

Why Gamma Distribution of Seismic Inter-Event Times: A Theoretical Explanation

Laxman Bokati¹, Aaron Velasco^{1,2}, and Vladik Kreinovich^{1,3}

¹Computational Science Program

²Department of Geological Sciences

³Department of Computer Science

University of Texas at El Paso

500 W. University

El Paso, Texas 79968, USA

lbokati@miners.utep.edu, aavelasco@utep.edu, vladik@utep.edu

Abstract

It is known that the distribution of seismic inter-event times is well described by the Gamma distribution. Recently, this fact has been used to successfully predict major seismic events. In this paper, we explain that the Gamma distribution of seismic inter-event times can be naturally derived from the first principles.

1 Formulation of the Problem

Gamma distribution of seismic inter-event times: empirical fact. Detailed analysis of the seismic inter-event times t – i.e., of times between the two consequent seismic events occurring in the same area – shows that these times are distributed according to the Gamma distribution, with probability density

$$\rho(t) = C \cdot t^{\gamma-1} \cdot \exp(-\mu \cdot t), \quad (1)$$

for appropriate values γ , μ , and C ; see, e.g., [1, 2].

Lately, there has been a renewed interest in this seemingly very technical result, since a recent paper [4] has shown that the value of the parameter μ can be used to predict a major seismic event based on the preceding foreshocks. Specifically, it turns out that more than 70% of major seismic events in Southern California could be predicted some time in advance – with an average of about two weeks in advance.

Why gamma distribution? This interest raises a natural question: why the inter-event times follow gamma distribution? In this paper, we provide a possible theoretical explanation for this empirical fact.

2 Our Explanation

Maximum entropy: general idea. In our explanation, we will use Laplace's Indeterminacy Principle, which is also known as the maximum entropy approach; see, e.g., [3]. The simplest case of this approach is when we have n alternatives, and we have no reasons to believe that one of them is more probable. In this case, a reasonable idea is to consider these alternatives to be equally probable, i.e., to assign, to each of these n alternatives, the same probability $p_1 = \dots = p_n$. Since the probabilities should add to 1, i.e., $\sum_{i=1}^n p_i = 1$, we thus get $p_i = 1/n$.

In this case, we did not introduce any new degree of certainty into the situation that was not there before – as would have happened, e.g., if we selected a higher probability for one of the alternatives. In other words, out of all possible probability distributions, i.e., out of all possible tuples (p_1, \dots, p_n) for which $\sum_{i=1}^n p_i = 1$, we selected the only with the largest possible uncertainty.

In general, it is known that uncertainty can be described by the *entropy* S , which in the case of finitely many alternatives has the form $S = - \sum_{i=1}^n p_i \cdot \ln(p_i)$, and in the case of a continuous random variable with probability density $\rho(x)$, for which sum becomes an integral, a similar form

$$S = - \int \rho(x) \cdot \ln(\rho(x)) dx. \quad (2)$$

Maximum entropy: examples. If the only information that we have about a probability distribution is that it is located somewhere on the given interval $[a, b]$, then the only constraint on the corresponding probability density function is that the overall probability over this interval is 1, i.e., that

$$\int_a^b \rho(x) dx = 1. \quad (3)$$

So, to apply the maximum entropy approach, we need to maximize the objective function (2) under the constraint (3). The usual way of solving such constraint optimization problem is to apply Lagrange multiplier method that reduces the original constraint optimization problem to an appropriate unconstrained optimization problem. In this case, this new problem means maximizing the expression

$$- \int \rho(x) \cdot \ln(\rho(x)) dx + \lambda \cdot \left(\int_a^b \rho(x) dx - 1 \right), \quad (4)$$

where the parameter λ – known as *Lagrange multiplier* – needs to be determined from the condition that the solution to this optimization problem satisfies the constraint (3).

For this problem, the unknowns are the values $\rho(x)$ corresponding to different x . Differentiating the expression (4) with respect to $\rho(x)$ and equating the derivative to 0, we get the following equation:

$$-\ln(\rho(x)) - 1 + \lambda = 0.$$

(Strictly speaking, we need to use variational differentiation, since the unknown is a function.) The above equation implies that $\ln(\rho(x)) = \lambda - 1$, and thus, that $\rho(x) = \text{const}$. So, we get a uniform distribution – in full accordance with the original idea that, since we do not have any reasons to believe that some points on this interval are more probable than others, we consider all these points to be equally probable.

If, in addition to the range $[a, b]$, we also know the mean value

$$\int_a^b x \cdot \rho(x) dx = m \tag{5}$$

of the corresponding random variable, then we need to maximize the entropy (2) under two constraints (3) and (5). In this case, the Lagrange multiplier methods leads to the unconstrained optimization problem of maximizing the following expression:

$$-\int \rho(x) \cdot \ln(\rho(x)) dx + \lambda \cdot \left(\int_a^b \rho(x) dx - 1 \right) + \lambda_1 \cdot \left(\int_a^b x \cdot \rho(x) dx - m \right). \tag{6}$$

Differentiating this expression with respect to $\rho(x)$ and equating the derivative to 0, we get the following equation:

$$-\ln(\rho(x)) - 1 + \lambda + \lambda_1 \cdot x = 0,$$

hence $\ln(\rho(x)) = (\lambda - 1) + \lambda_1 \cdot x$ and so, we get a (restricted) Laplace distribution, with the probability density $\rho(x) = C \cdot \exp(\mu \cdot x)$, where we denoted $C \stackrel{\text{def}}{=} \exp(\lambda - 1)$ and $\mu \stackrel{\text{def}}{=} \lambda_1$.

Comment. It is worth mentioning that if, in addition to the mean value, we also know the second moment of the corresponding random variable, then similar arguments lead us to a conclusion that the corresponding distribution is Gaussian. This conclusion is in good accordance with the ubiquity of Gaussian distributions.

What are the reasonable quantities in our problem. We are interested in the probability distribution of the inter-event time t . Based on the observations, we can find the mean inter-event time, so it makes sense to assume that we know the mean value of this time.

Usual (astronomical) time vs. internal time: general idea. This mean value is estimated if we use the usual (astronomical) time t , as measured, e.g., by rotation of the Earth around its axis and around the Sun. However, it is

known that many processes also have their own “internal” time – based on the corresponding internal cycles. For example, we can measure the biological time of an animal (or a person) by such natural cyclic activities as breathing or heartbeat. Usually, breathing and hear rate are more or less constant, but, e.g., during sleep, they slow down – as most other biological processes slow down. On the other hand, in stressful situations, e.g., when the animal’s life is in danger, all the biological processes speed up – including breathing and heart rate. To adequately describe how different biological characteristics change with time, it makes sense to consider them not only how they change in astronomical time, but also how they change in the corresponding internal time – measured not by number of Earth’s rotations around the sun, but rather in terms of number of heartbeats. An even more drastic slowdown occurs when an animal hibernates. In general, the system’s internal time can be sometimes slower than astronomical time, and sometimes faster.

Usual (astronomical) time vs. internal time: case of seismic events.

In our problem, there is a similar phenomenon: usually, seismic events are reasonably rare. However, the observations indicate that the frequency with which foreshocks appear increases when we get closer to a major seismic event. In such situation, the corresponding seismic processes speed up, so we can say that the internal time speeds up. In general, an internal time is often a more adequate description of the system’s changes than astronomical time. It is therefore reasonable to supplement the mean value of the inter-event time measured in astronomical time by the mean value of the inter-event time measured in the corresponding internal time.

How internal time depends on astronomical time: general idea.

To describe this idea in precise terms, we need to know how this internal time τ depends on the astronomical time. As we have mentioned, the usual astronomical time is measured by natural cycles, i.e., by processes which are periodic in terms of the time t . So, to find the expression for internal time, we need to analyze what cycles naturally appear in the studied system – and then define internal time in terms of these cycles.

To describe the system’s dynamics means to describe how the corresponding physical quantities $x(t)$ change with time t . In principle, in different physical situations, we can have different functions $x(t)$. In principle, to describe a general function, we need to have infinitely many parameters – e.g., we need to describe the values of this function at different moments of time. In practice, however, we can only have finitely many parameters. So, it is reasonable to consider finite-parametric families of functions. The simplest – and most natural – is to select some basic functions $e_1(t), \dots, e_n(t)$, and to consider all possible linear combinations of these functions, i.e., all possible functions of the type

$$x(t) = C_1 \cdot e_1(t) + \dots + C_n \cdot e_n(t), \tag{6}$$

where C_1, \dots, C_n are the corresponding parameters. This is indeed what is done in many situations: sometimes, we approximate the dynamics by polynomials

– linear combinations of powers t^k , sometimes we use linear combinations of sinusoids, sometimes linear combinations of exponential functions, etc.

How internal time depends on astronomical time: case of seismic events. The quality of this approximation depends on how adequate the corresponding basis functions are for the corresponding physical process. Let us analyze which families are appropriate for our specific problem: analysis of foreshocks preceding a major seismic event. In this analysis, we can use the fact that, in general, to transform a physical quantity into a numerical value, we need to select a starting point and a measuring unit. If we select a different starting point and/or a different measuring unit (e.g., minutes instead of seconds), we will get different numerical values for the same quantity.

For the inter-event times, the starting point is fixed: it is 0, the case when the next seismic events follows immediately after the previous one. So, the only remaining change is the change of a measuring unit. If we replace the original time unit with a one which is r times smaller, then all numerical values are multiplied by r , i.e., instead of the original value t , we get a new value $t_{\text{new}} = r \cdot t$. For example, if we replace minutes by seconds, then the numerical values of all time intervals are multiplied by 60, so that, e.g., 2.5 minutes becomes $60 \cdot 2.5 = 150$ seconds.

Some seismic processes are faster, some are slower. This means that, in effect, they differ by this slower-to-faster or faster-to-slower transformations $t \rightarrow r \cdot t$. We would like to have a general description that would fit all these cases. In other words, we would like to make sure that the class (6) remains the same after this “re-scaling”, i.e., that for each i and for each r , the re-scaled function $e_i(r \cdot t)$ belongs to the same class (6). In other words, we require that for each i and r , there exists values $C_{ij}(r)$ for which

$$e_i(r \cdot t) = C_{i1}(r) \cdot e_1(t) + \dots + C_{in}(r) \cdot e_n(t). \quad (7)$$

Let us solve the resulting systems of equations. Seismic waves may be changing fast but, in general, they are still smooth. It is therefore reasonable to consider only smooth functions $e_i(t)$. If we pick n different values t_1, \dots, t_n , then, for each r and for each i , we get a system of n linear equations for determining n unknowns $C_{i1}(r), \dots, C_{in}(r)$:

$$\begin{aligned} e_i(r \cdot t_1) &= C_{i1}(r) \cdot e_1(t_1) + \dots + C_{in}(r) \cdot e_n(t_1); \\ &\dots \\ e_i(r \cdot t_n) &= C_{i1}(r) \cdot e_1(t_n) + \dots + C_{in}(r) \cdot e_n(t_n). \end{aligned}$$

Due to Cramer’s rule, each component $C_{ij}(r)$ of the solution to this system of linear equations is a ratio of two determinants and is thus, a smooth function of the corresponding coefficients $e_i(r \cdot t_j)$ and $e_i(t_j)$. Since the function $e_i(t)$ is differentiable, we conclude that the functions $C_{ij}(r)$ are also differentiable.

Since all the functions $e_i(t)$ and $C_{ij}(r)$ are differentiable, we can differentiate both sides of the formula (7) with respect to r and get:

$$e'_i(r \cdot t) \cdot t = C'_{i1}(r) \cdot e_1(t) + \dots + C'_{in}(r) \cdot e_n(t),$$

where for each function f , the expression f' , as usual, denotes the derivative. In particular, for $r = 1$, we get

$$e'_i(t) \cdot t = c_{i1} \cdot e_1(t) + \dots + c_{in} \cdot e_n(t),$$

where we denoted $c_{ij} \stackrel{\text{def}}{=} C'_{ij}(1)$. For the auxiliary variable $T = \ln(t)$ for which $t = \exp(T)$, we have $dT = \frac{dt}{t}$, hence $\frac{de_i(t)}{dt} \cdot t = \frac{de_i(t)}{dT}$. So, for the auxiliary functions $E_i(T) \stackrel{\text{def}}{=} e_i(\exp(T))$, we get

$$E'_i(T) = c_{i1} \cdot E_1(T) + \dots + c_{in} \cdot E_n(T).$$

So, for the functions $E_i(T)$, we get a system of linear differential equations with constant coefficients. It is well known that a general solution to such a system is a linear combination of the expressions $T^k \cdot \exp(a \cdot T) \cdot \sin(\omega \cdot T + \varphi)$ for some natural number k and real numbers a , ω , and φ . Thus, each function $e_i(t) = E_i(\ln(t))$ is a linear combination of the expressions

$$(\ln(t))^k \cdot \exp(a \cdot \ln(t)) \cdot \sin(\omega \cdot \ln(t) + \varphi) = (\ln(t))^k \cdot t^a \cdot \sin(\omega \cdot \ln(t) + \varphi). \quad (8)$$

So, the general expression (6) is also a linear combination of such functions.

The periodic part of this expression is a sine or cosine function of $\ln(t)$, so we can conclude that for seismic processes, the internal time τ is proportional to the logarithm $\ln(t)$ of the astronomic time: $\tau = c \cdot \ln(t)$ for some constant c .

This explains the ubiquity of Gamma distributions. Indeed, suppose that we know the mean values m_t and m_τ of the astronomical time t and the mean value of the internal time $\tau = c \cdot \ln(t)$. This means that the corresponding probability density function $\rho(t)$, in addition to the usual constraint $\int \rho(t) dt$, also satisfies the constraints

$$\int t \cdot \rho(t) dt = m_t$$

and

$$\int c \cdot \ln(t) \cdot \rho(t) dt = m_\tau.$$

Out of all possible distributions satisfying these three inequalities, we want to select the one that maximizes entropy

$$- \int \rho(t) \cdot \ln(\rho(t)) dt.$$

To solve the resulting constraint optimization problem, we can apply the Lagrange multiplier method and reduce it to the unconstrained optimization problem of maximizing the expression:

$$-\int \rho(t) \cdot \ln(\rho(t)) dt + \lambda \cdot \left(\int \rho(t) dt - 1 \right) + \lambda_t \cdot \left(\int t \cdot \rho(t) dt - m_t \right) + \lambda_\tau \cdot \left(\int c \cdot \ln(t) \cdot \rho(t) dt - m_\tau \right),$$

for some values λ , λ_t , and λ_τ . Differentiating both sides with respect to each unknown $\rho(t)$, we conclude that

$$-\ln(\rho(t)) - 1 + \lambda + \lambda_t \cdot t + \lambda_\tau \cdot c \cdot \ln(t) = 0,$$

i.e., that

$$\ln(\rho(t)) = (\lambda - 1) + \lambda \cdot t + (\lambda_\tau \cdot c) \cdot \ln(t).$$

Exponentiating both sides, we get the desired Gamma distribution form (1).

$$\rho(t) = C \cdot \tau^{\gamma-1} \cdot \exp(\mu \cdot t),$$

with $C = \exp(\lambda - 1)$, $\gamma = \lambda_\tau \cdot c + 1$, and $\mu = \lambda_t$. Thus, we have indeed explained the ubiquity of the Gamma distribution.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

- [1] A. Corral, “Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes”, *Physical Review Letters*, 2004, Vol. 92, Paper 108501.
- [2] S. Heinzl, F. Scherbaum, and C. Beauval, “Estimating background activity based on interevent-time distribution”, *Bulletin of the Seismological Society of America*, 2006, Vol. 96, No. 1, pp. 313–320.
- [3] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.
- [4] D. T. Trugman and Z. E. Ross, “Pervasive foreshock activity across Southern California”, *Geophysical Research Letters*, 2019, Vol. 46, <https://doi.org/10.1029/2019GL083725>