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

Chon Van Le[1], Olga Kosheleva[2], and Vladik Kreinovich[2]

[1]International University of Ho-Chi-Minh City, Vietnam
lvchon@hcmiu.edu.vn

[2]University of Texas at El Paso, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

1. Gaussian distributions are ubiquitous

- Gaussian (normal) distributions are named after the great German mathematician and physicist Karl Friedrich Gauss (1777-1855).
- He discovered that these distributions adequately describe many real-world phenomena.
- Later, the ubiquity of these distributions got a theoretical explanation:
  - under reasonable conditions,
  - the distribution of the sum of a large number relatively small independent random variables is close to Gaussian.
- The more variables we add, the closer the resulting distribution to Gaussian.
- This result is known as the Central Limit Theorem.

2. Gaussian distributions are ubiquitous (cont-d)

- 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.

3. 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

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

4. From distributions to copulas (cont-d)

- 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 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.
- In all three cases, the same length is described by different numerical values.

5. From distributions to copulas (cont-d)

- 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.

6. From Gaussian distributions to Gaussian copulas

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

7. Gaussian distributions and Gaussian copulas in economics: initial successes

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

8. 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.

9. Gaussian distributions in economics: crisis (cont-d)

- 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 showed that their actual distribution was differen from Gaussian.

10. Mystery: distributions are not Gaussian, but Gaussian copulas still apply

- As we have mentioned, Gaussian copulas are derived from Gaussian distributions.
- The distributions turned out to be non-Gaussian.
- So, 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.
- How can we explain this?
- In this talk, we provide an explanation for this unexpected success of Gaussian copulas, an explanation that used fuzzy-related ideas.

11. 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 noise, where different noise components simply add to each other.
- However, economy is more complicated.

12. Why not Gaussian: let us analyze (cont-d)

- In economy, everything is interrelated.
- 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,
  - he joint effect of these two seemingly minor problems can bring the company into bankruptcy.

13. 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 not 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.
- So, unless either $a > 0$ or $b > 0$ is already a disaster (maximally possible effect), we should have $a < a * b$ and $b < a * b$.

14. So how can we describe this situation (cont-d)

- 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.

15. Operations with such properties are known

- The above properties are – almost exactly – the properties that define Achimedean "or"-operations (t-conorms) in fuzzy logic.
- It is known that all such operations have the form

$$a * b = f^{-1}(f(a) + f(b)) \text{ for some monotonic function } f(a).$$

- Here $f^{-1}(a)$ denotes the inverse function, i.e., the function for which $f(a) = b$ if and only $f^{-1}(b) = a$.

16. 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, to describe the effect,
  - instead of the values in the original scale $a, b, \ldots$,
  - we can use values $A \stackrel{\text{def}}{=} f(a)$, $B \stackrel{\text{def}}{=} f(b)$, ...
- 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.

17. This explains the ubiquity of Gaussian copulas (cont-d)

- 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.

18. 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.