ON THE UTILIZATION OF A KNOWN COEFFICIENT OF KURTOSIS IN THE ESTIMATION PROCEDURE OF VARIANCE

J. SINGH, B. N. PANDEY AND K. HIRANO

(Received June 28, 1971; revised Nov. 1, 1972)

1. Summary

Let μ and σ^2 be the unknown mean and variance of a population distribution, respectively. An estimator Y^* of σ^2 is developed which utilizes a priori information concerning the value of the population coefficient of kurtosis β_2. The estimator Y^* is shown to have a smaller mean-squared error than the usual unbiased estimator s^2 (that is, the sample variance). Furthermore in Table 1 we give the relative efficiencies of Y^* with respect to s^2, and in Table 2 the ranges of α that an another estimator $\tilde{Y}$, which uses the approximate value $\beta_2\alpha$ $(\alpha>0)$ instead of β_2 in the estimator Y^*, is more precise than s^2.

2. Introduction

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population with the unknown mean μ and variance σ^2. Searls [1] proposed an estimator $\bar{X}'$ of the population mean, deviding the sample total $\sum_{i=1}^{n} X_i$ by a scalar which is determined by minimizing the mean-squared error. Such an estimator $\bar{X}'$ is although biased, has smaller mean-squared error than the usual unbiased estimator $\bar{X}$ (that is, the sample mean). The value of the scalar depends upon the population coefficient of variation $v_0=\sigma/\mu$. The estimator $\bar{X}'$ may have its utility in those situations where an approximate or guessed value of the population coefficient of variation is available. Hirano [2] gave the relative efficiency of an estimator $\tilde{X}$, which uses $v=v_0\alpha$ instead of v_0 in $\bar{X}'$, with respect to $\bar{X}$, and the range of α that the estimator $\tilde{X}$ is more precise than $\bar{X}$.

In this paper we give an estimator Y^* of the population variance using a priori information about the population coefficient of kurtosis β_2, and discuss its properties in relation to the usual unbiased estimator s^2. Also we present the table of the relative efficiency of the estimator

$\tilde{Y}$, which uses the approximate value $\beta_2\alpha$ $(\alpha>0)$ instead of β_2 in the estimator Y^*, with respect to s^2 and furthermore the range of α that the estimator $\tilde{Y}$ is more precise than s^2.

3. Estimation that utilizes an exact value β_2

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population with the unknown mean μ and variance σ^2. Consider the estimator of σ^2

$$Y=M\sum_{i=1}^{n}(X_i-\bar{X})^2 \tag{1}$$

$$=M(n-1)s^2 \tag{2}$$

where $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ and M is a scalar.

We determine the scalar M to minimize the mean-squared error of Y. We have

$$\text{MSE}(Y)=\text{Var}(Y)+(\text{Bias } Y)^2 \tag{3}$$

where MSE (Y) is the mean-squared error of Y. It is well known that

$$E(s^2)=\sigma^2 \tag{4}$$

and

$$\text{Var}(s^2)=\frac{\mu_4}{n}+\frac{(3-n)\sigma^4}{n(n-1)} \tag{5}$$

where μ_4 is the fourth central moment. From (2) and (5)

$$\text{Var}(Y)=M^2(n-1)^2\,\text{Var}(s^2)=M^2(n-1)^2\left[\frac{\mu_4}{n}+\frac{(3-n)\sigma^4}{n(n-1)}\right]. \tag{6}$$

Now, in view of (4),

$$\text{Bias } Y=E(Y)-\sigma^2=M(n-1)\sigma^2-\sigma^2 .$$

Hence

$$(\text{Bias } Y)^2=\sigma^4[1-M(n-1)]^2 . \tag{7}$$

Therefore from (3), (6) and (7) we have

$$\text{MSE}(Y)=M^2(n-1)^2\left[\frac{\mu_4}{n}+\frac{(3-n)\sigma^4}{n(n-1)}\right]+\sigma^4[1-M(n-1)]^2 . \tag{8}$$

Differentiating (8) with respect to M and equating to zero we get

$$2M(n-1)^2\left[\frac{\mu_4}{n}+\frac{(3-n)\sigma^4}{n(n-1)}\right]-2\sigma^4(n-1)[1-M(n-1)]=0 . \tag{9}$$

From (9) we get

$$M=\frac{n}{n^2-2n+3+\beta_2(n-1)} \tag{10}$$

where $\beta_2=\mu_4/\sigma^4$. Further we have

$$\frac{\partial^2}{\partial M^2}\,\mathrm{MSE}\,(Y)=\frac{2(n-1)}{n}\sigma^4[\beta_2(n-1)+n^2-2n+3]\ . \tag{11}$$

If $n\geqq 2$, then all the terms in (11) are positive. We suppose that $n\geqq 2$ throughout in this paper. Hence $(\partial^2/\partial M^2)\,\mathrm{MSE}\,(Y)$ is always positive. Consequently the value of the scalar M which minimizes $\mathrm{MSE}\,(Y)$ is given by (10).

Here we define as follow the estimator Y^* with the scalar M given by (10) that minimizes $\mathrm{MSE}\,(Y)$;

$$Y^*=\frac{n}{\beta_2(n-1)+n^2-2n+3}\sum_{i=1}^{n}(X_i-\bar{X})^2\ . \tag{12}$$

Equations (8) and (10) give

$$\mathrm{MSE}\,(Y^*)=\frac{\sigma^4[\beta_2(n-1)+3-n]}{n^2-2n+3+\beta_2(n-1)}\ .$$

Since s^2 is an unbiased estimator of σ^2, we have

$$\mathrm{MSE}\,(s^2)=\mathrm{Var}\,(s^2)=\frac{\sigma^4}{n(n-1)}[\beta_2(n-1)+3-n]\ .$$

Hence the relative efficiency of Y^* with respect to s^2 is

$$\mathrm{REF}\,(Y^*)=\frac{\mathrm{MSE}\,(s^2)}{\mathrm{MSE}\,(Y^*)}=\frac{n^2-2n+3+\beta_2(n-1)}{n(n-1)}\ .$$

For different values of n and β_2 the relative efficiencies $\mathrm{REF}\,(Y^*)$ (%) are presented in Table 1.

Table 1 Relative efficiencies $\mathrm{REF}\,(Y^*)$ (%)

β_2	Sample size n						
	5	10	20	50	100	500	1000
1	110.0	102.2	100.5	100.1	100.0	100.0	100.0
2	130.0	112.2	105.5	102.1	101.0	100.2	100.1
3	150.0	122.2	110.5	104.1	102.0	100.4	100.2
4	170.0	132.2	115.5	106.1	103.0	100.6	100.3
5	190.0	142.2	120.5	108.1	104.0	100.8	100.4
6	210.0	152.2	125.5	110.1	105.0	101.0	100.5
7	230.0	162.2	130.5	112.1	106.0	101.2	100.6
8	250.0	172.2	135.5	114.1	107.0	101.4	100.7
9	270.0	182.2	140.5	116.1	108.0	101.6	100.8
10	290.0	192.2	145.5	118.1	109.0	101.8	100.9

The largest gains are obtained for small sample sizes.

4. Estimation that utilizes an approximate value $\beta_2\alpha$

In many practical cases, what we can obtain is the estimated value $\tilde{\beta}_2$ of the population coefficient of kurtosis β_2. Hence the assumption that we have only the approximate value $\tilde{\beta}_2=\beta_2\alpha$ $(\alpha>0)$ of the coefficient of kurtosis is reasonable. The estimator $\tilde{Y}$ of σ^2, which substitutes the approximate value $\tilde{\beta}_2=\beta_2\alpha$ for β_2 in the estimator Y^* given by (12), is defined by

$$\tilde{Y}=\frac{n}{\beta_2\alpha(n-1)+n^2-2n+3}\sum_{i=1}^{n}(X_i-\bar{X})^2 .$$

Now we are interested to obtain the range of α that the estimator $\tilde{Y}$ is more precise than the usual estimator s^2. Such a range of α is obtain from the following inequality;

$$\text{(13)} \qquad \mathrm{REF}\,(\tilde{Y})=\frac{\mathrm{MSE}\,(s^2)}{\mathrm{MSE}\,(\tilde{Y})}\geqq 1 .$$

After some calculation the inequality (13) is preserved and we have an inequality

$$\text{(14)} \qquad \begin{aligned}&(n-1)^2\beta_2^2[(n-1)\beta_2+3-n^2]\alpha^2\\&+2(n-1)\beta_2[(n-1)(n^2-2n+3)\beta_2+(n-3)^2]\alpha\\&+n^4-8n^3+24n^2-36n+27+(n-1)(-2n^3+9n^2-12n+9)\beta_2\geqq 0\end{aligned}$$

which is equivalent to the inequality (13). Then in Table 2 we give the ranges of α that satisfy the inequality (14) for different values of n and β_2.

Table 2 Ranges of $100\alpha\%$ which the relative efficiency REF $(\tilde{Y})$ has more than 1

β_2	Sample size n						
	5	10	20	50	100	500	1000
1	50.0-161.1	77.8-123.2	89.5-110.6	95.9-104.1	98.9-102.0	99.6-100.4	99.8-100.2
2	25.0-239.3	38.9-178.1	44.7-161.7	48.0-154.3	49.0-152.1	49.8-150.4	49.9-150.2
3	16.7-350.0	25.9-216.4	29.8-186.7	32.0-173.8	32.7-170.1	33.2-167.3	33.3-167.0
4	12.5-595.8	19.4-257.1	22.4-206.2	24.0-185.9	24.5-180.7	24.9-176.0	24.9-175.5
5	10.0-1810.0	15.6-307.9	17.9-224.5	19.2-195.0	19.6-187.1	19.9-181.4	20.0-180.7
6	8.3-	13.0-377.3	14.9-243.4	16.0-202.9	16.3-192.5	16.6-185.1	16.6-184.2
7	7.1-	11.1-481.7	12.8-263.9	13.7-210.0	14.0-197.0	14.2-187.9	14.3-186.8
8	6.3-	9.7-659.7	11.2-286.7	12.0-216.9	12.2-201.0	12.4-190.0	12.5-188.8
9	5.6-	8.6-1036.4	9.9-312.8	10.7-223.6	10.9-204.7	11.1-191.8	11.1-190.3
10	5.0-	7.8-2379.2	8.9-343.2	9.6-230.3	9.8-208.1	10.0-193.3	10.0-191.7

From Table 2 we can conclude that the estimator Y^* of σ^2 has robust efficiency for a considerable departure $\beta_2\alpha$ from the true value β_2 in the small sample cases or for the not so small β_2.

Acknowledgement

The authors wish to thank Dr. K. Takahashi for his useful comments.

BANARAS HINDU UNIVERSITY, INDIA
THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Searls, Donald J. (1964). The utilization of a known coefficient of variation in the estimation procedure, *Jour. Amer. Statist. Assoc.*, **59**, 1225-1226.
[2] Hirano, K. (1972). Using some approximately known coefficient of variation in estimating mean, *Proc. Inst. Stat. Math.* (In Japanese), Vol. 20-2, 61-64.