

Représentation uniforme des lois multidimensionnelles et mesures de dépendance

Younès Hillali

► **To cite this version:**

Younès Hillali. Représentation uniforme des lois multidimensionnelles et mesures de dépendance. [Rapport de recherche] RR-3704, INRIA. 1999. <inria-00072964>

HAL Id: inria-00072964

<https://hal.inria.fr/inria-00072964>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Représentation uniforme des lois
multidimensionnelles et mesures de dépendance*

Younès Hillali

N° 3704

Juin 1999

————— THÈME 4 —————

 *Rapport
de recherche*

Représentation uniforme des lois multidimensionnelles et mesures de dépendance

Younès Hillali*

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet SODAS

Rapport de recherche n° 3704 — Juin 1999 — 21 pages

Résumé : D’après Sklar (1973), toute loi de probabilité multidimensionnelle dont les marges sont continues peut s’écrire de façon unique par une fonction de répartition dont la masse est concentrée sur $[0, 1]^n$, appelée *copule* (Sklar 1973, Schweizer et Sklar 1983); “fonction de dépendance” (Galombos, 1978), et “représentation uniforme” (Kimeldorf et Sampson, 1975). Dans ce travail, on s’intéresse à une famille particulière de copules, dites copules archimédiennes. Nous donnons les conditions sous lesquelles une copule archimédienne multidimensionnelle définit une fonction de répartition de dimension n et nous présentons une généralisation d’une famille particulière de copules archimédiennes bidimensionnelles en dimension n . Nous examinons ses propriétés et nous montrons qu’il existe une correspondance naturelle entre les paramètres de cette famille et deux coefficients de dépendance multivariée. Cette correspondance servira à présenter deux procédures d’estimation de ces paramètres.

Mots-clés : Copules archimédiennes, dépendance multivariée, lois multidimensionnelles, représentation uniforme.

(Abstract: *pto*)

* Email : Younes.Hillali@inria.fr

Uniform representation of multivariate distributions and measures of association

Abstract: By Sklar (1973), every n -variate distribution function with continuous marginals can be represented by a distribution function on $[0, 1]^n$ with uniform marginals, referred to as a “copulas” (Sklar 1973, Schweizer and Sklar 1983); a “dependence function” (Galombos, 1978), and a “uniform representation” (Kimeldorf and Sampson, 1975). This paper is concerned with a special class of copulas called “Archimedean”, which includes the uniform representation of many multivariate distributions. In the bivariate case this family will be examined by Genest and Mackay (1986). We gives a generalization of the bivariate Archimedean copulas in the multivariate case. We also provide a generalization of Frank’s family which include as limiting cases the Fréchet upper bound (Fréchet, 1951) and the independence distribution, and we suggests two nonparametric estimation procedures of its parameters, based on the measures of association between random variables.

Key-words: Archimedean copulas, multivariate distributions, multivariate concordance, uniform representation.

1 Introduction

Ces dernières années, plusieurs auteurs se sont intéressés à la construction de fonctions de répartition multivariées dont les marges sont fixées. On peut citer, par exemple, les travaux de Krummenauer (1998), Johnson et al (1997), Chakak (1995), Koehler et Symanowski (1995), et Cuadras (1992). En effet, dans plusieurs situations, une modélisation par une distribution multivariée dont les marges sont des lois de Weibull, de Pareto, ou de Gamma, est plus justifiée qu'une loi normale multidimensionnelle dont la construction est plus simple ; pensons par exemple aux problèmes des valeurs extrêmes multivariées. L'objectif commun, alors, à tous ces travaux et au nôtre est celui de diversifier le choix de distributions possibles pour la modélisation ou la simulation de données multidimensionnelles correspondant à de telles situations.

Selon Sklar (1973), toute loi de probabilité multidimensionnelle F de marges F_1, \dots, F_n continues correspond de manière unique à une copule H , à savoir une fonction de répartition dont la masse est concentrée sur $[0, 1]^n$ et dont les marges sont uniformes. Cette copule est déterminée par la formule :

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

où $F_i^{-1}(x)$ représente l'infimum de l'ensemble $\{t \in \mathfrak{R} : F_i(t) \geq x\}$, $i = 1, \dots, n$. La fonction H représente la dépendance entre les lois marginales et ne présente aucune restriction sur leur nature.

Dans cet article, nous nous intéressons à une classe particulière de copules multidimensionnelles, à savoir les copules archimédiennes multidimensionnelles, qui ont été largement étudiées dans le cas bidimensionnel par Genest et Mackay (1986). Nous donnons les conditions sous lesquelles ces copules représentent des fonctions de répartition multidimensionnelles. Pour simplifier les difficultés d'ordre théorique ou d'ordre pratique engendrées par la construction des lois multidimensionnelles à marges unidimensionnelles, nous allons concentrer la deuxième partie de cet article à la généralisation d'une copule archimédienne bidimensionnelle particulière, dite copule de Frank (voir Genest (1987)), qui a suscité beaucoup d'intérêt dans la modélisation de lois bivariées. Nous présentons une étude détaillée de cette famille de lois multidimensionnelles, ainsi que des algorithmes permettant d'estimer ses paramètres.

Le reste de notre article se compose de cinq parties. À la section 2, nous présentons une généralisation des copules archimédiennes bidimensionnelles au cas multidimensionnel et nous traitons des propriétés mathématiques de cette nouvelle famille de lois multidimensionnelles. À la section 3, nous nous intéressons à un cas particulier, à savoir la famille de Frank. À la section 4 nous présentons deux méthodes d'estimation des paramètres de la famille de Frank en dimension 3, basées sur les coefficients de corrélation généralisés de Spearman et de Kendall. Nous décrivons, entre autres, un algorithme permettant de simuler des copules archimédiennes multidimensionnelles. Nous finirons par quelques remarques sur la généralisation d'autres familles de copules archimédiennes bidimensionnelles, ainsi que les résultats de la comparaison des deux procédures d'estimation des paramètres de la famille de Frank en dimension 3.

2 Copules archimédiennes multidimensionnelles

Par souci de simplification nous limitons notre discussion à la construction de lois multidimensionnelles dont les marges sont uniformes sur le carré unité $[0, 1]$. Une généralisation à d'autres formes de fonctions de répartition est immédiate en appliquant le principe de translation de Nataf (1962). Soient X_1, \dots, X_n , n variables aléatoires uniformes variant simultanément sur le carré unité $[0, 1]$, et H_n leur loi jointe. Soit Υ la classe des applications définies sur $[0, 1]$ à valeurs dans $[0, \infty[$, continues, strictement décroissantes et convexes, pour lesquelles $\Phi(1) = 0$. Considérons les fonctions H_Φ définies sur le rectangle $[0, 1]^2$ par :

$$H_\Phi(x, y) = \Phi^{-1}(\Phi(x) + \Phi(y)) \quad (2)$$

où Φ appartient à Υ et $H_\Phi(x, y) = 0$ si $\Phi(x) + \Phi(y) \geq \Phi(0)$. Pour qu'une fonction H_Φ de la forme (2) soit une fonction de répartition sur le carré unité $[0, 1]$, il faut et il suffit que $\Phi \in \Upsilon$ (voir le théorème 6.3.6 de Schweizer et Sklar (1983)). H_Φ est appelée copule archimédienne de dimension 2 (voir Genest et Mackay (1986)), associée au couple (X, Y) dont les fonctions de répartition marginales sont uniformes sur $[0, 1]$. Pour simplifier les notations, nous noterons simplement H une copule archimédienne H_Φ . Une première généralisation intuitive de H de la forme (2) à l'ordre $n \geq 3$ pourrait être de la forme :

$$H_n(x_1, \dots, x_n) = \Phi^{-1}(\Phi(x_1) + \dots + \Phi(x_n)), \quad 0 \leq x_1, \dots, x_n \leq 1 \quad (3)$$

D'après le théorème (6.3.6) de Schweizer et Sklar (1983), l'application de la forme (3) est une fonction de répartition pour tout $n \geq 3$ si et seulement si Φ^{-1} est complètement monotone, c'est-à-dire :

$$(-1)^k \frac{\partial^k}{\partial t^k} \Phi^{-1}(x) \geq 0, \forall k = 1, 2, \dots \quad (4)$$

Cependant, une modélisation des lois multidimensionnelles par une application de la forme (3) n'aura pas beaucoup d'intérêt, puisque chaque élément de cette famille est défini par une même application Φ et donc les dépendances entre les sous-ensembles de variables aléatoires de $\{X_1, \dots, X_n\}$ ne sont pas mises en valeurs, seule une dépendance globale est prise en compte. Pour cette raison, nous proposons d'étudier une famille, notée \mathcal{H} , de copules archimédiennes à l'ordre $n \geq 3$, proposée pour la première fois par Joe (1993), qui tient compte de toutes les dépendances entre les variables aléatoires X_1, \dots, X_n . Si on note H_n la copule archimédienne de dimension n du vecteur aléatoire (X_1, \dots, X_n) , alors cette famille est définie par :

$$H_n(x_1, \dots, x_n) = \Phi_{n-1}^{-1}(\Phi_{n-1}(H_{n-1}(x_1, \dots, x_{n-1})) + \Phi_{n-1}(x_n)) \quad (5)$$

où $H_{n-1}(x_1, \dots, x_{n-1}) = \Phi_{n-2}^{-1}(\Phi_{n-2}(H_{n-2}(x_1, \dots, x_{n-2})) + \Phi_{n-2}(x_{n-1}))$ avec $0 \leq x_1, \dots, x_n \leq 1$ et l'application Φ_i pour $1 \leq i \leq n-1$ appartient à la classe Υ .

L'objet principal de cet article est de construire une copule H_n qui soit une fonction de répartition de dimension $n \geq 3$.

D'après un résultat de Joe (1993), la copule H_n de type (5) est une fonction de répartition si et seulement si elle vérifie la condition suivante : pour tout i tel que $2 \leq i \leq n$ et toute constante k strictement positive, l'application Γ définie sur $[0, 1]$ à valeurs dans $[0, \infty[$ telle que : $\Gamma(t) = \exp(-k(\Phi_i \circ \Phi_{i-1}^{-1}(t)))$ est une transformée de Laplace.

Cette condition est généralement très difficile à vérifier, mais grâce à un théorème de Feller (1966) elle devient beaucoup plus simple et se résume par la proposition suivante :

Proposition 1 : Une application H_n pour $n \geq 3$ de type (5) est une fonction de répartition si et seulement si les applications $(-\Phi_{n-1} \circ \Phi_{n-2}^{-1}), \dots, (-\Phi_2 \circ \Phi_1^{-1})$ sont complètement monotones.

La démonstration de ce résultat se déduit du théorème de Feller (1966) et de la condition proposée par Joe (1993).

2.1 Propriétés des copules archimédiennes multidimensionnelles de la famille \mathcal{H}

Pour faciliter la compréhension des propriétés des copules de la forme (5), appartenant à la famille \mathcal{H} , on va se limiter à la dimension trois. Soit H une copule de dimension trois de type (5) engendrée par deux applications Φ_1 et Φ_2 appartenant à Υ et dont les lois marginales unidimensionnelles sont uniformes sur l'intervalle $[0, 1]$. D'après la procédure de construction des copules appartenant à la famille \mathcal{H} (voir équation (5)), H est définie par :

$$H(x_1, x_2, x_3) = \Phi_2^{-1}(\Phi_2[\Phi_1^{-1}(\Phi_1(x_1) + \Phi_1(x_2))] + \Phi_2(x_3)). \quad (6)$$

Puisque Φ_2^{-1} est définie de $[0, \infty[$ à valeurs dans $[0, 1]$, il convient de poser : $H(x_1, x_2, x_3) = 0$ lorsque $\Phi_2[\Phi_1^{-1}(\Phi_1(x_1) + \Phi_1(x_2))] + \Phi_2(x_3) \geq \Phi_2(0)$. Pour qu'une application H de type (6) soit une fonction de répartition de dimension trois sur $[0, 1]^3$, il faut et il suffit que l'application $-\Phi_2 \circ \Phi_1^{-1}$ soit complètement monotone. Cette condition s'écrit :

$$\frac{\partial^k(\Phi_2 \circ \Phi_1^{-1}(t))}{\partial t^k} \geq 0, \quad \forall k \geq 1.$$

Soient θ_1 et θ_2 les paramètres respectifs des applications Φ_{θ_1} et Φ_{θ_2} . On suppose, sans perte de généralité, que ces paramètres sont compris entre -1 et 1.

On note H_{θ_1, θ_2} une copule de dimension trois de type (6) engendrée par les deux applications Φ_{θ_1} et $\Phi_{\theta_2} \in \Upsilon$ telle que $-\Phi_{\theta_2} \circ \Phi_{\theta_1}^{-1}$ soit complètement monotone.

Soit H^0 la distribution correspondant au cas où X_1, X_2 et X_3 sont des variables aléatoires indépendantes : $H^0(x_1, x_2, x_3) = x_1 x_2 x_3$.

Soit H^1 (resp. H^{-1}) la distribution correspondant au cas où X_1, X_2 et X_3 sont parfaitement corrélées positivement (resp. la distribution correspondant au cas où X_1, X_2 et X_3 sont parfaitement corrélées négativement) :

$$H^1(x_1, x_2, x_3) = \min(x_1, x_2, x_3), \quad H^{-1}(x_1, x_2, x_3) = \max(x_1 + x_2 + x_3 - 2, 0).$$

H^1 et H^{-1} sont appelées respectivement la borne supérieure et inférieure de Fréchet (voir Fréchet (1951)). Les conditions suivantes sont une extension de celles de Kimeldorf et Simpson (1975) associées aux copules bidimensionnelles. Elles permettent d'assurer que les lois H^0 et H^1 sont un prolongement des copules de la

famille \mathcal{H} définie par l'équation (6) :

- 1) $H_{\theta_1, \theta_2}(x_1, x_2, x_3)$ est une fonction continue par rapport à θ_1 et θ_2 pour tout x_1, x_2, x_3 dans l'intervalle $[0, 1]$,
- 2) H_{θ_1, θ_2} est une copule absolument continue par rapport à la mesure de Lebesgue pour tout θ_1, θ_2 compris entre -1 et 1,
- 3) $\lim_{\theta_1, \theta_2 \rightarrow 1} H_{\theta_1, \theta_2} = H^1$, $\lim_{\theta_1, \theta_2 \rightarrow 0} H_{\theta_1, \theta_2} = H^0$.

La loi $H^{-1}(x_1, x_2, x_3) = \max(0, x_1 + x_2 + x_3 - 2)$ n'est pas prise en compte dans les conditions 1, 2 et 3 puisqu'elle n'est pas une copule et donc elle n'appartient pas à la famille \mathcal{H} .

2.2 Coefficients de dépendance et copules archimédiennes multidimensionnelles de la famille \mathcal{H}

Dans cette partie nous allons voir qu'il existe une correspondance naturelle entre les paramètres des copules archimédiennes de dimension n de la famille \mathcal{H} et deux mesures de dépendance multivariée, à savoir les coefficients de corrélation de Spearman et de Kendall généralisés.

2.2.1 Coefficients de la dépendance multivariée

Les mesures de la dépendance bivariée de Spearman et de Kendall s'étendent facilement à toute famille finie de variables aléatoires X_1, \dots, X_n .

Nous pouvons utiliser un nombre unique pour mesurer la dépendance collective (c'est le cas si on construit une loi multidimensionnelle à partir de ses marges par une copule de la forme (3)) ou $2^n - 1$ nombres pour tenir compte de toutes les distributions marginales (voir les travaux de Cuadras (1992)). Pour simplifier, nous nous limiterons au premier cas, c'est-à-dire à une mesure de la dépendance collective.

Soit H_n une copule de dimension n du vecteur aléatoire (X_1, \dots, X_n) de lois marginales uniformes sur $[0, 1]$ de loi jointe F telle que : $F(x_1, \dots, x_n) = H_n(x_1, \dots, x_n)$.

Les coefficients de corrélation de Spearman et de Kendall associés à un couple de variables aléatoires peuvent se généraliser à un vecteur aléatoire (X_1, \dots, X_n) , $n \geq 2$, de copule H_n (voir Joe (1990)), de la manière suivante :

$$\rho(H_n) = \frac{1}{(n+1)^{-1} - 2^{-n}} \left(\int_{[0,1]^n} x_1 \dots x_n dH_n(x_1, \dots, x_n) - 2^{-n} \right) \quad (7)$$

$$\tau(H_n) = \frac{1}{2^{n-1} - 1} \left(2^n \int_{[0,1]^n} H_n(x_1, \dots, x_n) dH_n(x_1, \dots, x_n) - 1 \right). \quad (8)$$

2.2.2 Coefficient de corrélation de Kendall d'une copule archimédienne de la famille \mathcal{H}

Si on considère trois variables aléatoires X , Y et Z de lois uniformes sur le carré unité et de copule H , alors, d'après les équations (7) et (8), les coefficients de corrélation de Spearman et de Kendall à l'ordre trois sont définis par :

$$\rho(H) = 8 \left(\int_{[0,1]^3} xyz dH - \frac{1}{8} \right) \quad (9)$$

$$\tau(H) = \frac{8}{3} \left(\int_{[0,1]^3} H dH - \frac{1}{8} \right) \quad (10)$$

Pour éviter les problèmes de dérivabilité on va supposer, dans ce qui suit, que Φ_{θ_1} et Φ_{θ_2} admettent trois dérivées continues sur $[0, 1]$, c'est-à-dire $\Phi_{\theta_1}, \Phi_{\theta_2} \in \Upsilon \cap C^3$.

Proposition 2 : Soit H_{θ_1, θ_2} une copule archimédienne de dimension trois et de type (6) engendrée par Φ_{θ_1} et $\Phi_{\theta_2} \in \Upsilon \cap C^3$ et vérifiant la condition nécessaire et suffisante de la proposition 1.

Alors, le coefficient de corrélation de Kendall généralisé mesurant la liaison entre les variables aléatoires X , Y et Z de loi jointe H_{θ_1, θ_2} est défini par :

$$\tau(H_{\theta_1, \theta_2}) = \frac{8}{3} \left(\int_0^1 \int_0^s I(t) \left[\frac{\Phi'_{\theta_1}(s)\Phi''_{\theta_2}(t)}{\Phi'_{\theta_1}(t)} - \frac{\Phi'_{\theta_1}(s)\Phi'_{\theta_2}(t)\Phi''_{\theta_1}(s)}{(\Phi'_{\theta_1}(t))^2} \right] - \frac{\Phi'_{\theta_1}(s)(\Phi'_{\theta_2}(t))^2}{\Phi'_{\theta_1}(t)} J(t) ds dt - \frac{1}{8} \right)$$

avec

$$I(t) = \frac{t}{\Phi_{\theta_2}(t)} \int_0^t \frac{1}{\Phi'_{\theta_2}(w)},$$

$$J(t) = \frac{t\Phi''_{\theta_2}(t)}{(\Phi'_{\theta_2}(t))^3} - \left(\frac{1}{\Phi'_{\theta_2}(t)} - \frac{1}{\Phi'_{\theta_2}(0)} \right)^2.$$

Preuve :

En utilisant les équations (6) et (10) on a :

$$\tau(H_{\theta_1, \theta_2}) = \frac{8}{3} \left(\int_{[0,1]^3} H_{\theta_1, \theta_2} dH_{\theta_1, \theta_2}^{ac} + \int_{[0,1]^3} H_{\theta_1, \theta_2} dH_{\theta_1, \theta_2}^s - \frac{1}{8} \right)$$

où $H_{\theta_1, \theta_2}^{ac}$ est la partie absolument continue par rapport à la mesure de Lebesgue de H_{θ_1, θ_2} et H_{θ_1, θ_2}^s sa partie singulière concentrée sur un ensemble de mesure nulle.

Il suffit donc d'intégrer $H_{\theta_1, \theta_2} dH_{\theta_1, \theta_2}^{ac}$ sur $[0, 1]^3$.

En utilisant les changements de variables :

$$w = H_{\theta_1, \theta_2}(x, y, z), t = \Phi_{\theta_1}^{-1}(\Phi_{\theta_1}(x) + \Phi_{\theta_1}(y))$$

on obtient le résultat souhaité.

3 Famille de Frank généralisée

Dans cette partie, nous proposons une famille de copules archimédiennes de dimension trois, appartenant à \mathcal{H} et généralisant la famille de Frank en dimension deux, largement traitée par Genest (1987). Nous donnerons la condition sous laquelle elle représente une fonction de répartition à trois dimensions concentrée sur $[0, 1]^3$ de marges uniformes. Nous examinerons deux estimateurs non-paramétriques de ses paramètres. Ces estimateurs sont basés sur les coefficients de corrélation de Spearman et de Kendall généralisés.

Soit H_γ la copule de Frank à deux dimensions définie pour $\gamma > 0, \gamma \neq 1$ et $x, y \in [0, 1]$ par :

$$H_\gamma(x, y) = \Phi_\gamma^{-1}(\Phi_\gamma(x) + \Phi_\gamma(y)) = \frac{\log(1 + \frac{(\gamma^x - 1)(\gamma^y - 1)}{(\gamma - 1)})}{\log(\gamma)}, \text{ avec } \Phi_\gamma(t) = -\log\left(\frac{1 - \gamma^t}{1 - \gamma}\right)$$

Une généralisation de H_γ en dimension trois, par une famille appartenant à \mathcal{H} , se présente sous la forme suivante :

$$H_{\gamma_1, \gamma_2}(x, y, z) = \Phi_{\gamma_2}^{-1}(\Phi_{\gamma_2}[\Phi_{\gamma_1}^{-1}(\Phi_{\gamma_1}(x) + \Phi_{\gamma_1}(y))] + \Phi_{\gamma_2}(z))$$

$$H_{\gamma_1, \gamma_2}(x, y, z) = \frac{\log\left(1 + \frac{(\gamma_2^{[\Phi_{\gamma_1}^{-1}(\Phi_{\gamma_1}(x) + \Phi_{\gamma_1}(y))] - 1)(\gamma_2^z - 1)}{(\gamma_2 - 1)}\right)}{\log(\gamma_2)} \quad (11)$$

où Φ_{γ_1} et Φ_{γ_2} sont deux applications de la classe Υ et définies de la même façon que Φ_{γ} .

Proposition 3 : La copule H_{γ_1, γ_2} définie dans l'équation (12) est une fonction de répartition de dimension trois, de marges uniformes sur $[0, 1]$, si et seulement si $\gamma_1 < \gamma_2$, avec $\gamma_1, \gamma_2 \neq 1$ et strictement positifs.

Preuve :

En appliquant la proposition (1) aux deux applications Φ_{γ_1} et Φ_{γ_2} associées à H_{γ_1, γ_2} , la condition nécessaire et suffisante se résume alors par :

$-\log(-1 + (-\exp(-x) + \exp(-x)\gamma_1 + 1)^\beta) - \log(-1 + \gamma_2)$ complètement monotone, avec $\beta = \frac{\log(\gamma_2)}{\log(\gamma_1)}$. Une étude de cette fonction donne la condition $\gamma_1 < \gamma_2$.

3.1 Propriétés de la famille de Frank généralisée

La borne supérieure de Fréchet à trois dimensions et la loi de l'indépendance sont un prolongement de la famille H_{γ_1, γ_2} . En effet :

$$\lim_{\gamma_1, \gamma_2 \rightarrow 1} H_{\gamma_1, \gamma_2}(x, y, z) = xyz,$$

$$\lim_{\gamma_1, \gamma_2 \rightarrow 0} H_{\gamma_1, \gamma_2}(x, y, z) = \min(x, y, z).$$

3.2 Coefficient de corrélation de Kendall généralisé de la famille de Frank de dimension 3

Soient X, Y et Z trois variables aléatoires de loi jointe la loi de Frank de dimension trois notée H_{γ_1, γ_2} , engendrée par deux applications Φ_{γ_1} et Φ_{γ_2} continues, strictement décroissantes et convexes, telles que $\gamma_1 < \gamma_2$. Si on note :

$$g(t) = \frac{-\gamma_2^{-t} + \gamma_2^{-t} \log(\gamma_2)t - 1}{\log(\gamma_2)^2}, \quad h(t) = g(t) - \frac{(\gamma_2^2)^{-t} + 2(\gamma_2^2)^{-t} \log(\gamma_2)t - 1}{2 \log(\gamma_2)^2},$$

$$I_1(\gamma_1, \gamma_2) = \int_0^1 \int_0^u \frac{-\Phi'_{\gamma_1}(u)\Phi''_{\gamma_2}(x)g(x)}{\Phi'_{\gamma_1}(x)} dx du,$$

$$I_2(\gamma_1, \gamma_2) = \int_0^1 \int_0^u \frac{\Phi'_{\gamma_1}(u)\Phi'_{\gamma_2}(x)g(x)}{\gamma_1^x} dx du,$$

$$I_3(\gamma_1, \gamma_2) = \int_0^1 \int_0^u \frac{-\Phi'_{\gamma_1}(u)(\Phi'_{\gamma_2}(x))^2 h(x)}{\Phi'_{\gamma_2}(x)} dx du.$$

Alors, en utilisant la proposition (2) et après plusieurs étapes de calcul, le coefficient de corrélation de Kendall entre X , Y et Z de loi H_{γ_1, γ_2} , noté $\tau(H_{\gamma_1, \gamma_2})$, est défini par :

$$\tau(H_{\gamma_1, \gamma_2}) = \frac{8}{3}(I_1(\gamma_1, \gamma_2) + I_2(\gamma_1, \gamma_2) + I_3(\gamma_1, \gamma_2) - \frac{1}{8}). \quad (12)$$

3.3 Coefficient de corrélation de Spearman généralisé de la famille de Frank de dimension 3

Soient X , Y et Z trois variables aléatoires de loi H_{γ_1, γ_2} . On pose les changements de variables suivants :

$$u = \Phi_{\gamma_1}^{-1}(\Phi_{\gamma_1}(x) + \Phi_{\gamma_2}(y)), w = H_{\gamma_1, \gamma_2}(x, y, z).$$

Soient g_1 , g_2 et g_3 trois fonctions définies sur $[0, 1]^3$ par :

$$g_1(x, u, w) = \frac{-\Phi'_{\gamma_1}(x)\Phi''_{\gamma_2}(u)}{\Phi'_{\gamma_1}(u)\gamma_2^w}, \quad g_2(x, u, w) = \frac{\Phi'_{\gamma_1}(x)\Phi'_{\gamma_2}(u)}{\gamma_1^u\gamma_2^w},$$

$$g_3(x, u, w) = \frac{-\Phi'_{\gamma_1}(x)(\Phi'_{\gamma_2}(u))^2(\gamma_2^{-w}(1 - 2\gamma_2^{-w}))}{\Phi'_{\gamma_1}(u)}.$$

Alors le coefficient de corrélation de Spearman entre X , Y et Z de loi jointe H_{γ_1, γ_2} s'écrit :

$$\rho(H_{\gamma_1, \gamma_2}) = 8 \left(\int_0^1 \int_0^x \int_0^u g_1(x, u, w) + g_2(x, u, w) + g_3(x, u, w) dx du dw - \frac{1}{8} \right). \quad (13)$$

La démonstration de ce résultat repose sur l'équation (9) et sur les changements de variables utilisés dans la proposition (2) (pour plus détails voir Hillali (1998)).

Avant d'étudier l'estimation des paramètres de la loi de Frank généralisée, nous présentons d'abord un algorithme qui s'avère utile dans le cadre d'études de simulation des copules archimédiennes multidimensionnelles. Pour simplifier les notations on suppose que les lois marginales sont uniformes sur $[0, 1]$.

Algorithme 1 :

L'algorithme suivant permet de générer une réalisation du vecteur aléatoire (X, Y, Z) de loi H de la forme (12) absolument continue, engendrée par deux applications Φ_{γ_1} et Φ_{γ_2} continues, décroissantes et convexes, et dont les paramètres vérifient la condition de la proposition (3) :

étape 1 : on génère trois variables aléatoires indépendantes X, U et T de lois uniformes sur l'intervalle $[0, 1]$;

étape 2 : on calcule $W_1 = (\Phi_{\gamma_1}^{-1})'(\Phi_{\gamma_1}(X)/U)$;

étape 3 : on pose $Y = \Phi_{\gamma_1}^{-1}(\Phi_{\gamma_1}(W_1) - \Phi_{\gamma_1}(X))$;

étape 4 : on calcule $W_2 = F^{-1}(T)$, avec F la fonction de répartition conditionnelle de la variable $W_2 = H(X, Y, Z)$ étant donné X et Y ;

étape 5 : on pose $Z = \Phi_{\gamma_2}^{-1}(\Phi_{\gamma_2}(W_2) - \Phi_{\gamma_2}(\Phi_{\gamma_1}^{-1}(\Phi_{\gamma_1}(X) + \Phi_{\gamma_1}(Y))))$.

Les méthodes d'estimation des paramètres de la loi H_{γ_1, γ_2} de type (12) peuvent être d'ordre paramétrique ou non-paramétrique. Dans les deux cas les calculs s'avèrent très difficiles, vu la complexité des formes des densités multidimensionnelles et des coefficients de corrélation généralisés de Spearman ou de Kendall. Nous allons essayer de donner quelques solutions et nous comparerons les deux méthodes présentées ci-dessous à l'aide de données simulées.

4 Estimation des paramètres de la loi de Frank généralisée

4.1 Méthode basée sur le coefficient de corrélation de Kendall

Les coefficients de corrélation de Spearman et de Kendall généralisés associés à H_{γ_1, γ_2} et présentés par les équations (13) et (14) se définissent comme fonctions des deux paramètres γ_1 et γ_2 ; d'où l'idée d'estimer ces paramètres à partir de ces deux coefficients d'association.

Soient X, Y et Z trois variables aléatoires, indépendantes ou non, de lois marginales uniformes sur $[0, 1]$ et de copule H_{γ_1, γ_2} .

Soit T le coefficient de corrélation de Kendall généralisé, calculé à partir de n réalisations des trois variables aléatoires X, Y et Z , estimé par le coefficient de

corrélation de Kendall moyen des couples (X, Y) , (X, Z) et (Y, Z) , défini par :

$$T = \frac{1}{3}(T_{XY} + T_{XZ} + T_{YZ}). \quad (14)$$

Alors, une bonne façon d'estimer γ_1 et γ_2 consiste à trouver deux paramètres $\tilde{\gamma}_1$ et $\tilde{\gamma}_2$ tels que :

$$\tau(H_{\tilde{\gamma}_1, \tilde{\gamma}_2}) = \tau(\tilde{\gamma}_1, \tilde{\gamma}_2) = T \quad (15)$$

où τ est le coefficient de corrélation de Kendall généralisé de la loi de Frank de dimension trois, défini par l'équation (13). L'équation (16) ne peut pas être résolue sachant que $\tilde{\gamma}_1$ et $\tilde{\gamma}_2$ sont tous les deux inconnus. Mais, puisque le paramètre γ_1 peut être interprété comme coefficient d'association des variables aléatoires X, Y de loi H_{γ_1} , alors, d'après un résultat de Genest (1987), un bon estimateur non-paramétrique $\tilde{\gamma}_1$ de γ_1 est solution de l'équation :

$$\tau(\tilde{\gamma}_1) = \frac{4(f(\tilde{\gamma}_1) + \log(\tilde{\gamma}_1))}{\log(\tilde{\gamma}_1)^2 + 1} = T_{X,Y} \quad (16)$$

avec $f(x) = \int_1^x \frac{\log(t)}{1-t} dt$ et $T_{X,Y}$ le coefficient de corrélation de Kendall entre X et Y calculé à partir de leurs réalisations respectives (x_1, \dots, x_n) et (y_1, \dots, y_n) , c'est-à-dire que $\tilde{\gamma}_1$ est la valeur de γ_1 pour laquelle le coefficient de corrélation de Kendall théorique est égal au coefficient de corrélation de Kendall calculé :

$$\tilde{\gamma}_1 = \tau^{-1}(T_{X,Y}).$$

Cette équation ne possède pas de solution explicite et le recours aux méthodes d'approximations successives est inévitable. Finalement, l'estimateur $\tilde{\gamma}_2$ de γ_2 sera la solution de l'équation suivante :

$$\tau(\tilde{\gamma}_1, \tilde{\gamma}_2) = T. \quad (17)$$

Remarque : Dans cette procédure, on ne peut pas estimer γ_2 à partir de $\tau(H_{\gamma_2})$ et ensuite estimer γ_1 comme solution de l'équation $\tau(\gamma_1, \tilde{\gamma}_2) = T$. En effet, le coefficient de corrélation théorique $\tau(H_{\gamma_2})$ ne dépend que de γ_2 , c'est-à-dire, il est indépendant du choix de la loi bidimensionnelle des couples (X, Z) ou (Y, Z) .

Par contre, le coefficient de corrélation de Kendall bidimensionnel, calculé à partir des réalisations, dépend du couple de variables aléatoires choisies.

Donc, le choix du couple (X, Z) de loi H_{γ_2} donne un estimateur $\tilde{\gamma}_2$ et le couple (Y, Z) donne un autre estimateur $\tilde{\gamma}'_2$. Cette méthode n'aboutit donc pas à un résultat cohérent. Cette procédure peut se résumer par l'algorithme suivant :

Algorithme 2 :

1. calculer le coefficient de corrélation de Kendall $T_{X,Y}$ des réalisations du couple (X, Y) ;
2. calculer le coefficient de corrélation de Kendall T à partir des réalisations des trois variables aléatoires, simulées selon la loi de Frank de dimension trois (voir équation 15);
3. résoudre l'équation $\tau(\gamma_1) = T_{X,Y}$ pour obtenir l'estimateur $\tilde{\gamma}_1$ de γ_1 ;
4. remplacer γ_1 par $\tilde{\gamma}_1$ et résoudre l'équation $\tau(\tilde{\gamma}_1, \gamma_2) = T$ pour obtenir l'estimateur $\tilde{\gamma}_2$ de γ_2 .

4.2 Méthode basée sur le coefficient de corrélation de Spearman

Un autre coefficient possible pour l'estimation des paramètres de H_{γ_1, γ_2} est le coefficient de corrélation de Spearman généralisé (voir équation (14)).

Soient (x_1, \dots, x_n) , (y_1, \dots, y_n) et (z_1, \dots, z_n) n réalisations des variables aléatoires X, Y et Z de loi jointe H_{γ_1, γ_2} et R_i^1, R_i^2 et R_i^3 les rangs respectifs de x_i, y_i et z_i pour $1 \leq i \leq n$. Alors le coefficient de corrélation de Spearman associé aux variables X, Y et Z noté ϱ est défini par (voir Joe (1990)) :

$$\varrho = \frac{n^{-1} \sum_i R_i^1 R_i^2 R_i^3 - ((n+1)/2)^3}{n^{-1} \sum_i i^3 - ((n+1)/2)^3} \quad (18)$$

Soit $\rho(\gamma_1, \gamma_2) = \rho(H_{\gamma_1, \gamma_2})$ le coefficient de corrélation de Spearman théorique des variables X, Y et Z de loi H_{γ_1, γ_2} . Alors, les estimateurs $\bar{\gamma}_1$ et $\bar{\gamma}_2$ de γ_1 et γ_2 sont solutions de l'équation :

$$\rho(H_{\gamma_1, \gamma_2}) : = \rho(\bar{\gamma}_1, \bar{\gamma}_2) = \varrho \quad (19)$$

Il est impossible de résoudre cette équation sachant que les deux paramètres γ_1 et γ_2 sont inconnus. Pour contourner ce problème, nous allons procéder en deux étapes,

en faisant appel aux lois marginales de H_{γ_1, γ_2} . Soit $\rho(H_{\gamma_1})$, qu'on notera par la suite $\rho(\gamma_1)$, le coefficient de corrélation de Spearman des variables aléatoires X et Y de loi H_{γ_1} .

D'après un résultat de Nelsen (1986), $\rho(H_{\gamma_1})$ et le paramètre d'association γ_1 des variables X et Y sont liés par la relation :

$$\rho(\gamma_1) = 1 + 12 \frac{1}{\log(\gamma_1)} (D_2(-\log(\gamma_1)) - D_1(-\log(\gamma_1))) \quad (20)$$

avec $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{\exp(t) - 1} dt$.

Si on note ϱ_1 le coefficient de corrélation de Spearman, calculé à partir des réalisations du couple (X, Y) , il paraît naturel qu'un estimateur $\bar{\gamma}_1$ de γ_1 soit solution de l'équation suivante :

$$\rho^{-1}(\varrho_1) = \bar{\gamma}_1. \quad (21)$$

Cette équation ne possède pas de solution explicite ; on a donc recours aux méthodes d'approximations successives d'intégrales, ce qui exige beaucoup de temps de calcul.

Connaissant le paramètre $\bar{\gamma}_1$ solution de l'équation (22), un estimateur $\bar{\gamma}_2$ de γ_2 est solution de l'équation :

$$\rho(\bar{\gamma}_1, \bar{\gamma}_2) = \varrho \quad (22)$$

où $\rho(\gamma_1, \gamma_2) = \rho(H_{\gamma_1, \gamma_2})$ et ϱ sont les coefficients de corrélation de Spearman à l'ordre 3 définis respectivement par les équations (14) et (19).

Les résultats des deux méthodes d'estimation de γ_1 et γ_2 ainsi qu'une comparaison de leurs performances sont présentés ci-dessous.

5 Simulations

Pour la famille de Frank généralisée en dimension trois, nous avons étudié deux méthodes non-paramétriques basées sur les coefficients de corrélation de Spearman et de Kendall généralisés à l'ordre trois.

Cette famille définie pour $0 < \gamma_1 < \gamma_2$ est donnée par :

$$H_{\gamma_1, \gamma_2}(x, y, z) = \frac{\log\left(1 + \frac{(\gamma_2^{[\Phi_{\gamma_1}^{-1}(\Phi_{\gamma_1}(x) + \Phi_{\gamma_1}(y))] - 1})(\gamma_2^z - 1)}{(\gamma_2 - 1)}\right)}{\log(\gamma_2)} \quad (23)$$

où $x, y, z \in [0, 1]$ et $\Phi_{\gamma_i}(t) = -\log\left(\frac{1 - \gamma_i^t}{1 - \gamma_i}\right)$, pour $i = 1, 2$.

Dans les tableaux qui suivent, les réalisations d'un vecteur aléatoire (X, Y, Z) de loi jointe H_{γ_1, γ_2} et de lois marginales uniformes sur $[0, 1]$ sont simulées à partir de l'algorithme (1) (voir section 3) ; E représente la moyenne des différentes estimations et Std l'écart type. Les paramètres $\tilde{\gamma}$ et $\bar{\gamma}$ représentent respectivement les estimateurs de la méthode du coefficient de corrélation de Kendall et de celle du coefficient de corrélation de Spearman.

$\gamma_1 = 0.4, \gamma_2 = 0.9$	$\bar{\gamma}_1$		$\bar{\gamma}_2$	
	E	Std	E	Std
$n = 10$	1.106	0.776	1.521	1.124
$n = 15$	1.095	0.295	1.466	0.638
$n = 25$	0.974	0.261	1.321	0.388
$n = 50$	0.942	0.223	1.264	0.335

TAB. 1: *Modèle de Frank généralisé : procédure du coefficient de corrélation de Spearman*

$\gamma_1 = 0.05, \gamma_2 = 0.1$	$\bar{\gamma}_1$		$\bar{\gamma}_2$	
	E	Std	E	Std
$n = 10$	0.121	0.182	0.756	1.056
$n = 15$	0.094	0.105	0.684	0.948
$n = 25$	0.076	0.094	0.591	0.902
$n = 50$	0.057	0.036	0.503	0.784

TAB. 2: *Modèle de Frank généralisé : procédure du coefficient de corrélation de Spearman*

$\gamma_1 = 0.01$	$\bar{\gamma}_1$	
	E	Std
$n = 10$	0.053	0.223
$n = 15$	0.049	0.190
$n = 25$	0.0250	0.0453
$n = 50$	0.0185	0.0448

TAB. 3: *Modèle de Frank en dimension 2 : procédure du coefficient de corrélation de Spearman*

$\gamma_1 = 0.4, \gamma_2 = 0.9$	$\tilde{\gamma}_1$		$\tilde{\gamma}_2$	
	E	Std	E	Std
$n = 10$	1.183	0.587	1.847	1.102
$n = 15$	1.224	0.321	1.701	0.814
$n = 25$	1.035	0.254	1.511	0.658
$n = 50$	0.961	0.210	1.390	0.421

TAB. 4: *Modèle de Frank généralisé : procédure du coefficient de corrélation de Kendall*

$\gamma_1 = 0.05, \gamma_2 = 0.1$	$\tilde{\gamma}_1$		$\tilde{\gamma}_2$	
	E	Std	E	Std
$n = 10$	0.134	0.0971	1.023	1.084
$n = 15$	0.0972	0.0925	0.928	0.951
$n = 25$	0.0754	0.084	0.841	0.911
$n = 50$	0.0542	0.0513	0.744	0.850

TAB. 5: *Modèle de Frank généralisé : procédure du coefficient de corrélation de Kendall*

5.1 Procédure d'estimation basée sur le coefficient de corrélation de Spearman

Comme nous l'avons précisé dans la section 4, cette méthode, basée sur les approximations numériques d'intégrales, donne des résultats contestables : soit une forte surestimation, soit une forte sous-estimation. Dans les simulations effectuées nous avons toujours observé une surestimation des paramètres, comme on peut le remarquer dans le tableau(1).

Pour de faibles valeurs de γ_1 (de l'ordre de 0.05), nous avons obtenu de mauvais estimateurs pour le paramètre γ_2 (une forte surestimation), par contre pour le paramètre γ_1 les résultats sont plus ou moins satisfaisants comme on peut le constater dans le tableau (2). Cette forte surestimation du paramètre γ_2 est due à des échantillons «pathologiques» dont le nombre n'est pas négligeable et pour lesquels on a obtenu des estimateurs fortement erronés pour γ_2 mais acceptables pour γ_1 , ce qui explique aussi la variabilité importante des estimations, notamment pour le tableau (2) : (0.00589, 5.241), (0.0875, 7.352) pour $\gamma_1 = 0.05$ et $\gamma_2 = 0.1$.

En résumé, la complexité du calcul du coefficient de corrélation de Spearman en dimension trois, qui engendre beaucoup d'approximations, rend cette méthode inefficace pour estimer les paramètres de la loi de Frank en dimension trois.

Paradoxalement, cette méthode donne des résultats satisfaisants dans le cas de la loi de Frank en dimension deux, comme on peut le constater dans le tableau (3) où nous avons simulé 100 échantillons de taille 10, 15, 25 et 50.

5.2 Procédure d'estimation basée sur le coefficient de corrélation de Kendall

La méthode du coefficient de corrélation de Kendall donne de moins bons résultats que la méthode précédente. Elle fournit des estimations aussi erronées que celles de la méthode du coefficient de corrélation de Spearman. On a observé de la même manière une surestimation dans tous les cas, surtout pour le paramètre γ_2 pour des valeurs faibles de γ_1 dans le tableau (5). Ces résultats ne sont pas du tout surprenants si on regarde de près la complexité de la formule du coefficient de corrélation de Kendall en dimension trois associé à la loi de Frank généralisée (voir équation 13) ; de plus cette méthode, testée par Genest (1987) dans le cas de la loi de Frank à deux dimensions où la complexité est peu apparente, a donné de mauvais résultats.

En général, les méthodes d'intégration numérique, en dehors de leurs applications à certains types de fonctions et de domaines d'intégration précis, restent peu fiables pour les problèmes d'estimation.

Ces simulations nous permettent de conclure qu'aucune méthode d'estimation des paramètres d'une loi multidimensionnelle n'est universelle. Selon la loi multidimensionnelle à traiter et ses caractéristiques : type de la vraisemblance, coefficient de corrélation de Spearman, coefficient de corrélation de Kendall, il faut choisir la méthode adaptée.

Sur le plan pratique, les difficultés et les problèmes typiquement numériques que soulèvent les algorithmes d'estimation des paramètres des lois multidimensionnelles (dimension trois et plus) exigent énormément d'attention, leur résolution n'est pas toujours aisée et le temps de calcul est fort important.

6 Conclusion

La généralisation d'autres familles de copules archimédiennes bidimensionnelles au cas multidimensionnel peut s'avérer inutilisable si la fonction de répartition correspondant à l'application $\Phi \in \Upsilon$ est trop compliquée, c'est-à-dire si on ne parvient pas à trouver une expression algébrique simple de Φ^{-1} . D'une manière générale la construction de lois multidimensionnelles présente des difficultés plutôt d'ordre pratique que d'ordre théorique : estimation des paramètres et expression de la densité multidimensionnelle.

Références

- Chakak, A. et Koehler, K.J. (1995). A strategy for constructing multivariate distributions. *Commun. Statist. Simula.*, 24 (3), 537-550.
- Cuadras, C.M (1992). Probability distribution with given multivariate marginals and given dependence structure. *Journal of multivariate analysis*, 45, 51-66.
- Feller, W. (1966). An introduction to Probability Theory and its Applications. *John Wiley Sons, New York*

- Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. univ. Lyon, sec. A. 14*, 53-57.
- Genest, C. (1987). Frank's family of bivariate distributions, *Biometrika*, 47, 549-555.
- Genest, C. et Mackay, R.J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *The Canadian journal of statistics*, 14, 145-159.
- Hillali, Y. (1998). Analyse et modélisation des données probabilistes : Capacités et lois multidimensionnelles. Thèse de doctorat de l'Université Paris IX-Dauphine.
- Joe, H. (1990). Multivariate concordance. *Journal of multivariate analysis*, 35, 12-30.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of multivariate analysis*, 46, 262-282.
- Johnson, N.L, Kotz, S. and Balakrishnan, N. (1997). Discrete multivariate distributions. *Wiley, New York*.
- Koehler, K.J. et Symanowski, J.T. (1995). Constructing multivariate distributions with specific marginal distributions. *Journal of multivariate analysis*, 55, 261-282.
- Krumpalauer, F. (1998). Representation of multivariate discrete distributions by probability generating functions. *Statistics and Probability letters*, 39, 327-331.
- Nataf, A. (1962). Détermination des Distributions de probabilités dont les marges sont données. *C. R. A. S. Paris, A 255*, 42-43.
- Nelsen, R.B. (1986). Properties of one-parameter family of bivariate distribution with specified marginals. *Communication in statistics, A. 15*, 3277-3285.
- Schweizer, B. et Sklar, A. (1983). Probabilistic metric spaces. *North-Holland, Amsterdam*.

Sklar, A. (1973). Random variables, joint distribution functions, and Copulas. *Kybernetika*, 9, 449-460.



Unit e de recherche INRIA Lorraine, Technop ole de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS L ES NANCY
Unit e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit e de recherche INRIA Rh one-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

 diteur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399