

Un contre-exemple à une conjecture de Hutchinson et Lai

A counterexample to a conjecture of Hutchinson and Lai

Patrick Munroe^a, Thomas Ransford^a, Christian Genest^{a,1}

^a*Département de mathématiques et de statistique, Université Laval,
1045, avenue de la Médecine, Québec (Québec), Canada G1V 0A6*

Abstract

En 1990, Hutchinson et Lai ont conjecturé que si un couple aléatoire (X, Y) est stochastiquement croissant en X et en Y , le rho de Spearman et le tau de Kendall sont tels que $1 + 3\tau \leq (\rho + 1)^2$. Cette conjecture est réfutée.

In 1990, Hutchinson and Lai conjectured that if a random pair (X, Y) is stochastically increasing in X and Y , then Spearman's rho and Kendall's tau are such that $1 + 3\tau \leq (\rho + 1)^2$. This conjecture is disproved.

Keywords: Bounds, copula, Kendall's τ , positive regression dependence, Spearman's ρ , stochastic increasingness.

2000 MSC: 62H20, 60E15

1. Introduction

Soit (X, Y) un couple de variables aléatoires continues dont la loi $H(x, y) = \Pr(X \leq x, Y \leq y)$ a pour marges $F(x) = \Pr(X \leq x)$ et $G(y) = \Pr(Y \leq y)$, où $x, y \in \mathbb{R}$. Pour mesurer le degré de dépendance entre X et Y , on fait

*. Corresponding author

Email addresses: `Patrick.Munroe.1@ulaval.ca` (Patrick Munroe),
`Thomas.Ransford@mat.ulaval.ca` (Thomas Ransford),
`Christian.Genest@mat.ulaval.ca` (Christian Genest)

souvent appel au rho de Spearman ou au tau de Kendall, définis par

$$\rho = -3 + 12 \iint F(x)G(y) dH(x, y), \quad \tau = -1 + 4 \iint H(x, y) dH(x, y).$$

La relation entre ρ et τ a fait l'objet de divers travaux depuis que Daniels <1950, 1951> et Durbin & Stuart <1951> ont montré que

$$\tau - (1 - \tau^2) \leq 3\tau - 2\rho \leq \tau + (1 - \tau^2).$$

Ces inégalités, dont Genest & Nešlehová <2009> ont récemment fourni une démonstration élémentaire, sont les meilleures possibles. Elles ne peuvent être améliorées que si des conditions particulières sont imposées sur la nature de la dépendance entre X et Y . Capéraà & Genest <1993> ont prouvé entre autres que $\rho \geq \tau \geq 0$ lorsque X est à la fois LTD et RTI par rapport à Y au sens de Lehmann <1966>.

Dans le cas où la loi de (X, Y) est stochastiquement croissante, des calculs empiriques ont amené Hutchinson & Lai <1990> à conjecturer que

$$-1 + \sqrt{1 + 3\tau} \leq \rho \leq \min(3\tau/2, 2\tau - \tau^2). \quad (1)$$

Par croissance stochastique, on entend le fait que les applications $t \mapsto \Pr(X > x|Y = t)$ et $t \mapsto \Pr(Y > y|X = t)$ sont toutes deux croissantes pour tous $x, y \in \mathbb{R}$. Cette condition se vérifie fréquemment en pratique.

Nelsen <1999, Exercice 5.36> a observé que $\rho \not\leq 3\tau/2$, mais les deux autres inégalités de l'énoncé (1) n'ont pu être confirmées ou infirmées à ce jour. De fait, Hürlimann <2003> a validé la conjecture de Hutchinson et Lai dans le cas des lois de valeurs extrêmes.

La présente note établit toutefois qu'en général, $-1 + \sqrt{1 + 3\tau} \not\leq \rho$. Le contre-exemple est fourni par une loi définie pour tous $x, y \in [0, 1]$ par

$$H_\theta(x, y) = x \wedge y \wedge (xy + \theta) = xy + (x \wedge y - xy) \wedge \theta$$

pour un certain $\theta \in [0, 1/4]$. Ici et dans la suite, $a \wedge b = \min(a, b)$. On montre d'abord au §2 que H_θ est une loi stochastiquement croissante pour tout $\theta \in [0, 1/4]$. On vérifie ensuite au §3 que $-1 + \sqrt{1 + 3\tau_\theta} > \rho_\theta$ lorsque $0 < \theta < e^{-12}/4 \approx 1.536 \times 10^{-6}$. D'autres propriétés de la famille de lois (H_θ) sont évoquées en conclusion.

2. H_θ est une copule stochastiquement croissante si $\theta \in [0, 1/4]$

Il s'avère que pour tout $\theta \in [0, 1/4]$, la fonction H_θ est une copule, c'est-à-dire une fonction de répartition à marges uniformes sur l'intervalle $[0, 1]$. Pour s'en assurer, on doit vérifier deux conditions : (i) $H_\theta(x, 0) = H_\theta(0, x) = 0$ et $H_\theta(x, 1) = H_\theta(1, x) = x$ pour tout $x \in [0, 1]$; (ii) $H_\theta(x_2, y_2) - H_\theta(x_2, y_1) - H_\theta(x_1, y_2) + H_\theta(x_1, y_1) \geq 0$ dès que $0 \leq x_1 \leq x_2 \leq 1$ et $0 \leq y_1 \leq y_2 \leq 1$.

On se convainc facilement que la condition (i) est vraie. Pour vérifier la condition (ii), on pose $z_{ij} = x_i \wedge y_j - x_i y_j$ pour $i, j \in \{1, 2\}$ et on montre que

$$\Delta = (z_{22} \wedge \theta) - (z_{21} \wedge \theta) - (z_{12} \wedge \theta) + (z_{11} \wedge \theta) + (x_2 - x_1)(y_2 - y_1) \quad (2)$$

est toujours positif. On considère pour ce faire six cas distincts :

1. Lorsque $z_{11} \geq \theta$ et $z_{22} \geq \theta$, on a

$$\Delta \geq \theta - \theta - \theta + \theta + (x_2 - x_1)(y_2 - y_1) \geq 0.$$

2. Lorsque $z_{11} \geq \theta$, $z_{22} < \theta$ et $x_2 \geq y_2$, on a

$$\Delta \geq z_{22} - z_{21} - \theta + \theta + (x_2 - x_1)(y_2 - y_1) = (1 - x_1)(y_2 - y_1) \geq 0.$$

3. Lorsque $z_{11} \geq \theta$, $z_{22} < \theta$ et $x_2 < y_2$, on a

$$\Delta \geq z_{22} - \theta - z_{12} + \theta + (x_2 - x_1)(y_2 - y_1) = (x_2 - x_1)(1 - y_1) \geq 0.$$

4. Lorsque $z_{11} < \theta$, $z_{22} \geq \theta$ et $x_1 \geq y_1$, on a

$$\Delta \geq \theta - z_{21} - \theta + z_{11} + (x_2 - x_1)(y_2 - y_1) = (x_2 - x_1)y_2 \geq 0.$$

5. Lorsque $z_{11} < \theta$, $z_{22} \geq \theta$ et $x_1 < y_1$, on a

$$\Delta \geq \theta - \theta - z_{12} + z_{11} + (x_2 - x_1)(y_2 - y_1) = x_2(y_2 - y_1) \geq 0.$$

6. Lorsque $z_{11} < \theta$ et $z_{22} < \theta$, on a

$$\begin{aligned} \Delta &\geq z_{22} - z_{21} - z_{12} + z_{11} + (x_2 - x_1)(y_2 - y_1) \\ &= x_2 \wedge y_2 - x_2 \wedge y_1 - x_1 \wedge y_2 + x_1 \wedge y_1 \geq 0. \end{aligned}$$

Puisque $\Delta \geq 0$ dans tous les cas, on conclut que H_θ est bien une copule.

Par ailleurs, au vu du Corollaire 5.2.11 de Nelsen <1999>, H_θ est stochastiquement croissante dans la mesure où les applications $t \mapsto H_\theta(x, t)$ et $t \mapsto H_\theta(t, y)$ sont concaves pour tous $x, y \in \mathbb{R}$. Ceci découle d'usages répétés du simple fait que si f et g sont des fonctions concaves d'une variable réelle, il en va de même de $f + g$ et de $f \wedge g$.

3. Le contre-exemple

On va maintenant montrer que (A) $0 \leq \rho_\theta \leq 12\theta$ et (B) $2\rho_\theta - 3\tau_\theta \leq 12\theta^2 \log(4\theta)$ pour tout $\theta \in [0, 1/4]$. Ceci permet de réfuter l'inégalité $\rho^2 + 2\rho - 3\tau \geq 0$, qui équivaut à celle de gauche dans l'énoncé (1). En effet, H_θ est une loi stochastiquement croissante pour laquelle, en vertu de (A) et (B),

$$\rho_\theta^2 + 2\rho_\theta - 3\tau_\theta \leq 144\theta^2 + 12\theta^2 \log(4\theta) = 12\theta^2 \{12 + \log(4\theta)\}.$$

Le membre de droite étant strictement négatif pour $0 < \theta < e^{-12}/4 \approx 1.536 \times 10^{-6}$, on peut conclure.

Pour démontrer (A), il suffit d'écrire

$$\rho_\theta = 12 \iint \{H_\theta(x, y) - xy\} dx dy = 12 \iint \{(x \wedge y - xy) \wedge \theta\} dx dy$$

et d'observer que la fonction sous l'intégrale de droite varie entre 0 et θ .

La démonstration de (B) est plus délicate. Pour la simplifier, on pose

$$A_\theta(x, y) = H_\theta(x, y) - xy = (x \wedge y - xy) \wedge \theta,$$

de sorte que

$$\rho_\theta = 12 \iint A_\theta(x, y) dx dy.$$

En faisant appel à l'équation 5.1.12 de Nelsen <1999>, on trouve

$$\begin{aligned} \tau_\theta &= 1 - 4 \iint \frac{\partial H_\theta}{\partial x}(x, y) \frac{\partial H_\theta}{\partial y}(x, y) dx dy \\ &= 1 - 4 \iint \left\{ y + \frac{\partial A_\theta}{\partial x}(x, y) \right\} \left\{ x + \frac{\partial A_\theta}{\partial y}(x, y) \right\} dx dy \\ &= -4 \iint \left\{ x \frac{\partial A_\theta}{\partial x}(x, y) + y \frac{\partial A_\theta}{\partial y}(x, y) + \frac{\partial A_\theta}{\partial x}(x, y) \frac{\partial A_\theta}{\partial y}(x, y) \right\} dx dy \\ &= 8 \iint A_\theta(x, y) dx dy - 4 \iint \frac{\partial A_\theta}{\partial x}(x, y) \frac{\partial A_\theta}{\partial y}(x, y) dx dy, \end{aligned}$$

la dernière identité découlant d'une intégration par parties. Par conséquent,

$$2\rho_\theta - 3\tau_\theta = 12 \iint \frac{\partial A_\theta}{\partial x}(x, y) \frac{\partial A_\theta}{\partial y}(x, y) dx dy.$$

Cette expression s'exprime comme la somme de deux termes correspondant aux domaines d'intégration $0 < x < y < 1$ et $0 < y < x < 1$. Par symétrie, ces deux intégrales sont égales. Ainsi,

$$\iint \frac{\partial A_\theta}{\partial x} \frac{\partial A_\theta}{\partial y} dx dy = 2 \int_0^1 \int_0^y \frac{\partial A_\theta}{\partial x} \frac{\partial A_\theta}{\partial y} dx dy.$$

Lorsque $0 < x < y < 1$, on trouve

$$A_\theta(x, y) = \begin{cases} x(1-y) & \text{si } x(1-y) < \theta, \\ \theta & \text{si } x(1-y) > \theta, \end{cases}$$

et donc

$$\frac{\partial A_\theta}{\partial x} = \begin{cases} 1-y & \text{si } x(1-y) < \theta, \\ 0 & \text{si } x(1-y) > \theta, \end{cases} \quad \frac{\partial A_\theta}{\partial y} = \begin{cases} -x & \text{si } x(1-y) < \theta, \\ 0 & \text{si } x(1-y) > \theta. \end{cases}$$

Il s'ensuit que

$$\begin{aligned} 2 \int_0^1 \int_0^y \frac{\partial A_\theta}{\partial x} \frac{\partial A_\theta}{\partial y} dx dy &= -2 \int_0^1 \int_0^{y \wedge \{\theta/(1-y)\}} (1-y)x dx dy \\ &= - \int_0^1 (1-y) \left(y \wedge \frac{\theta}{1-y} \right)^2 dy \\ &= - \int_0^1 \left\{ y^2(1-y) \wedge \frac{\theta^2}{1-y} \right\} dy. \end{aligned}$$

Puisque les racines de l'équation $y(1-y) = \theta$ sont $1/2 \pm \sqrt{1-4\theta}/2$, il vient

$$\begin{aligned} \int_0^1 \left\{ y^2(1-y) \wedge \frac{\theta^2}{1-y} \right\} dy &\geq \int_{(1-\sqrt{1-4\theta})/2}^{(1+\sqrt{1-4\theta})/2} \frac{\theta^2}{1-y} dy \\ &= \theta^2 \log \left(\frac{1+\sqrt{1-4\theta}}{1-\sqrt{1-4\theta}} \right) \\ &= \theta^2 \log \left\{ \frac{(1+\sqrt{1-4\theta})^2}{4\theta} \right\} \geq \theta^2 \log \left(\frac{1}{4\theta} \right). \end{aligned}$$

En combinant ces différents éléments, on conclut que $2\rho_\theta - 3\tau_\theta \leq 12\theta^2 \log(4\theta)$, ce qui achève la démonstration de (B).

4. Autres propriétés de H_θ

Les copules H_θ ont d'autres propriétés intéressantes. On note par exemple que $H_0(x, y) = xy$ et $H_{1/4}(x, y) = x \wedge y$ pour tous $x, y \in [0, 1]$; ainsi, H_0 est la copule d'indépendance et $H_{1/4}$ la borne supérieure de Fréchet–Hoeffding.

Quand $\theta \in (0, 1/4)$, la loi H_θ comporte des parties singulière et absolument continue. Soient

$$\ell_\theta = \frac{1}{2} - \frac{1}{2}\sqrt{1-4\theta}, \quad u_\theta = \frac{1}{2} + \frac{1}{2}\sqrt{1-4\theta}, \quad L_x = \frac{\theta}{1-x}, \quad U_x = 1 - \frac{\theta}{x}.$$

Le support de la partie singulière est alors donné par

$$\begin{aligned} \mathcal{S}_\theta = & \{(x, x) : 0 \leq x \leq \ell_\theta\} \cup \{(x, x) : u_\theta \leq x \leq 1\} \\ & \cup \{(x, L_x) : \ell_\theta < x < u_\theta\} \cup \{(x, U_x) : \ell_\theta < x < u_\theta\}. \end{aligned}$$

Le graphe de cet ensemble est représenté à la figure 1.

De plus, la composante absolument continue de H_θ vaut 1 sur l'ensemble $\{(x, y) : \ell_\theta < x < u_\theta \text{ et } L_x < y < U_x\}$ et s'annule partout ailleurs. La masse associée à la composante absolument continue est

$$\sqrt{1-4\theta} - 2\theta \log \left(\frac{1 + \sqrt{1-4\theta}}{1 - \sqrt{1-4\theta}} \right).$$

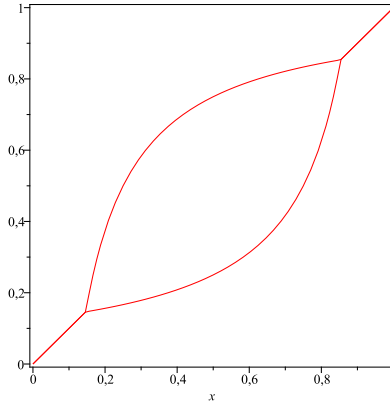


FIGURE 1: Ensemble sur lequel la partie singulière de H_θ est concentrée quand $\theta \in (0, 1/4)$. Set on which the singular part of C_θ is concentrated when $\theta \in (0, 1/4)$.

On note enfin que $H_{\theta_1}(x, y) \leq H_{\theta_2}(x, y)$ pour tous $x, y \in [0, 1]$, dès lors que $0 \leq \theta_1 \leq \theta_2 \leq 1/4$. La famille de lois (H_θ) est donc ordonnée en dépendance positive par quadrant, ce qui entraîne que l'application $\theta \mapsto \kappa(H_\theta)$ est croissante pour toute mesure de concordance κ au sens de la définition 5.1.7. du livre de Nelsen <1999>.

De fait, on vérifie sans peine que les applications

$$\theta \mapsto \rho_\theta = 1 - \sqrt{1 - 4\theta} + 10\theta\sqrt{1 - 4\theta} - 12\theta^2 \log \left(\frac{1 + \sqrt{1 - 4\theta}}{1 - \sqrt{1 - 4\theta}} \right),$$

et

$$\theta \mapsto \tau_\theta = 1 - \sqrt{1 - 4\theta} + 6\theta\sqrt{1 - 4\theta} - 4\theta^2 \log \left(\frac{1 + \sqrt{1 - 4\theta}}{1 - \sqrt{1 - 4\theta}} \right),$$

sont croissantes. Ces formules permettent de tracer le graphe de l'application $\theta \mapsto \rho_\theta^2 + 2\rho_\theta - 3\tau_\theta$; voir la figure 2. Le graphe de gauche y donne la fausse impression que la fonction est positive pour tout $\theta \in [0, 1/4]$; il faut restreindre le domaine à l'intervalle $[0, 1.5 \times 10^{-5}]$, comme sur le graphe de droite, pour que l'image soit concluante.

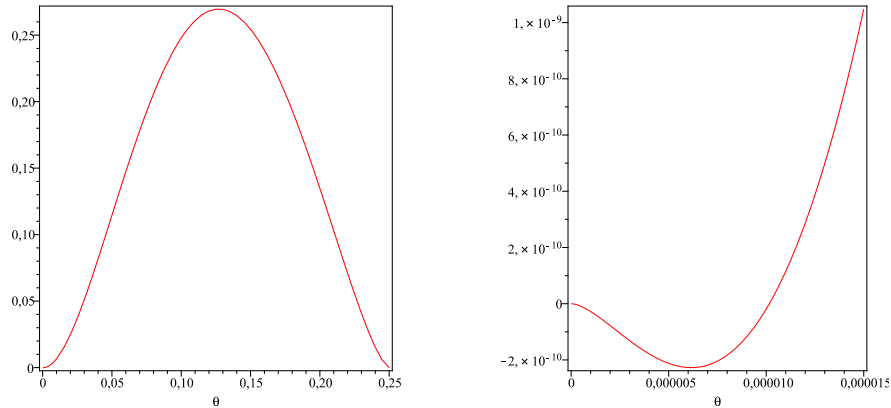


FIGURE 2: Graphe de $\rho_\theta^2 + 2\rho_\theta - 3\tau_\theta$ si $\theta \in [0, 1/4]$ (gauche) et $\theta \in [0, 1.5 \times 10^{-5}]$ (droite). Graph of $\rho_\theta^2 + 2\rho_\theta - 3\tau_\theta$ for $\theta \in [0, 1/4]$ (left) and $\theta \in [0, 1.5 \times 10^{-5}]$ (right).

La nature subtile du contre-exemple présenté ici laisse à penser que l'inégalité de gauche dans l'énoncé (1) est « presque vraie ». Il y a donc fort à parier qu'un argument encore plus délicat sera nécessaire pour établir ou

réfuter la dernière conjecture de Hutchinson & Lai <1990>, à savoir que $\rho \leq 2\tau - \tau^2$ pour les lois bivariées stochastiquement croissantes. Ceci pourra éventuellement faire l'objet de travaux ultérieurs.

Remerciements

Les auteurs remercient Mme Johanna Nešlehová pour ses commentaires et suggestions. Ces travaux ont été financés par le programme des Chaires de recherche du Canada, le Conseil de recherches en sciences naturelles et en génie du Canada, le Fonds québécois de la recherche sur la nature et les technologies, ainsi que l'Institut de finance mathématique de Montréal.

P. Capéraà, C. Genest, Spearman's ρ is larger than Kendall's τ for positively dependent random variables. *J. Nonparametr. Statist.* 2 (1993) 183–194.

H.E. Daniels, Rank correlation and population models. *J. Roy. Statist. Soc. Ser. B* 12 (1950) 171–181.

H.E. Daniels, Note on Durbin and Stuart's formula for $E(r_s)$. *J. Roy. Statist. Soc. Ser. B* 13 (1951) 310.

J. Durbin, A. Stuart, Inversions and rank correlation coefficients. *J. Roy. Statist. Soc. Ser. B* 13 (1951) 303–309.

C. Genest, J. Nešlehová, Analytical proofs of classical inequalities between Spearman's ρ and Kendall's τ . *J. Statist. Plann. Inference* 139 (2009) 3795–3798.

W. Hürlimann, Hutchinson–Lai's conjecture for bivariate extreme value copulas. *Statist. Probab. Lett.* 61 (2003) 191–198.

T.P. Hutchinson, C.D. Lai, *Continuous Bivariate Distributions, Emphasising Applications*. Rumsby Scientific, Adelaide, 1990.

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

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