TWO-PHASE SAMPLING FOR PPS ESTIMATION

M. S. CHIKKAGOUDAR

(Received Dec. 14, 1964; revised Nov. 29, 1966)

1. Introduction

In case of sampling with probability proportional to size (pps) or some other auxiliary variable x, which is positively correlated with the variable y under study, we require the advance knowledge of the data on the variable x. If such data is not readily available, it can be obtained by using the well-known technique of two-phase sampling, without appreciably adding to the cost of the enquiry. This scheme of two-phase sampling can be effected in any one of the following four ways:

I. Sampling without replacement in both the phases
II. Sampling with replacement in the first phase and without replacement in the second phase
III. Sampling with replacement in both the phases
IV. Sampling without replacement in the first phase and with replacement in the second phase.

Des Raj [3] developed the theory for the Scheme IV and obtained the variances of the estimators (of the population mean $\overline{Y}$) in the simple forms, which are appropriate for comparison with the large sample approximations to the corresponding variances in the case of two-phase sampling for ratio and regression estimation. The same author [2] dealt with the Scheme I, but the expressions for the variances are not in so compact forms as compared to those derived in case of the Scheme IV. Following Des Raj [3], we obtain here the estimators of the population mean $\overline{Y}$, and provide the variances of the estimators in the simple forms for the first two sampling schemes. Since the corresponding theory for the Scheme III is straightforward, it is omitted. Lastly, we illustrate the utility of this theory in two-phase sampling for stratification.

2. Estimation

Scheme I:

At the first phase an initial sample of size m, in which the variable x alone is measured, is drawn without replacement with equal probabilities. Then, at the second phase a subsample of size n, in which the variable y is measured, is drawn without replacement with probability proportional to the total x-value of the subsample.

For this sampling scheme, an unbiased estimator of the population mean $\bar{Y}$ is given by

$$\hat{\bar{Y}}_1=\frac{\bar{y}_n}{\bar{x}_n}\bar{x}_m \tag{1}$$

where $\bar{y}_n$ and $\bar{x}_n$ are the sample (of size n) means of the variables y and x respectively and $\bar{x}_m$ has a similar definition.

Variance of the estimator:

We have

$$V(\hat{\bar{Y}}_1)=E_1V_2(\hat{\bar{Y}}_1)+V_1E_2(\hat{\bar{Y}}_1) \tag{2}$$

where E_2 and V_2 stand for the conditional expectation and the conditional variance for a fixed first phase sample, whereas E_1 and V_1 stand for the expectation and the variance for varying first phase sample. Let $\sum_{s_k}^{\nu}{}_c$ denote the summation over all possible samples of size k from a set of ν units subject to the restriction c on the contents of the samples. We come across here with only two forms of restriction c: no restriction and the set s_k to contain the subset s_n of $n(<k)$ units. Then, the first part on the right-hand side of (2) is seen to be equal to

$$\begin{aligned} &\frac{1}{\binom{N}{m}\binom{m}{n}}\sum_{s_m}^{N}\sum_{s_n}^{m}\frac{\bar{y}_n^2}{\bar{x}_n}\bar{x}_m-E_1(\bar{y}_m^2) \\ &=\frac{1}{m\binom{N}{n}\binom{N-n}{m-n}}\sum_{s_n}^{N}\frac{\bar{y}_n^2}{\bar{x}_n}\left[\binom{N-n-1}{m-n}n\bar{x}_n-\binom{N-n-1}{m-n-1}N\bar{X}\right] \\ &\quad-\left[\left(\frac{1}{m}-\frac{1}{N}\right)S^2+\bar{Y}^2\right] \end{aligned} \tag{3}$$

where S^2 is the population mean square of the variable y. This result follows from the eq. (4) of Sampford [5]. Now, we know that

$$\frac{1}{\binom{N}{n}}\sum_{s_n}^{N}\bar{y}_n^2=\left(\frac{1}{n}-\frac{1}{N}\right)S^2+\bar{Y}^2. \tag{4}$$

On substituting from (4) in (3) and simplifying, we get

$$E_1V_2(\hat{\bar{Y}}_1)=\frac{N(m-n)}{m(N-n)}V_p \tag{5}$$

where

$$V_p=\frac{\bar{X}}{\binom{N}{n}}\sum_{s_n}^{N}\frac{\bar{y}_n^2}{\bar{x}_n}-\bar{Y}^2 \tag{6}$$

is the variance of an estimator of $\bar{Y}$, based on a sample of size n drawn with probability proportional to x without replacement, when the population mean $\bar{X}$ is known.

The second part on the right-hand side of (2) is

$$\mathrm{V}_1\mathrm{E}_2(\hat{\bar{Y}}_1)=\mathrm{V}_1(\bar{y}_m)=\left(\frac{1}{m}-\frac{1}{N}\right)S^2. \tag{7}$$

Adding (5) and (7), we get

$$\mathrm{V}(\hat{\bar{Y}}_1)=\frac{N(m-n)}{m(N-n)}V_p+\left(\frac{1}{m}-\frac{1}{N}\right)S^2. \tag{8}$$

Estimator of the variance:

We know that

$$\hat{\mathrm{V}}(\hat{\bar{Y}}_1)=\hat{\bar{Y}}_1^2-\mathrm{Est}(\bar{Y}^2). \tag{9}$$

Following the procedure adopted by Des Raj [2], an unbiased estimator of $\bar{Y}^2$ is obtained as

$$\mathrm{Est}(\bar{Y}^2)=\frac{\bar{x}_m}{\bar{x}_n}\left[\bar{y}_n^2-\left(\frac{1}{n}-\frac{1}{N}\right)s^2\right] \tag{10}$$

where s^2 is the mean square between the n units in the second phase sample. On substituting from (10) in (9), we get

$$\hat{\mathrm{V}}(\hat{\bar{Y}}_1)=\frac{\bar{x}_m}{\bar{x}_n}\left(\frac{\bar{x}_m}{\bar{x}_n}-1\right)\bar{y}_n^2+\left(\frac{1}{n}-\frac{1}{N}\right)\frac{\bar{x}_m}{\bar{x}_n}s^2. \tag{11}$$

Extension to multistage design:

In the case of multistage sampling, let t_i be an unbiased estimator of the ith unit total y_i, based on subsampling at the second and subsequent stages. Then, an unbiased estimator of $\bar{Y}$ is

$$\hat{\bar{Y}}_1'=\frac{\bar{t}_n}{\bar{x}_n}\bar{x}_m. \tag{12}$$

As we know, the variance of this estimator is made up of two components. The between-primary component is given by (8), and the within-primary component can be easily obtained by proceeding as in the case of obtaining (5). Thus, we have

(13) $$\mathrm{V}(\hat{\bar{Y}}_1')=\mathrm{V}(\hat{\bar{Y}}_1)+\frac{1}{N-n}\left[\left(\frac{1}{m}-\frac{1}{N}\right)\sum_{i=1}^{N}\sigma_i^2+\frac{N}{n}\left(\frac{1}{n}-\frac{1}{m}\right)\frac{\bar{X}}{\binom{N}{n}}\sum_{s_n}^{N}\frac{\sum_{i=1}^{n}\sigma_i^2}{\bar{x}_n}\right]$$

where σ_i^2 is the variance of t_i for a fixed primary unit i.

An unbiased estimator of the variance (13) is easily obtained, by following Des Raj [4], as

(14) $$\hat{\mathrm{V}}(\hat{\bar{Y}}_1')=\hat{\mathrm{V}}'(\hat{\bar{Y}}_1)+\frac{1}{Nn}\frac{\bar{x}_m}{\bar{x}_n}\sum_{i=1}^{n}\hat{\sigma}_i^2$$

where $\hat{\mathrm{V}}'(\hat{\bar{Y}}_1)$ is obtained by replacing y_i by t_i in $\hat{\mathrm{V}}(\hat{\bar{Y}}_1)$, given by (11) and $\hat{\sigma}_i^2$ is an unbiased estimator σ_i^2.

Scheme II:

At the first phase an initial sample of size m, in which the variable x alone is measured, is drawn with replacement with equal probabilities. The sampling at the second phase is similar to that in the Scheme I.

For this scheme, an unbiased estimator $\hat{\bar{Y}}_2$ of the population mean $\bar{Y}$ is structurally the same as $\hat{\bar{Y}}_1$, given by (1).

The variance of $\hat{\bar{Y}}_2$ consists of two parts as in (2), and the first part is equal to

(15) $$\mathrm{E}_1\mathrm{V}_2(\hat{\bar{Y}}_2)=\mathrm{E}_1\left[\frac{\bar{x}_m}{\binom{m}{n}}\sum_{s_n}^{m}\frac{\bar{y}_n^2}{\bar{x}_n}-\bar{y}_m^2\right]$$
$$=\frac{1}{N^m\binom{m}{n}}\sum_{s_m}^{N}{}^{*}\sum_{s_n}^{m}\frac{\bar{y}_n^2}{\bar{x}_n}\bar{x}_m-\left(\frac{N-1}{Nm}S^2+\bar{Y}^2\right)$$

where $\sum_{s_n}^{N}{}^{*}$ denotes the summation over all the N^k samples of size k ($=m$ or n), drawn with replacement from the N units in the population. Let λ_i be the number of times the ith unit appears in the initial sample and α_i be the number of times the ith unit reappears in the subsample. Also, let $\sum'$ stand for the summation over all the λ_i's such that $\lambda_i\geqq 0$ and $\sum_{i=1}^{N}\lambda_i=m$, $\sum''$ for the summation over all the α_i's such that $\alpha_i\geqq 0$ and $\sum_{i=1}^{N}\alpha_i=n$, and $\sum'''$ for the summation over all the λ_i's such that $\lambda_i\geqq\alpha_i$ and $\sum_{i=1}^{N}\lambda_i=m$. In these notations, the first term on the right-hand side of (15) can be written as

(16) $$\frac{1}{mnN^m\binom{m}{n}}\sum{}'\frac{m!}{\lambda_1!\cdots\lambda_N!}\sum{}''\binom{\lambda_1}{\alpha_1}\cdots\binom{\lambda_N}{\alpha_N}\frac{\left(\sum_{i=1}^N\alpha_i y_i\right)^2}{\left(\sum_{i=1}^N\alpha_i x_i\right)}\left(\sum_{i=1}^N\lambda_i x_i\right)$$

$$=\frac{1}{mnN^m}\sum{}''\frac{n!}{\alpha_1!\cdots\alpha_N!}\frac{\left(\sum_{i=1}^N\alpha_i y_i\right)^2}{\left(\sum_{i=1}^N\alpha_i x_i\right)}\sum{}'''\frac{(m-n)!}{(\lambda_1-\alpha_1)!\cdots(\lambda_N-\alpha_N)!}\left(\sum_{i=1}^N\lambda_i x_i\right).$$

On substituting for the conditional expectation of $(\lambda_i-\alpha_i)$ for a fixed α_i and noting that

(17) $$\frac{1}{N^n}\sum_{s_n}^{N}{}^*\bar{y}_n^2=\frac{N-1}{Nn}S^2+\bar{Y}^2$$

the quantity on the right-hand side of (16) reduces to

(18) $$\frac{m-n}{m}V_p'+\frac{N-1}{Nm}S^2+\bar{Y}^2$$

where

(19) $$V_p'=\frac{\bar{X}}{N^n}\sum_{s_n}^{N}{}^*\frac{\bar{y}_n^2}{\bar{x}_n}-\bar{Y}^2.$$

From (15) and (18), we have

(20) $$\mathrm{E}_1\mathrm{V}_2(\hat{\bar{Y}}_2)=\frac{m-n}{m}V_p'.$$

The second part of $\mathrm{V}(\hat{\bar{Y}}_2)$ is equal to

(21) $$\mathrm{V}_1\mathrm{E}_2(\hat{\bar{Y}}_2)=\mathrm{V}_1(\bar{y}_m)=\frac{N-1}{Nm}S^2.$$

Hence, adding (20) and (21), we get

(22) $$\mathrm{V}(\hat{\bar{Y}}_2)=\frac{m-n}{m}V_p'+\frac{N-1}{Nm}S^2.$$

Proceeding as in the case of $\hat{\mathrm{V}}(\hat{\bar{Y}}_1)$, an unbiased estimator of the variance (22) is obtained as

(23) $$\hat{\mathrm{V}}(\hat{\bar{Y}}_2)=\frac{\bar{x}_m}{\bar{x}_n}\left(\frac{\bar{x}_m}{\bar{x}_n}-1\right)\bar{y}_n^2+\frac{\bar{x}_m}{n\bar{x}_n}s^2.$$

Extension to multistage design:

In this case, an unbiased estimator $\hat{\bar{Y}}_2'$, of the population mean, is structurally the same as $\hat{\bar{Y}}_1'$, given by (12). But the variance of $\hat{\bar{Y}}_2'$ is found to be

$$\text{(24)} \qquad V(\hat{\hat{Y}}_2')=V(\hat{\hat{Y}}_2)+\left[\frac{m-n}{mn^2}\frac{\bar{X}}{N^n}\sum_{s_n}^{N}{}^{*}\frac{\sum_{i=1}^{n}\sigma_i^2}{\bar{x}_n}+\frac{1}{Nm}\sum_{i=1}^{N}\sigma_i^2\right].$$

The second part on the right-hand side of (24) can be easily obtained by following the method used in getting (18). An unbiased estimator of this variance is

$$\text{(25)} \qquad \hat{V}(\hat{\hat{Y}}_2')=\hat{V}'(\hat{\hat{Y}}_2)$$

where $\hat{V}'(\hat{\hat{Y}}_2)$ is obtained by replacing y_i by t_i in $\hat{V}(\hat{\hat{Y}}_2)$, given by (23).

When sampling in both the phases is with equal probabilities, it is found that the expected number of distinct units in the second phase sample of Scheme II is greater than the corresponding number in the second phase sample of Scheme IV (section 1). This suggests that Scheme II is better than Scheme IV in this case.

3. Two-phase sampling for stratification

In this procedure, the initial sample of size m, in which the variable x alone is measured, is classified into different strata according to the values of x. If m_i is the number of units falling in the ith stratum ($i=1, 2, \cdots, h$), a sample of n_i units, in which the variable y also is measured, is drawn from m_i units, in the usual practice, with equal probabilities such that $\sum_{i=1}^{h} n_i=n$. Here we study the sampling procedure, where the information on the variable x is used not only for stratification of the units, but also for selecting the samples from different strata with probability proportional to x. We consider the two cases where (a) the n_i units are drawn without replacement with probability proportional to their total x-value and (b) the n_i units are drawn with replacement with probability proportional to x, while the initial samples of m units are drawn, in both the cases, without replacement with equal probabilities.

Case (a):

Let N_i be the number of units in the ith stratum and $P_i=N_i/N$ be the proportion of units falling in the ith stratum. An unbiased estimator of P_i is given by the corresponding sample proportion $p_i=m_i/m$.

In this case, an unbiased estimator of the population mean $\bar{Y}=\sum_{i=1}^{h} P_i\bar{Y}_i$ is given by

(26) $$\hat{\hat{Y}}=\sum_{i=1}^{h} p_i \hat{\hat{Y}}_i=\sum_{i=1}^{h} p_i\left(\frac{\bar{y}_{n_i}}{\bar{x}_{n_i}}\bar{x}_{m_i}\right).$$

To obtain the variance of this estimator, we write

(27) $$V(\hat{\hat{Y}})=E[V(\hat{\hat{Y}}|m_1,\cdots,m_h)]+V[E(\hat{\hat{Y}}|m_1,\cdots,m_h)].$$

By using the variance-formula (8), the first part on the right-hand side of (27) is seen to be equal to

(28) $$E\left[\sum_{i=1}^{h} p_i^2 V(\hat{\hat{Y}}_i|m_i)\right]=E\left[\sum_{i=1}^{h} p_i^2\left\{\frac{N_i(m_i-n_i)}{m_i(N_i-n_i)}V_{pi}+\left(\frac{1}{m_i}-\frac{1}{N_i}\right)S_i^2\right\}\right]$$
$$=\frac{1}{m^2}\sum_{i=1}^{h}\left[\frac{N_i}{N_i-n_i}\{E(m_i^2)-n_i E(m_i)\}V_{pi}+\left\{E(m_i)-\frac{E(m_i^2)}{N_i}\right\}S_i^2\right]$$

where V_{pi} is obtained by affixing the subscript i to n, N, $\bar{X}$ and $\bar{Y}$ in V_p, given by (6) and S_i^2 is the mean square of the ith stratum. Since m_i is the number of units falling in the ith stratum (having N_i units), when m units are drawn from a total of N units in the population, m_i has a hypergeometric distribution. Hence, on substituting for $E(m_i)$ and $E(m_i^2)$ of this hypergeometric distribution in the right-hand side of (28), we get

(29) $$E[V(\hat{\hat{Y}}|m_1,\cdots,m_h)]$$
$$=\sum_{i=1}^{h}\left[\frac{N_i n_i}{N_i-n_i}\left\{\frac{1}{n_i}\left(\frac{N-m}{m(N-1)}P_iQ_i+P_i^2\right)-\frac{P_i}{m}\right\}V_{pi}+\left\{\frac{P_i}{m}-\frac{1}{N_i}\left(\frac{N-m}{m(N-1)}P_iQ_i+P_i^2\right)\right\}S_i^2\right]$$

where $Q_i=1-P_i$. The second part on the right-hand side of (27) is equal to

(30) $$V\left[\sum_{i=1}^{h} p_i\bar{Y}_i\right]=\sum_{i=1}^{h}\bar{Y}_i^2 V(p_i)+\sum_{i\neq i'}^{h}\bar{Y}_i\bar{Y}_{i'}\operatorname{Cov}(p_i,p_{i'}).$$

On substituting for $V(p_i)$ and $\operatorname{Cov}(p_i, p_{i'})$, we get

(31) $$V[E(\hat{\hat{Y}}|m_1,\cdots,m_h)]=\frac{N-m}{m(N-1)}\sum_{i=1}^{h}P_i(\bar{Y}_i-\bar{Y})^2.$$

Substituting from (29) and (31) into (27), we get the variance of the estimator. By adding and subtracting the quantity

(32) $$\left[\frac{N-m}{m(N-1)}P_iQ_i+P_i^2\right]\frac{S_i^2}{n_i}$$

this variance can be rewritten as

$$(33)\quad V(\hat{\hat{Y}})=\sum_{i=1}^{h} P_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2$$
$$+\frac{N-m}{m(N-1)}\left[\sum_{i=1}^{h} P_i Q_i\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2+\sum_{i=1}^{h} P_i(\bar{Y}_i-\bar{Y})^2\right]$$
$$-\frac{N}{m}\sum_{i=1}^{h} P_i^2\frac{N_i-1}{N_i-n_i}\left(\frac{m-1}{N-1}-\frac{n_i-1}{N_i-1}\right)\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2-V_{pi}\right].$$

The first two terms on the right-hand side constitute the variance of the estimator when sampling at both the phases is with equal probabilities (Ref. Sukhatme [6], p. 115). Since the sampling with probability proportional to x is generally more efficient than the sampling with equal probabilities, we assume that

$$(34)\qquad \left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2\geqq V_{pi}.$$

Hence, one of the conditions for the proposed sampling method to be more efficient than the usual method with equal selection probabilities is

$$(35)\qquad \max\left(\frac{n_i-1}{N_i-1}\right)<\frac{m-1}{N-1}$$

or, approximately

$$(36)\qquad \max\left(\frac{n_i}{N_i}\right)<\frac{m}{N}.$$

If the unknown N_i is replaced by its estimator Np_i, the condition (36) reduces to

$$(37)\qquad \max\left(\frac{n_i}{m_i}\right)<1$$

which is the fundamental assumption under which the above theory (for case (a)) of two-phase sampling for stratification holds good. Under the condition (35) we see that the effect of selecting the samples from different strata with probability proportional to x, is to reduce the variance by the quantity

$$(38)\qquad \frac{N}{m}\sum_{i=1}^{h} P_i^2\frac{N_i-1}{N_i-n_i}\left(\frac{m-1}{N-1}-\frac{n_i-1}{N_i-1}\right)\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2-V_{pi}\right].$$

In the case of sampling with equal probabilities, we have

$$\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2=V_{pi}$$

and hence the quantity (38) reduces to zero.

To find an unbiased estimator of the variance (33), we write

$$\text{(39)} \qquad \mathrm{V}(\hat{\hat{Y}})=\frac{N-m}{m(N-1)}\sum_{i=1}^{h}P_i(\bar{Y}_i-\bar{Y})^2+\mathrm{E}\Big[\sum_{i=1}^{h}p_i^2\mathrm{V}(\hat{\hat{Y}}_i|m_i)\Big].$$

Hence,

$$\text{(40)} \qquad \hat{\mathrm{V}}(\hat{\hat{Y}})=\frac{N-m}{m(N-1)}\,\mathrm{Est}\Big[\sum_{i=1}^{h}P_i(\bar{Y}_i-\bar{Y})^2\Big]+\sum_{i=1}^{h}p_i^2\hat{\mathrm{V}}(\hat{\hat{Y}}_i|m_i).$$

The estimator $\hat{\mathrm{V}}(\hat{\hat{Y}}_i|m_i)$ is obtained from (11) by affixing the subscript i to n, m, N and s^2. To obtain the estimator in the first term on the right-hand side of (40), we consider the quantity defined by

$$\text{(41)} \qquad s_b^2=\sum_{i=1}^{h}p_i(\hat{\hat{Y}}_i-\hat{\hat{Y}})^2=\sum_{i=1}^{h}p_i\hat{\hat{Y}}_i^2-\hat{\hat{Y}}^2.$$

Taking the expectations, we get

$$\text{(42)} \qquad \mathrm{E}(s_b^2)=\mathrm{E}\Big[\sum_{i=1}^{h}p_i\{\mathrm{V}(\hat{\hat{Y}}_i|m_i)+\bar{Y}_i^2\}\Big]-[\mathrm{V}(\hat{\hat{Y}})+\bar{Y}^2]$$
$$=\frac{N(m-1)}{m(N-1)}\sum_{i=1}^{h}P_i(\bar{Y}_i-\bar{Y})^2+\mathrm{E}\Big[\sum_{i=1}^{h}p_iq_i\mathrm{V}(\hat{\hat{Y}}_i|m_i)\Big]$$

where $q_i=1-p_i$. Hence, we have

$$\text{(43)} \qquad \mathrm{Est}\Big[\sum_{i=1}^{h}P_i(\bar{Y}_i-\bar{Y})^2\Big]=\frac{m(N-1)}{N(m-1)}\Big[\sum_{i=1}^{h}p_i(\hat{\hat{Y}}_i-\hat{\hat{Y}})^2-\sum_{i=1}^{h}p_iq_i\hat{\mathrm{V}}(\hat{\hat{Y}}_i|m_i)\Big].$$

On substituting from (43) in (40), we get

$$\text{(44)} \qquad \hat{\mathrm{V}}(\hat{\hat{Y}})=\frac{N-m}{m(N-1)}\Big[\sum_{i=1}^{h}p_i(\hat{\hat{Y}}_i-\hat{\hat{Y}})^2-\sum_{i=1}^{h}p_i\hat{\mathrm{V}}(\hat{\hat{Y}}_i|m_i)\Big]$$
$$+\frac{N(m-1)}{m(N-1)}\sum_{i=1}^{h}p_i^2\hat{\mathrm{V}}(\hat{\hat{Y}}_i|m_i).$$

In the case of sampling with equal probabilities, the variance estimator (44) reduces to the eq. (12.12) of Cochran ([1], p. 274).

Case (b):

In this case, an unbiased estimator of $\bar{Y}$ is

$$\text{(45)} \qquad \hat{\hat{Y}}'=\sum_{i=1}^{h}p_i\hat{\hat{Y}}_i'=\sum_{i=1}^{h}p_i\Big(\frac{\bar{x}_{m_i}}{n_i}\sum_{j=1}^{n_i}\frac{y_{ij}}{x_{ij}}\Big).$$

To find the variance of this estimator, we have

(46) $$V(\hat{\bar{Y}}')=E[V(\hat{\bar{Y}}'|m_1,\cdots,m_h)]+V[E(\hat{\bar{Y}}'|m_1,\cdots,m_h)] \\ =E\left[\sum_{i=1}^{h}p_i^2V(\hat{\bar{Y}}_i'|m_i)\right]+V\left[\sum_{i=1}^{h}p_i\bar{Y}_i\right].$$

The second part on the right-hand side of (46) remains the same as in the case (a) and is given by (31). Considering the second part, $V(\hat{\bar{Y}}_i'|m_i)$ is obtained by Des Raj [3] as

(47) $$V(\hat{\bar{Y}}_i'|m_i)=\frac{N_i(m_i-1)}{m_i(N_i-1)}\frac{V_{pi}''}{n_i}+\left(\frac{1}{m_i}-\frac{1}{N_i}\right)S_i^2$$

where

(48) $$V_{pi}''=\frac{\bar{X}_i}{N_i}\sum_{j=1}^{N_i}\frac{y_{ij}^2}{x_{ij}}-\bar{Y}_i^2.$$

Substituting from (47) in the first part on the right-hand side of (46), it becomes

(49) $$E[V(\hat{\bar{Y}}'|m_1,\cdots,m_h)] \\ =\frac{1}{m^2}\sum_{i=1}^{h}\left[\frac{N_i}{N_i-n_i}\{E(m_i^2)-E(m_i)\}\frac{V_{pi}''}{n_i}+\left\{E(m_i)-\frac{E(m_i^2)}{N_i}\right\}S_i^2\right] \\ =\sum_{i=1}^{h}\left[\frac{N_i}{N_i-1}\left\{\frac{N-m}{m(N-1)}P_iQ_i+P_i^2-\frac{P_i}{m}\right\}\frac{V_{pi}''}{n_i}\right. \\ \left.+\left\{\frac{P_i}{m}-\frac{1}{N_i}\left(\frac{N-m}{m(N-1)}P_iQ_i+P_i^2\right)\right\}S_i^2\right].$$

Substituting from (31) and (49) in (46), we get the variance of the estimator. By adding and subtracting the quantity

(50) $$\left[\frac{P_i}{m}-\frac{N-m}{m(N-1)}P_iQ_i-P_i^2\right]\frac{S_i^2}{n_i}$$

this variance can be rewritten as

(51) $$V(\hat{\bar{y}}')=\sum_{i=1}^{h}\frac{P_i^2}{n_i}\left(\frac{N_i-1}{N_i}S_i^2\right) \\ +\frac{N-m}{m(N-1)}\left[\sum_{i=1}^{h}P_i(\bar{Y}_i-\bar{Y})^2+\sum_{i=1}^{h}P_iQ_i\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2\right. \\ \left.+\frac{N-1}{N}\sum_{i=1}^{h}P_i\frac{(n_i-1)}{n_i}S_i^2\right]-\frac{N(m-1)}{m(N-1)}\sum_{i=1}^{h}\frac{P_i^2}{n_i}\left[\frac{N_i-1}{N_i}S_i^2-V_{pi}''\right].$$

It is found that the first two terms on the right-hand side of (51) form the variance of the estimator, when sampling at both the phases is with equal probabilities and at the second phase selection is made with replacement. In this case also, we assume that

$$\frac{N_i-1}{N_i}S_i^2 \geqq V_{pi}''. \tag{52}$$

Then, the effect of adopting the sampling with probability proportional to x is to reduce the variance by the quantity

$$\frac{N(m-1)}{m(N-1)}\sum_{i=1}^{h}\frac{P_i^2}{n_i}\left(\frac{N_i-1}{N_i}S_i^2 - V_{pi}''\right). \tag{53}$$

In the case of sampling with equal probabilities

$$V_{pi}''=\frac{N_i-1}{N_i}S_i^2$$

and hence, the quantity (53) reduces to zero. It may be noted that in this case there is no any rigid condition for the proposed sampling scheme (case (b)) to be more efficient than the corresponding equal probability sampling, except the one given by (52), which is usually satisfied.

Following the procedure used in the previous case, an unbiased estimator of the variance (51) is obtained as

$$\hat{V}(\hat{\bar{y}}')=\frac{N-m}{N(m-1)}\left[\sum_{i=1}^{h}p_i(\hat{\bar{Y}}_i'-\hat{\bar{Y}}')^2-\sum_{i=1}^{h}p_i\hat{V}(\hat{\bar{Y}}_i'\,|\,m_i)\right] + \frac{m(N-1)}{N(m-1)}\sum_{i=1}^{h}p_i^2\hat{V}(\hat{\bar{Y}}_i'\,|\,m_i) \tag{54}$$

where $\hat{V}(\hat{\bar{Y}}_i'\,|\,m_i)$ is given by Des Raj [3] as

$$\hat{V}(\hat{\bar{Y}}_i'\,|\,m_i)=\frac{\bar{x}_{m_i}^2}{n_i(n_i-1)}\left[\sum_{j=1}^{n_i}\frac{y_{ij}^2}{x_{ij}}-\frac{1}{n_i}\left(\sum_{j=1}^{n_i}\frac{y_{ij}}{x_{ij}}\right)^2\right] + \frac{N_i-m_i}{N_in_i(m_i-1)}\left[\bar{x}_{m_i}\sum_{j=1}^{n_i}\frac{y_{ij}^2}{x_{ij}}-\frac{\bar{x}_{m_i}^2}{n_i-1}\left\{\left(\sum_{j=1}^{n_i}\frac{y_{ij}}{x_{ij}}\right)^2-\sum_{j=1}^{n_i}\frac{y_{ij}^2}{x_{ij}^2}\right\}\right]. \tag{55}$$

Acknowledgements

The author is grateful to Dr. M. V. Jambunathan for his guidance and to the Government of India for the award of Research Training Scholarship.

KARNATAK UNIVERSITY, DHARWAR, INDIA

REFERENCES

[1] W. G. Cochran, *Sampling Techniques*, Asia Publishing House, 1962.
[2] Des Raj, "Ratio estimation in sampling with equal and unequal probabilities," *J. Indian Soc. Agric. Statist.*, 7 (1954), 127-138.

[3] Des Raj, "On double sampling for pps estimation," *Ann. Math. Statist.*, 35 (1964), 900-902.

[4] Des Raj, "Some remarks on a simple procedure of sampling without replacement," *J. Amer. Statist. Ass.*, 61 (1966), 391-396.

[5] M. R. Sampford, "Methods of cluster sampling with and without replacement for clusters of unequal sizes," *Biometrika*, 49 (1962), 27-40.

[6] P. V. Sukhatme, *Sampling Theory of Surveys with Applications*, Iowa State College Press, Ames, Iowa, 1953.