The use of Isotones for comparing Tests of Normality against Skew Normal Distributions

Florian Ketterer, *Institut für Stochastik,*
Universität Karlsruhe (TH), Englerstr. 2, 76128 Karlsruhe, Germany.
Email: Florian.Ketterer@gmx.de

Bernhard Klar[1], *Institut für Stochastik,*
Universität Karlsruhe (TH), Englerstr. 2, 76128 Karlsruhe, Germany.
Email: Bernhard.Klar@stoch.uni-karlsruhe.de

Norbert Henze, *Institut für Stochastik,*
Universität Karlsruhe (TH), Englerstr. 2, 76128 Karlsruhe, Germany.
Email: Norbert.Henze@stoch.uni-karlsruhe.de

Abstract

The problem of testing for multivariate normality has received much attention. Among the myriad of tests available, we confine ourselves to three affine invariant and simple to implement tests. In order to compare the power of these tests against skew-normal distributions we use Monte Carlo simulations and isotones, a graphical device introduced by Mudholkar *et al.* (J. Roy. Statist. Soc. B 53, 1991, 221–232). To this end, we generalize the notion of a profile, a deterministic ideal sample used to construct isotones, to the bivariate case.

AMS Subject Classifications: 62G10 and 62G30

Key words: Profiles, Isotones, Simulation, Bivariate Skew-Normal distribution, Goodness-of-fit test.

1 Introduction

Let $X_1, \ldots, X_n$ be independent copies of a d-dimensional random vector X, where $d \geq 1$ is a fixed integer. We assume that the distribution $\mathbb{P}^X$ of X has a Lebesgue density but is otherwise unknown. Writing

$$\mathcal{N}_d := \{\mathcal{N}_d(\mu, \Sigma) : \mu \in \mathbb{R}^d \text{ and } \Sigma \in \mathbb{R}^{d,d} \text{ nonsingular}\}$$

for the class of all nondegenerate d-variate normal distributions, this paper considers the use of profiles, introduced by Mudholkar *et al.* ([13])) in the case $d = 1$, for testing the hypothesis

$$H_0 : \mathbb{P}^X \in \mathcal{N}_d$$

against general alternatives. The problem of testing for multivariate normality has received much attention (for an overview, see e.g. [11])). Among the myriad of tests available, we confine ourselves to the tests of Mardia based on skewness and kurtosis and the BHEP tests (see Section 2). All these tests are affine invariant and simple to implement. Moreover, the BHEP tests are consistent against each fixed nonnormal alternative distribution. The paper is organized as

[1]Corresponding Author

follows. In Section 2, we introduce the test statistics under discussion. Section 3 deals with profiles, their generalization to more than one dimension and isotones, a tool for the graphical comparison of the power of goodness-of-fit tests. In Section 4, the tests of normality against skew-normal alternatives are compared using isotones and Monte Carlo simulations. The final section contains some concluding remarks.

2 Test statistics

In what follows, $\bar{X}_n = n^{-1}\sum_{j=1}^n X_j$ and $S_n = n^{-1}\sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)'$ denote the mean vector and the covariance matrix of $X_1, \ldots, X_n$, respectively. All vectors are column vectors, and the prime denotes transpose. We make the tacit assumption $n > d$ which entails the almost sure invertibility of S_n (see [4]). The scaled residuals are denoted by $Y_{n,j} = S_n^{-1/2}(X_j - \bar{X}_n)$, $j = 1, \ldots, n$. Here, $S_n^{-1/2}$ is the symmetric positive definite square root of S_n^{-1}.

2.1 Mardia's test based on multivariate skewness

Multivariate sample skewness in the sense of Mardia ([12]), defined as

$$b_{1,d}^{(n)} = \frac{1}{n^2}\sum_{j,k=1}^{n}\{Y_{n,j}'Y_{n,k}\}^3,$$

is a consistent estimator of the underlying population parameter

$$\beta_{1,d} = \mathbb{E}\{(X_1-\mu)'\Sigma^{-1}(X_2-\mu)\}^3.$$

Here, μ and Σ denote the mean and the covariance matrix of X, respectively. A test for multivariate normality based on $b_{1,d}^{(n)}$ rejects H_0 for large values of

$$T_{n,S} = \frac{nb_{1,d}^{(n)}}{6}.$$

Mardia [12] showed that the limit null distribution of $T_{n,S}$ is $\chi^2_{d(d+1)(d+2)/6}$. More general results on the asymptotic behavior of $T_{n,S}$ have been obtained by Baringhaus and Henze ([3]). Since the convergence of quantiles of the finite sample H_0-distribution of $T_{n,S}$ is fairly slow, we use empirical quantiles of $T_{n,S}$ based on 100 000 replications.

2.2 Mardia's test based on multivariate kurtosis

Multivariate sample kurtosis in the sense of Mardia ([12]) is defined as

$$b_{2,d}^{(n)} = \frac{1}{n}\sum_{j=1}^{n}\{Y_{n,j}'Y_{n,j}\}^2.$$

Obviously, $b_{2,d}^{(n)}$ is a consistent estimator of the population parameter $\beta_{2,d} = \mathbb{E}\{(X-\mu)'\Sigma^{-1}(X-\mu)\}^2$. A test of H_0 using $b_{2,d}^{(n)}$ is based on

$$T_{n,K} = \sqrt{n}\,\frac{b_{2,d}^{(n)} - d(d+2)}{\sqrt{8d(d+2)}}$$

with a two-sided rejection region. The limit null distribution of $T_{n,K}$ is standard normal ([12]). The limit law of $T_{n,K}$ under general alternatives was obtained in ([9]). Like for $T_{n,S}$, we also used Monte Carlo quantiles of $T_{n,K}$ based on 100 000 replications.

2.3 The BHEP test

Writing $\psi_n(t) = n^{-1}\sum_{j=1}^n \exp(it'Y_{n,j})$, $t \in \mathbb{R}^d$, for the empirical characteristic function of the scaled residuals and $\psi(t) = \exp(-\|t\|^2/2)$ for the characteristic function of the standard normal distribution in $\mathbb{R}^d$, the BHEP test rejects H_0 for large values of

$$T_{n,\beta} \;=\; \int_{\mathbb{R}^d} \left| \psi_n(t) - \exp\left(-\frac{\|t\|^2}{2}\right)\right|^2 \varphi_{d,\beta}(t)dt.$$

Here,

$$\varphi_{d,\beta}(t) \;=\; (2\pi\beta^2)^{-\frac{d}{2}} \exp\left(-\frac{\|t\|^2}{2\beta^2}\right), \quad t \in \mathbb{R}^d,$$

denotes the density of the distribution $\mathcal{N}_d(0,\beta^2 I_d)$ (I_d is the unit matrix of order d), and $\beta > 0$ is a fixed parameter (Henze and Zirkler ([8])). The limit null distribution of $nT_{n,\beta}$ is the distribution of an infinite sum of weighted independent Chi square variates ([8]). Henze and Wagner ([10]) and Gürtler ([5]) obtained the limit law of $T_{n,\beta}$ under contiguous alternatives and fixed alternatives to normality, respectively.

The test statistic can be written as

$$\begin{aligned} T_{n,\beta} \;=\; & \frac{1}{n^2}\sum_{j,k=1}^n \exp\left(-\frac{\beta^2}{2}\|Y_{n,j} - Y_{n,k}\|^2\right) \\ & -\frac{2}{n}(1+\beta^2)^{-d/2}\sum_{j=1}^n \exp\left(-\frac{\beta^2\|Y_{n,j}\|^2}{2(1+\beta^2)}\right) + (1+2\beta^2)^{-d/2}, \end{aligned}$$

which shows that a test of H_0 based on $T_{n,\beta}$ can be easily carried out. An alternative expression of $T_{n,\beta}$ as an L^2-distance between density estimators is given in [8]). In what follows, we choose the value $\beta = 1$. Empirical finite sample quantiles of $T_{n,1}$ were obtained by simulations (100 000 replications).

3 Profiles and isotones

Mudholkar *et al.* [13] propose a graphical procedure that allows of a comparison of competing tests for univariate normality against alternatives from a two-parametric family of distributions that includes the normal law. We briefly review this procedure and generalize the approach to the bivariate case.

The basic idea is to find a non-stochastic sample of a distribution that, in a certain sense, is optimal with regard to all possible samples of the same size. To this end, let F be a (univariate) distribution function. Following Mudholkar *et* *al.* [13], a profile from F is a sequence $(P_n)_{n\geq 1}$ of sets $P_n = \{x_{n1}, \ldots, x_{nn}\} \subset \mathbb{R}$ such that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\hat{F}_{n,P_n}(x) - F(x)| = 0.$$

Here,

$$\hat{F}_{n,P_n}(x) \;=\; \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\{x_{nj} \leq x\}, \quad x \in \mathbb{R},$$

is the empirical distribution function corresponding to the 'size n profile' P_n.

Writing $F^{-1}(p) = \inf\{x \in \mathbb{R} \,|\, F(x) \geq p\}$, $0 < p < 1$, for the quantile function of F, Mudholkar et al. [13] choose

$$P_n^* = P_n^*(F) = \left\{ F^{-1}\left(\frac{j-0.5}{n}\right) : j = 1, \ldots, n \right\}$$

as a size n profile if the distribution function F is continuous. Without proof, they state the following optimality properties of P_n^*: If P_n is any other size n profile of F, then

$$\sup_{x\in\mathbb{R}} |\hat{F}_{n,P_n}(x) - F(x)| \;\geq\; \sup_{x\in\mathbb{R}} |\hat{F}_{n,P_n^*}(x) - F(x)| \tag{1}$$

and

$$\mathbb{E}_F[\hat{F}_{n,P_n}(X) - F(X)]^2 \;\geq\; \mathbb{E}_F[\hat{F}_{n,P_n^*}(X) - F(X)]^2. \tag{2}$$

Here, X denotes a random variable having distribution function F.

The first property follows immediately from the observation that, for any continuous distribution function F, $\sup_{x\in\mathbb{R}} |\hat{F}_{n,P_n^*}(x) - F(x)| = 1/(2n)$.

Regarding assertion (2), note that

$$\mathbb{E}_F \hat{F}^2_{n,P_n}(X) = 1 - \frac{1}{n^2}\sum_{j=1}^{n}(2j-1)F(x_j),$$

$$\mathbb{E}_F\, \hat{F}_{n,P_n}(X)F(X) = \frac{1}{2} - \frac{1}{2n}\sum_{j=1}^{n} F(x_j)^2,$$

where $x_1 < x_2 < \ldots < x_n$ is any ordered size n profile of F. Therefore, putting $\xi_j = F(x_j)$ for $j = 1, \ldots, n$, we obtain

$$\mathbb{E}_F \hat{F}^2_{n,P_n}(X) - 2\,\mathbb{E}_F\, \hat{F}_{n,P_n}(X)F(X) = \frac{1}{n}\sum_{j=1}^{n}\xi_j^2 - \frac{1}{n^2}\sum_{j=1}^{n}(2j-1)\xi_j. \tag{3}$$

Minimizing the right hand side of (3) subject to the constraints $0 \leq \xi_1 < \ldots < \xi_n \leq 1$ yields $\xi_j = (j - 1/2)/n$, from which the second optimality property follows.

We now generalize the notion of a distribution profile to the bivariate case. To this end, let F be a bivariate distribution function. By analogy with the univariate case, a profile from F is a sequence $(P_{n,2})_{n\geq 1}$ of sets

$$P_{n,2} = P_{n,2}(F) = \{(x_{n1}, y_{n,1}), \ldots, (x_{nn}, y_{nn})\} \subset \mathbb{R}^2$$

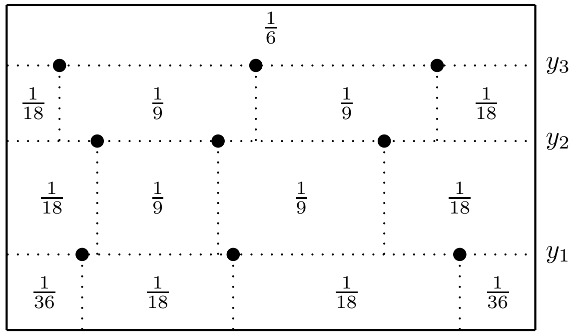

Figure 1: Distribution of mass of a bivariate size 9 profile $\tilde{P}_{9,2}$.

such that

$$\lim_{n\to\infty} \sup_{(x,y)\in\mathbb{R}^2} |\hat{F}_{n,P_{n,2}}(x,y) - F(x,y)| = 0,$$

where

$$\hat{F}_{n,P_{n,2}}(x,y) \; = \; \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\{x_{nj} \le x, y_{nj} \le y\}$$

$(x, y \in \mathbb{R})$ is the empirical distribution function of $(x_{n1}, y_{n1}), \ldots, (x_{nn}, y_{nn})$. The set $P_{n,2}$ is called a bivariate size n profile of F.

It is an open problem to find optimal bivariate profiles satisfying (1) or (2). However, the following procedure for obtaining bivariate profiles suggests itself. It is tailored to the case that n is a square number, i.e., $n = m^2$ for some integer m, and that F has a Lebesgue density. The idea is to generate a partition of $\mathbb{R}^2$ into n cells of equal probability. To be specific, let $F(x,y) = \mathbb{P}(X \le x, Y \le y)$ be the distribution function of a bivariate vector (X,Y) and $F_2(y) = \mathbb{P}(Y \le y)$ be the marginal distribution function of Y.

We put $y_k = F_2^{-1}\left(\frac{k-0.5}{m}\right)$ for $1 \le k \le m$, $y_0 = -\infty$, $y_{m+1} = \infty$, $x_{j0} = -\infty$, $x_{j,m+1} = \infty$ and $p_{jk} = \mathbb{P}(x_{jk} \le X \le x_{j+1,k+1}, y_j \le Y \le y_{j+1})$ if $0 \le j,k \le m$. Then we choose x_{jk} such that

$$p_{ij} = \begin{cases} \frac{1}{4m^2} & \text{if } (i,j) = (0,0) \\ \frac{1}{2m^2} & \text{if } (i,j) = (1,0), \ldots, (m-1,0) \\ \frac{1}{2m^2} & \text{if } (i,j) = (0,1), \ldots, (0,m-1) \\ \frac{1}{m^2} & \text{if } 1 \le i,j \le m-1 \end{cases}$$

Then, $\tilde{P}_{n,2}(F) = \{(x_{ij}, y_i)_{1\le i,j\le m}\}$ defines a bivariate size n profile that ensures, in a certain sense, equidistribution of the probability mass defined by F. However, we do not claim this profile to be optimal in the sense (1) or (2).

To illustrate the idea, Figure 1 shows the distribution of mass for the bivariate size 9 profile $\tilde{P}_{9,2}$. Figure 2 exhibits the bivariate size 100 profiles $\tilde{P}_{100,2}$ of the distributions $\mathcal{N}_2(0, I_d)$ (left) and $\mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.7\\ 0.7 & 1\end{pmatrix}\right)$ (right).

Isotones provide a graphical comparison of competing goodness-of-fit tests of normality under alternatives from a two-parametric family $\mathcal{F} = \{F(\cdot;\theta_1,\theta_2) :$

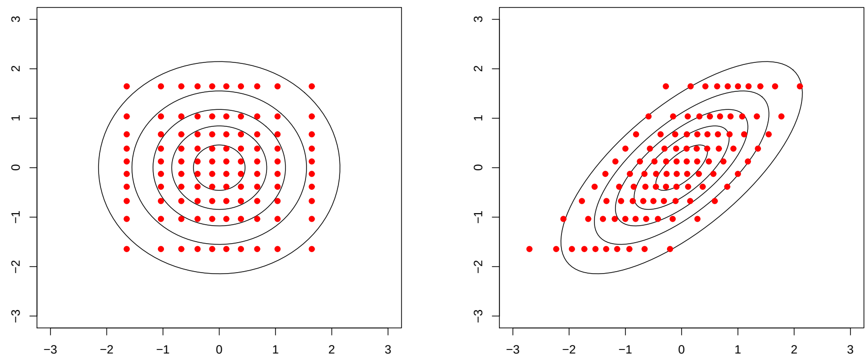

Figure 2: Size 100 profile $\tilde{P}_{100,2}$ of $\mathcal{N}_2(0, I_d)$ (*left*) and the bivariate centered normal law having unit variances and correlation $\rho = 0.7$ (*right*).

$\theta_1 \in \Theta_1, \theta_2 \in \Theta_2\}$ of distributions that includes the normal law for a special choice (θ_1^*, θ_2^*) of parameter values. Since the power of a goodness-of-fit test should increase with the distance of an alternative distribution from the null hypothesis, the p-values corresponding to samples from 'more distant' distributions should be smaller than p-values from samples originating from distributions that are nearer to the null hypothesis.

In what follows, our samples are the optimal profiles P_n^* in the univariate case and the profiles $\tilde{P}_{n,2}$ in the bivariate case. We define a mapping

$$p_{T_n} : \begin{cases} \Theta_1 \times \Theta_2 & \to [0,1] \\ (\theta_1, \theta_2) & \mapsto p_{T_n}(\theta_1, \theta_2) \end{cases}$$

that, for each parameter (θ_1, θ_2), assigns the p-value of a test statistic T_n, when computed on the data $\{x_1, \ldots, x_n\} = P_n^*$ in the univariate case and $\{(x_1, y_1), \ldots, (x_n, y_n)\} = \tilde{P}_{n,2}$ in the bivariate case. Note that these size n profiles are obtained from the distribution function $F(\cdot; \theta_1, \theta_2)$.

We now consider a 3-dimensional plot of the points $(\theta_1, \theta_2, p_{T_n}(\theta_1, \theta_2))$, $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$ in a cartesian coordinate system. For fixed $p \in (0,1)$, an *isotone* $\mathbb{I}(T_n, p)$ of a test T_n is the level curve at height p of the surface of p-values $p_{T_n}(\theta_1, \theta_2)$, $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$ in the θ_1-θ_2-plane.

Different tests are compared by means of the distance of isotones from the point (θ_1^*, θ_2^*), which represents the normal distribution. We will, however, only qualify two different tests if their isotones do not intersect.
In the latter case, we define a test T_n to be *better at level* p ($p \in (0,1)$) than a test $\tilde{T}_n$ if the distance of $\mathbb{I}(T_n, p)$ and (θ_1^*, θ_2^*) is smaller than the distance of the isotone $\mathbb{I}(\tilde{T}_n, p)$ and (θ_1^*, θ_2^*). A test T_n is *better* than a test $\tilde{T}_n$, if it is better than $\tilde{T}_n$ at level p for each $p \in (0,1)$.

The advantage of this graphical tool for comparing competing tests is that there is no need for large scale simulation studies, since both the optimal size

n profile P_n^* in the univariate case and the bivariate size n-Profil $\tilde{P}_{n,2}$ are non-stochastic. Notice however that the definition of P_n^* in the univariate case requires knowledge of the quantile function $F^{-1}(\cdot)$ of the alternative distribution. If F^{-1} is not known, numerical techniques must be used. In the bivariate case, the quantile function of the marginal distribution function of Y must be known or otherwise approximated by numerical procedures.

4 Results

4.1 Extended Skew-Normal distribution

In this subsection, we assess the power of the three tests introduced in Section 2 against the (univariate) extended skew-normal distribution $\mathcal{SNE}(\alpha,\tau)$ with density

$$f_{\alpha,\tau}(z) = \varphi(z)\frac{\Phi(\tau\sqrt{1+\alpha^2}+\alpha z)}{\Phi(\tau)}, \quad z \in \mathbb{R}$$

(see, e.g., Azzalini [1]). Putting $\tau = 0$, one obtains the (univariate) skew normal distribution $\mathcal{SN}(\alpha)$; setting $\alpha = 0$ yields the standard normal distribution $\mathcal{N}(0,1)$.

The following stochastic representation due to Henze [7] facilitates random number generation. Let T and V be independent random variables, where $V \sim \mathcal{N}(0,1)$ and T has a truncated normal distribution with density $f(t) = \varphi(t)/(1-\Phi(-\tau))$ for $t \geq -\tau$. Then,

$$Z = \frac{\alpha}{\sqrt{1+\alpha^2}}\cdot T + \frac{1}{\sqrt{1+\alpha^2}}\cdot V$$

has a $\mathcal{SNE}(\alpha,\tau)$-distribution.

In view of the affine invariance of the test statistics and the fact that $Z \sim \mathcal{SNE}(\alpha,\tau)$ implies $-Z \sim \mathcal{SNE}(-\alpha,\tau)$, it suffices to consider the case $\alpha \geq 0$.

The profiles of the $\mathcal{SNE}$ distribution and the corresponding isotones were computed for several sample sizes n and levels p. Figure 3 exhibits the typical behavior of the isotones for the case $n = 20$ and $p = 0.3$.

Since the isotone corresponding to the skewness test is closer to the line $\alpha = 0$ (i.e. the normal distribution) than the isotone of the BHEP test, the skewness test outperforms the latter. For the given alternatives, the kurtosis test ranks third.

To corroborate our findings, we conducted a Monte-Carlo study with sample sizes $n = 20$ and $n = 50$. Tables 1, 2 and 3 show the percentages of rejection of the three tests for $n = 50$ based on 10000 replications. The results for the case $n = 20$ are completely similar.

The simulation results reveal no fundamental difference between the BHEP and the skewness test, the latter being somewhat superior for negative values of τ. Again, the test based on kurtosis ranks third.

To sum up, both the simulation study and the graphical procedure based on isotones lead to similar conclusions regarding the power of the three tests against the extended skew-normal distribution.

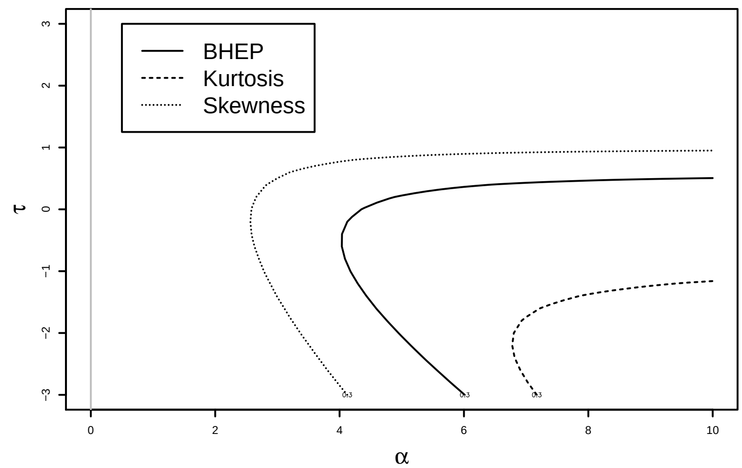

Figure 3: Isotones of normality tests against $\mathcal{SNE}$ alternatives, $n = 20$, $p = 0.3$

(α, τ)	-3	-2	-1	0	1	2	3
0	5.0	4.8	5.2	4.8	5.0	5.0	4.8
1	5.4	5.1	5.6	5.5	5.7	5.0	4.9
2	6.8	8.5	9.6	9.4	7.2	5.4	4.3
3	11.8	14.9	16.4	13.4	9.4	6.1	4.2
4	18.7	22.1	21.6	15.8	11.0	6.0	4.4
5	25.7	28.2	26.0	17.7	12.1	6.7	4.5
6	30.4	32.5	29.0	17.9	11.7	7.1	4.1
7	35.7	35.9	31.0	19.5	12.1	6.8	4.5
8	38.9	38.5	32.2	19.0	12.8	6.7	4.4
9	42.0	41.1	32.1	20.1	13.0	7.3	4.1
10	43.4	42.0	33.8	19.2	12.6	7.3	4.1

Table 1: Simulated power of the kurtosis test against the distribution $\mathcal{SNE}(\alpha, \tau)$, nominal level 5%, sample size 50

(α,τ)	-3	-2	-1	0	1	2	3
0	4.9	5.1	5.1	5.0	5.1	4.8	4.7
1	5.4	5.4	5.5	5.8	6.2	5.4	4.7
2	7.3	10.5	14.8	20.1	14.0	5.6	4.6
3	15.6	24.6	35.9	39.9	22.0	6.0	4.9
4	29.0	43.6	56.0	54.5	27.1	6.2	4.7
5	45.4	59.2	69.5	64.4	29.3	6.1	4.6
6	58.4	72.3	77.7	69.3	31.2	6.3	4.8
7	69.6	80.1	83.8	72.6	33.3	6.3	4.5
8	78.4	85.2	86.4	74.6	33.9	5.8	4.3
9	83.0	88.8	89.1	77.1	35.1	5.9	4.5
10	86.9	90.7	90.9	77.9	35.9	6.3	4.3

Table 2: Simulated power of the BHEP test against the distribution $\mathcal{SNE}(\alpha,\tau)$, nominal level 5%, sample size 50

(α,τ)	-3	-2	-1	0	1	2	3
0	5.0	5.1	5.0	5.1	5.3	4.9	4.7
1	5.0	5.4	6.2	6.4	6.4	5.1	4.3
2	9.2	13.3	18.4	22.7	14.8	5.0	4.2
3	20.2	30.3	41.1	41.0	21.7	5.6	4.2
4	36.2	49.3	59.1	52.9	25.9	5.9	4.1
5	52.2	64.5	70.4	61.5	26.9	5.6	4.0
6	64.5	74.6	76.7	65.0	29.3	6.1	3.8
7	72.9	80.6	81.9	68.0	29.8	5.7	3.8
8	80.1	85.1	84.5	69.6	31.3	5.7	4.0
9	84.1	88.3	86.0	71.1	31.7	6.1	3.6
10	87.1	89.7	87.9	72.0	32.2	6.0	3.8

Table 3: Simulated power of the skewness test against the distribution $\mathcal{SNE}(\alpha,\tau)$, nominal level 5%, sample size 50

4.2 Two-factor skew-normal distribution

As a second example, we consider the two-factor skew-normal ($\mathcal{TSN}$) distribution introduced by Gupta, Chen and Tang [6]. A d-dimensional random vector X has a $\mathcal{TSN}$ distribution with correlation matrix $\bar{\Omega} \in \mathbb{R}^{d,d}$ and parameter $\lambda \in \mathbb{R}^d, \Lambda = \mathrm{diag}(\delta_1, \ldots, \delta_d) \in \mathbb{R}^{d,d}$ with $\delta_i \geq 0$, if X has the density

$$f_{\lambda,\Lambda,\bar{\Omega}}(x) = 2\varphi_d(x;\bar{\Omega})\Phi\left(\frac{\lambda' x}{\sqrt{1 + x'\Lambda x}}\right), \quad x \in \mathbb{R}^d .$$

Here, $\varphi_d(\cdot;\bar{\Omega})$ stands for the density of $\mathcal{N}_d(0,\bar{\Omega})$, and Φ is the distribution function of the univariate standard normal distribution. This family is denoted by $\mathcal{TSN}_d(\lambda,\Lambda,\bar{\Omega})$; if $\bar{\Omega} = I_d$, we write $\mathcal{TSN}_d(\lambda,\Lambda)$.

Putting $G = \Phi$, $f_0(\cdot) = \varphi_d(\cdot,\bar{\Omega})$ and $w(z) = \frac{\lambda' z}{\sqrt{1+z'\Lambda z}}$ in Lemma 1 of [1], and noting that $w(-z) = -w(z)$, we obtain the following stochastic representation:

Proposition 4.1. *Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}_d(0,\bar{\Omega})$ be independent random variables. Then the random variable*

$$Z = \begin{cases} Y, & \text{if } X < \frac{\lambda' Y}{\sqrt{1+Y'\Lambda Y}}, \\ -Y, & \text{otherwise}, \end{cases}$$

has the density given in (4.2).

The special case $\Lambda = 0$ yields the d-dimensional skew-normal distribution introduced in [1], and denoted by $\mathcal{SN}_d(\lambda,\bar{\Omega})$; for $\lambda = 0$, the two-factor skew-normal distribution equals the $\mathcal{N}_d(0,\bar{\Omega})$-distribution.

Here, we treat the case $d = 1$ (and $\bar{\Omega} = 1$); hence, there are two parameters λ and Λ. In view of the affine invariance of the test statistics and the fact that $X \sim \mathcal{TSN}(\lambda,\Lambda)$ implies $-X \sim \mathcal{TSN}(-\lambda,\Lambda)$, we consider only nonnegative values of λ.

Figure 4 illustrates the typical behavior of the isotones for small p (with $n = 20$ and $p = 0.3$). For large values of p, one obtains a quite different picture, as Figure 5 shows for $p = 0.7$.

For small p, the isotones of the skewness test are closer to the line $\lambda = 0$ than those of the BHEP test; the isotones of the kurtosis test are even further away and do not fit within the display window. For large p, the isotones intersect repeatedly, and it is hardly possible to give any interpretation.

Figure 6 reveals the reason for this strange behavior. It shows a cross section of the surface of the the kurtosis statistic $T_{n,K}$ for $\Lambda = 4$ and the associated p-values $p_{T_{20,K}}$ which are not monotone in λ. Similar pictures are observed for other values of Λ and the other tests.

This non-monotone behavior can also be observed in simulations. As an example, Table 4 shows the empirical power of the kurtosis test for $n = 50$.

The results of a detailed simulation study (not shown here) for the sample sizes 20 and 50 give a clear picture: The BHEP test slightly outperforms the skewness test, and both tests are far better than the test based on kurtosis.

4.3 Bivariate Skew-Normal distribution

As third example, we consider the bivariate skew-normal distribution with $\bar{\Omega} = I_2$, denoted by $\mathcal{SN}_2(\alpha_1,\alpha_2)$. If the random vector (Z_1, Z_2) has a $\mathcal{SN}_2(\alpha_1,\alpha_2)$-

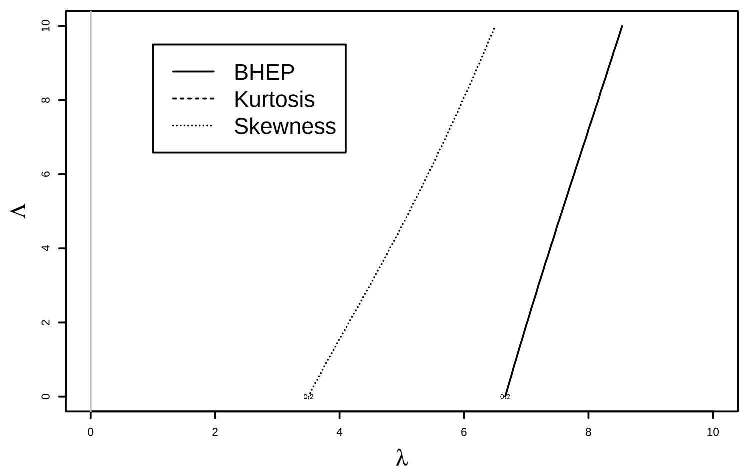

Figure 4: Isotones of normality tests against $\mathcal{TSN}$ alternatives, $n = 20$, $p = 0.2$

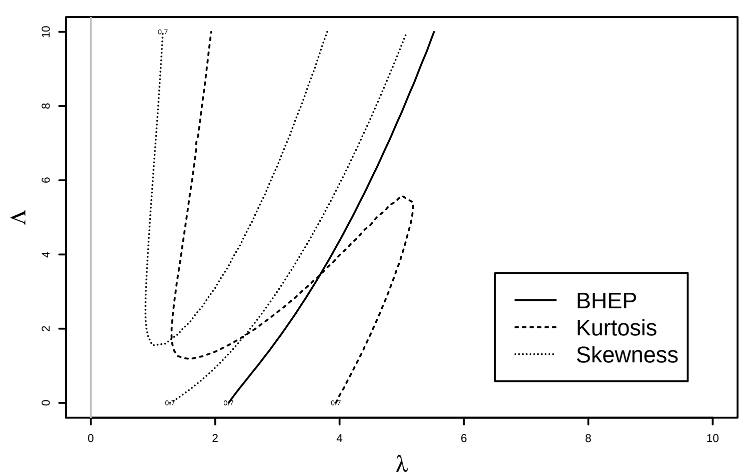

Figure 5: Isotones of normality tests against $\mathcal{TSN}$ alternatives, $n = 20$, $p = 0.7$

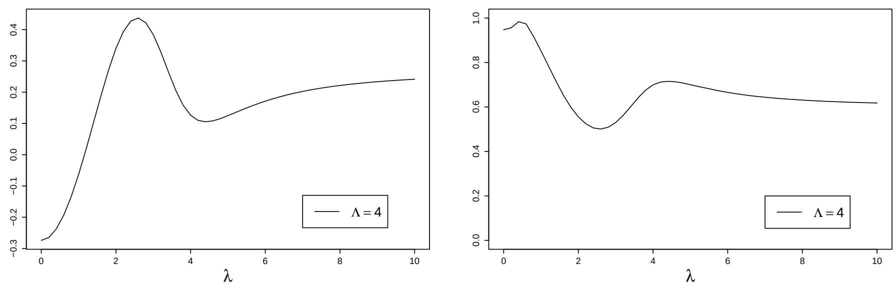

Figure 6: Cross section of the surface of the kurtosis statistic $T_{20,K}$ (*left*) and the corresponding p-values $p_{T_{20,K}}$ (*right*) of the $\mathcal{TSN}$-distribution for $\Lambda = 4$.

(λ, Λ)	0	1	2	3	4	5	6	7	8	9	10
0	5.0	4.7	4.9	4.7	4.7	5.1	5.3	5.0	5.4	4.8	5.2
1	5.6	8.0	7.4	6.4	5.6	5.4	5.7	5.6	5.6	5.7	5.5
2	9.2	14.2	16.3	16.0	15.4	13.1	11.7	10.7	10.7	9.3	8.9
3	13.9	13.8	18.0	22.5	24.0	23.5	22.8	21.6	19.9	19.5	17.9
4	16.1	15.0	16.1	19.4	22.6	24.4	27.3	27.4	27.6	27.5	27.1
5	16.8	17.4	16.5	17.5	20.0	22.1	24.8	26.3	28.6	28.9	29.1
6	18.5	17.8	18.6	18.5	18.6	19.2	20.7	23.2	25.2	27.3	27.5
7	19.2	18.9	18.3	19.0	18.6	18.3	19.7	20.4	21.5	23.0	24.7
8	19.4	19.8	19.9	18.8	19.5	19.9	19.0	19.9	20.7	21.0	21.6
9	19.8	20.3	19.3	20.2	20.0	19.5	19.8	19.1	20.7	20.3	20.9
10	20.1	19.7	19.7	19.6	20.0	19.3	20.3	20.3	19.9	19.5	20.0

Table 4: Simulated power of the kurtosis test against the distribution $\mathcal{TSN}(\lambda, \Lambda)$, nominal level 5%, sample size 50

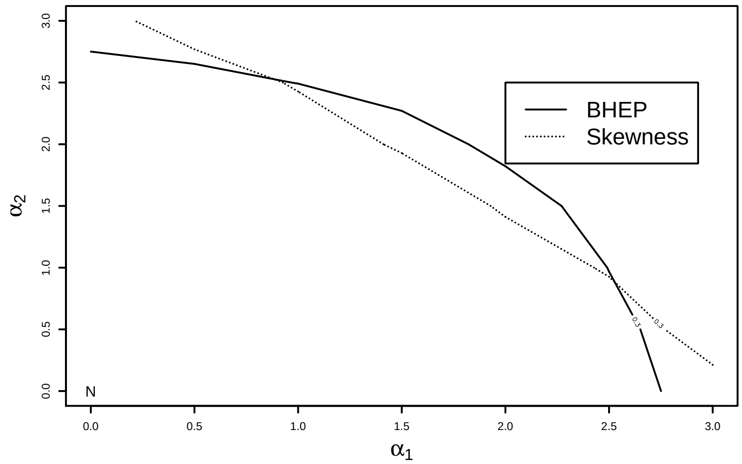

Figure 7: Isotones of normality tests against $\mathcal{SN}_2$ alternatives, $n = 100$, $p = 0.3$

distribution, the marginal distribution of Z_2 is $\mathcal{SN}(\bar{\alpha})$, where $\bar{\alpha} = \alpha_2/(\sqrt{1+\alpha_1^2})$ (Prop. 2 in Azzalini and Capitanio [2]). This result is useful for the computation of the bivariate n-Profile $\tilde{P}_{n,2}$. Since $(Z_1, -Z_2) \sim \mathcal{SN}_2(\alpha_1, -\alpha_2)$ and $(-Z_1, Z_2) \sim \mathcal{SN}_2(-\alpha_1, \alpha_2)$, we consider only nonnegative values of α_1 and α_2. The case $(\alpha_1, \alpha_2) = (0, 0)$ yields the bivariate normal distribution.

The kurtosis value $b_{2,2}^{(100)}$ for the profile of the normal distribution with $n = 100$ is 6.50 and thus not a value near 8, which is the theoretical value of $\beta_{2,2}$ under bivariate normality. This fact leads to a very small p-value under the hypothesis, and not a value near 1 as in all other cases. Apparently, the bivariate size n profile is not a typical sample from a normal distribution in view of the kurtosis statistic. Consequently, we have not computed isotones of the kurtosis test, and we also omit the kurtosis test from the simulations.

The typical behavior of the isotones of the remaining tests is illustrated in Figure 7 (with $n = 100$ and $p = 0.3$). The isotones of the skewness test are closer to $(0, 0)$ for alternatives with $\alpha_1 \approx \alpha_2$; hence, the skewness test is better for these parameter values. However, if the parameter values differ more, the BHEP test outperforms the skewness test.

Tables 5 and 6 show part of the results of a power study for the two tests with sample size $n = 100$.

Similar as the isotones indicate, the skewness test has higher power for $\alpha_1 \approx \alpha_2$, but only if α_1, α_2 are small. If one of the parameter values is larger than 4, the BHEP test outperforms the skewness test.

5 Discussion

The use of profiles and isotones is an interesting alternative to simulation studies for assessing the power of goodness-of-fit tests. In many cases, the conclusions drawn from both procedures are in good agreement, as the work of Mudholkar and co-workers [13], [14], [15] and our own results show. However, the following

(α_1, α_2)	0	1	2	3	4	5	6
0	4.9	6.2	21.1	46.0	63.4	74.2	80.9
1	6.0	9.5	26.6	49.6	65.7	73.8	79.9
2	21.2	27.0	42.4	57.2	68.6	76.0	81.2
3	46.1	50.0	57.8	67.0	73.0	78.8	82.7
4	64.2	66.0	69.4	74.1	77.5	81.2	83.7
5	74.6	75.0	76.5	78.8	81.4	82.5	85.3
6	79.6	79.6	81.3	82.5	83.6	85.4	85.8

Table 5: Simulated power of the BHEP test against the distribution $\mathcal{SN}_2(\alpha_1, \alpha_2)$, nominal level 5%, sample size 100

(α_1, α_2)	0	1	2	3	4	5	6
0	4.9	6.4	24.1	48.1	63.1	72.1	77.9
1	7.0	11.2	29.9	51.3	65.1	71.8	77.4
2	24.9	30.4	45.7	57.9	68.1	74.0	78.0
3	47.9	51.6	58.1	65.4	71.1	75.9	79.2
4	63.9	64.7	67.8	72.0	74.9	77.7	80.0
5	72.4	71.9	74.3	75.7	78.3	79.6	81.4
6	76.7	77.6	77.3	79.1	80.2	81.8	82.1

Table 6: Simulated power of the skewness test against the distribution $\mathcal{SN}_2(\alpha_1, \alpha_2)$, nominal level 5%, sample size 100

points should be kept in mind when using this procedure:

- In our experience, the qualitative behavior of isotones hardly changes with sample size, which is often different in power simulations. The reason seems to be that, even for a small sample size, the ecdf of a size n profile is a good approximation to the theoretical cdf. Hence, in the examples in Section 4, we have confined ourselves to using univariate profiles of size 20 and bivariate profiles of size 100.

- Since alternatives which are distant from the null hypotheses should be easier to detect, it is reasonable in general to compare different tests by means of the distance of isotones from the point or line which represents the normal distribution.

  However, as the $\mathcal{TSN}$ distribution shows, the test statistic and, hence, the p-values are not always monotone functions in that distance. In such cases, interpretation of isotones becomes difficult.

- The use of profiles and isotones seems especially appealing in higher dimensions given that time-consuming stochastic simulations are replaced by deterministic computations. On the other hand, defining multivariate profiles is demanding since it is not at all clear what an 'ideal sample' could be in two or more dimensions; furthermore, the construction of bivariate profiles as defined in Section 3 is time-consuming by itself since there are no closed-form expressions for most multivariate distribution functions.

References

[1] A. Azzalini (2005). The Skew-normal Distribution and Related Multivariate Families, Scand. J. Statist. **32**, 159-188.

[2] A. Azzalini, A. Capitanio (1998). Statistical Applications of the Multivariate Skew-Normal Distribution. J. Roy. Statist. Soc. **B 61**, 579-602.

[3] L. Baringhaus, N. Henze (1992). Limit Distributions for Mardia's Measure of Multivariate Skewness, Ann. Statist. **20**, 1889-1902.

[4] M.R. Eaton, M.D. Perlman (1973). The Non-Singularity of Generalized Sample Covariance Matrices. Ann. Statist. **1**, 710-717.

[5] N. Gürtler (2000). Asymptotic Results on the class of BHEP tests for Multivariate Normality with Fixed and Variable Smoothing Parameter (in German). Doctoral Dissertation, University of Karlsruhe (TH), Germany.

[6] A.K. Gupta, J.T. Chen, J. Tang (2007). A Multivariate Two-factor Skew Model. Statistics **41**, 301-307.

[7] N. Henze (1986). A Probabilistic Representation of the 'Skew-normal' Distribution. Scand. J. Statist. **13**, 271-275.

[8] N. Henze, B. Zirkler (1990). A Class of Invariant Consistent Tests for Multivariate Normality. Commun. Statist. A **19**, 3595-3617.

[9] N. Henze (1994). On Mardia's Kurtosis Test for Multivariate Normality, Commun. Statist.-Theory Meth. **23**, 1031-1045.

[10] N. Henze and Th. Wagner (1997). A new Approach to the BHEP Tests for Multivariate Normality. J. Multiv. Anal. **62**, 1-23.

[11] N. Henze (2002). Invariant Tests for Multivariate Normality: A critical review. Statist. Papers **43**, 467-506.

[12] K.V. Mardia (1970). Measures of Multivariate Skewness and Kurtosis with Applications. Biometrika **57**, 519-530.

[13] G.S. Mudholkar, G.D. Kollia, C.T. Lin, K.R. Patel (1991). A Graphical Procedure for Comparing Goodness-of-fit Tests. J. R. Statist. Soc. B **53**, 221-232.

[14] G.E. Wilding, G.S. Mudholkar (2007). Some modifications of the Z-tests of normality and their isotones. To appear in Statistical Methodology.

[15] G.E. Wilding, G.S. Mudholkar, G.D. Kollia (2007). Two Sets of Isotones for Comparing Tests of Exponentiality. J. Statist. Plann. Infer. **137**, 3815-3825.