ELSEVIER

Statistics & Probability Letters 34 (1997) 201-210

STATISTICS & PROBABILITY LETTERS

Testing independence by nonparametric kernel method

Ibrahim A. Ahmad[a,*], Qi Li[b]

[a] Division of Statistic, Northern Illinois University, Dekalb, IL 60115-2888, USA

[b] Department of Economics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

Received September 1995; revised August 1996

Abstract

Using nonparametric kernel estimation method, we propose a consistent test for independence of two random vectors based on the L_2 norm of difference between the joint density and the product of their marginals. A Monte Carlo study is carried out to examine the finite sample performance of the proposed test.

Keywords: Testing independence; Kernel estimation; Consistent tests; Monte Carlo simulation; Asymptotic normality

1. Introduction

Let $(X, Y)'$ be a $(p+q)\times 1$ random vector with joint distribution function (df)$F(x,y)$ defined on $\mathscr{S}\subset\mathscr{R}^{p+q}$ with a joint Lebesgue density $f(x,y)$. Further, let $F_1(x)(F_2(y))$ denote the marginal df of X (Y) with marginal density $f_1(x)(f_2(y))$. Consider the following hypothesis testing: $\mathrm{H}_0: f(x,y) = f_1(x)f_2(y)$ for almost all (x,y) against $\mathrm{H}_1: f(x,y) \neq f_1(x)f_2(y)$ on a set of positive measure. A measure of departure from H_0 is

$$I = \int\int [f(x,y) - f_1(x)f_2(y)]^2\,\mathrm{d}x\,\mathrm{d}y, \tag{1}$$

where the integrals are of dimensions p and q, respectively, and are taken over sets $\mathscr{S}_1\subset\mathscr{R}^p$ and $\mathscr{S}_2\subset\mathscr{R}^q$, $\mathscr{S}_1$ and $\mathscr{S}_2$ are the supports of X and Y, respectively. Note that $I\geqslant 0$ and $I = 0$ if and only if H_0 is true, hence I serves as a proper candidate for testing H_0. I can be written as

$$I = \int\int f(x,y)\,\mathrm{d}F(x,y) + \int f_1(x)\,\mathrm{d}F_1(x)\int f_2(y)\,\mathrm{d}F_2(y) - 2\int\int f_1(x)f_2(y)\,\mathrm{d}F(x,y).$$

Using a random sample $\{X_i, Y_i\}_{i=1}^n$, our test procedure is based on the following natural estimate of I:

$$\begin{aligned}\tilde{I}_n = {} & \int\int \hat{f}_n(x,y)\,\mathrm{d}F_n(x,y) + \int \hat{f}_{1n}(x)\,\mathrm{d}F_{1n}(x)\int \hat{f}_{2n}(y)\,\mathrm{d}F_{2n}(y) \\ & -2\int\int f_{1n}(x)f_{2n}(y)\,\mathrm{d}F_n(x,y),\end{aligned} \tag{2}$$

* Corresponding author.

PII S0167-7152(96)00183-6

where $\hat{f}_n(x,y) = (na^pb^q)^{-1}\sum_{i=1}^n k^x((x-X_i)/a)k^y((y-Y_i)/b)$, $\hat{f}_{1n}(x) = (na^p)^{-1}\sum_{i=1}^n k^x((x-X_i)/a)$, $\hat{f}_{2n}(y) = (nb^q)^{-1}\sum_{i=1}^n k^y((y-Y_i)/b)$, $F_n(x,y) = n^{-1}\sum_{i=1}^n I(X_i \leqslant x, Y_i \leqslant y)$, $F_{1n}(x) = n^{-1}\sum_{i=1}^n I(X_i \leqslant x)$ and $F_{2n}(y) = n^{-1}\sum_{i=1}^n I(Y_i \leqslant y)$. Writing $k_{ij}^x = k^x((X_i - X_j)/a)$ and $k_{ij}^y = k^y((Y_i - Y_j)/b)$, we can rewrite $\tilde{I}_n$ as follows:

$$\tilde{I}_n = (n^2a^pb^q)^{-1}\sum_i\sum_j k_{ij}^x k_{ij}^y + (n^4a^pb^q)^{-1}\left(\sum_i\sum_j k_{ij}^x\right)\left(\sum_l\sum_r k_{lr}^y\right)$$
$$-(2/n^3a^pb^q)\sum_i\sum_j\sum_l k_{ij}^x k_{jl}^y. \tag{3}$$

It can be shown that $\tilde{I}_n$ contains a degenerate U-statistic and several center terms. Using Lemma A.1 of the appendix, we will show that the U-statistic is of the order $O_p((na^{p/2}b^{q/2})^{-1})$ and has an asymptotically normal distribution. One can also show that the center terms are of the order $O_p((na^pb^q)^{-1})$, or $O_p((nb^q)^{-1})$ or $O_p((na^p)^{-1})$. Thus, some of the center terms have orders larger than or equal to that of the U-statistic. Hence, in order to derive the asymptotic distribution of $\tilde{I}_n$, these center terms need to be estimated and subtracted from $\tilde{I}_n$. However, in finite samples if the center terms and their estimates are not close, this can cause bias (sometimes substantial) and it is suggested that to do away with this bias, we modify $\tilde{I}_n$ and replace it by the following test statistic:

$$\hat{I}_n = (n^2a^pb^q)^{-1}\sum_i\sum_{j\neq i} k_{ij}^x k_{ij}^y + (n^4a^pb^q)^{-1}\left(\sum_i\sum_{j\neq i} k_{ij}^x\right)\left(\sum_l\sum_{r\neq l} k_{lr}^y\right)$$
$$-(2/n^3a^pb^q)\sum_i\sum_{j\neq i}\sum_{l\neq j} k_{ij}^x k_{jl}^y. \tag{4}$$

We will show under mild and usual assumptions on a, b, f, f_1, f_2, k^x and k^y, that $\hat{I}_n$ is, under H_0, asymptotically normal with mean zero and obtain its asymptotic variance and provide a consistent estimate of it.

The problem of testing independence using density estimation was first discussed by Rosenblatt (1975) and was improved by Rosenblatt and Wahlen (1992). Both have centering constants that depend on the unknown density f and/or its marginals f_1 and f_2. Both works are offered for the special case when $p = q = 1$ and $a = b$, and only considered the case of $f(x,y)$ is supported on all of $\mathscr{R}^2$. Also no Monte Carlo simulations were reported in these works, hence the finite sample performances of these tests are unknown. The present work may be viewed as further improvement to Rosenblatt and Wahlen (1992). We consider the general case where $\mathscr{S}_1$ and $\mathscr{S}_2$ are unions of convex subsets of $\mathscr{R}^p$ and $\mathscr{R}^q$, respectively ($\mathscr{S}_1 = \mathscr{R}^p$ and $\mathscr{S}_2 = \mathscr{R}^q$ are included as special case), and we allow $a \neq b$, also when $f_1(x)$ and/or $f_2(y)$ have compact support, we do not need to use special boundary kernel functions. Finally, we will investigate the finite sample performances of the proposed tests via Monte Carlo experiments.

2. Testing procedure for independence

In order to state our main results, first we state some needed assumptions.

(A1) $f(x,y)$, $f_1(x)$ and $f_2(y)$ are continuous and bounded on $\mathscr{S}$, $\mathscr{S}_1$ and $\mathscr{S}_2$, respectively. Also, $|f_1(x+u) - f_1(x)| \leqslant g_1(x)|u|$ for all $x, x+u \in \mathscr{S}_1$ and $|f_2(y+v) - f_2(y)| \leqslant g_2(y)|v|$ for all $y, y+v \in \mathscr{S}_2$. Both $g_1(X)$ and $g_2(Y)$ have finite expectations.

(A2) $k^x(\cdot):\mathscr{R}^p \to \mathscr{R}$ and $k^y(\cdot):\mathscr{R}^q \to \mathscr{R}$ are bounded, symmetric densities such that $\int k^x(u)uu'\,\mathrm{d}u = C_x I_p$, $\int k^y(v)vv'\,\mathrm{d}v = C_y I_q$, where C_x and C_y are positive constants and I_p and I_q are identity matrices.

(A3) As $n \to \infty$, $a \to 0$, $b \to 0$ and $na^p b^q \to \infty$.

Theorem 2.1. *Under* (A1)–(A3), *and under* H_0, *as* $n \to \infty$, $na^{p/2}b^{q/2}\hat{I}_n$ *is asymptotically normal with mean zero and variance* σ_0^2 *with*

$$\sigma_0^2 = 2R(f_1)R(f_2)R(k^x)R(k^y),$$

where $R(\phi) = \int \phi^2(u)\,\mathrm{d}u$. *With* $M(\hat{f}_{in}) = \int \hat{f}_{in}(x)\,\mathrm{d}F_{in}(x) = \frac{1}{n}\sum_{j=1}^n \hat{f}_{in}(X_j)$, $i = 1,2$; *a consistent estimate of* σ_0^2 *is* $\hat{\sigma}_0^2 = 2M(\hat{f}_{1n})M(\hat{f}_{2n})R(k^x)R(k^y)$, *leading to*

$$\hat{J}_n = na^{p/2}b^{q/2}\hat{I}_n/\hat{\sigma}_0 \text{ is asymptotically standard normal.}$$

The proof of Theorems 2.1 is given in the appendix.

Our Monte Carlo simulations showed (see below for details) the test statistic $\hat{J}_n$ has mean values (under H_0) quite close to zero, while a test with a center term (and subtracting the estimated center term) has substantial finite sample bias. These results clearly support the use of a test that does not have a center term. However, our Monte Carlo experiments also showed that the sample standard deviation of the $\hat{J}_n$ test is significantly smaller than 1, resulting in the estimated sizes are smaller than their nominal sizes. Hence, we propose a more accurate estimator for $\mathrm{var}(na^{p/2}b^{q/2}\hat{I}_n)$ given in the lemma below.

Lemma 2.1. *Another consistent estimate of* σ_0^2 *is given by*

$$\tilde{\sigma}_0^2 = 2\{M(\hat{f}_{1n})[R(k^x) - a^p\{2M(\hat{f}_{1n}^2) - (M(\hat{f}_{1n}))^2\}]\}\{M(\hat{f}_{2n})[R(k^y) - b^q\{2M(\hat{f}_{2n}^2) - (M(\hat{f}_{2n}))^2\}]\}.$$

Hence, $\hat{T}_n = na^{p/2}b^{q/2}\hat{I}_n/\tilde{\sigma}_0$ *is asymptotically standard normal as well. Moreover,*

$$\tilde{\sigma}_0^2 = \mathrm{var}(na^{p/2}b^{q/2}\hat{I}_n) + O_p(a^2+b^2) \text{ provided } n^{-1/2} = O(a^2+b^2).$$

The proof of Lemma 2.1 is also given in the appendix. Similar to the proof of Lemma 2.1, it is easy to show that $\hat{\sigma}_0^2 = \mathrm{var}(na^{p/2}b^{q/2}\hat{I}_n) + O_p(a+b)$, hence, $\tilde{\sigma}_0^2$ is a more accurate estimator of σ_0^2.

The consistency of our tests is given by the next theorem.

Theorem 2.2. *Under* (A1)–(A3), *and under* H_1, *as* $n \to \infty$, $P[\hat{J}_n > B_n] \to 1$ *and* $P[\hat{T}_n > B_n] \to 1$, *for any non-stochastic sequence* $\{B_n\}$ *such that* $B_n = o(na^{p/2}b^{q/2})$.

Proof. Theorem 2.2 follows from the facts that $\hat{I}_n \xrightarrow{p} I(>0)$ under H_1, and that $\hat{\sigma}_0^2 \xrightarrow{p} \sigma_0^2$ (under either H_0 or H_1).

Below we report on a small Monte Carlo simulation of $\hat{J}_n$ and $\hat{T}_n$ above. For ease of calculation, we take $p = q = 1$. We also computed a test statistic that has center terms, these center terms are estimated and subtracted from the original test. For $p = q = 1$ and $a = b$, one can show that $\tilde{I}_n = \hat{I}_n + (\tilde{I}_n - \hat{I}_n) = (na)^{-1}\sigma_0 Z + c(n) + o_p((na)^{-1})$, where Z denotes a standard-normal random variable and $c(n) = (na^2)^{-1}k^x(0)k^y(0) - (na)^{-1}\{k^y(0)Ef_1(X) + k^x(0)Ef_2(Y)\}$, the first part of the above result follows from Theorem 2.1, and the second part follows a similar proof of Rosenblatt and Wahlen (1992). Hence, the standardized test statistic (corresponds to $\tilde{I}_n$) is $\tilde{J}_n = na(\tilde{I}_n - \hat{c}(n))/\hat{\sigma}_0$, where $\hat{c}(n)$ is obtained from $c(n)$ with $Ef_1(X)$ $(Ef_2(Y))$ replaced by $n^{-1}\sum_{i=1}^n \hat{f}_{1n}(X_i)$ $(n^{-1}\sum_{i=1}^n \hat{f}_{2n}(Y_i))$. Then under H_0, $\tilde{J}_n$ is asymptotically standard normal. We first generate the standard bivariate normal random with mean zero, unit variance and a correlation ρ. Thus the

Table 1
Estimated sizes

	$\hat{J}_n$				$\hat{T}_n$				$\tilde{J}_n$			
n	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%
50	0.001	0.487	0.007	0.018	0.001	1.090	0.076	0.121	0.348	0.456	0.013	0.035
100	0.003	0.542	0.010	0.023	0.005	1.060	0.067	0.112	0.301	0.522	0.018	0.047
200	−0.001	0.599	0.017	0.034	−0.001	1.031	0.056	0.105	0.245	0.584	0.023	0.049

Table 2
Estimated power of the $\hat{J}_n$-test

	$\rho = 0.1$				$\rho = 0.3$				$\rho = 0.5$			
n	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%
50	0.040	0.490	0.007	0.021	0.320	0.630	0.041	0.079	1.04	0.910	0.219	0.337
100	0.077	0.571	0.014	0.034	0.602	0.769	0.099	0.182	1.98	1.16	0.553	0.690
200	0.115	0.643	0.021	0.048	1.09	0.946	0.243	0.363	3.66	1.49	0.925	0.967

Table 3
Estimated power of the $\hat{T}_n$-test

	$\rho = 0.1$				$\rho = 0.3$				$\rho = 0.5$			
n	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%
50	0.090	1.09	0.088	0.132	0.690	1.37	0.205	0.279	2.27	1.90	0.440	0.583
100	0.148	1.11	0.106	0.150	1.17	1.48	0.320	0.406	3.84	2.16	0.859	0.907
200	0.203	1.12	0.107	0.156	1.90	1.65	0.510	0.604	6.41	2.54	0.987	0.992

Table 4
Estimated power of the $\tilde{J}_n$-test

	$\rho = 0.1$				$\rho = 0.3$				$\rho = 0.5$			
n	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%
50	0.386	0.422	0.015	0.049	0.652	0.586	0.066	0.131	1.34	0.840	0.302	0.468
100	0.312	0.550	0.025	0.069	0.883	0.739	0.151	0.249	2.22	1.11	0.645	0.804
200	0.367	0.630	0.032	0.085	1.32	0.926	0.313	0.460	3.85	1.46	0.955	0.981

null hypothesis of independence corresponds to $\rho = 0$. For alternatives, we choose $\rho = 0.1$ or 0.3 or 0.5. We select k^x and k^y to be standard normal. The smoothing parameters are chosen by $a = b = cn^{-1/5}$, $c = 0.8$, or 1 or 1.2. The results for different c values are quite similar. Hence, we only report the case of $c = 1$ to save space. The sample sizes are $n = 50, 100, 200$. For $\rho = 0$, the number of replications are 5000 for all different values of n. For $\rho > 0$, the number of replications are 2000 for all cases. The estimated sizes and powers (the percentage of rejecting the null hypothesis) of $\hat{J}_n$, $\hat{T}_n$ and $\tilde{J}_n$ tests are given in Tables 1–4.

In Tables 1–4, Ave. and Std. are the mean value and standard deviation of the test statistic. Table 1 gives the estimated size. Observe first that both $\hat{J}_n$ and $\hat{T}_n$ have mean values quite close to zero. This is because we used 'leave-one-out' method to compute $\hat{J}_n$ and $\hat{T}_n$ (i.e., $i \neq j$ and $l \neq r$ in $k^x_{ij}k^y_{lr}$). In contrast, $\tilde{J}_n$

Table 5
Estimated sizes of the $\hat{T}_n$-test

	$n = 50$				$n = 100$				$n = 200$			
Distribution	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%
t(5)-distri.	0.004	1.04	0.078	0.118	0.041	1.09	0.068	0.112	0.001	1.02	0.062	0.108
Mixture	−0.023	1.07	0.069	0.111	−0.015	1.06	0.065	0.110	−0.013	0.998	0.059	0.105

Table 6
Estimated power of the $\hat{T}_n$-test (t(5)-distribution)

	$\rho = 0.1$				$\rho = 0.3$				$\rho = 0.5$			
n	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%
50	0.035	1.08	0.079	0.124	0.441	1.25	0.157	0.219	1.52	1.61	0.422	0.515
100	0.105	1.07	0.081	0.132	0.807	1.31	0.243	0.315	2.64	1.82	0.675	0.786
200	0.141	1.09	0.091	0.142	1.34	1.44	0.356	0.467	4.45	2.120	0.928	0.955

Table 7
Estimated power of the $\hat{T}_n$-test (mixture)

	$\rho = 0.1$				$\rho = 0.3$				$\rho = 0.5$			
n	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%	Ave.	Std.	5%	10%
50	0.167	1.08	0.995	0.154	0.588	1.19	0.182	0.261	1.49	1.47	0.424	0.513
100	0.336	1.11	0.118	0.195	1.01	1.26	0.292	0.377	2.49	1.59	0.688	0.769
200	0.579	1.13	0.172	0.257	1.71	1.34	0.480	0.596	4.23	1.78	0.940	0.968

has an obvious bias term due to the finite sample difference between $(\tilde{I}_n - \hat{I}_n)$ and $\hat{c}(n)$. We also observe that $\hat{\sigma}_0^2$ overestimates $\mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n)$, hence, the estimated standard deviations of $\tilde{J}_n$ and $\hat{J}_n$ are much less than unity for all cases considered. From Lemma 2.1, we know that $\hat{\sigma}_0^2 = \mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n) + O_p(a+b)$ and $\tilde{\sigma}_0^2 = \mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n) + O_p(a^2+b^2)$. Hence, $\tilde{\sigma}_0^2$ gives a more accurate estimate of $\mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n)$ and indeed Table 1 shows that the standard deviation of $\hat{T}_n$ are very close to unity for all cases.

We observe that the estimated sizes of the $\hat{J}_n$ and the $\tilde{J}_n$ tests are significantly smaller than their nominal sizes, mainly due to their over estimating $\mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n)$. Although their performances improve as n increases, they still substantially underestimate the nominal sizes for $n = 200$. Note that $\tilde{J}_n$ has slightly better estimated sizes than $\hat{J}_n$ because the positive bias term in $\tilde{J}_n$ helps it to take larger values than $\hat{J}_n$. In contrast, the $\hat{T}_n$ test gives much better estimated sizes, and its performance is quite satisfactory when $n = 200$. Recall that $\hat{\sigma}_0^2 = \mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n) + O_p(a+b)$ and $\tilde{\sigma}_0^2 = \mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n) + O_p(a^2+b^2)$. Therefore, we recommend using the more accurate estimator $\tilde{\sigma}_0^2$ in practice. Tables 2–4 report the estimated power for $\hat{J}_n$, $\hat{T}_n$ and $\tilde{J}_n$. As expected, we observe that $\hat{T}_n$ is the most powerful one among the three. The positive bias term in $\tilde{J}_n$ helps it to be more powerful than $\hat{J}_n$.

We also generated two non-normal bivariate random variables to examine the robustness of the above results to different distributions. We generated (i) both X and Y have bivariate t-distribution with degree of freedom equal 5. Under H_0, X and Y are independent, and under H_1, the correlation between Z_1 and Z_2 is ρ ($\rho = 0.1, 0.3, 0.5$), where Z_1 and Z_2 are standard normal variates that are defined from $X_1 = Z_1/\sqrt{V_1/5}$ and

$X_2 = Z_2/\sqrt{V_2/5}$, V_1 and V_2 are both $\chi^2(5)$-random variables and independent of each other. (ii) we generated a case where both X and Y are bivariate mixture normal random variable, with $X = pN(0,1)+(1-p)N(1,2)$ and $Y = pN(0,1)+(1-p)N(1,2)$ ($p = 0.4$). Under H_0, X and Y are independently generated, and under H_1, the covariance between X and Y is ρ ($\rho = 0.1, 0.3, 0.5$). The performances of all the three tests are quite similar to the case of normality of X and Y. For example, the $\hat{J}_n$ test significantly overestimated the $\mathrm{var}(na^{1/2}b^{1/2}\hat{I}_n)$, and hence it underestimated the sizes; also the $\tilde{J}_n$ test have significant bias due to the difference between the center term and its estimate. In comparison, the $\hat{T}_n$ performs much better. To save space, we only report the estimated size and power for the $\hat{T}_n$ test. The sample sizes are $n = 50$, 100, 200, the number of replications are 2000 and the smoothing parameters are chosen via $a = x_{sd}n^{-1/5}$ and $b = y_{sd}n^{-1/5}$, where x_{sd} and y_{sd} are the sample standard deviations of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, respectively. These results are given in Tables 5–7. The results are quite similar to those given in Tables 1 and 3, suggesting that our test is quite robust to different distributions of X and Y.

In summary we observe that the finite sample performances of $\tilde{J}_n$ is not satisfactory mainly because of (i) the bias due to the center terms, and (ii) its overestimate of $\mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n)$. The leave-one-out method successfully removes the bias from the $\hat{J}_n$ and the $\hat{T}_n$ tests. Also the more accurate estimator $\tilde{\sigma}_0^2$ gives quite accurate estimate of $\mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n)$. Thus, the $\hat{T}_n$ test overcomes both of the above problems.

An issue that has not been addressed in this paper is how to choose the smoothing parameters a and b optimally in the sense that the power of the proposed tests can be maximized. This is an important question that certainly deserves research efforts. We leave it for a future research project of ours.

Acknowledgements

We would like to thank two referees for their helpful comments that greatly improved the paper. Research of the first author was partially supported by a Fulbright Grant while research of the second author was supported by the Social Sciences and Humanities Research Council of Canada.

Appendix A

First we give a lemma that will be used to prove Theorem 2.1.

Lemma A.1. *Let*

$$U_n = \binom{n}{k}^{-1} \sum_{C} \psi_n(Z_{i_1},\ldots,Z_{i_k}) \tag{A.1}$$

be a k-order U*-statistic, where ψ_n is a symmetric function (exchangeable) that depends on n, $Z_1,\ldots,Z_n$ are independent and identically distributed random variables (or vectors), and the sum $\sum_C$ extends over all combinations $1 \leqslant i_1 < \cdots < i_k \leqslant n$ of $\{1,\ldots,n\}$. Assume that $E\{\psi_n(Z_1,\ldots,Z_k)|Z_1\} = 0$ almost surely and define $\psi_{nc}(z_1,\ldots,z_c) = E\{\psi_n(Z_1,\ldots,Z_k)|Z_1 = z_1,\ldots,Z_c = z_c\}$ ($c \leqslant k$) and their variances are $\sigma_{nc}^2 =$* $\mathrm{Var}\{\psi_{nc}(Z_1,\ldots,Z_c)\}$. *Further, define $G_n(Z_1,Z_2) = E[\psi_{n2}(Z_1,Z_3)\psi_{n2}(Z_2,Z_3)|Z_1,Z_2]$. If $E[\psi_n^2(Z_1,\ldots,Z_k)] < \infty$ for each n, $\sigma_{nc}^2/\sigma_{n2}^2 = o(n^{(c-2)})$ for $c = 3,\ldots,k$ (where $k \geqslant 3$) and as $n \to \infty$*

$$\frac{E[G_n^2(Z_1,Z_2)] + n^{-1}E[\psi_n^4(Z_1,Z_2)]}{\{E[\psi_{n2}^2(Z_1,Z_2)]\}^2} \to 0, \tag{A.2}$$

then nU_n is asymptotically normal with mean zero and var $\{k^2(k-1)^2\sigma_{n2}^2/2\}$.

Proof. This is Lemma B.4 of Fan and Li (1996). Note that when $k = 2$, Lemma A.1 reduces to Theorem 1 of Hall (1984).

Proof of Theorem 2.1. Recall the definition of $\hat{I}_n$ in (4), we can rewrite it as follows:

$$\hat{I}_n = \frac{1}{n^4 a^p b^q} \sum\sum\sum \sum_{i \neq j \neq l \neq r} k_{ij}^x [k_{ij}^y + k_{lr}^y - 2k_{jl}^y] + \hat{I}_{2n}$$

$$= \hat{I}_{1n} + \hat{I}_{2n}, \text{ say}, \tag{A.3}$$

where $\sum\sum\sum\sum_{i \neq j \neq l \neq r}$ denotes the sum over all different arrangement of i, j, l, r. Also, it is easy to see that $E|\hat{I}_{2n}| = (n^4 a^p b^q)^{-1} \mathrm{O}(n^3 a^p b^q) = \mathrm{O}(n^{-1})$. Hence, $\hat{I}_{2n} = \mathrm{O}_p(n^{-1})$. Denote $Z_i = (X_i, Y_i)$ and write $P_n(Z_i, Z_j, Z_l, Z_r) = \sum_{4!} k_{ij}^x [k_{ij}^y + k_{lr}^y - 2k_{jl}^y]$ with $\sum_{4!}$ extends over 4! different arrangements of i, j, l, r. Thus, $\hat{I}_{1n}$ can be written as a U-statistic as follows:

$$\hat{I}_{1n} = \binom{n}{4} \Big/ n^4 a^p b^q \left[\binom{n}{4}^{-1} \sum_{1 \leqslant i<j<l<r \leqslant n} P_n(Z_i, Z_j, Z_l, Z_r) \right]. \tag{A.4}$$

We will show that the conditions of Lemma A.1 above hold. First we show that $E[P_n(Z_i, Z_j, Z_l, Z_r)|Z_i] = 0$ almost surely, and if we put $P_{n2}(Z_1, Z_2) = E[P_n(Z_1, Z_2, Z_3, Z_4)|Z_1, Z_2]$, $G_n(Z_1, Z_2) = E[P_{n2}(Z, Z_1) P_{n2}(Z, Z_2)|Z_1, Z_2]$ and let σ_{nc}^2 be as in Lemma A.1 above (with $k = 4$) with $P_n(\cdot)$ replacing $\psi_n(\cdot)$, then we shall also show that $E[P_{n2}^4(Z_1, Z_2)] = \mathrm{O}(a^p b^q)$, $E[P_{n2}^2(Z_1, Z_2)] = \sigma_{n2}^2 = \mathrm{O}(a^p b^q)$, $E[G_n^2(Z_1, Z_2)] = \mathrm{O}(a^{3p} b^{3q})$, $\sigma_{n3}^2 = \mathrm{O}(a^p b^q)$ and $\sigma_{n4}^2 = \mathrm{O}(a^p b^q)$. Once the above are proved, then Lemma A.1 gives that $nU_n = n\binom{n}{4}^{-1} \sum_{1 \leqslant i<j<l<r \leqslant n} P_n(Z_i, Z_j, Z_l, Z_r)$ is asymptotically normal with zero mean and variance equal to $\{4^2(4-1)^2 \mathrm{Var}(P_{n2}(Z_1, Z_2))/2\}$ from which we deduce the asymptotic normality of $n a^{p/2} b^{q/2} \hat{I}_{1n} = ((\binom{n}{4})/n^4 a^{p/2} b^{q/2})[nU_n]$. Let us begin our computations. First, we need to compute $P_n(Z_i, Z_j)$. To do so, we need to write all 24 terms of $P_n(Z_i, Z_j, Z_l, Z_r)$. To help derive the expression of $P_n(Z_i, Z_j, Z_l, Z_r)$, we use $i = 1$, $j = 2$, $l = 3$, and $r = 4$. We distinguish four cases:

(I) (1234), (1243), (1324), (1342), (1423), (1432), i.e., $(ijlr)$, $(ijrl)$, $(iljr)$, $(ilrj)$, $(irjl)$, $(irlj)$;
(II) (2134), (2143), (2314), (2341), (2413), (2431), i.e., $(jilr)$, $(jirl)$, $(jlir)$, $(jlri)$, $(jril)$, $(jrli)$;
(III) (3124), (3142), (3214), (3241), (3412), (3421), i.e., $(lijr)$, $(lirj)$, $(ljir)$, $(ljri)$, $(lrij)$, $(lrji)$;
(IV) (4123), (4132), (4213), (4231), (4312), (4321), i.e., $(rijl)$, $(rilj)$, $(rjil)$, $(rjli)$, $(rlij)$, $(rlji)$.

Below we give the sum of each six terms of $k_{ij}^x [k_{lr}^y + k_{ij}^y - 2k_{jl}^y]$ for cases (I)–(IV):

Case (I): The sum $(1) = k_{ij}^x[k_{lr}^y + k_{ij}^y - 2k_{jl}^y] + k_{ij}^x[k_{rl}^y + k_{ij}^y - 2k_{jr}^y] + k_{il}^x[k_{jr}^y + k_{il}^y - 2k_{lj}^y] + k_{il}^x[k_{rj}^y + k_{il}^y - 2k_{lr}^y] + k_{ir}^x[k_{jl}^y + k_{ir}^y - 2k_{rj}^y] + k_{ir}^x[k_{lj}^y + k_{ir}^y - 2k_{rl}^y]$.

Case (II): The sum $(2) = k_{ji}^x[k_{lr}^y + k_{ji}^y - 2k_{il}^y] + k_{ji}^x[k_{rl}^y + k_{ji}^y - 2k_{ir}^y] + k_{jl}^x[k_{ir}^y + k_{jl}^y - 2k_{li}^y] + k_{jl}^x[k_{ri}^y + k_{jl}^y - 2k_{lr}^y] + k_{jr}^x[k_{il}^y + k_{jr}^y - 2k_{ri}^y] + k_{jr}^x[k_{li}^y + k_{jr}^y - 2k_{rl}^y]$.

Case (III): The sum $(3) = k_{li}^x[k_{jr}^y + k_{li}^y - 2k_{ij}^y] + k_{li}^x[k_{rj}^y + k_{li}^y - 2k_{ir}^y] + k_{lj}^x[k_{ir}^y + k_{lj}^y - 2k_{ji}^y] + k_{lj}^x[k_{ri}^y + k_{lj}^y - 2k_{jr}^y] + k_{lr}^x[k_{ij}^y + k_{lr}^y - 2k_{ri}^y] + k_{lr}^x[k_{ji}^y + k_{lr}^y - 2k_{rj}^y]$.

Case (IV): The sum $(4) = k_{ri}^x[k_{jl}^y + k_{ri}^y - 2k_{ij}^y] + k_{ri}^x[k_{lj}^y + k_{ri}^y - 2k_{il}^y] + k_{rj}^x[k_{il}^y + k_{rj}^y - 2k_{ji}^y] + k_{rj}^x[k_{li}^y + k_{rj}^y - 2k_{jl}^y] + k_{rl}^x[k_{ij}^y + k_{rl}^y - 2k_{li}^y] + k_{rl}^x[k_{ji}^y + k_{rl}^y - 2k_{lj}^y]$.

Next, we compute the expected value of each sum above conditional on Z_i and Z_j. Note that under H_0, X_i and Y_j are all independent, $i,j = 1,\dots,n$. Thus,

$$\begin{aligned}
E\{\mathrm{Sum}(1)|Z_i,Z_j\} &= E\{k^x_{ij}[k^y_{34} + k^y_{ij} - 2k^y_{j3}] + k^x_{ij}[k^y_{43} + k^y_{ij} - 2k^y_{j4}] \\
&\quad + k^x_{i3}[k^y_{j4} + k^y_{i3} - 2k^y_{j3}] + k^x_{i3}[k^y_{4j} + k^y_{i3} - 2k^y_{34}] \\
&\quad + k^x_{i4}[k^y_{j3} + k^y_{i4} - 2k^y_{4j}] + k^x_{i4}[k^y_{3j} + k^y_{i4} - 2k^y_{43}]|Z_i,Z_j\} \\
&= k^x_{ij}E[k^y_{34} + k^y_{ij} - 2k^y_{j3} + k^y_{43} + k^y_{ij} - 2k^y_{j4}|Y_i,Y_j] \\
&\quad + E[k^x_{i3}|X_i]E[k^y_{j4} + k^y_{i3} - 2k^y_{j3} + k^y_{4j} + k^y_{i3} - 2k^y_{34} \\
&\quad + k^y_{j3} + k^y_{i4} - 2k^y_{4j} + k^y_{3j} + k^y_{i4} - 2k^y_{43}|Y_i,Y_j] \\
&= 2k^x_{ij}E[k^y_{34} + k^y_{ij} - 2k^y_{j3}|Y_i,Y_j] + 4E[k^x_{i3}|X_i]E[k^y_{i3} - k^y_{34}|Y_i,Y_j]. \qquad \text{(A.5)}
\end{aligned}$$

In exactly a similar fashion one can easily show that

$$\begin{aligned}
E\{\mathrm{Sum}(2)|Z_i,Z_j\} &= 2k^x_{ji}E[k^y_{34} + k^y_{ji} - 2k^y_{i3}|Y_i,Y_j] + 4E[k^x_{j3}|X_j][k^y_{j3} - k^y_{34}|Y_i,Y_j], \\
E\{\mathrm{Sum}(3)|Z_i,Z_j\} &= 2E[k^x_{3i}|X_i]E[k^y_{j4} - k^y_{ij}|Y_i,Y_j] + 2E[k^x_{3j}|X_j]E[k^y_{i4} - k^y_{ji}|Y_i,Y_j] \\
&\quad + 2E[k^x_{34}]E\{[k^y_{ij} + k^y_{34} - k^y_{4i} - k^y_{4j}]|Y_i,Y_j\}
\end{aligned}$$

and

$$\begin{aligned}
E\{\mathrm{Sum}(4)|Z_i,Z_j\} &= 2E[k^x_{4i}|X_i|X_i]E[k^y_{j3} - k^y_{ij}|Z_i,Z_j] + 2E[k^x_{4j}|X_j][k^y_{i3} - k^y_{ji}|Y_i,Y_j] \\
&\quad + 2E[k^x_{43}][k^y_{ij} + k^y_{43} - k^y_{3i} - k^y_{3j}|Y_i,Y_j].
\end{aligned}$$

Hence, adding all four sums, we get that (after some simplifications):

$$\begin{aligned}
P_n(Z_i,Z_j) &= E\{P_n(Z_i,Z_j,Z_l,Z_r)|Z_i,Z_j\} \\
&= 4E\{[k^x_{ij} + k^x_{34} - k^x_{i3} - k^x_{j3}]|X_i,X_j\}E\{[k^y_{ij} + k^y_{34} - k^y_{i3} - k^y_{j3}]|Y_i,Y_j\}. \qquad \text{(A.6)}
\end{aligned}$$

Let us now show that $E\{P_n(Z_i,Z_j,Z_l,Z_r)|Z_i\} = 0$. We have

$$\begin{aligned}
E\{P_n(Z_i,Z_j,Z_l,Z_r)|Z_i\} &= E\{P_n(Z_i,Z_j)|Z_i\} \\
&= 4E\{[k^x_{i2} + k^x_{34} - k^x_{i3} - k^x_{23}]|X_i\}E\{[k^y_{i2} + k^y_{34} - k^y_{i3} - k^y_{23}]|Y_i\} \\
&= 0, \qquad \text{(A.7)}
\end{aligned}$$

since $E(k^x_{12}|X_1) = E(k^x_{13}|X_1)$ and $E(k^x_{34}) = E(k^x_{23})$ by identical distribution. Next, let us evaluate $E[P_n^2(Z_1,Z_2)]$:

$$\begin{aligned}
E[P_n^2(Z_1,Z_2)] &= 16E\{[E(k^x_{12} + k^x_{34} - k^x_{i1} - k^x_{23}|X_1,X_2)]^2[E(k^y_{12} + k^y_{34} - k^y_{13} - k^y_{23}|Y_1,Y_2)]^2\} \\
&= 16E\{[k^x_{12} + E(k^x_{34}) - E(k^x_{13}|X_1) - E(k^x_{23}|X_2)]^2\}E\{[k^y_{12} + E(k^y_{34}) - E(k^y_{13}|Y_1) - E(k^y_{23}|Y_2)]^2\} \\
&= 16a^pb^q\{Ef_1(X_1)Ef_2(Y_1)R(k^x)R(k^y) + o(1)\}, \qquad \text{(A.8)}
\end{aligned}$$

since using the fact that $[k_{12}^x + E(k_{34}^x) - E(k_{13}^x|X_1) - E(k_{23}^x|X_2)] = k_{12}^x + a^p[E(f_1(X_3)) - f_1(X_1) - f_1(X_2) + o(a^p)]$ we get

$$\begin{aligned}
&E\{[k_{12}^x + E(k_{34}^x) - E(k_{13}^x|X_1) - E(k_{23}^x|X_2)]^2\} \\
&\quad = E\{[k_{12}^x]^2 + a^{2p}[Ef_1(X_3) - f_1(X_1) - f_1(X_2) + o(a^p)]^2 \\
&\qquad + 2k_{12}^x a^p[Ef_1(X_3) - f_1(X_1) - f_1(X_2) + o(a^p)]\} \\
&\quad = E(k_{12}^x)^2 + O(a^{2p}) \\
&\quad = a^p Ef_1(X_1)R(k^x) + o(a^p).
\end{aligned} \tag{A.9}$$

It is easy to show that $E[P_{n2}^4(Z_1,Z_2)] = O(a^p b^q)$, $\sigma_{n3}^2 = O(a^p b^q)$ and $\sigma_{n4}^2 = O(a^p b^q)$. Finally, one can show after some algebra that $E[G_n^2(Z_1,Z_2)] = O(a^{3p}b^{3q})$, since a typical term in $G_n(Z_1,Z_2)$ involves $\int\int f_1(x)f_2(y)$ $k^x((X_1-x)/a)k^x((X_2-x)/a)k^y((Y_1-y)/b)k^y((Y_2-y)/b)\,dx\,dy = a^p b^q \int\int f_1(X_1-au)f_1(Y_2-bv)k^x(u)k^x$ $((X_2-X_1)/a+u)k^y(v)k^y((Y_2-Y_1)/b+v)\,du\,dv$. Hence by Lemma A.1 above, it follows that

$$na^{p/2}b^{q/2}\hat{I}_{1n} = \frac{\binom{n}{4}}{n^4 a^{p/2} b^{q/2}} \left[n \binom{n}{4}^{-1} \sum_{1 \leqslant i<j<l<k \leqslant n} P_n(Z_i,Z_j,Z_l,Z_k) \right], \tag{A.10}$$

converges to normal with mean zero and variance $((\binom{n}{4})^2/n^8 a^p b^q)2^{-1}4^2(4-1)^2[16a^p b^q \sigma_0^2] \to \sigma_0^2$.

We emphasize here that for the above result to hold, we do not need the support of $f_1(x)$ and $f_2(y)$ to be the whole Euclidean space $\mathscr{R}^p$ and $\mathscr{R}^q$. Clearly, $\hat{I}_{2n} = O_p(n^{-1})$ and $E(na^{p/2}b^{q/2}\hat{I}_{1n}) = 0$, regardless of the range of the supports of $f_1(x)$ and $f_2(y)$. Also, one can easily show that $a^{-p}E[(k_{12}^x)^2] = Ef_1(X)R(k^x) + o(1)$ even when $\mathscr{S}_1$ (the support of X) is a union of convex subsets of $\mathscr{R}^p$. Similarly, $b^{-q}E[(k_{12}^y)^2] = Ef_2(Y)R(k^y) + o(1)$ when $\mathscr{S}_2$ is a union of convex subsets of $\mathscr{R}^q$. Hence, $\mathrm{Var}(na^{p/2}b^{q/2}\hat{I}_{1n}) = \sigma_0^2 + o(1)$ when $\mathscr{S}$ is a union of convex subsets of $\mathscr{R}^{p+q}$ which includes the whole Euclidean space of $\mathscr{R}^{p-q}$ as a special case.

Proof of Lemma 2.1. We will only prove the case of $p = q = 1$. It is easy to see that $\hat{\sigma}_0^2 = \mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n) + O_p(a+b)$. We now show that $\tilde{\sigma}_0^2 = \mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n) + O(a^2+b^2)$ if $n^{-1/2} = O(a^2+b^2)$, and $f_1(x)$ and $f_2(y)$ are second-order differentiable with second-order derivatives bounded by functions that have finite expectations. We already observe that (see (A.9))

$$\begin{aligned}
&E\{[k_{12}^x + E(k_{34}^x) - E(k_{13}^x|X_1) - E(k_{23}^x|X_2)]^2\} \\
&\quad = E\{(k_{12}^x)^2 + a^2[E(f_1(X_3)) - f_1(X_1) - f_1(X_2) + O(a^2)]^2 \\
&\qquad + 2k_{12}^x a[E(f_1(X_3)) - f_1(X_1) - f_1(X_2)] + O(a^3)\},
\end{aligned}$$

and that $E[(k_{12}^x)^2] = aEf_1(X_1)R(k^x) + O(a^3)$. It is easy to see that

$$a^2E[E(f_1(X_3)) - f_1(X_1) - f_1(X_2)]^2 = a^2\{2E[f_1^2(X_1)] - (E[f_1(X_1)])^2\}$$

and

$$2aE\{k_{12}^x[Ef_1(X_3) - f_1(X_1) - f_1(X_2)]\} = 2a^2\{[E(f_1(X_1))]^2 - 2Ef_1^2(X_1) + O(a)\}.$$

Hence,

$$\begin{aligned}
&E\{[k_{12}^x + E(k_{34}^x) - E(k_{13}^x|X_1) - E(k_{23}^x|X_2)]^2\} \\
&\quad = a\{Ef_1(X_1)R(k^x) - a[2Ef_1^2(X_1) - (Ef_1(X_1))^2] + O(a^2)\}.
\end{aligned} \tag{A.11}$$

Similarly, we have

$$E\{[k_{12}^{y} + E(k_{34}^{y}) - E(k_{13}^{y}|X_1) - E(k_{23}^{y}|X_2)]^2\}$$
$$= b\{Ef_2(Y_1)R(k^y) - b[2Ef_2^2(Y_1) - (Ef_2(Y_1))^2] + \mathrm{O}(b^2)\}. \tag{A.12}$$

Substituting (A.11) and (A.12) into (A.8), we get

$$E[P_{n2}(Z_1,Z_2)] = 16ab\{Ef_1(X_1)R(k^x) - a[2Ef_1^2(X_1) - (Ef_1(X_1))^2]\}\{Ef_2(Y_1)R(k^y)$$
$$-b[2Ef_2^2(Y_1) - (Ef_2(Y_1))^2]\} + ab\mathrm{O}(a^2 + b^2). \tag{A.13}$$

Hence, if we let $E_n(f_1(X_1)) = \int \hat{f}_{1n}(x)\,\mathrm{d}F_{1n}(x)$ and $E_n(f_1^2(X_1)) = \int \hat{f}_{1n}^2(x)\,\mathrm{d}F_{1n}(x)$ and define estimates $E_nf_2(Y_1)$ and $E_nf_2^2(Y_1)$, similarly, we can see that plugging these estimates on the right-hand side of (A.13) gives $\tilde{\sigma}_0^2 = 2(ab)^{-1}16EP_{n2}^2(Z_1,Z_2) + \mathrm{O}_p(a^2 + b^2 + n^{-1/2})$, since under quite general conditions $E_n(f_i) = E(f_i) + \mathrm{O}_p(n^{-1/2})$, cf. Ahmad (1982). This last result together with $E|\hat{I}_{2n}| = \mathrm{O}(n^{-1})$ implies that $\tilde{\sigma}_0^2 = \mathrm{Var}(na^{1/2}b^{1/2}\hat{I}_n) + \mathrm{O}_p(a^2 + b^2)$ provided $n^{-1/2} = \mathrm{O}(a^2 + b^2)$.

References

Ahmad, I.A. (1982), Nonparametric estimation of the location and scale parameters based on density estimation, *Ann. Inst. Statist. Math.* **34**, 39-53.

Fan, Y. and Q. Li (1996), Consistent model specification tests: Omitted variables and semiparametric functional forms, *Econometrica*, **64**, 865-890.

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

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

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