Hacettepe Journal of Mathematics and Statistics
Volume 44 (1) (2015), 215 – 228

An optimal controlled selection procedure for sample coordination problem using linear programming and distance function

Neeraj Tiwari* and Akhil Chilwal†

Abstract

A number of procedures have been developed for maximizing and minimizing the overlap of sampling units in different/repeated surveys. The concepts of controlled selection, transportation theory and controlled rounding have been used to solve the sample co-ordination problem. In this article, we proposed a procedure for sample co-ordination problem using linear programming with the concept of distance function that facilitates variance estimation using the Horvitz-Thompson estimator. The proposed procedure can be applied to any two-sample surveys having identical universe and stratification. Some examples have been considered to demonstrate the utility of the proposed procedure in comparison to the existing procedures.

2000 AMS Classification: 62D05

Keywords: Controlled Sampling, Distance Function, Sample Coordination, Variance Estimation.

Received 21/06/2013 : Accepted 16/02/2014 Doi : 10.15672/HJMS.201477458

1. Introduction

In practical life, we face situations where the same population is sampled in various surveys so as to obtain information on variety of characters or to obtain current estimates of a characteristic of the population. There are certain applications for which samples are selected at the same time point, for two or more surveys for the same population. For example, a sample can be designed for collecting information about the education status

*Department of Statistics, Kumaun University, S.S.J. Campus, Almora-263601,Uttarakhand (INDIA)
Email: kumarn_amo@yahoo.com
†Department of Statistics, Kumaun University, S.S.J. Campus, Almora-263601,Uttarakhand (INDIA)
Email:akhil.stat@gmail.com

of the families and another sample can be designed to collect information about the income of the families for the same population. On the other hand, if we have improved data after conducting a survey, then it would be desirable to improve the stratification and measures of size. In each survey, it is possible that both the stratification and the measure of size (i.e. the selection probabilities) of the sampling units are different. With the help of updated data a redesign is attempted in which the old units remain the same but the stratification and the selection probabilities are changed. In the redesigning of a survey for the same population, the two samples must be selected sequentially since the designs pertain to different time points. Moreover, it may be considered desirable to contain as many old units as possible in the new sample so that the costs associated with hiring of new enumerators, providing training to the enumerators, etc., can be reduced. Mostly in all the surveys, the cost of sampling is roughly proportional to the total number of units sampled in surveys. Thus, if we select the same units twice instead of selecting two different units, it will reduce the cost of the survey. When the cost of the survey is limited, it is usually desirable to select the units which can be taken as a sample for both the surveys (in case of simultaneous as well as in sequential selection). It can be achieved by minimizing the number of different units in the union of the samples. This is known as the problem of maximization of overlap between the sampling units or the positive sample co-ordination problem. On the other hand, a situation also exists, where it is desirable to withdraw or minimize the likelihood of selecting the same unit in more than one survey. This kind of problem is known as minimization of overlap of sampling units or the problem of negative sample co-ordination.

The problem of co-ordination of sampling units has been a topic of interest for more than fifty years. Various procedures have been proposed by different researchers in order to solve the sample co-ordination problem. Early developments on this topic were due to Patterson (1950) and Keyfitz (1951). Raj (1956) introduced the sample co-ordination problem as a transportation problem in linear programming by considering one unit per stratum. Kish and Hess (1959), Fellegi (1963, 1966), Gray and Platek (1963) and Kish (1963) also proposed some procedures for sample co-ordination problem but these procedures were in general restricted to either two successive samples or to small sample size. To solve the problem in context of a large sample size, Kish and Scott (1971) proposed a procedure for sample co-ordination problem. Brewer et al. (1972) proposed the concept of permanent random number (PRN) for solving the sample co-ordination problem. Causey et al. (1985) proposed an optimum linear programming procedure for maximizing the expected number of sampling units which are common to the two designs, when the two sets of sample units are chosen sequentially. Ernst and Ikeda (1995) also proposed a linear programming procedure for overlap maximization under very general conditions. Ernst (1996) introduced a procedure for sample co-ordination problem, with one unit per stratum designs where the two designs may have different stratifications. Ernst (1998) also proposed a procedure for sample co-ordination problem with no restriction on the number of sample units per stratum, but the stratification must be identical. Both of these procedures proposed by Ernst used the controlled selection algorithm of Causey et al. (1985) and can be used for simultaneous as well as sequential sample surveys. Based on the procedures of Ernst (1996, 1998), Ernst and Paben (2002) introduced a new procedure for sample co-ordination problem, which has no restriction on the number of sample units selected per stratum and also does not require that the two designs have identical stratification. Deville and Tille (2000) used random partition of population to solve the sample co-ordination problem in repeated sample surveys. Matei and Tille (2006) introduced a methodology for sample co-ordination problem based on iterative proportional fitting (IPF), to compute the probability distribution of a bi-design. Their procedure can be applied to any type of sampling design for which it is possible to compute the probability distribution for the two samples. Matei and Skinner (2009) developed optimal sampling design for given unit inclusion probabilities in order to realize maximum co-ordination. Their method is based on the concept of controlled selection and some theoretical conditions on joint selection probability of two samples. Tiwari and Sud (2012) introduced a procedure for solving sample co-ordination problem using the multiple objective linear programming. There procedure is efficient but quite cumbersome in the sense that before applying the idea of nearest proportional to size design to obtain the desired controlled IPPS design they have to first obtain an appropriate uncontrolled IPPS design and then define a non-IPPS design which totally avoids the non-preferred samples to make their probabilities zero. The procedure of Tiwari and Sud (2012) can be used for situations where two surveys are conducted for the same population with identical stratification and the sample units are selected simultaneously for the two designs.

In this article, using the linear programming approach with distance function, we propose an improved method for sample co-ordination problem which maximize (or minimize) the overlap of sampling units between two designs. The proposed procedure got its inspiration from the sample co-ordination procedure of Ernst (1998). The basic concept of the proposed procedure is adopted from Ernst (1998), however the way of solving the controlled selection problem is different and made quite simple in the proposed procedure. The proposed procedure also facilitates variance estimation using Yates–Grundy (1953) form of Horvitz-Thompson (1952) variance estimator, a feature not available with the procedure of Ernst (1998). SAS 9.3 and MATLAB 10.0 windows version packages have been used to solve this problem.

In Section 2, we describe the preliminaries and notations adopted in this article. In Section 3, we discuss the proposed methodology for positive and negative sample co-ordination problem. In Section 4, some numerical examples have been considered to demonstrate the utility of the proposed procedure. Finally in Section 5, the findings of this article are summarized.

2. The Basic Notations and Preliminaries

Let us consider a two-dimensional population array A of N units, consisting of cells that have real numbers, a_{ij}, $(i = 1, \ldots, R, j = 1, \ldots, C)$. Suppose a sample of size n is to be obtained from this population. Let y be the characteristic under study, Y_{ij} the y-value for the i^{th} & j^{th} unit in the population $(i = 1, \ldots, R, j = 1, \ldots, C)$ and y_l the y-value for the l^{th} unit in the sample $(l = 1, \ldots, n)$. Let s_{ij} be each internal entry of a sample (s). Then s_{ij} equals either $[a_{ij}]$ or $[a_{ij}]+1$, where $[a_{ij}]$ is the integer part of a_{ij}. We have to consider a set of samples with selection probabilities that satisfy the constraints:

$$E(s_{ij} \mid i,j) = \sum_{i,j\in s, s\in S} \mathrm{s}_{ij} p(s) = n a_{ij} \tag{2.1}$$

and

$$\sum_{s\in S} p(s) = 1 \tag{2.2}$$

where S is the set of all possible samples $\{s\}$, and $p(s)$ is the selection probability of each sample s.

There can be many sets of probability distributions $p(s)$ satisfying Eq. (2.1) and Eq. (2.2), although only one set of probabilities can be used to obtain a solution to the sample co-ordination problem. We may consider an algorithm based on an appropriate and objective principle to find the solution that reflects the closeness of each sample s to A. For this purpose we consider the several following measures of closeness between A and s.

The ordinary distance, which is often called the Euclidean distance, given by

$$\xi_1(A,s) = [\sum_{i=1}^{R}\sum_{j=1}^{C}(a_{ij} - s_{ij})^2]^{1/2} \tag{2.3}$$

is the most common measure to define the distance between A and s, as it is easy to calculate.

Two other distance measures can also be used to define the distance between A and s. These are:

(i) Cosine Distance Function:

$$\xi_2(A,s) = 1 - \sum_{i=1}^{R}\sum_{j=1}^{C}\frac{a_{ij}s_{ij}}{\|A\|_2\|s\|_2} \tag{2.4}$$

(ii) Bray-Curtis Distance Function:

$$\xi_3(A,s) = \frac{\sum_{i=1}^{R}\sum_{j=1}^{C}(a_{ij}-s_{ij})}{\sum_{i=1}^{R}\sum_{j=1}^{C}(a_{ij}+s_{ij})} \tag{2.5}$$

We have applied these three distance functions given in Eq.(2.3), Eq.(2.4) and Eq.(2.5), to all co-ordination problems considered by us and found that the distance function ξ_2 given in Eq.(2.4) provides best result in terms of the value of the objective function. Therefore, we shall be using ξ_2 as the distance measure in this article.

Following the notations of Ernst (1998), we consider two sampling designs D_1 and D_2, with identical population and stratification, consisting N units, with S denoting one of the strata. We have to select the given number of sample units from the two designs. The selection probability of each unit in S is different for the two designs. First of all we consider the problem of maximizing the overlap of sampling units in D_1 and D_2 designs. For this purpose, the sample units are selected subject to the following conditions originally introduced by Ernst (1998):

(i) There are a predetermined number of units, n_l, selected from S for the D_l sample, $l = 1, 2$. That is, the sample size for each stratum and design combination is fixed.
(ii) The i^{th} unit in S is selected for the D_l sample with its assigned probability, denoted by $(\pi_i)_l$, $l = 1, 2$.
(iii) The expected value of the number of sample units common to the two designs is maximized.
(iv) The number of sample units common to any D_1 and D_2 samples is always within one of the maximum expected value.

In each stratum S, we applied the described procedure separately. First of all, we construct a real valued tabular array $W = (w_{ij})$, which is a two–dimensional array. Here (w_{ij}) represents the internal units of W. A tabular array is one in which the final row and column contain the marginal values (marginal values are the sum of internal values for each row and column). Array W is known as the controlled selection problem, as it specifies the probability and expected value conditions to be satisfied for the problem under consideration.

Ernst (1998) suggested that the problem of maximizing the overlap of sampling units for the two designs can be converted into the "Controlled Selection" problem $W = (w_{ij})$, where W is an $(N{+}1) \times 5$ array with N internal rows and 4 internal columns, here N is the number of units in the stratum universe. The internal units of W are computed for each internal row $i = 1, \ldots, N$ as follows:

$$w_{i3} = min[(\pi_i)_1, (\pi_i)_2] \tag{2.6}$$

$$w_{il} = (\pi_i)_l - w_{i3}, \qquad l = 1, 2 \tag{2.7}$$

$$w_{i4} = 1-\sum_{l=1}^{3} \mathrm{w}_{il} \tag{2.8}$$

$$w_{i5} = \sum_{l=1}^{4} \mathrm{w}_{il} \tag{2.9}$$

Array W can be considered as controlled selection problem. The first unit w_{i1} in the i^{th} internal row of this array denotes the probability that the i^{th} unit is in the sample D_1 but not in D_2; the second unit w_{i2} is the probability that the i^{th} unit is in the sample D_2 but not in D_1; the third unit w_{i3} is the probability that the i^{th} unit is in both the samples D_1 and D_2; and the fourth element w_{i4} is the probability that it is in neither sample. The marginals in the first four columns of the last row represents the expected number of units in the corresponding category.

The controlled selection problem W can be solved by constructing a sequence of integer valued tabular array, $M_1 = (m_{ij1})$, $M_2 = (m_{ij2})$,...., $M_u = (m_{iju})$, with the same number of rows and columns as W and associated probabilities p_1, p_2,..., p_u, which specify certain conditions. At last, a random array $M = (m_{ij})$, is then chosen among these u arrays using the indicated probability. Now we discuss the conditions which must be satisfied by this sequence of integer valued arrays. In each internal row of these arrays, one of the four internal columns has the value 1 and the other three have value 0. The value 1 in the first column indicates that the unit is only in the D_1 sample; value 1 in the second column indicates that the unit is only in the D_2 sample. Similarly, value 1 in the third column indicates that the unit is in both samples; and the value 1 in the forth column indicates that the unit is in neither sample.

Ernst (1998) has derived a set of conditions which, if met by the random array M, are sufficient to satisfy the conditions (i)-(iv). These conditions are as follows:

Condition (ii) will be satisfied if

$$p(m_{il} = 1) + p(m_{i3} = 1) = w_{il} + w_{i3} = (\pi_i)_l, \quad i = 1, ..., N, l = 1, 2 \tag{2.10}$$

Similarly, condition (*iii*) will be satisfied if

$$p(m_{i3} = 1) = w_{i3}, \quad i = 1, ..., N, \tag{2.11}$$

If it can be established that if

$$E(m_{il}) = \sum_{v=1}^{u} \mathrm{p}_v m_{ijv} = w_{ij}, \;\; i = 1, ..., N, j = 1, ..., 4 \tag{2.12}$$

then conditions (*ii*) and (*iii*) will hold, since (2.12) implies (2.10) and (2.11).

To establish condition (*i*), one only needs to show that

$$m_{(N+1)lv} + m_{(N+1)3v} = n_l, \quad l = 1, 2 \;\; v = 1, ..., u \tag{2.13}$$

Finally, to establish (*iv*), it is sufficient to show that

$$\mid m_{(N+1)3v} - w_{(N+1)3} \mid <1, \qquad v = 1, ..., u \tag{2.14}$$

here $w_{(N+1)3}$ is the maximum expected number of units which are common to the two samples and $m_{(N+1)3v}$ is the number of units common to the v^{th} possible sample.

Now the problem reduces to the solution of the controlled selection problem W in such a way as to satisfy the conditions (2.12)-(2.14). The solution of the controlled selection problem W, will then maximize the overlap of sampling units in the design D_1 and D_2. To find the solution of the controlled selection problem W, Ernst (1998) used the procedure of Causey et al. (1985) and showed that the solution obtained through the procedure of Causey et al. (1985) satisfied all the conditions of maximization of overlap. The procedure of Causey et al. (1985) is based on the theory of controlled rounding, developed by Cox and Ernst (1982). In general, a controlled rounding of an $(N+1) \times (L+1)$ tabular array $W = (w_{ij})$ to a positive integer base b is an $(N+1) \times (L+1)$ tabular array $M = (m_{ij})$ for which $m_{ij} = \lfloor w_{ij}/b \rfloor \; or (\lfloor w_{ij}/b \rfloor + 1)b$ for all i,j where $\lfloor x \rfloor$ denotes the greatest integer not exceeding x.

One drawback of the procedure of Ernst (1998) is that it is quite tedious in implementation. At each step of the procedure one has to obtain the zero-restricted controlled rounding of the adjusted array. Only after this the procedure of controlled selection can be achieved. The other drawback of this procedure is that the variance estimation is not possible in most of the cases due to non-fulfilment of the non-negativity condition $\pi_{ij} \leq \pi_i \pi_j$ of Y-G form of the H-T variance estimator, where π_i and π_j denote the first order inclusion probabilities and π_{ij} is the second order inclusion probability of the units i and j.

Recently, Tiwari and Sud (2012) proposed a procedure for solving sample coordination problem using the multiple objective linear programming. First of all, they constructed a two dimensional real valued array W, with internal units w_{ij}, as defined in (2.6)-(2.9). Using FORTRAN 77 program, they obtained a set A of all possible combinations of units according to the probabilities of the array W. The all possible combinations of units were nothing but the sequence of arrays, M_1, M_2,..., M_t. The probabilities p_1, p_2,..., p_t, satisfying the conditions (2.12)-(2.14), associated with these arrays, were also calculated. After that, they excluded all those arrays from the set A,which do not satisfy the condition (2.13). This set, termed as the set of non-preferred samples, was denoted by A_1. Next, they obtained an IPPS design $p(s)$ for each sample (s) in the set of all possible samples. This IPPS design is known as uncontrolled IPPS design as no restrictions were imposed on the selection probabilities. Tiwari and Sud (2012) used Sampford (1967) IPPS design to obtained the initial uncontrolled IPPS design p(s), as this design imposed only one restriction on selection probabilities. Using the initial IPPS design $p(s)$, they obtained a design $p_0(s)$ given by:

$$p_0(s) = \begin{cases} \frac{p(s)}{1-\sum_{s\in A_1} p(s)} & for s \in (A - A_1) \\ 0 & (otherwise) \end{cases}$$

The design p_0(s) assigned zero probability to the non-preferred samples and was termed as 'controlled design'. This design $p_0(s)$ is no longer an IPPS design. So, Tiwari and Sud (2012) proposed a new IPPS design $p^*(s)$, which is as near as possible to the design $p_0(s)$. To achieve the design $p^*(s)$, they minimized the directed distance D from the sampling design $p^*(s)$ to $p_0(s)$, given as:

$$D(p_0, p^*) = \sum_{s\in(A-A_1)} \frac{p^{*2}(s)}{p_0(s)} - 1$$

The maximization of overlap of units between the two designs was obtained through the solution of the controlled selection problem W. To find the solution of the controlled selection problem W, Tiwari and Sud (2012) applied the theory of multiple objective linear programming. With the help of this multiple objective linear programming they obtained an IPPS design that assigns zero probability to non-preferred samples. Tiwari and Sud (2012) also modified their procedure for the situations where minimization of overlap of sampling units was desirable. To minimize the overlap of sampling units, they redefine the internal units of W and made some changes in the objective function of the linear programming. Their procedure also provided the facility of variance estimation using the HT variance estimator.

3. Optimal Controlled Procedure

We propose a procedure for positive co-ordination problem which also provides the advantage of variance estimation using HT variance estimator. The proposed procedure is compared with the procedure of Tiwari and Sud (2012) and it was found that it provides better results than the procedure of Tiwari and Sud (2012). The linear programming approach in conjunction with a distance function was used in the proposed procedure for maximizing the probability of those sample combinations which consist of maximum number of overlapped sampling units. The non-negativity condition of HT estimator is also achieved in the proposed procedure to facilitate variance estimation through Horvitz-Thompson estimator. The proposed procedure is described as follows:

First of all, we obtain the $(N+1) \times 5$ array W with internal units as discussed in (2.6)-(2.9). Using 'combcoms' command in MATLAB 10.0, we obtain all possible combinations of units according to the probabilities of the array W. Let the set of all possible pairs of D_1 and D_2 samples be denoted by B. The set of all possible samples 'B' satisfy the conditions (2.12)-(2.14). In order to satisfy the condition (2.13), we neglect all those arrays from the set of all possible arrays which do not satisfy the condition (2.13). Let this set of arrays be denoted by B_1. The set B_1 shows the set of non-preferred samples. Now, we obtain an appropriate controlled inclusion probability proportional to size (IPPS) design $p(s)$, for each sample (s) in the set of all possible samples (B), using the values of the internal units (w_{ij}'s) of the array W. The design $p(s)$ assigns zero probability to the non preferred samples and is termed as a controlled IPPS design.

The maximization of overlap of units between the two designs D_1 and D_2 is obtained through the solution of the controlled selection problem W, which satisfies the conditions (2.12)-(2.14). This is achieved through the solution of the following linear programming problem:

$$\text{Maximize} \qquad \phi = \sum_{s \in B} \xi_2(A, s) p(s) \tag{3.1}$$

Subject to the constraints:

$$\begin{aligned}
&i)\ \sum_{s \in B - B_1} p(s) = 1 \\
&ii)\ \sum_{s \in B - B_1} p(s) m_{ij} = w_{ij}, \qquad i = 1, ..., N, j = 1, ..., 4 \\
&iii)\ \lfloor w_{ijv} \rfloor \leq m_{ijv} \leq \lfloor w_{ijv} \rfloor + 1, \qquad i = 1, ..., N; j = 1, ..., 4; v = 1, ..., u \\
&iv)\ p(s) \geq 0 \\
&v)\ \sum_{s \ni i,j} p(s) \leq (\pi_i)_l (\pi_j)_l, \qquad i < j = 1,...,\text{N};\ l = 1,\ 2 \\
&vi)\ \sum_{s \ni i,j} p(s) > 0, \qquad i < j = 1, .., N
\end{aligned} \tag{3.2}$$

where B-B_1 shows the set of sample combinations consisting of maximum number of overlapped sampling units, s represents a sample in the set of all possible samples generated through the $(N+1) \times 5$ array W and $\lfloor x \rfloor$ denotes the greatest integer not exceeding x.

In the proposed procedure, constraints (i) and (iv) in (3.2) are necessary for any sampling design. Constraints (ii) and (iii) are required to satisfy (2.11) and (2.13), respectively. Constraint (ii) also ensures that the resultant design is an IPPS design. Constraint (v) in (3.2) is desirable as it ensures the sufficient condition for non-negativity of the Y-G form of the HT variance estimator and constraint (vi) is desirable as it ensures unbiased variance estimation using HT estimator.

In many situations, it is often desirable to withdraw the selection of same unit for two or more surveys. In these situations, we have to minimize the overlap of sampling units for two or more surveys. The proposed procedure can be easily modified to minimize the overlap of sampling units. In order to minimize the overlap of sampling units, we have to redefine the internal units of the array W. For negative co-ordination, condition (2.6) is replaced by

$$w_{i3} = max(\pi_{i1} + \pi_{i2} - 1, 0) \tag{3.3}$$

Conditions (2.7), (2.8) and (2.9) will remain the same as for the case of maximization of overlap of sampling units. The objective function, in the case of minimization of overlap of sampling units is redefined as:

$$Maximize \qquad \phi = \sum_{s \in C} \xi_2(A,s)p(s) \tag{3.4}$$

where C denotes the set of all sample combinations, which consists of minimum number of overlapped sampling units.

One limitation of the proposed linear programming approach is that it becomes cumbersome when the population size is large, as the process of enumeration of all possible samples and formation of the objective function and constraints becomes quite tedious. With the help of faster computing techniques and modern statistical tools, there may not be much difficulty in using the proposed procedure for large populations. The proposed plan takes lesser computing time in comparison to the procedures of Ernst (1998) and Tiwari & Sud (2012).

The proposed procedure can be used for the situations when the two surveys are conducted for the same population with identical stratification. These two surveys can be conducted sequentially or simultaneously. There is no restriction on the number of units selected per stratum. The proposed procedure is superior to the procedures of Ernst (1998) and Tiwari and Sud (2012) as the proposed procedure maximizes the probability of those sample combinations which consists of maximum number of overlapped sampling units (in case of positive co-ordination) or minimize the probability of those sample combinations which consists of maximum number of overlapped sampling units (in case of negative co-ordination). The proposed procedure also ensures variance estimation using H-T variance estimator and in the situations, where the conditions of H-T estimator could not be satisfied, some alternative variance estimator can be used.

4. Empirical Evaluation

In this section, we shall present some empirical examples to demonstrate the utility of the proposed procedure. We also compare the proposed procedure with the procedures of Ernst (1998) and Tiwari and Sud (2012).

Example 1.1 (Maximization Case): Let consider the following example taken from Ernst (1998), with inclusion probabilities and values of characteristic Y (given in Table 4.1) for two sampling designs with 5 ($N = 5$) different units in each stratum.

Table 4.1

Inclusion probabilities of units

i	1	2	3	4	5
π_{i1}	0.6	0.4	0.8	0.6	0.6
π_{i2}	0.8	0.4	0.2	0.4	0.2

Consider that a sample of size 3 is to be selected for sampling design D_1 and a sample of size 2 for the sampling design D_2, then find the values of internal units of W. Using (2.6)-(2.9), the array W is obtained as:

	0.0	0.2	0.6	0.2	1
	0.0	0.0	0.4	0.6	1
W =	0.6	0.0	0.2	0.2	1
	0.2	0.0	0.4	0.4	1
	0.4	0.0	0.2	0.4	1
	1.2	0.2	1.8	1.8	5

Now we have to solve the above controlled selection problem with $4N$ $(= 20)$ internal units in W and $n = 5$, where n denotes the total number of sample units to be selected from the two designs. The set of all possible samples (B) consists of 15,504 samples. Out of these 15,504 samples, only 24 samples satisfy the condition (2.13). Therefore, all arrays M_l that belongs to the set $(B\text{-}B_1)$ consists of 24 samples given as:

Sample 1 0.0 0.2 0.0 0.0 0.0 0.0 0.4 0.0 0.6 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.4	Sample 2 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.6 0.6 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.2 0.0	Sample 3 0.0 0.2 0.0 0.0 0.0 0.0 0.4 0.0 0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.4 0.4 0.0 0.0 0.0	Sample 4 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.6 0.6 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0
Sample 5 0.0 0.0 0.6 0.0 0.0 0.0 0.4 0.0 0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.4	Sample 6 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.6 0.6 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.4	Sample 7 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.6 0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.2 0.0	Sample 8 0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.6 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.4
Sample 9 0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.2 0.0	Sample 10 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.6 0.6 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.2 0.0	Sample 11 0.0 0.2 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.2 0.0 0.0 0.0 0.4 0.0 0.0 0.0	Sample 12 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.2 0.0 0.2 0.0 0.0 0.0 0.4 0.0 0.0 0.0
Sample 13 0.0 0.0 0.6 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.4	Sample 14 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.2 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.4	Sample 15 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.2 0.2 0.0 0.0 0.0 0.0 0.0 0.2 0.0	Sample 16 0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.0 0.0 0.2 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.4
Sample 17 0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.2 0.0 0.0 0.0 0.0 0.0 0.2 0.0	Sample 18 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.6 0.0 0.0 0.2 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.2 0.0	Sample 19 0.0 0.0 0.6 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.4 0.4 0.0 0.0 0.0	Sample 20 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.4 0.4 0.0 0.0 0.0

Sample 21	Sample 22	Sample 23	Sample 24
0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0	0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.4 0.4 0.0 0.0 0.0	0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0	0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.6 0.0 0.0 0.2 0.0 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0

Here B-B_1 consists of the set of sample combinations which have two overlapped sampling units, that is, sample numbers 5, 6, 7, 8, 9, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24. Thus, the objective function (3.1) becomes:

Maximize $\phi = 0.2476{*}x[5] + 0.1874{*}x[6] + 0.223{*}x[7] + 0.3557{*}x[8] + 0.4013{*}x[9] + 0.3271{*}x[10] + 0.4013{*}x[13] + 0.3271{*}x[14] + 0.3705{*}x[15] + 0.5444{*}x[16] + 0.6115{*}x[17] + 0.5047{*}x[18] + 0.3557{*}x[19] + 0.2863{*}x[20] + 0.2863{*}x[21] + 0.4860{*}x[22] + 0.4860{*}x[23] + 0.4013{*}x[24]$

Applying the method discussed in Section 3 and solving the resultant linear programming problem through the SAS 9.3 and MATLAB 10.0 windows version packages, we obtain the controlled IPPS plan given in Table 4.2.

The objective function has the value: $\phi = 0.260624$

Table 4.2

Optimal controlled IPPS plan corresponding to proposed procedure

s	$p(s)$	s	$p(s)$	s	$p(s)$	s	$p(s)$
1	0.0	7	0.12	13	0.08	19	0.04
2	0.0	8	0.04	14	0.0	20	0.12
3	0.0	9	0.0	15	0.0	21	0.0
4	0.20	10	0.0	16	0.04	22	0.0
5	0.12	11	0.0	17	0.08	23	0.0
6	0.12	12	0.0	18	0.0	24	0.0

Ernst (1998) obtained the following solution for this problem:

$p_3 = 0.2$, $p_6 = 0.4$, $p_{18} = 0.2$, $p_{19} = 0.2$

Tiwari and Sud (2012) obtained the following solution for this problem:

$p_1 = 0.08$, $p_3 = 0.04$, $p_4 = 0.08$, $p_5 = 0.08$, $p_6 = 0.24$, $p_9 = 0.08$,
$p_{15} = 0.12$, $p_{20} = 0.16$, $p_{22} = 0.04$, $p_{23} = 0.08$

Using the procedures of Ernst (1998) and Tiwari and Sud (2012), we find the value of ϕ is 0.8 and 0.8, respectively. Thus, we observe that the value of ϕ for the proposed procedure is very small in comparison to the procedures of Ernst (1998) and Tiwari and Sud (2012). With the help of proposed procedure we can also estimate the value of variance using the Horvitz-Thompson variance estimator.

Example 1.2 (Minimization Case): Let us suppose the inclusion probabilities of Example 1.1 for the two sampling designs for 5 different units. For the sampling design

D_1, we have to select a sample size of 3, and a sample of size 2 for the sampling design D_2, in such a way that the overlap between the two designs is minimized. First of all we find the values of internal units of W. Using (2.7)-(2.9) and (3.3), W is obtained as:

$$W = \begin{array}{c|c} 0.2\ 0.4\ 0.4\ 0.0 & 1 \\ 0.4\ 0.4\ 0.0\ 0.2 & 1 \\ 0.8\ 0.2\ 0.0\ 0.0 & 1 \\ 0.6\ 0.4\ 0.0\ 0.0 & 1 \\ 0.6\ 0.2\ 0.0\ 0.2 & 1 \\ \hline 2.6\ 1.6\ 0.4\ 0.4 & 5 \end{array}$$

Solving the controlled selection problem with $N = 20$ and $n = 5$ the possible combinations satisfying condition (2.13) are given in Appendix. After solving this example with the help of proposed scheme, we obtain the controlled IPPS sampling plan given in Table 4.3.

Table 4.3

Optimal controlled IPPS plan corresponding to proposed scheme

s	$p(s)$	s	$p(s)$	s	$p(s)$	s	$p(s)$
1	0.12	5	0.0	9	0.04	13	0.0
2	0.0	6	0.08	10	0.0	14	0.20
3	0.08	7	0.16	11	0.0	15	0.0
4	0.12	8	0.0	12	0.2	16	0.0

The value of the objective function is: $\phi\ = 0.165016$

We also solve this example by the procedure of Ernst (1998), we get the following result:

$$p_1 = 0.4,\ p_6 = 0.2,\ p_{13} = 0.2,\ p_{16} = 0.2$$

Tiwari and Sud (2012) obtained the following solution for this problem:

$$p_2 = 0.08,\ p_3 = 0.04,\ p_4 = 0.16,\ p_5 = 0.12,\ p_7 = 0.16,\ p_8 = 0.04,$$
$$p_{11} = 0.08,\ p_{12} = 0.12,\ p_{14} = 0.20;$$

Following the procedures of Ernst (1998) and Tiwari and Sud (2012), the value of ϕ is same for both the procedures is 0.6.

5. Conclusion

In this article, we have proposed a linear programming approach with distance function as a weight for each sample, to obtain an optimum solution for the sample co-ordination problem. The proposed procedure is superior to the procedures of Ernst (1998) and Tiwari and Sud (2012) as it maximizes the probability of sample combinations having maximum number of overlapped samplin units (in case of positive co-ordination) or minimize the probability of sample combinations having maximum number of overlapped sampling units (in case of negative co-ordination). The proposed procedure also ensures variance estimation using Y-G (1953) form of H-T (1952) variance estimator as it satisfies the non-negativity condition of Horvitz-Thompson variance estimator through constraint (v) in Eq. (3.2). The proposed procedure takes lesser computing time in comparison to the procedures of Ernst (1998) and Tiwari and Sud (2012) and is found to be more advantageous than these procedures.

ACKNOWLEDGEMENT

The authors are thankful to the two referees and the Editor for their constructive comments, which led to considerable improvement in presentation of this manuscript.

Appendix

Example 1.2 (Minimization Case). For this example the all possible combinations are as follows:

Sample 1 0.0 0.4 0.0 0.0 0.0 0.4 0.0 0.0 0.8 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.6 0.0 0.0 0.0	Sample 2 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.6 0.0 0.0 0.0 0.6 0.0 0.0 0.0	Sample 3 0.2 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.2 0.0 0.0 0.6 0.0 0.0 0.0 0.6 0.0 0.0 0.0	Sample 4 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0 0.8 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.6 0.0 0.0 0.0
Sample 5 0.2 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.8 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.6 0.0 0.0 0.0	Sample 6 0.2 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.4 0.0 0.0 0.6 0.0 0.0 0.0	Sample 7 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.0 0.8 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.2 0.0 0.0	Sample 8 0.2 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.8 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.2 0.0 0.0
Sample 9 0.2 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.2 0.0 0.0	Sample 10 0.2 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.8 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.2 0.0 0.0	Sample 11 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.2 0.0 0.0 0.6 0.0 0.0 0.0 0.6 0.0 0.0 0.0	Sample 12 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.8 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.6 0.0 0.0 0.0
Sample 13 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.8 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.2 0.0 0.0	Sample 14 0.0 0.0 0.4 0.0 0.0 0.4 0.0 0.0 0.8 0.0 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.2	Sample 15 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.2	Sample 16 0.0 0.0 0.4 0.0 0.4 0.0 0.0 0.0 0.8 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.0 0.2

Now we apply the proposed model as follows:

Max. ϕ = 0.1563*x[1] + 0.3235*x[2] + 0.3622*x[3] + 0.2081*x[4] + 0.2409*x[5] + 0.4325*x[6] + 0.2409*x[7] + 0.2751*x[8] + 0.4792*x[9] + 0.3362*x[10];

After solving the above model, we find the desired results shown in example 1.2.

References

[1] Brewer, K. R. W., Early, L. J., and Joyce, S. F. *Selecting several samples from a single population.* Austral. J. Statist. 14, 231-239, 1972.

[2] Causey, B.D., Cox, L.H. and Ernst, L.R. *Application of transformation theory to statistical problem.* J. Amer. Statist. Assoc., 80, 903-909, 1985.

[3] Cox, L.H. and Ernst, L.R. *Controlled rounding.* INFOR 20, 423-432, 1982.

[4] Deville, J.C. and Tille, Y. *Selection of several unequal probability samples from the same population.* J. Statist. Plann. Infer. 86, 215-227, 2000.

[5] Ernst, L. R. *Maximizing the overlap of sampling units for two designs with simultaneous selection.* J. Office. Statist. 12, 33-45, 1996.

[6] Ernst, L. R. *Maximizing and minimizing overlap when selecting a large number of units per stratum simultaneously for two designs.* J. office. Statist. 14, 297-314, 1998.

[7] Ernst, L. R. and Ikeda, M. *A reduced size transportation algorithm for maximizing the overlap between surveys.* Surveys Methodology 21, 147-157, 1995.

[8] Ernst, L. R. and Paben, S. P. *Maximizing and minimizing the overlap when selecting any number of units per stratum simultaneously for two designs with different stratifications.* J. Offic. Statist. 18, 185-202, 2002.

[9] Fellegi, I. *Changing the probabilities of selection when two units are selected with PPS without replacement.* Proc. Soc. Statist. Sec. Washington: American Statistical Association. pp. 434-442, 1966.

[10] Fellegi, I.P. *Sampling with varying probabilities without replacement: Rotating and non-rotating samples.* J, Amer. Stat. Assoc., 58, 183-201, 1963.

[11] Gray, G. and Paltek, R. *Several methods of redesign area samples utilizing probabilities proportional to size when the sizes change significantly.* J. Amer. Statist. Assoc. 63, 1280-1297, 1963.

[12] Horvitz, D.G. and Thompson, D.J. *A generalization of sampling without replacement from finite universe.* J. Amer. Statist. Assoc., 47, 663-685, 1952.

[13] Keyfitz, N. *Sampling with probabilities proportional to size: Adjustment for changes in probabilities.* J. Amer. Statist. Assoc. 46, 105-109, 1951.

[14] Kish, L. *Changing strata and selection probabilities.* Proc. Soc. Statist. Sec. Washington: Amer. Statist. Assoc. pp. 124-131, 1963.

[15] Kish, L. and Hess, I. *Some sampling techniques for continuing surveys operations.* Proc. Soc. Statis. Sec. Washington: American Statistical Association. pp. 139-143, 1959.

[16] Kish, L. and Scott, A. *Retaining units after changing strata and probabilities.* J. Amer. Statist. Assoc. 66, 461-470, 1971.

[17] Matei, A. and Tille, Y. *Maximal and minimal sample co-ordination.* Sankhya Ind. J. Statist. 67, 590-612, 2006.

[18] Matei, A, and Skinner, C. *Optimal sample coordination using controlled selection.* J. Statist. Plann. Infer. 139, 3112-3121, 2009.

[19] Patterson, H. *Sampling on successive occasions with partial replacement of units.* J. Roy. Statist. Soc. B. 12, 241-255, 1950.

[20] Raj, D. *On the method of overlapping maps in sample surveys.* Sankhya Ind. Statist. 17, 89-98. 1956.

[21] Tiwari, N. and Sud, U.C. *An optimal procedure for sample coordination using multiple objective functions and nearest proportional to IPPS size sampling design.* Comm. Statist.-Theory and Methods, 41, 2014-2033, 2012.

[22] Yates, F. and Grundy, P.M. *Selection without replacement from within strata with probability proportional to size.* J. Roy. Statist. Soc., B15, 253-261, 1953.