A Possible Common Mechanism Behind Skew Normal Distributions in Economics and Hydraulic Fracturing-Induced Seismicity

Laxman Bokati
The University of Texas at El Paso, lbokati@miners.utep.edu

Aaron Velasco
The University of Texas at El Paso, aavelasco@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Kittawit Autchariyapanitkul
Maejo University, kittawit.a@mju.ac.th

Comments:
Technical Report: UTEP-CS-22-37

Recommended Citation

A Possible Common Mechanism Behind Skew Normal Distributions in Economics and Hydraulic Fracturing-Induced Seismicity

Laxman Bokati, Aaron Velasco, Vladik Kreinovich, and Kittawit Autchariyapanitkul

Abstract Many economic situations – and many situations in other application areas – are well-described by a special asymmetric generalization of normal distributions – known as skew-normal. However, there is no convincing theoretical explanation for this empirical phenomenon. To be more precise, there are convincing explanations for the ubiquity of normal distributions, but not for the transformation that turns normal into skew-normal. In this paper, we use the analysis of hydraulic fracturing-induced seismicity to show explain the ubiquity of such a transformation.

1 Formulation of the Problem

Ubiquity of normal distributions. In many practical situations, the empirical probability distribution is close to Gaussian (normal), i.e., to a distribution with the probability density

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right). \tag{1}$$

Laxman Bokati

Computational Science Program, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: lbokati@miners.utep.edu

Aaron Velasco

Department of Earth Sciences, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: aavelasco@utep.edu

Vladik Kreinovich

Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: vladik@utep.edu

Kittawit Autchariyapanitkul

Faculty of Economics, Maejo University, Chiang Mai, Thailand, e-mail: kittawit.a@mju.ac.th

There is a convincing explanation for this ubiquity. Indeed, in many practical situations, the random phenomenon is caused by a large number of small independent random factors, and it is known that in such situations, under some reasonable conditions, the probability distribution is close to Gaussian.

To be more precise, the corresponding result – known as the Central Limit Theorem – says as the number of factors increases, the distribution tends to Gaussian; see, e.g., [4].

Normal distributions are symmetric. One of the properties of the normal distribution is that it is symmetric with respect to inversion around the central point a.

To be more precise, if we consider two points x and x' on two different sides of the central point a and at the same distance from a, i.e., for which $x' - a = -(x - a)$, then we will get $f(x') = f(x)$.

Not all practical distributions are symmetric. Not all empirical distributions are Gaussian. In particular, practical distributions are often not symmetric.

How can we describe distributions which are not symmetric. One of the advantages of a normal distribution is that it depends only on two parameters: the mean a and the standard deviation σ. The more parameters, the more experimental results are needed to determine all of them; see, e.g., [4]. The fact that we only have two parameters to determine means that it is sufficient to have a small amount of experimental data to uniquely determine the actual distribution.

It is desirable to have a similar few-parametric family to describe a more general class of asymmetric distributions.

Skew-normal distributions. Several few-parametric families have been proposed for the purpose of describing asymmetric distributions. It turns out that in many economic and other problem, the most adequate description is provided by the so-called *skew-normal distribution* (see, e.g., [2]), i.e., the distribution with the probability density

$$g(x) = A \cdot (f(x))^\alpha \cdot \int^x f(y)\,dy, \tag{2}$$

where $f(x)$ is the probability density (1) of the normal distribution, and A is a normalizing factor, making sure that $g(x)$ is a probability density function, i.e., that $\int g(x)\,dx = 1$.

Why. A natural question is: why is the family (2) ubiquitous?

We know why normal distributions are ubiquitous – this follows from the Central Limit Theorem – but how can we explain the transformations (2) from normal distribution to skew-normal distribution?

What we do in this paper. In this paper, we provide a possible answer to this question by noticing that a similar transformation (2) naturally appears, in the first approximation, in another application area – namely, in the analysis of hydraulic fracturing-induced seismicity.

2 Hydraulic fracturing-induced seismicity: first-approximation description and the resulting explanation

What is hydraulic fracturing. Traditional methods of extracting oil and gas left a significant amount of it in the oil well.

To extract the remaining amount of oil and gas, practitioners use *hydraulic fracturing*. In this process, high-pressure hot liquid (water with some chemical added) is pumped into the wells. This liquid fractures the rocks that prevented oil and gas from coming out and thus, pushes oil and gas out.

Fracturing causes seismicity. In general, seismic activity is caused by stresses. Many of these stresses happen to be close to the well depths. Fracturing adds to these stresses, and thus, causes additional seismic activity; see, e.g., [3].

Comment. Most of the resulting seismic activity is minor: humans can feel it, but it does not cause serious damage.

How seismicity depends on fracturing: first approximation. In general, the amount of seismic activity $g(t)$ is correlated with the amount of fluid $f(t)$ pumped into the area.

To be more precise, this correlation comes with a delay – since it takes some time for the pumped liquid to reach the fault areas, where the stresses are location – but for the purposes to this paper, in the first approximation, we can ignore this delay.

In this delay-ignoring approximation, the value $g(t)$ of the quantity g at moment t depends on the value $f(t)$ of the quantity f at the same moment of time: $g(t) = F(f(t))$ for some function $F(f)$.

Scale-invariance. What should be the dependence $g = F(f)$ between the numerical values of the quantities f and g?

To answer this question, let us recall that the numerical values of most physical quantities depend on the choice of the measuring unit: if we choose a measuring unit which is λ times smaller than the previous one, then all numerical values are multiplied by λ. For example, if we replace meters with centimeters – a unit which is 100 times smaller – then 2 m becomes $100 \cdot 2 = 200$ cm.

In many cases – as in the case of length – there is no fundamental reason for selecting a measuring unit: all measuring units are equally reasonable. In such cases, it makes sense to require that the exact form of the dependence $g = F(f)$ should remain the same if we change the unit for f.

Of course, if we change the unit for f, we must appropriately change the unit for g. For example, if we want to preserve the formula for velocity $v = d/t$ as a ratio between distance d and time t, then, if we change the unit of distance from km to m, we need to also change the corresponding unit for velocity from km/h to m/h.

In general, such scale-invariance of the dependence $g = F(f)$ means that for every $\lambda > 0$, if we replace the numerical values of f with new values $f' = \lambda \cdot f$, then there should exist an appropriate value μ – depending on λ – for which $g = F(f)$ implies that $g' = F(f')$, where $g' \stackrel{\text{def}}{=} \mu \cdot g$.

Comment. In economics, scaling corresponds, e.g., to changing a currency in which we gauge incomes and prices.

Which dependencies are scale-invariant. If we substitute the expressions $f' = \lambda \cdot f$, $g' = \mu(\lambda) \cdot g$, and $g = F(f)$ into the formula $g' = F(f')$, we conclude that $\mu(\lambda) \cdot F(f) = F(\lambda \cdot f)$. It is known (see, e.g., [1]) that all measurable – in particular, all continuous – solutions of this functional equation have the form

$$F(f) = B \cdot f^{\alpha}, \tag{3}$$

for some constants B and α.

The amplitude of the seismic activity is proportional to overall amount of injected liquid. Newly injected liquid triggers the seismic activity, because its presence disrupts the situation at the fault.

In the beginning of the injection, only the newly injected liquid contributes to this disruption. However, as more and more liquid makes it way to the fault, the larger the overall volume of the liquid that disturbs the fault and thus, the larger the intensity of the resulting seismic activity.

From this viewpoint, the coefficient B – that described this intensity – grows with the overall amount $\int^{t} f(s)\,ds$ of the pumped liquid. In the first approximation, we can assume that this dependence is linear, i.e., that $B = A \cdot \int^{t} f(s)\,ds$ for some constant A. Substituting this expression for B into the formula (3), we conclude that

$$g(t) = A \cdot (f(t))^{\alpha} \cdot \int^{t} f(s)\,ds. \tag{4}$$

Conclusion of this section. Thus, the first-approximation description of hydraulic fracturing-induced seismicity leads to exactly the desired formula (1). In this derivation, we did not use any specific physical features of this phenomenon, just a general idea. Thus, we have indeed a general explanation for the empirical formula (1).

For example, in economics, we can make similar arguments about investing additional money into an economy: the current change depends on the amount invested now, but the coefficient of proportionality grows with the size of the economy – i.e., is, in the first approximation, proportional to the overall amount of money already invested: indeed, the larger the economy, the more efficiently it can use new funds.

Comment. It should be mentioned that with respect to seismicity, the formula (4) is a very crude approximation, an approximation that provides mostly the qualitative explanation. This is in contrast to many econometric applications, where the dependence (1) works very well.

This difference is easy to explain:

- in geosciences, we only have a limited number of measurements, while
- in economics, we have a large amount of data.

Even though such cases – in which the dependence (1) is reasonably exact – are as rare in economics as it is in geosciences:

- in economics, where we have a much larger amount of data, it is possible to find cases when this dependence works well, while
- in geosciences, where there is much less data, such cases may be difficult to find.

Acknowledgments

References

1. J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.
2. A. Azzalini and A. Capitanio, *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge, Massachusetts, 2013.
3. R. Schultz, R. J. Skoumal, M. R. Brudzinski, D. Eaton, B. Baptie, and W. Ellsworth, "Hydraulic fracturing-induced seismicity", *Reviews of Geophysics*, 2020, Vol. 58, Paper 000695.
4. D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures*, Chapman & Hall/CRC, London, UK, 2011.