

Estimations for a Uniform Scale Parameter in the Presence of an Outlier

Jungsoo Woo¹⁾, Changsoo Lee²⁾

Abstract

We shall propose several estimators and confidence intervals for the scale parameter in a uniform distribution with the presence of a generalized uniform outlier, and obtain mean squared errors(MSE) for their proposed estimators. And we shall compare numerical MSE's for the proposed several estimators of the scale parameter. Also, we shall compare numerically expected lengths of confidence intervals of the scale parameter in a uniform distribution with the presence of a generalized uniform outlier.

1. Introduction

The problem of outliers in random data sets is very interesting, important and common one. There are two basic mechanisms which give rise to samples which appear to have outliers. It is a matter of some importance which of the mechanisms generated any particular set of observations since this consideration certainly affects, or should affect, ones subsequent analysis of the data. Mechanism (i) : The data come from some heavy tailed distribution. There is no question that any observations is in any way erroneous. Mechanism(ii) : The data aries from two distributions. One of these, the basic distribution, generates good observations, while another, the contaminating distribution, generates contaminants. When this mechanism is appropriate, it may often be invoked in either of two ways. The first, mechanism (iia), specifies that in the sample of size n , exactly $n-k$ observations come form the basic distribution, and k from contaminating distribution. For this model to be of use presuppose a knowledge of k - the number of contaminants in the sample. More commonly, k is not known. A suitable model is the mechanism(iib) : with probability p , any given observation comes from the contaminating distribution. Here we shall consider problems of estimation for the sclae parameter in a uniform distribution with the presence of a generalized uniform 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

1) Professor, Department of Statistics & Mathematics, Yeungnam University, Kyungsan, 712-749, Korea

2) Full-Time Instructor, Department of Computer Information Engineering, Kyungwoon University,
Gumi, 730-850, Korea

gamma distribution in the presence of outliers. Woo & Ali(1996) considered parametric estimations for two parameter exponential model in the presence of unidentified outliers. Hossain & Nath(1997) compared several methods for estimating the parameters of the Burr XII distribution and investigated their performances in presence and in absence of outliers. Woo, Lee & Lee(1998) studied effects of an outlier for estimators in an uniform distribution.

In this paper, we shall propose several estimators and confidence intervals for the scale parameter in a uniform distribution with the presence of a generalized uniform outlier. And, we shall obtain mean squared errors(MSE) for their estimators and compare numerically efficiencies for proposed several estimators of the scale parameter in a uniform distribution. Also, we shall compare numerically expected lengths of proposed confidence intervals for the scale parameter in a uniform distribution with the presence of a generalized uniform outlier.

2. Estimates of Scale Parameter

The uniform distribution is given by

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

where θ is the scale parameter, denoted it by $UNIF(0, \theta)$. 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 the uniform distribution over $(0, \theta)$.

We shall consider the following situations : Assume X_1, \dots, X_n be independent random variables such that all but one of them are from $UNIF(0, \theta)$ and one remaining random variable is from $GUNIF(\alpha, \theta)$, where $GUNIF$ denotes a generalized uniform distribution with the density function ;

$$f(x; \alpha, \beta) = \frac{\alpha+1}{\beta^{\alpha+1}} x^\alpha, \quad 0 < x < \beta, \quad (\text{See Proctor(1987)}).$$

where α is known real number greater than -1.

Let $X_{(1)}, \dots, X_{(n)}$ be the corresponding order statistics for the random variables X_1, \dots, X_n .

Note that the complete and sufficient statistics for the scale parameter θ in an assumed uniform distribution is $X_{(n)}$. Our goal is considered several estimators of the scale parameter θ in a uniform distribution with the presence of a generalized uniform outlier based on $X_{(n)}$.

Let $Y_i = X_i/\theta$, $i = 1, \dots, n$. Then Y_1, \dots, Y_n are independent random variables such that all but one of them are from $UNIF(0, 1)$ and one remaining random variable is from $GUNIF(\alpha, 1)$. So let $Y_{(1)}, \dots, Y_{(n)}$ be the corresponding order statistics for Y_1, \dots, Y_n . Then

from the permanent theory(Vaught et al (1972), the density function of $Y_{(i)}$, $i=1, \dots, n$, is

$$f_i(y) = \binom{n-1}{i-1} \{ (i+\alpha)y^{i+\alpha-1}(1-y)^{n-i} \\ + (n-i)y^{i-1}(1-y)^{\alpha+1}(1-y)^{n-i-1}, \quad 0 < y < 1. \quad (2.1)$$

Especially, the density function of the largest order statistics $Y_{(n)}$ is

$$f_n(y) = (n+\alpha)y^{n+\alpha-1}, \quad 0 < y < 1,$$

which is $GUNIF(n+\alpha, 1)$.

And, from the permanent theory, the joint density function between $Y_{(i)}$ and $Y_{(j)}$,

$1 \leq i < j \leq n$, is

$$f_{i,j}(u, v) = C(n, i, j) \{ (i+\alpha)u^{i+\alpha-1}(v-u)^{j-i-1}(1-v)^{n-j} \\ + (j-i-1)u^{i-1}(v^{\alpha+1}-u^{\alpha-1})(v-u)^{j-i-2}(1-v)^{n-j} \\ + (\alpha+1)u^{i-1}(v-u)^{j-i-1}v^{\alpha}(1-v)^{n-j} \\ + (n-j)u^{i-1}(v-u)^{j-i-1}(1-v^{\alpha+1})(1-v)^{n-j-1} \}, \quad 0 < u < v < 1, \quad (2.2)$$

where $C(n, i, j) = (n-1)! / ((i-1)!(j-i-1)!(n-j)!)$.

Especially, the joint density function between the smallest order statistics $Y_{(1)}$ and the largest order statistics $Y_{(n)}$ is

$$f_{1,n}(u, v) = (n-1) \{ (1+\alpha)(v^{\alpha}+u^{\alpha})(v-u)^{n-2} \\ + (n-2)(v^{\alpha+1}-u^{\alpha+1})(v-u)^{n-3} \}, \quad 0 < u < v < 1.$$

From results (2.1) and (2.2), we can obtain moments of $Y_{(i)}^k$ and $Y_{(i)}^r \cdot Y_{(j)}^s$ as follows :

$$E(Y_{(i)}^k) = \frac{(k)_i}{(k)_n} - \frac{(\alpha+k)_i}{(\alpha+k+1)_n},$$

$$E(Y_{(i)}^r \cdot Y_{(j)}^s) = C(n, i, j)B(j+\alpha+r+s, n-j+1) \\ \cdot \{ (i+\alpha)B(i+\alpha+r, s+1) + (\alpha+1)B(i+r, \alpha+s+1) \\ + (j-i-1)[B(i+r, \alpha+s+2) - B(i+\alpha+r+1, s+1)] \} \quad (2.3) \\ + C(n, i, j)(n-j) \{ B(i+r, s+1)B(j+r+s, n-j) \\ - B(i+r, \alpha+s+2)B(j+\alpha+r+s+1, n-j) \},$$

where $(a)_b = \Gamma(a+b)/\Gamma(b)$ and $B(c, d)$ is a Beta function.

Here we consider point estimations of the scale parameter in a uniform distribution with the presence of a generalized uniform outlier. In the uniform distribution with an outlier, we shall define a first scale estimator which is the MLE for the scale parameter θ in the iid sample case as follows: $\widehat{\theta}_1 = X_{(n)}$.

From result (2.3), MSE for $\widehat{\theta}_1$ is

$$MSE(\widehat{\theta}_1) = \frac{2}{(n+\alpha+1)(n+\alpha+2)} \theta^2. \quad (2.4)$$

Next, we shall define a second scale estimator which is UMVUE for the scale parameter θ in the iid sample case as follows ; $\widehat{\theta}_2 = \frac{n+1}{n} X_{(n)}$.

From the result (2.3), MSE for $\widehat{\theta}_2$ is

$$MSE(\widehat{\theta}_2) = \frac{\alpha^3 + \alpha^2(n+2) + \alpha(n+1)^2 + n(n+1)^2}{n^2(n+\alpha+1)^2(n+\alpha+2)} \theta^2. \quad (2.5)$$

Also, we shall define a third scale estimator which is minimum risk estimator for the scale parameter θ in the iid sample case as follows ; $\widehat{\theta}_3 = \frac{n+2}{n+1} X_{(n)}$.

From the result (2.3), MSE for $\widehat{\theta}_3$ is

$$MSE(\widehat{\theta}_3) = \frac{(n+\alpha)(n+2)^2 + (\alpha-1)^2(n+\alpha+2)}{(n+1)^2(n+\alpha+1)^2(n+\alpha+2)} \theta^2. \quad (2.6)$$

As Johnson(1950) proposed estimator for the scale parameter θ in the uniform distribution over $(0, \theta)$, a fourth scale estimator is defined as follows ;

$$\widehat{\theta}_4 = 2^{1/n} \cdot X_{(n)}.$$

From result (2.1), we can obtain that

$$P\{ |\widehat{\theta}_4 - \theta| < |\widehat{\theta}_3 - \theta| \} = \begin{cases} \frac{2^{n+\alpha}}{\left(\frac{n+2}{n+1} + 2^{1/n}\right)^{n+\alpha}}, & n \leq 3 \\ 1 - \frac{2^{n+\alpha}}{\left(\frac{n+2}{n+1} + 2^{1/n}\right)^{n+\alpha}}, & n \geq 4. \end{cases} \quad (2.7)$$

From result (2.9), Table 1 shows numerical values of probability for sample sizes n=5(10)30

and various values of the shape parameter α . From Table 1, $\widehat{\theta}_4$ -estimator is a closer estimator of the uniform scale parameter θ with the presence of a generalized uniform outlier than $\widehat{\theta}_3$ -estimator, but from Table 2, $\widehat{\theta}_3$ -estimator has smaller MSE than $\widehat{\theta}_4$ -estimator at a neighborhood of 0.

Table 1. Probabilities for closeness criterion of $\widehat{\theta}_4$ with respect to $\widehat{\theta}_3$

$\alpha \backslash n$	5	10	15	20	25	30	50
-0.5	0.48258	0.52428	0.53926	0.54698	0.55169	0.55486	0.56128
0.5	0.55305	0.56006	0.56324	0.56501	0.56613	0.56691	0.56852
2.0	0.64118	0.60876	0.59689	0.59071	0.58693	0.58437	0.57916

From the result (2.3), MSE for $\widehat{\theta}_4$ is

$$MSE(\widehat{\theta}_4) = \left\{ \frac{(n+\alpha)2^{1/n}}{n+\alpha+2} - \frac{(n+\alpha)2^{1/n+1}}{n+\alpha+1} + 1 \right\} \theta^2. \quad (2.8)$$

Finally, as Gibbons(1975) proposed estimator for the scale parameter θ which is symmetric estimator in the iid sample case, we shall define a fifth scale estimator as follows ;

$$\widehat{\theta}_5 = X_{(1)} + X_{(n)}.$$

From the result (2.3), MSE for $\widehat{\theta}_5$ is

$$MSE(\widehat{\theta}_5) = \left\{ \frac{n-1}{n+1} + \frac{2(\alpha+1)}{n(n+\alpha+2)} - \frac{n+\alpha-1}{n+\alpha+1} \right. \\ \left. - \frac{2(\alpha+1)\Gamma(n)\Gamma(\alpha+1)}{(n+\alpha+1)(n+\alpha+2)\Gamma(n+\alpha+1)} \right\} \theta^2. \quad (2.9)$$

From results (2.4) through (2.9), estimators $\widehat{\theta}_i$ are asymptotically unbiased and MSE-consistent estimators for the scale parameter θ . Table 2 shows the numerical values of mean squared errors for proposed estimators for the scale parameter in an assumed uniform distribution with the presence of a generalized uniform outlier for sample sizes $n=5(5)25$, the scale parameter $\theta=1$ and various values of the shape parameter α . From Table 2, $\widehat{\theta}_2$ -estimator is more efficient than other proposed estimators of the scale parameter in an assumed uniform distribution with a generalized uniform outlier when values of α are smaller

than -0.5. When values of α are in a neighborhood of 0, $\widehat{\theta}_3$ -estimator is more efficient than other proposed estimators of the scale parameter. But, as values of α increase, $\widehat{\theta}_4$ -estimator is more efficient than other proposed estimators of the scale parameter. For all the sample sizes, $\widehat{\theta}_1$ -estimator of the scale parameter is worse in a sense of MSE than other proposed estimators when values for the shape parameter α are negative. But when values for the shape parameter α are positive, $\widehat{\theta}_5$ -estimator of the scale parameter is worse in a sense of MSE than other proposed estimators

Next, we shall consider a confidence interval for the scale parameter in an assumed uniform distribution with a generalized uniform outlier. From result (2.1), $T_n \equiv X_{(n)}/\theta$ is a pivotal quantity, we can obtain a $100(1-\gamma)$ % equal-tail confidence interval (CI_{eq}) for the scale parameter θ with a generalized uniform outlier as follows :

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

Therefore, the expected length of CI_{eq} for θ is

$$((\gamma/2)^{-\frac{1}{n+\alpha}} - (1-\gamma/2)^{-\frac{1}{n+\alpha}}) \cdot \frac{n+\alpha}{n+\alpha+1} \theta. \quad (2.10)$$

We can propose another $100(1-\gamma)$ % confidence interval (CI_a) for θ as follows ;

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

From result (2.3), the expected length of CI_a for θ is

$$(\gamma^{-\frac{1}{n+\alpha}} - 1) \cdot \frac{n+\alpha}{n+\alpha+1} \theta. \quad (2.11)$$

From results (2.10) and (2.11), Table 3 shows numerical values for expected lengths of CI_{eq} and CI_a for sample sizes $n=10(5)30$, levels of significance $\gamma=0.01, 0.05$ and 0.1 , scale parameter $\theta=1$ and various values of the shape parameter α . From Table 3, an expected length for CI_a is shorter than that for CI_{eq} for the uniform scale parameter with the presence a generalized uniform distribution.

Table 2. MSE's for proposed estimators of the scale parameter in an assumed uniform distribution with a generalized uniform outlier $GUNIF(\alpha, 1)$ ($\theta=1$).

α	n	MSE				
		$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_4$	$\widehat{\theta}_5$
-0.9	5	0.06428	0.03845	0.03902	0.03995	0.06012
	10	0.01784	0.00980	0.00985	0.01041	0.01664
	15	0.00822	0.00438	0.00439	0.00470	0.00768
	20	0.00471	0.00247	0.00246	0.00266	0.00440
	25	0.00305	0.00158	0.00158	0.00171	0.00285
-0.8	5	0.06203	0.03702	0.03742	0.03827	0.05579
	10	0.01750	0.00961	0.00965	0.01017	0.01578
	15	0.00812	0.00432	0.00433	0.00462	0.00735
	20	0.00467	0.00245	0.00246	0.00263	0.00424
	25	0.00302	0.00157	0.00157	0.00169	0.00276
-0.6	5	0.05787	0.03444	0.03452	0.03520	0.05098
	10	0.01686	0.00925	0.00926	0.00973	0.01509
	15	0.00791	0.00421	0.00422	0.00449	0.00716
	20	0.00458	0.00240	0.00241	0.00257	0.00417
	25	0.00298	0.00155	0.00155	0.00166	0.00273
-0.5	5	0.05594	0.03328	0.03321	0.03381	0.04970
	10	0.01656	0.00908	0.00908	0.00952	0.01500
	15	0.00782	0.00416	0.00416	0.00442	0.00717
	20	0.00453	0.00238	0.00238	0.00254	0.00419
	25	0.00296	0.00153	0.00153	0.00165	0.00275
-0.2	5	0.05071	0.03026	0.02975	0.03012	0.04796
	10	0.01569	0.00861	0.00857	0.00893	0.01505
	15	0.00753	0.00401	0.00400	0.00423	0.00728
	20	0.00441	0.00231	0.00231	0.00246	0.00428
	25	0.00289	0.00150	0.00150	0.00160	0.00282
0.2	5	0.04480	0.02709	0.02603	0.02612	0.04752
	10	0.01463	0.00806	0.00797	0.00822	0.01521
	15	0.00717	0.00383	0.00381	0.00399	0.00739
	20	0.00425	0.00223	0.00222	0.00235	0.00435
	25	0.00280	0.00146	0.00145	0.00155	0.00286
0.5	5	0.04102	0.02523	0.02378	0.02368	0.04753
	10	0.01391	0.00770	0.00757	0.00775	0.01524
	15	0.00692	0.00370	0.00367	0.00383	0.00740
	20	0.00413	0.00217	0.00216	0.00227	0.00436
	25	0.00274	0.00142	0.00142	0.00151	0.00286
1.0	5	0.03571	0.02285	0.02083	0.02043	0.04761
	10	0.01282	0.00717	0.00699	0.00705	0.01515
	15	0.00653	0.00351	0.00347	0.00357	0.00735
	20	0.00393	0.00208	0.00207	0.00216	0.00432
	25	0.00264	0.00138	0.00137	0.00144	0.00284

α	n	MSE				
		$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_4$	$\widehat{\theta}_5$
2.0	5	0.02777	0.02000	0.01697	0.01606	0.04761
	10	0.01098	0.00637	0.00608	0.00595	0.01473
	15	0.00584	0.00319	0.00313	0.00314	0.00713
	20	0.00362	0.00192	0.00190	0.00195	0.00421
	25	0.00246	0.00129	0.00128	0.00132	0.00277
4.0	5	0.01818	0.01818	0.01363	0.01194	0.04776
	10	0.00833	0.00541	0.00495	0.00446	0.01399
	15	0.00476	0.00275	0.00264	0.00250	0.00674
	20	0.00307	0.00169	0.00165	0.00162	0.00399
	25	0.00215	0.00115	0.00113	0.00113	0.00264
6.0	5	0.01282	0.01846	0.01282	0.01056	0.04844
	10	0.00653	0.00496	0.00437	0.00360	0.01360
	15	0.00395	0.00247	0.00233	0.00206	0.00648
	20	0.00264	0.00152	0.00147	0.00137	0.00383
	25	0.00189	0.00104	0.00102	0.00098	0.00254
8.0	5	0.00952	0.01942	0.01296	0.01027	0.04940
	10	0.00526	0.00478	0.00408	0.00310	0.01344
	15	0.00333	0.00231	0.00213	0.00176	0.00633
	20	0.00229	0.00141	0.00135	0.00119	0.00372
	25	0.00168	0.00097	0.00094	0.00086	0.00247

Table 3. Expected lengths of CI_{eq} and CI_a for the uniform scale parameter($\theta=1$).

γ	α	n	CI_{eq}	CI_a
0.01	-0.5	10	0.67509	0.56437
		15	0.41231	0.34971
		20	0.29673	0.25339
		25	0.23176	0.19868
		30	0.19013	0.16341
	0.5	10	0.59882	0.50264
		15	0.38251	0.32499
		20	0.28098	0.24016
		25	0.22203	0.19046
		30	0.18353	0.15781
	2.0	10	0.51200	0.43181
		15	0.34510	0.29385
		20	0.26025	0.22272
		25	0.20889	0.17933
		30	0.17446	0.15009

γ	α	n	CI_{eq}	CI_a
0.05	-0.5	10	0.42687	0.33542
		15	0.26937	0.21469
		20	0.19686	0.15796
		25	0.15513	0.12496
		30	0.12800	0.10338
	0.5	10	0.38214	0.30146
		15	0.25088	0.20029
		20	0.18680	0.15003
		25	0.14882	0.11995
		30	0.12368	0.09993
	2.0	10	0.33026	0.26176
		15	0.22746	0.18199
		20	0.17352	0.13953
		25	0.14027	0.11315
		30	0.11771	0.09516
0.1	-0.5	10	0.33052	0.24815
		15	0.21137	0.16100
		20	0.15545	0.11922
		25	0.12295	0.09468
		30	0.10170	0.07852
	0.5	10	0.29699	0.22388
		15	0.19718	0.15045
		20	0.14764	0.11334
		25	0.11802	0.09093
		30	0.09830	0.07593
	2.0	10	0.25780	0.19526
		15	0.17914	0.13699
		20	0.13730	0.10554
		25	0.11132	0.08584
		30	0.09361	0.07235

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] Hossain, A. M. and Nath, S. K.(1997). Estimation of Parameters in the Presence of Outlier for a Burr XII distribution, *Communications in Statistics, Theory and Method*, Vol. 26(3), 637–652.
 - [6] Johnson, N. L.(1950). On the Comparision of Estimators, *Biometrika*, Vol. 37, 281–287.
 - [7] Proctor, J.W.(1987), Estimation of Two Generalized Curves covering the pearson system, *Proceedings of ASA Computing Section*, pp. 287–292.
 - [8] Vaught, 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.
 - [9] 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.
 - [10] Woo, J. S., Lee, C. S., and Lee J. C.(1998). Effects of an Outlier for Estimators in a Uniform Distribution, *The Korean Communications in Statistics*, Vol. 5(3), 837–845.