How to adequately describe generic non-Gaussian multi-D distributions: sliced-normal approach and its natural generalization

Michael Beer[1], Olga Kosheleva[2], Vladik Kreinovich[3], and Uyen Hoang Pham[4]
[1]Institute for Risk and Reliability, Leibniz University Hannover, Callinstr. 34, 30167 Hannover, Germany, beer@irz.uni-hannover.de,
[2,3] Departments of [2]Teacher Education and [3]Computer Science
University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, olgak@utep.edu, vladik@utep.edu
[4]University of Economics and Law – Vietnam National University
Ho Chi Minh City, Vietnam, uyenph@uel.edu.vn

1. Gaussian distributions are ubiquitous

- In many practical situation, random quantities are normally distributed.
- For such distributions, $f(x_1,\ldots,x_n)=\exp(-Q(x_1,\ldots,x_n))$ for some quadratic function $Q(x_1,\ldots,x_n)$.
- This fact was first noticed by Gauss himself – after whom these distributions are named.
- Gauss noticed that a similar distribution was observed when he measured different quantities, ranging:
  - from distances and angles
  - to characteristics of the electromagnetic field.

2. Gaussian distributions are ubiquitous (cont-d)

- This ubiquity remains: e.g., a reasonably recent empirical study showed that:
  - in 60% of the measuring instruments,
  - the distribution of measurement error is close to Gaussian.
- This fact – and similar facts – explains why Gaussian distributions are also called *normal*.
- Most distributions are Gaussian; in this sense Gaussian distributions are a norm,

3. Why are Gaussian distributions ubiquitous?

- Normal distributions are so ubiquitous that many people do not even ask a why-question. There is a joke (not far from the truth).
- Mathematicians think that it is an empirical law – that all probability distributions are Gaussian.
- Engineers think that mathematicians have proved a theorem that all distributions are Gaussian.
- This is, of course, a joke, but there is, indeed, a mathematical theorem explaining why many empirical distributions are Gaussian.
- Fact: in many real-life situations, a random quantity is the joint effect of many small independent factors.
- For example, this is a typical situation with many high-accuracy measuring instruments.

4. Why are Gaussian distributions ubiquitous (cont-d)

- Once we eliminate all main components of the measurement error, what is left is a combination of many small independent factors.
- In such situations, the Central Limit Theorem proves that the resulting summary distribution is close to Gaussian.
- The more components are combined, the closer the summary distribution to Gaussian.
- This explains why many actual real-life probability distributions are close to Gaussian.

5. We often need to go beyond Gaussian distributions: general argument

- While many real-life distributions are close to Gaussian, many are not.
- Yes, the measurement error is normally distributed for 60% of measuring instruments.
- This means that in 40% of measuring instruments the probability distribution of measurement error is clearly non-Gaussian; so:
  - to analyze measurement errors – and many other random quantities,
  - we need to be able to describe and analyze non-Gaussian distributions.

6. We often need to go beyond Gaussian distributions: economic argument

- Non-Gaussian distributions are ubiquitous in many applications areas.
- It turns out that in economics and finance, the difference between Gaussian and non-Gaussian distribution is very important.
- Indeed, in science and engineering, we need to study non-Gaussian distributions to get a more accurate description of reality.
- So Gaussian distributions provide a good first approximation to reality.
- However, in economics, the situation is drastically different.
- Before the 2008 world-wide economic crisis, many economists:
  - followed the example of science and engineering and
  - assumed, in their recommendations, that most distributions are Gaussian.

7. We often need to go beyond Gaussian distributions: economic argument (cont-d)

- In particular, they assumed that the stock prices follow a normal distribution.
- In a normal distribution, with confidence $1 - 10^{-8}$, all the changes of a random variables do not exceed 6σ.
- Here σ denotes the standard deviation.
- So, they concluded that stock price cannot fall by more than 6σ – and made investment strategies based on this conclusion.
- As a result, when the stock market fell by more than 6σ:
  - most people were not prepared – since they did not expect such a drastic decrease, and
  - a serious crisis hit.

8. How to describe non-Gaussian distributions are described now: let us start

- It is known that a normal distribution is uniquely determined by its first and second moments.
- Vice versa, for any consistent combination of first and second moments, there is a normal distribution with exactly these moments.
- Thus, for normal distribution, the first two moments uniquely determine third and higher moments.
- In practice, as we have mentioned, distributions are often non-Gaussian.
- Thus, the third and higher moments are, in general, different from their Gaussian values.
- We want to find a model that correctly describes:
  - not only the first two empirical moments,
  - but also the third moment.

9. How to describe non-Gaussian distributions are described now: let us start (cont-d)

- So, we need to go beyond Gaussian models.
- Gaussian distributions are symmetric with respect to reversion over the center x_0: $x - x_0 \mapsto x_0 - x$.
- For such symmetric distributions, the third central moment is 0.
- So, a non-zero third central moment corresponds to the case when the distribution is not symmetric – i.e., skewed.
- Because of this, the corresponding phenomenon is known as *skewness*.

10. How skewness is taken into account

- There are several general models of distributions with given three moments.
- One of the most widely used is so-called *skew-normal* distribution.
- In particular, such distributions have been successfully used in economics.
- One of the main reasons why this particular distribution is successful is that it can be naturally derived based on commonsense requirements.

11. General case: general idea of the sliced-normal approach

- Skew-normal distributions describe the case when we take into account the first three moments.
- If we want to take into account more moments, we need to have a more general family of distributions.
- One of such successful families is the family of *sliced-normal* distributions.
- There are explanations of why reasonable requirements naturally leads to this family.
- The idea behind this family is straightforward.
- We perform some transformation

$$(x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_n) = (P_1(x_1, \ldots, x_n), \ldots, P_n(x_1, \ldots, x_n)).$$

12. General case: general idea of the sliced-normal approach (cont-d)

- We assume that:
  - the resulting values y_i are independent, and
  - each of these values y_i is normally distributed with mean 0 and standard deviation 1.
- In this case, we have an explicit expression for the probability density $f(x_1, \ldots, x_n)$ of the resulting distribution:

$$f(x_1, \ldots, x_n) = \exp\left(-\frac{\sum_{i=1}^{n} P_i^2(x_1, \ldots, x_n)}{2}\right).$$

- The functions $P_i(x_1, \ldots, x_n)$ that describe the transformations are selected among linear combinations

$$c_1 \cdot b_1(x_1, \ldots, x_n) + \ldots + c_m \cdot b_m(x_1, \ldots, x_n).$$

13. General case: general idea of the sliced-normal approach (cont-d)

- Here $b_j(x_1, \ldots, x_n)$ are given functions and c_j are the coefficients that we need to determine from the data.
- Good news about the sliced-normal distributions is that the corresponding likelihood function is convex.
- Thus, we can feasibly find the optimal values $c_1, \ldots, c_m$.
- At present, usually, as the functions $b_j(x_1, \ldots, x_n)$, we use *monomials*.
- Monomials are functions of the type $b_j(x_1, \ldots, x_n) = x_1^{d_{1j}} \cdot \ldots \cdot x_n^{d_{nj}}$.
- Here, d_{ij} are non-negative integers.

14. Need to use inverses

- In many practical situation, one can alternatively use either the original quantity x, or its inverse $1/x$, and both representations are useful.
- Here are some examples.
- In general, in physics, we use velocity v to describe how fast an object moves.
- However, in seismology, it is more convenient to use an inverse $1/v$ to velocity known as *slowness*.
- In electrical engineering, instead of resistance R, it is often helpful to use its inverse known as *conductivity*.
- A periodic process can be sometimes better described by its period, and sometimes by frequency – the inverse of a period.

15. Need to use inverses (cont-d)

- Car's fuel consumption can be described in two mutually inverse scales:
  - in liters per kilometer (or, in US units, in gallons per mile),
  - or in kilometers per liter (correspondingly, miles per gallon).
- There are many similar economic examples.
- The most natural is exchange rate.
- The amount of US dollars one gets from 1 Euro is the exact inverse to how many dollars one gets for 1 Euro.
- There are many other examples.
- Suppose that you have a working plant with a fixed number of workers.
- One of the tasks is to come up with an optimal number of low-level managers who can manage all these workers.

16. Need to use inverses (cont-d)

- In this case, a natural unit is number of managers per worker.
- In other cases, you only have a small group of skilled managers.
- In this case, a natural task is to find the optimal number of workers that these managers can supervise.
- If needed, we can add more workers later, when we train or hire more experienced managers.
- In this case, a natural unit is number of workers per manager.
- This is the exact inverse to the number of managers per worker.

17. This leads to a problem with the traditional sliced-normal approach

- We would like to have a general family of distributions that would be applicable both:
  - when we use the original quantity and
  - when we use its inverse.
- It would be weird to have a family:
  - that *can* be applied when we use a dollar-to-Euro exchange rate,
  - but that *cannot* be applied when we use the inverse Euro-to-dollar exchange rate.
- But this is exactly what happen if we use the usual sliced-normal approach.
- Indeed, this approach is based on having a transformation $(x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_n)$.

18. This leads to a problem with the traditional sliced-normal approach (cont-d)

- Here $y_i = P_i(x_1, \ldots, x_n)$ polynomially depend on the inputs and y_i are normally distributed.
- If we use an inverse $X_1 = 1/x_1$ instead of the first quantity, the expression of y_j in terms of X_1 is no longer a polynomial
- Indeed, even in the simplest case when $n = 1$ and $y_1 = P_1(x_1) = x_1$, we have $y_1 = 1/X_1$ – which is not a polynomial.

19. Resulting challenge

- We need to extend the traditional sliced-normal approach, so that:
  - the resulting family of transformation functions
  - be invariant with respect to the inverse $x_i \mapsto 1/x_i$.
- This means that:
  - if a mapping $x_1, \ldots, x_n \mapsto y_j = P_j(x_1, \ldots, x_n)$ is contained in this family,
  - then the following mapping should also belong to this family:

$$x_1, \ldots, x_n \mapsto P_j\left(x_1, \ldots, x_{i-1}, \frac{1}{x_i}, x_{i+1}, \ldots, x_n\right)$$

- We do not want to add too many functions.

20. Resulting challenge (cont-d)

- Indeed, one of the main advantages of the sliced-normal approach is that:
  - it enables to reduce the number of parameters
  - that is needed to represent generic non-Gaussian distributions.
- If we enlarge the family too much, we will lose this advantage.
- So, what we need is the smallest possible inversion-invariant extension of the class of polynomials.
- In this talk, we describe such smallest invariant extension.

21. The resulting generalization of the usual sliced-normal approach

- Let us first describe what we want in precise terms.
- We say that a family F of functions of n variables is *inversion-invariant* if:
  - for every function $P(x_1, \ldots, x_n) \in F$ and for every i,
  - the function $P\left(x_1, \ldots, x_{i-1}, \dfrac{1}{x_i}, x_{i+1}, \ldots, x_n\right)$ also belongs to F.
- We say that a family A is *smaller* then a family B if $A \subseteq B$.
- We say that a family C is the *inversion hull* of the family F if:
  - first, C is inversion-invariant and contains F, and
  - second, C is smaller than every other inversion-invariant family G containing F.
- In these terms, what we want is to find the inversion hull of the family F used in the sliced-normal approach.

22. The resulting generalization of the usual sliced-normal approach (cont-d)

- In particular, we want to find the inversion hull of the family of all linear combinations of several monomials.
- We will first present the general result, and then we will describe what will happen for monomials.

23. General result and its proof

- Let F be the family of all possible linear combinations of the functions $b_j(x_1, \ldots, x_n)$.
- Then, its inversion hull C is the set of all possible linear combinations of the functions $b_j(x_1^{\varepsilon_1}, \ldots, x_n^{\varepsilon_n})$, where $\varepsilon_i \in \{-1, 1\}$ for all i.
- **Proof.** Let us first show that all new functions $b_j(x_1^{\varepsilon_1}, \ldots, x_n^{\varepsilon_n})$ are included in any inversion-invariant extension G of F.
- Indeed, in each such extension, due to inversion invariance:
  - we can change one of the inputs from x_j to $1/x_j$, and
  - the function remains in G.
- We can perform this change for all the inputs x_i for which $\varepsilon_i = -1$.
- Then, we will conclude that the resulting new function is indeed contained in G.

24. Proof of the general result (cont-d)

- To complete the proof, it is sufficient to prove that the set of linear combinations of new functions is itself inversion invariant.
- Indeed, applying the inversion to a linear combination is equivalent to a linear combination of inverted functions.
- And these inverted functions are all in C.
- The proposition is proven.

25. Case of monomials: result and examples

- Let F be the family of all possible linear combinations of monomials $b_j(x_1, \ldots, x_n) = x_1^{d_{1j}} \cdot \ldots \cdot x_n^{d_{nj}}$.
- Then, its inversion hull is the family of all possible linear combinations of expressions $x_1^{\pm d_{1j}} \cdot \ldots \cdot x_n^{\pm d_{nj}}$ for all combinations of signs.
- Let $n = 1$ and the family F consists of linear combinations of monomials 1 and x_1.
- Then its inversion hull C is the set of all linear combinations of the expressions 1, x_1, and $1/x_1$.
- When $n = 2$ and the family F consists of linear combinations of monomials 1, x_1, x_2, and $x_1 \cdot x_2$.
- Then its inversion hull C is the set of all linear combinations of the expressions

$$1, \quad x_1, \quad \frac{1}{x_1}, \quad x_2, \quad \frac{1}{x_2}, \quad x_1 \cdot x_2, \quad \frac{x_1}{x_2}, \quad \frac{x_2}{x_1}, \text{ and } \frac{1}{x_1 \cdot x_2}.$$

26. Important comment

- There is an important particular case of a linear combination of monomials.
- It is when we approximate a function by the sum of the first few terms of its Taylor series

$$\sum_{i_1=0}^{\infty} \ldots \sum_{i_n=0}^{\infty} c_{i_1,\ldots,i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}.$$

- From this viewpoint, the inversion hull is a similar approximation to the *Laurent series*.
- In these series, each variable can be raised to negative powers as well:

$$\sum_{i_1=-\infty}^{\infty} \ldots \sum_{i_n=-\infty}^{\infty} c_{i_1,\ldots,i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}.$$

27. Example

- Let us again consider the case when $P_1(x_1) = 1/x_1$.
- Then $x_1 = P_1^{-1}(y_1) = 1/y_1$ is simply the inverse of a standard Gaussian random variable y_1 with mean 0 and standard deviation 1.
- According to the general formulas:
  - ased on the probability density function (pdf)
  $$f(y_1) = \frac{1}{2 \cdot \sqrt{\pi}} \cdot \exp\left(-\frac{y_1^2}{2}\right),$$
  - we can describe the pdf $g(x_1)$ as $g(x_1) = f(P_1(x_1)) \cdot |P_1'(x_1)|$, where $P_1'(x_1)$, as usual, denotes the derivative.
- in our specific case, $P_1'(x_1) = -1/x_1^2$, so
$$g(x_1) = \frac{1}{x_1^2} \cdot \frac{1}{2 \cdot \sqrt{\pi}} \cdot \exp\left(-\frac{1}{2x_1^2}\right).$$
- This expression tends to 0 both when $x_1 \to 0$ and when $x_1 \to \pm\infty$.

28. Example (cont-d)

- So, we get a bimodal distribution.
- In the traditional sliced-normal description, we would need a more complex expression to get a similar bimodal distribution.
- The possibility to get a simpler representation is an additional advantage of the proposed generalization.

29. Acknowledgments

This work was supported by:

- the AT&T Fellowship in Information Technology,
- the Institute for Risk and Reliability, Leibniz Universitaet Hannover, Germany,
- the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Focus Program SPP 100+ 2388, Grant Nr. 501624329,
- the European Union under the project ROBOPROX (No. CZ.02.01.01/00/22 008/0004590),
- the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand,
- the Ho Chi Minh City University of Banking, Vietnam, and
- Thang Long University, Hanoi, Vietnam.