International Journal of Sciences & Applied Research

International Journal of Sciences & Applied Research

www.ijsar.in

Parameter estimation in exponentiated exponential distribution based on Progressively Censored Survival Time Data

Kamil ALAKUŞ[1], Necati Alp ERİLLİ[2]*

[1]Department of Statistics, University of OndokuzMayis, Samsun, Turkey.
[2]Department of Econometrics, University of Cumhuriyet, Sivas, Turkey.

Correspondence Address: *Necati Alp ERİLLİ, Department of Econometrics, University of Cumhuriyet, Sivas, Turkey.

Abstract

Introduction: Survival time data measure the time to a certain event, such as failure, death etc. These times are subject to be random variables. These data have the property for non-negativeless and must have skewed distribution. The exponentiated exponential distribution can define as a generalization of the standard exponential distribution. It is obtained with Fisher information matrix under Type-II censoring for this distribution.

Case Report: The aim of this study is to investigate estimating of parameters in exponentiated exponential distribution based on progressively censored survival time data. Moreover likelihood based confidence intervals are also investigated in the same data frame. It is also illustrate the results using bile duct cancer data set.

Discussion: In this study we consider parameter(s) estimation in the EE distribution based on progressively censored samples. We also consider likelihood based confidence interval of the parameters in that distribution. We present a real data example where it is observed that the EE has a better fit compare to standard exponential, gamma, log-logistic, log-normal and Weibull.

Keywords: Exponentiated exponential distribution, hazard function, likelihood based confidence interval, maximum likelihood estimator, progressively censoring

Introduction

A new skewed failure time distribution named exponentiated exponential distribution (or generalized exponential distribution) has been introduced and studied by Gupta and Kundu (1999, 2001, 2003, 2004, 2007). It can be defined as a generalization of the standard exponential distribution. In particular, the EE distribution is defined by the cumulative distribution function (cdf) as given in Equation.1:

$$F(t;\alpha,\lambda)=(1-exp(-\lambda t))^{\alpha}, \quad t>0,\ \alpha \text{ and } \lambda>0 \qquad (1)$$

This is simply the αth power of the cdf standard exponential distribution. The mathematical properties of EE distribution have been studied in detail by Gupta and Kundu (2001), and Nadarajah and Kotz (2003). The aim of this study is to investigate estimators of the parameters in the EE distribution based on censored (progressively) survival time data using the method of maximum likelihood. The corresponding probability distribution function (pdf) as given below:

$$f(t;\alpha,\lambda)=\alpha\lambda e^{-\lambda t}\left(1-e^{-\lambda t}\right)^{\alpha-1},$$

$t>0$, α and $\lambda>0$ (2)

Here α is the shape parameter and λ is scale parameter. The standard exponential distribution is the particular case of (2) for $\alpha=1$.

The EE distribution shares an attractive physical interpretation. Suppose that the survival times of n-observations in a medical study are independently and identically distributed according to Equation.1 or 2. Then it follows the survival time of the study also has the EE distribution.

Since than, Exponentiated Exponential Distribution is studied for different types of data or distributions. Gauss et all. (2013) is studied The Exponentiated Generalized Class of Distributions. Nadarajahand Gupta (2007) is studied for exponentiated gamma distribution. Gauss and Cordeiro (2013) is studied for the class of Generalized Exponentiated Distributions. Nadarajah and Kotz (2006) is also give more about Exponentiated Type Distributions.

The Exponentiated Exponential (EE) distribution

α-powered standard exponential distribution function can be given in Equation.3:

$$f(t;\alpha,\lambda)=\alpha\lambda e^{-\lambda t}\left(1-e^{-\lambda t}\right)^{\alpha-1},$$

$t>0$, α and $\lambda>0$. (3)

The corresponding cdf and survival function of the EE distribution can given by Equation.4 and Equation.5:

$$F(t;\alpha,\lambda)=\left(1-exp(-\lambda t)\right)^{\alpha},$$

$t>0$, α and $\lambda>0$. (4)

$$S(t;\alpha,\lambda)=1-\left(1-exp(-\lambda t)\right)^{\alpha},$$

$t>0$, α and $\lambda>0$. (5)

Therefore hazard function defined as in Equation.6:

$$h(t;\alpha,\lambda)=\frac{f(t;\alpha,\lambda)}{S(t;\alpha,\lambda)}=$$

$$\alpha\lambda e^{-\lambda t}\left(1-e^{-\lambda t}\right)^{\alpha-1}\left(1-\left(1-e^{-\lambda t}\right)\right)^{\alpha-1}, t>0 \quad (6)$$

The hazard function of the EE is constant at λ when $\alpha=1$, increasing when $\alpha>1$, and decreasing when $\alpha<1$. Therefore the EE distribution has a nice physical interpretation.

Parameters estimation

One of the features of the survival data which renders standard methods inappropriate is that survival times are generally censored. This may be because the data from a study are to be analysed at a point in the when some observations are still alive. Alternatively, the survival time status of an observation at the time of analysis might not be known because that observation has been lost to follow-up. If the observation was last known to be alive or lost to follow-up at time c, the time c is called a censored survival time. This censoring occurs after the observation has been entered into a study that is, to the right of the last known survival time, and is therefore known as right censoring. This type of right-censoring is called progressively censoring. Consequently, right-censored survival time is then less than the actual, but unknown, survival time.

In the following sub-section we are going to give point estimators of the parameters based on censored survival time data using maximum likelihood method. Moreover we are going to give likelihood based confidence interval estimators of the parameters in the same data frame.

Point estimation of parameters

In this sub-section we consider estimation by the method of maximum likelihood for estimation of parameters. Now an uncensored random sample is $t_1,\cdots,t_d$, and censored sample be $c_1,\cdots,c_{n-d}$. An observation observed to survive at t contributes a term $f(t;\alpha,\lambda)$ to the likelihood, the density of survive at t. The contribution from an observation whose survival time is censored at c is $S(c;\alpha,\lambda)$, the probability of survival beyond c. The full likelihood from n independent observation, indexed by i, is then

$$L=\prod_{i=1}^{n}\left(f(t_i;\alpha,\lambda)\right)^{w_i}\prod_{i=1}^{n}\left(S(t_i;\alpha,\lambda)\right)^{(1-w_i)} \quad (7)$$

where the w_i is a dummy variable that is defined as follows

$$w_i=\begin{cases}1, & \text{if survival time uncensored}\\ 0, & \text{if survival time censored}\end{cases} \quad (8)$$

The log-likelihood is

$$\log L(t;\alpha,\lambda)=\sum_{i=1}^{n}w_i\log f(t_i;\alpha,\lambda) + \sum_{i=1}^{n}(1-w_i)\log S(t_i;\alpha,\lambda) \quad (9)$$

In terms of the observed survival or censoring time $x_i=\min(t_i,c_i)$, this becomes

$$\log L(x_i;\alpha,\lambda)=d\log(\alpha)+d\log(\lambda)-\lambda\sum_{i=1}^{n}w_ix_i+(\alpha-1)\sum_{i=1}^{n}w_i\log(1-e^{-\lambda x_i}) +\sum_{i=1}^{n}(1-w_i)\log\left(1-\left(1-e^{-\lambda x_i}\right)^{\alpha}\right) \quad (10)$$

The first order derivatives of (10) with respect to the two parameters are:

$$\frac{\partial\log L(.)}{\partial\alpha}=\frac{d}{\alpha}+\sum_{i=1}^{n}w_i\log\left(1-e^{-\lambda x_i}\right)-\sum_{i=1}^{n}(1-w_i)\left(1-e^{-\lambda x_i}\right)^{\alpha}\log\left(1-e^{-\lambda x_i}\right)\times\left(1-\left(1-e^{-\lambda x_i}\right)^{\alpha}\right)^{-1}$$

and

$$\frac{\partial\log L(.)}{\partial\lambda}=\frac{d}{\lambda}-\sum_{i=1}^{n}w_ix_i+(\alpha-1)\sum_{i=1}^{n}w_i\left(x_ie^{-\lambda x_i}\left(1-e^{-\lambda x_i}\right)^{-1}\right)-\alpha\sum_{i=1}^{n}(1-w_i)x_ie^{-\lambda x_i}\times\left(1-e^{-\lambda x_i}\right)^{\alpha-1}\left(1-\left(1-e^{-\lambda x_i}\right)^{\alpha}\right)^{-1}$$

Replacing to $e^{-\lambda x_i}$, and $(1-e^{-\lambda x_i})$ with l_i and f_i, respectively, then the last two equations can be written as follows:

$$\frac{\partial\log L(.)}{\partial\alpha}=\frac{d}{\alpha}+\sum_{i=1}^{n}w_i\log f_i-\sum_{i=1}^{n}(1-w_i)F_i\log f_iS_i^{-1} \quad (11)$$

$$\frac{\partial\log L(.)}{\partial\lambda}=\frac{d}{\lambda}-\sum_{i=1}^{n}w_ix_i+(\alpha-1)\sum_{i=1}^{n}w_ix_il_if_i^{-1}-\alpha\sum_{i=1}^{n}(1-w_i)x_il_if_i^{\alpha-1}S_i^{-1} \quad (12)$$

In equations (11), and (12) F_i and S_i represent cdf and survival function of *i*-th observation. Setting the last two expressions to zero and solving them simultaneously yields the maximum likelihood estimates of the two parameters.

Confidence interval estimation of parameters

In this sub-section we are going to give likelihood based confidence interval estimators of the parameters in the censored survival time data. Let ℓ_0 be log-likelihood function value according to maximum likelihood estimates from the EE distribution. Then for λ parameter $100(1-\alpha)\%$ confidence intervals estimates are given by $3.841=2(\ell_0-\ell_{\lambda L})$ and $3.841=2(\ell_0-\ell_{\lambda U})$, where $\ell_{\lambda L}$ and $\ell_{\lambda U}$ denotes lover and upper values of log-likelihood function when α value fixed. In similar way, for the parameter α, $100(1-\alpha)\%$ confidence intervals estimates are given by $3.841=2(\ell_0-\ell_{\alpha L})$ and $3.841=2(\ell_0-\ell_{\alpha U})$, where $\ell_{\alpha L}$ and $\ell_{\alpha U}$ denotes lover and upper values of log-likelihood function when λ value fixed. The value of 3.841 is %95 table values of the chi-square distribution when degrees of freedom 1.

Results

In this study we used Bile Duct Cancer data to illustrate the proposed method. Data are taken Fleming et al. (1980). They are use the data for comparison two survival distributions using modified Kolmogorov-Smirnov test statistic. Cox and Oakes (1984) was used the data for illustrating expectation and maximisation (EM) algorithm from gamma distribution with the group of radiation-drug therapy patients. We use the data for illustrating the EE distribution and comparing the other five distributions. For the group of radiation-drug therapy patients survival times are given in Table 1.

For comparing the EE distribution results with the other five distributions we use the following survivor functions: For standard exponential distribution, $S(t;\beta)=\exp(-t/\beta)$, for gamma distribution, $S(t;\alpha,\beta)=\exp(-t/\beta)\left(\sum_{k=0}^{\alpha-1}(t/\beta)^k! \quad \times k^{-1}\right)$, for log-logistic distribution, $S(t;\alpha,\beta)=\left(1+(t/\beta)^{\alpha}\right)^{-1}$, for log-normal distribution, $S(t;\mu,\sigma)=1-\Phi((\log t-\mu)/\sigma)$, and For Weibull distribution, $S(t;\alpha,\beta)=\exp\left(-(t/\beta)^{\alpha}\right)$.

We give all six distribution survival function in Fig. 1 with Kaplan-Meier estimate as a goodness-of-fit test. The figure shows that except standard exponential distribution the other five distributions fit these data well.

We have fitted standard exponential, the EE, gamma, log-logistic, log-normal and Weibull to this data set. We present the estimates of parameter(s) and median estimate of these six distributions in Table 2. In addition we also present the log-likelihood $\ell L(.)$ in the same table.

Comparing these six distributions we conclude that the best distribution of this data set is the EE distribution. Because its log-likelihood function has biggest value. We also test to the hypothesis $H_0:\alpha=1$ against $H_1:\alpha\neq 1$ the EE alternatives with likelihood ratio statistic. Then we have $\chi_H=2(127.0593-123.5766)=6.9654$, which asymptotically follows the chi-square distribution with one degrees of freedom under the null hypothesis. Thus we reject the null hypothesis at the level of 0.0083.

Table 1: Survival Times for 22 Patients. Survival times (in days) of 22 patients with bile duct cancer treated with combined drug and Radiation therapy (data from Fleming et al.).

Uncensored Survival Times	***30, 67, 95, 148, 170, 171, 176, 193, 200, 221, 243, 261, 262, 263, 399, 414, 446,464, 777***
Censored Survival Times	79, 82, 446

Table 2: Log-likelihood values from bile duct cancer data.

Distribution Name	**Estimated Scale Parameter**	**Estimated Shape Parameter**	**Log-Likelihood Value $\ell L(.)$**	**Estimated Median Value**
Standard Exponential	295.1052	None	-127.0593	204.5513
The EE	165.9948	2.5816	-123.5766	240.0550
Gamma	119.2764	2.3648	-123.6630	243.4571
Log-Logistic	234.9320	2.4862	-123.9324	234.9320
Log-Normal	5.421122	0.7432	-124.4434	226.1327
Weibull	315.7340	1.6401	-123.8524	252.5059

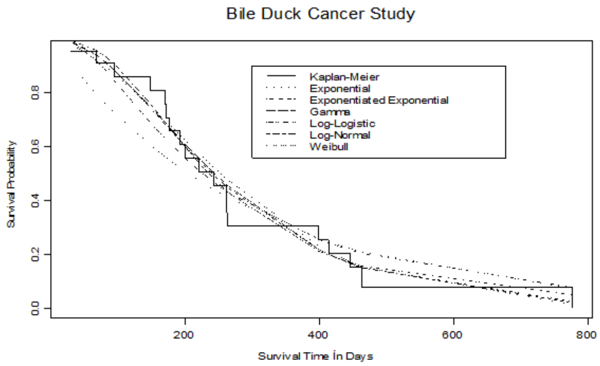

Fig. 1: Comparing survival probabilities with bile duck cancer.

Firstly, we estimate the two parameters using S-Plus written program and the corresponding results are $\hat{\alpha} = 2.581575$ and $\hat{\beta} = \hat{\lambda}^{-1} = 165.9673$. Secondly, we give %95 likelihood based confidence interval estimates as follows. The result are: For shape parameter α, $(1.6367, 3.8466)$, and for scale parameter λ $(4.4236e-3, \quad 8.0130e-3)$. The shape parameter is not including 1 then the same conclusion may be done as likelihood test statistic.

Discussion

One of the features of the survival data which renders standard methods inappropriate is that survival times are generally censored. This may be because the data from a study are to be analysed at a point in the when some observations are still alive. Alternatively, the survival time status of an observation at the time of analysis might not be known because that observation has been lost to follow-up. If the observation was last known to be alive or lost to follow-up at time c, the time c is called a censored survival time. This censoring occurs after the observation has been entered into a study that is, to the right of the last known survival time, and is therefore known as right censoring. The right-censored survival time is then less than the actual, but unknown, survival time.

In this study we consider parameter(s) estimation in the EE distribution based on progressively censored samples. We also consider likelihood based confidence interval of the parameters in that distribution. We present a real data example where it is observed that the EE has a better fit compare to standard exponential, gamma, log-logistic, log-normal and Weibull. Extensive studies are required to get other estimations. For example, confidence interval of parameters with other approaches (symmetrical), and survival probabilities.

References

Cox D. R., Oakes D., 1984. Analysis of survival data. Chapman & hall, London, UK.

Fleming T. R., O'Fallon J. R., O'Brien P. C., Harrington D. P., 1980. Modified Kolmogorv-Smirnov test procedures with application to arbitrarily right censored data. Biometrics 36, 607-26.

Gupta, R. D. and Kundu, D., 1999. Generalized Exponential Distributions. Australian and New Zealand Journal of Statistics, Vol. 41, No. 2, pp. 173-188.

Gupta, R. D. and Kundu, D., 2003. Discriminating Between the Weibull and the GE Distributions. Computational Statistics and Data Analysis, vol. 43, pp.179-196.

Gupta, R. D. and Kundu, D., 2004. Discriminating Between the Gamma and Generalized Exponential Distributions. Journal of Statistical Computation and Simulation, vol. 74, no. 2, pp. 107-121.

Gupta, R. D. and Kundu, D., 2007. Generalized exponential distribution: existing methods and recent developments. Journal of the Statistical Planning and Inference, vol. 137, no. 11, pp.3537 – 3547.

Gupta, R. D. and Kundu,D., 2001. Exponentiated Exponential Family; An Alternative to Gamma and Weibull. Biometrical Journal,vol. 33, no. 1, pp.117-130.

Nadarajah S., Kotz, S., 2003. On the exponentiated exponential distribution. (to appear in statistica).

Zheng G., 2002. On the fisher information matrix in type ii censored data form exponentiated exponential family. Biometrical journal 44, 353-57.

Nadarajah S., Kotz, S. The Exponentiated Type Distributions. Acta Applicandae Mathematica 92(2),97-111, 2006.

Gauss M. & Cordeiro M (2013). “The Exponentiated Generalized Class of Distributions”. Journal of Data Science. 11. 1-12.

Nadarajah, S. and Gupta, A. K. (2007). The exponentiated gamma distribution with application to drought data.Calcutta Statistical Association Bulletin 59, 29-54.

Gauss M. Cordeiro, G. M., Ortega, E.M.M, and da Cunha, D.C.C. The Exponentiated Generalized Class of Distributions. Journal of Data Science 11,(2013), 1-27.