based on Wald's tests is defined by $C_2(X)$ with φ replaced by $\widehat{\varphi}$ and $\chi^2_{2,\alpha}$ replaced by $\chi^2_{1,\alpha} = z^2_{\alpha/2}$, which results in the confidence interval

$$\{\mu : n(\bar{X}-\mu)^2 \le z^2_{\alpha/2}\widehat{\varphi}\} = [\bar{X} - z_{\alpha/2}\sqrt{\widehat{\varphi}/n},\ \bar{X} + z_{\alpha/2}\sqrt{\widehat{\varphi}/n}]$$

The confidence set based on the LR tests is defined by $C_1(X)$ with $\chi^2_{2,\alpha}$ and φ replaced by $\chi^2_{1,\alpha} = z^2_{\alpha/2}$ and $n^{-1}\sum_{i=1}^n (X_i-\mu)^2 = \widehat{\varphi} + (\bar{X}-\mu)^2$, respectively, which leads to the confidence interval

$$\begin{aligned}
& \{\mu : n + n\log(\widehat{\varphi} + (\bar{X}-\mu)^2) \le z^2_{\alpha/2} + n + n\log\widehat{\varphi}\} \\
= \; & \{\mu : \widehat{\varphi} + (\bar{X}-\mu)^2 \le \exp(\log\widehat{\varphi} + z^2_{\alpha/2}/n)\} \\
= \; & \{\mu : (\bar{X}-\mu)^2 \le \widehat{\varphi}[\exp(z^2_{\alpha/2}/n) - 1]\} \\
= \; & [\bar{X} - \sqrt{\widehat{\varphi}}\sqrt{\exp(z^2_{\alpha/2}/n) - 1},\ \bar{X} + \sqrt{\widehat{\varphi}}\sqrt{\exp(z^2_{\alpha/2}/n) - 1}]
\end{aligned}$$

The confidence set based on Rao's score tests is defined by $C_3(X)$ with $\chi^2_{2,\alpha}$ and φ replaced by $z^2_{\alpha/2}$ and $n^{-1}\sum_{i=1}^n (X_i-\mu)^2 = \widehat{\varphi} + (\bar{X}-\mu)^2$, respectively, which results in the confidence interval

$$\{\mu : n(\bar{X}-\mu)^2 \le z^2_{\alpha/2}[\widehat{\varphi} + (\bar{X}-\mu)^2]\}$$
$$= [\bar{X} - z_{\alpha/2}\sqrt{\widehat{\varphi}/n}\sqrt{1-z^2_{\alpha/2}/n},\ \bar{X} + z_{\alpha/2}\sqrt{\widehat{\varphi}/n}\sqrt{1-z^2_{\alpha/2}/n}]$$

Confidence intervals for quantiles

Let $X_1,...,X_n$ be i.i.d. from a continuous c.d.f. F on $\mathscr{R}$ and let $\theta = F^{-1}(p)$ be the pth quantile of F, $0 < p < 1$.
The general methods we previously discussed can be applied to obtain a confidence set for θ, but we introduce here a method that works particularly for quantile problems.
In fact, for any given α, it is possible to derive a confidence interval (or bound) for θ with confidence coefficient $1-\alpha$ (Exercise 84), but the computation of such a confidence interval may be cumbersome.
We focus on asymptotic confidence intervals for θ.
Our result is based on the following result due to Bahadur (1966).

Theorem 7.8 (refinement of Bahadur's representation)

Let $X_1,...,X_n$ be i.i.d. from a continuous c.d.f. F on $\mathscr{R}$ that is twice differentiable at $\theta = F^{-1}(p)$, $0 < p < 1$, with $F'(\theta) > 0$.
Let F_n be the empirical c.d.f.
Let $\{k_n\}$ be a sequence of integers satisfying $1 \le k_n \le n$ and $k_n/n = p + o((\log n)^\delta/\sqrt{n})$ for some $\delta > 0$.
Then

$$X_{(k_n)} = \theta + \frac{(k_n/n) - F_n(\theta)}{F'(\theta)} + O\left(\frac{(\log n)^{(1+\delta)/2}}{n^{3/4}}\right) \quad \text{a.s.}$$

Corollary 7.1

Assume the conditions in Theorem 7.8 and
$k_n/n = p + cn^{-1/2} + o(n^{-1/2})$ with a constant c.
Then

$$\sqrt{n}(X_{(k_n)} - F_n^{-1}(p)) \to_{a.s.} c/F'(\theta).$$

Using Corollary 7.1, we can obtain a confidence interval for θ with limiting confidence coefficient $1-\alpha$ (Definition 2.14) for any given α.

Corollary 7.2

Assume the conditions in Theorem 7.8.
Let $\{k_{1n}\}$ and $\{k_{2n}\}$ be two sequences of integers satisfying $1 \le k_{1n} < k_{2n} \le n$,

$$k_{1n}/n = p - z_{1-\alpha/2}\sqrt{p(1-p)/n} + o(n^{-1/2}),$$

and

$$k_{2n}/n = p + z_{1-\alpha/2}\sqrt{p(1-p)/n} + o(n^{-1/2}),$$

where $z_a = \Phi^{-1}(a)$. Then the confidence interval $C(X) = [X_{(k_{1n})}, X_{(k_{2n})}]$ has the property that $P(\theta \in C(X))$ does not depend on P and

$$\lim_{n\to\infty} \inf_{P\in\mathscr{P}} P\big(\theta \in C(X)\big) = \lim_{n\to\infty} P\big(\theta \in C(X)\big) = 1-\alpha.$$

Furthermore,

$$\text{the length of } C(X) = \frac{2z_{1-\alpha/2}\sqrt{p(1-p)}}{F'(\theta)\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{a.s.}$$

Proof

Note that $P(\theta \in C(X)) = P(X_{(k_{1n})} \leq \theta \leq X_{(k_{2n})}) = P(U_{(k_{1n})} \leq p \leq U_{(k_{2n})})$, where $U_{(k)}$ is the kth order statistic based on a sample $U_1, ..., U_n$ i.i.d. from the uniform distribution $U(0,1)$ (Exercise 84).
Hence, $P(\theta \in C(X))$ does not depend on P and $\lim_{n\to\infty} P\big(\theta \in C(X)\big) = \lim_{n\to\infty} \inf_{P\in\mathscr{P}} P\big(\theta \in C(X)\big)$.
By Corollary 7.1, Theorem 5.10, and Slutsky's theorem,

$$
\begin{aligned}
P(X_{(k_{1n})} > \theta) &= P\left(F_n^{-1}(p) - z_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{F'(\theta)\sqrt{n}} + o_p(n^{-1/2}) > \theta\right) \\
&= P\left(\frac{\sqrt{n}(F_n^{-1}(p) - \theta)}{\sqrt{p(1-p)}/F'(\theta)} + o_p(1) > z_{1-\alpha/2}\right) \\
&\to 1 - \Phi(z_{1-\alpha/2}) \\
&= \alpha/2.
\end{aligned}
$$

The first result follows, since similarly $P(X_{(k_{2n})} < \theta) \to \alpha/2$.
The result for the length of $C(X)$ follows directly from Corollary 7.1.

Remarks

- The confidence interval $[X_{(k_{1n})}, X_{(k_{2n})}]$ given in Corollary 7.2 is called Woodruff's (1952) interval.
- It has limiting confidence coefficient $1 - \alpha$, a property that is stronger than the $1 - \alpha$ asymptotic correctness.
- The length of Woodruff's interval is $X_{(k_{2n})} - X_{(k_{1n})}$.
  By the result in Corollary 7.2,

$$X_{(k_{2n})} - X_{(k_{1n})} = \frac{2z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}F'(\theta)} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{a.s.},$$

  This means

$$\frac{[X_{(k_{2n})} - X_{(k_{1n})}]^2}{4z_{\alpha/2}^2} = \frac{p(1-p)}{n[F'(\theta)]^2} + o\left(\frac{1}{n}\right) \quad \text{a.s.}$$

  Therefore, $[X_{(k_{2n})} - X_{(k_{1n})}]^2/(4z_{\alpha/2}^2)$ is a consistent estimator of the asymptotic variance of the sample pth quantile.

Remarks

- From Theorem 5.10, if $F'(\theta)$ exists and is positive, then

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to_d N\left(0, \frac{p(1-p)}{[F'(\theta)]^2}\right).$$

If the derivative $F'(\theta)$ has a consistent estimator $\widehat{d}_n$ obtained using some method such as one of those introduced in §5.1.3, then $\widehat{V}_n = p(1-p)/\widehat{d}_n^2$ is a consistent estimator of $p(1-p)/[F'(\theta)]^2$ and the method introduced in §7.3.1 can be applied to derive the following $1-\alpha$ asymptotically correct confidence interval:

$$C_1(X) = \left[\widehat{\theta}_n - z_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\widehat{d}_n\sqrt{n}}, \widehat{\theta}_n + z_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\widehat{d}_n\sqrt{n}}\right].$$

The length of $C_1(X)$ is asymptotically almost the same as Woodruff's interval.
However, $C_1(X)$ depends on the estimated derivative $\widehat{d}_n$ and it is usually difficult to obtain a precise estimator $\widehat{d}_n$.