Lecture 40: V-statistics and the weighted LSE

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

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

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

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

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

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

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,

$$V_n - \vartheta = mV_{n1} + \tfrac{m(m-1)}{2}V_{n2} + o_p(n^{-1}), \qquad (3)$$

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

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

(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 $\sum_{j=1}^{\infty}\lambda_j = 0$.

Example 3.28. Consider the estimation of μ^2, where $\mu = EX_1$.
From the results in §3.2, the U-statistic $U_n = \frac{1}{n(n-1)}\sum_{1\le i<j\le n} X_iX_j$ is unbiased for μ^2.
The corresponding V-statistic is simply $V_n = \bar{X}^2$.
If $\mu \ne 0$, then $\zeta_1 \ne 0$ and the asymptotic relative efficiency of V_n w.r.t. U_n is 1.
If $\mu = 0$, then

$$nV_n \to_d \sigma^2\chi^2_1 \qquad \text{and} \qquad nU_n \to_d \sigma^2(\chi^2_1 - 1),$$

where χ^2_1 is a random variable having the chi-square distribution χ^2_1.
Hence the asymptotic relative efficiency of V_n w.r.t. U_n is

$$E(\chi^2_1 - 1)^2/E(\chi^2_1)^2 = 2/3.$$

The weighted LSE

In the linear model

$$X = Z\beta + \varepsilon, \tag{4}$$

the unbiased LSE of $l^\tau\beta$ may be improved by a slightly biased estimator when $V = \mathrm{Var}(\varepsilon)$ is not $\sigma^2 I_n$ and the LSE is not BLUE.

Assume that Z is of full rank so that every $l^\tau\beta$ is estimable.
If V is known, then the BLUE of $l^\tau\beta$ is $l^\tau\breve{\beta}$, where

$$\breve{\beta} = (Z^\tau V^{-1} Z)^{-1} Z^\tau V^{-1} X \tag{5}$$

(see the discussion after the statement of assumption A3 in §3.3.1).
If V is unknown and $\hat{V}$ is an estimator of V, then an application of the substitution principle leads to a *weighted least squares estimator*

$$\hat{\beta}_w = (Z^\tau \hat{V}^{-1} Z)^{-1} Z^\tau \hat{V}^{-1} X. \tag{6}$$

The weighted LSE is not linear in X and not necessarily unbiased for β.
If the distribution of ε is symmetric about 0 and $\hat{V}$ remains unchanged when ε changes to $-\varepsilon$, then the distribution of $\hat{\beta}_w - \beta$ is symmetric about 0 and, if $E\hat{\beta}_w$ is well defined, $\hat{\beta}_w$ is unbiased for β.
In such a case the LSE $l^\tau\hat{\beta}$ may not be a UMVUE (when ε is normal), since $\mathrm{Var}(l^\tau\hat{\beta}_w)$ may be smaller than $\mathrm{Var}(l^\tau\hat{\beta})$.

Asymptotic properties of the weighted LSE depend on the asymptotic behavior of $\hat{V}$.
We say that $\hat{V}$ is consistent for V if and only if

$$\|\hat{V}^{-1}V - I_n\|_{\max} \to_p 0, \tag{7}$$

where $\|A\|_{\max} = \max_{i,j} |a_{ij}|$ for a matrix A whose (i,j)th element is a_{ij}.

Theorem 3.17. Consider model (4) with a full rank Z. Let $\breve{\beta}$ and $\hat{\beta}_w$ be defined by (5) and (6), respectively, with a $\hat{V}$ consistent in the sense of (7). Assume the conditions in Theorem 3.12. Then

$$l^\tau(\hat{\beta}_w - \beta)/a_n \to_d N(0,1),$$

where $l \in \mathcal{R}^p$, $l \neq 0$, and

$$a_n^2 = \mathrm{Var}(l^\tau\breve{\beta}) = l^\tau(Z^\tau V^{-1} Z)^{-1} l.$$

Proof. Using the same argument as in the proof of Theorem 3.12, we obtain that

$$l^\tau(\breve{\beta} - \beta)/a_n \to_d N(0,1).$$

By Slutsky's theorem, the result follows from

$$l^\tau\hat{\beta}_w - l^\tau\breve{\beta} = o_p(a_n).$$

Define

$$\xi_n = l^\tau (Z^\tau \hat{V}^{-1} Z)^{-1} Z^\tau (\hat{V}^{-1} - V^{-1})\varepsilon$$

and

$$\zeta_n = l^\tau [(Z^\tau \hat{V}^{-1} Z)^{-1} - (Z^\tau V^{-1} Z)^{-1}] Z^\tau V^{-1} \varepsilon.$$

Then

$$l^\tau \hat{\beta}_w - l^\tau \breve{\beta} = \xi_n + \zeta_n.$$

The result follows from $\xi_n = o_p(a_n)$ and $\zeta_n = o_p(a_n)$ (details are in the textbook).

Theorem 3.17 shows that as long as $\hat{V}$ is consistent in the sense of (7), the weighted LSE $\hat{\beta}_w$ is asymptotically as efficient as $\breve{\beta}$, which is the BLUE if V is known.
By Theorems 3.12 and 3.17, the asymptotic relative efficiency of the LSE $l^\tau \hat{\beta}$ w.r.t. the weighted LSE $l^\tau \hat{\beta}_w$ is

$$\frac{l^\tau (Z^\tau V^{-1} Z)^{-1} l}{l^\tau (Z^\tau Z)^{-1} Z^\tau V Z (Z^\tau Z)^{-1} l},$$

which is always less than 1 and equals 1 if $l^\tau \hat{\beta}$ is a BLUE (in which case $\hat{\beta} = \breve{\beta}$).

Finding a consistent $\hat{V}$ is possible when V has a certain type of structure.

Example 3.29. Consider model (4). Suppose that $V = \mathrm{Var}(\varepsilon)$ is a block diagonal matrix with the ith diagonal block

$$\sigma^2 I_{m_i} + U_i \Sigma U_i^\tau, \qquad i = 1, ..., k, \tag{8}$$

where m_i's are integers bounded by a fixed integer m, $\sigma^2 > 0$ is an unknown parameter, Σ is a $q \times q$ unknown nonnegative definite matrix, U_i is an $m_i \times q$ full rank matrix whose columns are in $\mathcal{R}(W_i)$, $q < \inf_i m_i$, and W_i is the $p \times m_i$ matrix such that $Z^\tau = (\ W_1\ W_2\ ...\ W_k\ )$.
Under (8), a consistent $\hat{V}$ can be obtained if we can obtain consistent estimators of σ^2 and Σ.
Let $X = (Y_1, ..., Y_k)$, where Y_i is an m_i-vector, and let R_i be the matrix whose columns are linearly independent rows of W_i. Then

$$\hat{\sigma}^2 = \frac{1}{n - kq} \sum_{i=1}^{k} Y_i^\tau [I_{m_i} - R_i (R_i^\tau R_i)^{-1} R_i^\tau] Y_i \tag{9}$$

is an unbiased estimator of σ^2.
Assume that Y_i's are independent and that $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$.
Then $\hat{\sigma}^2$ is consistent for σ^2 (exercise). Let $r_i = Y_i - W_i^\tau \hat{\beta}$ and

$$\hat{\Sigma} = \frac{1}{k} \sum_{i=1}^{k} \left[(U_i^\tau U_i)^{-1} U_i^\tau r_i r_i^\tau U_i (U_i^\tau U_i)^{-1} - \hat{\sigma}^2 (U_i^\tau U_i)^{-1} \right]. \tag{10}$$

It can be shown (exercise) that $\hat{\Sigma}$ is consistent for Σ in the sense that $\|\hat{\Sigma} - \Sigma\|_{\max} \to_p 0$ or, equivalently, $\|\hat{\Sigma} - \Sigma\| \to_p 0$ (see Exercise 116).