Nonparametric tests of independence between random vectors

R. Beran [a], M. Bilodeau [b,*] and P. Lafaye de Micheaux [c]

[a] *UC Davis Department of Statistics, 360 Kerr Hall, One Shields Ave., Davis, CA 95616, USA*

[b] *Département de mathématiques et de statistique, université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7*

[c] *Université Pierre Mendès France, Laboratoire de Statistique et Analyse de Données (LabSAD), BP 47 / F-38040 Grenoble Cedex 9, France*

Abstract

A non parametric test of the mutual independence between many numerical random vectors is proposed. This test is based on a characterization of mutual independence defined from probabilities of half-spaces in a combinatorial formula of Möbius. As such, it is a natural generalization of tests of independence between univariate random variables using the empirical distribution function. If the number of vectors is p and there are n observations, the test is defined from a collection of processes $R_{n,A}$, where A is a subset of $\{1, \ldots, p\}$ of cardinality $|A| > 1$, which are asymptotically independent and Gaussian. Without the assumption that each vector is one-dimensional with a continuous cumulative distribution function, any test of independence can not be distribution free. The critical values of the proposed test are thus computed with the bootstrap which is shown to be consistent. Another similar test, with the same asymptotic properties, for the serial independence of a multivariate stationary sequence is also proposed. The proposed test works when some or all of the marginal distributions are singular with respect to Lebesgue measure. Moreover, in singular cases described in Section 4, the test inherits useful invariance properties from the general affine invariance property.

Key words: Bootstrap, Gaussian process, independence, multivariate distribution, serial independence
1991 MSC: 62H15, 62G09, 60F05

* Corresponding author
Email addresses: beran@wald.ucdavis.edu (R. Beran),
bilodeau@dms.umontreal.ca (M. Bilodeau),
Pierre.Lafaye-de-Micheaux@upmf-grenoble.fr (P. Lafaye de Micheaux).

Preprint submitted to Journal of Multivariate Analysis *19 January 2007*

1 Introduction

Nonparametric tests of independence between random vectors are scarce in the literature. Puri and Sen [17] defined a class of association parameters between two vectors based on componentwise ranking which results in a non invariant statistic. Gieser and Randles [13] proposed an invariant test of independence between two vectors based on interdirections and obtained the asymptotic distribution, with Pitman asymptotic relative efficiencies, when both vectors follow an elliptically symmetric distribution. A generalization of the interdirection quadrant test is proposed in Um and Randles [19] who considered a test of pairwise independence among many elliptically contoured vectors. Cléroux *et al.* [5] derived a nonparametric test of no association between two vectors. It is a test of the independence between each variable in one vector with any variable in the other vector. Bilodeau and Lafaye de Micheaux [3] used the empirical characteristic function to test the mutual independence among p normally distributed vectors without specifying their joint distribution.

Earlier papers still of interest on nonparametric tests of independence between random variables based on the empirical cumulative distribution function (cdf) are those of Hoeffding [14] and Blum *et al.* [4]. In Hoeffding [14], the asymptotic distribution of these processes for testing independence between two variables is quite simple. In that case, the asymptotic covariance function is a product of two covariance functions of brownian bridges. This description gets more complicated when there are more than two variables. Blum *et al.* [4] proposed a modification of the edf process which preserves the product structure of the covariance function and they gave explicit expressions for the case of three random variables. Ghoudi *et al.* [12] characterized independence between p random variables with a Möbius transformation due to Deheuvels [9]. Let F be the joint cdf of $(X^{(j)})_{j=1}^{p}$ and $F^{(j)}$ denote the marginal cdf of $X^{(j)}$. Let $\mathcal{I}_p = \{A \subset \{1,\ldots,p\} : |A| > 1\}$, where $|A|$ is the cardinality of the set A. Note that the cardinality of $\mathcal{I}_p$ is $2^p - p - 1$. For any $t = (t^{(j)})_{j=1}^{p} \in \mathbb{R}^p$ and any $A \in \mathcal{I}_p$, define

$$\mu_A(t) = \sum_{B \subset A} (-1)^{|A \setminus B|} F(t^B) \prod_{j \in A \setminus B} F^{(j)}(t^{(j)}),$$

where $\prod_{\emptyset} = 1$. The vector $t^B = ((t^B)^{(j)})_{j=1}^{p} \in \mathbb{R}^p$ is defined as

$$(t^B)^{(j)} = \begin{cases} t^{(j)}, & j \in B; \\ \infty, & j \notin B. \end{cases}$$

They characterized independence as follows: $X^{(1)}, \ldots, X^{(p)}$ are independent if and only if $\mu_A(t) = 0$, for all $t \in \mathbb{R}^p$ and all $A \in \mathcal{I}_p$. This characterization was also given previously in a slightly different form in Deheuvels [8]. This led to the processes

$$V_{n,A}(t) = \sqrt{n} \sum_{B \subset A} (-1)^{|A \setminus B|} F_n(t^B) \prod_{j \in A \setminus B} F_n^{(j)}(t^{(j)}),$$

where F_n and $F_n^{(j)}$ denote, respectively, the joint empirical cdf and the marginal empirical cdf. The weak convergence of these processes can be stated as: $\{V_{n,A} : A \in \mathcal{I}_p\}$ converges weakly to $\{V_A : A \in \mathcal{I}_p\}$, where V_A are independent zero mean Gaussian processes with covariance function

$$\operatorname{cov}(V_A(s), V_A(t)) = \prod_{k \in A} \min\{F^{(k)}(s^{(k)}), F^{(k)}(t^{(k)})\} - F^{(k)}(s^{(k)})F^{(k)}(t^{(k)}).$$

The asymptotic processes being independent, it becomes tractable to consider all sets A simultaneously via test statistics such as $\sum_A T_{n,A}$ or $\max_A T_{n,A}$, where for a given set A, the Cramér-von Mises statistic

$$T_{n,A} = \int V_{n,A}^2(t) dF_n(t)$$

is used. Moreover, identification of subsets A with large values of $T_{n,A}$ can be used as a tool for finding dependent subsets of variables. The construction of a dependogram for this identification is illustrated in Genest and Rémillard [11].

The problem of serial independence is also treated. If $Y_1, Y_2, \ldots$ is a stationary sequence of random variables, the problem of serial independence is to determine whether p consecutive observations are independent. In this serial context, Ghoudi *et al.* [12] established that the same processes used in the non serial problem possess the same asymptotic properties.

This paper treats the two problems in a multivariate setting: $X^{(j)} \in \mathbb{R}^{d_j}$ in the non serial problem and $Y_j \in \mathbb{R}^q$ in the serial problem. Section 2 introduces two processes $R_{n,A}$ and $S_{n,A}$ obtained from the Möbius transformation of a process indexed by cartesian products of half-spaces. For recent applications of half-spaces in statistics, see Beran and Millar [2]. Unlike in the univariate case where Ghoudi *et al.* [12] assume continuous marginal distribution function for $X^{(j)}$ (or Y_j) to obtain a distribution free statistic, the multivariate statistic can not be distribution free, even in the continuous case. In Section 3, validity of bootstrap technology is established to obtain critical values from the bootstrap distribution. The wide range of applicability of the proposed methodology is illustrated in Section 4. The test works when all or some of the variables are singular with respect to Lebesgue measure and it inherits useful invariance properties from the general affine invariance property. This means that, in the examples of Section 4, one can recode discrete variables or apply a rotation to data on a sphere without affecting the conclusion. The main motivation for the Kolmogorov-Smirnov approach adopted is its easiness to yield the affine invariance property and its consequences in special cases treated in Section 4. A method of Fisher to combine p-values is proposed in Section 5. Section 6 presents the basic elements of the algorithm used to evaluate the test statistic. All the proofs are deferred to the Appendix.

2 Half-spaces and independence

The general multivariate case with $X^{(j)} \in \mathbb{R}^{d_j}$ is treated. Let $|\cdot|$ and $\langle \cdot, \cdot \rangle$ denote, respectively, euclidian norm and inner product in cartesian spaces $\mathbb{R}^{d_j}$. For $j = 1, \ldots, p$, let $\mathcal{S}_{d_j} = \{x^{(j)} \in \mathbb{R}^{d_j} : |x^{(j)}| = 1\}$ be the unit sphere in $\mathbb{R}^{d_j}$. For every $(s^{(j)}, t^{(j)}) \in \mathcal{S}_{d_j} \times \mathbb{R}$, define the half-space

$$H(s^{(j)}, t^{(j)}) = \{x^{(j)} \in \mathbb{R}^{d_j} : \langle s^{(j)}, x^{(j)} \rangle \leq t^{(j)}\}.$$

The collection of half-spaces in $\mathbb{R}^{d_j}$, which separate probabilities (Cramér and Wold, 1936), is denoted

$$\mathcal{F}^{(d_j)} = \{H(s^{(j)}, t^{(j)}) : (s^{(j)}, t^{(j)}) \in \mathcal{S}_{d_j} \times \mathbb{R}\}.$$

Let P be the joint probability for $(X^{(j)})_{j=1}^p$ and $P^{(j)}$ be the marginal probability for $X^{(j)}$. The following basic characterization of independence follows from characteristic functions: $X^{(1)}, \ldots, X^{(p)}$ are independent if and only if

$$P(\times_{j=1}^p H(s^{(j)}, t^{(j)})) = \prod_{j=1}^p P^{(j)}(H(s^{(j)}, t^{(j)})),$$

for all $(H(s^{(j)}, t^{(j)}))_{j=1}^p \in \mathcal{F}^{(d_1)} \times \cdots \times \mathcal{F}^{(d_p)}$.

Let $l^\infty(\mathcal{F})$, where $\mathcal{F} = \mathcal{F}^{(d_1)} \times \cdots \times \mathcal{F}^{(d_p)}$, be the set of all bounded functions on $\mathcal{F}$ metrized with the supremum norm $||\cdot||_{\mathcal{F}}$. The σ-algebra in $l^\infty(\mathcal{F})$ is that generated by open sets, *i.e.* the Borel σ-algebra. The independence half-space process in $l^\infty(\mathcal{F})$ is defined as

$$\sqrt{n}[\mathbb{P}_n(\times_{j=1}^p H(s^{(j)}, t^{(j)})) - \prod_{j=1}^p \mathbb{P}_n^{(j)}(H(s^{(j)}, t^{(j)}))],$$

where $\mathbb{P}_n$ is the empirical probability distribution of $X_1, \ldots, X_n$ i.i.d., and where $X_i = (X_i^{(j)})_{j=1}^p$. The asymptotic distribution of this process is difficult when $p \geq 3$, even in the univariate case ($d_j = 1$), see *e.g.* Blum *et al.* [4] and Ghoudi *et al.* [12].

However, if the Möbius transformation is applied, the equivalent criterion follows: $X^{(1)}, \dots, X^{(p)}$ are independent if and only if $\nu_A((s^{(j)}, t^{(j)})_{j=1}^p) = 0$ for all $(H(s^{(j)}, t^{(j)}))_{j=1}^p \in \mathcal{F}$ and all $A \in \mathcal{I}_p$, where

$$\nu_A((s^{(j)}, t^{(j)})_{j=1}^p) = \sum_{B \subset A} (-1)^{|A \setminus B|} P(\times_{j=1}^p H^B(s^{(j)}, t^{(j)})) \cdot \prod_{j \in A \setminus B} P^{(j)}(H(s^{(j)}, t^{(j)})).$$

Here, the notation

$$H^B(s^{(j)}, t^{(j)}) = \begin{cases} H(s^{(j)}, t^{(j)}), & j \in B; \\ \mathbb{R}^{d_j}, & j \notin B, \end{cases}$$

is used.

2.1 The non serial case

For each subset A, the Möbius independence half-space process in $l^\infty(\mathcal{F})$ is defined as

$$R_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \sqrt{n} \sum_{B \subset A} (-1)^{|A \setminus B|} \mathbb{P}_n(\times_{j=1}^p H^B(s^{(j)}, t^{(j)})) \cdot \prod_{j \in A \setminus B} \mathbb{P}_n^{(j)}(H(s^{(j)}, t^{(j)})).$$

A rejection region for an independence test is constructed by combining Kolmogorov test statistics for all subsets

$$\cup_{A \in \mathcal{I}_p} \{||R_{n,A}||_{\mathcal{F}} > r_A\}, \tag{1}$$

for some critical values r_A chosen to achieve an asymptotic preassigned global significance level α. This test is invariant under the group of affine linear transformations,

$$(X_i^{(j)})_{j=1}^p \mapsto (A^{(j)} X_i^{(j)} + b^{(j)})_{j=1}^p,$$

where $A^{(j)} : d_j \times d_j$ is any non singular matrix and $b^{(j)} \in \mathbb{R}^{d_j}$ is any vector. This comes as a consequence that $||R_{n,A}||_{\mathcal{F}}$ is invariant. In the univariate setting, half-space probabilities reduce to the distribution function. The process $R_{n,A}$ is thus a natural generalization of the process of Blum *et al.* [4] or Ghoudi *et al.* [12].

The weak convergence of the Möbius independence half-space processes is now described via the closely related processes $\breve{R}_{n,A}$ defined as

$$\breve{R}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^{p}) = \sqrt{n} \sum_{B \subset A} (-1)^{|A \setminus B|} \mathbb{P}_n(\times_{j=1}^{p} H^B(s^{(j)}, t^{(j)})) \cdot \prod_{j \in A \setminus B} P^{(j)}(H(s^{(j)}, t^{(j)})).$$

which differ from $R_{n,A}$ in that the empirical marginal probability $\mathbb{P}_n^{(j)}$ is replaced by the true and unknown $P^{(j)}$. The approach to weak convergence denoted by $\rightsquigarrow$ and used to deal with measurability issues is that of Hoffmann-Jørgensen [15] as described in van der Vaart and Wellner [20].

Theorem 1 *If $X^{(1)}, \ldots, X^{(p)}$ are independent, then*

$$\{\breve{R}_{n,A} : A \in \mathcal{I}_p\} \rightsquigarrow \{R_A : A \in \mathcal{I}_p\}.$$

The processes R_A are independent zero mean Gaussian processes with covariance function given by

$$C_A((s^{(j)}, t^{(j)})_{j=1}^{p}, (\tilde{s}^{(j)}, \tilde{t}^{(j)})_{j=1}^{p}) = \prod_{k \in A} \Big[P^{(k)}(H(s^{(k)}, t^{(k)}) \cap H(\tilde{s}^{(k)}, \tilde{t}^{(k)})) - P^{(k)}(H(s^{(k)}, t^{(k)}))P^{(k)}(H(\tilde{s}^{(k)}, \tilde{t}^{(k)})) \Big].$$

The next result asserts that the two processes $R_{n,A}$ and $\breve{R}_{n,A}$ are asymptotically equivalent. This comes as a consequence of the Glivenko-Cantelli theorem for half-spaces of Wolfowitz [22]; see also Dehardt [7] or Steele [18].

Theorem 2 *For every $A \in \mathcal{I}_p$, $||R_{n,A} - \breve{R}_{n,A}||_{\mathcal{F}} \to 0$, where convergence is in outer probability.*

The asymptotic significance level of the test (1) is given by

$$\alpha = 1 - \prod_{A \in \mathcal{I}_p} P\left\{||R_A||_{\mathcal{F}} \le r_A\right\}.$$

Thus, the critical values r_A in (1) can be chosen, see Genest and Rémillard [11], as the β-quantile of the distribution of $||R_A||_{\mathcal{F}}$, where $\beta = (1-\alpha)^{1/(2^p-p-1)}$. However, in our case, the distribution of R_A is no longer distribution free; it depends on the individual distribution of the marginals $X^{(k)}$, $k \in A$. Thus, in general, a different critical value r_A is required, even for different subsets A of the same cardinality. In the next section, it is shown that the critical values r_A can be approximated by the quantiles of the bootstrap distributions.

2.2 The serial case

The problem of testing for serial independence of a multivariate stationary sequence is now addressed. The test statistic in the serial context is very similar. Consider a stationary sequence $Y_1, Y_2, \ldots$ in $\mathbb{R}^q$, where Y_j is distributed according to the probability Q. For any fixed p, let $X_i = (Y_i, \ldots, Y_{i+p-1})$ and $X_i^{(j)} = Y_{i+j-1}$, where $i = 1, \ldots, n'$, $j = 1, \ldots, p$ and $n' = n-p+1$. For a given $A \in \mathcal{I}_p$, the process is

$$S_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \sqrt{n'} \sum_{B \subset A} (-1)^{|A \setminus B|} \mathbb{P}_n(\times_{j=1}^p H^B(s^{(j)}, t^{(j)})) \cdot \prod_{j \in A \setminus B} \mathbb{Q}_n(H(s^{(j)}, t^{(j)})).$$

where $\mathbb{P}_n$ is the empirical probability distribution of $X_1, \ldots, X_{n'}$ and $\mathbb{Q}_n$ is the empirical probability distribution of $Y_1, \ldots, Y_n$. Note that $\mathbb{P}_n$ assigns weights $\frac{1}{n'}$, rather than $\frac{1}{n}$ for $\mathbb{Q}_n$, to each observation. Here, the index set of the process is the p-fold cartesian product $\mathcal{F} = (\mathcal{F}^{(q)})^p$. In the serial context, the subset A and its translate, say $A + k$, lead essentially to the same process. Hence, the test proposed for serial independence of p consecutive observations has critical region

$$\cup_{A \in \mathcal{A}_p} \{||S_{n,A}||_{\mathcal{F}} > s_A\}, \tag{2}$$

where $\mathcal{A}_p = \{A \in \mathcal{I}_p : 1 \in A\}$ has now the cardinality $2^{p-1} - 1$.

Theorem 3 *Let $Y_1, Y_2, \ldots$ be i.i.d. Q. Then, for any fixed p,*

$$\{S_{n,A} : A \in \mathcal{A}_p\} \rightsquigarrow \{S_A : A \in \mathcal{A}_p\},$$

where S_A are independent zero mean Gaussian processes with covariance function given by

$$D_A((s^{(j)}, t^{(j)})_{j=1}^p, (\tilde{s}^{(j)}, \tilde{t}^{(j)})_{j=1}^p) = \prod_{k \in A} \Big[Q(H(s^{(k)}, t^{(k)}) \cap H(\tilde{s}^{(k)}, \tilde{t}^{(k)}))$$

$$-Q(H(s^{(k)},t^{(k)}))Q(H(\tilde{s}^{(k)},\tilde{t}^{(k)}))\Big].$$

The critical points s_A in the test (2) are the β-quantiles, $\beta = (1-\alpha)^{1/(2^{p-1}-1)}$, of the distributions of $||S_A||_{\mathcal{F}}$. These quantiles will also be approximated using the bootstrap distributions. However, unlike the non serial case, since the sequence is assumed to be stationary a common critical value s_A can be chosen for all subsets A of the same cardinality.

3 Bootstrap of $R_{n,A}$ and $S_{n,A}$

In the univariate setting, transformation of marginals to uniform variables shows that Kolmogorov-Smirnov or Cramér-von Mises statistics are distribution free. However, in the multivariate context, critical values for the tests are obtained by bootstrap methodology since the asymptotic distribution is no longer distribution free.

3.1 Bootstrap of $R_{n,A}$

Under the hypothesis of independence, the unknown parameters in the non serial case are the marginal probabilities $(P^{(j)})_{j=1}^p$. The nonparametric bootstrap distribution is constructed by sampling independently (the null distribution) from the empirical marginal distribution for each subvector. Recall that $\mathbb{P}_n^{(j)}$ is the empirical probability of the subvectors $X_1^{(j)},\dots,X_n^{(j)}$. A bootstrap sample is thus a sample $X_1^*,\dots,X_n^*$ i.i.d. $\mathbb{P}_n^{(1)}\times\cdots\times\mathbb{P}_n^{(p)}$. Template A in Beran [1], rephrased for tests, outlines a way to verify that the bootstrap distributions converge correctly and that the asymptotic rejection probability is as intended.

To this end, the semimetric d_R between two finite collections of probability distributions is defined through the quarter-space semimetric

$$\begin{aligned}&d_R\left((P^{(j)})_{j=1}^p,(Q^{(j)})_{j=1}^p\right)\\&=\sum_{j=1}^p\sup_{H_1,H_2\in\mathcal{F}^{(d_j)}}|P^{(j)}(H_1\cap H_2)-Q^{(j)}(H_1\cap H_2)|.\end{aligned}$$

The empirical marginal probabilities converge in this semimetric,

$$d_R\left((\mathbb{P}_n^{(j)})_{j=1}^p,(P^{(j)})_{j=1}^p\right)\to 0$$

in probability. This holds since $\mathcal{F}^{(d_j)}$, and thus $\mathcal{F}^{(d_j)} \cap \mathcal{F}^{(d_j)}$, are Vapnik-Červonenkis (VC) classes (van der Vaart and Wellner 1996, p. 147). Thus, they are also Glivenko-Cantelli classes.

The triangular array convergence for the Möbius independence processes is now established. Let $(P_n^{(j)})_{j=1}^p$, $n = 1, 2 \ldots$ be any sequence. Assume that the sample $X_{n1}, \ldots, X_{nn}$ is i.i.d. from the product probability $P_n^{(1)} \times \cdots \times P_n^{(p)}$ and let $\hat{\mathbb{P}}_n$ be the empirical probability of $X_{n1}, \ldots, X_{nn}$. Define the process

$$R^*_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \sqrt{n} \sum_{B \subset A} (-1)^{|A \setminus B|} \hat{\mathbb{P}}_n(\times_{j=1}^p H^B(s^{(j)}, t^{(j)})) \cdot \prod_{j \in A \setminus B} \hat{\mathbb{P}}_n^{(j)}(H(s^{(j)}, t^{(j)})).$$

Theorem 4 *If* $(P_n^{(j)})_{j=1}^p$, $n = 1, 2 \ldots$ *is any sequence such that*

$$d_R\left((P_n^{(j)})_{j=1}^p, (P^{(j)})_{j=1}^p\right) \to 0, \tag{3}$$

then $\{R^*_{n,A} : A \in \mathcal{I}_p\} \rightsquigarrow \{R_A : A \in \mathcal{I}_p\}$, *where the limiting distribution is as in Theorem 1.*

The last condition in Template A is the continuity of the limiting cdf. In our setting, it becomes the continuity of the cdf of $||R_A||_{\mathcal{F}}$. This follows from Proposition 2 in Beran and Millar [2].

3.2 Bootstrap of $S_{n,A}$

The asymptotic bootstrap distribution in the serial case is treated similarly. Let $\mathbb{Q}_n$ be the empirical distribution of $Y_1, \ldots, Y_n$. A bootstrap sequence following the null hypothesis is $Y_1^*, \ldots, Y_n^*$ i.i.d. $\mathbb{Q}_n$.

A similar semimetric between any two probabilities Q_1 and Q_2 on $\mathbb{R}^q$ is defined as

$$d_S(Q_1, Q_2) = \sup_{H_1, H_2 \in \mathcal{F}^{(q)}} |Q_1(H_1 \cap H_2) - Q_2(H_1 \cap H_2)|.$$

The Glivenko-Cantelli theorem for ergodic stationary sequences in Steele [18] gives the convergence $d_S(\mathbb{Q}_n, Q) \to 0$ in probability.

The triangular array convergence also holds here. Let Q_n, $n = 1, 2, \ldots$ be any sequence of probability distributions on $\mathbb{R}^q$. Assume that the sequence

$Y_{n1}, \ldots, Y_{nn}$ is i.i.d. Q_n. Construct as before the overlapping (m-dependent) sequence $X_{n1}, \ldots, X_{nn'}$, where $n' = n-p+1$. Let $\hat{\mathbb{Q}}_n$ be the the empirical probability of $Y_{n1}, \ldots, Y_{nn}$ and let $\hat{\mathbb{P}}_n$ be the empirical probability of $X_{n1}, \ldots, X_{nn'}$. The next theorem asserts that the triangular array process

$$S^*_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \sqrt{n'} \sum_{B \subset A} (-1)^{|A \setminus B|} \hat{\mathbb{P}}_n(\times_{j=1}^p H^B(s^{(j)}, t^{(j)})) \cdot \prod_{j \in A \setminus B} \hat{\mathbb{Q}}_n(H(s^{(j)}, t^{(j)}))$$

converges nicely.

Theorem 5 *If Q_n, $n = 1, 2 \ldots$ is any sequence such that $d_S(Q_n, Q) \to 0$, then $\{S^*_{n,A} : A \in \mathcal{A}_p\} \rightsquigarrow \{S_A : A \in \mathcal{A}_p\}$, where the limiting distribution is as in Theorem 3.*

The variable $||S_A||_{\mathcal{F}}$ has a continuous cdf just like the variable $||R_A||_{\mathcal{F}}$ before. Thus, the conditions of Template A are fulfilled.

The critical values s_A in (2) corresponding to subsets of the same cardinality are identical. In this case, for all the subsets A of a given cardinality, say $|A| = k$, it is suggested to amalgamate the $\binom{p}{k} \cdot B$ bootstrap values $||S^*_{n,A}||_{\mathcal{F}}$, where B is the number of bootstrap sequences. The common critical value s_A is estimated by the β-quantile, $\beta = (1-\alpha)^{1/(2^{p-1}-1)}$, of these amalgamated bootstrap values.

4 Examples and applications

A dependogram, a word coined by Genest and Rémillard [11], is a graphical display in which, for each subset A, a vertical bar is drawn of height corresponding to the values of $||R_{n,A}||_{\mathcal{F}}$. A star at the height given by the bootstrap approximation to the β-quantile, $\beta = (1-\alpha)^{1/(2^p-p-1)}$, of $||R_A||_{\mathcal{F}}$ is added to the graph. Subsets such that the vertical bar exceeds this quantile can be flagged for dependent variables. The lexicographic order of the subsets in the non serial dependogram for $p = 4$ is as in Table 1, whereas Table 2 gives the lexicographic order for the serial dependogram.

For testing serial independence of a stationary sequence, a similar dependogram can be constructed. As mentioned before, a single critical value s_A can be used for all the subsets with the same cardinality. This common critical value serves to draw an horizontal line over all the vertical bars corresponding to the subsets A of common cardinality $|A|$.

Subsets							
1	{1,2}	4	{2,3}	7	{1,2,3}	10	{2,3,4}
2	{1,3}	5	{2,4}	8	{1,2,4}	11	{1,2,3,4}
3	{1,4}	6	{3,4}	9	{1,3,4}		

Table 1
Lexicographic order of the subsets for $p = 4$ in the non serial dependogram.

Subsets							
1	{1,2}	3	{1,4}	5	{1,2,4}	7	{1,2,3,4}
2	{1,3}	4	{1,2,3}	6	{1,3,4}		

Table 2
Lexicographic order of the subsets for $p = 4$ in the serial dependogram.

All the dependograms in the examples to follow were done at the global significance level $\alpha = .05$. The critical values r_A (or s_A) were computed on the basis of $B = 2000$ bootstrap samples. Computations were done on a Pentium 4 with a CPU of 2 Ghz with a RAM of 1 Gb running under Windows XP. The elaborate programs, including the graphical interface, were written in R with C++ subroutines to compute the statistics. Computation times are reported in each example. They are reasonable for univariate problems, but they can be lengthy for small multivariate situations. Parallel programming which is suited to bootstrap methodology would be the next step to reduce computation times.

4.1 Dependence among four discrete variables

The asymptotic distribution of the Cramér-von Mises test of Deheuvels [9], Ghoudi *et al.* [12] or Genest and Rémillard [11] does not apply when some discrete components can not by all means be treated as continuous.

As an example, a sample of size $n = 100$ on four variables is generated from the following distribution. Firstly, W_1, W_3, W_4 and W_6 are independent Poisson(1) variables. Secondly, and independently from the first step, W_2 and W_5 are two independent Poisson(3) variables. The observed variables are $X^{(1)} = W_1 + W_2$, $X^{(2)} = W_2 + W_3$, $X^{(3)} = W_4 + W_5$, and $X^{(4)} = W_5 + W_6$. This yields a pair $(X^{(1)}, X^{(2)})$ independent of the pair $(X^{(3)}, X^{(4)})$ with each pair having a correlation of $\frac{3}{4}$. The dependogram evaluated in 6.6 mins is shown in Figure 1. It displays significant values for the subsets 1 and 6 which correspond, respectively, to the two subsets $A = \{1, 2\}$ and $A = \{3, 4\}$. The subset $A = \{1, 2, 3, 4\}$ yields also a significant value. The reason for this last significant value is made clear from a detailed analysis of the independence characterization. Under this specific model, all the quantities ν_A (given here without the half-space

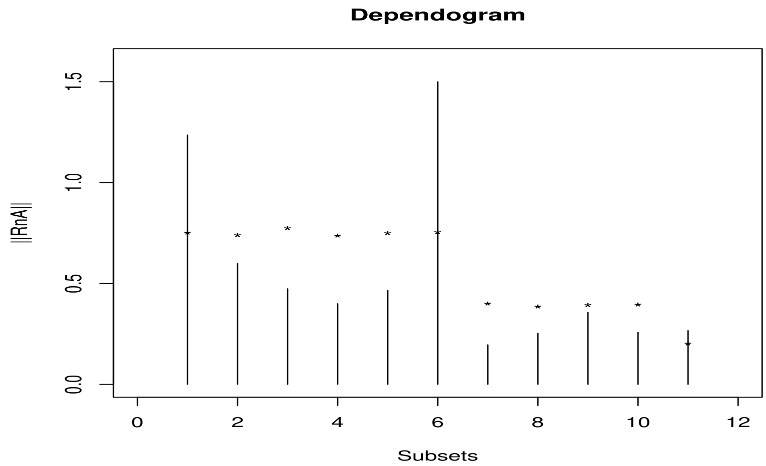

Fig. 1. The two structures of dependence are evident in subsets 1 and 6 which correspond, respectively, to the two subsets $A = \{1, 2\}$ and $A = \{3, 4\}$.

arguments) are null except $\nu_{\{1,2\}} = P^{(1,2)} - P^{(1)}P^{(2)}$, $\nu_{\{3,4\}} = P^{(3,4)} - P^{(3)}P^{(4)}$, and $\nu_{\{1,2,3,4\}} = (P^{(1,2)} - P^{(1)}P^{(2)})(P^{(3,4)} - P^{(3)}P^{(4)})$, where $P^{(i_1,\ldots,i_s)}$ is the joint probability of the variables corresponding to the indices. When the sample is large enough to draw enough power, the last subset will also come out as significant.

4.2 Dependence between three bivariate vectors

This example considers $n = 50$ observations on six variables W_i, $i = 1, \ldots, 6$, jointly distributed as a multivariate normal with mean vector 0 and covariance matrix

$$\left(\begin{array}{cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & .4 & .5 \\ 0 & 0 & 0 & 1 & .1 & .2 \\ \hline 0 & 0 & .4 & .1 & 1 & 0 \\ 0 & 0 & .5 & .2 & 0 & 1 \end{array}\right).$$

The structure of dependence among the three subvectors $X^{(1)} = (W_1, W_2)$ $X^{(2)} = (W_3, W_4)$ and $X^{(3)} = (W_5, W_6)$ is seen clearly from the third subset $A = \{2, 3\}$ in Figure 2 which required 3.8 hours of computations. The more powerful normal theory likelihood ratio test Λ could be used here. However, this test is very sensitive to non normality. In fact, it should not be used, even for heavy-tailed elliptically contoured distributions without a kurtosis

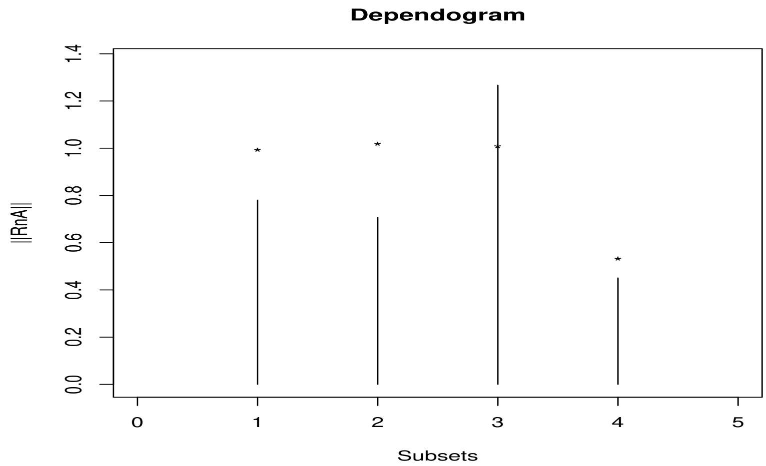

Fig. 2. The dependence between the last two subvectors shows up in the third subset $A = \{2, 3\}$.

correction; see Muirhead and Waternaux [16]. The proposed test shares with the likelihood ratio test the property of affine invariance.

4.3 4-dependent variables which are 2-independent and 3-independent

Define the uniform variable W on the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$. The vector $X = (X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)})$ is built from W through the indicator functions

$$X^{(1)} = \mathbb{I}\{W \in \{1, 2, 3, 5\}\}, \; X^{(2)} = \mathbb{I}\{W \in \{1, 2, 4, 6\}\}$$
$$X^{(3)} = \mathbb{I}\{W \in \{1, 3, 4, 7\}\}, \; X^{(4)} = \mathbb{I}\{W \in \{2, 3, 4, 8\}\}.$$

These four dependent binary variables are 2-independent or pairwise independent; they are also 3-independent. A sample of size $n = 100$ on these 4 variables was generated from which the dependogram in Figure 3 resulted in 6.1 mins. Note that affine invariance can be used to recode the data without affecting the resulting dependogram.

4.4 Serial independence of a binary sequence of zeros and ones

An i.i.d. sequence $W_1, \ldots, W_n$ of length $n = 100$ was used, where the binary variable W_i takes values 0 and 1 with probabilities 0.2 and 0.8, respectively. The product sequence Y_i is defined by $Y_i = W_i W_{i+3}$, $i = 1, \ldots, n-3$, which is dependent at lag 3. Figure 4 shows the dependograms (which took 8.25 mins) of the original sequence W_i and of the product sequence Y_i. Values of $p = 2$ or $p = 3$ could not possibly detect this dependence. The minimal value $p = 4$

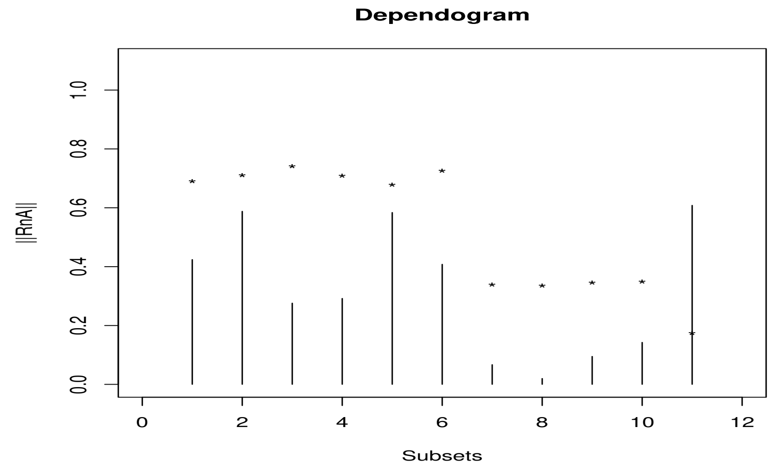

Fig. 3. This dependogram identifies the 4-dependence in the last subset $A = \{1, 2, 3, 4\}$. No other dependencies were declared significant.

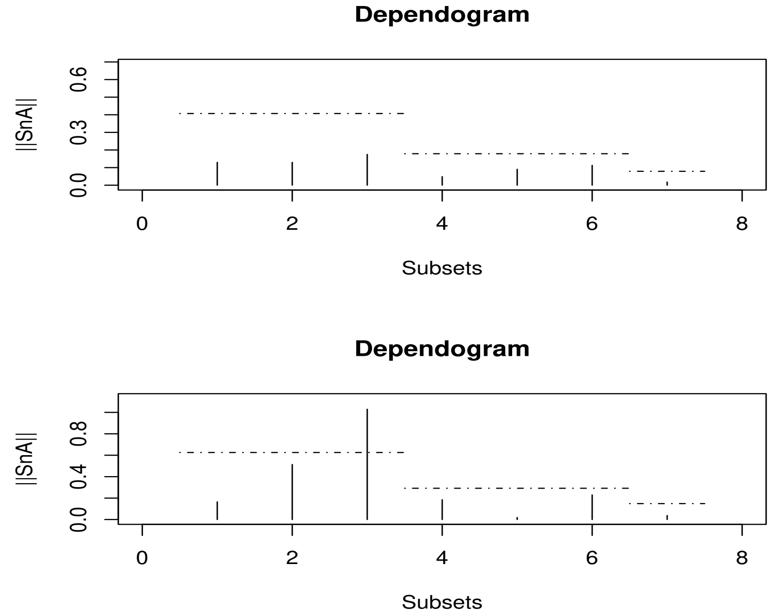

Fig. 4. The upper dependogram does not declare any serial dependence in the i.i.d. sequence W_i. The lower dependogram for the sequence Y_i exhibits a serial dependence at lag 3 through the subset 3 corresponding to $A = \{1, 4\}$.

was used, although a larger value could also have been used. The sequence could be recoded without any effect.

4.5 Serial independence of directional data

The number of test statistics for serial independence of a stationary sequence of observations on the sphere in $\mathbb{R}^q$ is very limited; see Watson [21]. For $(s^{(j)}, t^{(j)})$ in $\mathcal{S}_q \times [-1, 1]$, the half-space $H^{(j)} = \{x^{(j)} \in \mathcal{S}_q : \langle s^{(j)}, x^{(j)} \rangle \leq t^{(j)}\}$ becomes the polar cap with the pole located at $s^{(j)}$ and of size determined by the angle $\theta = \arccos(t^{(j)})$. Thus, the asymptotic covariance of the process $S_{n,A}$ is expressed in terms of probabilities of polar caps and intersections of polar caps. The test statistic based on $S_{n,A}$ is applicable in this context.

As an example of a sequence on the circle, an i.i.d. sequence U_i of length $n = 75$ was generated according to the bivariate normal distribution with mean vector 0 and covariance matrix I_2, the identity matrix. A sequence with serial dependence at lag 1 was created by letting $W_i = U_i + \sqrt{2}U_{i+1}$, $i = 1, \ldots, n-1$. This yields a correlation matrix $\text{cor}(W_i, W_{i+1}) = \frac{2}{3}I_2$ between two observations at lag 1. The final angular gaussian sequence on the circle is obtained by the normalization, $Y_i = W_i/|W_i|$. The minimal value $p = 2$ could detect this dependence, however the larger value $p = 3$ was used. The dependogram computed in 7.9 hours is shown in Figure 5.

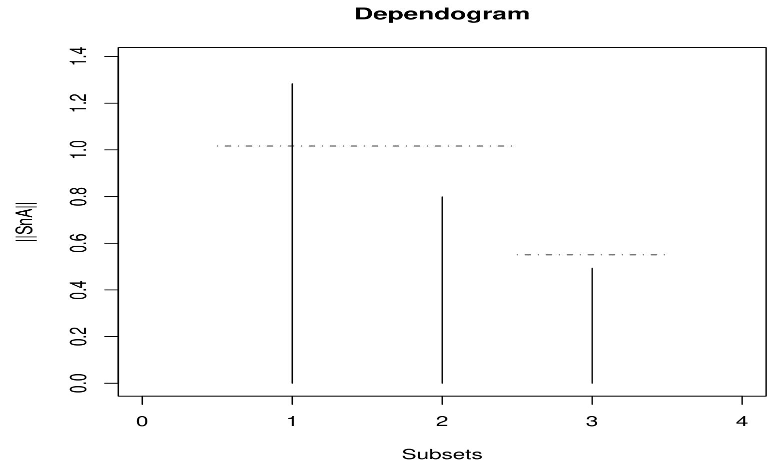

Fig. 5. The dependogram for the angular gaussian sequence Y_i on the circle exhibits a serial dependence at lag 1 through the first subset corresponding to $A = \{1, 2\}$.

The general technique presented yields also tests for serial independence of random axes (represented as random orthogonal projection matrices of rank one) or for serial independence of random rotations (represented as random orthogonal matrices). These tests and the directional example treated in detail are invariant under rotations of the coordinate system.

5 Fisher's combined p-values

Following Genest and Rémillard [11], if $F_{A,n}$ denotes the distribution function under the hypothesis of independence of $||R_{n,A}||_{\mathcal{F}}$, the p-values $1 - F_{A,n}(||R_{n,A}||_{\mathcal{F}})$ are approximately uniform on $[0, 1]$. Since the variables $||R_{n,A}||_{\mathcal{F}}$ are asymptotically independent (with a continuous limiting cdf), then combination of p-values in the manner of Fisher yields the overall test of independence

$$W_n = -2 \sum_{A \in \mathcal{I}_p} \log\{1 - F_{A,n}(||R_{n,A}||_{\mathcal{F}})\}$$

which should be approximately distributed as chi-square with $2(2^p - p - 1)$ degrees of freedom. However, $F_{A,n}$ being unknown, in practice, the test could be run as follows.

Step 1. Compute $||R_{n,A}||_{\mathcal{F}}$, for every $A \in \mathcal{I}_p$, from the original sample of size n.

Step 2. Generate $B = 2000$ (say) bootstrap samples of size n from the product measure $\mathbb{P}_n^{(1)} \times \cdots \times \mathbb{P}_n^{(p)}$. For each bootstrap sample, compute $||R^*_{n,A,i}||_{\mathcal{F}}$, $i = 1, \ldots, B$.

Step 3. Let

$$F^*_{A,n}(u) = \frac{1}{B} \sum_{i=1}^{B} \mathbb{I}\left\{||R^*_{n,A,i}||_{\mathcal{F}} \le u\right\}, \ u \ge 0.$$

Compute

$$\widehat{W}_n = -2 \sum_{A \in \mathcal{I}_p} \log\{1 - F^*_{A,n}(||R_{n,A}||_{\mathcal{F}})\}.$$

Step 4. An approximate p-value is given by

$$\frac{1}{B} \sum_{i=1}^{B} \mathbb{I}\{W^*_{n,i} \ge \widehat{W}_n\},$$

where

$$W^*_{n,i} = -2 \sum_{A \in \mathcal{I}_p} \log\{1 - F^*_{A,n}(||R^*_{n,A,i}||_{\mathcal{F}})\}, \ i = 1, \ldots, B.$$

Note that $F^*_{A,n}(||R^*_{n,A,i}||_{\mathcal{F}})$ is easily evaluated as R_i/B, where R_i is the rank of $||R^*_{n,A,i}||_{\mathcal{F}}$ among $||R^*_{n,A,1}||_{\mathcal{F}}, \ldots, ||R^*_{n,A,B}||_{\mathcal{F}}$.

A simulation was conducted, in the context of univariate discrete marginals, to investigate the speed of convergence to chi-square, as n increases. The assumed distribution is that of $(X^{(1)}, \ldots, X^{(p)})$ where all the p components are independent Poisson(1) variables. The simulation generates $M = 2000$ samples of size n from this assumed distribution. Then, for each of these M samples, a value $\widehat{W}_{n,i}$, $i = 1, \ldots, M$, was computed by going through steps 1-3 above. For a test of size $\alpha = 5\%$, the quantile $\widehat{W}_{n,\lfloor 0.95M \rfloor}$ is then compared to the 0.95 quantile of a chi-square distribution with $2(2^p - p - 1)$ degrees of freedom found in the last row of Table 3. Another way of making the comparison is to compute the empirical probability

$$P[\widehat{W}_n \geq c] \approx \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}\{\widehat{W}_{n,i} \geq c\},$$

where $c = \chi^2_{2(2^p-p-1)}(.95)$.

	$p=2$		$p=4$	
n	$\widehat{W}_{n,\lfloor 0.95M \rfloor}$	prob	$\widehat{W}_{n,\lfloor 0.95M \rfloor}$	prob
20	6.14	0.055	37.26	0.089
50	6.07	0.051	35.40	0.062
100	5.91	0.047	34.87	0.058
∞	5.99	0.050	33.92	0.050

Table 3
Critical points and empirical probabilities of Fisher's test with $\alpha = 5\%$ based on $M = 2000$ random samples of size n and $B = 2000$ bootstrap samples. Each of the n observations consists of p independent Poisson(1) variables.

For $p = 2$, the level of significance is reasonably satisfied even for values of n as small as $n = 20$. A larger value of n would be required when $p = 4$. A similar test can be devised for a test of serial independence based on $||S_{n,A}||_{\mathcal{F}}$. Presumably, still larger values of n would be required in this case since $n-p+1$ plays the role of n.

6 Numerical evaluation of $||R_{n,A}||_{\mathcal{F}}$

The sphere on $\mathbb{R}^1$ contains only two points ± 1. The sphere on $\mathbb{R}^d$ requires a discretization of its parametric space which is $(0,\pi)^{d-2} \times (0, 2\pi)$. The interval $(0,\pi)$ is discretized into N points and $(0, 2\pi)$ into $2N$ points. This gives $2N^{d-1}$ points on the sphere. For example, in a two components problem with $d_1 = 3$ and $d_2 = 2$, this gives, for $N = 10$, $200 \times 20 = 4000$ choices of directions $s^{(1)}$ and $s^{(2)}$. For each choice of directions, the $2^p - p - 1$ (singletons and the empty set excluded) processes are evaluated with the formula

$$\begin{aligned} R_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \prod_{k \in A} \Big[& \mathbb{I}\{X_i^{(k)} \in H(s^{(k)}, t^{(k)})\} \\ & - P_n^{(k)}(H(s^{(k)}, t^{(k)}))\Big]. \end{aligned}$$

For a given choice of $s^{(1)}, \ldots, s^{(p)}$, we need to evaluate p matrices of order $n \times n$ constructed as follows. For $s^{(k)}$, evaluate the n values $\langle X_j^{(k)}, s^{(k)} \rangle \equiv t_j^{(k)}$. The $n \times n$ matrix for $s^{(k)}$ has an element in position (i, j) given by

$$\mathbb{I}\{\langle X_i^{(k)}, s^{(k)} \rangle \leq t_j^{(k)}\}.$$

This matrix is then modified by subtracting to each element (i, j) the proportion of ones in column j. For a given subset A, one multiplies together the appropriate $|A|$ such matrices to obtain an $n \times n$ matrix Ψ_A (say). An n-vector ψ_A is then obtained by adding the rows of Ψ_A. The maximum value of $|R_{n,A}|$ (for the given choice of directions) is the largest (in absolute value) component of ψ_A divided by $\sqrt{n}$. The global max is obtained by varying the choice of directions. When changing directions one can rewrite over the previous $n \times n$ matrices used before.

A random search to reduce computational times can be done easily. One generates random vectors $s^{(k)}$ uniformly distributed over the unit spheres (normed multivariate $N_{d_k}(0, I)$). The rest of the algorithm is the same as with the non-random procedure.

A Proofs

The proofs of Theorems 1, 2 and 3 follow from modifications to those in Ghoudi *et al.* [12]. They are included for completeness.

Proof of Theorem 1. Following van der Vaart and Wellner [20, p. 35], weak convergence of the marginals and asymptotic tightness are established. As in Ghoudi *et al.* [12], the multinomial formula yields the equivalent representation as an i.i.d. sum

$$\begin{aligned}\breve{R}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^{p}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \prod_{k \in A} \Big[& \mathbb{I}\{X_i^{(k)} \in H(s^{(k)}, t^{(k)})\} \\ & - P^{(k)}(H(s^{(k)}, t^{(k)})) \Big] .\end{aligned} \tag{A.1}$$

Finite dimensional convergence for the pair

$$\Big(\breve{R}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^{p}), \breve{R}_{n,B}((\tilde{s}^{(j)}, \tilde{t}^{(j)})_{j=1}^{p})\Big)$$

thus follows from the central limit theorem. If $A = B$, the asymptotic covariance is that given in Theorem 1. However, if $A \neq B$, then there is at least one index in A which is not in B (or the converse). Then, for this index k,

$$E\left[\mathbb{I}\{X_i^{(k)} \in H(s^{(k)}, t^{(k)})\} - P^{(k)}(H(s^{(k)}, t^{(k)}))\right] = 0.$$

This proves weak convergence of the marginals to the appropriate Gaussian distribution. Next, another representation used in Ghoudi *et al.* [12] becomes

$$\breve{R}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \sum_{B \subset A} (-1)^{|A \setminus B|} \prod_{j \in A \setminus B} P^{(j)}(H(s^{(j)}, t^{(j)}))$$
$$\cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\prod_{j \in B} \mathbb{I}\{X_i^{(j)} \in H(s^{(j)}, t^{(j)})\} - \prod_{j \in B} P^{(j)}(H(s^{(j)}, t^{(j)}))\right],$$

for half-space processes. For each subset B, the last sum over i is an empirical process indexed by sets in the collection $\mathcal{F}_B = \times_{j \in B} \mathcal{F}^{(d_j)}$. Since each collection $\mathcal{F}^{(d_j)}$ is a Vapnik-Červonenkis class of index $d_j + 2$ (van der Vaart and Wellner [20, p. 152]), then $\mathcal{F}_B$ is also a VC-class of index $\sum_{j \in B} d_j + |B| + 1$ (van der Vaart and Wellner [20, p. 147]). Asymptotic tightness is thus satisfied for each B and since there is only a finite number of B's, the proof is complete. □

Proof of Theorem 2. The following expression holds

$$R_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) - \breve{R}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \sum_{B \subset A, B \neq \emptyset} (-1)^{|B|}$$
$$\cdot \prod_{j \in B} \left[\mathbb{P}_n^{(j)}(H(s^{(j)}, t^{(j)})) - P^{(j)}(H(s^{(j)}, t^{(j)}))\right] \cdot \breve{R}_{n,A \setminus B}((s^{(k)}, t^{(k)})_{k=1}^p).$$

Hence,

$$||R_{n,A} - \breve{R}_{n,A}||_{\mathcal{F}} \leq \sum_{B \subset A, B \neq \emptyset} \prod_{j \in B} ||\mathbb{P}_n^{(j)} - P^{(j)}||_{\mathcal{F}^{(d_j)}} \cdot ||\breve{R}_{n,A \setminus B}||_{\mathcal{F}}$$

The result follows because $||\breve{R}_{n,A \setminus B}||_{\mathcal{F}}$ converges weakly from Theorem 1 and since the Glivenko-Cantelli theorem for half-spaces in Wolfowitz [22] holds. □

Proof of Theorem 3. As in the non serial context, use the expression

$$\breve{S}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = \frac{1}{\sqrt{n'}} \sum_{i=1}^{n'} \prod_{k \in A} \left[\mathbb{I}\{X_i^{(k)} \in H(s^{(k)}, t^{(k)})\}\right.$$
$$\left. - Q(H(s^{(k)}, t^{(k)}))\right].$$

Weak convergence of the marginals is proved. Because of the overlapping of Y_j's in consecutive X_i's, the X_i's form an m-dependent sequence with $m = p-1$, see *e.g.* Ferguson [10, p. 69]. Thus, the central limit theorem for m-dependent sequences establishes that $\breve{S}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^{p})$ and $\breve{S}_{n,B}((\tilde{s}^{(j)}, \tilde{t}^{(j)})_{j=1}^{p})$ are asymptotically and jointly normal with asymptotic covariance $\sigma_{00} + 2\sigma_{01} + \cdots + 2\sigma_{0m}$, where

$$\sigma_{0u} = E\left\{\prod_{k\in A}\left[\mathbb{I}\{X_i^{(k)} \in H(s^{(k)}, t^{(k)})\} - Q(H(s^{(k)}, t^{(k)}))\right]\right.$$
$$\left.\cdot \prod_{k\in B}\left[\mathbb{I}\{X_{i+u}^{(k)} \in H(\tilde{s}^{(k)}, \tilde{t}^{(k)})\} - Q(H(\tilde{s}^{(k)}, \tilde{t}^{(k)}))\right]\right\}.$$

All of the above expectations are null unless $A = B$ (both in $\mathcal{A}_p$) and $u = 0$. Next, to establish asymptotic tightness, assume without loss of generality that n' is a multiple of p, say $n' = rp$. This amounts to neglecting at most $p - 1$ terms in the sequence. Rewrite the sequence $X_1, X_2, \ldots$ as an array with p rows, each consisting of r i.i.d. vectors,

$$\begin{array}{llll} X_1 & X_{1+p} & \cdots & X_{1+(r-1)p} \\ X_2 & X_{2+p} & \cdots & X_{2+(r-1)p} \\ \vdots & \vdots & \ddots & \vdots \\ X_p & X_{p+p} & \cdots & X_{p+(r-1)p}. \end{array}$$

Then, the expression

$$\breve{S}_{n,A}((s^{(j)}, t^{(j)})_{j=1}^{p}) = \frac{1}{\sqrt{p}}\sum_{h=1}^{p}\sum_{B\subset A}(-1)^{|A\setminus B|}\prod_{j\in A\setminus B} Q(H(s^{(j)}, t^{(j)}))$$
$$\cdot \frac{1}{\sqrt{r}}\sum_{i=0}^{r-1}\left[\prod_{j\in B}\mathbb{I}\{X_{pi+h}^{(j)} \in H(s^{(j)}, t^{(j)})\} - \prod_{j\in B} Q(H(s^{(j)}, t^{(j)}))\right]$$

establishes asymptotic tightness since for each pair (h, B) in finite number, the last sum over i is an empirical process indexed by a VC-class. Finally, the proof that $S_{n,A}$ and $\breve{S}_{n,A}$ are equivalent processes is based on the inequality

$$||S_{n,A} - \breve{S}_{n,A}||_{\mathcal{F}} \le \sum_{B\subset A, B\neq\emptyset} ||\mathbb{Q}_n - Q||_{\mathcal{F}^{(q)}}^{|B|} \cdot ||\breve{S}_{n,A\setminus B}||_{\mathcal{F}}$$

and is the same as for Theorem 2. □

Proof of Theorem 4. Define

$$\breve{R}^*_{n,A}((s^{(j)},t^{(j)})_{j=1}^p) = \sqrt{n}\sum_{B\subset A}(-1)^{|A\setminus B|}\hat{\mathbb{P}}_n(\times_{j=1}^p H^B(s^{(j)},t^{(j)}))$$
$$\cdot \prod_{j\in A\setminus B} P_n^{(j)}(H(s^{(j)},t^{(j)})).$$

With the representation (A.1), the Lindeberg condition for the triangular array of the i.i.d. random variables

$$\prod_{k\in A}\left[\mathbb{I}\{X_{ni}^{(k)}\in H(s^{(k)},t^{(k)})\} - P_n^{(k)}(H(s^{(k)},t^{(k)}))\right],\ i=1,2,\ldots$$

is applied. Thus, the finite dimensional convergence for the pair

$$\left(\breve{R}^*_{n,A}((s^{(j)},t^{(j)})_{j=1}^p), \breve{R}^*_{n,B}((\tilde{s}^{(j)},\tilde{t}^{(j)})_{j=1}^p)\right)$$

follows with the same limiting normal distribution as in Theorem 1. The other representation

$$\breve{R}^*_{n,A}((s^{(j)},t^{(j)})_{j=1}^p) = \sum_{B\subset A}(-1)^{|A\setminus B|}\left[\prod_{j\in A\setminus B} P_n^{(j)}(H(s^{(j)},t^{(j)}))\right] \cdot U_{n,B}((s^{(j)},t^{(j)})_{j\in B}), \tag{A.2}$$

where

$$U_{n,B}((s^{(j)},t^{(j)})_{j\in B}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\prod_{j\in B}\mathbb{I}\{X_{ni}^{(j)}\in H(s^{(j)},t^{(j)})\} - \prod_{j\in B}P_n^{(j)}(H(s^{(j)},t^{(j)}))\right], \tag{A.3}$$

is used to establish asymptotic tightness. Since the term between brackets in (A.2) converges as $n\to\infty$, it suffices to establish that the triangular array empirical process $U_{n,B}$ is asymptotically tight. This follows from Lemma 2.8.8 in van der Vaart and Wellner [20, p. 174] with the semimetric on $\times_{j\in B}\mathcal{F}^{(d_j)}$, for their condition (2.8.5) implied by our condition (3), defined with the symmetric difference between half-spaces

$$\rho_P\left((H(s^{(j)},t^{(j)}))_{j\in B},(H(\tilde{s}^{(j)},\tilde{t}^{(j)}))_{j\in B}\right)$$
$$= \sum_{j\in B}P^{(j)}\left[H(s^{(j)},t^{(j)})\triangle H(\tilde{s}^{(j)},\tilde{t}^{(j)})\right].$$

Finally, $||R^*_{n,A} - \breve{R}^*_{n,A}||_{\mathcal{F}} \to 0$ in outer probability follows as in Theorem 2 by establishing $||\hat{\mathbb{P}}^{(j)}_n - P^{(j)}_n||_{\mathcal{F}^{(d_j)}} \to 0$ in outer probability, for each $j \in A$. This last assertion follows again from Lemma 2.8.8 cited above whose conclusion is that $\sqrt{n}(\hat{\mathbb{P}}^{(j)}_n - P^{(j)}_n)$ converges weakly. □

Proof of Theorem 5. Define the process

$$\breve{S}^*_{n,A}((s^{(j)}, t^{(j)})^p_{j=1}) = \frac{1}{\sqrt{n'}} \sum_{i=1}^{n'} \prod_{k\in A} \Big[\mathbb{I}\{X^{(k)}_{ni} \in H(s^{(k)}, t^{(k)})\} \\ -Q_n(H(s^{(k)}, t^{(k)}))\Big] .$$

Asymptotic normality of the marginals is established by considering any linear combination

$$a\breve{S}^*_{n,A}((s^{(j)}, t^{(j)})^p_{j=1}) + b\breve{S}^*_{n,B}((\tilde{s}^{(j)}, \tilde{t}^{(j)})^p_{j=1}).$$

This linear combination involves the sum of the variables

$$L_{n,i} = a \prod_{j\in A} \Big[\mathbb{I}\{X^{(j)}_{ni} \in H(s^{(j)}, t^{(j)})\} - Q_n(H(s^{(j)}, t^{(j)}))\Big] \\ +b \prod_{j\in B} \Big[\mathbb{I}\{X^{(j)}_{ni} \in H(\tilde{s}^{(j)}, \tilde{t}^{(j)})\} - Q_n(H(\tilde{s}^{(j)}, \tilde{t}^{(j)}))\Big] ,\ i = 1, \dots, n',$$

which form a triangular array in which each row is an m-dependent sequence. At this point, the proof is a slight extension to triangular arrays of the classical proof of the central limit theorem for m-dependent sequences, see *e.g.* Theorem 11 in Ferguson [10, p. 70], in which intervenes Markov's idea of splitting the sum into big blocks and little blocks. Write $n' = s(k + m)$ and let

$$S'_{n,k} = \sum_{j=0}^{s-1} V_{n,k,j},\ \ S''_{n,k} = \sum_{j=0}^{s-1} W_{n,k,j},$$

where the big blocks of size k and the little blocks of size m are, respectively,

$$V_{n,k,j} = \sum_{i=1}^{k} L_{n,j(k+m)+i},\ \ W_{n,k,j} = \sum_{i=k+1}^{k+m} L_{n,j(k+m)+i}.$$

Then, the $V_{n,k,j}$, for $j = 0, \dots, s - 1$, are i.i.d. variables with distribution depending on k and n.

Firstly, it is shown that $S'_{n,k}/\sqrt{n}$ is asymptotically normal when k is fixed and

$n \to \infty$ (and thus $s \to \infty$). Since the variables $V_{n,k,j}$ are independent, the Lindeberg condition amounts to

$$\frac{1}{\mathrm{Var}(V_{n,k,j})} E\left[V_{n,k,j}^2 \mathbb{I}\left\{|V_{n,k,j}| \geq \epsilon\,(s\ \mathrm{Var}(V_{n,k,j}))^{\frac{1}{2}}\right\}\right] \to 0, \text{ as } n \to \infty.$$

Note that $|V_{n,k,j}| \leq k(|a| + |b|)$ and

$$\begin{aligned}
\frac{1}{k}\mathrm{Var}(V_{n,k,j}) \overset{n\to\infty}{\to} &\ a^2 \prod_{j\in A}[Q(H(s^{(j)}, t^{(j)})) - Q(H(s^{(j)}, t^{(j)}))^2] \\
+ &\ b^2 \prod_{j\in B}[Q(H(\tilde{s}^{(j)}, \tilde{t}^{(j)})) - Q(H(\tilde{s}^{(j)}, \tilde{t}^{(j)}))^2] \\
+ &\ \mathbb{I}\{A = B\}\, 2ab \prod_{j\in A}[Q(H(s^{(j)}, t^{(j)}) \cap H(\tilde{s}^{(j)}, \tilde{t}^{(j)})) \\
&\qquad\qquad -Q(H(s^{(j)}, t^{(j)}))Q(H(\tilde{s}^{(j)}, \tilde{t}^{(j)}))] \\
= &\ \mathrm{Var}[aS_A((s^{(j)}, t^{(j)})_{j=1}^p) + bS_B((\tilde{s}^{(j)}, \tilde{t}^{(j)})_{j=1}^p)] \\
\equiv &\ c.
\end{aligned}$$

Thus, the Lindeberg condition is satisfied and one may conclude that, as $n \to \infty$, $S'_{n,k}/\sqrt{s} \xrightarrow{D} N(0, ck)$ and also $S'_{n,k}/\sqrt{n} \xrightarrow{D} Z_k$, where $Z_k \sim N(0, ck/(k+m))$. Finally, if $k \to \infty$, then $Z_k \xrightarrow{D} N(0, c)$.

Secondly, it is established that $S''_{n,k}/\sqrt{n} \to 0$ in probability as $k \to \infty$, uniformly in n. Let $S_{n,m} = \sum_{i=1}^m L_{n,i}$. The inequalities

$$P\left[\frac{|S''_{n,k}|}{\sqrt{n}} > \delta\right] \leq \mathrm{Var}\left(S''_{n,k}\right)\frac{1}{n\delta^2} \leq \mathrm{Var}\left(S_{n,m}\right)\frac{1}{k\delta^2} \tag{A.4}$$

hold. It can be checked that $\mathrm{Var}\,(S_{n,m})$ converges to some positive constant as $n \to \infty$. Thus, one can find an upper bound (not depending on n) for (A.4) converging to 0 as $k \to \infty$.

Asymptotic tightness is established with the representation in the proof of Theorem 3

$$\begin{aligned}
\breve{S}^*_{n,A}((s^{(j)}, t^{(j)})_{j=1}^p) = &\frac{1}{\sqrt{p}}\sum_{h=1}^p \sum_{B\subset A}(-1)^{|A\setminus B|}\left[\prod_{j\in A\setminus B} Q_n(H(s^{(j)}, t^{(j)}))\right] \\
&\cdot W_{n,h,B}((s^{(j)}, t^{(j)})_{j\in B}),
\end{aligned}$$

where

$$W_{n,h,B}((s^{(j)},t^{(j)})_{j\in B}) = \frac{1}{\sqrt{r}} \sum_{i=0}^{r-1} \left[\prod_{j\in B} \mathbb{I}\{X_{n,pi+h}^{(j)} \in H(s^{(j)},t^{(j)})\} - \prod_{j\in B} Q_n(H(s^{(j)},t^{(j)})) \right].$$

Thus, it suffices that the processes $W_{n,h,B}$ be asymptotically tight, for all h and B, and this follows in the same manner as for the process $U_{n,B}$ in the proof of Theorem 4. The equivalence of the two processes, $||S_{n,A}^* - \breve{S}_{n,A}^*||_{\mathcal{F}} \to 0$ in outer probability, is also derived as in Theorem 4. □

References

[1] R. Beran, The impact of the bootstrap on statistical algorithms and theory, *Statistical Science* **18** (2003) 175–184.

[2] R. Beran, P. W. Millar, Confidence sets for a multivariate distribution, *The Annals of Statistics* **14** (1986) 431–443.

[3] M. Bilodeau, P. Lafaye de Micheaux, A multivariate empirical characteristic function test of independence with normal marginals, *Journal of Multivariate Analysis* **95** (2005) 345–369.

[4] J. R. Blum, J. Kiefer, M. Rosenblatt, Distribution free tests of independence based on the sample distribution function, *The Annals of Mathematical Statistics* **32** (1961) 485–498.

[5] R. Cléroux, A. Lazraq, Y. Lepage, Vector correlation based on ranks and a nonparametric test of no association between vectors, *Communications in Statistics Theory and Methods* **24** (1995) 713–733.

[6] H. Cramér, H. Wold, Some theorems on distribution functions, *Journal of the London Mathematical Society* **11** (1936) 290–294.

[7] J. Dehardt, Generalization of the Glivenko-Cantelli theorem, *The Annals of Mathematical Statistics* **42** (1971) 2050–2055.

[8] P. Deheuvels, Indépendance multivariée partielle et inégalités de Fréchet, in: *Studies in probability and related topics*, Nagard, Rome, 1983, pp. 145–155.

[9] P. Deheuvels, An asymptotic decomposition for multivariate distribution-free tests of independence, *Journal of Multivariate Analysis* **11** (1981) 102–113.

[10] T. S. Ferguson, *A course in large sample theory*, Chapman & Hall, London, 1996.

[11] C. Genest, B. Rémillard, Tests of independence and randomness based on the empirical copula process, *Test* **13** (2004) 335–369.

[12] K. Ghoudi, R. J. Kulperger, B. Rémillard, A nonparametric test of serial independence for time series and residuals, *Journal of Multivariate Analysis* **79** (2001) 191–218.

[13] P. W. Gieser, R. H. Randles, A nonparametric test of independence between two vectors, *Journal of the American Statistical Association* **92** (1997) 561–567.

[14] W. Hoeffding, A non-parametric test of independence, *The Annals of Mathematical Statistics* **19** (1948) 546–557.

[15] J. Hoffmann-Jørgensen, Stochastic processes on Polish spaces, *Various Publications Series*, **39**, Aarhus Universitet Matematisk Institut, Aarhus, 1991.

[16] R. J. Muirhead, C. M. Waternaux, Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika* **67** (1980) 31–43.

[17] M. L. Puri, P. K. Sen, *Nonparametric methods in multivariate analysis*, New York, Wiley, 1971.

[18] J. M. Steele, Empirical discrepancies and subadditive processes, *The Annals of Probability* **6** (1978) 118–127.

[19] Y. Um, R. H. Randles, A multivariate nonparametric test of independence among many vectors, *Journal of Nonparametric Statistics* **13** (2001) 699–708.

[20] A. W. van der Vaart, J. A. Wellner, *Weak convergence and empirical processes*, Springer Series in Statistics, Springer-Verlag, New York, 1996.

[21] G. S. Watson, *Statistics on spheres*, University of Arkansas Lecture Notes in the Mathematical Sciences, **6**, John Wiley & Sons Inc., New York, 1983.

[22] J. Wolfowitz, Generalization of the theorem of Glivenko-Cantelli, *The Annals of Mathematical Statistics* **25** (1954) 131–138.