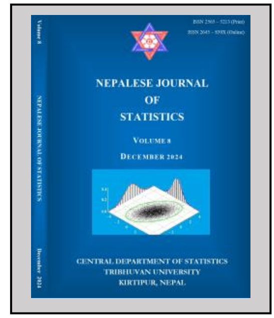

A New System of Inclusion Probability Proportional to Size Sampling Schemes of Two Units

Ajit Kumar Sabat[1], R. K. Sahoo[2] and L. N. Sahoo[3]*

Submitted: 26 August 2024; Accepted: 8 December 2024
Published online: 27 December 2024

DOI: https://doi.org/10.3126/njs.v8i1.73163

ABSTRACT

Background: One of the well-liked varying probability sampling methods is an inclusion probability proportional to size sampling scheme in which the first order inclusion probabilities are exactly proportional to size measures. Such schemes for sample size two have attracted attention of survey statisticians because of their simplicity in implementation, and nonnegative, unbiased and stable variance estimation of the Horvitz-Thompson estimator.

Objective: The purpose of the paper is to set forth a system or family of inclusion probability proportional to size sampling schemes of two units for estimating a finite population total.

Materials and Methods: Standard sampling techniques have been used to examine basic properties of the proposed family of sampling schemes as desired under Horvitz-Thompson framework. A numerical study, utilizing live data of 21 populations, has been conducted to evaluate relative performance of some member schemes of the family.

Results: The suggested family has been shown to satisfy almost all basic requirements of an inclusion probability proportional to size design and have flexible feature as it easily reduces to some of such existing designs. Three new designs, as special cases of the proposed family, have also been explored. Empirical results show that one of the new schemes comes out as the most efficient and the most stable among the comparable schemes.

Conclusion: Versatility property of the suggested system of sampling schemes along with fulfilment of fundamental needs of an inclusion probability proportional to size scheme will encourage for further research to detect other schemes of the system or outside the system with greater accuracy.

Keywords: Auxiliary character, HT estimator, inclusion probability, IPPS sampling scheme.

Address correspondence to the author: Asian School of Business Management University, Shiksha Vihar, Bhola, Chandaka, Bhubaneswar-754012, India,
Email: ajitsabat.1992@gmail.com[1]; Department of Statistics, Central University of Haryana, Jant Pali, Mahendragarh-123031, India, Email:ranjansahoostatistics@gmail.com[2];
Utkal University, Bhubaneswar–751004, India (Former Professor),
Email: lnsahoostatuu@rediffmail.com (Corresponding author email) [3*]

INTRODUCTION

Consider a surveyed population U of N $(< \infty)$ distinct and identifiable units with y_i as the measured value of the study variable y for the ith unit $(i = 1, 2, \ldots, N)$. Suppose that our aim is to estimate the unknown population total $Y = \sum_{i=1}^{N} y_i$ based upon a random sample s of fixed size n taken from U in accordance with an unequal probability sampling without replacement scheme with π_i and π_{ij} as the inclusion probability of the ith unit and the joint inclusion probability of the ith and jth units respectively. Under this configuration, Horvitz & Thompson (1952) introduced an unbiased estimator for Y that is frequently applied in survey analyses. This estimator, traditionally termed as HT estimator, is defined by

$$\hat{Y}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i}.$$

If size of the sample is fixed, then the variance of $\hat{Y}_{HT}$ is calculated by

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

(see, for example, Sarndal et al., 2003, p.43). As suggested by Sen (1952), and Yates & Grundy (1953) independently, $Var(\hat{Y}_{HT})$ is unbiasedly estimated by

$$v(\hat{Y}_{HT}) = \frac{1}{2} \sum_{i \neq j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2. \tag{2}$$

A sufficient condition for $v(\hat{Y}_{HT})$ to be non-negative is that $\pi_i \pi_j > \pi_{ij} > 0, \forall\, i \neq j$.

In many surveys, prior information is easily and cheaply available on an auxiliary character (variable) x assuming a known positive value x_i on the unit i and $X = \sum_{i=1}^{N} x_i$. Appreciable reduction in $Var(\hat{Y}_{HT})$ is then achievable by setting $\pi_i = np_i$, where $p_i = x_i / X$ is the initial probability of selection of the ith unit. This is of course only assured if the ratios y_i / x_i are roughly constant throughout the population. Such a scheme is popularly known as an inclusion probability proportional to size (IPPS or πps) sampling scheme or design. The estimator regularly used under the scheme is the HT estimator. Hence, according to the basic principles developed by Horvitz & Thompson (1952), an IPPS sampling design must satisfy $\sum_{i=1}^{N} \pi_i = n$, $\sum_{i \neq j}^{N} \pi_{ij} = (n-1)\pi_i$ and $\sum_i \sum_{j<i}^{N} \pi_{ij} = \frac{1}{2} n(n-1)$, recognized as its πps characteristics. Apart from these crucial properties, the scheme should produce an unbiased and nonnegative variance estimator of $\hat{Y}_{HT}$ as given in (2).

There had been substantial developments towards the formulation of various IPPS sampling schemes in the survey sampling literature (Durbin, 1953; Brewer, 1963; Sampford, 1967; Singh, 1978; Deshpande & Prabhu Ajgaonkar, 1982; Chao, 1982; Dey & Srivastava, 1987; Shahbas & Hanif, 2003; Senapati et al., 2006; Sahoo et al., 2006, 2007, 2011; Tiwari & Chilwal, 2013). Comprehensive review of different IPPS schemes along with their merits and demerits are also found in the textbooks by Brewer & Hanif (1983), Chaudhuri & Vos (1988), Mukhopadhyay (1996) and Arnab (2017). However, larger number of IPPS designs are confined to $n = 2$ only. The probable reasons are that calculation of π_{ij} becomes laborious when $n > 2$ and most schemes appear to be less productive than even probability proportional to size with replacement (PPSWR) scheme. However, IPPS sampling schemes of $n = 2$ are beneficial for stratified sampling with smaller stratum size (Chaudhuri & Vos, 1988, p. 148). The present study focuses attention on the development of a family or a system of IPPS sampling schemes for $n = 2$ that possesses acceptable properties under the HT model and provides an unbiased and non-negative Sen-Yates-Grundy estimator of $Var(\hat{Y}_{HT})$.

MATERIALS AND METHODS

Description of the suggested sampling scheme

Let $\phi(p_i)$ be a known function of p_i such that $\phi(p_i) > 0$ for all i. Corresponding to $N-$population units, consider a set of revised probabilities $\{P_1, P_2, \dots, P_N\}$ such that P_i is given by

$$P_i = \frac{(1-z_i)(2p_i - \alpha z_i)}{(1-2z_i)}, \; i = 1, 2, \dots, N, \tag{3}$$

where $z_i = \phi(p_i)/\sum_{i=1}^{N}\phi(p_i)$ and α is a known constant. In the usual practice α is determined from $\sum_{i=1}^{N} P_i = 1$ as

$$\alpha = \sum_{i=1}^{N}\frac{p_i}{1-2z_i} \Big/ \sum_{i=1}^{N}\frac{z_i(1-z_i)}{1-2z_i}. \tag{4}$$

Here, we remark that the $i-$th revised probability P_i is feasible only for those circumstances where $z_i < 1/2$ and $z_i < 2p_i/\alpha$ i.e., $z_i < \min(1/2, 2p_i/\alpha)\ \forall\ i$. But, in the actual practice, severity of these restrictions on z_i is supported by the characteristics of the expounded function ϕ.

We define the suggested sampling scheme for $n = 2$ in the following manner and recognize this new scheme as S_ϕ:

- The first unit in the sample, say i, is drawn with revised probability P_i and without replacement.
- The second unit in the sample, say j, is drawn with conditional probability $P_{j/i} = \frac{z_j}{1-z_i}$ from the rest $N-1$ population units.

Although z_i regulates the revised probabilities of selections of the units, it is heavily dependent on the selection of the non-negative function $\phi(p_i)$. From the ensuing examination of the properties of the generalized scheme S_ϕ undertaken in the next sub-section, it is also clear that for any choice of $\phi(p_i)$ the scheme meets the IPPS conditions. Hence, motivated by this it may be concluded that for different choices of $\phi(p_i)$, S_ϕ defines a family of IPPS sampling schemes. The expressions for the first and second order inclusion probabilities of the scheme are derived as follows.

We have

$$\pi_i = P_i + \sum_{j\neq i} P_j P_{i/j}$$
$$= \frac{(1-z_i)(2p_i-\alpha z_i)}{(1-2z_i)} + \sum_{j\neq i} \frac{(1-z_j)(2p_j-\alpha z_j)}{(1-2z_j)} \cdot \frac{z_i}{1-z_j}$$
$$= \frac{(1-z_i)(2p_i-\alpha z_i)}{(1-2z_i)} - \frac{z_i(2p_i-\alpha z_i)}{(1-2z_i)} + z_i \sum_{j=1}^{N} \left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)$$
$$= \frac{(1-2z_i)(2p_i-\alpha z_i)}{(1-2z_i)} + z_i \sum_{j=1}^{N} \left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)$$
$$= 2p_i - z_i \left[\alpha - 2\sum_{j=1}^{N} \frac{p_j}{1-2z_j} + \alpha \sum_{j=1}^{N} \frac{z_j}{1-2z_j}\right]. \tag{5}$$

From (4), we find

$$\sum_{i=1}^{N} \frac{p_i}{1-2z_i} - \alpha \sum_{i=1}^{N} \frac{z_i(1-z_i)}{1-2z_i} = 0$$
$$\Rightarrow \quad \sum_{i=1}^{N} \frac{p_i}{1-2z_i} - \frac{\alpha}{2} \sum_{i=1}^{N} \frac{z_i(1+1-2z_i)}{1-2z_i} = 0$$
$$\Rightarrow \quad \alpha - 2\sum_{i=1}^{N} \frac{p_i}{1-2z_i} + \alpha \sum_{i=1}^{N} \frac{z_i}{1-2z_i} = 0. \tag{6}$$

Hence, using (6) from (5) we have

$$\pi_i = 2p_i. \tag{7}$$

Further, by definition

$$\pi_{ij} = P_i P_{j/i} + P_j P_{i/j}.$$

Some algebra shows that

$$\pi_{ij} = \frac{(2p_i-\alpha z_i)z_j}{(1-2z_i)} + \frac{(2p_j-\alpha z_j)z_i}{(1-2z_j)}. \tag{8}$$

If we entertain the scheme with $p_i = \frac{1}{N}\ \forall\, i$, then we see that $z_i = \frac{1}{N}$, $\alpha = \frac{1}{N}$ and $P_i = \frac{1}{N}$. Finally, as is expected, we derive $\pi_i = \frac{1}{N}$ and $\pi_{ij} = \frac{2}{N(N-1)}$, which are the inclusion probabilities for $n = 2$ under simple random sampling without replacement (SRSWOR).

IPPS properties of S_ϕ

Let us now have an examination on the fascinating IPPS or πps properties of the suggested generalized sampling scheme S_ϕ.

(i) $\sum_{i=1}^{N} \pi_i = 2\sum_{i=1}^{N} p_i = 2$

(ii) $$\sum_{j\neq i} \pi_{ij} = \frac{(2p_i-\alpha z_i)}{(1-2z_i)} \sum_{j\neq i} z_j + z_i \sum_{j\neq i} \left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)$$
$$= \frac{(2p_i-\alpha z_i)}{(1-2z_i)} \left(\sum_{j=1}^{N} z_j - z_i\right) - \frac{(2p_i-\alpha z_i)z_i}{(1-2z_i)} + z_i \sum_{j=1}^{N} \left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)$$
$$= 2p_i - \alpha z_i + z_i \sum_{j=1}^{N} \left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)$$
$$= 2p_i - z_i \left[\alpha - 2\sum_{j=1}^{N} \frac{p_j}{1-2z_j} + \alpha \sum_{j=1}^{N} \frac{z_j}{1-2z_j}\right]. \tag{9}$$

Hence, utilizing expression (6) in (9), we have $\sum_{j\neq i} \pi_{ij} = 2p_i = \pi_i$.

 ISSN: 2565-5213 (Print); 2645-839X (Online)

(iii) $\sum_{i=1}^{N}\sum_{j<i}\pi_{ij} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij} = \frac{1}{2}\sum_{i=1}^{N}2p_i = 1.$

(iv) It remains to establish that $\pi_i\pi_j - \pi_{ij} > 0, i \neq j = 1, 2, \ldots, N$. But, for simplicity of notations, first we shall show that $\pi_1\pi_2 - \pi_{12} > 0$. Following Konijn (1973, p.253), we obtain

$$\begin{aligned}\pi_1\pi_2 - \pi_{12} &= \left(\pi_{12} + \textstyle\sum_{j>2}\pi_{1j}\right)\left(\pi_{12} + \sum_{j>2}\pi_{2j}\right) - \pi_{12}\\ &= \pi_{12}\left(1 - \textstyle\sum_{i>2}\sum_{j>i}\pi_{ij}\right) + \sum_{j>2}\pi_{1j}\sum_{j>2}\pi_{2j} - \pi_{12}\\ &= \textstyle\sum_{j>2}\pi_{1j}\sum_{j>2}\pi_{2j} - \pi_{12}\sum_{i>2}\sum_{j>i}\pi_{ij} \qquad (10)\end{aligned}$$

Further we have that

$$\begin{aligned}\textstyle\sum_{j>2}\pi_{1j}\sum_{j>2}\pi_{2j} &= \left[\sum_{j>2}\left\{\frac{(2p_1-\alpha z_1)z_j}{(1-2z_1)} + \frac{(2p_j-\alpha z_j)z_1}{(1-2z_j)}\right\}\right]\times\\ &\quad\left[\sum_{j>2}\left\{\frac{(2p_2-\alpha z_2)z_j}{(1-2z_2)} + \frac{(2p_j-\alpha z_j)z_2}{(1-2z_j)}\right\}\right]\\ &= \left[\frac{(2p_1-\alpha z_1)}{(1-2z_1)}\sum_{j>2}z_j + z_1\sum_{j>2}\left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)\right]\times\\ &\quad\left[\frac{(2p_2-\alpha z_2)}{(1-2z_2)}\sum_{j>2}z_j + z_2\sum_{j>2}\left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)\right]\\ &= \frac{(2p_1-\alpha z_1)(2p_2-\alpha z_2)}{(1-2z_1)(1-2z_2)}\left(\sum_{j>2}z_j\right)^2 + z_1z_2\left[\sum_{j>2}\left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)\right]^2\\ &\quad+\pi_{12}\sum_{j>2}z_j\sum_{j>2}\left(\frac{2p_j-\alpha z_j}{1-2z_j}\right), \qquad (11)\end{aligned}$$

and

$$\begin{aligned}\pi_{12}\textstyle\sum_{i>2}\sum_{j>i}\pi_{ij} &= \pi_{12}\sum_{i>2}\sum_{j>i}\left[\frac{(2p_i-\alpha z_i)z_j}{1-2z_i} + \frac{(2p_j-\alpha z_j)z_i}{1-2z_j}\right]\\ &= \pi_{12}\left[\sum_{j>2}\left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)\sum_{k>2}z_k - \sum_{j>2}\frac{(2p_j-\alpha z_j)z_j}{1-2z_j}\right]\\ &= \pi_{12}\sum_{j>2}\left[\left(\sum_{k>2}z_k - z_j\right)\left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)\right] \qquad (12)\end{aligned}$$

Using (11) and (12) in (10), we obtain that

$$\begin{aligned}\pi_1\pi_2 - \pi_{12} &= \frac{(2p_1-\alpha z_1)(2p_2-\alpha z_2)}{(1-2z_1)(1-2z_2)}\left(\textstyle\sum_{j>2}z_j\right)^2 + z_1z_2\left[\sum_{j>2}\left(\frac{2p_j-\alpha z_j}{1-2z_j}\right)\right]^2\\ &\quad+\pi_{12}\textstyle\sum_{j>2}\frac{(2p_j-\alpha z_j)z_j}{1-2z_j} > 0\,.\end{aligned}$$

Similarly, for any arbitrary i and j, it can be shown that

$$\begin{aligned}\pi_i\pi_j - \pi_{ij} &= \frac{(2p_i-\alpha z_i)(2p_j-\alpha z_j)}{(1-2z_i)(1-2z_j)}\left(\textstyle\sum_{k>2}z_k\right)^2 + z_iz_j\left[\sum_{k>2}\left(\frac{2p_k-\alpha z_k}{1-2z_k}\right)\right]^2\\ &\quad+\pi_{ij}\textstyle\sum_{k>2}\frac{(2p_k-\alpha z_k)z_k}{1-2z_k} > 0\,. \qquad (13)\end{aligned}$$

The above derivations corroborate that the recommended sampling scheme retains its πps properties and provides an unbiased estimator of the Sen-Yates-Grundy variance of the HT estimator that is also non-negative no matter what selection is made for $\phi(p_i)$.

RESULTS

Some individual cases of S_ϕ

Even though S_ϕ produces an infinite number of schemes, identification of each individual case is impossible. However, Table 1 displays some specific cases corresponding to some specific choices of $\phi(p_i)$ together with respective expressions for z_i, α, P_i and $P_{j/i}$ where $B = \sum_{i=1}^{N} \frac{p_i(1-p_i)}{(1-2p_i)}$, $\beta = \frac{1}{B}\sum_{i=1}^{N} \frac{p_i}{(1-2p_i)}$, $B_k = \sum_{i=1}^{N} \frac{z_{ki}(1-z_{ki})}{1-2z_{ki}}$ and $\alpha_k = \frac{1}{B_k}\sum_{i=1}^{N} \frac{p_i}{1-2z_{ki}}$, $k = 1, 2, 3, 4, 5, 6$. From the tabulated results, it is observed that the IPPS schemes of Midzuno (1952) and Brewer (1963), and those of Sahoo et al. (2006, 2007, 2011) are distinct members of S_ϕ. Nevertheless, the domain of S_ϕ is confined only to the said schemes. Some other such schemes also appear as particular cases of the family for other selections of $\phi(p_i)$. For instance, corresponding to three simple choices of $\phi(p_i)$ we identify three new schemes denoted by S_4, S_5 and S_6 as shown in Table 1.

Efficiency of S_ϕ

It would of course be interesting to explore the effectiveness of the introduced scheme in respect of some suitable performance measures. But our precedent discussions imply that the variance of $\hat{Y}_{HT}$ under the scheme relates to the selection of $\phi(p_i)$. This makes efficiency evaluation of S_ϕ compared to another scheme unfeasible unless a specific case of $\phi(p_i)$ is taken into consideration. Keeping this more appreciative point in mind and in view of the difficulties encountered in attempting to evaluate various results theoretically, a numerical study is undertaken to explore efficiencies of the eight sampling schemes $S_M, S_B, S_1, S_2, S_3, S_4, S_5$ and S_6 as particular cases of S_ϕ defined in Table 1. To make the numerical comparison more viable, two additional IPPS sampling plans due to Singh (1978) and Deshpande & Prabhu Ajgaonkar (1982), denoted by S_S and S_{DP} respectively, are also included in the study. IPPS schemes of Rao (1965), Durbin (1967) and Samford (1967) were not taken into consideration on the ground that their π_i and π_{ij} values for $n = 2$ are identical to those of the Brewer's scheme.

The goal of the present numerical study is twofold *i.e.*, to obtain an idea of what to expect in the relative performances of the member schemes of S_ϕ versus other IPPS designs outside the system (called as the non-member schemes), and to analyse to what extent the member schemes differ from each other on the ground of their relative performances. Here, we deal with two different performance measures:

(i) Efficiency when compared with the variance of the conventional estimator $\hat{Y}_{PPS} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{p_i}$ under the PPSWR scheme. We consider theoretical variance formula of the HT estimator given in (1) based on a scheme as a measure of its efficiency.

(ii) Stability of the variance estimator. Here we consider Hanurav's (1967) benchmark defined by $\varphi = \min \frac{\pi_{ij}}{\pi_i \pi_j} > \eta$, for η sufficiently away from zero, as a measure of stability of the variance estimator. However, to use other stability measures we may refer to Rao & Bayless (1969), Stehman & Overton (1994) and Sarndal (1996).

Table 1. Selections of $\phi(p_i)$ and the resulting sampling scheme.

Selection of $\boldsymbol{\phi(p_i)}$	$\boldsymbol{z_i}$	$\boldsymbol{\alpha}$	$\boldsymbol{P_i}$	$\boldsymbol{P_{j/i}}$	Sampling scheme
$\boldsymbol{\frac{1}{N}}$	$\frac{1}{N}$	$\frac{N}{N-1}$	$\frac{2(N-1)p_i-1}{N-2}$	$\frac{1}{N-1}$	Midzuno (1952) (S_M, say)
$\boldsymbol{p_i}$	p_i	β	$\frac{1}{B}\cdot\frac{p_i(1-p_i)}{(1-2p_i)}$	$\frac{p_j}{1-p_i}$	Brewer (1963) (S_B, say)
$\boldsymbol{\sqrt{p_i}}$	$\frac{\sqrt{p_i}}{\sum_{i=1}^{N}\sqrt{p_i}}=z_{1i}$	α_1	$\frac{(1-z_{1i})(2p_i-\alpha_1 z_{1i})}{(1-2z_{1i})}$	$\frac{z_{1j}}{1-z_{1i}}$	Sahoo et al. (2006)(S_1, say)
$\boldsymbol{p_i(1-p_i)}$	$\frac{p_i(1-p_i)}{\sum_{i=1}^{N}p_i(1-p_i)}=z_{2i}$	α_2	$\frac{(1-z_{2i})(2p_i-\alpha_2 z_{2i})}{(1-2z_{2i})}$	$\frac{z_{2j}}{1-z_{2i}}$	Sahoo et al. (2007)(S_2, say)
$\boldsymbol{p_i^2}$	$\frac{p_i^2}{\sum_{i=1}^{N}p_i^2}=z_{3i}$	α_3	$\frac{(1-z_{3i})(2p_i-\alpha_3 z_{3i})}{(1-2z_{3i})}$	$\frac{z_{3j}}{1-z_{3i}}$	Sahoo et al. (2011)(S_3, say)
$\boldsymbol{\frac{p_i}{1-p_i}}$	$\frac{\frac{p_i}{1-p_i}}{\sum_{i=1}^{N}\frac{p_i}{1-p_i}}=z_{4i}$	α_4	$\frac{(1-z_{4i})(2p_i-\alpha_4 z_{4i})}{(1-2z_{4i})}$	$\frac{z_{4j}}{1-z_{4i}}$	New sampling scheme(S_4, say)
$\boldsymbol{\frac{p_i}{2-p_i}}$	$\frac{\frac{p_i}{2-p_i}}{\sum_{i=1}^{N}\frac{p_i}{2-p_i}}=z_{5i}$	α_5	$\frac{(1-z_{5i})(2p_i-\alpha_5 z_{5i})}{(1-2z_{5i})}$	$\frac{z_{5j}}{1-z_{5i}}$	New sampling scheme(S_5, say)
$\boldsymbol{\frac{p_i}{1-2p_i}}$	$\frac{\frac{p_i}{1-2p_i}}{\sum_{i=1}^{N}\frac{p_i}{1-2p_i}}=z_{6i}$	α_6	$\frac{(1-z_{6i})(2p_i-\alpha_6 z_{6i})}{(1-2z_{6i})}$	$\frac{z_{6j}}{1-z_{6i}}$	New sampling scheme(S_6, say)

Description of the numerical study

Table 2 summarizes 21 populations whose data are used for this numerical study. Numerical values of the percentage relative efficiency (PRE) of the HT estimator under the ten comparable IPPS methods *viz.*, S_M, S_B, S_1, S_2, S_3, S_4, S_5, S_6, S_S and S_{DP} compared to $\hat{Y}_{PPS}$, and variance estimator stability parameter φ of these schemes are exhibited in Tables 3 and 4 respectively. To calculate relative efficiency of a sampling scheme, the exact variance formula for the full population has been used. But after determining the value of $\frac{\pi_{ij}}{\pi_i\pi_j}$ for all $\binom{N}{n}$ possible samples of $n=2$ drawn from a population, the φ −value of a scheme has been decided. For gaining better knowledge on the potency of the comparable schemes, entries for the best performed cases are boldly printed and those for second best performed cases are underlined. Discussions on the numerical findings of the study are precisely presented in the next section.

Table 2. Populations under study.

Pop. no.	Source	N	y	x
1	Sarndal et al., 2003, p. 660	50 clusters	total population in 1985	total population in 1975
2	Sukhatme et al., 1984, p. 67	25 villages	area under rice	cultivated area
3	Sukhatme et al., 1984, p. 297	89 circles	area under wheat	no. of villages
4	Cochran, 1977, p. 152	49 cities	inhabitants in 1930	inhabitants in 1920
5	Mukhopadhyay, 1996, p. 207	36 households	household income	household size
6	Mukhopadhyay, 1996, p. 193	20 jute mills	quantity of raw materials	no. of labourers
7	Murthy, 1977, p. 398	43 factories	no. of absentees	no. of workers
8	Murthy, 1977, p. 399	34 villages	area under wheat in 1964	cultivated area in 1961
9	Murthy, 1977, p. 399	34 villages	area under wheat in 1964	area under wheat in 1963
10	Murthy, 1977, p. 228	80 factories	output	no. of workers
11	Murthy, 1977, p. 228	80 factories	output	fixed capital
12	Murthy, 1977, p. 422	24 villages	no. of cattle in survey	no. of cattle in census
13	Singh & Chaudhary, 1986, p. 155	17 villages	no. of milch animals survey	no. of milch animals census
14	Singh & Singh Mangat, 1996, p. 199	24 teachers	blood pressure	age
15	Singh & Singh Mangat, 1996, p. 192	30 villages	rental value of irrigated land for current year	assessed rental value 5 years back
16	Singh & Singh Mangat, 1996, p. 193	24 wards	no. of dwellings occupied by tenants	no. of dwellings
17	Singh & Singh Mangat, 1996, p. 193	27 buffalos	milk yield after introduction of the new feed	milk yield before introduction of the new feed
18	Asok & Sukhatme, 1976	35 villages	acreage under oats in 1957	recorded acreage of crops and grass for 1947
19	Horvitz & Thompson (1952)	20 blocks	no. of households in a block	eye estimated no. of households in a block
20	Raj & Chandhok (1998, p.291)	20 wards	actual no. of households	eye estimated no. of households
21	Konijn, 1973, p. 49	16 families	expenditure on food	total expenditure

DISCUSSION

Findings on the efficiency

From the quantitative tabular values (Table 3), although for many cases efficiency of the HT estimator varies trivially from one method to another, efficiency differences between the member and non-member schemes of S_{ϕ} are noticeable. This means that, on the efficiency ground, S_S and S_{DP} appear to be inferior to rest eight schemes. Among the member schemes, only three schemes *i.e.*, S_3, S_4 and S_6 come out reasonably well whereas precision of rest five member schemes is not so significant. In this comparison, the new scheme S_4 turns out as the most efficient as it occupies first, second and third positions in 11, 5 and 2 populations respectively. On the same consideration, the schemes S_6 and S_3 may be regarded as second best and third best performers.

Findings on the stability

Numerical values of the stability parameter φ for the comparable schemes are shown in Table 4. We see that in respect of stability of the variance estimators, the schemes (except S_3, S_4 and S_6) behave irregularly and any distinction between them is not very clear cut. However, being the best performed ones in 7 populations, the two schemes S_4 and S_3 perform equally well and jointly emerge out as the most stable schemes. Contrastingly, although the new scheme S_6 do not work so well, based on the computed results it appears to be the second-best stable scheme.

CONCLUSION

Mathematical results derived in this investigation, establish that the proposed system not only works well under the limitations of IPPS requirements but also brings about other IPPS schemes. Results of the empirical evaluation on the efficiency and stability show that the overall performances of the non-member schemes are normally inferior to the member schemes under both criteria. Overall, we conclude that our new scheme S_6 may be accepted as the most suitable one among the comparable schemes on the grounds of both efficiency and stability although on the later ground it behaves like S_3. However, this cannot be established as a general conclusion because the comparative study utilizes data of 21 populations only and the efficiency gains between the comparable schemes are very small.

Table 3. PRE of the different sampling schemes.

Pop. no.	Sampling schemes									
	S_M	S_B	S_1	S_2	S_3	S_4	S_5	S_6	S_S	S_{DP}
1	103.519	<u>103.879</u>	103.499	102.907	102.948	**104.869**	102.872	102.879	102.023	102.179
2	104.866	105.701	105.538	105.696	105.798	**105.981**	105.702	<u>105.801</u>	105.734	105.701
3	**103.770**	100.949	102.595	100.950	102.571	<u>102.855</u>	102.748	100.906	100.923	100.949
4	100.061	100.372	100.680	100.374	100.529	<u>100.718</u>	100.472	**100.888**	100.359	100.372
5	101.302	102.087	101.460	102.291	**103.812**	102.005	102.286	<u>103.667</u>	101.243	101.287
6	106.144	106.137	<u>106.700</u>	106.142	106.493	**106.835**	106.046	106.037	106.004	106.037
7	<u>102.717</u>	102.612	102.471	102.610	102.549	**102.933**	102.612	102.698	102.331	102.412
8	103.487	103.562	103.393	103.267	103.471	<u>105.471</u>	103.501	**106.962**	103.213	103.162
9	102.011	104.018	103.428	<u>104.555</u>	102.954	**104.721**	104.020	104.349	101.081	101.018
10	108.201	108.135	108.232	108.342	<u>108.701</u>	**109.865**	108.466	108.289	108.101	108.143
11	112.047	112.985	112.304	111.511	**113.650**	113.281	109.474	<u>113.345</u>	100.972	101.781
12	104.672	105.180	105.151	104.224	104.979	<u>106.178</u>	104.980	**107.188**	105.127	105.187
13	106.724	<u>107.716</u>	106.718	106.716	106.722	107.515	106.716	**107.962**	106.315	106.216
14	102.832	103.989	103.994	104.012	<u>107.956</u>	**108.381**	104.184	104.189	104.210	104.189
15	103.760	103.647	103.656	103.618	**103.933**	103.617	<u>103.817</u>	103.795	103.612	103.617
16	104.507	104.767	104.761	104.854	<u>104.988</u>	**105.465**	104.867	104.857	104.690	104.567
17	106.231	103.850	104.711	103.863	<u>106.698</u>	103.846	103.847	**106.850**	103.733	103.650
18	104.333	102.876	103.998	104.701	104.776	**107.499**	104.459	<u>106.986</u>	101.432	101.654
19	104.656	104.543	104.778	103.765	104.564	**105.315**	104.456	<u>105.139</u>	103.991	104.567
20	105.989	106.222	105.234	106.453	106.849	**109.789**	106.786	<u>107.007</u>	105.123	104.435
21	108.224	107.345	107.765	106.342	**110.725**	<u>108.331</u>	108.006	108.115	106.332	106.411

Table 4. Stability parameter (φ) of the different sampling schemes.

Pop. no.	Sampling schemes									
	S_M	S_B	S_1	S_2	S_3	S_4	S_5	S_6	S_S	S_{DP}
1	0.4975	0.5013	0.5021	0.5022	**0.5173**	0.5164	0.5034	0.4987	0.5086	0.4978
2	0.5423	0.5354	0.5354	0.5365	**0.5463**	0.5371	0.5376	0.5387	0.5371	0.5372
3	0.5399	0.5386	0.5401	0.5377	0.5417	**0.5516**	0.5354	0.5457	0.5324	0.5347
4	0.5317	0.5315	0.5276	0.5323	0.5343	**0.5355**	0.5280	0.5256	0.5258	0.5277
5	**0.5293**	0.5169	0.5178	0.5206	0.5276	0.5232	0.5245	0.5241	0.5243	0.5265
6	0.5171	0.5132	0.5165	0.5036	**0.5280**	0.5205	0.5229	0.5176	0.5121	0.5121
7	0.5454	0.5434	0.5425	0.5441	0.5432	0.5470	0.5433	**0.5486**	0.5412	0.5400
8	0.5165	0.5229	0.5232	0.5316	0.5235	**0.5343**	0.5254	0.5170	0.5156	0.5212
9	0.4876	0.4996	0.4898	0.4906	0.4934	**0.5099**	0.4856	0.4917	0.4917	0.4923
10	0.5322	0.5323	0.5331	0.5324	**0.5374**	0.5315	0.5311	0.5312	0.5346	0.5314
11	0.5216	0.5243	0.5223	0.5254	0.5320	0.5262	**0.5339**	0.5246	0.5243	0.5278
12	0.5109	0.5175	0.5123	0.5104	0.5104	**0.5237**	0.5112	0.5112	0.5113	0.5113
13	0.5456	**0.5567**	0.5486	0.5487	0.5512	0.5553	0.5449	0.5500	0.5492	0.5478
14	0.5059	0.5087	0.5125	0.5111	**0.5172**	0.5073	0.5100	0.5133	0.5081	0.5060
15	0.5051	0.5062	0.5067	0.5071	0.5142	**0.5163**	0.5058	0.5089	0.5064	0.5050
16	0.5365	0.5321	**0.5432**	0.5298	0.5302	0.5348	0.5316	0.5289	0.5267	0.5256
17	0.5112	0.5112	0.5167	**0.5234**	0.5095	0.5085	0.5080	0.5109	0.5125	0.5075
18	0.5143	0.5165	0.5141	0.5140	**0.5325**	0.5245	0.5155	0.5151	0.5144	0.5143
19	0.5184	0.5264	0.5271	0.5320	**0.5362**	0.5182	0.5285	0.5134	0.5155	0.5087
20	0.5564	0.5576	0.5553	0.5546	0.5541	0.5601	0.5521	**0.5629**	0.5551	0.5543
21	0.5046	0.5065	0.5074	0.5071	0.5071	**0.5137**	0.5112	0.5065	0.5034	0.5090

ACKNOWLEDGEMENTS

The authors are grateful to the learned reviewers whose constructive comments led to an improvement on the earlier version of the paper. The authors are also thankful to the Editor-in-Chief for his general guidance given for the preparation of this paper.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTION

This paper is an unpublished part of the Ph.D. Thesis entitled "Some Classical Estimation Procedures on the Effective Use of Auxiliary Information in Survey Sampling" submitted by Ajit Kumar Sabat (First & principal author) to G.M. University, Sambalpur, Odisha, India in December 2022. Mr Sabat carried out his research work under the supervision of Dr R. K. Sahoo (Second author) and overall guidance of Prof L. N. Sahoo (Third & corresponding author).

FUNDING

No funding has been received.

DATA AVAILABILITY

Data available in various textbooks and published research papers (as referenced) are utilized in the empirical study of the paper.

ETHICAL STATEMENT

Various results embodied in the paper are original contributions of the authors. Authors also acknowledge that the paper has not been published or submitted for publication in any other journal.

REFERENCES

Arnab, R. (2017). *Survey sampling theory and applications*, Academic Press, Elsevier.

Asok, C., & Sukhatme, B. V. (1976). On Sampford's procedure of unequal probability sampling without replacement. *Journal of the American Statistical Association*, *71*, 912-918. https://doi.org/10.1080/01621459.1976.10480968

Brewer, K. R. W. (1963). Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics*, *5*, 93-105.

Brewer, K. R. W., & Hanif, M. (1983). *Sampling with unequal probabilities*, Lecture Notes in Statistics, Berlin: Springer-Verlag.

Chao, M. (1982). A general-purpose unequal probability sampling plan. *Biometrika*, *69*, 653-656. https://doi.org/10.1093/biomet/69.3.653

Chaudhuri, A., & Vos, J. W. E. (1988). *Unified theory and strategies of survey sampling*, Amsterdam; North Holland.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.), New York: John Wiley & Sons.

Deshpande, M. N., & Prabhu-Ajgaonkar, S. G. (1982). An IPPS (inclusion probability proportional to size) sampling scheme. *Statistica Neerlandica*, *36*, 209-212. https://doi.org/10.1111/j.1467-9574

Dey, A., & Srivastava, A. K. (1987). A sampling procedure with inclusion probabilities proportional to size. *Survey Methodology*, 13, 85-92.

Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. *Journal of the Royal Statistical Society*, *B15*, 262-269. https://doi.org/10.1111/j.2517-6161.1953.tb00141.x

Durbin, J. (1967). Design of multi-stage surveys for the estimation of sampling errors. *Applied Statistics*, *16*, 152-164. https://doi.org/10.2307/2985777

Hanurav, T. V. (1967). Optimum utilization of auxiliary information: Sampling of two units from stratum. *Journal of the Royal Statistical Society*, *B29*, 374-391.

Horvitz, D. G., & Thompson, D. J. (1952). A generalisation of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*, 663-685. https://doi.org/10.2307/2280784

Konijn, H. S. (1973). *Statistical theory of sample survey design and analysis*, Amsterdam: North-Holland.

Midzuno, H. (1952). On the sampling system with probability proportionate to sum of sizes. *Annals of the Institute of Statistical Mathematics*, *3*, 99-107. http://dx.doi.org/10.1007/BF02949779

Mukhopadhyay, P. (1996). *Theory and methods of survey sampling*, New Delhi: Prentice-Hall of India.

Murthy, M. N. (1977). *Sampling theory and methods*, Calcutta: Statistical Publishing Society.

Raj, D., & Chandhok, P. (1998). *Sample survey theory*, Narosa Publishing House.

Rao, J. N. K, & Bayless, D. L. (1969). An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum. *Journal of the American Statistical Association*, *64*, 540-549. https://doi.org/10.2307/2283637

Rao, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, *3*, 173-180.

Sahoo, L. N., Das, B. C., & Singh, G. N. (2006). A note on an IPPS sampling scheme. *Allgemeines Statistisches Archiv*, *90*, 385-393. https://doi.org/10.1007/s10182-006-0240-2

Sahoo, L. N., Mishra, G.. & Nayak, S.R. (2007). On a πps scheme of sampling of two units. *Brazilian Journal of Probability and Statistics*, *21*, 165-173. http://www.jstor.org/stable/43601096

Sahoo, L. N., Nayak, S. R., & Sahoo, R. K. (2011). A new IPPS sampling of two units. *Thailand Statistician*, *9*, 93-102. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34290

Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika, 54*, 499–513. https://doi.org/10.2307/2335041

Sarndal, C. E. (1996). Efficient estimators with simple variance in unequal probability sampling. *Journal of the American Statistical Association*, *91*, 1289-1300. https://doi.org/10.2307/2291747

Sarndal, C. E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling* (2nd ed.), Berlin: Springer-Verlag.

Sen, A. R. (1952). On the estimator of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, *5*, 119-127.

Senapati, S. C., Sahoo, L. N., & Mishra, G. (2006). On a scheme of sampling of two units with inclusion probability proportional to size. *Austrian Journal of Statistics*, *35*, 445-454. https://doi.org/10.17713/ajs.v35i4.354

Shahbas, M. Q., & Hanif, M. (2003). A simple procedure for unequal probability sampling without replacement sample size 2. *Pakistan Journal of Statistics*, *19*, 151-160.

Singh, D., & Chaudhary, F.S. (1986). *Theory and analysis of sample survey designs*, New York: John Wiley & Sons.

Singh, P. (1978). The selection of samples of two units with inclusion probability proportional to size. *Biometrika*, *65*, 450-454. https://doi.org/10.2307/2335231

Singh, R., & Singh Mangat, N. (1996). *Elements of survey sampling*, Amsterdam: Kluwer Academic Publishers.

Stehman, S., & Overton, V. (1994). Comparison of variance estimators of the Horvitz-Thompson estimator for randomized variable probability systematic sampling. *Journal of the American the Statistical Association*, *89*, 30-43. https://doi.org/10.1080/01621459.1994.10476443

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., & Asok, C., (1984). *Sampling theory of surveys with applications*, New Delhi: Indian Society of Agricultural Statistics.

Tiwari, N., & Chilwal, A. (2013). A simplified selection scheme for unequal probability sampling without replacement. *Thailand Statistician*, *11*, 133-141. http://statassoc.or.th

Yates, F., & Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society*, *B15*, 235-261.

Reference to this paper should be made as follows:

Sabat, A. K., Sahoo, R. K., & Sahoo, L. N. (2024). A new system of inclusion probability proportional to size sampling schemes of two units. *Nep. J. Stat, 8,* 33-46.