

# Second Order Expansions for High-Dimension Low-Sample-Size Data Statistics in Random Setting

Gerd Christoph <sup>1,2,†</sup>  and Vladimir V. Ulyanov <sup>2,3,\*,†</sup> 

<sup>1</sup> Department of Mathematics, Otto-von-Guericke University Magdeburg, 39016 Magdeburg, Germany; gerd.christoph@ovgu.de

<sup>2</sup> Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, 119991 Moscow, Russia

<sup>3</sup> Faculty of Computer Science, National Research University Higher School of Economics, 167005 Moscow, Russia

\* Correspondence: vulyanov@cs.msu.ru or vulyanov@hse.ru

† These authors contributed equally to this work.

Received: 12 May 2020; Accepted: 9 July 2020; Published: 14 July 2020



**Abstract:** We consider high-dimension low-sample-size data taken from the standard multivariate normal distribution under assumption that dimension is a random variable. The second order Chebyshev–Edgeworth expansions for distributions of an angle between two sample observations and corresponding sample correlation coefficient are constructed with error bounds. Depending on the type of normalization, we get three different limit distributions: Normal, Student’s *t*-, or Laplace distributions. The paper continues studies of the authors on approximation of statistics for random size samples.

**Keywords:** second order expansions; high-dimensional; low sample size; random sample size; Laplace distribution; Student’s *t*-distribution

**MSC:** 62E17 (Primary) 62H10; 60E05 (Secondary)

## 1. Introduction

Let  $\vec{X}_1 = (X_{11}, \dots, X_{1m})^T, \dots, \vec{X}_k = (X_{k1}, \dots, X_{km})^T$  be a random sample from  $m$ -dimensional population. The data set can be regarded as  $k$  vectors or points in  $m$ -dimensional space. Recently, there has been significant interest in a high-dimensional datasets when the dimension is large. In a high-dimensional setting, it is assumed that either (i)  $m$  tends to infinity and  $k$  is fixed, or (ii) both  $m$  and  $k$  tend to infinity. Case (i) is related to high-dimensional low sample size (HDLSS) data. One of the first results for HDLSS data appeared in Hall et al. [1]. It became the basis of research in mathematical statistics for the analysis of high-dimensional data, see, e.g., Fujikoshi et al. [2], which are an important part of the current data analysis fashionable area called *Big data*. Scientific areas where these settings have proven to be very useful include genetics and other types of cancer research, neuroscience, and also image and shape analysis. See a recent survey on HDLSS asymptotics and its applications in Aoshima et al. [3].

For examining the features of the data set, it is necessary to study the asymptotic behavior of three functions: the length  $\|\vec{X}_i\|$  of a  $m$ -dimensional observation vector, the distance  $\|\vec{X}_i - \vec{X}_j\|$  between any two independent observation vectors, and the angle  $\text{ang}(\vec{X}_i, \vec{X}_j)$  between these vectors at the

population mean. Assuming that  $\vec{X}_i$ 's are a sample from  $N(0, I_m)$ , it was shown in Hall et al. [1] that for HDLSS data the three geometric statistics satisfy the following relations:

$$\|\vec{X}_i\| = \sqrt{m} + \mathcal{O}_p(1), \quad i = 1, \dots, k, \quad (1)$$

$$\|\vec{X}_i - \vec{X}_j\| = \sqrt{2m} + \mathcal{O}_p(1), \quad i, j = 1, \dots, k, \quad i \neq j, \quad (2)$$

$$\text{ang}(\vec{X}_i, \vec{X}_j) = \frac{1}{2}\pi + \mathcal{O}_p(m^{-1/2}), \quad i, j = 1, \dots, k, \quad i \neq j, \quad (3)$$

where  $\|\cdot\|$  is the Euclidean distance and  $\mathcal{O}_p$  denotes the stochastic order. These interesting results imply that the data converge to the vertices of a deterministic regular simplex. These properties were extended for non-normal sample under some assumptions (see Hall et al. [1] and Aoshima et al. [3]). In Kawaguchi et al. [4], the relations (1)–(3) were refined by constructing second order asymptotic expansions for distributions of all three basic statistics. The refinements of (1) and (2) were achieved by using the idea of Ulyanov et al. [5] who obtained the computable error bounds of order  $\mathcal{O}(m^{-1})$  for the chi-squared approximation of transformed chi-squared random variables with  $m$  degrees of freedom.

The aim of the present paper is to study approximation for the third statistic  $\text{ang}(\vec{X}_1, \vec{X}_2)$  under generalized assumption that  $m$  is a realization of a random variable, say  $N_n$ , which represents the sample dimension and is independent of  $\vec{X}_1$  and  $\vec{X}_2$ . This problem is closely related to approximations of statistics constructed from the random size samples, in particular, to this kind of problem for the sample correlation coefficient  $R_m$ .

The use of samples with random sample sizes has been steadily growing over the years. For an overview of statistical inferences with a random number of observations and some applications, see Esquivel et al. [6] and the references cited therein. Gnedenko [7] considered the asymptotic properties of the distributions of sample quantiles for samples of random size. In Nunes et al. [8] and Nunes et al. [9], unknown sample sizes are assumed in medical research for analysis of one and more than one-way fixed effects ANOVA models to avoid false rejections, obtained when using the classical fixed size  $F$ -tests. Esquivel et al. [6] considered inference for the mean with known and unknown variance and inference for the variance in the normal model. Prediction intervals for the future observations for generalized order statistics and confidence intervals for quantiles based on samples of random sizes are studied in Barakat et al. [10] and Al-Mutairi and Raqab [11], respectively. They illustrated their results with real biometric data set, the duration of remission of leukemia patients treated by one drug. The present paper continues studies of the authors on non-asymptotic analysis of approximations for statistics based on random size samples. In Christoph et al. [12], second order expansions for the normalized random sample sizes are proved, see below Propositions 1 and 2. These results allow for proving second order asymptotic expansions of random sample mean in Christoph et al. [12] and random sample median in Christoph et al. [13]. See also Chapters 1 and 9 in Fujikoshi and Ulyanov [14].

The structure of the paper is the following. In Section 2, we describe the relation between  $\text{ang}(\vec{X}_1, \vec{X}_2)$  and  $R_m$ . We recall also previous approximation results proved for distributions of  $\text{ang}(\vec{X}_1, \vec{X}_2)$  and  $R_m$ . Section 3 is on general transfer theorems, which allow us to construct asymptotic expansions for distributions of randomly normalized statistics on the base of approximation results for non-randomly normalized statistics and for the random size of the underlying sample, see Theorems 1 and 2. Section 4 contains the auxiliary lemmas. Some of them have independent interest. For example, Lemma 3 on the upper bounds for the negative order moments of a random variable having negative binomial distribution. We formulate and discuss main results in Sections 5 and 6. In Theorems 3–8, we construct the second order Chebyshev–Edgeworth expansions for distributions of  $\text{ang}(\vec{X}_1, \vec{X}_2)$  and  $R_m$  in random setting. Depending on the type of normalization, we get three different limit distributions: Normal, Laplace, or Student's  $t$ -distributions. All proofs are given in the Appendix A.

## 2. Sample Correlation Coefficient, Angle between Vectors and Their Normal Approximations

We slightly simplify notation. Let  $\vec{X}_m = (X_1, \dots, X_m)^T$  and  $\vec{Y}_m = (Y_1, \dots, Y_m)^T$  be two vectors from an  $m$ -dimensional normal distribution  $N(0, I_m)$  with zero mean, identity covariance matrix  $I_m$  and the sample correlation coefficient

$$R_m = R_m(\vec{X}_m, \vec{Y}_m) = \frac{\sum_{k=1}^m X_k Y_k}{\sqrt{\sum_{k=1}^m X_k^2 \sum_{k=1}^m Y_k^2}}. \quad (4)$$

Under the null hypothesis  $H_0: \{\vec{X}_m \text{ and } \vec{Y}_m \text{ are uncorrelated}\}$ , the so-called null density  $p_{R_m}(y; n)$  of  $R_m$  is given in Johnson, Kotz and Balakrishnan [15], Chapter 32, Formula (32.7):

$$p_{R_m}(y; m) = \frac{\Gamma((m-1)/2)}{\sqrt{\pi} \Gamma((m-2)/2)} (1 - y^2)^{(m-4)/2} \mathbb{I}_{(-1,1)}(y)$$

for  $m \geq 3$ , where  $\mathbb{I}_A(\cdot)$  denotes indicator function of a set  $A$ .

- Note  $\mu = \mathbb{E}R_m = 0$  and  $\sigma^2 = \text{Var}(R_m) = 1/(m-1)$  for  $m \geq 2$ ,
- $R_2$  is two point distributed with  $P(R_2 = -1) = P(R_2 = 1) = 1/2$ ,
- $R_3$  is  $U$ -shaped with  $p_{R_3}(y; 3) = (1/\pi) (1 - y^2)^{-1/2} \mathbb{I}_{(-1,1)}(y)$  and
- $R_4$  is uniform with density  $p_{R_4}(y; 4) = 1/2 \mathbb{I}_{(-1,1)}(y)$ .
- Moreover, for  $m \geq 5$ , the density function  $p_{R_m}(y; m)$  is unimodal.

Consider now the standardized correlation coefficient

$$\bar{R}_m = \sqrt{m-c} R_m \quad (5)$$

with some correcting real constant  $c < m$  having density

$$p_{\bar{R}_m}(y; m, c) = \frac{\Gamma((m-1)/2)}{\sqrt{m-c} \sqrt{\pi} \Gamma((m-2)/2)} \left(1 - \frac{y^2}{m-c}\right)^{(m-4)/2} \mathbb{I}_{\{|y| < \sqrt{m-c}\}}(y), \quad (6)$$

which converges with  $c = \mathcal{O}(1)$  as  $m \rightarrow \infty$  to the standard normal density

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad y \in (-\infty, \infty)$$

and by Konishi [16], Section 4, Formula (4.1) as  $m \rightarrow \infty$ :

$$F_m^*(x, c) := P\left(\sqrt{m-c} R_m \leq x\right) = \Phi(x) + \frac{x^3 + (2(c-1) - 3)x}{4(m-c)} \varphi(x) + \mathcal{O}(m^{-3/2}), \quad (7)$$

where  $\Phi(x) = \int_{-\infty}^x \varphi(y) dy$  is the standard normal distribution function. Note that in Konishi [16] the sample size (in our case the dimension of vectors) is  $m+1$  and  $c = 1 + 2\Delta$  with Konishi's correcting constant  $\Delta$ . Moreover, (7) follows from the more general Theorem 2.2 in the mentioned paper for independent components in the pairs  $(X_k, Y_k)$ ,  $k = 1, \dots, m$ .

In Christoph et al. [17], computable error bounds of approximations in (7) with  $c = 2$  and  $c = 2.5$  of order  $\mathcal{O}(m^{-2})$  for all  $m \geq 7$  are proved:

$$\sup_x \left| P\left(\sqrt{m-2.5} R_m \leq x\right) - \Phi(x) - \frac{x^3 \varphi(x)}{4(m-2.5)} \right| \leq \frac{B_m}{(m-2.5)^2} \leq \frac{B}{m^2} \quad (8)$$

and

$$\sup_x \left| P\left(\sqrt{m-2} R_m \leq x\right) - \Phi(x) - \frac{(x^3 - x) \varphi(x)}{4(m-2)} \right| \leq \frac{B_m^*}{(m-2)^2} \leq \frac{B^*}{m^2} \quad (9)$$

where for some  $m \geq 7$  constants  $B_m$  and  $B_m^*$  are calculated and presented in Table 1 in Christoph et al. [17]: i.e.,  $B_7 = 1.875$ ,  $B_7^* = 2.083$  and  $B_{50} = 0.720$ ,  $B_{50}^* = 0.982$ .

Usually, the asymptotic for  $\bar{R}_m$  is (9), where  $c = 2$  since it is related to the  $t$ -distributed statistic  $\sqrt{m-2} R_m / \sqrt{1-R_m^2}$ . With the correcting constant  $c = 2.5$ , one term in the asymptotic in (8) vanishes.

In order to use a transfer theorem from non-random to random dimension of the vectors, we prefer (7) with  $c = 0$ . In a similar manner as proving (8) and (9) in Christoph et al. [17], one can verify the following inequalities for  $m \geq 3$ :

$$\sup_x \left| P(\sqrt{m} R_m \leq x) - \Phi(x) - \frac{(x^3 - 5x)}{4m} \varphi(x) \right| \leq C_1 m^{-2}. \quad (10)$$

Let us consider now the connection between the correlation coefficient  $R_m$  and the angle  $\theta_m$  of the involved vectors  $\vec{X}_m, \vec{Y}_m$ :

$$\theta_m = \text{ang}(\vec{X}_m, \vec{Y}_m). \quad (11)$$

Hall et al. [1] showed that under the given conditions

$$\theta_m = \frac{1}{2}\pi + \mathcal{O}_p(m^{-1/2}) \quad \text{as } m \rightarrow \infty,$$

where  $\mathcal{O}_p$  denotes the stochastic order. Since

$$\cos \theta_m = \frac{\|\vec{X}_m\|^2 + \|\vec{Y}_m\|^2 - \|\vec{X}_m - \vec{Y}_m\|^2}{2 \|\vec{X}_m\| \|\vec{Y}_m\|} = R_m(\vec{X}_m, \vec{Y}_m) = R_m,$$

the computable error bounds for  $\theta_m$  follows from computable error bounds for  $R_m$ .

For any fixed constant  $c < m$ , and arbitrary  $x$  with  $|x| < \sqrt{m-c} \pi/2$ , we obtain for the angle  $\theta_m : 0 < \theta_m < \pi$ :

$$\begin{aligned} P(\sqrt{m-c}(\theta_m - \pi/2) \leq x) &= P(\theta_m \leq \pi/2 + x/\sqrt{m-c}) \\ &= P(\cos \theta_m \geq \cos(\pi/2 + x/\sqrt{m-c})) \\ &= P(R_m \geq -\sin(x/\sqrt{m-c})) \\ &= P(\sqrt{m-c} R_m \leq \sqrt{m-c} \sin(x/\sqrt{m-c})) \end{aligned} \quad (12)$$

because  $R_m$  is symmetric and  $P(R_m \leq x) = P(-R_m \leq x)$ .

Equation (12) shows the connection between the correlation coefficient  $R_m$  and the angle  $\theta_m$  among the vectors involved. In Christoph et al. [17], computable error bound of approximation in (8) are used to obtain similar bound for the approximation of the angle between two vectors, defined in (11). Here, the approximation (10) and (12) with  $c = 0$  lead for any  $m \geq 3$  and for  $|x| \leq \pi \sqrt{m}/2$  to

$$\sup_x \left| P(\sqrt{m}(\theta_m - \frac{\pi}{2}) \leq x) - \Phi(x) - \frac{(1/3)x^3 - 5x}{4m} \varphi(x) \right| \leq C_1 m^{-2}. \quad (13)$$

Many authors investigated limit theorems for the sums of random vectors when their dimension tends to infinity, see, e.g., Prokhorov [18]. In (6) and (7), the dimension  $m$  of the vectors  $\vec{X}_m$  and  $\vec{Y}_m$  tends to infinity.

Now, we consider the correlation coefficient of vectors  $\vec{X}_m$  and  $\vec{Y}_m$ , where the non-random dimension  $m$  is replaced by a random dimension  $N_n \in \mathbb{N}_+ = \{1, 2, \dots\}$  depending on some natural parameter  $n \in \mathbb{N}_+$  and  $N_n$  is independent of  $\vec{X}_m$  and  $\vec{Y}_m$  for any  $m, n \in \mathbb{N}_+$ . Define

$$R_{N_n} = \frac{\sum_{k=1}^{N_n} X_k Y_k}{\sqrt{\sum_{k=1}^{N_n} X_k^2 \sum_{k=1}^{N_n} Y_k^2}}.$$

### 3. Statistical Models with a Random Number of Observations

Let  $X_1, X_2, \dots \in \mathbb{R} = (-\infty, \infty)$  and  $N_1, N_2, \dots \in \mathbb{N}_+ = \{1, 2, \dots\}$  be random variables on the same probability space  $(\Omega, \mathbb{A}, \mathbb{P})$ . Let  $N_n$  be a random size of the underlying sample, i.e., the random number of observations, which depends on parameter  $n \in \mathbb{N}_+$ . We suppose for each  $n \in \mathbb{N}_+$  that  $N_n \in \mathbb{N}_+$  is independent of random variables  $X_1, X_2, \dots$  and  $N_n \rightarrow \infty$  in probability as  $n \rightarrow \infty$ . Let  $T_m := T_m(X_1, \dots, X_m)$  be some statistic of a sample with *non-random sample size*  $m \in \mathbb{N}_+$ . Define the random variable  $T_{N_n}$  for every  $n \in \mathbb{N}_+$ :

$$T_{N_n}(\omega) := T_{N_n(\omega)}(X_1(\omega), \dots, X_{N_n(\omega)}(\omega)), \quad \omega \in \Omega,$$

i.e.,  $T_{N_n}$  is some statistic obtained from a random sample  $X_1, X_2, \dots, X_{N_n}$ .

The randomness of the sample size may crucially change asymptotic properties of  $T_{N_n}$ , see, e.g., Gnedenko [7] or Gnedenko and Korolev [19].

#### 3.1. Random Sums

Many models lead to random sums and random means

$$S_{N_n} = \sum_{k=1}^{N_n} X_k \quad \text{and} \quad M_{N_n} = \frac{1}{N_n} \sum_{k=1}^{N_n} X_k. \quad (14)$$

A fundamental introduction to asymptotic distributions of random sums is given in Döbler [20].

It is worth mentioning that a suitable scaled factor by  $S_{N_n}$  affects the type of limit distribution. In fact, consider random sum  $S_{N_n}$  given in (14). For the sake of convenience, let  $X_1, X_2, \dots$  be independent standard normal random variables and  $N_n \in \mathbb{N}_+$  be geometrically distributed with  $E(N_n) = n$  and independent of  $X_1, X_2, \dots$ . Then, one has

$$\mathbb{P}\left(\frac{1}{\sqrt{N_n}} S_{N_n} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \quad \text{for all } n \in \mathbb{N}, \quad (15)$$

$$\mathbb{P}\left(\frac{1}{\sqrt{\mathbb{E}(N_n)}} S_{N_n} \leq x\right) \rightarrow \int_{-\infty}^x \frac{1}{\sqrt{2}} e^{-\sqrt{2}|u|} du \quad \text{as } n \rightarrow \infty, \quad (16)$$

$$\mathbb{P}\left(\frac{\sqrt{\mathbb{E}(N_n)}}{N_n} S_{N_n} \leq x\right) \rightarrow \int_{-\infty}^x (2+u^2)^{-3/2} du \quad \text{as } n \rightarrow \infty. \quad (17)$$

We have three different limit distributions. The suitable scaled geometric sum  $S_{N_n}$  is standard normal distributed or tends to the Laplace distribution with variance 1 depending on whether we take the random scaling factor  $1/\sqrt{N_n}$  or the non-random scaling factor  $1/\sqrt{\mathbb{E}N_n}$ , respectively. Moreover, we get the Student distribution with two degrees of freedom as the limit distribution if we use scaling with the mixed factor  $\sqrt{\mathbb{E}(N_n)}/N_n$ . Similar results also hold for the normalized random mean  $M_{N_n} = \frac{1}{N_n} S_{N_n}$ .

Assertion (15) is obtained by conditioning and the stability of the normal law. Moreover, using Stein's method, quantitative Berry–Esseen bounds in (15) and (16) for arbitrary centered random

variables  $X_1$  with  $\mathbb{E}(|X_1|^3) < \infty$  were proved in (Chen et al. [21], Theorem 10.6), (Döbler [20] Theorems 2.5 and 2.7) and (Pike and Ren [22] Theorem 3), respectively. Statement (17) follows from (Bening and Korolev [23] Theorem 2.1).

First order asymptotic expansions are obtained for the distribution function of random sample mean and random sample median constructed from a sample with two different random sizes in Bening et al. [24] and in the conference paper Bening et al. [25]. The authors make use of the rate of convergence of  $\mathbb{P}(N_n \leq g_n x)$  to the limit distribution  $H(x)$  with some  $g_n \uparrow \infty$ . In Christoph et al. [12], second order expansions for the normalized random sample sizes are proved, see below Propositions 1 and 2. These results allow for proving second order asymptotic expansions of random sample mean in Christoph et al. [12] and random sample median in Christoph et al. [13].

### 3.2. Transfer Proposition from Non-Random to Random Sample Sizes

Consider now the statistic  $T_{N_n} = T_{N_n}(\vec{X}_{N_n}, \vec{Y}_{N_n})$ , where the dimension of the vectors  $\vec{X}_{N_n}, \vec{Y}_{N_n}$  is a random number  $N_n \in \mathbb{N}_+$ .

In order to avoid too long expressions and at the same time to preserve a necessary accuracy, we limit ourselves to obtaining limit distributions and terms of order  $m^{-1}$  in the following non-asymptotic approximations with a bounds of order  $m^{-a}$  for some  $a > 1$ .

We suppose that the following condition on the statistic  $T_m = T_m(\vec{X}_m, \vec{Y}_m)$  with  $\mathbb{E}T_m = 0$  is met for a non-random sample size  $m \in \mathbb{N}_+$ :

**Condition 1.** There exist differentiable bounded function  $f_2(x)$  with  $\sup_x |x f_2'(x)| < c_0$  and real numbers  $a > 1, C_1 > 0$  such that for all integer  $m \geq 1$

$$\sup_x \left| \mathbb{P}(m^\gamma T_m \leq x) - \Phi(x) - m^{-1} f_2(x) \right| \leq C_1 m^{-a}, \quad (18)$$

where  $\gamma \in \{-1/2, 0, 1/2\}$ .

**Remark 1.** Relations (10) and (13) give the examples of statistics such that Condition 1 is met. For other examples of multivariate statistics of this kind, see Chapters 14–16 in Fujikoshi et al. [2].

Suppose that the limiting behavior of distribution functions of the normalized random size  $N_n \in \mathbb{N}_+$  is described by the following condition.

**Condition 2.** There exist a distribution function  $H(y)$  with  $H(0+) = 0$ , a function of bounded variation  $h_2(y)$ , a sequence  $0 < g_n \uparrow \infty$  and real numbers  $b > 0$  and  $C_2 > 0$  such that for all integer  $n \geq 1$

$$\left. \begin{aligned} \sup_{y \geq 0} |\mathbb{P}(g_n^{-1} N_n \leq y) - H(y)| &\leq C_2 n^{-b}, & 0 < b \leq 1 \\ \sup_{y \geq 0} |\mathbb{P}(g_n^{-1} N_n \leq y) - H(y) - n^{-1} h_2(y)| &\leq C_2 n^{-b}, & b > 1 \end{aligned} \right\} \quad (19)$$

**Remark 2.** In Propositions 1 and 2 below, we get the examples of discrete random variables  $N_n$  such that Condition 2 is met.

Conditions 1 and 2 allow us to construct asymptotic expansions for distributions of randomly normalized statistics on the base of approximation results for normalized fixed-size statistics (see relation (18)) and for the random size of the underlying sample (see relation (19)). As a result, we obtain the following transfer theorem.

**Theorem 1.** Let  $|\gamma| \leq K < \infty$  and both Conditions 1 and 2 be satisfied. Then, the following inequality holds for all  $n \in \mathbb{N}_+$ :

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( g_n^\gamma T_{N_n} \leq x \right) - G_n(x, 1/g_n) \right| \leq C_1 \mathbb{E} \left( N_n^{-a} \right) + (C_3 D_n + C_4) n^{-b}, \quad (20)$$

$$G_n(x, 1/g_n) = \int_{1/g_n}^{\infty} \left( \Phi(xy^\gamma) + \frac{f_2(xy^\gamma)}{g_n y} \right) d \left( H(y) + \frac{h_2(y)}{n} \right), \quad (21)$$

$$D_n = \sup_x \int_{1/g_n}^{\infty} \left| \frac{\partial}{\partial y} \left( \Phi(xy^\gamma) + \frac{f_2(xy^\gamma)}{y g_n} \right) \right| dy, \quad (22)$$

where  $a > 1, b > 0, f_2(z), h_2(y)$  are given in (18) and (19). The constants  $C_1, C_3, C_4$  do not depend on  $n$ .

**Remark 3.** Later, we use only the cases  $\gamma \in \{-1/2, 0, 1/2\}$ .

**Remark 4.** The domain  $[1/g_n, \infty)$  of integration in (21) depends on  $g_n$ . Thus, it is not clear how  $G_n(x, 1/g_n)$  is represented as a polynomial in  $g_n^{-1}$  and  $n^{-1}$ . To overcome this problem (see (26)), we prove the following theorem.

**Theorem 2.** Under the conditions of Theorem 1 and the additional conditions on functions  $H(\cdot)$  and  $h_2(\cdot)$ , depending on the convergence rate  $b$  in (19):

$$H(1/g_n) \leq c_1 g_n^{-b}, \quad b > 0, \quad (23)$$

$$\left. \begin{array}{l} i: \quad \int_0^{1/g_n} y^{-1} dH(y) \leq c_2 g_n^{-b+1}, \\ ii: \quad h_2(0) = 0 \quad \text{and} \quad |h_2(1/g_n)| \leq c_3 n g_n^{-b}, \\ iii: \quad \int_0^{1/g_n} y^{-1} |h_2(y)| dy \leq c_4 n g_n^{-b}, \end{array} \right\} \quad \text{for } b > 1, \quad (24)$$

we obtain for the function  $G_n(x, 1/g_n)$  defined in (21):

$$\sup_x |G_n(x, 1/g_n) - G_{n,2}(x)| \leq C g_n^{-b} + \sup_x \left( |I_1(x, n)| \mathbb{I}_{\{b < 1\}}(b) + |I_2(x, n)| \right) \quad (25)$$

with

$$G_{n,2}(x) = \begin{cases} \int_0^{\infty} \Phi(xy^\gamma) dH(y), & 0 < b < 1, \\ \int_0^{\infty} \Phi(xy^\gamma) dH(y) + \frac{1}{g_n} \int_0^{\infty} \frac{f_2(xy^\gamma)}{y} dH(y), & b = 1 \\ \int_0^{\infty} \Phi(xy^\gamma) dH(y) + \frac{1}{g_n} \int_0^{\infty} \frac{f_2(xy^\gamma)}{y} dH(y) \mathbb{I}_{\{\gamma=0\}}(\gamma) + \frac{1}{n} \int_0^{\infty} \Phi(xy^\gamma) dh_2(y), & b > 1. \end{cases} \quad (26)$$

$$I_1(x, n) = \int_{1/g_n}^{\infty} \frac{f_2(xy^\gamma)}{g_n y} dH(y) \quad \text{for } b \leq 1 \quad \text{and} \quad I_2(x, n) = \int_{1/g_n}^{\infty} \frac{f_2(xy^\gamma)}{g_n n y} dh_2(y) \quad \text{for } b > 1.$$

**Remark 5.** The additional conditions (23) and (24) guarantee to extend the integration range from  $[1/g_n, \infty)$  to  $(0, \infty)$  of the integrals in (26).

Theorems 1 and 2 are proved in Appendix A.1.

#### 4. Auxiliary Propositions and Lemmas

Consider the standardized correlation coefficient (5) having density (6) with correcting real constant  $c = 0$  and standardized angle  $\sqrt{m}(\theta_m - \pi/2)$ , see (12). By (10) and (13) for  $m \geq 3$ , we have

$$\sup_x \left| P \left( \sqrt{m} R_m \leq x \right) - \Phi(x) - \frac{(x^3 - 5x)}{4m} \varphi(x) \right| \leq C_1 m^{-2}, \quad m \in \mathbb{N}_+, \quad (27)$$



and for the angle  $\theta_m$  between the vectors for  $|x| \leq \pi \sqrt{m}/2$

$$\sup_x \left| P\left(\sqrt{m}(\theta_m - \frac{\pi}{2}) \leq x\right) - \Phi(x) - \frac{(1/3)x^3 - 5x}{4m} \varphi(x) \right| \leq C_1 m^{-2}, \quad m \in \mathbb{N}_+, \quad (28)$$

where (27) and (28) for  $m = 1$  and  $m = 2$  are trivial and  $C_1$  does not depend on  $m$ .

Suppose  $f_2(x; a) = (a x^3 - 5x) \varphi(x)/4$  with  $a = 1$  or  $a = 1/3$  when (27) or (28) are considered. Since a product of polynomials in  $x$  with  $\varphi(x)$  is always bounded, numerical calculus leads to

$$\sup_x |x f_2'(x; a)| = \sup_x |x (a x^4 - (3a + 5)x^2 + 5)| \varphi(x)/4 \leq 0.4.$$

Condition 1 of the transfer Theorem 1 to the statistics  $R_m$  and  $\theta_m$  are satisfied with  $c_0 = 0.4$  and  $a = 2$ .

Next, we estimate  $D_n(x)$  defined in (22).

**Lemma 1.** Let  $g_n$  a sequence with  $0 < g_n \uparrow \infty$  as  $n \rightarrow \infty$ . Then, with some  $0 < c(\gamma, a) < \infty$ , we obtain with  $a = 1$  and  $a = 1/3$ :

$$D_n = \sup_x \int_{1/g_n}^{\infty} \left| \frac{\partial}{\partial y} \left( \Phi(x y^\gamma) + \frac{f_2(x y^\gamma; a)}{y g_n} \right) \right| dy \leq \frac{1}{2} + \frac{c(\gamma, a)}{4}.$$

In the next subsection, we consider the cases when the random dimension  $N_n$  is negative binomial distributed with success probability  $1/n$ .

#### 4.1. Negative Binomial Distribution as Random Dimension of the Normal Vectors

Let the random dimension  $N_n(r)$  of the underlying normal vectors be negative binomial distributed (shifted by 1) with parameters  $1/n$  and  $r > 0$ , having probability mass function

$$\mathbb{P}(N_n(r) = j) = \frac{\Gamma(j+r-1)}{\Gamma(j)\Gamma(r)} \left(\frac{1}{n}\right)^r \left(1 - \frac{1}{n}\right)^{j-1}, \quad j = 1, 2, \dots \quad (29)$$

with  $\mathbb{E}(N_n(r)) = r(n-1) + 1$ . Then,  $\mathbb{P}(N_n(r)/g_n \leq x)$  tends to the Gamma distribution function  $G_{r,r}(x)$  with the shape and rate parameters  $r > 0$ , having density

$$g_{r,r}(x) = \frac{r^r}{\Gamma(r)} x^{r-1} e^{-rx} \mathbb{I}_{(0,\infty)}(x), \quad x \in \mathbb{R}. \quad (30)$$

If the statistic  $T_m$  is asymptotically normal, the limit distribution of the standardized statistic  $T_{N_n(r)}$  with random size  $N_n(r)$  is Student's  $t$ -distribution  $S_{2r}(x)$  having density

$$s_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi} \Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \quad \nu > 0, \quad x \in \mathbb{R}, \quad (31)$$

with  $\nu = 2r$ , see Bening and Korolev [23] or Schluter and Trede [26].

**Proposition 1.** Let  $r > 0$ , discrete random variable  $N_n(r)$  have probability mass function (29) and  $g_n := \mathbb{E}N_n(r) = r(n-1) + 1$ . For  $x > 0$  and all  $n \in \mathbb{N}$  there exists a real number  $C_2(r) > 0$  such that

$$\sup_{x \geq 0} \left| \mathbb{P}\left(\frac{N_n(r)}{g_n} \leq x\right) - G_{r,r}(x) - \frac{h_{2,r}(x)}{n} \right| \leq C_2(r) n^{-\min\{r, 2\}}, \quad (32)$$

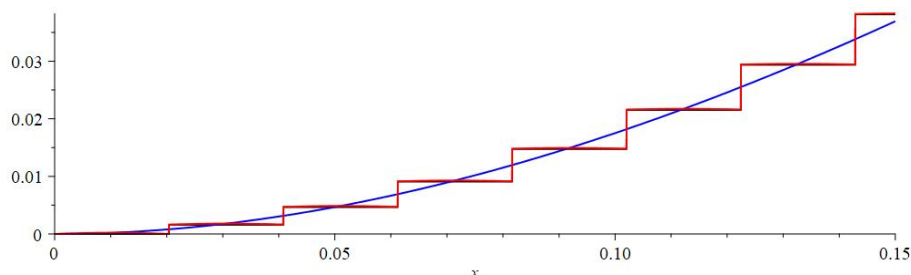
where

$$h_{2,r}(x) = \begin{cases} 0, & \text{for } r \leq 1, \\ \frac{g_{r,r}(x) ((x-1)(2-r) + 2Q_1(g_n x))}{2r}, & \text{for } r > 1. \end{cases}$$



$$Q_1(y) = 1/2 - (y - [y]) \text{ and } [.] \text{ denotes the integer part of a number.} \quad (33)$$

Figure 1 shows the approximation of  $\mathbb{P}(N_n(r) \leq (r(n-1) + 1)x)$  by  $G_{2,2}(x)$  and  $G_{2,2}(x) + h_2(x)/n$ .



**Figure 1.** Distribution function  $\mathbb{P}(N_n(r) \leq (r(n-1) + 1)x)$  (black line, almost covered by the red line), the limit law  $G_{2,2}(x)$  (blue line) and the second approximation  $G_{2,2}(x) + h_2(x)/n$  (red line) with  $n = 25$  and  $r = 2$ .

**Remark 6.** The convergence rate for  $r \leq 1$  is given in Bening et al. [24] or Gavrilenko et al. [27]. The Edgeworth expansion for  $r > 1$  is proved in Christoph et al. [12], Theorem 1. The jumps of the sample size  $N_n(r)$  have an effect only on the function  $Q_1(\cdot)$  in the term  $h_{2,r}(\cdot)$ .

The negative binomial random variable  $N_n$  satisfies Condition 2 of the transfer Theorem 1 with  $H(x) = G_{r,r}(x)$ ,  $h_2(x) = h_{2,r}(x)$ ,  $g_n = \mathbb{E}N_n(r) = r(n-1) + 1$  and  $b = \min\{r/2\}$ .

**Lemma 2.** In Theorem 2 the additional conditions (23) and (24) are satisfied with  $H(x) = G_{r,r}(x)$ ,  $h_2(x) = h_{2,r}(x)$ ,  $g_n = \mathbb{E}N_n(r) = r(n-1) + 1$  and  $b = \min\{r/2\}$ . Moreover, one has for  $\gamma \in \{-1/2, 0, 1/2\}$  and  $f_2(z; a) = (az^3 - 5z)\varphi(z)/4$ , with  $a = 1$  or  $a = 1/3$ :

$$|I_1(x, n)| = \begin{cases} \left| \int_{1/g_n}^{\infty} \frac{f_2(xy^\gamma; a)}{g_n y} dG_{r,r}(y) \right| \leq c_5 g_n^{-r} & r < 1, \\ \left| \int_{1/g_n}^{\infty} \frac{f_2(xy^\gamma; a)}{n y} dG_{1,1}(y) - n^{-1} f_2(x; a) \ln n \mathbb{I}_{\{\gamma=0\}}(\gamma) \right| \leq c_6 n^{-1}, & r = 1, \end{cases} \quad (34)$$

$$|I_2(x, n)| = \left| \int_{1/g_n}^{\infty} \frac{f_2(xy^\gamma; a)}{g_n n y} dh_{2,r}(y) \right| \leq \begin{cases} c_7 g_n^{-r}, & r > 1, r \neq 2, \\ (c_7 + c_8 \ln n \mathbb{I}_{\{\gamma=0\}}(\gamma)) g_n^{-2}, & r = 2. \end{cases} \quad (35)$$

Furthermore, we have

$$0 \leq g_n^{-1} - (rn)^{-1} \leq (r-1)(rn)^{-2} e^{-1/2} \quad \text{for } r > 1, \quad n \geq 2. \quad (36)$$

In addition to the expansion of  $N_n(r)$  a bound of  $\mathbb{E}(N_n(r))^{-a}$  is required, where  $m^{-a}$  is rate of convergence of Edgeworth expansion for  $T_m$ , see (18).

**Lemma 3.** Let  $r > 0$ ,  $\alpha > 0$  and the random variable  $N_n(r)$  is defined by (29). Then,

$$\mathbb{E}(N_n(r))^{-\alpha} \leq C(r) \begin{cases} n^{-\min\{r, \alpha\}}, & r \neq \alpha \\ \ln(n) n^{-\alpha}, & r = \alpha \end{cases} \quad (37)$$

and the convergence rate in case  $r = \alpha$  cannot be improved.

#### 4.2. Maximum of $n$ Independent Discrete Pareto Random Variables Is the Dimension of the Normal Vectors

Let  $Y(s) \in \mathbb{N}$  be discrete Pareto II distributed with parameter  $s > 0$ , having probability mass and distribution functions

$$\mathbb{P}(Y(s) = k) = \frac{s}{s+k-1} - \frac{s}{s+k} \quad \text{and} \quad \mathbb{P}(Y(s) \leq k) = \frac{k}{s+k}, \quad k \in \mathbb{N}, \quad (38)$$

which is a particular class of a general model of discrete Pareto distributions, obtained by discretization continuous Pareto II (Lomax) distributions on integers, see Buddana and Kozubowski [28].

Now, let  $Y_1(s), Y_2(s), \dots$ , be independent random variables with the same distribution (38). Define for  $n \in \mathbb{N}$  and  $s > 0$  the random variable

$$N_n(s) = \max_{1 \leq j \leq n} Y_j(s) \quad \text{with} \quad \mathbb{P}(N_n(s) \leq k) = \left( \frac{k}{s+k} \right)^n, \quad n \in \mathbb{N}. \quad (39)$$

It should be noted that the distribution of  $N_n(s)$  is extremely spread out on the positive integers. In Christoph et al. [12], the following Edgeworth expansion was proved:

**Proposition 2.** Let the discrete random variable  $N_n(s)$  have distribution function (39). For  $x > 0$ , fixed  $s > 0$  and all  $n \in \mathbb{N}$ , then there exists a real number  $C_3(s) > 0$  such that

$$\sup_{y>0} \left| \mathbb{P} \left( \frac{N_n(s)}{n} \leq y \right) - H_s(y) - \frac{h_{2;s}(y)}{n} \right| \leq \frac{C_3(s)}{n^2},$$

$$H_s(y) = e^{-s/y} \quad \text{and} \quad h_{2;s}(y) = s e^{-s/y} (s-1+2Q_1(ny)) / (2y^2), \quad y > 0 \quad (40)$$

where  $Q_1(y)$  is defined in (33).

**Remark 7.** The continuous function  $H_s(y) = e^{-s/y} \mathbb{I}_{(0,\infty)}(y)$  with parameter  $s > 0$  is the distribution function of the inverse exponential random variable  $W(s) = 1/V(s)$ , where  $V(s)$  is exponentially distributed with rate parameter  $s > 0$ . Both  $H_s(y)$  and  $\mathbb{P}(N_n(s) \leq y)$  are heavy tailed with shape parameter 1.

**Remark 8.** Therefore,  $\mathbb{E}(N_n(s)) = \infty$  for all  $n \in \mathbb{N}$  and  $\mathbb{E}(W(s)) = \infty$ . Moreover:

- First absolute pseudo moment  $v_1 = \int_0^\infty x |d(\mathbb{P}(N_n(s) \leq nx) - e^{-s/x})| = \infty$ ,
- Absolute difference moment  $\chi_u = \int_0^\infty x^{u-1} |\mathbb{P}(N_n(s) \leq nx) - e^{-s/x}| dx < \infty$  for  $1 \leq u < 2$ , see Lemma 2 in Christoph et al. [12].

On pseudo moments and some of their generalizations, see Chapter 2 in Christoph and Wolf [29].

**Lemma 4.** In Transfer Theorem 2, the additional conditions (23) and (24) are satisfied with  $H(y) = H_s(y) = e^{-s/y}$ ,  $h_2(y) = h_{2;s}(y) = s e^{-s/y} (s-1+2Q_1(ny)) / (2y^2)$ ,  $y > 0$ ,  $g_n = n$  and  $b = 2$ . Moreover, one has for  $|\gamma| \leq K < \infty$  and  $f_2(z; a) = (a z^3 - 5z) \varphi(z) / 4$ , with  $a = 1$  or  $a = 1/3$ :

$$I_2(x, n) = \left| \int_{1/n}^\infty \frac{f_2(x y^\gamma; a)}{n^2 y} d h_{2;s}(y) \right| \leq C(s) n^{-2}.$$

**Lemma 5.** For random size  $N_n(s)$  with probabilities (39) with reals  $s \geq s_0 > 0$  and arbitrary small  $s_0 > 0$  and  $n \geq 1$ , we have

$$\mathbb{E}(N_n(s))^{-\alpha} \leq C(s) n^{-\alpha}. \quad (41)$$

The Lemmas are proved in Appendix A.2.

## 5. Main Results

Consider the sample correlation coefficient  $R_m = R_m(\vec{X}_m, \vec{Y}_m)$ , given in (4) and the two statistics  $R_m^* = \sqrt{m} R_m$  and  $R_m^{**} = m R_m$  which differ from  $R_m$  by scaling factors. Hence, by (10),

$$\mathbb{P}(\sqrt{m} R_m \leq x) = \mathbb{P}(R_m^* \leq x) = \mathbb{P}\left(\frac{1}{\sqrt{m}} R_m^{**} \leq x\right) = \Phi(x) + \frac{(x^3 - 5x)}{4m} \varphi(x) + r(m) \quad (42)$$

with  $|r(m)| \leq Cm^{-2}$ .

Let  $\theta_m$  be the angle between the vectors  $\vec{X}_m$  and  $\vec{Y}_m$ . Contemplate the statistics  $\Theta_m = \theta_m - \pi/2$ ,  $\Theta_m^* = \sqrt{m}(\theta_m - \pi/2)$  and  $\Theta_m^{**} = m(\theta_m - \pi/2)$  which differ only in scaling. Then, by (13),

$$\mathbb{P}(\sqrt{m} \Theta_m \leq x) = \mathbb{P}(\Theta_m^* \leq x) = \mathbb{P}\left(\frac{1}{\sqrt{m}} \Theta_m^{**} \leq x\right) = \Phi(x) + \frac{((1/3)x^3 - 5x)}{4m} \varphi(x) + r^*(m)$$

with  $|r^*(m)| \leq Cm^{-2}$ .

Consider now the statistics  $R_{N_n}$ ,  $R_{N_n}^*$  and  $R_{N_n}^{**}$  as well as  $\Theta_{N_n}$ ,  $\Theta_{N_n}^*$  and  $\Theta_{N_n}^{**}$  when the vectors have random dimension  $N_n$ . The normalized statistics have different limit distributions as  $n \rightarrow \infty$ .

### 5.1. The Random Dimension $N_n = N_n(r)$ Is Negative Binomial Distributed

Let the random dimension  $N_n(r)$  be negative binomial distributed with probability mass function (29) and  $g_n = \mathbb{E}N_n(r) = r(n-1) + 1$ . "The negative binomial distribution is one of the two leading cases for count models, it accommodates the overdispersion typically observed in count data (which the Poisson model cannot)", see Schluter and Trede [26].

It follows from Theorems 1 and 2 and Proposition 1 that if limit distributions for  $\mathbb{P}(g_n^\gamma N_n(r)^{1/2-\gamma} R_{N_n(r)} \leq x)$  for  $\gamma \in \{1/2, 0, -1/2\}$  exist they are  $\int_0^\infty \Phi(xy^\gamma) dG_{r,r}(y)$  with densities given below in the proof of the corresponding theorems:

$$\frac{r^r}{\sqrt{2\pi}\Gamma(r)} \int_0^\infty y^{r-1/2} e^{-(xy^\gamma + ry)} dy = \begin{cases} s_{2r}(x) = \frac{\Gamma(r+1/2)}{\sqrt{2r\pi}\Gamma(r)} \left(1 + \frac{x^2}{2r}\right)^{-(r+1/2)}, & \gamma = 1/2, \\ \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, & \gamma = 0, \\ l_1(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, \text{ for } r = 1, & \gamma = -1/2, \end{cases} \quad (43)$$

where in case  $\gamma = -1/2$  for  $r \neq 1$  generalized Laplace distributions occur.

#### 5.1.1. Student's $t$ -Distribution

We start with the case  $\gamma = 1/2$  in Theorems 1 and 2. Consider the statistic  $\bar{R}_{N_n(r)} = \sqrt{g_n} R_{N_n(r)}$ . The limit distribution is the Student's  $t$ -distribution  $S_{2r}(x)$  with  $2r$  degrees of freedom with density (31).

**Theorem 3.** Let  $r > 0$  and (29) be the probability mass function of the random dimension  $N_n = N_n(r)$  of the vectors under consideration. If the representation (42) for the statistic  $R_m$  and the inequality (32) with  $g_n = \mathbb{E}N_n(r) = r(n-1) + 1$  hold, then there exists a constant  $C_r$  such that for all  $n \in \mathbb{N}_+$

$$\sup_x \left| \mathbb{P}\left(\sqrt{g_n} R_{N_n(r)} \leq x\right) - S_{2r,n}(x; 1) \right| \leq C_r \begin{cases} n^{-\min\{r, 2\}}, & r \neq 2, \\ \ln(n) n^{-2}, & r = 2, \end{cases} \quad (44)$$

where

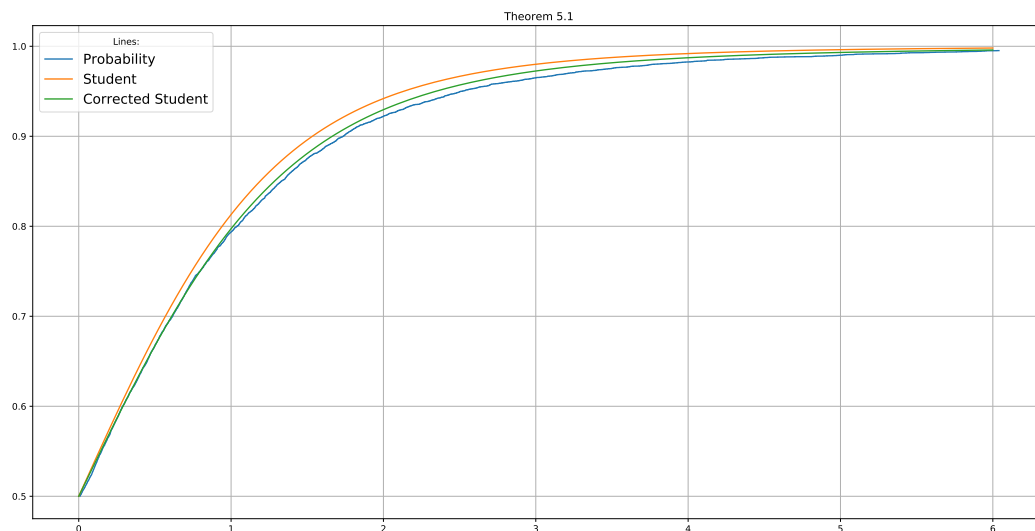
$$S_{2r,n}(x; a) = S_{2r}(x) + \frac{s_{2r}(x)}{rn} \left( ax^3 - \frac{10rx + 5x^3}{2r-1} + \frac{(2-r)(x^3 + x)}{4(2r-1)} \right). \quad (45)$$

Moreover, the scaled angle  $\theta_{N_n(r)}$  between the vectors  $\vec{X}_{N_n(r)}$  and  $\vec{Y}_{N_n(r)}$  allows the estimate

$$\sup_x \left| \mathbb{P} \left( \sqrt{g_n} (\theta_{N_n(r)} - \pi/2) \leq x \right) - S_{2r;n}(x; 1/3) \right| \leq C_r \begin{cases} n^{-\min\{r, 2\}}, & r \neq 2, \\ \ln(n) n^{-2}, & r = 2, \end{cases}$$

where  $S_{2r;n}(x; 1/3)$  is given in (45) with  $a = 1/3$ .

Figure 2 shows the advantage of the Chebyshev–Edgeworth expansion versus the limit law in approximating the empirical distribution function.



**Figure 2.** Empirical version of  $\mathbb{P} \left( \sqrt{g_n} R_{N_n(r)} \leq x \right)$  (blue line), limit Student law  $S_{2r}(x)$  (orange line) and second approximation  $S_{2r;n}(x; 1)$  (green line) for the correlation coefficient for pairs of normal vectors with random dimension  $N_{25}(2)$ . Here,  $x > 0, n = 25$  and  $r = 2$ .

**Remark 9.** The limit Student's  $t$ -distribution  $S_{2r}(x)$  is symmetric and a generalized hyperbolic distribution which can be written as a regularized incomplete beta function  $I_z(a, b)$ . For  $x > 0$ :

$$S_{2r}(x) = \int_{-\infty}^x s_{2r}(u) du = \frac{1}{2} \left( 1 + I_{2r/(x^2+2r)}(1/2, r) \right) \quad \text{and} \quad I_z(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z t^{a-1}(1-t)^{b-1}.$$

**Remark 10.** For integer values  $v = 2r \in \{1, 2, \dots\}$  the Student's  $t$ -distribution  $S_{2r}(x)$  is computable in closed form:

$$\begin{aligned} \text{the Cauchy law} \quad S_1(x) &= \frac{1}{2} + \frac{1}{\pi} \arctan(x), & S_2(x) &= \frac{1}{2} + \frac{x}{2\sqrt{2+x^2}}, \\ S_3(x) &= \frac{1}{2} + \frac{1}{\pi} \left( \frac{x}{\sqrt{3}(1+x^2/3)} + \arctan(x/\sqrt{3}) \right) & \text{and} \quad S_4(x) &= \frac{1}{2} + \frac{27(x^2+3)x(2x^2+9)}{8(3x^2+9)^{5/2}}. \end{aligned}$$

**Remark 11.** If the dimension of the vectors has the geometric distribution  $N_n(1)$ , then asymptotic distribution of the sample coefficient is the Student law  $S_2(x)$  with two degrees of freedom.

**Remark 12.** The Cauchy limit distribution occurs when the dimension of the vectors has distribution  $N_n(1/2)$ .

**Remark 13.** The Student's  $t$ -distributions  $S_{2r}(x)$  are heavy tailed and their moments of orders  $\alpha \geq 2r$  do not exist.

### 5.1.2. Standard Normal Distribution

Let  $\gamma = 0$  in the Theorems 1 and 2 examining the statistics  $R_{N_n(r)}^*$  and  $\Theta_{N_n(r)}^* = \sqrt{N_n(r)}(\theta_{N_n(r)} - \pi/2)$ .

**Theorem 4.** Let  $r > 0$  and  $N_n = N_n(r)$  be the random vector dimension having probability mass function (29). If the representation (42) for the statistic  $R_m$  and the inequality (32) with  $g_n = \mathbb{E}N_n(r) = r(n-1) + 1$  hold, then there exists a constant  $C_r$  such that for all  $n \in \mathbb{N}_+$

$$\sup_x \left| \mathbb{P} \left( \sqrt{N_n(r)} R_{N_n(r)} \leq x \right) - \Phi_{n,2}(x; 1) \right| \leq C_r \begin{cases} n^{-\min\{r, 2\}}, & r \neq 2, \\ \ln(n) n^{-2}, & r = 2, \end{cases} \quad (46)$$

where

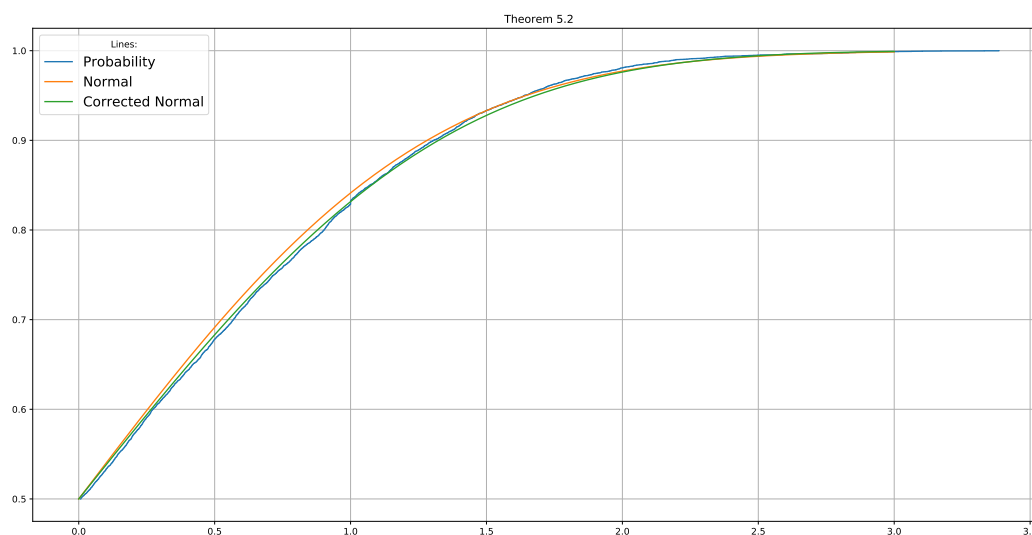
$$\Phi_{n,2}(x; a) = \Phi(x) + \frac{\varphi(x)}{n} \left( \frac{(ax^3 - 5x) \ln n}{4} \mathbb{I}_{\{r=1\}}(r) + \frac{\Gamma(r-1)(ax^3 - 5x)}{4\Gamma(r)} \mathbb{I}_{\{r>1\}}(r) \right). \quad (47)$$

Moreover, the scaled angle  $\theta_{N_n(r)}^*$  between the vectors  $\vec{X}_{N_n(r)}$  and  $\vec{Y}_{N_n(r)}$  allows the estimate

$$\sup_x \left| \mathbb{P} \left( \sqrt{N_n(r)} (\theta_{N_n(r)} - \pi/2) \leq x \right) - \Phi_{n,2}(x; 1/3) \right| \leq C_r \begin{cases} n^{-\min\{r, 2\}}, & r \neq 2, \\ \ln(n) n^{-2}, & r = 2, \end{cases}$$

where  $\Phi_{n,2}(x; 1/3)$  is given in (47) with  $a = 1/3$ .

Figure 3 shows that the second order Chebyshev–Edgeworth expansion approximates the empirical distribution function better than the limit normal distribution.



**Figure 3.** Empirical version of  $\mathbb{P} \left( \sqrt{N_n(r)} R_{N_n(r)} \leq x \right)$  (blue line), limit normal law  $\Phi(x)$  (orange line) and second approximation  $\Phi_{n,2}(x; 1)$  (green line) for the correlation coefficient for pairs of normal vectors with random dimension  $N_{25}(2)$ . Here,  $x > 0$ ,  $n = 25$  and  $r = 2$ .

**Remark 14.** When the distribution function of a statistic  $T_m$  without standardization tends to the standard normal distribution  $\Phi(x)$ , i.e.,  $\mathbb{P}(T_m \leq x) \rightarrow \Phi(x)$ , then the limit law for  $\mathbb{P}(T_{N_n} \leq x)$  remains the standard normal distribution  $\Phi(x)$ .

### 5.1.3. Generalized Laplace Distribution

Finally, we use  $\gamma = -1/2$  in Theorems 1 and 2 examining the statistic  $g_n^{-1/2} R_{N_n(r)}^{**}$ . Theorems 1 and 2 state that if there exists a limit distribution of  $\mathbb{P}\left(g_n^{-1/2} R_{N_n(r)}^{**} \leq x\right)$  as  $n \rightarrow \infty$  then it has to be a scale mixture of normal distributions with zero mean and gamma distribution:

$$L_r(x) = \int_0^\infty \Phi(xy^{-1/2}) dG_{r,r}(y)$$

having density, see formula (A9) in the proof of Theorem 5:

$$l_r(x) = \frac{r^r}{\Gamma(r)} \int_0^\infty \varphi(xy^{-1/2}) y^{r-3/2} e^{-ry} dy = \frac{2r^r}{\Gamma(r)\sqrt{2\pi}} \left(\frac{|x|}{\sqrt{2r}}\right)^{r-1/2} K_{r-1/2}(\sqrt{2r}|x|). \quad (48)$$

where  $K_\alpha(u)$  is the  $\alpha$ -order Macdonald function or  $\alpha$ -order modified Bessel function of the third kind. See, e.g., Oldham et al. [30], Chapter 51, or Kotz et al. [31], Appendix, for properties of these functions.

For integer  $r = 1, 2, 3, \dots$  these densities  $l_r(x)$ , so-called *Sargan densities*, and their distribution functions are computable in closed forms:

$$\left. \begin{aligned} l_1(x) &= \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} & \text{and } L_1(x) &= 1 - \frac{1}{2} e^{-\sqrt{2}|x|}, \quad x > 0 \\ l_2(x) &= \left(\frac{1}{2} + |x|\right) e^{-2|x|} & \text{and } L_2(x) &= 1 - \frac{1}{2}(1+x) e^{-2|x|}, \quad x > 0 \\ l_3(x) &= \frac{3\sqrt{6}}{16} \left(1 + \sqrt{6}|x| + 2x^2\right) e^{-\sqrt{6}|x|} & \text{and } L_3(x) &= 1 - \left(\frac{1}{2} + \frac{5\sqrt{6}x}{16} + \frac{3x^2}{8}\right) e^{-\sqrt{6}|x|}, \end{aligned} \right\} \quad (49)$$

where  $L_r(-x) = 1 - L_r(x)$  for  $x \geq 0$ .

The standard Laplace distribution is  $L_1(x)$  with variance 1 and density  $l_1(x)$  given in (49). Therefore, Sargan distributions are a kind of generalizations of the standard Laplace distribution.

**Theorem 5.** Let  $r = 1, 2, 3$  and (29) be probability mass function of the random dimension  $N_n = N_n(r)$  of the vectors under consideration. If the representation (42) for the statistic  $R_m$  and the inequality (32) for  $N_n(r)$  with  $g_n = \mathbb{E}N_n(r) = r(n-1) + 1$  hold, then there exists a constant  $C_r$  such that for all  $n \in \mathbb{N}_+$

$$\sup_x \left| \mathbb{P}\left(g_n^{-1/2} N_n(r) R_{N_n(r)} \leq x\right) - L_{n,2}(x; 1) \right| \leq C_r \begin{cases} n^{-\min\{r, 2\}}, & r \neq 2, \\ \ln(n) n^{-2}, & r = 2, \end{cases} \quad (50)$$

where

$$L_{n,2}(x; a) = \begin{cases} L_1(x), & r = 1, \\ L_2(x) + \frac{ax|x| - 5x\sqrt{2}}{2(n-1)+1} e^{-2|x|}, & r = 2, \\ L_3(x) + \frac{27}{24(n-1)+8} \left( \frac{ax^3}{\sqrt{2}} - \frac{5x|x|}{6} - \frac{5x}{6\sqrt{6}} \right) e^{-\sqrt{6}|x|} \\ \quad + \frac{9x}{2n} \left( \frac{1}{12\sqrt{6}} + \frac{|x|}{12} - \frac{x^2}{6\sqrt{6}} \right) e^{-\sqrt{6}|x|}, & r = 3. \end{cases} \quad (51)$$

For arbitrary  $r > 0$ , the approximation rate is given by:

$$\sup_x \left| \mathbb{P}\left(g_n^{-1/2} N_n(r) R_{N_n(r)} \leq x\right) - L_r(x) \right| \leq C_r n^{-\min\{r, 1\}}.$$

Moreover, the scaled angle  $N_n(r) \theta_{N_n(r)}$  between the vectors  $\vec{X}_{N_n(r)}$  and  $\vec{Y}_{N_n(r)}$  allows the estimate

$$\sup_x \left| \mathbb{P}\left(g_n^{-1/2} N_n(r) \theta_{N_n(r)} \leq x\right) - L_{n,2}(x; 1/3) \right| \leq C_r \begin{cases} n^{-\min\{r, 2\}}, & r \neq 2, \\ \ln(n) n^{-2}, & r = 2, \end{cases}$$

where  $L_{n;2}(x;1/3)$  is given in (51) with  $a = 1/3$ .

Figure 4 shows that the Chebyshev–Edgeworth expansion approaches the empirical distribution function better than the limit Laplace law.

**Remark 15.** One can find the distribution functions  $L_r(x)$  for arbitrary  $r > 0$  with formula 1.12.1.3 in Prudnikov et al. [32]:

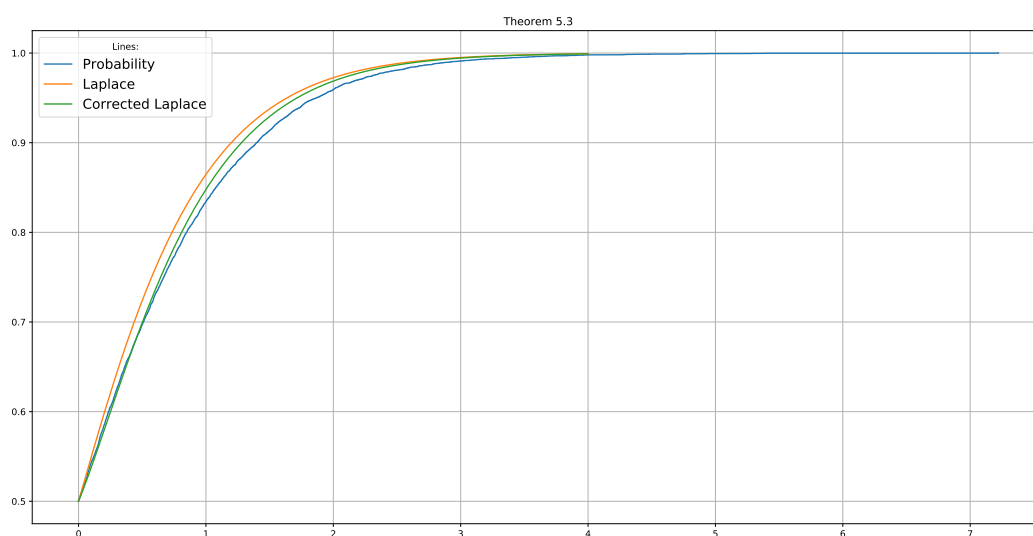
$$\begin{aligned} L_r(x) &= \frac{1}{2} + \frac{2r^r}{\sqrt{2\pi}\Gamma(r)} \int_0^x \left(\frac{|x|}{2r}\right)^{r-1/2} K_{r-1/2}(\sqrt{2r}|x|) dx \\ &= \frac{1}{2} + \frac{x}{(2r)^{(r-1/2)/2}} \left( K_{r-1/2}(\sqrt{2rx}) L_{r-3/2}(\sqrt{2rx}) + K_{r-3/2}(\sqrt{2rx}) L_{r-1/2}(\sqrt{2rx}) \right). \end{aligned}$$

where  $L_\alpha(x)$  are the modified Struve functions of order  $\alpha$ , for properties of modified Struve functions see, e.g., Oldham et al. [30], Section 57:13.

**Remark 16.** The function (48) as density of a mixture of normal distributions with zero mean and random variance  $W_r$  having gamma distribution  $\mathbb{P}(W_r \leq y) = G_{r,r}(y)$  is given also in Kotz et al. [31], Formula (4.1.32) with  $\tau = r$ ,  $\sigma = 1/\sqrt{r}$ , using Formula (A.0.4) with  $\lambda = -r + 3/2$  and the order-reflection formula  $K_{-\alpha}(x) = K_\alpha(x)$ . Such a variance gamma model is studied in Madan and Senata [33] for share market returns.

**Remark 17.** A systematic exposition about the Laplace distribution and its numerous generalization and diverse applications one finds in the useful and interesting monography by Kotz et al. [31]. Here, these generalized Laplace distributions  $L_1(x)$ ,  $L_2(x)$  and  $L_3(x)$  are the leading terms in the approximations of the sample correlation coefficient  $R_{N_n(r)}^{**}$  of two Gaussian vectors with negative binomial distributed random dimension  $N_n(r)$  and the angle  $\theta_{N_n(r)}^{**}$  between these vectors.

**Remark 18.** In Goldfeld and Quandt [34] and Missiakoulis [35] Sargans densities  $l_r(x)$  and distribution functions  $L_r(x)$  for arbitrary integer  $r = 1, 2, 3, \dots$  have been studied as an alternative to normal law in econometric models because they are computable in closed form, see also Kotz et al. [31], Section 4.4.3 and the references therein.



**Figure 4.** Empirical version of  $\mathbb{P}\left(g_n^{-1/2} N_n(r) R_{N_n(r)} \leq x\right)$  (blue line), limit Laplace law  $L_r(x)$  (orange line) and second approximation  $L_{n;2}(x;1)$  (green line) for the correlation coefficient for pairs of normal vectors with random dimension  $N_{25}(2)$ . Here,  $x > 0$ ,  $n = 25$  and  $r = 2$ .



### 5.2. The Random Dimension $N_n = N_n(s)$ Is the Maximum of $n$ Independent Discrete Pareto Random Variables

The random dimension  $N_n(s)$  has probability mass function (39): Since  $\mathbb{E}N_n(s) = \infty$  we choose  $g_n = n$  and consider again the cases  $\gamma = 1/2$ ,  $\gamma = 0$  and  $\gamma = -1/2$ .

It follows from Theorems 1 and 2 and Proposition 2 that if limit distributions for  $\mathbb{P}(g_n^\gamma R_{N_n(s)} \leq x)$  for  $\gamma \in \{1/2, 0, -1/2\}$  exist, they are  $\int_0^\infty \Phi(xy^\gamma) dH_s(y)$  with densities given below in the proof of the corresponding theorems

$$\frac{s}{\sqrt{2\pi}} \int_0^\infty y^{-3/2} e^{-(x^2 y^{2\gamma/2} + s/y)} dy = \begin{cases} l_{1/\sqrt{s}}(x) = \frac{\sqrt{2s}}{2} e^{-\sqrt{2s}|x|}, & \gamma = 1/2, \\ \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, & \gamma = 0, \\ s_2^*(x; \sqrt{s}) = \frac{1}{2\sqrt{2s}} \left(1 + \frac{x^2}{2s}\right)^{-3/2}, & \gamma = -1/2, \end{cases} \quad (52)$$

where  $s_2^*(x; \sqrt{s})$  is the density of the scaled Student's  $t$ -distribution  $S_2^*(x; \sqrt{s})$  with 2 degrees of freedom, see Definition B37 in Jackman [36], p.507. If  $Z$  has density  $s_2^*(x; \sqrt{s})$  then  $Z/\sqrt{s}$  has a classic Student's  $t$ -distribution with two degrees of freedom.

#### 5.2.1. Laplace Distribution

We start with the case  $\gamma = 1/2$  in Theorems 1 and 2. Consider the statistics  $\sqrt{n} R_{N_n(s)}$  and  $\sqrt{n}(\theta_{N_n(s)} - \pi/2)$ . The limit distribution is now the Laplace distribution

$$L_{1/\sqrt{s}}(x) = \frac{1}{2} + \frac{1}{2} \text{sign}(x) \left(1 - e^{-\sqrt{2s}|x|}\right) \quad \text{with density} \quad l_{1/\sqrt{s}}(x) = \frac{\sqrt{2s}}{2} e^{-\sqrt{2s}|x|}.$$

**Theorem 6.** Let  $s > 0$  and (39) be the probability mass function of the random dimension  $N_n = N_n(s)$  of the vectors under consideration. If the representation (42) for the statistic  $R_m$  and the inequality (32) with  $g_n = n$  hold, then there exists a constant  $C_s$  such that for all  $n \in \mathbb{N}_+$

$$\sup_x \left| \mathbb{P} \left( \sqrt{n} R_{N_n(s)} \leq x \right) - L_{1/\sqrt{s};n}(x; a) \right| \leq C_s n^{-2},$$

where

$$L_{1/\sqrt{s};n}(x; a) = L_{1/\sqrt{s}}(x) + \frac{l_{1/\sqrt{s}}(x)}{8sn} \left( a 2s x^3 - (4-s)x \left(1 + \sqrt{2s}|x|\right) \right). \quad (53)$$

Moreover, the scaled angle  $\theta_{N_n(s)}$  between the vectors  $\vec{X}_{N_n(s)}$  and  $\vec{Y}_{N_n(s)}$  allows the estimate

$$\sup_x \left| \mathbb{P} \left( \sqrt{n} (\theta_{N_n(s)} - \pi/2) \leq x \right) - L_{1/\sqrt{s};n}(x; 1/3) \right| \leq C_s n^{-2},$$

where  $L_{1/\sqrt{s};n}(x; 1/3)$  is given in (53) with  $a = 1/3$ .

#### 5.2.2. Standard Normal Distribution

Let  $\gamma = 0$  in the Theorems 1 and 2 examine the statistics  $R_{N_n(s)}^*$  and  $\Theta_{N_n(s)}^* = \sqrt{N_n(s)}(\theta_{N_n(s)} - \pi/2)$ .

**Theorem 7.** Let  $s > 0$  and  $N_n = N_n(s)$  be the random vector dimension having probability mass function (39). If the representation (42) for the statistic  $R_m$  and the inequality (32) with  $g_n = n$  hold, then there exists a constant  $C_s$  such that, for all  $n \in \mathbb{N}_+$

$$\sup_x \left| \mathbb{P} \left( \sqrt{N_n(s)} R_{N_n(s)}^* \leq x \right) - \Phi(x) - \frac{1}{4n} \varphi(x) s^2 (x^3 - 5x) \right| \leq C_s n^{-2},$$

Moreover, the scaled angle  $\theta_{N_n(s)}^*$  between the vectors  $\vec{X}_{N_n(s)}$  and  $\vec{Y}_{N_n(s)}$  allows the estimate

$$\sup_x \left| \mathbb{P} \left( \sqrt{N_n(s)} (\theta_{N_n(s)} - \pi/2) \leq x \right) - \Phi(x) - \frac{1}{4n} \varphi(x) s^2 \left( \frac{1}{3} x^3 - 5x \right) \right| \leq C_s n^{-2}.$$

### 5.2.3. Scaled Student's $t$ -Distribution

Finally, we use  $\gamma = -1/2$  in Theorems 1 and 2 examining the statistics  $n^{-1/2} N_n(s) R_{N_n(s)}$  and  $n^{-1/2} N_n(s) (\theta_{N_n(s)} - \pi/2)$ . The limit Scaled Student's  $t$ -Distribution  $S_2^*(x; \sqrt{s})$  with two degrees of freedom is a scale mixture of the normal distribution with zero mean and mixing exponential distribution  $1 - e^{-sy}$ ,  $y \geq 0$ , and it is representable in a closed form, see (A15) below in the proof of Theorem 8:

$$\begin{aligned} \int_0^\infty \Phi(x/\sqrt{y}) d e^{-s/y} &= \int_0^\infty \Phi(x/\sqrt{y}) s y^{-2} e^{-s/y} dy = \int_0^\infty \Phi(x\sqrt{z}) s e^{-sz} dz \\ &= \int_0^\infty \Phi(x\sqrt{z}) d(1 - e^{-sz}) = \frac{1}{2} + \frac{x/\sqrt{s}}{2\sqrt{2}\sqrt{1+x^2/(2s)}} = S_2^*(x) \end{aligned}$$

**Theorem 8.** Let  $s > 0$  and  $N_n = N_n(s)$  be the random vector dimension having probability mass function (39). If the representation (42) for the statistic  $R_m$  and the inequality (32) with  $g_n = n$  hold, then there exists a constant  $C_s$  such that for all  $n \in \mathbb{N}_+$

$$\sup_x \left| \mathbb{P} \left( n^{-1/2} N_n(s) R_{N_n(s)} \leq x \right) - S_{n,2}^*(x; 1) \right| \leq C_r n^{-2}, \quad (54)$$

where

$$S_{n,2}^*(x; \sqrt{s}; a) = S_2^*(x; \sqrt{s}) + \frac{(15a + 3s - 18)x^3 - 6xs(6-s)}{4n(x^2 + 2s)^2} s_2^*(x; \sqrt{s}) \quad (55)$$

Moreover, the scaled angle  $\theta_{N_n(s)}^*$  between the vectors  $\vec{X}_{N_n(s)}$  and  $\vec{Y}_{N_n(s)}$  allows the estimate

$$\sup_x \left| \mathbb{P} \left( n^{-1/2} N_n(s) \theta_{N_n(s)} \leq x \right) - S_{n,2}(x; \sqrt{s}; 1/3) \right| \leq C_s n^{-2},$$

where  $S_{n,2}(x; \sqrt{s}; 1/3)$  is given in (55) with  $a = 1/3$ .

Theorems 3 to 8 are proved in Appendix A.3.

## 6. Conclusions

The asymptotic distributions of the sample correlation coefficient of vectors with random dimensions are normal scale mixtures. From (43) and (52), one can conclude that random dimension and corresponding scaling have significant influence on limit distributions. A scale mixture of a normal distribution change the tail behavior of the distribution. Students  $t$ -Distributions have polynomial tails, as one class of heavy-tailed distributions, they can be used to model heavy-tail returns data in finance. The Laplace distributions have heavier tails than normal distributions.

**Author Contributions:** Conceptualization, G.C. and V.V.U.; methodology, V.V.U. and G.C.; writing—original draft, G.C. and V.V.U.; writing—review and editing, V.V.U. and G.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding. It was done within the framework of the Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, and HSE University Basic Research Programs.

**Acknowledgments:** The authors would like to thank the Managing Editor and the Reviewers for the careful reading of the manuscript and pertinent comments. Their constructive feedback helped to improve the quality of this work and shape its final form.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proofs of the Theorems and Lemmas

### Appendix A.1. Proofs of Theorems 1 and 2

**Proof of Theorem 1.** The proof follows along the similar arguments of the more general transfer Theorem 3.1 in Bening et al. [24]. Since in Theorem 3.1 in Bening et al. [24] the constant  $\gamma$  has to be non-negative and in our Theorem 1, we also need  $\gamma = -1/2$ , therefore we repeat the proof. Conditioning on  $N_n$ , we have

$$\mathbb{P}(g_n^\gamma T_{N_n} \leq x) = \mathbb{P}(N_n^\gamma T_{N_n} \leq x(N_n/g_n)^\gamma) = \sum_{m=1}^{\infty} \mathbb{P}(m^\gamma T_m \leq x(m/g_n)^\gamma) \mathbb{P}(N_n = m).$$

Using now (18) with  $\Phi_m(z) := \Phi(z) + m^{-1}f_2(z)$ , we find

$$\begin{aligned} \sum_{m=1}^{\infty} \left| \mathbb{P}(m^\gamma T_m \leq x(m/g_n)^\gamma) - \Phi_m(x(m/g_n)^\gamma) \right| \mathbb{P}(N_n = m) \\ \stackrel{(18)}{\leq} C_1 \sum_{m=1}^{\infty} m^{-a} \mathbb{P}(N_n = m) = C_1 \mathbb{E}(N_n^{-a}). \end{aligned} \quad (\text{A1})$$

Taking into account  $\mathbb{P}(N_n/g_n < 1/g_n) = \mathbb{P}(N_n < 1) = 0$ , we obtain

$$\begin{aligned} \sum_{m=1}^{\infty} \Phi_m(x(m/g_n)^\gamma) \mathbb{P}(N_n = m) &= \mathbb{E}(\Phi_{N_n}(x(N_n/g_n)^\gamma)) \\ &= \int_{1/g_n}^{\infty} \Delta_n(x, y) d\mathbb{P}\left(\frac{N_n}{g_n} \leq y\right) = G_n(x, 1/g_n) + I_1, \end{aligned}$$

where  $\Delta_n(x, y) := \Phi(xy^\gamma) + f_2(xy^\gamma)/(g_n y)$ ,  $G_n(x, 1/g_n)$  is defined in (21) and

$$I_1 = \int_{1/g_n}^{\infty} \Delta_n(x, y) d\left(\mathbb{P}\left(\frac{N_n}{g_n} \leq y\right) - H(y) - \frac{h_2(y) \mathbb{I}_{\{b>1\}}(b)}{n}\right).$$

Estimating integral  $I_1$ , we use the integration by parts for Lebesgue–Stieltjes integrals, the boundedness of  $f_2(z)$ , say  $\sup_z |f_2(z)| \leq c_1^*$ , and estimates (19)

$$\begin{aligned} |I_1| &\leq \sup_x \lim_{L \rightarrow \infty} |\Delta_n(x, y)| \left| \mathbb{P}(N_n/g_n \leq y) - H(y) - n^{-1}h_2(y) \mathbb{I}_{\{b>1\}}(b) \right| \Big|_{y=1/g_n}^{y=L} \\ &\quad + \sup_x \int_{1/g_n}^{\infty} \left| \frac{\partial}{\partial y} \Delta_n(x, y) \right| \left| \mathbb{P}(N_n/g_n \leq y) - H(y) - n^{-1}h_2(y) \mathbb{I}_{\{b>1\}}(b) \right| dy \\ &\leq (1 + c_1^*) C_2 n^{-b} + C_2 D_n n^{-b}, \end{aligned}$$

where  $D_n$  is defined in (22). Together with (A1), we obtain (20) and Theorem 1 is proved.  $\square$

**Proof of Theorem 2.** Using (23), we find for  $b > 0$

$$\int_0^{1/g_n} \Phi(xy^\gamma) dH(y) \leq \int_0^{1/g_n} dH(y) = H(1/\sqrt{g_n}) - H(0) \stackrel{(23)}{\leq} c_1 g_n^{-b}.$$

Let now  $b > 1$ . Since  $f_2(z)$  is supposed to be bounded, it follows from  $|f_2(z)| \leq c_1^* < \infty$  and (24i) that

$$\int_0^{1/g_n} |f_2(xy^\gamma)| y^{-1} dH(y) \leq c_1^* \int_0^{1/g_n} y^{-1} dH(y) \stackrel{(24i)}{\leq} c_1^* c_2 g_n^{-b+1}.$$

Integration by parts,  $|z|\varphi(z) \leq c^* = (2\pi e)^{-1/2}$ , (24ii) and (24iii) lead to

$$\left| \int_0^{1/g_n} \Phi(xy^\gamma) dh_2(y) \right| \leq |h_2(1/g_n)| + \gamma c^* \int_0^{1/g_n} y^{-1} |h_2(y)| dy \leq (c_3 + \gamma c^* c_4) n g_n^{-b}.$$

Theorem 2 is proved.  $\square$

## Appendix A.2. Proofs of Lemmas 1 to 5

**Proof of Lemma 1.** To estimate  $D_n$ , we consider three cases:

$$D_n = \sup_x |D_n(x)| = \max\{\sup_{x>0} |D_n(x)|, \sup_{x<0} |D_n(x)|, |D_n(0)|\}.$$

Let  $x > 0$ . Since  $\frac{\partial}{\partial y} \Phi(x y^\gamma) = \gamma x y^{\gamma-1} \varphi(x y^\gamma) \geq 0$ , we find

$$\int_{1/g_n}^{\infty} \left| \frac{\partial}{\partial y} \Phi(x y^\gamma) \right| dy = \int_{1/g_n}^{\infty} \gamma x y^{\gamma-1} \varphi(x y^\gamma) dy = \int_{x g_n^{-\gamma}}^{\infty} \varphi(u) du = \Phi(\infty) - \Phi(x g_n^{-\gamma}) \leq 1/2.$$

Consider now  $f_2(x y^\gamma; a) = (a(x y^\gamma)^3 - 5 x y^\gamma) \varphi(x y^\gamma) / 4$  with  $a = 1$  or  $a = 1/3$ . Then,

$$\frac{\partial}{\partial y} \left( \frac{f_2(x y^\gamma; a)}{y} \right) = \frac{Q_5(x y^\gamma; a)}{4 y^2}, \quad Q_5(z; a) = -(\gamma a z^5 - ((3a + 5) \gamma - a) z^3 + 5(\gamma - 1) z) \varphi(z). \quad (\text{A2})$$

Since  $\sup_z |Q_5(z; a)| \leq c(\gamma; a) < \infty$  and  $g_n^{-1} \int_{1/g_n}^{\infty} y^{-2} dy = 1$ , inequality (29) holds for  $x > 0$ . Taking into account  $|D_n(x)| = |D_n(-x)|$  and  $D_n(0) = 0$ , Lemma 1 is proved.  $\square$

**Proof of Lemma 2.** Using (30), we find  $G_{r,r}(1/g_n) \leq c_1 g_n^{-r}$  with  $c_1 = r^{r-1} / \Gamma(r)$ . For  $r > 1$ , then  $\int_0^{1/g_n} y^{-1} dG_{r,r}(y) \leq c_2 g_n^{-r+1}$  with  $c_2 = r^r / ((r-1)\Gamma(r))$ . Since  $g_{r,r}(0) = 0$ ,  $h_{2,r}(0) = 0$  and  $g_n \leq r n$  for  $r > 1$ , then (24ii) and (24iii) hold with  $c_3 = c_r^*$  and  $c_4 = c_r^* / (r-1)$ , where  $c_r^* = \frac{r^r}{2r\Gamma(r)} \sup_y \{e^{-ry} (|y-1||2-r|+1)\} < \infty$ .

It remains to prove the bounds in (34) and (35). Let first  $r < 1$ . With  $c_1^* = \sup_z |f_2(z; a)|$ , we find

$$|I_1(x, n)| \leq \frac{c_1^* r^r}{g_n \Gamma(r)} \int_{1/g_n}^{\infty} y^{r-2} dy \leq \frac{c_1^* r^r}{(r-1)\Gamma(r)} g_n^{-r} \quad \text{with} \quad c_5 = \frac{c_1^* r^r}{(r-1)\Gamma(r)}.$$

If  $r = 1$  with  $c_1^{**} = \sup_z \{|a z^2 - 5| \varphi(z/\sqrt{2})\}$ , we find  $|f_2(z; a)| \leq c_1^{**} |z| \varphi(z/\sqrt{2})$  and

$$|I_1(x, n)| \leq \frac{c_1^{**} |x|}{\sqrt{2} \pi n} \int_{1/n}^{\infty} y^{\gamma-1} e^{-(y+x^2 y^2 \gamma/4)} dy \quad \text{with} \quad \gamma \in \{-1/2, 0, 1/2\}.$$

For  $\gamma = 1/2$  using  $|x| (1 + x^2/4)^{-1/2} \leq 2$ , we obtain

$$|I_1(x, n)| \leq \frac{c_1^{**} |x|}{\sqrt{2} \pi n} \int_{1/n}^{\infty} y^{1/2-1} e^{-(1+x^2/4)y} dy \leq \frac{c_1^{**} |x| \Gamma(1/2)}{\sqrt{2} \pi (1 + x^2/4)^{1/2}} n^{-1} \leq c_6 n^{-1}, \quad c_6 = \sqrt{2} c_1^{**}.$$

If  $\gamma = -1/2$ , then Prudnikov et al. [37], formula 2.3.16.3, for  $x \neq 0$  leads to

$$I_1(x, n) \leq \frac{c_1^{**} |x|}{\sqrt{2} \pi n} \int_{1/n}^{\infty} y^{-1-1/2} e^{-(2y+x^2/(4y))} dy \leq \frac{c_1^{**} |x|}{\sqrt{2} \pi n} \frac{2\sqrt{\pi}}{|x|} e^{-\sqrt{2}|x|} \leq \frac{\sqrt{2} c_1^{**}}{n}, \quad c_6 = \sqrt{2} c_1^{**}.$$

Finally, if  $\gamma = 0$ , then  $f_2(x y^\gamma; a) = f_2(x; a)$  does not depend on  $y$ . Using now

$$0 \leq \ln n - \int_{1/n}^1 y^{-1} dG_{1,1}(y) = \int_{1/n}^1 \frac{1 - e^{-y}}{y} dy \leq 1 \quad \text{and} \quad \int_1^{\infty} y^{-1} dG_{1,1}(y) \leq e^{-1},$$

then (34) for  $r = 1$  holds with  $c_6 = c_1^* (1 + e^{-1})$ .

Let  $r > 1$ . Integration by parts for Lebesgue–Stieltjes integrals in  $I_2(x, n)$  in (35) and (A2) lead to

$$I_2(x, n) \leq \frac{1}{n g_n} \left( c_1^* g_n |h_{2,r}(1/g_n)| + \int_{1/g_n}^{\infty} \frac{|Q_5(x y^\gamma; a)|}{4 y^2} |h_{2,r}(y)| dy \right). \quad (\text{A3})$$

Since  $c(\gamma; a) = \sup_z |Q_5(z; a)| < \infty$  and with above defined  $c_r^*$ , we find

$$\int_{1/g_n}^{\infty} \frac{|h_{2,r}(y)|}{y^2} dy \leq c_r^* \int_{1/g_n}^{\infty} y^{r-3} dy = \frac{c_r^*}{(2-r)} g_n^{-r+2} \quad \text{for } 1 < r < 2$$

and with  $c_r^{**} = \frac{r^{r-1}}{2\Gamma(r)} \sup_y \{(e^{-ry/2} (|y-1| |2-r| + 1))\} < \infty$ , we obtain

$$\int_{1/g_n}^{\infty} \frac{|h_{2,r}(y)|}{y^2} dy \leq c_r^{**} \int_{1/g_n}^{\infty} y^{r-3} e^{-ry/2} dy \leq \frac{c_r^{**} \Gamma(r-2)}{(r/2)^{r-2}} \quad \text{for } r > 2.$$

Hence, we obtain (35) for  $r > 1, r \neq 2$  with some constant  $0 < c_7 < \infty$ .

For  $r = 2$ , the second integral in line above is an exponential integral. Therefore, we estimate the integral  $I_2(x, n)$  in (35) more precisely like in estimating  $I_1(x, n)$  above, taking into account the given function  $f_2(z; a)$ .

Using  $|h_{2,2}(y)| \leq 4y e^{-2y}$  and consider (A2), define  $P_4(z; a)$  by  $Q_5(z; a) = -z P_4(z; a) \varphi(z/\sqrt{2})$  with  $c_2^* = \sup_z |P_4(z; a)| \varphi(z/\sqrt{2}) < \infty$ , we obtain in (A3)

$$\int_{1/g_n}^{\infty} \frac{|Q_5(x y^\gamma)|}{4 y^2} |h_{2,2}(y)| dy \leq \frac{c_2^* |x|}{\sqrt{2} \pi} \int_{1/g_n}^{\infty} y^{\gamma-1} e^{-(2y+x^2 y^{2\gamma}/4)} dy.$$

We estimate the latter integral in the same way as  $I_1(x, n)$  for the two cases  $\gamma = 1/2$   $\gamma = -1/2$  and find (35) for  $r = 2$  with some constants  $0 < c_7 < \infty$ .

In order to prove (35) for  $r = 2$  and  $\gamma = 0$ , we consider for  $\alpha > 0$  the following inequalities:

$$\int_{1/g_n}^{\infty} y^{-1} e^{-\alpha y} dy \begin{cases} \leq \int_{1/g_n}^1 y^{-1} dy + \int_1^{\infty} e^{-\alpha y} dy \leq \ln g_n + \alpha^{-1} e^{-\alpha}, \\ \geq \int_{1/g_n}^1 y^{-1} e^{-\alpha y} dy \geq e^{-\alpha} \int_{1/g_n}^1 y^{-1} dy \geq e^{-\alpha} \ln g_n \end{cases}. \quad (\text{A4})$$

The upper bound in (A4) leads to (35) for  $r = 2, \gamma = 0$ , too. The lower bound in (A4) shows that the  $\ln n$ -term cannot be improved.

Bound (36) for  $n \geq 2, r > 1$  results from  $0 \leq \frac{1}{g_n} - \frac{1}{r n} = \frac{r-1}{r^2 n^2 (1 - (r-1)/(r n))} \leq \frac{2(r-1)}{r^2 n^2}$ .  $\square$

**Proof of Lemma 3.** Let  $r > 0$ . If  $n = 1$ , then  $\mathbb{P}(N_1(r) = 1) = 1$  and (37) holds with  $C(r) = 1$ . Let  $n \geq 2$  and  $\alpha > 0$

$$\mathbb{E}(N_n(r))^{-\alpha} = \frac{1}{n^r} \left( 1 + \sum_{k=2}^{\infty} \frac{\Gamma(k+r-1)}{k^\alpha \Gamma(r) \Gamma(k)} \left( 1 - \frac{1}{n} \right)^{k-1} \right).$$

It follows from the relations (49) and (50) with their corresponding bounds in the proof of Theorem 1 in Christoph et al. [12] that

$$\frac{\Gamma(k+r-1)}{\Gamma(r) \Gamma(k)} = \frac{1}{(k+r-1) B(r, k)} = \frac{k^{r-1}}{\Gamma(r)} \left( 1 + R_1(k) \right), \quad |R_1(k)| \leq \frac{c_1(r)}{k}. \quad (\text{A5})$$

For  $x \geq k \geq 2$  using  $(1 - 1/n)^x \leq e^{-x/n}$ , we find

$$\frac{k^{r-1} (1 - 1/n)^{k-1}}{k^\alpha} \leq \int_k^{k+1} \frac{x^r (1 - 1/n)^{x-2}}{(x-1)^{1+\alpha}} dx \leq 2^{3+\alpha} \int_k^{k+1} x^{r-3} e^{-x/n} dx.$$

Then, with  $c_2(r) = 2^{3+\alpha} (1 + c_1(r)) / \Gamma(r)$ , we obtain

$$\mathbb{E}(N_n(r))^{-\alpha} \leq c_2(r) n^{-r} J_r(n), \quad \text{where } J_r(n) = \int_1^{\infty} x^{r-\alpha-1} e^{-x/n} dx = n^{r-\alpha} \int_{1/n}^{\infty} y^{r-\alpha-1} e^{-y} dy.$$

Since  $J_r(n) \leq (\alpha - r)^{-1}$  for  $0 < r < \alpha$ ,  $J_r(n) \leq n^{r-\alpha} \Gamma(r - \alpha)$  for  $r > \alpha$  and using (A4) with  $r = \alpha$   $J_r(n) \leq \ln n + e^{-1}$ , the upper bound (37) is proved.

Let  $r = \alpha > 0$ . Considering the formula (A5),  $0 \leq \sum_{k=2}^{\infty} k^{-1} |R_1(k)| \leq c_1(r) \pi^2 / (6 \Gamma(r)) < \infty$ ,  $\sum_{(n)} := \sum_{k=2}^{n-1} k^{-1} \geq \ln n - \ln 2$  and  $\sum_{k=2}^{n-1} \frac{1 - (1 - 1/n)^{k-1}}{k} \leq \sum_{k=2}^{n-1} \frac{k-1}{kn} \leq 1$ , we find:

$$\mathbb{E}(N_n(r))^{-r} \geq \frac{1}{n^r \Gamma(r)} \left( \sum_{k=2}^{n-1} \frac{1}{k} \left(1 - \frac{1}{n}\right)^{k-1} - c_3 \right) \geq \frac{1}{n^r \Gamma(r)} \left( \sum_{k=2}^{n-1} \frac{1}{k} - c_4 \right) \geq \frac{1}{n^r \Gamma(r)} (\ln n - c_5),$$

where  $c_3 = c_1(r) \pi^2 / 6$ ,  $c_4 = 1 + c_3$  and  $c_5 = c_4 - \ln 2$ . Hence, the  $\ln n$ -term cannot be dropped.  $\square$

**Proof of Lemma 4.** The upper bounds in the estimates (23) and (24) with  $H_s(y)$ ,  $h_{2,s}(y)$  and  $I_2(x, n)$  given in (40) are  $C(s)e^{-sn/2}$ . For example, (24ii):

$$\int_0^{1/n} y^{-1} |h_{2,s}(y)| dy \leq s(s+1)/2 \int_0^{1/n} y^{-3} e^{-s/y} dy \leq (s+1)/(2s) \int_{sn}^{\infty} z e^{-z} dz \leq (s+1)/(2s) e^{-sn/2}. \quad \square$$

**Proof of Lemma 5.** Proceeding as in Bening et al. [24] using

$$\mathbb{P}(N_n(s) = k) = \left( \frac{k}{s+k} \right)^n - \left( \frac{k-1}{s+k-1} \right)^n = s n \int_{k-1}^k \frac{x^{n-1}}{(s+x)^{n+1}} dx$$

and Formula 2.2.4.24 in Prudnikov et al. [37], p. 298, then

$$\mathbb{E}(N_n^{-\alpha}) = s n \sum_{k=1}^{\infty} \frac{1}{k^{\alpha}} \int_{k-1}^k \frac{x^{n-1}}{(s+x)^{n+1}} dx \leq s n \int_0^{\infty} \frac{x^{n-1-\alpha}}{(s+x)^{n+1}} dx = s n B(n-\alpha, 1+\alpha).$$

Using  $B(n-\alpha, 1+\alpha) = \Gamma(1+\alpha) (n+1)^{-1+\alpha} (1 + R_1/n)$  with  $|R_1| \leq c < \infty$ , we obtain (41).  $\square$

### Appendix A.3. Proofs of Theorems 3 to 8

**Proof of Theorem 3.** Since the additional assumptions (23) and (24) in the transfer Theorem 2 for the limit Gamma distribution  $H(x) = G_{r,r}(x)$  of the normalized sample size  $N_n(r)$  are satisfied by Lemma 2 with  $b = r > 0$  and by Lemma 3 for  $\alpha = 2$ , it remains to calculate the integrals in (26). Define

$$\begin{aligned} J_1^*(x) &= \int_0^{\infty} \Phi(x\sqrt{y}) dG_{r,r}(y), & J_2^*(x) &= \int_0^{\infty} \frac{a(x\sqrt{y})^3 - 5x\sqrt{y} \varphi(x\sqrt{y})}{4y} dG_{r,r}(y), & \text{and} \\ J_3^*(x) &= \int_0^{\infty} \Phi(x\sqrt{y}) dh_{2,r}(y) & \text{with } h_{2,r}(y) &= \left( (y-1)(2-r) + 2Q_1((r(n-1)+1)y) \right) \frac{g_{r,r}(y)}{2r}, \end{aligned}$$

and  $Q_1(y) = 1/2 - (y - [y])$ . Then,

$$G_{2,n}(x; 0) = J_1^*(x) + \frac{J_2^*(x)}{g_n} + \frac{J_3^*(x)}{n} \quad \text{with } g_n = \mathbb{E}N_n(r) = r(n-1) + 1. \quad (\text{A6})$$

Using formula 2.3.3.1 in Prudnikov et al. [37], p. 322, with  $\alpha = r - 1/2$ ,  $r + 1/2$ ,  $p = 1 + x^2/(2r)$  and  $q = 1$ :

$$M_{\alpha}(x) = \frac{r^r}{\Gamma(r) \sqrt{2\pi}} \int_0^{\infty} y^{\alpha-1} e^{-(r+x^2/2)y} dy = \frac{\Gamma(\alpha) r^{r-\alpha}}{\Gamma(r) \sqrt{2\pi}} (1 + x^2/(2r))^{-\alpha} \quad (\text{A7})$$

we calculate the integrals occurring in (A6). Consider

$$\begin{aligned} \frac{\partial}{\partial x} J_1^*(x) &= \int_0^{\infty} y^{1/2} \varphi(x\sqrt{y}) g_{r,r}(y) dy = \frac{r^r}{\Gamma(r) \sqrt{2\pi}} \int_0^{\infty} y^{r-1/2} e^{-(r+x^2/2)y} dy \\ &= M_{r+1/2}(x) = s_{2r}(x) \quad \text{and} \quad J_1^*(x) = S_{2r}(x). \end{aligned}$$

The integral  $J_2^*(x)$  in (A6) we calculate again with (A7) using  $M_{r-1/2}(x) = s_{2r}(x) (2r + x^2)/(2r - 1)$  and  $M_{r+1/2}(x) = s_{2r}(x)$

$$\begin{aligned} J_2^*(x) &:= \frac{r^r}{\sqrt{2\pi} \Gamma(r)} \int_0^\infty \frac{1}{y} \left( a x^3 y^{3/2} - 5 x y^{1/2} \right) y^{r-1} e^{-(r+x^2/2)y} dy \\ &= \left( a x^3 M_{r+1/2}(x) - 5 x M_{r-1/2}(x) \right) = \left( a x^3 - \frac{10 r x + 5 x^3}{2r - 1} \right) s_{2r}(x). \end{aligned}$$

The integral  $J_3^*(x)$  in (A6) is the same as the integral  $J_4(x)$  in the proof of Theorem 2 in Christoph et al. [12] with the estimate

$$\sup_x \left| J_3^*(x) - \frac{(2-r)x(x^2+1)}{4r(2r-1)} s_{2r}(x) \right| \leq c(r) n^{-r+1}.$$

With (36), we proved (44).  $\square$

**Proof of Theorem 4.** By Lemma 2, the additional assumptions (23) and (24) in the transfer Theorem 2 are satisfied with the limit Gamma distribution  $H(x) = G_{r,r}(x)$  of the normalized sample size  $N_n(r)$  with  $b = r > 0$ . In Transfer Theorem 1 for  $T_{N_n} = \sqrt{N_n} R_{N_n}$ , the right-hand side of (20) is estimated by Lemma 1 and Lemma 3 for  $\alpha = 2$  for the case  $\gamma = 0$ . Then, we have by (21) with (35)

$$G_n(x, 1/g_n) = \Phi(x) \left( 1 - G_{r,r}(1/g_n) - n^{-1} h_{2,r}(1/g_n) \mathbb{I}_{\{r>1\}}(r) \right) + \frac{f_2(x; a)}{g_n} \int_{1/g_n}^\infty \frac{1}{y} dG_{r,r}(y) \mathbb{I}_{\{r>1\}}(r).$$

The estimates (23), (24i), (24ii), (34) and  $\int_0^\infty y^{-1} dG_{r,r}(y) = r \Gamma(r-1)/\Gamma(r)$  for  $r > 1$  lead to (46) with  $\Phi_{n,2}(x; 1)$  defined in (47). Thus, Theorem 4 is proved.  $\square$

**Proof of Theorem 5.** By Lemma 2, the additional assumptions (23) and (24) in the transfer Theorem 2 are satisfied with the limit Gamma distribution  $H(x) = G_{r,r}(x)$  of the normalized sample size  $N_n(r)$  with  $b = r > 0$ . In Transfer Theorem 1 for  $T_{N_n} = g_n^{-1/2} N_n R_{N_n}$ , the right-hand side of (20) is estimated by Lemma 1 and Lemma 3 for  $\alpha = 2$  for the case  $\gamma = -1/2$ . Then, we have in (25)

$$G_{2,n}(x; 0) = J_{1,r}^*(x) + \left( \frac{J_{2,r}^*(x)}{g_n} + \frac{J_{3,r}^*(x)}{n} \right) \mathbb{I}_{\{r>1\}}(r) \quad \text{with} \quad g_n = \mathbb{E}N_n(r) = r(n-1) + 1 \quad (\text{A8})$$

$$\begin{aligned} J_{1,r}^*(x) &= \int_0^\infty \Phi(x/\sqrt{y}) dG_{r,r}(y), \quad J_{2,r}^*(x) = \int_0^\infty \frac{(a x^3 y^{-3/2} - 5 x y^{-1/2}) \varphi(x/\sqrt{y})}{4 y} dG_{r,r}(y), \quad \text{and} \\ J_{3,r}^*(x) &= \int_0^\infty \Phi(x/\sqrt{y}) dh_{2,r}(y) \quad \text{with} \quad h_{2,r}(y) = \left( (y-1)(2-r) + 2Q_1((r(n-1)+1)y) \right) \frac{g_{r,r}(y)}{2r}. \end{aligned}$$

Consider formula 2.3.16.1 in Prudnikov et al. [37], p. 444:

$$I_\alpha := \int_0^\infty y^{\alpha-1} e^{-py-q/y} dy = 2 \left( \frac{q}{p} \right)^{\alpha/2} K_\alpha(2\sqrt{pq}) \quad p > 0, \quad q > 0,$$

where  $K_\alpha(u)$  is the  $\alpha$ -order Macdonald function (or  $\alpha$ -order modified Bessel function of the second kind), see, e.g., Oldham et al. [30], Chapter 51, for properties of these functions.

Let us calculate the integral  $J_{1,r}^*(x)$  occurring in (A8). Consider

$$\begin{aligned} \frac{d}{dx} J_{1,r}^*(x) &= \frac{r^r}{\sqrt{2\pi} \Gamma(r)} \int_0^\infty y^{r-3/2} e^{-ry-(x^2/(2y))} dy \\ &= \frac{2r^r}{\sqrt{2\pi} \Gamma(r)} \left( \frac{|x|}{2r} \right)^{r-1/2} K_{r-1/2}(\sqrt{2r}|x|) =: l_r(x). \end{aligned} \quad (\text{A9})$$



If  $\alpha = \pm 1/2, \pm 3/2, \pm 5/2, \dots$  the integral  $I_\alpha$  and consequently  $K_\alpha(x)$  are computable in closed-form expressions with formula 2.3.16.2 in Prudnikov et al. [37], p. 444:

$$I_m^* = \int_0^\infty y^{m-1/2} e^{-py-q/y} dy = (-1)^m \sqrt{\pi} \frac{\partial^m}{\partial p^m} \left( p^{-1/2} e^{-2\sqrt{pq}} \right), \quad p > 0, q > 0, m = 0, 1, 2, \dots \quad (\text{A10})$$

and with formula 2.3.16.3 in Prudnikov et al. [37], p. 444:

$$I_{-m}^* = \int_0^\infty y^{-m-1/2} e^{-py-q/y} dy = (-1)^m \sqrt{\frac{\pi}{p}} \frac{\partial^m}{\partial q^m} e^{-2\sqrt{pq}}, \quad p > 0, q > 0, m = 0, 1, 2, \dots$$

For  $r = 1, 2, 3$  using (A10) with  $m = r - 1$ , we find

$$l_r(x) = \frac{d}{dx} J_{1,r}^*(x) = \frac{r^r}{\Gamma(r) \sqrt{2\pi}} \int_0^\infty y^{r-3/2} e^{-ry-x^2/(2y)} dy = \frac{r^r}{\Gamma(r) \sqrt{2\pi}} I_{r-1}^*$$

and we obtain the densities  $l_r(x)$  in (49) with

$$I_m^*(x) = \begin{cases} \sqrt{2\pi} \frac{1}{|x|} e^{-\sqrt{2r}|x|}, & x \neq 0, & m = -1, \\ \sqrt{\pi} e^{-\sqrt{2r}|x|}, & & m = 0, \\ \sqrt{\pi} \left( \frac{1}{2r^{3/2}} + \frac{|x|\sqrt{2}}{2r} \right) e^{-\sqrt{2r}|x|}, & & m = 1, \\ \sqrt{\pi} \left( \frac{3}{4r^{5/2}} + \frac{3|x|\sqrt{2}}{4r^2} + \frac{|x|^2}{2r^{3/2}} \right) e^{-\sqrt{2r}|x|}, & & m = 2. \end{cases}$$

Consider now  $J_{2,r}(x)$  for  $r = 2$  and  $r = 3$ :

$$J_{2,r}^*(x) = \int_0^\infty \frac{(a x^3 y^{-3/2} - 5x y^{-1/2}) r^r y^{r-1} e^{-ry-x^2/(2y)}}{4y \sqrt{2\pi} \Gamma(r)} dy = \frac{r^r}{4 \sqrt{2\pi} \Gamma(r)} (a x^3 I_{r-3}^*(x) - 5x I_{r-2}^*(x)).$$

Hence,

$$J_{2,2}^*(x) = (a x |x| - 5x/\sqrt{2}) e^{-2|x|} \quad \text{and} \quad J_{2,3}^*(x) = \frac{27}{8} \left( \frac{a x^3}{\sqrt{2}} - 5x \left( \frac{1}{6\sqrt{6}} + \frac{|x|}{6} \right) \right) e^{-\sqrt{6}|x|}.$$

Integration by parts in the integral  $J_{3,r}^*$  in (A8) leads to

$$\begin{aligned} J_{3,r}^*(x) &:= \int_0^\infty \Phi(x y^{-1/2}) d(h_{2,r}(y)) = \frac{x}{2} \int_0^\infty y^{-3/2} \varphi(x y^{-1/2}) h_{2,r}(y) dy \\ &= \frac{r^r x}{2r \Gamma(r) \sqrt{2\pi}} \int_0^\infty y^{r-5/2} e^{-ry-x^2/(2y)} \left( (y-1)(2-r) + 2Q_1(g_n y) \right) dy, \\ &= \frac{r^{r-1} x}{2 \Gamma(r) \sqrt{2\pi}} \int_0^\infty y^{r-5/2} (y-1)(2-r) e^{-ry-x^2/(2y)} dy \\ &\quad + \frac{r^{r-1} x}{\Gamma(r) \sqrt{2\pi}} \int_0^\infty y^{r-5/2} Q_1(g_n y) e^{-ry-x^2/(2y)} dy = J_{3,r,1}(x) + J_{3,r,2}(x). \end{aligned}$$

Since  $J_{3,2,1}(x)$  vanishes, we calculate  $J_{3,3,1}(x)$ :

$$J_{3,3,1}(x) = \frac{9x}{2\sqrt{2\pi}} (I_1^*(x) - I_2^*(x)) = \frac{9x}{2} \left( \frac{1}{12\sqrt{6}} + \frac{|x|}{12} - \frac{|x|^2}{6\sqrt{6}} \right) e^{-\sqrt{6}|x|}.$$

It remains to estimate  $J_{3,2,2}(x)$  and  $J_{3,3,2}(x)$ . The function  $Q_1(y)$  is periodic with period 1:

$$Q_1(y) = Q_1(y+1) \quad \text{for all } y \in \mathbb{R} \quad \text{and} \quad Q_1(y) := 1/2 - y \quad \text{for } 0 \leq y < 1$$

It is right-continuous and has the jump 1 at every integer point  $y$ . The Fourier series expansion of  $Q_1(y)$  at all non-integer points  $y$  is

$$Q_1(y) = 1/2 - (y - [y]) = \sum_{k=1}^{\infty} \frac{\sin(2\pi k y)}{k\pi} \quad y \neq [y], \quad (\text{A11})$$

see formula 5.4.2.9 in Prudnikov et al. [37], p. 726, with  $a = 0$ .

Using the Fourier series expansion (A11) of the periodic function  $Q_1(y)$  and interchange sum and integral, we find

$$J_{3;r,2}^* = \frac{x}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \frac{1}{k} \int_0^{\infty} y^{r-5/2} e^{-ry-x^2/(2y)} \sin(2\pi k g_n y) dy. \quad (\text{A12})$$

First, we consider  $r = 2$ . Let  $p > 0, q > 0$  and  $b > 0$  be some real constants. Formula 2.5.37.4 in Prudnikov et al. [37], p. 453 reads

$$\int_0^{\infty} y^{-1/2} e^{-py-q/y} \sin(by) dy = \sqrt{\frac{\pi}{p^2+b^2}} e^{-2\sqrt{q}z_+} (z_+ \sin(2\sqrt{q}z_-) + z_- \cos(2\sqrt{q}z_-)). \quad (\text{A13})$$

with  $2z_{\pm}^2 = \sqrt{p^2+b^2} \pm p$ . Consider  $z_{\pm}$  with  $p = r, q = x^2/2, b = 2\pi k g_n, k \geq 1$  and  $n \geq 1$ : Then,

$$\sqrt{\frac{\pi}{p^2+b^2}} = \sqrt{\frac{\pi}{r^2+(2\pi k g_n)^2}} \leq \frac{\sqrt{\pi}}{2\pi k g_n}, \quad \sqrt{2}|x|z_+ e^{-\sqrt{2}|x|z_+} \leq e^{-1} \quad \text{and} \quad 0 < z_- \leq z_+$$

leads to

$$\begin{aligned} |J_{3;2,2}^*(x)| &\leq \frac{2|x|}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \frac{1}{k} \sqrt{\frac{\pi}{p^2+b^2}} e^{-2\sqrt{q}z_+} (z_+ \sin(2\sqrt{q}z_-) + z_- \cos(2\sqrt{q}z_-)) \\ &\leq \frac{2}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \frac{1}{k} \frac{\sqrt{\pi} \sqrt{2} e^{-1}}{2\pi k g_n} = \frac{1}{2\pi e g_n} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi}{12 e g_n}. \end{aligned}$$

Together with  $g_n \geq n$ , we find  $n^{-1}|J_{3;2,2}^*(x)| \leq C n^{-2}$ .

Consider finally  $J_{3;3,2}^*$  given in (A12). In order to estimate  $J_{3;3,2}^*(x)$ , we apply Leibniz's rule for differentiation under the integral sign with respect to  $p$  in (A13) and obtain

$$\int_0^{\infty} y^{1/2} e^{-py-q/y} \sin(by) dy = \frac{\partial}{\partial p} \left\{ \sqrt{\frac{\pi}{p^2+b^2}} e^{-2\sqrt{q}z_+} (z_+ \sin(2\sqrt{q}z_-) + z_- \cos(2\sqrt{q}z_-)) \right\}.$$

Simple calculation considering  $\sqrt{q} = |x|/\sqrt{2}$  and  $|x|^m e^{-\sqrt{2}|x|z_+} \leq \frac{m}{2^{m/2} z_+} \leq \frac{m}{2^{m/2} b^m}$  for  $m = 1, 2$ , leads to

$$|x| \frac{\partial}{\partial p} \left\{ \sqrt{\frac{\pi}{p^2+b^2}} e^{-2\sqrt{q}z_+} (z_+ \sin(2\sqrt{q}z_-) + z_- \cos(2\sqrt{q}z_-)) \right\} \leq \frac{C}{b^2} = \frac{C}{(2\pi k g_n)^2}$$

and we find equation (A12) with  $r = 3$  that  $n^{-1}|J_{3;3,2}^*| \leq C n^{-3}$  and (50) is proved. The approximation (52) holds since Lemmas 1, 2, and 3 are valid for arbitrary  $r > 0$ . Theorem 5 is proved.  $\square$

**Proof of Theorem 6.** By Lemma 4, the additional assumptions (23) and (24) in the transfer Theorem 2 are satisfied with the limit inverse exponential distribution  $H_s(y)$  and  $h_{2;s}(y)$  given in (40),  $g_n = n$  and  $b = 2$ . In Transfer Theorem 1, the right-hand side of (20) is estimated by Lemma 1 and Lemma 5 for  $\alpha = 2$  for the case  $\gamma = 1/2$ . Then, we have in (25) with (35)

$$G_{2;n}(x;0) = J_{1;s}^*(x) + n^{-1} J_{2;s}^*(x) + n^{-1} J_{3;s}^*(x),$$

where  $J_{1;s}^*(x) = \int_0^\infty \Phi(x\sqrt{y})de^{-s/y}$ ,  $J_{2;s}^*(x) = \int_0^\infty \frac{(ax^3y^{3/2} - 5x\sqrt{y})\varphi(x\sqrt{y})}{4y}de^{-s/y}$ ,  
 and  $J_{3;s}^*(x) = \int_0^\infty \Phi(x\sqrt{y})dh_{2;s}(y)$  with  $h_{2;s}(y) = se^{-s/y}(s-1+2Q_1(ny))/(2y^2)$ .

To obtain (53), we calculate the above integrals as in the proof of Theorem 5 in Christoph et al. [12]. Here, we use Formula 2.3.16.3 in Prudnikov et al. [37], p. 344 with  $p = x^2/2 > 0$ ,  $s > 0$ ,  $m = 1, 2$ :

$$\int_0^\infty \frac{e^{-x^2y/2}}{\sqrt{2\pi}y^{m-3/2}}dH_s(y) = \int_0^\infty \frac{se^{-x^2y/2-s/y}}{\sqrt{2\pi}y^{m+1/2}}dy = (-1)^m \frac{s}{|x|} \frac{\partial^m}{\partial s^m} e^{-\sqrt{2s}|x|}. \quad (\text{A14})$$

In the mentioned proof we obtained with (A14) for  $m = 1$

$$\int_0^\infty \Phi(x\sqrt{y})dH_s(y) = L_{1/\sqrt{s}}(x)$$

and with (A14) for  $m = 2$

$$n^{-1} \sup_x \left| J_{3;s}^*(x) - \frac{(1-s)x(1+\sqrt{2s}|x|)}{8s} l_{1/\sqrt{s}}(x) \right| \leq n^{-1} c(s) e^{-\sqrt{\pi sn}/2} \leq C(s) n^{-2}.$$

Again using (A14) with  $p = x^2/2 > 0$ ,  $s > 0$ ,  $m = 1, 2$  we find

$$\begin{aligned} J_{2;s}(x) &= \frac{s}{\sqrt{2\pi}} \int_0^\infty (ax^3y^{-1-1/2} - 5xy^{-2-1/2})e^{-(x^2y/2+s/y)}dy \\ &= \frac{2sax^3 - 5x(\sqrt{2s}|x| + 1)}{8s} l_{1/\sqrt{s}}(x). \end{aligned}$$

□

**Proof of Theorem 7.** By Lemma 4, the additional assumptions (23) and (24) in Transfer Theorem 2 are satisfied with the limit inverse exponential distribution  $H_s(y)$  and  $h_{2;s}(y)$  given in (40),  $g_n = n$  and  $b = 2$ . In Transfer Theorem 1, the right-hand side of (20) is estimated by Lemma 1 and Lemma 5 for  $\alpha = 2$  in the case  $\gamma = 0$ . Then, we have in (21) with (35)

$$G_n(x, 1/n) = \Phi(x) \left( 1 - e^{-sn} - n^{-1}h_{2;s}(1/n) \right) + \frac{f_2(x; a)}{n} \int_{1/n}^\infty \frac{1}{y} de^{-s/y}.$$

The estimates (24i), (24ii) for  $b = 2$  and  $\int_0^\infty y^{-1}de^{-s/y} = s \int_0^\infty y^{-3}e^{-s/y}dy = s^2 \int_0^\infty ze^{-z}dz = s^2$  lead to

$$\left| G_n(x, 1/g_n) - \Phi(x) - n^{-1}s^2f_2(x; a) \right| \leq C_s n^{-2}$$

and Theorem 7 is proved. □

**Proof of Theorem 8.** By Lemma 4, the additional assumptions (23) and (24) in Transfer Theorem 2 are satisfied with the limit inverse exponential distribution  $H_s(y)$  and  $h_{2;s}(y)$  given in (40),  $g_n = n$  and  $b = 2$ . In Transfer Theorem 1, the right-hand side of (20) is estimated by Lemma 1 and Lemma 5 for  $\alpha = 2$  in the case  $\gamma = -1/2$ . Then, we have in (21) with (35)

$$G_{2;n}(x; 0) = J_{1;s}^*(x) + n^{-1}J_{2;s}^*(x) + n^{-1}J_{3;s}^*(x),$$

where  $J_{1;s}^*(x) = \int_0^\infty \Phi(xy^{-1/2})de^{-s/y}$ ,  $J_{2;s}^*(x) = \int_0^\infty \frac{(ax^3y^{-3/2} - 5xy^{-1/2})\varphi(xy^{-1/2})}{4y}de^{-s/y}$ ,  
 and  $J_{3;s}^*(x) = \int_0^\infty \Phi(xy^{-1/2})dh_{2;s}(y)$  with  $h_{2;s}(y) = se^{-s/y}(s-1+2Q_1(ny))/(2y^2)$ .

To obtain (54), we calculate the above integrals:

$$\begin{aligned} \frac{\partial}{\partial x} \int_0^\infty \Phi(x\sqrt{y}) de^{-s/y} &= \frac{s}{\sqrt{2\pi}} \int_0^\infty y^{-3/2} e^{-(x^2/2+s)y} dy = \frac{s}{\sqrt{2\pi}} \int_0^\infty z^{1/2-1} e^{-(x^2/2+s)z} dz \\ &= \frac{1}{2\sqrt{2s}} \left(1 + \frac{x^2}{2s}\right)^{-3/2} = s_2^*(x; \sqrt{s}), \quad \text{and} \quad \int_0^\infty \Phi(x\sqrt{y}) de^{-s/y} = S_2^*(x; \sqrt{s}). \end{aligned} \quad (\text{A15})$$

Define  $K = (s + x^2/2)$ . With  $z = K/y$  and  $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz, \alpha > 0$ , we obtain

$$\begin{aligned} J_{2;s}^*(x) &= \frac{s}{4\sqrt{2\pi}} \int_0^\infty (ax^3 y^{-9/2} - 5xy^{-7/2}) e^{-K/y} dy = \frac{s K^{-7/2}}{4\sqrt{2\pi}} \int_0^\infty (ax^3 z^{5/2} - 5xz^{3/2} K) e^{-z} dz \\ &= \frac{s K^{-7/2}}{4\sqrt{2\pi}} (ax^3 \Gamma(7/2) - 5x K \Gamma(5/2)) = \frac{1}{4(x^2 + 2s)^2} (15(a-1)x^3 - 30xs) s_2^*(x; \sqrt{s}). \end{aligned}$$

Integration by parts in  $J_{3;s}^*(x)$  leads to

$$J_{3;s}^*(x) = \frac{x}{2\sqrt{2\pi}} \int_0^\infty y^{-3/2} e^{-x^2/(2y)} s y^{-2} e^{-s/y} ((s-1)/2 + Q_1(ny)) dy = J_{4;s}^*(x) + J_{5;s}^*(x),$$

where

$$J_{4;s}^*(x) = \frac{s(s-1)x}{4\sqrt{2\pi}} \int_0^\infty y^{-7/2} e^{-K/y} dy = \frac{s(s-1)x \Gamma(5/2)}{4\sqrt{2\pi} K^{5/2}} = \frac{3(s-1)x}{4(x^2 + 2s)} s_2^*(x; \sqrt{s})$$

and using the Fourier series expansion (A11) of the periodic function  $Q_1(y)$  and interchange sum and integral, we find

$$\begin{aligned} J_{5;s}^*(x) &= \frac{s x}{2\sqrt{2\pi}} \int_0^\infty y^{-7/2} e^{-K/y} Q_1(ny) dy = \frac{s x}{2\sqrt{2\pi}} \sum_{k=1}^\infty \frac{1}{k} \int_0^\infty y^{-7/2} e^{-K/y} \sin(2\pi k n y) dy \\ &= \frac{s x}{2\sqrt{2\pi}} \sum_{k=1}^\infty \frac{1}{k} \int_0^\infty y^{-7/2} e^{-K/y} \sin(2\pi k n y) dy. \end{aligned}$$

Integration by parts in the latter integral and  $|x|/\sqrt{K} \leq \sqrt{2}$  leads now to

$$\sup_x |J_{5;s}^*(x)| \leq \sup_x \frac{s|x|}{(2\pi)^{3/2} n} \sum_{k=1}^\infty \frac{1}{k^2} \int_0^\infty \left( \frac{7}{2} y^{-9/2} + K y^{11/2} \right) e^{-K/y} dy \leq c_s n^{-1}$$

with  $c_s = \frac{s\sqrt{2}}{(2\pi)^{3/2} n} \left( \frac{7\Gamma(11/2)}{2s^3} + \frac{\Gamma(13/2)}{s^4} \right) \frac{\pi^2}{6}$  and Theorem 8 is proved.  $\square$

## References

1. Hall, P.; Marron, J.S.; Neeman, A. Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 427–444. [\[CrossRef\]](#)
2. Fujikoshi, Y.; Ulyanov, V.V.; Shimizu, R. *Multivariate Statistics. High-Dimensional and Large-Sample Approximations*; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010.
3. Aoshima, M.; Shen, D.; Shen, H.; Yata, K.; Zhou, Y.-H.; Marron, J.S. A survey of high dimension low sample size asymptotics. *Aust. N. Z. J. Stat.* **2018**, *60*, 4–19. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Kawaguchi, Y.; Ulyanov, V.V.; Fujikoshi, Y. Asymptotic distributions of basic statistics in geometric representation for high-dimensional data and their error bounds (Russian). *Inform. Appl.* **2010**, *4*, 12–17.
5. Ulyanov, V.V.; Christoph, G.; Fujikoshi, Y. On approximations of transformed chi-squared distributions in statistical applications. *Sib. Math. J.* **2006**, *47*, 1154–1166. [\[CrossRef\]](#)
6. Esquivel, M.L.; Mota, P.P.; Mexia, J.T. On some statistical models with a random number of observations. *J. Stat. Theory Pract.* **2016**, *10*, 805–823. [\[CrossRef\]](#)

7. Gnedenko, B.V. Estimating the unknown parameters of a distribution with a random number of independent observations. (Probability theory and mathematical statistics (Russian)). *Tr. Tbil. Mat. Instituta* **1989**, *92*, 146–150.
8. Nunes, C.; Capistrano, G.; Ferreira, D.; Ferreira, S.S.; Mexia, J.T. Fixed effects ANOVA: an extension to samples with random size. *J. Stat. Comput. Simul.* **2014**, *84*, 2316–2328. [[CrossRef](#)]
9. Nunes, C.; Capistrano, G.; Ferreira, D.; Ferreira, S.S.; Mexia, J.T. Exact critical values for one-way fixed effects models with random sample sizes. *J. Comput. Appl. Math.* **2019**, *354*, 112–122. [[CrossRef](#)]
10. Barakat, H.M.; Nigm, E.M.; El-Adll, M.E.; Yusuf, M. Prediction of future generalized order statistics based on exponential distribution with random sample size. *Stat. Pap.* **2018**, *59*, 605–631. [[CrossRef](#)]
11. Al-Mutairi, J.S.; Raqab, M.Z. Confidence intervals for quantiles based on samples of random sizes. *Stat. Pap.* **2020**, *61*, 261–277. [[CrossRef](#)]
12. Christoph, G.; Monakhov, M.M.; Ulyanov, V.V. Second-order Chebyshev–Edgeworth and Cornish-Fisher expansions for distributions of statistics constructed with respect to samples of random size. *J. Math. Sci.* **2020**, *244*, 811–839. [[CrossRef](#)]
13. Christoph, G.; Ulyanov, V.V.; Bening, V.E. Second Order Expansions for Sample Median with Random Sample Size. *arXiv* **2019**, arXiv:1905.07765v2.
14. Fujikoshi, Y.; Ulyanov, V.V. *Non-Asymptotic Analysis of Approximations for Multivariate Statistics*; Springer: Singapore, 2020.
15. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1995; Volume 2.
16. Konishi, S. Asymptotic expansions for the distributions of functions of a correlation matrix. *J. Multivar. Anal.* **1979**, *9*, 259–266. [[CrossRef](#)]
17. Christoph, G.; Ulyanov, V.V.; Fujikoshi, Y. Accurate approximation of correlation coefficients by short Edgeworth–Chebyshev expansion and its statistical applications. In *Prokhorov and Contemporary Probability Theory*; Proceedings in Mathematics & Statistics 33; Shiryaev, A.N., Varadhan, S.R.S., Presman, E.L., Eds.; Springer: Heidelberg, Germany, 2013; pp. 239–260.
18. Prokhorov, Y.V. Limit theorems for the sums of random vectors whose dimension tends to infinity. *Teor. Veroyatnostei Primen.* **1990**, *35*, 751–753. English translation: *Theory Probab. Appl.* **1991**, *35*, 755–757. [[CrossRef](#)]
19. Gnedenko, B.V.; Korolev, V.Y. *Random Summation. Limit Theorems and Applications*; CRC Press: Boca Raton, FL, USA, 1996.
20. Döbler, C. New Berry-Esseen and Wasserstein bounds in the CLT for non-randomly centered random sums by probabilistic methods. *ALEA Lat. Am. J. Probab. Math. Stat.* **2015**, *12*, 863–902.
21. Chen, L.H.Y.; Goldstein, L.; Shao, Q.-M. *Normal Approximation by Stein’s Method. Probability and its Applications*; Springer: Heidelberg, Germany, 2011.
22. Pike, J.; Ren, H. Stein’s method and the Laplace distribution. *ALEA Lat. Am. J. Probab. Math. Stat.* **2014**, *11*, 571–587.
23. Bening, V.E.; Korolev, V.Y. On the use of Student’s distribution in problems of probability theory and mathematical statistics. *Theory Probab. Appl.* **2005**, *49*, 377–391. [[CrossRef](#)]
24. Bening, V.E.; Galieva, N.K.; Korolev, V.Y. Asymptotic expansions for the distribution functions of statistics constructed from samples with random sizes (Russian). *Inform. Appl.* **2013**, *7*, 75–83.
25. Bening, V.E.; Korolev, V.Y.; Zeifman, A.I. Asymptotic expansions for the distribution function of the sample median constructed from a sample with random size. In *Proceedings 30th ECMS 2016 Regensburg*; Claus, T., Herrmann, F., Manitz, M., Rose, O., Eds.; European Council for Modeling and Simulation: Regensburg, Germany, 2016; pp. 669–675.
26. Schluter, C.; Trede, M. Weak convergence to the Student and Laplace distributions. *J. Appl. Probab.* **2016**, *53*, 121–129. [[CrossRef](#)]
27. Gavrilenko, S.V.; Zubov, V.N.; Korolev, V.Y. The rate of convergence of the distributions of regular statistics constructed from samples with negatively binomially distributed random sizes to the Student distribution. *J. Math. Sci.* **2017**, *220*, 701–713. [[CrossRef](#)]
28. Buddana, A.; Kozubowski, T.J. Discrete Pareto distributions. *Econ. Qual. Control.* **2014**, *29*, 143–156. [[CrossRef](#)]
29. Christoph, G.; Wolf, W. *Convergence Theorems with a Stable Limit Law, Mathematical Research 70*; Akademie-Verlag: Berlin, Germany, 1992.

30. Oldham, K.B.; Myl, J.C.; Spanier, J. *An Atlas of Functions*, 2nd ed.; Springer Science + Business Media: New York, NY, USA, 2009.
31. Kotz, S.; Kozubowski, T.J.; Podgórski, K. *The Laplace distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*; Birkhäuser: Boston, MA, USA, 2001.
32. Prudnikov, A.P.; Brychkov, Y.A.; Marichev, O.I. *Integrals and Series, Volume 2: Special Functions*, 3rd ed.; Gordon & Breach Science Publishers: New York, NY, USA, 1992.
33. Madan, D.B.; Senata, E. The variance gamma (V.G.) model for share markets returns. *J. Bus.* **1990** *63*, 511–524. [[CrossRef](#)]
34. Goldfeld, S.M.; Quandt, R.E. Econometric modelling with non-normal disturbances. *J. Econom.* **1981**, *17*, 141–155. [[CrossRef](#)]
35. Missiakales, S. Sargan densities which one? *J. Econom.* **1983**, *23*, 223–233. [[CrossRef](#)]
36. Jackman, S. *Bayesian Analysis for the Social Sciences*; Wiley Series in Probability and Statistics; John Wiley & Sons, Ltd.: Chichester, UK, 2009.
37. Prudnikov, A.P.; Brychkov, Y.A.; Marichev, O.I. *Integrals and Series, Volume 1: Elementary Functions*, 3rd ed.; Gordon & Breach Science Publishers: New York, NY, USA, 1992.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).