Measuring non-linear dependence for two random variables distributed along a curve

Pedro Delicado* **and Marcelo Smrekar**†

Abstract We propose new dependence measures for two real random variables not necessarily linearly related. Covariance and linear correlation are expressed in terms of principal components and are generalized for variables distributed along a curve. Properties of these measures are discussed. The new measures are estimated using principal curves and are computed for simulated and real data sets. Finally, we present several statistical applications for the new dependence measures.

Keywords Dependence measures, independence tests, linearity tests, principal curves, Rényi's axioms, similarity measures for pairs of variables.

1 Introduction

Correlation coefficient and covariance are appropriate dependence measures between two random variables when they are linearly related. However, it is usual to find situations where this condition is not fulfilled.

Several works have introduced non-linear dependence measures between two random variables. Rényi (1959) enunciated seven properties which should be verified by any dependence measure between two random variables defined over the same probability space. These axioms were discussed and partially modified by Bell (1962), Scheizer and Wolff (1981) and Nelsen (2006), among others.

Nonparametric functional estimation methods have given rise to several definitions of dependence measures. Bjerve and Doksum (1993) defined a measure on the plane called a *correlation curve* (observe that this is not a scalar measure). It is based on locally estimated coefficients of nonparametric regression and measures local linear correlation conditioned on one variable. Therefore,

*Corresponding author. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. pedro.delicado@upc.edu

†Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.

the correlation curve is not symmetric and one of Rényi's axioms is violated. Bjerve and Doksum (1993) proposed a symmetric version. See also Doksum et al. (1994) for extension to several dimensions and Doksum and Froda (2000) for the smoothing parameter choice. Jones (1996) criticized correlation curves and defined local dependence measure based on nonparametric bivariate density estimation (see also Holland and Wang, 1987, and Wang, 1993). This local dependence function satisfies some of Rényi's axioms, but has the disadvantage of being a whole function over $\mathbb{R}^2$.

In this paper we propose two new dependence measures for two real random variables that are non-linearly related. First, we express correlation and covariance in terms of the first principal component (Section 2). Then, in Section 3 they are generalized for random variables in $\mathbb{R}^2$ distributed along a curve. In Section 4 we discuss the properties of these measures regarding Rényi's axioms. Section 5 shows how the new measures can be estimated using principal curves (Hastie and Stuetzle, 1989; Kégl et al., 2000; Delicado, 2001) and how they are applied to simulated data sets. The new coefficients are used in Section 6 to measure nonlinear dependences in a real data set concerning neighborhood education and age distribution in Barcelona. In Section 7 several statistical applications for the new dependence measures are presented and illustrated with this real data set. Finally, conclusions are provided in Section 8.

2 Linear relation measures in terms of principal components

Let (X, Y) be a two-dimensional random variable with variance matrix Σ. Let λ_1 and λ_2 ($\lambda_1 \geq \lambda_2$) be the eigenvalues of Σ, and let α be the angle between the eigenvector associated to λ_1 (the first principal component) and axis x. Remember that λ_i is the variance of the i-th principal component, for $i = 1, 2$.

Without loss of generality we can assume that the first eigenvector of Σ belongs to the first quadrant. Using the spectral decomposition of matrix Σ,

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}^T \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

Then $\mathrm{V}(X) = \sigma_X^2 = \lambda_1 \cos^2\alpha + \lambda_2 \sin^2\alpha$, $\mathrm{V}(Y) = \sigma_Y^2 = \lambda_1 \sin^2\alpha + \lambda_2 \cos^2\alpha$,

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = (\lambda_1 - \lambda_2) \cos\alpha \sin\alpha. \tag{1}$$

We can also express the correlation coefficient as a function of λ_1, λ_2 and α:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{(\lambda_1 - \lambda_2)\cos\alpha \sin\alpha}{(\lambda_1 \cos^2\alpha + \lambda_2 \sin^2\alpha)^{\frac{1}{2}} (\lambda_1 \sin^2\alpha + \lambda_2 \cos^2\alpha)^{\frac{1}{2}}}. \qquad (2)$$

Note that the expressions of covariance and correlation are symmetric in α and $(\pi/2 - \alpha)$, which is equivalent to saying that $\sigma_{XY} = \sigma_{YX}$ and $\rho_{XY} = \rho_{YX}$.

3 Non-linear dependence measures on $\mathbb{R}^2$

In this section we define dependency measures for two random variables jointly distributed in $\mathbb{R}^2$ along a one-dimensional curve. We use expressions (1) and (2) to define measures of local linear relationship; then by aggregating them we obtain global measures of dependence.

Let $c : I \subseteq \mathbb{R} \mapsto \mathbb{R}^2$ be a one-dimensional smooth curve in the plane. We assume that c is parameterized by the length of arch, or equivalently that c is unit speed: $\|c'(s)\| = 1$ for all $s \in I$. Let $v(s), s \in I$, be a unitary vector field orthogonal to c (that is, $c'(s)^T v(s) = 0$ for all $s \in I$). We define $\chi_c : I \times \mathbb{R} \mapsto \mathbb{R}^2$ by $\chi_c(s,t) = c(s) + tv(s)$. Assume that; (S,T) is jointly distributed in $A \subseteq I \times \mathbb{R}$ with density $h(S,T)$; that $E(T|S=s) = 0$ and $\mathrm{V}(S) > \mathrm{V}(T|S = s)$, and that $\chi_c : A \mapsto \chi_c(A)$ is a one-to-one application. Function χ_c is a particular case of the function defined by Hastie and Stuetzle (1989) in the proof of their *Proposition 6*, which has also been used in Delicado (2001), Delicado and Huerta (2003) and Delicado and Smrekar (2007). This latter paper includes the derivation of the expression of the density of (X,Y) in terms of the curve c and the density of (S,T). Necessary conditions for χ_c being one-to-one can be found in Hastie and Stuetzle (1989).

Definition 1 *Let (X,Y) be the bivariate random variable obtained as*

$$(X,Y) = \chi_c(S,T) : A \mapsto \mathbb{R}^2.$$

We say that (X,Y) is distributed along the curve $c(I)$, that $c(I)$ is the generating curve, and that (S,T) is the generating bivariate random variable.

The random variable (X,Y) can be described as generated by a random point on the curve $c(I)$ plus an orthogonal random noise. Then the curve $c(I)$ summarizes the structure of the (X,Y) distribution. For the particular case of $c(I)$ being a straight line, the statistical dependence between X and Y can be said to be linear, and it is well measured using covariance and correlation.

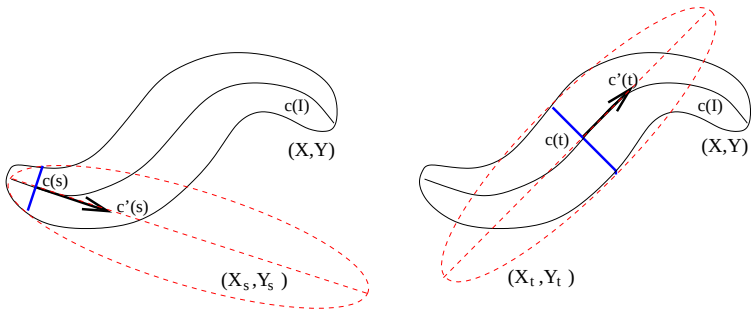

Figure 1: *Linearizing the distribution of* (X, Y) *around* $c(s)$ *and* $c(t)$.

In order to generalize linear dependence measures, we start by defining local measures (of variance and covariance) around a point $c(s)$ in $\mathbb{R}^2$. The underlying idea is to *linearize* the distribution of (X, Y) around $c(s)$; that is, we look for a random variable (X_s, Y_s) distributed along a straight line in such a way that the distributions of (X_s, Y_s) and (X, Y) are similar around $c(s)$. Figure 1 illustrates this linearization process for two points $c(s)$ and $c(t)$. Local measures are defined as follows.

Definition 2 *Let* $(X, Y) = \chi_c(S, T)$ *be a bivariate random variable distributed along the curve* $c(I)$. *For* $s \in I$, *let* $\alpha(s)$ *be the angle between* $c'(s)$ *and the abscissas axis. We define local variances of* X *and* Y *at* $c(s)$ *as*

$$\mathrm{LV}_X(s) = \mathrm{V}(S)\cos^2\alpha(s) + \mathrm{V}(T|S=s)\sin^2\alpha(s),$$

$$\mathrm{LV}_Y(s) = \mathrm{V}(S)\sin^2\alpha(s) + \mathrm{V}(T|S=s)\cos^2\alpha(s).$$

Local covariance at $c(s)$ *is defined as*

$$\mathrm{LCov}_{(X,Y)}(s) = \{\mathrm{V}(S) - \mathrm{V}(T|S=s)\}\cos\alpha(s)\sin\alpha(s),$$

and local correlation at $c(s)$ *as*

$$\mathrm{LCor}_{(X,Y)}(s) = \mathrm{LCov}_{(X,Y)}(s)/\{\mathrm{LV}_X(s)\mathrm{LV}_Y(s)\}^{1/2}.$$

Once local dependence measures have been defined, we aggregate them to obtain two global measures. It is important to notice that the local covariance and correlation have the sign of the curve slope $c'(s)$. So local measures can have different signs in different points of curve $c(I)$, and they may cancel out when they are aggregated. Therefore it is convenient to aggregate squared values of local measures and then take the square root. We propose to do aggregation by taking expectations with respect to the distribution of the random variable S.

Definition 3 *In the context of Definition 2, the Covariance of X and Y along their generating curve c is defined as*

$$\mathrm{CovGC}(X,Y) = \left(E_S[\{\mathrm{LCov}_{(XY)}(S)\}^2]\right)^{1/2},$$

and their Correlation along the curve c is

$$\mathrm{CorGC}(X,Y) = \left(E_S[\{\mathrm{LCor}_{(XY)}(S)\}^2]\right)^{1/2}.$$

These definitions are a generalization of the absolute value of covariance and correlation, respectively, as is established in the next Proposition.

Proposition 1 *Let $(X,Y) = \chi_c(S,T)$ be a bivariate random variable distributed along the curve $c(I)$. Assume that the curve $c(I)$ is in fact a straight line ($c'(s) = \beta$ for all $s \in I$, $\beta \in \mathbb{R}^2$, $\|\beta\| = 1$) and that $\mathrm{V}(T|S = s)$ does not depend on s. Then, for all $s \in I$, $\mathrm{LV}_X(s) = \mathrm{V}(X)$, $\mathrm{LV}_Y(s) = \mathrm{V}(Y)$,*

$$\mathrm{LCov}_{(X,Y)}(s) = \mathrm{Cov}(X,Y),\ \mathrm{LCor}_{(X,Y)}(s) = \mathrm{Cor}(X,Y),$$

$$\mathrm{CovGC}(X,Y) = |\mathrm{Cov}(X,Y)| \text{ and } \mathrm{CorGC}(X,Y) = |\mathrm{Cor}(X,Y)|.$$

Proof. The proof of this result is straightforward. The only point requiring some care is to show that the straight line $c(I)$ is in fact the first principal component of (X,Y). For its proof, let $Z = a_1X + a_2Y$, with $a_1^2 + a_2^2 = 1$, a normalized linear combination of X and Y. By Definition 1, there exist b_1 and b_2, with $b_1^2 + b_2^2 = 1$, such that $Z = a_1X + a_2Y = b_1S + b_2T$. Assuming that $\mathrm{V}(T) < \mathrm{V}(S)$, we have

$$\mathrm{V}(Z) = b_1^2\mathrm{V}(S) + b_2^2\mathrm{V}(T) \leq \mathrm{V}(S)$$

with equality if and only if $b_1 = 1$ and $b_2 = 0$. We conclude that the first principal component is the generating straight line $c(I)$. □

Three examples illustrate the computation of non-linear dependence measures.
Example 1: Data on a ring. Let (S,T) be a uniform random vector on $A = I \times J$ with $I = [-\pi, \pi)$ and $J = [-\varepsilon, \varepsilon]$, $\varepsilon \in (0,1)$. Let $c(s) = \{\cos(s), \sin(s)\}^T$ be the usual parametrization of the unit circumference S^1 in $\mathbb{R}^2$, and $\chi : A \mapsto \mathbb{R}^2$ the corresponding one-to-one function. Then $\chi(A) = \{x : x \in \mathbb{R}^2, 1 - \varepsilon \leq \|x\| \leq 1 + \varepsilon\}$. We define $(X,Y) = \chi(S,T)$. (X,Y) is generated as a uniform random point in the S^1 circumference plus an orthogonal uniform random noise (see Figure 2, graph bottom left). Observe that $c'(s) = \{-\sin(s), \cos(s)\} = \{\cos(s + \pi/2), \sin(s + \pi/2)\}$ and it follows that $\alpha(s) = s + \pi/2$. The square of the CovGC is

$$\mathrm{CovGC}^2(X,Y) = E_S\left([\{E_S(S^2) - E_T(T^2|S)\}\cos\alpha(S)\sin\alpha(S)]^2\right)$$

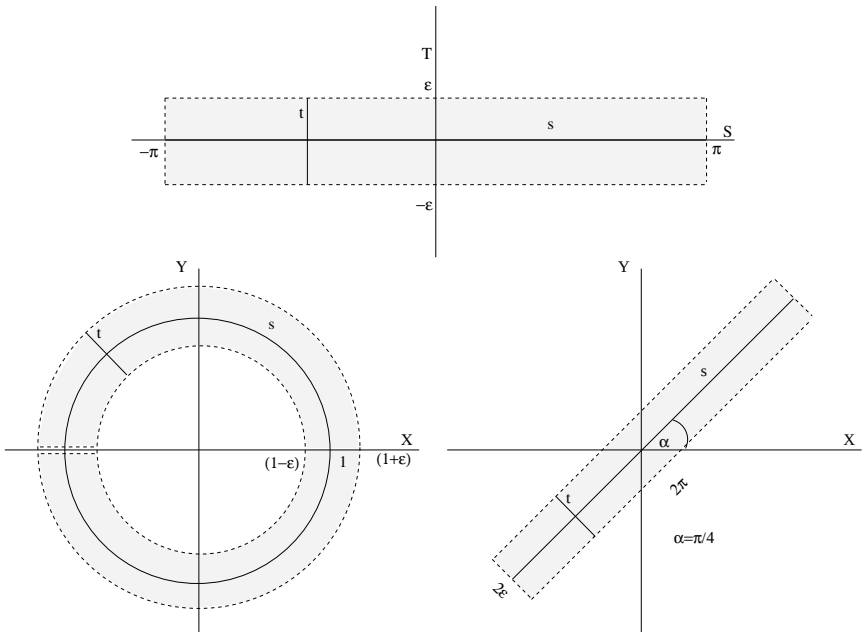

Figure 2: *Generating variables* (S,T) *(top graphic) and two examples of random variables* (X,Y) *distributed along a curve. In the bottom left graphic the generating curve is the circumference* S^1*, and in the bottom right graphic the generating curve is a straight line.*

$$= \left(\frac{\pi^2}{3} - \frac{\varepsilon^2}{3}\right)^2 \int_{-\pi}^{\pi} \left\{\cos\left(s+\frac{\pi}{2}\right)\sin\left(s+\frac{\pi}{2}\right)\right\}^2 \frac{1}{2\pi} ds = \frac{1}{8}\left(\frac{\pi^2}{3} - \frac{\varepsilon^2}{3}\right)^2.$$

Then,

$$\mathrm{CovGC}(X,Y) = \frac{1}{2\sqrt{6}}(\pi^2 - \varepsilon^2).$$

The square of CorGC is

$$\mathrm{CorGC}^2(X,Y) = \int_0^{2\pi} \frac{\left(\frac{\pi^2}{3} - \frac{\varepsilon^2}{3}\right)^2 \cos^2\alpha(s)\sin^2\alpha(s)}{\{\frac{\pi^2}{3}\cos^2\alpha(s) + \frac{\varepsilon^2}{3}\sin^2\alpha(s)\}\{\frac{\pi^2}{3}\sin^2\alpha(s) + \frac{\varepsilon^2}{3}\cos^2\alpha(s)\}} \frac{1}{2\pi} ds$$

$$= \int_0^{2\pi} \frac{(\pi^2-\varepsilon^2)^2 \sin^2 s \cos^2 s}{(\pi^2\sin^2 s + \varepsilon^2\cos^2 s)(\pi^2\cos^2 s + \varepsilon^2\sin^2 s)} \frac{1}{2\pi} ds = \frac{(\pi-\varepsilon)^2}{\pi^2+\varepsilon^2}.$$

Then,

$$\mathrm{CorGC}(X,Y) = \frac{\pi-\varepsilon}{\sqrt{\pi^2+\varepsilon^2}}.$$

Example 2: Data in a rectangle. We define two elements of $\mathbb{R}^2$: $a = (\pi/\sqrt{2}, \pi/\sqrt{2})$ and $b = (-\varepsilon/\sqrt{2}, \varepsilon/\sqrt{2})$. Let B be the rectangle delimited by the points $(a+b)$, $(a-b)$, $(-a+b)$ and $(-a-b)$, according to Figure 2, graph

bottom right. Let (S,T) and A be as in Example 1, but now with $\varepsilon \in (0,\pi)$. Let χ be a rigid rotation transformation (with angle $\alpha = \pi/4$) such that $\chi(A) = B$, and let $(X,Y) = \chi(S,T)$. Therefore (X,Y) is a uniform random variable in the rectangle B. Moreover (X,Y) is distributed along the line $x = y$. The corresponding value $\alpha(s) = \alpha = \pi/4$ for all s. Let us compute the squares of CovGC and CorGC:

$$\begin{aligned}
\mathrm{CovGC}(X,Y)^2 &= E_S[\{E_S(S^2) - E_T(T^2|S)\}\cos^2\alpha\sin^2\alpha] \\
&= \left(\frac{\pi^2}{3} - \frac{\varepsilon^2}{3}\right)^2 \int_{-\pi}^{\pi} \left(\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\right)^2 \frac{1}{2\pi} ds = \frac{1}{4}\left(\frac{\pi^2}{3} - \frac{\varepsilon^2}{3}\right)^2, \\
\mathrm{CorGC}(X,Y)^2 &= \int_0^{2\pi} \frac{\left(\frac{\pi^2}{3} - \frac{\varepsilon^2}{3}\right)^2 \cos^2\alpha\sin^2\alpha}{(\frac{\pi^2}{3}\cos^2\alpha + \frac{\varepsilon^2}{3}\sin^2\alpha)(\frac{\pi^2}{3}\sin^2\alpha + \frac{\varepsilon^2}{3}\cos^2\alpha)} \frac{1}{2\pi} ds \\
&= \frac{(\pi^2-\varepsilon^2)^2(1/2)(1/2)}{\{\pi^2(1/2)+\varepsilon^2(1/2)\}\{\pi^2(1/2)+\varepsilon^2(1/2)\}} = \frac{\left(\pi^2-\varepsilon^2\right)^2}{\left(\pi^2+\varepsilon^2\right)^2}.
\end{aligned}$$

Therefore

$$\mathrm{CovGC}(X,Y) = \frac{1}{2\sqrt{3}}(\pi^2 - \varepsilon^2),\ \mathrm{CorGC}(X,Y) = \frac{\pi^2-\varepsilon^2}{\pi^2+\varepsilon^2}.$$

The covariance along the generating curve in the rectangle (coinciding with its covariance) is $\sqrt{2}$ times that in the ring (Example 1). In both cases the distribution over the generating curve is the same, but local angles between the curves and the abscissas axis are different. The correlation along the generating curve in the ring is lower than that in the rectangle for all $\varepsilon \in (0,1)$.

Example 3: It is worthwhile to note the CorGC value is not always greater than the usual correlation coefficient, as the following example makes clear. Consider the following points in $\mathbb{R}^2$: $O = (0,0)^\tau, P = (0,-1)^\tau, Q = (1,0)^\tau$. Let (X,Y) be a uniform random variable in the set $B = \overline{PO} \cup \overline{OQ}$. Then $\mathrm{Cor}(X,Y) > 0$, but $\mathrm{CorGC}(X,Y) = 0$, because the angle between the generating curve (which is just the parameterization of the set B) and the abscissas axis is 0 or $\pi/2$.

4 Rényi's axioms

Rényi (1959) gives a list of seven desirable properties (known as Rényi's axioms) which should be satisfied by any measure, $\delta(\cdot,\cdot)$, of dependence between two random variables, X and Y, defined on the same probability space. Rényi's axioms are the following:

A. $\delta(X,Y)$ is defined for any pair of random variables X and Y, neither of them being constant with probability 1.

B. $\delta(X,Y) = \delta(Y,X)$.

C. $0 \leq \delta(X,Y) \leq 1$.

D. $\delta(X,Y) = 0$ if and only if X and Y are independent.

E. $\delta(X,Y) = 1$ if there is a strict dependence between X and Y, i.e. either $X = g(Y)$ or $Y = f(X)$, where $g(\cdot)$ and $f(\cdot)$ are Borel-measurable functions.

F. If the Borel-measurable functions $f(\cdot)$ and $g(\cdot)$ map the real line in a one-to-one way into itself, then $\delta\{f(X), g(Y)\} = \delta(X,Y)$.

G. If the joint distribution of X and Y is normal, then $\delta(X,Y) = |R(X,Y)|$, where $R(X,Y)$ is the correlation coefficient of X and Y.

This set of axioms has been considered too restrictive by various authors (see, e.g., Scheizer and Wolff, 1981, and Nelsen, 2006) who have proposed slight modifications. Scheizer and Wolff (1981), for instance, restrict their attention to pairs of continuously distributed random variables, modify axioms E (replacing "if" by "if and only if" and limiting f and g to be a.s. strictly monotone functions), F (see F' below) and G (they allow $\delta(X,Y)$ to be a strictly increasing function of the absolute value of $R(X,Y)$), and add a continuity axiom H.

Here we adapt Rényi's axioms A, E and F as follows (F' is borrowed from Scheizer and Wolff, 1981):

A'. $\delta(X,Y)$ is defined for any pair of random variables (X,Y) distributed along a curve according to Definition 1.

E'. Let (X,Y) be two random variables distributed along a curve c according to Definition 1, with generating variables (S,T). $\delta(X,Y) = 1$ if and only if there is a strict dependence between (X,Y) and S, that is, $X = c_1(S)$ and $Y = c_2(S)$, or equivalently T is identically 0.

F'. If $f(\cdot)$ and $g(\cdot)$ are strictly monotone almost surely (a.s.) on Range X and Range Y, respectively, then $\delta\{f(X), g(Y)\} = \delta(X,Y)$.

Observe that axioms A' and E' are well suited to random variables distributed along a curve with no noise.

The following Theorem checks which axioms are verified by CorGC defined in Section 3. Observe that the same axioms (except C, E' and G) are verified by CovGC as well. See also the remarks following the proof.

Theorem 1 CorGC *verifies axioms A', B, C, E' and G.* CorGC *also satisfies the following properties:*

D'$_1$. *If X and Y are independent, then* $\mathrm{CorGC}(X,Y)=0$.

D'$_2$. *Let (X,Y) be distributed along a curve according to Definition 1. If* $\mathrm{CorGC}(X,Y)=0$ *and S and T are independent, then X and Y are independent.*

CorGC *does not verify axiom F'.*

Proof. The definition of CorGC implies that it verifies A'. CorGC satisfies condition B because of the symmetrical character of the function $cos(\alpha)sin(\alpha)$ in α and $(\pi-\alpha)$.

Let us prove that CorGC verifies C. The $\mathrm{CorGC}(X,Y)\geq 0$ because CorGC is the expectation of a positive function of a random variable. We now see that $\mathrm{CorGC}(X,Y)\leq 1$. If $\cos\alpha(S)=0$ a.s. or $\sin\alpha(S)=0$ a.s. then $\mathrm{CorGC}(X,Y)=0\leq 1$. In other cases,

$$\mathrm{CorGC}(X,Y)^2=\int \frac{[\{\mathrm{V}(S)-\mathrm{V}(T|S=s)\}\cos\alpha(s)\sin\alpha(s)]^2}{\mathrm{LV}_X(s)\ \mathrm{LV}_Y(s)}f_S(s)ds$$

$$=\int \frac{\{\mathrm{V}(S)-\mathrm{V}(T|S=s)\}^2 f_S(s)}{\mathrm{V}(S)^2+\mathrm{V}(T|S=s)^2+\mathrm{V}(S)\mathrm{V}(T|S=s)\{\tan^2\alpha(s)+\tan^{-2}\alpha(s)\}}ds$$

$$\leq \int \frac{\{\mathrm{V}(S)-\mathrm{V}(T|S=s)\}^2}{\mathrm{V}(S)^2+\mathrm{V}(T|S=s)^2+2\mathrm{V}(S)\mathrm{V}(T|S=s)}f_S(s)ds$$

$$=\int \frac{\{\mathrm{V}(S)-\mathrm{V}(T|S=s)\}^2}{\{\mathrm{V}(S)+\mathrm{V}(T|S=s)\}^2}f_S(s)ds\leq \int f_S(s)ds=1,$$

then $0\leq \mathrm{CorGC}(X,Y)\leq 1$. We have used that the function $g(x)=x^2+(1/x)^2$ has its minimum in $\mathbb{R}^+$ at $x=1$.

In the previous derivation we see that $\mathrm{CorGC}(X,Y)=1$ if and only if $\mathrm{V}(T|S=s)=0$. This implies that axiom E' is verified by CorGC. It follows from Proposition 1 that in the bivariate normal case CovGC and CorGC coincide with the absolute value of covariance and linear correlation, respectively. So CorGC verifies G.

Now we prove D'$_1$ and D'$_2$. Let X and Y be independent. Assume that $\mathrm{V}(X)\geq \mathrm{V}(Y)$. Then the marginal variance on the curve $c(s)=\{s,E(Y)\}$ is greater than the marginal variance on the curve $d(t)=\{E(X),t\}$, and the (X,Y) distribution is along $c(s)=\{s,E(Y)\}$, with χ_c the identity function in $\mathbb{R}^2$. Therefore $\alpha(s)$ is constantly 0 and CorGC and CovGC are also null. This proves D'$_1$. According to Definition 1, $\mathrm{V}(S)>\mathrm{V}(T|S=s)$ for all s. So,

if CorGC and CovGC are zero then $\alpha(s) = 0$ for all s and $X = S$, $Y = T$. Therefore X and Y are independent and D'$_2$ is verified.

Property F' does not hold for CorGC. See Remark 2 below and Example 5 given in Section 5. □

REMARK 1. The measure CorGC almost verifies Rényi's axiom D, which is slightly stronger than D'$_1$ plus D'$_2$.

REMARK 2. Let us examine more thoroughly why CorGC does not verify axiom F', which means invariance against strictly monotone transformations of variables X and Y. First at all, given (X, Y) distributed along a curve (according to Definition 1), and $f(\cdot)$ and $g(\cdot)$ strictly monotone functions on Range X and Range Y, respectively, in general it is not guaranteed that $\{f(X), g(Y)\}$ are distributed along any curve. So CorGC$\{f(X), g(Y)\}$ may not exist. On the other hand, even if CorGC$\{f(X), g(Y)\}$ is well defined, we must not expect that CorGC$\{f(X), g(Y)\}$ = CorGC(X, Y), because in the definition of CorGC orthogonalities play a central role and in general are not preserved when transforming (Range $X \times$ Range Y) by $\{f(\cdot), g(\cdot)\}$.

From Scheizer and Wolff (1981) and Nelsen (2006) it follows that axiom F' obliges us to measure the dependence between X and Y from their copula C_{XY}, a distribution function on $[0, 1]\times[0, 1]$ verifying $F_{XY}(x, y) = C_{XY}\{F_X(x), F_Y(y)\}$, for all reals x, y (see Theorem 1 in Scheizer and Wolff, 1981) where F_X, F_Y and F_{XY} are the distribution functions of X, Y and (X, Y), respectively. Any dependence measure between two absolute continuous random variables X and Y satisfying axiom F' must depend only on C_{XY}: $\delta(X, Y) = \delta\{F_X(X), F_Y(Y)\} = \delta(U, V)$, with $U = F_X(X)$ and $V = F_Y(Y)$ uniforms on $[0, 1]$ and (U, V) having the same copula as (X, Y). This result has its counterpart when measuring dependence from a bivariate random sample: the sampling analogue of axiom F' implies that any sampling dependence measure may depend only on the ranks of the observations. Kendal's τ and rank correlation Spearman's ρ are examples of such sampling dependence measures.

In order to define a dependence measure between X and Y being related with CorGC and verifying axiom F', we should assume that $(U, V) = \{F_X(X), F_Y(Y)\}$ is distributed along a curve, and then use the measure given in Definition 3 on (U, V). Nevertheless we consider that it is not natural to impose that (U, V) (a random variable on $[0, 1] \times [0, 1]$ with uniform marginals) is distributed along a curve. From a practical point of view, in Section 5 we present a sampling version of CorGC, and the possibility exists of applying it to the ranks of a sample.

5 Estimating the new dependence measures

In previous sections we introduced the dependence measures CovGC and CorGC for bivariate random variables distributed along a curve. Now we deal with definitions of analogous concepts for random samples drawn from such random variables, that is, definitions of estimators of the population concepts. Following Definitions 2 and 3, we need to estimate several elements before computing estimations of CovGC and CorGC: the generating curve $c(I)$, the interval $I \subseteq \mathbb{R}$, where it is defined, the speed vectors $c'(s)$ (and the corresponding angles $\alpha(s)$), the variance of the generating variable S, the variance $\mathrm{V}(T|S = s)$ orthogonal to the curve at each point $c(s)$, and the distribution of the generating variable S.

The natural way to estimate I, $c(s)$ and $c'(s)$, $s \in I$, is by means of a principal curve fitting algorithm. For more information on principal curves see, for instance, Hastie and Stuetzle (1989), Kégl et al. (2000) or Delicado (2001), where three different concepts of principal curves are introduced. The algorithm proposed by Hastie and Stuetzle (1989) is implemented in the packages `princurve` and `pcurve` of R (R Development Core Team, 2005). A Java implementation for Kégl et al. (2000) is available on the web page of one of the authors (`http://www.iro.umontreal.ca/%7Ekegl/research/pcurves/`). The algorithm proposed by Delicado (2001) has been implemented in C++ (see Delicado and Huerta, 2003). It is available, with interfaces in both MATLAB and R, at `http://www-eio.upc.es/%7Edelicado/PCOP/`.

The afore-mentioned concepts of principal curves (and others that can be found in the literature) share an undesirable property: if a random variable (X, Y) is distributed along a curve $c(I)$ (as defined in Definition 1), then the curve $c(I)$ is not a principal curve according to any of those definitions (Delicado, 2001). Nevertheless, given a data set, the differences between the three estimated principal curves are usually small, and they are close to the generating curve. Therefore as the first step to computing CovGC and CorGC, we propose using any of these algorithms to fit a principal curve to a bivariate data set.

Once the curve has been estimated, the nonparametric estimation of the other required elements in the definition of CovGC and CorGC is straightforward. In this paper we use the afore-mentioned implementations of the proposals of Hastie and Stuetzle (1989) and Delicado (2001). We denote them by HS and PCOP, respectively. Routines in R to compute CovGC and CorGC are available at `http://www-eio.upc.es/%7Edelicado/PCOP/`.

Table 1: Numeric summary of the simulation results. Mean and standard deviation (in brackets) for 500 simulations.

Example	Population CorGC	Correl. Coef.	Estimated CorGC, HS	Estimated CorGC, PCOP
1, $\varepsilon = .2$	.9344	–	.9347 (.0087)	.9037 (.0753)
1, $\varepsilon = .4$	.8657	–	.8740 (.0144)	.8513 (.0357)
2, $\varepsilon = 1.32$	.7	.7001 (.0222)	.7055 (.04094)	.7035 (.0224)
4, $\rho = .7$	.7	.7011 (.0355)	.7102 (.0354)	.7014 (.0348)
4, $\rho = .85$	.85	.8506 (.0196)	.8558 (.0188)	.8508 (.0201)
5, $\rho = .85$	–	.8190 (.0221)	.8661 (.0199)	.8511 (.0265)

As an illustration of the use of the new sampling dependence measures, we apply them to a battery of simulated data sets. We focus on the estimation of CorGC.

Example 1: Data on a ring (continued). We generate 500 samples of 200 bivariate data following the distribution described in Example 1 (Section 3). We fix $\varepsilon = 0.2$ and $\varepsilon = 0.4$, which gives values of CorGC = 0.9344 and CorGC = 0.8657, respectively. A numeric summary of the simulation results appears in Table 1. For $\varepsilon = 0.4$, Figure 3 (left graph) shows the density function estimated from the 500 sampling CorGC values when both principal curve methods, HS and PCOP, are used. Both estimators of CorGC are biased (in different directions), but the MSE of the HS based estimator is lower than that of the PCOP based estimator. This is because the HS principal curve estimator fits closed generating curves better than the PCOP method.

Example 2: Data in a rectangle (continued). Now we use the uniform distribution on a rectangle described in Example 2 (Section 3). We generate 500 samples of 200 bivariate data from it. We fix $\varepsilon = 1.32$ in order to obtain a value of CorGC = 0.7, which in this case coincides with the population correlation coefficient. Figure 3 (right grap) shows the density function estimated from the 500 sampling CorGC values when usual correlation coefficient and both principal curve methods, HS and PCOP, are used. Table 1 provides a numeric summary of the results. The CorGC estimator based on PCOP and the absolute value of the sampling correlation coefficient are comparable in this case. The CorGC estimator based on HS is also approximately unbiased, but presents much more variability. Results for greater values of ε (not presented here) indicate that the PCOP based estimator is more suitable for this kind of data. The reason is that the HS principal curve estimator does not fit well data having a non-closed

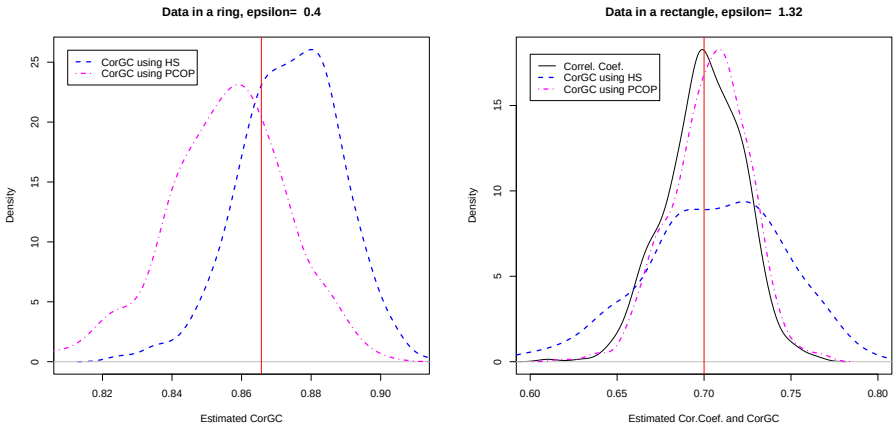

Figure 3: Nonparametric density estimation for 500 estimations of CorGC. Left panel: Data on a ring (Example 1). Right panel: Data in a rectangle (Example 2).

generating curve with compact support.

Example 4: Bivariate normal data. We now consider data coming from a bivariate normal distribution. Proposition 1 states that in this case the CovGC and the CorGC must agree with the absolute values of covariance and correlation coefficient, respectively. The results of the new estimators and the absolute values of the usual measures were compared in 500 samples of 200 bivariate normal data, with mean $\mu = (0, 0)^T$, unit variances and covariance equal to ρ. Therefore the population value of CovGC, CorGC, Cov and Cor are all equal to ρ. We show results for $\rho = 0.7$ and $\rho = 0.85$. A numeric summary of the results can be seen in Table 1. The nonparametric density estimations for the correlation coefficient (in absolute value) and the CorGC are represented in Figure 4, left panel, for $\rho = .85$. It can be seen that both CorGC estimators are comparable to the correlation coefficient as estimators of ρ. In fact in our experiment the MSE of the PCOP based CorGC estimator is slightly lower than that of the correlation coefficient when $\rho = 0.7$.

Example 5: Nonlinear transformation of bivariate normal data. The simulations described in Example 4 (for $\rho = 0.85$) were also used to compare the CorGC of the 500 samples with the CorGC of 500 samples nonlinearly transformed. We transform (x, y) to $\{f(x), g(y)\}$, with $f(x) = x$ and $g(y) = \text{sign}(y)\sqrt{|y|}$. Observe that this transformation is one of those considered in axiom F' (Section 4). Therefore if CorGC verified axiom F', we would find

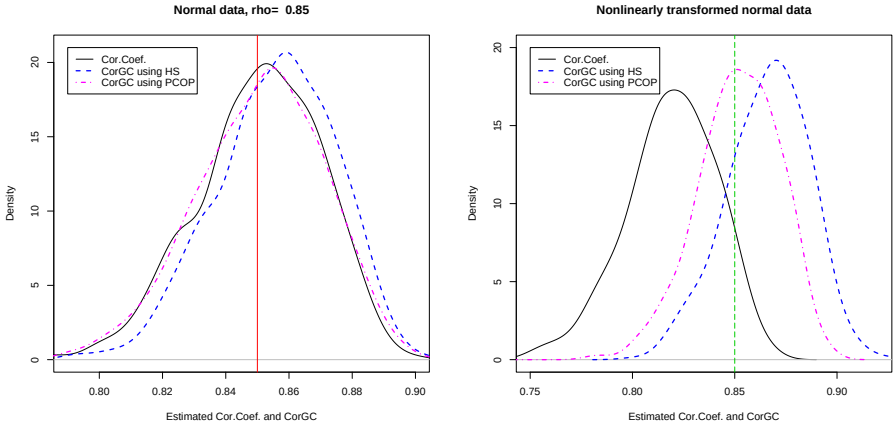

Figure 4: Nonparametric density estimation for 500 estimation of CorGC. Left panel: Bivariate normal data, $\rho = 0.85$ (Example 4). Right panel: Nonlinear transformation of bivariate normal data, $\rho = 0.85$ (Example 5).

that $\rho = \text{Cor}(X, Y) = \text{CorGC}(X, Y) = \text{CorGC}\{f(X), g(Y)\}$. One may observe that this is not true in this example. The results are summarized in Table 1, in Figure 4 (right panel) and in Figure 5. It is apparent that the HS based CorGC estimator is not in accord with to axiom F' (the same occurs for the correlation coefficient). This is not so clear for the PCOP based CorGC estimator. Nevertheless, both null hypotheses $\beta_0 = 0$, $\beta_1 = 1$ are rejected when a simple regression is fitted to the data drawn in the left panel of Figure 5. We conclude that axiom F' is not fulfilled by CorGC.

6 A real data analysis

In order to check the proposed dependence measures with a real data set, we consider data from the city of Barcelona regarding educational levels and age structure. Seven variables (listed in Table 2) are measured in 246 Zones of Study (ZRP, from the initials in Catalan), which are groupings of neighborhoods defined for statistical purposes. This city division coexists with other city divisions with lower and upper levels of aggregation. The data are obtained from the Department of Statistics of the Barcelona Municipal Council web page (`http://www.bcn.cat/estadistica/angles/index.htm`). We aggregated educational level categories into three levels and age distribution categories into four levels. Two of the original 248 ZRPs were clearly outliers in most of two-

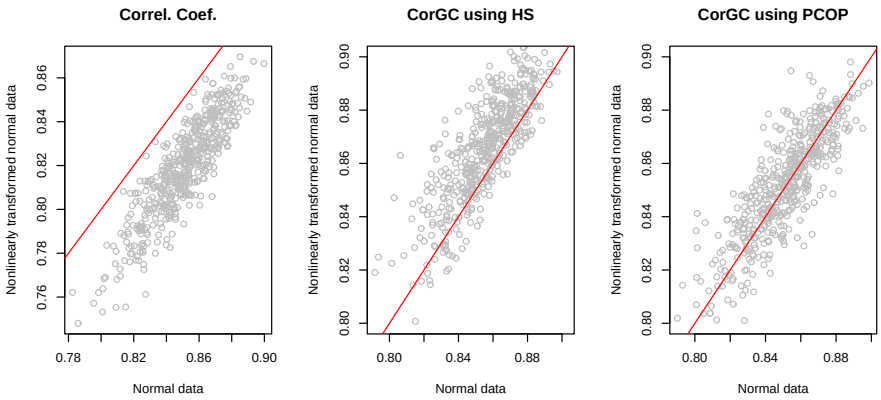

Figure 5: Joint distribution of correlation coefficient, HS based CorGC and PCOP based CorGC for 500 samples of a bivariate normal data, $\rho = 0.85$, and the same data nonlinearly transformed (Example 5).

dimensional marginal distributions. They are then removed, since otherwise they would distort the principal curve estimation. The seven variables are measuring proportions and their variability are comparable (with the exception of the slightly more disperse variables Primary and University). For this reason we decide not to apply any transformation to the original data. In cases where the dispersion varies greatly from variable to variable it would be advisable to transform the data before computing CorGC or CovGC (standardizing them or computing the sample ranks, for instance).

For an exploratory data analysis, the matrix of scatter-plots for the seven variables are shown in Figure 6. Principal curves are fitted to each pair of

Table 2: Variables observed in 246 neighborhoods (Zones of Study, ZRP) of the city of Barcelona.

Variable name	*Description:* Proportion of people in the ZRP with ...
Primary	... primary studies.
Secondary	... secondary studies.
University	... a university degree.
Age Group 1	... age under 14 years.
Age Group 2	... age between 15 and 24 years.
Age Group 3	... age between 25 and 64 years.
Age Group 2	... age over 65 years.

variables, using both HS (solid line) and PCOP (dashed line) methods. These graphs reveal the nonlinear nature of the dependence between most variables. It can be seen that the two principal curve algorithms generally agree. For each pair of variables, the CorGC is calculated using HS and PCOP algorithms. The absolute value of correlation coefficients and the two estimated CorGC values are given in Table 3 (upper diagonal entries).

There are many pairs of variables with a clear nonlinear joint distribution; some of them are also highly linearly related (i.e., Secondary and University) while others are uncorrelated (i.e., University and Age Group 3). The values of CorGC using HS and PCOP are usually similar (PCOP based coefficient taking in general smaller values), but there are examples where one is much higher than the other (i.e., Secondary and Age Group 3, or University and Age Group 4). This fact indicates that both methods are able to measure different features of the joint distribution.

7 Statistical applications of CorGC

In this Section we introduce some statistical applications of the CorGC coefficient. They are illustrated together with the Barcelona ZRP data (see Table 2).

7.1 Testing independence between two random variables

The CorGC coefficient (estimated by a principal curve fitting procedure) can be used as a test statistic for testing the null hypothesis of independence between two random variables, X and Y, against the alternative that (X, Y) are distributed along a curve. A random permutation mechanism (random assignment of observed y_j to observed x_i) allows us to approximate the null distribution of the test statistic.

As an example, this test procedure is used to test independence between pairs of variables in the Barcelona ZRP data set. The p-values (computed from 999 random permuted samples) for the independence tests using HS based CorGC and PCOP based CorGC as test statistics are shown in the lower diagonal entries of Table 3. The p-values for the incorrelation test using correlation coefficient (in absolute value) are also provided as a reference. This aids in assessing the advantage of using nonlinear dependence measures as test statistics.

There are many pairs of variables where the three independence test statistics lead to the same result (for instance, all three indicate that Primary and

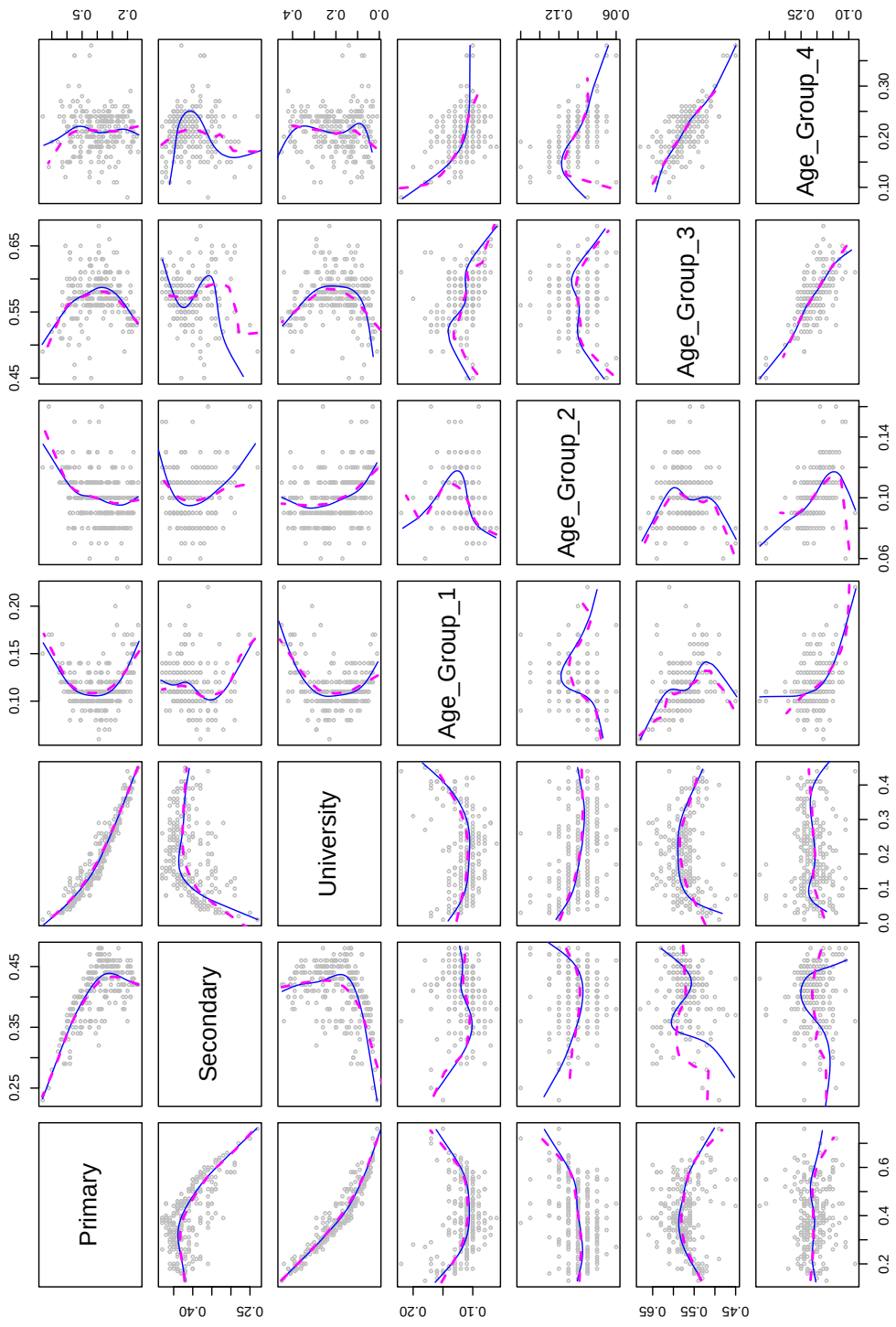

Figure 6: Scatter-plots matrix for the the Barcelona ZRP data.

Table 3: Upper diagonal entries (from top to bottom): Correlation coefficient (in absolute value), HS based CorGC and PCOP based CorGC for the Barcelona ZRP data. Lower diagonal entries (from top to bottom): p-values for the test of incorrelation and the tests of independence using correlation coefficient (in absolute value), HS based CorGC and PCOP based CorGC as test statistics, respectively.

	.6356	.9476	.1851	.2852	.0156	.0023
Primary	.6823	.9628	.6120	.4096	.5573	.2435
	.6338	.9625	.5857	.3683	.5194	.1892
.000		.3575	.0462	.0617	.0305	.0276
.000	Secondary	.6072	.2669	.5442	.6570	.6085
.000		.5357	.2280	.3770	.1623	.3536
.000	.000		.2083	.3184	.0096	.0041
.000	.081	University	.6283	.5119	.6109	.4249
.000	.000		.5588	.4064	.6120	.2905
.005	.501	.001	Age	.2391	.2803	.3872
.000	.889	.000	Group	.6512	.4205	.3982
.000	.686	.000	1	.5175	.4176	.4020
.000	.325	.000	.000	Age	.0362	.5290
.002	.000	.000	.002	Group	.2837	.5781
.000	.061	.000	.000	2	.2518	.4907
.807	.627	.874	.000	.563	Age	.6961
.002	.000	.000	.296	.147	Group	.7128
.000	.935	.000	.000	.260	3	.6605
.973	.680	.948	.000	.000	.000	Age
.463	.006	.532	.035	.000	.000	Group
.300	.696	.150	.001	.000	.000	4

University are not independent and that Primary and Age Group 4 can be considered independent). For pairs of uncorrelated variables having a U-shape joint distribution, the existence of non-linear dependence is detected by the CorGC based tests (this is the case for Primary and Age Group 3 or University and Age Group 3, for instance). There are also some pairs of variables where using HS based CorGC and PCOP based CorGC as independence test statistics do not lead to the same conclusion. For instance, Secondary and Age Group 3 are declared independent when using the PCOP based CorGC, but not when using the HS based CorGC. The opposite occurs for Age Group 1 and Age Group 3. This is in agreement with our remark at the end of Section 6 on the occasional inconsistency between both CorGC computation methods.

Finally, it should be noted that similar values in HS based CorGC and PCOP based CorGC do not correspond to similar p-values for testing independence. For instance, for Age Group 1 and Age Group 3 the CorGC values are .4205 and .4176, respectively, and the corresponding p-values are .213 and .000 (respectively). This fact reinforces our previous observations that HS based CorGC tend to be greater than PCOP based values.

7.2 Testing joint linear structure for two random variables

Let (X, Y) be a bivariate random variable. Now we wish to test the null hypothesis stating that the relation between X and Y, if any, is linear, against the alternative asserting that this relation is not just linear and that in fact X and Y are distributed along a curve (not being a straight line). A random sample $(x_i, y_i), i = 1, \ldots, n$ from (X, Y) is available.

Here *joint linear structure* for two random variables X and Y is understood as the type of relation between them that can be captured by fitting a straight line to the joint distribution of (X, Y). Observe that two independent variables can be said to be distributed along a straight line (with slope equal to 0 or infinity). Therefore the null hypothesis we are testing is equivalent to stating that X and Y are either linearly dependent or independent.

The procedure we propose for testing this null hypothesis is as follows. The first step is to transform the observed data $(x_i, y_i), i = 1, \ldots, n$, into their principal component scores, say $(u_i, v_i), i = 1, \ldots, n$. This transformation is just a rotation in $\mathbb{R}^2$. By construction, $(u_i, v_i), i = 1, \ldots, n$ are always incorrelated. In addition, under the null hypothesis, the data (u_i, v_i) can be considered as coming from a bivariate distribution (U, V), U and V being independent. So the second and last step is to apply the independence test described in Section

Table 4: P-values for the test of joint linear structure using HS based CorGC (upper diagonal entries) and PCOP based CorGC (lower diagonal entries) for the Barcelona ZRP data.

Primary	.000	.000	.000	.002	.000	.553
.000	Secondary	.000	.855	.000	.000	.001
.000	.000	University	.000	.006	.000	.550
.000	.763	.000	Ag.Gr.1	.000	.306	.000
.000	.006	.010	.035	Ag.Gr.2	.076	.243
.000	.972	.000	.436	.015	Ag.Gr.3	.574
.265	.927	.057	.574	.648	.583	Ag.Gr.4

7.1 to the data set $(u_i, v_i), i = 1, \ldots, n$.

This test procedure is used to test joint linear structure between pairs of variables in the Barcelona ZRP data set. Table 4 contains the p-values (computed from 999 random permuted samples) for this test using HS based CorGC (upper diagonal entries) and PCOP based CorGC (lower diagonal entries) as test statistics.

As expected, the null hypothesis of joint linear structure (which includes independence) is not rejected for pairs of variables where the independence was not previously rejected (see Section 7.1 and Table 3). This is the case not only for pairs of variables Primary and Age Group 4, Secondary and Age Group 1 and University and Age Group 4 using both HS and PCOP based CorGC as test statistic, but also Age Group 1 and Age Group 3, and Age Group 2 and Age Group 3 using the HS based statistic, and Secondary and Age Group 3, and Secondary and Age Group 4 using the PCOP based statistic. There is one case (Age Group 2 and Age Group 3 using PCOP) that does not adhere completely to this general rule, with no apparent explanation. We recommend testing first the independence hypothesis and then testing joint linear structure only for pairs of variables where the independence hypothesis is rejected.

There are three pairs of variables not previously declared independent, for which the null hypothesis of linear dependency is not rejected: Age Group 1 and Age Group 3, Age Group 2 and Age Group 4, and Age Group 3 and Age Group 4. The scatter-plots of these pairs of variables support these results, especially for the last case.

The test of joint linear structure proposed here has some points in common with the test for a linear relationship defined in Bowman and Azzalini (1997), Chapter 5: both procedures are aimed at testing the null hypothesis of linear

relationship. Nevertheless, the proposal of Bowman and Azzalini (1997) is based on nonparametric regression, and then the results depend on what variable has been specified as the response. Our proposal, on the other hand, is symmetric on the order of variables. Moreover, a permutation mechanism is not used for computing the p-value for the test (even though it would be possible to do so).

7.3 A similarity measure between pairs of variables

The absolute value of the correlation coefficient has traditionally been used as a similarity measure for pairs of variables (see Johnson and Wichern (2002), Chapter 12, for instance). Similarly, the CorGC coefficient can be considered as a similarity measure between two random variables. When this measure is computed for all pairs of marginals in a p-dimensional random variable (or data set), a similarity matrix S is obtained, with entries $s_{ij} \in [0,1]$, $i = 1 \ldots, p$, $j = 1 \ldots, p$. A standard way to obtain dissimilarities from a similarity measure (having diagonal entries equal to 1) is to define $d_{ij} \propto \sqrt{1 - s_{ij}}$ (see Johnson and Wichern (2002), Chapter 12, and references therein). Let $D = (d_{ij})$. In fact, this is the relation between a scalar product and a Euclidean distance defined on a vector space $\mathcal{A}$ with scalar product: if $s_{ij} = \langle a_i, a_j \rangle$, $d_{ij} = \langle a_i - a_j, a_i - a_j \rangle$ and $\langle a, a \rangle = 1$ for all $a \in \mathcal{A}$, then $d_{ij} \propto \sqrt{1 - s_{ij}}$.

The definition of CorGC does not guarantee that the similarity matrix S is positively defined. Therefore the dissimilarity matrix D can be non-Euclidean (a $p \times p$ dissimilarity matrix D is Euclidean if p points a_i in $\mathbb{R}^p$ exist such that the Euclidean distance between a_i and a_j is d_{ij} for all i, j; this property is equivalent to the positive definition of the corresponding similarity matrix S).

As an example, a similarity matrix S_{HS} can be constructed with entries $s_{ij} = s_{ji}$ and equal to the second row of entry (i, j) in Table 3. This is the similarity matrix containing the HS based CorGC coefficients for the seven variables in the Barcelona ZRP data set. The similarity matrix corresponding to the PCOP based CorGC, S_{PCOP}, takes the third row element in the entries of Table 3. With the first row element we define the similarity matrix corresponding to the absolute value of correlation coefficients, say $S_{|\rho|}$. None of these three similarity matrices are positive definite for our data set.

A similarity matrix containing CorGC coefficients (or the associated dissimilarity matrix) can be the base for later analysis as a cluster analysis for variables or non-metric multidimensional scaling (MDS). For instance, Figure 7 shows two planar configurations of the seven variables in the Barcelona ZRP data set. They are built by using non-metric MDS based on $S_{|\rho|}$ (left panel)

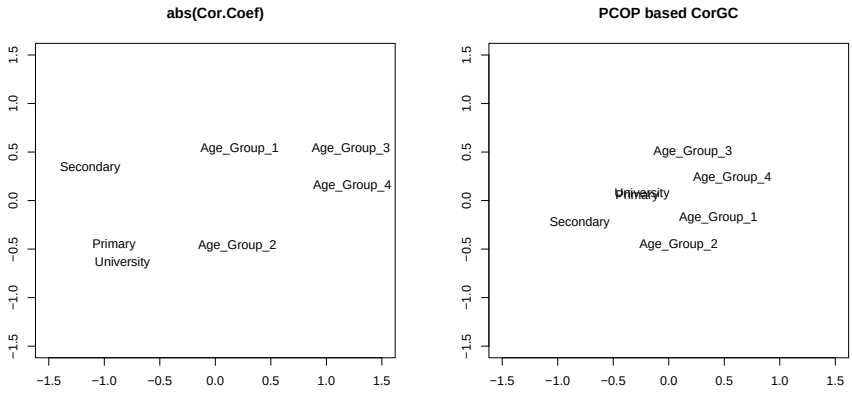

Figure 7: Two-dimensional non-metric MDS configurations for the seven variables in the Barcelona ZRP data. Absolute value of correlation coefficients (left panel) and PCOP based CorGC (right panel) are used as a similarity measure between variables.

and on S_{PCOP} (right panel). We use the function `isoMDS` from the R library `MASS` accompanying the book of Venables and Ripley (2002).

The planar configuration derived from $S_{|\rho|}$ is more scattered than that based on S_{PCOP}. This is because similarities between variables are stronger when nonlinear relations are taken into account. For instance, variables Primary and University are close when we use $S_{|\rho|}$ but they completely overlap in the map based on S_{PCOP} because their nonlinear dependence is stronger than that suggested by the correlation coefficient. Similarly, variables Secondary and Age Group 2 (with low correlation and clear nonlinear dependency) are closer in the S_{PCOP} map than in that based on $S_{|\rho|}$. The same is true for Age Group 1 and Age Group 2. These graphs show that the two groups of variables (education related variables on the one hand, and age variables on the other) are closer in the S_{PCOP}. This is in accordance with the fact that the relations between variables of both groups are mainly nonlinear.

8 Discussion

In this paper we present two new measures of dependence between two random variables distributed along a curve: the covariance and the correlation along the curve. We show that they verify a set of desirable properties closely related

to Rényi's axioms. The sampling version is also defined, based on the concept and estimators of principal curves. Their performance as estimators of the population concept is addressed by a simulation study. A real data set illustrates how the new measures can be used in several statistical applications, such as testing independence and linearity, or defining similarities between variables. Other applications could also be defined in a similar way as generalizations of canonical correlations or partial least squares. The methods described in the paper are implemented as functions in R and are available at the authors' web page.

Acknowledgements

Research partially supported by the Spanish Ministry of Education and Science and FEDER, MTM2006-09920, and by the EU PASCAL Network of Excellence, IST-2002-506778. We are grateful to Sonia Broner who provided us with the ZRP Barcelona data set.

References

Bell, C. (1962). Mutual information and maximal correlation as measures of dependence. *The Annals of Mathematical Statistics 33*, 587–595.

Bjerve, S. and K. Doksum (1993). Correlation curves: Measures of association as functions of covariate value. *The Annals of Statistic 21*, 890–902.

Bowman, A. W. and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press.

Delicado, P. (2001). Another look at principal curves and surfaces. *Journal of Multivariate Analysis 77*, 84–116.

Delicado, P. and M. Huerta (2003). Principal Curves of Oriented Points: Theoretical and computational improvements. *Computational Statistics 18*, 293–315.

Delicado, P. and M. Smrekar (2007, August 12-17). Mixture of nonlinear models: A Bayesian fit for principal curves. In *Proceedings of the International Joint Conference on Neural Networks*, Orlando, Florida, USA. ISBN: 1-4244-1380-X.

Doksum, K., S. Blyth, E. Bradlow, X. Meng, and H. Zhao (1994). Correlation curves as local measures of variance explained by regression. *Journal of American Statistical Association 89*, 571–582.

Doksum, K. and S. Froda (2000). Neighborhood correlation. *Journal of statistical planning and inference 91*, 267–294.

Hastie, T. and W. Stuetzle (1989). Principal curves. *Journal of the American Statistical Association 84*, 502–516.

Holland, P. and Y. Wang (1987). Dependence function for continuous bivariate densities. *Commun. Statist. 16*, 863–876.

Johnson, R. A. and D. W. Wichern (2002). *Applied multivariate statistical analysis* (5th ed.). Prentice Hall.

Jones, M. C. (1996). The local dependence function. *Biometrica 83*, 899–904.

Kégl, B., A. Krzyzak, T. Linder, and K. Zeger (2000). Learning and design of principal curves. *IEEE Trans. Pattern Analysis and Machine Intelligence 22*, 281–297.

Nelsen, R. B. (2006). *An Introduction to Copulas* (Second ed.). Springer Series in Statistics. New York: Springer.

R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Rényi, A. (1959). On measures of dependence. *Acta. Math. Acad. Sci. Hungar. 10*, 441–451.

Scheizer, B. and E. F. Wolff (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics 9*(4), 879–885.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). Springer.

Wang, Y. (1993). Construction of continuous bivariate density functions. *Statist. Sinica 3*, 173–187.