The Canadian Journal of Statistics
Vol. 41, No. 3, 2013, Pages 508–524
La revue canadienne de statistique

A new replicate variance estimator for unequal probability sampling without replacement

Emilio L. ESCOBAR[1]* and Yves G. BERGER[2]

[1]*Statistics Department, ITAM, Mexico City, Distrito Federal, Mexico*
[2]*Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, U.K.*

Key words and phrases: Gâteaux derivative; jackknife; pseudo-value; stratification; Taylor linearization.

MSC 2010: Primary 62D05; secondary 62F40.

Abstract: We propose a new replicate variance estimator suitable for differentiable functions of estimated totals. The proposed variance estimator is defined for any unequal-probability without-replacement sampling design, it naturally includes finite population corrections and it allows two-stage sampling. We show its design-consistency and its close relationship with linearization variance estimators. When estimating a total, the proposed estimator reduces to the Horvitz–Thompson variance estimator. Simulations suggest that the proposed variance estimator is more stable than its replicate competitors. *The Canadian Journal of Statistics* 41: 508–524; 2013 © 2013 Statistical Society of Canada

Résumé: Nous proposons un nouvel estimateur de variance ré-échantillonnés adapté à des fonctions dérivables de totaux estimés. L'estimateur de variance proposé est défini pour tous les plans d'échantillonnage à probabilités inégales sans remise. Il comprend naturellement les corrections de population finie et il peut s'appliquer à l'échantillonnage à deux degrés. Nous montrons sa convergence asymptotique sous le plan d'échantillonnage et sa relation avec les estimateurs linéarisés de variance. Lors de l'estimation d'un total, l'estimateur proposé se réduit à l'estimateur de variance de Horvitz-Thompson. Des simulations suggèrent que l'estimateur de variance proposée est plus stable que ses concurrents. *La revue canadienne de statistique* 41: 508–524; 2013 © 2013 Société statistique du Canada

1. INTRODUCTION

Replication methods for variance estimation such as the Jackknife, the Bootstrap and the Balanced Half-Sampling are very popular in practice (e.g., Shao & Tu, 1995; Davison & Hinkley, 1997; Lehtonen & Pahkinen, 2004; Wolter, 2007). However, there are only limited applications under unequal probability without-replacement sampling designs (e.g., Berger & Skinner, 2005; Berger & Rao, 2006; Berger, 2007).

Linearization is an alternative to replication methods (e.g., Robinson & Särndal, 1983; Binder, 1996; Deville, 1999; Demnati & Rao, 2004; Graf, 2011). Although slightly design-biased, the linearization variance estimators are more stable than its replication counterparts (e.g., Kish & Frankel, 1974; Kovar, Rao, & Wu, 1988; Shao & Tu, 1995, pp. 32, 69). There are different approaches for deriving linearization variance estimators that may give asymptotically equivalent but different estimators (Binder, 1996, pp. 17, 18). Deville (1999) proposes an approach based upon derivatives of the population parameter of interest. Demnati & Rao (2004) propose an approach which is based on derivatives of the estimator of the parameter of interest. Nevertheless,

* *Author to whom correspondence may be addressed.*
E-mail: emilio.lopez@itam.mx

linearization involves deriving analytic derivatives; a well-documented practical drawback (e.g., Shao & Tu, 1995, pp. 69, 281). Skinner (2004) and Demnati & Rao (2004, p. 21) raised the need of a replication estimator that overcomes that practical drawback.

We propose a new replicate variance estimator suitable for differentiable functions of estimated totals. The estimator is defined for any unequal-probability without-replacement sampling design and it naturally includes finite population corrections. The proposed approach consists in repeatedly perturbing the sampling weights. In Sections 3.1 and 3.2, we show that this novel approach can be interpreted in several ways depending on its configuration. Further, we show that the proposed replicate variance estimator is approximately equal to linearization variance estimators. Moreover, it can be seen as an approximation to the linearization estimators obtained by the Demnati & Rao (2004) approach.

We also show that it is asymptotically design-consistent and that it can handle two-stage sampling. For the Horvitz & Thompson (1952) point estimator, the proposed variance estimator reduces to the Horvitz & Thompson (1952) and the Sen (1953) and Yates & Grundy (1953) variance estimator.

2. THE CLASS OF POINT ESTIMATORS

Let $\mathcal{U} = \{1, \ldots, k, \ell, \ldots, N\}$ denote a finite population and let $\mathcal{S} = \{1, \ldots, n\} \subseteq \mathcal{U}$ denote a sample whose elements are randomly selected with an unequal probability sampling design without replacement. We assume full response. Consider the population parameter $\theta = h(t_1, \ldots, t_q, \ldots, t_Q)$, where $h(\cdot)$ is a smooth and differentiable function (e.g., Shao & Tu, 1995, ch. 2) of population totals t_q, $(q = 1, \ldots, Q)$ of Q survey variables, $t_q = \sum_{k\in\mathcal{U}} y_{qk}$, with y_{qk} denoting the value of the variable q for the unit $k \in \mathcal{U}$. Suppose that θ is estimated by its substitution point estimator $\hat{\theta} = h(\hat{t}_1, \ldots, \hat{t}_q, \ldots, \hat{t}_Q)$, where $\hat{t}_q$, is the Horvitz & Thompson (1952) point estimator $\hat{t}_q = \sum_{k\in\mathcal{S}} w_k\, y_{qk}$, with survey weights $w_k = 1/\pi_k$; where $\pi_k > 0$ is the inclusion probability of the unit k.

3. THE PROPOSED REPLICATE VARIANCE ESTIMATOR

We propose to estimate the variance of $\hat{\theta}$ by

$$\widehat{\mathrm{var}}(\hat{\theta})^{\mathrm{HT}}_{\mathrm{prop}} = \sum_{k\in\mathcal{S}} \sum_{\ell\in\mathcal{S}} \mathcal{D}_{k\ell}\; w_k \nu_k\; w_\ell \nu_\ell, \tag{1}$$

where $\mathcal{D}_{k\ell} = (\pi_{k\ell} - \pi_k\pi_\ell)/\pi_{k\ell}$, with $\pi_{k\ell} > 0$ denoting the joint inclusion probability of the units k and ℓ, and where

$$\nu_k = \frac{\hat{\theta} - \hat{\theta}^*_k}{\varrho_k} \tag{2}$$

with

$$\varrho_k = w_k^{1-\alpha_k}$$

for some $\alpha_k \geq 0$ (see Section 3.1), where $\hat{\theta}^*_k$ has the same functional form as $\hat{\theta}$ but using $\hat{t}^*_{qk}$ instead of $\hat{t}_q$, that is, $\hat{\theta}^*_k = h(\hat{t}^*_{1k}, \ldots, \hat{t}^*_{qk}, \ldots, \hat{t}^*_{Qk})$, with

$$\hat{t}^*_{qk} = \sum_{\ell\in\mathcal{S}} w^*_{\ell(k)}\, y_{q\ell}, \tag{3}$$

where

$$w^*_{\ell(k)} = \begin{cases} w_\ell & \text{if } \ell \neq k, \\ w_k - \varrho_k & \text{if } \ell = k. \end{cases}$$

Alternatively, with fixed sample size, we propose to use the estimator,

$$\widehat{\text{var}}(\hat{\theta})^{\text{SYG}}_{\text{prop}} = \frac{-1}{2} \sum_{k \in \mathcal{S}} \sum_{\ell \in \mathcal{S}} \mathcal{D}_{k\ell} \, (w_k \nu_k - w_\ell \nu_\ell)^2, \tag{4}$$

which is positive provided that the Sen–Yates–Grundy condition, $\mathcal{D}_{k\ell} < 0$, holds.

3.1. The Value of α_k

We originally developed the proposed replication variance estimators $\widehat{\text{var}}(\hat{\theta})^{\text{HT}}_{\text{prop}}$ and $\widehat{\text{var}}(\hat{\theta})^{\text{SYG}}_{\text{prop}}$ for $\alpha_k = 1$, Escobar & Berger (2011) jackknife variance estimator. However, to avoid restricting α_k, we explore its range of values. In Section 5, we show that Equations (1) and (4) are valid for any $\alpha_k \geq 0$. We recommend to use $\alpha_k = 1$ or $\alpha_k > 1$ (see below comments and Sections 3.3, 4.2, and 7).

Using $\alpha_k = 0$ gives $\varrho_k = w_k$ that corresponds to a naïve jackknife that deletes the unit k, that is, using $w^*_{\ell(k)} = 0$ if $\ell = k$. In Section 7 we show that this case produces biased and unstable estimates.

Using $\alpha_k = 1$ gives $\varrho_k = 1$. This implies that Equations (2) and (3) reduce respectively to $\nu_k = \hat{\theta} - \hat{\theta}^*_k$ and $\hat{t}^*_{qk} = \hat{t}_q - y_{qk}$, obtaining the Escobar & Berger (2011) jackknife. In this case, note that α_k (and therefore ϱ_k) is a constant free of k.

Using $\alpha_k > 1$ results in approximating the empirical influence function, that is, the Gâteaux (1919, p. 82) derivate, or the linearization variance estimators obtained by the Demnati & Rao (2004) approach. The larger the value of α_k, the closer the approximation (Sections 3.3, 4.2, and 7). This feature can be used when the derivatives involved in linearization become extremely cumbersome, for example, when $h(\cdot)$ is an implicit function (Shao & Tu, 1995, p. 29).

3.2. Example of a Total (the Underlying Idea)

Consider the case of estimating a total, $\hat{\theta} = \hat{t} = \sum_{k \in \mathcal{S}} w_k y_k$. The Equations (2) and (3) imply that $\nu_k = y_k$. Hence, Equations (1) and (4) reduce respectively to the Horvitz & Thompson (1952), and the Sen (1953); Yates & Grundy (1953) unbiased estimators of var($\hat{t}$). Note that this is true for any value of α_k.

3.2.1. Intuitive underlying idea

As suggested by a referee, we give an intuitive explanation. Consider the case $\alpha_k = 1$. Let $\mho$ be the artificial population obtained from expanding the $y_k \in s$ by their w_k, that is, the expanded sample. Accordingly, we are omitting y_k from $\mho$ and estimating θ via $\hat{\theta}^*_k = \hat{t} - y_k$ with a Bias$(\hat{\theta}^*_k) = -y_k$. Hence, $\nu_k = \hat{\theta} - \hat{\theta}^*_k = y_k$.

Comparing with the customary jackknife (Quenouille, 1956; Tukey, 1958) we have that: (i) The proposed replication removes units from $\mho$ instead of from s, (ii) The customary assumes the pseudo-values are unbiased and approx. i.i.d., whereas in the proposed the ν_k values are neither unbiased nor i.i.d. If $\alpha_k > 1$, the proposed deletes bits of units, that is, it perturbs the w_k by ϱ_k. The proposed variance estimator can thus be seen as a *post-expansion* or as a *delete-weight* jackknife.

From a Bootstrap (Efron, 1979) perspective, the proposed estimator can also be seen as a Bootstrap that deterministically subsamples n different subsets of size $\#(\mho)) - \varrho_k$ from $\mho$, instead of randomly subsampling (say) L resamples of size n from $\mho$. Note that there are at most n different pairs $w_k y_k$ in $\mho$.

As it has been mentioned, and as it will be shown, the proposed replication estimator can also be seen as an approximation to linearization estimators.

3.3. Example of a Ratio

Let $R = t_y/t_x$, be the parameter of interest, 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. Assume that R is estimated with the point estimator $\hat{R} = \hat{t}_y/\hat{t}_x$. A linearization variance estimator of $\hat{R}$ is given by (e.g., Demnati & Rao, 2004, Example 2.1),

$$\widehat{\text{var}}(\hat{R})_{Lin} = \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}} \mathcal{D}_{k\ell}\ w_k u_k\ w_\ell u_\ell \tag{5}$$

with

$$u_k = \frac{y_k - \hat{R}x_k}{\hat{t}_x}. \tag{6}$$

In this case, the ν_k values in Equation (2) are given by

$$\nu_k = u_k\left(\frac{\hat{t}_x}{\hat{t}_x - \varrho_k x_k}\right). \tag{7}$$

From Equations (6) and (7) we see that $\nu_k \approx u_k$, if $\hat{t}_x/(\hat{t}_x - \varrho_k x_k) \approx 1$. Thus, for large α_k, that is, small ϱ_k, we have that Equation (1) is a suitable approximation of Equation (5) and its sensitivity to highly-skewed weights should be low (see Section 7).

4. ALTERNATIVE ESTIMATORS FOR THE VARIANCE

4.1. Generalized Jackknife

The Campbell (1980) and Berger & Skinner (2005) generalized jackknife is defined by,

$$\widehat{\text{var}}(\hat{\theta})^{\text{CBS}}_{\text{Jack}} = \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}} \mathcal{D}_{k\ell}\ w_k\varepsilon_k\ w_\ell\varepsilon_\ell \tag{8}$$

with

$$\varepsilon_k = \left(\frac{1}{w_k} - \frac{1}{\hat{N}}\right)(\tilde{\theta} - \tilde{\theta}^{(k)}), \tag{9}$$

where $\tilde{\theta} = g(\tilde{\mu}_1, \ldots, \tilde{\mu}_q, \ldots, \tilde{\mu}_Q)$ is a function of Hájek (1971) mean estimators of Q variables with $\tilde{\mu}_q = \hat{t}_q/\hat{N}$, $\hat{N} = \sum_{k\in\mathcal{S}} w_k$ and where $\tilde{\theta}^{(k)} = g(\tilde{\mu}_1^{(k)}, \ldots, \tilde{\mu}_q^{(k)}, \ldots, \tilde{\mu}_Q^{(k)})$ has the same functional form as $\tilde{\theta}$ but replacing $\tilde{\mu}_q$ by $\tilde{\mu}_q^{(k)} = (\hat{t}_q - w_k y_k)/(\hat{N} - w_k)$. Note that the generalized jackknife (Eq. 8) is designed for functions of means whereas the proposed estimator (Eq. 1) is designed for functions of totals. Thus, the proposed estimator is more general.

4.1.1. Example of a ratio (revisited)

When $\hat{\theta} = \hat{R}$, we have that $\tilde{\theta} = \hat{\theta}$ and that the ε_k from Equation (9) are given by

$$\varepsilon_k = u_k\left(\frac{\hat{t}_x}{\hat{t}_x - w_k x_k}\right)\left(\frac{\hat{N} - w_k}{\hat{N}}\right). \tag{10}$$

From Equations (6), (7), and (10) we see that $\varepsilon_k \approx u_k$, if $(\hat{N} - w_k)/\hat{N} \approx 1$, and if $\hat{t}_x/(\hat{t}_x - w_k x_k) \approx 1$. Note that Equation (7) is a better approximation of Equation (6) than Equation (10). Hence, the proposed estimator (Eq. 1) should be as precise as Equation (5).

As suggested by a referee, we compute the above example for Poisson sampling. It can be shown that the Equations (5), (8), and (1) reduce respectively to

$$\widehat{\mathrm{var}}(\hat{R})_{\mathrm{PoiLin}} = \sum_{k \in \mathcal{S}} \frac{w_k - 1}{w_k} \, (w_k \, u_k)^2, \tag{11}$$

$$\widehat{\mathrm{var}}(\hat{R})^{\mathrm{CBS}}_{\mathrm{PoiJack}} = \sum_{k \in \mathcal{S}} \frac{w_k - 1}{w_k} \left(\frac{\hat{t}_x}{\hat{t}_x - w_k x_k} \right)^2 \left(\frac{\hat{N} - w_k}{\hat{N}} \right)^2 (w_k \, u_k)^2, \tag{12}$$

$$\widehat{\mathrm{var}}(\hat{R})^{\mathrm{HT}}_{\mathrm{PoiProp}} = \sum_{k \in \mathcal{S}} \frac{w_k - 1}{w_k} \left(\frac{\hat{t}_x}{\hat{t}_x - \varrho_k x_k} \right)^2 (w_k \, u_k)^2 \tag{13}$$

with u_k as defined in Equation (6). Thus, we can see that Equation (13) is a better approximation to Equation (11) than Equation (12), for uniformly negligible ϱ_k (large α_k).

4.2. Linearization Based on the Gâteaux Derivative

Let $\theta = T(M)$ be a functional where M is measure that allocates the unit mass to $k \in \mathcal{U}$ and let $\hat{\theta}$ be a functional $T(\hat{M})$, where $\hat{M}$ denotes a sample-based measure that allocates the mass w_k to the element $k \in \mathcal{S}$, and let δ_{y_k} be the degenerate point mass at y_k. The *empirical* influence function of $T(\cdot)$ (e.g., Davison & Hinkley, 1997, Section 2.7) in $\hat{M}$ (if it exists) is defined as

$$\mathrm{EIT}(\hat{M}, y_k) \;=\; \lim_{\zeta \to 0} \frac{T[\hat{M} + \zeta \delta_{y_k}] - T(\hat{M})}{\zeta},$$

which is the Gâteaux (1919) derivative of $T(\cdot)$ over the random measure $\hat{M}$, that is, the Demnati & Rao (2004) linearization approach (see Goga, Deville, & Ruiz-Gazen, 2009). A variance estimator is then given by Equation (1) after substituting ν_k by $\mathrm{EIT}(\hat{M}, y_k)$.

The approach proposed by Deville (1999, p. 197) estimates the population influence function $\mathrm{IT}(M, y_k)$ by $\mathrm{IT}(\hat{M}, y_k)$, that is, the influence function of $T(\cdot)$ in M evaluated at $M = \hat{M}$. Note that $\mathrm{IT}(\hat{M}, y_k)$ can be different from $\mathrm{EIT}(\hat{M}, y_k)$ which is the *empirical* influence function of $T(\cdot)$ in $\hat{M}$. Also note that, for the ratio (see Deville, 1999, p. 198), that approach does not always give the linearization estimator Equation (5) with u_k as in Equation (6).

Using Equation (2), we have that

$$\nu_k \;=\; \frac{T[\hat{M} + \zeta_k \delta_{y_k}] - T(\hat{M})}{\zeta_k}$$

with $\zeta_k = -\varrho_k$. Thus, ν_k can be interpreted as an approximation of $\mathrm{EIT}(\hat{M}, y_k)$. In other words, using the notation of Demnati & Rao (2004), ν_k approximates $z_k = \partial g(a_1, \ldots, a_N)/\partial a_k|_{(a_k = d_k)}$, where $\hat{\theta} = g(d_1, d_2, \ldots, d_N)$ with $d_k = w_k$ if $k \in \mathcal{S}$ and $d_k = 0$ otherwise, and constants $a_1, \ldots, a_N$. Again, large α_k (small ϱ_k) ensure that ζ_k is small. Note that the α_k can be defined such that ϱ_k is a constant free of k, for example, if $\alpha_k = 1$.

5. DESIGN-CONSISTENCY

We now set the validity of the proposed variance estimators Equations (1) and (4) under the Isaki & Fuller (1982) asymptotic framework. Consider a sequence of nested populations of increasing sizes $\{N_t : 0 < N_t < N_{t+1}, \forall t\}$. Consider also a sequence of non-necessarily nested samples of increasing sizes $\{n_t : n_t < n_{t+1}; n_t < N_t, \forall t\}$. Thus, if $t \to \infty$, it implies that $N_t \to \infty$ and $n_t \to \infty$, with $f = n_t/N_t$ a constant free of the limiting process. In what follows, we drop the index t to simplify the notation.

Consider the Horvitz & Thompson (1952) estimator $\hat{\mu}_q = \sum_{k\in\mathcal{S}} \bar{w}_k y_{qk}$ of the population mean $\mu_q = t_q/N, q = 1, \ldots, Q$ where $\bar{w}_k = w_k/N$. Thus, for the vector of means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_Q)^{\mathrm{T}}$ and the vector of estimators $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_Q)^{\mathrm{T}}$, the multivariate Horvitz-Thompson and Sen-Yates-Grundy design variances and variance estimators of $\hat{\boldsymbol{\mu}}$ are

$$\begin{aligned}\mathbf{var}(\hat{\boldsymbol{\mu}})_{\mathrm{HT}} &= \sum_{k\in\mathcal{U}}\sum_{\ell\in\mathcal{U}} \mathcal{D}_{k\ell}\, \pi_{k\ell}\, \bar{w}_k\bar{w}_\ell\, \boldsymbol{y}_k\boldsymbol{y}_\ell^{\mathrm{T}},\\ \widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{HT}} &= \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}} \mathcal{D}_{k\ell}\, \bar{w}_k\bar{w}_\ell\, \boldsymbol{y}_k\boldsymbol{y}_\ell^{\mathrm{T}},\end{aligned} \tag{14}$$

$$\begin{aligned}\mathbf{var}(\hat{\boldsymbol{\mu}})_{\mathrm{SYG}} &= \frac{-1}{2}\sum_{k\in\mathcal{U}}\sum_{\ell\in\mathcal{U}} \mathcal{D}_{k\ell}\, \pi_{k\ell}\, \{\bar{w}_k\boldsymbol{y}_k - \bar{w}_\ell\boldsymbol{y}_\ell\}\{\bar{w}_k\boldsymbol{y}_k - \bar{w}_\ell\boldsymbol{y}_\ell\}^{\mathrm{T}},\\ \widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{SYG}} &= \frac{-1}{2}\sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}} \mathcal{D}_{k\ell}\, \{\bar{w}_k\boldsymbol{y}_k - \bar{w}_\ell\boldsymbol{y}_\ell\}\{\bar{w}_k\boldsymbol{y}_k - \bar{w}_\ell\boldsymbol{y}_\ell\}^{\mathrm{T}}\end{aligned} \tag{15}$$

with $\boldsymbol{y}_k = (y_{1k}, \ldots, y_{Qk})^{\mathrm{T}}$. Now, assume the following regularity conditions:

(a) $\widehat{\mathrm{var}}(\hat{\theta})_L/\mathrm{var}(\hat{\theta})_L \to_p 1$, $\mathrm{var}(\hat{\theta})_L \neq 0$ where,

$$\begin{aligned}\mathrm{var}(\hat{\theta})_L &= \boldsymbol{\nabla}(\boldsymbol{\mu})^{\mathrm{T}}\mathbf{var}(\hat{\boldsymbol{\mu}})_{\mathrm{HT}}\boldsymbol{\nabla}(\boldsymbol{\mu}),\\ \widehat{\mathrm{var}}(\hat{\theta})_L &= \boldsymbol{\nabla}(\hat{\boldsymbol{\mu}})^{\mathrm{T}}\widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{HT}}\boldsymbol{\nabla}(\hat{\boldsymbol{\mu}}).\end{aligned} \tag{16}$$

Alternatively for fixed sample-size designs,

$$\begin{aligned}\mathrm{var}(\hat{\theta})_L &= \boldsymbol{\nabla}(\boldsymbol{\mu})^{\mathrm{T}}\mathbf{var}(\hat{\boldsymbol{\mu}})_{\mathrm{SYG}}\boldsymbol{\nabla}(\boldsymbol{\mu}),\\ \widehat{\mathrm{var}}(\hat{\theta})_L &= \boldsymbol{\nabla}(\hat{\boldsymbol{\mu}})^{\mathrm{T}}\widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{SYG}}\boldsymbol{\nabla}(\hat{\boldsymbol{\mu}}),\end{aligned} \tag{17}$$

where $\boldsymbol{\nabla}(\boldsymbol{x}) = (\partial h(\boldsymbol{\mu})/\partial\mu_1, \ldots, \partial h(\boldsymbol{\mu})/\partial\mu_Q)^{\mathrm{T}}_{\boldsymbol{\mu}=\boldsymbol{x}}$ is the gradient of $h(\cdot)$ at $\boldsymbol{x} \in \Re^Q$ with $h(\cdot)$ continuous and differentiable at $\boldsymbol{\mu}$.

(b) $\liminf \{n\ \mathrm{var}(\hat{\theta})_L\} > 0$.

(c) $n^{-1}\sum_{k\in\mathcal{S}} \bar{w}_k^{\tau}\bar{\varrho}_k^{\gamma}||\boldsymbol{y}_k||^{\tau+\gamma} = \mathcal{O}_p(n^{-(\tau+\gamma)}), \forall\tau \geq 2, \forall\gamma \geq 0$, where $\bar{\varrho}_k = \bar{w}_k^{1-\alpha_k}, \alpha_k \geq 0$, with $||\boldsymbol{A}|| = \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{1/2}$ the Euclidean norm.

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

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

(f) $\boldsymbol{\nabla}(\boldsymbol{x})$ is Lipschitz (Hölder) continuous of order δ, that is, $\|\boldsymbol{\nabla}(\boldsymbol{x}_1) - \boldsymbol{\nabla}(\boldsymbol{x}_2)\| \leq \lambda\, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{\delta}$, $\lambda > 0$ and $\delta > 0$ constants, $0 \leq \beta/2 < \delta$, for $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in the neighbourhood of $\boldsymbol{\mu}$.

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

The conditions (a), (c), (d), (e) and (f) are broadly similar to those proposed in Berger & Skinner (2005). Condition (a) is a necessary setting for proving the asymptotic design-consistency of replication variance estimators under standard arguments (e.g., Miller, 1964; Shao & Tu, 1995, Sections 2.1.1 and 3.1.5). This condition sets the existence of a consistent linearization variance estimator for the linearized variance. It holds in standard situations when it is possible to linearize the variance. Further details on the consistency of linearization variance estimators can be found, for example, in Robinson & Särndal (1983) and Särndal, Swensson, & Wretman (1992, Sections. 5.5 and 5.7). Regarding conditions (b) and (c), these are usual conditions utilized in asymptotic studies related to replication variance estimators with survey data (e.g., Shao & Tu, 1995, Section 6.4.1). Condition (b) holds in standard situations where the linearized variance decreases with rate n^{-1}; it holds, for example, when $\mathrm{var}(\hat{\theta})_L \geq \phi/n$ where ϕ is a non-negative constant. This inequality can be interpreted as an analogue to the Cramér–Rao lower bound. Condition (c) is a Lyapunov-type condition for the existence of moments and the behaviour of weights. This condition holds, for example, when the sampling weights w_k and the values of the y_k are bounded. Conditions (d) and (e) are mild conditions on the design similar to ones in Isaki & Fuller (1982, p. 91). These two conditions hold, for example, under simple random sampling where we obtain that $G_s = n^{-\beta}\{n/(n-1)\}(1-n/N)^2 = \mathcal{O}_p(1)$ and $H_s = 0$. Note that $G_s = \mathcal{O}_p(1)$ even if $\beta = 0$. Moreover, $H_s = 0$ as long as the Sen; Yates & Grundy (1953) condition $\mathcal{D}_{k\ell} < 0$ holds. It can be shown (see Berger, 2007; Escobar & Berger, 2013), that (d) and (e) hold under the Hájek (1964) maximum entropy sampling, the *rejective sampling* design (Hájek, 1981, Ch. 3, 7 and 14) or the *conditional Poisson sampling* design. Further, these two conditions allow two-stage sampling as shown in Escobar & Berger (2013). The conditions (f) and (g) are smoothness and differentiability requirements similar to those for the jackknife.

Theorem 1. *Under unequal probability sampling with fixed sample size, the regularity conditions (a)–(g) imply that Equation (4) is asymptotically design-consistent, that is,* $\widehat{var}(\hat{\theta})^{\mathrm{SYG}}_{\mathrm{prop}}/var(\hat{\theta})_L \rightarrow_p 1$.

A proof of Theorem 1 is given in the Appendix.

Corollary 1. *Under unequal probability sampling, if the regularity conditions (a)–(g) hold, then Equation (1) is asymptotically design-consistent. That is,* $\widehat{var}(\hat{\theta})^{\mathrm{HT}}_{\mathrm{prop}}/var(\hat{\theta})_L \rightarrow_p 1$.

The proof of Corollary 1 is also given in the Appendix. Hence, from the Theorem 1, Corollary 1 and the Slutsky's Theorem (e.g., Valliant, Dorfman, & Royall, 2000, p. 414), it follows that: $(\hat{\theta}-\theta)(\widehat{\mathrm{var}}(\hat{\theta})^{\mathrm{SYG}}_{\mathrm{prop}})^{-1/2} \rightarrow_d \mathrm{N}(0,1)$ and $(\hat{\theta}-\theta)(\widehat{\mathrm{var}}(\hat{\theta})^{\mathrm{HT}}_{\mathrm{prop}})^{-1/2} \rightarrow_d \mathrm{N}(0,1)$, when $\hat{\theta}$ is asymptotically Normal, thus yielding valid confidence intervals of $\hat{\theta}$ for θ.

6. TWO-STAGE SAMPLING

The proposed variance estimator can be used for two-stage sampling. For example, consider a self-weighted two-stage sampling design where n_I clusters are selected with inclusion probabilities π_{Ii} proportional to their sizes M_i, $(i = 1, \ldots, N_I)$. Within each selected cluster a simple random sample of m elements is drawn. Hence, the clusters' inclusion probabilities are $\pi_{Ii} = n_I M_i/N$ and the elements' inclusion probabilities are $\pi_k = n/N = f$.

By using the Hájek (1964, Eq. (5.27), p. 1511) approximation and by denoting $q_{Ii} = 1 - \pi_{Ii}$, the clusters' joint inclusion probabilities π_{Iij} are approximated by

$$\pi_{Iij} \approx \pi_{Ii}\,\pi_{Ij}\left\{1 - \frac{q_{Ii}q_{Ij}}{d}\right\} \quad (i \neq j = 1, \ldots, N_I), \tag{18}$$

where $d = \sum_{i\in\mathcal{U}} \pi_{Ii}\, q_{Ii}$. This approximation was originally developed for $d \to \infty$, that is, for $N_I \to \infty$ with a fixed m, under the maximum entropy sampling design (see Hájek, 1981, Chapters 3 and 6); namely Rejective Sampling. It requires that the used sampling design (for clusters) is of large entropy. Low entropy sampling designs (e.g., systematic probability proportional-to-size design) are not suitable for the above approximation. However, randomized systematic sampling is suitable as it is of large entropy. See Berger & Tillé (2009) for an overview. Berger (2011) gives sufficient conditions under which Hájek's results still hold for large entropy designs that do not possess maximum entropy.

Hence, given that the i-th cluster is selected, the elements' conditional inclusion probabilities are $\pi_{k|Ii} = m/M_i$ and $\pi_{k\ell|Ii} = m(m-1)/M_i(M_i-1)$. Using Equation (18), the elements' joint inclusion probabilities $\pi_{k\ell}$ are given by

$$\pi_{k\ell} \approx \begin{cases} \pi_{Ii}\,\pi_{k|Ii} = f & \text{if } (k=\ell) \in s_i, \\ \pi_{Ii}\,\pi_{k\ell|Ii} = f(m-1)/(M_i-1) & \text{if } (k \neq \ell) \in s_i, \\ \pi_{Iij}\,\pi_{k|Ii}\,\pi_{\ell|Ij} \approx f^2\{1-d^{-1}q_{Ii}q_{Ij}\} & \text{if } k \in s_i, \ell \in s_j, i \neq j, \end{cases}$$

where s_i denotes the sample of the i-th cluster. Substituting $\pi_{k\ell}$ in $\mathcal{D}_{k\ell}$, we obtain

$$\mathcal{D}_{k\ell} \approx \begin{cases} 1-f & \text{if } (k=\ell) \in s_i, \\ 1-\pi^*_{Ii} & \text{if } (k \neq \ell) \in s_i, \\ q_{Ii}\,q_{Ij}/(q_{Ii}\,q_{Ij}-d) & \text{if } k \in s_i,\ \ell \in s_j,\ i \neq j, \end{cases} \tag{19}$$

where $\pi^*_{Ii} = \pi_{Ii} m(M_i-1)/(m-1)M_i$. Thus, the proposed estimator is given by the Equations (1) or (4) with $\mathcal{D}_{k\ell}$ substituted by Equation (19).

The proposed variance estimators Equations (1) and (4) are consistent under self-weighted two-stage sampling when the regularity conditions of Section 5 hold. Hence, assuming that the customary conditions (a), (b), (c), (f), and (g) hold, it is only necessary to show that the conditions (d) and (e) hold.

Let $\beta = \log(n_I)/\log(n) < 1$ such that $n^\beta = n_I$ and let $f_I = n_I/N_I$. It can be shown that $|1-\pi^*_{Ii}| = \mathcal{O}(1)$ for $M_i \geq m \geq 2, \forall i = 1, \ldots, N_I$ and that $d = \mathcal{O}(N_I)$ as $q_{Ii} = \mathcal{O}(1)$. If $1-\pi^*_{Ii} > 0$, we have from Conditions (d) and (e), that $G_s = f_I^2 m^2 \mathcal{O}_p(n_I^{-1})$ and $H_s = m(m-1)\mathcal{O}_p(1)$. If $1-\pi^*_{Ii} < 0$, we have $G_s = f_I^2 m^2 \mathcal{O}_p(n_I^{-1}) + m(m-1)\mathcal{O}_p(1)$ and $H_s = 0$. Thus, G_s and H_s are $\mathcal{O}_p(1)$ when f_I and m are $\mathcal{O}(1)$. Note that this is also true if $\hat{d} = \sum_{i\in\mathcal{S}}(1-\pi_{Ii})$ is used instead of d, since $\hat{d} = \mathcal{O}_p(n_I)$. Thus, the conditions (d) and (e) hold.

7. SIMULATION STUDY

Consider the sugar cane farms dataset (Chambers & Dunstan, 1986) is a population frame of size $N = 338$. The variables of interest are: *Gross value of cane* (y_{1k}), and *Total farm expenditure* (y_{2k}). The parameter of interest is the ratio $R = t_1/t_2$, with true value $R = 1.58$, that is estimated by the point estimator $\hat{R} = \hat{t}_1/\hat{t}_2$. For selecting the samples and for computing the joint inclusion probabilities we use the Midzuno (1951) method. We consider four scenarios where the inclusion probabilities π_k are proportional to the variables:

- *Total cane harvested* (x_k), to obtain π_k correlated to y_{1k} and y_{2k}.
- *Square root of the total cane harvested* ($\sqrt{x_k}$), to obtain π_k mildly correlated to y_{1k} and y_{2k} through a non-linear relationship.
- *Variable with ones*. The π_k are equal in this case, $\pi_k = n/N$.
- *Generated variable* (ψ_k^{-1}) where $\psi_k \sim$ Log-Normal(0, 1/2), to use independent and randomly highly-skewed sampling weights w_k.

For each simulation, 1,000,000 samples were selected to compute: the empirical relative bias, RB $= \mathrm{B}[\widehat{\mathrm{var}}(\hat{R})]/\mathrm{var}(\hat{R})$, where $\mathrm{B}[\widehat{\mathrm{var}}(\hat{R})] = \mathrm{E}[\widehat{\mathrm{var}}(\hat{R})] - \mathrm{var}(\hat{R})$, and the empirical relative root mean square error RRMSE $= \{\mathrm{MSE}[\widehat{\mathrm{var}}(\hat{R})]\}^{1/2}/\mathrm{var}(\hat{R})$. The term $\mathrm{var}(\hat{R})$ is the empirical variance computed from the 1,000,000 observed values of $\hat{R}$. Computations were performed in R

2.15.2 (R Core Team, 2012) using some routines from the R packages 'sampling' (Tillé & Matei, 2012) and 'samplingVarEst' (Escobar & Barrios, 2012). The variance estimators used are:

- The Quenouille (1956) and Tukey (1958) standard jackknife,

$$\widehat{\text{var}}(\hat{\theta})_{\text{STD}} = \left(1 - \frac{n}{N}\right) \frac{n-1}{n} \sum_{k \in \mathcal{S}} (\hat{\theta}_{(k)} - \hat{\theta}_{(\cdot)})^2 \tag{20}$$

  with an ad hoc finite population correction, where $\hat{\theta}_{(k)}$ has the same functional form as $\hat{\theta}$ but using $\hat{t}_{q(k)}$ instead of $\hat{t}_q$, that is, $\hat{\theta}_{(k)} = h(\hat{t}_{1(k)}, \ldots, \hat{t}_{Q(k)})$, with $\hat{t}_{q(k)} = \sum_{\ell \neq k \in \mathcal{S}} w_\ell y_{q\ell}$, and $\hat{\theta}_{(\cdot)} = n^{-1} \sum_{k \in \mathcal{S}} \hat{\theta}_{(k)}$,
- The Campbell (1980) and Berger & Skinner (2005) generalized jackknife Equation (8),
- The proposed variance estimator Eqaution (1) with $\alpha_k = b_k$ and $\alpha_k = 0, 1, 2$ where $b_k = 1 + \log(n)/\log(w_k + 1/n)$, that is, $\varrho_k \approx n^{-1}$ and $\varrho_k = w_k, 1, w_k^{-1}$.
- The linearization variance estimator from Equation (5).

In Table 1, we see that the standard jackknife (Eq. 20) has increasing RB, in absolute value, for increasing n under the unequal probability scenarios $\pi_k \propto x_k$, $\sqrt{x_k}$ and ψ_k^{-1}, and it has a decreasing RB for increasing n with $\pi_k = n/N$. The proposed estimator (Eq. 1) with $\alpha_k = 0$, has the largest but always positive RB that decreases with increasing n. Further, in all four scenarios, we can observe that for increasing values of α_k, the RB of the proposed (Eq. 1) tends to replicate the RB of the linearization estimator (Eq. 5), confirming that Equation (1) is approximately equal to Equation (5) when $\alpha_k > 1$. The CBS generalized jackknife (Eq. 8) has a decreasing RB for increasing n. Note that the RB of the proposed estimator can be smaller than the RB of Equation (8) in absolute value. This is more noticeable with $\alpha_k \neq 0$ for the scenarios $\pi_k \propto x_k$ and $\sqrt{x_k}$. The estimator (Eq. 8) tends to be less biased than (Eqs. 5 and 1) with independent or highly skewed sampling weights. However, the RB of Equation (8) tends to be greater than Equations (5) and (1) with $\alpha_k > 0$ or $\alpha_k = b_k$ in the case where there is a non-linear relationship between the inclusion probabilities and the variables of interest.

Table 2 shows that the linearization estimator (Eq. 5) has the smallest RRMSE among all other variance estimators, except for $\pi_k \propto \psi_k^{-1}$ and $f = 0.201, 0.302$ where the standard estimator (Eq. 20) has the smallest RRMSE. Again, in all four scenarios, we can observe, for increasing values of α_k, that the RRMSE of the estimator (Eq. 1) tends to replicate the RRMSE of the linearization estimator (Eq. 5). In almost all cases, note that the proposed estimator (Eq. 1) with $\alpha_k \neq 0$ has smaller RRMSE than the estimator (Eq. 8). Thus, the estimator (Eq. 1) is more stable than the estimator (Eq. 8). This is more noticeable for small n, when the inclusion probabilities are poorly correlated with the variables of interest, or with highly skewed sampling weights. However, as previously suggested, the estimator from Equation (1) may become unstable if the extreme value $\alpha_k = 0$ is used.

8. DISCUSSION

We propose a new design-consistent replication variance estimator for any unequal-probability without-replacement sampling design. The proposed replication estimator is approximately equal to the linearization variance estimators proposed by Demnati & Rao (2004).

As it is suitable for functions of Horvitz & Thompson (1952) totals, the proposed estimator enjoys of broad applicability, being more general than the generalized jackknife (Campbell, 1980; Berger & Skinner, 2005) that is designed for functions of Hájek (1971) means. Our empirical results suggest that the proposed estimator is more stable than its competitors.

TABLE 1: Relative bias (%) of the Quenouille (1956) and Tukey (1958) standard jackknife (STD), the Campbell (1980) and Berger & Skinner (2005) generalized jackknife (CBS), the proposed replication variance estimator, and the Taylor linearization variance estimator.

		STD Jack.	CBS Jack.	Proposed replication (Eq. 1) with				Taylor Lin.
n	$f(\%)$	(Eq. 20)	(Eq. 8)	$\alpha_k = b_k$	$\alpha_k = 0$	$\alpha_k = 1$	$\alpha_k = 2$	(Eq. 5)
$\pi_k \propto x_k$								
2	0.6	18.8	10.9	−2.5	343.3	−2.1	−2.9	−2.9
4	1.2	4.8	4.2	−1.6	84.4	−1.1	−1.8	−1.8
7	2.1	2.7	1.8	−1.0	38.7	−0.3	−1.1	−1.1
10	3.0	2.1	1.1	−0.8	25.0	−0.1	−0.8	−0.9
17	5.0	2.1	0.5	−0.5	13.7	0.3	−0.5	−0.5
34	10.1	3.1	−0.2	−0.6	6.2	0.1	−0.5	−0.6
68	20.1	7.8	0.1	0.0	3.3	0.7	0.2	0.0
102	30.2	14.2	−0.1	0.0	2.2	0.7	0.2	0.0
$\pi_k \propto \sqrt{x_k}$								
2	0.6	15.0	12.3	−8.2	326.3	−7.9	−8.5	−8.5
4	1.2	5.9	11.7	−5.0	86.6	−4.5	−5.2	−5.2
7	2.1	3.2	6.7	−3.0	40.0	−2.3	−3.1	−3.2
10	3.0	2.5	4.6	−2.1	25.9	−1.4	−2.2	−2.2
17	5.0	2.0	2.5	−1.3	14.1	−0.5	−1.3	−1.4
34	10.1	2.6	1.1	−0.7	6.6	0.1	−0.7	−0.8
68	20.1	5.5	0.4	−0.5	3.1	0.4	−0.3	−0.5
102	30.2	10.0	0.3	−0.2	2.1	0.6	0.1	−0.2
$\pi_k = n/N$								
2	0.6	8.5	−3.6	−24.1	285.6	−23.8	−24.3	−24.3
4	1.2	6.8	5.1	−16.8	86.8	−16.4	−16.9	−17.0
7	2.1	5.1	4.7	−11.1	42.6	−10.5	−11.2	−11.2
10	3.0	4.0	3.9	−8.1	28.3	−7.4	−8.2	−8.2
17	5.0	2.5	2.4	−5.1	15.7	−4.3	−5.1	−5.2
34	10.1	1.0	1.0	−2.9	7.2	−2.0	−2.9	−3.0
68	20.1	0.9	0.9	−1.1	4.0	−0.1	−0.9	−1.1
102	30.2	0.2	0.2	−1.1	2.2	−0.2	−0.8	−1.1
$\pi_k \propto 1/\psi_k$								
2	0.6	7.0	−28.2	−41.3	252.9	−41.1	−41.5	−41.5
4	1.2	10.4	−12.8	−29.1	94.4	−28.6	−29.2	−29.2
7	2.1	6.8	−8.9	−21.6	47.7	−21.0	−21.6	−21.6
10	3.0	5.0	−7.0	−17.2	32.2	−16.6	−17.3	−17.3
17	5.0	3.1	−4.2	−11.5	19.1	−10.8	−11.5	−11.6
34	10.1	−0.3	−2.2	−6.9	9.5	−6.1	−6.8	−6.9
68	20.1	−4.9	−1.4	−4.3	4.6	−3.4	−4.1	−4.3
102	30.2	−9.2	−1.1	−3.2	3.1	−2.3	−3.0	−3.2

TABLE 2: Relative root mean-square error (%) of the Quenouille (1956) and Tukey (1958) standard jackknife (STD), the Campbell (1980) and Berger & Skinner (2005) generalized jackknife (CBS), the proposed replication variance estimator, and the Taylor linearization variance estimator.

		STD Jack.	CBS Jack.	Proposed replication (Eq. 1) with				Taylor Lin.
n	$f(\%)$	(Eq. 20)	(Eq. 8)	$\alpha_k = b_k$	$\alpha_k = 0$	$\alpha_k = 1$	$\alpha_k = 2$	(Eq. 5)
$\pi_k \propto x_k$								
2	0.6	202.1	189.5	150.4	814.9	151.0	149.9	149.9
4	1.2	101.7	102.9	88.3	199.2	88.8	88.1	88.1
7	2.1	68.7	68.5	63.2	101.3	63.6	63.1	63.1
10	3.0	55.0	54.7	51.7	72.4	52.1	51.7	51.7
17	5.0	40.4	40.0	38.8	47.5	39.1	38.8	38.8
34	10.1	27.4	26.8	26.4	29.3	26.6	26.4	26.4
68	20.1	20.0	18.1	18.0	19.0	18.1	18.0	18.0
102	30.2	20.3	14.3	14.3	14.8	14.4	14.3	14.3
$\pi_k \propto \sqrt{x_k}$								
2	0.6	197.4	184.8	144.7	774.9	145.2	144.3	144.3
4	1.2	108.9	118.3	89.1	209.7	89.7	89.0	89.0
7	2.1	74.0	78.3	65.9	108.3	66.4	65.8	65.8
10	3.0	59.1	61.3	54.5	77.3	54.9	54.5	54.4
17	5.0	42.9	43.5	40.7	50.0	41.1	40.7	40.7
34	10.1	28.8	28.2	27.4	30.4	27.7	27.4	27.4
68	20.1	19.7	17.9	17.7	18.7	17.9	17.7	17.7
102	30.2	17.6	13.3	13.2	13.7	13.3	13.2	13.2
$\pi_k = n/N$								
2	0.6	186.3	161.3	132.3	705.4	132.6	132.0	131.9
4	1.2	127.4	124.2	90.7	237.2	91.1	90.6	90.5
7	2.1	94.0	93.5	73.3	134.0	73.8	73.2	73.2
10	3.0	76.4	76.2	63.4	98.2	63.9	63.4	63.4
17	5.0	55.4	55.4	49.6	64.4	50.0	49.6	49.5
34	10.1	36.4	36.4	34.6	39.3	34.9	34.6	34.6
68	20.1	23.8	23.8	23.2	24.8	23.4	23.2	23.2
102	30.2	17.9	17.9	17.6	18.3	17.8	17.6	17.6
$\pi_k \propto 1/\psi_k$								
2	0.6	191.5	145.0	126.1	681.5	126.5	125.8	125.8
4	1.2	153.9	112.2	91.5	289.9	91.9	91.4	91.4
7	2.1	129.2	95.0	80.5	187.7	80.9	80.5	80.5
10	3.0	114.1	88.6	77.2	150.6	77.6	77.2	77.2
17	5.0	91.7	78.7	71.4	110.9	71.8	71.4	71.4
34	10.1	63.6	61.7	58.1	73.9	58.5	58.1	58.1
68	20.1	42.2	46.2	44.6	50.7	44.8	44.6	44.6
102	30.2	33.1	39.3	38.3	41.8	38.5	38.3	38.3

The proposed replicate estimator can be extended in a number of ways. For example, by embedding the Hájek (1964) approximation for the joint inclusion probabilities as in Berger (2007) and as in Escobar & Berger (2013) for self-weighted two-stage sampling; or by addressing non-response adjustments as in Berger & Rao (2006). Further, another possibility is to address the variance estimation of model parameters (e.g., Demnati & Rao, 2010).

ACKNOWLEDGEMENTS

The authors are grateful to Chris Skinner (London School of Economics, U.K.), Nikolaos Tzavidis (University of Southampton, U.K.) and Ray Chambers (University of Wollongong, Australia) and to two referees for helpful comments and suggestions. This research was supported by the Mexico's National Council for Science and Technology.

APPENDIX

Proof of Theorem 1 (and Corollary 1). We use standard arguments in proving design-consistency (e.g., Miller, 1964; Shao & Tu, 1995, Sections 2.1.1 and 3.1.5). Hence, from the mean value Theorem we have that,

$$\hat{\theta} - \hat{\theta}_k^* \;=\; h(\hat{\boldsymbol{\mu}}) - h(\hat{\boldsymbol{\mu}}_k^*) \;=\; \nabla(\boldsymbol{\xi}_k)^{\mathrm{T}}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_k^*) \;=\; \nabla(\hat{\boldsymbol{\mu}})^{\mathrm{T}}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_k^*) + r_k^*,$$

where $\boldsymbol{\xi}_k$ is a point between $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}_k^*$, and $r_k^* = \{\nabla(\boldsymbol{\xi}_k) - \nabla(\hat{\boldsymbol{\mu}})\}^{\mathrm{T}}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_k^*)$ is the remainder. Now, from the Equation (3) it can be shown that

$$\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_k^* \;=\; \bar{\varrho}_k\, \boldsymbol{y}_k, \tag{21}$$

where $\bar{\varrho}_k = \bar{w}_k^{1-\alpha_k}$ and $\bar{w}_k = w_k/N$, $\alpha_k \geq 0$. Combining with (Eq. 2) implies,

$$\nu_k \;=\; \nabla(\hat{\boldsymbol{\mu}})^{\mathrm{T}} \boldsymbol{y}_k \;+\; r_k, \tag{22}$$

where $r_k = \bar{\varrho}_k^{-1} r_k^* = \{\nabla(\boldsymbol{\xi}_k) - \nabla(\hat{\boldsymbol{\mu}})\}^{\mathrm{T}} \boldsymbol{y}_k$. For r_k, the Cauchy inequality implies

$$|r_k| \;\leq\; ||\nabla(\boldsymbol{\xi}_k) - \nabla(\hat{\boldsymbol{\mu}})||\; ||\boldsymbol{y}_k||. \tag{23}$$

As $\boldsymbol{\xi}_k$ is between $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}_k^*$ we have that $||\boldsymbol{\xi}_k - \hat{\boldsymbol{\mu}}|| \leq ||\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_k^*||$. This combined with condition (f) and (Eq. 21) imply, for constants $\lambda > 0$ and $0 \leq \beta/2 < \delta$, that

$$||\nabla(\boldsymbol{\xi}_k) - \nabla(\hat{\boldsymbol{\mu}})|| \;\leq\; \lambda ||\boldsymbol{\xi}_k - \hat{\boldsymbol{\mu}}||^{\delta} \;\leq\; \lambda ||\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_k^*||^{\delta} \;\leq\; \lambda \bar{\varrho}_k^{\delta} ||\boldsymbol{y}_k||^{\delta}. \tag{24}$$

Now, let

$$\tilde{r}_k = \bar{w}_k\, r_k, \tag{25}$$

$$\tilde{y}_k = \bar{w}_k\, \boldsymbol{y}_k^{\mathrm{T}} \nabla(\hat{\boldsymbol{\mu}}), \tag{26}$$

$$\Psi = n \operatorname{var}(\hat{\theta})_L. \tag{27}$$

By combining Equations (23), (24), and (25) and multiplying both sides by $\bar{w}_k$, we obtain

$$|\tilde{r}_k| \;\leq\; \lambda\, \bar{w}_k\, \bar{\varrho}_k^{\delta} ||\boldsymbol{y}_k||^{1+\delta}. \tag{28}$$

Using the Cauchy inequality and Condition (b) on Equations (26) and (27) give,

$$|\tilde{y}_k| \leq \bar{w}_k \, ||\boldsymbol{y}_k|| \, ||\nabla(\hat{\boldsymbol{\mu}})||, \tag{29}$$

$$\Psi^{-2} = \mathcal{O}(1). \tag{30}$$

By substituting Equation (22) in Equation (4) we obtain: $\widehat{\mathrm{var}}(\hat{\theta})^{\mathrm{SYG}}_{\mathrm{prop}} = A + 2(E - C) + D - B$, where

$$A = \nabla(\hat{\boldsymbol{\mu}})^{\mathrm{T}} \, \widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{SYG}} \, \nabla(\hat{\boldsymbol{\mu}}), \tag{31}$$

$$B = \sum\sum\nolimits_{k,\ell \in \mathcal{S}} \mathcal{D}_{k\ell} \, \tilde{r}_k \, \tilde{r}_\ell, \tag{32}$$

$$C = \sum\sum\nolimits_{k,\ell \in \mathcal{S}} \mathcal{D}_{k\ell} \, \tilde{r}_k \, \tilde{y}_\ell, \tag{33}$$

$$D = \sum\sum\nolimits_{k,\ell \in \mathcal{S}} \mathcal{D}_{k\ell} \, \tilde{r}_k^2, \tag{34}$$

$$E = \sum\sum\nolimits_{k,\ell \in \mathcal{S}} \mathcal{D}_{k\ell} \, \tilde{r}_k \, \tilde{y}_k \tag{35}$$

with $\widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{SYG}}$, $\tilde{r}_k$ and $\tilde{y}_k$ as Equations (15), (25) and (26). We have to show:

$$\frac{A}{\mathrm{var}(\hat{\theta})_L} \rightarrow_p 1, \tag{36}$$

$$\frac{B}{\mathrm{var}(\hat{\theta})_L} \rightarrow_p 0, \tag{37}$$

$$\frac{C}{\mathrm{var}(\hat{\theta})_L} \rightarrow_p 0, \tag{38}$$

$$\frac{D}{\mathrm{var}(\hat{\theta})_L} \rightarrow_p 0, \tag{39}$$

$$\frac{E}{\mathrm{var}(\hat{\theta})_L} \rightarrow_p 0. \tag{40}$$

If proving for Corollary 1, note that substituting Equation (22) in Equation (1) gives: $\widehat{\mathrm{var}}(\hat{\theta})^{\mathrm{HT}}_{\mathrm{prop}} = A + B + 2C$ with A as Equation (31) but replacing $\widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{SYG}}$ by $\widehat{\mathbf{var}}(\hat{\boldsymbol{\mu}})_{\mathrm{HT}}$ defined in Equation (14), and setting Equations (34) and (35) equal to zero. Thus, it would suffice to show Equations (36), (37) and (38).

Condition (a) implies the Equation (36). We now show Equation (37). From Equation (32) and Conditions (d) and (e) we have,

$$B = \frac{-1}{2} \sum_{k \in \mathcal{S}} \sum_{\ell \in \mathcal{S}} \mathcal{D}_{k\ell} (\tilde{r}_k - \tilde{r}_\ell)^2 + \frac{1}{2} \sum_{k \in \mathcal{S}} \sum_{\ell \in \mathcal{S}} \mathcal{D}_{k\ell} (\tilde{r}_k^2 + \tilde{r}_\ell^2) \leq \frac{B_1 + B_2}{2}, \tag{41}$$

where $B_1 = \sum\sum_{k,\ell \in \mathcal{S}} \mathcal{D}^-_{k\ell} (\tilde{r}_k - \tilde{r}_\ell)^2$ and $B_2 = \sum\sum_{k,\ell \in \mathcal{S}} \mathcal{D}^+_{k\ell} (\tilde{r}_k^2 + \tilde{r}_\ell^2)$. Now, using the Cauchy inequality on B_1, we have that $B_1^2 \leq G_s n^\beta \sum\sum_{k,\ell \in \mathcal{S}} (\tilde{r}_k - \tilde{r}_\ell)^4$, but as $\sum\sum_{k,\ell \in \mathcal{S}} (\tilde{r}_k - \tilde{r}_\ell)^4 = 2n \sum_{k \in \mathcal{S}} (\tilde{r}_k - \bar{r})^4 + 6 \left\{ \sum_{k \in \mathcal{S}} (\tilde{r}_k - \bar{r})^2 \right\}^2$ with $\bar{r} = n^{-1} \sum_{k \in \mathcal{S}} \tilde{r}_k$, we have that

$$B_1^2 \leq G_s \left[2n^{1+\beta} \sum_{k \in \mathcal{S}} (\tilde{r}_k - \bar{r})^4 + 6n^\beta \{ \sum_{k \in \mathcal{S}} (\tilde{r}_k - \bar{r})^2 \}^2 \right] \leq G_s (B_3 + B_4) \tag{42}$$

with $B_3 = 2n^{1+\beta}\sum_{k\in\mathcal{S}}\tilde{r}_k^4$ and $B_4 = 6n^{\beta}(\sum_{k\in\mathcal{S}}\tilde{r}_k^2)^2$. Hence, the Equation (42) and the Condition (d) imply that $B_1/\text{var}(\hat{\theta})_L \to_p 0$, if we show $(B_3+B_4)/\text{var}(\hat{\theta})_L^2 \to_p 0$. Thus, by using the Equation (28), we obtain

$$\frac{B_3+B_4}{\text{var}(\hat{\theta})_L^2} \le \frac{\lambda^4 n^{4+\beta}}{\Psi^2}\left[\frac{2}{n}\sum_{k\in\mathcal{S}}(\bar{w}_k\bar{\varrho}_k^{\delta}||\boldsymbol{y}_k||^{1+\delta})^4 + 6\{\frac{1}{n}\sum_{k\in\mathcal{S}}(\bar{w}_k\bar{\varrho}_k^{\delta}||\boldsymbol{y}_k||^{1+\delta})^2\}^2\right].$$

From Condition (f), we have $\beta < 4\delta$. This combined with Condition (c) and Equation (30) imply $(B_3+B_4)/\text{var}(\hat{\theta})_L^2 = n^{4+\beta}\mathcal{O}_p(n^{-4(1+\delta)}) + n^{4+\beta}\mathcal{O}_p(n^{-2(1+\delta)})^2$, that is,

$$(B_3+B_4)/\text{var}(\hat{\theta})_L^2 \to_p 0. \tag{43}$$

Thus, the Condition (d) and Equations (42) and (43) all together imply that

$$\frac{B_1}{\text{var}(\hat{\theta})_L} \to_p 0. \tag{44}$$

We now show that $B_2/\text{var}(\hat{\theta})_L \to_p 0$. As $\sum\sum_{k,\ell\in\mathcal{S}}(\tilde{r}_k^2+\tilde{r}_\ell^2)^2 = 2n\sum_{k\in\mathcal{S}}\tilde{r}_k^4 + 2(\sum_{k\in\mathcal{S}}\tilde{r}_k^2)^2$, we have by the Cauchy inequality that $B_2^2 \le H_s n^{\beta}\sum\sum_{k,\ell\in\mathcal{S}}(\tilde{r}_k^2+\tilde{r}_\ell^2)^2 = H_s(B_3+B_4/3)$. Thus, the Condition (e) and the Equation (43) imply that $B_2/\text{var}(\hat{\theta})_L \to_p 0$, which together with Equations (44) and (41) imply Equation (37). We now show Equation (38). By the triangle and Cauchy inequalities, Equation (33) implies

$$\begin{aligned}|C| &\le \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}|\mathcal{D}_{k\ell}||\tilde{r}_k||\tilde{y}_\ell| = \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}\mathcal{D}_{k\ell}^{-}|\tilde{r}_k||\tilde{y}_\ell| + \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}\mathcal{D}_{k\ell}^{+}|\tilde{r}_k||\tilde{y}_\ell|\\ &\le (G_s^{1/2}+H_s^{1/2})\tilde{C}^{1/2},\end{aligned}$$

where

$$\tilde{C} = n^{\beta}\sum_{k\in\mathcal{S}}\tilde{r}_k^2\sum_{\ell\in\mathcal{S}}|\tilde{y}_\ell|^2. \tag{45}$$

From Conditions (d) and (e), Equation (38) follows if we show $\tilde{C}/\text{var}(\hat{\theta})_L^2 \to_p 0$. Substituting Equations (28) and (29) in the Equation (45) implies

$$\frac{\tilde{C}}{\text{var}(\hat{\theta})_L^2} \le ||\nabla(\hat{\boldsymbol{\mu}})||^2\frac{\lambda^2 n^{4+\beta}}{\Psi^2}\left[\frac{1}{n}\sum_{k\in\mathcal{S}}(\bar{w}_k\,\bar{\varrho}_k^{\delta}||\boldsymbol{y}_k||^{1+\delta})^2\right]\left[\frac{1}{n}\sum_{\ell\in\mathcal{S}}\bar{w}_\ell^2||\boldsymbol{y}_\ell||^2\right]. \tag{46}$$

From Condition (f), $\beta < 2\delta$, which by Condition (c) and Equations (30) and (46) imply $\tilde{C}/\text{var}(\hat{\theta})_L^2 = n^{\beta}\mathcal{O}_p(n^{-2\delta})$; that implies Equation (38). We now show the Equation (39). Using the triangle and Cauchy inequalities on the Equation (34),

$$|D| \le \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}\mathcal{D}_{k\ell}^{-}|\tilde{r}_k|^2 + \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}\mathcal{D}_{k\ell}^{+}|\tilde{r}_k|^2 \le (G_s^{1/2}+H_s^{1/2})\tilde{D}^{1/2}, \tag{47}$$

where

$$\tilde{D} = n^{\beta}\sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}|\tilde{r}_k|^4 = n^{1+\beta}\sum_{k\in\mathcal{S}}|\tilde{r}_k|^4. \tag{48}$$

From Conditions (d) and (e), the Equation (39) follows if we show that $\tilde{D}/\mathrm{var}(\hat{\theta})^2_L \rightarrow_p 0$. By substituting the Equation (28) in Equation (48) we obtain,

$$\frac{\tilde{D}}{\mathrm{var}(\hat{\theta})^2_L} \;\leq\; \frac{\lambda^4\, n^{4+\beta}}{\Psi^2}\left[\frac{1}{n}\sum_{k\in\mathcal{S}}(\bar{w}_k\,\bar{\varrho}_k^{\,\delta}||\boldsymbol{y}_k||^{1+\delta})^4\right], \tag{49}$$

which by Condition (c), Equation (30) and as $\beta < 4\delta$, we have that $\tilde{D}/\mathrm{var}(\hat{\theta})^2_L = n^{\beta}\mathcal{O}_p(n^{-4\delta})$, implying the Equation (39). We now show the Equation (40). Using the triangle and the Cauchy inequalities on the Equation (35) gives

$$|E| \;\leq\; \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}\mathcal{D}^-_{k\ell}|\tilde{r}_k||\tilde{y}_k| + \sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}\mathcal{D}^+_{k\ell}|\tilde{r}_k||\tilde{y}_k| \;\leq\; (G_s^{1/2} + H_s^{1/2})\tilde{E}^{1/2},$$

where

$$\tilde{E} \;=\; n^{\beta}\sum_{k\in\mathcal{S}}\sum_{\ell\in\mathcal{S}}|\tilde{r}_k|^2|\tilde{y}_k|^2 \;=\; n^{1+\beta}\sum_{k\in\mathcal{S}}|\tilde{r}_k|^2|\tilde{y}_k|^2. \tag{50}$$

From Conditions (d) and (e), the Equation (40) follows if $\tilde{E}/\mathrm{var}(\hat{\theta})^2_L \rightarrow_p 0$. By substituting Equations (28) and (29) in the Equation (50),

$$\frac{\tilde{E}}{\mathrm{var}(\hat{\theta})^2_L} \;\leq\; ||\boldsymbol{\nabla}(\hat{\boldsymbol{\mu}})||^2\frac{\lambda^2\, n^{4+\beta}}{\Psi^2}\left[\frac{1}{n}\sum_{k\in\mathcal{S}}\bar{w}_k^{\,4}\bar{\varrho}_k^{\,2\delta}||\boldsymbol{y}_k||^{4+2\delta}\right]. \tag{51}$$

From Condition (f), we have $\beta < 2\delta$, which by Condition (c) and Equations (30) and (51) imply $\tilde{E}/\mathrm{var}(\hat{\theta})^2_L = n^{\beta}\mathcal{O}_p(n^{-2\delta})$, that is $\tilde{E}/\mathrm{var}(\hat{\theta})^2_L \rightarrow_p 0$. ■

BIBLIOGRAPHY

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

Berger, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pakistan Journal of Statistics* 27, 407–426.

Berger, Y. G. & Rao, J. N. K. (2006). Adjusted jackknife for imputation under unequal probability sampling without replacement. *Journal of the Royal Statistical Society Series B*, 68, 531–547.

Berger, Y. G. & Skinner, C. J. (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society Series B*, 67, 79–89.

Berger, Y. G. & Tillé, Y. (2009). Sampling with unequal probabilities. In *Sample Surveys: Design, Methods and Applications* (eds. D. Pfeffermann & C. R. Rao), 29A of *Handbook of Statistics*, 39–54. Amsterdam: Elsevier.

Binder, D. A. (1996). Linearization methods for single phase and two-phase samples: A cookbook approach. *Survey Methodology*, 22, 17–22.

Campbell, C. (1980). A different view of finite population estimation. *Proceeding of the Section on Survey Research Methods, American Statistical Association*, 319–324.

Chambers, R. L. & Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, 73, 597–604.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.

Demnati, A. & Rao, J. N. K. (2004). Linearization variance estimators for survey data. *Survey Methodology*, 30, 17–26.

Demnati, A. & Rao, J. N. K. (2010). Linearization variance estimators for model parameters from complex survey data. *Survey Methodology*, 36, 193–201.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193–203.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1–26.

Escobar, E. L. & Barrios, E. (2012). *SamplingVarEst: Sampling Variance Estimation*. R package version 0.9-1.

Escobar, E. L. & Berger, Y. G. (2011). Jackknife variance estimation for functions of Horvitz & Thompson estimators under unequal probability sampling without replacement. *Proceeding of the 58th World Statistics Congress*, International Statistical Institute, Dublin.

Escobar, E. L. & Berger, Y. G. (2013). A jackknife variance estimator for self-weighted two-stage samples. *Statistica Sinica*, 23, 595–613.

Gâteaux, R. (1919). Fonctions d'une infinité de variables indeépendantes. *Bulletin de la Société Mathématique de France*, 47, 70–96.

Goga, C., Deville, J.-C. & Ruiz-Gazen, A. (2009). Use of functionals in linearization and composite estimation with application to two-sample survey data. *Biometrika*, 96, 691–709.

Graf, M. (2011). Use of survey weights for the analysis of compositional data. In *Compositional Data Analysis: Theory and Applications*, Pawlowsky-Glahn, V. & Buccianti, A., editors, chap. 9. Wiley, Chichester.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35, 1491–1523.

Hájek, J. (1971). Comment on a paper by Basu, D. In *Foundations of Statistical Inference*, Godambe, V. P. & Sprott, D. A., editors, Holt, Rinehart & Winston, Toronto, p. 236.

Hájek, J. (1981). In *Sampling From a Finite Population*, Dupač, V., editor. Dekker, New York.

Horvitz, D. G. & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.

Isaki, C. T. & Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89–96.

Kish, L. & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society Series B*, 36, 1–37.

Kovar, J. G., Rao, J. N. K., & Wu, C. F. J. (1988). Bootstrap and other methods to measure errors in survey estimates. *The Canadian Journal of Statistics*, 16, 25–45.

Lehtonen, R. & Pahkinen, E. J. (2004). *Practical Methods for Design and Analysis of Complex Surveys*, 2nd ed., John Wiley & Sons, Chichester.

Midzuno, H. (1951). On the sampling system with probability proportionate to sum of sizes. *Annals of the Institute of Statistical Mathematics*, 3, 99–107.

Miller, R. G. (1964). A trustworthy jackknife. *The Annals of Mathematical Statistics*, 35, 1594–1605.

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

R Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robinson, P. M. & Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhya Series B*, 45, 240–248.

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

Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5, 119–127.

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

Skinner, C. J. (2004). Comment on the paper *Linearization variance estimators for survey data* by Demnati, A. & Rao, J. N. K. In *Survey Methodology*, 30, 17–26.

Tillé, Y. & Matei, A. (2012). *Sampling: Survey Sampling*. R package version 2.5.

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

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

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

Yates, F. & Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society Series B*, 15, 253–261.

Received 10 January 2012
Accepted 13 March 2013