The Effect of Outliers on the Estimation of the Parameters of the Exponentiated Exponential Distribution

Nahed Helmy

Faculty of Commerce

Al – Azhar University, Girls' Branceh

Abstract

This paper is concerned with estimating the parameters of the exponentiated exponential distribution using the maximum likelihood method in the presence of outliers. The effect of outliers on the estimation of the parameters of the exponentiated exponential distribution is studied. To satisfy this purpose random samples of different sizes are generated from the uniform distribution and transformed to the exponentiated exponential distribution.

Key words: *exponentiated exponential distribution, outliers*

1. Introduction

Ulhas (1989) derived the maximum likelihood estimators and moment estimators for samples from the Gamma distribution in the presence of outliers. These estimators are compared empirically when all the three parameters are unknown and when one of the three parameters is known; their bias and mean square error are investigated with the help of numerical technique.

Jeevanan and Nair (1993) discussed the estimation of the parameter and survival function of an exponential population. They assumed that the sample contains K outliers that are also exponentially distributed.

Brito, Chavez and *et.al.* (1997), studied the relationship between the connectivity of a mutual *k*-nearest-neighbor graph, and the presence of clustering structure and outliers for multivariate data sets. A test for detection of clustering structure and outliers is proposed and its performance is evaluated in simulated data.

Hossin and Nath (1997) dealt with unweighted least squares estimation of a two - parameter Burr XII distribution and compared the results with maximum likelihood. The performance of these estimates is examined with and without outliers through simulations studies. Also they obtained approximate confidence intervals for the parameters.

Dixit (2000) derived the maximum likelihood and moment estimators for samples from exponential distribution in the presence of outliers. These estimators are compared empirically when all the parameters are unknown; their bias is investigated with the help of numerical technique. He showed that these estimators are asymptotically unbiased.

Gupta and Kundu (2001) introduced and studied quite extensively the generalized exponential distribution. Generalized exponential distribution can be used as an alternative to gamma or Weibull distribution in many situations. They considered the maximum likelihood estimation of the different parameters of a generalized exponential distribution and discussed some of the testing of hypothesis problems. In this paper they considered five other estimation procedures and compared their performances through numerical simulations.

Nasiri (2010) dealt with the estimation of parameters of generalized exponential distribution with presence of outliers. The maximum likelihood and moment estimators are derived and compared empirically when all the parameters are unknown. The bias and mean square error are investigated with help of numerical techniques.

Deiri (2011) dealt with the problem of estimating reliability where Y has an exponential distribution with parameter λ and X has an exponential distribution in the presence of one outlier with parameters θ and β such that X and Y are independent. The moment and maximum likelihood estimators of reliability are derived and he found that the moment estimator of reliability is asymptotically unbiased estimator.

Dixit and Nooghabi (2011) derived the maximum likelihood and uniformly minimum variance unbiased estimators of the parameters of the probability density function, cumulative distribution function and *r*th moment for the Pareto distribution in the presence of outliers. It has been shown that maximum likelihood estimates are better than their uniformly minimum variance unbiased estimators . At the end these methods are illustrated with the help of real data from an insurance company.

Srivastava and Tanna (2011) proposed a preliminary test estimator of average life (scale parameter) in two parameter exponential distribution in presence of suspected outliers. They studied the risk properties of the estimators obtained using asymmetric loss function.

Deiri (2012) derived the moment and maximum likelihood estimators of parameters of the gamma distribution with presence of two outliers. These estimators are compared empirically when all the parameters are unknown. Their bias and mean square error are investigated with the help of numerical techniques. He showed that these estimators are asymptotically unbiased.

Makhdoom and Nasiri (2012) dealt with the estimation of parameters of the weighted exponential distribution with presence of outliers. The maximum likelihood and moment estimators are derived. These estimators are compared empirically using Monte Carlo simulation when all the parameters are unknown. The bias and mean square error are investigated with help of numerical techniques.

Maximum likelihood estimates of the parameters of the exponentiated exponential distribution are derived in Section 2. The effect of outliers on the estimation of the parameters of the exponentiated exponential distribution is examined in Section 3. Section 4 is devoted to a numerical study to illustrate the main results. Finally, main results and concluding remarks are introduced in Section 5.

2. Maximum Likelihood Estimates of the Parameters of the Exponentiated Exponential Distribution

The present section is devoted to estimate the parameters of the exponentiated exponential distribution using the maximum likelihood method.

The exponentiated exponential (EE) distribution with cumulative distribution function is given by

$$F(x,\theta,\lambda) = \left(1 - e^{-\lambda x}\right)^{\theta} \quad ,\theta,\lambda,x > 0 \qquad (2.1)$$

where $X \sim EE(\theta,\lambda)$ is used to denote that the random variable X follows an exponentiated exponential distribution with shape parameter θ and scale parameter λ.

The probability density function is then given by

$$f(x,\theta,\lambda) = \theta\lambda\left(1 - e^{-\lambda x}\right)^{\theta-1} e^{-\lambda x} \qquad (2.2)$$

In this section we are concerned with the maximum likelihood estimation (MLE) of the parameters θ,λ of $EE(\theta,\lambda)$. The related properties are discussed. Suppose that $X_1, X_2, \dots, X_n$ is a random sample of size n from EE distribution. The maximum likelihood function is given by

$$L(\theta,\lambda) = \theta^{n}\lambda^{n}\left(1 - e^{-\lambda x}\right)^{\theta-1} e^{-\lambda \Sigma x} \qquad (2.3)$$

The natural logarithm of the likelihood function is given by

$$\mathrm{Ln}(\theta,\lambda) = n\ln\theta + n\ln\lambda + (\theta - 1)\sum_{i=1}^{n} \ln\left(1 - e^{-\lambda X_i}\right) - \lambda \sum_{i=1}^{n} X_i \quad (2.4)$$

Differentiating equation (2.4) with respect to $\theta\ and\ \lambda$ and setting to zero we obtain

$$\frac{\partial L}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln\left(1 - e^{-\lambda X_i}\right) = 0 \qquad (2.5)$$

$$\frac{\partial L}{\partial \lambda} = \frac{n}{\lambda} + (\theta - 1)\sum_{i=1}^{n} \frac{X_i e^{-\lambda X_i}}{1 - e^{-\lambda X_i}} - \sum_{i=1}^{n} X_i = 0 \qquad (2.6)$$

The nonlinear likelihood equations (2.5) and (2.6) can be solved numerically .$\theta, \lambda$to obtain the MLE's of the parameters

3. The Effect of Outliers on the Estimation of the Parameters of the Exponentiated Exponential Distribution

To study the effect of outliers on the estimation of the parameters of the exponentiated exponential distribution, the following steps are introduced

1) The estimators $\hat{\theta}$ *and* $\hat{\lambda}$ of the parameters are obtained by solving equations(2.5) and (2.6) numerically

2) The main properties of the parameters $\hat{\theta}$ and $\hat{\lambda}$ are studied using the following formulae

$$\text{relative absolute bias } (\theta) = \frac{|bias|}{\theta} \qquad (2.7)$$

$$\text{relative absolute bias } (\lambda) = \frac{|bias|}{\lambda} \qquad (2.8)$$

$$\text{mean square error } (\theta) = \left(\text{bias}(\hat{\theta})\right)^2 + V(\hat{\theta}) \qquad (2.9)$$

$$\text{mean square error } (\lambda) = \left(\text{bias}(\hat{\lambda})\right)^2 + V(\hat{\lambda}) \qquad (2.10)$$

3) In the resulting data the lower and upper outliers are examined using the following formulae

Let the data set are ordered as follows

$$x_1 < x_2 < \cdots < x_n$$

The lower outlier, x_1, is defined as

$$x_1 = \frac{x_2 - x_1}{x_{n-1} - x_1} \qquad (2.11)$$

where

x_1: is the first value in the data set

x_2: is the second value in the data set

x_{n-1}: is the value of the $(n-1)$th order in the data set

The value of x_1 is compared with the tabulated value, $x_{n,\alpha}$ (Verma, *et. al.* (2006)) where n is the sample size and α is the level of significance. If x_1 is greater than $x_{n,\alpha}$, then x_1 is considered to be a lower outlier.

And the upper outlier x_n is defined as:

$$x_n = \frac{x_n - x_{n-1}}{x_n - x_1} \qquad (2.12)$$

where

x_n: is the largest value in the data set

The value of x_n is compared with the tabulated value, $x_{n,\alpha}$ (Verma *et. al.* (2006)) where n is the sample size and α is the level of significance. If x_n is greater than $x_{n,\alpha}$, then x_n is considered to be an upper.

4) If the data set contains a lower outlier, it is suggested to replace this value with x_2. And if the data set contains an upper outlier, it is suggested to replace this value with x_{n-1}.
5) The resulting data set, after the detection of outliers, is used to estimate the parameters $\hat{\theta}$ and $\hat{\lambda}$ of the exponentiated exponential distribution.
6) The main properties of the parameters $\hat{\theta}$ and $\hat{\lambda}$ are studied again.
7) If the relative absolute bias and the mean square error for the parameters are improved (i.e. have less values) then these outliers values will be appear to be inconsistent with the remainder of the data set. But if the relative absolute bias and the mean square error for the parameters are not improved , then the outlier values may be classified as a cluster that has a set of objects which may be treated as equivalent things in some way.

4. Numerical Study

Monte Carlo Simulation

Some numerical results based on the MLE's of $\mathrm{EE}(\theta,\lambda)$ distribution are obtained according the following steps

1) Given initial values θ_0, λ_0 , generate random samples of size $n(n = 30{,}50{,}100)$ from the $\mathrm{EE}(\theta,\lambda)$ by observing that if U is uniform $(0{,}1)$ then
$$X = \frac{-1}{\lambda}\ln\left(1 - \frac{\ln u}{\theta}\right)$$
which has the exponentiated exponential distribution.
2) For each sample size n and the initial values of shape and scale parameters (θ,λ) the MLE's of the parameters θ and λ can be obtained by solving the nonlinear equations (2.5) and (2.6) simultaneously using Mathcad iteration scheme.
3) Repeat steps (1) - (2) N times where N=10000 for $n = 30{,}50{,}100$.
4) The bias, variance and mean square error of the MLE's of the parameters (θ,λ) are computed by averaging over the N repetitions using equations (2.7) to (2.10).
The computation results are displayed in tables (4.1) and (4.2).
For the estimated parameters θ, λ it is assumed that the actual population values are $\theta = 3, \lambda = 1$.

Tables (4.1) and (4.2) display the estimation of the parameters, the variance, the relative absolute bias and the estimated mean square error of the maximum likelihood estimates of θ, λ for different sample sizes. In this section MLE's of the shape parameter, θ and scale parameter, λ of the exponentiated exponential distribution are obtained. MLE'S are found by solving the nonlinear equations (2.5) and (2.6) simultaneously.

4. Concluding Remarks

Table (4.1)

The Variance, Estimation, Relative Absolute Bias and Mean Square Error of the MLE's of the parameters of the Exponentiated Exponential Distribution $(\theta = 3, \lambda = 1)$

n	Case	parameter	Variance	Estimation	Relative absolute bias	MSE

30	Before treatment of outliers	θ	0.107	1.98	0.34	0.382
		λ	0.016	0.691	0.309	0.112
	After treatment of outliers	θ	0.181	2.03	0.323	0.374
		λ	0.016	0.707	0.293	0.105
50	Before treatment of outliers	θ	0.051	1.934	0.355	0.396
		λ	0.008	0.678	0.322	0.112
	After treatment of outliers	θ	0.055	1.961	0.346	0.378
		λ	0.009	0.688	0.312	0.106
100	Before treatment of outliers	θ	0.023	1.906	0.365	0.407
		λ	0.004	0.67	0.33	0.113
	After treatment of outliers	θ	0.024	1.919	0.36	0.397
		λ	0.004	0.674	0.326	0.11

Table (4.2)

The Variance, Estimation, Relative Absolute Bias and Mean Square Error of the MLE's of the parameters of the Exponentiated Exponential Distribution $(\theta = 3, \lambda = 0.5)$

n	Case	Parameter	Variance	Estimation	Relative absolute bias	MSE
	Before treatment of	θ	0.026	1.707	0.431	0.566

30	outliers	λ	0.0016	0.3	0.4	0.083
	After treatment of outliers	θ	0.027	1.73	0.423	0.547
		λ	0.0017	0.305	0.39	0.08
50	Before treatment of outliers	θ	0.015	1.693	0.436	0.574
		λ	0.00089	0.298	0.404	0.084
	After treatment of outliers	θ	0.016	1.706	0.431	0.563
		λ	0.00093	0.3	0.4	0.082
100	Before treatment of outliers	θ	0.0079	1.678	0.441	0.585
		λ	0.00044	0.296	0.408	0.084
	After treatment of outliers	θ	0.0079	1.685	0.438	0.579
		λ	0.0045	0.297	0.405	0.083

From tables (4.1) and (4.2) the following remarks are noticed:

For the sample sizes 30,50,100 the variance for the parameters θ, λ before the treatment of outliers is less than after the treatment of outliers but both the relative absolute bias and the mean square error before the treatment of outliers is larger than after the treatment of outliers.

As the sample size n increase the variance, the relative absolute bias and the mean square error of the parameters θ, λ decrease.

References

Brito, M.R. Chávez, E.L. Quiroz, A.J. and J.E. Yukich (1997) ,"Connectivity of the Mutual K – Nearest – Neighbor Graph in Clustering Outlier Detection" Statistics and Probability Letters, Vol. 35, No. 1, pp. 33 – 42.

Deiri, E.(2011)," Estimation of Reliability for Exponential Case in the Presence of One Outlier" Financial Mathematics and Applications Vol. 1 pp. 1 – 8.

***Deiri, E. (2012)*, "**Classical Estimator of Parameters of the Gamma Case with Presence of Two Outliers "Communications in Statistics - Simulation and Computation Vol. 41, Issue 5, pp. 590 - 597 .

Dixit ,U.J. (2000) ' Estimation of parameters of the exponential distribution in the presence of outliers generated from uniform distribution', International Conference on Statistics, Combinatorics and Related Areas - 7th International Conference of the Forum for Interdisciplinary Mathematics December 19-21, 2000.

Dixit ,U.J. and Nooghabi, M.J. (2011), "Efficient Estimation in the Pareto Distribution with the Personal Outliers" Statistical Methodology, Vol. 8, pp. 340 – 355.

Gupta, R.D. and Kaundu, D. (2001)," Generalized exponential distribution: different methods of estimations" Journal of Statistical Computation and Simulation, Vol. 69, pp. 315 – 337.

Hossian, A.M. and Nathm S.K. (1997) 'Estimation of Parameters in the Presence of Outliers for a Burr *XII* Distribution' Communication Statistical Theory - Methods, Vol. 26, no. 3, pp.637 - 652.

Jeevanad, E.S. and Nair, N.U. (1993),'Estimation of Exponential Parameters in the Presence of Outliers', Biometrical Journal, Vol.35, pp. 471 - 478.

Makhdoom,I. and Nasiri ,P.(2012), "Estimation Parameters of Weighted Exponential Distribution with Presence of Outliers Generated from Exponential Distribution" Journal of Applied Sciences, Vol.11, pp. 3811 – 3814.

Nasiri,P. (2010) 'Estimation of the Parameters of the Generalized Exponential Distribution in the Presence of One Outlier Generated from Uniform Distribution' , Applied Mathematical Sciences, Vol. 4, no. 48, pp. 2391 - 2404.

Srivastava, R. and Tanna, V. (2011), " Estimation of Average Life in Presence of Suspected Outliers in Exponential Life model Under Asymmetric Loss Function" Journal of Reliability and Statistical Studies ,Vol. 4 , pp. 107-117.

Ulhas J. Dixit (1989), "Estimation of parameters of the gamma distribution in the presence of outliers" Communications in Statistics – Theory and Methods, Vol. 18, pp. 3071 – 3085.

Verma, S.P. and Quiroz – Ruiz, A (2006),"Critical Values for Six Dixon Tests for Outliers in Normal Samples up to Sizes 100 and Applications in Science and engineering" Revista Mexicana de Ciencias Geologicas, Vol. 23, No. 2, pp. 133 - 161.