REVUE DE STATISTIQUE APPLIQUÉE

P. CAPÉRAÀ
A. I. GARRALDA-GUILLEM

Estimation de la variance asymptotique de statistiques linéaires de rang et applications

Revue de statistique appliquée, tome 48, nº 3 (2000), p. 75-93
<http://www.numdam.org/item?id=RSA_2000__48_3_75_0>

Rev. Statistique Appliquée, 2000, **XLVIII** (3), 75-93

ESTIMATION DE LA VARIANCE ASYMPTOTIQUE DE STATISTIQUES LINÉAIRES DE RANG ET APPLICATIONS

P. Capéraà*, A. I. Garralda-Guillem**

* *Département de mathématiques et de statistique*
Université Laval, Sainte-Foy (Québec) Canada G1K 7P4
email : caperaa@mat.ulaval.ca
** *Departamento de Matemática Aplicada*
Facultad de Ciencias, Universidad de Granada, Granada, España
email : agarral@goliat.ugr.es

RÉSUMÉ

Cet article concerne l'estimation de la variance asymptotique de statistiques linéaires de rang utilisées pour tester l'indépendance dans une loi bidimensionnelle. L'estimateur que nous proposons est fortement convergent. Au moyen de simulations de Monte-Carlo, nous montrons qu'il est un bon estimateur de la variance de statistiques linéaires de rang pour des tailles d'échantillons finies. Nous l'utilisons alors pour construire des tests afin de comparer le degré de dépendance de deux lois bidimensionnelles.

Mots-clés : *Lois bidimensionnelles, statistiques linéaires de rang, tests de rang, indépendance, variance asymptotique.*

ABSTRACT

This paper proposes an estimator of the asymptotic variance of linear rank statistics for testing the hypotheses of independence. This estimator is proved to be strongly consistent. Through simulations, it is also seen to perform reasonably well as an estimator of the variance of linear rank statistics for finite samples. It can thus be used to construct tests for comparing the degree of dependence between two bivariate distributions.

Keywords : *Bivariate distributions, linear rank statistics, rank tests, independence, asymptotic variance.*

1. Introduction

Les statistiques de rang sont souvent utilisées pour tester l'hypothèse d'indépendance dans une loi bidimensionnelle. En particulier, les statistiques très classiques de Spearman et de Kendall remplacent avantageusement le coefficient de corrélation échantillonnal de Pearson lorsque la loi des observations n'est pas une loi normale (Gokhale, 1968 et Behnen, 1971). Cependant, lorsqu'il s'agit de tester le degré d'association entre deux variables aléatoires ou de comparer deux lois bidimensionnelles selon leur force de dépendance, l'utilisation des statistiques de rang est beaucoup plus rare. C'est le cas, par exemple, du dernier problème cité pour lequel seul semble être utilisé le test construit à partir de la transformation z de Fisher du coefficient de corrélation de Pearson. Cependant, ce test est peu robuste relativement à l'absence de normalité bidimensionnelle comme le montrent les résultats de simulations présentés à la Section 4 de ce papier. Aussi est-il important de proposer d'autres tests pour de tels problèmes, car on a de plus en plus recours à des modèles bidimensionnels qui sont assez éloignés du modèle normal. Pour s'en persuader, il suffit de penser au développement récent des modèles de lois des valeurs extrêmes multivariées (voir Resnick, 1987, Coles et Tawn, 1991 et de Haan et de Ronde, 1998), et à l'utilisation de plus en plus fréquente des copules dans les domaines où l'étude de la structure de dépendance est le principal objectif (voir par exemple Bandeen-Roche et Liang, 1996). Les statistiques linéaires de rang devraient permettre la construction de techniques robustes pour étudier la dépendance dans ce type de modèles. À cette fin, il est nécessaire d'estimer la variance asymptotique de ces statistiques qui, sous certaines conditions de régularité, convergent en loi vers une loi normale (Ruymgaart, Shorack et van Zwet, 1972).

Dans cet article, nous proposons un estimateur de la variance asymptotique de statistiques linéaires de rang dont nous montrons, dans la Section 2, qu'il est fortement convergent. Afin d'évaluer sa validité et sa précision, il sera comparé, dans la Section 3, au moyen de simulations de Monte-Carlo, à l'estimateur jackknife correspondant dont on connaît les qualités dans de nombreuses situations. La Section 4 est consacrée à la comparaison de trois tests (Pearson, Kendall et Spearman) qui permettent de comparer la force de dépendance de deux lois bidimensionnelles. Enfin dans deux annexes on trouve, d'une part un schéma de la démonstration de la convergence presque sûre de l'estimateur proposé, et d'autre part une description des lois bidimensionnelles utilisées dans les simulations.

2. L'estimateur proposé

Soit $(X_1, Y_1), \ldots, (X_n, Y_n)$ un échantillon aléatoire provenant d'une loi continue bidimensionnelle H de marginales F et G. Dénotons respectivement par H_n, F_n et G_n la fonction de répartition échantillonnale de H, F et G. Soient R_i et S_i, $i = 1, \ldots, n$, les rangs de X_i et Y_i, respectivement, et considérons les statistiques linéaires de rang suivantes :

$$T_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{R_i}{n+1}\right) K\left(\frac{S_i}{n+1}\right) \tag{1}$$

où J et K sont deux fonctions score définies sur l'intervalle $(0, 1)$. Ruymgaart, Shorack et van Zwet (1972) ont établi sous certaines conditions de régularité la normalité asymptotique de $\sqrt{n}\,(T_n - \mu)/\sigma$, où

$$\mu = \iint_{\mathbb{R}^2} J\{F(x)\}K\{G(y)\}dH(x,y)$$

et

$$\begin{aligned}\sigma^2 = \operatorname{Var}\Big[& J\{F(X)\}K\{G(Y)\} \\ & + \iint_{\mathbb{R}^2} \{\phi_X(x) - F(x)\}J'\{F(x)\}K\{G(y)\}dH(x,y) \\ + & \iint_{\mathbb{R}^2} \{\phi_Y(y) - G(y)\}J\{F(x)\}K'\{G(y)\}dH(x,y)\Big] = \operatorname{Var}\{U(X,Y)\}\,, \quad (2)\end{aligned}$$

avec

$$\phi_X(x) = 0 \quad \text{si} \quad x < X \quad \text{et} \quad \phi_X(x) = 1 \quad \text{si} \quad x \geqslant X\,.$$

De leur côté, Genest, Ghoudi et Rivest (1995) ont montré que, sous certaines conditions de régularité, T_n converge presque sûrement vers μ.

Un estimateur simple et naturel de la variance asymptotique σ^2 serait la variance échantillonnale des variables $U(X_i, Y_i)$, $i = 1, \ldots, n$, si celles-ci étaient observables. Cependant, comme ce n'est pas le cas, nous considérons les pseudo-observations $U_n(X_i, Y_i)$, $i = 1, \ldots, n$, définies par

$$\begin{aligned} U_n(X_i,Y_i) = & J\{F_n^*(X_i)\}K\{G_n^*(Y_i)\} \\ & + \iint_{\mathbb{R}^2} \{\phi_{X_i}(x) - F_n^*(x)\}J'\{F_n^*(x)\}K\{G_n^*(y)\}dH_n(x,y) \\ & + \iint_{\mathbb{R}^2} \{\phi_{Y_i}(y) - G_n^*(y)\}J\{F_n^*(x)\}K'\{G_n^*(y)\}dH_n(x,y)\,, \end{aligned}$$

$$\text{où} \quad F_n^*(x) = \frac{n}{n+1}F_n(x) \quad \text{et} \quad G_n^*(y) = \frac{n}{n+1}G_n(y)\,.$$

Ainsi, nous proposons d'estimer la variance asymptotique σ^2 par la variance échantillonnale des pseudo-observations $U_n(X_i, Y_i)$, soit

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \left\{U_n(X_i,Y_i) - \overline{U}_n\right\}^2$$

$$\text{où} \qquad \overline{U}_n = \frac{1}{n}\sum_{i=1}^{n} U_n(X_i,Y_i)\,.$$

On peut noter en passant que cet estimateur est une fonction des rangs des observations X_i et Y_i seulement, et donc sa loi est indépendante des lois marginales F et G. Plus précisément, elle ne dépend que de la copule associée à la loi bidimensionnelle H, notion définie à l'Annexe 2.

La convergence presque sûre de cet estimateur est donnée dans le théorème suivant sous un ensemble de conditions de régularité C spécifié dans l'Annexe 1.

Théorème. – *Sous les conditions de régularité C, l'estimateur $\hat{\sigma}^2$ converge presque sûrement vers la variance asymptotique σ^2.*

La démonstration de ce théorème étant longue et technique, nous n'en donnons qu'un schéma dans l'Annexe 1. Le lecteur peut en trouver les détails dans la thèse de doctorat de A. I. Garralda-Guillem (1997), Universidad de Granada.

3. Comparaison de l'estimateur proposé avec l'estimateur jackknife

Dans cette section, nous rapportons les résultats de simulations de Monte-Carlo exécutées afin d'examiner la performance de

$$\hat{\sigma}(T_n) = \left[\frac{1}{n(n-1)} \sum_{i=1}^{n} \{U_n(X_i, Y_i) - \overline{U}_n\}^2\right]^{1/2}$$

comme estimateur de l'écart-type $\sigma(T_n)$ de statistiques linéaires de rang T_n. Plus précisément, nous considérons les deux statistiques suivantes, ayant respectivement une fonction score bornée et non bornée :

– le coefficient de corrélation de Spearman

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

– la statistique de van der Waerden

$$W_n = \frac{1}{n} \sum_{i=1}^{n} \Phi^{-1}\left(\frac{R_i}{n+1}\right) \Phi^{-1}\left(\frac{S_i}{n+1}\right)$$

où Φ^{-1} est la fonction quantile de la loi normale $N(0,1)$.

Afin de mieux cerner le comportement de $\hat{\sigma}(T_n)$, nous allons le comparer à celui de l'estimateur obtenu par la méthode du jackknife (voir par exemple Wolter, 1985) dont la validité a été montrée dans de nombreuses situations. Cet estimateur dénoté $\hat{\sigma}_J(T_n)$ est défini par

$$\hat{\sigma}_J(T_n) = \left\{\frac{n-1}{n} \sum_{i=1}^{n} \left(T_n^{(i)} - T_n^{(\cdot)}\right)^2\right\}^{1/2}$$

où $T_n^{(i)}, i = 1, \ldots, n$, dénote la statistique T_n calculée à partir des $n-1$ observations $(X_1, Y_1), \ldots, (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), \ldots, (X_n, Y_n)$, et $T_n^{(\cdot)}$ est la moyenne des n statistiques $T_n^{(i)}$.

Les résultats des simulations sont obtenus à partir de 5 000 échantillons pseudo-aléatoires provenant des distributions décrites dans l'Annexe 2 avec le coefficient $r = 0.25$ et 0.50 pour le premier groupe, et avec le coefficient de concordance de Kendall $\tau = 0.25$ et 0.50 pour le second groupe.

On trouve dans le Tableau 1 ci-dessous les moyennes $\bar{\hat{\sigma}}(T_n)$ et $\bar{\hat{\sigma}}_J(T_n)$ des estimations $\hat{\sigma}(T_n)$ et $\hat{\sigma}_J(T_n)$ obtenues à partir des 5 000 échantillons dans le cas où les variables X et Y sont indépendantes, pour des tailles d'échantillon de 25 et 50. Comme dans le cas de l'indépendance, l'écart-type de T_n a une expression assez simple (voir par exemple Hájek et Šidák, 1967, chap. 3), nous calculons sa valeur qui va servir, par comparaison, à évaluer les deux estimateurs. Finalement, on trouve entre parenthèses l'estimation de l'écart-type de $\hat{\sigma}(T_n)$ et de $\hat{\sigma}_J(T_n)$, respectivement.

TABLEAU 1

Valeurs de l'écart-type $\sigma(T_n)$ de T_n, des moyennes $\bar{\hat{\sigma}}(T_n)$ et $\bar{\hat{\sigma}}_J(T_n)$ des estimateurs $\hat{\sigma}(T_n)$ et $\hat{\sigma}_J(T_n)$, et des estimations multipliées par 10^2 de l'écart-type de ces estimateurs, à partir de 5 000 échantillons de taille $n = 25$, 50 issus d'une loi normale avec $r = 0$.

T_n		$\sigma(T_n)$	$\bar{\hat{\sigma}}(T_n)$	$\bar{\hat{\sigma}}_J(T_n)$
Spearman	$n = 25$	0.2041	0.2081 (2.128)	0.2151 (2.128)
	$n = 50$	0.1428	0.1441 (0.944)	0.1469 (0.951)
van der Waerden	$n = 25$	0.1609	0.1656 (2.530)	0.1686 (2.646)
	$n = 50$	0.1243	0.1243 (1.367)	0.1279 (1.510)

Les résultats apparaissant dans le Tableau 1 indiquent que, pour les deux tailles d'échantillon et les deux statistiques T_n, la moyenne des estimations $\hat{\sigma}(T_n)$ est légèrement plus proche de la valeur exacte $\sigma(T_n)$ que la moyenne des estimations obtenues par la méthode du jackknife. De plus, on constate que l'estimation de l'écart-type de $\hat{\sigma}(T_n)$ est, dans tous les cas, inférieure à celle de $\hat{\sigma}_J(T_n)$. Ceci tend à montrer que $\hat{\sigma}(T_n)$ est un estimateur moins variable, donc plus précis, que $\hat{\sigma}_J(T_n)$, au moins dans les cas étudiés. Il est assez remarquable, et aussi un peu étonnant, qu'avec une petite taille d'échantillon ($n = 25$), la validité de l'estimation de $\sigma(T_n)$ obtenue par l'estimateur $\hat{\sigma}(T_n)$ de l'écart-type asymptotique soit aussi bonne. En fait, le biais relatif de $\bar{\hat{\sigma}}(T_n)$ par rapport à $\sigma(T_n)$, soit $|\bar{\hat{\sigma}}(T_n) - \sigma(T_n)|/\sigma(T_n)$, est inférieur à 3% lorsque $n = 25$, et à 1% lorsque $n = 50$.

Dans le cas de la dépendance entre X et Y, on ne peut pas calculer la valeur exacte de l'écart-type $\sigma(T_n)$ de T_n qui, cette fois-ci, dépend de la copule C associée à la loi H. Aussi avons-nous pris comme référence la variance échantillonnale $s^2(T_n) = \frac{1}{4\,999}\sum_{i=1}^{5\,000}\left(T_{in} - \bar{T}_n\right)^2$ obtenue à partir de 5 000 échantillons pseudo-aléatoires de taille $n = 25$ et 50. Comme la variable $4\,999 s^2(T_n)/\sigma^2(T_n)$ est approximativement distribuée selon une loi de khi-deux, on peut alors construire un intervalle de confiance approximatif de niveau 0.95 pour $\sigma(T_n)$ qui est donné par

$$0.981 s(T_n) \leqslant \sigma(T_n) \leqslant 1.020 s(T_n)\,. \tag{3}$$

Ceci permet d'avoir un critère pour évaluer les estimateurs $\hat{\sigma}(T_n)$ et $\hat{\sigma}_J(T_n)$. Ainsi, on considérera qu'un estimateur est valide ou non selon que la moyenne des estimations obtenues à partir des 5 000 échantillons appartient ou non à l'intervalle.

TABLEAU 2

Valeurs de l'écart-type échantillonnal $s(T_n)$, des moyennes $\bar{\hat{\sigma}}(T_n)$ et $\bar{\hat{\sigma}}_J(T_n)$ pour la statistique T_n, obtenues à partir de 5 000 échantillons de taille $n = 50$, pour la loi normale bidimensionnelle N, les trois distributions de Pearson de type VII P(m), $m = 5$, 2.5 et 1.5, et la loi avec dépendance par couplage, avec $r = 0.25$ et 0.5.

T_n			N	$P(5)$	$P(2.5)$	$P(1.5)$	DC
Spearman	r	s	0.1355	0.1366	0.1321	0.1636	0.1405
	$=$	$\bar{\hat{\sigma}}$	0.1372	0.1371	0.1382*	0.1648	0.1382
	0.25	$\bar{\hat{\sigma}}_J$	0.1397*	0.1397*	0.1407*	0.1676*	0.1408
	r	s	0.1133	0.1163	0.1168	0.1498	0.1239
	$=$	$\bar{\hat{\sigma}}$	0.1158	0.1164	0.1189	0.1481	0.1239
	0.50	$\bar{\hat{\sigma}}_J$	0.1179*	0.1185	0.1210*	0.1506	0.1253
van der Waerden	r	s	0.1167	0.1182	0.1145	0.1637	0.1234
	$=$	$\bar{\hat{\sigma}}$	0.1174	0.1172	0.1185*	0.1599	0.1202*
	0.25	$\bar{\hat{\sigma}}_J$	0.1205*	0.1203	0.1214*	0.1684*	0.1241
	r	s	0.0953	0.0985	0.0993	0.1500	0.1096
	$=$	$\bar{\hat{\sigma}}$	0.0966	0.0976	0.1005	0.1430*	0.1086
	0.50	$\bar{\hat{\sigma}}_J$	0.0985*	0.0997	0.1025*	0.1495	0.1095

Dans le Tableau 2, on trouve les valeurs de $s(T_n)$, des moyennes $\bar{\hat{\sigma}}(T_n)$ et $\bar{\hat{\sigma}}_J(T_n)$ lorsque $n = 50$, pour les cinq distributions suivantes du premier groupe décrit dans l'Annexe 2.1 : la loi normale bidimensionnelle, dénotée par N, avec un coefficient de corrélation de Pearson $r = 0.25, 0.50$; les distributions de Pearson de type VII, dénotées par P(m), avec $m = 1.5, 2.5, 5$ et $r = 0.25, 0.50$; la distribution avec dépendance par couplage, dénotée par DC, où les trois variables U, V et W suivent la loi double-exponentielle standard, et γ est un réel tel que $r = \gamma^2/(1+\gamma^2)$ prenne respectivement la valeur 0.25 et 0.50. Dans le Tableau 3, on trouve les mêmes quantités $s(T_n)$ $\bar{\hat{\sigma}}(T_n)$ et $\bar{\hat{\sigma}}_J(T_n)$ lorsque $n = 50$, pour les trois copules décrites dans l'Annexe 2.2 et dénotées par CJ, GB et Fr. Pour chacune de ces copules, les paramètres sont calculés de telle façon que le coefficient de concordance de Kendall τ prenne respectivement la valeur 0.25 et 0.50.

TABLEAU 3

Valeurs de l'écart-type échantillonnal $s(T_n)$, des moyennes $\bar{\hat{\sigma}}(T_n)$ et $\bar{\hat{\sigma}}_J(T_n)$ pour la statistique T_n, obtenues à partir de 5 000 échantillons de taille $n = 50$, pour les copules de Cook et Johnson (CJ), Gumbel B (GB) et Frank (Fr) avec $\tau = 0.25$ et 0.50.

T_n			CJ	GB	Fr
Spearman	$\tau = 0.25$	s	0.1304	0.1338	0.1253
		$\bar{\hat{\sigma}}$	0.1295	0.1333	0.1275
		$\bar{\hat{\sigma}}_J$	0.1337*	0.1345*	0.1296*
	$\tau = 0.50$	s	0.9024	0.0995	0.0803
		$\bar{\hat{\sigma}}$	0.0872*	0.0984	0.0811
		$\bar{\hat{\sigma}}_J$	0.9221*	0.0983	0.0822*
van der Waerden	$\tau = 0.25$	s	0.1110	0.1163	0.1078
		$\bar{\hat{\sigma}}$	0.1069*	0.1173	0.1091
		$\bar{\hat{\sigma}}_J$	0.1141*	0.1172	0.1116*
	$\tau = 0.50$	s	0.0725	0.0871	0.0678
		$\bar{\hat{\sigma}}$	0.0692*	0.0876	0.0696*
		$\bar{\hat{\sigma}}_J$	0.0741*	0.0852*	0.0694*

Dans les tableaux 2 et 3, les valeurs de $\bar{\hat{\sigma}}(T_n)$ et $\bar{\hat{\sigma}}_J(T_n)$ suivies d'un astérisque correspondent à celles qui n'appartiennent pas à l'intervalle de confiance approximatif de $\sigma(T_n)$ donné en (3). Ainsi, on peut voir que la performance de l'estimateur $\hat{\sigma}(T_n)$ est légèrement supérieure à celle de l'estimateur jackknife puisque la moyenne $\bar{\hat{\sigma}}(T_n)$ appartient plus souvent à l'intervalle de confiance que $\bar{\hat{\sigma}}_J(T_n)$. On peut aussi remarquer que le degré de dépendance de la loi du couple (X, Y) ne semble pas avoir d'influence sur la précision des deux estimateurs considérés. Les résultats obtenus lorsque $n = 25$ sont pratiquement semblables à ceux de $n = 50$ et ne sont donc pas présentés ici.

Il découle de ces simulations que l'estimateur $\hat{\sigma}(T_n)$ se comporte légèrement mieux pour estimer l'écart-type $\sigma(T_n)$ de T_n que l'estimateur jackknife correspondant, pour les statistiques de Spearman et de van der Waerden. Cependant, d'autres résultats obtenus par simulation et qui n'apparaissent pas ici, ont montré que lorsque la fonction score de la statistique T_n est asymétrique et non bornée, $\hat{\sigma}(T_n)$ présentait un biais négatif pour des petites tailles d'échantillon comme $n = 25$, par exemple. C'est le cas, en particulier, de la statistique de Savage $\frac{1}{n}\sum_{i=1}^{n}\left(1+\log\frac{R_i}{n+1}\right)\left(1+\log\frac{S_i}{n+1}\right)$ pour laquelle l'estimateur jackknife $\hat{\sigma}_J(T_n)$ est préférable à l'estimateur proposé.

4. Comparaison de la dépendance de deux lois bidimensionnelles

Soient deux échantillons bidimensionnels de taille n_1 et n_2 respectivement, l'un provenant d'une loi $H_1(x, y)$ et l'autre d'une loi $H_2(x, y)$, ces deux lois étant continues et ayant les mêmes marges. Supposons que l'on veuille savoir si l'une de ces deux lois a une dépendance plus forte que l'autre au sens, par exemple, de la dépendance par quadrant (Yanagimoto et Okamoto, 1969). On est alors amené à tester l'hypothèse nulle $H_1 = H_2$ contre l'hypothèse $H_1 \leqslant H_2$. Comme il a été dit dans l'introduction, un test classique pour ce problème, et peut-être le seul utilisé, est celui faisant intervenir la statistique suivante de Pearson, après la transformation z de Fisher,

$$\left\{\frac{(n_1-3)(n_2-3)}{4(n_1+n_2-6)}\right\}^{1/2}\log\left\{\frac{(1+r_{2n_2})(1-r_{1n_1})}{(1+r_{1n_1})(1-r_{2n_2})}\right\}$$

où r_{in_i}, $i = 1, 2$, est le coefficient de corrélation de Pearson de l'échantillon i. En utilisant l'approximation normale, on rejette l'hypothèse nulle lorsque cette statistique est supérieure au quantile d'ordre $1-\alpha$ de la loi normale $N(0, 1)$. Dans le cas où H_1 et H_2 sont deux lois normales bidimensionnelles et où $n_1 = n_2$, ce test est équivalent au test du rapport des vraisemblances maximales (voir par exemple Kendall et Stuart, vol. 2, chap. 26, exercices 26.19-21). Cependant, lorsque les lois ne sont par normales, les performances de ce test ont été peu étudiées, du moins à la connaissance des auteurs. On peut s'attendre néanmoins à de faibles performances de sa part lorsqu'on est éloigné du modèle normal, comme c'est assez souvent le cas pour des méthodes paramétriques associées à ce modèle. Aussi allons-nous étudier ce test et le comparer, au moyen de simulations, à ceux déduits des statistiques de rang de Kendall et de Spearman. Pour ce qui est du premier de ces tests, nous considérons la statistique

$$\frac{(\tau_{2n_2}-\tau_{1n_1})}{4\left(\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}\right)^{1/2}}$$

où τ_{in_i}, $i = 1, 2$, est le coefficient de concordance de Kendall de l'échantillon i et $16S_i^2/n_i$ un estimateur de la variance de τ_{in_i} (Samara et Randles, 1988, Genest et Rivest, 1993). On déduit des résultats de ces auteurs que sous l'hypothèse nulle, cette statistique suit approximativement une loi normale $N(0,1)$. Pour le deuxième test de rang, nous considérons la statistique

$$\frac{(\rho_{2n_2} - \rho_{1n_1})}{\left\{\hat{\sigma}^2\left(\rho_{1n_1}\right) + \hat{\sigma}^2\left(\rho_{2n_2}\right)\right\}^{1/2}}$$

où ρ_{in_i} est le coefficient de corrélation de rang de Spearman de l'échantillon i, $\hat{\sigma}^2(\rho_{in_i})$ le carré de l'estimateur de l'écart-type associé proposé à la Section 3,

TABLEAU 4

Seuils observés des tests de Pearson (Pe), Spearman (Sp) et Kendall (Ke) pour les lois du premier groupe, obtenus à partir de 2 000 échantillons de tailles $n_1 = n_2 = 50$.

Lois	r	α	Pe	Sp	Ke
N	0.25	0.05	0.0550	0.0530	0.0530
		0.10	0.1120	0.1045	0.1045
	0.50	0.05	0.0500	0.0470	0.0485
		0.10	0.1030	0.0985	0.1020
$P(5)$	0.25	0.05	0.0480	0.0540	0.0535
		0.10	0.1080	0.1000	0.1010
	0.50	0.05	0.0555	0.0500	0.0535
		0.10	0.1040	0.1005	0.1035
$P(2.5)$	0.25	0.05	0.0520	0.0510	0.0545
		0.10	0.1015	0.1010	0.1025
	0.50	0.05	0.0485	0.0505	0.0505
		0.10	0.1125	0.1085	0.1045
$P(1.5)$	0.25	0.05	0.4015*	0.0520	0.0505
		0.10	0.4305*	0.0965	0.1005
	0.50	0.05	0.3990*	0.0515	0.0500
		0.10	0.4210*	0.1075	0.1040
DC	0.25	0.05	0.0750*	0.0595	0.0560
		0.10	0.1270*	0.1030	0.1005
	0.50	0.05	0.1110*	0.0505	0.0530
		0.10	0.1695*	0.1045	0.1030

statistique qui, d'après les résultats de Ruymgart, Shorack et van Zwet (1972) et du théorème de la Section 2, suit approximativement une loi normale $N(0,1)$ sous l'hypothèse nulle.

Les résultats des simulations que nous allons présenter sont obtenus à partir de 2 000 échantillons pseudo-aléatoires de tailles n_1 et n_2 égales à 50, et concernent le seuil observé et la puissance de ces trois tests dans les différentes situations décrites dans les Tableaux 4-7. Les lois simulées sont celles présentées à l'Annexe 2.

Examinons tout d'abord les résultats apparaissant dans les Tableaux 4 et 5.

TABLEAU 5

Seuils observés des tests de Pearson (Pe), de Spearman (Sp) et de Kendall (Ke) pour les copules du deuxième groupe avec des marges normales, double-exponentielle (De) et Cauchy, obtenus à partir de 2 000 échantillons de taille $n_1 = n_2 = 50$. *Les résultats pour les tests de Spearman et de Kendall sont indépendants des marges.*

Copules	τ	α	Pe Normale	Pe De	Pe Cauchy	Sp	Ke
CJ	0.25	0.05	0.0725*	0.1235*	0.1955*	0.0545	0.0530
		0.10	0.1270*	0.1785*	0.2395*	0.1030	0.0990
	0.50	0.05	0.0770*	0.1730*	0.3060*	0.0555	0.0460
		0.10	0.1310*	0.2230*	0.3410*	0.1030	0.0955
Fr	0.25	0.05	0.0520	0.0550	0.0615*	0.0440	0.0455
		0.10	0.0940	0.1065	0.0970*	0.0900	0.0885
	0.50	0.05	0.0525	0.0715*	0.1370*	0.0505	0.0525
		0.10	0.0970	0.1390*	0.1950*	0.0955	0.0955
GB	0.25	0.05	0.0690*	0.1185*	0.1575*	0.0575	0.0585
		0.10	0.1190*	0.1675*	0.1955*	0.1030	0.1020
	0.50	0.05	0.1105*	0.1600*	0.3055*	0.0480	0.0590
		0.10	0.1660*	0.2195*	0.3445*	0.1040	0.1085

Dans le Tableau 4, on trouve les seuils observés des trois tests pour les cinq lois du premier groupe (voir Annexe 2) avec $r = 0.25$ et 0.50, alors que dans le Tableau 5, on a ceux pour les lois du deuxième groupe avec $\tau = 0.25$ et 0.50, et ce pour un seuil nominal $\alpha = 0.05$ et 0.10. Afin d'évaluer la proximité d'un seuil observé au seuil nominal correspondant, on se sert de l'écart-type de la loi binomiale de paramètres $n = 2\,000$ et $p = \alpha$, ce qui donne 0.00487 pour $\alpha = 0.05$ et 0.00671 pour $\alpha = 0.10$. Nous considérons alors qu'un test respecte le seuil nominal α si son seuil observé appartient à l'intervalle de plus ou moins deux écart-types autour de α. Les seuils observés apparaissant dans les deux tableaux et qui sont suivis d'un astérisque sont ceux n'appartenant pas à l'intervalle (0.04026, 0.05974) pour $\alpha = 0.05$ et à l'intervalle (0.08658, 0.11342) pour $\alpha = 0.10$. On constate alors que les tests de Spearman et de Kendall respectent le seuil fixé, ce qui est somme toute normal pour des tests de rang, alors qu'il en est tout autrement dans plusieurs situations pour le test de Pearson. Plus précisément, il est intéressant de remarquer que même si les marges d'une loi sont normales, ce dernier test n'a pas obligatoirement un seuil robuste. C'est le cas en effet pour une loi qui a des marges normales et dont la copule est celle de Cook et Johnson ou de Gumbel B asymétrique. Il en résulte que la vérification de la normalité des marges d'une loi bidimensionnelle n'est pas suffisante pour assurer la validité de ce test.

TABLEAU 6

Puissances × 100 des tests de Pearson (PE), de Spearman (Sp) et de Kendall (Ke) pour des hypothèses nulles spécifiant $r = 0.25$ (partie gauche du tableau) et $r = 0.50$ (partie droite du tableau), pour les lois du premier groupe, obtenues à partir de 2 000 échantillons de tailles $n_1 = n_2 = 50$, avec un seuil nominal $\alpha = 0.05$.

Lois	r	Pe	Sp	Ke	r	Pe	Sp	Ke
N	0.25	5.50	5.30	5.30	0.5	5.00	4.70	4.85
	0.45	30.40	28.20	28.15	0.6	16.95	15.05	15.40
	0.65	82.60	75.70	76.25	0.7	47.15	39.95	41.10
	0.85	100.00	99.50	99.85	0.8	85.25	73.80	76.10
$P(5)$	0.25	4.80	5.40	5.35	0.5	5.55	5.00	5.35
	0.45	26.95	28.15	27.95	0.6	14.65	14.35	14.60
	0.65	73.05	71.95	72.10	0.7	38.20	36.50	37.15
	0.85	99.15	99.05	99.20	0.8	71.20	70.30	71.95
$P(2.5)$	0.25	5.20	5.10	5.45	0.5	4.85	5.05	5.05
	0.45	20.20	24.40	23.85	0.6	12.25	14.25	14.05
	0.65	49.50	68.95	68.75	0.7	23.65	32.50	32.80
	0.85	84.90	98.85	98.95	0.8	42.80	65.70	66.00
$P(1.5)$	0.25	40.15	5.20	5.60	0.5	39.90	5.15	5.00
	0.45	49.20	20.60	22.25	0.6	47.70	12.90	13.25
	0.65	60.45	52.10	56.55	0.7	49.25	25.45	27.70
	0.85	70.25	82.10	86.55	0.8	58.00	47.05	52.85
DC	0.25	7.50	5.95	5.60	0.5	11.10	5.05	5.30
	0.45	34.50	21.00	21.65	0.6	24.70	10.10	10.70
	0.65	75.40	52.30	53.30	0.7	48.00	23.30	24.45
	0.85	99.20	91.40	93.35	0.8	75.75	46.90	49.35

Les résultats apparaissant dans les Tableaux 6 et 7 concernent la puissance (multipliée par 100) des trois tests dans différentes situations, avec un seuil $\alpha = 0.05$. Dans le Tableau 6, on considère les lois du premier groupe pour lesquelles le paramètre r de la loi du deuxième échantillon varie de 0.25 (hypothèse nulle)[1] à 0.85 dans un premier cas, et de 0.50 (hypothèse nulle)[1] à 0.80 dans un deuxième cas.

On remarque tout d'abord que lorsque les observations suivent une loi normale, le test de Pearson domine les deux autres alors que pour la loi de Pearson de type VII avec $m = 5$, les trois tests sont équivalents. Lorsque $m = 2.5$, les deux tests de rang sont équivalents et dominent nettement celui de Pearson. Ce dernier ne respectant pas le seuil de 5% pour les deux dernières lois, sa puissance ne sera pas comparée à celle des deux autres. Quant à eux, on remarque au moyen d'un test de comparaison de proportions que le test de Kendall a une puissance légèrement supérieure à celle de Spearman.

TABLEAU 7

Puissances $\times$ 100 des tests de Pearson (Pe), de Spearman (Sp) et de Kendall (Ke) pour des hypothèses nulles spécifiant $\tau = 0.2$ (partie gauche du tableau) et $\tau = 0.5$ (partie droite du tableau) pour les copules du deuxième groupe, obtenues à partir de 2 000 échantillons de tailles $n_1 = n_2 = 50$, avec un seuil nominal $\alpha = 0.05$.

Copules	τ	Pe	Sp	Ke	τ	Pe	Sp	Ke
CJ	0.2	7.25	5.45	5.30	0.5	7.70	5.55	4.60
	0.4	33.20	29.65	29.00	0.6	27.75	23.25	22.95
	0.6	90.80	90.50	90.80	0.7	66.00	64.40	66.55
	0.8	100.00	100.00	100.00	0.8	94.45	95.95	97.05
Fr	0.2	5.20	4.40	4.55	0.5	5.25	5.05	5.25
	0.4	49.00	49.80	49.60	0.6	26.45	26.00	28.25
	0.6	97.00	97.60	98.15	0.7	74.20	75.15	79.80
	0.8	100.00	100.00	100.00	0.8	98.10	99.15	99.60
GB	0.2	6.90	5.75	5.85	0.5	11.05	4.80	5.90
	0.4	48.35	42.95	44.80	0.6	30.00	15.75	21.85
	0.6	90.75	87.70	93.10	0.7	49.10	33.05	54.60
	0.75	96.05	95.10	99.55	0.75	53.70	36.35	68.35

Le Tableau 7 présente les puissances pour les trois copules dans lesquelles le paramètre τ de la copule du deuxième échantillon varie de 0.2 (hypothèse nulle)[2] à 0.8 dans un premier cas, et de 0.5 (hypothèse nulle)[2] à 0.8 dans un deuxième cas. Cependant, pour la copule de Gumbel B asymétrique, $\tau = 0.8$ a été remplacé par $\tau = 0.75$ car il n'a pas été possible d'atteindre numériquement ce degré de dépendance en faisant varier les paramètres α, β et θ.

[1] La valeur de r associée à l'hypothèse nulle correspond à celle du premier échantillon.

[2] La valeur de τ associée à l'hypothèse nulle correspond à celle du premier échantillon.

Ces résultats montrent en premier lieu que pour la copule de Frank, les trois tests sont pratiquement équivalents, sauf pour $\tau = 0.7$ où le test de Kendall est légèrement supérieur. Pour les deux autres copules, le test de Pearson ne respecte pas le seuil de 5%. Dans le cas de la copule de Cook et Johnson, les deux tests de rang sont équivalents alors que pour la dernière copule, le test de Kendall domine le test de Spearman.

5. Conclusion

Dans ce travail, nous avons présenté un estimateur de la variance asymptotique de statistiques linéaires de rang utilisées dans les tests d'indépendance. Cet estimateur est fortement convergent et a montré, lors de simulations de Monte-Carlo, de bonnes propriétés de validité et de précision. En particulier, il est légèrement meilleur que l'estimateur jackknife correspondant, pour les statistiques de Spearman et de van der Waerden. Un tel estimateur permet, entre autres, de construire des intervalles de confiance pour les mesures de dépendance associées aux statistiques de rang, de construire des tests portant sur la force de dépendance de lois bidimensionnelles et sur la comparaison de deux lois bidimensionnelles selon leur degré de dépendance. Dans le cadre de ce dernier problème, nous avons comparé les tests de Spearman, de Kendall et de Pearson. Les résultats ont montré que ce dernier test n'a pas un seuil robuste lorsqu'on s'éloigne du modèle normal bidimensionnel, même si les lois marginales sont normales. Dans un tel cas, il devrait être remplacé par un des deux tests de rangs qui sont souvent équivalents avec cependant une légère supériorité du test de Kendall dans certaines situations. Ceci est peut-être à rapprocher des résultats obtenus par Woodworth (1970) qui montrent que l'efficacité relative de Bahadur du test d'indépendance de Kendall par rapport à celui de Spearman est légèrement supérieure à 1 dans le cas de la loi normale.

Bibliographie

BANDEEN-ROCHE K.J., LIANG K.-Y. (1996), Modelling failure-time associations in data with multiple levels of clustering, *Biometrika*, **83**, pp. 29-39.

BEHNEN K. (1971), Asymptotic optimality and ARE of certain rank order tests under contiguity. *Ann. Math. Statist.*, **42**, pp. 325-329.

BHUCHONGKUL S. (1964), A class of nonparametric tests for independence in bivariate populations. *Ann. Math. Statist.*, **35**, pp. 138-149.

CAPÉRAÀ P., FOUGÈRES A.-L., GENEST C. (1999), Bivariate distributions with given extreme value attractors, *J. Multivariate Anal.*, sous presse.

COLES S. G., TAWN J. A. (1991), Modelling extreme multivariate events, *J. Royal Statist. Soc. Ser. B*, **53**, pp. 377-392.

COOK R. D., JOHNSON M. E. (1981), A family of distributions for modelling nonelliptically symmetric multivariate data, *J. Roy. Statist. Soc. Ser. B*, **43**, pp. 210-218.

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

GARRALDA-GUILLEM A. I. (1997), Thèse de doctorat non publiée, Universidad de Granada, Espagne.

GENEST C. (1987), Frank's family of bivariate distributions, *Biometrika*, **74**, pp. 549-555.

GENEST C., GHOUDI K., RIVEST L.-P. (1995), A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82**, pp. 543-552.

GENEST C., MACKAY R. J. (1986a), Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données, *Canad. J. Statist.*, **14**, pp. 145-159.

GENEST C., MACKAY R. J. (1986b), The joy of copulas : bivariate distributions with uniform marginals, *Amer. Statist.*, **40**, pp. 280-283.

GENEST C., RIVEST L.-P. (1993), Statistical inference procedures for bivariate Archimedian copulas, *J. Amer. Statist. Assoc.*, **88**, pp. 1034-1043.

GOKHALE D. V. (1968), On asymptotic relative efficiencies of a class of rank tests for independence of two variables, *Ann. Inst. Statist. Math.*, **20**, pp. 255-261.

HAAN DE L., RONDE DE J. (1998), Sea and wind : multivariate extremes at work, *Extremes*, **1**, 7-45.

HÁJEK J., SĭDÁK Z. (1967), *Theory of Rank Tests*. Academic Press, New York.

JOHNSON M. E. (1987), *Multivariate Statistical Simulation*. John Wiley, New York.

KENDALL M. G., STUART A. (1961), *The Advanced Theory of Statistics*, Vol. 2. Charles Griffin & Company Limited.

PICKANDS J. (1981), Multivariate extreme value distributions, *Bull. Int. Statist. Inst.*, pp. 859-878.

RESNICK S. I. (1987), *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, New York.

RUYMGAART F. H., SHORACK G. R., VAN ZWET W. R. (1972), Asymptotic normality of nonparametric tests for independence, *Ann. Math. Statist.*, **43**, pp. 1122-1135.

SAMARA B., RANDLES R. H. (1988), A test for correlation based on Kendall's tau, *Commun. Statist.-Theory Meth.*, **17**, pp. 3191-3205.

SKLAR A. (1959), Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statistique, Univ. Paris*, **8**, pp. 229-231.

TAWN J. A. (1988), Bivariate extreme value theory : Models and estimation, *Biometrika*, **75**, pp. 397-415.

WOLTER K. M. (1985), *Introduction to Variance Estimation*, Springer-Verlag.

WOODWORTH G. H. (1970), Large deviations and Bahadur efficiency of linear rank statistics, *Ann. Math. Statist.*, **41**, pp. 251-283.

YANAGIMOTO T., OKAMOTO M. (1969), Partial orderings of permutations and monotonicity of a rank correlation statistic, *Ann. Inst. Statist. Math.*, **21**, pp. 489-506.

Annexe 1

Schéma de la démonstration du Théorème de la Section 2

Considérons les statistiques linéaires de rang données par (1) avec des fonctions scores J et K satisfaisant les conditions de régularité C suivantes (Ruymgaart *et al.*, 1972).

Conditions C **:**

(ii) J et K sont continues et croissantes sur $(0, 1)$;

(ii) J et K sont deux fois dérivables et J' et K' sont continues sur $(0, 1)$;

(iii) il existe des nombres réels $M > 0$, $a > 0$ et $b > 0$ satisfaisant les relations $a = \left(\frac{1}{2} - \delta\right)/p_0$ et $b = \left(\frac{1}{2} - \delta\right)/q_0$ pour $0 < \delta < 1/2$ et $p_0, q_0 > 1$ avec $p_0^{-1} + q_0^{-1} = 1$, tels que pour $i = 1, 2$, on a pour tout $0 < u < 1$

$$\left|J^{(i)}(u)\right| < M\{u(1-u)\}^{a+i} \quad \text{et} \quad \left|K^{(i)}(u)\right| < M\left\{u(1-u)\right\}^{b+i} .$$

Considérons la variance asymptotique σ^2 donnée par (2).

On déduit du théorème de Fubini, qui est applicable ici, que

$$\mathrm{E}\left\{\iint (\phi_X - F)J'(F)K(G)dH\right\} = \mathrm{E}\left\{\iint (\phi_Y - G)J(F)K'(G)dH\right\} = 0 .$$

On a alors

$$\begin{aligned}
\sigma^2 = \mathrm{Var}\{U(X,Y)\} &= \mathrm{E}[J^2\{F(X)\}K^2\{G(Y)\}] \\
&+ \mathrm{E}\left[\left\{\iint (\phi_X - F)J'(F)K(G)dH\right\}^2\right] \\
&+ \mathrm{E}\left[\left\{\iint (\phi_Y - G)J(F)K'(G)dH\right\}^2\right] \\
&+ 2\mathrm{E}\left[J\{F(X)\}K\{G(Y)\}\iint (\phi_X - F)J'(F)K(G)dH\right] \\
&+ 2\mathrm{E}\left[J\{F(X)\}K\{G(Y)\}\iint (\phi_Y - G)J(F)K'(G)dH\right] \\
&+ 2\mathrm{E}\left[\left\{\iint (\phi_X - F)J'(F)K(G)dH\right\}\left\{\iint (\phi_Y - G)J(F)K'(G)dH\right\}\right] \\
&- \mathrm{E}^2\left[J\{F(X)\}K\{G(Y)\}\right] .
\end{aligned}$$

D'une façon semblable, on peut exprimer $\hat{\sigma}^2$ par

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} U_n^2(X_i,Y_i) - \overline{U}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} J^2\{F_n^*(X_i)\}K^2\{G_n^*(Y_i)\} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left[\iint(\phi_{X_i} - F_n^*)J'(F_n^*)K(G_n^*)dH_n\right]^2 \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left[\iint(\phi_{Y_i} - G_n^*)J(F_n^*)K'(G_n^*)dH_n\right]^2 \\
&+ \frac{2}{n}\sum_{i=1}^{n}\left[J\{F_n^*(X_i)\}K\{G_n^*(Y_i)\}\iint(\phi_{X_i} - F_n^*)J'(F_n^*)K(G_n^*)dH_n\right] \qquad (4)\\
&+ \frac{2}{n}\sum_{i=1}^{n}\left[J\{F_n^*(X_i)\}K\{G_n^*(Y_i)\}\iint(\phi_{Y_i} - G_n^*)J(F_n^*)K'(G_n^*)dH_n\right] \\
&+ \frac{2}{n}\sum_{i=1}^{n}\left[\iint(\phi_{X_i}-F_n^*)J'(F_n^*)K(G_n^*)dH_n\iint(\phi_{Y_i}-G_n^*)J(F_n^*)K'(G_n^*)dH_n\right] \\
&- \left[\frac{1}{n}\sum_{i=1}^{n} J\{F_n^*(X_i)\}K\{G_n^*(Y_i)\}\right. \\
&+ \frac{1}{n}\sum_{i=1}^{n}\iint(\phi_{X_i} - F_n^*)J'(F_n^*)K(G_n^*)dH_n \\
&\left.+ \frac{1}{n}\sum_{i=1}^{n}\iint(\phi_{Y_i} - G_n^*)J(F_n^*)K'(G_n^*)dH_n\right]^2 .
\end{aligned}$$

Posons

$$\begin{aligned}
V(X,Y) &= J\{F(X)\}K\{G(Y)\}\,, \\
V_n(X_i,Y_i) &= J\{F_n^*(X_i)\}K\{G_n^*(Y_i)\}\,, \\
W_1(X) &= \iint(\phi_X - F)J'(F)K(G)dH\,, \\
W_{1n}(X_i) &= \iint(\phi_{X_i} - F_n^*)J'(F_n^*)K(G_n^*)dH_n\,, \\
W_2(Y) &= \iint(\phi_Y - G)J(F)K'(G)dH\,, \\
W_{2n}(Y_i) &= \iint(\phi_{Y_i} - G_n^*)J(F_n^*)K'(G_n^*)dH_n\,.
\end{aligned}$$

Pour établir la convergence presque sûre de $\hat{\sigma}^2$, il faut montrer que les termes linéaires et quadratiques de (4) convergent presque sûrement. Plus précisément, on prouve successivement que

1)
$$\frac{1}{n}\sum_{i=1}^{n} V_n(X_i,Y_i) \longrightarrow \mathrm{E}\{V(X,Y)\} \text{ p.s.},$$
$$\frac{1}{n}\sum_{i=1}^{n} W_{1n}(X_i) \longrightarrow \mathrm{E}\{W_1(X)\} = 0 \text{ p.s.},$$
$$\frac{1}{n}\sum W_{2n}(Y_i) \longrightarrow \mathrm{E}(W_2(Y)) = 0 \text{ p.s.},$$

d'où l'on peut conclure que $\overline{U}_n^2 \longrightarrow \mathrm{E}^2\{U(X,Y)\}$ p.s.;

2)
$$\frac{1}{n}\sum_{i=1}^{n} V_n^2(X_i,Y_i) \longrightarrow \mathrm{E}\{V^2(X,Y)\} \text{ p.s.},$$
$$\frac{1}{n}\sum_{i=1}^{n} W_{1n}^2(X_i) \longrightarrow \mathrm{E}\{W_1^2(X)\} \text{ p.s.},$$
$$\frac{1}{n}\sum_{i=1}^{n} W_{2n}^2(Y_i) \longrightarrow \mathrm{E}(W_2^2(Y)) \text{ p.s.},$$
$$\frac{1}{n}\sum_{i=1}^{n} V_n(X_i,Y_i)W_{1n}(X_i) \longrightarrow \mathrm{E}\left\{V(X,Y)W_1(X)\right\} \text{ p.s.},$$
$$\frac{1}{n}\sum_{i=1}^{n} V_n(X_i,Y_i)W_{2n}(Y_i) \longrightarrow \mathrm{E}\{V(X,Y)W_2(Y)\} \text{ p.s.},$$
$$\frac{1}{n}\sum_{i=1}^{n} W_{1n}(X_i)W_{2n}(Y_i) \longrightarrow \mathrm{E}\{W_1(X)W_2(Y)\},$$

d'où l'on déduit que

$$\frac{1}{n}\sum_{i=1}^{n} U_n^2(X_i,Y_i) \longrightarrow \mathrm{E}\{U^2(X,Y)\} \text{ p.s.},$$

ce qui complète la démonstration.

Annexe 2

Lois bidimensionnelles utilisées dans les simulations

Nous considérons deux groupes de lois bivariées. Dans le premier groupe, la force ou le degré de dépendance entre les variables X et Y est mesuré par un paramètre r, qui n'est autre que le coefficient de corrélation de Pearson lorsque les seconds moments de la loi existent. Dans le second groupe, la dépendance est paramétrisée par le coefficient de concordance de Kendall τ. On peut noter que pour ces deux groupes, si deux lois sont ordonnées par r ou τ, alors elles le sont selon l'ordre de dépendance par quadrant (Yanaginoto et Okamoto, 1969).

A.2.1 Premier groupe

Distributions bidimensionnelles de Pearson de type VII

Ces lois, appelées aussi lois de Student bidimensionnelles, ont des densités de la forme suivante (voir par exemple Johnson, 1987, chap. 6) :

$$f(u) = \frac{m-1}{\pi}\left|\sum\right|^{-1/2}\{1+(u-\mu)'\sum^{-1}(u-\mu)\}^{-m}, \mu \in \mathbb{R}^2$$

où $m > 1$ et $\sum$ est une matrice symétrique définie positive d'ordre 2, $|\sum|$ son déterminant et μ un vecteur de $\mathbb{R}^2$. Pour $m > 2$, on a $\mathrm{E}(X) = \mu$ et $\mathrm{Cov}(U) = \frac{2m-2}{2m-4}\sum$. En posant $m = 1.5$, on obtient la loi de Cauchy bidimensionnelle et lorsque m tend vers l'infini, la loi normale bidimensionnelle. Dans les simulations à venir, nous avons pris $\sum = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$. Des algorithmes pour simuler de telles lois sont décrits dans le chapitre 6 du livre de Johnson (1987).

Distributions avec dépendance par couplage

Ces lois sont celles des couples de variables aléatoires (X, Y) telles que

$$X = U + \gamma W\,, \quad Y = V + \gamma W$$

où U, V et W sont des variables mutuellement indépendantes et γ un nombre réel. Lorsque les moments d'ordre 2 existent, on a $r = \gamma^2/(1+\gamma^2)$. Ce type de lois a été utilisé, en particulier, pour étudier l'efficacité asymptotique des tests d'indépendance (voir par exemple Bhuchongkul, 1964, Hájek et Šidák, 1967). Pour les simulations, les trois variables U, V et W suivent une loi double-exponentielle standard.

A.2.2 Deuxième groupe : les copules

Soit H une loi bidimensionnelle avec pour lois marginales F et G. Si F et G sont continues, il existe une distribution unique C définie sur le carré unité, de lois marginales uniformes sur l'intervalle $[0, 1]$, telle que

$$\forall (x, y) \in \mathbb{R}^2\,, \quad H(x, y) = C\{F(x), G(y)\}\;.$$

Cette fonction C, introduite par Sklar (1959) et appelée une copule, caractérise la structure de dépendance de H. Une classe particulière de telles copules est la famille des copules Archimax qui ont la forme suivante (Capéraà, Fougères et Genest, 1999)

$$\forall (u, v) \in [0, 1]^2\,, \quad C_{\phi,A}(u, v) = \phi^{-1}\left[\{\phi(u)+\phi(v)\}\, A\left\{\frac{\phi(u)}{\phi(u)+\phi(v)}\right\}\right]$$

où ϕ est une fonction convexe, décroissante, définie sur $[0,1]$ et à valeurs dans $\mathbb{R}^+$, vérifiant $\phi(1)=0$ avec la convention $\phi(0)=\lim_{t\to 0+}\phi(t)$ et $\phi^{-1}(s)=0$ pour $s\geqslant\phi(0)$, et A est une fonction convexe de $[0,1]$ dans $[0,1]$ vérifiant $A(0)=A(1)=1$ et $\max(t,1-t)\leqslant A(t)\leqslant 1$.

L'intérêt de cette famille est qu'elle contient de nombreuses lois symétriques et asymétriques que l'on peut aisément simuler (un algorithme de simulation est donné dans Capéraà *et al.*, 1999). En particulier, lorsqu'on prend $A\equiv 1$, on retrouve les copules archimédiennes (Genest et MacKay, 1986 a,b) et lorsqu'on pose $\phi(t)=-\log t$, on retrouve alors les copules des valeurs extrêmes bidimensionnelles (Pickands, 1981). Pour toutes ces copules Archimax, le tau de concordance de Kendall s'exprime comme suit :

$$\tau_{\phi,A}=\tau_A+(1-\tau_A)\tau_\phi$$

$$\text{où}\qquad \tau_A=\int_0^1\frac{t(1-t)}{A(t)}\,dA'(t)\quad\text{et}\quad \tau_\phi=1+4\int_0^1\frac{\phi(t)}{\phi'(t)}\,dt\,.$$

Copules de Cook et Johnson (1981)

Ces copules ont la forme suivante pour tous $0\leqslant u,v\leqslant 1$,

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

$$\phi_\alpha(t)=\frac{t^{-\alpha}-1}{\alpha}\quad\text{et}\quad A\equiv 1\,.$$

Copules de Gumbel B asymétriques ou modèle logistique (Tawn, 1988)

Ces copules ont la forme suivante pour tous $0\leqslant u,v\leqslant 1$

$$C_{\alpha,\beta,\theta}(u,v)$$

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

$$\theta\geqslant 1\,,\quad 0<\alpha,\beta\leqslant 1\,,\quad \phi(t)=-\log t\quad\text{et}$$

$$A(t)=(1-\beta)+(\beta-\alpha)t+\{t^\theta\alpha^\theta+(1-t)^\theta\beta^\theta\}^{1/\theta}\,.$$

Copules de Frank (1979), (Genest, 1987)

Ces copules ont la forme suivante pour tous $0\leqslant u,v\leqslant 1$

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

$$\phi_\alpha(t)=-\log\frac{1-\alpha^t}{1-\alpha}\quad\text{et}\quad A\equiv 1\,.$$