A note on an IPPS sampling scheme

BY L. N. SAHOO, B. C. DAS AND G. N. SINGH*

SUMMARY: This paper presents a new IPPS sampling scheme possessing some desirable properties and providing an unbiased and non-negative variance estimator under H-T model. An empirical study is also undertaken to examine the performance of the scheme compared to some standard sampling schemes.

KEYWORDS: Inclusion probability, joint inclusion probability, unequal probability sampling. JEL C42.

1. INTRODUCTION

Let \((y_i, x_i), i = 1, 2, ..., N\) denote observations on \(i\)th unit of a finite population \(U\) in \((y, x)\) with totals \((Y, X)\), such that \(y\) is the survey variable and \(x\) is an auxiliary variable used as a size measure. Consider an estimation of \(Y\) based on a sample \(s\) of \(n\) units selected according to some unequal probability sampling without replacement scheme with \(\pi_i\) as the inclusion probability of \(i\)th unit and \(\pi_{ij}\) as the joint inclusion probability of \(i\)th and \(j\)th units. The most commonly used estimator in this situation is the Horvitz-Thompson (1952) estimator (H-T estimator) defined by

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

From the theory developed by Horvitz and Thompson (1952) \(\sum_{i=1}^{N} \pi_i = n\),

\[
\sum_{i \neq j}^{N} \pi_{ij} = (n - 1)\pi_i \quad \text{and} \quad \sum_{i}^{N} \sum_{j<i} \pi_{ij} = \frac{1}{2}n(n - 1). 
\]

Yates and Grundy (1953) derived expression for the variance of \(\hat{Y}_{HT}\). According to Sarndal et al. (2003), the variance of the estimator is actually calculated as

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

An unbiased estimator of \(\text{Var}(\hat{Y}_{HT})\) is given by

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

A sufficient condition for this expression to be always non-negative is that \(\pi_i \pi_j < \pi_{ij}\), \(i \neq j\).
It is well known that considerable reduction in the variance of $\hat{Y}_{HT}$ can be achieved by making $\pi_i = np_i$, where $p_i = x_i/X$, $i = 1, 2, ..., N$, is the initial probability of selection of $i$th unit. Such a scheme is called an inclusion probability proportional to size (IPPS) scheme. Brewer and Hanif (1983), and Chaudhuri and Vos (1988) have made elaborate discussions on a number of IPPS schemes. But many of these methods are restricted to $n = 2$, because calculation of $\pi_{ij}$ rapidly becomes cumbersome when $n > 2$. However, IPPS schemes with $n = 2$ are very much useful in stratified sampling where stratification is sufficiently ‘deep’, i.e., the number of strata (and their sizes) is such that a sample of size 2 per stratum meets the requirement on the total sample size. Indeed, the advantages of stratified sampling are exploited to the greatest possible extent if the number of strata is maximum, and $n = 2$ is the minimum sample size to estimate the variance within each stratum (cf, Chaudhuri and Vos, 1988, p. 148). For this reason, Yates and Grundy (1953) call the special case of samples of two units ‘practically most important’.

In this paper, we suggest a new IPPS sampling scheme for $n = 2$ having the above mentioned desirable properties in terms of $\pi_i$ and $\pi_{ij}$, and also performs well as compared to some popular probability proportional to size without replacement (PPSWOR) sampling schemes for a number of natural populations.

2. THE SUGGESTED SAMPLING SCHEME

Consider the set of revised probabilities $\{P_1, P_2, ..., P_N\}$ where $P_i$ is defined by

$$P_i = \frac{p_i(1 - z_i)(2 - \lambda/\sqrt{p_i})}{1 - 2z_i}, \quad i = 1, 2, ..., N,$$

such that $z_i = \sqrt{p_i}/\sum_{i=1}^{N} \sqrt{p_i}$ and

$$\lambda = \sum_{j=1}^{N} \frac{p_j}{1 - 2z_j} \left/ \sum_{j=1}^{N} \frac{\sqrt{p_j}(1 - z_j)}{1 - 2z_j} \right.,$$

determined by solving the equation $\sum_{i=1}^{N} P_i = 1$. It is true that computation of the revised probabilities is restricted only to those situations for which $z_i < \frac{1}{2}$ and $p_i > \lambda^2/4$ for all $i$. Here our motivation behind the definition of $P_i$ in (2) is to produce an efficient sampling plan by using a square root transformation of the initial probability $p_i$ under the H-T model.

Our suggested sampling scheme consists of the following steps:

Step I. Draw the first unit, say $i$, with revised probability $P_i$ and without replacement, $\lambda$ being given by (3).

Step II. Draw the second unit, say $j$, from the remaining $(N - 1)$ units with conditional probability