CHAPTER 14

Variance/Mean Square Estimation

14.1 INTRODUCTION

From survey data we very often estimate population parameters for various characteristics such as population mean, proportion, coefficient of variation, ratio of two characteristics, correlation coefficient, and so on. The variance or mean square error gives us an idea about the magnitude of the sampling error. The variance estimator is used to determine confidence interval of a parameter and testing of hypothesis related to it. It is also used to determine the optimal sample size for a survey design. The expression of variance of a linear estimator can be determined explicitly. Conditions of unbiasedness of variance estimations can also be derived easily. Although the variance is a nonnegative quantity, the unbiased estimators may not be so. If the variance estimator becomes negative, it cannot be used for interval estimation and testing of hypotheses. In this chapter we will study the method of determining variance of linear estimators along with their nonnegativity properties. The necessary condition for nonnegativity of the variance estimators has been established. However, no necessary and sufficient condition of existence of nonnegative variance estimators is available in general.

14.2 LINEAR UNBIASED ESTIMATORS

Let a sample s of size n be selected from a finite population $U=(U_1,\ldots, U_N)$ of N units with probability $p(s)$. Consider a linear homogeneous unbiased estimator for the population total Y

$$t(s, y) = \sum_{i\in s} b_{si} y_i \tag{14.2.1}$$

where b_{si}'s are constants satisfying unbiasedness condition.

$$\sum_{s\supset i} b_{si} p(s) = 1 \quad \forall\, i = 1, \ldots, N \tag{14.2.2}$$

Survey Sampling Theory and Applications
ISBN 978-0-12-811848-1
http://dx.doi.org/10.1016/B978-0-12-811848-1.00014-5

The variance of $t(s, y)$ is given by

$$\begin{aligned} V(t) &= V(s, y) \\ &= \sum_{i \in U} \alpha_{ii} y_i^2 + \sum_{i \neq} \sum_{j \in U} \alpha_{ij} y_i y_j \end{aligned} \tag{14.2.3}$$

where

$$\alpha_{ii} = \sum_{s \supset i} b_{si}^2 p(s) - 1 \text{ and } \alpha_{ij} = \sum_{s \supset ij} b_{si} b_{sj} p(s) - 1 \tag{14.2.4}$$

14.2.1 Conditions of Unbiased Estimation of Variance

Theorem 14.2.1

For a nonzero α_{ij}, a necessary condition of existence of an unbiased estimation of $V(t)$ is π_{ij}, the inclusion probability for the ith and jth units $(i \neq j)$ should be positive.

Proof

Let $\widehat{V}(s, y)$ be an estimator of $V(t)$ based on the selected sample s. If $\pi_{ij} = 0$, then the units i and j cannot occur together in a sample with positive probability and the estimator $\widehat{V}(s, y)$ will be free from the values of y_i and y_j. Hence the average $E\left[\widehat{V}(s, y)\right] = \sum \widehat{V}(s, y) p(s)$ does not involve y_i and y_j. Hence we cannot find an estimator $\widehat{V}(s, y)$ which is unbiased for $V(s, y)$ because α_{ij}, the coefficient of $y_i y_j$ in $V(s, y)$ is not zero.

Theorem 14.2.2

For $\alpha_{ij} \neq 0 \ \forall i \neq j$, a necessary and sufficient condition for a homogeneous quadratic estimator $\widehat{V}(s, y) = \sum_{i \in s} c_{ii}(s) y_i^2 + \sum_{i \neq} \sum_{j \in s} c_{ij}(s) y_i y_j$ to be unbiased for $V(t)$ is $\pi_{ij} > 0$ where $c_{ii}(s)$ and $c_{ij}(s)$ are known constants.

Proof

$\widehat{V}(s, y)$ is unbiased for $V(t)$ if and only if

$$E\left[\widehat{V}(s, y)\right] = V(s, y) \ \forall y \in R^N \tag{14.2.5}$$

The condition (14.2.5) holds if and only if the following conditions are satisfied

$$\sum_{s\supset i} c_{ii}(s)p(s) = \sum_{s} c_{ii}(s)I_{si}p(s) = \alpha_i \quad \forall i = 1, \ldots, N \tag{14.2.6}$$

and

$$\sum_{s\supset ij} c_{ij}(s)p(s) = \sum_{p} c_{ij}(s)p(s)I_{si}I_{sj} = \alpha_{ij} \quad \forall i \neq j = 1, \ldots, N \tag{14.2.7}$$

where $I_{si} = 1$ for $i \in s$ and $I_{si} = 0$ for $i \notin s$.

Because $\alpha_{ij} \neq 0$, Eq. (14.2.5) holds if and only if $I_{si}I_{sj} = 1$ for at least one pair i, j; $i \neq j$. In this situation π_{ij} becomes positive for $\forall i \neq j$.

Furthermore, if $\pi_{ij} > 0$, we can choose $c_{ii}(s)$ and $c_{ij}(s)$ in various ways so that conditions (14.2.6) and (14.2.7) are satisfied. The obvious choices are $c_{ii}(s) = \alpha_{ii}/\pi_i$ and $c_{ij}(s) = \alpha_{ij}/\pi_{ij}$, where π_i is the inclusion probability of the ith unit.

Remark 14.2.1

We can find several unbiased estimators of $V(t)$. For example, the following estimators $\widehat{V}_1(s, y)$ and $\widehat{V}_2(s, y)$ are both unbiased for $V(t)$.

$$\widehat{V}_1(s, y) = \sum_{i\in s} \frac{\alpha_{ii}}{\pi_i} y_i^2 + \sum_{i\neq} \sum_{j\in s} \frac{\alpha_{ij}}{\pi_{ij}} y_i y_j \tag{14.2.8}$$

$$\widehat{V}_2(s, y) = \frac{\sum_{i\in s} \frac{\alpha_{ii}}{\pi_i} M_i(s) y_i^2 + \sum_{i\neq} \sum_{j\in s} \frac{\alpha_{ij}}{\pi_{ij}} M_{ij}(s) y_i y_j}{p(s)} \tag{14.2.9}$$

where $M_i = \sum_s I_{si} =$ number of samples that contain the ith unit and $M_{ij} = \sum_s I_{si}I_{sj} =$ number of samples that contain both the ith and jth units.

Remark 14.2.2

Consider a linear systematic sampling with $N = 12$ and $n = 4$. Here the three possible samples are $s_1 = (1, 4, 7, 10)$, $s_2 = (2, 5, 8, 11)$, and $s_3 = (3, 6, 9, 12)$ each of which has a selection probability 1/3. Here the variance of the sample mean cannot be estimated unbiasedly because the second-order inclusion probabilities for some of the units, e.g., π_{12}, π_{15}, and π_{23} are exactly equal to zero.

14.3 NONNEGATIVE VARIANCE/MEAN SQUARE ESTIMATION

The variance or mean square error of an estimator is a nonnegative quantity, but its unbiased estimators are not always nonnegative. In this section we will establish necessary conditions of existence of nonnegative unbiased estimators of variance/mean square error of an estimator.

Let $T(s, y) = \sum_i c_{si} y_i$ be a linear estimator (not necessary unbiased) of the total Y, where c_{si}'s are constants free of y_i's and $c_{si} = 0$ for $i \notin s$. The mean square error of $T(s, y)$ is given by

$$\begin{aligned} M &= E[T(s, y) - Y]^2 \\ &= \left[\sum_{i=1}^{N} (c_{si} - 1) y_i\right]^2 \\ &= \sum_{i \in U} \beta_{ii} y_i^2 + \sum_{i \neq} \sum_{j \in U} \beta_{ij} y_i y_j \end{aligned} \tag{14.3.1}$$

where

$$\beta_{ii} = E(c_{si} - 1)^2 = \sum_{s \supset i} c_{si}^2 p(s) - 2 \sum_{s \supset i} c_{sj} p(s) + 1 \quad \text{and}$$

$$\beta_{ij} = E(c_{si} - 1)(c_{sj} - 1) = \sum_{s \supset ij} c_{si} c_{sj} p(s) - \sum_{s \supset i} c_{si} p(s) - \sum_{s \supset j} c_{sj} p(s) + 1 \tag{14.3.2}$$

We now present the following theorem proposed by Rao and Vijayan (1977).

Theorem 14.3.1

For a fixed sample of size n design, if M, the mean square error of $T(s, y)$, becomes zero when $z_i = y_i / w_i = k$, a constant, and $w_i(\neq 0)$'s are known for $i = 1, \ldots, N$, then

(i) The mean square error M can be written as

$$M = -\frac{1}{2} \sum_{i \neq} \sum_{j \in U} \beta_{ij} w_i w_j (z_i - z_j)^2$$

(ii) A nonnegative quadratic unbiased estimator of M is necessarily of the form

$$\widehat{M}(s, z) = -\frac{1}{2} \sum_{i \neq} \sum_{j \in s} \beta_{ij}(s) w_i\, w_j (z_i - z_j)^2 \tag{14.3.3}$$

where $\beta_{ij}(s)$'s are constants independent of y_i's, equal zero if $i, j \notin s$, and satisfy the unbiasedness condition

$$E\{\beta_{ij}(s)\} = \sum_{s \supset i,j} \beta_{ij}(s)p(s) \ = \ \beta_{ij} \text{ for } \forall\, i \neq j \in U$$

Proof

(i) The mean square error M given in Eq. (14.3.1) can be written as

$$M = \mathbf{y}' \, \mathbf{B} \, \mathbf{y} \tag{14.3.4}$$

where $\mathbf{y}' = (y_1, \ldots, y_N)$ and $\mathbf{B} = (\beta_{ij})$ is an $N \times N$ matrix with ith and jth element β_{ij}.

Because $M \geq 0$ for $\mathbf{y} \in R^N$, we find that the matrix $\mathbf{B}$ is a positive semidefinite and hence we can write

$$\mathbf{B} = \mathbf{H}'\mathbf{H} \tag{14.3.5}$$

Substituting $\mathbf{y} = k\mathbf{w} = k(w_1, \ldots, w_N)'$ in Eq. (14.3.4) and using the condition stated in the theorem we get

$$M = k^2\mathbf{w}'\mathbf{B}\mathbf{w} = 0$$
$$\text{i.e., } (\mathbf{H}\mathbf{w})'(\mathbf{H}\mathbf{w}) = 0$$
$$\text{i.e.,} \quad \mathbf{H}\mathbf{w} = 0$$
$$(\text{where } \mathbf{0} \text{ is a null column vector})$$
$$\text{i.e.,} \quad \mathbf{B}\mathbf{w} = 0$$
$$\text{i.e.,} \quad \beta_{ii}w_i^2 + w_i \sum_{j(\neq i) \in U} \beta_{ij}w_j = 0 \tag{14.3.6}$$

Now

$$\begin{aligned} M &= \sum_{i \in U} \beta_{ii}y_i^2 + \sum_{i \neq} \sum_{j \in U} \beta_{ij}y_i y_j \\ &= \sum_{i \in U} \beta_{ii}w_i^2 z_i^2 + \sum_{i \neq} \sum_{j \in U} \beta_{ij}w_i w_j z_i z_j \end{aligned} \tag{14.3.7}$$

$$(\text{writing } z_i = y_i/w_i)$$

Substituting Eq. (14.3.6) in Eq. (14.3.7) yields

$$M = -\sum_{i \neq} \sum_{j \in U} \beta_{ij} w_i w_j \left(z_i^2 - z_i z_j\right)$$

$$= -\frac{1}{2} \sum_{i \neq} \sum_{j \in U} \beta_{ij} w_i w_j (z_i - z_j)^2.$$

(ii) Because $\widehat{M}(s)$ is a nonnegative homogeneous quadratic unbiased estimator for M, we can write $\widehat{M}(s)$ as

$$\widehat{M}(s,y)=\sum_{i\in s}\beta_{ii}(s)y_i^2+\sum_{i\neq}\sum_{j\in s}\beta_{ij}(s)y_iy_j \tag{14.3.8}$$

where $\beta_{ii}(s)$ and $\beta_{ij}(s)$ are known constants independent of y_i's and satisfy the unbiasedness condition

$$\sum_s \widehat{M}(s,y)p(s)=M \tag{14.3.9}$$

Putting $y_i=kw_i \ \forall i=1,\ldots, N$, we find (from the condition of the theorem)

$$\sum_s \widehat{M}(s,w)p(s)=M=0 \tag{14.3.10}$$

Furthermore, the condition $\widehat{M}(s,w)\geq 0 \ \ \forall s$ with $p(s)>0$ yields

$$\widehat{M}(s,w)=0 \ \forall s \text{ with } p(s)>0$$

$$\text{i.e.,}\quad \sum_{i\in s}\beta_{ii}(s)w_i^2+\sum_{i\neq}\sum_{j\in s}\beta_{ij}(s)w_iw_j=0 \ \forall s \text{ with } \ p(s)>0 \tag{14.3.11}$$

Because $\widehat{M}(s,y)$ is a nonnegative quadratic form and $\widehat{M}(s,y)=0$ for $y_i=kw_i$ for $i\in s$, Eq. (14.3.6) yields

$$\beta_{ii}(s)w_i^2+w_i\sum_{j(\neq i)\in s}\beta_{ij}(s)w_j=0 \tag{14.3.12}$$

$$\text{i.e.,}\quad \beta_{ii}(s)w_i^2z_i^2=-w_iz_i^2\sum_{j(\neq i)\in s}\beta_{ij}(s)w_j$$

$$\begin{aligned}\text{i.e.,}\quad \sum_{i\in s}\beta_{ii}(s)w_i^2z_i^2&=-\sum_{i\in s}w_iz_i^2\sum_{j(\neq i)\in s}\beta_{ij}(s)w_j\\&=-\frac{1}{2}\sum_{i\neq}\sum_{j\in s}\beta_{ij}(s)w_iw_j\left(z_i^2+z_j^2\right)\end{aligned} \tag{14.3.13}$$

Substituting Eq. (14.3.13) in Eq. (14.3.8), we get

$$\widehat{M}(s,y)=\widehat{M}(s,z)=-\frac{1}{2}\sum_{i\neq}\sum_{j\in s}w_iw_j\beta_{ij}(s)(z_i-z_j)^2.$$

Remark 14.3.1

The constants $\beta_{ij}(s)$'s that satisfy $E\{\beta_{ij}(s)\} = \beta_{ij}$ can be chosen in various ways. The obvious choice is $\beta_{ij}(s) = \beta_{ij}/\pi_{ij}$, provided π_{ij}'s are easy to compute. Alternatively, we may choose $\beta_{ij}(s) = k_{ij}(s) = \beta_{ij}/\{M_{ij}p(s)\}$, where M_{ij} is the number of samples that contain both the ith and jth units. If all $\binom{N}{n}$ possible samples of size n have positive probabilities, then $M_{ij} = M_2 = \binom{N-2}{n-2}$. A more general choice given by Rao (1979) is $\beta_{ij}(s) = \beta_{ij}d_{ij}(s)/E\{d_{ij}(s)\}$, where $d_{ij}(s) = 0$ if the sample s does not contain both the units i and j and $E\{d_{ij}(s)\}$ is easy to compute.

Remark 14.3.2

For a fixed sample size $n = 2$ design, Lanke (1974a,b) relaxed the restriction of the class of quadratic estimators and has shown that any nonnegative unbiased estimator (not necessarily quadratic) of M is necessarily of the form (14.3.3). For $n > 2$, Vijayan (1975) proved that a nonnegative unbiased polynomial estimator of M is necessarily of the form (14.3.3).

14.3.1 Examples

14.3.1.1 Horvitz–Thompson Estimator

Consider the Horvitz–Thompson estimator $\widehat{Y}_{ht} = \sum_{i\in s}\frac{y_i}{\pi_i}$ based on a fixed sample size n design. The estimator $\widehat{Y}_{ht}$ is unbiased and becomes a constant (its variance becomes zero) when $y_i = k\pi_i$ for $i = 1,\ldots, N$ and k is a constant. Here $c_{si} = 1/\pi_i$ and $\beta_{ij} = \sum_{s\supset i,j}\frac{p(s)}{\pi_i\pi_j} - 1 = \frac{\pi_{ij}}{\pi_i\pi_j} - 1$. Hence, the variance of $\widehat{Y}_{ht}$ can be written as

$$V\left(\widehat{Y}_{ht}\right) = M = -\frac{1}{2}\sum_{i\neq}\sum_{j\in U}\beta_{ij}\pi_i\pi_j\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

A nonnegative unbiased variance estimator is necessarily of the form

$$\widehat{V}\left(\widehat{Y}_{ht}\right) = -\frac{1}{2}\sum_{i\neq}\sum_{j\in s}\beta_{ij}(s)\pi_i\pi_j\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

with $\sum_{s\supset i,j}\beta_{ij}(s)p(s) = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}$.

If we choose $\beta_{ij}(s)=\delta_{ij}$ independent of s, then the constraint $\delta_{ij}\sum_{s\supset i,j}p(s)=\dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$ yields $\delta_{ij}=\dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j\pi_{ij}}$. The variance estimator $\widehat{V}\left(\widehat{Y}_{ht}\right)$ reduces to Yates-Grundy's (1953) estimator viz.

$$\widehat{V}\left(\widehat{Y}_{ht}\right)=\widehat{V}_{YG}=\frac{1}{2}\sum_{i\neq}\sum_{j\in s}\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (14.3.14)$$

The estimator $\widehat{V}_{YG}$ is nonnegative for those samples for which $\pi_i\pi_j-\pi_{ij}\geq 0 \quad \forall i,\ j\in s$. However, for a sampling design with $\pi_i\pi_j-\pi_{ij}\geq 0\ \forall i,j\in U$, $\widehat{V}_{YG}\geq 0$.

For $n=2$, the probability of selecting the sample s, containing the units $i,\ j\ (i\neq j)$, is $p(s)=\pi_{ij}$. Hence the condition $\sum_{s\supset i,j}\beta_{ij}(s)p(s)=\dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$ reduces to $\beta_{ij}(s)\pi_{ij}=\dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$, i.e., $\beta_{ij}(s)=\dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$. Hence for $n=2$, the nonnegative variance estimator of $V\left(\widehat{Y}_{ht}\right)$ should be necessarily of the form

$$\widehat{V}\left(\widehat{Y}_{ht}\right)=\widehat{V}_{YG}=\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2$$

which is the Yates-Grundy (1953) estimator. Hence for $n=2$, $\widehat{V}_{YG}=\dfrac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\dfrac{y_i}{\pi_i}-\dfrac{y_j}{\pi_j}\right)^2$ becomes the unique nonnegative estimator of $V\left(\widehat{Y}_{ht}\right)$ in the class of quadratic unbiased estimators if and only if $\pi_i\pi_j-\pi_{ij}\geq 0\ \forall i,j\in U$.

Following Remark 14.3.2, we note that for $n=2$, any nonnegative unbiased estimator (not necessarily quadratic) necessarily becomes the Yates-Grundy (1953) estimator.

The choice of $\beta_{ij}(s)=\dfrac{\pi_i\pi_j-\pi_{ij}}{\pi_i\pi_jM_{ij}p(s)}$ yields an alternative unbiased estimator of $V\left(\widehat{Y}_{ht}\right)$ as

$$\widehat{V}^*\left(\widehat{Y}_{ht}\right)=\frac{1}{2p(s)}\sum_{i\neq}\sum_{j\in s}\frac{\pi_i\pi_j-\pi_{ij}}{M_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (14.3.15)$$

where M_{ij}'s are defined in Remark 14.3.1.

14.3.1.2 Hansen–Hurwitz Estimator

The Hansen–Hurwitz estimator based on PPSWR sampling design is given by

$$\widehat{Y}_{hh}=\frac{1}{n}\sum_{i\in s}n_i(s)\frac{y_i}{p_i}$$

where $n_i(s)$ is the number of times the ith unit appears in the sample s and $p_i(>0)$ is the normed size measure for the ith unit.

The estimator $\widehat{Y}_{hh}$ is unbiased for Y. The estimator $\widehat{Y}_{hh}$ becomes a constant if the y_i's are proportional to p_i's and in this case the variance of $\widehat{Y}_{hh}$ is zero. Here $c_{si} = \dfrac{n_i(s)}{np_i}$ and $\beta_{ij} = \dfrac{1}{n^2 p_i p_j} \sum_{s \supset i,j} n_i(s) n_j(s) p(s) - 1 = \dfrac{E[n_i(s)n_j(s)]}{n^2 p_i p_j} - 1 = -\dfrac{1}{n}$. Hence the variance of $\widehat{Y}_{hh}$ can be written as

$$V_{hh} = \frac{1}{2n} \sum_{i \neq} \sum_{j \in U} p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2$$

If we choose $\beta_{ij}(s) = \beta_{ij} \dfrac{n_i(s) n_j(s)}{E\{n_i(s) n_j(s)\}} = -\dfrac{n_i(s) n_j(s)}{n^2 (n-1) p_i p_j}$, we get an unbiased estimator of V_{hh} as

$$\begin{aligned} \widehat{V}_{hh}(1) &= \frac{1}{2n^2(n-1)} \sum_{i \neq} \sum_{j \in s} n_i(s) n_j(s) \left(\frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 \\ &= \frac{1}{n(n-1)} \sum_{i \in s} n_i(s) \left(\frac{y_i}{p_i} - \widehat{Y}_{hh} \right)^2 \end{aligned} \tag{14.3.16}$$

On the other hand, if we choose $\beta_{ij}(s) = -\beta_{ij} / \pi_{ij}$, with $\pi_{ij} = 1 - (1 - p_i)^n - (1 - p_j)^n + (1 - p_i - p_j)^n$, we get an alternative unbiased estimator of V_{hh} as

$$\widehat{V}_{hh}(2) = \frac{1}{2} \sum_{i \neq} \sum_{j \in s} \frac{p_i p_j}{\pi_{ij}} \left(\frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 \tag{14.3.17}$$

14.3.1.3 Murthy's Estimator

The Murthy's (1957) estimator for Y based on PPSWOR sampling scheme described in Section 5.3.2.1 is given by $\widehat{Y}_{mur} = \sum_{i \in s} \dfrac{p(s|i)}{p(s)} y_i$, where p_i is the probability of selection of the unit at the first draw and $p(s|i)$ is the conditional probability of the selection of the sample s given that the ith unit is selected at the first draw. Here $c_{si} = \dfrac{p(s|i)}{p(s)}$, $\sum_{s \supset i} c_{si} p(s) = 1$, and $\beta_{ij} = \omega_{ij} - 1$ with $\omega_{ij} = \sum_{s \supset i,j} \dfrac{p(s|i) p(s|j)}{p(s)}$. The estimator $\widehat{Y}_{mur}$ is unbiased for Y and if

$y_i = kp_i$, then $\widehat{Y}_{mur} = \frac{k}{p(s)} \sum_{i \in s} p_i p(s|i) = k =$ constant, because $\sum_{i \in s} p_i p(s|i) = p(s)$. Hence the variance of $\widehat{Y}_{mur}$ can be written as

$$V_{mur} = V\left(\widehat{Y}_{mur}\right) = \frac{1}{2} \sum_{i \neq} \sum_{j \in U} p_i p_j (1 - \omega_{ij}) \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \tag{14.3.18}$$

Clearly an unbiased estimator of V_{mur} is given by

$$\widehat{V}_{mur}(1) = \frac{1}{2} \sum_{i \neq} \sum_{j \in s} \frac{p_i p_j}{\pi_{ij}} (1 - \omega_{ij}) \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \tag{14.3.19}$$

The estimator (14.3.19) is found to be nonnegative (Rao, 1979).

Following Rao (1979), if we choose $\beta_{ij}(s) = c_{si} c_{sj} - \frac{p(s|i,j)}{p(s)} = \frac{p(s|i)p(s|j)}{\{p(s)\}^2} - \frac{p(s|i,j)}{p(s)}$ so that $\sum_{s \supset i,j} \beta_{ij}(s) p(s) = \beta_{ij}$, we get an alternative unbiased estimator of V_{mur} as

$$\widehat{V}_{mur}(2) = \frac{1}{2} \sum_{i \neq} \sum_{j \in s} \frac{p_i p_j}{\{p(s)\}^2} \{p(s)p(s|i,j) - p(s|i)p(s|j)\} \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \tag{14.3.20}$$

The estimator (14.3.20) is nonnegative as observed by Pathak and Shukla (1966).

For $n = 2$, $\widehat{V}_{\mathrm{mur}}(1)$ reduces to a unique nonnegative unbiased variance estimator proposed by Pathak and Shukla (1966) as

$$\widehat{V}_{mur}(1) = \frac{1}{2} \sum_{i} \sum_{j \in s} \frac{(1 - p_i)(1 - p_j)(1 - p_i - p_j)}{(1 - p_i - p_j)^2} \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \tag{14.3.21}$$

14.3.1.4 Unbiased Ratio Estimator

For Lahiri–Midzuno–Sen (1951, 1952, 1953) sampling scheme, the probability of selecting a sample of size n is $p(s) = \frac{x_s}{M_1 X}$, where $x_s = \sum_{i \in s} x_i$, $X = \sum_{i \in U} x_i$, $M_1 = \binom{N-1}{n-1}$, and $x_i(>0)$ is the size measure for the ith units (see Section 5.5). An unbiased estimator for the population total Y is given by

$$\widehat{Y}_{lms} = \frac{y_s}{x_s} X \text{ with } y_s = \sum_{i \in s} y_i$$

If $y_i = kx_i$ then the estimator $\widehat{Y}_{lms}$ becomes equal to kX, which is a constant and hence the variance of $\widehat{Y}_{lms}$ becomes zero. Here $c_{si} = X/x_s$ and

$$\beta_{ij} = X^2 \sum_{s \supset i,\, j} \frac{1}{x_s^2} p(s) - 1 = \frac{X}{M_1} \sum_{s \supset i,\, j} \frac{1}{x_s} - 1 \qquad (14.3.22)$$

So the variance of $\widehat{Y}_{lms}$ can be written as

$$V_{lms} = V\left(\widehat{Y}_{lms}\right) = -\frac{1}{2} \sum_{i \neq} \sum_{j \in U} \beta_{ij} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \qquad (14.3.23)$$

From Eq. (14.3.23), we immediately write an unbiased estimator of V_{lms} as

$$\widehat{V}_{lms}(1) = -\frac{1}{2} \sum_{i \neq} \sum_{j \in s} \frac{\beta_{ij} x_i x_j}{\pi_{ij}} \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \qquad (14.3.24)$$

where β_{ij} is given in Eq. (14.3.22).

Furthermore, writing $\beta_{ij} = \sum_{s \supset i,j} \left(\frac{1}{x_s} \frac{X}{M_1} - \frac{1}{M_2}\right)$ with $M_2 = \binom{N-2}{n-2}$, we can choose

$$\beta_{ij}(s) = \frac{1}{p(s)} \left(\frac{1}{x_s} \frac{X}{M_1} - \frac{1}{M_2}\right) = \frac{X}{x_s} \left(\frac{X}{x_s} - \frac{N-1}{n-1}\right) \qquad (14.3.25)$$

so that $\sum_{s \supset i,j} \beta_{ij}(s) p(s) = \beta_{ij}$. The choice of $\beta_{ij}(s)$ in Eq. (14.3.25) yields an alternative unbiased estimator of V_{lms} as

$$\widehat{V}_{lms}(2) = \frac{1}{2} \frac{X}{x_s} \left(\frac{N-1}{n-1} - \frac{X}{x_s}\right) \sum_{i \neq} \sum_{j \in s} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \qquad (14.3.26)$$

Rao and Vijayan (1977) proved that none of the estimators $\widehat{V}_{lms}(1)$ and $\widehat{V}_{lms}(2)$ are always nonnegative. No nonnegative variance estimators of V_{lms} are as yet available.

14.3.1.5 Ordinary Ratio Estimator

The ratio estimator for the population total Y based on a fixed sample s of size n under simple random sampling without replacement (SRSWOR) sampling is

$$\widehat{Y}_R = \frac{y_s}{x_s} X$$

where $y_s = \sum_{i \in s} y_i$ and $x_s = \sum_{i \in s} x_i$.

Here $c_{si} = (X/x_s)I_{si}$ and $\widehat{Y}_R$ becomes a constant if y_i is proportional to x_i. Hence we can write the mean square error of $\widehat{Y}_R$ as

$$M\left(\widehat{Y}_R\right) = -\frac{1}{2}\sum_{i\neq}\sum_{j\in U}\beta_{ij}x_ix_j\left(\frac{y_i}{x_i}-\frac{y_j}{x_j}\right)^2 \tag{14.3.27}$$

where

$$\begin{aligned}\beta_{ij} &= E(c_{si}c_{sj}) - E(c_{si}) - E(c_{sj}) + 1\\ &= X^2E\left(\frac{I_{si}I_{sj}}{x_s^2}\right) - XE\left(\frac{I_{si}}{x_s}\right) - XE\left(\frac{I_{sj}}{x_s}\right) + 1\end{aligned}$$

For SRSWOR $p(s) = 1\Big/\binom{N}{n}$ and hence we can write

$$\beta_{ij} = \binom{N}{n}^{-1}\left[X^2\sum_{s\supset i,j}\frac{1}{x_s^2} - X\sum_{s\supset i}\frac{1}{x_s} - X\sum_{s\supset j}\frac{1}{x_s} + \binom{N}{n}\right]$$

An unbiased estimator of $M\left(\widehat{Y}_R\right)$ is given by

$$\begin{aligned}\widehat{M}\left(\widehat{Y}_R\right) &= -\frac{1}{2}\sum_{i\neq}\sum_{j\in s}\frac{\beta_{ij}x_ix_j}{\pi_{ij}}\left(\frac{y_i}{x_i}-\frac{y_j}{x_j}\right)^2\\ &\left(\text{for SRSWOR } \pi_{ij} = \frac{N(N-1)}{n(n-1)}\right)\\ &= -\frac{1}{2}\frac{N(N-1)}{n(n-1)}\sum_{i\neq}\sum_{j\in s}\beta_{ij}x_ix_j\left(\frac{y_i}{x_i}-\frac{y_j}{x_j}\right)^2\end{aligned} \tag{14.3.28}$$

Remark 14.3.3

It is worth noting that the expressions of mean square error and its unbiased estimators given in Eqs. (14.3.27) and (14.3.28) are exact. The expressions of mean square error and its unbiased estimators given in Eqs. (8.4.7) and (8.4.8) are approximate and are valid when the sample size n is large.

14.3.1.6 Hartley–Ross Estimator

The Hartley–Ross (1954) estimator for the population total Y is given by

$$\widehat{Y}_{hr} = X\,\overline{r} + \frac{(N-1)n}{n-1}\left(\overline{y} - \overline{r}\,\overline{x}\right) \tag{14.3.29}$$

where $\bar{r} = \frac{1}{n}\sum_{i\in s}\frac{y_i}{x_i}$, $\bar{y} = \frac{1}{n}\sum_{i\in s} y_i$, $\bar{x} = \frac{1}{n}\sum_{i\in s} x_i$, and s is a sample of size n selected by SRSWOR method.

We can write

$$\widehat{Y}_{hr} = \sum_{i\in s} c_{si}y_i$$

where $c_{si} = \frac{1}{x_i}\left(\frac{X}{n} - \frac{N-1}{n-1}\bar{x}\right) + \frac{N-1}{n-1}$.

Because $\widehat{Y}_{hr}$ becomes constant when $y_i \propto x_i$, the variance of $\widehat{Y}_{hr}$ can be written as

$$V\left(\widehat{Y}_{hr}\right) = -\frac{1}{2}\sum_{i\neq}\sum_{j\in U}\beta_{ij}x_i x_j\left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \tag{14.3.30}$$

where $\beta_{ij} = E(c_{si} - 1)(c_{sj} - 1)$.

The expression of β_{ij} is very complex and was obtained by Robson (1957).

An unbiased estimator of $V\left(\widehat{Y}_{hr}\right)$ is

$$\widehat{V}\left(\widehat{Y}_{hr}\right) = -\frac{1}{2}\frac{N(N-1)}{n(n-1)}\sum_{i\neq}\sum_{j\in s}\beta_{ij}x_i x_j\left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2 \tag{14.3.31}$$

However, nonnegativity of the estimator $\widehat{V}\left(\widehat{Y}_{hr}\right)$ is not easy to check.

14.4 EXERCISES

14.4.1 Let $t(s) = \sum_{i\in s} b_{si}y_i$ be an unbiased estimator of the population total Y based on a sample s of size n selected with probability $p(s)$ and $b_{si} = 0$ for $i \notin s$. Show that the following estimators are unbiased for the variance of $t(s)$.

(i) $\widehat{V}_1 = \{t(s)\}^2 - \left(\sum_{i\in s} b_{si}y_i^2 + \sum_{i\neq}\sum_{j\in s}\frac{b_{si}b_{sj}}{\delta_{ij}}y_i y_j\right)$

(ii) $\widehat{V}_2 = \{t(s)\}^2 - \frac{1}{p(s)}\left(\sum_{i\in s}\frac{y_i^2}{M_i} + \sum_{i\neq}\sum_{j\in s}\frac{y_i y_j}{M_{ij}}\right)$

(iii) $\widehat{V}_3 = \sum_{i\in s} b_{si}(b_{si} - 1)y_i^2 + \sum_{i\neq}\sum_{j\in s} b_{si}b_{sj}\left(1 - \frac{1}{\delta_{ij}}\right)y_i y_j$

where $\delta_{ij} = E(b_{si}b_{sj})$, M_i and M_{ij} are, respectively, the total number of samples containing ith, and ith and $j(\neq i)$th units, and all are positive.

14.4.2 Show that the following are unbiased estimators of $V\left(\widehat{Y}_{ht}\right)$ based on a fixed effective size n sampling design.

(i) $\widehat{V}_1 = \dfrac{1}{2}\sum_{i\neq}\sum_{j\in s}\left(\dfrac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)(z_i - z_j)^2$

(ii) $\widehat{V}_2 = \dfrac{1}{2}\sum_{i\neq}\sum_{j\in s}\left(\dfrac{\pi_i\pi_j}{\pi_{ij}} - \dfrac{\pi_{ij}}{M_{ij}p(s)}\right)(z_i - z_j)^2$

(iii) $\widehat{V}_3 = \dfrac{1}{2p(s)}\sum_{i\neq}\sum_{j\in s}\left(\dfrac{\pi_i\pi_j - \pi_{ij}}{M_{ij}}\right)(z_i - z_j)^2$

(iv) $\widehat{V}_4 = \dfrac{1}{p(s)}\left[\sum_{i\in s}\left(\dfrac{1-\pi_i}{M_i\pi_i}\right)y_i^2 + \sum_{i\neq}\sum_{j\in s}\left(\dfrac{\pi_{ij} - \pi_i\pi_j}{M_{ij}\pi_i\pi_j}\right)y_iy_j\right]$

where $z_i = y_i/\pi_i$ and M_i and M_{ij} are as in Exercise 14.4.1.

14.4.3 Let s be a fixed sample of size n distinct units with selection probability $p(s)$ and α and β are fixed constants in [0,1]. Show that

$$\widehat{V}_{ht} = \frac{1}{2}\sum_{i\neq}\sum_{j\in s}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2\left[\pi_i\pi_j\left\{\frac{\alpha}{p(s)M_{ij}} + \frac{1-\alpha}{\pi_{ij}}\right\} - \pi_{ij}\left\{\frac{\beta}{p(s)M_{ij}} + \frac{1-\beta}{\pi_{ij}}\right\}\right]$$

is an unbiased estimator of the variance of $\widehat{Y}_{ht} = \sum_{i\in s}\dfrac{y_i}{\pi_i}$.

Furthermore, for the choices $\alpha = 0, \beta = 1$ and $\alpha = 1, \beta = 0$, show that sufficient condition for nonnegativity of $\widehat{V}_{ht}$ is $p(s) > \dfrac{\pi_{ij}^2}{\pi_i\pi_jM_{ij}}$ $\forall s \supset i, j$ and $p(s) < (\pi_i\pi_j)/M_{ij}$ $\forall s \supset i, j$, respectively, where M_{ij} is defined in Exercise 14.4.1 (Chaudhuri, 1981).

14.4.4 Show that the variance of the unbiased ratio estimator $\widehat{Y}_{lms} = \dfrac{y_s}{x_s}X$ for the population total Y based on an LMS can be estimated unbiasedly using the formula

$$\widehat{V}_{lms} = \frac{1}{2}\sum_{i\neq}\sum_{j\in s}p_ip_j\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2\left[\left\{\frac{\alpha}{\pi_{ij}} + \frac{(1-\alpha)}{M_{ij}p(s)}\right\} - \left\{\frac{\beta}{M_{ij}p(s)} + \frac{1-\beta}{\pi_{ij}}\right\}\frac{1}{M_i^2}\sum_{s\supset i,j}\frac{1}{p(s)}\right]$$

where M_i, M_{ij}, and α, β are as in Exercises 14.4.1 and 14.4.3, respectively (Chaudhuri, 1981).

14.4.5 Let s be a sample of size n selected with probability $p(s)$ by LMS sampling using x_i as a measure of size for the ith unit ($i = 1,\ldots, N$) of a population U of size N. Show that

(i) $t = \dfrac{y_s}{x_s}$ is an unbiased estimator for the population ratio $R = Y/X$, where $y_s = \sum_{i\in s} y_i$ and $x_s = \sum_{i\in s} x_i$,

(ii) The variance of t can be written as follows:

(a) $V_1(t) = \sum_{i\in U} \lambda_i y_i^2 + \sum_{i\neq}\sum_{j\in U} \lambda_{ij} y_i y_j$

(b) $V_2(t) = \sum_{i\in U} (nT_i - N) r_i^2 + \sum_{i<}\sum_{j\in U} (1 - T_{ij})(r_i - r_j)^2$

(c) $V_3(t) = \frac{1}{2}\sum_{i\neq}\sum_{j\in U} Q_{ij}$

(d) $V_4(t) = \frac{1}{2}\sum_{i\neq}\sum_{j\in U} R_{ij}$

where

$$\lambda_i = \frac{1}{MX}\sum_{s\supset i}\frac{1}{x_s} - \frac{1}{X^2},\ \lambda_{ij} = \frac{1}{MX}\sum_{s\supset i,j}\frac{1}{x_s} - \frac{1}{X^2},$$

$$M = \binom{N-1}{n-1},\ r_i = y_i/x_i,\ T_i = 1 + \lambda_i X^2,\ X = \sum_{i\in U} x_i,$$

$$T_{ij} = 1 + \lambda_{ij} X^2,\ Q_{ij} = \frac{T_i - 1}{N-1} r_i^2 + 2(T_{ij} - 1) r_i r_j + \frac{T_j - 1}{N-1} r_j^2 \text{ and}$$

$$R_{ij} = \left(\frac{T_{ij}}{n-1} - \frac{1}{N-1}\right) r_i^2 + 2(T_{ij} - 1) r_i r_j + \left(\frac{T_{ij}}{n-1} - \frac{1}{N-1}\right) r_j^2.$$

(iii) Show that $\widehat{V}_j(t)$ is unbiased for $V_j(t)$ for $j = 1, 2, 3, 4$, where

$$\widehat{V}_1(t) = \sum_{i\in U} \lambda_i y_i^2/\pi_i + \sum_{i\neq}\sum_{j\in U} \lambda_{ij} y_i y_j/\pi_{ij},$$

$$\widehat{V}_2(t) = \sum_{i\in s} (nT_i - N) r_i^2/\pi_i + \sum_{i<}\sum_{j\in s} (1 - T_{ij})(r_i - r_j)^2/\pi_{ij},$$

$$\widehat{V}_3(t) = \frac{1}{2}\sum_{i\neq}\sum_{j\in s} Q_{ij}/\pi_{ij} \text{ and } \widehat{V}_4(t) = \frac{1}{2}\sum_{i\neq}\sum_{j\in s} R_{ij}/\pi_{ij}$$

(π_i is inclusion probability for the ith unit and π_{ij} is the inclusion probability for the ith and $j(\neq i)$th units)

(iv) In case $y_i \geq 0$ for $i = 1, \ldots, N$, a sufficient condition of nonnegativity of $\widehat{V}_1(t)$ is $\lambda_{ij} \geq 0$ (Rao, 1972).
(v) Sufficient condition for nonnegativity of $\widehat{V}_2(t)$ is $T_i \geq N/n$ and $T_{ij} \leq 1$ for $i \neq j = 1, \ldots, N$ (Chaudhuri, 1976).
(vi) Sufficient condition for nonnegativity of $\widehat{V}_3(t)$ is $Q_{ij} \geq 0$ for $i \neq j = 1, \ldots, N$ (Rao, T. J., 1977a,b).
(vii) Sufficient condition for nonnegativity of $\widehat{V}_4(t)$ is $T_{ij} \geq \dfrac{n-1}{N-1}$ and $\left(T_{ij} - 1\right)^2 - \left(\dfrac{T_{ij}}{n-1} - \dfrac{1}{N-1}\right)^2 \leq 0$ for $i \neq j = 1, \ldots, N$ (Rao, T. J., 1977a,b).
(viii) Show that the sufficient conditions listed in (iv) to (vii) can never hold for $i \neq j = 1, \ldots, N$ except under trivial situation when $p_i = x_i/X = 1/N \;\; \forall i = 1, \ldots, N$ (Chaudhuri and Arnab, 1981).