Exact distributions under non-standard model assumptions

Wolf-Dieter Richter

Institute of Mathematics, University of Rostock, Ulmenstraße 69, Haus 3, D-18057 Rostock, Germany

Abstract. A rather general method is presented of how to derive the exact distribution of a function of a random vector if the vector does not satisfy standard model assumptions.

Keywords: Exact distribution, geometric measure representation, generalized method of indivisibles, non-Euclidean surface measure, power-exponential distribution, $l_{n,p}$-symmetric distribution, elliptically contoured distribution, functions of random vectors.
PACS: 02.30.Cj, 02.50.Ng, 02.50.Sk

INTRODUCTION

Exact probability distributions of many functions of random vectors were derived several decades ago in case of a Gaussian vector distribution and later in case of a spherical or an elliptically contoured vector distribution. One of the basic references in this area is [2].

The class of continuous $l_{n,p}$-symmetric probability laws with $p > 0$ extends the class of Gaussian spherical laws in a two-fold way. While the density level sets of spherical distributions are Euclidean spheres, the much more general $l_{n,p}$-spheres play the same role in the class of $l_{n,p}$-symmetric distributions. Thus even such clouds of randomly generated points may be probabilistically modeled which look similar like a cube ($p \to \infty$) or like a non-convex star body ($p \in (0,1)$). Moreover, the density generating function of an $l_{n,p}$-symmetric distribution has essential influence onto whether this distribution has light or heavy tails, besides the effects being caused already by the parameter p. Thus, $l_{n,p}$-symmetric distributions open new broad perspectives of probabilistic modeling.

In their real life studies in [10] and [4] the authors show that already power-exponential distributions which are special examples of the general class of $l_{n,p}$-symmetric distributions fit in some cases much more flexible to the given data than the Gaussian distribution. Applications of heavy tails may occur, e.g., in insurance and financial mathematics, while sometimes in reliability theory especially light tails may be of interest.

The fundamental method considered in this paper for deriving exact distributions of statistics in the case of $l_{n,p}$-symmetric sample distributions is to apply a geometric measure representation of this sample distribution. Such a measure representation was derived in [18]. In the same paper, there is also gained an overview of several applications of the geometric representation of the Gaussian law in [16] and in subsequent papers. In the present paper, recent applications of the geometric measure representation formula of $l_{n,p}$-symmetric distributions will be discussed.

Yet before Leibniz and Newton created calculus, the pupils of Galilei, Cavalieri and Torricelli, developed a general method of determining the content of an possibly complicated structured area if the content of a comparable area is already known. The main idea of this method was to divide each area into its so called indivisibles, then to measure them and finally to compare the results for both areas. Measuring the indivisibles, clearly, was based upon Euclidean arc-length, area-content or volume. For an introduction into the history of this development, we refer to [14] and [1].

In the last decades of the last century, many authors made use of the notion of uniform distribution on $l_{n,p}$-spheres, partly when studying a disintegration formula of the Lebesgue measure and more over of the multivariate p-power exponential distribution, see, e.g., [7], [21], [15], [22], [9] and [23]. Such disintegration method could be understood as a modern generalisation of the historical method of indivisibles if the mentioned uniform distribution would have a well defined geometric meaning. This, however, was not the case for a long time. For a certain detailed description of this problem, see [23]. There, the author gains a specific insight into the area of uniform distributions on $l_{n,p}$-spheres and also draws the readers attention to the for a long time unsolved basic problem of understanding the geometric nature of this distribution class.

A real geometric explanation of the notion of uniform distributions on $l_{n,p}$-spheres and more over of the class of $l_{n,p}$-symmetric distributions was given in [18] on the basis of a suitably chosen notion of non-Euclidean surface content. This was the starting point for the non-Euclidean generalization of the method of indivisibles which will be used throughout this paper.

NON-EUCLIDEAN GENERALIZATION OF THE METHOD OF INDIVISIBLES

For simplicity, we start with the description of the two-dimensional version of this method before giving its multivariate generalization. For more details, we refer to [18] and the papers cited therein. To start with, let A be a measurable area in R^2 with minimal and maximal distances $r_1 \geq 0$ and $r_2 < \infty$ from the origin, respectively, and $C(r)$ the Euclidean circle of radius r centered at the origin, then its area content $\lambda(A)$ satisfies the equation

$$\lambda(A) = \int_{r_1}^{r_2} l(C(r) \cap A) dr$$

where l means Euclidean arc-length. The sets $C(r) \cap A$, $r > 0$ can be considered as the indivisibles of A. Using their central projections $\frac{1}{r}[C(r) \cap A]$ onto the unit circle $C = C(1)$, one gets the disintegration formula

$$\lambda(A) = \int_{r_1}^{r_2} rl([\frac{1}{r}A] \cap C) dr$$

for the Lebesgue measure in R^2 where measuring arc-lengths is restricted to Borel measurable pieces of C. Let now $C_p(r) := \{(x,y) \in R^2 : (|x|^p + |y|^p)^{1/p} = r\}$, $r > 0$, $p > 0$ denote the $l_{2,p}$-circle of p-radius r and $C_p = C_p(1)$, then it follows from the co-area formula of measure theory that, unless for $p = 2$,

$$\lambda(A) \neq \int_{r_1}^{r_2} rl([\frac{1}{r}A] \cap C_p) dr.$$

Let a curve $\mathscr{C}$ be given by a parametric representation $t \to (x(t), y(t))$. If we change the quantity $(x'^2(t) + y'^2(t))^{1/2} dt$ from the definition of the curve's Euclidean arc-length by $(|x'(t)|^q + |y'(t)|^q)^{1/q} dt$ with $\frac{1}{p} + \frac{1}{q} = 1$ and denote the resulting non-Euclidean arc-length by l_q, then $\lambda(A)$ satisfies the new disintegration formula

$$\lambda(A) = \int_{r_1}^{r_2} rl_q([\frac{1}{r}A] \cap C_p) dr,\ p > 0.$$

The indivisibles of A are now the sets $A \cap C_p(r)$, $r > 0$. Clearly, $l_2 = l$ and $l_q \neq l$ if $q \neq 2$. Notice that $q \geq 1$ iff $p \geq 1$ and $q \in (-\infty, 0)$ iff $p \in (0,1)$. For more background information concerning the latter case, see [12].

Let $\mathscr{O}_q$ denote the correspondingly generalized non-Euclidean surface content, S_p the $l_{n,p}$-unit sphere, λ the Lebesgue measure and $\mathscr{B}^n$ the Borel σ-field in R^n, then

$$\lambda(A) = \int_{r_1}^{r_2} r^{n-1} \mathscr{O}_q([\frac{1}{r}A] \cap S_p) dr,\ A \in \mathscr{B}^n.$$

This formula reflects the non-Euclidean method of indivisibles for arbitrary finite dimension. Analogous formulas for probability measures will be given in the next section where we consider a method of weighted indivisibles.

GEOMETRIC REPRESENTATION OF $l_{n,p}$-SYMMETRIC DISTRIBUTIONS

Let us consider first the additional weight function $\varphi_p(x,y) = C_p^2 \exp\{-\frac{1}{p}(|x|^p + |y|^p)\}$, $(x,y) \in R^2$ for the two-dimensional case where $C_p = p^{1-1/p}/(2\Gamma(1/p))$. Notice that φ_p is the pdf of the exponential p-power distribution Φ_p in R^2. This distribution may be interpreted as a generalization of the Gaussian law as well as of the Laplace one where $p = 2$ or $p = 1$, respectively. It satisfies the non-Euclidean geometric measure representation

$$\Phi_p(A) = C_p^2 \int_0^{\infty} rl_q([\frac{1}{r}A] \cap C_p) \exp\{-\frac{r^p}{p}\} dr,\ A \in \mathscr{B}^2.$$

A powerful extension of the class of exponential power distributions is the class of continuous $l_{n,p}$-symmetric distributions having a pdf $\phi_{g,p}(x,y) = C(g,p)g(|x|^p + |y|^p)$, $(x,y) \in R^2$. This pdf is used for defining a generalized method of weighted indivisibles where it is assumed that the density generating function (dgf) $g : [0,\infty) \to [0,\infty)$ satisfies with $I_{2,g,p} = \int_0^\infty rg(r^p)dr$ the assumption $0 < I_{2,g,p} < \infty$. The probability measure $\Phi_{g,p}$ corresponding to $\phi_{g,p}$ allows the disintegration formula

$$\Phi_{g,p}(A) = C(g,p)\int_0^\infty rl_q([\frac{1}{r}A] \cap C_p)g(r^p)dr, A \in \mathscr{B}^2.$$

According to the geometric definition of a probability distribution where arc-length is measured by l_q instead of l, we introduce the uniform distribution $G_p(M) = l_q(M)/l_q(C_p)$ on $\mathscr{B}^2 \cap C_p$. Making use of this distribution, the $l_{2,p}$-circle intersection-percentage function (ipf) of the set $A \in \mathscr{B}^2$ is defined by $\mathscr{F}_p(A,r) = G_p([\frac{1}{r}A] \cap C_p)$, $r > 0$. It was originally for $p = 2$ introduces in [16] and allows to reformulate $\Phi_{g,p}$ as $\Phi_{g,p}(A) = C_2(g,p)l_q(C_p)\int_0^\infty r\mathscr{F}_p(A,r)g(r^p)dr, A \in B^2$. Here, $C_2(g,p)l_q(C_p) = 1/I_{2,g,p}$, hence from the geometric measure representation in [18] follows that

$$\Phi_{g,p}(A) = \frac{1}{I_{2,g,p}}\int_0^\infty r\mathscr{F}_p(A,r)g(r^p)dr, A \in \mathscr{B}^2. \tag{1}$$

The respective multivariate extension of this formula is then given with $I_{n,g,p} = \int_0^\infty r^{n-1}g(r^p)dr$ by

$$\Phi_{g,p}^{(n)}(A) = \frac{1}{I_{n,g,p}}\int_0^\infty r^{n-1}\mathscr{F}_p([\frac{1}{r}A] \cap S_p)g(r^p)dr, A \in \mathscr{B}^n. \tag{2}$$

APPLICATIONS

The geometric measure representation (2) can be used to evaluate the cdf of rather arbitrary statistics $T(X)$ where $X = (X_1, ..., X_n)^T \sim \Phi_{g,p}^{(n)}$ and $T : R^n \to R$. In this case, $P(T(X) < t) = P(X \in T^{-1}(-\infty,t)) = P^X(A(t))$ where P^X denotes the distribution on $\mathscr{B}^n$ which is induced by X and the set

$$A(t) = \{x \in R^n : T(x) < t\}, t \in R$$

is generated by the statistic T, i.e.,

$$P(T(X) < t) = \Phi_{g,p}^{(n)}(A(t)), t \in R.$$

Let us consider first again the two-dimensional case $n = 2$. The statistics $L(X) = a_1X_1 + a_2X_2$, $P(X) = X_1 \cdot X_2$ and $R(X) = X_1/X_2$ are considered in [11] for $p \in \{1,2\}$ and if $g(r) = \exp\{-r^p/p\}$, i.e. if the sample distribution is a Gaussian or Laplace one. Exact integral representations of the cdfs are derived immediately from formula (1). The key point for doing this is always to determine the respective ipf. Elementary geometric considerations indicate the cases of the p-radius r which one has to distinguish for the derivation of the ipf. The different areas of the p-radius for defining the ipf are then according to (1) reflected in the geometric integral representation formula of the cdf of a statistic. These representations as well as integral representations of the corresponding pdfs which follow directly by taking derivatives are numerically rather stable, especially in comparison with certain concurring other results. The case of arbitrary $p > 0$ has been dealt with in [13] where the statistics $M(X) = \max\{X_1, X_2\}$ as well as $P(X)$ and $R(X)$ are considered. Again, explicit integral representations of the cdfs and pdfs are derived from formula (1). Moreover, results of a simulation study are presented which indicate that rather large sample sizes are necessary to distinguish for real data sets between different values of p.

Let us recall that according to [6] there exist several representations of skewed normal and skewed elliptical distributions. In this respect, the authors in [8] open a new unified geometric perspective for dealing with skewed two-dimensional elliptically contoured distributions. According to their basic result, different representations of skewed elliptically contoured distributions can be compared with each other from one and the same geometric point of view.

A suitable non-Euclidean geometry has been detected in [20] for deriving a geometric measure representation of the heteroscedastic axes-aligned Gaussian law. As a consequence from this, in [5] a geometric parametrization of the regular Gaussian law is derived, the new parameters are estimated and their distributions are derived.

Let us shortly discuss now to the case of arbitrary finite dimension $n \geq 2$. In [17] and [18], p-or (p,g)-generalizations of the classical statistical distributions as χ^2-, Student- and Fisher-distribution are derived. More over, numerous applications of formula (2) are surveyed in [18] for the Gaussian and elliptically contoured case $p = 2$. As just to mention some of them we refer to remainder term estimation in the multivariate central limit theorem, large deviation asymptotic, probabilities of correct classification and exact modified Student test in non-linear regression. More over, the results of [19] are extended in [20] to the multivariate case. The paper [3] introduces the class of skewed $l_{n,p}$-symmetric distributions and opens a great variety of possibilities of future theoretical and applied work.

REFERENCES

1. K. Andersen, *Cavalieri's Method of Indivisibles,* Archive for History of Exact Sciences **31**, 1 (1984).
2. T. W. Anderson, *An Introduction to Multivariate Analysis*, Wiley, New York, (1984).
3. R. B. Arellano-Valle and W.-D. Richter, *On skewed continuous $l_{n,p}$-symmetric distributions,* Chilean Journal of Statistics, Accepted for print (2012).
4. M. Bethge and F. Sinz, *The conjoint effect of divisive normalization and orientation selectivity on redundancy reduction*, Advances in Neural Information Processing Systems 21: Proceedings of the 2008 Conference, 1521-1528 (2009).
5. T. Dietrich, S. Kalke and W.-D. Richter, *Stochastic representations and a geometric parametrization of the two-dimensional Gaussian law*, Submitted (2012).
6. M. G. Genton (ed.) , *Skew elliptical distributions and their applications. A journey beyound normality.* Chapman and Hall, Boca Raton (2004).
7. I.R. Goodman and S. Kotz *Multivariate $\theta-$generalized normal distributions,* J. Multivariate Anal. **3**, 204-219 (1973) .
8. T. Guenzel, W.-D. Richter, S. Scheutzow, K. Schicker and J. Venz, *Geometric approach to the skewed normal distribution,* Accepted for print (2012).
9. A. Gupta and D. Song, *L_p-norm spherical distributions.* J. Statist. Planning and Inference, **60**, 241-260 (1997).
10. A. Hyvaerinen and U. Koester, *Complex cell pooling and the statistics of natural images*, Network: Computation in Neural Systems, **18**, 2, 81-100 (2007).
11. S. Kalke, W.-D. Richter and F. Thauer, *Linear combinations, products and ratios of simplicial or spherical variates*, Communications in Statistics - Theory and Methods, Accepted for print (2012).
12. M. Moszyńska and W.-D. Richter, *Reverse triangle inequality. Antinorms and semi-antinorms,* Studia Scientarum Mathematicarum Hungarica **49**, (1), 120-138 (2012).
13. K. Mueller and W.-D. Richter, *Exact maximum, product and ratio distributions under non-standard assumptions,* Submitted (2012).
14. W. A. Nikiforowski and L. F. Freimann, *Wegbereiter der neuen Mathematik,* Verlag MIR, Moskau und VEB Fachbuchverlag Leipzig (1978).
15. S. T. Rachev and L. Rueschendorf, *Approximate independence of distributions on spheres and their stability properties.* Ann.Probab. **19**,3, 1311-1337 (1991).
16. W.-D.Richter, *Laplace-Gauss integrals, Gaussian measure asymptotic behaviour and probabilities of moderate deviations*, Z. Anal. Anw. **4** (3), 257-267 (1985).
17. W.-D.Richter, *Generalized spherical and simplicial coordinates*, J. Math. Anal. Appl. **336**, 1187-1202 (2007).
18. W.-D. Richter, *Continuous $l_{n,p}$-symmetric distributions.* Lithuanian Math. J., **49**, 1, 93-108 (2009).
19. W.-D.Richter, *Ellipses numbers and geometric measure representations*, Journal of Applied Analysis **17**, 165-179 (2011).
20. W.-D.Richter, *Geometric and stochastic representations for elliptically contoured distributions,* Communications in Statistics: Theory and Methods, Accepted for print (2012).
21. G. Schechtman and J. Zinn, *On the volume of the intersection of two $l_{n,p}$-balls,* Proc. Amer. Math. Soc. **110**, 1, 217-224 (1990).
22. D. Song and A. K. Gupta , *L_p-norm uniform distributions,* Proc. Amer. Math. Soc. **125**, no.2, 595-601 (1997).
23. P. J. Szablowski, *Uniform distributions on spheres in finite dimensional L_α and their generalizations* J. Multivariate Anal. **64**, 103-117 (1998).
24. More references, *can be found under* `http://www.math.uni-rostock.de/~richter/Publications.html`.