Tests for independence in nonparametric regression (supplement)

John H.J. Einmahl *Ingrid Van Keilegom*

Proof of (2.17) From (2.10) we have with high probability for large n and uniformly in x and y

$$\left.\begin{array}{l} \sqrt{n}(F_n(x,y) - \hat{F}_X(x)\hat{G}(y)) \\ \leq \widetilde{\alpha}_n\left(x, y + \frac{\log^2 n}{n}\right) - G(y)\alpha_n(x,\infty) \\ \quad -\hat{F}_X(x)\widetilde{\alpha}_n\left(\infty, y - \frac{\log^2 n}{n}\right) + 2C\frac{\log^2 n}{\sqrt{n}}, \\ \\ \sqrt{n}(F_n(x,y) - \hat{F}_X(x)\hat{G}(y)) \\ \geq \widetilde{\alpha}_n\left(x, y - \frac{\log^2 n}{n}\right) - G(y)\alpha_n(x,\infty) \\ \quad -\hat{F}_X(x)\widetilde{\alpha}_n\left(\infty, y + \frac{\log^2 n}{n}\right) - 2C\frac{\log^2 n}{\sqrt{n}}. \end{array}\right\} \tag{0.1}$$

Set $V_{n,0} = \sqrt{n}(F_n - \hat{F}_X\hat{G})$. From (2.12) and Proposition 2.1, we have, using the Skorohod construction for (2.12) (but keeping the same notation),

$$\sup_{\substack{x\in D_X \\ y\in I\!R}} |\alpha_n(x,y) - V(x,y)| \to 0 \text{ a.s.} \tag{0.2}$$

and

$$\sup_{\substack{x\in D_X \\ y\in I\!R}} |V_{n,0}(x,y) - V_0(x,y)| \to 0 \text{ a.s.} \tag{0.3}$$

Set

$$M(x,y) = F_X(x)G(y)(1-F_X(x))(1-G(y))$$

and

$$\hat{M}(x,y) = \hat{F}_X(x)\hat{G}(y)(1-\hat{F}_{X-}(x))(1-\hat{G}_-(y)).$$

Let $0 < \varepsilon < \frac{1}{4}$ be arbitrary and let $\delta(\varepsilon) > 0$ be a function of ε to be chosen later on, such that $\lim_{\varepsilon\downarrow 0}\delta(\varepsilon) = 0$. Denote with $q_{1\varepsilon}$ and $\widetilde{q}_{1\varepsilon}$ the $\delta(\varepsilon)$-th and $(1-\delta(\varepsilon))$-th quantiles of F_X, respectively, and with $q_{2\varepsilon}$, $\widetilde{q}_{2\varepsilon}$ the same quantiles of G. Write $S_\varepsilon = (q_{1\varepsilon}, \widetilde{q}_{1\varepsilon}) \times (q_{2\varepsilon}, \widetilde{q}_{2\varepsilon})$. We have

$$\left| \iint_{S_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{M}(x,y)} d\hat{F}_X(x)d\hat{G}(y) - \iint_{S_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x)dG(y) \right|$$

$$
\begin{aligned}
\le\;& \iint_{S_\varepsilon} \frac{|V_{n,0}^2(x,y) - V_0^2(x,y)|}{\hat{M}(x,y)} d\hat{F}_X(x) d\hat{G}(y) \\
&+ \iint_{S_\varepsilon} \frac{|M(x,y) - \hat{M}(x,y)|}{\hat{M}(x,y)M(x,y)} V_0^2(x,y) d\hat{F}_X(x) d\hat{G}(y) \\
&+ \left| \iint_{S_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} (d\hat{F}_X(x) d\hat{G}(y) - dF_X(x) dG(y)) \right|.
\end{aligned}
$$

From (0.3) and (0.2) we now see that the first and second term on the right converge to 0 a.s. The a.s. convergence to 0 of the third term follows from the Helly-Bray theorem.

Set $A_\varepsilon = I\!\!R^2 \backslash S_\varepsilon$. In view of what we just proved, it is now sufficient for the proof of (2.17) to show that for large n

$$
P\left(\iint_{A_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{M}(x,y)} d\hat{F}_X(x) d\hat{G}(y) \ge 17\varepsilon \right) \le 4\varepsilon \tag{0.4}
$$

and

$$
P\left(\iint_{A_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x) dG(y) \ge \varepsilon \right) \le \varepsilon. \tag{0.5}
$$

We have

$$
\frac{EV_0^2(x,y)}{M(x,y)} \;=\; \frac{G(y)(1-G(y)) + 2(H_1(y,y) - G^2(y)) + 2(H_2(y,y) - G^2(y))}{G(y)(1-G(y))} \le 5.
$$

So by the Markov inequality we obtain

$$
\begin{aligned}
& P\left(\iint_{A_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x) dG(y) \ge \varepsilon \right) \\
\le\;& \frac{1}{\varepsilon} E \iint_{A_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x) dG(y) = \frac{1}{\varepsilon} \iint_{A_\varepsilon} \frac{EV_0^2(x,y)}{M(x,y)} dF_X(x) dG(y) \\
\le\;& \frac{5}{\varepsilon} \iint_{A_\varepsilon} dF_X(x) dG(y) \le \frac{20\delta(\varepsilon)}{\varepsilon} \le \varepsilon,
\end{aligned}
$$

for $\delta(\varepsilon) \le \varepsilon^2/20$. This is (0.5).

Set $B_\varepsilon = A_\varepsilon \cap ((-\infty, m_1) \times (-\infty, m_2))$, with m_1 and m_2 the medians of F_X and G, respectively. Because of symmetry and because of (2.13) and (2.14), (0.4) follows from

$$
P\left(\iint_{B_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y) \ge \varepsilon \right) \le \varepsilon, \tag{0.6}
$$

for large n. Let Q_{1n} and Q_{2n} be the empirical quantile functions corresponding to $\hat{F}_X$ and $\hat{G}$ respectively, and set $a_n = \frac{1}{n^{3/4}}$. Trivially

$$\begin{aligned}
& n \int_{-\infty}^{Q_{2n}(a_n)} \int_{-\infty}^{Q_{1n}(a_n)} \frac{(F_n(x,y) - \hat{F}_X(x)\hat{G}(y))^2}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y) \\
\leq \; & 2n \int_{-\infty}^{Q_{2n}(a_n)} \int_{-\infty}^{Q_{1n}(a_n)} \frac{F_n^2(x,y) + \hat{F}_X^2(x)\hat{G}^2(y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y) \\
\leq \; & 4n\hat{F}_X(Q_{1n}(a_n))\hat{G}(Q_{2n}(a_n)) \overset{a.s.}{\leq} 4n\left(a_n + \frac{1}{n}\right)^2 \to 0 \quad (n \to \infty).
\end{aligned}$$

Now assume $x \geq Q_{1n}(a_n)$, $y \leq Q_{2n}(a_n)$ and $(x,y) \in B_\varepsilon$. We have

$$\frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} \leq 2n\left(\frac{\hat{G}(y)}{\hat{F}_X(x)} + \hat{F}_X(x)\hat{G}(y)\right).$$

Hence

$$\begin{aligned}
& \int_{Q_{1n}(a_n)}^{m_1} \int_{-\infty}^{Q_{2n}(a_n)} \frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{G}(y) d\hat{F}_X(x) \\
\leq \; & 2n \int_{Q_{1n}(a_n)}^{m_1} \left(\frac{1}{\hat{F}_X(x)} + \hat{F}_X(x)\right) d\hat{F}_X(x) \int_{-\infty}^{Q_{2n}(a_n)} \hat{G}(y) d\hat{G}(y) \\
\leq \; & n\left(1 + \log\frac{1}{\hat{F}_X(Q_{1n}(a_n))}\right)\hat{G}^2(Q_{2n}(a_n)) \\
\leq \; & n\left(1 + \log\frac{1}{a_n}\right)\left(a_n + \frac{1}{n}\right)^2 \to 0 \ \text{ a.s.}
\end{aligned}$$

Similarly we can deal with the region where $x \leq Q_{1n}(a_n)$, $y \geq Q_{2n}(a_n)$ and $(x,y) \in B_\varepsilon$. Set $T_\varepsilon = \{(x,y) \in B_\varepsilon : x \leq Q_{1n}(a_n) \text{ or } y \leq Q_{2n}(a_n)\}$. We have now shown that

$$P\left(\iint_{T_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y) \geq \frac{1}{2}\varepsilon\right) \leq \frac{1}{2}\varepsilon.$$

So it remains to consider that part of the integral in (0.6) where $Q_{1n}(a_n) \leq x \leq q_{1\varepsilon}$ and $Q_{2n}(a_n) \leq y \leq m_2$, or $Q_{2n}(a_n) \leq y \leq q_{2\varepsilon}$ and $Q_{1n}(a_n) \leq x \leq m_1$. This part of the proof is the crucial part. The proof for both regions, however, is rather similar, so we confine ourselves to studying the integral over the latter region, i.e. we will show

$$P\left(\int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}}\int\limits_{Q_{1n}(a_n)}^{m_1}\frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)}d\hat{F}_X(x)d\hat{G}(y)\geq\frac{1}{2}\varepsilon\right)\leq\frac{1}{2}\varepsilon. \tag{0.7}$$

Because of (0.1) and the fact that $(a+b+c+d)^2\leq 4(a^2+b^2+c^2+d^2)$, it suffices to prove (0.7) with $V_{n,0}$ replaced by $\widetilde{\alpha}_n\left(x,y\pm\frac{\log^2 n}{n}\right)$, $G(y)\alpha_n(x,\infty)$, $\hat{F}_X(x)\widetilde{\alpha}_n\left(\infty,y\pm\frac{\log^2 n}{\sqrt{n}}\right)$, $2C\frac{\log^2 n}{\sqrt{n}}$, respectively. Denote the resulting statements as (0.7a$\pm$), (0.7b), (0.7c$\pm$), (0.7d). For (0.7d) we have

$$\begin{aligned}
&\int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}}\int\limits_{Q_{1n}(a_n)}^{m_1}\frac{4C^2\log^4 n}{n\hat{F}_X(x)\hat{G}(y)}d\hat{F}_X(x)d\hat{G}(y)\\
\leq\;&\frac{4C^2\log^4 n}{n}\log\frac{1}{\hat{F}_X(Q_{1n}(a_n))}\log\frac{1}{\hat{G}(Q_{2n}(a_n))}\overset{a.s.}{\leq}\frac{4C^2\log^6 n}{n}\to 0.
\end{aligned}$$

For (0.7b) we see, writing $\alpha_{1n}(x)=\alpha_n(x,\infty)$,

$$\begin{aligned}
&\int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}}\int\limits_{Q_{1n}(a_n)}^{m_1}\frac{(G(y)\alpha_{1n}(x))^2}{\hat{F}_X(x)\hat{G}(y)}d\hat{F}_X(x)d\hat{G}(y)\\
=\;&\int\limits_{Q_{1n}(a_n)}^{m_1}\frac{\alpha_{1n}^2(x)}{\hat{F}_X(x)}d\hat{F}_X(x)\int\limits_{Q_{2n}(a_n)}^{q_{1\varepsilon}}\frac{G^2(y)}{\hat{G}(y)}d\hat{G}(y)\\
\leq\;&\left(\sup_{Q_{1n}(a_n)\leq x\leq m_1}\frac{|\alpha_{1n}(x)|}{\hat{F}_X^{\frac{1}{4}}(x)}\right)^2\int\limits_{Q_{1n}(a_n)}^{m_1}\frac{1}{\hat{F}_X^{\frac{1}{2}}(x)}d\hat{F}_X(x)\int\limits_{Q_{2n}(a_n)}^{q_{1\varepsilon}}\frac{G^2(y)}{\hat{G}(y)}d\hat{G}(y)\\
=\;&O_P(1)\int\limits_{Q_{2n}(a_n)}^{q_{1\varepsilon}}\hat{G}(y)d\hat{G}(y)=O_P(1)\cdot\delta^2(\varepsilon).
\end{aligned}$$

Recall that with arbitrarily high probability for large n

$$\hat{G}_1\left(y-\frac{\log^2 n}{n}\right)\leq\hat{G}(y)\leq\hat{G}_1\left(y+\frac{\log^2 n}{n}\right).$$

In the sequel it will also be used (for n a multiple of three) that $\hat{G}_1$ can be written as the average of three dependent empirical df's, each based on $\frac{n}{3}$ i.i.d. random variables with df G. Hence we can use the well-known properties of classical empirical df's and processes, in particular weak convergence of weighted empirical processes and results on the ratio of the empirical and the true df. For (0.7c+) we obtain ((0.7c−) can be dealt with similarly)

$$\int_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \int_{Q_{1n}(a_n)}^{m_1} \frac{\hat{F}_X^2(x)\widetilde{\alpha}_{2n}^2\left(y+\frac{\log^2 n}{n}\right)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x)d\hat{G}(y)$$

$$= \int_{Q_{1n}(a_n)}^{m_1} \hat{F}_X(x)d\hat{F}_X(x) \int_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \frac{\widetilde{\alpha}_{2n}^2\left(y+\frac{\log^2 n}{n}\right)}{\hat{G}(y)} d\hat{G}(y)$$

$$\leq \int_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \frac{\widetilde{\alpha}_{2n}^2\left(y+\frac{\log^2 n}{n}\right)}{G^{\frac{1}{2}}\left(y+\frac{\log^2 n}{n}\right)} \frac{G^{\frac{1}{2}}\left(y+\frac{\log^2 n}{n}\right)}{\hat{G}^{\frac{1}{2}}(y)} \frac{1}{\hat{G}^{\frac{1}{2}}(y)} d\hat{G}(y)$$

$$= O_P(1)\left(\sup_{Q_{2n}(a_n)\leq y\leq q_{2\varepsilon}} \frac{\left|\widetilde{\alpha}_{2n}\left(y+\frac{\log^2 n}{n}\right)\right|}{G^{\frac{1}{4}}\left(y+\frac{\log^2 n}{n}\right)}\right)^2 \int_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \frac{1}{\hat{G}^{\frac{1}{2}}(y)} d\hat{G}(y)$$

$$= O_P(1)\hat{G}^{\frac{1}{2}}(q_{2\varepsilon}) = O_P(1)\delta^{\frac{1}{2}}(\varepsilon).$$

Remains (0.7a+) ((0.7a−) can be dealt with similarly). In accordance with the above we let

$$\widetilde{F}_{n(1)}(x,y) = \frac{3}{n}\sum_{j=1}^{\frac{n}{3}} I(X_{3j-2:n}\leq x, V_{3j-2}\leq y)$$

and

$$\widetilde{\alpha}_{n(1)}(x,y) = \sqrt{\frac{n}{3}}\left(\widetilde{F}_{n(1)}(x,y) - F_X(x)G(y)\right);$$

$\widetilde{F}_{n(2)}, \widetilde{\alpha}_{n(2)}, \widetilde{F}_{n(3)}$ and $\widetilde{\alpha}_{n(3)}$ are defined similarly. Clearly for (0.7a+) , it is sufficient (use $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$) to consider $\widetilde{\alpha}_{n(1)}\left(x, y+\frac{\log^2 n}{n}\right)$ instead of $\widetilde{\alpha}_n\left(x, y+\frac{\log^2 n}{n}\right)$; $\widetilde{\alpha}_{n(2)}, \widetilde{\alpha}_{n(3)}$ can be dealt with similarly. So we need to show

$$P\left(\int_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \int_{Q_{1n}(a_n)}^{m_1} \frac{\widetilde{\alpha}_{n(1)}^2\left(x, y+\frac{\log^2 n}{n}\right)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x)d\hat{G}(y) \geq \frac{1}{20}\varepsilon\right) \leq \frac{1}{20}\varepsilon. \quad (0.8)$$

Set

$$\hat{F}_{X(1)}(x) = \widetilde{F}_{n(1)}(x,\infty) = \frac{3}{n}\sum_{j=1}^{\frac{n}{3}} I(X_{3j-2:n} \le x)$$

and

$$\hat{G}_{u(1)}(y) = \frac{1}{\left\lfloor \frac{n}{3}u \right\rfloor} \sum_{j=1}^{\lfloor \frac{n}{3}u \rfloor} I(V_{3j-2} \le y) \quad (0/0 = 0)\,.$$

Observe that

$$\widetilde{F}_{n(1)}(x,y) = \hat{F}_{X(1)}(x)\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y).$$

So, with $y_n = y + \frac{\log^2 n}{n}$,

$$\begin{aligned}
\widetilde{\alpha}_{n(1)}(x,y_n) &= \sqrt{\frac{n}{3}}(\hat{F}_{X(1)}(x)\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y_n) - F_X(x)G(y_n)) \\
&= \sqrt{\frac{n}{3}}\hat{F}_{X(1)}(x)(\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y_n) - G(y_n)) \\
&\quad + \sqrt{\frac{n}{3}}G(y_n)(\hat{F}_{X(1)}(x) - F_X(x)).
\end{aligned} \tag{0.9}$$

Since $\hat{F}_X(x) \le \hat{F}_{X(1)}(x) \le \hat{F}_X(x) + \frac{2}{n}$, for all x, we see similar as for (0.7b), that the latter term contributes $O_P(1)\delta^2(\varepsilon)$, to the integral in (0.8), where we used $(a+b)^2 \le 2(a^2+b^2)$. So it remains to consider the contribution of the first term. $\left(\left\lfloor \frac{n}{3}u \right\rfloor / \sqrt{\frac{n}{3}}\right)(\hat{G}_{u(1)}(y) - G(y))$ is the sequential empirical process based on $\frac{n}{3}$ i.i.d random variables with df G. Using the Komlós, Major and Tusnády (1975) Kiefer process approximation of the sequential empirical process, we have that there exists a Kiefer process K such that

$$\sup_{\substack{0\le u\le 1\\ y\in \mathbb{R}}} \left| \left(\left\lfloor \frac{n}{3}u \right\rfloor / \sqrt{\frac{n}{3}}\right)(\hat{G}_{u(1)}(y) - y) - \frac{K\left(G(y), \left\lfloor \frac{n}{3}u \right\rfloor\right)}{\sqrt{\frac{n}{3}}} \right| \overset{a.s.}{=} O\left(\frac{\log^2 n}{\sqrt{n}}\right).$$

Hence

$$\begin{aligned}
&\sup_{\substack{x\in D_X\\ y\in \mathbb{R}}} \left| \sqrt{\frac{n}{3}}\ \hat{F}_{X(1)}(x)(\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y) - G(y)) - \frac{K(G(y), \frac{n}{3}\hat{F}_{X(1)}(x))}{\sqrt{\frac{n}{3}}} \right| \\
&\overset{a.s.}{=} O\left(\frac{\log^2 n}{\sqrt{n}}\right).
\end{aligned} \tag{0.10}$$

Since $|\hat{F}_{X(1)}(x) - \hat{F}_X(x)| \le \frac{2}{n}$, for all x, we have by the law of the iterated logarithm for the ordinary empirical process that

$$\sup_{x\in D_X} |\hat{F}_{X(1)}(x) - \hat{F}_X(x)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{a.s.}$$

Using this, the relation between a Kiefer process and a bivariate Wiener process and Lemma 1.11.1 in Csőrgő and Révész (1981) on the oscillations of a bivariate Wiener process it follows immediately that

$$\sqrt{\frac{3}{n}} \sup_{\substack{x \in D_X \\ y \in I\!R}} \left| K(G(y), \frac{n}{3}\hat{F}_{X(1)}(x)) - K(G(y), \frac{n}{3}F_X(x)) \right| = O\left(\frac{1}{n^{1/5}}\right) \quad \text{a.s.} \tag{0.11}$$

Since $(K(x, \frac{n}{3}y))_{x,y} \stackrel{d}{=} (\sqrt{\frac{n}{3}}K(x,y))_{x,y}$, we obtain from F and G that for a Kiefer process $\widetilde{K}$

$$\sup_{\substack{x \in D_X \\ y \in I\!R}} \left| \sqrt{\frac{n}{3}}\hat{F}_{X(1)}(x)(\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y) - G(y)) - \widetilde{K}(G(y), F_X(x)) \right| \stackrel{a.s.}{=} O\left(\frac{1}{n^{1/5}}\right). \tag{0.12}$$

Now we turn back to (0.8), and in particular to the first term on the right in (0.9). From (0.12) we see that this term can be replaced by $\widetilde{K}(G(y_n), F_X(x))$; the remaining $O\left(\frac{1}{n^{1/5}}\right)$ can be treated as (0.7d) before. So it remains to consider

$$\begin{aligned} &\int_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \int_{Q_{1n}(a_n)}^{m_1} \frac{\widetilde{K}^2(G(y_n), F_X(x))}{F_X^{\frac{1}{2}}(x) G^{\frac{1}{2}}(y_n)} \frac{1}{\hat{F}_X^{\frac{1}{2}}(x)\hat{G}^{\frac{1}{2}}(y)} d\hat{F}_X(x) d\hat{G}(y) \\ \leq \quad & \left(\sup_{\substack{x \geq Q_{1n}(a_n) \\ y \geq Q_{2n}(a_n)}} \frac{\widetilde{K}(G(y_n), F_X(x))}{(F_X(x)G(y_n))^{\frac{1}{4}}} \right)^2 \hat{F}_X^{\frac{1}{2}}(m_1)\hat{G}^{\frac{1}{2}}(q_{2\varepsilon}). \end{aligned} \tag{0.13}$$

It follows from Corollary 1.12.2 in Csőrgő and Révész (1981) that the first term on the right of (0.13) is $O_P(1)$, so the right hand side of (0.13) is $O_P(1) \cdot \delta^{\frac{1}{2}}(\varepsilon)$.

Collecting everything we proved for (0.8) and (0.7), we now see that for a proper, small enough choice of $\delta(\varepsilon)$, we indeed proved (0.8) and (0.7). □

References

Csörgő, M. and Révész, P. (1981). *Strong approximations in probability and statistics.* Academic Press, New York.

Komlós, J. Major, P. and Tusnády G. (1975). An approximation of partial sums of independent rv's and the sample df, I. *Z. Warsch. Verw. Gebiete*, **32**, 111–131.