A NOTE ON A SPECIFICATION TEST OF INDEPENDENCE

by

Holger Dette and Natalie Neumeyer
Ruhr-Universität Bochum

Technical Report #98-06

Department of Statistics
Purdue University
West Lafayette, IN USA

1998

A NOTE ON A SPECIFICATION TEST OF INDEPENDENCE

by

Holger Dette and Natalie Neumeyer
Ruhr-Universität Bochum
Institut und Fakultät für Mathematik
44780 Bochum
Germany

Abstract

In a recent paper Zheng (1997a) proposed a new specification test of independence between two random vectors by the kernel method. He showed asymptotic normality under the hypothesis and local alternatives. The present work investigates the asymptotic distribution of the corresponding test statistic under fixed alternatives. In this case asymptotic normality of a standardized statistic is still valid but with a different rate of convergence.

Keywords: Test of independence, consistency, fixed alternatives, U-statistics

AMS 1991 Subject Classification: Primary 62G10; Secondary 62G07

1. Introduction

The verification of the independence of two random vectors is an important problem in statistical inference and numerous statistical procedures are based on this assumption. Early work on the problem of testing independence dates back to Hoeffding (1948a) or Blum, Kiefer and Rosenblatt (1961) and is based on the empirical distribution function; for more recent work we refer to Rosenblatt (1975), Robinson (1991) Rosenblatt and Whalen (1992) or Ahmad and Li (1997), who applied kernel density estimation to this problem. In a recent paper Zheng (1997a) described several drawbacks of these tests and proposed a new procedure for testing the independence of two random variables. He proved asymptotic normality of the corresponding test statistic under the hypothesis of independency and under local alternatives.

In the present paper we extend these results and establish the = asymptotic normality for a standardized version of Zheng's statistic under fixed alternatives. This provides further arguments in favor of Zheng's (1997a) approach because the asymptotic results of this paper can now also be used to = control the type II error of the test. From a practical point of view this is particularly important because the acceptance of the null will usually yield to a data analysis adapted to the independence assumption and is essential to control the corresponding error of such a procedure. Surprisingly it turns out that the rates of convergence for Zheng's (1997a) statistic are different in both cases. While under the hypotheses of independence the rate is $nh^{m/2}$ (here n denotes the sample size, h a bandwidth and m the dimension of the predictor), the rate of convergence under the alternative can be shown to be $n^{1/2}$ (independently of the bandwidth and the dimension of the predictor).

2. Zheng's Test of Independence

Following Zheng (1997a) let $Q(y,x)$ denote the joint distribution function of a random vector (Y,X) with values in $\mathbb{R}^{\ell \times m}$ and define $G(y|x)$ and $F(y)$ as the conditional distribution (given $X = x$) and marginal distribution function of Y respectively. The difference of these distribution functions will play an important role in this paper and is denoted by

$$\Delta(y|x) = G(y|x) - F(y). \tag{2.1}$$

The random variables X and Y are independent if and only if the = hypothesis

$$H_0: \ \Delta(y|x) = 0 \qquad \forall \ x, y \tag{2.2}$$

is valid. The alternative $\Delta(y|x) \neq 0$ will be denoted by H_1. The test of Zheng (1997a) is based on an estimator of the integral

$$W = \int E[\Delta^2(y|X)p(X)]w(y)dy$$

where p is the marginal density of X and w is a weight function. To be precise let $(Y_1, X_1), \ldots, (Y_n, X_n)$ denote a sample from the distribution Q and define

(2.3)

$$W_n = \frac{1}{n(n-1)} \sum_{i \neq 3Dj} \frac{1}{h^m} K\left(\frac{X_i - X_j}{h}\right) \int [I\{Y_i \leq y\} - F_n(y)][I\{Y_j \leq y\} - F_n(y)]w(y)dy$$

where $I\{\cdot\}$ denotes the indicator function, K is a kernel, h a bandwidth and F_n the empirical distribution function of $Y_1, \ldots, Y_n$. Under the hypothesis of independence (2.2) Zheng (1997a) showed that $nh^{m/2}W_n$ is asymptotically normal ($n \to \infty$, $h \to 0$, $nh^m \to \infty$) with mean 0 and variance

$$\sigma^2 = 2 \int K^2(u)du \int \int E[p(X)]\{F(y \wedge z) - F(y)F(z)\}^2 w(y)w(z)dydz,$$

where $y \wedge z = (y_1 \wedge z_1, \ldots, y_\ell \wedge z_\ell)^T$ denotes the minimum of the vectors $z = (z_1, \ldots, z_\ell)^T$ and $y = (y_1, \ldots, y_\ell)^T$. A similar result holds for local alternatives of the form $\Delta(y|x) - n^{-1/2}h^{-m/4}u(x, y)$ [see Theorem 3 of Zheng (1997a)]. A proof of these assertions requires the following basic assumptions.

(A1) The marginal density p of X and its first order derivatives are uniformly bounded. The conditional distribution function $G(y|x)$ has uniformly bounded first order derivatives with respect to x.

(A2) The weight function w is integrable and positive (a.e.).

(A3) The kernel K is a nonnegative, bounded, continuous and symmetric function on $\mathbb{R}^m$ such that $\int K(u)du = 1$ and $\int K(u)||u||^2 du < \infty$ (here $||\cdot||$ denotes the Euclidean norm on $\mathbb{R}^m$).

As a consequence of these results a consistent asymptotic level α test for the hypothesis of independence is obtained by rejecting H_0 whenever

$$|W_n| > nh^{m/2}\hat{\sigma}\, u_{1-\alpha}, \tag{2.4}$$

where $\hat{\sigma}^2$ is an appropriate estimator of the asymptotic variance and $u_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution.

The following result extends Zheng's (1997a) findings and determines the asymptotic distribution of his statistic in the case of = fixed alternatives. Throughout this paper $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ, variance σ^2 and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

Theorem 2.1. *If the assumptions (A1) – (A3) are satisfied, $h \to 0$, $nh^m \to \infty$, then under the alternative hypothesis $H_1 : \Delta(y|x) \neq 0$,*

$$\sqrt{n}(W_n - B_h) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda^2) \tag{2.5}$$

where the asymptotic bias and variance are given by

$$\begin{aligned} B_h &= \int\int \frac{1}{h}K\left(\frac{x_1 - x_2}{h}\right)\Delta(y|x_1)\Delta(y|x_2)p(x_1)p(x_2)dx_1dx_2w(y)dy \\ &= \int\int \Delta^2(y|x_1)p^2(x_1)dx_1dy + o(1) \\ &= E[\int \Delta^2(y|X)p(X)dy] + o(1) \end{aligned} \tag{2.6}$$

$$\begin{aligned} \lambda^2 = 4E&\left[\int\int \{G(y \wedge z|X) - G(y|X)G(z|X)\}\, v(z|X)v(y|X)w(y)w(z)dydz\right] \\ &+ 4\ Var\left[\int \Delta(y|X)v(y|X)w(y)dy\right] \end{aligned} \tag{2.7}$$

and $v(y|X)$ denotes the centered version of the random variable $\Delta(y|X)p(X)$ i.e.

$$v(y|X) = \Delta(y|X)p(X) - E[\Delta(y|X)p(X)]. \tag{2.8}$$

A standard calculation shows that the remainder in (2.6) is in general not of order $o(1/\sqrt{n})$ and as a consequence the measure of dependency W can in general not be used to center the statistic W_n under the alternative. However, if K is a kernel of order r [see e.g. Gasser, Müller, Mammitzsch (1985)], the marginal distribution function $G(y|x)$ and the marginal density $p(x)$ are r times continuously differentiable with bounded derivatives and $h = o(n^{-1/2r})$, then a straightforward calculation shows that one can replace B_h by W in (2.5), i.e.

$$\sqrt{n}(W_n - W) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda^2).$$

It should also be noted that the rates of convergence are different in Theorem 2.1 and Theorem 1 of Zheng (1997a). While under the null hypothesis of independency (and under local alternatives of order $n^{-1/2}h^{-m/4}$) the variance of W_n is of order $n^{-2}h^m$, the variance under fixed alternatives is of order n^{-1}. Therefore the asymptotic behaviour of W_n is similar as that of the statistics considered by Hoeffding (1948a) and Blum, Kiefer and Rosenblatt (1961) with a slightly different rate under the null hypothesis. As pointed out by Zheng (1997a) a test based on W_n has several advantages compared to the commonly used tests. Theorem 2.1 now provides a further argument in favor of the statistic W_n. The asymptotic results for fixed alternatives can be used in order to estimate the type II error of the test (2.4). From a practical point of view this is particularly important, because the acceptance of the null will yield to a subsequent data analysis adapted to the independence assumption. Consequently it is desirable to control the corresponding error of such a procedure.

The limiting distribution of W_n (under the hypothesis and alternative) can be derived under much weaker assumptions as commonly used in the literature [see e.g. Robinson (1991)]. However, if the density p of the marginal distribution of X has compact support, say C, and is positive on C, then an alternative test statistic may become appropriate, for which the asymptotics in Theorem 2.1 is more transparent. To be precise, let

$$(2.9) \qquad \hat{p}(x) = \frac{1}{nh^m} \sum_{j=3D1}^{n} K\left(\frac{x - X_j}{h}\right)$$

denote the usual density estimate of p and define

$$(2.10)\qquad W_n^* = \frac{1}{n(n-1)h^m} \sum_{i\neq 3Dj} \frac{1}{\hat{p}(X_i)} K\left(\frac{X_i - X_j}{h}\right) \int \left[I\{Y_i \le y\} - F_n(y)\right]\left[I\{Y_j \le y\} - F_n(y)\right] w(y)dy$$

Theorem 2.2. *If the assumption (A1)–(A3) are satisfied, $n \to \infty$, $h \sim n^{-1/(m+4)}$, $C =$ supp (p) is compact, p is positive on C, twice continuously differentiable with a uniformly bounded second derivative on C and the kernel K satisfies a Lipschitz condition, then we have the following asymptotic properties for the statistic W_n^* defined in $=$ (2.10).*

(a) If the hypothesis of independence is valid, then

$$(2.11)\qquad nh^{m/2} W_n^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$$

where the asymptotic variance is given by

$$(2.12)\qquad \tau_1^2 = 2\int K^2(u)du \int\int \{F(y \wedge z) - F(y)F(z)\}^2 w(y)w(z)dydz.$$

(b) Under the alternative $\Delta(y|x) \neq 0$ we have

$$(2.13)\qquad \sqrt{n}(W_n^* - B_h^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu^2)$$

where the asymptotic variance and bias are given by

$$(2.14)\qquad \begin{aligned} B_n^* &= \int\int \frac{1}{h} K\left(\frac{x_1 - x_2}{h}\right) \Delta(y|x_1)\Delta(y|x_2)p(x_2)dx_1dx_2w(y)dy \\ &= E\left[\int \Delta^2(y|X)w(y)dy\right] + o(1) \end{aligned}$$

and

$$(2.15)\qquad \begin{aligned} \mu^2 = 4E&\left[\int\int \{G(y \wedge z|X) - G(y|X)G(z|X)\}\Delta(z|X)\Delta(y|X)w(y)w(z)dydz\right] \\ &+ 4\ Var\left[\int \Delta^2(y|X)w(y) = dy\right]. \end{aligned}$$

The proof of Theorem 2.2 follows by the same arguments as given by Zheng (1997a) for part (a) and given in the proof of Theorem 2.1 for part (b) in the following section, observing that by results of Collomb and Härdle (1986) for every $\eta > 0$

$$\sup_{x \in C} |\hat{p}(x) - p(x)| = O_p(n^{-2/(4+m)+\eta}).$$

This allows us to replace the estimator $\hat{p}(X_i)$ in the statistic (2.10) by the random variable $p(X_i)$ and to work with a weighted version of W_n.

3. Proof of Theorem 2.1

For the sake of transparency, we restrict ourselves to the case of a one-dimensional predictor, i.e. $m = 1$; the general case follows exactly the same lines. Recall the definition of $\Delta(y|x)$ in (2.1), define $\varepsilon_i(y) = I\{Y_i \leq y\} - G(y|X_i)$,

$$\Gamma_n(y) = F_n(y) - F(y) = 3D\frac{1}{n}\sum_{k=1}^{n}(\varepsilon_k + \Delta(y|X_k)); \tag{3.1}$$

then we obtain the following decomposition for the statistic W_n in (2.3)

$$W_n = W_{n1} + W_{n2} + W_{n3} + 2W_{n4} - 2W_{n5} - 2W_{n6} \tag{3.2}$$

where

$$W_{n1} = \frac{1}{n(n-1)} \sum_{i \neq 3D j} \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \int \varepsilon_i(y)\varepsilon_j(y)w(y)dy \tag{3.3}$$

$$W_{n2} = \frac{1}{n(n-1)} \sum_{i \neq 3D j} \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \int \Delta(y|X_i)\Delta(y|X_j)w(y)dy \tag{3.4}$$

$$W_{n3} = \frac{1}{n(n-1)} \sum_{i \neq 3D j} \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \int \Gamma_n^2(y)w(y)dy \tag{3.5}$$

$$W_{n4} = \frac{1}{n(n-1)} \sum_{i \neq 3D j} \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \int \varepsilon_i(y)\Delta(y|X_j)w(y)dy \tag{3.6}$$

$$(3.7)\qquad W_{n5} = \frac{1}{n(n-1)} \sum_{i \neq 3Dj} \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \int \varepsilon_i(y)\Gamma_n(y)w(y)dy$$

$$(3.8)\qquad W_{n6} = \frac{1}{n(n-1)} \sum_{i \neq 3Dj} \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \int \Delta(y|X_i)\Gamma_n(y)w(y)dy.$$

¿From Zheng (1997a) we have

$$(3.9)\qquad W_{n1} = O_p\left(\frac{1}{n\sqrt{h}}\right); \quad W_{n3} = o_p\left(\frac{1}{n\sqrt{h}}\right); \quad W_{n5} = o_p\left(\frac{1}{n\sqrt{h}}\right)$$

as $n \to \infty$, and it remains to consider the terms W_{n2}, W_{n4}, W_{n6} which turn out to be of order $O(\frac{1}{\sqrt{n}})$. In order to show this assertion we first note that a straightforward calculation gives

$$(3.10)\qquad E[W_{n4}] = 0;$$

$$(3.11)\qquad E[W_{n6}] = \frac{2}{n} E\left[\int \Delta^2(y|X_1)p(X_1)w(y)dy\right] + o(\frac{1}{n})$$

and

$$(3.12)\qquad \begin{aligned} E[W_{n2}] &= E\left[\int \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \Delta(y|X_i)\Delta(y|X_j)w(y)dy\right] \\ &= E\left[\int \Delta^2(y|X_1)p(X_1)w(y)dy\right] + o(1). \end{aligned}$$

The calculation of the corresponding variances and covariances is actually more complicated. We have from (3.7) and (3.12)

$$\begin{aligned} E[W_{n2}^2] &= \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{i' \neq 3Dj'} \frac{1}{h^2} \int\int E\Bigg[K\left(\frac{X_i - X_j}{h}\right) K\left(\frac{X_{i'} - X_{j'}}{h}\right) \\ &\qquad \times \Delta(y|X_i)\Delta(y|X_j)\Delta(z|X_{i'})\Delta(z|X_{j'})\Bigg] w(y)w(z)dydz \\ &= (1 - \frac{4}{n})(E[W_{n2}])^2 + \frac{4}{nh^2} \int\int E\Bigg[K\left(\frac{X_1 - X_2}{h}\right) K\left(\frac{X_1 - X_3}{h}\right) \\ &\qquad \times \Delta(y|X_1)\Delta(y|X_2)\Delta(z|X_1)\Delta(z|X_3)\Bigg] w(y)w(z)dydz + o(\frac{1}{n}) \\ &= (1 - \frac{4}{n})(E[W_{n2}])^2 + \frac{4}{n} E\left[\{\int \Delta^2(y|X_1)p(X_1)w(y)dy\}^2\right] + o(\frac{1}{= n}) \end{aligned}$$

which gives

$$(3.13)\qquad \text{Var }(W_{n2}) = \frac{4}{n}\text{ Var }\left(\int \Delta^2(y|X_1)p(X_1)w(y)dy\right) + o(\frac{1}{n}).$$

Similarly,

$$(3.14)\quad E[W_{n4}^2] = \frac{1}{n^2(n-1)^2}\sum_{i\neq j}\sum_{i'\neq 3Dj'}\int\int \frac{1}{h^2}E\Bigg[K\left(\frac{X_i - X_j}{h}\right)K\left(\frac{X_{i'} - X_{j'}}{h}\right)$$
$$\times \Delta(y|X_j)\Delta(z|X_{j'})f(X_i, X_{i'}, X_j, X_{j'})\Bigg]w(y)w(z)dydz$$

where the conditional expectation

$$f(X_i, X_j, X_{i'}, X_{j'}) := E[\varepsilon_i(y)\varepsilon_{i'}(z)|X_i, X_j, X_{i'}, X_{j'}] = E[\varepsilon_i(y)\varepsilon_{i'}(y)|X_i, X_{i'}]$$

vanishes whenever $i \neq i'$ and is given by

$$(3.15)\qquad f(X_i, X_j, X_i, X_{j'}) = G(y\wedge z|X_i) - G(y|X_i)G(z|X_i) =: H(y, z|X_i)$$

otherwise (here the last equality defines $H(y, z|X_i)$). Observing (3.10) this implies for the variance of W_{n4}

$$\begin{aligned}\text{Var }(W_{n4}) &= \frac{1}{n}\int\int\frac{1}{h^2}E\left[K\left(\frac{X_1 - X_2}{h}\right)K\left(\frac{X_1 - X_3}{h}\right)\Delta(y|X_2)\Delta(z|X_3)H(y, z|X_1)\right]\\ (3.16)\qquad &\qquad\times w(y)w(z)dydz + o(\frac{1}{n})\\ &= \frac{1}{n}E\Bigg[\int\int\{G(y\wedge z|X_1) - G(y|X_1)G(z|X_1)\}\Delta(y|X_1)\Delta(z|X_1)\\ &\qquad\times p^2(X_1)w(y)w(z)dydz\Bigg] + o(\frac{1}{n})\end{aligned}$$

where the first equality follows from (3.14) by summing only over those pairs (i, j, i', j') for which $i = i'$. For the term W_{n6} we obtain from (3.8) and (3.1) $W_{n6} = W_{n6}^{(1)} + W_{n6}^{(2)} =$ where

$$(3.17)\qquad W_{n6}^{(1)} = \frac{1}{n^2(n-1)}\sum_{i\neq j}\sum_k \frac{1}{h}K\left(\frac{X_i - X_j}{h}\right)\int \Delta(y|X_i)\varepsilon_k(y)w(y)dy$$

$$W_{n6}^{(2)} = \frac{1}{n^2(n-1)} \sum_{i \neq j} \sum_k \frac{1}{h} K\left(\frac{X_i - X_j}{h}\right) \int \Delta(y|X_i)\Delta(y|X_k)w(y)dy. \tag{3.18}$$

This gives for the L^2-norm of the first term

$$\begin{aligned} E[(W_{n6}^{(1)})^2] &= \frac{1}{n^4(n-1)^2} \sum_{i \neq j} \sum_{i' \neq 3Dj'} \sum_k \frac{1}{h^2} \int\int E\Bigg[K\left(\frac{X_i - X_j}{h}\right) K\left(\frac{X_{i'} - X_{j'}}{h}\right) \\ &\qquad \times \Delta(y|X_i)\Delta(z|X_{i'})\varepsilon_k(y)\varepsilon_k(z)\Bigg] w(y)w(z)dydz \\ &= \frac{1}{n}\int\int E\left[\frac{1}{h}K\left(\frac{X_1 - X_2}{h}\right)\Delta(y|X_1)\right] E\left[\frac{1}{h}K\left(\frac{X_3 - X_4}{h}\right)\Delta(z|X_3)\right] \\ &\qquad \times E[\varepsilon_k(y)\varepsilon_k(z)]w(y)w(z)dydz + o(\frac{1}{n}) \\ &= \frac{1}{n}\int\int E[\Delta(y|X_1)p(X_1)]E[\Delta(z|X_2)p(X_2)]E[H(y,z|X_3)] \\ &\qquad \times w(y)w(z)dydz + o(\frac{1}{n}) \end{aligned}$$

where $H(y,z|X_3)$ is defined in (3.15). An analogous argument shows $E[W_{n6}^{(1)} W_{n6}^{(2)}] = o(\frac{1}{n})$ and

$$\begin{aligned} E[(W_{n6}^{(2)})^2] &= \frac{1}{n}\int\int E[\Delta(y|X_1)p(X_1)]E[\Delta(z|X_2)p(X_2)]E[\Delta(y|X_3)\Delta(z|X_3)] \\ &\qquad \times w(y)w(z)dydz + o(\frac{1}{n}) \end{aligned}$$

which yields in combination with (3.11)

$$\begin{aligned} \text{Var}\,(W_{n6}) &= \frac{1}{n}\int\int E[\Delta(y|X_2)p(X_2)]E[\Delta(z|X_3)p(X_3)]E[H(y,z|X_1)]w(y)w(z)dydz \\ &\quad + \frac{1}{n}E\left[\left\{\int \Delta(y|X_1)E[\Delta(y|X_2)p(X_2)]w(y)dy\right\}^2\right] + o(\frac{1}{n}) \end{aligned} \tag{3.19}$$

The remaining covariances can be treated by similar arguments and give

$$\text{Cov}\,(W_{n2}, W_{n4}) = o(\frac{1}{n}) \tag{3.20}$$

$$\begin{aligned} \text{Cov}\,(W_{n2}, W_{n6}) &= \frac{2}{n}\int\int E[\Delta^2(y|X_1)\Delta(z|X_1)\Delta(z|X_3)p(X_1)p(X_3)] \\ &\qquad \times w(y)w(z)dydz + o(\frac{1}{n}) \end{aligned} \tag{3.21}$$

$$\text{(3.22)} \qquad \begin{aligned} \text{Cov }(W_{n4}, W_{n6}) = \frac{1}{n} \int\int E\left[\Delta(y|X_1)\Delta(z|X_3)H(y,z|X_1)p(X_1)p(X_3)\right] \\ \times\, w(y)w(z)dydz + o(\frac{1}{n}). \end{aligned}$$

A combination of (3.21), (3.13), (3.16), (3.19)–(3.20) and straightforward (but tedious) algebra now yield

$$\text{Var }(W_n) = \frac{\lambda^2}{n} + o(\frac{1}{n})$$

where λ^2 is defined in (2.7).

Finally, the asymptotic normality of

$$\sqrt{n}(W_n - E[W_n]) = \sqrt{n}(W_{n2} - E[W_{n2}]) + 2\sqrt{n}W_{n4} - 2\sqrt{n}W_{n6} + o(1)$$

can be easily obtained by using the common arguments from the theory of U-statistics [see Hoeffding (1948b), Hall (1984) and Zheng (1997a,b)]. For the sake of simplicity we only consider the statistic W_{n4}; the assertion in the general case then follows by applying the Cramér-Wold device and similar arguments to the terms $W_{n6}, W_{n2} - E[W_{n2}]$. Define $Z_i = (X_i, Y_i)$; $i = 3D1, \ldots, n$, and a symmetric kernel by

$$U_n(Z_i, Z_j) = \frac{1}{2h}K\left(\frac{X_i - X_j}{h}\right)\int[\varepsilon_i(y)\Delta_j(y) + \Delta_i(y)\varepsilon_j(y)]w(y)dy,$$

then the statistic W_{n4} can be rewritten as

$$W_{n4} = \binom{n}{2}^{-1}\sum_{i<j} U_n(Z_i, Z_j).$$

Note that W_{n4} is a U-statistic with a kernel U_n depending on the sample size n as considered by Hall (1984). Moreover, $E[U_n(Z_i, Z_j)] = 0$ and it can also be shown that $E[U_n^2(Z_i, Z_j)] = o(n)$. From Lemma 3.1 in Zheng (1997b) we have

$$\text{(3.23)} \qquad W_{n4} - \hat{W}_{n4} = o_p(\frac{1}{\sqrt{n}})$$

where $\hat{W}_{n4}$ denotes the projection of the U-statistic W_{n4}, i.e.

$$\hat{W}_{n4} = \frac{2}{n}\sum_{i=3D1}^{n} E[U_n(Z_i, Z_j)|Z_i].$$

The asymptotic normality now follows from Ljapunoff's central limit theorem for triangular arrays of independent random variables,i.e.

$$\sqrt{n}\hat{W}_{n4} \xrightarrow{\mathcal{D}} \mathcal{N}(0, c^2)$$

where c^2 can be obtained from (3.16). The asymptotic normality of W_{n4} is now obvious by (3.23) and the discussion at the beginning of this paragraph completes the proof of Theorem 2.1. ∎

Acknowledgements. Parts of this paper were written while the first author was visiting Purdue University. This author would like to thank the Department of Statistics for its hospitality and the Deutsche Forschungsgemeinschaft for the financial support which made this visit possible. The authors are also grateful to Teena Seele who typed this paper with considerable technical expertise.

References

I. A. Ahmad, Q. Li (1997). Testing independence by nonparametric kernel method. *Statist. Probab. Lett.*, **34**, 201–210.

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

G. Collomb, W. Härdle (1986). Strong uniform convergence rates in robust time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Process. Appl.*, **23**, 77–89.

T. Gasser, H. G. Müller, V. Mammitzsch. Kernels for nonparametric curve estimation. *J. R. Statist. Soc. B*, **47**, 238–252.

P. Hall (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multiv. Anal.*, **14**, 1–16.

W. Hoeffding (1948a). A non-parametric test of independence. *Ann. Math. Statist.*, **19**, 546–557.

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

P. M. Robinson (1991). Consistent nonparametric entropy-based testing. Review of Economic studies, **58**, 437–453.

M. Rosenblatt (1975). A quadratic measure of deviation of two dimensional density estimates and a test of independence. *Ann. Statist.*, **3**, 1–14 [Correction: 10 (1982), 646].

M. Rosenblatt, B. E. Wahlen (1992). A nonparametric measure of independence under the hypothesis of independent components. *Statist. Probab. Lett.*, **15**, 245–252.

J. X. Zheng (1997a). A consistent specification test of independence. *Nonparam. Statist.*, **7**, 297–306.

J. X. Zheng (1997b). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, **75**, 263–289.