REVUE DE STATISTIQUE APPLIQUÉE

YOUNÈS HILLALI

# Test d'ajustement d'une loi bidimensionnelle : application à des données climatologiques

*Revue de statistique appliquée*, tome 49, nº 1 (2001), p. 79-95
<http://www.numdam.org/item?id=RSA_2001__49_1_79_0>

© Société française de statistique, 2001, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
*Article numérisé dans le cadre du programme*
*Numérisation de documents anciens mathématiques*
http://www.numdam.org/

*Rev. Statistique Appliquée*, 2001, **XLIX** (1), 79-95

# TEST D'AJUSTEMENT D'UNE LOI BIDIMENSIONNELLE : APPLICATION À DES DONNÉES CLIMATOLOGIQUES

Younès Hillali (*) (**)

*(*) INRIA Rocquencourt,*
*Domaine de Voluceau, B.P.105,*
*78153 Le Chesnay cedex, France*
*(**) Université de Paris-IX Dauphine, Ceremade,*
*Place du Maréchal De Lattre de Tassigny, 75775 Paris cedex 16, France*
*e-mail : Younes.Hillali@inria.fr*

## RÉSUMÉ

Si $F$ est la fonction de répartition d'une loi multivariée, de fonctions de répartition marginales $F_i, i = 1, ..., n$, il existe une application $H$, appelée copule ou fonction de dépendance, définie sur $[0,1]^n$ à valeurs dans $[0,1]$, telle que $F(x_1, ..., x_n) = H(F_1(x_1), ..., F_n(x_n))$. Cette application définit une distribution multivariée de marges uniformes sur $[0,1]$. Lorsque les marges sont continues, la fonction $F$ peut être représentée de manière unique par une copule $H$. Plusieurs familles de copules ont été présentées ces dernières années. Dans cet article, nous présentons une procédure d'ajustement d'une loi bidimensionnelle par une famille particulière de copules, dites archimédiennes, basée sur un résultat du à Genest *et al* [9]. Nous verrons que l'ajustement d'une fonction de répartition bidimensionnelle se résume à celui d'une fonction de répartition unidimensionnelle d'une pseudo-variable aléatoire $U = F(X_1, X_2)$ en comparant les versions paramétrique et non-paramétrique de sa distribution. Pour illustrer cette méthode nous présentons une application à des données climatologiques.

***Mots-clés :*** *Copules Archimédiennes, estimation non-paramétrique, lois bidimensionnelles, tau de Kendall.*

## ABSTRACT

If $F$ is a continuous $n$-variate cumulative distribution function with univariate margins $F_i, i = 1, ..., n$, then there exists a function $H$, called a copula or dependence function, from the unit n-cub to the unit interval such that $F(x_1, ..., x_n) = H(F_1(x_1), ..., F_n(x_n))$. The function $H$ defines a multivariate distribution function with uniform $(0,1)$ margins. If the marginal distributions are continuous, then there is a unique copula representation. Several families of copulas have been developed in recent years; but the choice of an appropriate copula which provides a suitable representation of the dependence structure between random variables, without any knowledge of its explicit form, has not been widely studied. In this paper, we present a procedure for fitting a special class of copulas, called Archimedean, to the bivariate distribution. This procedure is based on the result developed by Genest and Rivest [9]. Using

their result, we examine the degree of closeness of the parametric and non-parametric versions of the distribution function of the pseudo-observations $U = F(X_1, X_2)$. To illustrate this procedure, a climatic data set is analyzed.

***Keywords :*** *Archimedean copulas, non-parametric estimation, bivariate distributions, Kendall's tau.*

## 1. Introduction

Les méthodes de construction de lois multidimensionnelles à marges unidimensionnelles ou multidimensionnelles fixées ont été largement développées ces dernières années. Leur objectif est de diversifier le choix de distributions possibles pour la modélisation ou la simulation de données multidimensionnelles.

Citons d'une part, les travaux de Cuadras [5], qui permettent de générer des échantillons suivant une loi multidimensionnelle à marges multidimensionnelles données à partir d'une combinaison des lois marginales, pondérée par des paramètres liés à des coefficients d'association de type Kendall ou Spearman entre les variables aléatoires qui interviennent dans la modélisation. Cette méthode repose sur une connaissance préalable de la structure de la dépendance entre les variables aléatoires (matrice de corrélation et d'inter-corrélation). Nous trouvons d'autre part les travaux de Kimeldorf *et al* [17], Schweizer [19], Genest *et al* ([7], [8], [9]), Joe [16], Chakak et Koehler [3], Hillali [14], etc.

Ces auteurs proposent des stratégies de constructions de lois multidimensionnelles à marges unidimensionnelles basées sur le concept de copule développé par Sklar [22], que nous présenterons ci-dessous. Ces stratégies, comparées à celle de Cuadras [5], ne peuvent pas s'étendre au cas où les marges sont multidimensionnelles, mais ce cas ne fera pas l'objet de notre travail. Plus précisément, d'après une résultat de Genest *et al* [11], il est impossible de construire des lois à marges multidimensionnelles données à partir des copules.

Dans ce travail, nous nous limitons à la seconde stratégie de construction de lois multidimensionnelles basée sur les copules. En effet, d'après un résultat de Sklar [22], toute distribution de probabilité $F$ dont les marges $F_1$, ..., $F_n$ sont continues correspond de manière unique à une fonction de répartition $H$ dont la masse est concentrée sur le rectangle $[0,1]^n$ et dont les marges sont uniformes. Cette fonction de répartition est définie par :

$$H(x_1, ..., x_n) = F(F_1^{-1}(x_1), ..., F_n^{-1}(x_n)) \tag{1}$$

Dans cet article, nous nous intéressons à une méthode d'ajustement d'une loi bidimensionnelle possédant une représentation par une classe particulière de copules, à savoir les copules archimédiennes. Cette famille de copules représente une généralisation de plusieurs familles de lois bidimensionnelles connues. Nous examinons les propriétés élémentaires de cette famille de lois et nous présentons un résultat de base du à Genest et Rivest [9] qui nous permettra d'exposer clairement la procédure d'ajustement qui fera l'objet principal de ce travail.

Cet article se présente de la manière suivante : à la section 2, nous présentons une famille particulière de lois bidimensionnelles définies à partir des copules et nous donnons ses propriétés mathématiques. À la section 3 nous décrivons une procédure d'ajustement d'une telle famille de lois bidimensionnelles. À la section 4, nous présentons une application de cette méthode à des données climatologiques, dont l'objectif est de prévoir le risque engendré par les valeurs extrêmes de variables climatologiques sur le fonctionnement de réseaux de distributions par fibres optiques (RDFO), expérimentés au centre de recherche du CNET à Lannion. À la section 5 nous donnons les résultats obtenus, qui sont basés sur des tests d'ajustement graphique de nos données théoriques par rapport à quatre modèles de lois bidimensionnelles que nous décrivons par la suite. Nous terminons par des commentaires concernant les quatre modèles choisis et les difficultés de cette approche.

## 2. Copules archimédiennes bidimensionnelles

Soient $X$ et $Y$ deux variables aléatoires réelles de fonctions de répartition continues $F_1$ et $F_2$. Soit $K$ une fonction strictement croissante et continue, définie sur $[0,1]$ à valeurs dans $[0,1]$ telle que $K(0) = 0$ et $K(1) = 1$, vérifiant la relation suivante :

$$K(Pr\{X \leqslant x, Y \leqslant y\}) = K(Pr\{X \leqslant x\})K(Pr\{Y \leqslant y\}) \tag{2}$$

pour tous $0 \leqslant x, y \leqslant 1$. Si $K(t) = t$, les variables $X$ et $Y$ sont alors indépendantes au sens où on l'entend habituellement.

En posant $\Phi(t) = -\log(K(t))$ pour $0 < t \leqslant 1$, l'équation (2) devient :

$$\Phi(F(x,y)) = \Phi(F_1(x)) + \Phi(F_2(y)) = \Phi(H_\Phi(F_1(x), F_2(y))) \tag{3}$$

où $F(x,y)$ est la fonction de répartition du couple $(X,Y)$ et $H_\Phi$ une application définie sur $[0,1]^2$ à valeurs dans $[0,1]$, appelée copule *archimédienne*. L'équation (3) représente une généralisation naturelle d'un certain nombre de familles lois bidimensionnelles bien connues (voir Ali *et al* [1]).

### *2.1. Propriétés élémentaires*

Soient $X$ et $Y$ deux variables aléatoires continues de lois marginales $F_1$ et $F_2$ respectivement et de loi jointe $F$. Soit $\Psi$ la classe des applications $\Phi : [0,1] \longrightarrow [0,\infty]$ continues, strictement décroissantes et convexes, pour lesquelles $\Phi(1) = 0$ et $\Phi(0) = \infty$.

Toute fonction $\Phi \in \Psi$ possède une fonction inverse $\Phi^{-1} : [0,\infty] \longrightarrow [0,1]$ ayant les mêmes propriétés que $\Phi$, sauf que $\Phi^{-1}(0) = 1$ et $\Phi^{-1}(\infty) = 0$.

Considérons les fonctions définies sur le rectangle $[0,1]^2$ par :

$$H_\Phi(F_1(x), F_2(y)) = \Phi^{-1}[\Phi(F_1(x)) + \Phi(F_2(y))], \tag{4}$$

où $\Phi$ appartient à $\Psi$ et $H_\Phi(F_1(x), F_2(y)) = 0$ lorsque $\Phi(F_1(x)) + \Phi(F_2(y)) \geqslant \Phi(0)$.

D'après un résultat de Schweizer et Sklar [20], l'application $H_\Phi$ de type (4) est une fonction de répartition sur le carré unité si et seulement si $\Phi$ appartient à la classe $\Psi$. Cette application est unique si et seulement si $F_1$ et $F_2$ sont continues.

Il est intéressant de noter que dans la définition de l'application $\Phi$ il n'est pas nécessaire d'avoir $\Phi(0) = \infty$ pour que $\Phi$ génère une copule archimédienne de type (4) (voir famille de Genest et Ghoudi [10]). Dans ce cas, la copule archimédienne $H_\Phi$ de générateur $\Phi$ est définie par :

$$H_\Phi(u,v) = \Phi^{(-1)}[\Phi(u) + \Phi(v)], 0 \leqslant u, v \leqslant 1$$

où $\Phi^{(-1)}(t) = \Phi^{-1}(t)$ pour $t \in [0, \Phi(0)]$ et $\Phi^{(-1)}(t) = 0$ pour $t \in ]\Phi(0), \infty[$.

Pour des familles de copules $H_\Phi$ engendrées par des applications $\Phi \in \Psi$ dont les applications réciproques correspondant à $\Phi$ et à sa dérivée $\Phi'$ épousent une forme assez simple, il existe une correspondance naturelle entre les paramètres de ces familles et au moins trois mesures d'association entre entre les variables aléatoires $X$ et $Y$ de loi $H_\Phi$ (voir Schweizer et Wolf [21]), à savoir :

(i) le coefficient de corrélation de Spearman :

$$\rho(H_\Phi) = 12 \int_{[0,1]^2} xy \, dH_\Phi(x,y) - 3; \tag{5}$$

(ii) le coefficient de corrélation de Kendall :

$$\tau(H_\Phi) = 4 \int_{[0,1]^2} H_\Phi(x,y) \, dH_\Phi(x,y) - 1; \tag{6}$$

(iii) le coefficient de corrélation médiale de Blomqvist :

$$\beta(H_\Phi) = 4H_\Phi(\frac{1}{2}, \frac{1}{2}) - 1. \tag{7}$$

### *2.2. Exemples de lois bidimensionnelles engendrées par des copules archimédiennes*

Pour faciliter la compréhension de notre exposé, nous présentons quelques exemples de familles de lois qui sont de la forme (4).

**Famille de Gumbel [12]**

Cette famille, notée $G_\alpha$, est définie par :

$$|\log G_\alpha(x,y)|^{\alpha+1} = |\log F_1(x)|^{\alpha+1} + |\log F_2(y)|^{\alpha+1}$$

où $\alpha$ représente un coefficient d'association entre les deux variables aléatoires intervenant dans la modélisation. Pour que $G_\alpha$ soit une fonction de répartition, il faut et il suffit d'avoir $\alpha \geqslant 0$. L'application $\Phi_\alpha$ associée à cette famille est donnée par : $\Phi_\alpha(t) = |\log(t)|^{\alpha+1}$, pour $0 < t \leqslant 1$.

Le coefficient de corrélation de Kendall mesurant la liaison entre deux aléas $X$ et $Y$ de loi jointe $G_\alpha$ est donné par : $\tau(G_\alpha) = \dfrac{\alpha}{\alpha+1}$ (il suffit de remplacer $H_\Phi$ par $G_\alpha$ dans l'équation 6).

**Famille de Genest et Ghoudi [10]**

Cette famille est indicée par un paramètre réel $\alpha \in [0,1]$. Pour $\alpha > 0$ la fonction de répartition bidimensionnelle associée est définie pour tout $x, y \in [0,1]$ par :

$$H_\alpha(x,y) = \max[0, 1 - \{(1-x^\alpha)^{1/\alpha} + (1-y^\alpha)^{1/\alpha}\}^\alpha]^{1/\alpha}$$

L'application $\Phi_\alpha$ associée à cette famille est donnée par : $\Phi_\alpha(t) = (1-t^\alpha)^{1/\alpha}$.

De la même manière il existe une correspondance naturelle entre le paramètre $\alpha$ de cette famille et le coefficient de corrélation de Kendall entre deux variables aléatoires de loi $H_\alpha$ : $\tau(H_\alpha) = \dfrac{3\alpha-2}{\alpha-2}$

**Famille de Frank [6]**

La famille de Frank est définie pour tous $x, y \in [0,1]$ et $\alpha > 0$ par :

$$F_\alpha(x,y) = \log_\alpha[1 + \frac{(\alpha^x - 1)(\alpha^y - 1)}{(\alpha - 1)}]$$

On obtient la fonction de répartition $F_\alpha$ en posant $\Phi_\alpha(t) = -\log[\dfrac{1-\alpha^t}{1-\alpha}]$ dans l'équation (4).

**Famille de Clayton [4]**

Cette famille est engendrée par une application $\Phi_\alpha \in \Psi$ définie pour $\alpha > 0$ et $t \in [0,1]$ par : $\Phi_\alpha(t) = \dfrac{(t^{-\alpha}-1)}{\alpha}$. Sa forme générale est donnée par : $C_\alpha(x,y) = (x^{-\alpha} + y^{-\alpha} - 1)^{-1/\alpha}$.

## 3. Ajustement d'une loi bidimensionnelle par une copule archimédienne

Dans le paragraphe précédent, nous avons vu que toute loi de probabilité bidimensionnelle $F$ de deux variables aléatoires $X$ et $Y$ dont les marges $F_1$ et $F_2$ sont continues correspond de manière unique à une copule. Dans cette partie, nous

supposons que la copule associée à $F$ est une copule archimédienne, engendrée par une application $\Phi_\alpha$ appartenant à la classe $\Psi$ :

$$F(x,y) = H_{\Phi_\alpha}(F_1(x), F_2(y))$$

Le choix de la copule $H_{\Phi_\alpha}$ n'est pas arbitraire; il doit ajuster au mieux la fonction de répartition empirique bidimensionnelle associée à $F$.

Puisque l'estimation de $H_{\Phi_\alpha}$ est indépendante des lois marginales et pour simplifier notre exposé, nous allons nous restreindre au cas de deux variables aléatoires uniformes sur l'intervalle $[0,1]$. Dans ce cas, l'équation (4) devient :

$$H_{\Phi_\alpha}(x,y) = \Phi_\alpha^{-1}[\Phi_\alpha(x) + \Phi_\alpha(y)]. \tag{8}$$

Cette hypothèse n'est pas restrictive, il suffit d'utiliser le principe de translation de Nataf [18] pour se ramener à des distributions de probabilité quelconques.

Avant de détailler la procédure d'estimation de $H_{\Phi_\alpha}$, nous présentons un résultat de Genest et Rivest [9] qui nous sera utile par la suite.

**Proposition (Genest et Rivest [9]).** — *Soient $X$ et $Y$ deux variables aléatoires uniformes de copule archimédienne $H_{\Phi_\alpha}$ engendrée par une application $\Phi_\alpha$ appartenant à $\Psi$.*

*Cette copule est définie par :* $H_{\Phi_\alpha}(x,y) = \Phi_\alpha^{-1}(\Phi_\alpha(x) + \Phi_\alpha(y))$

*Soient*

$$V = \frac{\Phi_\alpha(X)}{\Phi_\alpha(X) + \Phi_\alpha(Y)}, U = H_{\Phi_\alpha}(X,Y) \text{ et } \lambda_{\Phi_\alpha}(u) = \frac{\Phi_\alpha(u)}{\Phi'_\alpha(u)}, 0 < u \leqslant 1.$$

*Alors :*

**1)** *$V$ est de loi uniforme sur* $[0,1]$,

**2)** *$U$ est de loi $S_{\Phi_\alpha}(u) = u - \lambda_{\Phi_\alpha}(u)$ sur* $[0,1]$,

**3)** *$U$ et $V$ sont indépendantes.*

L'un des résultats, que l'on peut tirer de cette proposition est que l'estimation de l'application $H_{\Phi_\alpha}$ se résume à celle d'une loi unidimensionnelle sur l'intervalle $[0,1]$ notée $S_{\Phi_\alpha}$ définie par : $S_{\Phi_\alpha}(u) = P\{H_{\Phi_\alpha}(x,y) \leqslant u\}$.

Soient $(x_1, ..., x_n)$ et $(y_1, ..., y_n)$ deux réalisations de $X$ et $Y$ de tailles $n$ de loi jointe $H_{\Phi_\alpha}$ de type (8). Soit $U_i$ la fonction de répartition empirique du couple d'observations $(x_i, y_i)$, définie par :

$$U_i = \frac{1}{n-1}\#\{(x_j, y_j) : x_j \leqslant x_i, y_j \leqslant y_i\}, 1 \leqslant i \leqslant n$$

Le choix de la division par $(n-1)$ dans la définition de $U_i$ est justifié par le fait qu'il est plus souhaitable d'obtenir un estimateur à valeurs dans $]0,1[$.

Un estimateur non-paramétrique de $S_{\Phi_\alpha}(U)$, noté $S_n(U)$, est donné par :

$$S_n(U) = \frac{1}{n}\sum_{i=1}^{n} \delta(U - U_i)$$

où $\delta(t)$ est la fonction de distribution de la masse du point à l'origine, égale à 1 pour $t \geqslant 0$ et à 0 sinon. D'après la proposition précédente, la pseudo-variable aléatoire $U = H_{\Phi_\alpha}(X, Y)$ est distribuée suivant la loi $S_{\Phi_\alpha}$ définie par :

$$S_{\Phi_\alpha}(u) = u - \lambda_{\Phi_\alpha}(u)$$

La fonction $\lambda_{\Phi_\alpha}$ est liée à l'application génératrice $\Phi_\alpha$ de $H_{\Phi_\alpha}$ par la relation :

$$\lambda_{\Phi_\alpha}(u) = \frac{\Phi_\alpha(u)}{\Phi'_\alpha(u)}, 0 < u \leqslant 1 \tag{9}$$

Il est donc plus intéressant d'estimer $\lambda_{\Phi_\alpha}(u)$ et de comparer l'estimateur obtenu à la fonction $\lambda_{\Phi_\alpha}(u)$ théorique qui est égale à $\dfrac{\Phi_\alpha(u)}{\Phi'_\alpha(u)}$. Puisque $S_n(u)$ est un estimateur de $S_{\Phi_\alpha}(u)$, alors un estimateur non-paramétrique de $\lambda_{\Phi_\alpha}$ est donné par :

$$\lambda_n(u) = u - S_n(u) \tag{10}$$

Connaissant un estimateur non-paramétrique $\lambda_n$ de $\lambda_{\Phi_\alpha}$, il faut le tester par rapport à la fonction $\lambda_{\Phi_\alpha}$ théorique. Or, cette dernière dépend du paramètre inconnu $\alpha$ de la fonction génératrice $\Phi_\alpha$. L'étape suivante consiste alors à estimer ce paramètre pour chaque famille de copules qu'on veut tester. Deux méthodes, au moins, sont possibles :

**i)** En utilisant la relation qui existe entre le coefficient de corrélation de Kendall d'une copule archimédienne $H_{\Phi_\alpha}$ et sa fonction génératrice $\Phi_\alpha$ (voir Genest [8]) :

$$\tau(\alpha) = \tau(H_{\Phi_\alpha}) = 4\int_0^1 \frac{\Phi_\alpha(t)}{\Phi'_\alpha(t)}\, dt + 1$$

Connaissant le coefficient de corrélation de Kendall calculé à partir de l'échantillon, noté $\tau_n$, défini par :

$$\tau_n = \frac{\displaystyle\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_{ij}Y_{ij}}{\dbinom{n}{2}}, \quad \text{avec} \quad \begin{cases} X_{ij} = 1 & \text{si } x_i \leqslant x_j; Y_{ij} = 1 \text{ si } y_i \leqslant y_j \\ X_{ij} = 0 & \text{si } x_i > x_j; Y_{ij} = 0 \text{ si } y_i > y_j. \end{cases}$$

l'estimateur $\hat{\alpha}$ de $\alpha$ est solution de l'équation : $\tau(\alpha) = \tau_n$

**ii)** En maximisant une pseudo-vraisemblance de $H_{\Phi_\alpha}$ donnée par :

$$l(\alpha) = \sum_{i=1}^{n} \log\left[h_{\Phi_\alpha}(F_{1n}(x_i), F_{2n}(y_i))\right]$$

où $h_{\Phi_\alpha}, F_{1n}, F_{2n}$ sont respectivement la densité de $H_{\Phi_\alpha}$ et les fonctions de répartition empiriques de $X$ et $Y$.

En remplaçant $\alpha$ par son estimateur $\hat{\alpha}$ dans l'équation (9), on obtient alors un estimateur paramétrique $\lambda_{\Phi_{\hat{\alpha}}}$ de $\lambda_{\Phi_\alpha}$.

**Remarque.** Nous avons étudié plusieurs algorithmes d'estimation des paramètres de $\Phi_\alpha$ et nous avons montré que la méthode d'estimation basée sur le coefficient de corrélation de Kendall et celle basée sur le maximum de la pseudo-vraisemblance, définie ci-dessus, donnent de bons résultats par rapport à d'autres méthodes paramétriques ou non-paramétriques, dans le cas de copules archimédiennes bidimensionnelles (voir Hillali *et al* [13]).

Finalement, si on suppose qu'une distribution bivariée $F$ à ajuster possède une représentation par une copule archimédienne de type (8), la procédure définie ci-dessus nous aide à choisir parmi plusieurs familles de copules archimédiennes celle qui ajuste au mieux nos données. Ce choix passe par une comparaison graphique de la fonction $\lambda$ théorique (équation 9) et de $\lambda$ empirique (équation 10); ce qui revient à comparer les versions paramétrique et non-paramétrique de la distribution de la variable non-observée $U = H_{\Phi_\alpha}(X, Y)$, c'est-à-dire les distributions $S_{\Phi_\alpha}(U)$ et $S_n(U)$.

Pour certifier le fait qu'une famille paramétrique de copules archimédiennes est la plus adaptée à nos données par rapport à d'autres modèles, deux approches sont possibles :

**1.** En utilisant des tests non-paramétriques basés sur des critères de type chi-deux, Kolmogorov-Smirnov, ou en calculant, par exemple, la distance entre les distributions $S_{\Phi_\alpha}(U)$ et $S_n(U)$ définie par : $\int (S_{\Phi_\alpha}(u) - S_n(u))^2 \, dS_n(u)$.

**2.** En comparant les estimateurs des paramètres du modèle obtenus par la méthode du maximum de vraisemblance associé aux distributions univariées et ceux obtenus par la méthode du maximum de vraisemblance opérant sur la distribution bivariée associée à chaque modèle de copules archimédiennes retenu. Concrètement, si on note $F$ la fonction de répartition d'un couple de variables aléatoires $(X_1, X_2)$ que nous souhaitons ajuster par une copule archimédienne $H_{\Phi_\alpha}$, définie par :

$$F(x_1, x_2, \theta_1, \theta_2, \alpha) = H_{\Phi_\alpha}(F_1(x_1, \theta_1), F_2(x_2, \theta_2), \alpha)$$

où $F_1$ et $F_2$ sont les fonctions de répartition univariées de $X_1$ et $X_2$, dépendant respectivement des paramètres (ou vecteurs de paramètres) $\theta_1$ et $\theta_2$ et $\alpha$ est le paramètre de la fonction génératrice $\Phi_\alpha$ de $H_{\Phi_\alpha}$, alors la seconde étape consiste à :

**2.1** déterminer les estimateurs $\hat{\theta_1}$ et $\hat{\theta_2}$ de $\theta_1$ et $\theta_2$ en maximisant les log-vraisemblances $L_i, i = 1, 2$ des distributions univariées de $X_1$ et $X_2$ définies

par :

$$L_i(\theta_i) = \sum_{j=1}^{n} \log f_i(x_i^j, \theta_i), i = 1, 2$$

où $(x_1^1, x_2^1), \ldots, (x_1^n, x_2^n)$ est un échantillon de $(X_1, X_2)$ et $f_1$ et $f_2$ sont les densités respectives de $X_1$ et $X_2$;

**2.2** déterminer un estimateur $\hat{\alpha}$ de $\alpha$ en maximisant la log-vraisemblance $L(\alpha, \hat{\theta_1}, \hat{\theta_2})$ de la loi jointe $F$ de densité $f$. $L$ et $f$ sont définis par :

$$L(\alpha, \hat{\theta_1}, \hat{\theta_2}) = \sum_{j=1}^{n} \log f(x_1^j, x_2^j, \hat{\theta_1}, \hat{\theta_2}, \alpha)$$

$$f(x_1^j, x_2^j, \theta_1, \theta_2, \alpha) = h_{\Phi_\alpha}\left[F_1(x_1^j, \theta_1), F_1(x_2^j, \theta_2)\right] \; f_1(x_1^j, \theta_1) f_2(x_2^j, \theta_2)$$

où $h_{\Phi_\alpha}$ est la densité de $H_{\Phi_\alpha}$. Sous des conditions de régularité, le vecteur $(\hat{\theta_1}, \hat{\theta_2}, \hat{\alpha})$ est solution de l'équation :

$$\left(\frac{\delta L_1}{\delta \theta_1}, \frac{\delta L_2}{\delta \theta_2}, \frac{\delta L}{\delta \alpha}\right) = (0, 0, 0)$$

**2.3** déterminer les estimateurs $\tilde{\theta_1}$, $\tilde{\theta_2}$ et $\tilde{\alpha}$ de $\theta_1$, $\theta_2$ et $\alpha$ en maximisant directement $L(\alpha, \theta_1, \theta_2)$. Sous des conditions de régularité, le vecteur $(\tilde{\theta_1}, \tilde{\theta_2}, \tilde{\alpha})$ est solution de l'équation :

$$\left(\frac{\delta L}{\delta \theta_1}, \frac{\delta L}{\delta \theta_2}, \frac{\delta L}{\delta \alpha}\right) = (0, 0, 0)$$

**2.4** pour chaque famille de copules archimédiennes retenue, après une comparaison graphique des distributions $S_{\Phi_\alpha}(U)$ et $S_n(U)$, comparer les estimateurs $(\hat{\theta_1}, \hat{\theta_2}, \hat{\alpha})$ et $(\tilde{\theta_1}, \tilde{\theta_2}, \tilde{\alpha})$ et en déduire le modèle le plus adéquat pour lequel ces deux estimateurs sont proches.

Dans l'application que nous présentons ci-dessous, le choix du modèle adéquat est basé, premièrement sur une comparaison graphique de $\lambda_{\Phi_\alpha}$ et de $\lambda_n$ et deuxièmement en comparant les estimateurs par la seconde approche citée ci-dessus.

## 4. Application à des données climatologiques

Dans cette partie nous allons présenter une application de la méthode d'ajustement d'une loi bidimensionnelle basée sur le concept des copules archimédiennes. L'objectif de cette application est de prévoir le risque engendré par les valeurs extrêmes de variables climatologiques sur le fonctionnement de réseaux de distributions par fibres optiques (RDFO), expérimentés au centre de recherche du CNET à Lannion.

Pour simplifier notre exposé, nous limitons notre étude à la modélisation de la loi bidimensionnelle de la variable température maximale journalière et de la variable représentant l'intervalle de temps qui s'est écoulé avant le fléchissement de cette température maximale, c'est-à-dire sa durée de vie. La dernière variable est considérée (par les chercheurs du CNET) comme étant la plus importante et la plus influente sur le fonctionnement des RDFO. En effet, une température maximale de longue durée peut avoir plus d'effets néfastes sur les RDFO qu'une température maximale qui lui est supérieure mais de durée moindre.

Le choix du maximum journalier de la température est basé sur l'hypothèse d'indépendance nécessaire pour l'application du théorème de la convergence asymptotique de la loi des valeurs extrêmes (voir par exemple Von-Mises [23]). Cette étude nous permettra de savoir avec quelle probabilité on peut atteindre, dans une saison fixée, une température $T$ ayant une telle durée.

Dans la suite, on note TM la variable représentant les valeurs maximales de la température (les mesures de la température sont recueillies toutes les 5 minutes depuis le 4 mars 1996 au site du CNET de Lannion) et DVTM la variable qui représente leur durée de vie. La première étape de l'étude du risque de la température maximale et de sa durée de vie sur le fonctionnement des RDFO est la modélisation des distributions univariées des variable TM et DVTM. Pour la variable TM nous avons montré que la loi généralisée des valeurs extrêmes, également appelée loi de Von-Mises et Jenkinsen [23] et [15], est la loi adéquate pour modéliser cette variable. Cette loi est définie par :

$$F_1(x, \alpha, x_0, k) = \exp(-(1 - k(\frac{x - x_0}{\alpha}))^{1/k})$$

La méthode itérative du maximum de vraisemblance, basée sur l'inverse de la matrice hessienne d'information, nous a permis de résoudre le problème d'estimation, plus difficile, pour ce modèle de valeurs extrêmes. Pour la variable DVTM nous avons prouvé, par des tests paramétriques et non-paramétriques, que le modèle de Weibull ajuste de manière très correcte la distribution empirique de cette variable. Les résultats de l'estimation des paramètres des lois univariées de ces deux variables sont données dans les tableaux 1 et 2. Pour plus de détails voir Hillali *et al* [13].

Pour modéliser la loi du couple de variables aléatoires température maximale et sa durée de vie, nous avons choisi quatre modèles de copules archimédiennes, les plus utilisés dans l'étude de simulations de lois bidimensionnelles. Ces modèles sont représentés dans le tableau 3 ainsi que leurs fonctions caractéristiques, à savoir : $\Phi_\alpha$, $\lambda$ et $\tau$. Nous avons supposé que la loi jointe $F$ des variables aléatoires TM et DVTM et de lois respectives $F_1$ et $F_2$ connues (voir Hillali *et al* [13]), correspond à une

copule archimédienne $H_{\Phi_\alpha}$ avec $\Phi_\alpha \in \Psi$, telle que :

$$F(x,y) = H_{\Phi_\alpha}(F_1(x), F_2(x)).$$

Nous avons appliqué la procédure d'estimation présentée au paragraphe précédent pour les quatre modèles retenus (voir tableau 3). Les résultats des estimations des paramètres des applications $\Phi_\alpha$ associées aux quatre modèles figurent dans le tableau 4. Dans le tableau 4, $\hat{\alpha}_F$, $\hat{\alpha}_C$, $\hat{\alpha}_{GB}$ et $\hat{\alpha}_{GD}$ représentent les estimateurs du paramètre $\alpha$ de l'application $\Phi_\alpha$ associée aux modèles de Frank, Clayton, Gumbel et Ghoudi respectivement. Ces paramètres sont estimés à partir d'un algorithme basé sur le coefficient de corrélation de Kendall et des méthodes d'intégration numérique (voir Hillali [14]).

Sur le plan pratique, les difficultés typiquement numériques que soulèvent cet algorithme exigent énormément d'attention, sa résolution n'est pas toujours aisée et le temps de calcul est fort important. Les résultats de la comparaison entre la fonction $\lambda$ empirique et $\lambda$ théorique pour chaque modèle sont représentés dans les figures 1 à 5.

TABLEAU 1

*Estimateurs des trois paramètres de la loi de la variable TM et leur écart type*

| mois-année | $\hat{\alpha}$ | $\hat{x}_0$ | $\hat{k}$ |
|---|---|---|---|
| jui-96 | 5.20 (0.027) | 16.679 (0.0487) | 0.229 (0.00115) |
| aut-96 | 4.722 (0.0194) | 16.167 (0.0382) | 0.191 (0.0083) |
| sep-96 | 4.452 (0.0195) | 10.695 (0.0366) | 0.157 (0.0011) |
| oct-96 | 3.824 (0.012) | 8.97 (0.0234) | 0.226 (0.0042) |
| nov-96 | 4.023 (0.0137) | 3.327 (0.027) | 0.128 (0.0058) |

TABLEAU 2

*Estimateurs des paramètres de la loi de la variable DVTM et leur écart type.*

| mois-année | $\beta$ | $\eta$ |
|---|---|---|
| jui-96 | 1.445 (0.0045) | 52.284 (1.259) |
| aut-96 | 1.387 (0.0043) | 51.039 (1.275) |
| sep-96 | 1.426 (0.0045) | 51.686 (1.281) |
| oct-96 | 1.461 (0.0043) | 49.323 (1.229) |
| nov-96 | 1.414 (0.0044) | 52.320 (1.335) |

TABLEAU 3

*Les quatre modèles de lois bidimensionnelles utilisés pour ajuster la loi du couple des variables aléatoires* (TM, DVTM).

$$D_1(\alpha) = \int_0^{\alpha} \left( \frac{t}{\alpha(\exp(t) - 1)} \right) \mathrm{d}t$$

| Modèle | $\Phi_\alpha$ | $-\lambda$ | $\tau$ |
|---|---|---|---|
| Clayton | $(x^{-\alpha} - 1)/\alpha$ | $x(1 - x^{\alpha})/\alpha$ | $\alpha/(\alpha + 2)$ |
| Frank | $\log\left(\frac{1 - \exp(-\alpha)}{1 - \exp(-\alpha x)}\right)$ | $\frac{1 - \exp(-\alpha x)}{\alpha \exp(-\alpha x)} \log\left(\frac{1 - \exp(-\alpha)}{1 - \exp(-\alpha x)}\right)$ | $1 + 4(D_1(\alpha) - 1)/\alpha$ |
| Gumbel | $(-\log(x))^{\alpha+1}$ | $-x \log(x)/(\alpha + 1)$ | $\alpha/(\alpha + 1)$ |
| Ghoudi | $(1 - x^{\alpha})^{1/\alpha}$ | $x(1 - x^{\alpha})/x^{\alpha}$ | $(3\alpha - 2)/(\alpha - 2)$ |

TABLEAU 4

*Estimation des paramètres des quatre modèles de lois bidimensionnelles choisies pour le couple de variables aléatoires (TM, DVTM).*

| mois-année | $\hat{\alpha}_F$ | $\hat{\alpha}_C$ | $\hat{\alpha}_{GB}$ | $\hat{\alpha}_{GD}$ |
|---|---|---|---|---|
| jui-96 | 26.513 | 0.497 | 0.248 | 0.799 |
| aut-96 | 229.041 | 0.650 | 0.325 | 0.851 |
| sep-96 | 198.108 | 0.643 | 0.321 | 0.848 |
| oct-96 | 15.304 | 0.440 | 0.220 | 0.781 |
| nov-96 | 9.510 | 0.386 | 0.193 | 0.765 |

TABLEAU 5

*Résultats de la comparaison des quatre familles de copules modélisant la loi jointe du couple (TM, DVTM) (on a indiqué par un point la famille de copule retenue à chaque date)*

| mois-année | Frank | Ghoudi | Clayton | Gumbel |
|---|---|---|---|---|
| jui-96 | • | | | |
| aut-96 | | | | • |
| sep-96 | | | | • |
| oct-96 | | | | • |
| nov-96 | | | | • |

## 5. Résultats

En analysant les figures 1 à 5 qui représentent une comparaison des estimateurs paramétrique $\lambda_{\Phi_{\hat{\alpha}}}$ et non-paramétrique $\lambda_n$ de $\lambda$ [1] et le tableau 4, on remarque qu'aucun modèle de copules archimédiennes bidimensionnelles définies au tableau 3 ne peut se généraliser sur les périodes d'étude du couple de variables aléatoires TM et DVTM, la valeur du paramètre $\alpha$, pour chaque modèle donné variant beaucoup d'un mois à l'autre.

Le modèle de Ghoudi est complètement rejeté, il ne s'adapte à aucun cas des lois jointes à modéliser. C'est une conclusion étonnante, si on se confie aux travaux de Genest et Ghoudi [10], qui précisent que cette famille mono-paramétrique de lois bidimensionnelles permet de modéliser tous les degrés de dépendance possible entre deux variables aléatoires réelles; en voici un contre exemple. Par contre, le modèle de Gumbel s'adapte mieux au couple des variables aléatoires TM et DVTM. Les résultats des autres modèles sont mitigés (voir tableau 5).

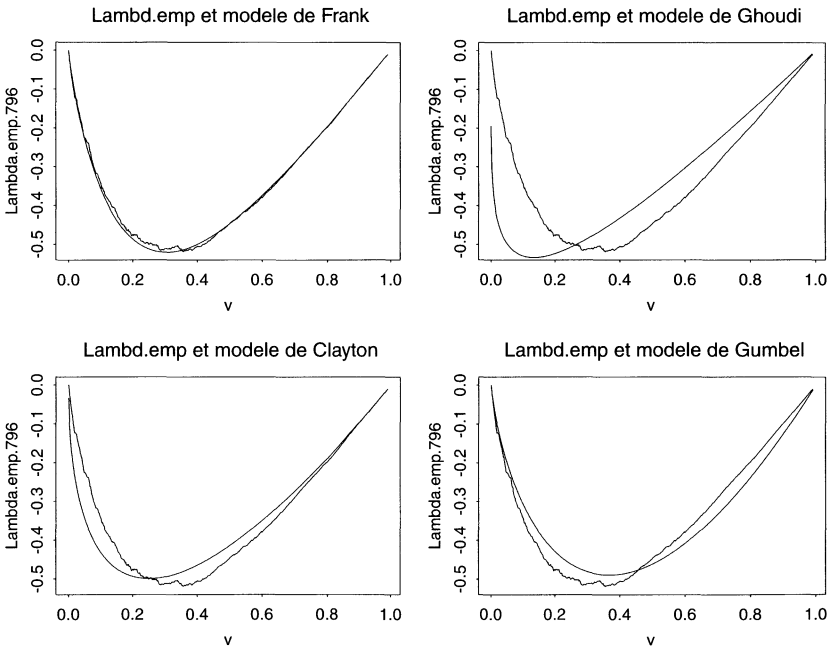


FIGURE 1

*Comparaison des estimateurs non-paramétrique et paramétrique $\lambda_n$ et $\lambda_{\Phi_{\hat{\alpha}}}$ de $\lambda$, associés aux quatre modèles pour le mois 7.96.*

[1] Dans les figures 1 à 5, la courbe lisse représente l'estimateur paramétrique $\lambda_{\Phi_{\hat{\alpha}}}$ et la courbe aux traits irréguliers représente l'estimateur non-paramétrique $\lambda_n$.

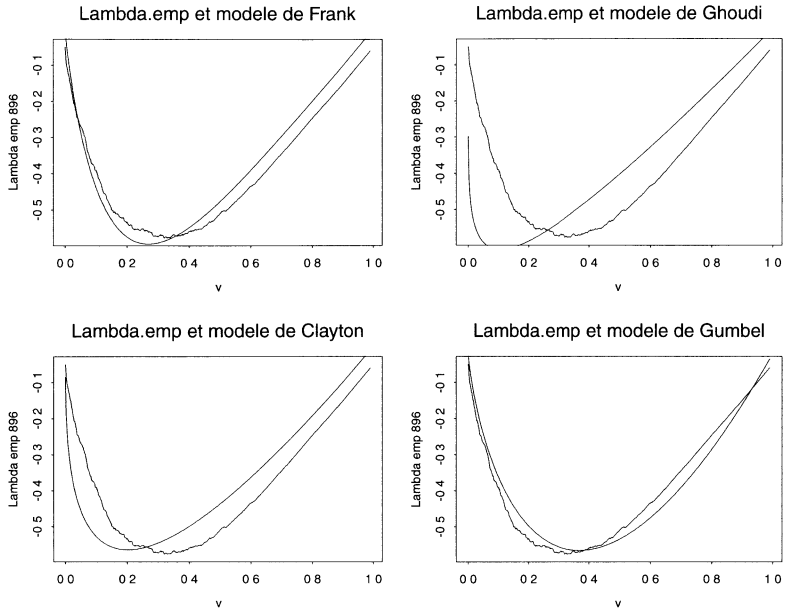


FIGURE 2

*Comparaison des estimateurs non-paramétrique et paramétrique $\lambda_n$ et $\lambda_{\Phi_{\hat{\alpha}}}$ de $\lambda$, associés aux quatre modèles pour le mois 8.96.*

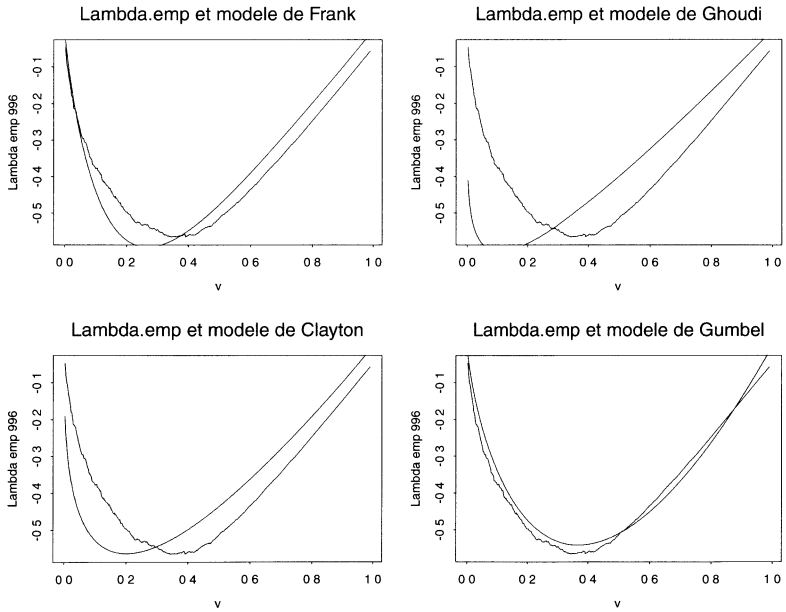


FIGURE 3

*Comparaison des estimateurs non-paramétrique et paramétrique $\lambda_n$ et $\lambda_{\Phi_{\hat{\alpha}}}$ de $\lambda$, associés aux quatre modèles pour le mois 9.96.*

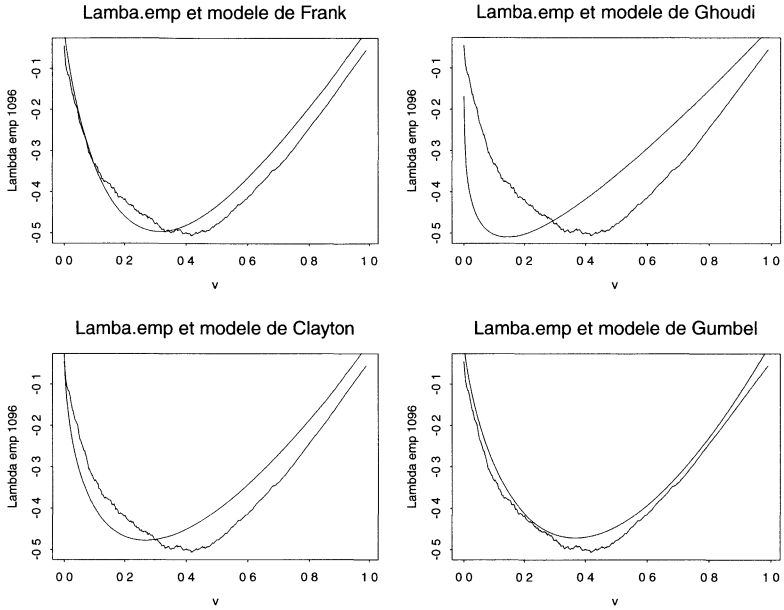


FIGURE 4

*Comparaison des estimateurs non-paramétrique et paramétrique $\lambda_n$ et $\lambda_{\Phi_{\hat{\alpha}}}$ de $\lambda$, associés aux quatre modèles pour le mois 10.96.*

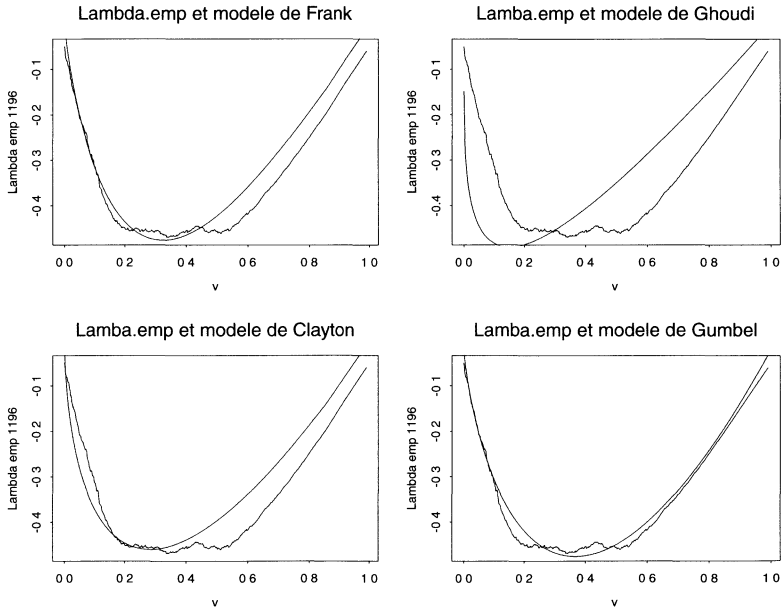


FIGURE 5

*Comparaison des estimateurs non-paramétrique et paramétrique $\lambda_n$ et $\lambda_{\Phi_{\hat{\alpha}}}$ de $\lambda$, associés aux quatre modèles pour le mois 11.96.*

## 6. Conclusion

Connaissant la loi bidimensionnelle adéquate du couple (TM, DVTM) et tous les paramètres de cette loi, on peut quantifier le risque d'atteindre une température maximale ou minimale de longue ou moyenne durée pour chaque mois de l'année. Cette étude pourrait s'étendre à toutes les variables ayant une influence sur le fonctionnement des RDFO.

Le choix des copules archimédiennes n'est pas restrictif, on pourrait choisir d'autres familles de copules non-archimédiennes pour construire des lois bidimensionnelles à marges unidimensionnelles données, mais les difficultés engendrées par l'étude de ces modèles sont multiples, aussi bien d'ordre théorique que d'ordre pratique. On pourrait, par exemple, s'inspirer des travaux de Carriere [2] pour définir une autre procédure d'ajustement de lois bidimensionnelles engendrées par des copules quelconques, dont les paramètres sont plutôt liés au coefficient de corrélation de Spearman qu'au coefficient de corrélation de Kendall, comme c'est le cas dans ce travail.

## Remerciements

Je remercie le Professeur Pierre Cazes pour sa disponibilité et ses commentaires.

## Références

[1] ALI M., MIKHAIL N. et HAQ N. S. (1978). *A class of bivariate distribution including the bivariate logistic given margins. Journal of multivariate analysis*, 8, pp. 405-412.

[2] CARRIERE J. F. (1994). "A large sample test for one parameter families of copulas". *Commun. Statisti. -Theory Meth.*, 23(5), pp. 1311-1317.

[3] CHAKAK A. et KOEHLER K.J. (1995). "A strategy for constructing multivariate distributions. *Commun. Statist. Simula.*, 24 (3), pp. 537-550.

[4] CLAYTON D. G. (1978). "A model for association in bivariate life tables and its application in epidemological studies of familial tendency in chronic disease incidence. *Biometrika*, 645, pp. 141-151.

[5] CUADRAS C. M. (1992). "Probability distribution with given multivariate marginals and given dependence structure. *Journal of multivariate analysis*, 45, pp. 51-66.

[6] FRANK M. J. (1979). "On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Math*, 19, pp. 53-77.

[7] GENEST C. et MACKAY J. (1986). « Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données », *The Canadian journal of statistics*, 4, pp. 145-159.

[8] GENEST C. (1987). "Frank's family of bivariate distributions". *Biometrika*, 47, pp. 549-555.

[9] GENEST C. et RIVEST P. L. (1993), "Statistical inference procedures for bivariate archimedean copulas", *JASA*, 88, pp. 1034-1043.

[10] GENEST C. et GHOUDI K. (1994). « Une famille de lois bidimensionnelles insolite ». *C. R. A. S. Paris*, I, pp. 351-354.

[11] GENEST C., QUESADA MOLINA J. J. et RODRIGUEZ LALLENA J. A. (1995). « De l'impossibilité de construire des lois à marges multidimensionnelles données à partir de copules ». *C. R. A. S. Paris*, I, pp. 723-726.

[12] GUMBEL E. J. (1961). "Bivariate logistic distributions". *J. Amer. Statist. Assoc.*, 56, pp. 335-349.

[13] HILLALI Y. , DIDAY E. et DHUIN C. (1997). « Étude des valeurs extrêmes multivariées des variables climatologiques et leur influence sur les RDFO ». Rapport de convention LISE CEREMADE (Université Paris IX-Dauphine) et CNET Lannion, novembre 1997.

[14] HILLALI Y. (1998). « Analyse et modélisation des données probabilistes : capacités et lois multidimensionnelles ». *Thèse de doctorat de troisième cycle, Université de Paris Dauphine*.

[15] JENKINSEN A.F. (1955). "The frequency distribution of the annual maximum (or minimum) value of meteorologie elements". *Quatr. J. Roy. Met. Soc*, 81, pp. 158-171.

[16] JOE H. (1993). "Parametric families of multivariate distributions with given margins". *Journal of multivariate analysis*, 46, pp. 262-282.

[17] KIMELDORF G. et SAMPSON A. R. (1975). "Uniform Representations of Bivariate Distributions". *Comm. Statist*, 4, pp. 617-627.

[18] NATAF A. (1962). « Détermination des Distributions de probabilités dont les marges sont données ». *C. R. A. S. Paris*, A 255, pp. 42-43.

[19] SCHWEIZER B. (1991). "Thirty years of Copulas, Advances in probability distributions given marginals : Beyond the Copulas, ed. by G. Dall'Aglio, S. Kotz and G. Salinetti". *Mathematics and its Applications*, 67, pp. 13-50.

[20] SCHWEIZER B. et SKLAR A. (1983). Probabilistic metric spaces. *North-Holland, Amsterdam.*

[21] SCHWEIZER B. et WOLF E. F. (1981). "On nonparametric measures of dependence for random variables". *The annals of Statistics,* 9, pp. 879-885.

[22] SKLAR A. (1973). "Random variables, joint distribution functions, and Copulas". *Kybernetika*, 9, pp. 449-460.

[23] VON MISES R. (1954). « La distribution de la plus grande de $n$ valeurs ». *In selected papers, Amer. Math. Soc*, II, pp. 271-294.