Extreme value statistics and thermodynamics of earthquakes: large earthquakes

Bernard H. Lavenda ([1])([2]) **and Elvio Cipollone** ([2])
([1]) *Università di Camerino (MC), Italy*
([2]) *National Centre for Research on Thermodynamics, ENEA-CR Casaccia, S. Maria di Galeria, Roma, Italy*

Abstract
A compound Poisson process is used to derive a new shape parameter which can be used to discriminate between large earthquakes and aftershock sequences. Sample exceedance distributions of large earthquakes are fitted to the Pareto tail and the actual distribution of the maximum to the Fréchet distribution, while the sample distribution of aftershocks are fitted to a Beta distribution and the distribution of the minimum to the Weibull distribution for the smallest value. The transition between initial sample distributions and asymptotic extreme value distributions shows that self-similar power laws are transformed into nonscaling exponential distributions so that neither self-similarity nor the Gutenberg-Richter law can be considered universal. The energy-magnitude transformation converts the Fréchet distribution into the Gumbel distribution, originally proposed by Epstein and Lomnitz, and not the Gompertz distribution as in the Lomnitz-Adler and Lomnitz generalization of the Gutenberg-Richter law. Numerical comparison is made with the Lomnitz-Adler and Lomnitz analysis using the same Catalogue of Chinese Earthquakes. An analogy is drawn between large earthquakes and high energy particle physics. A generalized equation of state is used to transform the Gamma density into the order-statistic Fréchet distribution. Earthquake temperature and volume are determined as functions of the energy. Large insurance claims based on the Pareto distribution, which does not have a right endpoint, show why there cannot be a maximum earthquake energy.

Key words *Gutenberg-Richter and Pareto laws – Fréchet and Gumbel distributions for energy and magnitude – compound Poisson processes – earthquake temperature and volume – upper order-statistics – maximum earthquake energy*

1. Introduction

The realization that the Gutenberg-Richter law belongs to the class of Pareto distribution functions (dfs) has facilitated the introduction of scaling concepts into earthquake prediction. Criticism has been lodged against the Gutenberg-Richter law insofar as it overestimates the likelihood of high-energy events (Lomnitz, 1974; Esteva, 1975). It was conjectured that the energy scale ceases to be valid at very large values and, with it, the validity of the Gutenberg-Richter law (Bolt, 1970; Otsuka, 1973; Chinnery and North, 1975). An upper cut-off on the Richter magnitude scale for each region could possibly be used to salvage the Gutenberg-Richter law on the high magnitude side (Cornell, 1968; Yegulalp and Kuo, 1974; Smith, 1976; Caputo, 1978). Yet the Gutenberg-Richter law is a statistical law, and one that is not based on any particular geographical area, or physical model. Consequently, if a «maximum magnitude» exists, it must be determined by statistical considerations – which reveal that there is no upper endpoint for either the Pareto or the Extreme Value (EV)

Mailing address: Prof. Bernard H. Lavenda, Università degli Studi di Camerino, Casella Postale 53, 62032 Camerino (MC), Italy; e-mail: lavenda@camserv.unicam.it

df – and not from considerations based on a specific geographical region, or physical model.

Such a physical model has been proposed by Lomnitz-Adler and Lomnitz (1978, 1979), based on strain accumulation and release at plate boundaries. The model assumes that large earthquakes form a Poisson process whose interoccurrence times are exponentially distributed. These times were implicitly equated with the accumulation time, which is proportional to the accumulation of potential slip across the boundary. Since the latter is proportional to the seismic moment, and the seismic moment is exponentially related to the magnitude, the model led to a double exponential df for the smallest value, known as the Gompertz df. The logarithmic transformation from magnitude to energy brings the negative shape parameter Weibull df for the smallest value into evidence (Kijko, 1982). Both dfs have right endpoints meaning that there are upper bounds on both the magnitude and energy for large earthquakes. Yet, no such «maximum magnitude» appears to exist ([1]).

The Gompertz df was presented as a generalization of the Gutenberg-Richter law, so that in a well defined limit it should reduce to the latter. An exponentially increasing function of the magnitude replaces the magnitude itself in the generalized magnitude-frequency expression. Only for small magnitudes would the two expressions coincide, since the Maclaurin series expansion of the exponent would, in general, diverge (Lomnitz-Adler and Lomnitz, 1978, 1979). However, it is the high magnitude region where the generalized Gutenberg-Richter law should be superior to the ordinary one, since the latter tends to overestimate the likelihood of high-energy events (Lomnitz-Adler and Lomnitz, 1978, 1979). Moreover, the energy-magnitude relation used to derive the Gompertz df was not the same as that proposed by Gutenberg and Richter (1949). The value of the slope parameter a (*cf.* eq. (4.6) below) was chosen to fit the data, which turned attention away from the strain energy and put into focus the potential seismic moment on the particular segment of the plate boundary. This incompatibility was subsequently re-addressed by Jones *et al.* (1982), who generalized the relation between the seismic moment and the accumulation time to justify the surprisingly small value of a which is necessary to fit the empirical data. The Gutenberg-Richter law is very sensitive to the logarithmic transformation between energy and magnitude. Large magnitudes imply extremely high energies; at such magnitudes, the Gutenberg-Richter relation «diverges for any magnitude-energy relation» (Lomnitz-Adler and Lomnitz, 1978, 1979).

The negative shape parameter Weibull df for the smallest value of the energy, or the Gompertz df for the smallest value of the magnitude, is an asymptotic df, and, as such, it is derivable from an initial sample df which belongs to its domain of attraction. There has been much ado about the putative universality of self-similar, homogeneous power laws, especially the Gutenberg-Richter law (Bak, 1996). No physical model is necessary to transform the homogeneous power law Pareto df into the exponential Fréchet df, which is one of three EV dfs. A nonscaling Fréchet df is close to a scaling, self-similar Pareto law in the upper tail. If the sample size is large, then the Fréchet df can be accurately fitted to the actual sample df of the maximum. In the limit, it would appear that large earthquakes are not free of natural scales. The Fréchet df is, however, a distribution for the largest – and not the smallest – value. Thus, there is no connection between the Gutenberg-Richter law and the Gompertz df. The Gompertz df, used in the statistical analysis of extreme life spans, has a right endpoint denoting maximum life expectancy or, equivalently, «maximum magnitude». If such a relation did exist (Kijko, 1982; Lomnitz-Adler and Lomnitz, 1982), it would defeat the purpose of generalizing the Gutenberg-Richter law, since it would introduce an upper bound on the magnitude, which could be considered as an upper limit to the validity of the Gutenberg-Richter relation itself.

([1]) Although Lomnitz (1994) acknowledges that the cumulative df of the magnitude is the Gumbel, and not the Gompertz, df, it does not follow that the Gumbel df «can be derived from first principles in various ways» from the physical model proposed by Lomnitz-Adler and Lomnitz (1978). The problem is that the energy is an exponential increasing function of the magnitude, and not a decreasing one (*cf.* Section 4).

Yet there appears to be nothing wrong with the fit of the Gompertz law to the empirical data – so long as one does not question the magnitude-energy relation used for the fitting. But, if the Gompertz law holds for the magnitude, then it is certainly not a generalization of the Pareto law for the energy. The solution to this paradox resides in the magnitude-energy relation itself. The particular transformation used by Lomnitz-Adler and Lomnitz (1978,1979) happens to set the same scale for both energy and magnitude. Consequently, large magnitudes were not transformed into extremely high energies. However, an exponential stretching of the scale will give a sample mean excess function which is an increasing function of the energy, and with a shape parameter greater than one. In addition, the sample hazard function decreases with increasing energy, which represents the «age» at which the system is known to have survived. The larger the magnitude, the less likely it is for an earthquake to occur at that magnitude. The behavior of these two functions implicates an EV df for the largest value, since contrary behavior would suggest an EV df for the smallest value, as the analysis of extreme life spans has shown. The latter maintains that the older the individual gets, the more likely he is to die.

One of the aims of this paper is to show that if the original Gutenberg-Richter energy-magnitude relation is used, the Gompertz law will no longer fit the empirical data. Rather, it will be found that the Gumbel df, which was originally proposed by Epstein and Lomnitz (1966), is the one that fits the empirical data for magnitude-frequency. The Pareto law for energy stands in the same relation to the Fréchet df as the Gutenberg-Richter law for magnitude is to the Gumbel df.

In order to make the paper as self-contained as possible, we briefly review EV theory and the statistical analysis of EV in Section 2. In Section 3, we treat the statistics of a compound Poisson process, where the number of jumps in a given time interval is Poisson regulated while the intensity of the jumps is governed by a Pareto law. This has the advantage of modifying the tectonic parameter in the Gutenberg-Richter or, equivalently, the shape parameter in the Pareto tail df. Positive values of the shape parameter correspond to the Pareto df, while negative values convert it into a Beta df. The Beta df has been employed by several authors to treat aftershock sequences in which the energy of any aftershock is bounded from above by the energy of the largest shock in the sequence (Utsu, 1960; Vere-Jones, 1966). This should provide a unified treatment of main shocks and aftershocks, with the added feature that whereas the sample exceedance df of main shocks fit to the Pareto df, or a Gutenberg-Richter law, the aftershock sample df for values of the energy not exceeding that of the primary shock should fit to a Beta df. This would imply that the logarithm of the number of aftershocks whose energy is less than, or equal to, the energy of the primary shock is a linear increasing function of the magnitude of the primary shock (*cf.* eq. (3.7) below and Lavenda (in preparation)). In Section 4, we draw an analogy between large earthquakes and high-energy particle physics in which a generalization of the thermal equation of state is used to transform the Gamma density into the density of order-statistics for the $\nu \geq 1$ largest energy. The Gamma density is intimately related to the central limit theorem; it approaches the normal df as the number of degrees-of-freedom increases without limit. The df of νth largest energy becomes the Gumbel df upon setting $\nu = 1$, thereby confirming the choice of the Gumbel df as the EV df for the largest value. An earthquake temperature is defined which yields a thermal equation of state that is reminiscent of a *constrained* thermodynamic system. The nonequilibrium constraint, or the difference between the earthquake temperature and the thermodynamic temperature, is responsible for the energy dependence of the earthquake volume. Numerical comparison is made in Section 5 with the results of Lomnitz-Adler and Lomnitz (1978, 1979) using the same Catalogue of Chinese earthquakes, 780 B.C. to 1973 A.D. In Section 6, an analogy is drawn between the maximum energy of an earthquake and the largest insurance claim in order to show that such an energy, or insurance claim, does not exist. The derivation of the compound Poisson process is given in the Appendix together with the derivation of the rate function from the large deviation principle.

2. A brief account of extreme values and their statistical analysis

In addition to exceedances beyond a given threshold, maxima can be extracted from blocks of data, each block being delineated by a given period of time or region of space ([2]). If the random variables, $X_1, ..., X_n$, are independent and identically distributed (iid) according to a common df F, then ([3])

$$P(\max_{i \leq n} X_i \leq x) =$$

$$= P(X_1 \leq x, \ldots, X_n \leq x) = F^n(x).$$

In a certain sense, $F^n(x)$ is close to a max-stable df $G(x)$. More specifically, if there exist sequences of constants a_n and b_n such that $F^n(a_n x + b_n) = F(x)$, then F belongs to the domain of attraction of max-stable df, $G(x)$. Nature has simplified matters considerably by limiting the number of EV classes to three: the Gumbel df, $G_1(x) = \exp(-\exp(-x))$, which has support on $(-\infty, \infty)$, the Fréchet df, $G_2(x) = \exp(-x^{-\gamma})$, with support on $(0, \infty)$ and a shape parameter $0 < \gamma < \infty$, and the Weibull df with a negative shape parameter, $G_3(x) = \exp(-(-x)^{-\gamma})$, with support on $(-\infty, 0)$.

The initial dfs F which belong respectively to the domain of attraction of the Gumbel, Fréchet, and the negative shape parameter Weibull dfs are the exponential df, $F_1(x) = 1 - \exp(-x)$, for $x \geq 0$, the Pareto df, $F_2(x) = 1 - x^{-\gamma}$, for $x \geq 1$ and a positive shape parameter, and the Beta df, $F_3(x) = 1 - (-x)^{-\gamma}$, for $-1 \leq x \leq 0$ and a negative shape parameter. The exponential and Pareto dfs are decreasing on their support. Beta densities with shape parameters $\gamma < -1$ also share this property, while those with $\gamma > -1$ are increasing functions. Beta densities of a positive variate and a negative shape parameter belong to the domain of attraction of the Weibull df for the smallest value.

The limiting dfs of minima are in a one-to-one correspondence with the EV dfs. They are therefore referred to as «converse» EV dfs. Instead of specifying the initial dfs, we specify their tails, or survivor functions, $\overline{F} = 1 - F$, so that if there exists a sequence of constants c_n and d_n then

$$P(\min_{i \leq n} X_i > c_n x + d_n) = \overline{F}^n(c_n x + d_n) = \overline{F}(x)$$

implies min-stability. The converse Gumbel df is the Gompertz df, $\widetilde{G}_1(x) = 1 - \exp(-e^x)$, with the same support as the Gumbel df, while the converse Weibull df, having a negative shape parameter, is $\widetilde{G}_3(x) = 1 - \exp(-x^{-\gamma})$, for a positive variate. In the older literature, the converse Weibull df is referred to as the Weibull df, or the «third asymptote» (Gumbel, 1958), since it was the Swedish engineer Weibull who used such a df to characterize statistically the principle that the «weakest link breaks the chain».

In addition to the location and scale parameters, μ and $\sigma > 0$, the Pareto and Beta dfs have a shape parameter γ. The Pareto tail df, with a positive variate greater than, or equal to, one, and Beta df, with a positive variate less than, or equal to, one, can be transformed into one another by changing the sign of the shape parameter; they fit closely with the upper tails of the Fréchet and the converse Weibull dfs, respectively. In order to distinguish between the two tail dfs, we note the following: the Pareto df has a left endpoint, while the Beta df has a right endpoint, representing «maximum age». The left endpoint of the Pareto df is the scale parameter, σ, corresponding to the threshold value. Although the Fréchet and Pareto densities possess similar upper tails, only the former is tied down to zero at the left endpoint. Henceforth we shall limit ourselves to positive variates only.

The mean excess, or the mean residual life, function of the Pareto df is a linearly increasing function, while it is a linearly decreasing function in the case of the Beta df. The hazard functions, or the failure rates, show contrary behavior: the hazard functions of the Pareto and Beta dfs decrease and increase, respectively.

([2]) See, for instance, Reiss and Thomas (1997) for the classification scheme, the statistical tools, and the references given therein.

([3]) If the X_i represent the energies of large earthquakes, then the iid property means that they are to be taken from declusterized catalogues. As Lomnitz (1994) points out, large earthquakes on the same fault are likely to be widely spaced in time, and large earthquakes occurring consecutively near to one another should be very rare.

Roughly speaking, the hazard function of the Pareto df implies that the longer the system is in operation the less likely it is to fail, while hazard function of the Beta df asserts that the older the system gets, the more prone it is to failure. The distinction is analogous to that of an «open», as opposed to a «closed», thermodynamic system; in the former there is no restriction on the energy whereas in the latter there is only a finite amount of energy available. We will now see how these tail dfs have been, and are, used in earthquake prediction.

3. Stochastic processes underlying the Gutenberg-Richter law

It is well-known that the Gutenberg-Richter relation can be converted into a scaling relation on a par with Pareto's law. If $N(X > x)$ is the number of exceedances of the scaled energy $x = E/E_0$, where E_0 is some conventionally fixed lower value, then both the Gutenberg-Richter and the Pareto laws can be expressed as (Lomnitz, 1974, 1994)

$$\ln N(X > n) = \ln n - \gamma \ln x \qquad x \geq 1 \qquad (3.1)$$

where n is the number of earthquakes greater than or equal to E_0. Normalizing the number of exceedances we get the probability of exceedance $N(X > x)/n = P(X > x)$. This allows the tectonic parameter in the Gutenberg-Richter law to be written as a fractal dimension

$$\gamma := \frac{-\ln P(X > x)}{\ln x} \qquad x \geq 1\,. \qquad (3.2)$$

If the random variables X_i are iid, the initial Pareto df belongs to the domain of attraction of a stable law for the largest value as either the sample size n, or the time interval τ, increases without limit, since they are related by $n = \kappa\tau$, where κ is a constant intensity. The larger the sample, the more likely it is to include high-energy events. The Gutenberg-Richter law tends to overestimate the likelihood of these high-energy events. This is why the maxima should fit better to an EV df.

The distinction between main shocks and aftershock, or foreshock, sequences can be made on the basis of the initial dfs which belong to the domain of attraction of the EV dfs or their converses. The characterizing property is the range of the shape parameter γ in (3.2). According to classical EV theory, there is no restriction on the exponent γ in Pareto's law, other than it be positive. Its maximum likelihood estimate is the inverse of the logarithm of the geometric mean of the observations. However, if we consider its characteristic function, or generating function (gf), we are immediately confronted by the fact that the shape parameter is either limited to the open interval (0,1), or to the open interval (1,2). As we discuss in the Appendix, the Pareto df in each of these intervals represents the intensity of small jumps: In the (0,1) interval, the jumps are necessarily positive, while in the (1,2) interval there is a «compensation» of jumps (de Finetti, 1975), which can be thought of as being responsible for the existence of a mean value. In both cases the variance is infinite. Moreover, the shape parameter can even become negative, resulting in a Beta df. The Beta df has been used by Utsu (1960) and Vere-Jones (1966) in the modeling of aftershock sequences. If the original shape parameter is to be considered positive, but limited to the two open intervals mentioned above, there is only one way to consider the transition between positive and negative shape parameters and that is by considering a compound Poisson process: the times the jumps take place are regulated by a Poisson process, but the intensity of the jumps themselves are distributed according to the Pareto df. This affords a continuous approximation to a discrete jump process, and it is independent of the Poisson regulatory process.

In the case of fractals, the range of the Hausdorff dimension depends upon the dimension in which the fractal is embedded. Consider the Cantor set which is embedded in one-dimension. The Hausdorff dimension γ lies in the open interval (0,1). Other types of fractals are embedded in two-dimensions, like the Sierpinski gasket. The Hausdorff dimension lies in the open interval (1,2). It would appear that larger values of γ would necessarily correspond to higher embedding dimensions, and there is no limit on the dimensionality.

The rates of large earthquakes and aftershock sequences show different dependencies on the energy. For large earthquakes, the rate decreases as the energy increases, while for aftershocks the rate decreases with the energy (Utsu, 1960; Vere-Jones, 1966). In the latter case, it would appear that the fractal dimension is negative. The rates of large earthquakes and aftershocks can be derived from classical EV theory. Using elementary limits, the probability of largest value M_n not exceeding a value x in a sample of size n is

$$P(M_n \le x) = \exp\{-nP(X > x)\} = \exp(-nx^{-\gamma}).$$

It says that if the exceedances are distributed according to a Pareto tail df, the df of the maximum value will be a Fréchet df with a shape parameter $\gamma > 0$. Moreover, if T_x represents the smallest time necessary for the energy of an earthquake to exceed a threshold value x, then

$$P(M_n \le x) = (P(T_x > \tau) \approx \exp(-\lambda_x \tau)$$

for large threshold values. The above two expressions can be compared by observing that $n = \kappa\tau$, where κ is a constant frequency. The rate is thus found to be

$$\lambda_x = \frac{\kappa}{x^\gamma}. \qquad (3.3)$$

The rates of large earthquakes *decrease* with increasing energies.

In contrast, the rate of aftershocks should decrease monotonically to zero with the energy (Vere-Jones, 1966). If the probability of not exceeding the threshold value $x \le 1$, which is now the ratio of the energy of a shock to the energy of the largest shock in the group (Utsu, 1960), is given by a Beta df

$$P(X \le x) = x^{-\gamma}, \qquad x \le 1 \qquad (3.4)$$

with a negative shape parameter γ, the tail df of the smallest value of the energy, W_n, will asymptotically be given by

$$P(W_n > x) = \exp(-nP(X \le x)).$$

This is the class of converse Weibull dfs. The probability that no value of the energy will fall below the threshold value x in the time τ is the same as the probability that the waiting time T_x for this to happen is greater than τ, *viz.*,

$$P(W_n > x) = P(T_x > \tau) \approx e^{-\lambda_x \tau}$$

where the rate is

$$\lambda_x = \kappa x^{|\gamma|}. \qquad (3.5)$$

The shock rate (3.5) decreases monotonically to zero with the energy. This is in direct contrast to the rate of large earthquakes (3.3). A fortiori, Utsu (1961) gives some grounds for considering a shape parameter $\gamma = -2$, which is a particular case of the converse Weibull class known as the Rayleigh df.

The existence of negative shape parameters leads to a new type of fractal dimension

$$-\gamma := \frac{\ln P(X \le x)}{\ln x} \qquad x \le 1 \qquad (3.6)$$

which is referred to as the «cluster» dimension (Hastings and Sugihara, 1993; Lavenda, 1996). Definition (3.6) follows directly from the Beta distribution (3.4). In contrast to the fractal dimension (3.2), in which there is «thinning out» of the geometrical object, (3.6) shows a « filling in» of the volume $x^{|\gamma|}$. The logarithm of the number of aftershocks not exceeding the energy x will be given by

$$\ln N(X \le x) = \ln n + \gamma \ln x \qquad (3.7)$$

where n is their total number. Therefore, if the rates of large earthquakes and aftershocks are given by (3.3) and (3.5), respectively, aftershocks will not follow the Gutenberg-Richter law (3.1) as previously suggested (Scheidegger, 1975), but, rather be characterized by a line of positive slope in a double logarithmic plot (Lavenda, in preparation).

The occurrence of large earthquakes is usually assumed from a Poisson process in time (Lomnitz-Adler and Lomnitz, 1978, 1979). That is, a large earthquake is a rare event, and it might be supposed that if we took a series of large samples, the frequencies of large earthquakes would follow a Poisson distribution. However, this is not necessarily the case, for all land areas are not equally exposed to risk. Risk inequality makes the simple Poisson process unsuitable for large earthquake prediction. Simple Poisson processes give consistently high estimates of a seismic hazard (Johnston and Nava, 1985, and references cited therein). The independence of Poisson events can be summarized as «no matter how long it has been since the last one, we are no closer to the next one».

Each jump in the Poisson process will not trigger an earthquake; rather, it is the gradual build-up of the accumulated strain energy that is finally liberated as a main shock. The cumulative number of jumps up to time τ, $\sum_{i=1}^{N(\tau)} X_i$ is a *compound* Poisson process, where the number of random variables $N(\tau)$ is a Poisson process with rate $\lambda_x > 0$. The probability that up to epoch τ the process has exactly n jumps is $P(N(\tau) = n) = e^{-\lambda_x \tau}(\lambda_x \tau)^n/n!$ In other words, the time the jumps take place is regulated by a Poisson process while the jumps themselves, X_i, are random variables whose intensity is distributed according to a *continuous* Pareto df.

In the Appendix we derive an upper bound on the df of the sample mean in terms of the entropy reduction

$$\Delta S(x) = -nx^{-\eta}/\eta. \tag{3.8}$$

By combining the Pareto df for the intensity of positive jumps with the Poisson process which regulates the time of the jumps, a compound Poisson process is constructed with a new shape parameter

$$\eta := \frac{\gamma}{1-\gamma}. \tag{3.9}$$

Positive values of η, implying values of $\gamma < 1$, again implicate the Pareto df for the distribution of the intensity of the jumps, while negative values of the shape parameter in the interval $1 < \gamma < 2$ indicate that the intensity of the jumps follows a Beta df. The limiting case of $\gamma = 2$ corresponds to the normal df. Unlike the original shape parameter, γ, there are no limitations in magnitude placed on the new shape parameter (3.9) as γ varies on either the open interval $(0,1)$ or the open interval $(1,2)$.

Initially, values of $\gamma < 1$ meant that the average tends to infinity with the sample size. Now, thanks to the compound Poisson process, upper limit has been reduced to 1/2. The average is likely to be much greater than any single component, which is possible only if the maximum term grows exceedingly large and dominates the average. Since the expectation of the ratio of the sum of random variables and the maximum term tends to a positive constant for $\gamma < 1$ (Feller, 1971), the same bound can be placed on the df of the maximum value as on the df of the sample mean. The latter will be determined in the Appendix.

4. Thermodynamics of large earthquakes

In systems with a variable number of particles, or events, the entropy is directly proportional to their number. In constrained thermodynamic systems, the reduction in entropy is proportional to either the negative of the number of exceedances over a given threshold, or the negative of the number of events less than or equal to a threshold value. The former applies to EV dfs, while the latter to converse EV dfs, or the dfs of the minimum (Lavenda, 1995). In the Appendix we identify the entropy reduction (3.8) with the rate function of the large deviation principle.

The entropy reduction (3.8) is concave, nonpositive and continuous. It tends to zero as the energy increases without limit, which coincides with the state of maximum disorder. A temperature,

$$\frac{d}{dx} S(x) = nx^{-(\eta+1)} = \beta \tag{4.1}$$

can be defined in accordance with the second law, where β is the inverse temperature measured in energy units. This temperature is to be associated with the reduction in the amount of stored elastic energy in the stress field when rupture occurs along a fault. The energy can be released in the form of radiation as seismic waves, dissipated into heat, or used partially to perform work. The thermal equation of state,

$$x = \left(\frac{n}{\beta}\right)^{1-\gamma} \qquad (4.2)$$

displays the fact that the energy grows more slowly than the temperature. This is a hallmark of a *constrained* thermodynamic system (Lavenda, 1995).

The temperature is proportional to the average kinetic energy. Ordinarily, the energy is greater than the average kinetic energy enabling the system to utilize the excess energy in forms other than translational motion. For instance, the energy increases as the fourth power of the temperature in thermal radiation. The energy can be used to create photons. In bound systems, the energy increases at a slower rate than the average kinetic energy of its constituents, like the slow build-up of elastic energy in the stress field which is generated in a limited region about the fault.

In fact, earthquakes bear a remarkable resemblance to the thermodynamics of high-energy particle physics (Lavenda, 1997). If the statistical physics of earthquakes followed the law of large numbers, the energy would be distributed as a Gamma density (Lavenda, 1991)

$$f(x) = \frac{\beta(\beta x)^{\nu-1}}{\Gamma(\nu)} e^{-\beta x}. \qquad (4.3)$$

According to the law of equipartition of energy, which is given by the first moment of (4.3), $\beta\bar{X} = \nu$, the average energy, $\bar{X}$, is equal the product of half the number of degrees-of-freedom, ν, and the absolute temperature. If we identify this number of half degrees-of-freedom with the number of exceedances above a threshold energy, the law of equipartition generalizes to

$$\beta x = N(X > x). \qquad (4.4)$$

Now, in order that the thermal equation of state hold, the number of events exceeding the energy x must be given by

$$N(X > x) = -\eta\Delta S(x).$$

The absolute value the entropy reduction is proportional to the number of events exceeding a given energy, which in turn is proportional to the rate (*cf.* eq. (3.3)).

If we consider (4.4) as the transformation of a variate $z = \beta x$, which obeys a Gamma density (4.3), then the variate x will be distributed according to

$$g(x) = \eta \frac{n^{\nu}}{\Gamma(\nu)} x^{-\nu\eta-1} e^{-n/x^{\eta}} \qquad (4.5)$$

which is the order-statistic density of the Fréchet df. For the largest value $\nu = 1$, (4.5) is the Fréchet density. The Gutenberg-Richter energy-magnitude relation (Gutenberg and Richter, 1949)

$$\ln x = ay + b \qquad (4.6)$$

where a and b are assumed to be universal constants, transforms the Fréchet order-statistic density (4.5) for the energy into the order-statistic Gumbel density

$$g(y) = A\frac{B^{\nu}}{\Gamma(\nu)} e^{-\nu Ay - Be^{-Ay}} \qquad (4.7)$$

for the magnitude y. A and B are positive constants with values $A = \eta a$ and $B = ne^{-\eta b}$. The mean magnitude of earthquakes with $y > 0$ is 1/A. This means that the smaller the shape parameter, η, the larger is the mean magnitude (*cf.* Section 6). For $\nu = 1$ (4.7) is the Gumbel density.

For arbitrary ν, (4.7) will have a maximum at

$$\tilde{y} = A^{-1} \ln\left(\frac{B}{\nu}\right)$$

which is the modal magnitude. Rearranging this expression gives

$$\ln N(Y > \tilde{y}) = \ln B - A\tilde{y}$$

where $N(Y > \tilde{y}) = \nu$. Thus the maximum modal magnitude is the magnitude that we can expect one earthquake of magnitude of $\tilde{y}$ or larger, in a τ year period since $N(Y > \tilde{y}) = 1$. In the τ year period there has been a total of n earthquakes.

The mean return period,

$$\mathcal{T}_y = e^{Ay}/B = 1/N(Y > y)$$

is the mean interval between large earthquakes (Gumbel, 1958). The mean return time of the most frequently observed maximum magnitude, $\tilde{y}$, is $\mathcal{T}_{\tilde{y}} = 1$ (*i.e.* $N(Y > \tilde{y}) = 1$). Finally,

$$H(y) = -\ln(1 - e^{-\kappa\tau e^{-\eta(ay+b)}})$$

is the cumulative hazard function for an earthquake of magnitude y in a τ year period. Its derivative is the hazard rate, or the mortality rate at «age» y. These expressions were tested against the California earthquake Catalogue for the period 1932-1962 and were found to provide an adequate basis for making predictions on the occurrences of the largest earthquake magnitudes in time (Epstein and Lomnitz, 1966).

When two nucleons with very large energies collide, energy is suddenly released into a small volume surrounding them. The volume is determined not by the exterior geometry of the vessel; rather, it is carved out by the redistribution of the energy among the various degrees-of-freedom present in the volume that have succeeded to reach thermal equilibrium (Fermi, 1950). In other words, the necessity to redistribute the sudden release of energy, and the thermalization of the various modes into which the energy has been released, determine the characteristic volume.

The elastic strain, prior to an earthquake, is uniformly distributed over a volume V which houses an elastic energy, $\mathcal{E} = 1/2\,\mu\epsilon^2 V$, where ϵ is the strain, and μ is the elastic modulus of the rock. When an earthquake occurs, a certain fraction of this energy is radiated away as seismic waves, while the rest is either dissipated as heat or partially converted into compressional or dilatational work. According to Benioff's elastic strain rebound theory, the strain is a logarithmic increasing function of time, and it is directly proportional to the square root of the radiated energy (Lomnitz, 1974). The constant of proportionality varies from earthquake to earthquake. This implies that the earthquake volume is constant if the fraction of radiated energy is independent of magnitude. However, it appears that the earthquake volume is a steeply increasing function of the magnitude (Lomnitz, 1974). Thus the strain should be independent of the magnitude (Scheidegger, 1975). The question is: how does the earthquake volume acquire a dependence on the magnitude?

In general, the mechanical equation of state relates two independent variables, say volume and temperature, to the pressure. If an additional equation of state is imposed, then only one variable can be varied independently. The thermal equation of state (4.1) is just such a constraint. It is analogous to a polytropic change in a stellar system in which the heat capacity is made to vary in a specified way during a quasistatic change. The «heat capacity» $c = dQ/dT$ has two extremes, $c = 0$ in which the quasistatic change is adiabatic, $dQ = 0$, and an isothermal change, $dT = 0$, in which it is infinite.

Taking into account the elastic energy stored in the rock, we introduce the scaled energy, $\varepsilon = \mathcal{E}/\mathcal{E}_R$ into the increment in the entropy

$$dS = \tilde{\beta}(dx - d\varepsilon) \qquad (4.8)$$

where $\mathcal{E}_R$ is the residual elastic energy that is uniformly distributed over a reduced volume, V_R, after the shocks have occurred. When both the elastic modulus and strain are assumed to be constant throughout the region, ε becomes the scaled earthquake volume $v = V/V_R$. The temperature, defined in (4.1) is, in general, much higher than the thermodynamic temperature $1/\tilde{\beta}$ appearing in (4.8). Their difference is proportional to the thermal gradient which is responsible for the constraint that is imposed upon the increment in the entropy given by (4.8).

For illustrative purposes, let us suppose that the thermodynamic temperature is given by the ideal thermal equation of state $\tilde{\beta} = \nu/x$, where the number of degrees-of-freedom, proportional to ν, can associate with faults, or subfaults, into which elastic energy has been stored. Then equating the increment in entropy (4.8) with that given in (4.1) and integrating result in

$$\ln x + \ln\left(1 + \frac{N(X > x)}{\nu(\eta - 1)}\right) = \ln \nu$$

where the arbitrary constant of integration has been set equal to zero. The ratio in the argument of the logarithm is proportional to the number of shocks which have an energy exceeding x to the number of the degrees-of-freedom in the earthquake volume V into which the elastic energy has been distributed. For sufficiently large x this will be a small number so that upon expanding the logarithm to first order and introducing the energy-magnitude relation, (4.6), we obtain

$$\ln \nu \approx ay + b + \frac{Be^{-Ay}}{\nu(\eta - 1)}.$$

Neglecting the last term because of its smallness in comparison with the other two terms, the above expression reduces to the empirically determined relation with a = 1.47 for the surface wave magnitude and b = 9.58 (Scheidegger, 1975). The ratio of the thermodynamic temperature to the earthquake temperature stand in the same proportion as the ratio of the number of shocks exceeding magnitude y to the number of degrees-of-freedom present in the earthquake volume.

5. The Catalogue of Chinese Earthquakes revisited

Lomnitz-Adler and Lomnitz (1978, 1979) proposed

$$\ln N(Y > y) = C - De^{ay} \qquad (5.1)$$

as a generalization of the Gutenberg-Richter law, where C and D are positive constants. The constant C was identified as the logarithm of the total number of events with magnitude y = 0 and greater, and a comes from the magnitude-energy relation (4.6), where a = 1.5 ln 10 and b = 11.8 ln 10 are supposedly universal constants.

Lomnitz-Adler and Lomnitz considered the small magnitude limit and expanded the exponential function in (5.1) in a Maclaurin series in powers of y. To first order, they recovered the Gutenberg-Richter relation. That is, the generalized relation (5.1) tends asymptotically to the original Gutenberg-Richter relation in the low magnitude range [4]. However, the Gutenberg-Richter relation, like the Pareto df, can only be fitted to the *upper* tail, and not to the lower one. As the authors correctly note, the exponential e^{ay}, which replaces the linear term in the original magnitude-frequency relation, diverges for large y.

Previous to the proposed generalization of the Gutenberg-Richter relation (5.1), Epstein and Lomnitz (1966) suggested

$$\ln G(y) = -Be^{-Ay}. \qquad (5.2)$$

Observing that the expected number of earthquakes which have a magnitude exceeding y is Be^{-Ay}, (5.2) can be written as

$$\ln N(Y > y) = \ln B - Ay.$$

Both (5.1) and (5.2) involve a double exponential df; (5.1) would be the logarithm of tail of the Gompertz df if C vanished, while (5.2) is the logarithm of the Gumbel df.

Unlike the Fréchet df, where the variate is always positive, double exponential dfs have an unlimited range. Nevertheless, even in cases where the variate must be positive, one may prefer to use a double exponential df. Recall that the coefficient B is proportional to the sample size, n. Preference of (5.2) as the df of magnitude to the Fréchet df for the distribution in energy is due to the observation that the probability of the largest magnitude to be negative, $\boldsymbol{P}(0) = \exp(-B)$, decreases with increasing n (Gumbel, 1958). Since this is also the probability of the smallest magnitude being positive, large sample sizes guarantee that the largest magnitudes are positive.

(4) Kijko (1982) assumes that the probability of an earthquake is given by a converse Weibull df for the smallest value of the energy. Introducing the Gutenberg-Richter magnitude-energy relation reduces it to the generalized Gutenberg-Richter law (18) with $C = 0$. Since the integral of the Weibull density over all values of the energy is finite, he concludes that there is no need to introduce an upper bound to the magnitude. Yet it is precisely because the converse Weibull df has an upper endpoint that the integral over all energies is finite. Although there is no need to specify the right endpoint for purposes of convergence, it nevertheless exists.

Because the Gompertz df derived by Lomnitz-Adler and Lomnitz (1978, 1979) is size independent, there is no guarantee that the magnitude will be limited to positive values. In order to ensure that the magnitude in their double exponential df will be positive, it would be necessary to introduce a location parameter which, if large enough, would make the probability of a value below the *a priori* lower bound negligible (Leadbetter *et al.*, 1983). This would require the introduction of some conventionally fixed magnitude in the energy-magnitude relation, and require a re-definition of n as the number of earthquakes equal or greater than that magnitude.

Instead of using the value $a = 3.454$ in (4.6), Lomnitz-Adler and Lomnitz found that the value $a = 0.17$ gave a better fit to the empirical data taken from the Catalogue of Chinese Earthquakes, 780 B.C.-1973 A.D. The catalogue has the fixed threshold value of magnitude 6 for which there are 629 events of greater or equal magnitude. Figure 1 shows that the sample mean excess function, or the mean residual life function, is a decreasing function of the energy. The straight line is obtained from a parametric maximum likelihood estimation using the statistical software XTREMES (Reiss and Thomas, 1997). The shape parameter was estimated to be

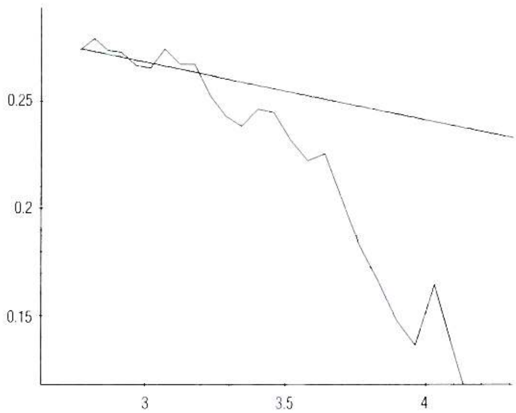

Fig. 1. The sample mean excess function as a function of the energy and its straight line parametric estimate of the Chinese Catalogue using the Lomnitz-Adler and Lomnitz best fit of the parameter $a = 0.17$.

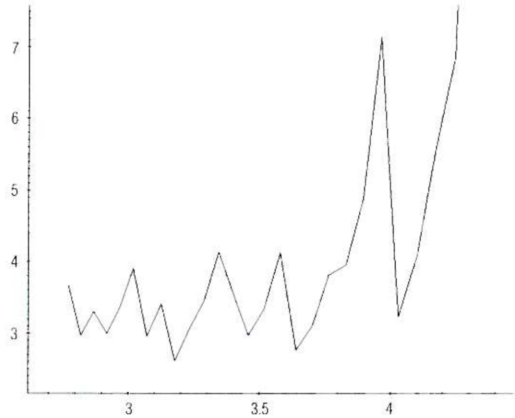

Fig. 2. The sample hazard function as a function of the energy for the Lomnitz-Adler and Lomnitz analysis of the Chinese Catalogue.

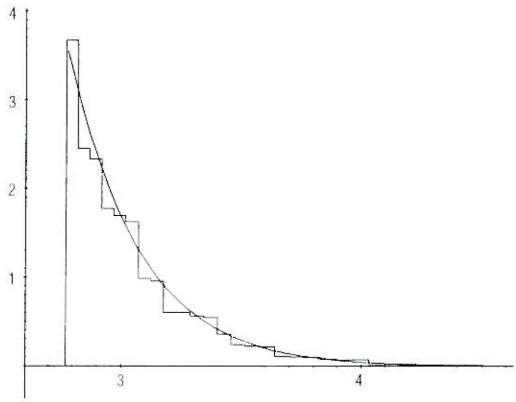

Fig. 3. The beta df for the energy fitted to the histogram using $a = 0.17$ in the magnitude-energy relation.

$\eta = 3.11$, or $\gamma = 0.757$, with location and scale parameters $\mu = 2.56$ and $\sigma = 0.34$, respectively. A straight line with negative slope would indicate modeling with a Beta df. The sample hazard function in fig. 2 tends to corroborate this conclusion, since it is an increasing function of the energy. In fig. 3, a Beta df is fitted to the histogram. If this Beta df is appropriate then there is a right endpoint to the distribution which is the «maximum energy» of an earthquake. Yet the deviations of the sample mean excess function and sample hazard function from straight line behavior, and the estimation of a positive shape parameter, rather than a negative one, cast doubt on the validity of modeling large earthquakes as a parametric df for the smallest value. Consequently, the motto that «the older you get the more probable is death» does not seem to apply to large earthquakes.

If the energy scale is stretched out using the logarithmic transformation between energy and magnitude (4.6), and the conventional values of a and b are employed, the picture changes dramatically. The sample mean excess function shown in fig. 4 has a straight line segment for magnitudes between 7.5 and 8.1. The parametric maximum likelihood estimate has a shape parameter, $\eta = 1.07$, or $\gamma = 0.52$, with a vanishing location parameter and scale parameter $\sigma = 101.73$. The minimum distance estimator between the sample and EV df supports a shape parameter $\eta = 1.04$, or $\gamma = 0.51$. Only generalized Pareto dfs have straight lines with positive slope for mean excess functions. This is corroborated by the sample hazard function, shown in fig. 5, which is a decreasing function of the energy. The smooth curve, η/x, with $\eta > 1$ and

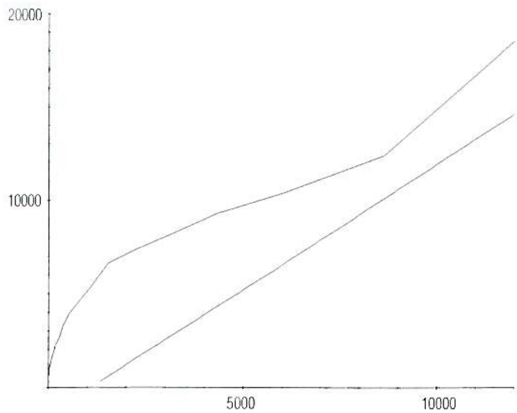

Fig. 4. The sample mean excess function and the straight line mean excess function estimated from the maximum likelihood method.

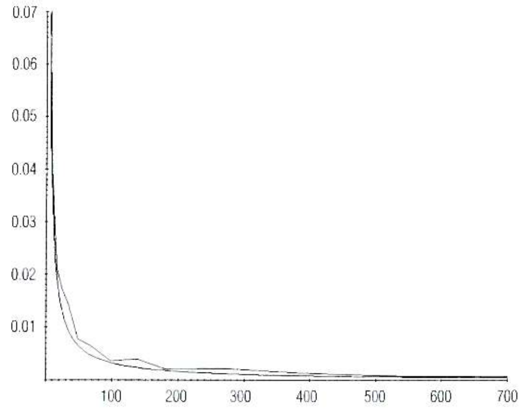

Fig. 5. The sample hazard function and the Pareto (smooth curve) hazard function.

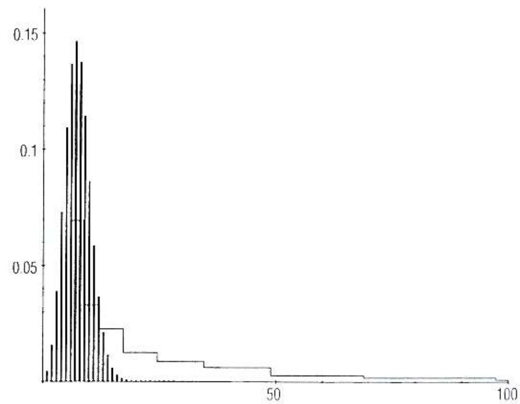

Fig. 6. The Poisson df is superimposed upon the histogram (energy in units of 10^{20} ergs).

$x > 1$, is the parametric hazard function of the Pareto df. For such processes, the longer the system is in operation, the less likely it is to fail.

Superimposed upon the histogram in fig. 6 is the Poisson df. It is visibly clear that the Poisson df cannot account for the long upper tail of the df. Alternatively, the Fréchet df is fitted to the histogram curve in fig. 7 using the minimum distance estimators for shape $\eta = 1.041$, location $\mu = 3.84$, and scale $\sigma = 1.55$, parameters. Employing the same estimators, the generalized Pareto tail df is seen to give a higher tail estimate. There is no right endpoint to these dfs so the event of «maximum magnitude» need not occur. Finally, the sample and estimated Fréchet quantile functions are shown in fig. 8. The qualitative similarity of the two curves reflects the goodness of fit of the Fréchet df to the sample df.

6. The «E_{max} catastrophe»

Consider the usual argument for a maximum energy (Knopoff and Kagan, 1977). The combination of the Gutenberg-Richter magnitude-frequency and magnitude energy relations gives a Pareto distribution, with a positive shape parameter α, for the probability of a scaled energy greater than $x > 1$,

$$\bar{F}(x) = x^{-\alpha}/\alpha \tag{6.1}$$

which has a heavy upper tail, especially if $\alpha < 1$. The «total amount of energy released by an earthquake» is

$$n\int_{x_{min}}^{x_{max}} \bar{F}(x)dx = \frac{n}{\alpha(1-\alpha)} x^{1-\alpha}\Big|_{x_{min}}^{x_{max}} . \tag{6.2}$$

For values of $\alpha < 1$, x_{max} must be finite while x_{min} can be zero, and since the Pareto and Fréchet dfs possess the same number of moments, the same holds for the Fréchet df. Moreover, the smaller the shape parameter, the larger the probability of a large value of the energy. Recall from Section 4 that $1/\alpha a$ is the mean magnitude of all earthquakes of magnitude $y > 0$.

We began by considering random variables which scale as $n^{1/\gamma}$. Their sum scales as $X_1 n^{1/\gamma}$, while their average has a $X_1 n^{(1-\gamma)/\gamma}$ scaling. If $\gamma < 1$, the average increases without limit with n. Now if these averages are considered as iid random variables, $S_n/n = (X_1 + ... + X_n)/n$, their average will have the same distribution as $S_1 n^{(1-2\gamma)/\gamma}$. In the event $\gamma > 1/2$, this factor will tend to zero as n does. However, in the case cited in Knopoff and Kagan (1977) $a \sim 1.5$ so that $\gamma = 3/7$, and, consequently, the factor

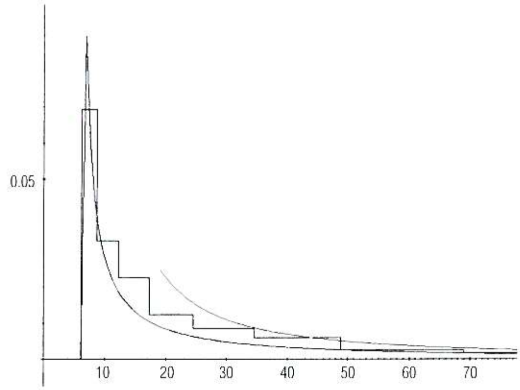

Fig. 7. The Fréchet density whose left endpoint is tied down at 6.16×10^{20} ergs, and the Pareto tail density are superimposed upon the histogram.

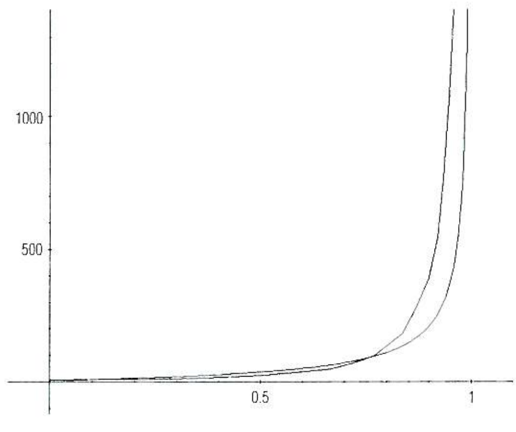

Fig. 8. Sample and Fréchet estimated (smooth curve) quantile functions.

$S_1 n^{(1-2\gamma)/\gamma}$, or equivalently $S_1 n^{(1-\eta)/\eta}$, will tend to infinity with n. The average is likely to be much larger than any of the variables which make up the average. Consequently there must be at least one component which grows exceedingly large and dominates the average. Given that $X_0 > x$, it is unlikely that any of the other X_i exceed this value. Even if the energies X_i were not independent, they become so as x increases. Hence the conditional probability for any $X_i > x$, given $X_0 > x$, tends to zero as x increases.

Why the condition $\gamma < 1/2$ makes the average of sums of random variables tend to increase without limit is due to a grouping effect where sums of random variables replace the random variables themselves. Without grouping, we require $\gamma < 1$, which corresponds to strictly stable dfs. Since strictly stable dfs do not have any moments, (6.2) cannot be used to determine the total energy that is released by an earthquake. If the intensity of the jumps follows a strictly stable Pareto law, the shape parameter $\gamma < 1$. But provided $\gamma > 1/2$, the shape parameter of the compound Poisson df $\eta > 1$, and consequently no maximum energy catastrophe occurs. The lower limit on the original shape parameter can be lowered continuously simply through the process of re-grouping.

The situation is analogous to insurance claims where a single large claim can control the entire portfolio (Teugels, 1984). The df of the claim size is assumed to be Pareto (6.1), and if $\alpha < 1$, the mean claim size (6.2) cannot be used to calculate the premium because it does not exist. To place an upper bound on (6.2) is artificial and meaningless. The problem becomes one of statistical estimation in which the hypothesis $\alpha > 1$ is to be tested against the alternative hypothesis $\alpha < 1$. If the null hypothesis is rejected then some kind of reinsurance should be considered.

Acknowledgements

This work was supported, in part, by Programma Accordo MICA-ENEA, and MURST ex-40% and 60%.

Appendix

In this section we will derive the compound Poisson process, and the large deviation principle expressed in terms of a particular rate function known as the entropy reduction. Regarding compound Poisson processes, Epstein and Lomnitz (1966) derived the Gumbel distribution from just such a process using the initial df rather than its convolution. That is, instead of considering the probability of a *sum* of iid random variables to be less than a given value, they considered the probability that *each* random variable is less than that value. The convolution makes a stronger statement, since the sum must be less than the given value, and it is the convolution that should be used to form a compound Poisson process (Feller, 1971). Considering the probability that each iid random variable is less than the threshold value is the same as asking for the probability that the maximum does not exceed that value (*cf.* Section 2), but averaging over all sample sizes using a Poisson df does not lead to a compound Poisson process. Rather, it gives the EV df (Epstein and Lomnitz, 1966). It is for this reason that we will work with the gf of the initial df since the probability of sums of random variables has a gf which is the product of the individual gfs.

Consider a random walk where we want to determine the probability that the system will be at a given position after j jumps have occurred. The time intervals between successive jumps correspond to independent variables each having an exponential density $\kappa e^{-\kappa\tau}$, where κ is the constant jump frequency. This is to say that the times in which the jumps take place are controlled by a Poisson process while the intensity of the small positive jumps are distributed according to a Pareto law, with shape parameter $\gamma < 1$. The substitution of a Pareto law for a discrete jump process means that we are considering a continuous approximation to the latter.

Since the number of jumps, j, that occurs in the time interval τ is regulated by a Poisson process, the gf of the compound Poisson process is

$$\mathcal{Q}_{\kappa\tau}(\beta) = \sum_{j=0}^{\infty} Q_1^j(\beta) \frac{(\kappa\tau)^j e^{-\kappa\tau}}{j!} = \exp\{\kappa\tau(Q_1(\beta) - 1)\}$$

where the gf of the random jumps is given by the Laplace transform

$$\mathcal{Q}_1(\beta) = \int_0^{\infty} e^{-\beta x}\, dF(x)$$

of the df F. We now specify that the df in the intensity of the jumps is given by a strictly stable Pareto tail df ($\gamma < 1$) so that the gf of the compound Poisson process will be given explicitly by

$$\ln \mathcal{Q}_{\kappa\tau}(\beta) = -\frac{\kappa\tau}{\Gamma(1-\gamma)}\int_0^{\infty}\frac{1-e^{-\beta x}}{x^{\gamma+1}}dx.$$

An integration by parts gives

$$\ln \mathcal{Q}_{\tau}(\beta) = -\kappa\tau\beta^{\gamma}/\gamma$$

which can easily be checked by differentiation. The gf is said to be infinitely divisible, meaning that any root, or power, of the gf is also a gf, $\mathcal{Q}_{\kappa\tau}(\beta) = \mathcal{Q}_1^{\kappa\tau}(\beta)$.

Póyla's conditions for any $\mathcal{Q}(\beta)$ to be a gf are: $\mathcal{Q}(0) = 0$, $\mathcal{Q}(\beta) \geq 0$ and decreasing, and that it be continuous convex. The latter implies $\gamma < 1$ and the corresponding dfs are said to be «strictly stable». Lévy extended the range to $1 < \gamma < 2$, and termed these dfs «quasi-stable». These stable dfs represent continuous, small jump processes: the strictly stable laws allow positive jumps only, while the quasi-stable laws show a «compensation» in the jumps where both positive and negative jumps occur (de Finetti, 1975). Strictly stable laws have no moments at all. Quasi-stable dfs have a mean value which can be thought of as being achieved through a compensation in forward (positive) and backward (negative) jumps. This prevents the process from running off to infinity. In regard to aftershock sequences, this would mean that no individual shock can have an energy greater than the primary shock, while no such restriction can be placed on large earthquakes if they obey strictly stable laws. Both strictly stable and quasi-stable laws have infinite variance; for if the variance were finite then the probability of a random walker to be at a given point after a larger number of jumps would tend to a normal df independent of the underlying jump probabilities.

We can now determine an upper bound for the cumulative probability distribution in an analogous way that the large deviation principle determines bounds on the tail of the df of the sample average. Setting $\kappa\tau = n$, the size of the sample, we can write the distribution of the sum, $S_n = X_1 + ... + X_n$, of iid random variables as (Varadhan, 1984)

$$\boldsymbol{P}_n(S_n/n \leq x) = \int_0^x d\boldsymbol{P}_n(z) \leq e^{\beta x}\int_0^x e^{-\beta z} d\boldsymbol{P}_n(z) \leq e^{\beta x}\int_0^{\infty} e^{-\beta z} d\boldsymbol{P}_n(z) = e^{\beta x}\left[\mathcal{Q}_1\left(\frac{\beta}{n}\right)\right]^n$$

for any $\beta > 0$. The last expression utilizes the fact that the gf is infinitely divisible.

For a fixed threshold value x, the function, $\beta x + n \ln \mathcal{Q}_1(\beta/n)$, is convex in β with an infimum $\beta(x)$ given by the solution of

$$x = -n\frac{d}{d\beta}\ln \mathcal{Q}_1(\beta(x)/n) = \left(\frac{n}{\beta}\right)^{1-\gamma}.$$

This is none other than the thermal equation of state (4.2). Introducing the value of $\beta(x)$ found in the above expression into $\beta x + n \ln \mathcal{Q}_1(\beta/n)$ results in the concave expression (3.8) for the entropy reduction. Consequently, the upper bound on the probability df can be expressed as a generalized Boltzmann principle for the large deviations (Lavenda, 1995)

$$\boldsymbol{P}(S/n \leq x) \leq e^{\Delta S(x)}.$$

Since the expectation of the ratio $\boldsymbol{S}_n/\boldsymbol{M}_n$ tends to a positive constant if $\gamma < 1$ (Feller, 1971), we may use the exponential of the entropy reduction as the upper limit for the probability of the maximal term $\boldsymbol{M}_n$ being less than or equal to x.

REFERENCES

BAK, P. (1996): *How Nature Works* (Springer-Verlag), 12-14.

BOLT, B.A. (1970): Causes of earthquakes, in *Earthquake Engineering* (Prentice-Hall), 21-45.

CAPUTO, M. (1978): The distribution of the faults and the maximum magnitude of a seismic region, in *Proceedings 1976 Friuli Earthquake and Antiseismic Design of Nuclear Installations, CNEN, Rome*, 709-736.

CHINNERY, M.A. and R.G. NORTH (1975): The frequency of very large earthquakes, *Science*, **109**, 1197-1198.

CORNELL, C.A. (1968): Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.*, **58**, 1583-1606.

DE FINETTI, B. (1975): *Theory of Probability*, vol. 2 (Wiley), 85-92.

EPSTEIN, B. and C. LOMNITZ (1966): A model for the occurrence of large earthquakes, *Nature*, **211**, 954-956.

ESTEVA, L. (1975): Geology and probability in the assessment of seismic risk, *Inst. Eng. UNAM, Rep. E13*, Mexico.

FELLER, W. (1971): *An Introduction to Probability and its Applications*, vol. II (Wiley), pp. 172.

FERMI, E. (1950): High energy nuclear events, *Prog. Theor. Phys.*, **5**, 570-583.

GUMBEL, E.J. (1958): *Statistics of Extremes* (Columbia University Press), 272 ff.

GUTENBERG, B. and C.F. RICHTER (1949): *Seismicity of the Earth and Associated Phenomena* (Princeton University Press), pp. 18.

HASTINGS, H.M. and G. SUGIHARA (1993): *Fractals: a User's Guide for the Natural Sciences* (Oxford University Press), 44-46.

JOHNSTON, A.C. and S.J. NAVA (1985): Recurrence rates and probability estimates for the New Madrid Seismic Zone, *J. Geophys. Res.*, **90**, 6737-6753.

JONES, I.F., L. MANSINHA and P.-Y. SHEN (1982): On the double exponential frequency-magnitude relation of earthquakes, *Bull. Seismol. Soc. Am.*, **72**, 2373-2375.

KIJKO, A. (1982): A comment on «A modified form of the Gutenberg-Richter magnitude-frequency relation» by J. LOMNITZ-ALDER and C. LOMNITZ, *Bull. Seismol. Soc. Am.*, **72**, 1759-1762.

KNOPOFF, L. and Y. KAGAN (1977): Analysis of the theory of extremes as applied to earthquake problems, *J. Geophys. Res.*, **82**, 5647-5657.

LAVENDA, B.H. (1991): *Statistical Physics: a Probabilistic Approach* (Wiley-Interscience), 170-173.

LAVENDA, B.H. (1995): *Thermodynamics of Extremes* (Albion), 6-8.

LAVENDA, B.H. (1996): Fractals, clusters, and order-statistics: a new relation between probability and entropy, *Int. J. Theor. Phys.*, **35**, 1993-2011.

LAVENDA, B.H. (1997): Thermodynamics at high energies, *Int. J. Theor. Phys.*, **36**, 1733-1744.

LEADBETTER, M.R., G. LINDGREN and H. ROOTZÉN (1983): *Extremes and Related Properties of Random Sequences and Processes* (Springer-Verlag), pp. 269.

LOMNITZ, C. (1974): *Global Tectonics and Earthquake Risk* (Elsevier), ch. 8.

LOMNITZ, C. (1994): *Fundamentals of Earthquake Prediction* (Wiley), pp. 161.

LOMNITZ-ADLER, J. and C. LOMNITZ (1978): A new magnitude-frequency relation, *Tectonophysics*, **49**, 237-245.

LOMNITZ-ADLER, J. and C. LOMNITZ (1979): A modified form of the Gutenberg-Richter magnitude-frequency relation, *Bull. Seismol. Soc. Am.*, **69**, 1209-1214.

LOMNITZ-ADLER, J. and C. LOMNITZ (1982): Reply to A. KIJKO, a comment on «A modified form of the Gutenberg-Richter magnitude-frequency relation», *Bull. Seismol. Soc. Am.*, **72**, 1763.

OTSUKA, M. (1973): Cutoff of seismic energy, *J. Phys. Earth (Tokyo)*, **21**, 119-123.

REISS, R-D. and M. THOMAS (1997): *Statistical Analysis of Extreme Values* (Birkhäauser), 12-16.

SCHEIDEGGER, A.E. (1975): *Physical Aspects of Natural Catastrophes* (Elsevier), pp. 42.

SMITH, S.W. (1976): Determination of the maximum earthquake magnitude, *Geophys. Res. Lett.*, **3**, 351-354.

TEUGELS, J.L. (1984): Extreme values in insurance mathematics, in *Statistical Extremes and Applications*, edited by J. TIAGO DE OLIVEIRA (Reidel), 253-259.

UTSU, T. (1961): A statistical study on the occurrence of aftershocks, *Geophys. Mag. (Tokyo)*, **30**, 521-605.

VARADHAN, S.R.S. (1984): *Large Deviations and Applications*, Soc. Ind. and Appl. Math., pp. 7

VERE-JONES, D. (1966): A Markov model for aftershock occurrence, *Pure Appl. Geophys.*, **64**, 31-42.

YEGULALP, T.M. and J.T. KUO (1974): Statistical prediction of the occurrence of maximum magnitude earthquakes, *Bull. Seismol. Soc. Am.*, **64**, 393-414.

(received March 15, 1999;
accepted December 14, 1999)