The estimation of b-value of the frequency-magnitude distribution and of its confidence intervals from binned magnitude data

S. Tinti[1], P. Gasperini[1,2]

[1]Dipartimento di Fisica e Astronomia "Augusto Righi", Università di Bologna, Italy

[2]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy

stefano.tinti@unibo.it, paolo.gasperini@unibo.it

Key points: Discrete magnitude distributions, resilience to magnitude incompleteness

Summary

The estimation of the slope (b-value) of the frequency magnitude distribution of earthquakes is based on a formula derived by Aki decades ago, assuming a continuous exponential distribution. However, as the magnitude is usually provided with a limited resolution, its distribution is not continuous but discrete. In the literature this problem was initially solved by an empirical correction (due to Utsu) to the minimum magnitude, and later by providing an exact formula such as that by Tinti and Mulargia, based on the geometric distribution theory. A recent paper by van der Elst showed that the b-value can be estimated also by considering the magnitude differences (which are proven to follow an exponential discrete Laplace distribution) and that in this case the estimator is more resilient to the incompleteness of the magnitude dataset.

In this work we provide the complete theoretical formulation including i) the derivation of the means and standard deviations of the discrete exponential and Laplace distributions; ii) the estimators of the decay parameter of the discrete exponential and trimmed Laplace distributions; and iii) the corresponding formulas for the parameter b. We further deduce iv) the standard 1-sigma confidence limits for the estimated b. Moreover, we are able v) to quantify the error associated with the Utsu minimum-magnitude correction.

We tested extensively such formulas on simulated synthetic datasets including complete catalogues as well as catalogues affected by a strong incompleteness degree such as aftershock sequences where the incompleteness is made to vary from one event to the next.

Plain language summary

The frequency distribution of the sizes (magnitudes) of earthquakes is particularly relevant for seismic hazard and forecasting. In particular, the slope (b-value) of linear relation existing between

the magnitude and the logarithm of the earthquake frequency has been proposed as an index of the state of stress within the Earth's interior and then of the state of preparation of a future damaging earthquake. In this work we provide a thorough formulation and detailed discussion of the methods by which the *b*-value and its uncertainty can be correctly estimated when the magnitudes of earthquakes are given with a limited resolution and partitioned in equal-size bins. The methods can be divided in two classes: methods analyzing binned magnitudes and methods analyzing binned magnitude differences. The goodness of the different methods is compared using simulated datasets including cases when a certain number of earthquakes are randomly eliminated from the datasets so that to reproduce the incompleteness observed in real data.

Introduction

The b-value of the frequency-magnitude distribution (FMD) (Gutenberg and Richter, 1944)

$$\log_{10} N = a + bM \tag{1}$$

is indicated by some researchers as a proxy of the level of differential stress within the Earth (Scholz, 1968, 2015, Amitrano, 2003) and thus as an index of the state of preparation of future strong earthquakes (Gulia and Wiemer, 2010, 2018, 2019, 2020). Some papers demonstrated that the b-value is negatively correlated with the rake of the focal mechanism (Shorlemmer et al. 2005, Petruccelli et al., 2018, Petruccelli et al., 2019a) and with the source depth (Spada et al., 2013, Petruccelli et al., 2019b). However, these results are controversial and others argued that b-value variations are statistically insignificant as they are due to artifacts of the methods used to determine it (Kagan 1999, 2002, 2003, Bird and Kagan, 2004).

One of the most critical aspects in b-value computations is the determination of the magnitude completeness threshold for the seismic dataset used (e.g. Woessner and Wiemer, 2005, Mignan and Woessner, 2012) as an underestimation of the threshold might bias (lowering) the estimated b-value, whereas an overestimation might reduce the size of the sample too much for a reliable b-value determination.

Aki (1965), assuming a continuous exponential distribution of magnitudes, deduced the formulas for the estimation of the b-value and of its standard confidence interval by the maximum likelihood method as

$$b = \frac{1}{\ln(10)\,(\bar{M} - M_c)} \tag{2}$$

$$\sigma_b = \frac{b}{\sqrt{N}} \tag{3}$$

where $\bar{M}$ is the average magnitude, M_c is the minimum (completeness) magnitude and N is the number of magnitudes in the sample. Eq. (2) was also derived by Utsu (1965) by the method of moments. Utsu (1966) evidenced that the value estimated by eq. (2) is biased (higher) when magnitudes are binned (usually to one decimal digit) and proposed an approximate correction to the original formula

$$b = \frac{1}{\ln(10)\,(\bar{M} - M_c + \delta)} \tag{4}$$

where δ is one half of the binning size (e.g. 0.05).

Studying in detail the statistical distribution of *b*, Shi and Bolt (1982) suggested the following formula for the confidence interval of the continuous distribution

$$\sigma_b = \ln(10) b^2 \sqrt{\frac{\sum_{i=1}^{N}(M_i - \bar{M})^2}{N(N-1)}} \tag{5}$$

Actually, if the magnitude data are binned, their distribution is not continuous anymore, but discrete and this implies changes in the estimators.

Bender (1983) analyzed the problem of estimating the *b*-value from magnitude grouped data and found that the maximum likelihood estimate of *b* is the value for which

$$\frac{q}{1-q} - \frac{nq^n}{1-q^n} = \sum_{i=1}^{n} \frac{(i-1)k_i}{N} \tag{6}$$

where $q = \exp[-b\ln(10)2\delta]$, k_i is the number of earthquakes in the *i*-th magnitude interval of width 2δ, and *n* is the number of magnitude intervals from M_c to the maximum magnitude of the dataset. An explicit expression for *b* was not derived by Bender (1983) and then it can be estimated only numerically.

Guttorp and Hopkins (1986) showed that the maximum likelihood estimate of *b* in case of magnitude data with limited accuracy 2δ is

$$b = \frac{1}{2\delta \ln(10)} \ln\left[1 + \frac{2\delta}{\bar{M} - M_c}\right] \tag{7}$$

which can also be written as

$$b = \frac{1}{2\delta \ln(10)} \ln\left[\frac{\bar{M} - M_c + 2\delta}{\bar{M} - M_c}\right] \tag{8}$$

Tinti and Mulargia (1987) derived the exact equation in case of grouped magnitudes in a paper focused on the confidence intervals, providing the form:

$$b = -\frac{1}{2\delta \ln(10)} \ln\left[\frac{\dfrac{(\bar{M} - M_c + \delta)}{2\delta} - 0.5}{\dfrac{(\bar{M} - M_c + \delta)}{2\delta} + 0.5}\right] \tag{9}$$

which is perfectly equivalent to eq. (8).

Marzocchi et al. (2020) suggested that, when data are binned, the *b*-value computed through the Utsu formula (4), say b_{Utsu}, has to be corrected by

$$b_{\text{corrected}} = \frac{1}{2\delta \ln(10)} \ln\left[\frac{1 + b_{Utsu}\delta \ln(10)}{1 - b_{Utsu}\delta \ln(10)}\right] \tag{10}$$

By substituting (4) in (10) we have

$$b_{\text{corrected}} = \frac{1}{2\delta \ln(10)} \ln\left[\frac{1 + \dfrac{1}{\ln(10)(\bar{M} - M_c + \delta)} \ln(10)\delta}{1 - \dfrac{1}{\ln(10)(\bar{M} - M_c + \delta)} \ln(10)\delta}\right] \tag{11}$$

$$= \frac{1}{2\delta \ln(10)} \ln\left[\frac{\dfrac{\bar{M} - M_c + \delta + \delta}{(\bar{M} - M_c + \delta)}}{\dfrac{\bar{M} - M_c + \delta - \delta}{(\bar{M} - M_c + \delta)}}\right]$$

which is exactly equivalent to eq. (8).

Van der Elst (2021) showed that in case of discretized data, the exact formula for estimating *b* is

$$b = \frac{1}{\delta \ln(10)} \coth^{-1}\left[\frac{1}{\delta}(\bar{M} - M_c + \delta)\right] \tag{12}$$

where $\coth^{-1}$is the inverse of the hyperbolic cotangent function. Recalling the definition of $\coth^{-1}$ it is easy to show that even such equation is equivalent to eq. (8).

Van der Elst (2021) also showed that the *b*-value can be consistently computed by the absolute magnitude differences $|\Delta M|$ (that follow the exponential discrete Laplace distribution) by

$$b = \frac{1}{2\delta \ln(10)} \operatorname{csch}^{-1}\left[\frac{1}{2\delta} \overline{|\Delta M|}\right] \tag{13}$$

where csch^{-1} is the inverse of the hyperbolic cosecant and $\overline{|\Delta M|}$ is the average of the absolute magnitude differences. Recalling the definition of csch^{-1}, equation (13) can also be written in terms of natural logarithm as

$$b = \frac{1}{2\delta \ln(10)} \ln\left[\frac{2\delta + \sqrt{4\delta^2 + \left(\overline{|\Delta M|}\right)^2}}{\overline{|\Delta M|}}\right] \tag{14}$$

We point out that, to compute magnitude differences, one can proceed essentially in two ways: in the first case, one computes the difference between the second and the first magnitude and then between the third and the second one and so on up to the last one:

$$|\Delta M|_i = |M_{i+1} - M_i| \ , \qquad i = 1,2, \dots, N-1 \tag{15}$$

This maximizes the number of data (in all *N*-1), but the differences are not independent from one another, and this might produce some statistical bias. In the second way, one computes the difference between the second and the first magnitude and then between the fourth and the third one and so on up to the last one

$$|\Delta M|_i = |M_{2i} - M_{2i-1}| \ , \qquad i = 1,2, \dots, N/2 \tag{16}$$

This grants that the differences are all independent from one another, but it halves the number of data.

As incompleteness also affects the Laplace distribution of magnitude differences, van der Elst (2021) suggested discarding all $\Delta M = 0$ and then only to consider absolute differences not lower than the binning size $\Delta M'_c = 2\delta$. In this case, he showed that the *b*-value estimator becomes formally equivalent to that of binned magnitudes of eq. (12):

$$b = \frac{1}{\delta \ln(10)} \coth^{-1}\left[\frac{1}{\delta}\left(\overline{|\Delta M|} - \Delta M'_c + \delta\right)\right] \tag{17}$$

provided that $\overline{|\Delta M|}$ and $\Delta M'_c$ replace $\overline{M}$ and M_c respectively. It is obvious that eq. (17) can be written in terms of natural logarithm as

$$b = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{|\Delta M|} - \Delta M'_c + 2\delta}{\overline{|\Delta M|} - \Delta M'_c}\right) \tag{18}$$

Van der Elst did not derive any expressions for the confidence intervals but suggested computing them by means of the bootstrap method (Hurvich and Tsai, 1989). He also asserted that the estimation of *b*-value is more stable and robust if only positive magnitude differences are used in eq. (18).

As van der Elst (2021) did not give much detail on his formulations, in this paper, we provide (see Appendices A-H): i) the complete theoretical derivation of the first two moments of the discrete exponential distribution; ii) the estimators of the decay parameter of the discrete exponential as well as of the discrete Laplace distributions, even in case of distribution trimming; iii) the corresponding formulas for estimating the parameter *b*. Moreover, we deduce iv) the standard one-sigma lower and upper confidence limits for the estimated *b* valid in case of discrete exponential variables as

$$b_1 = \frac{1}{2\delta \ln(10)} \ln\left[\frac{c + \sqrt{\frac{c}{N}}}{1 + \sqrt{\frac{c}{N}}}\right] \tag{19}$$

$$b_2 = \frac{1}{2\delta \ln(10)} \ln\left[\frac{c - \sqrt{\frac{c}{N}}}{1 - \sqrt{\frac{c}{N}}}\right] \tag{20}$$

where

$$c = \exp\left(2\delta \ln(10)\tilde{b}\right) = 10^{2\delta\tilde{b}} \tag{21}$$

and $\tilde{b}$ is the estimate of b. This applies to estimates made through (8) and through (18). In addition, we derive v) the one-sigma confidence limits also when *b* is estimated through the formula (14) that is for distributions of the absolute value of magnitude differences (see formulas (H20a) and (H20b) in the Appendix H). In the Appendix F, we demonstrate that vi) the Utsu correction (4) coincides with the expansion of the exact formula (8) truncated at the second order.

One of the main objectives of this paper is to evaluate the goodness of the various b-value estimators. To this purpose we perform a number of numerical simulations, with particular attention given to cases of incomplete magnitude datasets. The details on how we produce complete and incomplete synthetic datasets and we model aftershock sequences are given in Appendix I.

Evaluation index

To evaluate the goodness of the various methods, we introduce a significance level p we have devised specifically to this purpose, that, given a random sample, is suitable to measure how close a given characteristic value derived from the sample is to a given target value. In our case, the sample is the set of the M estimators $\tilde{b}_i$, $(i = 1,2,\dots,M)$, derived through one of the estimation formulas given above, and the characteristic value is the sample mean $\bar{b}_M$, while the target value is the b-value used to generate the M random datasets. To compute the index p, let's count the number M_+ of $\tilde{b}_i$ that are larger than $\bar{b}_M$ and the number L_+ that are larger than b. Further, let's count the number M_- of $\tilde{b}_i$ that are smaller than $\bar{b}_M$ and the number L_- that are smaller than b. Usually we expect that $M_+ + M_- = M$, and that $L_+ + L_- = M$, but it can happen that some of the $\tilde{b}_i$ accidentally equals $\bar{b}_M$ or b. Consequently, the sums $M_+ + M_-$ and $L_+ + L_-$ might be smaller than M, and it is more convenient to count all quantities separately. We define the performance index as:

$$
\begin{aligned}
p &= 1 \qquad if\ \bar{b}_M = b \\
p &= \frac{L_+}{M_+} \qquad if\ \bar{b}_M < b \\
p &= \frac{L_-}{M_-} \qquad if\ \bar{b}_M > b
\end{aligned}
\tag{22}
$$

The index p takes values between 0 and 1. The way p is computed, makes it a non-parametric index, very easy to obtain, and applicable to any kind of sample and to any pairs of variables one likes to compare. If the sample data number M is so large that one can approximate the sample frequency occurrences with a density probability curve, then the above criterion (22) can be expressed in terms of probability ratios as:

$$p = \frac{Prob(\tilde{b} > b)}{Prob(\tilde{b} > \bar{b}_M)} \quad if \;\; \bar{b}_M < b$$
$$p = \frac{Prob(\tilde{b} < b)}{Prob(\tilde{b} < \bar{b}_M)} \quad if \;\; \bar{b}_M > b \tag{23}$$

In case of variables $\tilde{b}_i$ following a Gaussian distribution with standard deviation σ, where $\bar{b}_M$ coincides with the distribution median and $Prob(\tilde{b} < \bar{b}_M) = Prob(\tilde{b} > \bar{b}_M) = 1/2$, the criterion simplifies further to:

$$p = 2\,Prob(\tilde{b} > b) \quad if \;\; \bar{b}_M < b; \quad p = 2\,Prob(\tilde{b} < b) \quad if \;\; \bar{b}_M > b \tag{24}$$

that can be easily expressed in terms of the error function of the normalized variables, and p identifies with the p-value of a two-sided Student's t test, i.e. $p = \left[1 - erf\left(\left|b - \bar{b}_M\right|/\sqrt{2}\sigma\right)\right]$.

In this paper, we will rely on the index p to judge the goodness of the estimators. We will use it like the p-value of a null-hypothesis test, where the null hypothesis is that the estimator is acceptable. If $p < \alpha$ where α is the significance level, then we reject the null hypothesis and consider the estimator unacceptable. In this analysis we take $\alpha = 0.05$. A further way we use p is assuming that the performance of the estimator method is an increasing function of p and that if the index of a method is larger than the index of another, then the former shows a better performance.

Results for complete datasets

We first compare the various estimators of eqs. (2), (4), (6), (8), (14) and (18) on complete binned datasets simulated by the procedure specified by the eqs. (I1)-(I3) given in the Appendix I. In this analysis the parameter M_{min} used to generate the random samples and the parameter M_c appearing in the estimator formulas are assumed to be equal. In Table 1 we report the average b-values $\bar{b}_M$ and the corresponding standard deviations S_M computed on a set of M=10,000 simulated samples each including N=1000 magnitudes, with binning size 2δ=0.1 and with b=1. When applying methods that use magnitudes (i.e. eqs. (2), (4), (6) and (8)), all samples have the same number of data (N=1000).

This is also true when magnitude differences are used (see eq.(14)). In this case, however, the data number depends on the way such differences are computed, since differences made through eq.(15) and through eq.(16) lead to N=999 and to N=500, respectively. When null differences are trimmed away (as is required by the estimator of eq. (18)), the number of remaining data changes from sample to sample, which gives the reason to introduce the mean number of data $\bar{N}$ in Table 1.

Table 1 – Estimates from complete simulated sets with *N*=1000, 2δ=0.1 and *b*=1

Estimator	**Eq.**	$\bar{b}_M$	S_M	$\bar{N}$	p
Aki (1965)	(2)	1.125907	0.039867	1000	0.000587
Aki (1965), Utsu (1966)	(4)	0.996582	0.031225	1000	0.913052
Bender (1983)	(6)	0.994843	0.031965	1000	0.869062
This paper, magnitudes	(8)	1.001003	0.031644	1000	0.974374
This paper, absolute differences by eq. (15)	(14)	1.001331	0.040499	999	0.974819
This paper, absolute differences by eq. (16)	(14)	1.001854	0.044692	500	0.966667
This paper, trimmed absolute differences by eq. (15)	(18)	1.001663	0.043709	885	0.970893
This paper, trimmed absolute differences by eq. (16)	(18)	1.002250	0.048326	443	0.963718

The results shown in Table 1 allow us to state that most methods reproduce the true *b*-value (b =1) reasonably well with the exception of the simple Aki formula (2) for which the estimated *b*-value is significantly different from the true one. Since the corresponding p is less than $\alpha = 0.05$, we conclude that the Aki estimator is not acceptable for binned data. We notice that the Utsu estimator (4) and the Bender method (6) give results with evaluation index p included in the interval $[0.85, 0.92]$, that is smaller than the index p of all other methods. This is the first clue that the methods based on the formulas (8), (14) and (18) are the most convenient ones. This analysis is corroborated by results obtained by varying the number of data *N* (100, 1000, 10000) and the theoretical *b*-value (0.7, 1.0, 1.5), that confirm the superiority of such methods (see Tables S1 to S9 in the supplementary material).

In Table 2 for different simulated complete datasets, we report *b*-values computed by eq. (8) and standard deviations estimated using various methods from the literature and also by eqs. (19) and (20) introduced in this paper. The latter ones are reported in the last column of the Table as the half amplitude of the confidence interval, i.e. as $\frac{1}{2}(b_2 - b_1)$. Notice that the mean value $\bar{b}_M$ is not the midpoint of the confidence interval and is always closer to the lower end (see the third-to and second- to-last column of the Table). It is remarkable that for all cases, the estimates based on Aki and on Shi-Bolt formulas as well as the ones based on eqs. (19) and (20) of this paper correspond very well to the standard deviation S_M computed from the simulated datasets.

Table 2 – Standard deviations for complete simulated sets with $\bar{b}_M$ estimated through eq.(8)

***b*-value**	N	$\bar{b}_M$	S_M	**Aki eq.(3)**	**Shi-Bolt eq.(5)**	$\bar{b}_M - b_1$ **eq.(19)**	$b_2 - \bar{b}_M$ **eq.(20)**	$\frac{1}{2}(b_2 - b_1)$
0.7	1000	0.700502	0.022256	0.022152	0.022080	0.021494	0.022903	0.022198
1	1000	1.000699	0.031843	0.031645	0.031507	0.030737	0.032758	0.031747
1.5	1000	1.501058	0.047887	0.047468	0.047132	0.046222	0.049287	0.047755

In Table 3 we report *b*-values computed for trimmed magnitude differences by eq. (18) and the corresponding standard deviations computed by means of eqs. (19) and (20), as specified earlier. The differences are computed by using eqs. (15) and (16). It is relevant to observe that when using independent differences (eq. 16), these values correspond well to the standard deviation S_M computed from the simulated datasets. On the contrary, when using non independent differences computed by eq. (15), there is an underestimation that amounts to about 22% for all the treated cases. The latter might be related to some sort of data correlation that reduces the number of "effective" independent data, say N_e, in the difference dataset, so that the calculated dispersion is less than the experimental one. Considering that the standard deviation scales inversely with the square root of the number of data, then the observed percentage increase of S_M corresponds to the decrease of about 40% in the number of effective data, i.e. $N_e \sim 0.67N$. Then, we conclude that to compute differences it is always preferable to use eq. (16), and this will be our choice in all the following computations shown in the paper and in the supplementary material.

Table 3 – Standard deviations for complete simulated sets with $\overline{b}_M$ estimated through trimmed absolute differences (Eq.18)

b-value	Eq.	N	$\overline{b}_M$	S_M	$\overline{b}_M - b_1$ eq.(19)	$b_2 - \overline{b}_M$ eq.(20)	$\frac{1}{2}(b_2 - b_1)$
0.7	(15)	999	0.700767	0.028808	0.021513	0.022923	0.022218
	(16)	500	0.701012	0.031554	0.030037	0.032859	0.031448
1	(15)	999	1.001103	0.041072	0.030764	0.032788	0.031776
	(16)	500	1.001422	0.044977	0.042951	0.047001	0.044976
1.5	(15)	999	1.501469	0.061124	0.046257	0.049327	0.047792
	(16)	500	1.501969	0.067012	0.064577	0.070719	0.067648

Table 4 - Estimates from complete simulated sets with N=1000, 2δ=0.5 and b=1

Estimator	Eq.	$\overline{b}_M$	S_M	$\overline{N}$	p
Aki (1965)	(2)	1.884281	0.106413	1000	0.000000
Aki (1965), Utsu (1966)	(4)	0.903155	0.024395	1000	0.000000
Bender (1983)	(6)	0.996548	0.033689	1000	0.916258
This paper, magnitudes	(8)	1.001296	0.033480	1000	0.966446
This paper, absolute differences	(14)	1.001698	0.041874	500	0.960875
This paper, trimmed absolute differences	(18)	1.004231	0.069217	240	0.954589

Even if, for most papers in the literature, the binning size is fixed to 0.1 as in Tables 1, 2 and 3, larger bins can be assumed when the magnitude resolution is wider, as it may occur for magnitudes derived from maximum macroseismic intensities.

In Table 4 (that coincides with Table S14 in the supplement) we show the results for a binning size 2δ=0.5. We can note that in this case even the Aki (1965) estimator as corrected by Utsu (1966) (eq. 4) significantly underestimates the simulated b-value, and that the corresponding evaluation index p does not pass the threshold. Therefore, it cannot be accepted as a valid estimator. This underestimation is observed even by varying the number of data N (100, 1000, 10000) and the theoretical b-value (0.7, 1.0, 1.5) (see Tables S10 to S18 in the supplementary material). This confirms that Utsu (1966) correction to the Aki (1965) formula is approximate and in not adequate for large bin sizes. Table 4 also confirms that the Bender method, though adequate, has a performance slight smaller than eq. (8) for magnitudes and formulas (14) and (18) for magnitude differences, but this might be due to a not fully accurate numerical minimization. It is relevant to note that the corresponding standard deviation S_M increase sensibly when passing from magnitudes to magnitude differences, since the number of data decreases, which means that the latter estimators can be considered less efficient.

Results for incomplete datasets

Experimental magnitude datasets are always affected by some degree of incompleteness. Therefore, evaluating how the estimators perform on incomplete data set is of paramount importance. Van der Elst (2021), on analyzing incomplete binned magnitude sequences, concluded that magnitude difference estimators are more robust if only the positive differences are used, since using also the negative differences might produce biased results. To check this conclusion, in the following computations we consider two further estimators, one based only on the trimmed positive differences and the other based only the trimmed negative ones in eq. (18). We build incomplete datasets by starting from 11,000 exponentially distributed magnitudes with $M_{min} = 0.4$ and decay parameter $b = 1$, and by applying the thinning method as explained in Appendix I with parameters

$\mu = 1$ and $\lambda = 0.2$. After the reduction the remaining magnitudes are less than 10%. More precisely the mean number of earthquakes per sample is $\bar{N}$=1093. The histogram of one of the simulated datasets is portrayed in Figure 1 and is manifestly incomplete. Seismologists usually refrain from estimating the b-value using all data of such incomplete dataset, and in common practice one discards the smaller magnitude classes that are visibly incomplete. Nonetheless, it is interesting to see the behavior of the estimators in the extreme case where M_c is taken to be equal to M_{min}, although it is clear from Figure 1 that such kind of samples are far from fitting an exponential distribution. It is worth stressing that $M_c$enters in all formulas based on magnitudes, namely (2), (4), (6) and (8), but not in the ones based on magnitude differences, i.e. (14) and (18). The results are shown in Table 5.

Table 5 - Incomplete simulated sets with $\mu = 1$, $\lambda = 0.2$, 2δ=0.1 and *b*=1, M_c=0.4

Estimator	Eq.	$\bar{b}_M$	S_M	$\bar{N}$	p
Aki (1965)	(2)	0.460944	0.006947	1093	0.000000
Aki (1965), Utsu (1966)	(4)	0.437711	0.006264	1093	0.000000
Bender (1983)	(6)	0.387450	0.024053	1093	0.000000
This paper, magnitudes	(8)	0.438082	0.006280	1093	0.000000
This paper, absolute differences	(14)	0.862855	0.032991	546	0.000000
This paper, trimmed absolute differences	(18)	0.890224	0.036483	506	0.005059
This paper, trimmed positive differences	(18)	0.892015	0.052275	253	0.052535
This paper, trimmed negative differences	(18)	0.891447	0.051621	247	0.044536

As expected, all methods provide estimates quite far from the true value, and correspondingly the index p lies below or close to the significance threshold $\alpha = 0.05$. But it is relevant to observe that on using differences one obtains better results, and even better on using strictly positive or strictly negative differences, although the improvement of the estimate is accompanied by an increment of the standard deviation.

In Table 6 we show the results considering exactly the same data sets as before, but assuming that M_c is equal to 1.1, corresponding to the maximum curvature of the frequency magnitude distribution, as suggested by Wiemer and Wyss (2000). It is clear that increasing the value of the completeness magnitude, improves the results for all methods. One common feature is that all estimates lie below the true value, with one exception. Indeed, the Aki (1965) estimator gives an overestimate and is seemingly better than the estimator corrected by Utsu (1966) and the exact formula using magnitudes. However, this is an artefact since the typical underestimation due to incompleteness is overcompensated for the Aki method by the effect of magnitude binning. Most importantly, observe that methods based on magnitude differences and either on the positive differences or on the negative differences produce the best results in terms of closeness to the true b and in terms of the evaluation index p. Such evidence seems to confirm the strategy suggested by van der Elst (2021) that using magnitude differences is preferable than using magnitudes. However, his claim that positive differences lead to better estimates than using negative differences is not so evident from this Table.

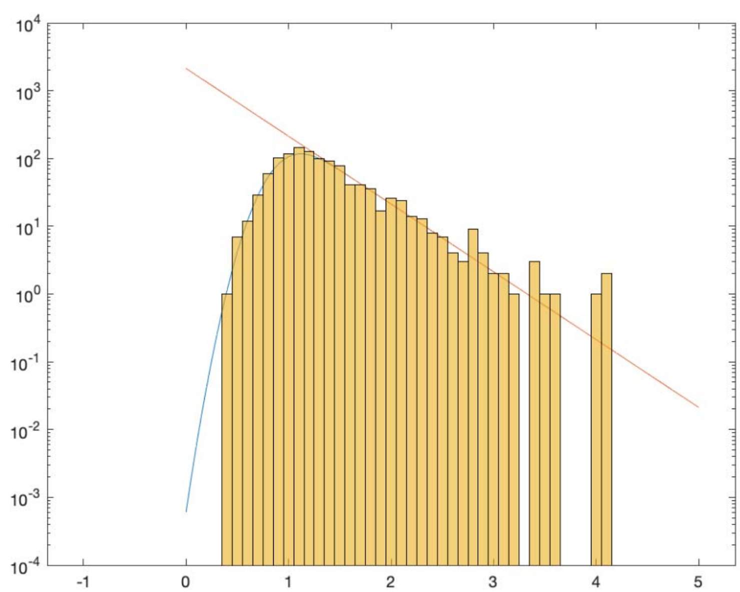

Figure 1 – Incomplete simulated set of N=1029 data with parameters: $\mu = 1$, $\lambda = 0.2$, $M_{min} = 0.4$, $2\delta = 0.1$, $b = 1$.

In Table 7 we show the results for the same parameters when the minimum magnitude M_c is increased up to 1.3 corresponding to the magnitude of maximum curvature plus 0.2, that is the way M_c is commonly set in the literature (Wiemer and Wyss, 2000, Mignan and Woessner, 2012).

Table 6 - Incomplete simulated sets with $\mu = 1$, $\lambda = 0.2$, 2δ=0.1, b=1 and M_c=1.1

Estimator	Eq.	$\bar{b}_M$	S_M	$\bar{N}$	p
Aki (1965)	(2)	1.026523	0.037518	786	0.478636
Aki (1965), Utsu (1966)	(4)	0.917912	0.029991	786	0.008349
Bender (1983)	(6)	0.911953	0.031097	786	0.006099
This paper, magnitudes	(8)	0.921364	0.030332	786	0.012220
This paper, absolute differences	(14)	0.973845	0.047540	393	0.582759
This paper, trimmed absolute differences	(18)	0.986348	0.051871	353	0.788916
This paper, trimmed positive differences	(18)	0.989979	0.074481	176	0.885868
This paper, trimmed negative differences	(18)	0.988299	0.073980	154	0.869655

Notice that increasing the value of the completeness magnitude has the effect of reducing the size of the samples and consequently of increasing the simulated standard deviation S_M. Further, it lowers the degree of incompleteness, and therefore leads to better estimates. As for the rest, most considerations made for Table 6 apply also to Table 7. Most precisely, we can see that now the Aki (1965) formula (2) clearly overestimates the b-value and is thus inadequate, whereas the other estimators using magnitudes as well as the estimators based on magnitude differences give correct results. In this case, the estimator using the trimmed positive differences seems to be slightly better than the one using negative differences but slightly worse than the one using the absolute differences, either trimmed or not.

Table 7 - Incomplete simulated sets with $\mu = 1$, $\lambda = 0.2$, 2δ=0.1, *b*=1 and M_c=1.3

Estimator	Eq.	$\bar{b}_M$	S_M	$\bar{N}$	p
Aki (1965)	(2)	1.107743	0.052196	541	0.027810
Aki (1965), Utsu (1966)	(4)	0.982229	0.041025	541	0.670191
Bender (1983)	(6)	0.976523	0.042257	541	0.579654
This paper, magnitudes	(8)	0.986471	0.041560	541	0.744560
This paper, absolute differences	(14)	0.998481	0.060113	270	0.979558
This paper, trimmed absolute differences	(18)	1.001747	0.064811	240	0.975911
This paper, trimmed positive differences	(18)	1.005584	0.094333	120	0.958555
This paper, trimmed negative differences	(18)	1.006768	0.092576	118	0.938444

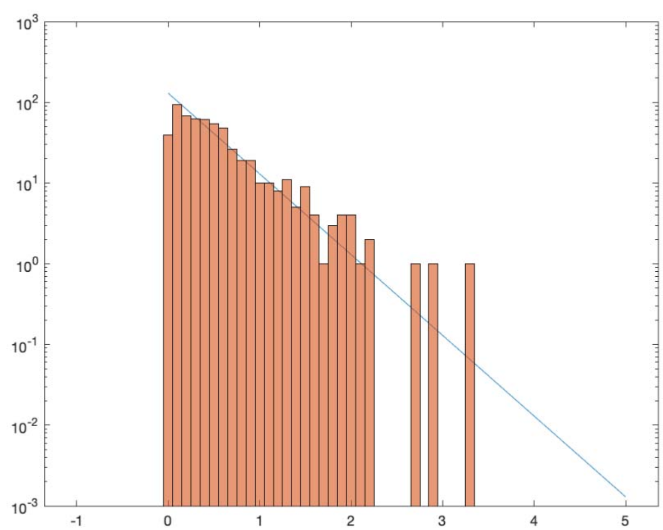

Figure 2 – Histogram of the absolute differences computed for the dataset shown in Figure 1. As expected its size is $N = 564$.

The resilience of trimmed differences estimators to magnitude incompleteness can be further proven through the following experiment. We take exactly the same datasets analyzed before and with estimates given in Table 5: they have average size $\bar{N} = 1093$ and have $M_c = 0.4$. They are strongly incomplete (see the exemplary sample in Figure 1). Let us estimate the b-value through the eq.(18) and change the value of $\Delta M'_c$. Figure 2 displays the histogram of the absolute differences of the same data set plotted in Figure 1. It includes also the null differences, that however will be neglected in the analysis. It is quite evident from this graph, that the histogram exhibits an exponential behavior much more than the corresponding graph of Figure 1, and it is much more so if the first column on the left hand side is discarded. This rises the expectation that methods based on differences, rather than on magnitudes are able to provide better estimates. This is confirmed in Table 8. The last three rows of Table 5 show the results for trimmed differences estimators (absolute, positive and negative) when we apply the basic trimming, which means that we set $\Delta M'_c = 2\delta$, so discarding the differences equal to 0. In Table 8 we show the results obtained by increasing $\Delta M'_c$ up to $\Delta M'_c = 10\delta$ (i.e. 5 times the bin size). In the Table we include also the outputs for $\Delta M'_c = 2\delta$ for the sake of comparison. Very remarkably, it can be seen that the deviations of the estimated b-values from the theoretical one progressively decrease when incrementing the amount of trimming, and at the end they become almost negligible for all of the three estimators. It is also worth stressing that increasing the trimming threshold $\Delta M'_c$ obviously reduces the number of available data and leads to larger standard deviations, but less dramatically than a similar increment is applied to the magnitude threshold M_c. Even in this experiment it can be observed that positive differences give results slightly better than negative differences and systematically better than absolute differences.

Results for incompleteness changing with time

It is known that incompleteness affects seismic catalogues systematically after a strong main shock, owing to the superposition of the waveforms in the recorded seismograms that prevents the correct location and size determination of many small shocks in the hours or days after the main shocks. It is also known that in these circumstances incompleteness is time dependent since it tends to decrease in time which means that the completeness magnitude can be modeled as a decreasing function of time.

Table 8 - Incomplete simulated sets generated with parameters $\mu = 1, \lambda = 0.2$, 2δ=0.1 and b=1, M_c=0.4 and estimated through eq.(18)

Estimator	$\Delta M'_c$	$\bar{b}_M$	S_M	$\bar{N}$	p
This paper, trimmed absolute differences	2δ	0.890224	0.036483	506	0.005059
This paper, trimmed positive differences	2δ	0.892015	0.052275	253	0.052535
This paper, trimmed negative differences	2δ	0.891447	0.051621	247	0.044536
This paper, trimmed absolute differences	4δ	0.927973	0.042749	428	0.101695
This paper, trimmed positive differences	4δ	0.930314	0.061307	214	0.267504
This paper, trimmed negative differences	4δ	0.929565	0.060234	212	0.254639
This paper, trimmed absolute differences	6δ	0.957032	0.049715	355	0.394683
This paper, trimmed positive differences	6δ	0.959837	0.071424	178	0.570807
This paper, trimmed negative differences	6δ	0.959462	0.070339	180	0.560562
This paper, trimmed absolute differences	8δ	0.977009	0.057063	290	0.683715
This paper, trimmed positive differences	8δ	0.980645	0.081274	145	0.812863
This paper, trimmed negative differences	8δ	0.980056	0.081047	146	0.807636
This paper, trimmed absolute differences	10δ	0.990306	0.064465	235	0.879491
This paper, trimmed positive differences	10δ	0.994635	0.091919	117	0.956170
This paper, trimmed negative differences	10δ	0.994486	0.092570	121	0.952622

We generate the synthetic data sets following the procedure described in the Appendix I. We set the magnitude of the main shock to m=5.6 in eq. (I4), p=1 and c=0.01 in eq. (I5) and T_E=5 days in eqs.

(I8-I13). We produce $M = 10{,}000$ aftershock sequences of $N = 40{,}000$ earthquakes that are made incomplete through the thinning mechanism, that is a probabilistic and time-dependent process. The magnitude mean of the thinning law $\mu(t)$ is given in eq.(I4) and λ is set to 0.2. Further, we compute the magnitude M_c using the criterion adopted for the analysis shown in Table 7, i.e. M_c is equal to the magnitude M_{mxc} corresponding to the maximum curvature of the sample frequency magnitude curve plus twice the bin size. Then, we eliminate all magnitudes smaller than M_c. The mean number of surviving data is $\bar{N} = 1041$. In Table 9 we present the results of the estimation when M_c is equal to 1.3. Our finding is that none of the estimators is performing very well. However, we observe that estimators based on magnitudes tend to severely underestimate the theoretical b, and the corresponding values of the evaluation index p is much smaller than the 5% significance level. Conversely, the estimators based on differences give results that can be considered satisfactory. A further remark is that one can note a slightly better performance for positive differences with respect to the negative ones.

Table 9 – Aftershock sequence with time-dependent incompleteness; parameters: $\lambda = 0.2$, *m*=5.6, *p*=1, *c*=0.01, T_E=5 days, N=40,000 (before thinning), 2δ=0.1 and *b*=1, M_c=1.3

Estimator	Eq.	$\bar{b}_M$	S_M	$\bar{N}$	p
Aki (1965)	(2)	0.835400	0.025265	1041	0.000000
Aki (1965), Utsu (1966)	(4)	0.762046	0.021019	1041	0.000000
Bender (1983)	(6)	0.752022	0.022715	1041	0.000000
This paper, magnitudes	(8)	0.764015	0.021183	1041	0.000000
This paper, absolute differences	(14)	0.952553	0.040146	520	0.244594
This paper, trimmed absolute differences	(18)	0.965537	0.043553	469	0.430854
This paper, trimmed positive differences	(18)	0.967733	0.061519	234	0.589812
This paper, trimmed negative differences	(18)	0.967363	0.062788	228	0.596715

To examine the performance of the estimators on aftershock sequences we have varied the number of data and the theoretical b-value (see Tables from S19 to S27 in the supplementary material). Notice that Table 9 coincides with Table S23. We have found confirmation that methods based on differences provide better results than methods based on magnitudes and that methods based on differences of the same sign (positive or negative) are better than the ones obtained when using absolute differences, and provide estimates very close to one another, with slight advantage of positive over negative datasets.

Discussion of the results

Computing the b-value of a magnitude catalogue or a magnitude sequence is a classical activity in standard seismicity analyses and is often used in advance studies. For complete datasets, when the magnitudes are taken as continuous exponentially distributed variables, the problem of estimating b and the related confidence intervals was given a definite solution respectively by Aki (1965) and by Shi and Bolt (1982) who made recourse to the chi-square distribution. When magnitudes are grouped in bins of equal size, the problem was also given a final solution by Guttorp and Hopkins (1986) for estimating b and by Tinti and Mulargia (1986 and 1987) to compute the corresponding confidence intervals. Catalogue incompleteness affects remarkably all the estimations, but this problem was long overlooked, since there was the believe that it regards only the smaller magnitudes range and that it is easily possible to find a threshold magnitude, M_c above which the data set is complete and is sufficiently large to allow accurate estimates. More sophisticated views emerged progressively as wells as criteria to find M_c (see Mignan and Woessner, 2012), but it was only recently that the problem was tackled by a very different point of view. It was van der Elst's (2021) who analyzing aftershock sequences where the completeness is known to change quite rapidly with time, proposed to consider datasets of differences of magnitudes in place of magnitudes and to base on them all the statistical inferences. This was a remarkable turning point in the discipline. He also suggested that using datasets of positive differences was the only correct way because the alternative choice of using absolute differences of magnitudes or negative differences leads to inaccurate estimates.

In this paper we have explored the potential of the new approach and we have considered only synthetic datasets since this is the best way to evaluate the performance of inferential estimators. Indeed, one can generate easily a very high number (10,000 in this paper) of pseudorandom samples of magnitudes of any reasonable size. First of all, we have noticed that passing from random variables to their random differences implies halving the size of the data set if one wants to keep data independency (see eq. (16)), which in principle is a cost that is hard to pay. On the other hand, making consecutive differences (see eq.(15)) does not alter significantly the set size (it passes from N to $N-1$), but it introduces an undesirable correlation in the data. The effect of such correlation has been shown in Table 3 where the standard deviation of the correlated differences was found to be larger than the theoretical one by an amount of about 22%. This implies that the variance increases by a factor 1.49 or correspondingly that the equivalent number of data is reduced. In this paper we have opted to work only with datasets formed by independent variables obtained through eq. (16), accepting to pay the cost of halving the dataset size. Consider that the cost is even larger if one uses only one-sign differences (either positive or negative) since in this case the database size collapses to one quarter.

To judge the goodness of the various estimators based on magnitudes and magnitude differences we have devised a new evaluation index denoted by p (eq.22), that can be interpreted as a non-parametric variant of the Student's t. Depending on the way it is defined, that is a sort of normalization, the index p, like the Student's t, tends to tolerate larger discrepancies for estimators with larger standard deviations, and does not prize the estimator efficiency. Bearing this in mind, we consider p a suitable index to evaluate estimators that are known a priori to operate on different dataset sizes like the ones working on magnitudes and on magnitude differences.

When treating complete binned datasets our analysis confirms that the classical estimator of eq. (8) works quite well and better than estimators for continuous magnitudes either in the original form (eq. (2)) or in the one corrected to account for binning (see eq. (4)). It is relevant to point out that also the methods based on magnitude differences and given in eqs. (14) and (18) provide equivalently good results.

The most important finding we obtain is that magnitude samples showing a very severe incompleteness transform to samples with an almost exponential decay when magnitudes are replaced by magnitude differences (see Figures 1 and 2). This seems to be a strong support to the strategy of using absolute differences or one-sign differences to evaluate b. Results reported in Tables from 5 to 7 show that increasing the magnitude value M_c in the traditional estimator (8) from the very low value of patent incompleteness to the magnitude of maximum curvature of the frequency magnitude distribution and even further, improves its performance very much. Nonetheless, from all those Tables it emerges that better evaluations are attained by the estimator (14) that applies to absolute differences and by the estimator (18) that applies to one-sign differences with the minimum possible trimming, i.e. the trimming involving only zero differences. Table 8 explores the performance of the estimator (18) when the magnitude dataset is very incomplete (M_c is assume to be very low) and the trimming is made progressively more substantial and it is found that increasing the trimming also improves the results. In virtue of these outcomes, one could play with M_c and with the amount of trimming $\Delta M'_c$ to optimize the estimates. However, refining this strategy is not the scope of this paper. On comparing the results of all Tables from 5 to 8, we observe that the most accurate estimates are reported in the last 4 rows of Table 7 and are obtained when $M_c = M_{mxc} + 4\delta$=1.3 and the trimming is the minimum, namely either null or equal to 2δ.

With this clue in mind, we have considered aftershock sequences. Their main feature is not only that they are strongly incomplete dataset, but also that the completeness magnitude changes quickly during the process, and formally it changes from one earthquake to the next (according to eq. (I4) in the Appendix I). Nonetheless, each sequence gives rise to a frequency magnitude histogram that can be examined as in the previous analysis, that is by establishing the maximum curvature magnitude

M_{mxc} and by determining M_c through the formula given above. The results of Tables S19-S27 in the supplementary material confirm the superiority of the formulas based on magnitude differences, but they also add that the estimator (18) including trimming is generally better than formula (14).

The second most important result of our analysis on the behavior of the estimators dealing with incomplete datasets is that the van der Elst's claim (2021) that in aftershock sequences one necessarily finds that positive magnitude differences are distributed exponentially much better than the negative ones and therefore the most successful estimator is the estimator (18) applied to the subsets of positive differences is not supported by enough evidence. As a matter of fact, we found that the estimates are very close to one another and that in some cases positive differences provide better values while in other cases the reverse is true.

A final remark regards the confidence limits. The van der Elst's contribution brought in the seismological arena the two new estimators (14) and (18), without providing however the related confidence intervals. The formula (18) is practically the same as (8) but applied to magnitude differences, since the supporting distribution (discrete Laplace) is the same. This means that the corresponding theoretical confidence intervals based on the geometric distribution are already known for any given number *N* of sample data (Tinti and Mulargia, 1986, 1987). In this paper we have provided explicit simple formulas, (19) and (20), to compute the 1-sigma confidence interval when the sample size is large enough that the distribution of the mean (either $\bar{M}$ or $\overline{|\Delta M|}$) approximates a Gaussian, which is normally the case in seismological practice. On the other hand, to authors' knowledge, the confidence intervals of the distribution underlying the estimator (14) have not yet been given a general theoretical solution in the seismological literature. In this context, and under the same assumption of Gaussian distribution, we propose the formulas (H20a and H20b) derived in the Appendix H to compute the endpoints of the 1-sigma confidence intervals.

Conclusions

This paper consisted in two parts and was inspired by the new approach due to van der Elst (2021) to deal with strongly incomplete data set of magnitudes to make estimates of the *b*-value. The theoretical part, mostly developed in the Appendixes recalls in a plane way the main properties of the binned magnitude distributions, more precisely the discrete exponential and the discrete Laplace distribution, the latter being analyzed also in its various variants of absolute differences (either including or not null differences) and one-sign differences. It is a systematic analysis reproducing known results, but also providing clarifications and leading to new outcomes, such as the expressions to compute the 1-sigma confidence limits of the estimators. The second part, chiefly illustrated in the main text and complemented by the supplementary material, is an attempt to evaluate the goodness of the classical estimators of the b-value compared to the ones based on magnitude differences. For estimators based on the distribution of magnitudes, exact formulations (eq. 8) are always preferable with respect to the approximate formula by Aki (1965) with Utsu (1966) correction (eq. 4), in particular when the bin size is larger than 0.1. The uncorrected formula by Aki (1965) (eq. 2) usually overestimates the theoretical b-value, but sometimes may deceptively appear to work well when, by chance, the overestimation due to the binning almost exactly compensates the underestimation due to incompleteness. In case of substantially incomplete catalogues, it was shown that the distributions of magnitude differences happen to be closer to an exponential decay. Therefore, estimators using magnitude differences (eqs. 14 and 18) are more robust with respect to magnitude incompleteness than those using magnitudes (eq. 8) and give correct b-values when the magnitude cutting threshold M_c is not lower than the magnitude of maximum curvature of the frequency magnitude distribution. Conversely, estimators using magnitudes (eq.8) give correct results only for M_c not lower than the magnitude of maximum curvature plus 0.2. The latter finding confirms the goodness of a common choice, made in current literature (Mignan and Woessner, 2012), to establish the magnitude completeness threshold.

Acknowledgements

This study was partially supported by the Real-time earthquake rIsk reduction for a reSilient Europe (RISE) project, funded by the European Union's Horizon 2020 research and innovation program under Grant Agreement Number 821115.

Open Research

Simulations are made using a Matlab code written by authors that, for reviewing purposes, is made available in the supporting information and will be provided in a public repository in case of acceptance for publication of the paper. The paper does not use any kind of data from public or private databases, but only data generated through the Matlab code quoted above

References

Aki, M. (1965). Maximum likelihood estimate of b in the formula log N = a – bM and its confidence limits, Bull. Earthquake Res. Inst., Tokyo Univ. 43, 237-239. https://doi.org/10.15083/0000033631.

Amitrano, D. (2003). Brittle–ductile transition and associated seismicity: Experimental and numerical studies and relationship with the b value, J. Geophys. Res. 108, no. B1, 2044, doi: 10.1029/2001JB000680.

Bender, B. (1983). Maximum likelihood estimation of *b* values from magnitude grouped data, Bull. Seism. Soc. Am., 73, 831-851.

Bird, P., & Kagan, Y. Y. (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings, Bull. Seismol. Soc. Am. 94, 2380–2399.

Gulia, L., & Wiemer, S. (2010). The influence of tectonic regimes on the earthquake size distribution: A case study for Italy, Geophys. Res. Lett. 37, L10305 , doi: 10.1029/2010GL043066.

Gulia, L., & Wiemer, S. (2019). Real-time discrimination of earthquake foreshocks and aftershocks, Nature 574, 193–199.

Gulia, L., Rinaldi, A. P., Tormann, T., Vannucci, G., Enescu, B., & S. Wiemer (2018). The effect of a mainshock on the size distribution of the aftershocks, Geophys. Res. Lett. 45, no. B1, doi: 10.1029/ 2018GL080619.

Gulia, L., Wiemer, S., & Vannucci, G. (2020). Pseudoprospective Evaluation of the Foreshock Traffic-Light System in Ridgecrest and Implications for Aftershock Hazard Assessment, Seismol. Res. Lett. 91, 2828–2842, doi: 10.1785/0220190307.

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California, Bull. Seism. Soc. Am., 34: 185-188. Doi: 10.1785/BSSA0340040185.

Guttorp, P. & Hopkins, D. (1986). On estimating varying *b* values, Bull. Seism. Soc. Am., 76, 889-895. doi: 10.1785/BSSA0760030889.

Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2006). Comparison of short-term and time-independent earthquake forecast models for southern California. Bulletin of the Seismological Society of America, 96(1), 90–106. Doi: 10.1785/0120050067.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples, Biometrika, 76, 297–307, doi:10.1093/biomet/76.2.297.

Kagan, Y. Y. (1999). Universality of the seismic moment-frequency relation, Pure Appl. Geophys. 155, 537–573.

Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. Statistical results, Geophys. J. Int. 148, 520–541.

Kagan, Y. Y. (2003). Accuracy of modern global earthquake catalogs, Phys. Earth Planet. In. 135, 173–209, doi: 10.1016/S0031-9201(02)00214-5.

Marzocchi, W., Spassiani, I., Stallone, A., & M. Taroni (2020). How to be fooled searching for significant variations of the b-value, Geophys. J. Int., 220, 1845–1856. doi: 10.1093/gji/ggz541

Mignan, A., & Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis, doi: 10.5078/corssa-00180805.

Ogata, Y. (1981). On Lewis' simulation method for point processes, IEEE Trans. Inform. Theory, IT-27, 23-31. doi: 10.1109/TIT.1981.1056305.

Ogata, Y. (1988), Statistical models for earthquake occurrences and residual analysis for point processes, J. Am. Stat. Assoc., 83, 9 – 27, doi:10.2307/2288914.

Ogata, Y., & Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, Geophys. J. Int., 113, 727-738. https://doi.org/10.1111/j.1365-246X.1993.tb04663.x.

Petruccelli, A., Vannucci, G., Lolli, B., & Gasperini, P. (2018). Harmonic fluctuation of the slope of the frequency–magnitude distribution (b-value) as a function of the angle of rake. Bull. Seismol. Soc. Am. 108. https://doi.org/10.1785/0120170328.

Petruccelli, A., Schorlemmer, D., Tormann, T., Rinaldi, A.P., Wiemer, S., Gasperini, P., & Vannucci, G. (2019a). The influence of faulting style on the size-distribution of global earthquakes, Earth Plan. Sc. Lett., 527, 115791, https://doi.org/10.1016/j.epsl.2019.115791.

Petruccelli, A., Gasperini, P., Tormann, T., Schorlemmer, D., Rinaldi, A. P., Vannucci, G., & Wiemer, S. (2019b). Simultaneous dependence of the earthquake-size distribution on faulting style and depth. Geophy. Research Lett., 4, 11,044–11,053. https://doi.org/10.1029/2019GL083997

Shi, Y., & Bolt, B. A. (1982). The standard error of the magnitude–frequency b-value, Bull. Seismol. Soc. Am. 72, 1677–1687. https://doi.org/10.1785/BSSA0720051677.

Spada, M., Tormann, T., Wiemer, S., & Enescu, B. (2013). Generic dependence of the frequency-size distribution of earthquakes on depth and its relation to the strength profile of the crust. Geophysical Research Letters, 40, 709–714. https://doi.org/10.1029/2012GL054198

Scholz, C. H. (1968). The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes, Bull. Seismol. Soc. Am. 58, 399–415.

Scholz, C. H. (2015). On the stress dependence of the earthquake b value, Geophys. Res. Lett. 42, 1399–1402, doi: 10.1002/2014GL062863.

Schorlemmer, D., Wiemer, S., & Wyss, M. (2005). Variations in earthquake- size distribution across different stress regimes, Nature 437, 539–542, doi: 10.1038/nature04094.

Tinti, S. & Mulargia, F. (1986). On the estimation of the geometric distribution parameter in seismology, Statistica, 46, 163-188.

Tinti, S. & Mulargia, F. (1987). Confidence intervals of b values for grouped magnitudes. Bull. Seism. Soc. Am., 77, 2125-2134.

Utsu, T. (1961). A statistical study of the occurrence of aftershocks, Geophys. Mag. 30, 521–605

Utsu, T. (1965). A method for Determining the Value of *b* in a Formula log *n* =*a*-*bM* showing the Magnitude-Frequency Relation for Earthquakes, Geophys. Bull. Hokkaido Univ. 13, 99-103, (in Japanese with English summary). https://doi.org/10.14943/gbhu.13.99.

Utsu, T. (1966). A statistical significance test of the difference in b-value between two earthquake groups, J. Phys. Earth 14, 34–37. https://doi.org/10.4294/jpe1952.14.37.

van der Elst, N. J. (2021). B-positive: A robust estimator of aftershock magnitude distribution in transiently incomplete catalogs. Journal of Geophysical Research: Solid Earth, 126, e2020JB021027. https://doi. org/10.1029/2020JB021027.

Wiemer, S., & Wyss. M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan, Bull. Seismol. Soc. Am. 90, no. 4, 859–869, doi: 10.1785/0119990114.

Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, Bull. Seismol. Soc. Am. 95, 684–698.

Appendix A - The continuous and discrete exponential distributions

Earthquake magnitudes, when taken as random variables, are supposed to follow an exponential distribution at least beyond a certain (completeness) magnitude threshold M_c. Generally, they are provided up to a few decimal digits (usually one) and therefore they can be naturally binned in classes of equal size, say 2δ. In the common practice, they are treated either as a discrete set of variables or as a continuous set. In the former case, if M_0 is the magnitude of the first class, the magnitude of the $i-th$ class is given by:

$$M_i = M_0 + 2\delta i \qquad (A1)$$

The integer i identifying the class is a discrete random variable obeying the probability distribution:

$$P_i = A(\alpha)e^{-\alpha i} \quad i = 0,1,2,\ldots \qquad (A2)$$

where α is assumed to be a positive decay parameter. Because P_i represents a probability for the random variable i, its distribution must satisfy the normalization condition, i.e. the sum of all probabilities must be equal to 1. By imposing it, we obtain:

$$\sum_{i=0}^{\infty} A(\alpha)e^{-\alpha i} = A(\alpha)\sum_{i=0}^{\infty} e^{-\alpha i} = \frac{A(\alpha)}{1-e^{-\alpha}} = 1 \qquad (A3)$$

It follows that $(A2)$ can be rewritten as:

$$P_i = (1-e^{-\alpha})\, e^{-\alpha i} \quad i = 0,1,2,\ldots \qquad (A4)$$

On the other hand, when treating the magnitude M as a continuous variable, its probability density function is given by:

$$P(M) = \beta e^{-\beta(M-M_c)} \quad M - M_c \geq 0 \qquad (A5a)$$

or

$$P(M) = \beta e^{-\beta(M-M_0+\delta)} \quad M - (M_0 - \delta) \geq 0 \qquad (A5b)$$

depending on the decay factor β. The formula $(A5b)$ is justified since, usually, the first value of the discrete set of magnitudes M_0 is taken as the midpoint of the first magnitude class, i.e. the one with the lower endpoint in M_c. This means that:

$$M_c = M_0 - \delta \tag{A6}$$

It can be shown that the decay factors α and β of the discrete and continuous distributions are linked by the relation:

$$\alpha = 2\delta\beta \tag{A7}$$

Indeed, if we consider the scaled variable:

$$y = \frac{M - M_c}{2\delta} \tag{A8}$$

then its probability density has the form:

$$P(y) = P(M)\frac{dM}{dy} = 2\delta\beta\, e^{-2\delta\beta y} = \alpha e^{-\alpha y} \quad y \geq 0 \tag{A9}$$

If we take only integer values of y, then the y axis results to be discretized with unitary bins, while the M axis happens to be discretized with a resolution that is finer and finer as 2δ is made smaller and smaller. Under these circumstances, the expression of P_i tends to the continuous counterpart $P(M)$, since the factor $(1 - e^{-\alpha})$ can be approximated by $2\delta\beta$.

In the following, the random variables we will consider are the continuous variable y defined in $(A8)$ and the discrete variable i defined in $(A1)$. We will see that all statistical formulas we will derive for the discrete variable i will tend to the corresponding formulas of the continuous variable y as the bin size 2δ becomes increasingly small. More specifically, if we approximate $e^{-\alpha}$ with 1 and $(1 - e^{-\alpha})$ with $2\delta\beta$, then the discrete-case expressions transform into the continuous-case ones.

Mean, variance and standard deviation

The formulas for the mean and variance of the continuous exponential distribution $(A9)$ are well known and will be given here for the sake of completeness. They are:

$$\mu_{CE} = \frac{1}{\alpha} \tag{A10}$$

$$var_{CE} = \frac{1}{\alpha^2}; \;\; \sigma_{CE} = \frac{1}{\alpha} \tag{A11}$$

where the subscript CE denotes a continuous exponential random variable. As regards the discrete distribution $(A4)$, we start with computing its mean μ_{DE} that is defined as:

$$\mu_{DE} = \sum_{i=0}^{\infty} i\, P_i = (1 - e^{-\alpha}) \sum_{i=1}^{\infty} i\, e^{-\alpha i} \tag{A12}$$

In order to compute the sum S_1 of the series in $(A12)$, we note that:

$$S_1 = \sum_{i=1}^{\infty} i\, e^{-\alpha i} = e^{-\alpha} \sum_{i=0}^{\infty} (i+1)\; e^{-\alpha i} = e^{-\alpha} \left(\sum_{i=0}^{\infty} i\, e^{-\alpha i} + \sum_{i=0}^{\infty} e^{-\alpha i} \right) \tag{A13}$$

In the last expression, we recognize that the first summation is S_1, while the second one is the sum of a geometric series. Therefore, the equation $(A13)$ becomes:

$$S_1 = e^{-\alpha} S_1 + \frac{e^{-\alpha}}{1 - e^{-\alpha}} \tag{A14}$$

This is an equation in the unknown S_1 with solution:

$$S_1 = \frac{e^{-\alpha}}{(1 - e^{-\alpha})^2} \tag{A15}$$

On substituting this expression in the definition $(A12)$, we eventually get:

$$\mu_{DE} = (1 - e^{-\alpha}) S_1 = \frac{e^{-\alpha}}{1 - e^{-\alpha}} \tag{A16}$$

The second moment, say $M_{2,DE}$, of the discrete exponential distribution is by definition:

$$M_{2,DE} = \sum_{i=0}^{\infty} i^2 P_i = (1 - e^{-\alpha}) \sum_{i=1}^{\infty} i^2 e^{-\alpha i} \tag{A17}$$

To evaluate the sum of the series, we can follow a procedure analogous to the one used earlier, that is:

$$S_2 = \sum_{i=1}^{\infty} i^2 e^{-\alpha i} = e^{-\beta} \sum_{i=0}^{\infty} (i+1)^2 e^{-\alpha i} = e^{-\alpha} \left(\sum_{i=0}^{\infty} i^2 e^{-\alpha i} + 2 \sum_{i=0}^{\infty} i\, e^{-\alpha i} + \sum_{i=0}^{\infty} e^{-\alpha i} \right) \tag{A18}$$

Remembering the values of the series in the last member of the above equation chain, we obtain the following equation in the unknown S_2:

$$S_2 = e^{-\alpha} \left(S_2 + 2 S_1 + \frac{1}{1 - e^{-\alpha}} \right) \tag{A19}$$

Its solution is:

$$S_2 = \frac{e^{-\alpha}}{1-e^{-\alpha}}\left(2S_1 + \frac{1}{1-e^{-\alpha}}\right) = \frac{e^{-\alpha}}{1-e^{-\alpha}}\left(\frac{2e^{-\alpha}}{(1-e^{-\alpha})^2} + \frac{1}{1-e^{-\alpha}}\right)$$

$$= \frac{e^{-\alpha}(1+e^{-\alpha})}{(1-e^{-\alpha})^3} \qquad (A20)$$

From the definition $(A17)$, we then obtain:

$$M_{2,DE} = (1-e^{-\alpha})S_2 = \frac{e^{-\alpha}(1+e^{-\alpha})}{(1-e^{-\alpha})^2} \qquad (A21)$$

The variance of a distribution can be computed from the mean and the second moment according to the formula:

$$var_{DE} = M_{2,DE} - (\mu_{DE})^2 \qquad (A21)$$

After substituting the respective values, it transforms to:

$$var_{DE} = \frac{e^{-\alpha}(1+e^{-\alpha})}{(1-e^{-\alpha})^2} - \frac{e^{-2\alpha}}{(1-e^{-\alpha})^2} = \frac{e^{-\alpha}}{(1-e^{-\alpha})^2}$$

$$= \frac{1}{4}\left(\operatorname{csch}\frac{\alpha}{2}\right)^2 \qquad (A23)$$

Consequently, the standard deviation σ_{DE} takes the form:

$$\sigma_{DE} = \frac{e^{-\frac{\alpha}{2}}}{1-e^{-\alpha}} = \frac{1}{2}\operatorname{csch}\frac{\alpha}{2} \qquad (A24)$$

We point out here that the formula $(A16)$ for the mean μ_{DE} is already known in the seismological literature and forms the basis for the estimator of b (8) given in the main text (Guttorp and Hopkins, 1986, Tinti and Mulargia, 1987, van der Elst, 2021). On the contrary, the expression $(A24)$ for the standard deviation of the discrete distribution is original and first derived in this paper.

Appendix B – The continuous and discrete distributions of the differences of exponential variables (Laplace distributions)

If we consider the scaled random variables y and z, both following the exponential distribution $(A9)$, then the random variable $w = z - y$ can be proven to obey the continuous Laplace distribution with density function defined as:

$$P(w) = \frac{\alpha}{2} e^{-\alpha|w|} \quad -\infty < w < +\infty \tag{B1}$$

It is a continuous density function with two symmetric exponential tails, and the same decay parameter α as the original distributions and it is known as Laplace distribution.

Let us now consider the differences of integer random variables i and j, both following the discrete exponential distribution $(A4)$ with the same parameter α. If they are independent variables, the joint probability distribution P_{ij} for the pair (i, j) is given by the product:

$$P_{i,j} = P_i P_j = (1 - e^{-\alpha})^2 e^{-\alpha(i+j)} \tag{B2}$$

Here the aim is to compute the probability that the difference takes a given value d. To this purpose, we have to sum up the probabilities of all the pairs where the difference is exactly equal to d. If we consider the Cartesian plane where i runs along the horizontal axis and j runs along the vertical axis, then the pairs exhibiting a constant difference between j and i can be found on straight lines parallel to the bisector of the first quadrant. Exactly on the bisector, the pairs have $i = j$ and the difference is identically zero. For the lines above the bisector, the difference is positive, whereas for the parallel lines lying below it, it is negative.

More formally, we introduce the random variable $d = j - i$, $d \in Z$, and compute its distribution P_d. First, we assume that $j \geq i$, and therefore that $d \geq 0$. Given d, all pairs (i, j) having difference equal to d, are of the type $(i, i + d)$ with $i \in N$. It follows that:

$$P_d = \sum_{i=0}^{\infty} P_i P_{i+d} = (1 - e^{-\alpha})^2 e^{-\alpha(i+i+d)} = (1 - e^{-\alpha})^2 e^{-\alpha d} \sum_{i=0}^{\infty} e^{-2\alpha i} \tag{B3}$$

Considering that the terms to be summed can be seen as the elements of a geometric series with constant ratio $e^{-2\alpha}$, we obtain the expression:

$$P_d = \frac{(1-e^{-\alpha})^2}{1-e^{-2\alpha}} e^{-\alpha d}, \;\; d \geq 0 \qquad (B4)$$

When $j \leq i$, following an analogous procedure, we can get a similar expression. Indeed, we should sum up all probabilities of the pairs $(j + |d|, j)$ getting the result:

$$P_d = \frac{(1-e^{-\alpha})^2}{1-e^{-2\alpha}} e^{-\alpha |d|}, \;\; d < 0 \qquad (B5)$$

Remembering the identity $1 - e^{-2\alpha} = (1 - e^{-\alpha})(1 + e^{-\alpha})$, both expressions $(B4)$ and $(B5)$ can be simplified to:

$$P_d = \frac{1-e^{-\alpha}}{1+e^{-\alpha}} e^{-\alpha |d|} = \tanh\frac{\alpha}{2} e^{-\alpha |d|} \quad -\infty < d < +\infty \qquad (B6)$$

where recourse is made to the identity:

$$\frac{1-e^{-\alpha}}{1+e^{-\alpha}} = \tanh\frac{\alpha}{2} \qquad (B7)$$

In the following the distribution $(B6)$ will be referenced to as discrete Laplace distribution.

<u>Mean, variance and standard deviation</u>

The computation of the mean of the continuous Laplace distribution μ_{CL} is straightforward, since $P(w) = P(-w)$). Here the subscript CL stands for <u>c</u>ontinuous <u>L</u>aplace distribution. Indeed, it is trivial to see that:

$$\mu_{CL} = \int_{-\infty}^{0} wP(w)dw + \int_{0}^{+\infty} wP(w)dw = -\int_{0}^{+\infty} wP(-w)dw \int_{0}^{+\infty} wP(w)dw = 0 \qquad (B8)$$

Owing to the vanishing of μ_{CL}, the second moment of the Laplace distribution $(B1)$ coincides with its variance:

$$var_{CL} = \int_{-\infty}^{+\infty} w^2 P(w)dw = 2\int_{0}^{+\infty} w^2 P(w)dw = \alpha \int_{0}^{+\infty} w^2 e^{-\alpha w} dw = \frac{2}{\alpha^2} \qquad (B9)$$

Hence, the corresponding standard deviation is:

$$\sigma_{CL} = \frac{\sqrt{2}}{\alpha} \tag{B10}$$

The mean μ_{DL} of the discrete distribution $(B6)$ is zero due to its symmetry around the origin (i.e. $P_{-d} = P_d$). Indeed:

$$\mu_{DL} = \sum_{d=-\infty}^{-1} d\, P_d + \sum_{d=1}^{\infty} d\, P_d = -\sum_{d=1}^{\infty} d\, P_{-d} + \sum_{d=1}^{\infty} d\, P_d = 0 \tag{B11}$$

And, as a consequence, its second moment and variance are coincident:

$$var_{DL} = \sum_{d=-\infty}^{\infty} d^2\, P_d = 2\sum_{d=1}^{\infty} d^2\, P_d = 2\frac{1-e^{-\alpha}}{1+e^{-\alpha}}\sum_{d=1}^{\infty} d^2 e^{-\alpha d} = 2\frac{1-e^{-\alpha}}{1+e^{-\alpha}} S_2 = \frac{2e^{-\alpha}}{(1-e^{-\alpha})^2}$$

$$= \frac{1}{2}\left(\operatorname{csch}\frac{\alpha}{2}\right)^2 \tag{B12}$$

The corresponding standard deviation results to be:

$$\sigma_{DL} = \frac{\sqrt{2}e^{-\frac{\alpha}{2}}}{1-e^{-\alpha}} = \frac{1}{\sqrt{2}}\operatorname{csch}\frac{\alpha}{2} \tag{B13}$$

On comparing expressions $(A11)$ with $(B9)$ and $(A23)$ with $(B12)$, it is worth noting that the variances of the Laplace distributions are twice larger than the corresponding variances of the exponential distributions, i.e.:

$$var_{CL} = 2var_{CE} \quad var_{DL} = 2var_{DE} \tag{B14}$$

Indeed, the results reached in this section could be anticipated simply by remembering that the mean and the variance of the difference of two independent random variables are respectively the difference of their mean and the sum of their variances, which entails that the resulting mean is zero and the resulting variance is twice as large.

Appendix C – The continuous and discrete one-sign differences distributions

If we restrict the attention only to one-sign differences, it is trivial to see that their distribution is exponential. Indeed, for the continuous case, the distribution $(B1)$ becomes:

$$P(w) = \alpha e^{-\alpha|w|} \quad -\infty < w \leq 0 \tag{C1a}$$

$$P(w) = \alpha e^{-\alpha w} \quad 0 < w < +\infty \tag{C1b}$$

Likewise, for the discrete case, the distribution $(B6)$ splits into:

$$P_d = (1 - e^{-\alpha})e^{-\alpha|d|} \quad -\infty < d \leq 0 \tag{C2a}$$

$$P_d = (1 - e^{-\alpha})e^{-\alpha d} \quad 0 \leq d < +\infty \tag{C2b}$$

It follows that the corresponding means, variances and standard deviation have the expressions $(A10)$ and $(A11)$ given in the Appendix A.

It is worth stressing that the result regarding the means was derived first by van der Elst (2021).

Appendix D – The continuous and discrete absolute differences distributions

Let us consider the absolute values of the differences, that are $|w|$ and $|d|$ respectively. It is worth outlining that for the continuous case the distribution is exponential, while for the discrete variables this is not true. In the former case, we can write:

$$P(|w|) = \alpha e^{-\alpha|w|} \; 0 \leq |w| \leq +\infty \tag{D1}$$

On the other hand, for the discrete variable $|d|$, we should distinguish the case of null differences from the others, and their probability distributions results to be:

$$P_0 = \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \tag{D2a}$$

$$P_{|d|} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} e^{-\alpha|d|} \quad 1 \leq |d| < +\infty \tag{D2b}$$

<u>Mean, variance and standard deviation</u>

The absolute values of the continuous differences are exponential variables and their statistical moments that are relevant in our context can be taken from the expressions displayed in the Appendix A. We can write them explicitly here below:

$$\mu_{CA} = \frac{1}{\alpha}\, , var_{CA} = \frac{1}{\alpha^2}\, , \sigma_{CA} = \frac{1}{\alpha} \tag{D3}$$

In the adopted notation the subscript CA stands for continuous absolute differences. The mean of the absolute values of the discrete differences is by definition given by:

$$\mu_{DA} = \sum_{|d|=1}^{\infty} |d| P_{|d|} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \sum_{i=1}^{\infty} i\, e^{-\alpha i} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} S_1 = \frac{2e^{-\alpha}}{(1 + e^{-\alpha})(1 - e^{-\alpha})} = \tag{D4}$$

In the above computations use has been made of the expression $(A14)$ for S_1.

Likewise, the second moment $M_{2,DA}$ is computed as:

$$M_{2,DA} = \sum_{|d|=1}^{\infty} |d|^2 P_{|d|} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \sum_{i=1}^{\infty} i^2\, e^{-\alpha i} = 2\frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} S_2 = \frac{2e^{-\alpha}}{(1 - e^{-\alpha})^2} \tag{D5}$$

where S_2 has been taken from $(A20)$. The variance can be computed as the difference between $M_{2,DA}$ and μ_{DA} squared, that is as:

$$var_{DA} = \frac{2e^{-\alpha}}{(1-e^{-\alpha})^2} - \frac{2e^{-2\alpha}}{(1+e^{-\alpha})^2(1-e^{-\alpha})^2} = \frac{2e^{-\alpha}(1+e^{-2\alpha})}{(1+e^{-\alpha})^2(1-e^{-\alpha})^2} \qquad (D6)$$

The related standard deviation is therefore given by:

$$\sigma_{DA} = \frac{\sqrt{2e^{-\alpha}(1+e^{-2\alpha})}}{(1+e^{-\alpha})(1-e^{-\alpha})} \qquad (D7)$$

It is relevant to observe that the variance of the absolute differences $(D6)$ is smaller than the variance of the discrete Laplace distribution $(B12)$, i.e.:

$$var_{DA} = var_{DL}\frac{1+e^{-2\alpha}}{(1+e^{-\alpha})^2} < var_{DL} \qquad (D8a)$$

since the adjusting factor is smaller than 1. Similarly, we can conclude that:

$$\sigma_{DA} < \sigma_{DL} \qquad (D8b)$$

Appendix E - The effect of trimming

For trimming we mean here the removal of all values below a predefined limit. Therefore, for the continuous variable y, we will consider only values $y \geq y' > 0$, and, likewise, for the continuous difference w we will take into account only values $w \geq w' > 0$ or $w \leq -w' < 0$. It is very easy to see that the distribution of y follows the exponential distribution:

$$P(y) = \alpha e^{-\alpha(y-y')} \quad y \geq y' \tag{E1}$$

We observe that trimming acts on the random variable y as a shift, which implies that the mean is incremented by an amount equal to the shift value, i.e.:

$$\mu_{T,CE} = \mu_{CE} + y' = \frac{1}{\alpha} + y' \tag{E2a}$$

while variance and standard deviation remain unchanged:

$$var_{T,CE} = var_{CE} = \frac{1}{\alpha^2} \qquad \sigma_{T,CE} = \sigma_{CE} = \frac{1}{\alpha} \tag{E2b}$$

Here the additional subscript T denotes the trimmed distribution. Further, it is immediate to observe that also the one-sign differences $w - w'$ and the absolute differences $|w - w'|$ follow an exponential distribution, that is:

$$P(w) = \alpha e^{-\alpha(w-w')} \quad w \geq w' > 0 \tag{E3a}$$

$$P(w) = \alpha e^{-\alpha|w-w'|} \quad w \leq -w' < 0 \tag{E3b}$$

$$P(|w|) = \alpha e^{-\alpha|w-w'|} \quad |w| \geq w' > 0 \tag{E3c}$$

Therefore, even for these distributions the mean results to be shifted by an amount equal to w', whereas variance and standard deviation do not change.

When considering the continuous Laplace distribution, appropriate for the differences, trimming is realized by considering the variables with absolute value larger than the threshold. The related density function is split into:

$$P(w) = \frac{\alpha}{2} e^{-\alpha(w-w')} \quad w \geq w' > 0 \tag{E4a}$$

$$P(w) = \frac{\alpha}{2} e^{-\alpha|w-w'|} \quad w \leq -w' < 0 \tag{E4b}$$

It is symmetric, centered in zero, and therefore, if we denote its mean by $\mu_{T,CL}$, we can write:

$$\mu_{T,CL} = 0 \tag{E5}$$

As for the variance, it identifies with the second moment and can be written as:

$$var_{T,CL} = 2\int_{w'}^{\infty} w^2 P(w) dw = \alpha \int_{w'}^{\infty} w^2 e^{-\alpha(w-w')} dw = \alpha e^{\alpha w'} \int_{w'}^{\infty} w^2 e^{-\alpha w} dw \tag{E6}$$

After a double integration by parts, the integral in the RHS can be computed analytically and $(E6)$ takes the form:

$$var_{T,CL} = \frac{1}{\alpha^2}[(1+\alpha w')^2 + 1] \tag{E7a}$$

with the corresponding standard deviation:

$$\sigma_{T,CL} = \frac{1}{\alpha}\sqrt{(1+\alpha w')^2 + 1} \tag{E7b}$$

Both expressions tend to the respective values $(B9)$ and $(B10)$ of the untrimmed distributions as w' tends to zero, i.e.:

$$var_{T,CL} \to var_{CL} \text{ and } \sigma_{T,CL} \to \sigma_{CL} \text{ as } w' \to 0 \tag{E8}$$

Notice further that both are increasing functions of w'.

As regards the discrete distributions, trimming is realized by considering only variables beyond specified integer thresholds, say i' and d'. Even in this case, the trimmed exponential distributions and the one-sign differences are exponential, i.e.:

$$P_i = (1-\alpha)e^{-\alpha(i-i')} \quad i \geq i' > 0 \qquad \text{positive differences} \tag{E9}$$

$$P_d = (1-\alpha)e^{-\alpha|d-d'|} \quad d \leq -d' < 0 \qquad \text{negative differences} \tag{E10a}$$

$$P_d = (1-\alpha)e^{-\alpha(d-d')} \quad d \geq d' > 0 \qquad \text{absolute differences} \tag{E10b}$$

So means are affected by trimming, whereas variances and standard deviations are not. For instance, for the exponential distribution we can write:

$$\mu_{T,DE} = \mu_{DE} + i' = \frac{e^{-\alpha}}{1-e^{-\alpha}} + y' \tag{E11a}$$

$$var_{T,DE} = var_{DE} = \frac{e^{-\alpha}}{(1-e^{-\alpha})^2} \quad \sigma_{T,DE} = \sigma_{DE} = \frac{e^{-\frac{\alpha}{2}}}{1-e^{-\alpha}} \tag{E11b}$$

We observe that trimming affects substantially the distribution of the absolute differences. Indeed, since trimming discards the value $d = 0$, the resulting distribution becomes exponential. It is worth to write it down explicitly:

$$P_{|d|} = (1-\alpha)e^{-\alpha(|d|-d')} \quad |d| \geq d' > 0 \tag{E12}$$

Its relevant statistical indices are quite different from the ones of the untrimmed distribution (see expressions $(D4)$, $(D6)$ and $(D7)$). They are:

$$\mu_{T,DA} = \frac{e^{-\alpha}}{1-e^{-\alpha}} + d' \quad var_{T,DA} = \frac{e^{-\alpha}}{(1-e^{-\alpha})^2} \quad \sigma_{T,DA} = \frac{e^{-\frac{\alpha}{2}}}{1-e^{-\alpha}} \tag{E13}$$

For the differences distributed according to the discrete Laplace distribution, trimming leads to the following expression for the probabilities:

$$P_d = \frac{1-e^{-\alpha}}{1+e^{-\alpha}} e^{-\alpha(|d|-d')} = \tanh\frac{\alpha}{2} e^{-\alpha(|d|-d')} \quad d \leq -d' < 0 \ \text{ or } \ d \geq d' > 0 \tag{E14}$$

It is a symmetric distribution with mean equal to zero, i.e.:

$$\mu_{T,DL} = 0 \tag{E15}$$

Its variance, being equal to its second moment, can be computed as:

$$var_{T,DL} = 2\sum_{d=d'}^{\infty} d^2 P_d = 2\frac{1-e^{-\alpha}}{1+e^{-\alpha}} e^{\alpha d'} \sum_{d=d'}^{\infty} d^2 e^{-\alpha d} = 2\frac{1-e^{-\alpha}}{1+e^{-\alpha}} e^{\alpha d'} \sum_{j=0'}^{\infty} (j+d')^2 e^{-\alpha j} \tag{E16}$$

The summation in the RHS of the last equation can be further elaborated:

$$\sum_{j=0'}^{\infty} (j+d')^2 e^{-\alpha j} = \sum_{j=0'}^{\infty} j^2 e^{-\alpha j} + 2d' \sum_{j=0'}^{\infty} j e^{-\alpha j} + d'^2 \sum_{j=0'}^{\infty} e^{-\alpha j}$$

$$= S_2 + 2d' S_1 + \frac{d'^2}{1-e^{-\alpha}} \tag{E17}$$

Combining $(E16)$ and $(E17)$ and remembering the expressions $(A15)$ and $(A20)$ respectively for S_1 and S_2, after some calculations we eventually get:

$$var_{T,DL} = \frac{2}{(1+e^{-\alpha})(1-e^{-\alpha})^2}\{e^{-\alpha} + [e^{-\alpha} + d'(1-e^{-\alpha})]^2\} \qquad (E18a)$$

$$\sigma_{T,DL} = \frac{1}{1-e^{-\alpha}}\sqrt{\frac{2}{(1+e^{-\alpha})}\{e^{-\alpha} + [e^{-\alpha} + d'(1-e^{-\alpha})]^2\}} \qquad (E18b)$$

It is worth noting that when d' is set equal to zero, both the above expressions transform into the corresponding untrimmed variables indices, that is var_{DL} and σ_{DL}.

Appendix F – Estimating the decay parameters by means of the mean method

For magnitudes obeying the Gutenberg-Richter formula (1) the decay parameter is b. If we opt for the canonical exponential expressions $(A5)$, the decay parameter is β. If we consider binned magnitudes, the decay parameter is α. Since these three parameters are linked by constant factors, we can estimate any one of them and very easily deduce the others. In this paper, the main attention goes to sequences of binned magnitudes and therefore here we concentrate on methods suitable to estimate α and only on discrete distributions. In this Appendix we will consider methods based on the mean value of the distributions. If we denote the generic mean by μ, and if it happens to depend on α, that is if $\mu = f(\alpha)$, then we can obtain α by means of the expression $\alpha = f^{-1}(\mu)$ where f^{-1} is the inverse function of f, provided that the inverse function exists. On the other hand, the mean of any distribution can be estimated from experimental data, and approximated by the sample mean value, the approximation being better and better as the data number N increases. The goodness of the estimate of μ reflects directly on how good the estimate of α is and will be treated later when addressing the confidence intervals. With this strategy in mind, we will consider separately the distributions treated so far, pointing out, however, that the method cannot be applied to the discrete Laplace distributions, either trimmed or untrimmed, because their mean μ_{DL} and $\mu_{T,DL}$ are identically zero, and thus not depending on α.

<u>Estimates based on exponential distributions</u>

The exponential distribution applies to binned trimmed or untrimmed magnitudes, as well as to binned trimmed or untrimmed one-sign magnitude differences, and also to binned trimmed absolute differences. As already stated, the untrimmed absolute differences follow a different distribution and will be addressed separately. In all these cases the formula for the mean can be written as (see Appendix E):

$$\mu = \frac{e^{-\alpha}}{1 - e^{-\alpha}} + k \qquad (F1)$$

where k is the trimming threshold and is equal to zero for untrimmed distributions.

The expression $(F1)$ can be inverted easily and leads to:

$$\alpha = \ln\left(\frac{\mu - k + 1}{\mu - k}\right) \tag{F2}$$

Interestingly, we can observe that the ratio in the formula $(F2)$ can be written also as:

$$\frac{\mu - k + 1}{\mu - k} = \frac{2\left(\mu - k + \frac{1}{2}\right) + 1}{2\left(\mu - k + \frac{1}{2}\right) - 1} = \frac{x + 1}{x - 1} \tag{F3a}$$

where we have posed:

$$x = 2\left(\mu - k + \frac{1}{2}\right) \tag{F3b}$$

Taking advantage of the identity:

$$\coth^{-1}(x) = \frac{1}{2}\ln\left(\frac{x + 1}{x - 1}\right) \tag{F4}$$

that links the natural logarithm with the inverse of the hyperbolic cotangent, the variable α in $(F2)$ can be alternatively given also as:

$$\alpha = 2\coth^{-1}\left(2\left(\mu - k + \frac{1}{2}\right)\right) \tag{F5}$$

We stress that in the above formulas α is the true value of the decay parameter. Therefore we can interpret formulas $(F2)$ and $(F5)$ as unbiased estimator of α, say $\tilde{\alpha}$, if we replace μ with its sample mean, since the sample mean tends to μ when the number of data in the sample increases.

In terms of binned magnitudes M_i given by $(A1)$ the above formulas $(F2)$ and $(F5)$ for the estimator $\tilde{\alpha}$ take the form:

$$\tilde{\alpha} = \ln\left(\frac{\bar{M} - M_k + 2\delta}{\bar{M} - M_k}\right) = 2\coth^{-1}\left(\frac{1}{\delta}(\bar{M} - M_k + \delta)\right) \tag{F6}$$

where $\bar{M}$ is the sample mean magnitude and

$$M_k = M_0 + 2\delta k \quad k \geq 0 \tag{F7}$$

is defined as the trimming threshold magnitude which coincides with the magnitude of the lowest bin if no trimming is applied.

Since $\alpha = 2\delta b \ln(10)$, thus the corresponding estimator of the decay parameter b is:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\bar{M} - M_k + 2\delta}{\bar{M} - M_k}\right) = \frac{1}{\delta \ln(10)} \coth^{-1}\left(\frac{1}{\delta}(\bar{M} - M_k + \delta)\right) \qquad (F8)$$

Observe that the expressions $(F8)$ coincide with the estimators (8) and (12) in the main text, where however we used a different notation and called M_k as M_c.

The logarithmic version of the above formula can be rewritten as:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(1 - \frac{2\delta}{\bar{M} - M_k}\right) \qquad (F9)$$

This is a version expandable in series. If we truncate the expansion to the second order, we obtain:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \left[\frac{2\delta}{\bar{M} - M_k} - \frac{1}{2}\frac{4\delta^2}{(\bar{M} - M_k)^2}\right] = \frac{1}{\ln(10)(\bar{M} - M_k)}\left(1 - \frac{\delta}{\bar{M} - M_k}\right) \qquad (F10)$$

It is interesting to observe that the formula (F10) when k=0, coincides with the first terms of the expansion of the expression (4) in the main text. Indeed:

$$\tilde{b} = \frac{1}{\ln(10)\,(\bar{M} - M_0 + \delta)} = \frac{1}{\ln(10)(\bar{M} - M_0)} \frac{1}{1 + \dfrac{\delta}{\bar{M} - M_0}} \approx \frac{1}{\ln(10)(\bar{M} - M_0)}\left(1 - \frac{\delta}{\bar{M} - M_0}\right)$$

Therefore, we can state that the estimator (4) is an approximation of the estimator for binned exponential magnitudes corrected at the second order in the variable $\delta/(\bar{M} - M_0)$.

When considering the binned magnitude differences, we come to analogous expressions for the estimator. If we denote by ΔM the magnitude differences, by $\overline{\Delta M}$ the related sample mean value, and by ΔM_k the trimming threshold, then for trimmed positive differences we obtain:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{\Delta M} - \Delta M_k + 2\delta}{\overline{\Delta M} - \Delta M_k}\right) \quad \Delta M \geq \Delta M_k = 2k\delta \quad k \geq 0 \qquad (F11)$$

For trimmed negative differences the formula is:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{|\Delta M|} - \Delta M_k + 2\delta}{\overline{\Delta M} - \Delta M_k}\right) \quad \Delta M \leq \Delta M_k = -2k\delta \quad k \geq 0 \qquad (F12)$$

Eventually for the trimmed absolute differences, we get:

$$\tilde{b} = \frac{1}{2\delta \ln(10)} \ln\left(\frac{\overline{|\Delta M|} - \Delta M_k + 2\delta}{\overline{|\Delta M|} - \Delta M_k}\right) \quad |\Delta M| \geq \Delta M_k = 2k\delta \quad k \geq 1 \qquad (F13)$$

All the above expressions $(F11) - (F13)$ can be also given in terms of the inverse hyperbolic cotangent, like in $(F8)$. They coincide with the formula given in the main text as (18), provided that we change notation replacing ΔM_k with $\Delta M'_c$.

<u>Estimates based on the untrimmed absolute differences distribution</u>

The mean μ_{DA} of the distribution of the untrimmed absolute differences is given by:

$$\mu_{DA} = \frac{2e^{-\alpha}}{(1+e^{-\alpha})(1-e^{-\alpha})} = \frac{2e^{-\alpha}}{1-e^{-2\alpha}} \qquad (F14)$$

It is an invertible function of α, as we will see. Indeed, the expression $(F14)$ can be transformed into:

$$\mu_{DA} e^{-2\alpha} + 2e^{-\alpha} - \mu_{DA} = 0 \qquad (F15a)$$

that can be interpreted as a quadratic equation in the unknown $e^{-\alpha}$, with roots:

$$e^{-\alpha} = \frac{-1 \pm \sqrt{1+{\mu_{DA}}^2}}{\mu_{DA}} = -\frac{1}{\mu_{DA}} \pm \sqrt{\frac{1}{{\mu_{DA}}^2} + 1} \qquad (F15b)$$

Of the two roots, only the positive one is an admissible solution, since the exponential in the LHS must be positive. Thus we can write:

$$\alpha = \ln\left(\frac{-1+\sqrt{1+{\mu_{DA}}^2}}{\mu_{DA}}\right)^{-1} = \ln\left(\frac{1+\sqrt{1+{\mu_{DA}}^2}}{\mu_{DA}}\right) = \ln\left(\frac{1}{\mu_{DA}} + \sqrt{\frac{1}{{\mu_{DA}}^2}+1}\right)$$

$$= \operatorname{csch}^{-1}(\mu_{DA}) \qquad (F16)$$

The last equality has been introduced in virtue of the following identity involving the natural logarithm and the inverse of the hyperbolic cosecant:

$$\operatorname{csch}^{-1}(x) = \ln\left(\frac{1}{x} + \sqrt{\frac{1}{x^2}+1}\right) \qquad (F17)$$

Substituting μ_{DA} with the sample mean, we obtain an unbiased estimator $\tilde{\alpha}$ and, in terms of the sample mean $\overline{|\Delta M|}$ of the absolute magnitude differences, we obtain an expression for $\tilde{\beta}$ that coincides with the formula (14) of the main text, i.e.:

$$\tilde{b} = \frac{1}{2\delta\ln(10)}\ln\left[\frac{2\delta + \sqrt{4\delta^2 + \left(\overline{|\Delta M|}\right)^2}}{\overline{|\Delta M|}}\right] = \frac{1}{2\delta\ln(10)}\operatorname{csch}^{-1}\left(\frac{\overline{|\Delta M|}}{2\delta}\right) \quad |\Delta M| \geq 0 \qquad (F18)$$

Appendix G - Estimating the decay parameters through the Maximum Likelihood method

The decay parameter can be estimated also by means of the Maximum Likelihood (ML) approach. As a general observation, the main conceptual difference between the ML method and the mean method is that the former applies to empirical samples, while the latter one uses relations proper of the theoretical distribution. However, provided that we replace the expected value of the distributions with the related sample means, the two methods are expected to lead to the same result. This is exactly what we will prove here, but we outline that there is an important caveat that we need to express for the estimator of the binned differences.

<u>Using samples of the discrete exponential distribution</u>

For the sake of simplicity, we will consider here only the untrimmed exponential distribution of binned magnitudes. Making recourse to the ML method, we introduce the Likelihood Function $L_N(\alpha)$ for a sample of N data $(i_1, i_2, .., i_N)$, that we write:

$$L_N(\alpha) = (1 - e^{-\alpha})^N \prod_{s=1}^{n} e^{-\alpha i_s} = (1 - e^{-\alpha})^N e^{-\alpha \sum_{s=1}^{s=N} i_s} = (1 - e^{-\alpha})^N e^{-\alpha N \bar{\iota}} \qquad i_s \geq 0 \quad (G1)$$

where $\bar{\iota}$ is the arithmetic mean of the sample.

The ML estimate of the parameter α is that value, say $\tilde{\alpha}$, that maximizes $L_N(\alpha)$ and can be found by solving the equation obtained by imposing that the derivative of $L_N(\alpha)$ with respect to α vanishes, i.e.:

$$\frac{dL_N(\alpha)}{d\alpha} = N(1 - e^{-\alpha})^{N-1} e^{-\alpha N \bar{\iota}} \frac{d}{d\alpha}\left(1 - e^{-\tilde{\alpha}}\right) - \alpha N \bar{\iota}(1 - e^{-\alpha})^N e^{-\alpha N \bar{\iota}} = 0 \qquad (G2)$$

Simplifying, we obtain that the estimator has to solve the equation:

$$\frac{d}{d\alpha}\left(1 - e^{-\tilde{\alpha}}\right) - \bar{\iota}\left(1 - e^{-\tilde{\alpha}}\right) = 0 \qquad (G3)$$

Therefore we get:

$$e^{-\tilde{\alpha}} - \bar{\iota}\left(1 - e^{-\tilde{\alpha}}\right) = 0 \qquad (G4)$$

with solution

$$\tilde{\alpha} = \ln\left(\frac{1+\bar{\iota}}{\bar{\iota}}\right) \qquad (G5)$$

that corresponds to the expression $(F2)$, once we pose $k = 0$ and substitute the theoretical mean μ with the sample mean $\bar{\iota}$.

<u>Using samples of the discrete Laplace distribution</u>

We have remarked that the Laplace distribution of the magnitude differences is unsuitable to the application of the mean method since its mean is identically zero. However, we can apply the ML method. Let us consider the Likelihood Function $L_N(\alpha)$ as the product of three functions $L_{N_+}(\alpha)$, $L_{N_-}(\alpha)$ and $L_{N_0}(\alpha)$, where $N = N_+ + N_- + N_0$ and where the subscripts denote the absolute frequencies of differences respectively greater than, smaller than and equal to zero. If we pose (see $(B6)$):

$$B(\alpha) = \tanh\frac{\alpha}{2} \qquad (G6)$$

then we can write for positive differences:

$$L_{N_+}(\alpha) = \big(B(\alpha)\big)^{N_+} e^{-\alpha\sum_{k=1}^{k=N_+} d_k} \quad d_k > 0 \qquad (G7a)$$

Analogously, for negative differences, we have:

$$L_{N_-}(\beta) = \big(B(\beta)\big)^{N_-} e^{-\beta\sum_{k=1}^{k=N_-} |d_k|} \quad d_k < 0 \qquad (G7b)$$

And for differences equal to zero:

$$L_{N_0}(\alpha) = \big(B(\alpha)\big)^{N_0} \qquad (G7c)$$

As a consequence, the Likelihood Function $L_N(\alpha)$ can be given the expression:

$$L_N(\alpha) = L_{N_+}(\alpha)\, L_{N_-}(\alpha)\, L_{N_0}(\alpha) = \big(B(\alpha)\big)^{N} e^{-\alpha N\overline{|d|}} \qquad -\infty < d < \infty \qquad (G8)$$

By imposing that the first derivative of $L_N(\alpha)$ with respect to α is equal to zero, we get the ML solving equation, that is:

$$\frac{dB(\tilde{\alpha})}{d\alpha} - \overline{|d|}B(\tilde{\alpha}) = 0 \qquad -\infty < d < \infty \qquad (G9)$$

Recalling the position $(G6)$ and recalling further the formula of the first derivative of the hyperbolic tangent, we get:

$$\frac{1}{2}\frac{1}{\left(\cosh\frac{\tilde{\alpha}}{2}\right)^2} = \overline{|d|}\tanh\frac{\tilde{\alpha}}{2} \quad -\infty < d < \infty \tag{G10}$$

After some calculations, the expression becomes:

$$\sinh\tilde{\alpha} = \frac{1}{\overline{|d|}} \quad -\infty < d < \infty \tag{G11}$$

which leads to the final expression for the estimator:

$$\tilde{\alpha} = \mathrm{csch}^{-1}\left(\overline{|d|}\right) \quad -\infty < d < \infty \tag{G12}$$

It is very important to stress that the formula $(G12)$ identifies with the formula $(F16)$ that resulted from the application of the mean method to the binned untrimmed absolute differences. Notice that if we had applied the ML approach to the distribution of the absolute differences, we would have obtained exactly the same result. So the question is: what is the distribution underlying the formula? The discrete Laplace distribution or the distribution of the absolute differences? The matter is relevant since it has an impact on the calculation of the confidence intervals. The answer can be given by observing that the formula $(G12)$ contains the mean of the absolute value of the differences, and therefore what matters are the properties of the sample mean of the untrimmed absolute differences.

Appendix H - Confidence intervals

The decay parameters α and b derived in the previous Appendix F are functions of the mean μ of a distribution of a discrete variable i with probability P_i and standard deviation σ. Let us say that:

$$p = g(\mu) \tag{H1}$$

where p denotes the parameter and g the function. The corresponding estimator $\tilde{p}$ has been computed through the same function g as:

$$\tilde{p} = g(\bar{\mu}_N) \tag{H2}$$

where $\bar{\mu}_N$ is the mean of an empirical sample of N data. The sample mean, being a linear combination of random variables, is in turn a random variable with expected value equal to μ and with standard deviation

$$\sigma_N = \frac{1}{\sqrt{N}}\,\sigma \tag{H3}$$

According to this view, the function g maps the random variable $\bar{\mu}_N$ into the random variable $\tilde{p}$. It is known that $\bar{\mu}_N$ tends to follow a Gaussian distribution $G(\mu, \sigma_N)$ as N increases, that peaks more and more around the true value μ. If we consider the one-sigma interval $I_{\mu_N} = [\bar{\mu}_N - \sigma_N, \bar{\mu}_N + \sigma_N]$, then the g-mapping induces a corresponding image interval $I_{p_N} = [p_{1,N}, p_{2,N}]$ in the estimator space, where, if the function g is monotonically decreasing as in our case, the extremes are given by:

$$p_{1,N} = g(\bar{\mu}_N + \sigma_N) \qquad p_{2,N} = g(\bar{\mu}_N - \sigma_N) \tag{H4}$$

If we call P_σ the probability that μ belongs to the interval I_{μ_N}, then, in virtue of the mapping, it results that the parameter p has the same probability to belong to the interval I_{p_N}. Formally it can be written that:

$$P_\sigma = P(\mu \in I_{\mu_N}) = P(p \in I_{p_N}) \tag{H5}$$

Further we can state that since $\bar{\mu}_N \in I_{\mu_N}$ by construction, then its image $\tilde{p} \in I_{p_N}$. Since, in general the function g is not linear, thus $\tilde{p}$ is not the midpoint of the interval. It is common practice to provide $\tilde{p}$ as the estimator of p and the extremes $p_{1,N}$ and $p_{2,N}$ of the image interval I_{p_N} as the one-sigma confidence interval. We stress that instead of $\tilde{p}$ as given in $(H2)$ it would be more correct to provide the midpoint of the image interval and its half-length as the result of the estimation process, i.e.:

$$\tilde{p}_N = \frac{1}{2}\left(p_{1,N} + p_{2,N}\right) \tag{H6}$$

$$\Delta\tilde{p}_N = \tfrac{1}{2}\left(p_{2,N} - p_{1,N}\right) \tag{H7}$$

Note that in the above formulas we assume to know σ that, through $(H3)$, would allow us to know σ_N. In practice, however, σ is not known. It could be estimated from the empirical standard deviation. Here we make the choice to estimate it as a function of the estimator $\tilde{p}$ given by $(H2)$. More specifically, the procedure to compute the confidence interval is:

1) compute $\bar{\mu}_N$ from the N-sample data;

2) calculate the estimator $\tilde{p}$;

3) obtain σ through a proper function of $\tilde{p}$, say $\sigma = \sigma(\tilde{p})$;

4) compute σ_N via $(H3)$;

5) calculate the extremes $p_{1,N}$ and $p_{2,N}$ by means of $(H4)$.

Confidence intervals for exponential distributions

As an illustrative example of the exponential distributions addressed in this paper, let us consider the discrete untrimmed exponential distribution and write the function g as:

$$\tilde{\alpha} = g(\bar{\mu}_N) = \ln\left(\frac{\bar{\mu}_N + 1}{\bar{\mu}_N}\right) \tag{H8}$$

Then we compute the standard deviation of an N-size sample in terms of the computed $\tilde{\alpha}$:

$$\sigma_N = \frac{1}{\sqrt{N}}\frac{e^{-\frac{\tilde{\alpha}}{2}}}{1 - e^{-\tilde{\alpha}}} \tag{H9}$$

The further step is to compute the extremes of the interval I_{σ_N}:

$$\bar{\mu}_N \pm \sigma_N = \frac{e^{-\tilde{\alpha}}}{1-e^{-\tilde{\alpha}}} \pm \frac{1}{\sqrt{N}}\frac{e^{-\frac{\tilde{\alpha}}{2}}}{1-e^{-\tilde{\alpha}}} = \frac{1}{c-1}\left(1 \pm \sqrt{\frac{c}{N}}\right) \qquad (H10)$$

In $(H10)$ use has been made of the inverse expression giving $\bar{\mu}_N$ in terms of $\tilde{\alpha}$. The last member of the above chain of equalities is obtained by first multiplying all numerators and denominators by the factor $c = e^{\tilde{\alpha}}$ and then by isolating the common factor $1/(c-1)$. The lower end of the interval I_{p_N} is:

$$p_{1,N} = \ln\left(\frac{\bar{\mu}_N + \sigma_N + 1}{\bar{\mu}_N + \sigma_N}\right) = \ln\left(\frac{\frac{1}{c-1}\left(1+\sqrt{\frac{c}{N}}\right)+1}{\frac{1}{c-1}\left(1+\sqrt{\frac{c}{N}}\right)}\right) = \ln\left(\frac{c+\sqrt{\frac{c}{N}}}{1+\sqrt{\frac{c}{N}}}\right) \qquad c = e^{\tilde{\alpha}} \qquad (H11)$$

Likewise, the upper end results to be:

$$p_{2,N} = \ln\left(\frac{\bar{\mu}_N - \sigma_N + 1}{\bar{\mu}_N - \sigma_N}\right) = \ln\left(\frac{c-\sqrt{\frac{c}{N}}}{1-\sqrt{\frac{c}{N}}}\right) \qquad c = e^{\tilde{\alpha}} \qquad (H12)$$

These calculations allow us to compute the one-sigma confidence interval for the decay parameter b, since we may determine the end points $b_{1,N}$ and $b_{2,N}$ as follows:

$$b_{1,N} = \frac{p_{1,N}}{2\delta\ln(10)} = \frac{1}{2\delta\ln(10)}\ln\left(\frac{c+\sqrt{\frac{c}{N}}}{1+\sqrt{\frac{c}{N}}}\right) \qquad (H13a)$$

$$b_{2,N} = \frac{p_{2,N}}{2\delta\ln(10)} = \frac{1}{2\delta\ln(10)}\ln\left(\frac{c-\sqrt{\frac{c}{N}}}{1-\sqrt{\frac{c}{N}}}\right) \qquad (H13b)$$

Here $c = e^{2\delta\ln(10)\tilde{b}} = 10^{2\delta\tilde{b}}$ and $\tilde{b}$ is given by the formula $(F8)$, that for untrimmed magnitudes is:

$$\tilde{b} = \frac{1}{2\delta\ln(10)}\ln\left(\frac{\bar{M} - M_0 + 2\delta}{\bar{M} - M_0}\right) \qquad (H14)$$

The estimated standard confidence interval can be computed as

$$\Delta b_- = \tilde{b} - \tilde{b}_{1,N} \tag{H15a}$$

and

$$\Delta b_+ = \tilde{b}_{2,N} - \tilde{b} \tag{H15b}$$

and finally

$$\Delta b = \frac{1}{2}\left(\tilde{b}_{2,N} - \tilde{b}_{1,N}\right) \tag{H15c}$$

The above formula are the ones proposed in the main text as (19) and (20).

Confidence intervals for the absolute difference distribution

The decay parameter α of the distribution of the absolute values of the differences is linked to the distribution mean through the formula $(F16)$ that therefore provides us with the function g:

$$\tilde{\alpha} = g(\bar{\mu}_N) = \ln\left(\frac{1 + \sqrt{1 + \bar{\mu}_N{}^2}}{\bar{\mu}_N}\right) = \mathrm{csch}^{-1}(\tilde{\alpha}) \tag{H16}$$

This formula was derived also by applying the ML method to the distribution of the differences, as noted before, but the standard deviation to use here is the one of the absolute differences shown in $(D7)$, while the formula $(B13)$ is unsuitable and would lead to incorrect evaluations. By using it, we can compute the sample standard deviation as:

$$\sigma_N = \frac{1}{\sqrt{N}} \frac{\sqrt{2e^{-\tilde{\alpha}}(1 + e^{-2\tilde{\alpha}})}}{(1 + e^{-\tilde{\alpha}})(1 - e^{-\tilde{\alpha}})} \tag{H17}$$

The endpoints of the interval I_{σ_N} are:

$$\bar{\mu}_N \pm \sigma_N = \frac{2e^{-\tilde{\alpha}}}{(1 + e^{-\tilde{\alpha}})(1 - e^{-\tilde{\alpha}})} \pm \frac{1}{\sqrt{N}} \frac{\sqrt{2e^{-\tilde{\alpha}}(1 + e^{-2\tilde{\alpha}})}}{(1 + e^{-\tilde{\alpha}})(1 - e^{-\tilde{\alpha}})} \tag{H18}$$

After some manipulations this formula can be given the following version:

$$\bar{\mu}_N \pm \sigma_N = \left(1 \pm \sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right) \mathrm{csch}\tilde{\alpha} \tag{H19}$$

In terms of the absolute difference magnitudes, the endpoints of the confidence interval are:

$$b_{1,N} = \frac{1}{2\delta \ln(10)} \ln\left[\frac{2\delta + \sqrt{4\delta^2 + (\mathrm{csch}\tilde{\alpha})^2 \left(1 + \sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)^2}}{\left(1 + \sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\mathrm{csch}\tilde{\alpha}}\right]$$

$$= \frac{1}{2\delta \ln(10)} \mathrm{csch}^{-1}\left(\left(1 + \sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\mathrm{csch}\tilde{\alpha}\right) \qquad (H20a)$$

$$b_{2,N} = \frac{1}{2\delta\ln(10)}\ln\left[\frac{2\delta + \sqrt{4\delta^2 + (\operatorname{csch}\tilde{\alpha})^2\left(1-\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)^2}}{\left(1-\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\operatorname{csch}\tilde{\alpha}}\right]$$

$$= \frac{1}{2\delta\ln(10)}\operatorname{csch}^{-1}\left(\left(1-\sqrt{\frac{\cosh\tilde{\alpha}}{N}}\right)\operatorname{csch}\tilde{\alpha}\right) \qquad (H20b)$$

For convenience we repeat here the corresponding estimator of $\tilde{b}$:

$$\tilde{b} = \frac{1}{2\delta\ln(10)}\ln\left[\frac{2\delta + \sqrt{4\delta^2 + \left(\overline{|\Delta M|}\right)^2}}{\overline{|\Delta M|}}\right] = \frac{1}{2\delta\ln(10)}\operatorname{csch}^{-1}\left(\frac{\overline{|\Delta M|}}{2\delta}\right) \qquad |\Delta M| \geq 0 \qquad (H21)$$

The associated amplitude of the confidence interval is deduced like in the previous example as:

$$\Delta b = \frac{1}{2}\left(\tilde{b}_{2,N} - \tilde{b}_{1,N}\right) \qquad (H22)$$

The above formulas are not provided in the main text.

Appendix I - Simulation of complete and incomplete magnitude datasets

To generate a complete random dataset of magnitudes $M \geq M_{min}$ with exponential distribution, we use the inverse exponential transformation

$$M = -\frac{\ln\{\text{rand}]0{:}1[\}}{b\ln(10)} + M_{min} \tag{I1}$$

where rand]0: 1[is a pseudo random number with uniform distribution in the interval]0: 1[.

The binning of magnitudes is obtained by

$$M_{binned} = \text{round}\left(\frac{M}{2\delta}\right)2\delta \tag{I2}$$

where round(x) indicates the closest integer to the argument value x and 2δ is the binning size. In such case, in order the simulated dataset be complete, the latter must include magnitudes down to $M_{min} - \delta$

$$M = -\frac{\ln\{\text{rand}]0{:}1[\}}{b\ln(10)} + M_{min} - \delta \tag{I3}$$

Therefore, eq. ($I3$) is the one adopted to generate all the complete magnitude datasets in the paper.

As suggested by Ogata and Katsura (1993), magnitude data incompleteness can be mimicked by a cumulative Gaussian probability distribution with mean μ and standard deviation λ

$$P(m \leq M|\mu, \lambda) = \frac{1}{\lambda\sqrt{2\pi}} \int_{-\infty}^{M} e^{-\frac{(m-\mu)^2}{2\lambda^2}}\, dm \tag{I4}$$

In this formulation the mean μ corresponds to the threshold magnitude at which P=0.5, that means that below it, the 50% of earthquakes cannot be correctly evaluated and are lost for the frequency magnitude analyses.

It can be introduced in the simulated dataset using the thinning method (Ogata, 1981), which consists in discarding the magnitudes for which an extracted random number in the interval]0: 1[is larger than the cumulative Gaussian probability ($I4$).

Van der Elst (2021) simulated datasets with time varying incompleteness as it may occur in the first hours or days after a strong main shock. For modelling such decaying incompleteness threshold, Helmstetter et al. (2006) proposed the empirical equation

$$m_c(t) = m - 4.5 - 0.75\log_{10}t \qquad (I5)$$

where m_c is the time dependent magnitude threshold of completeness, *m* is the magnitude of the mainshock and *t* is the time elapsed since the mainshock in days. The law (*I*5) is a decreasing function of time and implies that after a single day the threshold lowers down to $m - 4.5$. Van der Elst suggested to use equation (*I*5) to set the time varying mean $\mu(t)$ in eq. (*I*4), which entails the assumption that m_c is the magnitude below which half of the earthquakes are lost.

In order to simulate the time *t* of each shock after a main shock, we assumed a simple Omori-Utsu decay law (Utsu, 1961) with equation

$$r(t) = \frac{K}{(t + c)^p} \qquad (I6)$$

where $r(t)$ is the time varying rate (in shocks per day) of a non-homogeneous Poisson process, *p* and *c* are empirical parameters and *K* is a normalization factor depending on the number of shocks and on the considered time interval. Usually, $p \approx 1$ and *c* is of the order of some tens of minutes (about 0.01 days). The time integration

$$\tau = \int_0^t r(s)ds = F(t) \qquad (I7)$$

produces a set of transformed times which follow a stationary Poisson process with intensity 1 (Ogata, 1988).

Conversely, given a set of times τ_i generated according to a stationary Poisson process with intensity 1, the inverse integral transformation

$$t = F^{-1}(\tau) \qquad (I8)$$

corresponds to a non-homogeneous Poisson process with rate $r(t)$.

Moreover, it is often useful to generate sequences of exactly *N* events over a given time interval $[0, T_E]$

$$\int_0^{T_e} r(s)ds = N \tag{I9}$$

This implies that

$$K = \begin{Bmatrix} \dfrac{N(1-p)}{(T_E + c)^{1-p} - c^{1-p}} & p \neq 1 \\ \dfrac{N}{\ln(T_E + c) - \ln(c)} & p = 1 \end{Bmatrix} \tag{I10}$$

Then, the direct timescale transform is

$$\tau = F(t) = \int_0^t r(s)ds = \begin{Bmatrix} N \dfrac{(t + c)^{1-p} - (c)^{1-p}}{(T_E + c)^{1-p} - (c)^{1-p}} & p \neq 1 \\ \mathrm{N} \dfrac{\ln(t + c) - \ln(c)}{\ln(T_E + c) - \ln(c)} & p = 1 \end{Bmatrix} \tag{I11}$$

and the inverse timescale transform is

$$t = F^{-1}(\tau) = \begin{Bmatrix} \left[\tau \dfrac{(T_E + c)^{1-p} - (c)^{1-p}}{N} + (c)^{1-p}\right]^{1/(1-p)} - c & p \neq 1 \\ \exp\left[\tau \dfrac{\ln(T_E + c) - \ln(c)}{N} + \ln(c)\right] - c & p = 1 \end{Bmatrix} \tag{I12}$$

The set of stationary Poisson times withy intensity 1 can be generated by cumulating exponentially distributed interevent times (starting from $\tau_1 = -\ln\{1 - \mathrm{rand}]0{:}1[\,\}$)

$$\tau_i = \tau_{i-1} - \ln\{1 - \mathrm{rand}]0{:}1[\,\}, \qquad i = 2, N \tag{I13}$$