International Journal of Advanced Statistics and Probability, 5 (2) (2017) 87-90

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
doi: 10.14419/ijasp.v5i2.7834
Research paper

The exact extreme value distribution – applied study

Aisha Fayomi *, Neamat Qutb, Ohoud Al-Beladi

Department Of Statistics, Faculty Of Science / King Abdul-Aziz University, Saudi Arabia
**Corresponding author E-mail: afayomi@kau.edu.sa*

Abstract

Extreme value theory is used to develop models for describing the distribution of extreme events. Exact extreme value or compound distribution which is based on the theory of the maximum of random variables of random numbers is one of the most important models that are applicable in various situations, for instance of interest, it uses partial duration series (PDF) data to analyze extreme hydrological. As part of our earlier study, the parameters of this model were estimated by two methods, maximum likelihood (ML) and Bayesian-based on non-informative and informative priors. Moreover, a comparative study using simulated data showed that the Bayesian based on informative prior is the best estimation method. In this paper, a real data set taken from records of the largest daily rainfall data of Jeddah city in Saudi Arabia is used to fit the model when the parameters are estimated by Bayesian method. A comparative applied study indicates that the exact extreme value model under Bayesian estimates (BE) of its parameters provides appropriate fit for this data set and it is more applicable than the same model when the parameters are estimated by ML method and other three classical extreme value models.

Keywords: *Bayesian Estimation; Compound Distribution; Exact Extreme Value Distribution; Extreme Rainfall Events; Maximum Likelihood Estimation.*

1. Introduction

Extreme rainfall events pose a constant threat to society. Large scale flooding, after extended periods of heavy precipitation, cause disruption and loss of property and can lead to loss of lives. Therefore, the study of extreme rainfall is of major interest worldwide as it has important implications for life insurance, civil protection, local and regional planning and civil infrastructure design.

The two approaches widely used in data selection for extreme rainfall are annual maximum series (AMS) approach and peak over thresholds series (POT) approach, which is also known as partial duration series (PDS). The theoretical basis for development of the POT method was set by work of Todorovic [1]. He has formulated a stochastic model of floods in which the time of flood peak occurrence (and consequently, the number of peaks in a time interval), and the peak magnitudes are both random variables. Todorovic and Zelenhasic [2] have chosen Poisson process as the process of peak occurrence and exponential distribution as the distribution of peak magnitude.

In a study of Afif [3] the three classical extreme value distributions (Gumbel, Frechet and Weibull) are applied to rain on data from Jeddah city in Saudi Arabia during 31 years. A model for the form of an exponential distribution for the magnitude of rain above a pre-determined base level, compounded with a Poisson model for the occurrence of the exceedances is applied in her study to the PDS during 31 years (daily discharges over the high base level). The results of goodness of fit tests when the parameters are estimated by ML method of the four models to the observed data compared between the three classical extreme value distributions and compound distribution. The results showed that the compound distribution is more applicable than the three classical extreme value distributions for the rain data analysis. Qutb et al [4] discussed properties of the compound exponential-Poisson model, and they used the moments, maximum likelihood, and Bayesian methods to estimate the distribution parameters, when simulating and compared them, the results showed that Bayesian method - based on informative prior distributions - is the best estimation method.

The aim of this study is to evaluate the performances of Bayesian and ML methods in estimating the parameters of the exact extreme value model by using a real data set that appeared in Afif [3].

This paper is organized as follows. Section 2 introduces the compound exponential-Poisson model. Section 3 is dedicated to the description of the study area. In Section 4 application study and test of the model assumptions are investigated. In section 5, the parameters' estimation method will be explained. The validation of the exact extreme value distribution will be tested in Section 6. Finally, concluding remarks are given in Section 7.

2. The compound exponential-Poisson model

A random variable X is defined to have a compound exponential-Poisson distribution if its probability density function (pdf) is given by.

$$f(x)=\frac{\alpha}{\beta}\exp\left(-\frac{x}{\beta}\right)\exp\left[-\alpha\exp\left(-\frac{x}{\beta}\right)\right], \quad (1)$$

$-\infty < x < \infty$ where $\alpha,\beta > 0$.

Its cumulative distribution function (CDF) is

$$F(x)=\exp\left[-\alpha\exp\left(-\frac{x}{\beta}\right)\right] \quad (2)$$

3. Study area

Jeddah is a Saudi-Arabian city located on the cost of the Red Sea and is the major urban center of western Saudi Arabia. It is the largest city in Makkah Province, the largest sea port on the Red Sea, and the second-largest city in Saudi Arabia after the capital city, Riyadh. The population of the city currently stands at over 3.4 million. It is considered the commercial capital of Saudi Arabia.

Jeddah features an arid climate, unlike other Saudi Arabian cities; Jeddah retains its warm temperature in winter, while summer temperatures are very hot, often breaking the 40° C mark in the afternoon and dropping to 30° C in the evening.

The rainfall in Jeddah is few and rare, and usually occurs of small amounts in December, the rain reaching around 3 inches (7.6 mm). On 25 November 2009 (Wednesday evening), severe flooding affected the city and other areas of the Makkah area. Civil defense officials described it as the worst in 27 years. As of 29 November 2009, around 116 people Killed and more than 350 are missing. Some roads have been filled with less than a meter of water, and many victims are believed to have sunk into their cars. More than 3,000 cars were swept away or damaged. Rainfall continued on Thursday (November 26th) for four hours, with rainfall averaging about 90 millimeters, twice the year-long rainfall and the heaviest in Saudi Arabia in a decade. The timing of the flooding came two days before Eid al-Adha and during the Hajj in neighboring Makkah. On 26 January 2011, severe flooding also affected the city and other areas of Makkah. The high level of water floods hit twice the rate recorded in Jeddah 2009. Meanwhile, local witnesses said that the east of Jeddah has sunk, and the flood waters rush westward towards the Red Sea, and turn the streets to rivers again.

In this study, we used the real data given in Afif [3], which is a rain data have been collected from city of Jeddah in Saudi Arabia and covers the period of 1978 to 2008 (i.e. 31 years). This data obtained from Presidency of Meteorology and Environment (PME) of Saudi Arabia.

4. Application study and test of the model assumptions

To demonstrate the applicability of the exact extreme value (compound) distribution, Afif [3] applied the model to rain on data obtained from PME of Saudi Arabia. The rain data, in the form of PDS (daily discharges over a high base level), cover the period of 1978 to 2008 (i.e. 31 years) have been collected from city of Jeddah in Saudi Arabia.

4.1. The selection of the base level

It should be noted that the selection of the base level appears somewhat arbitrary. If a high base level is used, then the amount of data extracted will be few. On the other hand, a low base level will render less valid for the assumption behind the Poisson model for the occurrence of the exceedance. Afif [3] applied three base levels (40, 30 and 21 mm). The level of 30 mm - which was given by expert from PME - considered to be hazardous level and may be justifiable in terms of significant heavy rain. Other two levels (40 and 21 mm) have been chosen to study the effect of changing the base level on the goodness of fit of the three classical extreme value distributions and the exact extreme value distribution and its assumptions.

4.2. Test of the assumptions

The model has three assumptions: 1) The number of exceedances has Poisson distribution. 2) The exceedances have exponential distribution. 3) The exceedances are independent. These assumptions were tested using the rain data, and the results of these tests are discussed in the following subsections.

4.2.1. Goodness of fit test to Poisson distribution

In deriving the exact extreme value model, the Poisson distribution has been used to describe the process of occurrence of rain exceedances. In order to achieve this assumption, the number of exceedances for each year over the 31-year period is extracted from which a Poisson distribution model is assumed for the occurrences of these exceedances. An analysis, including the K-S goodness of fit test assuming Poisson distribution was performed. It should be noted that the p-values of these tests are 1, 0.950 and 0.891 respectively for the three base levels, which reinforce the validity of Poisson's assumption for the occurrences of the exceedances above the three selected base levels.

4.2.2. Goodness of fit test to exponential distribution

In deriving the exact extreme value model, the exponential distribution has been used to describe the magnitude of rain above a pre-determined base level. The exceedances above three base levels 40, 30 and 21 mm respectively, are obtained and fitted with an exponential distribution function. It should be noted that the goodness of fit tests reinforced the validity of exponential assumption for the exceedances above the three selected base levels.

4.2.3. Test of the independence

According to the exact extreme value distribution, the exceedances are assumed to be independent. To satisfy the third assumption, the water resources council guidelines require that flood peaks be separated by more than five days. This method may drop important data, especially for the small sample size (short period of collected data). The simplest method for detecting a lack of independence among observations within a sample is computing the serial correlation and testing to decide whether it is significantly different from zero. Serial correlation is defined as the simple correlation between adjacent observations ξ_t and ξ_{t-1} for t=1, 2… n-1. This correlation is called serial correlation with lag one because each observation is correlated with the next observation, see [5].

Hypothesis tests with significance levels 5% and 1% indicated that the exceedances are uncorrelated; the non-significant correlation coefficients for the three base levels are 0.258, 0.319 and 0.355 respectively. Hence we don't need to separate between the data points and lose very important information.

5. Bayesian estimation of the parameters

In this paper, the discussion will be focused more on the Bayesian method instead of ML method, which used by Afif [3].Bayesian Bayesian inference is an alternative to the classical statistical inference. With Bayesian approach, the parameters could themselves be random variables have probability density functions called prior distributions. Bayesian analysis combines the information at the data represented by the entire likelihood function with prior knowledge about the parameters, which may come from other data sets or a modeler's experience and physical intuition. Parameter estimation is made through the posterior distribution which is computed using Bayes' theorem.

$$\pi(\theta|\underline{x}) = \frac{l(\underline{x}|\theta)\pi(\theta)}{\int_{\theta} l(\underline{x}|\theta)\pi(\theta)\,d\theta} \tag{3}$$

Where $\pi(\theta|\underline{x})$ is the posterior distribution of the parameters θ, $l(\underline{x}|\theta)$ is , is the likelihood function, and $\pi(\theta)$ is the prior distribution of θ.

A Monte Carlo method is an algorithm that relies on repeated pseudo-random sampling for computation, and is therefore, stochastic (in contrast to opposed to deterministic). Monte Carlo methods are often used for simulation. The union of Markov

chains and Monte Carlo methods is called MCMC [6]. Metro-Hasting (MH) algorithm (Metropolis et al. [7], Hastings [8]) and Gibbs sampling (Geman and Geman [9]) are very famous and the most practical MCMC for simulation studies. MH used to obtain a sequence of random samples from a probability distribution for which direct sampling is difcult. Chib and Greenberg [10] and gamer man and Lopes [11] provide comprehensive preliminary details as well as an intensive development and applications of MCMC specifically on MH techniques. MH is a fundamental algorithm for many Markov chain simulation approaches while the Gibbs sampling is a good alternative if the full conditional distributions for each parameter are known. In Qutbet al. [4] we estimated the exact extreme value distribution parameters using ML and Bayesian-under informative prior distribution (inverted gamma distribution with one parameter)-methods. The estimates simulated by the Metro-Hasting function in the (MH adaptive) package. In this study, the BE is computed and displayed in Table 1 with the corresponding MLE of the parameters for the three base levels 40, 30 and 21 mm respectively. These estimates are used in the goodness of fit tests.

Table 1: Estimates of the Compound Exponential-Poisson Distribution Parameters

Method of estimation	Base level	Parameter	
		Shape, α	Scale, β
ML	40 mm	2.75	10.82
	30 mm	3.76	11.12
	21 mm	2.35	13.15
Bayesian	40 mm	2.99	8.06
	30 mm	4.80	8.98
	21 mm	2.65	11.17

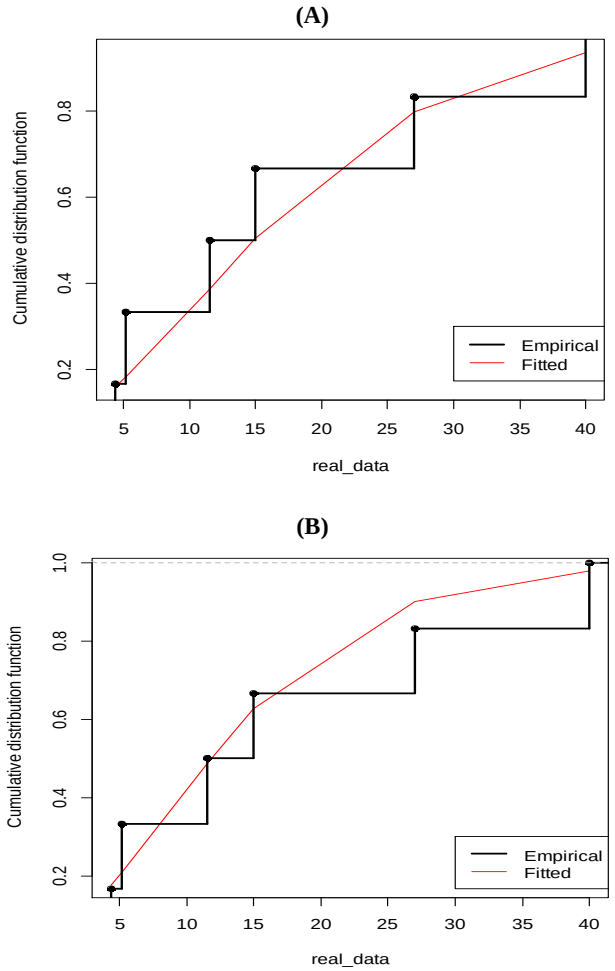

Fig. 1: The Plot for the Empirical and Fitted CDF of the Compound Distribution Using ML and Bayesian Estimates when Base Level=40 Mm. (A) Using MLE, (B) Using BE.

6. Validation of exact extreme value model

The exact extreme value (compound) model is used to fit largest exceedances, from the data of rain through 31 years period. The validity of the fitted model is checked by applying the Kolmogorov-Smirnov (K-S) test statistics on the fitted CDF and the empirical CDF. The fitted CDF is conducted by replacing the parameters with their MLE and BE as seen in Fig. 1, Fig. 2 and Fig. 3. The test statistic value of the K-S test using MLE and BE are recorded in Table 2.

The K-S test values indicate that the exact extreme value model under BE of its parameters provides appropriate fit for this data set and it is more applicable for the three base levels than the same model when the parameters are estimated by ML method.

Table 2: Goodness-of-Fit Test Statistic Value

Base level	Method of estimation	
	ML	Bayesian
40 mm	0.164	0.125
30 mm	0.172	0.143
21 mm	0.181	0.161

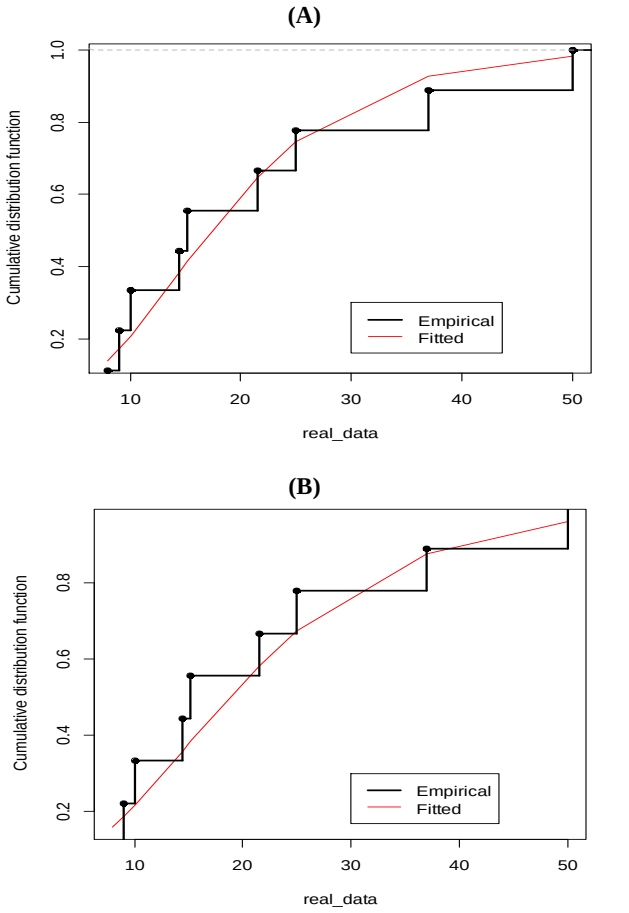

Fig. 2: The Plot for the Empirical and Fitted CDF of the Compound Distribution Using ML and Bayesian Estimates when Base Level=30 Mm. (A) Using MLE, (B) Using BE.

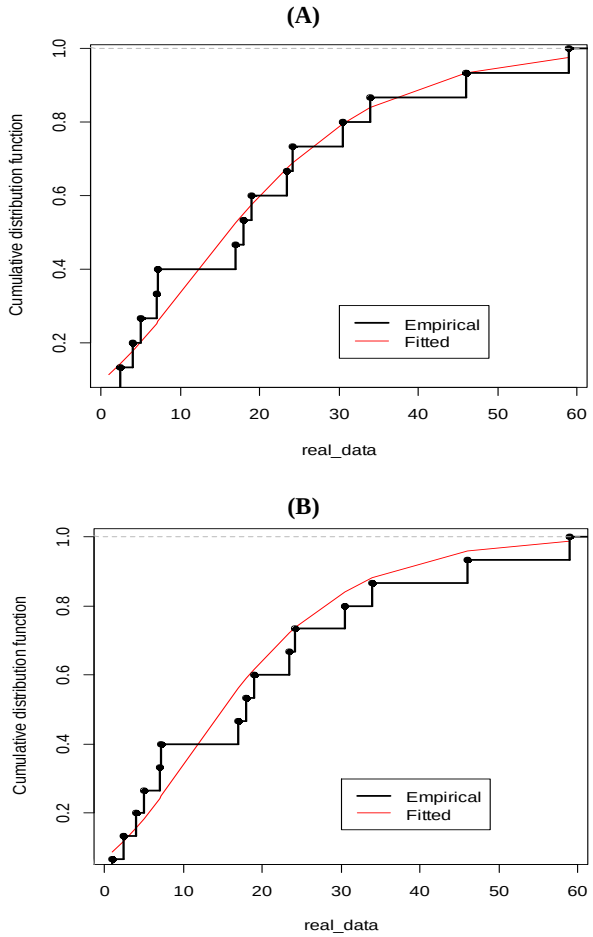

Fig. 3: The Plot for the Empirical and Fitted CDF of the Compound Distribution Using ML and Bayesian Estimates when Base Level=21 Mm. (A) Using MLE, (B) Using BE.

7. Conclusion

A comparative analysis of test statistic values of K-S test shows that, the exact extreme value model with BE of its parameters is more applicable to the rain data for three base levels than the model with MLE of its parameters.

From Afif [3] and our applied study, we can conclude that the exact extreme value model under BE of its parameters is the best fit to the rain data comparing with the same model when theparameters are estimated by ML method and the classical extreme value models (Gumbel, Frechet and Weibull).

References

[1] Todorovic, P.” Onsome problems involving random number of random variables”, *The Annals of Mathematical Statistics*, Vo.41, No.3, (1970), pp: 1059-1063. https://doi.org/10.1214/aoms/1177696981.

[2] Todorovic, P.and Zelenhasic,E. ”A stochastic model for flood analysis”,*Water Resources Research*, Vo .6, No.6, (1970), pp:1641-1648. https://doi.org/10.1029/WR006i006p01641.

[3] Afif, W. M. M. The Compound and Extreme Value Distributions. M.Sc. Thesis, King Abdul-Aziz University,Jeddah,(2011).

[4] Qutb, N., Fayomi, A., and Al-Beladi, O.”Estimation of the parameters of the exact extreme value distribution”,*International Organization of Scientific Research- Journal of Mathematics*, Vo .13, No.2, (2017), pp: 1-9.

[5] Kotb, N. S. A. Estimation of the Parameters of Compound Distribution. Ph. D. Thesis, Department of Statistics, Faculty of Commerce, Al-Azhar University, (2002).

[6] Robert, C. P. and Casella, G., *Monte Carlo Statistical Methods*, Springer- Verlag, (2004). https://doi.org/10.1007/978-1-4757-4145-2.

[7] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.”Equation of state calculations by fast computing machines”,*The Journal of Chemical Physics*, Vo.21, No.6, (1953), pp: 1087–1092. https://doi.org/10.1063/1.1699114.

[8] Hastings, W. K.”Monte Carlo sampling methods using Markov chains and their applications”,*Biometrika*, Vo.57, No.1, (1970), pp:97–109. https://doi.org/10.1093/biomet/57.1.97.

[9] Geman, S. and Geman, D”Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”,*IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vo.6,No.6, (1984),pp: 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

[10] Chib, S. and Greenberg, E.”Understanding the metropolis-hasting's algorithm”,*The American Statistician* , Vo.49, No.4, (1995), pp:327–335.

[11] Gamerman, D. and Lopes, H. F.,*Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference*,Chapman and Hall,(2006).