Eartquake Death Tolls
Leon Knopoff, Didier Sornette

▶ To cite this version:

HAL Id: jpa-00247166
https://hal.science/jpa-00247166
Submitted on 1 Jan 1995

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Earthquake Death Tolls

Leon Knopoff(1) and Didier Sornette(2)

(1) Department of Physics and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90095, USA
(2) Laboratoire de Physique de la Matière Condensée, CNRS URA 190, Université de Nice-Sophia Antipolis, B.P. 71, Parc Valrose, 06108 Nice Cedex 2, France

Received 12 August 1995, accepted 31 August 1995

Abstract. — In the risk and insurance literature, the (one-point) distributions of losses in natural disasters have been proposed to be characterized by ‘fat tail’ power laws, i.e. very large destruction may occur with a non-vanishing rate. A naive hypothesis of uncorrelated Poissonian occurrence would suggest that the losses are solely characterized by the properties of the underlying power law distributions, i.e. the longer we wait, the more dramatic will be the largest disaster, which could be as much as a finite fraction of the total population or the total wealth of a country. We find indeed that the numbers \( Z \) of deaths in the very largest earthquakes of this century can be described by a power law distribution \( P(Z) \propto Z^{-(1+\delta)} \) with \( \delta = 1.0 \pm 0.3 \), implying an unbounded behavior for the most devastating earthquakes. However, the distribution of the number of deaths per capita in each country in this century has a well-defined maximum value, suggesting that the naive extrapolation of the power law distribution is incorrect and that the understanding of correlations is necessary to ascertain the level of risk from natural disasters. The one-point distributions only provide an upper bound of the expected risk. We propose a speculative model to explain the correlations between deaths in large earthquakes and their countries of occurrence: we suggest that large ancient civilizations that have matured into large present-day populations were the beneficiaries of isolation from marauders due to the relative geographic protection by tectonic processes largely of an orogenic nature.

1. Earthquake Size Distribution

The estimation of risk due to natural disasters is of increasing importance in view of the growth of the world’s population and its concentration in areas prone to earthquakes, volcanic eruptions, typhoons or floods. In the risk and insurance literature, the (one-point) distributions of losses have been proposed to be characterized by ‘fat tail’ power laws, i.e. very large destruction may occur with a non-vanishing rate. How do these distributions evolve over long times and what are the possible consequences of correlations between events on long-term insurance policies? A naive hypothesis of uncorrelated Poissonian occurrence would suggest that the losses are solely characterized by the properties of the underlying power law
distributions, i.e. the longer we wait, the more dramatic will be the largest disaster, which could be as much as a finite fraction of the total population or the total wealth of a country. Is this a valid assumption? We address this question using rank-ordering statistics on the distribution of earthquake fatalities $Z$ derived from historical data, which has been proposed to be a robust measure of losses.

It is well-known that "small" earthquakes are distributed in size according to the Gutenberg-Richter power law, $P(E)\,\text{d}E \sim E^{-(1+2b/3)}$, where $E$ is the energy released in the earthquakes. For example, a detailed analysis [1] of the World-Wide Harvard earthquake catalog (1977 – 1995) shows that $b = 1.00 \pm 0.02$ for shallow earthquakes with magnitudes less than a value in the 7.1 to 7.6 range. Larger earthquakes can be also described by a power law distribution but with a larger exponent $b = 2.3 \pm 0.3$. The cross-over between these two populations is ill-defined and we caution that there is no direct evidence to confirm the hypothesis that the large-earthquake branch is indeed a power law. In fact, a gamma distribution (power law truncated by an exponential roll-off) also fits the entire suite of earthquake moments from the smallest to the largest satisfactorily [2]. Another well-documented system is that of the earthquakes of the Southern California catalog (1932 – 1995): in this case, $b = 1.00 \pm 0.02$ up to the largest magnitude observed, $M_W \approx 7.5$ [1]. We caution that all existing earthquake catalogs span a time scale which is significantly shorter than the time interval between the earthquakes with magnitudes in the range $M_W \approx 8.5 – 9.5$ known to occur.

2. Distribution of Death Tolls

A first step toward the assessment of the level of risk associated with earthquakes is to determine the distribution function of the sizes of these disasters. The destructive power of an earthquake results from a conjunction of several factors, including the magnitude and moment of the earthquake, the depth of focus, the orientation of the earthquake fault, the distribution of population in the neighborhood of the earthquake, site effects such as liquefaction, landslides, the nature of the soil, the quality of the infrastructure including failure of utility lifelines, etc. In addition, one should take into account collateral damage due to fire, tsunami, environmental contamination, etc.

The physical sizes of earthquakes, the economic losses and the deaths that result from earthquakes are each measures of magnitude of risk. As an introduction to the problem, consider first the number of human deaths in each earthquake worldwide in this century [3]. Let $Z_1 \geq Z_2 \geq \ldots \geq Z_N$ be the rank-ordered number of deaths in each earthquake. For example, $Z_1$ corresponds to the Tangshan, China earthquake of 1976. Although deaths due to earthquakes are rare occurrences, the observation of a few tens of the largest cases is sufficient to determine the distribution with good precision using rank-ordering techniques; these methods were originally introduced in the study of the statistics of languages [4] and then afterward in a variety of fields [5], including earthquakes [1]. The rank-ordering distribution is the same as the cumulative distribution but with an interchange of axes. However, the statistical analyses of the two are different: the statistics of the rank-ordering procedure is concerned with the uncertainties in the values of $Z$ while the cumulative distribution is concerned with the uncertainties of the order numbers, which are integers. Here, we focus attention on those extreme events which characterize the distribution at the large end of the tail. In the case of power law distributions of the largest events, the tail of the distribution is dramatically constrained. and a surprisingly good recovery of the exponent can be derived from an extremely small data set. The reliability of the retrieval of the exponent has been explored in detail in [1].
Fig. 1. — Log-log rank-ordering plot of the distribution of deaths per earthquake: $Z_n$ is the $n$th largest number of deaths per earthquake, in decreasing order. The best fit to the entries having ranks 1 to 25 is shown by the straight line, which defines a power law for the extreme tail of the distribution (see text).

In Figure 1 we show the log-log rank-order distribution of $Z$. The equation $\log Z_n = \frac{1}{\delta} \log n + C$, with $\delta = 1.0 \pm 0.3$ and $C$ a constant, i.e. that $Z_n \sim n^{-1/\delta}$ is a good fit up to rank around 25. We recover the distribution $P(Z) \sim Z^{-(1+\delta)}$ characterizing the very largest death tolls. This is obtained using the equation $N \int_{Z_n}^{\infty} P(Z) dZ \simeq n$, which expresses that there are typically $n$ values of $Z$ larger than or equal to $Z_n$ out of a set of $N$ entries (see [1] for a more precise derivation and the appendix). The error bar $\pm 0.3$ on the determination of the exponent $\delta$ results from the usual square root uncertainty of the maximum likelihood estimator [6], from a systematic bias resulting from the estimator itself, as well as the existence of a crossover to a different distribution for the larger ranks, which correspond to smaller death tolls (see [1] for discussion).

The above result suggests that the longer we wait, the more dramatic will be the largest disaster, which could be a significant fraction of the population of a country. If $\delta < 1$, then the average number $< Z >$ of deaths per earthquake is given by $< Z > = \int_1^{Z_1} ZP(Z) dZ \simeq \frac{\delta}{1-\delta} Z_1^{1-\delta}$, where $Z_1$ is the maximum observed size. Since $Z_1$ grows as $N^{1/\delta}$ with $N$, then $< Z >$ diverges as $< Z > \sim N^{1/\delta-1}$. Practically, this means that the number of deaths $Z_1$ due to the single most disastrous earthquake among the total of $N$ earthquakes accounts for a finite fraction of the total deaths in all earthquakes, which is $N < Z >$. Indeed, the above estimate gives the finite fraction $\frac{Z_1}{N < Z >} \simeq \frac{1-\delta}{\delta}$, confirming our statement; a more rigorous derivation is given in [7], p.169. Thus, for example if $\delta = 2/3$, the correct formula predicts that the largest event corresponds on average to $1/3$ of the total.
If $\delta$ turns out to be larger than 1, the average number $< Z >$ of deaths per earthquake is well-defined. However, its standard deviation $< Z^2 > - < Z >^2$ diverges, meaning that as time goes on there are larger and larger fluctuations. The practical consequence is that the probability of human extinction (or the probability of "ruin") remains important. These ideas, put in quantitative terms in [8], rely on the assumption that the power law distribution is limited only by the finite size of the population, and hence that the possibility of an extinction is finite even if quite small.

3. Per Capita Death Toll

We show that the above estimates may be pessimistic because they overlook the existence of correlations between different measures of destruction in earthquakes. We propose that a more sensible measure of the impact of earthquake destruction on human activity is a variable that increases with time over a long time, normalized by a relevant measure of wealth. Indeed, for governments and insurance companies [9], the fundamental question is the adequate coverage of the risk. One strategy is to divide the region at risk into districts and apply an estimate for each districts separately. Thus, in principle, one could focus the resources of the insuring agencies on the regions at greater risks. For the purposes of constructing an example, suppose the region at risk to be the entire earth, and the districts to be the individual countries. On the basis of our assumed local, i.e. national, strategy, we would then expect a greater risk for an earthquake prone country, than in a tectonically stable part of the world. While this is a trivial statement, and is easily extended to smaller regions at risk with smaller subdivisions into districts, the more important question is to estimate the risk to be applied to each district [10].

As an introduction to the problem of the development of a risk strategy that takes into account the cumulative value of loss, we take as our estimate of risk the per capita number of human deaths due to earthquakes in this century for each country, referred to the population in 1973 [11,12]. Our justification for this procedure is that what is important economically is not the absolute value of the loss but its magnitude relative to the surrounding wealth. Also, the analysis of single events can tell us something over long times, only if we make reasonable assumptions on their correlations in time and space. We attempt to obtain information regarding the important correlations between losses over a relatively long time.

Let $D_1 \geq D_2 \geq \ldots \geq D_n \geq \ldots \geq D_N$ be the rank-ordered per capita deaths for each country. For example, $D_1 = 9.36 \times 10^{-3}$, $D_2 = 8.39 \times 10^{-3}$ and $D_3 = 4.95 \times 10^{-3}$ correspond to Armenia, Turkmenia and Guatemala. Figure 2 shows three types of plots: a) log $D_n$ as a function of log $n$; b) $D_n$ as a function of log $n$; c) log $D_n$ as a function of $n$. From examination of Figure 2a, it is clear that the log-log plot is not linear, which would have defined a power law distribution. If we insist on fitting a power law distribution, we find that only about the first largest 10 entries can be fitted with exponent $\delta = 0.7 \pm 0.4$. Figure 2b shows a slightly better fit: a linear relation between $D_n$ and log $n$ is a reasonably good description of the data up to rank about 15. In this case, we get the distribution $P(D) \sim e^{-D/D_0} \text{ where } D_0 = (8 \pm 1) \times 10^{-3}$ Finally, Figure 2c shows a remarkable linear dependence of log $D_n$ as a function of $n$ for the first 37 countries out of the 47 countries with deaths due to earthquakes in this century. Beyond entry 37, there is a roll-off which may be attributable to finite size effects, since the last ten points correspond to countries having about 200 or less deaths, a figure that is very small compared to the number of deaths for the largest entries in the list, which are roughly $4 \times 10^5$ for China, $1.5 \times 10^5$ for Japan, $10^5$ for Italy, etc. The best fit is log $D_n = \log D_{\max} - \alpha n$, with $D_{\max} \simeq 10^{-2}$ and $\alpha \simeq 0.18$. Thus the distribution is $P(D) \sim D^{-1}$ for $D < D_{\max}$ and is 0 for $D > D_{\max}$. Let us stress that this distribution is completely different from a pure power law $D^{-(1+\delta)}$ with $\delta \rightarrow 0$ and with no cut-off. Even for $\delta$ as small as $10^{-2}$ (this is all the more true for larger $\delta$).
Fig. 2. — The $D_n$ are the rank-ordered per capita deaths for each country. a) $\log D_n$ as a function of $\log n$; b) $D_n$ as a function of $\log n$; c) $\log D_n$ as a function of $n$: in this case, the straight line fit defines the distribution $P(D) \sim D^{-1}$ for $D < D_{\text{max}}$ and 0 for $D > D_{\text{max}}$. 
Fig. 3. — Log-log rank-ordering of the distribution of national populations $p_N$ in 1994 ($p_1$ corresponds to China, etc.). The linear fit defines a power law distribution with exponent $\mu = 1 \pm 0.3$ for the 50 most populous countries.

the representation (a) showing log $D_n$ as a function of log $n$ qualifies as the best fit to $D^{-(1+\delta)}$, as have been checked with synthetic tests [13]. Therefore, the representation (c) showing a linear dependence of log $D_n$ as a function of $n$ is characteristic of a distribution with a finite upper cut-off. These results show that power law behavior with nonzero $\delta$ for the per capita distribution is the poorest fit of the three attempts. Despite the existence of the power law distribution $D^{-\delta}$ for the death toll per earthquake, the important result of this analysis is that there is a well-defined maximum number of deaths per capita $D_{\text{max}}$ that seems to have been adequately sampled even for the limited earthquake history of this century. The value of $D_{\text{max}}$ is mainly constrained by the entries having the large ranks up to 37 whose deaths per capita range from $10^{-3}$ to $10^{-5}$; nevertheless the entries with the lowest ranks, i.e. those with the largest numbers of deaths per capita, are quite compatible with this determination. While we have no explanation for the $D^{-1}$ dependence for $D < D_{\text{max}}$, we stress that the important result is the existence of the cut-off.

4. Discussion

To summarize, the distribution of the number of deaths per earthquake is a power law with exponent $\delta \approx 1$. The same law holds for the distribution of the total number of deaths due to earthquakes in each country, which is reasonable in view of the stability properties of sums of variables distributed with power laws (Lévy laws are stable laws with respect to addition). However, the distribution of death per capita for countries in this century has a well-defined upper limit, i.e. there is a sharp cut-off to the distribution. How can we rationalize this paradox?

A rank-ordering analysis of the distribution of national populations shown in Figure 3 shows that the probability $P(p)$ that a country has $p$ inhabitants is given by $P(p) \approx p^{-(1+\mu)}$ with $\mu \approx 1$ for ranks up to 50, i.e. for the 50 most populated countries. If the death toll per earthquake (or the total number of deaths per country) and the national population were
uncorrelated, we could simply deduce the distribution of the death toll $D$ per capita, using $D \sim Z/p$. The distribution of a variable which is the ratio of two variables, each with power law distributions, is itself a power law with an exponent equal to that of the variable in the numerator, since in an uncorrelated universe the largest value of the ratio occurs when the numerator explores the tail of its distribution while the denominator remains in the center of its own distribution (this argument is correct if $p$ is bounded from below and is nonzero).

Our result is in disagreement with this simple argument, but instead suggests that there is a strong correlation between the death toll in earthquakes and population. This might seem a posteriori a trivial statement but it is interesting to see it confirmed by a statistical analysis. This example provides a non-trivial case where important correlations destroy power laws, showing a possible outcome of the interplay between power law tails and correlation. The simplest correlation between country size and earthquake death toll is the obvious fact that death toll increases with population. A second type of correlations is that destruction and death are more probable in poor countries or in countries with poor construction standards. A more subtle correlation that we would like to suggest is that large populations are often found in tectonically active, earthquake prone areas. We speculate that the development of ancient civilizations, i.e. the development of large communal populations, took place because of geographical isolation, induced by tectonic processes of mountain building that presented natural defenses. Examples of such ancient civilizations are to be found in Anatolia (Turkey), China, Greece, India, Japan, Java (Indonesia), Mexico, Persia, Peru, Rome. The absence of tectonic defences in the plains promoted diffusion by marauders and hence the incubation processes necessary to develop large populations could not emerge. The growth of these populations has continued to this day. It was much later, in the middle ages, that global trade emerged as an even more potent force that allowed for the development of large populations in the plains, for example, of northern and western Europe. And the exploration and development of North America took place because of these more potent forces as well. We note that other studies have proposed correlations between earthquake active regions and the development of human populations [14] and between earthquakes and emerging infections [15] and it remains to be seen if these issues are components of the correlations that must be quantified for a global insurance policy. In this regard, we find that the distribution of population sizes of countries in the death toll earthquake list is very different from that of the complete set. As pointed out, this suggests that the occurrence of earthquakes may influence population growth.

In sum, our results stress the importance of taking into account correlations between the number of casualties in earthquake events in order to define a more relevant global insurance policy than has been considered to this time.

Acknowledgments

We are grateful to C. Vanneste and Y.Y. Kagan for useful discussions. This work has been partially supported by the CNRS-NSF International Cooperation program, and is Publication no. 223 of the Southern California Earthquake Center and Publication no. 4305 of the Institute of Geophysics and Planetary Physics, University of California, Los Angeles.

Appendix A

Consider a stochastic process in which the outcome $E$, which we call the energy generically, has the distribution

$$P(E)dE = \frac{C}{E^{1+\mu}}dE,$$  \hspace{1cm} (1)
with $E_{\min} = 1 < E < \infty$; then $C = \mu$. Suppose that $N$ events occur within a given time interval. Let $E_1 \geq E_2 \geq \ldots \geq E_n \geq \ldots \geq E_N$ be the energies of the events listed in descending order. The probability $F(E_n) dE_n$ that the energy for the nth order event be equal to $E_n$ within $dE_n$ is [5]:

$$F(E_n) dE_n = \left( \frac{N}{n} \right) \left( 1 - \int_{E_n}^{+\infty} P(E) dE \right)^{N-n} \left( \int_{E_n}^{+\infty} P(E) dE \right)^{n-1} P(E_n) dE_n$$ \hspace{1cm} (2)

for an arbitrary distribution $P(E)$. For the power law, $F(E_n)$ has a peak at $E_n^{\text{max}} = \left( \frac{\mu}{\mu_n + 1} \right)^{1/\mu}$, which recovers and makes precise the scaling law for the rank ordered values stated in the text. If $P(E)$ is given by the power law (1), $F(E_n)$ can be expanded around this maximum as

$$F(E_n) = F(E_n^{\text{max}}) - \frac{1}{2} \left| \frac{d^2 F(E_n)}{dE_n^2} \right|_{E_n^{\text{max}}} (E_n - E_n^{\text{max}})^2 + \ldots,$$

which allows us to get an estimate of the standard deviation of $E_n$ through the calculation of $\Delta E_n \equiv \langle (E_n - E_n^{\text{max}})^2 \rangle^{1/2}$, where the brackets indicate that the statistical average is taken. We obtain

$$\frac{\Delta E_n}{E_n^{\text{max}}} = \left( \frac{1}{\mu (n \mu + 1)} \right)^{1/2}.$$ \hspace{1cm} (3)

References

[12] Because the political boundaries of the world change with a relatively short time constant, the definition of a nation is ephemeral. We have considered the nations that emerged from the breakup of the Soviet Union as independent entities for these purposes. Yugoslavia, Indonesia, New Guinea and others are single entries in our tabulation.
[13] We thank C. Vanneste for these simulations.