Journal of Official Statistics, Vol. 21, No. 4, 2005, pp. 543–570

Evaluation of Variance Approximations and Estimators in Maximum Entropy Sampling with Unequal Probability and Fixed Sample Size

Alina Matei and Yves Tillé[1]

Recent developments in survey sampling allow one to quickly draw samples with unequal probability, maximum entropy and fixed sample size. The joint inclusion probabilities can be computed exactly. For this sampling design, 7 approximations and 20 estimators of variance have been computed. A large set of simulations shows that knowledge of the joint inclusion probabilities is not necessary in order to obtain an accurate variance estimator.

Key words: Variance; unequal probabilities; maximum entropy; fixed sample size; simulations.

1. Introduction

Two of the most commonly used variance estimators in unequal probability sampling design are the Horvitz-Thompson estimator (Horvitz and Thompson 1952) and the Sen-Yates-Grundy estimator (Yates and Grundy 1953; Sen 1953). Both estimators use joint inclusion probabilities. It is often a hard task to evaluate the joint inclusion probabilities. Maximum entropy sampling design with fixed sample size allows the fast and exact computation of them.

The maximum entropy sampling design with fixed sample size is one of the principal topics of the post-mortem book of Hájek (1981). The principal problem of the implementation of this design was the combinatory explosion of the set of all possible samples of fixed size. When it comes to implementing an algorithm for drawing a maximum entropy sample, a very important result has been given by Chen et al. (1994). They have shown that the maximum entropy sampling design can be presented as a parametric exponential family, and they have proposed an algorithm that makes it possible to pass from its parameter to the first-order and the joint inclusion probabilities and vice versa. In a manuscript paper, Deville (2000) has improved this algorithm. Chen et al. (1994) and Deville (2000) pointed out that a fast computation of the parameter makes it possible to employ three methods: rejective sampling, sequential sampling, and draw by draw sampling. Deville (2000) has shown that the joint inclusion probabilities can be computed exactly by means of a recursive method, without enumerating the possible samples. Using this method, the variance for the Horvitz-Thompson estimator of the total

[1] Statistics Group, University of Neuchâtel, Espace de l'Europe 4, CP 805, 2002 Neuchâtel, Switzerland. Email: alina.matei@unine.ch; yves.tille@unine.ch

Acknowledgments: The authors would like to thank the Associate Editor and the anonymous referees for their useful suggestions and corrections.

population can be computed exactly. The joint inclusion probabilities are also used to compute the Horvitz-Thompson and Sen-Yates-Grundy variance estimators. Aires (1999) has provided another method to find the exact expression of the inclusion probabilities in the case of the rejective sampling.

The maximum entropy sampling with fixed sample size design is relatively recent and therefore it is not yet sufficiently used in practice. This sampling design, however, has several interesting points:

a. Generally, the maximization of the entropy consists of defining a sampling design as randomly as possible. It is a high entropy situation according to Brewer's (2002, p. 146) definition when "the resulting relationship between the population and the sample follows no particular pattern" and we expect the variance estimators to perform well. In particular the simple random sampling without replacement and the Poisson sampling are maximum entropy sampling designs.
b. In the case of fixed sample size, all the samples have strictly positive probabilities of being selected. Therefore the joint inclusion probabilities are strictly positive.
c. The joint inclusion probabilities do not depend on the order of the units, and can be easily computed.
d. The algorithm to compute the inclusion probabilities is fast, and particularly convenient for making simulations.
e. Finally, a simple asymptotic argument allows constructing a family of variance approximations and a large set of variance estimators.

Our aim is to review and evaluate a large set of variance approximations and variance estimators. These are generally applicable to unequal probability designs. We test seven approximations and 20 estimators of variance in several cases of maximum entropy sampling by means of a set of simulations. The ratio of bias and the mean squared errors under the simulations are derived. Coverage rates of interval estimates for the 95% level are reported. The simulations indicate that knowledge of the joint inclusion probabilities is not necessary to construct a reasonable estimator of variance in the case of the maximum entropy sampling design with fixed sample size and unequal probabilities.

The article is organized as follows. In Section 2, the notation is defined and the maximum entropy sampling design is reviewed. Interest is then focused on the algorithm which allows the transition from the parameter of the exponential family to the first- and second-order inclusion probabilities and vice versa. In Sections 3 and 4, several approximation and estimator expressions for the variance are reviewed. In Sections 5 and 6, the empirical results are presented in order to compare the different methods of approximation or estimation to the true value of the variance. Section 7 presents the concluding remarks.

2. The Maximum Entropy Sampling Design

2.1. *Definition and notation*

Let $U = \{1, \ldots, k, \ldots, N\}$ be a finite population of size N. A sample s is a subset of U. A support $\mathscr{R}(U)$ is a set of samples of U. Let $\mathcal{S}(U) = \{s \subset U\}$ be the full support on U

with $\#S(U) = 2^N$, and let $S_n(U) = \{s \subset U | \#s = n\}$ be the sample support with fixed sample size equal to n. A sampling design $p(s) > 0, s \in \mathscr{R}(U)$ is a probability distribution on $\mathscr{R}(U)$ such that $\sum_{s \in \mathscr{R}(U)} p(s) = 1$. Let S be a random sample such that $Pr[S = s] = p(s)$. The first-order inclusion probability is defined by

$$\pi_k = \Pr[k \in S] = \sum_{s \in \mathscr{R}(U) | s \ni k} p(s), k \in U$$

and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k, \ldots, \pi_N)'$ is the vector of inclusion probabilities.

The entropy of a sampling design $p(.)$ on $\mathscr{R}(U)$ is given by

$$I(p) = -\sum_{s \in \mathscr{R}(U)} p(s) \log p(s)$$

If we calculate the sampling design on $\mathscr{R}(U)$ which maximizes the entropy under the restrictions given by fixed inclusion probabilities, we get

$$p(s, \mathscr{R}(U), \boldsymbol{\lambda}) = \frac{\exp \boldsymbol{\lambda}' \mathbf{s}}{\sum_{z \in \mathscr{R}(U)} \exp \boldsymbol{\lambda}' \mathbf{z}} \tag{1}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^N$ is the vector of Lagrange multipliers, and $\mathbf{s}$ is a vector of $\mathbb{R}^N$ such that

$$s_k = \begin{cases} 1 & \text{if } k \in s \\ 0 & \text{if } k \notin s \end{cases}$$

Chen et al. (1994) pointed out that (1) belongs to the exponential family and $\boldsymbol{\lambda}$ is its parameter. One of the characteristics of the exponential family is that there exists a one to one correspondence between the parameter and the expectation (on this topic, see for instance Brown, 1986, p. 74). The expectation is the inclusion probability vector

$$\boldsymbol{\pi} = \sum_{s \in \mathscr{R}(U)} \mathbf{s} p(s)$$

Remark 1. The sampling design which maximizes the entropy on the full support $S(U)$, when the inclusion probabilities π_k for all $k \in U$ are fixed, is the Poisson sampling design (see Hájek 1981, p. 30). The interest of the Poisson sampling is the independence between the selection of the units, which allows a very simple sequential implementation. The disadvantage of Poisson sampling is its random sample size. For this reason, fixed sample size methods ore often used.

2.2. *Maximum entropy sampling design with fixed sample size*

When the support is $S_n(U)$, the problem becomes more intricate, because the denominator of

$$p(s, S_n(U), \lambda) = \frac{\exp \boldsymbol{\lambda}' \mathbf{s}}{\sum_{z \in S_n(U)} \exp \boldsymbol{\lambda}' \mathbf{z}}, \quad s \in S_n(U)$$

cannot be simplified. For this reason, one might believe (were it not for the paper of Chen et al. 1994) that it is not possible to select a sample with this design without enumerating all the samples of $S_n(U)$.

Let $\boldsymbol{\pi}(\boldsymbol{\lambda}, n)$ be the vector of inclusion probabilities for the maximum entropy sampling design with fixed sample size equal to n. The first problem is the derivation of $\boldsymbol{\pi}(\boldsymbol{\lambda}, n)$ from $\boldsymbol{\lambda}$, which is theoretically given by

$$\boldsymbol{\pi}(\boldsymbol{\lambda}, n) = \frac{\sum_{s \in S_n(U)} \mathbf{s} \exp \boldsymbol{\lambda}'\mathbf{s}}{\sum_{s \in S_n(U)} \exp \boldsymbol{\lambda}'\mathbf{s}} \tag{2}$$

Unfortunately, Expression (2) becomes unfeasible to compute when U is large, because it becomes impossible to enumerate all the samples. Nevertheless, Chen et al. (1994) have shown a recursive relation between $\boldsymbol{\pi}(\boldsymbol{\lambda}, n-1)$ and $\boldsymbol{\pi}(\boldsymbol{\lambda}, n)$, which allows to pass from $\boldsymbol{\lambda}$ to $\boldsymbol{\pi}(\boldsymbol{\lambda}, n)$, without enumerating all the possible samples of $S(U)$.

Result 1. (Chen et al. 1994) For the first-order inclusion probabilities of the maximum entropy fixed sample size (size equal with n)

$$\pi_k(\boldsymbol{\lambda}, n) = n \frac{\exp \lambda_k [1 - \pi_k(\boldsymbol{\lambda}, n-1)]}{\sum_{\ell \in U} \exp \lambda_\ell [1 - \pi_\ell(\boldsymbol{\lambda}, n-1)]} \tag{3}$$

A proof of Result 1 is given in Appendix 1. Since $\pi_k(\boldsymbol{\lambda}, 0) = 0$, for all $k \in U$, this recursive relation allows a fast computation of the inclusion probability vector.

Another recursive relation (Deville 2000) allows to compute the joint inclusion probabilities.

Result 2. (Deville 2000) For the joint inclusion probabilities of the maximum entropy fixed sample size (size equal to n)

$$\pi_{k\ell}(\boldsymbol{\lambda}, n) = \frac{n(n-1) \exp \lambda_k \exp \lambda_\ell [1 - \pi_k(\boldsymbol{\lambda}, n-2) - \pi_\ell[\boldsymbol{\lambda}, n-2) + \pi_{kl}[\boldsymbol{\lambda}, n-2)]}{\sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} \exp \lambda_i \exp \lambda_j [1 - \pi_i(\boldsymbol{\lambda}, n-2) - \pi_j(\boldsymbol{\lambda}, n-2) + \pi_{ij}(\boldsymbol{\lambda}, n-2)]}$$

with

$$\pi_{k\ell}(\boldsymbol{\lambda}, 0) = \pi_{k\ell}(\boldsymbol{\lambda}, 1) = 0, \pi_{k\ell}(\boldsymbol{\lambda}, 2) = \frac{2 \exp \lambda_k \exp \lambda_\ell}{\sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} \exp \lambda_i \exp \lambda_j}, k, \ell \in U, k \neq \ell$$

A proof of Result 2 is given in Appendix 2.

In practice, the inclusion probabilities are generally fixed, and the main problem is to compute $\boldsymbol{\lambda}$ from a given inclusion probability vector $\boldsymbol{\pi}$. The knowledge of $\boldsymbol{\lambda}$ permits one to calculate the inclusion probabilities and the joint inclusion probabilities for the maximum entropy with fixed sample size design using Results 1 and 2. It is important to point out that the first-order inclusion probabilities of the Poisson design (which maximizes the entropy, but does not have a fixed sample size), denoted by $\tilde{\boldsymbol{\pi}}$, are not the same as the inclusion probabilities of fixed sample size design, denoted by $\boldsymbol{\pi}$. Deville (2000) has shown that $\tilde{\boldsymbol{\pi}}$ can be obtained by means of Algorithm 1, which is an application of the Newton method. It is straightforward $\lambda_k = \log [\tilde{\pi}_k / (1 - \tilde{\pi}_k)]$.

A justification of Algorithm 1 is given in Appendix 3.

Algorithm 1. Computation of $\tilde{\boldsymbol{\pi}}$

- Define

$$\boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n) = n \frac{\frac{\tilde{\pi}_k}{1-\tilde{\pi}_k}\{1 - \phi_k(\tilde{\boldsymbol{\pi}}, n-1)\}}{\sum_{\ell \in U} \frac{\tilde{\pi}_\ell}{1-\tilde{\pi}_\ell}\{1 - \phi_\ell(\tilde{\boldsymbol{\pi}}, n-1)\}} \quad \text{with} \quad \boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, 0) = 0$$

- Set $\tilde{\boldsymbol{\pi}}^{(0)} = \boldsymbol{\pi}$ and for $i = 1, 2, \ldots,$ until convergence

$$\tilde{\boldsymbol{\pi}}^{(i)} = \tilde{\boldsymbol{\pi}}^{(i-1)} + \boldsymbol{\pi} - \boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}^{(i-1)}, n) \tag{4}$$

2.3. The rejective algorithm

Let y_k be the variable of interest associated with the kth individual in the population, and let $x_k > 0$ be an auxiliary variable, which is known for all $k \in U$. The first-order inclusion probabilities are computed using the relation

$$\pi_k = \frac{n x_k}{\sum_{\ell \in U} x_\ell} \tag{5}$$

for all $k \in U$, where n is the sample size. If some $\pi_k > 1$, the value 1 is allocated to these units, and the inclusion probabilities are recalculated using (5) on the remaining units.

The rejective procedure follows from Result 3.

Result 3. For all constant $c \in \mathbb{R}$

$$p(s, S_n(U), \boldsymbol{\lambda}) = p(s, S(U), \boldsymbol{\lambda} + c1 | \#S = n) = \frac{p(s, S(U), \boldsymbol{\lambda} + c\mathbf{1})}{\sum_{s \in S_n(U)} p(s, S(U), \boldsymbol{\lambda} + c\mathbf{1})}$$

for all $s \in S_n(U)$, where $\mathbf{1}$ is a vector N ones.

The proof is obvious. The rejective method can thus be defined in Algorithm 2.

Since the constant c can be any real number, it should be chosen in order to maximize $1/\Pr(\#S = n)$, which can be achieved by using the Newton algorithm. A simpler way to fix the value of c is using a constant such that

$$\sum_{k \in U} \tilde{\pi}_k = \sum_{k \in U} \frac{\exp(\lambda_k + c)}{1 + \exp(\lambda_k + c)} = n \tag{6}$$

Note that Algorithm 1 provides $\tilde{\pi}_k$'s that have directly such properties.

Algorithm 2. Rejective Poisson sampling

1. Given $\boldsymbol{\pi}$, compute $\tilde{\boldsymbol{\pi}}$ with Algorithm 1; next compute (if needed) $\boldsymbol{\lambda}$ by

$$\lambda_k = \log \frac{\tilde{\pi}_k}{1 - \tilde{\pi}_k}$$

Eventually, vector λ_k can be normalized such that

$$\sum_{k \in U} \lambda_k = 0$$

2. Select a random sample $\tilde{S}$, using Poisson design $p(\tilde{s}, S(U), \boldsymbol{\lambda} + c\mathbf{1})$. If it is not equal to n, repeat the selection until it is equal to n.

3. Variance Approximations for Unequal Probability Sampling

A review of some variance approximations and variance estimators is presented below. Our aim is to compare different variance approximations as well as different variance estimators for the Horvitz-Thompson estimator

$$\widehat{Y}_\pi = \sum_{k \in S} \frac{y_k}{\pi_k}$$

of the total population

$$Y = \sum_{k \in U} y_k$$

The variance of the Horvitz-Thompson estimator $\widehat{Y}_\pi$ for a fixed sample size is (see Yates and Grundy, 1953; Sen, 1953)

$$\text{var}\,[\widehat{Y}_\pi] = -\frac{1}{2} \sum_{k \in U} \sum_{\substack{\ell \in U \\ \ell \neq k}} \left(\frac{y_k}{\pi_k} - \frac{y_\ell}{\pi_\ell} \right)^2 (\pi_{k\ell} - \pi_k \pi_\ell) \tag{7}$$

Seven variance approximations and twenty variance estimators have been compared using simulations. The notation for each approximation and each estimator is given in the parenthesis in the corresponding paragraph (e.g., $\text{var}_{\text{Hajek}_1}$ for the approximation Hájek 1). For simplicity, in the next formulae, the first-order inclusion probabilities $\pi_k(\lambda, n)$ are denoted by π_k, and the joint inclusion probabilities $\pi_{k\ell}(\lambda, n)$ are denoted by $\pi_{k\ell}$.

Result 3 shows that a sampling design $p(s)$ which maximizes the entropy and has the inclusion probabilities π_k can be viewed as a conditional Poisson sampling design $\tilde{p}(s)$ given that its sample size $\tilde{n}_S$ is fixed. If $\text{var}_{poiss}(.)$ denotes the variance and $\text{cov}_{poiss}(.)$ the covariance under the Poisson sampling $\tilde{p}(s)$ and var(.) the variance under the design $p(.)$, we can write

$$\text{var}\,(\widehat{Y}_\pi) = \text{var}_{poiss}(\widehat{Y}_\pi | \tilde{n}_S = n)$$

If we suppose that the couple $(\widehat{Y}_\pi, \tilde{n}_S)$ has a bivariate normal distribution (on this topic see Hájek 1964; Berger 1998a), we obtain

$$\text{var}_{poiss}(\widehat{Y}_\pi | \tilde{n}_S = n) = \text{var}_{poiss}(\widehat{Y}_\pi + (n - \tilde{n}_S)\beta)$$

where

$$\beta = \frac{\text{cov}_{poiss}(\tilde{n}_S, \widehat{Y}_\pi)}{\text{var}_{poiss}(\tilde{n}_S)}$$

$$\text{var}_{poiss}(\tilde{n}_S) = \sum_{k \in U} \tilde{\pi}_k (1 - \tilde{\pi}_k)$$

and

$$\text{cov}_{poiss}(\tilde{n}_S, \widehat{Y}_\pi) = \sum_{k \in U} \tilde{\pi}_k (1 - \tilde{\pi}_k) \frac{y_k}{\pi_k}$$

Defining $b_k = \tilde{\pi}_k(1 - \tilde{\pi}_k)$, we get the following general approximation of the variance for a sampling design with maximum entropy (see Deville and Tillé 2005; Tillé 2001, p. 117)

$$\text{var}_{approx}[\widehat{Y}_\pi] = \sum_{k \in U} \frac{b_k}{\pi_k^2} (y_k - y_k^*)^2 \tag{8}$$

where

$$y_k^* = \pi_k \beta = \pi_k \frac{\sum_{\ell \in U} b_\ell y_\ell / \pi_\ell}{\sum_{\ell \in U} b_\ell}$$

According to the values given to b_k, some variants of this approximation are obtained and presented below.

Hájek Approximation 1 ($\textbf{var}_{\textbf{Hajek}_1}$)

The most common value for b_k has been proposed by Hájek (1981)

$$b_k = \frac{\pi_k(1 - \pi_k)N}{N - 1} \tag{9}$$

(on this topic see also Rosén 1997; Tillé 2001).

Approximation under sampling with replacement ($\textbf{var}_{\textbf{repl}}$)

A simpler value for b_k could be

$$b_k = \pi_k \frac{N}{N - 1} \tag{10}$$

which leads to the variance under sampling with replacement.

Naive Approximation ($\textbf{var}_{\textbf{naive}}$)

A finite population correction can be added to (10) and thus

$$b_k = \pi_k \frac{N - n}{N} \frac{N}{N - 1} = \pi_k \frac{N - n}{N - 1}$$

in order to obtain the variance of simple random sampling without replacement in the case of equal inclusion probabilities.

Fixed-point Approximation ($\textbf{var}_{\textbf{Fix}}$)

Deville and Tillé (2005) have proposed solving the following equation system to find another approximation of b_k:

$$b_k - \frac{b_k^2}{\sum_{\ell \in U} b_\ell} = \pi_k(1 - \pi_k) \tag{11}$$

Since the equation system (11) is not linear, the coefficients b_k can be obtained by the fixed-point technique, using the following recurrence equation, until convergence:

$$b_k^{(i)} = \frac{(b_k^{(i-1)})^2}{\sum_{l \in U} b_l^{(i-1)}} + \pi_k(1 - \pi_k) \tag{12}$$

for $i = 0, 1, 2, 3, \ldots$, and using the initialization:

$$b_k^{(0)} = \pi_k(1 - \pi_k)\frac{N}{N-1}, k \in U$$

A necessary condition in order that a solution exists in the equation above is

$$\frac{\pi_k(1 - \pi_k)}{\sum_{\ell \in U} \pi_\ell(1 - \pi_\ell)} < \frac{1}{2}, \text{for all } k \in U$$

If the method is not convergent, consider the following variant, which uses one iteration and

$$b_k^{(1)} = \pi_k(1 - \pi_k)\left(\frac{N\pi_k(1 - \pi_k)}{(N-1)\sum_{\ell \in U} \pi_\ell(1 - \pi_\ell)} + 1\right)$$

Hartley-Rao Approximation 1 ($\mathrm{var}_{\text{H-Rao}_1}$)

An approximation of variance for the randomized systematic sampling was presented by Hartley and Rao (1962) (see also Brewer and Hanif 1983):

$$\begin{aligned}\mathrm{var}_{\text{H-Rao}_1}(Y) = {} & \sum_{k \in U} \pi_k\left(1 - \frac{n-1}{n}\pi_k\right)\left(\frac{y_k}{\pi_k} - \frac{Y}{n}\right)^2 \\ & - \frac{n-1}{n^2}\sum_{k \in U}\left(2\pi_k^3 - \frac{\pi_k^2}{2}\sum_{\ell \in U}\pi_\ell^2\right)\left(\frac{y_k}{\pi_k} - \frac{Y}{n}\right)^2 \\ & + \frac{2(n-1)}{n^3}\left(\sum_{k \in U}\pi_k y_k - \frac{Y}{n}\sum_{\ell \in U}\pi_\ell^2\right)^2\end{aligned}$$

Hartley-Rao Approximation 2 ($\mathrm{var}_{\text{H-Rao}_2}$)

In the same paper, Hartley and Rao (1962) have also suggested a simpler expression of variance (see also Brewer and Hanif 1983):

$$\mathrm{var}_{\text{H-Rao}_2}(Y) = \sum_{k \in U} \pi_k\left(1 - \frac{n-1}{n}\pi_k\right)\left(\frac{y_k}{\pi_k} - \frac{Y}{n}\right)^2 \tag{13}$$

Hájek Approximation 2 ($\mathbf{var}_{\mathbf{Hajek}_2}$)

Brewer (2002, p. 153) has used the following estimator, starting from Hájek (1964):

$$\mathrm{var}_{\mathrm{Hajek}_2}(Y) = \sum_{k \in U} \pi_k (1 - \pi_k) \left(\frac{y_k}{\pi_k} - \frac{\tilde{Y}}{n} \right)^2 \quad (14)$$

where $\tilde{Y} = \sum_{k \in U} a_k y_k$ and $a_k = n(1 - \pi_k) / \sum_{\ell \in U} \pi_\ell (1 - \pi_\ell)$

4. Variance Estimators

There are three classes of variance estimators. The first class is composed of the Horvitz-Thompson estimator (Horvitz and Thompson 1952) and the Sen-Yates-Grundy estimator (Yates and Grundy 1953; Sen 1953), which use the first-order and the joint inclusion probabilities. The second class uses only first-order inclusion probabilities for all $k \in S$ while in the third class, the variance estimators use only first-order inclusion probabilities, but for all $k \in U$.

4.1. First class of variance estimators

Horvitz-Thompson Estimator ($\widehat{\mathbf{var}}_{\mathbf{HT}}$)

The expression of this estimator is (see Horvitz and Thompson 1952)

$$\widehat{\mathrm{var}}_{\mathrm{HT}}[\widehat{Y}_\pi] = \sum_{k \in S} \frac{y_k^2}{\pi_k^2} (1 - \pi_k) + \sum_{k \in S} \sum_{\substack{\ell \in S \\ \ell \neq k}} \frac{y_k y_\ell}{\pi_k \pi_\ell \pi_{k\ell}} (\pi_{k\ell} - \pi_k \pi_\ell) \quad (15)$$

This estimator has several important drawbacks. In general, when the variable of interest $y_k \propto \pi_k$, $\mathrm{var}[\widehat{Y}_\pi] = 0$, but $\widehat{\mathrm{var}}_{\mathrm{HT}}$ is not necessarily equal to 0 in such a case. The Horvitz-Thompson estimator can also take negative values (on this topic see Cumberland and Royall 1981). For example, if $y_k = \pi_k$, for all $k \in U$, then $\mathrm{var}[\widehat{Y}_\pi] = 0$, and

$$\widehat{\mathrm{var}}_{\mathrm{HT}}[\widehat{Y}_\pi] = n^2 - \sum_{k \in S} \pi_k - \sum_{k \in S} \pi_k \sum_{\substack{\ell \in S \\ \ell \neq k}} \frac{\pi_\ell}{\pi_{k\ell}}$$

which is generally not null, but has a null expectation. Thus, negative values occur.

Sen-Yates-Grundy Estimator ($\widehat{\mathbf{var}}_{\mathbf{SYG}}$)

The expression of this estimator is (see Sen 1953; Yates and Grundy 1953)

$$\widehat{\mathrm{var}}_{\mathrm{SYG}}[\widehat{Y}_\pi] = \frac{1}{2} \sum_{k \in S} \sum_{\substack{\ell \in S \\ \ell \neq k}} \left(\frac{y_k}{\pi_k} - \frac{y_\ell}{\pi_\ell} \right)^2 \frac{\pi_k \pi_\ell - \pi_{k\ell}}{\pi_{k\ell}} \quad (16)$$

The Horvitz-Thompson and Sen-Yates-Grundy estimators are unbiased.

4.2. Second class of variance estimators

From expression (8), a general variance estimator can be derived (see Deville and Tillé 2005; Tillé 2001, p. 117):

$$\widehat{\mathrm{var}}[\widehat{Y}_\pi] = \sum_{k \in S} \frac{c_k}{\pi_k^2} (y_k - \widehat{y}_k^*)^2 \tag{17}$$

where

$$\widehat{y}_k^* = \pi_k \frac{\sum_{\ell \in S} c_\ell y_\ell / \pi_\ell}{\sum_{\ell \in S} c_\ell}$$

According to the choice of c_k in (17), various estimators have been proposed.

Deville Estimator 1 ($\widehat{\mathbf{var}}_{\mathbf{Dev1}}$)
Deville (1993) has proposed a simple value for c_k:

$$c_k = (1 - \pi_k) \frac{n}{n-1}$$

Deville Estimator 2 ($\widehat{\mathbf{var}}_{\mathbf{Dev2}}$)
In the same manuscript, Deville (1993) has suggested a more complex value (see also Deville 1999):

$$c_k = (1 - \pi_k) \left[1 - \sum_{k \in S} \left\{ \frac{1 - \pi_k}{\sum_{\ell \in S}(1 - \pi_\ell)} \right\}^2 \right]^{-1}$$

Variance under sampling with replacement ($\widehat{\mathbf{var}}_{\mathbf{repl}}$)
A simple value for c_k could be

$$c_k = \frac{n}{n-1} \tag{18}$$

which leads to the variance under sampling with replacement (see Särndal et al. 1992, expression 2.9.9, p. 53).

Naive Estimator ($\widehat{\mathbf{var}}_{\mathbf{naive}}$)
A finite population correction can be added to (18), resulting in

$$c_k = \frac{N-n}{N} \frac{n}{n-1}$$

in order to obtain the variance estimator of simple random sampling without replacement, in the case of equal inclusion probabilities.

Fixed-point Estimator ($\widehat{\mathbf{var}}_{\mathbf{Fix}}$)
Deville and Tillé (2005) have proposed using the following development in order to derive a value for c_k. The estimator defined in expression (17) can be written as

$$\widehat{\mathrm{var}}[\widehat{Y}_\pi] = \sum_{k \in S} \frac{y_k^2}{\pi_k^2} \left(c_k - \frac{c_k^2}{\sum_{\ell \in S} c_\ell} \right) - \frac{1}{\sum_{\ell \in S} c_\ell} \sum_{k \in S} \sum_{\substack{\ell \in S \\ \ell \neq k}} \frac{y_k y_\ell c_k c_\ell}{\pi_k \pi_\ell} \tag{19}$$

Using the formula (15) of $\widehat{\mathrm{var}}_{\mathrm{HT}}$, we can look for c_k which satisfies the equation

$$c_k - \frac{c_k^2}{\sum_{\ell \in S} c_\ell} = (1 - \pi_k)$$

These coefficients can be obtained by the fixed-point technique, using the following recurrence equation, until the convergence is fulfilled:

$$c_k^{(i)} = \frac{\left(c_k^{(i-1)}\right)^2}{\sum_{\ell \in S} c_\ell^{(i-1)}} + (1 - \pi_k)$$

for $i = 0, 1, 2, 3, \ldots$ and using the initialization

$$c_k^{(0)} = (1 - \pi_k)\frac{n}{n-1}, k \in S$$

A necessary condition in order that a solution exists in the equation above is:

$$\frac{1 - \pi_k}{\sum_{\ell \in S}(1 - \pi_\ell)} < \frac{1}{2}, \text{for all } k \in S$$

If the method is not convergent, consider the previous variant, which uses one iteration:

$$c_k^{(1)} = (1 - \pi_k)\left(\frac{n(1 - \pi_k)}{(n-1)\sum_{\ell \in S}(1 - \pi_\ell)} + 1\right)$$

Rosén Estimator ($\widehat{\mathbf{var}}_{\mathbf{R}}$)

Rosén (1991) suggested the following estimator (see also Ardilly 1994, p. 338):

$$\widehat{\mathrm{var}}_{\mathrm{R}}[\widehat{Y}_\pi] = \frac{n}{n-1}\sum_{k \in S}(1 - \pi_k)\left(\frac{y_k}{\pi_k} - A\right)^2$$

where

$$A = \frac{\sum_{k \in S} y_k \frac{1-\pi_k}{\pi_k^2} \log(1 - \pi_k)}{\sum_{k \in S} \frac{1-\pi_k}{\pi_k} \log(1 - \pi_k)} \tag{20}$$

Deville Estimator 3 ($\widehat{\mathbf{var}}_{\mathbf{Dev_3}}$)

Another proposal of Deville (1993) (see also Ardilly 1994, p. 338) is

$$\widehat{\mathrm{var}}_{\mathrm{Dev3}}[\widehat{Y}_\pi] = \frac{1}{1 - \sum_{k \in S} a_k^2}\sum_{k \in S}(1 - \pi_k)\left(\frac{y_k}{\pi_k} - \frac{\widehat{Y}_\pi}{n}\right)^2 \tag{21}$$

where

$$a_k = \frac{1 - \pi_k}{\sum_{k \in S}(1 - \pi_k)}$$

Estimator 1 ($\widehat{\mathbf{var}}_{\mathbf{1}}$)

We propose a new estimator which is defined as

$$\widehat{\mathrm{var}}_1[\widehat{Y}_\pi] = \frac{n(N-1)}{N(n-1)} \sum_{k \in S} \frac{b_k}{\pi_k^3} (y_k - \widehat{y}_k^*)^2 \tag{22}$$

where

$$\widehat{y}_k^* = \pi_k \frac{\sum_{\ell \in S} b_\ell y_\ell / \pi_\ell^2}{\sum_{\ell \in S} b_\ell / \pi_\ell}$$

and the coefficients b_k are defined in the same way as in Expression (12).

4.3. *Third class of variance estimators*

Berger Estimator ($\widehat{\mathbf{var}}_{\mathbf{Ber}}$)

Berger (1998b) has proposed using

$$c_k = (1 - \pi_k) \frac{n}{n-1} \frac{\sum_{k \in S} (1 - \pi_k)}{\sum_{k \in U} \pi_k (1 - \pi_k)}$$

in Expression (17).

Tillé Estimator ($\widehat{\mathbf{var}}_{\mathbf{T}}$)

An approximation of the joint inclusion probabilities by means of adjustment to marginal totals was described by Tillé (1996). Using this approximation and the Sen-Yates-Grundy estimator, the following estimator was developed:

$$\widehat{\mathrm{var}}_{\mathrm{T}}[\widehat{Y}_\pi] = \sum_{k \in S} \frac{y_k^2}{\pi_k \beta_k} \sum_{\ell \in S} \frac{\pi_\ell}{\beta_l} - \left(\sum_{k \in S} \frac{y_k}{\beta_k} \right)^2 - n \sum_{k \in S} \frac{y_k^2}{\pi_k^2} + \left(\sum_{k \in S} \frac{y_k}{\pi_k} \right)^2 \tag{23}$$

The coefficients β_k are calculated using the following algorithm:

$$\beta_k^{(0)} = \pi_k, \text{for all } k, \quad \beta_k^{(2i-1)} = \frac{(n-1)\pi_k}{\beta^{(2i-2)} - \beta_k^{(2i-2)}} \quad \text{and}$$

$$\beta_k^{(2i)} = \beta_k^{(2i-1)} \left(\frac{n(n-1)}{(\beta^{(2i-1)})^2 - \sum_{k \in U} (\beta_k^{(2i-1)})^2} \right)^{1/2}$$

where

$$\beta^{(i)} = \sum_{k \in U} \beta_k^{(i)}, i = 1, 2, 3, \ldots$$

The coefficients β_k are used to approximate the joint inclusion probabilities such that $\pi_{k\ell} \approx \beta_k \beta_\ell$. The convergence criterion is ensured by the marginal totals

$$\sum_{\substack{k \in U \\ k \neq \ell}} \pi_{k\ell} = \pi_\ell (n-1), \ell \in U$$

4.3.1. Some new estimators

Four new variance estimators (named Estimators 2, 3, 4, 5) can be constructed as follows.

Estimator 2 ($\widehat{\mathbf{var}}_2$)

$$\widehat{\mathrm{var}}_2[\widehat{Y}_\pi] = \frac{1}{1 - \sum_{k \in U} \frac{d_k^2}{\pi_k}} \sum_{k \in S} (1 - \pi_k) \left(\frac{y_k}{\pi_k} - \frac{\widehat{Y}_\pi}{n} \right)^2 \tag{24}$$

where

$$d_k = \frac{\pi_k (1 - \pi_k)}{\sum_{\ell \in U} \pi_\ell (1 - \pi_\ell)} \tag{25}$$

Estimator 3 ($\widehat{\mathbf{var}}_3$)

$$\widehat{\mathrm{var}}_3[\widehat{Y}_\pi] = \frac{1}{1 - \sum_{k \in U} \frac{d_k^2}{\pi_k}} \sum_{k \in S} (1 - \pi_k) \left(\frac{y_k}{\pi_k} - \frac{\sum_{\ell \in S} (1 - \pi_\ell) \frac{y_\ell}{\pi_\ell}}{\sum_{\ell \in S} (1 - \pi_\ell)} \right)^2 \tag{26}$$

where d_k is defined as in (25).

Estimator 4 ($\widehat{\mathbf{var}}_4$)

$$\widehat{\mathrm{var}}_4[\widehat{Y}_\pi] = \frac{1}{1 - \sum_{\ell \in U} b_\ell / n^2} \sum_{k \in S} \frac{b_k}{\pi_k^3} (y_k - \widehat{y}_k^*)^2 \tag{27}$$

where

$$\widehat{y}_k^* = \pi_k \frac{\sum_{\ell \in S} b_\ell y_\ell / \pi_\ell^2}{\sum_{\ell \in S} b_\ell / \pi_\ell} \tag{28}$$

and the coefficients b_k are defined in the same way as in Expression (9).

Estimator 5 ($\widehat{\mathbf{var}}_5$)

$$\widehat{\mathrm{var}}_5[\widehat{Y}_\pi] = \frac{1}{1 - \sum_{\ell \in U} b_\ell / n^2} \sum_{k \in S} \frac{b_k}{\pi_k^3} (y_k - \widehat{y}_k^*)^2 \tag{29}$$

where $\widehat{y}_k^*$ is defined as in (28) and the coefficients b_k are defined in the same way as in Expression (12).

4.3.2. The Brewer family

A set of high-entropy estimators was presented by Brewer (2002) and by Brewer and Donadio (2003). According to Brewer, a plausible sample estimator of the approximate design variance of the Horvitz-Thompson estimator, and one that can be constructed so as to be exactly design-unbiased under simple random sampling without replacement, is

$$\widehat{\text{var}}_{\text{Br}}[\widehat{Y}_\pi] = \sum_{k \in S} (e_k^{-1} - \pi_k) \left(\frac{y_k}{\pi_k} - \sum_{\ell \in S} \frac{y_\ell}{n \pi_\ell} \right)^2 \tag{30}$$

Four particular values for e_k were proposed (see Brewer 2002, pp. 152, 153, 158):

Brewer Estimator 1 ($\widehat{\mathbf{var}}_{\mathbf{Br_1}}$)

$$e_k = \frac{n-1}{n - \frac{\sum_{\ell \in U} \pi_\ell^2}{n}}$$

Brewer Estimator 2 ($\widehat{\mathbf{var}}_{\mathbf{Br_2}}$)

$$e_k = \frac{n-1}{n - \pi_k}$$

In this case, the estimator defined in (30) could have been placed in the second category, since it uses the inclusion probabilities only for the sample. In order, however, to keep the Brewer estimators in a single category, we place it here.

Brewer Estimator 3 ($\widehat{\mathbf{var}}_{\mathbf{Br_3}}$)

$$e_k = \frac{(n-1)/n}{1 - 2\pi_k/n + \frac{\sum_{\ell \in U} \pi_\ell^2}{n^2}}$$

Brewer Estimator 4 ($\widehat{\mathbf{var}}_{\mathbf{Br_4}}$)

$$e_k = \frac{(n-1)/n}{1 - \frac{(2n-1)\pi_k}{n(n-1)} + \frac{\sum_{\ell \in U} \pi_\ell^2}{n(n-1)}}$$

5. Simulations

Three data sets have been used for Monte-Carlo simulations: the mu284 population from Särndal et al. (1992), and two artificial populations. A set of 10,000 independent samples without replacement have been selected for each different sample size, $n = 10, 20$, and 40, using the rejective sampling. Table 3 gives the expected number of the rejected samples (which have sample sizes different from the fixed size n) under the simulations. From the mu284 population, two data items have been taken: the "revenues from 1985 municipal taxation" for the principal characteristic, and the "1985 population" for the auxiliary variable. Three observations (numbers 16, 114, 137) with large x_k were deleted from this population. Thus, $N = 281$, $Y = 53,151 \times 10^6$. The first artificial population was generated using the model $N = 100$, $x_k = k$, $y_k = 5x_k(1 + \varepsilon_k)$ (see Fig. 1), where $\varepsilon_k \sim N(0, 1/3)$, $k = 1, \cdots, N$. In this case, $Y = 25,482.917$. For the second artificial population the model used is $N = 100$, $x_k = k$, $y_k = 1/\pi_k$, $\pi_k = nx_k / \sum_{k \in U} x_k$, $k = 1, \cdots, N$ (see Fig. 2). In this case, for $n = 10$, $Y = 2,619.625$, for $n = 20$, $Y =$

$1,309.812$, and for $n = 40$, $Y = 654.906$. Three measures are used to compare the variance estimators:

- the ratio of bias

$$\mathrm{RB}(\widehat{\mathrm{var}}) = \frac{\mathrm{E}_{sim}(\widehat{\mathrm{var}}) - \mathrm{var}}{\sqrt{\mathrm{var}_{sim}(\widehat{\mathrm{var}})}}$$

  where $\mathrm{E}_{sim}()$ is the average calculated under simulations, $\mathrm{var}_{sim}()$ is the variance calculated under simulations, $\widehat{\mathrm{var}}$ is a variance estimator, and var is the true variance computed from Expression (7).
- the mean squared error

$$\mathrm{MSE}(\widehat{\mathrm{var}}) = \mathrm{var}_{sim}(\widehat{\mathrm{var}}) + (\mathrm{E}_{sim}(\widehat{\mathrm{var}}) - \mathrm{var})^2$$

- the coverage rate (CR) of an interval estimate for the 95% level.

The 95% confidence intervals for the value Y are computed using the t distribution with $n - 1$ degrees of freedom $[Y_\pi \pm t_{n-1,0.975}\sqrt{\widehat{\mathrm{var}}}]$. We use the 97.5 quantile of the t–distribution with $n - 1$ degrees of freedom instead of 1.96 even for $n = 40$ to improve the coverage rate (see Särndal et al. 1992, p. 281).

Table 1 (for the mu284 population), Table 2 (for the first artificial population) and Table 4 (for the second artificial population) summarize the performance of the approximations and estimators via simulations. The upper sections of Tables 1, 2, and 4 give the values of the variance approximations presented in Section 3, and the true value, var $[\widehat{Y}_\pi]$, computed from (7). The bottom sections of these tables give the values of RB, MSE, and CR for the variance estimators presented in Section 4. The ratio of the bias and the coverage rates are expressed in percentages. For clarity, the exponents are added in brackets (for example the Hájek approximation 1 is 3.808×10^{18} in the case of the mu284 population, $n = 10$).

6. Discussion of the Empirical Results

The reliable comparison between the different variance approximations is ensured by the fact that the true variance $\mathrm{var}[\widehat{Y}_\pi]$ can be calculated using Formula (7). Without any doubt,

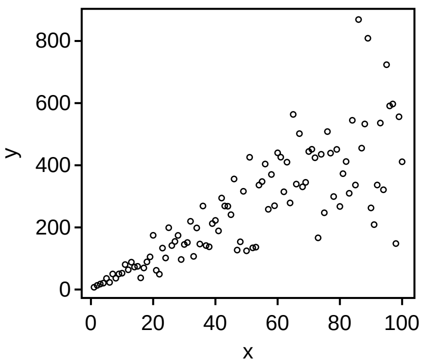

Fig. 1. Scatter plot for the first artificial population (x versus y)

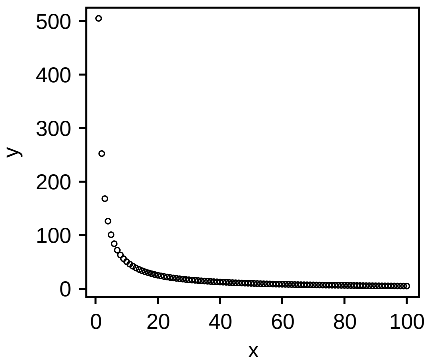

Fig. 2. Scatter plot for the second artificial population, $n = 10$ (x versus y)

the fixed-point approximation is the best. The approximations $\text{var}_{\text{Hajek}_1}$, $\text{var}_{\text{H-Rao}_1}$, $\text{var}_{\text{H-Rao}_2}$ and $\text{var}_{\text{Hajek}_2}$ are less precise. The worst results are given by var_{repl} (particularly in the case of the first two populations) and $\text{var}_{\text{naive}}$ (for all populations).

In the case of the variance estimators, the Horvitz-Thompson estimator has the largest MSE in Tables 1 and 2. In both populations, the variable of interest y_k and the auxiliary variable x_k are strongly correlated (for the mu284 population the coefficient of correlation is 0.99, and for the first artificial population 0.86). For the third population (which is badly adapted to the design), the correlation coefficient is approximately -0.40. In this case, $\widehat{\text{var}}_{\text{HT}}$ performed nearly the same as the other estimators studied. We are led to the same conclusion using an additional simulation study (results not shown in tables), where the variable of interest y_k and the auxiliary variable x_k are not correlated. As can be seen from the examples above, $\widehat{\text{var}}_{\text{HT}}$ has a large MSE in the cases where y_k and x_k are strongly correlated, which is the usual case in practice. An analytic study of the Horvitz-Thompson variance estimator is given in Stehman and Overton (1994).

Both population mu284 and the first artificial population arise from a structural model of the form $E(\mathbf{y}) = \beta\mathbf{x}$, $\text{var}(\mathbf{y}) = \sigma^2\mathbf{x}^2$. In such populations, for sufficiently small sample mean $\mathbf{x}$ and $\beta^2/\sigma^2 > 1$, Cumberland and Royall (1981) showed that $\widehat{\text{var}}_{\text{HT}}$ may take negative values. In artificial Population 1, $\beta^2/\sigma^2 = 9$. We included in Table 5 the number of times that $\widehat{\text{var}}_{\text{HT}} < 0$ among the 10,000 simulated samples. This could partially explain the large MSE for $\widehat{\text{var}}_{\text{HT}}$ in Population mu284 and artificial Population 1.

When it comes to the Sen-Yates-Grundy estimator as compared to the rest of the estimators (without taking into account $\widehat{\text{var}}_{\text{HT}}$, $\widehat{\text{var}}_{\text{repl}}$ and $\widehat{\text{var}}_{\text{naive}}$), which use only the first-order inclusion probabilities, we see that no big differences in the variance estimation are revealed by the simulations. However, for the second case and in Table 2, $\widehat{\text{var}}_{\text{SYG}}$ does not perform better than the other estimators. From the above and if we take seriously into account the fact that the Sen-Yates-Grundy estimator uses both first- and second-order inclusion probabilities (which makes it harder to compute), we find no reason why it should be preferred to the other estimators.

Table 1. Results of simulations for the mu284 population

		$n = 10\ (10^{18})$			$n = 20\ (10^{18})$			$n = 40\ (10^{18})$	
$\text{var}_{\text{Hajek}_1}$		3.808			1.778			1.005	
var_{Fix}		3.816			1.782			1.007	
$\text{var}_{\text{H-Rao}_1}$		3.818			1.788			1.059	
$\text{var}_{\text{H-Rao}_2}$		3.821			1.789			1.043	
$\text{var}_{\text{Hajek}_2}$		3.794			1.772			1.002	
var_{repl}		4.056			2.031			1.455	
$\text{var}_{\text{naive}}$		3.912			1.887			1.248	
True value		3.817			1.782			1.007	
	RB (%)	MSE (10^{36})	CR (%)	RB (%)	MSE (10^{35})	CR (%)	RB (%)	MSE (10^{34})	CR (%)
$\widehat{\text{var}}_{\text{HT}}$	−1.285	24.239	72.69	0.122	83.829	68.75	−0.657	225.587	71.87
$\widehat{\text{var}}_{\text{SYG}}$	−0.666	6.145	95.19	−0.013	6.041	95.20	−0.638	5.735	94.81
$\widehat{\text{var}}_{\text{Dev1}}$	−0.862	6.097	95.16	−0.343	5.939	95.27	−0.871	5.603	94.83
$\widehat{\text{var}}_{\text{Dev2}}$	−0.814	6.101	95.17	−0.173	5.947	95.27	−0.192	5.621	94.85
$\widehat{\text{var}}_{\text{repl}}$	−7.339	5.655	94.89	18.168	7.221	96.05	134.734	2.679	97.97
$\widehat{\text{var}}_{\text{naive}}$	3.414	6.466	95.45	13.307	6.800	95.85	90.026	1.333	97.04
$\widehat{\text{var}}_{\text{Fix}}$	−0.698	6.104	95.15	−0.054	5.948	95.20	−0.824	5.612	94.84
$\widehat{\text{var}}_{\text{R}}$	−0.835	6.098	95.17	−0.183	5.943	95.27	1.478	5.644	94.89
$\widehat{\text{var}}_{\text{Dev3}}$	−0.699	6.104	95.18	0.539	5.963	95.30	12.945	5.919	95.18
$\widehat{\text{var}}_{1}$	−0.494	6.129	95.22	0.139	5.969	95.29	−0.146	5.626	94.85
$\widehat{\text{var}}_{\text{Ber}}$	−0.697	6.141	95.19	−0.118	6.031	95.23	−0.762	5.719	94.81
$\widehat{\text{var}}_{\text{T}}$	−0.593	6.148	95.19	0.565	6.053	95.28	11.509	6.500	95.15
$\widehat{\text{var}}_{2}$	−0.694	6.105	95.18	0.546	5.964	95.30	12.949	5.918	95.18

Table 1. Continued

	$n = 10\ (10^{18})$			$n = 20\ (10^{18})$			$n = 40\ (10^{18})$		
$\widehat{\text{var}}_3$	−0.808	6.102	95.17	−0.166	5.948	95.28	−0.187	5.620	94.85
$\widehat{\text{var}}_4$	−1.429	6.104	95.14	−1.146	5.979	95.24	−2.112	5.755	94.77
$\widehat{\text{var}}_5$	−1.019	6.089	95.16	−0.645	5.929	95.25	−1.409	5.594	94.80
$\widehat{\text{var}}_{\text{Br1}}$	−0.869	6.093	95.20	0.177	5.944	95.29	12.746	5.885	95.17
$\widehat{\text{var}}_{\text{Br2}}$	−0.748	6.101	95.17	0.369	5.955	95.29	12.278	5.889	95.18
$\widehat{\text{var}}_{\text{Br3}}$	−0.627	6.109	95.17	0.561	5.966	95.30	11.809	5.894	95.14
$\widehat{\text{var}}_{\text{Br4}}$	−0.614	6.111	95.17	0.571	5.966	95.29	11.797	5.894	95.14

Table 2. Results of simulations for the first artificial population

	$n=10$ (10^6)			$n=20$ (10^6)			$n=40$ (10^6)		
$\mathrm{var}_{\mathrm{Hajek}_1}$		5.429			2.306			0.745	
$\mathrm{var}_{\mathrm{Fix}}$		5.444			2.312			0.746	
$\mathrm{var}_{\mathrm{H\text{-}Rao}_1}$		5.441			2.331			0.858	
$\mathrm{var}_{\mathrm{H\text{-}Rao}_2}$		5.455			2.324			0.758	
$\mathrm{var}_{\mathrm{Hajek}_2}$		5.374			2.283			0.737	
$\mathrm{var}_{\mathrm{repl}}$		6.245			3.123			1.563	
$\mathrm{var}_{\mathrm{naive}}$		5.621			2.499			0.938	
True value		5.444			2.312			0.746	
	RB (%)	MSE (10^{12})	CR (%)	RB (%)	MSE (10^{11})	CR (%)	RB (%)	MSE (10^{10})	CR (%)
$\widehat{\mathrm{var}}_{\mathrm{HT}}$	−0.758	7.504	93.71	0.358	8.528	93.09	−1.800	12.931	89.76
$\widehat{\mathrm{var}}_{\mathrm{SYG}}$	−0.827	6.979	94.80	0.669	5.663	94.89	−0.147	2.949	95.31
$\widehat{\mathrm{var}}_{\mathrm{Dev1}}$	−0.949	6.914	94.78	0.299	5.503	94.88	−2.338	2.582	95.19
$\widehat{\mathrm{var}}_{\mathrm{Dev2}}$	−0.887	6.917	94.78	0.557	5.512	94.88	−0.428	2.601	95.24
$\widehat{\mathrm{var}}_{\mathrm{repl}}$	6.444	7.254	95.11	73.064	12.878	97.08	285.872	68.302	99.33
$\widehat{\mathrm{var}}_{\mathrm{naive}}$	6.444	7.254	95.11	25.752	6.349	95.65	114.931	6.545	97.41
$\widehat{\mathrm{var}}_{\mathrm{Fix}}$	−0.895	6.936	94.78	0.581	5.534	94.89	−0.524	2.622	95.23
$\widehat{\mathrm{var}}_{\mathrm{R}}$	−0.936	6.914	94.78	0.346	5.504	94.88	−2.074	2.583	95.19
$\widehat{\mathrm{var}}_{\mathrm{Dev3}}$	−0.833	6.919	94.78	0.774	5.518	94.90	1.115	2.609	95.29
$\widehat{\mathrm{var}}_{1}$	−0.350	6.939	94.79	1.047	5.519	94.89	−1.576	2.587	95.19
$\widehat{\mathrm{var}}_{\mathrm{Ber}}$	−0.848	6.973	94.79	0.565	5.648	94.89	−1.165	2.917	95.28
$\widehat{\mathrm{var}}_{\mathrm{T}}$	−0.826	6.978	94.80	0.663	5.661	94.89	−0.258	2.921	95.29

Table 2. Continued

	$n = 10\ (10^6)$			$n = 20\ (10^6)$			$n = 40\ (10^6)$		
$\widehat{\text{var}}_2$	−0.826	6.920	94.78	0.787	5.518	94.90	1.152	2.610	95.29
$\widehat{\text{var}}_3$	−0.880	6.918	94.78	0.569	5.513	94.88	−0.392	2.602	95.24
$\widehat{\text{var}}_4$	−1.699	6.962	94.75	−0.781	5.601	94.84	−3.848	2.659	95.16
$\widehat{\text{var}}_5$	−1.052	6.893	94.77	−0.010	5.481	94.80	−3.147	2.572	95.17
$\widehat{\text{var}}_{\text{Br1}}$	−0.935	6.893	94.77	0.426	5.480	94.87	−1.003	2.559	95.23
$\widehat{\text{var}}_{\text{Br2}}$	−0.895	6.916	94.78	0.514	5.508	94.89	−0.791	2.588	95.22
$\widehat{\text{var}}_{\text{Br3}}$	−0.855	6.939	94.77	0.601	5.537	94.88	−0.581	2.617	95.24
$\widehat{\text{var}}_{\text{Br4}}$	−0.850	6.942	94.77	0.605	5.539	94.88	−0.576	2.618	95.25

Table 3. Expected number of the rejected samples under the simulations

	$n = 10$	$n = 20$	$n = 40$
mu284	6.643	9.457	12.372
artificial pop. 1	6.535	8.715	9.841
artificial pop. 2	6.391	8.634	10.013

Concerning the bias, the unbiased Horvitz-Thompson and Sen-Yates-Grundy estimators show nonzero bias due to the measurement error contingent on the finite size of the simulations.

The estimator with replacement and the naive estimator are highly biased in the first two populations and overestimate the variance. Therefore the coverage rates are very good in the case of these populations. In the third population, $\widehat{\text{var}}_{\text{repl}}$ and $\widehat{\text{var}}_{\text{naive}}$ perform better than all the other estimators concerning the RB and MSE, but we must take into account that this population is badly adapted to a real case.

The estimators (different from $\widehat{\text{var}}_{\text{HT}}, \widehat{\text{var}}_{\text{SYG}}, \widehat{\text{var}}_{\text{repl}}, \widehat{\text{var}}_{\text{naive}}$) which use only the first-order inclusion probabilities have similar performances and deserve consideration as practical alternatives. However, in the first population study, which is a real case, for $n = 40$, the estimators $\widehat{\text{var}}_{\text{Dev3}}, \widehat{\text{var}}_{T}, \widehat{\text{var}}_{2}, \widehat{\text{var}}_{\text{Br1}}, \widehat{\text{var}}_{\text{Br2}}, \widehat{\text{var}}_{\text{Br3}}$ and $\widehat{\text{var}}_{\text{Br4}}$ get highly biased.

Concerning the coverage rate, $\widehat{\text{var}}_{\text{HT}}$ gives poor coverage rates compared to all the other estimators in the first two populations. Its coverage ranges from 68.75% to 72.69% in the mu284 population and from 89.76% to 93.71% in artificial Population 1. The estimator $\widehat{\text{var}}_{\text{SYG}}$ gives better coverage percentages, and lies closer to the other estimators (without taking into account $\widehat{\text{var}}_{\text{repl}}$ and $\widehat{\text{var}}_{\text{naive}}$). In the first two populations the coverage rate is close to the nominal 95% for all the presented estimators (without $\widehat{\text{var}}_{\text{HT}}, \widehat{\text{var}}_{\text{repl}}, \widehat{\text{var}}_{\text{naive}}$). In the same populations, the estimators $\widehat{\text{var}}_{\text{repl}}$ and $\widehat{\text{var}}_{\text{naive}}$ give very rich coverage rates with coverage above the nominal 95%.

Artificial Population 2 is a special case: all the presented estimators give very poor coverage rates from 35% to 40%. This is due to the fact that "the exactness of the normal approximation used in computation of the 95% confidence interval depends significantly on the shape of the finite population" and "we can expect the approach to normality of the variable $(\widehat{Y}_\pi - Y)/\sqrt{\widehat{\text{var}}(\widehat{Y}_\pi)}$ to be slower" in the case of a highly skewed population, or with outlying values or other abnormal features (see Remark 2.11.2 in Särndal et al. 1992, p. 57). Fig. 3 gives the histogram for $\mathbf{y}$. Artificial Population 2 is highly skewed (for example for $n = 10$, $\gamma_1 = 6.019$) and has 12 outlying observations. We have deleted these 12 observations and we have rerun the simulations. Even if the nominal 95% was not reached, the CR was highly improved for all the presented estimators (for $n = 10$, $CR \approx 72\%$, for $n = 20$, $CR \approx 79\%$, and for $n = 40$, $CR \approx 83\%$).

Table 4. Results of simulations for the second artificial population

		$n = 10\ (10^8)$			$n = 20\ (10^7)$			$n = 40\ (10^6)$	
$\mathrm{var}_{\mathrm{Hajek}_1}$		1.575			1.966			2.458	
$\mathrm{var}_{\mathrm{Fix}}$		1.559			1.948			2.434	
$\mathrm{var}_{\mathrm{H\text{-}Rao}_1}$		1.559			1.948			2.438	
$\mathrm{var}_{\mathrm{H\text{-}Rao}_2}$		1.559			1.948			2.437	
$\mathrm{var}_{\mathrm{Hajek}_2}$		1.559			1.947			2.433	
$\mathrm{var}_{\mathrm{repl}}$		1.579			1.978			2.490	
$\mathrm{var}_{\mathrm{naive}}$		1.421			1.583			1.494	
True value		1.559			1.948			2.434	
	RB (%)	MSE (10^{19})	CR (%)	RB (%)	MSE (10^{16})	CR (%)	RB (%)	MSE (10^{14})	CR (%)
$\widehat{\mathrm{var}}_{\mathrm{HT}}$	0.900	1.075	36.13	0.105	6.844	36.98	0.299	5.623	38.73
$\widehat{\mathrm{var}}_{\mathrm{SYG}}$	0.995	1.078	36.12	0.098	6.829	36.87	0.260	5.584	38.60
$\widehat{\mathrm{var}}_{\mathrm{Dev1}}$	0.919	1.043	36.06	−0.028	6.603	36.84	0.007	5.314	38.47
$\widehat{\mathrm{var}}_{\mathrm{Dev2}}$	0.921	1.044	36.06	−0.020	6.618	36.84	0.055	5.364	38.52
$\widehat{\mathrm{var}}_{\mathrm{repl}}$	0.476	0.874	35.65	−0.241	6.229	37.35	0.167	5.443	40.34
$\widehat{\mathrm{var}}_{\mathrm{naive}}$	0.476	0.874	35.65	1.705	4.418	35.97	−6.356	2.070	35.46
$\widehat{\mathrm{var}}_{\mathrm{Fix}}$	0.989	1.074	36.12	0.105	6.844	36.97	0.300	5.624	38.65
$\widehat{\mathrm{var}}_{\mathrm{R}}$	0.921	1.044	36.06	−0.019	6.619	36.85	0.052	5.359	38.56
$\widehat{\mathrm{var}}_{\mathrm{Dev3}}$	0.930	1.048	36.08	0.015	6.679	36.89	0.221	5.532	38.91
$\widehat{\mathrm{var}}_{1}$	0.878	1.026	36.04	−0.096	6.486	36.77	−0.088	5.219	38.43
$\widehat{\mathrm{var}}_{\mathrm{Ber}}$	0.993	1.077	36.12	0.091	6.817	36.86	0.215	5.537	38.53
$\widehat{\mathrm{var}}_{\mathrm{T}}$	0.994	1.078	36.12	0.100	6.831	36.91	0.282	5.598	38.91
$\widehat{\mathrm{var}}_{2}$	0.929	1.048	36.07	0.014	6.676	36.89	0.216	5.527	38.90
$\widehat{\mathrm{var}}_{3}$	0.920	1.044	36.06	−0.022	6.615	36.84	0.051	5.359	38.52
$\widehat{\mathrm{var}}_{4}$	0.905	1.027	36.03	−0.055	6.498	36.78	−0.029	5.230	38.44

Table 4. Continued

	$n = 10\ (10^8)$			$n = 20\ (10^7)$			$n = 40\ (10^6)$		
$\widehat{\text{var}}_5$	0.862	1.019	35.97	−0.122	6.442	36.75	−0.124	5.184	38.43
$\widehat{\text{var}}_{\text{Br1}}$	0.869	1.022	35.99	−0.089	6.497	36.81	0.036	5.339	38.70
$\widehat{\text{var}}_{\text{Br2}}$	0.927	1.047	36.06	0.007	6.665	36.89	0.173	5.481	38.84
$\widehat{\text{var}}_{\text{Br3}}$	0.984	1.072	36.13	0.101	6.835	36.99	0.306	5.625	39.01
$\widehat{\text{var}}_{\text{Br4}}$	0.990	1.075	36.13	0.106	6.844	36.99	0.310	5.629	39.01

Table 5. Number of times that $\hat{\text{var}}_{HT} < 0$ *among 10,000 simulated samples*

	$n = 10$	$n = 20$	$n = 40$
mu284 population	2312	2708	2450
artificial population 1	61	65	278
artificial population 2	0	0	0

7. Conclusions

Using the method of Chen et al. (1994) and Deville (2000), the joint inclusion probabilities can be computed exactly for a maximum entropy sampling design with fixed sample size and unequal probabilities. The joint inclusion probabilities are used in the formulae of two variance estimators, the Horvitz-Thompson and the Sen-Yates-Grundy. An empirical study demonstrates that inferiority of $\widehat{\text{var}}_{HT}$ is restricted to populations having high correlation between the variable of interest y_k and the auxiliary variable x_k, and where $\widehat{\text{var}}_{HT} < 0$. Except in the case of these populations, $\widehat{\text{var}}_{HT}$ performs nearly the same as $\widehat{\text{var}}_{SYG}$. In the same case, these two estimators have a comportment similar to those of the estimators which use only the first-order inclusion probabilities (except $\widehat{\text{var}}_{repl}$ and $\widehat{\text{var}}_{naive}$). Under simulations, the estimators which use only the first-order inclusion probabilities (different from $\widehat{\text{var}}_{repl}$ and $\widehat{\text{var}}_{naive}$ which overestimate the variance) have similar performances, regardless of correlation between y_k and x_k. The use of first-order inclusion probabilities over the whole population and joint inclusion probabilities does not lead to more accurate variance estimators in the case of a maximum entropy sampling design with fixed sample size and unequal probabilities. So we recommend the use of a simple estimator such as Deville Estimator 1, and, in the approximation class, the fixed-point approximation.

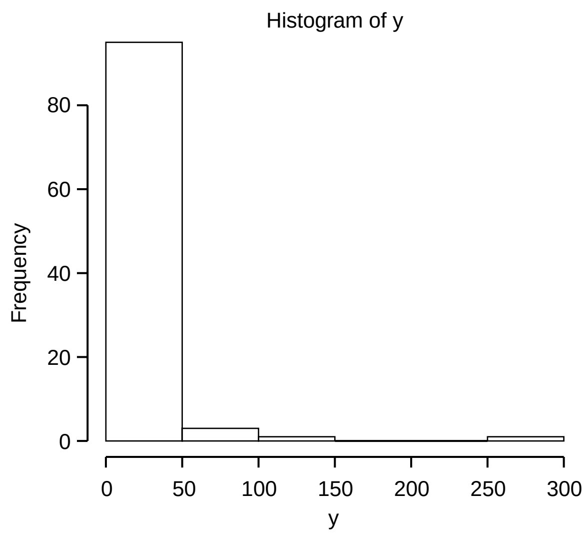

Fig. 3. The second artificial population, n = 10

Appendix 1

Proof of Result 1

If we note $C(\boldsymbol{\lambda}, S_n(U)) = \sum_{s \in S_n(U)} \exp \boldsymbol{\lambda}' \mathrm{s}$, then

$$\begin{aligned}
\pi_k(\boldsymbol{\lambda}, n) &= \frac{\sum_{s \in S_n(U)} s_k \exp \boldsymbol{\lambda}' \mathrm{s}}{C(\boldsymbol{\lambda}, S_n(U))} \\
&= \frac{\exp \lambda_k}{C(\boldsymbol{\lambda}, S_n(U))} \sum_{s \in S_{n-1}(U \setminus \{k\})} \exp \lambda' \mathrm{s} \\
&= \frac{\exp \lambda_k}{C(\boldsymbol{\lambda}, S_n(U))} \left(\sum_{s \in S_{n-1}(U)} \exp \boldsymbol{\lambda}' \mathrm{s} - \sum_{s \in S_{n-1}(U)} s_k \exp \boldsymbol{\lambda}' \mathrm{s} \right) \\
&= \frac{\exp \lambda_k C(\boldsymbol{\lambda}, S_{n-1}(U))}{C(\boldsymbol{\lambda}, S_n(U))} (1 - \pi_k(\boldsymbol{\lambda}, n-1))
\end{aligned}$$

Since $\sum_{k \in U} \pi_k(\boldsymbol{\lambda}, n) = n$, we finally get

$$\pi_k(\boldsymbol{\lambda}, n) = n \frac{\exp \lambda_k \{1 - \pi_k(\boldsymbol{\lambda}, n-1)\}}{\sum_{\ell \in U} \exp \lambda_\ell \{1 - \pi_\ell(\boldsymbol{\lambda}, n-1)\}}$$

Appendix 2

Proof of Result 2

If we note $C(\boldsymbol{\lambda}, S_n(U)) = \sum_{s \in S_n(U)} \exp \boldsymbol{\lambda}' \mathrm{s}$, then

$$\begin{aligned}
\pi_{k\ell}(\boldsymbol{\lambda}, n) &= \sum_{s \in S_n(U)} s_k s_\ell p(s) \\
&= \frac{\sum_{s \in S_n(U)} s_k s_\ell \exp \boldsymbol{\lambda}` \mathrm{s}}{C(\boldsymbol{\lambda}, S_n(U))} \\
&= \sum_{\substack{s \in S_n(U) \\ k, \ell \in s}} \frac{\prod_{j \in S} \exp \lambda_j}{C(\boldsymbol{\lambda}, S_n(U))} \\
&= \frac{\exp \lambda_k \exp \lambda_\ell \sum_{\substack{s \in S_{n-2}(U) \\ k, \ell \notin s}} \prod_{j \in s} \exp \lambda_j}{C(\boldsymbol{\lambda}, S_n(U))} \\
&= \exp \lambda_k \exp \lambda_\ell \Pr \{k, \ell \notin s | s \in S_{n-2}\} \frac{C(\boldsymbol{\lambda}, S_{n-2}(U))}{C(\boldsymbol{\lambda}, S_n(U))} \\
&= \exp \lambda_k \exp \lambda_\ell (1 - \pi_k(\boldsymbol{\lambda}, n-2) - \pi_\ell(\boldsymbol{\lambda}, n-2) \\
&\quad + \pi_{k\ell}(\boldsymbol{\lambda}, n-2)) \frac{C(\boldsymbol{\lambda}, S_{n-2}(U))}{C(\boldsymbol{\lambda}, S_n(U))}
\end{aligned}$$

Since $\sum_{k\in U}\sum_{\substack{\ell\in U\\ \ell\neq k}}\pi_{k\ell}(\boldsymbol{\lambda},n) = n(n-1)$, we finally get

$$\pi_{k\ell}(\boldsymbol{\lambda},n)=\frac{n(n-1)\ \exp\lambda_k\ \exp\lambda_\ell(1-\pi_k(\boldsymbol{\lambda},n-2)-\pi_\ell(\boldsymbol{\lambda},n-2)+\pi_{k\ell}(\boldsymbol{\lambda},n-2))}{\sum_{i\in U}\sum_{\substack{j\in U\\ j\neq i}}\exp\lambda_i\exp\lambda_j(1-\pi_i(\boldsymbol{\lambda},n-2)-\pi_j(\boldsymbol{\lambda},n-2)+\pi_{ij}(\boldsymbol{\lambda},n-2))}$$

$k,\ell\in U, k\neq\ell$

Appendix 3

Justification of Algorithm 1

Suppose that $\sum_{k\in U}\lambda_k = 0$, in order to have a unique definition of $\boldsymbol{\lambda}$. Indeed,

$$p(s, S_n(U), \boldsymbol{\lambda}) = p(s, S_n(U), \boldsymbol{\lambda}^*), \text{for all } s \in S_n(U)$$

when $\lambda_k^* = \lambda_k + c$ for any $c \in \mathbb{R}$. The inclusion probability vector can be written as a function of $\boldsymbol{\lambda}$ and n:

$$\boldsymbol{\pi}(\boldsymbol{\lambda}, n) = \sum_{s\in S_n(U)} p(s, S_n(U), \boldsymbol{\lambda})$$

Since $\sum_{k\in U}\pi_k = n$, $\boldsymbol{\pi}(\boldsymbol{\lambda}, n)$ is a one to one application from

$$\left\{\boldsymbol{\lambda}\in\mathbb{R}^N \middle| \sum_{k\in U}\lambda_k = 0\right\}$$

to

$$\left\{\boldsymbol{\pi}\in]0,1[^N \middle| \sum_{k\in U}\pi_k = n\right\}$$

Define $\boldsymbol{\pi}(\boldsymbol{\lambda}, n)$ as a function of $\tilde{\boldsymbol{\pi}}$, that will be denoted $\boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n)$, and

$$\boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n) = \boldsymbol{\pi}(\boldsymbol{\lambda}, n) = \frac{\sum_{s\in S_n(U)}\mathbf{s}\exp\boldsymbol{\lambda}'\mathbf{s}}{\sum_{s\in S_n(U)}\exp\boldsymbol{\lambda}'\mathbf{s}} = \frac{\sum_{s\in S_n(U)}\mathbf{s}\prod_{k\in s}\frac{\tilde{\pi}_k}{1-\tilde{\pi}_k}}{\sum_{s\in S_n(U)}\prod_{k\in s}\frac{\tilde{\pi}_k}{1-\tilde{\pi}_k}}$$

Since $\tilde{\boldsymbol{\pi}}$ can be derived from $\boldsymbol{\lambda}$ and vice versa (see Result 1), $\boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n)$ can be computed recursively by means of Expression (3)

$$\boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n) = n\frac{\frac{\tilde{\pi}_k}{1-\tilde{\pi}_k}\{1-\phi_k(\tilde{\boldsymbol{\pi}}, n-1)\}}{\sum_{\ell\in U}\frac{\tilde{\pi}_\ell}{1-\tilde{\pi}_\ell}\{1-\phi_\ell(\tilde{\boldsymbol{\pi}}, n-1)\}}$$

If the vector of inclusion probabilities $\boldsymbol{\pi}$ (such that $\sum_{k\in U}\pi_k = n$) is given, Chen et al. (1994) have proposed solving the equation

$$\boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n) = \boldsymbol{\pi}$$

in $\tilde{\boldsymbol{\pi}}$ by the Newton method, which gives the algorithm

$$\tilde{\boldsymbol{\pi}}^{(i)} = \tilde{\boldsymbol{\pi}}^{(i-1)} + \left|\frac{\partial \boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n)}{\partial \tilde{\boldsymbol{\pi}}}\right|^{-1}_{\tilde{\boldsymbol{\pi}}=\tilde{\boldsymbol{\pi}}^{(i-1)}} (\boldsymbol{\pi} - \boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}^{(i-1)}, n))$$

where $i = 1, 2, \ldots$ and with $\tilde{\boldsymbol{\pi}}^{(0)} = \boldsymbol{\pi}$. Unfortunately, the matrix

$$\left|\frac{\partial \boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}, n)}{\partial \tilde{\boldsymbol{\pi}}}\right|_{\tilde{\boldsymbol{\pi}}=\tilde{\boldsymbol{\pi}}^{(i-1)}} \tag{31}$$

is not easy to compute. However, Deville (2000) pointed out that Matrix (31) is very close to the identity matrix, which allows significantly simplifying the algorithm. Finally we can use

$$\tilde{\boldsymbol{\pi}}^{(i)} = \tilde{\boldsymbol{\pi}}^{(i-1)} + \boldsymbol{\pi} - \boldsymbol{\phi}(\tilde{\boldsymbol{\pi}}^{(i-1)}, n) \tag{32}$$

which makes is possible to pass quite quickly from $\boldsymbol{\pi}$ to $\tilde{\boldsymbol{\pi}}$ and thus to $\boldsymbol{\lambda}$. The number of operations needed to compute $\tilde{\boldsymbol{\pi}}$ is $O(N^2 \times n \ \times$ number of iterations i).

8. References

Aires, N. (1999). Algorithms to Find Exact Inclusion Probabilities for Conditional Poisson Sampling and Pareto πps Sampling Designs. Methodology and Computing in Applied Probability, 4, 457–469.

Ardilly, P. (1994). Les Techniques de Sondage. Paris: Technip. [In French]

Berger, Y. (1998a). Rate of Convergence for Asymptotic Variance for the Horvitz-Thompson Estimator. Journal of Statistical Planning and Inference, 74, 149–168.

Berger, Y. (1998b). Variance Estimation Using List Sequential Scheme for Unequal Probability Sampling. Journal of Official Statistics, 14, 315–323.

Brewer, K. (2002). Combined Survey Sampling Inference, Weighing Basu's Elephants. London: Arnold.

Brewer, K. and Donadio, M. (2003). The High Entropy Variance of the Horvitz-Thompson Estimator. Survey Methodology, 29, 189–196.

Brewer, K. and Hanif, M. (1983). Sampling with Unequal Probabilities, Lecture Notes in Statistics, Volume 15. New York: Springer Verlag.

Brown, L.D. (1986). Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory. Institute of Mathematical Statistics, Lecture Notes, Monograph Series, Volume 9, Hayward, CA.

Chen, X.-H., Dempster, A., and Liu, J. (1994). Weighted Finite Population Sampling to Maximize Entropy. Biometrika, 81, 457–469.

Cumberland, W. and Royall, R. (1981). Prediction Models in Unequal Probability Sampling. Journal of the Royal Statistical Society, Series B, 43, 353–367.

Deville, J.-C. (1993). Estimation de la Variance Pour les Enquêtes en Deux Phases. Note Interne Manuscrite. France: INSEE. [In French]

Deville, J.-C. (1999). Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques. Techniques d'enquête, Survey Methodology, 25, 193–204.

Deville, J.-C. (2000). Note Sur l'algorithme de Chen, Dempster et Liu. Technical Report, France CREST-ENSAI. [In French]

Deville, J.-C. and Tillé, Y. (2005). Variance Approximation Under Balanced Sampling. Journal of Statistical Planning and Inference, 128, 411–425.

Hájek, J. (1964). Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population. Annals of Mathematical Statistics, 35, 1491–1523.

Hájek, J. (1981). Sampling from a Finite Population. New York: Marcel Dekker.

Hartley, H. and Rao, J. (1962). Sampling with Unequal Probabilities and Without Replacement. Annals of Mathematical Statistics, 33, 350–374.

Horvitz, D. and Thompson, D. (1952). A Generalization of Sampling without Replacement from a Finite Universe. Journal of the American Statistical Association, 47, 663–685.

Rosén, B. (1991). Variance Estimation for Systematic pps-sampling. Report 1991:15, Statistics Sweden.

Rosén, B. (1997). On Sampling with Probability Proportional to Size. Journal of Statistical Planning and Inference, 62, 159–191.

Sen, A. (1953). On the Estimate of the Variance in Sampling with Varying Probabilities. Journal of Indian Society for Agricultural Statistics, 5, 119–127.

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

Stehman, V. and Overton, W.S. (1994). Comparison of Variance Estimators of the Horvitz-Thompson Estimator for Randomized Variable Probability Systematic Sampling. Journal of the American Statistical Association, 89, 30–43.

Tillé, Y. (1996). Some Remarks on Unequal Probability Sampling Designs Without Replacement. Annales d'Economie et de Statistique, 44, 177–189.

Tillé, Y. (2001). Théorie des sondages: echantillonnage et estimation en populations finies. Paris: Dunod. [In French]

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, 15, 235–261.

Received June 2003
Revised August 2004