Multivariate Nonparametric Tests of Independence

Y. Fan\(^2\), P. Lafaye de Micheaux\(^{1,2}\), S. Penev\(^2\), D. Salopek\(^2\)

\(^1\)Professor of Statistics  
CREST-ENSAI, France  
\(^2\)School of Mathematics and Statistics,  
University of New South Wales, Australia

Australian Statistical Conference  
Hotel Realm, Canberra  
December 8, 2016
An example of non-monotone dependence


- Pearson’s correlation is equal to 0.139, with a p-value of 0.095.
- Spearman’s rho is equal to -0.035, with a p-value of 0.675.
- Kendall’s tau is equal to -0.032, with a p-value of 0.574.

We see a pattern ⇒ another tool to measure dependence is needed.
[Bakirov et al., 2006], [Székely et al., 2007] developed a dependence measure between TWO random vectors, called distance correlation:

- $0 \leq d_{\text{Cor}}(x, y) \leq 1$;
- $d_{\text{Cor}}(x, y) = 1 \Rightarrow y = \alpha + Bx$ (also true for the empirical);
- $d_{\text{Cor}}(x, y) = 0 \Leftrightarrow x$ and $y$ independent $\Rightarrow x$ and $y$ non correlated;
- can detect any form of dependence!

**Note**: The empirical version of $d_{\text{Cor}}(x, y)$, noted $d_{\text{cor}}(x, y)$, has a simple formula built using empirical characteristic functions.

- Implemented in the energy R package.
Independence test based on the distance correlation

Let be given two observed samples $X : n \times p$ and $Y : n \times q$, each constituted of $n$ realizations of the random vectors $\mathbf{x}$ and $\mathbf{y}$. Define:

$$a_{kl} = |x_k - x_l|_p, \quad A_{kl} = a_{kl} - \bar{a}_k - \bar{a}_l + \bar{a}_*,$$
$$b_{kl} = |y_k - y_l|_q, \quad B_{kl} = b_{kl} - \bar{b}_k - \bar{b}_l + \bar{b}_*,$$

$$dcov^2(X, Y) = n^{-2} \sum_{k,l=1}^{n} A_{kl} B_{kl} \quad \text{and} \quad s_2 = n^{-2} \sum_{k,l=1}^{n} a_{kl} n^{-2} \sum_{k,l=1}^{n} b_{kl}.$$

We reject, at significance level at most $\alpha$ (use of a bound, or of random permutations), the hypothesis of independence between $\mathbf{x}$ and $\mathbf{y}$ if

$$\sqrt{n} \times dcov(X, Y) / \sqrt{s_2} > \Phi^{-1}(1 - \alpha/2).$$
Is it necessary to go further than pairwise dependence?

\[ \begin{align*}
\begin{cases}
    x_1 \\
x_2 \\
y
\end{cases} \quad 3 \text{ sample } (n \times 1) \text{ i.i.d. } N(0, 1) \quad \text{and} \quad x_3 = \begin{cases}
    +\text{abs}(y), & \text{if } x_1x_2 > 0; \\
    -\text{abs}(y), & \text{otherwise.}
\end{cases}
\end{align*} \]

\( x_1, x_2 \) and \( x_3 \) are pairwise independent ...
... but ...  
Values of $x_3$ such that $x_1 \times x_2 > 0$ are all positive!

0.6408445 1.6013778 1.7164682 0.4069357 1.7282346 ... 

Values of $x_3$ such that $x_1 \times x_2 \leq 0$ are all negative!

-0.7778154 -1.6473925 -0.1542662 -1.1756313 -1.1021281 ... 

If you give me the sign of $x_1 \times x_2$, I can give you the sign of $x_3$!
Going further than pairwise dependence

Hence, here the interaction with a third variable is relevant to understand the relation between birth rate and mortality rate.

* Credit: G. Boglioni
Towards mutual dependence between \( p \) sub-vectors

We thus need a test of independence between several variables.

We have generalized the *distance correlation* of [Bakirov et al., 2006] and [Székely et al., 2007] to obtain a test procedure for investigating the mutual dependence between \( p \geq 2 \) vectors. (First improvement over the distance correlation test)
Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a random sample, namely $n$ independent copies of

$$
\mathbf{x} = \begin{pmatrix} 
\mathbf{x}^1 \\
\vdots \\
\mathbf{x}^p 
\end{pmatrix} \in \mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}.
$$

Similarly, we note $\mathbf{t} = (t_1^T, \ldots, t_p^T)^T \in \mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}$ to denote a column vector of length $q := \sum_{\ell=1}^p q_\ell$.

We want to test the mutual independence of the $p$ sub-components of $\mathbf{x}$. This is equivalent to test the null hypothesis

$$
\mathcal{H}_0 : \varphi(\mathbf{t}) = \prod_{\ell=1}^p \varphi_\ell(t_\ell), \quad \forall \mathbf{t} \in \mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p},
$$

where $\varphi(\mathbf{t})$ is the true (unknown) characteristic function of $\mathbf{x}$ at point $\mathbf{t}$ and $\varphi_\ell(t_\ell)$ is the true (unknown) characteristic function of $\mathbf{x}_\ell$ at point $t_\ell$, $\ell = 1, \ldots, p$. 

Nonparametric Tests of dependence
A natural nonparametric empirical estimator of $\varphi(t)$ is

$$
\hat{\varphi}_n(t) = n^{-1} \sum_{j=1}^{n} e^{it^T X_j} = n^{-1} \sum_{j=1}^{p} e^{i t^\ell X^\ell_j}.
$$

Similarly, a nonparametric estimator of $\varphi_\ell(t_\ell)$ is given by

$$
\hat{\varphi}_{n,\ell}(t_\ell) = n^{-1} \sum_{j=1}^{n} e^{it_\ell^T X^\ell_j}.
$$

The test statistics (location+orthogonal invariant) we propose for testing $\mathcal{H}_0$ are

$$
n T_n(w) = \| \sqrt{n} D_n \|_w^2 = n \int_{\mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}} \left| D_n(t) \right|^2 w(t) \, dt,
$$

where

$$
D_n(t) = \hat{\varphi}_n(t) - \prod_{\ell=1}^{p} \hat{\varphi}_{n,\ell}(t_\ell) \in \mathbb{C}
$$

and where $w(t) = \prod_{\ell=1}^{p} v(t_\ell)$ are some properly chosen weight functions.
We obtained the following closed-form expression for our new measure of multivariate dependence

\[
 n T_n(w) := \int_{\mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}} \left| \sqrt{n} \left( \tilde{\varphi}_n(t) - \prod_{\ell=1}^p \tilde{\varphi}_{n,\ell}(t_{\ell}) \right) \right|^2 \prod_{\ell=1}^p v(t_{\ell}) dt
\]

\[
= n^{-1} \sum_{j,j'=1}^n \prod_{\ell=1}^p \xi_{j,j',\ell} - 2 \left\{ \sum_{j=1}^n \prod_{\ell=1}^p \left[ n^{-1} \sum_{j'=1}^n \xi_{j,j',\ell} \right] \right\} + n \prod_{\ell=1}^p \left[ n^{-2} \sum_{j,j'=1}^n \xi_{j,j',\ell} \right]
\]

where

\[
\xi_{j,j',\ell} = \int_{\mathbb{R}^{q_\ell}} \cos \left( t_{\ell}^T (x_j^{\ell} - x_{j'}^{\ell}) \right) v(t_{\ell}) dt_{\ell} \in \mathbb{R}.
\]

We also worked on other choices for the weight function with which we also obtained closed form expression. (Second improvement over the distance correlation test which considered only one specific weight function)

To construct the rejection region of the test, it would be great to know the (asymptotic) null distribution of \( n T_n(w) \). We have done this in two steps.
Step 1.

**Theorem 1**

If $x^1, \ldots, x^p$ are mutually independent, then \( \sqrt{n}D_n = \sqrt{n}(\hat{\phi}_n - \prod_{\ell=1}^{p} \hat{\phi}_{n,\ell}) \) converges (weak convergence of processes) in $C(\mathbb{R}^q, \mathbb{C})$ to a zero mean complex Gaussian process $D$ having covariance function given by

\[
C(s, t) = \mathbb{E}[D(s)D(t)] = \prod_{\ell=1}^{p} \varphi_{\ell}(-t_\ell + s_\ell) - \prod_{\ell=1}^{p} \left[ \varphi_{\ell}(-t_\ell) \varphi_{\ell}(s_\ell) \right] \left\{ 1 - p + \sum_{\ell=1}^{p} \frac{\varphi_{\ell}(-t_\ell + s_\ell)}{\varphi_{\ell}(-t_\ell) \varphi_{\ell}(s_\ell)} \right\}
\]

and pseudo-covariance function given by

\[
P(s, t) = \mathbb{E}[D(s)D(t)] = \prod_{\ell=1}^{p} \varphi_{\ell}(t_\ell + s_\ell) - \prod_{\ell=1}^{p} \left[ \varphi_{\ell}(t_\ell) \varphi_{\ell}(s_\ell) \right] \left\{ 1 - p + \sum_{\ell=1}^{p} \frac{\varphi_{\ell}(t_\ell + s_\ell)}{\varphi_{\ell}(t_\ell) \varphi_{\ell}(s_\ell)} \right\}.
\]
Sketch of proof: To prove that $\sqrt{n}D_n$ converges in $C(\mathbb{R}^q, \mathbb{C})$ to a complex Gaussian process $D$, we use

**Proposition 1.1 (Proposition 14.6 of [Kallenberg, 2002])**

Let $X, X_1, X_2, \ldots$ be random elements in $C(T, S)$ (endowed with the topology of uniform convergence on compacts), where $S$ is a metric space and $T$ is locally compact, second-countable, and Hausdorff (abbreviated as lcscH). Then $X_n \xrightarrow{d} X$ iff the convergence holds for the restrictions to arbitrary compact subsets $K \subset T$.

The metric $\rho$ of uniform convergence on all compact sets on $C(\mathbb{R}^q, \mathbb{C})$ is defined by

$$
\rho(x, y) := \sum_{j=1}^{\infty} 2^{-j} \frac{\rho_j(x, y)}{1 + \rho_j(x, y)},
$$

where $\rho_j(x, y) = \sup_{\|t\| \leq j} |x(t) - y(t)|$ is the usual sup norm.
Sketch of proof (cont’d) : It thus suffices to show that $r_j(\sqrt{n}D_n) \xrightarrow{d} r_j(D)$ in order to show the convergence $\sqrt{n}D_n \xrightarrow{d} D$, where, for all mapping $x \in C(\mathbb{R}^q, \mathbb{C})$, $r_j(x)$ is the restriction of $x$ to the ball $B^q_j$ of radius $j$.

We prove a straightforward generalization of [Csörgo, 1985, Theorem p. 294], who worked on the same processes as we did (defined for random variables, not random vectors though) but restricted to a compact set.

We also need his (technical) condition (∗) to be satisfied (mild tail condition on the underlying joint distribution).

It then remains to compute the moments of the limiting Gaussian, which is not too difficult.
Theorem 2

We have

\[ nT_n(w) = \| \sqrt{n}D_n \|_w^2 = \int_{\mathbb{R}^q_1 \times \cdots \times \mathbb{R}^q_p} |\sqrt{n}D_n(t)|^2 w(t) \, dt \overset{L}{\rightarrow} \sum_{k=1}^{\infty} \frac{\lambda_k}{2} \left( (1 + |p_k|) \xi_k + (1 - |p_k|) \eta_k \right) \]

Here \( \xi_k, \eta_k, k = 1, 2, \ldots, m, \ldots \) are independent pairs of independent chi-square random variables, each with one degree of freedom and \( \lambda_k \) (respectively \( f_k \)) are the eigenvalues (respectively eigenfunctions) of the integral operator \( O \) defined by

\[ O(f)(x) = \int_{\mathbb{R}^q} f(y) C(x, y) w(y) \, dy, \]

The values \( p_k \) are defined as

\[ p_k = \int f_k(s) f_k(-s) w(s) \, ds. \]
Sketch of proof: We use the nice properties of our limiting covariance function $C(\cdot, \cdot)$ (hermitian, positive definite) and a generalization [Ferreira and Menegatto, 2012] (to $\mathbb{R}^q$) of the Mercer’s theorem which states that there exists a sequence of non-negative real numbers $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and an orthonormal (with respect to $\langle \cdot, \cdot \rangle_w$) family $\{f_k\}$ such that

$$C(s, t) = \sum_{k=1}^{\infty} \lambda_k f_k(s) f_k(t),$$

where $\langle f, g \rangle_w = \int f(t) g(t) w(t) dt$. This enables us to write:

$$\int C(s, t) f_j(t) w(t) dt = \int \sum_{k=1}^{\infty} \lambda_k f_k(s) f_k(t) f_j(t) w(t) dt$$

$$= \sum_{k=1}^{\infty} \lambda_k f_k(s) \int f_j(t) f_k(t) w(t) dt$$

$$= \sum_{k=1}^{\infty} \lambda_k f_k(s) \delta_{jk} = \lambda_j f_j(s).$$

$\Rightarrow$ They are solutions of the integral operator $O$. 

Nonparametric Tests of dependence  ASC 2016 16/27
Sketch of proof (cont’d) : Now, an expansion of the complex Gaussian process \( D \) on our \( \{ f_k \} \) orthonormal basis gives:

\[
D(t) = \sum_{k=1}^{\infty} \langle D, f_k \rangle f_k(t).
\]

So

\[
\int |D(t)|^2 w(t) \, dt = \int \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \langle D, f_k \rangle \overline{\langle D, f_{k'} \rangle} f_k(t) \overline{f_{k'}(t)} \, dt = \sum_{k=1}^{\infty} |\langle D, f_k \rangle|^2
\]

where

\[
\langle D, f_k \rangle = \int D(t) \overline{f_k(t)} w(t) \, dt \sim CN_1(0, \lambda_k, \lambda_k p_k)
\]

(the complex normal distribution) where \( p_k = \int f_k(s) f_k(-s) w(s) ds \) and where

\[
|\langle D, f_k \rangle|^2 \sim \frac{\lambda_k}{2} \left( (1 + |p_k|) \chi_1^2 + (1 - |p_k|) \chi_1^2 \right)
\]

from [Ducharme et al., 2016, Theorem 3.3].
We have thus obtained the (asymptotic) null distribution of our test statistic $nT_n(w)$: infinite weighted sum of independent chi-square r.v.

Unfortunately, since

$$C(s,t) = \prod_{\ell=1}^{p} \varphi_{\ell}(-t_{\ell} + s_{\ell}) - \prod_{\ell=1}^{p} \left[ \varphi_{\ell}(-t_{\ell}) \varphi_{\ell}(s_{\ell}) \right] \left\{ 1 - p + \sum_{\ell=1}^{p} \frac{\varphi_{\ell}(-t_{\ell} + s_{\ell})}{\varphi_{\ell}(-t_{\ell}) \varphi_{\ell}(s_{\ell})} \right\}$$

is unknown (depends on the unknown marginals $\varphi_{\ell}(\cdot)$), then the weights $\lambda_k$'s, which are solutions (together with the $f_k$'s) of

$$\lambda f(s) = \int_{\mathbb{R}^q} f(t) C(s,t) w(t) dt$$

are also unknown and one cannot compute critical values for the test (quantiles of the distribution $\sum_{k=1}^{\infty} \frac{\lambda_k}{2} \left( (1 + |p_k|) \chi^2_1 + (1 - |p_k|) \chi^2_1 \right)$, using the CompQuadForm R package)

⇒ We need to (This will be the third improvement over the distance correlation test)

- estimate these $\lambda_k$'s;
- show that the use of these estimates $\hat{\lambda}_{n,k}$ has no (asymptotic) impact.
Theorem 3

For all $K \geq 1$, we have

$$\sum_{k=1}^{K} \frac{\lambda_k}{2} ((1 + |p_k|)\xi_k + (1 - |p_k|)\eta_k) - \sum_{k=1}^{K} \frac{\hat{\lambda}_k}{2} ((1 + |\hat{p}_k|)\xi_k + (1 - |\hat{p}_k|)\eta_k) \overset{P}{\rightarrow} 0,$$

where the $\hat{\lambda}_k$ (and $\hat{f}_k$) are the solutions of a discretized version of

$$\lambda f(x) = \int_{\mathbb{R}^q} f(y)C(x,y)w(y)dy$$

(1)

where the unknown $C(x,y)$ is replaced with

$$\hat{C}_n(x,y) = \prod_{\ell=1}^{p} \hat{\phi}_{n,\ell}(-x_{\ell} + y_{\ell}) - \prod_{\ell=1}^{p} [\hat{\phi}_{n,\ell}(-x_{\ell})\hat{\phi}_{n,\ell}(y_{\ell})] \left\{ 1 - p + \sum_{\ell=1}^{p} \frac{\hat{\phi}_{n,\ell}(-x_{\ell} + y_{\ell})}{\hat{\phi}_{n,\ell}(-x_{\ell})\hat{\phi}_{n,\ell}(y_{\ell})} \right\}. $$
Sketch of proof: One can write

\[ \sum_{j=1}^{N} \omega_j f(y_j) C(x, y_j) \xrightarrow{N \to \infty} \int_{\mathbb{R}^q} f(y) C(x, y) w(y) dy = \lambda f(x), \]

where the \( \omega_j \)'s are the weights of some cubature and where the \( y_j \)'s are nodes of the cubature. In the sequel, we suppose that \((\omega_j, y_j)\) can also be given as \((1/N, y_j)\) (i.e., a Monte-Carlo integration). So now we can write

\[ \sum_{j=1}^{N} \omega_j f(y_j) C(y_i, y_j) \approx \lambda f(y_i) \quad i = 1, \ldots, N. \]

This is the discretized version of the integral operator problem.
Sketch of proof (cont’d) : Let \( f_j = f(y_j), \ j = 1, \ldots, N \) and \( f = (f_1, \ldots, f_N)^T \). The system above can be written in matrix form as

\[
Cf = \lambda f \tag{2}
\]

where \( C \) is the \( N \times N \) (possibly random) matrix \( \left( C(y_i, y_j) \right)_{i,j} \).

Using results in [Rosasco et al., 2010], we can prove that the solutions (eigenvalues and eigenvectors) of (2) are good approximations (when \( N \) gets large) of those of (1) to estimate the \( p_k \).
Sketch of proof (cont’d) : But, since $C$ is unknown, we replace problem (2) with

$$\tilde{C}_n f = \lambda f$$

(3)

where $\tilde{C}_n$ is the $N \times N$ matrix $(\tilde{C}_n(y_i, y_j))_{i,j}$. Notice $n$ is not $N$.

Since $\tilde{C}_n$ is a strongly consistent estimator of $C$, we have that each element of $\tilde{C}_n$ converges (w.p.1, $n \to \infty$) to the corresponding element of $C$.

We can then use our Theorem 4 on next slide to prove that, for each $N$ fixed, the solutions of (3) converge (w.p.1, $n \to \infty$) to the solutions of (2).

Note : Even if $N$ is large, we need only to compute the first largest eigenvalues of the above Hermitian matrix (this is possible using the zhepvx Lapack Fortran subroutine).
**Theorem 4**

Let $N \in \mathbb{N}^*$ be a fixed integer. Let $A_n = (a_{ij,n})$ be a sequence of $(N \times N)$ stochastic complex matrices defined on some sample space $\Omega$, with (possibly complex) eigenvalues $\lambda_{(1),n}, \cdots, \lambda_{(N),n}$, counting multiplicity, ordered by the increasing value of their modulus. Let $A = (a_{ij})$ be a $(N \times N)$ positive definite stochastic complex matrix defined on $\Omega$ with (real positive) eigenvalues $0 < \lambda_{(1)} \leq \cdots \leq \lambda_{(N)}$, counting multiplicity. We note $f_{k,n}$ and $f_k$, $k = 1, \ldots, N$, the associated eigenvectors. If $A_n \overset{a.s.}{\longrightarrow} A$, i.e.

$$\forall i = 1, \ldots, N, \ \forall j = 1, \ldots, N, \ a_{ij,n} \overset{a.s.}{\longrightarrow} a_{ij}, \ \text{when } n \to \infty,$$

then

$$\lambda_{(k),n} \overset{a.s.}{\longrightarrow} \lambda_{(k)}, \ \text{and} \ \| (I_N - P_{A,k})f_{k,n} \|_2 \overset{a.s.}{\longrightarrow} 0, \ \forall k = 1, \ldots, N,$$

where $P_{A,k}$ is the orthogonal projector on the space spanned by all the orthonormal eigenvectors of $A$ associated with $\lambda_k$.

**Sketch of Proof** : Use characteristic polynomials and a general form of Rouché’s Theorem in complex analysis (e.g., [Rudin, 1987, p. 229]).
We implemented this test in the R package IndependenceTests.

Let us apply it on our data set.

```r
> require("IndependenceTests")
> X <- read.table("world_data.txt", header = TRUE); attach(X)
> X <- as.matrix(X); n <- nrow(X)
>
> head(X[,1:3])
  mortality natality gdp.capita
[1,] 14.89 25.47  3000
[2,] 14.46 10.72  8000
[3,] 14.44  8.92 18400
[4,] 14.33 33.38  1500
[5,] 14.31 10.00 24500
[6,] 14.28 36.60  2800
>
> set.seed(1); res <- mdcov(X, vecd = c(1, 1), a = 1, weight.choice = 1, N = 300, K = 200)
> res$mdcov
[1] 3.505156
> res$pvalue
[1] 5.66429e-10 # We reject the pairwise independence
>
> set.seed(1); res <- mdcov(X, vecd = c(1, 1, 1), a = 1, weight.choice = 1, N = 300, K = 200)
> res$mdcov
[1] 1.029831
> res$pvalue
[1] 0.00016071 # We reject the threewise independence
```
Note: our procedure generalizes the distance covariance.

```r
> res <- .C("Dcov1Cnormed", X, vecd = c(1L, 1L), a = 1.0, as.integer(n), p = 2L, weight.choice = 3L,
+ stat = 0.0, denom = 0.0, PACKAGE = "IndependenceTests")
> n * res$stat
[1] 302.3978
> sqrt(res$stat)
[1] 1.159306

> require("energy")
> set.seed(1) ; dcov.test(X[,1], X[,2], R = 100)

dCov test of independence

data: index 1, replicates 100
nV^2 = 302.4, p-value = 0.009901
sample estimates:
dCov
1.159306
```
A few perspectives

- Perform simulations to evaluate the power of this new test procedure against some competitors.

- Study optimality/efficiency, e.g. using results in [Nikitin, 1995].

- Build a test of multivariate serial dependence.
Thank you for your attention!


An extension of Mercer’s theory to $L^p$. *Positivity*, 16(2) :197–212.


