JEAN-FRANÇOIS PLANTE

À PROPOS D'UNE MESURE DE CORRÉLATION DES RANGS DE BLEST

Mémoire
présenté
à la Faculté des études supérieures
pour l'obtention
du grade de maître ès sciences (M. Sc.)

Département de mathématiques et de statistique
FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL

JUILLET 2002

© Jean-François Plante, 2002

RÉSUMÉ

La copule sous-jacente à un couple de variables aléatoires continues permet de décrire complètement le lien de dépendance qui les unit. Pour résumer sur une échelle numérique l'intensité de la dépendance représentée par une copule, on emploie en général des coefficients de corrélation tels que ceux de Pearson, Spearman et Kendall. En 2000, une nouvelle mesure de corrélation basée sur les rangs a été proposée par David C. Blest. Ce mémoire présente les propriétés du coefficient de Blest et en propose quelques variantes. Des simulations permettent de comparer la valeur du coefficient de Blest et de ses versions symétriques et duales en tant que statistiques de tests permettant d'éprouver l'hypothèse d'indépendance.

Jean-François Plante, étudiant

Christian Genest, directeur

AVANT-PROPOS

Je voudrais d'abord exprimer ma gratitude envers mes parents, frères et sœurs pour m'avoir permis de grandir dans un climat familial sain et dynamique. C'est grâce à eux que j'ai appris à partager et à persévérer. J'ai été chanceux de pouvoir profiter tous les soirs d'une atmosphère paisible et reposante à la maison, après une dure journée à l'école.

Plus particulièrement, je désire remercier mon père de m'avoir appris l'honnêteté et l'intégrité. La confiance dont il a fait preuve à mon égard dès ma tendre enfance m'a permis de m'épanouir. À ma mère, je veux dire merci pour son dévouement, son amour et son attention qui m'ont soutenu constamment.

Je voudrais aussi dire merci à tous les amis qui m'ont accompagné au cours de mes études. Grâce à eux, j'ai eu énormément de plaisir à évoluer dans un milieu scolaire dynamique et enrichissant.

Au plan scientifique, j'ai eu beaucoup de plaisir à travailler avec Christian Genest. Son habileté à comprendre les fondements des problèmes et à les transmettre est fascinante. Son enseignement et ses conseils m'ont certainement aidé à apprendre comment comprendre !

Je suis également reconnaissant envers M. Kilani Ghoudi de m'avoir donné accès à certains de ses programmes informatiques et d'avoir bien voulu être membre du jury avec M. Louis-Paul Rivest.

Une partie de ce travail a été financée par des octrois individuels et collectifs accordés à M. Christian Genest et à l'équipe du *Fond québécois de recherche sur la nature et les technologies* qu'il coordonne. Que lui et ses collègues en soient remerciés, ainsi que le *Conseil de recherches en sciences naturelles et en génie du Canada*, qui m'a accordé une bourse d'étude de deuxième cycle.

TABLE DES MATIÈRES

RÉSUMÉ **ii**

AVANT-PROPOS **iii**

INTRODUCTION **1**

CHAPITRE I. ON BLEST'S MEASURE OF RANK CORRELATION **3**

1.1 Introduction 3

1.2 Limiting Distribution of Blest's Coefficients 7

1.3 Properties of Blest's Index 10

1.4 A Symmetrized Version of Blest's Measure of Association . . . 13

1.5 Performance as a Test of Independence 19

1.5.1 Pitman Efficiency 19

1.5.2 Power Comparisons in Finite Samples 25

1.6 Conclusion 29

1.7 Appendix 30

CHAPITRE II. MESURES DE CORRÉLATION BASÉES SUR LES RANGS **33**

2.1 Terminologie 33

2.2 Formes des coefficients 33

2.2.1 Élimination des marges 34

2.2.2 Paramètre estimé et loi asymptotique 34

2.2.3 Degré des polynômes considérés 35
2.2.4 Rôle des constantes 35
2.3 Mesures considérées 36
2.4 Résumé des propriétés 38
2.4.1 Dualité et symétrie radiale 38
2.4.2 Dualité et test d'indépendance 39
2.4.3 Liens linéaires 39
2.4.4 Propriétés individuelles 40

CHAPITRE III. CALCULS EXPLICITES 43

3.1 Variance des coefficients sous l'hypothèse d'indépendance . . . 43
3.1.1 Rho de Spearman 43
3.1.2 Coefficient de Blest 45
3.1.3 Coefficient de Plantagenet 47
3.2 Corrélation entre les coefficients sous l'hypothèse d'indépendance 50
3.2.1 Blest vs Spearman 50
3.2.2 Plantagenet vs Spearman 52
3.2.3 Blest vs Plantagenet 53
3.3 Valeur théorique de certains paramètres 54
3.3.1 Copule de Farlie–Gumbel–Morgenstern généralisée . . . 55
3.3.2 Copule de Cuadras–Augé 58
3.4 Variance asymptotique 59
3.5 Efficacité relative asymptotique 63
3.5.1 Illustration de la première méthode : copule de Cuadras–Augé 65
3.5.2 Illustration de la deuxième méthode : copule de Clayton 66

CHAPITRE IV. SIMULATIONS **69**

4.1 Loi normale bivariée 70

4.2 Copule de Frank 73

4.3 Copule de Clayton 75

4.4 Copule de valeurs extrêmes de Gumbel 78

4.5 Copule de valeurs extrêmes de Galambos 80

4.6 Transposition du minimum 83

4.6.1 Copule (6) 83

4.6.2 Copule (7) 86

CONCLUSION **89**

BIBLIOGRAPHIE **90**

INTRODUCTION

Pour quantifier la dépendance entre deux variables, la mesure la plus classique est sans doute le coefficient de corrélation de Pearson. Malheureusement, cette statistique dépend des lois marginales des observations, ce qui en limite souvent les valeurs possibles et en complique l'interprétation.

Une façon efficace de contourner ce problème consiste à utiliser des coefficients basés sur les rangs tels que le ρ de Spearman et le τ de Kendall. Les coefficients basés sur les rangs présentent l'avantage de ne pas être influencés par les lois marginales des observations.

En 2000, David C. Blest a proposé un coefficient de corrélation basé sur les rangs qui ressemble au ρ de Spearman, mais qui a la particularité de pondérer différemment les écarts observés entre les classements induits par les deux variables, selon qu'ils se produisent dans les premiers ou les derniers rangs.

Ce mémoire complète le travail de Blest en précisant la loi asymptotique de son coefficient et en en proposant des versions symétriques et duales.

Le chapitre I présente l'ensemble des résultats obtenus. Il s'agit d'un article écrit en anglais qui est actuellement en cours d'évaluation.

Le chapitre II apporte quelques explications supplémentaires qui viennent compléter le contenu de l'article. On y présente un point de vue plus intuitif des résultats obtenus et on y retrouve un résumé des propriétés de chaque coefficient.

Des exemples de calculs explicites, omis dans l'article, sont donnés au chapitre III. La plupart d'entre eux ont été réalisés à l'aide du logiciel de calcul symbolique MAPLE. Dans chaque cas, le code utilisé est présenté.

Pour évaluer la performance du coefficient de Blest et de ses variantes à titre de statistiques permettant de tester l'hypothèse d'indépendance, des simulations ont été effectuées. Les résultats de cette étude de Monte–Carlo sont consignés au chapitre IV.

Enfin, la conclusion du mémoire soulève plusieurs questions sur les différentes mesures de corrélation qu'il est possible de construire à partir de polynômes et sur l'avantage des unes par rapport aux autres.

CHAPITRE I

ON BLEST'S MEASURE OF RANK CORRELATION

Christian GENEST and Jean-François PLANTE

Blest (2000, *Austral. and New Zealand J. Statist.*) a proposé une mesure non paramétrique de corrélation entre des aléas X et Y. Son coefficient, asymétrique dans ses arguments, pondère différemment les écarts observés dans les classements induits par ces variables selon qu'ils se produisent dans les premiers ou les derniers rangs. Les auteurs déterminent la loi limite de l'indice de Blest et en suggèrent des variantes symétriques dont ils explorent les mérites au moyen de calculs d'efficacité relative asymptotique et d'une étude de Monte–Carlo.

1.1 Introduction

Although Pearson's correlation coefficient is one of the most ubiquitous concepts in the scientific literature, it is now widely recognized, at least among statisticians, that the degree of stochastic dependence between random variables X and Y with joint cumulative distribution function H and marginals F and G is much more appropriately characterized and measured in terms

of the joint distribution C of the pair $(F(X), G(Y))$, viz.

$$C(u,v) = \mathrm{P}\{F(X) \le u, G(Y) \le v\}, \quad 0 \le u, v \le 1$$

whose marginals are uniform on the interval $(0,1)$. Indeed, most modern concepts and measures of dependence, not to mention stochastic orderings (see for example Joe 1997, Nelsen 1999 or Drouet–Mari & Kotz 2001), are functions of the so-called "copula" C, which is uniquely determined on Range$(F) \times$ Range(G) and hence everywhere in the special case where F and G are continuous, as will be assumed henceforth.

In particular, classical nonparametric measures of dependence such as Spearman's rho

$$\begin{aligned} \rho(X,Y) &= 12 \int_{\mathbb{R}^2} F(x)G(y)\,\mathrm{d}H(x,y) - 3 \\ &= 12 \int_{[0,1]^2} uv\,\mathrm{d}C(u,v) - 3 = 12 \int_0^1 \int_0^1 C(u,v)\,\mathrm{d}v\,\mathrm{d}u - 3 \quad (1) \end{aligned}$$

and Kendall's tau

$$\tau(X,Y) = 4 \int_{\mathbb{R}^2} H(x,y)\,\mathrm{d}H(x,y) - 1 = 4 \int_{[0,1]^2} C(u,v)\,\mathrm{d}C(u,v) - 1$$

are superior to Pearson's coefficient in that while they vanish when the variables are independent, they always exist and take their extreme values ± 1 when X and Y are in perfect (positive or negative) functional dependence, i.e., when either $Y = G^{-1}\{F(X)\}$ or $Y = G^{-1}\{1 - F(X)\}$ with probability one. Except in very special circumstances, these cases of complete dependence are not instances of *linear* dependence. Thus when considering random data $(X_1, Y_1), \ldots, (X_n, Y_n)$ from an unknown distribution H whose support

is all of $[0, \infty)^2$, for instance, the values of ρ and τ are unconstrained, whereas Pearson's correlation can only span an interval $[r, 1]$ whose lower bound $r > -1$ depends on the choice of marginals F and G. When the latter are unknown, it is thus difficult to know what to make of an observed Pearson correlation of -0.2, say.

Because copulas are margin-free and ranks are maximally invariant statistics of the observations under monotone transformations of the marginal distributions, ρ, τ and indeed all other copula-based measures of dependence (e.g., the index of Schweizer & Wolf 1981) should be estimated by functions of the ranks R_i and S_j of the X_i's and Y_j's. Letting F_n and G_n stand for the empirical distribution functions of X and Y respectively, the classical estimate for ρ is

$$\rho_n = \mathrm{corr}\{F_n(X), G_n(Y)\} = \frac{12}{n^3 - n} \sum_{i=1}^{n} R_i S_i - 3\,\frac{n+1}{n-1}$$

since $F_n(X_i) = R_i/n$ and $G_n(Y_j) = S_j/n$ for all $1 \le i, j \le n$. Likewise, τ is traditionally estimated by the scaled difference in the numbers of concordant and discordant pairs, or equivalently by

$$\tau_n = \frac{2}{n^2 - n} \sum_{1 \le i < j \le n} \mathrm{sign}(R_i - R_j)\ \mathrm{sign}(S_i - S_j).$$

In comparing ρ_n and τ_n in terms of their implicit weighing of differences $R_i - S_i$, Blest (2000) was led to propose an alternative measure of rank correlation that "... attaches more significance to the early ranking of an initially given order." Assume, for example, that X_i and Y_i represent the running times of sprinter $i = 1, \ldots, n$ in two successive track-and-field meetings. The

correlation in the pairs (R_i, S_i) then gives an idea of the consistency between the two rankings. However, differences in the top ranks would seem to be more critical, in that they matter in awarding medals. As a result, Blest suggests that these discrepancies should be emphasized, whereas all rank reversals are given the same weight in Spearman's or Kendall's coefficient.

To be specific, Blest's index is defined by

$$\nu_n = \frac{2n+1}{n-1} - \frac{12}{n^2-n} \sum_{i=1}^{n} \left(1 - \frac{R_i}{n+1}\right)^2 S_i.$$

The constants are rigged so that the coefficient varies between 1 and -1 and is most extreme when the rankings coincide ($S_i = R_i$) or are antithetic ($S_i = n + 1 - R_i$). Numerical results and calculations reported by Blest (2000) also indicate that his measure can discriminate more easily between individual permutations than either ρ_n or τ_n while being highly correlated with both of them. Furthermore, partial evidence is provided which indicates that the large-sample distribution of ν_n is normal.

The first objective of this paper is to show that ν_n is an asymptotically unbiased estimator of the population parameter

$$\begin{aligned} \nu(X,Y) &= 2 - 12 \int_{\mathbb{R}^2} \{1 - F(x)\}^2 G(y)\, \mathrm{d}H(x,y) \qquad (2) \\ &= 2 - 12 \int_{[0,1]^2} (1-u)^2 v\, \mathrm{d}C(u,v) \end{aligned}$$

and that $\sqrt{n}(\nu_n - \nu)$ converges in distribution to a centered normal random variable whose variance is specified in Section 1.2. While ν_n may be appropriate as an index of discrepancy between two rankings, it is pointed out

in Section 1.3 that since in general $\nu(X,Y) \neq \nu(Y,X)$, the parameter (2) estimated by ν_n is not a measure of concordance, in the sense of Scarsini (1984). The properties of a symmetrized version of ν_n are thus considered in Section 1.4, and the relative merits of this new statistic (and a natural complement thereof) for testing independence are then examined in Section 1.5 through asymptotic relative efficiency calculations and a small Monte Carlo simulation study. Brief concluding comments are given in Section 1.6 and the Appendix (Section 1.7) contains some technical details.

1.2 Limiting Distribution of Blest's Coefficients

Define $J(t) = (1-t)^2$ and $K(t) = t$ for all $0 \leq t \leq 1$, and for arbitrary integers n and $i \in \{1, \ldots, n\}$, let $J_n(t) = J\{i/(n+1)\}$ and $K_n(t) = K\{i/(n+1)\}$ if $(i-1)/n < t \leq i/n$. The large-sample behaviour of Blest's sample measure of association is clearly the same as that of

$$2 - \frac{12}{n} \sum_{i=1}^{n} \left(1 - \frac{R_i}{n+1}\right)^2 \frac{S_i}{n+1} \ ,$$

which may be written as

$$2 \ - 12 \int J_n(F_n) K_n(G_n) dH_n$$

in terms of the empirical cumulative distribution function H_n of H and its marginals F_n and G_n. Since the functions J and K and their derivatives are bounded on $[0,1]$, and in view of Remark 2.1 of Ruymgaart, Shorack & van Zwet (1972), a direct application of Theorem 2.1 of these authors implies that

$\sqrt{n}(\nu_n - \nu)$ is asymptotically normally distributed with mean and variance as specified below.

PROPOSITION 1. *Under the assumption of random sampling from a continuous bivariate distribution H with underlying copula C, $\sqrt{n}(\nu_n - \nu)$ converges weakly to a normal random variable with zero mean and the same variance as*

$$12\left[(1-U)^2 V - 2\int_U^1 (1-u)\mathrm{E}(V|U=u)\,\mathrm{d}u + \int_V^1 \mathrm{E}\{(1-U)^2|V=v\}\,\mathrm{d}v\right], \tag{3}$$

where the pair (U, V) is distributed as C. In particular, the variance of the latter expression equals $16/15$ when U and V are independent.

As is the case for Spearman's rho, compact algebraic formulas for ν can be found for relatively few models. Examples 1 and 2 illustrate, in two special cases of interest, the explicit calculations that can sometimes be made using a symbolic calculator such as MAPLE; see Chapter III for details. Example 3, which concerns the pervasive normal model, is somewhat more subtle.

EXAMPLE 1. Suppose that (X, Y) follows a Farlie–Gumbel–Morgenstern distribution with marginals F and G, viz.

$$H_\theta(x, y) = F(x)G(y) + \theta F(x)G(y)\{1 - F(x)\}\{1 - G(y)\}, \quad x, y \in I\!R$$

with parameter $\theta \in [-1, 1]$. Then

$$\nu_\theta(X, Y) = \rho_\theta(X, Y) = \frac{3}{2}\,\tau_\theta(X, Y) = \frac{\theta}{3}$$

and the variance of (3) equals

$$\frac{16}{15} - \frac{16}{63}\,\theta^2.$$

Example 2. For arbitrary $a, b \in [0,1]$ and $\theta \in [-1,1]$, let

$$\begin{aligned} H_{\theta,a,b}(x,y) &= F(x)G(y)\left[1+\theta\left\{1-F(x)^a\right\}\left\{1-G(y)^b\right\}\right] \\ &= \Pi\left\{F(x)^{1-a}, G(y)^{1-b}\right\} C_\theta\left\{F(x)^a, G(y)^b\right\}, \quad x, y \in I\!R \end{aligned}$$

where $C_\theta(u,v) = H_\theta\left\{F^{-1}(u), G^{-1}(v)\right\}$ is the Farlie–Gumbel–Morgenstern copula and $\Pi(u,v) = uv$ denotes the independence copula. By Khoudraji's device (see Genest, Ghoudi & Rivest 1998), $H_{\theta,a,b}$ is an asymmetric extension of the Farlie–Gumbel–Morgenstern distribution, recently considered in a somewhat more general form by Bairamov, Kotz & Bekçi (2000). For any pair (X,Y) distributed as $H_{\theta,a,b}$, one finds

$$\nu_{\theta,a,b}(X,Y) = \frac{2ab(a+5)\theta}{(a+2)(a+3)(b+2)}\,,$$

while

$$\rho_{\theta,a,b}(X,Y) = \frac{3ab\theta}{(a+2)(b+2)} = \frac{3}{2}\,\tau_{\theta,a,b}(X,Y).$$

Note that

$$\nu_{\theta,a,b}(X,Y) \neq \nu_{\theta,a,b}(Y,X) = \nu_{\theta,b,a}(X,Y),$$

unless of course $a = b$, in which case the copula is symmetric in its arguments. An explicit but long formula (not shown here) for the asymptotic variance of $\sqrt{n}(\nu_n - \nu)$ is also available in this case.

Example 3. Suppose that (X,Y) has a bivariate normal distribution and that $\mathrm{corr}(X,Y) = r \in [-1,1]$. Then

$$\nu_r(X,Y) = \rho_r(X,Y) = \frac{6}{\pi}\,\arcsin\left(\frac{r}{2}\right)$$

while

$$\tau_r(X,Y) = \frac{2}{\pi}\arcsin(r).$$

The fact that $\nu(X,Y) = \rho(X,Y)$ arises whenever a copula C is radially symmetric, i.e., when its associated survival function $\bar{C}(u,v) = 1 - u - v + C(u,v)$ satisfies the condition

$$\bar{C}(u,v) = C(1-u, 1-v), \quad 0 \le u, v \le 1. \tag{4}$$

Indeed, since the measure C is then invariant by the change of variable $(x,y) = (1-u, 1-v)$, one must have

$$\int_{[0,1]^2} (1-u)^2 v \,\mathrm{d}C = \int_{[0,1]^2} u^2(1-v)\,\mathrm{d}C.$$

Expanding both sides and using the fact that C has uniform marginals on $[0,1]$ yields

$$\int_{[0,1]^2} u^2 v \,\mathrm{d}C = \int_{[0,1]^2} uv\,\mathrm{d}C - \frac{1}{12}\ ,$$

from which it follows at once that $\nu(X,Y) = \rho(X,Y)$ (and hence equals $\nu(Y,X)$ as well). The formulas for $\rho_r(X,Y)$ and $\tau_r(X,Y)$ are standard normal theory ; see for example Exercise 2.14, p. 54 of Joe (1997).

Unfortunately, the variance of (3) can only be computed numerically in Example 3. Additional properties of $\nu(X,Y)$ are described next.

1.3 Properties of Blest's Index

According to Scarsini (1984), the following are fundamental properties that any measure κ of concordance should satisfy :

a) κ is defined for every pair (X, Y) of continuous random variables.

b) $-1 \leq \kappa(X, Y) \leq 1$, $\kappa(X, X) = 1$, and $\kappa(X, -X) = -1$.

c) $\kappa(X, Y) = \kappa(Y, X)$.

d) If X and Y are independent, then $\kappa(X, Y) = 0$.

e) $\kappa(-X, Y) = \kappa(X, -Y) = -\kappa(X, Y)$.

f) If $(X, Y) \prec (X^*, Y^*)$ in the positive quadrant dependence ordering, then $\kappa(X, Y) \leq \kappa(X^*, Y^*)$.

g) If $(X_1, Y_1), (X_2, Y_2), \ldots$ is a sequence of continuous random vectors that converges weakly to a pair (X, Y), then $\kappa(X_n, Y_n) \to \kappa(X, Y)$ as $n \to \infty$.

It is well known that both ρ and τ meet all these conditions, and it is easy to check that the index ν defined in (2) has properties a), b), d), f), and g). To show the latter two, it is actually more convenient to use the alternative representation

$$\nu(X, Y) = -2 + 24 \int_{[0,1]^2} (1 - u) C(u, v) \, \mathrm{d}u \, \mathrm{d}v,$$

which follows immediately from an extension of Hoeffding's identity (1) due to Quesada–Molina (1992).

As was already pointed out in Example 2, however, Blest's measure does not satisfy condition c). Furthermore, property e) is not verified either. For, if (X, Y) has copula C, then $C^*(u, v) = v - C(1-u, v)$ is the copula associated with the pair $(-X, Y)$, and hence

$$\nu(-X, Y) = \nu(X, Y) - 2\rho(X, Y), \tag{5}$$

so that $\nu(-X,Y) \neq -\nu(X,Y)$ except in the special case where $\nu(X,Y) = \rho(X,Y)$. Note, however, that $\nu(X,-Y) = -\nu(X,Y)$ by a similar argument involving the copula $C^{**}(u,v) = u - C(u,1-v)$ of the pair $(X,-Y)$.

The following simple example provides a concrete illustration of the failure of condition e) for Blest's index.

EXAMPLE 4. Suppose that X is uniform on the interval $(0,1)$ and that either

$$Y = \begin{cases} \theta - X & \text{if } 0 < X < \theta, \\ X & \text{if } \theta \leq X < 1. \end{cases} \tag{6}$$

or

$$Y = \begin{cases} X & \text{if } 0 < X \leq 1-\theta, \\ 2-\theta-X & \text{if } 1-\theta < X < 1. \end{cases} \tag{7}$$

for some constant $0 \leq \theta \leq 1$. The joint distribution of the pair (X,Y) is then a "shuffle of min" copula in the sense of Mikusiński, Sherwood & Taylor (1992), and the probability mass of this singular distribution is spread uniformly on two line segments depending on θ, as illustrated in Figure 1.

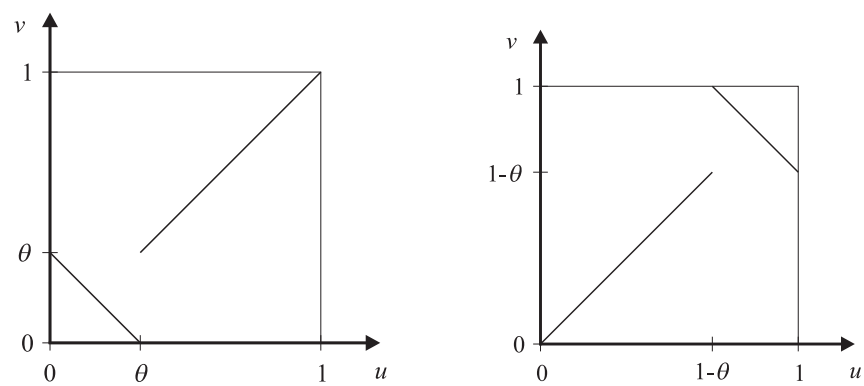


FIGURE 1. Support of the "shuffle of min" copulas (6) and (7) discussed in Example 4.

When (X, Y) is distributed as (6), one finds

$$\nu_\theta(X, Y) = 1 - 4\theta^3 + 2\theta^4,$$

while if (X, Y) is distributed as (7), one gets

$$\nu_\theta(X, Y) = 1 - 2\theta^4.$$

Had Blest's coefficient met condition e) above, the two expressions would have been equal, since the two supports displayed in Figure 1 are equivalent up to a rotation of 180 degrees about the point $(1/2, 1/2)$, i.e., a reflection through the line $u = 1/2$ followed by another reflection through the line $v = 1/2$.

1.4 A Symmetrized Version of Blest's Measure of Association

While the coefficient ν_n and its population equivalent $\nu(X, Y)$ may be appropriate as an index of discrepancy between a fixed order, say given by the X_i's, and an alternate order given by the other variable, the above discussion makes it clear that neither can be used to measure concordance.

There may be interest, however, in symmetrized versions of Blest's proposal, viz.

$$\xi_n = -\frac{4n+5}{n-1} + \frac{6}{n^3 - n} \sum_{i=1}^{n} R_i S_i \left(4 - \frac{R_i + S_i}{n+1}\right) \tag{8}$$

and its theoretical counterpart

$$\begin{aligned}\xi(X,Y) = \frac{\nu(X,Y)+\nu(Y,X)}{2} &= -4 + 6\int_{[0,1]^2} uv(4-u-v)\,\mathrm{d}C(u,v) \\ &= -2 + 12\int_{[0,1]^2} (2-u-v)C(u,v)\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

Because of an obvious connection between English history and the concatenation of their names, the authors jokingly refer to these quantities as the (empirical and theoretical) "Plantagenet" coefficients. Nevertheless, they believe that Blest should be credited for ξ_n and ξ.

As explained in Section 4 of Blest (2000), $\mathrm{E}(\nu_n) = 0$ under the null hypothesis of independence, and hence $\mathrm{E}(\xi_n) = 0$ as well. Under H_0, it is also a simple matter to prove (see Appendix for details) that

$$\mathrm{var}(\xi_n) = \frac{31n^2 + 60n + 26}{30(n+1)^2(n-1)}$$

and that $\mathrm{cov}(\rho_n, \xi_n) = \mathrm{var}(\rho_n) = 1/(n-1)$, so that

$$\mathrm{corr}(\rho_n, \xi_n) = \sqrt{\frac{30(n+1)^2}{31n^2+60n+26}} \to \sqrt{\frac{30}{31}} \approx 0.9837,$$

while $\mathrm{corr}(\rho_n, \nu_n) \to \sqrt{15/16} \approx 0.9682$.

Although it is not generally true that ξ_n is an unbiased estimator of ξ under other distributional hypotheses between X and Y, it is asymptotically unbiased, as implied by the following result.

Proposition 2. *Under the assumption of random sampling from a continuous bivariate distribution H with underlying copula C, $\sqrt{n}(\xi_n - \xi)$ converges*

weakly to a normal random variable with zero mean and the same variance as

$$6\left[UV(4-U-V)+\int_U^1 \left\{2(2-u)\mathrm{E}(V|U=u)-\mathrm{E}\left(V^2|U=u\right)\right\}\,\mathrm{d}u \right.$$
$$\left. +\int_V^1 \left\{2(2-v)\mathrm{E}(U|V=v)-\mathrm{E}\left(U^2|V=v\right)\right\}\,\mathrm{d}v\right],$$

where the pair (U,V) *is distributed as* C*. In particular, the variance of the latter expression equals* $31/30$ *when* U *and* V *are independent.*

Proof. The asymptotic behaviour of ξ_n is obviously the same as that of

$$-4+\frac{6}{n}\sum_{i=1}^{n}\frac{R_i}{n+1}\,\frac{S_i}{n+1}\left(4-\frac{R_i}{n+1}-\frac{S_i}{n+1}\right),$$

which may be written alternatively as

$$-4+6\int_{[0,1]^2} uv(4-u-v)\,\mathrm{d}C_n(u,v)$$

in terms of the rescaled empirical copula function, as defined by Genest, Ghoudi & Rivest (1995). The conclusion is then an immediate consequence of their Proposition A·1, upon choosing $J(u,v)=uv(4-u-v)$, $\delta=1/4$ and $M=p=q=2$, say, to satisfy conditions (i) and (ii) of their result. □

Proceeding as above but with $J(u,v)=auv+buv(4-u-v)$ with arbitrary reals a and b actually shows that any linear combination of $\sqrt{n}(\rho_n-\rho)$ and $\sqrt{n}(\xi_n-\xi)$ is normally distributed with zero mean and the same variance as

$$6\Bigg[UV\{2a+b(4-U-V)\}$$
$$+\int_U^1 \left[2\{a+b(2-u)\}\mathrm{E}(V|U=u)-b\mathrm{E}\left(V^2|U=u\right)\right]\,\mathrm{d}u$$
$$+\int_V^1 \left[2\{a+b(2-v)\}\mathrm{E}(U|V=v)-b\mathrm{E}\left(U^2|V=v\right)\right]\,\mathrm{d}v\Bigg]. \qquad (9)$$

Consequently, the joint distribution of $\sqrt{n}(\rho_n-\rho)$ and $\sqrt{n}(\xi_n-\xi)$ must be asymptotically normal with zero mean and a covariance matrix whose diagonal entries correspond to the choices $(a,b)=(1,0)$ and $(0,1)$. The limiting covariance between these two quantities can also be derived from these large-sample variances and that of the linear combination corresponding to $a=2$ and $b=-1$, for instance. These observations are formally gathered below.

Proposition 3. *Under the assumption of random sampling from a continuous bivariate distribution H with underlying copula C, $\sqrt{n}(\xi_n-\xi,\rho_n-\rho)'$ converges weakly to a normal random vector with zero mean and covariance matrix*

$$\begin{pmatrix} \sigma_\xi^2 & \kappa \\ \kappa & \sigma_\rho^2 \end{pmatrix} \quad \textit{with} \quad \kappa=\frac{4\sigma_\rho^2+\sigma_\xi^2-\sigma_{2\rho-\xi}^2}{4}\ ,$$

where σ_ξ^2, σ_ρ^2 and $\sigma_{2\rho-\xi}^2$ are the variances of (9) *corresponding to the choices $(a,b)=(1,0)$, $(0,1)$, and $(2,-1)$, and where the pair (U,V) is distributed as C. In particular, $\sigma_\xi^2=31/30$ and $\sigma_\rho^2=\kappa=1$ under independence.*

Remark 1. When condition (4) holds, making the change of variables $(x,y)=(1-u,1-v)$ in Equation (9) shows that one must then have $\sigma_\xi^2=\sigma_{2\rho-\xi}^2$ and

hence $\kappa = \sigma_\rho^2$ in Proposition 3. Models for which $C = \bar{C}$ include the Farlie–Gumbel–Morgenstern, the Gaussian, the Plackett (1965), and Frank's copula (Nelsen 1986; Genest 1987); note that the latter is the only Archimedean copula that is radially symmetric (see for example Nelsen 1999, p. 97).

Remark 2. Since the joint distribution of $\sqrt{n}(\rho_n - \rho)$ and $\sqrt{n}(\tau_n - \tau)$ is also known to be normal with limiting correlation equal to 1, Proposition 3 implies that

$$\lim_{n\to\infty} \operatorname{corr}(\xi_n, \tau_n) = \lim_{n\to\infty} \operatorname{corr}(\xi_n, \rho_n) = \sqrt{\frac{30}{31}}$$

under the null hypothesis of independence.

The following examples provide numerical illustrations of these various facts.

Example 1 (continued). If (X, Y) follows a Farlie–Gumbel–Morgenstern distribution with marginals F and G and parameter $\theta \in [-1, 1]$, then

$$\xi_\theta(X, Y) = \rho_\theta(X, Y) = \frac{3}{2}\,\tau_\theta(X, Y) = \frac{\theta}{3}$$

and

$$\sqrt{n}(\xi_n - \xi_\theta) \longrightarrow N\left(0, \frac{1}{450}\,\theta^3 - \frac{157}{630}\,\theta^2 + \frac{2}{225}\,\theta + \frac{31}{30}\right).$$

Furthermore, $\operatorname{cov}(\sqrt{n}\xi_n, \sqrt{n}\rho_n) \to 1 - 11\,\theta^2/45$ and

$$\operatorname{corr}(\xi_n, \rho_n) \longrightarrow \sqrt{\frac{3150 - 770\,\theta^2}{3255 + 28\,\theta - 785\,\theta^2 + 7\,\theta^3}}$$

as $n \to \infty$. The latter is a decreasing function of θ taking values in the interval $[0.9747, 0.9887]$.

Example 2 (continued). For any pair (X, Y) distributed as $H_{\theta,a,b}$, one finds

$$\xi_{\theta,a,b}(X, Y) = \frac{2ab(ab + 4a + 4b + 15)\theta}{(a+2)(a+3)(b+2)(b+3)}\ ,$$

which is (happily!) symmetric in a and b. An algebraic expression for the asymptotic variance of $\sqrt{n}(\xi_n - \xi)$ exists but is rather unwieldy.

Example 3 (continued). When the population is bivariate normal with (Pearson) correlation r, it was seen earlier that $\nu_r(X,Y) = \rho_r(X,Y)$, and hence $\xi_r(X,Y) = \rho_r(X,Y)$. Note, however, that ξ_n is not necessarily equal to ρ_n. In this case, numerical integration must be used to compute the asymptotic variance of $\sqrt{n}(\xi_n - \xi)$ or the correlation between that statistic and ρ_n.

Example 4 (continued). Whether the pair (X,Y) is distributed as shuffle of min (6) or (7), one has $\xi_\theta(X,Y) = \nu_\theta(X,Y)$, since both of these copulas are symmetric in their arguments. Accordingly, the symmetrized version ξ of Blest's index satisfies all the conditions listed by Scarsini (1984), except e).

If (X,Y) is distributed as (6),

$$\sqrt{n}(\xi_n - \xi_\theta) \to N\left[0, 16\,\theta^5(1-\theta)(3-2\theta)^2\right]$$

while if (X,Y) is distributed as (7), then

$$\sqrt{n}(\xi_n - \xi_\theta) \to N\left[0, 64\,\theta^7(1-\theta)\right].$$

In both cases, $\rho_\theta(X,Y) = 1 - 2\theta^3$ (because Spearman's rho satisfies condition e) and $\mathrm{corr}(\xi_n, \rho_n) \to 1$ as $n \to \infty$, for all values of $0 \le \theta \le 1$.

Remark 3. Should a nonparametric measure of dependence κ attach more significance to early ranks, as suggested by Blest (2000), it is clear from Example 4 that such a construct could not possibly satisfy condition e).

Indeed, if (U, V) and (U^*, V^*) followed distributions (6) and (7) with the same parameter, say $\theta < 1/2$, Blest's concept would imply that $\kappa(U, V) < \kappa(U^*, V^*)$, since it is plain from Figure 1 that reversals occur in small ranks under (6) while they occur in large ranks under (7). In view of property e), it would thus follow that

$$\kappa(U, V) < \kappa(U^*, V^*) = \kappa(-U^*, -V^*),$$

but this is impossible since $\kappa(-U^*, -V^*) = \kappa(1 - U^*, 1 - V^*) = \kappa(U, V)$ because these three pairs of variables have the same underlying copula.

1.5 Performance as a Test of Independence

In order to assess the potential of the symmetrized version of Blest's coefficient for testing against independence in bivariate models, it may be of interest to compare its performance with that of Spearman's rho and Kendall's tau, which are the two most common rank statistics used to this end. Asymptotic and finite-sample comparisons are presented in turn.

1.5.1 Pitman Efficiency

Using Proposition 3, it is a simple matter to compute Pitman's asymptotic relative efficiency (ARE) of tests T_ξ and T_ρ based on ξ_n and ρ_n, respectively. Given a family (C_θ) of copulas with $\theta = \theta_0$ corresponding to independence, standard theory (see for example Lehmann 1998, p. 371 ff.) implies that

$$\mathrm{ARE}\,(T_\xi, T_\rho) = \frac{30}{31} \left(\frac{\xi_0'}{\rho_0'}\right)^2,$$

where $\xi_0' \equiv \mathrm{d}\xi_\theta/\mathrm{d}\theta$ evaluated at $\theta = \theta_0$ and ρ_0' is defined *mutatis mutandis*. The factor 30/31 comes about because as pointed out in Proposition 3, the asymptotic variances of $\sqrt{n}\xi_n$ and $\sqrt{n}\rho_n$ are 31/30 and 1, respectively.

As is obvious from Remark 1,

$$\mathrm{ARE}\,(T_\xi, T_\rho) = \frac{30}{31} \approx 96.77\%$$

when the copula model under which Pitman's efficiency is computed is radially symmetric. There is thus no reason to base a test of independence on ξ_n (whose variance is larger than that of ρ_n) if the alternative satisfies condition (4), as is the case for the normal distribution and the Farlie–Gumbel–Morgenstern, Plackett and Frank copulas, for instance. Additional examples of explicit calculations are given below.

Example 5. Suppose that the copula of a pair (X, Y) is of the form

$$C_\theta(u, v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, \quad \theta > 0$$

with $C_0(u, v) = \lim_{\theta\to 0} C_\theta(u, v) = uv$ for all $0 \le u, v \le 1$. This Archimedean copula model, generally attributed to Clayton (1978), is quite popular in survival analysis, where it provides a natural bivariate extension of Cox's proportional hazards model; see for example Oakes (2001, §7.3). One can check easily that

$$\dot{C}_0(u, v) \equiv \lim_{\theta\to 0} \frac{\partial C_\theta(u, v)}{\partial\theta} = uv \log(u) \log(v),$$

so that

$$\rho_0' = 12 \int_0^1 \int_0^1 \dot{C}_0(u, v)\, \mathrm{d}u\, \mathrm{d}v = \frac{3}{4}\ ,$$

while

$$\xi_0' = 12 \int_0^1 \int_0^1 (2 - u - v)\dot{C}_0(u, v)\, \mathrm{d}u\, \mathrm{d}v = \frac{5}{6} \ ,$$

yielding $\mathrm{ARE}(T_\xi, T_\rho) = 1000/837 \approx 119.47\%$. Consequently, an improvement of some 20% can be achieved by using ξ_n instead of ρ_n for testing against independence. The same would be true of τ_n, since $\mathrm{ARE}(T_\tau, T_\rho) = 1$ in this case (and all subsequent ones considered here).

Example 6. Suppose that the pair (X, Y) is distributed as Gumbel's bivariate exponential distribution (Gumbel 1960a), whose copula (Nelsen 1999, p. 94) is

$$C_\theta(u, v) = uv \exp\{-\theta \log(u) \log(v)\}, \quad 0 < \theta < 1$$

with $C_0(u, v) = \lim_{\theta \to 0} C_\theta(u, v) = uv$. In this case, $\dot{C}_0(u, v) = -uv \log(u) \log(v)$, so that $\rho_0' = -3/4$ and $\xi_0' = -5/6$, and $\mathrm{ARE}(T_\xi, T_\rho) = 1000/837$, as above.

Example 7. Suppose that the pair (X, Y) follows a bivariate logistic distribution as defined by Ali, Mikhail & Haq (1978). The corresponding copula is then (Nelsen 1999, p. 25)

$$C_\theta(u, v) = \frac{uv}{1 + \theta(1 - u)(1 - v)} \ , \quad -1 < \theta < 1.$$

In this case, $\rho_0' = \xi_0' = 1/3$, whence $\mathrm{ARE}(T_\xi, T_\rho) = 30/31$, as is the case for radially symmetric copulas (which this one isn't).

Example 8. Suppose that the distribution of (X, Y) is a member of the bivariate exponential family of Marshall & Olkin (1967), considered as a prime example of "common shock" model in reliability theory. Its associated

copula, known as the generalized Cuadras–Augé copula (Nelsen 1999, §3.1.1), is

$$C_{a,b}(u,v) = \min\left(u^{1-a}v, uv^{1-b}\right)$$

with $0 \le a, b \le 1$. For this family, independence occurs whenever $\min(a,b) = 0$ and

$$\rho_{a,b} = \frac{3ab}{2a + 2b - ab}$$

with the convention that $\rho_{a,b} = 0$ when $a = b = 0$, so that

$$\left.\frac{\partial \rho'_{a,b}}{\partial b}\right|_{b=0} = \left.\frac{\partial \rho'_{a,b}}{\partial a}\right|_{a=0} = \frac{3}{2}$$

when the other parameter is fixed. A closed-form expression (not reproduced here) is also available for $\xi_{a,b}$, and symbolic calculation yields

$$\left.\frac{\partial \xi'_{a,b}}{\partial b}\right|_{b=0} = \left.\frac{\partial \xi'_{a,b}}{\partial a}\right|_{a=0} = \frac{4}{3}\ ,$$

whence $\mathrm{ARE}(T_\xi, T_\rho) = 640/837 \approx 76.46\%$.

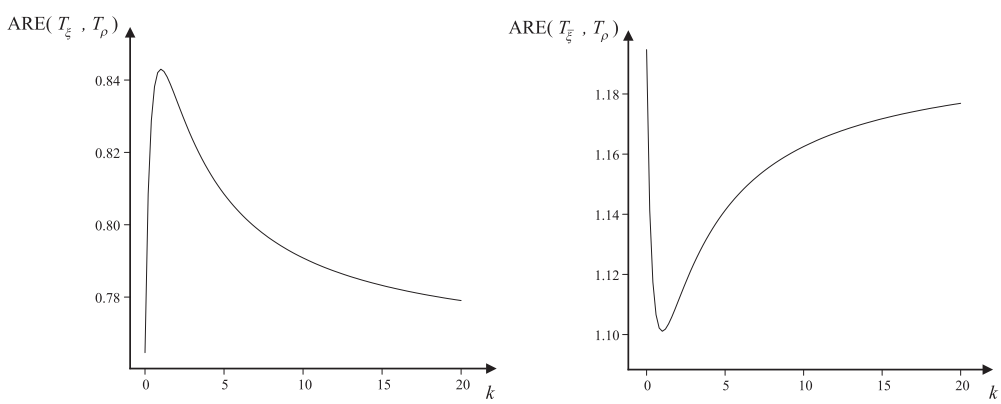


FIGURE 2. Pitman's asymptotic relative efficiency for the Marshall–Olkin copula of Example 8 when $b = ka$, plotted as a function of k; $\mathrm{ARE}(T_\xi, T_\rho)$ is depicted in the left panel, while $\mathrm{ARE}(T_{\bar{\xi}}, T_\rho)$ is displayed in the right panel.

As an additional example, consider the one-parameter family obtained by setting $b = ka$ for some fixed $0 < k < 1/a$. Then $\rho_0' = 3k/(2k+2)$ and

$$\xi_0' = \frac{k(8 + 19k + 8k^2)}{(2+3k)(3+2k)(1+k)} \ ,$$

so that Pitman's ARE is a function of the constant k that is plotted in the left panel of Figure 2. The ARE is seen to reach its minimum value of 640/837 when $k \to 0$ or ∞. Its maximum, viz., $392/465 \approx 84.30\%$, occurs when $k = 1$, which corresponds to the standard copula of Cuadras & Augé (1981).

Examples 5 to 8 show that a test of independence based on the symmetrized version of Blest's coefficient is sometimes preferable, but not always, to Spearman's test. Interestingly, however, it is generally possible to outperform the test based on ρ_n (or τ_n, since the latter two are usually equivalent asymptotically) by using either the test involving ξ_n or a similar procedure founded on the complementary statistic

$$\begin{aligned} \bar{\xi}_n &= -\frac{4n+5}{n-1} + \frac{6}{n^3-n} \sum_{i=1}^{n} \bar{R}_i \bar{S}_i \left(4 - \frac{\bar{R}_i + \bar{S}_i}{n+1}\right) \\ &= -\frac{2n+1}{n-1} + \frac{6}{n^3-n} \sum_{i=1}^{n} R_i S_i \left(\frac{R_i + S_i}{n+1}\right) \end{aligned}$$

defined as in (8), but with $\bar{R}_i = (n+1) - R_i$ and $\bar{S}_i = (n+1) - S_i$ instead of R_i and S_i, respectively.

Since using the reverse ranks amounts to working with the transformed data $(-X_1, -Y_1), \ldots, (-X_n, -Y_n)$, and in view of the discussion surrounding equa-

tion (5), it is plain that $\bar{\xi}_n$ is an asymptotically unbiased estimator of

$$\bar{\xi}(X,Y) = \xi(-X,-Y) = 2\rho(X,Y) - \xi(X,Y). \tag{10}$$

Calling on Proposition 2, one can also check readily that the limiting distribution of $\bar{\xi}_n$ is actually the same as that which ξ_n would have if the underlying dependence function were what Nelsen (1999, p. 28) calls the survival copula, i.e., $\hat{C}(u,v) = u+v-1+C(1-u,1-v)$. Furthermore, $\mathrm{var}(\bar{\xi}_n) = \mathrm{var}(\xi_n)$ for any sample size $n \geq 1$ under the null hypothesis of independence.

The right panel of Figure 2 shows $\mathrm{ARE}(T_{\bar{\xi}}, T_\rho)$ as a function of k for the generalized Cuadras–Augé copula of Example 8 with parameter $b = ka$. The curve reaches its minimum value of $512/465 \approx 110.11\%$ at $k = 1$. The ARE tends to $1000/837 \approx 119.47\%$ as $k \to 0$ or ∞.

More generally, it follows from relation (10) that $\bar{\xi}'_\theta = 2\rho'_\theta - \xi'_\theta$, whence

$$\begin{aligned} \frac{31}{30}\,\mathrm{ARE}\left(T_{\bar{\xi}}, T_\rho\right) &= \left(\frac{\bar{\xi}'_0}{\rho'_0}\right)^2 = \left(2 - \frac{\xi'_0}{\rho'_0}\right)^2 \\ &= \left\{2 - \sqrt{\frac{31}{30}\,\mathrm{ARE}\left(T_\xi, T_\rho\right)}\right\}^2, \end{aligned}$$

so that the two ARE's are monotone decreasing functions of each other. Accordingly,

$$\frac{31}{30}\,\max\left\{\mathrm{ARE}\left(T_\xi, T_\rho\right), \mathrm{ARE}\left(T_{\bar{\xi}}, T_\rho\right)\right\} = \max\left\{x, \left(2-\sqrt{x}\right)^2\right\},$$

where $x = (31/30) \times \mathrm{ARE}\left(T_\xi, T_\rho\right) \geq 0$. Since the right-hand side is minimized when $x = 1$,

$$\max\left\{\mathrm{ARE}\left(T_{\bar{\xi}}, T_\rho\right), \mathrm{ARE}\left(T_\xi, T_\rho\right)\right\} \geq \frac{30}{31} \approx 96.77\%.$$

Moreover, at least one of T_ξ or $T_{\bar{\xi}}$ provides an improvement over Spearman's test unless

$$\sqrt{x} = \frac{\xi_0'}{\rho_0'} \in \left(2 - \sqrt{\frac{31}{30}}, \sqrt{\frac{31}{30}}\right) \approx (0.9834, 1.0165).$$

From the above examples and the authors' experience with other copula models, it would appear that for "smooth" families of distributions, the largest possible asymptotic relative efficiency attainable with either T_ξ or $T_{\bar{\xi}}$ is 1000/837, i.e., when $\sqrt{x} = 9/10$ or $10/9$, as in Examples 5, 6 and 8. The exact conditions under which this occurs remain to be determined, however.

1.5.2 Power Comparisons in Finite Samples

To examine the performance of tests of independence based on ξ_n and $\bar{\xi}_n$, Monte Carlo simulations were carried out for various sample sizes and families of copulas spanning all possible degrees of association between stochastic independence ($\rho = 0$) and complete positive dependence ($\rho = 1$), whose underlying copula is the Fréchet upper bound $M(u, v) = \min(u, v)$.

Results reported herein are for pseudo-random samples of size $n = 25$ and 100 for the normal (which is archetypical of radially symmetric copulas), the Clayton (Example 5), and the bivariate extreme value distributions of Gumbel (1960b) and Galambos (1975). The latter two have copulas of the form

$$C_\theta(u, v) = \exp\left[\log(uv) A\left\{\frac{\log(u)}{\log(uv)}\right\}\right]$$

with

$$A(t) = \left\{t^\theta + (1-t)^\theta\right\}^{1/\theta}, \quad \theta \geq 1$$

for Gumbel's model and

$$A(t) = 1 - \left\{t^{-\theta} + (1-t)^{-\theta}\right\}^{-1/\theta}, \quad \theta \geq 0$$

for Galambos' model ; see for example Ghoudi, Khoudraji & Rivest (1998).

Figures 3 to 6 compare the power of the four (two-sided) tests under the four selected models when $n = 25$ (left panel) and $n = 100$ (right panel). These curves are based on 5000 replicates. In each case, the test statistic was standardized using its exact variance under the null hypothesis of independence and compared to the 95th centile of the asymptotic standard normal distribution.

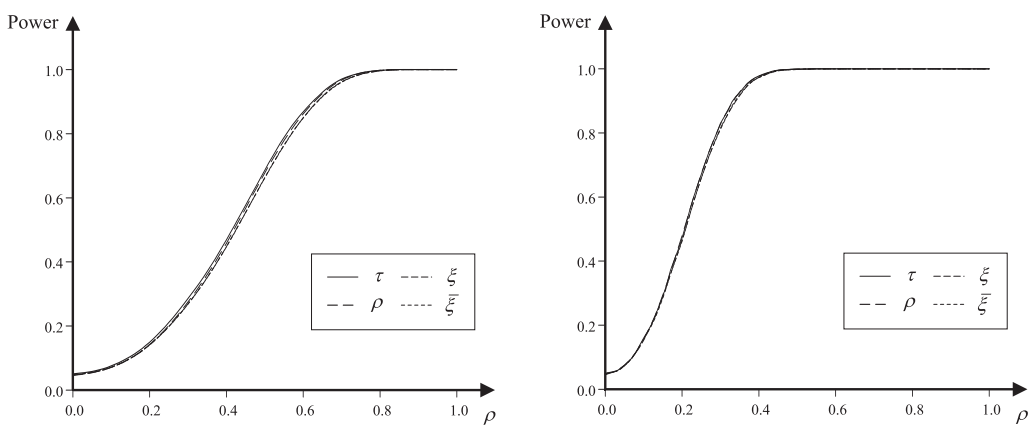


FIGURE 3. Power curve for rank tests of independence of level $\alpha = 5\%$ based on ρ_n, τ_n, ξ_n and $\bar{\xi}_n$, drawn as a function of Spearman's rho for random samples of size $n = 25$ (left panel) and $n = 100$ (right panel) from the bivariate normal distribution.

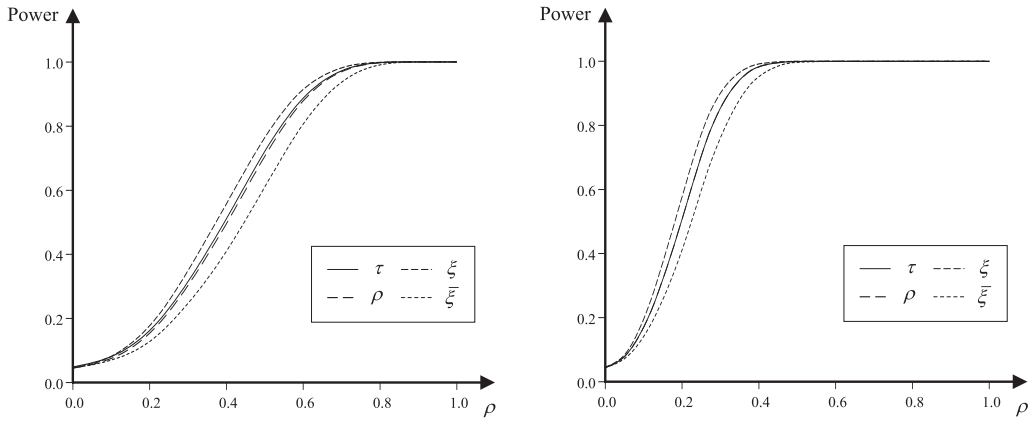


FIGURE 4. Power curve for rank tests of independence of level $\alpha = 5\%$ based on ρ_n, τ_n, ξ_n and $\bar{\xi}_n$, drawn as a function of Spearman's rho for random samples of size $n = 25$ (left panel) and $n = 100$ (right panel) from Clayton's bivariate copula.

According to Figure 3, there is very little to choose between the four procedures when the underlying dependence structure is normal, even when n is small. Although strictly speaking, the test based on Kendall's tau is best, its slight advantage tends to be attenuated as the sample size increases; and as expected, power generally increases with the sampling effort.

As suggested by Example 5, the test based on ξ_n should be preferable to those based on Spearman's rho or Kendall's tau in Clayton's model. This is confirmed in Figure 4, where the good performance of T_ξ is compensated by the comparative lack of power of $T_{\bar{\xi}}$, as already discussed in Section 5.1.

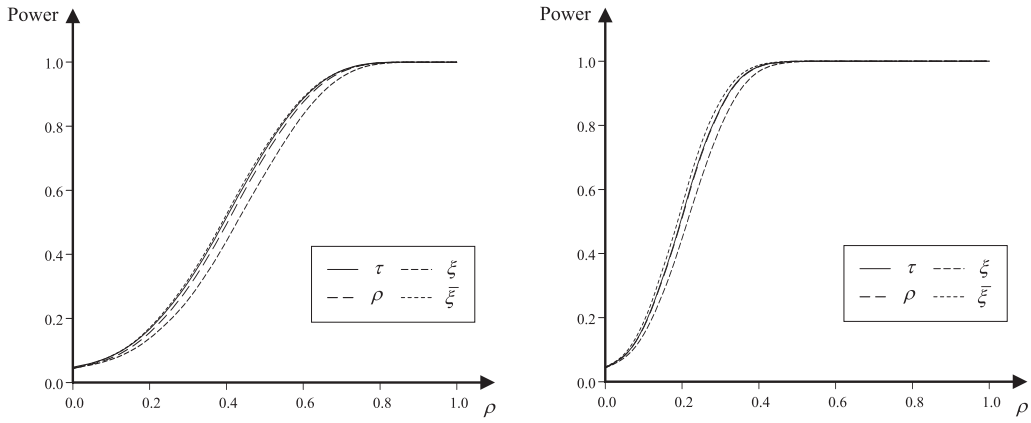


FIGURE 5. Power curve for rank tests of independence of level $\alpha = 5\%$ based on ρ_n, τ_n, ξ_n and $\bar{\xi}_n$, drawn as a function of Spearman's rho for random samples of size $n = 25$ (left panel) and $n = 100$ (right panel) from Gumbel's bivariate extreme value copula.

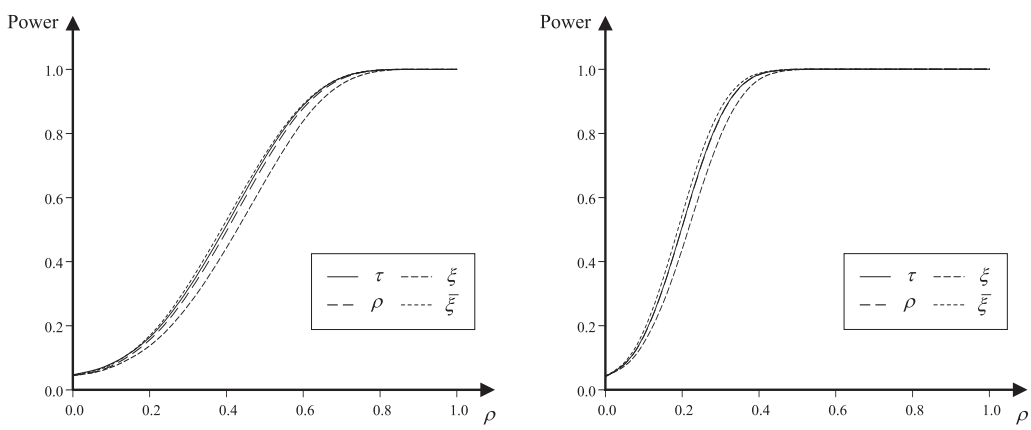


FIGURE 6. Power curve for rank tests of independence of level $\alpha = 5\%$ based on ρ_n, τ_n, ξ_n and $\bar{\xi}_n$, drawn as a function of Spearman's rho for random samples of size $n = 25$ (left panel) and $n = 100$ (right panel) from Galambos' bivariate extreme value copula.

Finally, Figures 5 and 6 provide two examples of extreme value distributions in which greater power accrues from the use of the statistic $\bar{\xi}_n$ than from either Spearman's rho, Kendall's tau or the symmetrized version of Blest's coefficient.

1.6 Conclusion

This paper has continued the work of Blest (2000) by showing that his coefficient is asymptotically normal with parameters for which an explicit form was given in several instances. A symmetric version of his measure was also proposed which is highly correlated with Spearman's rho while retaining Blest's idea that greater emphasis should be given to discrepancies in the first ranks. The new measure, whose limiting distribution is also normal, was compared to Spearman's rho and Kendall's tau as a test statistic for independence, both through simulations in small samples and in terms of asymptotic relative efficiency. It was shown that non-negligible improvements in power are possible, either when the test is based on the symmetrized version ξ_n of Blest's coefficient, or on a complementary statistic $\bar{\xi}_n$ involving reverse ranks.

It would be of interest, in future work, to characterize the type of alternatives to independence for which ξ_n is preferable to $\bar{\xi}_n$. A more ambitious project would be to identify polynomials $Q(u, v)$ and families of copulas for which an empirical coefficient of the form $\int Q(u, v)\, \mathrm{d}C_n(u, v)$ would be a powerful test statistic. Finally, in the spirit of Hallin & Puri (1992) or Ferguson, Genest & Hallin (2000), the merits of Blest's indices could also be investiga-

ted as a measure of serial dependence or as a test of randomness in a time series context.

1.7 Appendix

Explicit formulas for

$$\operatorname{var}(\xi_n) = \frac{36}{(n^3-n)^2} \operatorname{var}\left\{\sum_{i=1}^n R_i S_i \left(4 - \frac{R_i+S_i}{n+1}\right)\right\}$$

and

$$\operatorname{corr}(\xi_n, \rho_n) = \operatorname{corr}\left\{\sum_{i=1}^n R_i S_i \left(4 - \frac{R_i+S_i}{n+1}\right) \ , \ \sum_{i=1}^n R_i S_i\right\}$$

can be found under the assumption of independence through repeated use of the following elementary result.

Lemma. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from some continuous distribution H, and let $(R_1, S_1), \ldots, (R_n, S_n)$ be its associated set of ranks. If J, K, L and M are real-valued functions defined on the integers $\{1, \ldots, n\}$, then*

$$\mathrm{E}\left\{\sum_{i=1}^n J(R_i)K(S_i)\right\} = \frac{1}{n}\left\{\sum_{i=1}^n J(i)\right\}\left\{\sum_{j=1}^n K(j)\right\}.$$

and

$$\mathrm{E}\left\{\sum_{i\neq k} J(R_i)K(S_i)L(R_k)M(S_k)\right\}$$

$$= \frac{1}{n(n-1)}\left\{\sum_{i\neq k} J(i)L(k)\right\}\left\{\sum_{j\neq \ell} K(j)M(\ell)\right\}.$$

Proof. The first identity can be found, e.g., in the book of Hájek (1969, Theorem 24B, p. 117). The second one is undoubtedly known as well, but harder to locate. A proof is included here for completeness.

Without loss of generality, one may write

$$\sum_{i \neq k} J(R_i)K(S_i)L(R_k)M(S_k) = \sum_{i \neq k} J(i)K(q_{t,i})L(k)M(q_{t,k})$$

for a specific element $Q_t = (q_{t,1}, \ldots, q_{t,n})$ in the collection $Q = \{Q_1, \ldots, Q_{n!}\}$ of permutations of the vector $(1, \ldots, n)$. Under the hypothesis of independence, all points in Q are equally likely. Thus if Q_T denotes a random permutation in this set, one has

$$\begin{aligned} \mathrm{E}\left\{\sum_{i \neq k} J(R_i)K(S_i)L(R_k)M(S_k)\right\} &= \mathrm{E}\left\{\sum_{i \neq k} J(i)K(q_{T,i})L(k)M(q_{T,k})\right\} \\ &= \frac{1}{n!}\sum_{t=1}^{n!}\sum_{i \neq k} J(i)K(q_{t,i})L(k)M(q_{t,k}) \\ &= \frac{1}{n!}\sum_{i \neq k} J(i)L(k)\sum_{t=1}^{n!} K(q_{t,i})M(q_{t,k}) \end{aligned}$$

Now for arbitrary integers $i, j, k, \ell \in \{1, \ldots, n\}$ with $i \neq k$ and $j \neq \ell$, the event $(q_{t,i}, q_{t,k}) = (j, \ell)$ occurs exactly $(n-2)!$ times as t ranges over $1, \ldots, n!$. Therefore,

$$\sum_{t=1}^{n!} K(q_{t,i})M(q_{t,k}) = (n-2)!\sum_{j \neq \ell} K(j)M(\ell),$$

which yields the second identity. □

As an example of application,

$$\begin{aligned}
\mathrm{E}\left\{\sum_{i=1}^{n} R_i S_i \left(4 - \frac{R_i + S_i}{n+1}\right)\right\} &= 4\,\mathrm{E}\left(\sum_{i=1}^{n} R_i S_i\right) - \frac{2}{n+1}\,\mathrm{E}\left(\sum_{i=1}^{n} R_i^2 S_i\right) \\
&= \frac{4}{n}\left(\sum_{i=1}^{n} i\right)^2 - \frac{2}{n^3 - n}\left(\sum_{i=1}^{n} i\right)\left(\sum_{i=1}^{n} i^2\right) \\
&= \frac{n(n+1)(4n+5)}{6}\ ,
\end{aligned}$$

from which it follows that $\mathrm{E}(\xi_n) = 0$ under the assumption of independence. All other computations are similar and are easily performed with a symbolic calculator such as Maple.

CHAPITRE II

MESURES DE CORRÉLATION BASÉES SUR LES RANGS

L'objectif ambitieux de ce chapitre consiste à partager avec le lecteur l'intuition et les raisonnements qui sous-tendent les résultats présentés au chapitre I.

2.1 Terminologie

Au chapitre I, les axiomes de Scarsini offrent une liste de propriétés que devrait idéalement posséder toute mesure de concordance. Il est clair qu'une statistique ne répondant pas au moins aux axiomes a, b et d de Scarsini ne saurait être considérée comme "mesure de corrélation." Les seules mesures considérées dans la suite sont celles basées sur les rangs.

2.2 Formes des coefficients

Dans le chapitre précédent, les mesures de corrélation utilisées (à l'exception de τ) sont toutes de la forme

$$\kappa_n = a_n \sum_{i=1}^{n} p\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right) + b_n,$$

où $p(u, v)$ est un polynôme et a_n, b_n sont des constantes choisies de telle sorte que $\kappa_n \in [-1, 1]$. Dans tous les cas, les valeurs ± 1 ne sont prises que si les rangs R_i et S_i sont les mêmes ou sont antithétiques (c'est-à-dire $S_i = n + 1 - R_i$, pour $1 \leq i \leq n$).

Les quatre remarques suivantes présentent des simplifications qui permettent de se concentrer sur ce qui constitue vraiment un coefficient de corrélation des rangs.

2.2.1 Élimination des marges

Comme les rangs sont invariants sous toute transformation monotone croissante des marges, ils ne dépendent pas des lois marginales des observations. Ainsi, sans perte de généralité, on peut se limiter à considérer des lois dont les marges sont uniformes, à savoir des copules !

2.2.2 Paramètre estimé et loi asymptotique

David C. Blest a défini un coefficient de corrélation échantillonnal sans toutefois préciser sa valeur dans la population. En général, il est possible de déterminer quel paramètre estime un coefficient κ_n basé sur les rangs grâce à une généralisation d'un résultat de Ruymgaart, Shorack & van Zwet (1972) proposée par Genest, Ghoudi & Rivest (1995). On obtient en boni la loi asymptotique de κ_n.

Puisque le polynôme $p(u, v)$ est une fonction continue définie sur le compact $[0, 1]^2$, les conditions de régularité de leur proposition A·1 sont toujours

respectées. Ainsi,

$$\sqrt{n}(\kappa_n - \kappa) \xrightarrow{\text{loi}} N(0, \sigma^2)$$

avec

$$\kappa = a\,\mathrm{E}\{p(U,V)\} + b = a \int p(u,v)\,\mathrm{d}C(u,v) \;+\; b$$

et

$$\sigma^2 = a^2\,\mathrm{var}\Bigg[\, p(U,V) + \int \mathbb{1}(U \le u)\,\frac{\partial}{\partial u}\,p(u,v)\,\mathrm{d}u \\ + \int \mathbb{1}(V \le v)\,\frac{\partial}{\partial v}\,p(u,v)\,\mathrm{d}v\Bigg],$$

où $a = \lim_{n\to\infty} na_n$ et $b = \lim_{n\to\infty} b_n$. Bref, peu importe le polynôme $p(u,v)$ utilisé, κ_n estime $\kappa = a\,\mathrm{E}\{p(U,V)\} + b$ sans biais.

2.2.3 Degré des polynômes considérés

Tous les termes ne contenant pas d'interaction (c'est-à-dire les monômes u^k et v^k pour $k = 0, 1, \ldots$) constituent en fait des constantes pouvant être incluses dans b_n lorsque sommés ou intégrés. Par exemple, $p(u,v) = (1-u)^2 v$ et $p(u,v) = u^2 v - 2uv$ conduisent tous deux à la mesure de Blest.

2.2.4 Rôle des constantes

Comme les constantes a_n et b_n sont choisies pour que $\kappa_n \in [-1, 1]$, les multiples d'un polynôme mènent au même coefficient de corrélation. Par exemple, $p(u,v) = uv - u^2 v/2$ conduit aussi à la mesure de Blest.

En laissant de côté les constantes et les symboles de sommation, on écrit dans la suite

$$\kappa_n \sim p(U,V) \qquad \text{ou} \qquad \kappa \sim p(U,V)$$

pour signifier que κ est la mesure de corrélation basée sur $p(u,v)$, c'est-à-dire que

$$\kappa_n = a_n \sum_{i=1}^{n} p\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right) + b_n$$

et

$$\kappa = a \int p(u,v)\,\mathrm{d}C + b\ .$$

Cette notation plus sobre permet de mieux comprendre la nature des coefficients envisagés.

2.3 Mesures considérées

Les mesures étudiées dans ce mémoire sont

$$\begin{aligned}
\rho_n &\sim uv\,, \\
\nu_n &\sim (1-u)^2 v \ \sim\ uv(2-u)\,, \\
\overline{\nu}_n &\sim u^2 v \ \sim\ uv(u)\,, \\
\xi_n &\sim uv(4-u-v) \quad \text{et} \\
\overline{\xi}_n &\sim uv(u+v)\,.
\end{aligned}$$

Seul $\overline{\nu}_n$ n'est pas mentionné dans le chapitre I. Il s'agit du coefficient dual de ν_n, évoqué dans l'article de Blest.

L'écriture dépouillée des sommes et des constantes donne une meilleure vision de la nature des coefficients. Ainsi, ρ_n peut être interprété comme une simple mesure d'interaction, alors que ν_n, $\overline{\nu}_n$, ξ_n et $\overline{\xi}_n$ sont des versions pondérées de cette même interaction. Il est à noter que contrairement à ν_n et ξ_n, $\overline{\nu}_n$ et $\overline{\xi}_n$ accordent plus d'importance aux écarts observés dans les derniers rangs que dans les premiers. De telles mesures pourraient s'avérer utiles pour comparer deux jeux de données dans lesquels des inversions constatées dans les grandes valeurs importent beaucoup ; pensons par exemple aux notes accordées par deux juges lors d'une épreuve de patinage artistique.

Tel que mentionné précédemment, les termes n'impliquant qu'une seule variable peuvent être relégués aux constantes. Ainsi, il est toujours possible de mettre uv en évidence et d'écrire

$$\kappa_n \sim uv\, q(u,v)$$

pour un certain polynôme $q(u,v)$ qui pondère l'interaction. Malheureusement, tous les polynômes $q(u,v)$ ne conduisent pas à des mesures de corrélation acceptables. Une interprétation géométrique du comportement des coefficients liée à la forme de $q(u,v)$ permettrait d'appuyer l'intuition et aiderait sans doute à cerner les polynômes acceptables, mais aucun progrès significatif n'a été réalisé en ce sens dans le cadre de ce mémoire.

2.4 Résumé des propriétés

2.4.1 Dualité et symétrie radiale

Si une copule est radialement symétrique, les couples (U, V) et $(1-U, 1-V)$ sont de même loi. Cette propriété se traduit par le fait que (R_i, S_i) et $(\overline{R}_i, \overline{S}_i) = (n+1-R_i, n+1-S_i)$ ont la même distribution.

Soit κ_n une mesure de corrélation basée sur les rangs et $\overline{\kappa}_n$ le même coefficient calculé au moyen des rangs inversés $(\overline{R}_i, \overline{S}_i)$. On appelle $\overline{\kappa}_n$ le dual de κ_n. Les deux indices ont la même loi (puisqu'il s'agit de la même fonction calculée sur des variables aléatoires de même loi). Ainsi, on aura toujours

$$\mathrm{var}(\kappa_n) = \mathrm{var}(\overline{\kappa}_n)$$

et

$$\kappa = \mathrm{E}(\kappa_n) = \mathrm{E}(\overline{\kappa}_n) = \overline{\kappa}\ .$$

Comme les coefficients $\overline{\nu}_n$ et $\overline{\xi}_n$ sont les duaux de ν_n et ξ_n, les propriétés ci-haut continuent de valoir. En particulier, comme la copule d'indépendance est radialement symétrique, on a

$$\mathrm{var}(\nu_n) = \mathrm{var}(\overline{\nu}_n)\ , \quad \mathrm{var}(\xi_n) = \mathrm{var}(\overline{\xi}_n)$$

et

$$\nu = \overline{\nu} = \xi = \overline{\xi}$$

sous l'hypothèse d'indépendance.

2.4.2 Dualité et test d'indépendance

Soit

$$\gamma_n = \frac{\kappa_n + \overline{\kappa}_n}{2} \ .$$

Sous l'hypothèse d'indépendance, $\gamma = \kappa = \overline{\kappa}$. Par le même argument que celui utilisé au chapitre I, on a alors :

$$\max\left\{\mathrm{ARE}(T_\kappa, T_\gamma), \mathrm{ARE}(T_{\overline{\kappa}}, T_\gamma)\right\} \geq \frac{\mathrm{var}(\gamma_n)}{\mathrm{var}(\kappa_n)}$$

avec égalité si la copule est radialement symétrique car alors, $\kappa_0' = \gamma_0'$.

Dans le cas de ν_n et de ξ_n, $(\nu_n + \overline{\nu}_n)/2 = \left(\xi_n + \overline{\xi}_n\right)/2 = \rho_n$. Par conséquent, la borne obtenue compare les coefficients au ρ de Spearman, ce qui la rend beaucoup plus intéressante.

2.4.3 Liens linéaires

Voici quelques relations linéaires élémentaires entre les différents coefficients issus des travaux de Blest.

$$\boxed{\begin{aligned} \nu_n &= 2\rho_n - \overline{\nu}_n \\ \xi_n &= 2\rho_n - \overline{\xi}_n \\ \xi(X,Y) &= \frac{\nu(X,Y) + \nu(Y,X)}{2} \\ \overline{\xi}(X,Y) &= \frac{\overline{\nu}(X,Y) + \overline{\nu}(Y,X)}{2} \end{aligned}}$$

2.4.4 Propriétés individuelles

Pour faciliter d'éventuelles références, les tableaux suivants résument les propriétés du ρ de Spearman et des variantes du coefficient de Blest. La façon de déduire ces informations est expliquée aux chapitres I et III.

Rho de Spearman (ρ)

$$\rho_n = -3\,\frac{n+1}{n-1} + \frac{12}{n(n+1)(n-1)}\sum_{i=1}^{n} R_i S_i$$

$$\begin{aligned}\rho &= -3 + 12\int uv\,\mathrm{d}C(u,v)\\ &= -3 + 12\int C(u,v)\,\mathrm{d}u\,\mathrm{d}v\end{aligned}$$

Sous l'hypothèse d'indépendance entre X et Y, on trouve

$$\mathrm{var}\left(\sqrt{n}\,\rho_n\right) = \frac{n}{n-1} \overset{n\to\infty}{\longrightarrow} 1$$

Axiomes de Scarsini :

a	b	c	d	e	f	g
•	•	•	•	•	•	•

Coefficients de Blest (ν et $\overline{\nu}$)

$$\begin{aligned}
\nu_n &= \frac{2n+1}{n-1} - \frac{12}{n(n+1)^2(n-1)} \sum_{i=1}^{n} (n+1-R_i)^2 S_i \\
\overline{\nu}_n &= -\frac{2n+1}{n-1} + \frac{12}{n(n+1)^2(n-1)} \sum_{i=1}^{n} R_i^2 S_i
\end{aligned}$$

$$\begin{aligned}
\nu &= 2 - 12 \int (1-u)^2 v \, \mathrm{d}C(u,v) \\
&= -2 + 24 \int (1-u) C(u,v) \, \mathrm{d}u \, \mathrm{d}v
\end{aligned}$$

$$\begin{aligned}
\overline{\nu} &= -2 + 12 \int u^2 v \, \mathrm{d}C(u,v) \\
&= -4 + 24 \int u C(u,v) \, \mathrm{d}u \, \mathrm{d}v
\end{aligned}$$

Sous l'hypothèse d'indépendance entre X et Y, on trouve

$$\begin{aligned}
\mathrm{var}\left(\sqrt{n}\,\nu_n\right) = \mathrm{var}\left(\sqrt{n}\,\overline{\nu}_n\right) &= \frac{n(2n+1)(8n+11)}{15(n+1)^2(n-1)} \xrightarrow{n\to\infty} \frac{16}{15} \\
\mathrm{corr}(\nu_n, \rho_n) &= \sqrt{\frac{15(n+1)^2}{(2n+1)(8n+11)}} \xrightarrow{n\to\infty} \sqrt{\frac{15}{16}} \\
\mathrm{corr}(\nu_n, \overline{\nu}_n) &= \frac{14n^2+30n+19}{(2n+1)(8n+11)} \xrightarrow{n\to\infty} \frac{7}{8}
\end{aligned}$$

Axiomes de Scarsini :

a	b	c	d	e	f	g
•	•		•		•	•

Coefficients de Plantagenet (ξ et $\overline{\xi}$)

$$\begin{aligned}
\xi_n &= -\frac{4n+5}{n-1} + \frac{6}{n(n+1)(n-1)} \sum_{i=1}^{n} R_i S_i \left(4 - \frac{R_i + S_i}{n+1}\right) \\
\overline{\xi}_n &= -\frac{2n+1}{n-1} + \frac{6}{n(n+1)^2(n-1)} \sum_{i=1}^{n} R_i S_i \left(R_i + S_i\right)
\end{aligned}$$

$$\begin{aligned}
\xi &= -4 + 6 \int uv(4-u-v)\,\mathrm{d}C(u,v) \\
&= -2 + 12 \int (2-u-v)C(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
\overline{\xi} &= -2 + 6 \int uv(u+v)\,\mathrm{d}C(u,v) \\
&= -4 + 12 \int (u+v)C(u,v)\,\mathrm{d}u\,\mathrm{d}v
\end{aligned}$$

Sous l'hypothèse d'indépendance entre X et Y, on trouve

$$\begin{aligned}
\mathrm{var}\left(\sqrt{n}\,\xi_n\right) = \mathrm{var}\left(\sqrt{n}\,\overline{\xi}_n\right) &= \frac{n(31n^2+60n+26)}{30(n+1)^2(n-1)} \xrightarrow{n\to\infty} \frac{31}{30} \\
\mathrm{corr}(\xi_n, \rho_n) &= \sqrt{\frac{30(n+1)^2}{31n^2+60n+26}} \xrightarrow{n\to\infty} \sqrt{\frac{30}{31}} \\
\mathrm{corr}(\xi_n, \nu_n) &= \sqrt{\frac{31n^2+60n+26}{2(2n+1)(8n+11)}} \xrightarrow{n\to\infty} \sqrt{\frac{31}{32}} \\
\mathrm{corr}(\xi_n, \overline{\xi}_n) &= \frac{29n^2+60n+34}{31n^2+60n+26} \xrightarrow{n\to\infty} \frac{29}{31}
\end{aligned}$$

Axiomes de Scarsini :

a	b	c	d	e	f	g
•	•	•	•		•	•

CHAPITRE III

CALCULS EXPLICITES

Ce chapitre explicite quelques-uns des calculs qui ont conduit aux résultats des chapitres précédents.

3.1 Variance des coefficients sous l'hypothèse d'indépendance

Les calculs de cette section exploitent essentiellement le résultat de la section 1.7.

3.1.1 Rho de Spearman

Rappelons que

$$\rho_n = -3\,\frac{n+1}{n-1} + \frac{12}{n(n+1)(n-1)}\,\sum_{i=1}^{n} R_i S_i$$

et que

$$\mathrm{var}(\rho_n) = \frac{144}{n^2(n+1)^2(n-1)^2}\,\mathrm{var}\left(\sum_{i=1}^{n} R_i S_i\right)\ .$$

On doit donc calculer, à l'aide du résultat de la section 1.7 :

$$\mathrm{E}\left(\sum_{i=1}^{n} R_i S_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n} i\right)^2$$

et

$$\begin{aligned}
\mathrm{E}\left\{\left(\sum_{i=1}^{n} R_i S_i\right)^2\right\} &= \mathrm{E}\left(\sum_{i=1}^{n} R_i^2 S_i^2 + \sum_{i\neq j} R_i R_j S_i S_j\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} i^2\right)^2 + \frac{1}{n(n-1)}\left(\sum_{i\neq j} ij\right)^2 .
\end{aligned}$$

En se servant de MAPLE, on trouve :

```
> A1 :=factor(simplify(sum(i,i=1..n)^2/n)) ;
```

$$A1 := \frac{1}{4}\, n\,(n+1)^2$$

```
> B1 :=factor(simplify(sum(i^2,i=1..n)^2/n+
  sum(i*(sum(j,j=1..n)-i),i=1..n)^2/n/(n-1))) ;
```

$$B1 := \frac{1}{144}\,(9\,n^2 + 19\,n + 8)\,(n+1)^2\,n^2$$

```
> C1 :=factor(simplify(B1-A1^2)) ;
```

$$C1 := \frac{1}{144}\,n^2\,(n-1)\,(n+1)^2$$

```
> simplify(C1*144/(n^2*(n-1)^2*(n+1)^2)) ;
```

$$\frac{1}{n-1}$$

Par conséquent,

$$\boxed{\mathrm{var}(\rho_n) = \frac{1}{n-1}\ .}$$

3.1.2 Coefficient de Blest

Rappelons que

$$\nu_n = \frac{2n+1}{n-1} - \frac{12}{n(n+1)^2(n-1)} \sum_{i=1}^{n} (n+1-R_i)^2 S_i$$

et que

$$\overline{\nu}_n = -\frac{2n+1}{n-1} + \frac{12}{n(n+1)^2(n-1)} \sum_{i=1}^{n} R_i^2 S_i\ .$$

Sous l'hypothèse d'indépendance, (R_i, S_i) et $(\overline{R}_i, \overline{S}_i) = (n+1-R_i, n+1-S_i)$ ont la même distribution. Puisque $\overline{\nu}_n$ est égal à ν_n calculé à partir des rangs inversés,

$$\mathrm{var}(\nu_n) = \mathrm{var}(\overline{\nu}_n) = \frac{144}{n^2(n+1)^4(n-1)^2}\ \mathrm{var}\left(\sum_{i=1}^{n} R_i^2 S_i\right)\ .$$

En exploitant le résultat de la section 1.7, on trouve

$$\mathrm{E}\left(\sum_{i=1}^{n} R_i^2 S_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n} i^2\right)\left(\sum_{i=1}^{n} i\right)$$

et

$$\mathrm{E}\left\{\left(\sum_{i=1}^{n} R_i^2 S_i\right)^2\right\} = \mathrm{E}\left(\sum_{i=1}^{n} R_i^4 S_i^2 + \sum_{i\neq j} R_i^2 R_j^2 S_i S_j\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} i^4\right)\left(\sum_{i=1}^{n} i^2\right) + \frac{1}{n(n-1)}\left(\sum_{i\neq j} i^2 j^2\right)\left(\sum_{i\neq j} ij\right).$$

Des calculs assistés par Maple montrent que :

```
> A2 := factor(simplify(sum(i,i=1..n)*sum(i^2,i=1..n)/n)) ;
```

$$A2 := \frac{1}{12}\, n\,(n+1)^2\,(2\,n+1)$$

```
> B2 :=factor(simplify(sum(i^4,i=1..n)*sum(i^2,i=1..n)/n+
  sum(i^2*(sum(j^2,j=1..n)-i^2),i=1..n)*
  sum(i*(sum(j,j=1..n)-i),i=1..n)/n/(n-1))) ;
```

$$B2 := \frac{1}{2160}\,(2\,n+1)\,(30\,n^3+83\,n^2+63\,n+4)\,(n+1)^2\,n^2$$

```
> C2 :=factor(simplify(B2-A2^2)) ;
```

$$C2 := \frac{1}{2160}\, n^2\,(n-1)\,(2\,n+1)\,(8\,n+11)\,(n+1)^2$$

```
> simplify(C2*144/(n^2*(n-1)^2*(n+1)^4)) ;
```

$$\frac{1}{15}\,\frac{(2\,n+1)\,(8\,n+11)}{(n-1)\,(n+1)^2}$$

Donc,

$$\boxed{\mathrm{var}(\nu_n) = \mathrm{var}(\overline{\nu}_n) = \frac{(2\,n+1)\,(8\,n+11)}{15\,(n+1)^2\,(n-1)}\ .}$$

3.1.3 Coefficient de Plantagenet

Rappelons que

$$\xi_n = -\frac{4n+5}{n-1} - \frac{6}{n(n+1)(n-1)}\ \sum_{i=1}^{n} R_i S_i \left(4 - \frac{R_i+S_i}{n+1}\right)$$

et que

$$\overline{\xi}_n = -\frac{2n+1}{n-1} + \frac{6}{n(n+1)^2(n-1)}\ \sum_{i=1}^{n} R_i S_i\,(R_i+S_i)\ .$$

Comme précédemment, sous l'hypothèse d'indépendance, (R_i, S_i) a la même distribution que $(\overline{R}_i, \overline{S}_i)$ et $\overline{\xi}_n$ est le dual de ξ_n. Ainsi,

$$\mathrm{var}(\xi_n) = \mathrm{var}(\overline{\xi}_n) = \frac{36}{n^2(n+1)^4(n-1)^2}\ \mathrm{var}\left\{\sum_{i=1}^{n} R_i S_i (R_i+S_i)\right\}\ .$$

Or,

$$\begin{aligned}
\mathrm{E}\left\{\sum_{i=1}^{n} R_i S_i (R_i+S_i)\right\} &= \mathrm{E}\left(\sum_{i=1}^{n} R_i^2 S_i\right) + \mathrm{E}\left(\sum_{i=1}^{n} R_i S_i^2\right) \\
&= \frac{2}{n}\left(\sum_{i=1}^{n} i^2\right)\left(\sum_{i=1}^{n} i\right)
\end{aligned}$$

et

$$
\begin{aligned}
&\mathrm{E}\left[\left\{\sum_{i=1}^{n} R_i S_i (R_i + S_i)\right\}^2\right] \\
&\qquad = \mathrm{E}\left\{\sum_{i=1}^{n} R_i^2 S_i^2 (R_i + S_i)^2 + \sum_{i \neq j} R_i S_i (R_i + S_i)(R_j + S_j)\right\} \\
&\qquad = \mathrm{E}\Bigg\{\sum_{i=1}^{n} \left(R_i^4 S_i^2 + 2Ri^3 S_i^3 + R_i^2 S_i^4\right) + \\
&\qquad\qquad \sum_{i \neq j} \left(R_i^2 R_j^2 S_i S_j + R_i^2 R_j S_i S_j^2 + R_i R_j^2 S_i^2 S_j + R_i R_j S_i^2 S_j^2\right)\Bigg\} \\
&\qquad = \frac{2}{n}\left\{\left(\sum_{i=1}^{n} i^4\right)\left(\sum_{i=1}^{n} i^2\right) + \left(\sum_{i=1}^{n} i^3\right)^2\right\} \\
&\qquad\qquad + \frac{2}{n(n-1)}\left\{\left(\sum_{i \neq j} i^2 j^2\right)\left(\sum_{i \neq j} ij\right) + \left(\sum_{i \neq j} i^2 j\right)^2\right\}.
\end{aligned}
$$

À l'aide de Maple, on trouve :

```
> A3 :=factor(simplify(2*sum(i,i=1..n)*sum(i^2,i=1..n)/n)) ;
```

$$A3 := \frac{1}{6}\, n\, (n+1)^2\, (2\, n + 1)$$

```
> B3 :=factor(simplify(2*(sum(i^4,i=1..n)*
  sum(i^2,i=1..n)+sum(i^3,i=1..n)^2)/n+
  2*(sum(i^2*(sum(j^2,j=1..n)-i^2),i=1..n)*
  sum(i*(sum(j,j=1..n)-i),i=1..n)+
  sum(i^2*(sum(j,j=1..n)-i),i=1..n)^2)/n/(n-1))) ;
```

$$B3 := \frac{1}{1080}\,(120\,n^4 + 391\,n^3 + 419\,n^2 + 146\,n + 4)\,(n+1)^2\,n^2$$

```
> C3 :=factor(simplify(B3-A3^2)) ;
```

$$C3 := \frac{1}{1080}\,n^2\,(n-1)\,(31\,n^2 + 60\,n + 26)\,(n+1)^2$$

```
> simplify(C3*36/(n^2*(n-1)^2*(n+1)^4)) ;
```

$$\frac{1}{30}\,\frac{31\,n^2 + 60\,n + 26}{(n-1)\,(n+1)^2}$$

Bref,

$$\boxed{\mathrm{var}(\xi_n) = \mathrm{var}(\overline{\xi}_n) = \frac{31n^2 + 60n + 26}{30(n+1)^2(n-1)}\ .}$$

3.2 Corrélation entre les coefficients sous l'hypothèse d'indépendance

Sous l'hypothèse d'indépendance, les coefficients de Blest et de Plantagenet ainsi que leurs duaux estiment la même mesure, à savoir ρ. À l'aide du résultat de la section 1.7, la corrélation entre certains des coefficients peut être calculée. Les calculs sont présentés ci-après dans les cas les plus intéressants, mais peuvent être effectués de la même façon pour n'importe quel choix de coefficients.

Une simplification sera effectuée pour chacun des calculs. Elle est rappelée d'entrée de jeu :

$$\operatorname{corr}(aX + b, cY + d) = \operatorname{corr}(X, Y).$$

Les résultats des calculs de la section précédente seront utilisées à l'occasion, auquel cas les noms de variables déjà utilisés apparaîtront sans commentaires supplémentaires. Par exemple, `A1` fait référence à $\mathrm{E}(\sum R_i S_i)$ obtenu à la section 3.1.1.

3.2.1 Blest vs Spearman

Étant donné que $\overline{\rho}_n = \rho_n$ et que $(\overline{R}_i, \overline{S}_i)$ a la même distribution que (R_i, S_i) sous l'indépendance, on a

$$\operatorname{corr}(\nu_n, \rho_n) = \operatorname{corr}(\overline{\nu}_n, \rho_n) = \operatorname{corr}\left(\sum_{i=1}^{n} R_i^2 S_i, \sum_{i=1}^{n} R_i S_i\right) .$$

À l'aide du résultat de la section 1.7, on trouve

$$\begin{aligned}
\mathrm{E}\left\{\left(\sum_{i=1}^{n} R_i^2 S_i\right)\left(\sum_{i=1}^{n} R_i S_i\right)\right\} &= \mathrm{E}\left(\sum_{i=1}^{n} R_i^3 S_i^2 + \sum_{i\neq j} R_i^2 R_j S_i S_j\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} i^3\right)\left(\sum_{i=1}^{n} i^2\right) + \frac{1}{n(n-1)}\left(\sum_{i\neq j} i^2 j\right)\left(\sum_{i\neq j} ij\right) .
\end{aligned}$$

En exploitant les résultats de la section 3.1, on obtient à l'aide de Maple

```
> A4 :=factor(simplify(sum(i^3,i=1..n)*sum(i^2,i=1..n)/n+
  sum(i^2*(sum(j,j=1..n)-i),i=1..n)*
  sum(i*(sum(j,j=1..n)-i),i=1..n)/n/(n-1))) ;
```

$$A4 := \frac{1}{72}\,(3\,n^2 + 5\,n + 1)\,(n+1)^3\,n^2$$

```
> assume(n>1) ;
  B4 :=factor(simplify((A4-A1*A2)/sqrt(C1*C2),assume=real)) ;
```

$$B4 := \frac{(n+1)\,\sqrt{15}}{\sqrt{(2\,n+1)\,(8\,n+11)}}$$

Conséquemment,

$$\boxed{\mathrm{corr}(\nu_n, \rho_n) = \mathrm{corr}(\overline{\nu}_n, \rho_n) = \sqrt{\frac{15(n+1)^2}{(2n+1)(8n+11)}} \overset{n\to\infty}{\longrightarrow} \sqrt{\frac{15}{16}} .}$$

3.2.2 Plantagenet vs Spearman

On veut calculer

$$\mathrm{corr}(\xi_n, \rho_n) = \mathrm{corr}(\overline{\xi}_n, \rho_n) = \mathrm{corr}\left\{\sum_{i=1}^{n} R_i S_i (R_i + S_i), \sum_{i=1}^{n} R_i S_i\right\} .$$

La première égalité découle du fait que $\overline{\rho}_n = \rho_n$ et que (R_i, S_i) et $(\overline{R}_i, \overline{S}_i)$ ont la même distribution sous l'indépendance. À l'aide du résultat de la section 1.7,

$$\begin{aligned}
&\mathrm{E}\left[\left\{\sum_{i=1}^{n} R_i S_i (R_i + S_i)\right\}\left(\sum_{i=1}^{n} R_i S_i\right)\right] \\
&\qquad = \mathrm{E}\left\{\sum_{i=1}^{n} R_i^2 S_i^2 (R_i + S_i) + \sum_{i \neq j} R_i R_j S_i S_j (R_i + S_i)\right\} \\
&\qquad = \frac{2}{n}\left(\sum_{i=1}^{n} i^3\right)\left(\sum_{i=1}^{n} i^2\right) + \frac{2}{n(n-1)}\left(\sum_{i \neq j} i^2 j\right)\left(\sum_{i \neq j} ij\right) .
\end{aligned}$$

Avec Maple, on trouve :

```
> A5 :=2*factor(simplify(sum(i^3,i=1..n)*sum(i^2,i=1..n)/n+
  sum(i^2*(sum(j,j=1..n)-i),i=1..n)*
  sum(i*(sum(j,j=1..n)-i),i=1..n)/n/(n-1))) ;
```

$$A5 := \frac{1}{36}\,(3\,n^2 + 5\,n + 1)\,(n+1)^3\,n^2$$

```
> assume(n>1) ;
  B5 :=factor(simplify((A5-A1*A3)/sqrt(C1*C3),assume=real)) ;
```

$$B5 := \frac{(n+1)\sqrt{30}}{\sqrt{31\,n^2+60\,n+26}}$$

Par conséquent,

$$\boxed{\mathrm{corr}(\xi_n,\rho_n) = \mathrm{corr}(\overline{\xi}_n,\rho_n) = \sqrt{\frac{30(n+1)^2}{31n^2+60n+26}} \overset{n\to\infty}{\longrightarrow} \sqrt{\frac{30}{31}}\,.}$$

3.2.3 Blest vs Plantagenet

Comme (R_i, S_i) et $(\overline{R}_i, \overline{S}_i)$ ont la même distribution sous l'hypothèse d'indépendance,

$$\mathrm{corr}(\xi_n,\nu_n) = \mathrm{corr}(\overline{\xi}_n,\overline{\nu}_n) = \mathrm{corr}\left\{\sum_{i=1}^n R_iS_i(R_i+S_i), \sum_{i=1}^n R_i^2S_i\right\}\,.$$

En utilisant encore une fois le résultat de la section 1.7, il vient

$$\begin{aligned}
&\mathrm{E}\left[\left\{\sum_{i=1}^n R_iS_i(R_i+S_i)\right\}\left(\sum_{i=1}^n R_i^2S_i\right)\right] \\
&\quad = \mathrm{E}\left\{\sum_{i=1}^n R_i^3S_i^2(R_i+S_i) + \sum_{i\neq j} R_iR_j^2S_iS_j(R_i+S_i)\right\} \\
&\quad = \frac{1}{n}\left\{\left(\sum_{i=1}^n i^4\right)\left(\sum_{i=1}^n i^2\right) + \left(\sum_{i=1}^n i^3\right)^2\right\} \\
&\qquad + \frac{1}{n(n-1)}\left\{\left(\sum_{i\neq j} i^2j^2\right)\left(\sum_{i\neq j} ij\right) + \left(\sum_{i\neq j} i^2j\right)^2\right\}\,.
\end{aligned}$$

Le calcul explicite nous donne, à l'aide de MAPLE :

```
> A6 :=factor(simplify((sum(i^4,i=1..n)*sum(i^2,i=1..n)+
  sum(i^3,i=1..n)^2)/n+
  (sum(i^2*(sum(j^2,j=1..n)-i^2),i=1..n)*
  sum(i*(sum(j,j=1..n)-i),i=1..n)+
  sum(i^2*(sum(j,j=1..n)-i),i=1..n)^2)/n/(n-1))) ;
```

$$A6 := \frac{1}{2160}\,(120\,n^4 + 391\,n^3 + 419\,n^2 + 146\,n + 4)\,(n+1)^2\,n^2$$

```
> assume(n>1) ;
  B6 :=factor(simplify((A6-A2*A3)/sqrt(C2*C3),assume=real)) ;
```

$$B6 := \frac{1}{2}\,\frac{(31\,n^2 + 60\,n + 26)\,\sqrt{2}}{\sqrt{(2\,n+1)\,(8\,n+11)\,(31\,n^2 + 60\,n + 26)}}$$

et donc,

$$\boxed{\mathrm{corr}(\xi_n, \nu_n) = \sqrt{\frac{31n^2 + 60n + 26}{2(2n+1)(8n+11)}} \overset{n\to\infty}{\longrightarrow} \sqrt{\frac{31}{32}}\,.}$$

3.3 Valeur théorique de certains paramètres

La valeur théorique estimée par chacun des coefficients peut être obtenue à l'aide des formules données au chapitre II. L'utilisation de la forme simplifiée

à l'aide du résultat de Quesada–Molina (1992) évite de dériver la copule, ce qui simplifie les calculs et permet de traiter sans tracas les copules possédant une composante singulière. En fait, il suffit de calculer trois intégrales pour déduire tous les coefficients considérés, à savoir :

$$\iint C(u,v)\,\mathrm{d}u\,\mathrm{d}v\ , \qquad \iint uC(u,v)\,\mathrm{d}u\,\mathrm{d}v \quad \text{et} \quad \iint vC(u,v)\,\mathrm{d}u\,\mathrm{d}v\ .$$

En effet,

$$\begin{aligned}
\rho &= -3 + 12 \iint C(u,v)\,\mathrm{d}u\,\mathrm{d}v\ , \\
\nu &= -2 + 24 \iint (1-u)C(u,v)\,\mathrm{d}u\,\mathrm{d}v\ , \\
\overline{\nu} &= -4 + 24 \iint uC(u,v)\,\mathrm{d}u\,\mathrm{d}v\ , \\
\xi &= -2 + 12 \iint (2-u-v)C(u,v)\,\mathrm{d}u\,\mathrm{d}v\ , \\
\overline{\xi} &= -4 + 12 \iint (u+v)C(u,v)\,\mathrm{d}u\,\mathrm{d}v\ .
\end{aligned}$$

L'utilisation de Maple permet d'éviter de faire manuellement des calculs simples, mais parfois fastidieux. Étant donné que les commandes sont semblables d'un cas à l'autre, seuls deux exemples sont donnés ici en détails.

3.3.1 Copule de Farlie–Gumbel–Morgenstern généralisée

La copule de Farlie–Gumbel–Morgenstern généralisée, déjà présentée à l'exemple 2 du chapitre I, est définie par

$$C_{\theta,a,b}(u,v) = uv\left\{1 + \theta(1-u^a)(1-v^b)\right\}\ ,$$

où $a, b \in [0,1]$, $\theta \in [-1,1]$ et $(u,v) \in [0,1]^2$. À l'aide de MAPLE, on obtient facilement la valeur théorique des coefficients de corrélation en fonction des paramètres de la copule.

```
> assume(theta>-1,theta<1) :
  assume(a>0,a<1) :
  assume(b>0,b<1) :
  C :=(u,v)->u*v*(1+theta*(1-u^a)*(1-v^b)) ;
```

$$C := (u,\, v) \to u\, v\, (1 + \theta\, (1 - u^a)\, (1 - v^b))$$

```
> E :=simplify(int(int(C(u,v),u=0..1),v=0..1)) :
  Eu :=simplify(int(int(u*C(u,v),u=0..1),v=0..1)) :
  Ev :=simplify(int(int(v*C(u,v),u=0..1),v=0..1)) :
```

Spearman

```
> simplify(factor(-3+12*E)) ;
```

$$3\, \frac{\theta\, a\, b}{(2+b)\,(2+a)}$$

$$\boxed{\rho = \frac{3ab\theta}{(a+2)(b+2)}}$$

Blest

```
> simplify(factor(-2+24*(E-Eu))) ;
  simplify(factor(-4+24*Eu)) ;
```

$$2\,\frac{\theta\,a\,b\,(5+a)}{(2+b)\,(2+a)\,(3+a)}$$

$$4\,\frac{\theta\,a\,b}{(2+b)\,(3+a)}$$

$$\boxed{\nu = \frac{2ab(a+5)\theta}{(a+2)(a+3)(b+2)}\,, \qquad \overline{\nu} = \frac{4ab\theta}{(a+3)(b+2)}}$$

Plantagenet

```
> simplify(factor(-2+12*(2*E-Eu-Ev))) ;
  simplify(factor(-4+12*(Eu+Ev))) ;
```

$$2\,\frac{\theta\,a\,b\,(15+4\,b+a\,b+4\,a)}{(2+b)\,(2+a)\,(3+a)\,(3+b)}$$

$$2\,\frac{\theta\,a\,b\,(12+5\,b+2\,a\,b+5\,a)}{(2+b)\,(2+a)\,(3+a)\,(3+b)}$$

$$\boxed{\xi = \frac{2ab(ab+4a+4b+15)\theta}{(a+2)(a+3)(b+2)(b+3)}\,, \qquad \overline{\xi} = \frac{2ab(2ab+5a+5b+12)\theta}{(a+2)(a+3)(b+2)(b+3)}}$$

3.3.2 Copule de Cuadras–Augé

La copule de Cuadras–Augé s'écrit sous la forme :

$$C(u,v) = \min\left(u^{1-a}v, uv^{1-a}\right)$$

avec $a \in [0,1]$ et $(u,v) \in [0,1]^2$. Elle est un cas particulier de la copule de Marshall & Olkin (1967) présentée dans l'exemple 8 du chapitre I. Grâce à MAPLE, on obtient successivement :

```
> assume(a>0,a<1) :
  E :=simplify(int(simplify(int(u^(1-a)*v,u=v..1)),v=0..1)
  +int(simplify(int(u*v^(1-a),u=0..v)),v=0..1)) :
  Eu :=simplify(int(simplify(int(u^(2-a)*v,u=v..1)),v=0..1)
  +int(int(u^2*v^(1-a),u=0..v),v=0..1)) :
  Ev :=simplify(int(simplify(int(u^(1-a)*v^2,u=v..1)),
  v=0..1)+int(int(u*v^(2-a),u=0..v),v=0..1)) :
```

Spearman

```
> simplify(factor(-3+12*E)) ;
```

$$-3\,\frac{a}{-4+a}$$

$$\boxed{\rho = \frac{-3a}{a-4}}$$

Blest

```
> simplify(factor(-2+24*(E-Eu))) ;
  simplify(factor(-4+24*Eu)) ;
```

$$-2\,\frac{a\,(-7+a)}{(-4+a)\,(-5+a)}$$

$$-4\,\frac{a}{-5+a}$$

$$\boxed{\nu = \frac{-2a(a-7)}{(a-4)(a-5)}\,, \qquad \overline{\nu} = \frac{-4a}{a-5}}$$

Plantagenet

```
> simplify(factor(-2+12*(2*E-Eu-Ev))) ;
  simplify(factor(-4+12*(Eu+Ev))) ;
```

$$-2\,\frac{a\,(-7+a)}{(-4+a)\,(-5+a)}$$

$$-4\,\frac{a}{-5+a}$$

$$\boxed{\xi = \frac{-2a(a-7)}{(a-4)(a-5)}\,, \qquad \overline{\xi} = \frac{-4a}{a-5}}$$

3.4 Variance asymptotique

Étant donné la longueur du calcul de la variance asymptotique, un seul exemple sera fait en détails.

La copule de Farlie–Gumbel–Morgenstern est l'un des modèles de dépendance les plus simples faisant intervenir l'indépendance (lorsque $\theta = 0$). Cette copule est de la forme :

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v)$$

avec $(u, v) \in [0, 1]^2$ et $\theta \in [-1, 1]$. Pour les autres copules, les calculs peuvent se faire de la même façon. Par contre, les expressions peuvent rapidement devenir complexes et longues à calculer.

```
> C :=(u,v)->u*v*(1+theta*(1-u)*(1-v)) ;
  c :=simplify(factor(diff(C(u,v),u,v))) ;
```

$$C := (u,\, v) \rightarrow u\, v\, (1 + \theta\, (1 - u)\, (1 - v))$$

$$c := 1 + \theta - 2\, \theta\, v - 2\, \theta\, u + 4\, u\, v\, \theta$$

Espérances conditionnelles

```
> Eusv :=simplify(int(u*c,u=0..1)) :
  Eu2sv :=simplify(factor(int(u^2*c,u=0..1))) :
  Evsu :=simplify(factor(int(v*c,v=0..1))) :
  Ev2su :=simplify(factor(int(v^2*c,v=0..1))) :
```

Spearman

```
> f :=simplify(u*v+int(Evsu,u=u..1)+int(Eusv,v=v..1)) :
  E :=simplify(int(int(f*c,u=0..1),v=0..1)) :
  E2 :=simplify(int(int(f^2*c,u=0..1),v=0..1)) :
  simplify(144*(E2-E^2)) ;
```

$$1 - \frac{11}{45}\,\theta^2$$

$$\boxed{\sigma_\rho^2 = 1 - \frac{11}{45}\,\theta^2}$$

Blest

```
> f :=simplify((1-u)^2*v+int(-2*(1-u)*Evsu,u=u..1)+
  int(1-2*Eusv+Eu2sv,v=v..1)) :
  E :=simplify(int(int(f*c,u=0..1),v=0..1)) :
  E2 :=simplify(int(int(f^2*c,u=0..1),v=0..1)) :
  simplify(144*(E2-E^2)) ;
```

$$-\frac{16}{63}\,\theta^2 + \frac{16}{15}$$

$$\boxed{\sigma_\nu^2 = \frac{16}{15} - \frac{16}{63}\,\theta^2}$$

```
> f :=simplify(u^2*v+int(2*u*Evsu,u=u..1)+
  int(Eu2sv,v=v..1)) :
  E :=simplify(int(int(f*c,u=0..1),v=0..1)) :
  E2 :=simplify(int(int(f^2*c,u=0..1),v=0..1)) :
  simplify(144*(E2-E^2)) ;
```

$$-\frac{16}{63}\,\theta^2+\frac{16}{15}$$

$$\boxed{\sigma_{\overline{\nu}}^2=\frac{16}{15}-\frac{16}{63}\,\theta^2}$$

Plantagenet

```
> f :=simplify(u*v*(4-u-v)+int(2*(2-u)*Evsu-Ev2su,u=u..1)+
  int(2*(2-v)*Eusv-Eu2sv,v=v..1)) :
  E :=simplify(int(int(f*c,u=0..1),v=0..1)) :
  E2 :=simplify(int(int(f^2*c,u=0..1),v=0..1)) :
  simplify(36*(E2-E^2)) ;
```

$$\frac{2}{225}\,\theta+\frac{1}{450}\,\theta^3-\frac{157}{630}\,\theta^2+\frac{31}{30}$$

$$\boxed{\sigma_{\xi}^2=\frac{1}{450}\,\theta^3-\frac{157}{630}\,\theta^2+\frac{2}{225}\,\theta+\frac{31}{30}}$$

```
> f :=simplify(u*v*(u+v)+int(2*u*Evsu+Ev2su,u=u..1)+
  int(2*v*Eusv+Eu2sv,v=v..1)) :
  E :=simplify(int(int(f*c,u=0..1),v=0..1)) :
  E2 :=simplify(int(int(f^2*c,u=0..1),v=0..1)) :
  simplify(36*(E2-E^2)) ;
```

$$\frac{2}{225}\,\theta + \frac{1}{450}\,\theta^3 - \frac{157}{630}\,\theta^2 + \frac{31}{30}$$

$$\boxed{\sigma^2_{\overline{\xi}} = \frac{1}{450}\,\theta^3 - \frac{157}{630}\,\theta^2 + \frac{2}{225}\,\theta + \frac{31}{30}}$$

Comme la copule de Farlie–Gumbel–Morgenstern est radialement symétrique, (R_i, S_i) et $(\overline{R}_i, \overline{S}_i)$ ont la même distribution sous cette loi, ce qui implique que ν_n et ξ_n ont la même distribution que $\overline{\nu}_n$ et $\overline{\xi}_n$ respectivement. Il n'est donc pas étonnant que $\sigma^2_\nu = \sigma^2_{\overline{\nu}}$ et que $\sigma^2_\xi = \sigma^2_{\overline{\xi}}$.

3.5 Efficacité relative asymptotique

Sous l'hypothèse d'indépendance, la loi limite de toutes les mesures de corrélation proposées est normale. On peut se servir de ce résultat pour bâtir un test d'indépendance. En notant σ^2_κ la variance du coefficient de corrélation κ, on rejettera l'hypothèse d'indépendance si $|\kappa_n| > z_{\frac{\alpha}{2}}\sigma_\kappa$.

Pour comparer des tests basés sur différentes statistiques, on peut utiliser l'efficacité asymptotique de Pitman. Dans ce contexte, on doit choisir une

contre-hypothèse qui ne dépend que d'un paramètre continu (disons θ) incluant l'indépendance comme cas particulier. En notant T_ρ et T_κ les tests basés sur ρ et κ,

$$ARE(T_\kappa, T_\rho) = \frac{n_\rho}{n_\kappa}\,,$$

la proportion des tailles d'échantillons nécessaires pour obtenir la même puissance avec les deux tests lorsque la contre-hypothèse tend vers l'indépendance. Si l'indépendance est atteinte pour $\theta = 0$, on trouve dans Lehmann (1998, p. 375) que

$$ARE(T_\kappa, T_\rho) = \frac{\sigma_\rho^2}{\sigma_\kappa^2}\left(\frac{\kappa_0'}{\rho_0'}\right)^2\,,$$

où $\rho_0' = \frac{\partial}{\partial\theta}\rho_\theta\big|_{\theta=0}$ et $\kappa_0' = \frac{\partial}{\partial\theta}\kappa_\theta\big|_{\theta=0}$.

Pour les calculs explicites, la partie cruciale consiste à évaluer κ_0' et ρ_0'. Deux approches sont possibles ; elles sont toutes les deux illustrées ci-après à l'aide de Maple.

1. Utiliser les formes explicites du paramètre estimé (tel que calculé à la section 3.3) et calculer directement les dérivées.
2. Faire passer la dérivée sous l'intégrale dans les formules de la section 3.4 pour obtenir, par exemple

$$\rho_0' = 12\int_0^1\int_0^1 \frac{\partial}{\partial\theta}C_\theta(u,v)\bigg|_{\theta=0}\,\mathrm{d}u\,\mathrm{d}v\,.$$

3.5.1 Illustration de la première méthode : copule de Cuadras–Augé

En consultant la section 3.3.2 (ou la copule de Cuadras–Augé est brièvement présentée), on trouve

$$\rho = \frac{-3}{a-4}\ , \qquad \nu = \xi = \frac{-2a(a-7)}{(a-4)(a-5)} \qquad \text{et} \qquad \overline{\nu} = \overline{\xi} = \frac{-4a}{a-5}\ .$$

De la section 3.1, on sait aussi que

$$\sigma_\rho^2 = 1\ , \qquad \sigma_\nu^2 = \sigma_{\overline{\nu}}^2 = \frac{16}{15} \qquad \text{et} \qquad \sigma_\xi^2 = \sigma_{\overline{\xi}}^2 = \frac{31}{30}\ .$$

On a donc en main toutes les informations nécessaires pour calculer l'efficacité relative asymptotique des coefficients les uns par rapport aux autres.

```
> rho :=-3*a/(a-4) :
  nu :=-2*a*(a-7)/(a-4)/(a-5) :
  nubar :=-4*a/(a-5) :
  xi :=nu : xibar :=nubar :
```

Blest

```
> 15/16*subs(a=0,(diff(nu,a)/diff(rho,a))^2) ;
  15/16*subs(a=0,(diff(nubar,a)/diff(rho,a))^2) ;
```

$$\frac{49}{60}$$
$$\frac{16}{15}$$

$$\boxed{ARE(T_\nu, T_\rho) = \frac{49}{60} \approx 0.8167\ , \qquad ARE(T_{\overline{\nu}}, T_\rho) = \frac{16}{15} \approx 1.067}$$

Plantagenet

```
> 30/31*subs(a=0,(diff(xi,a)/diff(rho,a))^2) ;
  30/31*subs(a=0,(diff(xibar,a)/diff(rho,a))^2) ;
```

$$\frac{392}{465}$$
$$\frac{512}{465}$$

$$\boxed{ARE(T_\xi, T_\rho) = \frac{392}{465} \approx 0.8430\ , \qquad ARE(T_{\overline{\xi}}, T_\rho) = \frac{512}{465} \approx 1.1011}$$

3.5.2 Illustration de la deuxième méthode : copule de Clayton

L'exemple 5 du chapitre I concerne l'efficacité relative asymptotique lorsque la copule de Clayton sert de contre-hypothèse. Voici explicitement les calculs qui mènent au résultat.

```
> C :=(u^(-a)+v^(-a)-1)^(-1/a) :
  Cp :=limit(diff(C,a),a=0) ;
```

$$Cp := u\,v\ln(u)\ln(v)$$

```
> E :=int(int(Cp,u=0..1),v=0..1) :
  Eu :=int(int(u*Cp,u=0..1),v=0..1) :
  Ev :=int(int(v*Cp,u=0..1),v=0..1) :
```

```
> rho :=12*E :
  nu :=24*(E-Eu) :
  nubar :=24*Eu :
  xi :=12*(2*E-Eu-Ev) :
  xibar :=12*(Eu+Ev) :
```

Blest

```
> 15/16*(nu/rho)^2 ;
  15/16*(nubar/rho)^2 ;
```

$$\frac{125}{108}$$
$$\frac{20}{27}$$

$$\boxed{ARE(T_\nu, T_\rho) = \frac{125}{108} \approx 1.1574\ , \qquad ARE(T_{\overline{\nu}}, T_\rho) = \frac{20}{27} \approx 0.7407}$$

Plantagenet

```
> 30/31*(xi/rho)^2 ;
  30/31*(xibar/rho)^2 ;
```

$$\frac{1000}{837}$$
$$\frac{640}{837}$$

$$\boxed{ARE(T_\xi, T_\rho) = \frac{1000}{837} \approx 1.1947\ , \qquad ARE(T_{\overline{\xi}}, T_\rho) = \frac{640}{837} \approx 0.7646}$$

Comme mentionné dans le chapitre I, la mauvaise performance d'un coefficient est toujours accompagnée par une bonne performance de son dual.

CHAPITRE IV

SIMULATIONS

Des simulations ont été réalisées afin de déterminer le seuil et la puissance des différents tests d'indépendance à taille finie. Ces résultats visent à compléter les comparaisons effectuées au moyen de l'efficacité relative asymptotique de Pitman.

Pour chaque loi considérée, 4 tailles d'échantillon ont été utilisées ($n = 25$, 50, 100 et 250). Les contre-hypothèses sont exprimées en terme du coefficient de corrélation de Pearson afin de rendre les graphiques comparables d'une loi à l'autre. Pour la construction des graphiques de puissance, les valeurs de r ont été discrétisées avec un pas de 0.01. Pour chacune de ces valeurs, 5000 tests bilatéraux de seuil 5 % ont été effectués à partir des versions expérimentales des six mesures de corrélation suivantes : ρ, τ, ν, $\overline{\nu}$, ξ et $\overline{\xi}$. Chaque statistique de test a été standardisée à l'aide de sa variance exacte sous l'indépendance et le seuil observé a été calculé au moyen de la loi asymptotique.

Lorsque les courbes de puissance sont rapprochées ou se superposent, il devient très difficile de déterminer à quels coefficients elles correspondent. Souvent, les courbes associées à ν_n et ξ_n sont pratiquement confondues ; il en va de même pour celles de $\overline{\nu}_n$ et $\overline{\xi}_n$. De fait, elles ne sont distinguables que

si $n = 25$, auquel cas ν_n et $\overline{\nu}_n$ sont légèrement plus puissants que ξ_n et $\overline{\xi}_n$. Ainsi, seuls ρ_n, τ_n, ξ_n et $\overline{\xi}_n$ ont été retenus pour les graphiques.

La bibliothèque publique Ranlib a servi à générer les nombres pseudo-aléatoires univariés. En ce qui concerne les algorithmes spécifiques aux lois bivariées, ils sont précisés dans leurs sections respectives.

4.1 Loi normale bivariée

La loi normale bivariée comporte cinq paramètres, à savoir les moyennes et les variances des deux variables, ainsi que le coefficient r caractérisant le degré de corrélation qu'il y a entre elles ($r = 0$ correspondant à l'indépendance). Puisque la copule sous-jacente ne dépend que de r, on a généré des observations

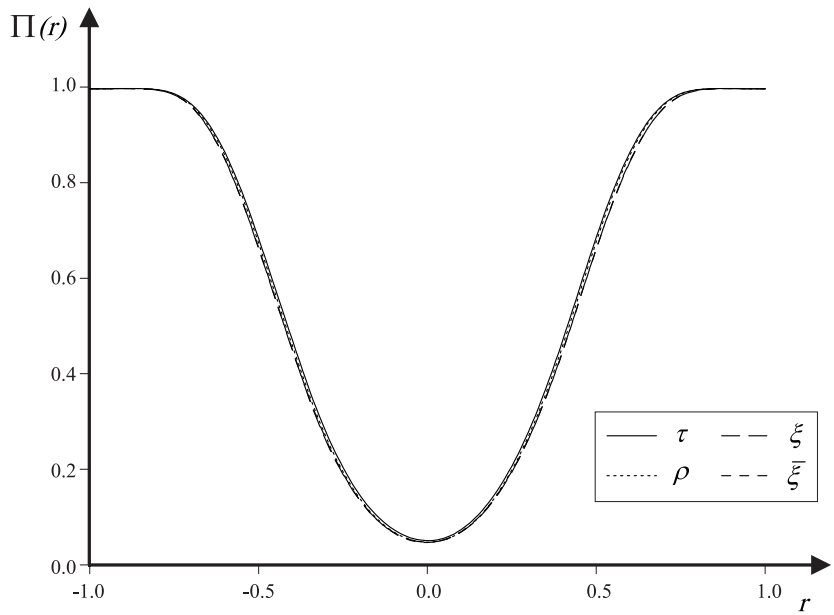
$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \overset{iid}{\sim} N_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \right\}$$

de densité

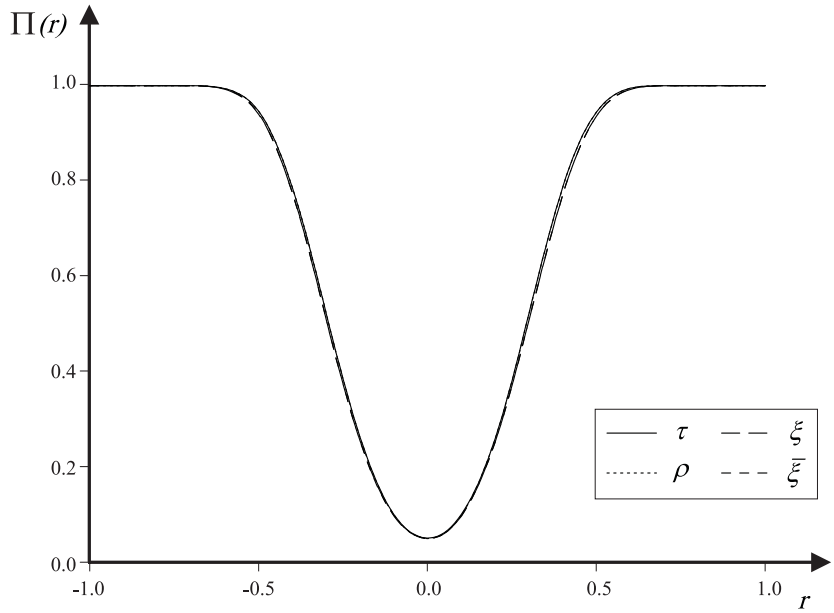
$$f(x,y) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(x^2 + y^2 - 2rxy\right) ,$$

où $r \in [-1, 1]$ et $(x, y) \in I\!R^2$. Une fonction de Ranlib a été utilisée pour générer ces observations.

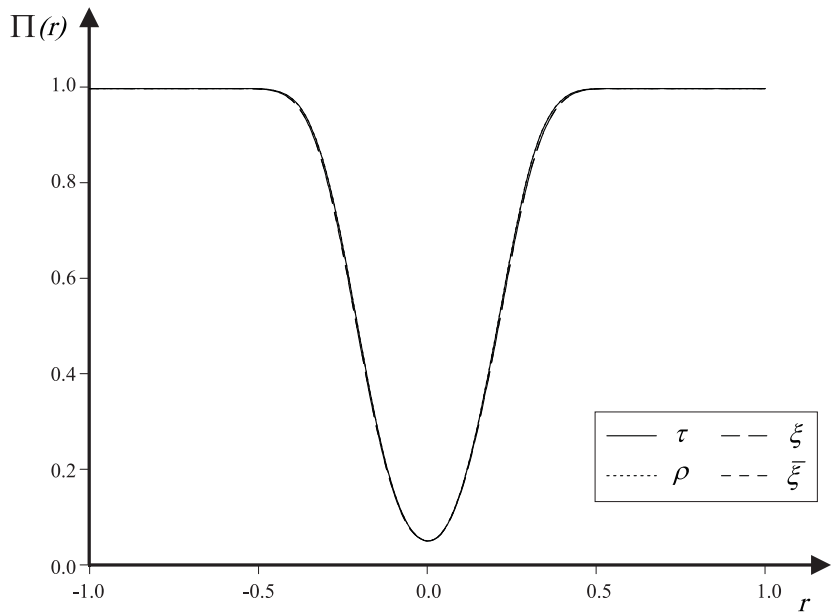
Tel que mentionné au chapitre I, ρ_n, ξ_n et $\overline{\xi}_n$ estiment tous la même valeur lorsque la loi est radialement symétrique. Le comportement très semblable de tous les coefficients observé dans ce cas n'est donc pas étonnant. La variance plus élevée des coefficients de Plantagenet explique qu'ils soient légèrement moins puissants (on voit apparaître les pointillés sous le trait plein).



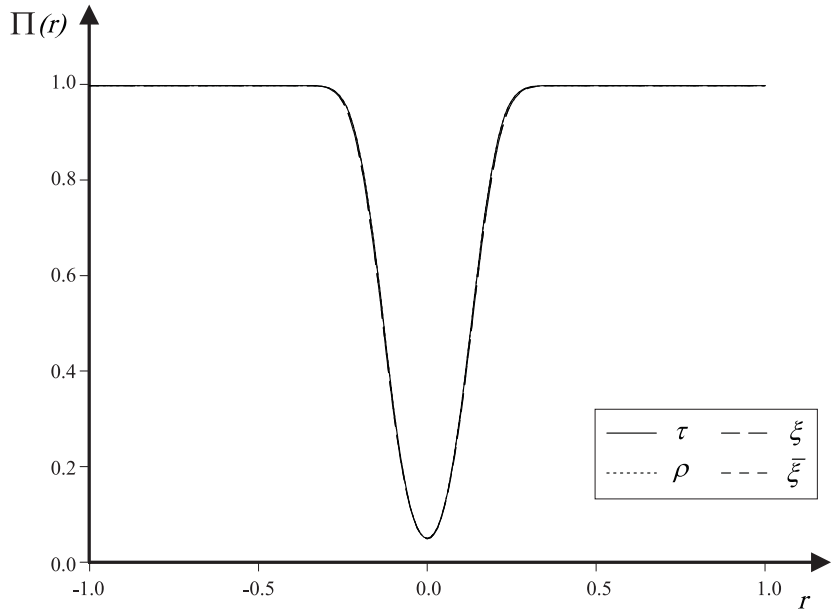
Loi normale, $n = 25$



Loi normale, $n = 50$



Loi normale, $n = 100$



Loi normale, $n = 250$

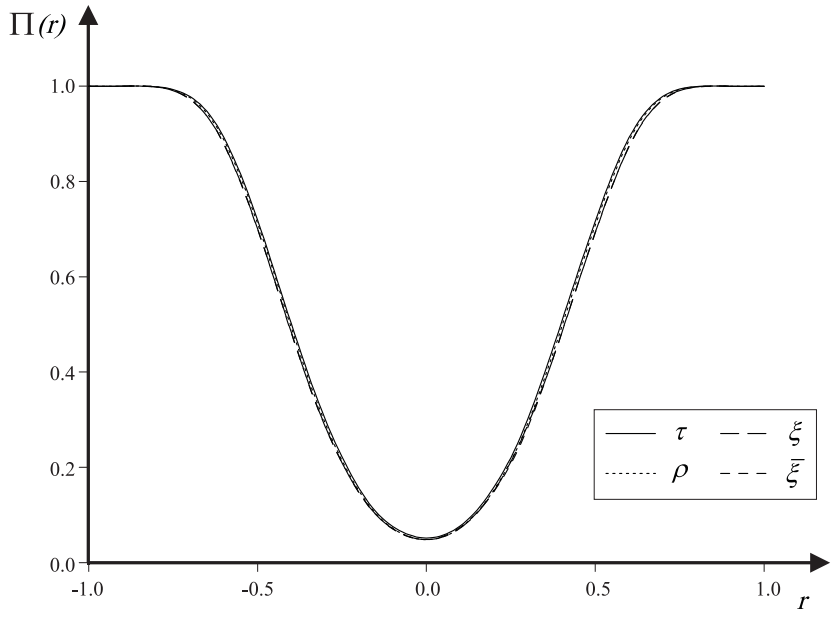
4.2 Copule de Frank

La copule de Frank (1979) s'écrit sous la forme

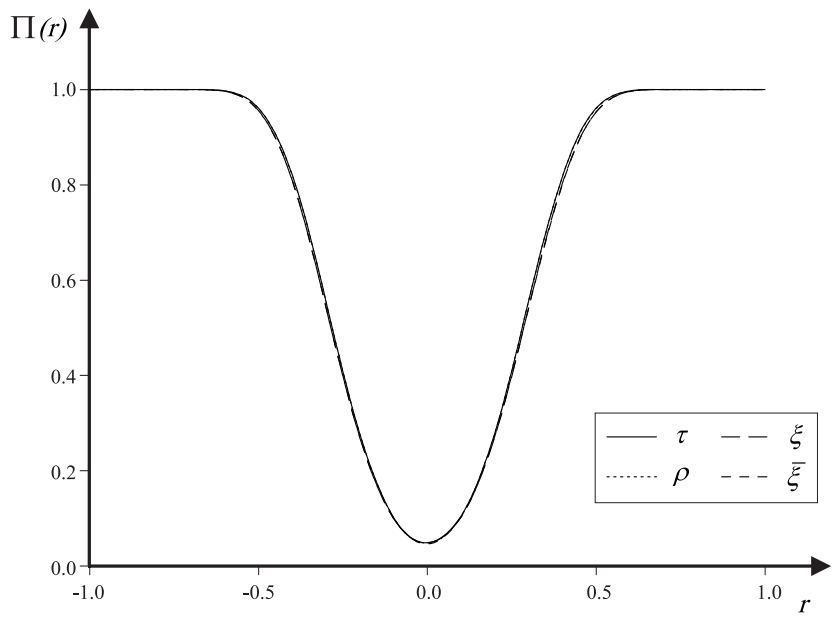
$$C_\alpha(u,v) = \log_\alpha \left\{ 1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha - 1} \right\} ,$$

où $\alpha > 0$ et $(u,v) \in [0,1]^2$. L'indépendance est obtenue lorsque $\alpha \to 1$. L'algorithme utilisé pour simuler des couples de telles variables est tiré de Genest & MacKay (1986).

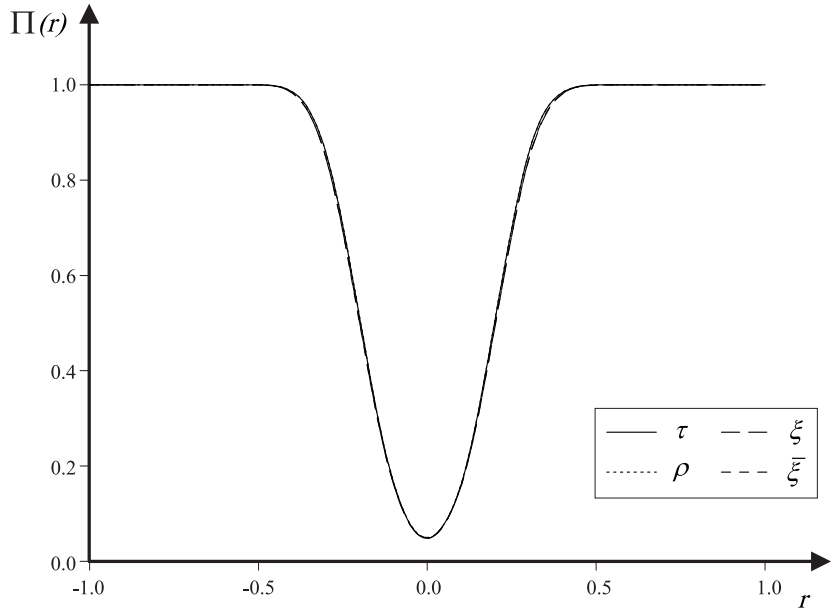
Comme les graphiques le montrent bien, le comportement des tests pour ce type de contre-hypothèse ressemble beaucoup à ce qui a été observé pour la loi normale. Ceci s'explique sans doute du fait que la copule de Frank est (la seule copule archimédienne) radialement symétrique.



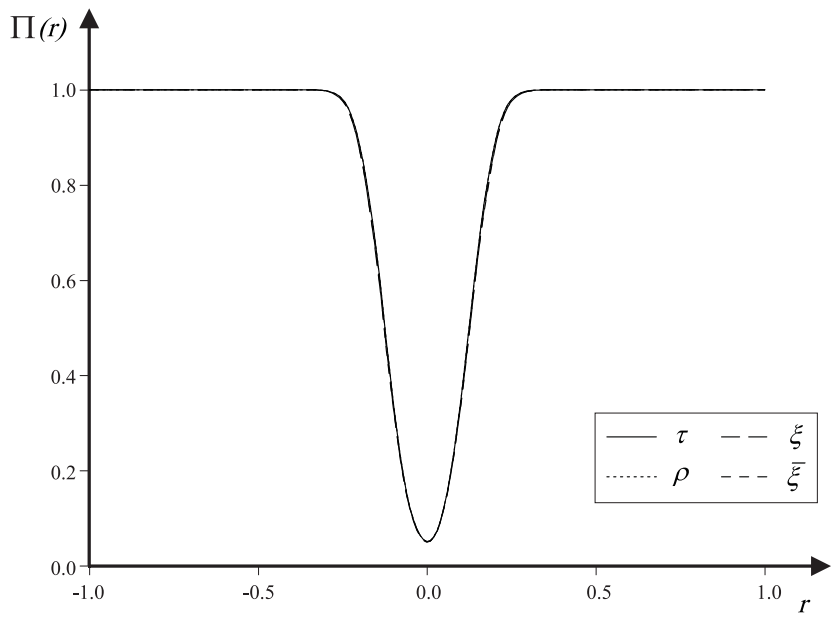
Copule de Frank, $n = 25$



Copule de Frank, $n = 50$



Copule de Frank, $n = 100$



Copule de Frank, $n = 250$

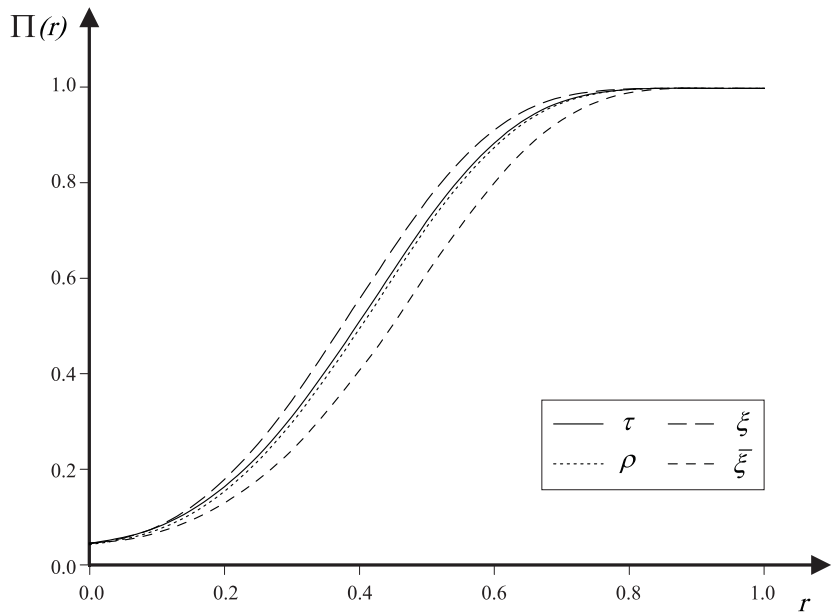
4.3 Copule de Clayton

La copule de Clayton fait l'objet de l'exemple 5 du chapitre I. Elle s'écrit sous la forme

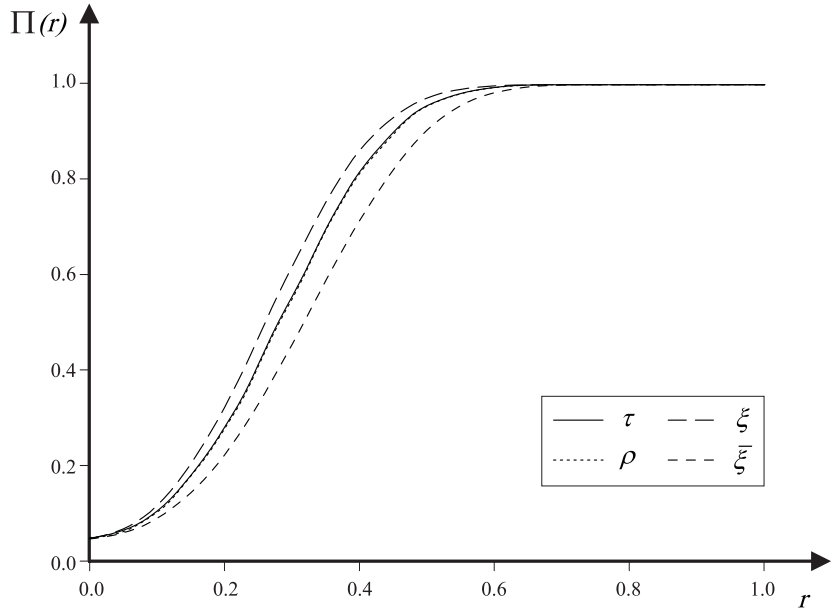
$$C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta} \ ,$$

où $\theta > 0$ et $(u,v) \in [0,1]^2$. L'indépendance est obtenue lorsque $\theta \to 0$. L'algorithme proposé par Marshall & Olkin (1988) a été utilisé pour générer les pseudo-aléas de cette loi.

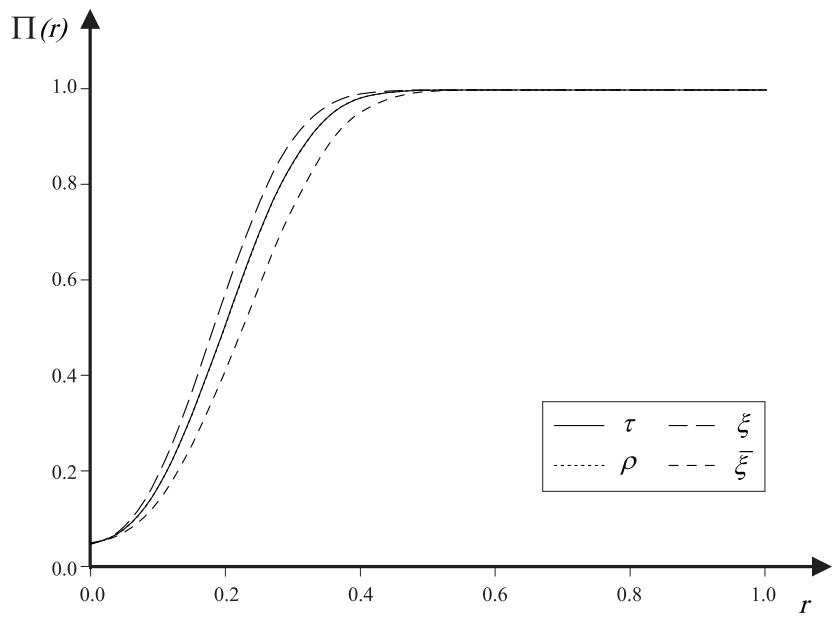
Les résultats de la simulation concordent avec l'efficacité relative asymptotique. En effet, le test basé sur ξ_n est plus puissant que celui basé sur ρ_n, alors que celui basé sur $\overline{\xi}_n$ est moins performant.



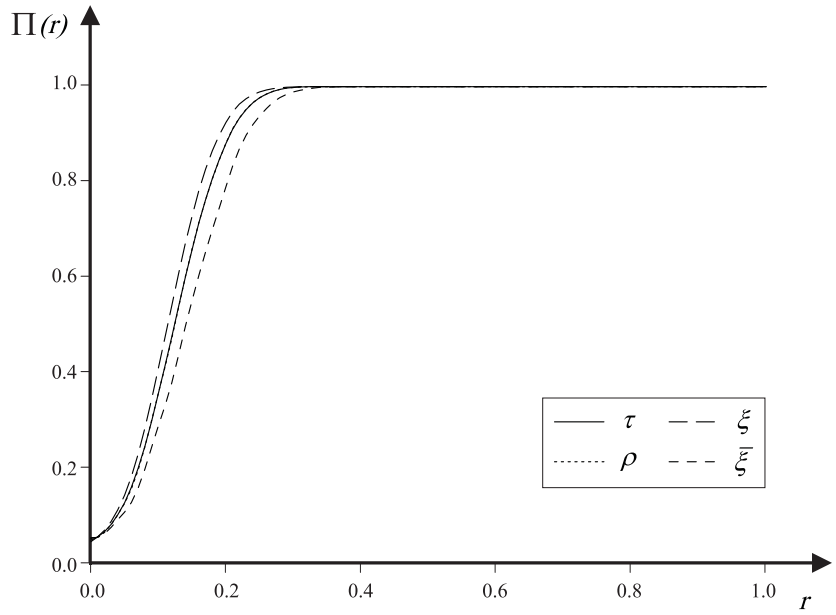
Copule de Clayton, $n = 25$



Copule de Clayton, $n = 50$



Copule de Clayton, $n = 100$



Copule de Clayton, $n = 250$

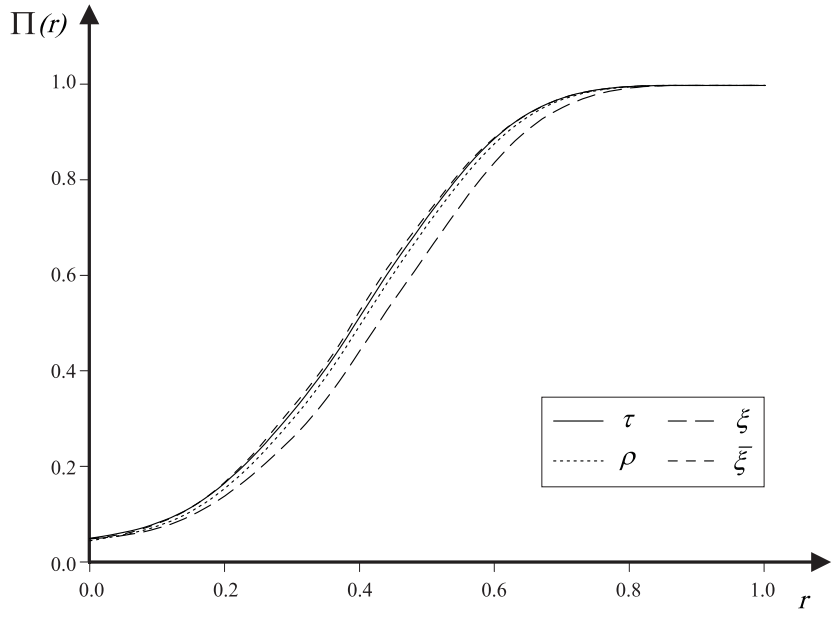
4.4 Copule de valeurs extrêmes de Gumbel

La copule de valeurs extrêmes de Gumbel s'écrit sous la forme

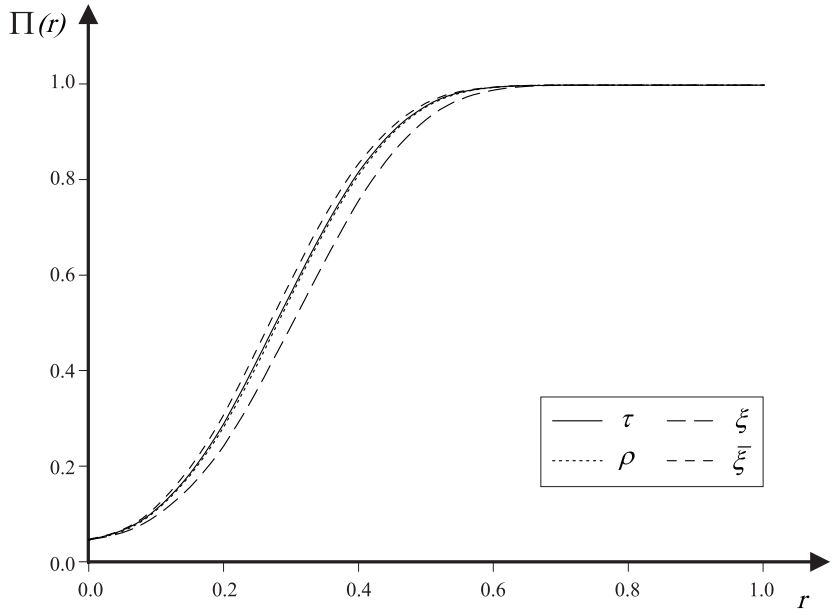
$$C_\theta(u,v) = \exp\left[-\left\{\left(\log\frac{1}{u}\right)^\theta + \left(\log\frac{1}{v}\right)^\theta\right\}^{1/\theta}\right],$$

où $\theta \geq 1$ et $(u,v) \in [0,1]^2$. L'indépendance correspond au cas $\theta = 1$. L'algorithme proposé par Genest & MacKay (1986) a été utilisé pour générer les couples de cette loi.

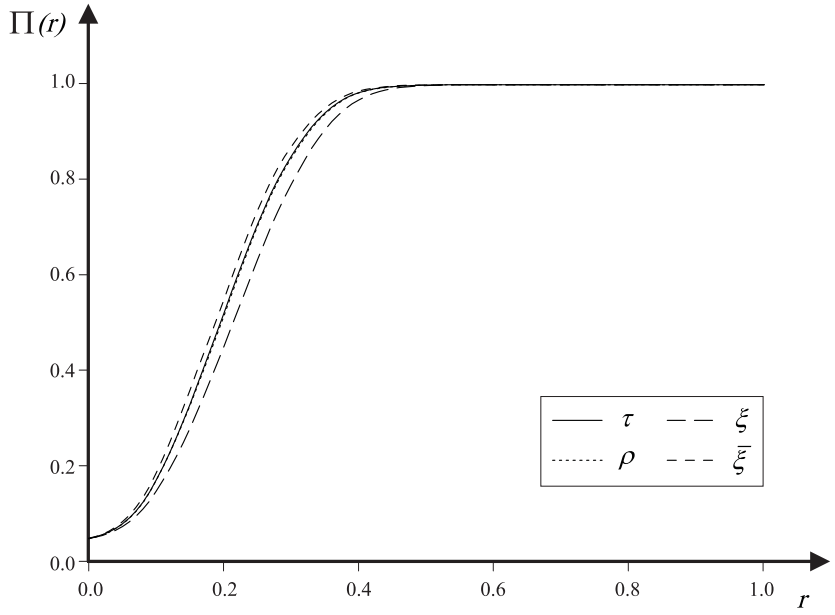
Dans ce cas-ci, le test basé sur $\overline{\xi}_n$ est le plus performant. De plus, le test fondé sur ξ_n se révèle moins puissant que celui qui s'appuie sur ρ_n.



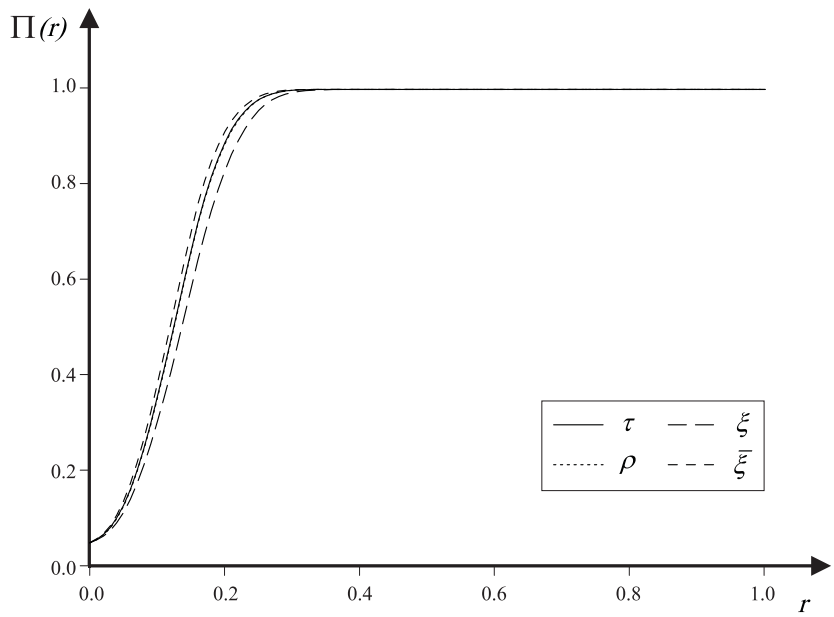
Copule de valeurs extrêmes de Gumbel, $n = 25$



Copule de valeurs extrêmes de Gumbel, $n = 50$



Copule de valeurs extrêmes de Gumbel, $n = 100$



Copule de valeurs extrêmes de Gumbel, $n = 250$

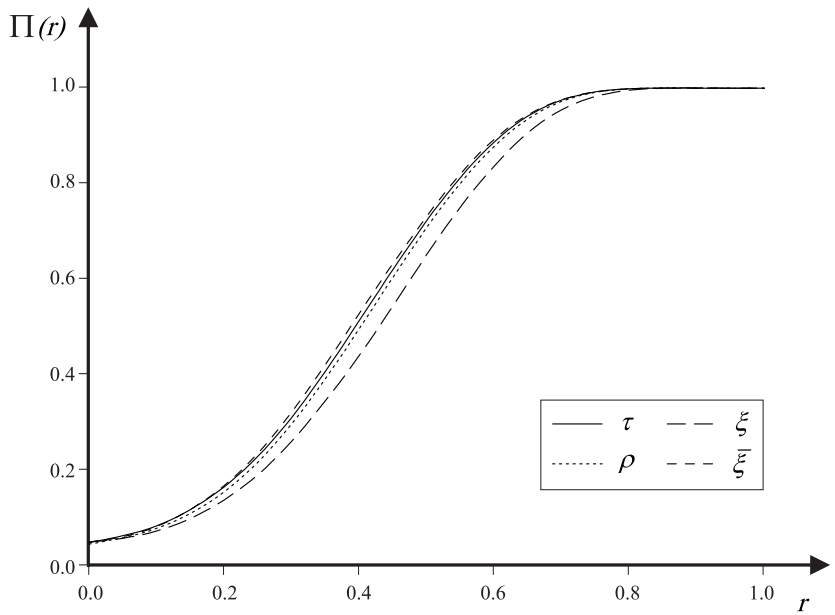
4.5 Copule de valeurs extrêmes de Galambos

La copule de Galambos est aussi de type extrême. Elle s'écrit

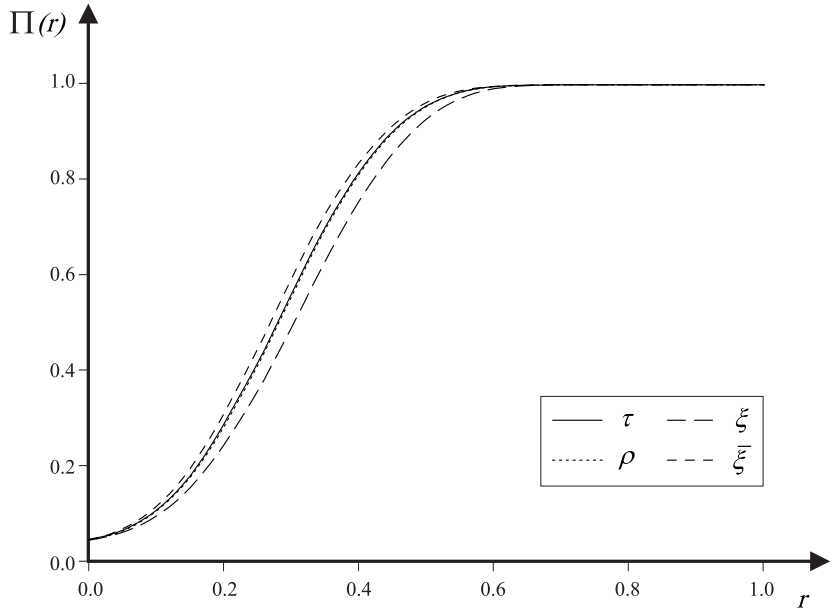
$$C_\theta(u,v) = uv \exp\left[\left\{\left(\log\frac{1}{u}\right)^{-\theta} + \left(\log\frac{1}{v}\right)^{-\theta}\right\}^{-1/\theta}\right],$$

où $\theta > 0$ et $(u,v) \in [0,1]^2$. On obtient l'indépendance quand $\theta \to 0$. L'algorithme présenté par Ghoudi, Khoudraji & Rivest (1998) a été utilisé pour la simulation de couples de cette loi.

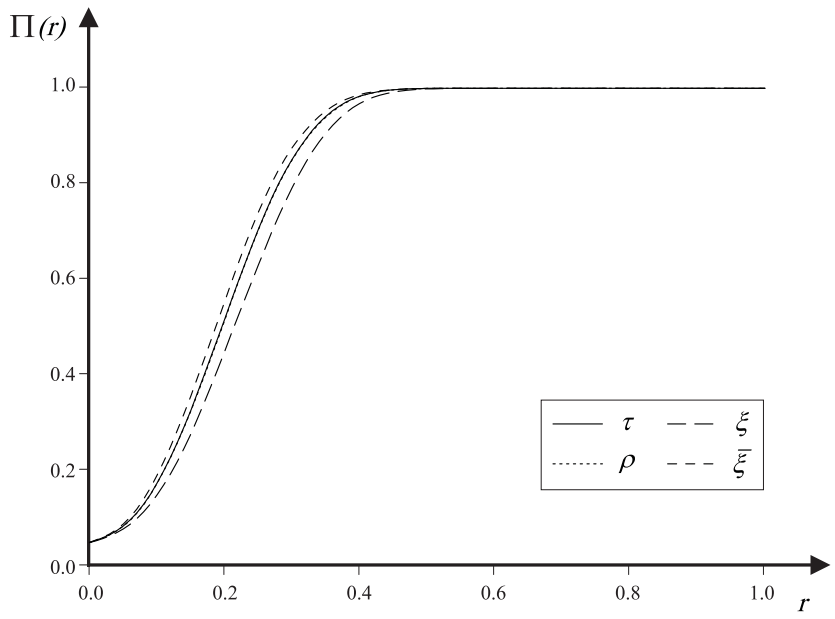
Les graphiques mettent en évidence la bonne performance du test basé sur $\overline{\xi}_n$. Comme dans le cas de la copule de Gumbel, le test basé sur ξ_n s'avère moins puissant.



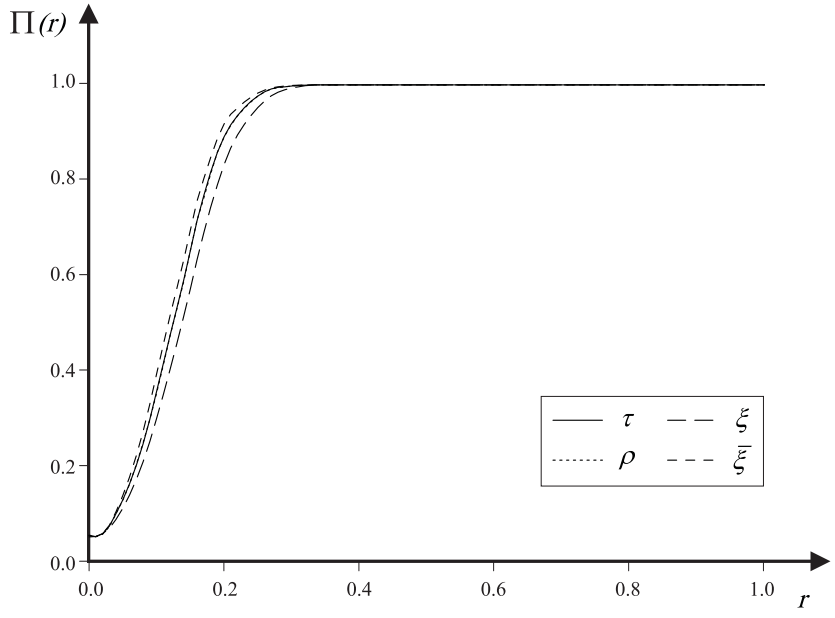
Copule de valeurs extrêmes de Galambos, $n = 25$



Copule de valeurs extrêmes de Galambos, $n = 50$



Copule de valeurs extrêmes de Galambos, $n = 100$



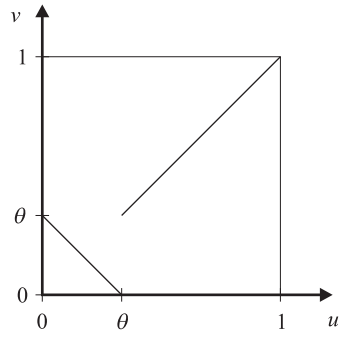
Copule de valeurs extrêmes de Galambos, $n = 250$

4.6 Transposition du minimum

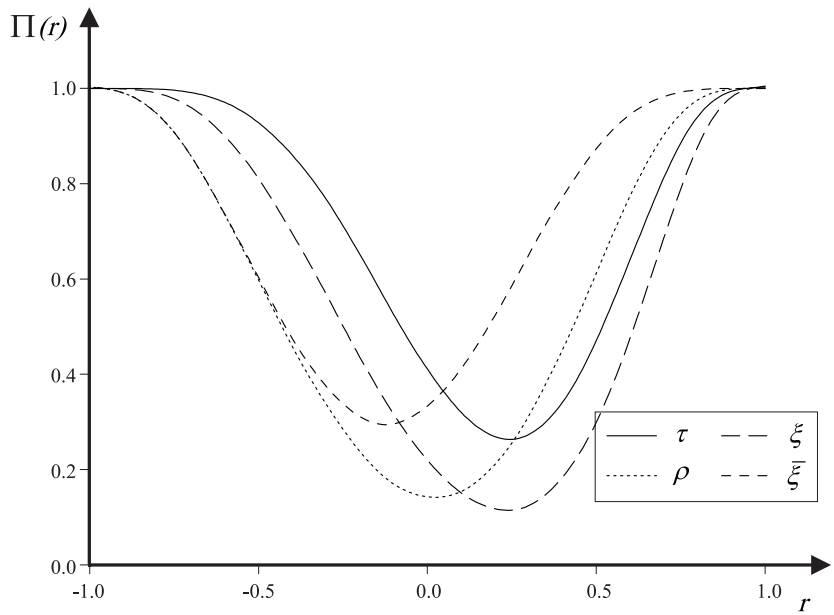
L'exemple 4 du chapitre I exploite deux copules de type transposition du minimum ("shuffle of min") pour démontrer que si l'on impose à un coefficient de corrélation de favoriser les premiers rangs, alors il ne saurait répondre à l'axiome e) de Scarsini. Même si ces deux copules peuvent difficilement modéliser quoi que ce soit en pratique et qu'elles n'incluent pas l'indépendance comme cas particulier, elles permettent de mieux comprendre le comportement des coefficients dans les tests d'indépendance.

4.6.1 Copule (6)

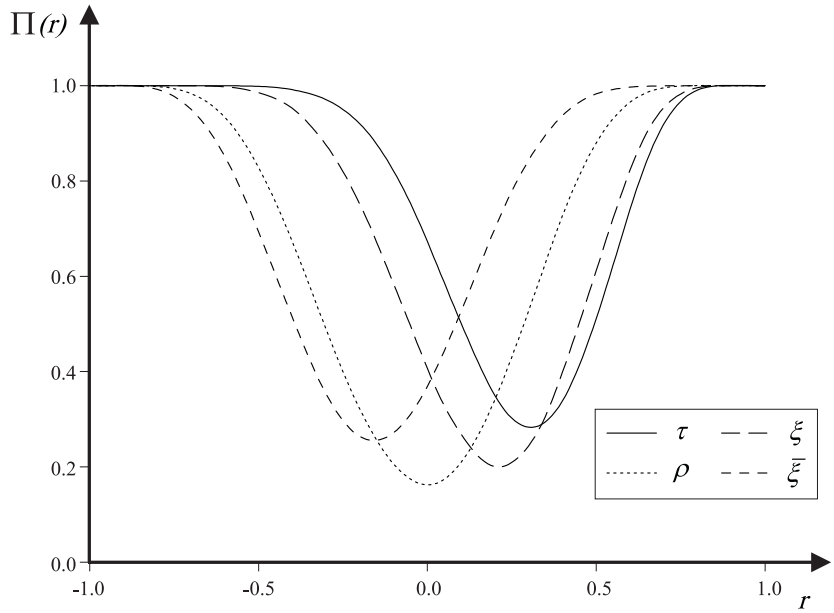
Rappelons que la copule (6) est celle dont le support est le suivant :



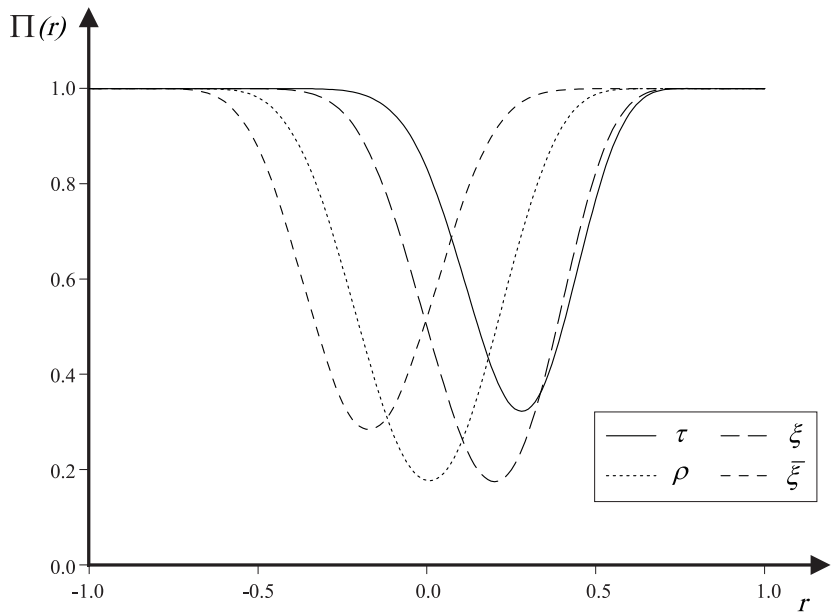
Étant donné que ξ punit plus sévèrement les différences observées dans les premiers rangs, on doit avoir $\xi < \rho$. Par conséquent, le test basé sur ξ_n rejette l'hypothèse nulle moins souvent que celui basé sur ρ ; il est donc moins puissant. Lorsque r devient très petit, la plupart des rangs se trouvent inversés et l'interprétation des résultats est moins évidente.



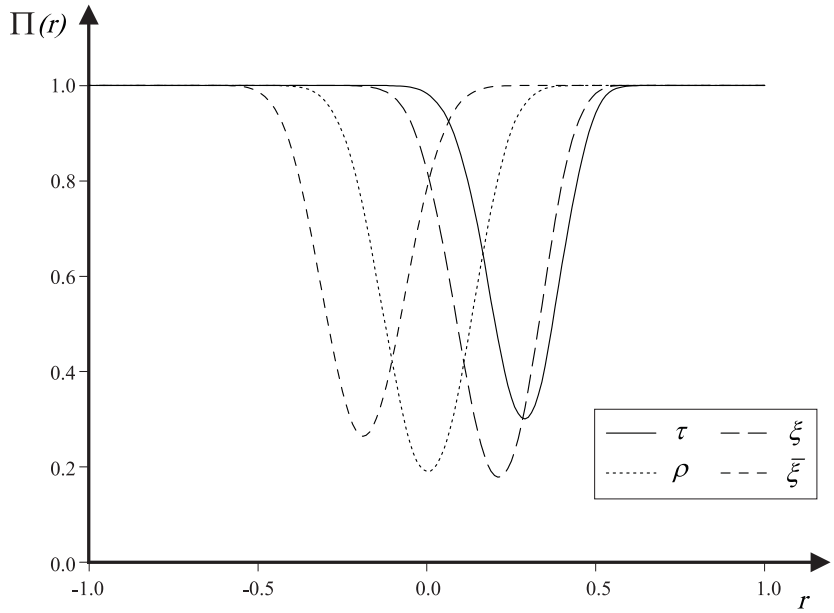
Copule (6), $n = 25$



Copule (6), $n = 50$



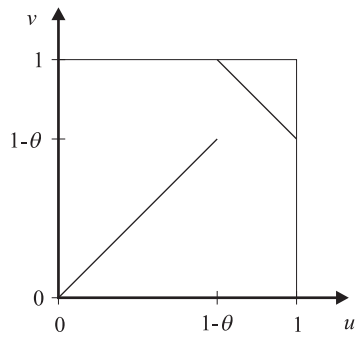
Copule (6), $n = 100$



Copule (6), $n = 250$

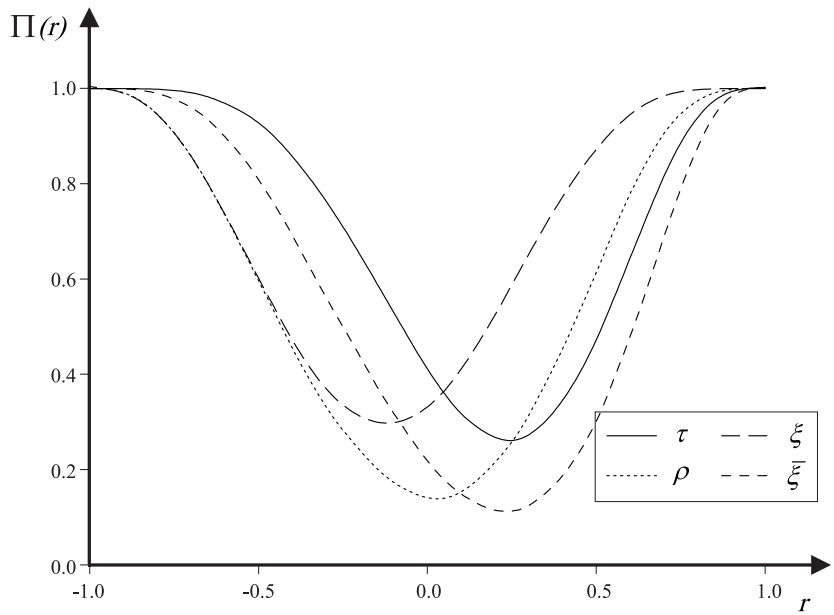
4.6.2 Copule (7)

Rappelons que la copule (7) est celle dont le support est le suivant :

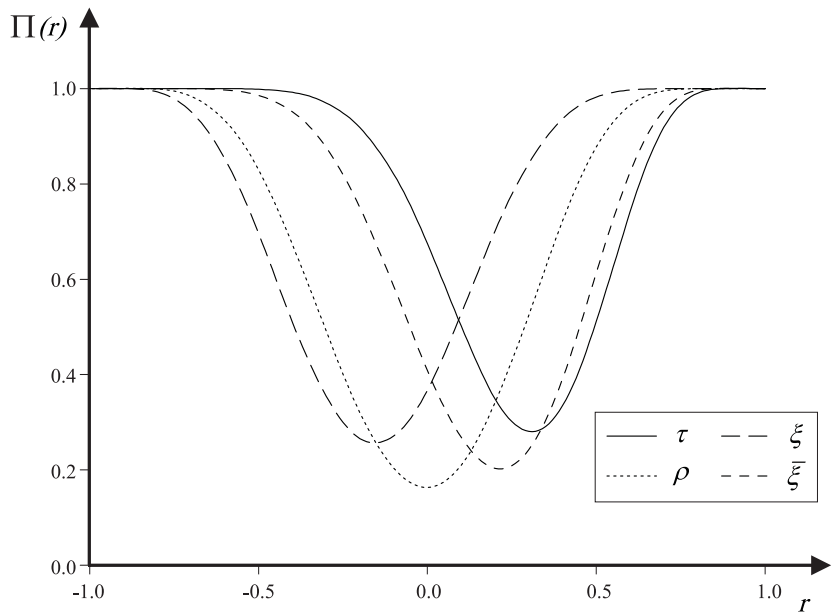


On peut constater que le test basé sur $\overline{\xi}_n$ adopte ici le même comportement que celui basé sur ξ_n dans l'exemple précédent, et ce, essentiellement pour les mêmes raisons.

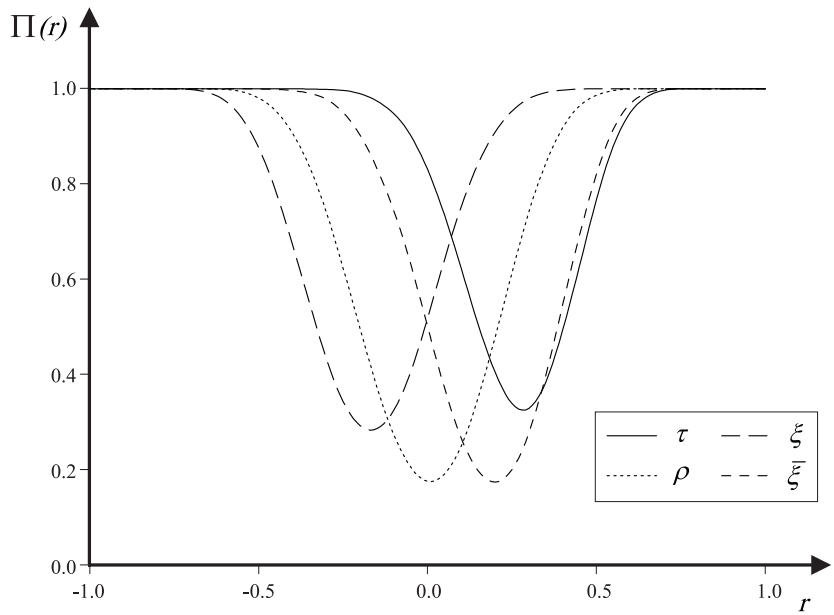
Autre remarque intéressante : de façon générale, le test basé sur τ_n se révèle plus puissant que celui fondé sur ρ_n pour les contre-hypothèses envisagées, bien que les deux tests soient asymptotiquement équivalents !



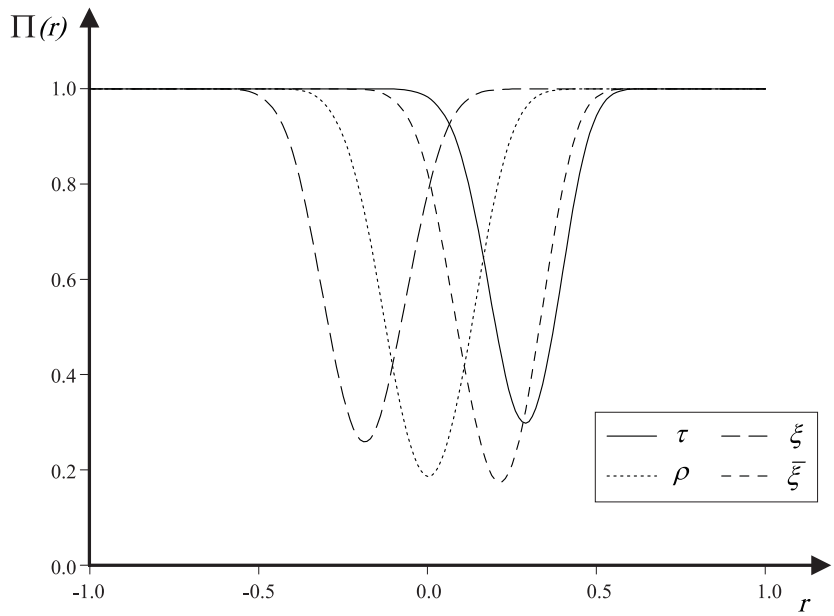
Copule (7), $n = 25$



Copule (7), $n = 50$



Copule (7), $n = 100$



Copule (7), $n = 250$

CONCLUSION

David C. Blest a proposé un coefficient de corrélation des rangs qui visait à affiner le ρ de Spearman. Ce mémoire complète son travail en précisant la loi asymptotique du coefficient de Blest et en en proposant des versions symétriques et duales.

Si la copule qui caractérise la dépendance d'un couple de variables aléatoires est radialement symétrique, le coefficient de Blest et ses variantes sont la plupart du temps des estimateurs (légèrement) inefficaces du ρ de Spearman. De façon plus générale, le lien entre ces coefficients duaux conduit à une borne minimale pour l'efficacité relative asymptotique de la meilleure variante du coefficient de Blest par rapport à ρ. Des simulations ont d'ailleurs permis de vérifier que les nouveaux coefficients pouvaient surpasser τ et ρ comme statistiques d'un test d'indépendance.

En formulant sa proposition, Blest a ouvert la porte à tout un ensemble de nouvelles mesures de corrélation basées sur des fonctions polynomiales des rangs. Une piste intéressante à explorer serait de vérifier quels polynômes débouchent sur des coefficients de corrélation acceptables et quels sont, géométriquement parlant, les avantages des uns par rapport aux autres. Une autre question en suspens est de savoir s'il existe des polynômes autres que uv qui permettent de construire un coefficient de corrélation respectant tous les axiomes de Scarsini. Il y a fort à parier que non !

BIBLIOGRAPHIE

M. M. Ali, N. N. Mikhail & M. S. Haq (1978). A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, 8, 405–412.

I. G. Bairamov, S. Kotz & M. Bekçi (2000). *New Generalized Farlie–Gumbel–Morgenstern Distributions and Concomitants of Order Statistics*. Technical Report, George Washington University, Washington, DC.

D. C. Blest (2000). Rank correlation — An alternative measure. *Australian and New Zealand Journal of Statistics*, 42, 101—111.

D. G. Clayton (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141–151.

C. M. Cuadras & J. Augé (1981). A continuous general multivariate distribution and its properties. *Communications in Statistics*, *Theory and Methods*, 10, 339–353.

D. Drouet–Mari & S. Kotz (2001). *Correlation and Dependence*. Imperial College Press, London.

T. S. Ferguson, C. Genest & M. Hallin (2000). Kendall's tau for serial dependence. *La revue canadienne de statistique*, 28, 587–604.

M. J. Frank (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, 19, 194–226.

J. Galambos (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70, 674–680.

C. Genest (1987). Frank's family of bivariate distributions. *Biometrika*, 74, 549–555.

C. Genest, K. Ghoudi & L.-P. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543–552.

C. Genest, K. Ghoudi & L.-P. Rivest (1998). Commentaire sur un article de E. W. Frees et de E. A. Valdez intitulé "Understanding relationships using copulas." *North American Actuarial Journal*, 2, 143–149.

C. Genest & R. J. MacKay (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *La revue canadienne de statistique*, 14, 145–159.

K. Ghoudi, A. Khoudraji & L.-P. Rivest (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *La revue canadienne de statistique*, 26, 187–197.

E. J. Gumbel (1960a). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55, 698–707.

E. J. Gumbel (1960b). Distributions des valeurs extrêmes en plusieurs dimensions. *Bulletin de l'Institut de statistique de l'Université de Paris*, 9, 171–173.

J. Hájek (1969). *A Course in Nonparametric Statistics*. Holden-Day, San Francisco, CA.

M. Hallin & M. L. Puri (1992). Rank test for time series analysis : A survey. Dans *New Directions in Time Series Analysis*, publié sous la direction de D. R. Brillinger, E. Parzen et M. Rosenblatt, pp. 112–153. Springer, New York.

H. Joe (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

E. L. Lehmann (1998). *Nonparametrics : Statistical Methods Based on Ranks, Première édition révisée*. Upper Prentice Hall, Saddle River, NJ.

A. W. Marshall & I. Olkin (1967). A generalized bivariate exponential distribution. *Journal of Applied Probability*, 4, 291–302.

A. W. Marshall & I. Olkin (1988). Families of multivariate distributions. *Journal of the American Statistical Society*, 83, 834–841.

P. Mikusiński, H. Sherwood & M. D. Taylor (1992). Shuffles of min. *Stochastica*, 13, 61–74.

R. B. Nelsen (1986). Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics*, *Theory and Methods*, 15, 3277–3285.

R. B. Nelsen (1999). *An Introduction to Copulas*. Lecture Notes in Statistics No 139. Springer, Berlin.

D. Oakes (2001). *Biometrika* centenary : Survival analysis. *Biometrika*, 88, 99–142.

R. L. Plackett (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60, 516–522.

J. J. Quesada–Molina (1992). A generalization of an identity of Hoeffding and some applications. *Journal of the Italian Statistical Society*, 1, 405–411.

F. H. Ruymgaart, G. R. Shorack & W. R. van Zwet (1972). Asymptotic normality of nonparametric tests for independence. *The Annals of Mathematical Statistics*, 43, 1122–1135.

M. Scarsini (1984). On measures on concordance. *Stochastica*, 8, 201–208.

B. Schweizer & E. F. Wolff (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9, 879–885.