Journal of Multivariate Analysis 132 (2014) 82–93

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A non-Gaussian multivariate distribution with all lower-dimensional Gaussians and related families

Subhajit Dutta*, Marc G. Genton

CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 6 April 2014
Available online 28 July 2014

AMS 2000 subject classifications:
primary 62J05
secondary 62P30

Keywords:
Distribution of products of random variables
Half-normal distribution
Sign function
Skewed distributions
Symmetric distributions

ABSTRACT

Several fascinating examples of non-Gaussian bivariate distributions which have marginal distribution functions to be Gaussian have been proposed in the literature. These examples often clarify several properties associated with the normal distribution. In this paper, we generalize this result in the sense that we construct a p-dimensional distribution for which any proper subset of its components has the Gaussian distribution. However, the joint p-dimensional distribution is inconsistent with the distribution of these subsets because it is not Gaussian. We study the probabilistic properties of this non-Gaussian multivariate distribution in detail. Interestingly, several popular tests of multivariate normality fail to identify this p-dimensional distribution as non-Gaussian. We further extend our construction to a class of elliptically contoured distributions as well as skewed distributions arising from selections, for instance the multivariate skew-normal distribution.

1. Introduction

In his fundamental paper, Bhattacharyya [8] stated the following assertions related to the bivariate Gaussian distribution: (i) normality and homoscedasticity of the component distributions, (ii) linearity of the regression lines, (iii) concentric spherical/elliptic contours, and (iv) normality of the marginal distributions. He was interested in framing a set of sufficient conditions that would be considered essential to determine the bivariate normal law. In the paper, he continued to give answers to these assertions, some of which he remarked "*appear to be interesting, others trivial*". In the statistics literature, Bhattacharyya's assertions have raised more questions, and they have been extended to various scenarios that are not necessarily Gaussian. Gelman and Meng [15] present results developed on Gaussian marginals with non-Gaussian conditional distributions, and the review paper by Arnold, Castillo and Sarabia [5] states conditionally skewed distributions with Gaussian marginals.

The assertion in statement (iv) on the normality of marginal distributions intrigues us, and we study it in more detail in this paper. It is well known that a pair of marginal distributions does not uniquely determine a bivariate distribution; for example, a bivariate distribution with Gaussian marginals need not necessarily be jointly Gaussian; see, e.g., [27]; [1, p. 37]; [14, p. 69]; and [10]. Kowalski's [19] paper with the references therein and sub-section 10.1 in [25] provide a collection of several such examples as well. These examples have interesting constructions, invoke questions regarding fundamental ideas of normality and correlation theory, and aid in improving the understanding of basic statistical theory. On a different note, there is an immense literature on the characterization of the normal distribution, but our interest is contrary to this fact, i.e., conditions which fail to determine the Gaussian law.

* Corresponding author.
E-mail addresses: subhajit.dutta@kaust.edu.sa, tijahbus@gmail.com (S. Dutta), marc.genton@kaust.edu.sa (M.G. Genton).

http://dx.doi.org/10.1016/j.jmva.2014.07.007

We first suggest a simple construction by considering $X_1^* = X_1 S_2$ and $X_2^* = X_2 S_1$, where X_1 and X_2 are independent and identically distributed (i.i.d.) standard normal variates; $S_1 = \mathrm{sign}(X_1)$ and $S_2 = \mathrm{sign}(X_2)$. Here, we define the univariate sign function as follows:

$$\mathrm{sign}(u) = \begin{cases} +1 & \text{if } u > 0, \\ -1 & \text{if } u < 0, \\ 0 & \text{if } u = 0. \end{cases}$$

The sign function is a key ingredient in our construction. It is quite easy to check that X_1^* and X_2^* are normally distributed. Note that $\mathrm{E}(e^{tX_1^*}) = \mathrm{E}(e^{tX_1 S_2}) = \{\mathrm{E}(e^{tX_1}) + \mathrm{E}(e^{-tX_1})\}/2 = e^{t^2/2}$ for any $t \in \mathbb{R}$. This follows from the facts that $X_1 \perp\!\!\!\perp X_2$, and by virtue of the symmetry $\mathrm{sign}(X_1) = \pm 1$ with an equal probability half. Here and henceforth, $X \perp\!\!\!\perp Y$ means that X and Y are independent random variables. However, (X_1^*, X_2^*) is non-normal because $X_1^* X_2^* = X_1 S_2 X_2 S_1 = (X_1 S_1)(X_2 S_2)$ is positive with probability 1. This is so because the function $u\ \mathrm{sign}(u)$ is positive for any u. Moreover, the support of this distribution is on the quadrants $(+,+)$ and $(-,-)$ only (see the left panel of Fig. 1 in Section 4.1). It is interesting to note that a similar idea was also discussed by Anderson [1, p. 37]. More specifically, he considered a bivariate normal distribution and proposed to shift probability mass from quadrant $(+,-)$ to quadrant $(+,+)$, and the same from quadrant $(-,+)$ to quadrant $(-,-)$. This is to be done in a way such that the marginal normality is preserved, but the resulting bivariate distribution with support only on the quadrants $(+,+)$ and $(-,-)$ is clearly non-normal. However, he did not give any specific construction in this discussion. In his book on counter-examples, Stoyanov [25, pp. 91–92] gives two examples of multivariate distributions, whose subsets have Gaussian distributions. The first example is expressed in terms of a density function, and has no stochastic representation. We will discuss the connection with the second example in Section 2, after we state our construction for the p-dimensional case with $p \geq 2$.

In this paper, our goal is to first extend the construction stated above, and hence build a result for normal distributions in the p-dimensional case. It turns out that the joint distribution of the p-dimensional variable has quite interesting properties, and we study them in detail. In the following sections, we study the symmetry of the joint distribution (Sections 3 and 4), its moments (Section 5) and other properties related to its probability distribution function (Section 6). A direction of extending normal distributions to other symmetric cases is by considering spherically and elliptically contoured distributions [13] and asymmetric distributions via selections [2]. In particular, we note that the multivariate skew-normal distribution [7] and other multivariate skewed distributions are special cases of these selections (Section 7). In Section 8, we generalize the above-mentioned result to these popular multivariate distributions. We conclude this study in Section 9 by posing some unresolved questions. The proofs of the theorems are given in the Appendix.

2. Gaussian lower-dimensional distributions

Let $X_1, \ldots, X_p$ be i.i.d. standard normal random variates. Define

$$X_k^* = X_k S_{k+1} \quad \text{for } 1 \leq k \leq p-1, \quad \text{and} \quad X_p^* = X_p S_1,$$

where $S_k = \mathrm{sign}(X_k)$ for $1 \leq k \leq p$. Define $\{k_1, \ldots, k_{p-1}\}$ to be a subset of size $p-1$ from $\{1, \ldots, p\}$, and $\mathbf{X}_{p-1,p}^* = (X_{k_1}^*, \ldots, X_{k_{p-1}}^*)^\top$. This notation treats $\mathbf{X}_{p-1,p}^*$ as a $(p-1)$-dimensional sub-vector of $\mathbf{X}_p^*$, and distinguishes it from the $(p-1)$-dimensional vector $\mathbf{X}_{p-1}^*$. We now state the main result of this study.

Theorem 1. *The vector $\mathbf{X}_p^* = (X_1^*, \ldots, X_p^*)^\top \sim F_p^*$ is such that for any subset $\{k_1, \ldots, k_{p-1}\}$ of $\{1, \ldots, p\}$,*

$$\mathbf{X}_{p-1,p}^* = (X_{k_1}^*, \ldots, X_{k_{p-1}}^*)^\top \sim \mathcal{N}_{p-1}(\mathbf{0}_{p-1}, \mathbf{I}_{p-1}) \quad \textit{but } \mathbf{X}_p^* \sim F_p^* \neq \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p),$$

where $\mathcal{N}_q$ denotes the joint q-dimensional standard normal distribution, $\mathbf{0}_q$ is the q-dimensional vector of zeros and $\mathbf{I}_q$ is the $q \times q$ identity matrix for any positive integer q.

Proof of Theorem 1. Consider the case when $\{k_1, \ldots, k_{p-1}\} = \{1, \ldots, p-1\}$, i.e., define $\mathbf{X}_{p-1,p}^* = (X_1^*, \ldots, X_{p-1}^*)^\top$. The corresponding $(p-1)$-dimensional vector is $(X_1, \ldots, X_{p-1})^\top$ with the vector of signs as $(S_2, \ldots, S_p)^\top$.

For any $\mathbf{t} \in \mathbb{R}^{p-1}$, consider the moment generating function of $\mathbf{X}_{p-1,p}^*$ as follows:

$$\begin{aligned} \mathrm{E}(e^{\mathbf{t}^\top \mathbf{X}_{p-1,p}^*}) &= \mathrm{E}\left(\prod_{k=1}^{p-1} e^{t_k X_k S_{k+1}}\right) \\ &= \mathrm{E}\left\{\prod_{k=2}^{p-1} e^{t_k X_k S_{k+1}} \mathrm{E}\left(e^{t_1 X_1 S_2}\right) \mid (X_2, \ldots, X_p)\right\} \\ &= \mathrm{E}\left\{\prod_{k=2}^{p-1} e^{t_k X_k S_{k+1}}\right\} e^{t_1^2 S_2^2/2} \quad [\because \mathrm{E}(e^{tZ}) = e^{t^2/2} \text{ for } Z \sim \mathcal{N}(0,1)] \end{aligned}$$

$$
\begin{aligned}
&= e^{t_1^2/2}\mathrm{E}\left\{\prod_{k=2}^{p-1} e^{t_k X_k S_{k+1}}\right\} \quad [\because \operatorname{sign}^2(u) = 1] \\
&= \mathrm{E}\left\{\prod_{k=3}^{p-1} e^{t_k X_k S_{k+1}} \mathrm{E}\left(e^{t_2 X_2 S_3}\right) \mid (X_3, \ldots, X_p)\right\} \\
&= e^{t_1^2/2 + t_2^2/2}\mathrm{E}\left\{\prod_{k=3}^{p-1} e^{t_k X_k S_{k+1}}\right\} \\
&\ \ \vdots \\
&= e^{t_1^2/2+\cdots+t_{p-2}^2/2}\ \mathrm{E}(e^{t_{p-1}X_{p-1}S_p}) \\
&= e^{t_1^2/2+\cdots+t_{p-2}^2/2}\ \left\{\frac{1}{2}\mathrm{E}(e^{-t_{p-1}X_{p-1}}) + \frac{1}{2}\mathrm{E}(e^{t_{p-1}X_{p-1}})\right\} \quad [\because S_p = \pm 1 \text{ w.p. } 1/2] \\
&= e^{t_1^2/2+\cdots+t_{p-1}^2/2} = e^{\mathbf{t}^\top \mathbf{t}/2}.
\end{aligned}
$$

This now implies that $\mathbf{X}^*_{p-1,p} \sim \mathcal{N}_{p-1}(\mathbf{0}_{p-1}, \mathbf{I}_{p-1})$. As per our construction, the probability of the joint product of the p random variables is $\Pr\left(X_1^* \cdots X_p^* = X_1 S_1 \cdots X_p S_p > 0\right) = 1$. Note that the support of F_p^* is $\{(u_1, \ldots, u_p) : u_1 \cdots u_p > 0\}$. Here, and henceforth $x \cdot y$ denotes $x \times y$. This is a proper subset of $\mathbb{R}^p$, and hence the distribution F_p^* is clearly non-Gaussian. This completes a part of the proof and gives the main idea underlying it. The proof is completed by considering other permutations, and this is presented in the Appendix. □

From the statement of Theorem 1, it is clear that the distribution of any subset of random variables of size less than $p-1$ is also Gaussian because the joint $(p-1)$-dimensional distribution is Gaussian with mean $\mathbf{0}_{p-1}$ and variance $\mathbf{I}_{p-1}$ (i.e., independent components). In particular, the marginal distributions are all Gaussians as well. Given $\mathbf{X}_p^*$, we now have any set of $p-1$ random variables $\mathbf{X}^*_{p-1,p}$ that are independent but jointly (or, p-wise) dependent (see [17] for pairwise independent random variables which are not jointly independent). It is important to note at this point that the argument of Theorem 1 is exclusive only to Gaussian random variables. This is the case because of a special property of the multivariate standard normal distribution, namely that a spherical distribution with the scale matrix as the identity matrix and independent components is necessarily Gaussian (see Theorem 4.11 in [13]).

Considering the indices of the random vector as $(1, \ldots, p)$, we chose the vector of signs to be $(2, \ldots, p, 1)$. But, the construction is quite flexible. We may further consider permutations of the sign vector for which we have a 'p-cycle' (see, e.g., [9]). One easy way to obtain these permutations is by shifting each index to the left (e.g., $(3, 4, \ldots, p, 1, 2)$, $(4, 5, \ldots, p, 1, 2, 3)$, etc.). In fact, we have a total of $p!/(p \cdot 1) = (p-1)!$ such choices for the construction of the p-dimensional sign vector. The idea of a 'p-cycle' avoids a complete cycle among any subset of size $p-1$ or smaller, which would not yield the desired construction (also see Section 7 for more details).

We conclude this section with a discussion of Example (ii) in [25, p. 92]. The author describes a stochastic representation for the joint p-dimensional distribution as follows (we put it in a summarized form, and use our notations for the sake of consistency): "*$X_j = \xi_j |Z_j|$ where $Z_1, \ldots, Z_p$ are i.i.d. standard normals and $(\xi_1, \ldots, \xi_p)$ are such that they are $(p-1)$-wise independent but* $\Pr(\xi_1 = 1, \ldots, \xi_p = 1) = P \neq 1/2^p$ *and not independent*". However, he does not suggest any specific way of constructing such probability distributions. Here, we give a specific choice of $(\xi_1, \ldots, \xi_p)$ in terms of $Z_1, \ldots, Z_p$. From our construction, it follows that ξ_k is $S_k S_{k+1}$ for $k = 1, \ldots, p-1$ and ξ_p as $S_p S_1$. Therefore, the probability P is $1/2^{p-1}$, which is clearly different from the independent case (when $P = 1/2^p$). Moreover, our construction comes from a different motivation, it has a simpler structure, and continues to attract independent interest as we demonstrate in the sections that follow.

3. Symmetry of $\mathbf{X}_p^*$ in p dimensions

Recall that the distribution of each of $X_1, \ldots, X_p$ is standard normal, and they are all independent. We therefore have $X_k \overset{d}{=} -X_k$ for any $1 \le k \le p$, which in turn implies that $\mathbf{X}_p \overset{d}{=} -\mathbf{X}_p$ for any p, where $\mathbf{X}_p = (X_1, \ldots, X_p)^\top$. This componentwise symmetry and independence now implies an important fact on mutual independence as follows:

$$|\mathbf{X}_p| = (|X_1|, \ldots, |X_p|)^\top \perp\!\!\!\perp \mathbf{S}_p = (S_1, \ldots, S_p)^\top. \tag{1}$$

This property has also been studied by Arellano-Valle, del Pino and San Martin [3, p. 113] for the more general class of componentwise symmetric distributions.

We have $X_k = |X_k| S_k$, and the componentwise symmetry of X_k also implies that $S_k \overset{d}{=} -S_k$ for $1 \le k \le p$, which in turn yields the following:

$$-X_k^* \overset{d}{=} -|X_k| S_k S_{k+1} = |X_k|(-S_k) S_{k+1} = |-X_k| S_k S_{k+1} = X_k^*,$$

for any $1 \le k \le p-1$, and by a similar argument we also get $-X_p^* \overset{d}{=} X_p^*$. Hence, the relation in (1) continues to hold for $\mathbf{X}_p^*$ and $\mathbf{S}_p^*$ (the sign vector of $\mathbf{X}_p^*$) as well.

Combining the above mentioned facts, we now have the following stochastic representation for the vector $\mathbf{X}_p^*$:

$$\begin{aligned}\mathbf{X}_p^* &= (X_1S_2, X_2S_3, \ldots, X_{p-1}S_p, X_pS_1)^\top \\ &= (|X_1|S_1S_2, |X_2|S_2S_3, \ldots, |X_{p-1}|S_{p-1}S_p, |X_p|S_pS_1)^\top. \end{aligned} \tag{2}$$

This representation is very useful and plays an important role in our study on the symmetry of the joint p-dimensional distribution.

Theorem 2. *For even p, the joint distribution is centrally symmetric, i.e.,* $\mathbf{X}_p^* \stackrel{d}{=} -\mathbf{X}_p^*$.

It is interesting to observe that although we have componentwise symmetry, we do not have central symmetry of $\mathbf{X}_p^*$ for odd values of p. We argue for this asymmetry in the proof of Theorem 2, which follows from the 'circular' definition of the sign vector, $\mathbf{S}_p$.

4. The joint distribution of $\mathbf{X}_p^*$

We now give some expressions for the distribution function, and as a consequence, we compute the density function of $\mathbf{X}_p^*$ (denoted by f_p^*). From the stochastic representation (2) and the independence Eq. (1), we derive the following relation for any $\mathbf{x} = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$:

$$\begin{aligned}F_p^*(\mathbf{x}) &= \Pr(X_1^* \le x_1, \ldots, X_p^* \le x_p) \\ &= \Pr(|X_1|S_1S_2 \le x_1, |X_2|S_2S_3 \le x_2, \ldots, |X_p|S_pS_1 \le x_p) \\ &= \frac{1}{2^p} \sum_{(s_1,\ldots,s_p)\in\{-1,+1\}^p} \Pr(|X_1|s_1s_2 \le x_1, |X_2|s_2s_3 \le x_2, \ldots, |X_p|s_ps_1 \le x_p) \\ &= \frac{1}{2^{p-1}} \sum_{\{(m_1,\ldots,m_p)\in\{-1,+1\}^p:\, m_1\cdots m_p>0\}} \Pr(m_1|X_1| \le x_1, m_2|X_2| \le x_2, \ldots, m_p|X_p| \le x_p).\end{aligned}$$

This follows from the fact that $\Pr(S_1S_2 = 1, \ldots, S_pS_1 = 1) = \Pr(S_1 = 1, \ldots, S_p = 1) + \Pr(S_1 = -1, \ldots, S_p = -1) = 1/2^p + 1/2^p = 1/2^{p-1}$. The other 2^p probabilities can all be worked out to be equal to $1/2^{p-1}$ in a similar way. Consider a partition of $\mathbb{R}^p$ as $\mathbb{R}_+^p \cup \mathbb{R}_-^p$, where $\mathbb{R}_+^p = \{(u_1, \ldots, u_p) : u_1 \cdots u_p > 0\}$ and $\mathbb{R}_-^p = \{(u_1, \ldots, u_p) : u_1 \cdots u_p < 0\}$.

Theorem 3. *The density function* f_p^* *for any* $p > 1$ *can be obtained as follows:*

$$f_p^*(\mathbf{x}) = \begin{cases} 2\phi(x_1)\cdots\phi(x_p) & \text{for } \mathbf{x} \in \mathbb{R}_+^p, \\ 0 & \text{for } \mathbf{x} \in \mathbb{R}_-^p. \end{cases}$$

Here, $\phi(x)$ *denotes the density function of the univariate standard normal distribution.*

As per our intuition and in view of the symmetry, f_2^* (the density function of $\mathbf{X}_2^*$) has support on the $(+, +)$ and $(-, -)$ quadrants of $\mathbb{R}^2$. One may also recall Anderson's [1, p. 37] construction in this context. However, the support of the distribution on $\mathbb{R}_+^p$ is much more complicated for $p > 2$. When $p = 3$, the support is on the four orthants $(+, +, +)$, $(+, -, -)$, $(-, +, -)$ and $(-, -, +)$ in $\mathbb{R}^3$.

Remark. If X_k are independent and symmetric about 0 with density function f_k for $k = 1, \ldots, p$, then $f_p^*(\mathbf{x}) = 2f_1(x_1)\cdots f_p(x_p)$ for $\mathbf{x} \in \mathbb{R}_+^p$. This also includes the special case when $\mathbf{X}_p = (X_1, \ldots, X_p)^\top$ is spherically symmetric.

4.1. The joint distribution F_2^* and the related copula

We state the functional form of the distribution function of $\mathbf{X}_2^*$ as follows (see proof of Theorem 2 for Case II in the Appendix):

$$F_2^*(\mathbf{x}) = \begin{cases} \{\Phi_+(x_1)\Phi_+(x_2) + 1\}/2, & \text{if } \mathbf{x} \in \mathbb{R}_+^2, \\ \Phi_-(x_2)/2, & \text{if } x_1 > 0,\ x_2 < 0, \\ \Phi_-(x_1)/2, & \text{if } x_1 < 0,\ x_2 > 0, \\ \Phi_-(x_1)\Phi_-(x_2)/2, & \text{if } \mathbf{x} \in \mathbb{R}_-^2. \end{cases}$$

Here $\Phi_+(x) = \Pr(|X| \le x) = 2\Phi(x) - 1$ for $x > 0$, and $\Phi_-(x) = \Pr(-|X| \le x) = 2 - 2\Phi(-x) = 2\Phi(x)$ for $x < 0$. Here, the notation $\Phi(x)$ denotes the cumulative univariate standard normal distribution function.

The distribution function, F_2^* is a function of the univariate distribution functions $\Phi(x_1)$ and $\Phi(x_2)$ only. So, the copula (see, e.g., [24]) of the two-dimensional random vector, $\mathbf{X}_2^*$, works out fairly easily to be:

$$C_2^*(u_1, u_2) = \begin{cases} (2u_1 - 1)(2u_2 - 1)/2 + 1/2, & \text{if } u_1 > 1/2,\ u_2 > 1/2, \\ u_2, & \text{if } u_1 > 1/2,\ u_2 < 1/2, \\ u_1, & \text{if } u_1 < 1/2,\ u_2 > 1/2, \\ 2u_1u_2, & \text{if } u_1 < 1/2,\ u_2 < 1/2. \end{cases}$$

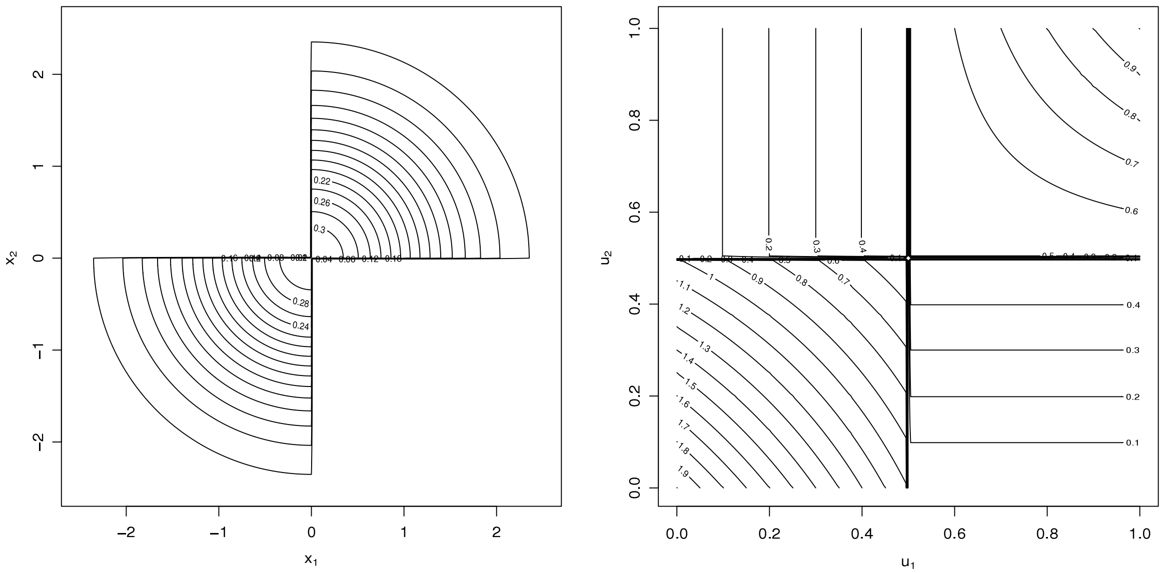

Fig. 1. For $p = 2$, we have the contour plot of the density function in the left panel, and that of the copula function in the right panel.

The copula is clearly symmetric due to the componentwise symmetry of the X_k^*'s. Moreover, $\mathbf{X}_2^*$ is centrally symmetric, which implies that the copula and its version based on the survival function are equal, i.e., the copula $C_2^*(u_1, u_2)$ is 'radially symmetric' about $(0, 0)$; see [24, p. 36]; [20]. For the general case with any p, we will have similar properties and the copula is 'radially symmetric' for any even value of p.

To get an idea of the joint distribution, we construct a contour plot using the form of the density function, f_2^*, as well as the contours for the bivariate copula function, C_2^* (see right panel of Fig. 1).

5. Moments of $\mathbf{X}_p^*$

At this point, recall the following stochastic representation of $\mathbf{X}_p^*$ stated in Eq. (2) for any $p \geq 2$:

$$\mathbf{X}_p^* = (|X_1|S_1S_2, |X_2|S_2S_3, \ldots, |X_{p-1}|S_{p-1}S_p, |X_p|S_pS_1)^\top.$$

By componentwise symmetry, we have $\mathrm{E}(X_k^*) = 0$, which further implies that $\mathrm{E}(\mathbf{X}_p^*) = \mathbf{0}_p$. Further, note that the variance satisfies $\mathrm{V}(X_k^*) = \mathrm{V}(|X_k|S_k) = \mathrm{V}(|X_k|)\mathrm{V}(S_k) = 1$ as $|X_k| \perp\!\!\!\perp S_k$ and $\mathrm{V}(S_k) = 1$ for all $1 \leq k \leq p$. For $p = 2$, $\mathrm{E}(X_1^*X_2^*) = \mathrm{E}(|X_1|\,|X_2|) = 2/\pi$. But, for $p > 2$, $\mathrm{E}(X_k^*X_{k'}^*) = \mathrm{E}(|X_k|S_kS_{k+1}|X_{k'}|S_{k'}S_{k'+1}) = 0$ for $1 \leq k \neq k' \leq p$ because $\mathrm{E}(S_k) = 0$ and independence condition (1) holds for $\mathbf{X}_p^*$. We therefore get

$$\mathrm{V}(\mathbf{X}_p^*) = \begin{cases} (1 - 2/\pi)\mathbf{I}_2 + (2/\pi)\mathbf{1}_2\mathbf{1}_2^\top, & \text{if } p = 2, \\ \mathbf{I}_p, & \text{if } p > 2. \end{cases}$$

Here, $\mathbf{1}_p$ denotes the p-dimensional column vector of 1's.

We now proceed to study the joint moments. Since the $(p - 1)$-dimensional or fewer distributions are all Gaussians, those moments match with the Gaussian distribution. But, for the joint p-dimensional case, we have quite different results. Firstly, note that

$$\mathrm{E}\left(\prod_{k=1}^p X_k^*\right) = \mathrm{E}\left(\prod_{k=1}^p |X_k|S_k^2\right) = \mathrm{E}\left(\prod_{k=1}^p |X_k|\right) = \prod_{k=1}^p \mathrm{E}(|X_k|) = (2/\pi)^{p/2},$$

as $S_k^2 = 1$ for any $1 \leq k \leq p$. So, the usual product moment matches with the standard normal distribution. We divide our study into the following three sub-cases:

1. If m_k is even for all $1 \leq k \leq p$, we have

$$\mathrm{E}\left(\prod_{k=1}^p X_k^{*m_k}\right) = \mathrm{E}\left(\prod_{k=1}^p |X_k|^{m_k}\right) = \prod_{k=1}^p \mathrm{E}(X_k^{m_k}) = \prod_{k=1}^p \frac{m_k!}{2^{m_k/2}(m_k/2)!},$$

as $S_k^{m_k} = 1$ for any $1 \leq k \leq p$. Therefore, all the even ordered moments also match with the standard p-variate normal distribution as well.

2. When m_k is odd for some of the k's (say, the set O) and even for the rest of the indices, we have
$$\mathrm{E}\left(\prod_{k=1}^{p} X_k^{*m_k}\right) = \mathrm{E}\left(\prod_{k\in\{1,\ldots,p\}\setminus O} |X_k|^{m_k}\right)\mathrm{E}\left(\prod_{k\in O} X_k^{*m_k}\right) = 0,$$
because of $(p-1)$-wise (which implies componentwise) independence of X_k^*s and $\mathrm{E}(X_k^*) = 0$ for the m_ks that belong to the set O.
3. If m_k is odd for all $1 \le k \le p$, we have
$$\begin{aligned}\mathrm{E}\left(\prod_{k=1}^{p} X_k^{*m_k}\right) &= \mathrm{E}\left(X_1 S_1^{m_p}\prod_{k=2}^{p} X_k^{m_k} S_k^{m_{k-1}}\right) = \mathrm{E}\left(\prod_{k=1}^{p} X_k^{m_k} S_k\right) = \mathrm{E}\left(\prod_{k=1}^{p} |X_k|^{m_k}\right)\\ &= \prod_{k=1}^{p}\mathrm{E}(|X_k|^{m_k}),\end{aligned}$$
since $S_k^{m_{k-1}} = S_k$ for any odd m_k, $2 \le k \le p$ and $S_1^{m_p} = S_1$ for any odd m_p.
The m-th order moment of the absolute value of a standard normal variate (say, Z) can be expressed as follows (see [12]):
$$\mathrm{E}(|Z|^m) = \begin{cases}(2r-1)\cdot(2r-3)\cdots 3\cdot 1, & \text{for } m = 2r,\\ \dfrac{2^r r!}{\sqrt{2\pi}}, & \text{for } m = 2r+1.\end{cases}$$
Using this fact, for odd m_k, we obtain the following expression:
$$\mathrm{E}\left(\prod_{k=1}^{p} X_k^{*m_k}\right) = \prod_{k=1}^{p}\frac{2^{(m_k-1)/2}\{(m_k-1)/2\}!}{\sqrt{2\pi}}.$$

Unlike the p-dimensional standard normal distribution, all the odd ordered moments of F_p^* are not zero because of the dependence among the components induced by the 'circular' definition of the sign vector, $\mathbf{S}_p$.

6. Other properties and inferences

We first study the conditional distributions. Define $\mathbf{X}_p^* = (\mathbf{X}_{m,p}^{*\top}, \mathbf{X}_{p-m,p}^{*\top})^\top$, where $\mathbf{X}_{m,p}^*$ is a $m \times 1$ random vector with the first m components of $\mathbf{X}_p^*$ for any $0 < m < p$. Now, consider the conditional density of the vector $\mathbf{X}_{m,p}^*$ given $\mathbf{X}_{p-m,p}^*$ (say, g_m^*) as follows:
$$g_m^*(\mathbf{x}_{m,p}|\mathbf{x}_{p-m,p}) = f_p^*(\mathbf{x})/f_m^*(\mathbf{x}_{p-m,p}) = 2\phi(x_1)\cdots\phi(x_p)/\{\phi(x_{m+1})\cdots\phi(x_p)\},$$
for $\mathbf{x} = (\mathbf{x}_{m,p}^\top, \mathbf{x}_{p-m,p}^\top)^\top \in \mathbb{R}_+^p$, $\mathbf{x}_{(1)}^\top$ is a $m \times 1$ vector, and $x_i \in \mathbb{R}$ for $m+1 \le i \le p$. Recall that $\mathbf{X}_{p-m,p}^* \sim \mathcal{N}_{p-m}(\mathbf{0}_{p-m}, \mathbf{I}_{p-m})$. Therefore,
$$g_m^*(\mathbf{x}_{m,p}|\mathbf{x}_{p-m,p}) = 2\phi(x_1)\cdots\phi(x_m), \quad \text{for } x_1\cdots x_m > 0.$$
It turns out that all the conditional distributions also belong to the same family as the p-dimensional distributions. Although all low-dimensional distributions are Gaussian, the joint as well as the conditional belong to the same family of non-Gaussian distributions.

The squared of signs is unity (i.e., $S_k^2 = 1$ for $k = 1, \ldots, p$) and $X_1, \ldots, X_p$ are i.i.d. standard normal. So, we have
$$\mathbf{X}_p^{*\top}\mathbf{X}_p^* = X_1^{*2} + \cdots + X_p^{*2} \stackrel{d}{=} X_1^2 + \cdots + X_p^2 \sim \chi_p^2,$$
although $\mathbf{X}_p^*$ is not normally distributed. Here χ_p^2 denotes the chi-squared distribution with p degrees of freedom (df). Interestingly, this result continues to hold for asymmetric distributions like the multivariate skew-normal distribution; see, e.g., [16,4] and references therein. In our case, the following more general result holds:
$$\mathbf{X}_p^{*\top}\Lambda_p\mathbf{X}_p^* = X_1^{*2}/\sigma_1^2 + \cdots + X_p^{*2}/\sigma_p^2 \stackrel{d}{=} X_1^2/\sigma_1^2 + \cdots + X_p^2/\sigma_p^2,$$
where $\Lambda_p = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. So, the quadratic form $\mathbf{X}_p^{*\top}\Lambda_p\mathbf{X}_p^*$ has the non-central chi-squared distribution; see [23, pp. 22–24].

At this point, we recall Mardia's [21] indices for testing normality of a sample based on the multivariate skewness parameter $\beta_{1,p} = \mathrm{E}\{(\mathbf{X}_p - \boldsymbol{\mu}_p)^\top\boldsymbol{\Sigma}_p{}^{-1}(\tilde{\mathbf{X}}_p - \boldsymbol{\mu}_p)\}^3$, where $\tilde{\mathbf{X}}_p \stackrel{d}{=} \mathbf{X}_p$, and they are independent. The multivariate kurtosis parameter is $\beta_{2,p} = \mathrm{E}\{(\mathbf{X}_p - \boldsymbol{\mu}_p)^\top\boldsymbol{\Sigma}_p{}^{-1}(\mathbf{X}_p - \boldsymbol{\mu}_p)\}^2$. For the p-dimensional standard normal distribution, we have $\beta_{1,p} = 0$, and $\beta_{2,p} = p(p+2)$, which follows from the fact that $\mathrm{E}(Y^2) = p(p+2)$ for $Y \sim \chi_p^2$. Here, we have $\boldsymbol{\mu}_p = \mathrm{E}(\mathbf{X}_p)$ and $\boldsymbol{\Sigma}_p = \mathrm{V}(\mathbf{X}_p)$.

For the random vector $\mathbf{X}_p^*$, we recall that $\boldsymbol{\mu}_p = \mathbf{0}_p$ and $\boldsymbol{\Sigma}_p = \mathbf{I}_p$ for $p > 2$. So, $\beta_{1,p} = \mathrm{E}(\mathbf{X}_p^{*\top}\tilde{\mathbf{X}}_p^*)^3 = \mathrm{E}(\sum_{k=1}^p X_k^*\tilde{X}_k^*)^3 = \mathrm{E}(\sum_{i_1,\ldots,i_p} C_{3,p}X_1^{*i_1}\tilde{X}_1^{*i_1}\cdots X_p^{*i_p}\tilde{X}_p^{*i_p})$, where $C_{3,p} = 3!/(i_1!\cdots i_p!)$. Now, $\beta_{1,p} = \sum_{i_1,\ldots,i_p} C_3\mathrm{E}(X_1^{*i_1})\mathrm{E}(\tilde{X}_1^{*i_1})\cdots\mathrm{E}(X_p^{*i_p})\mathrm{E}(\tilde{X}_p^{*i_p}) = 0$, because of independence and the fact that in each term, at least one of the i_k's is either 1 or 3 (also recall the discussion

Table 1
Power functions of different tests of multivariate normality for a sample of size 100 with 1000 replicates from F_p^* with varying values of p.

Dimension (p)	Mskew	Mkurt	SW	R	HZ	SR	mvS
$p = 2$	0.139	0.143	0.468	0.047	0.950	0.930	0.551
$p = 5$	0.040	0.043	0.640	0.072	0.156	0.085	0.048
$p = 10$	0.034	0.068	1	0.076	0.056	0.045	0.049
$p = 25$	0.010	0.451	1	0.090	0.034	0.055	0.050
$p = 50$	0	1	1	0.130	0	0.042	0.054
$p = 90$	0	1	1	0.170	0	0.047	0.051

on the moments of $\mathbf{X}_p^*$ in Section 5). Also, $\beta_{2,p} = \mathrm{E}(\mathbf{X}_p^{*\top}\mathbf{X}_p^*)^2 = p(p+2)$ for the random vector $\mathbf{X}_p^*$ as $\mathbf{X}_p^{*\top}\mathbf{X}_p^* \sim \chi_p^2$. It now follows that the values of $\beta_{1,p}$ and $\beta_{2,p}$ coincide with those of the p-dimensional standard normal distribution for any value of $p > 2$. Mardia's test therefore fails to identify this new distribution as being non-normal.

Projection pursuit techniques are often used to reject the hypothesis whether a given multivariate distribution is Gaussian or not. Theorem 3.6 in [11] states that every projection of the distribution of $\mathbf{X}_p^*$ is not Gaussian, almost surely. However, it is interesting to note that if we consider projections of $\mathbf{X}_p^*$ where at least one of the components is zero, then the distribution of this projection is Gaussian.

6.1. Results of tests of multivariate normality on F_p^*

To understand the distribution F_p^* in contrast with the standard p-dimensional Gaussian distribution, we carried out some tests of multivariate Gaussianity. The review paper by Mecklin and Mundfrom [22] presents a detailed discussion of several tests of multivariate normality. They divide the existing tests into four main categories: (i) procedures based on graphical plots and correlation coefficients, (ii) goodness-of-fit tests, (iii) tests based on measures of skewness and kurtosis, and (iv) consistent procedures based on the empirical characteristic function. Apart from Mardia's indices based on skewness (Mskew) and kurtosis (Mkurt) (category (iii)), we also used popular tests of multivariate normality like Shapiro–Wilks (SW), Royston (R) and a more recent proposal called mvShapiro (mvS) [28] (category (ii)). We also considered Henze–Zirkler (HZ) (category (iv)) and one based on energy statistics Szekely–Rizzo (SR) [26] (category (ii)). The results that we obtained are summarized in Table 1.

We expect the power to be close to 1 for all these tests and for any value of $p > 2$. However, the results in Table 1 differ from our expectations. From a theoretical standpoint, Mardia's skewness index fails to detect F_p^* as a non-normal distribution. Popular tests like HZ, SR and mvS strongly reject the claim of F_p^* to be Gaussian for $p = 2$, but the power decreases with increasing values of p. For the R test, strangely, the power grows with increasing p. Additionally, we see interesting and stronger results for Mkurt and SW. As expected, the power of Mkurt decreases with increasing p up to 10, but starts to increase and eventually shoots to 1 for p greater than 50. Surprisingly, SW performs differently from all other tests, and the power increases to 1 for $p = 10$ onwards. However, this result is actually 'false' because this test fails to detect Gaussianity in a data of size 100 generated from the p-dimensional standard normal distribution. We probably need to increase the sample size to obtain more accurate inferences in higher dimensions.

7. More permutations and the 'half-normal' distribution

A natural aspect to study is the joint p-dimensional distribution under permutations other than the 'p-cycles' mentioned in Section 2. For other permutations in which we have a 'cycle' of smaller size in the permutation, interestingly, the joint distribution can be worked out fairly easily with a stochastic representation similar to the one mentioned in Eq. (2). For any general permutation, whenever we have a 'p'-cycle' with $p' < p$, the components of the vector with this smaller 'p'-cycle' will have the joint distribution $F_{p'}^*$, while the remaining part, $p - p'$, will have the multivariate 'half-normal' distribution [18, p. 326]. It can be argued to be thus because the random variates in each of the $p - p'$ dimensions are multiplied by their respective sign. To get a clearer explanation, consider the case when $p = 3$ and the sign vector has indices (2, 1, 3), i.e., $p' = 2$. Then, we obtain the following:

$$\mathbf{X}_3^* = (X_1^*, X_2^*, X_3^*)^\top = (X_1S_2, X_2S_1, X_3S_3)^\top = (|X_1|S_1S_2, |X_2|S_2S_1, |X_3|)^\top.$$

The joint distribution will be supported on the orthants $(+, +, +)$ and $(-, -, +)$ of $\mathbb{R}^3$.

On the other hand, define $X_k^{**} = X_kS_k$ for $1 \le k \le p$. The sign vector $\mathbf{S}_p$, now has the natural indices $(1, \ldots, p)$, and we get $\mathbf{X}_p^{**} = (|X_1|, \ldots, |X_p|)^\top$. This is the p-dimensional 'half normal' distribution that has support on the orthant in $\mathbb{R}^p$ where all the components are positive. In the present context, we think that calling it the '$1/2^p$-normal distribution' will be more appropriate, while $\mathbf{X}_p^*$ with support on $\mathbb{R}_+^p$ (which includes half of $\mathbb{R}_p$) may be called the 'half-normal' distribution. To summarize, at one end, we have the distribution F_p^* (when we have a permutation with a complete 'p-cycle') and the p-dimensional 'half normal' distribution on the other (when we have the identity permutation). Considering permutations for the sign vector that take into account smaller cycles, we get a joint distribution that lies in a sense somewhere in-between these two distributions at the two 'extremities'.

8. Other distributions where all lower-dimensional distributions are the same

We now try to address Theorem 1 for distributions other than the standard normal distribution. First, define the new p-dimensional vector $\mathbf{Y}_p^*$ as follows:

$$\mathbf{Y}_p^* \stackrel{d}{=} \mathbf{a}_p + \mathbf{B}_p\mathbf{X}_p^*,$$

where $\mathbf{a}_p$ is a vector in $\mathbb{R}^p$, and $\mathbf{B}_p$ is a $p \times p$ symmetric, positive definite matrix for any $p \geq 2$. We may also consider $\mathbf{B}_p$ to be a random matrix of the simple form $Z\mathbf{I}_p$ with Z a positive random variable, independent of $\mathbf{X}_p^*$. Other general forms for the stochastic matrix $\mathbf{B}_p$ maybe $\mathrm{diag}(Z_1, \ldots, Z_p)$ with Z_k positive for $1 \leq k \leq p$.

Define $\mathbf{Y}_{p-1,p}^*$ based on $\mathbf{Y}_p^*$ by selecting a subset of size $p-1$ from the set $\{1, \ldots, p\}$. We say a distribution belongs to the same 'family of distributions' if that member can be obtained by a location and/or scale transformation of a standard family member. In other words, the class of distributions is closed under affine transformations.

Theorem 4. *The distributions of all the $(p-1)$-dimensional random vectors $\mathbf{Y}_{p-1,p}^*$ belong to the same family of distributions, but that of $\mathbf{Y}_p^*$ does not.*

We now state consequences of this result in several other popular and important classes of symmetric and asymmetric distributions derived from the standard normal distribution.

8.1. Elliptically contoured distributions

Define $\mathbf{a}_p = \boldsymbol{\mu}_p$ and $\mathbf{B}_p = \boldsymbol{\Sigma}_p^{1/2}$, where $\boldsymbol{\Sigma}_p^{1/2}$ is the unique symmetric square root of a symmetric and positive definite matrix, $\boldsymbol{\Sigma}_p$. The vector $\boldsymbol{\mu}_p$ is the p-dimensional location parameter, and $\boldsymbol{\Sigma}_p$ is the $p \times p$ scale matrix. Now, for any $(p-1)$-dimensional distribution, $\mathbf{Y}_{p-1,p}^* \sim \mathcal{N}_{p-1}(\boldsymbol{\mu}_{p-1}, \boldsymbol{\Sigma}_{p-1})$, although $\mathbf{Y}_p^*$ is not normally distributed with this correlation structure. Given that $\boldsymbol{\Sigma}_p$ is a covariance matrix, $\boldsymbol{\Sigma}_{p-1}$ is also a valid covariance matrix for any subset of size $p-1$ because it is symmetric by definition. The positive definiteness of $\boldsymbol{\Sigma}_{p-1}$ follows since $\boldsymbol{\Sigma}_{p-1}$ can be expressed as $\mathbf{D}\boldsymbol{\Sigma}_p\mathbf{D}^\top$, where $\mathbf{D}$ is a $(p-1) \times p$ matrix obtained by selecting rows of the identity matrix $\mathbf{I}_p$ with indices from the vector, $\mathbf{Y}_{p-1,p}^*$.

We further state constructions for other spherically contoured distributions:

1. (*Multivariate t-distribution*) Define $\mathbf{a}_p = \mathbf{0}_p$ and $\mathbf{B}_p = (Z_m/m)^{-1/2}\mathbf{I}_p$, where $Z_m \sim \chi_m^2$, independent of $\mathbf{X}_p^*$;
2. (*Scale mixture of multivariate normal distribution*) Define $\mathbf{a}_p = \mathbf{0}_p$ and $\mathbf{B}_p = w_p\mathbf{I}_p$, where w_p is a non-negative random variable, independent of $\mathbf{X}_p^*$.

Using affine transformations we can consider distributions that are elliptically symmetric. So, we now have a version of Theorem 4 for a class of elliptically symmetric distributions.

8.2. Selections

We first give a brief introduction of selection distributions. Consider $\mathbf{U} \in \mathbb{R}^q$ and $\mathbf{V} \in \mathbb{R}^p$ to be two random vectors, and denote by C a measurable subset of $\mathbb{R}^q$. Arellano-Valle et al. [2] define a selection distribution as the conditional distribution of $\mathbf{V}$ given $\mathbf{U} \in C$, denoted by $(\mathbf{V}|\mathbf{U} \in C)$. Now, $\mathbf{Y} \in \mathbb{R}^p$ has a selection distribution if $\mathbf{Y} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$.

Further, if $(\mathbf{U}, \mathbf{V})$ is jointly normal, i.e.,

$$\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \sim \mathcal{N}_{q+p}\left(\begin{bmatrix}\boldsymbol{\mu}_\mathbf{u}\\ \boldsymbol{\mu}_\mathbf{v}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_\mathbf{u} & \boldsymbol{\Sigma}_\mathbf{uv}\\ \boldsymbol{\Sigma}_\mathbf{uv}^\top & \boldsymbol{\Sigma}_\mathbf{v}\end{bmatrix}\right),$$

then $\mathbf{Y}$ is said to have a multivariate selection distribution based on normal distributions. Let $\mathbf{X}_0 \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{u})$ and $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{v} - \boldsymbol{\Sigma}_\mathbf{uv}^\top\boldsymbol{\Sigma}_\mathbf{u}^{-1}\boldsymbol{\Sigma}_\mathbf{uv})$ be independent random vectors. Define

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu}_\mathbf{v} + \boldsymbol{\Sigma}_\mathbf{uv}^\top\boldsymbol{\Sigma}_\mathbf{u}^{-1}\mathbf{X}_0[C - \boldsymbol{\mu}_\mathbf{u}] + \mathbf{X}, \tag{3}$$

where $\mathbf{X}_0[C - \boldsymbol{\mu}_\mathbf{u}] \stackrel{d}{=} (\mathbf{X}_0|\mathbf{X}_0 \in C - \boldsymbol{\mu}_\mathbf{u})$. Further, note that $(\mathbf{U}, \mathbf{V})^\top \stackrel{d}{=} (\boldsymbol{\mu}_\mathbf{u} + \mathbf{X}_0, \boldsymbol{\mu}_\mathbf{v} + \boldsymbol{\Sigma}_\mathbf{uv}^\top\boldsymbol{\Sigma}_\mathbf{u}^{-1}\mathbf{X} + \mathbf{X}_0)^\top$.

Now, recall the statement of Theorem 4 and the related discussion preceding it. Define $\mathbf{a}_p = \boldsymbol{\mu}_\mathbf{v} + \boldsymbol{\Sigma}_\mathbf{uv}^\top\boldsymbol{\Sigma}_\mathbf{u}^{-1}\mathbf{X}_0[C - \boldsymbol{\mu}_\mathbf{u}]$ and $\mathbf{B}_p = \mathbf{I}_p$. Note that $\mathbf{a}_p$ is a random vector and it has a Gaussian distribution, while $\mathbf{B}_p$ is a non-stochastic matrix. Moreover, selection distributions are closed under linear transformations. Using Eq. (3) with $\mathbf{X}_p^*$, we now have a version of Theorem 4 for selections.

We can use this construction, and results from the previous sub-section, to extend this result to scale mixtures of the selection normal distributions as well. Skew-symmetric distributions can be obtained by considering the dimension q to be 1, the selection subset C to be $\{u \in \mathbb{R} | u > \beta\}$ with $\beta \in \mathbb{R}$, given a symmetric random variable U, and a symmetric random vector, $\mathbf{V}$. We have $\Pr(U > 0) = 1/2$ due to its symmetry, and the function $\pi(\mathbf{x}) = \Pr(U > 0|\mathbf{V} = \mathbf{x})$ is called a skewing function, which satisfies $0 \leq \pi(\mathbf{x}) \leq 1$ and $\pi(-\mathbf{x}) = 1 - \pi(\mathbf{x})$. Using the choices of $\mathbf{a}_p$ and $\mathbf{B}_p$ mentioned above for this specific case therefore gives us Theorem 4 for multivariate skewed distributions [16,6].

9. Conclusions

In this work, we demonstrated some results related to the broad idea of characterization of the multivariate Gaussian distribution in terms of the distribution of its subsets. We also explored some issues related to inference, such as the impact of our construction on several tests of multivariate normality. However, some questions remain unanswered. The first question is how to characterize the class of p-dimensional distributions for which all $(p-1)$-dimensional distributions are Gaussian (also see the review paper by Arnold et al. [5]). We could also explore the link of the p-dimensional copula with the $(p-1)$-dimensional Gaussian copulas. Further, it would be interesting to study the distribution of new random variables obtained by multiplying $\mathbf{X}_p^*$ by its sign vector (as stated before Theorem 1). Another area of research could be to explore this construction for discrete symmetric distributions.

Acknowledgments

The authors would like to thank the Associate Editor for a careful reading, and suggestions to an earlier version of the paper. They additionally thank Prof. Reinaldo Arellano-Valle, Prof. Adelchi Azzalini, Prof. Juan Antonio Cuesta-Albertos and Dr. Raphaël Huser for useful directions, discussions and comments which led to an improvement of the paper. The first author is grateful to both Prof. Probal Chaudhuri for asking the question in Theorem 1, and Prof. Anil K. Ghosh who gave him the first ideas.

Appendix. Mathematical details

Proof of Theorem 1 (*Continued*)**.** Consider $\{j_1, \ldots, j_{p-1}\}$ to be a proper subset of size $p-1$ of the set $\{1, \ldots, p\}$ such that $j_1 < \cdots < j_{p-1}$. Since we draw $p-1$ ordered numbers out of p, there are three possibilities for the choice of the set $\{j_1, \ldots, j_{p-1}\}$. The set is either $\{1, \ldots, p-1\}$ or $\{2, \ldots, p\}$ or it is such that there exists a unique j_m such that $j_m - j_{m-1} = 2$ (there are $p-2$ such choices).

Define $\mathbf{X}_{p-1,p}^* = (X_{j_1}^*, \ldots, X_{j_{p-1}}^*)^\top$. We now look into the corresponding $(p-1)$-dimensional vector of signs $\mathbf{S}_{p-1}$, whose indices we denote as the set $\{i_1, \ldots, i_{p-1}\}$. For the first two possibilities, $\{i_1, \ldots, i_{p-1}\}$ is $\{2, \ldots, p\}$ and $\{3, \ldots, p, 1\}$, respectively. In the third, it is such that $i_k = j_k + 1$ if $j_k \neq p$ for $k = 1, \ldots, p-2$ and 1 if $i_{p-1} = p$. We have already argued when $\{j_1, \ldots, j_{p-1}\} = \{1, \ldots, p-1\}$, and call it **Case I**. The argument for the second possibility follows from Case I. We now give an argument for the third possibility, which appears to be a bit tricky.

Case II: The third possibility

$$
\begin{aligned}
\mathrm{E}(e^{\mathbf{t}^\top \mathbf{X}_{p-1,p}^*}) &= \mathrm{E}\left(\prod_{k=1}^{p-1} e^{t_k X_{j_k} S_{i_k}}\right) \\
&= \mathrm{E}\left\{\prod_{k=2,\neq m}^{p-1} e^{t_k X_{j_k} S_{i_k}} \mathrm{E}\left(e^{t_m X_{j_m} S_{i_m}}\right) \mid (X_{i_1}, \ldots, X_{i_{p-1}})\right\} \\
&= \mathrm{E}\left\{\prod_{k=2,\neq m}^{p-1} e^{t_k X_{j_k} S_{i_k}} e^{t_m^2 S_{i_m}^2/2}\right\} \quad [\because \mathrm{E}(e^{tZ}) = e^{t^2/2} \text{ for } Z \sim \mathcal{N}(0,1)] \\
&= e^{t_m^2/2} \mathrm{E}\left\{\prod_{k=2,\neq m}^{p-1} e^{t_k X_{j_k} S_{i_k}}\right\} \quad [\because \mathrm{sign}^2(u) = 1].
\end{aligned}
$$

Now, take a look at the sets $\{j_1, \ldots, j_{p-1}\} \setminus j_m$ and $\{i_1, \ldots, i_{p-1}\} \setminus i_m$. Without loss of generality, we call them $\{j_1, \ldots, j_{p-2}\}$ and $\{i_1, \ldots, i_{p-2}\}$ again. Now, either $\{j_1, \ldots, j_{p-2}\} = \{1, \ldots, p-2\}$ or there exists a unique j_n such that $j_n - j_{n-1} = 3$; in fact $n = m+1$. We therefore get

$$
\begin{aligned}
\mathrm{E}(e^{\mathbf{t}^\top \mathbf{X}_{p-1,p}^*}) &= e^{t_m^2/2} \mathrm{E}\left\{\prod_{k=1,\neq m+1}^{p-2} e^{t_k X_{j_k} S_{i_k}} \mathrm{E}(e^{t_n X_{j_{m+1}} S_{i_{m+1}}}) \mid (X_{i_1}, \ldots, X_{i_{p-2}})\right\} \\
&= e^{t_m^2/2 + t_{m+1}^2/2} \mathrm{E}\left\{\prod_{k=1,\neq m+1}^{p-2} e^{t_k X_{j_k} S_{i_k}}\right\} \\
&\ \ \vdots \\
&= e^{t_m^2/2 + t_{m+1}^2/2 + \cdots + t_{p-1}^2/2} \mathrm{E}\left\{\prod_{k=1}^{m-1} e^{t_k X_k S_{k+1}}\right\}.
\end{aligned}
$$

For Case I above for the set $\{1, \ldots, m-1\}$, we now have

$$\mathrm{E}(e^{\mathbf{t}^\top \mathbf{X}^*_{p-1,p}}) = e^{t_m^2/2+\cdots+t_{p-1}^2/2} \cdot e^{t_1^2/2+\cdots+t_{m-1}^2/2} = e^{\mathbf{t}^\top \mathbf{t}/2}.$$

This now implies that $\mathbf{X}^*_{p-1,p} \sim \mathcal{N}_{p-1}(\mathbf{0}_{p-1}, \mathbf{I}_{p-1})$ for any ordered subset of size $p-1$. The unordered case is a permutation of the ordered case and hence can be obtained by pre-multiplying $\mathbf{X}^*_{p-1,p}$ by a matrix obtained by permuting the rows of the identity matrix $\mathbf{I}_{p-1}$.

As per our construction, the probability of the distribution of the componentwise product of $\mathbf{X}^*_p$ is now

$$\Pr\left(X_1^* \cdots X_p^* = X_1S_1 \cdots X_pS_p > 0\right) = 1 \Rightarrow \mathbf{X}^*_p = (X_1^*, \ldots, X_p^*)^\top \sim F_p^*,$$

as u sign(u) is positive for any u. The support of F_p^* is $\mathbb{R}_p^+ = \{(u_1, \ldots, u_p) : u_1 \cdots u_p > 0\}$. This is clearly a proper subset of $\mathbb{R}^p$, and hence the distribution F_p^* is non-Gaussian. $\square$

Proof of Theorem 2. Firstly, recall that for general distributions, componentwise symmetry does not imply central symmetry. In particular, it holds when the components are independent. For odd p, we get

$$\begin{aligned}
-\mathbf{X}^*_p &= (-X_1S_2, -X_2S_3, -X_3S_4, \ldots, -X_{p-1}S_p, -X_pS_1)^\top \\
&\stackrel{d}{=} (-|X_1|S_1S_2, -|X_2|S_2S_3, -|X_3|S_3S_4, \ldots, -|X_{p-1}|S_{p-1}S_p, -|X_p|S_pS_1)^\top \\
&\stackrel{d}{=} (|X_1|(-S_1)S_2, |X_2|S_2(-S_3), |X_3|(-S_3)S_4, \ldots, |X_{p-1}|S_{p-1}(-S_p), |X_p|S_p(-S_1))^\top \\
&\stackrel{d}{=} (|-X_1|S_1S_2, |X_2|S_2S_3, |-X_3|S_3S_4, \ldots, |X_{p-1}|S_{p-1}S_p, |-X_p|(-S_p)S_1)^\top \\
&= (X_1^*, X_2^*, X_3^*, \ldots, X_{p-1}^*, -X_p^*)^\top \\
&\neq \mathbf{X}^*_p.
\end{aligned}$$

The main idea in the third step above is to always assign the negative sign to the sign with the odd index. The 'circular' definition of the sign vector $\mathbf{S}_p$ prevents the central symmetry.

For even p, we have the following:

$$\begin{aligned}
-\mathbf{X}^*_p &= (-X_1S_2, -X_2S_3, -X_3S_4, -X_4S_5, \ldots, -X_{p-1}S_p, -X_pS_1)^\top \\
&\stackrel{d}{=} (-|X_1|S_1S_2, -|X_2|S_2S_3, -|X_3|S_3S_4, -|X_4|S_4S_5, \ldots, -|X_{p-1}|S_{p-1}S_p, -|X_p|S_pS_1)^\top \\
&\stackrel{d}{=} (|X_1|(-S_1)S_2, |X_2|S_2(-S_3), |X_3|(-S_3)S_4, |X_4|S_4(-S_5), \ldots, |X_{p-1}|(-S_{p-1})S_p, |X_p|S_p(-S_1))^\top \\
&\stackrel{d}{=} (|-X_1|S_1S_2, |X_2|S_2S_3, |-X_3|S_3S_4, |X_4|S_4S_5, \ldots, |-X_{p-1}|S_{p-1}S_p, |X_p|S_pS_1)^\top \\
&\stackrel{d}{=} \mathbf{X}^*_p.
\end{aligned}$$

This completes the proof. $\square$

Proof of Theorem 3. Recall that a simplified expression for F_p^* now follows from (3) and the independence of the random variates as follows:

$$F_p^*(\mathbf{x}) = \frac{1}{2^{p-1}} \sum_{\{(m_1,\ldots,m_p)\in\{-1,+1\}^p : m_1\cdots m_p>0\}} \Pr(m_1|X_1| \le x_1, m_2|X_2| \le x_2, \ldots, m_p|X_p| \le x_p).$$

Consider the partition $\mathbb{R}^p = \mathbb{R}_+^p \cup \mathbb{R}_-^p$, where $\mathbb{R}_+^p = \{(u_1, \ldots, u_p) : u_1 \cdots u_p > 0\}$ and $\mathbb{R}_-^p = \{(u_1, \ldots, u_p) : u_1 \cdots u_p < 0\}$. Depending on whether p is odd or even, we break up these two sets accordingly in two very different ways.

Case I: For p *odd*, we first describe the set $\mathbb{R}_+^p$:

$$\mathbb{R}_+^p = \bigcup_{k=\{0,1,3,\ldots,p\}} \bigcup_{\{O_k\in\mathbf{P}_k\}} P_{O_k},$$

where $P_{O_k} = \{(u_1, \ldots, u_p) : u_i > 0 \text{ for } i \in O_k \text{ and } u_i < 0 \text{ for } i \in E_k\}$ with $O_k \subset \{1, \ldots, p\}$ of size k such that k is odd and E_k is such that $E_k \cup O_k = \{1, \ldots, p\}$. When $k = 0$, $P_{O_k} = P_+ = \{(u_1, \ldots, u_p) : u_i > 0 \text{ for } i = 1, \ldots, p\}$. Here and henceforth, $\mathbf{P}_k$ is the collection of all possible subsets of $\{1, 2, \ldots, p\}$ having size k.

Define the function $\Psi(\{1,\ldots,p\}, r) = \sum\cdots\sum_{1\le i_1<\cdots<i_r\le p}\prod_{j=1}^{r}\Phi_+(x_{i_j})$. In particular, for $r=1$, we have $\sum_{i=1}^{p}\Phi_+(x_i)$. We now give the expression of F_p^* for $\mathbf{x}\in\mathbb{R}_+^p$ as follows:

$$F_p^*(\mathbf{x}) = \begin{cases} \left\{\prod_{i=1}^{p}\Phi_+(x_i) + \sum_{r=1,3,\ldots,(p-2)}\Psi(\{1,\ldots,p\}, r)\right\}\Big/ 2^{(p-1)}, & \text{if } \mathbf{x}\in P_+,\\ \left\{\prod_{i\in O_k}\Phi_+(x_i)\prod_{j\in E_k}\Phi_-(x_j) + \sum_{r\in O_k}\Psi(O_k, r)\prod_{j\in E_k}\Phi_-(x_j)\right\}\Big/ 2^{(p-1)}, & \text{if } \mathbf{x}\in P_{O_k}. \end{cases}$$

Similarly, we now define a partition of $\mathbb{R}_-^p$ to be

$$\mathbb{R}_-^p = \bigcup_{k=\{0,2,4,\ldots,(p-1)\}}\ \bigcup_{\{E_k\in\mathbf{P}_k\}} P_{E_k},$$

where $P_{E_k} = \{(u_1,\ldots,u_p) : u_i > 0 \text{ for } i\in E_k \text{ and } u_i < 0 \text{ for } i \in O_k\}$ with $E_k\subset\{1,\ldots,p\}$ of size k such that k is even, and O_k is such that $E_k\cup O_k = \{1,\ldots,p\}$. When $k=0$, we have $P_{E_k} = P_- = \{(u_1,\ldots,u_p) : u_i < 0 \text{ for } i=1,\ldots,p\}$.

We now give the expression of F_p^* for $\mathbf{x}\in\mathbb{R}_-^p$ as follows:

$$F_p^*(\mathbf{x}) = \begin{cases} \left\{\sum_{r\in E_k}\Psi(E_k, r)\prod_{j\in O_k}\Phi_-(x_j)\right\}\Big/ 2^{(p-1)}, & \text{if } \mathbf{x}\in P_{E_k},\\ 0, & \text{if } \mathbf{x}\in P_-. \end{cases}$$

Case II: In a similar spirit, we have a representation of the set $\mathbb{R}_+^p$ for *even* p as follows:

$$\mathbb{R}_+^p = \bigcup_{k=\{0,2,4,\ldots,p\}}\ \bigcup_{\{E_k\in\mathbf{P}_k\}} P_{E_k},$$

where $P_{E_k} = \{(u_1,\ldots,u_p) : u_i > 0 \text{ for } i\in E_k \text{ and } u_i < 0 \text{ for } i\in E_k^c\}$ with $E_k\subset\{1,2,\ldots,p\}$ of size k is such that k is even, and O_k is such that $E_k\cup O_k=\{1,\ldots,p\}$. P_+ and P_- are as defined above. We now give the expression of F_p^* for $\mathbf{x}\in\mathbb{R}_+^p$ as follows:

$$F_p^*(\mathbf{x}) = \begin{cases} \left\{\prod_{i=1}^{p}\Phi_+(x_i) + \sum_{r=2,4,\ldots,(p-2)}\Psi(\{1,\ldots,p\}, r)\right\}\Big/ 2^{(p-1)}, & \text{if } \mathbf{x}\in P_+,\\ \prod_{i=1}^{p}\Phi_-(x_i)\Big/ 2^{(p-1)}, & \text{if } \mathbf{x}\in P_-,\\ \left\{\prod_{i\in E_k}\Phi_+(x_i)\prod_{j\in E_k^c}\Phi_-(x_j)\right\}\Big/ 2^{(p-1)}, & \text{if } \mathbf{x}\in P_{E_k}. \end{cases}$$

We also describe the set $\mathbb{R}_-^p$ below:

$$\mathbb{R}_-^p = \bigcup_{k=\{0,1,3,\ldots,(p-1)\}}\ \bigcup_{\{O_k\in\mathbf{P}_k\}} P_{O_k},$$

where $P_{O_k} = \{(u_1,\ldots,u_p) : u_i > 0 \text{ for } i\in O_k \text{ and } u_i < 0 \text{ for } i\in E_k\}$ with $O_k\subset\{1,2,\ldots,p\}$ of size k such that k is odd and E_k is such that $E_k\cup O_k = \{1,\ldots,p\}$. When $k=0$, we have $P_{O_k} = P_+$.

We now give the expression of F_p^* for $\mathbf{x}\in\mathbb{R}_-^p$ as follows:

$$F_p^*(\mathbf{x}) = \left\{\sum_{r\in E_k}\Psi(E_k, r)\prod_{j\in O_k}\Phi_-(x_j)\right\}\Big/ 2^{(p-1)}, \quad \text{if } \mathbf{x}\in P_{O_k}.$$

Note that $\frac{\partial\Phi(x)}{\partial x} = \phi(x)$ and $\phi(x) = \phi(-x)$ for any $x\in\mathbb{R}$. This now implies that $\frac{\partial\Phi_+(x)}{\partial x} = \frac{\partial\Phi_-(x)}{\partial x} = 2\phi(x)$ for any $x\in\mathbb{R}$. Also, note that

$$\frac{\partial^p\Psi(\{1,\ldots,p\}, r)}{\partial x_1\ldots\partial x_p} = 0,$$

since each term in the function Ψ involves a proper subset of each of the p variables $\Phi(x_1), \ldots, \Phi(x_p)$. Using the rule of differentiation with respect to the variables $x_1, \ldots, x_p$, the resulting density function, f_p^*, for any $p > 1$ can be obtained as follows:

$$f_p^*(\mathbf{x}) = \begin{cases} 2\phi(x_1)\cdots\phi(x_p) & \text{for } \mathbf{x} \in \mathbb{R}_+^p, \\ 0 & \text{for } \mathbf{x} \in \mathbb{R}_-^p. \end{cases}$$

Note that only the leading term in the expression of F_p^* in the set $\mathbb{R}_+^p$ contributes to the density function f_p^*. □

Proof of Theorem 4. Recall the definition $\mathbf{Y}_p^* \stackrel{d}{=} \mathbf{a}_p + \mathbf{B}_p\mathbf{X}_p^*$, where $\mathbf{a}_p$ is a vector in $\mathbb{R}^p$, and $\mathbf{B}_p$ is a $p \times p$ scale matrix for any $p \geq 2$. We define $\mathbf{Y}_{p-1,p}^*$ based on $\mathbf{X}_{p-1,p}^*$, which has a standard $(p-1)$-dimensional standard normal distribution. Note that linear transformations are again Gaussian random variates. Therefore, $\mathbf{Y}_{p-1,p}^*$ has a standard normal distribution, although $\mathbf{Y}_p^*$ will not be Gaussian because $\mathbf{X}_p^*$ has a non-Gaussian distribution.

Consider $\mathbf{B}_p = Z\mathbf{I}_p$, with Z a positive random variable, independent of $\mathbf{X}_p^*$. Then for any $(p-1)$-dimensional distribution, $\mathbf{Y}_{p-1,p}^*$ belongs to the same family of distributions because $\mathbf{X}_{p-1,p}^*$ has a standard normal distribution and the class of distributions is closed under affine transformations. But, $\mathbf{Y}_p^*$ does not because $\mathbf{X}_p^* \in \mathbb{R}_+^p$ with probability 1 by definition.

Further, assume $\mathbf{a}_p$ to be a random vector (independent of $\mathbf{X}_p^*$) and $\mathbf{Y}_{p-1,p}^*$, constructed from $\mathbf{X}_{p-1,p}^*$, belongs to the same family of distributions. By a similar argument as above, $\mathbf{Y}_p^*$ fails to belong to this class of distributions. □

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.
[2] R.B. Arellano-Valle, M.D. Branco, M.G. Genton, A unified view on skewed distributions arising from selections, Canad. J. Statist. 34 (2006) 581–601.
[3] R.B. Arellano-Valle, G. del Pino, E. San Martin, Definition and probabilistic properties of skew-distributions, Statist. Probab. Lett. 58 (2002) 111–121.
[4] R.B. Arellano-Valle, M.G. Genton, An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators, Ann. Inst. Statist. Math. 62 (2010) 363–381.
[5] B.C. Arnold, E. Castillo, J.M. Sarabia, Conditionally specified distributions: an introduction, Statist. Sci. 16 (2001) 249–274.
[6] A. Azzalini, with the collaboration of A. Capitanio, The Skew-Normal and Related Families, IMS Monographs Series, Cambridge University Press, 2014.
[7] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[8] A. Bhattacharyya, On some sets of sufficient conditions leading to the normal bivariate distribution, Sankhyā 6 (1943) 399–406. Proceedings of the Indian Statistical Conference.
[9] N.L. Biggs, Discrete Mathematics, Oxford University Press, New York, 2002.
[10] J.D. Broffitt, Zero correlation, independence, and normality, Amer. Statist. 40 (1986) 276–277.
[11] J.A. Cuesta-Albertos, E. del Barrio, R. Fraiman, C. Matran, The random projection method in goodness of fit for functional data, Comput. Statist. Data Anal. 51 (2007) 4814–4831.
[12] R.C. Elandt, The folded normal distribution: two methods of estimating parameters from moments, Technometrics 3 (1961) 551–562.
[13] K.T. Fang, S. Kotz, K.W. Ng, Symmetric Multivariate and Related Distributions, Chapman and Hall, London, New York, 1990.
[14] W. Feller, An Introduction to Probability Theory and its Applications. Vol. 2, second ed., John Wiley, New York, 1971.
[15] A. Gelman, X.L. Meng, A note on bivariate distributions that are conditionally normal, Amer. Statist. 45 (1991) 125–126.
[16] M.G. Genton, Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality. Edited Volume, Chapman & Hall/CRC, Boca Raton, FL, 2004, 416 p.
[17] A. Joffe, On a set of almost deterministic k-independent random variables, Ann. Probab. 2 (1974) 161–162.
[18] S. Kotz, N. Balakrishnan, N.L. Johnson, Continuous Multivariate Distributions, Models and Applications, John Wiley, New York, 2004.
[19] C.J. Kowalski, Non-normal bivariate distributions with normal marginals, Amer. Statist. 45 (1973) 103–106.
[20] B. Li, M.G. Genton, Nonparametric identification of copula structures, J. Amer. Statist. Assoc. 108 (2013) 666–675.
[21] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.
[22] C.J. Mecklin, D.J. Mundfrom, An appraisal and bibliography of tests for multivariate normality, Internat. Statist. Rev. 72 (2004) 123–138.
[23] R.J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley, New York, 2005.
[24] R.B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.
[25] J. Stoyanov, Counterexamples in Probability, Wiley, Chichester, 1997.
[26] G.J. Szekely, M.L. Rizzo, A new test for multivariate normality, J. Multivariate Anal. 93 (2005) 58–80.
[27] S.P. Vaswani, A pitfall in correlation theory, Nature 160 (1947) 405–406.
[28] J.A. Villaseñor-Alva, E. Gonzalez-Estrada, A generalization of Shapiro–Wilk's test for multivariate normality, Comm. Statist. Theory Methods 38 (2009) 1870–1883.