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

Michael Beer, Olga Kosheleva, Vladik Kreinovich, and Uyen Hoang Pham

Abstract In many practical situations, distributions are non-Gaussian. One of the most efficient ways to describe a general non-Gaussian distribution is to use the sliced-normal approach, when we find a polynomial transformation that transform the original multi-D distribution into a Gaussian one. However, this approach has a limitation. This limitations comes from the fact that in many cases, it should not matter whether we use the original quantity or its inverse. For example, when studying fluctuations of exchange rates, it should not matter whether we use dollar-to-Euro or Euro-to-dollar ratio. However, when we use the traditional sliced-normal approach, in these two cases, we get different results. In this paper, we show how to generalize the traditional sliced-normal approach so as to eliminate this difference.

1 Practical problem: need to deal with non-Gaussian distributions

Gaussian distributions are ubiquitous. In many practical situation, random quantities are normally distributed, i.e., distributed according to a Gaussian distribution,

Michael Beer
Institute for Risk and Reliability, Leibniz University Hannover, Callinstr. 34,
30167 Hannover, Germany, e-mail: beer@irz.uni-hannover.de

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

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

Uyen Hoang Pham
University of Economics and Law – Vietnam National University
Ho Chi Minh City, Vietnam, e-mail: uyenph@uel.edu.vn

with the probability density $f(x_1,\dots,x_n) = \exp(-Q(x_1,\dots,x_n))$ for some quadratic function $Q(x_1,\dots,x_n)$. This fact was first noticed by Gauss himself – after whom these distributions are named – who noticed that a similar distribution was observed when he measured different quantities, ranging from distances and angle to characteristics of the electromagnetic field.

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; see, e.g., [11, 12]. 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,

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) that:

- mathematicians think that it is an empirical law – that all probability distributions are Gaussian, while
- 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. The use of this theorem is based on the fact that 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: 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 (see, e.g., [16]) proves that the resulting summary distribution is close to Gaussian – and 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.

We often need to go beyond Gaussian distributions: general argument. While many real-life distributions are close to Gaussian, many are not. For example, the fact that the measurement error is normally distributed for 60% of measuring instruments 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.

We often need to go beyond Gaussian distributions: economic argument. Non-Gaussian distributions are ubiquitous in many applications areas. It turns out that in some application areas – such as 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 – but 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. 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, where 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.

2 How non-Gaussian distributions are described now: from skew-normal to sliced-normal approaches

Why start with skewness. It is known that a normal distribution is uniquely determined by its first and second moments – and, vice versa, for any consistent combination of first and second moments, there is a normal distribution with exactly these moments. Since for the normal distribution, the first two moments uniquely determine the whole distribution – and thus, 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.

So, if we want to find a model that correctly describes not only the first two empirical moments, but also the third moment, 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*.

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; see, e.g., [1]. In particular, such distributions have been successfully used in economics: see, e.g., [10, 19, 20]; see also [18]. One of the main reasons why this particular distribution is successful is that it can be naturally derived based on commonsense requirements; see, e.g., [9].

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; see, e.g., [2, 3, 4, 5, 6, 7, 8, 13]. An explanation of why reasonable requirements naturally leads to this family is given in [2].

The idea behind this family is straightforward: we perform some transformation

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

We assume that the resulting values y_i are independent, and that 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:

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

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

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

where $b_j(x_1,\ldots,x_n)$ 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, and thus (see, e.g., [14]), we can feasibly find the optimal values $c_1,\ldots,c_m$.

How sliced-normal approach is used now. At present, usually, as the functions $b_j(x_1,\ldots,x_n)$, we use *monomials*, i.e., functions of the type

$$b_j(x_1,\ldots,x_n) = x_1^{d_{1j}} \cdot \ldots \cdot x_n^{d_{nj}}$$

for some non-negative integers d_{ij}.

3 Research problem: limitations of the traditional sliced-normal approach

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*; see, e.g., [15, 17].
- 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.
- 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.

For example, if 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. 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 – 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 – which is the exact inverse to the number of managers per worker.

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 $(x_1,\ldots,x_n) \mapsto (y_1,\ldots,y_n)$, where $y_i = P_i(x_1,\ldots,x_n)$ 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.

Resulting challenge. In view of the above problem, we need to extend the traditional sliced-normal approach, so that the resulting family of transformation functions be invariant with respect to the inverse $x_i \mapsto 1/x_i$ – i.e., 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 mapping

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

should also belong to this family.

We do not want to add too many functions. 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.

What we do in this paper. In this paper, we describe such smallest invariant extension.

4 The resulting generalization of the usual sliced-normal approach

Let us first describe what we want in precise terms.

Definition 1.

- *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*

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

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 C is inversion-invariant and contains F, and is smaller than every other inversion-invariant family G containing F.*

In terms of this definition, what we want is to find the inversion hull of the family F used in the sliced-normal approach. In particular, we want to find the inversion hull of the family of all linear combinations of several monomials.

In this section, we will first present the general result, and then we will describe what will happen for monomials.

Proposition 1. *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. Once we perform this change for all the inputs x_i for which $\varepsilon_i = -1$, we will conclude that the resulting new function is indeed contained in G.

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 they are all in C. The proposition is proven.

Discussion. In particular, if we start with the traditional sliced-normal approach – where the functions b_j are monomials – we get the following result.

Proposition 2. *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.*

Example 1. When $n = 1$ 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

$$1,\ x_1,\ \text{and}\ \frac{1}{x_1}.$$

Example 2. 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

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

Comment. An important particular case of a linear combination of monomials is when we approximate a function by the sum of the first few terms of its Taylor series

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

From this viewpoint, the inversion hull is a similar approximation to the *Laurent series* – in which each variable can be raised to negative powers as well:

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

Example 1 (continued). 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, based 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 $x_1 \to \pm\infty$. 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 of the traditional sliced-normal approach.

Acknowledgments

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

References

1. A. Azzalini, *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge, UK, 2013.
2. M. Beer, O. Kosheleva, and V. Kreinovich, "Fusion of probabilistic knowledge as foundation for sliced-normal approach", *Proceedings of the 9th International Workshop on Reliable Engineering Computing REC'2021*, Taormina, Italy, May 16–20, 2021, pp. 408–418.
3. B. K. Colbert, L. G. Crespo, and M. M. Peet, "A sum of squares optimization approach to uncertainty quantification", In: *Proceedings of the 2019 American Control Conference ACC 2019*, Philadelphia, Pennsylvania, USA, July 10–12, 2019, pp. 5378–5384.
4. B. K. Colbert, L. G. Crespo, and M. M. Peet, " "A convex optimization approach to improving suboptimal hyperparameters of sliced normal distributions", *Proceedings of the 2020 American Control Conference ACC 2020*, Denver, Colorado, USA, July 1–3, 2020, pp. 4478–4483, doi: 10.23919/ACC45564.2020.9147403.
5. L. Crespo, "An introduction to sliced-normal distributions", In M. Beer and E. Zio, editors, *Proceedings of the 29th European Safety and Reliability Conference ESREL'2019*, Hannover, Germany, September 2019. Research Publishing, Singapore, p. 28.
6. L.-G. Crespo, B. K. Colbert, S. P. Kenny, and D. P. Giesy, "On the quantification of aleatory and epistemic uncertainty using Sliced-Normal distributions", *Systems and Control Letters*, 134, Paper 104560, 2019.
7. L.-G. Crespo, B. K. Colbert, S. P. Kenny, and D. P. Giesy, "Dimensionality reduction of sliced-normal sistributions", *IFAC-PapersOnLine*, 2020, Vol. 53, No. 2, pp. 7412–7417.
8. L.-G. Crespo and S. M. Snyder, "Polynomial approximations to the sliced-normal density", In: E. B. Abrahamsen, T. Aven, F. Bouder, R. Flage, and Marja Ylönen (eds.), *Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference ESREL/SEA-E*, Stavanger, Norway, June 15–19, 2025, pp. 187–194, doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P7789-cd
9. D. L. Gallegas Espinosa, O. Kosheleva, and V. Kreinovich, "Why skew-normal distributions and how they are related to ReLU activation function in deep learning", In: H. T. T. Nguyen, H. H. Phan, and V. N. Huynh (eds.), *Data Science in Finance and Accounting*, Springer, Cham, Switzerland, to appear.
10. Z. Ma, T. Wang, S. T. B. Choy, Z. Wei, and X. Zhu, "Extending the A Priori Procedure for estimating location parameter under multivariate skew normal settings", In: N. N. Thach, V. Kreinovich, D. T. Ha, and N. D. Trung (eds.), *Optimal Transport Statistics for Economics and Related Topics*, Springer, Cham, Switzerland, 2024, pp. 107–116.
11. P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991 (in Russian).
12. A. I. Orlov, "How often are the observations normal?", *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.

13. Th. Potthast, N. Kitahara, M. Broggi, F. Mett, and M. Beer, *Stochastic Model Updating Through Sliced-Normal Maps*, Available at SSRN: https://ssrn.com/abstract=5050218 or http://dx.doi.org/10.2139/ssrn.5050218.
14. R. T. Rockafeller, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1997.
15. P. M. Shearer, *Introduction to Seismology*, Cambridge University Press, Cambridge, UK, 2019.
16. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
17. S. Stein and N. Wysession, "An Introduction to Seismology, Earthquakes, and Earth Structure", Wiley-Blackwell, Hoboken, New Jersey, USA, 1991.
18. K. Suriya, T. Sudtasan, T. Wang, O. Lerma, and V. Kreinovich, "A natural simple model of scientists' strength leads to skew-normal distribution", *International Journal of Intelligent Technologies and Applied Statistics (IJITAS)*, 2015, Vol. 8, No. 2, pp. 153–158.
19. T. Tong, X. Chen, T. Wang, D. Trafimow, S. T. B. Choy, and C. Wang, "The A Priori Procedure (APP) for estimating standardized regression coefficients in multiple linear model with skew errors", In: V. Kreinovich, S. Sriboonchitta, and W. Yamaka (eds.), *Machine Learning for Econometrics and Related Topics*, Springer, Cham, Switzerland, 2024, pp. 143–160.
20. Z. Wang, T. Wang, D. Trafimow, S. T. B. Choy, and X. Chen, "Extending a priori procedure for simultaneously estimating location and scale parameters in the context of skew normal distributions", In: N. N. Thach, N. D. Trung, D. T. Ha, and V. Kreinovich (eds.), *Artificial Intelligence and Machine Learning for Econometrics: Applications and Regulation (and Related Topics)*, Springer, Cham, Switzerland, to appear.