The Korean Communications
in Statistics Vol. 5, No. 3, 1998
pp. 837-845

Effects of an Outlier for Estimators in a Uniform Distribution

Jungsoo Woo[1) , Changsoo Lee[2) and Jangchoon Lee[3)

Abstract

We shall propose several estimators and confidence intervals for the scale parameter in a uniform distribution with the presence of a unidentified outlier and obtain biases and mean squared errors for their proposed estimators. And we shall numerically compare the performances for the proposed several estimators of the sclae parameter. Also, we shall compare lengths of confidence intervals of the scale parameter in a uniform distribution through Monte Carlo methods.

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. Here we shall consider problems of estimation for the sclae parameter in a uniform distribution with the presence of a unidentified outlier.

Gather & Kale(1988) considered problems of estimating maximum likelihood estimator in the presence of outliers. Dixit(1991) studied the estimation for power of the scale parameter of the gamma distribution in the presence of outliers. Jeevanand & Nail(1994) and Woo & Ali(1996) considered parametric estimations for two parameter exponential model in the presence of unidentified outliers. Woo & Lee (1996) studied effects of identified outliers for parametric estimators in a Pareto distribution. Woo & Lee(1997 & 1998) studied problems for parametric estimations in an exponential distribution with an unidentified Pareto outlier and in a Pareto distribution with an unidentified exponential outlier.

In this paper, we shall propose several estimators and confidence intervals for the scale parameter in a uniform distribution with the presence of an unidentified outlier and derive

1. Professor, Department of Statistics, Yeungnam University, Kyungsan, 712-749, Korea
2. Full-Time Instructor, Department of Computer Information Sciences, Kyungwoon University, Gumi, 730-850, Korea
3 Assistant Professor, Department of Computer Sciences, Teagu Sciences Colledge, Teagu, Korea

exactly density functions and joint density functions for order statistics with nonidentically distributied random variates in a uniform distribution based on the permanent method. Also, we shall obtain the biases and mean squared errors(MSE) for their estimators and compare numerically efficiencies for the proposed several estimators of the scale parameter in a uniform distribution. Also, we shall compare lengths of confidence intervals for the scale parameter in a uniform distribution through Monte Carlo methods.

2. Estimates of Scale Parameter

The uinform distribution is given by

$$f(x;\theta)=\frac{1}{\theta},\quad 0<x<\theta, \tag{2.1}$$

where θ is the scale parameter, denoted by $UNIF(\theta)$. Gibbons(1974) investigated parametric estimators of the scale parameter in a population uniformly distributed over $(0,\theta)$. Fan(1991) studied properties of the order statistics of a uniform distribution over $(0,\theta)$.

We shall consider the following situation : Suppose $X_1,\cdots,X_n$ are independent random variables such that all but one of them are from $UNIF(\theta)$ and one remaining random variable is from $UNIF(a\theta)$, where $a(\neq 1)$ is known constant. Let $X_{(1)},\cdots,X_{(n)}$ be corresponding the order statistics for $X_1,\cdots,X_n$. Note that the complete and sufficient statistics for the scale parameter θ in an assumed uniform distribution with a unidentified outlier is $X_{(n)}$. Our goal is to estimate the scale parameter θ in a uniform distribution with the presence of an unidentified outlier.

From the permanent theory(Vaught et al(1972)), the density function of $X_{(i)}$, $1\le i\le n$, can be obtained as follow : Let $i\neq n$. For $a>1$,

$$f_i(x)=C(n,i)\frac{x^{i-1}(\theta-x)^{n-i-1}}{a\theta^n}\{n(a\theta-x)-(a-1)i\theta\},\quad 0<x<\theta, \tag{2.2}$$

for $a<1$,

$$f_i(x)=\begin{cases} C(n,i)\dfrac{x^{i-1}(\theta-x)^{n-i-1}\{n(a\theta-x)-(a-1)i\theta\}}{a\theta^n}, & 0<x<a\theta \\ C(n,i)\dfrac{(i-1)x^{i-2}(\theta-x)^{n-i}}{\theta^{n-1}}, & a\theta\le x<\theta. \end{cases} \tag{2.3}$$

And let $i=n$. For $a>1$,

$$f_n(x)=\begin{cases} \dfrac{nx^{n-1}}{a\theta^n}, & 0<x\le\theta \\ \dfrac{1}{a\theta}, & \theta<x<a\theta. \end{cases} \tag{2.4}$$

for $a<1$,

$$f_n(x)=\begin{cases}\dfrac{nx^{n-1}}{a\theta^n}, & 0<x\le a\theta\\[2mm] \dfrac{(n-1)x^{n-2}}{\theta^{n-1}}, & \theta<x<a\theta.\end{cases} \tag{2.5}$$

where $C(n,i)=\dfrac{(n-1)!}{(i-1)!\,(n-i)!}$.

Also, from the permanent theory, the joint density function for $X_{(i)}$ and $X_{(j)}$, $1\le i<j\le n$, can be obtained as follow : For $a>1$ and $j\ne n$,

$$\begin{aligned} f_{ij}(x,y)= &\ C(n,i,j)\{n(a\theta-y)-(a-1)j\theta\}\\ &\cdot\frac{x^{i-1}(y-x)^{j-i-1}(\theta-y)^{n-j-1}}{a\theta^n}, \qquad 0<x<y<\theta, \end{aligned} \tag{2.6}$$

for $a>1$ and $j=n$,

$$f_{in}(x,y)=\begin{cases}\dfrac{(n-1)!}{(i-1)!\,(n-i-1)!}\,\dfrac{nx^{i-1}(y-x)^{n-i-1}}{a\theta^n}, & 0<x<y\le\theta\\[2mm] \dfrac{(n-1)!}{(i-1)!\,(n-i-1)!}\,\dfrac{x^{i-1}(y-x)^{n-i-1}}{a\theta^n}, & 0<x<\theta<x<a\theta.\end{cases} \tag{2.7}$$

For $a<1$ and $i\ne 1$,

$$f_{ij}(x,y)=\begin{cases} C(n,i,j)\dfrac{x^{i-1}(y-x)^{j-i-1}(\theta-y)^{n-j-1}}{a\theta^n} & \\ \quad\cdot\{n(a\theta-y)-(a-1)j\theta\}, & 0<x<y<a\theta\\[2mm] C(n,i,j)\dfrac{x^{i-1}(y-x)^{j-i-2}(\theta-y)^{n-j}}{a\theta^n} & \\ \quad\cdot\{i(y-x)+(j-i-1)(a\theta-x)\}, & 0<x<a\theta<y<\theta\\[2mm] C(n,i,j)\dfrac{(i-1)x^{i-2}(y-x)^{j-i-1}(\theta-y)^{n-j-1}}{\theta^{(n-1)}}, & 0<a\theta<x<y<\theta. \end{cases} \tag{2.8}$$

for $a<1$ and $i=1$,

$$f_{1j}(x,y)=\begin{cases} C(n,1,j)\dfrac{(y-x)^{j-2}(\theta-y)^{n-j-1}}{a\theta^n} & \\ \quad\cdot\{n(a\theta-y)-(a-1)j\theta\}, & 0<x<y<a\theta\\[2mm] C(n,1,j)\dfrac{(y-x)^{j-3}(\theta-y)^{n-j}}{a\theta^n} & \\ \quad\cdot\{(y-x)+(j-2)(a\theta-x)\}, & 0<x<a\theta<y<\theta. \end{cases} \tag{2.9}$$

where $C(n, i, j) = \frac{(n-1)!}{(i-1)!\,(j-i-1)!\,(n-i)!}$.

Here we consider point estimations of the scale parameter in a uniform distribution with the presence of an unidentified outlier. In a uniform distribution, the ML estimator for the scale parameter θ is $\widehat{\theta}_1 = X_{(n)}$.

From the results (2.4) and (2.5), we can obtain the bias and MSE for $\widehat{\theta}_1$ as following :

$$BIAS[\widehat{\theta}_1] = \begin{cases} \left[-\frac{(a-1)n+a}{a(n+1)} + \frac{a^2-1}{2a} \right] \cdot \theta, & a>1 \\ -\frac{n-(a^n-1)}{n(n+1)}\,\theta, & a<1, \end{cases}$$

$$MSE[\widehat{\theta}_1] = \begin{cases} \left[\frac{(a-1)n(n+3)+2a}{a(n+1)(n+2)} + \frac{a^3-3a^2+2}{3a} \right] \cdot \theta^2, & a>1 \\ \left[\frac{2}{n(n+1)} + \frac{2a^n(n(a-1)-2)}{n(n+1)(n+2)} \right] \cdot \theta^2, & a<1. \end{cases} \tag{2.10}$$

Now, we consider the UMVUE for the scale parameter θ in an assumed uniform distribution as follows:

$$\widehat{\theta}_2 = \frac{n+1}{n} X_{(n)}.$$

From the results (2.4) and (2.5), we can obtain the bias and MSE for $\widehat{\theta}_2$ as following :

$$BIAS[\widehat{\theta}_2] = \begin{cases} \left[\frac{(a^2-1)(n+1)}{2an} - \frac{a-1}{a} \right] \cdot \theta, & a>1 \\ \frac{(a^n-1)}{n^2}\,\theta, & a<1, \end{cases}$$

$$MSE[\widehat{\theta}_2] = \begin{cases} \Big[\frac{(a-1)(n+1)[a^2(n+1)-(a+1)(2n-1)]}{3an^2} \\ \quad + \frac{(a-1)n(n+2)+1}{an(n+2)} \Big] \cdot \theta^2, & a>1 \\ \left[\frac{1}{n^2} + \frac{2a^n[(a-1)n+a-2]}{n^2(n+2)} \right] \cdot \theta^2, & a<1. \end{cases} \tag{2.11}$$

Next, we consider the minimum risk estimator(MRE) for the scale parameter θ in the assumed uniform distribution :

$$\widehat{\theta}_3 = \frac{n+2}{n+1} X_{(n)}.$$

From the results (2.4) and (2.5), we can obtain the bias and MSE for $\widehat{\theta}_3$ as following :

$$BIAS[\widehat{\theta}_3] = \begin{cases} \left[-\dfrac{(a-1)n(n+2)+a}{a(n+1)^2} + \dfrac{(a^2-1)(n+2)}{2a(n+1)} \right] \cdot \theta, & a>1 \\ \dfrac{(a^n-2)n+2(a^n-1)}{n(n+1)^2} \theta, & a<1, \end{cases}$$

$$MSE[\widehat{\theta}_3] = \begin{cases} \left[\dfrac{(a-1)(n+2)[a^2(n+2)-(a+1)(2n+1)}{3a(n+1)^2} + \dfrac{(a-1)n(n+2)+a}{a(n+1)^2} \right] \cdot \theta^2, & a>1 \\ \left[\dfrac{n^2+3n+4+2[(a-1)n-1](n+2)a^n}{n(n+1)^3} \right] \cdot \theta^2, & a<1. \end{cases} \quad (2.12)$$

By the definitions of the ordinary jackknife, we can obtain the jackknife estimator for $\widehat{\theta}_1$ as follows :

$$J(\widehat{\theta}_1) = \frac{2n-1}{n} X_{(n)} - \frac{n-1}{n} X_{(n-1)}.$$

From the results (2.2) through (2.8), we can obtain the bias and MSE of $J(\widehat{\theta}_1)$ as following :

$$BIAS[J(\widehat{\theta}_1)] = \begin{cases} \left[-\dfrac{2(a-1)n^2-(a-1)n+1}{an(n+1)} + \dfrac{(a^2-1)(2n-1)}{2an} + \dfrac{(a-1)(n-1)^2}{an^2(n+1)} \right] \cdot \theta, & a>1 \\ \left[-\dfrac{1}{n^2} + \dfrac{[(a-1)n^2+1]a^{n-1}}{n^2(n+1)} \right] \cdot \theta, & a<1, \end{cases}$$

$$MSE[J(\widehat{\theta}_1)] = \begin{cases} \left[\dfrac{4(a-1)n^4+(3-a)n^2+(3a-5)n+2}{an^2(n+1)(n+2)} + \dfrac{(a^3-1)(2n-1)^2}{3an^2} - \dfrac{(a^2-1)(2n-1)(2n^2-2n+1)}{an^3} - \dfrac{6(a-1)(n-1)^2}{an^2(n+1)(n+2)} \right] \cdot \theta^2, & a>1 \\ \left[\dfrac{2(n^2-n+1)}{n^3(n+1)} + \dfrac{n-1}{n^3} a^{n-1} - \dfrac{6n^2-6n+2}{n^3(n+1)} a^n + \dfrac{5n^3-5n+2}{n^3(n+1)(n+2)} a^{n+1} \right] \cdot \theta^2, & a<1. \end{cases} \quad (2.13)$$

Now, we consider a confidence estimation for the scale parameter θ in a uniform distribution with the presence of an unidentified outlier.
Since $X_{(n)}/\theta$ is a pivotal quantity, and has the density function as follows : For $a>1$

$$f(t)=\begin{cases}\frac{n}{a}t^{n-1}, & 0<x\le 1\\ \frac{1}{a}, & 1<x<a,\end{cases} \tag{2.14}$$

and for $a<1$

$$f(t)=\begin{cases}\frac{n}{a}t^{n-1}, & 0<x\le a\\ (n-1)t^{n-2}, & a<x<1.\end{cases} \tag{2.15}$$

So, from the results (2.14) and (2.15), we can obtain a 100(1- γ)% equal tail confidence interval for the scale parameter θ as follows : For $a>1$ and $2(1-1/a)\le\gamma$,

$$((a(2-\gamma)/2)^{-\frac{1}{n}}X_{(n)}\ ,\ (a\gamma/2)^{-\frac{1}{n}}X_{(n)}), \tag{2.16}$$

for $a>1$ and $2(1-1/a)>\gamma$,

$$((a(2-\gamma)/2)^{-1/n}X_{(n)}\ ,\ (a\gamma/2)^{-1}X_{(n)}).$$

For $a<1$ and $2(1-a^{n-1})\le\gamma$,

$$((a(2-\gamma)/2)^{-\frac{1}{n}}X_{(n)}\ ,\ (a\gamma/2)^{-\frac{1}{n}}X_{(n)},),$$

for $a<1$ and $2(1-a^{n-1})>\gamma$,

$$(((2-\gamma)/2)^{-\frac{1}{n}}X_{(n)}\ ,\ (\gamma/2)^{-\frac{1}{n-1}}X_{(n)}).$$

Also, we can obtain another 100(1- γ)% confidence interval for the scale parameter θ as follows : For $a>1$ and $1-1/a\le\gamma$,

$$(X_{(n)}\ ,\ [1-a(1-\gamma)]^{-\frac{1}{n}}X_{(n)}), \tag{2.17}$$

for $a>1$ and $1-1/a>\gamma$,

$$(a^{-1}X_{(n)}\ ,\ (a\gamma)^{-\frac{1}{n}}X_{(n)}).$$

For $a<1$ and $1-a^{n-1}\le\gamma$,

$$(X_{(n)}\ ,\ [1-a(1-\gamma)]^{-\frac{1}{n}}X_{(n)}),$$

for $a<1$ and $1-a^{n-1}>\gamma$,

$$(X_{(n)}\ ,\ (a\gamma)^{-\frac{1}{n-1}}X_{(n)}).$$

From the results (2.10) through (2.17), tables show the numerical values of mean squared errors for proposed four estimators and simulated values of confidence intervals for the scale parameter θ in an assumed uniform distribution with the presence of an unidentified outlier for sample sizes n=10(5)30, the scale parameter $\theta=1$ and various values of a. In the sense of mean squared error, the MRE $\widehat{\theta}_3$ is more efficient than other proposed estimators of the scale parameter in an assumed uniform distribution with an unidentified outlier as a approaches at 1. But as the values of a are larger than 1, the MLE $\widehat{\theta}_1$ is more efficient than other proposed estimators of the scale parameter. And as the values of a are smaller than 1, the UMVUE $\widehat{\theta}_2$ is more efficient than other proposed estimators of the scale parameter in an assumed uniform distribution with an unidentified outlier. From table 2, the lengths of second proposed interval of the scale parameter in an assumed uniform distribution with an unidentified outlier is shorter than that of equal tail confidence interval.

Table 1. MSE's for several proposed estimators of the scale parameter in an assumed uniform distribution with an unidentified outlier ($\theta = 1$).

a	n	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$J(\widehat{\theta}_1)$
1.0625	10	0.01433	0.00892	0.00872	0.01475
	15	0.00699	0.00430	0.00423	0.00765
	20	0.00415	0.00256	0.00253	0.00477
	25	0.00275	0.00172	0.00171	0.00332
	30	0.00197	0.00125	0.00124	0.00248
1.1250	10	0.01404	0.01074	0.01033	0.01836
	15	0.00711	0.00562	0.00548	0.01077
	20	0.00442	0.00366	0.00359	0.00763
	25	0.00311	0.00269	0.00265	0.00602
	30	0.00237	0.00214	0.00212	0.00508
1.2500	10	0.01628	0.01921	0.01818	0.03757
	15	0.01004	0.01232	0.01193	0.02865
	20	0.00763	0.00953	0.00933	0.02463
	25	0.00644	0.00809	0.00797	0.02253
	30	0.00578	0.00722	0.00714	0.02125
1.5000	10	0.03787	0.06083	0.05785	0.14288
	15	0.03268	0.04755	0.04633	0.13089
	20	0.03066	0.04172	0.04106	0.12524
	25	0.02967	0.03849	0.03809	0.12203
	30	0.02912	0.03646	0.03618	0.11997
2.0000	10	0.17424	0.26583	0.25620	0.70142
	15	0.17304	0.22937	0.22526	0.69019
	20	0.16883	0.21239	0.21013	0.68427
	25	0.16809	0.20261	0.20118	0.68069
	30	0.16767	0.19626	0.19528	0.67830
3.0000	10	0.89394	1.23167	1.19835	3.45130
	15	0.89134	1.11044	1.09592	3.49000
	20	0.89033	1.05424	1.04434	3.50774
	25	0.88984	1.01845	1.01331	3.51791
	30	0.88956	0.99615	0.99260	3.52451

Continued....

a	n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$J(\hat{\theta}_1)$
0.4000	10	0.01818	0.01000	0.01006	0.01654
	15	0.00833	0.00444	0.00446	0.00781
	20	0.00476	0.00250	0.00251	0.00453
	25	0.00307	0.00160	0.00160	0.00295
	30	0.00215	0.00111	0.00111	0.00208
0.5000	10	0.01817	0.00999	0.01006	0.01656
	15	0.00833	0.00444	0.00446	0.00781
	20	0.00476	0.00250	0.00251	0.00453
	25	0.00307	0.00160	0.00160	0.00295
	30	0.00215	0.00111	0.00111	0.00208
0.75000	10	0.01801	0.00990	0.00996	0.01860
	15	0.00832	0.00443	0.00445	0.00785
	20	0.00476	0.00249	0.00250	0.00454
	25	0.00307	0.00160	0.00160	0.00296
	30	0.00215	0.00211	0.00111	0.00208
0.8750	10	0.01738	0.00956	0.00959	0.01672
	15	0.00820	0.00437	0.00438	0.00792
	20	0.00473	0.00248	0.00248	0.00458
	25	0.00306	0.00159	0.00159	0.00297
	30	0.00214	0.00111	0.00111	0.00208
0.9375	10	0.01659	0.00912	0.00912	0.01597
	15	0.00796	0.00424	0.00425	0.00778
	20	0.00464	0.00243	0.00244	0.00456
	25	0.00303	0.00157	0.00157	0.00298
	30	0.00213	0.00110	0.00110	0.00209
0.96875	10	0.01597	0.00878	0.00875	0.01514
	15	0.00772	0.00412	0.00411	0.00750
	20	0.00453	0.00238	0.00237	0.00445
	25	0.00297	0.00154	0.00154	0.00293
	30	0.00209	0.00108	0.00108	0.00207

References

[1] 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*, Vol. 20(4), 1315-1328.

[2] Fan, D.Y.(1991). On a Property of the Order Statistics of the Uniform Distribution, *Communications in Statistics, Theory and Method*, Vol. 20, 1903-1909.

[3] Gibbons, J.D.(1974). Estimation of the unknown upper limit of uniform distribution, *Sankya, Series B*, Vol. 36, 29-40.

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

[5] Jeevanand, E.S. and Nail, N.U.(1994). Estimation of Exponential Parameters in the Presence of Outliers, *Biom. Journal*, Vol. 35, 471-478.

[6] 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.

[7] Woo, J.S. and Ali, M.M.(1996). Parametric Estimation of Two-parameter Exponential Model in the Presence of Unidentified Outliers, *Journal of Information & Optimization Sciences*, Vol. 17, 57-63.

[8] Woo, J.S. and Lee C.S.(1996). Effects of Identified Outliers for Parametric Estimators in a Pareto, *The Korean Communications in Statistics*, Vol. 3(1), 195-206.

[9] ________________(1997). Parametric Estimators in an Exponential Distribution with a Unidentified Pareto Outlier, To appear.

[10] ________________(1998). Parametric Estimators in a Pareto Distribution with a Unidentified Exponential Outlier, To appear.