PRODUCT METHOD OF ESTIMATION

By M. N. MURTHY

Indian Statistical Institute

SUMMARY. In this paper a method of estimation, which may be considered as the dual property of the ratio method of estimation, is suggested. The product estimator is of the form $(yx)/x$, where y and x are unbiased estimators of the parameters Y and X corresponding to the variate under consideration and the supplementary variate respectively based on any probability sample design. It is shown that an idea of the correlation coefficient between y and x and the coefficient of variation of x and y would help in determining whether an unbiased, product or ratio estimator is to be used in a particular situation. The question of estimation of the bias of the product estimator is also considered.

1. Introduction

Though the problem of improving upon an unbiased estimator by the method of ratio estimation using a suitable supplementary variate has received considerable attention in sampling theory, little attention seems to have been given to the dual situation of improving upon an unbiased estimator by using a suitable product estimator. However, Goodman (1960) has considered the question of obtaining variances and variance estimators for products of estimators. In this paper the product method of estimation is introduced in a general form. Suppose y and x are unbiased estimators of the parameters Y and X corresponding to the variate under consideration and the supplementary variate respectively and let the value of X be known. Then the product estimator is given by $(yx)/X$. This estimator is being suggested, as it is complementary to the ratio estimator $(y/x)X$ and hence is likely to be useful in situations, where the ratio estimator is not efficient, that is, where the estimators y and x are negatively correlated. The technique of obtaining almost unbiased estimators for non-linear parametric functions considered by Murthy (1962) is applied to the proposed product estimator to make it unbiased.

2. Product method of estimation

Let y and x be unbiased estimators of the parameters Y and X corresponding to the variate under consideration and the supplementary variate respectively based on any probability sample design and let the value of X be known. Then an alternative estimator of Y, which may be termed the product estimator, is given by

$$\hat{Y} = \frac{yx}{X}. \qquad \dots \quad (2.1)$$

For considering the efficiency of this estimator as compared to that of the unbiased estimator y, let us derive the expressions for the expected value and the mean square error of this estimator. Writing

$$y = Y(1+e_1) \text{ and } x = X(1+e_2), \text{ where } E(e_1) = E(e_2) = 0,$$

we get the bias and the mean square error of the product estimator $\hat{Y}$ as

$$B(\hat{Y}) = E\left(\frac{yx}{X} - Y\right)$$

$$= YE[(1+e_1)(1+e_2)-1]$$

$$= YV_{11} = Y\rho(y, x)\frac{\sigma_y\sigma_x}{YX} \quad \ldots \quad (2.2)$$

and

$$M(\hat{Y}) = E\left(\frac{yx}{X} - Y\right)^2$$

$$= Y^2E(e_1+e_2+e_1e_2)^2$$

$$= Y^2(V_{20}+2V_{11}+V_{02}+2V_{21}+2V_{12}+V_{22}) \quad \ldots \quad (2.3)$$

where

$$V_{ij} = E(e_1^i e_2^j) = \frac{E(y-Y)^i(x-X)^j}{Y^i X^j},$$

σ_y and σ_x are the standard errors of the estimators y and x respectively and $\rho(y, x)$ is the correlation coefficient between y and x. From (2.2) it can be seen that if y and x are uncorrelated, the product estimator is unbiased for Y. The variance estimators for products of estimators have been considered by Goodman (1960).

In case of sampling schemes such as simple random sampling or varying probability sampling with replacement or any other sampling scheme involving selection of independent sub-samples, the bias and the mean square error of the product estimator are of the form

$$B(\hat{Y}) = Y\frac{V'_{11}}{m} \quad \ldots \quad (2.4)$$

and

$$M(\hat{Y}) = Y^2\left[\frac{1}{m}(V'_{20}+2V'_{11}+V'_{02})+\frac{2}{m^2}(V'_{21}+V'_{12})+\frac{1}{m^3}\{V'_{22}+(m-1)V'_{20}V'_{02}\}\right] \quad \ldots \quad (2.5)$$

where V'_{ij} stands for V_{ij} given in (2.3) for a sample of one unit or for one sub-sample and m is the sample size or the number of sub-samples as the case may be. This shows that if m is large, the contribution of the bias to the mean square error would be negligible, and that the terms involving $1/m^2$ and $1/m^3$ in the mean square error may be neglected, in which case (2.5) reduces to

$$M(\hat{Y}) = \frac{Y^2}{m}(V'_{20}+2V'_{11}+V'_{02}). \quad \ldots \quad (2.6)$$

Thus we see that the bias of the product estimator (2.1) is given by

$$B(\hat{Y}) = YV_{11} = Y\rho(y,x)\frac{\sigma_y\sigma_x}{YX} = Y\rho(y,x)C_yC_x,$$

where C_y and C_x are the coefficients of variation of the estimators y and x respectively and that the mean square error correct to the second degree of approximation is

$$M(\hat{Y}) = Y^2(V_{20}+2V_{11}+V_{02}). \quad \ldots \quad (2.7)$$

In case of large samples, the bias relative to the parameter Y, that is, $B(\hat{Y})/Y$, is likely to be negligible since $|B(\hat{Y})|/Y \leqslant C_yC_x$ and this estimator will be more efficient than the unbiased estimator y if

$$M(\hat{Y}) < Y^2V_{20},$$

that is, if

$$V_{11} < -\tfrac{1}{2}V_{02},$$

which leads to the condition

$$\rho(y,x) < -\tfrac{1}{2}\frac{C_x}{C_y} \text{ in case both } Y \text{ and } X \text{ are positive or negative}$$

$$\ldots \quad (2.8)$$

and $$\rho(y,x) > +\tfrac{1}{2}\frac{C_x}{C_y} \text{ in case either } Y \text{ or } X \text{ is negative.}$$

In deriving this result, it is assumed that both Y and X are non-zero.

This result is of interest as it shows that for any given supplementary variate, one can decide whether to use the usual unbiased, ratio or product estimator on the basis of the correlation coefficient between the estimators y and x of the parameters corresponding to the variate under consideration and the supplementary variate respectively and of the correlation coefficients of y and x provided the sample size is large. In this connection it may be noted that the ratio estimator $(y/x)X$ is more efficient than the usual unbiased estimator y in case of large samples if $\rho(y, x) > +(C_y/2C_x)$ in case both Y and X are positive or negative and $\rho(y, x) < -(C_y/2C_x)$ in case either Y or X is negative. Hence the estimator to be used in a particular situation is the

$$\begin{aligned} &\text{product estimator} \quad \frac{yx}{X}, \quad \text{if} \quad -1 \leqslant \rho(y, x) < -\tfrac{1}{2}\frac{C_x}{C_y}, \\ &\text{unbiased estimator} \quad y, \quad \text{if} \quad -\tfrac{1}{2}\frac{C_x}{C_y} \leqslant \rho(y, x) \leqslant +\tfrac{1}{2}\frac{C_x}{C_y}, \qquad \ldots \quad (2.9) \\ &\text{ratio estimator} \quad \frac{y}{x}X, \quad \text{if} \quad +\tfrac{1}{2}\frac{C_x}{C_y} < \rho(y, x) \leqslant +1 \end{aligned}$$

in case both Y and X are positive or negative. If either Y or X is negative, the conditions for the product and the ratio estimators to be more efficient than the usual unbiased estimator will be reversed. That is, if either Y or X is negative, then the estimator to be used is the

$$\begin{aligned} &\text{product estimator} \quad \frac{yx}{X}, \quad \text{if} \quad +\tfrac{1}{2}\frac{C_x}{C_y} < \rho(y, x) \leqslant +1, \\ &\text{unbiased estimator} \quad y, \quad \text{if} \quad -\tfrac{1}{2}\frac{C_x}{C_y} \leqslant \rho(y, x) \leqslant +\tfrac{1}{2}\frac{C_x}{C_y}, \qquad \ldots \quad (2.10) \\ &\text{ratio estimator} \quad \frac{y}{x}X, \quad \text{if} \quad -1 \leqslant \rho(y, x) < -\tfrac{1}{2}\frac{C_x}{C_y}. \end{aligned}$$

Since the whole range of the correlation coefficient between the estimators y and x has been covered, this procedure will lead to better utilization of the available supplementary information in improving upon the ordinary unbiased estimator. Incidentally it may be noted that the expressions for the bias and the mean square error of the product estimators are exact unlike those of the ratio estimator and that their derivation does not require the rather restrictive assumptions such as x being non-zero, $\left|\frac{x-X}{X}\right| < 1$, etc. involved in the derivation of the corresponding expressions for the ratio estimator.

3. Unbiased product estimator

Suppose in a survey the sample is drawn in the form of n independent interpenetrating sub-samples. Let y_i and x_i be unbiased estimates of the parameters Y and X based on the i-th sub-sample, $(i = 1, 2, \ldots, n)$. The following two product estimators can be considered to estimate the parameter Y :

$$\hat{Y}_1 = \frac{\bar{y}\bar{x}}{X} = \frac{1}{X}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \quad \ldots \quad (3.1)$$

and

$$\hat{Y}_2 = \frac{1}{X}\left(\frac{1}{n}\sum_{i=1}^{n} y_i x_i\right). \quad \ldots \quad (3.2)$$

Applying result (2.2) to (3.1), we get

$$B(\hat{Y}_1) = Y\bar{V}_{11}$$

where $\bar{V}_{11} = E(\bar{y}-Y)(\bar{x}-X)/YX = \frac{1}{n^2}\sum_{i=1}^{n} E(y_i-Y)(x_i-X)/YX,$

since y_i and $x_j\,(j \neq i)$ are uncorrelated. Hence we have

$$B(\hat{Y}_1) = \frac{1}{n^2}\sum_{i=1}^{n} B\left(\frac{y_i x_i}{X}\right). \quad \ldots \quad (3.3)$$

The bias of $\hat{Y}_2$ is given by

$$B(\hat{Y}_2) = \frac{1}{n}\sum_{i=1}^{n} B\left(\frac{y_i x_i}{X}\right). \quad \ldots \quad (3.4)$$

Comparing (3.3) and (3.4), we see that the bias of $\hat{Y}_2$ is n times the bias of $\hat{Y}_1$, that is,

$$B(\hat{Y}_2) = nB(\hat{Y}_1). \quad \ldots \quad (3.5)$$

Since $E(\hat{Y}_2-\hat{Y}_1) = B(\hat{Y}_2)-B(\hat{Y}_1) = (n-1)\,B(\hat{Y}_1)$, an unbiased estimator of the bias of $\hat{Y}_1$ is given by

$$\hat{B}(\hat{Y}_1) = \frac{\hat{Y}_2-\hat{Y}_1}{n-1}. \quad \ldots \quad (3.6)$$

This unbiased estimator of the bias of the product estimator may be used to correct the estimator $\hat{Y}_1$ for its bias, thereby obtaining an unbiased product estimator $\hat{Y}_3$ given by

$$\hat{Y}_3 = \frac{n\hat{Y}_1 - \hat{Y}_2}{n-1}. \qquad \ldots \quad (3.7)$$

The conditions under which the unbiased product estimator $\hat{Y}_3$ is more efficient than the biased product estimator $\hat{Y}_1$ are the same as those given by Murthy and Nanjamma (1959) in case of obtaining an almost unbiased ratio estimator.

References

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

Murthy, M. N. and Nanjamma, N. S. (1959): Almost unbiased ratio estimates based on interpenetrating sub-sample estimates. *Sankhyā*, 21, 381-392.

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

Paper received : November, 1962.
Revised : October, 1963.