

THE USE OF THE TRANSPORTATION PROBLEM IN COORDINATING THE SELECTION OF SAMPLES FOR BUSINESS SURVEYS

Philip T. Reiss, Ioana Şchiopu-Kratina, Lenka Mach¹

ABSTRACT

We view the problem of maximizing or minimizing the expected overlap of two surveys as a transportation problem (TP) and give simple selection algorithms for solving it. We compare our method with the sequential SRSWOR for positive and negative coordination, the method of collocated samples and the PRN method with full-stratum rotation.

KEY WORDS: SRSWOR, transportation problem, NWCR, expected overlap, hypergeometric distribution, p.p.s. selection.

RÉSUMÉ

On regarde le problème de la maximisation ou de la minimisation du chevauchement espéré entre deux enquêtes comme un problème de transport. On donne des algorithmes de sélection simples pour solutionner ce problème. On compare notre méthode avec le sondage aléatoire simple séquentiel pour les coordinations positive et négative, la méthode des échantillons colloqués et la méthode des nombres aléatoires permanents avec rotation complète dans les strates.

MOTS CLÉS : Sondage aléatoire simple, problème de transport, règle du coin nord-ouest, chevauchement espéré, loi hypergéométrique, sélection ppt.

1. INTRODUCTION

We consider two surveys with stratified frames and simple random sampling without replacement (SRSWOR) designs in each stratum. We maximize or minimize the expected overlap of samples selected for the two surveys and preserve both designs as in (TP) below. We compare our method with some permanent random numbers (PRN) methods (Ernst (1999), Ohlsson (1995)): sequential SRSWOR, collocated sampling and the PRN method with full-stratum rotation. We use the Northwest Corner Rule (NWCR) solution given in Section 2 to select the sample(s) through a two-step selection algorithm. In Section 3 we maximize the expected overlap and give selection algorithms. We minimize the expected overlap in Section 4 and compare our method to some PRN methods in Section 5.

A sample design on a finite population is a probability space (U, S, P) , $P = (p_i)_{i=1, \dots, n}$, with samples $s_i \in S$, $i = 1, \dots, n$, as elementary events. We consider two designs (U, S, P) and (U', S', Q) , where $Q = (q_j)_{j=1, \dots, m}$. There always exists a probability distribution $X = (x_{ij})_{i, j \geq 1}$ with P and Q as marginals, i.e. a solution to the transportation problem:

$$(TP) \quad x_{ij} \geq 0, \quad \sum_{j=1}^m x_{ij} = p_i, \quad 1 \leq i \leq n, \quad \sum_{i=1}^n x_{ij} = q_j, \quad 1 \leq j \leq m, \quad \sum_{i=1}^n p_i = \sum_{j=1}^m q_j = 1.$$

We seek a solution to (TP) which maximizes or minimizes a function of interest, called the objective function. A first TP application to sampling was given by Des Raj (1956): two surveys are carried out on a dispersed population of villages and the number of villages visited once for both surveys must be maximized to minimize the cost of operations in the field. Another important application occurs when a sample is updated. We wish to retain a maximum number of units to

¹ Philip T. Reiss (ptr2003@columbia.edu), Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, 6th Floor, New York, NY, 10032. Ioana Şchiopu-Kratina (ioana.schiopu-kratina@statcan.ca), Statistics Canada, R.H.Coats 16th Floor, Ottawa, Ontario, K1A 0T6, and Lenka Mach (lenka.mach@statcan.ca), Statistics Canada, R.H. Coats Building, 11th Floor, Ottawa, Ontario, K1A 0T6.

compare estimates over time and also to reduce the cost of first contacts, i.e. maximize the expected number of “non-births” in the sample. On the other hand, the expected overlap of samples must be minimized to reduce response burden when several surveys are taken on the same population. The original TP has a large number of variables, but we deal instead with hypergeometric probabilities for configurations of samples. We solve well-defined mathematical problems and do not require an assignment of PRNs to units.

2. THE TRANSPORTATION PROBLEM FOR TWO SURVEYS

The NWCR gives a solution $X = (x_{ij})_{i,j \geq 1}$ to the TP and is a greedy algorithm, i.e. each x_{ij} is the maximum possible probability (mass) given the revised marginals $p'_i, q'_j, i, j \geq 1$. Each $x_{ij} \neq 0, i, j \geq 1$, is the left-over mass of the corresponding row or column and so the remainder of that row or column has 0 mass (see Example 1).

(NWCR): We set $x_{i0} = 0, x_{0j} = 0$ and, if x_{rs} has been defined for all pairs $(r, s), r \leq i, s \leq j, (r, s) \neq (i, j)$, then

$$x_{ij} = \min \left\{ p_i - \sum_{s=0}^{j-1} x_{is}, q_j - \sum_{r=0}^{i-1} x_{rj} \right\}, i, j \geq 1. \blacksquare$$

Proposition 1 (see also Hoffman (1985), p. 99): If $X(i, j), P(i)$ and $Q(j)$ represent the cumulative distributions, i.e.

$$P(i) = \sum_{k=1}^i p_k, Q(j) = \sum_{k=1}^j q_k, X(i, j) = \sum_{k=1}^i \sum_{l=1}^j x_{kl} \text{ and } X \text{ is given by NWCR, then } X(i, j) = \min \{P(i), Q(j)\}, i, j \geq 1.$$

Proof: For each $(i, j), X(i, j) \leq \min \{P(i), Q(j)\}$, so it suffices to show that $X(i, j) \geq \min \{P(i), Q(j)\}$. The statement is clearly true if $i = 1$ or $j = 1$. For $\min(i, j) > 1$, i.e. the complementary situation, we consider two cases.

Case 1: $x_{ij} \neq 0$. The left-over mass of p_i or that of q_j has been placed in cell (i, j) . We show that $X(i, j) \geq P(i)$ (or $Q(j) \geq \min \{P(i), Q(j)\}$). Note that this actually means $X(i, j) = P(i)$ (or $Q(j)$) because the opposite inequality always holds. Assume, without loss of generality, that the left-over mass of row i has been placed in (i, j) . We claim that all marginals p_k with $k < i$ would have been distributed within the block with lower right corner (i, j) , i.e. $X(i, j) \geq P(i)$. If this were not so, some p_k would not have been distributed by column j , where $k < i$. We now claim that the mass of column j would have been distributed by row k , which would imply that $x_{ij} = 0$, contradicting the assumption of *Case 1*. If $x_{kj} \neq 0$, it must have been the left-over portion of q_j that was placed in (k, j) since we would not have distributed all of p_k . If $x_{kj} = 0$, either the mass of row i or that of column j has been distributed before (k, j) . As before, it must have been the mass of the column j . In either situation, $x_{k'j} = 0$, for all $k' > k$ and so $x_{ij} = 0$ (since $i > k$), which is impossible in *Case 1*.

Case 2: $x_{ij} = 0$. By (NWCR), p_i (or q_j) has been distributed before column j (row i). Assume the distribution of q_j was completed in row $k < i$. Then, as above, we have $X(k, j) = Q(j) \leq X(i, j)$ (since $i > k$), so $X(i, j) \geq \min \{P(i), Q(j)\}$. ■

Definition 1: The matrix of overlaps $R = (\rho_{ij})_{i,j \geq 1}$ (or the function ρ) is supermodular (SM) if it satisfies:

$$(SM) \quad \rho_{ij} + \rho_{rs} \geq \rho_{is} + \rho_{rj}, \text{ for all } 1 \leq r \leq i, 1 \leq s \leq j.$$

Proposition 2 (Hoffman (1985)): If R is SM and X is given by NWCR, then the expected overlap $E_X(\rho) = \sum_{i,j} \rho_{ij} x_{ij}$ is a

maximum, i.e. $E_X(\rho) = \max \sum_{i,j} \rho_{ij} x'_{ij}$ over all solutions X' of (TP). ■

Definition 2: The matrix of overlaps R (or the function ρ) is submodular (sM) if it satisfies:

$$(sM) \quad \rho_{ij} + \rho_{rs} \leq \rho_{is} + \rho_{rj}, \text{ for all } 1 \leq r \leq i, 1 \leq s \leq j.$$

Example 1: Consider the following SM matrix, where the interior consists of entries of the objective function (“overlaps”) and the marginals are probabilities:

$$\begin{array}{c|cc} 1 & 1/2 & 1/2 \\ \hline 1/3 & 1 & 1 \\ 2/3 & 0 & 1 \end{array}$$

The NWCR gives: $x_{11} = \min \{1/2, 1/3\} = 1/3, x_{12} = 0, x_{21} = 1/6, x_{22} = 1/2$ with the maximum expected overlap of $5/6$. If we interchange the rows above, we obtain an sM matrix and the NWCR gives the minimum expected overlap of $1/2$. ■

Proposition 3: A matrix R is SM if and only if (SM') holds:

$$(SM') \quad \rho_{ij} + \rho_{i+1,j+1} \geq \rho_{i,j+1} + \rho_{i+1,j}, i=1,\dots,n-1, j=1,\dots,m-1. \blacksquare$$

Lemma 1: If $c_i, i = 1, \dots, n$, and $c'_j, j = 1, \dots, m$, are two decreasing sequences of positive numbers, then the $n \times m$ matrix with entries $\rho_{ij} = \min\{c_i, c'_j\}$ is SM.

Proof: By Proposition 3 it suffices to prove (SM'). By the hypotheses, $\rho_{i+1,j} = \min\{\rho_{ij}, c_{i+1}\}$ and $\rho_{i,j+1} = \min\{\rho_{ij}, c'_{j+1}\}$. Thus the right hand side of (SM') equals $\min\{\rho_{ij}, \min\{c_{i+1}, c'_{j+1}\}\} + \min\{\rho_{ij}, \max\{c_{i+1}, c'_{j+1}\}\}$, since one of c_{i+1}, c'_{j+1} is the minimum and the other is the maximum of the two. (SM') then becomes $\rho_{ij} + \min\{\rho_{ij}, c_{i+1}, c'_{j+1}\} \geq \min\{\rho_{ij}, \min\{c_{i+1}, c'_{j+1}\}\} + \min\{\rho_{ij}, \max\{c_{i+1}, c'_{j+1}\}\}$, which reduces to the trivial $\rho_{ij} \geq \min\{\rho_{ij}, \max\{c_{i+1}, c'_{j+1}\}\}$. \blacksquare

3. MAXIMUM OVERLAP

The selections are SRSWOR from a stratum common to two surveys. This applies to the case of distinct surveys and also when we select for the same survey after updates to the frame. We have N units in the first population, from which n are selected, and $N' = N - D + B = C + B$ units in the second, from which n' are selected. Here D units belong only to the first population (deaths), B only to the second (births) and C belong to both populations. If $s \in S$, $c(s)$ (or c) is its number of units from the common pool (non-deaths). The overlap of two samples is the number of units they have in common. Thus $c(s)$ is the maximum overlap that s could have with any $s' \in S'$. Similarly, $c(s')$ (or c') represents the non-births in s' . The samples from the old (first) design will form the rows of the matrix of overlaps. We group those with the same value of c and calculate the probabilities $p(c)$ of occurrence of each such group, which we call a super-row. These probabilities form a hypergeometric distribution. We order the super-rows in decreasing order of values of c . We perform the same operations in creating super-columns. The cells (c, c') of the matrix of overlaps have as entries the block-overlaps $\rho(c, c') = \min\{c, c'\}$ and so this matrix is SM by Lemma 1.

Example 2: Let $N = 6, n = 3, D = 3$ and $B = 2$. Then $N' = 5$, from which we select by SRSWOR $n' = 4$ units. In Tables 1 and 2, the left margin refers to the original survey while the top margin refers to the new survey. The margins of Table 1 are the possible values of c and c' , and the entries give the resulting maximal overlap, $\min\{c, c'\}$. The margins of Table 2 contain the hypergeometric probabilities corresponding to the super-rows and super-columns, which we denote $p(c)$ and $q(c')$; the entries are the joint probabilities assigned by (NWCR). There are two columns in Table 1 since $C = 3$ and $B = 2$ and so the maximum c' is $c'_M = 3$ and the minimum is $c'_m = 2$.

$c \setminus c'$	3	2
3	3	2
2	2	2
1	1	1
0	0	0

$p(c) \setminus q(c')$	8/20	12/20
1/20	1/20	0
9/20	7/20	1/10
9/20	0	9/20
1/20	0	1/20

Theorem 1: Consider two SRSWOR surveys. A solution to (TP) which maximizes the expected overlap is obtained in the following way: Form super-rows by grouping samples from the first survey which have the same value of c and calculate the probability of each super-row. Order the super-rows in descending order of values of c . Repeat the procedure with super-columns for the second survey, and form a matrix with these super-rows and super-columns. Distribute the mass of the super-rows and super-columns to the cells of this matrix (called blocks) using (NWCR). Distribute the mass of each block (c, c') equally among those of its cells (s, s') that have the maximum possible overlap within the block, $\rho(c, c')$. All other cells in the block receive 0 mass.

Proof: We show first that (TP) holds. We know that NWCR has the prescribed marginals, which are defined at the level of super-rows and super-columns. We show that, within each super-row (column), all rows (columns) receive the same mass. We are only concerned with entries (pairs of samples) which attain the maximum possible overlap within a block (the block-overlap) assumed, without loss of generality, to be $c < c'$. Each row (column) of the block has $n - c$ deaths ($n' - c'$ births). Consider a sample s in this block. It contains $\binom{C-c}{c'-c} \binom{B}{n'-c'}$ entries with overlap c , i.e. samples s' which

contain the same c units from the common pool as s . A similar formula holds for columns and the number of entries with overlap c in this block is the product of these. They receive an equal amount of mass and so each row and column within a block receives the same mass. Thus rows (columns) within a super-row (column) pick the same amount of mass from each block and therefore across all super-columns (rows) of the matrix. To show optimality it suffices to apply Proposition 2 at the level of blocks because the entire mass of a block is placed on entries with maximum overlap. The matrix of block-overlaps is SM (Lemma 1), so NWCR gives the maximum expected overlap by Proposition 2. ■

Example 3: We refer to Example 2 and expand below the block corresponding to cell (2, 2). The “marginals” are now the samples. We use u , d and b to denote the units from the common pool, deaths and births respectively. Here $c = c' = 2$. The top margins list all new samples with two births, while the left margins list all old samples which contain one death. Table 3 contains the actual overlap of pairs of samples, and Table 4 the probabilities of selection in our optimal solution.

Table 3				Table 4			
	$u_1u_2b_1b_2$	$u_3u_2b_1b_2$	$u_1u_3b_1b_2$		$u_1u_2b_1b_2$	$u_3u_2b_1b_2$	$u_1u_3b_1b_2$
$u_1u_2d_1$	2	1	1	$u_1u_2d_1$	1/90	0	0
$u_3u_2d_1$	1	2	1	$u_3u_2d_1$	0	1/90	0
$u_1u_3d_1$	1	1	2	$u_1u_3d_1$	0	0	1/90
$u_1u_2d_2$	2	1	1	$u_1u_2d_2$	1/90	0	0
$u_3u_2d_2$	1	2	1	$u_3u_2d_2$	0	1/90	0
$u_1u_3d_2$	1	1	2	$u_1u_3d_2$	0	0	1/90
$u_1u_2d_3$	2	1	1	$u_1u_2d_3$	1/90	0	0
$u_3u_2d_3$	1	2	1	$u_3u_2d_3$	0	1/90	0
$u_1u_3d_3$	1	1	2	$u_1u_3d_3$	0	0	1/90

Recall that the probability of the block is 1/10. We distribute it equally among the 9 pairs of samples with overlap of 2, so each receives a mass of 1/90. All other pairs in the block have probability 0 of being selected. ■

Assume now that a sample s has been drawn first. We need to select s' in the updated population so that the expected overlap is maximized. Theorem 1 applies to super-rows and super-columns while the actual selection is done for a sample selected in the first survey (a row). The selection could be adapted to the level of samples but this is not needed. One can show that for all samples in the same super-row the same super-column will be selected.

Selection algorithm 1: The conditions and notation are as in Proposition 1 and Theorem 1. The selected sample s belongs to super-row i . The super-rows have been ordered in decreasing order of values of c up to and including super-row i . We calculate $P(i-1)$ and $P(i)$ and define $X(j; i) = X(i, j) - X(i-1, j)$, $j = 1, \dots, m$. Note that $P(i) - P(i-1) = X(m; i)$ is the mass of the super-row i , p_i . To select the super-column which contains the new sample, we randomly pick $r = r(i) \leq p_i$, then create and label super-columns j in descending order of values of c' . We calculate $Q(j)$, then calculate $X(j; i)$, $j \geq 1$. The new sample s' will belong to super-column $j(r)$, where $j(r)$ is the smallest integer such that $r \leq X(j(r); i)$. Let $c = c(s)$, $c' = c(s')$ and $\rho = \min\{c, c'\}$. We now randomly select a sample s' from the subset $\{s': \rho(s, s') = \rho\}$ of the super-column $j(r)$. If $\rho = c < c'$ then there are c common units in s , which we retain. We select at random $c' - c$ more units from the $C - c$ common-pool units still available. We randomly select $n' - c'$ units from the B births to complete s' . If $\rho = c' < c$, we deselect randomly $c - c'$ common units from s , then complete s' by randomly selecting $n' - c'$ births. ■

Example 4: We consider Examples 2 and 3 and assume that $s = u_1u_2d_1$ was selected first, in super-row 2. Note that $P(1) = 1/20$, $P(2) = 1/2$. We pick a random number $r \leq 9/20$, say $r = 3/20$. Now $Q(1) = 8/20$ and $X(1; 2) = X(2, 1) - X(1, 1) = 7/20 > 3/20$. Thus the new sample is in super-column 1 with $c' = 3$, so we retain u_1u_2 and add u_3 from the common pool. We randomly select one of the two births, say b_2 , and the new sample is $s' = u_1u_2u_3b_2$. ■

The implementation of the selection algorithm is particularly simple when we are updating for births (or deaths) only or simply changing the sample size. We applied Corollary 1 below in selecting a sample for the Local Government Surveys.

Corollary 1 (Births only): Assume that $D = 0$, so we have only one super-row with $c(s) = n$ for all $s \in S$, and we determine the number of births from the appropriate hypergeometric distribution. We then select the units as above. ■

Selection algorithm 2 (Simultaneous selection for two surveys with sample sizes n and n'): We randomly pick $r \in [0, 1]$, then recursively create super-rows and label them in decreasing order of c . We calculate the cumulative super-row probabilities $P(i)$ and stop at the first $i = i(r)$ such that $r \leq P(i)$. We do the same for super-columns and obtain the super-column index $j(r)$. The pair of samples (s, s') is in the block labeled $(i(r), j(r))$ with overlap $\rho = \min\{c, c'\}$. If $\rho = c'$ then we select the samples as follows. Using an SRSWOR selection scheme, we first pick c' units from the common pool of C units. To complete the selection of the sample s' , we randomly select $n' - c'$ units from the B births. To complete the sample s , we randomly select $c - c'$ units from the remaining $C - c'$ common units and then $n - c$ more units from the “dead” units. The case $\min\{c, c'\} = c$ is similar and will be omitted. ■

4. MINIMIZING THE OVERLAP OF TWO SURVEYS

Consider surveys with a stratified SRSWOR design and a common stratification. For (s, s') with $c(s) = c$ and $c(s') = c'$, the minimum possible overlap is $\rho(s, s') = \max\{0, c(s) + c(s') - C\} = \rho(c, c')$, defined as the block overlap of (c, c') . Note that here, in order to obtain an sM matrix of overlaps, we must order the rows and the columns of the matrix in opposite order.

Theorem 2 (Minimizing the overlap of two surveys and selection algorithm): A solution to TP for two surveys which minimizes the expected overlap of samples is obtained in the following way: Form super-rows and super-columns as in Theorem 1. Order the super-rows, say, in increasing order of values of c . Then order the super-columns in decreasing order of values of c' . The matrix of overlaps formed with these as rows and columns is sM. Distribute the mass to the blocks by NWCR. Within each block, distribute the mass equally to cells with the block-overlap (minimum possible within the block). All other cells in the block receive mass 0. The selection of a sample within a block with block-overlap $\rho(c, c') = \max\{0, c + c' - C\}$ is done as follows. Using a SRSWOR selection scheme, we first pick ρ units from the pool of C common units. Next, we randomly select $c' - \rho$ units from the remaining units in the common pool for s' and a different set of $c - \rho$ units for s . We complete s' by randomly picking $n' - c'$ units from the B births. Similarly, we select $n - c$ deaths for s .

Proof: In view of previous results we only need to show that the matrix of overlaps is sM, for which we use an adaptation of Proposition 3. Assume that the super-rows were labeled starting with c_m and ending with c_M , while the super-columns were labeled starting with c'_M and ending with c'_m . For a block with overlap $\rho_{ij} = \max\{0, c_i + c'_j - C\}$, we distinguish three cases:

- (i) $c_i + c'_j > C$. Then $\rho_{i,j+1} = \rho_{ij} - 1$, $\rho_{i+1,j} = \rho_{ij} + 1$ and $\rho_{i+1,j+1} = \rho_{ij} + 1 - 1$, so (sM) holds with equality.
- (ii) $c_i + c'_j = C$. Then $\rho_{i+1,j+1} = \rho_{ij} = 0$, $\rho_{i+1,j} = \rho_{ij} + 1 = 1$ and $\rho_{i,j+1} = 0$, and so (sM) holds with strict inequality.
- (iii) $c_i + c'_j < C$. Then all four entries are 0, and so (sM) holds again with equality. ■

5. COMPARISON WITH OTHER METHODS

We compare our method to a class of PRN methods. In the PRN methods all units on the frame are independently assigned random numbers from the uniform distribution on $[0, 1]$, which are then retained permanently. The frame is sorted in the order of these PRNs. To select a sample, we choose a starting point $a \in [0, 1]$ and a direction (right or left). To select n units according to an SRSWOR design, we take the units corresponding to the first n PRNs to the right (or left) of a . This method is referred to as sequential SRSWOR. When the frame is updated, births are independently assigned newly generated PRNs and the dead units are discarded with their PRNs. Sequential SRSWOR for positive coordination of two surveys uses the same starting point (origin) and direction to select pairs of samples to maximize the expected overlap. For negative coordination (minimizing the overlap), distinct and preferably far apart origins a_1 and a_2 are chosen and the directions are either the same or opposite (Ohlsson (1995)). In collocated sampling a transformation is applied to evenly space the assigned PRNs. We use repeated collocated sampling as described by Srinath and Carpenter (1995). We view the full-stratum rotation method for negative coordination as in Ernst et al. (1998). Assume that a sequential SRSWOR sample has been selected on the first occasion by moving to the right. Let a_0 be the largest PRN associated with a unit in this sample. On the second occasion, a_0 becomes the starting point for the selection of the second sample. Note that a_0 is a random number, which may account for the selection bias of the method (Table 6 and

Proposition 1 of Ernst et al.). Some formal properties of the PRN methods, like the expected overlap, do not appear to have been studied.

To compare methods we conducted a simulation study and performed 100,000 repetitions using the data in Example 2. The three measures used for comparison are: E , V and the selection bias, where E is the expected overlap and V the corresponding variance. Let $Q' = (q'(s'))_{s' \in S'}$ be the design probability on the second occasion. Ernst et al. (1998) defined a selection bias by $\sum_{s' \in S'} b(s')q'(s')/n' - B/N'$, where $b(s')$ is the number of births in s' . It is the difference between the

expected sample proportion of births and the population proportion of births. Since our method requires (TP) to hold, we obtain an SRSWOR design on the second occasion with no selection bias. For positive coordination, the two compared methods perform equally well. They have no selection bias and attain the maximum expected overlap:

Table 5 – Positive coordination of two surveys

Method	E	V	Selection bias
Our method	1.5	0.45	0
Sequential SRSWOR	1.49	0.43	0

Our method for negative coordination performs very well. It has the smallest V and thus the overlap of each pair (s, s') selected by our method is likely to be close to E . The full-stratum rotation method has the smallest E but it does not produce an SRSWOR design on the second occasion. Repeated collocated sampling performs better than sequential SRSWOR but the associated sample size n' is a random variable.

Table 6 – Negative coordination of two surveys

Method	E	V	Selection bias
Our method	0.9	0.19	0
Sequential SRSWOR	0.91	0.52	0
PRN, full-stratum rotation	0.69	0.39	0.054
Repeated collocated	0.88	0.39	0

ACKNOWLEDGEMENT

The authors would like to thank Pierre Lavallée from Statistics Canada for having presented this material at the 2003 Annual Meeting of the SSC.

REFERENCES

- Ernst, L.R, Valliant, R. and Casady, R.J. (1998). Permanent and collocated random number sampling and the coverage of births and deaths. *Proceedings of the ASA Annual Meeting*, 457-462.
- Ernst, L.R. (1999). The maximization and minimization of sample overlap problems: a half century results. *Proceedings of the International Association of Survey Statisticians*.
- Hoffman, A.J. (1985). On greedy algorithms that succeed. In *Surveys in Combinatorics*, ed. Anderson, I. Cambridge U. Press, 97-112.
- Ohlsson, E. (1995). Coordination of samples using Permanent Random Numbers. In *Business Survey Methods*, eds. B.G. Cox, D. A. Binder, D. N. Chinnappa, A. Christianson, M. J. Colledge, and P. S. Kott. New York: John Wiley, 153-169.
- Raj, D.(1956). On the method of overlapping maps in sample surveys. *Sankhyā*, **17**, 89-98.
- Srinath, K.P. and Carpenter, R.M. (1995). Sampling methods for repeated business surveys. In *Business Survey Methods*, eds. B.G. Cox, D. A. Binder, D. N. Chinnappa, A. Christianson, M. J. Colledge, and P. S. Kott. New York: Wiley, 171-183.