Jackknife variance estimation for functions of Horvitz & Thompson estimators under unequal probability sampling without replacement

Escobar, Emilio L.
University of Southampton, Social Statistics
Southampton, SO17 1BJ, United Kingdom.
E-mail: Emilio.Lopez-Escobar@soton.ac.uk

Berger, Yves G.
University of Southampton, Southampton Statistical Sciences Research Institute
Southampton, SO17 1BJ, United Kingdom.
E-mail: Y.G.Berger@soton.ac.uk

The jackknife is a popular method in survey sampling which is widely used for standard error estimation (e.g. Shao & Tu (1995) and Wolter (2007)). The applicability and theoretical properties of jackknife variance estimators under unequal probabilities without-replacement sampling have been studied to a limited extent. Some examples are given by Campbell (1980), Berger & Skinner (2005), Berger & Rao (2006) and Berger (2007), who proposed jackknife variance estimators for functions of Hájek (1971) point estimators.

We propose two generalised jackknife variance estimators suitable for functions of Horvitz & Thompson (1952) point estimators. Regularity conditions under which the proposed estimators are design-consistent are also provided. These estimators are defined for without-replacement unequal-probability sampling designs and they naturally include finite population corrections. The proposed estimators are compared with the Campbell (1980) jackknife variance estimator for a ratio.

The class of point estimators

Let $\mathcal{U} = \{1, \ldots, k, l, \ldots, N\}$ denote a finite population and let $s = \{1, \ldots, n\}$ denote a sample whose elements are randomly selected with an unequal probability sampling design without replacement, $s \subseteq \mathcal{U}$, $n \leq N$. Assume that we are interested in the population parameter $\theta = h(t_1, \ldots, t_q, \ldots, t_Q)$ which is a function of population totals from Q survey variables, where $h(\cdot)$ is a smooth and differentiable function (e.g. Shao & Tu (1995), Chapter 2), $t_q = \sum_{k\in\mathcal{U}} y_{qk}$ with y_{qk} denoting the measurement of the q-th variable for unit $k \in \mathcal{U}$, $q = 1, \ldots, Q$. Further, assume we estimate θ by the substitution point estimator $\widehat{\theta} = h(\widehat{t}_1, \ldots, \widehat{t}_q, \ldots, \widehat{t}_Q)$ where $\widehat{t}_q = \sum_{k\in s} w_k y_{qk}$ is the Horvitz & Thompson (1952) point estimator of t_q, with survey weights $w_k = \pi_k^{-1}$ where π_k denotes the inclusion probability of unit k, $\pi_k > 0$, $\forall k \in \mathcal{U}$ and π_{kl} denotes the joint inclusion probabilities of units k and l, $\pi_{kl} > 0, \forall k, l \in \mathcal{U}$.

The proposed variance estimator

We propose to estimate the variance of $\widehat{\theta}$ by the jackknife variance estimator

$$v_{JHT} = \sum\sum_{(k,l)\in s} \mathcal{D}_{kl}\ \nu_k\ \nu_l, \tag{1}$$

with

$$\nu_k = w_k(\widehat{\theta} - \widehat{\theta}^{(k)}), \tag{2}$$

where $\mathcal{D}_{kl} = \pi_{kl}^{-1}\{\pi_{kl} - \pi_k\pi_l\}$, and $\widehat{\theta}^{(k)} = h(\widehat{t}_1^{(k)}, \ldots, \widehat{t}_q^{(k)}, \ldots, \widehat{t}_Q^{(k)})$ with

$$\widehat{t}_q^{(k)} = \sum_{(l\neq k)\in s} w_l y_{ql} + (w_k - 1)y_{qk}. \tag{3}$$

Alternatively, if the sampling design is of fixed sample size, we propose to estimate $\text{var}(\widehat{\theta})$ by

$$v_{JSYG} = -\frac{1}{2}\sum\sum_{(k,l)\in s} \mathcal{D}_{kl}\,(\nu_k - \nu_l)^2, \tag{4}$$

which is always positive provided $\mathcal{D}_{kl} < 0$ (e.g. Chao (1982)).

For the simplest case where $\widehat{\theta} = \widehat{t} = \sum_{k\in s} w_k y_k$, (2) and (3) imply $\nu_k = w_k(\widehat{t} - \widehat{t}^{(k)}) = w_k(\widehat{t} - (\widehat{t} - y_k)) = w_k y_k$. Hence, the proposed jackknifes (1) and (4) reduce, respectively, to the Horvitz & Thompson (1952), and the Sen (1953) and Yates & Grundy (1953) unbiased estimators of $\text{var}(\widehat{t})$.

Design-consistency

The design-consistency is set under the Isaki and Fuller (1982) asymptotic framework. Accordingly, consider a sequence of nested populations of size $\{N_t : 0 < N_t < N_{t+1}, \forall t\}$. Consider also a sequence of (non-necessarily nested) samples of size $\{n_t : n_t < n_{t+1}; n_t < N_t, \forall t\}$. Thus, $t \to \infty$ implies $N_t \to \infty$ and $n_t \to \infty$, with constant $f = n_t/N_t$. In what follows, the index t is dropped to simplify the notation.

In asymptotic studies, it is convenient to work with means instead of totals. Hence, re-define the weights w_k as $\tilde{w}_k = w_k/N, \forall k \in \mathcal{U}$, such that $\widehat{t}_q$ becomes the mean estimator $\tilde{\mu}_q = \sum_{k\in s} \tilde{w}_k y_{qk}$ for the population mean $\mu_q = t_q/N$, $q = 1, \ldots, Q$. Now, recall Results 2.8.1 and 2.8.2 from Särndal *et al.* (1992) and denote by $\mathbf{\Sigma}_{HT} = \sum\sum_{(k,l)\in\mathcal{U}} \mathcal{D}_{kl}\pi_{kl}\tilde{w}_k\tilde{w}_l \boldsymbol{y}_k\boldsymbol{y}_l^T$, $\widehat{\mathbf{\Sigma}}_{HT} = \sum\sum_{(k,l)\in s} \mathcal{D}_{kl}\tilde{w}_k\tilde{w}_l\boldsymbol{y}_k\boldsymbol{y}_l^T$, $\mathbf{\Sigma}_{SYG} = -\frac{1}{2}\sum\sum_{(k,l)\in\mathcal{U}} \mathcal{D}_{kl}\pi_{kl}\{\tilde{w}_k\boldsymbol{y}_k - \tilde{w}_l\boldsymbol{y}_l\}\{\tilde{w}_k\boldsymbol{y}_k - \tilde{w}_l\boldsymbol{y}_l\}^T$ and $\widehat{\mathbf{\Sigma}}_{SYG} = -\frac{1}{2}\sum\sum_{(k,l)\in s}\mathcal{D}_{kl}\{\tilde{w}_k\boldsymbol{y}_k - \tilde{w}_l\boldsymbol{y}_l\}\{\tilde{w}_k\boldsymbol{y}_k - \tilde{w}_l\boldsymbol{y}_l\}^T$ the multivariate Horvitz-Thompson and Sen-Yates-Grundy population and sample variances of the estimator $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_Q)^T$ for the population parameter $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_Q)^T$, with $\pi_{kl} > 0, \forall k, l \in \mathcal{U}$ where $\mathcal{D}_{kl} = \pi_{kl}^{-1}\{\pi_{kl} - \pi_k\pi_l\}$ and $\boldsymbol{y}_k = (y_{1k}, \ldots, y_{Qk})^T$. Consider the following regularity conditions:

(a) $v_L/V_L \to_p 1$, $V_L \neq 0$ with $V_L = \boldsymbol{\nabla}(\boldsymbol{\mu})^T\mathbf{\Sigma}_{HT}\boldsymbol{\nabla}(\boldsymbol{\mu})$, $v_L = \boldsymbol{\nabla}(\tilde{\boldsymbol{\mu}})^T\widehat{\mathbf{\Sigma}}_{HT}\boldsymbol{\nabla}(\tilde{\boldsymbol{\mu}})$ (for fixed sample size designs: $V_L = \boldsymbol{\nabla}(\boldsymbol{\mu})^T\mathbf{\Sigma}_{SYG}\boldsymbol{\nabla}(\boldsymbol{\mu})$, $v_L = \boldsymbol{\nabla}(\tilde{\boldsymbol{\mu}})^T\widehat{\mathbf{\Sigma}}_{SYG}\boldsymbol{\nabla}(\tilde{\boldsymbol{\mu}})$), where $\boldsymbol{\nabla}(\boldsymbol{x})$ is the gradient of $h(\cdot)$ at $\boldsymbol{x} \in \Re^Q$, $\boldsymbol{\nabla}(\boldsymbol{x}) = (\partial h(\boldsymbol{\mu})/\partial\mu_1, \ldots, \partial h(\boldsymbol{\mu})/\partial\mu_Q)^T_{\boldsymbol{\mu}=\boldsymbol{x}}$, $h(\cdot)$ is continuous and differentiable at $\boldsymbol{\mu}$.

(b) $\lim\inf\{n\ V_L\} > 0$.

(c) $n^{-1}\sum_{k\in s}\tilde{w}_k^{\tau}\,\|\boldsymbol{y}_k\|^{\tau} = \mathcal{O}_p(n^{-\tau}), \forall\tau \geq 2$, with $\|\boldsymbol{A}\| = \text{tr}(\boldsymbol{A}^T\boldsymbol{A})^{1/2}$ the Euclidean norm.

(d) $G_s = n^{-\beta}\sum\sum_{(k\neq l)\in s}(\mathcal{D}_{kl}^-)^2 = \mathcal{O}_p(1)$, with $0 \leq \beta < 1$, $\mathcal{D}_{kl}^- = -\mathcal{D}_{kl}$ if $\mathcal{D}_{kl} < 0$, 0 otherwise.

(e) $H_s = n^{-\beta}\sum\sum_{(k\neq l)\in s}(\mathcal{D}_{kl}^+)^2 = \mathcal{O}_p(1)$, with $0 \leq \beta < 1$, $\mathcal{D}_{kl}^+ = \mathcal{D}_{kl}$ if $\mathcal{D}_{kl} \geq 0$, 0 otherwise.

(f) $\boldsymbol{\nabla}(\boldsymbol{x})$ is Lipschitz (Hölder) continuous of order δ, i.e. $\|\boldsymbol{\nabla}(\boldsymbol{x}_1) - \boldsymbol{\nabla}(\boldsymbol{x}_2)\| \leq \lambda\,\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{\delta}$, $\lambda > 0$ constant, $\beta/2 < \delta \leq 1$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in neighbourhood of $\boldsymbol{\mu}$ (e.g. Shao and Tu (1995), page 43).

(g) $\|\boldsymbol{\nabla}(\tilde{\boldsymbol{\mu}})\| = \mathcal{O}_p(1)$.

Condition (a) sets the consistency of the linearisation variance estimator v_L for V_L (Särndal *et al.* (1992), Secs. 5.5 & 5.7)). Conditions (b) and (c) are typical (Shao & Tu (1995), pp. 258-260): (b) implies that V_L decreases with rate n^{-1} and (c) is a Lyapunov condition. Conditions (d) and (e) are mild requirements on the design, and (f) and (g) are usual smoothness conditions for jackknifes.

Theorem 1. *For fixed sample size designs, if regularity conditions (a)-(g) hold, then the variance estimator v_{JSYG} in (4) is asymptotically design-consistent for the approximate linearised variance $V_L \neq 0$, i.e. $v_{JSYG}/V_L \to_p 1$.*

Corollary 1. *If regularity conditions (a)-(g) hold, then the variance estimator v_{JHT} in (1) is asymptotically design-consistent for the approximate linearised variance $V_L \neq 0$, i.e. $v_{JHT}/V_L \to_p 1$.*

Corollary 2. *From Theorem 1, by Slutsky's theorem and asymptotic Normality of* $\widehat{\theta}$ *for* θ, *it follows* $\{v_{JHT}\}^{-1/2}(\widehat{\theta}-\theta) \rightarrow_d \boldsymbol{N}(0,1)$ *and* $\{v_{JSYG}\}^{-1/2}(\widehat{\theta}-\theta) \rightarrow_d \boldsymbol{N}(0,1)$. *Thus, both jackknife variance estimators,* v_{JHT} *and* v_{JSYG} *in (1) and (4), allow valid confidence intervals of* $\widehat{\theta}$ *for* θ.

Example: The ratio

We now illustrate how the proposed estimators works for the ratio point estimator. Let the parameter of interest be $R = t_y/t_x = \mu_y/\mu_x$, where $t_y = \sum_{k\in\mathcal{U}} y_k$ and $t_x = \sum_{k\in\mathcal{U}} x_k$ are the population totals of the variables y and x, and $\mu_y = t_y/N$ and $\mu_x = t_x/N$ are population means. Now, assume that R is estimated with the point estimator $\widehat{R} = \sum_{k\in s} w_k y_k / \sum_{k\in s} w_k x_k$, which can be thought either as a function of Horvitz-Thompson (1952) total estimators

$$\widehat{R} = \widehat{t}_y/\widehat{t}_x, \tag{5}$$

where $\widehat{t}_y = \sum_{k\in s} w_k y_k$, $\widehat{t}_x = \sum_{k\in s} w_k x_k$, or as a function of Hájek (1971) mean estimators

$$\widehat{R} = \breve{\mu}_y/\breve{\mu}_x, \tag{6}$$

where $\breve{\mu}_y = \widehat{t}_y/\widehat{N}$ and $\breve{\mu}_x = \widehat{t}_x/\widehat{N}$, with $\widehat{N} = \sum_{k\in s} w_k$. Hence, the proposed variance estimator v_{JHT} in (1) and the Campbell (1980) jackknife variance estimator, below in (7), are comparable as they estimate the variance of the same point estimator. From Berger & Skinner (2005), Campbell's estimator is defined as

$$v_{JC} = \sum\sum\nolimits_{(k,l)\in s} \mathcal{D}_{kl}\, \varepsilon_k\, \varepsilon_l, \tag{7}$$

where $\varepsilon_k = (1 - w_k/\widehat{N})(\breve{\theta} - \breve{\theta}^{(k)})$ and $\breve{\theta} = g(\breve{\mu}_1, \ldots, \breve{\mu}_p, \ldots, \breve{\mu}_P)$ is a function of Hájek (1971) mean estimators from P variables with $\breve{\mu}_p = \widehat{t}_p/\widehat{N}$, and where $\breve{\theta}^{(k)} = g(\breve{\mu}_1^{(k)}, \ldots, \breve{\mu}_p^{(k)}, \ldots, \breve{\mu}_P^{(k)})$ has the same functional form as $\breve{\theta}$ but using $\breve{\mu}_p^{(k)} = (\widehat{t}_p - w_k y_k)/(\widehat{N} - w_k)$ instead of $\breve{\mu}_p$.

It is well known (e.g. Särndal et al. (1992), Result 5.6.2), that the approximate linearised variance of $\widehat{R}$ is given by $V_L = \sum\sum_{(k,l)\in\mathcal{U}} \mathcal{D}_{kl}\pi_{kl}u_k u_l$ where $u_k = w_k(y_k - Rx_k)/t_x$. Besides, it is also known that an unbiased estimator of V_L is given by $v_L = \sum\sum_{(k,l)\in s}\mathcal{D}_{kl}\breve{u}_k\breve{u}_l$, where

$$\breve{u}_k = \frac{w_k}{\widehat{t}_x}\,(y_k - \widehat{R}x_k). \tag{8}$$

Henceforth, the quantities ν_k of the proposed jackknife variance estimator v_{JHT} in (1) are given by

$$\begin{aligned}
\nu_k &= w_k\left(\widehat{R} - \frac{\widehat{t}_y - y_k}{\widehat{t}_x - x_k}\right), \\
&= \frac{w_k}{\widehat{t}_x}\,(y_k - \widehat{R}x_k)\left(\frac{\widehat{t}_x}{\widehat{t}_x - x_k}\right), \\
&= \breve{u}_k\left(\frac{\widehat{t}_x}{\widehat{t}_x - x_k}\right), \qquad (9)
\end{aligned}$$

whereas the quantities ε_k of Campbell's jackknife v_{JC} in (7) are given by

$$\begin{aligned}
\varepsilon_k &= \left(1 - \frac{w_k}{\widehat{N}}\right)\left(\widehat{R} - \frac{\widehat{t}_y - w_k y_k}{\widehat{t}_x - w_k x_k}\right), \\
&= \frac{w_k}{\widehat{t}_x}\,(y_k - \widehat{R}x_k)\left(\frac{\widehat{t}_x}{\widehat{t}_x - w_k x_k}\right)\left(\frac{\widehat{N} - w_k}{\widehat{N}}\right), \\
&= \breve{u}_k\left(\frac{\widehat{t}_x}{\widehat{t}_x - w_k x_k}\right)\left(\frac{\widehat{N} - w_k}{\widehat{N}}\right). \qquad (10)
\end{aligned}$$

It can clearly be seen from (9) that $\nu_k \doteq \check{u}_k$ if $(\widehat{t}_x - x_k)^{-1}\widehat{t}_x \doteq 1$. On the other hand, from (10) we have that $\varepsilon_k \doteq \check{u}_k$ if $\widehat{N}^{-1}(\widehat{N} - w_k) \doteq 1$ and if $(\widehat{t}_x - w_k x_k)^{-1}\widehat{t}_x \doteq 1$.

Thus, the proposed jackknife estimator is a suitable approximation of the Linearisation variance estimator v_L. It is more accurate than Campbell's jackknife as (9) is less sensitive to (highly-skewed) weights than (10).

REFERENCES

Berger, Y.G. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities. Biometrika, 94, 953-964.

Berger, Y.G. & Rao, J.N.K. (2006). Adjusted jackknife for imputation under unequal probability sampling without replacement. J. R. Statist. Soc. B., 68, 531-547.

Berger, Y.G. & Skinner, C.J. (2005). A jackknife variance estimator for unequal probability sampling. J. R. Statist. Soc. B. 67, 1, 79-89.

Campbell, C. (1980). A different view of finite population estimation. Proc. Surv. Res. Meth. Sect. Am. Statist. Assoc. 319-324.

Chao, M.T. (1982). A general purpose unequal probability sampling plan. Biometrika. 69, 3, 653-656.

Hájek, J. (1971). Comment on a paper by Basu, D. in Foundations of Statistical Inference (Godambe, V.P. and Sprott, D.A. eds.). p. 236. Toronto: Holt, Rinehart and Winston.

Horvitz, D.G. & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J. Am. Statist. Assoc. 47, 663-685.

Isaki, C.T. & Fuller, W.A. (1982). Survey design under the regression superpopulation model. J. Am. Statist. Assoc. 77, 377, 89-96.

Miller, R.G. (1964). A trustworthy jackknife, Ann. Math. Statist. 35, 4, 1594-1605.

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

Robinson, P.M. & Särndal C.E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. Sankhyā: The Indian Journal of Statistics, B. 45, 2, 240-248.

Särndal, C.-E., Swensson, B. & Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer.

Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. J. Indian Soc. Agr. Statist. 5, 119-127.

Shao, J. & Tu, D. (1995). The Jackknife and Bootstrap. New York: Springer.

Tukey, J.W. (1958). Bias and confidence in not-quite large samples (abst.). Ann. Math. Statist. 29, 2, 614.

Tillé, Y. (2006). Sampling Algorithms. New York: Springer.

Valliant, R., Dorfman, A.H. & Royall, R.M. (2000). Finite Population Sampling and Inference: A Prediction Approach. New York: Wiley.

Wolter, K.M. (2007). Introduction to Variance Estimation. 2nd Ed. New York: Springer.

Yates, F. & Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. J. R. Statist. Soc. B. 15, 253-261.