Stat 710: Mathematical Statistics
Lecture 40

Jun Shao

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

Lecture 40: Simultaneous confidence intervals and Scheffe's method

So far we have studied confidence sets for a real-valued θ or a vector-valued θ with a finite dimension k.
In some applications, we need a confidence set for real-valued θ_t with $t \in \mathscr{T}$, where $\mathscr{T}$ is an index set that may contain infinitely many elements, for example, $\mathscr{T} = [0,1]$ or $\mathscr{T} = \mathscr{R}$.

Definition 7.6

Let X be a sample from $P \in \mathscr{P}$, let θ_t, $t \in \mathscr{T}$, be real-valued parameters related to P, and let $C_t(X)$, $t \in \mathscr{T}$, be a class of (one-sided or two-sided) confidence intervals.
(i) Intervals $C_t(X)$, $t \in \mathscr{T}$, are level $1-\alpha$ *simultaneous confidence intervals* for θ_t, $t \in \mathscr{T}$, iff

$$\inf_{P \in \mathscr{P}} P\left(\theta_t \in C_t(X) \text{ for all } t \in \mathscr{T}\right) \geq 1-\alpha.$$

The left-hand side of the above expression is the confidence coefficient of $C_t(X)$, $t \in \mathscr{T}$.

Lecture 40: Simultaneous confidence intervals and Scheffe's method

So far we have studied confidence sets for a real-valued θ or a vector-valued θ with a finite dimension k.
In some applications, we need a confidence set for real-valued θ_t with $t \in \mathscr{T}$, where $\mathscr{T}$ is an index set that may contain infinitely many elements, for example, $\mathscr{T} = [0,1]$ or $\mathscr{T} = \mathscr{R}$.

Definition 7.6

Let X be a sample from $P \in \mathscr{P}$, let θ_t, $t \in \mathscr{T}$, be real-valued parameters related to P, and let $C_t(X)$, $t \in \mathscr{T}$, be a class of (one-sided or two-sided) confidence intervals.
(i) Intervals $C_t(X)$, $t \in \mathscr{T}$, are level $1-\alpha$ *simultaneous confidence intervals* for θ_t, $t \in \mathscr{T}$, iff

$$\inf_{P \in \mathscr{P}} P\left(\theta_t \in C_t(X) \text{ for all } t \in \mathscr{T}\right) \geq 1-\alpha.$$

The left-hand side of the above expression is the confidence coefficient of $C_t(X)$, $t \in \mathscr{T}$.

Definition 7.6 (continued)

(ii) Intervals $C_t(X)$, $t \in \mathscr{T}$, are simultaneous confidence intervals for θ_t, $t \in \mathscr{T}$, with asymptotic confidence level $1-\alpha$ iff

$$\lim_{n\to\infty} P\left(\theta_t \in C_t(X) \text{ for all } t \in \mathscr{T}\right) \geq 1-\alpha.$$

Intervals $C_t(X)$, $t \in \mathscr{T}$, are $1-\alpha$ asymptotically correct iff the equality in the above expression holds.

If the index set $\mathscr{T}$ contains $k < \infty$ elements, then $\theta = (\theta_t, t \in \mathscr{T})$ is a k-vector and the methods studied in the previous sections can be applied to construct a level $1-\alpha$ confidence set $C(X)$ for θ.
If $C(X)$ can be expressed as $\prod_{t\in\mathscr{T}} C_t(X)$ for some intervals $C_t(X)$, then $C_t(X)$, $t \in \mathscr{T}$, are level $1-\alpha$ simultaneous confidence intervals.
This simple method, however, does not always work.

A simple method for the case of $\mathscr{T}$ containing finite many elements is introduced next.

Definition 7.6 (continued)

(ii) Intervals $C_t(X)$, $t \in \mathscr{T}$, are simultaneous confidence intervals for θ_t, $t \in \mathscr{T}$, with asymptotic confidence level $1-\alpha$ iff

$$\lim_{n\to\infty} P\big(\theta_t \in C_t(X) \text{ for all } t \in \mathscr{T}\big) \geq 1-\alpha.$$

Intervals $C_t(X)$, $t \in \mathscr{T}$, are $1-\alpha$ asymptotically correct iff the equality in the above expression holds.

If the index set $\mathscr{T}$ contains $k < \infty$ elements, then $\theta = (\theta_t, t \in \mathscr{T})$ is a k-vector and the methods studied in the previous sections can be applied to construct a level $1-\alpha$ confidence set $C(X)$ for θ.
If $C(X)$ can be expressed as $\prod_{t\in\mathscr{T}} C_t(X)$ for some intervals $C_t(X)$, then $C_t(X)$, $t \in \mathscr{T}$, are level $1-\alpha$ simultaneous confidence intervals.
This simple method, however, does not always work.

A simple method for the case of $\mathscr{T}$ containing finite many elements is introduced next.

Bonferroni's method

Bonferroni's method, which works when $\mathscr{T}$ contains $k < \infty$ elements, is based on the following simple inequality for k events $A_1, \ldots, A_k$:

$$P\left(\bigcup_{i=1}^{k} A_i\right) \leq \sum_{i=1}^{k} P(A_i) \quad \text{i.e.,} \quad P\left(\bigcap_{i=1}^{k} A_i^c\right) \geq 1 - \sum_{i=1}^{k} P(A_i).$$

For each $t \in \mathscr{T}$, let $C_t(X)$ be a level $1-\alpha_t$ confidence interval for θ_t.
If α_t's are chosen so that $\sum_{t\in\mathscr{T}} \alpha_t = \alpha$ (e.g., $\alpha_t = \alpha/k$ for all t), then Bonferroni's simultaneous confidence intervals are $C_t(X)$, $t \in \mathscr{T}$, since

$$P(\theta_t \in C_t(X), t \in \mathscr{T}) \geq 1 - \sum_{t\in\mathscr{T}} P(\theta_t \notin C_t(X)) = 1 - \sum_{t\in\mathscr{T}} \alpha_t = 1 - \alpha$$

Bonferroni's intervals are of level $1-\alpha$, but they are not of confidence coefficient $1-\alpha$ even if $C_t(X)$ has confidence coefficient $1-\alpha_t$ for any fixed t.

Bonferroni's method does not require that $C_t(X)$, $t \in \mathscr{T}$, be independent.

Bonferroni's method is easy to use, but may be too conservative.

Bonferroni's method

Bonferroni's method, which works when $\mathscr{T}$ contains $k < \infty$ elements, is based on the following simple inequality for k events $A_1, ..., A_k$:

$$P\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k P(A_i) \quad \text{i.e.,} \quad P\left(\bigcap_{i=1}^k A_i^c\right) \geq 1 - \sum_{i=1}^k P(A_i).$$

For each $t \in \mathscr{T}$, let $C_t(X)$ be a level $1 - \alpha_t$ confidence interval for θ_t.
If α_t's are chosen so that $\sum_{t \in \mathscr{T}} \alpha_t = \alpha$ (e.g., $\alpha_t = \alpha/k$ for all t), then Bonferroni's simultaneous confidence intervals are $C_t(X)$, $t \in \mathscr{T}$, since

$$P(\theta_t \in C_t(X), t \in \mathscr{T}) \geq 1 - \sum_{t \in \mathscr{T}} P(\theta_t \notin C_t(X)) = 1 - \sum_{t \in \mathscr{T}} \alpha_t = 1 - \alpha$$

Bonferroni's intervals are of level $1 - \alpha$, but they are not of confidence coefficient $1 - \alpha$ even if $C_t(X)$ has confidence coefficient $1 - \alpha_t$ for any fixed t.
Bonferroni's method does not require that $C_t(X)$, $t \in \mathscr{T}$, be independent.
Bonferroni's method is easy to use, but may be too conservative.

Scheffé's method

Consider the normal linear model

$$X = N_n(Z\beta, \sigma^2 I_n),$$

where β is a p-vector of unknown parameters, $\sigma^2 > 0$ is unknown, and Z is an $n \times p$ known matrix of rank $r \le p$.
Let L be an $s \times p$ matrix of rank $s \le r$.
Suppose that $\mathscr{R}(L) \subset \mathscr{R}(Z)$ and we would like to construct simultaneous confidence intervals for $t^\tau L\beta$, where $t \in \mathscr{T} = \mathscr{R}^s - \{0\}$.
Using the argument in Example 7.15, for each $t \in \mathscr{T}$, we can obtain the following confidence interval for $t^\tau L\beta$ with confidence coefficient $1-\alpha$:

$$\left[t^\tau L\widehat{\beta} - t_{n-r,\alpha/2}\widehat{\sigma}\sqrt{t^\tau Dt},\, t^\tau L\widehat{\beta} + t_{n-r,\alpha/2}\widehat{\sigma}\sqrt{t^\tau Dt} \right],$$

where $\widehat{\beta}$ is the LSE of β, $\widehat{\sigma}^2 = \|X - Z\widehat{\beta}\|^2/(n-r)$, $D = L(Z^\tau Z)^- L^\tau$, and $t_{n-r,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_{n-r}.
However, these intervals are not level $1-\alpha$ simultaneous confidence intervals for $t^\tau L\beta$, $t \in \mathscr{T}$.

Scheffé's method

Scheffé's (1959) method of constructing simultaneous confidence intervals for $t^\tau L\beta$ is based on the following equality (exercise):

$$x^\tau A^{-1} x = \max_{y \in \mathscr{R}^k, y \neq 0} \frac{(y^\tau x)^2}{y^\tau A y},$$

where $x \in \mathscr{R}^k$ and A is a $k \times k$ positive definite matrix.

Theorem 7.10

Assume the normal linear model.
Let L be an $s \times p$ matrix of rank $s \leq r$.
Assume that $\mathscr{R}(L) \subset \mathscr{R}(Z)$ and $D = L(Z^\tau Z)^- L^\tau$ is of full rank.
Then

$$C_t(X) = \left[t^\tau L\widehat{\beta} - \widehat{\sigma}\sqrt{s c_\alpha t^\tau D t}, t^\tau L\widehat{\beta} + \widehat{\sigma}\sqrt{s c_\alpha t^\tau D t} \right], \quad t \in \mathscr{T},$$

are simultaneous confidence intervals for $t^\tau L\beta$, $t \in \mathscr{T}$, with confidence coefficient $1-\alpha$, where $\widehat{\sigma}^2 = \|X - Z\widehat{\beta}\|^2/(n-r)$, $\mathscr{T} = \mathscr{R}^s - \{0\}$, and c_α is the $(1-\alpha)$th quantile of the F-distribution $F_{s,n-r}$.

Scheffé's method

Scheffé's (1959) method of constructing simultaneous confidence intervals for $t^\tau L\beta$ is based on the following equality (exercise):

$$x^\tau A^{-1}x = \max_{y\in\mathscr{R}^k, y\neq 0} \frac{(y^\tau x)^2}{y^\tau A y},$$

where $x \in \mathscr{R}^k$ and A is a $k\times k$ positive definite matrix.

Theorem 7.10

Assume the normal linear model.
Let L be an $s\times p$ matrix of rank $s \le r$.
Assume that $\mathscr{R}(L) \subset \mathscr{R}(Z)$ and $D = L(Z^\tau Z)^- L^\tau$ is of full rank.
Then

$$C_t(X) = \left[t^\tau L\widehat{\beta} - \widehat{\sigma}\sqrt{sc_\alpha t^\tau D t}, t^\tau L\widehat{\beta} + \widehat{\sigma}\sqrt{sc_\alpha t^\tau D t} \right], \quad t \in \mathscr{T},$$

are simultaneous confidence intervals for $t^\tau L\beta$, $t\in\mathscr{T}$, with confidence coefficient $1-\alpha$, where $\widehat{\sigma}^2 = \|X - Z\widehat{\beta}\|^2/(n-r)$, $\mathscr{T} = \mathscr{R}^s - \{0\}$, and c_α is the $(1-\alpha)$th quantile of the F-distribution $F_{s,n-r}$.

Proof

Note that $t^\tau L\beta \in C_t(X)$ for all $t \in \mathscr{T}$ is equivalent to

$$\frac{(t^\tau L\widehat{\beta} - t^\tau L\beta)^2}{s\widehat{\sigma}^2 t^\tau D t} \le c_\alpha \quad \text{for all } t \in \mathscr{T},$$

i.e.,

$$\max_{t\in\mathscr{T}} \frac{(t^\tau L\widehat{\beta} - t^\tau L\beta)^2}{s\widehat{\sigma}^2 t^\tau D t} \le c_\alpha.$$

From the inequality

$$x^\tau A^{-1} x = \max_{y\in\mathscr{R}^k, y\neq 0} \frac{(y^\tau x)^2}{y^\tau A y},$$

we have

$$\max_{t\in\mathscr{T}} \frac{(t^\tau L\widehat{\beta} - t^\tau L\beta)^2}{s\widehat{\sigma}^2 t^\tau D t} = \frac{(L\widehat{\beta} - L\beta)^\tau D^{-1}(L\widehat{\beta} - L\beta)}{s\widehat{\sigma}^2} = F.$$

Note that F has the F-distribution $F_{s,n-r}$.
Then

$$P\left(t^\tau L\beta \in C_t(X) \text{ for all } t \in \mathscr{T}\right) = P\left(F \le c_\alpha\right) = 1 - \alpha.$$

Remarks

- If the normality assumption is removed but conditions in Theorem 3.12 are assumed, then Scheffé's intervals in Theorem 7.10 are $1-\alpha$ asymptotically correct (exercise).
- The choice of the matrix L depends on the purpose of analysis.
  - One particular choice is $L = Z$, in which case $t^\tau L\beta$ is the mean of $t^\tau X$.
  - When Z is of full rank, we can choose $L = I_p$, in which case $\{t^\tau L\beta : t \in \mathscr{T}\}$ is the class of all linear functions of β.
  - Another L commonly used when Z is of full rank is the following $(p-1)\times p$ matrix:

$$L = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & 0 & \cdots & 0 & -1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}.$$

It can be shown (exercise) that for this L,

$$\left\{t^\tau L\beta : t \in \mathscr{R}^{p-1} - \{0\}\right\} = \left\{c^\tau\beta : c \in \mathscr{R}^p - \{0\}, c^\tau J = 0\right\},$$

where J is the p-vector of ones.

Remarks

- Functions $c^\tau \beta$ satisfying $c^\tau J = 0$ are called *contrasts*.
- Setting simultaneous confidence intervals for $t^\tau L\beta$, $t \in \mathscr{T}$, with the above L is the same as setting simultaneous confidence intervals for all nonzero contrasts.

Example 7.28 (Simple linear regression)

Consider the special case where

$$X_i = N(\beta_0 + \beta_1 z_i, \sigma^2), \qquad i = 1, ..., n,$$

and $z_i \in \mathscr{R}$ satisfying $S_z = \sum_{i=1}^n (z_i - \bar{z})^2 > 0$, $\bar{z} = n^{-1}\sum_{i=1}^n z_i$.
In this case, we are usually interested in simultaneous confidence intervals for the regression function $\beta_0 + \beta_1 z$, $z \in \mathscr{R}$.
Note that the result in Theorem 7.10 (with $L = I_2$) can be applied to obtain simultaneous confidence intervals for $\beta_0 y + \beta_1 z$, $t \in \mathscr{T} = \mathscr{R}^2 - \{0\}$, where $t = (y, z)$.
If we let $y \equiv 1$, Scheffé's intervals in Theorem 7.10 are

Remarks

- Functions $c^\tau\beta$ satisfying $c^\tau J = 0$ are called *contrasts*.
- Setting simultaneous confidence intervals for $t^\tau L\beta$, $t \in \mathscr{T}$, with the above L is the same as setting simultaneous confidence intervals for all nonzero contrasts.

Example 7.28 (Simple linear regression)

Consider the special case where

$$X_i = N(\beta_0 + \beta_1 z_i, \sigma^2), \qquad i = 1, ..., n,$$

and $z_i \in \mathscr{R}$ satisfying $S_z = \sum_{i=1}^n (z_i - \bar{z})^2 > 0$, $\bar{z} = n^{-1}\sum_{i=1}^n z_i$.
In this case, we are usually interested in simultaneous confidence intervals for the regression function $\beta_0 + \beta_1 z$, $z \in \mathscr{R}$.
Note that the result in Theorem 7.10 (with $L = I_2$) can be applied to obtain simultaneous confidence intervals for $\beta_0 y + \beta_1 z$, $t \in \mathscr{T} = \mathscr{R}^2 - \{0\}$, where $t = (y, z)$.
If we let $y \equiv 1$, Scheffé's intervals in Theorem 7.10 are

Example 7.28 (continued)

$$\left[\widehat{\beta}_0+\widehat{\beta}_1 z-\widehat{\sigma}\sqrt{2c_\alpha D(z)},\ \widehat{\beta}_0+\widehat{\beta}_1 z+\widehat{\sigma}\sqrt{2c_\alpha D(z)}\right], \quad z\in\mathscr{R}, \qquad (1)$$

where $D(z)=n^{-1}+(z-\bar{z})^2/S_z$.

Unless

$$\max_{z\in\mathscr{R}}\frac{(\widehat{\beta}_0+\widehat{\beta}_1 z-\beta_0-\beta_1 z)^2}{D(z)}=\max_{t=(y,z)\in\mathscr{T}}\frac{(\widehat{\beta}_0 y+\widehat{\beta}_1 z-\beta_0 y-\beta_1 z)^2}{t^\tau(Z^\tau Z)^{-1}t} \qquad (2)$$

holds with probability 1, where Z is the $n\times 2$ matrix whose ith row is the vector $(1,z_i)$, the confidence coefficient of the intervals in (1) is larger than $1-\alpha$.

We now show that (2) actually holds with probability 1 so that the intervals in (1) have confidence coefficient $1-\alpha$.

First,

$$P\left(n(\widehat{\beta}_0-\beta_0)+n(\widehat{\beta}_1-\beta_1)\bar{z}\neq 0\right)=1.$$

Second, it can be shown (exercise) that the maximum on the right-hand side of (2) is achieved at

Example 7.28 (continued)

$$t = \begin{pmatrix} y \\ z \end{pmatrix} = \frac{Z^\tau Z}{n(\widehat{\beta}_0 - \beta_0) + n(\widehat{\beta}_1 - \beta_1)\bar{z}} \begin{pmatrix} \widehat{\beta}_0 - \beta_0 \\ \widehat{\beta}_1 - \beta_1 \end{pmatrix}. \qquad (3)$$

Finally, (2) holds since y in (3) is equal to 1 (exercise).