

IMPROVED APPROXIMATION OF THE SUM OF RANDOM VECTORS BY THE SKEW NORMAL DISTRIBUTION

MARCUS C. CHRISTIANSEN,* *University of Ulm*
NICOLA LOPERFIDO,** *Università degli Studi di Urbino Carlo Bo*

Abstract

We study the properties of the multivariate skew normal distribution as an approximation to the distribution of the sum of n independent, identically distributed random vectors. More precisely, we establish conditions ensuring that the uniform distance between the two distribution functions converges to 0 at a rate of $n^{-2/3}$. The advantage over the corresponding normal approximation is particularly relevant when the summands are skewed and n is small, as illustrated for the special case of exponentially distributed random variables. Applications to some well-known multivariate distributions are also discussed.

Keywords: Central limit theorem; Cramer's condition; order of convergence; skew normal; skewness; third-order tensor

2010 Mathematics Subject Classification: Primary 60F05; 60B12
Secondary 15A69; 62E17

1. Introduction

The normal distribution is often used as an approximating distribution, due to the central limit theorem. Unfortunately, the speed of convergence cannot be faster than $n^{-1/2}$ in the general case. More precisely, let X_1, X_2, \dots be independent, identically distributed (i.i.d.) and standardized random vectors with values in \mathbb{R}^d . Let Q_n denote the distribution of $Y_n = n^{-1/2}(X_1 + \dots + X_n)$. According to the Berry–Esseen Theorem (see, e.g. Theorem 13.3 of Bhattacharya and Ranga Rao (1986)), there exists a finite constant $C > 0$ such that, for every Borel set $A \subset \mathbb{R}^d$, we have

$$|Q_n(A) - \Phi(A)| \leq n^{-1/2}C \tag{1}$$

if the random vectors X_1, X_2, \dots have finite fourth-order moments, where Φ is the d -dimensional standardized normal distribution. Now consider the two-point distributed random variables X_1, X_2, \dots with $\mathbb{P}(X_i = -1) = \frac{1}{2} = \mathbb{P}(X_i = 1)$. Then, for $n = 2m$, we have

$$\lim_{2m \rightarrow \infty} (2m)^{1/2} Q_{2m}(\{0\}) = \sqrt{\frac{2}{\pi}},$$

which means that, for even n , the distribution Q_n has points with probability mass of order $n^{-1/2}$. Since Φ is absolutely continuous, the absolute difference $|Q_n - \Phi|$ also has points

Received 17 May 2012; revision received 24 June 2013.

* Postal address: Institute of Insurance Science, University of Ulm, 89081 Ulm, Germany.

Email address: marcus.christiansen@uni-ulm.de

** Postal address: Dipartimento di Economia, Politica e Società, Università degli Studi di Urbino Carlo Bo, Via Saffi 42, 61029 Urbino, Italy.

with mass of order $n^{-1/2}$, and, consequently, the speed of convergence in (1) cannot be faster than $n^{-1/2}$.

In order to rule out purely discrete distributions, in particular the above mentioned two-point distributions, we might impose *Cramer's condition*, which holds for a probability measure Q on \mathbb{R}^d satisfying

$$\limsup_{\|t\| \rightarrow \infty} |\widehat{Q}(t)| < 1,$$

where \widehat{Q} is the characteristic function of a random vector with distribution Q . Cramer's condition is equivalent to

$$\sup_{\|t\| > b} |\widehat{Q}(t)| < 1$$

for all positive b ; see, e.g. Bhattacharya and Ranga Rao (1986, Section 20). If the random vectors X_1, X_2, \dots satisfy Cramer's condition and have finite fourth-order moments, the first-order Edgeworth expansion (see, e.g Theorem 20.1 of Bhattacharya and Ranga Rao (1986)) yields a finite constant $C > 0$ such that, for every Borel set $A \subset \mathbb{R}^d$, we have

$$\left| Q_n(A) - \Phi(A) - n^{-1/2} \frac{1}{6} \int_A \left(\sum_{(i,j,k) \in \{1, \dots, d\}^3} m_{ijk} x_i x_j x_k - \sum_{(i,j) \in \{1, \dots, d\}^2} 3m_{ij} x_j \right) \phi(x) dx \right| \leq n^{-1} C,$$

where $m_{ijk} = \mathbb{E}(X_{1i} X_{1j} X_{1k})$ for $i, j, k \in \{1, \dots, d\}$ and where X_{1i} is the i th entry of random vector X_1 . The absolute difference $|Q_n(A) - \Phi(A)|$ converges to 0 as n goes to ∞ , but the speed of convergence is not faster than $n^{-1/2}$, unless we further assume that the random vectors X_1, X_2, \dots are nonskewed, which happens when $m_{ijk} = 0$ for all triples $(i, j, k) \in \{1, \dots, d\}^3$. However, the assumption that X_1, X_2, \dots are nonskewed is too strong for many applications; see, e.g. financial applications (Adcock (2007)). Another solution is to approximate $Q_n(A)$ by

$$\Phi(A) + n^{-1/2} \frac{1}{6} \int_A \left(\sum_{(i,j,k) \in \{1, \dots, d\}^3} m_{ijk} x_i x_j x_k - \sum_{(i,j) \in \{1, \dots, d\}^2} 3m_{ij} x_j \right) \phi(x) dx$$

instead of $\Phi(A)$, which is the original intention of Edgeworth's expansion. Unfortunately, the elements of the approximating series are not necessarily probability distributions, and the approximation deteriorates in the distribution's tails, where it can even become negative; see Field and Ronchetti (1990, p. 14).

Since the problem lies in the skewness of Q_n , it is intuitively appealing to approximate Q_n with another skewed distribution. The skew normal distribution, being a simple and skewed generalization of the normal distribution, makes a natural candidate for this task. A random variable $Z \in \mathbb{R}^1$ has a skew normal distribution with location parameter ξ , scale parameter ψ , and shape parameter λ if its probability density function is

$$f(z; \xi, \psi, \lambda) = \frac{2}{\psi} \phi\left(\frac{z - \xi}{\psi}\right) \Phi\left(\lambda \frac{z - \xi}{\psi}\right), \tag{2}$$

where $\lambda, \xi, z \in \mathbb{R}$, $\psi \in \mathbb{R}^+$, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function (PDF) and the cumulative distribution function (CDF) of a standard normal distribution, respectively (Azzalini (1985)). When (2) is the PDF of Z we shall write $Z \sim \text{SN}(\xi, \psi, \lambda)$. The parameters ξ, ψ, λ do not in general equal the expectation μ , the standard deviation σ , skewness γ_1 , or

kurtosis β_2 . However, μ , σ , γ_1 , and β_2 are simple functions of ξ , ψ , λ . Azzalini and Dalla Valle (1996) generalized the skew normal distribution to the multivariate case by introducing the multivariate skew normal distribution. The distribution of a random vector $Z \in \mathbb{R}^d$ is d -dimensional skew normal with location parameter ξ , scale parameter Ω , and shape parameter α , that is $Z \sim \text{SN}(\xi, \Omega, \alpha)$, if its PDF is

$$f(z; \xi, \Omega, \alpha) = 2\phi_d(z - \xi; \Omega)\Phi(\alpha^\top(z - \xi)), \quad z, \xi, \alpha \in \mathbb{R}^d, \Omega \in \mathbb{R}^{d \times d}, \quad (3)$$

where $\Phi(\cdot)$ is the CDF of a standardized normal variable and $\phi_d(z - \xi; \Omega)$ is the density function of a d -dimensional normal distribution with mean ξ and variance Ω ; see Azzalini and Dalla Valle (1996). It is easy to check that (3) reduces to (2) in the univariate case, by letting $\psi = \Omega^{1/2}$ and $\lambda = \alpha\psi$. Also, all univariate marginals of (3) are of the form (2). Basic distributional properties of multivariate skew normal distributions are reviewed in De Luca *et al.* (2006). Applications and extensions of the skew normal distribution are discussed in the survey papers by Azzalini (2005), (2006), and Kotz and Vicari (2005). According to Kotz and Vicari (2005), ‘the field of skewed distributions has become—in our opinion—one of the most fruitful and promising areas in the development of statistical distribution theory and applications, during the last 20 years which does not so far require using advanced mathematical tools’.

In the literature several authors argue that the skew normal distribution offers a better approximation for Q_n than the normal distribution if Q is skewed. Gupta and Kollo (2003) propose a density expansion based on the multivariate skew normal distribution. Bartoletti and Loperfido (2010) present a similar expansion in the univariate case, for the mean of dependent data. However, both expansions are accurate up to an error which vanishes with speed $n^{-1/2}$, which is no faster than the approximation speed in the normal approximation. Chang *et al.* (2008) show some examples of binomial distributions which can be much better approximated by skew normal distributions than by normal distributions. However, no theoretical results are given to support their intriguing numerical results.

In the present paper we overcome the above mentioned problems by constructing an approximating series for $Q_n(A)$ from the class of skew normal distributions. We prove that, under Cramer’s condition and with existing fourth-order moments, the speed of convergence is at least $n^{-2/3}$. The rest of the paper is organized as follows. In Section 2 we introduce some basic concepts and notation. Sections 3 and 4 show some preliminary results and the main convergence result, respectively. In Section 5 we discuss a financial application. In Section 6 we use the sum of exponentially distributed random variables to illustrate the theoretical results in the previous sections. Section 7 contains some concluding remarks and hints for future research.

2. Tensor notation for third moments and third cumulants

Third-order tensors provide a natural tool for representing the skewness of a random vector. A (real) tensor $\mathcal{A} = \{a_{i_1 \dots i_p}\} \in \mathbb{R}^{n_1 \times \dots \times n_p}$ is a multidimensional array of numerical values identified by a vector of subscripts. We call p the order of \mathcal{A} and $\dim(\mathcal{A}) = (n_1, \dots, n_p)^\top$ the dimension of \mathcal{A} . Up to a choice of bases, a tensor is an element of a tensor product of vector spaces. Comon *et al.* (2008) and Qi *et al.* (2007) extensively review tensor properties and applications. Good reviews of basic concepts can also be found in Comon *et al.* (2008), Kolda and Bader (2009), and Kilmer and Martin (2011).

Intuitively, tensors generalize matrices. Up to a choice of bases, each matrix represents a linear transformation acting on a vector space. Similarly, up to a choice of bases, each tensor represents a multilinear transformation simultaneously acting on several vector spaces.

A multilinear operator is a function of several elements belonging to possibly different vector spaces, which is linear in each one of the elements themselves. For example, the determinant of a square matrix is a multilinear function of its columns. Similarities between matrices and tensors might be better appreciated by looking at third-order tensors and symmetric tensors. Third-order tensors might be regarded as linear operators acting on spaces of matrices; see Braman (2010). Similarly, matrices represent linear operators acting on spaces of vectors. Each symmetric tensor is uniquely associated to a polynomial in several variables; see Comon *et al.* (2008). Similarly, each quadratic form in several variables is uniquely associated to a symmetric matrix. As a statistical application, the third cumulant of a linear function $a^\top x$ of a d -dimensional random vector x is a third degree polynomial in a_1, \dots, a_d , characterized by the third cumulant of x .

Some tensors are of special interest; for example, symmetric, rank-one, and third-order tensors. A tensor is symmetric if it is invariant under permuting indices; see Comon *et al.* (2008) and Brachat *et al.* (2010). A tensor is of rank one if it may be represented as the Kronecker product of several vectors. In probability and statistics, third-order tensors are the natural tool for representing the skewness of a random vector $X = (X_1, \dots, X_d)^\top$ with mean $\mu = (\mu_1, \dots, \mu_d)^\top$ which satisfies $\mathbb{E}(|X_i^3|) < \infty$ for $i = 1, \dots, d$. The third cumulant of X is $\mathcal{K}_3 = \{\kappa_{ijk}\}$, where $\kappa_{ijk} = \mathbb{E}((X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k))$ for $i, j, k = 1, \dots, d$; see McCullagh (1987, p. 25). In order to use standard matrix techniques, it is often convenient to ‘flatten’ \mathcal{K}_3 into the matrix

$$\kappa_3(X) = \mathbb{E}((X - \mu) \otimes (X - \mu)^\top \otimes (X - \mu)),$$

which is also referred to as the third cumulant of X . The symbol ‘ \otimes ’ denotes the Kronecker (or tensor) product, and acts on matrices $A = \{a_{ij}\} \in \mathbb{R}^{p \times q}$ and $B = \{b_{ij}\} \in \mathbb{R}^{h \times k}$ by obtaining a block matrix $A \otimes B \in \mathbb{R}^{ph \times qk}$ whose i, j th block is the matrix $a_{ij}B$ (see, e.g. Mardia *et al.* (1979, p. 459). Third cumulants are symmetric tensors of order 3. De Luca and Loperfido (2012) review the main properties of third cumulants and also examine some third cumulants of rank one.

The following lemma shows in tensor notation how the third moment

$$\mu_3(X) = \mathbb{E}(X \otimes X^\top \otimes X)$$

of a random vector X relates to the third moment of its linear transformation $Y = AX$.

Lemma 1. *Let X be a d -dimensional random vector with finite third moment $\mu_3(X)$, and let A be a $k \times d$ real matrix. Then the third moment of AX is $\mu_3(AX) = (A \otimes A)\mu_3(X)A^\top$.*

Proof. We first recall some fundamental properties of the Kronecker product (see, e.g. Mardia *et al.* (1979, p. 460)).

(P1) The Kronecker product is associative, $(A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C$.

(P2) If $A, B, C,$ and D are matrices of such size that we can form the matrix products AC and BD then $(A \otimes B)(C \otimes D) = AC \otimes BD$.

(P3) If a and b are two vectors then $ab^\top, a \otimes b^\top,$ and $b^\top \otimes a$ denote the same matrix.

(P4) If $A, B,$ and C are matrices of such size that we can form the matrix product BC then $A \otimes BC = (A \otimes B)C$ and $C(A \otimes B) = CA \otimes B$.

By definition, the third moment of AX is

$$\mu_3(AX) = \mathbb{E}(AX \otimes X^T A^T \otimes AX).$$

As a direct consequence of P1 and P3, $\mu_3(AX)$ equals

$$\mathbb{E}(AX \otimes (X^T A^T \otimes AX)) = \mathbb{E}(AX \otimes AX \otimes X^T A^T).$$

We now use properties P1, P2, and P4 to represent $\mu_3(AX)$ as

$$\begin{aligned} \mathbb{E}((AX \otimes AX) \otimes X^T A^T) &= \mathbb{E}((A \otimes A)(X \otimes X) \otimes X^T A^T) \\ &= \mathbb{E}((A \otimes A)(X \otimes X \otimes X^T)A^T). \end{aligned}$$

Linear properties of the expected value and further application of P1 and P3 lead to

$$\mu_3(AX) = (A \otimes A)\mathbb{E}(X \otimes X \otimes X^T)A^T = (A \otimes A)\mathbb{E}(X \otimes X^T \otimes X)A^T.$$

The expected value in the above equation is the third moment of X , that is $\mu_3(X)$. Hence, we can write $\mu_3(AX) = (A \otimes A)\mu_3(X)A^T$ and complete the proof.

3. Preliminary results on the skew normal distribution

The following lemmas are instrumental in proving the main convergence theorem, which will be presented in the next section.

Let X_1, X_2, \dots be i.i.d. and standardized random variables with values in \mathbb{R}^d . Let Q_n denote the distribution of $Y_n = n^{-1/2}(X_1 + \dots + X_n)$. Let $\text{SN}(\cdot; \xi, \Omega, \alpha)$ denote the d -dimensional distribution of a multivariate skew normal random vector.

Lemma 2. *The third cumulant of X_1 is a third-order tensor of rank at most one, i.e.*

$$\kappa_3(X_1) = \gamma \otimes \gamma^T \otimes \gamma$$

for some $\gamma \in \mathbb{R}^d$, if and only if there exist parameters ξ_n, Ω_n , and α_n such that, for all

$$n > \|\gamma\|^6 \frac{(\pi - 2)^3}{2(4 - \pi)^2}, \tag{4}$$

the first three cumulants and the first three moments of $\text{SN}(\xi_n, \Omega_n, \alpha_n)$ and Y_n are equal. The parameters are

$$\delta_n = n^{-1/6} \frac{\pi^{1/2}}{(\sqrt{2}(4 - \pi))^{1/3}} \gamma, \quad \Omega_n = I + \frac{2}{\pi} \delta_n \delta_n^T, \quad \xi_n = -\sqrt{\frac{2}{\pi}} \delta_n,$$

and

$$\alpha_n = \frac{\Omega_n^{-1} \delta_n}{\sqrt{1 - \delta_n^T \Omega_n^{-1} \delta_n}}.$$

Proof. The d -dimensional skew normal random vector $Z_n \sim \text{SN}(\xi_n, \Omega_n, \alpha_n)$ has third cumulant

$$\kappa_3(Z_n) = \sqrt{\frac{2}{\pi}} \left(\frac{4}{\pi} - 1 \right) \delta_n \otimes \delta_n^T \otimes \delta_n,$$

where $\delta_n = \Omega_n \alpha_n / \sqrt{1 + \alpha_n^T \Omega_n \alpha_n}$; see, e.g. De Luca and Loperfido (2012). Since $\kappa_3(Y_n) = n^{-1/2} \kappa_3(X_1)$, for the equality of $\kappa_3(Z_n)$ and $\kappa_3(Y_n)$ we necessarily need $\kappa_3(X_1)$ to be a

third-order tensor of rank at most one and

$$\delta_n = n^{-1/6} \frac{\pi^{1/2}}{(\sqrt{2}(4 - \pi))^{1/3}} \gamma. \tag{5}$$

The first two cumulants of Y_n and Z_n are equal if and only if the first two moments are equal. Since Y_n has expectation 0 and a covariance matrix that equals the identity matrix, we have $\mathbb{E}(Y_n) = \mathbb{E}(Z_n)$ and $\mathbb{E}(Y_n Y_n^\top) = \mathbb{E}(Z_n Z_n^\top)$ if and only if

$$\xi_n + \sqrt{\frac{2}{\pi}} \delta_n = 0 \quad \text{and} \quad \Omega_n - \frac{2}{\pi} \delta_n \delta_n^\top = I;$$

see, e.g. De Luca *et al.* (2006). It remains to show that there exists an α_n such that $\delta_n = \Omega_n \alpha_n / \sqrt{1 + \alpha_n^\top \Omega_n \alpha_n}$ has the form (5). The α_n we are looking for can be defined by

$$\alpha_n = \frac{\Omega_n^{-1} \delta_n}{\sqrt{1 - \delta_n^\top \Omega_n^{-1} \delta_n}}$$

if Ω_n^{-1} exists and $1 - \delta_n^\top \Omega_n^{-1} \delta_n$ is strictly positive. By applying Sylvester’s determinat theorem, the determinant of Ω_n equals $1 + 2/\pi \delta_n^\top \delta_n$, which is strictly positive since $\delta_n^\top \delta_n$ is nonnegative. Also, with the help of Sylvester’s determinat theorem we can show that

$$1 - \delta_n^\top \Omega_n^{-1} \delta_n = \frac{\det(\Omega_n - \delta_n \delta_n^\top)}{\det(\Omega_n)}.$$

Since we already know that $\det(\Omega_n)$ is strictly positive, we have $1 - \delta_n^\top \Omega_n^{-1} \delta_n > 0$ if and only if

$$0 < \det(\Omega_n - \delta_n \delta_n^\top) = \det\left(I + \left(\frac{2}{\pi} - 1\right) \delta_n \delta_n^\top\right) = 1 + \left(\frac{2}{\pi} - 1\right) \delta_n^\top \delta_n,$$

where the second equality is again an application of Sylvester’s determinat theorem. By substituting δ_n with (5), we can show that the positivity of $1 + (2/\pi - 1) \delta_n^\top \delta_n$ is equivalent to (4). This completes the proof.

Note that $(\xi_n, \Omega_n, \alpha_n) \rightarrow (0, I, 0)$ as $n \rightarrow \infty$ and that $\text{SN}(0, I, 0)$ is the d -dimensional standard normal distribution.

Remark 1. The constant in (4) is slightly smaller than 1.01.

Lemma 3. *Under the assumptions of Lemma 2 but with the stronger condition*

$$n > m \|\gamma\|^6 \frac{(\pi - 2)^3}{2(4 - \pi)^2} \tag{6}$$

for an arbitrary but fixed $m > 1$, the characteristic function of $\text{SN}(\xi_n, \Omega_n, \alpha_n)$ has the form

$$e^{-t^\top t/2} \left(1 - i \frac{(t^\top \gamma)^3}{6n^{1/2}} + \frac{(t^\top \gamma)^4}{n^{2/3}} \frac{\pi - 3}{3(\sqrt{2}(4 - \pi))^{4/3}} \right) + R(t, n), \quad t \in \mathbb{R}^d, \tag{7}$$

where the remainder has an upper bound of

$$|R(t, n)| \leq \frac{C}{n^{5/6}} e^{-\varepsilon t^\top t} (\|t\|^5 + \dots + \|t\|^{23})$$

for positive constants C and ε .

Proof. The characteristic function of $\text{SN}(\xi_n, \omega_n, \alpha_n)$ is

$$\Psi_n(t) = e^{it^\top \xi_n - t^\top \Omega_n t / 2} \left(1 + i \int_0^{t^\top \delta_n} \sqrt{\frac{2}{\pi}} e^{u^2/2} du \right), \quad t \in \mathbb{R}^d.$$

Using the representations of ξ_n and Ω_n according to Lemma 2, we obtain

$$\Psi_n(t) = e^{-i\sqrt{2/\pi} t^\top \delta_n} e^{-t^\top t / 2} e^{-(1/\pi)(t^\top \delta_n)^2} \left(1 + i \int_0^{t^\top \delta_n} \sqrt{\frac{2}{\pi}} e^{u^2/2} du \right).$$

Since, for any real number x and nonnegative integer r , we have

$$\left| e^{ix} - \sum_{k=0}^r \frac{(ix)^k}{k!} \right| \leq \frac{|x|^{r+1}}{(r+1)!}$$

(see, e.g. Corollary 8.2 of Bhattacharya and Ranga Rao (1986)), we can show that

$$\begin{aligned} e^{-i\sqrt{2/\pi} t^\top \delta_n} &= 1 - i \left(\frac{2}{\pi} \right)^{1/2} \frac{(t^\top \delta_n)^1}{1} - \left(\frac{2}{\pi} \right)^{2/2} \frac{(t^\top \delta_n)^2}{2} \\ &\quad + i \left(\frac{2}{\pi} \right)^{3/2} \frac{(t^\top \delta_n)^3}{6} + \left(\frac{2}{\pi} \right)^{4/2} \frac{(t^\top \delta_n)^4}{24} + R_1(t, n), \end{aligned}$$

where the remainder $R_1(t, n)$ has an upper bound of

$$|R_1(t, n)| \leq \left(\frac{2}{\pi} \right)^{5/2} \frac{|t^\top \delta_n|^5}{120}.$$

From Taylor’s theorem we obtain

$$\begin{aligned} e^{-(1/\pi)(t^\top \delta_n)^2} &= 1 - \frac{1}{\pi} (t^\top \delta_n)^2 + \frac{1}{2\pi^2} (t^\top \delta_n)^4 + R_2(t, n), \\ \left(1 + i \int_0^{t^\top \delta_n} \sqrt{\frac{2}{\pi}} e^{u^2/2} du \right) &= 1 + i \sqrt{\frac{2}{\pi}} \frac{(t^\top \delta_n)^1}{1} + i \sqrt{\frac{2}{\pi}} \frac{(t^\top \delta_n)^3}{6} + R_3(t, n), \end{aligned}$$

where the remainders $R_2(t, n)$ and $R_3(t, n)$ have upper bounds of

$$\begin{aligned} |R_2(t, n)| &\leq |t^\top \delta_n|^7 + |t^\top \delta_n|^9, \\ |R_3(t, n)| &\leq e^{(t^\top \delta_n)^2/2} (3 + 6|t^\top \delta_n|^2 + |t^\top \delta_n|^4) \sqrt{\frac{2}{\pi}} \frac{|t^\top \delta_n|^5}{120}. \end{aligned}$$

Thus, by multiplying all three series, we can show that

$$\Psi_n(t) = e^{-t^\top t / 2} \left(1 - i \frac{(t^\top \gamma)^3}{6n^{1/2}} + \frac{(t^\top \gamma)^4}{n^{2/3}} \frac{\pi - 3}{3(\sqrt{2}(4 - \pi))^{4/3}} + R(t, n) \right),$$

where the remainder $R(t, n)$ has an upper bound of

$$|R(t, n)| \leq C_1 (|t^\top \delta_n|^5 + \dots + |t^\top \delta_n|^{23}) e^{|t^\top \delta_n|^2/2}$$

for some finite constant C_1 . The definition of δ_n , according to Lemma 2, and the Cauchy–Schwarz inequality lead to

$$|t^\top \delta_n|^2 \leq \|t\|^2 n^{-2/6} \frac{\pi^{2/2}}{(\sqrt{2}(4 - \pi))^{2/3}} \|\gamma\|^2 < \|t\|^2 m^{-2/6}$$

for all n that satisfy (6), and, hence, we have

$$e^{-t^\top t/2} |R(t, n)| \leq \frac{C_3}{n^{5/6}} (\|t\|^5 + \dots + \|t\|^{23}) e^{-(1-m^{-1/3})t^\top t/2}.$$

Since we assumed that m is strictly greater than 1, the difference $(1 - m^{-1/3})$ is strictly positive.

4. Main result

In this section we establish conditions ensuring that the uniform distance, with respect to convex Borel sets, between the distribution function of multivariate skew normal random vectors and the distribution of the sum of n i.i.d. random vectors converges to 0 at a rate of $n^{-2/3}$. The univariate case follows as a special case.

Theorem 1. *Assume that the random vectors X_1, X_2, \dots satisfy Cramer’s condition, have finite fourth-order moments, and a third-order tensor of rank at most one. Then there exists a finite constant $C > 0$ such that, for all convex Borel sets $A \subset \mathbb{R}^d$, we have*

$$|Q_n(A) - \text{SN}(A; \xi_n, \Omega_n, \alpha_n)| \leq \frac{C}{n^{2/3}}$$

for all n that satisfy (6), where $\xi_n, \Omega_n, \alpha_n$, and γ are defined as in Lemma 2.

Proof. Let U_n^1 and U_n^2 be the signed Borel measures on \mathbb{R}^d that have the characteristic functions

$$\begin{aligned} \Upsilon_n^1(t) &= e^{-t^\top t/2} \left(1 - i \frac{(t^\top \gamma)^3}{6n^{1/2}} + \frac{(t^\top \gamma)^4}{n^{2/3}} \frac{\pi - 3}{3(\sqrt{2}(4 - \pi))^{4/3}} \right), \\ \Upsilon_n^2(t) &= e^{-t^\top t/2} \left(1 - i \frac{(t^\top \gamma)^3}{6n^{1/2}} \right), \quad t \in \mathbb{R}^d. \end{aligned}$$

The Fourier inverse of $\Upsilon_n^1(t) - \Upsilon_n^2(t)$ is of the form $n^{-2/3} \exp\{-\|x\|^2/2\} p(x)$, $x \in \mathbb{R}^d$, where $p(x)$ is a multivariate polynomial over the coordinates of x and is of finite order; see, e.g. Bhattacharya and Ranga Rao (1986, Section 2.7). Since this is an integrable function with respect to X , we obtain

$$|U_n^2(A) - U_n^1(A)| \leq \frac{C_1}{n^{2/3}}, \tag{8}$$

where C_1 is a finite constant that does not depend on n and A . In the following we will denote by C_2, C_3, \dots further finite constants, which are also independent of n and A . According to Bhattacharya and Ranga Rao (1986, Equation (20.49)), we also have

$$|Q_n(A) - U_n^2(A)| \leq \frac{C_2}{n^1} \tag{9}$$

for all convex Borel sets A . For every Borel set A and $r \geq 0$, we write

$$A_1 = A \cap \{x \in \mathbb{R}^d : \|x\| \leq r\}, \quad A_2 = A \setminus A_1.$$

Since $|\Upsilon_n^1(t)|$ is integrable on \mathbb{R}^d , the signed measure U_n^1 has a Radon–Nikodym derivative u_n . Similarly, the absolute value of (7) is integrable, and, thus, $\text{SN}(\xi_n, \Omega_n, \alpha_n)$ has a Radon–Nikodym derivative s_n . Hence, by inverse Fourier transformation, we obtain

$$\begin{aligned} |U_n^1(A_1) - \text{SN}(A_1; \xi_n, \Omega_n, \alpha_n)| &\leq \int_{A_1} |u_n(x) - s_n(x)| \, dx \\ &\leq \lambda^d(A_1) \frac{1}{(2\pi)^d} \int |\Upsilon_n^1(t) - \Psi_n^1(t)| \, dt \\ &\leq r^d \frac{C_3}{n^{5/6}}, \end{aligned}$$

where Ψ_n is the characteristic function of $\text{SN}(\xi_n, \Omega_n, \alpha_n)$, and λ^d is the d -dimensional lebesgue measure. The Fourier inverse of $\Upsilon_n^1(t)$ is of the form

$$e^{-\|x\|^2/2} \left(1 + \frac{q(x)}{n^{1/2}} + \frac{p(x)}{n^{2/3}} \right), \quad x \in \mathbb{R}^d,$$

where $q(x)$ and $p(x)$ are multivariate polynomials over the coordinates of x and are of finite order; see, e.g. Bhattacharya and Ranga Rao (1986, Section 2.7). Thus,

$$|U_n^1(A_2)| \leq \int_{\{\|x\|>r\}} e^{-\|x\|^2/2} \left(1 + \frac{q(x)}{n^{1/2}} + \frac{p(x)}{n^{2/3}} \right) \, dx \leq C_4 e^{-r^2/4}$$

for $r \geq r_0$, where r_0 is a finite constant that does not depend on n and A . Let $\phi(x - \xi; \Omega)$ be the PDF of a multivariate normal distribution with expectation ξ and covariance matrix Ω , and let Φ be the CDF of a one-dimensional standard normal distribution. Since ξ_n and Ω_n converge monotonously to 0 and the identity matrix, respectively, we have

$$\begin{aligned} |\text{SN}(A_2; \xi_n, \Omega_n, \alpha_n)| &\leq \int_{\{\|x\|>r\}} 2\phi(x - \xi_n; \Omega_n) \Phi(\alpha_n^\top(x - \xi_n)) \, dx \\ &\leq \int_{\{\|x\|>r\}} 2\phi(x - \xi_n; \Omega_n) \, dx \\ &\leq C_5 e^{-r^2/C_6} \end{aligned}$$

for $r \geq \tilde{r}_0$ and $n \geq \tilde{n}_0$. Thus,

$$|U_n^1(A) - \text{SN}(A; \xi_n, \Omega_n, \alpha_n)| \leq r^d \frac{C_3}{n^{5/6}} + C_4 e^{-r^2/4} + C_5 e^{-r^2/C_6} \tag{10}$$

for all $r \geq \max\{r_0, \tilde{r}_0\}$ and $n \geq \tilde{n}_0$. We now take

$$r = (\max\{4, C_6\} \ln(n))^{1/2} + \max\{r_0, \tilde{r}_0\}.$$

Then, from (8), (9), and (10), we obtain

$$\begin{aligned} |Q_n(A) - \text{SN}(A; \xi_n, \Omega_n, \alpha_n)| &\leq \frac{C_1}{n^{2/3}} + \frac{C_2}{n^1} + (\max\{4, C_6\} \ln(n))^{d/2} \frac{C_3}{n^{5/6}} + \frac{C_4}{n} + \frac{C_5}{n} \\ &\leq \frac{C_7}{n^{2/3}} \end{aligned}$$

for all $n \geq n_0$, where n_0 is a finite constant that does not depend on A . Since the absolute difference $|Q_n(A) - \text{SN}(A; \xi_n, \Omega_n, \alpha_n)|$ is bounded by 2, the above estimate is also true for $n < n_0$ if we replace C_7 by $\max\{2n_0^{2/3}, C_7\}$. This completes the proof.

The proof uses the fact that the characteristic functions of Q_n and $\text{SN}(\xi_n, \Omega_n, \alpha_n)$ are equal up to the terms of order $n^{-1/2}$. The next higher order term in (7) is of order $n^{-2/3}$. We have no

free parameters left for the skew normal distribution, and so we are not able to fit the $n^{-2/3}$ -term to a corresponding term in the characteristic function of Q_n . This means that in general we cannot improve the speed of convergence beyond $n^{-2/3}$.

In the univariate case the rank-one assumption is satisfied for any random variable with finite third moment. Hence, the above theorem simplifies as follows.

Corollary 1. *Assume that the random variables X_1, X_2, \dots satisfy Cramer’s condition and have finite fourth-order moments. Then there exists a finite constant $C > 0$ such that, for all convex Borel sets $A \subset \mathbb{R}$, we have*

$$|Q_n(A) - \text{SN}_1(A; \xi_n, \psi_n, \lambda_n)| \leq \frac{C}{n^{2/3}}$$

for all n that satisfy (6), where $\psi_n = \Omega_n^{1/2}$, $\lambda_n = \alpha_n \psi_n$, and $\xi_n, \Omega_n, \alpha_n, \gamma$ are defined as in Lemma 2.

5. A class of multivariate distributions with rank-one cumulants

Theorem 1 in Section 4 assumes that the third cumulant of the summands is a rank-one tensor. A natural question to ask is whether this assumption is fulfilled for at least some multivariate probability distributions commonly used in probability and statistics. In this section we shall address the problem by considering some well-known statistical models purported to describe financial data. More precisely, we consider models of the form $X = W + \lambda U$, where λ is a d -dimensional real vector, U is a random variable, and W is a d -dimensional, centrally symmetric random vector, meaning that $W - \mathbb{E}(W)$ and $\mathbb{E}(W) - W$ are identically distributed; see Serfling (2006). Both W and U have finite third-order moments and are independent of each other.

Lemma 4. *Let $X = W + \lambda U$, where λ is a d -dimensional real vector, U is a random variable, and W is a d -dimensional, symmetric random vector. Also, let W and Z have finite third-order moments and be stochastically independent of each other. Then the third cumulant of X is a tensor of rank not greater than one.*

Proof. Without loss of generality we can assume that $\mathbb{E}(W)$ is the null vector and that $\mathbb{E}(U) = 0$, so that the third cumulant of X equals its third moment: $\mu_3(X) = \kappa_3(X)$. The i th cumulant of the sum of independent random vectors is the sum of the i th cumulants of the random vectors themselves. Hence, the third cumulant of X is $\kappa_3(X) = \kappa_3(W) + \kappa_3(\lambda U)$. By assumption, W and $-W$ are identically distributed, implying that the third cumulant of W is a null tensor and $\kappa_3(X) = \kappa_3(\lambda U)$. We now apply Lemma 1 to write $\kappa_3(\lambda U) = (\lambda \otimes \lambda) \kappa_3(U) \lambda^\top$. The same basic properties of the tensor product which we recalled in the proof of Lemma 1 lead to $\kappa_3(X) = \kappa_3(U) \lambda \otimes \lambda^\top \otimes \lambda$. Since $\kappa_3(U)$ is a scalar, $\kappa_3(X)$ is either of rank one or zero. The latter case occurs when either λ is a null vector or $\kappa_3(U)$ is 0. This completes the proof.

We shall now consider three examples of multivariate probability distributions fulfilling the assumptions of Lemma 4.

Example 1. (*Generalized skew normal distribution.*) The distribution of a d -dimensional random vector X is extended skew normal with location parameter ξ , scale parameter Ω , shape parameter δ , and normalizing parameter τ , that is, $X \sim \text{ESN}(\xi, \Omega, \delta, \tau)$ if the PDF of X is

$$f(x; \xi, \Omega, \delta, \tau) = \frac{\phi(x; \xi, \Omega)}{\Phi(\tau)} \Phi\left(\frac{\tau + \delta^\top \Omega^{-1}(x - \xi)}{\sqrt{1 - \delta^\top \Omega^{-1} \delta}}\right),$$

where $x, \xi, \delta \in \mathbb{R}^d$, $\tau \in \mathbb{R}$, Ω is a $d \times d$ symmetric, positive definite matrix, and $\delta^T \Omega^{-1} \delta < 1$; see, e.g. Arnold and Beaver (2002). The extended skew normal distribution fulfills the assumptions of Lemma 4, with W being a d -dimensional normal vector and U being a truncated normal random variable; see, e.g. Adcock (2010). Adcock (2007) thoroughly discusses financial applications of the generalized skew normal distribution.

Example 2. (*Normal gamma distribution.*) The distribution of a d -dimensional random vector X is normal gamma with parameters $\mu \in \mathbb{R}^d$, Σ a $d \times d$ positive definite matrix, $\lambda \in \mathbb{R}_0^d$, $\tau \in \mathbb{R}$, and $\nu \in \mathbb{R}_+$, that is, $X \sim \text{MNG}(\mu, \Sigma, \lambda, \tau, \nu)$, if the PDF of X is

$$f(x; \mu, \Sigma, \lambda, \nu) = \frac{\mathbb{E}(Y^\nu)}{\Gamma(\nu)} \Phi(\tau) \sqrt{\frac{\exp[\tau^2 - (x - \mu)^\top \Sigma^{-1} (x - \mu)]}{(\lambda^\top \Sigma^{-1} \lambda)^\nu (2\pi)^{n-1} |\Sigma|}},$$

where Y has the normal distribution $N(\tau, 1)$ truncated from below at 0, $\Gamma(\nu)$ is the gamma function evaluated at ν , and $\tau = [\lambda^\top \Sigma^{-1} (x - \mu) - 1] / \sqrt{\lambda^\top \Sigma^{-1} \lambda}$. The normal gamma distribution fulfills the assumptions of Lemma 4, when W is chosen to be a d -dimensional normal vector with mean μ and variance Σ , while U is a gamma random variable with scale parameter 1 and shape parameter ν ; see Adcock and Shutes (2012).

Example 3. (*Finite mixtures of normals.*) Consider a d -dimensional random vector X whose distribution is either $N(\mu_1, \Omega)$ with probability π_1 or $N(\mu_2, \Omega)$ with probability $\pi_2 = 1 - \pi_1$, that is a normal mixture with two homoscedastic components; see, e.g. McLachlan and Peel (2000). Again, this model fulfills the assumptions of Lemma 4, with $W \sim N(\mu_1, \Omega)$, $\lambda = \mu_2 - \mu_1$, and U satisfying $\mathbb{P}(U = 0) = 1 - \mathbb{P}(U = 1) = \pi_1$. Applications in financial econometrics of normal mixture models are reviewed in Haas *et al.* (2009).

6. A simple numerical example

Suppose that X_1, X_2, \dots take values in \mathbb{R}^1 . Then the third cumulant of X_1 always has a rank of at most one since it is a 1×1 matrix. In order to apply Theorem 1 it is sufficient to assume finite fourth-order moments and Cramer’s condition.

We now assess the adequacy of the skew normal approximation to the distribution of $S_n = X_1 + \dots + X_n$ when X_1, X_2, \dots are i.i.d. exponential random variables with PDF

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0, \end{cases}$$

where $\alpha > 0$. The exponential distribution is both very skewed and leptokurtic: its third and fourth standardized moments are 2 and 9, respectively. Hence, the normal approximation to the distribution of S_n might be unsatisfactory when the number of summands is small. We shall overcome the problem by approximating the distribution of S_n with a skew normal distribution with the same mean, variance, and skewness.

The distribution of S_n is gamma $\text{Ga}(\alpha, n)$ with PDF

$$f_n(x) = \begin{cases} \frac{\alpha^n x^{n-1} e^{-\alpha x}}{\Gamma(n)} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0. \end{cases}$$

Its mean, variance, and skewness are $\mu_n = n/\alpha$, $\sigma_n^2 = n/\alpha^2$, and $\gamma_{1,n} = 2/\sqrt{n}$, respectively. The approximating distribution is $SN(\xi_n, \psi_n, \lambda_n)$, where

$$\xi_n = \frac{n - \sqrt{nq_n}}{\alpha}, \quad \psi_n = \frac{\sqrt{n(1 + q_n)}}{\alpha}, \quad \lambda_n = \sqrt{\frac{\pi q_n}{2 + 2q_n - \pi q_n}}, \quad q_n = \sqrt[3]{\frac{16}{n(4 - \pi)^2}}.$$

Straightforward but tedious calculations show that the mean, variance, and skewness of $Ga(\alpha, n)$ equal those of $SN(\xi_n, \psi_n, \lambda_n)$. The lower bound (4) is smaller than $4.04/\alpha^6$.

Figures 1 and 2 graphically display the skew normal approximation to $Ga(3, 5)$. Figure 1 contains the PDF graphs of $Ga(3, 5)$, $SN(\xi_5, \psi_5, \lambda_5)$, and $N(\mu_5, \sigma_5^2)$. Figure 2 plots the CDF of $Ga(3, 5)$ versus the CDF of $SN(\xi_5, \psi_5, \lambda_5)$. Both figures show that the skew normal approximation to the gamma distribution is quite satisfactory even when the number of summands is very small. In particular, Figure 1 shows that the skew normal approximation performs better than the normal approximation, especially in the tails. Following Chang *et al.* (2008), we also compute the Kolmogorov distance between $Ga(3, 5)$ and $SN(\xi_5, \psi_5, \lambda_5)$, i.e. the maximum absolute difference between their CDFs. It was found to be 0.0201, which is quite small and consistent with the above mentioned visual findings. The Kolmogorov distance between $Ga(3, 5)$ and $N(\mu_5, \sigma_5^2)$ is almost three times bigger, being equal to 0.0596.

The skew normal approximation to the gamma distribution significantly improves when a few summands are added, as shown in Figures 3 and 4. Figure 3 contains the PDF graphs of $Ga(3, 8)$, $SN(\xi_8, \psi_8, \lambda_8)$, and $N(\mu_8, \sigma_8^2)$. Figure 4 plots the CDF of $Ga(3, 8)$ versus the CDF of $SN(\xi_8, \psi_8, \lambda_8)$. The Kolmogorov distance between the $Ga(3, 8)$ and $SN(\xi_8, \psi_8, \lambda_8)$ is 0.0095,

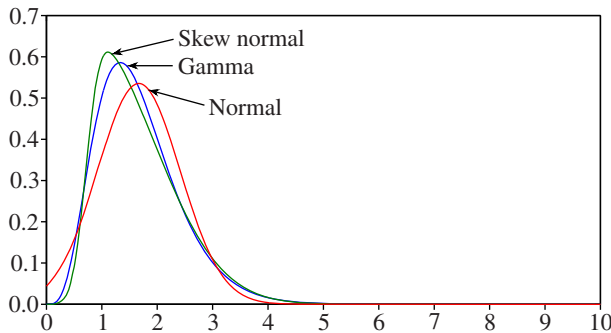


FIGURE 1: PDFs of $Ga(3, 5)$, $SN(\xi_5, \psi_5, \lambda_5)$, and $N(\mu_5, \sigma_5^2)$.

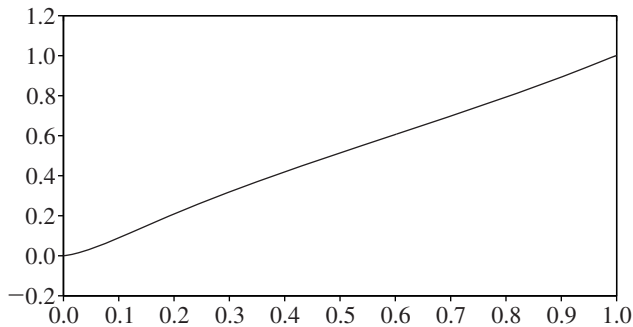


FIGURE 2: The CDF of $Ga(3, 5)$ versus the CDF of $SN(\xi_5, \psi_5, \lambda_5)$.

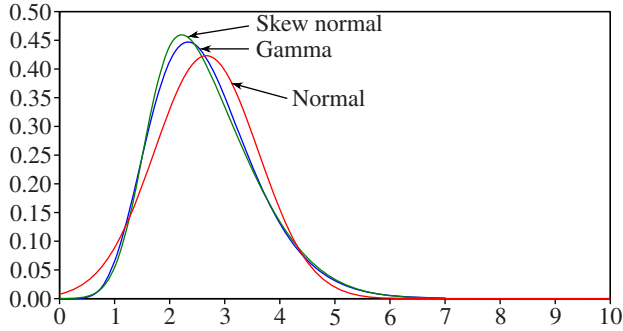


FIGURE 3: PDFs of $Ga(3, 8)$, $SN(\xi_8, \psi_8, \lambda_8)$, and $N(\mu_8, \sigma_8^2)$.

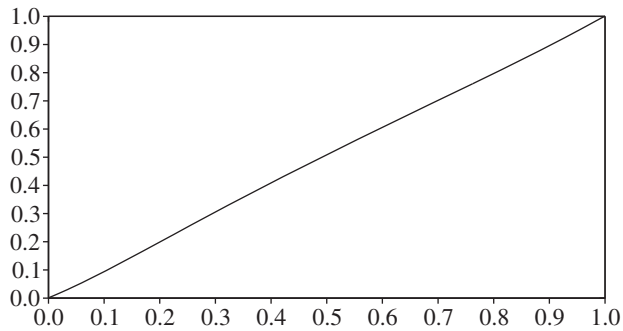


FIGURE 4: The CDF of $Ga(3, 8)$ versus the CDF of $SN(\xi_8, \psi_8, \lambda_8)$.

which is less than half the value in the previous case. On the other hand, the Kolmogorov distance between $Ga(3, 8)$ and $N(\mu_8, \sigma_8^2)$ has decreased by a much smaller proportion, being equal to 0.0471.

7. A simulation study

In this section we use Monte-Carlo simulation to compare the performance of the proposed approximation with the normal approximation, when the sample mean comes from a lognormal distribution $X = \exp(Z)$, where Z is a standard normal random variable. Well-known properties of the lognormal distribution imply that the mean, variance, skewness, and kurtosis of X are $\mu = \sqrt{e} \approx 1.6487$, $\sigma^2 = (e - 1)e \approx 4.6708$, $\gamma = (e + 2)\sqrt{e - 1} \approx 6.1849$, and $\beta = e^4 + 2e^3 + 3e^2 - 6 \approx 110.94$, respectively. Since X is very skewed and leptokurtic, and much more so than the exponential distribution, many summands are needed to obtain a satisfactory approximation. Unlike the exponential distribution, however, the PDF convolution of i.i.d. lognormal variables does not have a simple analytical form. Hence the need for simulation.

Simple but tedious calculations show that the skew normal approximation to the distribution of the standardized mean $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ of n independent variables distributed as X is $SN(\xi_n, \psi_n, \lambda_n)$, where

$$\xi_n = -\sqrt{\frac{2}{\pi}} \tilde{\gamma} \psi_n, \quad \psi_n = \sqrt{\frac{\pi}{\pi - 2\delta_n}}, \quad \lambda_n = \frac{\delta_n}{\sqrt{1 - \delta_n^2}},$$

TABLE 1: Results of the simulation study.

Number	Mean		Median		Maximum	
	Skew	Normal	Skew	Normal	Skew	Normal
50	16.216	26.512	13.624	26.338	34.008	49.364
100	6.924	19.616	6.837	18.673	13.760	36.636
150	4.052	17.048	3.859	16.160	7.848	32.960
200	3.075	14.764	2.882	14.076	6.296	27.797
250	2.776	12.434	2.126	12.323	7.066	23.359
300	2.273	11.668	1.674	11.116	5.954	21.194
350	1.705	11.523	1.634	10.765	3.989	21.550

and

$$\delta_n = \frac{\sqrt{\pi/2} \sqrt[3]{2\gamma/\sqrt{n}}}{\sqrt{\sqrt[3]{(4-\pi)^2} + \sqrt[3]{4\gamma^2/n}}}$$

We shall denote by $\Psi_n(x)$ the CDF of $SN(\xi_n, \psi_n, \lambda_n)$ evaluated at the real value x .

We simulate 100 000 samples of size $n = 50, 100, 150, 200, 250, 300, 350$ from X ; see Table 1. We compute the empirical CDF F_n of the standardized means $Y_i = \sqrt{n}(\bar{X}_{n,i} - \mu)/\sigma$, where $\bar{X}_{n,i}$ is the mean of the n observations in the i th sample, for $i = 1, \dots, 100\,000$. Finally, for each n , we compute the mean, the median, and the maximum of the absolute differences $|F_n(Y_i) - \Phi(Y_i)|$ and $|F_n(Y_i) - \Psi_n(Y_i)|$. Table 1 provides the simulations' results, multiplied by 1000 for better readability. As expected, both approximations improve with the sample size. However, the skew normal approximation always performs better than the normal approximation. The difference is more marked when considering the mean and the medians of the absolute differences.

8. An empirical example

In this section we shall use a real data set to illustrate the practical relevance of the proposed approximation.

Our application deals with particulate matter with an aerodynamic equivalent diameter of up to $10 \mu m$, commonly referred to as PM_{10} , whose harmful effects on human health are well documented ; see, e.g. Brunekreef and Holgate (2002). The data are annual means of daily PM_{10} concentrations collected by the Italian Institute for Environmental Protection and Research (ISPRA) during the year 2010 from 205 monitoring stations. All the selected stations were located in rural, residential, or natural areas, in order to detect contributions to PM_{10} distinct from urban and industrial activities. Concentrations are expressed in micrograms per cubic meter ($\mu g/m^3$) and then rounded to the nearest integer, without appreciable loss of accuracy.

Modelling the data by the normal distribution appears like a natural choice, motivated by the central limit theorem, since each is the mean of 365 daily values, which in turn are the means of intra-daily observations. However, a normality test based on the sample skewness, which equals -0.4180 , rejects the normality hypothesis at the 0.01 level. A possible explanation for the negative skewness is that locations of the monitoring stations leads to measurements which are biased towards low values of PM_{10} . A similar skewing mechanism, with the necessary modifications, was discussed by Loperfido and Guttorp (2008) with respect

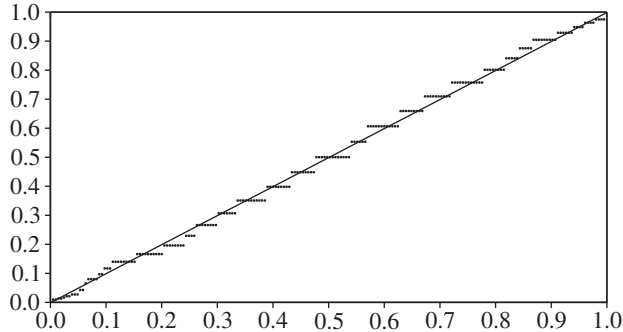


FIGURE 5: PP plot showing the empirical cumulative distribution versus the CDF of the approximating skew normal distribution (*dotted line*), and the CDF of the approximating skew normal distribution versus itself (*solid line*).

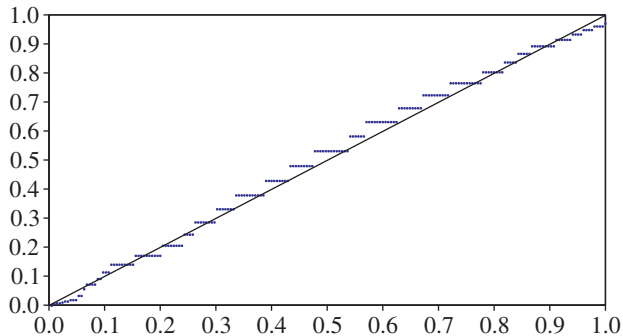


FIGURE 6: PP plot showing the empirical cumulative distribution versus the CDF of the approximating normal distribution (*dotted line*), and the CDF of the approximating normal distribution versus itself (*solid line*).

to networks primarily aimed at finding large values of air pollution. Apparently, the daily data are too skewed and the sample size too small for the normal approximation to be appropriate.

The Kolmogorov–Smirnov statistics for the skew normal and the normal fits are 0.0414 and 0.0648, respectively. Their values were obtained by substituting the true parameter values with their maximum likelihood estimates, as done by Mateu-Figueras *et al.* (2007) who also gave tables for the relevant quantiles of the sampling distributions of the above test statistics. Using these tables, we found that the p -values for the skew normal fit and the normal fit were higher than 0.25 and lower than 0.01, respectively. Visual comparison of the PP plots in Figure 5 (for the skew normal approximation) and Figure 6 (for the normal approximation) lead to similar results. We conclude that the skew normal approximation, motivated by theorems in the previous sections, improves upon the normal approximation, motivated by the central limit theorem.

9. Conclusion

We showed that sums of i.i.d. random variables that are skewed, satisfy Cramer’s condition, and have finite fourth-order moments are better approximated by skew normal distributions than normal distributions. We proved that the minimal speed of convergence with respect to

the uniform distance increases from $n^{-1/2}$ to $n^{-2/3}$. The results hold true for the multivariate case, under the additional rank-one assumption. Rank-one tensors are a fundamental concept in numerical linear algebra, where they are often used to approximate tensors of the same dimension but with a more complicated structure; see Kofidis and Regalia (2002), Zhang and Golub (2002), Wang and Ahuja (2004), and Qi (2011). It is interesting to note that rank-one tensors play a role in probability theory too. At the present time, we are unable to prove that the skew normal approximation is optimal. Rather, we conjecture that a distribution allowing for wider ranges of skewness and kurtosis might offer a better approximation.

A relevant question for future research is whether the i.i.d. assumption can be weakened to nonidentical random variables and/or dependent random variables. Such a result would provide a unified theoretical explanation for some well-known stylized features of financial time series. As an example, consider aggregational gaussianity, meaning that data taken at decreasing sampling frequencies appear to be increasingly normal; see Cont (2001) and Rydberg (2000). This feature is often explained by the central limit theorem. On the other hand, the multivariate skew normal distribution satisfactorily models the skewness of multivariate financial returns; see De Luca and Loperfido (2012) and Franceschini and Loperfido (2010). These empirical features lead us to conjecture that results in this paper might be extended to random vectors belonging to a multivariate stationary process under assumptions we are currently investigating.

Another interesting research topic concerns the way in which the adequacy of the approximation is measured. In this paper we measured it by the absolute value of the difference between the approximating and the approximated distribution, following Bhattacharya and Ranga Rao (1986) and Chang *et al.* (2008). However, there are cases where the relative difference between the two distributions appear to be more appropriate. The Kullback–Leibler distance formalizes this concept, both in the univariate and in the multivariate case. Also, Van Hulle (2005) showed how the third cumulant might be used to approximate the Kullback–Leibler distance between a normal and another distribution with finite fourth-order moments. It would then be interesting to investigate whether the skew normal approximation would be a significant improvement over the normal approximation when the adequacy of the approximation is measured by the Kullback–Leibler distance.

Acknowledgement

The authors would like to thank an anonymous referee for her/his useful comments, which significantly improved the quality of the present paper.

References

- ADCOCK, C. J. (2007). Extensions of Stein's lemma for the skew-normal distribution. *Commun. Statist. Theory Meth.* **36**, 1661–1671.
- ADCOCK, C. J. (2010). Asset pricing and portfolio selection based on the multivariate extended skew-Student-*t* distribution. *Ann. Operat. Res.* **176**, 221–234.
- ADCOCK, C. J. AND SHUTES, K. (2012). On the multivariate extended skew-normal, normal-exponential and normal-gamma distributions. *J. Statist. Theory Practice* **6**, 636–664.
- ARNOLD, B. C. AND BEAVER, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11**, 7–54.
- AZZALINI, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.
- AZZALINI, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.* **32**, 159–200.
- AZZALINI, A. (2006). Some recent developments in the theory of distributions and their applications. *Atti della XLIII Riunione Scientifica della Società Italiana di Statistica*, 51–64.
- AZZALINI, A. AND DALLA VALLE, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–726.

- BARTOLETTI, S. AND LOPERFIDO, N. (2010). Modelling air pollution data by the skew-normal distribution. *Stoch. Environ. Res. Risk Assess.* **24**, 513–517.
- BHATTACHARYA, R. N. AND RANGA RAO, R. (1986). *Normal Approximation and Asymptotic Expansions*. Robert E. Krieger, Malabar, FL.
- BRACHAT, J., COMON, P., MOURRAIN, B. AND TSIGARIDAS, E. (2010). Symmetric tensor decomposition. *Linear Algebra Appl.* **433**, 1851–1872.
- BRAMAN, K. (2010). Third-order tensors as linear operators on a space of matrices. *Linear Algebra Appl.* **433**, 1241–1253.
- BRUNEKREEFF, B. AND HOLGATE, S. T. (2002). Air pollution and health. *Lancet* **360**, 1233–1242.
- CHANG, C.-H., LIN, J.-J., PAL, N. AND CHIANG, M.-C. (2008). A note on improved approximation of the binomial distribution by the skew-normal distribution. *Amer. Statistician* **62**, 167–170.
- COMON, P., GOLUB, G., LIM, L.-H. AND MOURRAIN, B. (2008). Symmetric tensors and symmetric tensor rank. *SIAM J. Matrix Anal. Appl.* **30**, 1254–1279.
- CONT, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**, 223–236.
- DE LUCA, G. AND LOPERFIDO, N. (2012). Modelling multivariate skewness in financial returns: a SGARCH approach. *European J. Finance*, DOI:10.1080/1351847X.2011.640342.
- DE LUCA, G., GENTON, M. AND LOPERFIDO, N. (2006). A multivariate skew-GARCH Model. In *Econometric Analysis of Economic and Financial Time Series, Part A* (Adv. Econometrics **20**), eds T. B. Formby, D. Terrell and R. Carter Hill, JAI Press Inc., Bingley, UK, pp. 33–56.
- FIELD, C. A. AND RONCHETTI, E. (1990). *Small Sample Asymptotics*. Institute of Mathematical Statistics – Monograph Series, Hayward, CA.
- FRANCESCHINI, C. AND LOPERFIDO, N. (2010). A skewed GARCH-type model for multivariate financial time series. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance XII*, eds M. Corazza and C. Pizzi, Springer, Milan, pp. 143–152.
- GUPTA, A. K. AND KOLLO, T. (2003). Density expansions based on the multivariate skew-normal distribution. *Sankhya* **65**, 821–835.
- HAAS, M., MITTNIK, S. AND PAOLELLA, M. S. (2009). Asymmetric multivariate normal mixture GARCH. *Comput. Statist. Data Anal.* **53**, 2129–2154.
- KILMER, M. E. AND MARTIN, C. D. (2011). Factorization strategies for third-order tensors. *Linear Algebra Appl.* **435**, 641–658.
- KOFIDIS, E. AND REGALIA, P. A. (2002). On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM J. Matrix Anal. Appl.* **23**, 863–884.
- KOLDA, T. G. AND BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500.
- KOTZ, S. AND VICARI, D. (2005). Survey of developments in the theory of continuous skewed distributions. *Metron* **63**, 225–261.
- LOPERFIDO, N. AND GUTTORP, P. (2008). Network bias in air quality monitoring design. *Environmetrics* **19**, 661–671.
- MARDIA, K. V., KENT, J. T. AND BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, London.
- MATEU-FIGUERAS, G., PUIG, P. AND PEWSEY, A. (2007). Goodness-of-fit tests for the skew-normal distribution when the parameters are estimated from the data. *Commun. Statist. Theory Meth.* **36**, 1735–1755.
- MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall, London.
- McLACHLAN, G. AND PEEL, D. (2000). *Finite Mixture Models*. John Wiley, New York.
- QI, L. (2011). The best rank-one approximation ratio of a tensor space. *SIAM J. Matrix Anal. Appl.* **32**, 430–442.
- QI, L., SUN, W. AND WANG, Y. (2007). Numerical multilinear algebra and its applications. *Frontiers Math. China* **2**, 501–526.
- RYDBERG, T. H. (2000). Realistic statistical modelling of financial data. *Internat. Statist. Rev.* **68**, 233–258.
- SERFLING, R. J. (2006). Multivariate symmetry and asymmetry. In *Encyclopedia of Statistical Sciences*, 2nd edn, eds S. Kotz, C. B. Read, N. Balakrishnan and B. Vidakovic, Wiley, New York.
- VAN HULLE, M. M. (2005). Edgeworth approximation of multivariate differential entropy. *Neural Computation* **17**, 1903–1910.
- WANG, H. AND AHUJA, N. (2004). Compact representation of multidimensional data using tensor rank-one decomposition. In *Pattern Recognition, 2004*. (Proc. 17th Internat. Conf. Pattern Recognition **1**), IEEE, Red Hook, NY, pp. 44–47.
- ZHANG, T. AND GOLUB, G. H. (2001). Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.* **23**, 534–550.