Chapter 10

Variance Estimation

10.1 Introduction

Variance estimation is an important practical problem in survey sampling. Variance estimates are used in two purposes. One is the analytic purpose such as constructing confidence intervals or performing hypothesis testings. The other is descriptive purposed to evaluate the efficiency of the survey designs or estimates for planning surveys.

Desirable variance estimates should satisfy the following properties:

- It should be unbiased, or approximately unbiased.
- Variance of the variance estimator should be small. That is, the variance estimator is stable.
- It should not take negative values.
- The computation should be simple.

The HT variance estimator is unbiased but it can take negative values. Also, computing the joint inclusion probabilities π_{ij} can be cumbersome when the sample size is large.

Example 10.1. *Consider a finite population of size $N = 3$ with $y_1 = 16, y_2 = 21$ and $y_3 = 18$ and consider the following sampling design.*

Table 10.1: A sampling design

Sample (A)	$Pr(A)$	HT estimator	HT variance estimator
$A_1 = \{1,2\}$	0.4	50	206
$A_2 = \{1,3\}$	0.3	50	200
$A_3 = \{2,3\}$	0.2	60	-90
$A_4 = \{1,2,3\}$	0.1	80	-394

The sampling variance of HT estimator is 85. Note that HT variance estimator has expectation

$$206 \times 0.4 + 200 \times 0.3 + (-90) \times 0.2 + (-394) \times 0.1 = 85$$

but it can take negative values in some samples.

The variance estimator under PPS sampling is always nonnegative and can be computed without computing the joint inclusion probability. In practice, PPS sampling variance estimator is often applied as an alternative variance estimator even for without-replacement sampling. The resulting variance estimator can be written

$$\hat{V}_0 = \frac{1}{n(n-1)} \sum_{i \in A} \left(\frac{y_i}{p_i} - \hat{Y}_{HT} \right)^2 = \frac{n}{(n-1)} \sum_{i \in A} \left(\frac{y_i}{\pi_i} - \frac{1}{n} \hat{Y}_{HT} \right)^2, \tag{10.1}$$

where $p_i = \pi_i / n$. The following theorem express the bias of the simplified variance estimator in (10.1) as an estimator of the variance of HT estimator.

Theorem 10.1. *The simplified variance estimator in (10.1) satisfies*

$$E\left(\hat{V}_0\right) - Var\left(\hat{Y}_{HT}\right) = \frac{n}{n-1} \left\{ Var\left(\hat{Y}_{PPS}\right) - Var\left(\hat{Y}_{HT}\right) \right\} \tag{10.2}$$

where

$$Var\left(\hat{Y}_{PPS}\right) = \frac{1}{n} \sum_{i=1}^{N} p_i \left(\frac{y_i}{p_i} - Y \right)^2$$

and $p_i = \pi_i / n$.

Proof. Note that $\hat{V}_0$ satisfies

$$\begin{aligned} & E\left\{\sum_{k\in A}\left(\frac{y_k}{p_k}-Y+Y-\hat{Y}_{HT}\right)^2\right\} \\ = \; & E\left\{\sum_{k\in A}\left(\frac{y_k}{p_k}-Y\right)^2\right\}+2E\left\{\sum_{k\in A}\left(\frac{y_k}{p_k}-Y\right)\left(Y-\hat{Y}_{HT}\right)\right\}+E\left\{\sum_{k\in A}\left(Y-\hat{Y}_{HT}\right)^2\right\}. \end{aligned}$$

The first term is

$$\begin{aligned} E\left\{\sum_{k\in A}\left(\frac{y_k}{p_k}-Y\right)^2\right\} &= \sum_{k=1}^{N}\left(\frac{y_k}{p_k}-Y\right)^2\pi_k \\ &= n^2 Var\left(\hat{Y}_{PPS}\right) \end{aligned}$$

and, using

$$\sum_{k\in A}\left(\frac{y_k}{p_k}-Y\right)=n\left(\sum_{k\in A}\frac{y_k}{\pi_k}-Y\right)=n\left(\hat{Y}_{HT}-Y\right),$$

the second term equals to $-2nVar\left(\hat{Y}_{HT}\right)$ and the last term is equal to $nVar\left(\hat{Y}_{HT}\right)$, which proves the result. □

In many cases, the bias term in (10.2) is positive and the simplified variance estimator is conservatively estimate the variance. Under SRS, the relative bias of the simplified variance estimator (10.1) is

$$\frac{\hat{V}_0-Var\left(\hat{Y}_{HT}\right)}{Var\left(\hat{Y}_{HT}\right)}=\frac{n}{N-n} \tag{10.3}$$

and the relative bias is negligible when n/N is negligible.

The simplified variance estimator can be directly applicable to multistage sampling designs. Under multistage sampling design, the HT estimator of the total can be written

$$\hat{Y}_{HT}=\sum_{i\in A_I}\frac{\hat{Y}_i}{\pi_{Ii}}$$

where $\hat{Y}_i$ is the estimated total for the PSU i. The simplified variance estimator is then given by

$$\hat{V}_0=\frac{n}{(n-1)}\sum_{i\in A_I}\left(\frac{\hat{Y}_i}{\pi_{Ii}}-\frac{1}{n}\hat{Y}_{HT}\right)^2.$$

Under stratified multistage cluster sampling, the simplified variance estimator can be written

$$\hat{V}_0 = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(w_{hi} \hat{Y}_{hi} - \frac{1}{n_h} \sum_{j=1}^{n_h} w_{hj} \hat{Y}_{hj} \right)^2 \tag{10.4}$$

where w_{hi} is the sampling weight for cluster i in stratum h.

10.2 Linearization approach to variance estimation

When the point estimator is a nonlinear estimator such as ratio estimator or regression estimator, exact variance estimation for such estimates is very difficult. Instead, we often rely on linearization methods for estimating the sampling variance of such estimators.

Roughly speaking, if $\mathbf{y}$ is a p-dimensional vector and when $\bar{\mathbf{y}}_n = \bar{\mathbf{Y}}_N + O_p\left(n^{-1/2}\right)$ holds, the Taylor linearization of $g\left(\bar{\mathbf{y}}_n\right)$ can be written as

$$g\left(\bar{\mathbf{y}}_n\right) = g\left(\bar{\mathbf{Y}}\right) + \sum_{j=1}^{p} \frac{\partial g\left(\bar{\mathbf{Y}}\right)}{\partial y_j} \left(\bar{y}_j - \bar{Y}_j\right) + O_p\left(n^{-1}\right).$$

Thus, the variance of the linearized term of $g\left(\bar{\mathbf{y}}_n\right)$ can be written

$$V\left\{g\left(\bar{\mathbf{y}}_n\right)\right\} \doteq \sum_{i=1}^{p} \sum_{j=1}^{p} \frac{\partial g\left(\bar{\mathbf{Y}}\right)}{\partial y_i} \frac{\partial g\left(\bar{\mathbf{Y}}\right)}{\partial y_j} Cov\left\{\bar{y}_i, \bar{y}_j\right\}$$

and it is estimated by

$$\hat{V}\left\{g\left(\bar{\mathbf{y}}_n\right)\right\} \doteq \sum_{i=1}^{p} \sum_{j=1}^{p} \frac{\partial g\left(\bar{\mathbf{y}}_n\right)}{\partial y_i} \frac{\partial g\left(\bar{\mathbf{y}}_n\right)}{\partial y_j} \hat{C}\left\{\bar{y}_i, \bar{y}_j\right\}. \tag{10.5}$$

In summary, the linearization method for variance estimation can be described as follows:

1. Apply Taylor linearization and ignore the remainder terms.

2. Compute the variance and covariance terms of each component of $\bar{y}_n$ using the standard variance estimation formula.

3. Estimate the partial derivative terms ($\partial g / \partial y$) from the sample observation.

Example 10.2. *If the parameter of interest is $R = \bar{Y}/\bar{X}$ and we use*

$$\hat{R} = \frac{\bar{Y}_{HT}}{\bar{X}_{HT}}$$

to estimate R, we can apply the Taylor linearization to get

$$\hat{R} = R + \bar{X}^{-1}\left(\bar{Y}_{HT} - R\bar{X}_{HT}\right) + O_p\left(n^{-1}\right)$$

and the variance estimation formula in (10.5) reduces to

$$\hat{V}\left(\hat{R}\right) \doteq \bar{X}_{HT}^{-2}\hat{V}\left(\bar{Y}_{HT}\right) + \bar{X}_{HT}^{-2}\hat{R}^2\hat{V}\left(\bar{X}_{HT}\right) - 2\bar{X}_{HT}^{-2}\hat{R}\hat{C}\left(\bar{X}_{HT}, \bar{Y}_{HT}\right). \tag{10.6}$$

If the variances and covariances of $\bar{X}_{HT}$ and $\bar{Y}_{HT}$ are estimated by HT variance estimation formula, (10.6) can be estimated by

$$\hat{V}\left(\hat{R}\right) \doteq \frac{1}{\hat{X}_{HT}^2}\sum_{i\in A}\sum_{j\in A}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}\frac{e_i}{\pi_i}\frac{e_j}{\pi_j}$$

where $e_i = y_i - \hat{R}x_i$.

Using the result in Example 10.2, the variance of the ratio estimator $\hat{Y}_r = X\hat{R}$ is estimated by

$$\hat{V}\left(\hat{Y}_r\right) \doteq \left(\frac{X}{\hat{X}_{HT}}\right)^2\sum_{i\in A}\sum_{j\in A}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}\frac{e_i}{\pi_i}\frac{e_j}{\pi_j}, \tag{10.7}$$

which is obtained by multiplying $\hat{X}_{HT}^{-2}X^2$ to the variance formula in (8.13). Recall that the linearized variance estimator of the ratio estimator is given by (8.13). The variance estimator in (10.7) is asymptotically equivalent to the linearization variance estimator in (8.13), but it give more adequate measure of the conditional variance of the ratio estimator, as advocated by Royall and Cumberland (1981). More generally, Särndal, Swensson, and Wretman (1989) proposed using

$$\hat{V}\left(\hat{Y}_{GREG}\right) = \sum_{i\in A}\sum_{j\in A}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}\frac{g_ie_i}{\pi_i}\frac{g_je_j}{\pi_j} \tag{10.8}$$

as the conditional variance estimator of the GREG estimator of the form $\hat{Y}_{GREG} = \sum_{i\in A}\pi_i^{-1}g_iy_i$, where

$$g_i = \mathbf{X}'\left(\sum_{i\in A}\frac{1}{\pi_i c_i}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\frac{1}{c_i}\mathbf{x}_i \tag{10.9}$$

and

$$e_i = y_i - \mathbf{x}_i' \left(\sum_{i \in A} \frac{1}{\pi_i c_i} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i \in A} \frac{1}{\pi_i c_i} \mathbf{x}_i y_i.$$

The g_i in (10.9) is the factor to applied to the design weight to satisfy the calibration constraint.

Example 10.3. *We now consider variance estimation of the poststratification estimator in (8.28). To estimate the variance, the unconditional variance estimator is given by*

$$\hat{V}_u = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) \sum_{g=1}^{G} \frac{n_g - 1}{n - 1} s_g^2 \tag{10.10}$$

where $s_g^2 = (n_g - 1)^{-1} \sum_{i \in A_g} (y_i - \bar{y}_g)^2$. *On the other hand, the conditional variance estimator in (10.8) is given by*

$$\hat{V}_c = \left(1 - \frac{n}{N}\right) \frac{n}{n-1} \sum_{g=1}^{G} \frac{N_g^2}{n_g} \frac{n_g - 1}{n_g} s_g^2. \tag{10.11}$$

Note that the conditional variance formula in (10.11) is very similar to the variance estimation formula under the stratified sampling.

10.3 Replication approach to variance estimation

We now consider an alternative approach to variance estimation that is based on several replicates of the original sample estimator. Such replication approach does not use Taylor linearization, but instead generates several resamples and compute a replicate to each resample. The variability among the replicates is used to estimate the sampling variance of the point estimator. Such replication approach often uses repeated computation of the replicates using a computer. Replication methods include random group method, jackknife, balanced repeated replication, and bootstrap method. More details of the replication methods can be found in Wolter (2007).

10.3.1 Random group method

Random group method was first used in jute acreage surveys in Bengal (Mahalanobis; 1939). Random group method is useful in understanding the basic idea for replication approach to variance estimation. In the random group method, G random groups are constructed from the sample and the point estimate is computed for each random group sample and then combined to get the final point estimate. Once the final point estimate is constructed, its variance estimate is computed using the variability of the G random group estimates. There are two versions of the random group method. One is independent random group method and the other is non-independent random group method. We first consider independent random group method.

The independent random group method can be described as follows.

[Step 1] Using the given sampling design, select the first random group sample, denoted by $A_{(1)}$, and compute $\hat{\theta}_{(1)}$, an unbiased estimator of θ from the first random group sample.

[Step 2] Using the same sampling design, select the second random group sample, independently from the first random group sample, and compute $\hat{\theta}_{(2)}$ from the second random group sample.

[Step 3] Using the same procedure, obtain G independent estimate $\hat{\theta}_{(1)}, \cdots, \hat{\theta}_{(G)}$ from the G random group sample.

[Step 4] The final estimator of θ is

$$\hat{\theta}_{RG} = \frac{1}{G} \sum_{k=1}^{G} \hat{\theta}_{(k)} \tag{10.12}$$

and its unbiased variance estimator is

$$\hat{V}\left(\hat{\theta}_{RG}\right) = \frac{1}{G} \frac{1}{G-1} \sum_{k=1}^{G} \left(\hat{\theta}_{(k)} - \hat{\theta}_{RG}\right)^2 . \tag{10.13}$$

Since $\hat{\theta}_{(1)}, \cdots, \hat{\theta}_{(G)}$ are independently and identically distributed, the variance estimator in (10.13) is unbiased for the variance of $\hat{\theta}_{RG}$ in (10.12). Such independent random group method is very easy to understand and is applicable for complicated sampling designs for selecting $A_{(g)}$, but the sample allows for duplication of sample elements and the variance estimator may be unstable when G is small.

We now consider non-independent random group method, which does not allow for duplication of the sample elements. In the non-independent random group method, the sample is partitioned into G groups, exhaustive and mutually exclusive, denoted by $A = \cup_{g=1}^{G} A_{(g)}$ and then apply the sample estimation method for independent random group method, by treating the non-independent random group samples as if they are independent. The following theorem expresses the bias of the variance estimator for this case.

Theorem 10.2. *Let $\hat{\theta}_{(g)}$ be an unbiased estimator of θ, computed from the g-th random group sample $A_{(g)}$ for $g = 1, \cdots, G$. Then the random group variance estimator (**??**) satisfies*

$$E\left\{\hat{V}\left(\hat{\theta}_{RG}\right)\right\} - V\left(\hat{\theta}_{RG}\right) = -\frac{1}{G(G-1)} \sum\sum_{i \neq j} Cov\left(\hat{\theta}_{(i)}, \hat{\theta}_{(j)}\right) \tag{10.14}$$

Proof. By (10.12), we have

$$V\left(\hat{\theta}_{RG}\right) = \frac{1}{G^2}\left\{\sum_{i=1}^{G} V\left(\hat{\theta}_{(i)}\right) + \sum\sum_{i \neq j} Cov\left(\hat{\theta}_{(i)}, \hat{\theta}_{(j)}\right)\right\} \tag{10.15}$$

and, since

$$\sum_{i=1}^{G}\left(\hat{\theta}_{(i)} - \hat{\theta}_{RG}\right)^2 = \sum_{i=1}^{G}\left(\hat{\theta}_{(i)} - \theta\right)^2 - G\left(\hat{\theta}_{RG} - \theta\right)^2$$

and using $E\left(\hat{\theta}_{(i)}\right) = \theta$,

$$\begin{aligned}
E\left\{\sum_{i=1}^{G}\left(\hat{\theta}_{(i)} - \hat{\theta}_{RG}\right)^2\right\} &= \sum_{i=1}^{G} V\left(\hat{\theta}_{(i)}\right) - G \times V\left(\hat{\theta}_{RG}\right) \\
&= \left(1 - G^{-1}\right)\sum_{i=1}^{G} V\left(\hat{\theta}_{(i)}\right) - G^{-1}\sum\sum_{i \neq j} Cov\left(\hat{\theta}_{(i)}, \hat{\theta}_{(j)}\right)
\end{aligned}$$

which implies

$$E\left\{\hat{V}\left(\hat{\theta}_{RG}\right)\right\} = G^{-2}\sum_{i=1}^{G} V\left(\hat{\theta}_{(i)}\right) - G^{-2}\left(G-1\right)^{-1}\sum\sum_{i\neq j} Cov\left(\hat{\theta}_{(i)},\hat{\theta}_{(j)}\right).$$

Thus, using (10.15), we have (10.14). □

The right side of (10.14) is the bias of $\hat{V}\left(\hat{\theta}_{RG}\right)$ as an estimator for the variance of $\hat{\theta}_{RG}$. Such bias becomes zero if the sampling for random groups is a with-replacement sampling. For without-replacement sampling, the covariance between the two different replicates is negative and so the bias term becomes positive. The relative amount of the bias is often negligible. The following example compute the amount of the relative bias.

Example 10.4. *Consider a sample of size n obtained from the simple random sampling from a finite population of size N. Let $b = n/G$ be an integer value which is the size of $A_{(g)}$ such that $A = \cup_{g=1}^{G} A_{(g)}$. The sample mean of y obtained from $A_{(g)}$ is denoted by $\bar{y}_{(g)}$ and the overall mean of y is given by*

$$\hat{\theta} = \frac{1}{G}\sum_{g=1}^{G} \bar{y}_{(g)}.$$

In this case, $\bar{y}_{(1)},\cdots,\bar{y}_{(G)}$ are not independent distributed but identically distributed with the same mean. By (10.15), we have

$$Var\left(\hat{\theta}\right) = \frac{1}{G}V\left(\bar{y}_{(1)}\right) + \left(1-\frac{1}{G}\right)Cov\left(\bar{y}_{(1)},\bar{y}_{(2)}\right)$$

and since

$$V\left(\bar{y}_{(1)}\right) = \left(\frac{1}{b}-\frac{1}{N}\right)S^2,$$

we have

$$Cov\left(\bar{y}_{(1)},\bar{y}_{(2)}\right) = -\frac{1}{N}S^2.$$

Thus, the bias in (10.14) reduces to

$$\begin{aligned} Bias\left\{\hat{V}\left(\hat{\theta}_{RG}\right)\right\} &= -Cov\left(\bar{y}_{(1)},\bar{y}_{(2)}\right) \\ &= \frac{1}{N}S^2 \end{aligned}$$

which is often negligible. Thus, the random group variance estimator (10.13) can be safely used to estimate the variance of $\bar{y}_n$ in the simple random sampling.

Such random group method provides a useful computational tool for estimating the variance of the point estimators. However, the random group method is applicable only when the sampling design for $A_{(g)}$ has the same structure as the original sample A. The partition $A = \cup_{g=1}^{G} A_{(g)}$ that leads to unbiasedness of $\hat{\theta}_{(g)}$ is not always possible for complex sampling designs.

10.4 Jackknife method

Jackknife was first introduced by Quenouille (1949) to reduce the bias of ratio estimator and then was suggested by Tukey (1958) to be used for variance estimation. Jackknife is very popular in practice as a tool for variance estimation.

To introduce the idea of Quenouille (1949), suppose that n independent observations of (x_i, y_i) are available that are generated from a distribution with mean (μ_x, μ_y). If the parameter of interest is $\theta = \mu_y/\mu_x$, then the sample ratio $\hat{\theta} = \bar{x}^{-1}\bar{y}$ has bias of order $O\left(n^{-1}\right)$. That is, we have

$$E\left(\hat{\theta}\right) = \theta + \frac{C}{n} + O\left(n^{-2}\right).$$

If we delete the k-th observation and recompute the ratio

$$\hat{\theta}^{(-k)} = \left(\sum_{i \neq k} x_i\right)^{-1} \sum_{i \neq k} y_i,$$

we obtain

$$E\left(\hat{\theta}^{(-k)}\right) = \theta + \frac{C}{n-1} + O\left(n^{-2}\right).$$

Thus, the jackknife pseudo value defined by $\hat{\theta}_{(k)} = n\hat{\theta} - (n-1)\,\hat{\theta}^{(-k)}$ can be used to compute

$$\hat{\theta}_{(.)} = \frac{1}{n}\sum_{k=1}^{n} \hat{\theta}_{(k)} \tag{10.16}$$

which has bias of order $O\left(n^{-2}\right)$. Thus, the jackknife can be used to reduce the bias of nonlinear estimators.

Note that if $\hat{\theta} = \bar{y}$ then $\hat{\theta}_{(k)} = y_k$. Tukey (1958) suggested using $\hat{\theta}_{(k)}$ as approximate IID observation to get the following jackknife variance estimator.

$$\begin{aligned} \hat{V}_{JK}\left(\hat{\theta}\right) &\doteq \frac{1}{n}\frac{1}{n-1}\sum_{k=1}^{n}\left(\hat{\theta}_{(k)} - \bar{\theta}_{(.)}\right)^2 \\ &= \frac{n-1}{n}\sum_{k=1}^{n}\left(\hat{\theta}^{(-k)} - \bar{\theta}^{(.)}\right)^2 \end{aligned}$$

For the special case of $\hat{\theta}_n = \bar{y}$, we obtain

$$\hat{V}_{JK} = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \frac{1}{n}s_y^2$$

and the jackknife variance estimator is algebraically equivalent to the usual variance estimator of the sample mean under simple random sampling ignoring the finite population correction term.

If we are interested in estimating the variance of $\hat{\theta} = f(\bar{x}, \bar{y})$, using $\hat{\theta}^{(-k)} = f\left(\bar{x}^{(-k)}, \bar{y}^{(-k)}\right)$, we can construct

$$\hat{V}_{JK}\left(\hat{\theta}\right) = \frac{n-1}{n}\sum_{k=1}^{n}\left(\hat{\theta}^{(-k)} - \hat{\theta}\right)^2 \tag{10.17}$$

as the jackknife variance estimator of $\hat{\theta}$. The jackknife replicate $\hat{\theta}^{(-k)}$ is computed by using the same formula for $\hat{\theta}$ using the jackknife weight $w_i^{(k)}$ instead of the original weight $w_i = 1/n$, where

$$w_i^{(-k)} = \begin{cases} (n-1)^{-1} & \text{if } i \neq k \\ 0 & \text{if } i = k. \end{cases}$$

To discuss the asymptotic property of the jackknife variance estimator in (10.17), we use the following Taylor expansion, which is often called second-type Taylor expansion.

Lemma 10.1. *Let $\{X_n, W_n\}$ be a sequence of random variables such that*

$$X_n = W_n + O_p(r_n)$$

where $r_n \to 0$ as $n \to \infty$. If $g(x)$ is a function with s-th continuous derivatives in the line segment joining X_n and W_n and the s-th order partial derivatives are bounded, then

$$g(X_n) = g(W_n) + \sum_{k=1}^{s-1} \frac{1}{k!} g^{(k)}(W_n)(X_n - W_n)^k + O_p(r_n^s)$$

where $g^{(k)}(a)$ is the k-th derivative of $g(x)$ evaluated at $x = a$.

Now, since $\bar{y}^{(-k)} - \bar{y} = (n-1)^{-1}(\bar{y} - y_k)$, we have

$$\bar{y}^{(-k)} - \bar{y} = O_p\left(n^{-1}\right).$$

For the case of $\hat{\theta} = f(\bar{x}, \bar{y})$, we can apply the above lemma to get

$$\begin{aligned} \hat{\theta}^{(-k)} - \hat{\theta} &= \frac{\partial f}{\partial x}(\bar{x}, \bar{y})\left(\bar{x}^{(-k)} - \bar{x}\right) \\ &\quad + \frac{\partial f}{\partial y}(\bar{x}, \bar{y})\left(\bar{y}^{(-k)} - \bar{y}\right) + o_p\left(n^{-1}\right) \end{aligned}$$

so that

$$\begin{aligned} &\frac{n-1}{n} \sum_{k=1}^{n} \left(\hat{\theta}^{(-k)} - \hat{\theta}\right)^2 \\ = &\left\{\frac{\partial f}{\partial x}(\bar{x}, \bar{y})\right\}^2 \frac{n-1}{n} \sum_{k=1}^{n} \left(\bar{x}^{(-k)} - \bar{x}\right)^2 + \left\{\frac{\partial f}{\partial y}(\bar{x}, \bar{y})\right\}^2 \frac{n-1}{n} \sum_{k=1}^{n} \left(\bar{y}^{(-k)} - \bar{y}\right)^2 \\ &+2\left\{\frac{\partial f}{\partial x}(\bar{x}, \bar{y})\right\}\left\{\frac{\partial f}{\partial y}(\bar{x}, \bar{y})\right\} \frac{n-1}{n} \sum_{k=1}^{n} \left(\bar{x}^{(-k)} - \bar{x}\right)\left(\bar{y}^{(-k)} - \bar{y}\right) + o_p\left(n^{-1}\right). \end{aligned}$$

Thus, the jackknife variance estimator is asymptotically equivalent to the linearized variance estimator. That is, the second-type Taylor linearization leads to

$$\begin{aligned} \hat{V}_{JK}\left(\hat{\theta}\right) &= \left\{\frac{\partial f}{\partial x}(\bar{x}, \bar{y})\right\}^2 \hat{V}(\bar{x}) + \left\{\frac{\partial f}{\partial y}(\bar{x}, \bar{y})\right\}^2 \hat{V}(\bar{y}) \\ &\quad +2\left\{\frac{\partial f}{\partial x}(\bar{x}, \bar{y})\right\}\left\{\frac{\partial f}{\partial y}(\bar{x}, \bar{y})\right\} \hat{C}(\bar{x}, \bar{y}) + o_p\left(n^{-1}\right). \end{aligned}$$

The above jackknife method is constructed under simple random sampling. For stratified multistage sampling, jackknife replicates can be constructed by deleting a cluster for each replicate. Let

$$\hat{Y}_{HT} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi} \hat{Y}_{hi}$$

be the HT estimator of Y under the stratified multistage cluster sampling. The jackknife weights are constructed by deleting a cluster sequentially as

$$w_{hi}^{(-gj)} = \begin{cases} 0 & \text{if } h = g \text{ and } i = j \\ (n_h - 1)^{-1} n_h w_{hi} & \text{if } h = g \text{ and } i \neq j \\ w_{hi} & \text{otherwise} \end{cases}$$

and the jackknife variance estimator is computed by

$$\hat{V}_{JK}\left(\hat{Y}_{HT}\right) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left(\hat{Y}_{HT}^{(-hi)} - \frac{1}{n_h} \sum_{j=1}^{n_h} \hat{Y}_{HT}^{(-hj)} \right)^2 \tag{10.18}$$

where

$$\hat{Y}_{HT}^{(-gj)} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi}^{(-gj)} \hat{Y}_{hi}.$$

The following theorem presents the algebraic property of the jackknife variance estimator in (10.18).

Theorem 10.3. *The jackknife variance estimator in (10.18) is algebraically equivalent to the simplified variance estimator in (10.4).*

If $\hat{\theta} = f\left(\hat{X}_{HT}, \hat{Y}_{HT}\right)$ then the jackknife replicates are constructed as $\hat{\theta}^{(-gj)} = f\left(\hat{X}_{HT}^{(-gj)}, \hat{Y}_{HT}^{(-gj)}\right)$. In this case, the jackknife variance estimator is computed as

$$\hat{V}_{JK}\left(\hat{\theta}\right) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left(\hat{\theta}^{(-hi)} - \frac{1}{n_h} \sum_{j=1}^{n_h} \hat{\theta}^{(-hj)} \right)^2 \tag{10.19}$$

or, more simply,

$$\hat{V}_{JK}\left(\hat{\theta}\right) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left(\hat{\theta}^{(-hi)} - \hat{\theta} \right)^2. \tag{10.20}$$

The asymptotic properties of the above jackknife variance estimator under stratified multistage sampling can be established. For references, see Krewski and Rao (1981) or Shao and Tu (1995).

Example 10.5. *We now revisit Example 10.3 to estimate the variance of poststratification estimator using the jackknife. For simplicity, assume SRS for the sample.*

The poststratification estimator is computed as

$$\hat{Y}_{post} = \sum_{g=1}^{G} N_g \bar{y}_g$$

where $\bar{y}_g = n_g^{-1} \sum_{i \in A_g} y_i$. *Now, the k-th replicate of* $\hat{Y}_{post}$ *is*

$$\hat{Y}_{post}^{(-k)} = \sum_{g \neq h} N_g \bar{y}_g + N_h \bar{y}_h^{(-k)}$$

where $\bar{y}_h^{(-k)} = (n_h - 1)^{-1} (n_h \bar{y}_h - y_k)$. *Thus,*

$$\begin{aligned} \hat{Y}_{post}^{(-k)} - \hat{Y}_{post} &= N_h \left(\bar{y}_h^{(-k)} - \bar{y}_h \right) \\ &= N_h (n_h - 1)^{-1} (\bar{y}_h - y_k) \end{aligned}$$

and

$$\begin{aligned} \hat{V}_{JK}\left(\hat{Y}_{post}\right) &= \frac{n-1}{n} \sum_{k=1}^{n} \left(\hat{Y}_{post}^{(-k)} - \hat{Y}_{post} \right)^2 \\ &= \frac{n-1}{n} \sum_{g=1}^{G} N_g^2 (n_g - 1)^{-1} s_g^2, \end{aligned}$$

which, ignoring n/N *term, is asymptotically equivalent to the conditional variance estimator in (10.11).*

10.5 Other methods

(Skip)

Reference

Krewski, D. and Rao, J.N.K. (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics* **9**, 1010-1019.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353-360.

Royall, R. M. and Cumberland, W. G. (1981). An empirical study of the ratio estimator and estimators of its variance. *Journal of the American Statistical Association* **76**, 66-77.

Särndal, C. E., Swensson, B., and Wretman, J. H. (1989). The weighted residual technique for estimation the variance of the general regression estimator of the finite population total, *Biometrika* **76**, 527-537.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples (abstract). *Annals of Mathematical Statistics* **29**, 614.

Wolter, K. M. (2007). *Introduction to Variance Estimation: Second Edition*. New York: Springer-Verlag.