A NOTE ON ESTIMATION OF VARIANCE IN EXPONENTIAL DENSITY

By

V.P. Ojha

University of Gorakhpur, U.P.

(Received : July, 1980)

Summary

Another estimator T of the variance θ^2 of the exponential distribution is presented which reduces the effect of large true observations and is better than OJHA and SRIVASTAVA [1] estimator S_t^2 for a wide range of cutoff point t for all sample sizes.

Introduction

In the one-parameter exponential distribution, the variance θ^2 is the square of the values of the mean θ. Pandey and Singh [2] proposed a biased estimator $y^*=\frac{n^2\bar{y}^2}{(n^2+5n+6)}$ as the estimator of the variance θ^2, where $\bar{y}=\frac{1}{n}\sum_{j=1}^{n} y_j$ is the sample mean and showed that y^* is better, in the sense of min. mean squared error, than the usual unbiased estimator

$$S^2=\frac{1}{n-1}\left[\sum_{j=1}^{n}(y_j-\bar{y})^2\right] \text{ and } \hat{\theta}^2=\frac{n\bar{y}^2}{n+1}.$$

Since the distribution is positive valued only and positively skewed, therefore, a sampler may occasionally get a few or more observations from the right tail of the distribution and consequently get considerably higher estimates of the parameters using unbiased estimators. The phenomenon is more pronounced in small samples.

In order to reduce the effect of these offending observations, Searls [3] has suggested an estimator ;

$$\bar{y}_t=\frac{1}{n}\left[\sum_{j=1}^{r} y_j+(n-r)\,t\right]$$

of the mean θ which is constructed by replacing all the observations greater than a predetermined Cutoff point t by the value of t itself. Here r is the number of observations less than or equal to t. It is further assumed that the sampler has some indications of the possible range in which the data may fall, which helps him to choose cut off point t and thus decide the 'bigness' of the sampled material. In the similar circumstances and for the same purpose, Ojha and Srivastava [1] proposed an estimator

$$S_t^2=\frac{1}{n-1}\left[\sum_{j=1}^{r} y_j^2+(n-r)\,t-n\bar{y}_t^2\right]$$

of the variance θ^2 and showed that S_t^2 fares better than the unbiased estimator S^2 for a wide range of values of t.

In the present note we propose another estimator

$$T=\frac{n\bar{y}_t^2}{n+1}$$

of the variance θ^2 and study its properties.

2. Estimator

Let $y_1, y_2, y_3 \ldots\ldots y_n$ be a random sample of size n from an exponential population with p.d.f.

$$f(y,\theta)=\frac{1}{\theta}\,e^{-y/\theta} \qquad ; y>0,\ \theta>0. \qquad \ldots(1)$$

with c.d.f. $F(y)$.

The proposed estimator T of the variance θ^2 is

$$T=\frac{n}{n+1}\bar{y}_t^2 \qquad \ldots(2)$$

with

$$\bar{y}_t=\frac{1}{n}\left[\sum_{j=1}^{r} y_i+(n-r)\,t\right], \qquad \ldots(3)$$

where r is the number of observations with values less than equal to t and t is the cutoff point where the distribution $F(y)$ get truncated on the right.

Using the Eqns, (2.7) and (3.9) and notations of Ojha and Srivastava [1], we get

$$E[\bar{y}_t^2]=\frac{1}{n}[p(\sigma_t^2+\mu_t^2)+qt^2-(p\mu_t+qt)^2]+(p\mu_t+qt)^2 \quad (4)$$

and

$$E[(y_t^2)^2=\frac{1}{n^3}[(p\alpha_{4,t}+qt^4)+4(n-1)(p\alpha_{3,t}+qt^3)$$

$$+6(n-1)(n-2\ \{p(\sigma_t^2+\mu_t^2)+qt^2\}(p\mu_t+qt)^2$$

$$+(n-1)(n-2)(n-3)(p\mu_t+qt)^4+3(n-1)\{p(\sigma_t^2+\mu_t^2)$$

$$+qt^2\}^2] \quad (5)$$

giving

Bias $(T)=E(T)-\theta^2$

$$=\frac{1}{n+1}[\sigma_2^*+n(p\mu_t+qt)^2]-\theta^2 \quad (6)$$

and

$$MSE(T)=\frac{1}{n(n+1)^2}\cdot M_T^* \quad (7)$$

where

$$M_T^*=\mu_4^x+4\,n\mu_3^*(p\mu_t+qt)+4\sigma_2^*(p\mu+qt)^2$$

$$+(2n-3)\sigma_4^*+\left[\frac{1}{n+1}\,\sigma_2^*\{+n(p\mu_t+qt)^2\}-\theta^2\right]^2 \quad (8)$$

Here, $p\,\mu_t+qt=p\theta$

$$\lambda=\sigma_2^*/\theta^2 \qquad (0<\lambda\leqslant 1) \quad (9)$$

$$\sigma_2^*=p(\sigma_t^2+\mu_t^2)+qt^2-(p\mu_t+qt)^2$$

$$=[1-2q\,\frac{t}{\theta}-q^2]\theta^2 \quad (10)$$

$$\mu_3^* = (p\,\alpha_{3,t} + qt^3) - 3\{p(\sigma_t^2 + \mu_t^2) + qt^2\}(p\mu_t + qt)$$

$$+2(p\mu_t + qt)^3$$

$$= [3q\,\lambda - 3q\left(\frac{t}{\theta}\right)^2 + p^2(q+2)]\,\theta^3 \tag{11}$$

$$\mu_4^* = (p\alpha_{4,t} + qt^4) - 4(p\alpha_{3,t} + qt^3)(p\mu_t + qt)$$

$$+6\{p(\sigma_t^2 + \mu_t^2) + qt^2\}(p\mu_t + qt)^2 - 3(p\mu_t + qt)^4$$

$$= [-3\lambda^2 + 12\lambda - 4q\left(\frac{t}{\theta}\right)^3]\,\theta^4 \tag{12}$$

and

$$\lambda = \sigma_2^*/\theta^2 \tag{13}$$

So that

$$\text{Bias}\,(T) = \frac{1}{n+1}[\lambda + np^2]\,\theta^2 - \theta^2 \tag{14}$$

$$MSE\,(T) = \frac{\theta^4}{n(n+1)}\,M_T^* \tag{15}$$

wiih

$$M_T^* = (n-1)(\lambda + np^2)^2 - [2n^2 + 2n - 12 - 5(n-1)\lambda](\lambda + np^2)$$

$$+n(n+1)^2 - 4q\left(\frac{t}{\theta}^3\right) - 12\,n\,q(t/\theta)^2 \tag{16}$$

By defining relative efficiency $REF\,(T, \frac{n}{n+1}\,\bar{y}^2)$ of the estimator T w.r.t. $\hat{\theta}^2 = \frac{n\,\bar{y}^2}{n+1}$ as

$$REF\left(T, \frac{n\bar{y}^2}{n+1}\right) = \frac{\text{Var}\left(\frac{n\,\bar{y}^2}{n+1}\right)}{MSE\,(T)} \qquad \text{...(17)}$$

we get,

$$REF\left(T, \frac{n\bar{y}^2}{n+1}\right) = \frac{4\,n^2 + 10n + 6}{M_T^*} \qquad \text{...(18)}$$

Since,

$$\text{Var}\,(\hat{\theta}^2) = \frac{4n+6}{n(n+1)}\theta^4 \qquad \text{...(19)}$$

The table I below gives the relative efficiency of the estimator T for samples of sizes 5, 10, 20, 50 and 100 as compared to an unbiased estimator $\hat{\theta}^2=\frac{n\bar{y}^2}{n+1}$ for the exponential distribution. In conformity with the estimator S_t^2 of Ojha and Srivastava [1] it is seen that the gains in relative efficiency are achieved for a wide range of t and the gains become modest for large sample sizes and for large values of t.

TABLE 1

Ref. Efficiencies (%) of T w.r.t. $\frac{n\bar{y}^2}{n+1}$

Values of t/θ	*Sample size*				
	5	*10*	*20*	*50*	*100*
1	127.77	104.65	57.15	21.90	11.02
2	316.69	235.38	166.27	91.70	52.89
3	196.44	177.52	163.27	142.90	121.43
4	140.62	132.70	128.28	124.40	121.17
5	118.19	114.23	112.16	110.78	109.78
6	108.42	106.34	105.16	105.19	104.34
7	103.93	102.27	102.30	101.37	101.86
8	101.86	101.27	100.99	100.83	100.78
9	100.85	100.56	100.43	100.35	100.32
10	100.38	100.25	100.17	100.14	100.13

In order to compare T with S_t^2 of Ojha and Srivastava [1], the Relative efficiency of T w.r.t. S_t^2 is defined as

$$REF(T, S_t^2)=\frac{MSE(S_t^2)}{MSE(T)}$$

$$=\frac{(n+1)^2}{n-1}\,\frac{M_S^*}{M_T^*} \qquad \ldots(20)$$

where

$$MSE\ (S_t^2)=\frac{1}{n\ (n-1)}\ M_S^* \qquad \ldots(21)$$

and

$$M_T^x=(n^2-5n+6)\ \lambda^2-2(n-1)\ (n-6)\ \lambda+n(n-1) - 4\ (n-1)\ qt^3/\theta^3 \qquad \ldots(22)$$

The following Table 2 gives the values of relative efficiency of estimator T w.r.t. estimator S_t^2, as earlier, for sample sizes 5, 10, 20, 50 and 100. It is seen that estimator T is better than the estimator S_t^2 for all values of t and sample sizes except for the sample size to be less than 20 with t nearer to 3 to 5 times mean. Further, the $REF\ (T, S_t^2)$ increases with the sample sizes n and the values of t.

TABLE 2

Ref. Efficiencies (%) of estimator T w.r.t. estimator S_t^2

Value of t/θ	*Sample size n*				
	5	*10*	*20*	*50*	*100*
1	112.38	190.21	211.94	205.78	208.09
2	134.53	180.73	262.39	359.76	414.51
3	79.47	87.92	116.46	198.51	306.27
4	129.60	95.67	98.53	115.35	144.95
5	130.93	125.16	123.46	126.33	131.52
6	155.39	151.41	150.46	151.52	151.73
7	172.36	170.07	169.93	170.28	170.61
8	183.03	181.90	182.40	183.06	183.37
9	189.19	188.83	189.71	190.60	189.96
10	192.59	192.09	193.72	194.76	195.17

REFERENCES

[1] OJHA, V.P. and Srivastava S.R. (1979) : An Estimator of the Population Variance in the Presence of True Observation *Journ. Ind. Soc. Agri. Stat.* Vol. 31 No. 1, 77-84.

[2] Pandey, B.N. and Singh, J. (1977) : A note on Estimation of Variance in Exponential density Sankhya Series B, Vol. 39 No. 3, 294-98.

[3] Searls, D.T. (1966) : An Estimator for a Population Mean which reduces the effect of large True Observations. *Journ. Amer. Stat. Assocn.*, Vol. 61, 1200-1204.