응용통계연구 제9권 2호
1994년 9월 pp. 313-321

Jackknife Parametric Estimation in the Two Parameter Exponential Model with an Identified Outlier

Jungsoo Woo , Changsoo Lee[1)]

Abstract

When a single identified outlier in a small sample is presented, the small sample properties of the MLE's and its jackknife estimators of the location and scale parameters in an assumed exponential model will be considered by the method of permanent theory.

1. Introduction

The problem of estimating parameters in the presence of contaminated observations, which possibly occur as outlier has been studied extensively. Outliers may arise in a variety of life testing situation ; (i) the timing mechanism for one of the items fails, yielding an overestimated lifetime ; (ii) one of the items is accidently subjected to excessively low or incorrectly measured stress, yielding an excessively long life ; (iii) a failed item is replaced one or more items but the replacement is inadvertently overlooked.

Many authors (Kale & Sinha(1971), Chikkagouder & Kunchur(1980), Rauhut (1982), and Gather(1986)) considered the problem of estimation for the parameters of an exponential model in the presence of an outlying observation. Recently, Gather and Kale(1988) considered the maximum likelihood estimator(MLE) of parameters in the exponential family under k-outlier model, and Woo(1994) have considered parametric estimation of two-parameter exponential model in the presence of unidentified outliers.

Here, we consider the parametric estimation of the location and scale parameters in the two-parameter exponential model when a single identified identified outlier in a small sample is present. The exact mathematical formulas of density function and joint density function of order statistics with non-identically distributed variates in an assumed exponential model are derived by the method of permanent theory (Bapat and Beg(1989)). By the pdf of order statistics, we obtain means and variances for the MLE's and its jackknife esrimators of the location and scale parameters in the two-parameter exponential model with an identified outlier in the small sample. Through the numerical evaluation of

1) Department of Statistics, Yeungnam University, Kyungsan, 712-749, KOREA.

biases and mean square errors (MSE) of the MLE's and its jackknife estimators for the two parameters in an assumed exponential model with an identified outlier, the small sample properties for their estimators can be compared in the sense of bias and MSE.

2. Parametric estimation

In life testing research the most widely exploited model is the exponential model with p.d.f.

$$f(x\ ;\ \mu,\sigma)\ =\ \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right),\qquad x>\mu,\quad \mu,\ \sigma>0,$$

where μ and σ are reffered as the location and scale parameters, respectively, denoted it by X ~ EXP(μ,σ).

Consider the following situation. Suppose experimemtal animals are subjected to massive doses of radiation and their survival times are recorded. During the administering of the radiation, it is known that exactly one animal received a dose of radiation for inexcess to other. It is not known which animal is so subjected. It may then be assumed that the expected survival time of the overdoesed animal is different than the other (n-1) animals. Assuming these animals have survival times that are exponentially distributed, this is an experimemtial situation that is an example of the problem under consideration. Chikkagoudal and Kunchur(1980) and Gross, Hunt, and Odeh(1986) considered this situation with emphasis on the estimation of parameters.

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables where all but one of them are from EXP(μ, σ), but one remaining random variable is from EXP($\mu, b\cdot\sigma$), b is a positive constant. Before the start of the experiment, we have no prior knowledge as to which one of these n is the outlier. Let $Y_1 \le Y_2 \le \cdots \le Y_n$ be the order statistics of the sample. Then the marginal pdf's of Y_i, i = 1,2, $\cdots$,n, can be obtained by the method of permanent theory (cf. Vaughn and Venables, 1972).

Vaughn and Venables showed that the joint pdf of $Y_1, Y_2, \cdots, Y_n$ at the point $(y_1, y_2, \cdots, y_n)$ may be written as the permanent of the matrix of the marginal pdf's as follows; $f_{1,2,\cdots,n}(y_1,y_2,\cdots,y_n) = \text{per}\ |F|$, where $o \le y_1 \le y_2 \le \cdots \le y_n$ and F is the $n\times n$ matrix

$$F = \begin{pmatrix} f_1(y_1) & f_2(y_1) & \cdots & f_n(y_1) \\ \vdots & \vdots & \cdots & \vdots \\ f_1(y_n) & f_2(y_n) & \cdots & f_n(y_n) \end{pmatrix}$$

The permanent of an $n\times n$ square matrix A, denoted as per$|A|$, has the same definition as its determinant except that all signs are positive. Thus per | A | is the sum of n!-terms. Each term in this sum is formed as the product of n terms obtained by choosing an element from each row and column. Let f_o and F_o be the pdf and cdf of an identified outlier random variable, respectively. From the permanent theory, the pdf's of Y_i, i = 1,2, ⋯ ,n, are obtained as

$$f_{Y_i}(y_i)=\sum_{k=0}^{i-1}(-1)^k\binom{n-1}{i-1}\binom{i-1}{k}\left[\frac{b(n-i)+1}{b\sigma}\right]\exp\left\{-\frac{b(n-i+k)}{b\sigma}(x-\mu)\right\}$$
$$+\sum_{k=0}^{i-2}(-1)^k(i-1)\binom{n-1}{i-1}\binom{i-2}{k}\sigma^{-1}\left\{\exp\left\{-\frac{n-i+k+1}{\sigma}(x-\mu)\right\}\right.$$
$$\left.-\exp\left\{-\frac{b(n-i+k+1)+1}{b\sigma}(x-\mu)\right\}\right\},\qquad 0<y_i. \tag{2.1}$$

Therefore, the moment generating functions(mgf) of Y_i, i = 1,2, ⋯,n, are obtained as

$$M_{Y_i}(t)=\sum_{k=0}^{i-1}(-1)^k\binom{n-1}{i-1}\binom{i-1}{k}e^{t\mu}\left\{\frac{b(n-i)+1}{b(n-i+k)+1}\right\}\left\{1-\frac{b\sigma}{b(n-i+k)+1}t\right\}^{-1}$$
$$+\sum_{k=0}^{i-2}(-1)^k\binom{n-1}{i-1}\binom{i-2}{k}e^{t\mu}\left\{\frac{1}{n-i+k+1}\left(1-\frac{\sigma}{n-i+k+1}t\right)^{-1}\right.$$
$$\left.-\frac{b}{b(n-i+k+1)+1}\left(1-\frac{b\sigma}{b(n-i+k+1)+1}t\right)^{-1}\right\}. \tag{2.2}$$

From the mgf of Y_i, we can obtain the first and second moments of Y_i, i=1,2, ⋯ ,n, as follows ;

$$E(Y_i)=\sum_{k=0}^{i-1}(-1)^k\binom{n-1}{i-1}\binom{i-1}{k}\left\{\frac{b(n-i)+1}{b(n-i+k)+1}\right\}\left\{\frac{b\sigma}{b(n-i+k)+1}\right\}$$
$$+\sum_{k=0}^{i-2}(-1)^k(i-1)\binom{n-1}{i-1}\binom{i-1}{k}\left\{\frac{1}{n-i+k+1}\left(\frac{\sigma}{n-i+k+1}+\mu\right)\right. \tag{2.3}$$
$$\left.-\frac{b}{b(n-i+k+1)+1}\left(\frac{b\sigma}{b(n-i+k+1)+1}+\mu\right)\right\}\equiv\mu_{i:n}$$

$$
\begin{aligned}
E(Y_i^2) = & \sum_{k=0}^{i-1}(-1)^k \binom{n-1}{i-1}\binom{i-1}{k}\left\{\frac{b(n-i)+1}{b(n-i+k)+1}\right\} \\
& \left\{2\left(\frac{b\sigma}{b(n-i+k)+1}\right)^2+2\mu\frac{b\sigma}{b(n-i+k)+1}+\mu^2\right\} \\
+ & \sum_{k=0}^{i-2}(-1)^k (i-1)\binom{n-1}{i-1}\binom{i-1}{k} \\
& \left\{\frac{1}{n-i+k+1}\left[2\left(\frac{\sigma}{n-i+k+1}\right)^2+2\mu\left(\frac{\sigma}{n-i+k+1}\right)+\mu^2\right]\right. \\
& -\frac{b}{b(n-i+k+1)+1}\left[2\left(\frac{b\sigma}{b(n-i+k+1)+1}\right)^2\right. \\
& \left.\left.+2\mu\left(\frac{b\sigma}{b(n-i+k+1)+1}\right)+\mu^2\right]\right\}.
\end{aligned}
\tag{2.4}
$$

Therefore, we can obtain the variances of Y_i, i = 1,2, ⋯,n, denoted by $\sigma^2_{i:n}$.

Here we consider the parametric estimations of μ and σ in an assumed exponential model when a single identified outlier is present. In an assumed exponential model, the MLE's of μ and σ are given by

$$
\begin{aligned}
\hat{\mu} = & \; Y_1, \\
\hat{\sigma} = & \begin{cases} \dfrac{1}{n}\left\{\sum_{i=1}^{n-1}Y_i+\dfrac{1}{b}Y_n\right\}-\dfrac{1}{n}\left\{n-1+\dfrac{1}{b}\right\}Y_1, & 1<b \\ \dfrac{1}{n}\left\{\sum_{i=2}^{n}Y_i+\dfrac{1}{b}Y_1\right\}-\dfrac{1}{n}\left\{n-1+\dfrac{1}{b}\right\}Y_1 & 1>b. \end{cases}
\end{aligned}
\tag{2.5}
$$

(to applying the method of Gather and Kale(1988)).

From the results (2.3) and (2.4), we can obtain the means and variances of $\hat{\mu}$ and $\hat{\sigma}$ as follows ; $E(\hat{\mu}) = b\sigma/(b(n-1)+1)$,

$$
E(\hat{\sigma}) = \begin{cases} \dfrac{1}{n}\left\{\sum_{i=1}^{n-1}\mu_{i:n}+\dfrac{1}{b}\mu_{n:n}\right\}-\dfrac{1}{n}\left\{n-1+\dfrac{1}{b}\right\}\mu_{1:n}, & 1<b \\ \dfrac{1}{n}\left\{\sum_{i=2}^{n}\mu_{i:n}+\dfrac{1}{b}\mu_{1:n}\right\}-\dfrac{1}{n}\left\{n-1+\dfrac{1}{b}\right\}\mu_{1:n}, & 1>b \end{cases}
\tag{2.6}
$$

and

$$VAR(\hat{\mu}) = \left\{ \frac{b\sigma}{b(n-1)+1} \right\}^2, \tag{2.7}$$

$$VAR(\hat{\sigma}) = \begin{cases} \frac{1}{n^2} \left\{ \sum_{i=2}^{n-1} \{2(n-i)-1+\frac{2}{b}\}\sigma^2_{i:n} + \frac{1}{b^2}\sigma^2_{n:n} \right. \\ \quad \left. -(n+\frac{1}{b}-2)^2\left\{\frac{b\sigma}{b(n-1)+1}\right\}^2 \right\} & 1<b \\ \frac{1}{n^2}\left\{ \sum_{i=2}^{n} \{2(n-i)+1\}\sigma^2_{i:n} - (n-1)^2\left\{\frac{b\sigma}{b(n-1)}+1\right\}^2 \right\} & 1>b, \end{cases}$$

respectively.

By the definitions of ordinary jackknife, we can obtain the jackknife estimators of $\hat{\mu}$ and $\hat{\sigma}$ as follows ; $J(\hat{\mu}) = \frac{2n-1}{n}Y_1 - \frac{n-1}{n}Y_2$

$$J(\hat{\sigma}) = \begin{cases} \frac{1}{n}\{\sum_{i=1}^{n-1} Y_i + \frac{1}{b}Y_n\} - \frac{1}{n}\{2(n-1)+\frac{1}{b}\}Y_1 + \frac{1}{n}\{n-2+\frac{1}{b}\}Y_2, & 1<b \\ \frac{1}{n}\{\sum_{i=2}^{n} Y_i + \frac{1}{b}Y_1\} - \frac{1}{n}\{2(n-1)+\frac{1}{b}\}Y_1 + \frac{1}{n}\{n-2+\frac{1}{b}\}Y_2, & 1>b. \end{cases} \tag{2.8}$$

From the results (2.3) and (2.4), the means and variances of $J(\widehat{\mu})$ and $J(\widehat{\sigma})$ are obtained as follows; $E[J(\widehat{\mu})] = (2n-1)\mu_{1:n}/n - (n-1)\mu_{2:n}/n$,

$$E[J(\widehat{\sigma})] = \begin{cases} \frac{1}{n}\{\sum_{i=1}^{n-1}\mu_{i:n} + \frac{1}{b}\mu_{n:n}\} - \frac{1}{n}\{2(n-1)+\frac{1}{b}\}\mu_{1:n} \\ \quad + \frac{1}{n}\{n-2+\frac{1}{b}\}\mu_{2:n} & 1<b \\ \frac{1}{n}\{\sum_{i=2}^{n}\mu_{i:n} + \frac{1}{b}\mu_{1:n}\} - \frac{1}{n}\{2(n-1)+\frac{1}{b}\}\mu_{1:n} \\ \quad + \frac{1}{n}\{n-2+\frac{1}{b}\}\mu_{2:n} & 1>b \end{cases} \tag{2.9}$$

and

$$VAR[J(\widehat{\mu})] = \frac{2n-1}{n^2}\left\{\frac{b\sigma}{b(n-1)+1}\right\}^2\sigma^2_{1:n} + \left(\frac{n-1}{n}\right)^2\sigma^2_{2:n},$$

$$VAR[J(\hat{\sigma})] = \begin{cases} \frac{1}{n^2}\ \{\sum_{i=3}^{n-1}\{2(n-i)-1+\frac{2}{b}\}\sigma^2_{i:n}+\frac{1}{b^2}\sigma^2_{n:n} \\ \quad +(n-1+\frac{1}{n})(3n-7+\frac{3}{b})\sigma^2_{2:n} \\ \quad -(2n-3+\frac{1}{b})(2n-5+\frac{3}{b})\sigma^2_{1:n}\} & 1<b \\ \frac{1}{n^2}\ \{\sum_{i=3}^{n}\{2(n-i)+1\}\sigma^2_{i:n} \\ \quad +(n-1+\frac{1}{b})(3n-5+\frac{1}{b})\sigma^2_{2:n} \\ \quad -4(n-1)(n-2+\frac{1}{b})\sigma^2_{1:n}\} & 1>b. \end{cases} \tag{2.10}$$

3. Numerical comparisions

From the results (2.6) through (2.10), we can evaluate the numerical values of biases and MSE's for the MLE's and their jackknife estimators for the location and scale parameters in an assumed exponential model with the presence of an identified outlier in a small sample. The numerical values of biases and MSE's for the estimators are given in Tables 1 and 2, for n = 5(5)20, b = 2 (or 1/2) and $\mu=0$, $\sigma=1$. From the Tables, under the assumed exponential model and the small sample sizes, the jackknife estimators are more useful in the bias reduction than the MLE's of the two parameters, and in the sense of mean square error, the jackknife estimator $J(\hat{\mu})$ is more efficient than MLE $\hat{\mu}$, but the MLE $\hat{\sigma}$ is more efficient than the jackknife estimator $J(\hat{\sigma})$.

Appendix

Table 1. The biases and MSE's for the MLE and its jackknife estimator of the location parameter μ.($\mu=0$, $\sigma=1$)

n	b	$\hat{\mu}$		$J(\hat{\mu})$	
		Bias	MSE	Bias	MSE
5	2	0.22222	0.09876	-0.00317	0.01778
	1/2	0.16666	0.05555	-0.00666	0.01004
10	2	0.10526	0.02216	-0.00031	0.00211
	1/2	0.09091	0.01653	-0.00091	0.00157
15	2	0.06896	0.00951	-0.00008	0.00061
	1/2	0.02650	0.00781	-0.00027	0.00050
20	2	0.05128	0.00526	-0.00003	0.00025
	1/2	0.04716	0.00454	-0.00012	0.00022

Table 2. The biases and MSE's for the MLE and its jackknife estimator of the scale parameter σ.($\mu=0$, $\sigma=1$)

n	b	$\hat{\sigma}$		$J(\hat{\sigma})$	
		Bias	MSE	Bias	MSE
5	2	0.28960	0.21969	0.11460	0.26651
	1/2	0.26667	0.21200	0.16667	0.29472
10	2	0.16928	0.10875	0.07238	0.17546
	1/2	0.14091	0.10354	0.02979	0.11391
15	2	0.12376	0.07187	0.05948	0.07339
	1/2	0.09583	0.06833	0.02440	0.07274
20	2	0.09772	0.05247	0.04904	0.05282
	1/2	0.07136	0.04947	0.01872	0.05206

References

[1] Bapat, R.B. and Beg, M.I. (1989). Order Statistics for Nonidentically Distributed Variables and Permanents, *Sankhya,* Vol. 51, 79-93.

[2] Chikkagouder, M.S. and Kunchur, S.H. (1980). Estimation of the Mean of an Exponential Distribution in the Presence of an Outlier, *Canadian Journal of Statistics,* Vol. 8, 59-63.

[3] Gather, U. (1986). Robust Estimation of the Mean of the Exponential Distribution in the Outlier Situation, *Communications in Statisitcs, Theory and Methods,* Vol. 15, 2323-2345.

[4] Gather, U. and Kale, B.K. (1988). Maximum Likelihood Estimation in the Presence of Outliers, *Communications in Statisitcs, Theory and Method,* Vol. 17, 3767-3784.

[5] Gray, D.P. and Schicany, W.R. (1972). The Generalized Jackknife Statistics, Marcel Dekker, Inc., 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. Exponentially Distributed, *Communications in Statisitcs, Theory and Methods,* Vol. 15, 1113-1123.

[7] Kale, B.K. and Sinha, S.K. (1971). Estimation of Expected Life in the Presence of an Outlier Observation, *Technometrics,* Vol. 13, 755-759.

[8] Rauhut, B.O. (1982). Estimation of the Mean of an Exponential Distribution with an Outlying Observation, *Communications in Statistics, Theory and Methods,* Vol. 11, 1439-1452.

[9] Vaughn, R.J. and Venables, W.N. (1972). Permanent Expressions for Order Statistic Densities, *Journal of the Royal Statistical Society,* Ser. B, Vol. 34, 308-310.

[10] Woo, J.S. (1994). Parametric Estimation of Two-Parameter Exponential Model in the Presence of Unidentified Outliers, To appear in Korean Communications in Statistics, 1.

하나의 확실한 이상점을 갖는 지수모형에서 모수에 대한 잭나이프 추정

우정수, 이창수[1)]

요약

소표본에서 하나의 확실한 이상점을 갖는 지수모형에서 permanent 이론을 도입하여 표본들의 순서통계량의 분포를 정확한 형태로 유도하고, 이 결과를 이용하여 가정된 지수모형의 위치모수와 측도모수에 대한 최우추정량과 그 잭나이프추정량을 편의와 평균제곱오차면에서 두 추정량의 소표본 성질을 비교하였다.

1) (712-749) 경상북도 경산시 대동 214-1, 영남대학교 이과대학 통계학과.