ISSN: 0067-2904

Modeling of medium-Sized magnitudes of earthquakes

Saleemah H. Jasaim AlSaady
Department of Astronomy and Space, College of Science, University of Baghdad , Baghdad, Iraq

Abstract

In this work, it have earthquakes data been analytized as a result of ground water injection attendant to dig wells and oil, which affects the pressure pores in the rock and is according to cause a medium-sized earthquake that have already occurred more than 20 times. according to the publication of the Department of Attenba, the dangers of earthquakes and movement flooring in Guy-Greenbrie-United area of one of United States of America cities, where these data represent a monitor magnitudes of earthquakes from September 2008 the September 2011, we conducted statistical analysis and create a model for earthquakes flooring that was chosen for four distributions and results i the optima reveal the best distribution is the Burr X11 then General Extreme value, and lognormal distribution, and normal distribution. We applied the maximum likelihood method for estimating the parameters of these distribution.

Keywords: Burr Distribution , general Extreme value Distributions, log normal Distribution, normal Distribution, QQ plots

نمذجة الزلازل الأرضية متوسطة الشدة

سليمه حمادي جاسم الساعدي
قسم علوم الفضاء والفلك، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

في هذا العمل تم تحليل بيانات الزلازل الأرضية متوسطة الشدة نتيجة حقن الارض بالمياه المصاحبة لحفر الابار النفطية والتي يؤثر على الضغط المسام في الصخور ويحتمل ان تتسبب في حدوث الزلازل متوسطة الحجم التي حدثت بالفعل لا كثر من 20مره حسب ما نشر من دائرة التنبؤ واخطار بالزلازل و الحركة الأرضية في منطقةGuy–Greenbrier–United لا حدى المدن الولايات المتحدة الأمريكية حيث تمثل هذه البيانات رصد مقادير الزلازل في ايلول 2008الى ايلول 2011 ، اجرينا تحليلا احصائيا ولإيجاد نموذج للزلازل الأرضية تم اختيار اربعة توزيعات هو توزيع بيير والتوزيع العام للقيم المتطرفة وتوزيع اللوغاريتمي الطبيعي التوزيع الطبيعي وظهرت النتائج على التوالي ان توزيع بيير افضل نموذج ثم توزيع العام للقيم المتطرفة ثم توزيع اللوغاريتمي الطبيعي وبعده التوزيع الطبيعي . ولقد استخدمنا طريقة الامكان الاعظم في تقدير معلمات تلك التوزيعات .

Introduction

Earthquakes are twitching and move and ripple violently to the earth's surface followed by liberation of energy from the earth's crust, and this energy is generated as a result of occur displacement between crustal rocks through block and that occurs as a result of continuous exposure to cramps and large pressure. The world Author Paul Denton first proposed between the size of the earthquake and

Email: salemahjasm@yahoo.com

the frequency of occurrence in a showing paper published that showed the size of the event and the frequency of occurrence is common significantly [1], despite the fact that the values of a and b may differ significantly from region to region or more. Through the Poisson distribution of earthquakes of moderate, which is larger than the M7 to measure tremors and medium events have had time Tkar-madh 20 days while the possibility of obtaining 50%, according to the meteorological community in their areas and in the last time also repeated event for 14 days and was tolerable 90% [2]. Andrzej Kijko gave the Japanese the world research shows the possibility of estimating Poisson distribution parameters (. B and λ) data at risk of earthquakes and discovered distributed the first researcher able to determine the size of earthquakes in Japan [2]. Andre Kijko pre. M.A.Sellove, sented models for the occurrence and amounts -zlszl floor. Is compared to models of the data the Japanese set for years 1885-1980 using likelihood methods to get the best model, through changing the time scale to investigate the deviation in the data model. these were tactics graphic traditional associated with fixed Poisson can be used with a time scale converted into elementary operations, effective use of such an analysis makes possible to find a set of data that has not been obtained in the form of features. Based on these analyzes, it is being investigated the usefulness of this occurrence of earthquakes to predict a major earthquake [3], where these data represent a monitor magnitudes of earthquakes from September 2008 the September 2011, we conducted statistical analysis and to create a model for earthquakes flooring that was chosen for four distributions and results i the optima reveal the best distribution is the Burr X11 then General Extreme value, and lognormal distribution, and normal distribution. We using the maximum likelihood method.

2. Parametric modeling of earthquakes magnitudes

four parametric distributions are considered to represent the. The selected distributions are chosesuch that its limiting distribution is unbounded; that is, their upper endpoint tends to infinity.

2.1 Burr (Type XII) distribution

The Burr distribution is an important model in many fields such as business, engineering, quality control and other fields. Further, the Burr distribution having logistic and Weibull as special sub-models, is a very popular distribution for modeling lifetime data and for modeling phenomenon with monotone failure rates. Moreover, the mixtures of distributions are many interesting in various scientific fields such as physics, biology and medicine among others Burr system of distributions was constructed in 1942 by Irving W. Burr. is a very popular distribution for modeling function of the BXII can be written in closed form, probability density function [4]:

$$p(x)= 1-\left(1+\frac{x}{\alpha})^{c}\right)\wedge -k \qquad x>0,\ c>0,\ k>0, \alpha>0 \quad \ldots\ldots(1)$$

and cumulative distribution function:

$$F(x;c,k) =\frac{ck}{\alpha}\left[\frac{x}{\alpha}\right]^{c-1}\left(1+(\frac{x}{\alpha})^{c})\right)\wedge^{k-1} \quad x>0,\ c>0,\ k>0, \alpha>0 \quad \ldots\ldots (2)$$

Note when *c*=1, the Burr distribution becomes the, often referred to as the

The Burr Type XII distribution is a member of a system of continuous distributions introduced by which comprises 12 distributions.

2.2 Generalized Extreme Value (GEV) distribution

Let Mn = max (Y_1, . . . , Yn) where Yi are i.i.d. from a continuous cumulative distribution function F. Suppose we can find normalizing constants $a_n > 0$ and b_n such that

$$P\left(\frac{M_n-b_n}{a_n}\leq y\right)\to G(y) \quad \ldots\ldots(3)$$

as $n\to\infty$ • , where G is some proper distribution function. Then G is necessarily one of three possible types of limiting distribution functions which have been called the Gamble type, Fr´echet type and Waybill type. Nowadays, it is realized that it is more convenient to consider the generalized extreme value (GEV) distribution, which holds the three types as special cases. It has a cumulative distribution function of the form

$$G(y) = \exp\{-\left[1+\frac{\xi(y-\mu)}{\sigma}\right]_{+}^{-\frac{1}{\xi}}\} \quad \ldots\ldots(4)$$

where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter, , and h_+ = max(h, 0).

Differentiation of (7) gives the probability density function of the GEV distribution

$$g(y) = \frac{1}{\sigma}[1 + \xi\left(\frac{y-\mu}{\sigma}\right)]^{-(\frac{1}{\xi}+1)} \exp\{-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\} \quad \text{.......................................(5)}$$

The (GEV) distribution is very flexible and is the distribution that is recommended for use in many studies concerning extreme events. [5, 6]

2.3 Log normal distribution

If X is distributed as a normal, then log(X) is distributed as a lognormal,
The lognormal distribution is very important for neuroscience because the result of multifactorial, multiplicative processes gives a variable with a lognormal distribution. That is, when there eared a large number of causes, each roughly equal in effect, then the product of those causes results in a lognormal variable There is no need to develop mathematically the properties of this distribution .Rather, the importance for the neuron scientists is to recognize when variables are distributed as a lognormal. log-normal distribution. A random variable which is log-normally distributed takes only positive real values. The distribution is occasionally referred to as the Galton distribution or Galton's distribution, after Francis Galton. The log-normal distribution also has been associated with other names, such as McAlister, Gibrat and Cobb–Douglas.A log-normal process is the statistical realization of the multiplicative product of many independent random variables, each of which is positive. This is justified by considering the central limit theorem in the log domain [7].
The probability density function is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x}\exp\left[-\frac{1}{2}\left(\frac{\ln x - m}{\sigma}\right)^2\right] \quad \text{..(6)}$$

$-\infty < x > \infty, -\infty < M > \infty, \infty \quad > \partial >$
and cumulative distribution function

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_0^x \frac{1}{t}\exp\left[-\frac{1}{2}\left(\frac{\ln t - m}{\sigma}\right)^2\right]\mathrm{d}x = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln x - m}{\sigma\sqrt{2}}\right)\right] \quad \text{...............(7).}$$

2.4 Normal distribution

The normal, a continuous distribution, is the most important of all the distributions.
It is widely used and even more widely abused. Its graph is bell-shaped. You see the bell curve in almost all disciplines. Some of these include psychology, business, economics, the sciences, nursing, and, of course, mathematics. The most useful continuous distribution, and one that occurs frequently, is the normal distribution. The probability density functions of normal random variables are symmetric single peaked, bell-shaped density curves. Data sets occurring in nature will often have such a bell-shaped distribution, as measurements on many random variables are closely approximated by a normal probability distribution .The normal distribution is useful because of the central limit theorem. In its most general form, under some conditions (which include finite variance), it states that averages of random variables independently drawn from independent distributions converge in distribution to the normal, that is, become normally distributed when the number of random variables is sufficiently large Moreover, many results and methods (such as propagation of uncertainty and least squares parameter fitting) can be derived analytically in explicit form when the relevant variables are normally distributed. The normal distribution is sometimes informally called the bell curve. However, many other distributions are bell-shaped (such as the Cauchy, Student's t, and logistic distributions). The terms Gaussian function and Gaussian bell curve are also ambiguous because they sometimes refer to multiples of the normal distribution that cannot be directly interpreted in terms of probabilities. [7]
The probability density of the normal distribution is:

$$\text{p(x)} = \left(\frac{1}{\sqrt{2\sigma^2\pi}}\right)e^{-\frac{(X-\mu)^\wedge 2}{2\sigma^2}} \text{...(8)}$$

and cumulative distribution function

$$\text{F(x)} = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \text{...(9)}$$

3. Maximum Likelihood Estimation

In 1922 the greatest possible estimation method (Fisher) gave Fisher It was one of the most modalities, (Maximum likelihood function) (MLE).In the twentieth century the importance of the science of statistics because it gave the best estimate of the parameter of the And efficiency . We will use the maximum likelihood method in estimating the parameters of the selected models because it is one of the best methods with high efficiency. It is one of the most widely used methods of statistical estimation has become generally accepted as the most robust procedure. The most important step to study the MLE is to evaluate the sample joint distribution which are also called the likelihood function. In the case of identical and independent sample, the likelihood function is just the product of margin AL density of individual sample. Numerical maximum likelihood estimation requires maximization of the log–likelihood function or equivalently minimization of its negation. The most important step to study the MLE is to evaluate the sample joint distribution which are also called the likelihood function. In the case of identic calando independent sample, the likelihood function is just the product of marginal density of individual sample. However, in the study of time series analysis, the dependence structure of observation is species and it is not correct to use the product of marginal density to evaluate the likelihood function.1 To evaluate the sample likelihood, the use of conditional density is needed as seen in the following.

3.2 Burr Distributions MLE

The likelihood function of the burr model is given by

$$\log(\mathrm{x}) = \alpha^n + \beta^n \prod \prod_{i=1}^{n} x^{n-1} \prod_{i=1}^{n} (1 + x_{i=1)}^{\alpha}) - \beta - 1 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots . . (10)$$

And taken the natural logarithm of both sides the equation becomes as follows

$$\log(x)=\mathrm{n}\log(\alpha)+\mathrm{n}\log(\beta)(\alpha-1)\sum_{i=1}^{n\,l} xi + (-\beta-1)\sum_{i=1}^{n}(1+x_{i0)}^{\alpha} \dots\dots\dots\dots\dots\dots\dots(11)$$

Implicit differentiation sides as yours to make 0 have the following equation

$$\frac{n}{\beta^n} - \sum_{i=1}^{n} \log(1 + x_i^{\alpha}) = 0 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots . (12)$$

And solve these equations using numerical methods because they are non-linear and extract values of

3.1 GEV MLE distribution

The likelihood function of the GEV model is given by

$$L(\mu,\sigma,\xi) = (\frac{1}{\sigma})^n \prod_{i=1}^{n} [1 + \xi\left(\frac{y_i - \mu}{\sigma}\right)]^{-(\frac{1}{\xi}+1)} exp \sum_{i=1}^{n} \{-\left[1 + \xi\left(\frac{y_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\} \dots\dots\dots\dots\dots\dots . \dots . (13)$$

The negative log-likelihood function is

$$-\log(\iota) = n\ log\ \sigma + \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\log\left[1+\xi\left(\frac{y_i-\mu}{\sigma}\right)\right] + \sum_{i=1}^{n}[1+\xi\left(\frac{y_i-\mu}{\sigma}\right)]^{-1/\xi} \dots\dots\dots\dots\dots\dots\dots....(14)$$

$$\prod_{i=1}^{n} f(t1; a, \beta)$$

And solve these equations using numerical methods because they are non-linear and extract values of estimates

3.3 Log Normal MLE

The likelihood function of the log normal model is given by

$$L(x;\mu,\sigma) = \prod_{Xi}^{n}(\frac{1}{Xi})N(l\mathrm{n}\ \mathrm{x};\mu;б)\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\ \dots(15)$$

where by L we denote the probability density function of the log-normal distribution and by N that of the normal distribution. Therefore, using the same indices to denote distributions, we can write the log-likelihood function thus:

L (μ,б|x1;$x2\dots\dots xn)=-\sum_k lnXk + LN(\mu,б$|Ln x2,…….Ln nx)…………………………………..…(16)

ι=constant+$\sum_k lnXk + LN(\mu,б$|Ln x2,…….Ln nx)...(17)

3.4 normal distribution MLE

Normal random variables having mean and variance . The probability density function of a generic term of the sequence is

The mean and the variance are the two parameters that need to be estimated.

The likelihood function is The log-likelihood function is [8]

L (μ,$б^2$|x1;$x2\dots\dots xn)=-\frac{n}{2}Ln(2\piб^2)-\frac{1}{2б^2}\sum_{j=1}^{n}Xj-\mu$...(18)

$\iota(\mu,\sigma^2;X1\dots xn)=(2\pi\sigma^2)^{-2n}\exp\left[\frac{1}{2б^2}\right]\sum_{j=1}^{n}(Xj-\mu)^2\cdot\dots\dots$ (19)

4. Empirical Results and Analysis

4.1. Fitting earthquakes magnitudes Maxima

The data used for the analysis, which represents magnitudes of earthquakes as a result of water injection used in the drilling of wells in the city of raster Gre enbrie US cities in September 2008 for September 2011.data are found in [9].

Table 1- Fitting Results

#	Distribution	Parameters
2	Burr	k=1.1094 =5.3171 C =2.0104
3	Gen. Extreme Value	a=-0.01471 =0.54417 =1.7477
5	Lognormal	=0.32997 =0.66661
6	Normal	=0.70922 =2.054

In practical applications, an appropriate estimator may be selected according to goodness of fit tests rather than on theoretical considerations. Therefore, an important step is to test whether the resulting model fits the observations. Table- 2 lists the maximum likelihood estimators along with the goodness of fit statistics for each model. According to the goodness of fit statistics the four models seems to be accepted.

Table 2- Fitting Results and Goodness of Fit Statistics

Distribution	**Kolmogorov Smirnov**		**Anderson Darling**		**Chi-Squared**	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Burr x11	0.05245	1	2.3308	1	45.277	1
Gen. Extreme value	0.06296	2	2.451	2	55.831	2
Lognormal	0.06326	3	4.9294	3	55.842	3
Normal	0.10864	4	13.938	4	94.521	4

4.2. QQ plots

Given an ordered sample of independent observations $X_{(l)} \le X_{(2)} \le \cdots \le X_{(n)}$ from a population with estimated distribution function $\hat{F}$, a quantile plot consists of the points

$$\left\{\left(\hat{F}^{-1}\left(\frac{i}{n+1}\right),x_{(i)}\right):i=1,\dots,n\right\}$$

If $\hat{F}$ is *a reasonable estimate of F, then the quantile plot should also consist of* points close to the unit diagonal [10].

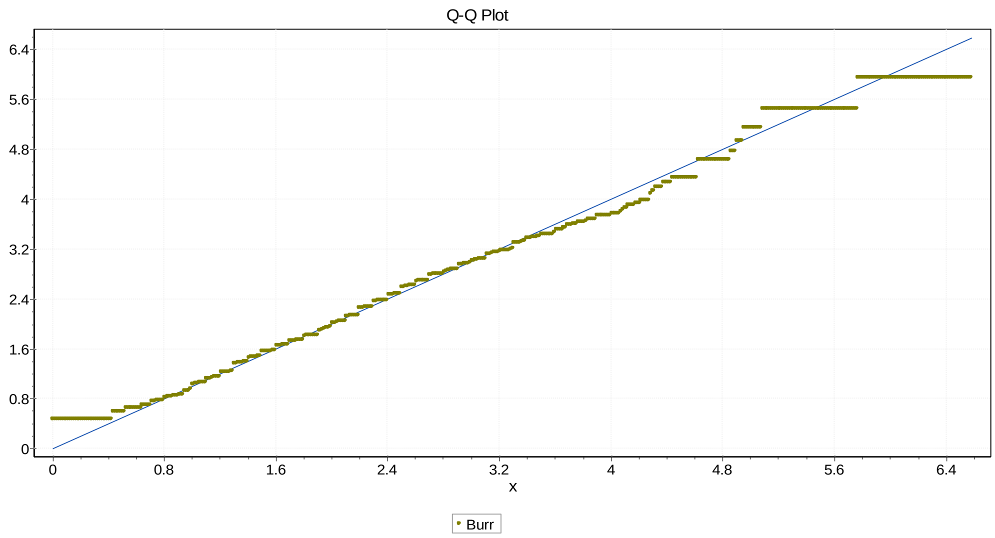

Figure 1- A quantile – quantile plot for the fitted Burr model to earthquake magnitude

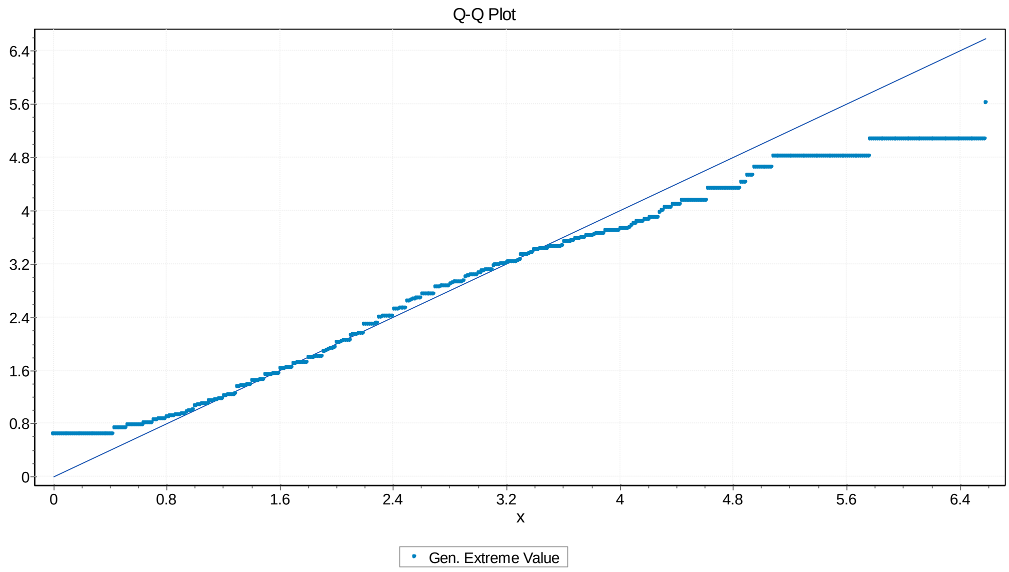

Figure 2- A quantile – quantile plot for the fitted model Gen Extreme Value to earthquake magnitude

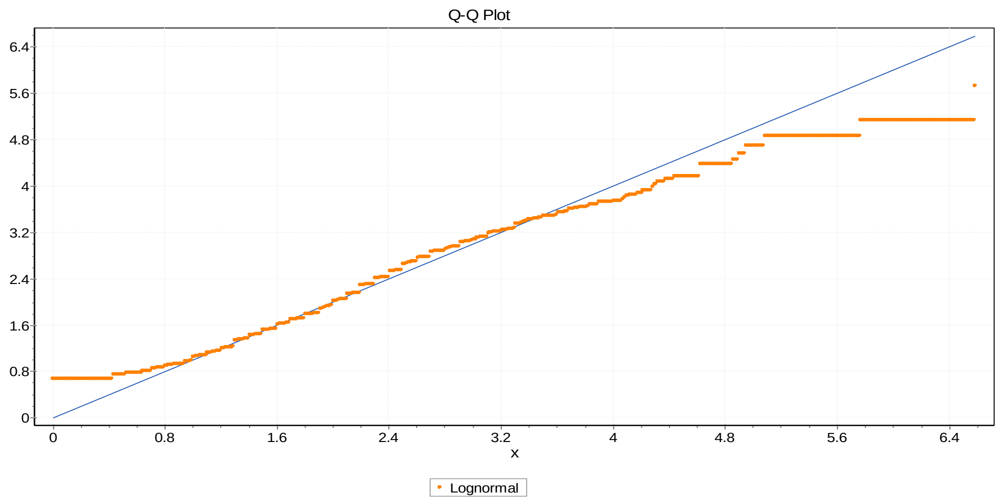

Figure 3- A quintile – quintile plot for the fitted mod lognormal Value to earthquake magnitude

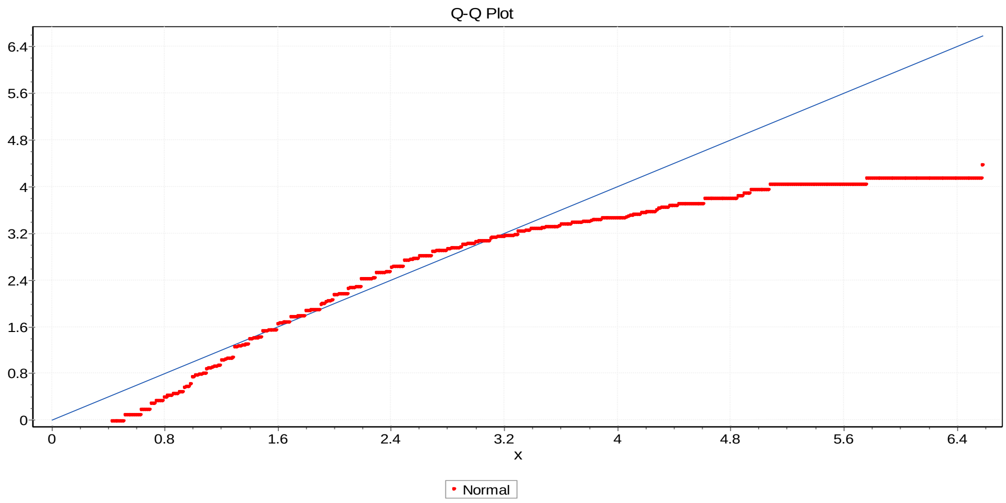

Figure 4- A quintile – quintile plot for the fitted normal model to earthquake magnitude

Results and Conclusions

1. According to the QQ plots given in Figures- (1, 2, 3) and (4) , it is clear that the Burr distribution is fitted better than the GEN Extreme value distributions.
2. Conclusions based on the goodness of fit tests (Kolmogorov Smirnov, Anderson Darling, and Chi-Squared) shows that the Burr distribution is fitted better followed by the GEN Extreme value distributions. And hence lognormal distribution, and normal distribution
3. we can conclude that the shape parameter ξ is actually different
4. They can conduct this study in Iraq case of availability of the data

References

1. Dawood , A., S. **2009**. Probability Analysis of Extreme Monthly Rainfall in Mosul City, North of Iraq. *Marsh Bulletin*, 4(1), pp: 60-74.
2. Author Paul Denton, J. **2004**. How frequent are earthquakes .Natural Environment research conncil
3. Andrzej Kijko. **2004.** Estimation of the Maximum Earthquake Magnitude, *m*.P *ure and applied geophysics August*, 161(8) pp: 1655-1681
4. Kijko, A. and SELLO VLL, M.A. **2006**. Estimation of earthquake hazard parameters from incomplete data files. Part Incorporation of magnitude heterogeneity ,seismological society of America ,Issue 0037-1106, pp: 1655-1681
5. Mohan D. Pant and Todd C. Headrick **2005**. A Method for Simulating Burr Type III and Type XII Distributions through *L*-Moments and *L* Correlations *University of Texas at Arlington Southern* Illin.Beirlant, J; Y. Goegebeur; J. Segers; J. Teugels Statistics of Published in *ISRN Applied Mathematics*, 2013, Article ID 191604 at doi: 10.1155/2013/191604
6. Amir Agha Kouchak and Nasrin Nasrollahi. **2010**. Semi-parametric and Parametric Inference of Extreme Value Models for Rainfall Data, Water Resour Manage, 24, pp:1229–1249
7. John Norstad **2011** The Normal and Lognormal Distributions , *1999 Updated: November 3,*
8. Ananth Sankar and Chin-Hui Lee A Maximum-likelihood A pproach to stochastic Matching for *Robust Speech Reecognition AT&T Bell Loboratories Murray Hill,NJ 07974 2009*
9. Najah Abd . **2013** Induced Seismicity of Central . Arkansas, by, December . *Unpblish D.PH.thesis* University of Arkansas at Little Rock.
10. Tasnim .H.K. AlBaldawi and Zaina Alzuabid **. 2015** Statistical Analysis of Extreme Rainfall Data in Baghdad City**.** Unpblish MSc.Thesis , Baghdad University, Baghdad, Iraq.