Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Statistics & Probability Letters 65 (2003) 263–268
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Kendall distribution functions[☆]

Roger B. Nelsen[a,*], José Juan Quesada-Molina[b], José Antonio Rodríguez-Lallena[c], Manuel Úbeda-Flores[c]

[a] Department of Mathematical Sciences, Lewis & Clark College, Portland, OR 97219, USA
[b] Departamento de Matemática Aplicada, Universidad de Granada, Granada, Spain
[c] Departamento de Estadística y Matemática Aplicada, Universidad de Almería, Almería, Spain

Received June 2002

Abstract

If X and Y are continuous random variables with joint distribution function H, then the Kendall distribution function of (X, Y) is the distribution function of the random variable $H(X, Y)$. Kendall distribution functions arise in the study of stochastic orderings of random vectors. In this paper we study various properties of Kendall distribution functions for both populations and samples.

MSC: primary 60E05; secondary 62H05; 62E10

Keywords: Copulas; Distribution functions; Kendall's tau; Stochastic orderings

1. Introduction

The *Kendall stochastic ordering* $\prec_K$ of continuous random vectors (X_1, Y_1) and (X_2, Y_2), with distribution functions H_1 and H_2, respectively, is defined as $(X_1, Y_1) \prec_K (X_2, Y_2)$ if and only if $H_1(X_1, Y_1) \prec_{\text{st}} H_2(X_2, Y_2)$, where $\prec_{\text{st}}$ denotes the ordinary stochastic ordering for (one-dimensional) random variables (Capéraà et al., 1997). If we let K_i denote the distribution function of the random variable $H_i(X_i, Y_i)$, then

$$(X_1, Y_1) \prec_K (X_2, Y_2) \quad \text{if and only if } K_1(t) \geqslant K_2(t) \quad \text{for all } t \text{ in } \mathbf{R}. \tag{1}$$

Kendall's name is associated with this ordering since the population version of the measure of association known as Kendall's tau can be expressed (Genest and Rivest, 1993, 2001) as

[☆] Research supported by the Spanish C. I. C. Y. T. Grant (PB98-1010), the Junta de Andalucía, and the institutions of the authors.

[*] Corresponding author. Fax: +1-503-768-7668.
E-mail address: nelsen@lclark.edu (R.B. Nelsen).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2003.08.002

$\tau(X,Y) = 3 - 4\int_0^1 K(t)\,dt$. While in Genest and Rivest (1993) and Capéraà et al. (1997) the distribution function K (of $H(X,Y)$) is called a "decomposition of Kendall's tau," we shall call it the *Kendall distribution function of* (X,Y). This function also appears in Genest and Rivest (2001) and Nelsen et al. (2001) as a bivariate probability integral transform.

In this paper, we study various properties of Kendall distribution functions and their consequences. After some preliminaries concerning copulas, we use the Bertino family of copulas to show that every distribution function satisfying the properties of a Kendall distribution function is the Kendall distribution function of some pair of random variables. We also examine the equivalence relation on the set of copulas induced by Kendall distribution functions. In the final section, we study empirical Kendall distribution functions and their relationships to the ordinary sample version of Kendall's tau.

2. Preliminaries

As is often the case when dealing with bivariate distributions, the use of copulas simplifies matters. A (two-dimensional) *copula* is a function $C : \mathbf{I}^2 \to \mathbf{I} = [0,1]$ which satisfies (a) $C(t,0) = C(0,t) = 0$ and $C(t,1) = C(1,t) = t$ for all t in $\mathbf{I}$, and (b) $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geqslant 0$ for all u_1,u_2,v_1,v_2 in $\mathbf{I}$ such that $u_1 \leqslant u_2$ and $v_1 \leqslant v_2$. Equivalently, a copula is the restriction to $\mathbf{I}^2$ of a continuous bivariate distribution whose margins are uniform on $\mathbf{I}$. Recall from *Sklar's Theorem* (Sklar, 1959) that any bivariate distribution function H with marginal distribution functions F and G can be written as $H(x,y) = C(F(x),G(y))$, where C is a copula. M and W denote the copulas for the Fréchet–Hoeffding upper and lower bounds, respectively, which for any copula C satisfy $W(u,v) = \max(u+v-1,0) \leqslant C(u,v) \leqslant \min(u,v) = M(u,v)$ for all u,v in $\mathbf{I}$. For continuous random variables X and Y, each one is almost surely an increasing (decreasing) function of the other if and only if their copula is $M(W)$. The copula of any pair of independent continuous random variables is $\Pi(u,v) = uv$. For further details, see Nelsen (1999).

As a consequence of Sklar's Theorem, the Kendall distribution function K of (X,Y) depends only on the copula C of X and Y, since if $\mu_H(\mu_C)$ denotes the measure induced on $\mathbf{R}^2$ $(\mathbf{I}^2)$ by $H(C)$, then for any t in $\mathbf{I}$,

$$K(t) = \mu_H(\{(x,y) \in \mathbf{R}^2 | H(x,y) \leqslant t\}) = \mu_C(\{(u,v) \in \mathbf{I}^2 | C(u,v) \leqslant t\}) \tag{2}$$

(Nelsen et al., 2001). So if U and V are random variables uniformly distributed on $\mathbf{I}$ whose joint distribution function is C, the copula of X and Y, then (X,Y) and (U,V) have the same Kendall distribution function. We also note that $\tau(X,Y) = 4E[C(U,V)] - 1$ (Genest and Rivest, 1993). As a consequence, we will often refer to the "Kendall distribution function of C," and write K_C for K in (2). If the copulas of (X_1,Y_1) and (X_2,Y_2) are C_1 and C_2, respectively, we will rewrite the left side of (1) as $C_1 \prec_K C_2$, thus ordering the set of copulas via their Kendall distribution functions.

3. Basic properties of Kendall distribution functions

As a consequence of (2), the Kendall stochastic ordering $\prec_K$ in (1) is a "nonparametric" ordering, in the sense that it depends only on the copulas C_1 and C_2 of (X_1,Y_1) and (X_2,Y_2), respectively. Another such ordering is the *positive quadrant dependence* ordering $\prec_{\text{pqd}}$: $(X_1,Y_1) \prec_{\text{pqd}} (X_2,Y_2)$

if and only if $C_1(u,v) \leqslant C_2(u,v)$ on $\mathbf{I}^2$ (so named since (X,Y) is positive quadrant dependent (PQD) if $C \geqslant \Pi$). In spite of the apparent similarity in form of these two orders ($C_1(U,V) \prec_{\mathrm{st}} C_2(U,V)$ for $\prec_K$, and $C_1(u,v) \leqslant C_2(u,v)$ for $\prec_{\mathrm{pqd}}$), it is known (Capéraà et al., 1997) that $\prec_K$ does not imply $\prec_{\mathrm{pqd}}$. We now show that $\prec_{\mathrm{pqd}}$ does not imply $\prec_K$.

Example 3.1. Let C be the copula given by $C(u,v) = \min(M(u,v), 1/4 + W(u,v))$, that is, C is the copula whose probability mass is uniformly distributed on three line segments in $\mathbf{I}^2$, one from $(0,0)$ to $(1/4,1/4)$, one from $(1/4,3/4)$ to $(3/4,1/4)$, and one from $(3/4,3/4)$ to $(1,1)$. Then $\Pi \prec_{\mathrm{pqd}} C$. However, $K_C(t) = \max(t, (3/4)\lfloor t + 3/4 \rfloor)$ and $K_\Pi(t) = t - t\ln t$ (Nelsen et al., 2001), so that $K_C(1/e) = 3/4 > 2/e = K_\Pi(1/e)$, i.e., it is not true that $\Pi \prec_K C$.

The Kendall stochastic ordering induces a positive dependence property (similar to PQD) for bivariate vectors known as *positive K-dependence* (PKD) (Averous and Dortet-Bernadet, 2002): (X,Y) is PKD if their copula C satisfies $C \succ_K \Pi$, or equivalently, if $K_C(t) \leqslant K_\Pi(t) = t - t\ln t$. Example 3.1 illustrates that PQD does not imply PKD; the following example shows that PKD does not imply PQD.

Example 3.2. Let C be the copula given by $C(u,v) = \min(M(u,v), \max(0, u-1/3, v-1/3, u+v-2/3))$, that is, C is the copula whose probability mass is uniformly distributed on three line segments in $\mathbf{I}^2$, one from $(0,1/3)$ to $(1/3,2/3)$, one from $(1/3,0)$ to $(2/3,1/3)$, and one from $(2/3,2/3)$ to $(1,1)$. Then $K_C(t) = \min(2t, \max(t, 2/3))$, and thus $K_C(t) \leqslant K_\Pi(t)$. However, $C(1/3,1/3) = 0 < 1/9 = \Pi(1/3,1/3)$, i.e., it is not true that $C \succ_{\mathrm{pqd}} \Pi$.

Since every copula C satisfies $W \leqslant C \leqslant M$ on $\mathbf{I}^2$, the Frechet–Hoeffding bounds M and W are the upper and lower bounds for the set of copulas with respect to the PQD ordering. The same is true for the Kendall ordering: $W \prec_K C \prec_K M$ for every copula C, or equivalently, $t = K_M(t) \leqslant K_C(t) \leqslant K_W(t) = 1$ for all t in $\mathbf{I}$ (Capéraà et al., 1997). This observation, along with $C(0,0) = 0$, establishes

Theorem 3.1. *Let C be a copula, and K_C its Kendall distribution function. Then* (a) $t \leqslant K_C(t)$ *for all t in* $\mathbf{I}$, *and* (b) $K_C(0^-) = 0$.

We now show that the properties in Theorem 3.1 actually characterize Kendall distributions functions, that is, that if F is *any* right-continuous distribution function which satisfies $t \leqslant F(t)$ on $\mathbf{I}$ and $F(0^-) = 0$, then there exists a copula C such that the Kendall distribution function of C is F, i.e., $K_C = F$. Genest and Rivest (1993) proved this result for distribution functions F satisfying a further condition—$F(t^-) > t$ for t in $(0,1)$—using Archimedean copulas. To provide a construction without the restriction that $F(t^-) > t$ for t in $(0,1)$, we use a family of functions introduced by Bertino (1977). Let $\delta : \mathbf{I} \to \mathbf{I}$ be a function such that $\delta(1) = 1$, $\delta(t) \leqslant t$, and $0 \leqslant \delta(t_2) - \delta(t_1) \leqslant 2(t_2 - t_1)$ for t, t_1 and t_2 in $\mathbf{I}$ with $t_1 < t_2$; and for u, v in $\mathbf{I}$ set

$$B_\delta(u,v) = \min(u,v) - \min(s - \delta(s) | \min(u,v) \leqslant s \leqslant \max(u,v)). \tag{3}$$

Each B_δ is a copula; for t in $\mathbf{I}$, $B_\delta(t,t) = \delta(t)$; if $\delta(t) = t$, $B_\delta = M$; and if $\delta(t) = \max(2t - 1, 0)$, $B_\delta = W$ (Fredricks and Nelsen, 2002). We also let $\delta^{(-1)}$ denote the cadlag inverse of δ, i.e.,

$\delta^{(-1)}(t) = \sup\{u|\delta(u) \leqslant t\}$ for t in $\mathbf{I}$. The following lemma presents the Kendall distribution function of the Bertino copula B_δ.

Lemma 3.2. *Let B_δ be the copula given by (3). Then for t in $\mathbf{I}$, $K_{B_\delta}(t) = 2\delta^{(-1)}(t) - t$.*

Proof. Let B_δ be given by (3), fix t in $\mathbf{I}$, and set $S_B = \{(u,v) \in \mathbf{I}^2 | B_\delta(u,v) \leqslant t\}$ and $S_M = \{(u,v) \in \mathbf{I}^2 | \min(u,v) \leqslant \delta^{(-1)}(t)\}$. Since $\mu_{B_\delta}(S_M) = 2\delta^{(-1)}(t) - t$ (Nelsen et al., 2001), we need only show that $\mu_{B_\delta}(S_B) = \mu_{B_\delta}(S_M)$, which we accomplish by showing that $S_B \subseteq S_M$ and $\mu_{B_\delta}(S_M \setminus S_B) = 0$. But $B_\delta(u,v) \leqslant t$ implies $\delta(\min(u,v)) \leqslant t$, whence $\min(u,v) \leqslant \delta^{(-1)}(t)$, from which it follows that $S_B \subseteq S_M$. Now assume $u \leqslant v$ (the case $v < u$ is similar and is omitted). Since the portion of the boundary of $\{(u,v) \in \mathbf{I}^2 | B_\delta(u,v) < t\}$ contained in $(0,1)^2$ is a nonincreasing set (a set S in $\mathbf{R}^2$ is *nonincreasing* if for any (a,b) and (c,d) in S, $a < c$ implies $b \geqslant d$), to show that $\mu_{B_\delta}(S_M \setminus S_B) = 0$ it will suffice to show that $\mu_{B_\delta}([u, \delta^{(-1)}(t)] \times [v,1]) = 0$ for all (u,v) such that $u \in [t, \delta^{(-1)}(t)]$ and $v = \min(s|B_\delta(u,s) = t)$. With such u and v, $B_\delta(u,v) = t$, $[u, \delta^{(-1)}(t)] \times [v,1] \subseteq [u,v] \times [v,1]$, and we claim that $\mu_{B_\delta}([u,v] \times [v,1]) = 0$. First, $\mu_{B_\delta}([u,v] \times [v,1]) = v - u - \delta(v) + t$, and from the definition of v, $B_\delta(u,r) < t$ for r in $[u,v)$, so that $\min(s - \delta(s)|s \in [u,r]) > u - t$. Hence $r - \delta(r) > u - t$ for r in $[u,v)$. But since $B_\delta(u,v) = t$, $\min(s - \delta(s)|s \in [u,v]) = u - t$, so that $v - \delta(v) = u - t$. Hence $\mu_{B_\delta}([u,v] \times [v,1]) = 0$, which completes the proof. □

We now have:

Theorem 3.3. *Let F be a right-continuous distribution function such that $F(0^-) = 0$ and $F(t) \geqslant t$ for all t in $\mathbf{I}$. Then there exists a copula C such that $K_C(t) = F(t)$ for all t.*

Proof. Let F satisfy the hypotheses above, and let α and δ be the functions defined on $\mathbf{I}$ by $\alpha(t) = [t + F(t)]/2$ and $\delta(t) = \sup\{s \in \mathbf{I} | \alpha(s) \leqslant t\}$. It is immediate that $\delta(1) = 1$, $\delta(t) \leqslant t$, $\alpha(\delta(t)) \geqslant t$, and $\delta(t_1) \leqslant \delta(t_2)$ for t_1 and t_2 in $\mathbf{I}$ with $t_1 < t_2$. Furthermore, $\delta^{(-1)}(t) = \alpha(t)$ for t in $\mathbf{I}$. If $\delta(t_1) = \delta(t_2)$, then it is immediate that $\delta(t_2) - \delta(t_1) \leqslant 2(t_2 - t_1)$. Suppose that $\delta(t_1) < \delta(t_2)$. If $\delta(t_1) \leqslant r < \delta(t_2)$, then $\alpha(r) \leqslant t_2$. Hence $r - \delta(t_1) \leqslant r - \delta(t_1) + F(r) - F(\delta(t_1)) = 2[\alpha(r) - \alpha(\delta(t_1))] \leqslant 2(t_2 - t_1)$, and thus $\delta(t_2) - \delta(t_1) \leqslant 2(t_2 - t_1)$. Therefore, B_δ, as given by (3), is a copula; and $K_{B_\delta}(t) = 2\delta^{(-1)}(t) - t = 2\alpha(t) - t = F(t)$. □

Kendall distribution functions induce an equivalence relation $\equiv_K$ on the set $\mathbf{C}$ of copulas: if C_1 and C_2 are copulas with Kendall distribution functions K_1 and K_2, respectively, then $C_1 \equiv_K C_2$ if and only if $K_1(t) = K_2(t)$ for all t in $\mathbf{I}$. The following corollary illustrates that the set of Bertino copulas is a system of distinct representatives for the equivalence classes of $\equiv_K$.

Corollary 3.4. *Each equivalence class of the equivalence relation $\equiv_K$ on $\mathbf{C}$ contains a unique Bertino copula.*

Proof. The proof of Theorem 3.3 shows that each equivalence class contains at least one Bertino copula. If both B_δ and B_γ belong to the same equivalence class, then $2\delta^{(-1)}(t) - t = 2\gamma^{(-1)}(t) - t$ for all t in $\mathbf{I}$, from which it follows that $\delta = \gamma$, hence $B_\delta = B_\gamma$. □

Remark. Theorem 3.3 can also be proved by another method, showing that for each distribution function F satisfying the hypotheses of the theorem, there is an associative copula, i.e., a copula C such that $C(C(u,v),w) = C(u,C(v,w))$ for u,v,w in $\mathbf{I}$, such that the Kendall distribution function of C coincides with F. As a consequence, each equivalence class of $\equiv_K$ contains a unique associative copula (see Úbeda Flores (2001) for details).

As noted in Section 1, the population version of the measure of association known as Kendall's tau for continuous random variables X and Y, whose copula is C, is expressible in terms of the Kendall distribution function of C: $\tau(X,Y) = \tau_C = 3 - 4\int_0^1 K(t)\,dt$. We use this result to show that the equivalence classes of $\equiv_K$ containing the copulas M and W for the Fréchet–Hoeffding bounds are singletons. Let K_1, K_2 and τ_1, τ_2 denote the Kendall distribution functions and the values of Kendall's tau associated with copulas C_1 and C_2, respectively.

Theorem 3.5. *If $C_1 \equiv_K M$, then $C_1 = M$; and if $C_2 \equiv_K W$, then $C_2 = W$.*

Proof. Suppose $C_1 \equiv_K M$ and $C_2 \equiv_K W$. Then $K_1(t) = t$ and $K_2(t) = 1$ on $\mathbf{I}$, and hence $\tau_1 = 1$ and $\tau_2 = -1$. But M and W are the unique copulas for which Kendall's tau equals 1 and -1, respectively, hence $C_1 = M$ and $C_2 = W$. □

With the notation preceding the above proof, note that if $C_1 \prec_K C_2$, then $\tau_1 \leqslant \tau_2$ (Capéraà et al., 1997). However, the reverse implication does not hold. For example, if C_1 is Π and C_2 is the copula from Example 3.1, then $\tau_1 = 0 < 1/2 = \tau_2$, yet $C_1 \prec_K C_2$ does not hold.

4. The empirical Kendall distribution function

Let $\{(x_k, y_k)\}_{k=1}^n$ denote a sample of size n from a continuous distribution, and let $x_{(i)}$ and $y_{(j)}$, $1 \leqslant i,j \leqslant n$, denote the order statistics from the sample. Then the *empirical copula* C' and the *empirical Kendall distribution function* $K_{C'}$ are defined as

$$C'(i/n, j/n) = (1/n)(\text{number of points}(x_k, y_k) \text{ such that } x_k \leqslant x_{(i)} \text{ and } y_k \leqslant y_{(j)});$$

and for all t,

$$K_{C'}(t) = (1/n)(\text{number of pairs } (x_k, y_k) \text{ whose ranks } (i,j) \text{ satisfy } C'(i/n, j/n) \leqslant t).$$

A pair (x_k, y_k) and (x_m, y_m) of points in the sample are *concordant* if $x_k < x_m$ and $y_k < y_m$ or $x_k > x_m$ and $y_k > y_m$; and *discordant* if $x_k < x_m$ and $y_k > y_m$ or $x_k > x_m$ and $y_k < y_m$. We let t_n denote the value of Kendall's tau for the sample, i.e.,

$$t_n = [(\text{number of concordant pairs}) - (\text{number of discordant pairs})] \Big/ \binom{n}{2}. \tag{4}$$

Analogous to $\tau = 4E[C(U,V)] - 1$ and $\tau = 3 - 4\int_0^1 K(t)\,dt$ for the population value of Kendall's tau, we have

Theorem 4.1. *Let* $\{(x_k, y_k)\}_{k=1}^{n}$ *denote a sample of size n from a continuous distribution, and let* t_n *denote the value of Kendall's tau for the sample. Then*

$$\text{(a)} \quad t_n = \frac{4}{n-1} \sum C'(i/n, j/n) - \frac{n+3}{n-1},$$

where the sum is over the n points in the sample; and

$$\text{(b)} \quad t_n = 3 - \frac{4n}{n-1} \int_0^1 K_{C'}(t)\,\mathrm{d}t.$$

Proof. Part (a) readily follows from the observations that $\sum C'(i/n, j/n) = 1 +$(number of concordant pairs)$/n$ and (4), $t_n = (4/n(n-1))$(number of concordant pairs)-1. For part (b), we note that since $K_{C'}$ is a step function, $\int_0^1 K_{C'}(t)\,\mathrm{d}t = (1/n)\sum_{m=1}^{n-1} K_{C'}(m/n)$. In this sum, a sample point whose ranks are (i, j) is counted $n-m$ times when $C'(i/n, j/n) = m/n$, and thus $\sum_{m=1}^{n-1} K_{C'}(m/n) = (1/n)\sum(n - nC'(i/n, j/n))$, where the last sum is over the n points in the sample. Invoking the result in (a) and simple algebra establishes (b). □

References

Averous, J., Dortet-Bernadet, J.-L., 2002. Dependence for Archimedean copulas and aging properties of their generating functions, preprint.

Bertino, S., 1977. Sulla dissomiglianza tra mutabili cicliche. Metron 35, 53–88.

Capéraà, P., Fougères, A.-L., Genest, C., 1997. A stochastic ordering based on a decomposition of Kendall's tau. In: Beneš, V., Štěpán, J. (Eds.), Distributions with Given Marginals and Moment Problems. Kluwer Academic Publishers, Dordrecht, pp. 81–86.

Fredricks, G.A., Nelsen, R.B., 2002. The Bertino family of copulas. In: Cuadras, C., Fortiana, J., Rodríguez Lallena, J.A. (Eds.), Distributions with Given Marginals and Statistical Modelling. Kluwer Academic Publishers, Dordrecht, pp. 81–92.

Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. J. Amer. Statist. Assoc. 88, 1034–1043.

Genest, C., Rivest, L.-P., 2001. On the multivariate probability integral transformation. Statist. Probab. Lett. 53, 391–399.

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

Nelsen, R.B., Quesada Molina, J.J., Rodríguez Lallena, J.A., Úbeda Flores, M., 2001. Distribution functions of copulas: a class of bivariate probability integral transforms. Statist. Probab. Lett. 54, 277–282.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231.

Úbeda Flores, M., 2001. Cópulas y cuasicópulas: interrelaciones y nuevas propiedades. Aplicaciones. Ph.D. Dissertation, Servicio de Publicaciones de la Universidad de Almería, Spain.