

SHORT REPORT

Open Access

# High-dimensional star-shaped distributions



Wolf-Dieter Richter 

Correspondence:  
wolf-dieter.richter@uni-rostock.de  
University of Rostock, Institute of  
Mathematics, Ulmenstraße 69,  
Rostock, Germany

## Abstract

Stochastic representations of star-shaped distributed random vectors having heavy or light tail density generating function  $g$  are studied for increasing dimensions along with corresponding geometric measure representations. Intervals are considered where star radius variables take values with high probability, and the derivation of values of distribution functions of  $g$ -robust statistics is proved to be based upon considering random events whose probability is asymptotically negligible if the dimension of the sample vector is approaching infinity. Moreover, a principal component representation of  $p$ -generalized elliptically contoured  $p$ -generalized Gaussian distributions is discussed.

**Keywords:** Increasing dimension, Star-shaped distribution, Star radius distribution, Star-uniform distribution,  $p$ -generalized elliptically contoured distribution, Principal component representation,  $g$ -robust statistic, Indeterminate form

## Introduction

Among the frequently obtained impressions one gets from analyzing high-dimensional data sets are that an observation point's distance from the zero element of the sample space is likely to belong to a certain interval from the positive real line, away from zero, and that the distribution of the direction of the vector seems to be close, in a certain sense, to a uniform distribution on the set of all directions that are observable from a certain center. The first observation can be reflected from a probabilistic point of view by a measure concentration type property including what is done in (Biau and Mason 2015) and (Vershynin 2016), and the second is part of background for testing uniformity on high-dimensional spheres, see e.g. (Cutting et al. 2017), possibly after projecting data points onto spheres as, e.g., in (Banerjee and Ghosh 2004).

In situations of the described type, it may be reasonable to model the data, or their residuals after fitting to a model, by multivariate star-shaped distributions. In this regard, (Balkema and Embrechts 2007) and (Balkema et al. 2010) discover conditions ensuring that star-shaped distributions with the Gauss-exponential law being one of the most known examples appear as limit laws in certain high-risk scenarios.

Distributions from the class of star-shaped distributions are flexible with respect to convexity or radial concavity, allow different variability of probability mass along different directions of the sample space and are able to model light and heavy distribution centers and tails. Because there is no natural number being representative for large dimensions, one might like to consider sequences or schemas of series of  $n$ -dimensional vectors with

$n$  approaching infinity. However, for simplicity of notation, we instead consider here just a single random vector  $X$  taking values in  $\mathbb{R}^n$  and assume afterwards that  $n$  is tending to infinity in formulas holding for  $X$ .

Let us recall at this point the following general aspect of uni- or multivariate asymptotic probabilistic analysis being of particular importance, for example, in large deviation theory, but not exclusively there. Studying the limit behavior of certain sequences of distributions on specific subsets of their ranges of definition and comparing it to how the appearing limit law itself behaves on the same sets needs to precisely know the latter one. In this respect, it is an independent problem to study the behavior limit laws show on the sets of interest. Similarly, if a sequence of distributions of increasing dimension is approximated in a certain part of its range of definition by a high-dimensional star-shaped limit law then studying the latter one is an independent problem being in the core of interest of the present note.

With the agreement of considering just one single vector  $X$  of dimension  $n$ , particular questions concerned by the buzzword 'big data' are approached in the present short note by reflecting above mentioned impressions gained from data in the language of probability distributions. To be more specific, we are dealing here with star-shaped distributions in  $\mathbb{R}^n$  and correspondingly distributed vectors. Such vector allows a stochastic representation as a product of a random generalized radius variable  $R$  and a random vector  $U$  being star-uniformly distributed on a star-sphere and independent of  $R$ , as well as a corresponding geometric measure representation. Some consequences which can be drawn from these representations in case of increasing dimension are studied. In particular, a representation of  $p$ -generalized elliptically contoured distributions is considered from the point of view of principal components.

The paper is structured as follows. In "Preliminaries" section, we present preliminary facts on star-shaped distributions including the notions of star surface content measure and star-uniform distribution on a star sphere. "A principal component representation" section deals with the particular class of  $p$ -generalized elliptically contoured distributions and it is studied there how they apply to modeling high-dimensional data. "A measure concentration property" section is then aimed to consider typical intervals where  $R$  takes values if  $X$  is star-shaped distributed, and in "On  $g$ -robust statistics" section distributions of univariate statistics are described which can basically be derived from star-uniformly distributed vectors. Such distributions are not affected by whether  $X$  has a density generating function  $g$  generating light or heavy distribution tails and is therefore called  $g$ -robust. The derivation of values of distribution functions of  $g$ -robust statistics is proved to be based upon considering random events whose probability is asymptotically negligible if the dimension of the sample vector is approaching infinity.

### Preliminaries

Let  $K \subset \mathbb{R}^n$  be a star body having the origin in its interior and assume that the Minkowski functional  $h_K$  of  $K$  is positively homogeneous of degree one. We call  $K(r) = rK$  and its boundary  $S(r) = rS$  the star ball and star sphere of star radius  $r > 0$ , respectively. If  $g : [0, \infty) \rightarrow [0, \infty)$  satisfies  $0 < I(n, g) < \infty$  where  $I(n, g) = \int_0^\infty r^{n-1} g(r) dr$  then it is called a density generating function. In such case,

$$\varphi_{g,K}(x) = C(g, K)g(h_K(x)), x \in \mathbb{R}^n$$

is called a star-shaped density and  $K$  its contour defining star body. The corresponding probability measure is denoted  $\Phi_{g,K}$  and the normalizing constant allows the representation

$$C(g, K) = \frac{1}{\mathfrak{D}_S(S)I(n, g)}$$

where  $\mathfrak{D}_S(S)$  means the star-generalized surface content of  $S$ , see (Richter 2014). If the additional assumption  $C(g, K) = 1$  is satisfied then  $g$  is called a density generator. In the following example, we recall an explicit analytical representation of the star-generalized surface content measure  $O_S$  in case  $S$  is a  $p$ -generalized ellipsoid with main axes of half lengths  $a_1, \dots, a_n$  and indicate relationships to representations in other particular cases.

**Example 1** Let  $K = \{x \in \mathbb{R}^n : |x|_{a,p} \leq 1\}$  where  $|x|_{a,p} = \left(\sum_{i=1}^n \left|\frac{x_i}{a_i}\right|^p\right)^{1/p}$ ,  $a = (a_1, \dots, a_n)^T$ ,  $a_i > 0, i = 1, \dots, n, p > 0$ , and  $S^{+(-)} = S \cap \{x : x_n > (<)0\}$  the upper (lower) half of the  $(a, p)$ -ellipsoid  $S = \{x : |x|_{a,p} = 1\}$ . Then

$$\mathfrak{D}_S(A) = a_n \left( \int_{G(A \cap S^+)} + \int_{G(A \cap S^-)} \right) \frac{d(x_1, \dots, x_{n-1})}{\left(1 - \sum_{i=1}^{n-1} \left|\frac{x_i}{a_i}\right|^p\right)^{1-1/p}}, A \in \mathfrak{B}(S) \tag{1}$$

where  $G(A \cap S^{+(-)}) = \{\vartheta \in \mathbb{R}^{n-1} : \exists \eta = \eta(\vartheta) \text{ s.t. } (\vartheta^T, \eta)^T \in A \cap S^{+(-)}\}$ ,  $\mathfrak{B}(S) = \mathfrak{B}^n \cap S$  and  $\mathfrak{B}^n$  denotes the Borel  $\sigma$ -field in  $\mathbb{R}^n$ .

- If  $p = 2$  and  $a = 1_n = (1, \dots, 1)^T$  then  $\mathfrak{D}_S(A)$  is the Euclidean surface content of the measurable subset  $A$  of  $S$ .
- If  $p = 1$  then  $\mathfrak{D}_S(A)$  can be considered as a particular polyhedral generalized surface content of  $A$ .
- If  $\mathfrak{D}_{S,\infty}(A)$  is defined as the limit of  $\mathfrak{D}_S(A), A \in \mathfrak{B}(S)$  as  $p \rightarrow \infty$  then  $\mathfrak{D}_{S,\infty}$  can be considered as another particular polyhedral generalized surface content measure. For the whole class of polyhedral generalized surface content measures, see (Richter and Schicker 2017).
- Generalizations of representation (1) hold true for all cases where  $K$  is a ball with respect to any norm or antinorm, see (Richter 2015).

The next example deals with the asymptotic behavior of star surface content and volume of star spheres and star balls or ellipsoids, respectively, if dimension is approaching infinity.

**Example 2** [a] It is well known that if  $S$  is the Euclidean unit sphere then  $\mathfrak{D}_S(S) = \omega_n$  where  $\omega_n = 2\pi^{n/2}/\Gamma(\frac{n}{2})$  is the Euclidean surface content of  $S$ . It is known that  $\arg \sup_n \omega_n = 7$  and that  $\omega_n$  is monotonously decreasing starting from this value, see e.g. (Loskot and Beaulieu 2007). Moreover, according to Stirling’s formula,

$$\mathfrak{D}_S(S) \sim \frac{\sqrt{2e}}{1 + \frac{1}{6n}} \left(\frac{2\pi e}{n}\right)^{(n-1)/2} \text{ as } n \rightarrow \infty$$

meaning that the ratio of the quantity on the left hand side divided by that of the right hand side tends to one if  $n$  tends to infinity. Obviously,  $\mathfrak{D}_S(S)$  is tending to zero quite fast as  $n \rightarrow \infty$ .

(b) If  $S$  is the  $l_{n,p}$ -sphere having unit star radius,  $p > 0$ , then the star surface content of  $S$  is known to be  $\mathfrak{D}_S(S) = \omega_{n,p}$  where  $\omega_{n,p} = 2^n (\Gamma(\frac{1}{p}))^n / (p^{n-1} \Gamma(\frac{n}{p}))$ . Note that

$$\mathfrak{D}_S(S) \sim \frac{p}{\sqrt{2\pi}} \left(\frac{p}{n}\right)^{\frac{n}{p}-\frac{1}{2}} \left(\frac{2\Gamma(\frac{1}{p})}{p}\right)^n e^{\frac{n}{p}}, n \rightarrow \infty.$$

Let  $\Omega_{n,p} = \frac{\omega_{n,p}}{n}$  denote the volume of the  $l_{n,p}$ -ball. The asymptotic relations following from the latter one,

$$\Omega_{n,p}^{\frac{p}{n}} \sim \frac{pe \left[\frac{2}{p}\Gamma\left(\frac{1}{p}\right)\right]^p}{n} \text{ and } \Omega_{n,p}^{\frac{p}{n \ln n}} \sim \frac{1}{e}, n \rightarrow \infty,$$

generalize two results given in (Chen and Lin 2014) for the particular Euclidean case  $p = 2$ .

(c) It is well known that the star surface content of the  $(a, p)$ -ellipsoid  $S = \{x \in \mathbb{R}^n : |x|_{a,p} = 1\}$  is  $\mathfrak{D}_S(S) = a_1 \cdots a_n \omega_{n,p}$ , thus

$$\mathfrak{D}_S(S) \sim a_1 \cdots a_n \frac{p}{\sqrt{2\pi}} \left(\frac{p}{n}\right)^{\frac{n}{p}-\frac{1}{2}} \left(\frac{2\Gamma(\frac{1}{p})}{p}\right)^n e^{\frac{n}{p}}, n \rightarrow \infty. \tag{2}$$

If a random vector  $X$  follows the star-shaped density  $\varphi_{g,K}$  then it allows the stochastic representation

$$X \stackrel{d}{=} R \cdot U$$

meaning that  $X$  is distributed as  $R \cdot U$ . The nonnegative random variable  $R$  is independent of the random vector  $U$ ,  $R$  has density function

$$f_R(r) = I(n, g)^{-1} r^{n-1} g(r) I_{[0, \infty)}(r)$$

and  $U$  has the star-uniform distribution

$$\omega_S(A) = \frac{O_S(A)}{O_S(S)}, A \in \mathfrak{B}(S).$$

Here,  $I_{[0, \infty)}(r) = 1$  if  $r \geq 0$  and  $I_{[0, \infty)}(r) = 0$  otherwise. Accordingly, the geometric measure representation of star-shaped distribution laws reads

$$\Phi_{g,K}(B) = \frac{1}{I(n, g)} \int_0^\infty r^{n-1} g(r) \mathfrak{F}_S(B, r) dr, B \in \mathfrak{B}^n$$

where

$$\mathfrak{F}_S(B, r) = \frac{\mathfrak{D}_S\left(\left[\frac{1}{r}B\right] \cap S\right)}{\mathfrak{D}_S(S)} = \omega_S\left(\left[\frac{1}{r}B\right] \cap S\right)$$

is the star sphere intersection proportion function of the set  $B$ . Let the Minkowski functional of  $K$  be denoted  $h_K$  then

$$R \stackrel{d}{=} h_K(X),$$

and  $R$  is called the star radius of  $X$ .

### A principal component representation

In this section we study to what extent a particular class of continuous multivariate star-shaped distributions applies to modeling high-dimensional data. To be specific, we consider  $p$ -generalized elliptically contoured  $p$ -generalized Gaussian distributions.

It will be shown which way principal component analysis can be used to identify those components of such star-shaped distributed vectors being of major importance for the modeling process. In particular it turns out that covariance ellipsoids being  $l_2$ -ellipsoids have the same main axes as density level set ellipsoids being  $l_p$ -ellipsoids.

Let  $X_i, i = 1, \dots, n$  be independent random variables correspondingly following the densities

$$f_p(x; \mu_i, \sigma_i) = \frac{C_p}{\sigma_i} \exp \left\{ -\frac{|x - \mu_i|^p}{p\sigma_i^p} \right\}, x \in \mathbb{R}$$

where  $C_p = p^{1-1/p} / (2\Gamma(\frac{1}{p}))$ . Then  $\mathbb{E}X_i = \mu_i$  and  $V(X_i) = \kappa\sigma_i^2$  are expectation and variance of  $X_i$ , respectively,  $i = 1, \dots, n$ . Here and in what follows we use the notation

$$\kappa(n) = p^{\frac{2}{p}} \frac{\Gamma(\frac{n+2}{2})}{\Gamma(\frac{n}{2})} \text{ and } \kappa = \kappa(1).$$

Let  $\sigma = (\sigma_1, \dots, \sigma_n)^T, X = (X_1, \dots, X_n)^T$  and  $\mu = (\mu_1, \dots, \mu_n)^T$ , then  $X - \mu$  allows the stochastic representation

$$X - \mu \stackrel{d}{=} R_{\sigma,p} U_{\sigma,p}$$

where  $R_{\sigma,p} = |X - \mu|_{\sigma,p}$  and  $U_{\sigma,p} = \frac{1}{R_{\sigma,p}}(X - \mu)$  are independent. The star radius  $R_{\sigma,p}$  follows the  $p$ -generalized *Chi*-distribution with  $n$  d.f. having according to (Richter 2007) the density

$$f(r) = \frac{r^{n-1} e^{-\frac{r^p}{p}}}{p^{\frac{n}{p}-1} \Gamma(\frac{n}{p})} I_{[0,\infty)}(r),$$

and the stochastic basis vector  $U_{\sigma,p}$  is star-uniformly distributed on the star sphere  $S = E_{\sigma,p}$  from Example 2(c) with  $a_i = \sigma_i, i = 1, \dots, n, U_{\sigma,p} \sim \omega_{\sigma,p}$ . Thus, using notation in (Richter 2014),  $X$  belongs to the class  $EC_{\sigma,p,\mu,I_n}$  and its density generating function can be chosen as  $g(r) = g_p(r) = e^{-\frac{r^p}{p}} I_{[0,\infty)}(r)$ . Note that  $\mathbb{E}U_{\sigma,p} = 0_n$  and, because  $\mathbb{E}R_{\sigma,p}^2 = \kappa(n)$ ,

$$\text{cov}(U_{\sigma,p}) = \frac{\kappa}{\kappa(n)} D^2$$

where

$$D = \text{diag}(\sigma_1, \dots, \sigma_n).$$

It follows that  $\mathbb{E}X = \mu$  and  $\text{cov}(X) = \kappa D^2$ . Now, denote an orthogonal  $n \times n$ -matrix  $O = (O_{i,j})_{i,j=1,\dots,n}$  and let the transpose of its  $i$ 'th row be  $O_i = (O_{i,1}, \dots, O_{i,n})^T$ . The random vector  $Y = O(X - \mu)$  follows the  $p$ -generalized elliptically contoured density

$$f_Y(y) = \frac{C_p^n}{\sigma_1 \cdots \sigma_n} \exp \left\{ -\frac{1}{p} |O^T y|_{\sigma,p}^p \right\}, y \in \mathbb{R}^n, \tag{3}$$

that is  $Y$  belongs to the class  $EC_{\sigma,p,0_n,O}$ . Note that  $\mathbb{E}Y = 0_n$  and  $\text{cov}(Y) = \mathbb{E}YY^T = \Sigma = \kappa O D^2 O^T = (\sigma_{i,j})_{i,j=1,\dots,n}$  where

$$\sigma_{i,j} = \text{cov}(Y_i, Y_j) = \kappa O_i^T D^2 O_j, 1 \leq i, j \leq n.$$

Thus,  $f_Y = \varphi_{g_p,OK}$  and the boundary of  $K$  is  $S = E_{\sigma,p}$ . Note that  $OK$  is a star body having the properties introduced at the beginning of this section. The covariance ellipsoid of  $Y$  is

$$C(\Sigma) = \{x \in \mathbb{R}^n : x^T \Sigma^{-1} x = 1\} = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n \frac{|\Pi_{O_i} x|^2}{\sigma_i^2} = \kappa \right\}$$

where  $\Pi_y x$  means the orthogonal projection of  $x$  into the linear space spanned up by  $y$ . The main axes of  $C(\Sigma)$  belong to the spaces spanned up by the vectors  $O_i$  and have half lengths of size  $\sqrt{\kappa}\sigma_i, i = 1, \dots, n$ , respectively. Moreover, the set  $C(\Sigma)$  is symmetric with respect to any of the lines  $L_i = \mathcal{L}\{O_i\}, i = 1, \dots, n$ . The latter holds also true for the  $p$ -generalized ellipsoids

$$OE_{\sigma,p} = \{Ox \in \mathbb{R}^n : |x|_{\sigma,p} = 1\}$$

because  $E_{\sigma,p}$  is symmetric with respect to the lines

$$L_i^* = \{x \in \mathbb{R}^n : x = \lambda e_i, \lambda \in \mathbb{R}\},$$

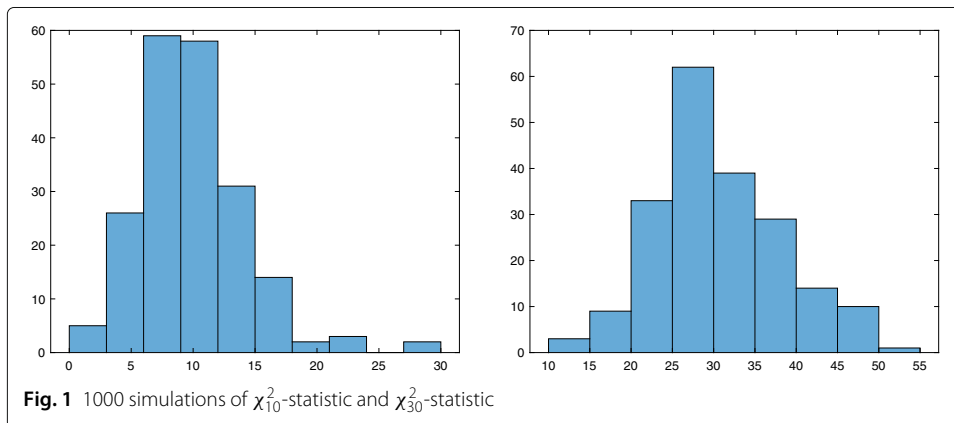
for  $i = 1, \dots, n$ , and

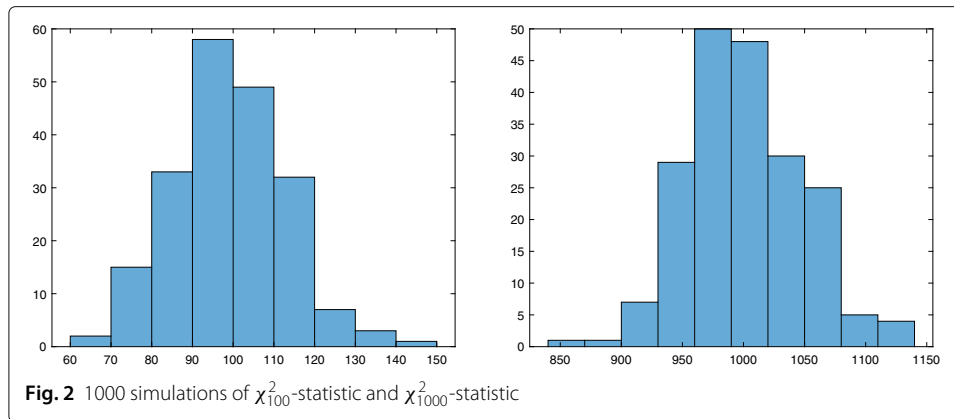
$$OL_i^* = \{\lambda Oe_i : \lambda \in \mathbb{R}\} = \{\lambda O_i : \lambda \in \mathbb{R}\} = L_i$$

with  $e_1, \dots, e_n$  being the standard orthonormal basis vectors in  $\mathbb{R}^n$ . One may chose the variances according to the maximization procedure from principal component analysis,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ . For high-dimensional data holds  $\sigma_n \rightarrow 0$  as  $n \rightarrow \infty$ . This circumstance allows to introduce data reduction. Therefore, methods from this area of statistical analysis apply to model data of arbitrary fixed or increasing dimension by  $p$ -generalized elliptically contoured distributions if  $p \geq 1$ . If  $p \in (0, 1)$ , certain maximization principles from PCA are to be changed with corresponding minimization principles. In this sense, representation (3) may be called a principal component representation of  $p$ -generalized elliptically contoured  $p$ -generalized Gaussian densities. With slight changes, the density generating function  $g = g_p$  may be replaced in (3) with an arbitrary one for representing general  $p$ -generalized elliptically contoured densities.

**A measure concentration property**

A  $\chi^2$ -distributed random variable with  $n$  d.f. can be considered as the square of the Euclidean norm of an  $n$ -dimensional standard Gaussian vector. The probability mass of such vector is the more concentrated in a relatively small shell having radius of order  $\sqrt{n}$  the larger the vector’s dimension is, see Figs. 1 and 2. Observing the empirical ”relative concentration numbers” 30/10, 45/30, 90/100 and 300/1000 one may argue that suitably defined numbers might even converge to zero in some sense. This will be proved here within the even more general frame of  $\chi^p$ -distributions. For definitions and properties of these distributions we refer to (Richter 2007; 2009; 2014; 2015; 2016).





If a multivariate distribution converges to the standard Gaussian law then the square of the Euclidean norm of the correspondingly distributed vector  $X$ , i.e. the square of the Euclidean radius  $R = |X|_{1,2}$  of such vector, will tend under some additional assumption to the  $\chi^2$ -distribution with  $n$  d.f. . Let now  $X$  follow a star-shaped distribution  $\Phi_{g,K}$ , what can we say then about the behavior of (the suitably defined power of) its generalized (star) radius? In this section we derive typical intervals where star radius variables of high-dimensional star-shaped vectors take values.

**Proposition 1** For  $\delta > 0$  chosen such that

$$\alpha = \frac{1}{\delta^2} \left( \frac{I(n+2, g)I(n, g)}{I^2(n+1, g)} - 1 \right)$$

is approaching zero as dimension  $n$  is tending to infinity, and independently of the shape defining star body  $K$ , the typical behavior of the random star radius of a vector following the star-shaped distribution  $\Phi_{g,K}$  with density generating function  $g$  is described by

$$P \left( (1 - \delta) \frac{I(n+1, g)}{I(n, g)} \leq R \leq (1 + \delta) \frac{I(n+1, g)}{I(n, g)} \right) \geq 1 - \alpha.$$

*Proof* Obviously,

$$\mathbb{E}R^k = \frac{I(n+k, g)}{I(n, g)}.$$

Now, Tschebyscheff's inequality applies  $\square$

According to (Biau and Mason 2015), the behavior of  $|X|_{1,p}$  as  $n$  increases is called the distance concentration phenomenon in the computational learning literature. For sums of independent random variables or matrices, sharper concentration inequalities of exponential type are proved in (Vershynin 2016).

For more details on moments of  $p$ -spherical random vectors, see (Arellano-Valle and Richter 2012), for an asymmetric situation if  $p = 1$  see (Henschel and Richter 2002). The following corollary deals with a class of light tailed high-dimensional star-shaped distributions.

**Corollary 1** Let  $K$  be any star body as introduced in "Preliminaries" section. If the density generating function of a high-dimensional star-shaped distribution  $\Phi_{g,K}$  is that of Kotz type with parameters  $s > 0, t > 0, k > 1 - n$ ,

$$g(r) = r^{k-1} e^{-tr^s} I_{[0,\infty)}(r),$$

and

$$\delta \gg \frac{1}{\sqrt{n}}$$

(meaning that  $\delta\sqrt{n} \rightarrow \infty$  as  $n \rightarrow \infty$ ) then, for sufficiently large  $n$ , there holds

$$P\left( (1 - \delta) \frac{\Gamma\left(\frac{n+k}{s}\right)}{t^{1/s} \Gamma\left(\frac{n+k-1}{s}\right)} \leq R \leq (1 + \delta) \frac{\Gamma\left(\frac{n+k}{s}\right)}{t^{1/s} \Gamma\left(\frac{n+k-1}{s}\right)} \right) \geq 1 - \alpha.$$

*Proof* First we check that  $\alpha$  tends to zero: because

$$\frac{I(n+2, g)I(n, g)}{I^2(n+1, g)} = 1 + \frac{1}{s(n+k)} + O\left(\frac{1}{n^2}\right),$$

where  $O(\cdot)$  means Landau's big O symbol, it follows

$$\alpha = \frac{1}{\delta^2} \left[ \frac{1}{s(n+k)} + O\left(\frac{1}{n^2}\right) \right], n \rightarrow \infty.$$

Such  $\alpha$  approaches zero as  $n$  tends to infinity if  $\delta^2 \cdot n \rightarrow \infty$  for  $n \rightarrow \infty$ . Finishing the proof, we finally observe that

$$\frac{I(n+1, g)}{I(n, g)} = \frac{\Gamma\left(\frac{k+n}{s}\right)}{t^{1/s} \Gamma\left(\frac{k+n-1}{s}\right)} \square$$

□

**Remark 1** On using Stirling's formula, it can be seen that

$$\frac{\Gamma\left(\frac{n+k}{s}\right)}{\Gamma\left(\frac{n+k-1}{s}\right)} = \left(\frac{n+k}{s}\right)^{1/s} \left(1 + O\left(\frac{1}{n}\right)\right) \text{ as } n \rightarrow \infty. \tag{4}$$

Before turning to the case of heavy distribution tails, we note that asymptotic relation (4) makes the statement of Corollary 1 more specific in the sense that

$$I_{Kt}(R) = \left( (1 - \delta) \left(\frac{n+k}{ts}\right)^{1/s}, (1 + \delta) \left(\frac{n+k}{ts}\right)^{1/s} \right)$$

is a reasonable interval where  $R$  takes values with high probability if we are given a high-dimensional star-shaped vector with density generating function of Kotz type and fixed or increasing dimension  $n$ . The most essential role is played here by parameter  $s$  which basically determines the relative heaviness or lightness of the distribution tails.

**Corollary 2** Let  $K$  be any star body as introduced in "Preliminaries" section. If the density generating function of a star-shaped distribution  $\Phi_{g,K}$  is that of Pearson type VII with parameters  $s > 0, k > n + 2$  [12] where  $k - n \rightarrow \infty$  as  $n \rightarrow \infty$ ,

$$g(r) = \left(1 + \frac{r}{s}\right)^{-k} I_{[0,\infty)}(r),$$

and

$$\delta \gg \frac{1}{\sqrt{n}} \text{ as well as } \delta \gg \frac{1}{\sqrt{k-n}},$$



then, for sufficiently large  $n$ , with probability greater or equal to  $1 - \alpha$ , the star-radius  $R$  of the high-dimensional vector  $X$  belongs to the interval

$$I_{P7}(R) = \left( (1 - \delta) \frac{sn}{k - n - 1}, (1 + \delta) \frac{sn}{k - n - 1} \right).$$

*Proof* Checking that  $\alpha$  tends to zero, we find that

$$\frac{I(n + 2, g)I(n, g)}{I^2(n + 1, g)} = 1 + \frac{1}{n} + \frac{1}{k - n - 2} + \frac{1}{n(k - n - 2)}.$$

The proof is finished by observing that

$$\frac{I(n + 1, g)}{I(n, g)} = \frac{sn}{k - n - 1} \quad \square$$

$\square$

**Remark 2** Let  $\delta = \delta(n) \rightarrow +0$  and  $\alpha = \alpha(n) \rightarrow +0$  as  $n \rightarrow \infty$  such that  $n\delta^2(n) \rightarrow \infty$  and assume that in the situation of Corollary 2 there holds additionally that  $(k - n)\delta^2(n) \rightarrow \infty$ . The statements of Corollaries 1 and 2 can be reformulated then as

$$P(R > (1 + \delta)\mathbb{E}R \text{ or } R < (1 - \delta)\mathbb{E}R) = O(\alpha(n)) \text{ as } n \rightarrow \infty$$

where Landau's symbol  $O$  defined for the asymptotic relation  $f(n) = O(g(n))$ ,  $n \rightarrow \infty$  guaranties the existence of a constant  $C$  such that for all  $n$  there holds  $|f(n)/g(n)| \leq C$ . Moreover, in the situation of Corollaries 1 and 2 we have that

$$\mathbb{E}R = \left( \frac{n + k}{ts} \right)^{1/s} \left( 1 + O\left(\frac{1}{n}\right) \right) \text{ and } \mathbb{E}R = \frac{sn}{k - n - 1},$$

respectively.

**Remark 3** Let us finally mention that because in fact we are considering sequences or even schemas of series of vectors and distributions, the assumption  $k - n \rightarrow \infty$  stated in Corollary 2 is not contradictory. Instead, it ensures a certain variability of the result. Moreover, we remark that (Henschel 2001) and (Henschel and Richter 2002) study the exact distribution of  $R$  in case of simplicially contoured vectors (or  $l_{n,1}$ -spherical vectors having nonnegative components). General  $l_{n,p}$ -spherical vectors and their star radius  $R$  are studied in (Richter 2009), (Arellano-Valle and Richter 2012) and (Richter 2014), tables of corresponding exact quantiles of  $R^p$  and  $R$  are to be found in (Müller and Richter 2016) and (Richter 2016).

### On $g$ -robust statistics

If the distribution of a statistic does not depend on the density generating function  $g$  of a star-shaped sample vector density  $\varphi_{g,K}$  then it is commonly called  $g$ -robust. It is well known that Student and Fisher type statistics possess besides the  $g$ -robustness property further optimality properties. Here we will see that decisions based upon such statistics are done by closer analyzing random events whose probability is asymptotically negligible if the dimension of the sample vector is approaching infinity.

To be more concrete, in this section, we describe a class of statistical distribution functions, derived from a star-shaped sample vector, the ratio representations of whose values are asymptotically negligible as vector dimension increases unboundedly. It turns out, e.g., that classical and generalized Student and Fisher distributions belong to this class.

Let a random vector  $X$  follow the distribution law  $\Phi_{g,K}, X \sim \Phi_{g,K}$ , and the sets

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

being generated by a statistic  $T : \mathbb{R}^n \rightarrow \mathbb{R}$  such that the equation

$$\mathfrak{F}_S(B(t), r) = \mathcal{C}(t), \forall t$$

is satisfied where  $\mathcal{C}(t) \in [0, 1]$  does not depend on  $r$ . A statistic  $T$  of this type is  $g$ -robust. It follows by the geometric representation proved for star-shaped distributions in (Richter 2014) that

$$P(T(X) < t) = \Phi_{g,K}(B(t)) = \frac{\mathfrak{V}_S(B(t) \cap S)}{\mathfrak{V}_S(S)}. \tag{5}$$

The distribution law of  $T(X)$  is determined already by that of vector's  $X$  central projection onto the star sphere  $S$ . Thus, for  $t \in \mathbb{R}$ , the probability  $P(T(X) < t)$  is defined by the  $\omega_S$ -value of the random event  $B(t) \cap S$  in the geometric probability space  $(S, \mathfrak{B}(S), \omega_S)$ , and its representation in (5) will be called its ratio representation. It can be observed that many star spheres show the asymptotic behaviour

$$\mathfrak{V}_S(S) \rightarrow 0 \text{ as } n \rightarrow \infty. \tag{6}$$

For a particular case of such type, see Example 2(a). The statistical model concerned in this case is dealing with independent and homoscedastic random variables. In case of increasing dimensions, we are confronted then with sequences of probability spaces with asymptotically negligible set  $S$ .

**Example 3** *In case  $S$  is the unit  $l_{n,p}$ -sphere, condition (6) is satisfied.*

In the following example we restrict our consideration to the  $n$ -dimensional standard Gaussian law  $\Phi_{g,K} = \Phi$  where  $g(r) = e^{-r^2/2}$  and  $K = \{x : x_1^2 + \dots + x_n^2 \leq 1\}$ .

**Example 4** *A set  $A \subset \mathbb{R}^n$  belongs to the class  $\mathfrak{A}(\text{dir}, \text{dist})$  if there exist functions  $e_A : [0, \infty) \rightarrow S_n(1)$  and  $R_A : [0, \infty) \rightarrow [0, \infty)$  satisfying the following two assumptions:*

$\mathfrak{A}1)$  *The set  $A$  allows the representation*

$$A \cap S_n(r) = H_n(e_A(r), R_A(r)) \cap S_n(r), r > 0,$$

where  $H_n(e, R) = \{x \in \mathbb{R}^n : \Pi_e x = \lambda e, \lambda \geq R\}$  is a halfspace and  $S_n(r)$  the Euclidean sphere of radius  $r$ .

$\mathfrak{A}2)$  *The function  $\mathfrak{C} : [0, \infty) \rightarrow \mathbb{R}^n$  with*

$$\mathfrak{C}(r) = R_A(r)e_A(r), r \geq 0$$

*is a piecewise continuous curve such that  $A$  becomes a Borel set.*

*The functions  $e_A$  and  $R_A$  are called directional type and distance type functions of the set  $A$ , respectively. If  $A \in \mathfrak{A}(\text{dir}, \text{dist})$  then*

$$\begin{aligned} \Phi(A) &= \frac{C(n-1)}{\sqrt{2\pi}} \int_0^\infty r^{n-1} e^{-r^2/2} \int_0^{\alpha^*(r)} (\sin \alpha)^{n-2} d\alpha dr, \\ \alpha^*(r) &= \arctan \left( (r/R_A(r))^2 - 1 \right)^{1/2} \in (0, \pi), r > 0 \end{aligned}$$

where  $C(n) = 2^{1-n/2} / \Gamma(n/2)$ , see (Richter 1995). If the function  $r \rightarrow \alpha^*(r)$  is constant then

$$\Phi(A) = \frac{\omega_{n-1}}{\omega_n} \int_0^{\alpha^*} (\sin \alpha)^{n-2} d\alpha$$

where

$$\frac{\omega_{n-1}}{\omega_n} \sim \sqrt{\frac{n}{2\pi}}, n \rightarrow \infty.$$

**Example 5** Let  $\frac{1}{\sigma}X, \sigma > 0$ , be a standard Gaussian distributed random vector in  $R^n$ . The statistic

$$T_{\epsilon, \mathcal{N}} = \frac{(X, \epsilon)}{\|\Pi_{\mathcal{N}}X\|/\sqrt{k}}$$

is known to be Student distributed with  $k$  d.f. for all  $\epsilon \in S_n(1)$  and all  $k$ -dimensional linear subspaces  $\mathcal{N}$  of  $R^n$  such that  $\epsilon \perp \mathcal{N}$  and  $k \leq n - 1$ .

Let  $A = B(t) = \{T_{\epsilon, \mathcal{N}} < t\}$ . Then  $A \in \mathfrak{A}(\text{dir}, \text{dist})$  where  $e_A(r) = \epsilon$ , the distance type function is  $R_{B(t)}(r) = \tilde{t}r/(\tilde{t}^2 + 1)^{1/2}$ ,  $\tilde{t} = t/\sqrt{n - 1}$  and the function  $\alpha^*(r) = \arctan(1/\tilde{t})$  is constant. Evaluating with  $k = n - 1$  the limit of  $\Phi(A)$  as  $n \rightarrow \infty$  leads to the well known result  $\Phi_{0,1}(t)$  where  $\Phi_{0,1}$  denotes the cumulative distribution function of the univariate standard normal distribution.

For similar properties of a corresponding exact Student test in nonlinear regression, see (Ittrich 2000) and (Ittrich and Richter 2005).

**Example 6** For a related consideration on the  $p$ -generalized Fisher statistic, see (Richter 2009).

**Remark 4** If one is interested in avoiding the asymptotic negligibility of  $S$  in case of increasing dimension one may leave the class of statistical models dealing with independent homoscedastic observations. Density level sets of sample vectors having heteroscedastic components may be star-shaped. If  $S$  is an  $(a, p)$ -ellipsoid then, according to asymptotic relation (2), condition (6) may be violated and even asymptotically stabilizing.

**Acknowledgement**

The author likes to thank all the Reviewers, Associated Editor and Editor for their valuable comments.

**Funding**

Not applicable.

**Availability of data and material**

Not applicable.

**Competing interests**

The author declares no conflict of interest

Received: 5 September 2018 Accepted: 21 May 2019

Published online: 06 June 2019

**References**

Arellano-Valle, R. B., Richter, W.-D.: On skewed continuous  $I_{n,p}$ -symmetric distributions. *Chil. J. Stat.* **3**(2), 193–212 (2012)  
 Balkema, A., Embrechts, P., Nolde N.: Meta densities and the shape of their sample clouds. *J. Multivar. Anal.* **101**, 1738–1754 (2010)  
 Balkema, G., Embrechts, P.: High Risk Scenarios and Extremes. A geometric approach. European Mathematical Society Publishing House, Zürich (2007)

- Banerjee, A., Ghosh, J.: Frequency sensitive competitive learning for scalable balanced clustering on high-dimensional hypersurfaces. *IEEE Neural Netw.* **15**, 702–719 (2004)
- Biau, G., Mason, D.: High-dimensional p-norms. In: Hallin P Mason, Steinebach (eds) *Mathematical statistics and limit theorems*, pp. 21–40. Springer, Cham, (2015)
- Chen, C. P., Lin, L.: Inequalities for the volume of the unit ball in  $\mathbb{R}^n$ . *Mediterr. J. Math.* **11**, 299–314 (2014)
- Cutting, C., Paidaveine, D., Verdebout, T.: Testing uniformity on high-dimensional spheres. *Anna Stat.* **45**(3), 1024–1058 (2017)
- Henschel, V.: *Ausgewählte lineare Modelle in simplizial konturiert verteilten Grundgesamtheiten*. GCA-Verlag, Herdecke (2001)
- Henschel, V., Richter, W.-D.: Geometric generalization of the exponential law. *J. Multivar. Anal.* **81**(2), 189–204 (2002)
- Ittrich, C.: *Exakte Methoden in Regressionsmodellen mit einem nichtlinearen Parameter und sphärisch symmetrischen Fehlern*. Shaker, Aachen (2000)
- Ittrich, C., Richter, W.-D.: Exact tests and confidence regions in nonlinear regression. *Statistics.* **39**, 13–42 (2005)
- Loskot, P., Beaulieu, N. C.: On monotonicity of the hypersphere volume and area. *J. Geo.* **87**, 96–98 (2007). <https://doi.org/10.1007/s00022-007-1891-1>
- Müller, K., Richter, W.-D.: Extreme value distributions for dependent jointly  $L_{n,p}$ -symmetrically distributed random variables. *Depend. Model.* **4**, 30–62 (2016). <https://doi.org/10.1515/demo-2016-0002>
- Richter, W.-D.: A geometric approach to the Gaussian law. In: Mammitsch/Schneeweiß (ed.) *Symposia Gaussiana*, pp. 25–45. Walter de Gruyter & Co., Berlin, Conf. B. (1995)
- Richter, W.-D.: Generalized spherical and simplicial coordinates. *J. Math. Anal. Appl.* **335**, 1187–1202 (2007). <https://doi.org/10.1016/j.jmaa.2007.03.047>
- Richter, W.-D.: Continuous  $L_{n,p}$ -symmetric distributions. *Lith. Math. J.* **49**(1), 93–108 (2009). <https://doi.org/10.1007/s10986-009-9030-3>
- Richter, W.-D.: Geometric disintegration and star-shaped distributions. *J. Stat. Distrib. Appl.* **1**(20) (2014). <https://doi.org/10.1186/s40488-014-0020-6>
- Richter, W.-D.: Convex and radially concave contoured distributions. *J. Probab. Stat.*, [56]12 (2015). <https://doi.org/10.1155/2015/165468>, Article ID 165468
- Richter, W.-D.: Exact inference on scaling parameters in norm and antinorm contoured sample distributions. *J. Stat. Distrib. Appl.* **3**(8) (2016). <https://doi.org/10.1186/s40488-016-0046-z>
- Richter, W.-D., Schicker, K.: Polyhedral star-shaped distributions. *J. Probab. Stat.*, 35 (2017). <https://doi.org/10.1155/2017/7176897>
- Vershynin, R.: Four lectures on probabilistic methods for data science 41 (2016). arXiv 1612.06661. Cornell University. <https://arxiv.org/abs/1612.06661>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---