*Distributions with Fixed Marginals and Related Topics*
IMS Lecture Notes – Monograph Series Vol. 28, 1996

# NONPARAMETRIC MEASURES OF MULTIVARIATE ASSOCIATION

BY ROGER B. NELSEN
*Lewis and Clark College*

We study measures of multivariate association which are generalizations of the measures of bivariate association known as Spearman's rho and Kendall's tau. Since the population versions of Spearman's rho and Kendall's tau can be interpreted as measures of average positive (and negative) quadrant dependence and average total positivity of order two, respectively (Nelsen, 1992), we extend these ideas to the multivariate setting and derive measures of multivariate association from averages of orthant dependence and multivariate total positivity of order two. We examine several properties of these measures, and present examples in three and four dimensions.

**1. Introduction.** The purpose of this paper is to present three measures of multivariate association which are derived from multivariate dependence concepts. We begin in Section 2 by reviewing the main results for the bivariate case: Spearman's rho is a measure of average quadrant dependence; while Kendall's tau is a measure of average total positivity of order two. Relationships between measures of multivariate association and concordance are discussed in Section 3. In Section 4 we construct two measures of multivariate association from two generalizations of quadrant dependence – upper orthant dependence and lower orthant dependence, and examine properties of these measures in Section 5. In Section 6 we construct a measure from multivariate total positivity of order two, and in Section 7 we examine properties of this measure.

**2. The Bivariate Case.** Before proceeding, we adopt some notation and review some definitions. Let $X$ and $Y$ denote continuous random variables (r.v.'s) with joint distribution function (d.f.) $H$ and marginal d.f.'s $F$ and $G$. Let $C$ denote the copula of $X$ and $Y$, the function $C : \boldsymbol{I}^2 \to \boldsymbol{I} = [0,1]$ given by $H(x,y) = C(F(x), G(y))$. We will also denote the joint and marginal densities of $X$ and $Y$ by $h$, $f$, and $g$, respectively.

---

AMS 1991 Subject Classification: Primary 62H20

Key words and phrases: copula, Kendall's tau, multivariate total positivity of order two, orthant dependence, Spearman's rho.

The r.v.'s $X$ and $Y$ are said to be *positively quadrant dependent*, written PQD, (Lehmann, 1966) iff the probability that they are simultaneously large is at least as great as it would be were $X$ and $Y$ independent, i.e., iff

$$P(X > x, Y > y) \geq P(X > x)P(Y > y) \text{ for all } x \text{ and } y. \tag{2.1}$$

In terms of d.f.'s, $X$ and $Y$ are PQD iff $1 - F(x) - G(y) + H(x,y) \geq [1 - F(x)][1 - G(y)]$, or equivalently, iff $H(x,y) \geq F(x)G(y)$ (i.e., iff $P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y)$). So, in a sense, the expression $H(x,y) - F(x)G(y)$ measures "local" quadrant dependence at each point $(x,y)$ in $\boldsymbol{R}^2$.

The population version of Spearman's rho is given by

$$\rho_s = 12 \int_{\boldsymbol{R}^2} [H(x,y) - F(x)G(y)]dF(x)dG(y),$$

and hence $\frac{1}{12}\rho_s$ represents a measure of average quadrant dependence, where the average is with respect to the marginal distributions of $X$ and $Y$ (Nelsen, 1992). Invoking the probability transforms $u = F(x)$ and $v = G(y)$ and the copula $C$, the above expressions (as well as ones to follow) simplify. The measure of local quadrant dependence becomes $C(u,v) - uv$ (for $u$ and $v$ in $\boldsymbol{I}^2$), and $\rho_s = 12 \int_{\boldsymbol{I}^2} [C(u,v) - uv]dudv$.

In a similar fashion, the population analog of Kendall's tau is related to the dependence property known as *total positivity of order two.* $X$ and $Y$ are totally positive of order two (written $TP_2$) iff their joint density $h$ satisfies $h(x_1,y_1)h(x_2,y_2) \geq h(x_1,y_2)h(x_2,y_1)$ for all $x_1$, $x_2$, $y_1$, $y_2$ in $\boldsymbol{R}$ such that $x_1 < x_2$ and $y_1 < y_2$. Thus $h(x_1,y_1)h(x_2,y_2) - h(x_1,y_2)h(x_2,y_1)$ measures "local" $TP_2$ for $X$ and $Y$. Let $t$ denote its average for $-\infty < x_1 < x_2 < \infty$ and $-\infty < y_1 < y_2 < \infty$, i.e., let

$$t = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} \int_{-\infty}^{y_2} [h(x_1,y_1)h(x_2,y_2) - h(x_1,y_2)h(x_2,y_1)]dy_1dx_1dy_2dx_2.$$

It can be shown that $\tau = 2t$ where $\tau$ is the population version of Kendall's tau,

$$\tau = 4 \int_{\boldsymbol{R}^2} H(x,y)dH(x,y) - 1 = 4 \int_{\boldsymbol{I}^2} C(u,v)dC(u,v) - 1. \tag{2.2}$$

See Nelsen (1992) for details.

**3. Concordance.** In the bivariate setting, two pairs of r.v.'s $(X_1,Y_1)$ and $(X_2,Y_2)$ are *concordant* if $X_1 < X_2$ and $Y_1 < Y_2$ or if $X_1 > X_2$ and $Y_1 > Y_2$; and *discordant* if $X_1 < X_2$ and $Y_1 > Y_2$ or if $X_1 > X_2$ and $Y_1 < Y_2$. As is well known (Kruskal, 1958), the population version of Kendall's $\tau$ is the probability that two independent observations of the r.v.'s $X$ and $Y$ (with d.f. $H$) are concordant, scaled to be 0 when $X$ and $Y$ are independent and 1 when the d.f. of $X$ and $Y$ is the Fréchet upper bound. This can be seen by noting that

$$\begin{aligned} P(X_1 < X_2, Y_1 < Y_2 \text{ or } X_1 > X_2, Y_1 > Y_2) &= 2P(X_1 < X_2, Y_1 < Y_2) \\ &= 2\int_{\boldsymbol{R}^2} P(X < x, Y < y)dH(x,y) = 2\int_{\boldsymbol{R}^2} H(x,y)dH(x,y) \end{aligned}$$

and observing that this integral appears in (2.2).

For notation in the multivariate case, let $X_1, X_2, \cdots, X_n$ be continuous r.v.'s with marginal d.f.'s $F_1, F_2, \cdots F_n$ respectively, joint d.f. $H$, and copula $C : \boldsymbol{I}^n \to \boldsymbol{I}$ given by $H(x_1, x_2, \cdots, x_n) = C(F_1(x_1), F_2(x_2), \cdots, F_n(x_n))$. Also let $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$, $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$, and let $\boldsymbol{X} > \boldsymbol{x}$ denote the component-wise inequality $X_i > x_i$, $i = 1, 2, \cdots, n$. In a recent paper Joe (1990) studied a family of measures of multivariate concordance given by

$$\tau^* = \sum_{k=n'}^{n} w_k P(\boldsymbol{D} \in B_{k,n-k}) \tag{3.1}$$

where $\boldsymbol{D} = \boldsymbol{X} - \boldsymbol{Y}$, $B_{k,n-k}$ is the set of $\boldsymbol{x}$ in $\boldsymbol{R}^n$ with $k$ positive and $n-k$ negative or $k$ negative and $n-k$ positive components; $n' = [\frac{n+1}{2}]$, and the coefficients $w_k$ satisfy some technical restrictions. At one extreme, $w_k = 1 - \frac{4k(n-k)}{n(n-1)}$ and $\tau^*$ is the average of the $\binom{n}{2}$ pairwise Kendall's $\tau$'s for the components of $X$. At the other extreme, $w_n = 1$ and $w_k = \frac{-1}{2^{n-1}-1}$ for $k < n$, in which case $\tau^* = \frac{1}{2^{n-1}-1}(2^{n-1}P(\boldsymbol{X} < \boldsymbol{Y} \text{ or } \boldsymbol{X} > \boldsymbol{Y}) - 1)$. In Section 6 we will see that this measure of concordance is also a measure of average total positivity of order two.

We further note that $\rho_n^+$ and $\rho_n^-$, which will appear in Section 4, are also discussed in Joe (1990), where they appear as the scaled expected values $E(F_1(x_1)F_2(x_2)\cdots F_n(x_n))$ and $E(\overline{F}_1(x_1)\overline{F}_2(x_2)\cdots \overline{F}_n(x_n))$, respectively.

**4. Two Measures of Average Orthant Dependence.** There are two standard ways to generalize positive quadrant dependence to the multivariate situation (Shaked, 1982). The first is a generalization of (2.1): $X$ is *positively upper orthant dependent* (PUOD) iff $P(\boldsymbol{X} > \boldsymbol{x}) \geq \prod_{i=1}^{n} P(X_i > x_i)$ and *negatively upper orthant dependent* (NUOD) when $\geq$ is replaced by $\leq$. The second is similar: $\boldsymbol{X}$ is *positively lower orthant dependent* (PLOD) iff $P(\boldsymbol{X} \leq \boldsymbol{x}) \geq \prod_{i=1}^{n} P(X_i \leq x_i)$ and *negatively lower orthant dependent* (NLOD) when $\geq$ is replaced by $\leq$. When $n = 2$, positive upper orthant dependence and positive lower orthant dependence are the same, both reducing to positive quadrant dependence (Lehmann, 1966); but for $n \geq 3$ they are distinct concepts. For example (when $n = 3$), if $\boldsymbol{X}$ assumes the four values $(1,1,1), (1,0,0), (0,1,0)$ and $(0,0,1)$ each with probability $\frac{1}{4}$, then it is easy to verify that $\boldsymbol{X}$ is PUOD but not PLOD. (Note that $P(\boldsymbol{X} \leq \boldsymbol{0}) = 0$ while $P(X_1 \leq 0)P(X_2 \leq 0)P(X_3 \leq 0) = \frac{1}{8}$.)

As with quadrant dependence, we can view the expression $P(\boldsymbol{X} > \boldsymbol{x}) - \prod_{i=1}^{n} P(X_i > x_i)$ as a measure of "local" upper orthant dependence at each point $\boldsymbol{x}$ in $\boldsymbol{R}^n$. If we set $U_i = F_i(X_i)$ for $i = 1, 2, \cdots, n$ and $\boldsymbol{U} = (U_1, U_2, \cdots, U_n)$, then each $U_i$ is uniform on $\boldsymbol{I}$ and the d.f. of $\boldsymbol{U}$ is the copula $C$. Furthermore, $\boldsymbol{X}$ is PUOD (NUOD) iff $\boldsymbol{U}$ is PUOD (NUOD), or equivalently, iff $P(\boldsymbol{U} > \boldsymbol{u}) \geq (\leq) \prod_{i=1}^{n}(1-u_i)$. Thus $P(\boldsymbol{U} > \boldsymbol{u}) - \prod_{i=1}^{n}(1-u_i)$ also measures local upper orthant dependence for $\boldsymbol{X}$. Let $p$ denote its average:

$$p = \int_{\boldsymbol{I}^n} [P(\boldsymbol{U} > \boldsymbol{u}) - \prod_{i=1}^{n}(1-u_i)]du_1du_2\cdots du_n.$$

Now let $\boldsymbol{V}$ be a random vector with $n$ independent components each uniformly distributed on $\boldsymbol{I}$. Then

$$\begin{aligned}\int_{\boldsymbol{I}^n} [P(\boldsymbol{U} > \boldsymbol{u})]du_1du_2\cdots du_n &= P(\boldsymbol{U} > \boldsymbol{V}) \\ &= P(\boldsymbol{V} < \boldsymbol{U}) = \int_{\boldsymbol{I}^n}[P(\boldsymbol{V} < \boldsymbol{u})]dC(\boldsymbol{u}) = \int_{\boldsymbol{I}^n} u_1u_2\cdots u_n dC(\boldsymbol{u}),\end{aligned}$$

and since $\int_{\boldsymbol{I}^n} \prod_{i=1}^{n}(1-u_i)du_1du_2\cdots du_n = \left(\frac{1}{2}\right)^n$, we have

$$p = \int_{\boldsymbol{I}^n} u_1u_2\cdots u_n dC(\boldsymbol{u}) - \left(\frac{1}{2}\right)^n.$$

If $\boldsymbol{X}$ is a vector of independent r.v.'s with marginal d.f.'s $F_1, F_2, \cdots, F_n$, then $H(x_1, x_2, \cdots, x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n)$, and hence the copula for such an $\boldsymbol{X}$, which we will denote by $\Pi(\boldsymbol{u})$, is given by $\Pi(u_1, u_2, \cdots, u_n) = u_1u_2\cdots u_n$. In this case $p = 0$. When $H(x_1, x_2, \cdots, x_n) = \min(F_1(x_1), F_2(x_2), \cdots, F_n(x_n))$, the Fréchet upper bound for distributions with margins $F_1, F_2, \cdots, F_n$, then the copula is $M(\boldsymbol{u}) = \min(u_1, u_2, \cdots, u_n)$. In this case $P(\boldsymbol{U} > \boldsymbol{u}) = \min(1 - u_1, 1 - u_2, \cdots, 1 - u_n)$, so that $\int_{\boldsymbol{I}^n}[P(\boldsymbol{U} > \boldsymbol{u})]du_1du_2\cdots du_n = \dfrac{1}{n+1}$. Since $C(\boldsymbol{u}) \leq M(\boldsymbol{u})$ for every $C$, it follows that $p \leq \frac{1}{n+1} - \left(\frac{1}{2}\right)^n$. So if we divide $p$ by this constant, we will have a measure which is 1 when the d.f. of $\boldsymbol{X}$ is the

Fréchet upper bound (and still 0 when the components of $\boldsymbol{X}$ are independent). Hence we make the following

DEFINITION 4.1. *A measure of multivariate association $\rho_n^+$ derived from average upper orthant dependence for the random vector $\boldsymbol{X}$ with copula $C(\boldsymbol{u})$ is given by:*

$$\rho_n^+ = \frac{n+1}{2^n-(n+1)}(2^n \int_{\boldsymbol{I}^n} u_1 u_2 \cdots u_n dC(\boldsymbol{u}) - 1);$$

*or, equivalently,*

$$\rho_n^+ = \frac{n+1}{2^n-(n+1)}(2^n \int_{\boldsymbol{I}^n} \Pi(\boldsymbol{u}) dC(\boldsymbol{u}) - 1).$$

In a similar fashion, we can define another measure $\rho_n^-$ from average lower orthant dependence. Recall that $\boldsymbol{X}$ is PLOD (NLOD) iff $P(\boldsymbol{X} \le \boldsymbol{x}) \ge (\le) \prod_{i=1}^{n} P(X_i \le x_i)$, that is, iff $P(\boldsymbol{U} \le \boldsymbol{u}) \ge (\le) \prod_{i=1}^{n} u_i$. So $P(\boldsymbol{U} \le \boldsymbol{u}) - \prod_{i=1}^{n} u_i$ measures "local" positive and negative lower orthant dependence; let $q$ denote its average;

$$q = \int_{\boldsymbol{I}^n} [P(\boldsymbol{U} \le \boldsymbol{u}) - \prod_{i=1}^{n} u_i] du_1 du_2 \cdots du_n;$$

from which it readily follows that

$$q = \int_{\boldsymbol{I}^n} C(\boldsymbol{u}) du_1 du_2 \cdots du_n - \left(\frac{1}{2}\right)^n.$$

As with $p, q = 0$ when the components of $\boldsymbol{X}$ are independent, and $q = \frac{1}{n+1} - \left(\frac{1}{2}\right)^n$ when the d.f. of $\boldsymbol{X}$ is the Fréchet upper bound. So we make the following

DEFINITION 4.2. *A measure of multivariate association $\rho_n^-$ derived from average lower orthant dependence for the random vector $\boldsymbol{X}$ with copula $C(\boldsymbol{u})$ is given by:*

$$\rho_n^- = \frac{n+1}{2^n-(n+1)}(2^n \int_{\boldsymbol{I}^n} C(\boldsymbol{u}) du_1 du_2 \cdots du_n - 1);$$

or, equivalently,

$$\rho_n^- = \frac{n+1}{2^n-(n+1)}(2^n \int_{\boldsymbol{I}^n} C(\boldsymbol{u}) d\Pi(\boldsymbol{u}) - 1).$$

**5. Properties of $\rho_n^+$ and $\rho_n^-$ and Examples.**

1. A lower bound for $\rho_n^+$ and $\rho_n^-$ is $\frac{2^n-(n+1)!}{n![2^n-(n+1)]}$. Since the Fréchet lower bound for distributions with marginal d.f.'s $F_1, F_2, \cdots, F_n$ is $\max(F_1(x_1)+F_2(x_2)+\cdots+F_n(x_n)-n+1,0)$, it follows that $P(\boldsymbol{U}>\boldsymbol{u}) \geq \max(1-u_1-u_2-\cdots-u_n,0)$. Thus $\int_{\boldsymbol{I}^n}[P(\boldsymbol{U}>\boldsymbol{u})]du_1du_2\cdots du_n \geq \int_{\boldsymbol{I}^n}[\max(1-u_1-u_2-\cdots-u_n,0)]du_1du_2\cdots du_n = \dfrac{1}{(n+1)!}$, so that $\rho_n^+ \geq \frac{n+1}{2^n-(n+1)}\left(\frac{2^n}{(n+1)!}-1\right)$. Similarly, since $C(\boldsymbol{u}) \geq \max(u_1+u_2+\cdots+u_n-n+1,0)$, then $\int_{\boldsymbol{I}^n} C(\boldsymbol{u})du_1du_2\cdots du_n \geq \int_{\boldsymbol{I}^n}\max(u_1+u_2+\cdots+u_n-n+1,0)du_1du_2\cdots du_n = \frac{1}{(n+1)!}$, and we obtain the same lower bound for $\rho_n^-$. Since the Fréchet lower bound is not a d.f. for $n \geq 3$, this bound may well fail to be best possible.

2. For $n=2$, both $\rho_2^+$ and $\rho_2^-$ reduce to Spearman's $\rho_s$ discussed in Section 2.

3. For $n=3$, the inclusion-exclusion principle yields

$$P(\boldsymbol{U}>\boldsymbol{u})+C(\boldsymbol{u}) = 1-u_1-u_2-u_3+C(u_1,u_2,1)+C(u_1,1,u_3)+C(1,u_2,u_3).$$

Integrating over $\boldsymbol{I}^3$ gives

$$\frac{\rho_3^++1}{8}+\frac{\rho_3^-+1}{8} = 1-3\cdot\frac{1}{2}+\frac{\rho_{XY}+3}{12}+\frac{\rho_{XZ}+3}{12}+\frac{\rho_{YZ}+3}{12},$$

where $\rho_{XY}$, $\rho_{XZ}$ and $\rho_{YZ}$ denote Spearman's $\rho_s$ for the two r.v.'s displayed as the subscript. It follows that

$$\frac{1}{2}(\rho_3^++\rho_3^-) = \frac{1}{3}(\rho_{XY}+\rho_{XZ}+\rho_{YZ}).$$

Similar expressions can be obtained for $n \geq 4$.

4. Suppose that the distribution of $\boldsymbol{X}$ is *radially symmetric* (Nelsen, 1993), that is, there is a point $\boldsymbol{a}$ in $\boldsymbol{R}^n$ such that $P(\boldsymbol{X} \leq \boldsymbol{a}-\boldsymbol{x}) = P(\boldsymbol{X} > \boldsymbol{a}+\boldsymbol{x})$ for all $\boldsymbol{x}$ in $\boldsymbol{R}^n$. It follows that $P(\boldsymbol{U} \leq \boldsymbol{u}) = P(\boldsymbol{U} > \boldsymbol{1}-\boldsymbol{u})$, so that $\rho_n^+ = \rho_n^-$.

EXAMPLE 1. Let $X$, $Y$, and $Z$ be r.v.'s, each uniform on $\boldsymbol{I}$, such that the probability mass in $\boldsymbol{I}^3$ is uniformly distributed on the faces of the tetrahedron with vertices (1, 0, 0), (0, 1, 0), (0, 0, 1), and (1, 1, 1). For this distribution, $\rho_3^+ = \frac{2}{15}$ and $\rho_3^- = -\frac{2}{15}$. Furthermore, since $X$, $Y$, and $Z$ are pairwise independent, $\rho_{XY} = \rho_{XZ} = \rho_{YZ} = 0$. Thus a common measure of multivariate dependence, the average of the pairwise Spearman's $\rho_s$'s, is 0 for $(X,Y,Z)$. But the fact that $\rho_3^+$ is positive and $\rho_3^-$ is negative indicates some degree of positive upper orthant dependence and negative lower orthant dependence – indeed, $P\left(X>\frac{1}{2}, Y>\frac{1}{2}, Z>\frac{1}{2}\right) = \frac{3}{16}$ and $P\left(X<\frac{1}{2}, Y<\frac{1}{2}, Z<\frac{1}{2}\right) = \frac{1}{16}$, but both of these probabilities are $\frac{1}{8}$ when $X$, $Y$, and $Z$ are mutually independent.

EXAMPLE 2. Suppose $X$, $Y$, and $Z$ have the trivariate Farlie-Gumbel-Morgenstern (FGM) distribution on $\boldsymbol{I}^3$ with d.f. $H(x,y,z) = xyz[1+\alpha(1-y)(1-z)+\beta(1-x)(1-z)+\gamma(1-x)(1-y)+\delta(1-x)(1-y)(1-z)]$; where the four parameters $\alpha$, $\beta$, $\gamma$, and $\delta$ are each in $[-1,1]$ satisfying the inequalities $1+\epsilon_1\alpha+\epsilon_2\beta+\epsilon_3\gamma \geq |\delta|$ for $\epsilon_i = \pm 1$, $\epsilon_1\epsilon_2\epsilon_3 = 1$. The univariate margins are uniform on $\boldsymbol{I}$ while the bivariate margins are FGM, and the pairwise Spearman's $\rho_s$'s are $\rho_{XY} = \frac{1}{3}\gamma$, $\rho_{XZ} = \frac{1}{3}\beta$, and $\rho_{YZ} = \frac{1}{3}\alpha$. Thus the average of the pairwise $\rho_s$'s is $\frac{1}{9}(\alpha+\beta+\gamma)$, however

$$\rho_3^+ = \frac{1}{9}(\alpha+\beta+\gamma) - \frac{1}{27}\delta \text{ and } \rho_3^- = \frac{1}{9}(\alpha+\beta+\gamma) + \frac{1}{27}\delta.$$

EXAMPLE 3. Copulas for trivariate Cuadras-Augé distributions are weighted geometric means of the copula for independent r.v.'s and the copula for the Fréchet upper bound: $C_\theta(x,y,z) = [\min(x,y,z)]^\theta (xyz)^{1-\theta}$, $\theta \in \boldsymbol{I}$. The univariate margins are uniform on $\boldsymbol{I}$ and the bivariate margins are Cuadras-Augé distributions with parameter $\theta$. This distribution is both PUOD and PLOD and $\rho_{XY} = \rho_{XZ} = \rho_{YZ} = 3\theta/(4-\theta)$; however

$$\rho_3^+ = \frac{\theta(11-5\theta)}{(3-\theta)(4-\theta)} \quad \text{and} \quad \rho_3^- = \frac{\theta(7-\theta)}{(3-\theta)(4-\theta)}.$$

EXAMPLE 4. Let $X$, $Y$, and $Z$ be r.v.'s, each uniform on $\boldsymbol{I}$, such that the probability mass in $\boldsymbol{I}^3$ is uniformly distributed on two triangles, one with vertices (1, 0, 0), (0, 1, 0), and (0, 0, 1); and one with vertices (1, 1, 0), (1, 0, 1), and (0, 1, 1). Here $\rho_3^+ = \rho_3^- = 0$. Note that $X$, $Y$, and $Z$ are again pairwise independent and that $(X,Y,Z)$ is radially symmetric about $\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)$.

**6. A Measure of Average Multivariate Total Positivity of Order Two.** The multivariate version of the $TP_2$ property is called multivariate total positivity of order two (Karlin and Rinott, 1980): A distribution with joint density $h$ is *multivariate totally positive of order two* ($MTP_2$) iff for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\boldsymbol{R}^n$,

$$h(\boldsymbol{x} \vee \boldsymbol{y})h(\boldsymbol{x} \wedge \boldsymbol{y}) \geq h(\boldsymbol{x})h(\boldsymbol{y})$$

*where*

$$\boldsymbol{x} \vee \boldsymbol{y} = (\max(x_1,y_1), \max(x_2,y_2), \cdots, \max(x_n,y_n))$$

*and*

$$\boldsymbol{x} \wedge \boldsymbol{y} = (\min(x_1,y_1), \min(x_2,y_2), \cdots, \min(x_n,y_n)).$$

Thus $h(\boldsymbol{x}\vee\boldsymbol{y})h(\boldsymbol{x}\wedge\boldsymbol{y}) - h(\boldsymbol{x})h(\boldsymbol{y})$ measures "local" $MTP_2$ for a distribution with density $h$, and we let $T$ denote its average:

$$T = \int_{\boldsymbol{R}^n}\int_{\boldsymbol{R}^n}[h(\boldsymbol{x}\vee\boldsymbol{y})h(\boldsymbol{x}\wedge\boldsymbol{y}) - h(\boldsymbol{x})h(\boldsymbol{y})]dx_1dx_2\cdots dx_ndy_1dy_2\cdots dy_n.$$

Let $s_i = F_i(x_i)$, $t_i = F_i(y_i)$, and $h(\boldsymbol{x}) = c(F_1(x_1), F_2(x_2), \cdots, F_n(x_n))$ $f_1(x_1)f_2(x_2)\cdots f_n(x_n)$ where $f_i$ is the (marginal) density of $X_i$ and $c(\boldsymbol{u}) = \partial^n C(\boldsymbol{u})/\partial u_1\partial u_2\cdots\partial u_n$. Then

$$T = \int_{\boldsymbol{I}^n}\int_{\boldsymbol{I}^n}[c(\boldsymbol{s}\vee\boldsymbol{t})c(\boldsymbol{s}\wedge\boldsymbol{t}) - c(\boldsymbol{s})c(\boldsymbol{t})]ds_1ds_2\cdots ds_ndt_1dt_2\cdots dt_n.$$

Now let $\boldsymbol{u} = \boldsymbol{s}\vee\boldsymbol{t}$ and $\boldsymbol{v} = \boldsymbol{s}\wedge\boldsymbol{t}$. Then $\boldsymbol{v}\le\boldsymbol{u}$ and $dv_idu_i = ds_idt_i$ for $i = 1, 2, \cdots, n$; hence

$$T = \int_0^1\cdots\int_0^1\int_0^1\int_0^{u_n}\cdots\int_0^{u_2}\int_0^{u_1}\sum_{k=0}^{n}\sum_{\substack{S\subseteq N\\|S|=k}}[c(\boldsymbol{u})c(\boldsymbol{v}) - c(\boldsymbol{w})c(\boldsymbol{z})]$$
$$dv_1dv_2\cdots dv_ndu_1du_2\cdots du_n$$

where $N = \{1, 2, \cdots, n\}$,

$$w_i = \begin{cases} u_i & \text{if } i\in S,\\ v_i & \text{if } i\notin S,\end{cases} \quad\text{and}\quad z_i = \begin{cases} v_i & \text{if } i\in S,\\ u_i & \text{if } i\notin S.\end{cases}$$

Evaluation of the $n$ inner integrals yields

$$T = \int_{\boldsymbol{I}^n}[2^nC(\boldsymbol{u})c(\boldsymbol{u}) - \sum_{k=0}^{n}\sum_{\substack{S\subseteq N\\|S|=k}}\partial_S^kC(\boldsymbol{u})\partial_{N-S}^{n-k}C(\boldsymbol{u})]du_1du_2\cdots du_n,$$

where $\partial_S^kC(\boldsymbol{u})$ denotes the $k$th order mixed partial derivative of $C(\boldsymbol{u})$ with respect to the $k$ variables whose subscripts are in $S$. But since the double sum in the above expression is simply $\frac{\partial^n}{\partial u_1\partial u_2\cdots\partial u_n}C^2(\boldsymbol{u})$, we have

$$T = 2^n\int_{\boldsymbol{I}^n}C(\boldsymbol{u})c(\boldsymbol{u})du_1du_2\cdots du_n - \int_{\boldsymbol{I}^n}\frac{\partial^n}{\partial u_1\partial u_2\cdots\partial u_n}C^2(\boldsymbol{u})du_1du_2\cdots du_n.$$

The second multiple integral above is $C^2(\mathbf{1}) = 1$, and thus

$$T = 2^n\int_{\boldsymbol{I}^n}C(\boldsymbol{u})dC(\boldsymbol{u}) - 1.$$

As with $p$ and $q$ in Section 4, $T = 0$ when the components of $X$ are independent, and when the d.f. of $\boldsymbol{X}$ is the Fréchet upper bound, $T = 2^{n-1} - 1$. So we make the following

DEFINITION 6.1. *A measure of multivariate association $\tau_n$ derived from average multivariate total positivity of order two for the random vector $\boldsymbol{X}$ with copula $C(\boldsymbol{u})$ is given by:*

$$\tau_n = \frac{1}{2^{n-1}-1}(2^n \int_{\boldsymbol{I}^n} C(\boldsymbol{u})dC(\boldsymbol{u}) - 1).$$

## 7. Properties of $\tau_n$ and Examples.

1. A lower bound for $\tau_n$ is $\frac{-1}{2^{n-1}-1}$ (since $\int_{\boldsymbol{I}^n} C(\boldsymbol{u})dC(\boldsymbol{u}) \geq 0$.)

2. For $n = 2$, $\tau_2$ reduces to Kendall's $\tau$ as presented in Section 2.

3. For $n = 3$, $\tau_3 = \frac{1}{3}(\tau_{XY} + \tau_{XZ} + \tau_{YZ})$, where $\tau_{XY}$, $\tau_{XZ}$ and $\tau_{YZ}$ denote Kendall's $\tau$ for the two r.v.'s displayed as the subscript. This follows from the fact that the family of measures of multivariate concordance studied by Joe (1990) includes both $\tau_n$ and the average of the pairwise Kendall's $\tau$'s; but has only one member when $n = 3$.

4. As noted in Section 3, $\tau_n$ is a scaled probability of concordance. This follows from the observation that $2\int_{\boldsymbol{I}^n} C(\boldsymbol{u})dC(\boldsymbol{u}) = P(\boldsymbol{X} < \boldsymbol{Y} \text{ or } \boldsymbol{X} > \boldsymbol{Y})$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent each with d.f. $H$.

EXAMPLE 5. Let $X, Y$, and $Z$ be r.v.'s on $\boldsymbol{I}^3$ with a density $h(x,y,z) = 4$ in the two cubes $[0, \frac{1}{2}]^3$ and $[\frac{1}{2}, 1]^3$, and 0 elsewhere. Then $\rho_3^+ = \rho_3^- = \frac{3}{4}$ while $\tau_3 = \frac{1}{2}$.

EXAMPLE 6. *Now let $X, Y$, and $Z$ be r.v.'s on $\boldsymbol{I}^3$ with a density $h(x,y,z) = 2$ in the four cubes $[\frac{1}{2}, 1] \times [0, \frac{1}{2}] \times [0, \frac{1}{2}]$, $[0, \frac{1}{2}] \times [\frac{1}{2}, 1] \times [0, \frac{1}{2}]$, $[0, \frac{1}{2}] \times [0, \frac{1}{2}] \times [\frac{1}{2}, 1]$, and $[\frac{1}{2}, 1]^3$; and 0 elsewhere. Now $\rho_3^+ = \frac{1}{8}$, $\rho_3^- = -\frac{1}{8}$, while $\tau_3 = 0$. Note that in this example $X, Y$, and $Z$ are pairwise independent.*

EXAMPLE 7. *Let $X, Y, Z$, and $W$ be r.v.'s on $\boldsymbol{I}^4$ with d.f. $H(x,y,z,w) = xyzw[1 + \theta(1-x)(1-y)(1-z)(1-w)]$ and density $h(x,y,z,w) = 1 + \theta(1 - 2x)(1-2y)(1-2z)(1-2w)$, where $|\theta| \leq 1$. Any three of these four r.v.'s are mutually independent; however all four are not unless $\theta = 0$, and*

$$\rho_4^+ = \rho_4^- = \frac{5}{891}\theta \text{ and } \tau_4 = \frac{2}{567}\theta.$$

## References

JOE, H. (1990). Multivariate concordance. *J. Multivariate Anal.* **35**, 12–30.

KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities – I. *J. Multivariate Anal.* **10**, 467–498.

KRUSKAL, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **53**, 814–861.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–1153.

NELSEN, R. B. (1992). On measures of association as measures of positive dependence. *Statist. Probab. Lett.* **14**, 269–274.

NELSEN, R. B. (1993). Some concepts of bivariate symmetry. *Nonparametric Statist.* **3**, 95–101.

SHAKED, M. (1982). A general theory of some positive dependence notions. *J. Multivariate Anal.* **12**, 199–218.

DEPARTMENT OF MATHEMATICAL SCIENCES
LEWIS AND CLARK COLLEGE
PORTLAND, OR 97219-7899
nelsen@lclark.edu