SORT 30 (1) January-June 2006, 55-84

The importance of being the upper bound in the bivariate family*

C. M. Cuadras

University of Barcelona

Abstract

Any bivariate cdf is bounded by the Fréchet-Hoeffding lower and upper bounds. We illustrate the importance of the upper bound in several ways. Any bivariate distribution can be written in terms of this bound, which is implicit in logit analysis and the Lorenz curve, and can be used in goodness-of-fit assesment. Any random variable can be expanded in terms of some functions related to this bound. The Bayes approach in comparing two proportions can be presented as the problem of choosing a parametric prior distribution which puts mass on the null hypothesis. Accepting this hypothesis is equivalent to reaching the upper bound. We also present some parametric families making emphasis on this bound.

MSC: 60E05, 62H17, 62H20, 62F15

Keywords: Hoeffding's lemma, Fréchet-Hoeffding bounds, given marginals, diagonal expansion, logit analysis, goodness-of-fit, Lorenz curve, Bayes test in 2×2 tables.

1 Introduction

Several concepts and equations play an important role in statistical science. We prove that the bivariate upper Fréchet bound and the maximal Hoeffding correlation are two related expressions which, directly or implicitly, are quite useful in probability and statistics.

* Dedicated to the memory of Joan Augé (1919-1993).

Address for correspondence: C.M. Cuadras. Universitat de Barcelona. Diagonal, 645. 08023 Barcelona (Spain).
E-mail: ccuadras@ub.edu

Received: February 2006
Accepted: June 2006

Let X, Y be two random variables with continuous joint cumulative distribution function (cdf) $H(x, y)$ and marginal cdf's $F(x), G(y)$. Assuming finite variances, Hoeffding (1940) proved that the covariance in terms of the cdf's is given by

$$\mathrm{Cov}(X, Y) = \int_{R^2} (H(x, y) - F(x)G(y))dxdy. \tag{1}$$

Then he proved that the correlation coefficient

$$\rho_H(X, Y) = \mathrm{Cov}(X, Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

satisfies the inequality

$$\rho^- \leq \rho_H \leq \rho^+,$$

where ρ^-, ρ^+ are the correlation coefficients for the bivariate cdf's

$$H^-(x, y) = \max\{F(x) + G(y) - 1, 0\} \quad \text{and} \quad H^+(x, y) = \min\{F(x), G(y)\},$$

respectively.

In another seminal paper Fréchet (1951) proved the inequality

$$H^-(x, y) \leq H(x, y) \leq H^+(x, y), \tag{2}$$

where H^- and H^+ are the so-called lower and upper Fréchet-Hoeffding bounds. If H reaches these bounds then the following functional relations hold between the random variables:

$$\begin{aligned} F(X) &= 1 - G(Y), \quad \text{(a.s.) if} \quad H = H^-, \\ F(X) &= G(Y), \qquad \text{(a.s.) if} \quad H = H^+. \end{aligned}$$

The distributions H^-, H^+ and $H = FG$ (stochastic independence) are examples of cdf's with marginals F, G. The construction of distributions when the marginals are given is a topic of increasing interest – see, for example, the proceedings edited by Cuadras, Fortiana and Rodriguez-Lallena (2002).

Note that H^- and H^+ are related by

$$H^+(x, y) = F(x) - H^-(x, G^{-1}(1 - G(y)),$$

and that the $p-$dimensional generalization of (2) is

$$H^-(x_1, \ldots, x_p) \leq H(x_1, \ldots, x_p) \leq H^+(x_1, \ldots, x_p),$$

where $H(x_1, \ldots, x_p)$ is a cdf with univariate marginals $F_1, \ldots, F_p$ and

$$\begin{aligned} H^-(x_1, \ldots, x_p) &= \max\{F_1(x_1) + \ldots + F_p(x_p) - (p-1), 0\}, \\ H^+(x_1, \ldots, x_p) &= \min\{F_1(x_1), \ldots, F_p(x_p)\}. \end{aligned}$$

However, if $p > 2$, in general only H^+ is a cdf, see Joe (1997). Thus we may focus our study on the Fréchet-Hoeffding upper bound.

The aim of this paper is to present some relevant aspects of H^+, which may generate any bivariate cdf and is implicit in some statistical problems.

2 Distributions in terms of upper bounds

Hoeffding's formula (1) was extended by Cuadras (2002a) as follows. Let us suppose that the ranges of X, Y are the intervals $[a, b], [c, d] \subset \overline{R}$, respectively. Thus $F(a) = G(c) = 0$, $F(b) = G(d) = 1$. Let $\alpha(x), \beta(y)$ be two real functions of bounded variation defined on $[a, b], [c, d]$, respectively. If $\alpha(a)F(a) = \beta(c)G(c) = 0$ and the covariance between $\alpha(X), \beta(Y)$ exists, it can be obtained from

$$\mathrm{Cov}(\alpha(X), \beta(Y)) = \int_a^b \int_c^d (H(x, y) - F(x)G(y)) d\alpha(x) d\beta(y). \tag{3}$$

Suppose that the measure $dH(x, y)$ is absolutely continuous with respect to $dF(x)dG(y)$ and that

$$\int_a^b \int_c^d (dH(x, y))^2 / dF(x)dG(y) < \infty.$$

Then the following diagonal expansion

$$dH(x, y) - dF(x)dG(y) = \sum_{k \geq 1} \rho_k a_k(x) b_k(y) dF(x)dG(y) \tag{4}$$

exists, where ρ_k, $a_k(X)$, $b_k(Y)$ are the canonical correlations and variables, respectively (see Hutchinson and Lai, 1991).

Let us consider the upper bounds

$$F^+(x, y) = \min\{F(x), F(y)\}, \quad G^+(x, y) = \min\{G(x), G(y)\},$$

and the symmetric kernels

$$K(s, t) = F^+(s, t) - F(s)F(t), \quad L(s, t) = G^+(s, t) - G(s)G(t).$$

Then using (3) and integrating (4), we can obtain the following expansion

$$H(x,y) = F(x)G(y) + \sum_{k\geq 1} \rho_k \int_a^b K(x,s)da_k(s) \int_c^d L(t,y)db_k(t),$$

which shows the generating power of the upper bounds (see Cuadras, 2002b, 2002c). Thus we can consider the nested family

$$H_n(x,y) = F(x)G(y) + \sum_{k=1}^{n} \rho_k \int_a^b K(x,s)da_k(s) \int_c^d L(t,y)db_k(t),$$

by taking generalized orthonormal sets of functions (a_k) and (b_k) with respect to F and G. It is worth noting that it can exist a non-countable class of canonical correlations and functions (Cuadras, 2005a).

3 Correspondence analysis on the upper bound

Correspondence analysis (CA) is a multivariate method to visualize categorical data, typically presented as a two-way contingency table $\mathbf{N}$. The distance used in the graphical display of the rows (and columns) of $\mathbf{N}$ is the so-called chi-square distance between the profiles of rows (and between the profiles of columns). This method is described in Benzécri (1973) and Greenacre (1984), and it can be interpreted as the discrete version of (4) – see also Cuadras *et al.* (2000).

Let $\mathbf{N} = (n_{ij})$ be an $I \times J$ contingency table and $\mathbf{P} = n^{-1}\mathbf{N}$ the correspondence matrix, where $n = \sum_{ij} n_{ij}$. Let $\mathbf{r} = \mathbf{P1}$, $\mathbf{D}_r$ =diag($\mathbf{r}$), $\mathbf{c} = \mathbf{P}^\mathsf{T}\mathbf{1}$, $\mathbf{D}_c$ =diag($\mathbf{c}$), the vectors and diagonal matrices with the marginal frequencies of $\mathbf{P}$.

CA uses the singular value decomposition

$$\mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}^\mathsf{T})\mathbf{D}_c^{-1/2} = \mathbf{U}\mathbf{D}_\sigma\mathbf{V}^\mathsf{T}, \tag{5}$$

where $\mathbf{D}_\sigma$ is the diagonal matrix of singular values in descending order, and $\mathbf{U}$ and $\mathbf{V}$ have orthonormal columns. To represent the I rows of $\mathbf{N}$ we may take as principal coordinates the rows of $\mathbf{A} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\sigma$. Similarly, to represent the J columns of $\mathbf{N}$ we may use the principal coordinates contained in the rows of $\mathbf{B} = \mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D}_\sigma$. CA has the advantage that we can perform a joint representation of rows and columns, called the symmetric representation, as a consequence of the transition relations

$$\mathbf{A} = \mathbf{D}_r^{-1}\mathbf{P}\mathbf{B}\mathbf{D}_\sigma^{-1}, \quad \mathbf{B} = \mathbf{D}_c^{-1}\mathbf{P}^\mathsf{T}\mathbf{A}\mathbf{D}_\sigma^{-1}. \tag{6}$$

Let us apply CA on the upper bound. Consider the $I \times I$ triangular matrix

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

and similarly the $J \times J$ matrix $\mathbf{M}$. The cumulative joint distribution is $\mathbf{H} = \mathbf{LPM}^{\mathsf{T}}$ and the cumulative marginals are $\mathbf{R} = \mathbf{Lr}$ and $\mathbf{C} = \mathbf{Mc}$. The $I \times J$ matrix $\mathbf{H}^{+} = (h_{ij}^{+})$ with entries

$$h_{ij}^{+} = \min\{\mathbf{R}(i), \mathbf{C}(j)\}, \quad i = 1, \cdots, I, \ j = 1, \cdots, J,$$

contains the cumulative upper bound for table $\mathbf{N}$. The correspondence matrix for this bound is

$$\mathbf{P}^{+} = \mathbf{L}^{-1}\mathbf{H}^{+}(\mathbf{M}^{\mathsf{T}})^{-1}.$$

For instance, if $I = J = 2$ and $\mathbf{r} = (s, 1-s)^{\mathsf{T}}$, $\mathbf{c} = (t, 1-t)^{\mathsf{T}}$, then $\mathbf{R} = (\mathbf{s}, 1)^{\mathsf{T}}$, $\mathbf{C} = (t, 1)^{\mathsf{T}}$ and

$$\mathbf{H}^{+} = \begin{bmatrix} \min\{s,t\} & s \\ t & 1 \end{bmatrix}, \quad \mathbf{P}^{+} = \begin{bmatrix} \min\{s,t\} & s - \min\{s,t\} \\ t - \min\{s,t\} & 1 - s - t + \min\{s,t\} \end{bmatrix}.$$

For a geometric study of Fréchet-Hoeffding bounds in $I \times J$ probabilistic matrices, see Nguyen and Sampson (1985). For a probabilistic study with discrete marginals (binomial, Poisson), see Nelsen (1987).

Table 1: *Survey combining staff-groups with smoking categories (left) and upper bound correspondence matrix (right).*

	Original table Smoking category				*Upper bound* Smoking category			
Staff	(0)	(1)	(2)	(3)	(0)	(1)	(2)	(3)
SM	4	2	3	2	0.057	0	0	0
JM	4	3	7	4	0.093	0	0	0
SE	25	10	12	4	0.166	0.098	0	0
JE	18	24	33	13	0	0.135	0.321	0
SC	10	6	7	2	0	0	0	0.129

Example 1 Table 1, left, (Greenacre, 1984) reports a cross-tabulation of staff-groups (SM=Senior Managers, JM=Junior Managers, SE=Senior Employers, JE=Junior Employers, SC=Secretaries) by smoking category (none(0). light(1), medium(2), heavy(3)) for 193 members of a company. CA on Table 1, right, which contains the

relative frequency upper bound, provides Figure 1. This table is quasi-diagonal. Note the proximity of the rows to the columns, specially along the first dimension.

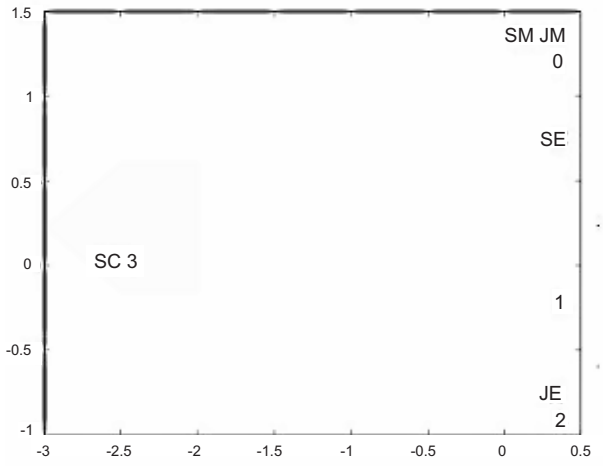

Figure 1: *Symmetric correspondence analysis representation of the upper bound in Table 1, right.*

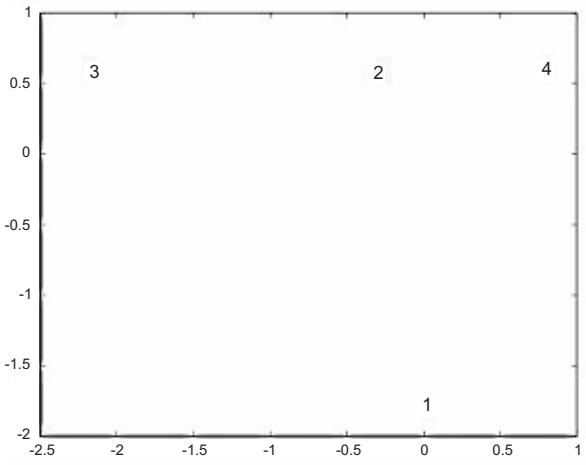

Figure 2: *Symmetric correspondence analysis representation of the upper bound in Table 2, right. Rows and columns are represented on coincident points.*

Example 2 CA is now performed on Table 2, left, an artificial 4×4 table whit the same marginals. Figure 2 exhibits the representation of the relative frequency upper bound. Now this table is diagonal. Note that rows and columns are placed on coincident points.

Table 2: *Artificial contingency table with the same margin frequencies (left) and upper bound correspondence matrix (right).*

	Original table				*Upper bound*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
(1)	8	6	2	8	0.24	0	0	
(2)	6	0	4	10	0	0.20	0	
(3)	6	4	0	4	0	0	0.14	
(4)	4	10	8	20	0	0	0	0.42

4 Orthogonal expansions

Here we work only with a r. v. X with range $[a, b]$, continuous cdf F, and the above symmetric kernel $K(s,t) = \min\{F(s), F(t)\} - F(s)F(t)$. This kernel is the covariance of the stochastic process $\mathbf{X} = \{X_t,\ t \in [a,b]\}$, where X_t is the indicator of $[X > t]$. If the trace

$$\mathrm{tr}(K) = \int_a^b F(t)(1 - F(t))dt$$

is finite, K can be expanded as

$$K(s,t) = \sum_{k\geq 1} \lambda_k \psi_k(s)\psi_k(t),$$

where $\psi_k, \lambda_k,\ k \geq 1$, are the eigenfunctions and eigenvalues related to the integral operator defined by K. Let us consider the integrals

$$f_k(x) = \int_a^x \psi_k(t)dt.$$

Direct application of (3) shows that $(f_k(X))$ is a sequence of mutually uncorrelated random variables:

$$\mathrm{Cov}(f_i(X), f_j(X)) = \begin{cases} 0 & \text{if } i \neq j, \\ \lambda_i & \text{if } i = j. \end{cases}$$

These variables are principal components of $\mathbf{X}$ and $f_1(X)$ characterizes the distribution of X (Cuadras 2005b).

Examples of principal components $f_n(X)$ and the corresponding variances λ_n are:

1. $(\sqrt{2}/(n\pi))(1 - \cos n\pi X)$, $\lambda_n = 1/(n\pi)^2$, if X is $[0,1]$ uniform.
2. $\left[2J_0(\xi_n \exp(-X/2)) - 2J_0(\xi_n)\right]/\xi_n J_0(\xi_n)$, $\lambda_n = 4/\xi_n^2$, if X is exponential with

unit mean, where ξ_n is the $n-$th positive root of J_1 and J_0, J_1 are the Bessel functions of the first order.

3. $(n(n+1))^{-1/2}[L_n(F(X)) + (-1)^{n+1}\sqrt{2n+1}]$, $\lambda_n = 1/(n(n+1))$, if X is standard logistic, where (L_n) are the Legendre polynomials on $[0,1]$.

4. $c_n[X\sin(\xi_n/X) - \sin(\xi_n)]$, $\lambda_n = 3/\xi_n^2$, if X is Pareto with $F(x) = 1 - x^{-3}$, $x > 1$, where $c_n = 2\xi_n^{-1/2}(2\xi_n - \sin(2\xi_n))^{-1/2}$ and $\xi_n = \tan(\xi_n)$.

Assuming a finite, we can expand $\mathbf{X}$ as $X_t = \psi_1(t)f_1(X) + \psi_2(t)f_2(X) + \ldots$ and from $X_t = X_t^2$, integrating X_t on $[a,b]$ we have $X = a + \int_a^b X_t dt = a + \int_a^b X_t^2 dt$, and the variable X can be expanded in two ways

$$X = a + \sum_{k\geq 1} f_k(b)f_k(X) = a + \sum_{k\geq 1} f_k(X)^2,$$

where the convergence is in the mean-square sense. See Cuadras and Fortiana (1995), Cuadras *et al.* (2006) for other expansions, and Cuadras and Cuadras (2002) for applications in goodness-of-fit assessment. These expansions depend on a countable set of functions, again related to the upper bound.

5 Logit and probit analysis

The upper bound is implicit in some transformations. Suppose that F, the cdf of X, is unknown, whereas Y follows the logistic distribution $G(\alpha + \beta y)$, where

$$G(y) = 1/(1 + \exp(-y)), \quad -\infty < y < +\infty.$$

We may take G as a "model" for F in the sense that H, the cdf of (X,Y), attains the upper bound $H^+(x,y) = \min\{F(x), G(\alpha+\beta y)\}$. In other words, we assume the functional relation $F(X) = G(\alpha + \beta Y)$, with F unknown and G standard logistic (Figure 3). This gives rise to the logistic transformation

$$\ln\left(\frac{F(x)}{1 - F(x)}\right) = \alpha + \beta y. \tag{7}$$

If the plot of $\ln[F(x)/(1 - F(x))]$ against y is almost linear, then the data fit the upper bound. The probit transformation arises similarly by considering the $N(0,1)$ distribution.

In logit and probit analysis applied in bioassay, the user observes the proportion of $[X > x]$ for x fixed, i.e., observes $F(x)$ rather than X. Then (7) is used, where the parameters α, β should be estimated. Thus the outcomes arise from the above random process $\mathbf{X} = \{X_t,\ t \in [a,b]\}$. It is also worth noting that $F(X)$ is the first principal

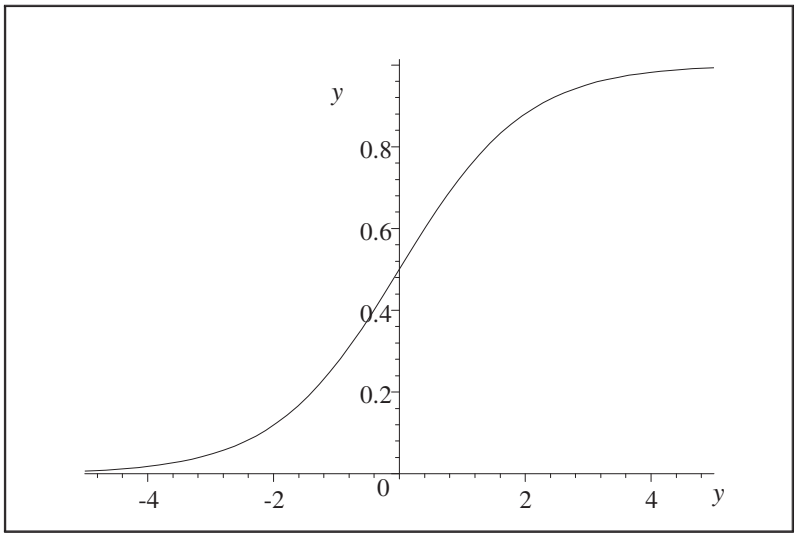

Figure 3: *The logistic curve illustrates the support of the upper bound when a distribution function is considered logistic as a model.*

component of $\mathbf{X}$ only if X is logistic, see Cuadras and Lahlou (2000). Thus logit is better than probit and both transformations can be viewed as a consequence of using the upper bound.

6 Given the regression curve

When Y is increasing in X, an ideal and quite natural relation between X and Y is $F(X) = G(Y)$. Therefore, to predict Y given X when H is unknown, a reasonable way is to use the Fréchet-Hoeffding upper bound. This gives rise to the regression curve

$$y = G^{-1} \circ F(x).$$

Of course, H^+ puts all mass concentrated on this curve.

Let us construct a cdf H_θ with this (or any in general) regression curve. If the ranges of X, Y are the intervals $[a, b]$, $[c, d]$, and $\varphi : [a, b] \to [c, d]$ is an increasing function, the following family

$$H_\theta(x, y) = \theta F(\min\{x, \varphi^{-1}(y)\}) + (1 - \theta)F(x)J_\theta(y), \quad 0 \le \theta < \theta^+, \tag{8}$$

where θ^+ is given below, is a bivariate cdf with marginals F, G, provided that

$$J_\theta(y) = [G(y) - \theta F(\varphi^{-1}(y))]/(1 - \theta)$$

is a cdf. The regression curve is linear in φ and $H_\theta(x, y)$ has a singular part with mass on the curve $y = \varphi(x)$. It can be proved that $P[Y = \varphi(X)] = \theta$ (see Cuadras, 1992, 1996).

With $\varphi = G^{-1} \circ F$ equation (8) reduces to

$$H_\theta(x, y) = \theta F(\min\{x, F^{-1} \circ G(y)\}) + (1 - \theta)F(x)G(y), \quad 0 \le \theta \le 1. \tag{9}$$

and the upper bound is attained at $\theta = 1$.

Next, let us find the covariance and prove an inequality. Let $\psi = \varphi^{-1}$, suppose that ψ' exists and X, Y have densities f, g (Lebesgue measure). Differentiation of $J_\theta(y)$ gives $g(y) - \theta f(\psi(y))\psi'(y) > 0$, hence θ is bounded by

$$\theta^+ = \inf_{y\in[c,d]} \{\frac{g(y)}{f(\psi(y))\psi'(y)}\},$$

where we write "ess inf" if necessary. From (3) we have

$$\begin{aligned} \mathrm{Cov}_{H_\theta}(X, Y) &= \int_a^b \int_c^d \theta(F(\min\{x, \psi(y)\}) - F(x)F(\psi(y))d(x)d\varphi(y) \\ &= \theta \mathrm{Cov}(X, \varphi(X)). \end{aligned} \tag{10}$$

Thus the following inequality holds:

$$\inf_{y\in[c,d]} \{\frac{g(y)}{f(\psi(y))\psi'(y)}\}\rho(X, \varphi(X)) \le \rho^+.$$

In particular, if $\varphi(x) = x$ and f, g have the same support $[a, b]$, we obtain

$$\max[\inf_{x\in[a,b]} \{\frac{f(x)}{g(x)}\}, \inf_{x\in[a,b]} \{\frac{g(x)}{f(x)}\}] \le \rho^+.$$

7 Parent distribution of a data set

Let $\chi = \{x_1, x_2, \ldots, x_N\}$ be a sample of X with unknown cdf F, and let F_N be the empirical cdf. We are interested in ascertaining the parent distribution of χ. This problem has been widely studied assuming that F belongs to a finite family of cdf's $\{F_1, \ldots, F_n\}$ (see Marshall *et al.*, 2001).

The maximum Hoeffding correlation is a good similarity measure between two cdf's. Assuming the variables standardized, it can be computed by

$$\rho^+(F_i, F_j) = \int_0^1 F_i^{-1}(u)F_j^{-1}(u)du.$$

Thus, a distance between F_i and F_j, which lies between 0 and $\sqrt{2}$, is given by

$$d_{ij} = \sqrt{2(1 - \rho^+(F_i, F_j))}.$$

We can also compute the correlation and distance between data and any theoretical cdf. Then the $(n + 1) \times (n + 1)$ matrix $D = (d_{ij})$ is a Euclidean distance matrix and we can perform a metric scaling in order to represent the set $\{F_1, \ldots, F_n, F_N\}$ using the two first principal axes (see Mardia *et al.*, 1979). The graphic display may give an indication of the underlying distribution of the sample.

There are other distances between distributions, e.g., the Kolmogorov distance

$$d(F_i, F_j) = \sup_{-\infty<x<\infty} |F_i(x) - F_j(x)|,$$

(Marshall *et al.*, 2001) and the Wasserstein distance (del Barrio *et al.*, 2000)

$$\mathcal{W}_{ij} = \int_0^1 [F_i^{-1}(u) - F_j^{-1}(u)]^2 du,$$

which can be used for the same purpose. However $d(F_i, F_j)$ and $\mathcal{W}_{ij}$ may not give Euclidean distance matrices and are not invariant under affine transformation of the variables. On the other hand, $\mathcal{W}_{ij}$ is directly related to the maximum correlation (see Cuadras and Cuadras, 2002).

Fortiana and Grané (2002, 2003) refined this approach. They used some statistics based on this maximum correlation and obtained asymptotic and exact tests for testing the exponentiality and the uniformity of a sample, which compare with other goodness-of-fit statistics.

Example 3 Suppose that χ is the $N = 50$ sample of $X =$ "sepal length" of *Iris setosa,* the well-known data set used by R. A. Fisher to illustrate discriminant analysis (see Mardia *et al.*, 1979). Suppose the following statistical models:

$$\{U(\text{uniform}), E(\text{exponential}), N(\text{normal}), G(\text{gamma}), LN(\text{log-normal})\}.$$

The matrix of maximum correlations is reported in Table 3. Figure 4 is the metric scaling representation of probability models and data. The closest model is N, so we may decide that this data is drawn from a normal distribution.

The uniform distribution U, the second closest distribution to the data, may be another candidate. To decide between normal and uniform we may proceed as follows.

First, we assume normality and perform the integral transformation $y = \Phi(x)$ on the standardized sample, where Φ is the $N(0, 1)$ cdf, and correlate the transformed sample χ^*, say, with the principal components $(f_n(X))$, see Section 4. Let r_k =Cor$(\chi^*, f_k(U))$

Table 3: *Maximum Hoeffding correlations among several distributions and data. U (uniform), E (exponential), N (normal), G (gamma), LN (log-normal).*

	U	*E*	*N*	*GA*	*LN*	*Data*
U	1					
E	0.8660	1				
N	0.9772	0.9032	1			
G	0.9472	0.9772	0.9730	1		
LN	0.6877	0.8928	0.7628	0.8716	1	
Data	0.9738	0.8925	0.9871	0.9660	0.7452	1

be the coefficient of correlation between χ^*, with empirical cdf F_N^*, and $f_k(U)$, where the correlation is taken with respect to the upper bound $H_N^+(x,u) = \min[F_N^*(x), u]$. The theoretical correlations are:

$$\rho_k = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ 4\sqrt{6}/(k\pi^2) & \text{if } k \text{ is even.} \end{cases}$$

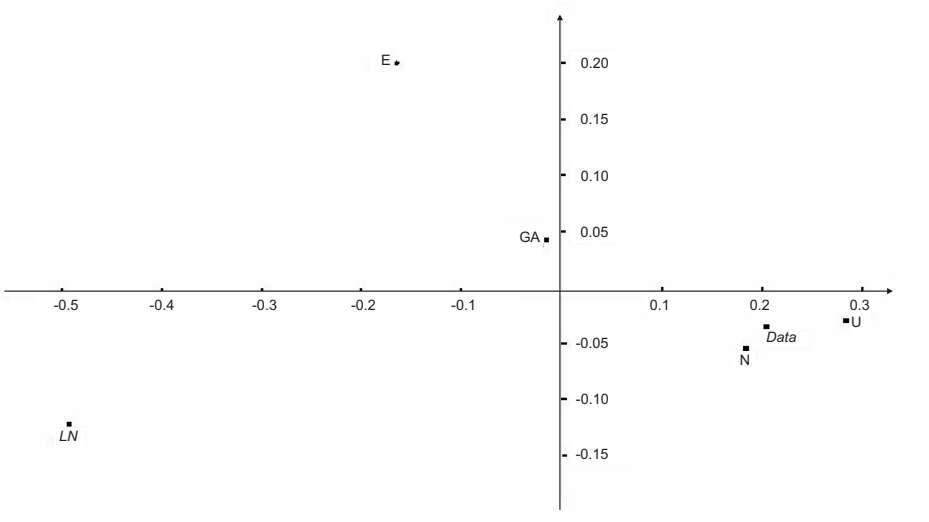

Figure 4: *Metric scaling representation of a sample (Data) and the distributions uniform (U), exponential (E), gamma (GA), normal (N) and log-normal (LN), using the maximum correlation between two distributions (Cuadras and Fortiana, 1994).*

It can be proved that $\rho^+(\chi^*, U) = \sum_{k\geq 1} \rho_k r_k$. Thus

$$\rho^+(\chi^*, U) = (4\sqrt{6}/\pi^2) \sum_{k=0}^{\infty} (r_{2k+1})/(2k+1)^2.$$

This agreement coefficient $\rho^+(\chi^*, U)$ between the transformed sample and the uniform distribution has an expansion similar to the expansions of Cramér-von Mises and Anderson-Darling statistics used in goodness-of-fit.

Second, we perform analogous computations for the original sample χ assuming that it is drawn from an uniform distribution, the alternative model. This gives Table 4, where $\rho^+(\chi, U) = 0.9738$, $\rho^+(\chi^*, U) = 0.9915$, $\rho^+(\chi^*, f_1(U)) = 0.9792$, etc. These results give support to the normality of the sample. See Cuadras and Fortiana (1994), Cuadras and Lahlou (2000) and Cuadras and Cuadras (2002) for further aspects of this graphic test.

Table 4: *Maximum correlations between normal and uniform and correlations among principal directions and data.*

	Theoretical	Normal	Uniform
ρ^+	1	0.9915	0.9738
ρ_1	0.9927	0.9792	0.9435
ρ_2	0	0.0019	−0.0102
ρ_3	0.1103	0.1713	0.2847
ρ_4	0	−0.0277	−0.0324

8 Kendall's tau and Spearman's rho

Any proposal of coefficient of stochastic dependence between X and Y should evaluate the difference between the joint cdf H and the independence FG. Thus

$$A(X, Y) = c \int_{\mathbf{R}^2} (H(x, y) - F(x)G(y))d\mu,$$

where c is a normalizing constant and μ is a suitable measure. The maximum value for $A(X, Y)$, where H has margins F, G, is attained at the upper bound:

$$\max A(X, Y) = c \int_{\mathbf{R}^2} (H^+(x, y) - F(x)G(y))d\mu.$$

However, as proposed by Hoeffding (1940), it is quite convenient to have a coefficient "scale invariant", that is, it should remain unchanged by monotonic transformations of X and Y. The integral transformation $u = F(x)$ and $v = G(y)$ is a monotonic transformation that provides the copula C_H, i.e., a bivariate cdf with uniform marginals on $\mathbf{I} = [0, 1]$, such that

$$H(x, y) = C_H(F(x), G(y)).$$

This copula exists (Sklar's theorem) and is unique if F, G are continuous. Thus we can construct bivariate distributions $H = C(F, G)$ with given univariate marginals F, G by

using copulas C. ("Copula" as a function which links marginals was coined by Sklar (1959). The same concept was called "uniform representation" by G. Kimeldorf and A. Sampson in 1975, and "dependence function" by P. Deheuvels and J. Galambos in 1978).

Kendall's τ and Spearman's ρ_s are coefficients of dependence computed from the copula C_H by using $d\mu = dC_H$ and $d\mu = dudv$, respectively. They are defined by

$$\begin{aligned}\tau &= 4\int_{\mathbf{I}^2}(C_H(u,v) - uv)\, dC_H(u,v) \\ &= 4\int_{\mathbf{I}^2} C_H(u,v)\, dC_H(u,v) - 1,\end{aligned}$$

and

$$\begin{aligned}\rho_s &= 12\int_{\mathbf{I}^2}(C_H(u,v) - uv)\, dudv \\ &= 12\int_{\mathbf{I}^2} C_H(u,v) dudv - 3.\end{aligned}$$

Then $\tau = \rho_s = 1$ when $H = H^+$, that is, when $F(X) = G(Y)$ (a.s.).

Spearman's ρ_s is Pearson's correlation between $F(X)$ and $G(Y)$ and can be 0 even if there is stochastic dependence. For example, $\rho_s = 0$ for the copula $C = uv + \theta(2u^3 - 3u^2 + u)(2v^3 - 3v^2 + v)$, $|\theta| \leq 1$. On the other hand, $F(X)$ is the first principal dimension for the logistic distribution (Section 4). Then we may extend ρ_s by obtaining Pearson's correlation $\mathrm{Cor}(f_1(X), g_1(Y))$ between the principal dimensions $f_1(X)$ and $g_1(Y)$. This may improve the measure of stochastic dependence between X and Y, with applications to testing independence (Cuadras, 2002b, 2002c).

Table 5: *Some parametric families, their properties and the upper bound.*

Family	Spearman	Kendall	Constant	Archimedian	Upper bound
FGM	yes	yes	yes	no	no
Normal	yes	yes	yes	no	yes
Plackett	yes	no	yes	no	yes
Cuadras-Augé	yes	yes	yes	no	yes
Regression (Fréchet)	yes	yes	yes	no	yes
Clayton-Oakes	yes	yes	yes	yes	yes
AMH	yes	yes	yes	yes	no
Frank	yes	yes	no	yes	yes
Raftery	yes	yes	no	no	yes
Gumbel-Barnett	no	no	no	yes	no
Gumbel-Hougaard	no	yes	yes	yes	yes
Joe	no	no	no	yes	yes

9 Some bivariate families

In this section we present some parametric families of bivariate distributions, in terms of F, G rather than copulas, as some aspects such as constant quantity and regression are not well manifested with uniform marginals. To get the corresponding copula simply replace F, G by u, v. For instance, for the FGM family the copula is $C_\alpha = uv[1+\alpha(1-u)(1-v)]$. References for these families can be found in Cuadras (1992, 1996, 2002, 2005), Druet-Mari and Kotz (2001), Hutchinson and Lai (1991), Joe (1997), Kotz *et al.* (2000), Mardia (1970) and Nelsen (1999). Table 5 summarizes some aspects, e.g., whether or not ρ_s and τ can be given in closed form and the family contains the upper bound.

9.1 FGM

The Farlie-Gumbel-Morgenstern family provides a simple and widely used example of distribution H with marginals F, G. This family does not reach H^+ and can be seen as the first term in the diagonal expansion (4).

1. Cdf : $H_\alpha = FG[1+\alpha(1-F)(1-G)], \quad -1 \le \alpha \le +1.$
2. Constant quantity : $\alpha = (H-FG)/[FG(1-F)(1-G)].$
3. Spearman: $\rho_s = \alpha/3.$
4. Kendall: $\tau = 2\alpha/9.$
5. Maximal correlation : $\rho_1 = |\alpha|/3.$
6. Fréchet-Hoeffding bounds : $H^- < H_{-1} < H_0 = FG < H_{+1} < H^+.$

9.2 Normal

Let N_ρ and n_ρ be the cdf and pdf, respectively, of the standard bivariate normal with correlation coefficient ρ. The distribution H_ρ is obtained by the "translation method", as described by Mardia (1970).

1. Cdf : $H_\rho = N_\rho(\Phi^{-1}F, \Phi^{-1}G), \quad -1 \le \rho \le +1.$
2. Constant quantity : $\frac{\rho}{1-\rho^2} = \dfrac{\partial^2 \log n_\rho}{\partial x \partial y}$ (normal marginals).
3. Spearman: $\rho_s = \frac{6}{\pi}\arcsin(\rho/2).$
4. Kendall: $\tau = \frac{2}{\pi}\arcsin(\rho).$
5. Maximal correlation : $\rho_1 = \rho.$
6. Fréchet-Hoeffding bounds : $H_{-1} = H^- < H_0 = FG < H_1 = H^+.$

9.3 Plackett

The Plackett family arises in the problem of correlating two dichotomized variables X, Y, when the ranges are divided into four regions and the correlation is computed as a function of the association parameter ψ. Then H is defined such that

$$\psi = \frac{H(1 - F - G + H)}{(F - H)(G - H)},$$

is constant. $\psi \geq 0$ is the cross product ratio in 2×2 contingency tables.

1. Cdf: $H_\psi = \left[S - \left\{S^2 - 4\psi(\psi - 1)FG\right\}^{1/2}\right] / \{2(\psi - 1)\}, \quad \psi \geq 0,$ where $S = 1 + (F + G)(\psi - 1)$.
2. Constant quantity: $\psi = H\,(1 - F - G + H)\,/[(F - H)(G - H)]$.
3. Spearman: $\rho_s = \dfrac{\psi + 1}{\psi - 1} - \dfrac{2\psi}{(\psi - 1)^2} \ln \psi$.
4. Kendall: τ not in closed form.
5. Fréchet-Hoeffding bounds: $H_0 = H^- < H_1 = FG < H_\infty = H^+$.

9.4 Cuadras-Augé

The Cuadras-Augé family is obtained by considering a weighted geometric mean of the independence distribution and the upper Fréchet-Hoeffding bound. The corresponding copula is $C_\theta = (\min\{u, v\})^\theta\,(uv)^{1-\theta}$. Although obtained independently by Cuadras and Augé (1981), C_θ is the survival copula of the Marshall and Olkin (1967) bivariate distribution when the variables are exchangeable. (The survival copula for H is $C_{\overline{H}}$ such that $\overline{H} = C_{\overline{H}}(\overline{F}, \overline{G})$, where $\overline{F} = 1 - F$, $\overline{G} = 1 - G$, $\overline{H} = 1 - F - G + H$). The canonical correlations for this family constitutes a continuous set. Kimeldorf and Sampson (1975) proposed a copula C_λ also related to Marshall-Olkin. C_λ is given in Block and Sampson (1988) in the form $C_\lambda = u + v - 1 + (1 - u)^\lambda (1 - v)^\lambda \min\{(1 - u)^\lambda, (1 - v)^\lambda\}$, $0 \leq \lambda \leq 1$. See Muliere and Scarsini (1987) for an unified treatment of C_θ, C_λ and other related copulas. A generalization of C_θ, proposed by Nelsen (1991), is $C_{\alpha,\beta} = \min\{u^\alpha, v^\beta\} u^{1-\alpha} v^{1-\beta}$ for $\alpha, \beta \in [0, 1]$.

1. Cdf : $H_\theta = (\min\{F, G\})^\theta\,(FG)^{1-\theta}, \quad 0 \leq \theta \leq 1.$
2. Constant quantity: $\theta = \ln(H/FG)/\ln(H^+/FG)$.
3. Spearman: $\rho_s = 3\theta/(4 - \theta)$.
4. Kendall: $\tau = \theta/(2 - \theta)$.

5. Maximal correlation: $\rho_1 = \theta$.
6. Fréchet-Hoeffding bounds: $H_0 = FG < H_1 = H^+$.

9.5 Regression (Fréchet)

A family with a given correlation coefficient $0 \leq r \leq 1$ can be constructed taking (X, X) with probability r and (X, Y), where X, Y are independents, with probability $(1 - r)$. The cdf is then

$$H^*(x, y) = rF(\min\{x, y\}) + (1 - r)F(x)G(y).$$

A generalization is the regression family H_r defined below (Cuadras, 1992). Family (9) is a particular case. This family extends the weighted mean of the upper bound and independence, $H_\theta = \theta H^+ + (1-\theta)FG$, proposed by Fréchet (1951) and studied by Nelsen (1987) and Tiit (1986). Note that $H_\theta \neq H_r$ have the same copula. See also Section 6.

1. Cdf : $H_r(x, y) = rF(\min\{x, y\}) + (1 - r)F(x)J(y), \quad 0 \leq r < 1,$
   where $J(y) = [G(y) - rF(y)] / (1 - r)$ is a univariate cdf.
2. Spearman: $\rho_s = r$.
3. Kendall: $\tau = r(r + 2)/3$.
4. Constant quantity : $r = [H(x, y) - F(x)G(y)]/[F(\min\{x, y\}) - F(x)F(y)]$.
5. Fréchet-Hoeffding bounds : $H_0 = FG < H_1 \leq H^+$, with $H_1 = H^+$ if $F = G$.

9.6 Clayton-Oakes

The Clayton-Oakes distribution is a bivariate model in survival analysis, which satisfies for all failure times s and t, the equation

$$h(s, t)\bar{H}(s, t) = \frac{1}{c}\int_s^\infty h(u, v)du \int_t^\infty h(u, v)dv$$

where $\bar{H} = 1 - F - G + H$ and h is the density.

1. Cdf: $H_c = \max\{(F^{-c} + G^{-c} - 1)^{-1/c}, 0\}, \quad -1 \leq c < \infty.$
2. Spearman: $\rho_s = 12 \int_0^1 \int_0^1 (u^{-c} + v^{-c})^{-1/c} dudv - 3$ (see Hutchinson and Lai, 1991, p. 240).

3. Kendall: $\tau = c/(c+2)$.
4. Constant quantity : $c^{-1} = h\bar{H}/(\frac{\partial}{\partial x}\bar{H}\frac{\partial}{\partial y}\bar{H})$.
5. Fréchet-Hoeffding bounds : $H_{-1} = H^- < H_0 = FG < H_\infty = H^+$.

This family is also known as: 1) Kimeldorf and Sampson, 2) Cook and Johnson, 3) Pareto.

9.7 Frank

Frank's copula $C = -\theta^{-1}\ln(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$ arises in the context of associative functions. It is characterized by the property that $C(u,v)$ and $\hat{C}(u,v) = u+v-C(u,v)$ are associative, that is, $C(C(u,v),w) = C(u,C(v,w))$ and similarly for $\hat{C}$. Statistical aspects of Frank's family were given by Nelsen (1986) and Genest (1987).

1. Cdf: $H_\theta = -\theta^{-1}\ln(1 + \dfrac{(e^{-\theta F}-1)(e^{-\theta G}-1)}{e^{-\theta}-1})$, $\quad -\infty \le \theta \le \infty$.
2. Spearman: $\rho_s = 1 - (12/\theta)[D_1(\theta) - D_2(\theta)]$.
3. Kendall: $\tau = 1 - (4/\theta)[1 - D_1(\theta)]$, where $D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^t-1}dt$.
4. Fréchet-Hoeffding bounds: $H_{-\infty} = H^- < H_0 = FG < H_\infty = H^+$.

9.8 AMH

The Ali-Mikhail-Haq distribution is obtained by considering the odds in favour of failure against survival. Thus $(1-F)/F = K$ must be non-increasing and $F = 1/(1+K)$. The bivariate extension is $H = 1/(1+L)$, where L is the corresponding bivariate odds function. Some conditions to H gives the model below.

1. Cdf: $H_\alpha = FG/[1 - \alpha(1-F)(1-G)]$, $\quad -1 \le \alpha \le +1$.
2. Constant quantity: $\alpha = (H - FG)/[H(1-F)(1-G)]$.
3. Spearman: $\rho_s = -\frac{12(1+\alpha)}{\alpha^2}\mathrm{diln}(1-\alpha) - \frac{24(1-\alpha)}{\alpha^2}\ln(1-\alpha) - \frac{3(\alpha+12)}{\alpha}$, where $\mathrm{diln}(1-\alpha) = \int_0^\alpha x^{-1}\ln(1-x)dx$ is the dilogarithmic function.
4. Kendall: $\tau = \frac{3\alpha-2}{3\alpha} - \frac{2(1-\alpha)^2}{3\alpha^2}\ln(1-\alpha)$.
5. Fréchet-Hoeffding bounds: $H^- < H_{-1} < H_0 = FG < H_1 < H^+$.

9.9 Raftery

The Raftery distribution is generated by considering

$$X = (1-\theta)Z_1 + JZ_3, \quad Y = (1-\theta)Z_2 + JZ_3,$$

where Z_1, Z_2 and Z_3 are independent and identically distributed exponential with $\lambda > 0$, and J is Bernoulli of parameter θ, independent of Z's. This distribution, via Sklar's theorem, generates a family.

1. Cdf: $H_\theta = H^+ + \dfrac{1-\theta}{1+\theta}(FG)^{1/(1-\theta)}\{1 - [\max\{F,G\}]^{-(1+\theta)/(1-\theta)}\}, \quad 0 \le \theta \le 1.$
2. Spearman: $\rho_s = \dfrac{\theta(4-3\theta)}{(2-\theta)^2}.$
3. Kendall: $\tau = \dfrac{2\theta}{3-\theta}.$
4. Fréchet-Hoeffding bounds: $H_0 = FG < H_1 = H^+.$

9.10 Archimedian copulas

For the independence copula $C = uv$ we have $-\ln C = -\ln u - \ln v$. A fruitful generalization of this additivity, related to $\varphi(t) = -\ln t$, is the key idea for defining the so-called Archimedian copulas (Genest and MacKay, 1987). The cdf is described by a function $\varphi : \mathbf{I} \to [0, \infty)$ such that

$$\varphi(1) = 0, \quad \varphi'(t) < 0, \quad \varphi''(t) > 0,$$

for all $0 < t < 1$, conditions which guarantee that φ has inverse. These copulas are defined as

$$\begin{array}{ll} C(u,v) & = \varphi^{-1}[\varphi(u) + \varphi(v)] \quad \text{if } \varphi(u) + \varphi(v) \le \varphi(0), \\ & = 0 \qquad\qquad\qquad\qquad\quad \text{otherwise.} \end{array}$$

For example, the AMH copula $C = uv/[1 - \alpha(1-u)(1-v)]$ satisfies

$$1 + (1-\alpha)\frac{1-C}{C} = \left[1 + (1-\alpha)\frac{1-u}{u}\right]\left[1 + (1-\alpha)\frac{1-v)}{v}\right]$$

i.e., the above relation for $\varphi(t) = \ln[1 + (1-\alpha)(1-t)/t] = \ln[\{1 - \alpha(1-t)\}/t]$.

Archimedian copulas play an important role because they have interesting properties. For instance:

1. Probability density: $c(u,v) = -\varphi''(C(u,v))\,\varphi'(u)\varphi'(v)/\left[\varphi'(C(u,v))\right]^3$.
2. Kendall's tau: $\tau = 4\int_0^1 [\varphi(t)/\varphi'(t)]dt + 1$.
3. C has a singular component if and only if $\varphi(0)/\varphi'(0) \neq 0$. In this case

$$P[\varphi(U) + \varphi(V) = \varphi(0)] = -\frac{\varphi(0)}{\varphi'(0)}.$$

Table 6: *Basic copulas and some Archimedian families. Kendall's tau can not be given in closed form for Gumbel-Barnett and Joe.*

Copula	cdf	$\varphi(t)$	
Lower bound	$C^- = \max\{u+v-1, 0\}$	$(1-t)$	
Independence	$C^0 = uv$	$-\ln t$	
Upper bound	$C^+ = \min\{u, v\}$	Not archimedian	
Family	**cdf**	$\varphi(t)$	**Bounds**
Gumbel-Barnett	$FG\exp(-\theta \ln F \ln G)$ $0 < \theta \leq 1$	$\ln(1-\theta \ln t)$	$C_0 = C^0$ $C_1 < C^+$
Gumbel-Hougaard	$\exp(-[(-\ln F)^\theta + (-\ln G)^\theta]^{1/\theta})$ $1 \leq \theta < \infty, \quad \tau = 1 - 1/\theta$	$(-\ln t)^\theta$	$C_1 = C^0$ $C_\infty = C^+$
Joe	$1 - [\overline{F}^\theta + \overline{G}^\theta - \overline{F}^\theta \overline{G}^\theta]^{1/\theta}$ $1 \leq \theta < \infty, \overline{F} = 1 - F$	$-\ln[1-(1-t)^\theta]$	$C_1 = C^0$ $C_\infty = C^+$

Table 6 summarizes the Archimedian property for the three basic copulas and three Archimedian families. The Gumbel-Hougaard family can be obtained by compounding. The copula C_H is the only extreme-value distribution (i.e., C_H^n is also a cdf) which is Archimedian. The constant quantity is $\ln H(x,x)/\ln F(x)$ if $F = G$.

9.11 Shuffles of Min

Can complete dependence be very close to independence? Apparently not, as the opposite of stochastic independence between X and Y is the relation $Y = \varphi(X)$ (a.s.), where φ is a one-to-one function. When φ is monotonic non-decreasing, the distribution of (X, Y) is the upper bound $H^+ = \min\{F, G\}$, so $\rho_s = \tau = 1$.

Let us consider the related copula $C^+ = \min\{u, v\}$. A family of copulas, called shuffles of Min, has interesting properties and can be constructed from C^+. The support of this copula can be described informally by placing the mass of C^+ on $\mathbf{I}^2$, which is cut vertically into a finite number of strips. The strips are then shuffled with some of them

flipped around the vertical axes of symmetry and then reassembled to form the square again. A formal definition is given in Mikusinski *et al.* (1992).

If the copula of (X, Y) is a shuffle of Min, then it can be arbitrarily close to the independence copula $C^0 = uv$. It can be proved that, for any $\varepsilon > 0$, there exists a shuffle of Min C_ε such that $\sup_{u,v\in\mathbf{I}} |C_\varepsilon(u, v) - uv| < \varepsilon$. Statistically speaking, we may have a bivariate sample, where (x, y) are completely dependent, but being impossible to distinguish from independence. This family is even dense, that is, we may approximate any copula by a shuffle of Min.

10 Additional aspects

Here we consider more statistical and probabilistic concepts where the bivariate upper bound is also present.

10.1 Multivariate generation

Any bivariate cdf H can be generated by a copula C, i.e., $H = C(F, G)$, where F, G are univariate cdf's.

One is tempted to use multivariate marginals F, G of dimensions p and q and a bivariate copula C to construct $H = C(F, G)$. But, is H a cdf? As proved by Genest *et al.* (1995), the answer is no, except for the independence copula $C^0 = uv$. In particular, the upper bound is not useful for this purpose. For instance, if F_1, F_2, G are univariate cdf's for (X_1, X_2, Y) with $F = F_1F_2$, and we consider $H = \min\{F, G\}$, then $\min\{F_i, G\}$ is the distribution of $(X_i, Y), i = 1, 2$. Therefore, $F_1(X_1) = G(Y)$ and $F_2(X_2) = G(Y)$, which contradicts the independence of X_1 and X_2.

10.2 Distances between distributions

If X and Y have univariate cdf's F and G and joint cdf H, we can define a distance between X and Y (and between F and G) by using

$$d_\alpha(X, Y) = E_H|X - Y|^\alpha,$$

assuming that $E(X^\alpha)$ and $E(Y^\alpha)$ exist. For $\alpha > 1$ it can be proved (see Dall'Aglio, 1972) that the minimum of $d_\alpha(X, Y)$ when H has marginals F, G is obtained when $H^+ = \min\{F, G\}$. The case $\alpha = 2$ corresponds to the maximum correlation ρ^+ and was proved by Hoeffding (1940).

Several authors (J. Bass, S. Cambanis, R. L. Dobrushin, G. H. Hardy, C. L. Mallows, A. H. Tchen, S. S. Vallender, W. Whitt and others) have considered the extreme bounds

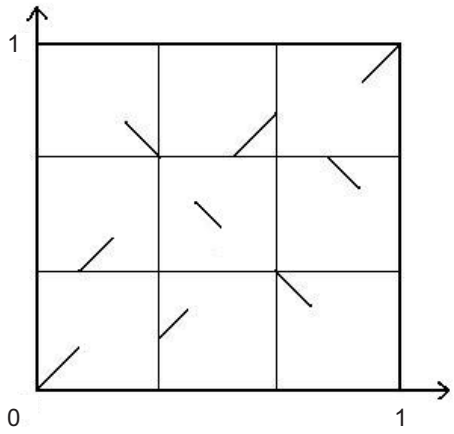

Figure 5: *A simple example of the support of a Shuffle of Min.*

for $E_H XY$, $E_H|X - Y|^\alpha$, $E_H f(|X - Y|)$ with $f'' > 0$, $\int \varphi(x,y)dH(x,y)$ where φ is superadditive (i.e., $\varphi(x,y) + \varphi(x',y') > \varphi(x',y) + \varphi(x,y')$ for all $x' > x$ and $y' > y$) and $E_H k(X,Y)$ when $k(x,y)$ is a quasi-monotone function (property quite similar to superadditivity). Thus, the supremum of $E_H XY$, $\int \varphi dH$ and $E_H k(X,Y)$ are achieved by the upper bound H^+. For instance, the supremum of $E_H k(X,Y)$ is

$$E_{H^+} k(X,Y) = \int_0^1 k(F^{-1}(u), G^{-1}(u))du.$$

See Tchen (1980).

10.3 Convergence in probability

The upper bound can also be applied to study the convergence in distribution and probability. Suppose that (X_n) is a sequence of r.v.'s with cdf's (F_n), which converges in probability to X with cdf F. Then it can be proved that this occurs if and only if

$$H_n(x,y) \to \min\{F(x), F(y)\},$$

where H_n is the joint cdf of (X_n, X). See Dall'Aglio (1972).

10.4 Lorenz curve and Gini coefficient

The Lorenz curve is a graphical representation of the distribution of a positive r. v. X. It is used to study the distribution of income. If X with cdf F ranges in (a,b), this curve is

defined by

$$L(y) = \frac{\int_a^y x dF(x)}{\int_a^b x dF(x)},$$

and can be given in terms of F

$$L(F) = \frac{\int_0^u F^{-1}(v)dv}{\int_0^1 F^{-1}(v)dv}.$$

The Lorenz curve is a convex curve in $\mathbf{I}^2$ under the diagonal from $(0,0)$ to $(1,1)$. Deviation from this diagonal indicates social inequality, see Figure 6. A global measure of inequality is the Gini coefficient $\mathcal{G}$, defined as twice the area between the curve and the diagonal:

$$\mathcal{G} = 1 - 2\int_0^1 L(F)dF.$$

For example, if X is Pareto with cdf $F(x) = 1 - (x/a)^c$ if $x > a$ then

$$L(F) = 1 - (1-F)^{1-1/c} \quad \text{and} \quad \mathcal{G} = 1/(2c-1).$$

The optimum social equality corresponds to $c = \infty$ and the maximum inequality to $c = 1$. However, for $c = 1$ the mean does not exist.

Assuming that X has finite variance $\sigma^2(X)$, Gini's coefficient can also be expressed as

$$\begin{aligned}
\mathcal{G} &= \int_a^b \int_a^b |x-y| dF(x)dF(y) \\
&= 2\int_a^b F(t)(1-F(t))dt \\
&= 4\mathrm{Cov}(X, F(X)) \\
&= \frac{2}{\sqrt{3}}\sigma(X)\mathrm{Cor}(X, F(X)).
\end{aligned}$$

But $\mathrm{Cor}(X, F(X))$ is the maximum Hoeffding correlation between X and U, where U is uniformly distributed. Thus, if $\sigma^2(X)$ exists, the maximum social inequality is $\mathcal{G} = 1/3$ and is attained when X is uniform, that is, when poor, middle and rich classes have the same proportions. Note that Pareto with $c = 2$ also gives $\mathcal{G} = 1/3$, but in this case $\sigma^2(X)$ does not exist.

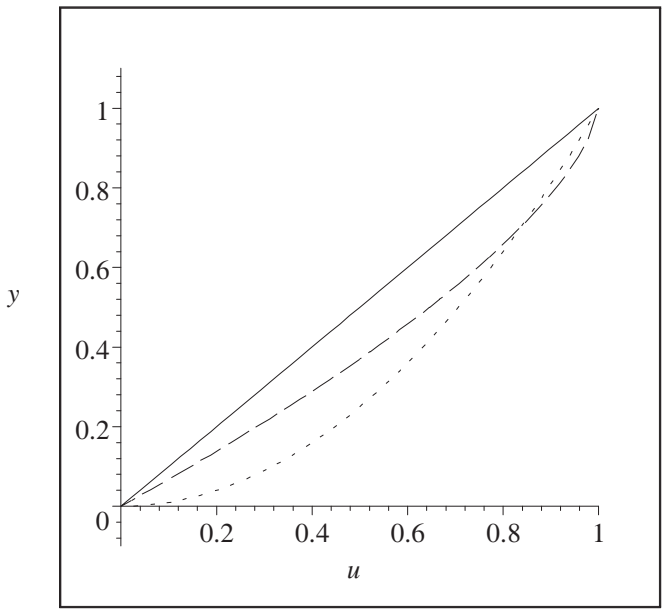

Figure 6: *Lorenz curve. Diagonal (solid line), Pareto with $c = 3$ (dash line) and uniform (dots line). The dots curve indicates maximum inequality (assuming finite variance) and this curve is related to the upper bound.*

10.5 Triangular norms and quasi-copulas

The theory of triangular norms (T-norms) is used in the study of associative functions, probabilistic metric spaces and fuzzy sets. See Schweizer and Sklar (1983), Alsina *et al.* (2006).

A T-norm T is a mapping from $\mathbf{I}^2$ into $\mathbf{I}$ such that $T(u, 1) = u$, $T(u, v) = T(v, u)$, $T(u_1, v_1) \leq T(u_2, v_2)$ whenever $u_1 \leq u_2$, $v_1 \leq v_2$, and $T(T(u, v), w) = T(u, T(v, w))$.

Examples of T-norms are $C^- = \max\{u + v - 1, 0\}$, $C^+ = \min\{u, v\}$, $C^0 = uv$ and Z defined by $Z(a, 1) = Z(1, a) = a$ and $Z(a, b) = 0$ otherwise. Note that Z is not a copula. It is readily proved that

$$Z < C^- < C^0 < C^+.$$

Thus the bivariate upper bound is the supremum of the partial ordered set of the T-norms.

A quasi-copula $Q(u, v)$ is a function $Q : \mathbf{I}^2 \to \mathbf{I}$ satisfying $Q(0, v) = Q(u, 0) = 0$, $Q(u, 1) = Q(1, u) = u$, Q is non-decreasing in each of its arguments, and Q satisfies the Lipschiptz's condition

$$|Q(u', v') - Q(u, v)| \leq |u' - v'| + |u - v|.$$

The quasi-copulas were introduced by Alsina *et al.* (1993) to study operations on univariate distributions not derivable from corresponding operations on random variables on the same probability space. For example, if X and Y are independent with cdf's F and G, the convolution $F * G(x) = \int F(x-y)dG(y)$ provides the cdf of $X + Y$. However, the geometric mean $\sqrt{FG}$ can not be the cdf of the random variable $K(X, Y)$, for a Borel-measurable function K (see Nelsen, 1999). Any copula is a quasi-copula and again the bivariate upper bound is the supremum of the partial ordered set of the quasi-copulas.

11 Bayes tests in contingency tables

We show here that the upper bound is also related to the test of comparing two proportions from a Bayesian perspective.

The problem of choosing intrinsic priors to perform an objective analysis is discussed in Casella and Moreno (2004). We choose a prior distribution which puts positive mass on the null hypothesis. This distribution depends on a positive parameter measuring dependence, which can be estimated via Pearson's contingency coefficient. Thus this test can be approached by the chi-square test and improved by obtaining the Bayes factor. Once again, the upper bound appears in this context.

Suppose that k_1, k_2 are binomial independent $B(n_1, p_1), B(n_2, p_2)$, respectively. We consider the test of hypothesis

$$H_0 : p_1 = p_2 \quad vs. \quad H_1 : p_1 \neq p_2. \tag{11}$$

Writing $k_i' = n_i - k_i, i = 1, 2$, the classic (asymptotic) approach is based on the chi-square statistic

$$\chi_1^2 = n\phi^2, \tag{12}$$

where $n = n_1 + n_2$ and

$$\phi^2 = (k_1k_2' - k_1'k_2)^2/[(k_1 + k_2)(k_1' + k_2')(k_1 + k_1')(k_2 + k_2')]$$

is the squared phi-coefficient.

Let us suppose that (p_1, p_2) is an observation of a random vector (P_1, P_2), with support $\mathbf{I}^2$, following one of the following copulas:

$$\begin{aligned} C_1(p_1, p_2) &= \theta_1 \min\{p_1, p_2\} + (1 - \theta_1)(p_1p_2), \quad 0 \leq \theta_1 \leq 1, \\ C_2(p_1, p_2) &= \min\{p_1, p_2\}^{\theta_2}(p_1p_2)^{1-\theta_2}, \quad 0 \leq \theta_2 \leq 1, \quad (p_1, p_2) \in \mathbf{I}^2. \end{aligned}$$

C_1 is the copula related to the regression family, see (8), and was implicit in Fréchet

(1951). Copula C_2, proposed by Cuadras and Augé (1981), is the survival copula of the Marshall-Olkin distribution. These copulas satisfy (see also Sections 9.4 and 9.5):

1. There is independence for $\theta_i = 0$ and functional dependence for $\theta_i = 1$, that is,

$$P_1 = P_2 \text{ (a.s.) if } \theta_i = 1.$$

2. The pdf's with respect to the measure $\nu = \mu^2 + \mu_1$ are

$$\begin{aligned} c_1(p_1, p_2) &= (1 - \theta_1) + \theta_1 I_{\{p_1 = p_2\}}, \\ c_2(p_1, p_2) &= (1 - \theta_2) \max\{p_1, p_2\}^{-\theta_2} + \theta_2 p_1^{(1-\theta_2)} I_{\{p_1 = p_2\}}, \end{aligned}$$

   where $I_{\{p_1=p_2\}}$ is the indicator function, μ^2 and μ_1 are the Lebesgue measures on $\mathbf{I}^2$ and the line $p_1 = p_2$, respectively. Thus:

$$\int_0^{p_1} \int_0^{p_2} c_i(u_1, u_2) d\nu = C_i(p_1, p_2), \quad i = 1, 2.$$

3. These distributions have a singular part:

$$P_{C_1}[P_1 = P_2] = \theta_1, \quad P_{C_2}[P_1 = P_2] = \frac{\theta_2}{2 - \theta_2}.$$

4. The parameter θ measures stochastic dependence. Actually θ is the correlation coefficient for C_1 (see (10) for $\varphi(x) = x$) and the maximum correlation for C_2 (Cuadras, 2002a):

$$\theta_1 = \mathrm{Cor}_{C_1}(P_1, P_2), \qquad \theta_2 = \max_{\psi_1, \psi_2} \mathrm{Cor}_{C_2}(\psi_1(P_1), \psi_2(P_2)).$$

With these prior distributions (11) can be expressed as

$$H_0 : \theta = 1 \quad vs. \quad H_1 : 0 \le \theta < 1, \tag{13}$$

where θ is either θ_1 or θ_2. Note that to accept H_0 is equivalent to say that the copula reaches the upper Fréchet-Hoeffding bound.

As it has been presented in Section 9, there are other parametric copulas reaching the upper bound $\min\{p_1, p_2\}$, i.e., the hypothesis H_0. However in most of these copulas the probability of having $p_1 = p_2$ is zero and the estimation of the dependence parameter is not available from the contingency table.

Inference on the parameter θ_2 is discussed in Ruiz-Rivas and Cuadras (1988) and Ocaña and Ruiz-Rivas (1990). However, if only the sufficient statistic (k_1, k_2) is

available, in view of the above property, we may take the sample correlation between indicators of the events in a 2×2 contingency table. The square of this correlation is just the squared phi-coefficient ϕ^2. Accordingly, we may estimate θ by ϕ and the classic approach for testing (13) is again by means of (12).

Testing (11) or (13) may be improved using the Bayesian perspective. The likelihood function is

$$L(k_1, k_2; p_1, p_2) = p_1^{k_1}(1-p_1)^{k'_1} p_2^{k_2}(1-p_2)^{k'_2}.$$

Under $H_0 : p_1 = p_2 = p$ this function reduces to

$$L(k_1, k_2; p_1, p_2) = p^{k_1+k_2}(1-p)^{k'_1+k'_2}.$$

The Bayes factor in testing (11), expressed as $(p_1, p_2) \in \omega$ *vs.* $(p_1, p_2) \in \Omega - \omega$, where θ is interpreted as another unknown parameter, is

$$B_i = \frac{\int_\omega L(k_1, k_2; p_1, p_2) dC_i(p_1, p_2)}{\int_{\Omega-\omega} L(k_1, k_2; p_1, p_2) dC_i(p_1, p_2)}, \quad i = 1, 2.$$

Thus we obtain for copulas C_1 and C_2

$$B_1 = \frac{\int_0^1 p^{k_1+k_2}(1-p)^{k'_1+k'_2} dp}{\int_0^1 \int_0^1 (1-\theta_1) p_1^{k_1}(1-p_1)^{k'_1} p_2^{k_2}(1-p_2)^{k'_2} dp_1 dp_2 d\theta_1},$$

$$B_2 = \frac{\int_0^1 p^{k_1+k_2}(1-p)^{k'_1+k'_2} dp}{\int_0^1 \int_0^1 \int_0^1 (1-\theta_2) p_1^{k_1}(1-p_1)^{k'_1} p_2^{k_2}(1-p_2)^{k'_2} \max\{p_1, p_2\}^{-\theta_2} dp_1 dp_2 d\theta_2}.$$

High and low values of B_1 and B_2 give evidence for H_0 and H_1, respectively.

Finally, we can approach the more general hypothesis

$$H_0 : p_2 = \phi(p_1) \quad vs. \quad H_1 : p_2 \neq \phi(p_1),$$

where ϕ is a monotonic function, by using the family (8) as prior distribution, which also puts positive mass to the null hypothesis.

Example 4 Table 7 is a 2×2 contingency table summarizing the results of comparing surgery with radiation therapy in treating cancer, and was used by Casella and Moreno (2004). For this table we obtain

$$\widehat{\theta} = 0.1208,\ \chi_1^2 = 0.599,\ B_1 = 5.982,\ B_2 = 5.754.$$

The Bayes factors B_1, B_2, as well as χ^2_1, give support to the null hypothesis (the proportions are equal).

Table 7: *Contingency table combining treatment and cancer.*

	Cancer controlled	Cancer not controlled	
Surgery	21	2	23
Radiation therapy	15	3	18
	36	5	41

References

Alsina, C., Frank, M. J. and Schweizer, B. (2006). *Associative Functions: Triangular Norms and Copulas*. World Scientific, Singapore.

Alsina, C., Nelsen, R. B., and Schweizer, B. (1993). On the characterization of a class of binary operations on distribution functions. *Statistics and Probability Letters*, 17, 75-89.

Benzécri, J. P. (1973). *L'Analyse des Données. I. La Taxinomie. II. L'Analyse des Correspondances*. Dunod, Paris.

Block, H. W. and Sampson, A. R. (1988). Conditionally ordered distributions. *Journal of Multivariate Analysis,* 27, 91-104.

Casella, G. and Moreno, E. (2004). Objective Bayesian analysis of contingency tables. *Technical Report*, 2002-023. Department of Statistics, University of Florida.

Cuadras, C. M. (1992). Probability distributions with given multivariate marginals and given dependence structure. *Journal of Multivariate Analysis,* 42, 51-66.

Cuadras, C. M. (1996). A distribution with given marginals and given regression curve. In *Distributions with Fixed Marginals and Related Topics* (Eds. L. Rüschendorf, B. Schweizer and D. Taylor), pp. 76-84, IMS Lecture Notes-Monograph Series, Vol. 28, Hayward.

Cuadras, C. M. (2002*a*). On the covariance between functions. *Journal of Multivariate Analysis,* 81, 19-27.

Cuadras, C. M. (2002*b*). Correspondence analysis and diagonal expansions in terms of distribution functions. *Journal of Statistical Planning and Inference,* 103, 137-150.

Cuadras, C. M. (2002*c*). Diagonal distributions via orthogonal expansions and tests of independence. In *Distributions with Given Marginals and Statistical Modelling,* (Eds. C. M. Cuadras, J. Fortiana and J. A. Rodriguez-Lallena), pp. 35-42, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Cuadras, C. M. (2005*a*). Continuous canonical correlation analysis. *Research Letters in Information and Mathematical Sciences*, 8, 97-103.

Cuadras, C. M. (2005*b*). First principal component characterization of a continuous random variable. In *Advances on Models, Characterizations and Applications* (Eds. N. Balakrishnan, I. Bairamov and O. Gebizlioglu), pp. 189-199, Chapman & Hall/CRC-Press, New York.

Cuadras, C. M. and Augé, J. (1981). A continuous general multivariate distribution and its properties. *Communications in Statistics-Theory and Methods,* A10, 339-353.

Cuadras, C. M. and Cuadras, D. (2002). Orthogonal expansions and distinction between logistic and normal. In *Goodness-of-fit Tests and Model Validity,* (Eds. C. Huber-Carol, N. Balakrishnan, M. S. Nikulin and M. Mesbah), pp. 327-339, Birkhaüser, Boston.

Cuadras, C. M. and Fortiana, J. (1994). Ascertaining the underlying distribution of a data set. In *Selected Topics on Stochastic Modelling,* (Eds. R. Gutierrez and M. J. Valderrama), pp. 223-230, World Scientific, Singapore.

Cuadras, C. M. and Fortiana, J. (1995). A continuous metric scaling solution for a random variable. *Journal of Multivariate Analysis,* 52, 1-14.

Cuadras, C. M., Fortiana, J. and Greenacre, M. J. (2000). Continuous extensions of matrix formulations in correspondence analysis, with applications to the FGM family of distributions. In *Innovations in Multivariate Statistical Analysis* (Eds. R. D. H. Heijmans, D. S. G. Pollock and A. Satorra.), pp. 101-116, Kluwer Ac. Publ., Dordrecht.

Cuadras, C. M., Fortiana, J. and Rodriguez-Lallena, J. A. (Eds.) (2002). *Distributions with Given Marginals and Statistical Modelling,* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Cuadras, C. M. and Lahlou, Y. (2000). Some orthogonal expansions for the logistic distribution. *Communications in Statistics-Theory and Methods,* 29, 2643-2663.

Cuadras, C. M., Cuadras, D. and Lahlou, Y. (2006). Principal directions for the general Pareto distribution. *Journal of Statistical Planning and Inference*, 136, 2572-2583.

Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposia Mathematica*, 9, 131-150.

del Barrio, E., Cuesta-Albertos, J. A. and Matrán, M. (2000). Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *TEST*, 9, 1-96.

Druet-Mari, D. and Kotz, S. (2001). *Correlation and Dependence*. Imperial College Press, London.

Fortiana, J. and Grané, A. (2002). A scale-free goodness-of-fit statistic for the exponential distribution based on maximum correlations. *Journal of Statistical Planning and Inference,* 108, 85-97.

Fortiana, J. and Grané, A. (2003). Goodness-of-fit tests based on maximum correlations and their orthogonal decompositions. *Journal of the Royal Statisitcal Society, Series B,* 65, 115-126.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges son données. *Annales de l'Université de Lyon, Série* 3, 14, 53-77.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74, 540-555.

Genest, C. and MacKay, J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician*, 40, 280-283.

Genest, C., Quesada-Molina, J. J. and Rodriguez-Lallena, J. A. (1995). De l'impossibilté de construire des lois a marges données a partir de copules. *Comptes Rendus Academie Sciences Paris*, 320, 723-726.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis.* Academic Press, London.

Hoeffding, W. (1940). Maszstabinvariante Korrelationstheorie. *Schriften des Mathematischen Institiuts und des Instituts für Angewandte Mathematik der Universitat Berlin,* 5, 179-233.

Hutchinson, T. P. and Lai, C. D. (1991). *The Engineering Statistician's Guide to Continuous Bivariate Distributions.* Rumsby Scientific Pub., Adelaide.

Joe, H. (1997). *Multivariate Models and Dependence Concepts,* Chapman and Hall, London.

Kimeldorf, G. and Sampson, A. (1975). Uniform representation of bivariate distributions. *Communications in Statistics,* 4, 617-627.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions*. Wiley, New York.

Mardia, K. V. (1970). *Families of Bivariate Distributions*. Griffin, London.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press, London.

Marshall, A. W., and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62, 30-44.

Marshall, A. W., Meza, J. C. and Olkin, I. (2001). Can data recognize the parent distribution? *Journal of Computational and Graphical Statistics,* 10, 555-580.

Mikusisnki, P., Sherwood, H., Taylor, M. D. (1992). Shuffles of Min. *Stochastica*, 13, 61-74.

Muliere, P. and Scarsini, M. (1987). Characterization of Marshall-Olkin type class of distributions. *Annals Institute Statistical Mathematics*, 39, 429-441.

Nelsen, R. B. (1986). Properties of a one-parameter family of bivariate distributions with given marginals. *Communications in Statistics-Theory and Methods,* 15, 3277-3285.

Nelsen, R. B. (1987). Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics-Theory and Methods,* 16, 199-208.

Nelsen, R. B. (1991). Copulas and Association. In *Advances in Probability Distributions with Given Marginals* (Eds. G. Dall'Aglio, S. Kotz and G. Salinetti), pp. 51-74, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Nelsen, R. B. (1999). *An Introduction to Copulas.* Springer, New York.

Nguyen, T. T. and Sampson, A. R. (1985). The geometry of certain fixed marginal probability distributions. *Linear Algebra and Its Applications*, 70, 73-87.

Ocaña, J. and Ruiz-Rivas, C. (1990). Computer generation and estimation in a one-parameter system of bivariate distributions with specified marginals. *Communications in Statistics-Simulation and Computation,* 19, 37-55

Ruiz-Rivas, C. and Cuadras, C. M. (1988). Inference properties of a one-parameter curved family of distributions with given marginals. *Journal of Multivariate Analysis,* 27, 447-456.

Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces*. North-Holland, New York.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *The Annals of Probability*, 8, 814-827.

Tiit, E. (1986). Random vectors with given arbitrary marginal and given correlation matrix. *Acta et Commentationes Universitatis Tartuensis*, 733, 14-39.