Scientific
Research
Publishing

Estimating the Probability of Earthquake Occurrence and Return Period Using Generalized Linear Models

Shrestha Noora

Department of Statistics, Tribhuvan University, Kathmandu, Nepal
Email: shresthanoora@gmail.com

Abstract

In this paper, the frequency of an earthquake occurrence and magnitude relationship has been modeled with generalized linear models for the set of earthquake data of Nepal. A goodness of fit of a statistical model is applied for generalized linear models and considering the model selection information criterion, Akaike information criterion and Bayesian information criterion, generalized Poisson regression model has been selected as a suitable model for the study. The objective of this study is to determine the parameters (a and b values), estimate the probability of an earthquake occurrence and its return period using a Poisson regression model and compared with the Gutenberg-Richter model. The study suggests that the probabilities of earthquake occurrences and return periods estimated by both the models are relatively close to each other. The return periods from the generalized Poisson regression model are comparatively smaller than the Gutenberg-Richter model.

Keywords

Earthquake, Generalized Linear Model, Gutenberg-Richter Relation, Poisson Regression, Seismic Hazard

1. Introduction

The earthquake is the supreme terrifying and harsh phenomena of nature that can do significant damages to infrastructure and cause the death of people. Nepal is one of the paramount catastrophe prone countries in the world. Nepal situated in the center of the Himalayan range, lies in between 80°4' to 88°12' east longitude and 26°22' to 30°27' north latitude (MoHA & DP Net, 2015). Less than

10% of earthquakes happen within seismic plates, but remaining 90% are commonly found in the plate periphery (Lamb & Jones, 2012). The entire region of Nepal is likely to experience devastating earthquakes as it lies between two seismically energetic Indian and Eurasian tectonic plates (MoUD, 2016).

Nepal has a long history of numerous earthquakes and has experienced great earthquakes in the past two centuries with moment magnitudes $M_w = 7$ and greater. The recorded earthquake in the history of Nepal was on 7th June 1255 AD with magnitude $M_w = 7.7$. On 16th January 1934 AD, an earthquake called Nepal Bihar Earthquake, hit Nepal and its surrounding regions with $M_w = 8.4$ magnitude. The fatality figures were the highest for any recorded earthquake in the history of Nepal (MoHA & DP Net, 2015; MoUD, 2016). The latest earthquake experienced in Nepal was on 25th April 2015 at 11:56 am local time. The earthquake of magnitude 7.8 M_w, called Gorkha Earthquake, hit at Barpark located 82 kilometers northwest of Nepal's capital of Kathmandu affecting millions of citizens (USGS, 2016). This event has been the most powerful earthquake disaster to strike Nepal since the earthquake in 1934, tracked by many aftershocks, the largest being $M_w = 7.3$ magnitude on 12th May 2015. The devastating earthquake included about 9000 fatalities, 23,000 injuries, more than 500,000 destroyed houses, and 270,000 damaged houses (Lamb & Jones, 2012; NPC, 2015). People worldwide desire to know the likelihood of earthquakes but neither physical nor statistical models are adequate for predictions and other analysis of seismic pattern (Konsuk & Aktas, 2013; Vere-Jones, Ben-Zion, & Zuniga, 2005). There is a little evidence of failure of earthquake prediction, but this does not deny the need to look forward and decrease the hazard and loss of life (Nava, Herrera, Frez, & Glowacka, 2005).

This paper anticipated to deal with the questions 1) What is the frequency-magnitude relationship of earthquake in this region? 2) Every how many years (in average) an earthquake occurs with magnitude ≥ M? 3) What is the probability of an occurrence of at least one earthquake of magnitude ≥ M in the next "t" years? The aim of the earthquake prediction is to aware people about the possible devastating earthquakes timely enough to allow suitable reaction to the calamity and reduce the loss of life and damage from the earthquake occurrence (Vere-Jones et al., 2005; Nava et al., 2005).

Some researchers believed that the most analysis of seismic hazards is sensitive to inaccuracies in the earthquake catalogue. It is assumed that the long-term earthquake catalogue is not homogeneous and the regular earthquakes, which might include foreshocks and aftershocks of characteristic events, follow Gutenberg-Richter frequency magnitude relationship (Wyss, Shimazaki, & Ito, 1999; Kagan, 1993). In seismology, the Gutenberg-Richter relation is mainly used to find the association between the frequency and magnitude of the earthquake occurrence because the distributions of earthquakes in any areas of the planet characteristically satisfy this relation (Gutenberg & Richter, 1954; Gutenberg & Richter, 1956).

The earlier research papers have applied the generalized linear models (GLM), which included Poisson regression, negative-binomial, and gamma regression models, for an earthquake hazard analysis. Konsuk and Aktas (2013) analyzed that the magnitude random variable is distributed as the exponential distribution. The relation between magnitude and frequency is characterized using the Gutenberg Richter function. Shrey and Baker (2011) fitted logistic regression model by maximum likelihood method using generalized linear model for predicting the probability of near fault earthquake ground motion pulses and their period. Turker and Bayrak (2016) estimated an earthquake occurrence probability and the return period in ten regions of Turkey using the Gutenberg Richter model and the Poisson model.

Several studies mentioned that the generalized linear model is used to include a common method for computing parameter estimates, and it also provides significant results for the estimation probabilities of earthquake occurrence and recurrence periods, which are considered as significant parameters of seismic hazard related studies (Nava et al., 2005; Shrey & Baker, 2011; Turker & Bayrak, 2016). This study is noteworthy on its own from the Statistical and Geoscience perspectives on fitting the models to the earthquake data of Nepal. The important seismic parameters (a and b values) of Gutenberg Richter (GR) relationship and generalized linear models are examined by studying the past earthquake data. It is also intended to estimate the probability of an earthquake occurrence and its return periods of occurring earthquakes in the future "t" years using GR relationship and compared with the Poisson model.

2. Materials and Methods

The earthquake data are obtained from the National Seismological Centre, Department of Mines and Geology, Kathmandu, Nepal, which covers earthquakes from 25th June 1994 through 29th April 2019. It includes epicenter, latitude, longitude, stations, reporting time, and date. The most important factors affecting the seismic hazard in this region are taken into account such as frequency, magnitude, probability of exceedance, and return period of earthquake (Sebastiano, 2012). The statistical analysis has been accomplished using IBM SPSS 23.0 for Mac OS. In this study, the magnitude values, measured in local magnitude (M_L), 4.0 or greater are used for earthquake data. The local magnitude is the logarithm of maximum trace amplitude recorded on a Wood-Anderson seismometer, located 100 km from the epicenter of the earthquake (Sucuogly & Akkar, 2014). The available data are tabulated for the frequency distribution of magnitude $4 \leq M \leq 7.6$ and the number of earthquakes for "t" years. The Kolmogorov Smirnov goodness of fit test and the Anderson Darling test is used to check the normality assumption of the data (Gerald, 2012).

The Kolmogorov Smirnov test statistics is defined by

$$D = \max\left[F(y_i) - \frac{i-1}{n}, \frac{i}{n} - F(y_i) \right]. \tag{1}$$

where, F is the theoretical cumulative distribution of the distribution being tested.

The Anderson Darling test statistics is defined by

$$A^2 = -n - S \qquad (2)$$

where, $S = \sum_{i=1}^{n} \frac{(2i-1)}{n} \left[\ln F(y_i) + \ln(1 - F(y_{n+1-i})) \right]$.

Here, F is the cumulative distribution function of the specified distribution and n is the sample size. The Durbin Watson test is used to measure the auto-correlation in residuals from regression analysis.

The Durbin Watson test statistics is calculated using

$$DW = \frac{\sum_{i=1}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}, \quad 0 \le DW \le 4 \qquad (3)$$

where, e_i are residuals from ordinary least squares regression (Gerald, 2012).

2.1. Gutenberg-Richter (GR) Relation

Gutenberg and Richter (1954) have suggested an expression for the magnitude and frequency of earthquake events larger than magnitude (M). It states that the logarithm of the frequency is linearly dependent on the magnitude of the earthquake. The relation is generally fitted to the data that are available for any region of the globe. The Gutenberg Richter relation is

$$\log_{10} N = a - bM \text{ or } N = 10^{(a-bM)}. \qquad (4)$$

where, N is a number of earthquakes having magnitude larger than M during a time period "t", $\log N$ is a logarithm of the number of earthquakes with magnitude M, "a" is a constant that measures the total number of earthquakes at the given source or measure of seismic activity, and "b" is a slope of regression line or measure of the small versus large events.

The annual frequency of exceeding the M event magnitude is computed dividing the number of events N by the "t" years,

$$N_1(M) = \frac{N(M)}{t}. \qquad (5)$$

Taking logarithm on both sides of Equation (5) we get,

$$\log N_1(M) = \log N(M) - \log t = a - bM - \log t. \qquad (6)$$

The probability of occurrence of at least one earthquake of magnitude $\ge M$ in the next "t" years is

$$P(t) = 1 - e^{-N_1(M) \times t}. \qquad (7)$$

The number of years, in an average, an earthquake occurs with magnitude $\ge M$ is given by

$$T_R = \frac{1}{N_1(M)} \qquad (8)$$

where, $N_1(M) = 10^{(a-bM-\log t)}$ (Gutenberg & Richter, 1954, 1956).

2.2. Generalized Linear Models

In the present study, generalized linear models (GLM) are applied as it basically eliminates the scaling problem compared to conventional regression models. An important characteristic of GLM is that it assumes the observations are independent. The other assumption about the error structure is that there is, a single error term in the model. The normality and constant variance properties are not a compulsion for the error component. GLM is most commonly used to model count data. The selection of measurement scale is a significant feature of model selection; for example, in this study, transformed scale, such as $\log N$ and $\ln N$ are assumed to be better for additivity of systematic effects (McCullagh & Nelder, 1989).

Let $y_1, y_2, \cdots, y_n$ be the independent response observations with mean $\mu_1, \mu_2, \cdots, \mu_n$ respectively. The random element Y has an independent normal distribution with constant variance σ^2 and $E(Y) = \mu_i$. The systematic component: covariates $x_1, x_2, \cdots, x_n$ produce a linear predictor η_i. The link between the random and systematic components is $\mu_i = \eta_i$.

The generalized linear model is made up of a linear predictor

$$\eta_i = \beta_0 + \sum \beta_i x_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{9}$$

and two functions 1) a link function that describes how the mean, $E(Y) = \mu_i$, depends on the linear predictor $g(\mu_i) = \eta_i$, $i = 1, 2, \cdots, n$ and 2) a variance function that describes how the variance, Var(Y) depends on the mean, Var(Y) = $\phi V(\mu_i)$, where the dispersion parameter ϕ is a constant (McCullagh & Nelder, 1989; Dobson & Barnett, 2008).

The dependent variable y_i is a count (number of earthquake occurrence), such that $y_i = 0, 1, 2, \cdots, n$. Hence, a rational probability model for count data is frequently the Poisson distribution.

The probability function of a Poisson distribution is given by

$$f(y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \cdots, n \tag{10}$$

where, the parameter $\mu_i > 0$. The mean and variance of Poisson distribution are equal to the parameter μ. Nevertheless, this statement may not be true and occasionally over dispersion or under dispersion conditions can be observed. When the observed variance is greater than the variance of a theoretical model, over dispersion happens. In the existence of over dispersion, the generalized negative binomial regression model (GNBR) offers an alternative to the generalized Poisson regression model (GPR). If the observed variability is significantly smaller than the predicted variance or under dispersion, Gamma models are more appropriate. GLM allows choosing the suitable model fit on the basis of dispersion parameters and model fit criteria. The procedures of model fitting are 1) model selection 2) parameter estimation and 3) prediction of future values (McCullagh & Nelder, 1989; Kokonendji, 2014).

2.2.1. Model Selection

The model selection information criteria that are based on likelihood functions and applications to the parametric model based problems are 1) Akaike information criterion (AIC): AIC procedure is generally considered to select the model that minimizes $AIC = -2LL + 2d$, where LL is the maximized log likelihood of the model given "n" observation, "d" is the dimension of a model. Since the likelihood function's value is multiplied by -2, ignoring the second component, the model with the minimum AIC is the one with the highest value of the likelihood function. 2) Bayesian information criterion or Schwarz information (BIC): It is also a widespread model selection principle. It selects the model that minimizes $BIC = -2LL + d \log n$. The best model is the one that provides the minimum AIC and BIC (Fabozzi, Focardi, Rachev, Arshanapalli, & Markus, 2014).

2.2.2. A Goodness of Fit

The goodness of fit of a statistical model is continued to explain how well it fits a set of observed values "y" by a set of fitted values $\hat{\mu}$ derived from the model. It tests the hypothesis as H_0: The model fits, and H_1: The model does not fit. It also reviews the inconsistency between observed values and the expected value because a small discrepancy may be acceptable, but not the larger one (McCullagh & Nelder, 1989). The deviance residual is considered for the generalized measure of discrepancy. The residual sum of squares is the deviance for Normal distribution and is given by $\sum (y - \hat{\mu})^2$. For Poisson regression, the deviance is G^2, which is minus twice the log likelihood ratio.

G^2 is also called likelihood ratio statistic and is defined as

$$G^2 = -2 \sum y_i \log \frac{\hat{\mu}_i}{y_i} \tag{11}$$

where, y_i is the observed values and $\hat{\mu}_i$ is the fitted value. The small value of G^2 indicates that the model fits well (Bishop, Fienberg, & Holland, 2007).

The other significant measure of discrepancy is the generalized Pearson Chi Square statistics, which is given by

$$\chi^2 = \frac{\sum (y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{12}$$

where, $V(\hat{\mu}_i)$ is the estimated variance function for the distribution concerned. The Pearson Chi square statistics for the Normal distribution is the residual sum of squares, where as for the Poisson distribution it is the Pearson Chi square statistics, and is given by

$$\chi^2 = \frac{\sum (y_i - \hat{\mu}_i)^2}{(\hat{\mu}_i)} \tag{13}$$

where, y_i is the observed value, and $\hat{\mu}_i$ is the expected value under the assumption that null hypothesis is true, i.e. the assumed model is a good one. X^2 and G^2 are both measure how closely the model fits the observed data. The null hypothesis is rejected if the values of X^2 and G^2 are large enough. After selecting the

model, the unknown parameters have to be estimated. The probability of exceedance in a time period "t", described by a Poisson distribution, is given by the relationship: $P(t) = 1 - e^{-N(M)t}$. The return period of earthquake is a statistical measurement representing the average recurrence interval over an extensive period of time and is calculated using the relation $T_R = 1/N(M)$ (Madsen & Thyregod, 2010; Raymond, Montgomery, Vining, & Robinson, 2010; Shroder & Wyss, 2014).

3. Results and Discussion

The number of occurrence of earthquakes (n) is a count data and the parametric statistics for central tendency, mean = 26 and median = 6 are calculated. It is observed that the most of the values are less than 26; hence, the average value cannot be deliberated as the true representation of the data. It can also be perceived that the data is positively skewed and lacks symmetry; and thus the normality assumption has been severely violated. Therefore, to convert the non-normal data to the normal log transformation of cumulative frequency of earthquakes $\log N$ is used.

The hypothesis for normality testing is

H_0: The data follow a specified distribution and

H_1: The data do not follow a specified distribution.

In order to check the distribution of the transformed variable, first of all Kolmogorov Smirnov test is applied.

Table 1 displays the Kolmogorov Smirnov test statistics for testing specified distribution of data. The p-value is not significant (0.147 > 0.05) and failed to accept H_1 for $\log N$, which displayed that normality, exists in the data. In addition, $\ln N$ also statistically fitted to the Poisson distribution, the p-values is not significant (0.629 > 0.05). The very severe limitation of the Kolmogorov Smirnov test is that the distribution must be fully specified, i.e. the parameters are known. If location, scale and shape parameters are estimated from the available data, the critical region of this test is no longer valid (Gerald, 2012). Therefore, the Anderson Darling test is used to observing normality of the data.

The Anderson Darling test is not available in SPSS version 23 and hence it is calculated using Anderson Darling normality test calculator for excel. The result is displayed in Table 2. The p-value = 0.09505 > 0.05 indicates normality.

The Durbin-Watson test is used to determine whether there is evidence of first order autocorrelation in the data and result presented in Table 3. The hypothesis for the Durbin Watson test is H_0: There are no first order autocorrelation and H_1: The first order correlation exists.

The small value of the D-W score (0.596 < 2) indicates a positive first order autocorrelation, which is assumed to be a common occurrence in this case. Hence, it can be concluded that the observations are linearly independent. The correlation value R = 0.995 specifies that there is a very high degree of association between the magnitude and occurrence of the earthquake. The higher value

Table 1. Kolmogorov Smirnov (K-S) test.

Variables	Model	Parameter Estimates		Goodness of Fit Test	
		Mean	SD	K-S	p-value
Log N	Normal	1.39	0.964	0.126	0.147
n	Normal	26	42.6	0.304	0.000
Ln N	Poisson	3.24	1.8	0.749	0.629

Table 2. Anderson Darling (AD) test.

No. of data points	Sample mean	Sample sigma	AD test statistic	p-value
37	1.391	0.964	0.6264	0.09505

Variable: log$N$.

Table 3. Durbin Watson (D-W) test.

Model	R	R Square	Adjusted R^2	Std. Error of the Estimate	D-W
1	0.995	0.991	0.990	0.09407	0.596

Predictors: (Constant), M. Dependent Variable: log$N$.

of coefficient of determination (R^2 = 0.991) portrayed, the magnitude of earthquake explained 99.1% of the variation in occurrence of earthquake while 0.9% were due to other variables that were not included in the model. According to the results, it is observed that log$N$ and ln$N$ can be considered as dependent variables for Gutenberg-Richter model and generalized Poisson regression model or negative binomial regression model respectively.

The model selection criterion for generalized linear models is illustrated in Table 4. The cumulative frequency of earthquake (N) is divided by the time period (t) and used as a response variable in generalized linear models to select a suitable model. It demonstrates the values of AIC, and BIC for model selection which are reasonably smaller for the GPR model than the normal and GNBR. Hence, the generalized Poisson regression model is considered as the suitable model to fit the data. The significant measures of discrepancy for the Poisson regression model is deviance residual (value/df = 0.170) and generalized Pearson Chi square statistics (value/df = 0.110). These values measure how diligently the model fits the observed data. Furthermore, the generalized Poisson regression model is detected to be the best model to fit the data because 1) it was suitable for count data of earthquake occurrences, 2) model information criterion AIC and BIC are fewer, and 3 deviance and Pearson Chi square statistics are less than one. After selecting the model, the unknown parameters are estimated.

The estimated parameters of the Gutenberg Richter relationship are demonstrated in Table 5. The GR relationship of the earthquakes that had occurred in time period t = 25 years is expressed as log$N$ = 6.532 − 0.887M, where, N is the number of earthquakes ≥ M, log$N$ is the dependent variable, M is the predictor. The model provides the important parameters of the earthquake such as

Table 4. Model selection criterion for GLM.

Model Selection Criterion	Probability Distribution		
	Normal	Poisson	Negative Binomial
Log likelihood (LL)	−119.330	−40.090	−60.504
Akaike's information criterion (AIC)	244.661	84.181	125.009
Bayesian information criterion (BIC)	249.493	87.403	128.231

Table 5. Parameter estimation for Gutenberg Richter model.

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi Square	df	p-value
Intercept	6.532	0. 0831	6.370	6.695	6182.335	1	0.000
M	−0.887	0.0141	−0.914	−0.859	3960.184	1	0.000
Scale	0.008	0.0019	0.005	0.013			

$a = 6.532$, b $= -0.887$, $a' = a - \log(b\ln10) = 6.22$, $a_1 = a - \log(t) = 5.13$, and $a_1' = a' - \log(t) = 4.82$.

Table 6 displays the estimated parameters in the generalized Poisson regression model and is given by $\ln N = 15.06 - 2.04M$, where, $\ln N$ is the response variable. The other significant parameters of the earthquake are obtained: $a = 15.06$, $b = -2.04$, $a' = 13.513$, $a_1 = 11.84$, and $a_1' = 10.29$.

The relationship between frequency and magnitude of an earthquake ≥ 4 using GR model and GPR model is shown in **Figure 1**.

The seismic risk expressed in percentage and the return period of the earthquake in years in the Gutenberg Richter model is illustrated in **Table 7**. The GR relation is $\log N(M) = 6.532 - 0.887M$. The annual frequency of exceeding the M event magnitude is $N_1(M) = N(M)/t = N(M)/25$. Taking logarithm on both sides, $\log N_1(M) = \log N(M) - \log t = \log N(M) - \log 25 = 6.532 - 0.887M - 1.398 = 5.134 - 0.887*M$. For magnitude 7.5, $\log N_1(M \geq 7.5) = 5.134 - 0.887*7.5 = -1.5185$. Now, $N_1(M \geq 7.5) = 10^{(-1.5185)} = 0.030305$. Hence, the return period for 7.5 magnitude is given by $T_R(M \geq 7.5) = 1/N_1(M) = 32.99$ years. Similarly, the return period for magnitude 6 and 7 are calculated as 1.54 and 11.88 years. The probability of occurrence of at least one earthquake of magnitude ≥ M in the next "t" years, is obtained by the relation, $P(t) = 1 - e^{\left[-N_1(M)*t\right]}$. For illustration, when $M = 7.5$ and $t = 50$ years, $P(t) = 1 - e^{(-0.030305*50)} = 78\%$, which is the probability of exceedance in 50 years.

The probability of exceedance expressed in percentage and the return period of an earthquake in years for the Poisson regression model is shown in **Table 8**. The GPR relation obtained is $\ln N = 15.06 - 2.04M$. The annual frequency of exceeding the M event magnitude for 7.5 M_L is calculated as $N_1(M) = \exp(a - bM - \ln t) = 0.031$. The probability of occurrence of at least one earthquake of magnitude ≥ 7.5 within 50 years is obtained as 79% and the return period is 31.78

Table 6. Parameter estimation for generalized Poisson regression model.

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi Square	df	p-value
Intercept	15.06	0.1263	14.812	15.307	14207.391	1	0.000
M	−2.04	0.0283	−2.095	−1.984	5198.546	1	0.000

Table 7. Probability of exceedance (%) and return period using GR model.

M	$N_1(M)$	Probability of exceedance (%) for t years						T_R (Years)
		t = 1	t = 5	t = 10	t = 20	t = 50	t = 100	
4.0	38.55	100	100	100	100	100	100	0.03
4.5	13.89	100	100	100	100	100	100	0.07
5.0	5.00	99.3	100	100	100	100	100	0.20
5.5	1.80	83.5	100	100	100	100	100	0.56
6.0	0.65	47.7	96	100	100	100	100	1.54
6.5	0.23	20.8	69	90	99	100	100	4.28
7.0	0.08	8.1	34	57	81	99	100	11.88
7.5	0.03	3.0	14	26	45	78	95	32.99
7.6	0.02	2.4	12	22	39	71	92	40.47

Table 8. Probability of exceedance (%) and return period using GPR Model.

M	$N_1(M)$	Probability of exceedance (%) for t years						T_R (Years)	Theoretical Return Period
		t = 1	t = 5	t = 10	t = 20	t = 50	t = 100		
4.0	39.69	100	100	100	100	100	100	0.0252	1
4.5	14.31	100	100	100	100	100	100	0.0699	1
5.0	5.16	99	100	100	100	100	100	0.1938	1.01
5.5	1.86	84	100	100	100	100	100	0.5373	1.18
6.0	0.67	49	97	100	100	100	100	1.4901	2.05
6.5	0.24	21	70	91	99	100	100	4.1325	4.65
7.0	0.09	8	35	58	83	99	100	11.460	11.97
7.5	0.03	3	15	27	47	79	96	31.78	32.28
7.6	0.03	3	12	23	40	72	92	38.97	39.48

years. The theoretical return period is the reciprocal of the probability that the event will be exceeded in any one year. There is a statistical statement that on an average, a 10 years event will appear once every ten years and the same process may be true for 100 year event. The theoretical values of return period in Table 8 are slightly greater than the estimated return periods.

Figure 2 demonstrates the probability of earthquake occurrence (%) for different time periods in years using GR and GPR models. In GR model, the
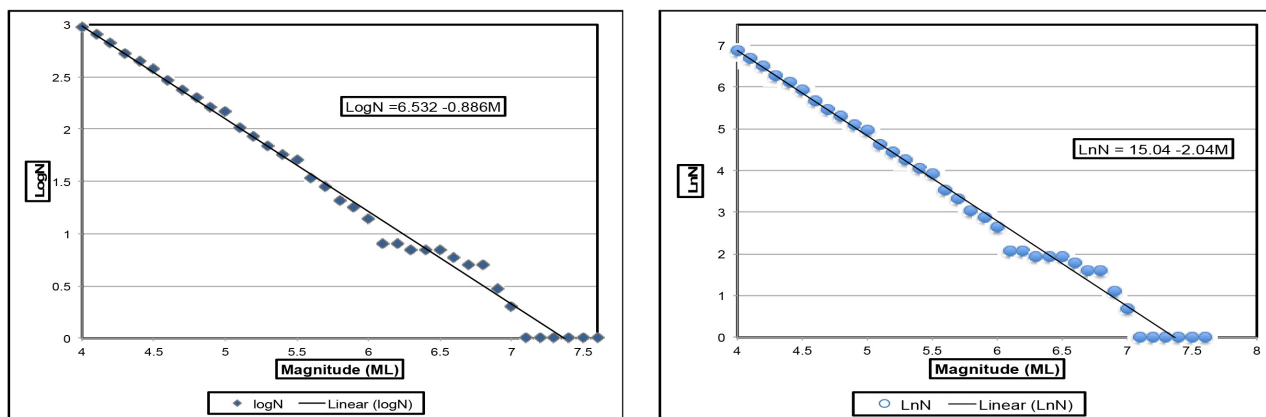
Figure 1. Magnitude (M_L)-frequency relation using GR and GPR models.
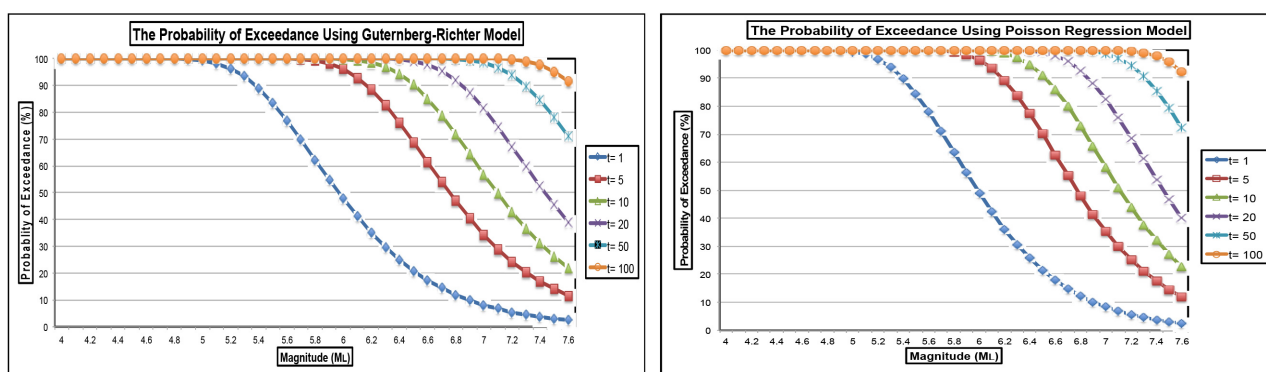

Figure 2. The probability of exceedance (%) for t years using GR and GPR models.

probability of an earthquake incident of magnitude less than 6 is almost certainly in the next 10 years and more, with the return period 1.54 years. In GPR model, the probability of the earthquake event of magnitude less than 5.5 is almost certainly in the next 5 years and more, with the return period 0.537 years (196 days).

The probability of exceedance using the GR model is found to be less than the results obtained from the GPR model for magnitude higher than 6.0. Likewise, the return periods obtained from both the models are slightly close to each other. The return period values of GPR model are comparatively less than that of the GR model. The probability of exceedance in 10 years with magnitude 7.6 for GR and GPR models is 22% and 23% and the return periods are 40.47 years and 38.99 years respectively. The estimated values depict that the probability of exceedance increases when the time period increases. It can also be noticed that the return period of the earthquake is larger for the higher magnitudes.

4. Conclusion

The frequency magnitude relationship of the earthquake data of Nepal modelled with the Gutenberg Richter (GR) model is $\log N = 6.532 - 0.887M$ and with generalized Poisson regression (GPR) model is $\ln N = 15.06 - 2.04M$. The parame-

ters a and b values for GR and GPR models are ($a = 6.532$, $b = -0.887$) and ($a =15.06$, $b = -2.04$) respectively. The return periods from GPR model are moderately smaller than that of GR model. In GR model, the return period for 7.5, 7 and 6 magnitudes are 32.99 years, 11.88 years and 1.54 years respectively. In GPR model, the return period for 7.5, 7 and 6 magnitudes are 31.78 years, 11.46 years, and 1.49 years respectively.

This study suggests that the probability of earthquake occurrence produced by both the models is close to each other. In GR model, the probability of earthquake occurrence of at least one earthquake of magnitude ≥ 7.5 in the next 10 years is 26% and the magnitude ≥ 6.5 is 90%. The probability of exceedance of magnitude 6 or lower is 100% in the next 10 years. Similarly, in GPR model, the probability of earthquake occurrence of at least one earthquake of magnitude ≥ 7.5 in the next 10 years is 27% and the magnitude ≥ 6.5 is 91%. The probability of exceedance of magnitude 6 or lower is 100% in the next 10 years.

However, it is very important to understand that the estimated probability of an earthquake occurrence and return period are statistical predicted values, calculated from a set of earthquake data of Nepal. All the parameters required to describe the seismic hazard are not considered in this study. The earthquake catalogue has 25 years of data so the predicted values of return period and the probability of exceedance in 50 years and 100 years cannot be accepted with reasonable confidence. Actually, nobody knows that when and where an earthquake with magnitude ≥ M will occur with probability 1% or more. Nevertheless, the outcome of this study will be helpful for the preparedness planning to reduce the loss of life and property that may happen due to earthquakes because Nepal lies in the high seismic region. Further research can be conducted considering other rational earthquake hazard parameters for different regions that are prone to earthquake occurrence.

Data Availability

The data studied in this paper is the earthquake data from the National Seismological Centre, Department of Mines and Geology, Kathmandu, Nepal, which covers earthquakes from 25th June 1994 through 29th April 2019. It is an open access data available on the website http://seismonepal.gov.np/earthquakes.

Conflicts of Interest

The author declares no conflicts of interest regarding publication of this paper.

References

Bishop, M. Y., Fienberg, E. S., & Holland, W. P. (2007). *Discrete Multivariate Analysis: Theory and Applications.* New York: Springer.

Dobson, A. J., & Barnett, A. G. (2008). *An Introduction to Generalized Linear Models* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC Press.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., Arshanapalli, B. G., & Markus, H. (2014). *The Basics of Financial Econometrics.* Hoboken, NJ: Wiley.

https://doi.org/10.1002/9781118856406

Gerald, K. (2012). *Statistics for Management and Economics Abbreviated* (9th ed.). Boston, MA: South Western Cengage Learning.

Gutenberg, B., & Richter, C. F. (1954). *Seismicity of the Earth and Related Phenomena.* Princeton, NJ: Princeton University Press.

Gutenberg, B., & Richter, C. F. (1956). Magnitude and Energy of Earthquakes. *Annali di Geofisica, 9,* 1-15.

Kagan, Y. Y. (1993). Statistics of Characteristic Earthquakes. *Bulletin of the Seismological Society of America, 83,* 7-24.

Kokonendji, C. C. (2014). Over-and Under Dispersion Models. In N. Balakrishnan (Ed.), *Methods and Applications of Statistics in Clinical Trials: Planning Analysis and Inferential Methods* (pp. 506-526). New York: John Wiley & Sons. https://doi.org/10.1002/9781118596333.ch30

Konsuk, H., & Aktas, S. (2013). Estimating the Recurrence Periods of Earthquake Data in Turkey. *Open Journal of Earthquake Research, 2,* 21-25. https://doi.org/10.4236/ojer.2013.21002

Lamb, M. R., & Jones, B. K. (2012). *United States Geological Survey Natural Hazards Response: Geological Survey Fact Sheet 2012-3061.* Reston, VA: US Geological Survey. https://doi.org/10.3133/fs20123061

Madsen, H., & Thyregod, P. (2010). *Introduction to General and Generalized Linear Models.* Boca Raton, FL: Chapman & Hall/CRC Press.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models, Second Edition, Monographs on Statistics and Applied Probability 37.* New York: Chapman and Hall.

MoHA and DP Net (2015). *Nepal Disaster Report 2015.* Nepal: Government of Nepal, Ministry of Home Affairs and Disaster Preparedness Network Nepal.

MoUD (2016). *Seismic Retrofitting Guidelines of Buildings in Nepal.* Nepal: Government of Nepal, Ministry of Urban Development.

Nava, F. A., Herrera, C., Frez, J., & Glowacka, E. (2005). Seismic Hazard Evaluation using Markov Chains: Application to the Japan Area. *Pure and Applied Geophysics, 162,* 1347-1366. https://doi.org/10.1007/s00024-005-2673-z

NPC (2015). *Nepal Earthquake 2015: The Post Disaster Needs Assessment.* Nepal: Government of Nepal, National Planning Commission.

Raymond, M. H., Montgomery, D. C., Vining, G. G., & Robinson, J. T. (2010). *Generalized Linear Models with Applications in Engineering and the Sciences.* Hoboken, NJ: John Wiley & Sons.

Sebastiano, D. A. (2012). *Earthquake Research and Analysis-Statistical Studies, Observations and Planning.* Croatia: InTech Web.org.

Shrey, S. K., & Baker, J. W. (2011). Regression Models for Predicting the Probability of Near-fault Earthquake Ground Motion Pulses, and Their Period. In *Proceedings of 11th International Conference on Applications of Statistics and Probability in Civil Engineering* (8 p.). Zurich, Switzerland.

Shroder, J. F., & Wyss, M. (2014). *Earthquake Hazard, Risk, and Disasters.* Amsterdam, Netherlands: Elsevier.

Sucuogly, H., & Akkar, S. (2014). *Basic Earthquake Engineering from Seismology to Analysis and Design.* New York: Springer.

Turker, T., & Bayrak, Y. (2016). A Poisson Method Application to the Assessment of the Earthquake Hazard in the North Anatolian Fault Zone. *Challenge Journal of Structural*

Mechanics, 2, 109-121. https://doi.org/10.1063/1.4945828

USGS (2016). *M 7.8 Nepal Earthquake 2015—A Small Push to Mt. Everest.* Reston, VA: United States Geological Survey. https://www.usgs.gov/natural-hazards/earthquake-hazards/science/m78-nepal-earthquake-2015-a-small-push-mt-everest?qt-science_center_objects=0#qt-science_center_objects

Vere-Jones, D., Ben-Zion, Y., & Zuniga, R. (2005). Statistical Seismology. *Pure and Applied Geophysics, 162,* 1023-1026. https://doi.org/10.1007/s00024-004-2659-2

Wyss, M., Shimazaki, K., & Ito, A. (1999). Seismicity Patterns, Their Statistical Significance, and Physical Meaning. *Pure and Applied Geophysics, 155,* 203-205. https://doi.org/10.1007/978-3-0348-8677-2_1