AN ESTIMATOR FOR VARIANCE OF A NORMAL POPULATION WHEN PRIOR INFORMATION ON THE MEAN IS AVAILABLE

By

S.R. SRIVASTAVA AND D.C. SINGH*

Banaras Hindu University, Varanasi

(Received : December, 1979)

SUMMARY

An estimator of variance of a normal population when two guessed values of mean are available has been considered and its efficiency is examined.

1. INTRODUCTION

Consider a random sample $x_1, x_2, \ldots, x_n$ of size n from a normal population with unknown mean μ and variance σ^2. If σ^2 is known, Pradhan [2] developed a $(\alpha+\beta)_{min}$ test of $H_1 : \mu=\mu_0$ against $H_2 : \mu=\mu_1(\mu_1>\mu_0)$ which minimizes the sum of the probabilities of two types of errors, given by

$$\text{Reject } H_1 \text{ if } \bar{x} \geqslant \frac{\mu_0+\mu_1}{2}, \qquad \ldots(1)$$

where $\bar{x}$ is the sample mean. Singh and Pandey [3] showed that test (1) has the same property even if σ^2 is not known.

Estimator of σ^2 using prior information in the form of a guessed value of μ have been considered by Davis and Arnold [1], and Srivastava and Singh [4]. In this paper we propose an estimator of the variance σ^2 using μ_0 and μ_1 as guessed (somehow) values of μ. Its properties are studied and relative efficiency is discussed.

* Present address :P.G. College, Varanasi.

2. The Proposed Estimator $\hat{\sigma}^2$

The proposed estimator $\hat{\sigma}^2$ of σ^2 is

$$\hat{\sigma}^2 = \begin{cases} ks^2+(1-k)\, w_0^2 \text{ if } \bar{x} < \dfrac{\mu_0+\mu_1}{2} \\ ks^2+(1-k)\, w_1^2; \text{ otherwise,} \end{cases} \qquad \ldots(2)$$

where

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i-\bar{x})^2, \quad w_j^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i-\mu_j)^2, \; j=0, 1$$

and k $(0 \leqslant k \leqslant 1)$ is a constant chosen by the experimenter according to his belief in guessed values. It is possible to write (2) as follows :

$$\tilde{\sigma}^2 = \begin{cases} \dfrac{n-(1-k)}{n} s^2+(1-k)(\bar{x}-\mu_0)^2 \text{ if } \bar{x} < \dfrac{\mu_0+\mu_1}{2} \\ \dfrac{n-(1-k)}{n} s^2+(1-k)(\bar{x}-\mu_1)^2; \text{ otherwise} \end{cases}$$

3. Bias, MSE and Relative Efficiency of $\tilde{\sigma}^2$

The expected value of $\tilde{\sigma}^2$ is

$$E(\tilde{\sigma}^2) = E\left[\left\{\frac{n-(1-k)}{2} s^2+(1-k)(\bar{x}-\mu_0)^2\right\} \Big/ \bar{x} < \frac{\mu_0+\mu_1}{2}\right]$$

$$\cdot P\left[\bar{x} < \frac{\mu_0+\mu_1}{2}\right] + E\left[\left\{\frac{n-(1-k)}{n} s^2+(1-k)(\bar{x}-\mu_1)^2\right\} \Big/\right.$$

$$\left.\bar{x} \geqslant \frac{\mu_0+\mu_1}{2}\right] P\left[\bar{x} \geqslant \frac{\mu_0+\mu_1}{2}\right]$$

$$= \frac{n-(1-k)}{n} \sigma^2+(1-k)$$

$$\left[\int_{-\infty}^{\frac{\mu_0+\mu_1}{2}} (\bar{x}-\mu_0)^2 f(\bar{x}) d\bar{x}) + \int_{\frac{\mu_0+\mu_1}{2}}^{\infty} (\bar{x}-\mu_1)^2 f(\bar{x})\, d\bar{x} \right], \qquad \ldots(3)$$

where

$$f(\bar{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{n}{2}\left(\frac{\bar{x}-\mu}{\sigma}\right)^2}$$

Simplifying (3), we get

$$E(\tilde{\sigma}^2)=\sigma^2+\frac{\sigma^2}{n}(1-k)\Big[\delta_1^2-2(\delta_1-\delta_0)\phi(a)-(\delta_1^2-\delta_0^2)\Phi(a)\Big],$$

where

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2},\quad \Phi(t)=\int_{-\infty}^{t}\phi(x)\,dx,$$

$$a=\frac{\delta_0+\delta_1}{2}\text{ and }\delta_i=\frac{(\mu_i-\mu)\sqrt{n}}{\sigma},\ i=0,1.$$

Hence, the bias in $\tilde{\sigma}^2$ expressed as a function σ^2 given by

$$\frac{E(\tilde{\sigma}^2)-\sigma^2}{\sigma^2}=\frac{(1-k)}{n}\Big[\delta_1^2-2(\delta_1-\delta_0)\phi(a)-(\delta_1^2-\delta_0^2)\Phi(a)\Big]$$

Now

$$E(\tilde{\sigma}^4)=E\left[\left\{\frac{n-(1-k)}{n}s^2+(1-k)(\bar{x}-\mu_0)^2\right\}^2\Big/\bar{x}<\frac{\mu_0+\mu_1}{2}\right]$$

$$\cdot P\left[\bar{x}<\frac{\mu_0+\mu_1}{2}\right]+E\left[\left\{\frac{n-(1-k)}{n}s^2+(1-k)(\bar{x}-\mu_1)^2\right\}^2\Big/\right.$$

$$\left.\bar{x}\geqslant\frac{\mu_0+\mu_1}{2}\right]P\left[\bar{x}\geqslant\frac{\mu_0+\mu_1}{2}\right]$$

$$=\left(\frac{n-(1-k)}{n}\right)^2\left(\frac{2\sigma^4}{n-1}+\sigma^4\right)+(1-k)^2$$

$$\left[\int_{-\infty}^{\frac{\mu_0+\mu_1}{2}}(\bar{x}-\mu_0)^4 f(\bar{x})\,d\bar{x}+\int_{\frac{\mu_0+\mu_1}{2}}^{-\infty}(\bar{x}-\mu_1)^4 f(\bar{x})\,d\bar{x}\right]$$

$$+2(1-k)\,\sigma^2\left(\frac{n-(1-k)}{n}\right)$$

$$\left[\int_{-\infty}^{\frac{\mu_0+\mu_1}{2}}(\bar{x}-\mu_0)^2 f(\bar{x})\,d\bar{x}+\int_{\frac{\mu_0+\mu_1}{2}}^{\infty}(\bar{x}-\mu_1)^2 f(\bar{x})\,d\bar{x}\right]\quad\ldots\quad(4)$$

Simplifying (4), we obtain

$$E(\sigma^4)=\left(\frac{n-(1-k)}{n}\right)^2\left(\frac{2\sigma^4}{n-1}+\sigma^4\right)+\frac{(1-k)^2\sigma^4}{n^2}$$

$$\cdot[3+6\delta_1^2+\delta_1^4-\{4(\delta_1-\delta_0)(2+a^2)-6(\delta_1^2-\delta_0^2)a+$$

$$4(\delta_1^3-\delta_0^3)\}\,\phi(a)-\{6(\delta_1^2-\delta_0^2)+(\delta_1^4-\delta_0^4)\}\Phi(a)]$$

$$+2(1-k)\left(\frac{n-(1-k)}{n}\right)\frac{\sigma^4}{n}[1+\delta_1^2-2(\delta_1-\delta_0)\phi(a)-$$

$$(\delta_1^2-\delta_0^2)\,\Phi(a)].$$

So,

$$MSE(\tilde{\sigma}^2)=E\,(\tilde{\sigma}^4)-2\sigma^2E\,(\tilde{\sigma}^2)+\sigma^4$$

$$=\frac{2\sigma^4}{n-1}\left[\left(\frac{n-(1-k)}{n}\right)^2+\frac{(n-1)\,(1-k)^2}{2n^2}\{2+4\delta_1^2+\delta_1^4\right.$$

$$-[4(\delta_1-\delta_0)\,(1+a^2)-6(\delta_1^2-\delta_0^2)a+4(\delta_1^3-\delta_0^3)]\phi(a)$$

$$\left.-[4(\delta_1^2-\delta_0^2)+(\delta_1^4-\delta_0^4)]\,\Phi\,(a)\}\right] \quad ..\ (5)$$

The relative efficiency of $\tilde{\sigma}^2$ to s^2 is given by

$$e=\frac{V(s^2)}{MSE(\tilde{\sigma}^2)}$$

$$=\left[\left(\frac{n-(1-k)}{2}\right)^2+\frac{(n-1)(1-k)^2}{2n^2}\{2+4\delta_1^2+\delta_1^4\right.$$

$$-[4(\delta_1-\delta_0)(1+a^2)-6(\delta_1^2-\delta_0^2)a+4(\delta_1^3-\delta_0^3)]\phi(a)-$$

$$\left.-[4(\delta_1^2-\delta_0^2)+(\delta_1^4-\delta_0^4)]\Phi(a)\}\right]^{-1}$$

Differentiating (5) *w.r.* to k and putting the derivative equal to zero, we get

$$k=\frac{(n-1)(\triangle-2)}{2+(n-1)\triangle},$$

where

$$\triangle=2+4\delta_1^2+\delta_1^4-[4(\delta_1-\delta_0)(1+a^2)-6(\delta_1^2-\delta_0^2)a+4(\delta_1^3-\delta_0^3)]\phi(a)$$

$$-\ [4(\delta_1^2-\delta_0^2)+(\delta_1^4-\delta_0^4)]\Phi(a)$$

which is the value of k for which MSE $(\tilde{\sigma}^2)$ is minimum. The relative efficiency will be greater than one if

$$k'<k\leqslant 1$$

where

$$k'=\frac{2-4n+(n-1)\triangle}{2+(n-1)\triangle}$$

4. Recommendations

In order to compare the efficiency of the proposed estimator $\tilde{\sigma}^2$ with that of s^2, we have computed the values of e for $n=10$, 15, 30 $k=0.3(0.2)0.9$ and for set of values of δ_0 and δ_1 which are given in Table 1 to 4. As the loss (gain) in efficiency is small where n is large (see Table 1—4), we have computed the values of k' for only small value of $n=10$, in Table 5 (see Appendix). From Table 5, we conclude that for fixed n if $-1 \leqslant \delta_0 \leqslant 0$, $0 \leqslant \delta_1 \leqslant 1$ and $(\delta_0=-2,\ \delta_1=2)$ any value of k leads to improvement of $\tilde{\sigma}^2$ over s^2, which may also seen from Table 1 to 4. In above said ranges of δ_0 and δ_1 the gain in efficiency decreases as k increases in general. As the value of k increases, the ranges of δ_0 and δ_1 for which $\tilde{\sigma}^2$ is more efficient than s^2 also increases. Therefore we suggest to choose k in the following manner :

(*i*) If μ is expected to lie very near to μ_0 or μ_1, any value of k may be chosen.

(*ii*) If μ is not expected to differ much from μ_0 or μ_1, k may be chosen around 0.5.

(*iii*) In other situations, *i.e.*, μ may differ too much from μ_0 or μ_1, k should be taken equal to 1.

Acknowledgement

The authors are grateful to the referees for their suggestions which lead to the improvement of the paper.

References

[1] Davis, R.L. and Arnold, J.C. ; An Efficient Preliminary Test Estimator for the Variance of a Normal population when the Mean is Unknown, *Biometrika*, Vol. **57** (1970), pp. 674-677.

[2] Pradhan, Meena ; On Testing a Simple Hypothesis Against a Simple Alternative Making the Sum of the Probabilities of Two Types of Errors Minimum, *Jour. Ind. Statist. Assoc.*, Vol. **6** (1968), pp. 149-159.

[3] Singh, J. and Pandey M. ; A Note on Testing of a Normal Population When Variance is Unknown, *Jour. Ind. Stast. Assoc.*, Vol. 16 (1978), pp. 145-146.

[4] Srivastava, S.R. and Singh, R.D. : The preliminery Test Estimator for the Variance of a Normal Population *Proceeding of National Academy of Science*, **49** (1979). pp. 142-148.

Values of e

TABLE 1

$k=0.3$

n	δ_0 \ δ_1	−3	−2	−1	0	1	2	3
10		0.2866	0.6194	0.9820	1.1135	1.1562	0.8240	0.5167
15	−4	0.3674	0.7007	0.9853	1.0732	1.0403	0.8697	0.6064
30		0.5282	0.8182	0.9913	1.0355	1.0194	0.9270	0.7474
10			0.6193	0.9820	1.1200	1.1125	0.9724	0.7259
15	−3		0.7007	0.9853	1.0774	1.0726	0.9786	0.7916
30			0.8181	0.9913	1.0375	1.0352	0.9877	0.8792
10				0.9813	1.1200	1.1404	1.0974	0.6300
15	−2			0.9848	1.0774	1.0904	1.0631	0.7101
30				0.9910	1.0375	1.0437	1.0306	0.8247
10					1.1155	1.1383	1.1109	1.1125
15	−1				1.0745	1.0891	1.0715	1.0726
30					1.0361	1.0430	1.0347	1.0352
10						1.0692	1.1200	1.1200
15	0					1.0444	1.0774	1.0774
30						1.0214	1.0375	1.0375
10							0.9813	0.9820
15	1						0.9848	0.9853
30							0.9910	0.9913
10								0.6217
15	2							0.7028
30								0.8196

TABLE 2

$k=0.5$

n	δ_0 \ δ_1	-3	-2	-1	0	1	2	3
10		0.4462	0.7782	1.0196	1.0876	1.0625	0.9256	0.6903
15	-4	0.5377	0.8341	1.0116	1.0569	1.0404	0.9457	0.7622
30		0.6914	0.9059	1.0051	1.0277	1.0196	0.9703	0.8602
10			0.7782	1.0196	1.0908	1.0872	1.0143	0.8590
15	-3		0.8341	1.0116	1.0590	1.0566	1.0079	0.8956
30			0.9059	1.0051	1.0288	1.0276	1.0032	0.9429
10				1.0192	1.0908	1.1006	1.1077	0.7867
15	-2			1.0113	1.0590	1.0653	1.0519	0.8409
30				1.0049	1.0287	1.0318	1.0253	0.9100
10					1.0886	1.0996	1.0854	1.0872
15	-1				1.0576	1.0647	1.0561	1.0555
30					1.0281	1.0315	1.0273	1.0276
10						1.0656	1.0908	1.0908
15	0					1.0425	1.0590	1.0588
30						1.0206	1.0287	1.0288
10							1.0192	1.0196
15	1						1.0113	1.0116
30							1.0049	1.0051
10								0.7801
15	2							0.8356
30								0.9068

TABLE 3

$k=0.7$

n	δ_0 \ δ_1	—3	—2	—1	0	1	2	3
10		0.7028	0.9271	1.0319	1.0560	1.0473	0.9951	0.8791
15	—4	0.7730	0.9473	1.0205	1.0357	1.0309	0.9953	0.9121
30		0.8676	0.9714	1.0099	1.0203	1.0151	0.9970	0.9519
10			0.9271	1.0319	1.0570	1.0558	1.0299	0.9661
15	—3		0.9473	1.0205	1.0374	1.0366	1.0192	0.9751
30			0.9714	1.0099	1.0184	1.0180	1.0092	0.9864
10				1.0318	1.0570	1.0603	1.0534	0.9314
15	—2			1.0204	1.0374	1.0396	1.0349	0.9504
30				1.0099	1.0184	1.0195	1.0172	0.9731
10					1.0563	1.0600	1.0555	1.0558
15	—1				1.0369	1.0394	1.0364	1.0366
30					1.0181	1.0194	1.0179	1.0180
10						1.0484	1.0570	1.0570
15	0					1.0316	1.0374	1.0374
30						1.0155	1.0184	1.0184
10							1.0318	1.0319
15	1						1.0204	1.0205
30							1.0099	1.0099
10								0.9281
15	2							0.9480
30								0.9718

TABLE 4

$k=0.9$

n	δ_0 \ δ_1	—3	—2	—1	0	1	2	3
10	—4	0.9674	1.0046	1.0171	1.0196	1.0187	1.0130	0.9981
15		0.9768	1.0027	1.0113	1.0130	1.0124	1.0084	0.9982
30		0.9877	1.0012	1.0056	1.0065	1.0061	1.0040	0.9988
10			1.0046	1.0171	1.0197	1.0196	1.0163	1.0203
15	—3		1.0027	1.0113	1.0131	1.0130	1.0111	1.0061
30			1.0012	1.0056	1.0065	1.0064	1.0055	1.0029
10				1.0170	1.0197	1.0200	1.0193	1.0052
15	—2			1.0112	1.0131	1.0133	1.0128	1.0031
30				1.0056	1.0065	1.0066	1.0064	1.0014
10					1.0196	1.0200	1.0196	1.0196
15	—1				1.0130	1.0133	1.0130	1.0130
30					1.0065	1.0066	1.0064	1.0065
10						1.0188	1.0197	1.0197
15	0					1.0124	1.0131	1.0131
30						1.0062	1.0065	1.0065
10							1.0170	1.0171
15	1						1.0112	1.0113
30							1.0056	1.0056
10								1.0047
15	2							1.0028
30								1.0012

TABLE 5

Values of k' $n=10$

δ_0 \ δ_1	—3	—2	—1	0	1	2	3
—4	0.9627	0.8701	0.3810	—1.5727	—0.2150	0.7228	0.9089
3		0.8701	0.3809	—1.9819	—1.5219	0.4181	0.8107
—2			0.3838	—1.9791	—4.8020	—0.9266	0.8653
—1				—1.6895	—4.2964	—1.4390	—1.5218
0					—0.3023	—1 .9791	—1.9819
1						0.3838	0.3809
2							0.8691