Why Gaussian Copulas Are Ubiquitous in Economics: Fuzzy-Related Explanation

Chon Van Le, Olga Kosheleva, and Vladik Kreinovich

Abstract In many real-life situations, deviations are caused by a large number of independent factors. It is know that in such situations, the distribution of the resulting deviations is close to Gaussian, and thus, that the copulas – that describe the multi-D distributions as a function of 1-D ones – are also Gaussian. In the past, these conclusions were also applied to economic phenomena, until the 2008 crisis showed that in economics, Gaussian models can lead to disastrous consequences. At present, all economists agree that the economic distributions are not Gaussian – however, surprisingly, Gaussian copulas still often provide an accurate description of economic phenomena. In this paper, we explain this surprising fact by using fuzzy-related arguments.

1 Formulation of the Problem

Gaussian distributions are ubiquitous. Gaussian (normal) distributions are named after the great German mathematician and physicist Karl Friedrich Gauss (1777-1855) who discovered that these distributions adequately describe many real-world phenomena.

Later, the ubiquity of these distributions got a theoretical explanation: it turns out that under reasonable conditions, the distribution of the sum of a large number relatively small independent random variables is close to Gaussian – and the more variables we add, the closer the resulting distribution to Gaussian. This result is known as the Central Limit Theorem; see, e.g., [11].

Chon Van Le
International University of Ho-Chi-Minh City, Vietnam, e-mail: lvchon@hcmiu.edu.vn

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA
e-mail: olgak@utep.edu, vladik@utep.edu

In many real-life phenomena, what we observe is the result of the joint effect of many small factors – e.g., what we view as noise during measurement is caused by a large number of small independent factors. Not surprisingly, in the majority of measuring instruments, the distribution of measurement error is indeed close to Gaussian; see, e.g., [9, 10].

From distributions to copulas. Central Limit Theorem implies that both 1-D and multi-D distributions are Gaussian.

For 1-D distributions, the most widely used ways of describing such distributions are probability density functions $f(x)$ and cumulative distribution functions (cdfs)

$$F(x) \stackrel{\text{def}}{=} \text{Prob}(X \le x).$$

In the multi-D case, it is also possible to use probability density functions $f(x_1,\ldots,x_n)$ and cumulative distribution functions

$$F(x_1,\ldots,x_n) \stackrel{\text{def}}{=} \text{Prob}(X_1 \le x_1 \,\&\, \ldots \,\&\, X_n \le x_n).$$

However, in the multi-D case there is another convenient way of describing the distribution: by describing:

- *marginal* cfds $F_i(x_i) \stackrel{\text{def}}{=} \text{Prob}(X_i \le x_i)$, and
- a function $C(p_1,\ldots,p_n)$ for which

$$F(x_1,\ldots,x_n) = C(F_1(x_1),\ldots,F_n(x_n));$$

  this function $C(p_1,\ldots,p_n)$ is known as a *copula*.

The advantage of such a copula-based representation is related to the fact that often, there are different scale for measuring each quantity x_i. For example, we can measure length in meters, in centimeters, or in inches – and in all three cases, the same length is described by different numerical values. If we re-scale one of the variables – or even several variables, then:

- the probability density function $f(x_1,\ldots,x_n)$ and the cumulative distribution function $F(x_1,\ldots,x_n)$ change,
- but the copula remains the same.

This scaling-invariance is one of the main reasons why copulas are actively used in many applications.

From Gaussian distributions to Gaussian copulas. For each multi-D family of distributions, there is a corresponding family of copulas – copulas corresponding to distributions from this family. In particular, copulas corresponding to multi-D Gaussian distributions are known as *Gaussian copulas*.

Gaussian distributions and Gaussian copulas in economics: initial successes. In the past, in line with the above general idea, specialists in economics also used normal distributions (and the corresponding Gaussian copulas) to describe economic phenomena – and used them reasonably successfully.

Gaussian distributions in economics: crisis. One of the properties of a normal distribution is that deviations from the mean which are larger then 3 standard deviations are extremely rare: they occur in 0.1% of the cases. Deviations larger than 6 standard deviations are even rarer: they occur once in 100 million cases.

So, if we assume that an economic process – such as stock market prices – is normally distributed, we can safely ignore the possibility that these prices will go down by more than 6 standard deviations. This is exactly what financial folks assumed – and then came the 2008 crisis, when the prices unexpected dropped even more than 6 standard deviations. This was a disaster, quite a few companies relying on the Gaussian-derived stability of stock market went bankrupt, economies tanked.

Statistician almost immediately found out what went wrong: a detailed analysis of the behavior of stock marker prices and other economic characteristics showed that their actual distribution was differen from Gaussian; see, e.g., [2, 5].

Mystery: distributions are not Gaussian, but Gaussian copulas still apply. As we have mentioned, Gaussian copulas are derived from Gasussian distributions. So, since the distributions turned out to be non-Gaussian, it was natural to expect that the copulas would turn out to be non-Gaussian as well. But, strangely, in many case, Gaussian copulas still provide a very accurate description of economic phenomena; see, e.g., [3] and references therein. How can we explain this?

What we do in this paper. In this paper, we provide an explanation for this unexpected success of Gaussian copulas, an explanation that used fuzzy-related ideas.

2 Analysis of the Problem and the Resulting Explanation

Why not Gaussian: let us analyze. Economic deviations are also caused by a large number of small independent events, so why do not we get a Gaussian distribution here? The Central Limit Theorem – that explains Gaussian distributions – assumes that the joint effect of two small factors is equal to the sum of the effects of each of these factors. In other words, it assumes that the factors do not interact with each other.

This assumption may be true for the noise, where different noise components simply add to each other. However, economy is more complicated. In economy, everything is interrelated, and the joint effect of two factors is, in general, different from a simple sum of the effects of individual factors. For example, for a small company, inflation may be an annoying but possible-to-live-with problem, and tax increase may be also not pleasant but tolerable, but the joint effect of these two seemingly minor problems can bring the company into bankruptcy.

So how can we describe this situation? The above argument shows that in economics, to adequately describe the joint effect of several factors, we cannot use addition, we must use some other operation $a * b$. What are the natural properties of such an operation?

- First, the joint effects of two or three factors should depend on the order in which we combine these factors. So, we should have $a*b=b*a$ (commutativity) and $a*(b*c)=(a*b)*c$ (associativity).
- Second, if one of these factors is missing – e.g., if $a=0$ – the joint effect should simply coincide with another one: $0*b=b$.
- The joint effect should larger than each of the effects, i.e., unless either $a>0$ or $b>0$ is already a disaster (maximally possible effect), we should have $a<a*b$ and $b<a*b$.
- Finally, small changes in a and b should cause small changes in $a*b$. In other words, the function $a,b\mapsto a*b$ should be continuous.

Operations with such properties are known. The above properties are – almost exactly – the properties that define Achimedean "or"-operations (t-conorms); see, e.g., [1, 4, 6, 7, 8, 12] in fuzzy logic.

It is known that all such operations have the form $a*b=f^{-1}(f(a)+f(b))$ for some monotonic function $f(a)$, where $f^{-1}(a)$ denotes the inverse function, i.e., the function for which $f(a)=b$ if and only $f^{-1}(b)=a$.

This explains the ubiquity of Gaussian copulas. Indeed, the formula $a*b=f^{-1}(f(a)+f(b))$ can be equivalently described as $f(a*b)=f(a)+f(b)$. Thus, in general,

$$f(a_1*\ldots*a_n)=f(a_1)+\ldots+f(a_n).$$

So, if, to describe the effect, instead of the values in the original scale $a,b,\ldots$, we will use values $A\stackrel{\text{def}}{=}f(a)$, $B\stackrel{\text{def}}{=}f(b)$, ..., then in this new scale, the joint effect A of several factors $A_1,\ldots,A_n$ is simply equal to the sum of the individual effects

$$A=A_1+\ldots+A_n.$$

Thus, in this new scale, the joint effect is simply the sum of individual effects. So, by the Central Limit Theorem, the distribution of the joint effect is Gaussian. Therefore, the corresponding copula is Gaussian as well.

We have already mentioned that while non-linear re-scaling changes the marginal distributions, it does not change the copula. Thus, while marginal distributions are non-Gaussian, the copula remains Gaussian.

This is exactly the strange phenomenon that we have been trying to explain – now we have an explanation.

Acknowledgments

This work was supported 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),

- the AT&T Fellowship in Information Technology,
- the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. C. Donnelly and P. Embrechts, “The devil is in the tails: actuarial mathematics and the subprime mortgage crisis”, *ASTIN Bulletin*, 2010, Vol. 40, No. 1, pp. 1–33.
3. N. Kaewsompong and T. Chitkasame, “Economic policy uncertainty and stock-bond correlations: evidence from the Thailand market”, In: S. Sriboonchitta, V. Kreinovich, and W. Yamaka (eds.), *Credible Asset Allocation, Optimal Transport Methods, and Related Topics*, Springer, Cham, Switzerland, 2022, to appear.
4. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
5. A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, New Jersey, 2015.
6. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
7. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
8. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
9. P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991, in Russian.
10. A. I. Orlov, “How often are the observations normal?”, *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.
11. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
12. L. A. Zadeh, “Fuzzy sets”, *Information and Control*, 1965, Vol. 8, pp. 338–353.