ALMOST UNBIASED RATIO ESTIMATES BASED ON INTERPENETRATING SUB-SAMPLE ESTIMATES

By M. N. MURTHY
and
N. S. NANJAMMA
Indian Statistical Institute, Calcutta

SUMMARY. In this paper a technique is developed to estimate the bias of an ordinary ratio estimate to a given degree of approximation on the basis of the interpenetrating sub-sample estimates. This estimate of the bias is used to correct the ratio estimate for its bias, thereby obtaining an almost unbiased ratio estimate.

1. Introduction

In large scale sample surveys, the method of ratio estimation is used to estimate various ratios. It is also used to estimate totals where supplementary information is available, since under certain circumstances usually met with, it is more efficient than the conventional methods of obtaining unbiased estimates. But a satisfactory treatment of the bias and error of a ratio estimate is not yet available. However different sampling and estimation procedures have been given which provide unbiased ratio estimates. In this paper, two different types of ratio estimates based on estimates obtained from n independent, and interpenetrating sub-samples have been compared from the points of view of bias (to the second degree of approximation) and mean square error (to the fourth degree of approximation). This study helps in obtaining an estimate of the bias of the ratio estimate, for any probability sampling design. Once the bias is estimated, the ratio estimate can be corrected to give an unbiased ratio estimate (unbiased to the second degree of approximation). The gain in precision of this unbiased ratio estimate as compared with the biased one has been studied.

The above results can be generalised to estimate the bias of a ratio estimate to any degree of approximation, using a series of ratio estimates based on a number of independent and interpenetrating sub-samples. These generalised results, for the particular case where the estimates of the variates in question are distributed in the bivariate normal form are given in sections 7 and 8 of this paper.

2. Approximations for the bias and mean square error of a ratio estimate

Let $\hat{y}$ and $\hat{x}$ be unbiased estimates of y and x, the population totals of two characteristics, based on any probability sample. $\hat{y}/\hat{x}$ can be considered as an estimate of the ratio $R = y/x$. This estimate is consistent but biased. Assuming that $\frac{|\hat{x}-x|}{x} < 1$, and neglecting terms of degree greater than two in the expansion of $\left(1+\frac{\hat{y}-y}{y}\right)\left(1+\frac{\hat{x}-x}{x}\right)^{-1}$, it can be shown that the bias of $\hat{y}/\hat{x}$

$$B(\hat{y}/\hat{x}) = \frac{1}{x^2}\,[R\,\mathrm{var}(\hat{x}) - \mathrm{cov}\,(\hat{x}, \hat{y})]. \qquad \ldots \quad (2.1)$$

The mean square error of $\hat{y}/\hat{x}$, to the fourth degree of approximation, is

$$M(\hat{y}/\hat{x}) = R^2 \left\{ \left(\frac{\mu_{02}}{y^2} - \frac{2\mu_{11}}{xy} + \frac{\mu_{20}}{x^2} \right) + 2\left(\frac{2\mu_{21}}{x^2 y} - \frac{\mu_{12}}{xy^2} - \frac{\mu_{30}}{x^3} \right) + 3\left(\frac{\mu_{22}}{x^2y^2} - \frac{2\mu_{31}}{x^3 y} + \frac{\mu_{40}}{x^4} \right) \right\} \quad \ldots \ (2.2)$$

where $\mu_{ij} = \mathcal{E}\,[(\hat{x}-x)^i\,(\hat{y}-y)^j]$.

If the sample size is fairly large, the assumption $\frac{|\hat{x}-x|}{x} < 1$ can be considered to be valid. Further x usually denotes the number of persons, or households,or some such characteristic for which we expect reliable estimates with a good design. For simple random sampling, a large number of empirical studies have shown that generally if the sample size is greater than 30, the assumption that $\frac{|\hat{x}-x|}{x} < 1$ is valid; and that the contribution of the higher degree terms to the bias and variance of the ratio estimate will be negligible.

3. Comparison of two different ratio estimates

Let $(y_i,\ x_i)$ be unbiased estimates of the population totals y and x, from the i-th independent interpenetrating sub-sample ($i = 1, 2, \ldots, n$). The following two ratio estimates can be considered as estimates of $R = y/x$

(i) $$R_1 = \frac{y_1+y_2+\ldots+y_n}{x_1+x_2+\ldots+x_n}$$

(ii) $$R_n = \frac{1}{n}\left(\frac{y_1}{x_1}+\frac{y_2}{x_2}+\ldots+\frac{y_n}{x_n}\right)$$

Applying result (2.1) to $R_1 = \left(\dfrac{\sum_{i=1}^{n} y_i/n}{\sum_{i=1}^{n} x_i/n} \right)$, we get the bias of R_1,

$$B(R_1) = B_1 = \frac{1}{x^2}\left[R \text{ var}\left(\frac{\Sigma x_i}{n}\right) - \text{cov}\left(\frac{\Sigma y_i}{n}, \frac{\Sigma x_i}{n}\right)\right]$$

since $\mu_{11}(x_i, y_j) = 0$ for $i \neq j$

$$= \frac{1}{n^2x^2} \sum_{i=1}^{n} \left\{ R\mu_{20}(x_i) - \mu_{11}(x_i, y_i) \right\}$$

$$= \frac{1}{n^2} \left\{ \sum_{i=1}^{n} B\left(\frac{y_i}{x_i}\right)\right\}. \quad \ldots \ (3.1)$$

Bias of R_n, $$B(R_n) = B_n = \frac{1}{n}\left\{ \sum_{i=1}^{n} B\left(\frac{y_i}{x_i}\right)\right\}. \quad \ldots \ (3.2)$$

Comparing (3.1) and (3.2), we note that the bias of R_n is n times that of R_1, to the second degree of approximation.

We now compare the mean square errors of R_1 and R_n to the fourth degree of approximation, assuming that the sub-sample sizes are the same (as is the case generally) so that

$$B\left(\frac{y_i}{x_i}\right) = B$$

$$\mu_{rs}(x_i, y_i) = \mu_{rs}$$

and

$$M\left(\frac{y_i}{x_i}\right) = M \text{ for all } i.$$

By applying result (2.2) to R_1 and simplifying, we obtain,

$$M(R_1) = M_1 = \frac{R^2}{n}\left\{\left(\frac{\mu_{02}}{y^2}+\frac{\mu_{20}}{x^2}-\frac{2\mu_{11}}{xy}\right)+\frac{2}{n}\left(\frac{2\mu_{21}}{x^2y}-\frac{\mu_{30}}{x^3}-\frac{\mu_{12}}{xy^2}\right)+\right.$$

$$+\frac{3}{n^2}\left(\frac{\mu_{40}}{x^4}+\frac{\mu_{22}}{x^2y^2}-\frac{2\mu_{31}}{x^3y}\right)+\frac{3(n-1)}{n^2}\left(\frac{3\mu_{20}^2}{x^4}+\frac{\mu_{20}\mu_{02}}{x^2y^2}+\frac{2\mu_{11}^2}{x^2y^2}-\frac{6\mu_{20}\mu_{11}}{x^3y}\right)$$

$$= \frac{M}{n}-\frac{n-1}{n^2}A \qquad \dots \quad (3.3)$$

where

$$A = R^2\left[2\left(\frac{2\mu_{21}}{x^2y}-\frac{\mu_{30}}{x^3}-\frac{\mu_{12}}{xy^2}\right)+\frac{3(n+1)}{n}\left(\frac{\mu_{40}}{x^4}+\frac{\mu_{22}}{x^2y^2}-\frac{2\mu_{31}}{x^3y}\right)-\right.$$

$$\left.-\frac{3}{n}\left(\frac{3\mu_{20}^2}{x^4}+\frac{\mu_{20}\mu_{02}}{x^2y^2}+\frac{2\mu_{11}^2}{x^2y^2}-\frac{6\mu_{20}\mu_{11}}{x^3y}\right)\right].$$

and

$$M(R_n) = M_n = \mathcal{E}(R_n-R)^2 = \mathcal{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{x_i}-R\right)\right]^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathcal{E}\left(\frac{y_i}{x_i}-R\right)^2+\frac{1}{n^2}\sum_{i\neq j}B\left(\frac{y_i}{x_i}\right)B\left(\frac{y_j}{x_j}\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}M\left(\frac{y_i}{x_i}\right)+\frac{1}{n^2}\sum_{i\neq j}^{n}B\left(\frac{y_i}{x_i}\right)B\left(\frac{y_j}{x_j}\right) = \frac{M}{n}+\frac{(n-1)B^2}{n}$$

$$\dots \quad (3.4)$$

from (3.3) and (3.4) $M_n = M_1+\frac{n-1}{n^2}A+\frac{n-1}{n}B^2$.

Comparison of M_n and M_1 is dificult in general. Hence it is assumed that $\hat{x}$ and $\hat{y}$ are distributed in the bivariate normal form in which case the bias and mean square error of $\hat{y}/\hat{x}$ reduce to

$$B = Rc_x(c_x-\rho c_y)$$

$$M = R^2[(c_y^2-2\rho c_xc_y+c_x^2)(1+3c_x^2)+6c_x^2(c_x-\rho c_y)^2]$$

Further $$A = 3R^2c_x^2[(c_y^2-2\rho c_xc_y+c_x^2)+2(c_x-\rho c_y)^2], \text{ which is } >0$$

where $$c_x^2 = \frac{\mu_{20}}{x^2},\ c_y^2 = \frac{\mu_{02}}{y^2},$$

and ρ = correlation coefficient between $\hat{x}$ and $\hat{y}$.

$\therefore$ The mean square error of R_n is greater than that of R_1. Thus R_1 is better than R_n from the considerations of both bias and mean square error.

4. ESTIMATION OF THE BIAS OF THE RATIO ESTIMATE

An unbiased estimate of the bias of the ratio estimate to the second degree of approximation, is given below.

$$\mathcal{E}(R_1) = R+B_1$$

$$\mathcal{E}(R_n) = R+B_n$$

$$\therefore \mathcal{E}(R_n-R_1) = B_n-B_1.$$

But $$B_n = nB_1$$

$$\therefore \mathcal{E}(R_n-R_1) = (n-1)B_1.$$

$\therefore \quad \hat{B}_1 = \dfrac{R_n-R_1}{n-1}$ is an unbiased estimate of B_1, the bias of R_1 to the second degree of approximation.

The variance of the estimate of bias of R_1 is given by

$$V(\hat{B}_1) = \frac{1}{(n-1)^2}(V_1+V_n-2\rho_{R_1R_n}\sqrt{V_1V_n})$$

where $$V_1 = \text{Variance of } R_1 = M_1-B_1^2$$

$$V_n = \text{Variance of } R_n = M_n-B_n^2$$

and $$\rho_{R_1R_n} = \text{Correlation coefficient of } R_1 \text{ and } R_n.$$

$\therefore$ using (3.1), (3.2), (3.3) and (3.4), we get

$$V_1 = V_n+\frac{n-1}{n^2}(B^2-A)$$

$$\therefore \quad V(\hat{B}_1) = \frac{V_n}{(n-1)^2}(\alpha^2-2\rho_{R_1R_n}\alpha+1) \quad \ldots \quad (4.1)$$

where $$\alpha^2 = 1+\frac{n-1}{n^2}\frac{B^2-A}{V_n}.$$

If $\bar{x}$ and $\bar{y}$ are bivariate normally distributed

$$A-B^2 = 3c_x^2R^2[(c_y^2-2\rho c_x c_y+c_x^2)+\tfrac{1}{2}(c_x-\rho c_y)^2]$$

which is greater than or equal to 0.

It follows that $\alpha^2 \leqslant 1$ and therefore $V_1 < V_n$. Thus it may be observed that R_1 is a better estimate than R_n from the point of view of bias, mean square error as well as variance.

The expression for the correlation coefficient between R_1 and R_n to the fourth degree of approximation is

$$\rho_{R_1R_n} = \frac{n\left\{\dfrac{n^2+3n+(1+\rho)(n+1)c^2}{(n+c^2)^2}-\dfrac{n+1}{n}-\dfrac{c^2(1-\rho)}{n}\right\}}{(11c^2-5c^2\rho+2)^{\frac{1}{2}}(11c^2-5c^2\rho+2n)^{\frac{1}{2}}} \qquad \ldots \quad (4.2)$$

under the assumption that

(a) $\bar{x}$ and $\bar{y}$ are bivariate normally distributed with the same coefficient of variation $(c_x = c_y = c)$ and

(b)[1] $\qquad (x_1^2+x_1x_2+\ldots+x_1x_n) < 2[V(\bar{x})+n\,\bar{x}^2]$.

In the above expression ρ stands for the correlation coefficient of $\bar{x}$ and $\bar{y}$. If c^2 is small, $\rho_{R_1R_n}$ will be nearly equal to 1.

The coefficient of variation of the estimate of bias may be large; still it may be possible to get a ratio estimate corrected for its bias which is more efficient than the biased one.

5. An almost unbiased ratio estimate

Since we have obtained an estimate of bias of R_1, that can be used to correct R_1 for its bias, and we get an almost unbiased ratio estimate.

$$R_c = \left(\frac{n^2R_1-R_n}{n-1}\right)$$

---

[1] This assumption was necessary to derive the $\mathcal{E}(R_1R_n)$, for

$$\rho_{R_1R_n} = \frac{\mathcal{E}(R_1R_n)-\mathcal{E}(R_1)\mathcal{E}(R_n)}{(V(R_1)V(R_n))^{\frac{1}{2}}}$$

$$\mathcal{E}(R_1R_n) = \mathcal{E}\left(\frac{y_1^2+y_1y_2+\ldots+y_1y_n}{x_1^2+x_1x_2+\ldots+x_1x_n}\right)$$

$$= \frac{V(\bar{y})+n\bar{y}^2}{V(\bar{x})+n\bar{x}^2}+\frac{(V(\bar{y})+n\bar{y}^2)V(\epsilon')-(V(\bar{x})+n\bar{x}^2)\,\mathrm{cov.}(\epsilon,\epsilon')}{(V(\bar{x})+n\bar{x}^2)}$$

where $\qquad s = (y_1^2+y_1y_2+\ldots+y_1y_n)-(V(\bar{y})+n\bar{y}^2)$

$\qquad s' = (x_1^2+x_1x_2+\ldots+x_1x_n)-(V(\bar{x})+n\bar{x}^2)$.

We say it is an 'almost' unbiased estimate because it is unbiased only to the second degree of approximation. The variance of the corrected estimate is

$$V(R_c) = \frac{1}{(n-1)^2}\left(n^2V_1+V_n-2n\rho_{R_1R_n}\sqrt{V_1V_n}\right)$$

$$= \frac{V_n}{(n-1)^2}\left(n^2\alpha^2-2n\rho_{R_1R_n}\alpha+1 \right) \qquad \dots \quad (5.1)$$

The gain in precision due to using R_{1c} instead of R_1 is given by

$$G(R_c) = \frac{M_1-V(R_c)}{M_1} = 1-\frac{n^2\alpha^2-2n\rho_{R_1R_n}\alpha+1}{(n-1)^2(\alpha^2+z^2)} \qquad \dots \quad (5.2)$$

where $z^2 = \frac{B^2}{n^2V_n} = \frac{B^2}{nV}$, B and V being the bias and variance of the ratio estimate based on one sub-sample. It may be noted that

$$\frac{|B|}{V^{\frac{1}{2}}} < c_x.$$

where c_x is the coefficient of variation of the estimate $\hat{x}$ based on one sub-sample. If the sub-sample size is large c_x will be small. Hence z^2 can be neglected. It is to be noted that neglecting z^2 does not amount to neglecting bias. The gain in precision can be written as

$$G(R_c) = 1-\frac{n^2\alpha^2-2n\rho_{R_1R_n}\alpha+1}{(n-1)^2\alpha^2}. \qquad \dots \quad (5.3)$$

Further the expression in (5.2) is greater than that in (5.3).

$$G(R_c) > 0, \text{ if } (n-1)^2\alpha^2-(n^2\alpha^2-2n\rho_{R_1R_n}\alpha+1) > 0$$

$$\text{i.e., if } (2n-1)\alpha^2-2n\rho_{R_1R_n}\alpha+1 < 0.$$

which will be true if α lies between the roots of the equation

$$(2n-1)\alpha^2-2n\rho_{R_1R_n}\alpha+1 = 0 \qquad \dots \quad (5.4)$$

(i.e.) if α lies between $\dfrac{n\rho_{R_1R_n}\pm(n^2\rho^2_{R_1R_n}-2n+1)^{\frac{1}{2}}}{(2n-1)}$.

For given values of α and $\rho_{R_1R_n}$, the minimum value of n which makes $G(R_c) > 0$ is given by

$$n = \left[\frac{(1-\alpha^2)}{2\alpha(\rho_{R_1R_n}-\alpha)}\right]+1 \qquad \dots \quad (5.5)$$

It can be seen that $G(R_c)$ will be positive only if $\rho_{R_1R_n} > \alpha$. Further for given values of α and $\rho_{R_1R_n}$ where $\rho_{R_1R_n} > \alpha$, the value of n which maximises the gain is

$$n = \frac{(1-\rho_{R_1R_n}\alpha)}{\alpha(\rho_{R_1R_n}-\alpha)} . \qquad \dots \quad (5.6)$$

For given values of n and $\rho_{R_1R_n}$ the value of α which maximises the gain is

$$\alpha = \frac{1}{n\rho_{R_1R_n}} .$$

A table showing for given values of $\rho_{R_1R_n}$ and α, the minimum value of n required to make the gain positive, the optimum n and the maximum gain are given below.

MINIMUM AND MAXIMUM VALUES OF $G(R_c)$ WITH THE CORRESPONDING VALUES OF n FOR DIFFERENT $\rho_{R_1R_n}$ AND α WHERE $\rho_{R_1R_n} > \alpha$.

sl. no.	α	$\rho_{R_1R_n}$	minimum		maximum	
			n	$G(R_c)$	n	$G(R_c)$
(0)	(1)	(2)	(3)	(4)	(5)	(6)
1	0.6	0.7	6	0.0089	10	0.0192
2		0.8	3	0.0550	4	0.0988
3		0.9	2	0.0889	3	0.3056
4	0.7	0.8	4	0.0113	7	0.0260
5		0.9	2	0.1020	3	0.1684
6	0.8	0.9	3	0.0409	4	0.0486

6. Empirical study

In section 5 we have discussed the efficiency of the corrected estimate R_c as compared to that of R_1. There it has been pointed out that under certain circumstances R_c will be a better estimate of R than R_1. In this section we give an example where the variance of R_c turned out to be less than the mean square error of R_1

The data for this study consist of the village-wise figures for the number of households and the number of persons attending the village market for a sample of 300 villages scattered over a wide region. Treating this as the population,two samples of size 30 villages are drawn systematically with independent random starts to estimate the number of persons going to market per household. From these two sub-samples the estimates R_1 and R_n are calculated. Then the corrected estimate is found by estimating the bias of R_1.

For the purpose of this study, all the possible pairs of systematic samples are enumerated. Then for each of the pairs R_1 R_n, and R_n-R_1 are calculated. The variance of R_c, the corrected estimate and the mean square error of R_1 are determined. The results are given below.

$$\text{Population ratio} = 7.6857$$
$$E(R_1) = 7.9211 \text{ and } E(R_n) = 8.1401$$
$$B(R_1) = 0.2354 \text{ and } B(R_n) = 0.4544$$

It is to be noted that B_1 is almost half of B_n. This may be taken as indicating that second degree approximation is good enough. The variance of R_c and mean square error of R_1 are

$$V(R_c) = 8.9992 \text{ and } M(R_1) = 9.6144$$
$$\rho_{R_1 R_n} = 0.9856 \text{ and } \alpha = 0.8871$$
$$G(R_c) = 6.4\%$$

7. Comparison of a series of ratio estimates

When n, the number of independent and interpenetrating sub-samples is a multiple of 2, 3, ... and k, we can construct the following series of ratio estimates.

$$R_m = \frac{1}{m}\left\{\frac{\sum\limits_{i=1}^{\frac{n}{m}} y_i}{\sum\limits_{i=1}^{\frac{n}{m}} x_i} + \frac{\sum\limits_{i=\frac{n}{m}+1}^{2\frac{n}{m}} y_i}{\sum\limits_{i=\frac{n}{m}+1}^{2\frac{n}{m}} x_i} + \ldots + \frac{\sum\limits_{i=(m-1)\frac{n}{m}+1}^{n} y_i}{\sum\limits_{i=(m-1)\frac{n}{m}+1}^{n} x_i}\right\}$$

where $m = 1, 2, 3, \ldots, k, n$.

$$= \frac{1}{m}\left\{\sum_{j=1}^{m} (R_{n/m})_j\right\} \text{ where } (R_{n/m})_j = \frac{\sum\limits_{i=(j-1)\frac{n}{m}+1}^{j\frac{n}{m}} y_i}{\sum\limits_{i=(j-1)\frac{n}{m}+1}^{j\frac{n}{m}} x_i}.$$

It may be noted that there are $\dfrac{n!}{m!\left\{\left(\dfrac{n}{m}\right)!\right\}^m}$ different ways of partitioning the n sub-samples into m partitions each containing n/m sub-samples.

In practice the situation may arise where $(\hat{x}, \hat{y})$ are approximately distributed in the bivariate normal form. In such a case we may make use of the properties of the bivariate normal distribution for writing down the expressions for the bias and mean square error of the ratio estimate. It is to be noted that the infinite series for the bias and mean square error are divergent. As has been rightly pointed out by Cramér and Kendall in their books, in statistical practice one is interested not so much in the convergence properties of the infinite series representing a function but in finding out whether the first few terms of that series will give a good approximation to the function.

Naturally in a finite population where the estimate $\hat{x}$ does not take the value 0, the bias and the mean square error of the ratio estimate $\hat{y}/\hat{x}$ will be finite quantities. The formal expressions for the bias and mean square error in terms of infinite series under the assumption of bivariate normality are considered here. The problem as to how many terms are to be taken to obtain a desired degree of approximation in different situations is yet to be fully investigated. So in discussing the bias, only terms of degree greater than $2k$ are neglected where k is any finite number.

$$\text{Bias of } \frac{\hat{y}}{\hat{x}} \text{ is given by } B(\hat{y}/\hat{x}) = R(c_x-\rho c_y)\sum_{j=1}^{k}\frac{(2j)!}{2^j j!}c_x^{2j-1} = \sum_{j=1}^{k} A_j \quad \dots \quad (7.1)$$

where
$$A_j = R(c_x-\rho c_y)\frac{(2j!)}{2^j j!}c_x^{2j-1}.$$

Mean square error of $\frac{\hat{y}}{\hat{x}}$ to the fourth degree of approximation is given by

$$M\left(\frac{\hat{y}}{\hat{x}}\right) = R^2\{(1+3c_x^2)(c_y^2-2\rho c_x c_y+c_x^2)+6c_x^2(c_x-\rho c_y)^2\}. \quad \dots \quad (7.2)$$

$$\text{Bias of } R_m,\ B(R_m) = B_m = \left(\frac{m}{n}\right)^{\frac{1}{2}} R(c_x-\rho c_y)\sum_{j=1}^{k}\frac{(2j)!}{2^j j!}\frac{c_x^{2j-1}}{\left(\frac{n}{m}\right)^{j-\frac{1}{2}}} = \sum_{j=1}^{k}\frac{m^j}{n^j}A_j = \sum_{j=1}^{k} m^j\eta_j \quad \dots \quad (7.3)$$

where $\quad \eta_j = \dfrac{A_j}{n^j}$.

From (7.3) it follows that the biases of the series of ratio estimates have the same sign, and the absolute magnitude of the bias increases with m. From (7.2) the mean square error of R_m to the fourth degree of approximation is given by

$$M(R_m) = M_m = \frac{R}{n}(c_y^2-2\rho c_x c_y+c_x^2)+\frac{m}{n^2}A+\frac{m(m-1)}{n^2}B^2$$

$$= \frac{M-A}{n}+\frac{m}{n^2}A+\frac{m(m-1)}{n^2}B^2$$

where A, B and M have been defined in section 3. Since $A > 0$, M_m is an increasing function of m.

Further $M_m - M_{m-1} = \dfrac{A}{n^2}+\dfrac{2(m-1)}{n^2}B^2$ which also increases as m increases.

8. Estimation of the bias of a series of ratio estimates

The bias of R_m to the $(2k)$-th degree of approximation can be estimated as given below, from n independent and interpenetrating sub-samples, provided[1] n is a multiple of 2, 3, 4 ... and k, where $\hat{x}$ and $\hat{y}$ are distributed in the bivariate normal form. (The bias, when $\hat{x}$ and $\hat{y}$ are not bivariate normally distributed can also be estimated by adopting a similar procedure).

The bias of R_m to the $2k$-th degree of approximation is given by

$$B = \left(\sum_{j=1}^{k} m^j \eta_j \right) \quad m = 1, 2, \ldots k, n \quad \text{[see (7.3)]} \qquad \ldots \quad (7.4)$$

$$\mathcal{E}(R_m) = R + B_m$$

$$\therefore \qquad \mathcal{E}(R_m - R_1) = B_m - B_1 = \sum_{j=1}^{k} (m^j - 1)\eta_j \qquad \ldots \quad (7.5)$$

Let $$D_m = R_m - R_1$$

From (7.5) $$\underset{1\times k}{\mathcal{E}(D)} = \underset{1\times k}{(\eta)} \quad \underset{1\times k}{(\Lambda)}$$

where $$D = (D_2, D_3, \ldots, D_k, D_n)$$

$$\eta = (\eta_1, \eta_2, \ldots, \eta_k)$$

$$\Lambda = \begin{pmatrix} 2-1 & 3-1 & \ldots, & k-1 & n-1 \\ 2^2-1 & 3^2-1 & \ldots, & k^2-1 & n^2-1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 2^k-1 & 3^k-1 & \ldots, & k^k-1 & n^k-1 \end{pmatrix}$$

$$\therefore \quad \eta = \mathcal{E}(D)\, \Lambda^{-1},$$

From (5.6) $$\underset{1\times k}{(B)} = \underset{1\times k}{(\eta)} \quad \underset{k\times k}{(\Lambda+\epsilon)}$$

where $$B = (B_2, B_3 \ldots B_k, B_n)$$

and ϵ is a $(k\times k)$ matrix whose elements are all equal to 1.

$$\therefore \quad B = \mathcal{E}[D] + \mathcal{E}[D]\, \Lambda^{-1}\epsilon.$$

But $$\Lambda^{-1}\epsilon = \begin{pmatrix} s_1 & s_1 & \ldots & s_1 \\ s_2 & s_2 & \ldots & s_2 \\ \cdot & \cdot & \cdot & \cdot \\ s_k & s_k & \ldots & s_k \end{pmatrix}$$

where s_m is the sum of the elements in the m-th row of Λ^{-1}.

$\therefore$ An estimate of (B), $(\hat{B}) = D + D\Lambda^{-1}\epsilon = D + D[S](1)$

where $$[S] = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix} \text{ and } (1) = (1, 1, \ldots 1)$$

$$\therefore \quad \hat{B}_m = \sum_j D_j S_{j-1} + D_m \quad \text{where } j = 2, 3 \ldots k, n, \text{ and } s_{n-1} = s_k \quad \ldots \quad (7.6)$$

Hence the corrected estimate is given by $R_c = R_m - \hat{B}_m$.

[1] This condition is only sufficient but not necessary. The estimate of bias and hence the corrected estimate may be obtained even if it is not a multiple of 2, 3, . . . k by considering the series of ratio estimates defined over over-lapping partitions of the n sub-samples.

Particular cases :

$$(1)\quad k=1: R_1 = \frac{\sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n} x_i}, R_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{x_i}\right)$$

$\therefore$ $$B_m = m\eta_1 = \frac{m}{n} A_1. \text{ where } m = 1, n.$$

Since $$\Lambda = (n-1),\ \Lambda^{-1} = \frac{1}{n-1} \text{ and } [S] = \frac{1}{n-1}$$

$\therefore$ $$B_n = (R_n - R_1) + (R_n - R_1)\frac{1}{n-1} = \frac{n}{n-1}(R_n - R_1)$$

But $$B_n - B_1 = \mathcal{E}(R_n - R_1)$$

$\therefore$ $$\hat{B}_1 = \hat{B}_n - (R_n - R_1) = \frac{n}{n-1}(R_n - R_1) - (R_n - R_1) = \frac{R_n - R_1}{n-1}$$

$$(2)\quad k=2: R_1 = \frac{\sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n} x_i}, R_2 = \left\{\frac{\sum\limits_{i=1}^{n/2} y_i}{\sum\limits_{i=1}^{n/2} x_i} + \frac{\sum\limits_{i=(n/2)+1}^{n} y_i}{\sum\limits_{i=(n/2)+1}^{n} x_i}\right\} \text{ and } R_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{x_i}\right)$$

where n is a multiple of 2.

Here $$B_m = m\eta_1 + m^2\eta_2 = \frac{m}{n}A_1 + \frac{m^2}{n^2}A_2\,;\ m = 1, 2 \text{ and } n.$$

$$\Lambda = \begin{pmatrix} 2-1 & n-1 \\ 2^2-1 & n^2-1 \end{pmatrix} = \begin{pmatrix} 1 & n-1 \\ 3 & n^2-1 \end{pmatrix}$$

$\therefore$ $$\Lambda^{-1} = \frac{1}{(n-1)(n-2)}\begin{pmatrix} n^2-1 & -(n-1) \\ -3 & 1 \end{pmatrix}$$

$\therefore$ $$s_1 = \frac{n}{n-2},\ s_2 = \frac{-2}{(n-1)(n-2)}$$

$\therefore$ $$(\hat{B}_2, \hat{B}_n) = (R_2 - R_1, R_n - R_1) + (R_2 - R_1, R_n - R_1)\begin{pmatrix} \frac{n}{n-2} \\ \frac{-2}{(n-1)(n-2)} \end{pmatrix}(1,1)$$

$$\therefore \qquad \hat{B}_2 = (R_2-R_1)+\frac{n}{n-2}(R_2-R_1)-\frac{2(R_n-R_1)}{(n-1)(n-2)}$$

$$= \frac{-2n}{n-1}R_1+\frac{2(n-1)}{n-2}\;R_2-\frac{2R_n}{(n-1)(n-2)}$$

similarly $$\hat{B}_n = \frac{-2n}{n-1}R_1+\frac{n}{n-2}R_2+\frac{n(n-3)R_n}{(n-1)(n-2)}$$

and $$\hat{B}_1 = -\frac{(n+1)}{n-1}R_1+\frac{nR_2}{n-2}-\frac{2R_n}{(n-1)(n-2)}\;.$$

Hence the almost unbiased estimate in this case is

$$R_e = \frac{2nR_1}{n-1}-\frac{nR_2}{n-2}+\frac{2R_n}{(n-1)(n-2)}\;.$$

It may be noted that the results given above will also be obtained when an estimate of the bias to the third degree of approximation is considered, in the case when $\mathcal{Z}$ and $\mathcal{Y}$ are not distributed in the bivariate normal form.

The authors wish to thank Prof. D. B. Lahiri for his guidance in preparing this paper.

REFERENCES

CRAMER, H. (1946): *Mathematical Methods of Statistics.* Princeton University Press.

FIELLER, E. C. (1932): The distribution of the index in a normal bivariate population. *Biometrika*, 24, 428–440.

KENDALL, M. G. (1948): *Advanced Theory of Statistics*, Vol. I., Charles Griffin & Co.

SUKHATME, P. V. (1953): *Sampling Theory of Surveys with Applications.* The Iowa State College Press.

Paper received : November, 1956.

Revised : March, 1959.