Thailand Statistician
January 2012; 10(1) : 69-86
http://statassoc.or.th
Contributed paper

Stratified Inverse Sampling for Rare Populations

Prayad Sangngam* [a] and Prachoom Suwattee [b]

[a] Department of Statistics, Faculty of Science, Silpakorn University, Nakhon Pathom, 73000, Thailand.
[b] Department of Statistics, School of Applied Statistics, National Institute of Development Administration, Bangkok, 10240, Thailand.
* Author for correspondence; e-mail: prayad@su.ac.th

Received: 22 April 2011
Accepted: 2 March 2012

Abstract

This paper considers stratified inverse sampling with four variations from each stratum, namely inverse random sampling with replacement, inverse random sampling without replacement, inverse probability proportional to size (PPS) sampling with replacement and inverse PPS sampling without replacement. Unbiased estimators of the mean of a study variable in the whole population and the number of units in a class of interest together with their unbiased variance estimators are given. Estimation of the mean per unit in the class of interest is also presented. A simulation study is employed to study the properties of these sampling designs and the results indicate that inverse sampling without replacement is more efficient than inverse sampling with replacement. Inverse PPS sampling gives higher efficiencies of the estimates than inverse random sampling when correlation coefficient between auxiliary and study variables is large. When the number of sampled units in a class of interest increases, the variance and mean squared error of the estimate decreases.

Keywords: inverse sampling, stratified sampling, unequal probability sampling.

1. Introduction

This paper introduces stratified inverse sampling designs which are appropriate when the population under study contains only a small fraction of units having a particular characteristic of interest. Such population is called a rare population. In examples of rare population surveys, it is desirable to estimate the total number of the animals, trees and plants with a special characteristic in a given forest. A survey may apply to sampling a human population at risk for HIV/AIDS. A sampling design might be concerned with a minority population, specific age/sex group such as males aged 18 to 24, the persons with rare diseases. Methods of sampling rare populations have been reviewed by Kalton and Anderson [1]. A wide variety of methods has been used for sampling rare populations including two-stage cluster, two-phase, network, snowball, multiple-frames and stratified sampling. The use of adaptive cluster sampling is described in Thompson and Seber [2]. One disadvantage of these sampling designs is that a sample may contain the units without a characteristic of interest. In order to obtain a given number of sampled units with characteristics of interest, Haldane [3] considered an inverse sampling with equal probabilities with replacement. An unbiased estimator of the prevalence of a characteristic was given. However, the formula for the variance of the estimator was complicated and an unbiased estimator of the variance was not given. Finney [4] gave an unbiased estimator of the variance of Haldane's estimator. Mikulski and Smith [5], Sathe [6] derived upper and lower bounds of the variance of Haldane's unbiased estimator. Salehi and Seber [7] found an unbiased estimator of the population mean and an unbiased estimator of its variance when an inverse sampling was without replacement. Christman and Lan [8] developed inverse adaptive cluster sampling. Three stopping rules were considered which depended on the number of units of interest. They found an unbiased estimator of the population total but an unbiased estimator of its variance was not considered. Salehi and Seber [9] developed general inverse adaptive cluster sampling. They presented an unbiased estimator of the population total under sampling without replacement and an unbiased estimator of the variance was also given. The sampling schemes aforementioned were considered under sampling with equal probabilities. Greco and Naddeo [10] considered inverse sampling with unequal probabilities with replacement. An unbiased estimator of the population total and an unbiased estimator of its variance were also given. Sangngam and Suwattee [11] considered inverse sampling with unequal probabilities and without replacement by modifying Midzuno's scheme [12].

Another method of reducing the variance of an estimator is to stratify a population into strata with varying prevalence of characteristic of interest. Usually, stratified random sampling with a disproportional allocation of the sample size is applied for rare population. In this paper, we consider variations of stratified inverse sampling for a rare population. The parameters to be estimated are the mean of a study variable in the whole population, the number of units with characteristic of interest and the mean of a study variable in a class of interest. A simulation study is employed to compare the four inverse sampling under consideration.

2. Notations

Let $U=\{u_1, u_2,\ldots,u_N\}$ denote a finite population of N distinct and identifiable units. Assume that the population consists of two subpopulations, C and $\overline{C}$ with cardinality M and $N-M$, respectively. It is assumed that units belonging to classes C or $\overline{C}$ are unknown before sampling. The population is stratified into L strata. A subpopulation consists of N_h units for $h=1,2,\ldots,L$ where $\sum_{h=1}^{L} N_h = N$. Let u_{hj} denote the j-th unit in the stratum h and y_{hj} a study value of the unit u_{hj}. A stratum is partitioned into C_h and $\overline{C}_h$ with cardinality M_h and $N_h - M_h$ respectively where $M_h > 2$ and $\sum_{h=1}^{L} M_h = M$. Let n_h be the sample size in stratum h and m_h be the number of sample units in the class C_h. For example, $C_h=\{u_{hj}: y_{hj} \ge b\}$ and $\overline{C}_h=\{u_{hj}: y_{hj} < b\}$, where b is a given constant. In stratum h, let z_{hj} be the selection probability of a unit j, $z_{Ch}=\sum_{j\in Ch} z_{hj}$ be the probability of selecting a unit in C_h, $z_{\overline{C}h}=\sum_{j\in \overline{C}h} z_{hj} = 1 - z_{Ch}$ the probability of selecting a unit in $\overline{C}_h$, $\sigma^2_{Ch}=\frac{1}{z_{Ch}}\sum_{j\in Ch} z_{hj}\left(\frac{y_{hj}}{z_{hj}}-\frac{Y_{Ch}}{z_{Ch}}\right)^2$ the variance of a study variable for units in class C_h, $\sigma^2_{\overline{C}h}=\frac{1}{z_{\overline{C}h}}\sum_{j\in \overline{C}h} z_{hj}\left(\frac{y_{hj}}{z_{hj}}-\frac{Y_{\overline{C}h}}{z_{\overline{C}h}}\right)^2$ the variance of study variable for units in class $\overline{C}_h$, $\overline{y}_{Ch}=\frac{1}{m_h}\sum_{j\in s_{Ch}} \frac{y_{hj}}{z_{hj}}$ the mean of study values per probability of selection for sampled units in a set s_{Ch}, $\bar{y}_{\bar{C}h} = \frac{1}{n_h - m_h} \sum_{j \in s_{\bar{C}h}} \frac{y_{hj}}{z_{hj}}$ the mean of study values per probability of selection for sampled units in a set $s_{\bar{C}h}$, $\hat{\sigma}^2_{Ch} = \frac{1}{m_h - 1} \sum_{j \in s_{Ch}} \left(\frac{y_{hj}}{z_{hj}} - \bar{y}_{Ch} \right)^2$ the variance of sampled units in a set s_{Ch} and $\hat{\sigma}^2_{\bar{C}h} = \frac{1}{n_h - m_h - 1} \sum_{j \in s_{\bar{C}h}} \left(\frac{y_{hj}}{z_{hj}} - \bar{y}_{\bar{C}h} \right)^2$ the variance of sampled units in a set $s_{\bar{C}h}$. From definition of $\hat{\sigma}^2_{Ch}$ and $\hat{\sigma}^2_{\bar{C}h}$, assume that $m_h > 1$ and $n_h - m_h > 1$. The number of units in the class of interest is given by $M = \sum_{h=1}^{L} M_h$. The population total of a study variable is denoted by $Y = \sum_{h=1}^{L} \sum_{j=1}^{N_h} y_{hj}$. The population mean of a study variable is $\bar{Y} = Y/N$. The total of a study variable in a class of interest is denoted by $Y_C = \sum_{h=1}^{L} \sum_{j \in Ch} y_{hj}$. The mean of a study variable in C is given by $\bar{Y}_C = Y_C/M$.

3. Stratified Inverse PPS Sampling

3.1 Stratified Inverse PPS Sampling with Replacement

In stratum h, a unit u_{hj} is selected with probability z_{hj} and with replacement until the sample contains m_h units (including replicates) from class C_h where m_h is fixed in advance. The sample consists of m_h units from C_h and $n_h - m_h$ units from $\bar{C}_h$. The $s_h = \left(i_1, i_2, \ldots, i_{n_h}\right)$ denotes the sample from stratum h. The sample s_h can be partitioned into two sets s_{Ch} and $s_{\bar{C}h}$ of units from C_h and $\bar{C}_h$, respectively.

Theorem 3.1 Under stratified inverse PPS sampling with replacement, an unbiased estimator of the population mean is

$$\bar{y}_{st} = \frac{1}{N} \sum_{h=1}^{L} \left[\hat{P}_h \, \bar{y}_{Ch} + (1 - \hat{P}_h) \bar{y}_{\bar{C}h} \right], \tag{3.1}$$

where $\hat{P}_h = \dfrac{m_h - 1}{n_h - 1}$. The variance of $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L}\left[\left(\frac{Y_{Ch}}{z_{Ch}} - \frac{Y_{\bar{C}h}}{z_{\bar{C}h}}\right)^2 V\left(\hat{P}_h\right) + \frac{\sigma_{Ch}^2}{m_h}E\left(\hat{P}_h^2\right) + \frac{\sigma_{\bar{C}h}^2}{m_h - 1}E\left[\hat{P}_h\left(1-\hat{P}_h\right)\right]\right], \quad (3.2)$$

where $Y_{Ch} = \sum_{j\in Ch} y_{hj}$ and $Y_{\bar{C}h} = \sum_{j\in \bar{C}h} y_{hj}$. For $m_h > 2$, an unbiased estimator of the variance is

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L}\left[\left(\bar{y}_{Ch} - \bar{y}_{\bar{C}h}\right)^2 \hat{V}\left(\hat{P}_h\right) + \frac{\hat{\sigma}_{Ch}^2}{m_h}\hat{P}_h^* + \frac{\hat{\sigma}_{\bar{C}h}^2}{m_h - 1}\left(\hat{P}_h - \frac{m_h - 1}{m_h - 2}\hat{P}_h^*\right)\right], \quad (3.3)$$

where $\hat{V}\left(\hat{P}_h\right) = \dfrac{\hat{P}_h\left(1-\hat{P}_h\right)}{n_h - 2}$ and $\hat{P}_h^* = \dfrac{(m_h - 1)(m_h - 2)}{(n_h - 1)(n_h - 2)}$.

Proof: Greco and Naddeo [10] showed that $\hat{Y}_h = \hat{P}_h\,\bar{y}_{Ch} + (1-\hat{P}_h)\bar{y}_{\bar{C}h}$ is an unbiased estimator of $Y_h = \sum_{j=1}^{N_n} y_{hj}$. They also showed that $V\left(\hat{Y}_h\right) = \left(\dfrac{Y_{Ch}}{z_{Ch}} - \dfrac{Y_{\bar{C}h}}{z_{\bar{C}h}}\right)^2 V\left(\hat{P}_h\right) + \dfrac{\sigma_{Ch}^2}{m_h}E\left(\hat{P}_h^2\right) + \dfrac{\sigma_{\bar{C}h}^2}{m_h - 1}E\left[\hat{P}_h\left(1-\hat{P}_h\right)\right]$ and an unbiased estimator of $V\left(\hat{Y}\right)$ is $\hat{V}\left(\hat{Y}_h\right) = \left(\bar{y}_{Ch} - \bar{y}_{\bar{C}h}\right)^2 \hat{V}\left(\hat{P}_h\right) + \dfrac{\hat{\sigma}_{Ch}^2}{m_h}\hat{P}_h^* + \dfrac{\hat{\sigma}_{\bar{C}h}^2}{m_h - 1}\left(\hat{P}_h - \dfrac{m_h - 1}{m_h - 2}\hat{P}_h^*\right)$. We get that $E(\bar{y}_{st}) = \dfrac{1}{N}\sum_{h=1}^{L}E\left(\hat{Y}_h\right) = \dfrac{1}{N}\sum_{h=1}^{L}\sum_{j=1}^{N_h} y_{hj} = \bar{Y}$, since the estimates are unbiased in the individual strata. Because of the samples are drawn independently in different strata, the covariance between $\hat{Y}_h$ and $\hat{Y}_k$ equals to 0 for $h \neq k = 1,2,\ldots,L$. This gives the results (3.2) and (3.3). Note that when the number m_h increases the variance of $\bar{y}_{st}$ decreases.

Corollary 3.1 An unbiased estimator of M is

$$\hat{M}_{st} = \sum_{h=1}^{L}\hat{P}_h\left(\frac{1}{m_h}\sum_{j\in s_{Ch}}\frac{1}{z_{hj}}\right), \quad (3.4)$$

with the variance

$$V\left(\hat{M}_{st}\right) = \sum_{h=1}^{L}\left[\frac{M_h^2}{z_{Ch}^2}V\left(\hat{P}_h\right) + \frac{\sigma_{Ch}'^2}{m_h}E\left(\hat{P}_h^2\right)\right], \quad (3.5)$$

where $\sigma'^2_{Ch} = \frac{1}{z_{Ch}} \sum_{j \in Ch} z_{hj} \left(\frac{1}{z_{hj}} - \frac{M_h}{z_{Ch}} \right)^2$ **.** For $m_h > 2$, an unbiased estimator of $V\left(\hat{M}_{st}\right)$ is

$$\hat{V}\left(\hat{M}_{st}\right) = \sum_{h=1}^{L} \left[\left(\frac{1}{m_h} \sum_{j \in Ch} \frac{1}{z_{hj}} \right)^2 \hat{V}\left(\hat{P}_h\right) + \frac{\hat{\sigma}'^2_{Ch}}{m_h} \hat{P}^*_h \right], \tag{3.6}$$

where $\hat{\sigma}'^2_{Ch} = \frac{1}{m_h - 1} \sum_{j \in s_{Ch}} \left(\frac{1}{z_{hj}} - \frac{1}{m_h} \sum_{j \in s_{Ch}} \frac{1}{z_{hj}} \right)^2$.

Proof: Define y'_{hj} to be 1 if the unit belongs to C_h and 0 if the unit belongs to $\bar{C}_h$. Then $M_h = \sum_{j=1}^{N_h} y'_{hj}$ and $M = \sum_{h=1}^{L} M_h$ is estimated unbiasedly by $\hat{M}_{st}$ in (3.4). For the study value y'_{hj}, taking on values 1 or 0, the variance, σ^2_{Ch} and sampled variance, $\hat{\sigma}^2_{Ch}$ are replaced by σ'^2_{Ch} and $\hat{\sigma}'^2_{Ch}$ respectively.

Corollary 3.2 An unbiased estimator of the total of a study variable in class C is $\hat{Y}_{Cst} = \sum_{h=1}^{L} \hat{P}_h \bar{y}_{Ch}$ with the variance

$$V\left(\hat{Y}_{Cst}\right) = \sum_{h=1}^{L} \left[\left(\frac{Y_{Ch}}{z_{Ch}} \right)^2 V\left(\hat{P}_h\right) + \frac{\sigma^2_{Ch}}{m_h} E\left(\hat{P}^2_h\right) \right]. \tag{3.7}$$

For $m_h > 2$, an unbiased estimator of the variance of $\hat{Y}_{Cst}$ is

$$\hat{V}\left(\hat{Y}_{Cst}\right) = \sum_{h=1}^{L} \left[\bar{y}^2_{Ch} \hat{V}\left(\hat{P}_h\right) + \frac{\hat{\sigma}^2_{Ch}}{m_h} \hat{P}^*_h \right]. \tag{3.8}$$

Proof: Define new variable y''_{hj} to be y_{hj} if the unit belongs to C_h and to be 0 if the unit belongs to class $\bar{C}_h$. Then $Y_{Ch} = \sum_{j=1}^{N_h} y''_{hj}$ and $\hat{Y}_{Cst}$ is an unbiased estimator of $Y_C = \sum_{h=1}^{L} Y_{Ch}$. For the study value y''_{hj}, taking on values y_{hj} or 0, we get that $Y_{\bar{C}h} = 0$, $\sigma^2_{\bar{C}h} = 0$, $\bar{y}_{\bar{C}h} = 0$ and $\hat{\sigma}^2_{\bar{C}h} = 0$.

Theorem 3.2 $\bar{y}_{Cst} = \dfrac{\hat{Y}_{Cst}}{\hat{M}_{st}}$ is a biased estimator of $\bar{Y}_C$ with bias $B(\bar{y}_{Cst}) = \dfrac{-Cov(\bar{y}_{Cst}, \hat{M}_{st})}{M}$ and $\dfrac{|B(\bar{y}_{Cst})|}{[V(\bar{y}_{Cst})]^{\frac{1}{2}}} \le CV(\hat{M}_{st})$. An approximate mean squared error of the estimator is $MSE(\bar{y}_{Cst}) \approx \dfrac{1}{M^2}\left[\bar{Y}_C^2 V(\hat{M}_{st}) + V(\hat{Y}_{Cst}) - 2\bar{Y}_C Cov(\hat{M}_{st}, \hat{Y}_{Cst})\right]$

where $Cov(\hat{M}_{st}, \hat{Y}_{Cst}) = \sum_{h=1}^{L}\left[\dfrac{M_h Y_{Ch}}{z_{Ch}^2} V(\hat{P}_h) + \dfrac{\sigma_{Ch}''^2}{m_h} E(\hat{P}_h^2)\right]$ and

$\sigma_{Ch}''^2 = \dfrac{1}{z_{Ch}} \sum_{j \in C} z_{hj}\left(\dfrac{1}{z_{hj}} - \dfrac{M_h}{z_{Ch}}\right)\left(\dfrac{y_{hj}}{z_{hj}} - \dfrac{Y_{Ch}}{z_{Ch}}\right)$.

Proof: $Cov(\bar{y}_{Cst}, \hat{M}_{st}) = E(\bar{y}_{Cst}\hat{M}_{st}) - E(\bar{y}_{Cst})E(\hat{M}_{st}) = E\left(\dfrac{\hat{Y}_C}{\hat{M}_{st}}\hat{M}_{st}\right) - M\,E(\bar{y}_{Cst})$. Hence,

$E(\bar{y}_{Cst}) = \bar{Y}_C - \dfrac{Cov(\bar{y}_{Cst}, \hat{M}_{st})}{M}$.

So $E(\bar{y}_{Cst}) - \bar{Y}_C = \dfrac{-Cov(\bar{y}_{Cst}, \hat{M}_{st})}{M}$ and $\dfrac{|B(\bar{y}_{Cst})|}{[V(\bar{y}_{Cst})]^{\frac{1}{2}}} = \left|\dfrac{Corr(\bar{y}_{Cst}, \hat{M}_{st})}{M}\right|\left[\dfrac{V(\bar{y}_{st})V(\hat{M}_{st})}{V(\bar{y}_{st})}\right]^{\frac{1}{2}}$

$\le \dfrac{[V(\hat{M}_{st})]^{\frac{1}{2}}}{M} = CV(\hat{M}_{st})$. The approximate mean squared error of $\bar{y}_{Cst}$ can derive by using linearization method.

3.2 Stratified Inverse PPS Sampling without Replacement

When the probability proportional to size without replacement at each draw is taken, we get a PPS sample without replacements. Since the selection probabilities change from draw to draw, to estimate the population total, we can use the Horvitz-Thompson estimator. Unfortunately, inclusion probabilities from inverse sampling depend on the unknown parameter M. It is not easy to use the Horvitz-Thomson estimator. Salehi and Seber [7] proved that the Murthy's estimator can be applied to inverse sampling. From stratum h, the first unit is drawn using the probability of selection z_{hj} and the remaining units are drawn one by one with equal probabilities and without replacement until the sample contains m_h units from class C_h. Assume that $m_h \le M_h$.

The sample is $s_h = \{i_1, i_2, \ldots, i_{n_h}\}$ where n_h is the sample size. Let s_{Ch} denote the set of units from class C_h and $s_{\bar{C}h}$ the set of units from $\bar{C}_h$ with cardinalities m_h and $n_h - m_h$, respectively, where $s_{Ch} \cap s_{\bar{C}h} = \phi$ and $s_{Ch} \cup s_{\bar{C}h} = s_h$.

Theorem 3.3 In stratified inverse PPS sampling without replacement, an unbiased estimator of $\bar{Y}$ is

$$\bar{y}_{st} = \frac{1}{N}\sum_{h=1}^{L}\left[\frac{(m_h - 1)}{m_h \sum_{j\in s_h} z_{hj} - \sum_{j\in s_{Ch}} z_{hj}} \sum_{j\in s_{Ch}} y_{hj} + \frac{m_h}{m_h \sum_{j\in s_h} z_{hj} - \sum_{j\in s_{Ch}} z_{hj}} \sum_{j\in s_{\bar{C}h}} y_{hj}\right]. \tag{3.9}$$

The variance of $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L}\left\{\sum_{i=1}^{N_h-1}\sum_{j=i+1}^{N_h}\left[1 - \sum_{s_h \ni i,j} \frac{P(s_h \mid i)P(s_h \mid j)}{P(s)}\right]\left(\frac{y_{hi}}{z_{hi}} - \frac{y_{hj}}{z_{hj}}\right)^2 z_{hi}z_{hj}\right\}. \tag{3.10}$$

For $m_h > 2$, an unbiased estimator of $V(\bar{y}_{st})$ is

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L}\frac{1}{\left(m_h \sum_{i\in s_h} z_{hi} - \sum_{i\in s_{Ch}} z_{hi}\right)}\left\{k_{h1}\sum_{i\in s_{Ch}}\sum_{j<i\in s_{Ch}}\left(\frac{y_{hi}}{z_{hi}} - \frac{y_{hj}}{z_{hj}}\right)^2 z_{hi}z_{hj} + k_{h2}\sum_{i\in s_{Ch}}\sum_{j\in s_{\bar{C}h}}\left(\frac{y_{hi}}{z_{hi}} - \frac{y_{hj}}{z_{hj}}\right)^2 z_{hi}z_{hj}\right.$$

$$\left. + k_{h3}\sum_{i\in s_{\bar{C}h}}\sum_{j<i\in s_{\bar{C}h}}\left(\frac{y_{hi}}{z_{hi}} - \frac{y_{hj}}{z_{hj}}\right)^2 z_{hi}z_{hj}\right\}, \tag{3.11}$$

where $k_{h1} = \frac{(N_h - 1)(m_h - 2)}{n_h - 2} - \frac{(m_h - 1)^2}{\left(m_h \sum_{i\in s_h} z_{hi} - \sum_{i\in s_{Ch}} z_{hi}\right)}$,

$$k_{h2} = \frac{(N_h - 1)(m_h - 1)}{n_h - 2} - \frac{m_h (m_h - 1)}{\left(m_h \sum_{i\in s_h} z_{hi} - \sum_{i\in s_{Ch}} z_{hi}\right)}$$

and $k_{h3} = \frac{(N_h - 1)m_h}{n_h - 2} - \frac{m_h^2}{\left(m_h \sum_{i\in s_h} z_{hi} - \sum_{i\in s_{Ch}} z_{hi}\right)}$.

Proof: Murthy's unbiased estimator of the population total in stratum h is

$$\hat{Y}_h = \sum_{j=1}^{n_h} y_{hj} \frac{P(s_h \mid j)}{P(s_h)}, \tag{3.12}$$

where n_h is the sample size. Under stratified inverse PPS sampling without replacement,

$$\frac{P(s_h \mid j)}{P(s_h)} = \frac{P(s_h, j)}{z_{hj} P(s_h)} = \begin{cases} \dfrac{m_h - 1}{m_h \sum_{j\in s_h} z_{hj} - \sum_{j\in s_{Ch}} z_{hj}} & ; j \in s_{Ch} \\ \dfrac{m_h}{m_h \sum_{j\in s_h} z_{hj} - \sum_{j\in s_{Ch}} z_{hj}} & ; j \in s_{\bar{C}h} \end{cases}.$$

By substituting $P(s_h \mid j)/P(s_h)$ into (3.12), an unbiased estimator of Y_h is obtained. This estimator does not depend on the order of selection of the units. The variance of $\hat{Y}_h$ is given by

$$V(\hat{Y}_h) = \sum_{i=1}^{N_h} \sum_{j<i}^{N_h} \left[1 - \sum_{s_h \ni i,j} \frac{P(s_h \mid i) P(s_h \mid j)}{P(s_h)} \right] \left(\frac{y_{hi}}{z_{hi}} - \frac{y_{hj}}{z_{hj}} \right)^2 z_{hi} z_{hj} . \tag{3.13}$$

An unbiased estimator of the variance of estimator $\hat{Y}_h$ is

$$\hat{V}(\hat{Y}_h) = \sum_{i=1}^{n_h} \sum_{j<i}^{n} \left[\frac{P(s_h \mid i,j)}{P(s_h)} - \frac{P(s_h \mid i) P(s_h \mid j)}{\left[P(s_h)\right]^2} \right] \left(\frac{y_{hi}}{z_{hi}} - \frac{y_{hj}}{z_{hj}} \right)^2 z_{hi} z_{hj} , \tag{3.14}$$

where $P(s_h \mid i,j)$ refers to the probability of getting sample s given that the i-th and j-th units are selected in any order in the first two draws. From stratified inverse PPS sampling without replacement, for $m_h > 2$,

$$\frac{P(s_h \mid i,j)}{P(s_h)} = \begin{cases} \dfrac{(m_h - 2)(N_h - 1)}{(n_h - 2)\left(m_h \sum_{i\in s_h} z_{hi} - \sum_{i\in s_{Ch}} z_{hi} \right)} & ; i, j \in s_{Ch} \\ \dfrac{(m_h - 1)(N_h - 1)}{(n_h - 2)\left(m_h \sum_{i\in s_h} z_{hi} - \sum_{i\in s_{Ch}} z_{hi} \right)} & ; i \in s_{Ch} \text{ and } j \in s_{\bar{C}h} \\ \dfrac{m_h (N_h - 1)}{(n_h - 2)\left(m_h \sum_{i\in s_h} z_{hi} - \sum_{i\in s_{Ch}} z_{hi} \right)} & ; i, j \in s_{\bar{C}h} \end{cases}.$$

By substituting the expressions $\frac{P(s_h \mid j)}{P(s_h)}$ and $\frac{P(s_h \mid i,j)}{P(s_h)}$ into (3.14), an unbiased estimator of the variance of $\hat{Y}_h$ is obtained. Because of the independence of samples across strata, we get (3.9) and (3.11).

Corollary 3.3 In stratified inverse PPS sampling without replacement, an unbiased estimator of Y_C is given by

$$\hat{Y}_{Cst} = \sum_{h=1}^{L}\left[\frac{(m_h - 1)}{m_h \sum_{j\in s_h} z_{hj} - \sum_{j\in s_{Ch}} z_{hj}} \sum_{j\in s_{Ch}} y_{hj}\right], \tag{3.15}$$

with the variance

$$V\left(\hat{Y}_{Cst}\right) = \sum_{h=1}^{L}\left\{\sum_{i=1}^{N_h-1}\sum_{j=i+1}^{N_h}\left[1 - \sum_{s_h \ni i,j}\frac{P(s_h \mid i)P(s_h \mid j)}{P(s_h)}\right]\left(\frac{y''_{hi}}{z_{hi}} - \frac{y''_{hj}}{z_{hj}}\right)^2 z_{hi}z_{hj}\right\}, \tag{3.16}$$

where $\sum_{s_h \ni i,j}$ refers to summation over all samples, s_h which contains the i-th and j-th units. For $m_h > 2$, an unbiased estimator of $V\left(\hat{Y}_{Cst}\right)$ is

$$\hat{V}\left(\hat{Y}_{Cst}\right) = \sum_{h=1}^{L}\frac{1}{\left(m_h \sum_{j\in s_h} z_{hj} - \sum_{j\in s_{Ch}} z_{hj}\right)}\left\{k_{h1}\left[\left(\sum_{j\in s_{Ch}} z_{hj}\right)\left(\sum_{j\in s_{Ch}}\frac{y_{hj}^2}{z_{hj}}\right) - \left(\sum_{j\in s_{Ch}} y_{hj}\right)^2\right]\right.$$

$$\left. + k_{h2}\left[\left(\sum_{j\in s_{\bar{C}h}} z_{hj}\right)\left(\sum_{j\in s_{Ch}}\frac{y_{hj}^2}{z_{hj}}\right)\right]\right\}. \tag{3.17}$$

Proof: We have $Y_{Ch} = \sum_{j=1}^{N_h} y''_{hj}$ and $Y_C = \sum_{h=1}^{L} Y_{Ch}$ is estimated unbiasedly by $\hat{Y}_{Cst}$ in (3.15). For the study value y''_{hj}, taking on values y_{hj} or 0, we get that the variance of $\hat{Y}_{Cst}$ and its unbiased estimate are given by expressions (3.16) and (3.17), respectively.

Corollary 3.4 In stratified inverse PPS sampling without replacement, an unbiased estimator of M is given by

$$\hat{M}_{st} = \sum_{h=1}^{L}\frac{m_h(m_h - 1)}{m_h \sum_{j\in s_h} z_{hj} - \sum_{j\in s_{Ch}} z_{hj}}. \tag{3.18}$$

For $m_h > 2$, an unbiased variance estimator of $\hat{M}_{st}$ is given by

$$\hat{V}\left(\hat{M}_{st}\right)=\sum_{h=1}^{L}\frac{1}{\left(m_h\sum_{j\in s_h}z_{hj}-\sum_{j\in s_{Ch}}z_{hj}\right)}\left\{k_{h1}\left[\left(\sum_{h\in s_{Ch}}z_{hj}\right)\left(\sum_{j\in s_{Ch}}\frac{1}{z_{hj}}\right)-m_h^2\right]\right.$$

$$\left.+k_{h2}\left(\sum_{j\in s_{\bar{C}h}}z_{hj}\right)\left(\sum_{j\in s_{Ch}}\frac{1}{z_{hj}}\right)\right\}. \qquad (3.19)$$

Proof: We have $M_h=\sum_{j=1}^{N_h}y'_{hj}$ and $M=\sum_{h=1}^{L}M_h$ is estimated by expression (3.18). So $\hat{M}_{st}$ is an unbiased estimator of M. For the study value y'_{hj}, taking on values 1 or 0, an unbiased estimator of $V\left(\hat{M}_{st}\right)$ is obtained as expressions (3.19).

Theorem 3.4 Under stratified inverse PPS sampling without replacement, the bias of a ratio estimator, $\bar{y}_{Cst}=\frac{\hat{Y}_{Cst}}{\hat{M}_{st}}$ of $\bar{Y}_C$ is given by $B(\bar{y}_{Cst})=\frac{-Cov\left(\bar{y}_{Cst},\hat{M}_{st}\right)}{M}$ and its bound is $\frac{\left|B(\bar{y}_{Cst})\right|}{\left[V(\bar{y}_{Cst})\right]^{\frac{1}{2}}}\le CV\left(\hat{M}_{st}\right)$. An approximate mean squared error of the estimator $\bar{y}_{Cst}$ is given by $MSE(\bar{y}_{Cst})\approx\frac{1}{M^2}\left[\bar{Y}_C^2V\left(\hat{M}_{st}\right)+V\left(\hat{Y}_{Cst}\right)-2\bar{Y}_CCov\left(\hat{Y}_{Cst},\hat{M}_{st}\right)\right]$ where

$$Cov\left(\hat{Y}_{Cst},\hat{M}_{st}\right)=\sum_{h=1}^{L}\sum_{i=1}^{N}\sum_{j<i}^{N}\left[1-\sum_{s_h\ni i,j}\frac{P(s_h\,|\,i)P(s_h\,|\,j)}{P(s_h)}\right]\left(\frac{y''_{hi}}{z_{hi}}-\frac{y''_{hj}}{z_{hj}}\right)\left(\frac{x'_{hi}}{z_{hi}}-\frac{x'_{hj}}{z_{hj}}\right)z_{hi}z_{hj}\,,$$

$$y''_{hj}=\begin{cases}y_{hj} & ;j\in C_h\\ 0 & ;j\in\bar{C}_h\end{cases}\quad\text{and}\quad x'_{hj}=\begin{cases}1 & ;j\in C_h\\ 0 & ;j\in\bar{C}_h\end{cases}.$$

Proof: The proof is similarly to that in theorem 3.2.

4. Stratified Inverse Random Sampling

4.1 Stratified Inverse Random Sampling with Replacement

If the probabilities of selection are equal to $z_{hj}=\frac{1}{N_h}$ for $j=1,2,\ldots,N_h$ and $h=1,2,\ldots,L$, stratified inverse PPS sampling with replacement reduces to stratified inverse random sampling with replacement. Estimators under stratified inverse PPS sampling are reduced to the following forms. An unbiased estimator of $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \left[\hat{P}_h \bar{y}_{Ch} + (1-\hat{P}_h)\bar{y}_{\bar{C}h} \right], \tag{4.1}$$

where $W_h = \frac{N_h}{N}$, $\hat{P}_h = \frac{m_h - 1}{n_h - 1}$, $\bar{y}_{Ch} = \frac{1}{m_h}\sum_{j\in s_{Ch}} y_{hj}$ and $\bar{y}_{\bar{C}h} = \frac{1}{m_h}\sum_{j\in s_{\bar{C}h}} y_{hj}$. The variance of $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \left[\left(\bar{Y}_{Ch} - \bar{Y}_{\bar{C}h}\right)^2 V\left(\hat{P}_h\right) + \frac{\sigma_{Ch}^2}{m_h} E\left(\hat{P}_h^2\right) + \frac{\sigma_{\bar{C}h}^2}{m_h - 1} E\left[\hat{P}_h\left(1-\hat{P}_h\right)\right]\right], \tag{4.2}$$

where $\bar{Y}_{Ch} = \frac{1}{M_h}\sum_{j\in Ch} y_{hj}$, $\bar{Y}_{\bar{C}h} = \frac{1}{N_h - M_h}\sum_{j\in \bar{C}h} y_{hj}$, $\sigma_{Ch}^2 = \frac{1}{M_h}\sum_{j\in Ch}\left(y_{hj} - \bar{Y}_{Ch}\right)^2$ and

$\sigma_{\bar{C}h}^2 = \frac{1}{N_h - M_h}\sum_{j\in \bar{C}h}\left(y_{hj} - \bar{Y}_{\bar{C}h}\right)^2$**.** An unbiased estimator of the variance is

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \left[\left(\bar{y}_{Ch} - \bar{y}_{\bar{C}h}\right)^2 \hat{V}\left(\hat{P}_h\right) + \frac{\hat{\sigma}_{Ch}^2}{m_h}\hat{P}_h^* + \frac{\hat{\sigma}_{\bar{C}h}^2}{m_h - 1}\left(\hat{P}_h - \frac{m_h - 1}{m_h - 2}\hat{P}_h^*\right)\right], \tag{4.3}$$

where $\hat{V}\left(\hat{P}_h\right) = \frac{\hat{P}_h\left(1-\hat{P}_h\right)}{n_h - 2}$, $\hat{P}_h^* = \frac{(m_h - 1)(m_h - 2)}{(n_h - 1)(n_h - 2)}$, $\hat{\sigma}_{Ch}^2 = \frac{1}{m_h - 1}\sum_{j\in s_{Ch}}\left(y_{hj} - \bar{y}_{Ch}\right)^2$ and

$\hat{\sigma}_{\bar{C}h}^2 = \frac{1}{n_h - m_h - 1}\sum_{j\in s_{\bar{C}h}}\left(y_{hj} - \bar{y}_{\bar{C}h}\right)^2$. Estimators of $\bar{Y}_C$ and M are derived in the same ways.

4.2 Stratified Inverse Random Sampling without Replacement

If the probabilities of selection are equal to $z_{hj} = \frac{1}{N_h}$ for $h = 1, 2, \ldots, N_h$ and $j = 1, 2, \ldots, L$, stratified inverse PPS sampling without replacement reduces to stratified inverse random sampling without replacement. An unbiased estimator of $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \left[\hat{P}_h \bar{y}_{Ch} + (1-\hat{P}_h)\bar{y}_{\bar{C}h} \right], \tag{4.4}$$

where $\hat{P}_h = \frac{m_h - 1}{n_h - 1}$, $\bar{y}_{Ch} = \frac{1}{m_h}\sum_{j\in s_{Ch}} y_{hj}$ and $\bar{y}_{\bar{C}h} = \frac{1}{m_h}\sum_{j\in s_{\bar{C}h}} y_{hj}$. The variance of $\bar{y}_{st}$ is

$$V\left(\bar{y}_{st}\right)=\sum_{h=1}^{L}W_h^2\left[\left(\bar{Y}_{Ch}-\bar{Y}_{\bar{C}h}\right)^2 V\left(\hat{P}_h\right)+\left(1-\frac{m_h}{M_h}\right)\frac{S_{Ch}^2}{m_h}E\left(\hat{P}_h^2\right)\right.$$

$$\left.+\frac{S_{\bar{C}h}^2}{m_h-1}E\left[\hat{P}_h\left(1-\hat{P}_h\right)\left(1-\frac{n_h-m_h}{N_h-M_h}\right)\right]\right], \qquad (4.5)$$

where $\bar{Y}_{Ch}=\frac{1}{M_h}\sum_{j\in Ch}y_{hj}$, $\bar{Y}_{\bar{C}h}=\frac{1}{N_h-M_h}\sum_{j\in\bar{C}h}y_{hj}$, $S_{Ch}^2=\frac{1}{M_h-1}\sum_{j\in Ch}\left(y_{hj}-\bar{Y}_{Ch}\right)^2$ and

$S_{\bar{C}h}^2=\frac{1}{N_h-M_h-1}\sum_{j\in\bar{C}h}\left(y_{hj}-\bar{Y}_{\bar{C}h}\right)^2$ **.** An unbiased estimator of the variance is

$$\hat{V}\left(\bar{y}_{st}\right)=\sum_{h=1}^{L}W_h^2\left[\left(\bar{y}_{Ch}-\bar{y}_{\bar{C}h}\right)^2\hat{V}\left(\hat{P}_h\right)+\frac{\hat{S}_{Ch}^2}{m_h}\left(\hat{P}_h^*-\frac{m_h}{N_h}\hat{P}_h\right)\right.$$

$$\left.+\hat{S}_{\bar{C}h}^2\left(\frac{\hat{P}_h\left(1-\hat{P}_h\right)}{m_h-1}-\frac{1-\hat{P}_h}{N_h}-\frac{\hat{V}\left(\hat{P}_h\right)}{n_h-m_h}\right)\right], \qquad (4.6)$$

where $\hat{V}\left(\hat{P}_h\right)=\frac{\hat{P}_h\left(1-\hat{P}_h\right)}{n_h-2}$, $\hat{P}_h^*=\frac{(m_h-1)(m_h-2)}{(n_h-1)(n_h-2)}$, $\hat{S}_{Ch}^2=\frac{1}{m_h-1}\sum_{j\in s_{Ch}}\left(y_{hj}-\bar{y}_{Ch}\right)^2$ and

$\hat{S}_{\bar{C}h}^2=\frac{1}{n_h-m_h-1}\sum_{j\in s_{\bar{C}h}}\left(y_{hj}-\bar{y}_{\bar{C}h}\right)^2$. Estimators of $\bar{Y}_C$ and M are similarly derived. Note that the variance $V\left(\bar{y}_{st}\right)$ decreases when the number of sampled units in C_h increases.

5. Simulation Study

To compare the estimators for the four sampling designs described above, simulation is used. Six populations, each of size $N=10,000$, were created from the model[1] where the correlation coefficients (ρ) between the study variable (y) and the auxiliary variable (x) are $\rho=0.1,0.2,0.5,0.6,0.8$ and 0.9. Each population is divided into five strata. The class of interest is a set of units with values of a study variable greater than or equal to the 90-th percentile of each population data set, i.e. $C_h=\left\{u_{hj}:y_{hj}\geq P_{90}\right\}$.

---

[1] The model $y_i=\alpha x_i+\left(1-\alpha\right)w_i$ where $x_i\sim N\left(500,100^2\right)$ and $w_i\sim N\left(500,100^2\right)$ is used. The value of α depends on the correlation coefficient between x and y.

From each population, $10,000$ Monte Carlo samples were drawn using the four stratified inverse sampling schemes. The values of m_h are set to be 2, 5, 8 and 10 percent of M_h. The estimates $\tilde{V}(\hat{\theta})$ and $M\tilde{S}E(\hat{\theta})$ are calculated as $\tilde{V}(\hat{\theta}) = \sum_{j=1}^{10,000} \frac{(\hat{\theta}_j - \tilde{\theta})^2}{10,000-1}$ and $M\tilde{S}E(\hat{\theta}) = \sum_{j=1}^{10,000} \frac{(\hat{\theta}_j - \theta)^2}{10,000}$ where $\tilde{\theta} = \frac{1}{10,000}\sum_{j=1}^{10,000} \hat{\theta}_j$ is the mean of the estimates of θ from 10,000 samples. The formula of relative efficiency of two estimators under sampling design A and sampling design B is $R.E.(\hat{\theta}) = \frac{\tilde{V}(\hat{\theta}_A)}{\tilde{V}(\hat{\theta}_B)}$ where $\hat{\theta}_A$ and $\hat{\theta}_B$ are estimators under sampling designs A and B, respectively. For biased estimator, we replace $\tilde{V}(\hat{\theta})$ with $M\tilde{S}E(\hat{\theta})$.

Table 5.1. Comparison of the Relative Efficiencies of the Estimates under Stratified Inverse Sampling with and without Replacement

ρ	% of M_h	**Stratified Inverse Random Sampling with to without Replacement**			**Stratified Inverse PPS Sampling with to without Replacement**		
		$R.E.(\bar{y}_{st})$	$R.E.(\hat{M}_{st})$	$R.E.(\bar{y}_{Cst})$	$R.E.(\bar{y}_{st})$	$R.E.(\hat{M}_{st})$	$R.E.(\bar{y}_{Cst})$
0.1	2	1.030	1.004	0.995	1.026	1.016	1.013
	5	1.054	1.071	1.059	1.063	1.035	1.054
	8	1.084	1.079	1.070	1.078	1.068	1.076
	10	1.128	1.123	1.088	1.119	1.114	1.116
0.2	2	1.053	1.031	1.020	1.051	1.047	1.018
	5	1.038	1.034	1.063	1.053	1.046	1.055
	8	1.107	1.104	1.071	1.101	1.058	1.056
	10	1.118	1.092	1.124	1.125	1.095	1.105
0.5	2	1.015	1.008	1.022	1.086	1.072	1.014
	5	1.051	1.032	1.051	1.199	1.064	1.054
	8	1.118	1.131	1.100	1.288	1.079	1.073
	10	1.083	1.079	1.083	1.335	1.117	1.094

ρ	% of M_h	**Stratified Inverse Random Sampling with to without Replacement**			**Stratified Inverse PPS Sampling with to without Replacement**		
		R.E.$(\bar{y}_{st})$	R.E.$(\hat{M}_{st})$	R.E.$(\bar{y}_{Cst})$	R.E.$(\bar{y}_{st})$	R.E.$(\hat{M}_{st})$	R.E.$(\bar{y}_{Cst})$
0.6	2	1.024	1.023	1.032	1.097	1.021	1.025
	5	1.067	1.052	1.038	1.270	1.106	1.039
	8	1.027	1.041	1.131	1.335	1.102	1.078
	10	1.102	1.122	1.137	1.469	1.141	1.120
0.8	2	1.050	1.038	1.030	1.232	0.947	1.004
	5	1.061	1.062	1.052	1.430	1.021	1.011
	8	1.007	1.040	1.074	1.466	1.067	1.075
	10	1.138	1.121	1.104	1.529	1.128	1.088
0.9	2	1.036	1.055	1.042	1.301	1.059	0.995
	5	1.038	1.062	1.040	1.501	1.153	1.011
	8	1.083	1.080	1.087	1.505	1.140	1.034
	10	1.051	1.067	1.107	1.532	1.098	1.053

Table 5.1. indicates that the estimates under stratified inverse random sampling without replacement are more efficient than those with replacement. The estimates under stratified inverse PPS sampling without replacement have higher efficiencies than those under stratified inverse PPS sampling with replacement.

Table 5.2. Comparison of the Relative Efficiencies of the Estimates under Stratified Inverse Random Sampling and Stratified Inverse PPS Sampling

ρ	% of M_h	**Stratified Inverse PPS Sampling to Random Sampling with Replacement**			**Inverse PPS Sampling to Random Sampling without Replacement**		
		R.E.$(\bar{y}_{st})$	R.E.$(\hat{M}_{st})$	R.E.$(\bar{y}_{Cst})$	R.E.$(\bar{y}_{st})$	R.E.$(\hat{M}_{st})$	R.E.$(\bar{y}_{Cst})$
0.1	2	0.911	0.984	0.983	0.908	0.995	1.000
	5	0.891	1.013	1.026	0.898	0.979	1.020
	8	0.908	1.014	1.001	0.903	1.003	1.006
	10	0.900	1.019	0.976	0.893	1.011	1.002

ρ	% of M_h	Stratified Inverse PPS Sampling to Random Sampling with Replacement			Stratified Inverse PPS Sampling to Random Sampling without Replacement		
		$R.E.(\bar{y}_{st})$	$R.E.(\hat{M}_{st})$	$R.E.(\bar{y}_{Cst})$	$R.E.(\bar{y}_{st})$	$R.E.(\hat{M}_{st})$	$R.E.(\bar{y}_{Cst})$
0.2	2	0.929	1.006	1.000	0.927	1.022	0.998
	5	0.895	1.001	1.008	0.909	1.012	1.000
	8	0.912	1.030	0.999	0.907	0.987	0.985
	10	0.915	0.998	1.007	0.920	1.001	0.990
0.5	2	0.930	1.009	1.009	0.995	1.074	1.001
	5	0.838	1.027	0.998	0.957	1.058	1.001
	8	0.811	1.081	1.045	0.934	1.031	1.019
	10	0.733	1.005	0.999	0.903	1.040	1.009
0.6	2	0.914	1.045	1.007	0.980	1.043	1.001
	5	0.802	1.000	1.004	0.954	1.052	1.005
	8	0.734	1.013	1.051	0.954	1.072	1.002
	10	0.683	1.045	1.016	0.910	1.063	1.001
0.8	2	1.102	1.152	1.026	1.293	1.052	1.000
	5	0.924	1.045	1.042	1.244	1.004	1.001
	8	0.857	1.063	1.018	1.248	1.090	1.019
	10	0.868	1.036	1.040	1.166	1.042	1.025
0.9	2	1.639	1.141	1.067	2.059	1.145	1.018
	5	1.431	1.100	1.051	2.069	1.195	1.022
	8	1.352	1.054	1.055	1.878	1.112	1.004
	10	1.332	1.088	1.054	1.943	1.120	1.003

Table 5.2. shows that when the correlation coefficient (ρ) is greater than 0.8, the efficiencies of the estimates under stratified PPS sampling with replacement are larger than those under stratified inverse random sampling with replacement. When the correlation coefficient is greater than 0.7 the relative efficiencies of the estimates under stratified PPS sampling without replacement are larger than those under stratified inverse random sampling without replacement. However, the relative efficiencies of the estimates decrease when m_h increases.

6. Conclusion

This paper considers stratified inverse sampling with four variations. The parameters estimated are the population mean of a study variable $(\bar{Y})$, the number of units in a class of interest (M) and the mean from units in a class of interest $(\bar{Y}_C)$. Unbiased estimators of $\bar{Y}$ and M are given and unbiased estimates of their variances are found. The biased estimate of $\bar{Y}_C$ is presented. When the number of sampled units in a class of interest increases, the variances or mean squared errors of the estimators decrease. The results of the simulation study indicate that stratified inverse random sampling without replacement have a smaller estimates of variances or mean squared error of estimators than that of estimates under random sampling with replacement. Stratified inverse PPS sampling without replacement gives higher efficient estimators than PPS sampling with replacement. Stratified inverse PPS sampling designs should be used in a survey when the auxiliary variable is highly correlated to the study variable.

References

[1] Kalton, G. and Anderson, D. Sampling rare populations. Journal of the Royal Statistical Society, Series A, 1986; 149: 65-82.

[2] Thompson, S. K., and Seber, G. A. F. Adaptive Sampling. New York: Wiley 1996.

[3] Haldane, J. B. S. On a Method of estimating frequencies, Biometrika, 1945; 33: 222-225.

[4] Finney, D.J. On a method of estimating frequencies. Biometrika, 1949; 36: 223-234.

[5] Mikulski, P.W. and Smith, P.J. A variance bound for unbiased estimation in inverse sampling. Biometrika, 1976; 63: 216-217.

[6] Sathe, Y. S. Sharper variance bounds for unbiased estimation in inverse sampling. Biometrika, 1977; 64: 425-426.

[7] Salehi, M.M. and Seber, G.A.F. A new proof of Murthy's estimator with applies to sequential sampling. Aust NZ J Stat, 2001; 43: 281-286.

[8] Christman, M.C. and Lan, F. Inverse adaptive cluster sampling. Biometrics, 2001; 57: 1096-1105.

[9] Salehi, M.M. and Seber, G.A.F. A general inverse sampling scheme and its application to adaptive cluster sampling. Aust NZ J Stat, 2004; 46: 483-494.

[10] Greco, L. and Naddeo, S. Inverse sampling with unequal selection probabilities. Communications in Statistics: Theory and Methods, 2007; 36: 1039-1048.

[11] Sangngam, P. and Suwattee, P. Modified sampling scheme in inverse sampling without replacement. International Conference on Computational and Statistical Science, June 11-12, 2010. Manila, Philippines, pp. 580-584.

[12] Midzuno, H. On the sampling system with probability proportionate to sum of sizes. AISM, 1952; 3: 99-107.