International Journal of Statistics and Probability; Vol. 5, No. 5; September 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Quartile Method Estimation of Two-Parameter Exponential Distribution Data with Outliers

Entisar A. Elgmati[1], Nadia B. Gregni[1]

[1] Department of Statistics, Faculty of Science, University of Tripoli, Tripoli, Libya

Correspondence: Entisar A. Elgmati, Department of Statistics, University of Tripoli, Tripoli, P.O.Box 13008, Libya. E-mail: eelgmati@uot.edu.ly, eelgmati@hotmail.com

Received: May 26, 2016 Accepted: June 19, 2016 Online Published: August 9, 2016

doi:10.5539/ijsp.v5n5p12 URL: http://dx.doi.org/10.5539/ijsp.v5n5p12

Abstract

Several methods have been used to estimate the unknown parameters in the two-parameter exponential distribution. Here we have considered two of these methods, maximum likelihood method and median-first order statistics method. However, in the presence of outliers these methods are not valid. In this paper we propose two approaches that deal with this situation. The idea is based on using first and third quartile instead of the minimum statistics. We investigated the parameters estimate using these methods through simulation study. The new method gives similar results under the normal situation and much better results when the data has outliers.

Keywords: exponential distribution, maximum likelihood estimation, median method, quartile method.

1. Introduction

The two-parameter exponential distribution is widely used in applied statistics since it has many applications in real life. It can be used to model the data such as the service times of agents in a system, the extreme values of annual snowfall or rainfall, the time it takes before your next telephone call, the time until a radioactive particle decays, and the distance between mutations on a DNA strand. Also it is well-suited to model the constant hazard rate portion of the bathtub curve used in reliability theory, etc.

The probability density function of the two-parameter exponential distribution EXP(α, β), where α is a scale parameter and β is a location parameter, is given by

$$\begin{aligned} f(x;\alpha,\beta) &= \frac{1}{\alpha} e^{-(x-\beta)/\alpha} \quad , \quad x \ge \beta \quad \alpha > 0 \\ &= 0 \quad , \quad otherwise \end{aligned} \tag{1}$$

Estimation, predictions and inferential issues for the exponential distribution have been studied by several authors. Cohen and Helm (1973) used several methods (BLUE, MLE, ME, MVUE and MME) to estimate the parameters. The robust M-estimation method used by Peter (1974) to estimate the scale parameter. Lawless (1977) studied a confidence interval for the scale parameter and obtained a prediction interval for a future observation. Two new classes of confidence interval for the scale parameter proposed by Petropoulos (2011). While Lai and Augustine (2012) obtained interval estimates for the threshold (location) parameter and derived a predictive function for the two parameter. Afify (2004) and Muhammad and Ahmed (2011) reviewed and compared several methods for estimating the two-parameter exponential distribution.

In this paper we study the parameters estimate of the two-parameter exponential distribution in the presence of the outliers. A method based on the quartiles is introduced to deal with this situation. Also comparisons between this method and other methods, maximum likelihood and median-first order statistics are considered. Simulation studies are used to illustrate the accuracy of the proposed method estimates and to compare the method with other methods used to estimate the parameters of the exponential distribution with and without outliers in the data.

2. Methodology

2.1 Maximum Likelihood Estimators (MLE)

Let $X_1, X_2, ..., X_n$ be a random sample of size n from exponential distribution as defined in (1). The maximum likelihood estimates of α and β are

And

$$\hat{\alpha} = \overline{X} - \frac{\hat{\beta}}{n}$$

$$\hat{\beta} = \min(X_i) = X_{(1)}$$

2.2 Median-First Order Statistics Method (MOS)

The Median first order statistics method was proposed by Afify (2004), to estimate the unknown parameters in (1). This method is based on the assumption that

$$E(X_{(1)}) = X_{(1)}$$

Thus

$$X_{(1)} = \beta + \frac{\alpha}{n} \tag{2}$$

then solving (2) with the median equation $X_{med} = \beta + \alpha \ln 2$ yields the following estimates

$$\hat{\alpha} = \frac{med - X_{(1)}}{\ln 2 - \frac{1}{n}}$$

and

$$\hat{\beta} = \frac{med - n\,X_{(1)} ln2}{1 - n \ln 2}$$

2.3 Quartile Method (QM)

Since the previous two methods uses the smallest observation $X_{(1)}$ to estimate the location parameter β and the scale parameter α, therefore in the case that the minimum statistics is an outlier the estimates will be invalid. Our method is based on two approaches, the first one uses third and first quartile (Q_3&Q_1) while the second one based on third quartile and the median (Q_3&M).

The first and third quartile of two-parameter exponential distribution is given by:

$$Q_1 = \beta - \alpha \ln(\frac{3}{4}) \tag{3}$$

$$Q_3 = \beta + \alpha \ln 4 \tag{4}$$

Now solving (3) and (4) give the following estimates:

$$\hat{\alpha} = \frac{Q_3 - Q_1}{\ln 3}$$

$$\hat{\beta} = Q_3 + \hat{\alpha} \ln 4$$

Where Q_1 and Q_3 are the first and third quartile of the sample respectively.

Now considering the median and the third quartile case, the median can be written as

$$med = \beta + \alpha \ln 2 \tag{5}$$

From (4) and (5) we get

$$\hat{\alpha}^* = \frac{Q_3 - med}{\ln 2}$$

$$\hat{\beta}^* = 2med - Q_3$$

Where med is median of the sample.

3. Simulation Study

In this simulation study all our results are based on 1000 simulations in each case. Samples of size n=(20, 50, 100, 150) were simulated from the exponential distribution with different values for α and β but we only present the results for the values of (α , β) being (1,2) because of similarity.

Table 1 gives the estimates of the two parameters (α , β) using MLE, MOS and QM methods with the mean square errors (MSE's) being in parenthesis. Also the total deviation (TD) mentioned in Afify (2004) was used as a measurement to compare these methods. From the table, one can notice that our proposed method, either using Q3&M or Q3&Q1, gives estimates that too close to the true value especially when the sample size increases. However comparing the proposed method with the other methods being included in this paper, the other ones have smaller MSE's and TD's. But since the regular MLE and the MOS chooses the minimum of the sample to estimate location parameter, this leads to a problem in estimating when the minimum of the sample is an outlier. Thus another simulation study was done with an outlier being included in the data and the results of the estimates using the methods presented earlier with their MSE's and TD's are reported in Table 2.

Table 1. The estimates and MSE's of the two-parameters for different sample size (n) using MLE, MOS and QM methods.

$\alpha=1$ $\beta=2$	MLE		MOS		QM Q3&M		QM Q3&Q1	
n	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
20	0.9674 (0.0519)	2.0496 (0.0049)	1.0548 (0.1202)	1.9969 (0.0027)	0.9366 (0.1798)	2.0788 (0.1341)	0.9495 (0.1013)	2.0609 (0.0291)
TD	**0.0574**		**0.05325**		**0.10280**		**0.08095**	
50	0.9811 (0.0201)	2.0192 (0.0007)	1.0136 (0.0446)	1.9989 (0.0004)	0.9746 (0.0794)	2.0259 (0.0588)	0.9778 (0.0435)	2.0215 (0.0112)
TD	**0.0285**		**0.01305**		**0.03835**		**0.03295**	
100	0.9893 (0.0101)	2.0098 (0.0002)	1.0069 (0.0232)	1.9997 (0.0001)	0.9857 (0.0418)	2.0144 (0.0316)	0.9869 (0.0226)	2.0127 (0.0057)
TD	**0.0156**		**0.00675**		**0.02150**		**0.01945**	
150	0.9912 (0.0065)	2.0064 (0.0001)	1.0062 (0.0128)	1.9997 (0.0001)	0.9760 (0.0269)	2.0206 (0.0181)	0.9839 (0.0149)	2.0097 (0.0035)
TD	**0.0120**		**0.00605**		**0.03430**		**0.02095**	

From Table 2, it is clear that for all sample sizes being used, the MLE and MOS give invalid estimate for both parameters while the QM method performed very well. Therefore in case of outliers our proposed method has strong superiority over the other methods since it does not affected by outliers in finding the parameters estimate. Note that the proposed method works well also when the maximum value of the sample is an outlier.

Table 2. The estimates and MSE's for the two-parameters, when there is an outlier in the data, using MLE, MOS and QM methods and different sample size (n).

$\alpha=1$ $\beta=2$	MLE		MOS		QM Q3&M		QM Q3&Q1	
n	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
20	3.0261 (4.1572)	0.001 (3.9960)	4.1763 (10.1988)	-0.2078 (4.8747)	0.9479 (0.1829)	2.0299 (0.1317)	0.9619 (0.0987)	2.0105 (0.0257)
TD	**3.0256**		**4.28020**		**0.06705**		**0.04335**	
50	2.9895 (3.9789)	0.001 (3.9960)	3.9728 (8.8804)	-0.0785 (4.3199)	0.9676 (0.0736)	2.0046 (0.0529)	0.9707 (0.0409)	2.0002 (0.0094)
TD	**2.9890**		**4.01205**		**0.03470**		**0.02940**	
100	3.001 (4.0140)	0.001 (3.9960)	3.9286 (8.5968)	-0.0383 (4.1546)	0.9882 (0.0423)	1.9998 (0.0302)	0.9893 (0.0218)	1.9984 (0.0052)
TD	**3.0005**		**3.94775**		**0.01190**		**0.01150**	
150	2.9987 (4.0018)	0.001 (3.9960)	3.9190 (8.5353)	-0.0251 (4.1011)	0.9887 (0.0296)	2.0060 (0.0211)	0.9899 (0.0158)	2.0043 (0.0037)
TD	**2.9982**		**3.93155**		**0.01430**		**0.01225**	

4. Conclusion

In this paper, we have considered the estimation problem of the unknown parameters of two-parameter exponential distribution when there is an outlier in the data. The maximum likelihood method and median-first order statistics method in addition to our proposed method (quartile method) were used for estimating the unknown location and scale parameters of two parameter exponential distribution. A comparison between these methods were done through simulation study. It is observed that the performance of the quartile method was satisfactory under the normal situation and become more accurate when sample size increases. Furthermore it worked very well when an outlier presents in the data while other methods were invalid.

References

Afify, E. E. (2004). Linear and Nonlinear regression of exponential distribution. *Inter. Statist., online Statistics Journal.*

Cohen, A. C., & Helm, F. R. (1973). Estimation in Exponential Distribution. *Technometrics*, *15*, 415-418. http://dx.doi.org/10.1080/00401706.1973.10489054

Lai, J., & Augustine, C. W. (2012). Interval estimation of the two-parameter exponential distribution. *Journal of Probability and Statistics*, Article ID 734575. http://dx.doi.org/10.1155/2012/734575

Lawless, J. F. (1977). Prediction intervals for the two parameter exponential distribution. *Technometrics*, *19*(4), 469-472. http://dx.doi.org/10.1080/00401706.1977.10489587

Muhammad, Z. R., & Ahmed, S. A. (2011). Estimation accuracy of exponential distribution parameters. *Pakistan Journal of Statistics and Operation Research*, *7*(2), 217-232.

Peter, F. T. (1974). Huber-Sense Robust M-Estimation of a Scale Parameter, with Application to the Exponential Distributio. *JASA*, *74*, 147-152.

Petropoulos, C. (2011). New classes of improved confidence intervals for the scale parameter of a two parameter exponential distribution. *Statistical Methodology*, *8*(4), 401-410. http://dx.doi.org/10.1016/j.stamet.2011.03.002

Copyrights