International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:6, No:1, 2012

Confidence Intervals for Double Exponential Distribution: A Simulation Approach

M. Alrasheedi*

***Abstract*—**The double exponential model (DEM), or Laplace distribution, is used in various disciplines. However, there are issues related to the construction of confidence intervals (CI), when using the distribution.In this paper, the properties of DEM are considered with intention of constructing CI based on simulated data. The analysis of pivotal equations for the models here in comparisons with pivotal equations for normal distribution are performed, and the results obtained from simulation data are presented.

***Keywords*—**Confidence intervals, double exponential model, pivotal equations, simulation

I. Introduction

THERE are many different probability distributions with analytical probability density functions (PDF) that are used according to the research discipline. The most commonly applied PDF is normal distribution, which describes a wide range of different processes. However, many processes can be described more precisely using the double exponential model (DEM) [1-3]. Recently, interest in the Laplace distribution has grown due to its potential transforming applicationin financial functions[3]. For instance, the difference of two independent two parameter exponential variables follows double exponential distribution, and the logarithm of ratios of uniform or Pareto variables follows the DEM, as well [1].

DEM has not been used extensively as a model due, in part, to the lack of available statistical techniques available for this distribution. From the experimenter's point of view, DEM is often not used because of a sharp peak in the center of the PDF; however,despite this anomaly, it is often the preferred model with exponential tails [1],[3]. In addition, it has been suggested that to investigate the properties of real data,we need to first perform several goodness-of-fit tests. An example of one such test is described by [3].

While the evaluation of CI is important, it is trivial in cases of normally distributed values [1]. In such, we can use auxiliary distributions, such as Student, Laplace, and χ^2. The goal of the present study is to analyze evaluation methods of CI for DEM. In addition, DEM data simulation, pivot construction, and analysis of the results are also performed in this study.

II. Double Exponential Distribution Analysis

The probability density function of DEM is defined as

$$f(x,\eta,\theta) = \frac{1}{2\theta} e^{-\frac{|x-\eta|}{\theta}} \qquad (1)$$

DEM is specific class of distribution:

$$f(\beta) = \frac{1}{\Gamma\left[1+\frac{1}{2}(1+\beta)\right] 2^{1+\frac{1}{2}(1+\beta)}\theta} \exp\left\{-\frac{1}{2}\left|\frac{x-\eta}{\theta}\right|^{\frac{2}{1+\beta}}\right\} \qquad (2)$$

This class of distributions includes normal ($\beta = 0$) and DEM ($\beta = 1$), which is symmetrical in scale and location parameters. Thus, we can compare the PDFs of these two distributions:

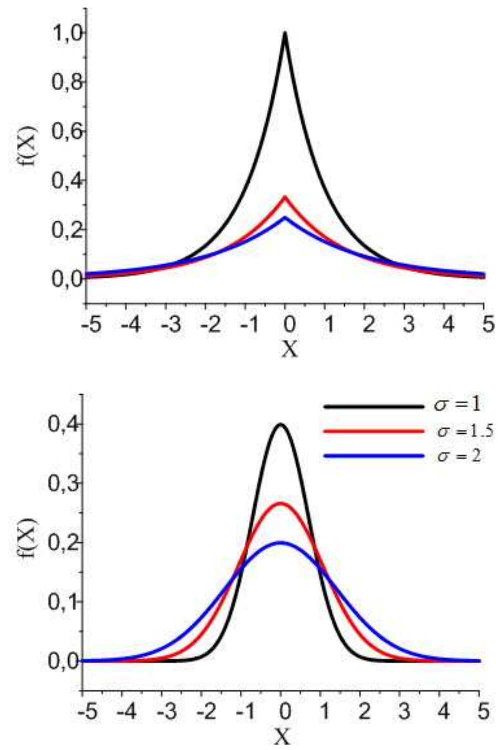

Fig. 1 Probability density functions of DEM and Gauss distributions

One can see from Figure 1 that the DEM has heavier tails than the normal distribution. It is therefore evident that increasing scale parameters leads to agradual decreasing of probability density function.

III. Simulation of Data with Double Exponential Probability Density Function

To generate the numbers distributed with double exponentials, we require a simple random number generator

* M. Alrasheedi is with the School of Business, King Faisal University, Hofuf, Alhasa 31982 Saudi Araiba (phone: 966-5-50335999; fax: 966-3-5812530; e-mail: malrasheedy@ kfu.edu.sa).

ISSN: 2517-9934
Vol:6, No:1, 2012

and the equation for the probability density function. The distribution function for DEM is:

$$F(x)=\begin{cases}\frac{1}{2}e^{\frac{x-\eta}{\theta}}, x\le\eta\\ 1-\frac{1}{2}e^{-\frac{x-\eta}{\theta}}, x>\eta\end{cases} \quad (3)$$

Now, we can find X from the equation for $F(x)$:

$$x=\begin{cases}\theta\ln F+\eta, F<1/2\\ -\theta\ln(2(1-F))+\eta, F>1/2\end{cases} \quad (4)$$

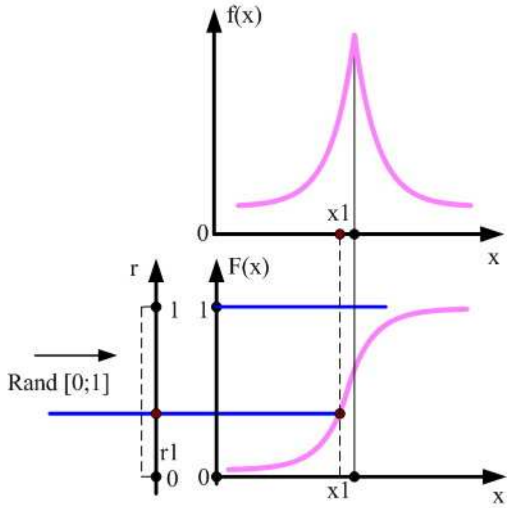

Fig. 2 DEM generation scheme

For each random number, we simply calculate the x value using the above equation. This scheme generation is shown in Figure 2. We can use this scheme in cases when the probability density function is defined analytically and it is simple to derive the equation for variable x. Next, the random numbers in the interval can be easily generated. To check this DEM generating algorithm, the theoretical and experimental results for DEM with $\eta=15$ and $\theta=3$, which should generate the theoretical PDF as shown in (1), are compared. For experiment 1, we'll use the method described above.

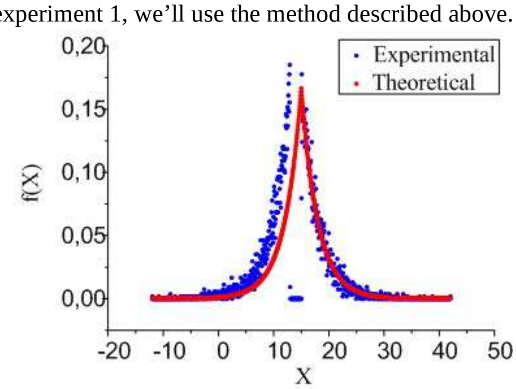

Fig. 3 Theoretical and experimental probability density functions for $\eta=15,\theta=3$

The correlation of the experimental and theoretical results confirms the accuracy of our approach.

IV. Analysis of DEM Data and Construction of Confidence Intervals

A. Evaluation of Pivots

After we have generated the data with DEM, consider the time series obtained from our data using the following equations:

$$W=\frac{\tilde{X}-\eta}{\frac{1}{n}\sum\left|X_i-\tilde{X}\right|}, \quad V=\frac{1}{n\theta}\sum\left|X_i-\tilde{X}\right| \quad (5)$$

In this experiment, $\tilde{X}$ is the median value. In the case of even data, the array length is defined as:

$$\tilde{X}=\frac{1}{2}\left(Array[N/2]+Array[N/2+1]\right) \quad (6)$$

The distribution of W is shown in Figure 4.

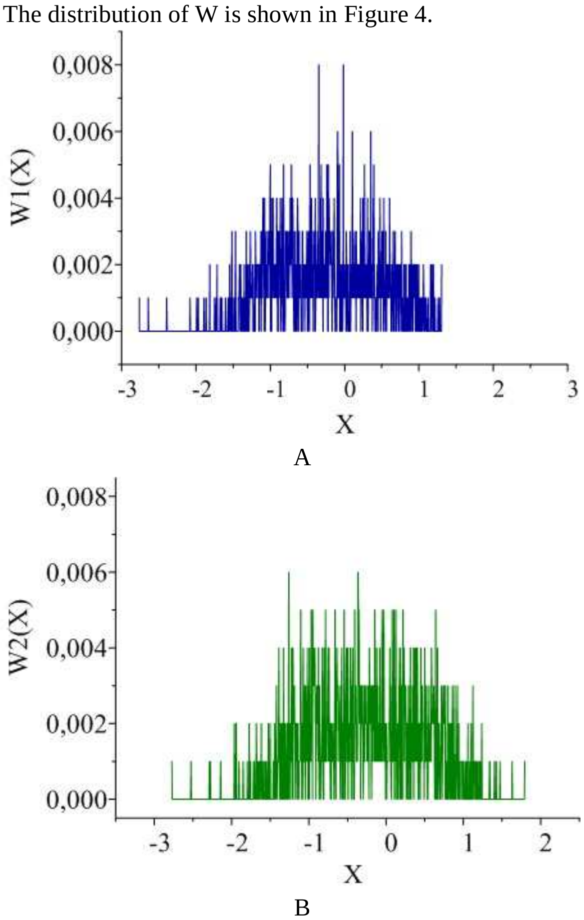

Fig. 4 W-distribution: A($\eta=15,\theta=3$), B($\eta=6,\theta=1$)

After comparing Figure 4A and Figure 4B, we can conclude that the W-distribution is independent of the DEM parameters

(η,θ). The percentiles of W-distribution are obtained from the simulated data.

The V-distribution is shown in Figure 5.

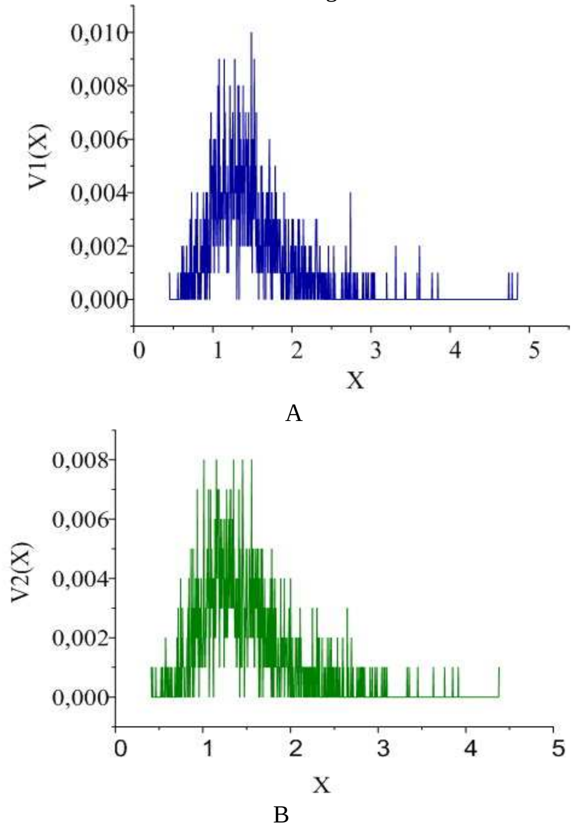

Fig. 5 V-distribution: A($\eta=15,\theta=3$), B($\eta=6,\theta=1$)

It is apparent from Figure 5 that the V-distribution doe snot depend on DEM parameters (η,θ), either.

B. Confidence Intervals for DEM Distribution

Confidence intervalconstruction is an important part of the statistical inference thatrefers to obtaining statements such as $P(a(X_1...X_n)\le\theta\le b(X_1...X_n))=\gamma$, where γ is typically chosen to be 0.9, 0.95, or 0.99. In other words, we use the CI for the estimation of an unknown general parameter using only a sample of the data with the confidence probability γ.The principle of the CI is shown in Figure 6.

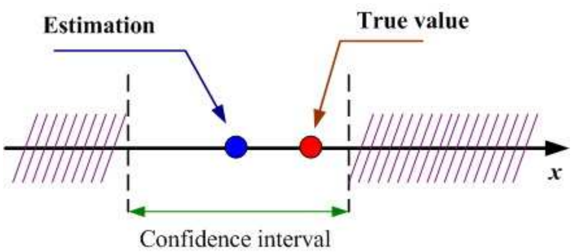

Fig. 6 Confidence interval and true value of an unknown parameter

One of the most useful methods for constructing CI is the method of pivotal quantities. A pivot is a function of $(\delta, X_1, X_2..X_n)$, whose distribution does not depend on the parameter δ. In the case of DEM distribution, the parameters are η and θ. As long as the distributions of W and V do not depend on corresponding parameters, W-distribution and V-distribution should be used as pivots for estimation of the population mean and variance.Unlike DEM distribution, Normal distribution (N) is well studied and the CI algorithm construction for the latter distribution is simple. The pivots for N are the t-distribution for the mean and χ^2-distribution for the variance, which are a consequence of the properties of N [5].

The CI for the population mean is defined as:

$$P\left(\bar{x}-t_\gamma\frac{s}{\sqrt{n}}<\eta<\bar{x}+t_\gamma\frac{s}{\sqrt{n}}\right)=\gamma, \qquad (7)$$

where $\bar{x}$ is sample mean, S is sample variance, and $t_\gamma(n-1,\gamma)$ is the value from the t-distribution table corresponding to the confidence probability (γ) and number of degrees of freedom (n).

The CI for the population variance is defined as:

$$P\left(\frac{(n-1)s^2}{{\chi_2}^2}<\theta<\frac{(n-1)s^2}{{\chi_1}^2}\right)=\gamma, \qquad (8)$$

where χ_1^2,χ_2^2 are the values from the χ^2-distribution table, which correspond to the number of degrees of freedom (n) and probabilities, $\frac{1+\gamma}{2},\frac{1-\gamma}{2}$ [5].

To construct the CI for DEM distribution, the pivotal distributions W and V, which are obtained from the given data sample using equations (5), should be used. The CI for the population mean is defined as:

$$P\left(a\le\frac{\tilde{X}-\eta}{\frac{1}{n}\sum\left|X_i-\tilde{X}\right|}\le b\right)=\gamma \qquad (9)$$

In order to construct CI with the corresponding percentiles of pivot distribution W, the following equation should be used:

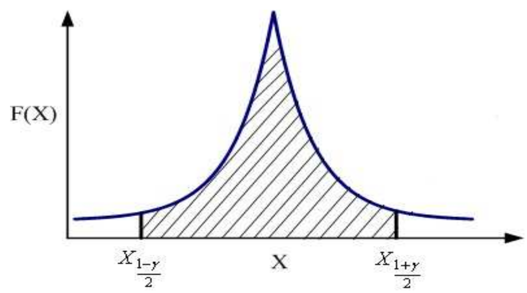

Fig. 7 Percentiles for the CI corresponding to γ probability

Finally, we obtain the equation for η estimation:

$$\left(\tilde{X} - X_{\frac{1+\gamma}{2}} \frac{\sum_{i=1}^{n} \left| X_i - \tilde{X} \right|}{n}, \tilde{X} - X_{\frac{1-\gamma}{2}} \frac{\sum_{i=1}^{n} \left| X_i - \tilde{X} \right|}{n} \right) \quad (10)$$

Similar to the aforementioned described procedure, we can construct the CI for the population variance using the pivotal distribution V:

$$\left(\frac{\sum_{i=1}^{n} \left| X_i - \tilde{X} \right|}{n X_{\frac{1+\gamma}{2}}}, \frac{\sum_{i=1}^{n} \left| X_i - \tilde{X} \right|}{n X_{\frac{1-\gamma}{2}}} \right) \quad (11)$$

C. Asymptotic Confidence Intervals for DEM Distribution Based on χ^2-Distribution

In this section, we will construct the asymptotic CI for the DEM distributed data. Consider the following equations:

$$T_i = \frac{2|X_i - \eta|}{\theta} \quad (12)$$

and

$$X_i = \pm \frac{\theta}{2} T_i + \eta \quad (13)$$

The absolute value of Jacobian of transformation is:

$$\left| \frac{d}{dt_i} \left(\pm \frac{\theta}{2} T_i + \eta \right) \right| = \frac{\theta}{2} \quad (14)$$

Using the existing distribution of independent value X with PDF $f(X)$ and the relationship with dependent variable $Y = g(X)$, we can obtain the equation for the PDF of dependent variable Y [5]:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \quad (15)$$

For our case, g is defined in (13). Thus, we obtain:

$$f(t_i) = \frac{1}{2\theta} \exp\left\{ \frac{-\left| \frac{\theta}{2} t_i + \eta - \eta \right|}{\theta} \right\} \frac{\theta}{2} +$$

$$+ \frac{1}{2\theta} \exp\left\{ \frac{-\left| \frac{\theta}{2} t_i + \eta - \eta \right|}{\theta} \right\} \frac{\theta}{2} == \frac{1}{2} \exp(-\frac{t_i}{2}) \quad (16)$$

This equation is a density function of χ^2-distribution with 2 degrees of freedom. We can use the property of gamma distribution [5] regarding the sum of independent gamma-variables. Considering that the χ^2-distribution is a partial case of gamma distribution, the same property is accurate for this distribution.

Therefore, $\sum_i \frac{2|X_i - \eta|}{\theta}$ follows the χ^2-distribution, and this distribution can be used as a pivot for the construction of the asymptotical CI.

V. Simulation and Discussion

A. Examining the pivotal quantities

In order to prove the reliability of the described techniques for the construction of CI, it is important to consider an example with a simulated data sample size of 10x1000. The obtained experimental distribution from our data sample is shown in Figure 3. The 2.5 and 97.5 percentiles of obtained pivotal equations (5) for simulated data with $\eta = 15, \theta = 3$ are:

$$W_{2.5\%} = -1.521,\ W_{97.5\%} = 0.729$$

The corresponding percentiles of the t-distribution with 9 degrees of freedom are:

$$t_{2.5\%} = -2.2622, t_{97.5\%} = 2.2622$$

From these data, we can see that the 2.5 percentile of W is greater than the corresponding percentile of t-distribution and the 97.5 percentile of W is smaller than the 97.5 percentile of t-distribution. The considered percentiles for V-distribution are:

$$V_{2.5\%} = 1.04, V_{97.5\%} = 2.782$$

Now, we can construct the CI for the population mean and variance based on W, V, and also on the t and χ^2-distributions.

Consider the following small sample of simulation data:

10.5889, 11.7205, 4.7902, 11.8991, 10.4191, 15.2304, 10.1076, 16.5358, 21.9029, 11.1247

The median, mean, and standard deviation of the sample are 12.669, 12.4319, and 4.3315, respectively. The 95% CI for the population mean and variance based on both obtained pivots and also on standard t and χ^2-squared, which are used for normal distribution, can be evaluated. For the simulated data, we obtain the following data using equations (7-11):

$$10.2126 \le \eta \le 17.794,\ 1.2111 \le \theta \le 3.2399$$

based on W and V. The CI, if we treat the data as normally-distributed, are:

$$9.3292 \le \eta \le 15.5347,\ 2.1095 \le \theta \le 5.5989$$

based on t and χ^2-squared distributions. In last equation we have used the fact that $\theta = \sqrt{\frac{\sigma^2}{2}}$ after applying (8).

We can see that, in our case, the intervals for η are almost equal, but for θ we obtain narrower intervals for W and V distributions. Unlike t and χ^2-squared distributions, W and V distributions are more precise for the DEM data because they are constructed directly from the simulated sample.

B. Construction of Confidence Intervals for the Population Variance from the Simulated Data

To demonstratethat the use of pivotal equations for normal distribution is incorrect for DEM cases, one shouldconstruct 1000 CI for each 10 values ofsimulated data. In this study, we use the equations (7) and (8) to construct the CI. After applying these estimators, we obtained 843 points, which captures the actualmean value of $\eta = 15$, thus, in our case, the true coverage level is 84.3%. For the variance we obtained 261 points, which captures the true value of variance $\sigma^2 = 2\theta^2$ that corresponds to 26.1% of the true confidence level.

Now, let's examine the result obtained in 5.3. We can calculate the CI for the population variance based on χ^2-squared distribution with 20 degrees of freedom for 10 points from our simulation sample. Using the pivot $2\sum_i \frac{|X_i - \hat{\eta}|}{\theta}$ we have:

$$P\left(\chi^2_{2n,0.025} < \frac{2\sum_i |X_i - \hat{\eta}|}{\theta} < \chi^2_{2n,0.975} \right) = 0.95$$

Hence, the asymptotic confidence interval for θ is:

$$\frac{2\sum_i |X_i - \hat{\eta}|}{\chi^2_{2n,0.975}} \le \theta \le \frac{2\sum_i |X_i - \hat{\eta}|}{\chi^2_{2n,0.25}} =$$

$$= \left(\frac{67.3899}{\chi^2_{2n,0.975}}, \frac{67.3899}{\chi^2_{2n,0.025}} \right) = (1.9722, 7.0265)$$

In this example, we usethe percentiles of the χ^2 distribution with 20 degrees of freedom. It is therefore apparent that both asymptotic confidence interval and the one based on bootstrappingprimarily agree and capture the true value of the population variance.

VI. Conclusion

In this paper, we have investigated the properties of the DEM. The pivot equations for this distribution were obtained based on the simulated data. We have analyzed the techniques for the construction of the CIfor both DEM and normal distributions. Our study reveals that the trivial usage of the pivots of normal distribution is incorrect if applied to DEM distributed data. Also, we have demonstrated that for population variance estimation,the χ^2-distribution with 20 degrees of freedom, which is a very important result for practical applications, can be used.These approachesmay be used as a theoretical basis for the analysis of experimental data.

References

[1] S. Kotz, T. Kozubowski, K. Podgorski, The Laplace Distribution and Generalizations, Birkhauser Boston, 2001, Ch. 2-4.
[2] N. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, Vol. 2, Wiley-Interscience, 1995, Ch. 24 .
[3] Y. R. Gel, "Test of fit for a Laplace distribution against heavier tailed alternatives", Computational Statistics and Data Analysis, Vol. 54, no. 4, pp. 958-965, 2010.
[4] N. Ekstrand, B. Smeets, "Weighting of double exponential distributed data in lossless image compression", Lund University. Data compression conference, March 30-April 01, 1998.
[5] J. Rice, Mathematical statistics and data analysis, 3rd edition, Thomson Brooks/Cole, 2007, p. 275-280.

M. Alrasheedi is PhD in Operations Research,Dept. of Quantitative Methods, School of Business, King Faisal University, Saudi Arabia. Research Interests: Applications of Quantitative Methods in Management and Finance, Applied Operations Research.