Sankhyā : The Indian Journal of Statistics
1987, Volume 49, Series B, Pt. 3, pp. 283–306.

OPTIMAL INTEGRATION OF TWO OR THREE PPS SURVEYS WITH COMMON SAMPLE SIZE $n > 1$

By K. KRISHNAMOORTHY and SUJIT KUMAR MITRA

Indian Statistical Institute

SUMMARY. We consider a plan P for integration of k surveys for the special case of a sample size one for each survey and n independent repetitions of P so as to ensure a sample size n for each survey. We restrict our attention only to the plans of this type which we denote by P^n. A plan is called optimal if it minimizes the expected number of distinct units in the integrated survey. It is shown that when $k = 2$ and P is obtained through the Mitra-Pathak algorithm then P is indeed optimal in the above sense. The same is also true for $k = 3$ if $\theta_2 \leqslant 1$. We recall that $\theta_2 = \sum_{j=1}^{N} P_{(2)j}$ where P_{ij} is the probability of selecting the j-th population unit as specified by the i-th survey and $P_{(1)j} \leqslant P_{(2)j} \leqslant P_{(3)j}$ are the ordered values when P_{1j}, P_{2j} and P_{3j} are arranged in increasing order. When $\theta_2 > 1$ we identify a plan P which is optimal for $n = 1$ and has the following properties: P^n is optimal for sufficiently large sample size n. A sufficient condition is stated under which P^n is optimal for all sample sizes n. Numerical computation shows that even when P^n is not optimal the loss in using P^n is numerically insignificant.

1. Introduction

The algorithms for optimal integration of two or three surveys in Mitra and Pathak (1984) and the ones modified in Krishnamoorthy and Mitra (1986) to suit other cost functions essentially refer to optimality in the context of a sample size one drawn from each of the population. The object of the present paper is to present some results for optimal integration for a general sample size n when observations are drawn with probability proportional to size and with replacement. For two surveys the problem of optimal integration in the context of general sample size n was posed and satisfactorily solved by Keyfitz (1951) and Lahiri (1954) for a somewhat different cost function. Des Raj (1956) formulated this problem as a linear programming problem. This approach is further explored in Arthanari and Dodge (1981) and more recently in Causey, Cox and Ernst (1985) who apparently are unaware of the work of Arthanari and Dodge.

AMS (1980) *subject classification:* 62D05.

Key words and phrases: Algorithm, configuration, integrated survey, optimal integration, residual mass, majorisation.

We shall consider the case where the cost of the integrated survey depends exclusively on ν, the number of distinct population units that required to be studied, and is infact a linear function of ν with a positive slope. As we noted earlier this case has already been satisfactorily solved by Keyfitz (1951). We show that independent repetitions of Mitra-Pathak algorithm infact gives optimal results for a general sample size n. The argument infact extends itself for a fairly large class of situations encountered in respect of three surveys where we note that independent repetitions of Mitra-Pathak algorithm indeed produces optimal results. The same however may not be said about certain other classes and our research efforts in this paper are directed to these sub-classes. We are able to isolate two integration plans that broadly come under plans derivable through the Mitra-Pathak algorithm which seem to play a very crucial role here. One of them can be easily shown to be optimal for large sample sizes. We conjecture that between themselves the two will cover the entire range of sample sizes $n > 1$. Our Theorem 5 shows that when these two plans are identical, then the common plan is indeed optimal for all sample sizes.

2. Notations and some preliminary results

Consider a finite population of N units serially numbered $1, 2, \ldots, N$. Let $\mathcal{S}$ denote the set $\{1, 2, \ldots, N\}$. It is proposed to carry out k separate surveys on this population. Let P_{ij} denote the probability that the j-th population unit is included in the i-th survey and X_i denote the random variable associated with the i-th survey such that $P(X_i = j) = P_{ij}$ on $\mathcal{S} (1 \leqslant i \leqslant k, 1 \leqslant j \leqslant N)$ and $\sum_{j=1}^{N} P_{ij} = 1$. An integrated survey is a joint probability distribution of random variables $X_1, X_2, \ldots, X_k$ on $\mathcal{S}^k$, the k-th cartesian power of $\mathcal{S}$, which realizes for X_i the same marginal distribution as the one determined by the i-th survey. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$ be the observed sample in the integrated survey and $\nu(\boldsymbol{x})$ denote the number of distinct integers appearing in the k coordinates of $\boldsymbol{x}$. An integrated survey is called optimal if it minimizes $E\,\nu(\boldsymbol{X})$.

A matrix $((a_{ij}))_{k \times N}$ of nonnegative numbers will be called a configuration if the row totals are all equal. In the configuration of P_{ij}'s, let $P_{(i)j}$ denote the i-th smallest entry in j-th column and let

$$\theta_i = \sum_{j=1}^{N} P_{(i)j}, \qquad i = 1, 2, \ldots, k.$$

Further, let

$$\mathcal{S}_i = \{\boldsymbol{x} : \nu(\boldsymbol{x}) = i\}, i = 1, 2, \ldots, k.$$

We record here the definition of majorization and a theorem concerning the same which we shall make use of later in this paper. For $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{R}^n$, let the coordinates be arranged in a nondecreasing order and the ordered coordinate values be denoted by $x_{(1)}, x_{(2)} \ldots, x_{(n)}, x_{(1)} \leqslant x_{(2)} \leqslant \ldots < x_{(n)}$. The n-tuple $\boldsymbol{x}$ is said to be majorized by the n-tuple $\boldsymbol{y}$($\boldsymbol{y}$ majorizes $\boldsymbol{x}$) if

$$\sum_{j=i}^{n} x_{(j)} \leqslant \sum_{j=i}^{n} y_{(j)}, \quad i = 2, 3, \ldots, n,$$

and

$$\sum_{j=1}^{n} x_{(j)} = \sum_{j=1}^{n} y_{(j)}.$$

Theorem 1 : *Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{R}^n$. If $\boldsymbol{y}$ majorizes $\boldsymbol{x}$ then for all convex functions g,*

$$\sum_{i=1}^{n} g(x_i) \leqslant \sum_{i=1}^{n} g(y_i).$$

For a proof of Theorem 1, see Marhsall and Olkin (1979, page 115).

Consider a plan P for integration of k surveys for the special case of a sample size one for each survey. Let P_j denote the probability that the j-th population unit is selected for atleast one of the k surveys. We have seen in Mitra and Pathak (1984) that the expected number of distinct units is equal to $\sum_{j=1}^{N} P_j$. The following lemma can be similarly established.

Lemma 1 : *If the plan is independently repeated n times to achieve the desired sample size n for each survey then the expected number of distinct units in the integrated survey is given by*

$$E\,\nu_n = \sum_{j=1}^{N} (1-(1-P_j)^N).$$

Since we propose to consider only plans of this type any integration plan can henceforth be identified with the vector $\boldsymbol{P} = (P_1, P_2, \ldots, P_N)$. We have seen in Krishnamoorthy and Mitra (1986) that the vector $\boldsymbol{P}$ is not unique even for the optimal plans derived from Mitra-Pathak algorithms. Let $\mathcal{P}$ denote the class of such optimal integration plans $\boldsymbol{P}$.

The following lemma can be easily established.

Lemma 2 : *The set $\mathcal{P}$ is a closed convex set.*

We have seen in Mitra and Pathak (1984) and more explicitly in Krishnamoorthy and Mitra (1986) that for a plan belonging to $\mathcal{P}$

$$P_{(3)j} \leqslant P_j \leqslant P_{(3)j} + P_{(2)j} - P_{(1)j}$$

for every j.

The next lemma gives an upper bound for $P_{(3)j}$ in terms of θ_2.

Lemma 3 : $P_{(3)j} \leqslant 2-\theta_2$ *for all* j.

Proof: Lemma 3 is trivially true if $\theta_2 \leqslant 1$. Consider the case $\theta_2 > 1$. For some j, let

$$P_{(3)j}+\theta_2-1 > 1.$$

In the j-th column of the stochastic matrix of P_{ij}'s, assume without any loss of generality, that $P_{1j} = P_{(1)j}$, $P_{2j} = P_{(2)j}$ and $P_{3j} = P_{(3)j}$. Since $P_{(3)a} = \max(P_{1a}, P_{2a}, P_{3a})$ for all a,

$$\begin{aligned} \sum_{a \neq j} (P_{(3)a}-P_{(1)a}) & \\ \geqslant & \sum_{a \neq j} (P_{1a}-P_{(1)a}) \\ = & \sum_{a=1}^{N} (P_{1a}-P_{(1)a}) \\ = & 1-\theta_1. \end{aligned}$$

Adding this inequality with the previous one, we get

$$\theta_3+\theta_2-1-\theta_1+P_{(1)j} > 2-\theta_1$$

which implies that

$$P_{(1)j} > 3-\theta_3-\theta_2 = \theta_1$$

which is impossible since $\theta_1 = \sum_{a=1}^{N} P_{(1)a}$.

This completes the proof of Lemma 3.

In the following theorem we show that for any predetermined choice of probabilities of selection of the N units, subject to certain conditions there exists a corresponding optimal integration plan.

Theorem 2 : *Consider a stochastic matrix for three surveys for which* $\theta_2 > 1$. *Let* $e_1, e_2, \ldots, e_N$ *be numbers such that*

$$P_{(3)j} \leqslant e_j \leqslant P_{(3)j}+\min\{P_{(2)j}-P_{(1)j},\ \theta_2-1\}$$

and

$$\sum_{j=1}^{N} e_j = 2-\theta_1.$$

Then there exists an optimal integration plan for which $P_j = e_j$, $j = 1, 2, \ldots, N$.

Proof: By Lemma 3 it is seen that $e_j \leqslant 1$ for all j.

Let us consider the configuration as it stands after the smallest entries are zeroed out in all the columns. Each row total in this configuration is now equal to $1-\theta_1$. Assume without loss of generality that in the j-th column $P_{1j} = P_{(1)j}, P_{2j} = P_{(2)j}$ and $P_{3j} = P_{(3)j}$. Suppose that $P_{(2)j}-P_{(1)j} \geqslant \theta_2-1$ and $e_j = P_{(3)j}+\theta_2-1$. Then the condition $\sum_{j=1}^{N} e_j = 2-\theta_1$ implies that $e_i = P_{(3)i}, i = 1, 2, ..., N(i \neq j)$. In this case Mitra-Pathak algorithms can be applied so that $P_i = e_i, i = 1, 2, ..., N$. Let $P_{(2)i}-P_{(1)i} < \theta_2-1$ for all i, and $e_j = P_{(3)j}+P_{(2)j}-P_{(1)j}-\delta_j$. For P_j to be equal to e_j it is necessary that the points of the type (x, j, j), $x \neq j$, in $\mathcal{S}_2$ should have a total mass of $\delta_j = P_{(3)j}+P_{(2)j}-P_{(1)j}-e_j$. Out of the available masses in the configuration we have committed ourselves an amount δ_j from both $P_{(3)j}-P_{(1)j}$ and $P_{(2)j}-P_{(1)j}$. What remains, namely $P_{(3)j}-P_{(1)j}-\delta_j$ and $P_{(2)j}-P_{(1)j}-\delta_j$ we shall call them residual masses which will play a crucial role in determining the odd member x in the triplet (x, j, j). The condition $\sum_{j=1}^{N} e_j = 2-\theta_1$ is equivalent to $\sum_{j=1}^{N} \delta_j = 1-\theta_1$. To prove Theorem 2 it is therefore suffices to show that the available residual masses are just sufficient to fix all the odd members in this plan.

As in Theorem 4 of Krishnamoorthy and Mitra (1986), let Γ_{ik} denote the set of indices of those columns for which the i-th row contains k-th smallest column entry $(i, k = 1, 2, 3)$. The total demand for residual masses in row 1 is thus seen to be equal to $\sum_{j\epsilon\Gamma_{11}} \delta_j$ and the total committed mass in row 1 is equal to $\sum_{j\epsilon(\Gamma_{12}\cup\Gamma_{13})} \delta_j$. Since the first row total is $1-\theta_1$ the available residual mass in row 1 is equal to

$$
\begin{aligned}
1-\theta_1-\sum_{j\epsilon(\Gamma_{12}\cup\Gamma_{13})} \delta_j &= \sum_{j=1}^{N} \delta_j - \sum_{j\epsilon(\Gamma_{12}\cup\Gamma_{13})} \delta_j \\
&= \sum_{j\epsilon(\Gamma_{11}\cup\Gamma_{12}\cup\Gamma_{13})} \delta_j - \sum_{j\epsilon(\Gamma_{12}\cup\Gamma_{13})} \delta_j \\
&= \sum_{j\epsilon\Gamma_{11}} \delta_j.
\end{aligned}
$$

The available residual masses in row 1 is thus just sufficient to meet the demand.

The same argument applies to other rows.

Our next lemma shows that $\mathcal{P}$ is essentially a complete class in the sense that for any integration plan $\boldsymbol{P}$ that is outside the class $\mathcal{P}$ there exists a plan $\boldsymbol{P}^*$ in the class $\mathcal{P}$ such that $P_j^* \leqslant P_j$ for every j.

Lemma 4 : *Consider a stochastic matrix of three surveys with* $\theta_2 > 1$. *For every plan* $\mathbf{P} \notin \mathcal{P}$ *there exists a plan* $\mathbf{P}^*$ *in* $\mathcal{P}$ *such that*

$$P_j^* \leqslant P_j$$

for every j *and* $P_j^* < P_j$ *for some* j.

Proof: Note that, when $\theta_2 > 1$, for an optimal plan $E\nu_n = 2-\theta_1$. Since the plan $\mathbf{P}$ is not an optimal one

$$\sum_{j=1}^{N} P_j > 2-\theta_1$$

and for some j, $P_j > P_{(3)j}$. Reduce those P_j's (for which $P_j > P_{(3)j}$) to $P_{(3)j}$ or to some $a_j > P_{(3)j}$ such that the new P_j's (call them P_j^*) add upto $2-\theta_1$. Then, Theorem 2 ensures the existence of the plan $\mathbf{P}^* = (P_1^*, P_2^*, \ldots, P_N^*)$, $P_j^* \leqslant P_j$ for all j and $P_j^* < P_j$ for some j.

Let $\min\{P_{(2)j}-P_{(1)j}, \theta_2-1\}$ be denoted by Δ_j.

Lemma 5 : *Consider a stochastic matrix for 3 surveys with* $\theta_2 > 1$. $\mathbf{P} = (P_1, P_2, \ldots, P_N)$ *is an extreme point in* $\mathcal{P}$ *if and only if* $P_j = P_{(3)j}$ or $P_j = P_{(3)j}+\Delta_j$ *for all but at most one* j. *Further, if* J_1 *denotes the set of integers for which* $P_j = P_{(3)j}$ *and* J_2 *denotes the set of integers for which* $P_j = P_{(3)j}+\Delta_j$ *then*

$$\sum_{j \epsilon J_1} (P_{(2)j}-P_{(1)j}) \leqslant 1-\theta_1$$

and

$$\sum_{j \epsilon J_2} (P_{(2)j}-P_{(1)j}) \leqslant \theta_2-1.$$

Proof: Let $\mathbf{P} = (P_1, P_2, \ldots, P_N)$ be an optimal integration plan such that $J_3 = \mathcal{S}-(J_1 \bigcup J_2)$ contains at most one integer. Without loss of generality assume that $J_1 = \{2, 3, \ldots, m\}$, $J_2 = \{m+1, m+2, \ldots, N\}$ and $J_3 = \{1\}$. Let $\mathbf{P}^*$ and $\mathbf{P}'$ be two vectors in $\mathcal{P}$ such that

$$\alpha \mathbf{P}^* + (1-\alpha)\mathbf{P}' = \mathbf{P}$$

for some α in $(0,1)$. Since $P_j = P_{(3)j}$ $(2 \leqslant j \leqslant m)$ and $P_j = P_{(3)j}+\Delta_j$, $(m+1 \leqslant j \leqslant N)$, $\alpha P_j^*+(1-\alpha)P_j' = P_j$ implies that

$$P_j^* = P_j' = P_j, \quad j = 2, 3, \ldots, N. \qquad \ldots \quad (1)$$

Again as $\sum_1^N P_j^* = \sum_1^N P_j' = \sum_1^N P_j$, (1) implies that

$$P_1^* = P_1' = = P_1.$$

Thus, we have

$$P^* = P' = P$$

and so P is an extreme point.

We now suppose that J_3 contains more than one integer, say, $J_3 = \{1, 2, ..., i\}$. Then, write

$$P_1 = P_{(3)1}+\varepsilon_1, \; 0 < \varepsilon_1 < \Delta_1.$$

$$P_2 = P_{(3)2}+\varepsilon_2, \; 0 < \varepsilon_2 < \Delta_2.$$

Choose the numbers ϕ_1 and ϕ_2 such that

$$\varepsilon_1 < \phi_1 < \min(\Delta_1, \varepsilon_1+\varepsilon_2)$$

$$\varepsilon_2 < \phi_2 < \min(\Delta_2, \varepsilon_1+\varepsilon_2)$$

and define

$$P_1^* = P_{(3)1}+\phi_1, \; P_2^* = P_{(3)2}+\varepsilon_2-(\phi_1-\varepsilon_1)$$

$$P_1' = P_{(3)1}+\varepsilon_1-(\phi_2-\varepsilon_2), \; P_2' = P_{(3)2}+\phi_2$$

and

$$\boldsymbol{P}^* = (P_1^*, P_2^*, P_3, ..., P_N)$$

$$\boldsymbol{P}' = (P_1', P_2', P_3, ..., P_N).$$

Clearly the plans $\boldsymbol{P}^*$ and $\boldsymbol{P}'$ belong to $\mathcal{P}$ and their existence is guaranteed by Theorem 2. $\boldsymbol{P}$ can be written as

$$\boldsymbol{P} = \alpha\boldsymbol{P}^*+(1-\alpha)\boldsymbol{P}'$$

where

$$\alpha = (\phi_2-\varepsilon_2)/(\phi_1-\varepsilon_1+\phi_2-\varepsilon_2).$$

Thus, if J_3 contains more than one integer, P can not be an extreme point. We next show that

$$\sum_{j \in J_2} (P_{(2)j}-P_{(1)j}) \leqslant \theta_2-1.$$

Let $l \in J_3$ and

$$P_l = P_{(3)l}+\varepsilon_l, \quad 0 < \varepsilon_l < P_{(2)l}-P_{(1)l}.$$

Since $\boldsymbol{P}$ is an optimal integration plan and $\theta_3+\theta_2+\theta_1 = 3$.

$$\sum_{j=1}^{N} P_j = 2-\theta_1 = \theta_3+\theta_2-1. \qquad \text{... (2)}$$

Write

$$\sum_{j=1}^{N} P_j = \sum_{j \in J_1} P_{(3)j}+\sum_{j \in J_2} (P_{(3)j}+\Delta_j)+P_{(3)l}+\varepsilon_l.$$

$$= \sum_{j=1}^{N} P_{(3)j}+\sum_{j \in J_2} \Delta_j+\varepsilon_l$$

$$= \theta_3+\sum_{j \in J_2} \Delta_j+\varepsilon_l \qquad \text{... (3)}$$

Equations (2) and (3) imply that

$$\sum_{j \in J_2} \Delta_j + \varepsilon_l = \theta_2 - 1. \quad \ldots \quad (4)$$

As $\varepsilon_l \geqslant 0$ ($\varepsilon_l = 0 \Leftrightarrow J_3$ is an empty set) from (4) we have

$$\sum_{j \in J_2} \Delta_j \leqslant \theta_2 - 1.$$

Hence $\Delta_j = P_{(2)j} - P_{(1)j}$ for each $j \in J_2$ and

$$\sum_{j \in J_2} \{P_{(2)j} - P_{(1)j}\} \leqslant \theta_2 - 1.$$

Similarly, writing

$$\sum_{j \in J_1} (P_{(2)j} - P_{(1)j}) = \sum_{j=1}^{N} (P_{(2)j} - P_{(1)j}) - \left\{ \sum_{j \in J_2} (P_{(2)j} - P_{(1)j}) + \varepsilon_l \right\} - (P_{(2)l} - P_{(1)l} - \varepsilon_l)$$

$$= \theta_2 - \theta_1 - (\theta_2 - 1) - (P_{(2)l} - P_{(1)l} - \varepsilon_l) \text{ (using (4))}$$

$$= 1 - \theta_1 - (P_{(2)l} - P_{(1)l} - \varepsilon_l)$$

and using the relation $(P_{(2)l} - P_{(1)l} - \varepsilon_l) \geqslant 0$, we prove

$$\sum_{j \in J_1} (P_{(2)j} - P_{(1)j}) \leqslant 1 - \theta_1.$$

Lemma 6 : *Consider a stochastic matrix for which $\theta_2 > 1$. Let $\boldsymbol{P} = (P_1, P_2, \ldots, P_N)$ be an extreme point in $\mathcal{P}$ such that $P_1 \leqslant P_2 \leqslant \ldots \leqslant P_N$. For any i, $1 \leqslant i \leqslant N$,*

$$\sum_{j=i}^{N} P_j \leqslant \sum_{j=i}^{N} P_{(3)j} + \theta_2 - 1 \quad \ldots \quad (5)$$

and

$$\sum_{j=1}^{i} P_j \geqslant \sum_{j=1}^{i} (P_{(3)j} + P_{(2)j} - P_{(1)j}) - (1 - \theta_1). \quad \ldots \quad (6)$$

Proof : Since $\boldsymbol{P}$ is an optimal plan, the relation $\sum_{j=1}^{N} P_j = 2 - \theta_1 = \theta_3 + \theta_2 - 1$ and the inequality $P_j \geqslant P_{(3)j}$ (for all j) imply that $P_j = P_{(3)j} + \alpha_j$ where $0 \leqslant \alpha_j \leqslant \min\{P_{(2)j} - P_{(1)j}, \theta_2 - 1\}$ for all j and $\sum_{j=1}^{N} \alpha_j = \theta_2 - 1$. Therefore we have

$$\sum_{j=i}^{N} P_j = \sum_{j=i}^{N} (P_{(3)j} + \alpha_j) \leqslant \sum_{j=i}^{N} P_{(3)j} + \theta_2 - 1.$$

Define J_1, J_2 and J_3 as in Lemma 5 and let $A_i = \{1, 2, \ldots, i\}$. Then

$$\sum_{j \in A_i} P_j = \sum_{j \in (A_i \cap J_1)} P_j + \sum_{j \in (A_i \cap J_2)} P_j + \sum_{j \in (A_i \cap J_3)} P_j$$

$$= \sum_{j \in (A_i \cap J_1)} P_{(3)j} + \sum_{j \in (A_i \cap J_2)} (P_{(3)j} + P_{(2)j} - P_{(1)j})$$

$$+ \sum_{j \in (A_i \cap J_3)} (P_{(3)j} + P_{(2)j} - P_{(1)j}) - (1 - \theta_1 - \sum_{j \in J_1} (P_{(2)j} - P_{(1)j}))$$

$$\geqslant \sum_{j \in A_i} (P_{(3)j} + P_{(2)j} - P_{(1)j}) - (1 - \theta_1)$$

since J_3 contains at most one integer and

$$\sum_{j \in J_1} (P_{(2)j} - P_{(1)j}) \geqslant \sum_{j \in (A_i \cap J_1)} (P_{(2)j} - P_{(1)j}).$$

3. Main results

In the initial configuration of P_{ij}'s for three surveys let $\theta_2 \leqslant 1$.

The following theorem shows that the plan derived through the Mitra-Pathak algorithm, which is optimal for a sample size one, is also optimal for a general sample size n when observations are drawn with probability proportional to size and with replacement.

Theorem 3 : *For $\theta_2 \leqslant 1$, the plan $\boldsymbol{P}^* = (P_1^*, \ldots, P_N^*)$ obtained through the Mitra-Pathak algorithm is optimal for a general sample size n.*

Proof : Since for any plan $\boldsymbol{P}$, $P_j \geqslant P_{(3)j}$, $j = 1, 2, \ldots, N$, from Lemma 1, we have

$$E\nu_n = \sum_{j=1}^{N} (1-(1-P_j)^n) \geqslant \sum_{j=1}^{N} (1-(1-P_{(3)j})^n).$$

We also know that Mitra-Pathak algorithm applied to the configuration of P_{ij}'s, when $\theta_2 \leqslant 1$, gives a plan $\boldsymbol{P}^*$ such that $P_j^* = P_{(3)j}$ for all j. Therefore $\boldsymbol{P}^*$ is optimal.

We next consider the case $\theta_2 > 1$. We describe here two plans, namely, $\boldsymbol{P}_t = (P_{t1}, \ldots, P_{tN})$ and $\boldsymbol{P}_b = (P_{b1}, \ldots, P_{bN})$ which can be derived through the Mitra-Pathak algorithms. Later in Theorem 5 we show that if the two plans are identical, that is, $\boldsymbol{P}_t = \boldsymbol{P}_b = \boldsymbol{P}$, then the common plan $\boldsymbol{P}$ is optimal for a general sample size n.

Plan $\boldsymbol{P}_t$: For a $j \in \mathcal{S}$, we determine the value of P_j as follows :

Consider the initial configuration $((P_{ij}))_{3\times N}$. Let t_N denote the column of $((P_{ij}))_{3\times N}$ which maximizes

$$\{P_{(3)j}+\min(P_{(2)j}-P_{(1)j},\ \theta_2-1)\}$$

over $j \in \mathcal{S}$. Define

$$P_{t_N} = P_{(3)t_N}+P_{(2)t_N}-P_{(1)t_N}, \text{if } P_{(2)t_N}-P_{(1)t_N} < \theta_2-1$$
$$= P_{(3)t_N}+\theta_2-1, \qquad \text{otherwise.}$$

Similarly, if t_k denotes the column which attains

$$\xi_{t_k} = \max_{\substack{1\leqslant u\leqslant N \\ u\neq t_{k+1},\ldots,u\neq t_N}} \left[P_{(3)u}+\min\left\{P_{(2)u}-P_{(1)u}, \theta_2-1-\sum_{j=k+1}^{N}(P_{(2)t_j}-P_{(1)t_j})\right\}\right]$$

then define

$$P_{t_k} = \xi_{t_k}$$

as long as
$$\theta_2-1-\sum_{j=k+1}^{N}(P_{(2)t_j}-P_{(1)t_j}) \geqslant 0.$$

Let m_t be an integer such that

$$\theta_2-1-\sum_{j=m_t+2}^{N}(P_{(2)t_j}-P_{(1)t_j}) \geqslant 0$$

and

$$\theta_2-1-\sum_{j=m_t+1}^{N}(P_{(2)t_j}-P_{(1)t_j}) < 0.$$

For $1 \leqslant k \leqslant m_t$, if t_k denotes the column which attains

$$\xi_{t_k} = \max_{\substack{1\leqslant j\leqslant N \\ j\neq t_N,\ldots,j\neq t_{k+1}}} \{P_{(3)j}\}$$

then define

$$P_{t_k} = \xi_{t_k} = P_{(3)t_k}.$$

Thus, we construct the plan $\boldsymbol{P}_t$ as $\boldsymbol{P}_t = (P_{t_1}, P_{t_2}, \ldots, P_{t_N})$ where $P_{t_1} \leqslant P_{t_2} \leqslant \ldots \leqslant P_{t_N}$ and $\sum_{j=1}^{N} P_{t_j} = 2-\theta_1$.

Plan $\boldsymbol{P}_b$: We here determine the values of P_j's as follows :

Let b_1 denote the column of $((P_{ij}))_{3\times N}$ which attains

$$\xi_{b_1} = \min_{1\leqslant j\leqslant N} \{P_{(3)j}+P_{(2)j}-P_{(1)j}-\min(P_{(2)j}-P_{(1)j},\ 1-\theta_1)\}.$$

Define

$$P_{b_1} = \xi_{b_1}.$$

Since $P_{(2)j}-P_{(1)j} \leqslant 1-\theta_1$ for all j, $P_{b_1} = \min_{1 \leqslant j \leqslant N} \{P_{(3)j}\}$. If b_k denotes the column which attains

$$\xi_{b_k} = \min_{\substack{1 \leqslant u \leqslant N \\ u \neq b_1, \ldots, u \neq b_{k-1}}} \left[P_{(3)u}+P_{(2)u}-P_{(1)u}-\min \left\{ P_{(2)u}-P_{(1)u}, \right.\right.$$

$$\left.\left. 1-\theta_1-\sum_{j=1}^{k-1} (P_{(2)b_j}-P_{(1)b_j}) \right\} \right]$$

then define

$$P_{b_k} = \xi_{b_k}$$

as long as $1-\theta_1-\sum_{j=1}^{k-1} (P_{(2)b_j}-P_{(1)b_j}) \geqslant 0$.

Let m_b denote an integer such that

$$1-\theta_1-\sum_{j=1}^{m_b} (P_{(2)b_j}-P_{(1)b_j}) \geqslant 0$$

and

$$1-\theta_1-\sum_{j=1}^{m_b+1} (P_{(2)b_j}-P_{(1)b_j}) < 0.$$

For $m_b+2 \leqslant k \leqslant N$, if b_k denotes the column which attains

$$\xi_{b_k} = \min_{\substack{1 \leqslant j \leqslant N \\ j \neq b_1, \ldots, j \neq b_{k-1}}} \{P_{(3)j}+P_{(2)j}-P_{(1)j}\}$$

then define

$$P_{b_k} = \xi_{b_k} = P_{(3)b_k}+P_{(2)b_k}-P_{(1)b_k}.$$

Thus, we construct the plan $\boldsymbol{P}_b$ as $\boldsymbol{P}_b = (P_{b_1}, \ldots, P_{b_N})$ where

$$P_{b_1} \leqslant P_{b_2} \leqslant \cdots \leqslant P_{b_N} \text{ and } \sum_{j=1}^{N} P_{b_j} = 2-\theta_1.$$

Theorem 2 ensures that the plans $\boldsymbol{P}_t$ and $\boldsymbol{P}_b$ can be derived through the Mitra-Pathak algorithms.

Suppose that the two plans are identical. That is,

$$\boldsymbol{P}_b = \boldsymbol{P}_t = \boldsymbol{P}^*.$$

Without loss of generality assume that

$$\boldsymbol{P}^* = (P_1^*, P_2^*, \ldots, P_N^*) \text{ and } m_t = m_b = m.$$

Then

$$P^*_j = \begin{cases} P_{(3)j}, & 1 \leqslant j \leqslant m \\ P_{(3)j}+q_j, & j = m+1 \\ P_{(3)j}+P_{(2)j}-P_{(1)j}, & m+2 \leqslant j \leqslant N \end{cases}$$

where
$$q_j = \theta_2 - 1 - \sum_{a=m+2}^{N} (P_{(2)a}-P_{(1)a})$$
$$= \sum_{a=1}^{m+1} (P_{(2)a}-P_{(1)a})-(1-\theta_1).$$

Theorem 4 : *Consider a stochastic matrix for 3 surveys with* $\theta_2 > 1$. *Let* $\mathbf{P}_l = (P_{l_1}, P_{l_2} \ldots, P_{l_N})$ *be an extreme point in* $\mathcal{P}$ *such that* $P_{l_1} \leqslant P_{l_2} \leqslant \ldots \leqslant P_{l_N}$. *Then*

$$\sum_{j=i}^{N} P^*_j \geqslant \sum_{j=i}^{N} P_{l_j},\ i = m+2, m+3, \ldots, N \qquad \ldots (7)$$

and
$$\sum_{j=1}^{i} P^*_j \leqslant \sum_{j=1}^{i} P_{l_j},\ i = 1, 2, \ldots, m. \qquad \ldots (8)$$

where P^*_j *is the j-th component of* $\mathbf{P}^*$.

Proof : In order to save the space and avoid notational complexity we here prove only the particular case

$$\sum_{j=i}^{N} P^*_j \geqslant \sum_{j=i}^{N} P_{l_j},\ i = N-2, N-1, N \qquad \ldots (9)$$

of (7). The argument for proving (7) is exactly similar to the one for (9). Write

$$\eta_j = 1-\theta_1- \sum_{a=1}^{j} (P_{(2)a}-P_{(1)a}),\ \ j = 1, 2, \ldots, m$$

and
$$\gamma_j = \sum_{a=j}^{N} (P_{(2)a}-P_{(1)a}), \qquad j = m+2, \ldots, N.$$

We first prove that
$$P^*_N+P^*_{N-1} \geqslant P_{l_N}+P_{l_{N-1}}. \qquad \ldots (10)$$

Since $P^*_N = P_{l_N}$, it follows from the definition of P_{l_N} that

$$P^*_N \geqslant P_{l_N}. \qquad \ldots (11)$$

If $P^*_{N-1} \geqslant P_{l_{N-1}}$, trivially (10) holds and so we assume that

$$P^*_{N-1} < P_{l_{N-1}}. \qquad \dots \quad (12)$$

Let $A = \{1, 2, \dots, m+1\}$, $B_i = \{i, i+1, \dots, N\}$ and $D_i = \mathcal{S}-(A \cup B_i)$. Since $P_{t_{N-1}} = P^*_{N-1}$, the definition of $P_{t_{N-1}}$ implies that

$$P^*_{N-1} \geqslant P_{(3)j}+P_{(2)j}-P_{(1)j} \geqslant P_j$$

for any $j \in D_N$. So the set $\{l_N, l_{N-1}\} \not\subset D_N$ otherwise (12) will not hold.

Case i : Without loss of generality assume that

$$l_{N-1} \in A \text{ and } l_N = N.$$

For any $l_i \in A$, if $P^*_{N-1} < P_{l_i}$, it follows from the definition of $P_{t_{N-1}}$ that

$$P_{(2)l_i}-P_{(1)l_i} > \theta_2-1-\gamma_N$$

and

$$P^*_{N-1} = P_{t_{N-1}} \geqslant P_{(3)l_i}+\theta_2-1-\gamma_N.$$

Therefore, (12) implies that

$$P^*_{N-1} \geqslant P_{(3)l_{N-1}}+\theta_2-1-\gamma_N.$$

Since $l_N = N$, $P_{(3)N} = P_{(3)l_N}$ and we have

$$P^*_{N-1}+P_{(3)N} \geqslant P_{(3)l_{N-1}}+P_{(3)l_N}+\theta_2-1-\gamma_N$$

which implies

$$P^*_{N-1}+P^*_N \geqslant P_{l_{N-1}}+P_{l_N}$$

because $P_{(3)N}+\gamma_N = P^*_N$ and from Lemma 6,

$$P_{(3)l_{N-1}}+P_{(3)l_N}+\theta_2-1 \geqslant P_{l_N}+P_{l_{N-1}}.$$

Case ii : Let $\{l_N, l_{N-1}\} \subset A$ and without loss of generality assume that $l_{N-1} = i < j = l_N$. Suppose that

$$P_{(3)N} < P_{(3)i}.$$

Since $P^*_N = P_{b_N}$, the definition of P_{b_N} implies that

$$P_{(3)N}+P_{(2)N}-P_{(1)N}-\eta_{i-1} \geqslant P_{(3)i} \qquad \dots \quad (13)$$

where

$$\eta_{i-1} \geqslant P_{(2)i}-P_{(1)i}+\eta_{j-1} \quad (\text{because } i, j \in A \text{ and } i < j). \qquad \dots \quad (14)$$

Adding (13) and (14), we get

$$P_N^* \geqslant P_{(3)i}+P_{(2)i}-P_{(1)i}+\eta_{j-1}. \quad \dots \quad (15)$$

If $j \leqslant m$, then

$$P_{N-1}^* \geqslant P_j^* = P_{(3)j} \quad \dots \quad (16)$$

and

$$\eta_{j-1} \geqslant P_{(2)j}-P_{(1)j} \quad \dots \quad (17)$$

Combining (15), (16) and (17), we get

$$P_N^*+P_{N-1}^* \geqslant P_{(3)j}+P_{(2)j}-P_{(1)j}+P_{(3)i}+P_{(2)i}-P_{(1)i} \geqslant P_{l_N}+P_{l_{N-1}}.$$

If $j = m+1$, then

$$P_{N-1}^* \geqslant P_j^* = P_{(3)j}+P_{(2)j}-P_{(1)j}-\eta_{j-1} \quad \dots \quad (18)$$

and adding (15) to (18), we get

$$P_N^*+P_{N-1}^* \geqslant P_{l_N}+P_{l_{N-1}}.$$

We now suppose that

$$P_{(3)N} \geqslant P_{(3)i} \ (i = l_{N-1}). \quad \dots \quad (19)$$

The assumptions that $l_N \epsilon A$ and $P_{N-1}^* < P_{l_N}$ imply that

$$P_{N-1}^* \geqslant P_{(3)l_N}+\theta_2-1-\gamma_N^*. \quad \dots \quad (20)$$

Thus, as in the case *i*, adding (19) and (20), we obtain

$$P_N^*+P_{N-1}^* \geqslant P_{l_N}+P_{l_{N-1}}.$$

We next show that

$$P_{N-2}^*+P_{N-1}^*+P_N^* \geqslant P_{l_{N-2}}+P_{l_{N-1}}+P_{l_N}. \quad \dots \quad (21)$$

(21) holds obviously if $P_{N-2}^* \geqslant P_{l_{N-2}}$. So, we let

$$P_{N-2}^* < P_{l_{N-2}}. \quad \dots \quad (22)$$

Notice that, under the assumption (22), none of the units l_N, l_{N-1} l_{N-2} is from the set D_{N-1}.

Case i (a): With loss of generality assume that

$$l_N = N, \ l_{N-1} = N-1 \text{ and } l_{N-2} \epsilon A.$$

Since $P_{t_{N-2}} = P_{N-2}^* < P_{l_{N-2}}$, the definition of $P_{t_{N-2}}$ implies that

$$P_{N-2}^* \geqslant P_{(3)l_{N-2}}+\theta_2-1-\gamma_{N-1} \quad \dots \quad (23)$$

and from the above assumption

$$P_{(3)N}+P_{(3)N-1} = P_{(3)l_N}+P_{(3)l_{N-1}}. \quad \dots \quad (24)$$

Adding (23) and (24), we get

$$P^*_{N-2}+P_{(3)N}+P_{(3)N-1}+\gamma_{N-1} \geqslant P_{(3)l_N}+P_{(3)l_{N-1}}+P_{(3)l_{N-2}}+\theta_2-1$$

which implies

$$P^*_{N-2}+P^*_{N-1}+P^*_N \geqslant P_{l_{N-2}}+P_{l_{N-1}}+P_{l_N}$$

since $P_{(3)N}+P_{(3)N-1}+\gamma_{N-1} = P^*_N+P^*_{N-1}$ and from Lemma 6

$$P_{(3)l_N}+P_{(3)l_{N-1}}+P_{(3)l_{N-2}}+\theta_2-1 \geqslant P_{l_N}+P_{l_{N-1}}+P_{l_{N-2}}.$$

Case ii (a): We here assume that

$$l_N = N \text{ and } \{l_{N-1}, l_{N-2}\} \subset A$$

$l_N = N$ implies that $P^*_N \geqslant P_{l_N}$.

Using a similar argument as given in case (*ii*) it can be easily shown that

$$P^*_{N-1}+P^*_{N-2} \geqslant P_{l_{N-1}}+P_{l_{N-2}}$$

and so

$$P^*_N+P^*_{N-1}+P^*_{N-2} \geqslant P_{l_{N-1}}+P_{l_{N-2}}+P_{l_N}.$$

Case iii: Let $\{l_{N-2}, l_{N-1}, l_N\} \epsilon A$ and without loss of generality, let $l_{N-2} = i$, $l_{N-1} = j$, $l_N = k$ and $i < j < k$. Suppose that

$$P_{(3)N} < P_{(3)i}.$$

Since $P^*_N = P_{b_N}$, it follows from the definition of P_{b_N} that

$$P_{(3)N}+P_{(2)N}-P_{(1)N}-\eta_{i-1} \geqslant P_{(3)i} \quad \dots \quad (25)$$

where

$$\eta_{i-1} \geqslant P_{(2)i}-P_{(1)i}+P_{(2)j}-P_{(1)j}+\eta_{k-1} \quad \dots \quad (26)$$

(because $i, j, k \,\epsilon\, A$, $i < j < k$).

Adding (25) and (26), we have

$$P^*_N \geqslant P_{(3)i}+P_{(2)i}-P_{(1)i}+P_{(2)j}-P_{(1)j}+\eta_{k-1} \quad \dots \quad (27)$$

for $k \leqslant m$,

$$P^*_{N-1} \geqslant P^*_j = P_{(3)j}$$

$$P^*_{N-2} \geqslant P^*_k = P_{(3)k} \quad \dots \quad (28)$$

and

$$\eta_{k-1} \geqslant P_{(2)k}-P_{(1)k}.$$

From the inequalities given in (27) and (28), it can be easily seen that

$$P^*_N+P^*_{N-1}+P^*_{N-2} \geqslant P_{l_N}+P_{l_{N-1}}+P_{l_{N-2}}.$$

For $k = m+1$,

$$P^*_{N-1} \geqslant P^*_j = P_{(3)j}$$

$$P^*_{N-2} \geqslant P^*_k = P_{(3)k}+P_{(2)k}-P_{(1)k}-\eta_{k-1} \quad \dots \quad (29)$$

and combining (27) and (29), we get

$$P^*_N+P^*_{N-1}+P^*_{N-2} \geqslant P_{l_N}+P_{l_{N-1}}+P_{l_{N-2}}.$$

Suppose that $P_{(3)N} \geqslant P_{(3)i}$ and $P_{(3)N-1} < P_{(3)j}$.

Again using a similar argument as given above, it can be easily shown that

$$P^*_{N-1} \geqslant P_{(3)j}+P_{(2)j}-P_{(1)j}+\eta_{k-1}$$

and for $k \leqslant m+1$

$$P^*_{N-1}+P^*_{N-2} \geqslant P_{l_{N-1}}+P_{l_{N-2}}. \quad \dots \quad (30)$$

Thus, from (11) and (30), we have

$$P^*_N+P^*_{N-1}+P^*_{N-2} \geqslant P_{l_N}+P_{l_{N-1}}+P_{l_{N-2}}.$$

Finally, as the case $P_{(3)N} \geqslant P_{(3)i}$ and $P_{(3)N-1} \geqslant P_{(3)j}$ is similar to case $i(a)$, we omit the proof of this case.

We now show that

$$\sum_{j=1}^{i} P^*_j \leqslant \sum_{j=1}^{i} P_{l_j}, \qquad i = 1, 2, \dots, m.$$

For the same reasons given earlier, we prove only the particular case

$$\sum_{j=1}^{i} P^*_j \geqslant \sum_{j=1}^{i} P_{l_j}, \qquad i = 1, 2, 3.$$

We first prove that

$$P^*_1+P^*_2 \leqslant P_{l_1}+P_{l_2} \quad \dots \quad (31)$$

Since $P^*_1 = P_{b_1}$, from the definition of P_{b_1} it follows that

$$P^*_1 \leqslant P_{l_1}. \quad \dots \quad (32)$$

Inequality (31) holds obviously when $P_2^* \leqslant P_{l_2}$ and so let

$$P_2^* > P_{l_2}. \quad \ldots \quad (33)$$

Let $A_i = \{1, 2, \ldots, i\}$. Since $P_2^* = P_{b_2}$ is the minimum of $P_{(3)j}$'s $(2 \leqslant j \leqslant m)$, under the assumption (33) the set $\{l_1, l_2\} \not\subset \{2, 3, \ldots, m\}$.

Case I: Without loss of generality, let $l_1 = 1$, $l_2 \in B_{m+1}$. For any $l_i \in B_{m+1}$, if $P_{b_2} > P_{l_i}$, then the definition of P_{b_2} implies that

$$P_{(2)l_i} - P_{(1)l_i} > 1 - \theta_1 - (P_{(2)1} - P_{(1)1})$$

and

$$P_{b_2} \leqslant P_{(3)l_i} + P_{(2)l_i} - P_{(1)l_i} - (1 - \theta_1 - P_{(2)1} + P_{(1)1}).$$

Since $P_{b_2} = P_2^*$, (33) implies that

$$P_2^* \leqslant P_{(3)l_2} + P_{(2)l_2} - P_{(1)l_2} - (1 - \theta_1 - P_{(2)1} + P_{(1)1}). \quad \ldots \quad (34)$$

As $\quad l_1 = 1,\ P_{(3)l_1} = P_{(3)1},\ P_{(2)1} - P_{(1)1} = P_{(2)l_1} - P_{(1)l_i}. \quad \ldots \quad (35)$

Combining (34) and (35), we get

$$P_2^* + P_{(3)1} \leqslant P_{(3)l_1} + P_{(3)l_2} + \sum_{j=1}^{2} P_{(2)l_j} - P_{(1)l_j} - (1 - \theta_1) \quad \ldots \quad (36)$$

which implies that

$$P_2^* + P_1^* \leqslant P_{l_1} + P_{l_2}$$

since $P_1^* = P_{(3)1}$ and from Lemma (6) the r.h.s. of (36) is less than or equal to $P_{l_1} + P_{l_2}$.

Case II: Let $\{l_1, l_2\} \in B_{m+1}$. Without loss of generality assume that $l_1 = i < j = l_2$. We first suppose that

$$P_{(2)1} - P_{(1)1} \geqslant \theta_2 - 1 - \sum_{a=j+1}^{N} (P_{(2)a} - P_{(1)a})$$

$$= \theta_2 - 1 - \gamma_{j+1}.$$

Since $P_1^* = P_{t_1}$, according to the plan $\boldsymbol{P}_t$, we have

$$P_{(3)1} + \theta_2 - 1 - \gamma_{j+1} \leqslant P_{(3)j} + P_{(2)j} - P_{(1)j} \quad \ldots \quad (37)$$

and as $i \in B_{m+1}$,

$$P_2^* \leqslant P_i^* = P_{(3)i} + P_{(2)i} - P_{(1)i}. \quad \ldots \quad (38)$$

It follows from (37) and (38) that

$$P_1^*+P_2^* \leqslant P_{(3)j}+P_{(3)i}+P_{(2)j}-P_{(1)j}+P_{(2)i}-P_{(1)i}-(\theta_2-1-\gamma_{j+1}). \quad \ldots \quad (39)$$

If $i \geqslant m+2$, then

$$P_{(2)j}-P_{(1)j}+P_{(2)i}-P_{(1)i} \leqslant \theta_2-1-\gamma_{j+1}$$

and so from (39), we get

$$P_1^*+P_2^* \leqslant P_{(3)j}+P_{(3)i} \leqslant P_{l_2}+P_{l_1}.$$

If $i = m+1$, then

$$P_2^* \leqslant P_i^* = P_{(3)i}+\theta_2-1-\gamma_{i+1} \quad \ldots \quad (40)$$

and

$$\theta_2-1-\gamma_{j+1} \geqslant P_{(2)j}-P_{(1)j}+(\theta_2-1-\gamma_{i+1}). \quad \ldots \quad (41)$$

Adding (37) and (40), and using (41), we obtain

$$P_{(3)1}+P_2^* = P_1^*+P_2^* \leqslant P_{(3)j}+P_{(3)i} \leqslant P_{l_2}+P_{l_1}.$$

We now suppose that

$$P_{(2)1}-P_{(1)1} < \theta_2-1-\gamma_{j+1}.$$

According to the plan $\boldsymbol{P}_t$,

$$P_{(3)1}+P_{(2)1}-P_{(1)1} \leqslant P_{(3)j}+P_{(2)j}-P_{(1)j}. \quad \ldots \quad (42)$$

and since $i \in B_{m+1}$, according to the plan P_b,

$$P_2^* = P_{(3)2} \leqslant P_{(3)i}+P_{(2)i}-P_{(1)i}-(1-\theta_1-P_{(2)1}+P_{(1)1}). \quad \ldots \quad (43)$$

Adding (42) and (43) and after some simplifications we get

$$P_1^*+P_2^* = P_{(3)1}+P_{(3)2} \leqslant P_{(3)l_1}+P_{(3)l_2}+\sum_{j=1}^{2}(P_{(2)l_j}-P_{(1)l_j})-(1-\theta_1)$$

$$\leqslant P_{l_1}+P_{l_2} \text{ (from Lemma 6).}$$

We next prove that

$$P_1^*+P_2^*+P_3^* \leqslant P_{l_1}+P_{l_2}+P_{l_3} \quad \ldots \quad (44)$$

which holds obviously when $P_3^* \leqslant P_{l_3}$. So let

$$P_3^* > P_{l_3}. \quad \ldots \quad (45)$$

Note that under the assumption (45), the set $\{l_1, l_2, l_3\} \not\subset \{3, 4, \ldots, m\}$.

Case I (*a*) : Without loss of generality assume that

$$l_1 = 1,\ l_2 = 2 \text{ and } l_3 \in B_{m+1}.$$

According to the plan P_b,

$$P_3^* = P_{b_3} \leqslant P_{(3)l_3}+P_{(2)l_3}-P_{(1)l_3}-(1-\theta_1-\sum_{j=1}^{2}(P_{(2)j}-P_{(1)j})).$$

Adding $P_{(3)1}+P_{(3)2}$ to both sides of the above inequality, we get

$$P_3^*+P_{(3)1}+P_{(3)2} \leqslant \sum_{j=1}^{3}(P_{(3)l_j}+P_{(2)l_j}-P_{(1)l_j})-(1-\theta_1)$$

which implies that

$$P_1^*+P_2^*+P_3^* \leqslant P_{l_1}+P_{l_2}+P_{l_3} \qquad \text{(from Lemma 6).}$$

Case II (a): Let $l_1 \in A_2$ and $\{l_2, l_3\} \subset B_{m+1}$. Further, without loss of generality let $l_2 = i < j = l_3$. Since $l_1 \in A_2$, we have

$$P_1^* \leqslant P_{l_1}.$$

Using a similar argument as given for the case II, it can be easily shown that

$$P_2^* \leqslant P_{(3)j}+P_{(2)j}-P_{(1)j}-(\theta_2-1-\gamma_{j+1})$$

and for $i \geqslant m+1$

$$P_2^*+P_3^* \leqslant P_{l_2}+P_{l_3}$$

and hence

$$P_1^*+P_2^*+P_3^* \leqslant P_{l_1}+P_{l_2}+P_{l_3}.$$

Case III: Let $\{l_1, l_2, l_3\} \subset B_{m+1}$ and without loss of generality assume that $l_1 = i$, $l_2 = j$, $l_3 = k$ and $i < j < k$. Further, let

$$P_{(2)1}-P_{(1)1} \geqslant \theta_2-1-\gamma_{k+1}.$$

According to the plan $\boldsymbol{P_t}$.

$$P_{(3)1}+\theta_2-1-\gamma_{k+1} \leqslant P_{(3)k}+P_{(2)k}-P_{(1)k} \qquad \ldots \quad (46)$$

which implies

$$P_1^* \leqslant P_{(3)k}+P_{(2)k}-P_{(1)k}-(\theta_2-1-\gamma_{k+1}). \qquad \ldots \quad (47)$$

Since $j \in B_{m+1}$,

$$P_2^* \leqslant P_j^* = P_{(3)j}+P_{(2)j}-P_{(1)j}. \qquad \ldots \quad (48)$$

If $i \geqslant m+2$, then

$$P_3^* \leqslant P_i^* = P_{(3)i}+P_{(2)i}-P_{(1)i} \qquad \ldots \quad (49)$$

and

$$\theta_2-1-\gamma_{k+1} \geqslant P_{(2)i}-P_{(1)i}+P_{(2)j}-P_{(1)j}+P_{(2)k}-P_{(1)k}. \qquad \ldots \quad (50)$$

Adding (47), (48) and (49), and using the relation (50), we get

$$P_1^*+P_2^*+P_3^* \leqslant P_{(3)i}+P_{(3)j}+P_{(3)k}$$
$$= P_{(3)l_1}+P_{(3)l_2}+P_{(3)l_3} \qquad \ldots \quad (51)$$

If $i = m+1$, then

$$P_3^* \leqslant P_i^* = P_{(3)i}+(\theta_2-1-\gamma_{i+1}) \quad \cdots \quad (52)$$

and

$$\theta_2-1-\gamma_{k+1} \geqslant P_{(2)j}-P_{(1)j}+P_{(2)k}-P_{(1)k}+(\theta_2-1-\gamma_{i+1}). \quad \cdots \quad (53)$$

Adding (47), (48) and (52), and using (53), we can establish (51). We now suppose that

$$P_{(2)1}-P_{(1)1} < \theta_2-1-\gamma_{k+1}$$

and

$$P_{(2)2}-P_{(1)2} \geqslant \theta_2-1-\gamma_{j+1}.$$

As in the above case one can easily show that

$$P_2^* \leqslant P_{(3)j}+P_{(2)j}-P_{(1)j}-(\theta_2-1-\gamma_{j+1})$$

and for $i \geqslant m+1$.

$$P_2^*+P_3^* \leqslant P_{(3)j}+P_{(3)i} \leqslant P_{l_1}+P_{l_2}.$$

Since $P_1^* \leqslant P_{l_3}$ (always holds), we have

$$P_1^*+P_2^*+P_3^* \leqslant P_{l_1}+P_{l_2}+P_{l_3}.$$

Finally, we consider the case

$$P_{(2)1}-P_{(1)1} < \theta_2-1-\gamma_{k+1}$$

and

$$P_{(2)2}-P_{(1)2} < \theta_2-1-\gamma_{j+1}.$$

Then, according to the plan $\mathbf{P}_t$, we have

$$P_{(3)1}+P_{(2)1}-P_{(1)1} \leqslant P_{(3)l_3}+P_{(2)l_3}-P_{(1)l_3} \quad \cdots \quad (54)$$

and

$$P_{(3)2}+P_{(2)2}-P_{(1)2} \leqslant P_{(3)l_2}+P_{(2)l_2}-P_{(1)l_2} \quad \cdots \quad (55)$$

Since $P_3^* = P_{b_3}$, the definition of P_{b_3} implies that

$$P_3^* \leqslant P_{(3)l_1}+P_{(2)l_1}-P_{(1)l_1}-\Big(1-\theta_1-\sum_{j=1}^{2}(P_{(2)j}-P_{(1)j})\Big). \quad \cdots \quad (56)$$

Adding (54), (55) and (56), we get

$$P_1^*+P_2^*+P_3^* \leqslant \sum_{j=1}^{3}(P_{(3)l_j}+P_{(2)l_j}-P_{(1)l_j})-(1-\theta_1)$$

$$\leqslant P_{l_1}+P_{l_2}+P_{l_3} \qquad \text{(from Lemma 6)}$$

We now are in a position to prove the following theorem.

Theorem 5 : *The plan* $\mathbf{P}^*$ *is optimal in the context of a general sample size* n.

Proof : Let $\mathcal{P}'$ denote the set of extreme points in the class $\mathcal{P}_v$. Since $\mathcal{P}$ is complete and convex, to prove that P^* is an optimal plan it is enough to show that it is optimal in the set $\mathcal{P}'$.

Let $P = (P_1, P_{21}, ..., P_N) \in \mathcal{P}'$. The inequalities (7) and (8) are equivalent to

$$\sum_{j=i}^{N} P^*_{(j)} \geqslant \sum_{j=i}^{N} P_{(j)}, \quad i = 2, 3, ..., N.$$

which together with

$$\sum_{j=1}^{N} P^*_j = \sum_{j=1}^{N} P^*_j = \sum_{j=1}^{N} P_j$$

imply that P^* majorizes any $P \in \mathcal{P}'$. Since $(1-x)^n$ is a convex function of $x(0 \leqslant x \leqslant 1)$, from Theorem 1, we have

$$\sum_{j=1}^{N} (1-P^*_j)^N \geqslant \sum_{j=1}^{N} (1-P_j) \Rightarrow$$

$$\sum_{j=1}^{N} (1-(1-P^*_j)^n) \leqslant \sum_{j=1}^{N} (1-(1-P_j)^n)$$

for all $P \in \mathcal{P}'$.

We conclude this paper with the following examples.

Example 1 shows that any arbitrary plan derived through Mitra-Pathak algorithm need not be better than the usual one if the surveys were carried out independently.

Example 1. Consider the following stochastic matrix for 3 surveys.

TABLE 1

$i \backslash j$	1	2	3	4	
	values of P_{ij}				
1	0.0	0.5	0.0	0.5	$\theta_1 = 0.0$
2	0.6	0.0	0.2	0.2	$\theta_2 = 1.2$
3	0.4	0.4	0.2	0.0	$\theta_3 = 1.8$

Mitra-Pathak algorithm gives the following plan :

$$P_{211} = 0.1, P_{411} = 0.3, P_{212} = 0.2, P_{232} = 0.2, P_{443} = 0.2$$

and

$$P_1 = 0.6, P_2 = 0.5, P_3 = 0.4, P_4 = 0.5 \quad ... \quad (57)$$

where P_{ijk} denotes $P(X_1 = i, X_2 = j, X_3 = k)$.

If the surveys were carried out independently, then for the integrated plan, obtained as the product of the marginal distributions of X_1, X_2 and X_3, it can be seen that

$$P_j = 1-(1-P_{1j})(1-P_{2j})(1-P_{3j}), \quad j = 1, 2, \ldots, N.$$

Therefore, for the present example,

$$P_1 = 0.76, \quad P_2 = 0.7, \quad P_3 = 0.36, \quad P_4 = 0.6. \qquad \ldots \quad (58)$$

For $n \geqslant 7$, the value Ev_n of the plan (57) is greater than that of the plan (58).

In Example 1 note that the plans P_t and P_b are identical. That is,

$$P_{t_4} = P_{b_4} = 0.8, \; P_{t_3} = P_{b_3} = 0.5, \; P_{t_2} = P_{b_2} = 0.5, \; P_{t_1} = P_{b_1} = 0.2$$

where $t_4 = 1$, $t_3 = 2$, $t_2 = 4$ and $t_1 = 3$.

We next give an example where the plans P_t and P_b are not identical.

Example 2 : Consider the stochastic matrix given in Table 2.

TABLE 2

	values of P_{ij}									
$j \backslash i$		2	3	4	5	6	7	8	9	
1	.10	0.0	.15	0.0	.20	.20	0.0	.05	.30	$\theta_1 = 0.0$
2	.10	.10	0.0	.20	.15	0.0	.15	.30	0.0	$\theta_2 = 1.15$
3	0.0	.12	.15	.15	0.0	.21	.27	0.0	.10	$\theta_3 = 1.85$

Plan : $\boldsymbol{P_b}$

$$P_{b_1} = .10, P_{b_2} = .12, P_{b_3} = .15, P_{b_4} = .20, P_{b_5} = .20,$$

$$P_{b_6} = .21, P_{b_7} = .27 \; P_{b_8} = .35, P_{b_9} = .40$$

where $b_j = j$, $j = 1, 2, \ldots, 9$.

Plan $\boldsymbol{P_t}$:

$$P_{t_1} = .10, P_{t_2} = .12, P_{t_3} = .15, P_{t_4} = .20, P_{t_5} = .20,$$

$$P_{t_6} = .21, P_{t_7} = .30 \; P_{t_8} = .30, P_{t_9} = .42$$

where $t_j = j$, for $j = 1, 2, \ldots, 6$, $t_7 = 8$, $t_8 = 9$ and $t_9 = 7$.

We compute the value Ev_n of the plans P_b and P_t for $n = 2, 3, \ldots, 10$ and present in the following table.

TABLE 3

	values of Ev_n	
n	Plan P_b	Plan P_t
2	3.4736	3.4726
3	4.5787	4.5773
4	5.4214	5.4201
5	6.0739	6.0732
6	6.5862	6.5863
7	6.9935	6.9942
8	7.3208	7.3221
9	7.5865	7.5881
10	7.8040	7.8057

The above table values show that Ev_n of the plan P_b is greater than that of the plan P_t for $2 \leqslant n \leqslant 5$ and less than that of the plan P_t for $n \geqslant 6$. Also note that the absolute difference between them is numerically insignificant for all $n \geqslant 2$.

The following example shows that the plans P_t and P_b are not identical but the value Ev_n of the plan P_b is smaller than that of the plan P_t for all $n \geqslant 2$.

Example 3 : Consider the following stochastic matrix for 3 surveys.

TABLE 4

$i \backslash j$	1	2	3	4	5	6	7	8	9	
1	.10	0.0	.15	0.0	.20	.20	0.0	.05	.30	$\theta_1 = 0.0$
2	.10	.10	0.0	.20	.15	0.0	.15	.30	0.0	$\theta_2 = 1.15$
3	0.0	.12	.15	.15	0.0	.22	.26	0.0	.10	$\theta_3 = 1.85$

(values of P_{ij})

$$\textit{Plan } P_b : \; P_{b_1} = .10, \; P_{b_2} = .12, \; P_{b_3} = .15, \; P_{b_4} = .20, \; P_{b_5} = .20,$$

$$P_{b_6} = .22, \; P_{b_7} = .26, \; P_{b_3} = .35, \; P_{b_9} = .40$$

where $b_j = j$ for $j = 1, 2, \ldots, 9$.

$$\textit{Plan } \mathbf{P}_t: \quad P_{t_1} = .10, P_{t_2} = .12, P_{t_3} = .15, P_{t_4} = .20, P_{t_5} = .20,$$

$$P_{t_6} = .22, P_{t_7} = .30, P_{t_8} = .30, P_{t_9} = .41$$

where $t_j = j$ for $j = 1, 2, \ldots, 6$, $t_7 = 8$, $t_8 = 9$ and $t_9 = 7$.

Numerical computation shows that

$$\sum_{j=1}^{9} (1-(1-P_j)^n) < \sum_{j=1}^{9} (1-(1-P_j)^n)$$

and the difference between them is numerically insignificant for all $n \geqslant 2$.

References

Arthanari, T. S. and Dodge (1981): *Mathematical Programming in Statistics*, Wiley, New York.

Causey, B. D., Cox, L. H. and Ernst, L. R. (1985): Applications of transportation theory to statistical problems. *J. Amer. Statist. Assoc.*, **80**, 903-909.

Krishnamoorthy, K. and Mitra, S. K. (1986): Cost robustness of an algorithm for optimal integration of surveys. *Sankhyā*, Ser. B., **48**, 233-245.

Keyfitz, N. (1951): Sampling with probability proportional to size: adjustment for changes in probabilities. *J. Amer. Statist. Assoc.*, **46**, 105-109.

Lahiri, D. B. (1954): Technical paper on some aspects of the development of the sample design. *Sankhyā*, **14**, 264-316.

Marshall, A. W. and Olkin, I. (1979): *Inequalities: Theory of Majorization and its Applications*, Academic press, New York.

Mitra, S. K. and Pathak, P. K. (1984): Algorithms for optimal integration two or three surveys. *Scand. J. Statist.*, **11**, 257-263.

Raj, D. (1957): On the method of overlapping maps in sample surveys. *Sankhyā* **17**, 89-98.

Paper received: August, 1986.
Revised: May, 1987.