Statistical Papers **48**, 329-336 (2007)

Statistical Papers

Extremes of Nonexchangeability

Roger B. Nelsen

Department of Mathematical Sciences, Lewis & Clark College, Portland, Oregon 97219 USA

Received: February 3, 2005; revised version: April 7, 2005

Summary

For identically distributed random variables X and Y with joint distribution function H, we show that the supremum of $|H(x,y)-H(y,x)|$ is $1/3$. Using copulas, we define a measure of nonexchangeability, and study maximally nonexchangeable random variables and copulas. In particular, we show that maximally nonexchangeable random variables are negatively correlated in the sense of Spearman's rho.
Keywords: Copula; Nonexchangeable random variables: Spearman's rho.

1. Introduction

Since their introduction by de Finetti in 1930, exchangeable random variables have found important applications in many areas of statistics: limit laws, extreme value theory, Bayesian statistics, stochastic processes, etc. (see (Galambos, 1982) for a survey of properties and applications of exchangeable random variables). Formally, a pair X, Y of random variables are *exchangeable* if the vectors (X,Y) and (Y,X) have the same distribution. Equivalently, if H denotes the joint distribution function of X and Y, then X and Y are exchangeable if $H(x,y) = H(y,x)$ for all real x and y. Consequently, exchangeable random variables must be identically distributed. However, little attention seems to have been focused on the ways in which identically distributed random variables can fail to be exchangeable, and the resulting dependence structure.

When X and Y are not exchangeable, $H(x,y) \neq H(y,x)$ for some x and y, and the supremum of $|H(x,y)-H(y,x)|$ can be used to measure the "amount" or "degree" of nonexchangeabilty of X and Y. This quantity can equal 1—for example, when the supports of X and Y are subsets of (0,1) and (1,2) respectively, then $H(1,2) = 1$ and $H(2,1) = 0$. If X and Y are continuous and identically distributed (as we shall now assume), the set of values of $|H(x,y) - H(y,x)|$ for real x and y is the same as the set of values of $|C(u,v) - C(v,u)|$ for u and v in $\mathbf{I}$ = [0,1], where C denotes the *copula* of X and Y. In general, the copula of X and Y is defined implicitly by the relationship $H(x,y) = C(F(x),G(y))$, where F and G denote the distribution functions of X and Y, respectively. For identically distributed random variables (i.e., when $F = G$), exchangeability is equivalent to the symmetry of the copula: $C(u,v) = C(v,u)$ for all u,v in $\mathbf{I}$.

In a sense, the relationship between exchangeability and nonexchangeability is analogous to the relationship between independence and dependence. While there is but one copula for independent random variables (namely $C(u,v) = uv$), there is but one class of copulas for exchangeable random variables (the symmetric copulas). At the other extreme, there are several forms of "complete" or "maximal" dependence—examples include: (i) X and Y are *mutually completely dependent* (Lancaster, 1963) if each is almost surely an invertible function of the other; and (ii) X and Y are *comonotone* (*countermonotone*) if each random variable is almost surely an increasing (decreasing) function of the other, or equivalently, if the copula of X and Y is $M(u,v) = \min(u,v)$ ($W(u,v) = \max(0,u+v-1)$). In this paper we shall construct a measure of nonexchangeability, investigate one form of "maximal" nonexchangeability, and study the dependence structure of maximally nonexchangeable identically distributed random variables.

A few comments on notation: $V_C([a,b]\times[c,d])$ denotes the *C-volume* of the rectangle $[a,b]\times[c,d]$, i.e. the probability mass the copula C assigns to that rectangle, computed via the formula $V_C([a,b]\times[c,d])$ = $C(b,d) - C(a,d) - C(b,c) + C(a,c)$; C^T denotes the *transpose* of C, given by $C^T(u,v) = C(v,u)$; and $\prec$ denotes the pointwise partial order for copulas, i.e., $C \prec C'$ if and only if $C(u,v) \le C'(u,v)$ for all u,v in $\mathbf{I}$.

The following two copulas will play an important role in the sequel:

$$C_1(u,v) = \min\left(u, v, (u - 2/3)^+ + (v - 1/3)^+\right), \tag{1}$$

$$C_2(u,v) = \max\left(0, u + v - 1, 1/3 - (1/3 - u)^+ - (2/3 - v)^+\right), \tag{2}$$

where $x^+ = \max(x,0)$. Both C_1 and C_2 are singular, the support of C_1 consists of two line segments in $\mathbf{I}^2$, one connecting $(0,1/3)$ to $(2/3,1)$ and one connecting $(2/3,0)$ to $(1,1/3)$; while the support of C_2 consists of three line segments in $\mathbf{I}^2$, one connecting $(0,2/3)$ to $(1/3,1/3)$, another connecting $(1/3,1)$ to $(2/3,2/3)$, and a third connecting $(2/3,1/3)$ to $(1,0)$. See Figure 1.

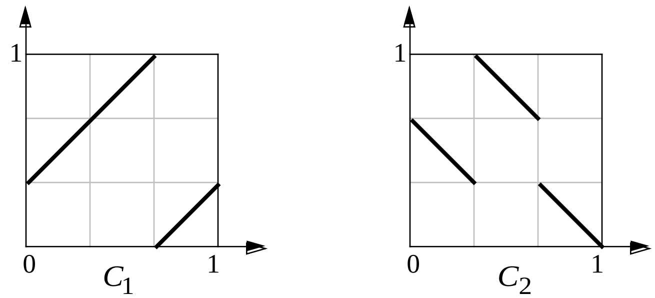

Figure 1. The supports of C_1 and C_2

2. A Measure of Nonexchangeability

We first show that the quantity $|C(u,v)-C(v,u)|$ has an attainable upper bound of $1/3$.

Lemma 2.1. *For any copula C and any u,v in* **I**,

$$|C(u,v)-C(v,u)| \le \min(u,v,1-u,1-v,|u-v|). \tag{3}$$

Proof. Assume $u \le v$, it then suffices to show that $|C(u,v)-C(v,u)| \le \min(u,1-v,v-u)$. First note that $|C(u,v)-C(v,u)| \le \max(C(u,v),C(v,u)) \le u$. Also, $C(u,v)-C(v,u) \le u - C(v,u) \le 1 - v$, where the last inequality follows from $V_C([v,1]\times[u,1]) \ge 0$. Similarly $C(v,u)-C(u,v) \le 1-v$, thus $|C(u,v)-C(v,u)| \le 1-v$. Finally, $C(u,v)-C(v,u) \le C(u,v)-C(u,u) \le v-u$, since $V_C([u,1]\times[u,v]) \ge 0$; likewise $C(v,u)-C(u,v) \le v-u$. Hence $|C(u,v)-C(v,u)| \le v-u$, which completes the proof, as the case $v \le u$ is analogous. □

The graph of $z = \min(u,1-v,v-u)$ for (u,v) satisfying $0 \le u \le v \le 1$ consists of the upper three faces of a tetrahedron whose base is the triangle $0 \le u \le v \le 1$ in the *u*-*v* plane and whose vertex is the point $(1/3,2/3,1/3)$. An analogous result holds for graph of $\min(v,1-u,u-v)$ for (u,v) satisfying $0 \le v \le u \le 1$. Thus $\min(u,v,1-u,1-v,|u-v|) \le 1/3$ for all u,v in **I**. Furthermore, for the copulas C_1 and C_2 in (1) and (2), $C_1(1/3,2/3) - C_1(2/3,1/3) = C_2(1/3,2/3) - C_2(2/3,1/3) = 1/3$, which proves

Theorem 2.2. *For any copula C*,

$$\sup_{u,v\in \mathbf{I}} |C(u,v)-C(v,u)| \le \frac{1}{3}, \tag{4}$$

and the inequality is best-possible.

Corollary 2.3. *For any bivariate distribution function H with identical margins*,

$$\sup_{x,y\in(-\infty,\infty)} |H(x,y)-H(y,x)| \le \frac{1}{3},$$

and the inequality is best-possible.

The inequality in Theorem 2.2 motivates the following definition:

Definition 2.4. Let C be a copula. The *degree of nonexchangeability* of C, denoted $\delta(C)$, is given by

$$\delta(C) = 3 \sup_{u,v \in \mathbf{I}} \left| C(u,v) - C(v,u) \right|. \tag{5}$$

The degree of nonexchangeability of a copula can also be viewed as a measure of copula asymmetry, since $\delta(C) = 0$ if and only if $C = C^T$.

Example 2.1. Let $\{C_{\alpha,\beta}\}$ be the family of asymmetric copulas given by

$$C_{\alpha,\beta}(u,v) = uv + uv(1-u)(1-v)\left[\alpha + (\beta - \alpha)v(1-u)\right]$$

for $|\alpha| \le 1$, $(1/2)\left[\alpha - 3 - (9 + 6\alpha - 3\alpha^2)^{1/2}\right] \le \beta \le 1$, and $\alpha \ne \beta$. The quantity $\left|C_{\alpha,\beta}(u,v) - C_{\alpha,\beta}(v,u)\right|$ achieves a maximum value of $|\alpha - \beta|\sqrt{5}/125$ at the points $(u,v) = \left((5-\sqrt{5})/10, (5+\sqrt{5})/10\right)$ and $\left((5+\sqrt{5})/10, (5-\sqrt{5})/10\right)$, and thus $\delta\left(C_{\alpha,\beta}\right) = 3|\alpha - \beta|\sqrt{5}/125$. The values of α and β that maximize $|\alpha - \beta|$ are 1 and $-1-\sqrt{3}$, respectively, and hence the member of this family with the largest degree of nonexchangeability is $C_{1,-1-\sqrt{3}}$, with $\delta\left(C_{1,-1-\sqrt{3}}\right) \cong 0.20$.

Example 2.2. Let $\{C_\theta\}$ be the family of asymmetric copulas given by

$$C_\theta(u,v) = \min\left(u, \frac{\theta}{1-\theta}v + \frac{1-2\theta}{1-\theta}(u+v-1)^+\right)$$

for $\theta \in [0,1/2]$. Each member of this family is singular, with probability mass θ uniformly distributed on the line segment joining (0,0) to $(\theta,1-\theta)$ and on the line segment joining $(\theta,1-\theta)$ to (1,1); and mass $1 - 2\theta$ uniformly distributed on the line segment joining $(\theta,1-\theta)$ to (1,0). See Figure 2. Note that $C_0(u,v) = W(u,v)$ and $C_{1/2}(u,v) = M(u,v)$. For members of this family, the quantity $\left|C_\theta(u,v) - C_\theta(v,u)\right|$ achieves a maximum value of $\theta(1-2\theta)/(1-\theta)$ at the points $(u,v) = (\theta,1-\theta)$ and $(1-\theta,\theta)$. The value of θ which maximizes $\theta(1-2\theta)/(1-\theta)$ is $\theta = 1-\sqrt{2}/2$, and hence the member of this family with the largest degree of nonexchangeability is $C_{1-\sqrt{2}/2}$, with $\delta\left(C_{1-\sqrt{2}/2}\right) = 3\left(3-2\sqrt{2}\right) \cong 0.515$.

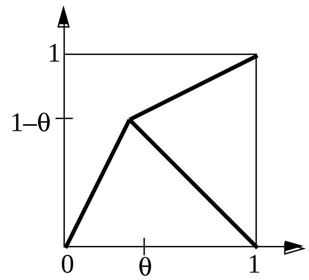

Figure 2. The support of C_θ for $\theta \in (0,1/2)$

If X and Y are identically distributed continuous random variables with copula C, then X and Y are exchangeable if and only if $\delta(C) = 0$. At the other extreme, we have

Definition 2.5. If X and Y are identically distributed continuous random variables with copula C and $\delta(C) = 1$, then X and Y are *maximally nonexchangeable*, and C is called a *maximally nonexchangeable copula*.

Observe that the copulas C_1 and C_2 in (1) and (2) are maximally nonexchangeable. In the next section we study properties of the set of maximally nonexchangeable copulas.

3. Maximally Nonexchangeable Random Variables and Copulas

Let $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, and $\mathbf{C}_4$ denote the following sets of copulas:

$$\mathbf{C}_1 = \{C | C(2/3,1/3) = 0\},\ \mathbf{C}_2 = \{C | C(1/3,2/3) = 1/3\},$$

$$\mathbf{C}_3 = \{C | C(1/3,2/3) = 0\},\ \mathbf{C}_4 = \{C | C(2/3,1/3) = 1/3\}.$$

Note that $\mathbf{C}_1$ and $\mathbf{C}_4$ are disjoint, as are $\mathbf{C}_2$ and $\mathbf{C}_3$; a copula C is in $\mathbf{C}_1$ ($\mathbf{C}_2$) if and only if C^T is in $\mathbf{C}_3$ ($\mathbf{C}_4$), and C_1 and C_2 from (1) and (2) are in $\mathbf{C}_1 \cap \mathbf{C}_2$.

With these four sets, we can describe the set of maximally nonexchangeable copulas, find bounds for that set, and examine a measure of association for maximally nonexchangeable random variables.

Theorem 3.1. *Let* $\mathbf{C}$ *denote the set of maximally nonexchangeable copulas, i.e.,* $\mathbf{C} = \{C | \delta(C) = 1\}$. *Then*:

(a) $\mathbf{C} = (\mathbf{C}_1 \cap \mathbf{C}_2) \cup (\mathbf{C}_3 \cap \mathbf{C}_4)$, *i.e.,* $C \in \mathbf{C}$ *if and only if either* $C(1/3,2/3) = 1/3$ *and* $C(2/3,1/3) = 0$, *or* $C(1/3,2/3) = 0$ *and* $C(2/3,1/3) = 1/3$;

(b) *for every* $C \in \mathbf{C}$, $C \in \mathbf{C}_1 \cap \mathbf{C}_2$ *if and only if* $C_2 \prec C \prec C_1$, *and* $C \in \mathbf{C}_3 \cap \mathbf{C}_4$ *if and only if* $C_2^T \prec C \prec C_1^T$, *where* C_1 *and* C_2 *are given in* (1) *and* (2);

(c) *every maximally nonexchangeable copula has the same diagonal section* $C(u,u)$, *i.e., for any* $C \in \mathbf{C}$, $C(u,u) = (u - 1/3)^+ + (u - 2/3)^+$.

Proof. (a) If $\delta(C) = 1$, then there exists a point (u_0,v_0) in $\mathbf{I}^2$ with $u_0 < v_0$ such that either (i) $C(u_0,v_0) - C(v_0,u_0) = 1/3$ or (ii) $C(v_0,u_0) - C(u_0,v_0) = 1/3$. We show that in case (i), $C \in \mathbf{C}_1 \cap \mathbf{C}_2$, the proof that $C \in \mathbf{C}_3 \cap \mathbf{C}_4$ in case (ii) is analogous and omitted. Assume that $C(u_0,v_0) - C(v_0,u_0) = 1/3$. From (3) and (4), we have $\min(u_0, 1 - v_0, v_0 - u_0) = 1/3$, and hence $(u_0,v_0) = (1/3, 2/3)$. So $C(1/3,2/3) - C(2/3,1/3) = 1/3$, but since $C(1/3,2/3) \le 1/3$ and $C(2/3,1/3) \ge 0$, it follows that $C(1/3,2/3) = 1/3$ and $C(2/3,1/3) = 0$. Hence $C \in \mathbf{C}_1 \cap \mathbf{C}_2$.
(b) Let $C \in \mathbf{C}$, and assume $C_2 \prec C \prec C_1$. Hence $C(1/3,2/3) = 1/3$ and $C(2/3,1/3) = 0$, thus $C \in \mathbf{C}_1 \cap \mathbf{C}_2$. Now let $C \in \mathbf{C}_1 \cap \mathbf{C}_2$. An upper bound for any copula in $\mathbf{C}_1$ is C_1 (Nelsen,1999; Theorem 3.2.2), but C_1 is an element of both $\mathbf{C}_1$ and $\mathbf{C}_2$, hence it is the least upper bound for $\mathbf{C}_1 \cap \mathbf{C}_2$. Similarly, a lower bound for any copula in $\mathbf{C}_2$ is C_2, but C_2 is an element of both $\mathbf{C}_1$ and $\mathbf{C}_2$, hence it is the greatest lower bound for $\mathbf{C}_1 \cap \mathbf{C}_2$. The case $C \in \mathbf{C}_3 \cap \mathbf{C}_4$ is similar and omitted.
(c) Since $C_1(u,u) = C_2(u,u) = (u - 1/3)^+ + (u - 2/3)^+$, it follows from (b) that for every $C \in \mathbf{C}_1 \cap \mathbf{C}_2$, $C(u,u) = (u - 1/3)^+ + (u - 2/3)^+$. The same result holds for $C \in \mathbf{C}_3 \cap \mathbf{C}_4$ since $C^T(u,u) = C(u,u)$. □

Note that as a consequence of Theorem 3.1, one-third of the probability mass associated with any maximally nonexchangeable copula in $\mathbf{C}_1 \cap \mathbf{C}_2$ is contained in each of the squares $[0,1/3] \times [1/3,2/3]$, $[1/3,2/3] \times [2/3,1]$, and $[2/3,1] \times [0,1/3]$; and one-third of the probability mass associated with any maximally nonexchangeable copula in $\mathbf{C}_3 \cap \mathbf{C}_4$ is contained in each of the squares $[0,1/3] \times [2/3,1]$, $[1/3,2/3] \times [0,1/3]$, and $[2/3,1] \times [1/3,2/3]$.

If continuous random variables X and Y have copula C, the population version of the measure of association known as Spearman's rho, which we denote by $\rho_{X,Y}$ or $\rho(C)$, is given by (Schweizer and Wolff, 1981; Nelsen, 1999)

$$\rho_{X,Y} = \rho(C) = 12\int_0^1\int_0^1 C(u,v)\,du\,dv - 3.$$

Unlike the case of exchangeable random variables, where nothing can be said about the magnitude of rho (or other measures of association, for that matter), maximally nonexchangeable random variables must be negatively correlated, in the sense of Spearman's rho:

Corollary 3.2. *Let X and Y be maximally nonexchangeable continuous random variables. If* $\rho_{X,Y}$ *denotes Spearman's rho for X and Y, then* $\rho_{X,Y} \in [-5/9, -1/3]$.

Proof. Let C denote the copula of X and Y. Then $\delta(C) = 1$, and as a consequence of Theorem 3.1(b), we need only evaluate $\rho(C_1)$ and $\rho(C_2)$, since $C \prec C'$ implies $\rho(C) \le \rho(C')$, and $\rho(C) = \rho(C^T)$. Elementary calculus readily establishes that $\rho(C_1) = -1/3$ and $\rho(C_2) = -5/9$. □

Other measures of association yield different results. The population version of the medial correlation coefficient (or Blomqvist's beta) for continuous random variables with copula C is given by $\beta(C) = 4C(1/2,1/2) - 1$. When C is a maximally nonexchangeable copula we have $C(1/2,1/2) = 1/6$ [see Theorem 3.1(c)], and hence $\beta(C) = -1/3$ for any pair of maximally nonexchangeable continuous random variables. However, maximally nonexchangeable random variables may be positively correlated when we employ other measures of association. For continuous random variables X and Y with copula C, the population version of Kendall's tau, which we denote by $\tau_{X,Y}$ or $\tau(C)$, is given by (Nelsen, 1999)

$$\tau_{X,Y} = \tau(C) = 1 - 4\int_0^1\int_0^1 \frac{\partial}{\partial u}C(u,v)\frac{\partial}{\partial v}C(u,v)\,du dv.$$

Since $C \prec C'$ implies $\tau(C) \le \tau(C')$ and $\tau(C) = \tau(C^T)$, and since $\tau(C_1) = +1/9$ and $\tau(C_2) = -5/9$, we have $\tau_{X,Y} \in [-5/9, +1/9]$ when X and Y are maximally nonexchangeable continuous random variables.

4. Concluding Remarks

The degree of nonexchangeability $\delta(C)$ in (5) is a normalized L_∞ distance between the graphs of $z = C(u,v)$ and $z = C(v,u)$. Clearly any other suitably normalized measure of distance between the surfaces $z = C(u,v)$ and $z = C(v,u)$ will yield a measure of nonexchangeability. For example, the L_p distance, $1 \le p < \infty$, given by

$$\left(k_p\int_0^1\int_0^1 |C(u,v) - C(v,u)|^p\,du dv\right)^{1/p},$$

where k_p is chosen so that the distance lies in [0,1], is such a measure.

Other forms of asymmetry can be treated in a fashion similar to nonexchangeability. For example, a copula C is *radially symmetric* (the form of symmetry exhibited by elliptically contoured distributions, among others) if $C(u,v) = \overline{C}(u,v)$ for all u,v in $\mathbf{I}$, where $\overline{C}$ denotes the survival copula associated with C, i.e., $\overline{C}(u,v) = u + v - 1 + C(1-u,1-v)$. Thus the quantity $\left|C(u,v) - \overline{C}(u,v)\right|$ measures radial asymmetry. Using methods analogous to those introduced in Section 2, it can be shown that

$$\left|C(u,v) - \overline{C}(u,v)\right| \le \min\left(u, v, 1-u, 1-v, \max\left(|u-v|, |u+v-1|\right)\right),$$

and that the right-hand side of this inequality attains a maximum of $1/3$ at the points $(1/3,1/3)$, $(1/3,2/3)$, $(2/3,1/3)$ and $(2/3,2/3)$. Hence, just as in the case of nonexchangeability,

$$\sup_{u,v \in \mathbf{I}} \left| C(u,v) - \overline{C}(u,v) \right| \leq \frac{1}{3},$$

and the inequality is best-possible since $\sup_{u,v \in \mathbf{I}} \left| C_1(u,v) - \overline{C}_1(u,v) \right|$ = $\sup_{u,v \in \mathbf{I}} \left| C_2(u,v) - \overline{C}_2(u,v) \right|$ = $1/3$, where C_1 and C_2 are given in (1) and (2). Thus a “degree of radial asymmetry” can be defined analogous to $\delta(C)$ in (5), and copulas that achieve the maximum value of this measure can be studied.

References

1. Galambos J (1982) Exchangeability, in Encyclopedia of Statistical Sciences, Vol. 2, Kotz S and Johnson NL, editors. John Wiley & Sons, New York, 573-577
2. Lancaster HO (1963) Correlation and complete dependence of random variables. Ann. Math. Statist. **34**, 1315-1321
3. Nelsen RB (1999) An Introduction to Copulas. Springer, New York
4. Schweizer B, Wolff EF (1981) On nonparametric measures of dependence for random variables. Ann. Statist. **9**, 870-885