On the interrelation between Kendall's τ and Spearman's ρ

Wolfgang Trutschnig
Department for Mathematics
University Salzburg
www.trutschnig.net

International Statistics Festival Ulm
Ulm, 2018-09-18

Introductory remarks & notation:

- Suppose that $X, Y \sim \mathcal{U}(0,1)$. Then the distribution function A of (X,Y), restricted to $[0,1]^2$, is called a (two-dimensional) *copula*.
- A probability measure μ on $\mathcal{B}([0,1]^2)$ is called *doubly stochastic* if we have
$$\mu(E \times [0,1]) = \mu([0,1] \times E) = \lambda(E)$$
for all $E \in \mathcal{B}([0,1])$.
- Copulas are distribution functions (restricted to $[0,1]^2$) of doubly stochastic measures.
- $\mathcal{C}$ denotes the family of all (two-dimensional) copulas, $\mathcal{P}_\mathcal{C}$ the family of all doubly stochastic measures ($A \in \mathcal{C} \longleftrightarrow \mu_A \in \mathcal{P}_\mathcal{C}$).
- Examples: $M(x,y) = \min\{x,y\}, \Pi(x,y) = xy, W(x,y) = \max\{x+y-1, 0\}$ are copulas.

- In 1951 Fréchet studied the question, how the family $\mathcal{F}_{F,G}$ of ALL joint distribution functions H that have F and G as marginals looks like.

Theorem (Two-dimensional version of Sklar's theorem, 1959)

Suppose that X and Y are random variables with continuous distribution functions F and G, and let H denote their joint distribution function. Then there exists a unique copula A such that

$$H(x,y) = A\Big(F(x), G(y)\Big) \tag{1}$$

holds for all $x, y \in \mathbb{R}$. In other words: Copulas are the link between multivariate distribution functions and their marginals.

- If T and S are strictly increasing transformations on $Supp(\mathbb{P}^X)$ and $Supp(\mathbb{P}^Y)$ resp., then A is also the copula of $(T \circ X, S \circ Y)$.
- Consequence: All scale-invariant dependence between X and Y is captured by the copula A underlying (X, Y).
- Copula pop-up naturally in various problems.

Two examples where copulas naturally pop-up

- Consider the following situation:
- $X \sim F$ and $Y \sim G$ describe default times of firms (or obligors) or lifetimes of electronic components; **we do not know the distribution of** (X, Y).
- In risk management one key quantity is the probability of a joint default, i.e. $\mathbb{P}(X = Y)$.
- As before, let $\mathcal{F}_{F,G}$ denote the Fréchet class of F, G, i.e. the class of all two-dimensional d.f. H having F and G as marginals.
- We want to know the worst-case-scenario, i.e. the quantity

$$\sup_{H \in \mathcal{F}_{F,G}} \mu_H(\Gamma(id)). \tag{2}$$

- Setting $T = G \circ F^-$ eq. (2) simplifies to

$$\sup_{H \in \mathcal{F}_{F,G}} \mu_H(\Gamma(id)) = \sup_{A \in \mathcal{C}} \mu_A(\Gamma(T)) =: \overline{w}_T. \tag{3}$$

- The following result can be shown (directly or via optimal transport).

Theorem (Mroz, T., Fernández-Sánchez, ?)

Suppose that $T : [0,1] \to [0,1]$ *is non-decreasing. Then there exists a copula* A *such that*

$$\overline{w}_T = \mu_A(\Gamma(T)) = \int_{[0,1]} \min\left\{T'(x), 1\right\} d\lambda(x). \tag{4}$$

- For an arbitrary, measurable $T : [0,1] \to [0,1]$ we get the same formula with T' replaced by F_T' (the derivative of the distribution function F_T' of T).

- The two most important *concordance measures* - Kendall's τ and Spearman's ρ - only depend on the underlying copula.
- As before: $X \sim F, Y \sim G$.
- Kendall's τ of (X, Y) is defined by

$$\tau(X, Y) = \mathbb{P}\Big((X_1 - X_2)(Y_1 - Y_2) > 0\Big) - \mathbb{P}\Big((X_1 - X_2)(Y_1 - Y_2) < 0\Big),$$

  whereby (X_1, Y_1), (X_2, Y_2) are samples from (X, Y).
- Spearman's ρ is the Pearson correlation coefficient of the random variables $U := F \circ X$ and $V := G \circ Y$, i.e.

$$\rho(X, Y) = 12\Big(\mathbb{E}(UV) - \tfrac{1}{4}\Big).$$

- Then we have

$$\tau(X, Y) = 4 \int_{[0,1]^2} A d\mu_A - 1, \quad \rho(X, Y) = 12 \int_{[0,1]^2} A d\lambda_2 - 3 \qquad (5)$$

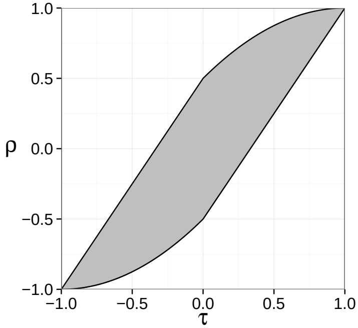


Figure: Classical τ-ρ-region Ω_0 determined by the inequalities by Daniels and Durbin & Stuart.

- Known since the 1950s:
- Daniels' inequality (JRSS-B, 1950):

$$|3\tau - 2\rho| \leq 1$$

- Durbin & Stuart's inequality (JRSS-B, 1951):

$$\frac{(1+\tau)^2}{2} - 1 \leq \rho \leq 1 - \frac{(1-\tau)^2}{2}$$

- The inequalities determine the *classical τ-ρ region* Ω_0.
- Ω_0 is convex and compact.

What was known

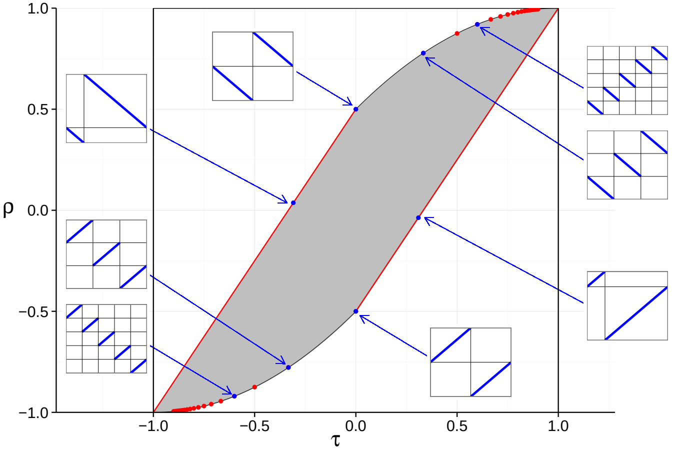
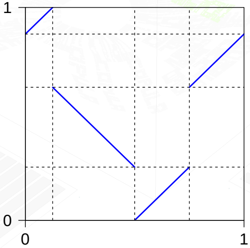


Figure: The region Ω_0, $\pm p_n$ in red, where $p_n := (-1 + \frac{2}{n}, -1 + \frac{2}{n^2})$, $n \geq 2$.

Definition

Shuffle of M $A \in \mathcal{C}$ is called *shuffle of M* if there exists a λ-preserving, piecewise linear function $h : [0,1] \to [0,1]$ with slope ± 1 such that the mass of A is concentrated on the graph of h. A shuffle is called *straight* is all segments have slope $+1$. $\mathcal{C}_{\mathcal{S}^+}$ will denote the family of all straight shuffles.

- The segments do not necessarily have the same length.
- Each straight shuffle can be expressed in terms of a permutation $\pi \in \sigma_n$ and an element $u \in \Delta_n$ (Δ_n...n-dim unit simplex).



Main objective (2014): Try to determine the exact τ-ρ-region

$$\begin{aligned}\Omega &= \Big\{(\tau(X,Y),\rho(X,Y)) : \; X,Y \text{ continuous r.v.}\Big\} \\ &= \Big\{(\tau(A),\rho(A)) : \; A \in \mathcal{C}\Big\} \end{aligned} \tag{6}$$

- Do we have $\Omega = \Omega_0$?
- Determine which distributions yield points on the boundary of Ω.
- Simple idea: Work with straight shuffles since $\mathcal{C}_{\mathcal{S}^+}$ is dense in $(\mathcal{C}, d_\infty)$ and τ, ρ are continuous w.r.t. d_∞

- We studied the classical proofs and more recent ones, and tried to improve them $\rightarrow$ no success.
- We ran numerous simulations (straight shuffles) $\rightarrow$ no examples at which the parabolic inequality is sharp other than $\pm p_n$.
- Points closest to the boundary were all of the from of **prototypes**: '$n-1$ stripes of equal length and one shorter stripe in decreasing order'.
- ...is that it?

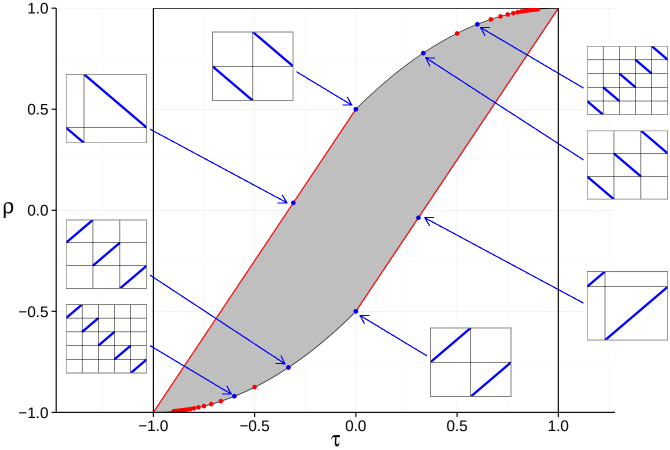


Figure: The region Ω_0, $\pm p_n$ in red, where $p_n := (-1 + \frac{2}{n}, -1 + \frac{2}{n^2})$, $n \geq 2$.

- Define $\Phi_n : [-1 + \frac{2}{n}, -1 + \frac{2}{n-1}] \to [-1, 1]$:

$$\Phi_n(x) = -1 - \frac{4}{n^2} + \frac{3}{n} + \frac{3x}{n} - \frac{n-2}{\sqrt{2}n^2\sqrt{n-1}}(n-2+nx)^{3/2} \tag{7}$$

- Combine the functions Φ_n to one function Φ

$$\Phi(x) = \begin{cases} -1 & \text{if } x = -1, \\ \Phi_n(x) & \text{if } x \in \left[-1 + \frac{2}{n}, -1 + \frac{2}{n-1} \right] \text{ for some } n \geq 2. \end{cases} \tag{8}$$

- Key step: Prove

$$\Omega \subseteq \left\{ (x, y) \in [-1, 1]^2 : \Phi(x) \leq y \leq -\Phi(-x) \right\} =: \Omega_\Phi. \tag{9}$$

via induction on the number of stripes.

- Use a homotopy argument to show $\Omega = \Omega_\Phi$.

What is new

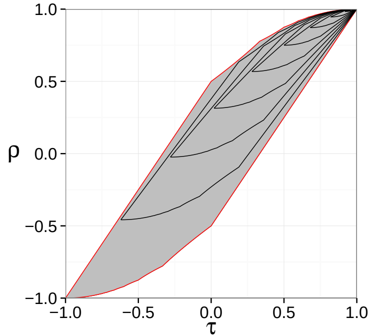


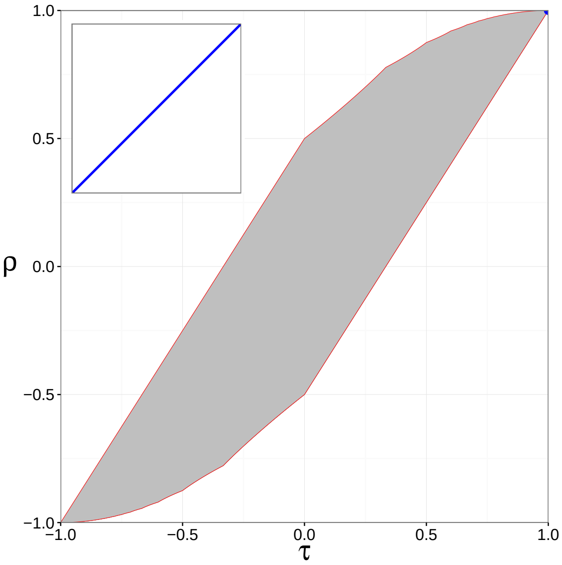
Figure: The curves $\gamma_s(t) = H(s,t)$ for H being the homotopy shrinking $\partial\Omega$ into $(1,1)$.

Theorem (Schreyer, Paulin, T., JRSS-B, 2017)

The exact τ-ρ-region Ω fulfills

$$\Omega = \Big\{(x,y) \in [-1,1]^2 : \Phi(x) \leq y \leq -\Phi(-x)\Big\}.$$

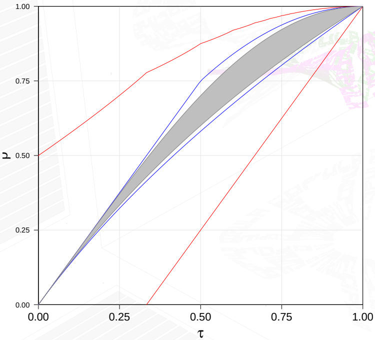
What is new



What is new

- Durbin & Stuart's inequality is not sharp outside the points $\pm p_n$.
- Ω is not convex.
- For every point $(x, y) \in \Omega$ there is a shuffle A of M such that we have $(\tau(A), \rho(A)) = (x, y)$.
- In other words: For each $(x, y) \in \Omega$ there exist (mutually) completely dependent random variables X, Y with $(\tau(X, Y), \rho(X, Y)) = (x, y)$.
- Complete dependence everywhere in Ω.

Extreme-Value Copulas

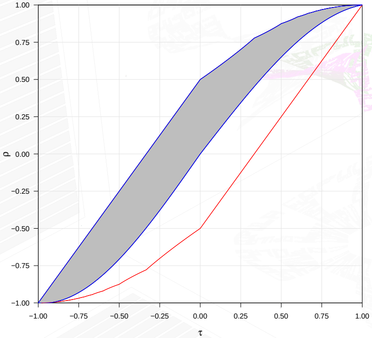


- Red: boundary of Ω
- Blue: Hutchinson-Lai inequ. $-1+\sqrt{1+3\tau(A_a)} \le \rho(A_a) \le \min\left\{\frac{3\tau(A_a)}{2}, 2\tau(A_a)-\tau(A_a)^2\right\}$
- Gray: conjectured exact τ-ρ-region for EVC

Conjectured sharp inequalities for EVCs:

$$\frac{3\tau(A_a)}{2+\tau(A_a)} \le \rho(A_a) \le \frac{3\tau(A_a)}{2+\tau(A_a)^3} - \frac{1}{3}(1-\tau(A_a))^2\tau(A_a)^4$$

Archimedean Copulas



- Red: boundary of Ω
- Gray: conjectured exact τ-ρ-region for Arch. Copulas
- Upper boundary coincides wit the one of Ω

Conjectured lower bound:

$$\rho \geq \begin{cases} \frac{7\tau - 2\tau^3}{5}, & \text{if } \tau \in [0,1], \\ \frac{31\tau - 11\tau^3}{20}, & \text{if } \tau \in [-1,0]. \end{cases}$$