MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Skewness and kurtosis as locally best invariant tests of normality

Akimichi TAKEMURA, Muneya MATSUI and Satoshi KURIKI

METR 2006–47 August 2006

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Skewness and kurtosis as locally best invariant tests of normality

AKIMICHI TAKEMURA
Graduate School of Information Science and Technology
University of Tokyo
MUNEYA MATSUI
Department of Mathematics, Keio University
and
SATOSHI KURIKI
The Institute of Statistical Mathematics

August, 2006

Abstract

Consider testing normality against a one-parameter family of univariate distributions containing the normal distribution as the boundary, e.g., the family of t-distributions or an infinitely divisible family with finite variance. We prove that under mild regularity conditions, the sample skewness is the locally best invariant (LBI) test of normality against a wide class of asymmetric families and the kurtosis is the LBI test against symmetric families. We also discuss non-regular cases such as testing normality against the stable family and some related results in the multivariate cases.

Keywords and phrases: generalized hyperbolic distribution, infinitely divisible distribution, normal mixture, outlier detection, stable distribution.

1 Introduction

In 1935, E.S. Person remarked:

> "...it seems likely that for large samples and when only small departures from normality are in question, the most efficient criteria will be based on the moment coefficients of the sample, e.g. on the values of $\sqrt{\beta_1}$ and β_2."

Surprisingly this statement has never been formally proved, although there exists large literature on testing normality and sampling distributions of the skewness and the kurtosis. See Thode (2002) for a comprehensive survey on tests of normality. The purpose of this paper is to give a proof of this statement for fixed sample size ($n \geq 3$) under general regularity conditions for a wide class of alternatives, including the normal mixture alternatives and the infinitely divisible alternatives with finite variance. Technically all the necessary ingredients are already given in the literature. Therefore the merit of this paper is to give a clear statement and a proof of this basic fact in a unified framework and also to consider some non-regular cases, in particular testing normality against the stable family.

In fact "non-regular" may not be an appropriate term, because by considering contamination type alternatives, we see that there are functional degrees of freedom in constructing an alternative family and the locally best invariant test against the family. Therefore by "small departure" we are excluding contamination type departures from normality. See our discussion at the end of Section 2.

In this paper we are concerned with testing the null hypothesis that the true distribution belongs to the normal location scale family, against the alternatives of other location scale families. We are mainly interested in invariant testing procedures with respect to the location and the scale changes of the observations. In the context of outlier detection, Ferguson (1961) proved that the skewness and the kurtosis are the locally best invariant tests of normality for slippage type models of outliers. In Ferguson's setting, the proportion of outliers can be substantial but the the amount of slippage tends to zero. In establishing the LBI property, Ferguson (1961) derived the basic result (see Proposition 1 below) on the likelihood ratio of the maximal invariant under the location-scale transformation. The same result was given in Section II.2.2 of Hájek & Šidák (1967). Uthoff (1970, 1973) used the result to derive the best invariant tests of normality against some specific alternatives. See also Section 3.2 of Hájek et al. (1999). A general result on the likelihood ratio of maximal invariant was given in Wijsman (1967, 1990) and it led to some important results of Kariya et al. (Kuwana & Kariya (1991), Kariya & George (1994, 1995)) in the multivariate setting.

In Ferguson (1961)'s setting of outlier detection, if the number of outliers are distributed according to the binomial distribution, the problem of outlier detection is logically equivalent to testing normality against mixture alternatives. Therefore the LBI property of the skewness and the kurtosis against mixture alternatives is a straightforward consequence of Ferguson (1961). However Ferguson's result has not been interpreted in this manner. In this paper we establish the LBI property of the skewness and the kurtosis in a more general setting and treat the normal mixture model as an example.

In testing multivariate normality, even if we restrict ourselves to invariant testing procedures, there is no single LBI test, because the maximal invariant moments are multidimensional (e.g. Takemura (1993)). Furthermore the invariance can be based on the full general linear group or the triangular group. This distinction leads to different results, because the invariance with respect to the triangular group preserves certain multivariate one-sided alternatives, whereas the general linear group does not. In Section 6 we discuss these points in a setting somewhat more general than considered by Kariya and his coauthors.

The organizations of this paper is as follows. In Section 2 we state our main theorem concerning the locally best invariant test of normality against one-sided alternatives. We also discuss Laplace approximation to the integral in LBI for large sample sizes n. In Section 3 we show that our theorem applies in particular to the normal mixture family and the infinitely divisible family. In Section 4 as an important non-regular case we consider testing against the stable family. In Section 5 we compare locally best invariant test and tests based on profile likelihood. Finally in Section 6 we discuss generalizations of our main theorem to multivariate cases.

2 Locally best invariant test of univariate normality

Let

$$f_{a,b}(x;\theta) = \frac{1}{b} f\left(\frac{x-a}{b};\theta\right), \qquad -\infty < a < \infty,\ b > 0,\ \theta \ge 0,$$

denote a one-parameter family of location-scale densities with the shape parameter θ. We simply write $f(x;\theta) = f_{0,1}(x;\theta)$ for the standard case $(a,b) = (0,1)$. We assume that $\theta = 0$ corresponds to the normal density

$$f(x;0) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big).$$

Based on i.i.d. observations $x_1,\dots,x_n$ from $f_{a,b}(x;\theta)$ we want to test the null hypothesis of normality:

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_1 : \theta > 0. \tag{1}$$

Here we are testing normality ($\theta = 0$) against the one-sided alternatives ($\theta > 0$). If we are concerned about heavier tail than the normal as the alternatives, this is a natural setting. However suppose that we are concerned about asymmetry and we do not know whether the distribution may be left-skewed or right-skewed under the alternatives. In this case we should test normality against two-sided alternatives and then (1) is not a suitable formulation. In this paper for simplicity we only consider one-sided alternatives, thus avoiding the consideration of unbiasedness of tests.

It should be noted that there exists an arbitrariness in choosing a standard member $((a,b) = (0,1))$ from a location-scale family. For the normal family we usually choose the standard normal density $\phi(x)$ as the standard member. Note however that in Section 4 we take $N(0,2)$ as the standard member in considering the stable alternatives for notational convenience. Given a particular choice of standard members $f(x;\theta)$, $\theta \ge 0$, we can choose another smooth set of standard members as

$$f_{a(\theta),b(\theta)}(x;\theta) = \frac{1}{b(\theta)} f\left(\frac{x-a(\theta)}{b(\theta)};\theta\right), \tag{2}$$

where $a(\theta), b(\theta)$ are smooth function of θ and $(a(0), b(0)) = (0, 1)$. This arbitrariness does not matter if we use invariant testing procedures. However as in the case of normal mixture distributions in Section 3.1, it is sometimes convenient to resolve this ambiguity in an appropriate manner. Details on parametrization is discussed in Appendix B.

As mentioned above we are primarily interested in invariant testing procedures. A critical region R is invariant if

$$(x_1, \ldots, x_n) \in R \quad \Leftrightarrow \quad (a + bx_1, \ldots, a + bx_n) \in R, \quad -\infty < \forall a < \infty,\ \forall b > 0.$$

Fix a particular alternative $\theta_1 > 0$. We state the following basic result (Theorem b in Section II.2.2 of Hájek & Šidák (1967), Section 2 of Ferguson (1961)) on the most powerful invariant test against θ_1.

Proposition 1. *The critical region of the most powerful invariant test for testing $H_0 : \theta = 0$ against $H_1 : \theta = \theta_1 > 0$ is given by*

$$\frac{\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n f(a + bx_i\,; \theta_1) b^{n-2} da db}{\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n f(a + bx_i\,; 0) b^{n-2} da db} > k \tag{3}$$

for some $k > 0$.

Note that the values $(x_1, \ldots, x_n)$ can be replaced by any maximal invariant of the location-scale transformation, since the ratio in (3) is invariant. For our purposes it is most convenient to replace x_i, $i = 1, \ldots, n$, by the standardized value

$$z_i = \frac{x_i - \bar{x}}{s}, \quad s^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2. \tag{4}$$

Then $\sum_{i=1}^n z_i = 0$ and $\sum_{i=1}^n z_i^2 = n$ and

$$\prod_{i=1}^n f(a + bz_i\,; 0) = \frac{1}{(2\pi)^{n/2}} \exp\Big(- \frac{n(a^2 + b^2)}{2} \Big).$$

Therefore, as in (26) of Section II.2.2 of Hájek & Šidák (1967), the denominator of (3) becomes the following constant:

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n f(a + bz_i\,; 0) b^{n-2} da db \quad = \quad \frac{\Gamma((n-1)/2)}{2n^{n/2} \pi^{(n-1)/2}}.$$

Since we are considering a fixed sample size n, this constant can be ignored in (3) and the rejection region is written as

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n f(a + bz_i\,; \theta_1) b^{n-2} da db > k'. \tag{5}$$

We now consider $\theta = \theta_1$ close to 0. For a while we proceed formally. Throughout this paper we assume that $l(x;\theta) = \log f(x;\theta)$ is continuously differentiable with respect to θ including the boundary $\theta = 0$. Then

$$l(x;\theta) = l(x;0) + l_\theta(x;0)\theta + o(\theta),$$

where

$$l_\theta(x;\theta) = \frac{\partial}{\partial\theta} \log f(x;\theta)$$

is the score function. Therefore

$$f(x;\theta) = f(x;0)\exp(l_\theta(x;0)\theta + o(\theta)) = f(x;0)(1 + l_\theta(x;0)\theta) + o(\theta)$$

and

$$\begin{aligned}\prod_{i=1}^n f(a+bz_i\,;\theta) &= \prod_{i=1}^n f(a+bz_i\,;0)(1+\sum_{i=1}^n l_\theta(a+bz_i;0)\theta) + o(\theta)\\ &= \frac{1}{(2\pi)^{n/2}} \exp\Big(-\frac{n(a^2+b^2)}{2}\Big)(1+\sum_{i=1}^n l_\theta(a+bz_i;0)\theta) + o(\theta).\end{aligned}$$

It follows that for small $\theta = \theta_1$ the rejection region (5) can be approximately written as

$$T(z_1,\ldots,z_n) = \int_0^\infty \int_{-\infty}^\infty \sum_{i=1}^n l_\theta(a+bz_i;0)\exp\Big(-\frac{n(a^2+b^2)}{2}\Big)b^{n-2}dadb > k''. \quad (6)$$

In order to justify the above derivation we assume the following convenient regularity condition.

Assumption 1. *For some* $\epsilon > 0$

$$\int_0^\infty \int_{-\infty}^\infty g(a,b;\epsilon)^n \exp\Big(-\frac{n(a^2+b^2)}{2}\Big)b^{n-2}dadb < \infty,$$

where

$$g_n(a,b;\epsilon) = \sup_{|z|\le\sqrt{n},\,0\le\theta\le\epsilon} \frac{|\frac{\partial}{\partial\theta}f(a+bz;\theta)|}{f(a+bz;0)}.$$

Under this regularity condition we have the following theorem.

Theorem 1. *Under Assumption 1 the unique rejection region of the locally best invariant test of normality* $H_0 : \theta = 0$ *vs.* $H_1 : \theta > 0$ *is given by (6), provided that* $P_0(T(z_1,\ldots,z_n) = k'') = 0$ *under* H_0.

A straightforward proof is given in Appendix A.1. Note that the statement of this theorem is slightly complicated by the requirement that $P_0(T(z_1,\ldots,z_n) = k'') = 0$ under

H_0. We need this requirement because if $P_0(T(z_1,\dots,z_n)=k'')>0$, in order to maximize the local power we have to look at $O(\theta^2)$ terms in the expansion of $f(x;\theta)$ around $\theta=0$.

A particularly simple result is obtained when $l_\theta(x;0)$ is a polynomial of degree k in x. In this case $l_\theta(a+bz_i;0)$ is a polynomial in a,b and z_i and $l_\theta(a+bz_i;0)$ is written as

$$l_\theta(a+bz_i;0)=p_0(a,b)z_i^k+p_1(a,b)z_i^{k-1}+\cdots+p_k(a,b), \tag{7}$$

where $p_0(a,b),\dots,p_k(a,b)$ are polynomials in a and b. Denote the standardized l-th central moment by

$$\tilde m_l=\frac{m_l}{s^l}=\frac{1}{n}\sum_{i=1}^n z_i^l.$$

Then average of (7) is written as

$$\frac{1}{n}\sum_{i=1}^n l_\theta(a+bz_i;0)=p_0(a,b)\tilde m_k+\cdots+p_{k-3}(a,b)\tilde m_3+p_{k-2}(a,b)+p_k(a,b).$$

Furthermore the integral $\int_0^\infty\int_{-\infty}^\infty p_j(a,b)\exp(-n(a^2+b^2)/2)b^{n-2}dadb$ can be explicitly evaluated. See Appendix C. In particular if $l_\theta(x;0)$ is a third degree polynomial, then (6) is equivalent the standardized sample skewness of the observations. Now consider the case that $l_\theta(x;0)$ is a fourth degree polynomial without odd degree terms. Then

$$\int_{-\infty}^\infty a^{2l+1}\exp\Big(-\frac{na^2}{2}\Big)da=0$$

implies that $\int_{-\infty}^\infty p_{k-3}(a,b)da=0$ in (7). Therefore (6) is equivalent the standardized sample kurtosis. We now have the following corollary.

Corollary 1. *Assume the same regularity condition as in Theorem 1. If the score function $l_\theta(x;0)$ is a third degree polynomial in x, then the locally best invariant test of normality is given by the standardized sample skewness. If $l_\theta(x;0)$ is a fourth degree polynomial in x without odd degree terms, then the locally best invariant test of normality is given by the standardized sample kurtosis.*

In the next section we show that in two important cases, $l_\theta(x;0)$ is a third degree polynomial for asymmetric alternatives and is a fourth degree polynomial in x without odd degree terms for symmetric alternatives.

For general score function the integral (6) may not be easy to evaluate. Although in this paper we are considering fixed n, we here discuss Laplace approximation to the integral (6) for large n. Let A denote a random variable having the distribution $N(0,1/n)$ and let B denote the random variable such that $B/\sqrt{n}$ has the χ-distribution with $n-1$ degrees of freedom. Then as $n\to\infty$, (A,B) converges to $(0,1)$ in distribution (or equivalently in probability). Note that except for the normalizing constant, the integral in (6) can be written as $E[\sum_{i=1}^n l_\theta(A + Bz_i; 0)]$. Under mild regularity conditions, for large n, this expectation is simply approximated by putting $(A, B) = (0, 1)$:

$$\tilde{T}(z_1, \ldots, z_n) = \sum_{i=1}^{n} l_\theta(z_i; 0) \tag{8}$$

It is easily shown that this is in fact the Laplace approximation (e.g. Bleistein & Handelsman (1986)) to the integral in (6). We call $\tilde{T}$ approximate LBI for testing normality. Under mild regularity conditions, the approximate LBI and the LBI should be asymptotically equivalent.

In Appendix A.1 of Kuriki & Takemura (2001) it is shown that the test based on the k-th standardized sample cumulant is asymptotically equivalent to the test based on $\sum_{i=1}^n H_k(z_i)$, where H_k is the k-th Hermite polynomial. We see that the k-th standardized sample cumulant is characterized as an approximate LBI for the case that the score function is given by $H_k(x)$. See a further discussion in Section 5. When n is not too large, we may consider evaluating $E[\sum_{i=1}^n l_\theta(A + Bz_i; 0)]$ by numerical integration or by Monte Carlo sampling.

For the rest of this section we make several remarks on the above results. In the location-scale transformation $x_i \mapsto a + bx_i$ we might allow $b \neq 0$ to be negative. The maximal invariant is $\boldsymbol{z} = (z_1, \ldots, z_n)'$ with $\boldsymbol{z}$ identified with $-\boldsymbol{z}$, or more compactly it is $\boldsymbol{z}\boldsymbol{z}'$. Then an invariant critical region can not depend on a sign preserving function ψ of $\boldsymbol{z}$ (i.e. $\psi(-\boldsymbol{z}) = -\psi(\boldsymbol{z})$). In particular it can not depend on the skewness m_3 itself, although it can depend on $|m_3|$. In the univariate case, allowing $b < 0$ is somewhat unnatural and we have so far only considered $b > 0$. However in the multivariate case the invariance with respect to the full general linear group corresponds to allowing $b < 0$ in the univariate case. We discuss this point further in Section 6.

Let $g(x)$ be a probability density. By an ϵ-contamination alternative we mean a density of the form

$$f(x; \epsilon) = (1 - \epsilon)\phi(x) + \epsilon g(x) = \phi(x) + \epsilon(g(x) - \phi(x)).$$

Letting $\theta = \epsilon$, we see

$$l_\theta(x; 0) = \frac{g(x)}{\phi(x)} - 1.$$

Therefore as long as $g(x) = \phi(x)(1 + l_\theta(x; 0))$ is a probability density, we can construct a one-parameter contamination family of alternatives such that $T(z_1, \ldots, z_n)$ in (6) is the LBI with this score function $l_\theta(x; 0)$. By "small departures from normality" Pearson (1935) probably did not have a contamination alternative in mind. In our setting the sample size n is fixed. If ϵ is much smaller than $1/n$, we actually have no observation from $g(x)$ with probability close to 1. In this sense a contamination family seems to possess certain non-regularity as a family containing the normal distribution.

3 Normal mixture family and infinitely divisible family of distributions

In this section we discuss two general classes of alternatives such that the score function at the normal distribution is a polynomial and Corollary 1 is applicable. The first is the normal mixture family and the second is the infinitely divisible family with finite variance.

3.1 Normal mixture family

Suppose that the mean μ and the variance σ^2 of the normal distribution $N(\mu, \sigma^2)$ has the prior distribution $g(\mu, \sigma^2; \theta)$, $\theta \geq 0$, such that g degenerates to the point mass at $(0, 1)$ as $\theta \to 0$. For simplicity write $\tau = 1/\sigma^2 - 1$. Then as $\theta \to 0$, both μ and τ converge to 0 in distribution. The marginal density is given by

$$f(x; \theta) = \int_{-1}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-(\tau + 1)\frac{(x - \mu)^2}{2}) h(\mu, \tau; \theta) d\mu d\tau,$$

where $h(\mu, \tau; \theta) = (\tau + 1)^2 g(\mu, 1/(1 + \tau); \theta)$. Consider the expansion

$$\begin{aligned} \exp(-(\tau + 1)\frac{(x - \mu)^2}{2}) &= \exp(-\frac{x^2}{2}) \exp(-(\tau + 1)\frac{\mu^2}{2}) \exp((\tau + 1)x\mu - \tau\frac{x^2}{2}) \\ &= \exp(-\frac{x^2}{2}) \exp(-(\tau + 1)\frac{\mu^2}{2}) \Big(1 + ((\tau + 1)x\mu - \tau\frac{x^2}{2}) \\ &\quad + \frac{1}{2}((\tau + 1)x\mu - \tau\frac{x^2}{2})^2 + \cdots\Big). \end{aligned} \tag{9}$$

The term $\exp(-(\tau + 1)\frac{\mu^2}{2})$ can be absorbed into $h(\mu, \tau; \theta)$ and can be ignored. Also the constant term (i.e. terms not involving x) in the expansion can be ignored. Now from (36) of Appendix B it follows that without loss of generality we can choose the prior distribution in such a way that the expected values of the coefficients of x and x^2 vanish. Therefore in (9) we only need to consider the cubic or higher degree terms in x in the expansion. Relevant terms on the right-hand side of (9) are

$$\exp(-\frac{x^2}{2})\big[-\frac{1}{2}\mu\tau x^3 + \frac{1}{8}\tau^2 x^4 + \frac{1}{6}\mu^3 x^3 - \frac{1}{4}\mu^2\tau x^4 + \frac{1}{24}\mu^4 x^4\big]. \tag{10}$$

If only the scale parameter is mixed, i.e. if $\mu \equiv 0$, then the dominant term is $(1/8)\tau^2 x^4$. The primary example of this case is the family of t-distributions with $m = 1/\theta$ degrees of freedom, where the mixing distribution for the scale is the inverse Gamma distribution. From the above consideration it follows that the LBI test against the t-family is given by the standardized sample kurtosis. On the other hand if only the location parameter is mixed, i.e. $\tau \equiv 0$ and $E_g(\mu^3) \neq 0$, then the LBI test is given by the standardized sample skewness.

More interesting case is that μ and τ is of the same order and the LBI test involves both skewness and kurtosis simultaneously. This happens in a limiting case of "normal variance-mean mixture." In the normal variance-mean mixture, X given $Y = y$ is normal with mean $a + by$, $b \neq 0$, and variance y:

$$X \mid Y = y \;\sim\; N(a + by, y), \qquad Y \sim g(y, \theta).$$

Now assume that Y degenerates to a constant as $\theta \to 0$. Since we are considering location-scale invariant tests, we can assume that $Y \to 1$ in distribution and $a = -b$. Writing $\mu = b(y - 1)$ we have

$$\tau = \frac{1}{y} - 1 = -\frac{\mu}{\mu + b} = -\frac{\mu}{b} + o(|\mu|), \qquad \text{or} \quad \mu = -b\tau + o(|\tau|). \tag{11}$$

Therefore μ and τ become proportional as $\theta \to 0$. In the following subsection we look at the generalized hyperbolic distribution as an example of this case.

3.1.1 The case of the generalized hyperbolic distribution

Generalized hyperbolic distribution (GH distribution) was introduced by Barndorff-Nielsen (1977). Detailed explanations including applications of GH distributions are given in Barndorff-Nielsen & Shephard (2001), Eberlein (2001) or Masuda (2002). From Eberlein (2001) the density is written as

$$\begin{aligned} &f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) \\ &\quad = a(\lambda, \alpha, \beta, \delta) \left(\delta^2 + (x-\mu)^2\right)^{(\lambda - \frac{1}{2})/2} K_{\lambda - \frac{1}{2}} \left(\alpha \sqrt{\delta^2 + (x-\mu)^2}\right) \exp\left(\beta(x - \mu)\right), \end{aligned} \tag{12}$$

where

$$a(\lambda, \alpha, \beta, \delta) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda - \frac{1}{2}} \delta^\lambda K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})}$$

is the normalizing constant and K_λ is the modified Bessel function of the third kind with index λ:

$$K_\lambda(z) = \frac{1}{2} \int_0^\infty y^{\lambda - 1} \exp\left(-\frac{1}{2} z \left(y + y^{-1}\right)\right) dy, \quad z > 0.$$

The parameter space is given by

$$-\infty < \mu, \lambda < \infty, \quad \alpha > |\beta|,$$

with the additional boundaries $\{\delta = 0, \lambda > 0\}$ and $\{\alpha = |\beta|, \lambda < 0\}$.

GH distribution can be characterized as a normal variance-mean mixture using the generalized inverse Gaussian distributions (GIG distributions) as the mixing distribution. Let $X \mid Y = y$ be distributed as $N(\mu + \beta y, y)$ and let Y have the generalized inverse Gaussian distribution with parameters λ, δ, and $\gamma = \sqrt{\alpha^2 - \beta^2}$. The density of Y is written as

$$f_{GIG}(y; \lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\delta\gamma)} y^{\lambda - 1} \exp\left(-\frac{1}{2}\left(\frac{\delta^2}{y} + \gamma^2 y\right)\right), \quad y > 0, \tag{13}$$

where the parameter space is given by $\gamma, \delta > 0$, $-\infty < \lambda < \infty$, with the additional boundaries $\{\delta = 0, \lambda > 0\}$ and $\{\gamma = 0, \lambda < 0\}$.

In (13) let $\delta \to \infty$ and $\gamma \to \infty$ such that $\gamma/\delta \to \bar{c}$, then it is easily seen that Y degenerates to $\bar{c}$. Therefore GH distribution converges to $N(\mu + \beta\bar{c}, \bar{c})$ as $\delta \to \infty$ and $\gamma \to \infty$ such that $\gamma/\delta \to \bar{c}$. As above we can assume $\bar{c} = 1$ and $\mu = -\beta$ without loss of generality. We also assume that β is fixed. For simplicity let $\delta = \gamma$. Then (13) is written as

$$f_{GIG}(y; \lambda, \gamma) = \frac{1}{2K_\lambda(\gamma^2)} y^{\lambda-1} \exp\left(-\frac{\gamma^2}{2}\left(\frac{1}{y} + y\right)\right).$$

Note that this density has exponentially small tails at $y = 0$ and $y = \infty$. Therefore term by term integration in (9) is justified.

By (11), the main term in (10) is simply given as

$$\exp(-\frac{x^2}{2})(\frac{\beta}{2}x^3 + \frac{1}{8}x^4)\tau^2.$$

It follows that the rejection region of the LBI test (for a fixed β) is given by

$$c_{n+2}\sum_{i=1}^{n} z_i^4 + 4\beta c_{n+1}\sum_{i=1}^{n} z_i^3 > k,$$

where

$$c_l = \int_0^\infty x^l e^{-nx^2/2} dx = \frac{2^{(l-1)/2}}{n^{(l+1)/2}}\Gamma(\frac{l+1}{2}). \tag{14}$$

We see that the LBI test involves both the skewness and the kurtosis simultaneously and the weight depends on the value of β.

3.2 Infinitely divisible family

Here we consider an infinitely divisible family with finite variance. The characteristic function of an infinitely divisible random variable X with mean 0 and variance 1 can be written as

$$\phi(t) = \exp[\int_{-\infty}^{\infty} (e^{itu} - 1 - itu)\frac{1}{u^2}\mu(du)], \tag{15}$$

where the Lévy measure μ can be taken as a probability measure. Here we assume that X possesses moments up to an appropriate order. Since moments of the Lévy measure μ are the cumulants of X, existence of moments of X up to an appropriate order is equivalent to the existence of moments of μ to the same order. For example if Y has the exponential distribution, the characteristic function of $X = Y - 1$ can be written as (15) with $\mu(du) = ue^{-u}, u > 0$, (Example 8.10 of Sato (1999)) and for the double-exponential distribution with variance 1, $\mu(du) = |u|e^{-\sqrt{2}u}, -\infty < u < \infty$.

Now we introduce the time parameter $m = 1/\theta$ and consider a Lévy process $X(m)$, where $X = X(1)$ has the characteristic function (15). Furthermore we standardize the variance as $X(m)/\sqrt{m}$. Then by the central limit theorem $X(m)/\sqrt{m}$ converges to $N(0,1)$ as $m \to \infty$. The characteristic function of $X(m)/\sqrt{m}$ is written as

$$\phi_m(t) = \phi(t/\sqrt{m})^m = \exp[\int_{-\infty}^{\infty} m(e^{iut/\sqrt{m}} - 1 - \frac{iut}{\sqrt{m}})\frac{1}{u^2}\mu(du)]. \tag{16}$$

Recalling the fact $|e^{ix} - (1 + ix + (ix)^2/2 + \cdots + (ix)^k)/k| \le |x|^{k+1}/(k+1)!$ for all real x, we can expand the integrand in (16) as

$$m(e^{iut/\sqrt{m}} - 1 - \frac{iut}{\sqrt{m}}) = -\frac{t^2}{2} + \frac{(it)^3}{6\sqrt{m}}u + \frac{(it)^4}{24m}u^2 + o(1/m)$$

up to an appropriate order and integrate it term by term. Then

$$\phi_m(t) = \exp\Big(-\frac{t^2}{2} + \frac{\kappa_3}{6\sqrt{m}}(it)^3 + \frac{\kappa_4}{24m}(it)^4\Big)(1 + o(1/m)), \tag{17}$$

where $\kappa_j = \int_{-\infty}^{\infty} u^j \mu(du)$ is the j-th cumulant of X. Note that (17) is formally the same as the usual Edgeworth expansion of the cumulant generating function of m i.i.d. random variables. By considering a Lévy process, we can allow m to be fractional and we have a family of distributions $\{X(m)/\sqrt{m}\}$ indexed by the continuous parameter $m = 1/\theta$. By the usual Edgeworth expansion, the density function of $X(m)/\sqrt{m}$ is given as

$$f(x; 1/m) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}(1 + \frac{\kappa_3}{6\sqrt{m}}H_3(x) + \frac{\kappa_4}{24m}H_4(x) + \frac{\kappa_3^2}{72m}H_6(x)) + o(1/m),$$

where $H_j(x)$ is the j-th Hermite polynomial. We now see that i) if $\kappa_3 \neq 0$ then the LBI test is given by the sample skewness and ii) if $\kappa_3 = 0$ and $\kappa_4 \neq 0$ then the LBI test is given by the standardized sample kurtosis.

As examples consider the centered exponential distribution and the double-exponential distribution discussed at the beginning of this section. In the former case we test normality against the family of normalized Gamma distributions and the LBI test is given by the standardized sample skewness. In the latter case, the characteristic function of $X(m)/\sqrt{m}$ is given by

$$\phi_m(t) = \Big(1 - \frac{t^2}{2m}\Big)^{-m}$$

This is a dual family of distributions to t-family in the sense of Dreier & Kotz (2002). The LBI test against this family is given by the sample kurtosis, as in the case of t-family.

4 Testing against the stable family

In this section as an important non-regular case we consider testing against the stable family. The characteristic function of a general stable distribution ($\alpha \neq 1$) is given by

$$\Phi(t) = \Phi(t; \mu, \sigma, \alpha, \beta) = \exp\left(-|\sigma t|^\alpha \left\{1 + i\beta(\mathrm{sgn}t)\tan\left(\frac{\pi\alpha}{2}\right)(|\sigma t|^{1-\alpha} - 1)\right\} + i\mu t\right),$$

where μ is the location, σ is the scale, β is the "skewness" and α is the characteristic exponent. The parameter space is given by

$$-\infty < \mu < \infty,\ \sigma > 0,\ 0 < \alpha \le 2,\ |\beta| \le 1.$$

For the standard case $(\mu, \sigma) = (0, 1)$ we simply write the characteristic function as

$$\Phi(t; \alpha, \beta) = \exp\left(-|t|^{\alpha}\left\{1 + i\beta(\mathrm{sgn}t)\tan\left(\frac{\pi\alpha}{2}\right)(|t|^{1-\alpha} - 1)\right\}\right). \tag{18}$$

This is Zolotarev's (M) parameterization (see p.11 of Zolotarev (1986)). The corresponding density is written as $g(x; \mu, \sigma, \alpha, \beta)$ and $g(x; \alpha, \beta)$ in the standard case.

Letting $\alpha = 2$ in (18) we obtain $N(0, 2)$. For convenience let $\theta = 2 - \alpha$, $\mu = a$, $\sigma = b$ and we write

$$f(x; \theta) = g(x; a, b, 2 - \theta, \beta),$$

where $f(x; 0)$ corresponds to $N(0, 2)$. For this section we take $N(0, 2)$ as the standard member of the normal location-scale family. In the following we fix β and for each β we consider LBI for $H_0 : \theta = 0$ vs $H_1 : \theta > 0$. This is similar to the case of generalized hyperbolic distributions. In particular for $\beta = 0$ we are testing normality against the symmetric stable family, which is important in practice.

It can be shown that we can differentiate $g(x; \alpha, \beta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\Phi(t; \alpha, \beta)dt$ under the integral sign and the score function is written as

$$l_\theta(x; 0) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\partial}{\partial\alpha}\Phi(t; 2, \beta)dt. \tag{19}$$

In particular for $\beta = 0$

$$l_\theta(x; 0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos(tx)\log|t|\ t^2 e^{-t^2}dt.$$

The non-regularity of stable family lies in the fact that this score function has a very heavy tail. In fact in Matsui (2005) it is shown that for large $|x|$

$$l_\theta(x; 0) = O\Big(\exp\big(\frac{x^2}{4}\big)|x|^{-3}\Big).$$

Thus under $N(0, 2)$, $E[l_\theta(x; 0)] = 0$ exists but $E[l_\theta(x; 0)^2] = \infty$ diverges. This corresponds to the fact that as $\alpha \uparrow 2$, the Fisher information $I_{\alpha\alpha}$ diverges to infinity. Matsui (2005) gives a detailed analysis of the Fisher information matrix for the general stable distribution close to the normal distribution.

Although Assumption 1 does not hold for this case and we have to give a separate proof, the following theorem holds.

Theorem 2. *In the general stable family consider testing $H_0 : \alpha = 2$ vs. $H_1 : \alpha < 2$ for fixed β. Then the locally best invariant is given by (6), where the score function is given in (19).*

The proof of this theorem is very technical and it is given in Appendix A.2. Note that score function puts extremely heavy weights to outlying observations and this test can be considered as an outlier detection test. This is intuitively reasonable, because the stable distribution with $\alpha < 2$ does not possess a finite variance.

5 Tests based on the profile likelihood

In this section we consider tests based on the profile likelihood, where the location and the scale parameters are estimated by the maximum likelihood. We show that the LBI test and the test based on the profile likelihood are different in general except for the case that the score function is a third degree polynomial or a fourth degree polynomial without odd degree terms. Our argument in this section is formal and we implicitly assume enough regularity conditions so that our formal argument is justified.

Consider a density close to a normal distribution of the form

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)\big\{1 + \theta h\big(\frac{x-\mu}{\sigma}\big) + o(\theta)\big\}, \tag{20}$$

where h is some smooth function.

We estimate μ and σ by the maximum likelihood under the null and under the alternative and take the ratio of the maximized likelihoods. θ is considered to be fixed in the estimation. Since the maximum likelihood estimator is location-scale equivariant, we obtain an invariant testing procedure. Under the null hypothesis of normal distribution the maximum likelihood estimates are $\hat\mu = \bar x$ and $\hat\sigma^2 = s^2$. Under the alternative, an approximation to $\hat\mu$ and $\hat\sigma^2$ to the order of $O(\theta)$ is easily derived as

$$\hat\mu = \bar x - \theta s \frac{1}{n}\sum_{i=1}^n h'(z_i) + o(\theta), \qquad \hat\sigma^2 = s^2(1 - \theta \frac{1}{n}\sum_{i=1}^n z_i h'(z_i)) + o(\theta). \tag{21}$$

Let $L(\hat\mu, \hat\sigma^2)$ denote the log-likelihood under the alternative and let $L(\bar x, s^2)$ denote the log-likelihood under the null. Then substituting (21) into (20) we obtain

$$\begin{aligned} L(\hat\mu,\hat\sigma^2) &= -\frac{1}{2}\sum_{i=1}^n (x_i - \bar x)^2 \frac{1}{s^2}(1 + \theta\frac{1}{n}\sum_{i=1}^n z_i) - n\log s + \frac{\theta}{2}\frac{1}{n}\sum_{i=1}^n z_i h'(z_i) + o(\theta) \\ &= L(\bar x, s^2) - \frac{\theta}{2}\frac{n-1}{n}\sum_{i=1}^n z_i h'(z_i) + o(\theta). \end{aligned}$$

Hence the test based on the profile likelihood ratio has the rejection region

$$\sum_{i=1}^n z_i h'(z_i) > k. \tag{22}$$

On the other hand, as discussed in Section 2, for large n the Laplace approximation to the integral in (6) implies that the LBI is asymptotically equivalent to

$$\sum_{i=1}^n h(z_i) > k'. \tag{23}$$

We see that (22) and (23) are generally different even asymptotically. It should be noted that if h is a third degree polynomial or a fourth degree polynomial without odd degree terms, then both the profile likelihood procedure and the LBI procedure reduce to the sample skewness and the sample kurtosis.

6 Multivariate extensions

In this section we consider multivariate extensions of our results. A comprehensive survey on invariant tests of multivariate normality is given in Henze (2002).

For a column vector $a \in \mathbb{R}^p$ and a $p \times p$ nonsingular matrix B, let

$$f_{a,B}(x;\theta) = \frac{1}{|\det B|} f(B^{-1}(x-a);\theta), \quad x \in \mathbb{R}^p, \tag{24}$$

be a one-parameter family with the shape parameter θ. As in the univariate case, we assume that

$$f(x;0) = \frac{1}{(2\pi)^{p/2}} \exp(-\|x\|^2/2),$$

where $\|\cdot\|$ denotes the standard Euclidean norm in $\mathbb{R}^p$. Based on the i.i.d. samples $x_1, \dots, x_n$ from $f_{a,B}(x;\theta)$, we discuss invariant testing procedures for testing the normality

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_1 : \theta > 0.$$

Write $X = (x_1, \dots, x_n)' \in \mathbb{R}^{n\times p}$. Consider the group

$$\mathbb{R}^p \times GL(p) = \{(a,B) \mid a \in \mathbb{R}^p,\ B \in \mathbb{R}^{p\times p},\ \det B \neq 0\}$$

endowed with the product $(a_1, B_1)\cdot(a_2, B_2) = (B_2 a_1 + a_2, B_2 B_1)$. This group acts on the sample space $\mathbb{R}^{n\times p}$ as

$$(a,B)X = 1_n a' + XB', \quad (a,B) \in \mathbb{R}^p \times GL(p), \tag{25}$$

where $1_n = (1,\dots,1)' \in \mathbb{R}^n$. For each θ fixed, the action (25) induces the transitive action on the parameter space. In other words, the model (24) is a transformation model with the parameter (a,B). Thus, it is natural to consider invariant procedures under the action (25).

Let $LT(p)$ be the set of $p\times p$ lower triangular matrices with positive diagonal elements. Let $\bar{x} = \sum_{i=1}^n x_i/n$ and $S = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'/n$ be the sample mean vector and the sample covariance matrix. Let $T \in LT(p)$ be the Cholesky root of S so that $S = TT'$. Let

$$Z = (z_1, \dots, z_n)' = (X - 1_n \bar{x}')(T')^{-1} \tag{26}$$

($z_i = T^{-1}(x_i - \bar{x})$, $i = 1, \dots, n$). It is easy to see that a maximal invariant under the action (25) is

$$W = ZZ' = (X - 1_n\bar{x}')S^{-1}(X - 1_n\bar{x}')',$$

and we can choose a cross section $\tilde{Z} = \tilde{Z}(X) = (\tilde{z}_1, \dots, \tilde{z}_n)' \in \mathbb{R}^{n\times p}$ as a unique decomposition of $W = \tilde{Z}\tilde{Z}'$ in some appropriate way. Note that $\tilde{Z} = ZQ'$, or $\tilde{z}_i = Qz_i$, for some $p \times p$ orthogonal matrix Q. The following is a multivariate extension of Proposition 1.

Proposition 2. *Under the group action of* $\mathbb{R}^p \times GL(p)$*, the critical region of the most powerful invariant test for testing* $H_0 : \theta = 0$ *against* $H_1 : \theta = \theta_1 > 0$ *is given by*

$$\frac{\int_{GL(p)} \int_{\mathbb{R}^p} \prod_{i=1}^n f(a + Bz_i; \theta_1) |\det B|^{n-p-1} da dB}{\int_{GL(p)} \int_{\mathbb{R}^p} \prod_{i=1}^n f(a + Bz_i; 0) |\det B|^{n-p-1} da dB} > k \tag{27}$$

for some $k > 0$*, where* $da = \prod_{i=1}^p da_i$ *and* $dB = \prod_{i,j=1}^p db_{ij}$ *are the Lebesgue measures of* $\mathbb{R}^p$ *and* $\mathbb{R}^{p\times p}$*, respectively.*

Proof. The Jacobian of the transformation $X \mapsto (a, B)X = 1_n a' + XB'$ is $(\det B)^n$. The left invariant measure of $\mathbb{R}^p \times GL(p)$ is $(\det B)^{-(p+1)} da dB$. From Theorem 4 of Wijsman (1967), the critical region is

$$\frac{\int_{GL(p)} \int_{\mathbb{R}^p} \prod_{i=1}^n f(a + B\tilde{z}_i; \theta_1) |\det B|^{n-p-1} da dB}{\int_{GL(p)} \int_{\mathbb{R}^p} \prod_{i=1}^n f(a + B\tilde{z}_i; 0) |\det B|^{n-p-1} da dB} > k,$$

which is equivalent to (27). □

Next consider the subgroup

$$\mathbb{R}^p \times LT(p) = \{(a, T) \mid a \in \mathbb{R}^p,\ T \in LT(p)\}$$

of $\mathbb{R}^p \times GL(p)$. This also acts on the sample space $\mathbb{R}^{n\times p}$ with the same action (25) with $GL(p)$ replaced by $LT(p)$. For this group, the induced action on the parameter (a, B) in the model (24) is not transitive anymore. However, when we consider a subclass of (24) that

$$\begin{aligned} f_{a,B}(x; \theta) &= \frac{1}{|\det B|} h(\|B^{-1}(x - a)\|^2; \theta) \\ &= \frac{1}{\sqrt{\det(BB')}} h((x - a)'(BB')^{-1}(x - a); \theta) \end{aligned} \tag{28}$$

(h is a function), the action on the parameter (a, BB') is transitive, and invariant testing procedures under the group $\mathbb{R}^p \times LT(p)$ may be more appropriate in some cases.

For the action of $\mathbb{R}^p \times LT(p)$, Z in (26) is a maximal invariant, and we can use Z itself as a cross section.

The most powerful invariant test under the action of $\mathbb{R}^p \times LT(p)$ is given as follows.

Proposition 3. *Under the group action of* $\mathbb{R}^p \times LT(p)$*, the critical region of the most powerful invariant test for testing* $H_0 : \theta = 0$ *against* $H_1 : \theta = \theta_1 > 0$ *is given by*

$$\frac{\int_{LT(p)} \int_{\mathbb{R}^p} \prod_{i=1}^n f(a + Tz_i; \theta_1) da \prod_{i=1}^p t_{ii}^{n-i-1} \prod_{i\ge j} dt_{ij}}{\int_{LT(p)} \int_{\mathbb{R}^p} \prod_{i=1}^n f(a + Tz_i; 0) da \prod_{i=1}^p t_{ii}^{n-i-1} \prod_{i\ge j} dt_{ij}} > k'$$

for some $k' > 0$*, where* $T = (t_{ij}) \in LT(p)$*.*

Proof. The Jacobian of the transformation $X \mapsto (a, T)X = 1_n a' + XT'$ is $(\det T)^n = \prod_{i=1}^n t_{ii}^n$. The left invariant measure of $\mathbb{R}^p \times LT(p)$ is $da \prod_{i=1}^p t_{ii}^{-(i+1)} \prod_{i \geq j} dt_{ij}$. The proposition follows from Theorem 4 of Wijsman (1967). □

From Propositions 2 and 3, under similar conditions to Assumption 1, the LBI test can be derived by integrating the score function $\sum_{i=1}^n \ell_\theta(a+Bz_i; 0)$ with respect to (a, B). In the rest of this section, we examine a particular case where

$$\prod_{i=1}^n f(a + Bz_i; \theta) = \prod_{i=1}^n f(a + Bz_i; 0)\{1 + \theta \sum_{i=1}^n \ell_\theta(a + Bz_i; 0) + o(\theta)\}$$

with

$$\ell_\theta(x; 0) = p_0\|x\|^4 + p_1\|x\|^2 + p_2.$$

This holds, for example, when $f_{a,B}(x; \theta)$ is of the form of (28) with

$$h(y; \theta) = \begin{cases} \dfrac{\Gamma((p + \theta^{-1})/2)}{(\theta^{-1}\pi)^{p/2}\Gamma(\theta^{-1}/2)}(1 + \theta y)^{-(p+\theta^{-1})/2} & (\theta > 0) \\ \dfrac{1}{(2\pi)^{p/2}} \exp(-y/2) & (\theta = 0) \end{cases}$$

(multivariate t distribution with θ^{-1} degrees of freedom). We restrict our attention to the case $p_0 > 0$ for simplicity.

Assumption 2.
(i)

$$\frac{\frac{\partial}{\partial\theta} f(a + Bz; \theta)|_{\theta=0}}{f(a + Bz; 0)} = p_0\|z\|^4 + p_1\|z\|^2 + p_2 \quad (p_0 > 0).$$

(ii) For some $\epsilon > 0$,

$$\int_{GL(p)} \int_{\mathbb{R}^p} g(a, B; \epsilon)^n \exp(-\frac{n}{2}\|a\|^2 - \frac{n}{2}\mathrm{tr}(B'B))|\det B|^{n-p-1} da dB < \infty,$$

where

$$g(a, B; \epsilon) = \sup_{\|z\| \leq 1, 0 \leq \theta \leq \epsilon} \frac{|\frac{\partial}{\partial\theta} f(a + Bz; \theta)|}{f(a + Bz; 0)}.$$

Theorem 3. *Under Assumption 2, the rejection region of the LBI test for testing normality $H_0 : \theta = 0$ vs. $H_1 : \theta = \theta_1 > 0$ under the action of $\mathbb{R}^p \times GL(p)$ is given by*

$$\sum_{i=1}^n \|z_i\|^4 > k.$$

The rejection region of the LBI test under the action of $\mathbb{R}^p \times LT(p)$ is given by

$$(n+p+2)(n+p)\sum_{i=1}^{n}\|z_i\|^4 - 2(n+p+2)\sum_{i=1}^{n}\sum_{j,k=1}^{p}\max(j,k)z_{ij}^2 z_{ik}^2$$
$$-2(n+p)\sum_{i=1}^{n}\sum_{j,k=1}^{p}\min(j,k)z_{ij}^2 z_{ik}^2 + 4\sum_{i=1}^{n}(\sum_{j=1}^{p} j z_{ij}^2)^2 > k',$$

where z_{ij} is the jth element of z_i.

The lemma below is used in proving Theorem 3. This is easily proved by some standard Jacobian formulas in the multivariate analysis (e.g., page 86 of Muirhead (1982)).

Lemma 1. *Let $Sym(p)$ denote the set of $p \times p$ real symmetric matrices. Define a map $\varphi : \mathbb{R}^{p\times p} \to Sym(p)$ by $\varphi(B) = nB'B$. Then, for any measurable set $D \subset Sym(p)$,*

$$\int_{\varphi(B)\in D} \exp(-\frac{n}{2}\mathrm{tr}(B'B))|\det B|^{n-p-1}dB \propto \int_D \exp(-\frac{1}{2}\mathrm{tr}\, C)(\det C)^{\frac{1}{2}(n-p-2)}dC,$$

where $C = (c_{ij}) \in Sym(p)$ and $dC = \prod_{i\geq j} dc_{ij}$.

Proof of Theorem 3. Note first that $\sum_{i=1}^n \|a+Bz_i\|^2 = n\|a\|^2+n\mathrm{tr}(B'B)$ because $\sum_{i=1}^n z_i = 0$ and $\sum_{i=1}^n z_i z_i' = nI_p$. The second and the third terms of ℓ_θ are irrelevant to z_i's.

In the case of $\mathbb{R}^p \times GL(p)$, the rejection region is of the form $\sum_{i=1}^n I(z_i) > k$, where

$$I(z) = \int_{GL(p)}\int_{\mathbb{R}^p} \|a+Bz\|^4 \exp(-\frac{n}{2}\|a\|^2 - \frac{n}{2}\mathrm{tr}(B'B))|\det B|^{n-p-1}dadB.$$

By Lemma 1 the integral of a function of $B'B$ can be replaced by taking expectation with respect to the Wishart distribution $nB'B \sim W_p(n-1, I_p)$. On the other hand, the integration with respect to a is regarded as the expectation with respect to $\sqrt{n}a \sim N_p(0, I_p)$. Note that for the Wishart matrix $C \sim W_p(n-1, I_p)$, it holds that

$$E[z'Cz] = (n-1)\|z\|^2, \qquad E[(z'Cz)^2] = (n-1)(n+1)\|z\|^4.$$

By taking expectations of

$$\begin{aligned}\|a+Bz\|^4 &= (\|a\|^2 + 2a'Bz + z'B'Bz)^2 \\ &= (z'B'Bz)^2 + 2\|a\|^2(z'B'Bz) \\ &\quad +(\text{terms of odd degrees in } a) + (\text{a term independent of } z),\end{aligned}$$

we see that $\sum_{i=1}^n I(z_i)$ is proportional to $\sum_{i=1}^n \|z_i\|^4 + \text{const}$.

In the case of $\mathbb{R}^p \times LT(p)$, the rejection region is of the form $\sum_{i=1}^n I(z_i) > k'$, where

$$I(z) = \int_{LT(p)}\int_{\mathbb{R}^p} \|a+Tz\|^4 \exp(-\frac{n}{2}\|a\|^2 - \frac{n}{2}\mathrm{tr}(T'T))da\prod_{i=1}^{n} t_{ii}^{n-i-1}\prod_{i\geq j} dt_{ij}.$$

The integration with respect to T is reduced to taking expectations $nt_{ii}^2 \sim \chi^2_{n-i-1}$ and $\sqrt{n}t_{ij} \sim N(0,1)$ $(i > j)$. The details are given in Appendix D. □

A Proofs

A.1 Proof of Theorem 1

By the mean value theorem

$$f(x;\theta) = f(x;0) + \theta \frac{\partial}{\partial\theta} f(x;\theta^*),$$

where $0 < \theta^* = \theta^*(x) < \theta$. Then

$$\begin{aligned}\prod_{i=1}^n f(a+bz_i;\theta) &= \prod_{i=1}^n \Big(f(a+bz_i;0) + \theta \frac{\partial}{\partial\theta} f(a+bz_i;\theta^*)\Big) \\ &= \prod_{i=1}^n f(a+bz_i;0) \times (1 + \theta \sum_{i=1}^n \frac{\frac{\partial}{\partial\theta} f(a+bz_i;\theta^*)}{f(a+bz_i;0)} + \theta^2 R)\end{aligned}$$

where

$$R = \sum_{l=2}^n \theta^{l-2} \sum_{1\le i_1<\cdots<i_l\le n} \frac{\frac{\partial}{\partial\theta} f(a+bz_{i_1};\theta^*)}{f(a+bz_{i_1};0)} \cdots \frac{\frac{\partial}{\partial\theta} f(a+bz_{i_l};\theta^*)}{f(a+bz_{i_l};0)}.$$

By Assumption 1

$$\int_0^\infty \int_{-\infty}^\infty |R| \exp\Big(-\frac{n(a^2+b^2)}{2}\Big) b^{n-2} da db < \infty. \tag{29}$$

Furthermore by the continuous differentiability of $f(x;\theta)$ with respect to θ and the dominated convergence theorem we have

$$\begin{aligned}&\int_0^\infty \int_{-\infty}^\infty \sum_{i=1}^n \frac{\frac{\partial}{\partial\theta} f(a+bz_i;\theta^*)}{f(a+bz_i;0)} \exp\Big(-\frac{n(a^2+b^2)}{2}\Big) b^{n-2} da db \\ &\quad \to \int_0^\infty \int_{-\infty}^\infty \sum_{i=1}^n l_\theta(a+z_i;0) \exp\Big(-\frac{n(a^2+b^2)}{2}\Big) b^{n-2} da db \qquad (\theta \to 0).\end{aligned}$$

Now the theorem follows by the standard argument on the locally most powerful test (e.g. Section 4.8 of Cox & Hinkley (1974)). □

A.2 Proof of Theorem 2

In the proof, $M > 0$ denotes some suitable constant. Since Assumption 1 is not applicable for Theorem 2, we have to prove the finiteness of (29) by a separate argument. It suffices to prove that for each subsequence $1 \le i_1 < \cdots < i_l \le n$

$$\int_0^\infty \int_{-\infty}^\infty \Big| \frac{\partial}{\partial\theta} f(a+bz_{i_1};\theta^*) \cdots \frac{\partial}{\partial\theta} f(a+bz_{i_l};\theta^*) \Big| \exp\left(-\frac{1}{4}\sum_{k\ne i_j}(a+bz_k)^2\right) b^{n-2} da db < \infty.$$

Without loss of generality consider $i_1 = 1, \ldots, i_l = l$ and write

$$W_l = \left| \frac{\partial}{\partial\theta} f(a + bz_1; \theta^*) \cdots \frac{\partial}{\partial\theta} f(a + bz_l; \theta^*) \right| \exp\left(-\frac{1}{4} \sum_{k=l+1}^{n} (a + bz_k)^2 \right) b^{n-2}.$$

For evaluations of W_l we need the following property of the score function of general stable distributions. It follows from Lemma 3.1 of Matsui (2005).

Lemma 2. *For $\alpha = 2 - \theta \neq 1$, $|(\partial/\partial\theta) f(x;\theta)|$ is bounded and uniformly continuous in x. Furthermore as $\theta = 2 - \alpha \downarrow 0$, there exist $M > 0$, $x_0 > 0$, such that*

$$\left| \frac{\partial}{\partial\theta} f(x;\theta) \right| \leq M \cdot |x|^{\theta - 3} \log |x|, \qquad \forall |x| \geq x_0.$$

The integrability of W_l for $l \leq n-1$ follows from that of W_n, since $\exp(-1/4x^2) \leq M \cdot |\partial/\partial\theta f(x;\theta^*)|$ from Lemma 2. However, the integrability of W_n needs a very detailed argument. We replace a by $r = a + bz_1$, then W_n becomes

$$W_n(r,b) \equiv \prod_{k=1}^{n} \left| \frac{\partial}{\partial\theta} f(r + b(z_k - z_1); \theta^*) \right| b^{n-2}. \tag{30}$$

Note that $z_k - z_1 \neq 0$ implies

$$\begin{aligned} &\exists c > 0 \quad \text{s.t.} \quad \forall k \neq 1 \quad |c(z_k - z_1)| > 2, \\ &b > cx > c|r| \ \Rightarrow\ |r + b(z_j - z_1)| > x. \end{aligned} \tag{31}$$

Now we divide the integral of (30) into three parts

$$\left(\int_{|r| \leq x_0} \int_0^\infty + \int_{|r| > x_0} \int_{b \leq c|r|} + \int_{|r| > x_0} \int_{b > c|r|} \right) W_n(r,b) dr db \equiv I_1 + I_2 + I_3.$$

Using Lemma 2 and (31) in I_1, we have

$$\begin{aligned} I_1 \ \leq\ & \int_{|r| \leq x_0} \int_{b \leq cx_0} W_n(r,b) dr db \\ & \qquad + M \cdot \int_{|r| \leq x_0} \int_{b > cx_0} \max_k \left(\frac{\log |r + b(z_k - z_1)|}{|r + b(z_k - z_1)|^{3-\theta^*}} \right)^{n-1} b^{n-2} dr db < \infty \\ <\ & \infty. \end{aligned}$$

For I_2 the following lemma is useful.

Lemma 3. *Suppose that $\{z_k \neq 0 : k \in n,\ z_k \neq z_j\}$ are given. Then*

$$\frac{W_n(r,b)}{\left| \frac{\partial}{\partial\theta} f(r;\theta^*) \right|} = \prod_{k=2}^{n} \left| \frac{\partial}{\partial\theta} f(r + b(z_k - z_1); \theta^*) \right| \cdot b^{n-2} \tag{32}$$

is bounded in $-\infty < r < \infty$ and $b > 0$.

Proof. Assume that (32) is not bounded. Choose a sequence of (r, b) such that (32) diverges to ∞. Since the terms in the absolute value on the right-hand side are bounded, b has to go to ∞. By the assumption we can choose r such that for some k,

$$|r + b(z_k - z_1)| < c' b^\gamma,$$

where $c' > 0$ is a constant and $0 \le \gamma < 1$ (otherwise (32) converges to 0 as $b \uparrow \infty$ from Lemma 2). Then for $k \neq l$ we have

$$|r + b(z_k - z_1) - \{r + b(z_l - z_1)\}| = b|z_k - z_l|.$$

Hence as $b \uparrow \infty$

$$|r + b(z_l - z_1)| > x_0. \tag{33}$$

Furthermore, since $c' b^\gamma < b|z_l - z_1|$ for sufficiently large b, the triangular equality gives

$$\left| \frac{r}{\sqrt{b}} + \sqrt{b}(z_l - z_1) \right| \ge \left| c' b^{\gamma - 1/2} - \sqrt{b}|z_k - z_l| \right| \uparrow \ \infty, \quad \text{as } b \uparrow \infty.$$

Therefore, from Lemma 2 and (33), as $b \uparrow \infty$ the left-hand side of (32) approaches

$$\begin{aligned} &\left| \frac{\partial}{\partial \theta} f(r + b(z_k - z_1); \theta^*) \right| \cdot \prod_{l \neq k} \frac{\log |r + b(z_l - z_1)|}{|r + b(z_l - z_1)|^{3-\theta^*}} \cdot b^{n-2} \\ &\le M \cdot \prod_{l \neq k} \frac{\log |r + b(z_l - z_1)|}{|r + b(z_l - z_1)|^{1-\theta^*}} \frac{1}{|r/\sqrt{b} + \sqrt{b}(z_l - z_1)|^2} \quad \downarrow \ 0, \end{aligned}$$

regardless of selection of k. This is a contradiction and the proof is over. □

By Lemma 3 we get

$$\begin{aligned} I_2 \quad &\le \sup_{r,b} \left\{ W_n(r, b) \Big/ \left| \frac{\partial}{\partial \theta} f(r; \theta^*) \right| \right\} \cdot \int_{|r| > x_0} \int_{b \le c|r|} \left| \frac{\partial}{\partial \theta} f(r; \theta^*) \right| dbdr \\ &\le M \cdot \int_{|r| > x_0} \left| \frac{\partial}{\partial \theta} f(r; \theta^*) \right| \cdot 2c|r| dr \ < \infty. \end{aligned}$$

Finally for I_3 from Lemma 2 and (31),

$$I_3 \le M \cdot \int_{|r| > x_0} \frac{\log r}{|r|^{3-\theta^*}} \int_{b > c|r|} \max_k \left(\frac{\log |r + b(z_k - z_1)|}{|r + b(z_k - z_1)|^{3-\theta^*}} \right)^{n-1} b^{n-2} dbdr.$$

Since for large $x > 0$, $(\log x)^{n-1} \le x$, we have

$$\int_{b > c|r|} \left(\frac{\log |r + b(z_k - z_1)|}{|r + b(z_k - z_1)|^{3-\theta^*}} \right)^{n-1} b^{n-2} db \le M \cdot \int_{b > c|r|} |r + b(z_k - z_1)|^{-(n-1)(3-\theta^*)+1} b^{n-2} db.$$

The right-hand side is bounded by the equation 2.111, 2 on p.67 of Gradshteyn & Ryzhik (2000):

$$\int \frac{x^l}{z_1^m}dx = \frac{x^l}{z_1^{m-1}(l+1-m)b'} - \frac{na'}{(l+1-m)b'}\int \frac{x^{l-1}}{z_1^m}dx,$$

where $z_1 = a' + b'x$ and a', b' are constants. By induction we obtain

$$\begin{aligned}\int \frac{x^l}{(a'+b'x)^m}dx &= -\frac{x^l}{(m-l-1)(a'+b'x)^{m-1}b'} \\ &\quad -\sum_{k=1}^{l}\frac{l(l-1)\cdots(l+1-k)a'^k x^{l-k}}{(m-l-1)\cdots(m-l-1+k)(a'+b'x)^{m-1}b'^{k+1}}.\end{aligned}$$

Letting $a' = r$, $b' = (z_k - z_1)$, $m = \lfloor (n-1)(3-\theta^*)\rfloor - 1$, $l = n-2$, in the equation above and utilizing Lemma 2, we have

$$\int_{b>c|r|}\left(\frac{\log|r+b(z_k-z_1)|}{|r+b(z_k-z_1)|^{3-\theta^*}}\right)^{n-1} b^{n-2}db \le M \cdot \frac{n-1}{|r|^{\lfloor (n-1)(3-\theta)\rfloor - n}}.$$

Since the right-hand side is integrable with respect to $r > x_0$, we have $I_3 < \infty$. This completes the proof.

B Question of parametrization

Here we briefly discuss how to choose $a(\theta)$ and $b(\theta)$ in (2). Write $l(x;\theta) = \log f(x;\theta)$. Under the assumption that the 3×3 Fisher information matrix exists at $(a(\theta), b(\theta), \theta)$, it is convenient to determine $(a'(\theta), b'(\theta))$ in such a way that $(d/d\theta)l_{a(\theta),b(\theta)}(x;\theta)$ is orthogonal to the location-scale family in the sense of Fisher information, i.e.

$$\int \frac{d}{d\theta}l_{a(\theta),b(\theta)}(x;\theta)\frac{\partial}{\partial a}l_{a(\theta),b(\theta)}(x;\theta)f_{a(\theta),b(\theta)}(x;\theta)dx = 0 \tag{34}$$

$$\int \frac{d}{d\theta}l_{a(\theta),b(\theta)}(x;\theta)\frac{\partial}{\partial b}l_{a(\theta),b(\theta)}(x;\theta)f_{a(\theta),b(\theta)}(x;\theta)dx = 0 \tag{35}$$

These give a system of differential equations for $(a(\theta), b(\theta))$.

Actually we are only concerned in the neighborhood of the normal distribution and we only consider determining $(a'(0), b'(0))$. At $\theta = 0$, $l(x;0) = -(1/2)\log(2\pi) - x^2/2$. Therefore

$$\frac{\partial}{\partial a}l_{a(\theta),b(\theta)}(x;\theta)_{|\theta=0} = x, \quad \frac{\partial}{\partial b}l_{a(\theta),b(\theta)}(x;\theta)_{|\theta=0} = x^2$$

$$\frac{d}{d\theta}l_{0,1}(x;\theta) = -\frac{1}{b'(0)} + b'(0)x^2 + a'(0) + l_\theta(x;0)$$

and (34), (35) reduce to

$$\int (-\frac{1}{b'(0)} + b'(0)x^2 + a'(0)x + l_\theta(x;0))x^k\phi(x)dx = 0, \qquad k = 1, 2,$$

which can be solved for $a'(0)$ and $b'(0)$.

Note that we do not necessarily have to explicitly solve for $a'(0)$ and $b'(0)$. Instead for theoretical developments we can use the fact that the standard member $f(x;\theta)$ can be chosen in such a way that

$$\int l_\theta(x;0)x^k\phi(x)dx = 0, \qquad k = 1,2. \tag{36}$$

When $l_\theta(x;0)$ is a polynomial in x, (36) shows that we can choose $l_\theta(x;0)$ such that it is cubic or of higher degree in x. This is enough for simplifying our treatment of mixing distribution in Section 3.1.

C Details in the case of polynomial score function

Here we write out coefficients of LBI in the case of polynomial score function (cf. (7)). Suppose that $l_\theta(x;0)$ is given as

$$l_\theta(x;0) = c_0x^k + c_1x^{k-1} + \cdots + c_k = \sum_{j=0}^{k} c_{k-j}x^j.$$

Then

$$\sum_{i=1}^{n} l_\theta(a+bz_i;0) = \sum_{i=1}^{n}\sum_{j=0}^{k} c_{k-j}(a+bz_i)^j = n\sum_{j=0}^{k} c_{k-j}\sum_{l=0}^{j}\binom{j}{l}a^lb^{j-l}\tilde{m}_{j-l}.$$

Using (14) for even l we have

$$\int_0^\infty\int_{-\infty}^\infty a^lb^{j-l}\exp(-\frac{n(a^2+b^2)}{2})b^{n-2}dadb = \frac{2^{(n+j-2)/2}}{n^{(n+j)/2}}\Gamma(\frac{l+1}{2})\times\Gamma(\frac{n+j-l-1}{2}).$$

For odd l the integral is zero. Also we only consider $j-l\geq 3$. Hence the LBI test statistic is given as

$$\sum_{j=3}^{k} c_{k-j}\left(\frac{2}{n}\right)^{(n+j-2)/2}\sum_{l=0,\ l:\text{even}}^{j-3}\binom{j}{l}\Gamma(\frac{l+1}{2})\times\Gamma(\frac{n+j-l-1}{2})\tilde{m}_{j-l}. \tag{37}$$

D Moments of $z'T'Tz$

Let $T = (t_{ij}) \in LT(p)$ be a random matrix whose diagonal and lower off-diagonal elements are independently distributed as $t_{ii} \sim \chi_{m+p-i}$, $t_{ij} \sim N(0,1)$ $(i > j)$, where $m > 0$ is a constant. Let $z = (z_1,\ldots,z_p)' \in \mathbb{R}^p$ be a constant vector. In this section we evaluate the expectations

$$R_p(z) = E[z'T'Tz], \qquad S_p(z) = E[(z'T'Tz)^2]$$

required in proving Theorem 3.

Write $z_2 = (z_i)_{2\leq i\leq p}$, $t_{21} = (t_{i1})_{2\leq i\leq p}$ and $T_{22} = (t_{ij})_{2\leq i,j\leq p}$. Then z and T are represented as block matrices

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \qquad T = \begin{pmatrix} t_{11} & 0 \\ t_{21} & T_{22} \end{pmatrix}.$$

Note that

$$\begin{aligned} z'T'Tz &= (z_1, z_2') \begin{pmatrix} t_{11} & t_{21}' \\ 0 & T_{22}' \end{pmatrix} \begin{pmatrix} t_{11} & 0 \\ t_{21} & T_{22} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= z_1^2(t_{11}^2 + t_{21}'t_{21}) + 2z_1 t_{21}' T_{22} z_2 + z_2' T_{22}' T_{22} z_2. \end{aligned}$$

By taking the expectation with respect to $t_{11}^2 \sim \chi^2_{m+p-1}$, $t_{21} \sim N_{p-1}(0, I_{p-1})$, we have

$$\begin{aligned} R_p(z) &= z_1^2(m+p-1+p-1) + R_{p-1}(z_2) \\ &= z_1^2(m+2p-2) + R_{p-1}(z_2, \ldots, z_p) \\ &= \sum_{i=1}^{p} z_i^2(m+2p-2i). \end{aligned}$$

Also,

$$\begin{aligned} (z'T'Tz)^2 &= \{z_1^2(t_{11}^2 + t_{21}'t_{21}) + 2z_1 t_{21}' T_{22} z_2 + z_2' T_{22}' T_{22} z_2\}^2 \\ &= z_1^4(t_{11}^2 + t_{21}'t_{21})^2 + 4z_1^2(t_{21}' T_{22} z_2)^2 + (z_2' T_{22}' T_{22} z_2)^2 \\ &\quad +4z_1^3(t_{11}^2 + t_{21}'t_{21}) t_{21}' T_{22} z_2 + 2z_1^2(t_{11}^2 + t_{21}'t_{21}) z_2' T_{22}' T_{22} z_2 \\ &\quad +4z_1 t_{21}' T_{22} z_2 z_2' T_{22}' T_{22} z_2. \end{aligned}$$

Noting that $E[(\chi^2_\nu)^2] = \nu(\nu+2)$, we have

$$\begin{aligned} S_p(z) &= z_1^4(m+2p-2)(m+2p) + 4z_1^2 R_{p-1}(z_2) + S_{p-1}(z_2) \\ &\quad +2z_1^2(m+2p-2) R_{p-1}(z_2) \\ &= z_1^4(m+2p-2)(m+2p) + 2z_1^2(m+2p) R_{p-1}(z_2) + S_{p-1}(z_2) \\ &= (z_1^2, \ldots, z_p^2) A_p \begin{pmatrix} z_1^2 \\ \vdots \\ z_p^2 \end{pmatrix}, \end{aligned}$$

where

$$A_p = \left(\begin{array}{c|c} (m+2p)(m+2p-2) & * \\ \hline (m+2p)(m+2p-4) & \\ (m+2p)(m+2p-6) & \\ \vdots & A_{p-1} \\ (m+2p)m & \end{array} \right)$$

$$
\begin{aligned}
&= \begin{pmatrix} m+2p & m+2p & \cdots & m+2p \\ m+2p & m+2p-2 & \cdots & m+2p-2 \\ \vdots & \vdots & & \\ m+2p & m+2p-2 & & m+2 \end{pmatrix} \odot \begin{pmatrix} m+2p-2 & m+2p-4 & & m \\ m+2p-4 & m+2p-4 & & m \\ & & & \vdots \\ m & m & \cdots & m \end{pmatrix} \\
&= [m+2p+2-2\min(i,j)] \odot [m+2p-2\max(i,j)]_{1\le i,j\le p} \\
&= [(m+2p+2)(m+2p) - 2(m+2p+2)\max(i,j) - 2(m+2p)\min(i,j) \\
&\quad +4\max(i,j)\min(i,j)]_{1\le i,j\le p}.
\end{aligned}
$$

Here $\odot$ denotes the elementwise multiplication of matrices. This means

$$
\begin{aligned}
S_p(z) &= (m+2p+2)(m+2p)(\sum_{i=1}^{p} z_i^2)^2 - 2(m+2p+2)\sum_{i,j=1}^{p} \max(i,j) z_i^2 z_j^2 \\
&\quad -2(m+2p)\sum_{i,j=1}^{p} \min(i,j) z_i^2 z_j^2 + 4(\sum_{i=1}^{p} i z_i^2)^2.
\end{aligned}
$$

References

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Philosophical Transactions of the Royal Society of London, Series A, Math and Physical Science* **353**, 401–419.

Barndorff-Nielsen, O. E. & Shephard, N. (2001). Modelling by Lévy processess for financial econometrics. In *Lévy Processes*, O. E. Barndorff-Nielsen & T. Mikosch, eds. Boston, MA: Birkhäuser, pp. 283–318.

Bleistein, N. & Handelsman, R. A. (1986). *Asymptotic Expansions of Integrals.* New York: Dover.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

Dreier, I. & Kotz, S. (2002). A note on the characteristic function of the t-distribution. *Statist. Probab. Lett.* **57**, 221–224.

Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance. In *Lévy Processes*, O. E. Barndorff-Nielsen & T. Mikosch, eds. Boston, MA: Birkhäuser, pp. 319–336.

Ferguson, T. S. (1961). On the rejection of outliers. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, J. Neyman, ed. University of California Press, pp. 253–287.

Gradshteyn, I. S. & Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products., Sixth Edition.* San Diego, CA: Academic Press Inc.

Hájek, J. & Šidák, Z. (1967). *Theory of Rank Tests.* New York: Academic Press.

Hájek, J., Šidák, Z. & Sen, P. K. (1999). *Theory of Rank Tests.* Probability and Mathematical Statistics. San Diego, CA: Academic Press Inc.

HENZE, N. (2002). Invariant test for multivariate normality: A critical review. *Statistical Papers* **43**, 467–506.

KARIYA, T. & GEORGE, E. I. (1994). Locally best invariant tests for multivariate normality in curved families with μ known. In *Multivariate analysis and its applications*, T. W. Anderson, K.-T. Fang & I. Olkin, eds. Institute of Mathematical Statistics, pp. 311–322.

KARIYA, T. & GEORGE, E. I. (1995). LBI tests for multivariate normality in curved families and Mardia's test. *Sankhya, Series A, Indian Journal of Statistics* **57**, 440–451.

KURIKI, S. & TAKEMURA, A. (2001). Tail probabilities of the maxima of multilinear forms and their applications. *Ann. Statist.* **29**, 328–371.

KUWANA, Y. & KARIYA, T. (1991). LBI tests for multivariate normality in exponential power distributions. *Journal of Multivariate Analysis* **39**, 117–134.

MASUDA, H. (2002). Analytical properties of GIG and GH distributions. *Proc. Inst. Statist. Math.* **50**, 165–199. (in Japanese).

MATSUI, M. (2005). Fisher information matirix of genaral stable distributions close to the normal distribution. *Mathematical Methods of Statistics* **14**, 224–251.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.

PEARSON, E. S. (1935). A comparison of β_2 and Mr Geary's w_n criteria. *Biometrika* **27**, 333–352.

SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, vol. 68 of *Cambridge Studies in Advanced Mathematics.* Cambridge: Cambridge University Press. Translated from the 1990 Japanese original, Revised by the author.

TAKEMURA, A. (1993). Maximally orthogonally invariant higher order moments and their application to testing elliptically-contouredness. In *Proceedings of the Third Pacific Area Statistical Conference*, K. Matusita, M. L. Puri & T. Hayakawa, eds. VSP International Science Publishers, pp. 225–235.

THODE, H. C. (2002). *Testing for Normality.* New York: Marcel Dekker Inc.

UTHOFF, V. A. (1970). An optimum test property of two well-known statistics. *J. Amer. Statist. Assoc.* **65**, 1597–1600.

UTHOFF, V. A. (1973). The most powerful scale and location invariant test of the normal versus the double exponential. *Ann. Statist.* **1**, 170–174.

WIJSMAN, R. A. (1967). Cross-sections of orbits and their application to densities of maximal invariants. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics.* Berkeley, Calif.: Univ. California Press, pp. 389–400.

WIJSMAN, R. A. (1990). *Invariant Measures on Groups and Their Use in Statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 14. Hayward, CA: Institute of Mathematical Statistics.

ZOLOTAREV, V. M. (1986). *One-Dimensional Stable Distributions.* American Mathematical Society Translations of Mathematical Monographs, Vol. 65. Providence, RI.: American Mathematical Society. Translation of the original 1983 (in Russian).