Deriving Asymptotic Properties of Survey Sampling Estimators

Ismael Flores Cervantes[1]
[1]Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract
Most survey sampling estimators derive their large sample properties by establishing an asymptotic equivalence to the Horvitz-Thompson (HT) estimator If the proposed estimator is asymptotically equivalent to the HT estimator, then it inherits the HT estimator asymptotic properties such as design consistency with a limiting normal distribution. Although this approach is valid, it does not provide insights on the proposed estimator's efficiency in small samples. As a result, most papers include simulation studies to examine these properties empirically.

We take a different approach and show that methods from classical asymptotic theory for estimators as functions of random variables can be used to derive the asymptotic property of survey sampling estimators. The focus of this approach is the discrete random vector of the sample membership indicators as the only stochastic component of the estimator. The use of discrete multivariate statistics and matrix operations reduces the derivation of the expressions of the estimator and its asymptotic properties to an algebraic problem while providing new insights into its properties. We illustrate these methods by deriving the variance, variance estimator, and determining the sufficient conditions for the HT estimator and its variance estimator to be design consistent.

Keywords: Large sample theory, function of random variables, discrete multivariate statistics, finite population, sample design

1. Introduction

In this paper, we derive the large sample properties of survey sampling estimators using the principle and tools from classical asymptotic theory (Polansky, 2011; Lehmann, 1999). We extend an idea developed by Tillé (2006) and postulate that sample designs are uniquely defined as a multivariate discrete random variable for the sample membership indicator with an expected value and a variance-covariance matrix with specific properties that determine the design. The observed sample is a realization of this multivariate discrete distribution. Defining survey-sampling estimators as functions of the random sample membership indicators facilitates the study of the large sample properties of current estimators and the derivation of the expression of the variance and estimate of the variance of new estimators. This approach requires familiarity with modern matrix notation and matrix operations, and provides new insights into the performance of estimators without the use of simulations.

The rest of this paper is organized as follows. In Section 2, we summarize the current approach for studying the large sample properties of survey estimators based on the work of Isaki and Fuller (1982). Section 3 presents the ideas for the proposed approach, while

Section 4 and 5 describe the approach in detail. Sections 6 to 10 illustrate the use of the proposed framework to study the large sample properties of the Horvitz-Thompson (HT) estimator and derive the formulas of the variance and variance estimator. Section 11 presents the conclusions and final thoughts.

2. Current Framework for Studying the Asymptotic Properties of Survey Sampling Estimators

Isaki and Fuller (1982) is the seminal paper that established the standard for studying the large-sample properties of estimators in survey sampling theory. Before summarizing their approach, we introduce the concepts and notation of their approach (see Fuller, 2009).

Let $\mathcal{F}$ be the finite population of known size N defined as the entire set $\mathcal{F}=\left(y_1,...,y_N\right)$ where $\mathbf{y}=\left(y_1,...,y_N\right)^{\mathrm{T}}$ is the variable of interest defined for all the elements of $\mathcal{F}$ where each element is identified by the label U_k where $U=\{1,...,N\}$. The population $\mathcal{F}$ is sampled according to a single-stage sample design $p(s)$ where a sample of size n is drawn without replacement. The sample design determines the first-order probability of inclusion of each unit of the frame denoted as $\pi_k \in (0,1)$ for $k \in U$ and $\boldsymbol{\Delta}=\left[\pi_{kl}-\pi_k\pi_l\right]\in\mathbb{R}^{N\times N}$ is the variance-covariance matrix of the sample design where $\pi_{kl}\in\mathbb{R}$ is the second-order probability of inclusion of elements k and l defined as the probability that the 2-tuple (k,l) are both selected in the sample.

We are interested in estimating the population total of the y, defined as $Y=\sum_{k\in U} y_k$. In the simplest case (without the use of any auxiliary information from the population), we can estimate the total Y using the HT estimator defined as $\hat{Y}_{HT}=\sum_{k\in U} d_k y_k s_k$ where d_k is the sampling weight computed as $d_k=\pi_k^{-1}$ and $s_k\in\{0,1\}$ is the sample membership indicator where $s_k=1$ if the unit k is selected in the sample and $s_k=0$ otherwise (Horvitz and Thompson, 1952).

In the Isaki and Fuller (1982) theoretical framework for the asymptotic analysis of design-based estimators, the existence of an indexed sequence of nested finite populations, $\{\mathcal{F}_N\}_{N=1}^{\infty}$ is assumed, with increasing sizes $N_1,\ldots N_N$ where $\mathcal{F}_1\subset\mathcal{F}_2\subset\ldots\subset\mathcal{F}_N$, $N_1<N_2<\ldots<N_N$, with labels for each element in the population $\{U_N=\{1,...,N_N\}\}_{N=1}^{\infty}$. The framework also assumes that a sequence of associated probability samples $\{A_N\}_{N=1}^{\infty}$ where a sample is drawn from each fine population in the sequence according to a sequence of sample designs $\left\{p_N\left(A_N=a_N\right)\right\}_{N=1}^{\infty}$. Both the finite population size N_N and sample

size n_N increase to infinity but that the ratio is finite, since by definition, $\lim_{\substack{N\to\infty \\ n\to\infty}} \frac{n_N}{N_N} = f_N = f \in (0,1)$.

Isaki and Fuller (1982) show that the HT estimator meets the conditions stated in the following two lemmas:

- Lemma 1: $\hat{Y}_{HT,N} - Y_N = \mathcal{O}_p\left(1/n_N^{\delta}\right)$ for $\delta > 0$.
- Lemma 2: There is a sequence of sample designs $\left\{p\left(A_N = a_N\right)\right\}_{N=1}^{\infty}$ with inclusion probabilities $\pi_{k,N}$ for $k \in (1,\ldots,N)$ such as $\mathbb{E}\left(\hat{Y}_{HT,N} - Y_N\right) = \mathcal{O}(1/n_N)$.

The HT estimator is said to be design consistent because it meets these two conditions. These lemmas are closely related to the definition of design-consistent estimator $\hat{\theta}$ for the finite population parameter θ given by Särndal, Swensson, and Wretman (1992):

- The sequence of the estimator $\hat{\theta}_N$ is asymptotically unbiased for the population parameter $\hat{\theta}_N$ if

$$\lim_{N\to\infty}\left[\mathbb{E}\left(\hat{\theta}_N\right) - \theta_N\right] = 0.$$

- The sequence of the estimator $\hat{\theta}_N$ is consistent for the population parameter $\hat{\theta}_N$ if for any fixed $\varepsilon > 0$,

$$\lim_{N\to\infty} \Pr\left[\left|\hat{\theta}_N - \theta_N\right| > \varepsilon\right] = 0.$$

Although these conditions define design-consistent estimators, whether an estimator is design consistent does not only depend on the sequence of sample designs but also on how the sequence of the outcome $\left\{Y_N = \left\{y_1,\ldots,y_N\right\}\right\}_{N=1}^{\infty}$ is specified as $N \to \infty$ (Särndal, Swensson, and Wretman, 1992).

Breidt and Opsomer (2017) expand this and provide two sufficient conditions for the design consistency of the HT estimator that reflect the limiting behavior of the sequence of the outcome $\left\{Y_N\right\}_{N=1}^{\infty}$:

- Assuming that $\lim_{N\to\infty} \frac{n}{N} = f \in (0,1)$, then for all N, $\min_{k\in U} \pi_k \geq \lambda_1 \geq 0$ and $\limsup_{N\to\infty} n \max_{k,l\in U, k\neq l} \left|\Delta_{kl}\right| < \infty$; and
- The outcome variables y_k for $k \in U$ satisfy $\limsup_{N\to\infty} N^{-1} \sum_{k\in U} y_k^2 < \infty$.

If these conditions are met, then the upper bound of the variance of the HT estimator is

$$\mathbb{V}\left(\hat{Y}_{HT}\right) \leq \frac{1}{N\lambda_1}\sum_{k=1}^{N_N}\frac{y_{Nk}^2}{N} + \frac{\max\limits_{k\neq l\in U_N}\left|\Delta_{N,kl}\right|}{\lambda_1^2}\left(\sum_{k=1}^{N_N}\frac{\left|y_k\right|}{N}\right)^2, \tag{2.1}$$

which converges to zero as $N \to \infty$. In other words, $\mathbb{V}\left(\hat{Y}_{HT}\right) = \mathcal{O}\left(\frac{1}{N}\right)$ (Fuller, 2009).

The previous results have used in the literature as the building block for deriving the asymptotic estimators defined as functions of the HT estimator. This approach does not address the situations where the sequence for the outcome $\left\{Y_N\right\}_{N=1}^{\infty}$ interacts with the sequence of sample designs. For example, in πps sample designs (Särndal, Swensson, and Wretman, 1992), the probability of inclusion π_k may be a function of auxiliary variables related to the outcome y_k. Estimates from sample designs where π_k is proportional to y_k are more efficient. An implicit assumption in (2.1) is that the probabilities of inclusion are independent of the outcome y_k.

3. Alternative Framework for the Study of Survey Sampling Estimators

Most of the literature related to the large sample properties of more complex estimators is based on the results of the HT estimator described in the previous section. However, these proofs are lengthy and technically difficult (Knottnerus, 2009). Many proofs are done using a piecewise approach where the properties of the components of the estimator are analyzed separately using different criteria for upped bounds producing expressions such as equation (2.1) that are be difficult to interpret and to derive for other estimators. The current approach is also not informative for the comparison of the asymptotic properties of different estimators and sample designs. Consequently, most papers include simulation studies to examine their properties empirically.

We propose a different framework for the study of the large sample properties of survey sampling estimators. We combine and extend existing ideas to formalize the definition of the sample design and survey sampling estimators. The proposed framework is based on standard asymptotic theory to derive the statistical properties of finite population estimators. Although in some textbooks this approach is called infinite population theory, this is a misnomer since this theory also address the asymptotic properties of finite discrete random variables (see Polansky, 2011; Lehmann, 1999). The second idea is the extension of the methods from the standard asymptotic theory proposed by Cornfield (1944) for proving the properties of finite population estimators (see Section 2.9 in Cochran, 1977). The proposed framework relies heavily on modern matrix notation and algebra. This idea is suggested by Dol, Steerneman, and Wansbeek (1996), who show the convenience of the use of a number of matrix-algebra results to determine the sufficient conditions for the consistency and the rate of convergence of the HT estimator. We expand the matrix notation to include element-wise operations (i.e., Hadamard operations such as product and division). The use of matrix algebra reduces the derivation of the formulas for expected values and variance to a simple algebraic exercise while providing new insights into the properties of the estimators. Finally, we use Tillé's (2006) approach, which redefines the concept of sample design as a multivariate discrete random vector of the sample membership indicators.

The proposed framework is easily extended and can be used for deriving new estimators, their variances, and variance estimators. Applications of this framework are described in Flores Cervantes (2019). The framework can also be extended to include the effect of nonresponse when it is modeled as a random variable with a well-defined distribution.

4. Alternative Definition of a Sample Design

We begin by defining single-stage sample designs where the sample is drawn without replacement. Any sample design can be uniquely defined as follows:

Let $\mathbf{S} \in \{0,1\}^N$ be a vector-valued random variable with a discrete multivariate distribution consisting of N random sample membership indicators $\mathbf{S} = (S_1,...,S_N)^{\mathrm{T}}$, with $S_k = s_k$ where s_k is the realization of S_k and $S_k = s_k = 1$ if the unit k is selected in the sample, and $S_k = s_k = 0$ otherwise. The probability mass function (*pmf*) of $\mathbf{S}$ for a single-stage sample design without replacement is

$$\mathbf{p}(\mathbf{S} = \mathbf{s}) = \exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{S} - \alpha(\lambda Q)\right), \tag{4.1}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^N$, Q is the support of $\mathbf{p}(\mathbf{S})$, and $\alpha(\lambda Q)$ is a function that ensures that the cumulative of $\mathbf{p}(\mathbf{S} = \mathbf{s})$ is one. The expected value of $\mathbf{S}$, $\boldsymbol{\pi} \in (0,1)^{N\times 1}$, is

$$\boldsymbol{\pi} = \mathbb{E}(\mathbf{S}) = \left(\sum_{S_1\in\{0,1\}} S_1 p(S_1),..., \sum_{S_N\in\{0,1\}} S_N\, p(S_N)\right)^{\mathrm{T}} = (\pi_1,...,\pi_N)^{\mathrm{T}},$$

where $\boldsymbol{\pi} = [\pi_k] \in (0,1]^N$ is the vector of the first-order inclusion probabilities $\pi_k > 0$ [1] for $k \in U$. The second moment of $\mathbf{S}$, $\boldsymbol{\Pi} \in \mathbb{R}^{N\times N}$, as

$$\boldsymbol{\Pi} = \mathbb{E}\left(\mathbf{S}\mathbf{S}^{\mathrm{T}}\right) = [\pi_{kl}],$$

where $\boldsymbol{\Pi}$ is the matrix with the second-order probabilities of inclusion π_{kl} of elements k and l. Combining the previous results, the variance-covariance matrix of $\mathbf{S}$, $\boldsymbol{\Delta} \in \mathbb{R}^{N\times N}$, is

$$\begin{aligned}\boldsymbol{\Delta} &= \mathbb{E}\left(\mathbf{S}\mathbf{S}^{\mathrm{T}}\right) - \mathbb{E}(\mathbf{S})\mathbb{E}(\mathbf{S})^{\mathrm{T}} \\ &= \boldsymbol{\Pi} - \boldsymbol{\pi}\boldsymbol{\pi}^{\mathrm{T}}\end{aligned},$$

[1] In order to be a Lebesgue measure, $\pi_k > 0$.

where $\boldsymbol{\Delta} = [\Delta_{kl}] = [\pi_{kl} - \pi_k \pi_l]$ for $k, l \in U$.

Not all multivariate discrete random vectors $\mathbf{S}$ are useful sample designs. We are interested in those random vectors $\mathbf{S}$ such that $\boldsymbol{\pi}$ and $\boldsymbol{\Pi}$, $0 < \pi_k \le 1$ for $k \in U$ and $\pi_{kl} > 0$ in $\boldsymbol{\Pi}$. These two conditions define an *estimable design*. These conditions are needed because the survey estimators (including variance estimators) are expanded by the inverse of π_k and π_{kl} of the sampled units. For example, the HT estimator and its variance estimator are

$$\hat{Y}_{HT} = (\mathbf{y} \oslash \boldsymbol{\pi})^{\mathrm{T}} \mathbf{s} \text{ and } \mathbb{V}\left(\hat{Y}_{HT}\right) = (\mathbf{y} \odot \mathbf{s} \oslash \boldsymbol{\pi})(\boldsymbol{\Delta} \oslash \boldsymbol{\Pi})(\mathbf{y} \odot \mathbf{s} \oslash \boldsymbol{\pi})^{\mathrm{T}},$$

where the operators $\odot$ and $\oslash$ are the Hadamard-Schur or element-wise matrix product and division (Horn and Johnson, 2013). The definition of an estimable design is similar to the concept of a *measurable design* (Särndal, Swensson, and Wretman, 1992).

In order to determine the regularity condition of the large sample properties of the estimators, we rely on the properties of the covariance matrix $\boldsymbol{\Delta}$ of estimable designs. $\boldsymbol{\Delta}$ is a Hermitian matrix with the following properties:

(a.) A real (square) symmetric matrix;
(b.) A normal matrix such that $\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathrm{T}} = \boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Delta}$;
(c.) A matrix that can be diagonalized by a unitary matrix with real elements on the diagonal (finite-dimensional spectral theorem); and
(d.) A matrix with real and linearly independent eigenvalues.

Additional properties of the variance-covariance matrix $\boldsymbol{\Delta}$ depend on the type of sample designs as described in the next section.

Example 1:

A commonly used sampling design is systematic sampling (Cochran, 1977), however, although there is a *pmf* that defines systematic sampling, there are 2-tuples (k, l) with the second-order probabilities of selection with 0 values. In other words, there are pairs of units that are never selected in the same sample. Although the HT estimator $\hat{Y}_{HT} = \sum_{k=U} s_k d_k y_k$ of the total of the population $Y = \sum_{k=U} y_k$ for a systematic sample design is defined, the estimator of the variance cannot be computed. In practice, practitioners assumed that the sample is drawn with replacement. In this case, the assumed design has a *pmf* with the off-diagonals of $\boldsymbol{\Pi}$ defined as $\pi_{kl} = \pi_k \pi_l$. The elements of the Hadamard division $\boldsymbol{\Delta} \oslash \boldsymbol{\pi}$ is $\frac{\Delta_{kl}}{\Pi_{kl}} = \frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}}$ when the element is not in the diagonal is 0, or $\frac{\Delta_{kk}}{\prod_{kk}} = \frac{\pi_k (1 - \pi_k)}{\pi_k} = 1 - \pi_k$ otherwise.

Example 2:

Suppose there is a sample design defined by the *pmf* of $\mathbf{S}$ for a population of size $N=100$, where the $\mathbb{E}\left(S_k\right)=0$ for $k \geq 90$. In this example, the 90th to 100th units of the population do not have a positive value of selections. The HT estimator for the total $Y=\sum_{k \in U} Y_k$ is not defined (i.e., division by zero for units with $\pi_k=0$) and this design is estimable. In practice, the HT estimator can be computed for the subset of units with positive values; however, in this case, the HT estimator produces an estimate of the total $Y^*=\sum_{k \in\{1,\ldots,89\}} Y_k$, which is a biased estimator of Y.

5. Types of Sample Designs

We use the variance of the sum of the elements of $\mathbf{S}$ to classify the sample designs. Let $Z: \mathbb{R}^N \mapsto \mathbb{R}$ be the vector-to-scalar function defined as $Z=Z(\mathbf{S})=\mathbf{1}^{\mathrm{T}} \mathbf{S}$. The variance of Z is $\mathbb{V}(Z)=\mathbf{1}^{\mathrm{T}} \mathbf{\Delta} \mathbf{1}$, which is directly derived using the standard rules for variance computation for random vectors (see Gallager, 2013). In this case, $Z(\mathbf{S})$ is a linear function of S_k for $k \in U$ since it can be written as the linear combination of the elements of the random vector $\mathbf{S}$ as

$$Z=\sum_{k \in U} a_k S_k = a_1 S_1+\ldots+a_N S_N,$$

where $a_k=1$ for $k \in U$.

The variance of the sample size, $\mathbb{V}(Z)$, corresponds to the grand sum of $\mathbf{\Delta}$ (i.e., the sum of all the elements of $\mathbf{\Delta}$) and can be decomposed as the sum of the contribution of the variances and covariance of the elements S_k as

$$\mathbb{V}(Z)=\mathbf{1}^{\mathrm{T}} \mathbf{\Delta} \mathbf{1}=\sum_{k \in U} \mathbb{V}\left(S_k\right)+\sum_{k, l \in U, k \neq l} \mathbb{C}\left(S_k, S_l\right). \tag{5.1}$$

This expression has an intuitive meaning. Each element of $\mathbf{S}$, S_k, contributes to the total variance $\mathbb{V}(Z)$ through the variance component, $\mathbb{V}\left(S_k\right)$, and the sum of the covariance with the other elements $\sum_{l \in U, k \neq l} \mathbb{C}\left(S_k, S_l\right)$. When the sample design is Bernoulli or Poisson sampling, then $\mathbb{C}\left(S_k, S_l\right)=0$ for all $k \neq l \in U$. In this case, the variance of the sample size is the sum of the contributions of the variance of each unit in the frame as $\mathbb{V}(Z)=\sum_{k \in U} \pi_k\left(1-\pi_k\right)$.

The random vector $\mathbf{S}$ represents a fixed sample size design if $\mathbb{V}\left(\mathbf{1}^{\mathrm{T}}\mathbf{S}\right)=0$. Some examples of fixed sample size designs are simple random sampling, Sampford, Midzuno-Sen, Sumter, and Tillé sampling (see Tillé, 2006). These designs have the following properties:

(a) $\boldsymbol{\Delta}$ is positive semidefinite.

(b) If $\lambda_{\min}(\boldsymbol{\Delta}) \le \lambda_{N-1} \le \ldots \le \lambda_2 \le \lambda_{\max}(\boldsymbol{\Delta})$ are the ordered eigenvalues of $\boldsymbol{\Delta}$, then $\lambda_{\min}(\boldsymbol{\Delta})=0$; that is, the eigenvalues $\lambda_k(\boldsymbol{\Delta})$ for $k \in U$ are nonnegative.

(c) $\mathbf{1}\mathrm{row}_k\,\boldsymbol{\Delta}=0$ and $\mathbf{1}^{\mathrm{T}}\mathrm{col}_k\,\boldsymbol{\Delta}=0$ for $k \in U$, and $\mathrm{Tr}(\mathbf{I}\boldsymbol{\Delta})=0$— that is, the sums of rows, the sum of columns, and the total sum of the elements of $\boldsymbol{\Delta}$ — is zero.

(d) The sample size is $n=\mathbf{1}^{\mathrm{T}}\boldsymbol{\pi}$.

When the sample size is fixed, the sum of the diagonal of $\boldsymbol{\Delta}$, $\sum_{k\in U}\pi_k(1-\pi_k)$, has the same value as the sum of the off-diagonal elements, $\sum_{k\in U}\sum_{l\in U}(\pi_{kl}-\pi_k\pi_l)$. This equality can be proved using the properties of the variance-covariance matrix $\boldsymbol{\Delta}$ for fixed-size sample designs listed above.

The discrete random vector $\mathbf{S}$ with parameters $\mathbb{E}(\mathbf{S})=\boldsymbol{\pi}$ and $\mathbb{C}(\mathbf{S})=\boldsymbol{\Delta}$ is a random sample size design if $\mathbb{V}\left(\mathbf{1}^{\mathrm{T}}\mathbf{S}\right)\neq 0$. Some examples of random size designs are Bernoulli, and Poisson sampling, see Tillé (2006). Although this type of sampling is less frequently implemented in practice, random size designs are especially useful for modeling nonresponse. The properties of the random sample size designs are:

(a) $\boldsymbol{\Delta}$ is positive definite with all eigenvalues $\lambda_k(\boldsymbol{\Delta})>0$ for $k \in U$.

(b) $\boldsymbol{\Delta}=\mathrm{diag}(\boldsymbol{\pi})$ because $\pi_{kl}=\pi_k\pi_l$ in $\boldsymbol{\Delta}$ for $k,l \in U : k \neq l$.

(c) The row and column sums are $\mathbf{1}^{\mathrm{T}}\mathrm{row}_k\,\boldsymbol{\Delta}=\pi_k$, $\mathbf{1}\mathrm{col}_l\,\boldsymbol{\Delta}=\pi_l$ for $k,l \in U$, and $\mathrm{Tr}(\mathbf{I}\boldsymbol{\Delta})=n$, where n is the expected sample size, $n=\mathbb{E}\left(\mathbf{1}^{\mathrm{T}}\mathbf{S}\right)$.

(d) $\mathbb{V}\left(\mathbf{1}^{\mathrm{T}}\mathbf{S}\right)=\mathbf{1}^{\mathrm{T}}\boldsymbol{\Delta}\mathbf{1}=\mathbf{1}^{\mathrm{T}}\left(\boldsymbol{\pi}\odot(1-\boldsymbol{\pi})\right)$.

(e) Let $\mathbf{s}\in\{0,1\}^{N\times 1}$ be the vector of one realization of $\mathbf{S}$, $\mathbf{S}=\mathbf{s}$, then the observed sample size n_o is $n_o=\mathbf{1}^{\mathrm{T}}\mathbf{s}$.

(f) If $\lambda_{\min}(\boldsymbol{\Delta}) \le \lambda_{N-1} \le \ldots \le \lambda_2 \le \lambda_{\max}(\boldsymbol{\Delta})$ are the ordered eigenvalues of the variance-covariance matrix $\boldsymbol{\Delta}$, then the eigenvalues are the first-order probability of inclusion $\boldsymbol{\pi}$. The largest eigenvalue of $\boldsymbol{\Delta}$ is $\lambda_{\max}(\boldsymbol{\Delta})=\arg\max_{k\in U}\{\pi_k\}$.

6. Functions of the Random Vector S

We now explore two basic functions of the random vector $\mathbf{S}$ using results from multivariate standard statistical limit theory.

6.1 Function for the Mean Vector of Random Vectors

Let $\mathbf{Z}:\mathbb{R}^N \to \mathbb{R}^N$ be a vector-valued function defined as $\mathbf{Z}(\mathbf{S}) = \frac{1}{N}\sum_{k=1}^{N}\mathbf{S}_k$ where $\mathbf{S}_k$ is the k-th realization of the random vector $\mathbf{S}$ for $k \in \{1,...,N\}$. The random vector $\mathbf{Z}$ is the average of N vectors $\mathbf{S}_k$ of size N each. This function is a typical example found in statistical limit theory textbooks (e.g., Polansky, 2011). Define $\{\mathbf{Z}_N\}_{N=1}^{\infty}$ as the sequence of estimators $\mathbf{Z}$. Then

(a) $\lim_{N\to\infty} \mathbb{E}(\mathbf{Z}_N \mid \mathcal{F}) = \boldsymbol{\pi}$.

(b) $\mathbb{V}(\mathbf{Z}_N \mid \mathcal{F})$ is bounded, $\mathbb{V}(\mathbf{Z}_N \mid \mathcal{F}) = \mathcal{O}\left(\frac{1}{N}\right)$.

(c) Following from (a) and (b), the sequence of estimators $\{\mathbf{Z}_N\}_{N=1}^{\infty}$ is consistent for $\boldsymbol{\pi}$ (weak convergence; Polansky, 2011).

6.2 Function for the Mean of the Elements of the Random Vector S

The second function is defined in terms of the elements of a single vector Z that increases in size N in the sequence as $N \to \infty$. This is in contrast to the first function in Section 6.1, where the number and size of the averaged vectors increases as the population size increases. Let $\mathrm{Z}:\mathbb{R}^N \to \mathbb{R}$ be a vector-to-scalar valued function $\mathrm{Z}(\mathbf{S}) = \frac{1}{N}\mathbf{1}^{\mathrm{T}}\mathbf{S}$. This function differs from the one in the previous section because Z is the average of the N elements S_k of a single realization of $\mathbf{S}$. The function Z is the expected overall sampling rate. To study the asymptotic properties of Z, let $\{Z_N\}_{N=1}^{\infty}$ be the sequence of estimators Z. The expected value and variance of this sequence are

$$\mathbb{E}(Z_N) = \frac{1}{N}\mathbf{1}_N^{\mathrm{T}}\boldsymbol{\pi}_N \text{, and} \tag{6.1}$$

$$\mathbb{V}(Z_N) = \frac{1}{N^2}\mathbf{1}_N^{\mathrm{T}}\boldsymbol{\Delta}_N\mathbf{1}_N . \tag{6.2}$$

The function $\mathrm{Z}(\mathbf{S})$ is not as common and the elements $S_k \in \mathbf{S}$ are not required to have the same expected value, that is, $\mathbb{E}(S_k) \neq \mathbb{E}(S_l)$ for $k \neq l$ and $k,l \in U$, and the 2-tuples (k,l) can be correlated (they are not independent).

In our approach, we can simplify these expressions depending on the type of sample design. If $\mathbf{S}$ is a fixed sample size design, then $\mathbb{E}(Z)=\frac{1}{N}\sum_{k\in U}\pi_k=\frac{n}{N}=f$, where $n=\sum_{k\in N}\pi_k$ is the sample size, and f is the overall sampling rate. In these designs, because $\mathbb{V}(Z)=0$, there is no need to find an upper bound for the sequence of estimators $\{Z_N\}_{N=1}^{\infty}$.

On the other hand, if $\mathbf{S}$ is a random sample size design, then the sequence $\{Z_N\}_{N=1}^{\infty}$ converges to the expected sample size n. To obtain the upper bound of the variance $\mathbb{V}(Z_N)$, we apply the standard rules for variances of random vectors, inequalities for quadratic forms of Hermitian matrices, and inequalities for eigenvalues in terms of matrix norms. So

$$\begin{aligned}\mathbb{V}(Z_N)&=\frac{1}{N^2}\mathbf{1}_N^{\mathrm{T}}\mathbf{\Delta}_N\mathbf{1}_N\\&=\frac{1}{N^2}Q_{\mathbf{\Delta}_N}(\mathbf{1}_N)\\&\le\frac{1}{N^2}\lambda_{\max}(\mathbf{\Delta}_N)\|\mathbf{1}_N\|_2^2=\frac{\lambda_{\max}(\mathbf{\Delta}_N)}{N}\end{aligned},\qquad(6.3)$$

where $Q_{\mathbf{\Delta}_N}(\mathbf{1}_N)=\mathbf{1}_N^{\mathrm{T}}\mathbf{\Delta}_N\mathbf{1}_N$ is the quadratic form of the vector $\mathbf{1}_N$ with respect to the matrix $\mathbf{\Delta}_N$, $\lambda_{\max}(\mathbf{\Delta}_N)$ is the maximum eigenvalue of the matrix $\mathbf{\Delta}_N$, and $\|\mathbf{1}_N\|_2^2$ is the squared L^2-norm of the vector $\mathbf{1}_N$, where $\|\mathbf{1}_N\|_2^2=\sum_{k\in N}1^2=N$. The variance $\mathbb{V}(Z_N)$ of the sampling rate is bounded by a function that depends on the largest eigenvalue of $\mathbf{\Delta}_N$, $\lambda_{\max}(\mathbf{\Delta}_N)$.

In sample designs where the sample draws are independent (e.g., for $k\neq l, k,l\in U$), then the variance-covariance matrix is $\mathbf{\Delta}_N=\mathrm{diag}\left(\boldsymbol{\pi}_N\odot(\mathbf{1}_N-\boldsymbol{\pi}_N)\right)$. Since for diagonal matrices the eigenvalues are the elements of the diagonal, the largest or maximum eigenvalue is

$$\lambda_{\max}(\mathbf{\Delta}_N)=\max_{k\in U_N}\arg\{\Delta_{N,kk}\}=\max_{k\in U_N}\arg\left\{\pi_{N,k}(1-\pi_{N,k})\right\}.\qquad(6.4)$$

The eigenvalue $\lambda_{\max}(\mathbf{\Delta}_N)$ is a function of $\pi_{N,k}$. Sometimes it is desirable to have a bound that does not depend on the first-order inclusion probabilities. This upper bound can be found by noticing that $\lambda_{\max}(\mathbf{\Delta}_N)$ is the variance of a random variable with a Bernoulli distribution with the parameter $\pi=\max_{k\in U_N}\arg\left\{\pi_{N,k}(1-\pi_{N,k})\right\}$. Since the possible values of

π are constrained between 0 and 1, then $\pi(1-\pi)$ has a global maximum at $\pi = \frac{1}{2}$. Then, the upper bound of the variance of the sequence $\{Z_N\}_{N=1}^{\infty}$ is

$$\mathbb{V}(Z_N) \le \frac{K_N}{N} = \mathcal{O}\left(\frac{1}{N}\right), \tag{6.5}$$

where $K_N = 0.5$. The expression (6.5) assumes that $K_N = \mathcal{O}(1)$, which is true since, by definition, the vector $\mathbf{S}$ is an estimable design where $\lim_{N\to\infty} \lambda_{\max}(\mathbf{\Delta}_N) > 0$ and $\lim_{N\to\infty} \lambda_{\max}(\mathbf{\Delta}_N) < 1$ for any N.

7. Linear Functions of the Elements of the Random Vector S

We now introduce a constant vector $\mathbf{a} \in \mathbb{R}^N$ in the function Z. Let $\mathbf{a} = [a_k] \in \mathbb{R}^N$ be a vector of constants, and let $Z : \mathbb{R}^N \to \mathbb{R}$ be the function of $\mathbf{S}$ defined as $Z(\mathbf{S}) = \frac{1}{N}\mathbf{a}^{\mathrm{T}}\mathbf{S}$, the linear combination of the sample membership indicators s_k for $k \in U$ of the form

$$Z(\mathbf{S}) = \frac{1}{N} a_1 s_1 + \ldots + \frac{1}{N} a_N s_N .$$

To study the asymptotic properties of this estimator, we define the sequence of estimators $\{Z_N\}_{N=1}^{\infty}$ and apply the rules used in Section 6. The expected value and variance of the sequence $\{Z_N\}_{N=1}^{\infty}$ are

$$\mathbb{E}(Z_N) = \frac{1}{N}\mathbf{a}_N^{\mathrm{T}}\boldsymbol{\pi}_N, \text{ and} \tag{7.1}$$

$$\mathbb{V}(Z_N) = \frac{1}{N^2}\mathbf{a}_N^{\mathrm{T}}\mathbf{\Delta}_N\mathbf{a}_N = \frac{1}{N^2}\mathrm{Q}_{\mathbf{\Delta}_N}(\mathbf{a}_N) \le \frac{\lambda_{\max}(\Delta_N)}{N}\frac{\|\mathbf{a}_N\|_2^2}{N}, \tag{7.2}$$

where $\|\mathbf{a}_N\|_2^2$ is the square of the L^2-norm of $\mathbf{a}_N$, $\|\mathbf{a}_N\|_2^2 = \sum_{k\in N} a_{Nk}^2$. The upper bound of $\mathbb{V}(Z_N)$ is a function of the largest eigenvalue of the variance-covariance matrix $\mathbf{\Delta}_N$. Replacing $\lambda_{\max}(\Delta_N)$ by $K_N \ge \lambda_{\max}(\Delta_N)$, the upper bound is

$$\mathbb{V}(Z_N) \le \frac{K_N}{N}\frac{\|\mathbf{a}_N\|^2}{N},$$

where K_N can be any vector-induced matrix norms for $\mathbf{\Delta}_N$. Instead of using the matrix-induced norms, we can use the largest eigenvalue of $\mathbf{\Delta}_N$ directly for these designs since the variance-covariance matrix is $\mathbf{\Delta}_N = \text{diag}\left(\boldsymbol{\pi}_N \odot \left(1-\boldsymbol{\pi}_N\right)\right)$. The largest eigenvalue of $\mathbf{\Delta}_N$ is

$$\lambda_{\max}\left(\Delta_N\right) = \max_{k \in U_N} \arg\left\{\left[\mathbf{\Delta}_{Nkk}\right]\right\} = \max_{k \in U_N} \arg\left\{\left[\pi_{Nk}\left(1-\pi_{Nk}\right)\right]\right\}. \tag{7.3}$$

As in the previous section, we can use an upper bound of the largest eigenvalue noting that $\lambda_{\max}\left(\Delta_N\right)$ is the variance of a Bernoulli random variable with the parameter $\pi = \max_{k \in U_N} \arg\left\{\pi_{N,k}\left(1-\pi_{N,k}\right)\right\}$, which has a global maximum at $\pi = \frac{1}{2}$. After combining these results, the upper bound of the sequence of estimators $Z\left(\mathbf{S}\right)$ is

$$\mathbb{V}\left(Z_N\right) \leq \frac{K_N}{N} \frac{\left\|\mathbf{a}_N\right\|^2}{N} = \mathcal{O}\left(\frac{1}{N}\right) \mathcal{O}\left(1\right) = \mathcal{O}\left(\frac{1}{N}\right), \tag{7.4}$$

where $K_N = \frac{1}{2}$. The upper bound of $Z\left(\mathbf{S}\right)$ is of the order $\mathcal{O}\left(\frac{1}{N}\right)$ after applying Slutsky's theorem and assuming that $\frac{\left\|\mathbf{a}_N\right\|_2^2}{N} = \mathcal{O}\left(1\right)$. The expression (7.4) assumes that $\lambda_{\max}\left(\Delta_N\right) = \mathcal{O}\left(1\right)$ as $N \to \infty$, which is true since, by definition, $\mathbf{S}$ is an estimable design.

8. The Horvitz-Thompson Estimator as a Linear Function of the Elements of the Random Vector S

We have already derived some of the properties of the HT estimator of the population mean $\bar{Y} = \frac{1}{N}\mathbf{1}^{\mathrm{T}}\mathbf{y}$, defined as

$$\hat{\bar{Y}}_{HT} = Z\left(\mathbf{S}\right) = \frac{1}{N}\left(\mathbf{d} \odot \mathbf{y}\right)^{\mathrm{T}} \mathbf{S}, \tag{8.1}$$

because $\hat{\bar{Y}}_{HT}$ is the linear function $Z\left(\mathbf{S}\right)$ from the previous section but with the vector of constants $\mathbf{a}$ defined as $\mathbf{a} = \mathbf{d} \odot \mathbf{y}$, where $\mathbf{d} \in \mathbb{R}^{N \times 1}$ is the vector with the sampling weights

$$\mathbf{d} = \mathbf{1} \oslash \mathbb{E}\left(\mathbf{S}\right) = \mathbf{1} \oslash \boldsymbol{\pi} = \boldsymbol{\pi}^{\odot -1} = \left[d_k\right] = \left[\pi_k^{-1}\right],$$

where $\boldsymbol{\pi}^{\odot -1}$ is the Hadamard inverse of $\boldsymbol{\pi}$ or the Hadamard division of $\mathbf{1}$ by $\boldsymbol{\pi}$, and $\mathbf{y} \in \mathbb{R}^{N \times 1}$ is the vector with the outcome $\mathbf{y} = \left[y_k\right]$ for $k \in U$. Let $\left\{\hat{\bar{Y}}_{HT,N}\right\}_{N=1}^{\infty}$ be the sequence of HT estimators defined in (8.1); then the expected value and variance are

$$\mathbb{E}\left(\hat{\bar{Y}}_{HT,N}\right) = \frac{1}{N}\mathbf{1}_N^{\mathrm{T}} \mathbf{y}_N = \bar{Y} \text{ and} \tag{8.2}$$

$$\mathbb{V}\left(\hat{\bar{Y}}_{HT,N}\right)=\frac{1}{N^2}\left(\mathbf{d}_N\odot\mathbf{y}_N\right)^{\mathrm{T}}\Delta_N\left(\mathbf{d}_N\odot\mathbf{y}_N\right). \tag{8.3}$$

To study the large sample properties of the variance of HT estimator in (8.3), we reparametrize it using the variable $\breve{\mathbf{S}}$, defined as follows:

Let $\breve{\mathbf{S}}:\mathbb{R}^N\to\mathbb{R}^N$ be a vector-to-vector valued function of $\mathbf{S}$, where $\breve{\mathbf{S}}=\mathbf{d}\odot\mathbf{S}$. The expected value of $\breve{\mathbf{S}}$ is

$$\mathbb{E}\left(\breve{\mathbf{S}}\right)=\mathbf{d}\odot\mathbb{E}\left(\mathbf{S}\right)=\mathbf{d}\odot\boldsymbol{\pi}=\mathbf{1}. \tag{8.4}$$

Since $\breve{\mathbf{S}}$ is a random vector, we can compute the variance-covariance matrix of $\breve{\mathbf{S}}$, $\boldsymbol{\Delta}_{\breve{\mathbf{S}}}\in\mathbb{R}^{N\times N}$, as

$$\begin{aligned}\mathbb{V}\left(\breve{\mathbf{S}}\right)&=\boldsymbol{\Delta}_{\breve{\mathbf{S}}}=\mathbf{d}^{\mathrm{T}}\mathbb{V}\left(\mathbf{S}\right)\mathbf{d}\\&=\mathbf{d}\boldsymbol{\Delta}\mathbf{d}^{\mathrm{T}}=\boldsymbol{\Delta}\oslash\mathbf{d}^{\odot 2}=\left[\frac{d_k d_l}{d_{kl}}-1\right],\end{aligned} \tag{8.5}$$

where $\mathbf{d}^{\odot 2}$ is the Hadamard product $\mathbf{d}\odot\mathbf{d}$. The variance of the sequence of HT estimators, $\left\{\hat{\bar{Y}}_{HT,N}\right\}_{N=1}^{\infty}$, can be rewritten in terms of $\breve{\mathbf{S}}$ as

$$\begin{aligned}\mathbb{V}\left(\bar{Y}_{HT,N}\right)&=\frac{1}{N^2}\mathbf{y}_N^{\mathrm{T}}\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}\mathbf{y}_N=\frac{1}{N^2}\mathbf{Q}_{\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}}\left(\mathbf{y}_N\right)\\&\le\frac{\lambda_{\max}\left(\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}\right)}{N}\frac{\left\|\mathbf{y}_N\right\|_2^2}{N}\end{aligned}, \tag{8.6}$$

where $\mathbf{Q}_{\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}}\left(\mathbf{y}_N\right)$ is the quadratic form of the vector $\mathbf{y}_N$ with respect to the matrix $\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}$, and $\lambda_{\max}\left(\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}\right)$ is the largest eigenvalue of the matrix $\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}$. Then the upper bound of the sequence of HT estimators is a function of $\lambda_{\max}\left(\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}\right)$, the largest eigenvalue of the reparametrized covariance matrix $\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}$. We can refine the upper bound by replacing $\lambda_{\max}\left(\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}\right)$ by its upper bound $K_N\ge\lambda_{\max}\left(\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}\right)$ using any of the matrix norms induced by the vector 1-norm, ∞-norm, or Frobenius norm as

$$K_N = \begin{cases} \left\| \boldsymbol{\Delta}_{\breve{\mathbf{S}},N} \right\|_1 = \max_{l \in U_N} \sum_{k=1}^{N} \left| \Delta_{\breve{\mathbf{S}},N,kl} \right| = \max_{l \in U_N} \sum_{k=1}^{N} \left| d_{Nk} d_{Nl} d_{Nkl}^{-1} - 1 \right| & \text{1- norm} \\ \left\| \boldsymbol{\Delta}_{\breve{\mathbf{S}},N} \right\|_\infty = \max_{k \in U_N} \sum_{l=1}^{N} \left| \Delta_{\breve{\mathbf{S}},N,kl} \right| = \max_{k \in U_N} \sum_{l=1}^{N} \left| d_{Nk} d_{Nl} d_{Nkl}^{-1} - 1 \right| & \infty\text{-norm} \\ \left\| \boldsymbol{\Delta}_{\breve{\mathbf{S}},N} \right\|_F = \left[\operatorname{tr}\left(\boldsymbol{\Delta}_{\breve{\mathbf{S}},N}^{\mathrm{T}} \boldsymbol{\Delta}_{\breve{\mathbf{S}},N} \right) \right]^{1/2} & \text{Frobenius norm} \end{cases} .$$

For random sample size designs, the matrix norm K_N can be simplified to

$$K_N = \underset{k \in U_N}{\arg\max} \left\{ \Delta_{\breve{\mathbf{S}},N} \right\} = d_{N\max} - 1 > d_{N\max} ,$$

since $\boldsymbol{\Delta}_{\breve{\mathbf{S}},N} = \operatorname{diag}\left(\mathbf{d}_N - 1 \right)$ for these designs. The upper bound of the sequence $\left\{ \hat{\bar{Y}}_{HT,N} \right\}_{N=1}^{\infty}$ is a function of the maximum sampling weight $d_{k,N}$, not the maximum $\pi_{k,N}$ as in the estimator in Section 7. The upper bound of the variance of $\left\{ \hat{\bar{Y}}_{HT,N} \right\}_{N=1}^{\infty}$ is

$$\mathbb{V}\left(\bar{Y}_{HT,N} \right) \le \frac{K_N}{N} \frac{\left\| \mathbf{y}_N \right\|_2^2}{N} = \mathcal{O}\left(\frac{1}{N} \right) \mathcal{O}(1) = \mathcal{O}\left(\frac{1}{N} \right) , \tag{8.7}$$

where $\left\| \mathbf{y}_N \right\|_2^2$ is the square of the Euclidian norm of $\mathbf{y}_N$, $\left\| \mathbf{y}_N \right\|_2^2 = \mathbf{y}_N^{\mathrm{T}} \mathbf{y}_N$. The order of the variance of the HT estimator, $\mathbb{V}\left(\bar{Y}_{HT,N} \right)$, is $\mathcal{O}\left(N^{-1} \right)$ after applying Slutsky's theorem. Two implicit assumptions in (8.7) are $K_N = \mathcal{O}(1)$ and $\frac{\left\| \mathbf{y}_N \right\|_2^2}{N} = \mathcal{O}(1)$.

For random size designs, the matrix-induced norms are functions of the sum of the elements; for example, for the 1-norm, these are sums by rows, $\sum_{k=1}^{N} \left| d_{k,N} d_{l,N} d_{kl,N}^{-1} - 1 \right|$, of $\left\| \boldsymbol{\Delta}_{\breve{\mathbf{S}},N} \right\|_1$. The terms to sum to determine the row with the largest value for the 1-norm are listed in the Table 1.

Table 1: Elements of the matrix $\boldsymbol{\Delta}_{\breve{S},N}$

	Column l			
Row k	1	2	…	N
1	$\left\|d_{1,N}-1\right\|$	$\left\|\frac{d_{1,N}d_{2,N}}{d_{12,N}}-1\right\|$	...	$\left\|\frac{d_{1,N}d_{N,N}}{d_{1N,N}}-1\right\|$
2	$\left\|\frac{d_{2,N}d_{1,N}}{d_{21,N}}-1\right\|$	$\left\|d_{2,N}-1\right\|$	…	$\left\|\frac{d_{2,N}d_{N,N}}{d_{2N,N}}-1\right\|$
…	⋮	⋮	⋮	⋮
N	$\left\|\frac{d_{N,N}d_{1,N}}{d_{N1,N}}-1\right\|$	…	…	$\left\|d_{N,N}-1\right\|$

Since the summands are absolute values, the largest row sum corresponds to the row k with the largest weight $d_{\max}=\arg\max_{k\in U_N}\{\mathbf{d}_N\}$. Then the upper bound or 1-norm is

$$K_{\max,N}=\left|\frac{d_{\max,N}d_{1,N}}{d_{\max 1,N}}-1\right|+...+\left|d_{\max,N}-1\right|+...+\left|\frac{d_{\max,N}d_{N,N}}{d_{\max N,N}}-1\right|.$$

After rewriting the norm in terms of the inclusion probabilities, we obtain

$$K_{\max,N}=\left|d_{\max,N}-1\right|+\left|\frac{\pi_{\max 1,N}}{\pi_{\max,N}\pi_{1,N}}-1\right|+...+\left|\frac{\pi_{\max N,N}}{\pi_{\max,N}\pi_{N,N}}-1\right|.$$

Since we assume that $\mathbf{S}$ is an estimable design where the sums by rows of $\boldsymbol{\Delta}$ are equal to 0 for $N\to\infty$ (see properties of Hermitian matrix for this design in Section 5), then $\pi_{kl,N}$ converges to $\pi_{k,N}\pi_{l,N}$ for all $k\neq l\in U$. This equality is needed in order to maintain the row sum in $\boldsymbol{\Delta}$ as the population size increases. Combining these results, then the upper bound of the 1-norm can be written as

$$K_N>\arg\max_{k\in U_N}\{\mathbf{d}_N\}=d_{\max,N}.$$

After substituting K_N in (8.2), we obtain the same expression in (8.7) for the upper bound of the variance of HT estimator for designs with random sample sizes.

Comparing the expression of the upper bound from Breidt and Opsomer (2017) in (2.1), we notice that the first term matches (8.7), because $0<\lambda_1\leq\min_{k\in U}\pi_k=\max_{k\in U}d_k$. Note that by definition $\mathbf{S}$ is an estimable design, then $\pi_k>0$ for all $k\in U$; therefore there is no need for the lower bound λ_1 to be greater than zero in the first term of the equation (2.1). The second term of (2.1) is not needed because it goes to 0 as $N \to \infty$.

9. The Design Consistency of the Horvitz-Thompson Estimator in Sample Designs When $\boldsymbol{\pi}$ and $\mathbf{y}$ Are Related

A more complete study of the asymptotic properties of an estimator requires examining the limiting behavior of all the quantities that are used to compute the sequence of estimators as $N \to \infty$ and $n \to \infty$. This is where the main difference arises between the current approach and our approach. First, we do not consider the sequence of sample size separately from the design because, by definition, the sample size is determined by the design as the sum of the probabilities of inclusion $\sum_{k \in U_N} \pi_{k,N}$. In other words, the increasing sample size when $n \to \infty$ is not arbitrary since it depends by the sample design $\mathbf{S}$. Second, the proposed framework for large simple analysis considers three sequences: the population size, the outcome $\mathbf{y}$, and the design $\mathbf{S}$ through the probability of inclusion (determined by the design) that may be determined by a sequence of auxiliary variables $\mathbf{x}$ related to $\mathbf{y}$. However, to avoid inconsistencies resulting from multiple sequences converging to infinity, we express all tow of these sequences in terms of the order of the population size N; for example, $Y_N = \sum_{k \in U_N} y_{k,N} = \mathcal{O}(N)$ and $\frac{\sum_{k \in N} |x_K|}{N} = \mathcal{O}\left(\frac{1}{N^{1/2}}\right)$. Defining the sequences as functions of the order of the population size N enables us to use the algebraic rules for the order of a function $\mathcal{O}$ –the Bachmann-Landau order operator– when studying the large sample properties of the estimators.

The results for the upper bounds of the variance of the HT estimator from the literature and those presented in the previous section assume that the sample design $\mathbf{S}$ is independent of the outcome sample $\mathbf{y}$. However, in practice samples are designed with the goal of producing efficient estimates (with minimum variance) of the outcome of interest. A stratified, *pps* (probability-proportional-to-size with replacement), or *πps* (π proportional to size without replacement) sample designs may be preferable. These designs make use use of an auxiliary variable $\mathbf{x} \in \mathbb{R}_{>0}^{N \times 1}$, $\mathbf{x} = [x_k]$ that is known for all $k \in U$ that is related to the outcome; for example, $y_k \propto x_k^{\gamma}$ where $\gamma \neq 0$. Sampling design that produces estimators with increased efficiency derive the first-order probabilities of inclusion $\boldsymbol{\pi}$ based on the auxiliary variable $\mathbf{x}$. For example, we can have a very efficient designs when $y_k \propto x_k$ if the probabilities of selection are be defined as

$$\boldsymbol{\pi} = n\mathbf{x} / X \,,$$

where n is the expected sample size $n = \mathbf{1}^{\mathrm{T}} \boldsymbol{\pi}$ and X is the population total of x computed as $X = \mathbf{1}^{\mathrm{T}} \mathbf{x}$.

To study the properties of the HT estimator in designs with random sample sizes when $\mathbf{y}$ and $\boldsymbol{\pi}$ are related, we use the general expression of the upper bound of the variance of the HT estimator in (8.3), without separating the product $\mathbf{y}_N \odot \mathbf{d}_N$ in the quadratic form as

$$\mathbb{V}\left(\bar{Y}_{HT,N}\right) \leq \frac{1}{N^2}\left\|\left(\mathbf{y} \odot \mathbf{d}\right)^{\odot 2} \odot \operatorname{diag}\left(\boldsymbol{\Pi}\right)\right\|^2. \tag{9.1}$$

In the following designs, we assume that $X_N = \sum_{k \in U_N} x_{k,N} = \mathcal{O}\left(N\right)$ and $\boldsymbol{\pi} = n\mathbf{x} / X$; that is, the auxiliary variables are of the same order as the population size.

We examine first the case when $\gamma = 1$ and $y_k = cx_k$. In *πps* designs, designs with fixed sample sizes, the variance of the HT estimator $\mathbb{V}\left(\bar{Y}_{HT,N}\right) = 0$. In designs with random sample sizes, the variance $\mathbb{V}\left(\bar{Y}_{HT,N}\right)$ in terms of y_k after simplification is

$$\mathbb{V}\left(\bar{Y}_{HT,N}\right) \leq \frac{1}{n}\frac{\sum_{k \in U}\left|Y_k\right|}{N} = \mathcal{O}\left(\frac{1}{n}\right), \tag{9.2}$$

which converges to zero since, by definition, $\frac{\sum_{k \in U}\left|Y_k\right|}{N} = \mathcal{O}\left(1\right)$. The sufficient conditions in expression (9.2) are more general than those in Section 2. When $\boldsymbol{\pi}$ is closely related to $\mathbf{y}$, the HT estimator is design consistent as to long as the first moment of the absolute values of y is defined and not the second moment as in (2.1).

We now examine the case where $\gamma = -1$ and $y_k = \frac{c}{x_k}$. The expression of the variance of the HT estimator after simplification is

$$\mathbb{V}\left(\bar{Y}_{HT,N}\right) \leq \frac{1}{n}\frac{\sum_{k \in U}\left|Y_k\right|^3}{N}, \tag{9.3}$$

which converges to zero when $\frac{\sum_{k \in U}\left|Y_k\right|^3}{N} = \mathcal{O}\left(1\right)$. The sufficient condition for the HT estimator, in this case, is more constrained than in expressions (9.2) and in (2.1) in Section 2. When $\boldsymbol{\pi}$ is inversely related to $\mathbf{y}$, the estimator is design-consistent if the third moment of the absolute values of $\mathbf{y}$ is defined.

The results in (9.2) and (9.3) show that the sufficiency conditions for the HT estimator to be design consistent when $\mathbf{y}$ and $\mathbf{x}$ are related depend on both orders of the sequence of the outcome $\mathbf{y}$ and auxiliary information $\mathbf{x}$ when the latter is used to compute the probabilities of inclusion $\boldsymbol{\pi}$. These results are easier to derive using the proposed approach presented in this paper.

10. The Variance Estimator of the Horvitz-Thompson Estimator

In this section, we derive the variance estimator of the HT estimator. The variance estimator of the HT estimator of the mean $\bar{Y}$ is derived from (8.7) after replacing $\boldsymbol{\Delta}$ by $\tilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta} \oslash \boldsymbol{\Pi}$ as

$$\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right) = \frac{1}{N^2}\left(\mathbf{y} \odot \mathbf{d} \odot \mathbf{S}\right)^{\mathrm{T}} \breve{\boldsymbol{\Delta}}\left(\mathbf{y} \odot \mathbf{d} \odot \mathbf{S}\right). \tag{10.1}$$

The matrix $\tilde{\boldsymbol{\Delta}}$ is the sample expanded matrix $\boldsymbol{\Delta}$ which is the element-wise division of $\boldsymbol{\Delta}$ by the probabilities of inclusion in $\boldsymbol{\Pi}$ (i.e., a generalization of the HT estimator for matrices). We reparametrize $\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)$ as a sum of the new variable $\psi_{kl} = \frac{y_k}{\pi_k}\frac{y_l}{\pi_l}\Delta_{kl}$ expanded by π_{kl}, similar to an HT estimator with the variable ψ_{kl} as

$$\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right) = \frac{1}{N^2}\sum_{k\in U}\sum_{l\in U}\frac{\psi_{kl}}{\pi_{kl}}. \tag{10.2}$$

We continue reparametrizing (10.2) using the following variables:

- $\boldsymbol{\psi} \in \mathbb{R}^{N\times N}$, where $\boldsymbol{\psi} = \left(\mathbf{y} \oslash \boldsymbol{\pi}\right)^{\mathrm{T}} \boldsymbol{\Delta}\left(\mathbf{y} \oslash \boldsymbol{\pi}\right)$, the matrix representation of ψ_{kl}.
- $\mathbf{S}_2 \in \mathbb{R}^{N\times N}$, a matrix with the sample membership indicators of the 2-tuples (k,l), where $\mathbb{E}\left(\mathbf{S}_2\right) = \boldsymbol{\Pi}$, the matrix with the second-order probability of inclusion π_{kl}.
- $\boldsymbol{\Delta}_{\mathbf{S}_2} \in \mathbb{R}^{N^2\times N^2}$, the covariance matrix of $\mathbf{S}_2$, where $\boldsymbol{\Delta}_{\mathbf{S}_2} = \left[\pi_{klmn} - \pi_{kl}\pi_{mn}\right]$ and π_{klmn} is the fourth-order inclusion probability of the 4-tuples (k,l,m,n).
- To avoid tensor notation (i.e., multidimensional matrices), we vectorize $\boldsymbol{\psi}$ and $\boldsymbol{\Pi}$ as $\mathrm{vec}\left(\boldsymbol{\psi}\right) \in \mathbb{R}^{N^2}$, $\mathrm{vec}\left(\boldsymbol{\Pi}^{\odot -1}\right) \in \mathbb{R}^{N^2}$ (Magnus and Neudecker, 1999). The expression of $\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)$ with the reparametrized variables is

$$\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right) = \frac{1}{N^2}\mathrm{vec}\left(\boldsymbol{\psi}\right)^{\mathrm{T}} \mathrm{vec}\left(\boldsymbol{\Pi}^{\odot -1} \odot \mathbf{S}_2\right). \tag{10.3}$$

The expected value is

$$\begin{aligned}\mathbb{E}\left(\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)\right) &= \frac{1}{N^2}\operatorname{vec}(\boldsymbol{\psi})^{\mathrm{T}}\,\mathbb{E}\left(\operatorname{vec}\left(\boldsymbol{\Pi}^{\odot -1}\odot\mathbf{S}_2\right)\right)\\ &= \frac{1}{N^2}\operatorname{vec}(\boldsymbol{\psi})^{\mathrm{T}}\operatorname{vec}\left(\boldsymbol{\Pi}^{\odot -1}\odot\mathbb{E}(\mathbf{S}_2)\right)\\ &= \frac{1}{N^2}\operatorname{vec}(\boldsymbol{\psi})^{\mathrm{T}}\operatorname{vec}\left(\boldsymbol{\Pi}^{\odot -1}\odot\boldsymbol{\Pi}\right)\\ &= \frac{1}{N^2}\operatorname{vec}(\boldsymbol{\psi})^{\mathrm{T}}\mathbf{1}_{N^2} = \mathbb{V}\left(\bar{Y}_{HT}\right)\end{aligned}. \quad (10.4)$$

Therefore, $\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)$ is an unbiased estimator of $\mathbb{V}\left(\bar{Y}_{HT}\right)$.

To study the limiting distribution and bounds of the estimator $\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)$ as $N\to\infty$, we derive the expression of $\mathbb{V}\left(\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)\right)$ following the same procedures from the previous sections.

$$\begin{aligned}\mathbb{V}\left(\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)\right) &= \frac{1}{N^4}\mathbb{V}\left(\operatorname{vec}(\boldsymbol{\psi}\oslash\boldsymbol{\Pi})^{\mathrm{T}}\operatorname{vec}(\mathbf{S}_2)\right) =\\ &= \frac{1}{N^4}\operatorname{vec}(\boldsymbol{\psi}\oslash\boldsymbol{\Pi})^{\mathrm{T}}\,\mathbb{V}\left(\operatorname{vec}(\mathbf{S}_2)\right)\operatorname{vec}(\boldsymbol{\psi}\oslash\boldsymbol{\Pi})\\ &= \frac{1}{N^4}\operatorname{vec}(\boldsymbol{\psi}\oslash\boldsymbol{\Pi})^{\mathrm{T}}\boldsymbol{\Delta}_{\mathbf{S}_2}\operatorname{vec}(\boldsymbol{\psi}\oslash\boldsymbol{\Pi})\\ &= \frac{1}{N^4}Q_{\boldsymbol{\Delta}_{\mathbf{S}_2}}\left(\operatorname{vec}(\boldsymbol{\psi}\oslash\boldsymbol{\Pi})\right)\\ &\le \frac{\lambda_{\max}\left(\boldsymbol{\Sigma}_{\mathbf{S}_2}\right)}{N^3}\frac{\|\mathbf{y}\odot\mathbf{y}\|_2^2}{N}\end{aligned},$$

where $\lambda_{\max}\left(\boldsymbol{\Sigma}_{\mathbf{S}_2}\right)$ is the largest eigenvalue of the matrix

$$\boldsymbol{\Sigma}_{\mathbf{S}_2} = \boldsymbol{\Delta}_{\mathbf{S}_2}\odot\boldsymbol{\Delta}^{\odot 2}\oslash\boldsymbol{\pi}^{\odot 2}\oslash\boldsymbol{\Pi}^{\odot 2},$$

with the element $\boldsymbol{\Sigma}_{klmn,\mathbf{S}_2} = \frac{\left(\pi_{kl}-\pi_k\pi_l\right)^2}{\pi_k^2\pi_l^2}\frac{\left(\pi_{klmn}-\pi_{kl}\pi_{mn}\right)}{\pi_{kl}^2}$.

An upper bound $K \ge \lambda_{\max}\left(\boldsymbol{\Sigma}_{\mathbf{S}_2}\right)$ is obtained using the vector induced matrix norms in $\boldsymbol{\Sigma}_{\mathbf{S}_2}$ as

$$K = \begin{cases} \left\| \boldsymbol{\Sigma}_{\mathbf{S}_2} \right\|_1 = \max_{l \in U} \sum_{k=1}^{N} \left| \boldsymbol{\Sigma}_{\mathbf{S}_2,kl} \right| & \text{1- norm} \\ \left\| \boldsymbol{\Sigma}_{\mathbf{S}_2} \right\|_\infty = \max_{k \in U} \sum_{l=1}^{N} \left| \boldsymbol{\Sigma}_{\mathbf{S}_2,kl} \right| & \infty\text{-norm} \\ \left\| \boldsymbol{\Sigma}_{\mathbf{S}_2} \right\|_F = \left[\operatorname{tr}\left(\boldsymbol{\Sigma}_{\mathbf{S}_2,kl}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{S}_2,kl} \right) \right]^{1/2} & \text{Frobenius norm} \end{cases} .$$

The main difficulty of identifying an upper bound for K is that it requires examining the elements of $\boldsymbol{\Sigma}_{\mathbf{S}_2}$. The third- and fourth-order π_{klm} π_{klmn} of inclusion probabilities (π_{klm} and π_{klmn}) are not available or are difficult to compute for some complex designs.

On the other hand, for random sample size designs, we can refine the value of K since $\boldsymbol{\Sigma}_{\mathbf{S}_2}$ is a diagonal matrix, where $\boldsymbol{\Sigma}_{\mathbf{S}_2} = \left[(d_k - 1)^3 \right] = \left[\left(\pi_k^{-1} - 1 \right)^3 \right]$. K is the maximum sampling weight which is equivalent to the smallest π_k. Assuming that $\frac{\|\mathbf{y} \odot \mathbf{y}\|_2^2}{N} = \mathcal{O}(1)$, then, after using Slutsky's theorem,

$$\mathbb{V}\left(\hat{\mathbb{V}}\left(\bar{Y}_{HT} \right) \right) \le \frac{K}{N^3} \frac{\|\mathbf{y} \odot \mathbf{y}\|_2^2}{N} = \mathcal{O}\left(\frac{1}{N^3} \right) \mathcal{O}(1) = \mathcal{O}\left(\frac{1}{N^3} \right). \tag{10.5}$$

$\hat{\mathbb{V}}\left(\bar{Y}_{HT} \right)$ is bounded in probability and $\lim_{N \to \infty} \hat{\mathbb{V}}\left(\bar{Y}_{HT,N} \right) = \lim_{N \to \infty} \mathbb{V}\left(\bar{Y}_{HT,N} \right) = 0$. The expression in (10.5) implicitly assumes that $\frac{\|\mathbf{y} \odot \mathbf{y}\|_2^2}{N}$ is $\mathcal{O}(1)$. This ratio can be written as

$$\frac{\|\mathbf{y} \odot \mathbf{y}\|_2^2}{N} = \frac{\left\| \mathbf{y}^{\odot 2} \right\|_2^2}{N} = \frac{\sum_{k=1}^{N} \left(y_k^2 \right)^2}{N} = \frac{\sum_{k=1}^{N} y_k^4}{N} = \mathcal{O}(1), \tag{10.6}$$

which is the fourth population moment of y.

Breidt and Opsomer (2017) do not provide an explicit expression for the upper bound of $\hat{\mathbb{V}}\left(\bar{Y}_{HT} \right)$ similar to (10.5). However, they list two sufficient conditions, D3 and D4, for design consistency of $\hat{\mathbb{V}}\left(\bar{Y}_{HT} \right)$. The condition D3 is $\arg\min_{k,l \in U_N} \{\pi_{kl}\} \ge \lambda_2 > 0$, which has the parameter λ_2 as a lower bound so the smallest π_{kl} is not zero. This parameter is not needed because the sample designs in the sequence $\{\mathbf{S}_N\}_{N=1}^{\infty}$ are assumed to estimable; therefore $\pi_{kl} > 0$ for all $k, l \in U$. The condition D4 matches equation (10.6).

Unlike the previous section, where we analyzed the sufficient conditions for consistency by examining the relationship between $\mathbf{y}$ and $\boldsymbol{\pi}$, we illustrate how the speed of convergence varies and the situations where $\hat{\mathbb{V}}\left(\bar{Y}_{HT}\right)$ does not become zero as $N \rightarrow \infty$. We begin by substituting x_k^2 by y_k^4 in $\left\|\mathbf{y} \odot \mathbf{y}\right\|_2^2$, then an upper bound of $\hat{\mathbb{V}}\left(\bar{Y}_{HT,N}\right)$, in terms of the population mean $\bar{Y}_N$, is $\hat{\mathbb{V}}\left(\bar{Y}_{HT,N}\right) \leq K\bar{X}_N^2 = K_N N^2 \bar{Y}_N^4$. If we define $\left\{\mathbf{y}_N\right\}_{N=1}^{\infty}$ as a sequence of real constants, $\mathbf{y}_N \in \mathbb{R}^N$ where $\bar{Y}_N = \mathcal{O}\left(N^P\right)$, then the value of p such as $\hat{\mathbb{V}}\left(\bar{Y}_{HT,N}\right)$ does not converge, e.g., $\hat{\mathbb{V}}\left(\bar{Y}_{HT,N}\right) \geq \mathcal{O}_p\left(1\right)$, is $p \geq -\frac{1}{2}$. If $-\frac{3}{4} < p < -\frac{1}{2}$, then $\mathbb{V}\left(\bar{Y}_{HT,N}\right)$ converges at a slower rate than $\mathcal{O}_p\left(N^{-1}\right)$; if $p < -\frac{3}{4}$, $\mathbb{V}\left(\bar{Y}_{HT,N}\right)$ converges at a faster rate than $\mathcal{O}_p\left(N^{-1}\right)$.

11. Final Thoughts

We have presented a systematic framework that facilitates the study of the large properties of the survey sampling estimators by focusing on the sample design as a multivariate random vector with the sample membership indicators with a well-defined *pmf*. The proposed approach not only provides a systematic method for determining the sufficient conditions for design consistency but also facilitates the derivation of new estimators, their variances, and variance estimators. In this framework, all survey estimators are functions (linear or nonlinear) of these elements of the random vector of the membership sample indicators, and standard statistical tools for functions of random variables can be used to study their properties. Furthermore, the proposed framework enables us to extend the sufficient conditions of the large sample properties of the HT estimator and its variance estimator not reported in the literature. We have shown that the sufficient conditions for HT estimators to be design consistent also depend on the relationship between the outcome and the probabilities of inclusion when the later are derived using auxiliary variables related to the outcome. This relationship has not been accounted for in the current literature. Analyses based on the presented approach of more complex estimators such as the Hájek, ratio, GREG (generalized regression), and poststratified estimators, among others, are presented in Flores Cervantes (2019). Future research will address the extension to sample designs with replacement and with multiple stages. Another research area is the extension of the framework to address nonresponse weighting adjustments.

References

Breidt, F. J., and J. D. Opsomer. 2017. Model-assisted survey estimation with modern prediction techniques. *Statistical Science,* 32(2), 190-205.

Cochran, W. 1977. *Sampling techniques* (3rd ed.). New Delhi, India: Wiley & Sons.

Cornfield, J. 1944. On samples from finite populations. *Journal of the American Statistical Association*, 39(226), 236-239. Retrieved from https://www.jstor.org/stable/2279953

Dol, W., T. Steerneman, and T. Wansbeek. 1996. Matrix algebra and sampling theory: the case of the Horvitz-Thompson estimator. *Linear Algebra and Its Applications,* 237/238, 225-238.

Flores Cervantes, I. 2019. *A unifying parametric framework for estimating finite population totals from complex samples (Doctoral dissertation).* College Park, MD: University of Maryland.

Fuller, W. A. 2009. *Sampling statistics.* Hoboken, NJ: John Wiley & Sons.

Gallager, R. G. 2013. *Stochastic processes theory for applications.* Cambridge, England: University Press.

Horn, R., and C. Johnson. 2013. *Matrix analysis* (2nd ed.). New York, NY: Cambridge University Press.

Horvitz, D., and D. Thompson. 1952. A generalisation of sampling without replacement from a finite universe. *Journal of the American Statistical Association,* 47, 663-685.

Isaki, C. T., and W. A. Fuller. 1982. Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77(377), 89-96.

Knottnerus, P. 2009. *On asymptotic distributions in random sampling from finite populations.* The Hague, The Netherlands: Statistics Netherlands.

Lehmann, E. 1999. *Elements of large-sample theory.* New York, NY: Springer-Verlag.

Magnus, J. R., and H. Neudecker. (1999). *Matrix differential calculus with applications in statistics and econometrics.* Chichester, England: Wiley.

Polansky, A. 2011. *Introduction to statistical limit theory.* Boca Raton, FL: Chapman & Hall/CRC.

Särndal, C., B. Swensson, and J. Wretman. 1992. *Model assisted survey sampling.* New York, NY: Springer.

Tillé, Y. 2006. *Sampling algorithms.* Springer.