A Probability Sampling Approach for Variance Minimization

Sun-Woong Kim, Dongguk University, Steven G. Heeringa and Peter W. Solenberger, University of Michigan
Sun-Woong Kim, Dongguk University, Jung Gu Pil Dong 3 Ga 26, Seoul, South Korea 100-715
Steven Heeringa, Institute for Social Research, 426 Thompson Street, Ann Arbor, Michigan 48106

Key Words: Statistical efficiency, inclusion probability, non-linear programming

1. Introduction

A number of techniques for probability sampling without replacement (SWOR) have been suggested, although it is not clear which method is consistently superior in terms of statistical efficiency. Rao and Bayless (1969) empirically studied the stability of estimators of the population total for a variety of methods of unequal probability SWOR when selecting two units per stratum. One of their major conclusions is that when a stable estimator is required, Murthy's (1957) method is preferred over the methods of Lahiri (1951), Raj (1956), Rao, Hartley and Cochran (1962), Brewer (1963), Fellegi (1963), Hanurav (1967), and probability proportional to size (PPS) sampling with replacement.

One the other hand, Jessen (1969) proposed four interesting sampling schemes. One of them, labeled method 4, shows high efficiency in comparisons of variances of estimators relative to those of alternative SWOR selection schemes, including some of the above-mentioned methods.

However, Jessen's method may be difficult to employ in practical problems due to the arbitrariness and complexities of trials to determine the inclusion probabilities that are required for the variance formula for the estimator of the total population.

In this paper, we first review Jessen's method. Second, we suggest two probability sampling schemes using non-linear programming approaches to overcome certain disadvantages in carrying out Jessen's method. Finally, we illustrate the practicality and statistical efficiency of our methods through application to several examples from the literature.

2. Review of Jessen's method

Assume that the sampler has auxiliary information, that we will label the absolute measure of size, for each unit in a finite population consisting of N units. Let X_i, $i = 1, \cdots, N$ denote the absolute size measure of unit i. The relative measure of size for unit i, P_i is defined by

$$P_i = X_i / \sum_{j=1}^{N} X_j \qquad (2.1)$$

Then π_i, which is called the first-order inclusion probability, is expressed as

$$\pi_i = \sum_{s \in S, i \in s} P(s) = nP_i \ . \qquad (2.2)$$

where S is a set of all possible samples and $P(s)$ denotes the selection probability of a sample, s, of n specified units in the population.

The second-order inclusion probability, π_{ij} for the i th and j th units, $j \neq i$ is defined as

$$\pi_{ij} = \sum_{s \in S, i, j \in s} P(s) \ . \qquad (2.3)$$

Horvitz and Thompson (1952) proposed the following unbiased estimator of the population total and its variance,

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{Y_i}{\pi_i} \qquad (2.4)$$

and

$$Var\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + 2\sum_{i=1}^{N}\sum_{j>i}^{N} \pi_{ij} \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j} - Y^2 \ , \quad (2.5)$$

where Y_i is the characteristic of interest for the i th unit, and $Y = \sum_{i=1}^{N} Y_i$. Y_{HT} is commonly referred to as the Horvitz-Thompson estimator.

Equation (2.5) may be expressed in the alternate form

$$Var\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N}\sum_{j>i}^{N} \left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \ , \quad (2.6)$$

which was first derived by Yates and Grundy (1953).

Jessen (1969) examined the influence of the second-order inclusion probability, π_{ij} on the variance of Horvitz-Thompson estimator represented in (2.6). His method is described as follows.

Let

$$W_{ij} = \pi_i \pi_j - \pi_{ij} \ . \qquad (2.7)$$

Then (2.8) below is derived using (2.7)

$$W = \sum_{i=1}^{N}\sum_{j>i}^{N} W_{ij}$$

$$= \sum_{i=1}^{N} \sum_{j>i}^{N} \pi_i \pi_j - \frac{1}{2} \sum_{i=1}^{N} \pi_i \left(\sum_{i=1}^{N} \pi_i - 1 \right)$$

$$= \frac{n - \sum_{i=1}^{N} \pi_i^2}{2} \qquad (2.8)$$

Since the number of all possible samples is $N!/(2!(N-2)!)$, the average of all W_{ij} s can be expressed as:

$$\overline{W} = \left(n - \sum_{i=1}^{N} \pi_i^2 \right) \Big/ N(N-1) . \qquad (2.9)$$

Then Equation (2.6) can be written in another form

$$Var\left(Y_{HT}\right) = \overline{W} \sum_{i=1}^{N} \sum_{j>i}^{N} \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2 . \qquad (2.10)$$

However, using Equation (2.10) is almost impossible in practice because in real sampling problems the X_i s are typically unequal, though $W_{ij} = \overline{W}$ is one hypothetical possibility.

Jessen proposed the method using W_{ij} $\overline{W}$ instead of $W_{ij} = \overline{W}$ when selecting samples of size $n = 2$. His method consists of the following steps.

Step 1. Calculate the π_i s and $\overline{W}$ by using the absolute size measure of each unit in a finite population.

Step 2. Compute π_{ij}^* s, regarded as the first approximation to the desired π_{ij} s that minimizes Equation (2.6), where

$$\pi_{ij}^* = \pi_i \pi_j - \overline{W} . \qquad (2.11)$$

Step 3. Check if π_{ij}^* do meet the following requirement for all i, j.

$$\sum \pi_{ij}^* = \pi_k , \ i, j, k = 1, \cdots, N , \ j > i , \quad (2.12)$$

where the sum is over all π_{ij} s containing the subscript k for either i or j.

Step 4. If there is significant discrepancy, the final second-order inclusion probabilities are determined by using the $(N-1) \times N$ tableau (See Jessen (1969, Table 3.2)) so that the following approximation holds instead of Equation (2.12).

$$\sum \pi_{ij}^* \quad \pi_k , \ k = 1, \cdots, N \qquad (2.13)$$

With respect to the variance of the estimator of the population total, Jessen empirically compared his three methods, that is, methods 3 and 4 and the method using $\overline{W}$ to alternatives including PPS sampling with replacement and the methods of Raj (1956), Yates and Grundy (1953), Hartley and Rao (1962), Narain (1951), Cochran (1963), Rao, Hartley and Cochran (1962), and Horvitz and Thompson (1952). In the results of comparisons using the two examples from Yates and Grundy (1953) and Cochran (1963), the Raj and Narain methods perform best. Jessen's method 4 has the second highest average efficiency for three populations in each example. The paper points out that the example problems vary widely in terms of the variances of the population totals of interest and it may be beneficial to favor a method that is robust, one that is least sensitive to population characteristics. Furthermore, Jessen points out that among the ten alternatives tested Raj's and Narain's methods are perhaps the most difficult to apply in practical settings.

Considering these results and Jessen's argument for the desirability of a robust estimator Jessen's method might be preferred over the alternatives included in that study.

However, there are some disadvantages in using Jessen's method. First of all, it is difficult to employ in practical problems due to the arbitrariness of the trials to satisfy the requirement specified in Equation (2.13) which requires finding the final second-order inclusion probabilities using the $(N-1) \times N$ tableau in the Step 4. In addition, Jessen's method is limited to samples of size $n = 2$.

In the next section we propose new non-linear programming approaches to avoid these problems and describe how to implement them.

3. Non-linear programming approaches

A non-linear programming (NLP) problem is specified as

minimize $f(\underline{x})$ (3.1)

subject to $g_i(\underline{x}) \geq 0$ for all $i = 1, \cdots, l$,   (3.2)

$h_i(\underline{x}) = 0$ for all $i = 1, \cdots, m$,   (3.3)

where $\underline{x}$ is a vector of t components $x_1, x_2, \cdots, x_t$, the function f is called the objective function. The expressions, $g_i(\underline{x}) \geq 0$ and $h_i(\underline{x}) = 0$ are called the inequality constraint and the equality constraint respectively.

The solution to a NLP requires finding a feasible point $\underline{x}^*$ such that $f(\underline{x}) \geq f(\underline{x}^*)$ for each feasible point $\underline{x}$. The feasible point is called the optimal

solution to the problem. The maximization of a NLP problem can be also considered. When the objective function and all constraints are linear, the problem, which is the special case of a NLP problem, is called a linear programming (LP) problem.

We may consider the following NLP approach which is quite straightforward since Jessen's method finds π_{ij}^* s roughly satisfying (2.13) and $W_{ij} \quad \overline{W}$. Designate a set of π_{ij} such that the following objective function is minimized:

$$\sum_{i=1}^{N}\sum_{j>i}^{N}\left(W_{ij}-\overline{W}\right)^2 = \sum_{i=1}^{N}\sum_{j>i}^{N}\left(\left(\pi_i\pi_j-\pi_{ij}\right)-\overline{W}\right)^2, (3.4)$$

subject to the inequality constraint,

$$\pi_{ij} \geq 0, \ j > i = 1, \cdots, N, \qquad (3.5)$$

and the equality constraint,

$$\sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i, \ i = 1, \cdots, N. \qquad (3.6)$$

The components $x_1, x_2, \cdots, x_n$ in the general NLP problem specification are the second-order inclusion probabilities, π_{ij} s, and (3.1), (3.2) and (3.3) correspond to (3.4), (3.5) and (3.6) respectively.

Note that since $\overline{W}$ is a constant, minimization of (3.4) is equivalent to minimizing

$$\sum_{i=1}^{N}\sum_{j>i}^{N}W_{ij}^2 = \sum_{i=1}^{N}\sum_{j>i}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)^2. \qquad (3.7)$$

(3.7) is also equivalent to

$$\sum_{i=1}^{N}\sum_{j>i}^{N}\left(\pi_{ij}^2 - 2\pi_i\pi_j\pi_{ij}\right) \qquad (3.8)$$

due to the fixed values π_i and π_j.

By using the NLP approach that consists of (3.4) or (3.7) or (3.8) as a nonlinear objective function and the linear constraints of (3.5) and (3.6), we can achieve optimal control over the second-order inclusion probabilities π_{ij} for reducing the variance of Horvitz-Thompson estimator of Equation (2.6) without the repeated trials of Jessen's method. But the optimization does not mean that the NLP approach solution always yields smaller variances than Jessen's method. However, the proposed NLP approach does present the exact solution to minimize the objective function under the constraints. We will refer to this NLP method as alternative I.

While alternative I is an effective alternative to Jessen's method, we also consider a second, more direct approach (alternative II).

From (3.4) we can introduce the objective function

$$\sum_{i=1}^{N}\sum_{j>i}^{N}W_{ij} = \sum_{i=1}^{N}\sum_{j>i}^{N}\pi_i\pi_j - \sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij}. \qquad (3.9)$$

Since the first term on the right-hand is a fixed value, minimization of (3.9) amounts to maximization of

$$\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij}. \qquad (3.10)$$

Thus the simpler function, which is a linear form having equal weight for each second-order inclusion probability, can be used as the objective function instead of (3.4) under the same constraints as the alternative I NLP specification. In this case where all possible second-order inclusion probabilities are enumerated in (3.10), the NLP algorithm is not maximizing individual π_{ij} s but simply finding solutions meeting the constraints. Note that the objective function (3.10) is now indicated by

$$\sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{s \in S, i, j \in s} P(s) = \frac{1}{2}(n-1)\sum_{i=1}^{N}\sum_{s \in S, i \in s} P(s)$$

$$= \frac{1}{2}n(n-1)\sum_{s \in S} P(s) \qquad (3.11)$$

and in this function the selection probabilities of individual samples have the equal weight of $1/2\,n(n-1)$.

In order to implement our alternatives I and II, we use the SAS/OR NLP procedure in order to optimize non-linear or linear objective functions under the linear constraints. A variety of NLP algorithms are available and SAS/OR (2001) provides the capability details to choose an optimization algorithm.

For alternative II, LP can also be used since both the objective function and the constraints are linear. Unlike Jessen's method, the NLP approaches would not be restricted to the sample of size $n = 2$. When the sample size n is two, most NLP algorithms will yield the same solutions.

Since the second-order inclusion probabilities that are the solution set under the NLP approaches determine the $P(s)$, as shown in Equation (2.3), we can obtain the selection probability for each sample by solving those equations for $P(s)$. We can select a sample s with the specified $P(s)$ by using the well-known method of cumulative sums.

4. Numerical examples

We have chosen several example problems from the literature to empirically evaluate the NLP approaches, the alternatives I and II, and compare

their performance to Jessen's method. Table 4.1 shows the first example from Yates and Grundy (1953). There are three artificial populations each consisting of $N = 4$ units. The value of the variable of interest, Y_i, and the relative size of each unit, P_i, are given. Note that ranges of Y_i/P_i for the three example populations are not extreme. Two units from each population are selected as a sample s.

Table 4.1 Description of Three Populations

Unit	$i:$	1	2	3	4
Relative Size	$P_i:$	0.1	0.2	0.3	0.4
Population A	$Y_i:$	0.5	1.2	2.1	3.2
	$Y_i/P_i:$	5	6	7	8
Population B	$Y_i:$	0.8	1.4	1.8	2.0
	$Y_i/P_i:$	8	7	6	5
Population C	$Y_i:$	0.2	0.6	0.9	0.8
	$Y_i/P_i:$	2	3	3	2

Prior to empirical evaluations for the new sampling schemes and other selected schemes, we need to focus on a visible comparison between the solutions on the second-order inclusion probabilities π_{ij} s from the Jessen's method and the alternatives I and II. Table 4.2 presents the results obtained by using the three methods. The result for Jessen's method is from Jessen (1969).

Table 4.2 Solutions of Three Methods

Units	π_{ij}		
(i, j)	Jessen's Method	Alternative I	Alternative II
1, 2	0.010	0.013	0.000
1, 3	0.050	0.053	0.050
1, 4	0.140	0.133	0.150
2, 3	0.140	0.133	0.150
2, 4	0.250	0.253	0.250
3, 4	0.410	0.413	0.400

As described in the section 3, alternative I uses the NLP approach to get over the disadvantages such as the arbitrariness and complexities in implementing Jessen's method. Thus we may expect that the solutions for the second-order inclusion probabilities from the two methods should be similar. This expectation is confirmed in Table 4.2 which shows very similar values for the two sets of the second-order inclusion probabilities. Also, note that the solutions are very different from Jessen's and the Alternative I solutions.

A comparison for the variance of the population total of the two NLP alternatives with those of the

Jessen's method, Yates and Grundy (1953), Raj (1956), Hartley and Rao (1962) and PPS sampling with replacement is given in Table 4.3. The Yates and Grundy method uses an iterative procedure to decide a set of "adjusted measures of size" which approximately satisfy

$$\sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i = kX_i , \qquad (4.1)$$

where k is a proportional constant.

The Raj's method first adopts the assumption that

$$Y_i = \alpha + \beta X_i , \qquad (4.2)$$

where there is no information on the actual values of α and β.

Then it obtains an optimal set of π_{ij} minimizing the variance of the form (2.5) using linear programming methods. Hartley and Rao's scheme arranges the units in a random order and then systematically selects a sample of n using the "progressive sums of nP_i ." For sampling with probability proportional to size with replacement see the pages 252-253 of Cochran (1977).

The variances in Table 4.3 are provided from Raj (1956) for Raj's method and the Yates and Grundy method, Hartley and Rao (1962) for Hartley and Rao method and sampling with probability proportional to size with replacement and Jessen (1969) for Jessen's method. The correlation between the size measure X_i and the characteristic of interest Y_i for the three populations are respectively: 0.995, 0.976, 0.876.

Table 4.3 Comparison of Variances

Pop.	$Var(Y)$						
	PPS	R	$\bar{J}$	AI	AII	YG	HR
A	0.500	0.200	0.245	0.253	0.225	0.323	0.367
B	0.500	0.200	0.245	0.253	0.225	0.269	0.367
C	0.125	0.100	0.070	0.067	0.075	0.057	0.033
Average	0.375	0.167	0.187	0.191	0.175	0.216	0.256
Rel. Eff.	100	225	201	196	214	173	147

Note. *PPS* : Sampling with probability proportional to size with replacement

    R : Raj's method

    $\bar{J}$: Jessen's method (average over three trials)

    AI : Alternative I

    AII : Alternative II

    YG : Yates and Grundy method

    HR : Hartley and Rao method

To choose a preferred method, we consider two criteria: smaller average variances across problems (efficiency); low sensitivity of the method to population characteristics (robustness). By these criteria, Raj's method seems to be most preferable because it has the highest statistical efficiency. Note, however, that Raj's method is unlikely to be efficient if the relation between X_i and Y_i is not linear. Also, his method like the Jessen's method is restricted to samples of size $n = 2$.

Alternative II is clearly better than others except for Raj's method. Jessen's method and alternative I have a small difference for the variance owing to the similar solutions shown in Table 4.2.

The second example is available from Cochran (1977). Table 4.4 shows three artificial populations that have five units and different characteristics of interest, as well as the different relative sizes for the units. For reference the three populations have correlation between the variable of interest and the measure of size of 0.873, 0.997 and 0.275 respectively. The sample size for this example is $n = 2$.

Table 4.4 Description of Three Populations

Unit	i :	1	2	3	4	5
Rel. Size	P_i :	0.1	0.1	0.2	0.3	0.3
Pop. A	Y_i :	0.3	0.5	0.8	0.9	1.5
	Y_i/P_i :	3	5	4	3	5
Pop. B	Y_i :	0.3	0.3	0.8	1.5	1.5
	Y_i/P_i :	3	3	4	5	5
Pop. C	Y_i :	0.7	0.6	0.4	0.9	0.6
	Y_i/P_i :	7	6	2	3	2

Table 4.5 describes a comparison of our NLP approaches to useful sampling schemes described by Brewer (1963), Murthy (1957) and Rao, Hartley and Cochran (1962), as well as to simple random sampling and PPS sampling with replacement.

Brewer's scheme, which is relevant only for $n = 2$, draws the first unit with probability proportional to $P_i(1 - P_i)/(1 - 2P_i)$, and selects the second unit with probability $P_i/(1 - P_j)$, where j is the first unit drawn. Murthy's scheme draws the successive units with probabilities P_i, $P_j/(1 - P_i)$, $P_k/(1 - P_i - P_j)$, and so on. Rao, Hartley and Cochran's method first divides the units in the population into n random groups for a sample of size n and selects one unit from each group.

The variances for the other schemes except for alternatives I and II are from Cochran (1977).

Table 4.5 Comparison of Variances

Pop.	$Var(Y)$						
	SRS	PPS	B	AI	AII	M	RHC
A	1.575	0.400	0.246	0.247	0.278	0.267	0.320
B	2.715	0.320	0.248	0.247	0.184	0.237	0.256
C	0.248	1.480	1.251	1.290	1.160	1.130	1.184
Average	1.513	0.733	0.582	0.594	0.541	0.545	0.587
Rel. Eff.	100	206	260	254	280	278	258

Note. *SRS* : Simple random sampling

PPS : Sampling with probability proportional to size with replacement

B : Brewer's method

AI : Alternative I

AII : Alternative II

M : Murthy's method

RHC : Rao, Hartley and Cochran method

As presented in Table 4.5, alternative II is most efficient followed by Murthy's method and then Brewer's method. Alternative I is less efficient, but it is much better than simple random sampling and sampling with probability proportional to size with replacement. Also, for these example problems there is only a small difference in efficiency between alternative I and the Rao, Hartley and Cochran method.

5. Conclusion

In this paper, we suggest two probability sampling schemes using NLP approaches. Alternative I avoids the practical problems encountered in Jessen's method and provides the optimum solution for the second-order inclusion probabilities. Based on the tests described here, NLP alternative II appears to be preferred over alternative I and other sampling schemes with respect to both statistical efficiency and practicality of use.

Proposed sampling schemes would be useful in the stratified multistage cluster sampling design, where two clusters are drawn from each stratum.

Some empirical comparisons will be followed for the cases in which the sample size is greater than two. In future research, we are going to examine the stability and the non-negativity of the Yates and Grundy (1953)'s variance estimator for the Horvitz-Thompson estimator of population totals.

References

Brewer, K. R. W. (1963). "A model of systematic sampling with unequal probabilities," *Australian Journal of Statistics*, 5, 5-13.

Cochran, W. G. (1963). *Sampling Techniques*, 2nd Ed. New York: John Wiley and Sons.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd Ed. New York: John Wiley and Sons.

Fellegi, I. P. (1963). "Sampling with varying probabilities without replacement: rotating and non-rotating samples," *Journal of the American Statistical Association*, 58, 183-201.

Hanurav, T. V. (1967). "Optimum utilization of auxiliary information: πPS sampling of two units from a stratum," *Journal of the Royal Statistical Society, Series B*, 29, 374-391.

Hartley, H. O. and Rao, J. N. K. (1962). "Sampling with unequal probabilities and without replacement," *Annals of Mathematical Statistics*, 33, 350-374.

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

Jessen, R. J. (1969). "Some methods of probability non-replacement sampling," *Journal of the American Statistical Association*, 64, 175-193.

Lahiri, D. B. (1951). "A method for sample selection providing unbiased ratio estimates," *International Statistical Institute Bulletin*, 33, No. 2, 133-140.

Murthy, M. N. (1957). "Ordered and unordered estimators in sampling without replacement," *Sankhya*, 18, 379-390.

Narain, R. D. (1951). "On sampling without replacement with varying probabilities," *Journal of the Indian Society of Agricultural Statistics*, 3, 169-175.

Raj, Des (1956). "A note on the determination of optimum probabilities in sampling without replacement," *Sankhya*, 17, 197-200.

Rao, J. N. K. and Bayless, D. L. (1969). "An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum," *Journal of the American Statistical Association*, 64, 540-559.

Rao, J. N. K., Hartley, H. O., and Cochran, W. G. (1962). "On a simple procedure of unequal probability sampling without replacement," *Journal of the Royal Statistical Society, Series B*, 24, 482-491.

SAS/OR (2001). *User's Guide: Mathematical Programming*, Version 8, Cary, NC: SAS Institute Inc.

Yates, F. and Grundy, P. M. (1953). "Selection without replacement from within strata with probability proportional to size," *Journal of the Royal Statistical Society, Series B*, 15, 253-261.