An Empirical Comparison of Efficiency between Optimization and Non-optimization Probability Sampling of Two Units from a Stratum

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: Unequal Probability Sampling, Inclusion Probabilities, IPPS Sampling, Nonlinear Programming, Linear Constraints

1. Introduction

Prior to Hansen and Hurwitz (1943) there were a number of multi-stage sample design developments involving the equal probability selection of primary sampling units (PSUs) from strata with subsampling of elements from the selected PSUs. These authors first suggested unequal probability selection of PSUs with probability proportional to size (PPS).

With regard to the number of PSUs, n, to be selected from each primary stage stratum in multistage sampling designs, "two-per-stratum" or "paired-selection" designs (Kish, 1965) have many applications and advantages in variance estimation for sample estimates.

Since the early 1940s, survey statisticians have developed a large number of sampling procedures that can be used to select n=2 PSUs from each stratum including PPS with replacement (PPSWR) sampling and 50 PPS without replacement (PPSWOR) sampling procedures that Brewer and Hanif (1983) described in detail. It is a fact that among these many alternatives only a very few PPSWOR sampling procedures have been recommended in the published literature and broadly used in practice. We may choose Brewer's (1963) method, Hanurav's (1967) method and Murthy's (1957) method. These approaches are called inclusion probability proportional to size (IPPS or πPS) sampling methods.

Jessen (1969), Rao and Bayless (1969) and Cochran (1977) empirically studied the efficiencies of the variances or variance estimators for different unequal probability sampling methods for paired primary stage selections. Rao and Bayless (1969) extended this work, covering more sampling schemes and populations.

More recently Kim, Heeringa and Solenberger (2003, 2004) suggested nonlinear programming (NLP) approaches that are IPPS sampling methods. These methods adopt "whole sample" procedures in which the units are not individually drawn but an optimized selection probability is specified for each possible total sample configuration by using the nonlinear programming algorithms. These alternatives may be called optimization sampling methods, while others that use algorithms that do not focus on the selection probability of each complete sample may be called non-optimization sampling methods.

In this paper, we begin by describing the principle of IPPS sampling. Second, we discuss the desirable features of IPPS sampling methods. Third, we review the NLP approaches we developed in 2003 and 2004. Those are easy to implement by using some publicly available software such as SAS/OR® (SAS, 2001). Fourth, we suggest some conditions to improve efficiency relative to certain existing sampling strategies and given these conditions, we show when it is feasible for NLP approaches to improve on existing methods. Finally, we examine the efficiency of the variances and variance estimators for the optimization and non-optimization sampling methods using the natural populations given in Rao and Bayless (1969).

2. Principle of IPPS Sampling

Consider a finite population U consisting of N distinct and identifiable units $u_1, u_2, \cdots, u_N$. Given a sampling scheme to select a specified sample, s , of n units in the population U , define a function $p(s)$, which indicates the selection probability of the sample. Then we call $p(\cdot)$ a sampling design or sampling plan.

The probability that the i th unit, u_i , is included in a sample of size n under a certain sampling design , the so-called the first-order inclusion probability, π_i is defined by

$$\pi_i = \sum_{i \in s, s \in S} p(s) , \qquad (2.1)$$

where S is the collection of all possible samples from U .

Similarly, the probability of selecting pairs of units $\left(u_i, u_j\right)$ over all samples, the second-order inclusion probability, is defined by

$$\pi_{ij} = \sum_{i,j \in s, s \in S} p(s) \ . \qquad (2.2)$$

Note that in the case of $n = 2$ PSUs per stratum, this simplifies to:

$$\pi_{ij} = p(s) \, . \qquad (2.3)$$

A common focus in much of sampling theory is estimation of the population total $Y = \sum_{i=1}^{N} y_i$ based on a sample, where y_i is the value of the characteristic of interest for element, u_i . Horvitz and Thompson (1952) considered three subclasses of linear estimators of the population total when sampling without replacement. One of them is denoted by

$$\hat{Y} = \sum_{i=1}^{n} \alpha_i y_i \, , \qquad (2.4)$$

where α_i is a fixed weight for an element, u_i , selected from the population.

Also, they suggested the linear estimator, often called the H-T estimator of the population total and given by

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \, . \qquad (2.5)$$

As proved by Godambe (1955), the best linear estimator, the unbiased estimator having minimum variance, does not exist uniquely for the entire class of the known linear estimators and H-T estimator is the best linear estimator within the subclass (2.4).

By using the usual definition of the variance of the estimator, that is, $Var\left(Y_{HT}\right) = E\left(Y_{HT} - E(Y_{HT})\right)^2$, the variance of the H-T estimator is obtained as follows:

$$Var_{HT}\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N} \frac{y_i^2 (1-\pi_i)}{\pi_i} + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{y_i y_j (\pi_i \pi_j - \pi_{ij})}{\pi_i \pi_j} \, , \qquad (2.6)$$

where $\pi_i > 0$ for all i .

Then an unbiased estimator from a sample s of (2.6) is given by

$$Var_{s,HT}\left(Y_{HT}\right) = \sum_{i=1}^{n} \frac{y_i^2 (1-\pi_i)}{\pi_i^2} + 2 \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{y_i y_j (\pi_i \pi_j - \pi_{ij})}{\pi_i \pi_j \pi_{ij}} \qquad (2.7)$$

where $\pi_i > 0$ and $\pi_{ij} > 0$ for all i and j .

Since (2.7) does not vanish even if all y_i/π_i are equal and can give a negative value, Yates and Grundy (1953) proposed the well-known revised version of (2.6) and (2.7) given by

$$Var_{YG}\left(Y_{HT}\right) = \sum_{i=1}^{N} \sum_{j>i}^{N} \left(\pi_i \pi_j - \pi_{ij}\right) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \qquad (2.8)$$

and

$$Var_{s,YG}\left(Y_{HT}\right) = \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \, , \qquad (2.9)$$

respectively.

Note that if $\pi_i = n\, y_i/Y$, in other words, if the first-order inclusion probability π_i is exactly proportional to the y_i , then the H-T estimator (2.5) has zero variance and the sampling design $p(\cdot)$ will be best.

But since the y_i are usually unknown in practice, this design is purely theoretical. Instead, we may assume that an auxiliary variable x_i is available, whose value is known for each unit in the population and closely correlated with the y_i . Also, we may assume that the information on any other supplementary variables for the units is lacking and there is no advanced knowledge of the values of the y_i .

The x_i can then serve as the size measure of the i th unit. Then $p_i = x_i / X$, where $X = \sum_{j=1}^{N} x_j$, represents the relative measure of size for u_i and accordingly

$$\pi_i = np_i \, . \qquad (2.10)$$

The sampling designs satisfying (2.10) (or the π_i are proportional to the x_i) are called IPPS sampling designs.

To reduce the variance or variance estimator of the H-T estimator in these situations, rather than addressing the squared terms $\left(y_i/\pi_i - y_j/\pi_j\right)^2$ in (2.8) or (2.9) we may focus on the terms:

$$\pi_i \pi_j - \pi_{ij} \, , \; j > i = 1, \cdots, N \, . \qquad (2.11)$$

Note that those terms often vary widely and are highly dependent on the sampling schemes used. As a result, an IPPS sampling strategy that minimizes values of these non-squared terms (2.11) is essential.

3. Desirable Requirements in IPPS Sampling

Based on (2.8) and (2.9), we can identify the following properties of an efficient IPPS sampling method:

(i)The π_i are strictly proportional to the x_i ;

(ii) The sample size n is fixed;

(iii) $\pi_{ij} > 0$ for all i, j ($i \neq j$);

(iv) $\pi_i \pi_j - \pi_{ij} > 0$ for all i, j ($i \neq j$);

(v) $\pi_{ij} \big/ \pi_i \pi_j > c$ for all i, j ($i \neq j$), where the value of c is positive and as far from zero as possible.

Of the above, (i) is fundamental for the IPPS sampling designs. (ii) is preferred by survey samplers and (iii) is a condition required for unbiased variance estimation. Also, (iv) and (v) are needed for non-negativity and the stability of the variance estimator (2.8), respectively. We may say that (i), (ii), (iii) and (iv) are crucial conditions and the last one (v) is optional, but important for efficiency.

4. Review of NLP Approaches

The construction of IPPS sampling schemes that satisfy all five of these conditions is not a simple matter. In 2003 and 2004 (Kim, et al., 2003, 2004) we suggested NLP IPPS approaches that meet these requirements and are easy to implement in selecting two primary units per stratum. Those approaches were developed based on the following concepts.

When constructing a sampling scheme, our main objective is that it will lead to the following features: (a) a smaller variance, (b) a smaller variance estimator, (c) a non-negative and stable variance estimator.

Jessen (1969) proposed four interesting sampling schemes. One of them, an IPPS sampling method called Method 4, was designed to reduce the variance (2.8). It shows high efficiency in comparisons of variances of estimators but it is difficult to employ in practice because of the arbitrary and complex sequence of manual trials needed to determine the second-order inclusion probabilities. Also, it is limited to the samples of size $n = 2$.

Hanurav (1967) and Nigam, Kumar and Gupta (1984) suggested alternative IPPS sampling schemes that provide a non-negative and stable variance estimator. The method of Hanurav (1967) is available for only $n = 2$. The method proposed by Nigam, et al. uses binary block designs in an experimental design approach. Their method involves considerable trial and error to carry out even for cases where the sample size is small.

Note that these approaches have attempted to address at most one of the objectives that is, (a) or (c), and the second one (b) has not even been tried. It may reflect that fact that focusing on two or more of these features would make the sampling scheme too complicated or require trade-offs between them, resulting in a poor result for each.

In fact it seems that addressing two or three of these features simultaneously might be almost impossible. But for this purpose we can use NLP, which is available in some software such as SAS/OR®.

The idea for some parts of the following three NLP approaches stems from Jessen's (1969) method. They address multiple features in the set (a)-(c) and result in optimized sample designs. We now show how those features can be achieved simultaneously using NLP.

Consider a NLP problem specified as :

$$\textit{Minimize} \text{ (or } \textit{Maximize}\text{) } f(\underline{z}) \tag{4.1}$$

$$\text{subject to } g_t(\underline{z}) \geq 0 \text{ for all } t = 1, \cdots, k, \tag{4.2}$$

$$h_t(\underline{z}) = 0 \text{ for all } t = k+1, \cdots, l, \tag{4.3}$$

where $\underline{z}$ is a vector of m decision variables, { $z_1, z_2, \cdots, z_m$ }, and the function f is called the objective function with inequality constraints (4.2) and equality constraints (4.3).

The optimum solution to the NLP problem is a point $\underline{z}^*$ such that $f(\underline{z}) \geq f(\underline{z}^*)$ (or $f(\underline{z}) \leq f(\underline{z}^*)$) for each feasible point $\underline{z}$.

Now we start with the designation of sampling schemes for achieving both (a) and (c).

Jessen's (1969) method 4 for $n = 2$ is an IPPS sampling with:

$$W_{ij} \approx \overline{W}, \tag{4.4}$$

where $W_{ij} = \pi_i \pi_j - \pi_{ij}$ in (2.11) and a constant $\overline{W}$ is the average of all possible W_{ij} s:

$$\overline{W} = \left(n - \sum_{i=1}^{N} \pi_i^2 \right) \Big/ N(N-1). \tag{4.5}$$

Note that each trial of method 4 yields a different W_{ij}, and inappropriately approximates $\overline{W}$ because the π_{ij} are manually adjusted to achieve the following relation:

$$\sum_{j(\neq i)}^{N} \pi_{ij} = \pi_i, \quad i = 1, \cdots, N. \tag{4.6}$$

We may consider the first alternative NLP approach below which optimizes the π_{ij}.

Designate the NLP problem for $n \geq 2$:

Minimize

$$\sum_{i=1}^{N} \sum_{j>i}^{N} \left(W_{ij} - \overline{W} \right)^2 = \sum_{i=1}^{N} \sum_{j>i}^{N} \left(\left(\pi_i \pi_j - \pi_{ij} \right) - \overline{W} \right)^2, \tag{4.7}$$

subject to the linear inequality constraints,

$$c_{NLP} \pi_i \pi_j < \pi_{ij} \leq \pi_i \pi_j, \quad j > i = 1, \cdots, N, \tag{4.8}$$

where c_{NLP} is a real number between 0 and 1,

and the linear equality constraints,

$$\sum_{j\neq i}^{N}\pi_{ij}=(n-\mathbf{1})\pi_i\ ,\ i=1,\cdots,N\ . \qquad (4.9)$$

The variables $z_1,z_2,\cdots,z_n$ in the NLP problem specification are the second-order inclusion probabilities, π_{ij} s, and (4.1), (4.2) and (4.3) correspond to (4.7), (4.8) and (4.9), respectively. The constraints (4.8) are the combined expression of (iii), (iv) and (v) among the desirable requirements in IPPS sampling.

Note that since $\overline{W}$ is a constant and $\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij}=n(n-1)/2$, the minimization of the objective function (4.7) is equivalent to

$$Minimize\ \sum_{i=\mathbf{1}}^{N}\sum_{j>i}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)^{\mathbf{2}}, \qquad (4.10)$$

or

$$Minimize\ \sum_{i=\mathbf{1}}^{N}\sum_{j>i}^{N}\left(\pi_{ij}^{\mathbf{2}}-\mathbf{2}\pi_i\pi_j\pi_{ij}\right), \qquad (4.11)$$

which does not depend on $\overline{W}$.

Note that (4.10) is related to (2.11). This NLP method to find the optimum π_{ij} s by minimizing (4.7) or (4.10) or (4.11) under the linear constraints (4.8) and (4.9) will be called NLP I.

From (4.7) we can also introduce the objective function

$$\sum_{i=\mathbf{1}}^{N}\sum_{j>i}^{N}\left(W_{ij}-\overline{W}\right), \qquad (4.12)$$

which amounts to

$$\sum_{i=\mathbf{1}}^{N}\sum_{j>i}^{N}\pi_i\pi_j-\sum_{i=\mathbf{1}}^{N}\sum_{j>i}^{N}\pi_{ij}\ . \qquad (4.13)$$

Since the first term in (4.13) is a fixed value, minimization of (4.13) reduces to maximization of

$$\sum_{i=\mathbf{1}}^{N}\sum_{j>i}^{N}\pi_{ij}\ . \qquad (4.14)$$

In this case where all possible second-order inclusion probabilities are enumerated, since (4.14) is equal to $n(n-1)/2$, as mentioned above, the NLP problem does not maximize (4.14) but simply finds solutions, π_{ij} s, meeting the constraints. We will call this approach NLP II.

Next we introduce NLP III for constructing a sampling design that has all features (a), (b) and (c).

In order to obtain the sampling design providing the smaller variance estimators, we consider the problem:

$$Minimize\ \sum_{s\in S}Var_{s,SYG}\left(Y_{HT}\right). \qquad (4.15)$$

In theory the hypothetical situation where the non-squared factor in (2.9) is equal to a constant can be considered to succeed in solving the problem (4.15). In other words, if

$$(\pi_i\pi_j-\pi_{ij})/\pi_{ij}=(\pi_i\pi_j/\pi_{ij})-\mathbf{1}\approx\Pi-\mathbf{1}, \qquad (4.16)$$

where Π is a constant, preferably $\Pi\approx 1$,
(4.15) reduces to

$$\Pi^*\sum_{s\in S}\left[\sum_{i=\mathbf{1}}^{n}\sum_{j>i}^{n}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^{\mathbf{2}}\right], \qquad (4.17)$$

where $\Pi^*=\Pi-\mathbf{1}$.

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 (4.15) we may consider to use

$$Minimize\ \sum_{s\in S}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}-\Pi^*\right), \qquad (4.18)$$

which is equivalent to

$$Minimize\left[\frac{(N-\mathbf{2})!}{(n-\mathbf{2})!(N-n)!}\sum_{i}^{N}\sum_{j>i}^{N}\left(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}-\Pi^*\right)\right]. \qquad (4.19)$$

Note that (4.19) amounts to

$$Minimize\ \sum_{i}^{N}\sum_{j>i}^{N}\frac{\pi_i\pi_j}{\pi_{ij}}\ . \qquad (4.20)$$

The solution to the NLP III problem achieving (4.20) under the same constraints used for NLP I or NLP II would be successful for (b) and (c). In addition, since the weighted mean of $Var_{s,SYG}\left(Y_{HT}\right)$ is equal to (2.8), that is,

$$E\left[Var_{s,SYG}\left(Y_{HT}\right)\right]=\sum_{s\in S}\left[Var_{s,SYG}\left(Y_{HT}\right)\right]\ p(s)$$
$$=Var_{SYG}\left(Y_{HT}\right), \qquad (4.21)$$

the minimization of the variance estimators is directly related to the minimization of the variance. Thus (a) can be also achieved.

To implement NLP I, NLP II and NLP III, we use the SAS/OR NLP procedure, which optimizes nonlinear (or linear) objective functions under the linear constraints. A variety of NLP algorithms are available and SAS/OR (2001) provides the dcoumentation to choose an optimization algorithm.

5.Criteria to Improve Efficiency Relative to Existing Sampling Strategies

With respect to the variances as well as variance estimators, the linear inequality constraints (4.8) in these NLP approaches have some relationship to PPSWR sampling, Brewer's (1963) method, Hanurav's (1967) method and Murthy's (1957) method. These relationships can be summarized as a simple form involving a specific value of c_{NLP} in (4.8) needed to achieve better efficiency relative to those methods. The relationships can be proved through the following theorems for $n=2$.

Theorem 5.1

The variance of the H-T estimator in NLP approaches is smaller than that for the estimator $Y_{PPS} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}$ in PPSWR sampling if for i, j,

$$\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}, \tag{5.1}$$

where $\pi_{ij,NLP}$ indicates the second-order inclusion probabilities in NLP approaches.

Proof.

If $n=2$,

$$Var_{YG}\left(Y_{HT}\right) = \sum_{i=1}^{N}\sum_{j>i}^{N} p_i p_j\left(1-\frac{\pi_{ij,NLP}}{4p_i p_j}\right)\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2 \tag{5.2}$$

and

$$Var\left(Y_{PPS}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j>i}^{N} p_i p_j\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2. \tag{5.3}$$

If $1-\frac{\pi_{ij,NLP}}{4p_i p_j} < \frac{1}{2}$ for all i, j, which reduces to $\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}$, then

$$Var_{YG}\left(Y_{HT}\right) < Var\left(Y_{PPS}\right). \tag{5.4}$$

Theorem 5.2

In NLP approaches the smaller variance estimator of the H-T estimator compared to the estimator Y_{PPS} in PPSWR sampling is achieved if for i, j,

$$\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}. \tag{5.5}$$

Proof.

For $n=2$,

$$Var_{s,YG}\left(Y_{HT}\right) = \left(\frac{p_i p_j}{\pi_{ij,NLP}}-\frac{1}{4}\right)\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2 \tag{5.6}$$

and

$$Var\left(Y_{PPS}\right) = \frac{1}{4}\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2. \tag{5.7}$$

If $\frac{p_i p_j}{\pi_{ij,NLP}}-\frac{1}{4} < \frac{1}{4}$ for all i, j, which is equal to $\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}$, then

$$Var_{SYG}\left(Y_{HT}\right) < Var\left(Y_{PPS}\right). \tag{5.8}$$

Similarly, we can show the following theorems for the methods of Brewer and Hanurav that are IPPS sampling procedures.

Theorem 5.3

The admissible condition for the second-order inclusion probabilities to obtain the variance or variance estimator of the H-T estimator in NLP approaches that is comparable to that for Brewer's method is:

$$\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}, \tag{5.9}$$

for all i, j.

Proof.

$\pi_{ij,B}$, the second-order inclusion probabilities in Brewer's method, is denoted by

$$\pi_{ij,B} = \frac{2p_i p_j}{\frac{1}{2}\left(1+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right)}\frac{1-p_i-p_j}{\left(1-2p_i\right)\left(1-2p_j\right)}. \tag{5.10}$$

Also,

$$\pi_{ij,B} = \frac{1}{2}\pi_i\pi_j K, \tag{5.11}$$

where $K = \frac{1}{\left(1+\sum_{i=1}^{N}\frac{p_i}{1-2p_i}\right)}\left(\frac{1}{\left(1-2p_i\right)}+\frac{1}{\left(1-2p_j\right)}\right).$

(5.12)

Since $K>1$ or $K<1$,

$$\pi_{ij,B} > \frac{1}{2}\pi_i\pi_j \tag{5.13}$$

or

$$\pi_{ij,B} < \frac{1}{2}\pi_i\pi_j. \tag{5.14}$$

Hence it is clear from (5.13) and (5.14) that the admissible condition is (5.9).

Theorem 5.4

The admissible condition for the second-order inclusion probabilities to obtain the variance or variance estimator of the H-T estimator in NLP approaches that is at least comparable to that for Hanurav's method is:

$$\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}\,, \qquad (5.15)$$

for all i, j.

Proof.

Let $\pi_{ij,H}$ indicate the second-order inclusion probabilities in Hanurav's method. These are given on page 384, Hanurav (1967) and

$$\pi_{ij,H} = \frac{1}{2}\pi_i\pi_j\alpha\,, \qquad (5.16)$$

$$\text{where } \alpha = \frac{\left[1-\frac{1}{2}\beta\Big/\left(1-p_{(N)}\right)\right]^2}{(1-\beta)}\,, \qquad (5.17)$$

$$\beta = \frac{2(1-p_{(N)})(p_{(N)}-p_{(N-1)})}{1-p_{(N)}-p_{(N-1)}}\,, \qquad (5.18)$$

and $p_{(N)}$ is the largest p_i and $p_{(N-1)}$ is the second largest p_i.

Since $\alpha < 1$,

$$\frac{1}{2}\pi_i\pi_j < \pi_{ij,H}\,. \qquad (5.19)$$

To achieve the given purpose, (5.15) is required from (5.19).

Theorem 5.5

The admissible condition for the second-order inclusion probabilities that the H-T estimator in NLP approaches has at least comparable variance with the estimator $Y_M = \dfrac{\sum_{i=1}^{n} p(s|i)y_i}{p(s)}$ in Murthy's method is

$$\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}\,, \qquad (5.20)$$

for all i, j.

Proof.

For $n = 2$, the variance of the H-T estimator in NLP approaches is given by (5.2) and the variance of Y_M is expressed as

$$Var\left(Y_M\right) = \sum_{i=1}^{N}\sum_{j>i}^{N} p_i p_j \frac{1-p_i-p_j}{2-p_i-p_j}\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2 . \qquad (5.21)$$

See pages 263-264, Cochran (1977).

Comparing (5.2) and (5.21), if for all i, j,

$$1-\frac{\pi_{ij,NLP}}{4p_ip_j} < \frac{1-p_i-p_j}{2-p_i-p_j}\,, \qquad (5.22)$$

(5.2) is smaller than (5.21).

Further (5.22) becomes

$$c_M\pi_i\pi_j < \pi_{ij,NLP}\,, \qquad (5.23)$$

where $c_M = \dfrac{1}{2-p_i-p_j}$.

But since $\frac{1}{2} < c_M < 1$, (5.23) gives (5.20).

Theorem 5.6

The admissible condition for the second-order inclusion probabilities that the H-T estimator in NLP approaches has at least comparable variance estimator with the estimator Y_M in Murthy's method is

$$\frac{1}{2}\pi_i\pi_j < \pi_{ij,NLP}\,, \qquad (5.24)$$

for all i, j.

Proof.

For $n = 2$, the variance estimator of the H-T estimator in NLP approaches is given by (5.6) and the variance estimator of Y_M is denoted by

$$Var\left(Y_M\right) = \frac{(1-p_i)(1-p_j)(1-p_i-p_j)}{(2-p_i-p_j)^2}\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2 . \qquad (5.25)$$

See page 264, Cochran (1977).

From (5.6) and (5.25), if for all i, j,

$$\frac{1}{c_{s,M}+1}\pi_i\pi_j < \pi_{ij,NLP}\,, \qquad (5.26)$$

$$\text{where } c_{s,M} = \frac{4(1-p_i)(1-p_j)(1-p_i-p_j)}{(2-p_i-p_j)^2}\,, \qquad (5.27)$$

(5.6) is smaller than (5.25).

But since $\frac{1}{2} < \frac{1}{c_{s,M}+1} < 1$, the admissible condition (5.24) is given from (5.26).

Corollary 5.1

The common condition for the second-order inclusion probabilities that the NLP approaches are not to be inferior to PPSWR sampling, Brewer's method, Hanurav's method and Murthy's method is

$$c_{NLP}\pi_i\pi_j < \pi_{ij,NLP}\,, \text{ for all } i, j\,, \qquad (5.27)$$

where $c_{NLP} = 0.5$.

Proof.

This is clear from (5.1), (5.5), (5.9), (5.15), (5.20) and (5.24).

Remark 5.1

The linear inequality constraints (4.8) in the NLP approaches are not only the combined condition of (iii), (iv) and (v) among the desirable requirements in

IPPS sampling but also the admissible condition not to be less efficient than other methods.

6. Feasibility of NLP approaches

Though it is clear that the condition (5.27) is theoretically desirable in using NLP approaches, the feasibility of this condition should be verified. We examined NLP approaches for the selected 16 natural populations in Rao and Bayless (1969). Table 6.1 shows the comparison of δ s, the minimum values of $\pi_{ij}/\pi_i\pi_j$ in three IPPS methods: the NLP approaches, Brewer's method and Hanurav's method.

Table 6.1 Comparison of δ s for 3 IPPS Methods

No.	*CV*	N	B	H
1	0.14	0.54*	0.53	0.54*
2	0.17	0.54*	0.53	0.53
3	0.30	0.51*	0.50	0.51*
4	0.40	0.51*	0.49	0.50
5	0.43	0.52*	0.49	0.50
6	0.44	0.50*	0.47	0.50*
7	0.46	0.51*	0.48	0.50
8	0.50	0.51*	0.48	0.50
9	0.52	0.50*	0.48	0.50*
10	0.59	0.51*	0.45	0.49
11	0.65	0.51*	0.47	0.49
12	0.65	0.52*	0.44	0.47
13	0.71	0.47	0.49	0.50*
14	0.91	0.51*	0.45	0.49
15	0.93	0.50*	0.39	0.48
16	0.98	0.51*	0.43	0.50

Note. *CV* : Coefficient of variation for auxiliary variable x_i
N: NLP I, NLP II and NLP III
B: Brewer's method
H: Hanurav's method
$*$: The largest value
δ s for B and H are from Rao and Bayless (1969).

Note that except for the 13th population, the δ s in NLP approaches are equal to 0.5 or larger than 0.5. This indicates that $c_{NLP} = 0.5$ in (5.27) is mostly feasible in practice. Also, note that δ s in NLP approaches are mostly larger than the ones in Brewer's method and mostly equal to or larger than the ones in Hanurav's method. As a result NLP approaches can be successful in achieving the smaller values of the non-squared terms (2.11) compared to other IPPS methods.

7. Comparison of Efficiency

Following the feasibility study on NLP approaches above, this section first provides an empirical comparison of variances measured by the percent gain in variance over Brewer's method and given by

$$\left[\left(Variance\ \ for\,Brewer's\ \ method\right)/Variance\right)\right)-1\right]\times 100\cdot \tag{7.1}$$

Table 7.1 Comparison of variances

No.	*CV*	N1	N2	N3	H	M	R	P
1	0.14	+0	+0	+0	+0	1	1	–10
2	0.17	–0	–0	–0	–0	+0	+0	–11
3	0.30	+0	+0	+0	–0	+0	+0	–5
4	0.40	+0	+0	+0	+0	–1	–2	–7
5	0.43	–1	–1	–0	–0	1	1	–7
6	0.44	+0	+0	+0	+0	–0	–2	–7
7	0.46	+0	+0	+0	+0	+0	–0	–6
8	0.50	–0	+0	–0	–0	1	–0	–5
9	0.52	+0	+0	+0	+0	–0	–2	–8
10	0.59	–0	–0	3	1	–2	–6	–17
11	0.65	1	+0	2	+0	–0	–3	–10
12	0.65	–1	–1	–0	–1	6	5	–7
13	0.71	–1	–0	–1	–0	+0	–1	–4
14	0.91	–1	–1	–1	–0	4	3	–3
15	0.93	1	1	2	1	7	3	–9
16	0.98	+0	+0	–1	+0	6	4	–3

Note. N1: NLP I, N2: NLP II, N3: NLP III, H: Hanurav's method, M: Murthy's method, R: method of Rao, Hartley and Cochran, P: PPSWR
See Table 6.1 for CVs.

As shown in Table 7.1, among the three NLP approaches, the third is slightly better. The third NLP approach is slightly better than non-optimization methods such as those of Brewer, Hanurav and Rao, Hartley and Cochran (1962) and compares favorably with Murthy's method when the coefficient of variation for auxiliary variable is smaller.

Table 7.2 shows comparison of the percent gain in variance estimator over Brewer's method and given by

$$\left[\left(\frac{CV^2\ \left(variance\ estimator\ in\ Brewer's\ \ method\right)}{CV^2\left(variance\ estimator\right)}\right)-1\right]\times 100 \tag{7.2}$$

Table 7.2 indicates that Murthy's method and the Rao, Hartley and Cochran's method are best. Among the three NLP approaches, the third is slightly better. Also, the third is slightly better than Brewer's method and compares favorably with Hanurav's method.

Table 7.2 Comparison of variance estimators

No.	*CV*	N1	N2	N3	H	M	R	P
1	0.14	+0	+0	+0	+0	3	5	–12
2	0.17	–1	–1	–1	–0	3	7	–12
3	0.30	+0	+0	1	+0	1	2	–5
4	0.40	2	1	1	1	–3	–5	–13
5	0.43	–2	–2	–0	–0	6	10	–6
6	0.44	–0	–0	–0	–0	1	1	–9
7	0.46	–0	–1	+0	–0	5	9	+0
8	0.50	–2	–1	–1	–0	4	8	–3
9	0.52	–2	–1	–2	–1	4	9	–3
10	0.59	–7	–5	–12	2	13	20	–6
11	0.65	–4	–3	–2	–2	4	5	–9
12	0.65	5	5	4	–1	16	27	3
13	0.71	+0	–1	2	1	2	4	–3
14	0.91	1	1	1	2	8	15	2
15	0.93	27	23	27	17	38	75	36
16	0.98	16	15	17	15	22	39	19

Note. See Table 7.1

8. Discussion

Under the principle of IPPS sampling, the NLP approaches exactly satisfy the desirable requirements. We proved that there exists an appropriate condition in NLP approaches in order for the approach not to be less efficient than other alternatives. It seems that this condition will hold in most empirical problems.

The optimization sampling method using NLP also appears to be better than other IPPS methods such as the ones of Brewer and Hanurav with respect to efficiencies. The optimization method would be more efficient when the strata have smaller coefficients of variation in measures of size for the primary sampling units. Although Murthy's method is not inferior, NLP approaches may be preferred to his method since these give self-weighting.

In our continuing research on NLP methods we will deal with the cases where the sample size is greater than two PSUs per stratum.

References

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

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.

Godambe, V. P. (1955). "A unified theory of sampling from finite populations," *Journal of the Royal Statistical Society, Series B*, 17, 269-278.

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

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.

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.

Kim, S. W., Heeringa, S. G., and Solenberger, P. W. (2004). "Inclusion Probability Proportional to Size Sampling: A Nonlinear Programming Approach to Ensure a Nonnegative and Stable Variance Estimator," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3821-3828.

Kish, L. (1965). *Survey Sampling*, New York: John Wiley and Sons.

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.

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.

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