ON THE ESTIMATION OF RATIO AND PRODUCT OF THE POPULATION PARAMETERS

By M. P. SINGH

Indian Statistical Institute

SUMMARY. In this paper two estimators for the estimation of ratios and two for the estimation of products, which utilise information on supplementary characteristics, have been proposed. The specific conditions under which the proposed ratio and product estimators are more efficient than the usual ratio and product estimators respectively, have been obtained for any sample design. An example is given for illustration. The use of multi-supplementary variables is also briefly discussed.

1. Introduction

The estimation of ratio or product of two parameters is of considerable importance in practice. For instance, it is as important and sometimes more important to estimate the proportions of population under different means of livelihood and total crop production (i.e. product of cultivated area and yield rate) in a region than estimating the total population and total cultivated area. In estimating ratios and products the commonly adopted practice is to take the ratio and product respectively of the unbiased estimates of the two parameters as an estimator. In case the true value of one of the parameters is known one may feel that it is sufficient to estimate the other parameter only and obtain an estimate of the true ratio or product by dividing (for ratio) or multiplying (for product) this estimator by the known value of the other parameter. However, even if the true value of one of the parameters is known, in many situations it is more efficient to use its unbiased estimator in estimating the ratio or the product.

In this paper two estimators have been proposed for estimating the ratio or the product of two parameters. They are shown to be more efficient than usual ratio or product estimator under certain conditions. The results derived here are quite general in nature and are applicable to any sample design and many of the parameters commonly met with in practice. An illustrative example has been given for the case of simple random sampling in Section 5. In the last section use of more than one supplementary variable has been briefly discussed.

2. Formulation of the estimators

We shall consider the general procedure of the estimation of the population parameters given by Nanjamma, Murthy and Sethi (1959) in connection with obtaining unbiased ratio estimators and by Murthy (1963a) in giving generalised unbiased estimators. The population parameters expressible in the form

$$F_i = \sum_{\alpha_i \in A_i} f_i(\alpha_i) \qquad i = 1, 2, 3 \qquad \ldots \quad (2.1)$$

are considered, where $f_i(\alpha_i)$ is single-valued set function defined over the class A_i of sets α_i whose elements belong to the finite population $\mathscr{X}$.

For any sample design $D = D(S, P)$ where S is the set of samples s and P the probability measure defined over it, such that $\sum_{s \in S} P(s) = 1$, assuming that the sample space is so specified that each $\alpha_i \epsilon A_i$ occurs in at least one s and each s contains at least one set $\alpha_i \epsilon A_i$; the generalised unbiased estimator of $F_i (i = 1, 2, 3)$ is given as

$$\hat{F}_i = \sum_{\alpha_i \epsilon s} f_i(\alpha_i)\, \phi_i(s, \alpha_i)/P(s) \qquad \ldots \quad (2.2)$$

where
$$\sum_{s \supset \alpha_i} \phi_i(s, \alpha_i) = 1.$$

In the present paper we shall confine ourselves to the estimation of the ratio $R(= F_1/F_2)$ and the product $P(= F_1 \cdot F_2)$ of the parameters F_1 and F_2. The commonly adopted method for estimating them is to find the ratio $(\hat{R})$ and the product $(\hat{P})$ respectively of the corresponding estimates $\hat{F}_1$ and $\hat{F}_2$. It is well known that $\hat{R}$ and $\hat{P}$ estimate the true ratio R and product P respectively, efficiently only if $f_1(\alpha_1)$ and $f_2(\alpha_2)$ are highly positively correlated in the former and highly negatively correlated in the latter case. Here two new estimators for estimating R and P have been proposed, which can be efficiently used even if the above conditions are not met.

Let us define the estimators of R as

$$R_1^* = \hat{R}\left(\frac{\hat{F}_3}{F_3}\right) = \frac{\Sigma f_1(\alpha_1)\phi_1(s, \alpha_1)}{\Sigma f_2(\alpha_2)\, \phi_2(s, \alpha_2)} \; \frac{\Sigma f_3(\alpha_3)\phi_3(s, \alpha_3)}{P(s)F_3} \qquad \ldots \quad (2.3)$$

$$R_2^* = \hat{R}\left(\frac{F_3}{\hat{F}_3}\right) = \frac{\Sigma f_1(\alpha_1)\, \phi_1(s, \alpha_1)P(s) \cdot F_3}{\Sigma f_2(\alpha_2)\, \phi_2(s, \alpha_2)\Sigma f_3(\alpha_3)\phi_3(s, \alpha_3)} \qquad \ldots \quad (2.4)$$

and the estimators of P as

$$P_1^* = \hat{P}\left(\frac{\hat{F}_3}{F_3}\right) = \frac{\Sigma f_1(\alpha_1)\, \phi_1(s, \alpha_1)\, \Sigma f_2(\alpha_2)\, \phi_2(s, \alpha_2)\, \Sigma f_3(\alpha_3)\phi_3(s, \alpha_3)}{F_3[P(s)]^3} \qquad \ldots \quad (2.5)$$

$$P_2^* = \hat{P}\left(\frac{F_3}{\hat{F}_3}\right) = \frac{\Sigma f_1(\alpha_1)\, \phi_1(s, \alpha_1)\Sigma f_2(\alpha_2)\phi_2(s, \alpha_2)}{\Sigma f_3(\alpha_3)\phi_3(s, \alpha_3)\, P(s)} \cdot F_3 \qquad \ldots \quad (2.6)$$

summation being taken over each $\alpha_i \epsilon s$. The particular ratio and products of interest can be easily obtained by assigning specific values to α_i, $f_i(\alpha_i)$, $\phi_i(s, \alpha_i)$ and $P(s)$.

It is to be noted here that the parameter F_3 given in (2.1) is defined for the supplementary variable and that information on $f_3(\alpha_3)$ related to $f_1(\alpha_1)$ and $f_2(\alpha_2)$ is available in the survey for each $\alpha_i \epsilon A_i$ in $\mathscr{X}$. In case F_3 is not known in advance and if it is easier and less costly to observe $f_3(\alpha_3)$ than $f_1(\alpha_1)$ and $f_2(\alpha_2)$, then a two-phase sample may be used. The procedure consists in selecting a large first-phase sample S' from a given population with some probability scheme and observing the value of $f_3(\alpha_3)$ for all $\alpha_3 \epsilon A_3$ which are contained in S'. A second-phase sample is drawn from S' with some probability scheme to observe $f_1(\alpha_1)$ and $f_2(\alpha_2)$. The estimators for the two-phase sample can easily be defined on the similar lines.

These estimators like usual ratio and product estimators are biased. The bias and mean square error can be obtained by considering the Taylor's expansion and considering only terms upto the second degree. Let us write

$$\hat{F}_i = F_i(1+\epsilon_i) \quad \ldots \quad (2.7)$$

such that $E(\epsilon_i) = 0$ and $|\epsilon_i < 1|$ for $i = 1, 2, 3$. Let us denote $B(\hat{R})$, $B(R_1^*)$, $B(R_2^*)$, $B(\hat{P})$, $B(P_1^*)$, $B(P_2^*)$ as the bias and $M(\hat{R})$, $M(R_1^*)$, $M(R_2^*)$, $M(\hat{P})$, $M(P_1^*)$ and $M(P_2^*)$ as the mean square error of the corresponding estimators. We get expressions for biases as

$$B(\hat{R}) = R\left[\frac{V(\hat{F}_2)}{F_2^2} - \frac{\text{cov}(\hat{F}_1, \hat{F}_2)}{F_1F_2}\right] \quad \ldots \quad (2.8)$$

$$B(R_1^*) = B(\hat{R}) + R\left[\frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} - \frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.9)$$

$$B(R_2^*) = B(\hat{R}) + R\left[\frac{V(\hat{F}_3)}{F_3^2} - \frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} + \frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.10)$$

$$B(\hat{P}) = \text{cov}(\hat{F}_1, \hat{F}_2) \quad \ldots \quad (2.11)$$

$$B(P_1^*) = B(\hat{P}) + P\left[\frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} + \frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.12)$$

$$B(P_2^*) = B(\hat{P}) + P\left[\frac{V(\hat{F}_3)}{F_3^2} - \frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} - \frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.13)$$

and mean square error as

$$M(\hat{R}) = R^2\left[\frac{V(\hat{F}_1)}{F_1^2} + \frac{V(\hat{F}_2)}{F_2^2} - \frac{2\,\text{cov}(\hat{F}_1, \hat{F}_2)}{F_1F_2}\right] \quad \ldots \quad (2.14)$$

$$M(R_1^*) = M(\hat{R}) + R^2\left[\frac{V(\hat{F}_3)}{F_3^2} + 2\cdot\frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} - 2\,\frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.15)$$

$$M(R_2^*) = M(\hat{R}) + R^2\left[\frac{V(\hat{F}_3)}{F_3^2} - 2\,\frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} + 2\,\frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.16)$$

$$M(\hat{P}) = P^2\left[\frac{V(\hat{F}_1)}{F_1^2} + \frac{V(\hat{F}_2)}{F_2^2} + 2\,\frac{\text{cov}(\hat{F}_1, \hat{F}_2)}{F_1F_2}\right] \quad \ldots \quad (2.17)$$

$$M(P_1^*) = M(\hat{P}) + P^2\left[\frac{V(\hat{F}_3)}{F_3^2} + 2\,\frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} + 2\,\frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.18)$$

$$M(P_2^*) = M(\hat{P}) + P^2\left[\frac{V(\hat{F}_3)}{F_3^2} - 2\,\frac{\text{cov}(\hat{F}_1, \hat{F}_3)}{F_1F_3} - 2\,\frac{\text{cov}(\hat{F}_2, \hat{F}_3)}{F_2F_3}\right] \quad \ldots \quad (2.19)$$

Proceeding on the lines given by Murthy (1963) the bias of these estimators can be estimated using interpenetrating sub-sample estimates and using these estimates of bias the estimators can be made almost unbiased.

3. COMPARISON OF THE ESTIMATORS

Comparing the mean square of error of the estimators R_1^* and R_2^* with that of $\hat{R}$ it will be observed that R_1^* and R_2^* would be more efficient than $\hat{R}$ if the following conditions (3.1) and (3.2) respectively are satisfied, that is if

$$\rho_{13}(C_1/C_3)-\rho_{23}(C_2/C_3) < -\frac{1}{2} \qquad \ldots \quad (3.1)$$

and

$$\rho_{13}(C_1/C_3)-\rho_{23}(C_2/C_3) > \frac{1}{2} \qquad \ldots \quad (3.2)$$

where $C_i^2 = V(\hat{F}_i)/F_i^2$ is the square of coefficient of variation of $\hat{F}_i$, ρ_{ij} $(i \neq j = 1, 2, 3)$ is the correlation coefficient between F_i and F_j; and $\text{cov}(\hat{F}_i, \hat{F}_j)/F_i F_j = C_i C_j \rho_{ij}$.

Similarly comparing the mean square errors of P_1^* and P_2^* with that of $\hat{P}$ we find that P_1^* and P_2^* would be more efficient than $\hat{P}$ if the following conditions (3.3) and (3.4) respectively are satisfied, that is if

$$\rho_{13}(C_1/C_3)+\rho_{23}(C_2/C_3) < -\frac{1}{2} \qquad \ldots \quad (3.3)$$

and

$$\rho_{13}(C_1/C_3)+\rho_{23}(C_2/C_3) > \frac{1}{2} \qquad \ldots \quad (3.4)$$

It may be noted here that these conditions as such do not depend on the value of ρ_{12} but they do depend on the sign of the parameters F_1, F_2 and F_3; and therefore the conditions get changed according to the changes in the sign of the parameters. The above conditions have been derived assuming that F_1, F_2 and F_3 are all positive or all the three are negative.

4. CONFIGURATIONAL REPRESENTATION OF THE CONDITIONS

For simplicity and better understanding of the regions of preference of the proposed estimators, let us assume

$$C_1 = C_2 = C_3 = C. \qquad \ldots \quad (4.1)$$

Now under this assumption the conditions (3.1) and (3.2) for R_1^* and R_2^* to be more efficient than $\hat{R}$ can much better be illustrated with the help of configuration SR given in Figure 1.

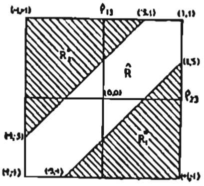

Fig. 1.

From the above configuration it is clear that whenever it is possible to choose a supplementary character $f_3(x_3)$ for all $\alpha_3 \in A_3$ such that the pair (ρ_{13}, ρ_{23}) lies in either of the regions R_1^* and R_2^*, then under the assumption (4.1), the corresponding estimator improves over the usual ratio estimator $\hat{R}$, irrespective of the correlation between $f_1(x_1)$ and $f_2(\alpha_2)$. It would be observed that the relative increase in precision is appreciably high when $f_1(x_1)$ and $f_2(x_2)$ are negatively correlated. As an example we shall consider the two extreme situations

$$\text{(i)} \quad \rho_{12} = -1.0,\ \rho_{13} = -1.0 \text{ and } \rho_{23} = +1.0.$$

In this case the point (ρ_{13}, ρ_{23}) lies in R_1^* and we get

$$B(\hat{R}) = 2RC^2,\ B(R_1^*) = 0$$

$$M(\hat{R}) = 4\cdot(R\,C)^2 \text{ and } M(R_1^*) = \frac{M(\hat{R})}{4}.$$

$$\text{(ii)} \quad \rho_{12} = -1.0,\ \rho_{13} = +1.0,\ \rho_{23} = -1.0$$

in which case we prefer R_2^* and we get,

$$B(\hat{R}) = 2RC^2,\ B(R_2^*) = \frac{B(\hat{R})}{2} \text{ and } M(R_2^*) = \frac{M(\hat{R})}{4}.$$

It may be noted that for $\rho_{12} = +1.0$, $\hat{R}$ is always superior to both R_1^* and R_2^*, that is in that case the configuration SR is never realised.

Similarly the set of conditions (3.3) and (3.4) under the assumption (4.1) can much better be effected by the configuration Sp given in Figure 2 below.

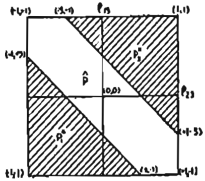

Fig. 2.

Again from the above configuration it is evident that whenever it is possible to get $f_3(\alpha_3)$ such that the pair (ρ_{13}, ρ_{23}) lies in either of the regions P_1^* and P_2^*, the corresponding estimator improves over the usual product estimator. It can easily be verified that the gain in precision is comparatively higher when $f_1(a_1)$ and $f_2(a_2)$ are found to be positively correlated. For example consider the cases

(i) $\rho_{12} = +1.0$, $\rho_{13} = +1.0$ and $\rho_{23} = +1.0$

(ii) $\rho_{12} = +1.0$, $\rho_{13} = -1.0$ and $\rho_{23} = -1.0$.

In the first case $M(P_2^*) = \dfrac{M(\hat{P})}{4}$ with $B(P_2^*) = 0$ and in the second case $M(P_1^*) = \dfrac{M(\hat{P})}{4}$ with $B(P_1^*) = 0$. It may be noted that when $\rho_{12} = -1.0$, usual product estimator is always superior than P_1^* and P_2^*.

5. An example

The data for all 61 blocks of Ahmedabad city Ward No.1 (*Khadia* I) taken from 1961 population census have been considered for the purpose of this study. It is intended to determine the proportion (R) of 'total females employed (F_1)' to the 'total female population (F_2)'. The supplementary characteristic chosen for this purpose is the 'females in services (F_3) (*group IX of population census*). For this population we find that

$$F_1 = 455 \quad \bar{F}_1 = 7.46 \quad C_1'^2 = 0.5046 \quad \rho_{12} = 0.0388$$

$$F_2 = 19198 \quad \bar{F}_2 = 205.54 \quad C_2'^2 = 0.0379 \quad \rho_{13} = 0.7737$$

$$F_3 = 324 \quad \bar{F}_3 = 5.31 \quad C_3'^2 = 0.5737 \quad \rho_{23} = -0.0474$$

where $\bar{F}_i = F_i/61$, denotes the average population of the blocks for the corresponding characteristics under consideration and $C_i^2 = kC_i'^2$, where $C_i'^2$ stands for square of the coefficient of variation for the i characteristics and k is a constant given by $\dfrac{N-n}{(N-1)n}$ where N and n are respectively the number of blocks in the population and sample. It is assumed here that the sample is drawn with equal probability without replacement.

Obviously the estimator R_2^* is to be preferred in this case since for this example the condition (3.2) is being satisfied. Substituting these values of C_i^2 and ρ_{ij} in equations (2.14) and (2.16), we get,

$$M(\hat{R}) = k'(0.5318)$$

and

$$M(R_2^*) = k'(0.2542)$$

which gives the relative gain in precision of R_2^* as 209.2%. It is to be noted here that this relative precision is independent of sample size since $k'(= k\,R^2)$ is constant for both the estimation procedures. It is also observed that estimator R_1^* is to be preferred when the ratio of interest is (F_2/F_1) and $\hat{R}$ is to be preferred in case it is intended to find the ratio R as (F_3/F_1).

6. Use of multi-supplementary variables

Let us suppose that in some surveys information on $(p+2)$ real-valued set function $f_i(\alpha_i)$ defined over the class $A_i(i = 1, 2, \ldots, p+2)$ are available and we are interested in estimation of either ratio $R(= F_1/F_2)$ or product $P(= F_1 \cdot F_2)$. The parameter F_i is of the form (2.1) that is

$$F_i = \sum_{\alpha_i \in A_i} f_i(\alpha_i) \qquad \text{for } i = 1, 2, \ldots (p+2).$$

We can improve the usual ratio and product estimators by considering the estimators R_p^* and P_p^* defined by

$$R_p^* = \hat{R} \cdot \prod_{i=3}^{q+2} F_i^* \prod_{i=q+3}^{p+2} (F_i^*)^{-1} \qquad \ldots \quad (6.1)$$

and

$$P_p^* = \hat{P} \prod_{i=3}^{q+2} F_i^* \prod_{i=q+3}^{p+2} (F_i^*)^{-1} \qquad \ldots \quad (6.2)$$

where $F_i^* = \hat{F}_i/F_i$, provided the values of $F_i(i = 3, 4, \ldots, p+2)$ are exactly known.

The mean square error (m.s.e.) of these estimators considering only the terms upto second degree are obtained by similar approach given in Section 2. The m.s.e. are

$$M(R_p^*) = M(\hat{R})+R^2\Big[M_p-2\sum_{i=3}^{q+2} C_2C_i\rho_{2i}+2\sum_{j=q+3}^{p+2} C_2C_j\rho_{2j}\Big] \qquad \ldots \quad (6.3)$$

and
$$M(P^*) = M(\hat{P})+P^2\Big[M_p+2\sum_{i=3}^{q+2} C_2C_i\rho_{2i}-2\sum_{j=q+3}^{p+2} C_2C_j\rho_{2j}\Big] \qquad \ldots \quad (6.4)$$

where
$$M_p = \sum_{i=3}^{p+2} C_i^2+2\sum_{j(<k)=3}^{q+2} C_jC_k\rho_{jk}-2\sum_{j(<k)=q+3}^{p+2} C_jC_k\rho_{jk}$$

$$+2\sum_{i=3}^{q+2} C_1C_i\rho_{1i}-2\sum_{j=q+3}^{p+2} C_1C_j\rho_{1j}-2\sum_{i<j}\sum C_iC_j\rho_{ij}. \qquad \ldots \quad (6.5)$$

The exact conditions under which R_p^* and P_p^* may be more efficient than $\hat{R}$ and $\hat{P}$ respectively, is complicated. However it is observed from equations (5.3) and (5.4) that sufficient condition for R_p^* to be more efficient than R is that $f_2(\alpha_2)$ should be positively correlated with any q of the $f_i(\alpha_i)$'s and negatively correlated with the remaining $(p-q)$ of $f_i(\alpha_i)$'s where as the condition for P_p^* to be more efficient than $\hat{P}$ get reversed. The configurational representation showing the distinct region of preference is also difficult to get; however, these conditions in particular cases can be obtained as indicated in Sections 4 and 5.

ACKNOWLEDGEMENT

The author wishes to express his gratitude to Dr. M. N. Murthy for his guidance and to Professor D. B. Lahiri for his encouragement in the preparation of this paper.

REFERENCES

GOODMAN, L. A. (1960): On the exact variance of products. *Jour. Amer. Stat. Ass.*, 55, 708-713.

HANSEN, M. H., HURWITZ, W. N. and MADOW, W. G. (1953): *Sample Survey Methods and Theory*, 1, *John Wiely and Sons*, New York.

MURTHY, M. N. (1962): Almost unbiased estimators based on interpenetrating sub-samples. *Sankhyā*, Series A, 24, 303-314.

——— (1963): Generalised unbiased estimation in sampling from finite populations. *Sankhyā*, Series B, 25, 245-262.

——— (1964): Product method of estimation. *Sankhyā*, Series A, 26, 69-74.

MURTHY, M. N., NANJAMMA, N. S. and SETHI, V. K. (1959): Some sampling systems providing unbiased ratio estimators. *Sankhyā*, 21, 299-314.

Paper received : August, 1965.

Paper revised : October, 1965.