Sankhyā: The Indian Journal of Statistics
1997, Volume 59, Series A, Pt. 1, pp. 42-59

TESTING BIVARIATE INDEPENDENCE AND NORMALITY

By WILBERT C.M. KALLENBERG
University of Twente, Enschede
TERESA LEDWINA*
and
EWARYST RAFAJLOWICZ
Technical University of Wrocław, Wrocław

SUMMARY. In many statistical studies the relationship between two random variables X and Y is investigated and in particular the question whether X and Y are independent and normally distributed is of interest. Smooth tests may be used for testing this. They consist of several components, the first measuring linear (in)dependence, the next ones correlations of higher powers of X and Y. Since alternatives are not restricted to bivariate normality, not only linear (in)dependence is of interest and therefore all components come in. Moreover, normality itself is also tested by the smooth test. It is well-known that choosing the number of components in a smooth test is an important issue. Recently, data driven methods are developed for doing this. The resulting new test statistics for testing independence and normality are introduced in this paper. For a very large class of alternatives, including also independent X and Y with nonnormal marginals, consistency is proved. Monte Carlo results show that the data driven smooth test behaves very well for finite sample sizes.

1. Introduction

In many statistical studies the relationship between two random variables X and Y is studied and in particular the question whether X and Y are independent is of interest. Assuming bivariate normality the problem reduces to testing the null hypothesis that the correlation coefficient equals 0 against

---

Paper received. September 1996; revised February 1997.

AMS (1991) *subject classification.* 62H15, 62 H20, 62 E25.

Keywords and phrases. Correlation, model selection, Hermite polynomials, consistency, Monte Carlo study.

* The research of Teresa Ledwina was supported by Grant KBN 350 044. Programming work has been done by Krzysztof Bogdan under support of KBN 350 044 Grant.

the alternative hypothesis that it is unequal to 0. The obvious test statistic is Pearson's sample correlation coefficient and under bivariate normality the problem is extensively studied and well understood, cf. e.g. Bickel and Doksum (1977, p. 219 ff), Lehmann (1986, p. 248 ff).

If we do not assume bivariate normality, Pearson's sample correlation coefficient may be replaced e.g. by its nonparametric counterpart, called Spearman's rank correlation coefficient. However, this is two steps at once. Both under the null hypothesis and under alternatives the bivariate normality assumption is abandoned. In many situations we want to keep normality under the null hypothesis, but want to be on guard against nonnormality. Therefore, nonnormal alternatives are considered as well. As a consequence, we do not only investigate *linear* (in)dependence (measured by the correlation coefficient of X and Y, being the only remaining parameter in case of bivariate normality), but take also other forms of (in)dependence into account, in particular correlations of higher powers of X and Y.

To tackle the problem with this in mind we propose smooth test statistics. They consist of several components, the first measuring linear (in)dependence, the next ones correlations of higher powers of X and Y. Recently, there is renewed interest in smooth tests for goodness of fit, cf. e.g. Rayner and Best (1989, 1990), Milbrodt and Strasser (1990), Eubank and LaRiccia (1992), Kaigh (1992), Inglot, Kallenberg and Ledwina (1994). Smooth tests are recommended for their flexibility, in contrast to classical tests as Kolmogorov-Smirnov and Cramér-von Mises tests, for which tests there are only few directions of deviations from the null hypothesis with reasonable asymptotic power. Since we want to consider all kind of nonnormal alternatives as well, here we are interested in a broad range of alternatives and hence we apply smooth tests.

Smooth tests for testing bivariate independence and normality are proposed by Javitz (1975, Section 9) and Koziol (1979), while the Rao score test in Section 5 of Mardia and Kent (1991) is related to it. Further we mention the tests for independence, given in Eubank, LaRiccia and Rosenstein (1987).

It is well-known that choosing the number of components in a smooth test is an important issue. Recently, data driven methods are developed for doing this, cf. Inglot, Kallenberg and Ledwina (1997) and references therein. The method can be described as follows. Exponential families of smooth alternatives are introduced. By Schwarz's selection rule a suitable dimension is chosen. In this appropriate exponential family the score test is applied. This method, introduced by Ledwina (1994), turned out to be very succesful as is seen by the Monte Carlo results and the theoretical optimality results in Kallenberg and Ledwina (1995 a,b,c,1997), Inglot and Ledwina (1996) and Inglot, Kallenberg and Ledwina (1997).

It is the aim of this paper to extend the method to testing bivariate independence and normality. Several new problems arise. The involved orthonormal systems used in previous research concerning data driven smooth tests, consist of bounded functions. In this paper the *unbounded* Hermite polynomials are used. Even at the very beginning the unboundedness leads to problems, because of nonexistence of the moment generating function, thus disturbing the straightforward introduction of the exponential family of smooth alternatives. Nevertheless, smooth tests can be defined. This is done in Section 2. Also the selection rule is presented in this section.

Section 3 shows the asymptotic null distribution of the selection rule and the data driven smooth test statistic. It turns out that the limiting distribution of the test statistic is simply chi-squared one. Since the first order approximation is not very precise for common sample sizes, a more accurate approximation is given.

The omnibus character of the test is shown in Section 4, where its consistency is proved, both at all kinds of dependent alternatives and at nonnormal alternatives.

The finite sample case is discussed in Section 5, where Monte Carlo results show its power behaviour at a great variety of alternatives. Moreover, the data driven smooth test is compared with the test based on Pearson's sample correlation coefficient and with the modified Hoeffding test for bivariate independence as introduced by Blum, Kiefer and Rosenblatt (1961). It turns out that the data driven smooth test behaves very well for testing bivariate independence and normality with a possible large gain in power in case of nonnormal marginals and only slightly less power for alternatives with normal marginals. In view of the simulation results and the consistency the data driven smooth test is recommended for testing bivariate independence and normality.

2. Test statistic

Let $(X_1, Y_1,) \dots, (X_n, Y_n)$ be i.i.d. r.v.'s with d.f. D. We want to test the null hypothesis that X_i and Y_i are independent and normally distributed. So we have

$$H_0 : D(x,y) = \Phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{y-\nu}{\tau}\right), \quad x, y \in \mathbb{R},$$

where Φ is the standard normal d.f. and μ, ν, σ, τ are nuisance parameters.

The first step in defining smooth tests is the introduction of smooth alternatives. Obviously many choices are possible. One of these is the choice of the orthonormal system. Under the null hypothesis we have normality. The Hermite polynomials are orthonormal on the standard normal density, which means that $\int H_i(x)H_j(x)\varphi(x)dx = 0$ if $i \neq j$ and 1 if $i = j$, where φ is the standard normal density and H_j the j^{th} normalized Hermite polynomial, defined by

$$H_j(x) = \frac{(-1)^j}{\sqrt{j!}} \frac{1}{\varphi(x)} \frac{d^j\varphi(x)}{dx^j}. \qquad \dots(2.1)$$

We list some further properties of H_j used in this paper (for proofs of those properties see e.g. Sansone (1959))

$$\begin{aligned}\int H_j(x)\varphi(x)dx &= 0, \quad H_j'(x) = j^{1/2}H_j(x), |H_j(x)| < 1.087e^{x^2/4} \\ xH_j'(x) &= jH_j(x) + \{j(j-1)\}^{1/2}H_{j-2}(x).\end{aligned} \qquad \ldots(2.2)$$

Hence, $H_i(X)$ and $H_j(X)$ have for $i \neq j$ expectation 0, variance 1 and are uncorrelated when X has a normal distribution with expectation 0 and variance 1.

The orthonormality on φ makes the choice of this orthonormal system a natural one for the testing problem under consideration. Moreover, it implies that the linear independence of X and Y is investigated firstly, followed by correlations of higher powers of X and Y, cf. (2.4) and (2.5). Note also that Koziol's (1979) smooth test statistic (cf. (2.4)) for testing bivariate independence and normality, although introduced on an intuitive ground, is based upon Lancaster's canonical representation of the bivariate normal density in terms of Hermite polynomials. For an insightful treatment of such representations see Sarmanov and Bratoeva (1969).

Here we rederive Koziol's (1979) test statistic. Following the classical way of introducing smooth alternatives we consider functions

$$\exp\left\{\sum_{j=1}^{k}\theta_j H_j\left(\frac{x-\mu}{\sigma}\right)H_j\left(\frac{y-\nu}{\tau}\right)\right\}\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\tau}\varphi\left(\frac{y-\nu}{\tau}\right). \qquad \ldots(2.3)$$

Unfortunately, in many cases (2.3) is not integrable. For instance, if $k = 3$, (2.3) is never integrable if $\theta_3 \neq 0$. Therefore, an exponential family of smooth alternatives with densities w.r.t. Lebesgue measure on $\mathbb{R} \times \mathbb{R}$ proportional to (2.3), can not be defined. This problem also arises in deriving the smooth test for testing normality in the goodness-of-fit problem, when using Hermite polynomials. But it seems that Rayner and Best (1989, p. 81) in their derivation have overlooked it. The problem has been mentioned by Mardia and Kent (1991, p. 356). As they state, the score statistic makes sense even when (2.3) is not integrable. Since technically the smooth alternatives are only used to derive an appropriate test statistic, i.e. the score statistic, this solves the problem. (The reason that the lack of integrability of (2.3) is not so important might be, that our main interest is not on extreme large values of $|(x-\mu)/\sigma|$ and $|(y-\nu)/\tau|$.)

Therefore, we proceed as if (2.3) is a family of densities and as the second step apply the score statistic for testing $\theta = 0$, where $\theta = (\theta_1, \ldots, \theta_k)'$ and the transpose of a vector is denoted by $'$. The score statistic is given by the Koziol statistic, cf. Koziol (1979) p. 862,

$$K_k = K_k(\hat\mu, \hat\nu, \hat\sigma, \hat\tau) = \sum_{j=1}^{k}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_j\left(\frac{X_i-\hat\mu}{\hat\sigma}\right)H_j\left(\frac{Y_i-\hat\nu}{\hat\tau}\right)\right\}^2, \qquad \ldots(2.4)$$

where

$$\hat{\mu} = \bar{X} = n^{-1}\sum_{i=1}^{n} X_i\,, \quad \hat{\sigma} = \left\{ n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \right\}^{1/2},$$
$$\hat{\nu} = \bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i\,, \quad \hat{\tau} = \left\{ n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \right\}^{1/2}$$

are the maximum likelihood estimators under H_0. The first term in K_k is nothing else than n times the square of Pearson's sample correlation coefficient, thus measuring the linear (in)dependence. The following terms are measuring sample correlations of higher powers of X and Y.

The next step in defining data driven tests concerns the selection of the dimension k. We take the modified Schwarz's rule, similar to the one presented in Inglot, Kallenberg and Ledwina (1997), given by

$$S2 = \min\{k : 1 \le k \le d(n),\ K_k - k\log n \ge K_j - j\log n,\ j = 1,\dots,d(n)\}. \quad \dots (2.5)$$

Here $d(n)$ is a sequence of numbers tending to infinity as $n \to \infty$.

Finally, the test statistic is obtained by inserting $S2$ for k and the null hypothesis is rejected for large values of

$$KS2 = K_{S2}.$$

3. Asymptotic null distribution of $S2$ and $KS2$

The following theorem gives the asymptotic null distribution of $S2$ and $KS2$. By χ_1^2 we denote a r.v. with a chi-squared distribution with one degree of freedom.

THEOREM 3.1. *Assume that $d(n)$ tends to infinity and that $d(n) = o(logn)$. Under H_0 we have*

$$S2 \to 1 \ \textit{in probability}$$

and

$$KS2 \xrightarrow{D} \chi_1^2.$$

PROOF. It is easily seen from (2.4) that K_k is location-scale invariant. In view of (2.5) the same holds for $S2$ and $KS2$. Therefore, under H_0 we may assume that X_i and Y_i are $N(0,1)$-distributed and independent. The latter will be denoted by P_0 and E_0 for the probability measure and expectation, respectively.

Introduce also an auxiliary statistic

$$\mathcal{I}_k = \sum_{j=1}^{k} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_j(X_i) H_j(Y_i) \right\}^2$$

and set $a_k = (k-1)\log n$.

By the Markov inequality we have for any $k = 2, 3, \ldots$, cf. also (2.2),

$$P_0\left(\mathcal{I}_k \geq \tfrac{1}{4}a_k\right) \leq 4a_k^{-1} E_0 \mathcal{I}_k = 4a_k^{-1}k. \qquad \ldots(3.1)$$

The Appendix gives the proof that for $d(n)$ satisfying $d(n) = o(\{n/\log n\}^{1/6})$, we have

$$\lim_{n\to\infty} \sum_{k=2}^{d(n)} P_0\left(\left|K_k^{1/2} - \mathcal{I}_k^{1/2}\right| \geq \tfrac{1}{2}a_k^{1/2}\right) = 0. \qquad \ldots(3.2)$$

Since here we assume $d(n) = o(\log n)$, (3.2) holds. Therefore, for any $k \in \{2, \ldots, d(n)\}$, (3.1), (3.2) and $d(n) = o(\log n)$ imply

$$\begin{aligned} P_0(S2 \geq 2) &\leq \sum_{k=2}^{d(n)} P_0(K_k^{1/2} \geq a_k^{1/2}) \leq \sum_{k=2}^{d(n)} P_0\left(\mathcal{I}_k^{1/2} \geq \tfrac{1}{2}a_k^{1/2}\right) + o(1) \\ &\leq \sum_{k=2}^{d(n)} 4a_k^{-1}k + o(1) = o(1). \end{aligned} \qquad \ldots(3.3)$$

Since under P_0 $K_1 \xrightarrow{D} \chi_1^2$, the proof is easily completed. ▱

REMARK 3.1. The proof of Theorem 3.1 consists of two steps. The first one is replacing in $K_k(\hat{\mu}, \hat{\nu}, \hat{\sigma}, \hat{\tau})$ the estimators by the true values of the corresponding parameters, cf. (3.2). For this part we require $d(n) = o\left(\{n/\log n\}^{1/6}\right)$. The second step concerns the moderate deviation probabilities in (3.3). Often for this purpose, the moment generating function of the involved statistic is used. However, as mentioned already in Section 2, the moment generating function of $H_j(X)H_j(Y)$ does not exist on $(0, \infty)$ for $j \geq 3$. Another way to prove moderate deviation probabilities is to use the existence of certain moments. Although all moments of $H_j(X)H_j(Y)$ exist, absolute moments of order greater than 2 grow rapidly with j. Therefore, in (3.1) the second moment is used, resulting in the condition $d(n) = o(\log n)$. ▱

REMARK 3.2. The first-order approximation of $KS2$ induced by Theorem 3.1 is not very accurate. The same phenomenon occurs in data driven goodness-of-fit tests and can be overcome by using a second-order approximation. We follow the line of argument given in Section 4 of Kallenberg and Ledwina (1995a) and refer to it for more details. The idea is that $KS2 \geq K_1$ and therefore $P_0(KS2 \leq x)$ is overestimated by its first-order approximation, which is the χ_1^2-approximation. Replacing

$$n^{-1/2}\sum_{i=1}^{n} H_j\left(\frac{X_i-\hat{\mu}}{\hat{\sigma}}\right) H_j\left(\frac{Y_i-\hat{\nu}}{\hat{\tau}}\right), \quad j=1,2,$$

by U_j, $j=1,2$, where U_1, U_2 are independent and $N(0,1)$-distributed, we may approximate $P_0(KS2 \leq x)$ by

$$Pr(U_1^2 \leq x)Pr(U_2^2 \leq \log n) + Pr(U_1^2 + U_2^2 \leq x,\ U_2^2 > \log n).$$

Here the first term refers to $P_0(KS2 \leq x,\ S2 = 1)$, with moreover the event $\{S2 = 1\}$ essentially restricted to "dimension 1 beats dimension 2". The second term concerns $P_0(KS2 \leq x_1,\ S2 = 2)$ with $\{S2 = 2\}$ essentially restricted to dimension 2 beats dimension 1. A further simplification leads to the proposed approximation

$$\begin{array}{ll} \{2\Phi(\sqrt{x})-1\}\{2\Phi(\sqrt{\log n})-1\} & \text{if } x \leq \log n \\ \{2\Phi(\sqrt{x})-1\}\{2\Phi(\sqrt{\log n})-1\} + 2\Phi(-\sqrt{\log n}) & \text{if } x \geq 2\log n \\ \text{linearize} & \text{if } \log n \leq x \leq 2\log n. \end{array} \qquad \ldots(3.4)$$

It is seen that this approximation works very well by calculating approximation (3.4) at the simulated critical value 5.525 for $n = 50$ and $\alpha = 0.05$, giving 0.943 which is very close to $1-\alpha$. ▱

4. Consistency

We consider alternatives $(X_1, Y_1), \ldots, (X_n, Y_n)$, which are i.i.d. r.v.'s with probability measure $\mathbb{P}$, satisfying $E_{\mathbb{P}}\phi_j(X, Y; \beta) \neq 0$ for some j, where

$$\beta = (\mu, \nu, \sigma, \tau)' = \left(E_{\mathbb{P}}X, \sqrt{\mathrm{var}_{\mathbb{P}}X},\ E_{\mathbb{P}}Y, \sqrt{\mathrm{var}_{\mathbb{P}}Y}\right)',$$
$$\phi_j(x, y; \beta) = H_j(\sigma^{-1}(x-\mu))H_j(\tau^{-1}(y-\nu))$$

and $E_{\mathbb{P}}$, $\mathrm{var}_{\mathbb{P}}$ denote the expectation and variance, respectively, taken under $\mathbb{P}$. (Note that under H_0 we have $E\phi_j(X, Y; \beta) = 0$ for all j.) Let J be the smallest j for which $E_{\mathbb{P}}\phi_j(X, Y; \beta) \neq 0$.
So we have

$$E_{\mathbb{P}}\phi_1(X, Y; \beta) = \ldots = E_{\mathbb{P}}\phi_{J-1}(X, Y; \beta) = 0, E_{\mathbb{P}}\phi_J(X, Y; \beta) \neq 0. \qquad \ldots(4.1)$$

Implicitly it is assumed that $\mathrm{var}_{\mathbb{P}}X < \infty$ and $\mathrm{var}_{\mathbb{P}}Y < \infty$. Moreover, by speaking on $E_{\mathbb{P}}\phi_J(X, Y; \beta)$ it is also reasonable to assume that $E_{\mathbb{P}}|XY|^J <$

$\infty, E_{\mathbb{P}}|X|^{J} < \infty$ and $E_{\mathbb{P}}|Y|^{J} < \infty$.
Therefore we assume

$$E_{\mathbb{P}}|XY|^{J} < \infty \text{ and } E_{\mathbb{P}}|X|^{\max(J,2)} < \infty, E_{\mathbb{P}}|Y|^{\max(J,2)} < \infty. \qquad \ldots (4.2)$$

The consistency of the test statistic $KS2$ is given by the following theorem.

THEOREM 4.1. *Let $\mathbb{P}$ be any alternative satisfying (4.1) and (4.2). Then, if $d(n)$ tends to infinity,*

$$\mathbb{P}(S2 \geq J) \to 1, \ KS2 \xrightarrow{\mathbb{P}} \infty$$

and hence, if $d(n) = o(\log n)$, $KS2$ is consistent against any alternative of the form (4.1), satisfying (4.2), i.e. against any alternative $\mathbb{P}$ for which (4.2) holds and

$$E_{\mathbb{P}} H_j \left(\frac{X - \mu}{\sigma} \right) H_j \left(\frac{Y - \nu}{\tau} \right) \neq 0$$

for some j.

REMARK 4.1. The class of alternatives for which consistency is proved is very large. Essentially any correlation between the same powers of X and Y is detected. Moreover, also the normality, which is part of the null hypothesis, is tested. When X and Y are independent, but their marginal distributions are both nonnormal, then also consistency is obtained. Therefore, the proposed test is really an omnibus test for testing whether X_i and Y_i are independent and normally distributed, thus guarding against dependence as well as nonnormality. ▱

PROOF OF THEOREM 4.1. By location and scale invariance of the test statistic $KS2$, we may standardize and therefore assume w.l.o.g. that the true value of β equals $\beta_0 = (0, 0, 1, 1)'$. Define for any $0 < \delta < \frac{1}{2}$ the set

$$B_\delta = \{(\beta_1, \beta_2, \beta_3, \beta_4) : |\beta_1| < \delta, |\beta_2| < \delta, |\beta_3 - 1| < \delta, |\beta_4 - 1| < \delta\}.$$

Because H_j is a polynomial of degree j, there exists for any j a constant c_j, not depending on δ, such that $\hat{\beta} \in B_\delta$ implies

$$n^{-1} \sum_{i=1}^{n} \left| \phi_j(X_i, Y_i; \hat{\beta}) - \phi_j(X_i, Y_i; \beta_0) \right| \leq \delta c_j n^{-1} \sum_{i=1}^{n} (|X_i|^j + 1)(|Y_i|^j + 1).$$

Since δ may be chosen arbitrarily small, it now easily follows from (4.2) and the law of large numbers that for any fixed j

$$n^{-1} \sum_{i=1}^{n} \left| \phi_j(X_i, Y_i; \hat{\beta}) - \phi_j(X_i, Y_i; \beta_0) \right| \xrightarrow{\mathbb{P}} 0$$

and

$$n^{-1}\sum_{i=1}^{n}\phi_j(X_i,Y_i;\beta_0) \xrightarrow{\mathbb{P}} E_{\mathbb{P}}\phi_j(X,Y;\beta_0).$$

Therefore, for any $j \in \{1,\ldots,J-1\}$ we have, cf. (2.4) and (4.1),

$$n^{-1}K_j \xrightarrow{\mathbb{P}} 0 \text{ and } n^{-1}K_J \xrightarrow{\mathbb{P}} \{E_{\mathbb{P}}\phi_J(X,Y;\beta_0)\}^2 > 0. \qquad \ldots(4.3)$$

Since for any $j \in \{1,\ldots,J-1\}$

$$S2 = j \quad \text{implies} \quad n^{-1}K_j - jn^{-1}\log n \geq n^{-1}K_J - Jn^{-1}\log n,$$

(4.3) yields $\mathbb{P}(S2 = j) \to 0$ for any $j \in \{1,\ldots,J-1\}$ and thus $\mathbb{P}(S2 \geq J) \to 1$. For any $j \geq J$ it follows from (2.4) that $K_j \geq K_J$. By (4.3) $K_J \xrightarrow{\mathbb{P}} \infty$ and hence we get $KS2 \xrightarrow{\mathbb{P}} \infty$. The proof of Theorem 4.1 is now easily completed, using Theorem 3.1. ▱

5. Simulations

To see how well the new test based on $KS2$ works in the finite sample case, a simulation study has been performed. For a wide range of alternatives the new test is compared to the test based on Pearson's sample correlation coefficient and a version of Hoeffding's test for independence. Define

$$R = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})/(n\hat{\sigma}\hat{\tau}),$$

Pearson's sample correlation coefficient. The corresponding test is based on $|R|$. The modified Hoeffding's test has been introduced by Blum, Kiefer and Rosenblatt (1961). The test statistic is given by

$$n\int[F_n(x,y) - F_n(x,\infty)F_n(\infty,y)]^2 dF_n(x,y)$$

with

$$F_n(x,y) = n^{-1}\sum_{i=1}^{n} I\{X_i \leq x,\ Y_i \leq y\}$$

and $I\{A\}$ the indicator of the set A.

The following alternatives are presented (note that by location and scale invariance of the test statistics only standardized alternatives need to be investigated). We use the following abbreviations BD: Bickel and Doksum (1977), J: Johnson (1987), CJ: Cook and Johnson (1986).

alternative	reference	density- or contourplots
1. Bivariate normal $N(0,0,1,1,\rho)$	BD pp. 22-28 J pp. 49-55	J pp. 51-54
2. Generalized Burr-Pareto-logistic (α,β) a. $\beta=-1$ b. $\beta=1$	CJ p. 127 J pp. 170-179	J pp. 174, 177
3. Morgenstern (α)	J pp. 180-190	J pp. 185-187
4. Normal mixture	J pp. 55-62	
a. $0.5N(0,0,1,1,\rho)+$ $0.5N(1,1,1,1,0.9)$		J pp. 56, 58-61
b. $0.5N(0,0,1,1,\rho)+$ $0.5N(0,0,1,1,-\rho)$		
5. Lognormal (σ_1,σ_2,ρ)	J pp. 63, 65-69	
a. $\sigma_1=1,\sigma_2=1$		J p. 64
b. $\sigma_1=0.05,\sigma_2=0.5$		J p. 68
6. Sinh^{-1}-normal $(0,1,\mu_2,\sigma_2,\rho)$	J pp. 63, 70-76	
a. $\mu_2=0,\sigma_2=1$		J p. 74
b. $\mu_2=2,\sigma_2=0.5$		J p. 75
7. Logit-normal $(0,0.5,0,\sigma_2,\rho)$	J pp. 63, 76-82	
a. $\sigma_2=1$		J p. 81
b. $\sigma_2=2$		J p. 84
8. Pearson type VII (m) with $\mu=0$ and $\Sigma=I$	J pp. 117-121	J pp. 119-122

The alternatives 1, 2 and 3 have normal marginals. For $\rho=0$ the alternatives 5, 6 and 7 are independent. For $\rho=0$ the alternative 4b has correlation 0, but there is still dependence. The alternative 8 also represents uncorrelated but dependent components. Since the alternatives 4, 5, 6 and 7 have no normal marginals, even for $\rho=0$ they do not belong to H_0. The alternatives 1, 2, 3, 4, 5a, 6a and 8 are symmetric in X and Y, 5b, 6b and 7 are not. The alternatives 1 and 8 are elliptically contoured distributions.

These properties show that the alternatives indeed give a wide class of alternatives to the hypothesis of independence and bivariate normality.

To facilitate the perception of the power behaviour of the three tests under

consideration, the results are presented in the following figures. Note that the parameter ρ appearing in Fig. 2, 3, 4 and 5, and in the corresponding alternatives 4-7, is not the correlation coefficient of X and Y, but of the bivariate normal distribution used in the definition of these alternatives.)
For all figures curves correspond to: $KS2$ test: ————, Pearson's sample correlation coefficient test: — — — — and Hoeffding's test: - - - - - - - - .

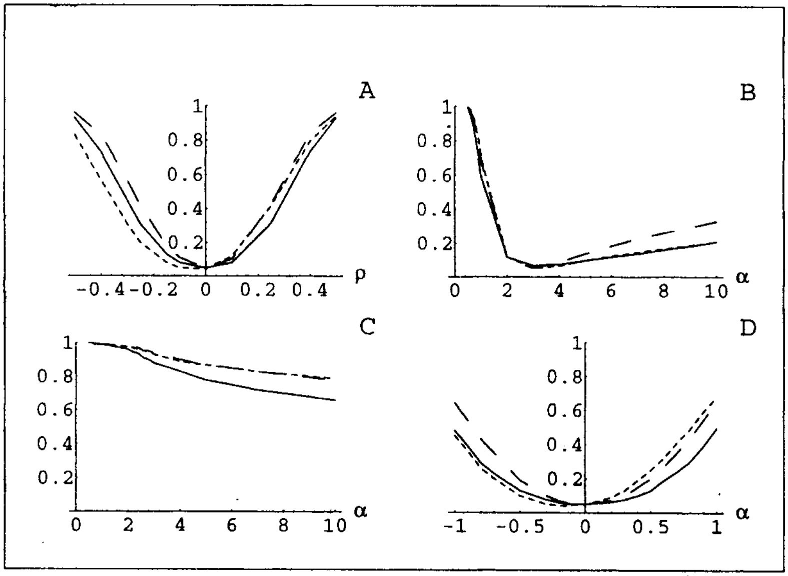

Figure 1: Powers of the tests for alternatives with normal marginals.
Plot A: N(0,0,1,1,ρ), ρ varying;
Plot B: Generalized Burr-Pareto-logistic (α,-1), α varying;
Plot C: Generalized Burr-Pareto-logistic (α,1), α varying;
Plot D: Morgenstern(α), α varying.

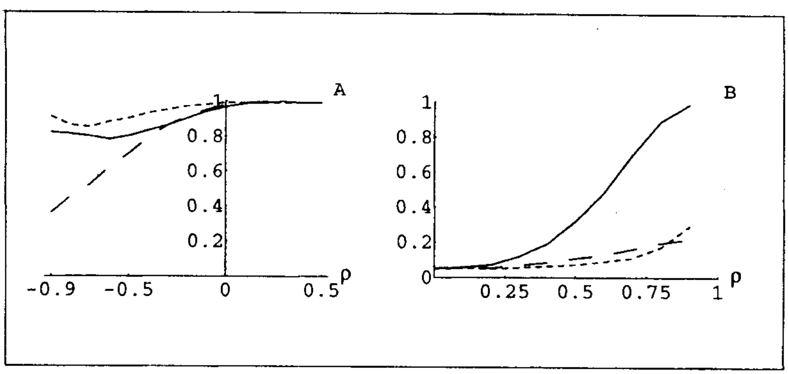

Figure 2: Powers of the tests for normal mixture alternatives.
Plot A: 0.5 N(0,0,1,1,ρ) + 0.5 N(1,1,1,1,0.9), ρ varying;
Plot B: 0.5 N(0,0,1,1,ρ) + 0.5 N(0,0,1,1,-ρ), ρ varying.

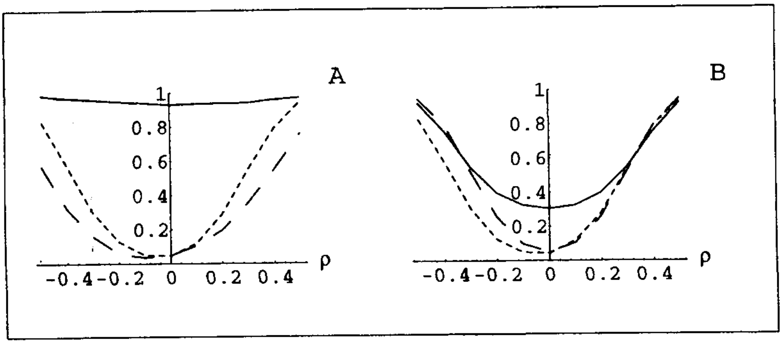

Figure 3: Powers of the tests for Johnson's lognormal $(\sigma_1, \sigma_2, \rho)$ alternatives.
Plot A: $\sigma_1 = 1$, $\sigma_2 = 1$, ρ varying;
Plot B: $\sigma_1 = 0.05$, $\sigma_2 = 0.5$, ρ varying.

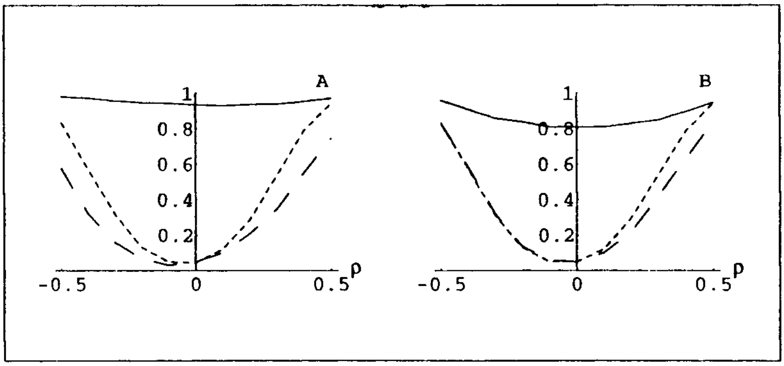

Figure 4: Powers of the tests for Johnson's $\sinh^{-1}$-normal $(0, 1, \mu_2, \sigma_2, \rho)$ alternatives.
Plot A: $\mu_2 = 0$, $\sigma_2 = 1$, ρ varying;
Plot B: $\mu_2 = 2$, $\sigma_2 = 0.5$, ρ varying.

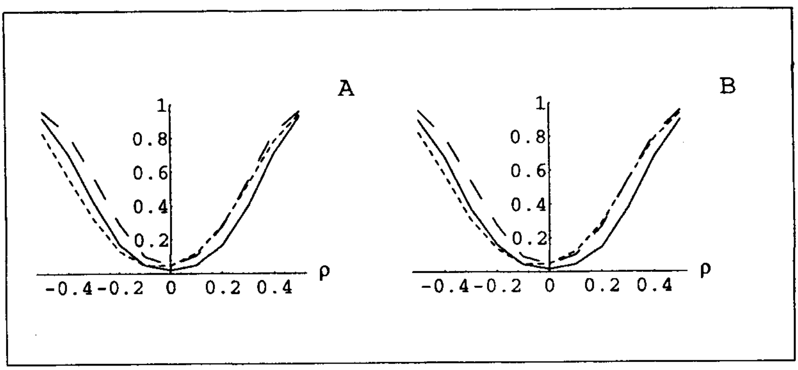

Figure 5: Powers of the tests for Johnson's logit-normal $(0, 0.5, 0, \sigma_2, \rho)$ alternatives.
Plot A: $\sigma_2 = 1$, ρ varying;
Plot B: $\sigma_2 = 2$, ρ varying.

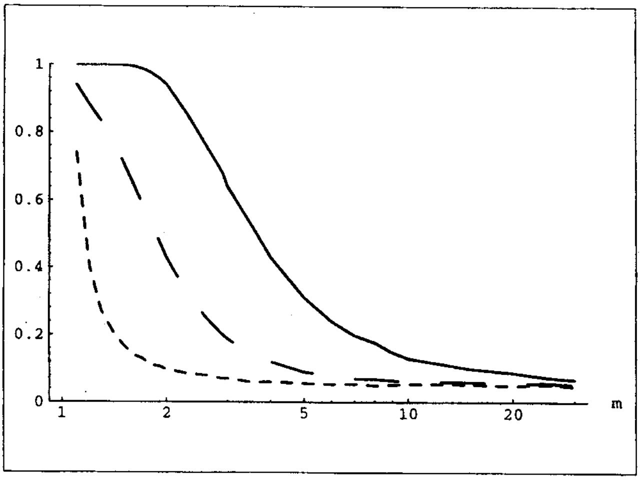

Figure 6: Powers of the tests for Pearson type VII (m) alternatives, m varying (plotted in the log scale).

From these figures and the other simulations which have been performed, we come to the following conclusions.

1. The data driven smooth test behaves very well for testing bivariate independence and normality.
2. For nonnormal marginals a large gain in power may occur when using the data driven smooth test, especially in case of small correlation, cf. Fig. 2, 3, 4 and 6.
3. For normal marginals, cf. Fig. 1, the power of Pearson's sample correlation coefficient test and Hoeffding's test is only slightly larger.
4. The test based on Pearson's sample correlation coefficient and Hoeffding's test are quite comparable, except for some normal mixtures (cf. alternative 4a in Fig. 2) and Pearson's type VII (cf. Fig. 6).
5. Although not presented in the figures, it turns out that the behaviour of Koziol's test with *fixed k* is very unreliable with large differences in power for different choices of k.
6. In contrast to Koziol's test the data driven smooth test is very stable w.r.t. $d(n)$, unless $d(n)$ is very small (smaller than 5 e.g. for $n = 50$).

REMARK 5.1 For alternative 4b the correlation coefficient equals 0, while still having dependence between X and Y. Nevertheless, there is not much difference in power behaviour between Pearson's sample correlation coefficient test and Hoeffding's test. Both have much lower power than the data driven smooth test, cf. Fig. 2. A similar situation occurs for alternative 8, cf. Fig. 6.

A classical example of correlation 0 and dependence of X and Y is the case where X is standard normal and $Y = X^2$. The simulated power of Pearson's sample correlation coefficient test in that case equals 0.38 for $n = 50$, while Hoeffding's test and the data driven smooth test have simulated power 1.00. For $Y = X^{-2}$ an even more extreme situation takes place. Simulated powers are 0 and 1, respectively. ▱

Appendix

PROOF OF (3.2). The proof will be given via several lemmas. We use the following notation

$$\begin{aligned}
\beta &= (\beta_1, \beta_2, \beta_3, \beta_4)' = (\mu, \nu, \sigma, \tau)', \beta_0 = (0,0,1,1)', \\
\phi_j(x, y; \beta) &= H_j\left(\frac{x-\mu}{\sigma}\right) H_j\left(\frac{y-\nu}{\tau}\right), \\
Y_n(\beta) &= n^{-1/2}\sum_{i=1}^{n}(\phi_1(X_i, Y_i; \beta), \ldots, \phi_k(X_i, Y_i; \beta))', \\
R1_j &= (\hat\beta - \beta_0)'\left(n^{-1/2}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\phi_j(X_i, Y_i; \beta)\Big|_{\beta=\beta_0}\right), R1 = (R1_1, \ldots, R1_k)', \\
R2_j &= \tfrac{1}{2}n^{-1/2}\sum_{i=1}^{n}(\hat\beta - \beta_0)'\frac{\partial^2}{\partial\beta\partial\beta'}\phi_j(X_i, Y_i; \beta)\Big|_{\beta=\xi}(\hat\beta - \beta_0), \\
R2 &= (R2_1, \ldots, R2_k)', a_k = (k-1)\log n.
\end{aligned}$$

Note that P_0 means that X_i and Y_i are $N(0,1)$-distributed and independent. We assume in this appendix (apart from Lemma A.2) that

$$d(n) = o(\{n/\log n\}^{1/6}). \qquad \ldots (A.1)$$

LEMMA A.1. *For some ξ between $\hat\beta$ and β_0 we have*

$$n^{-1/2}\sum_{i=1}^{n}\phi_j(X_i, Y_i; \hat\beta) = n^{-1/2}\sum_{i=1}^{n}\phi_j(X_i, Y_i; \beta_0) + R1_j + R2_j.$$

PROOF. Taylor expansion of $\phi_j(X_i, Y_i; \beta)$ around $\beta = \beta_0$. ▱

Writing $\|\cdot\|$ for the Euclidean norm we get

$$K_k^{1/2} = \|Y_n(\hat\beta)\| \text{ and } \mathcal{I}_k^{1/2} = \|Y_n(\beta_0)\|.$$

Hence by Lemma A.1

$$\left|K_k^{1/2} - \mathcal{I}_k^{1/2}\right| = \left|\|Y_n(\hat\beta)\| - \|Y_n(\beta_0)\|\right| \le \|R1\| + \|R2\|. \qquad \ldots (A.2)$$

Let B_n be the set

$$|\hat{\beta}_t - \beta_{0t}| \le 2n^{-1/2}\{\log d(n)\}^{1/2}, \quad t = 1, \dots, 4.$$

The next lemmas show that $R1$ and $R2$ are on this set large only with small probability.

LEMMA A.2. *Assume* $d(n) = o(n^{1/3})$. *We have*

$$\lim_{n\to\infty} \sum_{k=2}^{d(n)} P_0\left(\|R1\| \ge \tfrac{1}{4}a_k^{1/2},\, B_n\right) = 0.$$

PROOF. Let

$$U_j = (U_{1j}, \dots, U_{4j})' = n^{-1/2}\sum_{i=1}^{n} \frac{\partial}{\partial\beta}\phi_j(X_i, Y_i; \beta)\,\Big|_{\beta=\beta_0},$$

then $R1_j = (\hat{\beta} - \beta_0)'U_j$ and hence on the set B_n the Schwarz inequality yields

$$|R1_j|^2 \le 16n^{-1}\log d(n) \sum_{t=1}^{4} U_{tj}^2.$$

By the Markov inequality we get

$$P_0(\|R1\|^2 \ge \frac{1}{16}a_k,\, B_n) \le P_0\left(\sum_{j=1}^{k}\sum_{t=1}^{4} U_{tj}^2 \ge \frac{1}{256}\,\frac{na_k}{\log d(n)}\right) \le$$
$$\frac{256\log d(n)}{na_k}\sum_{j=1}^{k}\sum_{t=1}^{4} E_0U_{tj}^2.$$

Since $E_0U_{tj} = 0$, we have $E_0U_{tj}^2 = \mathrm{var}_0 U_{tj} = j$ if $t = 1, 3$ and $j(2j-1)$ if $t = 2, 4$, as is easily seen by direct calculation (cf. (2.2)). Therefore,

$$\sum_{k=2}^{d(n)} P_0\left(\|R1\| \ge \tfrac{1}{4}a_k^{1/2},\, B_n\right) \le \frac{1024d(n)^3\log d(n)}{n\log n},$$

which tends to 0 because $d(n) = o(n^{1/3})$. ▱

LEMMA A.3. *We have*

$$\lim_{n\to\infty} \sum_{k=2}^{d(n)} P_0\left(\|R2\| \ge \tfrac{1}{4}a_k^{1/2},\, B_n\right) = 0.$$

PROOF. It follows by straightforward calculations (cf. (2.2)) that on the set $|\mu| < \delta, |\nu| < \delta, |\sigma - 1| < \delta, |\tau - 1| < \delta$ with $\delta < 0.1$ we have for all $t, u \in \{1, \ldots, 4\}$

$$\left| \frac{\partial^2}{\partial\beta_t \partial\beta_u} \phi_j(x, y; \beta) \right| \le 10 j^2 \exp\left\{ \left(\frac{x^2}{4} + \frac{y^2}{4} \right) (1 + 4\delta) \right\}.$$

Hence, setting $b_n = 1 + 8n^{-1/2}\{\log d(n)\}^{1/2}$, we have on the set B_n for n sufficiently large

$$|R2_j| \le \tfrac{1}{2} n^{-1/2} \sum_{i=1}^{n} 16 \left\{4n^{-1} \log d(n)\right\} 10 j^2 \exp\{\tfrac{1}{4} b_n (X_i^2 + Y_i^2)\}.$$

This implies, using the Markov inequality,

$$\begin{aligned} &\sum_{k=2}^{d(n)} P_0 \left(\|R_2\| \ge \tfrac{1}{4} a_k^{1/2}, B_n \right) \\ &\le \sum_{k=2}^{d(n)} P_0 \left(n^{-1} \sum_{i=1}^{n} \exp\{\tfrac{1}{4} b_n (X_i^2 + Y_i^2)\} \ge \frac{\sqrt{n \log n}}{1280\sqrt{2} k^2 \log d(n)} \right) \\ &\le 640\sqrt{2} (n \log n)^{-1/2} d(n)^3 \log d(n) E_0 \exp\{\tfrac{1}{4} b_n (X_i^2 + Y_i^2)\}, \end{aligned}$$

which tends to 0 by (A.1). ▱

The next lemma shows that B_n has large probability.

LEMMA A.4. *If* $2n^{-1/2}\{logd(n)\}^{1/2} < \frac{1}{2}$, *then*

$$\sum_{k=2}^{d(n)} P_0 \left(|\hat\beta_t - \beta_{0t}| > 2n^{-1/2} \{\log d(n)\}^{1/2} \text{ for some } t \in \{1, \ldots, 4\} \right) \le \frac{10}{d(n)}.$$

PROOF. Let $\delta > (2\pi n)^{-1/2}$, then

$$P_0 (|\hat\mu| > \delta) = P_0 \left(|\bar X \sqrt{n}| > \delta\sqrt{n} \right) \le \frac{e^{-n\delta^2/2}}{\sqrt{2\pi}\delta\sqrt{n}} \le e^{-n\delta^2/2}.$$

Writing $n\hat\sigma^2$ as a sum of i.i.d. r.v.'s and using Markov's inequality and some simple algebra it can be shown that for $0 < \delta < \frac{1}{2}$ we have

$$P_0 (|\hat\sigma - 1| > \delta) \le 4e^{-n\delta^2/2}.$$

Since $(\hat\mu, \hat\sigma)$ and $(\hat\nu, \hat\tau)$ have the same distribution, the result follows. ▱

In view of (A.2) we get

$$\sum_{k=2}^{d(n)} P_0\left(\left|K_k^{1/2} - \mathcal{I}_k^{1/2}\right| \geq \tfrac{1}{2}a_k^{1/2}\right) \leq \sum_{k=2}^{d(n)} P_0\left(\|R1\| \geq \tfrac{1}{4}a_k^{1/2}, B_n\right)$$
$$+\sum_{k=2}^{d(n)} P_0\left(\|R2\| \geq \tfrac{1}{4}a_k^{1/2}, B_n\right) + \sum_{k=2}^{d(n)} P_0(B_n^c).$$

Combination of Lemma A.2, A.3 and A.4 completes the proof of (3.2).

References

BICKEL, P.J. and DOKSUM, K.A. (1977). *Mathematical Statistics, Basic Ideas and Selected Topics*. Holden-Day, Inc., Oakland, California.

BLUM, J.R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485-498.

COOK, R.D. and JOHNSON, M.E. (1986). Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. *Technometrics* **28**, 123-131.

EUBANK, R.L. and LARICCIA, V.N. (1992). Asymptotic comparison of Cramér-von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20**, 2071-2086.

EUBANK, R.L., LARICCIA, V.N. and ROSENSTEIN, R.B. (1987). Test statistics derived as components of Pearson's phi-squared distance measure. *J. Amer. Statist. Assoc.* **82**, 816-825.

INGLOT, T., KALLENBERG, W.C.M. and LEDWINA, T. (1994). Power approximations to and power comparison of certain goodness-of-fit tests. *Scand. J. Statist.* **21**, 131-145.

– – –– (1997). Data driven smooth tests for composite hypotheses. *Ann. Statist.* to appear.

INGLOT, T. and LEDWINA, T. (1996). Asymptotic optimality of data driven Neyman's tests for uniformity. *Ann. Statist.* **24**, 1982-2019.

JAVITZ, H.S. (1975). Generalized smooth tests of goodness of fit, independence, and equality of distributions. Unpublished thesis, Univ. of California, Berkeley.

JOHNSON, M.E. (1987). *Multivariate Statistical Simulation*, Wiley, New York.

KAIGH, W.D. (1992). EDF and EQF orthogonal component decompositions and tests of uniformity. *Nonparametric Statist.* **1**, 313-334.

KALLENBERG, W.C.M. and LEDWINA, T. (1995a). On data driven Neyman's tests. *Probab. Math. Statist.* **15**, 409-426.

– – –– (1995b). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.* **23**, 1594-1608.

– – –– (1995c). Data driven smooth tests for composite hypotheses: Comparison of powers. Manuscript.

– – –– (1997). Data driven smooth tests when the hypothesis is composite. *J. Amer. Statist. Assoc.* to appear.

KOZIOL, J.A. (1979). A smooth test for bivariate independence. *Sankhyā* Ser. B **41**, 260-269.

LEDWINA, T. (1994). Data-driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89**, 1000-1005.

LEHMANN, E.L. (1986). *Testing Statistical Hypotheses*, Second Edition. Wiley, New York.

MARDIA, K.V. and KENT, J.T. (1991). Rao score tests for goodness of fit and independence. *Biometrika* **78**, 355-363.

MILBRODT, H. and STRASSER, H. (1990). On the asymptotic power of the two-sided Kolmogorov-Smirnov test. *J. Statist. Plann. Inference* **26**, 1-23.

RAYNER, J.C.W. and BEST, D.J. (1989). *Smooth Tests of Goodness of Fit*, Oxford University Press, New York.

– – – – (1990). Smooth tests of goodness of fit: An overview. *Internat. Statist. Rev.* **58**, 9-17.

SANSONE, G. (1959). *Orthogonal Functions*, Interscience, New York.

SARMANOV, O.V. and BRATOEVA, Z.N. (1967). Probabilistic properties of bilinear expansions of Hermite polynomials. *Theory Probab. Appl.* **12**, 470-481.

FACULTY OF APPLIED MATHEMATICS
P.O. BOX 217
7500 AE ENSCHEDE
THE NETHERLANDS

INSTITUTE OF MATHEMATICS
TECHNICAL UNIVERSITY OF WROCŁAW
50-370 WROCŁAW
WYBRZEŻE WYSPIAŃSKIEGO 27