

SCIENTIFIC REPORTS

OPEN

Deviation from power law of the global seismic moment distribution

Isabel Serra¹ & Álvaro Corral^{1,2}

Received: 08 August 2016
Accepted: 30 November 2016
Published: 05 January 2017

The distribution of seismic moment is of capital interest to evaluate earthquake hazard, in particular regarding the most extreme events. We make use of likelihood-ratio tests to compare the simple Gutenberg-Richter power-law (PL) distribution with two statistical models that incorporate an exponential tail, the so-called tapered Gutenberg-Richter (Tap) and the truncated gamma, when fitted to the global CMT earthquake catalog. Although the Tap distribution does not introduce any significant improvement of fit respect the PL, the truncated gamma does. Simulated samples of this distribution, with parameters $\beta = 0.68$ and $m_c = 9.15$ and reshuffled in order to mimic the time occurrence of the order statistics of the empirical data, are able to explain the temporal heterogeneity of global seismicity both before and after the great Sumatra-Andaman earthquake of 2004.

The Gutenberg-Richter (GR) law is not only of fundamental importance in statistical seismology¹ but also a cornerstone of non-linear geophysics² and complex-systems science³. It simply states that, for a given region, the magnitudes of earthquakes follow an exponential probability distribution. As the (scalar) seismic moment is an exponential function of magnitude, when the GR law is expressed in terms of the former variable, it translates into a power-law distribution^{4,5}, i.e.,

$$f(M) \propto \frac{1}{M^{1+\beta}}, \quad (1)$$

with M seismic moment, $f(M)$ its probability density, (fulfilling $\int_{\forall M} f(M) dM = 1$), the sign “ \propto ” denoting proportionality, and the exponent $1 + \beta$ taking values close to 1.65. This simple description provides rather good fits of available data in many cases^{6–9}, with, remarkably, only one free parameter, β . A totally equivalent characterization of the distribution uses the survivor function (or complementary cumulative distribution), defined as

$$S(M) = \int_M^\infty f(M') dM', \quad (2)$$

for which the GR power law takes the form $S(M) \propto 1/M^\beta$.

The power-law distribution has important physical implications, as it suggests an origin from a critical branching process or a self-organized-critical state^{3,10,11}. Nevertheless, it presents also some conceptual difficulties, due to the fact that the mean value $\langle M \rangle$ provided by the distribution turns out to be infinite^{4,12}. These elementary considerations imply that the GR law cannot be naively extended to arbitrarily large values of M , and one needs to introduce additional parameters to describe the tail of the distribution, coming presumably from finite-size effects. However, a big problem is that the change from power law to a faster decay seems to take place at the highest values of M that have been observed, for which the statistics are very poor¹³.

Kagan⁷ has enumerated the requirements that an extension of the GR law should fulfil; in particular, he considered, among other: (i) the so called tapered (Tap) Gutenberg-Richter distribution (also called Kagan distribution¹⁴), with a survivor function given by

$$S_{tap}(M) \propto e^{-M/\theta} / M^\beta \quad (3)$$

and (ii) the (left-) truncated gamma (TrG) distribution, for which the density is

$$f_{trg}(M) \propto e^{-M/\theta} / M^{1+\beta}. \quad (4)$$

¹Centre de Recerca Matemàtica, Edifici C, Campus Bellaterra, E-08193 Barcelona, Spain. ²Departament de Matemàtiques, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Barcelona, Spain. Correspondence and requests for materials should be addressed to I.S. (email: iserra@crm.cat) or Á.C. (email: acorral@crm.cat)

Note that both expressions have essentially the same functional form, but the former refers to the survivor function and the later to the density. As $f(M) = -dS(M)/dM$, differentiation of $S_{tap}(M)$ in (i) shows the difference between both distributions. In both cases, parameter θ represents a crossover value of seismic moment, signalling a transition from power law to exponential decay; so, θ gives the scale of the finite-size effects on the seismic moment. The corresponding value of (moment) magnitude (sometimes called corner magnitude) can be obtained from $m_c = \frac{2}{3}(\log_{10} \theta - 9.1)$, when the seismic moment is measured in $N \cdot m^{15,16}$.

Kagan⁷ also argues that available seismic catalogs do not allow the reliable estimation of θ , except in the global case (or for large subsets of this case), in particular, he recommends the use of the centroid moment tensor (CMT) catalog^{17,18}. From his analysis of global seismicity, and comparing the values of the likelihoods, Kagan⁷ concludes that the tapered GR distribution gives a slightly better fit than the truncated gamma distribution, for which in addition the estimation procedure is more involving. In any case, the β -value seems to be universal (at variance with θ), see also refs 9, 19 and 20.

Nevertheless, the data analyzed by Kagan⁷, from 1977 to 1999, comprises a period of relatively low global seismic activity, with no event above magnitude 8.5; in contrast, the period 1950–1965 witnessed 7 of such events²¹. Starting with the great Sumatra–Andaman earthquake of 2004, and following since then with 5 more earthquakes with $m \geq 8.5$ (up to the time of submitting this article), the current period seems to correspond to the past higher levels of activity.

Main *et al.*²² and Bell *et al.*²³ have re-examined the problem of the seismic moment distribution including recent global data (shallow events only). Using a Bayesian information criterion (BIC), Bell *et al.*²³ compare the plain GR power law with the tapered GR distribution, and conclude that, although the tapered GR gives a significantly better fit before the 2004 Sumatra event, the occurrence of this changes the balance of the BIC statistics, making the GR power law more suitable; that is, the power law is more parsimonious, or simply, is enough for describing global shallow seismicity when the recent mega-earthquakes are included in the data. Similar results have been published in ref. 24.

In the present paper we revisit the problem with more recent data, including also the truncated gamma distribution, using other statistical tools, and reaching somewhat different conclusions: when data includes periods of high seismic activity, indeed, the tapered GR distribution does not introduce any significant improvement with respect to the power law²³, but the truncated gamma does.

Data, Models and Maximum Likelihood Estimation

As Main *et al.*²² and Bell *et al.*²³, we analyze the global CMT catalog^{17,18}, in our case for the period between January 1, 1977 and October 31, 2013, with the values of the seismic moment converted into $N \cdot m$ ($1 \text{ dyn} \cdot \text{cm} = 10^{-7} N \cdot m$). We restrict to shallow events (depth < 70 km) and, in order to avoid incompleteness, to magnitude $m > 5.75$ (equivalent to $M > 5.3 \cdot 10^{17} N \cdot m$), as Main *et al.*²² and Bell *et al.*²³. This yields 6150 events.

As statistical tools, we use maximum likelihood estimation (MLE) for fitting, and likelihood-ratio (LR) tests for comparison of different fits. Maximum likelihood estimation is the best-accepted method in order to fit probability distributions, as it yields estimators which are invariant under re-parameterizations, and which are asymptotically efficient for regular models, in particular for exponential families²⁵ (the three models under consideration here are regular, and the PL and the TrG belong to the exponential family). When maximum likelihood is used under a wrong model, what one finds is the closest model to the true distribution in terms of the Kullback-Leibler divergence²⁵.

Model selection tests based on the likelihood ratio have the advantage that the ratio is invariant with respect to changes of variables (if these are one-to-one²⁵). Moreover, for comparing the fit of models in pairs, LR test is preferable in front of the computation of differences in BIC or AIC (Akaike information criterion), as the test relies on the fact that the distribution of the LR is known, under a suitable null hypothesis, which provides a significance level (or level of risk) to its value. So, LR tests constitute probability-based model selection (in contrast to BIC and AIC). But note that the log-likelihood-ratio is equal to the difference of BIC or AIC when the number of parameters of the two models is the same.

In order to perform MLE it is necessary to specify the densities of the distributions, including the normalization factors. In our case, all distributions are defined for M above the completeness threshold a , i.e., for $M > a$, being zero otherwise (as mentioned above, a is fixed to $5.3 \times 10^{17} N \cdot m$). For the power-law (PL) distribution (which yields the GR law for the distribution of M) Eq. (1) reads

$$f_{pl}(M; \beta) = \frac{\beta}{a} \left(\frac{a}{M} \right)^{1+\beta}, \quad (5)$$

with $\beta > 0$. For the tapered Gutenberg–Richter,

$$f_{tap}(M; \beta, \theta) = \left[\frac{\beta}{a} \left(\frac{a}{M} \right)^{1+\beta} + \frac{1}{\theta} \left(\frac{a}{M} \right)^\beta \right] e^{-(M-a)/\theta}, \quad (6)$$

with $\beta > 0$ and $\theta > 0$. And for the left-truncated (and extended to $\beta > 0$) gamma distribution;

$$f_{trg}(M; \beta, \theta) = \frac{1}{\theta \Gamma(-\beta, a/\theta)} \left(\frac{\theta}{M} \right)^{1+\beta} e^{-M/\theta}, \quad (7)$$

with $-\infty < \beta < \infty$ and $\theta > 0$, and with $\Gamma(\gamma, z) = \int_z^\infty x^{\gamma-1} e^{-x} dx$ the upper incomplete gamma function, defined for $z > 0$ when $\gamma < 0$.

		$\hat{\beta}$	$\hat{\theta}$ (N·m)	\hat{m}_c	l (M in N·m)	$l - l_{pl}$	$l_{trg} - l_{tap}$
PL	MLE	0.685	∞	∞	−268466.609		
	s.e.	0.009					
Tap	MLE	0.684	3.3×10^{22}	8.94	−268465.315	1.294	
	s.e.	0.009	2.6×10^{22}	0.23			
TrG	MLE	0.681	6.7×10^{22}	9.15	−268464.844	1.765	0.471
	s.e.	0.009	6.6×10^{22}	0.27			

Table 1. Maximum likelihood estimation of the parameters with their standard errors (s.e.) and maximum value of the log-likelihood function, $l = \ln \hat{L}$ when the PL, Tap, and TrG distributions are fitted to the seismic moment of shallow CMT earthquakes, using the whole data set ($N = 6150$). The standard error for $\hat{\beta}$ and $\hat{\theta}$ is computed from the Fisher information matrix and corresponds to one standard deviation of the distribution of each parameter. The standard error for \hat{m}_c is computed from that of $\hat{\theta}$ using the delta method⁴².

We summarize the parameterization of the densities as $f(M; \Theta)$, where $\Theta = \{\beta, \theta\}$ for the Tap and TrG distributions and $\Theta = \beta$ for the power law. Note that for the TrG distribution, it is clear that the exponent β is a shape parameter and θ is a scale parameter; in fact, these parameters play the same role in the Tap distribution, which turns out to be a mixture of two truncated gamma distributions, one with shape parameter β and the other with $\beta - 1$, but with common scale parameter θ . Exactly,

$$f_{tap}(M; \beta, \theta) = \left[\beta \Gamma(-\beta, a/\theta) f_{trg}(M; \beta, \theta) + \Gamma(1 - \beta, a/\theta) f_{trg}(M; \beta - 1, \theta) \right] \left(\frac{a}{\theta} \right)^\beta e^{a/\theta} \quad (8)$$

(in our case, the contribution of the second TrG will be only about 0.14%). In contrast, the power law lacks a scale parameter. In all cases the completeness threshold a is a truncation parameter, but it is kept fixed and is not a free parameter, therefore.

Other authors consider the upper truncated power-law distribution^{13,26}, given by $f(M; \beta, \theta) \propto 1/M^{1+\beta}$ for $a < M < \theta$, and zero otherwise; then θ becomes a truncation parameter. We disregard this model because such an abrupt truncation is unphysical⁷, because the occurrence of one single earthquake with size larger than the resulting value of θ invalidates the selected model, and because the fact that the support of the distribution involves the unknown parameter θ leads to a violation of the regularity conditions for which standard likelihood theory holds²⁵.

The knowledge of the probability densities allows the direct computation of the likelihood function as $L(\Theta) = \prod_{i=1}^N f(M_i; \Theta)$, where M_i are the N observational values of the seismic moment. Maximization of the likelihood function with respect the values of the parameters leads to the maximum-likelihood estimation $\hat{\Theta}$ of these parameters, with $\hat{L} = L(\hat{\Theta})$ the value of the likelihood at its maximum. Note that the independence assumption that is implicit in the expression for $L(\Theta)$ arises in fact as the maximum-entropy solution when there is no information about dependence²⁷. If the data cannot be considered independent, the MLE results will just describe a marginal distribution $f(M; \Theta)$ of the sample under consideration, and inference about the underlying population will not be possible, as the sample may be not representative of the population. In any case, the results of MLE for our three models are reported in Table 1, and an illustration of the corresponding fits is provided in the Supplementary Information (SI). Although the TrG model has the highest likelihood one has to perform a proper model comparison.

Model Comparison

A powerful method for comparison of pairs of models is the likelihood-ratio test, specially suitable when one model is nested within the other, which means that the first model is obtained as a special case of the second one. This is the case of the power-law distribution with respect to the other two distributions; indeed, the power law is nested both within the Tap and within the truncated gamma, as taking $\theta \rightarrow \infty$ in any of the two leads to the power-law distribution. This is easily seen taking into account that $S_{tap}(M) = (a/M)^\beta e^{-(M-a)/\theta}$, or just performing the limit in the expression for $f_{tap}(M)$ above. For the truncated gamma distribution, when doing the $\theta \rightarrow \infty$ limit in $f_{trg}(M)$ one needs to use that, for $\gamma < 0$, $z^\gamma/\Gamma(\gamma, z) \rightarrow -\gamma$ when $z \rightarrow 0$, see ref. 28 for $\gamma \neq -1, -2, \dots$

Given two probability distributions, 1 and 2, with 1 nested within 2, the likelihood-ratio test evaluates \hat{L}_2/\hat{L}_1 , where \hat{L}_2 is the likelihood (at maximum) of the “bigger” or “full” model (either Tap or TrG) and \hat{L}_1 corresponds to the nested or null model (power law in our case). Taking logarithms we get the log-likelihood-ratio

$$\mathcal{R} = \ln \frac{\hat{L}_2}{\hat{L}_1} = l_2 - l_1, \quad (9)$$

with $l_j = \ln \hat{L}_j = \sum_{i=1}^N \ln f_j(M_i; \hat{\Theta}_j)$, where f_j denotes the probability density function of the distribution j for every $j = 1, 2$, and the MLE corresponds to $\hat{\Theta}_1 = \hat{\beta}_1$ and $\hat{\Theta}_2 = \{\hat{\beta}_2, \hat{\theta}_2\}$. In order to compare the fit provided by the two distributions, it is necessary to characterize the distribution of \mathcal{R} .

Let n_1 and n_2 be the number of free parameters in the models 1 and 2, respectively. In general, if the models are nested, and under the null hypothesis that the data comes from the simpler model, the probability distribution of

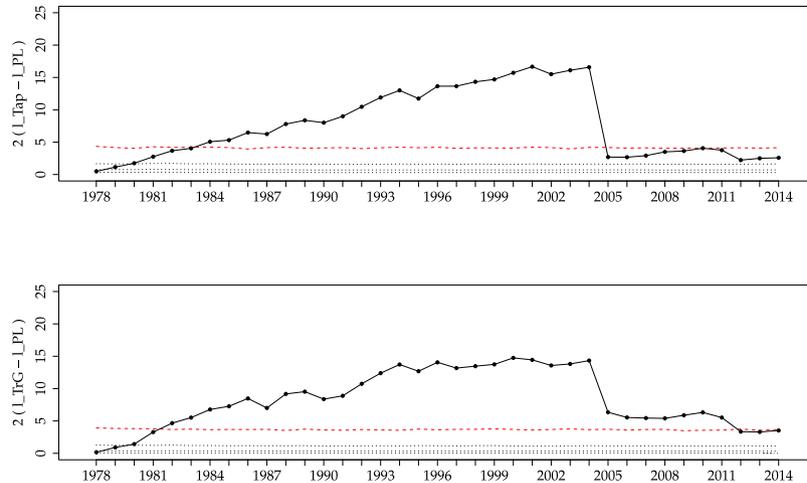


Figure 1. Results of the likelihood-ratio tests. The points (joined by lines) denote the value of the statistic $2\mathcal{R}$ for the empirical data. Lines show different percentiles of the distribution of $2\mathcal{R}$ for 10000 simulations of the power-law null hypothesis with the same number of data (dotted black lines: first, second, and third quartiles), including the critical value of the test (at level 0.05, dashed red line). The abscissa corresponds to the ending point of a time window starting always in Jan 1, 1977. Note that the year is considered a continuous variable (not a categorical variable), so, the time window ending on Dec 31, 2004 takes value $2004.99\dots \approx 2005$. (a) Tap distribution versus power law. (b) Truncated gamma versus power law.

the statistic $2\mathcal{R}$ in the limit $N \rightarrow \infty$ is a chi-squared distribution with degrees of freedom equal to $n_2 - n_1 > 0$. So, for $n_2 = n_1 + 1$,

$$2\mathcal{R} > 3.84 \quad (10)$$

with a level of risk equal to 0.05. Note that the chi-squared distribution provides a penalty for model complexity as the “range” or “scale” of the distribution is given directly by the number of the degrees of freedom. This likelihood-ratio test constitutes the best option to choose among models 1 and 2, in the sense that it has a convergence to its asymptotic distribution faster than any other test²⁹. The null and alternative hypotheses correspond to accept model 1 or 2, respectively, although the acceptance of model 1 does not imply the rejection of 2, it is simply that the “full” model 2 does not bring any significant improvement with respect the simpler model 1, which is more parsimonious.

On the other hand, when the nesting of distribution 1 within 2 takes place in such a way that the space of parameters of the former one lies within a boundary of the space of parameters of distribution 2, the approach just explained for the asymptotic distribution of $2\mathcal{R}$ is not valid^{30,31}. This happens when testing both the Tap or the TrG distributions in front of the power-law distribution, as the $\theta \rightarrow \infty$ limit of the latter corresponds to the boundary of the parameter space of the two other distributions, and then, what one should obtain for $2\mathcal{R}$ is a mixture of a chi-square and a Dirac delta function. Nevertheless, this latter result is also unapplicable in our case, as the power-law distribution does not fulfil the sufficient conditions stated in ref. 30, due to the divergence of the second moment³². This illustrates part of the difficulties of performing proper model selection when fractal-like distributions are involved³³. In order to obtain the distribution of $2\mathcal{R}$ and from there the p -values of the LR tests, we are left to the simulation of the null hypothesis. We advance that the results seem to indicate that the distribution of $2\mathcal{R}$, for high percentiles, is close to chi-square with one degree of freedom, so that Eq. (10) is approximately valid, but we lack a theoretical support for this fact.

Let us proceed, using this method, by comparing the performance of the power-law and Tap fits when applied to the global shallow seismic activity, for time windows starting always in 1977 and ending in the successive times indexed by the abscissa in Fig. 1(a) (as in ref. 23). The log-likelihood-ratio of these fits (times 2), is shown in the figure together with the critical region of the test. In agreement with Bell *et al.*²³, we find that: (i) the power-law fit can be safely rejected in front of the Tap distribution for any time window ending between 1984 and before 2004; and (ii) the results change drastically after the occurrence of the great 2004 Sumatra earthquake, for which the power law cannot be rejected at the 0.05 level. So, for parsimony reasons, the power law becomes preferable in front of the Tap distribution for time windows ending later than 2004. The fact that, for these time windows, the Tap distribution cannot be distinguished from the power law is also in agreement with previous results showing that the contour lines in the likelihood maps of the Tap distribution are highly non-symmetric and may be unbounded for smaller levels of risk^{7,24,34}.

When we compare the power-law fit with the truncated gamma, using the same test, for the same data, the results are more significant, see Fig. 1(b). The situation previous to 2004 is nearly the same, with an extremely poor performance of the power law; but after 2004, despite a big jump again in the value of the likelihood ratio, the power law remains non-acceptable, at the 0.05 level. It is only after the great Tohoku earthquake of 2011 that the p -value of the test enters slightly into the non-rejection region, but keeping values very close to the 0.05 limit.

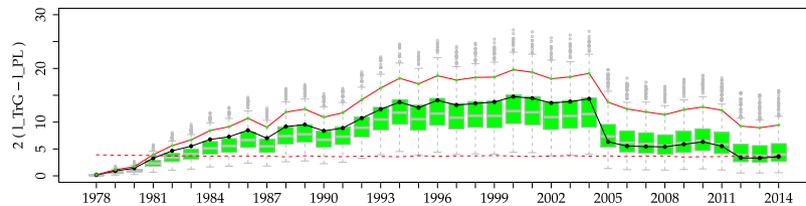


Figure 2. Comparison of the empirical values of the statistic $2\mathcal{R} = 2(l_{trg} - l_{pl})$ (points with lines, shown also in Fig. 1) with those resulting from 1000 simulations of the TrG distribution (boxplots) using the final parameters of Table 1 (i.e., $\beta = 0.681$ and $m_c = 9.15$). The 95th percentile of the boxplots is also shown, in continuous red. Simulated seismic moments are reshuffled as explained in the text to make the comparison possible. The agreement between empirical data and simulations is very remarkable. The red dashed line is the same as in Fig. 1. Remember that the central lines in the boxplots represent the three quartiles of the distribution of $2\mathcal{R}$.

From here we conclude that, in order to find an alternative to the power-law distribution, the truncated gamma distribution is a better option than the Tap distribution, as it is more clearly distinguishable from the power law (for this particular data).

At this point, a direct comparison between these two distributions (Tap and TrG) seems pertinent. In this case we may use the likelihood-ratio test of Vuong for non-nested models^{35,36}. As the number of parameters is the same for both models, their log-likelihood-ratio coincides with the difference in BIC or AIC, but the LR procedure incorporates a statistical test which specifies the distribution of the statistic under consideration. Unfortunately, the results are inconclusive, as no significance difference shows up. This is not surprising if one considers that the LR test for non-nested models is less powerful than the LR test for nested models used above.

In order to check the possible influence of the different heterogeneous populations present in global seismicity, associated to different tectonic zones, we have separately analyzed subduction zones, similarly as done in ref. 37, using Flinn-Engdahl's regionalization⁶. The results for the LR tests are qualitatively the same, with the main difference that the values of $l_{trg} - l_{pl}$ become somewhat smaller (not shown); nevertheless, as long as a time window of several years is considered, the power-law hypothesis can always be rejected except after the Tohoku earthquake. The resulting MLE parameters for the TrG are $\hat{\beta} = 0.649$ and $\hat{m}_c = 9.23$ ($\hat{\theta} = 8.78 \times 10^{22}$ N·m) for $N = 4067$ events. Then, the slightly larger value of \hat{m}_c with respect the global case (Table 1) makes the power law a bit harder to reject.

Simulated Data with Temporal Reshuffling

As we have seen, in contrast to the Tap, the TrG distribution does bring an improvement with respect the PL, so, we concentrate on further comparisons between TrG and PL. With the purpose of gaining further insight, we simulate random samples following the truncated gamma distribution, with the parameters $\hat{\beta}_{trg}$ and $\hat{\theta}_{trg}$ obtained from MLE of the complete dataset (Table 1), with the same truncation parameter a and number of points ($N = 6150$) also. To avoid that the conclusions depend on the time correlation of magnitudes in the empirical data, we reshuffle the simulated data in such a way that the temporal occurrence of the order statistics in the seismic moment is the same as for the empirical data; in other words, the largest simulated event is assigned to take place at the time of the 2011 Tohoku earthquake (the largest of the CMT catalog²³), the second largest at the time of the 2004 Sumatra event, and so on. In this way, we model earthquake seismic moments as arising from a gamma distribution with fixed parameters, with occurrence times given by the empirical times, and with practically the same seismic-moment correlations as the empirical data.

We simulate 1000 datasets with $N = 6150$ each. The results, summarized in Fig. 2 using boxplots³⁸, show that the behaviour of the empirical data is not atypical in comparison with this gamma modelling. In nearly all time windows the empirical data lies in between the first and third quartile of the simulated data, although before 2004 the empirical values are close to the third quartile whereas after 2004 they lay just below the median. This leads us to compute the statistics of the jump in the log-likelihood-ratio between 2004 and 2005. The estimated probability of having a jump larger than the empirical value is around 4.5%, which is not far from what one could accept from the gamma modelling explained above. Thus, a TrG distribution, with fixed parameters, is able to reproduce the empirical findings, if the peculiar time ordering of magnitude of the real events is taken into account. Notice also that, although the simulated data come from a TrG distribution, they are not distinguishable from a power law for about half of the simulations of the last time windows, as the critical region is close to the median indicated by the boxplots.

We can also compare the evolution of the estimated parameters for the empirical dataset and for the reshuffled TrG simulations, with a good agreement again, see Fig. 3. There, it is clear that although the exponent β reaches very stable values relatively soon (around 1990), the scale parameter θ (equivalent to m_c) is largely unstable, and the occurrence of the biggest events makes its value increase.

As a complementary control we invert the situation, simulating 1000 synthetic power-law datasets with $\beta = 0.685$ (Table 1), $a = 5.3 \times 10^{17}$ N·m, and $N = 6150$, for which the same time reshuffling is performed, in such a way that the order of the order statistics is the same. In this case, the results of the simulations lead, on average, to much smaller values of the log-ratio in comparison with the empirical data, which corresponds to the limit of rejection for many time windows, see Fig. 4. So, a power-law distribution with temporal reshuffling cannot

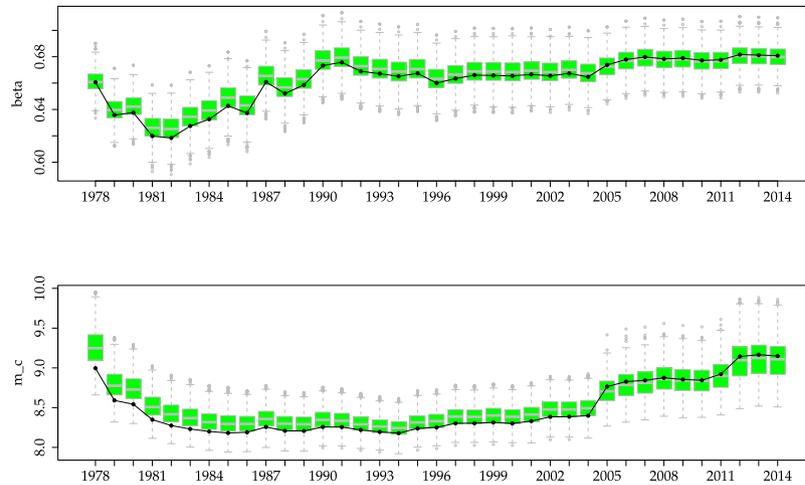


Figure 3. Comparison between empirical estimated parameters $\hat{\beta}$ and \hat{m}_c for the TrG distribution (points with lines) and the estimations for the 1000 simulations of Fig. 2 (i.e., TrG with $\beta=0.681$ and $m_c=9.15$ with temporal reshuffling, boxplots). The different stability of both parameters is apparent, as well as the similarity between data and simulations. (a) $\hat{\beta}$. (b) \hat{m}_c .

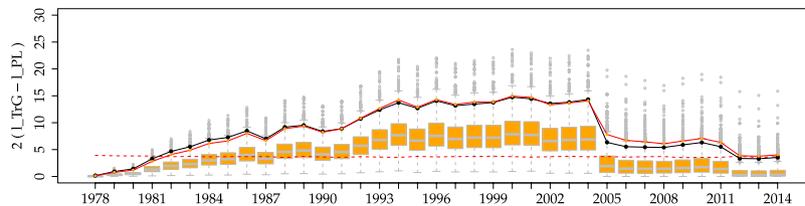


Figure 4. As Fig. 2, but simulating a power law with parameter $\beta=0.685$ (Table 1) instead of a TrG distribution. The reshuffling is also as in Fig. 2, as explained in the text. The simulations lead, on average, to values of the likelihood ratio smaller than the empirical ones. Note that the difference with Fig. 1(b) is that there (i) there is no reshuffling and (ii) the value of β in the simulations corresponds to the obtained $\hat{\beta}$ for each time window.

account for the empirical results as clearly as a truncated gamma distribution. Doing the same with a Tap distribution one finds something in between, see SI.

Discussion

Testing different statistical models for the distribution of seismic moment of global shallow seismicity (using the CMT catalog) we have found that, in contrast to the Tap distribution, the truncated gamma brings significant improvement with respect to the power law. Moreover, in order to reproduce the time evolution of the statistical results, it suffices that independent seismic moments following a truncated gamma distribution with fixed parameters $\beta=0.68$ and $m_c=9.15$ are reshuffled so that the peculiar empirical time sequence of magnitudes is maintained (note that after reshuffling independence is broken). So, despite the fact that the future occurrence of more and larger mega-earthquakes could significantly change the value of parameter m_c ¹³, the current value is enough to explain the available data. Although ref. 13 claims that no less than 45,000 events are necessary for the reliable estimation of m_c , our simulations with 6150 events indicate otherwise, see for instance the last boxplot for the estimation of m_c in Fig. 3, which yields a mean value of 9.11, with a standard deviation of 0.24, totally consistent with the results in Table 1. We conclude that the fundamental problem in the estimation of m_c is not the number of available data but the temporal heterogeneity of the seismic moment distribution. We have also found, with a similar reshuffling procedure, that a power-law distribution cannot account for the empirical findings. Direct comparison of Figs 2 and 4 shows how the TrG distribution outperforms the power law. Additionally, it would be very interesting to investigate if the high values of the likelihood ratio attained before the 2004 Sumatra event could be employed to detect the end of periods of low global seismic activity. Certainly, more case studies would be necessary for that purpose.

As extra arguments in favour of the truncated gamma distribution in front of the tapered GR, we can bring not statistical evidence but physical plausibility and statistical optimality. On the one hand, the former distribution can be justified as coming from a branching process that is slightly below its critical point^{12,39}. Further reasons that may support the truncated gamma are that this arises (i) as the maximum entropy outcome under the constraints of fixed (arithmetic) mean and fixed geometric mean of the seismic moment⁴⁰; (ii) as the closest to the power law, in terms of the Kullback-Leibler divergence, when the mean seismic moment is fixed⁴¹; and (iii) as a stable distribution under a fragmentation process with a power-law transition rate⁴¹. We are not aware of similar

theoretical support in favour of the Tap distribution. On the other hand, it is straightforward to check that the truncated gamma belongs to the exponential family²⁵, in contrast to the Tap distribution. And it is well known that estimators in the exponential family achieve the Cramér-Rao lower bound for any sample size, in contrast to other regular models, where the bound is only achieved asymptotically.

References

1. T. Utsu. Representation and analysis of earthquake size distribution: a historical review and some new approaches. *Pure Appl. Geophys.* **155**, 509–535 (1999).
2. B. D. Malamud. Tails of natural hazards. *Phys. World* **17**(8), 31–35 (2004).
3. P. Bak. *How Nature Works: The Science of Self-Organized Criticality*. Copernicus, New York (1996).
4. L. Knopoff & Y. Kagan. Analysis of the theory of extremes as applied to earthquake problems. *J. Geophys. Res.* **82**, 5647–5657 (1977).
5. A. Corral. Scaling and universality in the dynamics of seismic occurrence and beyond. In A. Carpinteri & G. Lacidogna, editors *Acoustic Emission and Critical Phenomena*, pages 225–244 Taylor and Francis, London (2008).
6. Y. Y. Kagan. Universality of the seismic moment-frequency relation. *Pure Appl. Geophys.* **155**, 537–573 (1999).
7. Y. Y. Kagan. Seismic moment distribution revisited: I. statistical results. *Geophys. J. Int.* **148**, 520–541 (2002).
8. A. Deluca & A. Corral. Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. *Acta Geophys.* **61**, 1351–1394 (2013).
9. Y. Y. Kagan. *Earthquakes: Models, Statistics, Testable Forecasts*. Wiley (2014).
10. D. Vere-Jones. A branching model for crack propagation. *Pure Appl. Geophys.* **114**, 711–725 (1976).
11. I. Main. Statistical physics, seismogenesis, and seismic hazard. *Rev. Geophys.* **34**, 433–462 (1996).
12. A. Corral and F. Font-Clos. Criticality and self-organization in branching processes: application to natural hazards. In M. Aschwanden, editor *Self-Organized Criticality Systems*, pages 183–228 Open Academic Press, Berlin (2013).
13. G. Zöller. Convergence of the frequency-magnitude distribution of global earthquakes: Maybe in 200 years. *Geophys. Res. Lett.* **40**, 3873–3877 (2013).
14. D. Vere-Jones, R. Robinson & W. Yang. Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation. *Geophys. J. Int.* **144**(3), 517–531 (2001).
15. H. Kanamori. The energy release in great earthquakes. *J. Geophys. Res.* **82**(20), 2981–2987 (1977).
16. H. Kanamori & E. E. Brodsky. The physics of earthquakes. *Rep. Prog. Phys.* **67**, 1429–1496 (2004).
17. G. Ekstrom, M. Nettles & A. M. Dziewonski. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Phys. Earth Planet. Int.* **200–201**, 1–9 (2012).
18. T. A. Chou, A. M. Dziewonski & J. H. Woodhouse. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J. Geophys. Res.* **86**, 2825–2852 (1981).
19. C. Godano & F. Pingue. Is the seismic moment-frequency relation universal? *Geophys. J. Int.* **142**, 193–198 (2000).
20. Y. Y. Kagan. Earthquake size distribution: Power-law with exponent $\beta \equiv 1/2$? *Tectonophysics*. **490**, 103–114 (2010).
21. T. Lay. Why giant earthquakes keep catching us out. *Nature* **483**, 149–150 (2012).
22. I. G. Main, L. Li, J. McCloskey & M. Naylor. Effect of the Sumatran mega-earthquake on the global magnitude cut-off and event rate. *Nature Geosci.* **1**, 142 (2008).
23. A. F. Bell, M. Naylor & I. G. Main. Convergence of the frequency-size distribution of global earthquakes. *Geophys. Res. Lett.* **40**, 2585–2589 (2013).
24. E. L. Geist & T. Parsons. Undersampling power-law size distributions: effect on the assessment of extreme natural hazards. *Nat. Hazards* **72**, 565–595 (2014).
25. Y. Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford UP, Oxford (2001).
26. M. Holschneider, G. Zöller & S. Hainzl. Estimation of the maximum possible magnitude in the framework of a doubly truncated Gutenberg-Richter model. *Bull. Seismol. Soc. Am.* **101**(4), 1649–1659 (2011).
27. T. Broderick, M. Dudk, G. Tkacik, R. E. Schapireb & W. Bialek. Faster solutions of the inverse pairwise Ising problem. *arXiv* 0712.2437 (2007).
28. NIST Digital Library of Mathematical Functions. 2014. <http://dlmf.nist.gov/8.7#E3>.
29. P. McCullagh & D. R. Cox. Invariants and likelihood ratio statistics. *Ann. Statist.* **14**(4), 1419–1430 (1986).
30. S. G. Self & K.-Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **82**, 605–610 (1987).
31. C. J. Geyer. On the asymptotics of constrained M -estimation. **22**(4), 1993–2010 (1994).
32. J. del Castillo & A. López-Ratera. Saddlepoint approximation in exponential models with boundary points. *Bernoulli* **12**(3), 491–500 (2006).
33. Y. Y. Kagan. Why does theoretical physics fail to explain and predict earthquake occurrence? In P. Bhattacharyya & B. K. Chakrabarti, editors *Modelling Critical and Catastrophic Phenomena in Geoscience*, Lecture Notes in Physics 705, pages 303–359 Springer, Berlin (2006).
34. Y. Y. Kagan & F. Schoenberg. Estimation of the upper cutoff parameter for the tapered Pareto distribution. *J. Appl. Probab.* **38A**, 158–175 (2001).
35. Q. H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**(2), 307–33 (1989).
36. A. Clauset, C. R. Shalizi & M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).
37. Y. Y. Kagan, P. Bird & D. D. Jackson. Earthquake patterns in diverse tectonic zones of the globe. *Pure Appl. Geophys.* **167**(6), 721–741 (2010).
38. B. Rosner. *Fundamentals of Biostatistics*. Cengage Learning (Boston), 8th edition (2016).
39. K. Christensen & N. R. Moloney. *Complexity and Criticality*. Imperial College Press, London (2005).
40. I. G. Main & P. W. Burton. Information theory and the earthquake frequency-magnitude distribution. *Bull. Seismol. Soc. Am.* **74**(4), 1409–1426 (1984).
41. D. Sornette & A. Sornette. General theory of the modified Gutenberg-Richter law for large seismic moments. *Bull. Seismol. Soc. Am.* **89**(4), 1121–1130 (1999).
42. G. Casella & R. L. Berger. *Statistical Inference*. Duxbury, Pacific Grove CA, 2nd edition (2002).

Acknowledgements

We are grateful to J. del Castillo, Y.Y. Kagan, I.G. Main, M. Naylor, and F. Schoenberg for their feedback. Research expenses were founded by projects FIS2012-31324, FIS2015-71851-P, and MTM2012-31118 from Spanish MINECO, 2014SGR-1307 from AGAUR, and the Collaborative Mathematics Project from La Caixa Foundation (I.S.).

Author Contributions

Both authors discussed the problem. I.S. performed the statistical analysis. A.C. wrote a draft of the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Serra, I. and Corral, Á. Deviation from power law of the global seismic moment distribution. *Sci. Rep.* 7, 40045; doi: 10.1038/srep40045 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017