Jour. Ind. Soc. Ag. Statistics
50(1), 1997 : 11–17

An Approach for Almost Separation of Bias Precipitates

Housila P. Singh and Raj K. Gangele
Vikram University, Ujjain, M.P.
(Received : September, 1994)

SUMMARY

The present investigation deals with the problem of estimating population variance using known coefficient of variation. A funnel associated with a filter-paper to filter the bias precipitates appearing in the estimators of population variance is proposed. An empirical study is also given.

Key words: Population variance, Linear variety, Asymptotic mean square error, Bias.

1. Introduction

In many situations, the problem of estimating population variance σ^2 assumes importance but has not received much attention. However a few authors have paid their attention towards this problem using prior information on some parameters such as coefficient of variation, skewness, kurtosis of the character under study e.g. Singh *et al.* [5], Lee [2], Singh [6], Searls and Intarapanich [4] and Gangele [1].

Motivated by Singh and Singh [7] we suggest a technique to extract the bias precipitates from the estimators of variance by using a funnel connected with a filter-paper. The apparatus consists of linear variety of estimators and linear restriction. We shall observe that the reactants (statistical constants) used for bias filtration depend on the shape parameters $\sqrt{\beta_1}$ and β_2. However in case of normal parent statistical constants or chemicals depend only on coefficient of variation which is somewhat of a simple type of apriori information.

2. Notations and Expectations

Let $y_1, y_2, ..., y_n$ be a random sample of size n drawn from a population with unknown mean μ ($\neq$ 0) and unknown variance $\sigma^2 > 0$. For the sake of simplicity we assume that population under investigation is infinite. Let

$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$ and $s^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{(n-1)}$ be unbiased estimators of μ and σ^2 respectively. We write

$$e_0 = \frac{(\bar{y} - \mu)}{\mu}, \quad e_1 = \frac{(s^2 - \sigma^2)}{\sigma^2}$$

such that

$$E(e_0) = E(e_1) = 0$$

and

$$E(e_0^2) = \frac{C^2}{n}, \quad E(e_1^2) = \frac{(\beta_2 - 1)}{n} \text{ and } E(e_0 e_1) = \frac{\sqrt{\beta_1}\, C}{n}$$

where $C = (\sqrt{\mu_2}/\mu) = \frac{\sigma}{\mu}$, $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, $\beta_2 = \frac{\mu_4}{\mu_2^2}$ and μ_r, ($r = 2, 3, 4$) is the r-th central moment.

3. *Linear Variety*

When the population coefficient of variation 'C' is known, we consider $d_1 = s^2$, $d_2 = s^2(\hat{C}/C)$ and $d_3 = s^2(C/\hat{C})$ such that $d_i \in D$, ($i = 1, 2, 3$) where D denotes the set of all possible estimators for estimating the population variance σ^2 and $\hat{C}(= s/\bar{y})$ is the consistent estimate of coefficient of variation 'C'. By definition, the set D will be a linear variety if

$$d_g = \sum_{i=1}^{3} g_i d_i \in D \tag{3.1}$$

for

$$\sum_{i=1}^{3} g_i = 1 \tag{3.2}$$

and

$$g_i \in R; \quad (i = 1, 2, 3)$$

where g_i ($i = 1, 2, 3$) denotes the amount of the reactants used for bias precipitates' filtration and R stands for the set of real numbers.

4. *Mean Square Error*

Using (3.2), the relation (3.1) in terms of e_0 and e_1 may be written as

$$d_g = \sigma^2[1 + e_1 + (g_2 - g_3)(\frac{1}{2}e_1 - e_0)] + O(e^2) \tag{4.1}$$

Let us choose

$$g_2 - g_3 = g \text{ (say, another constant)} \tag{4.2}$$

Then

$$d_g = \sigma^2[1 + e_1 + g(\frac{1}{2}e_1 - e_0)] + O(e^2)$$

or

$$d_g - \sigma^2 = \sigma^2[e_1 + g(\frac{1}{2}e_1 - e_0)] + O(e^2) \tag{4.3}$$

squaring both sides of (4.3), retaining terms upto second powers of e's and then taking expectation, we have the mean square error (MSE) of d_g to the first degree of approximation,

$$MSE(d_g) = \left(\frac{\sigma^4}{4n}\right)\left[4(\beta_2 - 1) + g^2(4C^2 - 4\sqrt{\beta_1}C + \beta_2 - 1) + 4g(\beta_2 - 2\sqrt{\beta_1}C - 1)\right] \tag{4.4}$$

which is minimized for

$$g = -\frac{2(\beta_2 - 2\sqrt{\beta_1}\,C - 1)}{(4C^2 - 4\sqrt{\beta_1}C + \beta_2 - 1)} = g_0 \text{ (say)} \tag{4.5}$$

Subtracting (4.5) in (4.4) we obtain the minimum MSE of d_g as

$$\text{min.MSE}(d_g) = \left(\frac{4\sigma^4}{n}\right)\frac{(\beta_2 - \beta_1 - 1)C^2}{(4C^2 - 4\sqrt{\beta_1}C + \beta_2 - 1)} \tag{4.6}$$

or

$$\text{min.MSE}(d_g) = \left(\frac{4\sigma^4}{n}\right)\frac{(\beta_2 - \beta_1 - 1)C^2}{[(\beta_2 - \beta_1 - 1) + (2C - \sqrt{\beta_1})^2]} \tag{4.7}$$

5. *Funnel for Estimators*

From (2.2), (4.2) and (4.5), we have

$$\sum_{i=1}^{3} g_i = 1 \tag{5.1}$$

and

$$g_2 - g_3 = g_0 \tag{5.2}$$

where g_0 is given in (4.5).

From (5.1) and (5.2), we have three unknowns to be determined from only two equations. It is, therefore, not possible to find out unique values for the amount of reactants g_i ($i = 1, 2, 3$) to be used for filtration. So our funnel is handicapped and can not be used without filter-paper. In order to obtain unique values for the reactants g_i's ($i = 1, 2, 3$) we shall connect a filter-paper with the funnel by imposing a linear constraint as

$$\sum_{i=1}^{3} g_i B(d_i) = 0 \tag{5.3}$$

where $B(d_i)$ denotes the bias in the i-th ($i = 1, 2, 3$) estimator, d_i, of population variance. The expression (5.1), (5.2) and (5.3) may also be expressed as

$$\begin{bmatrix} 0 & 1 & -1 \\ 1 & 1 & 1 \\ B(d_1) & B(d_2) & B(d_3) \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} g_0 \\ 1 \\ 0 \end{bmatrix} \tag{5.4}$$

or

$$A_{3\times 3} \quad G_{3\times 1} = B_{3\times 1}$$

The solution of the system of equations gives the values of g_i's ($i = 1, 2, 3$) separating the bias precipitates within the proposed linear variety in (3.1) if $|A| \neq 0$.

Theorem: The system of equations given by (5.4) will have a unique solution if

$$\{B(d_3) + B(d_2)\} \neq 2B(d_1) \tag{5.5}$$

Proof. The proof of the theorem follows if we set $|A| \neq 0$.

6. *Bias Separation of Order* $O(n^{-1})$

We shall now outline the manner in which one can use the funnel connected with filter-paper to separate the bias precipitates of order $O(n^{-1})$ for the estimator d_g in (3.1). The biases of the estimators d_i, (i = 1, 2, 3) to the first degree of approximation, are respectively given by

$$B(d_1) = 0 \tag{6.1}$$

$$B(d_2) = \left(\frac{\sigma^2}{8n}\right)\left[8C^2 - 12\sqrt{\beta_1}C + 3(\beta_2 - 1)\right] \tag{6.2}$$

$$B(d_3) = -\left(\frac{\sigma^2}{8n}\right)\left[\beta_2 - 4\sqrt{\beta_1}C - 1\right] \tag{6.3}$$

It is verified from (6.1) to (6.3) that (5.5) is true and then solving the equation we get the unique solution as

$$g_1 = \left[1 + g_0 + \frac{g_0}{D}(\beta_2 - 4\sqrt{\beta_1}C - 1)\right] \tag{6.4}$$

$$g_2 = -\frac{g_0(\beta_2 - 4\sqrt{\beta_1}C - 1)}{2D} \tag{6.5}$$

$$g_3 = -g_0\left[1 + \frac{(\beta_2 - 4\sqrt{\beta_1}C - 1)}{2D}\right] \tag{6.6}$$

where $g_0 = -\dfrac{2(\beta_2 - 2\sqrt{\beta_1}C - 1)}{(4C^2 - 4\sqrt{\beta_1}C + \beta_2 - 1)}$ and $D = (4C^2 - 4\sqrt{\beta_1}C + \beta_2 - 1)$

Use of these g_i's (i = 1, 2, 3) filtrates the bias upto terms of order $O(n^{-1})$. The same process may be repeated by considering $B(d_i)$; (i = 2, 3) to the order $O(n^{-2})$, if the bias in d_g is to be reduced to the order $O(n^{-3})$ and so on.

6.1 Normal Distribution

For this distribution, we have

$$\beta_1 = 0 \text{ and } \beta_2 = 3 \tag{6.7}$$

Thus the expressions in (6.4), (6.5), (6.6) and (4.6) respectively reduce to:

$$g_1 = \left[1 - \frac{2}{(1+2C^2)} - \frac{2}{(1+2C^2)^2}\right] \tag{6.8}$$

$$g_2 = \frac{1}{(1+2C^2)^2} \tag{6.9}$$

$$g_3 = \left[\frac{2}{(1+2C^2)} + \frac{1}{(1+2C^2)^2}\right] \tag{6.10}$$

and

$$\text{min.Var}\,(d_g) = \left(\frac{4\sigma^4}{n}\right)\frac{C^2}{(1+2C^2)} \tag{6.11}$$

The variance of usual unbiased estimator s^2 to terms of order $O(n^{-1})$ in normal parent is given by

$$\text{Var}\,(s^2) = \frac{2\sigma^4}{n} \tag{6.12}$$

Thus the relative efficiency of d_g with respect to s^2 is

$$\text{RE}(d_g, s^2) = 1 + \frac{1}{2C^2} \tag{6.13}$$

which shows that d_g is always more efficient than usual unbiased estimator s^2.

It is to be noted that in case of normal parent, only knowledge of coefficient of variation 'C' is sufficient for the proposed estimator d_g to be more efficient than usual unbiased estimator s^2. Further, the reactants depend only on the coefficient of variation.

7. *Empirical Study*

For the purpose of illustration, we consider the example cited in Maiti ([3], p. 83).

Data under consideration has been taken from 1961 census, West Bengal, District Census Hand Book, Midnapore (Census of India, 1961). The character y denotes the village population. The required parameters are:

$$\beta_1 = 7.789,\ \beta_2 = 14.541 \text{ and } C^2 = 0.918$$

Proceeding on the same lines indicated in the article, it may easily be seen that

$$g_1 = -2.61199, \quad g_2 = 0.54882, \quad g_3 = 3.06317$$

Using these values of g_i's ($i = 1, 2, 3$), one can eliminate the bias to the order $O(n^{-1})$ respectively, in the estimator d_g at (3.1). In practice, one can use the values of β_1, β_2 and C from past data or pilot study in formulating the estimator d_g.

ACKNOWLEDGEMENT

Authors are thankful to the referee for his valuable suggestions towards the improvement of the paper.

REFERENCES

[1] Gangele, R.K., 1992. Estimation of variance and circular probable error. M.Phil. dissertation submitted to Vikram University, Ujjain, M.P., India.

[2] Lee, K.H., 1981. Estimation of variance of mean using known coefficient of variation. *Comm. Statist.—Theory Methods*, A(10) 5, 503-514.

[3] Maiti, P., 1982. Some contributions to the sampling theory using apriori and aposteriori information. Ph.D. thesis submitted to Indian Statistical Institute, Calcutta, India.

[4] Searls, D.T. and Intarapanich, P., 1990. A note on an estimator for the variance that utilizes the kurtosis. *Amer. Statist.*, 44(4), 295-296.

[5] Singh, J., Pandey, B.N. and Hirano, K., 1973. On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. *Ann. Inst. Statist. Math.*, 25, 51.

[6] Singh, H.P., 1984. Use of auxiliary information in estimation procedures for some population parameters. Ph. D. thesis submitted to Indian School of Mines, Dhanbad, Bihar, India.

[7] Singh, S. and Singh, R., 1993. A new method: Almost separation of bias precipitates in sample surveys. *Jour. Ind. Stat. Assoc.*, 31, 99-105.