

# Three-parameter generalized exponential distribution in earthquake recurrence interval estimation

Sumanta Pasari · Onkar Dikshit

Received: 23 January 2013 / Accepted: 13 February 2014 / Published online: 5 March 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** The purpose of this article is to study the three-parameter (scale, shape, and location) generalized exponential (GE) distribution and examine its suitability in probabilistic earthquake recurrence modeling. The GE distribution shares many physical properties of the gamma and Weibull distributions. This distribution, unlike the exponential distribution, overcomes the burden of memoryless property. For shape parameter  $\beta > 1$ , the GE distribution offers increasing hazard function, which is in accordance with the elastic rebound theory of earthquake generation. In the present study, we consider a real, complete, and homogeneous earthquake catalog of 20 events with magnitude above 7.0 (Yadav et al. in *Pure Appl Geophys* 167:1331–1342, 2010) from northeast India and its adjacent regions (20°–32°N and 87°–100°E) to analyze earthquake inter-occurrence time from the GE distribution. We apply the modified maximum likelihood estimation method to estimate model parameters. We then perform a number of goodness-of-fit tests to evaluate the suitability of the GE model to other competitive models, such as the gamma and Weibull models. It is observed that for the present data set, the GE distribution has a better and more economical representation than the gamma and Weibull distributions. Finally, a few conditional probability curves (hazard curves) are presented to demonstrate the significance of the GE distribution in probabilistic assessment of earthquake hazards.

**Keywords** Recurrence interval · Memoryless · Generalized (exponentiated) exponential distribution · Conditional probability · Northeast India

## 1 Introduction

It has long been observed that moderate to great earthquakes usually occur in a repetitive manner (Rikitake 1976; Utsu 1984; Kagan and Jackson 1991; Faenza et al. 2008; Working Group 2008; Chen et al. 2013). Various probability distributions, namely the exponential

---

S. Pasari (✉) · O. Dikshit  
Department of Civil Engineering, Indian Institute of Technology, Kanpur 208016, India  
e-mail: spasari@iitk.ac.in; sumanta.pasari@gmail.com

(Poisson), gamma, lognormal, and Weibull distributions, are regularly used to model interevent times of such earthquakes (Utsu 1984; Cornell and Winterstein 1986; Nishenko and Buland 1987; Anagnos and Kiremidjian 1988; Parvez and Ram 1997; SSHAC 1997; Yadav et al. 2010; Yazdani and Kowsari 2011; Chen et al. 2013; Pasari and Dikshit 2013). These distributions, though very popular because of their easy interpretation and robust application in several fields, have certain drawbacks. For instance, the exponential distribution, which is also connected to the discrete Poisson distribution, possesses the memoryless property, which contradicts the physics of the earthquake-generating mechanism as illustrated in the ‘elastic rebound theory’ (Reid 1910). Similarly, computation of the distribution function or survival function of a gamma distribution is difficult when the shape parameter is non-integer (Johnson et al. 1995). The closed form of the distribution function of the sum of independent and identically distributed (i.i.d.) Weibull or lognormal random samples is hardly available. This implies that the distribution of the mean of i.i.d. Weibull or lognormal random samples is difficult to obtain (Gupta and Kundu 1999). Besides, the Weibull and lognormal distributions do not show a reproductive (hereditary) property. Therefore, we often end up with situations where studying of other probability models may be required.

Apart from the exponential (Poisson), gamma, lognormal, and Weibull distributions, previous studies have used the triple exponential model (Kijko and Sellevoll 1981), Brownian passage time distribution (Matthews et al. 2002), Pareto distribution (Kagan and Schoenberg 2001; SSHAC 1997), Rayleigh distribution (Yazdani and Kowsari 2011), negative binomial distribution (Dionysiou and Papadopoulos 1992), generalized gamma distribution (Bak et al. 2002), and a few other distributions (SSHAC 1997; Working Group 2008) to determine the underlying pattern of earthquake interevent times and subsequently to strengthen the concept of empirical recurrence modeling. Nevertheless, the most appropriate and versatile distribution function for earthquake recurrence modeling still remains an open research question.

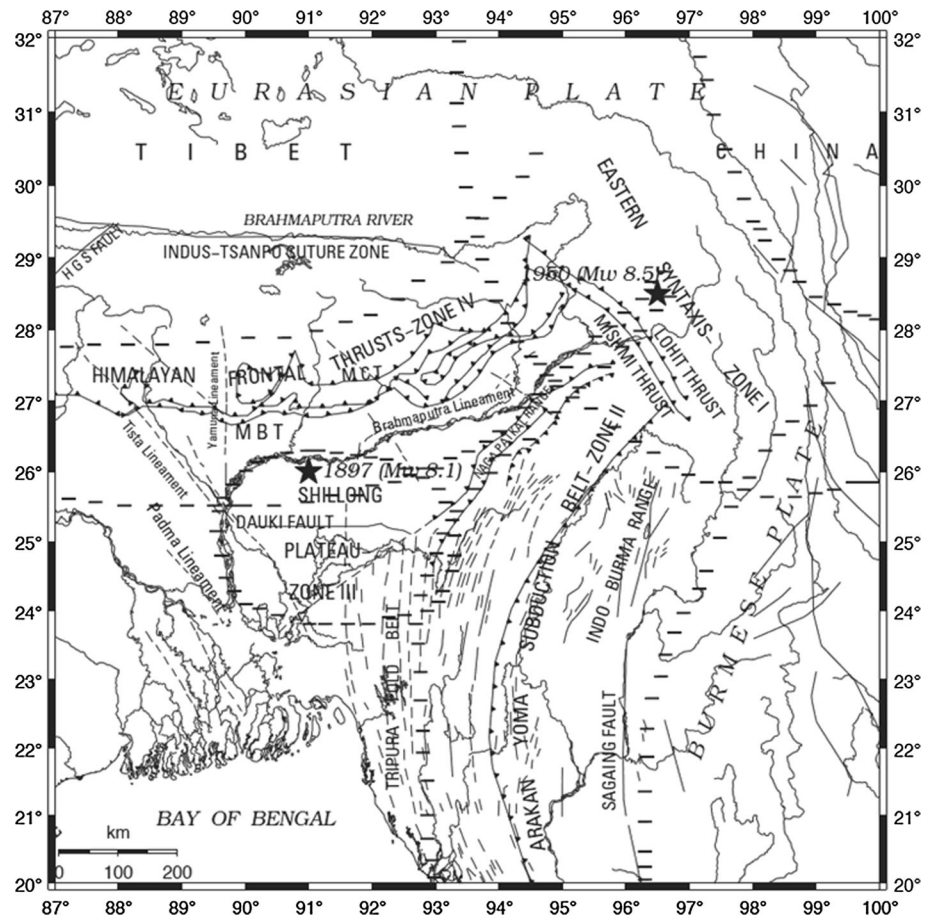
### 1.1 Scope and objective

Earthquake recurrence modeling in northeast India using the exponential, gamma, lognormal, and Weibull models was carried out previously by Parvez and Ram (1997), Yadav et al. (2010), and Pasari and Dikshit (2013). This article uses the GE distribution to estimate the recurrence time of large earthquakes.

The GE distribution (Gupta and Kundu 1999) is a particular member of the general class of exponentiated distributions proposed by Gupta et al. (1998) as  $F(t) = [G(t)]^\beta$ , where  $G(t)$  is the base distribution and  $\beta > 0$  is a shape parameter. It is also known as the exponentiated exponential distribution (Gupta and Kundu 2007). The GE distribution shares many physical properties of gamma and Weibull distributions. This distribution, as an important tool for lifetime data analysis, has widespread applications in the field of medical and biological research (Gupta and Kundu 1999, 2007). However, no study has been carried out to date to explore the suitability of the GE distribution in the field of natural hazards. In view of this, the present research aims to introduce the three-parameter GE model in seismic recurrence studies and to investigate its effectiveness in earthquake recurrence modeling of large earthquakes. The efficacy of the GE model is determined by comparing it to the popular gamma and Weibull models that have been serving as leading earthquake recurrence models for the last few decades. The overall methodology for the present study is arranged in three steps: model description, parameter estimation, and model comparison. Toward the end, we also provide a few conditional probability curves (using the GE distribution) to assess future earthquake hazards in the study region.

1.2 Seismotectonic settings of the study area and earthquake data file description

We consider the northeast part of India and its adjoining regions (20°–32°N and 87°–100°E) for the present study. This region falls under seismic zones V (most seismically active zone), IV (high), and III (moderate) on the seismic zonation map of India (BIS 2002). A number of active thrust faults (Fig. 1), namely main boundary thrust (MBT), main central thrust (MCT), Lohit thrust, Misami thrust, and Sagaing thrust, are present in the study area (Gupta et al. 1986; Yadav et al. 2009). This region has experienced many large earthquakes in the past (Thingbaijam et al. 2008). Among these, two massive great earthquakes (marked in Fig. 1), namely the Shilong plateau earthquake of June 12, 1897 (Mw 8.1) and the upper Assam earthquake of 15 August 1950 (Mw 8.5) caused extensive loss of life and property in the Indian subcontinent (Oldham 1899; Poddar 1950; Molnar and Pandey 1989; Bilham and England 2001). A detailed discussion of the general



**Fig. 1** Seismotectonic map of northeast India and its adjacent regions consisting of several faults, thrusts, lineaments, and structural features; two stars mark the epicentral locations of the 1897 Shilong earthquake and 1950 Assam earthquake. The map also highlights four active seismogenic source zones: eastern syntaxis-zone I, Arakan-Yoma subduction belt-zone II, Shilong plateau-zone III, and Himalayan frontal thrusts including MCT and MBT-zone IV (modified after Gupta et al. 1986 and Yadav et al. 2009)

**Table 1** List of  $M \geq 7.0$  earthquakes that occurred (since 1846) in the study region (after Yadav et al. 2010)

S. No.	Date						Location		Focal depth (km)	Magnitude
	Year	Month	Day	Hour	Min	Sec	Latitude (N)	Longitude (E)		
1	1846	12	10				26.00	93.00		7.5
2	1868	6	30				24.50	91.50		7.5
3	1885	1	1				25.40	90.00		7.3
4	1897	6	12	11	5		25.90	91.90	60	8.7
5	1908	12	12				26.50	97.00		7.5
6	1912	5	23	2	24		21.00	97.00	25	7.9
7	1918	7	8	10	22		24.50	91.00	60	7.6
8	1923	9	9	22	3	42.00	25.25	91.00		7.1
9	1931	1	27	20	9		25.60	96.80	60	7.6
10	1943	10	23				21.50	93.50		7.2
11	1946	9	12	15	20		23.50	96.00	60	7.5
12	1947	7	29	13	43		28.50	94.00	60	7.9
13	1950	8	15	14	9		28.50	96.70		8.6
14	1951	11	18	9	35	45.00	30.50	91.00		8.0
15	1954	3	21	23	42		24.40	95.20	180	7.5
16	1961	2	4	8	51		24.90	93.34	141	7.6
17	1976	5	29	14	0	18.50	24.53	98.71	10	7.0
18	1988	8	6	0	36	24.60	25.15	95.13	91	7.2
19	1991	1	5	14	57	11.50	23.61	95.90	20	7.1
20	1995	7	11	21	46	39.78	21.97	99.20	12	7.1

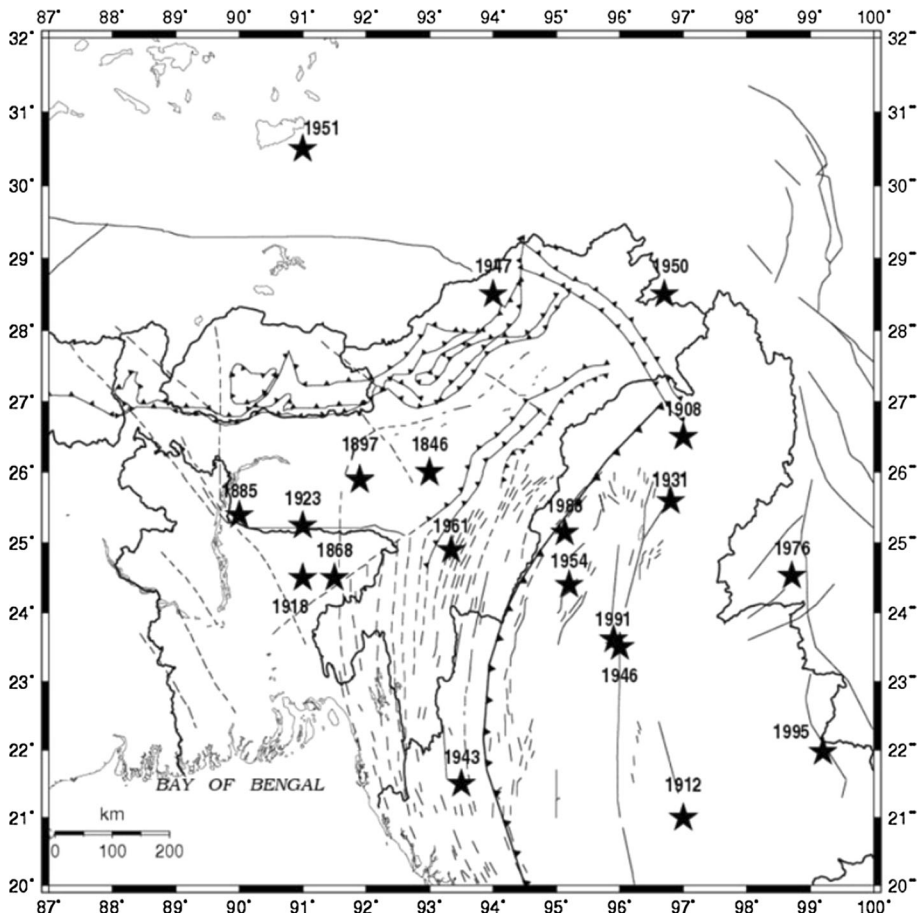
seismotectonic setting, historical seismicity, and earthquake losses of northeast India may be found in Gupta et al. (1986), Nandy (1986), Kayal (1996), Thingbaijam et al. (2008), and the references therein.

We use a real, complete, and homogeneous earthquake catalog (Yadav et al. 2010) of 20 events ( $M \geq 7.0$ ) spanning the period 1846–1995. These events are listed in Table 1, and their geographical epicentral locations are shown in Fig. 2. At this point, it can be noted that no major earthquake ( $M \geq 7.0$ ) has occurred in the study region since 1995. Thus, the present catalog accounts all main shocks with magnitude  $M \geq 7.0$  for the period 1846–2013.

It is evident from Table 1 that most of the Himalayan earthquakes are of shallow-type (focal depth  $< 80$  km) earthquakes and thus more hazardous. The present study, however, aims to estimate future earthquake occurrences purely on the basis of empirical modeling of earthquake event gaps (time); hence, it does not consider any kind of social, positional, geological, or geophysical influences in the analysis.

## 2 Preliminaries and the GE model description

Let  $T$  be a positive random variable of inter-occurrence times of successive events with the density function  $f(t)$ , distribution function  $F(t)$ , and hazard function  $h(t)$ . Let  $t_e$  and  $\tau$  denote the elapsed time (time beyond the last occurrence) and the residual time (remaining time to



**Fig. 2** Epicentral locations of  $M \geq 7.0$  earthquakes (as listed in Table 1) that occurred in northeast India and its surrounding regions (after Yadav et al. 2009, 2010)

a future occurrence), respectively. Knowing the elapsed time, the residual time is a random variable. Let  $P(\tau|t_e)$  be the conditional probability of earthquake occurrence, i.e., the probability of an earthquake to occur during time interval  $t_e$  and  $t_e + \tau$  given that no earthquake has been triggered during last  $t_e$  year(s).

### 2.1 Three-parameter generalized exponential (GE) distribution

The three-parameter generalized (exponentiated) exponential distribution (Gupta and Kundu 1999) uses two-parameter exponential distribution  $F_{Exp}(t; \alpha, \gamma)$  as the base distribution. Thus, the distribution function ( $F_{GE}$ ) of  $T \sim GE(\alpha, \beta, \gamma)$  can be defined (Gupta et al. 1998) as

$$F_{GE}(t; \alpha, \beta, \gamma) = [F_{Exp}(t; \alpha, \gamma)]^\beta = \left(1 - e^{-\frac{t-\gamma}{\alpha}}\right)^\beta \quad (t > \gamma, \alpha > 0, \beta > 0) \quad (1)$$

The GE distribution is controlled by three parameters—the scale parameter  $\alpha(> 0)$  that restricts the spread of  $t$ , the shape parameter  $\beta(> 0)$  that determines the appearance of the distribution, and the location parameter  $\gamma(< t)$  that controls the range of the distribution.

The shape parameter  $\beta$ , among all three parameters, plays the most important role in GE model description. It is easy to note that if  $\beta = 1$ , GE distribution exactly coincides with its base distribution, i.e., the two-parameter exponential distribution. Therefore, the GE distribution, like the gamma and Weibull distributions, is an extension (generalization) of classical exponential distribution. The density function  $f_{GE}(t)$  and the hazard function  $h_{GE}(t)$  of  $T \sim GE(\alpha, \beta, \gamma)$  are given below.

$$f_{GE}(t; \alpha, \beta, \gamma) = \frac{\beta}{\alpha} \left(1 - e^{-\frac{t-\gamma}{\alpha}}\right)^{\beta-1} e^{-\frac{t-\gamma}{\alpha}} \quad (t > \gamma > 0, \alpha > 0, \beta > 0) \tag{2}$$

$$h_{GE}(t; \alpha, \beta, \gamma) = \frac{\beta \left(1 - e^{-\frac{t-\gamma}{\alpha}}\right)^{\beta-1} e^{-\frac{t-\gamma}{\alpha}}}{1 - \left(1 - e^{-\frac{t-\gamma}{\alpha}}\right)^{\beta}} \quad (t > \gamma > 0, \alpha > 0, \beta > 0) \tag{3}$$

The expression for the corresponding conditional probability  $P_{GE}(\tau|t_c)$  for  $GE(\alpha, \beta, \gamma)$  is given as

$$P_{GE}(\tau|t_c) = P(T \leq t_c + \tau | T > t_c) = \frac{\left(1 - e^{-\frac{t_c-\gamma+\tau}{\alpha}}\right)^{\beta} - \left(1 - e^{-\frac{t_c-\gamma}{\alpha}}\right)^{\beta}}{1 - \left(1 - e^{-\frac{t_c-\gamma}{\alpha}}\right)^{\beta}} \tag{4}$$

$(t > \gamma > 0, \alpha > 0, \beta > 0)$

Since the present study also aims to compare GE distribution with gamma and Weibull distributions, we mention the gamma and Weibull density functions below. The associated distribution functions and hazard functions can be easily calculated from the respective density functions.

$$f_{Weibull}(t; \alpha, \beta) = \frac{\beta}{\alpha^{\beta}} t^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \quad (t > 0, \alpha > 0, \beta > 0) \tag{5}$$

$$f_{Gamma}(t; \alpha, \beta) = \frac{1}{\Gamma(\beta)} \frac{t^{\beta-1}}{\alpha^{\beta}} e^{-\frac{t}{\alpha}} \quad (t > 0, \alpha > 0, \beta > 0)$$

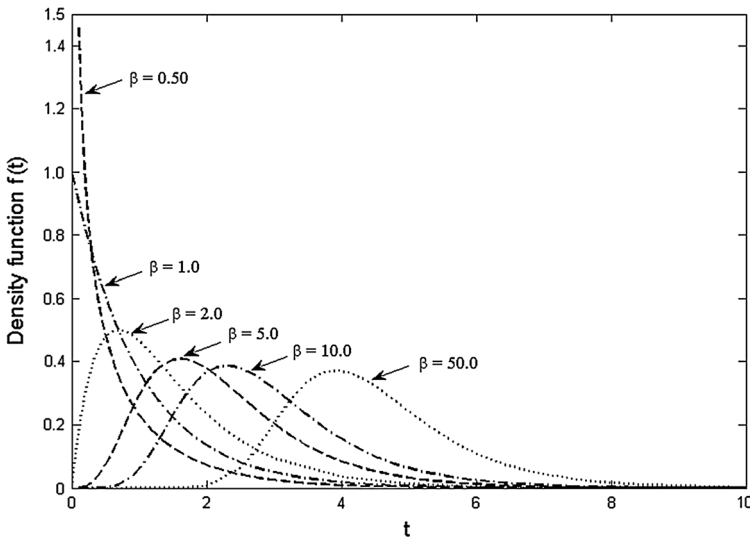
where

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt \quad (z > 0) \tag{6}$$

### 2.2 Model properties

The  $GE(\alpha, \beta, \gamma)$  density function assumes a variety of shapes depending on the shape parameter  $\beta$  (Fig. 3). It is monotonically decreasing for  $\beta \leq 1$ , and for  $\beta > 1$ , it is unimodal, skewed, and right-tailed, similar to Weibull or gamma density functions.

The hazard functions of the GE distribution, like gamma and Weibull hazard functions, assume various shapes depending on the shape parameter  $\beta$ . Specifically, for fixed  $(\alpha, \gamma)$ , the GE hazard function is increasing for  $\beta > 1$  and decreasing for  $\beta < 1$  (Gupta and Kundu 1999). For  $\beta = 1$ , the hazard function becomes constant, and the distribution follows the memoryless property. The different shapes of hazard function provide salient information about the instantaneous rate of failure of an earthquake reliability system and thus are of specific interest to the scientists who study and try to predict earthquakes (Matthews et al. 2002). The plots of the GE hazard function corresponding to different values of  $\beta$  are shown in Fig. 4.



**Fig. 3** Different shapes of the  $GE(1, \beta, 0)$  density function for different values of shape parameter  $\beta$

The moment generating function of  $T \sim GE(\alpha, \beta, \gamma)$  is given (see Appendix 1) as

$$M_T(x) = E(e^{xT}) = \gamma e^{x\gamma} \frac{\Gamma(\beta + 1)\Gamma(1 - \alpha x)}{\Gamma(\beta + 1 - \alpha x)} \quad (\alpha x < 1) \tag{7}$$

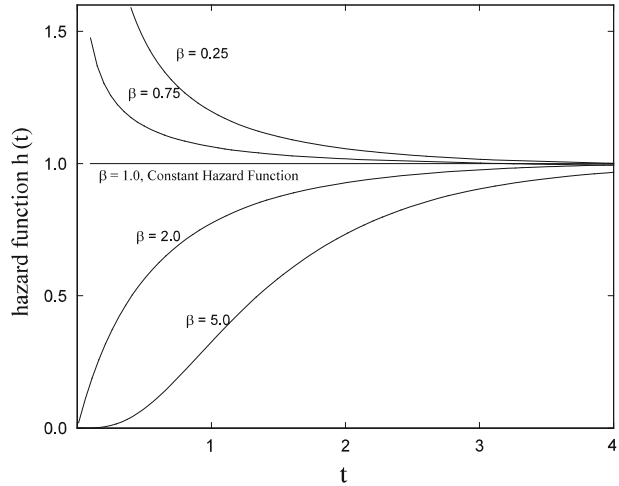
We use  $E(T^n) = \frac{d^n M_T(x)}{dx^n}(0)$  to obtain respective moments (about origin) of  $T \sim GE(\alpha, \beta, \gamma)$ . The mean and variance of the GE distribution are calculated as  $\gamma + \alpha[\psi(\beta + 1) - \psi(1)]$  and  $\alpha^2[\psi'(1) - \psi'(\beta + 1)]$ , respectively;  $\psi(x)$  and  $\psi'(x)$  denote the digamma function and its first derivate. It is observed (Gupta and Kundu 1999) that the mean and variance of the  $GE(\alpha, \beta, \gamma)$  are increasing functions of  $\beta$  (for fixed  $\alpha$ ). More specifically, the variance of the GE distribution increases to  $\frac{\pi^2 \alpha}{6}$ , unlike the variance of gamma distribution, which tends to infinity (as  $\beta$  increases), and the variance of Weibull distribution, which approximately equals  $\frac{\pi^2 \alpha}{6\beta^2}$  (for large values of  $\beta$ ). We also observe that the  $GE(1, \beta, 0)$  is unimodal with mode at  $\log \beta$  if  $\beta > 1$  and at 0 if  $\beta \leq 1$ . In addition,  $GE(1, \beta, 0)$  has its median at  $-\ln(1 - (0.5)^{\frac{1}{\beta}})$ . For large values of  $\beta$ , each of mean, median, and mode of  $GE(1, \beta, 0)$  approximately equals  $\log \beta$  (Gupta and Kundu 1999, 2007). A comprehensive representation of mean, mode, median, variance, and coefficient of variation (CV) is illustrated in Fig. 5.

It can be noted that the GE moment generating function  $M_T(x)$  is not very convenient to handle; thus, the sum of i.i.d. GE random variables is difficult to obtain. As a result, the GE distribution, like the Weibull distribution, does not support the reproductive property (Gupta and Kundu 2007).

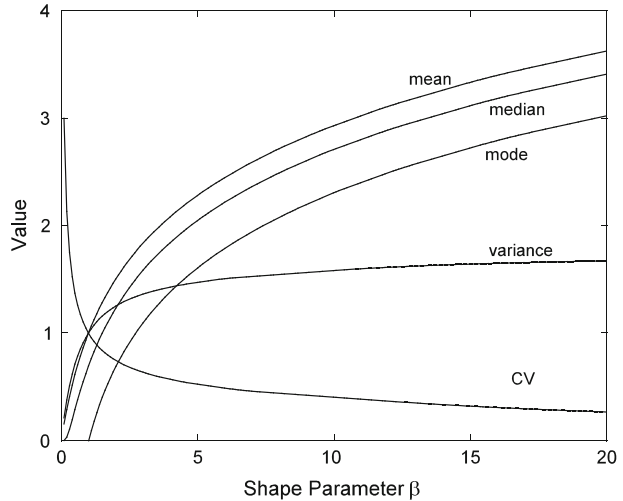
### 3 Inference from the modified maximum likelihood estimation (MMLE)

We apply the modified maximum likelihood estimation (MMLE) method to estimate GE model parameters. The MMLE method is entirely based on the classical maximum

**Fig. 4** Different shapes of the hazard function for different values of  $\beta$ , 0 hazard function for different values of  $\beta$



**Fig. 5** Shapes of mean, mode, median, variance, and coefficient of variation (CV) for the  $GE(1, \beta, 0)$  distribution for different values of shape parameter  $\beta$



likelihood estimation (MLE) method proposed by Fisher in 1912 (Aldrich 1997). If  $\{t_1, t_2, \dots, t_n\}$  is a random sample from  $GE(\alpha, \beta, \gamma)$  distribution, then the log-likelihood function ( $\ln L_{GE}$ ) is

$$\ln L_{GE}(\alpha, \beta, \gamma; t_1, t_2, \dots, t_n) = n \ln \beta - n \ln \alpha - \sum_{i=1}^n \left( \frac{t_i - \gamma}{\alpha} \right) + (\beta - 1) \sum_{i=1}^n \ln \left( 1 - e^{-\frac{t_i - \gamma}{\alpha}} \right) \tag{8}$$

The associated log-likelihood equations are

$$n\alpha + (\beta - 1) \sum_{i=1}^n \left( \frac{(t_i - \gamma)e^{-\frac{t_i - \gamma}{\alpha}}}{1 - e^{-\frac{t_i - \gamma}{\alpha}}} \right) - \sum_{i=1}^n (t_i - \gamma) = 0 \tag{9}$$

$$\frac{n}{\beta} + \sum_{i=1}^n \ln \left( 1 - e^{-\frac{t_i - \gamma}{\alpha}} \right) = 0 \tag{10}$$



$$n\alpha - \alpha(\beta - 1) \sum_{i=1}^n \left( \frac{e^{-\frac{t_i - \gamma}{\alpha}}}{1 - e^{-\frac{t_i - \gamma}{\alpha}}} \right) = 0 \tag{11}$$

The estimated parameters are the solution of Eqs. (9–11) satisfying the constraint that  $\gamma < t_i \forall i$ . However, the regularity conditions, existence, and uniqueness of the solution of Eqs.(9–11) require extensive investigation (Gupta and Kundu 1999, 2007; Raqab and Ahsanullah 2001). Therefore, in practice, we first estimate the location parameter  $\gamma$  as

$$\hat{\gamma} = \min_i t_i = t_{(1)} \tag{12}$$

and then the other parameters are estimated. It is observed that for fixed  $(\alpha, \beta)$ , the likelihood function ( $L_{GE}$ ) is a monotonically increasing function of  $\gamma$ . Further,  $\gamma$  can only assume values between 0 and  $t_{(1)}$ . Thus, the best possible estimate for  $\gamma$  is  $\hat{\gamma} = t_{(1)}$ . However, by doing so, the concept of searching the parameter values that maximize the likelihood function ( $L_{GE}$ ) or log-likelihood function ( $\ln L_{GE}$ ) is violated. This is precisely the reason for the adopted method to be called the local or the modified MLE (MML).

Once  $\hat{\gamma}$  is obtained, all data points  $\{t_1, t_2, \dots, t_n\}$  are shifted right or left to the abscissa on the basis of  $\hat{\gamma}$  and the minimum shifted point (which is 0) is discarded from the set of shifted points. This is essential because for  $\beta < 1$ , the functional value of likelihood function ( $L_{GE}$ ) does not exist [ $L_{GE}$  goes to infinity when the value of the location parameter approaches  $t_{(1)}$ ]; hence, the underlying parameter estimation technique becomes questionable. The shape and scale parameters are estimated from the remaining  $(n - 1)$  shifted data points  $\{t_{(2)} - t_{(1)}, t_{(3)} - t_{(1)}, \dots, t_{(n)} - t_{(1)}\}$ .

The modified log-likelihood function ( $\ln L'_{GE}$ ) is now obtained as

$$\begin{aligned} \ln L'_{GE}(\alpha, \beta, 0; t'_2, t'_3, \dots, t'_n) &= (n - 1) \ln \beta - (n - 1) \ln \alpha - \sum_{i=2}^n \frac{t'_i}{\alpha} \\ &+ (\beta - 1) \sum_{i=2}^n \ln \left( 1 - e^{-\frac{t'_i}{\alpha}} \right) \end{aligned} \tag{13}$$

In the above expression,  $t'_i$  denotes the shifted data point, i.e.,  $t'_i = t_{(i)} - t_{(1)}$ . The corresponding log-likelihood equations are given below.

$$(n - 1)\alpha + (\beta - 1) \sum_{i=2}^n \left( \frac{t'_i e^{-\frac{t'_i}{\alpha}}}{1 - e^{-\frac{t'_i}{\alpha}}} \right) - \sum_{i=2}^n t'_i = 0 \tag{14}$$

$$\frac{n - 1}{\beta} + \sum_{i=2}^n \ln \left( 1 - e^{-\frac{t'_i}{\alpha}} \right) = 0 \tag{15}$$

A simple manipulation of the above equations gives rise to the following Eq. (16) in terms of a single variable  $\alpha$ , which can be easily solved numerically or using any standard software packages such as MAPLE, MATLAB, and R.

$$(n - 1)\alpha - \left( \frac{n - 1}{\sum_{i=2}^n \ln \left( 1 - e^{-\frac{t'_i}{\alpha}} \right)} + 1 \right) \sum_{i=2}^n \left( \frac{t'_i e^{-\frac{t'_i}{\alpha}}}{1 - e^{-\frac{t'_i}{\alpha}}} \right) - \sum_{i=2}^n t'_i = 0 \tag{16}$$

Alternatively, we can estimate  $\alpha$  directly from Eq. (13) by maximizing

$$\begin{aligned}
 g(\alpha) &= \ln L'_{\text{GE}}(\alpha, \beta(\alpha), 0; t'_2, t'_3, \dots, t'_n) \\
 &= K - (n-1) \ln \left( - \sum_{i=2}^n \ln \left( 1 - e^{-\frac{t'_i}{\alpha}} \right) \right) - (n-1) \ln \alpha - \sum_{i=2}^n \ln \left( 1 - e^{-\frac{t'_i}{\alpha}} \right) - \sum_{i=2}^n \frac{t'_i}{\alpha} \\
 K &\equiv \text{constant term}
 \end{aligned} \tag{17}$$

Two approaches to solve (17) are provided in Appendix 2.

Table 2 provides the estimated model parameter values of the GE, gamma, and Weibull distributions. The maximum likelihood estimation methods of the gamma and Weibull models can be found in Johnson et al. (1995).

From Table 2, we see that the estimated shape parameter  $\beta$  of the GE model is greater than 1, meaning the GE density function for the present earthquake catalog is unimodal, right-tailed, and log-concave, and the estimated GE hazard function is monotonically increasing.

#### 4 Model selection

We use three well-established model selection criteria, namely the maximum likelihood criterion and its modification, known as the Akaike information criterion (AIC), the Kolmogorov-Smirnov (K-S) minimum distance criterion, and the chi-square criterion to appraise the suitability of GE distribution in comparison to gamma and Weibull distributions (Pasari and Dikshit 2013).

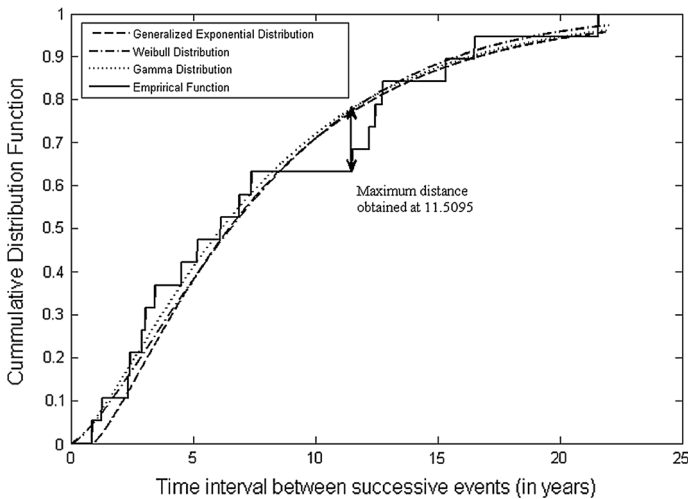
The maximum likelihood criterion uses the maximum log-likelihood value ( $\ln L$ ) to determine the best suitable model. This criterion, however, assumes that the number of parameters ( $k$ ) in each competitive model is the same. To overcome this limitation, several modifications have been proposed. Among these, the Akaike information criterion (Akaike 1974) defined by  $AIC = 2k - 2 \ln L$  has been widely accepted. The AIC involves a penalty function to account for model complexity due to the unequal number of parameters. The model with the minimum AIC value is tagged as the most suitable model. The log-likelihood values and AIC values of each distribution are listed in Table 4.

The Kolmogorov-Smirnov (K-S) minimum distance criterion uses K-S distances between empirical distribution function and probability distribution function to choose the most appropriate model. The K-S test belongs to the family of non-parametric and distribution-free goodness-of-fit tests (Johnson et al. 1995). A step-by-step procedure to obtain the K-S value is described in Appendix 3, and the calculated K-S distances are listed in Table 4. The associated K-S test plot is given in Fig. 6.

The minimum chi-square criterion is one of the oldest conventional techniques for model selection. This criterion uses the observed and expected frequencies of class intervals to calculate the chi-square value (see Appendix 4). However, in the chi-square test, there is no specified rule to choose the number and width of class intervals (Johnson et al. 1995). For this reason, we choose a uniform interval size of 3 years (6 classes: <3, 3–6, 6–9, 9–12, 12–15, and >15) to calculate the chi-square value. According to the chi-square test, the model with the minimum chi-square value is considered the best model. The different chi-square values along with their observed and expected frequencies are summarized in Table 3.

**Table 2** Parameter estimation from the (modified) maximum likelihood estimation technique

Probability model	Parameter	MLE/MMLE
GE	$\alpha$	6.038483
	$\beta$	1.374836
	$\gamma$	0.876712
Gamma	$\alpha$	5.218260
	$\beta$	1.499642
Weibull	$\alpha$	8.559719
	$\beta$	1.356785



**Fig. 6** The K-S test plot as a difference between the estimated cumulative distribution function for each test model with the empirical distribution function

Table 4 shows that the GE distribution, among three competitive distributions, has the minimum AIC value. This strongly suggests that the GE distribution is more economical as compared to gamma or Weibull distribution. In addition, we observe that the GE distribution has the minimum K-S distance value. This implies that the nonparametric K-S criterion also suggests GE distribution as the most appropriate distribution to represent the present earthquake catalog. The chi-square criterion, on the other hand, suggests the gamma distribution to be the best fitted model for the present data set.

A close look at the K-S curve (Fig. 6) substantiates that all three distributions cross each other, and all of them are quite close to the empirical distribution function. Therefore, at this stage, it is difficult to decide which one of these three competitive distributions performs better in a global sense. As a matter of fact, we avoid claiming that the GE is the most appropriate distribution for recurrence modeling. Rather, we argue that the GE distribution can be effectively used as a practical alternative to the gamma and Weibull distributions in earthquake recurrence modeling and associated problems. In the following section, we apply the GE model to generate a number of estimated conditional probability curves to measure earthquake hazards in the study region.

**Table 3** The chi-square criterion for different models

Model	Interval	Observed frequency	Expected frequency	$\chi^2_{\text{value}}$
GE	<3	5	3.57	4.12
	3–6	4	5.24	
	6–9	3	3.74	
	9–12	1	2.44	
	12–15	3	1.54	
	>15	3	2.47	
Gamma	<3	5	4.47	3.33
	3–6	4	4.80	
	6–9	3	3.52	
	9–12	1	2.35	
	12–15	3	1.50	
	>15	3	2.36	
Weibull	<3	5	4.07	2.76
	3–6	4	4.68	
	6–9	3	3.73	
	9–12	1	2.61	
	12–15	3	1.67	
	>15	3	2.23	

**Table 4** Model comparison using three criteria: the maximum log-likelihood criterion ( $\ln L$ ), AIC, minimum K-S distance criterion, and minimum chi-square ( $\chi^2_{\text{value}}$ ) criterion

Model	Criteria	Statistical methods Values
GE	$\ln L$	–53.4086
	AIC	112.8172
	$\chi^2_{\text{value}}$	3.3562
	K-S value <sup>a</sup>	0.1400 (11.5095)
Gamma	$\ln L$	–56.9075
	AIC	117.8149
	$\chi^2_{\text{value}}$	2.7116
	K-S value <sup>a</sup>	0.1482 (11.5095)
Weibull	$\ln L$	–56.8651
	AIC	117.7303
	$\chi^2_{\text{value}}$	2.7594
	K-S value <sup>a</sup>	0.1440 (11.5095)

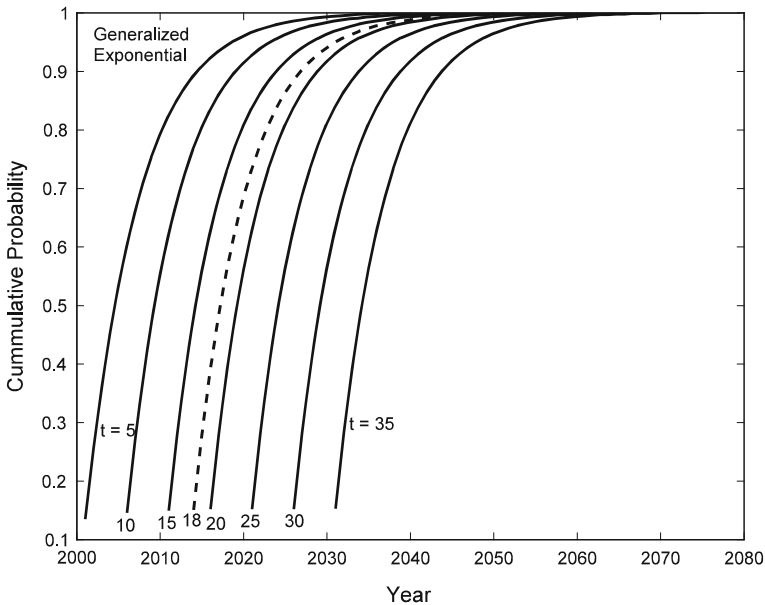
<sup>a</sup> Number in parentheses is the abscissa value (approximately) where the K-S distance is achieved

### 5 Earthquake hazard assessment

Following Utsu (1984), we assess earthquake hazards in terms of the estimated recurrence interval and conditional probability values. Subsequently, we generate a number of conditional probability curves (hazard curves) for different combination of elapsed time ( $t_e$ ) and residual time ( $\tau$ ). These estimated hazard curves play a significant role in seismic zonation and microzonation, evaluation of existing building codes, city planning and

**Table 5** Conditional probabilities of an earthquake of magnitude  $M \geq 7.0$  to occur in the next  $\tau$  year(s), given that no earthquake has occurred in the last  $t_e$  year(s) since the last occurrence of 1995

$\tau$	$t_e$ elapsed time							
	5 (2000)	10 (2005)	15 (2010)	18 (2013)	20 (2015)	25 (2020)	30 (2025)	35 (2030)
3	0.360	0.380	0.387	0.389	0.390	0.391	0.391	0.391
6	0.600	0.619	0.625	0.627	0.628	0.629	0.629	0.630
9	0.753	0.767	0.771	0.773	0.773	0.774	0.775	0.775
12	0.848	0.858	0.861	0.862	0.862	0.863	0.863	0.863
15	0.907	0.913	0.915	0.916	0.916	0.916	0.917	0.917
18	0.943	0.947	0.948	0.949	0.949	0.949	0.949	0.950
21	0.966	0.968	0.969	0.969	0.969	0.969	0.969	0.969
24	0.979	0.980	0.981	0.981	0.981	0.981	0.981	0.981
27	0.987	0.988	0.988	0.988	0.989	0.989	0.989	0.990
30	0.992	0.993	0.993	0.993	0.993	0.993	0.993	0.993



**Fig. 7** The estimated conditional probability curves (hazard curves) for elapsed time  $t_e = 5, 10, \dots, 35$  years using the GE distribution for earthquake events  $M \geq 7.0$  in the study region. The dotted line corresponds to an elapsed time of 18 years, i.e., 2013

infrastructure development, designing of highway/railway bridges, location choice of nuclear power plants, schools, and hospitals, and many other engineering applications (Baker 2008; Yadav et al. 2009, 2010).

The expected mean recurrence interval of an earthquake ( $M \geq 7.0$ ) from the GE model is  $8.23 \pm 6.40$  years in comparison to  $7.83 \pm 6.39$  years from the gamma model and  $7.84 \pm 5.84$  years from the Weibull model. The estimated cumulative probability (from the GE model) of a large  $M \geq 7.0$  magnitude earthquake by 2015 is 0.94, which is

critically high. The same is found to be 0.95 and 0.96 from the gamma and Weibull models, respectively. On the other hand, the conditional probability (from the GE model) reaches 0.80–0.90 after about 10–14 years (2023–2027) and 0.90–0.95 after about 14–18 years (2027–2031) for an elapsed time of 18 years (i.e., 2013). A comprehensive list of estimated conditional probabilities using the GE model is presented in Table 5, and the corresponding conditional probability curves (hazard curves) are schematically shown in Fig. 7.

## 6 Summary and Conclusions

This article has presented a study on the three-parameter generalized (exponentiated) exponential distribution (Gupta and Kundu 1999) and has investigated its scope in seismic recurrence studies. The purpose was to increase the choices of potential models to estimate earthquake interevent times. A detailed description of the GE model, its parameter estimation, and model selection techniques was provided. It was observed that the GE distribution, unlike the gamma distribution, has a tractable distribution function or survival function, which makes the GE distribution computationally much easier to handle than the gamma distribution. The GE distribution, like the gamma or Weibull distribution, offers both monotonically increasing and decreasing hazard functions, which play a significant role in seismic reliability analysis. This facility, however, was not previously available from an exponential (Poisson) distribution that only provides a constant failure rate. The moment generating function of the GE distribution is quite intractable. As a consequence, the GE distribution, unlike the gamma distribution and like the Weibull distribution, does not preserve the hereditary (reproductive) property.

For illustrative purposes, we used a real, homogeneous, and complete earthquake catalog of 20 events ( $M \geq 7.0$ ) from northeast India and its adjoining regions. The estimated (MMLE) shape parameter ( $\hat{\beta} = 1.374836$ ) reveals that the underlying GE distribution is unimodal and right-tailed, and the corresponding hazard function increases monotonically. In order to compare the GE distribution with the gamma and Weibull distributions, we applied three model selection criteria, namely, the maximum likelihood criterion and its extension (AIC), the K-S minimum distance criterion, and the chi-square criterion. It was observed that two criteria, namely the maximum likelihood test (AIC) and the K-S test, suggest that the GE model has comparatively better fitting for the present data set. This shows the efficacy of the GE model as a practical alternative to other popular probability models for analysis by seismologists and earthquake professionals. We have also presented a few conditional probability curves (hazard curves) for elapsed time  $t_e = 5, 10, \dots, 35$  years. These curves indicate very high chances of future earthquakes in the study region. The conditional probability of a large magnitude event by 2023–2027 and by 2027–2040 reaches to 0.80–0.90 and 0.90–0.99, respectively.

The present study has demonstrated the use of GE distribution in probabilistic earthquake recurrence modeling of northeast India and its adjoining regions. However, more work is needed to confirm the suitability of GE distribution for a broad spectrum of the earthquake catalog from different parts of the globe.

**Acknowledgments** We thank Prof. Debasis Kundu of IIT Kanpur for clarifying many doubts related to the GE distribution. We also thank Dr. R.B.S. Yadav of Kurukshetra University for his suggestions. We are pleased to thank two anonymous reviewers and the editor-in-chief Prof. Thomas Glade for their constructive

comments and useful suggestions for improving the present work. Financial support to S.P. by CSIR, India, is duly acknowledged.

**Appendix 1**

We put  $\alpha = 1, \gamma = 0$  in (3) for the sake of simplicity. Also, we replace the random variable  $T$  by  $U$ . Then, the corresponding density function becomes

$$f_{GE}(u; 1, \beta, 0) = \beta e^{-u}(1 - e^{-u})^{\beta-1} \quad (u > 0, \beta > 0) \tag{18}$$

Let  $M_U(x)$  denote the moment generating function (mgf). So, by definition

$$M_U(x) = E(e^{xU}) = \beta \int_0^\infty e^{(x-1)u} (1 - e^{-u})^{\beta-1} du = \beta \frac{\Gamma(\beta)\Gamma(1-x)}{\Gamma(\beta+1-x)} \quad (x < 1) \tag{19}$$

Therefore, the moment generating function  $M_T(x)$  of  $T(= \alpha U + \gamma) \sim GE(\alpha, \beta, \gamma)$  is obtained as

$$M_T(x) = E(e^{xT}) = \gamma e^{x\gamma} \frac{\Gamma(\beta+1)\Gamma(1-\alpha x)}{\Gamma(\beta+1-\alpha x)} \quad (\alpha x < 1) \tag{20}$$

**Appendix 2**

It is observed in Gupta and Kundu (1999) that  $g(\alpha)$  in Eq. (17) is unimodal. Thus, in order to find its maximum value, we differentiate  $g(\alpha)$  with respect to  $\alpha$  and equate the resultant expression to zero. This yields the following equation in  $\alpha$ .

$$g'(\alpha) = -(n-1) \frac{\sum_{i=2}^n \left( \frac{t_i e^{-\frac{t'_i}{\alpha}}}{1 - e^{-\frac{t'_i}{\alpha}}} \right)}{\sum_{i=2}^n \ln \left( 1 - e^{-\frac{t'_i}{\alpha}} \right)} + (n-1)\alpha - \sum_{i=2}^n \left( \frac{t_i e^{-\frac{t'_i}{\alpha}}}{1 - e^{-\frac{t'_i}{\alpha}}} \right) - \sum_{i=2}^n t'_i = 0 \tag{21}$$

Equation (21) can be solved in various ways, such as numerical techniques (e.g., fixed point iteration and the Newton-Raphson method) or by using any standard one-dimensional non-linear equation solver package. For completeness, we have provided schemes of the fixed point iteration and the Newton-Raphson methods below.

(a) Fixed point iteration

We first write  $g'(\alpha) = 0$  as  $h(\alpha) = \alpha$  where,

$$h(\alpha) = \left[ \frac{\sum_{i=2}^n \left( \frac{t_i e^{-\frac{t'_i}{\alpha}}}{1 - e^{-\frac{t'_i}{\alpha}}} \right)}{\sum_{i=2}^n \ln \left( 1 - e^{-\frac{t'_i}{\alpha}} \right)} + \frac{1}{n-1} \sum_{i=2}^n \left( \frac{t'_i e^{-\frac{t'_i}{\alpha}}}{1 - e^{-\frac{t'_i}{\alpha}}} \right) + \frac{1}{n-1} \sum_{i=2}^n t'_i \right] \tag{22}$$

Then, we apply the scheme of fixed point iteration as

$$\alpha_{j+1} = h(\alpha_j); \quad j = 1, 2, \dots \tag{23}$$

## (b) Newton-Raphson method

Equation (21) can also be solved by the Newton-Raphson scheme for non-linear equation. The scheme is given as

$$\alpha_{j+1} = \alpha_j - \frac{g'(\alpha_j)}{g''(\alpha_j)}; \quad j = 1, 2, \dots \quad (24)$$

**Appendix 3**

In the K-S test, we first construct the empirical distribution function  $H_n$  for  $n$  i.i.d. random variables  $T_1, T_2, \dots, T_n$  as

$$H_n(t) = \frac{1}{n} \sum_{i=1}^n I_{T_i \leq t} \quad (25)$$

Here,  $I_{T_i \leq t}$  is the indicator function, equals 1 if  $T_i \leq t$ , and otherwise equals to 0. This makes  $H_n(t)$  a step function. Suppose we have two competitive models  $F$  and  $G$ . Then, the corresponding K-S distances are calculated as

$$\begin{aligned} D_1 &= \sup_{-\infty < t < \infty} |H_n(t) - F(t)| \\ D_2 &= \sup_{-\infty < t < \infty} |H_n(t) - G(t)| \end{aligned} \quad (26)$$

In the above expression,  $\sup_t$  denotes the supremum of the set of distances. If  $D_1 < D_2$ , we choose model  $F$ ; otherwise, we choose model  $G$ .

**Appendix 4**

For simplicity, we assume two competitive models  $F$  and  $G$  to describe the chi-square criterion. We further assume that  $f(t; \tilde{\theta})$  and  $g(t; \tilde{\varphi})$  are the corresponding fitted models of  $F$  and  $G$ . In the chi-square test, we first divide the range of sample observations into  $k$  equal parts and record its observed frequencies. We then compute expected frequencies for all  $k$  parts using fitted models. Suppose the observed frequencies are  $n_1, n_2, \dots, n_k$  and expected frequencies are  $f_1, f_2, \dots, f_k$  and  $g_1, g_2, \dots, g_k$  respectively, then we compute the chi-square distances between and  $\{t_1, t_2, \dots, t_n\}, f(t; \tilde{\theta})$  and  $\{t_1, t_2, \dots, t_n\}, g(t; \tilde{\varphi})$  as

$$\begin{aligned} \chi_{f,\text{data}}^2 &= \sum_{i=1}^k \frac{(n_i - f_i)^2}{f_i} \\ \chi_{g,\text{data}}^2 &= \sum_{i=1}^k \frac{(n_i - g_i)^2}{g_i} \end{aligned} \quad (27)$$

If  $\chi_{f,\text{data}}^2 < \chi_{g,\text{data}}^2$  then we choose model  $F$ ; otherwise, we choose model  $G$ . The same approach can now be extended and used to prioritize a number of competitive models.



## References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Auto Control* 19(6):716–723
- Aldrich J (1997) RA Fisher and the making of maximum likelihood 1912–1922. *Statist Sci* 12(3): 162–176
- Anagnos T, Kiremidjian AS (1988) A review of earthquake occurrence models for seismic hazard analysis. *Probab Eng Mech* 3(1):1–11
- Bak P, Christensen K, Danon L, Scanlon T (2002) Unified scaling law for earthquakes. *Phys Rev Lett* 88(17):178501–178504
- Baker JW (2008) An introduction to probabilistic seismic hazard analysis. [http://www.stanford.edu/~bakerjw/Publications/Baker\\_\(2008\)\\_Intro\\_to\\_PSHA\\_v1\\_3.pdf](http://www.stanford.edu/~bakerjw/Publications/Baker_(2008)_Intro_to_PSHA_v1_3.pdf). Accessed on 25 July 2013
- Bilham R, England P (2001) Plateau pop-up during the 1897 earthquake. *Nature* 410:806–809
- BIS (2002) IS 1893 (part 1)–2002: Indian standard criteria for earthquake resistant design of structures, part I—general provisions and buildings. Bureau of Indian Standards, New Delhi
- Chen C, Wang JP, Wu YM, Chan CH (2013) A study of earthquake inter-occurrence distribution models in Taiwan. *Nat Hazards* 69(3):1335–1350
- Cornell CA, Winterstein S (1986) Applicability of the Poisson earthquake occurrence model. *Seismic Hazard Methodology for the Central and Eastern United States*, EPRI Research Report, p 101
- Dionysiou DD, Papadopoulos GA (1992) Poissonian and negative binomial modeling of earthquake time series in the Aegean area. *Phys Earth Planet Inter* 71:154–165
- Faenza L, Marzocchi W, Serretti P, Boschi E (2008) On the spatio-temporal distribution of  $M$  7.0 + worldwide seismicity. *Tectonophysics* 449:97–104
- Gupta RD, Kundu D (1999) Generalized exponential distributions. *Aust N Z J Stat* 41(2):173–188
- Gupta RD, Kundu D (2007) Generalized exponential distributions: existing theory and recent developments. *J Stat Plan Inference* 137(11):3537–3547
- Gupta HK, Rajendran K, Singh HN (1986) Seismicity of the northeast India region: Part I: the data base. *J Geol Soc India* 28:345–365
- Gupta RC, Gupta PL, Gupta RD (1998) Modeling failure time data by Lehmann alternatives. *Comm Statist Theory Methods* 27:887–904
- Johnson NL, Kotz S, Balakrishnan N (1995) *Continuous univariate distributions*, 2nd edn. Wiley, New York, vol 2, p 756
- Kagan YY, Jackson DD (1991) Long-term earthquake clustering. *Geophys J Int* 104:117–133
- Kagan YY, Schoenberg F (2001) Estimation of the upper cutoff parameter for the tapered Pareto distribution. *J Appl Prob* 38A:158–175
- Kayal JR (1996) Earthquake source process in Northeast India: a review. *Him Geol* 17:53–69
- Kijko A, Sellevoll MA (1981) Triple exponential distribution, a modified model for the occurrence of large earthquakes. *Bull Seism Soc Am* 71:2097–2101
- Matthews MV, Ellsworth WL, Reasenberg PA (2002) A Brownian model for recurrent earthquakes. *Bull Seism Soc Am* 92(6):2233–2250
- Molnar P, Pandey MR (1989) Rupture zones of the great earthquakes in the Himalayan region. *Proc Indian Acad Sci, Earth Planet Sci* 98(1):61–70
- Nandy DR (1986) Tectonics, seismicity and gravity of North eastern India and adjoining region. *Geol Surv India Memoir* 119:13–16
- Nishenko S, Buland R (1987) A generic recurrence interval distribution for earthquake forecasting. *Bull Seism Soc Am* 77:1382–1389
- Oldham RD (1899) Report on the great earthquake of 12 June 1897. *Memory Geol Soc India, Geol Surv India* 29:379
- Parvez IA, Ram A (1997) Probabilistic assessment of earthquake hazards in the north-east Indian peninsula and Hindukush regions. *Pure Appl Geophys* 149:731–746
- Pasari S, Dikshit O (2013) Impact of three-parameter Weibull models in probabilistic assessment of earthquake hazards. *Pure Appl Geophys*. doi:10.1007/s00024-013-0704-8
- Poddar MC (1950) The Assam earthquake of 15th August 1950. *Indian Miner* 4:167–176
- Raqab MZ, Ahsanullah M (2001) Estimation of the location and scale parameters of generalized exponential distribution based on order statistics. *J Stat Comput Simul* 69:104–124
- Reid HF (1910) The mechanics of the earthquake, the California earthquake of April 18, 1906. Report of the State Investigation Commission, Carnegie Institution of Washington, Washington, Vol. 2
- Rikitake T (1976) Recurrence of great earthquakes at subduction zones. *Tectonophysics* 35:305–362
- SSHAC (Senior Seismic Hazard Analysis Committee), Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts (1997). US Nuclear Regulatory Commission Report. CR-6372, Washington, DC, p 888

- Thingbaijam KKS, Nath SK, Yadav A, Raj A, Walling MY, Mohanty WK (2008) Recent seismicity in northeast India and its adjoining region. *J Seism* 12:107–123
- Utsu T (1984) Estimation of parameters for recurrence models of earthquakes. *Bull Earthq Res Inst Univ Tokyo* 59:53–66
- Working Group on California Earthquake Probabilities (2008) The uniform California earthquake rupture forecast, Version 2 (UCERF 2): USGS open file rep., 2007–1437 and California geological survey special report 203 (<http://pubs.usgs.gov/of/2007/1437/>)
- Yadav RBS, Bormann P, Rastogi BK, Das MC, Chopra S (2009) A homogeneous and complete earthquake catalog for northeast India and the adjoining region. *Seism Res Lett* 80(4):609–627
- Yadav RBS, Tripathi JN, Rastogi BK, Das MC, Chopra S (2010) Probabilistic assessment of earthquake recurrence in northeast India and adjoining regions. *Pure Appl Geophys* 167:1331–1342
- Yazdani A, Kowsari M (2011) Statistical prediction of the sequence of large earthquakes in Iran. *IJE Trans B Appl* 24(4):325–336