RATIO ESTIMATION IN SAMPLING WITH EQUAL AND UNEQUAL PROBABILITIES

By Des Raj

Indian Statistical Institute, Calcutta

1. Introduction

In sampling designs use of supplementary information through ratio estimation is known to bring about considerable reduction in the sampling variance of the estimates over those obtained without using such information. But the difficulty involved in the use of these estimates is that they are subject to an unknown bias and no exact result for the bias or sampling variance of the estimates is available. Current treatment of this topic brings in certain approximations the justification for which still remains unestablished, thereby reducing the value for practical application of the techniques involved. For example, Cochran (1953) assumes that the sample mean $\bar{x}$ is close to the population mean $\bar{X}$ so that to this degree of approximation the ratio estimate is assumed to be unbiased for calculating an approximate expression for its variance. There appears to be no rigorous basis on which certain terms are neglected, especially when the sample size cannot be made indefinitely large. As pointed out by Cochran, the results are believed to be true for large samples but it is not known when the sample size should be considered large enough. Sukhatme (1954) assumes that the contribution of terms involving powers, in the deviation of the sample mean from the population mean, higher than the second is negligible and obtains a first approximation to the bias and variance of the estimate and also obtains improved approximations by retaining more terms. Koop (1951) found by an ingenious device an approximate expression for the bias of the ratio estimate containing moments and product moments no higher than the fourth degree. He also obtained the same expression for the bias by using a simpler method but nevertheless making an additional assumption that the sample mean (for the auxiliary character) is smaller than twice the corresponding population mean; which shows that the additional assumption made is not necessary.

If the entire population is divided up into a large number of strata and the sample size within each stratum is small (as is usually the case) one feels diffident about using the ratio estimate within strata as it is based on a small sample and therefore there is a risk of introducing serious bias in the estimates. Leaving aside the question of bias which may possibly be negligible in large samples, no unbiased method of estimation of the sampling variance is available in literature. There are however certain methods which furnish approximate estimates of the sampling variance and these methods are currently being used for the estimation of error.

This situation led workers in the field to explore the possibilities of modifying the sampling scheme so that the estimate (retaining the character of a ratio estimate) becomes unbiased. Lahiri (1951) gave such a sampling scheme and also presented a convenient procedure for actually drawing the sample. Earlier Midzuno (1950) had given a similar sampling scheme in which the probability of sampling a certain number of units is proportionate to the sum of their sizes; incidentally in this scheme an unbiased estimate of a particular general form turned out to be usual ratio estimate. Horvitz and Thompson (1952) state that in Midzuno's scheme an unbiased estimate of the sampling variance of the estimate cannot be obtained from the sample elements except in the trivial case of equal probability for each sample combination. One of the objects of the present paper is to obtain such estimates appropriate to the different sampling designs. For small sample use, when the usual ratio estimate may not be used on account of the serious bias involved, the estimates discussed in this paper may be of particular interest. Before considering these problems we shall make certain observations on the usual (biased) ratio estimate.

2. Some Further Remarks on the Biased Ratio Estimate

There are two points about the usual ratio estimate which require some clarification. It is generally stated that it is the amount of correlation between y and x which determines whether the ratio estimate is superior to the simple average. One may get the impression that if there is perfect correlation between y and x, the ratio estimate would be always superior. The fact is that even in case of linear regression the issue does not depend entirely on the correlation coefficient. For example, suppose that

$$y = \alpha + \beta x \qquad (1)$$

so that there is perfect correlation between y and x. Then

$$V\left(\frac{\bar{y}}{\bar{x}}\right) = \alpha^2 V\left(\frac{1}{\bar{x}}\right) \qquad (2)$$

so that the ratio estimate is superior to the simple average if

$$\frac{Nn}{N-n}\ \frac{X^2}{\sigma_x^2}\ V\left(\frac{1}{\bar{x}}\right) < \frac{\beta^2}{\alpha^2} \qquad (3)$$

where

N = size of the finite population,

n = sample size,

X = total of the population for the auxiliary variate x,

σ_x^2 = variance of the population for the auxiliary variate.

Now for a given population of x's, the left-hand side of (3) is fixed and α, β can be suitably chosen (say by taking α sufficiently large) so that inequality (3) does not hold. Thus although there is perfect correlation between y and x the ratio estimate may be worse than the simple average. One can set about finding an exact condition under which the ratio estimate may be superior to the simple average. If the x's be assumed to be non-random variables while the y's are random variables (not an unrealistic assumption) and that for a given x we have

$$E(y) = \alpha + \beta x, \quad V(y) = \sigma^2 = \sigma_y^2 (1 - \rho^2) \qquad (4)$$

it can be shown that the ratio estimate is superior if

$$\alpha^2 V\left(\frac{1}{\bar{x}}\right) + \frac{\sigma_y^2 (1 - \rho^2)}{n} E\left(\frac{1}{\bar{x}^2}\right) < \frac{N-n}{Nn} \; \frac{\sigma_y^2}{X^2}. \qquad (5)$$

In case the regression line passes through the origin, the criterion is

$$\rho^2 > 1 - \frac{1}{X^2 E\left(\frac{1}{\bar{x}^2}\right)}. \qquad (6)$$

The application of these criteria is rather difficult in practice. The second point to be noted is that it is sometimes stated that for infinite populations when the regression of y on x passes through the origin and the variance of y for a given x is proportional to x, the ratio estimate is the best linear unbiased estimate. The proof of this result is made to depend on an application of Markoff theorem. In the opinion of the present author the application of Markoff theorem is not strictly justified in this case since for the i-th unit in the sample x_i (the auxiliary variate) like y_i (the variate under study) is generally a random variate and not a known constant as is assumed in the Markoff set-up.

3. Unbiased Ratio Estimate

Suppose the object is to estimate the aggregate Y of a variate y for a finite population of N units when we possess in advance information on another variate x for the different units in the population. If we select a sample s_n with probability proportional to

$\left(\sum_{1}^{n} x_i\right)_{s_n}$, the proposed ratio estimate is

$$\hat{y}_{urs} = \frac{(\Sigma\, y_i)_{sn}}{(\Sigma\, x_i)_{s_n}}\, X \tag{7}$$

where '*urs*' stands for unbiased ratio estimate for single sampling designs. So far as the actual sampling procedure for obtaining s_n is concerned, one straightforward method would consist in preparing a list of all possible samples of size n drawn from the entire population, determining the total size $\left(\sum_{1}^{n} x_i\right)$ of each of the possible samples and selecting one sample with probability proportional to the total size determined before. This is exactly equivalent to selecting a unit with probability proportionate to its size where the totality of units is the totality of all possible samples of size n drawn from a population of size N. But the labour involved in this procedure can be avoided to a great extent by following an equivalent rule given by Lahiri (1951):

(i) Take a random sample s_n of n observations (without replacement with equal probability) and find its aggregate size $(\Sigma\, x_i)_{s_n}$, (ii) choose a random number between zero and the sum of the n largest units (or any number greater than it), (iii) if the chosen random number exceeds the aggregate size $(\Sigma\, x_i)_{s_n}$ of s_n reject the sample and replace it, otherwise accept, (iv) if rejected, repeat the operation until a selection is made.

It is easy to see that the determination of the total size is not needed for all possible samples but only for the samples under trial. Another simple procedure for drawing the sample s_n in this case would consist in drawing the first unit in the sample with probabilities proportionate to size (of the auxiliary variate) and the remaining $n - 1$ units with equal probability without replacement.

4. Some Properties of the Unbiased Ratio Estimate

We shall now briefly study some of the properties of the proposed ratio estimate. It is easy to see that $\hat{y}_{urs}$ is unbiased for estimating Y. Regarding the variance of the estimate we have

$$V(\hat{y}_{urs}) = \frac{X}{{}^{N-1}C_{n-1}} \Sigma' \frac{\{(\Sigma\, y_i)_{sn}\}^2}{(\Sigma\, x_i)_{s_n}} - Y^2 \tag{8}$$

where Σ' denotes summation over all possible samples s_x. To obtain an unbiased estimate of the sampling variance, we give the following lemma:

Lemma 1.—An unbiased estimate of Y^2 is provided by

$$G_1 = \frac{1}{P(s_n)} \left[\frac{(\Sigma y_i^2)_{sn}}{{}^{N-1}C_{n-1}} + 2 \frac{\left(\sum_{j>i} y_i y_j\right)_{sn}}{{}^{N-2}C_{n-2}} \right] \tag{9}$$

where

$$P(s_n) = \frac{(\Sigma x_i)_{sn}}{X\, {}^{N-1}C_{n-1}}$$

is the probability with which s_n will be selected. The proof of the lemma follows from the observation that

$$E(G_1) = \Sigma' \left[\frac{(\Sigma y_i^2)_{sn}}{{}^{N-1}C_{n-1}} + 2 \frac{\left(\sum_{j>i} y_i y_j\right)_{sn}}{{}^{N-2}C_{n-2}} \right] \tag{10}$$

and

$$\Sigma' (\Sigma y_i^2)_{sn} = {}^{N-1}C_{n-1} \sum_1^N Y_i^2, \tag{11}$$

$$\Sigma' \left(\sum_{j>i} y_i y_j \right)_{sn} = {}^{N-2}C_{n-2} \sum_{j>i} Y_i Y_j.$$

It follows from Lemma 1 that an unbiased estimate of $V(\hat{y}_{urs})$ is provided by

$$\hat{V}(\hat{y}_{urs}) = \hat{y}_{urs}^2 - \frac{X\, {}^{N-1}C_{n-1}}{(\Sigma x_i)_{sn}} \left[\frac{(\Sigma y_i^2)_{sn}}{{}^{N-1}C_{n-1}} + 2 \frac{\left(\sum_{j>i} y_i y_j\right)_{sn}}{{}^{N-2}C_{n-2}} \right]. \tag{12}$$

It may be noted that the estimator (12) may take on negative values for certain samples. This difficulty of the variance estimates becoming negative in sampling with unequal probabilities has been encountered by Sen (1952) and Yates and Grundy (1953). To study the broad class of estimates of which the estimate (7) is a member, we state the following lemma without proof:

Lemma 2.—If from a finite population of size N, a sample s_n of size n be selected with probability $P(s_n)$, the only unbiased estimate of the aggregate of the population of the form

$$\lambda_{sn} (\Sigma y_i)_{sn} \tag{13}$$

is

$$\hat{y} = \frac{1}{{}^{N-1}C_{n-1}} \frac{(\Sigma y_i)_{sn}}{P(s_n)}, \tag{14}$$

the variance of the estimate being given by

$$V(\hat{y}) = \left(\frac{1}{{}^{N-1}C_{n-1}} \right)^2 \Sigma' \frac{\{(\Sigma y_i)_{sn}\}^2}{P(s_n)} - Y^2. \tag{15}$$

Further

$$V(\hat{y}) = 0$$

when

$$P(s_n) = \frac{(\Sigma y_i)_{s_n}}{{}^{N-1}C_{n-1} Y} \cdot \tag{16}$$

We thus find that, if the probability with which a sample s_n is selected be proportional to $(\Sigma y_i)_{s_n}$, the variance of the estimate would be zero. This result, however, is not of practical interest since if the y's be known in advance, the sample would be unnecessary. The result suggests, however, that if the items are relatively stable through time, the most recently available previous values of the y's may be the best measures of size to adopt. Thus, if the y's are reasonably correlated with x, values of which on different units are already available, and if the regression of y on x passes through the origin, the proposed estimate, $\hat{y}_{urs}$ obtained by substituting

$$P(s_n) = \frac{(\Sigma x_i)_{s_n}}{{}^{N-1}C_{n-1} X} \tag{17}$$

in (14) may be considerably more efficient than the comparable estimate based on the simple arithmetic mean of the y's, which does not take advantage of the correlation between y and x. As an illustration, we consider the population of twenty blocks in Ames, Iowa, given by Horvitz and Thompson (1952). It is required to estimate the total number of households in the portion of the city represented by these twenty blocks by taking a sample of two blocks. Auxiliary information regarding the eye-estimated number of households in the different blocks is known. Since the true regression passes through the origin in this case and the correlation coefficient ($\rho^2 = \cdot 75$) is sufficiently high, the estimator $\hat{y}_{urs}$ should be expected to be of considerable efficiency. In fact, the variance of $\hat{y}_{urs}$ is 3,579 while the variance of the simple average is 16,219. The usual ratio estimate is not comparable, since in samples of size two it will be subject to a serious bias.

5. Applications to Stratified Sampling

In case the design used is stratified sampling, one method of estimating the population total is to make independent estimates of the total for each stratum using the unbiased ratio estimate $\hat{y}_{urs}$ and add these estimates. The variance of the estimate will be obtained by adding up individual variances given by (8) over the different strata and an unbiased estimate of the variance of the final estimate will be obtained by adding up the estimated variances given by (12) over the different strata. Another method would be to draw the sample over the whole set of strata such that the probability that s_n be selected be proportional to

$$\left(\sum_1^k N_i \bar{x}_i\right)_{s_n} \tag{18}$$

where $\bar{x}_i$ is the sample mean based on sample size n_i in the i-th stratum of size N_i and k is the number of strata. An unbiased ratio estimate is then provided by

$$\hat{y}_{st} = X \frac{\left(\sum_1^k N_i \bar{y}_i\right)_{s_n}}{\left(\sum_1^k N_i \bar{x}_i\right)_{s_n}}. \tag{19}$$

It is obvious that in this case

$$P(s_n) = \frac{\left(\sum_1^k N_i \bar{x}_i\right)_{s_n}}{[XN']} \tag{20}$$

where

$$N' = {}^{N_1}C_{n_1} \; {}^{N_2}C_{n_2} \ldots {}^{N_k}C_{n_k}. \tag{21}$$

The variance of $\hat{y}_{st}$ is given by

$$V(\hat{y}_{st}) = \frac{X}{N'} \Sigma' \frac{\left\{\left(\sum_1^k N_i \bar{y}_i\right)_{s_n}\right\}^2}{\left(\sum_1^k N_i \bar{x}_i\right)_{s_n}} - Y^2 \tag{22}$$

and an unbiased estimate of $V(\hat{y}_{st})$ is provided by

$$\hat{V}(\hat{y}_{st}) = \hat{y}_{st}^2 - \frac{X}{\left(\sum_1^k N_i \bar{x}_i\right)_{s_n}} \left[\sum_{i=1}^{k} \frac{N_i}{n_i} \sum_{l=1}^{ni} y_{il}^2 + 2 \sum_{i=1}^{k} \sum_{m>l=1}^{ni} \frac{N_i}{n_i} \frac{N_i - 1}{n_i - 1} y_{il} y_{im} + 2 \sum_{j>i=1}^{k} \sum_{l=1}^{n_i} \sum_{m=1}^{n_j} \frac{N_i}{n_i} \frac{N_j}{n_j} y_{il} y_{jm}\right]. \tag{23}$$

6. Estimation of Proportions and Ratios

Suppose it is required to estimate the total number (or proportion) of individuals in a population belonging to a class C. Let the population consist of N clusters, the i-th cluster containing x_i elements out of which y_i elements belong to C. If a sample of n clusters is taken with probability proportional to $(\Sigma x_i)_{s_n}$, an unbiased estimate of the total number of elements belonging to C is provided by the ratio estimate $\hat{y}_{urs}$. Similar remarks apply if the population has been stratified.

If it be desired to estimate the ratio $R = Y/X$ where information on x is not initially available but has to be ascertained by enquiry, we may either take n observations (with replacement) with probabilities proportionate to the x's or take a sample s_n with probability proportional to $(\Sigma x_i)_{s_n}$. In the former case, an unbiased estimate of R is given by

$$\hat{y}_{pps} = \frac{1}{n} \sum_{1}^{n} \frac{y_i}{x_i} \tag{24}$$

whose variance is

$$V(\hat{y}_{pps}) = \frac{1}{n} \sum_{i=1}^{N} \frac{x_i}{X} \left(\frac{y_i}{x_i} - R \right)^2 \tag{25}$$

and estimated variance is

$$\hat{V}(\hat{y}_{pps}) = \frac{1}{n(n-1)} \sum_{1}^{n} \left(\frac{y_i}{x_i} - \hat{y}_{pps} \right)^2. \tag{26}$$

In the latter case the unbiased estimate is

$$\frac{(\Sigma y_i)_{sn}}{(\Sigma x_i)_{sn}}$$

whose variance and estimated variance have already been studied.

7. Extension to Multistage Designs

We have so far considered applications of the unbiased ratio estimate to unistage designs. Since, however, large-scale surveys are generally based on multistage designs it would be appropriate here to extend the unbiased ratio estimate to such designs. Suppose the population consists of N first stage units out of which n first stage units are selected so that the probability that a sample s_n of the units is selected be proportional to $(\Sigma x_i)_{s_n}$. The variate x refers to some known measure of size of the units. Suppose further that for the i-th first stage unit in the population there is an estimator T_i (based on sampling at the second and subsequent stages) of the total y_i of that unit such that

$$E(T_i) = y_i, \; V(T_i) = \sigma_i^2 = E(s_i^2). \tag{27}$$

For estimating the population total Y, we use the estimator

$$\hat{y}_{urm} = X \frac{(\Sigma\, T_i)_{s_n}}{(\Sigma\, x_i)_{s_n}} \tag{28}$$

where '*urm*' stands for unbiased ratio estimate in multistage designs. We have for given s_n,

$$E\,(\hat{y}_{urm}) = X \frac{(\Sigma\, y_i)_{s_n}}{(\Sigma\, x_i)_{s_n}}$$

so that (28) is unbiased for estimating Y. The variance of $\hat{y}_{urm}$ is given by

$$V\,(\hat{y}_{urm}) = V\,(\hat{y}_{urs}) + \frac{X}{{}^{N-1}C_{n-1}} \Sigma' \frac{(\Sigma\, \sigma_i^2)_{s_n}}{(\Sigma\, x_i)_{s_n}} \tag{29}$$

where $V\,(\hat{y}_{urs})$ is given by (8). To obtain an unbiased estimate of the sampling variance, we note that

$$E\left[\frac{1}{{}^{N-1}C_{n-1}} \frac{(\Sigma\, T_i^2)_{s_n}}{P\,(s_n)}\right] = \sum_1^N Y_i^2 + \sum_1^N \sigma_i^2, \tag{30}$$

$$E\left[\frac{1}{{}^{N-2}C_{n-2}} \frac{\left(\sum_{j>i} T_i T_j\right)_{s_n}}{P\,(s_n)}\right] = \sum_{j>i=1}^{N} Y_i Y_j, \tag{31}$$

$$E\left[\frac{1}{{}^{N-1}C_{n-1}} \frac{(\Sigma\, s_i^2)_{s_n}}{P\,(s_n)}\right] = \sum_1^N \sigma_i^2. \tag{32}$$

where $P\,(s_n)$ is given by (17). We have

$$E\,(G_2) = Y^2 \tag{33}$$

where

$$G_2 = \frac{1}{P\,(s_n)}\left[\frac{(\Sigma\, T_i^2)_{s_n}}{{}^{N-1}C_{n-1}} + 2\frac{\left(\sum_{j>i} T_i T_j\right)_{s_n}}{{}^{N-2}C_{n-2}} - \frac{(\Sigma\, s_i^2)_{s_n}}{{}^{N-1}C_{n-1}}\right]. \tag{34}$$

Hence an unbiased estimate of $V\,(\hat{y}_{urm})$ is provided by

$$\hat{V}\,(\hat{y}_{urm}) = \hat{y}_{urm}^2 - G_2. \tag{35}$$

Comparing this with the corresponding estimator (12) in the unistage case, we arrive at the following important rule for estimating the variance in the multistage case:

"The estimate of variance in multistage sampling is the sum of two parts. The first part is equal to the esitmate of variance calculated on the assumption that the first stage units have been measured without error. The second part is obtainable from the population total estimate itself by substituting the estimated variances for the estimates of the totals of the units."

8. Extension to Two-Phase or Double Sampling Designs

We have so far restricted the discussion to the case when supplementary information on a variate x is already available. In case, however, supplementary information is not available in advance it may be considered desirable to devote a part of the budget towards collection of such informaiton by a preliminary sample and utilise the information so collected in getting a more precise estimate of the main character under study. If the usual ratio estimator (Cochran, 1953) is used, the same difficulties come up as appeared in the case of single sampling designs. We shall modify the sampling procedure to get an unbiased ratio estimate for such designs and shall give an exact expression and an unbiased estimate for its variance. Suppose there is a finite population

$$u_1, u_2, \ldots, u_N \tag{36}$$

consisting of N units. The object is to estimate the mean

$$\bar{Y} = \frac{Y}{N}$$

of a certain variate y for this population. In the first phase, we shall draw without replacement with equal probabilities a preliminary sample $s_{n'}$ of size n' in which the auxiliary variate x alone is measured. In the second phase we shall select from the preliminary sample drawn a sub-sample s_n of size n with probability proportionate to $(\Sigma x_i)_{s_n}$ and in this sample the variate y is measured (the variate x was already measured in the preliminary sample). We shall denote by $(\bar{x}')_{s_{n'}}$ the sample mean (for x) obtained from the preliminary sample $s_{n'}$. The estimate suggested is

$$\hat{y}_{urd} = \frac{(\Sigma y_i)_{s_n}}{(\Sigma x_i)_{s_n}} (\bar{x}')_{s_{n'}} \tag{37}$$

where '*urd*' stands for unbiased ratio estimate in double sampling designs. We have for given $s_{n'}$

$$E(\hat{y}_{urd}) = (\bar{y}')_{s_{n'}}$$

where the right-hand side is the sample mean (for y) based on the n' observations in the preliminary sample. Since $s_{n'}$ was selected at random, the overall expectation of $\hat{y}_{urd}$ for all possible samples of size

n' from the population given by (36) will be $\bar{Y}$, so that $\hat{y}_{urd}$ is unbiased for estimating $\bar{Y}$. The variance of the estimate is given by

$$V(\hat{y}_{urd}) = \frac{1}{n} \frac{1}{{}^{n'}C_n} \frac{1}{{}^{N}C_{n'}} \Sigma'' (\bar{x}')_{s_{n'}} \Sigma' \frac{\{(\Sigma y_i)_{s_n}\}^2}{(\Sigma x_i)_{s_n}} - \bar{Y}^2 \qquad (38)$$

where Σ' denotes summation over all possible samples s_n from $s_{n'}$ and Σ'' denotes summation over all possible samples $s_{n'}$ from the population (36). To get an unbiased estimate of the sampling variance, we notice the following:

$$E\left[\frac{(\Sigma y_i^2)_{s_n}}{(\Sigma x_i)_{s_n}} (\bar{x}')_{s_{n'}}\right] = \frac{1}{N} \sum_{1}^{N} Y_i^2, \qquad (39)$$

$$E\left[\frac{\left(\sum_{j>i} y_i y_j\right)_{s_n}}{(\Sigma x_i)_{s_n}} (\bar{x}')_{s_{n'}}\right] = \frac{n-1}{N(N-1)} \left(\sum_{j>i=1}^{N} Y_i Y_j\right). \qquad (40)$$

We thus have

$$E(G_3) = \bar{Y}^2 \qquad (41)$$

where

$$G_3 = \frac{(\bar{x}')_{s_{n'}}}{N} \left[\frac{(\Sigma y_i^2)_{s_n}}{(\Sigma x_i)_{s_n}} + 2\,\frac{N-1}{n-1}\,\frac{\left(\sum_{j>i} y_i y_j\right)_{s_n}}{(\Sigma x_i)_{s_n}}\right]. \qquad (42)$$

Hence an unbiased estimate of $V(\hat{y}_{urd})$ is provided by

$$\hat{V}(\hat{y}_{urd}) = \hat{y}_{urd}^2 - G_3. \qquad (43)$$

In case the two samples drawn are independent, *i.e.*, the second sample s_n is not a sub-sample of $s_{n'}$ (although this is not a realistic situation), it is easy to see that the estimate (37) continues to be unbiased for estimating $\bar{Y}$. For the variance of the estimate in this case we have

$$V(\hat{y}_{urd}) = \left[\sigma_x^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \bar{X}^2\right] \left[\frac{1}{N\bar{X}} \frac{1}{{}^{N-1}C_{n-1}} \Sigma' \frac{\{(\Sigma y_i)_{s_n}\}^2}{(\Sigma x_i)_{s_n}}\right] - \bar{Y}^2. \qquad (44)$$

It is easy to see that an unbiased estimate of the sampling variance is still provided by (43).

9. Summary

The problem considered is estimation of the mean or total of a character for a finite population in sampling designs making use of ratio estimation. The inadequacy of current treatment of the bias or

sampling variance of the usual ratio estimate is discussed. Modified sampling procedures appropriate to unistage, stratified, multistage and multiphase designs are given which eliminate the bias of the usual ratio estimate. For such designs exact expressions and unbiased estimates for the variances of the estimates proposed are derived. A brief discussion regarding the optimum character of the usual ratio estimate is also included.

REFERENCES

1. Cochran, W. G. .. *Sampling Techniques*, John Wiley and Sons, New York, 1953.

2. Horvitz, D. G. and Thompson, D. J. "A generalisation of sampling without replacement from a finite universe," *Jour. Amer. Stat. Assoc.*, 1952, **47**, 663–85.

3. Koop, J. C. .. "A note on the bias of the ratio estimate," *Bull. Internat. Stat. Inst.*, 1951, **33**, 141–46.

4. Lahiri, D. B. .. "A method of sample selection providing unbiased ratio estimates," *ibid.*, 1951, **33**, 133–40.

5. Midzuno, H. .. "An outline of the theory of sampling systems," *Annals of the Inst. Stat. Math.* (Japan), 1950, **1**, 149–56.

6. Sen, A. R. .. "Further developments of the theory and application of the selection of primary sampling units with special reference to the North Carolina agricultural population," *Ph.D. Thesis*, University of North Carolina, 1952.

7. Sukhatme, P. V. .. *Sampling Theory of Surveys with Applications*, Indian Society of Agricultural Statistics (India) and Iowa State College Press (U.S.A.), 1954.

8. Yates, F. and Grundy, P. M. "Selection without replacement from within strata with probability proportionate to size," *Jour. Roy. Stat. Soc.*, 1953, Series B, **15**, 253–61.