SEMI-MARKOV MODELS FOR SEISMIC HAZARD ASSESSMENT IN CERTAIN AREAS OF GREECE

Votsi I.[1], Limnios N.[2], Tsaklidis G.[1] and Papadimitriou E.[3]

[1] *Aristotle University of Thessaloniki, Department of Mathematics, 54124 Thessaloniki, Greece,*
evotsi@math.auth.gr, tsaklidi@math.auth.gr

[2] *Université de Technologie de Compiègne, Laboratoire de Mathématiques Appliquées,*
60206 Compiègne, France, nikolaos.limnios@utc.fr

[3] *Aristotle University of Thessaloniki, Department of Geology, 54124 Thessaloniki, Greece.*
ritsa@geo.auth.gr

Abstract

The long-term probabilistic seismic hazard is studied through the application of semi-Markov model. In this model a sequence of earthquakes is considered as a Markov process and the waiting time distributions depend only on the type of the last and the next event. The principal hypothesis of the model is the property of one-step memory, according to which the probability of moving to any future state depends only on the present state. The model under consideration defines a continuous-time, discrete-state stationary process in which successive state occupancies are governed by the transition probabilities of the Markov process. The space of states is considered to be finite and the process started far in the past has achieved stationarity. Firstly, a non-parametric method is applied in order to determine the waiting times. Then, the waiting times derived by means of the exponential and Weibull distributions will be compared to each other, as well as with the actual waiting times. Thus, the probability of occurrence of the anticipated earthquakes of a specific magnitude scale is calculated. The models are applied to an historical catalogue for Northern Aegean Sea.

Key words*: Semi-Markov process, semi-Markov kernel, sojourn times, earthquakes.*

1. Introduction

Several analytical models have been proposed to represent the process of earthquake occurrence. Some of them are based on empirical observations of precursory phenomena, others on physical modelling of the earthquake process, and a third class on statistical analysis of patterns of seismicity.

Renewal models have been used for a quarter century to forecast the time of the next large earthquake on a specific fault segment, where large shocks occur repeatedly at approximately regular time intervals. The most common is the Poisson model, which assumes spatial and temporal independence of all earthquakes, including the maximum earthquake associated with a fault or within a region. In this model, it is assumed that the occurrence of one earthquake does not affect the probability of an occurrence of another earthquake in the same magnitude scale at the same location in the next unit of time. The time-independence feature characterizing the

Ψηφιακή Βιβλιοθήκη Θεόφραστος - Τμήμα Γεωλογίας. Α.Π.Θ.

homogeneous Poisson model, demonstrates that earthquakes are implicitly random events; however, this feature is not appealing, because almost all of our seismological instincts favour earthquake cycle models, in which strain, and hence hazard, builds up slowly from one major earthquake to the next. In this case, the probability of a large earthquake occurrence should be small immediately after a large earthquake, and then grows with time. Thus there is a need for a time-dependent "forecasting" approach to hazard assessment. The simplest time-dependent model is the non-homogeneous Poisson model. This model is not appropriate for the estimation of seismic hazard in long time intervals, because the hazard should be updated when a new earthquake occurs.

A number of statistical models that are based on other distributions have been used including Gaussian (Rikitake, 1974), Weibull (Hagiwara, 1974; Utsu, 1984; Rikitake, 1999), gamma (Udias and Rice, 1975; Utsu, 1984), double exponential (Utsu, 1972) and lognormal (Nishenko and Bouland, 1987) distributions. All these distributions have two parameters and represent the distribution of time intervals fairly well, but none of them is more consistent with the underlying process of earthquake generation. The disadvantage of these models is their lack of seismological basis. They are inadequate to characterize the occurrence of moderate to large earthquakes on a fault at specific locations. It has been noted that the occurrence of moderate to large earthquakes on a fault is non-random in space and time (Sykes, 1971 for earthquakes in subduction zones). Other stochastic models used to represent the sequences of events and probabilities of earthquake occurrence include double Poisson (Kameda and Ozaki, 1979), Markov (Veneziano and Cornell, 1974), semi-Markov (Cluff et al., 1980) and regenerative point processes (Vere-Jones and Ozaki, 1982). A model that considers the non-random character of earthquake size and recurrence time is the semi-Markov model proposed by Cluff et al. (1980). In this model a parametric method was applied for the estimates of the recurrence intervals.

However the characterisation of future earthquake occurrence is difficult because of the lack of a sufficient number of data. The application of parametric methods provides estimators with several attractive asymptotic properties but these estimators present inconvenience when the censored time or the sample size is small. Since applications of parametric methods presuppose certain conditions with respect to the sample size, non-parametric methods will be used in order overcome this difficulty and apply semi-Markov models with reliable results.

The problem of statistical inference for semi-Markov processes is of increasing interest in recent literature. There is a growing literature concerning inference problems for continuous-time semi-Markov processes. For instance, Moore and Pyke (1968) studied empirical and maximum likelihood estimators for semi-Markov kernel; Lagakos et al. (1978) obtained the non-parametric maximum likelihood estimator for the kernel of a finite state semi-Markov process with some absorbing states; Gill (1980) constructed an estimator for the kernel of a finite state semi-Markov kernel, using counting processes; Akritas and Roussas (1979) studied the asymptotic local normality; Ouhbi and Limnios (1999, 2001) studied empirical estimators for non-linear functionals of finite semi-Markov kernels.

2. Geological setting

The Aegean (Greece) and the surrounding lands, form a region of intense crustal deformation. The kinematic of the deformation is controlled by the southwestward motion of southern Aegean relative to Europe, the active Hellenic subduction in the south, the westward motion of
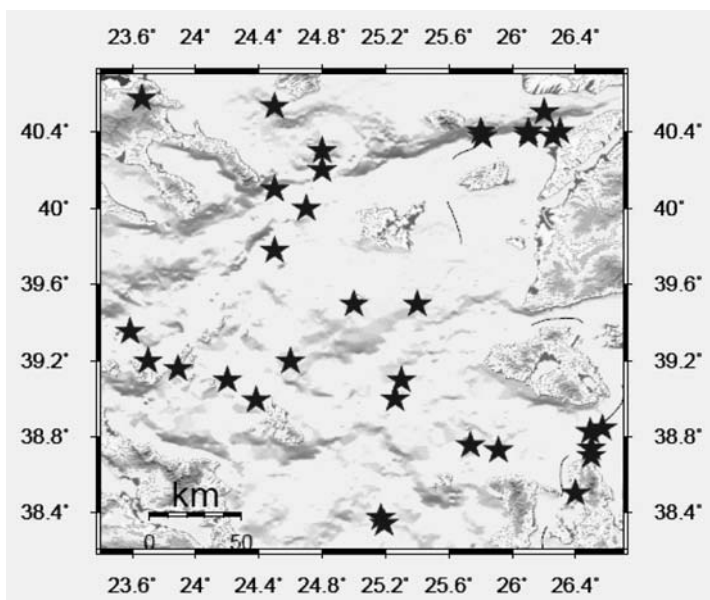
Fig.1 : Spatial distribution of Epicenters for the Earthquake Catalogue.

Anatolia in the east and the continental collision between northwestern Greece and the Apulian platform in the west. The area of northern Aegean Sea has experienced many destructive earthquakes as indicated by both instrumental data and historic information. It constitutes the northern boundary of the south Aegean plate (Papazachos et al. 1998) and is a continuation of the western part of the North Anatolian fault. Strike-slip dextral faulting dominates this region as the North Anatolian fault prolongs into the north Aegean area, where it bifurcates into two main branches of NE-SW trend. Parallel secondary faults are also recognized from seismicity and fault-plane solutions of recent strong earthquakes.

This area has frequently experienced large (M>=7.0) earthquakes, some of them occurring very close in time. The exact locations and magnitudes are of low precision before 1950, so inclusion of the events after this year is chosen. The region is bounded by latitudes 23.5 °E and 26.7ºE, longitudes 38.3°E and 40.6ºE, and the catalogue concerns earthquakes with magnitude M5.50 over the period 1950-2007. This catalogue is characterized by the desirable features of completeness, homogeneity and accuracy (Fig. 1).

Because it is implicit in the formulation that earthquakes lower the regional stress, and hence reduce the probability of immediately subsequent events, the model refers to main-sequence events only. For this reason aftershocks must be carefully identified and removed from the data before numerical fitting begins, by means of Reasenberg's (1985) declustering algorithm. The events of the catalogue are divided into three states. State 1 includes the events with magnitude $5.5 \leq M \leq 5.6$; State 2 includes events with magnitude $5.7 \leq M \leq 6$ and State 3 includes events with magnitude .

3. Methods of Estimation

3.1. Non-Parametric Method

In this section we briefly summarize the main definitions from the theory of semi-Markov

Ψηφιακή Βιβλιοθήκη Θεόφραστος - Τμήμα Γεωλογίας. Α.Π.Θ.

processes which are directly useful for our purposes. Let us consider a Markov renewal process (MRP), $(J,S) = (J_n, S_n)_{n\geq0}$ defined on a probability complete space, where $(J_n)_{n\geq0}$ is a Markov chain with values state space $E = \{1, 2, \ldots, s\}$, $(S_n)_{n\geq0}$ are the jump times which take values in R^+, $J_0, J_1, \ldots, J_n$, are the consecutive states to be visited by the MRP and $X_0 = 0, X_1, X_2, \ldots$ defined by $X_n = S_n - S_{n-1}$ for $n \geq 1$, are the sojourn times in these states. The stochastic behaviour of the MRP is determined completely by its initial law and its semi-Markov kernel defined respectively by $P(J_0 = k) = p(k)$ and $P(J_{n+1} = j, X_{n+1} \leq x / J_0, J_1, \ldots, J_n, X_1, X_2, \ldots, X_n) = Q_{ij}(x)$ on

the event $\{J_n = i\}$, for all $x \in S$ and $j \in E$. The probabilities are the transition probabilities of the Markov process $(J_n)_{n\geq0}$

Let us now consider the distribution function associated with the sojourn time in state i before

going to state j and defined as

Let $N_i(t)$ be the number of visits of $(J_n)_{n\geq0}$ to state $j \in E$ up to time t, and $N_{ij}(t)$ be the number of transitions from state i to state j up to time t, that is

and

Let us also define the distribution function and its mean value m_i, which is the

mean sojourn time of Z in state i.

3.1.1. Empirical Estimator of the Semi-Markov Kernel

The observation of a sample path of a semi-Markov process in the time interval $[0, t]$ is described as $H_t = \{J_0, J_1, \ldots, J_{N(t)}, X_1, X_2, \ldots, X_{N(t)}\}$.

In order to estimate the semi-Markov kernel of a finite state space semi-Markov process by observing one or several sample paths in the time interval under consideration, let as define the following empirical estimator of the semi-Markov kernel:
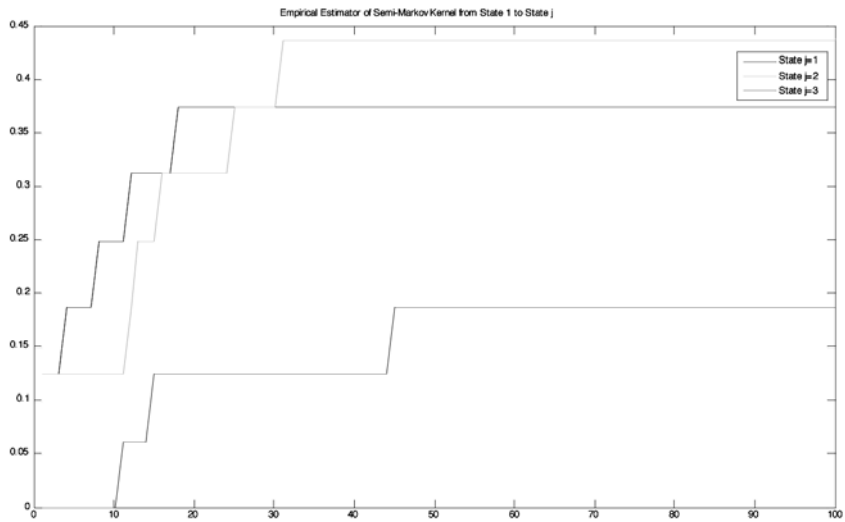
Then can be written as where

,

and

Fig. 2: Empirical Estimator for Semi-Markov Kernel from State 1 to State j, j=1, 2, 3.
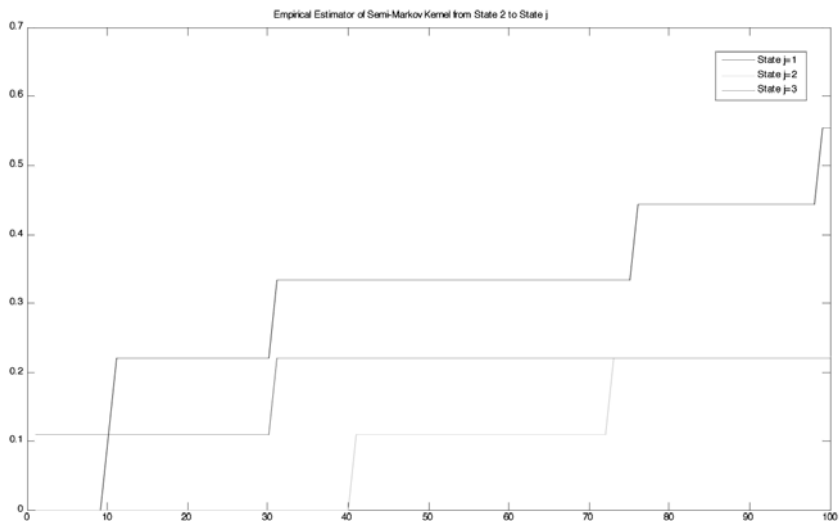

Fig. 3: Empirical Estimator for Semi-Markov Kernel from State 2 to State j, j=1, 2, 3.
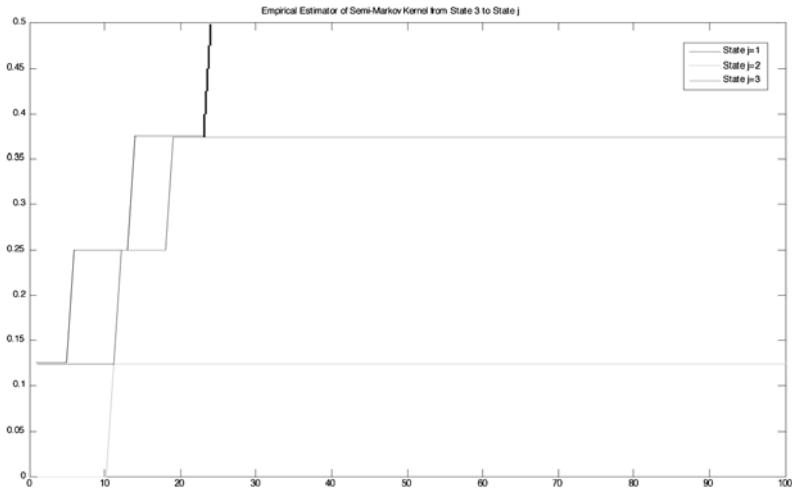
Fig. 4: Empirical Estimator for Semi-Markov Kernel from State 3 to State j, j=1, 2, 3.

Table 1.

Transitions	95% Confidence intervals for the transition probabilities	Point estimates for the transition probabilities
$1 \rightarrow 1$	$(0.3405, 0.7559)$	0.56
$1 \rightarrow 2$	$(0.2439, 0.6593)$	0.44
$2 \rightarrow 1$	$(0.3249, 0.7938)$	0.57
$2 \rightarrow 2$	$(0.2060, 0.6749)$	0.43

are the estimators of the transition probabilities and state transition functions.

In the case of two states

State 1: includes events with magnitude $5.5 \leq M \leq 5.7$

State 2: includes events with magnitude, $M \geq 5.8$

The point estimates and the 95% confidence intervals for the transition probabilities are presented in Table 1:

3.1.2. Empirical Estimator of Stationary Distribution for the Semi-Markov Process

Here we will assume that the embedded Markov chain (J_n) irreducible with invariant distribution ν_i and that $m_i < \infty$ for all $i \in E$, where m_i is the mean sojourn time in state i. The empirical estimator for the stationary distribution of the semi-Markov process is:

,

Ψηφιακή Βιβλιοθήκη Θεόφραστος - Τμήμα Γεωλογίας. Α.Π.Θ.

Table 2.

State	$\hat{v}_i(T)$	$\hat{m}_i(T)$	$\hat{\pi}_i(T)$
State 1	0.4848	12.3750	0.3084
State 2	0.2727	40.4444	0.5670
State 3	0.2425	10.0000	0.1246

where   is an estimator of the stationary distribution of the embedded Markov chain and
        is an estimator of the mean sojourn time in state *i*. The mean sojourn time in state *i* is

                        . We will use the empirical estimator of the stationary distribution of the
embedded Markov chain, defined as

In Table 2 empirical estimators of the stationary distribution for the embedded Markov chain, the
mean sojourn time and the stationary distribution for the semi-Markov process are presented.

3.2. Parametric Method

In order to use a parametric method of estimation and due to the small sample size, we define
only two different states:
State 1: includes events with magnitude $5.5 \leq M \leq 5.7$
State 2: includes events with magnitude $M \leq 5.8$.

We present below the numbers of observed transitions in the dataset, and the mean sojourn
times in between, in months (rounded):

Number of transitions: Means of sojourn times:
                                                                          .

Let us suppose that the sojourn times follow exponential distributions, namely their probabil-
ity density functions are of the form

The parameters' estimation is achieved through the maximum likelihood estimation method as follows.

Table 3.

	Parameters λ_{ij}
$1 \rightarrow 1$	30.0867
$1 \rightarrow 2$	10.6667
$2 \rightarrow 1$	14.0208
$2 \rightarrow 2$	28.5722

Ψηφιακή Βιβλιοθήκη Θεόφραστος - Τμήμα Γεωλογίας. Α.Π.Θ.

Table 4.

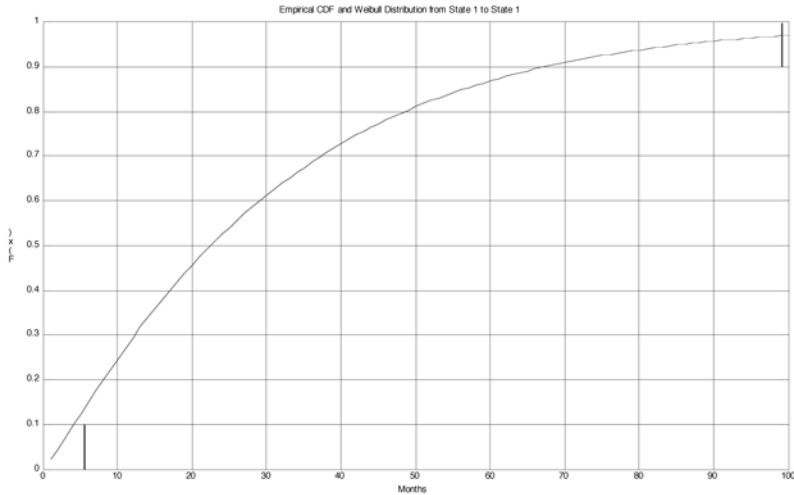| Transitions	Parameters a_{ij}	Parameters λ_{ij}
$1 \rightarrow 1$	31.2758	1.0947
$1 \rightarrow 2$	9.5418	1.0000
$2 \rightarrow 1$	12.6088	1.0000
$2 \rightarrow 2$	29.0511	1.0488

Fig. 5: Weibull Cumulative Distribution and Empirical Cumulative Distribution for Transitions from State 1 to State 1.

Supposing that the sojourn times follow Weibull distributions, their probability density functions are given by

The Weibull distribution is desirable because the associated hazard rate increases in time if $\lambda_{ij} > 1$. Actually the longer is the waiting time for transition from one state to another the higher is the probability that the transition happens. In order to estimate the parameters of the above distributions the maximum likelihood estimation method is used.

In figure 5 we present the empirical cumulative distribution and the theoretical cumulative distribution $F_{11}(t)$ of the sojourn time in state 1 before a new earthquake of the first class (i.e. state 1) occurs. Similar diagrams can be derived also for the functions $F_{12}(t)$, $F_{21}(t)$ and $F_{22}(t)$.

Assuming Weibull distributed sojourn times, the probabilities of successive earthquake occurrences in any time interval $(0, t)$, $t>0$, can be estimated using the cumulative distributions $F_{ij}(t)$. Taking into account the hazard rate functions (for the case of Weibull distribution)

we get that

for $\Delta(t)$ sufficiently small. So, the conditional probability of occurrence of an earthquake of magnitude class j, after 1 month and 6 months from the last event of magnitude class i, is given in the following tables:

Table 5.

Number of Months	Probability of next event in State 1 if last event in State 1	Probability of next event in State 1 if last event in State 2
1	0.0253	0.1076
6	0.1796	0.4378

Table 6.

Number of Months	Probability of next event in State2 if last event in State 1	Probability of next event in State 2 if last event in State 2
1	0.1318	0.0306
6	0.5546	0.2006

References

Alvarez, E. E., 2005. Estimation in Stationary Markov Renewal Processes, with Application to Earthquake Forecasting in Turkey. *Methodology and Computing in Applied Probability* 7, 119-130.

Alvarez, E. E., 2005. Smoothed nonparametric estimation in window censored semi-Markov processes. *Journal of Statistical Planning and Inference* 131, 209-229.

Anderson, A. W., Goodman, L. A., 1957. Statistical inference about Markov chains. *Ann. Math. Statist.* 28, 89-110.

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N., 1993. *Statistical Models Based on Counting Processes*. New York: Springer.

Balakrishnan, N., Limnios, N., Papadopoulos, C., 2001. Basic probabilistic models in reliability. In: Balakrishnan, N., Rao, C. R., eds. *Handbook of Statistics* 20. Amsterdam: North-Holland.

Cluff, L. S., Patwardhan, A. S. & Coppersmith, K. J., 1980. Estimating the probability of occurrences of surface faulting earthquakes on the Wasatch fault zone, Utah. *Bulletin of Seismological Society of America* 70, 5, 1463-1478.

Console, R., Rhodes, D. A., Murru, M, Evison, F. F., Papadimitriou, E. E. & Karakostas, V. G., 2006. Comparative performance of time-invariant, long-range and short-range forecasting models on the earthquake catalogue of Greece. *Journal of Geophysical Research* 111, B09304.

Garavaglia, E., Pavani, R, 2009. About Earthquake Forecasting by Markov Renewal Processes. *Methoodol. Comput. Appl. Probab.*

Greenwood, P. E., Wefelmeyer, W., 1996. Empirical Estimators for Semi-Markov Processes. *Mathematical Methods of Statistics* 5, 299-315.

Janssen, J. Limnios, N., (eds). 1999. *Semi-Markov Models and Applications*. Dordrecht: Kluwer Academic.

Kiremidjian, A. S., Anagnos, T., 1984. Stochastic slip-predictable model for earthquake occurrences. *Bulletin of Seismological Society of America* 72, 2, 739-755.

Limnios, N., Ouhbi, B., 2006. Nonparametric estimation of some important indicators in reliability for semi-Markov processes, *Statistical Methodology* 3, 341-350.

Limnios, N., Ouhbi, B., 2003. Nonparametric reliability estimation of semi-Markov processes, *Journal of Statistical Planning and Inference* 109, 155-165.

Limnios, N., Ouhbi, B. & Sadek, A., 2005. Empirical estimator of stationary distribution for semi- Markov processes. *Communications in Statistics-Theory and Methods* 34, 287-995.

Limnios, N., Opriçan, G., 2001. Semi-Markov Processes and Reliability, Birkhäuser, Boston.

Ouhbi, B., Limnios, N., 2001. Non-parametric Estimation for Semi-Markov Processes Based on its Hazard Rate Functions. *Statistical Inference for Stochastic Processes* 2, 151-173.

Limnios, N., Ouhbi, B., 2003. Empirical estimators and related functions for semi-Markov systems. In: Linqvist, B., Doksum, K., (eds). Mathematical and Statistical Methods in Reliability. Word Scientific.

Papadimitriou, E. E., Sykes, L. R., 2001. Evolution of the stress field in the northern Aegean Sea (Greece). *Geophysical Journal International* 146, 747-759.