



Wojciech Gamrot 

University of Economics in Katowice, College of Management, Department of Statistics,  
Econometrics and Mathematics, wojciech.gamrot@ue.katowice.pl

## A Stopping Rule for Simulation-Based Estimation of Inclusion Probabilities

**Abstract:** Design-based estimation of finite population parameters such as totals usually relies on the knowledge of inclusion probabilities characterising the sampling design. They are directly incorporated into sampling weights and estimators. However, for some useful sampling designs, these probabilities may remain unknown. In such a case, they may often be estimated in a simulation experiment which is carried out by repeatedly generating samples using the same sampling scheme and counting occurrences of individual units. By replacing unknown inclusion probabilities with such estimates, design-based population total estimates may be computed. The calculation of required sample replication numbers remains an important challenge in such an approach. In this paper, a new procedure is proposed that might lead to the reduction in computational complexity of simulations.

**Keywords:** Horvitz-Thompson estimator, inclusion probabilities, simulation, precision

**JEL:** C83, C63

## 1. Introduction

Following Särndal, Swensson and Wretman (1992: 5), we shall represent a finite population as a set of unit indices  $U = \{1, \dots, N\}$ . Values of a fixed characteristic for corresponding population units are represented by a vector  $\mathbf{y} = [y_1, \dots, y_N]'$ . The parameter under study is the population total (Hedayat, Sinha, 1991: 2):

$$t = \sum_{i \in U} y_i = \mathbf{1}' \mathbf{y}. \quad (1)$$

The unordered sample space may be represented by a matrix:

$$\mathbf{A} = [\mathbf{a}_{ij}] = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \quad (2)$$

whose each  $i$ -th row:

$$\mathbf{a}_i = [a_{i1}, \dots, a_{iN}] \quad (3)$$

represents one possible sample with  $a_{ij} = 1$  when this sample contains the  $j$ -th unit and  $a_{ij} = 0$  otherwise. The matrix  $\mathbf{A}$  has  $N$  columns and  $Z = 2^N$  rows representing all possible sequences of zeros and ones of the length  $N$ , including an empty sample represented by a sequence of  $N$  zeros and a *census* represented by  $N$  ones. A vector of corresponding sample sizes may be calculated as:

$$\mathbf{n} = [n_1, \dots, n_Z]' = \mathbf{A} \mathbf{1}. \quad (4)$$

Let an unordered sample:  $s \subseteq U$  be drawn from  $U$ . The sample composition may be characterised by a vector of sample membership indicators (Tillé, 2006: 8):

$$\mathbf{I}(s) = [I_1(s), \dots, I_N(s)], \quad (5)$$

where

$$I_i(s) = \begin{cases} 0 & \text{for } i \in s \\ 1 & \text{for } i \notin s \end{cases} \quad (6)$$

The sampling is equivalent to choosing a certain (say  $i$ -th) row of  $\mathbf{A}$  so that  $\mathbf{I}(s) = \mathbf{a}_i$ . It may be done according to a sampling design:

$$\mathbf{P} = [P_1, \dots, P_Z]', \quad (7)$$

which associates a selection probability  $P_i \in [0,1]$  with every row  $\mathbf{a}_i$  for  $i \in U$ , so that  $\sum P_i = 1$ . The expectation of the vector  $\mathbf{I}$  may be expressed as:

$$\boldsymbol{\delta} = [\pi_1, \dots, \pi_N]^\top = \mathbf{P}\mathbf{A}. \quad (8)$$

Elements of the vector  $\boldsymbol{\pi}$  are called *first-order inclusion probabilities* because  $\pi_i = Pr\{i \in s\}$  for  $i \in U$ . Let us also introduce a vector of corresponding weights:

$$\mathbf{d} = [\pi_1^{-1}, \dots, \pi_N^{-1}]^\top. \quad (9)$$

One may also define a matrix of second-order inclusion probabilities (Tillé, 2006: 17) as:

$$\Pi = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1N} \\ \vdots & \ddots & \vdots \\ \pi_{N1} & \cdots & \pi_{NN} \end{bmatrix} = \mathbf{A}'\text{diag}(\mathbf{P})\mathbf{A}, \quad (10)$$

where  $\pi_{ij} = Pr\{i, j \in s\}$ . This lets us express the covariance matrix of the vector  $\mathbf{I}$  as:

$$\mathbf{C} = \Pi - \boldsymbol{\pi}\boldsymbol{\pi}'. \quad (11)$$

The size of the sample  $s$  may be expressed as:

$$\mathbf{n}(s) = \mathbf{l}'\mathbf{I}(s). \quad (12)$$

Denote sampled elements as:

$$s = \{i_1, \dots, i_{n(s)}\}. \quad (13)$$

For any vector  $\mathbf{u} = [u_1, \dots, u_N]$ , let:

$$\mathbf{u}(s) = [u_{i_1}, \dots, u_{i_{n(s)}}]. \quad (14)$$

This lets us define sample vectors:  $\mathbf{y}(s)$ ,  $\boldsymbol{\pi}(s)$ ,  $\mathbf{d}(s)$  which are obtained by omitting elements corresponding to zeros in  $\mathbf{I}(s)$  respectively in  $\mathbf{y}$ ,  $\boldsymbol{\pi}$ ,  $\mathbf{d}$ . For known  $\boldsymbol{\pi}$ , the design-unbiased Horvitz-Thompson (HT) estimator of  $t$  may be expressed in the form (cf. Narain, 1951; Horvitz, Thompson, 1952):

$$\hat{t}(s) = \mathbf{d}'\text{diag}(\mathbf{I}(s))\mathbf{y} \quad (15)$$

or equivalently:

$$\hat{t}(s) = \mathbf{d}(s)\mathbf{y}(s). \quad (16)$$

## 2. Simulation-based estimation

To calculate the HT estimator, first order inclusion probabilities are needed. However, many sampling procedures are too complicated to calculate them. In particular, this is true for spatial sampling (Barabesi, Fattorini, Ridolfi, 1997; Fattorini, Ridolfi, 1997), order sampling schemes, especially the Pareto scheme (Rosén, 1997), rejective sampling (Wywiał, 2003; Boistard, Lopuhaä, Ruiz-Gazen, 2012; Yu, 2012), and sequential sum-quota sampling schemes (Pathak, 1976; Kremers, 1985). A particular example is the greedy sampling scheme (Gamrot, 2014: 223) where costs of sampling individual units vary but are known in advance, and the survey budget is restricted. Individual units are drawn to the sample sequentially, one-by-one, with equal probabilities, from a gradually shrinking pool of still-affordable units. In the most pessimistic case, the calculation of inclusion probabilities would require analysing all permutations of units, which is unfeasible. If inclusion probabilities do not depend on sample observations, then Fattorini (2006; 2009) proposes to perform a simulation experiment. It is carried out by generating a large number  $R$  of sample replications  $\check{s}_1, \dots, \check{s}_R$ . Empirical counts of unit occurrences are then calculated as:

$$\mathbf{m} = [m_1, \dots, m_N]' = \sum_{i \in \{1 \dots R\}} \mathbf{I}(\check{s}_i). \quad (17)$$

This enables the calculation of empirical inclusion probabilities:

$$\hat{\pi} = [\hat{\pi}_1, \dots, \hat{\pi}_N]' = \frac{\mathbf{m}}{R} \quad (18)$$

and empirical weights:

$$\hat{\mathbf{d}} = [\hat{\pi}_1^{-1}, \dots, \hat{\pi}_N^{-1}]'. \quad (19)$$

By omitting elements corresponding to non-sampled units respectively in  $m, \hat{\pi}, \hat{\mathbf{d}}$  one may then obtain empirical quantities  $m(s), \hat{\pi}(s), \hat{\mathbf{d}}(s)$  associated with the realised sample  $s$ . This leads to the calculation of the empirical HT estimator in the form:

$$\hat{\hat{t}}(s) = \hat{\mathbf{d}} \text{diag}(\mathbf{I}(s)) \mathbf{y} \quad (20)$$

or equivalently:

$$\hat{\hat{t}}(s) = \hat{\mathbf{d}}(s) \mathbf{y}(s). \quad (21)$$

### 3. Setting up the stopping rule

In order to establish a sufficient value of replication number  $R$  that guarantees a required precision of simulation-based estimates, Fattorini (2006; 2009) proposes the accuracy criterion:

$$Q(R) = \Pr \left\{ \left| \hat{t}(s) - \hat{t}(s) \right| < \varepsilon \hat{t}(s) \right\} \quad (22)$$

and finds its upper bound on the basis of Bennet's inequality. On this basis, he proposes a formula for the sufficient value of  $R$ . Later, Gamrot (2013) attempted to improve over that using asymptotic approximations based on a normal distribution, Chernoff-Hoeffding inequality, and pre-calculated tables of exact probabilities for the restricted maximum likelihood estimator. However, the relative deviation of the empirical HT estimator  $\hat{t}(s)$  from its "true" value  $\hat{t}(s)$  that would be calculated for known inclusion probabilities has a complex distribution. The construction of an upper bound for it requires the pessimistic assumption of possible high correlation among sample membership indicators. This leads to very conservative replication numbers, which results in long calculation time.

In what follows, it is demonstrated that these pessimistic assumptions are often overly conservative, and may be improved upon. The value of the empirical HT estimator depends on the count vector  $m$ . Let  $\Omega$  be a set of such values of this vector for which the condition

$$\left| \frac{\hat{t}(s) - \hat{t}(s)}{\hat{t}(s)} \right| < \varepsilon \quad (23)$$

is satisfied so that:

$$Q(R) = \Pr \{ m \in \dot{\cup} \} . \quad (24)$$

Hence, instead of examining the scalar distribution of  $\hat{t}(s)$  one may investigate a much simpler, multivariate distribution of  $m$ . As an introductory example, let us consider a population of size  $N = 3$ , with the sample space and sampling design:

$$A_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad P_0 = \begin{bmatrix} p_{000} \\ p_{001} \\ p_{010} \\ p_{011} \\ p_{100} \\ p_{101} \\ p_{110} \\ p_{111} \end{bmatrix}. \quad (25)$$

When a sample  $s$  corresponding to the sample indicator vector  $I(s) = [0, 1, 1]$  is drawn, all the columns of the matrix  $A$  which contain zeros and correspond to non-sampled units may be disregarded in our analysis because the HT estimator does not depend on these units and corresponding inclusion probabilities. Hence, it is sufficient to consider a reduced sample space and sampling design:

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \quad P = \begin{bmatrix} p_{00} \\ p_{01} \\ p_{10} \\ p_{11} \end{bmatrix} = \begin{bmatrix} p_{000} + p_{100} \\ p_{001} + p_{101} \\ p_{010} + p_{110} \\ p_{011} + p_{111} \end{bmatrix} \quad (26)$$

with the reduced sample indicator vector  $I(s) = [1, 1]$ . Such reduction may be carried out for populations and samples of any size. However, in a bivariate case, the distribution of the vector  $\mathbf{m} = [m_1, m_2]'$  for  $R$  sample replications is particularly simple and takes the form (Beyer, 1987: 532):

$$Pr(m_1, m_2) = \sum_a D(a) p_{11}^a p_{10}^{(m_1-a)} p_{01}^{(m_2-a)} p_{00}^{R-m_1-m_2+a}, \quad (27)$$

where:

$$D(a) = \frac{R!}{a!(m_1-a)!(m_2-a)!(R-m_1-m_2+a)!}. \quad (28)$$

In the following examples, we will now discuss in more detail some interesting special cases of such a bivariate distribution.

**Example 1.** Let the sample  $s$  of size 2 corresponding to the indicator vector  $I(s) = [1, 1]$  be drawn from the population according to the sampling design  $P = [0.1, 0.3, 0.5, 0.1]'$ . Let us consider four possible sample outcomes:  $\mathbf{y} = [1, 4]', \mathbf{y} = [1, 2]', \mathbf{y} = [1, 1]'$  and  $\mathbf{y} = [2, 1]'$ . Assume that  $\varepsilon = 0.1$  and  $R = 100$ . Distributions of the count vector  $\mathbf{m}$  for all four designs along with resulting values of the accuracy criterion  $Q(R)$  are shown in Figure 1. It is clearly visible that the region  $\Omega$  shifts when

values of the study variable change. This results in higher or lower probabilities  $Q(R)$ .

The dependency of  $Q(R)$  on sample observations of the study variable are not the only important effect. The following example illustrates another one:

**Example 2.** Let us consider four sampling designs  $P_1, P_2, P_3, P_4$  given by Table 1, together with corresponding values of the correlation coefficient  $\rho$  between the two sample membership indicators corresponding to the first and second unit. Let the sample  $s$  of size 2 corresponding to the sample membership indicator vector  $I(s) = [1, 1]$  be drawn from the population and let the following values of the study variable be observed:  $y = [1, 1]$ . Assume that  $\varepsilon = 0.1$  and  $R = 100$ . Distributions of the count vector  $m$  for all four designs along with resulting values of the accuracy criterion  $Q(R)$  are shown in Figure 2.

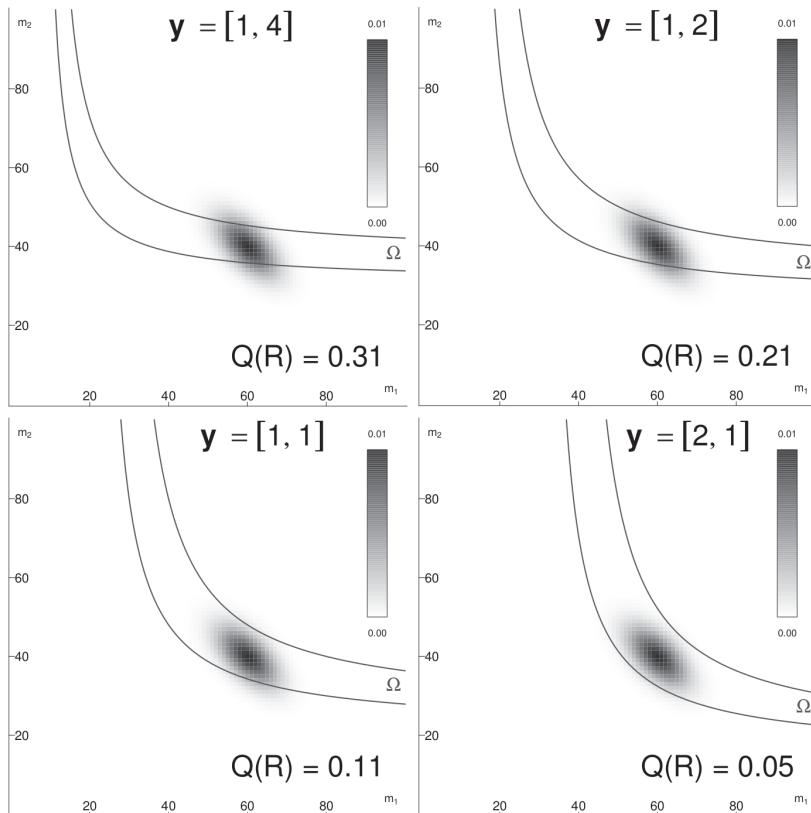
Table 1. Distributions of the count vector and correlation coefficients of sample membership indicators for certain sampling designs

	$P_1$	$P_2$	$P_3$	$P_4$
$p_{00}$	0.49	0.35	0.25	0.10
$p_{01}$	0.01	0.15	0.25	0.40
$p_{10}$	0.01	0.15	0.25	0.40
$p_{11}$	0.49	0.35	0.25	0.10
$\rho$	0.96	0.40	0.00	-0.6

Source: own elaboration

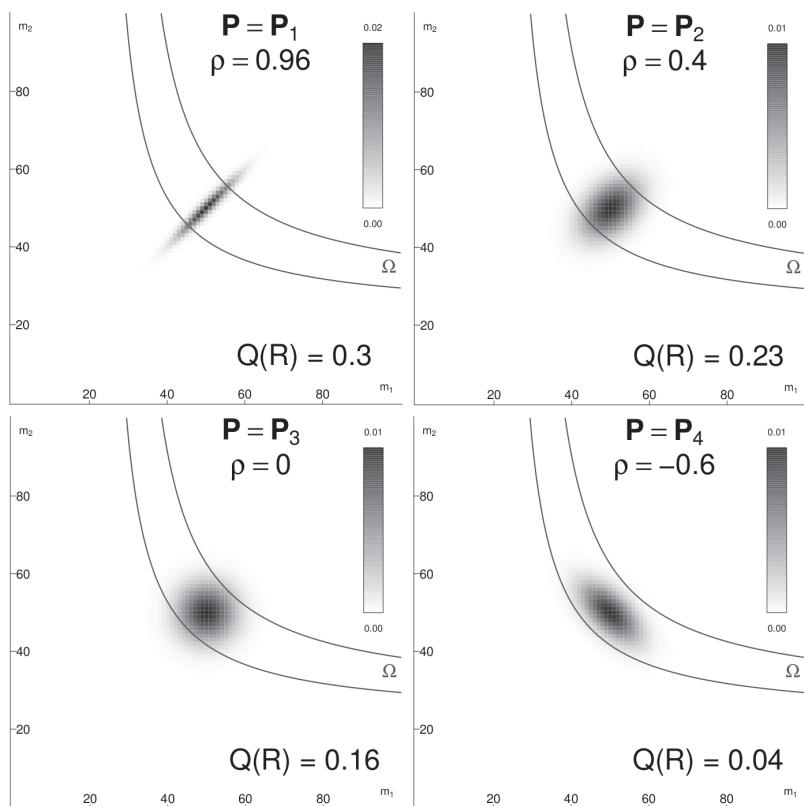
Despite its simplicity, the example shows that the probability  $Q(R)$  depends on the correlation  $\rho$  between sample membership indicators. It is high for the least favourable case extreme of positive correlation that is tacitly assumed in the derivation of known stopping rules. In reality, however, it is often much lower. Taking this effect into account, one might construct tighter bounds for  $Q(R)$  and obtain a stopping rule which gives lower required replication numbers.

This effect remains in force when more than two elements are drawn to the sample, as shown in the last example.

Figure 1. Location of the  $\Omega$  region for various sample observations of the study variable

Source: own elaboration

**Example 3.** Consider the population of size  $N = 3$  and the general sampling design  $\mathbf{P} = [b, a, a, a, a, a, a, b]'$ , where  $b = (1 - 6a)/2$ , that is parameterised by the constant  $a \in (0, 1/6)$ . The special feature of this design is that the correlation coefficient between all sample membership indicators takes the same value (say  $\rho$ ). Probabilities of drawing all the possible samples obtained for varying values of  $a$  and resulting  $\rho$ -values are shown in Table 2. Let us assume that all the three population elements are drawn to a sample resulting in the sample observation of the study variable:  $y = [1, 2, 3]$ . Let us also assume that  $\varepsilon = 0.15$  and  $R = 50$ . The empirical distribution of 5000 realisations of the count vector  $\mathbf{m}$ , along with boundaries of the  $\Omega$  region for  $a = 0.01, 0.05, 0.125, 0.166$ , are shown in Figure 3. It is clearly visible that the negative correlation among sample membership indicators – and even lack thereof – improves the accuracy criterion as compared with the worst case of perfect positive correlation.

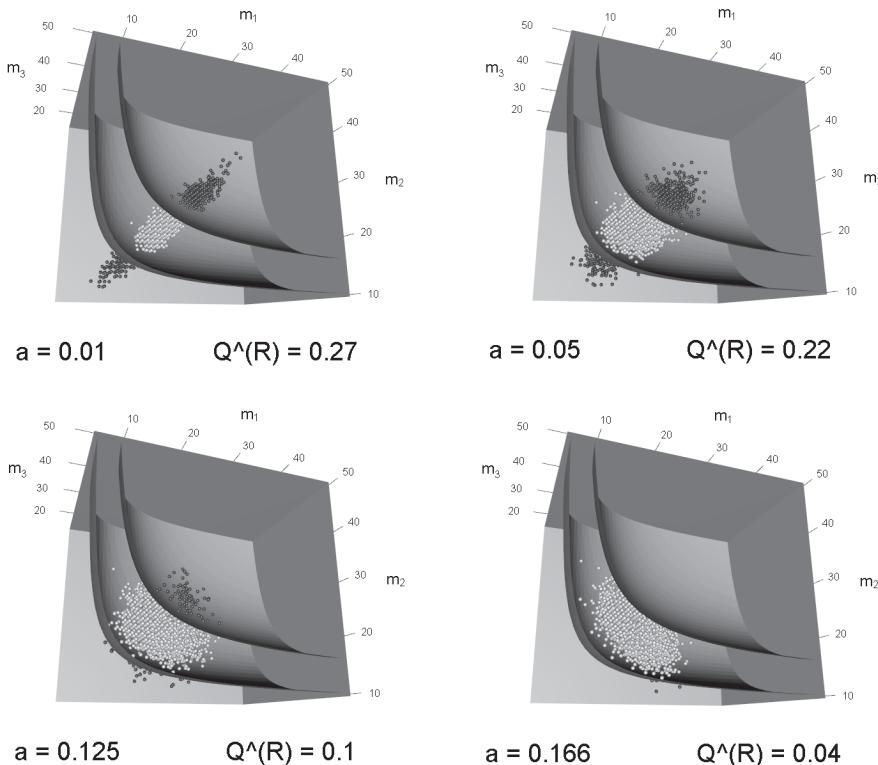
Figure 2. Distribution of the simulation count vector  $m$  for various sampling designs and  $n(s) = 2$ 

Source: own elaboration

Table 2. Sample drawing probabilities and correlation coefficients between sample membership indicators for various values of the parameter  $a$ 

<b><math>a</math></b>	<b>0.01</b>	<b>0.05</b>	<b>0.125</b>	<b>0.166</b>
$p_{000}$	0.47	0.35	0.125	0.002
$p_{001}$	0.01	0.05	0.125	0.166
$p_{010}$	0.01	0.05	0.125	0.166
$p_{011}$	0.01	0.05	0.125	0.166
$p_{100}$	0.01	0.05	0.125	0.166
$p_{101}$	0.01	0.05	0.125	0.166
$p_{110}$	0.01	0.05	0.125	0.166
$p_{111}$	0.47	0.35	0.125	0.002
$\rho$	0.92	0.6	0.00	-0.328

Source: own elaboration

Figure 3. Distribution of the simulation count vector  $\mathbf{m}$  for various sampling designs and  $n(s) = 3$ 

Source: own elaboration

## 4. The proposed stopping rule

The accuracy criterion depends on correlations among sample membership indicators. However, it is not reasonable to expect these correlations to be known when inclusion probabilities (defined as moments of their distributions) remain unknown. To account for correlations, the simulation may be divided into two phases. In the first phase,  $R_1$  sample replications  $\check{s}_1, \dots, \check{s}_{R_1}$  are generated. Occurrences of individual units and pairs of units are counted resulting in the following counts:

$$\mathbf{m}_1 = \sum_{i \in \{1, \dots, R_1\}} \mathbf{I}(\check{s}_i). \quad (29)$$

$$\mathbf{M}_1 = \begin{bmatrix} m_{11} & \cdots & m_{1N} \\ \vdots & \ddots & \vdots \\ m_{N1} & \cdots & m_{NN} \end{bmatrix}' = \sum_{i \in \{1, \dots, R_1\}} \mathbf{I}\left(\check{s}_i\right) \mathbf{I}\left(\check{s}_i\right)'. \quad (30)$$

Then the estimates of first and second order inclusion probabilities are obtained as:

$$\hat{\pi}_1 = \frac{\mathbf{m}_1}{R_1}, \quad (31)$$

$$\hat{\Pi}_1 = \frac{\mathbf{M}_1}{R_1} \quad (32)$$

This enables estimation of the covariance matrix  $\mathbf{C}$  which characterises a joint distribution of sample membership indicators as:

$$\hat{\mathbf{C}} = \hat{\Pi}_1 - \hat{\pi}_1 \hat{\pi}_1' \quad (33)$$

while  $\hat{\pi}_1$  is obviously an estimate of its expectation vector  $\boldsymbol{\pi}$ . In the second phase of simulation,  $R_2$  sample replications  $\check{s}_{R_1+1}, \dots, \check{s}_{R_1+R_2}$  are generated. Hence, in the whole simulation, a total of  $R = R_1 + R_2$  sample replications  $\check{s}_1, \dots, \check{s}_R$  are generated, which leads to the calculation of the final count vector  $\mathbf{m}$  and the empirical HT estimator according to expressions (17)–(21). Capabilities of contemporary computers make it possible to set  $R_1$ ,  $R_2$  and  $R$  quite large (in the order of millions) without much effort so that the distribution of  $\mathbf{m}$  tends to multivariate normal:  $N(R\boldsymbol{\pi}, R^2\mathbf{C})$  as shown by Krzyśko (2000: 31). After the first phase of the simulation, it may be approximated by  $N(R\hat{\boldsymbol{\pi}}, R^2\hat{\mathbf{C}})$ . Realisations of this distribution may be easily and quickly generated in large quantities, for example, by using algorithms described by Zieliński and Wieczorkowski (1997). This enables the estimation of the probability  $Q(R)$  associated with *any* value of  $R$  by counting what percentage of these pseudo-random realisations falls outside the  $\Omega$  region (with the unknown ‘true’ statistic  $\hat{t}(s)$  approximated by  $\hat{t}(s)$  based on  $R_1$  replications). Such estimation is easily repeated for various candidate values of  $R$  because generated replications of multivariate normal distribution may be reused. The re-calculation boils down to a few matrix operations (addition, multiplication, division of corresponding elements) which are easily serialised and optimised. The well-known golden-section or Newton-Raphson algorithms may hence be applied to find the minimum sufficient number  $R$  before the second phase of simulation is initiated.

## 5. Conclusions

According to the approach sketched in the last section, one relatively simple, but very time-consuming simulation, is replaced with a more complex but potentially faster procedure. Instead of calculating a conservative number of replications and then generating them all, a more subtle approach is proposed. After initially generating some  $R_1$  sample replications, the auxiliary nested but fast simulation is executed to establish the required total number  $R$  of replications accounting for correlations among sample membership indicators. Then the second, possibly quite a small batch of  $R - R_1$  replications, is generated, and the empirical HT estimator may finally be evaluated. The nested fast simulation step may eventually be repeated more times when more and more replications are available to make the initial assessment of  $R$  more reliable. It may also be done after all replications are generated to verify that their number is indeed sufficient. Further studies are needed to confirm whether the proposed procedure indeed produces substantial speeding-up of the whole simulation process.

### Acknowledgements

This paper was presented during the MSA 2019 conference which financed its publication. Organisation of the international conference “Multivariate Statistical Analysis 2019” (MSA 2019) was supported from resources for popularisation of scientific activities of the Minister of Science and Higher Education in the framework of agreement No 712/P-DUN/202019.

### References

- Barabesi L., Fattorini L., Ridolfi G. (1997), *Two-Phase Surveys of Elusive Populations*, [in:] *Proceedings of the Statistics Canada Symposium 97: New Directions of Surveys and Censuses*, Statistics Canada, Ottawa, pp. 285–287.
- Beyer W.H. (1987), *CRC Standard Mathematical Tables*, CRC Press, Boca Raton.
- Boistard H., Lopuhaä H.P., Ruiz-Gazen A. (2012), *Approximation of rejective sampling inclusion probabilities and application to higher order correlations*, “Electronic Journal of Statistics”, no. 6, pp. 1967–1983.
- Fattorini L. (2006), *Applying the Horvitz-Thompson Criterion in Complex Designs: A Computer-Intensive Perspective For Estimating Inclusion Probabilities*, “Biometrika”, vol. 93(2), pp. 269–278.
- Fattorini L. (2009), *An Adaptive Algorithm for Estimating Inclusion Probabilities and Performing the Horvitz-Thompson Criterion in Complex Designs*, “Computational Statistics”, vol. 24(4), pp. 623–639.
- Fattorini L., Ridolfi G. (1997), *A Sampling Design for Areal Units Based on Spatial Variability*, “Metron”, no. 55, pp. 59–72.
- Gamrot W. (2013), *On exact computation of minimum sample sizes for restricted estimation of a binomial parameter*, “Journal of Statistical Planning and Inference”, no. 143, pp. 852–866.

- Hedayat A.S., Sinha B.K. (1991), *Design and Inference in Finite Population Sampling*, Wiley, New York.
- Horvitz D.G., Thompson D.J. (1952), *A Generalization of Sampling Without Replacement from a Finite Universe*, "Journal of the American Statistical Association", no. 47, pp. 663–685.
- Kremers W.K. (1985), *The Statistical Analysis of Sum-Quota Sampling*, unpublished PHD thesis, Cornell University, Ithaca.
- Krzyśko M. (2000), *Wielowymiarowa analiza statystyczna*, Uniwersytet Adama Mickiewicza, Poznań.
- Narain R.D. (1951), *On sampling without replacement with varying probabilities*, "Journal of the Indian Society of Agricultural Statistics", no. 3, pp. 169–175.
- Pathak K. (1976), *Unbiased estimation in Fixed-Cost Sequential Sampling Schemes*, "Annals of Statistics", vol. 4(5), pp. 1012–1017.
- Rosén B. (1997), *On Sampling with Probability Proportional to Size*, "Journal of Statistical Planning and Inference", no. 62, pp. 159–191.
- Särndal C.E., Swensson B., Wretman J.H. (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- Tillé Y. (2006), *Sampling Algorithms*, Springer-Verlag, New York.
- Wwywał J. (2003), *On conditional sampling strategies*, "Statistical Papers", no. 44, pp. 397–419.
- Yu Y. (2012), *On the inclusion probabilities in some unequal probability sampling plans without replacement*, "Bernoulli", vol. 18(1), pp. 279–289.
- Zieliński W., Wieczorkowski R. (1997), *Komputerowe generatory liczb losowych*, WNT, Warszawa.

### **Reguła stopu dla estymacji prawdopodobieństw inkluzji na drodze symulacyjnej**

**Streszczenie:** Estymacja parametrów populacji skończonych i ustalonych, prowadzona w ramach podejścia randomizacyjnego, zazwyczaj wymaga znajomości prawdopodobieństw inkluzji charakteryzujących schemat losowania próby. Są one bezpośrednio wykorzystywane w celu wyznaczenia wag przypisanych poszczególnym wylosowanym jednostkom i uwzględniane podczas obliczania estymatorów. Jednak dla pewnych użytecznych schematów losowania pozostają nieznane. W takim wypadku możliwe jest ich wyznaczenie na drodze symulacyjnej, poprzez wielokrotne losowanie prób z wykorzystaniem tego samego schematu losowania i zliczanie wystąpień poszczególnych jednostek populacji. Zastępując nieznane prawdopodobieństwa inkluzji oszacowaniami uzyskanymi w wyniku takiego eksperymentu, otrzymuje się oszacowania wartości globalnej badanej cechy populacji. Szczygólnym wyzwaniem podczas takiego postępowania jest wyznaczenie liczby replikacji próby, zapewniającej wymaganą precyzyję estymacji. W niniejszym artykule proponowana jest nowa procedura, która może przyczynić się do zmniejszenia złożoności obliczeniowej eksperymentu symulacyjnego.

**Słowa kluczowe:** estymator Horvitza-Thompsona, prawdopodobieństwo inkluzji, symulacja, precyzyja

**JEL:** C83, C63

	<p>© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>)</p>
	<p>Received: 2020-04-27; verified: 2020-06-12. Accepted: 2020-09-17</p>
 <b>Member since 2018</b> JM13703	<p>This journal adheres to the COPE's Core Practices <a href="https://publicationethics.org/core-practices">https://publicationethics.org/core-practices</a></p>