

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

Paolo Gasperini<sup>1</sup> and Stefano Tinti<sup>1</sup>

<sup>1</sup>Università di Bologna

December 7, 2022

## Abstract

The estimation of the slope ( $b$ -value) of the frequency magnitude distribution of earthquakes is usually based on a formula derived decades ago under the hypothesis of continuous exponential distribution of magnitudes. However, as the magnitude is provided with a limited resolution (one decimal digit usually), its distribution is not continuous but discrete. In the literature this problem is solved mostly by applying an empirical correction to the minimum magnitude of the dataset depending on the binning size, but a recent paper recalled that this solution is only approximate and proposed an exact formula. The same paper further showed that the  $b$ -value can be estimated also by considering the positive 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 the derivation of i) 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 by the methods of the mean as well as of the maximum likelihood; and iii) the corresponding formulas for the parameter  $b$ . We further deduce iv) the standard confidence limits for the estimated  $b$ . Moreover, we are able v) to quantify the error associated with the formula including the Utsu minimum-magnitude correction. We tested such formulas on simulated synthetic datasets including cases with a certain amount of incompleteness.

## Hosted file

essoar.10512750.1.docx available at <https://authorea.com/users/563757/articles/610905-the-estimation-of-b-value-of-the-frequency-magnitude-distribution-and-of-its-confidence-intervals-from-binned-magnitude-data>

## Hosted file

Supporting\_information.docx available at <https://authorea.com/users/563757/articles/610905-the-estimation-of-b-value-of-the-frequency-magnitude-distribution-and-of-its-confidence-intervals-from-binned-magnitude-data>

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

S. Tinti<sup>1</sup>, P. Gasperini<sup>1,2</sup>

<sup>1</sup>Dipartimento di Fisica e Astronomia “Augusto Righi”, Università di Bologna, Italy

<sup>2</sup>Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy

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

Key points:

- Estimators of decay parameter of the discrete exponential and trimmed Laplace distributions by methods of the mean and maximum likelihood.
- Formulas for estimating the parameter b and the standard confidence limits for the estimated b.
- Test of such formulas using simulated set with or without incompleteness.

### **Abstract**

The estimation of the slope (b-value) of the frequency magnitude distribution of earthquakes is usually based on a formula derived decades ago under the hypothesis of continuous exponential distribution of magnitudes. However, as the magnitude is provided with a limited resolution (one decimal digit usually), its distribution is not continuous but discrete. In the literature this problem is solved mostly by applying an empirical correction to the minimum magnitude of the dataset depending on the binning size, but a recent paper recalled that this solution is only approximate and proposed an exact formula. The same paper further showed that the b-value can be estimated also by considering the positive 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 the derivation of i) 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 by the methods of the mean as well as of the maximum likelihood; and iii) the corresponding formulas for the parameter b. We further deduce iv) the standard confidence limits for the estimated b. Moreover, we are able v) to quantify the error associated with the formula including the Utsu minimum-magnitude correction. We tested such formulas on simulated synthetic datasets including cases with a certain amount of incompleteness.

### **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 discretised in equal-size bins. The performance 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.

## 1 Introduction

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

$$\log_{10} N = a + bM \quad (1)$$

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), although 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

$$\underline{\underline{b = \frac{1}{\ln(10)(\overline{M} - M_c)}}} \quad (2)$$

$$\underline{\underline{b = \frac{1}{\ln(10)(\overline{M} - M_c)}}} \quad (2)$$

$$\underline{\underline{\sigma_b = \frac{b}{\sqrt{N}}}} \quad (3)$$

where  $\overline{M}$  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

$$\underline{\underline{b = \frac{1}{\ln(10)(\overline{M} - M_c + \delta)}}} \quad (4)$$

where  $\delta$  is one half of the binning size (e.g. 0.05).

Studying in detail the statistical distribution of  $b$ , Shi and Bolt (1982) derived a more accurate formula for the confidence interval

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

Actually, if the magnitude data are binned, their distribution is not continuous anymore, but discrete and this implies changes in the estimators. Marzocchi et al. (2020) suggested that, when data are binned, the  $b$ -value computed through the Utsu formula (4) has to be corrected by

$$\underline{\underline{b_{\text{corrected}} = \frac{1}{2 \ln(10)} \ln \left[ \frac{1+b\delta}{1-b\delta} \frac{\ln(10)}{\ln(10)} \right]}} \quad (6)$$

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

$$\underline{\underline{b = \frac{1}{\ln(10)} \coth^{-1} \left[ \frac{1}{\delta} (\overline{M} - M_c + \delta) \right]}} \quad (7)$$

where  $\coth^{-1}$  is the inverse of the hyperbolic cotangent function. He also showed that the  $b$ -value can be consistently computed by the absolute magnitude dif-

ferences  $|\Delta M|$  (which follow the exponential discrete Laplace distribution) by

$$\underline{\underline{b = \frac{1}{2 \ln(10)} \operatorname{csch}^{-1} \left[ \frac{1}{2\delta} \overline{|\Delta M|} \right]}} \quad (8)$$

where  $\operatorname{csch}^{-1}$  is the inverse of the hyperbolic cosecant and  $\overline{|\Delta M|}$  is the average of the absolute magnitude differences.

We point out that, in order 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 magnitudes and then between the third and the second and so on up to the last one:

$$\underline{\underline{|M|_i = |M_{i+1} - M_i| \quad , \quad i = 1, 2, \dots, N-1}} \quad (9)$$

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 magnitudes and then between the fourth and the third and so on up to the last one

$$\underline{\underline{|M|_i = |M_{2i} - M_{2i-1}| \quad , \quad i = 1, 2, \dots, N/2}} \quad (10)$$

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. (7):

$$\underline{\underline{b = \frac{1}{\ln(10)} \operatorname{coth}^{-1} \left[ \frac{1}{\delta} \left( \overline{|\Delta M|} - \Delta M'_c + \delta \right) \right]}} \quad (11)$$

provided that  $\overline{|\Delta M|}$  and  $\Delta M'_c$  replace  $\overline{M}$  and  $M_c$  respectively. 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. (11).

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, and iii) the corresponding formulas for estimating the parameter  $b$ . Moreover, we further deduce iv) the standard one-sigma confidence limits for the estimated  $b$  as

$$\overline{b_1 = \frac{1}{2\delta \ln(10)} \ln \left[ \frac{c + \sqrt{\frac{c}{N}}}{1 + \sqrt{\frac{c}{N}}} \right]} \quad (12)$$

$$\overline{b_2 = \frac{1}{2\delta \ln(10)} \ln \left[ \frac{c - \sqrt{\frac{c}{N}}}{1 - \sqrt{\frac{c}{N}}} \right]} \quad (13)$$

where

$$\underline{\underline{c = \exp(2\delta \ln(10)b) = 10^{2\delta b}}} \quad (14)$$

In Appendix F, we provide also the formulas given here below involving only natural logarithm functions:

$$\overline{b = \frac{1}{2\delta \ln(10)} \ln \left( \frac{\overline{M} - M_c + 2\delta}{\overline{M} - M_c} \right)} \quad (15)$$

$$b = \frac{1}{2\delta \ln(10)} \ln \left[ \frac{2\delta + \sqrt{4\delta^2 + (|\Delta M|)^2}}{|\Delta M|} \right] \quad (16)$$

$$\overline{b = \frac{1}{2\delta \ln(10)} \ln \left( \frac{|\Delta M| - \Delta M'_c + 2\delta}{|\Delta M| - \Delta M'_c} \right)} \quad (17)$$

They are fully equivalent respectively to eqs. (7), (8) and (11), proposed by van der Elst, but do not include inverse hyperbolic functions that might be unavailable in some computational environments.

In Appendix F, we demonstrate the remarkable result that v) the Utsu correction (4) coincides with the expansion of the exact formula (15) truncated at the second order.

We stress also that our theoretical analysis allows us to show a further result (see Appendix I), that is that the correct eq. (15) coincides with the eq. (6) proposed by Marzocchi et al. (2020).

In order to show the efficiency of the various  $b$ -value estimators, 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 are given in Appendix L.

## 2 Comparing the $b$ -values computed by different formulations

We first compare the various estimators of eqs. (2), (4), (7), (8), (11), (15), (16)

and (17) on complete datasets simulated by eq. (L1b) and binned by eq. (L2). In Table 1 we report the average  $b$ -values and their standard deviations  $\sigma_b$  computed on a set of 10000 simulated datasets each including  $N=1000$  magnitudes, with binning size  $\Delta=0.1$  and with  $b=1$ . We also report the significance level  $p$  of the two-sided Student's t-test, testing the equality of the average  $b$ -value  $\bar{b}$  to the theoretical one  $b_s$ ,

$$t = \frac{|\bar{b} - b_s|}{\sigma_b} \quad (18)$$

with  $\bar{N} - 1$  degrees of freedom, being  $\bar{N}$  the average number of data used for the estimation in the magnitude datasets. This test is not fully rigorous because it assumes that the distribution of  $b$ -values is Normal and this is reasonably true only for large samples (with more than about 2000-5000 data). The correct test to be applied should be somehow similar to that proposed by Utsu (1966), which does not make any assumption on the statistical distribution of  $b$ . Unfortunately, the Utsu (1966) test is designed to compare the  $b$ -values of two samples, not to test if a sample mean corresponds to a given value. Furthermore, we use in the test average values of standard deviations and number of data, taken over the 10000 datasets, instead of values referring to a specific single dataset. Our approximate approach may provide anyway some sort of quantitative evaluation of the ability of each estimator to reproduce the simulated  $b$ -value, particularly when the  $p$  value is very small (e.g.  $\ll 0.01$ , allowing to safely discard the  $H_0$  hypothesis and thus suggesting that  $\bar{b} \neq b_s$ ) or when it is very large (e.g.  $\gg 0.1$ , suggesting to not discard the  $H_0$  hypothesis and then that  $\bar{b} \approx b_s$ ).

The results shown in Table 1 allow us to assert that most methods reproduce the true  $b$ -value ( $b_s = 1$ ) reasonably well with the exception of the simple Aki (1965) formula (2) for which the estimated  $b$ -value is significantly different from the true one. These results are confirmed even by varying the number of data  $N$  (100, 1000, 10000) and the theoretical  $b$ -value (0.7, 1.0, 1.5) (see Tables S1 to S9 in the supplementary material).

**Table 1 – Estimates from complete simulated sets with  $N=1000$ ,  $\Delta=0.1$  and  $b=1$**

Estimator	Eq.	$\bar{b}$	$\sigma_b$	$\bar{N}$	$p$
Aki (1965)	(2)	1.125526	0.040110	1000	0.001802
Aki (1965), Utsu (1966)	(4)	0.996282	0.031422	1000	0.905831
Van der Elst (2021), magnitudes	(7)	1.000699	0.031843	1000	0.982485
Van der Elst (2021), absolute differences by eq. (10)	(8)	1.001422	0.044977	500	0.974800
Van der Elst (2021), absolute differences by eq. (9)	(8)	1.001103	0.041072	999	0.978578
Van der Elst (2021), trimmed absolute differences by eq. (10)	(11)	1.001801	0.048492	443	0.970397
Van der Elst (2021), trimmed absolute differences by eq. (9)	(11)	1.001455	0.044096	885	0.973682
This paper, magnitudes	(15)	1.000699	0.031843	1000	0.982485

Estimator	Eq.	$\bar{b}$	$b$	$\bar{N}$	$p$
This paper, absolute differences by eq. (10)	(16)	1.001422	0.044977	500	0.974800
This paper, absolute differences by eq. (9)	(16)	1.001103	0.041072	999	0.978578
This paper, trimmed absolute differences by eq. (10)	(17)	1.001801	0.048492	443	0.970397
This paper, trimmed absolute differences by eq. (9)	(17)	1.001455	0.044096	885	0.973682

As eqs. (7), (8) and (11) are exactly equivalent to eqs. (15), (16) and (17) respectively, in the following we will consider only the latter ones. It is worth stressing that the approach used to compute magnitude differences (eq. 9 or 10) does not affect much the estimated  $b$ -value. However, we will see that it influences significantly the confidence intervals estimated by eq. (12) and (13).

In Table 2 for different simulated complete datasets, we report  $b$ -values computed by eq. (15) and confidence intervals estimated using various methods from the literature and also by eqs. (12) and (13) introduced in this paper. For all cases, the estimates based on eqs. (12) and (13) well correspond to the standard deviation  $\sigma_b$  computed from the simulated datasets and to confidence limits computed by other methods.

In Table 3 we report  $b$ -values computed for trimmed magnitude differences by eq. (17) and confidence intervals computed by eqs. (12) and (13), when differences are computed using eqs. (9) and (10). It is to note that confidence intervals estimated by independent differences (eq. 10) well correspond to the standard deviation  $\sigma_b$  computed from simulated datasets, whereas when using non independent differences computed by eq. (9) there is an underestimation by about 22%. The latter might be related to some sort of data correlation that reduces the number of “effective” independent data in the difference dataset. Then, we conclude that to compute differences it is always preferable to use eq. (10), and this will be our choice in the following computations.

**Table 2 – One-sigma confidence intervals for complete simulated sets**

$b$ - value	N	$\bar{b}$	$b$	Aki eq.(3)	Shi- Bolt eq.(5)	$\bar{b}-b_1$ eq.(12)	$b_2-\bar{b}$ eq.(13)	$\frac{1}{2}(b_2-b_1)$
----------------	---	-----------	-----	---------------	------------------------	--------------------------	--------------------------	------------------------

**Table 3 – One-sigma confidence intervals for complete simulated sets**

$b$ - value	Eq.	N	$\bar{b}$	$b$	$\bar{b}_-$ eq.(12)	$\bar{b}_+$ eq.(13)	$\frac{1}{2}(b_2-b_1)$
	(10)						

$b$ - value	Eq.	N	$\bar{b}$	$b$	$\bar{b}_-$ eq.(12)	$\bar{b}_+$ eq.(13)	$\frac{1}{2}(b_2-b_1)$
	(9)						
	(10)						
	(9)						
	(10)						
	(9)						

Even if, for the large majority of 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 we show the results for a binning size  $2=0.5$ . We can note that in this case the Aki (1965) estimator as corrected by Utsu (1966) (eq. 4) significantly underestimates the simulated  $b$ -value. 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 the Utsu (1966) correction to the Aki (1965) formula is approximate and only the exact formulas (7) and (15) provide the correct results in all cases.

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

Estimator	Eq.	$\bar{b}$	$b$	$\bar{N}$	$p$
Aki (1965)	(2)	1.883026	0.106794	1000	0.000000
Aki (1965), Utsu (1966)	(4)	0.902860	0.024514	1000	0.000079
This paper, magnitudes	(15)	1.000895	0.033628	1000	0.978777
This paper, absolute differences	(16)	1.001087	0.042059	500	0.979395
This paper, trimmed absolute differences	(17)	1.004389	0.069159	240	0.949451

As van der Elst (2021) asserts that, when analyzing incomplete datasets, the estimators are more robust if only the positive differences are used in eq. (11), in the following computations we consider two further estimators using only the trimmed positive and only the trimmed negative magnitude differences. In Table 5 we show the results of simulations of incomplete datasets built by using  $\mu = 1$ ,  $\lambda = 0.2$ ,  $N=1093$  (11000 before thinning) (see Appendix L) and by setting  $M_c=0.4$ , corresponding to the minimum magnitude of the simulated datasets. The histogram of one of the simulated datasets is portrayed in Fig. 1. We can see that all the estimators underestimate the theoretical  $b$  ( $b_s=1$ ), even if those based on differences seem to work slightly better.

**Table 5 - Incomplete simulated sets with  $\mu = 1$ ,  $\lambda = 0.2$ ,  $N=1093$  (11000 before thinning),  $2=0.1$  and  $b=1$ ,  $M_c=0.4$**

Estimator	Eq.	$\bar{b}$	$b$	$\bar{N}$	$p$
Aki (1965)	(2)	0.460944	0.006947	1093	0.000000
Aki (1965), Utsu (1966)	(4)	0.437711	0.006264	1093	0.000000
This paper, magnitudes	(15)	0.438082	0.006280	1093	0.000000
This paper, absolute differences	(16)	0.862855	0.032991	546	0.000037
This paper, trimmed absolute differences	(17)	0.890224	0.036483	506	0.002752
This paper, trimmed positive differences	(17)	0.892015	0.052275	253	0.039878
This paper, trimmed negative differences	(17)	0.891447	0.051621	247	0.036469

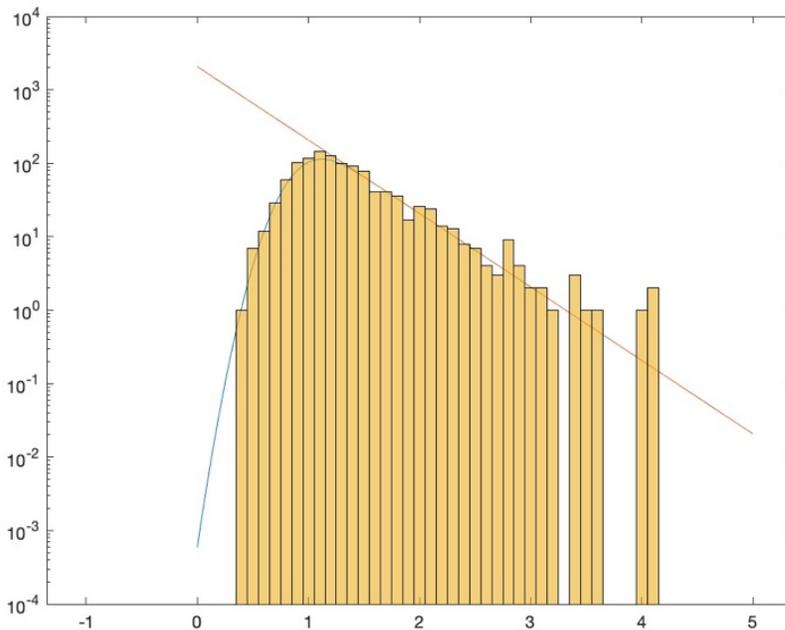


Fig. 1 – Incomplete simulated dataset with  $\mu = 1$ ,  $\lambda = 0.2$ ,  $N=1093$ ,  $2 = 0.1$  and  $b=1$ .

In Table 6 we show the results for the same simulation parameters when the minimum magnitude  $M_c$  is set to 1.1 corresponding to the maximum curvature (Wiemer and Wyss, 2000) of the FMD. In this case the simple Aki (1965) estimator would seem to give the correct result and to be better of the estimator corrected by Utsu (1966) and of the exact formula using magnitudes. But this is an artefact produced by the overestimation due to the binning which almost perfectly compensates the underestimation due to incompleteness. The other estimator based on magnitudes underestimate  $b$ , whereas all estimators based on magnitude differences give reasonably correct results. The performance ap-

pears slightly better for estimators based on trimmed differences (eq. 17) and among them we can note a slightly better performance of positive differences with respect to absolute and negative differences. Such evidence would seem to confirm the claim by van der Elst (2021) but from our results the preference for positive differences is very subtle and not so clear as in his paper.

**Table 6 - Incomplete simulated sets with  $\alpha = 1$ ,  $\beta = 0.2$ ,  $N=928$ ,  $\delta = 0.1$ ,  $b=1$  and  $M_c=1.1$**

Estimator	Eq.	$\bar{b}$	$b$	$\bar{N}$	$p$
Aki (1965)	(2)	1.026523	0.037518	786	0.479606
Aki (1965), Utsu (1966)	(4)	<i>0.917912</i>	<i>0.029991</i>	<i>786</i>	<i>0.006209</i>
This paper, magnitudes	(15)	<i>0.921364</i>	<i>0.030332</i>	<i>786</i>	<i>0.009704</i>
This paper, absolute differences	(16)	0.973845	0.047540	393	0.582509
This paper, trimmed absolute differences	(17)	0.986348	0.051871	353	0.792564
This paper, trimmed positive differences	(17)	0.989979	0.074481	176	0.893128
This paper, trimmed negative differences	(17)	0.988299	0.073980	154	0.874506

In Table 7 we show the results for the same parameters when the minimum magnitude  $M_c$  is set to 1.3 corresponding to the magnitude of maximum curvature plus 0.2, that is the way  $M_c$  is commonly set in literature (Wiemer and Wyss, 2000, Mignan and Woessner, 2012). We can see that now the simple Aki (1965) formula (2) clearly overestimates the  $b$ -value, whereas all other estimators using either magnitudes or differences give correct results, including the Aki-Utsu one (4). In this case, the estimator using the trimmed positive differences seems to be slightly better than the one using positive difference but slightly worse than the one using the trimmed absolute differences.

**Table 7 - Incomplete simulated sets with  $\alpha = 1$ ,  $\beta = 0.2$ ,  $N=640$ ,  $\delta = 0.1$ ,  $b=1$  and  $M_c=1.3$**

Estimator	Eq.	$\bar{b}$	$b$	$\bar{N}$	$p$
Aki (1965)	(2)	1.107743	0.052196	541	0.039020
Aki (1965), Utsu (1966)	(4)	0.982229	0.041025	541	0.664903
This paper, magnitudes	(15)	0.986471	0.041560	541	0.744912
This paper, absolute differences	(16)	0.998481	0.060113	270	0.979862
This paper, trimmed absolute differences	(17)	1.001747	0.064811	240	0.978515
This paper, trimmed positive differences	(17)	1.005584	0.094333	120	0.952895
This paper, trimmed negative differences	(17)	1.006768	0.092576	118	0.941846

The resilience of trimmed differences estimators to magnitude incompleteness can be improved by increasing the value of  $M'_c$  in eq. (17). In Table 8 we show the results obtained with the strongly incomplete set with  $M_c=0.4$  as in Table 5

but with increasing  $\Delta M'_c$  for trimmed differences estimators (absolute, positive and negative). For  $M'_c = 4\delta$  (twice the bin size), the  $p$  values of the Student's  $t$  test for both the positive and negative differences become larger than 10%. For  $M'_c = 10\delta$  (5 times the bin size) the deviations of the estimated  $b$ -values from the theoretical one becomes almost negligible for all of the three estimators. Note that increasing the trimming thresholds  $M'_c$  obviously reduces the number of available data but less dramatically than a similar increasing of the magnitude threshold  $M_c$ . Even in this case we have a preference for positive differences but still relatively weak.

**Table 8 - Incomplete simulated sets with  $\alpha = 1$ ,  $\beta = 0.2$ ,  $N=1093$ ,  $\sigma = 0.1$  and  $b=1$ ,  $M_c=0.4$**

Estimator	Eq.	$M'_c$	$\bar{b}$	$b$	$\bar{N}$	$p$
This paper, trimmed absolute differences	(17)	$4\delta$	0.927973	0.042749	428	0.092740
This paper, trimmed positive differences	(17)	$4\delta$	0.930314	0.061307	214	0.256954
This paper, trimmed negative differences	(17)	$4\delta$	0.929565	0.060234	212	0.243569
This paper, trimmed absolute differences	(17)	$6\delta$	0.957032	0.049715	355	0.388007
This paper, trimmed positive differences	(17)	$6\delta$	0.959837	0.071424	178	0.574610
This paper, trimmed negative differences	(17)	$6\delta$	0.959462	0.070339	180	0.565130
This paper, trimmed absolute differences	(17)	$8\delta$	0.977009	0.057063	290	0.687316
This paper, trimmed positive differences	(17)	$8\delta$	0.980645	0.081274	145	0.812110
This paper, trimmed negative differences	(17)	$8\delta$	0.980056	0.081047	146	0.805975
This paper, trimmed absolute differences	(17)	$10\delta$	0.990306	0.064465	235	0.880598
This paper, trimmed positive differences	(17)	$10\delta$	0.994635	0.091919	117	0.953556
This paper, trimmed negative differences	(17)	$10\delta$	0.994486	0.092570	121	0.952602

Finally, in Table 9 we show the effect on various estimators of a time-variable (decreasing) incompleteness as is known to occur after a strong main shock, owing to the superposition of the waveforms of many aftershocks that prevents the correct location and sizing of many small shocks in the hours or days after the main shocks. We set the magnitude of the main shock to  $m=4$  in eq. (L4),  $p=1$  and  $c=0.01$  in eq. (L5) and  $T_E=5$  days in (L8). The number of data  $N$  is the same of the magnitude dataset.

The estimators based on magnitudes tend to underestimate the theoretical  $b$ , whereas those based on differences give correct results. Even in this case we can note a slightly better performance for positive differences with respect to the negative ones. By varying the number of data and the theoretical  $b$  (see Tables from S19 to S27 in the supplementary material), positive differences are in general slightly better than negative differences but in one case they are worse ( $N=98$ ,  $b=1.5$ ). Both positive and negative differences are in all cases slightly better than absolute differences (positive and negative). However, the preference for one estimator with respect to another is in general quite weak.

**Table 9 - Incomplete (time dependent) simulated sets with  $\alpha = 1$ ,**

$= 0.2$ ,  $m=4$ ,  $p=1$ ,  $c=0.01$ ,  $T_E=5$  days,  $N=1031$  (30000 before thinning),  $2=0.1$  and  $b=1$ ,  $M_c=1.3$

Estimator	Eq.	$\bar{b}$	$b$	$\bar{N}$	$p$
Aki (1965)	(2)	1.000924	0.033805	1031	0.978183
Aki (1965), Utsu (1966)	(4)	0.897406	0.027167	1031	0.000159
This paper, magnitudes	(15)	0.900628	0.027462	1031	0.000311
This paper, absolute differences	(16)	0.982578	0.042176	515	0.679720
This paper, trimmed absolute differences	(17)	0.988949	0.045195	460	0.806931
This paper, trimmed positive differences	(17)	0.992715	0.065535	229	0.911588
This paper, trimmed negative differences	(17)	0.989501	0.064272	220	0.870380

### 3 Conclusions

We derived the complete formulations for the exact estimators of the  $b$ -value and of its confidence intervals of the frequency-magnitude distribution (FMD) from datasets with binned magnitudes. Such estimators are derived considering the discrete exponential distribution of magnitudes and the discrete Laplace distribution of magnitude differences and are equivalent to those recently proposed by van der Elst (2021). We also derive the first two moments of both distributions and the estimators of standard one-sigma confidence intervals of  $b$ -value.

To test the accuracy of all estimators to well reproduce the theoretical  $b$ -value, we simulated synthetic datasets with variable size and  $b$ -value, without or with a certain amount of incompleteness.

For estimators based on the distribution of magnitudes, exact formulations (eqs. 7 and 15) 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.

Estimators using magnitude differences (eqs. 8, 11, 16 and 17) are more robust with respect to magnitude incompleteness than those using magnitudes (eqs. 7 and 15) and give correct  $b$ -values when the magnitude cutting threshold  $M_c$  is not lower than the magnitude of maximum curvature of the FMD. Conversely, estimators using magnitudes (eqs. 7 and 15) 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.

Magnitude differences to be used in eq. (8, 11, 16 and 17) have to be computed

so as to be independent from each other (eq. 10) particularly for computing the analytical confidence intervals (eq. 12 and 13). The latter are found to well correspond to the standard deviations of  $b$ -values of simulated datasets.

Estimators based on trimmed magnitude differences (discarding differences smaller than the binning size or than a larger amount) are more accurate with respect to untrimmed ones. If the size of trimming difference threshold  $M'_c$  is increased up to about 5 times the binning size, the estimators become almost independent of magnitude incompleteness.

Estimators based on trimmed positive differences seem to reproduce the theoretical  $b$ -value better than those based on trimmed absolute and positive differences but the improvement is relatively small and not so clear as claimed by van der Elst (2021).

### 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 simple Matlab code written by authors, which is not made available publicly because it is not suitable for general usage. The paper does not use any kind of data.

### 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.
- 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. <https://doi.org/10.1785/BSSA0340040185>.
- 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>.
- Petrucelli, 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>.
- Petrucelli, 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>.

Petrucelli, 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. *Geophys. 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/BSA0720051677>.

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.

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 catalogs: 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  $2\delta$ . 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:

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

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

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

where  $\alpha$  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:

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

It follows that (A2) can be rewritten as:

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

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

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

or

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

depending on the decay factor  $\beta$ . 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  $\alpha$  and  $\beta$  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  $2\delta$  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  $2\delta$  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:

$$\mu_{\text{CE}} = \frac{1}{\alpha} \quad (\text{A10})$$

$$\text{var}_{\text{CE}} = \frac{1}{\alpha^2}; \quad \sigma_{\text{CE}} = \frac{1}{\alpha} \quad (\text{A11})$$

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

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

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

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

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:

$$S_1 = e^{-\alpha} S_1 + \frac{e^{-\alpha}}{1 - e^{-\alpha}} \quad (\text{A14})$$

This is an equation in the unknown  $S_1$  with solution:

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

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

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

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

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

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

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

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

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

Its solution is:

$$S_2 = \frac{e^{-\alpha}}{1 - e^{-\alpha}} \left( 2S_1 + \frac{1}{1 - e^{-\alpha}} \right) = \frac{e^{-\alpha}}{1 - e^{-\alpha}} \left( \frac{2e^{-\alpha}}{(1 - e^{-\alpha})^2} + \frac{1}{1 - e^{-\alpha}} \right) = \frac{e^{-\alpha} (1 + e^{-\alpha})}{(1 - e^{-\alpha})^3}$$

From the definition (A17), we then obtain:

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

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

$$\text{var}_{DE} = M_{2,DE} - (\mu_{DE})^2 \quad (A21)$$

After substituting the respective values, it transforms to:

$$\text{var}_{DE} = \frac{e^{-\alpha} (1 + e^{-\alpha})}{(1 - e^{-\alpha})^2} - \frac{e^{-2\alpha}}{(1 - e^{-\alpha})^2} = \frac{e^{-\alpha}}{(1 - e^{-\alpha})^2} = \frac{1}{4} \left( \text{csch} \frac{\alpha}{2} \right)^2$$

Consequently, the standard deviation  $\sigma_{\text{DE}}$  takes the form:

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

### 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:

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

It is a continuous density function with two symmetric exponential tails, and the same decay parameter  $\alpha$  as the original distributions.

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

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

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  $d = j - i$ ,  $d \in \mathbb{Z}$ , 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  $i \in \mathbb{N}$ . It follows that:

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

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:

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

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:

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

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

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

where recourse is made to the identity:

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

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

### Mean, variance and standard deviation

The computation of the mean of the continuous Laplace distribution  $\mu_{\text{CL}}$  is straightforward, since  $P(w) = P(-w)$ . Here the subscript CL stands for continuous Laplace distribution. Indeed, it is trivial to see that:

$$\mu_{\text{CL}} = \int_{-\infty}^0 wP(w)dw + \int_0^{+\infty} wP(w)dw = - \int_0^{+\infty} wP(-w)dw + \int_0^{+\infty} wP(w)dw = 0 \quad (B8)$$

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

$$\text{var}_{\text{CL}} = \int_{-\infty}^{+\infty} w^2 P(w)dw = 2 \int_0^{+\infty} w^2 P(w)dw = \alpha \int_0^{+\infty} w^2 e^{-\alpha w} dw = \frac{2}{\alpha^2} \quad (B9)$$

Hence, the corresponding standard deviation is:

$$\sigma_{\text{CL}} = \frac{\sqrt{2}}{\alpha} \quad (\text{B10})$$

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

$$\mu_{\text{DL}} = \sum_{d=-\infty}^{-1} d P_d + \sum_{d=1}^{\infty} d P_d = - \sum_{d=1}^{\infty} d P_{-d} + \sum_{d=1}^{\infty} d P_d = 0 \quad (\text{B11})$$

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

$$\text{var}_{\text{DL}} = \sum_{d=-\infty}^{\infty} d^2 P_d = 2 \sum_{d=1}^{\infty} d^2 P_d = 2 \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \sum_{d=1}^{\infty} d^2 e^{-\alpha d} = 2 \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} S_2 = \frac{2e^{-\alpha}}{(1 - e^{-\alpha})^2} = \frac{1}{2} \left( \text{csch} \frac{\alpha}{2} \right)^2$$

The corresponding standard deviation results to be:

$$\sigma_{\text{DL}} = \frac{\sqrt{2}e^{-\frac{\alpha}{2}}}{1 - e^{-\alpha}} = \frac{1}{\sqrt{2}} \text{csch} \frac{\alpha}{2} \quad (\text{B13})$$

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.:

$$\text{var}_{\text{CL}} = 2\text{var}_{\text{CE}} \quad \text{var}_{\text{DL}} = 2\text{var}_{\text{DE}} \quad (\text{B14})$$

### 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:

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

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

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

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

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

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

### 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:

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

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:

$$P_0 = \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} \quad (D2a)$$

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

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

$$\mu_{CA} = \frac{1}{\alpha}, \quad \text{var}_{CA} = \frac{1}{\alpha^2}, \quad \sigma_{CA} = \frac{1}{\alpha} \quad (D3)$$

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:

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

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:

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

$$\text{var}_{\text{DA}} = \frac{2e^{-\alpha}}{(1 - e^{-\alpha})^2} - \frac{2e^{-2\alpha}}{(1 + e^{-\alpha})^2 (1 - e^{-\alpha})^2} = \frac{2e^{-\alpha} (1 + e^{-2\alpha})}{(1 + e^{-\alpha})^2 (1 - e^{-\alpha})^2} \quad (D6)$$

The related standard deviation is therefore given by:

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

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.:

$$\text{var}_{\text{DA}} = \text{var}_{\text{DL}} \frac{1 + e^{-2\alpha}}{(1 + e^{-\alpha})^2} < \text{var}_{\text{DL}} \quad (D8a)$$

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

$$\sigma_{\text{DA}} < \sigma_{\text{DL}} \quad (D8b)$$

## 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:

$$\text{var}_{T,CE} = \text{var}_{CE} = \frac{1}{\alpha^2} \quad \sigma_{T,CE} = \sigma_{CE} = \frac{1}{\alpha} \tag{E2b}$$

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:

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

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:

$$\text{var}_{T,CL} = 2 \int_{w'}^{\infty} w^2 P(w) dw = \alpha \int_{w'}^{\infty} w^2 e^{-\alpha(w-w')} dw = \alpha e^{\alpha w'} \int_{w'}^{\infty} w^2 e^{-\alpha w} dw \quad (E6)$$

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

$$\text{var}_{T,CL} = \frac{1}{\alpha^2} \left[ (1 + \alpha w')^2 + 1 \right] \quad (E7a)$$

with the corresponding standard deviation:

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

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

$$\text{var}_{T,CL} \rightarrow \text{var}_{CL} \quad \text{and} \quad \sigma_{T,CL} \rightarrow \sigma_{CL} \quad \text{as} \quad w' \rightarrow 0 \quad (E8)$$

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.:

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

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

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

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

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

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

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:

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

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

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

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

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

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

$$\mu_{T,DL} = 0 \quad (E15)$$

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

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

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

$$\sum_{j=0'}^{\infty} (j + d')^2 e^{-\alpha j} = \sum_{j=0'}^{\infty} j^2 e^{-\alpha j} + 2d' \sum_{j=0'}^{\infty} j e^{-\alpha j} + d'^2 \sum_{j=0'}^{\infty} e^{-\alpha j} = S_2 + 2d' S_1 + \frac{d'^2}{1 - e^{-\alpha}}$$

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

$$\text{var}_{T,DL} = \frac{2}{(1 + e^{-\alpha})(1 - e^{-\alpha})^2} \left\{ e^{-\alpha} + [e^{-\alpha} + d'(1 - e^{-\alpha})]^2 \right\} \quad (E18a)$$

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

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  $\text{var}_{DL}$  and  $\sigma_{DL}$ .

## 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  $\beta$ . If we consider binned magnitudes, the decay parameter is  $\alpha$ . 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  $\alpha$  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  $\mu$ , and if it happens to depend on  $\alpha$ , that is if  $\mu = f(\alpha)$ , then we can obtain  $\alpha$  by means of the expression  $\alpha = f^{-1}(\mu)$  where  $f^{-1}$  is the inverse function of  $f$ , under the hypothesis the inverse 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  $\mu$  reflects directly on how good the estimate of  $\alpha$  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  $\mu_{DL}$  and  $\mu_{T,DL}$  are identically zero, and thus not depending on  $\alpha$ .

### Estimates based on exponential distributions

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) :

$$\mu = \frac{e^{-\alpha}}{1 - e^{-\alpha}} + k \tag{F1}$$

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 estimator  $\tilde{\alpha}$  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  $\alpha$  is the true value of the decay parameter. Therefore we obtain an unbiased estimator of  $\alpha$ , say  $\tilde{\alpha}$ , if we replace  $\mu$  with its sample mean, since the sample mean tends to  $\mu$  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:

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

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

$$M_k = M_0 + 2\delta k \quad k \geq 0 \quad (F7)$$

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:

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

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

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

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

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

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:

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

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  $\delta/(\overline{M} - M_0)$ .

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

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

For negative differences the formula is:

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

Eventually for the trimmed absolute differences, we get:

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

All the above expressions (F11)–(F13) can be also given in terms of the inverse hyperbolic cotangent, like in (F8).

#### **Estimates based on the untrimmed absolute differences distribution**

The mean  $\mu_{\text{DA}}$  of the distribution of the untrimmed absolute differences is given by:

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

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

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

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

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

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:

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

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

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

Substituting  $\mu_{\text{DA}}$  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, the corresponding estimator for  $\tilde{b}$  turns out to be:

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

## 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.

### Using samples of the discrete exponential distribution

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  $(i_1, i_2, \dots, i_N)$ , that we write:

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

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

The ML estimate of the parameter  $\alpha$  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  $\alpha$  vanishes, i.e.:

$$\frac{dL_N(\alpha)}{d\alpha} = N (1 - e^{-\alpha})^{N-1} e^{-\alpha N \bar{i}} \frac{d}{d\alpha} (1 - e^{-\alpha}) - \alpha N \bar{i} (1 - e^{-\alpha})^N e^{-\alpha N \bar{i}} = 0 \quad (G2)$$

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

$$\frac{d}{d\alpha} (1 - e^{-\alpha}) - \bar{i} (1 - e^{-\alpha}) = 0 \quad (G3)$$

Therefore we get:

$$e^{-\alpha} - \bar{i} (1 - e^{-\alpha}) = 0 \quad (G4)$$

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

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

### Using samples of the discrete Laplace distribution

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)):

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

then we can write for positive differences:

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

Analogously, for negative differences, we have:

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

And for differences equal to zero:

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

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

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

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

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

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

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

After some calculations, the expression becomes:

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

which leads to the final expression for the estimator:

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

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  $\alpha$  and  $b$  derived in the previous Appendix F are functions of the mean  $\mu$  of a distribution of a discrete variable  $i$  with probability  $P_i$  and standard deviation  $\sigma$ . Let us say that:

$$p = g(\mu) \quad (H1)$$

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

$$\tilde{p} = g(\overline{\mu}_N) \quad (H2)$$

where  $\overline{\mu}_N$  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  $\mu$  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  $\mu$ . 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:

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

If we call  $P_\sigma$  the probability that  $\mu$  belongs to the interval  $I_{\mu_N}$ , 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.:

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

$$\tilde{p}_N = \frac{1}{2} (p_{2,N} - p_{1,N}) \tag{H7}$$

Note that in the above formulas we assume to know  $\sigma$  that, through (H3), would allow us to know  $\sigma_N$ . In practice, however,  $\sigma$  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  $\sigma$  through a proper function of  $\tilde{p}$ , say  $\sigma = \sigma(\tilde{p})$ ;
- 4) compute  $\sigma_N$  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:

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

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

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

The further step is to compute the extremes of the interval  $I_{\sigma_N}$  :

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

In (H10) use has been made of the 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:

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

Likewise, the upper end results to be:

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

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:

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

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

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:

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

The estimated standard confidence interval can be computed as

$$b_- = \tilde{b} - \tilde{b}_{1,N} \quad (H15a)$$

and

$$b_+ = \tilde{b}_{2,N} - \tilde{b} \quad (H15b)$$

and finally

$$b = \frac{1}{2} (\tilde{b}_{2,N} - \tilde{b}_{1,N}) \quad (H15c)$$

### Confidence intervals for the absolute difference distribution

The decay parameter  $\alpha$  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  $g$ :

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

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:

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

The endpoints of the interval  $I_{\sigma_N}$  are:

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

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

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

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

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

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

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

$$b = \frac{1}{2} (\tilde{b}_{2,N} - \tilde{b}_{1,N}) \quad (H22)$$

**Appendix I – Derivation of eq. (15) from eq. (6) (Marzocchi et al., 2020)**

Let's start from eq. (6) of the main text:

$$\underline{\underline{b_{\text{corrected}} = \frac{1}{2\delta \ln(10)} \ln \left[ \frac{1+b \ln(10)\delta}{1-b \ln(10)\delta} \right]}} \quad (I1)$$

where  $b$  is computed by eq. (4) in which  $M_c$  is diminished by  $\delta$ . By substituting (4) in (I1) we have

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

This corresponds exactly to eq. (15).

**Appendix L - Simulation of complete and incomplete magnitude datasets**

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

$$\underline{\underline{M = -\frac{\ln\{\text{rand}\}0:1\}}{b \ln(10)} + M_c}} \quad (L1a)$$

where  $\text{rand}\}0:1[$  is a pseudo random number with uniform distribution in the interval  $]0:1[$ .

The binning of magnitudes is obtained by

$$\underline{\underline{M_{\text{binned}} = \text{round} \left( \frac{M}{2\delta} \right) 2\delta}} \quad (L2)$$

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

$$\underline{\underline{M = -\frac{\ln\{\text{rand}\}0:1\}}{b \ln(10)} + M_c - \delta}} \quad (L1b)$$

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

$$\underline{\underline{p(M|\mu, \lambda) = \frac{1}{\lambda\sqrt{2\pi}} \int_{-\infty}^M e^{-\frac{(m-\mu)^2}{2\lambda^2}} dm \quad (L3)}}$$

In this formulation the mean  $\mu$  corresponds to the threshold magnitude at which  $p=0.5$ , that means that above it, the 50% of earthquakes are correctly located and sized by the seismic network.

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 (L3).

Van der Elst also 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

$$\underline{\underline{m_c = m - 4.5 - 0.75t \quad (L4)}}$$

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. Van der Elst suggested to use equation (L4) to set the time varying mean  $\mu(t)$  in eq. (L3).

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

$$\underline{\underline{r(t) = \frac{K}{(t+c)^p} \quad (L5)}}$$

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 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

$$\underline{\underline{\tau = \int_0^t r(s)ds = F(t) \quad (L6)}}$$

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

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

$$\underline{\underline{t = F^{-1}(\tau) \quad (L7)}}$$

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  $[0, T_E]$

$$\underline{\underline{\int_0^{T_E} r(s)ds = N \quad (L8)}}$$

This implies that

$$\underline{\underline{\begin{array}{c} p \\ p = 1 \\ (L9) \end{array}}}$$

Then, the direct timescale transform is

$$\underline{\underline{\begin{array}{c} p \\ N \quad p = 1 \\ (L10) \end{array}}}$$

and the inverse timescale transform is

$$\underline{\underline{\begin{array}{c} p \\ + (c) - c \quad p = 1 \\ (L11) \end{array}}}$$

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

$$\underline{\underline{\tau_i = \tau_{i-1} - \ln\{1 - \text{rand}\}0 : 1[ \}, \quad i = 2, N \quad (L12)}}$$