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

Laxman Bokati\textsuperscript{1}, Aaron Velasco\textsuperscript{2}, Vladik Kreinovich\textsuperscript{1,3}, and Kittawit Autchariyapanitkul\textsuperscript{4}

\textsuperscript{1}Computational Science Program, Departments of \textsuperscript{2}Earth Sciences and \textsuperscript{3}Computer Science University of Texas at El Paso, El Paso, Texas 79968, USA
lbokati@miners.utep.edu, aavelasco@utep.edu, vladik@utep.edu
\textsuperscript{4}Faculty of Economics, Maejo University, Chiang Mai, Thailand, kittawit.a@mju.ac.th
1. Ubiquity of normal distributions

- In many practical situations, the empirical probability distribution is close to Gaussian, with the probability density

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

- 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.
- 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 says as the number of factors increases, the distribution tends to Gaussian.
- This result is known as the Central Limit Theorem.
2. 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 $f(x') = f(x)$. 
3. 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 \( \mu \) and the standard deviation \( \sigma \).
- The more parameters, the more experimental results are needed to determine all of them.
- It is therefore desirable to have a similar few-parametric family to describe a more general class of asymmetric distributions.
4. 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 skew-normal distribution.
- This distribution has the probability density \( g(x) = A \cdot (f(x))^\alpha \cdot \int_x^\infty f(y) \, dy \), where:
  - \( f(x) \) is the probability density of the normal distribution, and
  - \( A \) is a normalizing factor, making sure that \( g(x) \) is a probability density function, i.e., that \( \int g(x) \, dx = 1 \).
- A natural question is: why is this family ubiquitous?
- We know why normal distributions are ubiquitous – this follows from the Central Limit Theorem.
- But how can we explain the above transformations from normal distribution to skew-normal distribution?
5. What we do in this talk

- In this talk, we provide a possible answer to this question.
- Namely, we notice that:
  - a similar transformation naturally appears, in the first approximation, in another application area,
  - namely, in the analysis of hydraulic fracturing-induced seismicity.
6. 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.
7. 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.

• Most of the resulting seismic activity is minor: humans can feel it, but it does not cause serious damage.
8. 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.

- Indeed, it takes some time for the pumped liquid to reach the fault areas, where the stresses are located.

- However, for the purposes of 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:

$$g(t) = F(f(t)) \text{ for some function } F(f).$$
9. 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 \( \lambda \) times smaller than the previous one,
  - then all numerical values are multiplied by \( \lambda \).

- 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.
10. Scale-invariance (cont-d)

- 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 unit for velocity from km/h to m/h.
11. Scale-invariance (cont-d)

- 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 $\mu$ – depending on $\lambda$ – for which $g = F(f)$ implies that $g' = F(f')$, where $g' \stackrel{\text{def}}{=} \mu \cdot g$.

- In economics, scaling corresponds, e.g., to changing a currency in which we gauge incomes and prices.
12. 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 that all measurable – in particular, all continuous – solutions of this functional equation have the form

$$F(f) = B \cdot f^\alpha,$$

for some constants $B$ and $\alpha$. 
13. 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 above formula, we conclude that $g(t) = A \cdot (f(t))^\alpha \cdot \int^t f(s) \, ds$. 
14. Conclusions

- In hydraulic fracturing-induced seismicity, we get exactly the desired formula

\[ g(x) = A \cdot (f(x))^\alpha \cdot \int_0^x f(y) \, dy. \]

- We did not use any specific features of this phenomenon, so we have a general explanation for the above formula.

- In economics, we can make similar arguments about, e.g., investing 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.

- The larger the economy, the more efficiently it can use new funds.
15. Discussion

- For seismicity, the above formula is a very crude approximation.
- This is in contrast to many econometric applications, where this dependence 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 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.
16. 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-1834620 and HRD-2034030 (CAHSI Includes).
- It was also supported by the AT&T Fellowship in Information Technology.
- It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.