**Modelling Earthquake data using some lifetime distributions**

***Caner TANIŞ1,\*[C:\Users\Abdullah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ORCID-iD_icon-16x16.gif](https://orcid.org/0000-0003-0090-1661)***

*1 Science Faculty, Department of Statistics, Çankırı Karatekin University, Çankırı, Turkey*

|  |
| --- |
| **Abstract**  In this study, we provide statistical inferences about earthquake data by modelling some lifetime distributions. We consider five lifetime distributions (Weibull, exponentiated Exponential [1], exponentiated Weibull [2], generalized Lindley [3], and Power Lindley [4]) to model earthquake data. We consider two earthquake data sets in this paper. The first data consists of 20 observations denoting the magnitudes of earthquakes in the Kuşadası bay on 23 November 2020 while the second data set includes the magnitudes of earthquakes in the Kuşadası bay on 24 November 2020. The maximum likelihood method is used to estimate the unknown parameters of these distributions. We estimate the average magnitude of earthquakes via the maximum likelihood estimates of the parameters of five lifetime distributions.    Keywords: Weibull distribution, Exponentiated Exponential distribution, Maximum likelihood estimation, Real data analysis. |
|  |

1. **Introduction**

On October 30, 2020, an earthquake occurred in Kuşadası Bay with a magnitude calculated as 6.9 by the Kandilli Observatory and Earthquake Research Institute. Hundreds of aftershocks occurred after the earthquake [5,6]. As a result of this earthquake, a total of 117 people lost their lives and 1,034 people were injured [7]. Many buildings were destroyed in Bayraklı and Bornova districts of İzmir [8].

In the literature, many statistical distributions are used to model real-life data in many fields such as biology, chemistry, engineering, and medical sciences, etc. Some of the commonly used lifetime distributions are Weibull, Lindley, and various modified versions of these distributions. In this study, unlike other studies in the literature, estimates of the mean magnitude of earthquakes will be provided using some known lifetime distributions such as Weibull, exponential Exponential [1], exponentiated Weibull [2], generalized Lindley [3] and Power Lindley distribution [4]. The main aim of this paper is to model earthquake data via these lifetime distributions and estimate the average magnitude of the earthquakes.

The rest of this study is organized as follows: In the second section, we present the mentioned lifetime distributions. In Section 3, the maximum likelihood estimators (MLEs) of the parameters of the examined distributions are obtained. In Section 4, we present two data applications to determine the optimal model for each dataset and compute the estimated average magnitude of earthquakes. Also, the selection criteria to compare the fits of the models to data sets are given in same section. The results are given in Section 5.

1. **Modelling Methodology**
   1. **Weibull distribution**

The Weibull distribution is one of the popular lifetime distributions. The Weibull is very useful in modeling lifetime data obtained in various fields. The cumulative distribution function (CDF) and probability density function (PDF) of the Weibull distribution are given by

 (1)

 (2)

respectively, where,  is the shape parameter,  is the scale parameter and .

* 1. **Exponentiated Exponential distribution**

Exponentiated Exponential (EE) distribution was introduced by Gupta and Kundu [1]. The CDF and PDF of the EE distribution are given as follows:

 (3)

 (4)

respectively, where  and .

* 1. **Exponentiated Weibull distribution**

Exponentiated Weibull (EW) distribution was proposed by Pal et al. [2]. The EW distribution is a generalization of the Weibull distribution. The CDF and PDF of the EW distribution are given by

 (5)

 (6)

respectively, where and  are shape parameters,  is scale parameter and . The EW distribution is reduced the Weibull distribution for in (5).

* 1. **Generalized Lindley distribution**

Lindley distribution was suggested by Lindley [9]. The CDF and PDF of Lindley distribution are

 (7)

 (8)

respectively, where  and .

Generalized Lindley (GL) distribution is introduced by Nadarajah et al. [3]. The CDF and PDF of the GL distribution are given by

 (9)

 (10)

respectively, where,  ve .

* 1. **Power Lindley distribution**

Power Lindley distribution was proposed by Ghitany et al. [4]. The Power Lindley was obtained by considering a transformation  in (7)-(8). Thus, the CDF and PDF of Power Lindley distribution are given by

 (11)

 (12)

respectively, where  and [4]. Ghitany et al. [4] emphasized that Power Lindley distribution (PL) is a mixture of Weibull  and generalized gamma  distribution with mixing proportion.

1. **Point Estimation**

In this section, we present the maximum likelihood estimators (MLEs) of the parameters of the examined lifetime distributions in Section 2.

Let  be a random sample from the Weibull distribution.The log-likelihood function is given by

 (13)

The MLEs of the  and  parameters are the values that maximize the  function in (13). The MLEs of the  and  parameters can be obtained by simultaneous solution of the nonlinear equations created by taking the derivatives of the  function according to the  and  parameters and equating them to zero.

, be a random sample from the EE distribution. The log-likelihood function is given by

 (14)

[1].

The MLEs of the  and  parameters are the values that maximize the  function in (14). The MLEs of the  and  parameters can be obtained by simultaneous solution of the nonlinear equations created by taking the derivatives of the  function according to the  and  parameters and equating them to zero.

, be a random sample from the EW distribution. The log-likelihood function is given by

 (15)

[2].

The MLEs of the  and parameters are the values that maximize the  function in (15). The MLEs of the  and  parameters can be obtained by simultaneous solution of the nonlinear equations created by taking the derivatives of the  function according to the  and  parameters and equating them to zero.

, be a random sample from the GL distribution. The log-likelihood function is given by

 (16)

[3].

The MLEs of the  and  parameters are the values that maximize the  function in (16). The MLEs of the  and  parameters can be obtained by simultaneous solution of the nonlinear equations created by taking the derivatives of the  function according to the  and  parameters and equating them to zero.

, be a random sample from the PLdistribution. The log-likelihood function is given by

 (17)

[4].

The MLEs of the  and  parameters are the values that maximize the  function in (17). The MLEs of the  and  parameters can be obtained by simultaneous solution of the nonlinear equations created by taking the derivatives of the  function according to the  and  parameters and equating them to zero.

In this study, the **optim** function in the **R** program and the BFGS algorithm, which was first studied by Fletcher [10] were used to solve the related likelihood equations.

1. **Model Evaluation**

In this section, we present two earthquake data sets and some selection criteria to compare the fits of models to data sets. We consider some selection criteria such as Akaike information criterion (AIC), Bayesian information criterion (BIC), Anderson-Darling statistics (A\*), Cramér-von Mises statistics (W\*), Kolmogorov-Smirnov statistics (K-S), and p-values (A\*, W\*, KS) for earthquake data analysis. These measures are given by

, (18)

, (19)

, (20)

, (21)

, (22)

respectively, where  denotes  order statistics,  denotes to the number of parameters,  is the sample size,  is the value of the log-likelihood function,  denotes empirical distribution function, and  refers to the CDF of the examined model.

* 1. **Data Description**

In this subsection, we present two real data sets including the magnitudes of the earthquakes in Kuşadası bay. The first data set refers to the magnitudes of the earthquakes on November 23, 2020 while the second data set includes the magnitudes of the earthquakes on November 24, 2020. Some descriptive statistics of the data sets are given in Table 1.

**Table 1.** Some descriptive statistics of the earthquake data sets

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Data Set*** | ***n*** | ***Min.*** | ***Max.*** | ***Mean*** | ***Var.*** | ***CS*** | ***CK*** |
| ***1*** | 20 | 1.2 | 2.5 | 1.780 | 0.17 | 0.308 | -1.013 |
| ***2*** | 28 | 1.3 | 2.5 | 1.785 | 0.077 | 0.519 | 0.402 |

Min.: Minimum, Max.: Maximum, Var.: Variance,

CS: Coefficient of Skewness, CK: Coefficient of Kurtosis

1. **Results**

In this section, the results of the earthquake data analysis are presented. Table 2 contains the maximum likelihood estimates and their standard errors (SEs) of the parameters of the distributions fitted to the data given in Section 2. Table 3 shows the comparison statistics used to compare the models fitted to the data sets.

**Table 2.** The MLEs and SEs of the parameters of the fitted models for the earthquake datasets

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Data set | Model |  |  |  |  |  |  |
| 1 | Weibull | 4.8565 | 1.9424 | - | 0.8392 | 0.0947 | - |
| EE | 2.9180 | 101.5109 | - | 0.5271 | 80.9166 | - |
| EW | 1.2362 | 1.6353 | 14.0620 | 2.7924 | 1.8372 | 52.4211 |
| GL | 77.8533 | 3.2482 | - | 61.7207 | 0.5364 | - |
| PL | 3.5973 | 0.1856 | - | 0.5254 | 0.0715 | - |
| 2 | Weibull | 6.6413 | 1.9049 | - | 0.9009 | 0.0575 | - |
| EE | 1100.3067 | 4.2306 | - | 1077.9688 | 0.6193 | - |
| EW | 1.0281 | 2.1623 | 22.3760 | 1.6700 | 1.5381 | 58.1391 |
| GL | 4.5730 | 825.2685 | - | 0.6171 | 794.5009 | - |
| PL | 4.9163 | 0.0889 | - | 0.5911 | 0.0364 | - |

**Table 3.** The selection criteria for the earthquake datasets

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data set | Model | AIC | BIC | KS | A\* | W\* | p-value (KS) | p-value (A\*) | p-value (W\*) |
| 1 | Weibull | 24.7687 | 26.7601 | 0.1271 | 0.4265 | 0.0614 | 0.9031 | 0.8203 | 0.8111 |
| EE | 23.7226 | 25.7140 | **0.1012** | 0.3139 | 0.0383 | **0.9866** | 0.9265 | 0.9459 |
| EW | 25.5679 | 28.5551 | 0.1014 | 0.3129 | 0.0384 | 0.9863 | 0.9274 | 0.9453 |
| GL | **23.6953** | **25.6867** | 0.1013 | **0.3120** | **0.0380** | 0.9864 | **0.9281** | **0.9472** |
| PL | 24.9289 | 26.9203 | 0.1258 | 0.4152 | 0.0581 | 0.9095 | 0.8318 | 0.8316 |
| 2 | Weibull | 14.0751 | 16.7395 | 0.1241 | 0.5293 | 0.0727 | 0.7813 | 0.7156 | 0.7388 |
| EE | 10.0468 | 12.7112 | 0.1170 | 0.3096 | 0.0541 | 0.8383 | 0.9302 | 0.8550 |
| EW | 11.3317 | 15.3283 | **0.0952** | **0.2270** | **0.0380** | **0.9613** | **0.9812** | **0.9460** |
| GL | **9.9860** | **12.6504** | 0.1160 | 0.3034 | 0.0530 | 0.8457 | 0.9351 | 0.8618 |
| PL | 12.7474 | 15.4118 | 0.1121 | 0.4093 | 0.0566 | 0.8730 | 0.8383 | 0.8397 |

Bold text indicates the best model

From the Table 3, it can be concluded that except for the KS test statistic and the its p value, the GL distribution is the best fitted model to first data set according to other selection criteria while the EE distribution has more fit than other models acoording to KS statistics and its p-value for the first data set. On the other hand, we observe that except for the AIC and BIC, the EW distribution is the best fitted model to second data set according to other selection criteria while the GL distribution has more fit than other models according to AIC and BIC for the second data set.

|  |  |
| --- | --- |
|  |  |

**Figure 1.** The fitted CDFs (on left) and PDFs (on right) for data set 1

|  |  |
| --- | --- |
|  |  |

**Figure 2.** The fitted CDFs (on left) and PDFs (on right) for data set 2

**Table 4.** The estimated average magnitude of the earthquakes for datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Data Set*** | ***Weibull*** | ***EE*** | ***EW*** | ***GL*** | ***PL*** |
| ***1*** | 1.7804 | 1.7828 | 1.7784 | 1.7822 | 1.7765 |
| ***2*** | 1.7769 | 1.7919 | 1.7855 | 1.7916 | 1.7810 |

Figures 1-2 illustrate the fitted CDFs and PDFs for two datasets. Table 4 provides the estimates of the average magnitude of the earthquakes for data sets. These estimates are computed by using the expected values of the distributions under MLEs in Table 2. From Table 4, it can be concluded that the estimates are very close the true mean of the samples in Table 1.

**References**

1. Gupta, R. D., Kundu, D., (2001), Generalized exponential distribution: different method of estimations, Journal of Statistical Computation and Simulation, 69 (4), 315-337.
2. Pal, M., Masoom M.A., Jungsoo W. (2006). Exponentiated weibull distribution. *Statistica* 66 (2) 139-147.
3. Nadarajah, S., Bakouch, H. S., Tahmasbi, R. (2011). A generalized Lindley distribution. Sankhya B, 73(2), 331-359.
4. Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N., Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference. Computational Statistics & Data Analysis, 64, 20-33.
5. <https://jfm.sakarya.edu.tr/sites/jfm.sakarya.edu.tr/file/30_Ekim_2020_Kusadasi_Korfezi_depremi_bilgi_notu.pdf>,Utkucu, M., Budakoğlu, E., Doğan, E. 30 EKİM 2020 KUŞADASI KÖRFEZİ DEPREMİ BİLGİ NOTU, Access date: August 22, 2022.
6. <http://www.koeri.boun.edu.tr/sismo/2/wp-content/uploads/2020/10/20201030_izmir_V1.pdf>, Access date: August 22, 2022.
7. <https://www.hurriyet.com.tr/gundem/izmirdeki-depremde-enkazdan-cikartilan-mahir-tahirler-hayatini-kaybetti-41672849>, Access date: August 22, 2022.
8. <https://www.hurriyet.com.tr/galeri-izmir-depremi-son-dakika-iste-busenin-enkaz-altindan-kurtarildigi-anlar-canli-yayinda-iletisime-gecilmisti-41650162>, Access date: August 22, 2022.
9. Lindley, D.V., 1958. Fiducial distributions and Bayes’ theorem. Journal of the Royal Statistical Society, Series A 20, 102–107.
10. Fletcher R. (1987). Practical methods of optimization. John and Sons, Chichester.