Acta Universitatis Carolinae. Mathematica et Physica

Z. Schenková

Extreme value statistics of earthquakes

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 23 (1982), No. 1, 73--79

Persistent URL: http://dml.cz/dmlcz/142486

Extreme Value Statistics of Earthquakes

Z. SCHENKOVÁ
Geophysical Institute, Czechoslovak Academy of Sciences, Prague*)

Received 8 December 1981

The theory of largest values was applied to the data in the European and Balkan earthquake catalogues considering the first and the third asymptotic distributions. Gumbel's distribution of the first type holds for extreme values derived from both normal and exponential populations. The extreme value approach in spite of its limitations, is a suitable method for obtaining statistically defined information on occurrence probabilities or return periods of large earthquakes.

Первое и третье асимптотическое распределение теории экстремальных значений было применимо к статистике землетрясений европейского и балканского каталога. Распределение Гумбеля первого типа удовлетворяет эксремальным значениям если магнитуды землетрясений распределены как по нормальному так по показательному законам. Теория эксремальных значений вопреки своим ограничениям является удобным методом для получения статистически определенной информации о вероятности появления или периода повторения сильных землетрясений.

První a třetí asymptotické rozdělení největších hodnot bylo aplikováno na data z evropského a balkánského katalogu zemětřesení. Gumbelovo rozdělení prvního typu vyhovuje extrémním hodnotám odvozeným jak z normální tak exponenciální populace. Teorie extrémních hodnot se jeví navzdory svým omezením vhodnou metodou pro získání statisticky definované informace o pravděpodobnosti výskytu nebo periodě opakování velkých zemětřesení.

Although it is understood that there is nothing inherently random about the origin of earthquakes, it has not yet been possible to predict in a deterministic way timing, magnitude and location of future earthquakes. The magnitude-frequency relation $N(M)$ and the most likely maximum magnitude M_{max} for each region are the basic parameters for defining, in a probabilistic sense, the future seismic activity of each source region. The $N(M)$ distributions must be limited on both sides for physical reasons. The upper limit M_{max} or I_{max} determines the activity threshold and consequently the highest possible seismic effects. It can be estimated in several ways [1], [2], e.g. using the characteristic of the medium, size of the faults, seismotectonic analysis, strength of the material, thickness of the seismoactive layer, isostatic anomalies, the range of oscillation of the Benioff curve, empirical relation

*) 141 31 Praha 4, Boční II., čp. 1401, Czechoslovakia.

between M_{max} and the depth of focus, or by correlating observed M_{max} and seismic activity A, by extrapolating the magnitude-frequency relation and finally also by the extreme value statistics.

The theory of extreme values is founded on the following assumptions, which naturally limit the application of the results and must be considered in their interpretation:

(a) the conditions prevailing in the past must be valid also in the future,
(b) the observed largest events in a given interval are independent,
(c) the behaviour of the largest earthquakes in a given interval in the future will be similar to that of the past.

The second part of the theory, which is called the asymptotic theory of extremes, deals with asymptotic forms.

The stability postulate leads to three and only three asymptotic distributions of extremes and each assumes a specified behaviour of the absolute large values of the variable. This stability postulate may be described as follows: Let us assume that there are N samples each of the size n. From each sample the largest value is taken. Now the maximum of the N samples of size n is at the same time the maximum in a sample of Nn. Therefore the distribution of the largest values in a sample of size Nn should be the same as the distribution of the largest values in a simple of size n, except for a linear transformation. The asymptotic theory differs from the exact theory in the fact that it is still valid if a few neighbouring observations are dependent, which is quite favourable in the case of earthquakes. Asymptotic theory can be used if the initial distribution is unknown and the extreme observations are the only information available.

In the occurrence of maximum magnitude earthquakes only the first and third asymptotic distributions are usually considered [3]–[8]. The first asymptotic distribution of extremes assumes an unlimited variable from the right which is contrary to the commonly held belief on the existence of an upper magnitude limit which cannot be exceeded within a given volume of material of certain physical properties and under given stress distribution [3]–[6]. The first distribution also assumes that its upper tail falls off in an exponential manner. Therefore the first distribution holds for extreme values derived from both normal and exponential populations [6].

Under the assumption of the initial exponential distribution of magnitudes

$$H(x) = 1 - \exp(-x), \quad x \geqq 0 \tag{1}$$

the largest magnitudes x display a cumulative distribution of a double exponential

$$^{\mathrm{I}}F(x) = \exp[-\exp(-y)], \quad y = \alpha_n(x - u_n), \quad \alpha_n > 0. \tag{2}$$

where α_n is the extremal intensity function, u_n is the characteristic largest value, $^{\mathrm{I}}F(u_n) = 1/e$.

The mode of the largest reduced magnitudes $\tilde{y}_n = \log n$ and therefore the probability of the most probable largest magnitude $F(\tilde{x}_n)$ has the form

$$F(\tilde{x}_n) = \exp\left(-\frac{1}{n}\right). \tag{3}$$

The probability of the most probable smallest magnitude $F(\tilde{x}_1)$ is the solution of equation

$$n = \frac{1}{F(\tilde{x}_1)} + \frac{1}{F(\tilde{x}_1)\log F(\tilde{x}_1)} - \frac{1}{\log F(\tilde{x}_1)}. \tag{4}$$

The computation of the probability ${}^{\mathrm{I}}F(\hat{x}_j)$ of the j-th most probable largest magnitude can be simplified by introducing the following two assumptions:

$$\frac{j-1}{n} \leqq {}^{\mathrm{I}}F(\hat{x}_j) \leqq \frac{j}{n} \tag{5}$$

$${}^{\mathrm{I}}F(\hat{x}_{j+1}) - {}^{\mathrm{I}}F(\hat{x}_j) = K(n), \quad j = 1, 2, \ldots, n-1, \tag{6}$$

where $K(n)$ only depends on n. This second assumption has been derived from the observed frequencies where the differences amount to $1/n$. The probabilities $F(\tilde{x}_1)$ and $F(\tilde{x}_n)$ are determined from equations (3) and (4) and, according to equation (7), the other $(n-2)$ probabilities ${}^{\mathrm{I}}F(\hat{x}_j)$ must be evenly distributed between $F(\tilde{x}_1)$ and $F(\tilde{x}_n)$. Therefore the probability of the most probable j-th largest magnitude can be determined from the equation

$${}^{\mathrm{I}}F(\hat{x}_j) = F(\tilde{x}_1) + \frac{j-1}{n-1}\left[F(\tilde{x}_n) - F(\tilde{x}_1)\right]. \tag{7}$$

The extreme normal distribution of magnitudes can be expressed as

$${}^{\mathrm{IN}}F(x) = \left[1 + \Phi(x)\right]/2, \tag{8}$$

where $\Phi(x)$ stands for the Gaussian integral.

The probability of the most probable largest magnitude ${}^{\mathrm{N}}F(\tilde{x}_n)$ is then obtained as the solution of the equation

$$n - 1 = 2\tilde{x}_n\left[1 + \Phi(\tilde{x}_n)\right]/2\ {}^{\mathrm{IN}}f(\tilde{x}_n), \tag{9}$$

where ${}^{\mathrm{IN}}f(\tilde{x}_n)$ is the probability density function of $\tilde{x}_n$. The values of $[1 + \Phi(\tilde{x}_n)]$ are given in tables of normal probability functions for selected numerical values of the modes of the largest and smallest values of $\tilde{x}_n$ and $\tilde{x}_1$. The value of n can be determined from equation (9). The probabilities of the most probable largest and the most probable smallest observation ${}^{\mathrm{IN}}F(\tilde{x}_n)$ and ${}^{\mathrm{IN}}F(\tilde{x}_1)$, are obtained from equation (8) as a function number of observations n. With a view to the symmetry of the normal distribution ${}^{\mathrm{IN}}F(\tilde{x}_n) = 1 - {}^{\mathrm{IN}}F(\tilde{x}_1)$, the relation for computing the probability of the j-th largest magnitude in a given time interval (7) is simplified to read

$$ {}^{\mathrm{IN}}F(\hat{x}_j) = {}^{\mathrm{IN}}F(\tilde{x}_1) + [(j-1)/(n-1)]\,[1 - 2\,{}^{\mathrm{IN}}F(\tilde{x}_1)] \,. \tag{10} $$

The third asymptotic distribution of the largest values [7], [8] is defined by the formula

$$ {}^{\mathrm{III}}F(x) = \exp\left[-((\omega - x)/(\omega - u_n))^{k_n}\right], \tag{11} $$
$$ k_n > 0\,, \quad x \leqq \omega\,, \quad u_n < \omega\,, $$

where ω is the upper limit of largest values, k_n is the shape parameter, u_n is the characteristic largest value and ${}^{\mathrm{III}}F(u_n) = 1/e$ and ${}^{\mathrm{III}}F(\omega) = 1$. It means that there exists an upper threshold in the distribution of the largest values. The third asymptotic distribution is related to the first by a logarithmic transformation.

Parameters ω, u and k can be estimated by several methods [7] – by the classical method of moments or by the largest observed magnitude $x_N = M_{max.obs.}$ based on the stability postulate above mentioned or by the method of the least squares. For the estimate of the parameters by the first two methods, all values of x_i $(i = 1, \ldots, N)$ must be available. However, the classical method of moments does not guarantee that the upper limit ω is larger than the largest earthquake magnitude observed in N years.

For the estimate of parameters by the method of the least squares let us consider $x_1, x_2, \ldots, x_N$ to be the observed maximum magnitudes in a given time interval in a given region arranged in order of increasing magnitude, and $p_1, p_2, \ldots, p_N$ to be the corresponding plotting positions. The plotting position p_m of the m-th observation is defined by

$$ p_m = m/(N+1)\,, \tag{12} $$

where $m = 1, 2, \ldots, N$ is the rank of sample x_i and N is the number of observations. If the curve should pass through all of the plotted observation points, the following would hold for each p_i and x_i

$$ p_i = \exp\left[-((\omega - x_i)/(\omega - u))^k\right]. \tag{13} $$

In general, however, for each set of u, ω, and k, this curve will be at some distance from every x_i. Thus for every p_i

$$ p_i = \exp\left[-((\omega - x_i')/(\omega - u))^k\right] \tag{14} $$

will hold for a different x_i', $(i = 1, 2, \ldots, N)$. Then

$$ x_i' = \omega - (\omega - u)\, z_i^{\lambda}\,, \tag{15} $$

where $z_i = -\ln p_i$, $\lambda = 1/k$. The problem is to minimize

$$ \eta = \sum_{i=1}^{N} (x_i - x_i')^2 \tag{16} $$

for $\omega > 0$, $0 < u < \omega$, $k > 0$.

When considering the earthquake hazard in planning or in structural design the probabilistic estimate of the largest earthquake magnitudes in the next n years using the earthquake data from the past N years is a useful information. The procedure of the statistics of extremes provides the expected and the most probable largest earthquake magnitudes (mode) among other quantities.

On the assumption that the third asymptotic distribution of the largest values represents well the distribution of the largest earthquake magnitude in one year intervals, the largest magnitude in n years has a probability function $^{III}F(x_n)$ given by the relation

$$^{III}F(x_n) = {}^{III}F^n(x) = \exp\left[-n((\omega - x_n)/(\omega - u))^k\right]. \tag{17}$$

The expected largest magnitude during the next n years is

$$\bar{x}_n = \omega - (\omega - u)\,\Gamma(1 + \lambda)\,(1/n)^\lambda \tag{18}$$

and the most probable largest earthquake magnitude $\tilde{x}_n$ during the next n years is

$$\tilde{x}_n = \omega - (\omega - u)\left[(1 - \lambda)/n\right]^\lambda, \text{ where } \lambda = 1/k. \tag{19}$$

The return period of earthquakes with magnitudes equal or greater than a given threshold value is defined for the first and the third asymptotic distributions as

$$T(x) = [1 - F(x)]^{-1} \tag{20}$$

assuming that the observations are equidistant in time.

There are, however, limitations employing resulting values of the theory of extremes. Extreme value methods are unreliable for the estimation of return times greater than about one-half of the time span covered by the catalogue. In general, the extreme value statistics deals with the analysis of the extremes of a distribution and with the forecasting of further extremes. Thus the graphical or numerical treatment can supply useful information on the return periods of earthquakes or the possibility of estimating their largest magnitudes which will be exceeded with a given probability.

The theory of largest values was applied to the data in the European [9] and Balkan [10] catalogues. Examples of both asymptotic distribution of the largest values are given in Figs. 1 and 2.

The deviation of the observed points from the approximating line or curve is the measure of the fitness of the statistical model to the actual physical process.

Our analysis of the European data shown that:

(1) the return periods corresponding to the normal extreme distribution are larger than the return periods corresponding to the extreme double exponential type of distribution (for hazard calculation, however, the exponential form is still prefered),

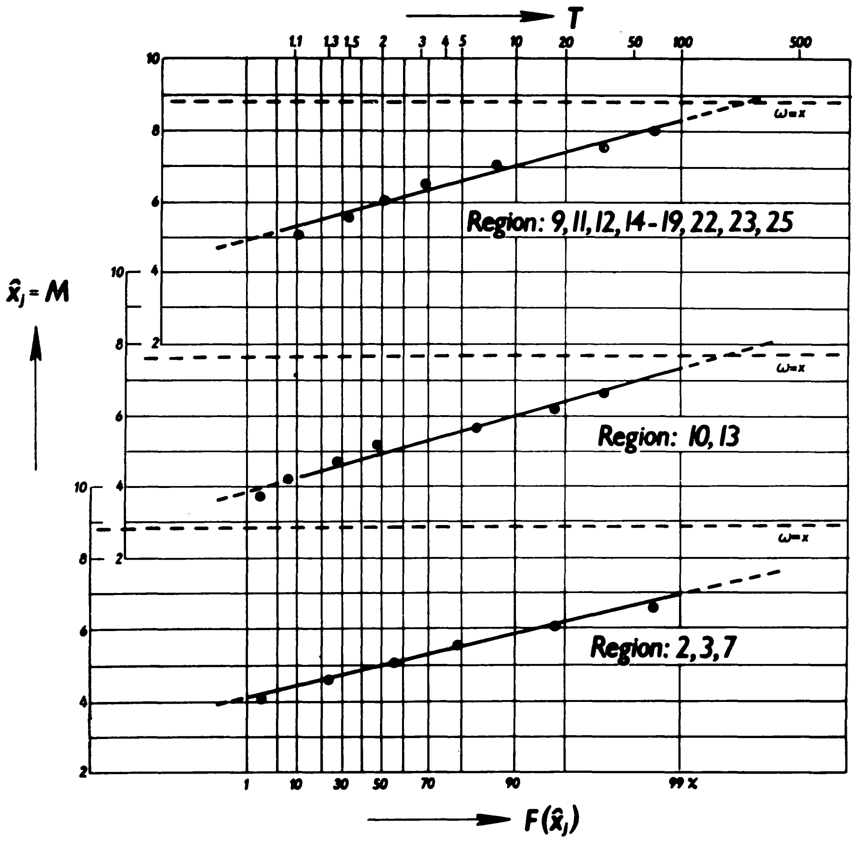

Fig. 1. The first asymptototic distribution of the largest values for three Balkan provinces (9, 11, 12, 14—19, 22, 23, 25 — Bulgaria and the Eastern part of the Aegean region; 10, 13 — Albania and West Macedonia; 2, 3, 7 — North-Western Yugoslavia).

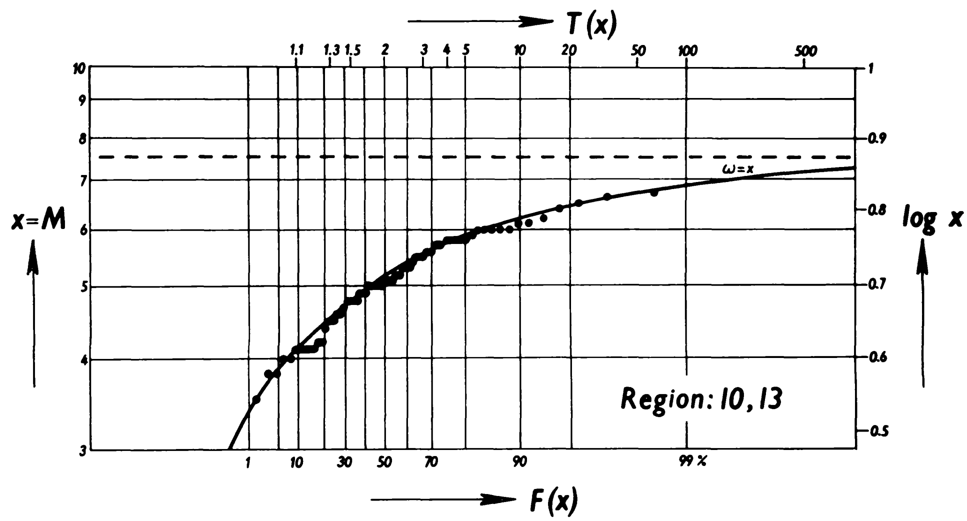

Fig. 2. The third asymptotic distribution of the largest values for Albania and West Macedonia.

(2) the third asymptotic distribution leads to larger return periods in comparison with the first distribution (i.e. to a lower hazard) and fits the observed data more closely.

References

[1] SCHENKOVÁ, Z., KÁRNÍK, V.: Sravneniye metodov opredeleniya maksimalnykh vozmozhnykh zemletryaseniy, Izv. AN SSSR, Fizika Zemli, No 11, 1974, 118.

[2] KÁRNÍK, V., SCHENKOVÁ, Z., SCHENK, V.: Seismic Risk Estimates for Central European Sites, Proc. of the 7th World Conf. Earthq. Engn., Sept. 8—13, 1980, Istanbul, Geoscience Aspects, Part I, 339.

[3] KÁRNÍK, V., SCHENKOVÁ, Z.: Application of the Theory of Largest Values to Earthquake Occurrence in the Balkan Region. Studia geoph. et geod., *18*, 1974, 134.

[4] SCHENKOVÁ, Z., SCHENK, V.: Return Periods of Earthquakes and Trends of Seismic Activity, Pure and Appl. Geophys., *113*, 1975, 683.

[5] SCHENKOVÁ, Z., KÁRNÍK, V.: Application of the Largest Value Theory to Balkan Earthquakes, Proc. of the Seminar on Seismic Zoning Maps, Skopje 27 Oct.—4 Nov. 1975, *1*, 1976, 193.

[6] SCHENKOVÁ, Z.: Influence of the Shape of the Exponential Distribution of the Largest Values on the Estimation of Earthquake Maxima, Publ. Inst. Geophys. Pol. Acad. Sc., A—5, *116*, 1977, 305..

[7] SCHENKOVÁ, Z., KÁRNÍK, V.: The Third Asymptotic Distribution of Largest Magnitudes in the Balkan Earthquake Provinces, Pure and Appl. Geophys., *116*, 1978, 1314.

[8] KÁRNÍK, V., SCHENKOVÁ, Z.: The Third Asymptotic Distribution in Earthquake Statistics, Proc. of the Symp. Analysis of Seismicity and on Seismic Risk, Liblice, 17—22 Oct., 1977; Academia, Praha 1978, 335.

[9] KÁRNÍK, V.: Seismicity of the European Area, Part I and II; Academia, Prague, 1968 and 1971.

[10] SHEBALIN, N. V., KÁRNÍK, V., HADŽIEVSKI, D. (Editors): Catalogue of Earthquakes of the Balkan Region, Part I and II, UNESCO, Skopje 1974.