Geofísica Internacional (2011) 50-2: 191-197

Original paper

Prediction of the next earthquake in the Mexican subduction zone and NAFZ using the predictive distribution

Alex S. Papadopoulos

Received: July 22, 2009; accepted: January 10, 2011; published on line: March 29, 2011

Resumen

La estimación del tiempo de espera para el siguiente terremoto, en una región sísmica específica, es una de las tareas más útiles y difíciles para los científicos dedicados a estudiar sismos. En este artículo estudiamos los tiempos de espera entre terremotos occurridos en dos distintas regiones sísmicas: Michoacán en México y Notre en Anatolia cerca de Izmit, Turquía. La prueba de Kolmogorov-Smirnov muestra que los tiempos de espera entre terremotos obedecen a una distribución exponencial. Con base en este resultado se utiliza una distribución probabilística para predecir el tiempo de espera del próximo terremoto. Se utiliza la media y la mediana de la distribución para estimar el tiempo de espera para el siguiente sismo; más aún, se puede predecir intervalos de tiempo para el siguiente terremoto. Se concluye que la mediana es mejor estimador que la media.

Palabras clave: distribución de predicción probabilística, media, mediana, intervalo de predicción, prueba de Kolmogorov-Smirnov, modelo exponencial.

Abstract

The estimation of the time of the next earthquake, in a given seismic region, is one of the most useful and difficult tasks for scientists who study and predict earthquakes. In this study, we examine the previous times between earthquakes in two seismic regions, Michoacán in Mexico and the North Anatolian seismic region near Izmit, Turkey. The Kolmogorov-Smirnov test shows that the recurrence times follow the exponential distribution. Based on this finding the predictive probability distribution of the time for the upcoming earthquake is derived. The mean and median of the predictive distribution are used to estimate the time for the next earthquake; furthermore prediction intervals for the time of the upcoming earthquake are derived. It is found that the median estimator is a better estimator than the mean.

Key words: predictive probability distribution, mean, median, prediction interval, Kolmogorov-Smirnov test, exponential model.

A. S. Papadopoulos
Department of Mathematics
Yeditepe University
Istanbul, Turkey
*Corresponding author: apapadopoulos@yeditepe.edu.tr

Introduction

The study of earthquake prediction has become the focus of many scientific communities and governmental agencies all over the world, and many prediction models have been proposed. The earthquake of September 19, 1985 in Michoacán, Mexico was a great natural disaster that caused over 10,000 deaths and left a large number of inhabitants homeless. Similarly, the earthquake that took place near Izmit, Turkey on August 19, 1999 caused the death of 17,255 people; many more were injured, and countless buildings were destroyed or heavily damaged. It is well documented that earthquakes are one of the most lethal natural disasters and have affected the history of mankind. Thus the prediction of the next earthquake, on a given seismic fault, is of great importance. One of the models used for prediction is the stochastic model. Stochastic models take in consideration previous seismic activities for forecasting the next earthquake on a specific fault. The usual method of stochastic forecasting is to fit a probability distribution function to the known times between earthquakes and then use this distribution to predict the time of the next earthquake. Utsu (1972) used the double exponential distribution, while the Weibull was used by Hagiwara (1974) and Ritikake (1974). The lognormal was utilized by Nishenko and Bulland (1987) and the gamma by Utsu (1984).

The seismic activity of large earthquakes (M ≥7) in and around the Michoacán (Mexico) fault segment was analyzed by Mikumo *et al.* (1998) and by Sergio G. Ferraes (2003). Specifically, Ferráes used the Weibull, Raleigh and Pareto distributions to predict the next large earthquake in the Michoacán fault-segment and concluded that the Pareto model predicts best. The earthquakes of size 5 or greater at the North Anatolian Fault Zone (NAFZ) in Turkey have been statistically analyzed by Yilmaz, Erişoğlu and Çelik (2004). In particular, they used the Weibull, Log-normal, Log-logistic, Exponential and Gamma distributions to fit the seismic data. Based on the Kolmogorov-Smirnov test statistic, they concluded that the Weibull distribution is the most appropriate to describe the earthquake occurrence probability.

Several authors have used Bayes theorem to model earthquakes. For example, Ferráes (1985) used a Bayesian model to predict strong earthquakes in the Hellenic arc and Ferráes (1985, 1986) also used Bayes' theorem to predict the interarrival times of strong earthquakes felt in Mexico City. Furthermore, Tsapanos *et al.* (2003) used Bayesian statistics in seismic hazard modeling.

The purpose of this study is to analyze data in and around the Michoacán fault-segment and also the NAFZ seismic data. To be more specific, the previous times between earthquakes will be used to predict the time of the next earthquake. The predictive distribution for the time of the next earthquake will be derived and will be used to forecast the next seismic activity in terms of point estimators and in terms of a prediction interval. The prediction interval gives the probability that the next earthquake will be contained in a specific time interval for a given probability. The point estimators will be the mean and median of the predictive distribution. These estimators will be compared using the mean absolute deviation to decide which one estimates better the time of the next earthquake.

In this presentation, it will be shown that the time between earthquakes of magnitude M ≥7 in and around the Michoacán fault between 1911 and 1986 follow the exponential distribution and will be analyzed using the above two approaches. Furthermore, the recurrence times for the earthquakes in NAFZ of magnitude M ≥6 from 1901 to 2003, can adequately be described by an exponential model and will be analyzed similarly.

Predictive distribution for the exponential model

Let the random variable X be defined as the time elapsed between two earthquakes. It will be assumed that X follows the exponential pdf with parameter θ, i.e.

$$f(x,\ \theta) = \theta \exp(-\theta x) \quad x \geq 0,\ \theta > 0 \qquad (1)$$

Let $\underline{x} = (x_1, x_2, ..., x_n)$ be n observed independent observations from the exponential distribution and X_{n+1} be an additional observation to be taken independently of $\underline{x}$. The predictive distribution of X_{n+1}, denoted as $p(x_{n+1} \mid \underline{x})$, is the probability distribution of X_{n+1} given that we have observed the past n times, $\underline{x} = (x_1, x_2, ..., x_n)$, it is a conditional probability distribution. Our aim is to use the predictive distribution to estimate the next future observation. Lawless (1972) has shown that the distribution of $X_{n+1}/\bar{x}$, is the F distribution with 2 and 2n degrees of freedom,

$$X_{n+1}/\bar{x} \sim F(2,2n) \qquad (2)$$

where $\bar{x}$ is the sample mean of the n observations. Because the degrees of freedom of the F-distribution are 2 and 2n, the predictive probability density of X_{n+1} simplifies to

$$p(x_{n+1} \mid \underline{x}) = (1/\bar{x})\ (\sum_{i=1}^{n} x_i / (x_{n+1} + \sum_{i=1}^{n} x_i))^{n+1} \qquad (3)$$

This function is a decaying function and as n approaches infinity $p(x_{n+1} \mid \underline{x})$, approaches the exponential function. The predictive distribution takes into consideration that if we have observed n independent observations from the exponential that the next observation is not necessarily also exponential.

The expected value and variance of X_{n+1} are given as

$$E\{X_{n+1}\} = (\sum_{i=1}^{n} x_i) / (n-1) \text{ , and} \quad (4)$$

$$Var\{X_{n+1}\} = (\bar{x})^2 n^2 / \{(n-1)(n-2)\} \quad (5)$$

Thus one can use the mean of the predictive distribution to estimate the next recurrence, let

$$x_e = (\sum_{i=1}^{n} x_i) / (n-1) \quad (6)$$

denote the estimated time of the upcoming earthquake using the expected value of the predictive distribution.

Another approach to estimate the time of the next earthquake is to use the median of the predictive distribution. The median of $p(x_{n+1} \mid \underline{x})$ is derived as follows,

$$Pr = (X_{n+1} < x_m) = 0.50 \text{ or} \quad (7)$$

$$\int_0^{x_m} (1/\bar{x}) \sum_{i=1}^{n} x_i / (x_{n+1} + \sum_{i=1}^{n} x_i))^{n+1} \, dx = 0.5 \quad (8)$$

And after some calculations the value of x_m is given as

$$x_m = (\sum_{i=1}^{n} x_i) (0.50^{-1/n}-1) \quad (9)$$

In addition to obtaining point estimates for the time of the next future earthquake, one can derive an interval estimate which is based on the prediction distribution known as a prediction interval. Since $X_{n+1}/\bar{x} \sim F(2,2n)$ the a 100(1-α)% prediction interval for $X_{n+1}/\bar{x}$ is obtained from the probabilistic statement,

$$Pr\,(F_{\alpha/2,2,2n} < X_{n+1} / \bar{x} < F_{1-\alpha/2,2,2n}) = 1 - \alpha \quad (10)$$

And the lower and upper prediction limits for X_{n+1} are given as

$$c_l = \bar{x}\, F_{\alpha/2,2,2n} \text{ and,} \quad (11)$$

$$c_u = \bar{x}\, F_{1-\alpha/2,2,2n} \quad (12)$$

Because the F-distribution has 2 and 2n degrees of freedom, the values of c_l and c_u can also be expressed as follows,

$$c = (\sum_{i=1}^{n} x_i) ((1-\alpha/2)^{-1/n}-1) \quad (13)$$

$$c_u = (\sum_{i=1}^{n} x_i) ((\alpha/2)^{-1/n}-1) \quad (14)$$

It should be noted that the predictions limits computed with equations (13) and (14) do not the need the F tables.

Analysis of the Michoacan and NAFZ seismic data

In this section we will use the findings of the previous section to estimate the time of the next large earthquake in the Michoacán fault-segment and in the North Anatolian Fault Zone. The Kolmogorov-Smirnov test shows that the seismic data in both seismic regions follow the Poisson process. It is well known that times between Poisson occurrences follow the exponential distribution. But, since we are using times between earthquakes to derive the predictive distribution, details are given about.

In analyzing large earthquakes (M≥7) in the Michoacán fault-segment the seven large earthquakes studied by Ferráes (2003) are used. These earthquakes are the large earthquakes in the northern segments of the Mexican subduction zone between the Rivera and Orozeo zones and the earthquake that occurred in the Petalián region on March 4, 1979. The table below, Table 1, is reproduced from the Ferráes (2003) study.

Using n recurrence times, $(x_1, x_2, ..., x_n)$, where n = 4,5,6 the parameter θ of the exponential distribution is estimated. Let $\hat{\theta}$ denote the maximum likelihood estimate of θ. The Kolmogorov-Smirnov test was performed on the first n (n = 4,5,6) recurrence times, and in all cases at the α = 0.05 level of significance the null hypothesis that the data follow the exponential distribution can not be rejected. Table 2, gives estimates for the parameter θ, the Kolmogorov–Smirnov test statistic, and the critical value for α = 0.05 in order to show that the data are adequately described by an exponential distribution.

Next we will apply the predictive equations (6) and (9) to predict the time of the next earthquake and equations (13,14) to obtain 90% prediction limits for the last three earthquakes, namely for the years 1981, 1985 and 1986. Table 3a summarizes the findings. For each year, the observed value of the recurrence time, the

Table 1

List of large earthquakes that have occurred in the Michoacán fault-segment, including also the Petalián earthquake

Event No (n)	Year	Date (years)	Latitude (˚N)	Longitude (˚W)	Recurrence time (X_i)	Magnitude
0	1911	1911.43	17.5	102.5	---	7.7
1	1941	1941.29	18.8	102.9	29.86	7.7
2	1973	1973.08	18.4	103.2	31.79	7.5
3	1979	1979.21	17.46	101.46	6.13	7.62
4	1981	1981.82	17.8	102.3	2.61	7.3
5	1985	1985.72	18.1	102.7	3.9	8.21
6	1986	1986.33	18.4	103.0	0.61	7.0

Table 2

Estimator for θ, test statistic and critical values for the first n observations

n	$\hat{\theta}$	Kolmogorov-Smirnov test statistic	critical value (α =0.05)
4	0.0568	0.3167	0.624
5	0.0673	0.2660	0.563
6	0.0801	0.2786	0.519

estimated recurrence time using the mean and the median approaches, and prediction limits are given. Table 3b, gives prediction times and limits for the next upcoming earthquake.

In order to compare the point estimators x_e and x_m, the approach recommended by Sterling and Pollock (1986, p. 338) will be used. Namely, we will compute the mean absolute deviation between the observed and the predicted value. Since we have two estimators, we will compute two mean absolute deviations; the one with the smaller value is better. To be specific, let

$$d_e = \frac{\sum_{i=1}^{n} | x_{io} - x_{ie} |}{n}$$

be mean absolute deviation for the mean estimator, and let

$$d_m = \frac{\sum_{i=1}^{n} | x_{io} - x_{ie} |}{n}$$

denote the mean absolute deviation for the median estimator. From the results of Table 3a, we have that d_e = 14.980 and d_m = 9.172.

Table 3a

Observed, estimated recurrence times and 90% prediction limits for the Michoacán fault-segment

n	Year	Observed x_o	Predicted (mean est) x_e	Predicted (median est) x_m	Lower Prediction Limit c_l	Upper Prediction Limit c_u
4	1981	2.610	33.890	17.617	1.169	116.203
5	1985	3.900	23.463	13.318	0.908	78.467
6	1986	0.610	18.573	11.047	0.766	60.960

Table 3b

Predicted times and 90% prediction limits for the upcoming earthquake in the Michoacán fault-segment

n	Year	Observed x_o	Predicted (mean est) x_e	Predicted (median est) x_m	Lower Prediction Limit c_l	Upper Prediction Limit c_U
7	--------	-------	14.980	9.172	0.643	48.501

Table 4

List of large earthquakes that have occurred in the North Anatolian Fault Zone

Event No (n)	Data	Date (years)	Latitude (˚N)	Longitude (˚W)	Recurrence time (X_i)	Magnitude
0	04/12/1905	1905.924	39	39	-----	6.8
1	22/02/1909	1909.144	40	38	3.200	6.3
2	24/01/1916	1916.064	40.27	36.83	6.920	7.1
3	18/05/1929	1929.376	40.2	37.9	13.312	6.1
4	21/07/1938	1938.552	39.44	33.79	9.176	6.6
5	27/12/1939	1939.989	39.8	39.51	1.437	7.9
6	22/04/1940	1940.304	39.64	35.25	0.315	6.2
7	20/12/1942	1942.969	40.87	36.47	2.665	7
8	20/06/1943	1943.467	40.85	30.51	0.499	6.5
9	26/11/1943	1943.901	41.05	33.72	0.434	7.2
10	02/02/1944	1944.088	41.41	32.69	0.187	7.2
11	13/08/1951	1951.616	40.88	32.87	7.528	6.9
12	07/09/1953	1953.680	41.09	33.01	2.064	6
13	23/02/1956	1956.147	39.89	30.49	2.468	6.4
14	01/06/1957	1957.414	40.67	31	1.267	7.1
15	22/07/1967	1967.555	40.67	30.69	10.141	6.8
16	03/09/1968	1968.669	41.81	32.39	1.114	6.5
17	13/03/1992	1992.196	39.72	39.63	23.527	6.8
18	12/11/1999	1999.862	40.81	31.19	7.666	7.4
19	06/06/2000	2000.428	40.7	32.98	0.566	6
20	27/01/2003	2003.072	39.48	39.77	2.644	6

The same analysis is performed for the North Anatolian Fault Zone seismic data. Table 4 is similar to Table 1 and gives the event, year, date in years, Latitude and Longitude, recurrence time and magnitude of the earthquake.

Using the same approach as we did for the Michoacán data, the twenty recurrence times in the NAFZ are analyzed. The Kolmogorov-Smirnov test was performed on the first n, (n=10,11,…,20) recurrence, times and in all cases the null hypothesis that the data follow the exponential distribution can not be rejected. Table 5 gives estimates for the parameter θ, the Kolmogorov–Smirnov test statistic, and the critical value for α=0.05 in order to show that the data are adequately described by an exponential distribution.

Using the same approach as for the Michoacán data, the time of the next earthquake was predicted using the mean and median point estimators. Furthermore, prediction limits were obtained for the last eleven earthquakes, namely from 2/02/1944 to 27/01/2003. Table 6a summarizes the findings. For each year, the observed value of the recurrence time, the estimated recurrence times and prediction limits are given. Table 6b gives prediction times and prediction limits for the next upcoming earthquake.

As it was done for the Michoacán data we will compute the mean absolute deviations for the two point estimators from the results showing in Table 6a. We conclude that d_e = 2.606 and d_m = 0.867.

Finally, it should be mentioned that the Kolmogorov-Smirnov test showed that the seismic data in both regions follow the Poisson process. The details of these tests are not presented, because for the derivation of the predictive distribution we need to show that the times between earthquakes follow the exponential distribution.

Table 5

Estimator for θ, test statistic and critical values for the first n observations

n	$\hat{\theta}$	Kolmogorov-Smirnov test statistic	critical value (α =0.05)
10	0.2620	0.2776	0.409
11	0.2407	0.2505	0.391
12	0.2513	0.2156	0.375
13	0.2588	0.1866	0.361
14	0.2719	0.1589	0.349
15	0.2434	0.1524	0.338
16	0.2549	0.1453	0.327
17	0.1971	0.1797	0.318
18	0.1916	0.1557	0.309
19	0.2011	0.1701	0.301
20	0.2059	0.1778	0.294

Table 6a

Observed, estimated recurrence times and 90% prediction limits for the NAFZ data

n	Year	Observed x_o	Predicted (mean est) x_e	Predicted (median est) x_m	Lower Prediction Limit c_1	Upper Prediction Limit c_u
10	02/02/1944	0.187	4.747	3.040	0.217	14.999
11	13/08/1951	7.528	4.240	2.739	0.196	13.330
12	07/09/1953	2.064	4.569	2.972	0.214	14.303
13	23/02/1956	2.468	4.341	2.840	0.205	13.542
14	01/02/1957	1.267	4.185	2.751	0.199	13.016
15	22/07/1967	10.141	3.961	2.613	0.189	12.286
16	03/09/1968	1.114	4.402	2.915	0.211	13.624
17	13/03/1992	23.527	4.183	2.778	0.201	12.920
18	12/11/1999	7.666	5.392	3.590	0.261	16.625
19	06/06/2000	0.566	5.526	3.688	0.268	17.010
20	01/27/2003	2.644	5.250	3.511	0.255	16.139

Table 6b

Predicted values and 90% prediction limits for the upcoming earthquake in NAFZ

n	Year	Observed x_o	Predicted (mean est) x_e	Predicted (median est) x_m	Lower Prediction Limit c_1	Upper Prediction Limit c_u
21	--------	--------	5.113	3.426	0.249	15.698

Conclusions

From Table 3a and from the values of the mean absolute deviations d_e = 14.980 and d_m = 9.172, we conclude that the median estimator for the next occurrence of an earthquake in the Michoacán segment is better than the one obtained using the mean estimator. By saying "better", it is meant that on the average is closer to the actual value. In all three situations the predictive interval contains the observed recurrence time. The length of a confidence interval depends on the variance of the distribution and for 1981, 1985 and 1986 the variances, computed from equation (5), are 2297.1, 852.8 and 459.9 respectively and thus we have long intervals. Furthermore, the computed prediction interval of 0.64 to 48.5 years is very long. This information is not of practical use and shows that statistical methods are not enough to predict the next earthquake with some kind of reasonable accuracy.

In analyzing the North Anatolian Fault Zone seismic data, we have similar findings. From Table 6a, along with the values of the mean absolute deviations d_e = 4.891 and d_m = 3.040, we conclude again that the median estimator is better than the mean estimator. The prediction intervals contain the observed recurrence times ten times, and one time the recurrence time is outside the interval. This is consistent with a 90% prediction interval, namely the interval to contain 90% of the time, the observed value. The length of the intervals for the NAFZ data is shorter than the ones in the Michoacán segment, this is happening because the variance of the data is smaller.

The major drawback of using the exponential distribution to analyze times between earthquakes is its memoryless property, namely, the fact that the time elapsed since the last earthquake does not affect our estimations of the time to the next earthquake. This contradicts the seismic gap theory that on a certain fault the earthquake hazard is small after a large earthquake and increases with time. The predictive density used in this study, does not have the lack of memory property and it is more appealing in estimating the time of the next earthquake using either a point or an interval estimate. In using a probability density function to estimate the time of the next earthquake is a tool that the practitioner can use along with other techniques to try to estimate with some kind of accuracy the time of the next earthquake.

Acknowledgments

The author is grateful to the reviewers for their constructive comments and suggestions.

Bibliography

Ferraes S.G., 1985, The Bayesian Probabilistic Prediction of Strong Earthquakes in Hellenic Arc. *Tectophys.*, II, 3-5, 339-354.

Ferraes S.G., 1986, Bayes Theorem and the Probabilistic Prediction of Interarrival Times From Strong Earthquakes Felt in Mexico City. *J. Phys. Earth*, 34, 71-83.

Ferraes S.G., 1988, The Optimum Bayesian Probability Procedure and the Prediction of Strong Earthquakes Felt in Mexico City. *Pure Appl. Geophys*., 127 (4), 561-571.

Ferraes S.G., 2003, Probabilistic Prediction of the Next Large Earthquake in the Michoacán Fault-Segment of the Mexican Subduction Zone. *Geofísica Internacional*, 42, 1, 69-81.

Hagiwara Y., 1974, Probability of Earthquake Occurrence as Obtained from a Weibull Distribution Analysis of Crucial Strain, *Tectonophysics,* 23, 318-323.

Lawless J.F., 1972. On Prediction Intervals for Samples from the Exponential Distribution and Prediction Limits for System Survival. *Sankya, Ser* B, 34, 1-14.

Mikumo T., Miyatake T. Santoyo MA., 1998, Dynamic Rupture of Asperities and Stress Change during a Sequence of Large Interplate Eartquakes in the Mexican Subduction Zone. *Bull. Seismol. Sec. Am*. 88, 3, 668-702.

Ninshenko SP., Bulland RA., 1987, Generic Recurrence Interval Distribution for Earthquake Forecasting. *Bull. Seismol. Sec. Am*., 77, 1382-1399.

Ritikake T., 1974, Probability of an Earthquake Occurrence as Estimated from Crustal Strain. *Tectonophysics*, 23, 299-312.

Sterking T.D., Pollack S.V., 1976, Introduction to Statistical Data Processing, Prentice Hall, Inc., Englewood Cliffs, NJ.

Tsapanos T.M., Papadopoulos G.A, Galanis O.C., 2003, Time Independent Seismic Hazard Analysis of Greece Deducted from Bayesian Statistics. *Nat. Haz. Earth Sys. Sci.*, 3, 129-134.

Upadhyay S.K., Pandey M., Dec 1989, Prediction Limits for an Exponential Distribution: A Bayes Predictive Distribution Approach. *IEEE Transac. Reliab.*, 38, 5, 599-602.

Utsu T., 1972, Aftershocks and earthquake statistics (IV). Journal of the Faculty of Science, Hokkaido University Series VII Geophysic, 4, 1-42.

Utsu T., 1984, Estimation of parameters for recurrence models of earthquakes. Bulletin of the Earthquake Research Institute, University of Tokyo, 59, 53-66.

Yilmaz V., Erişoğlu M., Çelik HM., 2004. Probabilistic Prediction of the Next Earthquake in the NAZF (North Anatolian Fault Zone), Turkey. *Dogus Uni. Dergisi*, 5, 2, 243-250.