JYVÄSKYLÄ STUDIES IN COMPUTER SCIENCE, ECONOMICS AND STATISTICS

29

Annaliisa Kankainen

Consistent Testing of Total Independence Based on the Empirical Characteristic Function

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 1995

JYVÄSKYLÄ STUDIES IN COMPUTER SCIENCE, ECONOMICS AND STATISTICS 29

Annaliisa Kankainen

Consistent Testing of Total Independence Based on the Empirical Characteristic Function

Esitetään Jyväskylän yliopiston yhteiskuntatieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Villa Ranan Paulaharju-salissa
joulukuun 2. päivänä 1995 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Social Sciences of the University of Jyväskylä,
in the Building Villa Rana, Paulaharju Hall, on December 2, 1995 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 1995

Consistent Testing of Total Independence Based on the Empirical Characteristic Function

JYVÄSKYLÄ STUDIES IN COMPUTER SCIENCE, ECONOMICS AND STATISTICS 29

Annaliisa Kankainen

Consistent Testing of Total Independence Based on the Empirical Characteristic Function

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 1995

Editor
Esko Leskinen, PhD
Department of Statistics
University of Jyväskylä

URN:ISBN:978-951-39-8328-4
ISBN 978-951-39-8328-4 (PDF)
ISSN 0357-9921

ISBN 951-34-0617-2
ISSN 0357-9921

Jyväskylä University Printing House and
Sisäsuomi Oy, Jyväskylä 1995

ABSTRACT

Consistent testing of total independence based on the empirical characteristic function / Annaliisa Kankainen
Jyväskylä: University of Jyväskylä, 1995, 90 p.
(Jyväskylä Studies in Computer Science, Economics and Statistics
ISSN 0357-9921; 29)
ISBN 951-34-0617-2
(diss.)

There are several tests for testing independence of two variables, but a shortage of tests that can be used to test total independence of several variables. The hypothesis of total independence H_0 can be expressed in a simple manner in terms of the characteristic function, therefore the new test developed here is based on the empirical characteristic function. The new test statistic is an integral transformation of the empirical stochastic process constructed in accordance with the hypothesis of total independence. The asymptotic distributions of the test statistic under H_0 and under the alternative hypothesis H_1 are derived, when the weight function in the expression defining the test statistic satisfies certain conditions. The test is scale and location invariant and consistent. Also a nonparametric modification of the test for continuous variables is considered. The simulation study carried out here shows that the new tests, corresponding to two different weight functions, have almost as large empirical powers as the Blum–Kiefer–Rosenblatt test of independence for continuous data, where the dependence between the variables is linear. They have higher empirical powers, when the dependence is nonlinear. The new tests can also be applied for discrete data. The simulation study also shows that the weight function has some effect on the empirical powers. As an example we study the independence of variables in a data of functional capacity of retired women, and compare the results of the new tests to the model that is found by the GLIM–program, based on log–linear models. In another example we test the independence of estimated factor scores, which are derived by varimax–rotation from the Finnish ITPA data and from a data of school achievements of the pupils on 6th and 9th grade at the comprehensive schools in Jyväskylä.

Key words: Total independence, empirical characteristic function, consistency, nonparametric test, asymptotic distribution.

ACKNOWLEDGEMENTS

My supervisor Docent Antti Penttinen introduced me to the empirical generating functions and I want to thank him for his support during my work. I am also gratefull to Professor Nikolai Ushakov for the valuable discussions. My sincere thanks are also due to my colleagues at the Department of Statistics.

I want to thank Professor Alexandra Ushakova and professor Hannu Niemi who read the manuscript carefully and made valuable comments and suggestions.

Professor Esko Leskinen and Docent Antti Penttinen provided the data for the examples.

I would also like to express my gratitude to my husband Pertti Niemi for drawing the graphs and supporting me also otherwise and to my daughters Katri, Minna and Sanni for their patience and I dedicate this study to them.

The financial support of the Ellen and Artturi Nyyssönen Foundation is appreciated.

Jyväskylä, October 1995

Annaliisa Kankainen

Contents

1 INTRODUCTION

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a set of outcomes, $\mathcal{F}$ is a σ–field of subsets of Ω, and P is a probability measure on $\mathcal{F}$. A random variable $X(\omega)$ is a transformation of Ω into the real line $\mathbf{R}$ such that inverseimages $X^{-1}(B)$ of Borel sets B are elements of $\mathcal{F}$. A random vector is a d–tuple $\mathbf{X} = (X_1, \ldots, X_d)$ of random variables defined on $(\Omega, \mathcal{F})$. The probability distribution of the random vector $\mathbf{X}$ is uniquely characterized by the distribution function

$$F_X(\mathbf{x}) = \Pr(\{\omega | X_1(\omega) \leq x_1, \ldots, X_d(\omega) \leq x_d\}), \ \mathbf{x} = (x_1, \ldots, x_d) \in \mathbf{R}^d.$$

This is also known as the probability law of $\mathbf{X}$. Two random vectors $\mathbf{X}$ and $\mathbf{Y}$, defined on possibly different probability spaces, have the same law if their distribution functions are the same.

Let $F(\mathbf{x})$ be a distribution function. Its characteristic function is defined by

$$c(\mathbf{t}) = E[\exp(i < \mathbf{t}, \mathbf{X} >)] = \int_{\mathbf{R}^d} e^{i<\mathbf{t},\mathbf{X}>} dF(\mathbf{x}), \ \mathbf{t} \in \mathbf{R}^d,$$

where $< \mathbf{t}, \mathbf{X} >= \sum_{k=1}^{d} t_k X_k$ is the inner product in $\mathbf{R}^d$. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from a distribution $F(\mathbf{x})$ and the empirical distribution function

$$F_n(x_1, \ldots, x_d) = \frac{1}{n} \sum_{j=1}^{n} \prod_{k=1}^{d} I_{\{X_{jk} \leq x_k\}}, \ \mathbf{X}_j = (X_{j1}, \ldots, X_{jd}),$$

where

$$I_{\{X_{jk} \leq x_k\}} = \begin{cases} 1, & \text{if } X_{jk} \leq x_k \\ 0, & \text{if } X_{jk} > x_k. \end{cases}$$

The empirical characteristic function is defined as

$$c_n(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n} \exp(i < \mathbf{t}, \mathbf{X}_j >) = \int_{\mathbf{R}^d} e^{i<\mathbf{t},\mathbf{X}_j>} dF_n(\mathbf{x}),\ \mathbf{t} \in \mathbf{R}^d.$$

It can be considered as an additional data summary. The sequence $c_n(\mathbf{t})$ converges to $c(\mathbf{t})$ in every compact set in $\mathbf{R}^d$ and due to consistency, $c_n(\mathbf{t})$ can be used instead of the empirical distribution function, $F_n(\mathbf{x})$, in various nonparametric estimation and testing problems, especially when the model (hypothesis) has a natural representation in terms of the characteristic function.

Tests based on the empirical characteristic function are distribution free in the sense that they can be applied with only very mild assumptions concerning the tail of the distribution of the variable. But the distribution of the test statistic under H_0 is generally not independent of the distribution function $F(\mathbf{x})$, in case of tests based on nontrivial functionals of the empirical characteristic function.

One way to use the empirical characteristic function is to select a predetermined value $\mathbf{t}_0$ (e.g. Heathcote, 1972; Csörgő, 1985) or values $\mathbf{t}_1, \ldots, \mathbf{t}_m$, (e.g. Koutrouvelis, 1980, 1981, 1985) and then evaluate the test statistic at these values. Now the difficulty is, how to choose the value $\mathbf{t}_0$ or the values $\mathbf{t}_1, \ldots, \mathbf{t}_m$. Another approach is to use an integral statistic like Feuerverger and Mureika (1977), who suggest a statistic of the form

$$I = \int_{\mathbf{R}} [\mathrm{Im} c_n(t)]^2 dG(t),$$

where $G(\cdot)$ is a distribution function which is symmetric about the origin, to be used in testing for symmetry of the distribution. Epps and Pulley (1983) use the empirical characteristic function to test normality and Baringhaus and Henze (1988) extend their test to the multivariate case. This test is based on the integral statistic

$$I = n\int_{\mathbf{R}^d} |c_n(\mathbf{t}) - c_0(\mathbf{t})|^2 g(\mathbf{t}) d\mathbf{t},$$

where $c_n(\mathbf{t})$ is the empirical characteristic function of the standardized variable $\mathbf{Y}_j = \mathbf{S}_n^{-\frac{1}{2}}(\mathbf{X}_j - \bar{\mathbf{X}})$ and $c_0(\mathbf{t}) = \exp(-\frac{1}{2}|\mathbf{t}|^2)$, where $|\mathbf{t}| = (< \mathbf{t}, \mathbf{t} >)^{\frac{1}{2}}$, and $g(\mathbf{t})$ is a weight function.

The aim of this paper is to develop a class of tests for total independence of several random variables based on the integral statistic of the empirical characteristic function.

Several tests for testing independence of two random variables exist (e.g. Hoeffding, 1948b; Rosenblatt, 1975; De Wet, 1980; Feuerverger, 1993). There are fewer tests for independence of several random variables (e.g. Blum, Kiefer and Rosenblatt, 1961; Deheuvels, 1981; Csörgő, 1985). The reason for this could be that in some classical parametric models for multivariate distributions already the bivariate independence between coordinates implies total independence of all coordinates. This argument is not always satisfied, and therefore tests for verifying the independence hypothesis against arbitrary alternative are needed.

In testing for total independence Hoeffding (1948b), Blum et al. (1961), De Wet (1980) and Deheuvels (1981) suggest the use of a stochastic process which is defined through the empirical distribution function

$$P_n(\mathbf{x}) = n^{\frac{1}{2}} \left\{ F_n(\mathbf{x}) - \prod_{k=1}^{d} F_{nk}(x_k) \right\}, \tag{1.1}$$

where F_{nk} is the marginal distribution function associated with the kth component of the empirical distribution function F_n.

The test Rosenblatt (1975) suggests is parallel to the one of Hoeffding (1948b), but is based on $f_n(x_1, x_2)$, a kernel estimate of the probability density function, instead of the empirical distribution function $F_n(\mathbf{x})$. His test can also be extended to multivariate case. Rosenblatt reports that tests based on density estimates are typically less powerfull than those based on sample distribution functions.

Csörgő (1985) and Feuerverger (1993) present a similar approach based on the empirical characteristic function $c_n(\mathbf{t})$. Let $c_{nk}(t_k)$ denote the empirical marginal characteristic function corresponding to the kth margin. When total independence holds, the stochastic process

$$S_n(\mathbf{t}) = n^{\frac{1}{2}} \{c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)\} = S_n^{(1)}(\mathbf{t}) + iS_n^{(2)}(\mathbf{t}) \tag{1.2}$$

converges under certain regularity conditions in distribution to $S_F(\mathbf{t})$, where $S_F(\cdot)$ is a zero mean complex Gaussian process (see Chapter 3 for more details).

In this paper we introduce a test of total independence based on the empirical characteristic function. In fact, the test statistic is of the form

$$T_n = n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d(\mathbf{t}), \tag{1.3}$$

where $g(\mathbf{t})$ is a weight function. We show that for $g(\mathbf{t})$ satisfying certain regularity conditions, the asymptotic distribution of the test statistic under H_0 will be a weighted sum of $\chi^2(1)$ variables. The distribution can be approximated by $\gamma\chi^2(\beta)$, where γ and β are set to match the limits of expectation and variance of the test statistic T_n $(n \to \infty)$. Under H_1 the distribution of $\frac{T_n}{n}$ can be approximated by a normal distribution.

The reason for developing a new test was that it is not generally enough to test the independence of variables only in pairs and there seemed to be rather few tests for testing total independence. The test by Blum et al. (1961) is useful for continuous variables, but difficulties may appear in case of discrete variables. In Csörgő's test (Csörgő, 1985) one has to maximize the complex variance function of the $S_F(\mathbf{t})$ process (see Chapter 3 or Csörgő, 1985). However, in practice there may be difficulties in carrying out the maximization.

The test Csörgő (1985) introduces is based on the statistic

$$S_n(\mathbf{t}^{(0)}) = n^{\frac{1}{2}}\{c_n(\mathbf{t}^{(0)}) - \prod_{k=1}^{d} c_{nk}(t_k^{(0)})\},$$

where $\mathbf{t}^{(0)}$ is the value that maximizes the complex variance function of the zero mean complex Gaussian process $S_F(\mathbf{t})$ in a compact connected set $K \in \mathbf{R}^d$, see Csörgő (1985) for the details.

Csörgő's test is inconsistent in the general case. In order to show it, one has to construct an example of a characteristic function such that

$$c(\mathbf{t}^{(0)}) = \prod_{k=1}^{d} c_k(t_k^{(0)}),$$

but

$$c(\mathbf{t}) \not\equiv \prod_{k=1}^{d} c_k(t_k).$$

Let $d = 2$ and

$$c(t_1, t_2) = e^{-|t_1+t_2|},$$

then obviously

$$c(t_1, t_2) \not\equiv c(t_1, 0)c(0, t_2).$$

The complex variance function, that is maximized, is

$$\begin{aligned}
\sigma^2(t_1, t_2) &= 1 - |c(t_1,0)|^2|c(0,t_2)|^2 \\
&\quad - \left[\left(1 - |c(t_1,0)|^2\right)|c(0,t_2)|^2 + \left(1 - |c(0,t_2)|^2\right)|c(t_1,0)|^2\right] \\
&= \left(1 - |c(t_1,0)|^2\right)\left(1 - |c(0,t_2)|^2\right)
\end{aligned}$$

hence, for any compact set of the form

$$K = \{(t_1, t_2)| -a \le t_1 \le b,\ -a \le t_2 \le b,\ a, b > 0,\ a < b\},$$

$t_1^{(0)} = t_2^{(0)} = b$, and therefore

$$c(t_1^{(0)}, t_2^{(0)}) = e^{-2b} = c(t_1^{(0)}, 0)c(0, t_2^{(0)}).$$

Now the test based on the maximum variance statistic (Csörgő, 1985) would suggest accepting the independence hypothesis, although the variables are dependent.

2 CHARACTERISTIC AND EMPIRICAL CHARACTERISTIC FUNCTIONS

2.1 Characteristic functions

We first recall some properties of univariate characteristic functions and then consider some generalizations into the multivariate case. When defining the test statistic T_n in (1.3), one may choose a weight function which is positive on some interval containing origin and zero outside the interval. Therefore, we also recall sufficient conditions for a characteristic function to be uniquely determined by its values on an interval containing origin.

2.1.1 Univariate characteristic functions

The characteristic function $c(t)$ of a distribution function $F(x)$ is defined by

$$c(t) = \int_{\mathbf{R}} e^{itx} dF(x).$$

Its decomposition to real and complex parts is

$$c(t) = E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)] = U(t) + iV(t).$$

Another decomposition for $c(t)$ is

$$c(t) = b_1 c_d(t) + b_2 c_{ac}(t) + b_3 c_s(t)$$

with $b_1,\ b_2,\ b_3 \geq 0$ and $b_1 + b_2 + b_3 = 1$. Here $c_d(t),\ c_{ac}(t)$ and $c_s(t)$ are the characteristic functions of the (purely) discrete, absolutely continuous and singular parts of $F(x)$ respectively. We recall that:

1. If $b_1 = 1$, then $c(t)$ is the characteristic function of a (purely) discrete distribution. Then $c(t)$ is almost periodic (Cuppens, 1975, p. 206), and
$$\limsup_{|t|\to\infty} |c(t)| = 1.$$

2. If $b_2 = 1$, then $c(t)$ belongs to an absolutely continuous distribution and
$$\lim_{|t|\to\infty} |c(t)| = 0.$$

3. If $b_3 = 1$, then $c(t)$ is the characteristic function of a singular distribution and
$$L = \limsup_{|t|\to\infty} |c(t)|$$
may be any number between zero and one.

According to *the uniqueness theorem* for characteristic functions (Lukacs, 1970, Theorem 3.1.1) the characteristic function determines the distribution completely: Two distribution functions $F_1(x)$ and $F_2(x)$ are identical if and only if their characteristic functions $c^{(1)}(t)$ and $c^{(2)}(t)$ are identical.

The continuity theorem (Lukacs, 1970, Theorem 3.6.1) says that, if we have a sequence of distribution functions $\{F_n(x)\}$, and $\{c_n(t)\}$ is the sequence of corresponding characteristic functions, then the sequence $\{F_n(x)\}$ converges weakly to a distribution function $F(x)$ if and only if the sequence $\{c_n(t)\}$ converges for every t to a function $c(t)$ which is continuous at $t = 0$. The limiting function $c(t)$ is then the characteristic function of $F(x)$.

The continuity theorem on an interval for characteristic functions (Loéve, 1977, p. 224), says that if $c_n(t) \to c_a(t)$ when $t \in (-a, a)$ and $c_a(t)$ is continuous at $t = 0$, then $c_a(t)$ extends to a characteristic function $c(t)$, $t \in \mathbf{R}$; if the extension $c(t)$ is unique, then $c_n(t) \to c(t)$, in every $t \in \mathbf{R}$.

It is important to check the uniqueness of the extension, because it is possible to have two characteristic functions which agree on an interval $(-a, a)$ but are different outside this interval. For example, in Theorem 4.3.1 by Lukacs (1970) Pólya's conditions for continuous distributions are stated: If $c(t)$ is a real–valued and continuous function defined for all real t satisfying the following conditions

1. $c(0) = 1$
2. $c(-t) = c(t)$
3. $c(t)$ convex for $t > 0$
4. $\lim_{t\to\infty} c(t) = 0$,

then $c(t)$ is the characteristic function of an absolutely continuous distribution $F(x)$.

And Pólya's conditions for lattice distributions are in Theorem 4.3.2 by Lukacs (1970): If $c(t)$ is a real–valued function which satisfies the following conditions

1. $c(0) = 1$

2. $c(-t) = c(t)$

3. $c(t)$ is convex and continuous in the interval $(0, r)$

4. $c(t)$ is periodic with period $2r$

5. $c(r) = 0,\ c(t) \geq 0$ in $[0, r]$,

then $c(t)$ is the characteristic function of a lattice distribution.

Now if $c(t)$ is a characteristic function satisfying the Pólya's conditions for lattice distributions, for example $c(t) = 1 - |\sin(t)|$, with $r = \frac{\pi}{2}$, and

$$c^{(1)}(t) = \begin{cases} c(t), & \text{if } |t| \leq r \\ 0, & \text{if } |t| > r, \end{cases}$$

then $c^{(1)}(t)$ satisfies the Pólya's conditions for continuous distributions and is therefore a characteristic function of an absolutely continuous distribution. Now $c(t)$ and $c^{(1)}(t)$ coincide on the interval $[-r, r]$, but are different outside the interval and obviously they belong to different distributions.

Following Loéve (1977, p. 225) we denote $z = t + iy \in \mathbf{C}$. A function $h(z) = \int_{\mathbf{R}} e^{izx} dF(x)$ is regular[1] in a disc $|z| < Y$, $Y > 0$, if and only if for every positive $y < Y$, $\int_{\mathbf{R}} e^{y|x|} dF(x)$ is finite. According to Loéve (1977, p. 225), if $h(z)$ is regular in the disc $|z| < Y$ or in the rectangle $|t| < T$, $T > 0$, $|y| < Y$, $Y > 0$, then $h(z)$ is regular on a strip $|y| < Y$.

On the other hand, a characteristic function $c(t)$ is said to be analytic if there exists a function $h(z)$ of the complex variable $z = t + iy$, which is regular in the disc $|z| < Y$, $Y > 0$, and a constant $\delta > 0$ such that $h(t) = c(t)$ for $|t| < \delta$. And using the extension theorem in Loéve (1977, p. 225), it can be said that if the characteristic function $c(t)$ is analytic, then its restriction $c_a(t)$ to an interval $(-a, a)$ determines the characteristic function also outside the interval uniquely. For example Normal distribution and Poisson distribution both have analytic characteristic functions. On the other hand Cauchy distribution has a characteristic function which is not analytic. Lukacs (1970, Theorem 7.2.1), formulates the analyticity in terms of the distribution function $F(x)$: The characteristic function $c(t)$ of a distribution function $F(x)$ is analytic if and only if there exists a positive constant Y such that the relation

$$1 - F(x) + F(-x) = O(e^{-yx}) \text{ as } x \to \infty$$

holds for all positive $y < Y$, where Y may be infinite.

[1] A complex valued function is regular (holomorfic) in U if it is differentiable in each point of an open subset $U \subseteq \mathbf{C}$ (Burckel, 1979).

2.1.2 Multivariate characteristic functions

The multivariate characteristic function is defined as

$$c(\mathbf{t}) = \int_{\mathbf{R}^d} e^{i<\mathbf{t},\mathbf{x}>} dF(\mathbf{x}),\ \mathbf{t} \in \mathbf{R}^d.$$

Let $c_k(t_k) = c(0,\ldots,0,t_k,0,\ldots,0)$ denote the kth marginal characteristic function corresponding to the marginal distribution $F_k(x_k)$. Then all the results in Chapter 2.1.1 are applicable for every $c_k(t_k)$, $k = 1,\ldots,d$.

Cuppens (1975) presents some multivariate results for d–variate finite signed measures. Formulating Theorem 2.2.2 in Cuppens (1975) for distribution functions we have *the uniqueness theorem* for multivariate characteristic functions: Let F_1 and F_2 be two distribution functions, then $F_1 \equiv F_2$ if and only if $c^{(1)}(\mathbf{t}) = c^{(2)}(\mathbf{t})$ for all $\mathbf{t} \in \mathbf{R}^d$, where $c^{(1)}(\mathbf{t})$ and $c^{(2)}(\mathbf{t})$ are the characteristic functions corresponding to F_1 and F_2 respectively.

The probability measure (the distribution) corresponding to a characteristic function can also be determined. Theorem 1.6.2 by Lukacs and Laha (1964) is called *the inversion theorem in* $\mathbf{R}^d$: Let $\mathbf{X}$ be a random vector with probability measure P_X and characteristic function $c(\mathbf{t})$ and let $\mathbf{U}$ be the d–dimensional interval which contains all points $(\xi_1,\ldots,\xi_d)$ such that $a_j < \xi_j < a_j + h_j$, $j = 1,\ldots,d$. Then

$$P_X(\mathbf{U}) = \lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{-T}^{T} \cdots \int_{-T}^{T} \left[\prod_{j=1}^{d} \frac{1 - e^{-it_jh_j}}{it_j}\right] e^{-i<\mathbf{a},\mathbf{t}>} c(\mathbf{t}) d\mathbf{t},$$

provided that $\mathbf{U}$ is a continuity set of P_X.

Theorem 1.6.3 by Lukacs and Laha (1964) says that if $c(\mathbf{t})$ is absolutely integrable over $\mathbf{R}^d$, then the corresponding distribution function F is absolutely continuous and its density function is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} e^{-i<\mathbf{t},\mathbf{x}>} c(\mathbf{t}) d\mathbf{t}.$$

Theorem 1.6.4 by Lukacs and Laha (1964) states the independence of random variables in terms of characteristic functions. Accordingly, if $\mathbf{X} = (X_1,\ldots,X_d)$ is a random vector with characteristic function $c(\mathbf{t})$ and if $c_k(t_k)$ is the characteristic function of the component X_k, then the components $X_1,\ldots,X_d$ are independent if and only if the relation

$$c(\mathbf{t}) = \prod_{k=1}^{d} c_k(t_k)$$

holds for all $\mathbf{t} = (t_1,\ldots,t_d) \in \mathbf{R}^d$.

A distribution function F is convolution of two distribution functions F_1 and F_2, $F = F_1 * F_2$, if

$$F(\mathbf{y}) = \int_{\mathbf{R}^d} F_1(\mathbf{y} - \mathbf{x}) dF_2(\mathbf{x}).$$

Formulating Theorem 2.5.4 by Cuppens (1975) in terms of the distribution function F we obtain *the convolution theorem* saying that F is a convolution of F_1 and F_2 if and only if the corresponding characteristic functions satisfy the relation

$$c(\mathbf{t}) = c^{(1)}(\mathbf{t})c^{(2)}(\mathbf{t}) \text{ for all } \mathbf{t} \in \mathbf{R}^d.$$

For *the continuity theorem* for multivariate characteristic functions, see Theorems 2.6.8 and 2.6.9 by Cuppens (1975).

A d–variate characteristic function $c(\mathbf{t})$ is analytic if there exists a function $h(\mathbf{z})$ defined in $\mathbf{z} = \mathbf{t} + i\mathbf{y} \in \mathbf{C}^d$ with complex values which is regular in some neighbourhood of the origin and a positive constant δ such that

$$c(\mathbf{t}) = h(\mathbf{t}), \ |\mathbf{t}| < \delta.$$

Theorem 3.3.1 in Cuppens (1975) says that if a characteristic function $c(\mathbf{t})$ is analytic, the corresponding function $h(\mathbf{z})$ is regular, in fact, in a convex tube

$$\mathbf{R}^d + i\mathbf{Y} = \left\{\mathbf{z} \in \mathbf{C}^d | \mathrm{Im}\mathbf{z} \in \mathbf{Y}\right\},$$

where $\mathbf{Y}$ is the interior of the set $\left\{\mathbf{y} \in \mathbf{R}^d | \int e^{-<\mathbf{y},\mathbf{x}>} dF(\mathbf{x}) < \infty\right\}$. It appears that $\mathbf{Y}$ is a convex set in $\mathbf{R}^d$. Furthermore, the representation

$$h(\mathbf{z}) = \int_{\mathbf{R}^d} e^{i<\mathbf{z},\mathbf{x}>} dF(\mathbf{x}), \ \mathbf{z} \in \mathbf{R}^d + i\mathbf{Y},$$

holds. According to Theorem 3.3.2 in Cuppens (1975), if $\mathbf{Y} \subseteq \mathbf{R}^d$ is open, convex and contains the origin, then a characteristic function $c(\mathbf{t})$ is analytic in $\mathbf{R}^d + i\mathbf{Y}$ if and only if the integral $\int_{\mathbf{R}^d} e^{-<\mathbf{y},\mathbf{x}>} dF(\mathbf{x})$ converges for any $\mathbf{y} \in \mathbf{Y}$. It follows from Corollary 3.3.3 in Cuppens (1975) (by choosing $\theta = e_k$, the kth unit vector in $\mathbf{R}^d$, $k = 1, \ldots, d$) that the analyticity of $c(\mathbf{t})$ implies the analyticity of all the marginal characteristic functions $c_k(t_k)$, $k = 1, \ldots, d$.

To sum up: If two analytic characteristic functions agree in some interval $(-\mathbf{a}, \mathbf{a}) \subset \mathbf{R}^d$, then they agree on $\mathbf{R}^d$, in particular if $c(\mathbf{t}) = \prod_{k=1}^d c_k(t_k)$ in some interval $(-\mathbf{a}, \mathbf{a}) \subset \mathbf{R}^d$, then $c(\mathbf{t}) = \prod_{k=1}^d c_k(t_k)$, $\mathbf{t} \in \mathbf{R}^d$.

An example of a characteristic function, that satisfies the condition $c(\mathbf{t}) = \prod_{k=1}^d c_k(t_k)$ in some interval $[-\mathbf{a}, \mathbf{a}] \subset \mathbf{R}^d$, but $c(\mathbf{t}) \not\equiv \prod_{k=1}^d c_k(t_k)$, is

$$c(t_1, t_2) = \begin{cases} \frac{1}{2}(1 - |t_1 + 2|)(1 - |t_2 + 2|), & |t_1 + 2| \le 1, \ |t_2 + 2| \le 1, \\ (1 - |t_1|)(1 - |t_2|), & |t_1| \le 1, \ |t_2| \le 1, \\ \frac{1}{2}(1 - |t_1 - 2|)(1 - |t_2 - 2|), & |t_1 - 2| \le 1, \ |t_2 - 2| \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

It is clear that

$$c(t_1, t_2) = c(t_1, 0)c(0, t_2), \ |t_1| \le 1, |t_2| \le 1,$$

but for example

$$c(2, 2) = \frac{1}{2} \ne 0 = c(2, 0)c(0, 2).$$

The bivariate density function of the corresponding variables X_1, X_2 is

$$f(x_1, x_2) = \frac{2}{\pi^2} \frac{(1-\cos x_1)}{x_1^2} \frac{(1-\cos x_2)}{x_2^2} \cos^2(x_1 + x_2). \tag{2.1}$$

2.2 Empirical characteristic functions

We first recall the definition of the univariate empirical characteristic function and then study the convergence of the multivariate empirical characteristic functions in more detail.

2.2.1 Univariate empirical characteristic functions

Let $X_1, X_2, \ldots$ be independent random variables having a common distribution function $F(x)$ and characteristic function $c(t)$. The empirical distribution function is

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} I_{\{X_j \le x\}},$$

where

$$I_{\{X_j \le x\}} = \begin{cases} 1, & \text{if } X_j \le x \\ 0, & \text{if } X_j > x. \end{cases}$$

The empirical characteristic function is defined by

$$c_n(t) = \int_{\mathbf{R}} e^{itx} dF_n(x) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}.$$

(Cramér, 1966, p. 342 calls it the characteristic function of the sample.) For any fixed t, $c_n(t)$ is an average of bounded independent identically distributed random variables having mean $c(t)$. Therefore, as a consequence of the strong law of large numbers, $c_n(t)$ converges almost surely to $c(t)$. Feuerverger and Mureika (1977) show further that

$$\lim_{n \to \infty} \sup_{|t| \le M} |c_n(t) - c(t)| \stackrel{a.s.}{=} 0, \tag{2.2}$$

for fixed $M < \infty$.

2.2.2 Multivariate empirical characteristic functions

Let us consider the multivariate empirical characteristic functions in more detail. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of independent d–dimensional random vectors with common distribution function $F(\mathbf{x})$, $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbf{R}^d$, and characteristic function

$$c(\mathbf{t}) = \int_{\mathbf{R}^d} e^{i<\mathbf{t},\mathbf{x}>} dF(\mathbf{x}), \ \mathbf{t} = (t_1, \ldots, t_d) \in \mathbf{R}^d.$$

Let $F_n(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}^d$, be the empirical distribution function based on a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, and let

$$c_n(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^{n} e^{i<\mathbf{t},\mathbf{X}_j>} = \int_{\mathbf{R}^d} e^{i<\mathbf{t},\mathbf{x}>} dF_n(\mathbf{x}), \ \mathbf{t} = (t_1, \ldots, t_d) \in \mathbf{R}^d,$$

$$c_{nk}(t_k) = c_n(0, \ldots, 0, t_k, 0, \ldots, 0) = \frac{1}{n} \sum_{j=1}^{n} e^{it_k X_{jk}}, \ t_k \in \mathbf{R}$$

be the d–variate empirical characteristic function and, respectively, the kth empirical marginal characteristic function. Limit theorem (2.2) is also valid for multivariate case: on each bounded set $\mathbf{K} \subset \mathbf{R}^d$ we have

$$\lim_{n\to\infty} \sup_{\mathbf{t}\in\mathbf{K}} |c_n(\mathbf{t}) - c(\mathbf{t})| \stackrel{a.s.}{=} 0.$$

Csörgő and Totik (1983) have shown that, for a sequence of finite real numbers $\{M_n\}$ such that $\lim_{n\to\infty} (\frac{\log M_n}{n}) = 0$, one has

$$\lim_{n\to\infty} \sup_{|\mathbf{t}|\le M_n} |c_n(\mathbf{t}) - c(\mathbf{t})| \stackrel{a.s.}{=} 0$$

and, in addition, if $\lim_{|t_k|\to\infty} |c(\mathbf{t})| = 0$ for some k, $1 \le k \le d$, and $\limsup_{n\to\infty} (\frac{\log M_n}{n}) > 0$, then there exists a positive constant ϵ such that

$$\limsup_{n\to\infty} \Pr\{ \sup_{|\mathbf{t}|\le M_n} |c_n(\mathbf{t}) - c(\mathbf{t})| \ge \epsilon\} > 0.$$

In other words, $M_n = \exp(o(n))$ is not only the best possible rate in general for almost sure convergence, but if any faster decreasing sequence $\{M_n\}$ is taken, then even stochastic convergence can't be retained for any characteristic function vanishing at infinity along at least one path.

3 HYPOTHESIS OF TOTAL INDEPENDENCE

The total independence of the variables $X_1, \ldots, X_d$ implies the pairwise independence of these variables, but the reverse is not generally true. Therefore we now consider the hypothesis of total independence when the distributions of the variables are not known. Let $d \geq 2$ be a fixed integer and let $\mathbf{X}_j = (X_{j1}, \ldots, X_{jd})$, $1 \leq j \leq n$, be independent d-dimensional random vectors with a common unknown distribution function F. F_k is the common univariate (marginal) distribution function of all the kth components $X_{1k}, \ldots, X_{nk}$, $1 \leq k \leq d$. We are concerned with testing the hypothesis

$$H_0 : F(x_1, \ldots, x_d) = \prod_{k=1}^{d} F_k(x_k), \ (x_1, \ldots, x_d) \in \mathbf{R}^d,$$

of total independence. Rosenblatt (1975) writes this hypothesis in terms of probability density function

$$H_0 : f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k), \ (x_1, \ldots, x_d) \in \mathbf{R}^d,$$

and it can also be written as a similar decomposition of the characteristic function. Let $c(\mathbf{t})$, $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbf{R}^d$, be the characteristic function corresponding to F and $c_k(t_k) = c(0, \ldots, 0, t_k, 0, \ldots, 0)$ be the kth marginal characteristic function corresponding to F_k, $1 \leq k \leq d$. Then the null hypothesis is equivalent to

$$H_0 : \ c(t_1, \ldots, t_d) = \prod_{k=1}^{d} c_k(t_k), \ (t_1, \ldots, t_d) \in \mathbf{R}^d.$$

We can make use of the process

$$S_n(\mathbf{t}) = n^{\frac{1}{2}}\{c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)\} = S_n^{(1)}(\mathbf{t}) + iS_n^{(2)}(\mathbf{t})$$

in testing the hypothesis of total independence. It is parallel to the $P_n(\cdot)$ process in (1.1).

To be able to determine the asymptotic behaviour of the process $S_n(\mathbf{t})$, we need to study the behaviour of the d–variate empirical characteristic process

$$Y_n(\mathbf{t}) = n^{\frac{1}{2}}\{c_n(\mathbf{t}) - c(\mathbf{t})\} = \int_{\mathbf{R}^d} e^{i<\mathbf{t},\mathbf{x}>} d\mathbf{B}_n(\mathbf{x}),$$

where

$$\mathbf{B}_n(\mathbf{x}) = n^{\frac{1}{2}}\{F_n(\mathbf{x}) - F(\mathbf{x})\}$$

is the d–variate empirical process, where F_n stands for the empirical distribution function corresponding to F.

Let us consider a necessary and sufficient condition for the weak convergence of the $Y_n(\mathbf{t})$ process on any compact set $\mathbf{K}$ in $\mathbf{R}^d$. Let $\mathcal{C}(\mathbf{K})$ be the Banach space of continuous complex valued functions on $\mathbf{K}$ with the usual sup–norm. $Y_n(\cdot)$ restricted to $\mathbf{K}$ is a random element taking values in $\mathcal{C}(\mathbf{K})$ for each n. Write $c(\mathbf{t}) = U(\mathbf{t}) + iV(\mathbf{t})$, and consider a zero mean complex valued d–variate Gaussian random field $Y_F(\mathbf{t}) = R(\mathbf{t}) + iI(\mathbf{t})$, $\mathbf{t} = (t_1, \ldots, t_d)$ having the same covariance matrix as $Y_n(\mathbf{t})$ for each n. Then

$$\begin{aligned}
\sigma^{11}(\mathbf{s},\mathbf{t}) &= E[R(\mathbf{s})R(\mathbf{t})] = \tfrac{1}{2}[U(\mathbf{s}-\mathbf{t}) + U(\mathbf{s}+\mathbf{t})] - U(\mathbf{s})U(\mathbf{t}) \\
\sigma^{12}(\mathbf{s},\mathbf{t}) &= E[R(\mathbf{s})I(\mathbf{t})] = \tfrac{1}{2}[-V(\mathbf{s}-\mathbf{t}) + V(\mathbf{s}+\mathbf{t})] - U(\mathbf{s})V(\mathbf{t}) \\
\sigma^{21}(\mathbf{s},\mathbf{t}) &= E[I(\mathbf{s})R(\mathbf{t})] = \tfrac{1}{2}[V(\mathbf{s}-\mathbf{t}) + V(\mathbf{s}+\mathbf{t})] - V(\mathbf{s})U(\mathbf{t}) \\
\sigma^{22}(\mathbf{s},\mathbf{t}) &= E[I(\mathbf{s})I(\mathbf{t})] = \tfrac{1}{2}[U(\mathbf{s}-\mathbf{t}) - U(\mathbf{s}+\mathbf{t})] - V(\mathbf{s})V(\mathbf{t}).
\end{aligned}$$

We can also write

$$Y_F(\mathbf{t}) = \int_{\mathbf{R}^\mathrm{d}} e^{i<\mathbf{t},\mathbf{x}>} d\mathbf{B}_F(\mathbf{x}), \tag{3.1}$$

where $\mathbf{B}_F(\mathbf{x})$ is a d–variate Brownian bridge process associated with the distribution function F, see Shorack and Wellner (1986, p. 30).

Just like in the univariate case (Feuerverger and Mureika, 1977; Csörgő, 1981a; Marcus, 1981), the finite dimensional distributions of $Y_n(\cdot)$ converge to those of $Y_F(\cdot)$ as a consequence of the multidimensional central limit theorem. But $Y_n(\cdot)$ does not always converge weakly in $\mathcal{C}(\mathbf{K})$ to $Y_F(\cdot)$, since the latter process can be almost surely discontinuous for certain distributions F.

Csörgő (1981b) proves that $Y_n(\cdot)$ converges weakly to $Y_F(\cdot)$ in $\mathcal{C}(\mathbf{K})$ if and only if

$$\int_0^1 \frac{\overline{\psi}(h)}{h(\log\frac{1}{h})^{\frac{1}{2}}} dh < \infty, \tag{3.2}$$

where

$$\overline{\psi}(h) = \sup\{y|0 \le y \le 1, \lambda_d\{\mathbf{t}| \sup_{\mathbf{t}\in\mathbf{K}} |\mathbf{t}| < \tfrac{1}{2}, \psi(\mathbf{t}) < y\} < h\},$$

with λ_d standing for the d–dimensional Lebesgue measure, is the nondecreasing rearrangement of the function $\psi(\mathbf{t}) = (1 - U(\mathbf{t}))^{\frac{1}{2}}$. This condition is a mild tail condition on the underlying distribution and is satisfied if

$$\int_{\mathbf{R}^d} (\log^+(|\mathbf{x}|))^{1+\epsilon} dF(\mathbf{x}) < \infty \tag{3.3}$$

holds for arbitrary small $\epsilon > 0$. The condition (3.2) is generally not satisfied if the integral in (3.3) is finite only with $\epsilon = 0$.

We return to the hypothesis of total independence and the related empirical process $S_n(\mathbf{t})$ in (1.2). Let us also consider the zero mean complex Gaussian process

$$S_F(\mathbf{t}) = Y_F(\mathbf{t}) - \sum_{k=1}^{d} Y_F(0,\ldots,0,t_k,0,\ldots,0) \prod_{\substack{m=1\\ m\neq k}}^{d} c_m(t_m),$$

where $Y_F(\mathbf{t})$ is in (3.1). We have $E[Y_F(\mathbf{s})\overline{Y_F(\mathbf{t})}] = c(\mathbf{s}-\mathbf{t}) - c(\mathbf{s})c(-\mathbf{t})$ and therefore under H_0 the complex variance function of $S_F(\mathbf{t})$ is

$$\begin{aligned}\sigma^2(\mathbf{t}) = E|S_F(\mathbf{t})|^2 &= E[S_F^{(1)}(\mathbf{t})^2] + E[S_F^{(2)}(\mathbf{t})^2] \\ &= 1 - \prod_{k=1}^{d} |c_k(t_k)|^2 - \sum_{k=1}^{d}\{1 - |c_k(t_k)|^2\} \prod_{\substack{m=1\\ m\neq k}}^{d} |c_m(t_m)|^2.\end{aligned}$$

It can be estimated by

$$\sigma_n^2(\mathbf{t}) = 1 - \prod_{k=1}^{d} |c_{nk}(t_k)|^2 - \sum_{k=1}^{d}\{1 - |c_{nk}(t_k)|^2\} \prod_{\substack{m=1\\ m\neq k}}^{d} |c_{nm}(t_m)|^2. \tag{3.4}$$

Since $c_n(\cdot)$ is a strongly uniformly consistent estimator of $c(\cdot)$ one has

$$\sup_{\mathbf{t}\in\mathbf{K}} |\sigma_n^2(\mathbf{t}) - \sigma^2(\mathbf{t})| \overset{a.s.}{\to} 0.$$

Csörgő (1985) proves that under H_0 $S_n(\cdot)$ converges weakly in $\mathcal{C}(\mathbf{K})$ to $S_F(\cdot)$ if and only if the condition (3.2) holds.

4 THE TEST STATISTIC AND ITS LARGE SAMPLE MOMENTS

Let $\mathbf{X}_j = (X_{j1}, \ldots, X_{jd})$, $1 \leq j \leq n$, be i.i.d. d–dimensional ($d \geq 2$) random vectors with a common unknown distribution function F and let F_k be the marginal distribution function of the kth components. If $c(\mathbf{t})$, $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbf{R}^d$, is the characteristic function corresponding to F and $c_k(t_k) = c(0, \ldots, 0, t_k, 0, \ldots, 0)$ is the kth marginal characteristic function corresponding to F_k, $1 \leq k \leq d$, then the hypothesis of total independence can be expressed as

$$H_0: \; c(t_1, \ldots, t_d) = \prod_{k=1}^{d} c_k(t_k), \;\; (t_1, \ldots, t_d) \in \mathbf{R}^d.$$

Let $S_n(\cdot)$ be the empirical process in (1.2). We propose a test statistic of total independence of the form

$$\begin{aligned} T_n &= \int_{\mathbf{R}^d} |S_n(\mathbf{t})|^2 g(\mathbf{t}) d\mathbf{t} \\ &= n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d\mathbf{t}, \end{aligned} \tag{4.1}$$

where $g(\mathbf{t})$ is a weight function.

The choice of the weight function $g(\mathbf{t})$ needs an understanding of the behaviour of the characteristic and the empirical characteristic function. The following reasoning shows, that in some neighbourhood of the origin, the empirical characteristic function is a precise estimator of the characteristic function. Every characteristic function $c(\mathbf{t})$ is continuous and $c(\mathbf{0}) = 1$, therefore there exists a neighbourhood of origin in which $c(\mathbf{t})$ and every $c_k(t_k)$ are different from zero. It can also be shown that

$$E[|c_{nk}(t_k) - c_k(t_k)|^2] = \frac{1}{n}\left(1 - |c_k(t_k)|^2\right). \tag{4.2}$$

For any arbitrary characteristic function the expectation in (4.2) is small in some neighbourhood of the origin and converges to zero when $t_k \to 0$. On the other hand it is known that the tail behaviour of a distribution is reflected by the behaviour of its characteristic function at the origin.

As a consequence, the weight function should assign high weight in every marginal on an interval around the origin. The length of the interval for each marginal will depend inversely on the scale of the data corresponding to that marginal. Therefore the weight function should satisfy the conditions

$$\begin{cases} g1_k(t_k) > 0, & t_k \in (-a_k, a_k) \\ g1_k(t_k) = 0, & t_k \notin (-a_k, a_k), \text{ for all } k = 1, \ldots, d \end{cases} \tag{4.3}$$

or

$$\begin{cases} g2_k(t_k) > 0, & t_k \in \mathbf{R} \\ g2_k(t_k) \to 0, & \text{when } |t_k| \to \infty, \text{ for all } k = 1, \ldots, d. \end{cases} \tag{4.4}$$

If it is known that the characteristic function corresponding to the data is analytic, the weight function in (4.3) can be used, otherwise it is better to use (4.4). This is a consequence of the results concerning uniqueness of an extension of a characteristic function from an interval $(-\mathbf{a}, \mathbf{a})$ to $\mathbf{R}^d$, see Chapters 2.1.1 and 2.1.2. For example if observations are simulated from the distribution defined in (2.1), the empirical powers corresponding to the test statistic

$$T1_n = n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g1(\mathbf{t}) d\mathbf{t},$$

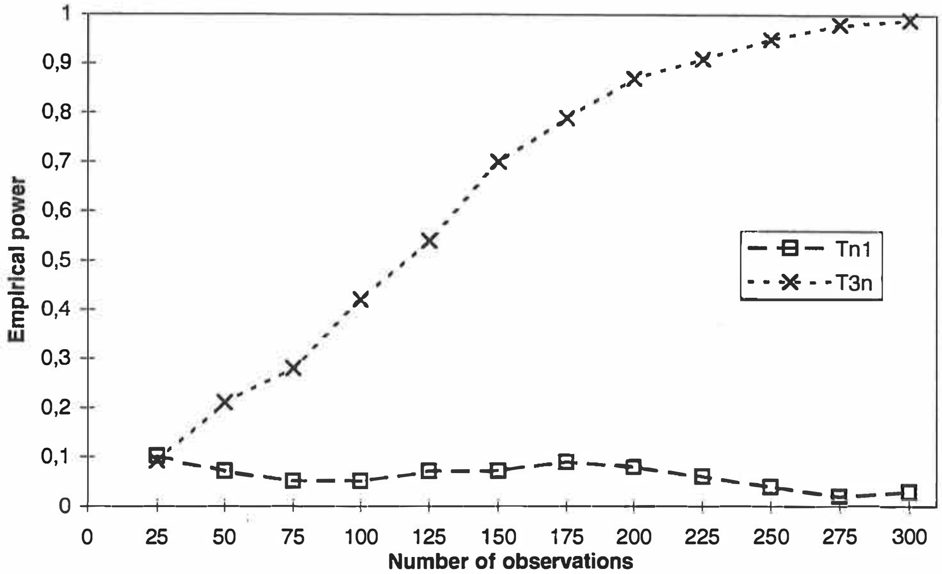
where

$$g1(\mathbf{t}) = \begin{cases} \frac{1}{2^d}, & t_k \in (-1, 1),\ k = 1, \ldots, d \\ 0, & \text{elsewhere} \end{cases}$$

don't converge to one as $n \to \infty$. But as Picture 4.1 shows, the empirical powers corresponding to the test statistic

$$T3_n = n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g3(\mathbf{t}) d\mathbf{t},$$

where $g3(\mathbf{t}) = \prod_{k=1}^{d} \frac{1}{\sqrt{20\pi}} e^{-\frac{1}{20} t_k^2}$, $\mathbf{t} \in \mathbf{R}^d$ converge to one (n >300).



PICTURE 4.1 The empirical powers of the tests based on the statistics $T1_n$ and $T3_n$ corresponding to 100 simulated data from the distribution defined in (2.1).

Practical considerations for the weight function are that

1. $g(\mathbf{t})$ is real valued
2. it provides a closed form for the integral (4.1)
3. we may write $g(\mathbf{t}) = \prod_{k=1}^{d} g_k(t_k)$
4. $0 < \int_R g_k(t_k)dt_k < \infty$.

   In the derivation of the limiting distribution of the test statistic under H_0 we need additionally

5. $g_k(t_k)$, $k = 1, \ldots, d$, to be even functions.

The data is studentized (observed values of each component of the vector valued variable are divided by the observed standard deviation of that component) to obtain a test statistic which is scale and location invariant.

For computation of the test statistic it is covenient to write

$$T_n = n \left[\frac{1}{n^2} \sum_{j=1}^{n} \sum_{m=1}^{n} \prod_{k=1}^{d} \int_{\mathbf{R}} e^{it_k\left(\frac{X_{jk}-X_{mk}}{S_k}\right)} g_k(t_k)dt_k \right.$$
$$- \frac{2}{n^{d+1}} \sum_{j=1}^{n} \prod_{k=1}^{d} \sum_{m=1}^{n} \int_{\mathbf{R}} e^{it_k\left(\frac{X_{jk}-X_{mk}}{S_k}\right)} g_k(t_k)dt_k$$

$$+\frac{1}{n^{2d}}\prod_{k=1}^{d}\sum_{j=1}^{n}\sum_{m=1}^{n}\int_{\mathbf{R}} e^{it_k\left(\frac{X_{jk}-X_{mk}}{S_k}\right)} g_k(t_k)dt_k\Bigg].$$

Because $g_k(t_k)$, $k = 1, \ldots, d$, are assumed to be even functions, further calculations show that

$$\begin{aligned} T_n &= n\Bigg[\frac{1}{n^2}\sum_{j=1}^{n}\sum_{m=1}^{n}\prod_{k=1}^{d}\int_{\mathbf{R}}\cos\left(t_k\left(\tfrac{X_{jk}-X_{mk}}{S_k}\right)\right)g_k(t_k)dt_k \\ &\quad -\frac{2}{n^{d+1}}\sum_{j=1}^{n}\prod_{k=1}^{d}\sum_{m=1}^{n}\int_{\mathbf{R}}\cos\left(t_k\left(\tfrac{X_{jk}-X_{mk}}{S_k}\right)\right)g_k(t_k)dt_k \\ &\quad +\frac{1}{n^{2d}}\prod_{k=1}^{d}\sum_{j=1}^{n}\sum_{m=1}^{n}\int_{\mathbf{R}}\cos\left(t_k\left(\tfrac{X_{jk}-X_{mk}}{S_k}\right)\right)g_k(t_k)dt_k\Bigg]. \end{aligned}$$

Next the first two large sample moments of T_n are derived. Because $\int_{\mathbf{R}} g(t_k)dt_k < \infty$, the integrals in the formula of the test statistic are uniformly and absolutely convergent and therefore the order of integration can be changed when calculating the expectations. Thus, under H_0 when $n \to \infty$ we have

$$\begin{aligned} E[T_n] &= \int_{\mathbf{R}^d} E|S_n(\mathbf{t})|^2 g(\mathbf{t})d\mathbf{t} \\ &\to \int_{\mathbf{R}^d} g(\mathbf{t})d\mathbf{t} - \prod_{k=1}^{d} E1_k - \sum_{k=1}^{d}\left(\int_{\mathbf{R}} g_k(t_k)dt_k - E1_k\right)\prod_{\substack{m=1\\ m\neq k}}^{d} E1_m \\ &= \int_{\mathbf{R}^d}\sigma^2(\mathbf{t})g(\mathbf{t})d\mathbf{t}, \end{aligned}$$

where

$$E1_k = \int_{\mathbf{R}} |c_k(t_k)|^2 g_k(t_k)dt_k.$$

These integrals can be estimated from the data by replacing $c_k(\cdot)$s by their empirical counterparts, hence

$$\widehat{E1_k} = \int_{\mathbf{R}} |c_{nk}(t_k)|^2 g_k(t_k)dt_k.$$

The second moment and further the variance of the test statistic under H_0 when $n \to \infty$ are derived using the expectation

$$\begin{aligned} E\left[Y_F(\mathbf{s})\overline{Y_F(\mathbf{u})}Y_F(\mathbf{t})\overline{Y_F(\mathbf{v})}\right] &= c(\mathbf{s}-\mathbf{u})c(\mathbf{t}-\mathbf{v}) + c(\mathbf{s}+\mathbf{t})c(-\mathbf{u}-\mathbf{v}) \\ &+c(\mathbf{s}-\mathbf{v})c(\mathbf{t}-\mathbf{u}) - c(\mathbf{s}-\mathbf{u})c(\mathbf{t})c(-\mathbf{v}) - c(\mathbf{s}+\mathbf{t})c(-\mathbf{u})c(-\mathbf{v}) \\ &-c(\mathbf{s}-\mathbf{v})c(\mathbf{t})c(-\mathbf{u}) - c(\mathbf{s})c(-\mathbf{u}+\mathbf{t})c(-\mathbf{v}) - c(\mathbf{s})c(-\mathbf{u}-\mathbf{v})c(\mathbf{t}) \\ &-c(\mathbf{s})c(-\mathbf{u})c(\mathbf{t}-\mathbf{v}) + 3c(\mathbf{s})c(-\mathbf{u})c(\mathbf{t})c(-\mathbf{v}), \end{aligned}$$

and the result is

$$\begin{aligned}\mathrm{Var}[T_n] &= \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} E\left[|S_n(\mathbf{t})|^2|S_n(\mathbf{s})|^2\right] g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s} - (E[T_n])^2 \\ &\to 2\prod_{k=1}^{d} E1_k^2 - 4\prod_{k=1}^{d} E2_k + 2\prod_{k=1}^{d} E3_k \\ &\quad +2\sum_{k=1}^{d}(E3_k - E1_k^2)\prod_{\substack{m=1\\m\neq k}}^{d} E1_m^2 \\ &\quad -4\sum_{k=1}^{d}(E3_k - E2_k)\prod_{\substack{m=1\\m\neq k}}^{d} E2_m \\ &\quad +2\sum_{k=1}^{d}\sum_{\substack{l=1\\l\neq k}}^{d}(E1_k^2E1_l^2 - 2E2_kE1_l^2 + E2_kE2_l)\prod_{\substack{m=1\\m\neq k,l}}^{d} E1_m^2.\end{aligned}$$

Here $E1_k$ is as before,

$$\begin{aligned}E2_k &= \int_{\mathbf{R}}\int_{\mathbf{R}} c_k(t_k+s_k)c_k(-t_k)c_k(-s_k)g_k(t_k)g_k(s_k)dt_kds_k \\ &= \int_{\mathbf{R}}\int_{\mathbf{R}} c_k(-t_k+s_k)c_k(t_k)c_k(-s_k)g_k(t_k)g_k(s_k)dt_kds_k \\ &= \int_{\mathbf{R}}\int_{\mathbf{R}} c_k(t_k-s_k)c_k(-t_k)c_k(s_k)g_k(t_k)g_k(s_k)dt_kds_k \\ &= \int_{\mathbf{R}}\int_{\mathbf{R}} c_k(-t_k-s_k)c_k(t_k)c_k(s_k)g_k(t_k)g_k(s_k)dt_kds_k\end{aligned}$$

and

$$\begin{aligned}E3_k &= \int_{\mathbf{R}}\int_{\mathbf{R}} |c_k(t_k+s_k)|^2 g_k(t_k)g_k(s_k)dt_kds_k \\ &= \int_{\mathbf{R}}\int_{\mathbf{R}} |c_k(t_k-s_k)|^2 g_k(t_k)g_k(s_k)dt_kds_k.\end{aligned}$$

The estimators are

$$\widehat{E2_k} = \int_{\mathbf{R}}\int_{\mathbf{R}} c_{nk}(t_k+s_k)c_{nk}(-t_k)c_{nk}(-s_k)g_k(t_k)g_k(s_k)dt_kds_k$$

and

$$\widehat{E3_k} = \int_{\mathbf{R}}\int_{\mathbf{R}} |c_{nk}(t_k+s_k)|^2 g_k(t_k)g_k(s_k)dt_kds_k.$$

The estimates of the limits of expectation and variance of T_n are needed in approximating the asymptotic distribution of the test statistic under H_0 in Chapter 5.2.

5 SOME LARGE SAMPLE PROPERTIES OF THE TEST STATISTIC

The test defined by the statistic T_n is distribution free in the sense that all one has to assume about the distribution of the d–variate vector of variables is that its characteristic function satisfies the condition (3.2) and if the weight function $g1(\mathrm{t})$ in (4.3) is used, then the characteristic function should be analytic. Furthermore, the use of studentized data makes the test both location and scale invariant. In this chapter we derive the asymptotic distribution of the test statistic under H_1 and H_0, so that we are able to determine the size and P-values and, on the other hand, the power of the test. We also show that the test is consistent and that large values of the test statistic suggest rejection of the hypothesis of total independence. A test statistic for total independence in terms of kernel estimators of its density and marginal density functions is also constructed, in a special case. It appears that this test is equivalent to the one defined by T_n provided the weight function is properly chosen. Also a nonparametric modification of the test in case of continuous variables is considered.

5.1 The asymptotic distribution of the test statistic under H_1

In deriving the asymptotic distribution of T_n we can make use of the well–known asymptotic results on U–statistics (Hoeffding, 1948a). A representation of T_n in terms of three U–statistics is of the form

$$T_n = n\left[U_1 + \frac{B1_n}{\sqrt{n}} - 2U_2 - 2\frac{B2_n}{\sqrt{n}} + U_3 + \frac{B3_n}{\sqrt{n}}\right],$$

where $B1_n$, $B2_n$, $B3_n$ are random variables with $\lim_{n\to\infty} E[Bi_n^2] = 0$, $i = 1, 2, 3$. The U–statistics are

$$\begin{aligned} U_1 &= \frac{1}{\binom{n}{2}} \sum_{j=1}^{n} \sum_{m>j}^{n} \int_{\mathbf{R}^d} \left[\frac{1}{2} \exp\left\{ i \sum_{k=1}^{d} t_k \frac{X_{jk}}{S_k} \right\} \exp\left\{ -i \sum_{k=1}^{d} t_k \frac{X_{mk}}{S_k} \right\} \right. \\ &\quad \left. + \frac{1}{2} \exp\left\{ i \sum_{k=1}^{d} t_k \frac{X_{jk}}{S_k} \right\} \exp\left\{ -i \sum_{k=1}^{d} t_k \frac{X_{mk}}{S_k} \right\} \right] g(\mathbf{t}) d\mathbf{t} \\ &= \frac{1}{\binom{n}{2}} \sum_{j=1}^{n} \sum_{m>j}^{n} \int_{\mathbf{R}^d} \phi_1(X_j, X_m) g(\mathbf{t}) d\mathbf{t}, \end{aligned}$$

$$\begin{aligned} U_2 &= \frac{1}{\binom{n}{d+1}} \sum_{j=1}^{n} \sum_{m_1>j}^{n} \cdots \sum_{m_d>m_{d-1}}^{n} \\ &\quad \int_{\mathbf{R}^d} \left[\frac{1}{(d+1)!} \sum_{p_1} \exp\left\{ i \sum_{k=1}^{d} t_k \frac{X_{jk}}{S_k} \right\} \exp\left\{ -i \sum_{k=1}^{d} t_k \frac{X_{p_1 k}}{S_k} \right\} \right. \\ &\quad \left. + \frac{1}{(d+1)!} \sum_{p_2} \sum_{l=1}^{d} \exp\left\{ i \sum_{k=1}^{d} t_k \frac{X_{m_l k}}{S_k} \right\} \exp\left\{ -i \sum_{k=1}^{d} t_k \frac{X_{p_2 k}}{S_k} \right\} \right] g(\mathbf{t}) d\mathbf{t} \\ &= \frac{1}{\binom{n}{d+1}} \sum_{j=1}^{n} \sum_{m_1>j}^{n} \cdots \sum_{m_d>m_{d-1}}^{n} \int_{\mathbf{R}^d} \phi_2(X_j, X_{m_1}, \ldots, X_{m_d}) g(\mathbf{t}) d\mathbf{t}, \end{aligned}$$

where $\sum_{p_1}$ stands for summation over all permutations of $\{m_1, \ldots, m_d\}$ and $\sum_{p_2}$ summation over all permutations of $\{j, m_1, \ldots, m_d\} \setminus \{m_l\}$,

$$\begin{aligned} U_3 &= \frac{1}{\binom{n}{2d}} \sum_{j_1=1}^{n} \cdots \sum_{j_d>j_{d-1}}^{n} \sum_{m_1>j_d}^{n} \cdots \sum_{m_d>m_{d-1}}^{n} \\ &\quad \int_{\mathbf{R}^d} \left[\frac{1}{(2d)!} \sum_{p_3} \exp\left\{ i \sum_{k=1}^{d} t_k \frac{X_{p_3 k}}{S_k} \right\} \exp\left\{ -i \sum_{k=1}^{d} t_k \frac{X_{p_3 k}}{S_k} \right\} \right] g(\mathbf{t}) d\mathbf{t} \\ &= \frac{1}{\binom{n}{2d}} \sum_{j_1=1}^{n} \cdots \sum_{j_d>j_{d-1}}^{n} \sum_{m_1>j_d}^{n} \cdots \sum_{m_d>m_{d-1}}^{n} \\ &\quad \int_{\mathbf{R}^d} \phi_3(X_{j_1}, \ldots, X_{j_d}, X_{m_1}, \ldots, X_{m_d}) g(\mathbf{t}) d\mathbf{t}, \end{aligned}$$

and $\sum_{p_3}$ is summation over all permutations of $\{j_1, \ldots, j_d, m_1, \ldots, m_d\}$.

Let

$$\theta_1 = E[\phi_1(X_j, X_m)] = E[U_1] = \int_{\mathbf{R}^d} c(\mathbf{t}) c(-\mathbf{t}) g(\mathbf{t}) d\mathbf{t},$$

$$\theta_2 = E[\phi_2(X_j, X_{m_1}, \ldots, X_{m_d})] = E[U_2] = \int_{\mathbf{R}^d} c(\mathbf{t}) \prod_{k=1}^{d} c_k(-t_k) g(\mathbf{t}) d\mathbf{t},$$

and

$$\begin{aligned}\theta_3 &= E[\phi_3(X_{j_1},\ldots,X_{j_d},X_{m_1},\ldots,X_{m_d})] = E[U_3] \\ &= \int_{\mathbf{R}^d}\prod_{k=1}^{d} c_k(t_k)c_k(-t_k)g(\mathbf{t})d\mathbf{t}.\end{aligned}$$

If $U_1' = U_1 + \frac{B1_n}{\sqrt{n}}$, $U_2' = U_2 + \frac{B2_n}{\sqrt{n}}$ and $U_3' = U_3 + \frac{B3_n}{\sqrt{n}}$, then all the conditions of Theorem 7.3 in Hoeffding (1948a) hold. Therefore, when $n \to \infty$, the joint distribution of $\sqrt{n}(U_1'-\theta_1), \sqrt{n}(U_2'-\theta_2), \sqrt{n}(U_3'-\theta_3)$ tends to normal distribution with zero mean and covariance matrix

$$\sigma_{1(kl)} = m(k)m(l)\zeta_1^{kl},\ k,l = 1,2,3,$$

where $m(1) = 2$, $m(2) = d+1$, $m(3) = 2d$. To recall the definition of ζ_1^{kl} we first introduce

$$\psi^{(l)}(x_1,\ldots,x_{m(l)}) = \phi_l(x_1,\ldots,x_{m(l)}) - \theta_l,$$

and

$$\psi_c^{(l)}(x_1,\ldots,x_c) = E\{\psi^{(l)}(x_1,\ldots,x_c,X_{c+1},\ldots,X_{m(l)}\},\ c = 1,\ldots,m(l),$$

$l = 1,2,3$. Then ζ_c^{kl}s are defined as

$$\zeta_c^{kl} = E\{\psi_c^{(k)}(X_1,\ldots,X_c)\psi_c^{(l)}(X_1,\ldots,X_c)\},\ k,l = 1,2,3.$$

For more details see Hoeffding (1948a). Now, if $\mathbf{U}' = (U_1',U_2',U_3')$, $\theta = (\theta_1,\theta_2,\theta_3)$ and $f(\mathbf{U}') = U_1' - 2U_2' + U_3'$, then according to the corollary in Serfling (1980, p. 124)

$$\sqrt{n}\,(f(\mathbf{U}') - f(\theta)) \xrightarrow{\mathcal{D}} N(0,\sigma_1^2),$$

as $n \to \infty$, where

$$\sigma_1^2 = \sum_{k=1}^{3}\sum_{l=1}^{3}\sigma_{1(kl)}\left.\frac{\partial f}{\partial u_k'}\right|_{\mathbf{u}'=\theta}\left.\frac{\partial f}{\partial u_l'}\right|_{\mathbf{u}'=\theta}, \tag{5.1}$$

and

$$\begin{aligned}f(\theta) &= \theta_1 - 2\theta_2 + \theta_3 \\ &= \int_{\mathbf{R}^d}|c(\mathbf{t}) - \prod_{k=1}^{d} c_k(t_k)|^2 g(\mathbf{t})d\mathbf{t}.\end{aligned}$$

To sum up we have the next result.

Theorem 5.1 *Under H_1,*

$$\frac{T_n}{n} \xrightarrow{\mathcal{D}} N(f(\theta),\frac{\sigma_1^2}{n}),\ n \to \infty.$$

The variance (5.1) can be calculated as follows
(here $\mathbf{t}_l = (0,\ldots,0,t_l,0,\ldots,0)$):

$$\sigma_1^2 = 4\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}+\mathbf{s})c(-\mathbf{t})c(-\mathbf{s})g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-8\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}+\mathbf{s})c(-\mathbf{t})\prod_{k=1}^{d} c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$+4\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}+\mathbf{s})\prod_{k=1}^{d} c(-\mathbf{t}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-8\sum_{l=1}^{d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}-\mathbf{t}_l)c(-\mathbf{t})c(\mathbf{s})c(\mathbf{t}_l-\mathbf{s}_l)\prod_{\substack{k=1\\k\neq l}}^{d} c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$+8\sum_{l=1}^{d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}-\mathbf{t}_l)c(\mathbf{s})c(\mathbf{t}_l-\mathbf{s}_l)c(-\mathbf{t}_l)\cdot$$

$$\cdot\prod_{\substack{k=1\\k\neq l}}^{d} c(-\mathbf{t}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$+8\sum_{l=1}^{d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}-\mathbf{t}_l)c(-\mathbf{t})c(\mathbf{t}_l+\mathbf{s}_l)c(-\mathbf{s}_l)\cdot$$

$$\cdot\prod_{\substack{k=1\\k\neq l}}^{d} c(\mathbf{s}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-8\sum_{l=1}^{d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}-\mathbf{t}_l)c(\mathbf{t}_l+\mathbf{s}_l)c(-\mathbf{t}_l)c(-\mathbf{s}_l)\cdot$$

$$\cdot\prod_{\substack{k=1\\k\neq l}}^{d} c(-\mathbf{t}_k)c(\mathbf{s}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$+4\sum_{l=1}^{d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t})c(\mathbf{s})c(-\mathbf{t}_l-\mathbf{s}_l)\prod_{\substack{k=1\\k\neq l}}^{d} c(-\mathbf{t}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-8\sum_{l=1}^{d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t})c(-\mathbf{t}_l+\mathbf{s}_l)c(-\mathbf{s}_l)\prod_{\substack{k=1\\k\neq l}}^{d} c(-\mathbf{t}_k)c(\mathbf{s}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$+4\sum_{l=1}^{d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t}_l+\mathbf{s}_l)c(-\mathbf{t}_l)c(-\mathbf{s}_l)\cdot$$

$$\cdot\prod_{\substack{k=1\\k\neq l}}^{d} c(\mathbf{t}_k)c(-\mathbf{t}_k)c(\mathbf{s}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-4\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t})c(-\mathbf{t})c(\mathbf{s})c(-\mathbf{s})g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$+8(d+1)\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t})c(-\mathbf{t})c(\mathbf{s})\prod_{k=1}^{d} c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-(12d+4)\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t})c(\mathbf{s})\prod_{k=1}^{d} c(-\mathbf{t}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-8d\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t})c(-\mathbf{t})\prod_{k=1}^{d} c(\mathbf{s}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$+16d\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} c(\mathbf{t})\prod_{k=1}^{d} c(-\mathbf{t}_k)c(\mathbf{s}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}$$

$$-4d\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} \prod_{k=1}^{d} c(\mathbf{t}_k)c(-\mathbf{t}_k)c(\mathbf{s}_k)c(-\mathbf{s}_k)g(\mathbf{t})g(\mathbf{s})d\mathbf{t}d\mathbf{s}.$$

It is obvious that $f(\mathbf{U}') = \frac{T_n}{n}$ is not an unbiased estimator of $f(\theta)$, however,

$$\lim_{n\to\infty}\sqrt{n}\left(E\left[\frac{T_n}{n}\right] - f(\theta)\right) = 0.$$

Under H_0 the expectation $f(\theta)$ and the variance σ_1^2 are both zero and the distribution is degenerated at zero. Therefore other methods are needed to find the asymptotic distribution of T_n under H_0.

5.2 The asymptotic distribution of the test statistic under H_0

To derive the asymptotic distribution of the test statistic T_n under H_0 we need yet another representation for it in terms of the process $S_n(\mathbf{t})$ in (1.2). In fact,

$$\begin{aligned}
T_n &= \int_{\mathbf{R}^d} |S_n(\mathbf{t})|^2 g(\mathbf{t})d\mathbf{t} \\
&= \int_{\mathbf{R}^d}\left\{\left[S_n^{(1)}(\mathbf{t})\right]^2 + \left[S_n^{(2)}(\mathbf{t})\right]^2\right\} g(\mathbf{t})d\mathbf{t} \\
&= \int_{\mathbf{R}^d}\left\{\left[S_n^{(1)}(\mathbf{t}) + S_n^{(2)}(\mathbf{t})\right]^2 - 2S_n^{(1)}(\mathbf{t})S_n^{(2)}(\mathbf{t})\right\} g(\mathbf{t})d\mathbf{t} \\
&= \int_{\mathbf{R}^d}\left[S_n^{(1)}(\mathbf{t}) + S_n^{(2)}(\mathbf{t})\right]^2 g(\mathbf{t})d\mathbf{t},
\end{aligned}$$

because $S_n^{(1)}(-\mathbf{t}) = S_n^{(1)}(\mathbf{t})$ and $S_n^{(2)}(-\mathbf{t}) = -S_n^{(2)}(\mathbf{t})$, and hence for any even function $g(\mathbf{t})$

$$\int_{\mathbf{R}^d} S_n^{(1)}(\mathbf{t})S_n^{(2)}(\mathbf{t})g(\mathbf{t})d\mathbf{t} = 0.$$

Under H_0, when (3.2) holds, $S_n(\mathbf{t})$ converges in distribution to $S_F(\mathbf{t})$ in every compact set in $\mathbf{R}^d$. According to Corollary 1 in Billingsley (1968, p. 31), $W_n(\mathbf{t}) = \left[S_n^{(1)}(\mathbf{t}) + S_n^{(2)}(\mathbf{t})\right] \xrightarrow{\mathcal{D}} W(\mathbf{t}) = \left[S_F^{(1)}(\mathbf{t}) + S_F^{(2)}(\mathbf{t})\right]$ in every compact set in $\mathbf{R}^d$.

It is obvious that $W(\mathbf{t})$ is a real valued continuous–parameter second order random process with zero mean and continuous covariance function

$$\begin{aligned} K(\mathbf{s},\mathbf{t}) &= E\left[W(\mathbf{s})W(\mathbf{t})\right] \\ &= E\left[S_F^{(1)}(\mathbf{s})S_F^{(1)}(\mathbf{t})\right] + E\left[S_F^{(1)}(\mathbf{s})S_F^{(2)}(\mathbf{t})\right] \\ &+ E\left[S_F^{(2)}(\mathbf{s})S_F^{(1)}(\mathbf{t})\right] + E\left[S_F^{(2)}(\mathbf{s})S_F^{(2)}(\mathbf{t})\right], \end{aligned}$$

where the expectations can be found in Csörgő (1985).

The next theorem holds for all nonnegative and integrable weight functions $g(\mathbf{t})$.

Theorem 5.2 *Under H_0*

$$T_n = \int_{\mathbf{R}^d} W_n(\mathbf{t})^2 g(\mathbf{t}) d\mathbf{t} \xrightarrow{\mathcal{D}} T = \int_{\mathbf{R}^d} W(\mathbf{t})^2 g(\mathbf{t}) d\mathbf{t}.$$

Proof. Let

$$T_u = \int_{-\mathbf{u}}^{\mathbf{u}} W(\mathbf{t})^2 g(\mathbf{t}) d\mathbf{t}, \ \mathbf{u} = (u_1, \ldots, u_d), \ u_k > 0, \ k = 1, \ldots, d$$

and

$$T_{un} = \int_{-\mathbf{u}}^{\mathbf{u}} W_n(\mathbf{t})^2 g(\mathbf{t}) d\mathbf{t}, \ \mathbf{u} = (u_1, \ldots, u_d), \ u_k > 0, \ k = 1, \ldots, d.$$

Because under H_0 $W_n(\mathbf{t}) \xrightarrow{\mathcal{D}} W(\mathbf{t})$ in every compact set of $\mathbf{R}^d$, Corollary 1 in Billingsley (1968, p. 31) yields $T_{un} \xrightarrow{\mathcal{D}} T_u$ as $n \to \infty$. Using formula (5) in Billingsley (1968)p. 223 one derives for every n, by taking $X = T_n - T_{un}$, that

$$\Pr\left(T_n - T_{un} \geq \alpha\right) \leq \frac{1}{\alpha} E\left[T_n - T_{un}\right], \ \alpha > 0.$$

It means that

$$\Pr\left(T_n - T_{un} \geq \alpha\right) \leq \int_{\mathbf{R}^d \setminus [-\mathbf{u},\mathbf{u}]} E\left[W_n(\mathbf{t})^2\right] g(\mathbf{t}) d\mathbf{t},$$

where $E\left[W_n(\mathbf{t})^2\right] = \sigma_n^2(\mathbf{t}) + o(n)$, $\sigma_n^2(\mathbf{t})$ is in (3.4), is uniformly bounded, therefore

$$\lim_{\mathbf{u}\to\infty} \limsup_{n\to\infty} \Pr\left(T_n - T_{un} \geq \alpha\right) = 0, \ \alpha > 0.$$

The same reasoning, with $X = T - T_u$, yields $T_u \xrightarrow{P} T$, as $\mathbf{u} \to \infty$, and therefore $T_u \xrightarrow{\mathcal{D}} T$, as $\mathbf{u} \to \infty$.

All the conditions in Theorem 4.2 in Billingsley (1968) are satisfied, hence $T_n \xrightarrow{\mathcal{D}} T$, as $n \to \infty$. □

The next result is an extension of results by Feuerverger and Mureika (1977). In the proof, however, the multivariate version of the Karhunen–Loéve expansion (Ash, 1965, Appendix) is needed. Suppose that the distribution of the variables is such that dependence in $\mathbf{R}^d$ implies dependence in $(-\mathbf{a}, \mathbf{a})$, see Chapter 2, and let the weight function be $g1(\mathbf{t})$ having the properties stated in (4.3).

Theorem 5.3 *The test statistic*

$$T1_n = \int_{-\mathbf{a}}^{\mathbf{a}} W_n(\mathbf{t})^2 g1(\mathbf{t}) d\mathbf{t}$$

is asymptotically distributed as

$$T1 = \int_{-\mathbf{a}}^{\mathbf{a}} W(\mathbf{t})^2 g1(\mathbf{t}) d\mathbf{t} = \sum_{j=1}^{\infty} \lambda_j Z_j^2,$$

where $\{\lambda_j\}$ *is the solution of the eigenvalue equation*

$$\lambda_j e_j(\mathbf{t}) = \int_{-\mathbf{a}}^{\mathbf{a}} e_j(\mathbf{s}) K(\mathbf{s}, \mathbf{t}) [g1(\mathbf{t}) g1(\mathbf{s})]^{\frac{1}{2}} d\mathbf{s}$$

and the distribution of $T1$ *is a weighted sum of independent* $\chi^2(1)$ *variables. The characteristic function of the distribution is*

$$c_{T1}(v) = \prod_{j=1}^{\infty} (1 - 2iv\lambda_j)^{\frac{1}{2}}.$$

The solution of the eigenvalue equation is not straightforward when $d = 2$ (Feuerverger, 1993). For $d > 2$ we can't even find $K1(\mathbf{s})$ and $K2(\mathbf{t})$ satisfying $K(\mathbf{s}, \mathbf{t}) = K1(\mathbf{s}) K2(\mathbf{t})$, therefore the critical values of the distribution are approximated. A simple way to approximate the distribution is to use

$$T1 \sim \gamma \chi^2(\beta),$$

see Johnson and Kotz (1970, p. 165), where γ and β are set to match the limits of expectation and variance of the test statistic. More precisely,

$$\begin{aligned} \beta &= \frac{2(E[T1])^2}{\mathrm{Var}[T1]} \\ \gamma &= \frac{\mathrm{Var}[T1]}{2E[T1)}, \end{aligned}$$

where $E[T1] = \lim_{n \to \infty} E[T1_n]$ and $\mathrm{Var}[T1] = \lim_{n \to \infty} \mathrm{Var}[T1_n]$.

An alternative way to bypass the problem is to estimate the fractiles of the asymptotic distribution of $T1_n$ by the bootstrap method (Efron and Tibshirani, 1993).

Due to Theorem 5.3 also the distribution of $T2_u$ is a weighted sum of independent $\chi^2(1)$ variables,

$$T2_u = \int_{-\mathbf{u}}^{\mathbf{u}} W(\mathbf{t})^2 g2(\mathbf{t}) d\mathbf{t} = \sum_{j=1}^{\infty} \lambda_j Z_j^2, \ \mathbf{u} = (u_1, \ldots, u_d), \ u_k > 0, \ k = 1, \ldots d,$$

where $\{\lambda_j\}$ is the solution of the eigenvalue equation

$$\lambda_j e_j(\mathbf{t}) = \int_{-\mathbf{u}}^{\mathbf{u}} e_j(\mathbf{s}) K(\mathbf{s}, \mathbf{t}) [g2(\mathbf{t}) g2(\mathbf{s})]^{\frac{1}{2}} d\mathbf{s}.$$

The characteristic function of the distribution is

$$c_{T2_u}(v) = \prod_{j=1}^{\infty} (1 - 2iv\lambda_j)^{\frac{1}{2}}.$$

On the other hand due to Theorem 5.2, under H_0

$$T2_n = \int_{\mathbf{R}^d} W_n(\mathbf{t})^2 g2(\mathbf{t}) d\mathbf{t} \xrightarrow{\mathcal{D}} T2 = \int_{\mathbf{R}^d} W(\mathbf{t})^2 g2(\mathbf{t}) d\mathbf{t}.$$

And because $W(\mathbf{t})^2 g2(\mathbf{t})$ is integrable, $T2_u \xrightarrow{P} T2$ and so $\Pr\{T2_u \leq q\} \approx \Pr\{T2 \leq q\}$, when $|u_k| > M_k$, $k = 1, \ldots, d$ for some positive constants M_k. Hence the asymptotic distribution of $T2_n$ can be approximated by the distibution of $T2_u$.

The approximation $\gamma\chi^2(\beta)$ with $\beta = 2(E[T2])^2/\mathrm{Var}[T2]$ and $\gamma = \mathrm{Var}[T2]/2E[T2]$ can be used, because $E[T2_u] \approx E[T2] = \lim_{n\to\infty} E[T2_n]$ and $\mathrm{Var}[T2_u] \approx \mathrm{Var}[T2] = \lim_{n\to\infty} \mathrm{Var}[T2_n]$.

5.3 Asymptotical power considerations

Under H_1 the distribution of $\frac{T_n}{n}$ can be approximated by $N(f(\theta), \frac{\sigma_1^2}{n})$. Then

$$\Pr\{T_n \geq q | H_1\} \approx \Phi\left(\frac{\sqrt{n} f(\theta) - \frac{q}{\sqrt{n}}}{\sigma_1}\right),$$

which can be made arbitrarily close to one for every alternative by taking n large enough (when $f(\theta) \neq 0$: the weight function having the properties stated in (4.3) or $c(\mathbf{t})$ is analytic if the weight function is of the form (4.4)).

Under H_0 $T1_n$ is asymptotically distributed as a weighted sum of independent $\chi^2(1)$ variables and the limiting distribution of $T2_n$ can be approximated by a distribution of weighted sum of independent $\chi^2(1)$ variables.

The critical values of the asymptotic distribution can be calculated using the approximation

$$\Pr\{T_n \geq q | H_0\} \approx 1 - F_{\gamma\chi^2(\beta)}(q),$$

with $\beta = \frac{2(E[T])^2}{\mathrm{Var}[T]}$ and $\gamma = \frac{\mathrm{Var}[T]}{2E[T]}$,

$$E[T] = \lim_{n\to\infty} E[T1_n],\ \mathrm{Var}[T] = \lim_{n\to\infty} \mathrm{Var}[T1_n]$$

for $T1_n$ and

$$E[T] \approx \lim_{n\to\infty} E[T2_n],\ \mathrm{Var}[T] \approx \lim_{n\to\infty} \mathrm{Var}[T2_n]$$

for $T2_n$.

The approximative size α critical region of the test T_n is

$$C = \{T_n | T_n \geq q_\alpha\}$$

where $F_{\gamma\chi^2(\beta)}(q_\alpha) = 1 - \alpha$.

The next two theorems show consistency of the test T_n.

Theorem 5.4 *If the underlying distribution $F(\mathbf{x})$ satisfies the condition*

$$\int_{\mathbf{R}^d} e^{\delta|\mathbf{x}|} dF(\mathbf{x}) < \infty$$

for some $\delta > 0$ then, under H_1,

$$T_n \overset{a.s.}{\to} \infty,\ n \to \infty, \tag{5.2}$$

for any nonnegative weight function $g(\mathbf{t})$ such that

$$0 < \int_{\mathbf{R}^d} g(\mathbf{t}) d\mathbf{t} < \infty. \tag{5.3}$$

Proof. Let random variables $X_1, \ldots, X_d$ be dependent, then

$$c(\mathbf{t}) \not\equiv \prod_{k=1}^{d} c_k(t_k).$$

According to Theorem 3.3.2 and Corollary 3.3.3 in Cuppens (1975) both characteristic functions $c(\mathbf{t})$ and $c_0(\mathbf{t}) = \prod_{k=1}^{d} c_k(t_k)$ are analytic, therefore their extensions from interval $(-\mathbf{a}, \mathbf{a})$ to $\mathbf{R}^d$ are unique. The Lebesque measure of the set, where

$$c(\mathbf{t}) = \prod_{k=1}^{d} c_k(t_k)$$

is zero. Thus

$$\int_U |c(\mathbf{t}) - \prod_{k=1}^{d} c_k(t_k)| d\mathbf{t} > 0 \tag{5.4}$$

for any set U having positive Lebesgue measure.

It follows from (5.3) that there exists a set U_0 of positive Lebesgue measure such that

$$g(\mathbf{t}) > 0, \ \mathbf{t} \in U_0. \tag{5.5}$$

From (5.4) and (5.5) we obtain

$$\int_{\mathbf{R}^d} |c(\mathbf{t}) - \prod_{k=1}^{d} c_k(t_k)| g(\mathbf{t}) d\mathbf{t} \geq \int_{U_0} |c(\mathbf{t}) - \prod_{k=1}^{d} c_k(t_k)| g(\mathbf{t}) d\mathbf{t} > 0. \tag{5.6}$$

Further, it is known that (Csörgő, 1981b)

$$\sup_{\mathbf{t} \in B} |c_n(\mathbf{t}) - c(\mathbf{t})| \overset{a.s.}{\to} 0, \ n \to \infty \tag{5.7}$$

and

$$\sup_{\mathbf{t} \in B} \left| \prod_{k=1}^{d} c_{nk}(t_k) - \prod_{k=1}^{d} c_k(t_k) \right| \overset{a.s.}{\to} 0, \ n \to \infty \tag{5.8}$$

for any bounded set B. Together with (5.3) this means that

$$\int_{\mathbf{R}^d} |c_n(\mathbf{t}) - c(\mathbf{t})| g(\mathbf{t}) d\mathbf{t} \overset{a.s.}{\to} 0, \ n \to \infty \tag{5.9}$$

and

$$\int_{\mathbf{R}^d} \left| \prod_{k=1}^{d} c_{nk}(t_k) - \prod_{k=1}^{d} c_k(t_k) \right| g(\mathbf{t}) d\mathbf{t} \overset{a.s.}{\to} 0, \ n \to \infty. \tag{5.10}$$

Besides,

$$\begin{aligned} \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)| g(\mathbf{t}) d\mathbf{t} \geq \\ \int_{\mathbf{R}^d} |c(\mathbf{t}) - \prod_{k=1}^{d} c_k(t_k)| g(\mathbf{t}) d\mathbf{t} - \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - c(\mathbf{t})| g(\mathbf{t}) d\mathbf{t} \\ - \int_{\mathbf{R}^d} \left| \prod_{k=1}^{d} c_{nk}(t_k) - \prod_{k=1}^{d} c_k(t_k) \right| g(\mathbf{t}) d\mathbf{t}. \end{aligned} \tag{5.11}$$

Using (5.6) and (5.9)–(5.11), we obtain

$$\liminf_{n \to \infty} \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)| g(\mathbf{t}) d\mathbf{t} > 0$$

and hence

$$\liminf_{n \to \infty} \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d\mathbf{t} > 0. \tag{5.12}$$

This is due to Cauchy–Schwarz–Bunjakovski inequality,

$$\int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d\mathbf{t} \geq \frac{\left(\int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)| g(\mathbf{t}) d\mathbf{t}\right)^2}{\int_{\mathbf{R}^d} g(\mathbf{t}) d\mathbf{t}}.$$

The limit (5.2) follows immediately from (5.12). □

Theorem 5.5 *If the integrable weight function $g(\mathbf{t})$ satisfies the condition*

$$g(\mathbf{t}) > 0 \tag{5.13}$$

for all $\mathbf{t} \in \mathbf{R}^d$, then

$$T_n \overset{a.s.}{\to} \infty, \ n \to \infty, \tag{5.14}$$

for any underlying distribution $F(\mathbf{x})$.

Proof. Since

$$c(\mathbf{t}) \not\equiv \prod_{k=1}^{d} c_k(t_k),$$

there exists an open set U of positive Lebesgue measure such that

$$\inf_{\mathbf{t} \in U} |c(\mathbf{t}) - \prod_{k=1}^{d} c_k(t_k)| > 0.$$

Hence, due to (5.13),

$$\int_{\mathbf{R}^d} |c(\mathbf{t}) - \prod_{k=1}^{d} c_k(t_k)|^2 g(\mathbf{t}) d\mathbf{t} > 0.$$

Together with (5.7), (5.8) and (5.11) this means that

$$\liminf_{n \to \infty} \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d\mathbf{t} > 0,$$

which implies (5.14). □

In Chapter 6.3 there are some empirical results concerning the power of the test.

5.4 Equivalent test in a special case

A test statistic for total independence in terms of kernel estimators of its density and marginal density functions is constructed, in a special case. It appears that this test is equivalent to the one defined by T_n provided the weight function is properly chosen. It also appears that this test is very close to the one of Rosenblatt (1975), but in Rosenblatt's test there are some additional constraints concerning the functions involved.

Let X be a random variable whose distribution function $F(x)$ is absolutely continuous and probability density function is $f(x)$. $F(x)$ can be estimated by empirical distribution function $F_n(x)$ and $f(x)$ by the kernel method leading to

$$f_n(x) = \frac{1}{nh(n)} \sum_{j=1}^{n} v\left(\frac{x - X_j}{h(n)}\right),$$

where $h(n) \to 0$ and $nh(n) \to \infty$, when $n \to \infty$, and v is a suitably chosen kernel function (Parzen, 1962, theorem A1). For a d–dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ we may write

$$f_n(\mathbf{x}) = \frac{1}{nh^d(n)} \sum_{j=1}^{n} v\left(\frac{\mathbf{x} - \mathbf{X}_j}{h(n)}\right), \tag{5.15}$$

where $h(n) \to 0$ and $nh^d(n) \to \infty$, when $n \to \infty$. Let $v(\mathbf{x})$ be a probability density function of the form

$$v(\mathbf{x}) = \prod_{k=1}^{d} v_k(x_k),$$

where $v_k(x_k)$ is the kth marginal probability density function. Let $F_{nk}(x_k)$ and $f_{nk}(x_k)$ be the marginal empirical distribution functions and probability density functions respectively, and let $c_n(\mathbf{t})$ and $c_{nk}(t_k)$ be the empirical characteristic and empirical marginal characteristic functions corresponding to these distributions.

The hypothesis of total independence as written in terms of the density function

$$H_0 : \; f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_{nk}(x_k).$$

In testing for independence one can use the test statistic

$$D_n = n \int_{\mathbf{R}^d} [f_n(\mathbf{x}) - \prod_{k=1}^{d} f_{nk}(x_k)]^2 w(\mathbf{x}) d\mathbf{x}, \tag{5.16}$$

where $w(\mathbf{x})$ is a weight function.

Let us denote

$$v^{(n)}(\mathbf{x}) = \frac{1}{h^d(n)} v\left(\frac{\mathbf{x}}{h(n)}\right)$$
$$v_k^{(n)}(x_k) = \frac{1}{h(n)} v_k\left(\frac{x_k}{h(n)}\right)$$

and let $V^{(n)}(\mathbf{x})$ and $V_k^{(n)}(x_k)$ be the distribution functions corresponding to probability densities $v^{(n)}(\mathbf{x})$ and $v_k^{(n)}(x_k)$, respectively. Then according to (5.15) $f_n(\mathbf{x})$ is the probability density function corresponding to $F_n * V^{(n)}(\mathbf{x})$, the convolution of $F_n(\mathbf{x})$ and $V^{(n)}(\mathbf{x})$. Let $\varphi_n(\mathbf{t})$ and $\varphi_{nk}(t_k)$ be the characteristic functions of $V^{(n)}(\mathbf{x})$ and $V_k^{(n)}(x_k)$. Then the characteristic function of $f_n(\mathbf{x})$ is $c_n(\mathbf{t})\varphi_n(\mathbf{t})$.

Theorem 5.6 *For weight functions $g(\mathbf{t}) = |\varphi_n(\mathbf{t})|^2$ and $w(\mathbf{x}) = 1$, the tests based on*

$$T_n = n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d\mathbf{t}$$

and D_n in (5.16) are equivalent. In fact,

$$T_n \equiv (2\pi)^{\frac{d}{2}} D_n.$$

Proof. Using the Plancherel's equation we get

$$\begin{aligned} T_n &= n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d\mathbf{t} \\ &= n \int_{\mathbf{R}^d} |c_n(\mathbf{t})\varphi_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)\varphi_{nk}(t_k)|^2 d\mathbf{t} \\ &= n(2\pi)^{\frac{d}{2}} \int_{\mathbf{R}^d} \left[f_n(\mathbf{x}) - \prod_{k=1}^{d} f_{nk}(x_k) \right]^2 d\mathbf{x} \\ &= (2\pi)^{\frac{d}{2}} D_n. \ \square \end{aligned}$$

Remark: Rosenblatt's test statistic (Rosenblatt, 1975), is very much the same as $\frac{D_n}{n}$ here, but there are some additional constraints concerning the functions involved, see the assumptions $a1 - a4$ in Rosenblatt (1975). The asymptotic distribution of Rosenblatt's test is normal and the asymptotic distribution of the test statistic $\frac{D_n}{n}$ is normal under H_1, but under H_0 it is degenerated at zero, because the variance of $\frac{D_n}{n}$ converges to zero so fast that $\sqrt{n}$ is no longer the appropriate normalizing factor; it tends to infinity too slowly. The difference in asymptotical behaviour can be explained by the assumption $a4$, which says that for the Rosenblatt's test the weight function $w(\mathbf{x})$ should be integrable. When the D_n test is constructed, the weight function is $w(\mathbf{x}) = 1$ and therefore in Theorem 2 (and its generalization for d>2) by Rosenblatt (1975) for the D_n test the term $A(n) \neq o(h(n))$.

5.5 A nonparametric modification of the test

Additional restrictions, concerning the tail of the underlying distribution, seem to be the disadvantage of the new test. Now a consistent and asymptotically distribution free modification of the test is considered. The idea is to transform the data so that the new variables have analytic characteristic functions (like in Feuerverger and Mureika, 1977; Feuerverger, 1993).

Let $f_1(x),\ldots,f_d(x)$ be continuous, bounded and strictly monotone functions. Then the independence of random variables $X_1,\ldots,X_d$ is equivalent to the independence of the random variables $f_1(X_1),\ldots,f_d(X_d)$, because

$$\Pr(X_1 \in A_1,\ldots,X_d \in A_d) = \Pr(f_1(X_1) \in f_1(A_1),\ldots,f_d(X_d) \in f_d(A_d)) \tag{5.17}$$

for any Borel sets $A_1,\ldots,A_d$.

The relationship between independence (or dependence) of random variables $X_1,\ldots,X_d$ and that of $f_1(X_1),\ldots,f_d(X_d)$ is not only qualitative, but quantitative as well. Indeed, let the independence measure be

$$\begin{aligned}\tau(X_1,\ldots,X_d) &= \tau(X_1,\ldots,X_d|\mathcal{A}) \\ &= \sup\left|\Pr\left(\bigcap_{i=1}^{d}\{X_i \in A_i\}\right) - \prod_{k=1}^{d}\Pr(X_i \in A_i)\right|,\end{aligned}$$

where supremum is taken over all collections of sets $A_1,\ldots,A_d$ from a fixed subclass $\mathcal{A}$ of Borel sets of the real line. It is reasonable to take for $\mathcal{A}$ either the class of all intervals (or half-lines) or the class of all Borel sets. In particular, random variables $X_1,\ldots,X_d$ are independent if and only if $\tau(X_1,\ldots,X_d) = 0$.

It easy to see that

$$\tau(X_1,\ldots,X_d) = \tau(f_1(X_1),\ldots,f_d(X_d)).$$

Thus testing for independence of $X_1,\ldots,X_d$ is equivalent with testing for independence of $f_1(X_1),\ldots,f_d(X_d)$.

On the other hand, in the case of variables $f_1(X_1),\ldots,f_d(X_d)$, according to Theorem 5.7 the consistency of the test based on the statistic T_n holds for any possible weight function, without any restrictions concerning the tail of the distribution. Therefore, the test statistic T_n can be replaced by the test statistic ($\mathbf{f} = (f_1,\ldots,f_d)$)

$$T_n^{(\mathbf{f})} = n\int_{\mathbf{R}^d} |c_n^{(\mathbf{f})}(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}^{(f_k)}(t_k)|^2 w(\mathbf{t})d\mathbf{t}, \tag{5.18}$$

where

$$c_n^{(\mathbf{f})}(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n} e^{i<\mathbf{t},\mathbf{f}(\mathbf{X}_j)>} = \frac{1}{n}\sum_{j=1}^{n} e^{i(t_1 f_1(X_{j1}),\ldots,t_d f_d(X_{jd}))}$$

and

$$c_{nk}^{(f_k)}(t_k) = \frac{1}{n}\sum_{j=1}^{n} e^{it_k f_k(X_{jk})}.$$

Theorem 5.7 *For any underlying distribution function $F(\mathbf{x})$, any continuous, bounded and strictly monotone functions $f_1(X_1), \ldots, f_d(X_d)$ and any nonnegative function $w(\mathbf{t})$ such that*

$$0 < \int_{\mathbf{R}^d} w(\mathbf{t}) d\mathbf{t} < \infty, \tag{5.19}$$

the following relation holds

$$T_n^{(\mathbf{f})} \stackrel{a.s.}{\to} \infty, \ n \to \infty, \tag{5.20}$$

under H_1, i.e. the test based on the test statistic in (5.17) is consistent.

The proof of the theorem is analogous with the proof of Theorem 5.3.

Sometimes it is practical to use a random function $\mathbf{f}_n$ depending on n, therefore a generalization of Theorem 5.7 is needed.

Theorem 5.8 *Let the conditions of Theorem 5.7 be satisfied. Consider a sequence of random functions $\mathbf{f}_1, \mathbf{f}_2, \ldots$ such that, with probability 1, $\mathbf{f}_n \to \mathbf{f}$ uniformly, i.e.*

$$\Pr\left(\max_{1 \le k \le d} \sup_{x_k} |f_{nk}(x_k) - f_k(x_k)| \to 0, \ n \to \infty \right) = 1.$$

Then

$$T_n^{(\mathbf{f}_n)} \stackrel{a.s.}{\to} \infty, \ n \to \infty.$$

The proof of this theorem is analogous with the proofs of Theorems 5.3 and 5.7. One just has to take into account that

$$\sup_{\mathbf{t} \in B} |c_n^{(\mathbf{f}_n)}(\mathbf{t}) - c_n^{(\mathbf{f})}(\mathbf{t})| \stackrel{a.s.}{\to} 0, \ n \to \infty,$$

$$\sup_{\mathbf{t} \in B} \left| \prod_{k=1}^{d} c_{nk}^{(f_{nk})}(t_k) - \prod_{k=1}^{d} c_{nk}^{(f_k)}(t_k) \right| \stackrel{a.s.}{\to} 0, \ n \to \infty$$

for any bounded set B and

$$\begin{aligned}
\int_{\mathbf{R}^d} |c_n^{(\mathbf{f}_n)}(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}^{(f_{nk})}(t_k)| w(\mathbf{t}) d\mathbf{t} \ \ge \ & \int_{\mathbf{R}^d} |c_n^{(\mathbf{f})}(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}^{(f_k)}(t_k)| w(\mathbf{t}) d\mathbf{t} \\
& - \int_{\mathbf{R}^d} |c_n^{(\mathbf{f}_n)}(\mathbf{t}) - c_n^{(\mathbf{f})}(\mathbf{t})| w(\mathbf{t}) d\mathbf{t} \\
& - \int_{\mathbf{R}^d} |c_{nk}^{(f_{nk})}(t_k) - \prod_{k=1}^{d} c_{nk}^{(f_k)}(t_k)| w(\mathbf{t}) d\mathbf{t}.
\end{aligned}$$

When the underlying distribution function $F(\mathbf{x})$ is known and continuous, then for any $k = 1, \ldots, d$, $F_k(x_k)$ can be taken for the function $f_k(x_k)$.

Then each random variable $F_k(X_k)$ is uniformly distributed on the interval $[0,1]$ and hence the test based on the statistic $T_n^{(\mathbf{f})}$ is distribution free. See Table A.1 in Appendix for the aproximative critical values for the tests based on $T1_n$ and $T2_n$, when the two weight functions introduced in Chapter 6 are used.

When the underlying distribution is unknown, like it usually is, then the marginal distribution functions $F_{nk}(x_k)$ are taken for the functions $f_{nk}(x_k)$.

It has to be noticed, however, that this modification is argumented only for continuous distributions ($F(\cdot)$ has to be continuous and strictly monotone). The simulation study shows, however, that the powers of the test $T_n^{(F_n)}$, which uses transformed data, are quite similar for the discrete and continuous data that are used, when they are compared to the powers of the test, which uses original data. It can also be seen that, when the data is transformed, more observations are needed, because the empirical powers corresponding to tests using transformed data converge to one slowly.

The values of the test statistics

$$T1_n = n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g1(\mathbf{t})d\mathbf{t},$$

where

$$g1(\mathbf{t}) = \begin{cases} \frac{1}{2^d}, & t_k \in (-1,1),\ k = 1,\dots,d \\ 0, & \text{elsewhere,} \end{cases}$$

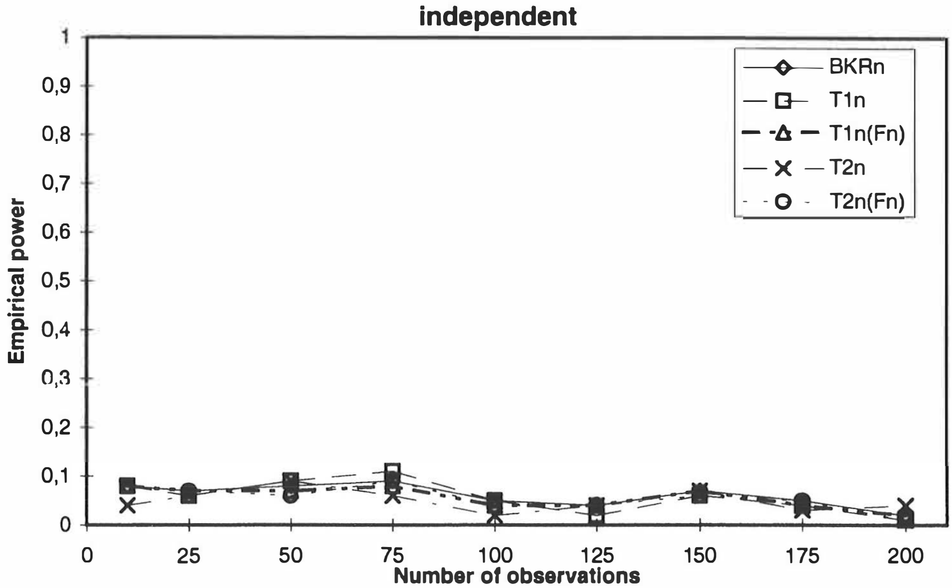
and

$$T2_n = n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g2(\mathbf{t})d\mathbf{t},$$

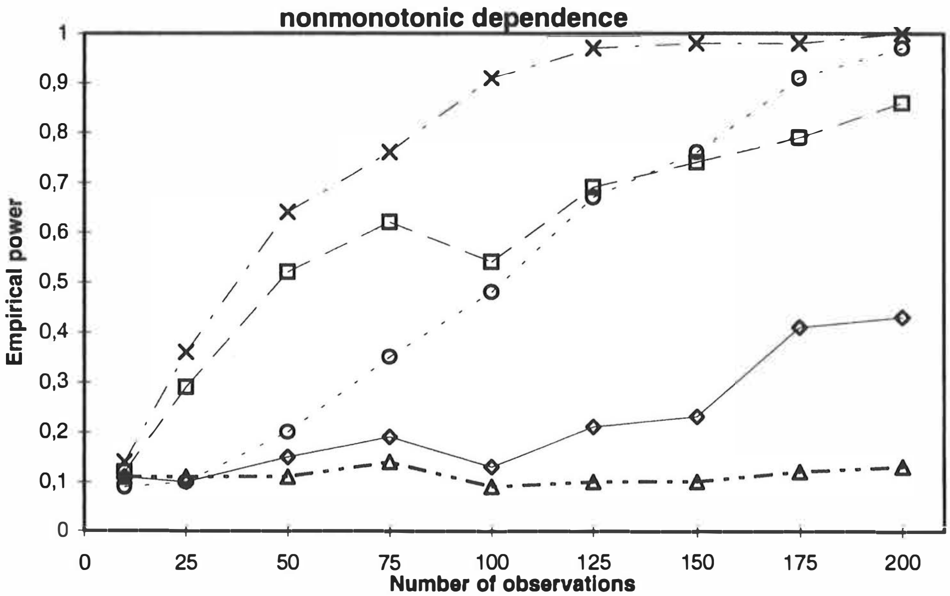
where

$$g2(\mathbf{t}) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t_k^2},\ \mathbf{t} \in \mathbf{R}^d,$$

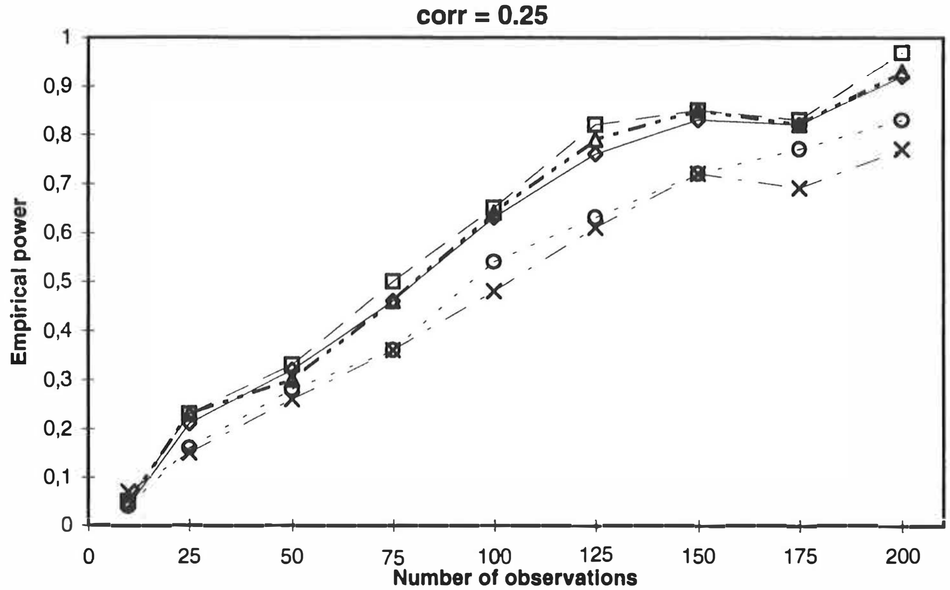
and the corresponding modifications $T1_n^{(F_n)}$ and $T2_n^{(F_n)}$ are calculated for eight different data, when there are two (d=2) and five (d=5) variables. (The descriptions of the data are after the pictures.) The empirical powers can be seen in the Pictures 5.1–5.16. There are also the empirical powers corresponding to the Blum–Kiefer–Rosenblatt test (see Chapter 6). Because the BKR_n test has different empirical sizes in case of discrete data, when five variables are involved, its empirical powers are not included in the pictures corresponding to those data.



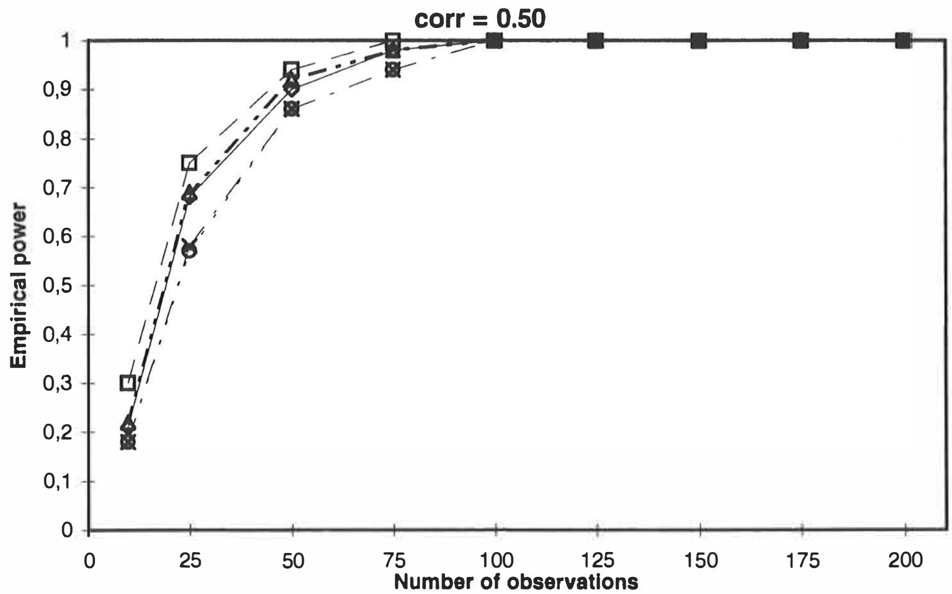
PICTURE 5.1 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where variables are independent and $N(0,1)$ distributed (data a).



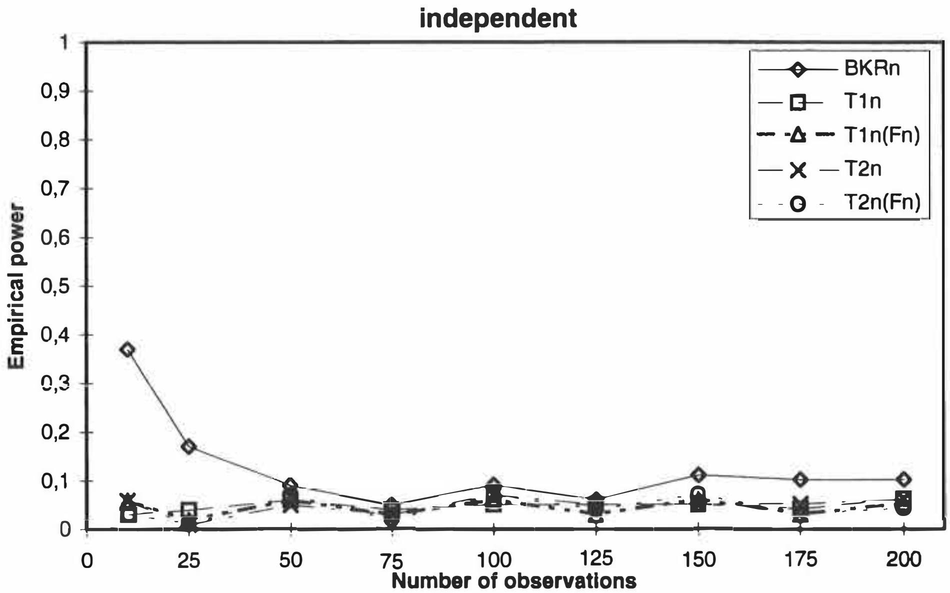
PICTURE 5.2 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where originally the independent variables are from $N(0,1)$ distribution and are transformed so that the variables are nonmonotonicly dependent (data b).



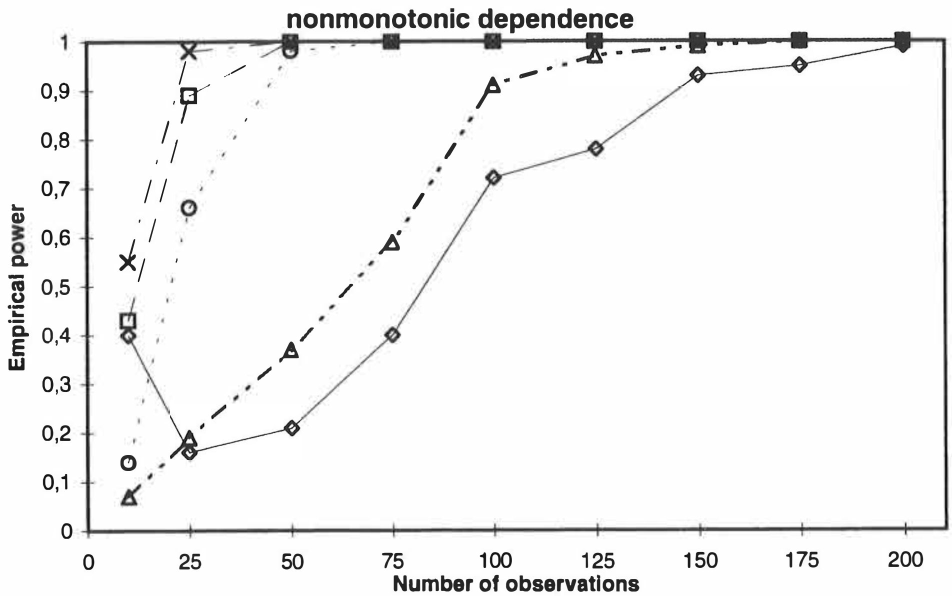
PICTURE 5.3 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where originally the independent variables are from $N(0,1)$ distribution and are transformed so that corr($\mathbf{X}_k, \mathbf{X}_l$)=0.25, $k \neq l$, $k, l = 1, \ldots, d$ (data c).



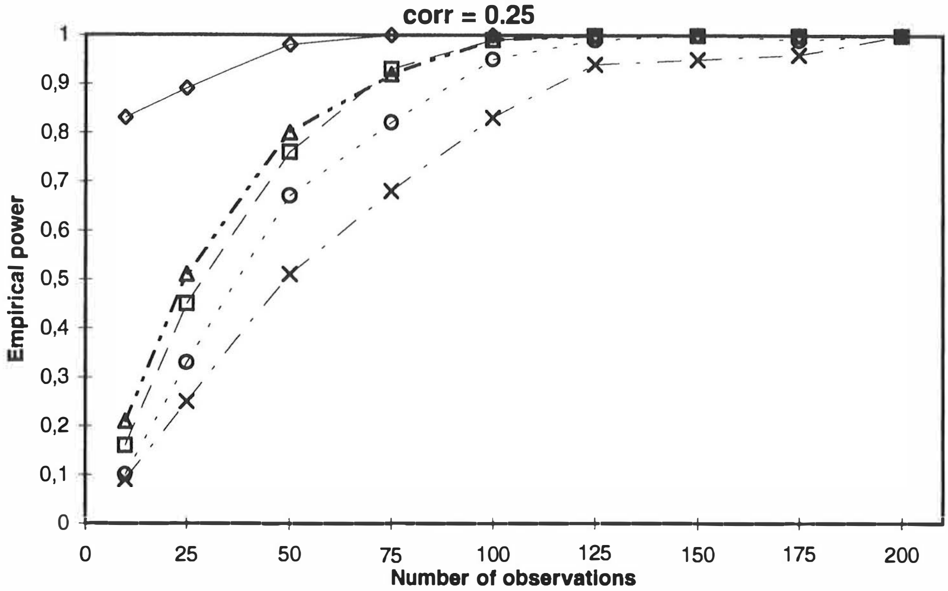
PICTURE 5.4 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where originally the independent variables are from $N(0,1)$ distribution and are transformed so that corr($\mathbf{X}_k, \mathbf{X}_l$)=0.50, $k \neq l$, $k, l = 1, \ldots, d$ (data d).



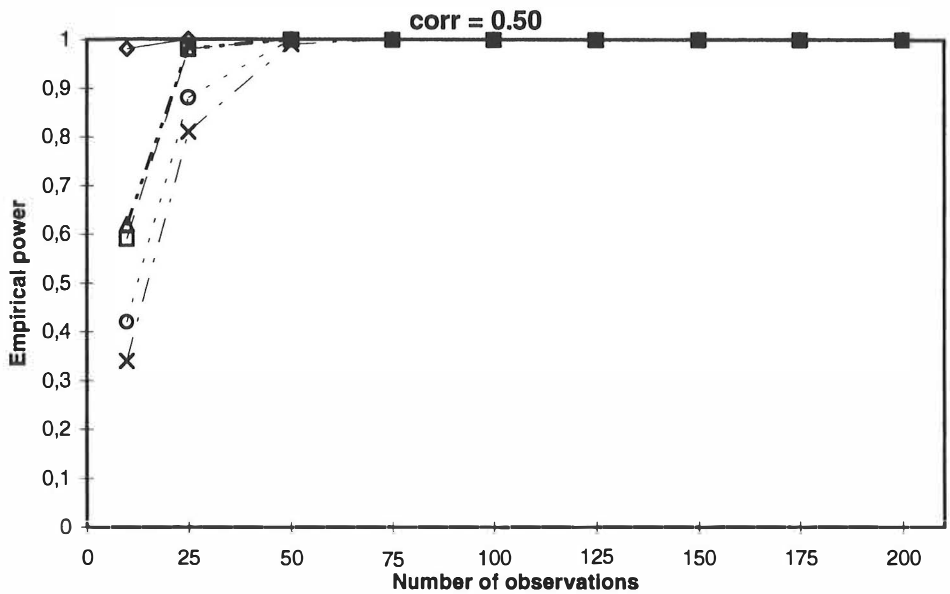
PICTURE 5.5 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where variables are independent and $N(0,1)$ distributed (data a).



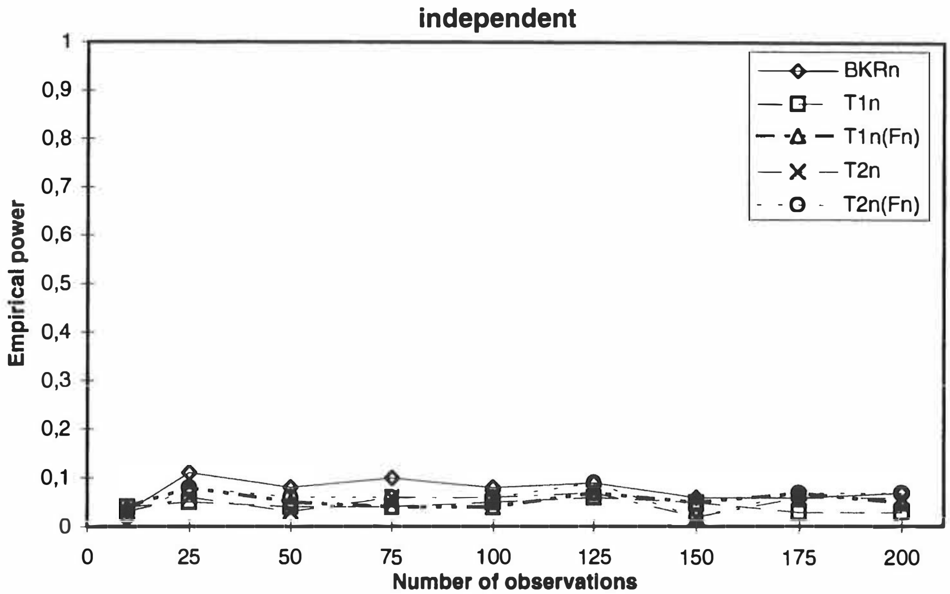
PICTURE 5.6 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where originally the independent variables are from $N(0,1)$ distribution and are transformed so that the variables are nonmonotonicly dependent (data b).



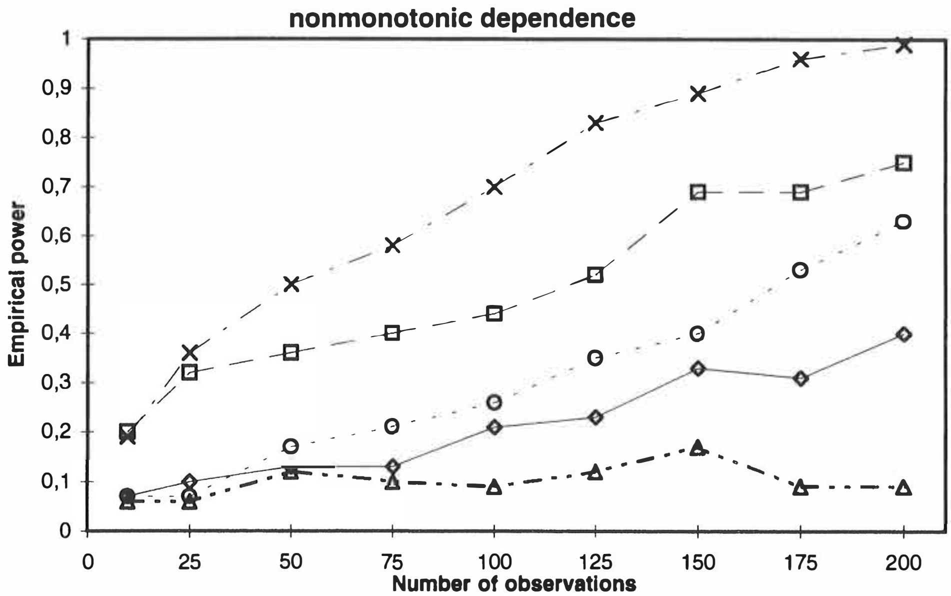
PICTURE 5.7 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where originally the independent variables are from $N(0,1)$ distribution and are transformed so that corr($\mathbf{X}_k, \mathbf{X}_l$)=0.25, $k \neq l$, $k, l = 1, \ldots, d$ (data c).



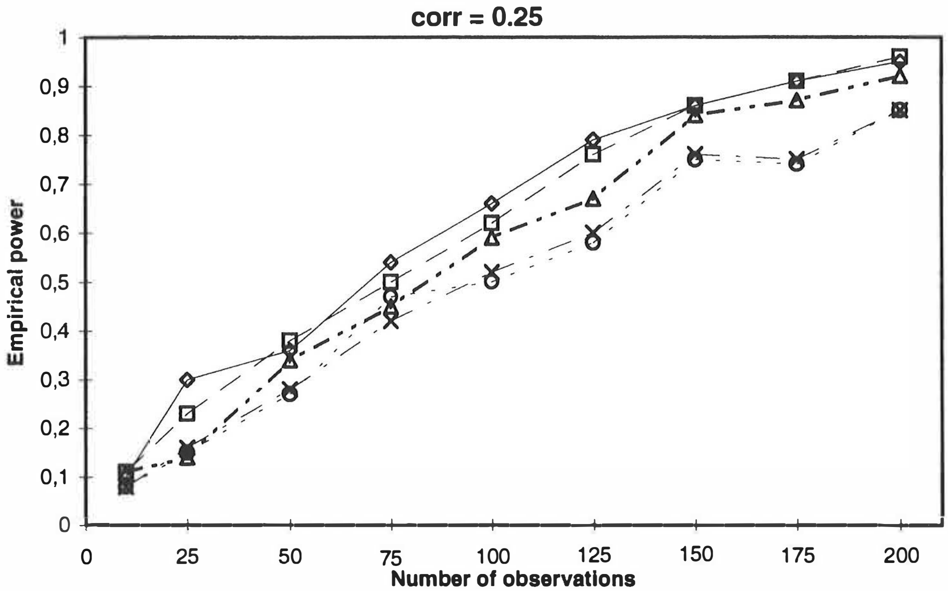
PICTURE 5.8 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where originally the independent variables are from $N(0,1)$ distribution and are transformed so that corr($\mathbf{X}_k, \mathbf{X}_l$)=0.50, $k \neq l$, $k, l = 1, \ldots, d$ (data d).



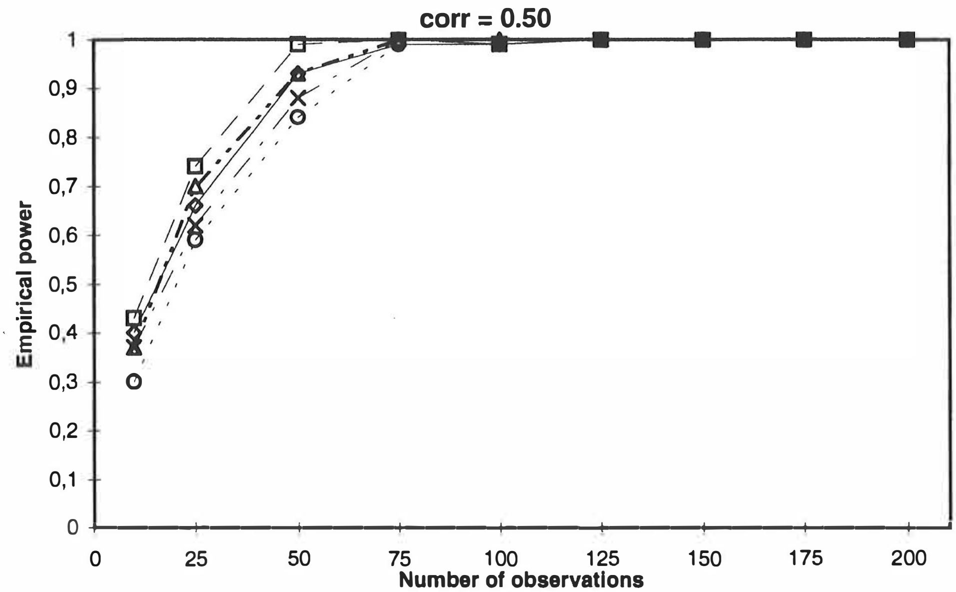
PICTURE 5.9 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where variables are independent and $Po(2)$ distributed (data e).



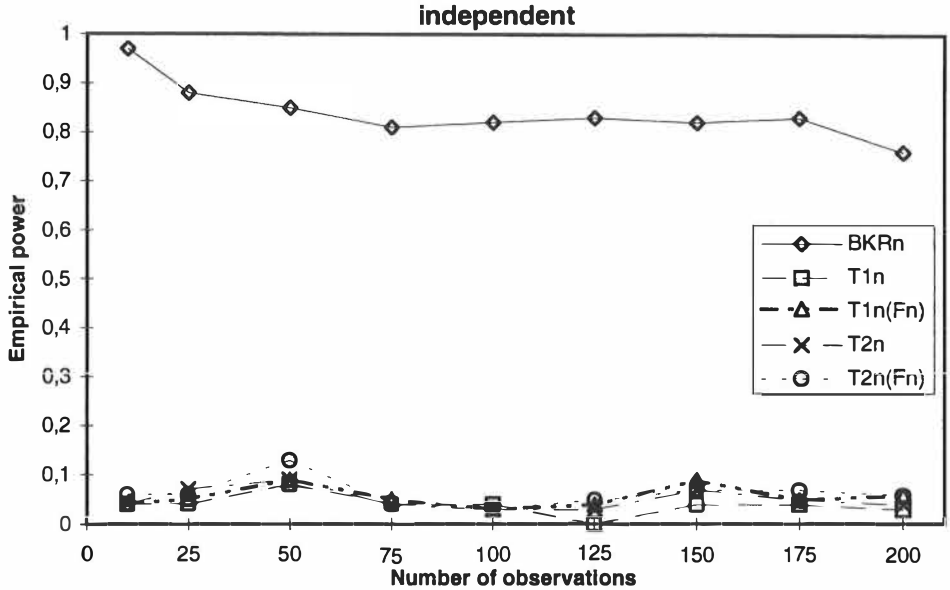
PICTURE 5.10 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where originally the independent variables are from $Po(2)$ distribution and are transformed so that the variables are nonmonotonicly dependent (data f).



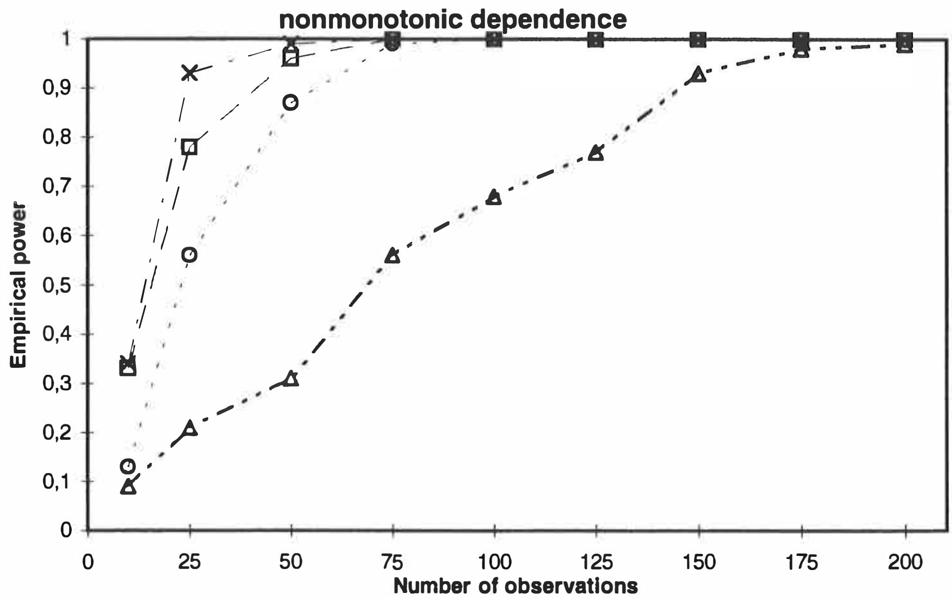
PICTURE 5.11 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where originally the independent variables are from $Po(2)$ distribution and are transformed so that $\text{corr}(\mathbf{X}_k, \mathbf{X}_l)=0.25$, $k \neq l$, $k, l = 1, \ldots, d$ (data g).



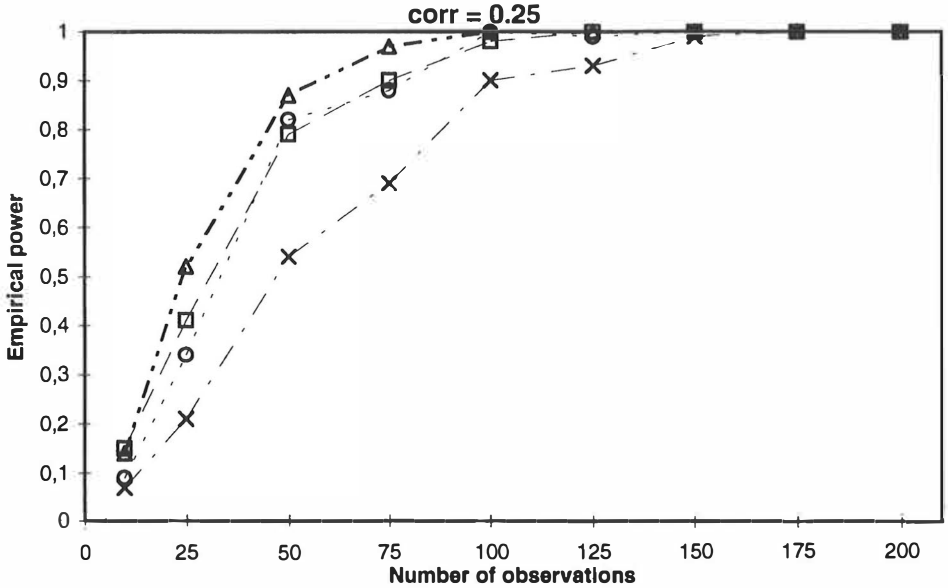
PICTURE 5.12 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of two variables, where originally the independent variables are from $Po(2)$ distribution and are transformed so that $\text{corr}(\mathbf{X}_k, \mathbf{X}_l)=0.50$, $k \neq l$, $k, l = 1, \ldots, d$ (data h).



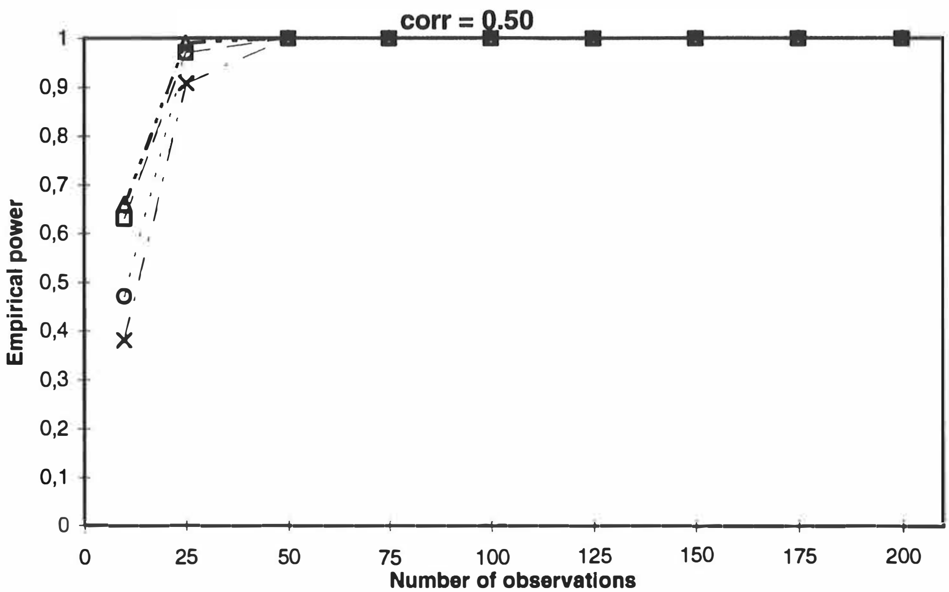
PICTURE 5.13 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where variables are independent and $Po(2)$ distributed (data e).



PICTURE 5.14 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where originally the independent variables are from $Po(2)$ distribution and are transformed so that the variables are nonmonotonicly dependent (data f).



PICTURE 5.15 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where originally the independent variables are from $Po(2)$ distribution and are transformed so that corr($\mathbf{X}_k, \mathbf{X}_l$)=0.25, $k \neq l$, $k, l = 1, \ldots, d$ (data g).



PICTURE 5.16 The empirical powers of the tests based on the statistics BKR_n, $T1_n$, $T1_n^{(F_n)}$, $T2_n$ and $T2_n^{(F_n)}$, corresponding to 100 simulated data of five variables, where originally the independent variables are from $Po(2)$ distribution and are transformed so that corr($\mathbf{X}_k, \mathbf{X}_l$)=0.50, $k \neq l$, $k, l = 1, \ldots, d$ (data h).

100 simulations are made for d=2 and d=5 of the following different data:

(a) $X_k,\ k=1,\ldots,d$, are independent $N(0,1)$ variables.

(b) $Z_0,\ Z_k,\ k=1,\ldots,d$, are independent $N(0,1)$ variables and the data are generated by

$$X_k = Z_0 Z_k,\ k=1,\ldots,d,$$

the variables are dependent, but $\mathrm{corr}(X_k,X_l)=0,\ k\neq l,\ k,l=1,\ldots,d$.

(c) $Z, Y_k,\ k=1,\ldots,d$, are independent $N(0,1)$ variables and the data come from

$$X_k = \alpha Y_k + (1-\alpha)Z,\ k=1,\ldots,d,$$

where α=$\frac{3-\sqrt{3}}{2}$, and $\mathrm{corr}(X_k,X_l)=0.25,\ k\neq l,\ k,l=1,\ldots,d$.

(d) $Z, Y_k,\ k=1,\ldots,d$, are independent $N(0,1)$ variables and the data are generated by

$$X_k = \alpha Y_k + (1-\alpha)Z,\ k=1,\ldots,d,$$

where α=0.5, and $\mathrm{corr}(X_k,X_l)=0.5,\ k\neq l,\ k,l=1,\ldots,d$.

(e) $X_k,\ k=1,\ldots,d$, are independent $Po(2)$ variables. (Then after studentizing, the new data are $X_k'=\frac{X_k}{\sqrt{2}}$, where $X_k\sim Po(2)$.)

(f) $Z_0,\ Z_{1k},\ Z_{2k},\ k=1,\ldots,d$, are independent $Po(2)$ variables and the data are generated by

$$X_k = Z_0(Z_{1k}-Z_{2k}),\ k=1,\ldots,d,$$

the variables are dependent, but $\mathrm{corr}(X_k,X_l)=0,\ k\neq l,\ k,l=1,\ldots,d$.

(g) $Z, Y_k,\ k=1,\ldots,d$, are independent $Po(\frac{1}{2})$ variables, $U_k,\ k=1,\ldots,d$ are independent $Po(1)$ variables, and the data come from

$$X_k = \frac{\alpha}{2}\left(Y_k+U_k\right)+(1-\alpha)\,Z,\ k=1,\ldots,d,$$

where α=$\frac{2}{3}$, and $\mathrm{corr}(X_k,X_l)=0.25,\ k\neq l,\ k,l=1,\ldots,d$. (Then after studentizing, the new data are $X_k'=\frac{Y_k+U_k+Z}{\sqrt{2}}$, where $Y_k+U_k+Z\sim Po(2)$.)

(h) $Z, Y_k,\ k=1,\ldots,d$, are independent $Po(1)$ variables and the data are generated by

$$X_k = \alpha Y_k + (1-\alpha)Z,\ k=1,\ldots,d,$$

where α=0.5, and $\mathrm{corr}(X_k,X_l)=0.5,\ k\neq l,\ k,l=1,\ldots,d$. (Then after studentizing, the new data are $X_k'=\frac{Y_k+Z}{\sqrt{2}}$, where $Y_k+Z\sim Po(2)$.)

6 EMPIRICAL COMPARISONS OF THE TESTS

In this chapter we choose two different weight functions and determine explicitely the corresponding test statistics $T1_n$ and $T2_n$. We also consider, in order to make comparisons, the test statistic BKR_n that Hoeffding (1948b) has introduced for two variables (a vector of dimesion d=2). Blum et al. (1961) have generalized this test in multivariate case, and Cotterill and Csörgő (1985) have calculated the critical values. Then some simulations are made to compare the empirical powers of the $T1_n$, $T2_n$ and BKR_n tests. The hypothesis of total independence is tested in 11 simulated data of different nature. In three data the variables (actually the components of the vector valued variable) are totally independent and in the other data, there are different kinds of dependencies between the variables. There are continous, $N(0,1)$, and discrete, $Po(2)$, variables, the number of components d varies from 2 to 10 and sample sizes are n=10, 50, 100 and 200.

6.1 Two different weight functions

The first weight function is product-uniform,

$$g1(\mathbf{t}) = \prod_{k=1}^{d} g1_k(t_k) = \begin{cases} \prod_{k=1}^{d} \frac{1}{2a_k}, & t_k \in (-a_k, a_k),\ k = 1,\ldots,d \\ 0, & \text{elsewhere.} \end{cases}$$

With studentized data, the test statistic is then of the form

$$T1_n \quad = n \quad \left[\frac{1}{n^2} \sum_{j=1}^{n} \sum_{m=1}^{n} \prod_{k=1}^{d} \frac{\sin\left[a_k \left(\frac{X_{jk}-X_{mk}}{S_k}\right)\right]}{a_k \left(\frac{X_{jk}-X_{mk}}{S_k}\right)} \right.$$

$$-\frac{2}{n^{d+1}}\sum_{j=1}^{n}\prod_{k=1}^{d}\sum_{m=1}^{n}\frac{\sin\left[a_k\left(\frac{X_{jk}-X_{mk}}{S_k}\right)\right]}{a_k\left(\frac{X_{jk}-X_{mk}}{S_k}\right)}$$
$$+\frac{1}{n^{2d}}\prod_{k=1}^{d}\sum_{j=1}^{n}\sum_{m=1}^{n}\frac{\sin\left[a_k\left(\frac{X_{jk}-X_{mk}}{S_k}\right)\right]}{a_k\left(\frac{X_{jk}-X_{mk}}{S_k}\right)}\Bigg].$$

The use of T_n is always based on studentition of the data, so the standard deviation is one for all variables. Thus, the values $a_k = 1$, $k = 1,\dots,d$, can be used, without any loss of generality.

Often in practice the distributions of the variables are unknown and the values $E1_k$, $E2_k$ and $E3_k$ are estimated by $\widehat{E1}_k$, $\widehat{E2}_k$ and $\widehat{E3}_k$, see Chapter 4. The estimates of the limits of expectation and variance of the test statistic can be calculated and then the size α critical values of the asymptotic distribution of $T1_n$ under H_0 can be approximated using $T1 \sim \hat{\gamma}\chi^2(\hat{\beta})$, where $\hat{\gamma}$ and $\hat{\beta}$ are set to match the estimated limits of expectation and variance of $T1_n$.

Let the second weight function be product–Gaussian,

$$g2(\mathrm{t}) = \prod_{k=1}^{d}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t_k^2},\ t_k \in \mathbf{R},\ k = 1,\dots,d.$$

Then, with studentized data, the test statistic will be

$$T2_n = n\ \Bigg[\frac{1}{n^2}\sum_{j=1}^{n}\sum_{m=1}^{n}\prod_{k=1}^{d}e^{-\frac{1}{2}\left(\frac{X_{jk}-X_{mk}}{S_k}\right)^2}$$
$$-\frac{2}{n^{d+1}}\sum_{j=1}^{n}\prod_{k=1}^{d}\sum_{m=1}^{n}e^{-\frac{1}{2}\left(\frac{X_{jk}-X_{mk}}{S_k}\right)^2}$$
$$+\frac{1}{n^{2d}}\prod_{k=1}^{d}\sum_{j=1}^{n}\sum_{m=1}^{n}e^{-\frac{1}{2}\left(\frac{X_{jk}-X_{mk}}{S_k}\right)^2}\Bigg].$$

The estimates for the limits of expectation and variance of $T2_n$ can be calculated using the estimates $\widehat{E1}_k$, $\widehat{E2}_k$ and $\widehat{E3}_k$. Then the limiting distribution of $T2_n$ under H_0 can be approximated by $\hat{\gamma}\chi^2(\hat{\beta})$.

In Appendix we have presented the values γ and β and the theoretical limits of expectations of the test statistics $T1_n$ and $T2_n$ under H_0 and also the approximative critical values of their asymptotic distributions under H_0 corresponding to sizes α=0.05, 0.01 and 0.001 in Table A.2, when all the variables are from $N(0,1)$ distribution, in Table A.3, when all the variables are from $Po(2)$ distribution and in Table A.4, when all the odd numbered variables are from $N(0,1)$ distribution and all the even numbered variables are from $Po(2)$ distribution. The limiting distribution can also be approximated by the bootstrap method and in Tables A.2, A.3 and A.4 in Appendix we have these approximative size α critical values from 1000 resamples of the original sample, where sample size n=200. The resamples are taken separately for each component of the vector of variables so that the components remain independent. For details of the bootstrap method see Efron and Tibshirani (1993).

While comparing the expectations of the test statistics, which are approximated by $\gamma\chi^2(\beta)$ or by the bootstrap method, one can see that the corresponding values are very close to each other. The difference between these corresponding values is some one-hundreths or some one-thousands of the corresponding standard deviation. So it seems that it doesn't make any difference which one of these approximations is used in testing the independence hypothesis. The same conclusion can be obtained by looking at the (t and b) empirical powers of the tests in Tables 6.1 to 6.9.

6.2 Blum–Kiefer–Rosenblatt test statistic

Hoeffding (1948b) developes a test of independence for two variables which is based on the empirical process

$$P_n(x_1, x_2) = n^{\frac{1}{2}} \left\{F_n(x_1, x_2) - F_{n1}(x_1)F_{n2}(x_2)\right\},$$

where F_n, F_{n1} and F_{n2} are empirical distribution functions corresponding to continuous distributions F, F_1 and F_2. Blum et al. (1961) generalizes this test to multivariate case, and the process is then

$$P_n(\mathbf{x}) = n^{\frac{1}{2}} \left\{F_n(\mathbf{x}) - \prod_{k=1}^{d} F_{nk}(x_k)\right\},$$

where

$$F_n(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{k=1}^{d} I_{\{X_{jk}\le x_k\}},\ \mathbf{X_j} = (X_{j1}, \ldots, X_{jd}),$$

$$I_{\{X_{jk}\le x_k\}} = \begin{cases} 1, & \text{if } X_{jk} \le x_k \\ 0, & \text{if } X_{jk} > x_k, \end{cases}$$

and

$$F_{nk}(x_k) = \frac{1}{n}\sum_{j=1}^{n} I_{\{X_{jk}\le x_k\}}.$$

The test statistic is of Cramér–von Mises type and is

$$BKR_n = \int_{\mathbf{R}^d} (P_n(\mathbf{x}))^2 \prod_{k=1}^{d} dF_{nk}(x_k),\ d \ge 2.$$

It can also be written in the form

$$
\begin{aligned}
BKR_n \; = n \; & \left[\frac{1}{n^2} \sum_{j=1}^{n} \sum_{m=1}^{n} \prod_{k=1}^{d} \left(\frac{1}{n} \sum_{l=1}^{n} I_{\{X_{jk} \le x_{lk}\}} I_{\{X_{mk} \le x_{lk}\}} \right) \right. \\
& - \frac{2}{n^{d+1}} \sum_{j=1}^{n} \prod_{k=1}^{d} \sum_{m=1}^{n} \left(\frac{1}{n} \sum_{l=1}^{n} I_{\{X_{jk} \le x_{lk}\}} I_{\{X_{mk} \le x_{lk}\}} \right) \\
& \left. + \frac{1}{n^{2d}} \prod_{k=1}^{d} \sum_{j=1}^{n} \sum_{m=1}^{n} \left(\frac{1}{n} \sum_{l=1}^{n} I_{\{X_{jk} \le x_{lk}\}} I_{\{X_{mk} \le x_{lk}\}} \right) \right] .
\end{aligned}
$$

The limiting distribution of BKR_n under H_0 is a weighted sum of $\chi^2(1)$ variables. Hoeffding (1948b), Blum et al. (1961) and Deheuvels (1981) study the limiting distribution and give suggestions how to calculate the test statistic and the critical values. Cotterill and Csörgő (1985) tabulates the critical values of the asymptotic distribution of BKR_n, when the number of variables d varies from 2 to 20. Blum et al. (1961) also points out that the results concerning the size and minimum power of the test are not significantly affected if discontinuous variables are admitted.

6.3 Power comparisons using simulation experiments

We perform simulations to compare the empirical powers of the tests $T1_n$, $T2_n$ and BKR_n. We also want to know how much the empirical powers of $T1_n$ and $T2_n$ differ, when the limiting distribution is estimated using data or, alternatively, using the knowledge that the variables are from $N(0,1)$ or $Po(2)$ distributions.

Sample sizes are $n = 10, 50, 100$ and 200, and the number of variables in the data d varies from 2 to 10. For each combination of n and d there are 11 different simulated data, in the first three of them the variables are totally independent and in the others they are dependent:

1. $X_k,\ k = 1, \ldots, d$, are independent $N(0,1)$ variables.

2. $X_k,\ k = 1, \ldots, d$, are independent $Po(2)$ variables. (Then after studentizing, the new data are $X'_k = \frac{X_k}{\sqrt{2}}$, where $X_k \sim Po(2)$.)

3. $X_k,\ k = 1, \ldots, d$, are independent variables, odd numbered are from $N(0,1)$ distribution and even numbered are from $Po(2)$ distribution.

4. $Z, Y_k,\ k = 1, \ldots, d$, are independent $N(0,1)$ variables and the data are generated by
$$X_k = \alpha Y_k + (1-\alpha) Z,\ k = 1, \ldots, d,$$
where α=0.5, and $\mathrm{corr}(X_k, X_l) = 0.5,\ k \neq l,\ k,l = 1, \ldots, d.$

5. Z, Y_k, $k = 1, \ldots, d$, are independent $Po(1)$ variables and the data are generated by
$$X_k = \alpha Y_k + (1-\alpha)Z, \; k = 1, \ldots, d,$$
where α=0.5, and $\mathrm{corr}(X_k, X_l) = 0.5$, $k \neq l$, $k, l = 1, \ldots, d$. (Then after studentizing, the new data are $X'_k = \frac{Y_k+Z}{\sqrt{2}}$, where $Y_k + Z \sim Po(2)$.)

6. Z, Y_k, $k = 1, \ldots, d$, are independent $N(0,1)$ variables and the data come from
$$X_k = \alpha Y_k + (1-\alpha)Z, \; k = 1, \ldots, d,$$
where $\alpha = \frac{3-\sqrt{3}}{2}$, and $\mathrm{corr}(X_k, X_l) = 0.25$, $k \neq l$, $k, l = 1, \ldots, d$.

7. Z, Y_k, $k = 1, \ldots, d$, are independent $Po(\frac{1}{2})$ variables, U_k, $k = 1, \ldots, d$ are independent $Po(1)$ variables, and the data come from
$$X_k = \frac{\alpha}{2}(Y_k + U_k) + (1-\alpha)Z, \; k = 1, \ldots, d,$$
where $\alpha = \frac{2}{3}$, and $\mathrm{corr}(X_k, X_l) = 0.25$, $k \neq l$, $k, l = 1, \ldots, d$. (Then after studentizing, the new data are $X'_k = \frac{Y_k+U_k+Z}{\sqrt{2}}$, where $Y_k + U_k + Z \sim Po(2)$.)

8. X_k, $k = 2, \ldots, d$, are independent $N(0,1)$ variables and $X_1 = X_2$.

9. X_k, $k = 2, \ldots, d$, are independent $Po(2)$ variables and $X_1 = X_2$.

10. Z_0, Z_k, $k = 1, \ldots, d$, are independent $N(0,1)$ variables and the data are generated by
$$X_k = Z_0 Z_k, \; k = 1, \ldots, d,$$
the variables are dependent, but $\mathrm{corr}(X_k, X_l) = 0$, $k \neq l$, $k, l = 1, \ldots, d$.

11. Z_0, Z_{1k}, Z_{2k}, $k = 1, \ldots, d$, are independent $Po(2)$ variables and the data are generated by
$$X_k = Z_0(Z_{1k} - Z_{2k}), \; k = 1, \ldots, d,$$
the variables are dependent, but $\mathrm{corr}(X_k, X_l) = 0$, $k \neq l$, $k, l = 1, \ldots, d$.

When the sample sizes are n=10, 50 and 100, 1000 simulations and, when $n = 200$, 100 simulations are made, for all $d = 2, \ldots, 10$. The values of the test statistics BKR_n, $T1_n$ and $T2_n$ are calculated in the 11 simulated data. Then the values are compared to size α=0.05 critical values of their limiting distributions and the decisions about rejecting H_0 are made. Usually in practice the distributions of the variables are not known, and the limiting distributions of $T1_n$ and $T2_n$ can be approximated by $\widehat{\gamma}\chi^2(\widehat{\beta})$, where $\widehat{\gamma}$ and $\widehat{\beta}$ are derived by using the estimates $\widehat{E1_k}$, $\widehat{E2_k}$ and $\widehat{E3_k}$, see Chapter 4, in calculating the estimates $\widehat{E[T1]}$ and $\widehat{\mathrm{Var}[T1]}$ or $\widehat{E[T2]}$ and $\widehat{\mathrm{Var}[T2]}$, respectively.

Theoretical limits of the expectations and variances of the test statistics $T1_n$ and $T2_n$ and the averages of the test statistics, their estimated expectations and variances from the simulations, for the number of variables $d = 2, \ldots, 10$ and the sample sizes n=10, 50, 100 and 200 are in Tables A.5, A.6 and A.7 in Appendix. In Table A.5 all the variables are $N(0,1)$ distributed, in Table A.6 all the variables are $Po(2)$ distributed and in Table A.7 all the odd numbered variables are $N(0,1)$ distributed and all the even numbered variables are $Po(2)$ distributed. The estimated expectations are close to the corresponding theoretical values, when n is greater than 10. With the increase of the number of variables (dimension of the vector valued variable), one needs more observations to ensure that the estimated expectations are close to the corresponding theoretical values. (Some of the differences for larger d may be caused by the precision in the calculations.) For discrete variables a slightly smaller number of observations is needed as compared to continuous variables. And for the test $T2_n$ more observations are needed than for $T1_n$ to ensure that the estimated expectation is close to the theoretical expectation.

The size of the error caused by the use of the estimated expectation and variance to find the critical values of the test statistic is not large. The same conclusion can be made by studying the empirical sizes and powers (e and t) of the tests statistics in Tables 6.1 to 6.9.

The averages of the test statistics $T1_n$ and $T2_n$ for the dependent data from 4 to 11 are in Appendix in Tables A.8 and A.9, respectively. When n is only 10 the average of the test statistic is smaller than the size 0.05 critical value of the corresponding null distribution (except for data 8 and 9) and therefore the results are not satisfactory. One can also deduce that when the dimension of the vector of variables is low, more observations are needed to be able to make right conclusion about dependence. In data 8 and 9 there is only one dependent pair of variables and the others are independent, therefore with those data more observations are needed when there are more variables. Under H_0 the expectation of the test statistic can't be greater than one and the standard deviation is very small; if the test statistic is greater than one H_0 can be rejected right away. And from Tables A.8 and A.9 it can be seen that the averages of the test statistics are greater than one in many combinations of n and d.

The empirical sizes of the tests (corresponding the data 1, 2 and 3) and the empirical powers of the tests corresponding to different kinds of dependencies in the data from 4 to 11 are in Tables 6.1 to 6.9. For the test statistics $T1_n$ and $T2_n$ there are the empirical sizes and empirical powers corresponding to the critical values from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), the critical values from theoretical $\gamma\chi^2(\beta)$ approximation (t) and the critical values approximated by the bootstrap method (b) for data 4 to 11.

The empirical sizes (and powers) of the $T1_n$ and $T2_n$ tests are similar for continuous and discrete variables and the powers of the tests $T1_n$ and $T2_n$ can be compared for all dependent data. For data 4, 5, 6 and 7, where all the variables are linearly dependent, and for data 8 and 9, where there is only one dependent pair of variables $X_1 = X_2$ and the others are independent, the test $T1_n$ has slightly greater powers than $T2_n$ for the same number of observations for all number of variables d applied in the simulations. But for data 10 and 11, where all the variables are nonmonotonicly dependent, the test $T2_n$ has greater powers than $T1_n$ for the same n for all d applied in the simulations.

The simulation study also shows that if there are more variables, the empirical powers of $T1_n$ and $T2_n$ will tend to 1 with less observations, except for data 8 and 9 of course.

For discrete data the $T1_n$ and $T2_n$ tests can be compared to the BKR_n test only when there are just two variables, d=2, because if $d > 2$ for discrete variables the empirical size of the BKR_n test is not even close to 0.05. Now when d=2 the empirical powers of the three tests are quite close to one another, only for data 10 and 11, nonmonotonic dependence, the test $T1_n$ and even more the test $T2_n$ have greater powers than BKR_n test.

When d >2, only the empirical powers of the continuous, $N(0,1)$, case can be compared and it seems that the powers of $T1_n$ and $T2_n$ tests are small when n=10 and then BKR_n test has greater powers for all data except for nonmonotonic dependence data 10. When n is at least 50, the BKR_n test has sligtly greater empirical powers at least when n and d are smaller, but for data 8, where there is only one dependent pair of variables and the other variables are independent, the $T1_n$ and $T2_n$ tests have greater powers especially when there are more variables, d is larger. For nonmonotonic dependence data 10 the BKR_n test has lowest powers and $T2_n$ test has highest powers for the same n for all d applied in simulations.

Cotterill and Csörgő (1985) recommend the BKR_n test to be used, because the behaviour of its power function against all alternatives is far superior to that of the other nonparametric tests of independence. Now because the empirical powers of the tests $T1_n$ and $T2_n$ already for reasonably small n values are similar to those of the BKR_n test and even greater when there is nonmonotonic dependence, we can recommend the $T1_n$ and the $T2_n$ tests for testing total independence of variables.

TABLE 6.1 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=2, when $x_{0.05}$ is the size 0.05 critical value from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=2		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.037	0.030	0.028	0.027	0.032	0.034	0.042
	50	0.045	0.051	0.042	0.039	0.041	0.045	0.055
	100	0.062	0.050	0.050	0.047	0.051	0.054	0.063
	200	0.05	0.04	0.03	0.03	0.05	0.05	0.07
2	10	0.071	0.055	0.056	0.063	0.041	0.045	0.057
	50	0.063	0.046	0.047	0.055	0.049	0.051	0.053
	100	0.078	0.064	0.061	0.066	0.056	0.059	0.063
	200	0.05	0.05	0.05	0.07	0.04	0.03	0.04
3	10	0.055	0.051	0.044	0.042	0.028	0.043	0.054
	50	0.067	0.045	0.049	0.048	0.040	0.044	0.047
	100	0.056	0.049	0.045	0.044	0.051	0.051	0.056
	200	0.07	0.04	0.04	0.03	0.06	0.07	0.08
4	10	0.256	0.321	0.285	0.267	0.199	0.212	0.250
	50	0.929	0.949	0.948	0.946	0.852	0.850	0.878
	100	0.996	0.998	0.998	0.998	0.989	0.988	0.990
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	0.297	0.314	0.303	0.323	0.219	0.236	0.248
	50	0.935	0.954	0.943	0.948	0.879	0.878	0.888
	100	0.997	0.998	0.998	0.998	0.990	0.990	0.990
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.099	0.100	0.092	0.086	0.073	0.091	0.109
	50	0.351	0.389	0.380	0.367	0.275	0.278	0.308
	100	0.633	0.654	0.652	0.646	0.503	0.497	0.528
	200	0.90	0.92	0.92	0.91	0.84	0.82	0.84
7	10	0.131	0.123	0.122	0.133	0.080	0.094	0.105
	50	0.401	0.396	0.396	0.415	0.286	0.291	0.306
	100	0.648	0.652	0.661	0.681	0.499	0.500	0.513
	200	0.90	0.92	0.92	0.93	0.80	0.81	0.82
8	10	1.000	1.000	1.000	1.000	1.000	0.999	0.999
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	10	0.990	1.000	1.000	1.000	1.000	0.995	0.996
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.067	0.211			0.221		
	50	0.110	0.447			0.634		
	100	0.184	0.611			0.890		
	200	0.35	0.87			0.99		
11	10	0.093	0.185			0.207		
	50	0.117	0.349			0.481		
	100	0.208	0.493			0.768		
	200	0.39	0.70			0.95		

TABLE 6.2 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=3, when $x_{0.05}$ is the size 0.05 critical value from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=3		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.103	0.035	0.028	0.024	0.042	0.039	0.029
	50	0.050	0.064	0.062	0.044	0.044	0.045	0.031
	100	0.040	0.057	0.056	0.050	0.058	0.057	0.045
	200	0.06	0.04	0.06	0.03	0.04	0.05	0.03
2	10	0.183	0.043	0.040	0.042	0.037	0.050	0.041
	50	0.163	0.054	0.055	0.059	0.054	0.060	0.047
	100	0.157	0.053	0.054	0.058	0.052	0.053	0.045
	200	0.22	0.07	0.07	0.07	0.07	0.06	0.06
3	10	0.155	0.044	0.037	0.030	0.045	0.052	0.045
	50	0.070	0.055	0.052	0.047	0.048	0.048	0.047
	100	0.078	0.045	0.048	0.039	0.059	0.064	0.055
	200	0.08	0.05	0.04	0.04	0.04	0.04	0.04
4	10	0.715	0.451	0.410	0.389	0.285	0.285	0.254
	50	1.000	0.994	0.994	0.994	0.973	0.969	0.959
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(0.724)	0.449	0.438	0.450	0.288	0.314	0.272
	50	(0.998)	0.997	0.997	0.997	0.973	0.970	0.964
	100	(1.000)	1.000	1.000	1.000	1.000	0.999	0.999
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.311	0.125	0.110	0.096	0.076	0.087	0.067
	50	0.718	0.579	0.573	0.547	0.379	0.377	0.333
	100	0.948	0.893	0.892	0.882	0.713	0.711	0.672
	200	1.00	0.99	0.99	0.99	0.96	0.96	0.96
7	10	(0.389)	0.144	0.139	0.146	0.086	0.106	0.095
	50	(0.784)	0.588	0.585	0.599	0.421	0.431	0.388
	100	(0.961)	0.894	0.900	0.903	0.743	0.745	0.719
	200	(1.00)	0.99	0.99	1.00	0.97	0.97	0.96
8	10	0.673	0.988	0.930	0.902	0.839	0.800	0.745
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(0.857)	0.980	0.939	0.943	0.855	0.822	0.792
	50	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.128	0.295			0.332		
	50	0.161	0.777			0.954		
	100	0.300	0.966			1.000		
	200	0.77	1.00			1.00		
11	10	(0.201)	0.246			0.261		
	50	(0.358)	0.640			0.859		
	100	(0.635)	0.898			0.994		
	200	(0.99)	1.00			1.00		

TABLE 6.3 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=4, when $x_{0.05}$ is the size 0.05 critical value from estimated $\hat{\gamma}\chi^2(\hat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=4		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.139	0.035	0.026	0.030	0.043	0.041	0.042
	50	0.051	0.052	0.047	0.055	0.050	0.045	0.047
	100	0.059	0.045	0.049	0.051	0.050	0.053	0.054
	200	0.05	0.03	0.03	0.05	0.02	0.02	0.02
2	10	0.505	0.038	0.036	0.036	0.042	0.049	0.045
	50	0.386	0.054	0.052	0.053	0.048	0.050	0.041
	100	0.419	0.056	0.056	0.056	0.048	0.054	0.044
	200	0.38	0.05	0.05	0.05	0.06	0.05	0.05
3	10	0.319	0.035	0.026	0.022	0.039	0.045	0.044
	50	0.195	0.052	0.051	0.047	0.055	0.058	0.054
	100	0.197	0.058	0.056	0.050	0.057	0.062	0.055
	200	0.27	0.09	0.09	0.08	0.09	0.09	0.08
4	10	0.913	0.496	0.445	0.466	0.333	0.339	0.341
	50	1.000	0.996	0.995	0.995	0.984	0.985	0.985
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(0.946)	0.575	0.545	0.546	0.342	0.365	0.354
	50	(1.000)	0.999	0.999	0.999	0.987	0.987	0.985
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.617	0.155	0.134	0.144	0.092	0.103	0.105
	50	0.919	0.696	0.687	0.705	0.417	0.410	0.414
	100	0.997	0.958	0.958	0.961	0.802	0.797	0.798
	200	1.00	1.00	1.00	1.00	0.99	0.99	0.99
7	10	(0.755)	0.157	0.132	0.132	0.088	0.104	0.096
	50	(0.966)	0.688	0.689	0.690	0.503	0.502	0.487
	100	(0.997)	0.960	0.956	0.956	0.852	0.843	0.839
	200	(1.00)	1.00	1.00	1.00	0.99	0.98	0.98
8	10	0.602	0.624	0.531	0.584	0.494	0.501	0.504
	50	0.993	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(0.886)	0.635	0.586	0.586	0.464	0.500	0.470
	50	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.227	0.389			0.421		
	50	0.218	0.963			0.997		
	100	0.463	0.999			1.000		
	200	0.99	1.00			1.00		
11	10	(0.422)	0.277			0.328		
	50	(0.714)	0.889			0.987		
	100	(0.945)	0.994			1.000		
	200	(1.00)	1.00			1.00		

TABLE 6.4 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=5, when $x_{0.05}$ is the size 0.05 critical value from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=5		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.293	0.042	0.030	0.027	0.027	0.026	0.025
	50	0.081	0.055	0.048	0.045	0.041	0.043	0.043
	100	0.076	0.047	0.044	0.042	0.049	0.047	0.046
	200	0.07	0.04	0.05	0.04	0.05	0.06	0.06
2	10	0.954	0.046	0.038	0.041	0.037	0.048	0.041
	50	0.837	0.050	0.054	0.060	0.042	0.052	0.040
	100	0.794	0.054	0.055	0.064	0.061	0.059	0.052
	200	0.73	0.02	0.02	0.02	0.03	0.04	0.03
3	10	0.687	0.042	0.029	0.033	0.042	0.046	0.041
	50	0.292	0.059	0.058	0.061	0.052	0.054	0.048
	100	0.268	0.058	0.058	0.060	0.058	0.051	0.048
	200	0.29	0.03	0.04	0.04	0.06	0.07	0.05
4	10	0.971	0.558	0.492	0.482	0.341	0.323	0.321
	50	1.000	1.000	1.000	1.000	0.989	0.988	0.988
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(1.000)	0.579	0.539	0.564	0.349	0.370	0.348
	50	(1.000)	1.000	1.000	1.000	0.993	0.993	0.993
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.822	0.161	0.123	0.119	0.080	0.087	0.086
	50	0.978	0.796	0.786	0.783	0.519	0.511	0.508
	100	1.000	0.982	0.981	0.981	0.860	0.856	0.853
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	10	(0.989)	0.197	0.168	0.182	0.093	0.113	0.100
	50	(0.997)	0.779	0.780	0.784	0.551	0.564	0.536
	100	(1.000)	0.975	0.975	0.976	0.889	0.889	0.875
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
8	10	0.637	0.406	0.310	0.302	0.263	0.267	0.263
	50	0.864	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(0.996)	0.408	0.356	0.386	0.284	0.319	0.301
	50	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.319	0.499			0.575		
	50	0.279	0.995			1.000		
	100	0.632	1.000			1.000		
	200	1.00	1.00			1.00		
11	10	(0.821)	0.376			0.461		
	50	(0.954)	0.982			1.000		
	100	(0.999)	1.000			1.000		
	200	(1.00)	1.00			1.00		

TABLE 6.5 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=6, when $x_{0.05}$ is the size 0.05 critical value from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=6		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.484	0.042	0.023	0.024	0.034	0.044	0.043
	50	0.113	0.054	0.052	0.053	0.053	0.053	0.047
	100	0.103	0.044	0.044	0.045	0.056	0.054	0.048
	200	0.12	0.09	0.09	0.09	0.08	0.08	0.07
2	10	1.000	0.052	0.035	0.033	0.033	0.044	0.048
	50	0.997	0.055	0.055	0.055	0.052	0.054	0.061
	100	0.999	0.043	0.048	0.046	0.048	0.051	0.054
	200	0.99	0.02	0.02	0.02	0.02	0.02	0.03
3	10	0.988	0.036	0.027	0.026	0.050	0.054	0.059
	50	0.738	0.055	0.055	0.055	0.048	0.049	0.058
	100	0.644	0.061	0.059	0.059	0.051	0.052	0.058
	200	0.60	0.03	0.03	0.03	0.05	0.06	0.06
4	10	0.994	0.573	0.505	0.510	0.341	0.339	0.325
	50	1.000	1.000	1.000	1.000	0.995	0.995	0.994
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(1.000)	0.601	0.556	0.550	0.356	0.375	0.384
	50	(1.000)	1.000	1.000	1.000	0.998	0.997	0.998
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.926	0.177	0.132	0.135	0.073	0.075	0.071
	50	0.997	0.827	0.816	0.819	0.522	0.515	0.503
	100	1.000	0.995	0.995	0.995	0.877	0.873	0.864
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	10	(1.000)	0.211	0.176	0.174	0.103	0.116	0.118
	50	(1.000)	0.832	0.828	0.823	0.582	0.595	0.603
	100	(1.000)	0.992	0.992	0.991	0.933	0.934	0.937
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
8	10	0.743	0.279	0.198	0.204	0.177	0.171	0.162
	50	0.758	1.000	1.000	1.000	0.999	0.999	0.999
	100	0.945	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(1.000)	0.302	0.247	0.239	0.189	0.221	0.234
	50	(1.000)	1.000	0.999	0.999	0.998	0.997	0.997
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.477	0.586			0.669		
	50	0.366	1.000			1.000		
	100	0.782	1.000			1.000		
	200	1.00	1.00			1.00		
11	10	(0.975)	0.487			0.573		
	50	(0.992)	0.999			1.000		
	100	(1.000)	1.000			1.000		
	200	(1.00)	1.00			1.00		

TABLE 6.6 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=7, when $x_{0.05}$ is the size 0.05 critical value from estimated $\hat{\gamma}\chi^2(\hat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=7		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.592	0.054	0.030	0.032	0.048	0.049	0.053
	50	0.159	0.054	0.051	0.051	0.049	0.051	0.061
	100	0.098	0.057	0.053	0.053	0.049	0.046	0.054
	200	0.07	0.07	0.07	0.07	0.07	0.08	0.08
2	10	1.000	0.037	0.029	0.019	0.036	0.042	0.030
	50	1.000	0.043	0.047	0.037	0.054	0.056	0.042
	100	1.000	0.043	0.044	0.030	0.050	0.052	0.039
	200	1.00	0.03	0.03	0.02	0.06	0.05	0.04
3	10	1.000	0.047	0.025	0.015	0.039	0.044	0.037
	50	0.880	0.042	0.039	0.023	0.053	0.053	0.041
	100	0.825	0.050	0.047	0.031	0.045	0.041	0.030
	200	0.77	0.08	0.08	0.05	0.05	0.06	0.04
4	10	1.000	0.563	0.486	0.489	0.316	0.316	0.319
	50	1.000	1.000	1.000	1.000	0.997	0.996	0.996
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(1.000)	0.620	0.563	0.542	0.366	0.392	0.370
	50	(1.000)	1.000	1.000	1.000	0.998	0.997	0.997
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.978	0.190	0.134	0.136	0.096	0.095	0.099
	50	0.998	0.859	0.851	0.852	0.541	0.536	0.556
	100	1.000	0.996	0.996	0.996	0.880	0.878	0.884
	200	0.99	1.00	1.00	1.00	1.00	1.00	1.00
7	10	(1.000)	0.202	0.164	0.137	0.106	0.116	0.103
	50	(1.000)	0.865	0.859	0.843	0.619	0.619	0.600
	100	(1.000)	0.997	0.997	0.996	0.935	0.940	0.927
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
8	10	0.778	0.231	0.162	0.164	0.124	0.129	0.137
	50	0.640	0.999	0.998	0.998	0.967	0.968	0.969
	100	0.846	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(1.000)	0.231	0.169	0.133	0.146	0.168	0.144
	50	(1.000)	0.996	0.996	0.996	0.980	0.978	0.971
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.608	0.641			0.749		
	50	0.417	1.000			1.000		
	100	0.822	1.000			1.000		
	200	1.00	1.00			1.00		
11	10	(0.997)	0.545			0.664		
	50	(1.000)	1.000			1.000		
	100	(1.000)	1.000			1.000		
	200	(1.00)	1.00			1.00		

TABLE 6.7 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=8, when $x_{0.05}$ is the size 0.05 critical value from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=8		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.697	0.044	0.017	0.012	0.033	0.037	0.035
	50	0.220	0.044	0.043	0.027	0.050	0.049	0.047
	100	0.140	0.057	0.056	0.038	0.037	0.040	0.036
	200	0.11	0.02	0.02	0.02	0.02	0.02	0.02
2	10	1.000	0.054	0.038	0.030	0.047	0.055	0.046
	50	1.000	0.045	0.046	0.039	0.052	0.060	0.043
	100	1.000	0.050	0.051	0.042	0.066	0.065	0.060
	200	1.00	0.05	0.05	0.05	0.08	0.07	0.07
3	10	1.000	0.043	0.027	0.023	0.036	0.042	0.042
	50	1.000	0.042	0.040	0.039	0.054	0.056	0.055
	100	0.999	0.051	0.050	0.048	0.055	0.055	0.052
	200	1.00	0.11	0.11	0.11	0.09	0.08	0.07
4	10	0.999	0.626	0.541	0.501	0.292	0.295	0.286
	50	1.000	1.000	1.000	1.000	0.990	0.990	0.989
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(1.000)	0.645	0.585	0.565	0.333	0.360	0.340
	50	(1.000)	1.000	1.000	1.000	0.998	0.998	0.998
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.993	0.198	0.132	0.114	0.086	0.094	0.088
	50	0.999	0.873	0.870	0.848	0.517	0.516	0.508
	100	1.000	0.999	0.999	0.995	0.872	0.872	0.866
	200	1.00	1.00	1.00	1.00	0.99	0.99	0.99
7	10	(1.000)	0.217	0.159	0.150	0.091	0.101	0.093
	50	(1.000)	0.882	0.873	0.859	0.617	0.627	0.599
	100	(1.000)	0.997	0.997	0.995	0.949	0.948	0.940
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
8	10	0.829	0.205	0.122	0.087	0.104	0.107	0.102
	50	0.623	0.991	0.987	0.980	0.911	0.908	0.898
	100	0.740	1.000	1.000	1.000	1.000	1.000	1.000
	200	0.96	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(1.000)	0.180	0.132	0.114	0.111	0.127	0.114
	50	(1.000)	0.989	0.987	0.980	0.910	0.899	0.887
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.651	0.740			0.817		
	50	0.488	1.000			1.000		
	100	0.845	1.000			1.000		
	200	1.00	1.00			1.00		
11	10	(1.000)	0.611			0.705		
	50	(1.000)	1.000			1.000		
	100	(1.000)	1.000			1.000		
	200	(1.00)	1.00			1.00		

TABLE 6.8 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=9, when $x_{0.05}$ is the size 0.05 critical value from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=9		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.768	0.048	0.024	0.022	0.042	0.046	0.047
	50	0.285	0.050	0.046	0.042	0.048	0.054	0.059
	100	0.186	0.051	0.046	0.044	0.057	0.058	0.062
	200	0.18	0.06	0.05	0.03	0.07	0.07	0.07
2	10	1.000	0.052	0.027	0.026	0.036	0.039	0.032
	50	1.000	0.057	0.054	0.051	0.049	0.049	0.037
	100	1.000	0.059	0.058	0.052	0.050	0.055	0.041
	200	1.00	0.03	0.03	0.03	0.03	0.04	0.02
3	10	1.000	0.037	0.018	0.015	0.019	0.024	0.032
	50	1.000	0.045	0.042	0.033	0.031	0.030	0.037
	100	1.000	0.061	0.060	0.053	0.062	0.063	0.077
	200	1.00	0.06	0.05	0.05	0.03	0.03	0.04
4	10	1.000	0.611	0.520	0.506	0.299	0.310	0.323
	50	1.000	1.000	1.000	1.000	0.997	0.996	0.997
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(1.000)	0.659	0.604	0.595	0.340	0.368	0.331
	50	(1.000)	1.000	1.000	1.000	0.998	0.998	0.997
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.994	0.190	0.128	0.119	0.068	0.073	0.081
	50	1.000	0.904	0.891	0.885	0.483	0.483	0.499
	100	1.000	0.998	0.997	0.997	0.895	0.892	0.901
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	10	(1.000)	0.237	0.188	0.183	0.090	0.106	0.083
	50	(1.000)	0.898	0.896	0.889	0.639	0.647	0.607
	100	(1.000)	0.998	0.998	0.998	0.954	0.951	0.943
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
8	10	0.883	0.145	0.075	0.068	0.070	0.075	0.083
	50	0.626	0.964	0.959	0.952	0.747	0.747	0.761
	100	0.678	1.000	1.000	1.000	1.000	1.000	1.000
	200	0.86	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(1.000)	0.142	0.095	0.091	0.086	0.102	0.081
	50	(1.000)	0.951	0.940	0.936	0.761	0.769	0.713
	100	(1.000)	1.000	1.000	1.000	0.998	0.998	0.998
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.721	0.782			0.827		
	50	0.541	1.000			1.000		
	100	0.850	1.000			1.000		
	200	1.00	1.00			1.00		
11	10	(1.000)	0.668			0.768		
	50	(1.000)	1.000			1.000		
	100	(1.000)	1.000			1.000		
	200	(1.00)	1.00			1.00		

TABLE 6.9 Probabilities $\Pr\{T_n \in C\} = \Pr\{T_n > x_{0.05}\}$ for d=10, when $x_{0.05}$ is the size 0.05 critical value from estimated $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation (e), from theoretical $\gamma\chi^2(\beta)$ approximation (t) and approximated by bootstrap method (b).

d=10		BKR_n	$T1_n$			$T2_n$		
Data	n		e	t	b	e	t	b
1	10	0.778	0.033	0.010	0.010	0.017	0.021	0.025
	50	0.330	0.052	0.046	0.047	0.055	0.055	0.062
	100	0.239	0.048	0.041	0.041	0.054	0.054	0.056
	200	0.16	0.03	0.03	0.03	0.05	0.05	0.05
2	10	1.000	0.057	0.036	0.036	0.025	0.038	0.031
	50	1.000	0.058	0.054	0.056	0.055	0.060	0.048
	100	1.000	0.066	0.064	0.065	0.071	0.076	0.053
	200	1.00	0.07	0.06	0.07	0.05	0.05	0.03
3	10	1.000	0.053	0.027	0.027	0.021	0.024	0.021
	50	1.000	0.057	0.051	0.047	0.052	0.052	0.048
	100	1.000	0.052	0.050	0.046	0.055	0.057	0.052
	200	1.00	0.09	0.09	0.09	0.06	0.06	0.05
4	10	1.000	0.631	0.552	0.553	0.277	0.289	0.294
	50	1.000	1.000	1.000	1.000	0.996	0.996	0.996
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	10	(1.000)	0.620	0.563	0.566	0.374	0.405	0.379
	50	(1.000)	1.000	1.000	1.000	0.996	0.997	0.996
	100	(1.000)	1.000	1.000	1.000	1.000	1.000	1.000
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
6	10	0.998	0.204	0.136	0.136	0.060	0.070	0.071
	50	1.000	0.899	0.886	0.887	0.506	0.509	0.521
	100	1.000	0.998	0.998	0.998	0.875	0.875	0.878
	200	1.00	1.00	1.00	1.00	0.99	0.99	0.99
7	10	(1.000)	0.249	0.187	0.191	0.096	0.112	0.104
	50	(1.000)	0.905	0.899	0.900	0.604	0.613	0.586
	100	(1.000)	0.999	0.999	0.999	0.941	0.946	0.933
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
8	10	0.878	0.134	0.070	0.070	0.064	0.071	0.072
	50	0.616	0.914	0.902	0.902	0.611	0.608	0.621
	100	0.636	1.000	1.000	1.000	0.988	0.988	0.988
	200	0.82	1.00	1.00	1.00	1.00	1.00	1.00
9	10	(1.000)	0.148	0.100	0.104	0.071	0.080	0.072
	50	(1.000)	0.890	0.884	0.888	0.615	0.623	0.581
	100	(1.000)	1.000	0.999	0.999	0.987	0.986	0.985
	200	(1.00)	1.00	1.00	1.00	1.00	1.00	1.00
10	10	0.781	0.833			0.868		
	50	0.557	1.000			1.000		
	100	0.824	1.000			1.000		
	200	1.00	1.00			1.00		
11	10	(1.000)	0.759			0.788		
	50	(1.000)	1.000			1.000		
	100	(1.000)	1.000			1.000		
	200	(1.00)	1.00			1.00		

6.4 Applications

Attitude, activity and chronic diseases of aging people. The first data consist of discrete variables and originate in the studies of physical exercise and health of retired women descripted in Heikkinen, Lahtinen and Lehtoranta (1972). The data consist of 254 observations of the three variables:

NCD	=	Number of chronic diseases, 1=some, 2=none
$POSA$	=	Participation in official social activities, 1=participates, 2=doesn't participate
PA	=	Positive attitude in life, 1=doesn't have positive attitude, 2=has positive attitude.

The $2 \times 2 \times 2$ contingency table of the observations is

	NCD			
	1		2	
	POSA			
PA	1	2	1	2
1	58	12	9	2
2	96	49	18	10

Using the asymptotic theory of log-linear models and applying the GLIM–program (Aitkin, Anderson, Francis and Hinde, 1989; McCullagh and Nelder, 1989), we obtain the results in Tables 6.10 and 6.11.

TABLE 6.10 The log-linear models and the corresponding P–values of the variables *NCD*, *POSA* and *PA*.

Model	Deviance	df	P–value
PA*NCD*POSA	0	0	1.0000
PA*NCD+PA*POSA+NCD*POSA	0.0002	1	0.9887
PA*NCD+PA*POSA	0.0454	2	0.9776
PA*NCD+NCD*POSA	8.0420	2	0.0179
PA*POSA+NCD*POSA	0.2471	2	0.8838
NCD+PA*POSA	0.3387	3	0.9526
POSA+PA*NCD	8.1336	3	0.0433
PA+NCD*POSA	8.3353	3	0.0396
PA+NCD+POSA	8.3353	4	0.0771

For other graphical models P=0.0000

TABLE 6.11 Statistical significance of the interaction terms formed among the set of variables *NCD*, *POSA* and *PA*.

Term	Deviance	df	P–value
PA.NCD.POSA	0.0002	1	0.9887
PA.NCD	0.2933	1	0.5581
PA.POSA	8.0882	1	0.0045
NCD.POSA	0.0916	1	0.7622

The parsimonius statistically sufficient model is the independence model $PA{+}NCD{+}POSA$ (scaled deviance=8.4269, df=4, P=0.0771). However, the pairwise interaction term $PA.POSA$ is statistically significant (scaled deviance=8.0882, df=1, P=0.0045). Hence the model $PA + NCD + POSA + PA.POSA$ is chosen, see Tables 6.10 and 6.11. If only the independence model $PA{+}NCD{+}POSA$ is estimated (which is not according to the principles of model–building) one would accept the hypothesis of total independence using the 5% significance level! The right conclusion is drawn from the fact that there is one significant interaction term.

Next the tests $T1_n$ and $T2_n$ are applied for the hypothesis of total independence which is the model $PA{+}NCD{+}POSA$ in terms of log–linear models. The results are in Table 6.12.

TABLE 6.12 The values of the test statistics $T1_n$ and $T2_n$ and the corresponding P–values in testing the independence of the variables *NCD*, *POSA* and *PA*.

Variables	$T1_n$	P–value	$T2_n$	P–value
PA, NCD, POSA	0.4322	0.0370	0.8785	0.0216
PA, NCD	0.0054	0.7609	0.0088	0.7609
PA, POSA	0.5226	0.0058	1.1120	0.0058
NCD, POSA	0.0171	0.5915	0.0286	0.5915

Our both tests reveal immediately that the three variables PA, NCD, $POSA$ are not totally independent. And when independence is tested in pairs it is found that PA and NCD are independent, NCD and $POSA$ are independent, but PA and $POSA$ are dependent, see Table 6.12. The results for pairwise interactions agree with the ones given by the deviance test statistic applied in the log–linear modelling. On the other hand there is a contradiction for the test of complete independence of three variables.

Are (varimax–rotated) estimated factor scores independent? In the second example, our data consist of estimated factors, which are obtained by varimax rotation from the actual data. The corresponding theoretical factors are orthogonal (totally independent), but as one sees their estimates are dependent.

Data: The Finnish ITPA data (Kuusinen and Leskinen, 1986), which is used in the studies about the relations between intellectual abilities and school achievement in longitudinal data. The pupils on 6th and 9th grade at the comprehensive schools in Jyväskylä (Kuusinen and Leskinen, 1988).

Three blocks of estimated factor scores (varimax-rotated) are studied. The three factors in the first block are derived from the Finnish ITPA data. The estimated factors are based on 215 obsevations and they are labelled

$$\begin{aligned} VER &= \text{verbal} \\ VIS &= \text{visual} \\ EXP &= \text{expression.} \end{aligned}$$

Their correlation matrix is

$$\begin{pmatrix} 1.000 & & \\ 0.2061 & 1.000 & \\ 0.2388 & 0.0838 & 1.000 \end{pmatrix}.$$

The estimated three factors in the second block are based on 215 observations of 6th grade pupils at the comprehensive schools in Jyväskylä and they are called

$$\begin{aligned} LING6 &= \text{linguistic factor on 6th grade} \\ ARSC6 &= \text{arts and sciences factor on 6th grade} \\ MATH6 &= \text{mathematics factor on 6th grade,} \end{aligned}$$

and their correlation matrix is

$$\begin{pmatrix} 1.000 & & \\ 0.2399 & 1.000 & \\ 0.2351 & 0.3818 & 1.000 \end{pmatrix}.$$

The estimated three factors in the third block are based on 215 observations of 9th grade pupils at the comprehensive schools in Jyväskylä and they are interpreted as

$$\begin{aligned} LING9 &= \text{linguistic factor on 9th grade} \\ ARSC9 &= \text{arts and sciences factor on 9th grade} \\ MATH9 &= \text{mathematics factor on 9th grade.} \end{aligned}$$

The correlation matrix is

$$\begin{pmatrix} 1.000 & & \\ 0.0644 & 1.000 & \\ 0.1711 & 0.1799 & 1.000 \end{pmatrix}.$$

The correlation matrixes are not identity matrixes, see the criticism about the use of the varimax method in Leskinen and Kuusinen (1991).

The results of the tests $T1_n$ and $T2_n$, also $T1_n^{(F_n)}$ and $T2_n^{(F_n)}$, see Chapter 5.5, concerning the independence of the variables within the blocks are in Table 6.13.

TABLE 6.13 The values of the test statistics $T1_n$ and $T2_n$, also $T1_n^{(F_n)}$ and $T2_n^{(F_n)}$, and the corresponding P–values in testing the independence of the factors VER, VIS, EXP and $LING6$, $ARSC6$, $MATH6$ and $LING9$, $ARSC9$, $MATH9$.

Variables	data	$T1_n$	P–value	$T2_n$	P–value
VER, VIS, EXP	original	0.7049	0.0000	0.9258	0.0002
	F_n	0.8825	0.0001	1.2364	0.0001
VER, VIS	original	0.3256	0.0059	0.4899	0.0158
	F_n	0.3309	0.0145	0.5173	0.0348
VER, EXP	original	0.5578	0.0000	0.8656	0.0000
	F_n	0.7568	0.0001	1.2470	0.0001
VIS, EXP	original	0.1304	0.1255	0.2962	0.1328
	F_n	0.1291	0.1648	0.3104	0.1782
LING6, ARSC6, MATH6	original	1.7050	0.0000	2.0627	0.0000
	F_n	2.2822	0.0000	2.7808	0.0000
LING6, ARSC6	original	0.6555	0.0000	1.1555	0.0000
	F_n	0.8811	0.0000	1.5124	0.0000
LING6, MATH6	original	0.4982	0.0004	0.8133	0.0004
	F_n	0.8070	0.0006	1.0735	0.0003
ARSC6, MATH6	original	1.3306	0.0000	1.7871	0.0000
	F_n	1.7395	0.0000	2.5742	0.0000
LING9, ARSC9, MATH9	original	0.9323	0.0000	1.6455	0.0000
	F_n	1.0809	0.0000	1.7311	0.0000
LING9, ARSC9	original	0.0962	0.2303	0.4224	0.0405
	F_n	0.1537	0.1218	0.4602	0.0557
LING9, MATH9	original	0.3789	0.0028	0.6140	0.0039
	F_n	0.4532	0.0035	0.7273	0.0060
ARSC9, MATH9	original	0.5726	0.0003	1.3149	0.0000
	F_n	0.6878	0.0002	1.4597	0.0000

In the first block variables are not totally independent and further studies show that variables VER, VIS are dependent, variables VER, EXP are dependent, but variables VIS, EXP are independent. In the second block all variables are dependent also in pairs. And in the third block variables $LING9, ARSC9$ are independent according to $T1_n$, $T1_n^{(F_n)}$ and $T2_n^{(F_n)}$ tests and dependent according to $T2_n$ test. A possible explanation for the contradiction is that there is a slight nonlinear dependence between those variables, the variable pairs $LING9, MATH9$ and $ARSC9, MATH9$ are dependent.

7 RECOMMENDATIONS AND REMARKS

We have proved that if condition (3.2) holds, the test based on the statistic

$$T_n = n \int_{\mathbf{R}^d} |c_n(\mathbf{t}) - \prod_{k=1}^{d} c_{nk}(t_k)|^2 g(\mathbf{t}) d\mathbf{t},$$

is consistent against all aternatives, when $g(\mathbf{t})$ is a weight function introduced in (4.4) satisfying the conditions 1 to 5 in Chapter 4. The asymptotic distribution of the test statistic can be approximated by a normal distribution under H_1 and under H_0 it is asymptotically distributed as a weighted sum of independent $\chi^2(1)$ variables. We have also shown that if the variables in data have analytic characteristic function, one may use a weight function introduced in (4.3) satisfying the conditions 1 to 5 in Chapter 4. Because the data is studentized, the test T_n is location and scale invariant.

The simulation experiments show that the empirical powers of both tests $T1_n$ and $T2_n$ converge to one quite rapidly. And because there isn't much difference between the empirical powers corresponding to the data consisting of continuous variables or on the other hand discrete variables, we can recommend the test T_n for data consisting of arbitrary variables. The sample size n must be at least 50, and if one wants to be sure that even slight dependencies in the data are detected, then more observations are needed.

There is no restriction for the number of variables independence of which is tested simultaneously, but when the number of variables d increases, the expectation of the test statistic converges to one and the variance of the test statistic converges to zero, and there will be difficulties with the precision in calculations.

The calculation of the test statistic and the estimates of its expectation and variance requires a considerable amount of CPU time, especially when the sample size n increases, so it might be useful to try to find some approximation of the test statistic, which would need less time for calculations, if one is often dealing with large data. Now sometimes time can be saved by using the knowledge that under H_0 the expectation of the test statistic is always smaller than one ($\int_{\mathbf{R}^d} g(\mathbf{t})d\mathbf{t}$) and the standard deviation is very small, so the H_0 hypothesis can be rejected immediately if the value of the test statistic is greater than one, at least when d is not very large.

In spite of the time consuming calculations we recommend the use of the approximation $\widehat{\gamma}\chi^2(\widehat{\beta})$ for the limiting distribution of the test statistic under H_0, because then tables of critical values for different numbers of variables d, weight functions and distributions of the variables are not needed. On the other hand the bootstrap method is also an asymptotic method and one would not benefit much by using it for small numbers of observations n. Also the calculation of the critical values of the test statistics by bootstrap method requires a lot more CPU time, than the calculation of the critical values by the $\widehat{\gamma}\chi^2(\widehat{\beta})$ approximation.

The data can also be transformed, so that the values of the empirical distribution function are used instead of the original data. Then the uniform distribution is used in deriving tables of critical values for the test statistic, with different numbers of variables d and some different weight functions. Then the calculation of the expectations and variances of the test statistic is not needed every time the test is used, but more obsevations are needed than with the original data.

A APPENDIX

For the test statistic $T1_n$ we have $E1 = 0.7510$, $E2 = 0.5834$ and $E3 = 0.6444$ for $Po(2)$ distribution and by numerical integration we get $E1 = 0.7351$, $E2 = 0.5503$ and $E3 = 0.6158$ for $U(0,1)$ distribution and $E1 = 0.7468$, $E2 = 0.5761$ and $E3 = 0.6367$ for $N(0,1)$ distribution. Now the limits of expectation and variance of the test statistic are calculated and then the size α=0.05, 0.01 and 0.001 critical values of the asymptotic distribution of $T1_n$ under H_0 are approximated using $T1 \sim \gamma\chi^2(\beta)$, where γ and β are set to match the limits of expectation and variance of $T1_n$.

For test statistic $T2_n$ we have $E1 = 0.5858$, $E2 = 0.3633$ and $E3 = 0.4571$ for $Po(2)$ distribution and by numerical integration we get $E1 = 0.5548$, $E2 = 0.3175$ and $E3 = 0.4268$ for $U(0,1)$ distribution and $E1 = 0.5774$, $E2 = 0.3536$ and $E3 = 0.4472$ for $N(0,1)$ distribution. Using these values the limits of expectation and variance of $T2_n$ are calculated and the limiting distribution of $T2_n$ under H_0 is approximated by $\gamma\chi^2(\beta)$.

The approximative size α=0.05, 0.01 and 0.001 critical values of the limiting distribution under H_0 are also calculated by the bootstrap method from 1000 resamples of the original sample, where sample size n=200.

The γ and β values and the limits of expectations of the test statistics $T1_n$ and $T2_n$ and also the approximated size α=0.05, 0.01 and 0.001 critical values of their asymptotic distributions under H_0 are in Table A.1, when all the variables are from $U(0,1)$ distribution (modified test), in Table A.2, when all the variables are from $N(0,1)$ distribution (data 1), in Table A.3, when all the variables are from $Po(2)$ distribution (data 2) and in Table A.4, when all the odd numbered variables are from $N(0,1)$ distribution and all the even numbered variables are from $Po(2)$ distribution (data 3).

We made 1000 simulations when the sample size n=10, 50 and 100 and 100 simulations when $n = 200$, for all numbers of variables $d = 2, \ldots, 10$, and calculated the values of the test statistics $T1_n$, $T2_n$. The averages of the test statistics, their estimated expectations and variances from the simulations and also the theoretical expectations and variances are in Tables A.5, A.6 and A.7. The bolded averages differ significantly from the corresponding theoretical values. And the averages of the test statistics for the simulations of the dependent data from 4 to 11 are in Tables A.8 and A.9 . The bolded averages are smaller than the corresponding size 0.05 critical values.

TABLE A.1 The limits of expectations, γ and β values of $T1_n$ and $T2_n$ and the approximated size α=0.05, 0.01 and 0.001 critical values of their asymptotic distributions under H_0 approximated by $\gamma\chi^2_\alpha(\beta)$ and by bootstrap method, when all the variables are $U(0,1)$ distributed.

$N(0,1)$ distribution

$T1_n$	d	Method	γ	β	$E[T1]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0450	1.6	0.0699	0.2292	0.3666	0.5665
		Bootstrap			0.0628	0.1920	0.3563	0.4740
	3	$\gamma\chi^2(\beta)$	0.0318	5.4	0.1728	0.3731	0.5036	0.6792
		Bootstrap			0.1765	0.3988	0.5553	0.8389
	4	$\gamma\chi^2(\beta)$	0.0223	12.8	0.2862	0.4938	0.6120	0.7640
		Bootstrap			0.2904	0.4909	0.6855	0.9382
	5	$\gamma\chi^2(\beta)$	0.0155	25.6	0.3975	0.5963	0.7007	0.8308
		Bootstrap			0.3964	0.5931	0.6966	0.8971
	6	$\gamma\chi^2(\beta)$	0.0108	46.4	0.4999	0.6820	0.7725	0.8827
		Bootstrap			0.5055	0.6987	0.7883	0.9064
	7	$\gamma\chi^2(\beta)$	0.0074	79.4	0.5902	0.7523	0.8297	0.9224
		Bootstrap			0.5931	0.7494	0.8508	0.9615
	8	$\gamma\chi^2(\beta)$	0.0051	131.1	0.6678	0.8089	0.8745	0.9521
		Bootstrap			0.6686	0.8139	0.8953	1.0364
	9	$\gamma\chi^2(\beta)$	0.0035	211.3	0.7329	0.8540	0.9090	0.9733
		Bootstrap			0.7276	0.8499	0.8975	0.9913
	10	$\gamma\chi^2(\beta)$	0.0023	334.9	0.7869	0.8895	0.9352	0.9883
		Bootstrap			0.7866	0.8956	0.9271	0.9448
$T2_n$	d	Method	γ	β	$E[T2]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0501	4.0	0.1982	0.4716	0.6608	0.9200
		Bootstrap			0.1941	0.4709	0.7375	1.0421
	3	$\gamma\chi^2(\beta)$	0.0258	16.2	0.4181	0.6853	0.8330	1.0207
		Bootstrap			0.4067	0.6615	0.8744	1.4207
	4	$\gamma\chi^2(\beta)$	0.0130	46.1	0.6012	0.8208	0.9300	1.0630
		Bootstrap			0.5921	0.8186	1.0009	1.1856
	5	$\gamma\chi^2(\beta)$	0.0065	113.9	0.7365	0.9040	0.9824	1.0755
		Bootstrap			0.7283	0.8998	0.9950	1.1063
	6	$\gamma\chi^2(\beta)$	0.0031	264.0	0.8304	0.9528	1.0078	1.0719
		Bootstrap			0.8303	0.9663	1.0326	1.1999
	7	$\gamma\chi^2(\beta)$	0.0015	593.8	0.8929	0.9799	1.0179	1.0617
		Bootstrap			0.8918	0.9884	1.0395	1.0814
	8	$\gamma\chi^2(\beta)$	0.0007	1320.1	0.9334	0.9940	1.0200	1.0497
		Bootstrap			0.9328	0.9963	1.0324	1.0678
	9	$\gamma\chi^2(\beta)$	0.0003	2928.8	0.9591	1.0006	1.0183	1.0384
		Bootstrap			0.9580	0.9996	1.0188	1.0647
	10	$\gamma\chi^2(\beta)$	0.0001	6516.6	0.9751	1.0033	1.0152	1.0287
		Bootstrap			0.9740	1.0049	1.0179	1.0604

TABLE A.2 The limits of expectations, γ and β values of $T1_n$ and $T2_n$ and the approximated size α=0.05, 0.01 and 0.001 critical values of their asymptotic distributions under H_0 approximated by $\gamma\chi^2_\alpha(\beta)$ and by bootstrap method, when all the variables are from $N(0,1)$ distribution.

Po(2) distribution

$T1_n$	d	Method	γ	β	$E[T1]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0277	2.3	0.0641	0.1825	0.2747	0.4054
		Bootstrap			0.0648	0.1871	0.2996	0.4313
	3	$\gamma\chi^2(\beta)$	0.0208	7.7	0.1598	0.3133	0.4077	0.5322
		Bootstrap			0.1638	0.3245	0.4634	0.6626
	4	$\gamma\chi^2(\beta)$	0.0155	17.2	0.2671	0.4324	0.5232	0.6382
		Bootstrap			0.2651	0.4231	0.5207	0.6354
	5	$\gamma\chi^2(\beta)$	0.0115	32.5	0.3739	0.5382	0.6225	0.7265
		Bootstrap			0.3739	0.5413	0.6444	0.7457
	6	$\gamma\chi^2(\beta)$	0.0084	56.2	0.4736	0.6295	0.7059	0.7982
		Bootstrap			0.4753	0.6279	0.7002	0.8442
	7	$\gamma\chi^2(\beta)$	0.0061	91.6	0.5629	0.7064	0.7743	0.8553
		Bootstrap			0.5628	0.7055	0.7750	0.9157
	8	$\gamma\chi^2(\beta)$	0.0044	144.2	0.6408	0.7698	0.8293	0.8997
		Bootstrap			0.6375	0.7849	0.8325	0.9227
	9	$\gamma\chi^2(\beta)$	0.0032	221.5	0.7072	0.8213	0.8729	0.9333
		Bootstrap			0.7094	0.8256	0.8952	0.9595
	10	$\gamma\chi^2(\beta)$	0.0023	335.0	0.7631	0.8625	0.9069	0.9584
		Bootstrap			0.7635	0.8622	0.9093	0.9907
$T2_n$	d	Method	γ	β	$E[T2]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0302	5.9	0.1786	0.3764	0.5033	0.6733
		Bootstrap			0.1784	0.3563	0.4827	0.7719
	3	$\gamma\chi^2(\beta)$	0.0172	22.4	0.3849	0.5917	0.7019	0.8399
		Bootstrap			0.3878	0.6150	0.7274	0.9607
	4	$\gamma\chi^2(\beta)$	0.0096	58.8	0.5635	0.7446	0.8330	0.9397
		Bootstrap			0.5646	0.7433	0.8693	1.0907
	5	$\gamma\chi^2(\beta)$	0.0052	134.5	0.7010	0.8473	0.9151	0.9954
		Bootstrap			0.6963	0.8483	0.9373	1.0693
	6	$\gamma\chi^2(\beta)$	0.0028	288.5	0.8003	0.9129	0.9634	1.0222
		Bootstrap			0.7993	0.9172	0.9662	1.0364
	7	$\gamma\chi^2(\beta)$	0.0014	600.8	0.8690	0.9531	0.9899	1.0323
		Bootstrap			0.8647	0.9502	0.9844	1.0257
	8	$\gamma\chi^2(\beta)$	0.0007	1237.1	0.9154	0.9767	1.0031	1.0333
		Bootstrap			0.9115	0.9782	1.0041	1.0472
	9	$\gamma\chi^2(\beta)$	0.0004	2543.4	0.9459	0.9900	1.0087	1.0300
		Bootstrap			0.9452	0.9884	1.0076	1.0513
	10	$\gamma\chi^2(\beta)$	0.0002	5249.4	0.9658	0.9970	1.0102	1.0251
		Bootstrap			0.9655	0.9963	1.0109	1.0280

TABLE A.3 The limits of expectations, γ and β values of $T1_n$ and $T2_n$ and the approximated size α=0.05, 0.01 and 0.001 critical values of their asymptotic distributions under H_0 approximated by $\gamma\chi^2_\alpha(\beta)$ and by bootstrap method, when all the variables are from $Po(2)$ distribution.

				$Po(2)$ distribution				
$T1_n$	d	Method	γ	β	$E[T1]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0280	2.2	0.0620	0.1793	0.2715	0.4026
		Bootstrap			0.0612	0.1722	0.2791	0.5070
	3	$\gamma\chi^2(\beta)$	0.0213	7.3	0.1551	0.3085	0.4036	0.5294
		Bootstrap			0.1494	0.3031	0.4208	0.5200
	4	$\gamma\chi^2(\beta)$	0.0161	16.2	0.2600	0.4265	0.5186	0.6356
		Bootstrap			0.2595	0.4261	0.5754	0.7239
	5	$\gamma\chi^2(\beta)$	0.0120	30.3	0.3651	0.5318	0.6179	0.7244
		Bootstrap			0.3688	0.5244	0.6458	0.7661
	6	$\gamma\chi^2(\beta)$	0.0090	51.8	0.4637	0.6230	0.7015	0.7967
		Bootstrap			0.4614	0.6255	0.6912	0.8184
	7	$\gamma\chi^2(\beta)$	0.0066	83.7	0.5526	0.7002	0.7704	0.8544
		Bootstrap			0.5585	0.7149	0.7814	0.9655
	8	$\gamma\chi^2(\beta)$	0.0048	130.5	0.6304	0.7641	0.8262	0.8996
		Bootstrap			0.6381	0.7738	0.8305	0.8963
	9	$\gamma\chi^2(\beta)$	0.0035	198.5	0.6973	0.8162	0.8704	0.9339
		Bootstrap			0.6970	0.8187	0.8769	1.0083
	10	$\gamma\chi^2(\beta)$	0.0025	297.1	0.7537	0.8582	0.9050	0.9594
		Bootstrap			0.7545	0.8566	0.9220	0.9830
$T2_n$	d	Method	γ	β	$E[T2]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0316	5.4	0.1715	0.3707	0.5005	0.6752
		Bootstrap			0.1636	0.3596	0.5191	0.7422
	3	$\gamma\chi^2(\beta)$	0.0183	20.4	0.3725	0.5830	0.6963	0.8389
		Bootstrap			0.3786	0.6001	0.7691	0.8773
	4	$\gamma\chi^2(\beta)$	0.0104	53.0	0.5491	0.7355	0.8271	0.9382
		Bootstrap			0.5546	0.7432	0.8571	1.2628
	5	$\gamma\chi^2(\beta)$	0.0057	119.8	0.6871	0.8393	0.9103	0.9946
		Bootstrap			0.6792	0.8475	0.9469	1.1019
	6	$\gamma\chi^2(\beta)$	0.0031	253.4	0.7881	0.9067	0.9601	1.0224
		Bootstrap			0.7805	0.9035	0.9837	1.0393
	7	$\gamma\chi^2(\beta)$	0.0017	519.2	0.8591	0.9487	0.9880	1.0334
		Bootstrap			0.8620	0.9557	0.9945	1.0766
	8	$\gamma\chi^2(\beta)$	0.0009	1050.2	0.9077	0.9738	1.0023	1.0350
		Bootstrap			0.9113	0.9782	1.0174	1.0424
	9	$\gamma\chi^2(\beta)$	0.0004	2118.4	0.9402	0.9882	1.0087	1.0320
		Bootstrap			0.9436	0.9937	1.0163	1.0561
	10	$\gamma\chi^2(\beta)$	0.0002	4284.5	0.9616	0.9960	1.0106	1.0271
		Bootstrap			0.9649	0.9989	1.0188	1.0482

TABLE A.4 The limits of expectations, γ and β values of $T1_n$ and $T2_n$ and the approximated size α=0.05, 0.01 and 0.001 critical values of their asymptotic distributions under H_0 approximated by $\gamma\chi^2_\alpha(\beta)$ and by bootstrap method, when the even numbered variables are from $Po(2)$ distribution and the odd numbered variables are from $N(0,1)$ distribution.

$N(0,1)$ and $Po(2)$ distributions

$T1_n$	d	Method	γ	β	$E[T1]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0279	2.3	0.0630	0.1809	0.2731	0.4040
		Bootstrap			0.0625	0.1831	0.2975	0.4046
	3	$\gamma\chi^2(\beta)$	0.0210	7.5	0.1583	0.3117	0.4063	0.5312
		Bootstrap			0.1591	0.3204	0.4429	0.5712
	4	$\gamma\chi^2(\beta)$	0.0158	16.7	0.2636	0.4295	0.5209	0.6369
		Bootstrap			0.2599	0.4376	0.5245	0.6088
	5	$\gamma\chi^2(\beta)$	0.0117	31.6	0.3704	0.5357	0.6207	0.7257
		Bootstrap			0.3705	0.5308	0.6105	0.7201
	6	$\gamma\chi^2(\beta)$	0.0087	53.9	0.4686	0.6263	0.7037	0.7975
		Bootstrap			0.4653	0.6269	0.7187	0.7743
	7	$\gamma\chi^2(\beta)$	0.0063	88.2	0.5585	0.7037	0.7726	0.8549
		Bootstrap			0.5620	0.7220	0.7817	0.9077
	8	$\gamma\chi^2(\beta)$	0.0046	137.2	0.6356	0.7669	0.8277	0.8996
		Bootstrap			0.6369	0.7730	0.8661	0.9606
	9	$\gamma\chi^2(\beta)$	0.0033	211.0	0.7028	0.8190	0.8718	0.9335
		Bootstrap			0.7072	0.8236	0.8726	0.9813
	10	$\gamma\chi^2(\beta)$	0.0024	315.5	0.7584	0.8604	0.9059	0.9589
		Bootstrap			0.7607	0.8626	0.8989	0.9786
$T2_n$	d	Method	γ	β	$E[T2]$	α=0.05	α=0.01	α=0.001
	2	$\gamma\chi^2(\beta)$	0.0309	5.7	0.1750	0.3735	0.5018	0.6741
		Bootstrap			0.1700	0.3613	0.4934	1.0056
	3	$\gamma\chi^2(\beta)$	0.0176	21.7	0.3808	0.5888	0.6700	0.8394
		Bootstrap			0.3840	0.5989	0.7463	0.9037
	4	$\gamma\chi^2(\beta)$	0.0100	55.9	0.5563	0.7401	0.8300	0.9389
		Bootstrap			0.5663	0.7459	0.8752	1.0726
	5	$\gamma\chi^2(\beta)$	0.0054	128.5	0.6955	0.8441	0.9132	0.9950
		Bootstrap			0.6953	0.8497	0.9303	1.0810
	6	$\gamma\chi^2(\beta)$	0.0029	270.4	0.7943	0.9099	0.9618	1.0223
		Bootstrap			0.7879	0.9054	0.9567	1.0648
	7	$\gamma\chi^2(\beta)$	0.0015	564.4	0.8649	0.9512	0.9891	1.0328
		Bootstrap			0.8636	0.9589	0.9931	1.0334
	8	$\gamma\chi^2(\beta)$	0.0008	1139.9	0.9116	0.9753	1.0028	1.0342
		Bootstrap			0.9104	0.9763	1.0061	1.0896
	9	$\gamma\chi^2(\beta)$	0.0004	2345.0	0.9434	0.9892	1.0087	1.0309
		Bootstrap			0.9442	0.9859	1.0093	1.0292
	10	$\gamma\chi^2(\beta)$	0.0002	4742.6	0.9637	0.9965	1.0104	1.0260
		Bootstrap			0.9657	0.9978	1.0099	1.0309

TABLE A.5 The averages of the test statistics $T1_n$ and $T2_n$, their estimated expectations and estimated variances from the simulations of data 1. (The bolded averages differ significantly from the corresponding theoretical values.)

d	n	$\overline{T1_n}$	$\widehat{E[T1]}$	$\widehat{\mathrm{Var}[T1]}$	$\overline{T2_n}$	$\widehat{E[T2]}$	$\widehat{\mathrm{Var}[T2]}$
2	10	0.0584	**0.0565**	0.0038	0.1703	**0.1620**	0.0141
	50	0.0645	0.0625	0.0036	0.1792	0.1749	0.0113
	100	0.0636	0.0633	0.0036	0.1797	0.1773	0.0112
	200	0.0631	0.0637	0.0036	0.1818	0.1778	0.0109
Theoretical			0.0641	0.0036		0.1786	0.0108
3	10	0.1556	**0.1417**	0.0071	0.3853	**0.3512**	0.0176
	50	0.1618	0.1563	0.0068	0.3829	0.3780	0.0141
	100	0.1616	0.1580	0.0067	0.3847	0.3818	0.0137
	200	0.1492	0.1590	0.0067	0.3681	0.3831	0.0134
Theoretical			0.1598	0.0067		0.3849	0.0132
4	10	0.2585	**0.2404**	0.0090	0.5639	**0.5255**	0.0153
	50	0.2672	0.2618	0.0085	0.5611	**0.5558**	0.0116
	100	0.2655	0.2647	0.0084	0.5616	0.5598	0.0112
	200	0.2718	0.2659	0.0084	0.5521	0.5618	0.0110
Theoretical			0.2671	0.0083		0.5635	0.0108
5	10	0.3647	**0.3401**	0.0096	0.6960	**0.6638**	0.0111
	50	0.3774	**0.3672**	0.0088	0.6980	**0.6934**	0.0080
	100	0.3735	0.3706	0.0087	0.6953	0.6970	0.0076
	200	0.3869	0.3726	0.0087	0.7192	0.6995	0.0075
Theoretical			0.3739	0.0086		0.7010	0.0073
6	10	0.4639	**0.4351**	0.0091	0.7990	**0.7675**	0.0072
	50	0.4703	**0.4660**	0.0082	0.7995	**0.7936**	0.0049
	100	0.4683	0.4700	0.0081	0.8003	0.7973	0.0047
	200	0.4860	0.4715	0.0080	0.8074	0.7983	0.0046
Theoretical			0.4736	0.0080		0.8003	0.0044
7	10	0.5562	**0.5226**	0.0081	0.8688	**0.8424**	0.0043
	50	0.5606	**0.5554**	0.0072	0.8697	**0.8641**	0.0028
	100	0.5635	0.5592	0.0071	0.8689	0.8664	0.0027
	200	0.5673	0.5610	0.0070	0.8697	0.8677	0.0026
Theoretical			0.5629	0.0069		0.8690	0.0025
8	10	0.6307	**0.6001**	0.0068	0.9143	**0.8946**	0.0025
	50	0.6411	**0.6331**	0.0059	0.9163	**0.9111**	0.0015
	100	0.6413	0.6370	0.0058	0.9162	0.9132	0.0014
	200	0.6332	0.6392	0.0058	0.9091	0.9145	0.0014
Theoretical			0.6408	0.0057		0.9154	0.0014
9	10	0.7008	**0.6676**	0.0055	0.9448	**0.9296**	0.0014
	50	0.7033	**0.7000**	0.0047	0.9449	**0.9430**	0.0008
	100	0.7082	0.7036	0.0046	0.9460	0.9444	0.0008
	200	0.7091	0.7055	0.0046	0.9465	0.9452	0.0007
Theoretical			0.7072	0.0045		0.9459	0.0007
10	10	0.7531	**0.7256**	0.0043	0.9647	**0.9541**	0.0007
	50	0.7638	**0.7563**	0.0037	0.9664	**0.9635**	0.0004
	100	0.7621	0.7595	0.0036	0.9647	0.9646	0.0004
	200	0.7553	0.7614	0.0035	0.9640	0.9652	0.0004
Theoretical			0.7631	0.0035		0.9658	0.0004

TABLE A.6 The averages of the test statistics $T1_n$ and $T2_n$, their estimated expectations and estimated variances from the simulations of data 2. (The bolded averages differ significantly from the corresponding theoretical values.)

d	n	$\overline{T1_n}$	$\overline{\widehat{E[T1]}}$	$\overline{\widehat{\text{Var}[T1]}}$	$\overline{T2_n}$	$\overline{\widehat{E[T2]}}$	$\overline{\widehat{\text{Var}[T2]}}$
2	10	0.0634	**0.0559**	0.0038	0.1740	**0.1579**	0.0141
	50	0.0627	0.0612	0.0037	0.1704	0.1689	0.0114
	100	0.0638	0.0617	0.0036	0.1696	0.1709	0.0112
	200	0.0568	0.0617	0.0035	0.1718	0.1710	0.0110
Theoretical			0.0620	0.0035		0.1715	0.0108
3	10	0.1568	**0.1418**	0.0074	0.3816	**0.3491**	0.0187
	50	0.1559	0.1532	0.0069	0.3711	0.3679	0.0145
	100	0.1561	0.1541	0.0068	0.3732	0.3703	0.0141
	200	0.1600	0.1552	0.0068	0.3715	0.3726	0.0135
Theoretical			0.1551	0.0066		0.3725	0.0136
4	10	0.2598	**0.2395**	0.0094	0.5565	**0.5204**	0.0163
	50	0.2594	0.2572	0.0087	0.5485	0.5433	0.0122
	100	0.2583	0.2585	0.0085	0.5522	0.5463	0.0118
	200	0.2641	0.2594	0.0085	0.5473	0.5474	0.0116
Theoretical			0.2600	0.0084		0.5491	0.0114
5	10	0.3633	**0.3394**	0.0099	0.6931	**0.6588**	0.0118
	50	0.3688	0.3613	0.0091	0.6880	0.6823	0.0085
	100	0.3661	0.3638	0.0090	0.6947	0.6847	0.0082
	200	0.3618	0.3639	0.0089	0.6850	0.6857	0.0080
Theoretical			0.3651	0.0088		0.6871	0.0079
6	10	0.4680	**0.4343**	0.0095	0.7947	**0.7625**	0.0078
	50	0.4610	0.4593	0.0086	0.7873	**0.7835**	0.0054
	100	0.4655	0.4621	0.0085	0.7901	0.7861	0.0051
	200	0.4429	0.4625	0.0084	0.7752	0.7871	0.0050
Theoretical			0.4637	0.0083		0.7881	0.0049
7	10	0.5515	**0.5210**	0.0084	0.8617	**0.8374**	0.0047
	50	0.5522	0.5481	0.0076	0.8605	**0.8551**	0.0032
	100	0.5526	0.5507	0.0075	0.8603	0.8574	0.0030
	200	0.5459	0.5515	0.0074	0.8590	0.8581	0.0029
Theoretical			0.5526	0.0073		0.8591	0.0028
8	10	0.6349	**0.5991**	0.0071	0.9128	**0.8910**	0.0028
	50	0.6322	0.6263	0.0064	0.9099	**0.9047**	0.0018
	100	0.6335	0.6285	0.0063	0.9096	0.9062	0.0017
	200	0.6463	0.6302	0.0062	0.9153	0.9075	0.0016
Theoretical			0.6304	0.0061		0.9077	0.0016
9	10	0.6973	**0.6667**	0.0058	0.9422	**0.9272**	0.0016
	50	0.6967	0.6931	0.0051	0.9414	**0.9378**	0.0010
	100	0.7014	0.6953	0.0050	0.9399	0.9390	0.0009
	200	0.7014	0.6958	0.0050	0.9411	0.9394	0.0009
Theoretical			0.6973	0.0049		0.9402	0.0008
10	10	0.7571	**0.7250**	0.0045	0.9633	**0.9520**	0.0008
	50	0.7567	**0.7496**	0.0040	0.9626	**0.9599**	0.0005
	100	0.7586	0.7521	0.0039	0.9635	0.9609	0.0005
	200	0.7635	0.7530	0.0039	0.9621	0.9613	0.0004
Theoretical			0.7537	0.0038		0.9616	0.0004

TABLE A.7 The averages of the test statistics $T1_n$ and $T2_n$, their estimated expectations and estimated variances from the simulations of data 3. (The bolded averages differ significantly from the corresponding theoretical values.)

d	n	$\overline{T1_n}$	$\widehat{E[T1]}$	$\widehat{Var[T1]}$	$\overline{T2_n}$	$\widehat{E[T2]}$	$\widehat{Var[T2]}$
2	10	0.0600	**0.0562**	0.0038	0.1725	**0.1599**	0.0141
	50	0.0622	0.0618	0.0036	0.1724	0.1718	0.0113
	100	0.0631	0.0625	0.0036	0.1744	0.1739	0.0112
	200	0.0580	0.0627	0.0035	0.1801	0.1744	0.0110
Theoretical			0.0630	0.0035		0.1750	0.0108
3	10	0.1546	**0.1417**	0.0072	0.3844	**0.3505**	0.0180
	50	0.1600	0.1554	0.0069	0.3792	0.3745	0.0142
	100	0.1578	0.1567	0.0067	0.3854	0.3777	0.0138
	200	0.1603	0.1576	0.0067	0.3766	0.3795	0.0135
Theoretical			0.1583	0.0066		0.3808	0.0134
4	10	0.2580	**0.2400**	0.0092	0.5581	**0.5230**	0.0158
	50	0.2626	0.2595	0.0086	0.5559	0.5500	0.0119
	100	0.2629	0.2617	0.0085	0.5605	0.5533	0.0115
	200	0.2767	0.2626	0.0084	0.5661	0.5545	0.0113
Theoretical			0.2636	0.0083		0.5563	0.0111
5	10	0.3695	**0.3399**	0.0097	0.6991	**0.6620**	0.0114
	50	0.3715	0.3649	0.0090	0.6937	**0.6886**	0.0082
	100	0.3726	0.3680	0.0088	0.6943	0.6921	0.0078
	200	0.3681	0.3690	0.0088	0.7009	0.6938	0.0077
Theoretical			0.3704	0.0087		0.6955	0.0075
6	10	0.4676	**0.4347**	0.0093	0.7990	**0.7653**	0.0075
	50	0.4728	**0.4630**	0.0085	0.7951	**0.7889**	0.0051
	100	0.4704	0.4660	0.0083	0.7950	0.7917	0.0049
	200	0.4796	0.4672	0.0082	0.7976	0.7930	0.0048
Theoretical			0.4686	0.0081		0.7943	0.0047
7	10	0.5520	**0.5219**	0.0082	0.8650	**0.8401**	0.0045
	50	0.5570	**0.5524**	0.0074	0.8665	**0.8604**	0.0030
	100	0.5598	0.5555	0.0072	0.8655	0.8624	0.0028
	200	0.5707	0.5574	0.0072	0.8724	0.8640	0.0027
Theoretical			0.5585	0.0071		0.8649	0.0027
8	10	0.6328	**0.5996**	0.0069	0.9131	**0.8928**	0.0026
	50	0.6365	**0.6297**	0.0062	0.9125	**0.9079**	0.0016
	100	0.6369	0.6329	0.0060	0.9136	0.9099	0.0016
	200	0.6456	0.6345	0.0060	0.9134	0.9108	0.0015
Theoretical			0.6356	0.0059		0.9116	0.0015
9	10	0.6949	**0.6672**	0.0056	0.9424	**0.9286**	0.0015
	50	0.6998	**0.6969**	0.0049	0.9417	**0.9407**	0.0009
	100	0.7066	0.6998	0.0048	0.9449	0.9420	0.0008
	200	0.6990	0.7010	0.0047	0.9396	0.9426	0.0008
Theoretical			0.7028	0.0047		0.9434	0.0008
10	10	0.7564	**0.7253**	0.0044	0.9638	**0.9529**	0.0008
	50	0.7590	**0.7528**	0.0038	0.9635	**0.9616**	0.0005
	100	0.7602	0.7558	0.0037	0.9641	0.9627	0.0004
	200	0.7729	0.7574	0.0037	0.9662	0.9634	0.0004
Theoretical			0.7584	0.0036		0.9637	0.0004

TABLE A.8 The averages of the test statistic $T1_n$ from the simulations of the dependent data 4 to 11. (The bolded averages are smaller than the corresponding size 0.05 critical values.)

		Data							
d	n	4	5	6	7	8	9	10	11
2	10	**0.1343**	**0.1380**	**0.0813**	**0.0827**	0.4327	0.4391	**0.0978**	**0.0932**
	50	0.5019	0.5332	**0.1716**	**0.1756**	2.1107	2.1202	**0.1214**	**0.1106**
	100	0.9904	0.9855	0.2897	0.2827	4.2263	4.2073	**0.1498**	**0.1318**
	200	1.7867	1.9420	0.4800	0.5251	8.4747	8.3971	0.1970	**0.1628**
3	10	**0.2989**	**0.3053**	**0.1938**	**0.1968**	0.4254	0.4287	**0.2297**	**0.2159**
	50	1.0069	1.0309	0.3758	0.3804	1.6879	1.6996	0.3306	**0.2873**
	100	1.8898	1.9374	0.6096	0.6257	3.2592	3.2557	0.4574	0.3830
	200	3.6568	3.6736	1.0838	1.0838	6.3954	6.4165	0.6680	0.5706
4	10	**0.4368**	0.4661	**0.3173**	**0.3139**	0.4530	0.4572	**0.3676**	**0.3420**
	50	1.3823	1.4103	0.5542	0.5685	1.4002	1.4079	0.6141	0.5324
	100	2.5085	2.6342	0.8731	0.8899	2.5640	2.5832	0.9409	0.7581
	200	4.8413	4.9403	1.4897	1.5971	4.9026	4.8930	1.5377	1.2370
5	10	0.5623	0.5820	**0.4199**	**0.4243**	**0.5008**	**0.5048**	**0.5059**	**0.4734**
	50	1.6299	1.7400	0.7221	0.7260	1.2094	1.2222	0.9542	0.8159
	100	2.9686	3.1217	1.0764	1.1020	2.0937	2.0893	1.5152	1.2268
	200	5.4631	5.9240	1.7399	1.9201	3.8467	3.8637	2.6219	2.1091
6	10	0.6664	0.6791	**0.5187**	**0.5256**	**0.5622**	**0.5675**	**0.6274**	**0.5947**
	50	1.7862	1.9072	0.8244	0.8666	1.0931	1.0912	1.3260	1.1312
	100	3.2052	3.4454	1.2050	1.2844	1.7522	1.7548	2.1741	1.7627
	200	6.0017	6.4949	1.9343	2.0203	3.0397	3.0934	3.9409	3.1544
7	10	0.7415	0.7743	**0.6074**	**0.6139**	**0.6277**	**0.6247**	0.7400	0.7030
	50	1.8460	2.0196	0.9115	0.9457	1.0155	1.0267	1.7337	1.4478
	100	3.2733	3.5508	1.2791	1.3677	1.5232	1.5132	2.8633	2.3840
	200	5.9366	6.8117	2.0239	2.1713	2.4947	2.5664	5.2210	4.1906
8	10	0.8125	0.8447	**0.6806**	**0.6845**	**0.6858**	**0.6832**	0.8584	0.8044
	50	1.9022	2.0776	0.9684	1.0127	0.9832	0.9817	2.0754	1.7594
	100	3.2832	3.6324	1.3037	1.4266	1.3475	1.3609	3.6218	2.9658
	200	6.0202	6.8108	1.9850	2.3587	2.0793	2.1124	6.7140	5.1504
9	10	0.8623	0.8936	**0.7406**	**0.7474**	**0.7321**	**0.7303**	0.9471	0.8844
	50	1.9163	2.0722	1.0055	1.0588	0.9546	0.9554	2.4667	2.0535
	100	3.1869	3.5819	1.3196	1.4496	1.2395	1.2354	4.3481	3.5151
	200	5.7768	6.5823	1.9520	2.2080	1.7823	1.7943	8.0775	6.4610
10	10	0.9097	0.9248	**0.7864**	**0.7981**	**0.7776**	**0.7820**	1.0293	0.9724
	50	1.8912	2.0755	1.0288	1.0702	0.9502	0.9476	2.8242	2.3388
	100	3.1009	3.5055	1.3195	1.4374	1.1504	1.1594	5.0723	4.0528
	200	5.4805	6.4507	1.9405	2.1579	1.5454	1.6095	9.5115	7.4306

TABLE A.9 The averages of the test statistic $T2_n$ from the simulations of the dependent data 4 to 11. (The bolded averages are smaller than the corresponding size 0.05 critical values.)

		Data							
d	n	4	5	6	7	8	9	10	11
2	10	**0.2759**	**0.2780**	**0.2041**	**0.2008**	0.8289	0.8427	**0.2165**	**0.2153**
	50	0.6891	0.7537	**0.2987**	**0.3051**	3.7373	3.7836	**0.2938**	**0.2581**
	100	1.2656	1.3163	0.4244	0.4286	7.4234	7.4461	0.4168	**0.3566**
	200	2.3183	2.4762	0.6252	0.7148	14.8294	14.8021	0.6329	0.5177
3	10	**0.5218**	**0.5238**	**0.4160**	**0.4152**	0.7272	0.7360	**0.4633**	**0.4419**
	50	1.1659	1.2370	**0.5634**	**0.5791**	2.4164	2.4536	0.7455	0.6519
	100	1.9665	2.1331	0.7488	0.7945	4.5489	4.5951	1.1156	0.9544
	200	3.5801	3.9138	1.1490	1.1999	8.7760	9.9193	1.8236	1.5681
4	10	**0.6987**	**0.7106**	**0.5976**	**0.5938**	0.7509	0.7450	**0.6682**	**0.6364**
	50	1.3750	1.4552	**0.7405**	0.7674	1.7186	1.7516	1.2531	1.1021
	100	2.2096	2.4739	0.9394	0.9920	2.9365	3.0202	2.0062	1.6926
	200	3.8924	4.3197	1.3172	1.4870	5.3744	5.5398	3.5815	2.8580
5	10	**0.8091**	**0.8147**	**0.7228**	**0.7223**	**0.7942**	**0.7963**	**0.8437**	**0.8066**
	50	1.4198	1.5556	0.8665	0.8832	1.3604	1.3782	1.7550	1.5193
	100	2.1786	2.4902	1.0389	1.0882	2.0638	2.1206	2.9644	2.4538
	200	3.6400	4.3050	1.3527	1.5624	3.4904	3.5955	5.3285	4.2936
6	10	**0.8885**	**0.8933**	**0.8140**	**0.8153**	**0.8472**	**0.8466**	0.9724	0.9405
	50	1.4148	1.5326	0.9294	0.9565	1.1735	1.1825	2.1916	1.8862
	100	2.0665	2.3639	1.0704	1.1439	1.5866	1.6160	3.7647	3.1014
	200	3.4567	3.9787	1.3405	1.4629	2.3854	2.4903	7.0494	5.7161
7	10	**0.9288**	**0.9454**	**0.8815**	**0.8779**	**0.8930**	**0.8916**	1.0802	1.0396
	50	1.3590	1.4706	0.9690	0.9920	1.0759	1.0895	2.6110	2.2102
	100	1.9017	2.1586	1.0747	1.1381	1.3195	1.3420	4.4665	3.7455
	200	2.9077	3.5248	1.3038	1.4287	1.7875	1.8881	8.2597	6.7016
8	10	**0.9570**	**0.9705**	**0.9231**	**0.9200**	**0.9284**	**0.9256**	1.1544	1.1030
	50	1.3081	1.4102	0.9859	1.0066	1.0391	1.0416	2.8588	2.4377
	100	1.7349	1.9445	1.0611	1.1218	1.1730	1.1922	5.0732	4.1982
	200	2.5831	3.0914	1.2153	1.3759	1.4466	1.4923	9.5688	7.4446
9	10	**0.9768**	**0.9860**	**0.9489**	**0.9481**	**0.9504**	**0.9497**	1.1944	1.1546
	50	1.2457	1.3342	0.9951	1.0155	1.0136	1.0160	3.1175	2.6356
	100	1.5688	1.7505	1.0511	1.1012	1.0948	1.1012	5.5950	4.4828
	200	2.2232	2.6529	1.1654	1.2680	1.2499	1.2778	10.3601	8.3521
10	10	**0.9872**	1.0010	**0.9663**	**0.9677**	**0.9669**	**0.9671**	1.2504	1.1899
	50	1.1890	1.2631	1.0010	1.0134	1.0049	1.0058	3.2987	2.7295
	100	1.4455	1.6117	1.0404	1.0757	1.0486	1.0571	5.9077	4.7602
	200	1.9267	2.2953	1.1253	1.2027	1.1386	1.1656	11.1021	8.6157

REFERENCES

Aitkin, M., Anderson, D., Francis, B., Hinde, J. 1989. *Statistical modelling in GLIM.* Oxford: Oxford Science Publications.

Ash, R. 1965. *Information theory.* (3rd edition) New York: Wiley.

Baringhaus, L., Henze, N. 1988. A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* **35**, 339–348.

Billingsley, P. 1968. *Convergence of probability measures.* New York: Wiley.

Blum, J. R., Kiefer, J., Rosenblatt, M. 1961. Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485–498.

Burckel, R. B. 1979. *An introduction to classical complex analysis I.* Stuttgart: Birkhäuser.

Cotterill, D. S., Csörgő, M. 1985. On the limiting distribution of and critical values for the Hoeffding, Blum, Kiefer, Rosenblatt independence criterion. *Statist. & Dec.* **3**, 1–48.

Cramér, H. 1966. *Mathematical methods of statistics.* (11th printing) Uppsala: Almqvist & Wiksells.

Csörgő, S. 1981a. Limit behaviour of the empirical characteristic function. *Ann. Prob.* **9** (1), 130–144.

Csörgő, S. 1981b. Multivariate empirical characteristic functions. *Zeit. Wahrscheinl.-theorie* **55**, 203–229.

Csörgő, S. 1985. Testing for independence by the empirical characteristic function. *J. Multivariate Anal.* **16**, 290–299.

Csörgő, S., Totik, V. 1983. On how long interval is the empirical characteristic function uniformly consistent ? *Acta Sci. Math.* **45**, 141–149.

Cuppens, R. 1975. *Decomposition of multivariate probability.* New York: Academic Press.

Deheuvels, P. 1981. An asymptotic decomposition for multivariate tests of independence. *J. Multivariate Anal.* **11**, 102–113.

De Wet, T. 1980. Cramér-von Mises tests for independence. *J. Multivariate Anal.* **10**, 38–50.

Efron, B., Tibshirani, R. J. 1993. *An introduction to the bootstrap.* New York: Chapman & Hall.

Epps, T. W., Pulley, L. B. 1983. A test for normality based on the empirical characteristic function. *Biometrika* **70**, 723–726.

Feuerverger, A. 1993. A consistent test for bivariate dependence. *The Int. Stat. Rev.* **61**, 419–433.

Feuerverger, A., Mureika, R. A. 1977. The empirical characteristic function and its applications. *Ann. Statist.* **5**, 88–97.

Heathcote, C. R. 1972. A test of goodness of fit for symmetric random variables. *Austral. J. Statist.* **14**, 172–181.

Heikkinen, E., Lahtinen, A., Lehtoranta, R. 1972. 66–vuotiaiden jyväskyläläisten liikunta- ja terveystutkimus. Perusraportti. *Jyväskylän yliopiston kansanterveyden laitoksen julkaisuja* **7**. (In Finnish.)

Hoeffding, W. 1948a. A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325.

Hoeffding, W. 1948b. A nonparametric test for independence. *Ann. Math. Statist.* **19**, 546–557.

Johnson, N. L., Kotz, S. 1970. *Continuous univariate distributions 2.* New York: Houghton Mifflin Company.

Koutrouvelis, I. A. 1980. A goodness-of-fit test based on the empirical characteristic function. *Biometrika* **67**, 238–240.

Koutrouvelis, I. A. 1981. Tests for normality in stable laws. *Commun. Statist.* **A10**, 581–595.

Koutrouvelis, I. A. 1985. Distribution free procedures for location and symmetry inference problems based on the empirical characteristic function. *Scand. J. Statist.* **12**, 257–269.

Kuusinen, J., Leskinen, E. 1986. Älykkyys ja koulumenestys. *Psykologia* **4**, 243–248. (In Finnish.)

Kuusinen, J., Leskinen, E. 1988. Koulumenestyksen rakenne ja pysyvyys peruskoulun 6. ja 9. luokilla. *Kasvatus* **19** (1), 22–31. (In Finnish.)

Leskinen, E., Kuusinen, J. 1991. Faktorianalyysin käytöstä kasvatustieteellisessä tutkimuksessa. *Kasvatus* **22** (4), 289–297. (In Finnish.)

Loéve, M. 1977. *Probability theory I.* (4th edition) New York: Springer-Verlag.

Lukacs, E. 1970. *Characteristic functions.* (2nd edition) London: Griffin.

Lukacs, E., Laha, R. G. 1964. *Applications of characteristic functions.* London: Griffin.

McCullagh, P., Nelder, J.A. 1989. *Generalized linear models.* (2nd edition) London: Chapman & Hall.

Marcus, M. B. 1981. Weak convergence of the empirical characteristic function. *Ann. Prob.* **9**, 194–210.

Parzen, E. 1962. On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**, 1065–1076.

Rosenblatt, M. 1975. A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** (1), 1–14.

Shorack, G. R., Wellner, J. A. 1986. *Empirical processes with applications to statistics.* New York: Wiley.

Serfling, R. J. 1980. *Approximation theorems of mathematical statistics.* New York: Wiley.

YHTEENVETO

Kahden muuttujan välistä riippumattomuutta on mahdollista testata monilla eri testeillä, mutta usean muuttujan täydellisen riippumattomuuden testaamisessa on vähemmän vaihtoehtoja. Yleisesti käytännössä testataan vain parittaiset riippumattomuudet muuttujajoukossa. Se ei kuitenkaan takaa täydellistä riippumattomuutta. Koska täydellisen riippumattomuuden hypoteesi voidaan kirjoittaa luonnollisella tavalla karakteristisen funktion avulla, perustuukin nyt kehitetty testi empiiriseen karakteristiseen funktioon. Empiiristä karakteristista funktiota on käytetty myös parametrien estimoinnissa ja esimerkiksi yhteensopivuuden, symmetrisyyden ja riippumattomuuden testaamisessa. Uusi testisuure on integraalimuunnos riippumattomuushypoteesia vastaavasta empiirisestä stokastisesta prosessista.

Tähän työhön on kerätty joitakin karakteristisen ja empiirisen karakteristisen funktion ominaisuuksia. Lisäksi on tarkasteltu riippumattomuushypoteesiin liittyvän empiirisen stokastisen prosessin asymptoottista käyttäytymistä. Testisuureen jakaumat on johdettu H_0–hypoteesin ja myös vaihtoehtoisen H_1–hypoteesin vallitessa. Testauksessa tosin käytetään approksimaatiota rajajakaumalle, koska se on helposti laskettavissa otoksen perusteella. Tällöin kriittisiä arvoja ei tarvitse taulukoida ja testi on käytössä täysin muuttujien jakaumista riippumaton. Lisäksi on todistettu, että testi on tarkentuva. Testisuureen jakaumassa esiintyvälle painofunktiolle on annettu joitakin ehtoja, joiden tulee olla voimassa, mutta esimerkiksi standardoidun normaalijakauman tiheysfunktio on sovelias ja suhteellisen hyvin toimiva vaihtoehto painofunktioksi. Jos painofunktiona käytetään tasajakauman tiheysfunktiota, täytyy aineiston muuttujista tietää, että niiden jakaumilla on analyyttiset karakteristiset funktiot.

Jatkuville muuttujille on tarkasteltu myös testin jakaumasta riippumatonta muunnosta, jolloin painofunktiollekaan ei tarvitse asettaa rajoituksia.

Simulaatiokokeilla on vertailtu Blum–Kiefer–Rosenblattin BKR_n testiä (Blum et al. 1961) ja uusia (kaksi erilaista painofunktiota) testejä empiirisen voimakkuusfunktion arvojen perusteella. Uudet testit ovat vertailukelpoisia BKR_n testin kanssa ja varsinkin diskreetillä datalla on suositeltavampaa käyttää uutta testiä. BKR_n testillä tuntuu olevan ongelmia diskreetin aineiston kanssa, mutta uudet nyt kehitetyt testit toimivat yhtä hyvin jatkuvalla ja diskreetillä aineistolla. Painofunktion valinnalla on jonkin verran vaikutusta siihen, kuinka nopeasti empiirisen voimakkuusfunktion arvot konvergoivat ykköseen.

Muuttujien riippumattomuutta on tutkittu kahdessa esimerkissä. Ensimmäinen aineisto on eläkeikäisten jyväskyläläisten naisten elinoloista, terveydentilasta ja elämäntyylistä sekä niiden yhteydestä vanhenemiseen (Heikkinen et al., 1972). Toisessa esimerkissä tutkitaan estimoitujen varimax menetelmällä saatujen faktoripistemäärien riippumattomuutta. Ensimmäisessä ryhmässä on kolme faktoria, jotka on saatu suomalaisesta ITPA aineistosta (Kuusinen and Leskinen, 1986) ja kahdessa muussa kolmen faktorin ryhmässä faktorit on laskettu aineistosta, jonka avulla tutkitaan koulumenestyksen rakennetta ja pysyvyyttä peruskoulun 6. ja 9. luokilla (Kuusinen and Leskinen, 1988).