ON ESTIMATION OF VARIANCE IN NORMAL DISTRIBUTION

By

B. N. Pandey

Banaras Hindu University, Varanasi-5

(Received : October, 1980)

SUMMARY

In this note some estimators for estimating μ^2 and σ^2 in normal distribution are proposed and compared with the usual and maximum likelihood estimators. It is shown by simulation technique that these estimators have gain in efficiency over the maximum likelihood estimator as σ increases and sample size is small.

1. Introduction

Consider a normal distribution with mean μ and variance $\sigma^2 = c^2\mu^2$ where C is the coefficient of variation. Searls (1964) introduced the idea of utilizing known coefficient of variation as an apriori information in the estimation procedure of mean. Khan (1968) has proposed an estimator for μ when C is known. Govindrajula Sahai (1972) have also suggested some estimators for μ and μ^2 assuming C as known. If C is unknown, the minimum variance unbiased estimator for μ^2 is

$$U = y^{-2} - \frac{s^2}{n}$$

Where, $$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (yi - \bar{y})^2, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} yi$$

and n is the sample size. For smaller values of n, U may be negative, so Das (1975) suggested a biased estimator for μ^2 as

$$D = \bar{y}^2 \left[1 + \frac{s^2}{n\bar{y}^2} \right]^{-1}$$

and has studied its large sample and small sample properties. To obtain an estimator which has same mean squared error as D for larger sample size but has smaller bias than D, we have proposed an estimator

$$P = \bar{y}^2 \left[1 + \frac{s^2}{n\bar{y}^2} \left(1 + \frac{s^2}{n\bar{y}^2}\right) \right]^{-1}$$

in Section 2 and have studied its large sample properties. Since we are interested in estimating $\sigma^2=C^2\mu^2$, the estimator will be $\hat{\sigma}^2=\hat{C}^2\hat{\mu}^2$. On the basis of D and P, the estimator for σ^2, which have been proposed in Section 3, are

$$P_1=s^2\left[1+\frac{s^2}{n\bar{y}^2}\right]^{-1}$$

and

$$P_2=s^2\left[1+\frac{s^2}{n\bar{y}^2}\right]\left(1+\frac{s^2}{n\bar{y}^2}\right)\Bigg]^{-1}.$$

To study the small sample properties of these estimators, in Section 3, we have used simulation technique which gives only some indication of their true properties. We have compared these two estimators with the maximum likelihood estimator

$$L=\frac{n-1}{n}\,s^2.$$

1. ESTIMATORS P AND ITS LARGE SAMPLE PROPERTIES

Let $y_1, y_2, \ldots, y_n$ be a random sample of size n from a normal population with mean μ and variance $\sigma^2=C^2\mu^2$. The proposed estimator for μ^2 is

$$P=\bar{y}^2\left[1+\frac{s^2}{n\bar{y}^2}\left(1+\frac{s^2}{n\bar{y}^2}\right)\right]^{-1}. \qquad \ldots (1)$$

Since $\bar{y}$ and s^2 are unbiased and consistent estimators for μ and σ^2 respectively, we can write $\bar{y}=\mu+u$ and $s^2=\sigma^2+v$. Here u and v are random variables and are such that $E(u)=E(v)=0$. Thus, P can be written as

$$P=\mu^2\left(1+\frac{u}{\mu}\right)^2\left[1+\frac{C^2}{n}\left(1+\frac{v}{\sigma^2}\right)\left(1+\frac{u}{\mu}\right)^{-2}\right.$$
$$\left.+\frac{C^4}{n^2}\left(1+\frac{v}{\sigma^2}\right)^2\left(1+\frac{u}{\mu}\right)^{-4}\right]^{-1}. \qquad \ldots (2)$$

Now, expanding the right hand side of equation (2), retaining the terms of order $O(n^{-2})$ and simplifying, we get

$$E(P)=\mu^2 \qquad \ldots (3)$$

$$B(P)=0 \qquad \ldots (4)$$

and

$$M(P)=\mu^4\left[\frac{4C^2}{n}+\frac{2C^4}{n^2}\right]. \qquad \ldots (5)$$

Here E, B and M stand for expectation, bias and mean squared error respectively. The usual estimator for μ^2 is $\bar{y}^2$, which has

$$E(\bar{y}^2)=\mu^2\ (1+C^2/_n) \qquad \ldots(6)$$

$$B(\bar{y}^2)=\frac{C^2}{n}\ \mu^2 \qquad \ldots(7)$$

and $$M(\bar{y}^2)=\mu^4\left[\frac{4C^2}{n}+\frac{3C^4}{n^2}\right]. \qquad \ldots(8)$$

The minimum variance unbiased estimator for μ^2 is $U=\bar{y}^2-s^2/n$ which has

$$E(U)=\mu^2 \qquad \ldots(9)$$

$$B(U)=0 \qquad \ldots(10)$$

and $$M(U)=\mu^4\left[\frac{4C^2}{n}+\frac{2C^4}{n^2}\right]. \qquad \ldots(11)$$

The proposed estimator by Das (1975) is D which has

$$E(D)=\mu^2\left[1+\frac{C^4}{n^2}\right] \qquad \ldots(12)$$

$$B(D)=\frac{C^4}{n^2}\ \mu^2 \qquad \ldots(13)$$

and $$M(D)=\mu^4\left[\frac{4C^2}{n}+\frac{2C^4}{n^2}\right]. \qquad \ldots(14)$$

From equations (4), (7), (10) and (13), we have

$$0=B(P)=B(U)\leqslant B(D)\leqslant B(\bar{y}^2)=\frac{C^2}{n}\ \mu^2. \qquad \ldots(15)$$

From equations (5), (8), (11) and (14) we have

$$M(P)=M(U)=M(D)\leqslant M(\bar{y}^2) \qquad \ldots(16)$$

Thus for larger sample sizes, the estimator P, which is nearly unbiased, is preferable.

2. Estimators P_1 and P_2 and their Properties

The population variance σ^2 is equal to $C^2\mu^2$. Its estimators will be $\hat{\sigma}^2=\hat{C}^2\ \hat{\mu}^2$. In previous Section we have considered different estimators for μ^2. If wr eonsider $\hat{C^2}=\frac{s^2}{n\bar{y}^2}$, the possible estimators for σ^2 will be

$$T_1=s^2 \qquad \ldots(17)$$

$$T_2=s^2=\frac{s^4}{n\bar{y}^2} \qquad \ldots(18)$$

$$P_1=s^2\left[1+\frac{s^2}{n\bar{y}^2}\right]^{-1}. \qquad \ldots(19)$$

and
$$P_2 = s^2 \left[1 + \frac{s^2}{n\bar{y}^2} \left(1 + \frac{s^2}{n\bar{y}^2} \right) \right]^{-1}. \qquad ...(20)$$

Here T_1 is the usual unbiased estimator for σ^2. T_2 may be negative for smaller values of n, so we have not considered this for our purpose. The maximum likelihood estimator for σ^2 is L which has $M(L) = \frac{2n-1}{n^2} \sigma^4 \leqslant M(s^2)$. Therefore, we have considered the maximum likclihood estimator L for our purpose but not s^2.

The proposed estimators P_1 and P_2 are functions of $\bar{y}$ and s^2, which are independently distributed in the normal distribution. But for smaller values of n, it will be difficult to obtain the exact expressions for bias and mean squared errors of these estimators. One can apply two procedures namely: (*i*) Simulation technique, and (*ii*) Quadrature technique, for computing the biases and mean squared errors of these estimators. It should be noted that the results obtained by simulation technique will give indication of the true properties of these estimators. So we have used simulation technique to see the true properties.

(*a*) *Simulation results :*

We have generated 1,000 random samples of size 5 from $N(10, 2^2)$, $N(10, 8^2)$ and $N(10, 10^2)$ and have calculated the relative biases and the relative efficiencies of P_1 and P_2 with respect to L.

TABLE

Relative biases and Relative Efficiencies of P_1, P_2 with respect to L for $n=5$

Population sampled	*Relative bias of P_1*	*Relative bias of P_2*	*Relative bias of L*	*REF (P_1, L)*	*REF (P_2, L)*
$N(10, 2^2)$	.008999	.007330	−.2000	82.75	86.98
$N(14, 8^2)$	−.1236	−.1488	−.2000	112.57	116.93
$N(10, 10^2)$	−.2890	−.3563	−.2000	122.63	120.59

The above table shows that the relative biases in the estimators P_1 and P_2 decreases monotonically as the coefficient of variation increases and are greator than the relative bias in L. However, the relative biases in P_1 and P_2 are less than that of L in magnitude for smaller values of C. Moreover, the magnitude of relative bias in P_1 is less than that of P_2 for moderate values of C. It can further be seen that for moderate values of C, there are gains in efficiency for eestimators P_1 and P_2 and the gains increase as the coefficient of variation increases. For smaller values of C the estimator P_2 has smaller relative bias and mean squared error than P_1. Thus estimator P_2 is preferable for smaller values of C and n.

(*b*) *Large sample results :*

Since $\bar{y}$ and s^2 are consistent estimators for μ and σ^2, respectively, therefore we can write $\bar{y}=\mu+u$ and $s^2=\sigma^2+v$. Now,

$$P_1=\sigma^2\left(1+\frac{v}{\sigma^2}\right)\left[1-\frac{C^2}{n}\left(1+\frac{v}{\sigma^2}\right)\left(1+\frac{u}{\mu}\right)^{-2}\right]^{-1} \quad \ldots(21)$$

and

$$P_2=\sigma^2\left(1+\frac{v}{\sigma^2}\right)\left[1+\frac{C^2}{n}\left(1+\frac{v}{\sigma^2}\right)\left(1+\frac{u}{\mu}\right)^{-2}+\frac{C^4}{n^2}\left(1+\frac{v}{\sigma^2}\right)^2\left(1+\frac{u}{\mu}\right)^{-4}\right]^{-1}. \quad \ldots(22)$$

Expanding the right hand sides of P_1, P_2, retaining the terms of order $0(\bar{n}^2)$ and simplifying, we get,

$$B(P_1)=-\sigma^2\left[\frac{C^2}{n}+\frac{2C^2}{n^2}+\frac{2C^1}{n^2}\right] \quad \ldots(23)$$

$$B(P_2)=-\sigma^2\left[\frac{C^2}{n}+\frac{2C^2}{n^2}+\frac{3C^4}{n^2}\right] \quad \ldots(24)$$

and

$$M(P_1)=M(P_2)=\frac{2\sigma^4}{n}\left[1+\frac{C^2}{2n}(C^2-8)\right]. \quad \ldots(25)$$

Thus we see that for larger value of n, P_1 and P_2 have gain efficiency for smaller values of C.

Thus we may conclude that the estimators P_1 and P_2 are preferable than the maximum likelihood estimator L if

(*i*) Sample size is small but coefficient of variation is large ;

(*ii*) Sample size is large but coefficient of variation is small.

ACKNOWLEDGEMENT

The author is thankful to the referee for his useful suggestions.

References

Searls, D.T. (1964) : "The utilization of a known coefficient of variation in the estimation procedure", *Jour. Amer. Stat. Assoc.* **59**, pp. 1225-26.

Khan, R.A. (1968) : "A note on estimation of the mean of a normal distribution with known coefficient of Variation", *Jour. Amer. Stat. Assoc.*, **62** pp. 1039-41.

Govindrajulu, Z. and Sahai, H. (1972) : "Estimation of parameters of a normal distribution with known coefficient of variation" *Rep. Stat. App. JUSE,* **19**, pp. 85-98.

Das, B. (1975) : "On estimation of μ^2 in normal distribution", *Cal. Stat. Assoc. Bull.*, **24**, pp. 135-38.