Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles *

Kilani Ghoudi
Département de mathématiques et d'informatique
Université du Québec à Trois-Rivières
Trois-Rivières, Québec G9A 5H7

Abdelhaq Khoudraji
Département de mathématiques, Faculté des sciences Semlalia
Université Cadi Ayyad, Boulevard du Prince Moulay Abdellah
B. P. S15, Marrakech, Maroc

Louis-Paul Rivest
Département de mathématiques et de statistique
Université Laval, Ste-Foy
Québec G1K 7P4

Titre abrégé: Copules de valeurs extrêmes

Le 5 février 1997

*Le financement de ce travail de recherche a été assuré en partie par le Conseil de recherches en sciences naturelles et en génie du Canada, par le Fonds pour la formation de chercheurs et l'aide à la recherche du Gouvernement de Québec, ainsi que le Fonds institutionnel de la recherche de l'Université du Québec à Trois-Rivières.

Propriétés statistiques des copules de valeurs extrêmes bidimensionelles

Keywords and phrases: Galambos' distribution, Goodness of fit, Gumbel's dependence function, Jackknife variance estimator,Multivariate Extreme Value Distributions, Simulation, U-statistic.
AMS 1991 subject classification: Primary 62H05; secondary 62E10.

Résumé

Soit (X,Y) un couple de variables aléatoires dont la fonction de répartition $H(x,y)$ est une loi des valeurs extrêmes bidimensionnelles. Si F_1 et F_2 sont les lois marginales de X et Y, on a $H(x,y) = C(F_1(x), F_2(y))$, où C est une copule de valeurs extrêmes bidimensionnelles. Dans un premier temps, on détermine la distribution conjointe de $Z = \log(F_1(X))/\log(F_1(X)F_2(Y))$ et de $W = C(F_1(X), F_2(Y))$. Ce résultat a plusieurs applications intéressantes. Il permet d'abord de construire un algorithme relativement simple pour simuler des lois de valeurs extrêmes bidimensionnelles. De plus quelque soit la copule des valeurs extrêmes C, ce résultat montre également que la loi marginale de $W = C(F_1(X), F_2(Y))$ appartient à une famille de distributions indicée par un paramètre. Cette observation permet de construire un test d'ajustement pour déterminer si une copule C appartient à la famille des copules de valeurs extrêmes.

Abstract

Let (X,Y) be a bivariate random vector whose distribution function $H(x,y)$ belongs to the class of bivariate extreme value distributions. If F_1 and F_2 are the marginals of X and Y, then $H(x,y) = C(F_1(x), F_2(y))$, where C is a bivariate extreme value dependence function. This paper gives the joint distribution of the random variables $Z = \log(F_1(X))/\log(F_1(X)F_2(Y))$ and $W = C(F_1(X), F_2(Y))$. Using this distribution, an algorithm to generate random variables having bivariate extreme value distribution is presented. Furthermore, it is shown that for any bivariate extreme value dependence function C, the distribution of the random variable $W = C(F_1(X), F_2(Y))$ belongs to a monoparametric family of distributions. This property is used to derive goodness of fit statistics to determine whether a copula belongs to an extreme value family.

1 Introduction

Dans plusieurs domaines d'applications de la statistique, comme par exemple en hydrologie ou en environnement, l'étude des grandes valeurs d'un échantillon revêt un intérêt particulier. Ainsi, lors de la construction d'un barrage hydroélectrique sur une rivière, les valeurs maximales des crues printanières des affluents de la rivière sont des données très importantes qui interviennent dans la planification du barrage. Les modèles statistiques pour les valeurs extrêmes sont des outils précieux pour comprendre le comportement des grandes valeurs d'un échantillon.

La théorie mathématique des modèles univariés pour les valeurs extrêmes est bien développée (voir Galambos, 1987). Les modèles multidimensionnels des valeurs extrêmes sont les distributions limites, lorsque la taille d'échantillon tend vers l'infini, de la loi conjointe des maxima marginaux, convenablement normalisés, dans un échantillon de p variables aléatoires. Si $F_1(x)$,..., $F_p(y)$ sont les lois marginales des valeurs extrêmes, que nous supposerons non dégénérées, les distributions limites sont du type $C(F_1(x_1), F_2(x_2), ..., F_p(x_p))$, où C est une copule de valeurs extrêmes. Rappelons qu'une copule est une fonction de répartition dont les marges sont uniformes (Sklar, 1959). Dans le cas bidimensionnel, Geoffroy (1958), Tiago de Olivera (1958) et Sibuya (1960) ont donné la forme générale des copules de valeurs extrêmes. Elles s'expriment comme suit,

$$C(u,v) = \exp\left[(\log u + \log v) A\left(\frac{\log u}{\log u + \log v}\right)\right]; \quad 0 < u, v < 1 \tag{1.1}$$

en terme d'une fonction convexe A définie sur $[0,1]$, satisfaisant

$$\max\{v, 1-v\} \leq A(v) \leq 1. \tag{1.2}$$

Plusieurs familles paramétriques de modèles pour A ont été suggérées; voir, par exemple, Anderson et Nadarajah (1993). Tawn (1988) discute de l'estimation des paramètres des lois bidimensionnelles pour des valeurs extrêmes. Pickands (1981) a donné la forme générale des copules de valeurs extrêmes multidimensionnels. Joe (1990) et (1994) présentent des familles paramétriques pour la loi conjointe de plus de deux valeurs extrêmes.

Soit (X, Y) un couple de variables aléatoires dont la fonction de répartition est donnée par $H(x,y) = C(F_1(x), F_2(y))$, où C est donnée par (**??**) pour une certaine fonction A. La Section 2 donne la loi conjointe des variables aléatoires

$$Z = \frac{\log(F_1(X))}{\log(F_1(X)F_2(Y))} \quad \text{et} \quad W = C(F_1(X), F_2(Y)).$$

Ce résultat est utilisé à la Section 3 pour construire un algorithme simple pour simuler des lois de valeurs extrêmes bidimensionnelles. Il permet ensuite d'élaborer, à la Section 4, un test d'ajustement pour les copules de valeurs extrêmes bidimensionnelles.

Tous les résultats obtenus dans ce travail sont indépendants des marges. On peut donc supposer, sans perte de généralité, que les marges de X et de Y sont uniformes; i.e. $F_1(x) = F_2(x) = x$, pour $x \in (0,1)$. On a alors $H(x,y) \equiv C(x,y)$.

2 Calcul de la loi conjointe de Z et de W

Dans cette section (X,Y) représente un couple de variables aléatoires distribué selon $C(x,y)$, on écrira $(X,Y) \sim C$, où C est donné par (**??**) pour une certaine fonction A. Les démonstrations des deux propositions suivantes sont reléguées en annexe.

Proposition 1 *Si $(X,Y) \sim C$, alors la loi conjointe de X et de $Z = \log X/\log(XY)$ est donnée par*

$$P(Z \leq z, X \leq x) = G(z,x) = \left\{ z + z(1-z)\frac{A'(z)}{A(z)} \right\} x^{\frac{A(z)}{z}} \text{ où } 1 \geq z, x \geq 0$$

et $A'(z)$ dénote la dérivée à droite de la fonction A au point z.

Puisque A est une fonction convexe satisfaisant (**??**), la dérivée à droite de A est définie dans $[0,1)$ et satisfait $-1 \leq A'(z) \leq 1$. Par extension, on définit $A'(1)$ comme étant le supremum, sur $(0,1)$, de $A'(z)$. En posant $u = 1$ dans le résultat précédent, on obtient la loi marginale de Z,

$$P(Z \leq z) = G_Z(z) = z + z(1-z)\frac{A'(z)}{A(z)}, \ 1 \geq z \geq 0. \tag{2.1}$$

Ce résultat a été démontré par Deheuvels (1991), dans le cas où A admet une dérivée seconde. Capéraà, Fougères et Genest (1996) utilisent (**??**) pour construire une estimation semi-paramétrique de A à partir d'un estimateur de G.

Proposition 2 *Si $(X,Y) \sim C$, alors la loi conjointe de $Z = \log X/\log(XY)$ et de $W = C(X,Y)$ est donnée par :*

$$\begin{aligned} F(z,w) &= (w - w\log w)\left\{z + z(1-z)\frac{A'(z)}{A(z)} - \int_0^z \frac{t(1-t)}{A(t)}dA'(t)\right\} \qquad (2.2) \\ &+ w\int_0^z \frac{t(1-t)}{A(t)}dA'(t) \qquad 1 \geq z, w \geq 0. \end{aligned}$$

Dans le cas de l'indépendance, on a $A(t) \equiv 1$ ainsi, $F(z,w) = z(w - w\log w)$. Les variables aléatoires Z et W sont alors indépendantes; Z suit un loi uniforme sur (0,1) et W est distribuée comme le produit de deux uniformes indépendantes sur $(0,1)$. Lorsque A admet une dérivée seconde dans $(0,1)$, F est absolument continue. Par contre, lorsque A' a des sauts, ces sauts correspondent à des singularités dans la loi de Z. Par exemple, lorsque $A(t) = \max(t, 1-t)$, Z prend la valeur 1/2 avec probabilité 1 et la loi de W est uniforme sur $(0,1)$.

Pour une copule de valeurs extrêmes quelconque, la loi marginale de $C(X,Y)$, s'obtient de (**??**) en posant $z = 1$. On a

$$P(C(X,Y) \leq w) = K_\tau(w) = w - (1-\tau)w\log w,$$

où

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)}dA'(t). \qquad (2.3)$$

L'expression simple obtenue pour K_τ permet de calculer facilement le tau de Kendall, $\tau(X,Y)$, une mesure de la dépendance entre deux variables aléatoires qui ne dépend pas des marges. On a

$$\tau(X,Y) = 4E\left\{C(X,Y)\right\} - 1 = \tau,$$

où τ est défini par (**??**). Cette formule permet de calculer le tau de Kendall d'une copule de valeurs extrêmes.

Considérons maintenant quelques exemples. Les familles de fonctions de répartition étudiées dans ce travail, de même que les fonctions A associées sont données au Tableau 1. Les fonctions de répartition de Gumbel de type A et B sont répertoriées par Johnson et Kotz (1972), voir aussi Hutchinson et Lai (1990). La famille de Galambos apparaît dans Galambos (1975) alors que la famille MO généralise une famille introduite par Marshall et Olkin (1967), voir aussi Nelsen (1991). La distribution de Gumbel de type B et la distribution de Galambos sont parfois appelées respectivement loi logistique et loi logistique négative (Tawn, 1988). Le tau de Kendall, évalué à l'aide de (**??**), de même que la fonction de répartition marginale de Z et sa densité sont donnés au Tableau 2, pour les quatre familles du Tableau 1.

Pour la famille de Gumbel de type B, des calculs élémentaires montrent que les variables W et Z de la Proposition 2 sont indépendantes. On aurait pu obtenir ce résultat à partir de la caractérisation des copules archimédiennes donnée par Genest et Rivest (1993), voir aussi Genest et McKay (1986). La famille de Gumbel de type A ne permet pas d'obtenir une dépendance élevée entre X et Y. La valeur maximale du tau de Kendall pour cette famille, tel que présenté au Tableau 2, est de $8 \cdot 3^{-1/2} \cdot \arctan(3^{-1/2}) - 2 \approx 0.4184$. Notons que lorsque $\theta = 1$, la copule de Galambos est égale à la copule de Gumbel de type A, avec $\theta = 1$. La formule pour le tau de Kendall présentée au tableau 2 pour la famille de Galambos est obtenue en faisant le changement de variable, $s = t^\theta/[t^\theta + (1-t)^\theta]$ dans (**??**). Dans la famille MO, la dérivée de A est constante sauf pour un saut de $\theta + \beta$ à $t = \beta/(\beta + \theta)$. La distribution F est singulière. On peut facilement montrer que $\Pr\{Z = \beta/(\beta+\theta)\} = \theta\beta/(\theta + \beta - \theta\beta)$.

Notons finalement que la Proposition 2 ne semble pas se généraliser facilement à p dimensions. En effet, nous n'avons pu trouver une forme simple pour la loi conjointe de $(\log X_j)/(\log X_1 + ... + \log X_p)$, j=1,...,p et de $C(X_1, ..., X_p)$ lorsque $(X_1, ..., X_p)$ est distribué selon une copule de valeurs extrêmes $C(x_1, ..., x_p)$.

3 Simulation de valeurs extrêmes bidimensionnelles

De nombreux articles étudient l'estimation paramétrique (Tawn, 1988; Joe, 1994) ou semi-paramétrique (Deheuvels, 1991; Capéraà et coll., 1996) de copules de valeurs extrêmes. Pour étudier les méthodes d'estimation par simulation, il est utile de pouvoir générer des couples de variables aléatoires extrêmes. Cette section montre comment utiliser la Proposition **??** pour simuler un couple de variables aléatoires (X, Y) selon une distribution $C(x, y)$ dans la classe (**??**). Si on sait simuler (X, Y) selon la loi $C(x, y)$, et si F_1 et F_2 sont des lois de valeurs extrêmes unidimensionnelles, alors le couple $(F_1^{-1}(X), F_2^{-1}(Y))$ est distribué selon une loi de valeurs extrêmes bidimensionnelle. L'algorithme proposé ici permet ainsi de simuler une loi de valeurs extrêmes bidimensionnelle.

Pour simplifier la présentation, supposons que A admet une dérivée seconde et donc que la distribution F est absolument continue. Dans ce cas, Z est absolument continue et a une densité $g_Z(z)$ que l'on peut calculer en dérivant (**??**). La distribution conditionnelle de W étant donné Z peut alors s'écrire

$$F(w|z) = \frac{1}{g_Z(z)} \frac{d}{dz} F(z, w) =$$
$$w \left\{ \frac{z(1-z)A''(z)}{A(z)g_Z(z)} \right\} + (w - w \log w) \left\{ 1 - \frac{z(1-z)A''(z)}{A(z)g_Z(z)} \right\}.$$

Etant donné Z, la loi de W est donc uniforme sur $(0, 1)$ avec probabilité $p(Z)$ et égale à la loi du produit de deux uniformes indépendantes sur $(0, 1)$ avec probabilité $1 - p(Z)$, où

$$p(z) = \frac{z(1-z)A''(z)}{A(z)g_Z(z)}.$$

Notons que puisque $g_Z(z)$ est la dérivée de (**??**), $0 \leq p(z) \leq 1$.

Dans la discussion qui suit, U_0, U_1, U_2 et U_3 représentent des variables aléatoires indépendantes uniformément distribuées sur $(0, 1)$. Pour simuler un couple de variables aléatoires (X, Y) selon la loi $C(x, y)$ donnée par (**??**), avec fonction de dépendance A, on peut utiliser la procédure suivante:

1. Simuler Z selon la loi $G_Z(z)$ donnée par (**??**).

2. Sachant Z, prendre $W = U_1$ avec probabilité $p(Z)$ et $W = U_1 U_2$ avec probabilité $1 - p(Z)$.

3. Poser $X = W^{Z/A(Z)}$ et $Y = W^{(1-Z)/A(Z)}$.

Considérons quelques exemples. Pour la copule de Gumbel de type B, d'après le Tableau 2, $G_Z(z) = m_r(z)$. Puisque $m_r^{-1}(t) = m_{1/r}(t)$, à l'étape 1 de l'algorithme il suffit de prendre $Z = m_{1/r}(U_0)$. De plus à l'étape 2, $p(Z) \equiv 1 - 1/r$. Cet algorithme est équivalent à celui obtenu en écrivant C comme une copule archimédienne (Genest et Rivest, 1993).

Lorsque la distribution $G_Z(z)$ n'est pas inversible on suggère, pour simuler Z, d'utiliser la méthode du rejet (Ross, 1985 p. 438). Cette méthode permet de simuler Z à partir de Z_1, une variable aléatoire de densité $g_1(z)$ que l'on sait simuler et telle que $g_Z(z)/g_1(z)$ soit borné par une constante c sur $(0, 1)$. La méthode du rejet utilise le résultat suivant: étant donné que $U_0 \leq g_Z(Z_1)/[cg_1(Z_1)]$, la loi conditionnelle de Z_1 est $G_Z(z)$. Lorsque c est voisin de 1, cet algorithme permet de simuler G_Z efficacement; par contre, lorsque c est grand, la condition $U_0 \leq g_Z(Z_1)/[cg_1(Z_1)]$ est rarement satisfaite et l'algorithme est peu performant.

Lorsque la dépendance entre X et Y est faible, on peut, pour simuler Z, utiliser l'algorithme du rejet en prenant $g_1(z) = 1$ pour $z \in (0, 1)$. Pour la copule de Gumbel de type A, on a $g_Z(z) \leq g_Z(1/2) = (4 + \theta)/(4 - \theta)$, tandis que pour la famille de Galambos, de patients calculs montrent que $g_Z(z) \leq g_Z(1/2) = 1 + (\theta + 1)/(2^{1+1/\theta} - 1)$. Dans l'algorithme du rejet on peut donc prendre $c = g_Z(1/2)$ pour ces deux familles. La borne c varie entre 1 et 5/3 lorsque θ va de 0 à 1 dans les deux familles. Cependant, lorsque θ augmente au delà de 1, c devient rapidement très grande pour la famille de Galambos. L'algorithme de simulation est alors inefficace.

Dans des cas de forte dépendance, la loi de Z est concentrée autour de 1/2. Le maximum c de $g_Z(z)$ est alors élevé, il faut donc modifier l'algorithme précédent pour le rendre plus efficace. La transformation $m_\gamma(Z)$, définie au Tableau 2, donne, avec $\gamma > 1$, une variable aléatoire moins concentrée autour de 1/2 que Z. On peut donc générer, à partir d'une loi uniforme, la variable $T = m_\gamma(Z)$ pour un choix judicieux de γ et prendre $Z = m_{1/\gamma}(T)$. Pour la famille de Galambos $\gamma = \theta$ est un bon choix. La loi de $T = m_\theta(Z)$ est alors donnée par $H_T(t)$ qui est définie au Tableau 2. On peut montrer que sa densité satisfait $h_T(t) \leq h_T(1/2) = 1/\theta + (\theta + 1)/[\theta(2^{1+1/\theta} - 1)]$. La valeur de $c = h(1/2)$ décroît de 5/3 à 1 lorsque θ va de 1 à ∞. La faible valeur de c montre que l'utilisation de la transformation $m_\theta(\cdot)$ donne un algorithme de simulation performant pour la famille Galambos. Khoudraji (1995) donne une variante de cet algorithme ayant une borne c un peu plus faible.

Pour simuler des copules de la famille MO, on peut adapter l'algorithme présenté ici au cas où la loi de Z est singulière (Khoudraji, 1995). Cependant la forme simple de C pour cette copule permet la construction d'une méthode directe de simulation.

En effet, on montre facilement que la loi conjointe de $X = \max(U_1^{1/(1-\theta)}, U_3^{1/\theta})$ et de $Y = \max(U_2^{1/(1-\beta)}, U_3^{1/\beta})$ est bien la copule MO de paramètres (θ, β). Ce résultat a une généralisation intéressante: si (X_0, Y_0) et (X_1, Y_1) sont distribués selon les copules de valeurs extrêmes C_0 et C_1, alors, la loi conjointe de $\max(X_0^{1/(1-\theta)}, X_1^{1/\theta})$ et $\max(Y_0^{1/(1-\beta)}, Y_1^{1/\beta})$ est également une copule de valeurs extrêmes. Cette observation permet, entre autres, de générer facilement, à partir des algorithmes pour la Gumbel de type B et la Galambos, des observations de versions asymétriques des lois logistique et logistique négative (Tawn, 1990).

Pour certaines lois de valeurs extrêmes apparaissant dans la littérature, telles les modèles bilogistique et bilogistique négatif, la fonction A dans (**??**) n'a pas une forme explicite. L'application de l'algorithme de simulation proposé ici à ces modèles est problématique. Elle nécessite des travaux supplémentaires.

4 Tests d'ajustement pour des copules de valeurs extrêmes bidimensionnelles

Soit $\{(X_i, Y_i) : i = 1, ..., n\}$ un échantillon bivarié d'une loi H. Pour savoir si H appartient à la famille des distributions de valeurs extrêmes bidimensionnelles, on peut d'abord vérifier que les marges sont des lois de valeurs extrêmes à l'aide de tests d'ajustement univariés (Coronel-Brizio, Lockhart et Stephens, 1996). Toutefois, il faut également s'assurer que la copule de H appartient à la famille (**??**). Cette section suggère une statistique qui permet de faire cette vérification et précise sa loi asymptotique.

A la Section 2, on a montré que pour toute copule de valeurs extrêmes C, la loi de $W = C(X, Y)$ est donnée par $K_\tau(w) = w - (1-\tau)w \log w$. Nous formulons la conjecture que la famille K_τ caractérise (**??**). En d'autres termes si la loi conjointe de X et de Y est $C(x, y)$ et si la loi de $C(X, Y)$ est égale à $K_\tau(w)$ pour un certain τ dans $(0, 1)$, alors, si la conjecture est vraie, C est une copule de valeurs extrêmes. Genest et Rivest (1989) présentent une preuve partielle de ce résultat, pour les copules archimédiennes. En effet, la seule copule archimédienne qui soit une copule de valeurs extrêmes est la copule de Gumbel de type B, car les copules archimédiennes sont caractérisées par la loi de $C(X, Y)$ (Genest et Rivest, 1993).

Dans la mesure où la conjecture précédente est vraie, on peut tester si une copule appartient à la famille (**??**), en considérant l'hypothèse $H_0 : K(w) = K_\tau(w),\ \tau \in (0,1)$. On peut construire un test d'ajustement simple à l'aide des deux premiers moments de K_τ. Si W est distribuée selon $K_\tau(w)$, alors, quelque soit τ, $8E\{W\} - 9E\{W^2\} - 1 = 0$. Les moments de W peuvent facilement être estimés. En effet, si $\delta\{(x_1,y_1),(x_2,y_2)\} = 1$ si $x_1 \geq x_2$ et $y_1 \geq y_2$ et à 0 sinon, alors, si (i,j,k) sont des indices distincts, $E[\delta\{(X_i,Y_i),(X_j,Y_j)\}] = E\{W\}$ et

$$E[\delta\{(X_i,Y_i),(X_j,Y_j)\}\delta\{(X_i,Y_i),(X_k,Y_k)\}] = E\{W^2\}.$$

Ces résultats sont démontrés en calculant des espérances conditionnelles, étant donné (X_i,Y_i). Une estimation non biaisée de $8E\{W\} - 9E\{W^2\} - 1$ est donc donnée par la statistique suivante,

$$\begin{aligned} S_n = {} & \frac{8}{n(n-1)} \sum_{i \neq j}^{n} \delta\{(X_i,Y_i),(X_j,Y_j)\} \\ & - \frac{9}{n(n-1)(n-2)} \sum_{i \neq j \neq k}^{n} \delta\{(X_i,Y_i),(X_j,Y_j)\}\delta\{(X_i,Y_i),(X_k,Y_k)\} - 1. \end{aligned} \tag{4.1}$$

Pour étudier la distribution asymptotique et pour estimer la variance de S_n, il est utile de noter que cette dernière est U-statistique d'ordre 3, dont le noyau symétrique est donné par

$$\begin{aligned} \psi\{(x_1,y_1),(x_2,y_2),(x_3,y_3)\} = {} & \frac{8}{6}[\delta\{(x_1,y_1),(x_2,y_2)\} + \delta\{(x_1,y_1),(x_3,y_3)\} \\ & + \delta\{(x_2,y_2),(x_1,y_1)\} + \delta\{(x_2,y_2),(x_3,y_3)\} + \delta\{(x_3,y_3),(x_1,y_1)\} \\ & + \delta\{(x_3,y_3),(x_2,y_2)\}] - 3[\delta\{(x_1,y_1),(x_2,y_2)\}\delta\{(x_1,y_1),(x_3,y_3)\} \\ & + \delta\{(x_2,y_2),(x_1,y_1)\}\delta\{(x_2,y_2),(x_3,y_3)\} \\ & + \delta\{(x_3,y_3),(x_1,y_1)\}\delta\{(x_3,y_3),(x_2,y_2)\}] - 1. \end{aligned}$$

On écrit alors

$$S_n = \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} \psi\{(X_i,Y_i),(X_j,Y_j),(X_k,Y_k)\}.$$

Notons $\psi_1(X_i,Y_i) = E[\psi\{(X_i,Y_i),(X_j,Y_j),(X_k,Y_k)\}|(X_i,Y_i)]$. Posons $\mu = 8E(W) - 9E(W^2) - 1$ et appelons σ^2 la variance de $\psi_1(X,Y)$. La théorie asymptotique des

U-statistiques (voir Lee, 1990) implique que $\sqrt{n}(S_n - \mu)$ converge en distribution vers une normale de moyenne 0 et de variance $9\sigma^2$. En particulier, $\mu = 0$ si la copule C appartient à la classe des copules de valeurs extrêmes.

Pour estimer la variance de S_n, on propose l'utilisation de la méthode du jackknife (Lee, 1990, chapitre 5). Une formule relativement simple pour cet estimateur est donnée par

$$\hat{V}_{Jack}(S_n) = \frac{n-1}{n} \sum_{i=1}^{n} (S_n^{(-i)} - S_n)^2,$$

où $S_n^{(-i)}$ est la statistique S_n calculée sur un échantillon de taille $n-1$, obtenue en excluant la ième observation. $\hat{V}_{Jack}(S_n)$ est convergent (Lee, 1990, p. 221), de plus il a un léger biais positif dans des petits échantillons (Lee, 1990, p. 219). On peut donc espérer que le test statistique qui rejette H_0 lorsque $|S_n|/\sqrt{\hat{V}_{Jack}(S_n)} > z_{\alpha/2}$ sera conservateur dans de petits échantillons. Ici z_α dénote le quantile d'ordre $(1-\alpha)$ de la loi normale standard.

Pour les contre-hypothèses satisfaisant $\mu = 8E\{W\} - 9E\{W^2\} - 1 \neq 0$, le test proposé est convergent. On a la convergence pour plusieurs familles de copules. Considérons, par exemple, celle de Clayton (1978). Pour cette famille $K(v) = v + v(1 - v^\alpha)/\alpha$ (Genest and Rivest, 1993) et $8E\{W\} - 9E\{W^2\} - 1 = 2\alpha/((\alpha+2)(\alpha+3))$, ce qui est non-nul lorsque α est positif. En fait la copule de Clayton est archimédienne. Puisque les copules de cette classe sont caractérisées par leurs fonctions K, le test suggéré est, règle générale, convergent pour des contre-hypothèses archimédiennes.

Pour étudier le comportement de la statistique du test dans de petits échantillons, ses seuils exacts ont été calculés par simulation (10000 répétitions) pour quelques distributions. Les résultats apparaissent au Tableau 3. On peut constater que lorsque le tau de Kendall est faible, le seuil empirique est voisin du seuil nominal alors que, pour des taus de Kendall élevés, le seuil empirique est inférieur au seuil nominal ce qui fait que le test est conservateur. Notons qu'un autre estimateur de variance, obtenu à partir des résultats de Barbe, Genest, Ghoudi et Rémillard (1996), a été étudié par simulation. Dans de petits échantillons il donne un test libéral lorsque le tau de Kendall est faible. L'estimateur de variance donné par la méthode du jackknife nous semble préférable car il donne toujours un test conservateur.

Pour déterminer si les maxima annuels des niveaux de la mer dans le Lowestoft et le Sheerness, étudiés par Tawn (1988), suivent une distribution de valeurs extrêmes, Le test de cette section donne une statistique standardisée de $z_{obs} = -0.146$. L'hypothèse que la copule sous-jacente appartient à la classe (**??**) n'est pas rejetée. Il est donc acceptable de modéliser ces données à l'aide d'une copule de valeurs extrêmes.

Notons finalement que ce test se généralise facilement à l'étude de l'ajustement d'une copule de valeurs extrêmes à p composantes. En effet, le vecteur S_n des $p(p-1)/2$ statistiques (**??**), pour toutes les paires de composantes possibles, est une U-statistique multidimensionnelle. On peut utiliser la méthode du jackknife pour estimer sa matrice de variances-covariances $v(S_n)$. Sous l'hypothèse nulle, la loi de $S_n^t v(S_n)^{-1} S_n$ suit approximativement une loi du khi-deux à $p(p-1)/2$ degrés de liberté, ce qui permet de construire un test d'ajustement.

Annexe

Démonstration de la Proposition ??: Puisque la loi marginale de X est uniforme, $C_{10}(x,y)$, la dérivée partielle de $C(x,y)$ par rapport à x, est en fait la loi conditionnelle de Y étant donné que $X = x$. Cette loi est donnée par

$$C_{10}(x,y) = \frac{1}{x}\left\{A\left(\frac{\log x}{\log x + \log y}\right) + \frac{\log y}{\log x + \log y} A'\left(\frac{\log x}{\log x + \log y}\right)\right\} C(x,y).$$

En conditionnant par rapport à X, on obtient

$$G(x,z) = P(X \leq x, Z \leq z) = \int_0^x C_{10}(t, t^{\frac{1-z}{z}})dt.$$

On évalue facilement cette intégrale pour démontrer la Proposition 1. ♣

La démonstration de la Proposition **??** fait appel à la fonction q définie de $[0,1] \to [0,1]$ par $t \mapsto t/A(t)$. Notons que q est strictement croissante sur son support. En effet, sa dérivée à droite est donnée par

$$q'(t) = \frac{A(t) - tA'(t)}{A^2(t)}.$$

Cette dernière est toujours positive car $A(t) > t$, et $A'(t) < 1$ sur le support de q. Notons cependant que si $A(t) = t$ dans un intervalle du type $(1-\epsilon, 1)$, $q(t)$ est constante et égale à 1 dans cet intervalle.

Démonstration de la Proposition ??: En écrivant $W = C(X,Y) = X^{A(Z)/Z}$, on déduit les égalités suivantes,

$$\begin{aligned} F(z,w) &= P\left(Z \le z, X^{A(Z)/Z} \le w\right) \\ &= P\left(q(Z) \le \min\left(q(z), \frac{\log X}{\log w}\right)\right). \end{aligned} \tag{A.1}$$

Si $A(t) = t$ dans un intervalle du type $(1-\epsilon, 1)$, alors $\{q(Z) \le q(z)\} = \{Z \le z\}$ sauf lorsque $z \in (1-\epsilon, 1)$. Dans ce cas $\{q(Z) \le q(z)\} = \{1 - \epsilon < Z \le 1\}$. Cependant, on a alors $G_Z(1-\epsilon) = 1$ et (**??**) est toujours vraie. La fonction $q(t) = t/A(t)$ étant croissante sur son support, le minimum dans l'équation précédente est égal à $q(z)$ seulement si $X \le w^{z/A(z)}$. Ainsi l'équation (**??**) devient

$$\begin{aligned} F(z,w) &= \int_0^{w^{\frac{z}{A(z)}}} P(Z \le z | X = x) dx + \\ &\quad \int_{w^{\frac{z}{A(z)}}}^1 P\left\{Z \le q^{-1}\left(\frac{\log x}{\log w}\right) | X = x\right\} dx \\ &= G\left(w^{\frac{z}{A(z)}}, z\right) + \int_{w^{\frac{z}{A(z)}}}^1 G_{10}\left\{x, q^{-1}\left(\frac{\log x}{\log w}\right)\right\} dx, \end{aligned} \tag{A.2}$$

où $G_{10}(x,z)$ est la dérivée partielle par rapport à x de la fonction $G(x,z)$ définie à la Proposition 1 et $q^{-1}(t)$ est l'inverse de la fonction $q(t)$.

Le premier terme de (**??**) est facilement évalué. Pour évaluer le deuxième terme, on fait le changement de variable $t = q^{-1}(\log x / \log w)$. Puisque $x = w^{t/A(t)}$, le jacobien est donné par

$$dx = \log w \frac{A(t) - tA'(t)}{A(t)^2} w^{t/A(t)} dt.$$

Ainsi, sachant que

$$G_{10}(w^{t/A(t)}, t) = \{A(t) + (1-t)A'(t)\} w^{1 - \frac{t}{A(t)}},$$

le deuxième terme de (**??**) est égal à

$$\begin{aligned}
&\int_z^0 \{A(t)+(1-t)A'(t)\}\, w\log w \left\{\frac{A(t)-tA'(t)}{A^2(t)}\right\} dt \\
= \quad & -w\log w\left\{z+\int_0^z \frac{(1-2t)A'(t)}{A(t)}dt-\int_0^z t(1-t)\left(\frac{A'(t)}{A(t)}\right)^2 dt\right\}. \qquad \text{(A.3)}
\end{aligned}$$

En regroupant les équations (**??**) et (**??**), on conclut que $P\left(Z\le z, W\le w\right)$ s'écrit

$$\begin{aligned}
& w\left\{z+z(1-z)\frac{A'(z)}{A(z)}\right\} \\
& \quad -w\log w\int_z^0\left\{1+(1-2t)\frac{A'(t)}{A(t)}-t(1-t)\left(\frac{A'(t)}{A(t)}\right)^2\right\}dt \\
= \quad & w\left\{z+z(1-z)\frac{A'(z)}{A(z)}\right\} \\
& \quad +w\log w\left\{-z+\int_0^z \frac{t(1-t)}{A(t)}dA'(t)-z(1-z)\frac{A'(z)}{A(z)}\right\}. \qquad \text{(A.4)}
\end{aligned}$$

♣

Bibliographie

Anderson, C. W. et Nadarajah, S. (1993). Environmental factors affecting reservoir safety. Dans *Statistics for the Environment* Publié sous la direction de V. Barnett et de F. Turkman, Wiley New York, 163-182

Barbe, P., Genest, C., Ghoudi, K. et Rémillard, B. (1996). On Kendall's process. *Journal of Multivariate Analysis*, **58**, 197-229.

Capéraà, P., Fougères, A.-L. et Genest, C. (1996). A nonparametric estimation procedure for bivariate extreme value copulas. Soumis pour publication.

Clayton, D. G. (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.

Coronel-Brizio, H. F., Lockhart, R.A. et Stephens, M. A. (1996). Tests of fit for the generalized extreme value distribution. Soumis pour publication.

Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters*, **12**, 429-439.

Galambos, J. (1975). Order statistics from sample from multivariate distributions. *J. Amer. Statist. Assoc.*, **70**, 674-680.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics.* 2nd ed., Kreir, Melbourne, FL.

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

Genest, C. et Rivest, L.-P. (1989). A characterization of Gumbel's family of extreme value distributions. *Statistics and Probability Letters*, **8**, 207-211.

Genest, C. et Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, **88**, 1034-1043.

Geoffroy, J. (1958). Contribution à la théorie des valeurs extrêmes. *Publications de l'Institut Statistique de l'Université de Paris*, **7**, 37-185.

Hutchinson, T. P. et Lai, C. D. (1990). *Continuous Bivariate Distributions, Emphasizing Applications.* Adelaide: Rumsby Scientific Publishing.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics and Probability Letters*, **9**, 75-81.

Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *La revue canadienne de statistique.* **22**, 47-64.

Johnson, N. L. et Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley & Sons: New York

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation des copules aux valeurs extrêmes.* Thèse de doctorat. Université Laval.

Lee, H. A. (1990) *U-statistics. Theory and Practice.* Marcel Dekker: New York

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

Nelsen, R. B. (1991). Copulas and association. Dans *Advances in Probability Distributions with Given Marginals.* Publié sous la direction de G. Dall'Aglio, S. Kotz & G. Salinetti. Kluwer Academic Press, 51-74.

Pickands, J. (1981). Multivariate extreme value distributions. Dans *Proceedings of the 43rd Session of the International Statistical Institute (Buenos Aires)*, 859-878.

Ross, S. M. (1985). *Introduction to Stochastic Processes. Third Edition.* John Wiley: New York

Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, **11**, 195-210.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'institut de statistique de l'Université de Paris*, **8**, 229-231.

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

Tawn, J.A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245-253.

Tiago de Olivera, J. (1958). Extremal distributions. *Revista da Fac. Ciencas, Univ. Lisboa*, **A8**, 299-310.

Famille	$A(t)$	Domaine	$C(x,y)$
Gumbel B	$\{(1-t)^r + t^r\}^{1/r}$	$r \geq 1$	$\exp\left(-\{(-\log x)^r + (-\log y)^r\}^{1/r}\right)$
Gumbel A	$\theta t^2 - \theta t + 1$	$1 \geq \theta \geq 0$	$xy \exp\left(-\theta \frac{\log x \log y}{\log x + \log y}\right)$
Galambos	$1 - \left(t^{-\theta} + (1-t)^{-\theta}\right)^{-1/\theta}$	$\theta \geq 0$	$xy \exp\{(-\log x)^{-\theta} + (-\log y)^{-\theta}\}^{-1/\theta}$
MO	$\max\{1-\theta t, 1-\beta(1-t)\}$	$1 \geq \theta, \beta \geq 0$	$x^{1-\theta} y^{1-\beta} \min(x^\theta, y^\beta)$

Tableau 1. Quatre familles de lois de valeurs extrêmes.

Famille	τ	$G_Z(z)$	$g_Z(z)$
Gumbel B	$1-1/r$	$m_r(z)^a$	$r\dfrac{z^{r-1}(1-z)^{r-1}}{[z^r+(1-z)^r]^2}$
Gumbel A	$\dfrac{8\arctan\left\{\sqrt{\theta/(4-\theta)}\right\}}{\sqrt{\theta(4-\theta)}}-2$	$1-z+\dfrac{2z-1}{\theta z^2-\theta z+1}$	$\dfrac{2-\theta(2z^2-2z+1)}{(\theta z^2-\theta z+1)^2}-1$
Galambos	$\tau(\theta)^b$	$H_T\left[m_\theta(z)\right]\ ^c$	$\theta\dfrac{(1-z)^{\theta-1}}{z^{\theta+1}}m_\theta(z)^2 h_T\left[m_\theta(z)\right]\ ^d$
MO	$\dfrac{\theta\beta}{\theta+\beta-\theta\beta}$	fonction discontinue	pas définie

Tableau 2. Tau de Kendall, fonction de répartition G_Z et sa densité g_Z pour quatres familles de lois de valeurs extrêmes.

^{a}où $m_r(z)=z^r/\{z^r+(1-z)^r\}$

^{b}où $\tau(\theta)=\dfrac{\theta+1}{\theta}\displaystyle\int_0^1\left\{\frac{1}{s^{1/\theta}}+\frac{1}{(1-s)^{1/\theta}}-1\right\}^{-1}ds$

^{c}où $H_T(t)=1-\dfrac{t^{-1/\theta}-t}{(1-t)^{-1/\theta}+t^{-1/\theta}-1}$

^{d}où $h_T(t)=\dfrac{d}{dt}H_T(t)=\dfrac{[1/t^{1+1/\theta}-1][1/(1-t)^{1+1/\theta}-1]}{\theta[1/t^{1/\theta}+1/(1-t)^{1/\theta}-1]^2}+\dfrac{\theta+1}{\theta[1/t^{1/\theta}+1/(1-t)^{1/\theta}-1]}$

Famille	τ	$n = 20$	$n = 30$	$n = 50$	$n = 100$
Gumbel B	0.25	0.0435	0.0495	0.0436	0.0455
	0.50	0.0376	0.0385	0.0401	0.0465
	0.75	0.0170	0.0210	0.0259	0.0336
Galambos	0.25	0.0460	0.0443	0.0468	0.0471
	0.50	0.0347	0.0363	0.0405	0.0490
	0.75	0.0199	0.0213	0.0282	0.0378
MO	0.25	0.0454	0.0486	0.0485	0.0468
	0.50	0.0483	0.0507	0.0484	0.0464
	0.75	0.0349	0.0372	0.0475	0.0484

Tableau 3: Seuil empirique du test d'ajustement avec un seuil nominal de 5%.