Inclusion Probability Proportional to Size Sampling: A Nonlinear Programming Approach to Ensure a Nonnegative and Stable Variance Estimator

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: Sen-Yates-Grundy Variance Estimator, Second-order Inclusion Probabilities, Binary Block Designs, Nonlinear Programming

1. Introduction

Since 1943 when Hansen and Hurwitz first introduced the use of probability proportional to size (PPS) sampling, a number of procedures for selecting samples without replacement have been developed. Many of them are reviewed and compared in Brewer and Hanif (1983). One of the popular methods is an Inclusion Probability Proportional to Size (IPPS) sampling schemes used in combination with the Horvitz-Thompson (H-T) (1952) estimator for the population total.

Jessen (1969) proposed four interesting IPPS sampling schemes and examined their properties. Nigam, Kumar, Gupta (1984) suggested an IPPS sampling scheme which is closely related to the methods of Jessen and provides a more stable variance estimator. However, their method uses binary block designs in an experimental design approach that involves considerable trial and error to carry out even for the cases where the sample size is small.

In this paper, we first suggest an IPPS sampling scheme for nonnegative and stable variance estimation. The method is simple to implement because it is structured as a nonlinear programming problem consisting of a nonlinear objective function and some linear constraints having flexible features. Second, we introduce two other IPPS sampling schemes that are originally developed by Kim, Heeringa and Solenberger (2003). They also adopt constraints similar to those of the first approach to guarantee the non-negativity and stability of the variance estimator. Third, we presents several strict constraints in nonlinear programming approaches that always yield the smaller variance estimator compared to methods such as PPS sampling with replacement, Murthy (1957)'s method and Brewer (1963)'s method, although they are restricted to the cases where the sample size is two. Finally, we illustrate statistical efficiency of the variance estimator as well as variance of the H-T estimator for our methods by applying them to an example problem in the literature.

2. Variance Estimator of Horvitz-Thompson Estimator

Let x_i represent the size measure of the unit i in a population of N distinct and identifiable units and the relative measure of size for unit i denote p_i by

$$p_i = x_i / X \,, \tag{2.1}$$

where $X = \sum_{j=1}^{N} x_j$.

Further, let the selection probability of a sample, s, of n specified units in the population be denoted by $p(s)$. The function $p(\cdot)$ is often called the sampling design or sampling plan. Then the probability of selecting unit i, so-called the first-order inclusion probability, π_i is defined by

$$\pi_i = \sum_{i \in s, s \in S} p(s) \,, \tag{2.2}$$

where S is a set of all possible samples.

Similarly, the second-order inclusion probability indicating the total probability of selecting units i and j, π_{ij} is defined by

$$\pi_{ij} = \sum_{i,j \in s, s \in S} p(s) \;. \tag{2.3}$$

The H-T estimator of the population total is

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \,, \tag{2.4}$$

where y_i is the value of the characteristic of interest for the unit i .

The variance of the H-T estimator is also given as follows:

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

Therefore, an unbiased estimator of (2.5) is given by

$$\widehat{Var}\left(\hat{Y}_{HT}\right)=\sum_{i=1}^{n}\frac{y_i^2(1-\pi_i)}{\pi_i^2}-2\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{y_i y_j(\pi_i\pi_j-\pi_{ij})}{\pi_i\pi_j\pi_{ij}}\ . \quad (2.6)$$

A different form of (2.6) is given by Sen (1953) and Yates and Grundy (1953). This estimator is

$$\widehat{Var}_{SYG}\left(\hat{Y}_{HT}\right)=\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \quad (2.7)$$

based on a more suitable form of the variance of (2.3)

$$Var_{SYG}\left(\hat{Y}_{HT}\right)=\sum_{i=1}^{N}\sum_{j>i}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2. \quad (2.8)$$

The squared factor in the Sen-Yates-Grundy variance estimator (2.7) (or the variance of the H-T estimator (2.8)) will be smaller when the variation between the y_i/π_i is small, for example, when the π_i is proportional to the x_i and the y_i are also proportional to the x_i .

The $(\pi_i\pi_j-\pi_{ij})/\pi_{ij}$ factor in (2.7), often varies widely and can be negative and unstable, depending on the sampling design. Hence achieving non-negativity and stability of the Sen-Yates-Grundy variance estimator may be essential in creating a sampling design. In addition, the second-order inclusion probability π_{ij} must be larger than zero with respect to unbiased variance estimation. These are often called the desirable properties for variance estimation.

3. The Suggested Approaches

In a sample survey we usually assume that the chosen size measures of the units in the population may be approximately proportional to the values of the characteristic of interest. We introduce the following three nonlinear programming (NLP) approaches for constructing a sampling design that not only has the smaller variance estimates (or the smaller variance) but also achieves the desirable properties in variance estimation.

(1) Approach I

We may be interested in reducing the variation between the non-squared factor in (2.7) so that we obtain the sampling design providing the smaller estimated variances (or producing the smaller bounds on the error of estimation) over all possible samples. In order to do so, we first consider the following problem:

$$Minimize\sum_{s\in S}\widehat{Var}_{SYG}\left(\hat{Y}_{HT}\right), \quad (3.1)$$

where Σ represents summation over all possible samples.

In theory the hypothetical situation where the non-squared factor in (2.7) is equal to a constant can be considered to succeed in solving the problem (3.1). In other words, if

$$(\pi_i\pi_j-\pi_{ij})/\pi_{ij}=(\pi_i\pi_j/\pi_{ij})-1=\Pi-1, \quad (3.2)$$

where Π is a constant,

(3.1) reduces to

$$\Pi^*\sum_{s\in S}\left[\sum_{i=1}^{n}\sum_{j>i}^{n}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2\right], \quad (3.3)$$

where $\Pi^*=\Pi-1$.

But since the size measures of the units in the population are different, achieving $\pi_i\pi_j/\pi_{ij}=\Pi$ for all possible samples is impossible. Thus to achieve (3.1) we may consider to use

$$Minimize\sum_{s\in S}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}-\Pi\right), \quad (3.4)$$

which is equivalent to

$$Minimize\left[\frac{(N-2)!}{(n-2)!(N-n)!}\sum_{i}^{N}\sum_{j>i}^{N}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}-\Pi\right)\right]. \quad (3.5)$$

Note that (3.5) amounts to

$$Minimize\sum_{i}^{N}\sum_{j>i}^{N}\frac{\pi_i\pi_j}{\pi_{ij}} \quad (3.6)$$

The Approach I consists of the following steps:

First, establish the nonlinear objective function in (3.6), that is, $\sum_{i}^{N}\sum_{j>i}^{N}\pi_i\pi_j/\pi_{ij}$.

Second, we add the following two constraints to the objective function:

(i) Bounded constraints

$$c\pi_i\pi_j < \pi_{ij} \le \pi_i\pi_j\,,\;\; j > i = 1,\cdots,N\;, \qquad (3.7)$$

where c is a real number between 0 and 1.

(ii) IPPS constraints

$$\sum_{j\neq i}^{N}\pi_{ij} = (n-1)\pi_i\;,\;\; i = 1,\cdots,N\;. \qquad (3.8)$$

It is noted that the constraint (3.7) is required for both non-negativity and stability of the Sen-Yates-Grundy variance estimator (2.7).

Finally, by using a NLP algorithm we find a set of the second-order inclusion probabilities, the optimum solution to the NLP problem consisting of the objective function (3.6), the constraints, (3.7) and (3.8), and an appropriate value of c which determines the level of stability of the variance estimator.

Note that since the second-order inclusion probabilities are expressed as the sum of the selection probabilities, as in (2.3), we can obtain the selection probability of each sample, *p(s)*, from the equations of those second-order inclusion probabilities.

(2) Approaches II and III

One of the purposes of the Approach I is to minimize the estimated variance, not necessarily the variance of the H-T estimator. But it is directly related to the minimization of the variance because the Sen-Yates-Grundy variance estimator (2.7), is an unbiased estimator of the variance of the H-T estimator (2.8) and the relation is expressed as

$$E\left[\widehat{Var}_{SYG}\left(\hat{Y}_{HT}\right)\right] = \sum_{s\in S}\left[\widehat{Var}_{SYG}\left(\hat{Y}_{HT}\right)\right]\, p(s)$$
$$= Var_{SYG}\left(\hat{Y}_{HT}\right). \qquad (3.9)$$

So the NLP approach to minimize the variance of the H-T estimator can also result in minimizing the estimated variances of samples selected under the design solution.

The two NLP approaches to minimize the variance have been developed by Kim, Heeringa and Solenberger (2003) (See pages 2169-2170 for the details), but the previous methods did not guarantee the non-negativity and stability of the variance estimator. However, to achieve non-negativity and stability it is only necessary to add the bounded constraints (3.7) and the IPPS constraints (3.8) to each of the following objective functions proposed by them.

$$Minimize\sum_{i=1}^{N}\sum_{j>i}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)^2 \qquad (3.10)$$

$$Maximize\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij} \qquad (3.11)$$

Accordingly, under the same constraints as in the Approach I we can find the solution to (3.10) or (3.11). Here we call those NLP Approaches II and III, respectively.

The bounded constraints (3.7) can be expressed as

$$c < \delta_{ij} \le 1\,,\;\; j > i = 1,\cdots,N\;, \qquad (3.12)$$

where $\delta_{ij} = \dfrac{\pi_{ij}}{\pi_i\pi_j}$.

Given this, obtaining a sampling design that achieves *MINMAX* δ_{ij} , which indicates the maximum among the minimum δ_{ij} s from the possible NLP solutions for each approach, may be preferable since it results in better stability for the variance estimator.

A variety of NLP software is available for implementing the three approaches. We use SAS/OR® software, specifically the NLP Procedure to optimize those objective functions under the certain constraints. Refer to SAS/OR® (2001) for the details.

The choice of an appropriate value of c that yields a set of the second-order inclusion probabilities achieving *MINMAX* δ_{ij} is not difficult. By repeating the steps as in the Approach I and applying some reasonable rules we can easily find the maximum value of c that still permits a solution to the NLP problem.

4. Strict Constraints for NLP approaches

PPS sampling with replacement or Murthy (1957)'s method or Brewer (1963)'s method are established sample selection methods. These methods have been widely used in many statistical organizations and by a lot of survey samplers.

The flexibility in the use of the value of c in the bounded constraints (3.7) in our NLP approaches is important since we can determine the value of the constraints that always yield the smaller variance estimator compared to those popular methods. We present those constraints as follows. Here the sample size is restricted to $n = 2$.

Since $\pi_i = 2p_i$, the Sen-Yates-Grundy variance estimator (2.7) can be written

$$\widehat{Var}_{SYG}\left(\hat{Y}_{HT}\right) = \left(\frac{p_i p_j}{\pi_{ij}} - \frac{1}{4}\right)\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2, \quad j > i = 1, \cdots, N \qquad (4.1)$$

For PPS sampling with replacement, the variance estimator is

$$\widehat{Var}\left(\hat{Y}_{PPS}\right) = \frac{1}{4}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2, \qquad (4.2)$$

where $\hat{Y}_{PPS} = \frac{1}{2}\left(\frac{y_i}{p_i} + \frac{y_j}{p_j}\right)$, $j > i = 1, \cdots, N$

The variance estimator for Murthy (1957)'s method is given by

$$\widehat{Var}\left(\hat{Y}_M\right) = \frac{(1-p_i)(1-p_j)(1-p_i-p_j)}{(2-p_i-p_j)^2}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2, \qquad (4.3)$$

where $\hat{Y}_M = \frac{1}{2-p_i-p_j}\left[(1-p_j)\frac{y_i}{z_i} + (1-p_i)\frac{y_j}{z_j}\right]$, $j > i = 1, \cdots, N$. Refer to Cochran (1977), page 264.

By comparison of (4.1) and (4.2), for all samples clearly

$$\widehat{Var}_{SYG}\left(\hat{Y}_{HT}\right) < \widehat{Var}\left(\hat{Y}_{PPS}\right) \qquad (4.4)$$

when

$$\frac{p_i p_j}{\pi_{ij}} - \frac{1}{4} < \frac{1}{4}. \qquad (4.5)$$

An alternative form of (4.5) can be given by

$$c_{PPS}\ \pi_i \pi_j < \pi_{ij}, \qquad (4.6)$$

where $c_{PPS} = 0.5$.

Hence when we use $c = 0.5$ in the bounded constraints (3.7) in our NLP approaches, it is expected that we can always obtain the smaller variance estimator than in PPS sampling with replacement.

Similarly, from (4.1) and (4.3), we see that if

$$\frac{p_i p_j}{\pi_{ij}} - \frac{1}{4} < \frac{(1-p_i)(1-p_j)(1-p_i-p_j)}{(2-p_i-p_j)^2}, \qquad (4.7)$$

then

$$\widehat{Var}_{SYG}\left(\hat{Y}_{HT}\right) < \widehat{Var}\left(\hat{Y}_M\right). \qquad (4.8)$$

Since (4.7) can be expressed as

$$c_{ij,M}\ \pi_i \pi_j < \pi_{ij}, \qquad (4.9)$$

where $c_{ij,M} = \frac{1}{p_{ij,M} + 1}$ and

$$p_{ij,M} = \frac{4(1-p_i)(1-p_j)(1-p_i-p_j)}{(2-p_i-p_j)^2}, \qquad (4.10)$$

using the non-constant $c_{ij} = \frac{1}{p_{ij,M} + 1}$, $j > i = 1, \cdots, N$ in the NLP approaches, which is not a constant, can always yield the smaller variance estimator than in Murthy's method.

Considering $p_{ij,M} < 1$ for all i, j in (4.10), the following relation is easily derived.

$$c_{PPS} < c_{ij,M} \qquad (4.11)$$

Thus the use of $c_{ij,M}$ gives more strict constraints than in the use of $c_{PPS} = 0.5$.

On the other hand, Brewer's method is an IPPS sampling scheme like our NLP approaches, and (2.7) is used as the formula of the variance estimator. To provide the smaller variance estimator than in Brewer's method, we need to use the constraints

$$c_{ij,B}\ \pi_i \pi_j < \pi_{ij}, \qquad (4.12)$$

where $c_{ij,B} = \frac{1}{4D}\left(\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right)$ and $D=\sum_{i=1}^{N}\frac{p_i(1-p_i)}{1-2p_i}$.

Since $c_{ij,B}$ is not always larger than 0.5, we cannot say that $c_{ij,B}$ always yields stricter constraints than $c_{PPS}=0.5$.

We have derived these strict constraints to illustrate what is required for our NLP approaches to yield smaller variance estimates than in the well-known three methods. Successful NLP solutions that apply constraints that match or exceed the values of c_{PPS} or $c_{ij,M}$ or $c_{ij,B}$ can achieve the smaller variance estimates as well as greater stability of the variance estimator. But note that whether or not the solutions of the NLP approaches exist subject to the strict constraints may depend on the sampling design problem, that is, the size measures of the units in the population.

5. An Illustration

We have chosen a numerical example given by Yates and Grundy (1953) to examine the results from the suggested NLP approaches with respect to the desirable properties for the Sen-Yates-Grundy variance estimator as well as the reductions in the variance of H-T estimator. Table 5.1 presents the three artificial populations of $N=4$ showing the relative size of each unit, p_i and the value of the characteristic of each unit, y_i . A sample of size $n=2$ is selected from each population. Note that when $n=2$, $\pi_{ij}=p(s)$. The correlation coefficients between the size measure p_i and the value of the characteristic y_i for the three populations are respectively: 0.995, 0.976, 0.876.

Table 5.1 Three Populations of $N=4$

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
Population B	y_i :	0.8	1.4	1.8	2.0
Population C	y_i :	0.2	0.6	0.9	0.8

The π_{ij} and the δ_{ij} obtained from Approach I, Approach II and Approach III as the value of c is increased by 0.100 from $c=0.000$ to $c=0.400$ are shown in Table 5.2, Table 5.3 and Table 5.4, respectively. Note that the results for Approach I are identical for $c=0.100$ to $c=0.400$ and those in Approach II are same when $c=0.000$ and $c=0.100$. The π_{ij} s from the three NLP approaches are very different from each other, yielding different δ_{ij} s, except for $c=0.400$ in Approaches II and III. But the pairs of units having minimum δ_{ij} are same for Approaches II and III.

Table 5.2 π_{ij} and δ_{ij} from Approach I

	c				
	0.000	0.100	0.200	0.300	0.400
π_{12} (δ_{12})	0.0000 (0.0000*)	0.0451 (0.5638)	0.0451 (0.5638)	0.0451 (0.5638)	0.0451 (0.5638)
π_{13} (δ_{13})	0.0742 (0.6181)	0.0570 (0.4752)	0.0570 (0.4752)	0.0570 (0.4752)	0.0570 (0.4752)
π_{14} (δ_{14})	0.1258 (0.7864)	0.0979 (0.6117)	0.0979 (0.6117)	0.0979 (0.6117)	0.0979 (0.6117)
π_{23} (δ_{23})	0.1258 (0.5243)	0.0979 (0.4078*)	0.0979 (0.4078*)	0.0979 (0.4078*)	0.0979 (0.4078*)
π_{24} (δ_{24})	0.2742 (0.8568)	0.2570 (0.8032)	0.2570 (0.8032)	0.2570 (0.8032)	0.2570 (0.8032)
π_{34} (δ_{34})	0.4000 (0.8333)	0.4451 (0.9273)	0.4451 (0.9273)	0.4451 (0.9273)	0.4451 (0.9273)

Note. * : minimum δ_{ij}

Table 5.3 π_{ij} and δ_{ij} from Approach II

	c				
	0.000	0.100	0.200	0.300	0.400
π_{12} (δ_{12})	0.0133 (0.1667*)	0.0133 (0.1667*)	0.0160 (0.2000*)	0.0240 (0.3000*)	0.0320 (0.4000*)
π_{13} (δ_{13})	0.0533 (0.4444)	0.0533 (0.4444)	0.0520 (0.4333)	0.0480 (0.4000)	0.0480 (0.4000*)
π_{14} (δ_{14})	0.1333 (0.8333)	0.1333 (0.8333)	0.1320 (0.8250)	0.1280 (0.8000)	0.1200 (0.7500)
π_{23} (δ_{23})	0.1333 (0.5556)	0.1333 (0.5556)	0.1320 (0.5500)	0.1280 (0.5333)	0.1200 (0.5000)
π_{24} (δ_{24})	0.2533 (0.7917)	0.2533 (0.7917)	0.2520 (0.7875)	0.2480 (0.7750)	0.2480 (0.7750)
π_{34} (δ_{34})	0.4133 (0.8611)	0.4133 (0.8611)	0.4160 (0.8667)	0.4240 (0.8833)	0.4320 (0.9000)

Note. * : minimum δ_{ij}

Table 5.4 π_{ij} and δ_{ij} from Approach III

	c				
	0.000	0.100	0.200	0.300	0.400
π_{12} (δ_{12})	0.0000 (0.0000*)	0.0080 (0.1000*)	0.0160 (0.2000*)	0.0240 (0.3000*)	0.0320 (0.4000*)
π_{13} (δ_{13})	0.0500 (0.4167)	0.0515 (0.4292)	0.0480 (0.4000)	0.0445 (0.3708)	0.0480 (0.4000*)
π_{14} (δ_{14})	0.1500 (0.9375)	0.1405 (0.8781)	0.1360 (0.8500)	0.1315 (0.8219)	0.1200 (0.7500)
π_{23} (δ_{23})	0.1500 (0.6250)	0.1405 (0.5854)	0.1360 (0.5667)	0.1315 (0.5479)	0.1200 (0.5000)
π_{24} (δ_{24})	0.2500 (0.7813)	0.2515 (0.7859)	0.2480 (0.7750)	0.2445 (0.7641)	0.2480 (0.7750)
π_{34} (δ_{34})	0.4000 (0.8333)	0.4080 (0.8500)	0.4160 (0.8667)	0.4240 (0.8833)	0.4320 (0.9000)

Note. $*$: minimum δ_{ij}

As described in the section 3, since we may prefer a sampling design giving *MINMAX* δ_{ij} in the sense that it would provide the more stable variance estimator, the maximum of available values of c should be used. For this problem, $c = 0.454$ is the largest value that still yields a NLP solution. Table 5.5 shows the solutions of the three approaches when $c = 0.454$ is applied. Although it seems that those from Approach I and Approaches II and III are different each other, especially for the *MINMAX* δ_{ij}, the solutions are nearly identical.

Table 5.5 π_{ij} and δ_{ij} from Three Approaches When Using $c = 0.454$

	AI	AII	AIII
π_{12} (δ_{12})	0.0365 (0.4563)	0.0363 (0.4538**)	0.0363 (0.4538**)
π_{13} (δ_{13})	0.0545 (0.4542**)	0.0545 (0.4542)	0.0545 (0.4542)
π_{14} (δ_{14})	0.1090 (0.6813)	0.1092 (0.6825)	0.1092 (0.6825)
π_{23} (δ_{23})	0.1090 (0.4542**)	0.1092 (0.4550)	0.1092 (0.4550)
π_{24} (δ_{24})	0.2545 (0.7953)	0.2545 (0.7953)	0.2545 (0.7953)
π_{34} (δ_{34})	0.4365 (0.9094)	0.4363 (0.9090)	0.4363 (0.9090)

Note. AI: Approach I, AII: Approach II, AIII: Approach III
$**$: MINMAX δ_{ij}

We do not deal here with the application of the strict constraints that always provide the smaller variance estimator than in PPS sampling with replacement, Murthy's method and Brewer's method because in this case the maximum of $c = 0.454$ is less than $c_{PPS} = 0.5$ in (4.6). So without using those constraints we compare our suggested approaches to those three methods.

Table 5.6 shows a comparison of the π_{ij} and the δ_{ij} of five IPPS sampling schemes, that is, Jessen's two methods, the two sampling plans from Nigam et al.'s method and Brewer's method. The results in the methods are quite different from those in Table 5.5. And since each minimum δ_{ij} for these comparative methods is smaller than in Approaches I, II and III, the NLP approaches are expected to yield the more stable variance estimator. Note that Jessen's Method 2 even provides zero π_{ij} and the non-negativity of variance estimator is not guaranteed because some values of δ_{ij} are larger than one.

Table 5.6 π_{ij} and δ_{ij} from Five Selected Methods

	J2	J3	N2	N3	B
π_{12} (δ_{12})	0.2000 (2.5000)	0.0666 (0.8325)	0.0500 (0.6250)	0.0400 (0.5000)	0.0277 (0.3465*)
π_{13} (δ_{13})	0.0000 (0.0000*)	0.0667 (0.5558)	0.0500 (0.4167*)	0.0600 (0.5000)	0.0535 (0.4455)
π_{14} (δ_{14})	0.0000 (0.0000*)	0.0667 (0.4169)	0.1000 (0.6250)	0.1000 (0.6250)	0.1188 (0.7426)
π_{23} (δ_{23})	0.0000 (0.0000*)	0.0666 (0.2775*)	0.1000 (0.4167*)	0.1000 (0.4167*)	0.1188 (0.4950)
π_{24} (δ_{24})	0.2000 (0.6250)	0.2667 (0.8334)	0.2500 (0.7813)	0.2600 (0.8125)	0.2535 (0.7921)
π_{34} (δ_{34})	0.6000 (1.2500)	0.4667 (0.9723)	0.4500 (0.9375)	0.4400 (0.9167)	0.4277 (0.8911)

Note. J2: Jessen (1969)'s method 2
J3: Jessen (1969)'s method 3
N2:Nigam et al. (1984)'s method (the second sampling plan)
N3:Nigam et al. (1984)'s method (the third sampling plan)
B: Brewer (1963)'s method
$*$: minimum δ_{ij}

The results for schemes J2, J3, N2 and N3 are from Jessen (1969), page 183 and Nigam et al. (1984), page 567, respectively.

In Table 5.7 we compare the stabilities using the following formula of coefficient of variation (CV) of variance estimators based on the π_{ij} in Table 5.5 and 5.6:

$$CV\left(\widehat{Var}(\hat{Y})\right) = \sqrt{Var\left[\widehat{Var}(\hat{Y})\right]} \Big/ E\left[\widehat{Var}(\hat{Y})\right]$$

$$= \sqrt{E\left[\widehat{Var}(\hat{Y})\right]^2 - \left[Var(\hat{Y})\right]^2} \Big/ Var(\hat{Y}) \qquad (4.1)$$

We added the two PPS sampling schemes of PPS sampling with replacement and Murthy's method in the table to compare with IPPS sampling methods.

Table 5.7 Comparison of Stabilities of Estimated Variances

Pop.	$CV\left(\widehat{Var}(\hat{Y})\right)$							
	PPS	AI	AII	AIII	J3	N3	B	M
A	1.600	1.244	1.242	1.242	2.127	1.349	1.160	1.015
B	1.600	1.244	1.242	1.242	2.127	1.349	1.160	1.015
C	1.000	1.481	1.483	1.483	1.473	1.392	1.623	0.713
Ave.	1.400	1.323	1.322	1.322	1.909	1.363	1.314	0.914
Rel. Sta.	100	106	106	106	73	103	107	153

Note. PPS: probability proportional to size sampling with Replacement
M: Murthy (1957)'s method
Ave.: Average of the CV for the three populations
Rel. Sta.: Relative Stability

See the notes in Table 5.5 and 5.6 for the others.

When comparing the relative stability, Murthy's method is best, while Jessen's Method 3 is considerably less efficient. Our NLP approaches and Brewer's method are almost similar and better than Nigam et al.'s method and PPS sampling method.

We have not shown the CV in terms of $c = 0.000$ to $c = 0.400$ in NLP approaches. Some of them yield better stability of variance estimators than when $c = 0.454$, but we may prefer using the maximum ($c = 0.454$) because we would be mainly interested in the reduction of the variation for the non-squared factor in (2.7) at the beginning of sample design.

Also, there may exist a tradeoff between the CV of variance estimator and variance of the H-T estimator in NLP approaches. It comes from the fact that a change of the objective function by a change of the value of c yields a change in the π_{ij}, and the increase (decrease) of the variance $Var(\hat{Y})$ in (4.1) by a change of the π_{ij} results in the decrease (increase) of the CV in (4.1). Note that the objective function is an increasing function of the value of c

Table 5.8 presents a comparison of variances used in the calculation of the CV of the variance estimator in Table 5.7. Brewer's method is the best. The NLP approaches are more efficient than Murthy's method and Nigam et al.'s method. The PPS sampling is the worst.

Table 5.8 Comparison of Variances

Pop.	$Var(\hat{Y})$							
	PPS	AI	AII	AIII	J3	N3	B	M
A	0.500	0.300	0.300	0.300	0.367	0.310	0.282	0.312
B	0.500	0.300	0.300	0.300	0.367	0.310	0.282	0.312
C	0.125	0.055	0.055	0.055	0.033	0.050	0.059	0.070
Ave.	0.375	0.218	0.218	0.218	0.256	0.223	0.208	0.232
Rel. Eff.	100	172	172	172	146	168	180	162

Note. Rel. Eff.: Relative Efficiency
See the notes in Table 5.5, Table 5.6 and Table 5.7 for the others.

In summary, the NLP approaches exactly achieve the non-negativity of the Sen-Yates-Grundy variance estimator and the sampling design having $MINMAX\ \delta_{ij}$ would provide the more stable variance estimator. As shown above, with respect to the stability of the variance estimator as well as the variance of the estimator they are better than Jessen's methods, Nigam et al.'s method and PPS sampling with replacement and comparable with Brewer's method and Murthy's method.

6. Conclusion

We have suggested three NLP approaches. They surely achieve the desirable properties such as non-negativity and stability in variance estimation. They are very flexible for $n = 2$ as well as $n > 2$ and can be easily implemented by using some publicly available software including SAS/OR.

There may be a tradeoff between the stability of the variance estimator and variance of the estimator if we would like to use the higher value of 'c' in the bounded constraints. But the variance for a sampling design having *MINMAX* δ_{ij} may be lower than in probability proportional to size sampling with replacement.

Developing a software application based on the suggested NLP approaches and SAS/OR NLP procedure and checking the efficiencies of some NLP algorithms provided in SAS/OR are highly recommended.

In future research, we will draw empirical comparisons to examine the tradeoff between the variance estimator and variance of the estimator for samples of different size and populations.

References

Brewer, K. R. W. and Hanif, M. (1983). *Sampling with unequal probabilities*, New York: Springer-Verlag.

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

Hansen, M. H. and Hurwitz, W. N. (1943). "On the theory of sampling from finite populations," *The Annals of Mathematical Statistics*, 14, 333-362.

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.

Kim, S. W., Heeringa, S. G., and Solenberger, P. W. (2003). "A probability sampling approach for variance minimization," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, 2168-2173.

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

Nigam, A. K., Kumar, P. and Gupta, V. K. (1984). "Some methods of inclusion probability proportional to size sampling," *Journal of the Royal Statistical Society, Series B*, 46, 564-571.

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

Sen, A. R. (1953). "On the estimate of variance in sampling with varying probabilities," *Journal of the Indian Society of Agricultural Statistics*, 5, 119-127.

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.