ON THE UTILIZATION OF STOCHASTIC A PRIORI INFORMATION IN ESTIMATION

P.C. SAXENA*

Banaras Hindu University, Varanasi-5

(*Received in September, 1971*)

1. INTRODUCTION

Recently Singh and Roy [3] proposed a procedure for estimating a population parameter when its a priori value, quite close to the true one, is known. Usually this a priori value is derived from earlier empirical investigations or from other samples, and in such situations it may be regarded as a random variable. For instance the a priori information, obtained from pilot studies or respective surveys, may specify the estimate of the population parameter together with its standard error (s.e). Quite often the a priori information may simply consist of the probable range of the parameter, which may be based on past experience or some theoretical consideration. For example, it may be expected that the value of the parameter may lie, with almost certainty, between 2 and 10. Dalenius [1] recommended the use of a simple or modified average, depending upon the expected skewness of the distribution of the parameter of these limits as true non-stochastic a priori value. We, however, treat them as, say, two-sigma limits estimate $\pm$ 2 s.e. (which in this case is $6\pm2\times2$) so that the estimated value of the parameter is 6 with a s.e. equal to 2. Under such specifications, an estimation procedure is proposed for combining the stochastic a priori information with the sample information and its efficiency is examined. The a priori value which is taken as a random variable is assumed to be independent of the sample characteristic under consideration.

2. THE RESULT

Let $\hat{\theta}_a$ be the estimated value, known a priori, of the parameter θ, and $\hat{\theta}_s$ be the estimator based on the sample. Then the following estimator combining $\hat{\theta}_a$ and $\hat{\theta}_s$ is proposed :

$$\tilde{\theta}_c = K\hat{\theta}_s + (1-K)\hat{\theta}_a \qquad ...(1)$$

where k is a scalar.

*Presently the author is Reader at the International Institute for Population Studies, Bombay-400088.

The bias of the estimator $\tilde{\theta}_c$ is

$$B(\tilde{\theta}_c)=KB(\hat{\theta}_s)+(1-K)B(\hat{\theta}_a) \quad ...(2)$$

and the mean squared error is

$$M(\tilde{\theta}_c)=K^2 M(\hat{\theta}_s)+(1-K)^2 M(\hat{\theta}_a) \quad ...(3)$$

The scalar K is so determined as to render $M(\tilde{\theta}_c)$ a minimum.

Differentiating $M(\tilde{\theta}_c)$ with respect to K and equating to zero we get

$$K=\frac{M(\hat{\theta}_a)}{M(\hat{\theta}_a)+M(\hat{\theta}_s)}\equiv K_{min} \text{ (say)} \quad ...(4)$$

It is easy to show that this value of K minimizes $M(\tilde{\theta}_c)$ and in that case the minimum value of $M(\tilde{\theta}_c)$ is

$$M_{min}(\tilde{\theta}_c)=\frac{M(\hat{\theta}_s)M(\hat{\theta}_a)}{M(\hat{\theta}_s)+M(\hat{\theta}_a)}=K_{min}\, M(\hat{\theta}_s) \quad ...(5)$$

which clearly demonstrates that $\tilde{\theta}_c$ is more efficient than $\hat{\theta}_s$.

If we choose

$$K=\frac{B(\hat{\theta}_a)}{B(\hat{\theta}_v)-B(\hat{\theta}_s)} \quad ...(6)$$

then $\tilde{\theta}_c$ becomes unbiased, but does not make $M(\tilde{\theta}_c)$ a minimum.

Singh and Roy [3] considered the estimator

$$\hat{\theta}_c=K\hat{\theta}_s+(1-K)\theta_a^* \quad ...(7)$$

where θ_v^* is a fixed given value of θ.

The expressions for bias and mean squared error of $\hat{\theta}_c$ are

$$B(\hat{\theta}_c)=KB(\hat{\theta}_s)+(1-K)\triangle \quad ...(8)$$

$$M(\hat{\theta}_c)=K^2M(\hat{\theta}_s)+(1-K)^2\triangle^2+2K(1-K)\triangle B(\hat{\theta}_s) \quad ...(9)$$

where

$$\triangle=\overset{*}{\theta}_a-\theta \quad ...(10)$$

The value of K that minimizes $M(\hat{\theta}_c)$ is

$$K'_{min}=\frac{\triangle[\triangle-B(\hat{\theta}_s)]}{[M(\hat{\theta}_s)-B^2(\hat{\theta}_s)]+[\triangle-B(\hat{\theta}_s)]^2} \quad ..(11)$$

which when substituted in (9) gives

$$M_{min}(\hat{\theta}_c)=\frac{\triangle^2[M(\hat{\theta}_s)-B^2(\hat{\theta}_s)]}{[M(\hat{\theta}_s)-B^2(\hat{\theta}_s)]+[\triangle-B(\hat{\theta}_s)]^2} \quad ...(12)$$

However, an elegant and explicit condition regarding the efficiency of $\tilde{\theta}_c$ as compared to $\hat{\theta}_c$ cannot be derived from the expressions of bias and mean squared error.

Let us consider a special case in which $\hat{\theta}_a$ and $\hat{\theta}_s$ are assumed to be unbiased estimators of θ. Then we have,

$$B(\tilde{\theta}_c)=0 \quad ...(13)$$

$$B(\hat{\theta}_c)=(1-K)\triangle \quad ...(14)$$

$$M(\tilde{\theta}_c)=V(\tilde{\theta}_c)=K^2V(\hat{\theta}_s)+(1-K)^2V(\hat{\theta}_a) \quad ..(15)$$

$$M(\hat{\theta}_c)=K^2V(\hat{\theta}_s)+(1-K)^2\triangle^2 \quad ...(16)$$

Thus from (15) and (16) the estimator $\tilde{\theta}_c$ will be more effiicient then $\hat{\theta}_c$ if

$$\triangle^2\geqslant V(\hat{\theta}_a) \quad ...(17)$$

or

$$(\overset{*}{\theta}_a-\theta)^2\geqslant E(\hat{\theta}_a-\theta)^2$$

It may be added that in practice the calculation of K min or K' min, as given by (5) and (11) respectively, is not possible, as they involve population values. However, when approximate values of $\frac{M(\hat{\theta}_s)}{\theta^2}$ and $\frac{M(\hat{\theta}_a)}{\theta^2}$ are known, say C_s and C_s we may take

$$K=\frac{C_a}{C_a+C_s} \qquad \ldots(18)$$

and then the mean squared error of $\tilde{\theta}_c$ is

$$\left(\frac{C_a}{C_a+C_s}\right)^2 M(\hat{\theta}_s)+\left(\frac{C_s}{C_a+C_s}\right)^2 M(\hat{\theta}_a) \qquad \ldots(19)$$

which is smaller than the mean squared error of $\hat{\theta}_s$ if

$$\frac{M(\hat{\theta}_a)}{M(\hat{\theta}_s)}<\left(1+2\frac{C_a}{C_s}\right) \qquad \ldots(20)$$

This condition is quite commonly achieved in actual practice.

Acknowledgement

The author is thankful to Mr. V.K. Srivastava for his help in preparing this note.

References

1. Dalenius, T. (1965) : Current trends in the development of Sample Survey Theory and Methods.
2. Searls, Donald, T. (1964) : The Utilization of a known coefficient of variation in the estimation procedure. Jour. Amer. Stat. Asson.' 59. pp. 1225-6.
3. Singh, M.P. & Roy, A.S. (1970) : An Estimator using a priori value of the Parameter in Survey Sampling, Jour. Ind. Soc. Agr. Stat. Vol. XXII, No. 2; pp. 11-20.