Transition Probabilities between Seismic Regions

C. Lomnitz

(Received 1967 April 10)

Summary

Earlier work on preferred patterns of transition of earthquake activity between seismic regions has been inconclusive. The problem is restated in terms of 51 seismic regions. A series of 4164 earthquakes covering the period 1904–1954 is analyzed for randomness of transition between consecutive events. Using the multinomial distribution as a null hypothesis, contingency tables are derived for the sequence of events $M \geq 6.5$ and $M \geq 7.0$. It is found that the hypothesis of randomness is rejected for both sequences at the 0.975 significance level. This result is attributed to the presence of aftershocks.

1. Introduction

Earlier studies on the stochastic nature of the earthquake process have focused on the distribution of seismic events in time. To this end it is assumed that seismic data collected from a certain region can be represented by a one-dimensional time series (see, for instance, review by Lomnitz 1966). Studies involving the distribution of earthquakes in space are rare.

The present study is related to the work of Yamaguti (1933), who examined the time series of world earthquakes for evidence of preferred transitions between geographical regions. He found that there was a tendency for earthquakes in certain regions to be preceded by shocks in other, usually quite distant, regions. This result was in agreement with an opinion held, among others, by Omori (1907). Patterns of regional sequencing of earthquakes have since been reported on several occasions, more recently by Tamrazyan (1962) and by Båth (1966); the latter author reports a west-to-east progression of epicentres among contiguous regions in the Alpide Belt.

The work of Yamaguti was criticized by Whipple (1934), on the ground that the observations were actually consistent with the probabilities theoretically derived from binomial trials. In addition, however, Whipple also pointed out that the body of data used by Yamaguti (420 earthquakes for the period 1900–1931) could be construed as representing "consecutive" events only in a highly arbitrary manner. The magnitude parameter had not been invented at the time of publication of the papers by Yamaguti and Whipple.

While this is a valid criticism it leaves open the question originally posed by Yamaguti. For if the data had been properly selected, it is conceivable that a significant dependence in the transition probabilities might have resulted. It seemed appropriate to repeat the Yamaguti experiment with a more homogeneous and extensive body of data, using the magnitude parameter defined by Richter (1935).
2. Description of data used

The present study utilizes a list of 4164 earthquakes compiled from the work of Gutenberg & Richter (1954) for the epoch 1904–1952, plus a supplement of shocks covering the period to the end of 1954.

The regionalization used is also due to Gutenberg & Richter (1954). It consists of 51 regions defined on the basis of seismicity; the relative frequency of earthquakes among these regions is highly variable. Some of the largest regions (e.g. Central Pacific Ocean, Region 39) are nearly empty of shocks, while small regions (New Hebrides or Moluccas) are among the most active. Fig. 1 shows the boundaries of the regions as used.

3. Data analysis

The definition of 'consecutive earthquakes' can only be given in terms of earthquake magnitude. For homogeneity and completeness of the data on a worldwide basis the minimum magnitude level we can use is $M \geq 6.5$. The total number of earthquakes above this magnitude level was 2086.

Let $N$ be the total number of events classified among 51 regions, so that

$$\sum_{i=1}^{51} n_i = N,$$

where $n_i$ is the number of events in region $i$. We desire to test the hypothesis that any pair of consecutive earthquakes, $A$ and $B$, are independent of the regions in which they occur. This hypothesis may be expressed as follows:

$$\Pr\{A \text{ occurs in } i, B \text{ occurs in } j\} = \Pr\{ A \text{ occurs in } i\} \cdot \Pr\{B \text{ occurs in } j\},$$

or

$$p_{ij} = p_i p_j$$

where $p_{ij}$ is the probability of a consecutive transition from region $i$ to $j$, and $p_i, p_j$ are the probabilities of occurrence of an event in these regions.

For the multinomial distribution the maximum likelihood estimator of $p_i$ is (Brownlee 1965):

$$p_i = \frac{n_i}{N}$$

and the maximum likelihood estimators of the $p_{ij}$ become, from (3) and (4):

$$p_{ij} = \frac{n_{ij} n_j}{N^2}.$$  \hspace{1cm} (5)

If $n_{ij}$ is the number of transitions from region $i$ to $j$, the value of chi-square for a multinomial distribution of $51 \times 51$ variables is

$$\chi^2 = \sum_{i=1}^{51} \sum_{j=1}^{51} \frac{(n_{ij} - Np_{ij})^2}{Np_{ij}}$$

with $(51^2 - 1)$ initial degrees of freedom. Substituting from (5) we obtain

$$\chi^2 = \sum_{i=1}^{51} \sum_{j=1}^{51} \frac{(n_{ij} - n_i n_j/N)^2}{n_i n_j/N}$$

Since $(51 - 1)$ parameters $p_i$ have been estimated from the data, the number of degrees of freedom is actually $\nu = (51^2 - 1 - 51 - 1)$, or 2548.
FIG. 1. Seismic regions of the world (from Gutenberg & Richter 1954).
Table 1 represents a contingency table of the number of transitions observed. The figure in the $i$th row and the $j$th column is $n_{ij}$, the number of transitions from regions $i$ to region $j$. The last row gives the column totals $n_i$ (which are equal to the row totals). Finally, the value of equation (7) is

$$\chi^2 = 3005.16.$$  

This value is to be compared to the distribution of $\chi^2(2548)$ as our null hypothesis. For a number of degrees of freedom greater than 100 we may use the normal approximation. The value $\sqrt{2\chi^2}$ is approximately normally distributed with mean $\sqrt{(2v-1)}$ and unit variance. Thus we find that the 97.5 percentile of $\chi^2(2548)$ is approximately (Brownlee 1965, p. 84):

$$\chi^2_{0.025}(2548) \approx \frac{1}{2}[\sqrt{(2 \times 2548 - 1) + 1.96}]^2 = 2693$$

which is lower than the value computed from the data. Hence the null hypothesis is rejected at the 97.5% significance level.

Let us repeat the calculations for a higher threshold of magnitude, say $M \geq 7.0$. The total number of events is now only 980. Table 2 shows the contingency table for these data, in the same format as for Table 1. The computed value of $\chi^2$ is now 2348.62, which is at the lower end of the distribution. Using the same procedure as above we find

$$\chi^2_{0.025}(2548) \approx \frac{1}{2}[\sqrt{(2 \times 2548 - 1) - 1.96}]^2 = 2414$$

which is higher than the computed value. Thus the null hypothesis is again rejected at the 97.5% significance level.

4. Conclusions

By testing the spatial transition pattern of pairs of consecutive earthquakes against a multinomial model, it was found that the probability of the transitions being independent of the epicentral regions was less than 2.5% in a $\chi^2$ test. This result was obtained for world earthquakes of the 1904-1954 epoch at two different magnitude levels: $M \geq 6.5$ and $M \geq 7.0$.

We conclude that a non-random pattern of transition between epicentral regions is definitely present. The above analysis gives no clue as to the nature of this pattern. However, a tentative interpretation may be reached as follows:

If the unknown pattern consists in some geotectonic relationship between different regions, such a pattern should not become apparent by means of any simple transformation of the contingency matrix. Consider now the distribution of column totals (or row totals) of the contingency matrix. The values of these totals fluctuate strongly (between $n_{35} = 0$ to $n_{19} = 195$, in the case of Table 1). Hence the value of the trace $D$ of the contingency matrix would be expected to lie somewhere near the centre of the distribution of the $n_i$. This diagonal sum may be written

$$D = \sum_{i=1}^{51} n_{ii}. \quad (8)$$

From Table 1 we obtain $D_{6.5} = 180$; similarly, from Table 2, $D_{7.0} = 95$. Instead of falling close to the mean values of $n_i$, these diagonal sums are found to be extreme values in both cases, as may be seen by comparing them with the actual column totals.

In the case of Table 1, $D$ includes 8.64% of all observations, as compared to 9.35% for the largest column total; in the case of Table 2, $D$ totals 9.69% of all observations, which equals the largest value of $n_i$.

This indicates an abnormally high incidence of transitions of the $i-i$ type, independently of any transition pattern which may exist between different regions.
This repetition of earthquakes in a given region is a known feature of the aftershock process.

Thus, in spite of the relatively high threshold magnitude of the data, the effect of aftershock occurrence is still quite significant. It is therefore necessary to use a probability model which does not neglect the occurrence of aftershocks, if it is desired to test earthquake data for preferential (non-random) transitions between different geographical regions.

Acknowledgment

I am indebted to Helen W. Freedman and Arnoldo Hax for helpful discussion and criticism.

Seismographic Stations,
University of California,
Berkeley,
California.
1967 April.

References