Lecture 19: Construction of unbiased or approximately unbiased estimators and method of moments

Survey samples from a finite population

Let $\mathscr{P} = \{1, ..., N\}$ be a finite population of interest

For each $i \in \mathscr{P}$, let y_i be a value of interest associated with unit i

Let $\mathbf{s} = \{i_1, ..., i_n\}$ be a subset of distinct elements of $\mathscr{P}$, which is a sample selected with selection probability $p(\mathbf{s})$, where p is known.

The value y_i is observed if and only if $i \in \mathbf{s}$.

$Y = \sum_{j=1}^{N} y_j$ is the unknown population total of interest.

Define

$$\pi_i = \text{probability that } i \in \mathbf{s}, \quad i = 1, ..., N.$$

Horvitz-Thompson estimators

- It gives a general method to obtain unbiased estimators.
- All we need is the inclusion probability π_i, which is known in sample surveys since $p(\mathbf{s})$ is known.

Theorem 3.15.

(i) (Horvitz-Thompson). If $\pi_i > 0$ for $i = 1, ..., N$ and π_i is known when $i \in \mathbf{s}$, then $\widehat{Y}_{ht} = \sum_{i \in \mathbf{s}} y_i/\pi_i$ is an unbiased estimator of the population total Y.

(ii) Define

$$\pi_{ij} = \text{probability that } i \in \mathbf{s} \text{ and } j \in \mathbf{s}, \quad i = 1, ..., N, j = 1, ..., N.$$

Then

$$\mathrm{Var}(\widehat{Y}_{ht}) = \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} y_i y_j \qquad (1)$$

$$= \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2. \qquad (2)$$

- Horvitz-Thompson's idea: inverse probability weighting
- The unbiasedness of the sample mean under simple random sampling without replacement is a special case of Theorem 3.15.
- Extension: $\mathscr{P}$ is a sample of size N and y_i is missing if $i \notin \mathbf{s}$
- If π_i is unknown, we need to replace it by an estimator.

Proof.

(i) Let $a_i = 1$ if $i \in \mathbf{s}$ and $a_i = 0$ if $i \notin \mathbf{s}$, $i = 1, ..., N$.
Then $E(a_i) = \pi_i$ and

$$E(\widehat{Y}_{ht}) = E\left(\sum_{i=1}^{N} \frac{a_i y_i}{\pi_i}\right) = \sum_{i=1}^{N} y_i = Y.$$

(ii) Since $a_i^2 = a_i$,

$$\mathrm{Var}(a_i) = E(a_i) - [E(a_i)]^2 = \pi_i(1 - \pi_i).$$

$$\mathrm{Cov}(a_i, a_j) = E(a_i a_j) - E(a_i)E(a_j) = \pi_{ij} - \pi_i \pi_j, \quad i \neq j.$$

Then

$$\begin{aligned}
\mathrm{Var}(\widehat{Y}_{ht}) &= \mathrm{Var}\left(\sum_{i=1}^{N} \frac{a_i y_i}{\pi_i}\right) \\
&= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2} \mathrm{Var}(a_i) + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{y_i y_j}{\pi_i \pi_j} \mathrm{Cov}(a_i, a_j) \\
&= \sum_{i=1}^{N} \frac{1 - \pi_i}{\pi_i} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} y_i y_j.
\end{aligned}$$

Proof (continued)

Hence (1) follows.
To show (2), note that

$$\sum_{i=1}^{N} \pi_i = n \qquad \text{and} \qquad \sum_{j=1,\dots,N, j\neq i} \pi_{ij} = (n-1)\pi_i,$$

which implies

$$\sum_{j=1,\dots,N, j\neq i} (\pi_{ij} - \pi_i\pi_j) = (n-1)\pi_i - \pi_i(n-\pi_i) = -\pi_i(1-\pi_i).$$

Hence

$$\begin{aligned} \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} y_i^2 &= \sum_{i=1}^{N} \sum_{j=1,\dots,N, j\neq i} (\pi_i\pi_j - \pi_{ij}) \frac{y_i^2}{\pi_i^2} \\ &= \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i\pi_j - \pi_{ij}) \left(\frac{y_i^2}{\pi_i^2} + \frac{y_j^2}{\pi_j^2} \right) \end{aligned}$$

and, (2) follows from (1).

How do we get an unbiased estimator of $\mathrm{Var}(\widehat{Y}_{ht})$?

Using Horvitz-Thompson's idea, the following estimators are unbiased:

$$v_1 = \sum_{i\in\mathbf{s}} \frac{1-\pi_i}{\pi_i^2} y_i^2 + 2\sum_{i\in\mathbf{s}} \sum_{j\in\mathbf{s},j>i} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j\pi_{ij}} y_i y_j$$

$$v_2 = \sum_{i\in\mathbf{s}} \sum_{j\in\mathbf{s},j>i} \frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2.$$

Simple random sampling without replacement

For simple random sampling without replacement,

$$\pi_i = E(a_i) = P(a_i=1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

$$\pi_{ij} = E(a_i a_j) = P(a_i=1, a_j=1) = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

$$\widehat{Y}_{ht} = \frac{N}{n}\sum_{i\in\mathbf{s}} y_i = \frac{N}{n}\sum_{i=1}^{N} a_i y_i = N(\text{the sample mean})$$

$$
\begin{aligned}
\mathrm{Var}(\widehat{Y}_{ht}) &= \sum_{i=1}^{N} \frac{1-\frac{n}{N}}{\frac{n}{N}} y_i^2 + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \frac{\frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2}}{\frac{n^2}{N^2}} y_i y_j \\
&= \frac{N-n}{n}\sum_{i=1}^{N} y_i^2 - \frac{2(N-n)}{n(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N} y_i y_j \\
&= \frac{N-n}{n}\left[\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}\sum_{i\neq j} y_i y_j\right] \\
&= \frac{N-n}{n}\frac{N}{N-1}\sum_{i=1}^{N}\left(y_i - \frac{Y}{N}\right)^2 \\
&= \left(1-\frac{n}{N}\right)\frac{N^2}{n}\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \frac{Y}{N}\right)^2 \\
&= N^2\left(1-\frac{n}{N}\right)\frac{S^2}{n}
\end{aligned}
$$

$\frac{n}{N}$ is called the finite sample fraction and $1-\frac{n}{N}$ is called the finite sample correction.
S^2 is called the population variance.

For simple random sampling without replacement, variance estimators v_1 and v_2 are the same.

Note that

$$
\begin{aligned}
v_1 &= \sum_{i\in\mathbf{s}} \frac{1}{\frac{n}{N}} \frac{\frac{n}{N} - \frac{n^2}{N^2}}{\frac{n^2}{N^2}} y_i^2 + 2\sum_{i\in\mathbf{s}} \sum_{j\in\mathbf{s}, j>i} \frac{1}{\frac{n(n-1)}{N(N-1)}} \frac{\frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2}}{\frac{n^2}{N^2}} y_i y_j \\
&= \frac{N(N-n)}{n^2} \sum_{i\in\mathbf{s}} y_i^2 - \frac{N(N-n)}{n^2(n-1)} \sum_{i,j\in\mathbf{s}, i\neq j} y_i y_j \\
&= \frac{N(N-n)}{n(n-1)} \sum_{i\in\mathbf{s}} y_i^2 - \frac{N(N-n)}{n^2(n-1)} \sum_{i,j\in\mathbf{s}} y_i y_j \\
&= \frac{N(N-n)}{n(n-1)} \left[\sum_{i\in\mathbf{s}} y_i^2 - \frac{1}{n} \left(\sum_{i\in\mathbf{s}} y_i \right)^2 \right] \\
&= N^2 \left(1 - \frac{n}{N}\right) \frac{s^2}{n}
\end{aligned}
$$

where

$$s^2 = \frac{1}{n-1} \sum_{i \in \mathbf{s}} (y_i - \bar{y})^2$$

is called the sample variance.

Since $E(v_1) = \text{Var}(\widehat{Y}_{ht})$, we have shown that $E(s^2) = S^2$.

Since s^2 is symmetric in its arguments, the early result implies that s^2 is a UMVUE of S^2 and v_1 is a UMVUE of $\text{Var}(\widehat{Y}_{ht})$, under simple random sampling without replacement.

To finish, we note that

$$\begin{aligned} v_2 &= v_1 + \sum_{i,j \in \mathbf{s}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{y_i^2}{\pi_i^2} \\ &= v_1 + \sum_{i,j \in \mathbf{s}} \left(\frac{1}{\pi_i^2} - \frac{1}{\pi_{ij}} \right) y_i^2 \\ &= v_1 + \frac{N^2}{n} \sum_{i \in \mathbf{s}} y_i^2 - \frac{N}{n} \sum_{i \in \mathbf{s}} y_i^2 - \frac{N(N-1)}{n(n-1)} \sum_{i,j \in \mathbf{s}, j \neq i} y_i^2 \\ &= v_1 \end{aligned}$$

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 $\vartheta = g(\theta)$ 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.

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

Assume that g is differentiable and

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

Then, by Theorem 2.6,

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

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 $X_1,...,X_n$ are i.i.d. random variables from P_θ, $\theta \in \Theta \subset \mathscr{R}^k$, and $E|X_1|^k < \infty$.

Let $\mu_j = EX_1^j$ be the jth moment of P and let

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

be the jth *sample moment*, which is an unbiased estimator of μ_j, $j = 1,...,k$.

Typically,

$$\mu_j = h_j(\theta), \qquad j = 1,...,k, \tag{3}$$

for some functions h_j on $\mathscr{R}^k$.

By substituting μ_j's on the left-hand side of (3) by the sample moments $\widehat{\mu}_j$, we obtain a *moment estimator* $\widehat{\theta}$, i.e., $\widehat{\theta}$ satisfies

$$\widehat{\mu}_j = h_j(\widehat{\theta}), \qquad j = 1, ..., k,$$

which is a sample analogue of (3).

This method of deriving estimators is called the *method of moments*.

An important statistical principle, the *substitution principle*, is applied in this method.

Let $\widehat{\mu} = (\widehat{\mu}_1, ..., \widehat{\mu}_k)$ and $h = (h_1, ..., h_k)$.

Then $\widehat{\mu} = h(\widehat{\theta})$.

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

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

If possible, we always choose a solution $\widehat{\theta}$ in the parameter space Θ.

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 $\widehat{\theta} = g(\widehat{\mu})$ for a function g.

If h^{-1} exists, then $g = h^{-1}$.

If g is continuous at $\mu = (\mu_1, ..., \mu_k)$, then $\widehat{\theta}$ is strongly consistent for θ, since $\widehat{\mu}_j \to_{a.s.} \mu_j$ by the SLLN.

If g is differentiable at μ and $E|X_1|^{2k} < \infty$, then $\widehat{\theta}$ is asymptotically normal, by the CLT and Theorem 1.12, and

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

where V_μ 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 $\widehat{\theta}$ is

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

Example 3.24

Let $X_1, ..., X_n$ be i.i.d. from a population P_θ indexed by the parameter $\theta = (\mu, \sigma^2)$, where $\mu = EX_1 \in \mathscr{R}$ and $\sigma^2 = \mathrm{Var}(X_1) \in (0, \infty)$.

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

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

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

If X_i is normal, then $\widehat{\theta}$ 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 $\widehat{\theta}$ is not sufficient and can often be improved.

Example 3.25

Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on (θ_1, θ_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 $\widehat{\mu}_1 = EX_1$ and $\widehat{\mu}_2 = EX_1^2$ and substituting θ_1 in the second equation by $2\widehat{\mu}_1 - \theta_2$ (the first equation), we obtain that

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

which is the same as

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

Since $\theta_2 > EX_1$, we obtain that

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

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

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

Example 3.26

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

$$EX_1 = kp$$

and

$$EX_1^2 = kp(1-p) + k^2p^2,$$

we obtain the moment estimators

$$\widehat{p} = (\widehat{\mu}_1 + \widehat{\mu}_1^2 - \widehat{\mu}_2)/\widehat{\mu}_1 = 1 - \tfrac{n-1}{n} S^2/\bar{X}$$

and

$$\widehat{k} = \widehat{\mu}_1^2/(\widehat{\mu}_1 + \widehat{\mu}_1^2 - \widehat{\mu}_2) = \bar{X}/(1 - \tfrac{n-1}{n} S^2/\bar{X}).$$

The estimator $\widehat{p}$ is in the range of $(0,1)$.
But $\widehat{k}$ may not be an integer.
It can be improved by an estimator that is $\widehat{k}$ 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

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

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

From the SLLN, $\widehat{c}_j$'s are strongly consistent.

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

$$\sqrt{n}\left(\widehat{c}_2 - c_2, ..., \widehat{c}_k - c_k\right) \to_d N_{k-1}(0, D)$$

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

$$c_{i+j+2} - c_{i+1} c_{j+1} - (i+1) c_i c_{j+2} - (j+1) c_{i+2} c_j + (i+1)(j+1) c_i c_j c_2.$$