

Stat 710: Mathematical Statistics

Lecture 38

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 38: Asymptotic confidence sets and likelihoods

Asymptotic criterion

In some problems, especially in nonparametric problems, it is difficult to find a reasonable confidence set with a given confidence coefficient or confidence level $1 - \alpha$.

A common approach is to find a confidence set whose confidence coefficient or confidence level is nearly $1 - \alpha$ when the sample size n is large.

A confidence set $C(X)$ for θ has asymptotic confidence level $1 - \alpha$ if $\liminf_n P(\theta \in C(X)) \geq 1 - \alpha$ for any $P \in \mathcal{P}$ (Definition 2.14).

If $\lim_{n \rightarrow \infty} P(\theta \in C(X)) = 1 - \alpha$ for any $P \in \mathcal{P}$, then $C(X)$ is a $1 - \alpha$ *asymptotically correct* confidence set.

Note that asymptotic correctness is not the same as having limiting confidence coefficient $1 - \alpha$ (Definition 2.14).

Asymptotically pivotal quantities

A known Borel function of (X, θ) , $\mathfrak{R}_n(X, \theta)$, is said to be *asymptotically pivotal* iff the limiting distribution of $\mathfrak{R}_n(X, \theta)$ does not depend on P . Like a pivotal quantity in constructing confidence sets (§7.1.1) with a given confidence coefficient or confidence level, an asymptotically pivotal quantity can be used in constructing asymptotically correct confidence sets.

Most asymptotically pivotal quantities are of the form $\widehat{V}_n^{-1/2}(\widehat{\theta}_n - \theta)$, where $\widehat{\theta}_n$ is an estimator of θ that is asymptotically normal, i.e.,

$$V_n^{-1/2}(\widehat{\theta}_n - \theta) \rightarrow_d N_k(0, I_k),$$

\widehat{V}_n is a consistent estimator of the asymptotic covariance matrix V_n . The resulting $1 - \alpha$ asymptotically correct confidence sets are

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

where $\chi_{k,\alpha}^2$ is the $(1 - \alpha)$ th quantile of the chi-square distribution χ_k^2 . If θ is real-valued ($k = 1$), then $C(X)$ is a confidence interval. When $k > 1$, $C(X)$ is an ellipsoid.

Asymptotically pivotal quantities

A known Borel function of (X, θ) , $\mathfrak{R}_n(X, \theta)$, is said to be *asymptotically pivotal* iff the limiting distribution of $\mathfrak{R}_n(X, \theta)$ does not depend on P . Like a pivotal quantity in constructing confidence sets (§7.1.1) with a given confidence coefficient or confidence level, an asymptotically pivotal quantity can be used in constructing asymptotically correct confidence sets.

Most asymptotically pivotal quantities are of the form $\widehat{V}_n^{-1/2}(\widehat{\theta}_n - \theta)$, where $\widehat{\theta}_n$ is an estimator of θ that is asymptotically normal, i.e.,

$$V_n^{-1/2}(\widehat{\theta}_n - \theta) \rightarrow_d N_k(0, I_k),$$

\widehat{V}_n is a consistent estimator of the asymptotic covariance matrix V_n . The resulting $1 - \alpha$ asymptotically correct confidence sets are

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

where $\chi_{k,\alpha}^2$ is the $(1 - \alpha)$ th quantile of the chi-square distribution χ_k^2 . If θ is real-valued ($k = 1$), then $C(X)$ is a confidence interval. When $k > 1$, $C(X)$ is an ellipsoid.

Example 7.20 (Functions of means)

Suppose that X_1, \dots, X_n are i.i.d. random vectors having a c.d.f. F on \mathcal{R}^d and that the unknown parameter of interest is $\theta = g(\mu)$, where $\mu = E(X_1)$ and g is a known differentiable function from \mathcal{R}^d to \mathcal{R}^k , $k \leq d$.

From the CLT, Theorem 1.12, $\hat{\theta}_n = g(\bar{X})$ satisfies

$$V_n^{-1/2}(\hat{\theta}_n - \theta) \rightarrow_d N_k(0, I_k)$$

$$V_n = [\nabla g(\mu)]^\tau \text{Var}(X_1) \nabla g(\mu) / n$$

A consistent estimator of the asymptotic covariance matrix V_n is

$$\hat{V}_n = [\nabla g(\bar{X})]^\tau S^2 \nabla g(\bar{X}) / n.$$

Thus,

$$C(X) = \{\theta : \|\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta)\|^2 \leq \chi_{k,\alpha}^2\},$$

is a $1 - \alpha$ asymptotically correct confidence set for θ .

Example 7.22 (Linear models)

Consider linear model $X = Z\beta + \varepsilon$, where ε has i.i.d. components with mean 0 and variance σ^2 .

Assume that Z is of full rank and that the conditions in Theorem 3.12 hold.

It follows from Theorem 1.9(iii) and Theorem 3.12 that for the LSE $\hat{\beta}$,

$$V_n^{-1/2}(\hat{\beta} - \beta) \rightarrow_d N_p(0, I_p)$$

$$V_n = \sigma^2(Z^T Z)^{-1}$$

A consistent estimator for V_n is $\hat{V}_n = (n - p)^{-1} SSR(Z^T Z)^{-1}$ (see §5.5.1).

Thus, a $1 - \alpha$ asymptotically correct confidence set for β is

$$C(X) = \{\beta : (\hat{\beta} - \beta)^T (Z^T Z) (\hat{\beta} - \beta) \leq \chi_{p, \alpha}^2 SSR / (n - p)\}.$$

Note that this confidence set is different from the one in Example 7.9 derived under the normality assumption on ε .

Discussions

The method of using asymptotically pivotal quantities can also be applied to parametric problems.

Note that in a parametric problem where the unknown parameter θ is multivariate, a confidence set for θ with a given confidence coefficient may be difficult or impossible to obtain.

Asymptotically correct confidence sets for θ can also be constructed by inverting acceptance regions of asymptotic tests for testing $H_0 : \theta = \theta_0$ versus some H_1 .

Asymptotic efficiency

Typically, in a given problem there exist many different asymptotically pivotal quantities that lead to different $1 - \alpha$ asymptotically correct confidence sets for θ .

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

Discussions

The method of using asymptotically pivotal quantities can also be applied to parametric problems.

Note that in a parametric problem where the unknown parameter θ is multivariate, a confidence set for θ with a given confidence coefficient may be difficult or impossible to obtain.

Asymptotically correct confidence sets for θ can also be constructed by inverting acceptance regions of asymptotic tests for testing $H_0 : \theta = \theta_0$ versus some H_1 .

Asymptotic efficiency

Typically, in a given problem there exist many different asymptotically pivotal quantities that lead to different $1 - \alpha$ asymptotically correct confidence sets for θ .

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

Proposition 7.4

Let

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

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

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

where \widehat{V}_{jn} is consistent for V_{jn} , $j = 1, 2$.

If $\text{Det}(V_{1n}) < \text{Det}(V_{2n})$ for sufficiently large n , where $\text{Det}(A)$ is the determinant of A , then

$$P(\text{vol}(C_1(X)) < \text{vol}(C_2(X))) \rightarrow 1.$$

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

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

Proposition 7.4

Let

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

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

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

where \widehat{V}_{jn} is consistent for V_{jn} , $j = 1, 2$.

If $\text{Det}(V_{1n}) < \text{Det}(V_{2n})$ for sufficiently large n , where $\text{Det}(A)$ is the determinant of A , then

$$P(\text{vol}(C_1(X)) < \text{vol}(C_2(X))) \rightarrow 1.$$

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

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

Asymptotic efficiency

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

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

If $\hat{\theta}_n$ 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 $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is a parametric family dominated by a σ -finite measure, where $\Theta \subset \mathcal{R}^k$.

Consider $\theta = (\vartheta, \varphi)$ and confidence sets for ϑ with dimension r .

Asymptotic efficiency

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

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

If $\hat{\theta}_n$ 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 $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is a parametric family dominated by a σ -finite measure, where $\Theta \subset \mathcal{R}^k$.

Consider $\theta = (\vartheta, \varphi)$ and confidence sets for ϑ with dimension r .

Parametric likelihoods

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

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

where $\ell(\hat{\theta}) = \sup_{\theta \in \Theta} \ell(\theta)$, $\ell(\vartheta, \hat{\varphi}_\vartheta) = \sup_{\varphi} \ell(\vartheta, \varphi)$, 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 $\chi_{r,\alpha}^2$, the $(1 - \alpha)$ th quantile of the chi-square distribution χ_r^2 , then

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

is a $1 - \alpha$ asymptotically correct confidence set.

Note that this confidence set and the one given by

$$C(X) = \{\theta : \|\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta)\|^2 \leq \chi_{k,\alpha}^2\}$$

are generally different.

Parametric likelihoods

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 \mathcal{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

$$A(\vartheta_0) = \{x : (\hat{\vartheta} - \vartheta_0)^\tau \{C^\tau [I_n(\hat{\theta})]^{-1} C\}^{-1} (\hat{\vartheta} - \vartheta_0) \leq \chi_{r,\alpha}^2\},$$

where $\hat{\theta} = (\hat{\vartheta}, \hat{\varphi})$ 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

$$C(X) = \{\theta : \|\hat{V}_n^{-1/2}(\hat{\vartheta} - \vartheta)\|^2 \leq \chi_{k,\alpha}^2\}$$

with $\hat{V}_n = C^\tau [I_n(\hat{\theta})]^{-1} C$.

Parametric likelihoods

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 \mathcal{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

$$A(\vartheta_0) = \{x : (\hat{\vartheta} - \vartheta_0)^\tau \{C^\tau [I_n(\hat{\theta})]^{-1} C\}^{-1} (\hat{\vartheta} - \vartheta_0) \leq \chi_{r,\alpha}^2\},$$

where $\hat{\theta} = (\hat{\vartheta}, \hat{\varphi})$ 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

$$C(X) = \{\theta : \|\hat{V}_n^{-1/2}(\hat{\vartheta} - \vartheta)\|^2 \leq \chi_{k,\alpha}^2\}$$

with $\hat{V}_n = C^\tau [I_n(\hat{\theta})]^{-1} C$.

Parametric likelihoods

When $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$, Rao's score test has acceptance region

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

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, \dots, 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

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

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

Parametric likelihoods

When $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$, Rao's score test has acceptance region

$$A(\vartheta_0) = \{x : [s_n(\vartheta_0, \hat{\varphi}_{\vartheta_0})]^T [I_n(\vartheta_0, \hat{\varphi}_{\vartheta_0})]^{-1} s_n(\vartheta_0, \hat{\varphi}_{\vartheta_0}) \leq \chi_{r,\alpha}^2\},$$

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, \dots, 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

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

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

Example 7.23 (continued)

The confidence set based on LR tests is equal to

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

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

When $Y = 0$, $(1-p)^n$ is strictly decreasing and, therefore, $C_1(X) = (0, \bar{p}]$ with $0 < \bar{p} < 1$.

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

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

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

Example 7.23 (continued)

The confidence set based on LR tests is equal to

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

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

When $Y = 0$, $(1-p)^n$ is strictly decreasing and, therefore, $C_1(X) = (0, \bar{p}]$ with $0 < \bar{p} < 1$.

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

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

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

Example 7.23 (continued)

Note that

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

and

$$[s_n(p)]^2 [I_n(p)]^{-1} = \frac{(Y-pn)^2}{p^2(1-p)^2} \frac{p(1-p)}{n} = \frac{n(\hat{p}-p)^2}{p(1-p)}.$$

Hence, the confidence set obtained by inverting acceptance regions of Rao's score tests is

$$C_3(X) = \{p : n(\hat{p}-p)^2 \leq p(1-p)\chi_{1,\alpha}^2\}.$$

It can be shown (exercise) that $C_3(X) = [p_-, p_+]$ with

$$p_{\pm} = \frac{2Y + \chi_{1,\alpha}^2 \pm \sqrt{\chi_{1,\alpha}^2 [4n\hat{p}(1-\hat{p}) + \chi_{1,\alpha}^2]}}{2(n + \chi_{1,\alpha}^2)}.$$