Lecture 28: V-statistics, functions of unbiased estimators, and method of moments

V-statistics
Let $X_1, \ldots, X_n$ be i.i.d. from $P$.
For every U-statistic $U_n$ as an estimator of $\vartheta = E[h(X_1, \ldots, X_m)]$, there is a closely related V-statistic defined by

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

As an estimator of $\vartheta$, $V_n$ is biased; but the bias is small asymptotically. For a fixed $n$, $V_n$ may be better than $U_n$ in terms of the mse.

Example
Corresponding to $U_n = \frac{2}{n(n-1)} \sum_{i<j} X_i X_j$, the V-statistic is

$$V_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j = \bar{X}^2$$

a function of the sample mean $\bar{X}$. 
Proposition 3.5

Let $V_n$ be defined by (1).

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

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

(ii) Assume that $E[h(X_{i_1}, \ldots, X_{i_m})]^2 < \infty$ for all $1 \leq i_1 \leq \cdots \leq i_m \leq m$. 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 $V_n$.

Theorem 3.16

Let $V_n$ be a V-statistic with $E[h(X_{i_1}, \ldots, X_{i_m})]^2 < \infty$ for all $1 \leq i_1 \leq \cdots \leq i_m \leq m$.

(i) If $\zeta_1 = \text{Var}(h_1(X_1)) > 0$, then $\sqrt{n}(V_n - \var) \rightarrow_d N(0, m^2 \zeta_1)$. 


Theorem 3.16 (continued)

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

$$n(V_n - \vartheta) \overset{d}{\to} \frac{m(m-1)}{2} \sum_{j=1}^{\infty} \lambda_j \chi_{1j}^2,$$

where $\chi_{1j}^2$'s and $\lambda_j$'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

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

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

(see Lemma 3.2).

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

An exactly unbiased estimator may not exist, or is hard to obtain. We often derive asymptotically unbiased estimators. Functions of sample means are popular estimators.

Functions of unbiased estimators

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

Note that \( E(T_n) = Eh(U_n) \) may not exist.

Assume that \( g \) is differentiable and

\[
c_n(U_n - \theta) \rightarrow_d Y.
\]

Then, by Theorem 2.6,

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

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).
Method of moments

The method of moments is the oldest method of deriving asymptotically unbiased estimators, which may not be the best estimators, but they are simple and can be used as initial estimators.

Consider a parametric problem where $X_1, \ldots, X_n$ are i.i.d. random variables from $P_{\theta}$, $\theta \in \Theta \subset \mathbb{R}^k$, and $E|X_1|^k < \infty$.

Let $\mu_j = EX_j^i$ be the $j$th moment of $P$ and let

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$$

be the $j$th sample moment, which is an unbiased estimator of $\mu_j$, $j = 1, \ldots, k$.

Typically,

$$\mu_j = h_j(\theta), \quad j = 1, \ldots, k,$$

for some functions $h_j$ on $\mathbb{R}^k$. 
Method of moments

By substituting $\mu_j$’s on the left-hand side of (2) by the sample moments $\hat{\mu}_j$, we obtain a moment estimator $\hat{\theta}$, i.e., $\hat{\theta}$ satisfies

$$\hat{\mu}_j = h_j(\hat{\theta}), \quad j = 1, \ldots, k,$$

which is a sample analogue of (2). This method of deriving estimators is called the method of moments. An important statistical principle, the substitution principle, is applied in this method.

Let $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_k)$ and $h = (h_1, \ldots, h_k)$. Then $\hat{\mu} = h(\hat{\theta})$.

If the inverse function $h^{-1}$ exists, then the unique moment estimator of $\theta$ is $\hat{\theta} = h^{-1}(\hat{\mu})$.

When $h^{-1}$ does not exist (i.e., $h$ is not one-to-one), any solution of $\hat{\mu} = h(\hat{\theta})$ is a moment estimator of $\theta$.

If possible, we always choose a solution $\hat{\theta}$ in the parameter space $\Theta$. 
Method of moments

In some cases, however, a moment estimator does not exist (see Exercise 111).
Moment estimators may not be unique.
We usually use moments with the lowest possible order.

Assume that $\hat{\theta} = g(\hat{\mu})$ for a function $g$.
If $h^{-1}$ exists, then $g = h^{-1}$.
If $g$ is continuous at $\mu = (\mu_1, \ldots, \mu_k)$, then $\hat{\theta}$ is strongly consistent for $\theta$, since $\hat{\mu}_j \to a.s. \mu_j$ by the SLLN.
If $g$ is differentiable at $\mu$ and $E|X_1|^{2k} < \infty$, then $\hat{\theta}$ is asymptotically normal, by the CLT and Theorem 1.12, and

$$\text{amse}_{\hat{\theta}}(\theta) = n^{-1}[\nabla g(\mu)]^\tau V_{\mu} \nabla g(\mu),$$

where $V_{\mu}$ is a $k \times k$ matrix whose $(i,j)$th element is $\mu_{i+j} - \mu_i \mu_j$.
Furthermore, the $n^{-1}$ order asymptotic bias of $\hat{\theta}$ is

$$(2n)^{-1}\text{tr}\left(\nabla^2 g(\mu) V_{\mu}\right).$$
Example 3.24

Let $X_1, ..., X_n$ be i.i.d. from a population $P_\theta$ indexed by the parameter $\theta = (\mu, \sigma^2)$, where $\mu = EX_1 \in \mathbb{R}$ and $\sigma^2 = \text{Var}(X_1) \in (0, \infty)$. This includes cases such as the family of normal distributions, double exponential distributions, or logistic distributions (Table 1.2, page 20).

Since $EX_1 = \mu$ and $EX_1^2 = \text{Var}(X_1) + (EX_1)^2 = \sigma^2 + \mu^2$, setting $\hat{\mu}_1 = \mu$ and $\hat{\mu}_2 = \sigma^2 + \mu^2$ we obtain the moment estimator

$$\hat{\theta} = \left( \overline{X}, \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right) = \left( \overline{X}, \frac{n-1}{n} S^2 \right).$$

Note that $\overline{X}$ is unbiased, but $\frac{n-1}{n} S^2$ is not.

If $X_i$ is normal, then $\hat{\theta}$ is sufficient and is nearly the same as an optimal estimator such as the UMVUE. On the other hand, if $X_i$ is from a double exponential or logistic distribution, then $\hat{\theta}$ is not sufficient and can often be improved.
Example 3.25

Let $X_1, \ldots, X_n$ be i.i.d. from the uniform distribution on $(\theta_1, \theta_2)$, $-\infty < \theta_1 < \theta_2 < \infty$.

Note that

$$EX_1 = (\theta_1 + \theta_2)/2 \quad \text{and} \quad EX_1^2 = (\theta_1^2 + \theta_2^2 + \theta_1 \theta_2)/3.$$ 

Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$ and substituting $\theta_1$ in the second equation by $2\hat{\mu}_1 - \theta_2$ (the first equation), we obtain that

$$(2\hat{\mu}_1 - \theta_2)^2 + \theta_2^2 + (2\hat{\mu}_1 - \theta_2)\theta_2 = 3\hat{\mu}_2,$$

which is the same as

$$(\theta_2 - \hat{\mu}_1)^2 = 3(\hat{\mu}_2 - \hat{\mu}_1^2).$$

Since $\theta_2 > EX_1$, we obtain that

$$\hat{\theta}_2 = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} + \sqrt{\frac{3(n-1)}{n}} S^2$$

$$\hat{\theta}_1 = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} - \sqrt{\frac{3(n-1)}{n}} S^2.$$ 

These estimators are not functions of the sufficient and complete statistic $(X_{(1)}, X_{(n)})$. 
Example 3.26

Let \( X_1, \ldots, X_n \) be i.i.d. from the binomial distribution \( Bi(p, k) \) with unknown parameters \( k \in \{1, 2, \ldots \} \) and \( p \in (0, 1) \).

Since

\[ EX_1 = kp \]

and

\[ EX_1^2 = kp(1 - p) + k^2 p^2, \]

we obtain the moment estimators

\[ \hat{p} = \frac{\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2}{\hat{\mu}_1} = 1 - \frac{n-1}{n} S^2 / \bar{X} \]

and

\[ \hat{k} = \frac{\hat{\mu}_1^2}{(\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2)} = \frac{\bar{X}}{(1 - \frac{n-1}{n} S^2 / \bar{X})}. \]

The estimator \( \hat{p} \) is in the range of \((0, 1)\).

But \( \hat{k} \) may not be an integer.

It can be improved by an estimator that is \( \hat{k} \) rounded to the nearest positive integer.
Nonparametric problems

Consider the estimation of the central moments

\[ c_j = E(X_1 - \mu_1)^j = \sum_{t=0}^{j} \binom{j}{t} (-\mu_1)^t \mu_{j-t}, \quad j = 2, ..., k. \]

the moment estimator of \( c_j \) is

\[ \hat{c}_j = \sum_{t=0}^{j} \binom{j}{t} (-\bar{X})^t \hat{\mu}_{j-t} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^j, \quad j = 2, ..., k, \]

which are sample central moments, (\( \hat{\mu}_0 = 1 \)).

From the SLLN, \( \hat{c}_j \)'s are strongly consistent.

If \( E|X_1|^{2k} < \infty \), then

\[ \sqrt{n} (\hat{c}_2 - c_2, ..., \hat{c}_k - c_k) \rightarrow_d N_{k-1}(0, D) \]

where the \((i,j)\)th element of the \((k-1) \times (k-1)\) matrix \( D \) is

\[ c_{i+j+2} - c_{i+1}c_{j+1} - (i+1)c_{i}c_{j+2} - (j+1)c_{i+2}c_{j} + (i+1)(j+1)c_{i}c_{j}c_{2}. \]