Brazilian Journal of Probability and Statistics (2005), 19, pp. 125–137.
©Associação Brasileira de Estatística

On grade transformation and its implications for copulas

Magdalena Niewiadomska-Bugaj[1] and **Teresa Kowalczyk**[2]
[1] *Western Michigan University*
[2] *Polish Academy of Sciences*

Abstract: As multivariate distributions with uniform one-dimensional margins, copulas provide very convenient models for studying dependence structure with tools that are scale-free. Each copula (n-copula) represents the whole class of continuous bivariate (multivariate) distributions from which it has been obtained when one-dimensional marginals were transformed by their cdf's. The similar property, however, does not hold when the original distributions are discrete, or mixed discrete-continuous. After the transformation by marginal cdf's the copula is not uniquely defined and consequently cannot be used for dependence studies analogously as in the continuous case. In this paper we will discuss grade transformation which, being an extension of probability integral transform, enables unique copula (or n-copula) representation of a multivariate distribution of any type for which cdf can be defined. We will also present how formulas of Kendall's τ, Spearman's ρ, and Gini correlation can be written for variables continuous or not using smoothed cdf's instead of original ones.

Key words: Copula, Gini correlation, grade transformation, Kendall's τ, Spearman's ρ.

1 Introduction

It is well known that when one-dimensional margins of a continuous multivariate distribution are transformed by their cdf's then one obtains a copula (n-copula) determined by a joint distribution in a unit square $[0,1]^2$ (or n-cube $[0,1]^n$) with margins having U[0, 1] distribution. It is also well known that such property does not hold in a discrete case. Therefore various attempts of "continuisation" of random variables were made over the years, with grade transformation being one of them. This approach is very general, applicable to all distributions for which the cdf can be defined. The so-called grade transformation, proposed by Szczesny in (1991), results in a continuous, uniformly distributed variables and therefore can be considered an extension of probability integral transform.

2 Kendall's τ, Spearman's ρ and Gini correlation as functions of smoothed cdf's.

Smoothed cumulative distribution function was introduced by Kowalczyk and Niewiadomska-Bugaj in (1998) as follows:

Definition 2.1 Let X, Y $(X_1, \ldots, X_n)$ be jointly distributed random variables with a joint cdf $H(x, y)$ $(H_n(x_1, \ldots, x_n))$ and marginal cdf's $F(x)$ and $G(y)$ $(F_i(x_i))$ respectively, and let $F(x-) = \lim_{t \to x, t<x} F(t)$, and similar limits for G and H, (H_n) be defined analogously. Then smoothed cdf functions $\tilde{F}, \tilde{H}, \tilde{H}_n$ are defined

$$\tilde{F}(x) \stackrel{\text{def}}{=} \frac{1}{2}[F(x) + F(x-)] \tag{2.1}$$

$$\tilde{H}(x, y) \stackrel{\text{def}}{=} \frac{1}{4}[H(x, y) + H(x-, y) + H(x, y-) + H(x-, y-)] \tag{2.2}$$

$$\begin{aligned} \tilde{H}_n(x_1, \ldots, x_n) \stackrel{\text{def}}{=} & \frac{1}{2^n}[H(x_1, x_2, \ldots, x_n) + H(x_1-, x_2, \ldots, x_n) \\ & + H(x_1, x_2-, \ldots, x_n) + H(x_1-, x_2-, \ldots, x_n) \\ & + \ldots + H(x_1-, x_2-, \ldots, x_n-)]. \end{aligned} \tag{2.3}$$

Definition (2.3) is the most general, with (2.2) and (2.1) being its special cases, and (2.2) implying (2.1). Notice that smoothed cdf's of non-continuous variables are no longer cdf's of any distribution since they are neither right nor left continuous. They, however, coincide with cdf's at points of their continuity. Using smoothed cdf's allows for formulas applicable for any type of variables.

Proposition 2.1 *Let X and Y be jointly distributed random variables with a joint cdf $H(x, y)$ and marginal cdf's $F(x)$ and $G(y)$ respectively. Then for smoothed cdf's $\tilde{F}(x), \tilde{G}(y), \tilde{H}(x, y)$ defined as in Definition 2.1 conditions (i) - (iii) are satisfied.*
(i) $E(\tilde{F}(X)) = \frac{1}{2} \leq E(F(X)), Var(\tilde{F}(X)) = \frac{1}{12} - \frac{1}{12}\int [F(x) - F(x-)]^2 dF(x)$,
(ii) $E(\tilde{F}(X)\tilde{G}(Y)) \leq E(F(X)G(Y))$,
(iii) $E(\tilde{H}(X, Y)) \leq E(H(X, Y))$
with equalities holding in the case of continuous distributions.

Proof. In general $EF(X-) = 0.5 - 0.5\int [F(x) - F(x-)]dF(x) \leq 0.5 \leq 0.5 + \int [F(x) - F(x-)]dF(x) = E(F(X))$ what implies (i). The rest of the proof is also immediate here, since by its definition smoothed cdf never exceeds an original cdf, and also if the original cdf is continuous, then the smoothed cdf and original cdf coincide.

In the case of continuous variables, Spearman's rho, Kendalls tau and Gini correlation (also known as Gini association index, see S. Das Gupta, 1999)) can be expressed in terms of the cdf's as

$$\tau = 4E(H(X, Y)) - 1 \tag{2.4}$$

$$\rho^* = 12\text{Cov}(F(X), G(Y)) \tag{2.5}$$

$$\Gamma = \frac{\mathrm{Cov}(Y, F(X))}{\mathrm{Cov}(Y, G(Y))} \tag{2.6}$$

The following theorem provides formulas for these indices as functions of smoothed cdf's, applicable for continuous as well as discrete, and mixed variables.

Proposition 2.2 *Regardless of the type of the jointly distributed variables X and Y, Kendall's τ, Spearman's ρ and Gini correlation Γ can be obtaind as*

$$\tau = 4E(\tilde{H}(X,Y)) - 1 \tag{2.7}$$

$$\rho^* = 12\,Cov(\tilde{F}(X), \tilde{G}(Y)) \tag{2.8}$$

$$\Gamma = \frac{Cov(Y, \tilde{F}(X))}{Cov(Y, \tilde{G}(Y))} \tag{2.9}$$

Proof. Let distribution of (X, Y) be discrete.
(i) Let $(X_1, Y_1), (X_2, Y_2)$ be two independent realizations of (X, Y). Then

$$\begin{aligned}
\tau &= \Pr(\text{concordance}) - \Pr(\text{discordance}) \\
&= P(X_1 < X_2, Y_1 < Y_2) + P(X_1 > X_2, Y_1 > Y_2) - P(X_1 < X_2, Y_1 > Y_2) \\
&\quad -P(X_1 > X_2, Y_1 < Y_2) \\
&= E[H(X-, Y-) + 1 - F(X) - G(Y) + H(X, Y)] \\
&\quad -E[F(X-) - H(X-, Y) + G(Y-) - H(X, Y-)] \\
&= E[H(X-, Y-) + H(X, Y-) + H(X-, Y) + H(X, Y)] \\
&\quad -E[F(X) + F(X-)] - E[G(Y) + G(Y-)] + 1 \\
&= 4E[\tilde{H}(X, Y)] - 2E[\tilde{F}(X)] - 2E[\tilde{G}(Y)] + 1 \\
&= 4E[\tilde{H}(X, Y)] - 1.
\end{aligned} \tag{2.10}$$

(ii) Spearman's ρ^* is defined as 3[Pr(concordance) - Pr(discordance)] of two pairs of random variables with identical corresponding marginal distributions, but in the second pair, marginal variables are independent.

$$\begin{aligned}
\rho^* &= 3[\Pr(\text{concordance}) - \Pr(\text{discordance})] \\
&= 3[P(X_1 < X_2, Y_1 < Y_2) + P(X_1 > X_2, Y_1 > Y_2) - P(X_1 < X_2, Y_1 > Y_2) \\
&\quad -P(X_1 > X_2, Y_1 < Y_2)] \\
&= 3E_{(X,Y)}[F(X-)G(Y-) + (1 - F(X))(1 - G(Y)) - F(X-)(1 - G(Y)) \\
&\quad -G(Y-)(1 - F(X)] \\
&= 3E_{(X,Y)}[(F(X-) + F(X))(G(Y-) + G(Y)) - (F(X-) + F(X)) \\
&\quad -(G(Y-) + G(Y)) + 1] \\
&= 3E_{(X,Y)}[4\tilde{F}(X)\tilde{G}(Y) - 2\tilde{F}(X) - 2\tilde{G}(Y) + 1] \\
&= 12E_{(X,Y)}[\tilde{F}(X)\tilde{G}(Y) - 0.25] \\
&= 12\mathrm{Cov}(\tilde{F}(X), \tilde{G}(Y)).
\end{aligned} \tag{2.11}$$

(iii) Let us now assume that (X, Y) are jointly distributed and that $P(X = x_i, Y = Y_j) = p_{ij}, P(X = x_i) = p_{i+}, P(Y = y_j) = p_{+j}, i = 1, \ldots, m, j = 1, \ldots, n$, where $x_1 < \cdots < x_m, y_1 < \cdots < y_n$. Then Lorenz curve for variable Y is defined as

$$L_Y(p) = \frac{\sum_{k=0}^{j-1} p_{+k} y_k + (p - \sum_{k=0}^{j-1} p_{+k}) y_j}{E(Y)} \tag{2.12}$$

where $\sum_{k=0}^{j-1} p_{+k} \leq p \leq \sum_{k=0}^{j} p_{+k}$, $p_{+0} = y_0 = 0$ (see Schechtman and Yitzaki, 1987, Das Gupta, 1999). The so-called Lorenz area between the curve $L_Y(p)$ given by (2.12) and diagonal connecting points $(0, 0)$ and $(1, 1)$ in $[0, 1]^2$ can be obtained as

$$\begin{aligned}
& \frac{1}{2} - \sum_{j=1}^{n} p_{+j} \times \frac{\sum_{k=0}^{j-1} p_{+k} y_k + \sum_{k=0}^{j} p_{+k} y_k}{2E(Y)} \\
= & \frac{E(Y) - \sum_{j=1}^{n} p_{+j} (\sum_{k=0}^{j-1} p_{+k} y_k + \sum_{k=0}^{j} p_{+k} y_k)}{2E(Y)} \\
= & \frac{E(Y) - (\sum_{j=1}^{n} \sum_{k=0}^{j-1} p_{+j} p_{+k} y_k + \sum_{j=1}^{n} \sum_{k=0}^{j} p_{+j} p_{+k} y_k)}{2E(Y)} \\
= & \frac{E(Y) - (\sum_{k=1}^{n-1} p_{+k} y_k \sum_{j=k+1}^{n} p_{+j} + \sum_{k=1}^{n-1} p_{+k} y_k \sum_{j=k}^{n} p_{+j})}{2E(Y)} \\
= & \frac{E(Y) - \{\sum_{k=1}^{n} p_{+k} y_k [1 - G(y_k)] + \sum_{k=1}^{n} p_{+k} y_k [1 - G(y_{k-1}]\}}{2E(Y)} \\
= & \frac{E(Y) - \{E(Y) - E[G(Y)Y] + E(Y) - E[G(Y-)Y]\}}{2E(Y)} \\
= & \frac{2E[\tilde{G}(Y)Y] - E(Y)}{2E(Y)} \\
= & \frac{2\{E[Y\tilde{G}(Y)] - E(Y)E[\tilde{G}(Y)]\}}{2E(Y)} = \frac{\mathrm{Cov}[Y, \tilde{G}(Y)]}{E(Y)}
\end{aligned} \tag{2.13}$$

The numerator of the Gini correlation is the Lorenz area obtained from so-called pseudo Lorenz curve given by the following formula

$$L_{Y|X}(p) = \frac{\sum_{k=0}^{i-1} p_{k+} E(Y|x_i) + (p - \sum_{k=0}^{i-1} p_{k+}) E(Y|x_i)}{E(Y)}$$

and where $\sum_{k=0}^{i-1} p_{k+} \leq p \leq \sum_{k=0}^{i} p_{k+}$. By the similar algebra as above, we obtain the corresponding Lorenz area equal

$$\begin{aligned}
& \frac{1}{2} - \sum_{i=1}^{m} p_{i+} \times \frac{\sum_{k=0}^{i-1} p_{k+} E(Y|x_k) + \sum_{k=0}^{i} p_{k+} E(Y|x_k)}{2E(Y)} \\
= & \frac{E(Y) - [\sum_{k=1}^{m} p_{k+} E(Y|x_k)(1 - F(x_k)) + \sum_{k=1}^{m} p_{k+} E(Y|x_k)(1 - F(x_{k-1}))]}{2E(Y)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{E(Y) - 2E(E(Y|X)) + \sum_{k=1}^{m} 2\tilde{F}(x_k)E(Y|x_k)p_{k+}}{2E(Y)} \\
&= \frac{2\sum_{k=1}^{m} \tilde{F}(x_k)p_{k+}(\sum_{j=1}^{n} y_j p_{kj}/p_{k+}) - E(Y)}{2E(Y)} \\
&= \frac{\sum_{k=1}^{m}\sum_{j=1}^{n} \tilde{F}(x_k)y_j - E(Y)E[\tilde{F}(X)]}{E(Y)} = \frac{\mathrm{Cov}[Y, \tilde{F}(X)]}{E(Y)} \qquad (2.14)
\end{aligned}
$$

Gini correlation, Γ, defined as a ratio of two Lorenz areas given by (2.14) and (2.13) respectively, equals

$$\Gamma = \frac{\mathrm{Cov}(Y, \tilde{F}(X))}{\mathrm{Cov}(Y, \tilde{G}(Y))}$$

what completes the proof.

3 Grade transformation and analog of Sklar's theorem

Let X be a random variable with a cdf $F(x)$, Then the cumulative distribution of a variable X transformed by its cdf can be expressed as

$$
\begin{aligned}
P(F(X) \le u) &= \int I(F(x), u)dF(x) \qquad (3.1) \\
&\le \int I(x, F^{-1}(u))dF(x) \qquad (3.2)
\end{aligned}
$$

where

$$I(a, b) = \begin{cases} 1 & \text{if } a \le b \\ 0 & \text{otherwise,} \end{cases}$$

$F^{-1}(t) = \inf\{x : F(x) \ge t\} = \sup\{x : F(x) < t\}$, and equality holds for points of continuity of function F.

Let us now substitute randomizing transformation F^* defined as (see Szczesny, 1991)

$$F^*(u, x) = \begin{cases} 1 & \text{if } F(x) \le u \\ (u - F(x-))/(F(x) - F(x-) & \text{if } F(x-) \le u < F(x) \\ 0 & \text{if } F(x-) > u \end{cases} \qquad (3.3)$$

for identity function I.

Unique, regardless of the distribution type of original variable, transformation of a random variable by its cdf is called grade transformation. If now each univariate variable in the multivariate distribution is transformed by a grade transformation generated by its own cdf, one obtains a joint continuous distribution

$$H_n^*(v_1, \ldots, v_n) = \int \cdots \int F_1^*(v_1, x_1) \cdots F_n^*(v_n, x_n) dH_n(x_1, \ldots, x_n)$$

that in the bivariate case has cdf

$$H^*(v_1, v_2) = \int\int F_1^*(v_1, x_1) F_2^*(v_2, x_2) dH(x_1, x_2) \tag{3.4}$$

Following examples illustrate how the grade density h^* is obtained:

Example 3.1 (i) Let X and Y be jointly distributed discrete variables such that $P(X = x_i, Y = y_j) = p_{ij}, i = 1, 2, j = 1, 2$ and $p_{11} + p_{12} + p_{21} + p_{22} = 1$. Then the joint grade density (the joint density of variables $F^*(X)$ and $G^*(Y)$) is uniform, but most likely different in each of the rectangles in the unit square $[0, 1]^2$ (see Figure 1). One can check by simple computations that the margins have $U[0, 1]$ distributions.

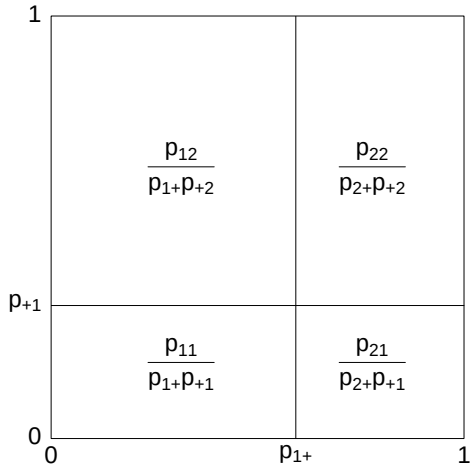

Figure 1 *Grade density obtained from a discrete distribution.*

(ii) Let now X and Y be jointly distributed discrete variables such that $P(X = x_1, Y = y_1) = p_{11}, P(X = x_2, Y = y_2) = p_{22}, P(X = x_1, Y = y_3) = p_{13}, P(X = x_2, Y = y_3) = p_{23}$ and $p_{11} + p_{22} + p_{13} + p_{23} = 1$.This distribution is determined on the Cartesian product $\{x_1, x_2\} \times \{y_1, y_2, y_3\}$ with at least two probabilities equal zero ($p_{12} = p_{21} = 0$). The grade density will be obtained similarly as in part (i) above but will be equal 0 in all rectangles related to zero probabilities.

(iii) Let us now consider random variables X and Y such that $P(X = x_0, Y = y_0) = p > 0$, and otherwise the distribution is continuous with density $h(x, y)$ over a certain region D ($\int\int_D h(x, y)dxdy = 1 - p$). Figure 2 shows grade density in a unit square $[0, 1]^2$. Grade density $h^*(u, v)$ equals $\frac{p}{p \times p} = \frac{1}{p}$ for $F(x_0-) \leq u \leq F(x_0), G(y_0-) \leq v \leq G(y_0)$ (black area), it equals 0 when $u < F(x_0), G(y_0-) \leq$

$v \leq G(y_0)$ or $u > F(x_0), G(y_0-) \leq v \leq G(y_0)$ (white areas), and finally it equals

$$\begin{array}{rll} h(F^{-1}(u), G^{-1}(v)) & \text{when} & 0 \leq u < F(x_0-), 0 \leq v < G(y_0-), \\ h(F^{-1}(u), G^{-1}(v-p)) & \text{when} & 0 \leq u < F(x_0-), G(y_0) < v \leq 1, \\ h(F^{-1}(u-p), G^{-1}(v)) & \text{when} & F(x_0) < u \leq 1, 0 \leq v < G(y_0-), \\ \text{and } h(F^{-1}(u), G^{-1}(v)) & \text{when} & F(x_0) < u \leq 1, G(y_0) < v \leq 1 \end{array}$$

(grey areas). In the case of more than one point (x, y) with a positive probability mass, more complex structure would be obtained, with more uniform distribution squares, and corresponding rectangles with the grade density equal zero.

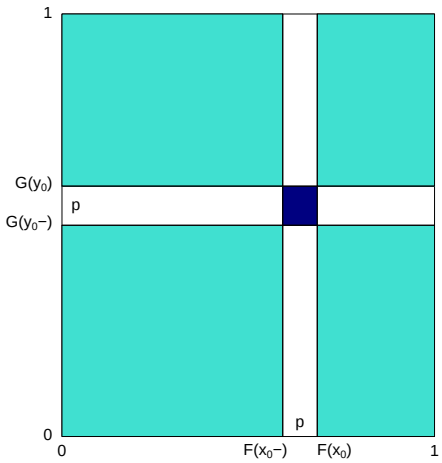

Figure 2 *Grade density of a mixed discrete-continuous distribution.*

For any cdf $H(x, y)$, formula (3.4) uniquely defines grade cdf H^* with $U[0, 1]$ margins. At points of continuity of marginal cdf's F and G

$$H^*(F(x), G(y)) = H(x, y). \tag{3.5}$$

The following proposition is an analog to Sklar's theorem (Sklar, 1959).

Proposition 3.1 *Let H be a joint cdf with margins F and G. Then the grade cdf H^*, uniquely determined by (3.4), is a copula. Moreover, for the points of continuity of F and G, the resulting unique copula coincides with that given by Sklar (1959) as*

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

Proof. The proposition is an immediate consequence of (3.3), (3.4), and (3.5).

It should be mentioned here that grade transformation can be seen as analogous to a simple interpolation used by Schweitzer and Sklar (1974), and later by Nelsen (1999), as an example of a possible extention of any subcopula to a copula, when marginal distributions are not continuous. Due to its interesting properties, including for example invariance of Kendall's τ and Spearman's ρ, the grade transformation is playing a special role in statistics enabling unified approach to analysis of continuous, discrete or categorical data.

4 Invariance under grade transformation

Several important properties of bivariate distributions are invariant under grade transformation. Some of the important ones are stated in the Proposition 4.1 below.

Proposition 4.1 *Let X and Y be jointly distributed random variables with a joint cdf $H(x,y)$ and marginal cdf's $F(x)$ and $G(y)$ respectively. Then*

(i) Spearman's ρ^* is invariant under grade transformation,

$$\rho^*(X,Y) = 12Cov(\tilde{F}(X), \tilde{G}(Y)) = 12Cov(F^*(X), G^*(Y)) = \rho^*(X^*, Y^*) \tag{4.1}$$

(ii) Kendall's τ is invariant under grade transformation,

$$\tau(X,Y) = 4E(\tilde{H}(X,Y)) - 1 = 4E(H^*(X,Y)) - 1 = \tau(X^*, Y^*). \tag{4.2}$$

(iii) TP_2 property is invariant under grade transformation

(iv) If additionally variables X_1, Y_1 are jointly distributed with the joint cdf $H_1(x,y)$ such that marginal cdf's $F_1(x) = F(x)$ and $G_1(y) = G(y)$, then concordance ordering of H and H_1 implies concordance ordering of copulas H^* and H_1^* obtained by their grade transformation (3.3).

$$H(x,y) \leq H_1(x,y) \Rightarrow H^*(u,v) \leq H_1^*(u,v)$$

Proof: (i) and (ii) are immediate consequences of Lemma 4.2 below. We include it without proof since it requires tedious, elementary algebra.

Lemma 4.2 *(a)* $E[F^*(X)] = E[\tilde{F}(X)] = 0.5$,
(b) $E[F^*(X)G^*(Y)] = E[\tilde{F}(X)\tilde{G}(Y)]$,
(c) $E[H^*(X,Y)] = E[\tilde{H}(X,Y)]$.

(iii) The definition of property TP_2 (total positivity of order two) given by Kimeldorf and Sampson (1987, 1989) requires that

$$\begin{aligned} &P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2)P(x_3 \leq X \leq x_4, y_3 \leq Y \leq y_4) \\ &-P(x_1 \leq X \leq x_2, y_3 \leq Y \leq y_4)P(x_3 \leq X \leq x_4, y_1 \leq Y \leq y_2) \geq 0 \end{aligned} \tag{4.3}$$

for any $x_1 \le x_2 \le x_3 \le x_4$ and $y_1 \le y_2 \le y_3 \le y_4$. If u_1, u_2, u_3, u_4 and v_1, v_2, v_3, v_4 are such that there exist x_1, x_2, x_3, x_4 and y_1, y_2, y_3, y_4 such that $u_i = F(x_i), v_i = G(y_i)$, where (x_i, y_j) are points of continuity of H for $i = 1, \ldots, 4, j = 1$, then TP_2 property is preserved under grade transformation since

$$P(u_i \le X^* \le u_{i+1}, v_j \le Y^* \le v_{j+1}) = P(x_i \le X \le x_{i+1}, y_j \le Y \le y_{j+1})$$

for $i = 1, 3, j = 1, 3$. If the condition above is not satisfied then there exists at least one pair x_0 (or y_0) and u (v) such that $F(x_0-) < u \le F(x_0)$ ($G(y_0-) < v \le G(y_0)$). Without any loss of generality, let us assume that $P(X = x_0) > 0$ and that there is only one such value in the distribution of (X, Y). We also assume that u_i, v_i, $i = 1, \ldots, 4$ are selected in such a way, that $F(x_0-) < u_2 \le F(x_0)$ (so $x_2 = F^{-1}(u_2) = x_0$). Then grade density

$$h^*(u, v) = w(v) \tag{4.4}$$

for any u, $F(x_0-) \le u \le F(x_0), 0 \le v \le 1$, and the form of function w will depend on the distribution of probability $P(X = x_0)$ along the line $X = x_0$.

Consequently, for

$$\alpha = \frac{u - F(x_0-)}{F(x_0) - F(x_0-)}$$

and because any value $x \ne x_0$ is a point of continuity, we have

$$\begin{aligned} P(u_1 \le X^* \le u_2, v_1 \le Y^* \le v_2) &= P(u_1 \le X^* \le F(x_0-), v_1 \le Y^* \le v_2) \\ &+ \alpha P(F(x_0-) < X^* \le F(x_0), v_1 \le Y^* \le v_2) \end{aligned}$$

and

$$\begin{aligned} P(u_1 \le X^* \le u_2, v_3 \le Y^* \le v_4) &= P(u_1 \le X^* \le F(x_0-), v_3 \le Y^* \le v_4) \\ &+ \alpha P(F(x_0-) < X^* \le F(x_0), v_3 \le Y^* \le v_4) \end{aligned}$$

and the left hand side of the condition (4.3) becomes

$$\begin{aligned} &P(u_1 \le X^* \le F(x_0-), v_1 \le Y^* \le v_2) P(x_3 \le X \le x_4, y_3 \le Y \le y_4) \\ &-P(u_1 \le X^* \le F(x_0-), v_3 \le Y^* \le v_4) P(x_3 \le X \le x_4, y_1 \le Y \le y_2) \\ &+\alpha[P(F(x_0-) < X^* \le F(x_0), v_1 \le Y^* \le v_2) P(x_3 \le X \le x_4, y_3 \le Y \le y_4) \\ &-P(F(x_0-) < X^* \le F(x_0), v_3 \le Y^* \le v_4) P(x_3 \le X \le x_4, y_1 \le Y \le y_2)] \\ &= P(x_1 \le X < x_2 = x_0, y_1 \le Y \le y_2) P(x_3 \le X \le x_4, y_3 \le Y \le y_4) \\ &-P(x_1 \le X < x_2 = x_0, y_3 \le Y \le y_4) P(x_3 \le X \le x_4, y_1 \le Y \le y_2) \\ &+\alpha[P(x_0 \le X \le x_0, y_1 \le Y \le y_2) P(x_3 \le X \le x_4, y_3 \le Y \le y_4) \\ &-P(x_0 \le X \le x_0, y_3 \le Y \le y_4) P(x_3 \le X \le x_4, y_1 \le Y \le y_2)] \end{aligned} \tag{4.5}$$

and it is nonnegative because of TP_2 property of (X, Y) and passing to the limit.

In the case of more than one value of X that has positive probability analogous resoning should be applied to each of them. If additionally there exists such y_0,

that $P(X = x_0, Y = y_0) > 0$ then the density $w(v)$ in (4.4) will become a constant independent on v.

(iv) If two joint continuous distributions with cdf's $H(x, y)$ and $H_1(x, y)$ and identical corresponding marginal cdf's $(F(x) = F_1(x), G(y) = G_1(y)$ for any $x, y)$ have concordance ordering property, that is for any x, y

$$H(x, y) \leq H_1(x, y)$$

then for copulas H^* and H_1^* resulting from a grade transformation (3.3) also have concordance ordering property, since

$$H^*(u, v) = H(x, y) \leq H_1(x, y) = H_1^*(u, v)$$

where $x = F^{-1}(u)$ and $y = G^{-1}(v)$.

If now two joint distributions are either both mixed discrete-continuous or both discrete (same marginal distributions make it impossible for distributions H and H_1 to be of different types) then the joint density for their corresponding copulas H^* and H_1^* is defined in the same rectangles, and consequently H^* and H_1^* can be expressed as linear combinations (with the same nonnegative weights) of cdf's H and H_1 respectively, and the concordance ordering will be preserved. To be more specific, let us consider the most general case, and take (u, v) such that there exist x_0 and y_0 for which $F(x_0-) < u < F(x_0)$ and $G(y_0-) < v < G(y_0)$. Next, let us define α and β as

$$\alpha = \frac{u - F(x_0-)}{F(x_0) - F(x_0-)} \text{ and } \beta = \frac{v - G(y_0-)}{G(y_0) - G(y_0-)}.$$

Then

$$\begin{aligned} H^*(u, v) &= H^*[F(x_0-), G(y_0-)] \\ &\quad +\alpha\{H^*[F(x_0), G(y_0-)] - H^*[F(x_0-), G(y_0-)]\} \\ &\quad +\beta\{H^*[F(x_0-), G(y_0)] - H^*[F(x_0-), G(y_0-)]\} \\ &\quad +\alpha\beta\{H^*[F(x_0), G(y_0)] - H^*[F(x_0), G(y_0-)] \\ &\quad -H^*[F(x_0-), G(y_0)] + H^*[F(x_0-), G(y_0-)]\} \\ &= (1-\alpha)(1-\beta)H^*[F(x_0-), G(y_0-)] + \beta(1-\alpha)H^*[F(x_0-), G(y_0)] \\ &\quad +\alpha(1-\beta)H^*[F(x_0), G(y_0-)] + \alpha\beta H^*[F(x_0), G(y_0)] \end{aligned} \tag{4.6}$$

If now the distributions of (X, Y), and (X_1, Y_1) are discrete, then there exist x_a and y_a such that $F(x_0-) = F(x_a)$, where x_a is the largest value of X that is less than x_0 and $P(X = x_a) > 0$, similarly $G(y_0-) = G(y_a)$. Then H^* given by (4.6) equals

$$\begin{aligned} H^*(u, v) &= (1-\alpha)(1-\beta)H^*[F(x_a), G(y_a)] + \beta(1-\alpha)H^*[F(x_a), G(y_0)] \\ &\quad +\alpha(1-\beta)H^*[F(x_0), G(y_a)] + \alpha\beta H^*[F(x_0), G(y_0)] \qquad (4.7) \\ &= (1-\alpha)(1-\beta)H[x_a, y_a] + \beta(1-\alpha)H[x_a, y_0] \\ &\quad +\alpha(1-\beta)H[x_0, y_a] + \alpha\beta H[x_0, y_0] \\ &\leq (1-\alpha)(1-\beta)H_1[x_a, y_a] + \beta(1-\alpha)H_1[x_a, y_0] \\ &\quad +\alpha(1-\beta)H_1[x_0, y_a] + \alpha\beta H_1[x_0, y_0] = H_1^*(u, v) \end{aligned}$$

and concordane ordering of (X, Y) and (X_1, Y_1), implies concordance ordering of (X^*, Y^*) and (X_1^*, Y_1^*), so concordance ordering property is preserved under grade transformation.

If, however, the joint distributions are mixed discrete-continuous, and $P(X = x_0, Y = y_0) > 0$, but in the neighborhood of the point (x_0, y_0) the distribution is continuous, then

$$\begin{aligned} H^*[F(x_0), G(y_0-)] &= H^*[F(x_0-), G(y_0-)], \\ H^*[F(x_0-), G(y_0)] &= H^*[F(x_0-), G(y_0-)] \end{aligned}$$

and formula (4.6) becomes

$$\begin{aligned} H^*(u, v) &= (1-\alpha\beta)H^*[F(x_0-), G(y_0-)] + \alpha\beta H^*[F(x_0), G(y_0)] \\ &= (1-\alpha\beta)lim_{x\to x_0(x<x_0), y\to y_0(y<y_0)} H(x, y) + \alpha\beta H(x_0, y_0) \quad (4.8) \end{aligned}$$

and consequently the concordance ordering is also preserved, what completes the proof.

After we presented the idea and basic properties of the grade transformation, we would like to comment on relations between the cdf (F), smoothed cdf $(\tilde{F})$, and grade transformation (F^*).

In the process of discretizing random variables, values from a certain range are usually represented by a single value with a probability mass equal to cumulated probability of values it substitutes. Such representing value can be selected several ways, sometimes it is a midpoint, but often it is a "median" value for a given range.

In the process of continuisation a probability mass is "spread" in the neighborhood of its corresponding value. How big is the neighborhood, and in what fashion the probability is distributed generally depends on the purpose the continuisation serves. One possible approach is to distribute the probability mass uniformly within $[x_0 - a, x_0 + a]$ with the associated value x_0 in the middle. If now the width of the interval decreases ($[x_0 - a_n, x_0 + a_n], a_n \to 0$ as $n \to \infty$) then as a limit of corresponding cdf's one obtains a smoothed cdf $(\tilde{F})$ not a cdf of the original distribution (F). Because of the "smoothed" cdf being a limiting case it is now clear why the formulas that include continuous transformations of cdf's are identical in the discrete and continuous cases when each cdf is substituted by its smoothed version, as in formulas in Proposition 4.1 (i) and (ii).

Grade transformation, on the other hand, transforms variables by their cdf's, uniformly stretching probability mass in points of discontinuity. If the smoothed cdf $(\tilde{F})$ is used, then the probability mass $P(X = x_0)$ is uniformly distributed around a $\tilde{F}(x_0)$ value, i.e., within the interval $[\tilde{F}(x_0-), \tilde{F}(x_0+)]$ which is identical with the interval $[F(x_0-), F(x_0)]$.

5 Final remarks

Few comments are in order here. First, the idea and attempts of smoothing cumulative distribution function of a discrete random variable are not new. For

example, the so-called mid P-value used in the analysis of categorical data, recommended by Lancaster in (1961), was nothing else but using $\tilde{F}$ instead of F in computations, although the idea of $\tilde{F}$ was not introduced explicitly.

Another comment we would like to make is that although studying the dependence structure of multivariate distributions by studying corresponding copulas obtained by the grade transformation is not the only possibility, it is a very general and convenient approach, applicable to distributions of any type for which the cdf can be defined. There is no restriction on the variable type, like for example requiring random variable to take values from the set of natural numbers, or for the set of possible values to be finite (Tchen, 1980).

Finally, even though it is not the topic of this paper, it should be mentioned that grade transformations enable so-called grade data analysis where behavior and relationship of jointly distributed variables is studied and exemplified by variety of curves. Several of the concepts related to dependence are illustrated by curves and surfaces in $[0,1]^2$ (or in general in n-cube $[0,1]^n$) which makes them easier to compare and interpret. For detailed information on grade methods of data analysis see Kowalczyk et al. (2004).

(Received March, 2004. Accepted June, 2004.)

References

Das Gupta, S. (1999). Gini association and pseudoLorenz cure. *Commun. Statist. Theor. Meth.*, **28**, 2181-2199.

Kimeldorf, G. and Sampson, A. (1987). Positive dependence orderings. *Ann. Inst. Statist. Math.*, **39**, 113-128.

Kimeldorf, G. and Sampson, A. (1989). A framework for positive dependence. *Ann. Inst. Statist. Math.*, **41**, 31-45.

Kowalczyk, T. and Niewiadomska-Bugaj, M. (1998). Grade correspondence analysis based on Kendall's tau. In *Proceedings of the IFCS-98 Conference*, Rome, Italy.

Kowalczyk, T., Pleszczynska, E. and Ruland, F. (eds) (2004). *Grade Models and Methods for Data Analysis.* Studies in Fuzziness and Soft Computing, Springer-Verlag.

Lancaster, H. O. (1961). Significance tests in discrete distributions. *J. Amer. Statist. Assoc.*, **56**, 223-234.

Nelsen, R. (1999). *An Introduction to Copulas.* New York: Springer-Verlag.

Schechtman, E. and Yitzaki, S. (1987). A measure of association based on Gini's mean difference. *Commun. Statist. Theor. Meth.*, **16**, 207-231.

Schweizer, B. and Sklar, A. (1974). Operations on distributions functions not derivable from operations on random variables. *Studia Math.*, **52**, 43-52.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Uni. Paris*, **8**, 229-231.

Szczesny, W. (1991). On the Performance of a Discriminant Function. *Journal of Classification*, **8**, 201-215.

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

Magdalena Niewiadomska-Bugaj
Department of Statistics
Western Michigan University
USA
E-mail: m.bugaj@wmich.edu

Teresa Kowalczyk
Institute of Computer Science
Polish Academy of Sciences
Warsaw, Poland
E-mail: tkow@ipipan.waw.pl