STATISTICAL TESTS FOR SCALING IN THE INTER-EVENT TIMES OF EARTHQUAKES IN CALIFORNIA

ÁLVARO CORRAL

Centre de Recerca Matemàtica, Edifici Cc, Campus UAB, E-08193 Bellaterra, Barcelona, Spain
AConcall at crm dot es

We explore in depth the validity of a recently proposed scaling law for earthquake inter-event time distributions in the case of the Southern California, using the waveform cross-correlation catalog of Shearer *et al.* Two statistical tests are used: on the one hand, the standard two-sample Kolmogorov-Smirnov test is in agreement with the scaling of the distributions. On the other hand, the one-sample Kolmogorov-Smirnov statistic complemented with Monte Carlo simulation of the inter-event times, as done by Clauset *et al.*, supports the validity of the gamma distribution as a simple model of the scaling function appearing on the scaling law, for rescaled inter-event times above 0.01, except for the largest data set (magnitude greater than 2). A discussion of these results is provided.

Keywords: Statistical seismology; scaling; goodness-of-fit tests; complex systems.

1. Introduction

In the last years considerable attention has been addressed to the distribution of inter-event times in natural hazards, in particular earthquakes[1,2,3,4,5,6,7,8,9,10], but also in human responses[11,12,13] and social behavior[14,15,16,17]. In some of these systems, the shape of the inter-event-time distribution for events above a certain threshold in size is independent on the threshold. Indeed, let τ denote the inter-event time, defined as the time between consecutive events above a size threshold s and let $D_s(\tau)$ be its probability density, then we can write

$$D_s(\tau) = R_s f(R_s \tau), \tag{1}$$

where f is a scaling function that provides the shape of $D_s(\tau)$ and R_s is the occurrence rate for events above s, providing the scale of $D_s(\tau)$ (and R_s^{-1} provides the scale for τ).

If the size distribution follows a power law (which is not always the case[10]), then $R_s \propto 1/s^\beta$ (where β is the exponent of the cumulative size distribution), and then

$$D_s(\tau) = \hat{f}(\tau/s^\beta)/s^\beta, \tag{2}$$

2   *Álvaro Corral*

which turns out to be a scaling law, equivalent to those obtained in the study of critical phenomena[18]. The law reflects a scale-invariant condition: there exist a change of scale in τ and s (a linear transformation) that does not lead to any change in the statistical properties of the process, at least regarding the inter-event-time probability density. The function $\hat{f}$ is just the scaling function f, except for proportionality constants.

Why is this scaling law of some relevance or interest? In general, when events are removed from a point process (as it is done in our case by raising the size threshold), the resulting inter-event-time distribution changes with respect to the original one, and a scaling law as Eq. (1) does not apply. However, for high enough thresholds s (for the extreme events that are of interest in hazard assessment studies), and when the events are randomly removed, it is expected that the resulting time process tends to a Poisson process (which means that the occurrence of the extreme events is independent on the history of the process, and just some multi-faced dice, thrown in continuous time, decides if an event takes place or not). From the point of view of statistical physics, the Poisson process constitutes a trivial fixed-point solution of the renormalization equations describing the *thinning* or decimation performed in event occurrence when the size threshold is raised[19,20]. For event occurrence on a large spatial scale, as for instance worldwide earthquakes, there is a second reason to expect exponential inter-event-time distributions: the pooled output of several time processes (i.e., China seismicity, superimposed to Japan seismicity, etc...) tends to a Poisson process if the processes are independent[21].

It is therefore surprising not only that the scaling function f is not exponential, but also that a non-exponential scaling function exists. In the case of earthquakes[4] (and fractures[6,7]) f is approximated by the so called gamma distribution, with parameters γ and a,

$$f(x) = \frac{1}{a\Gamma(\gamma, m/a)} \left(\frac{a}{x}\right)^{1-\gamma} e^{-x/a}, \ \text{ for } x = R_s\tau \geq m, \tag{3}$$

where $\Gamma(\gamma, m/a)$ is the complement of the incomplete gamma function (not normalized), $\Gamma(\gamma, u) \equiv \int_u^\infty u^{\gamma-1}e^{-u}du$. The cutoff value m is not considered a free parameter but fixed and the scale parameter a is not independent but can be obtained from the value of γ and m taking into account that $\langle x \rangle = \langle R_s\tau \rangle = \int_m^\infty xf(x)dx = 1$ (using that R_s is the inverse of the mean inter-event time). For stationary seismicity, as well as from fracture and nanofracture experiments, the shape parameter γ turns out to be close to 0.7, see Refs. 4, 6, 7. The reason to disregard $x-$values below m is due, on the one hand, to the incompleteness of seismic catalogs on the shortest time scales and to the existence errors in the determination of the inter-event times when these are small, and on the other hand to the breakdown of the stationarity condition in those short time scales by small aftershock sequences.

The usual way to establish the validity of a scaling law such as Eq. (1) is by plotting the different rescaled quantities together (in our case inter-event-time distributions for different thresholds) and judge visually if they collapse onto a single

curve or not. It would be nice if one could put some numbers into the quality of the scaling and the fit of the scaling function and test their limits of validity. Let us note that Kagan has argued that one of the reasons because theoretical physics has failed not only to predict but to explain earthquake occurrence is due to the poor use of statistics by the researchers in the field[22]. Indeed, "the quality of current earthquake data statistical analysis is low. Since little or no study of random and systematic errors is performed, most published statistical results are artifacts." We believe this criticism has applicability beyond the case of statistical seismology.

In this paper we will first use the Kolmogorov-Smirnov two-sample test in order to evaluate the fulfillment of the inter-event-time scaling law (1) in Southern California seismicity. Next, the goodness of the fit of the scaling function (3) to the rescaled inter-event-time densities will be tested by adapting the procedure introduced by Clauset *et al.*[23], consisting in maximum likelihood estimation of parameters, Kolmogorov-Smirnov one-sample statistic evaluation, and Monte Carlo simulation of the inter-event times in order to compute the distribution of the statistic.

2. Data

The seismological data used will be the Southern California waveform cross-correlation catalog of Shearer *et al.*[24] (for which, as far as the author knows, no study has published plain inter-event-time distributions $D_s(\tau)$, nevertheless, see also Refs. 25, 26). The catalog spans the years 1984-2002 (included), containing 77034 earthquakes with magnitude $M \geq 2$. Notice that we will use M as a measure of size, although by the Gutenberg-Richter law it is not power-law distributed but exponentially distributed[27]. In order to recover a power-law distribution one has to deal with the seismic moment, or the energy, which are exponential functions of the magnitude.

We will concentrate in earthquake occurrence under stationary conditions. It is well know that earthquakes trigger more earthquakes with a rate that changes in time following the Omori law[27]. In general, this breaks stationarity, as the rate of occurrence is not constant in time; however, at (relatively) large scales the resulting superposition of time-varying rates yields a constant rate, as it happens in worldwide seismic occurrence, and also for Southern California in certain time periods in which the largest earthquakes do not occur, see Fig. 1 of Ref. 28. Precisely for this reason, inter-event-times for stationary seismicity are more reliable than for non-stationary periods, as the large earthquakes present in the latter case prevent the detection of the small ones[29], which has dramatic consequences in the computation of the inter-event times.

The stationary time periods under consideration in this paper are (refining those in Ref. 28, following Ref. 30): $1984 - 1986.5$, $1990.3 - 1992.1$, $1994.6 - 1995.6$, $1996.1 - 1996.5$, $1997 - 1997.6$, $1997.75 - 1998.15$, $1998.25 - 1999.35$, $2000.55 - 2000.8$, $2000.9 - 2001.25$, $2001.6 - 2002$, $2002.5 - 2003$, where time is measured in years, 1

4   *Álvaro Corral*

year = 365.25 days (every 4 years an integer value in years corresponds to the true starting of the year) .

3. Testing Scaling

A simple way to quantify the validity of the scaling hypothesis in probability distributions can be obtained from the two-sample Kolmogorov-Smirnov (KS) test, which compares two empirical distributions. The procedure begins with the calculation of the maximum difference, in absolute value, between the rescaled cumulative distributions of the two data sets, i.e.,

$$d_{kl} \equiv \max_{\forall x} |P_k(x) - P_l(x)|; \tag{4}$$

as we have more than two data sets, we label them with indices k and l. The empirical cumulative distribution functions $P_k(x)$ are calculated as the fraction of observations in data set k below value x, and constitute an estimation of the theoretical cumulative distribution function, $F_k(x) \equiv \text{Prob[variable } < x] = \int_m^x D_k(x) dx$.

Obviously, the difference d_{kl} is randomly distributed, and therefore we can refer to it as a statistic. The key element of the KS test is that when the data sets k and l come indeed from the same underlying distribution $F(x) \equiv F_k(x) = F_l(x)$, the distribution of the KS statistic d_{kl} turns out to be independent on the form of $F(x)$ and can be easily computed. Therefore, the resulting value of d_{kl} can be considered as small or large by comparison with its theoretical distribution. Under the null hypothesis that both data sets come from the same distribution, the probability that the KS statistic is larger than the obtained empirical value d_{kl} gives the so called $p-$value, which constitutes the probability of making an error if the null hypothesis is rejected. The formulas for the probability distribution of d_{kl} are simple enough and are given by Press *et al.*[31], depending only on the number of data N in each of the sets; so, approximately, for large N_e,

$$p = \text{Prob [KS statistic } > d \text{]} = Q([\sqrt{N_e} + 0.12 + 0.11/\sqrt{N_e}]d), \tag{5}$$

with Q a decreasing function taking values between 1 and 0 (see Ref. 31) and N_e an effective number of data points (the "reduced" number, or one half of the harmonic mean of the number of data). Nevertheless, in order to calculate the $p-$value it is simpler to use the numerical routines provided in the same reference[31] (in particular, the routine called `probks`).

Notice that we have to compare the distributions of seismicity for $M \geq M_k$ and $M \geq M_l$ after rescaling, i.e., as a function of $R_k\tau$ and $R_l\tau$, respectively (otherwise, without rescaling, the distributions cannot be the same). In order to do that, for each data set, we first calculate the mean value of the inter-event time, $\langle\tau\rangle_k = R_k^{-1}$, and then, we disregard inter-event time values such that $R_k\tau < m$. The elimination of the smallest values increases the mean value of the remaining rescaled inter-event times, so, we repeat the procedure: we recalculate the mean inter-event time and rescale again the data by the new rate, disregarding those values below m.

The resulting data set has a mean value very close to one. We will assume that this procedure does not invalidate the applicability of the formulas we use for the calculation of the $p-$value.

The rescaled inter-event-time cumulative distributions for different magnitude ranges are shown at Fig. 1, ranging from $M \geq 2$ to $M \geq 4$, fixing $m = 0.01$. The scaling seems rather good, except for the case $M \geq 4$. Table 1 shows the KS statistic for each pair of distributions, as well as the corresponding $p-$values. Due to their high values (in all cases larger than 0.18 but in some others larger than 0.95), the null hypothesis cannot be rejected and each pair of data sets are compatible with the same underlying distribution, and therefore we have to agree with the scaling hypothesis (within statistical significance). Let us note that the $p-$value, being itself originated by a random set, is a random quantity (when different data sets are considered), and it turns out that the distribution of p is uniform, between 0 and 1. So, there is no reason to prefer $p = 0.9$ in front of $p = 0.2$. Only small enough values of p should lead to the rejection of the null hypothesis.

The results for the same data using $m = 0.001$ (which increases N), also shown in Table 1, are again in concordance with the scaling hypothesis, being the smallest $p-$value for this case larger than 0.14. Even for $m = 10^{-4}$ all the $p-$values are above 0.2, except for some of the pairs involving the set with $M \geq 2$. The behavior of d_{kl} when m is changed, which, following Ref. 23 should be a guide to chose the value of m (although in a different type of test, see next section), is not clear in this case.

4. Testing the Scaling Function

A different statistical test regards the goodness of the fit applied to some data. For instance, we can ask whether Eq. (3) is a good approximation to the empirical scaled distributions of inter-event times. Here, we will adapt the method of Ref. 23 to the kind of distributions that we are interested in.

First, a fit has to be performed. A usual way of proceed in the case of long-tailed distributions is to minimize the squared differences between the empirical density and the theoretical density in logarithmic scale; however, this method shows some problems and involves the arbitrary estimation of the density; other problems arises if one fits the cumulative distribution[23]. In contrast, maximum likelihood estimation avoids these difficulties by working directly with the "raw" data.

In order to be more general, let us consider the distribution given by the probability density,

$$D(x) = \frac{\delta}{a\Gamma(\gamma/\delta, (m/a)^{\delta})} \left(\frac{a}{x}\right)^{1-\gamma} e^{-(x/a)^{\delta}}, \text{ for } x \geq m, \qquad (6)$$

which constitutes the so called generalized gamma distribution, with shape parameters γ and δ and scale parameter a. We consider γ and δ greater than zero, the opposite case can be considered as well but then the function Γ has to be replaced by its complementary function (and multiplied by -1, as $\delta < 0$). The cutoff value m

6   *Álvaro Corral*

could be fixed to zero, but, as we have mentioned, for our data it is convenient to consider $m > 0$.

The $n-$th moment of the distribution is given by

$$\langle x^n \rangle = a^n \frac{\Gamma\left(\frac{\gamma+n}{\delta}, \frac{m^\delta}{a^\delta}\right)}{\Gamma\left(\frac{\gamma}{\delta}, \frac{m^\delta}{a^\delta}\right)}, \tag{7}$$

for $\gamma > 0$ and $\delta > 0$. Notice that a particular case is given by the scaling function $f(x)$ appearing in Eq. (1), for which $\langle x \rangle \equiv 1$, and only two of the three parameters are free; nevertheless we will not make use of that restriction for estimating the parameters.

The method of maximum likelihood estimation is based on the calculation of the likelihood function L, see Ref. 23. This is given by (or, in order to avoid dimensional problems, proportional to) the probability per unit of x^N that the data set comes from a particular distribution, given the values of its parameters i.e.,

$$L(\gamma, \delta, a) = \frac{Prob[x_1, x_2, \dots x_N | \gamma, \delta, a]}{dx_1, dx_2, \dots dx_N} \simeq \prod_{i=1}^{N} D(x_i | \gamma, \delta, a), \tag{8}$$

where N is the number of data and we make explicit the dependence of the probability density on its parameters. The last step assumes that each value x_i is independent on the rest. Naturally, this is not always the case (we know that earthquake inter-event times are correlated[32,28,33]) and then the maximum likelihood method provides an estimation of the distribution that generates the dataset in consideration but it may be that the dataset is not representative of the process we are studying (due to correlations, the phase space may not be evenly sampled).

It is more practical to work with the log-likelihood, ℓ, which is the logarithm of the likelihood; dividing also by N,

$$\ell(\gamma, \delta, a) \equiv \frac{\ln L(\gamma, \delta, a)}{N} = \frac{1}{N} \sum_{i=1}^{N} \ln D(x_i | \gamma, \delta, a), \tag{9}$$

which, notice, can be understood as a kind of estimator of the entropy of the distribution from the available data (with a missing -1 sign). In the case of the generalized gamma distribution (6) it is easy to get that

$$\ell(\gamma, \delta, a) = \ln \delta - \ln \Gamma\left(\frac{\gamma}{\delta}, \left(\frac{m}{a}\right)^\delta\right) + \gamma \ln \frac{G}{a} - \left(\frac{A(\delta)}{a}\right)^\delta, \tag{10}$$

where we have omitted a term $-\ln G$ that is independent on the parameters of the distribution, and we have introduced G as the geometric mean of the data, $\ln G \equiv (\sum \ln x_i)/N$, and $A(\delta)$ as what we may call the $\delta-$power mean, $A(\delta) \equiv \sqrt[\delta]{\sum x_i^\delta/N}$ (which, in contrast to G, depends on the value of the parameter δ; for instance, for $\delta = 1$, A is the arithmetic mean, but for $\delta = -1$, A is the harmonic mean).

The best estimate of the parameters would be that that maximizes the likelihood, or, equivalently, the log-likelihood. The previous expression is too complicated to

be maximized analytically, and it is too complicated to differentiate even (in fact, we would need to compute the derivative of the incomplete gamma function). So, we will perform a direct numerical maximization (in particular we will use the numerical routine `amoeba` from Ref. 31; the function Γ can be computed from the same source using routines `gammq` and `gammln`).

Fixing $\delta \equiv 1$, from which we recover Eq. (3) as a model of the distribution (which yields only one free parameter and has the advantage of being compatible with a Poisson process in the limit of long times), the resulting values of the parameters γ and a, obtained from maximum likelihood estimation, are given in Table 2. In all cases, except for $M \geq 4$, and if the cutoff m is not too small, the values of γ are close to 0.7.

Once we have obtained the estimators of the parameters, we can ask about their meaning. Maximum-likelihood estimation does not mean that it is likely that the data comes from the proposed theoretical distribution, with those parameters. In fact, maximum likelihood can be minimum unlikelihood, i.e., we are taking the less bad option among those provided by the *a priori* assumed probability model. In order to address this issue it is necessary to perform a goodness-of-fit test.

Following Ref. 23 we can employ again the Kolmogorov-Smirnov test, this time for one sample. The KS statistic is, similarly as before

$$d \equiv \max_{\forall x}|P(x) - F(x)| \tag{11}$$

where $P(x)$ is the empirical cumulative distribution of the data, defined in the previous section, and $F(x)$ is the theoretical proposal. For the distribution of Eq. (6),

$$F(x) \equiv \int_m^x D(x)dx = 1 - \frac{\Gamma(\gamma/\delta, (x/a)^\delta)}{\Gamma(\gamma/\delta, (m/a)^\delta)} \;\; \text{for } x \geq m. \tag{12}$$

The resulting values of d for our problem are also shown in Table 2. Now we can apply the recipe of Clauset *et al.* in order to select the most appropriate value of the cutoff m, which consists in selecting the value which minimizes d. Comparing between 0.003, 0.01, and 0.03, it seems clear that we should chose $m = 0.01$.

At this point we could proceed as in the previous section, using the formulas for the distribution of d. However, that only would be right if we were not estimating $F(x)$ from the data (if we were comparing with a theory free of parameters for instance). In order to know the distribution of the statistic d when the data are generated by the model with the parameters obtained by maximum likelihood estimation, we will use Monte Carlo simulations. Indeed, generating data from the theoretical distribution, we can repeat the whole process to obtain the statistical behavior of d when the null hypothesis is true (when the data come from the proposed theoretical distribution), and we can do it many times, in order to get significant statistics.

Schematically, the process for the calculation of p consists of the multiple iteration of the following steps:

(1) Simulate synthetic data s from the distribution given by Eq. (3) using the parameters γ and a obtained before for the empirical data
(2) Estimate the parameters γ_s and a_s by fitting the synthetic data s to Eq. (3) (proceeding in the same way as described above for the empirical data, see Eq. (10) and so on).
(3) Evaluate the KS statistic for the distribution of synthetic data s [generated in (1) with parameters γ and a] and the theoretical distribution with parameters γ_s and a_s [calculated in (2)], i.e., $d_s = \max_{\forall x} |P_s(x|\gamma, a) - F(x|\gamma_s, a_s)|$

We will obtain synthetic inter-event times from the gamma distribution by generating a table of the cumulative distribution. As the probability of an event has to be the same independently on the random variable we assign to that event, then, $u = F(x)$, where u is a uniform random number between zero and one and is also its own cumulative distribution. We can calculate numerically the function $F(x)$ (thanks to the numerical recipes `gammq` and `gammln`[31]), but are unable to calculate its inverse (at least, in a reasonable computer time), so we will tabulate the values of $F(x)$, for selected values of x in log scale (this is to deal with the multiple time scales that appear in the process, described by Eq. (3) or (12) when $\gamma < 1$ and $m \ll 1$). To be concrete, $x(k) = me^{\alpha k}$, where $k = 0, 1, \ldots$ and α is just a constant. Then, when a uniform value u is generated we can obtain the corresponding value of x by looking at the table and interpolating (or extrapolating) using the closest values of $u(k) = F(x(k))$.

For the case of our interest, the $p-$values calculated in this way, using 1000 randomly generated samples (which yield an uncertainty of about 3 % in p), are included in Table 2. Taking $m = 0.01$ (the value arising by the application of Clauset *et al.* recommendation), we cannot reject the hypothesis that the data set comes from the theoretical distribution with maximum likelihood parameters, except for $M \geq 2$, which yields $p = 0.032$, which is beyond the usual onset of acceptance of the null hypothesis, $p = 0.05$. Figure 2 illustrates the reason of the rejection. Indeed, although the theoretical distribution is very close to the empirical one, the difference is large enough for the high number of data involved. Although Eq. (5) is not valid in this case, we can use it as an approximation and see how for large N the statistic d scales as $1/\sqrt{N}$ ($N_e = N$ here). As the mode of the distribution Q in Eq. (5) is around 0.735 and practically all the probability is contained below 2, this means that we can expect $d < 2/\sqrt{N}$. So, for large N, d tends to zero, and the KS test is able to detect any small difference between the proposed theoretical distribution and the "true" distribution. This means that the test is not adequate if we are just interested to find an approximation to the true distribution, as only the "true" distribution is not rejected for a sufficient number of data. For comparison, we show in Table 3 the results for an exponential scaling function, which is clearly rejected except for $M \geq 4$.

5. Discussion

As another alternative, note that we have tested separately the validity of the scaling law and the adequacy of the scaling function given by Eq. (3). We could take advantage of the scaling behavior to fit and test the goodness of fit of the scaling function. For instance, we could combine all rescaled data sets (for all values of the minimum magnitude) and proceed as in Sect. 3 for this combined data set. The problem is that, by virtue of the Gutenberg-Richter law, when the minimum magnitude is raised in one unit, the number of events decreases by a factor 10, and therefore, data sets with large minimum magnitudes are under-represented. Perhaps we could just truncate the samples in order that all of them had the same number of data, but that would lead to a tremendous wasting of information.

The surprising character of the scaling law (1) when the scaling function is not exponential has lead to some criticisms by Molchan[34] and Saichev and Sornette[35]. The latter authors propose that, for the so called ETAS model, the scaling law is not valid, and one has a very slow variation of the inter-event-time distribution when the magnitude threshold is raised. Although we have tested that the scaling law is consistent with the data within statistical significance, this does not mean that we should reject Saichev and Sornette result. Nevertheless, the simplicity of the scaling hypothesis makes it the most adequate model for seismicity, at least as a null model to contrast with other hypothesis. On the other hand, other seismicity models have been recently proposed, which, in contrast to the ETAS model, are fully scale invariant and one would expect that are characterized by scaling inter-event-time distributions[36,37,38,39,40].

Saichev and Sornette also provide a pseudo-scaling function to which inter-event-time distributions can be fit[34,35]. In principle, the very same procedure used in our paper can be applied directly in order to fit the parameters of the Saichev-Sornette function and test the goodness-of-fit of the outcome. It is expected that the use of this new function, which has more parameters than Eq. (3) and models better the left tail of the distribution, could lead to the reduction of the cutoff m above which the functions are fit. We leave this task for future research.

Acknowledgements

References

1. P. Bak, K. Christensen, L. Danon, and T. Scanlon. Unified scaling law for earthquakes. *Phys. Rev. Lett.* 88, 178501, 2002.
2. A. Corral. Local distributions and rate fluctuations in a unified scaling law for earthquakes. *Phys. Rev. E* 68, 035102, 2003.
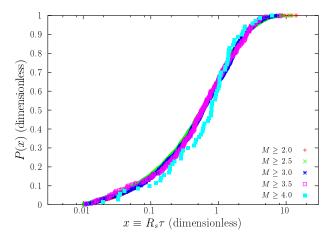
10   *Álvaro Corral*

Fig. 1.   Rescaled inter-event-time cumulative distributions for Southern California stationary seismicity, fixing minimum $x-$value $m = 0.01$. The collapse of the distributions is an indication of scaling, in agreement with the results of the KS tests performed.

3. A. Corral. Universal local versus unified global scaling laws in the statistics of seismicity. *Physica A* 340, 590–597, 2004.
4. A. Corral. Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Phys. Rev. Lett.* 92, 108501, 2004.
5. A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin. Long-term memory: a natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Phys. Rev. Lett.* 94, 048701, 2005.
6. J. Åström, P. C. F. Di Stefano, F. Pröbst, L. Stodolsky, J. Timonen, C. Bucci, S. Cooper, C. Cozzini, F.v. Feilitzsch, H. Kraus, J. Marchese, O. Meier, U. Nagel, Y. Ramachers, W. Seidel, M. Sisti, S. Uchaikin, and L. Zerle. Fracture processes observed with a cryogenic detector. *Phys. Lett. A* 356, 262–266, 2006.
7. J. Davidsen, S. Stanchits, and G. Dresen. Scaling and universality in rock fracture. *Phys. Rev. Lett.* 98, 125502, 2007.
8. E. L. Geist and T. Parsons. Distribution of tsunami interevent times. *Geophys. Res. Lett.* 35, L02612, 2008.
9. M. Baiesi, M. Paczuski, and A. L. Stella. Intensity thresholds and the statistics of the temporal occurrence of solar flares. *Phys. Rev. Lett.* 96, 051103, 2006.
10. A. Corral, L. Telesca, and R. Lasaponara. Scaling and correlations in the dynamics of forest-fire occurrence. *Phys. Rev. E* 77, 016101, 2008.
11. T. Nakamura, K. Kiyono, K. Yoshiuchi, R. Nakahara, Z. R. Struzik, and Y. Yamamoto. Universal scaling law in human behavioral organization. *Phys. Rev. Lett.* 99, 138103, 2007.
12. I. Osorio, M. G. Frei, D. Sornette, J. Milton, and Y.-C. Lai. Epileptic seizures: Quakes of the brain? *http://arxiv.org*, 0712.3929, 2007.
13. A. Corral, R. Ferrer-i-Cancho, G. Boleda, and A. Díaz-Guilera. Universal complex structures in written language. *http://arxiv.org*, 0901.2924, 2009.
14. A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature* 435, 207–211, 2005.
15. K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stanley. Scaling and memory
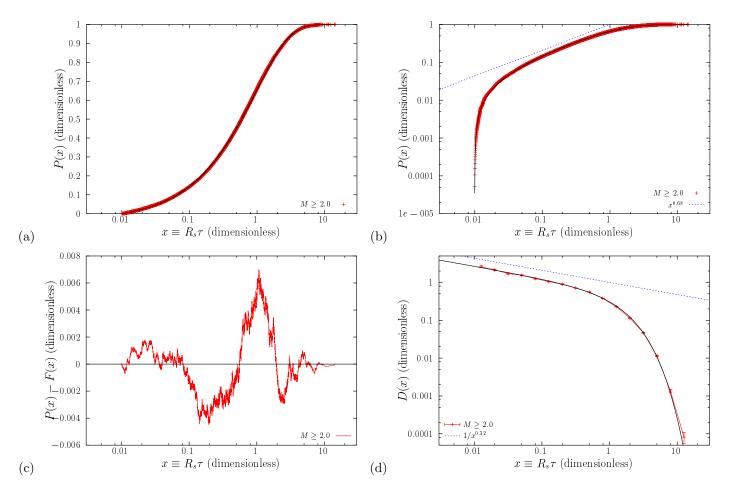
Fig. 2. (a) Rescaled inter-event-time cumulative distribution for Southern California stationary seismicity with $M \geq 2$, fixing minimum $x-$value $m = 0.01$, together with the fit obtained by maximum likelihood estimation. The $p-$value corresponding to the KS statistic, determined by Monte Carlo simulation turns out to be as small as 0.032, although the fit is visually acceptable. (b) Same as before, in log-log scale, together with a pure power law with the same exponent. (c) Difference between the distribution and its fit, $P(x) - F(x)$, which yields the KS statistic when its absolute value is maximized. (d) The corresponding probability density, estimated with 5 bins per decade, for comparison. Also the best fit and a pure power law are shown.

in volatility return intervals in financial markets. *Proc. Natl. Acad. Sci. USA* 102, 9424–9428, 2005.

16. U. Harder and M. Paczuski. Correlated dynamics in human printing behavior. *Physica A* 361, 329–336, 2006.

17. K.-I. Goh and A.-L. Barabási. Burstiness and memory in complex systems. *Europhys. Lett.* 81, 48002, 2008.

18. H. E. Stanley. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev. Mod. Phys.* 71, S358–S366, 1999.

19. A. Corral. Renormalization-group transformations and correlations of seismicity. *Phys. Rev. Lett.* 95, 028501, 2005.
20. A. Corral. Point-occurrence self-similarity in crackling-noise systems and in other complex systems. *J. Stat. Mech.* 2009, P01022, 2009.
21. D. R. Cox and P. A. W. Lewis. *The Statistical Analysis of Series of Events*. Methuen, London, 1966.
22. Y. Y. Kagan. Why does theoretical physics fail to explain and predict earthquake occurrence? In P. Bhattacharyya and B. K. Chakrabarti, editors, *Modelling Critical and Catastrophic Phenomena in Geoscience*, Lecture Notes in Physics, 705, pages 303–359. Springer, Berlin, 2006.
23. A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, accepted, 2009.
24. P. Shearer, E. Hauksson, and G. Lin. Southern California hypocenter relocation with waveform cross- correlation, part 2: Results using source-specific station terms and cluster analysis. *Bull. Seismol. Soc. Am.* 95, 904–915, 2005. http://www.data.scec.org/ftp/catalogs/SHLK/, file SHLK_1_02.txt.
25. A. Corral. Universal earthquake-occurrence jumps, correlations with time, and anomalous diffusion. *Phys. Rev. Lett.* 97, 178501, 2006.
26. J. Davidsen, P. Grassberger, and M. Paczuski. Earthquake recurrence as a record breaking process. *Geophys. Res. Lett.* 33, L11304, 2006.
27. H. Kanamori and E. E. Brodsky. The physics of earthquakes. *Rep. Prog. Phys.* 67, 1429–1496, 2004.
28. A. Corral. Dependence of earthquake recurrence times and independence of magnitudes on seismicity history. *Tectonophys.* 424, 177–193, 2006.
29. A. Helmstetter, Y. Y. Kagan, and D. D. Jackson. Comparison of short-term and time-independent earthquake forecast models for Southern California. *Bull. Seismol. Soc. Am.* 96, 90–106, 2006.
30. A. Corral. Structure of earthquake occurrence in space, time and magnitude. *Terra Nova* 19, 337–343, 2007.
31. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in FORTRAN*. Cambridge University Press, Cambridge, 2nd edition, 1992.
32. V. N. Livina, S. Havlin, and A. Bunde. Memory in the occurrence of earthquakes. *Phys. Rev. Lett.* 95, 208501, 2005.
33. S. Lennartz, V. N. Livina, A. Bunde, and S. Havlin. Long-term memory in earthquakes and the distribution of interoccurrence times. *Europhys. Lett.* 81, 69001, 2008.
34. G. Molchan. Interevent time distribution in seismicity: a theoretical approach. *Pure Appl. Geophys.* 162, 1135–1150, 2005.
35. A. Saichev and D. Sornette. "Universal" distribution of interearthquake times explained. *Phys. Rev. Lett.* 97, 078501, 2006.
36. D. Vere-Jones. A class of self-similar random measure. *Adv. Appl. Prob.* 37, 908–914, 2005.
37. A. Saichev and D. Sornette. Vere-Jones's self-similar branching model. *Phys. Rev. E* 72, 056122, 2005.
38. E. Lippiello, C. Godano, and L. de Arcangelis. Dynamical scaling in branching models for seismicity. *Phys. Rev. Lett.* 98, 098501, 2007.
39. D. L. Turcotte, J. R. Holliday, and J. B. Rundle. *Geophys. Res. Lett.* 34, L12303, 2007.
40. J. R. Holliday, D. L. Turcotte, and J. B. Rundle. *Physica A* 387, 933–943, 2008.

Table 1. KS statistic d (below the diagonal) and corresponding $p-$value (above diagonal, in percentage) for rescaled Southern-California stationary–seismicity inter-event-time distributions, with lower cutoffs $m = 0.01$ (top) and $m = 0.001$ (bottom). The scaling hypothesis cannot be rejected.

$m = 0.01$	N	$M \geq 2.0$	$M \geq 2.5$	$M \geq 3.0$	$M \geq 3.5$	$M \geq 4.0$
$M \geq 2.0$	18870	-	26.1%	36.0%	97.3%	21.3%
$M \geq 2.5$	4953	0.016	-	63.2%	86.3%	18.3%
$M \geq 3.0$	1184	0.028	0.024	-	95.7%	22.8%
$M \geq 3.5$	309	0.028	0.035	0.032	-	20.8%
$M \geq 4.0$	70	0.125	0.129	0.126	0.138	-

$m = 0.001$	N	$M \geq 2.0$	$M \geq 2.5$	$M \geq 3.0$	$M \geq 3.5$	$M \geq 4.0$
$M \geq 2.0$	19821	-	41.6%	43.8%	14.3%	31.9%
$M \geq 2.5$	5187	0.014	-	57.1%	29.6%	31.1%
$M \geq 3.0$	1268	0.025	0.024	-	68.0%	33.5%
$M \geq 3.5$	340	0.062	0.054	0.044	-	28.1%
$M \geq 4.0$	76	0.108	0.110	0.110	0.124	-

Table 2. Maximum likelihood parameters γ and a, KS statistic d and corresponding $p-$value (in percentage, determined by Monte Carlo simulation) for rescaled Southern-California stationary-seismicity inter-event-time distributions, using several values of the minimum value m.

$m = 0.03$	N	γ	a	d	$p-$value
$M \geq 2.0$	18009	0.68	1.35	0.008	1.2%
$M \geq 2.5$	4669	0.67	1.38	0.007	84.0%
$M \geq 3.0$	1122	0.73	1.29	0.021	25.7%
$M \geq 3.5$	287	0.79	1.22	0.034	55.7%
$M \geq 4.0$	69	0.89	1.07	0.089	16.3%

$m = 0.01$	N	γ	a	d	$p-$value
$M \geq 2.0$	18870	0.68	1.41	0.007	3.2%
$M \geq 2.5$	4953	0.64	1.50	0.009	37.3%
$M \geq 3.0$	1184	0.69	1.41	0.021	21.8%
$M \geq 3.5$	309	0.67	1.45	0.029	73.6%
$M \geq 4.0$	70	0.95	1.05	0.082	29.3%

$m = 0.003$	N	γ	a	d	$p-$value
$M \geq 2.0$	19466	0.65	1.51	0.009	0.0%
$M \geq 2.5$	5102	0.62	1.57	0.011	11.2%
$M \geq 3.0$	1247	0.59	1.65	0.029	1.4%
$M \geq 3.5$	328	0.56	1.74	0.046	9.0%
$M \geq 4.0$	74	0.72	1.37	0.103	5.8%

Table 3.   Same as the previous table, for the exponential distribution ($\gamma \equiv 1$).

$m = 0.01$	N	γ	a	d	$p-$value
$M \geq 2.0$	18870	1	0.99	0.072	0.0%
$M \geq 2.5$	4953	1	0.99	0.084	0.0%
$M \geq 3.0$	1184	1	1.00	0.077	0.0%
$M \geq 3.5$	309	1	1.00	0.079	0.4%
$M \geq 4.0$	70	1	1.00	0.077	59.5%