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

Terms of use:
© Univerzita Karlova v Praze, 1982

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project DML-CZ: The Czech Digital Mathematics Library http://project.dml.cz
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.

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 \( M_{\text{max}} \) and the depth of focus, or by correlating observed \( M_{\text{max}} \) 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 sample 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

\[
H(x) = 1 - \exp (-x), \quad x \geq 0
\]

the largest magnitudes \( x \) display a cumulative distribution of a double exponential

\[
1F(x) = \exp \left[ -\exp (-y) \right], \quad y = \alpha_n(x - u_n), \quad \alpha_n > 0.
\]

where \( \alpha_n \) is the extremal intensity function, \( u_n \) is the characteristic largest value, \( 1F(u_n) = 1/e \).
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 \( F(\tilde{x}_i) \) 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 \( \mathcal{F}(\tilde{x}_j) \) of the \( j \)-th most probable largest magnitude can be simplified by introducing the following two assumptions:
\[
\frac{j - 1}{n} \leq \mathcal{F}(\tilde{x}_j) \leq \frac{j}{n} \tag{5}
\]
\[
\mathcal{F}(\tilde{x}_{j+1}) - \mathcal{F}(\tilde{x}_j) = K(n), \quad j = 1, 2, \ldots, n - 1, \tag{6}
\]
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 \( \mathcal{F}(\tilde{x}_j) \) 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
\[
\mathcal{F}(\tilde{x}_j) = F(\tilde{x}_1) + \frac{j - 1}{n - 1} [F(\tilde{x}_n) - F(\tilde{x}_1)]. \tag{7}
\]

The extreme normal distribution of magnitudes can be expressed as
\[
\mathcal{F}(x) = \left[1 + \Phi(x)\right]/2, \tag{8}
\]
where \( \Phi(x) \) stands for the Gaussian integral.

The probability of the most probable largest magnitude \( \mathcal{F}(\tilde{x}_n) \) is then obtained as the solution of the equation
\[
n - 1 = 2\tilde{x}_n\left[1 + \Phi(\tilde{x}_n)\right]/2 \mathcal{F}(\tilde{x}_n), \tag{9}
\]
where \( \mathcal{F}(\tilde{x}_n) \) is the probability density function of \( \tilde{x}_n \). The values of \( \left[1 + \Phi(\tilde{x}_n)\right] \) 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 \( \mathcal{F}(\tilde{x}_n) \) and \( \mathcal{F}(\tilde{x}_1) \), are obtained from equation (8) as a function number of observations \( n \). With a view to the symmetry of the normal
distribution \( \text{INF}(\hat{x}_n) = 1 - \text{INF}(\hat{x}_1) \), the relation for computing the probability of the \( j \)-th largest magnitude in a given time interval (7) is simplified to read

\[
\text{INF}(\hat{x}_j) = \text{INF}(\hat{x}_1) + \left[ (j - 1)/(n - 1) \right] \left[ 1 - 2^{\text{INF}(\hat{x}_1)} \right].
\] (10)

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

\[
\text{III}(x) = \exp \left[ -\left( (\omega - x)/(\omega - u_n) \right)^{k_n} \right],
\] \( k_n > 0 \), \( x \leq \omega \), \( u_n < \omega \),

where \( \omega \) is the upper limit of largest values, \( k_n \) is the shape parameter, \( u_n \) is the characteristic largest value and \( \text{III}(u_n) = 1/e \) and \( \text{III}(\omega) = 1 \). 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 \( \omega \), \( u \) and \( k \) can be estimated by several methods \[7\] — by the classical method of moments or by the largest observed magnitude \( x_N = M_{\text{max.abs}} \) 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 \( \omega \) 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

\[
p_m = m/(N + 1),
\] \( m = 1, 2, \ldots, N \) 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 \( x_i \)

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

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

\[
p_i = \exp \left[ -\left( (\omega - x'_i)/(\omega - u) \right)^k \right]
\] (14)

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

\[
x'_i = \omega - (\omega - u) z'_i,
\] (15)

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

\[
\eta = \sum_{i=1}^{N} (x_i - x'_i)^2
\] (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 \( \text{III}F(x_n) \) given by the relation

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

The expected largest magnitude during the next \( n \) years is

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

and the most probable largest earthquake magnitude \( \bar{x}_n \) during the next \( n \) years is

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

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

\[
T(x) = \left[ 1 - F(x) \right]^{-1}
\]

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 preferred),
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