On the Variance of the Ratio Estimator

By T. J. RAO, Manchester [1])

Summary: In an earlier paper [RAO 1966] an exact expression for the variance of the ratio estimator under the MIDZUNO-SEN sampling scheme is obtained and here we study some of the interesting properties of the coefficients involved in this expression which depend on the auxiliary information. Use of these coefficients is made of in finding out an exact expression for the Bias and Mean Square Error of the ratio estimator under Simple Random Sampling With-Out Replacement (SRSWOR) scheme.

1. Introduction

The usual ratio estimator $\sum_{i\epsilon s} y_i / \sum_{i\epsilon s} x_i$ can be made unbiased by using the MIDZUNO [1952], SEN [1952] sampling scheme, which consists in drawing the first unit of the sample with probability proportional to size and the rest of the $(n-1)$ units of the sample by SRSWOR from the $(N-1)$ units of the population. An exact expression for the variance of the ratio estimator under this scheme is given by RAO [1966]

$$V(\hat{R}) = \sum_{i=1}^{N} \lambda_i Y_i^2 + \sum_{i \neq j}^{N} \sum^{N} \lambda_{ij} Y_i Y_j, \tag{1.1}$$

where
$$\lambda_i = \left\{ \frac{1}{\binom{N-1}{n-1} X} \right\} \sum_{\lambda} (X_i + X_\lambda^i)^{-1} - \frac{1}{X^2},$$

X_λ^i being the sum of the λ^{th} set of $n-1$ distinct X's not equal to X_i, the summation being taken over the $\binom{N-1}{n-1}$ such sets; and

$$\lambda_{ij} = \left\{ \frac{1}{\binom{N-1}{n-1} X} \right\} \sum_{\lambda} (X_i + X_j + X_\lambda^{ij})^{-1} - \frac{1}{X^2},$$

X_λ^{ij} being the sum of the λ^{th} set of $n-2$ distinct X's other than X_i and X_j, and the summation over λ being taken over the $\binom{N-2}{n-2}$ such sets.

---

[1]) T. J. RAO, Statistical Laboratory, Department of Mathematics, University of Manchester, Manchester. (On leave from the Indian Statistical Institute)

For the estimator of the population total Y given by $(\sum_{i\in s} y_i / \sum_{i\in s} x_i)\, X$, we have

$$V(\hat{Y}) = \sum_{i=1}^{N} (T_i - 1)\, Y_i^2 + \sum_{i\neq j}^{N}\sum^{N} (T_{ij} - 1)\, Y_i\, Y_j \qquad (1.2)$$

where
$$T_i = \frac{X}{\binom{N-1}{n-1}} \sum_{\lambda} (X_i + X_{\lambda}^{(i)})^{-1}$$

and
$$T_{ij} = \frac{X}{\binom{N-1}{n-1}} \sum_{\lambda} (X_i + X_j + X_{\lambda}^{(ij)})^{-1}.$$

The notation used above is the same as that of Rao [1967]. It is easy to note that

$$T_i = \lambda_i X^2 + 1$$

and

$$T_{ij} = \lambda_{ij} X^2 + 1.$$

In the next section we consider some of the estimators of $V(\hat{R})$.

2. Estimation of Variance

An unbiased estimate of $V(\hat{R})$ is given by Des Raj, Sen [1955]

$$\hat{V}_D(\hat{R}) = \frac{\hat{R}^2 - \sum_{i\in s} y_i^2 + 2\,\dfrac{N-1}{n-1} \sum_{\substack{i>j\\ \in s}} y_i\, y_j}{N\, n\, \bar{x}\, \bar{X}}.$$

But, $\hat{V}_D(\hat{R})$ can take sometimes negative values, and since the variance function is always positive it is desirable to have an estimator of the variance which is always positive definite. An attempt in this direction, in the context of the present estimator is due to Roy Chaudhury and his estimate is given by –

$$\hat{V}_R(\hat{R}) = \frac{1}{n\,\bar{X}^2} \left\{ m_2(y) - 2\,\frac{\bar{y}}{\bar{x}}\, m_{11}(x, y) + \left(\frac{\bar{y}}{\bar{x}}\right)^2 m_2(x) \right\},$$

where $m_2(y) = \dfrac{N-n}{N-1}\dfrac{1}{n} \sum_1^n (y_i - \bar{y})^2$, $m_{11}(x, y) = \dfrac{N-n}{N-1}\dfrac{1}{n} \sum_1^n (x_i - \bar{x})(y_i - \bar{y})$

and $m_2(x)=\frac{N-n}{N-1}\cdot\frac{1}{n}\sum_1^n(x_i-\bar{x})^2$. The main drawback of this estimate is that it is biased, despite its being positive definite. In fact,

$$E(\hat{V}_R(\hat{R}))=V\left(\frac{\bar{y}}{\bar{x}}\right)+0(n^{-3/2}).$$

For further details we refer to ROY CHAUDHURY.

Next we propose another estimator for the variance having the advantage of being unbiased, which can be suggested as a direct consequence from the expression of the variance.

Theorem 2.1: An unbiased estimator of the variance of $\hat{R}$ is given by

$$\hat{V}(\hat{R})=\sum_{i\in s}\lambda_i\, y_i^2/\pi_i'+\sum_{(i\neq j)\in s}\sum\lambda_{ij}\, y_i\, y_j/\pi_{ij}',$$

where π_i' is the probability of inclusion of the ith unit in the sample and π_{ij}' is the probability of joint inclusion of the units U_i and U_j in the sample for the MIDZUNO-SEN [SEN 1953] sampling scheme given by

$$\pi_i'=\frac{n-1}{N-1}+\frac{N-n}{N-1}\cdot\frac{X_i}{X}\quad\text{and}\quad \pi_{ij}'=\frac{n-1}{N-1}\cdot\frac{n-2}{N-2}+\frac{n-1}{N-1}\cdot\frac{N-n}{N-2}\left(\frac{X_i}{X}+\frac{X_j}{X}\right).$$

Proof: It can be easily seen that $E\,\hat{V}(\hat{R})=V(\hat{R})$ and hence etc.

Lemma 2.1: λ_i is positive for all i.

Proof: From the definition of λ_i we have

$$\lambda_i X^2+1=\frac{X}{\binom{N-1}{n-1}}\left(\frac{1}{X_i+X_1+\cdots+X_{n-1}}+\frac{1}{X_i+X_2+\cdots+X_n}+\cdots \binom{N-1}{n-1}\text{ terms}\right)$$

= inverse of the harmonic mean of

$$\frac{X_i+X_1+\cdots+X_{n-1}}{X},\ \frac{X_i+X_2+\cdots+X_n}{X},\cdots\binom{N-1}{n-1}\text{ terms}$$

> inverse of the arithmetic mean of these $\binom{N-1}{n-1}$ quantities

$$=\left\{\frac{\binom{N-1}{n-1}X_i+\binom{N-2}{n-2}(X-X_i)}{\binom{N-1}{n-1}X}\right\}^{-1}$$

$$= \left\{P_i + \frac{n-1}{N-1}(1-P_i)\right\}^{-1}$$

$$= \frac{1}{\pi_i'}$$

$$> 1$$

$$\therefore \ \lambda_i > 0 \quad \text{for all } i.$$

Theorem 2.2: A sufficient condition for $\hat{V}(\hat{R})$ to be positive is that $\lambda_{ij} \geq 0$ for all i, j.

Proof: Follows from Theorem 2.1 and Lemma 2.1.

In the next section we study some interesting properties of the T_i and T_{ij} coefficients that occur in the variance expression (1.2) (or equivalently λ_i and λ_{ij} of (1.1)).

3. Some properties of the coefficients T_i and T_{ij}

Theorem 3.1:

a) $\sum_{j \neq i}^{N} T_{ij} = (n-1)\, T_i$

b) $\sum_{i=1}^{N} T_i X_i = \frac{N X}{n}$, where $X = \sum_{1}^{N} X_i$

c) $\sum_{i=1}^{N} T_i X_i^2 + \sum_{i \neq j}^{N} \sum^{N} T_{ij} X_i X_j = X^2$

d) $T_i \geq \frac{1}{\pi_i'}$

e) $T_{ij} \geq \frac{1}{\pi_{ij}'} \left(\frac{n-1}{N-1}\right)^2$

and f) $T_i X_i \geq T_j X_j$ if and only if $X_i \geq X_j$.

Proof: (a), (b) and (c) can be verified from the definitions by proper rearrangement of the terms.

(d) is already proved in Lemma 2.1.

To prove (e) we observe that $\frac{N-1}{n-1} T_{ij}$ is the inverse of the harmonic mean of $\frac{X_i + X_j + X_\lambda^{ij}}{X}$, λ varying over the $\binom{N-2}{n-2}$ terms, $\geq$ inverse of the arithmetic mean of these terms $\frac{X_i + X_j + X_\lambda^{ij}}{X} = \left(\frac{N-1}{n-1}\pi_{ij}'\right)^{-1}$, where π_{ij}' is defined in Theorem 2.1 and this establishes (e).

(f) can be easily proved by considering

$$T_iX_i - T_jX_j = \frac{X}{\binom{N-1}{n-1}}\left\{\sum_{\lambda} X_i(X_i + X_\lambda^i)^{-1} - \sum_{\lambda} X_j(X_j + X_\lambda^j)^{-1}\right\}$$

$$= \frac{X}{\binom{N-1}{n-1}}\left\{\sum_{\substack{\lambda \\ \lambda \ni j}} X_i(X_i + X_\lambda^i)^{-1} + \sum_{\lambda} X_i(X_i + X_\lambda^{ij})^{-1}\right.$$

$$\left. - \sum_{\substack{\lambda \\ \lambda \ni i}} X_j(X_j + X_\lambda^j)^{-1} - \sum_{\lambda} X_j(X_j + X_\lambda^{ji})^{-1}\right\}$$

$$= \frac{X}{\binom{N-1}{n-1}}\left\{(X_i - X_j)\sum_{\substack{\lambda \\ \lambda \ni j}} (X_i + X_\lambda^i)^{-1} + \sum_{\lambda}\left[\left(1 + \frac{X_\lambda^{ij}}{X_i}\right)^{-1} - \left(1 + \frac{X_\lambda^{ji}}{X_j}\right)^{-1}\right]\right\}$$

≥ 0 iff $X_i \geq X_j$.

Remark: An elegant proof for (c) follows by putting $y_i = x_i$ for all i in (1.2) and setting the expression equal to zero.

It is to be noted that (b), (c) and (f) are quite useful in the problem of choice of a strategy for the method of ratio estimation and we refer to Rao [1967].

4. An exact expression for the Bias and Mean Square Error of the ratio estimator

Approximate expressions for the Bias and Mean Square Error of the ratio estimator were given by Hansen, Hurwitz and Madow, Cochran, Sukhatme, Murthy and Nanjamma and others using certain assumptions which are mostly valid for large samples. A different derivation is due to Koop who gets an expression for the bias of the ratio estimator to the order of n^{-4}. In this section we use the knowledge of auxiliary information and construct T_i and T_{ij} coefficients of higher order and show how they are useful in getting exact expressions for the Bias and Mean Square Error, which are given in a compact form.

Analagous to T_i and T_{ij} define

$$T_i^{(\alpha)} = \frac{X}{\binom{N-1}{n-1}}\sum_{\lambda}(X_i + X_\lambda^i)^{-\alpha}$$

and

$$T_{ij}^{(\alpha)} = \frac{X}{\binom{N-1}{n-1}} \sum_{\lambda} (X_i + X_j + X_{\lambda}^{ij})^{-\alpha}$$

and we denote $T_i^{(1)}$ and $T_{ij}^{(1)}$ by T_i and T_{ij} respectively.

Theorem 4.1: For a simple random sample of size n selected without replacement from a population of size N,

$$B\left(\frac{\bar{y}}{\bar{x}}\right) = \frac{n}{N\,X}\left(\sum_{i=1}^{N} T_i Y_i\right) - R$$

and

$$\text{M.S.E.}\left(\frac{\bar{y}}{\bar{x}}\right) = \frac{n}{N\,X}\left(\sum_{i=1}^{N} T_i^{(2)} Y_i^2 + \sum_{i \neq j}\sum T_{ij}^{(2)} Y_i Y_j\right) - 2R\frac{n}{N\,X}\sum_{i=1}^{N} T_j Y_i + R^2.$$

Proof:

$$B\left(\frac{\bar{y}}{\bar{x}}\right) = E\left(\frac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i}\right) - R$$

$$= \binom{N}{n}^{-1} \sum_{s\in S}\left(\frac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i}\right) - R$$

$$= \frac{n}{N\,X}\sum_{i=1}^{N} T_i Y_i - R.$$

$$\text{M.S.E.}\left(\frac{\bar{y}}{\bar{x}}\right) = E\left(\frac{\bar{y}}{\bar{x}} - R\right)^2$$

$$= \sum_{s\in S}\left(\frac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i} - R\right)^2 P_s$$

$$= \binom{N}{n}^{-1}\left\{\sum_{s\in S}\frac{(\sum_{i\in s} y_i)^2}{(\sum_{i\in s} x_i)^2} - 2R\sum_{s\in S}\left(\frac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i}\right) + \sum_{s\in S} R^2\right\}$$

$$= \binom{N}{n}^{-1}\left\{\sum_{i=1}^{N} Y_i^2\left(\sum_{\lambda}(X_i + X_{\lambda}^{i})^{-2}\right) + \sum_{i\neq j}^{N}\sum^{N} Y_i Y_j\left(\sum_{\lambda}(X_i + X_j + X_{\lambda}^{ij})^{-2}\right)\right.$$

$$\left. - 2R\sum_{i=1}^{N} Y_i\left(\sum_{\lambda}(X_i + X_{\lambda}^{i})^{-1}\right) + \binom{N}{n} R^2\right\}$$

$$= \frac{n}{N\,X}\left(\sum_{i=1}^{N} T_i^{(2)} Y_i^2 + \sum_{i\neq j}^{N}\sum^{N} T_{ij}^{(2)} Y_i Y_j\right) - 2R\frac{n}{N\,X}\sum_{i=1}^{N} T_i Y_i + R^2.$$

Remark: By proper rearrangement of terms one can show that

$$\text{a)}\quad \sum_{j\neq i}^{N} T_{ij}^{(2)} = (n-1)\, T_i^{(2)} \quad \text{and}$$

$$\text{b)}\quad \sum_{i=1}^{N} T_i^{(2)} X_i^2 + \sum_{i\neq j}^{N}\sum^{N} T_{ij}^{(2)} X_i X_j = \frac{N X}{n}$$

and we can now write the expressions for Bias and Mean Square Error in an alternate form, given by

$$B\left(\frac{\bar{y}}{\bar{x}}\right) = \frac{\sum_{i=1}^{N} T_i Y_i}{\sum_{i=1}^{N} T_i X_i} - R$$

and

$$\text{M.S.E.}\left(\frac{\bar{y}}{\bar{x}}\right) = \frac{\sum_{i=1}^{N} T_i^{(2)} Y_i^2 + \sum_{i\neq j}^{N}\sum^{N} T_{ij}^{(2)} Y_i Y_j}{\sum_{i=1}^{N} T_i^{(2)} X_i^2 + \sum_{i\neq j}^{N}\sum^{N} T_{ij}^{(2)} X_i X_j} - 2R \frac{\sum_{i=1}^{N} T_i Y_i}{\sum_{i=1}^{N} T_i X_i} + R^2.$$

Acknowledgement

The author is grateful to Dr. M. N. Murthy for constant encouragement.

References

Cochran, W. G.: Sampling Techniques. Wiley, New York, 1953.

Des Raj: On sampling with probabilities proportionate to size. Ganita, **6**, 175–182, 1955.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow: Sample Survey Methods and Theory. Wiley, New York, 1953.

Koop, J. C.: A note on the bias of the ratio estimate. Bull. Int. Stat. Inst., **33**, 141–146, 1951.

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

Murthy, M. N., and N. S. Nanjamma: Almost unbiased ratio estimates based on interpenetrating subsample estimates. Sankhya, **21**, 381–392, 1959.

Rao, T. J.: On the variance of the ratio estimator for Midzuno-Sen sampling scheme. Metrika, **10**, 89–91, 1966.

– On the choice of a strategy for ratio method of estimation. Jour. Roy. Stat. Soc., B, **29**, 392–397, 1967.

Roy Chaudhury, D. K.: Sampling with varying probabilities. Thesis submitted to the Indian Statistical Institute for the award of AISI, 1956.

Sen, A. R.: Present status of probability sampling and its use in estimation of farm characteristics (abstract). Econometrics, **20**, 103, 1952.

– On the estimate of variance in sampling with varying probabilities. Jour. Ind. Soc. Agr. Stat., **5**, 119–127, 1953.

– On the selection of n primary sampling units from a stratum structure (n > 2), Ann. Math. Stat., **26**, 744–751, 1955.

Sukhatme, P. V.: Sampling theory of surveys with applications. Ind. Soc. of Agr. Stat. and Iowa State Coll. Press, 1954.