Stat 709: Mathematical Statistics
Lecture 40

Jun Shao

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

Lecture 40: Functions of unbiased estimators and V-statistics

Functions of unbiased estimators

If the parameter to be estimated is $\vartheta = g(\theta)$ with a vector-valued parameter θ and U_n is a vector of unbiased estimators of components of θ, then $T_n = g(U_n)$ is often asymptotically unbiased for ϑ.
Assume that g is differentiable and

$$c_n(U_n - \theta) \to_d Y.$$

Then

$$\mathrm{amse}_{T_n}(P) = E\{[\nabla g(\theta)]^\tau Y\}^2 / c_n^2$$

(Theorem 2.6).
Hence, T_n has a good performance in terms of amse if U_n is optimal in terms of mse (such as the UMVUE or BLUE).

Example 3.22

Consider a polynomial regression of order p:

$$X_i = \beta^\tau Z_i + \varepsilon_i, \qquad i = 1, ..., n,$$

where $\beta = (\beta_0, \beta_1, ..., \beta_{p-1})$, $Z_i = (1, t_i, ..., t_i^{p-1})$, and ε_i's are i.i.d. with mean 0 and variance $\sigma^2 > 0$.

Suppose that the parameter to be estimated is $t_\beta \in \mathscr{T} \subset \mathscr{R}$ such that

$$\sum_{j=0}^{p-1} \beta_j t_\beta^j = \max_{t \in \mathscr{T}} \sum_{j=0}^{p-1} \beta_j t^j .$$

Note that $t_\beta = g(\beta)$ for some function g.

Let $\widehat{\beta}$ be the LSE of β.

Then the estimator $\widehat{t_\beta} = g(\widehat{\beta})$ is asymptotically unbiased and its amse can be derived under some conditions.

Example 3.22

Consider a polynomial regression of order p:

$$X_i = \beta^\tau Z_i + \varepsilon_i, \qquad i = 1, ..., n,$$

where $\beta = (\beta_0, \beta_1, ..., \beta_{p-1})$, $Z_i = (1, t_i, ..., t_i^{p-1})$, and ε_i's are i.i.d. with mean 0 and variance $\sigma^2 > 0$.

Suppose that the parameter to be estimated is $t_\beta \in \mathscr{T} \subset \mathscr{R}$ such that

$$\sum_{j=0}^{p-1} \beta_j t_\beta^j = \max_{t \in \mathscr{T}} \sum_{j=0}^{p-1} \beta_j t^j.$$

Note that $t_\beta = g(\beta)$ for some function g.

Let $\widehat{\beta}$ be the LSE of β.

Then the estimator $\widehat{t_\beta} = g(\widehat{\beta})$ is asymptotically unbiased and its amse can be derived under some conditions.

Example 3.23

In the study of the reliability of a system component, we assume that

$$X_{ij} = \vec{\theta}_i^{\tau} z(t_j) + \varepsilon_{ij}, \quad i = 1, ..., k, \; j = 1, ..., m.$$

Here X_{ij} is the measurement of the ith sample component at time t_j;
$z(t)$ is a q-vector whose components are known functions of the time t;
$\vec{\theta}_i$'s are unobservable random q-vectors that are i.i.d. from $N_q(\theta, \Sigma)$, where θ and Σ are unknown;
ε_{ij}'s are i.i.d. measurement errors with mean zero and variance σ^2;
$\vec{\theta}_i$'s and ε_{ij}'s are independent.

As a function of t, $\vec{\theta}^{\tau} z(t)$ is the degradation curve for a particular component and $\theta^{\tau} z(t)$ is the mean degradation curve.
Suppose that a component will fail to work if $\vec{\theta}^{\tau} z(t) < \eta$, a given critical value.
Assume that $\vec{\theta}^{\tau} z(t)$ is always a decreasing function of t.

Example 3.23

In the study of the reliability of a system component, we assume that

$$X_{ij} = \vec{\theta}_i^\tau z(t_j) + \varepsilon_{ij}, \quad i = 1, ..., k, \; j = 1, ..., m.$$

Here X_{ij} is the measurement of the ith sample component at time t_j;
$z(t)$ is a q-vector whose components are known functions of the time t;
$\vec{\theta}_i$'s are unobservable random q-vectors that are i.i.d. from $N_q(\theta, \Sigma)$, where θ and Σ are unknown;
ε_{ij}'s are i.i.d. measurement errors with mean zero and variance σ^2;
$\vec{\theta}_i$'s and ε_{ij}'s are independent.

As a function of t, $\vec{\theta}^\tau z(t)$ is the degradation curve for a particular component and $\theta^\tau z(t)$ is the mean degradation curve.
Suppose that a component will fail to work if $\vec{\theta}^\tau z(t) < \eta$, a given critical value.
Assume that $\vec{\theta}^\tau z(t)$ is always a decreasing function of t.

Example 3.23 (continued)

Then the reliability function of a component is

$$R(t) = P(\vec{\theta}^\tau z(t) > \eta) = \Phi\left(\frac{\theta^\tau z(t) - \eta}{s(t)}\right),$$

where $s(t) = \sqrt{[z(t)]^\tau \Sigma z(t)}$ and Φ is the standard normal distribution function.

For a fixed t, estimators of $R(t)$ can be obtained by estimating θ and Σ, since Φ is a known function.

It can be shown (exercise) that the BLUE of θ is the LSE

$$\widehat{\theta} = (Z^\tau Z)^{-1} Z^\tau \bar{X},$$

where Z is the $m \times q$ matrix whose jth row is the vector $z(t_j)$, $X_i = (X_{i1}, ..., X_{im})$, and $\bar{X}$ is the sample mean of X_i's.

Example 3.23 (continued)

Then the reliability function of a component is

$$R(t) = P(\vec{\theta}^\tau z(t) > \eta) = \Phi\left(\frac{\theta^\tau z(t) - \eta}{s(t)}\right),$$

where $s(t) = \sqrt{[z(t)]^\tau \Sigma z(t)}$ and Φ is the standard normal distribution function.

For a fixed t, estimators of $R(t)$ can be obtained by estimating θ and Σ, since Φ is a known function.

It can be shown (exercise) that the BLUE of θ is the LSE

$$\widehat{\theta} = (Z^\tau Z)^{-1} Z^\tau \bar{X},$$

where Z is the $m \times q$ matrix whose jth row is the vector $z(t_j)$, $X_i = (X_{i1}, ..., X_{im})$, and $\bar{X}$ is the sample mean of X_i's.

Example 3.23 (continued)

The estimation of Σ is more difficult.

It can be shown (exercise) that a consistent (as $k \to \infty$) estimator of Σ is

$$\widehat{\Sigma} = \frac{1}{k}\sum_{i=1}^{k}(Z^\tau Z)^{-1}Z^\tau(X_i - \bar{X})(X_i - \bar{X})^\tau Z(Z^\tau Z)^{-1} - \widehat{\sigma}^2(Z^\tau Z)^{-1},$$

where

$$\widehat{\sigma}^2 = \frac{1}{k(m-q)}\sum_{i=1}^{k}[X_i^\tau X_i - X_i^\tau Z(Z^\tau Z)^{-1}Z^\tau X_i].$$

Hence an estimator of $R(t)$ is

$$\widehat{R}(t) = \Phi\left(\frac{\widehat{\theta}^\tau z(t) - \eta}{\widehat{s}(t)}\right),$$

where $\widehat{s}(t) = \sqrt{[z(t)]^\tau \widehat{\Sigma} z(t)}$.

Example 3.23 (continued)

It is apparent that $\widehat{R}(t)$ can be written as $g(\bar{Y})$ for a function

$$g(y_1, y_2, y_3) = \Phi\left(\frac{y_1 - \eta}{\sqrt{y_2 - y_1^2 - y_3[z(t)]^\tau (Z^\tau Z)^{-1} z(t)}}\right).$$

$Y_{i1} = X_i^\tau Z(Z^\tau Z)^{-1} z(t)$
$Y_{i2} = [X_i^\tau Z(Z^\tau Z)^{-1} z(t)]^2$
$Y_{i3} = [X_i^\tau X_i - X_i^\tau Z(Z^\tau Z)^{-1} Z^\tau X_i]/(m-q)$
$Y_i = (Y_{i1}, Y_{i2}, Y_{i3})$
Suppose that ε_{ij} has a finite fourth moment, which implies the existence of $\mathrm{Var}(Y_i)$.
The amse of $\widehat{R}(t)$ can be derived (exercise).

V-statistics

Let $X_1,...,X_n$ be i.i.d. from P.

For every U-statistic U_n as an estimator of $\vartheta = E[h(X_1,...,X_m)]$, there is a closely related *V-statistic* defined by

$$V_n = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(X_{i_1},...,X_{i_m}). \quad (1)$$

As an estimator of ϑ, V_n is biased; but the bias is small asymptotically as the following results show.

For a fixed sample size n, V_n may be better than U_n in terms of their mse's.

Proposition 3.5

Let V_n be defined by (1).

(i) Assume that $E|h(X_{i_1},...,X_{i_m})| < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$. Then the bias of V_n satisfies

$$b_{V_n}(P) = O(n^{-1}).$$

(ii) Assume that $E[h(X_{i_1},...,X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$. Then the variance of V_n satisfies

$$\mathrm{Var}(V_n) = \mathrm{Var}(U_n) + O(n^{-2}),$$

where U_n is the U-statistic corresponding to V_n.

To study the asymptotic behavior of a V-statistic, we consider the following representation of V_n in (1):

$$V_n = \sum_{j=1}^{m} \binom{m}{j} V_{nj},$$

where

$$V_{nj} = \vartheta + \frac{1}{n^j} \sum_{i_1=1}^{n} \cdots \sum_{i_j=1}^{n} g_j(X_{i_1}, ..., X_{i_j})$$

is a “V-statistic" with

$$\begin{aligned} g_j(x_1,...,x_j) = h_j(x_1,...,x_j) &- \sum_{i=1}^{j} \int h_j(x_1,...,x_j) dP(x_i) \\ &+ \sum_{1 \le i_1 < i_2 \le j} \int \int h_j(x_1,...,x_j) dP(x_{i_1}) dP(x_{i_2}) - \cdots \\ &+ (-1)^j \int \cdots \int h_j(x_1,...,x_j) dP(x_1) \cdots dP(x_j) \end{aligned}$$

and $h_j(x_1,...,x_j) = E[h(x_1,...,x_j,X_{j+1},...,X_m)]$.

Using an argument similar to the proof of Theorem 3.4, we can show that

$$EV_{nj}^2 = O(n^{-j}), \qquad j = 1, ..., m,$$

provided that $E[h(X_{i_1}, ..., X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$.
Thus,

$$V_n - \vartheta = mV_{n1} + \frac{m(m-1)}{2} V_{n2} + o_p(n^{-1}),$$

which leads to the following result similar to Theorem 3.5.

Theorem 3.16

Let V_n be given by (1) with $E[h(X_{i_1}, ..., X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$.

(i) If $\zeta_1 = \mathrm{Var}(h_1(X_1)) > 0$, then

$$\sqrt{n}(V_n - \vartheta) \to_d N(0, m^2 \zeta_1).$$

Using an argument similar to the proof of Theorem 3.4, we can show that

$$EV_{nj}^2 = O(n^{-j}), \qquad j = 1, ..., m,$$

provided that $E[h(X_{i_1}, ..., X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$. Thus,

$$V_n - \vartheta = mV_{n1} + \frac{m(m-1)}{2} V_{n2} + o_p(n^{-1}),$$

which leads to the following result similar to Theorem 3.5.

Theorem 3.16

Let V_n be given by (1) with $E[h(X_{i_1}, ..., X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$.

(i) If $\zeta_1 = \mathrm{Var}(h_1(X_1)) > 0$, then

$$\sqrt{n}(V_n - \vartheta) \to_d N(0, m^2\zeta_1).$$

Theorem 3.16 (continued)

(ii) If $\zeta_1 = 0$ but $\zeta_2 = \mathrm{Var}(h_2(X_1, X_2)) > 0$, then

$$n(V_n - \vartheta) \to_d \frac{m(m-1)}{2} \sum_{j=1}^{\infty} \lambda_j \chi^2_{1j},$$

where χ^2_{1j}'s and λ_j's are the same as those in Theorem 3.5.

Theorem 3.16 shows that if $\zeta_1 > 0$, then the amse's of U_n and V_n are the same.

If $\zeta_1 = 0$ but $\zeta_2 > 0$, then an argument similar to that in the proof of Lemma 3.2 leads to

$$\mathrm{amse}_{V_n}(P) = \frac{m^2(m-1)^2\zeta_2}{2n^2} + \frac{m^2(m-1)^2}{4n^2}\left(\sum_{j=1}^{\infty}\lambda_j\right)^2$$

$$= \mathrm{amse}_{U_n}(P) + \frac{m^2(m-1)^2}{4n^2}\left(\sum_{j=1}^{\infty}\lambda_j\right)^2$$

(see Lemma 3.2).

Hence U_n is asymptotically more efficient than V_n, unless $\sum_{j=1}^{\infty}\lambda_j = 0$.

Theorem 3.16 (continued)

(ii) If $\zeta_1 = 0$ but $\zeta_2 = \mathrm{Var}(h_2(X_1, X_2)) > 0$, then

$$n(V_n - \vartheta) \to_d \frac{m(m-1)}{2} \sum_{j=1}^{\infty} \lambda_j \chi^2_{1j},$$

where χ^2_{1j}'s and λ_j's are the same as those in Theorem 3.5.

Theorem 3.16 shows that if $\zeta_1 > 0$, then the amse's of U_n and V_n are the same.

If $\zeta_1 = 0$ but $\zeta_2 > 0$, then an argument similar to that in the proof of Lemma 3.2 leads to

$$\begin{aligned} \mathrm{amse}_{V_n}(P) &= \frac{m^2(m-1)^2\zeta_2}{2n^2} + \frac{m^2(m-1)^2}{4n^2}\left(\sum_{j=1}^{\infty}\lambda_j\right)^2 \\ &= \mathrm{amse}_{U_n}(P) + \frac{m^2(m-1)^2}{4n^2}\left(\sum_{j=1}^{\infty}\lambda_j\right)^2 \end{aligned}$$

(see Lemma 3.2).

Hence U_n is asymptotically more efficient than V_n, unless $\sum_{j=1}^{\infty}\lambda_j = 0$.