ORDERED AND UNORDERED ESTIMATORS IN SAMPLING WITHOUT REPLACEMENT

M. N. MURTHY

Indian Statistical Institute, Calcutta

1. Introduction

In the last decade the probem of sampling with varying probabilities without replacement has received considerable attention and a number of estimators have been proposed. Das (1951) and Des Raj (1956) have given some ordered estimators, that is, estimators which take into account the order in which the units are drawn. In this paper it is shown that corresponding to any biased or unbiased ordered estimator there exists an unordered estimator which is more efficient than the former. The technique of improving the ordered estimators by unordered ones is also explained. This method is applied to the set of estimators given by Das and to one set of Des Raj's estimators which provide unbiased estimates of the population total and also to the unbiased variance estimators considered by them. (x_{si}, v_{si}, x'_{si}, v'_{si}, of Sections 3 and 4). Compared to other known estimators the estimator of Des Raj has two remarkable properties : (i) x_{si} is more efficient than the corresponding estimator for sampling with replacement as shown by Roy Choudhury (1956) and (ii) v_{si} is non-negative. It is shown that the technique of unordering preserves a fortiori the first property and retains the second property in two particular cases. In sampling the first unit with varying probability and the rest with equal probability without replacement, unordering of Das' estimator yields the familiar unbiased ratio estimator.

2. Unordered estimator

In sampling n units without replacement from a finite population of N units, there will be $\binom{N}{n}$ unordered samples (s). Corresponding to any unordered sample (s) of size n units, there will be $n!$ ordered samples (si). Let x_{si} $\{s = 1, 2, \ldots \binom{N}{n}; i = 1, 2, \ldots, M (= n!)\}$ be an estimator of a population parameter θ based on the ordered sample (si). Consider a scheme of selection in which the probability of selecting the ordered sample (si) is p_{si}. Then the probability p_s of getting the unordered sample (s) is the sum of the probabilities of getting the ordered samples corresponding to (s)

That is,

$$p_s = \sum_{i=1}^{M} p_{si}.$$

Theorem 1: *If* $\hat{\theta}_0 = x_{si}$ *and* $\hat{\theta}_U = \sum_{i=1}^{M} x_{si}\ p'_{si}$ *(where* $p'_{si} = p_{si}/p_s$*) are estimators of the population parameter* θ *then,*

$$\text{(i)} \quad E(\hat{\theta}_U) = E(\hat{\theta}_0)$$

$$\textit{and} \quad \text{(ii)} \quad V(\hat{\theta}_U) \leqslant V(\hat{\theta}_0).$$

where E and V stand for expectation and variance respectively.

Proof :

$$E(\theta_U) = \sum_{s=1}^{\binom{N}{n}} \theta_U p_s = \sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}\, p_{si} = E(\theta_0).$$

The variances of the estimators θ_0 and θ_U are given by

$$V(\theta_0) = \left\{ \sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}^2\, p_{si} - \left(\sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}\, p_{si} \right)^2 \right\} \qquad \dots \quad (2.1)$$

$$V(\theta_U) = \left\{ \sum_{s=1}^{\binom{N}{n}} \left(\sum_{i=1}^{M} x_{si}\, p'_{si} \right)^2 p_s - \left(\sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}\, p_{si} \right)^2 \right\}. \qquad \dots \quad (2.2)$$

Therefore,

$$V(\theta_0) - V(\theta_U) = \left\{ \sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} \left(x_{si} - \sum_{i=1}^{M} x_{si}\, p'_{si} \right)^2 p_{si} \right\}. \qquad \dots \quad (2.3)$$

This shows that the variance of the unordered estimator θ_U is less than or equal to that of the ordered estimator θ_0.

Corollary :

(*i*) *The mean square error of θ_U is less than or equal to that of θ_0.*

(*ii*) *If $\hat{V}_0(\theta_0) = v_{si}$ is an ordered estimator of $V(\theta_0)$, then $\hat{V}_U(\theta_0)$, an unordered estimator of $V(\theta_0)$, which has a lesser mean square error than $\hat{V}_0(\theta_0)$ is given by*

$$\hat{V}_U(\theta_0) = \left\{ \sum_{i=1}^{M} v_{si}\, p'_{si} \right\}. \qquad \dots \quad (2.4)$$

(*iii*) *An estimator $\hat{V}_U(\theta_U)$ of the variance of θ_U is given by*

$$\hat{V}_U(\theta_U) = \left\{ \theta_U^2 - \sum_{i=1}^{M} (x_{si}^2 - v_{si}) p'_{si} \right\}. \qquad \dots \quad (2.5)$$

(*iv*) *The 2k-th moment of θ_U is less than or equal to that of θ_0.*

That is, $$E(\theta_U - E(\theta_U))^{2k} \leqslant E(\theta_0 - E(\theta_0))^{2k}.$$

It may be noted that an estimator based on $m (m = 2, \dots, M-1)$ ordered samples drawn from the M ordered samples corresponding to the unordered sample (s) with probability proportional to the sum of their probabilities will be more efficient than the ordered estimator θ_0.

3. Unordering of Des Raj's estimators

Let a sample of size n be drawn from a finite population of N units with varying probabilities without replacement. Suppose the probabilities of selection at the first draw are

$$\{p_j\},\ (j = 1, 2, \ldots, N),\ p_j > 0,\ \sum_{j=1}^{N} p_j = 1. \qquad \ldots \quad (3.1)$$

The scheme of selection of a unit at a particular draw depends on the units already drawn in the sample and not on the order in which they were drawn. For instance, the probabilities of selection at the third draw given that the k-th and l-th units have already been chosen in the first two draws will be given by

$$\left\{ \frac{p_j}{1-p_k-p_l} \right\},\ (j \neq k \neq l).$$

Let $(y_1, y_2, \ldots, y_n)$ and $(p_1, p_2, \ldots, p_n)$ be the values of the units arranged in the order of selection in the sample drawn according to the above scheme and their respective initial probabilities. One of the sets of estimators given by Des Raj in this situation is given by

$$\left.\begin{aligned} x_{si1} &= \left\{ \frac{y_1}{p_1} \right\}, \\ x_{si2} &= \left\{ y_1 + \frac{y_2}{p_2}(1-p_1) \right\} \\ x_{sin} &= y_1+y_2+\cdots+y_{n-1}+\left(\frac{y_n}{p_n}\right)(1-p_1-p_2-\cdots-p_{n-1}) \end{aligned}\right\} \qquad \ldots \quad (3.2)$$

Each of the above estimators is unbiased for the population total and, therefore,

$$\bar{x}_{si} = \frac{1}{n}\left\{ \sum_{j=1}^{n} x_{sij} \right\}, \qquad \ldots \quad (3.3)$$

is also so. By making use of the fact that x_{sij} and $x_{sij'}$ $(j \neq j')$ are uncorrelated, Des Raj was able to get a non-negative estimator of the variance of $\bar{x}_{si}$ which is given by

$$\hat{V}_0(\bar{x}_{si}) = v_{si} = \frac{1}{n(n-1)}\left\{ \sum_{j=1}^{n} (x_{sij}-\bar{x}_{si})^2 \right\} \qquad \ldots \quad (3.4)$$

By applying the earlier theorem to $\bar{x}_{si}$ and v_{si} we get more efficient estimators of the population total and the variance of $\bar{x}_{si}$ respectively.

Theorem 2: *Unordering of the ordered estimator,*

$$\theta_0 = \sum_{j=1}^{n} c_j x_{ij}, \sum_{j=1}^{n} c_j = 1, x_{ij} = y_1 + y_2 + \cdots + y_{j-1} + \frac{y_j}{p_j}(1 - p_1 - \cdots p_{j-1}), \quad \ldots \quad (3.5)$$

yields an unordered estimator which is independent of the set c_j, $\left\{ \sum_{j=1}^{n} c_j = 1 \right\}$, *namely,*

$$\frac{\sum_{l=1}^{n} y_l P(s/l)}{P(s)}, \quad \ldots \quad (3.6)$$

where $P(s/l)$ is the conditional probability of getting the unordered sample (s) given that l-th unit has been selected at the first draw and $p(s)$ is the unconditional probability of getting (s).

Proof: Let $P(s/l_1, l_2, \ldots, l_j)$ denote the conditional probability of getting the unordered sample (s) given that l_1-th, l_2-th, ..., l_j th units have been selected in the first j draws.

The coefficient of y_l in the estimator got by unordering θ_0 in the usual way (Theorem 1) is given by $\frac{1}{P(s)}$ times

$$\sum_{j=1}^{n} c_j \left[\sum_{k=1}^{j-1} \left\{ \sum_{\substack{l_1, l_2, \ldots l_{k-1} \\ l_1 \neq l_2 \neq \ldots l_{k-1} \neq l}} \frac{p_l p_{l_1} p_{l_2} \cdots p_{l_{k-1}}}{(1-p_{l_1})(1-p_{l_1}-p_{l_2})\ldots(1-p_{l_1}-p_{l_2}-\ldots p_{l_{k-1}})} \times P(s/l\, l_1\, l_2 \ldots l_{k-1}) \right\} \right.$$

$$\left. + \sum_{\substack{l_1, l_2, \ldots, l_{j-1} \\ l_1 \neq l_2 \neq \ldots \neq l_{j-1} \neq l}} \frac{p_l p_{l_1} p_{l_2} \cdots p_{l_{j-1}}}{(1-p_{l_1})(1-p_{l_1}-p_{l_2})\ldots(1-p_{l_1}-p_{l_2}-\ldots-p_{l_{j-1}})} \times P(s/l\, l_1\, l_2 \ldots l_{j-1}) \times \frac{1-p_{l_1}-p_{l_2}-\ldots p_{l_{j-1}}}{p_l} \right] \ldots \quad (3.7)$$

The coefficient of c_1 in the above expression is $P(s/l)$. The theorem will be proved if we show that the coefficient of c_{j+1} is equal to that of c_j in the expression (3.7).

The first $(j-1)$ terms in both the coefficients are the same. The j-th and $(j+1)$th terms in the coefficient of c_{j+1} reduce to the j-th term in the coefficient of c_j because of the equality

$$P(s/l\, l_1\, l_2 \ldots l_{j-1}) = \sum_{\substack{l_j \\ l_j \neq l_{j-1} \neq \ldots \neq l_1 \neq l}} \frac{p_{l_j}}{(1-p_l-p_{l_1}-\ldots p_{l_{j-1}})} \times P(s/l\, l_1\, l_2 \ldots, l_j).$$

Therefore,

$$\theta_s = \bar{x}_s = \frac{\sum_{i=1}^{n} y_i P(s/i)}{P(s)}$$

The estimator given in (3.6) is a special case of a general class of estimators considered by Nanjamma, Murthy and Sethi, in an unpublished paper submitted to *Sankhyā*.

Since the unbiased estimators x_{sij} and $x_{sij'}$ $(j \neq j')$ of the population total Y are uncorrelated, we get

$$\sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} (x_{sij}\, x_{sij'})\, p_{si} = Y^2.$$

As this is true for all the $\binom{n}{2}$ pairs of the n estimators,

$$\sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} \Big\{ \sum_{j>j'}^{n} x_{sij}\, x_{sij'} \Big\}\, p_{si} = \binom{n}{2} Y^2.$$

Hence an unbiased estimator of Y^2 is given by

$$(\hat{Y^2}) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{M} \Big(\sum_{j>j'}^{n} x_{sij}\, x_{sij'} \Big)\, p_{si}. \quad \ldots \quad (3.8)$$

From this it follows that an unbiased estimator of the variance of $\bar{x}_s$ is

$$\hat{V}_U(\bar{x}_s) = \bar{x}_s^2 - \frac{1}{\binom{n}{2}} \sum_{i=1}^{M} \Big(\sum_{j>j'}^{n} x_{sij}\, x_{sij'} \Big)\, p_{si}$$

$$= \frac{1}{[P(s)]^2} \Big[\sum_{i=1}^{n} P(s/i)\{P(s/i) - P(s)\} y_i^2 + 2 \sum_{i>i'}^{n} \{P(s/i)\, P(s/i') - P(s) P(s/ii')\} y_i\, y_{i'} \Big] \ldots \quad (3.9)$$

The following particular cases of equations (3.6) and (3.9) will now be considered.

(i) *Simple random sampling without replacement.* In this case the estimators x_{si}, $\bar{x}_s$ and their variances are given by

$$x_{si} = \frac{1}{n} \Big\{ \sum_{j=1}^{n} (N+n+1-2j) y_j \Big\}, \quad \ldots \quad (3.10)$$

$$\bar{x}_s = \frac{N}{n} \Big\{ \sum_{j=1}^{n} y_j \Big\}, \quad \ldots \quad (3.11)$$

$$V(x_{si}) = \frac{\sigma^2}{n} \Big\{ N(N-n) + \frac{n^2-1}{3} \Big\}, \quad \ldots \quad (3.12)$$

and
$$V(\bar{x}_s) = \frac{\sigma^2}{n} N(N-n), \qquad \ldots (3.13)$$

where
$$\sigma^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N}(y_i-\bar{y})^2\right\}.$$

Therefore,
$$V(x_n) - V(\bar{x}_s) = \frac{n^2-1}{3n}\,\sigma^2. \qquad \ldots (3.14)$$

Comparison of the above expressions for the variances shows that $V(\bar{x}_s) < V(x_n)$. This is otherwise obvious also as $N\bar{y}$ is known to be the best unbiased linear estimator of Y. It is interesting to note that the variance estimator of $\bar{x}_s$ given in equation (3.9) reduces to the estimator commonly used, namely,

$$\hat{V}_U(\bar{x}_s) = N(N-n)\,\frac{s^2}{n}, \qquad \ldots (3.15)$$

where
$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}.$$

It may be pointed out that the divergence of Des Raj's estimator from the best unbiased linear estimator led to a search for a more efficient estimator.[1]

(ii) *Sampling of two units with varying probabilities without replacement.* This case is of importance as in actual practice one will, in general, be choosing two units from each stratum in stratified sampling. In this case x_{si} and $\bar{x}_s$ are given by

$$x_{si} = \tfrac{1}{2}\left\{(1+p_1)\frac{y_1}{p_1} + (1-p_1)\frac{y_2}{p_2}\right\}, \qquad \ldots (3.16)$$

and
$$\bar{x}_s = \frac{1}{2-p_1-p_2}\left\{(1-p_2)\frac{y_1}{p_1} + (1-p_1)\frac{y_2}{p_2}\right\}. \qquad \ldots (3.17)$$

The sampling variances of these two estimators are given by

$$V(x_{si}) = \tfrac{1}{4}\left\{\sum_{s=1}^{\binom{N}{2}} p_1 p_2 (2-p_1-p_2)\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2\right\}, \qquad \ldots (3.18)$$

and
$$V(\bar{x}_s) = \left\{\sum_{s=1}^{\binom{N}{2}} p_1 p_2 \frac{(1-p_1-p_2)}{(2-p_1-p_2)}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2\right\}. \qquad \ldots (3.19)$$

[1] Lahiri conjectured that Des Raj's estimators can be improved by weighting the different ordered estimators by their respective probabilities and in fact suggested the estimator given in (3.17)

Therefore,
$$V(x_{st})-V(x_s) = \tfrac{1}{4}\left\{\sum_{s=1}^{\binom{N}{2}} p_1\, p_2 \frac{(p_1+p_2)^2}{(2-p_1-p_2)}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2\right\}. \qquad \dots \quad (3.20)$$

In the case of simple random sampling without replacement the expression (3.18) above becomes

$$V(x_{st})-V(x_s) = \frac{\sigma^2}{2},$$

which is a particular case of the expression (3.12).

The estimator of variance of x_{st} given by Des Raj is,

$$\hat{V}_0(x_{st}) = v_{st} = \tfrac{1}{4}\,(1-p_1)^2 \left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2. \qquad \dots \quad (3.21)$$

By applying the earlier theorem to this v_{st}, we get a more efficient estimator of the variance, namely,

$$\hat{V}_U(x_{st}) = \tfrac{1}{4}\,(1-p_1)(1-p_2) \left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2, \qquad \dots \quad (3.22)$$

and this also is non-negative. Substituting the relevant values in equation (3·7), we get an unbiased estimator of the variance of x_s, namely,

$$\hat{V}_U(x_s) = \frac{(1-p_1)(1-p_2)(1-p_1-p_2)}{(2-p_1-p_2)^2}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2, \qquad \dots \quad (3.23)$$

which is always non-negative.

4. Unordering of Das' estimators

With the notation adopted in section 3, the set of estimators proposed by Das is given by,

$$x'_{st1} = \left(\frac{y_1}{p_1}\right)$$

$$x'_{st2} = \frac{1}{(N-1)}\,\frac{(1-p_1)}{p_1}\left(\frac{y_2}{p_2}\right)$$

$$x'_{stn} = \frac{(1-p_1)(1-p_1-p_2)\dots(1-p_1-p_2-\dots-p_{n-1})}{(N-1)(N-2)\dots(N-n+1)p_1\,p_2\dots p_{n-1}}\left(\frac{y_n}{p_n}\right). \qquad \dots \quad (4.1)$$

Each of the above estimators is unbiased for the population total Y and so is their mean

$$x'_{st} = \frac{1}{n}\sum_{j=1}^{n} x'_{stj}. \qquad \dots \quad (4.2)$$

An unbiased estimator of the variance of z'_{it} is

$$\hat{V}_{o}(z'_{s1}) = v'_{it} = z'^{2}_{it} - \left\{ \frac{1}{n} \sum_{j=1}^{n} \left(x'_{nj}\, y_j \right) + \frac{N-1}{\binom{n}{2}} \sum_{j > j'=1}^{n} (x'_{nj}\, y_{j'}) \right\}. \quad \ldots \quad (4.3)$$

In sampling the first unit with varying probability and the remaining $(n-1)$ units from $(N-1)$ units with equal probability without replacement, the estimator z'_{it} becomes

$$z_{it} = \frac{1}{n} \left\{ \frac{\sum_{i=1}^{n} y_i}{p_1} \right\}. \quad \ldots \quad (4.4)$$

Applying the Theorem 1 to this z'_{it}, we get the unordered estimator

$$z'_{s} = \left\{ \frac{\sum_{i=1}^{n} y_i}{\sum_{j=1}^{n} p_j} \right\}. \quad \ldots \quad (4.5)$$

This shows that the above unbiased ratio estimator is more efficient than the estimator given by Das. Lahiri (1951) and Midjuno (1952) have given sampling procedures which lead to the above estimator.

Substituting the relevant values in equation (2.5) we get an unbiased variance estimator of $\bar{z}'_{s}$, namely,

$$\hat{V}_{U}(z'_{s}) = z'^{2}_{s} - \left\{ (N-1) \frac{\left(\sum_{j=1}^{n} y_j \right)^2 - (N-n) \left(\sum_{j=1}^{n} y_j^2 \right)}{(n-1) \left(\sum_{j=1}^{n} p_j \right)} \right\} \quad \ldots \quad (4.6)$$

5. Numerical Examples

To study the relative performance of the ordered and unordered estimators, the following population given by Yates and Grundy (1953) will be considered.

unit	p	y	y/p
1	.1	0.5	5
2	.2	1.2	6
3	.3	2.1	7
4	.4	3.2	8
total	1.0	7.0	

This was deliberately chosen by them as being more extreme than will normally be encountered in practice. The object is to estimate the population total by selecting two units. Two schemes of selection will be considered:

Case (i): first unit with varying probability and second unit with equal probability without replacement;

Case (ii): both the units with varying probabilities without replacement.

For the prupose of comparison, the estimator and the variance estimator proposed by Horvitz and Thompson (1952) together with the variance estimator given by Yates and Grundy are also considered.

For the sake of convenience, the expressions for the different estimators given in subsequent tables are given below. Suppose (y_1, y_2) is the ordered sample drawn from a finite population of N units.

Case (i):

$$z_{st} = \frac{1}{2}\left\{ (1+p_1)\left(\frac{y_1}{p_1}\right) + (N-1)y_2 \right\}. \quad \ldots (5.1)$$

$$\hat{V}_0(z_{st}) = v_{st} = \frac{1}{4}\left\{(1-p_1)\left(\frac{y_1}{p_1}\right) - (N-1)y_2\right\}^2. \quad \ldots (5.2)$$

$$\hat{V}_U(\bar{z}_{st}) = \frac{v_{s1}\, p_1 + v_{s2}\, p_2}{p_1+p_2}. \quad \ldots (5.3)$$

$$z_s = \frac{1}{2(p_1+p_2)}\left[\left\{ 1+p_1+(N-1)p_2\right\} y_1 + \left\{1+p_2+(N-1)p_1\right\} y_2\right]. \quad (5.4)$$

$$\hat{V}_U(\bar{z}_s) = \hat{V}_U(z_{st}) - \frac{(z_{s1}-z_s)^2\, p_1 + (z_{s2}-z_s)^2 p_2}{p_1+p_2}. \quad \ldots (5.5)$$

$$z'_{st} = \frac{y_1+y_2}{2p_1}. \quad \ldots (5.6)$$

$$\hat{V}_0(z'_{st}) = v'_{st} = z'^2_{st} - \frac{1}{2p_1}\left\{y_1^2+y_2^2+2(N-1)y_1y_2\right\} \quad \ldots (5.7)$$

$$\hat{V}_U(z'_{st}) = \frac{v'_{s1}\, p_1 + v'_{s2}\, p_2}{p_1+p_2}. \quad \ldots (5.8)$$

$$z_s = \left\{\frac{y_1+y_2}{p_1+p_2}\right\}. \quad \ldots (5.9)$$

$$\hat{V}_U(z'_s) = z'^2_s - \left\{\frac{y_1^2+y_2^2+2(N-1)\, y_1\, y_2}{p_1+p_2}\right\}. \quad \ldots (5.10)$$

$$\bar{y}_{HT} = \left(\frac{y_1}{\pi_1}\right) + \left(\frac{y_2}{\pi_2}\right) \text{ where } \pi_1 = \frac{1}{N-1}\,\{(N-2)p_1+1\}$$

and

$$\pi_2 = \frac{1}{N-1}\,\{(N-2)p_2+1\}. \quad \ldots (5.11)$$

$$\hat{V}_{HT}(\hat{y}_{HT}) = (1-\pi_1)\left(\frac{y_1^2}{\pi_1^2}\right)+(1-\pi_2)\left(\frac{y_2^2}{\pi_2^2}\right)+2\ \frac{\pi_{12}-\pi_1\pi_2}{\pi_{12}}\ \frac{y_1\,y_2}{\pi_1\pi_2},$$

where
$$\pi_{12} = \frac{p_1+p_2}{(N-1)}. \qquad \ldots \quad (5.12)$$

$$\hat{V}_{YG}(\hat{y}_{HT}) = \frac{\pi_1\pi_2-\pi_{12}}{\pi_{12}}\left(\frac{y_1}{\pi_1}-\frac{y_2}{\pi_2}\right)^2. \qquad \ldots \quad (5.13)$$

Case (ii):

$$z_{sl} = \tfrac{1}{2}\left\{(1+p_1)\left(\frac{y_1}{p_1}\right)+(1-p_1)\left(\frac{y_2}{p_2}\right)\right\}.$$

$$\hat{V}_0(z_{sl}) = \tfrac{1}{4}(1-p_1)^2\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2.$$

$$\hat{V}_U(z_{sl}) = \tfrac{1}{4}(1-p_1)(1-p_2)\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2.$$

$$z_s = \frac{1}{2-p_1-p_2}\left\{(1-p_2)\left(\frac{y_1}{p_1}\right)+(1-p_1)\left(\frac{y_2}{p_2}\right)\right\}.$$

$$\hat{V}_U(z_s) = \frac{(1-p_1)(1-p_2)(1-p_1-p_2)}{(2-p_1-p_2)^2}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2.$$

$$z'_{sl} = \tfrac{1}{2}\left\{\left(\frac{y_1}{p_1}\right)+\frac{1}{N-1}\left(\frac{1-p_1}{p_1}\right)\left(\frac{y_2}{p_2}\right)\right\}. \qquad \ldots \quad (5.14)$$

$$\hat{V}_0(z'_{sl}) = v'_{sl} = z'^2_{sl}-\tfrac{1}{2}\left\{\left(\frac{y_1^2}{p_1}\right)+\frac{1}{N-1}\frac{(1-p_1)}{p_1}\left(\frac{y_2^2}{p_2}\right)+\frac{2(1-p_1)}{p_1}\ \frac{y_1\,y_2}{p_2}\right\}. \quad (5.15)$$

$$\hat{V}_U(z'_s) = \frac{1}{2-p_1-p_2}\left\{(1-p_2)v'_{s1}+(1-p_1)v'_{s2}\right\}. \qquad \ldots \quad (5.16)$$

$$z'_s = \frac{1}{2-p_1-p_2}\left\{(1-p_2)z'_{s1}+(1-p_1)z'_{s2}\right\}. \qquad \ldots \quad (5.17)$$

UNBIASED ESTIMATE OF ERROR VARIANCE CASE (i)

sample	$\hat{V}_0(x_{st})$	$\hat{V}_U(x_{st})$	$\hat{V}_U(x_s)$	$\hat{V}_0(x'_{st})$	$\hat{V}_U(x'_{st})$	$\hat{V}_U(x'_s)$	$\hat{V}_{BT}(\bar{y}_{BT})$	$\hat{V}_{YO}(\bar{y}_{BT})$
12	.20	1.88	1.87	45.80	18.49	14.48	−1.11	1.51
13	.81	2.37	2.30	114.20	28.93	14.85	2.27	4.33
14	6.50	3.46	3.25	241.80	45.38	14.58	6.45	7.34
21	2.72	1.88	1.87	4.84	18.49	14.48	−1.11	1.51
23	.56	.48	.44	15.64	3.44	1.62	.77	.92
24	5.76	2.16	1.94	34.20	2.63	−4.09	3.93	3.05
31	2.89	2.37	2.30	.51	28.93	14.85	2.27	4.33
32	.42	.48	.44	−4.70	3.44	1.62	.77	.92
34	3.52	2.69	2.60	−13.59	−20.01	−21.20	3.01	.72
41	2.72	3.48	3.25	−3.72	45.38	14.58	6.45	7.34
42	.30	2.16	1.94	−13.16	2.63	−4.09	3.93	3.05
43	.56	2.69	2.60	−24.82	−20.01	−21.20	3.01	.72
true error variance	2.223	2.223	2.103	9.701	9.701	0.363	2.884	2.884
variance of estimated error variance	1.0912	.8357	.7543	2583.03	491.72	194.33	4.7317	5.5115

UNBIASED ESTIMATE OF ERROR VARIANCE CASE (ii)

sample	$\hat{V}_0(x_{st})$	$\hat{V}_U(x_{st})$	$\hat{V}_U(x_s)$	$\hat{V}_0(x'_{st})$	$\hat{V}_U(x'_{st})$	$\hat{V}_U(x'_s)$	$\hat{V}_{BT}(\bar{y}_{BT})$	$\hat{V}_{YO}(\bar{y}_{BT})$
12	.20	.18	.17	93.20	49.60	42.95	−6.20	.41
13	.81	.63	.59	114.20	48.17	34.12	−4.70	1.62
14	1.82	1.22	1.08	134.60	47.92	27.39	−.58	2.79
21	.16	.18	.17	10.84	49.60	42.95	−6.20	.41
23	.16	.14	.12	11.78	2.55	1.71	−3.79	.36
24	.64	.48	.39	10.38	−3.07	−5.03	1.21	1.08
31	.49	.63	.59	−3.18	48.17	34.12	−4.70	1.62
32	.12	.14	.12	−5.62	2.55	1.71	−3.79	.36
34	.12	.10	07	−12.80	−15.07	−15.25	5.02	.18
41	.81	1.22	1.08	−9.86	47.92	27.39	−.58	2.79
42	.36	.48	.39	−13.16	−3.07	−5.03	1.21	1.08
43	.09	.10	.07	−17.01	−15.07	−15.25	5.02	.18
true error variance	0.365	0.365	0.312	5.435	5.435	1.107	0.823	0.823
variance of estimated error variance	.1575	.1236	.1004	1542.49	600.18	350.73	14.8092	6811

$$\hat{V}_U(x_i') = \hat{V}_U(x_{i_1}') - \frac{1}{2-p_1-p_2}\{(1-p_2)(x_{i_1}'-x_i')^2+(1-p_1)(x_{i_2}'-x_i')^2\}, \quad \ldots \quad (5.18)$$

$$\hat{y}_{HT} = \left(\frac{y_1}{\pi_1}\right) + \left(\frac{y_2}{\pi_2}\right) \text{ where } \pi_1 = p_1\left\{1+\sum_{j\neq 1}^{N}\left(\frac{p_j}{1-p_j}\right)\right\},$$

and

$$\pi_2 = p_2\left\{1+\sum_{j\neq 2}^{N}\left(\frac{p_j}{1-p_j}\right)\right\}. \quad \ldots \quad (5.19)$$

$$\hat{V}_{HT}(\hat{y}_{HT}) = (1-\pi_1)\left(\frac{y_1^2}{\pi_1^2}\right) + (1-\pi_2)\left(\frac{y_2^2}{\pi_2^2}\right)^2 + \frac{\pi_{12}-\pi_1\pi_2}{\pi_{12}}\left(\frac{y_1}{\pi_1}\right)\left(\frac{y_2}{\pi_2}\right),$$

where

$$\pi_{12} = \left\{\frac{p_1 p_2(2-p_1-p_2)}{(1-p_1)(1-p_2)}\right\},$$

$$\hat{V}_{YG}(\bar{y}_{HT}) = \frac{\pi_1\pi_2-\pi_{12}}{\pi_{12}}\left(\frac{y_1}{\pi_1}-\frac{y_2}{\pi_2}\right)^2 \quad \ldots \quad (5.20)$$

The results given in the above table show that for this population unordering of Das' estimators in the above two cases yields estimators which are much more efficient than the corresponding ordered estimators. Of the three unordered unbiased estimators of the population total, namely, x_s, x_s' and $\bar{y}_{HT}$, x_s' in case (i) and x_s in case (ii) have the least variance. It can also be seen that it is possible to improve substantially on the ordered variance estimators by unordering them.

The author is grateful to Prof. D. B. Lahiri for his valuable suggestions in preparing the paper.

REFERENCES

DAS, A. C. (1951): Two phase sampling and sampling with varying probabilities. *Bull. Int. Stat. Inst.*, 33, 105–112.

DES RAJ (1956): Some estimators in sampling with varying probabilities without replacement. *J. Amer. Stat. Ass.*, 51, 274, 269-284.

HORVITZ, D. G. & THOMPSON, D. J. (1952): A generalisation of sampling without replacement from a finite universe. *J. Amer. Stat. Ass.*, 47, 663–85.

LAHIRI, D. B. (1951): A method of sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.*, 33, 133–140.

MIDZUNO, H. (1952): On the sampling system with probability proportional to the sum of sizes. *Ann. Inst. Stat. Math.*, (Japan), 3, 99–108.

ROY CHOWDHURY, D. K. (1956): Sampling with varying probabilities (part of thesis submitted for the Associationship of the Indian Statistical Institute).

YATES, F. AND GRUNDY, P. M. (1953): Selection without replacement from within strata with probability proportional to size. *J. Roy. Stat. Soc.*, B, 15, 253-61.

Paper received: April 1957.