한국통계학회 논문집 제1권1호
1994년 12월 pp. 75-80

Parametric Estimation of Two-Parameter Exponential Model in the Presence of Unidentified Outliers

Jungsoo Woo[1)]

Abstract

Parametric estimators of two parameter exponential distribution in the presence of unidentified outliers are proposed, and the means and variances of the estimators are obtained as the exact forms.

1. Introduction

Many authors considered the problems of parametric estimation in the exponential and gamma distributions.There are many situations in which it is reasonable to assume that the items may not be homogeneous and hence the assumption of i.i.d. random variables may be unrealistic, and then the model may have to be modified suitably.

Gather and Kale(1988),Dixit(1989 & 1991),and Rao and et al(1991) considered the problem of parametric estimation in an exponential, a gamma, and a generalized gamma distributions in the presence of outliers.

Here we consider parametric estimation of the location and the scale parameters in an assumed same parameter exponential distribution in the presence of (n-k) unidentified outliers. The means and variances of estimators for the location and the scale parameter in the assumed exponential model are derived, and also the mean and variance of a variance estimator in the assumed exponential model.

Unbiased estimators for several parameters in the assumed exponential model are proposed, and then their variances are obtained as the exact forms.

Next, in Martz(1982), the reliable life time of the reliability in the assumed exponential model is considered, the ML estimator and an unbiased estimator for the reliable life time are given, and their means and variances are obtained.

Finally,we can numerically compare the MLE and proposed unbiased estimators of the scale,the location,the variance,and the reliable life time in the assumed exponential model as numerical evaluations of the mean squared errors(MSE) are carried out.

2. Parametric Estimations

Assume $X_1, \cdots, X_n$ be independent exponential random variables,

1) Department of Statistics, Yeungnam University, Kyongsan, KOREA.

$$X_i \sim \text{EXP}(\mu,\sigma)\ ,\quad i=1,\ 2,\ \cdots,k$$

$$X_j \sim \text{EXP}(\mu,\sigma/b)\ ,\qquad j=k+1,\ k+2,\ \cdots,\ n\ .$$

where $b>0$, $\mu\geq 0$,and $\sigma>0$, EXP denotes an exponential distribution with the location parameter μ and scale parameter σ,and b is a given known constant.

Let $Y_i = (X_i-\mu)/\sigma$, $i=1,\ 2,\ \cdots,\ n$ and $S=\sum_{i=1}^{n} Y_i$.

From formulas in Oberbettinger(1973 & 1974), the pdf of S can be obtained as:

$$f(s)=b^{n-k}e^{-s}s^{n-1}F(n-k,n;(1-b)s)/\Gamma(n),\qquad s>0 \tag{2.1}$$

where F is the hypergeometric function.

From formula 7.621(5) in Gradshteyn(1980) and formula 131.27 in Abramowitz(1970), then it can be shown that

$$\int_0^\infty f(s)ds=1\ ,$$

from which, we can get a useful formula to apply later:

$$\int_0^\infty e^{-s}s^{n-1}F(n-k,n;(1-b)s)ds=\Gamma(n)/b^{n-k},\qquad b>0 \tag{2.2}$$

From the formula (2.2), the moment generating function(mgf) of S is

$$M(t)=(1-t)^{-k}(1-t/b)^{n-k},\quad \text{if}\quad t<\min(1,b) \tag{2.3}$$

From the mgf of S, we can obtain the r-th moment of S

$$E(S^r) = \sum_{j=0}^{r}\binom{r}{j}(k)_{r-j}(n-k)_j/b^j, \tag{2.4}$$

where, $(a)_0=1$, $(a)_r=a(a+1)(a+2)\cdots(a+r-1)$.

From the result (2.4), especially if r=1 and 2, then we obtain the mean and variance of S :

$$E(S) = \frac{(bk+n-k)}{b}\quad\text{and}\quad VAR(S)=\frac{(b^2k+n-k)}{b^2} \tag{2.5}$$

Let $Y_{(1)}=\min\{Y_1,\cdots,Y_n\}$ and $Y_{(n)}=\max\{Y_1,\cdots,Y_n\}$.

Then the pdf's of $Y_{(1)}$ and $Y_{(n)}$ are obtained as follows:

$$f_1(y_1) = (bn-bk+k)\cdot\exp(-(bn-bk+k)y_1),\quad y_1>0$$

and

$$f_n(y_n) = \frac{d}{dy_n}(1-e^{-y_n})^k(1-e^{-by_n})^{n-k},\qquad y_n>0\ .$$

And hence,

$$E(Y_{(1)}{}^r) = \frac{r!}{(bn-bk+k)^r}\ ,\quad r=1,2,\ \cdots\ . \tag{2.6}$$

Assume it were a wrong model, an assumed same parameter exponential model, denoted it by EXP(μ,σ). The MLE's of μ and σ are given by

$$\hat{\mu}=X_{(1)} \quad \text{and} \quad \hat{\sigma}=\overline{X}-X_{(1)}, \quad \overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i .$$

From the results (2.4) and (2.6), we can obtain the means and variances of $\hat{\mu}$ and $\hat{\sigma}$:

$$E(\hat{\mu}) = \mu+\frac{\sigma}{k+nb-kb}, \quad E(\hat{\sigma}) = \sigma\frac{(bk+n-k)(k+nb-kb)-nb}{nb(k+nb-kb)}, \tag{2.7}$$

$$VAR(\hat{\mu}) = \frac{\sigma^2}{(k+(n-k)b)^2} \text{ ,and } VAR(\hat{\sigma}) = [\frac{kb^2+n-k}{n^2b^2}+\frac{n-2}{n(k+(n-k)b)^2}]\sigma^2 . \tag{2.8}$$

From the result (2.7), unbiased estimators for μ and σ are given by

$$\hat{\mu}=(1+D)X_{(1)}-D\overline{X} \text{ or } X_{(1)}-D(\overline{X}-X_{(1)}) \text{ ,and } \hat{\sigma}=(k+nb-nk)\cdot D\cdot(\overline{X}-X_{(1)}),$$

where, $D=nb[(kb+n-k)(k+nb-kb)-nb]^{-1}$.

From the Appendix in Gross and et al (1986) and the result (2.8), the variances of $\hat{\mu}$ and $\hat{\sigma}$ are obtained as follows:

$$VAR(\hat{\mu})=\sigma^2[\frac{kb^2+n-k}{n^2b^2}D^2+\frac{n+(n-2)D^2}{n(nb-kb+k)^2}],$$

$$VAR(\hat{\sigma})=\sigma^2[\frac{kb^2+n-k}{n^2b^2}+\frac{n-2}{n(k+nb-kb)^2}]\cdot(k+nb-kb)^2\cdot D^2 . \tag{2.9}$$

The MLE of the variance in the assumed exponential model is given by

$$\hat{\sigma}^2=(\overline{X}-X_{(1)})^2 .$$

From the Appendix in Gross and et al (1986) and the results (2.4), (2.5), (2.6), and (2.7), the mean and variance of $\hat{\sigma}^2$ can be obtained as follows:

$$E(\hat{\sigma}^2)=\sigma^2[\frac{1}{n^2}(k(k+1)+\frac{2}{b}k(n-k)+\frac{1}{b^2}(n-k)(n-k+1))$$
$$+\frac{2}{n}(\frac{n-1}{(k+nb-kb)^2}-\frac{n-k+bk}{b(k+nb-kb)})]\equiv D_3\cdot\sigma^2 , \tag{2.10}$$

$$VAR(\hat{\sigma}^2)=\sigma^4\Big\{\frac{1}{n^4}[k(k+1)(k+2)(k+3)+\frac{4}{b}k(k+1)(k+2)(n-k)+\frac{6}{b^2}k(k+1)(n-k)(n-k+1)$$
$$+\frac{4}{b^3}k(n-k)(n-k+1)(n-k+2)+\frac{1}{b^4}(n-k)(n-k+1)(n-k+2)(n-k+3)]$$
$$-\frac{4}{n^3}(k+nb-nk)^{-1}[k(k+1)(k+2)+\frac{3}{b}k(k+1)(n-k)+\frac{3}{b^2}k(n-k)(n-k+1)$$
$$+\frac{1}{b^3}(n-k)(n-k+1)(n-k+2)]+(\frac{12}{n^2}-\frac{12}{n^3})(k+nb-nk)^{-2}[k(k+1)+\frac{2}{b}k(n-k)$$
$$+\frac{1}{b^2}(n-k)(n-k+1)]+(-\frac{24}{n}+\frac{48}{n^2}-\frac{24}{n^3})(k+nb-nk)^{-3}[k+\frac{1}{b}(n-k)]$$
$$+(24-\frac{72}{n}+\frac{72}{n^2}-\frac{24}{n^3})[k+nb-nk]^{-4}-D_3^2\Big\}\equiv D_4\cdot\sigma^4 . \tag{2.11}$$

Let $\tilde{\sigma}^2=(\overline{X}-X_{(1)})^2/D_3$. Then $\tilde{\sigma}^2$ is an unbiased estimator of σ^2, and hence

$$VAR(\tilde{\sigma}^2) = \frac{D_4}{D_3^{\,2}} \cdot \sigma^4.$$

For given $0<r_0<1$, the reliable life time of the assumed exponential model is $t(r_0)=\mu-\sigma\cdot\ln r_0$ (see Martz(1982)). The MLE of the reliable life time, $t(r_0)$, is given as follow: $\hat{t}(r_0)=X_{(1)}-\ln r_0\cdot(\overline{X}-X_{(1)})$.

From the results (2.7) and (2.8), the mean and variance of $\hat{t}(r_0)$ are obtained as

$$E(\hat{t}(r_0))=\mu+\frac{D-\ln r_0}{D\cdot(k+nb-kb)}\cdot\sigma$$

and

$$VAR(\hat{t}(r_0))=\left[\frac{n+(n-2)(\ln r_0)^2}{n(k+nb-kb)^2}-\frac{kb^2+n-k}{n^2b^2}(\ln r_0)^2\right]\cdot\sigma^2 \qquad (2.12)$$

An unbiased estimator of the reliable life time in the assumed exponential model is given by

$$\tilde{t}(r_0)=X_{(1)}-D\cdot[1+\ln r_0\cdot(k+nb-kb)]\cdot(\overline{X}-X_{(1)}).$$

From the result (2.8), the variance of $\tilde{t}(r_0)$ is obtained by

$$VAR[\tilde{t}(r_0)]=\sigma^2\left\{\frac{1}{(k+nb-kb)^{\;2}}+D^2[1+(\ln r_0)(k+nb-kb)\;]^2\cdot\left[\frac{n-2}{n(n+nb-kb)^2}+\frac{kb^2+n-k}{n^2b^2}\right]\right\}. \qquad (2.13)$$

3. The Numerical Comparison

From the results (2.7) through (2.14), Table shows the exact numerical evaluation of mean square errors of eight proposed estimators in an assumed same parameter exponential model when unidentified outliers exist, for the sample size n=30 and b=2(or 1/2).

In the sense of mean squared error, every proposed unbiased estimator will be more recommended than the MLE of the parameter in an assumed same parameter exponential model, although a few unidentified outliers exist.

Table. The mean square errors of several proposed estimators

Estimator b	#	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\sigma}$	$\tilde{\sigma}$	$\hat{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{t}(1/4)$	$\tilde{t}(1/4)$
1/2	0	.0022 (100)	.0012 (100)	.0355 (100)	.0368 (100)	.2800 (100)	.2973 (100)	.0673 (100)	.0684 (100)
	1	.0023 (96)	.0012 (100)	.0377 (94)	.0378 (97)	.3235 (87)	.2998 (99)	.0748 (90)	.0703 (97)
	2	.0024 (92)	.0012 (100)	.0421 (84)	.0386 (95)	.3805 (74)	.3015 (99)	.0865 (78)	.0717 (95)
	3	.0025 (88)	.0013 (92)	.0487 (73)	.0392 (94)	.4522 (62)	.3026 (98)	.1024 (66)	.0728 (94)
	4	.0026 (85)	.0013 (92)	.0574 (62)	.0397 (93)	.5397 (52)	.3033 (98)	.1225 (55)	.0737 (93)
	5	.0026 (85)	.0014 (86)	.0682 (52)	.0401 (92)	.6441 (43)	.3035 (98)	.1469 (46)	.0744 (92)
2	0	.0022 (100)	.0012 (100)	.0355 (100)	.0368 (100)	.2800 (100)	.2973 (100)	.0673 (100)	.0684 (100)
	1	.0021 (105)	.0011 (109)	.0359 (99)	.0370 (99)	.2648 (106)	.2965 (100)	.0676 (100)	.0689 (99)
	2	.0020 (110)	.0010 (120)	.0367 (97)	.0372 (99)	.2522 (111)	.2958 (101)	.0678 (99)	.0693 (99)
	3	.0018 (122)	.0010 (120)	.0381 (93)	.0375 (95)	.2424 (116)	.2953 (101)	.0684 (98)	.0698 (98)
	4	.0017 (129)	.0009 (133)	.0400 (89)	.0377 (94)	.2353 (119)	.2951 (101)	.0708 (95)	.0703 (97)
	5	.0016 (138)	.0008 (150)	.0425 (84)	.0379 (94)	.2307 (121)	.2951 (101)	.0744 (90)	.0708 (97)

Where, # is the number of outliers, and numbers in the brackets represent relative efficiencies with respect to the regular model for n=30

References

[1] Abramowitz, M. and Stegun, I.A.(1970). *Handbooks of Mathematical functions*, New York.

[2] Dixit, V.J.(1989), Estimation of Parameters of the Gamma Distribution in the Presence of Outliers, *Communications in Statistics, Theory and Method*, 19(8), 3071-3085.

[3] Dixit, V.J.(1991). On the Estimation of Power of the Scale Parameter in the Gamma Distribution in the Presence of Outliers, *Communications in Statistics, Theory and Method*, 20(4), 1315-1328.

[4] Gather, V. and Kale, B.K.(1988). MLE in the Presence of Outliers, *Communications in Statistics, Theory and Method*, 17(11), 3767-3784.

[5] Gradsheyn, I.G. and Ryzhik, I.M.(1980). *Tables of Integrals, Series and Products*, Academic Press, New York.

[6] Gross, A.J., Hunt, H.H., and Odeh, R.E.(1986). The Correlation Coefficient between the Smallest and Largest Observation when (n-1) of the n Observations are i.i.d., *Communications in Statistics, Theory and Method*, 15(4), 1113-1123.

[7] Martz, H.F. and Waller, R.A.(1982). *Bayesian Reliability Analysis*, John Wiley and Sons, New York.

[8] Oberhettinger, F.(1974). *Tables of Mellin Transforms*, Springer-Verlag, New York.

[9] Oberhrttinger, F. and Baddii, L.(1973). *Tables of LaplacesTransforms*, Springer-Verlag, New York.

[10] Rao, A.V., Kantam, R.R.L. and Navasimham, V.L.(1991). Linear Estimation of Location and and Scale Parameters in the Generalized Gamma Distribution, *Communications in Statistics,Theory and Method*, 20(12), 3823-2848.