# A cautionary note on the Hanurav-Vijayan sampling algorithm

Guillaume Chauvet

# A cautionary note on the Hanurav-Vijayan sampling algorithm

Guillaume Chauvet*

March 9, 2021

**Abstract**

We consider the Hanurav-Vijayan sampling design, which is the default method programmed in the `SURVEYSELECT` procedure of the `SAS` software. We prove that it is equivalent to the Sunter procedure, but is capable of handling any set of inclusion probabilities. We prove that the Horvitz-Thompson estimator is not generally consistent under this sampling design. We propose a conditional Horvitz-Thompson estimator, and prove its consistency under a non-standard assumption on the first-order inclusion probabilities. Since this assumption seems difficult to control in practice, we recommend not to use the Hanurav-Vijayan sampling design.

## 1 Introduction

The Hanurav-Vijayan method (Vijayan, 1968) makes it possible to select a sample with probabilities proportional to size. This is the default method programmed in the `SURVEYSELECT` procedure of the `SAS` software for unequal probability sampling. It is therefore routinely used, see for example Langlet et al. (2003); Kulathinal et al. (2007); Myrskylä (2007); Jang et al. (2010); Zhao (2011);

*Univ Rennes, ENSAI, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France, email: chauvet@ensai.fr

Chauvet and Vallée (2020); Xiong and Higgins (2020).

This sampling algorithm has a number of interesting features. The procedure is of fixed-size, the required first-order inclusion probabilities are exactly respected, and the second-order inclusion probabilities are strictly positive and may be computed. However, the statistical properties of estimators arising from the Hanurav-Vijayan method remain poorly studied, which may be due to the fairly intricate description of the method. This is the purpose of this paper.

After describing the main notations and assumptions in section 2, we present the Hanurav-Vijayan method in section 3. We prove that it is equivalent to the so-called Sunter sequential procedure, but that it can handle any set of first-order inclusion probabilities. The consistency of the Horvitz-Thompson estimator is studied in Section 4. In particular, we prove that the Horvitz-Thompson is not generally consistent under the Hanurav-Vijayan method, unless a non-standard condition on the first-order inclusion probabilites is respected. This condition requires that the $n$ largest inclusion probabilities are very close to each other. A conditional Horvitz-Thompson estimator is suggested in Section 5, and its consistency is established under a weaker condition. The results of the simulation study in section 6 support our findings. We conclude in section 7.

## 2 Notation

We consider a finite population $U = \{1, \ldots, N\}$. Denote by $\pi = (\pi_1, \ldots, \pi_N)^\top$ a vector of probabilities, with $0 < \pi_k < 1$ for any unit $k \in U$, and with $n = \sum_{k \in U} \pi_k$ the expected sample size. We

suppose that the population $U$ is ordered with respect to the inclusion probabilities, i.e.

$$\pi_1 \leq \ldots \leq \pi_N. \tag{2.1}$$

We note $\pi_l^+ = \sum_{k=1}^{l} \pi_k$ for the cumulated inclusion probabilities up to unit $l$. A random sample is selected in $U$ by means of a without-replacement sampling design with parameter $\pi$, i.e. such that $E(I) = \pi$, where

$$I \quad = \quad (I_1, \ldots, I_N)^\top \tag{2.2}$$

is the vector of sample membership indicators. We are interested in the estimation of the total $t_y = \sum_{k \in U} y_k$ for a variable of interest $y_k$.

Throughout the paper, we will consider the following assumptions:

VA1: There exists some constant $C_1$ such that:

$$\frac{1}{N} \sum_{k \in U} y_k^2 \quad \leq \quad C_1.$$

SD1: We have $n \to \infty$ as $N \to \infty$, and there exists some constant $f \in ]0,1[$ such that $N^{-1}n \to f$.

There exists some constants $0 < \lambda_1 \leq \Lambda_1$ such that for any $k \in U$:

$$\lambda_1 \frac{n}{N} \leq \quad \pi_k \quad \leq \Lambda_1 \frac{n}{N}.$$

SD2: There exists some function $h(n, N) \to 0$ such that

$$\max_{i=1,\ldots,n-1} \{\pi_{N-n+i+1} - \pi_{N-n+i}\} \quad \leq \quad h(n, N). \tag{2.3}$$

The assumption (VA1) is related to the variable of interest, which is assumed to have a finite moment of order 2. The assumption (SD1) is related to the sampling design, and also defines the

asymptotic framework. It is assumed that all the first-order inclusion probabilities are of order $n/N$. The assumptions (VA1) and (SD1) are standard. The assumption (SD2) is more unusual, and states that the $n$ largest inclusion probabilities are sufficiently close to each other. Note that from the identity

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\{\pi_{N-n+i+1}-\pi_{N-n+i}\} \quad = \quad \frac{\pi_N-\pi_{N-n+1}}{n-1},$$

the mean value of these differences is of order $N^{-1}$ under assumption (SD1). It would therefore seem natural to use $h(n,N)=N^{-1}$ in assumption (SD2). In any case, we prove in Section 4 that under the Hanurav-Vijayan sampling process, $h(n,N)$ needs to be of smaller order for the Horvitz-Thompson estimator to be generally consistent.

## 3 Hanurav-Vijayan procedure

The sampling algorithm proposed by Vijayan (1968) is a generalization of a procedure by Hanurav (1967). Vijayan (1968) considered the specific case of unequal probability sampling with probabilities proportional to size. The description in Algorithm 1 is more general, since it can be applied for any set $\pi$ of inclusion probabilities. We have also simplified the presentation, to express the intermediary quantities needed in the sampling process in terms of the inclusion probabilities only.

The Hanurav-Vijayan procedure is split into two phases. During the first phase, an integer $n'$ is randomly selected in $\{1,\ldots,n\}$ and a new vector $\pi(0)$ of inclusion probabilities is obtained. The $n-n'$ units with the larger inclusion probabilities ($k > N-n+n'$) are selected, while the $n'$ remaining units with the larger inclusion probabilities ($N-n+1 < k \leq N-n+n'$) are given the same value $\frac{n'\pi_{N-n+1}}{\pi^+_{N-n}+n'\pi_{N-n+1}}$. During the second phase, a sample of size $n'$ is selected among the

**Algorithm 1** Hanurav-Vijayan procedure: draw by draw algorithm

Phase 1:

- Select an integer $n'$ with probabilities

$$\delta_i = (\pi_{N-n+i+1} - \pi_{N-n+i}) \frac{\pi^+_{N-n} + i\pi_{N-n+1}}{\pi^+_{N-n}} \text{ for } i \in \{1, \ldots, n\}, \tag{3.1}$$

  where $\pi_{N+1} = 1$. We note $N' = N - (n - n')$.

- Take $\pi(0) = \{\pi_1(0), \ldots, \pi_N(0)\}^\top$, such that

$$\pi_k(0) = \begin{cases} \frac{n'\pi_k}{\pi^+_{N-n} + n'\pi_{N-n+1}} & \text{if } k \leq N - (n-1), \\ \frac{n'\pi_{N-n+1}}{\pi^+_{N-n} + n'\pi_{N-n+1}} & \text{if } N - (n-1) < k \leq N', \\ 1 & \text{if } k > N'. \end{cases} \tag{3.2}$$

Phase 2: In the population $U' = \{1, \ldots, N'\}$, select a sample of size $n'$ as follows:

- Initialize with $i_0 = 0$.

- For $j = 1, \ldots, n'$, select one unit $i_j$ from $\{i_{j-1}+1, \ldots, N-n+j\}$ with probabilities proportional to

$$a^j_{i_{j-1}+1} = \frac{n' - j + 1}{n'} \pi_{i_{j-1}+1}(0) \tag{3.3}$$

  for unit $i_{j-1} + 1$ and

$$a^j_k = \prod_{l=i_{j-1}+1}^{k-1} \left\{ 1 - (n' - j) \frac{\pi_l(0)}{n' - \pi^+_l(0)} \right\} \times \frac{n' - j + 1}{n'} \pi_k(0) \tag{3.4}$$

  for $k = i_{j-1} + 2, \ldots, N - n + j$, where $\pi^+_l(0) = \sum_{k=1}^{l} \pi_k(0)$.

The final sample is: $S = \{i_1, \ldots, i_{n'}, N' + 1, \ldots, N\}$.

remaining units through a draw by draw procedure. The algorithm is of fixed size by construction. We have $E\{\pi(0)\} = \pi$, and conditionally on $\pi(0)$ the sampling in $U'$ is performed with inclusion probabilities $\pi(0)$ (see Vijayan, 1968, Theorem 1). Therefore, the original set of inclusion probabilities $\pi$ is exactly respected. We denote by $\pi_{kl}(0) = E\{I_k I_l | \pi(0)\}$ the second-order inclusion probability of units $k, l \in U'$ during Phase 2, conditionally on $\pi(0)$.

The random rounding in Phase 1 ensures that $\pi_{N-n+1}(0) = \pi_{N-n+2}(0) = \ldots = \pi_{N-n+n'}(0)$, which is necessary for the suitability of the draw by draw procedure in Phase 2. This is an early example of the splitting method later theorized by Deville and Tillé (1998). Note that if $\pi_{N-n+1} = \pi_N$, we obtain $n' = n$ with probability 1 and $\pi(0) = \pi$, which means that Phase 1 is not needed. For example, this occurs when sampling with equal probabilities, in which case the Hanurav-Vijayan procedure is equivalent to simple random sampling.

The second phase of the Hanurav-Vijayan algorithm may be more simply implemented in terms of a sequential procedure, presented in Algorithm 2. Proposition 1 states that both sampling algorithms are equivalent. The proof is given in Appendix A. The second phase of Algorithm 2 is a generalization of the selection-rejection method (Fan et al., 1962) for unequal probability sampling, known as the Sunter procedure (Sunter, 1977, 1986). The Sunter procedure is known to be non-exact, in the sense that it cannot be directly applied to any set of inclusion probabilities (e.g. Tillé, 2011, Section 6.2.8). The first phase of the Hanurav-Vijayan algorithm makes the Sunter algorithm applicable in full generality. It is remarkable that this solution was proposed ten years before the sequential procedure was introduced by Sunter (1977). Another possible generalization is proposed in Deville and Tillé (1998).

**Proposition 1.** *Algorithms 1 and 2 lead to the same sampling design.*

## 4 Horvitz-Thompson estimator

In this section, we are interested in the Horvitz-Thompson (HT) estimator

$$\hat{t}_{y\pi} = \sum_{k\in S} \frac{y_k}{\pi_k}. \tag{4.1}$$

We make use of the indicator

$$D_1(\pi) = \frac{1}{n}\sum_{i=1}^{n-1}(n-i)(\pi_{N-n+i+1} - \pi_{N-n+i}), \tag{4.2}$$

which can be seen as a measure of distance between the $n$ largest inclusion probabilities. We also use the notation

$$\xi(0) = \sum_{k\in U} \frac{y_k}{\pi_k}\left\{\pi_k(0) - \pi_k\right\}. \tag{4.3}$$

We have

$$\begin{aligned} V(\hat{t}_{y\pi}) &= VE\left\{\hat{t}_{y\pi}|\pi(0)\right\} + EV\left\{\hat{t}_{y\pi}|\pi(0)\right\} \qquad (4.4)\\ &\geq VE\left\{\hat{t}_{y\pi}|\pi(0)\right\} = V\left\{\xi(0)\right\} = \sum_{i=1}^{n}\delta_i E\left\{\xi(0)^2|n'=i\right\}\\ &\geq \delta_n\left[\left\{\frac{n}{\pi^+_{N-n}+n\pi_{N-n+1}} - 1\right\}\sum_{k=1}^{N-n} y_k + \sum_{k=N-n+1}^{N} y_k\left\{\frac{n\pi_{N-n+1}}{\pi_k(\pi^+_{N-n}+n\pi_{N-n+1})} - 1\right\}\right]^2, \end{aligned}$$

where the last line in (4.4) is obtained by keeping the case $i = n$ only. The inequality (4.4) gives the basic idea of why the HT-estimator may be inconsistent. The term $V\left\{\xi(0)\right\}$ is due to the randomization in Phase 1, which is needed for the suitability of the sampling in Phase 2. In some cases, this variability does not vanish as $n \to \infty$, as stated in Proposition 2.

**Algorithm 2** Hanurav-Vijayan-Sunter procedure: sequential algorithm

Phase 1:

- Select an integer $n'$ with probabilities

$$\delta_i \quad = \quad (\pi_{N-n+i+1} - \pi_{N-n+i}) \frac{\pi^+_{N-n} + i\pi_{N-n+1}}{\pi^+_{N-n}} \text{ for } i \in \{1, \ldots, n\},$$

  where $\pi_{N+1} = 1$. We note $N' = N - (n - n')$.

- Take $\pi(0) = \{\pi_1(0), \ldots, \pi_N(0)\}^\top$, such that

$$\pi_k(0) \quad = \quad \begin{cases} \frac{n'\pi_k}{\pi^+_{N-n} + n'\pi_{N-n+1}} & \text{if } k \leq N - (n-1), \\ \frac{n'\pi_{N-n+1}}{\pi^+_{N-n} + n'\pi_{N-n+1}} & \text{if } N - (n-1) < k \leq N', \\ 1 & \text{if } k > N'. \end{cases}$$

Phase 2: In the population $U' = \{1, \ldots, N'\}$, select a sample of size $n'$ as follows. Initialize with $n_0 = 0$. For $t = 1, \ldots, N' - 1$:

- take $I_t = 1$ with probability $\pi_t(t-1)$, and $n_t = n_{t-1} + I_t$,

- compute $\pi(t) = \{\pi_1(t), \ldots, \pi_{N'}(t)\}^\top$ such that

$$\pi_k(t) \quad = \quad \begin{cases} \pi_k(t-1) & \text{if } k \leq t-1, \\ I_t & \text{if } k = t, \\ (n' - n_t)\frac{\pi_k(0)}{n' - \pi^+_t(0)} & \text{if } k > t, \end{cases} \tag{3.5}$$

  where $\pi^+_t(0) = \sum_{k=1}^t \pi_k(0)$.

The vector of sample membership indicators is $I = \{\pi_1(N'-1), \ldots, \pi_{N'}(N'-1), 1, \ldots, 1\}^\top$.

**Proposition 2.** *Suppose that assumption (SD1) holds, and that there exists some constant* $0 < \lambda_2$ *such that*

$$\lambda_2 \frac{n}{N} \quad \leq \quad D_1(\pi). \tag{4.5}$$

*Suppose that there exists some constants* $c_2 > 0$ *and* $C_2$ *such that*

$$c_2 \leq \frac{1}{N-n} \left| \sum_{k=1}^{N-n} y_k \right| \quad \textit{and} \quad \frac{1}{n} \sum_{k=N-n+1}^{N} |y_k| \leq C_2. \tag{4.6}$$

*If*

$$\frac{c_2}{C_2} \quad > \quad \frac{1}{\lambda_2(1-f)} \frac{\Lambda_1}{\lambda_1} \left(1 + \frac{1}{\lambda_1}\right), \tag{4.7}$$

*where the constants* $f$, $\lambda_1$ *and* $\Lambda_1$ *are defined in assumption (SD1), then there exists some constant* $C > 0$ *such that:*

$$V(N^{-1}\hat{t}_{y\pi}) \quad \geq \quad (1 - \pi_N) C N^{-2} n^2. \tag{4.8}$$

Proposition 2 states that if the indicator $D_1(\pi)$ is too large, we can always find variables of interest satisfying assumption (VA1) and such that the HT-estimator is not consistent, since the second term in the right-hand side of (4.8) is bounded away from 0. The proof is given in Appendix B. The ratio $c_2/C_2$ may be thought of as a measure of balance of the total $t_y$ between the $N-n$ first units and the $n$ last units: the HT-estimator is not consistent if the total $t_y$ is too highly concentrated on the $N-n$ first units.

**Proposition 3.** *Suppose that assumptions (VA1) and (SD1) hold, and that assumption (SD2) holds with* $h(n, N) = o(N^{-1})$. *Then*

$$V(N^{-1}\hat{t}_{y\pi}) \quad = \quad o(1). \tag{4.9}$$

Proposition 3 states that the HT-estimator is consistent if the $n$ largest inclusion probabilities are sufficiently close, namely if assumption (SD2) holds with $h(n,N) = o(N^{-1})$. This assumption can not be dropped. For example, if there is a constant lag of order $N^{-1}$ between these probabilities, namely if there exists some constant $0 < \lambda$ such that

$$\pi_{N-n+i+1} - \pi_{N-n+i} \quad = \quad \frac{\lambda}{N},$$

then $D_1(\pi) = \frac{\lambda}{2}\frac{n-1}{N}$. Therefore, equation (4.5) in Proposition 2 holds, and there are some variables of interest such that (VA1) holds but the HT-estimator is not consistent.

From a look at the proof of Proposition 3, the assumption (SD2) is needed to control the term $VE(\hat{t}_{y\pi}|\pi(0))$, which is due to the first phase in Algorithm 1. To remove this variability, it is possible to work conditionally on $\pi(0)$. This is the purpose of the next section.

## 5 Conditional Horvitz-Thompson estimator

We are interested in the conditional Horvitz-Thompson (CHT) estimator, defined as

$$\hat{t}_{y\pi}(0) \quad = \quad \sum_{k \in S} \frac{y_k}{\pi_k(0)}. \tag{5.1}$$

This estimator makes use of the set of inclusion probabilities $\pi(0)$ obtained after Phase 1 of Algorithm 1. It may be rewritten as

$$\hat{t}_{y\pi}(0) \quad = \quad \sum_{k \in U'} \frac{y_k}{\pi_k(0)} I_k + \sum_{k > N'} y_k, \tag{5.2}$$

which leads to

$$E\left\{\hat{t}_{y\pi}(0)|\pi(0)\right\} \quad = \quad \sum_{k \in U'} y_k + \sum_{k > N'} y_k = t_y. \tag{5.3}$$

This is therefore an unbiased estimator for $t_y$, conditionally on $\pi(0)$.

**Proposition 4.** *Suppose that assumptions (VA1) and (SD1) hold, and that assumption (SD2) holds with* $h(n,N) = o\left(\frac{1}{\ln(n)}\right)$. *Then*

$$V\{N^{-1}\hat{t}_{y\pi}(0)\} \quad = \quad o(1). \tag{5.4}$$

*If the assumption (SD2) holds with* $h(n,N) = O\left(\frac{1}{n\ln(n)}\right)$, *then the CHT-estimator is* $\sqrt{n}$*-consistent.*

The proof of Proposition 4 is given in Appendix D. We clearly need a weaker assumption on the difference of the largest inclusion probabilities. Anyway, we need these differences to be no greater than $O\left(\frac{1}{n\ln(n)}\right)$ to ensure the usual $\sqrt{n}$-consistency, which is still demanding.

Another advantage of the CHT-estimator is that the variance may be easily estimated. From the corollary of Theorem 1 in Vijayan (1968), there is an explicit expression for the conditional second-order inclusion probabilities, which is restated in Proposition 5. Note that an incorrect factor of $\frac{1}{2}$ was indicated in equation (5.5) by Vijayan (1968), see Chaudhuri and Vos (1988).

**Proposition 5.** *(Vijayan, 1968) For* $k = 1, \ldots, N'$, *we note*

$$p_k(0) = \frac{\pi_k(0)}{n'} \quad \textit{and} \quad P_k(0) = \frac{\pi_k(0)}{n' - \pi_k^+(0)}.$$

*For* $k < l = 1, \ldots, N'$, *we have*

$$\pi_{kl}(0) \quad = \quad n'(n'-1)\{1 - P_1(0)\} \ldots \{1 - P_{k-1}(0)\}P_k(0)p_l(0). \tag{5.5}$$

*For* $k = 1, \ldots, N'$ *and* $l = N'+1, \ldots, N$, *we have*

$$\pi_{kl}(0) \quad = \quad \pi_k(0). \tag{5.6}$$

*For $k < l = N' + 1, \ldots, N$, we have*

$$\pi_{kl}(0) \quad = \quad 1. \tag{5.7}$$

The second-order inclusion probabilities $\pi_{kl}(0)$ are strictly positive, and satisfy the Sen-Yates-Grundy conditions (see Vijayan, 1968, Theorem 3). Therefore, the Sen-Yates-Grundy variance estimator is unbiased and takes positive values only. In Theorem 2 of Vijayan (1968), these probabilities are averaged to obtain the unconditional second-order inclusion probabilities for the HT estimator. However, this involves computing the $\pi_{kl}(0)$'s for each of the $n$ possible cases for the integer $n'$, which is cumbersome if $n$ is large.

## 6 Simulation study

We conduct a simulation study to illustrate the properties of the Horvitz-Thompson (HT) estimator and of the conditional Horvitz-Thompson (CHT) estimator. The set-up is inspired from Chauvet (2020). We generate 2 populations of size $N$, each consisting of an auxiliary variable $x$ and 4 variables of interest $y_1, \ldots, y_4$. The $x$-values are generated according to the model

$$x_k \quad = \quad \alpha + \eta_k. \tag{6.1}$$

In the first population, we use $\alpha = 8$ and $\eta_k$ is generated according to a Gamma distribution with shape and scale parameters 4 and 0.5. In the second population, we use $\alpha = 7$ and $\eta_k$ is generated according to a log-normal distribution with parameters 1.0 and 0.35. This leads to a mean of approximately 10 and a standard deviation of approximately 1 for the variable $x$ in both populations.

Given the $x$-values, the variables of interest are generated according to the following models:

$$\begin{aligned}
\texttt{linear} : y_{1k} &= \alpha_{10} + \alpha_{11}(x_k - \mu_x) + \sigma_1\ \epsilon_k, \\
\texttt{quadratic} : y_{2k} &= \alpha_{20} + \alpha_{21}(x_k - \mu_x)^2 + \sigma_2\ \epsilon_k, \\
\texttt{exponential} : y_{3k} &= \exp\{\alpha_{30} + \alpha_{31}(x_k - \mu_x)\} + \sigma_3\ \epsilon_k, \\
\texttt{bump} : y_{4k} &= \alpha_{40} + \alpha_{41}(x_k - \mu_x)^2 - \alpha_{42}\exp\left\{-\alpha_{43}(x_k - \mu_x)^2\right\} + \sigma_4\ \epsilon_k,
\end{aligned} \tag{6.2}$$

where $\mu_x$ is the population mean of $x$, and $\epsilon_k$ follows a standard normal distribution. The parameters are chosen in order to obtain a mean of approximately 20 and a standard deviation of approximately 3 for each variable of interest.

In each population, we compute inclusion probabilities proportional to $x$, according to the formula

$$\pi_k = n\frac{x_k}{\sum_{l\in U} x_l}. \tag{6.3}$$

We use ten different population sizes, ranging from $N = 2,000$ to $N = 20,000$, and a sampling fraction of 20% for each population. This leads to sample sizes ranging from $n = 400$ to $n = 4,000$. For example, when $N = 20,000$, the inclusion probabilities range between 0.16 and 0.32 when $x$ is generated by means of the Gamma distribution, and between 0.15 and 0.38 when $x$ is generated by means of the log-normal distribution.

We consider the indicator $D_1(\pi)$ defined in equation (4.2), and the additional indicators

$$\begin{aligned}
D_2(\pi) &= N \times \max_{i=1,\ldots,n-1}\{\pi_{N-n+i+1} - \pi_{N-n+i}\}, \\
D_3(\pi) &= \ln(n) \times \max_{i=1,\ldots,n-1}\{\pi_{N-n+i+1} - \pi_{N-n+i}\}.
\end{aligned}$$

If the assumption (SD2) is respected with $h(n, N) = o(N^{-1})$ (see Proposition 2), then $D_1(\pi) = o(1)$

and $D_2(\pi) = o(1)$, and they should therefore tend to 0 as $n$ increases. If the assumption (SD2) is respected with $h(n, N) = o(1/\ln(n))$ (see Proposition 3), then $D_3(\pi) = o(1)$ and $D_3(\pi)$ should therefore tend to 0 as $n$ increases. We have plotted these indicators in terms of the sample size $n$ in Figure 1. Neither of them decreases as $n$ increases. The indicator $D_1(\pi)$ is approximately constant, and so is the indicator $D_3(\pi)$ for large sample sizes ($n \geq 2,000$). The indicator $D_3(\pi)$ is clearly increasing with $n$. This supports the apparent difficulties for controlling the closeness of the largest inclusion probabilities via the assumption (SD2).

We consider the estimation of the population mean $\mu_y = N^{-1} \sum_{k \in U} y_k$. We select $B = 10,000$ samples by means of the HVS sampling algorithm. For each sample and each variable of interest, we consider the population mean $\mu_y = N^{-1} \sum_{k \in U} y_k$. We compute the Horvitz-Thompson estimator of the mean $\hat{\mu}_{y\pi} = N^{-1} \hat{t}_{y\pi}$, and the conditional estimator of the mean $\hat{\mu}_{y\pi}(0) = N^{-1} \hat{t}_{y\pi}(0)$. For each estimator $\hat{\mu}_y$ and for a given sample size $n$, we compute the Monte-Carlo variance

$$V_{MC,n}(\hat{\mu}_y) = \frac{1}{B} \sum_{b=1}^{B} \left\{ \hat{\mu}_y(s_b) - \frac{1}{B} \sum_{c=1}^{B} \hat{\mu}_y(s_c) \right\}^2, \tag{6.4}$$

with $\hat{\mu}_y(s_b)$ the estimator of the mean computed on the $b$-th sample. We also compute the Monte-Carlo variance ratio

$$RV_{MC,n}(\hat{t}_y) = \frac{V_{MC,n}(\hat{t}_y)}{V_{MC,n-400}(\hat{t}_y)}. \tag{6.5}$$

If the estimator $\hat{t}_y$ is consistent, the Monte-Carlo variance is expected to decrease as the sample size increases, and the Monte-Carlo variance ratios should be lower than 1.

The simulation results for the HT-estimator are presented in Table 1. In 17 out of 72 cases the variance ratio $RV_{MC,n}$ is greater than 1, indicating that the variance increases as $n$ increases. In

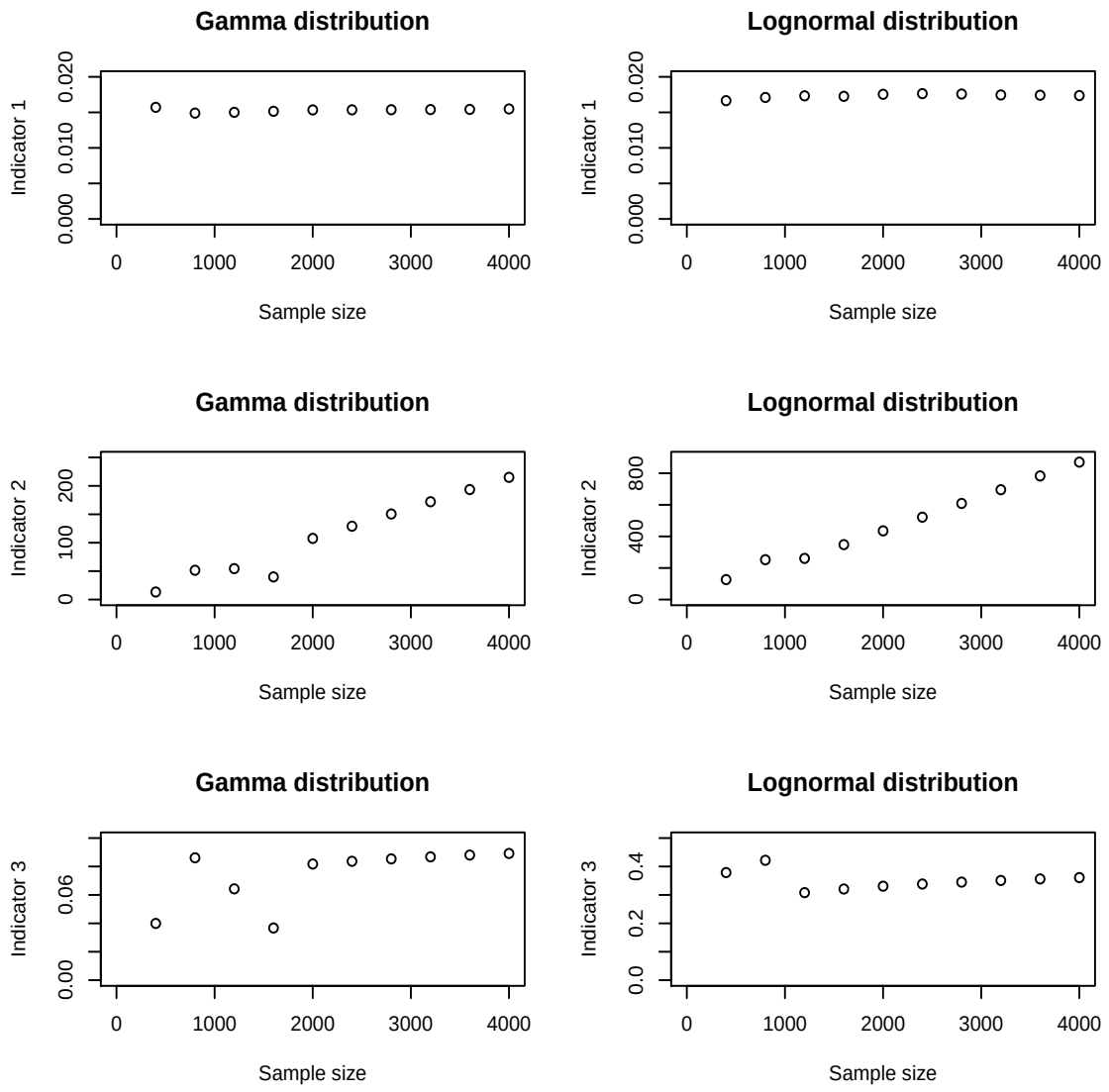


Figure 1: Indicators $D_1(\pi)$, $D_2(\pi)$ and $D_3(\pi)$ in function of the sample size $n$ with an auxiliary variable generated according to a gamma distribution (lhs) and by a lognormal distribution (rhs)

Population 1, the behavior of $\hat{\mu}_{y\pi}$ is particularly poor for `quadratic`, since the variance is of the same order with $n = 400$ ($27.84 \times 10^{-3}$) and $n = 4,000$ ($19.41 \times 10^{-3}$). In Population 2, the behavior of $\hat{\mu}_{y\pi}$ is particularly poor for `exponential`, since the variance is of the same order with $n = 400$ ($26.77 \times 10^{-3}$) and $n = 4,000$ ($21.75 \times 10^{-3}$). This supports the results in Section 4.

The simulation results for the CHT-estimator are presented in Table 2. The variance ratio $RV_{MC,n}$ is lower than 1 in 69 out of 72 cases, $RV_{MC,n}$ being lower than 1.03 in the three remaining cases. In almost all cases, the variance obtained with $n = 4,000$ is roughly one tenth of the variance obtained with $n = 400$, as could be expected. This supports the consistency result obtained in Proposition 4, even if the assumption (SD2) in Proposition 3 is not exactly respected (see the indicator $D_3(\pi)$ plotted in Figure 1). We note that the CHT-estimator is not necessarily more efficient than the HT-estimator. For `linear`, the variance of the HT-estimator is systematically lower.

# 7 Conclusion

In this paper, we have studied the Hanurav-Vijayan sampling algorithm. We have proposed a sequential characterization of the method, making the link with Sunter's procedure. We have also shown that to ensure the consistency of the Horvitz-Thompson estimator, or of an alternative conditional Horvitz-Thompson estimator, we need to control the closeness between the largest inclusion probabilities. This seems rather difficult to achieve in practice. On the other hand, alternative unequal probability sampling methods programmed in the `SURVEYSELECT` procedure lead to a consistent Horvitz-Thompson under the sole assumptions (VA1) and (SD1). This is the case for the Sampford method (Sampford, 1967) or Chromy's method (Chromy, 1979; Chauvet, 2020), for example. Therefore, we recommend that SAS users consider one of these two methods

instead.

## Acknowledgments

I would like to thank Todd Donahue for his careful reading of the manuscript.

## References

Chaudhuri, A. and Vos, J. (1988). Unified theory and strategies of survey sampling. Technical report.

Chauvet, G. (2020). A note on chromy's sampling procedure. *Journal of Survey Statistics and Methodology*.

Chauvet, G. and Vallée, A.-A. (2020). Inference for two-stage sampling designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):797–815.

Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pages 401–406.

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

Fan, C., Muller, M. E., and Rezucha, I. (1962). Development of sampling plans by using sequential (item by item) selection techniques and digital computers. *Journal of the American Statistical Association*, 57(298):387–402.

Hanurav, T. (1967). Optimum utilization of auxiliary information: $\pi$ps sampling of two units from a stratum. *Journal of the Royal Statistical Society: Series B (Methodological)*, 29(2):374–391.

Jang, L., Provost, M., and Sherk, A. (2010). Challenges in the design of the canadian community health survey on healthy aging. In *Proceedings of the Joint Statistical Meetings, American Statistical Association*, pages 2452–2466.

Kulathinal, S., Karvanen, J., Saarela, O., and Kuulasmaa, K. (2007). Case-cohort design in practice–experiences from the morgam project. *Epidemiologic Perspectives & Innovations*, 4(1):1–17.

Langlet, É. R., Faucher, D., and Lesage, É. (2003). An application of the bootstrap variance estimation method to the canadian participation and activity limitation survey. In *Proceedings of the Joint Statistical Meetings, American Statistical Association*, pages 2299–2306.

Myrskylä, M. (2007). *Generalised regression estimation for domain class frequencies*. PhD thesis, University of Helsinki.

Sampford, M. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54(3-4):499–513.

Sunter, A. (1977). List sequential sampling with equal or unequal probabilities without replacement. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 26(3):261–268.

Sunter, A. (1986). Solutions to the problem of unequal probability sampling without replacement. *International Statistical Review/Revue Internationale de Statistique*, pages 33–50.

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

Vijayan, K. (1968). An exact $\pi$-ps sampling scheme—generalization of a method of hanurav. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(3):556–566.

Xiong, Y. and Higgins, M. J. (2020). The benefits of probability-proportional-to-size sampling in cluster-randomized experiments. *arXiv preprint arXiv:2002.08009*.

Zhao, Y. (2011). Estimating the size of an injecting drug user population. *World Journal of AIDS*, 1(03):88.

# A Proof of Proposition 1

Let $k_1, \ldots, k_{n'}$ denote the $n'$ units successively selected during Phase 2 of Algorithm 2. It is sufficient to prove that their probability distribution is the same as that of the units $i_1, \ldots, i_{n'}$ successively selected during Phase 2 of Algorithm 1. The proof is by induction.

We begin with the probability distribution of $k_1$. We have

$$Pr(k_1 = 1) \quad = \quad Pr(I_1 = 1) = \pi_1(0) = a_1^1.$$

Also, for $k \in \{1, \ldots, N - n + 1\}$, we have

$$\begin{aligned}
Pr(k_1 = k) &= Pr(I_1 = \ldots = I_{k-1} = 0, I_k = 1) \\
&= \left\{ \prod_{l=1}^{k-1} \left( 1 - n' \frac{\pi_l(0)}{n' - \pi_{l-1}^+(0)} \right) \right\} \left\{ n' \frac{\pi_k(0)}{n' - \pi_{k-1}^+(0)} \right\} \\
&= \frac{\prod_{l=1}^{k-1} (n' - \pi_{l-1}^+(0) - n'\pi_l(0))}{n' \left\{ \prod_{l=2}^{k-1} (n' - \pi_{l-1}^+(0)) \right\} \{n' - \pi_{k-1}^+(0)\}} \{n'\pi_k(0)\} \\
&= \frac{\prod_{l=1}^{k-1} (n' - \pi_{l-1}^+(0) - n'\pi_l(0))}{\prod_{l=1}^{k-1} (n' - \pi_l^+(0))} \{\pi_k(0)\} \\
&= \left[ \prod_{l=1}^{k-1} \left\{ 1 - (n' - 1) \frac{\pi_l(0)}{n' - \pi_l^+(0)} \right\} \right] \times \pi_k(0) = a_k^1.
\end{aligned}$$

From equations (3.3) and (3.4), $i_1$ and $k_1$ have the same distribution.

Now, suppose that units $k_1, \ldots, k_{j-1}$ have been selected. We have

$$\begin{aligned} Pr(k_j = k_{j-1}+1 | k_1, \ldots, k_{j-1}) &= Pr(I_{k_{j-1}+1} = 1 | k_1, \ldots, k_{j-1}) \\ &= Pr(I_{k_{j-1}+1} = 1 | n'_{k_{j-1}} = j-1) \\ &= (n'-j+1)\frac{\pi_{k_{j-1}+1}(0)}{n' - \pi^+_{k_{j-1}}(0)} \\ &= \left\{ \frac{n'}{n' - \pi^+_{k_{j-1}}(0)} \right\} a^j_{k_{j-1}+1}. \end{aligned}$$

Also, for $k \in \{k_{j-1}+2, \ldots, N-n+j\}$, we have

$$\begin{aligned} Pr(k_j = k | k_1, \ldots, k_{j-1}) &= Pr(I_{k_{j-1}+1} = \ldots = I_{k-1} = 0, I_k = 1 | n'_{k_{j-1}} = j-1) \\ &= \left\{ \prod_{l=k_{j-1}+1}^{k-1} \left( 1 - (n'-j+1)\frac{\pi_l(0)}{n' - \pi^+_{l-1}(0)} \right) \right\} \left\{ (n'-j+1)\frac{\pi_k(0)}{n' - \pi^+_{k-1}(0)} \right\} \\ &= \frac{\prod_{l=k_{j-1}+1}^{k-1} \{n' - \pi^+_{l-1}(0) - (n'-j+1)\pi_l(0)\}}{\{n' - \pi^+_{k_{j-1}}(0)\} \left\{ \prod_{l=k_{j-1}+2}^{k-1} (n' - \pi^+_{l-1}(0)) \right\} \{n' - \pi^+_{k-1}(0)\}} \{n'-j+1\}\pi_k(0) \\ &= \frac{\prod_{l=k_{j-1}+1}^{k-1} \{n' - \pi^+_{l-1}(0) - (n'-j+1)\pi_l(0)\}}{\prod_{l=k_{j-1}+1}^{k-1} (n' - \pi^+_l(0))} \times \frac{n'-j+1}{n' - \pi^+_{k_{j-1}}(0)}\pi_k(0) \\ &= \left[ \prod_{l=k_{j-1}+1}^{k-1} \left\{ 1 - (n'-j)\frac{\pi_l(0)}{n' - \pi^+_l(0)} \right\} \right] \times \frac{n'-j+1}{n' - \pi^+_{k_{j-1}}(0)}\pi_k(0) \\ &= \left\{ \frac{n'}{n' - \pi^+_{k_{j-1}}(0)} \right\} a^j_k. \end{aligned}$$

From equations (3.3) and (3.4), $i_j$ and $k_j$ have the same conditional distribution. This completes the proof.

## B Proof of proposition 2

We note $\Delta = \Delta_1 + \Delta_2$, where

$$\Delta_1 = \left\{ \frac{n}{\pi^+_{N-n} + n\pi_{N-n+1}} - 1 \right\} \sum_{k=1}^{N-n} y_k, \tag{B.1}$$

$$\Delta_2 = \sum_{k=N-n+1}^{N} \frac{y_k}{\pi_k} \left\{ \frac{n\pi_{N-n+1}}{\pi^+_{N-n} + n\pi_{N-n+1}} - \pi_k \right\}. \tag{B.2}$$

By using equation (4.5) and the inequality

$$\begin{aligned} \frac{n}{\pi^+_{N-n} + n\pi_{N-n+1}} - 1 &= \frac{\sum_{k=N-n+1}^{N} \pi_k - n\pi_{N-n+1}}{\pi^+_{N-n} + n\pi_{N-n+1}} \\ &\geq \frac{\sum_{k=N-n+2}^{N} (\pi_k - \pi_{N-n+1})}{n} \\ &= \frac{1}{n} \sum_{i=1}^{n-1} (n-i)(\pi_{N-n+i+1} - \pi_{N-n+i}), \end{aligned}$$

we first obtain

$$|\Delta_1| \geq \lambda_2 \frac{n}{N} \left| \sum_{k=1}^{N-n} y_k \right|. \tag{B.3}$$

Also, from the assumption (SD1) and from the inequality

$$\left| \frac{n\pi_{N-n+1}}{\pi^+_{N-n} + n\pi_{N-n+1}} \right| \leq \frac{n}{N} \frac{\pi_N}{\pi_1},$$

we obtain

$$|\Delta_2| \leq \frac{\Lambda_1}{\lambda_1} \left( 1 + \frac{1}{\lambda_1} \right) \sum_{k=N-n+1}^{N} |y_k| \tag{B.4}$$

This leads to

$$|\Delta| \geq |\Delta_1| - |\Delta_2| \geq n \left\{ \lambda_2 (1-f) c_2 - \frac{\Lambda_1}{\lambda_1} \left( 1 + \frac{1}{\lambda_1} \right) C_2 \right\}, \tag{B.5}$$

and the result follows from equation (4.7) and from the inequality

$$\delta_n = (1 - \pi_N) \frac{\pi^+_{N-n} + n\pi_{N-n+1}}{\pi^+_{N-n}} \geq (1 - \pi_N).$$

## C Proof of proposition 3

Making use of Theorem 3 in Vijayan (1968), we have

$$\begin{aligned} V\{\hat{t}_{y\pi}|\pi(0)\} &\leq \sum_{k\in U'}\pi_k(0)\left\{\frac{y_k}{\pi_k}-\frac{1}{n'}\sum_{l\in U'}\frac{y_k}{\pi_k}\right\}^2 \leq \sum_{k\in U}\pi_k(0)\left\{\frac{y_k}{\pi_k}\right\}^2 \\ \Rightarrow EV\{\hat{t}_{y\pi}|\pi(0)\} &\leq \sum_{k\in U}\frac{y_k^2}{\pi_k}, \end{aligned} \tag{C.1}$$

and from assumptions (VA1) and (SD1), $EV\{\hat{t}_{y\pi}|\pi(0)\} = O(N^2n^{-1})$.

We also have

$$VE\{\hat{t}_{y\pi}|\pi(0)\} = V\{\xi(0)\} = \sum_{i=1}^{n-1}\delta_i E\{\xi(0)^2|n'=i\} + \delta_n E\{\xi(0)^2|n'=n\}. \tag{C.2}$$

For $i<n$, we obtain from the assumptions that $\delta_i = o(N^{-1})$ and $E\{\chi(0)^2|n'=i\} = O(N^4n^{-2})$, so that the first term in the rhs of (C.2) is $o(N^3n^{-1}) = o(N^2)$. We can also write

$$\begin{aligned} E\{\chi(0)^2|n'=n\} &= \left[\left\{\frac{n}{\pi^+_{N-n}+n\pi_{N-n+1}}-1\right\}\left\{\sum_{k=1}^{N-n}y_k+\pi_{N-n+1}\sum_{k=N-n+1}\frac{y_k}{\pi_k}\right\}\right. \\ &\quad \left. - \sum_{k=N-n+1}^{N}\frac{y_k}{\pi_k}(\pi_k-\pi_{N-n+1})\right]^2 \\ &\leq 2\left\{\frac{n}{\pi^+_{N-n}+n\pi_{N-n+1}}-1\right\}^2\left\{\sum_{k=1}^{N-n}y_k+\pi_{N-n+1}\sum_{k=N-n+1}\frac{y_k}{\pi_k}\right\}^2 \\ &\quad + 2\left\{\sum_{k=N-n+1}^{N}\frac{y_k}{\pi_k}(\pi_k-\pi_{N-n+1})\right\}^2. \end{aligned} \tag{C.3}$$

From the identity

$$\frac{n}{\pi^+_{N-n}+n\pi_{N-n+1}}-1 = \frac{\sum_{k=N-n+2}^{N}(\pi_k-\pi_{N-n+1})}{\pi^+_{N-n}+n\pi_{N-n+1}} = \frac{\sum_{i=1}^{n-1}(n-i)(\pi_{N-n+i+1}-\pi_{N-n+i})}{\pi^+_{N-n}+n\pi_{N-n+1}}$$

and from the assumptions, the first term in the rhs of (C.3) is $o(n^2)$, while the second term in the rhs of (C.3) is $o(N^2)$. From (C.2), we obtain that $VE\{\hat{t}_{y\pi}|\pi(0)\} = o(N^2)$. This completes the proof.

# D Proof of Proposition 4

## Preliminary result

**Lemma 1.** *Suppose that Assumptions (SD1) and (SD2) hold. Then some constants $C_3$ and $C_4$ exist such that*

$$E\left(\frac{1}{n'}\right) \leq C_3 h(n,N)\ln(n) + \frac{C_4}{N}.$$

## Proof

We have

$$\begin{aligned} E\left(\frac{1}{n'}\right) &= \sum_{i=1}^{n} \frac{\delta_i}{i} \\ &= \sum_{i=1}^{n} \left(\pi_{N-n+i+1} - \pi_{N-n+i}\right) \frac{\pi^+_{N-n} + i\pi_{N-n+1}}{\pi^+_{N-n}} \frac{1}{i} \\ &= \sum_{i=1}^{n} \frac{\pi_{N-n+i+1} - \pi_{N-n+i}}{i} + \frac{\pi_{N-n+1}(1-\pi_{N-n+1})}{\pi^+_{N-n}}, \end{aligned}$$

which gives the result.

## Proof of Proposition 4

First note that from equation (5.3), we have

$$V\left\{\hat{t}_{y\pi}(0)\right\} = EV\left\{\hat{t}_{y\pi}(0)|\pi(0)\right\}.$$

By using Theorem 3 in Vijayan (1968), we have

$$\begin{aligned} V\left\{\hat{t}_{y\pi}(0)|\pi(0)\right\} &= -\frac{1}{2}\sum_{k\neq l\in U'}\{\pi_{kl}(0)-\pi_k(0)\pi_l(0)\}\left\{\frac{y_k}{\pi_k(0)}-\frac{y_l}{\pi_l(0)}\right\}^2 \\ &\leq \frac{1}{2n'}\sum_{k\neq l\in U'}\pi_k(0)\pi_l(0)\left\{\frac{y_k}{\pi_k(0)}-\frac{y_l}{\pi_l(0)}\right\}^2 \\ &= \sum_{k\in U'}\pi_k(0)\left\{\frac{y_k}{\pi_k(0)}-\frac{1}{n'}\sum_{l\in U'}y_k\right\}^2 \\ &\leq \sum_{k\in U'}\frac{y_k^2}{\pi_k(0)}. \end{aligned}$$

By using the inequality

$$\pi_{N-n}^{+}+n'\pi_{N-n+1} \leq \pi_{N-n}^{+}+n\pi_{N-n+1}\leq \pi_N^{+}=n, \tag{D.1}$$

we obtain under assumption (H1) that for any $k\in U'$:

$$\pi_k(0) \geq f_0\frac{n'}{N}, \tag{D.2}$$

and the result follows from Assumption (H2) and Lemma 1.

Table 1: Monte-Carlo variance ($V_{MC,n}$) and Monte-Carlo variance ratio ($RV_{MC,n}$) for the Horvitz-Thompson (HT) estimator, for two populations and four variables of interest

| | Population 1 (Gamma distribution) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample size | 400 | 800 | 1,200 | 1,600 | 2,000 | 2,400 | 2,800 | 3,200 | 3,600 | 4,000 |
| | | | | | `linear` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 8.10 | 4.37 | 2.65 | 1.99 | 1.64 | 1.30 | 1.18 | 0.92 | 0.83 | 0.80 |
| ($RV_{MC,n}$) | | (0.54) | (0.61) | (0.75) | (0.83) | (0.79) | (0.91) | (0.77) | (0.91) | (0.96) |
| | | | | | `quadratic` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 27.84 | 28.14 | 18.42 | 20.22 | 15.90 | 16.84 | 24.68 | 14.43 | 12.31 | 19.41 |
| ($RV_{MC,n}$) | | (1.01) | (0.65) | (1.10) | (0.79) | (1.06) | (1.47) | (0.58) | (0.85) | (1.58) |
| | | | | | `exponential` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 9.42 | 5.55 | 4.77 | 3.58 | 3.68 | 3.33 | 3.84 | 3.12 | 2.72 | 3.38 |
| ($RV_{MC,n}$) | | (0.59) | (0.86) | (0.75) | (1.03) | (0.90) | (1.15) | (0.81) | (0.87) | (1.24) |
| | | | | | `bump` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 37.23 | 16.40 | 12.31 | 9.45 | 7.11 | 6.47 | 5.51 | 4.33 | 4.23 | 3.92 |
| ($RV_{MC,n}$) | | (0.44) | (0.75) | (0.77) | (0.75) | (0.91) | (0.85) | (0.79) | (0.98) | (0.93) |
| | Population 2 (Log-normal distribution) | | | | | | | | | |
| Sample size | 400 | 800 | 1,200 | 1,600 | 2,000 | 2,400 | 2,800 | 3,200 | 3,600 | 4,000 |
| | | | | | `linear` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 8.42 | 4.04 | 2.94 | 2.16 | 1.72 | 1.32 | 1.24 | 1.05 | 0.95 | 0.77 |
| ($RV_{MC,n}$) | | (0.48) | (0.73) | (0.74) | (0.80) | (0.77) | (0.94) | (0.85) | (0.90) | (0.81) |
| | | | | | `quadratic` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 37.61 | 36.57 | 24.70 | 37.25 | 27.30 | 17.26 | 19.51 | 29.01 | 24.15 | 27.77 |
| ($RV_{MC,n}$) | | (0.97) | (0.68) | (1.51) | (0.73) | (0.63) | (1.13) | (1.49) | (0.83) | (1.15) |
| | | | | | `exponential` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 26.77 | 27.11 | 20.49 | 28.29 | 22.09 | 15.00 | 16.08 | 23.31 | 19.20 | 21.75 |
| ($RV_{MC,n}$) | | (1.01) | (0.76) | (1.38) | (0.78) | (0.68) | (1.07) | (1.45) | (0.82) | (1.13) |
| | | | | | `bump` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 37.53 | 19.38 | 13.05 | 9.90 | 8.25 | 6.48 | 6.19 | 5.56 | 4.98 | 4.28 |
| ($RV_{MC,n}$) | | (0.52) | (0.67) | (0.76) | (0.83) | (0.78) | (0.95) | (0.90) | (0.90) | (0.86) |

Table 2: Monte-Carlo variance ($V_{MC,n}$) and Monte-Carlo variance ratio ($RV_{MC,n}$) for the conditional Horvitz-Thompson (CHT) estimator, for two populations and four variables of interest

| | Population 1 (Gamma distribution) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample size | 400 | 800 | 1,200 | 1,600 | 2,000 | 2,400 | 2,800 | 3,200 | 3,600 | 4,000 |
| | | | | | `linear` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 9.11 | 5.16 | 3.22 | 2.40 | 1.92 | 1.46 | 1.43 | 1.06 | 1.03 | 0.99 |
| ($RV_{MC,n}$) | | (0.57) | (0.62) | (0.75) | (0.80) | (0.76) | (0.98) | (0.74) | (0.96) | (0.96) |
| | | | | | `quadratic` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 12.46 | 8.31 | 4.16 | 3.34 | 2.50 | 2.06 | 1.91 | 1.55 | 1.34 | 1.29 |
| ($RV_{MC,n}$) | | (0.67) | (0.50) | (0.80) | (0.75) | (0.82) | (0.93) | (0.81) | (0.86) | (0.96) |
| | | | | | `exponential` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 10.21 | 5.80 | 3.68 | 2.73 | 2.22 | 1.67 | 1.65 | 1.23 | 1.19 | 1.13 |
| ($RV_{MC,n}$) | | (0.57) | (0.63) | (0.74) | (0.81) | (0.75) | (0.99) | (0.75) | (0.97) | (0.95) |
| | | | | | `bump` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 37.07 | 30.31 | 12.62 | 10.69 | 7.22 | 6.53 | 5.40 | 4.32 | 4.34 | 3.96 |
| ($RV_{MC,n}$) | | (0.82) | (0.42) | (0.85) | (0.67) | (0.91) | (0.83) | (0.80) | (1.01) | (0.91) |
| | Population 2 (Log-normal distribution) | | | | | | | | | |
| Sample size | 400 | 800 | 1,200 | 1,600 | 2,000 | 2,400 | 2,800 | 3,200 | 3,600 | 4,000 |
| | | | | | `linear` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 10.80 | 5.11 | 3.50 | 2.50 | 2.01 | 1.69 | 1.47 | 1.29 | 1.29 | 0.91 |
| ($RV_{MC,n}$) | | (0.47) | (0.68) | (0.71) | (0.81) | (0.84) | (0.87) | (0.88) | (1.00) | (0.70) |
| | | | | | `quadratic` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 16.10 | 7.50 | 4.95 | 3.57 | 2.98 | 2.32 | 2.07 | 2.08 | 1.83 | 1.27 |
| ($RV_{MC,n}$) | | (0.47) | (0.66) | (0.72) | (0.83) | (0.78) | (0.89) | (1.01) | (0.88) | (0.69) |
| | | | | | `exponential` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 10.61 | 4.90 | 3.37 | 2.43 | 1.99 | 1.55 | 1.38 | 1.31 | 1.24 | 0.86 |
| ($RV_{MC,n}$) | | (0.46) | (0.69) | (0.72) | (0.82) | (0.78) | (0.89) | (0.95) | (0.94) | (0.69) |
| | | | | | `bump` | | | | | |
| $V_{MC,n}$ ($\times 10^{-3}$) | 40.49 | 21.34 | 13.10 | 8.85 | 7.84 | 7.13 | 5.93 | 4.62 | 4.75 | 3.60 |
| ($RV_{MC,n}$) | | (0.53) | (0.61) | (0.68) | (0.89) | (0.91) | (0.83) | (0.78) | (1.03) | (0.76) |