Universität Stuttgart

Fachbereich Mathematik

Strongly consistent nonparametric tests of conditional independence

László Györfi, Harro Walk

Preprint 2011/023

Universität Stuttgart

Fachbereich Mathematik

Strongly consistent nonparametric tests of conditional independence

László Györfi, Harro Walk

Preprint 2011/023

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik.uni-stuttgart.de/preprints

ISSN **1613-8309**

Abstract
A simple and explicit procedure for testing the conditional independence of two multi-dimensional random variables given a third random vector is described. The associated L_1-based test statistic is defined when the empirical distribution of the variables is restricted to finite partitions. Distribution-free strong consistency is proved.

Keywords: Conditional independence, nonparametric test, partition,
distribution-free strong consistency.

1 Introduction

Consider an $\mathbb{R}^d \times \mathbb{R}^{d'} \times \mathbb{R}^{d''}$-valued random vector (X, Y, Z). We are interested in testing the null hypothesis that X and Y are conditionally independent given Z:

$$\mathcal{H}_0 : \mathbb{P}\{X \in A, Y \in B \mid Z\} = \mathbb{P}\{X \in A \mid Z\}\mathbb{P}\{Y \in B \mid Z\} \text{ a.s.}$$

for arbitrary Borel sets A, B.

There are many independence testing approaches in the statistics literature. For $d = d' = 1$, an early nonparametric test for independence, due to Hoeffding (1948), Blum et al. (1961), De Wet (1980), Cotterill and Csörgő (1985), is based on the notion of differences between the joint distribution function and the product of the marginals. The associated independence test is consistent under appropriate assumptions. Rosenblatt (1975) defined the statistic as the L_2 distance between the joint density estimate and the product of marginal density estimates. A second approach is to base the test on characteristic functions. One characteristic function-based independence test uses as its statistic the difference between the joint empirical characteristic function and the product of the marginal empirical characteristic functions at a particular point, chosen according to a variance maximization argument in Csörgő (1985). An alternative approach is to take a smoothed difference between the joint characteristic function and the product of the marginals, as proposed by Feuerverger (1993). Gretton and Györfi (2010) introduced nonparametric tests of independence of random vectors and proved their strong consistencies. As further literature we mention Bakirov et al. (2006), Beran et al. (2007), Dauxois and Nkiet (1998), Fukumizu et al. (2007), Gieser and Randles (1997), Gretton et al. (2005), Kankainen (1995), Genest et al. (2007), Dette and Neumeyer (2000) with references.

For nonparametric testing conditional independence, in the context of index functions Linton and Gozalo (1996) proposed and investigated tests of Kolmogorov-Smirnov and Cramér-von Mises type with a generalization of empirical distribution functions using rectangles, while Song (2009) used Rosenblatt transforms. Under density assumption Su and White (2008a, 2008b), used Hellinger metric and characteristic functions, respectively, relaxing the independence assumption on the sample. An information theory based test for conditional independence is due to Kullback (1968) and was applied to learning Bayesian networks, e.g., by de Campos (2006).

In this note we extend the test of Gretton and Györfi (2010) with L_1-based statistic to testing conditional independence and show distribution-free strong consistency.

2 Main result

A sample of $\mathbb{R}^d \times \mathbb{R}^{d'} \times \mathbb{R}^{d''}$-valued independent and identically distributed (i.i.d.) random vectors $(X_1, Y_1, Z_1), \dots, (X_n, Y_n, Z_n)$ is given such that the common distribution is the same as that of (X, Y, Z)

Introduce the following empirical distributions:

$$\nu_n(A, B, C) = \frac{\#\{i : (X_i, Y_i, Z_i) \in A \times B \times C, i = 1, \dots, n\}}{n},$$

$$\nu_{n,1}(A, C) = \frac{\#\{i : (X_i, Z_i) \in A \times C, i = 1, \dots, n\}}{n},$$

$$\nu_{n,2}(B,C) = \frac{\#\{i : (Y_i, Z_i) \in B \times C, i = 1, \dots, n\}}{n},$$

and

$$\mu_{n,3}(C) = \frac{\#\{i : Z_i \in C, i = 1, \dots, n\}}{n},$$

for any Borel subsets A, B and C.

Given the finite partitions $\mathcal{P}_n = \{A_{n,1}, \dots, A_{n,m_n}\}$ of $\mathbb{R}^d$ and $\mathcal{Q}_n = \{B_{n,1}, \dots, B_{n,m'_n}\}$ of $\mathbb{R}^{d'}$ and $\mathcal{R}_n = \{C_{n,1}, \dots, C_{n,m''_n}\}$ of $\mathbb{R}^{d''}$, we define the L_1 test statistic as

$$\begin{aligned} L_n &= \sum_{A \in \mathcal{P}_n, B \in \mathcal{Q}_n, C \in \mathcal{R}_n} \left| \frac{\nu_n(A,B,C)}{\mu_{n,3}(C)} - \frac{\nu_{n,1}(A,C)}{\mu_{n,3}(C)} \frac{\nu_{n,2}(B,C)}{\mu_{n,3}(C)} \right| \mu_{n,3}(C) \\ &= \sum_{A \in \mathcal{P}_n, B \in \mathcal{Q}_n, C \in \mathcal{R}_n} \left| \nu_n(A,B,C) - \frac{\nu_{n,1}(A,C)\nu_{n,2}(B,C)}{\mu_{n,3}(C)} \right|. \end{aligned}$$

For testing a simple hypothesis versus a composite alternative, Györfi and van der Meulen (1990) introduced a related goodness of fit test statistic L_n defined as

$$L_n(\mu_{n,1}, \mu_1) = \sum_{A \in \mathcal{P}_n} |\mu_{n,1}(A) - \mu_1(A)|.$$

Beirlant et al. (2001), and Biau and Györfi (2005) proved that, for all $0 < \varepsilon$,

$$\mathbb{P}\{L_n(\mu_{n,1}, \mu_1) > \varepsilon\} \le 2^{m_n} e^{-n\varepsilon^2/2}. \tag{1}$$

Theorem 1 *Under $\mathcal{H}_0$, for all $0 < \varepsilon_1$, $0 < \varepsilon_2$ and $0 < \varepsilon_3$,*

$$\begin{aligned} &\mathbb{P}\{L_n > \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4\} \\ \le\ & 2^{m_n \cdot m'_n \cdot m''_n} e^{-n\varepsilon_1^2/2} + 2^{m'_n \cdot m''_n} e^{-n\varepsilon_2^2/2} + 2^{m_n \cdot m''_n} e^{-n\varepsilon_3^2/2} + 2^{m''_n} e^{-n\varepsilon_4^2/2}. \end{aligned}$$

Proof. Introduce the notations

$$\nu(A,B,C) = \mathbb{P}\{X \in A, Y \in B, Z \in C\},$$

$$\nu_1(A,C) = \mathbb{P}\{X \in A, Z \in C\},$$
$$\nu_2(B,C) = \mathbb{P}\{Y \in B, Z \in C\}$$

and

$$\mu_3(C) = \mathbb{P}\{Z \in C\}.$$

We bound L_n according to

$$\begin{aligned} L_n \le & \sum_{A \in \mathcal{P}_n, B \in \mathcal{Q}_n, C \in \mathcal{R}_n} |\nu_n(A,B,C) - \nu(A,B,C)| \\ &+ \sum_{A \in \mathcal{P}_n, B \in \mathcal{Q}_n, C \in \mathcal{R}_n} \left| \nu(A,B,C) - \frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)} \right| \\ &+ \sum_{A \in \mathcal{P}_n, B \in \mathcal{Q}_n, C \in \mathcal{R}_n} \left| \frac{\nu_1(A,C)}{\mu_3(C)} \nu_2(B,C) - \frac{\nu_1(A,C)}{\mu_3(C)} \nu_{n,2}(B,C) \right| \\ &+ \sum_{A \in \mathcal{P}_n, B \in \mathcal{Q}_n, C \in \mathcal{R}_n} \left| \frac{\nu_1(A,C)}{\mu_3(C)} \nu_{n,2}(B,C) - \frac{\nu_{n,1}(A,C)}{\mu_{n,3}(C)} \nu_{n,2}(B,C) \right|. \end{aligned}$$

Under the null hypothesis $\mathcal{H}_0$, the second term of the right hand side is 0, while the third term is equal to

$$\begin{aligned} & \sum_{B\in\mathcal{Q}_n, C\in\mathcal{R}_n} \left(\sum_{A\in\mathcal{P}_n} \frac{\nu_1(A,C)}{\mu_3(C)}\right) |\nu_2(B,C)-\nu_{n,2}(B,C)| \\ = & \sum_{B\in\mathcal{Q}_n, C\in\mathcal{R}_n} |\nu_2(B,C)-\nu_{n,2}(B,C)|. \end{aligned}$$

Concerning the fourth term, we have that

$$\begin{aligned} & \sum_{A\in\mathcal{P}_n, C\in\mathcal{R}_n} \left(\sum_{B\in\mathcal{Q}_n} \nu_{n,2}(B,C)\right) \left|\frac{\nu_1(A,C)}{\mu_3(C)} - \frac{\nu_{n,1}(A,C)}{\mu_{n,3}(C)}\right| \\ = & \sum_{A\in\mathcal{P}_n, C\in\mathcal{R}_n} \mu_{n,3}(C) \left|\frac{\nu_1(A,C)}{\mu_3(C)} - \frac{\nu_{n,1}(A,C)}{\mu_{n,3}(C)}\right| \\ \le & \sum_{A\in\mathcal{P}_n, C\in\mathcal{R}_n} |\nu_1(A,C)-\nu_{n,1}(A,C)| \\ & + \sum_{C\in\mathcal{R}_n} \left(\sum_{A\in\mathcal{P}_n} \nu_1(A,C)\right) \left|\frac{\mu_{n,3}(C)}{\mu_3(C)} - 1\right| \\ = & \sum_{A\in\mathcal{P}_n, C\in\mathcal{R}_n} |\nu_1(A,C)-\nu_{n,1}(A,C)| \\ & + \sum_{C\in\mathcal{R}_n} |\mu_{n,3}(C)-\mu_3(C)|. \end{aligned}$$

Under the null hypothesis $\mathcal{H}_0$, these inequalities yield the bound

$$\begin{aligned} L_n \le & \sum_{A\in\mathcal{P}_n, B\in\mathcal{Q}_n, C\in\mathcal{R}_n} |\nu_n(A,B,C)-\nu(A,B,C)| \\ & + \sum_{B\in\mathcal{Q}_n, C\in\mathcal{R}_n} |\nu_2(B,C)-\nu_{n,2}(B,C)| \\ & + \sum_{A\in\mathcal{P}_n, C\in\mathcal{R}_n} |\nu_1(A,C)-\nu_{n,1}(A,C)| \\ & + \sum_{C\in\mathcal{R}_n} |\mu_{n,3}(C)-\mu_3(C)|. \end{aligned}$$

Thus, (1) implies that

$$\begin{aligned} & \mathbb{P}\{L_n > \varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4\} \\ \le & \mathbb{P}\left\{\sum_{A\in\mathcal{P}_n, B\in\mathcal{Q}_n, C\in\mathcal{R}_n} |\nu_n(A,B,C)-\nu(A,B,C)| > \varepsilon_1\right\} \\ & +\mathbb{P}\left\{\sum_{B\in\mathcal{Q}_n, C\in\mathcal{R}_n} |\nu_2(B,C)-\nu_{n,2}(B,C)| > \varepsilon_2\right\} \\ & +\mathbb{P}\left\{\sum_{A\in\mathcal{P}_n, C\in\mathcal{R}_n} |\nu_1(A,C)-\nu_{n,1}(A,C)| > \varepsilon_3\right\} \\ & +\mathbb{P}\left\{\sum_{C\in\mathcal{R}_n} |\mu_{n,3}(C)-\mu_3(C)| > \varepsilon_3\right\} \\ \le & 2^{m_n\cdot m_n'\cdot m_n''}e^{-n\varepsilon_1^2/2} + 2^{m_n'\cdot m_n''}e^{-n\varepsilon_2^2/2} + 2^{m_n\cdot m_n''}e^{-n\varepsilon_3^2/2} + 2^{m_n''}e^{-n\varepsilon_4^2/2}. \end{aligned}$$

■

The technique of Theorem 1 yields a distribution-free strong consistent test of homogeneity, which rejects the null hypothesis if L_n becomes large. The concept of strong consistent test is quite unusual: it means that both on $\mathcal{H}_0$ and on its complement the test makes a.s. no error after a random sample size. In other words, denoting by $\mathbf{P}_0$ (*resp.* $\mathbf{P}_1$) the probability under the null hypothesis (*resp.* under the alternative), we have

$$\mathbf{P}_0\{\text{rejecting } \mathcal{H}_0 \text{ for only finitely many } n\} = 1$$

and

$$\mathbf{P}_1\{\text{accepting } \mathcal{H}_0 \text{ for only finitely many } n\} = 1.$$

In a real life problem, for example, where we get the data sequentially, we only receive the data once, and should make the best possible inference from these data. Strong consistency means that the single sequence of inferences we make from the data received is a.s. perfect if the sample size is large enough. This concept is close to the definition of discernability introduced by Dembo and Peres (1994). Devroye and Lugosi (2002), Biau and Györfi (2005), and Gretton and Györfi (2010) provide further discussion and references. We emphasize the fact that the test presented in Corollary 1 is entirely distribution-free, i.e., the probability distributions ν, ν_1 and ν_2 are completely arbitrary.

Corollary 1 *Consider the test which rejects $\mathcal{H}_0$ when*

$$L_n > c_1\left(\sqrt{\frac{m_n m'_n m''_n}{n}} + \sqrt{\frac{m'_n m''_n}{n}} + \sqrt{\frac{m_n m''_n}{n}} + \sqrt{\frac{m''_n}{n}}\right) \approx c_1\sqrt{\frac{m_n m'_n m''_n}{n}},$$

where

$$c_1 > \sqrt{2\ln 2} \approx 1.177.$$

Assume that the conditions

$$\lim_{n\to\infty} \frac{m_n m'_n m''_n}{n} = 0, \tag{2}$$

and

$$\lim_{n\to\infty} \frac{m''_n}{\ln n} = \infty. \tag{3}$$

are satisfied. Then, under $\mathcal{H}_0$, after a random sample size the test makes a.s. no error. Under the alternative hypothesis, assume that the sequences of partitions $\mathcal{P}_n$ and $\mathcal{Q}_n$ are nested, and that for any sphere S centered at the origin

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n,\, A\cap S\neq 0} \operatorname{diam}(A) = 0, \tag{4}$$

and

$$\lim_{n\to\infty} \max_{B\in\mathcal{Q}_n,\, B\cap S\neq 0} \operatorname{diam}(B) = 0, \tag{5}$$

and

$$\lim_{n\to\infty} \max_{C\in\mathcal{R}_n,\, C\cap S\neq 0} \operatorname{diam}(C) = 0. \tag{6}$$

Then after a random sample size the test makes a.s. no error.

Proof.

STEP 1. Under $\mathcal{H}_0$, we obtain from Theorem 1 a non-asymptotic bound for the tail of the distribution of L_n, namely

$$\begin{aligned}
&\mathbb{P}\left\{L_n > c_1\left(\sqrt{\frac{m_n m'_n m''_n}{n}} + \sqrt{\frac{m'_n m''_n}{n}} + \sqrt{\frac{m_n m''_n}{n}} + \sqrt{\frac{m''_n}{n}}\right)\right\} \\
\leq\ & 2^{m_n m'_n m''_n} e^{-c_1^2 m_n m'_n m''_n/2} + 2^{m'_n m''_n} e^{-c_1^2 m'_n m''_n/2} + 2^{m_n m''_n} e^{-c_1^2 m_n m''_n/2} + 2^{m''_n} e^{-c_1^2 m''_n/2} \\
\leq\ & e^{-(c_1^2/2-\ln 2) m_n m'_n m''_n} + e^{-(c_1^2/2-\ln 2) m'_n m''_n} + e^{-(c_1^2/2-\ln 2) m_n m''_n} + e^{-(c_1^2/2-\ln 2) m''_n} \\
\leq\ & 4e^{-(c_1^2/2-\ln 2) m''_n}.
\end{aligned}$$

Therefore the condition (3) implies

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{ L_n > c_1 \left(\sqrt{\frac{m_n m'_n m''_n}{n}} + \sqrt{\frac{m'_n m''_n}{n}} + \sqrt{\frac{m_n m''_n}{n}} + \sqrt{\frac{m''_n}{n}} \right) \right\} < \infty,$$

and the proof under the null hypothesis is completed by the Borel-Cantelli lemma.
STEP 2. For the result under the alternative hypothesis, we first apply the triangle inequality

$$L_n \quad = \sum_{A\in\mathcal{P}_n, B\in\mathcal{Q}_n, C\in\mathcal{R}_n} \left| \nu(A,B,C) - \frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)} \right| + V_n,$$

where

$$\begin{aligned} |V_n| \quad \le \quad & \sum_{A\in\mathcal{P}_n, B\in\mathcal{Q}_n, C\in\mathcal{R}_n} |\nu_n(A,B,C) - \nu(A,B,C)| \\ & + \sum_{B\in\mathcal{Q}_n, C\in\mathcal{R}_n} |\nu_2(B,C) - \nu_{n,2}(B,C)| \\ & + \sum_{A\in\mathcal{P}_n, C\in\mathcal{R}_n} |\nu_1(A,C) - \nu_{n,1}(A,C)| \\ & + \sum_{C\in\mathcal{R}_n} |\mu_{n,3}(C) - \mu_3(C)| . \end{aligned}$$

According to Step 1, we get that

$$\liminf_{n\to\infty} L_n = \liminf_{n\to\infty} \sum_{A\in\mathcal{P}_n, B\in\mathcal{Q}_n, C\in\mathcal{R}_n} \left| \nu(A,B,C) - \frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)} \right|$$

a.s., since (2) implies $V_n \to 0$ a.s. The conditional distribution of (X,Y) given Z is denoted by ν_z:

$$\nu_z(A,B) = \mathbb{P}\{X\in A, Y\in B \mid Z=z\}$$

while $\mu_{1,z}$ and $\mu_{2,z}$ stand for the conditional distributions of X and Y given Z, respectively:

$$\nu_{1,z}(A) = \mathbb{P}\{X\in A \mid Z=z\}$$

and

$$\nu_{2,z}(B) = \mathbb{P}\{Y\in B \mid Z=z\}.$$

Let L^* be the expected total variation distance of ν_z and $\nu_{1,z}\times\nu_{2,z}$:

$$L^* = \mathbb{E}\{\sup_F |\nu_Z(F) - \nu_{1,Z}\times\nu_{2,Z}(F)|\},$$

where the supremum is taken over all Borel subsets F of $\mathbb{R}^d\times\mathbb{R}^{d'}$. Exactly, under the alternative hypothesis $L^*>0$ holds. By (2) it suffices to prove that by conditions (4), (5) and (6),

$$\liminf_{n\to\infty} \sum_{A\in\mathcal{P}_n, B\in\mathcal{Q}_n, C\in\mathcal{R}_n} \left| \nu(A,B,C) - \frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)} \right| \ge 2L^* > 0,$$

and therefore

$$\liminf_{n\to\infty} L_n \ge 2L^* > 0 \tag{7}$$

a.s.

The sequences of partitions $\mathcal{P}_n$ and $\mathcal{Q}_n$ are nested, then for any $n\ge m$ we have that

$$\begin{aligned} & \sum_{A\in\mathcal{P}_n, B\in\mathcal{Q}_n, C\in\mathcal{R}_n} \left| \nu(A,B,C) - \frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)} \right| \\ \ge \quad & \sum_{A\in\mathcal{P}_m, B\in\mathcal{Q}_m, C\in\mathcal{R}_n} \left| \nu(A,B,C) - \frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)} \right|, \end{aligned}$$

therefore it suffices to show that

$$\liminf_{m\to\infty}\liminf_{n\to\infty}\sum_{A\in\mathcal{P}_m,B\in\mathcal{Q}_m,C\in\mathcal{R}_n}\left|\nu(A,B,C)-\frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)}\right|\geq 2L^*>0. \tag{8}$$

Introduce the notation

$$C_n(z)=C_{n,j}\text{ if }z\in C_{n,j}.$$

For each $A\in\mathcal{P}_m, B\in\mathcal{Q}_m$, we get

$$\begin{aligned}
&\sum_{C\in\mathcal{R}_n}\left|\nu(A,B,C)-\frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)}\right|\\
&= \int\left|\frac{\int_{C_n(z)}\nu_w(A,B)\mu_3(dw)}{\mu_3(C_n(z))}-\frac{\int_{C_n(z)}\nu_{1,w}(A)\mu_3(dw)}{\mu_3(C_n(z))}\frac{\int_{C_n(z)}\nu_{2,w}(B)\mu_3(dw)}{\mu_3(C_n(z))}\right|\mu_3(dz)\\
&= \int|\nu_z(A,B)-\nu_{1,z}(A)\nu_{2,z}(B)|\,\mu_3(dz)+W_n,
\end{aligned}$$

where

$$\begin{aligned}
|W_n| \leq\ & \int\left|\frac{\int_{C_n(z)}\nu_w(A,B)\mu_3(dw)}{\mu_3(C_n(z))}-\nu_z(A,B)\right|\mu_3(dz)\\
&+\int\left|\frac{\int_{C_n(z)}\nu_{1,w}(A)\mu_3(dw)}{\mu_3(C_n(z))}-\nu_{1,z}(A)\right|\mu_3(dz)\\
&+\int\left|\frac{\int_{C_n(z)}\nu_{2,w}(B)\mu_3(dw)}{\mu_3(C_n(z))}-\nu_{2,z}(B)\right|\mu_3(dz)
\end{aligned}$$

because of $\nu_{1,w}\leq 1$, $\nu_{2,w}\leq 1$ for all w. The functions $w\to\nu_w(A,B)$, $w\to\nu_{1,w}(A)$, $w\to\nu_{2,w}(B)$ can be approximated in $L_1(\mu_3)$ by continuous functions with compact support (cf., e.g., Theorem A.1 in Györfi et al. (2002)). This and condition (6) imply $W_n\to 0$. Because the partitions $\mathcal{P}_m$ and $\mathcal{Q}_m$ are finite, we obtain

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{A\in\mathcal{P}_m,B\in\mathcal{Q}_m,C\in\mathcal{R}_n}\left|\nu(A,B,C)-\frac{\nu_1(A,C)\nu_2(B,C)}{\mu_3(C)}\right|\\
&= \sum_{A\in\mathcal{P}_m,B\in\mathcal{Q}_m}\int|\nu_z(A,B)-\nu_{1,z}(A)\nu_{2,z}(B)|\,\mu_3(dz).
\end{aligned} \tag{9}$$

For any fixed z, from (4) and (5) we get that

$$\lim_m\sum_{A\in\mathcal{P}_m,B\in\mathcal{Q}_m}|\nu_z(A,B)-\nu_{1,z}(A)\nu_{2,z}(B)|=2\sup_F|\nu_z(F)-\nu_{1,z}\times\nu_{2,z}(F)|$$

(cf. Barron et al. (1992)). Therefore Fatou lemma implies that

$$\begin{aligned}
&\liminf_m\int\left(\sum_{A\in\mathcal{P}_m,B\in\mathcal{Q}_m}|\nu_z(A,B)-\nu_{1,z}(A)\nu_{2,z}(B)|\right)\mu_3(dz)\\
&\geq \int\liminf_m\sum_{A\in\mathcal{P}_m,B\in\mathcal{Q}_m}|\nu_z(A,B)-\nu_{1,z}(A)\nu_{2,z}(B)|\,\mu_3(dz)\\
&= \mathbb{E}\left\{2\sup_F|\nu_Z(F)-\nu_{1,Z}\times\nu_{2,Z}(F)|\right\}\\
&= 2L^*,
\end{aligned} \tag{10}$$

which, by (9) yields (8), and the proof is complete. ∎

References

References

Bakirov, N., Rizzo, M., Székely, G., 2006. A multivariate nonparametric test of independence. J. Multivariate Anal. 97, 1742-1756.

Barron, A.R., Györfi, L., van der Meulen, E. C., 1992. Distribution estimation consistent in total variation and in two types of information divergence. IEEE Trans. Information Theory 38, 1437-1454.

Beirlant, J., Devroye, L., Györfi, L., Vajda, I., 2001. Large deviations of divergence measures on partitions. J. Statist. Plann. Inference 93, 1-16.

Beran, R., Bilodeau, M., Lafaye de Micheaux, P., 2007. Nonparametric tests of independence between random vectors. J. Multivariate Anal. 98, 1805-1824.

Biau, G., Györfi, L., 2005. On the asymptotic properties of a nonparametric l_1-test statistic of homogeneity. IEEE Trans. Information Theory 51, 3965-3973.

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.

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

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

Dauxois, J., Nkiet, G.M., 1998. Nonlinear canonical analysis and independence tests. Ann. Statist. 26, 1254-1278.

de Campos, L.M., 2006. A scoring function for learning Bayesian networks based on mutual information and conditional independence tests. J. Mach. Learning Res. 7, 2149-2187.

Dembo, A., Peres, Y., 1994. A topological criterion for hypothesis testing. Ann. Statist. 22, 106-117.

Dette, H., Neumeyer, N., 2000. A note on a specification test of independence. Metrika 51, 133-144.

Devroye, L., Lugosi, G., 2002. Almost sure classification of densities. J. Nonpar. Statist. 14, 675-698.

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

Feuerverger, A., 1993. A consistent test for bivariate dependence. Internat. Statist. Rev. 61, 419-433.

Fukumizu, K., Bach, F., Gretton, A., 2007. Statistical consistency of kernel canonical correlation analysis. J. Mach. Learning Res. 8, 361-383.

Genest, C., Quessy, J.F., Remillard, B., 2007. Asymptotic local efficiency of Cramér-von Mises tests for multivariate independence. Ann. Statist. 35, 166-191.

Gieser, P., Randles, R., 1997. A nonparametric test of independence between two vectors. J. Amer. Statist. Assoc. 92, 561-567.

Gretton, A., Györfi, L., 2010. Consistent nonparametric tests of independence. J. Mach. Learning Res. 11, 1391-1423.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., Schölkopf, B., 2005. Kernel methods for measuring independence. J. Mach. Learning Res. 6, 2075- 2129.

Györfi, L., Kohler, K., Krzyżak, A., Walk, H., 2002. A Distribution-Free Theory of Nonparametric Regression. Springer, New York.

Györfi, L., van der Meulen, E.C., 1990. A consistent goodness of fit test based on the total variation distance. In: Roussas, G. (Ed.), Nonparametric Functional Estimation and Related Topics. Kluwer, Dordrecht, pp. 631-645.

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

Kankainen, A., 1995. Consistent Testing of Total Independence Based on the Empirical Characteristic Function. PhD thesis, University of Jyväskylä.

Kullback, S., 1968. Information Theory and Statistics. Dover Publications, London.

Linton, O., Gozalo, P., 1996. Conditional independence restrictions: testing and estimation. Cowles Foundation Discussion Papers 1140, 47 pp.

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

Song, K., 2009. Testing conditional independence via Rosenblatt transforms. Ann. Statist. 37, 4011-4045.

Su, L., White, H., 2008a. A nonparametric Hellinger metric test for conditional independence. Econometric Theory 24, 829-864.

Su, L., White, H., 2008b. A consistent characteristic function-based test for conditional independence. J. Econometrics 143, 123-142.

László Györfi
Department of Computer Science and Information Theory
Budapest University of Technology and Economics 1521 Stoczek u. 2, Budapest, Hungary
E-Mail: gyorfi@szit.bme.hu

Harro Walk
Department of Mathematics
Universität Stuttgart Pfaffenwaldring 57, D-70569 Stuttgart, Germany
E-Mail: walk@mathematik.uni-stuttgart.de

Erschienene Preprints ab Nummer 2007/001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

2011/025 *Felber, T.; Jones, D.; Kohler, M.; Walk, H.:* Weakly universally consistent static forecasting of stationary and ergodic time series via local averaging and least squares estimates

2011/024 *Jones, D.; Kohler, M.; Walk, H.:* Weakly universally consistent forecasting of stationary and ergodic time series

2011/023 *Györfi, L.; Walk, H.:* Strongly consistent nonparametric tests of conditional independence

2011/022 *Ferrario, P.G.; Walk, H.:* Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors

2011/021 *Eberts, M.; Steinwart, I.:* Optimal regression rates for SVMs using Gaussian kernels

2011/020 *Frank, R.L.; Geisinger, L.:* Refined Semiclassical Asymptotics for Fractional Powers of the Laplace Operator

2011/019 *Frank, R.L.; Geisinger, L.:* Two-term spectral asymptotics for the Dirichlet Laplacian on a bounded domain

2011/018 *Hänel, A.; Schulz, C.; Wirth, J.:* Embedded eigenvalues for the elastic strip with cracks

2011/017 *Wirth, J.:* Thermo-elasticity for anisotropic media in higher dimensions

2011/016 *Höllig, K.; Hörner, J.:* Programming Multigrid Methods with B-Splines

2011/015 *Ferrario, P.:* Nonparametric Local Averaging Estimation of the Local Variance Function

2011/014 *Müller, S.; Dippon, J.:* k-NN Kernel Estimate for Nonparametric Functional Regression in Time Series Analysis

2011/013 *Knarr, N.; Stroppel, M.:* Unitals over composition algebras

2011/012 *Knarr, N.; Stroppel, M.:* Baer involutions and polarities in Moufang planes of characteristic two

2011/011 *Knarr, N.; Stroppel, M.:* Polarities and planar collineations of Moufang planes

2011/010 *Jentsch, T.; Moroianu, A.; Semmelmann, U.:* Extrinsic hyperspheres in manifolds with special holonomy

2011/009 *Wirth, J.:* Asymptotic Behaviour of Solutions to Hyperbolic Partial Differential Equations

2011/008 *Stroppel, M.:* Orthogonal polar spaces and unitals

2011/007 *Nagl, M.:* Charakterisierung der Symmetrischen Gruppen durch ihre komplexe Gruppenalgebra

2011/006 *Solanes, G.; Teufel, E.:* Horo-tightness and total (absolute) curvatures in hyperbolic spaces

2011/005 *Ginoux, N.; Semmelmann, U.:* Imaginary Khlerian Killing spinors I

2011/004 *Scherer, C.W.; Köse, I.E.:* Control Synthesis using Dynamic D-Scales: Part II — Gain-Scheduled Control

2011/003 *Scherer, C.W.; Köse, I.E.:* Control Synthesis using Dynamic D-Scales: Part I — Robust Control

2011/002 *Alexandrov, B.; Semmelmann, U.:* Deformations of nearly parallel G_2-structures

2011/001 *Geisinger, L.; Weidl, T.:* Sharp spectral estimates in domains of infinite volume

2010/018 *Kimmerle, W.; Konovalov, A.:* On integral-like units of modular group rings

2010/017 *Gauduchon, P.; Moroianu, A.; Semmelmann, U.:* Almost complex structures on quaternion-Kähler manifolds and inner symmetric spaces

2010/016 *Moroianu, A.; Semmelmann,U.:* Clifford structures on Riemannian manifolds

2010/015 *Grafarend, E.W.; Kühnel, W.:* A minimal atlas for the rotation group $SO(3)$

2010/014 *Weidl, T.:* Semiclassical Spectral Bounds and Beyond

2010/013 *Stroppel, M.:* Early explicit examples of non-desarguesian plane geometries

2010/012 *Effenberger, F.:* Stacked polytopes and tight triangulations of manifolds

2010/011 *Györfi, L.; Walk, H.:* Empirical portfolio selection strategies with proportional transaction costs

2010/010 *Kohler, M.; Krzyżak, A.; Walk, H.:* Estimation of the essential supremum of a regression function

2010/009 *Geisinger, L.; Laptev, A.; Weidl, T.:* Geometrical Versions of improved Berezin-Li-Yau Inequalities

2010/008 *Poppitz, S.; Stroppel, M.:* Polarities of Schellhammer Planes

2010/007 *Grundhöfer, T.; Krinn, B.; Stroppel, M.:* Non-existence of isomorphisms between certain unitals

2010/006 *Höllig, K.; Hörner, J.; Hoffacker, A.:* Finite Element Analysis with B-Splines: Weighted and Isogeometric Methods

2010/005 *Kaltenbacher, B.; Walk, H.:* On convergence of local averaging regression function estimates for the regularization of inverse problems

2010/004 *Kühnel, W.; Solanes, G.:* Tight surfaces with boundary

2010/003 *Kohler, M; Walk, H.:* On optimal exercising of American options in discrete time for stationary and ergodic data

2010/002 *Gulde, M.; Stroppel, M.:* Stabilizers of Subspaces under Similitudes of the Klein Quadric, and Automorphisms of Heisenberg Algebras

2010/001 *Leitner, F.:* Examples of almost Einstein structures on products and in cohomogeneity one

2009/008 *Griesemer, M.; Zenk, H.:* On the atomic photoeffect in non-relativistic QED

2009/007 *Griesemer, M.; Moeller, J.S.:* Bounds on the minimal energy of translation invariant n-polaron systems

2009/006 *Demirel, S.; Harrell II, E.M.:* On semiclassical and universal inequalities for eigenvalues of quantum graphs

2009/005 *Bächle, A, Kimmerle, W.:* Torsion subgroups in integral group rings of finite groups

2009/004 *Geisinger, L.; Weidl, T.:* Universal bounds for traces of the Dirichlet Laplace operator

2009/003 *Walk, H.:* Strong laws of large numbers and nonparametric estimation

2009/002 *Leitner, F.:* The collapsing sphere product of Poincaré-Einstein spaces

2009/001 *Brehm, U.; Kühnel, W.:* Lattice triangulations of E^3 and of the 3-torus

2008/006 *Kohler, M.; Krzyżak, A.; Walk, H.:* Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps

2008/005 *Kaltenbacher, B.; Schöpfer, F.; Schuster, T.:* Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems

2008/004 *Leitner, F.:* Conformally closed Poincaré-Einstein metrics with intersecting scale singularities

2008/003 *Effenberger, F.; Kühnel, W.:* Hamiltonian submanifolds of regular polytope

2008/002 *Hertweck, M.; Höfert, C.R.; Kimmerle, W.:* Finite groups of units and their composition factors in the integral group rings of the groups $PSL(2,q)$

2008/001 *Kovarik, H.; Vugalter, S.; Weidl, T.:* Two dimensional Berezin-Li-Yau inequalities with a correction term

2007/006 *Weidl, T.:* Improved Berezin-Li-Yau inequalities with a remainder term

2007/005 *Frank, R.L.; Loss, M.; Weidl, T.:* Polya's conjecture in the presence of a constant magnetic field

2007/004 *Ekholm, T.; Frank, R.L.; Kovarik, H.:* Eigenvalue estimates for Schrödinger operators on metric trees

2007/003 *Lesky, P.H.; Racke, R.:* Elastic and electro-magnetic waves in infinite waveguides

2007/002 *Teufel, E.:* Spherical transforms and Radon transforms in Moebius geometry

2007/001 *Meister, A.:* Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions