A note on the consistency of the Narain-Horvitz-Thompson estimator

Guillaume Chauvet *

December 10, 2014

Abstract

For the Narain-Horvitz-Thompson estimator to have usual asymptotic properties such as consistency, some conditions on the sampling design and on the variable of interest are needed. Cardot et al. (2010) give some sufficient conditions for the mean square consistency, but one of them is usually difficult to prove or does not hold for some unequal probability sampling designs. We propose alternative conditions for the mean square consistency of the Narain-Horvitz-Thompson estimator. A specific result is also proved in case when a martingale sampling algorithm is used, which implies consistency under a fast algorithm for the cube method.

Keywords: Cube method; Martingale algorithm; Mean-square consistency; Multinomial sampling; Sen-Yates-Grundy conditions.
1 Introduction

When a random sample $S$ is selected inside a finite population $U$, the Narain (1951)-Horvitz-Thompson (1952) estimator $\hat{t}_y\pi$ if often used for the total $t_y = \sum_{k\in U} y_k$ of some variable of interest. For the Narain-Horvitz-Thompson estimator to have usual asymptotic properties, such as asymptotic normality or consistency, some conditions on the sampling design and on the variable of interest are needed. Following the approach in Robinson and Särndal (1983) and Breidt and Opsomer (2000), Cardot et al. (2010) give sufficient conditions for the mean square consistency. However, one of these conditions is related to the second-order inclusion probabilities and is usually difficult to prove for unequal probability sampling designs.

In this note, we propose alternative conditions for the mean square consistency of the Narain-Horvitz-Thompson estimator, i.e. under which

$$E \left\{ N^{-1}(\hat{t}_y\pi - t_y) \right\}^2 = O(n^{-1})$$

with $N$ the population size. The proposed conditions are usually easier to prove, and are known to hold for several sampling designs with unequal probabilities. We also give conditions under which the Narain-Horvitz-Thompson is consistent in mean square under a martingale sampling algorithm, which implies consistency under a fast algorithm for the cube method (Deville and Tillé, 2004). Our asymptotic framework is that described in Isaki and Fuller (1982). We assume that the population $U$ belongs to a nested sequence $\{U_t\}$ of finite populations with increasing sizes $N_t$, and that the population vector of values $y_{Ut} = (y_{1t}, \ldots, y_{N_t})^T$ belongs to a sequence $\{y_{Ut}\}$ of $N_t$-vectors. For simplicity, the index $t$ will be suppressed in what follows and all limiting processes will be taken as $t \to \infty$. 
2 Finite population framework

We note \( \pi = (\pi_1, \ldots, \pi_N)^\top \) a \( N \)-vector of probabilities. Let \( p(\cdot) \) denote a sampling design in \( U \) with parameter \( \pi \), that is, such that the expected number of draws for unit \( k \) in the sample equals \( \pi_k > 0 \). Let \( n = \sum_{k \in U} \pi_k \) denote the integer average sample size. A random sample \( S \), with or without repetitions, is selected in \( U \) by means of the sampling design \( p(\cdot) \). The total \( t_y \) is unbiasedly estimated by

\[
\hat{t}_y = \sum_{k \in U} \frac{y_k}{\pi_k} I_k, \tag{2}
\]

with \( I = (I_1, \ldots, I_N)^\top \) and \( I_k \) the number of times that unit \( k \) is selected in the sample. The variance of \( \hat{t}_y \) is

\[
V(\hat{t}_y) = \sum_{k, l \in U} \frac{y_k y_l}{\pi_k \pi_l} Cov(I_k, I_l). \tag{3}
\]

If \( p(\cdot) \) is a without-replacement sampling design, a same unit \( k \) may appear only once in the sample and \( I_k \) is a sample membership indicator. Formula (2) yields the Narain-Horvitz-Thompson estimator \( \hat{t}_{y\pi} \) whose variance is

\[
V(\hat{t}_{y\pi}) = \sum_{k \in U} \left( \frac{y_k}{\pi_k} \right)^2 \pi_k (1 - \pi_k) + \sum_{k \neq l \in U} \frac{y_k y_l}{\pi_k \pi_l} (\pi_{kl} - \pi_k \pi_l), \tag{4}
\]

with \( \pi_{kl} \) the probability that units \( k \) and \( l \) are selected jointly in \( S \). Poisson sampling (Hájek, 1964) is a particular without-replacement sampling design, obtained when the vector \( I \) of sample membership indicators is obtained from \( N \) independent Bernoulli trials. In such case, the variance of the Narain-Horvitz-Thompson estimator is

\[
V(\hat{t}_{y\pi}) = \sum_{k \in U} \left( \frac{y_k}{\pi_k} \right)^2 \pi_k (1 - \pi_k), \tag{5}
\]
which is the first term of the variance in (1) for any without-replacement sampling design.

If \( p(\cdot) \) is a with-replacement sampling design, a same unit \( k \) may appear several times in the sample and formula (2) yields the [Hansen and Hurwitz (1953)](#) estimator \( \hat{t}_{yHH} \). Multinomial sampling is a particular with-replacement sampling design, obtained when the sample \( S \) is obtained from \( n \) independent draws, some unit \( k \) being selected with probability \( n^{-1} \pi_k \) at each draw. In such case, the variance of the Hansen-Hurwitz estimator is

\[
V(\hat{t}_{yHH}) = \sum_{k \in U} \pi_k \left( \frac{y_k}{\pi_k} - \frac{t_y}{n} \right)^2.
\] (6)

With-replacement sampling designs are less common in surveys. We therefore confine our attention to without-replacement sampling designs and to the Narain-Horvitz-Thompson estimator \( \hat{t}_{y\pi} \). However, the variance obtained under multinomial sampling will be a useful benchmark to prove the mean square consistency.

### 3 Sufficient conditions for mean-square consistency

From (4), we obtain

\[
V(\hat{t}_{y\pi}) \leq N^2 \left( \frac{1}{N \min_{l \in U} \pi_l} + \frac{\max_{k \neq l \in U} |\pi_{kl} - \pi_l\pi_k|}{(\min_{l \in U} \pi_l)^2} \right) \times \frac{1}{N} \sum_{k \in U} y_k^2.
\] (7)

This directly leads to Proposition 3.1 below.

**Proposition 3.1** [Cardot et al., 2010](#). Assume that the following conditions hold:

- \( H1. \) We assume that \( \lim_{t \to \infty} \frac{n}{N} = f \in ]0, 1[. \)

- \( H2. \) We assume that \( \min_{k \in U} \pi_k \geq \lambda_1 > 0. \)
**H3.** The variable $y$ has a bounded second moment, i.e. there exists some constant $C_1$ such that $N^{-1} \sum_{k \in U} y_k^2 \leq C_1$.

**H4:** We have $\limsup_{t \to \infty} n \max_{k \neq l \in U} |\pi_{kl} - \pi_k \pi_l| < \infty$.

Then (1) holds and the Narain-Horvitz-Thompson estimator is consistent in mean square.

The assumptions in Proposition 3.1 are essentially the same as that in Cardot et al. (2010), except for the assumption (H3) which was replaced with

**H3b.** The variable $y$ is bounded, i.e. there exists some constant $C_1$ such that $|y_k| \leq C_1$.

Cardot et al. (2013) noticed however that (H3b) could be weakened to (H3). As noted by Breidt and Opsomer (2000), the assumption (H4) holds for stratified simple random sampling. This property also holds for rejective sampling (Hájek, 1964, Boistard et al., 2012) and its Sampford-Durbin modification (Hájek and Dupac, 1981). However, this property is rather difficult to prove for other sampling designs with unequal probabilities.

When the variable $y$ has non-negative values, a first proposal is to replace (H4) with

**H4b:** there exists some constant $a \geq 0$ such that for any vector $\pi$ of inclusion probabilities, we have for any $k \neq l \in U$:

$$\pi_{kl} \leq \left(1 + \frac{a}{n}\right) \times \pi_k \pi_l.$$  \hspace{1cm} (8)

From (4), this leads to

$$V(\hat{t}_{y\pi}) \leq N^2 \left(\frac{1}{N} \min_{l \in U} \pi_l + \frac{a}{n}\right) \times \frac{1}{N} \sum_{k \in U} y_k^2.$$  \hspace{1cm} (9)
Proposition 3.2 Assume that (H1)-(H3) and (H4b) hold, and that the variable $y$ has non-negative values. Then (7) holds and the Narain-Horvitz-Thompson estimator is consistent in mean square.

The assumption (H4) will hold in particular with $a = 0$ when the sampling design satisfies the so-called Sen (1953)-Yates-Grundy (1953) conditions, namely $\pi_{kl} \leq \pi_k \pi_l$ for any $k \neq l \in U$. This property holds for stratified simple random sampling, and for several sampling algorithms with unequal probability such as Poisson sampling, the Midzuno method, the elimination method, Chao’s method and the pivotal method (Deville and Tillé, 1998); the Sampford design (Gabler, 1981, 1984); the conditional Poisson sampling design (Chen et al., 1994).

In the case when the variable of interest may take both negative and non-negative values, we can consider the alternative condition that

\begin{equation}
\text{(H4c): for any vector } \pi \text{ of inclusion probabilities, the variance of the Narain-Horvitz-Thompson estimator under the sampling design } p(\cdot) \text{ with parameter } \pi \text{ is no greater than the variance of the Hansen-Hurwitz estimator under multinomial sampling with parameter } \pi. \end{equation}

Under (H4c), it follows from (6) that for any variable $y$

\begin{equation}
V(\hat{t}_y) \leq \sum_{k \in U} \pi_k \left( \frac{y_k}{\pi_k} \right)^2 \leq \frac{N}{\min_{i \in U} \pi_i} \times \frac{1}{N} \sum_{k \in U} y_k^2. \tag{10} \end{equation}

Proposition 3.3 Assume that (H1)-(H3) and (H4c) hold. Then (7) holds and the Narain-Horvitz-Thompson estimator is consistent in mean square.

The assumption (H4b) holds for simple random sampling, and for several sampling algorithms with unequal probability such as the Sampford design (Gabler, 1981, 1984), the conditional Poisson sampling design (Qualité, 2008), Chao’s method (Sengupta, 1989), the elimination method
Deville and Tillé (1998) and pivotal sampling (Chauvet and Ruiz-Gazen, 2014). Note that in case of pivotal sampling, numerous second-order inclusion probabilities are usually equal to zero (Deville and Tillé, 1998), so that assumption (H4) does not hold while (H4b) and (H4c) are respected.

4 Consistency for a martingale sampling algorithm

A martingale sampling algorithm proceeds in steps $i = 0, \ldots, T$ from $\pi(0) = \pi$ the vector of inclusion probabilities to $\pi(T) = I$ the final vector of sample membership indicators, such that the sequence $\{\pi(i)\}_{i=0}^{T}$ is a discrete-time martingale with $\pi(i) \in [0, 1]^N$ for any $i = 0, \ldots, T$; see Tillé (2011) and Breidt and Chauvet (2011).

Under a martingale sampling algorithm, we have $I - \pi = \sum_{i=0}^{T} \delta(i)$, where $\{\delta(i)\}_{i=0}^{T}$ are the innovations of the martingale. Since these innovations are not correlated, we have

$$V(I - \pi) = \sum_{i=0}^{T} V[\delta(i)] = E \left[ \sum_{i=0}^{T} \delta(i)\delta(i)^\top \right]. \quad (11)$$

We can write $\hat{t}_y - t_y = \tilde{y}^\top (I - \pi)$ where $\tilde{y} = (\pi_1^{-1} y_1, \ldots, \pi_N^{-1} y_N)^\top$. From (11), we obtain

$$V(\hat{t}_y - t_y) = E \left[ \sum_{i=0}^{T} \sum_{k,l \in U} \frac{y_k y_l}{\pi_k \pi_l} \delta_k(i)\delta_l(i) \right]. \quad (12)$$

We note

$$U_i = \{ k \in U; \delta_k(i) \neq 0 \} \quad (13)$$

the random subset of units in $U$ that are affected by step $i$. Also, we note $C = \max_{i=0,\ldots,T} \text{Card}(U_i)$. 7
From (12), we obtain

\[ V(\hat{t}_y - t_y) = E \left[ \sum_{i=0}^{T} \sum_{k,l \in U_i} \frac{y_k y_l}{\pi_k \pi_l} \delta_k(i) \delta_l(i) \right] \]

\[ \leq E \left[ \sum_{i=0}^{T} \sum_{k,l \in U_i} \frac{|y_k|}{\pi_k} \times \frac{|y_l|}{\pi_l} \right] \]

\[ \leq \left( \frac{\max_{k \in U} |y_k|}{\min_{k \in U} \pi_k} \right)^2 \times C^2 \times E(T). \]  \( \text{(14)} \)

**Proposition 4.1** Assume that assumptions (H1)-(H2) and (H3b) hold. Assume that \( C = O(1) \) and that \( E(T) = O(N) \). Then (1) holds and the Narain-Horvitz-Thompson estimator is consistent in mean square.

Note that in Proposition 4.1, the stronger condition (H3b) on the variable \( y \) is needed. Proposition 4.1 is in particular useful when the sample \( S \) is selected by means of the cube method ([Deville and Tillé, 2004]). Suppose that a \( q \)-vector \( x_k \) of auxiliary variables is known at the design stage for any unit \( k \in U \). The \( N \times q \) matrix \( A = (x_k/\pi_k)_{k \in U} \) is called the matrix of constraints.

The cube method enables to select samples such that the set of balancing equations

\[ \sum_{k \in S} \frac{x_k}{\pi_k} = t_x \]  \( \text{(15)} \)

is respected, at least approximately. A fast procedure for balanced sampling proposed by Chauvet and Tillé (2006, 2007) is described in Algorithm 1. At any step \( i \), \( U_i \subset \{1, \ldots, N\} \) denotes the set of the \( q + 1 \) first columns of \( A \) such that \( u_k(i) \) is not an integer. This is also the set of the \( q + 1 \) first units in the population \( U \) that are still neither selected nor rejected at step \( i \). Also, \( A_i \) denotes the sub-matrix of \( A \) containing the columns in \( U_i \). From the definition of \( u(i) \) and \( \delta(i) \) in Algorithm 1 we have \( C \leq q + 1 \). Also, it is easily shown that \([(q + 1)^{-1}N] \leq T \leq N\), with \([(q + 1)^{-1}N] \) the largest integer smaller than \((q + 1)^{-1}N\). Proposition 4.2 below is thus an immediate consequence of Proposition 4.1.
Algorithm 1 A fast procedure for the cube method
First initialize at \( \pi(0) = \pi \). Next, at time \( i = 0, \ldots, T \), repeat the following steps:

1. If there exists some vector \( v(i) \neq 0 \) such that \( v(i) \in Ker(A_i) \), then:

   (a) Take any such vector \( v(i) \) (random or not), and take \( u(i) \) such that

   \[
   u_k(i) = \begin{cases} 
   v_k(i) & \text{if } k \in U_i, \\
   0 & \text{otherwise}. 
   \end{cases}
   \]

   Compute \( \lambda_1^*(i) \) and \( \lambda_2^*(i) \), the largest values of \( \lambda_1(i) \) and \( \lambda_2(i) \) such that

   \[
   0 \leq \pi(i) + \lambda_1(i)u(i) \leq 1 \quad \text{and} \quad 0 \leq \pi(i) - \lambda_2(i)u(i) \leq 1.
   \]

   (b) Take \( \pi(i + 1) = \pi(i) + \delta(i) \), where

   \[
   \delta(i) = \begin{cases} 
   \lambda_1^*(i)u(i) & \text{with probability } \lambda_2^*(i)/\{\lambda_1^*(i) + \lambda_2^*(i)\}, \\
   -\lambda_2^*(i)u(i) & \text{with probability } \lambda_1^*(i)/\{\lambda_1^*(i) + \lambda_2^*(i)\}.
   \end{cases}
   \]

2. Otherwise, drop the last column from the matrix \( A_i \) and go back to Step 1.

Proposition 4.2 Assume that assumptions (H1)-(H2) and (H3b) hold. Assume that the sample \( S \) is selected by means of Algorithm [1] and that \( q = O(1) \). Then \( V \left\{ N^{-1}(\hat{t}_y - t_y) \right\} = O(n^{-1}) \) and the Narain-Horvitz-Thompson estimator is consistent in mean square.

Other implementations of the cube method are possible, for which Proposition [1] may not be suitable to obtain the mean-square consistency. For the general balanced procedure described in Algorithm 8.3 in Tillé (2011), we have \( U_i = \{ k \in U; \pi_k(i - 1) \notin \{0,1\} \} \) which means that all the units that are still neither selected nor definitely rejected at step \( i - 1 \) are possibly affected at step \( i \). This leads to \( \mathcal{C} = N \), so that the assumptions for Proposition [1] are not fulfilled.
References

Boistard, H., Lopuhaä, H., and Ruiz-Gazen, A. (2012). Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electronic Journal of Statistics*, 6:1967–1983.

Breidt, F. and Chauvet, G. (2011). Improved variance estimation for balanced samples drawn via the cube method. *Journal of Statistical Planning and Inference*, 141(1):479–487.

Breidt, F. and Opsomer, J. (2000). Local polynomial regression estimators in survey sampling. *Annals of Statistics*, pages 1026–1053.

Cardot, H., Chaouch, M., Goga, C., and Labruère, C. (2010). Properties of design-based functional principal components analysis. *Journal of Statistical Planning and Inference*, 140(1):75–91.

Cardot, H., Goga, C., and Lardin, P. (2013). Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electronic Journal of Statistics*, 7:562–596.

Chauvet, G. and Ruiz-Gazen, A. (2014). A comparison of pivotal sampling and unequal probability sampling with replacement. *Submitted*.

Chauvet, G. and Tillé, Y. (2006). A fast algorithm for balanced sampling. *Computational Statistics*, 21(1):53–62.

Chauvet, G. and Tillé, Y. (2007). Application of fast sas macros for balancing samples to the selection of addresses. *Case Studies in Business, Industry and Government Statistics*, 1:173–182.
Chen, X.-H., Dempster, A., and Liu, J. (1994). Weighted finite population sampling to maximize entropy. *Biometrika*, 81(3):457–469.

Deville, J.-C. and Tillé, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika*, 85(1):89–101.

Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: the cube method. *Biometrika*, 91(4):893–912.

Gabler, S. (1981). A comparison of sampford’s sampling procedure versus unequal probability sampling with replacement. *Biometrika*, pages 725–727.

Gabler, S. (1984). On unequal probability sampling: sufficient conditions for the superiority of sampling without replacement. *Biometrika*, 71(1):171–175.

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

Hájek, J. and Dupac, V. (1981). *Sampling from a finite population*. Marcel Dekker New York.

Hansen, M. H. and Hurwitz, W. N. (1953). Sample survey methods and theory. vol. i.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

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

Narain, R. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 3:169–175.
Qualité, L. (2008). A comparison of conditional poisson sampling versus unequal probability sampling with replacement. *Journal of Statistical Planning and Inference*, 138(5):1428–1432.

Robinson, P. and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 240–248.

Sen, A. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5(1194):127.

Sengupta, S. (1989). On chao’s unequal probability sampling plan. *Biometrika*, 76(1):192–196.

Tillé, Y. (2011). *Sampling algorithms*. Springer.

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