

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Estimation bayésienne nonparamétrique de
copules

par

Simon Guillotte

Département de mathématiques et de statistique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Statistique

mars 2008

© Simon Guillotte, 2008



Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

**Estimation bayésienne nonparamétrique de
copules**

présentée par

Simon Guillotte

a été évaluée par un jury composé des personnes suivantes :

Alejandro Murua

(président-rapporteur)

François Perron

(directeur de recherche)

Mylène Bédard

(membre du jury)

Christian Genest

(examineur externe)

Alejandro Murua

(représentant du doyen de la FES)

Thèse acceptée le:

19 Février 2008

SOMMAIRE

Cette thèse est le fruit de deux projets sur lesquels j'ai travaillé durant les deux dernières années. Ces projets ont mené à deux articles, le premier a été accepté à la Revue Canadienne de Statistique et le second sera soumis prochainement.

Premier Article

Toute distribution bivariée dont les marges sont continues peut être décomposée via sa copule et ses distributions marginales. Dans le cas des distributions de valeurs extrêmes, la copule est caractérisée par une fonction de dépendance tandis que chacune des marges est définie par trois paramètres. Dans cet article, nous proposons une approche bayésienne pour l'estimation simultanée de la fonction de dépendance et des lois marginales. Un modèle non paramétrique est construit pour la fonction de dépendance et un algorithme MCMC à sauts réversibles pour le calcul de l'estimateur bayésien est proposé. Des comparaisons sont effectuées avec les estimateurs classiques. Nous observons que notre estimateur est plus performant que les compétiteurs en terme de l'intégrale des erreurs quadratiques moyennes. Ceci est une qualité importante, surtout lorsque l'on travaille avec des valeurs extrêmes. Enfin, nous illustrons l'utilité de notre approche en l'appliquant sur un jeu de données hydrologiques et effectuons des prévisions en utilisant un algorithme à sauts réversibles.

Second Article

Toute distribution bivariée dont les marges sont continues peut être décomposée via sa copule et ses distributions marginales. Nous construisons un sous-espace

d'approximation de dimension finie, à l'intérieur de l'espace des copules. Les éléments de ce sous-espace sont paramétrisés par une matrice doublement stochastique. L'estimation est donc effectuée à l'intérieur de l'espace des matrices doublement stochastiques également connu sous le nom de polytope de Birkhoff. À l'aide du modèle, nous pouvons construire des estimateurs qui demeurent à l'intérieur de la classe des copules. Une approche bayésienne est considérée. Cette dernière a l'avantage de permettre l'estimation simultanée des lois marginales et de la copule lorsque l'intérêt porte sur la distribution bivariée. Les estimateurs de Bayes associés à plusieurs lois a priori sont étudiés à l'aide de méthodes Monte-Carlo par chaînes de Markov. Une simulation est effectuée en considérant des modèles paramétriques connus pour la copule et les marges. Dans plusieurs cas, les résultats indiquent que certains estimateurs bayésiens considérés sont plus performants que les estimateurs classiques en terme de l'intégrale des erreurs quadratiques moyennes.

MOTS CLÉS

Bayes, copules, distribution bivariée de valeurs extrêmes, fonction de dépendance, Gibbs, matrices doublement stochastiques, MCMC, Metropolis-Hastings, non paramétrique, polytope de Birkhoff, prévisions, sauts réversibles.

SUMMARY

This thesis is based on two different projects that I have worked on during the last two years. These projects lead to two papers, the first one is accepted for publication in the Canadian Journal of Statistics and the second one should be submitted soon.

First Article

A bivariate distribution with continuous margins can be decomposed via a copula and its marginal distributions. In the case of extreme value distributions, the copula is characterized by a dependence function while each of the margins depends on three parameters. In this article we propose a Bayesian approach for the simultaneous estimation of the dependence function and the parameters defining the marginal distributions. A nonparametric model is constructed for the dependence function and a reversible jump MCMC algorithm is proposed for numerical evaluations of the Bayesian estimator. Comparisons are made with classical nonparametric estimators through an extensive simulation. We observe that our estimator outperforms every competitor in terms of mean integrated squared error, especially for small sample sizes, a much desirable quality when working with extreme values. Finally, we end this article by illustrating our estimation method on a hydrological data set and we make predictions based on the predictive distribution using a reversible jump MCMC algorithm.

Second Article

A bivariate distribution with continuous margins can be decomposed via a copula and its marginal distributions. On the space of copula functions, we construct a finite dimensional approximation subspace which is parameterized by a doubly

stochastic matrix. The estimation is done in the space of doubly stochastic matrices also known as the Birkhoff polytope. The model enables the construction of estimators which belong to the class of copulas. A Bayesian approach is undertaken which allows for simultaneous estimation of the copula and the marginal distributions. Estimators based on various priors are analyzed via Markov chain Monte Carlo methodology. A rather extensive simulation experiment is carried out using data generated from parametric families of copulas. In many cases, the results indicate that the estimators obtained from the Bayesian approach outperforms the standard kernel approach in terms of mean integrated squared error.

KEY WORDS

Bayes, Birkhoff polytope, Bivariate extreme value distribution, copulas, dependence function, doubly stochastic matrices, Gibbs, MCMC, Metropolis-Hastings, nonparametric, predictions, reversible jumps.

TABLE DES MATIÈRES

Sommaire	iii
Summary	v
Remerciements	x
Introduction	2
0.1. Les Copules.....	2
0.1.1. Copules archimédiennes.....	5
0.1.2. Copules de valeurs extrêmes.....	6
0.2. Méthodes de Monte-Carlo par chaînes de Markov.....	7
0.2.1. Le problème.....	7
0.2.2. Terminologie.....	7
0.2.3. Algorithme de Metropolis-Hastings.....	10
0.2.4. Algorithme de Metropolis-Hastings à sauts réversibles.....	11
0.2.5. Algorithme hybride Metropolis-Gibbs.....	14
0.3. Principales contributions de la thèse.....	15
0.3.1. Estimation des copules de valeurs extrêmes.....	15
0.3.2. Estimation des copules générales.....	16
Bibliographie	19
Chapitre 1. A Bayesian Estimator for the Dependence Function of a Bivariate Extreme Value Distribution	21
1.1. Introduction.....	24

1.2. The model	26
1.2.1. Known margins case	27
1.2.1.1. The approximation space	27
1.2.1.2. The Bayesian estimator	27
1.2.2. Unknown margins case	28
1.3. Simulation Experiments	29
1.4. Application to a hydrological data set	31
Discussion and practical considerations	37
Acknowledgments	39
Appendix	39
1.4.1.	39
1.4.2.	41
Bibliographie	44
Chapitre 2. Bayesian Estimation of a Bivariate Copula Function	46
2.1. Introduction	49
2.2. The approximation space for the copula function	51
2.2.1. Construction	51
2.2.2. The prior	53
2.3. The approximation space for the marginal distributions	55
2.3.1. Construction	55
2.3.2. The prior	56
2.4. Simulation Experiments	56
2.5. Appendix	58
2.5.1. Proofs	58
2.5.2. Algorithms	61

Bibliographie	66
Conclusion	68

REMERCIEMENTS

Je tiens d'abord à remercier mes parents, Hélène et Benoit, qui m'ont encouragé tout au long de mes études et plus particulièrement au doctorat. Ils ont toujours été présents pour moi, ils s'informaient et s'intéressaient à chacune des étapes du parcours. Ils ont su me reconforter dans les moments plus difficiles en m'apportant un support moral inégalé. Cette thèse leur est dédiée.

Mon directeur, François Perron, a été un mentor pour moi. Au-delà des connaissances, il m'a transmis un savoir-faire et m'a appris ce qu'était la recherche. Harmonisant originalité et rigueur scientifique, sa compréhension profonde de la statistique et sa vivacité d'esprit font de lui un chercheur exceptionnel. Je suis privilégié de l'avoir eu comme superviseur, m'offrant de la disponibilité à souhait et s'intéressant constamment à mes travaux. Il a toujours su m'offrir d'excellentes pistes lorsque mes idées accrochaient ou étaient moins claires. Il m'a également beaucoup encouragé avec ses précieux "c'est bon... lâche pas!" ou encore "c'est original!", qui étaient de véritables stimulants pour moi. J'espère devenir un chercheur à son image.

Je voudrais remercier Monsieur Roch Roy qui a cru en moi dès le début et qui m'a encouragé à entreprendre ces études doctorales. Sans lui, rien de ceci n'aurait vu le jour. Je tiens également à remercier Christian Léger pour ses bons conseils, Yves Lepage pour de nombreuses lettres de référence et Alejandro Murua pour sa grande gentillesse et son intérêt pour de futures collaborations. Aussi, la toute nouvelle professeure Mylène Bédard pour ses nombreux encouragements et le partage de son expérience en tant que récente finissante. Enfin, Jonathan Taylor qui était de passage à Montréal, pour l'intérêt qu'il m'a démontré concernant mes travaux et son désir de collaborer dans un futur rapproché.

Ayant effectué la totalité de mes études au département de mathématiques et statistique, il est temps pour moi d'aller découvrir de nouveaux horizons. Je garderai cependant un excellent souvenir du département qui possède un corps professoral dynamique. J'aimerais ainsi remercier tous les professeurs que j'ai eus, plus particulièrement Dana Schlomiuk, Martin Goldstein, André Giroux, Abraham Broer, Richard Duncan, Paul Gauthier, Khalid Benabdallah, Jacques Bélair, François Perron, Roch Roy, Yves Lepage, Christian Léger et Pierre Duchesne, pour leur enthousiasme et leur enseignement stimulant. On retrouvera sûrement un peu de chacun d'entre vous dans ma façon d'enseigner.

J'aimerais enfin remercier la Faculté des Études Supérieures de l'Université de Montréal, le Laboratoire de Statistique, mon directeur de recherche et l'ISM pour leur support financier.

INTRODUCTION

Mes travaux de recherche ont touché à plusieurs sujets, notamment aux copules, à l'estimation bayésienne et aux méthodes Monte-Carlo par chaînes de Markov. Dans cette section, j'introduis le lecteur à la terminologie ainsi qu'aux divers algorithmes et résultats sur lesquels reposent mes travaux. Je termine ensuite la section en présentant les apports originaux de la thèse. Bien qu'elle soit nécessaire à la bonne compréhension des deux articles qui font l'objet de cette thèse, cette introduction n'est pas exhaustive. Le lecteur est donc encouragé à consulter la liste de références au besoin.

0.1. LES COPULES

Afin de faire ressortir le concept de copule, je présente d'abord un exemple illustratif. Supposons qu'une même expérience soit répétée à deux reprises, de manière indépendante. On s'intéresse à une statistique X dont la fonction de répartition est F . Supposons de plus que les résultats x_1 et x_2 de la statistique X nous soient initialement indisponibles. On nous communique alors x_1 , puis on nous demande d'inférer le résultat x_2 . Sans aucune autre information concernant la nature de x_1 , nous pouvons supposer que les variables aléatoires x_1 et x_2 sont indépendantes et identiquement distribuées ayant F comme fonction de répartition marginale commune. Notre inférence sera alors basée sur la loi conjointe qui a comme fonction de répartition

$$F(x_1, x_2) = F(x_1)F(x_2).$$

Si dans une autre situation, l'on nous informe que x_1 représente toujours le plus petit et x_2 le plus grand des deux résultats, alors la relation conjointe entre x_1

et x_2 sera différente et notre inférence sera plutôt basée sur la loi conjointe ayant comme fonction de répartition

$$\begin{aligned}
F(x_1, x_2) &= 2F(\min\{x_1, x_2\})F(x_2) - F(\min\{x_1, x_2\})^2, \\
&= 2 \min\{F(x_1), F(x_2)\}F(x_2) - \min\{F(x_1), F(x_2)\}^2, \\
&= 2 \min\{1 - \sqrt{1 - (2F(x_1) - F(x_1)^2)}, \sqrt{F(x_2)^2}\} \sqrt{F(x_2)^2} \\
&\quad - \min\{1 - \sqrt{1 - (2F(x_1) - F(x_1)^2)}, \sqrt{F(x_2)^2}\}^2.
\end{aligned}$$

En notant que dans la seconde situation les lois marginales de x_1 et x_2 ont comme fonctions de répartition $2F - F^2$ et F^2 respectivement, on remarque que dans les deux cas la loi conjointe dépend des fonctions de répartition marginales par l'intermédiaire d'une certaine fonction. On peut alors se demander si ce sera toujours le cas et la réponse à cette question est affirmative. En effet, le théorème de Sklar (1959), énoncé ci-bas, affirme que toute fonction de répartition conjointe s'écrit en terme de ses fonctions de répartition marginales et d'une certaine fonction de dépendance liant celles-ci ensemble. Cette fonction de dépendance est appelée copule.

Définition 0.1.1. *Une copule (bivariée) est une fonction C définie sur le carré $S = [0, 1] \times [0, 1]$ satisfaisant :*

- les conditions aux bords

$$C(0, v) = 0 = C(u, 0), C(u, 1) = u, C(1, v) = v, \quad (u, v) \in S,$$

- la monotonie

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0, \quad u_1 \leq u_2, v_1 \leq v_2.$$

Les copules sont des fonctions lipschitziennes, de constante 1, et forment une famille équicontinue. Afin d'énoncer le théorème de Sklar, il est convenable de définir la fonction de répartition inverse généralisée d'une variable aléatoire.

Définition 0.1.2. *Soit $F : [-\infty, +\infty] \rightarrow [0, 1]$, une fonction de répartition. La fonction de répartition inverse généralisée $F^{-1} : [0, 1] \rightarrow [-\infty, +\infty]$ est définie par*

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad u \in [0, 1].$$

Cette fonction possède notamment les propriétés suivantes

- (1) $F^{-1}(u) \leq x \iff u \leq F(x)$,
- (2) $F \circ F^{-1}(u) \geq u$, $u \in [0, 1]$, avec égalité si et seulement si $u \in \text{Im}(F)$.

En deux dimensions, le théorème de Sklar s'énonce ainsi

Théorème 0.1.1. *Soient X_1 et X_2 des variables aléatoires de fonctions de répartition marginales F_1 et F_2 et soit F , la fonction de répartition conjointe de (X_1, X_2) . Il existe une copule C telle que*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)).$$

De plus, C est uniquement déterminée sur $\text{Im}(F_1) \times \text{Im}(F_2)$ par

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)).$$

En particulier, la copule est unique si F_1 et F_2 sont continues.

Le théorème de Sklar se généralise aisément aux dimensions supérieures. La famille des copules est uniformément bornée

$$\max\{0, u + v - 1\} \leq C(u, v) \leq \min\{u, v\}, \quad (u, v) \in S \quad (0.1.1)$$

et $W(u, v) = \max\{0, u + v - 1\}$, $M(u, v) = \min\{u, v\}$ sont appelées copules frontières de Fréchet-Hoeffding, voir Nelsen (2006). Ainsi, la structure de dépendance d'un vecteur aléatoire est une fonction se situant quelque part entre W et M . En fait, étant donné un vecteur aléatoire (X_1, X_2) , chacune des variables X_1 et X_2 est presque sûrement exprimable comme fonction strictement décroissante (croissante) de l'autre si et seulement si $C = W$ (M), voir Schweizer and Sklar (2005). Le lecteur est référé à Nelsen (2006) pour la généralisation des résultats ci-haut aux dimensions supérieures. Bien que l'inégalité (0.1.1) soit toujours valide, la borne inférieure W ne demeure plus une copule pour $n \geq 3$.

Développer des outils permettant de quantifier la dépendance entre des variables aléatoires est d'un intérêt primordial en sciences appliquées. Une mesure de dépendance courante est la corrélation linéaire. Pour la famille des lois binormales, cette mesure est particulièrement intéressante puisque seule la connaissance de celle-ci et des variances suffit pour la caractériser. De plus, pour cette famille de lois, une corrélation nulle est équivalente à l'indépendance. Ces propriétés ne sont

cependant plus vraies lorsque l'on quitte la famille des lois binormales. La corrélation linéaire possède certaines autres limitations, notamment du fait qu'elle soit définie seulement si les variances sont elles-mêmes finies. Les actuaires travaillant avec des distributions à queues épaisses peuvent faire face à de telles situations. Une autre difficulté de cette mesure provient du fait qu'elle ne soit pas invariante par rapport aux transformations strictement croissantes non linéaires des marges. En fait, ceci signifie qu'un simple changement d'échelle, logarithmique par exemple, peut modifier la valeur de la corrélation. Le prochain résultat, dont la démonstration est immédiate, nous montre d'autre part que la copule est invariante par rapport aux transformations strictement croissantes des marges.

Théorème 0.1.2. *Soient X_1 et X_2 des variables aléatoires de fonctions de répartition marginales F_1 et F_2 continues et de copule C . Si φ_1 et φ_2 sont des fonctions strictement croissantes, alors C est également la copule associée au vecteur $(\varphi_1(X_1), \varphi_2(X_2))$.*

Une conséquence de ce résultat est le fait que toute fonctionnelle de la loi conjointe, qui est invariante par rapport aux transformations strictement croissantes des marges est, en fait, une fonctionnelle de la copule, voir Schweizer and Sklar (2005). Les versions échantillonnales de telles statistiques ne dépendront alors que des rangs.

0.1.1. Copules archimédiennes

Cette famille de copules est parmi les plus utilisées dans la pratique, notamment en finance et en actuariat. Elles sont de la forme

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)),$$

où ϕ est une fonction dite génératrice, définie sur l'intervalle $[0, 1]$, continue, convexe, décroissante, avec $\phi(1) = 0$. On note que ϕ est unique à une constante multiplicative près. La copule de Gumbel en est un exemple, sa fonction génératrice est donnée par

$$\phi(t) = (-\log t)^\theta, \theta \geq 1.$$

Le nom de cette famille provient du fait que l'opération $*$: $(u, v) \mapsto u * v = C(u, v)$ définit une structure de semi-groupe abélien sur $[0, 1]$. Ce semi-groupe est ordonné et satisfait la propriété archimédienne : $\forall u, v \in (0, 1)$, il existe $n \in \mathbb{N}$ tel que $u^n < v$.

Plusieurs estimateurs de telles copules apparaissent dans la littérature et une approche bayésienne fut tout récemment proposée, voir Lambert (2007). L'auteur propose un estimateur pour la fonction $\lambda(t) = \phi(t)/\phi'(t)$, il est basé sur la représentation

$$\phi(t) = \exp\left(-\int_t^1 \frac{d\tau}{\lambda(\tau)}\right), 0 < t \leq 1.$$

0.1.2. Copules de valeurs extrêmes

Une autre famille de copules très utilisée concerne la distribution des valeurs extrêmes d'un échantillon. Dans le premier chapitre de la thèse, nous proposons un estimateur bayésien non paramétrique pour cette famille. Nous la décrivons brièvement dans ce qui suit.

Tout comme la loi normale apparaît comme distribution limite des moyennes échantillonnales normalisées, les lois bivariées de valeurs extrêmes apparaissent comme distributions limites des maximums ou minimums échantillonnaux normalisés. Plus précisément, soient $(X_1, Y_1), \dots, (X_n, Y_n)$, des vecteurs aléatoires indépendants et identiquement distribués, et posons $M_n = (\max X_i, \max Y_i)$. Une fonction de répartition bivariée F est dite de valeurs extrêmes s'il existe des suites normalisantes $a_n = (a_n^{(1)}, a_n^{(2)}) > 0$ et $b_n = (b_n^{(1)}, b_n^{(2)}) \in \mathbb{R}^2$, telles que $(M_n - b_n)/a_n \rightarrow F$ en loi. Notons qu'il peut exister plusieurs suites normalisantes pour lesquelles on a la convergence faible et le choix d'une suite particulière n'affecte que les paramètres de position et d'échelle des lois limites marginales. Ainsi, le choix d'une suite particulière n'a aucune importance en pratique, étant donné que l'estimation s'effectue en utilisant la suite dénormalisée, voir Coles (2001). On a la caractérisation suivante, voir Deheuvels (1984)

Théorème 0.1.3. *Une loi bivariée est une loi de valeurs extrêmes si et seulement si :*

1. ses distributions marginales sont des lois de valeurs extrêmes univariées, soient Gumbel, Fréchet ou Weibull, formant une seule famille paramétrisée par trois paramètres : de forme, de position et d'échelle,

2. sa copule s'écrit

$$C(u, v) = \exp \left\{ \log(uv) A \left(\frac{\log(u)}{\log(uv)} \right) \right\}, \forall u, v \in [0, 1], \quad (0.1.2)$$

où $A(\cdot)$ est convexe avec $\max(t, 1-t) \leq A(t) \leq 1$, $0 \leq t \leq 1$.

Une copule de la forme (0.1.2) ci-haut est dite copule de valeurs extrêmes. Un exemple de telles copules est donné par la famille paramétrique de Gumbel. Remarquablement, la famille paramétrique de Gumbel caractérise les copules qui sont à la fois archimédienne et de valeurs extrêmes, voir Genest and Rivest (1989).

0.2. MÉTHODES DE MONTE-CARLO PAR CHAÎNES DE MARKOV

0.2.1. Le problème

Supposons que l'on désire simuler des observations d'une certaine densité de probabilité π , mais que sa complexité rend la tâche difficile, voire même impossible avec les méthodes standards. Les méthodes de Monte-Carlo par chaînes de Markov (MCMC) offrent une solution indirecte à ce problème. Elles se basent sur l'idée qu'il est plus facile de construire une chaîne de Markov ergodique qui converge vers la loi stationnaire π , que de simuler directement de π . Cette construction est possible grâce au populaire algorithme introduit par Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953), dans un contexte de physique statistique et généralisé ensuite par Hastings, voir Hastings (1970).

0.2.2. Terminologie

Une chaîne de Markov X_n , $n \geq 0$ sur un espace d'états \mathcal{X} (typiquement \mathbb{R}^d) est définie à l'aide d'une mesure de probabilité initiale μ pour X_0 et d'un noyau de transition K : pour $x \in \mathcal{X}$, $B \subseteq \mathcal{X}$, $K(x, B)$ représente la probabilité que la chaîne passe de x à B en une itération. Plus généralement, le noyau de la n^e

itération est défini par

$$K^n(x, B) = \int_{\mathcal{X}} K^{n-1}(x, dy)K(y, B), n \geq 1,$$

avec $K^0(x, dy) = \delta_x(dy)$, la mesure de Dirac au point x .

Définition 0.2.1. Soit π une mesure de probabilité sur \mathcal{X} et K un noyau de transition. La mesure de probabilité π est dite invariante pour K si

$$\pi(dy) = \int_{\mathcal{X}} K(x, dy)\pi(dx).$$

Une mesure invariante est également appelée stationnaire. On note que si X_0 est de loi π , alors X_n est de loi π , pour tout $n \geq 0$. Meyn and Tweedie (1993), présentent des conditions pour lesquelles :

- on a l'existence d'une unique mesure de probabilité stationnaire π ,
- \bar{X}_n converge vers $\int_{\mathcal{X}} x \pi(dx)$,
- X_n converge vers π , en loi.

Une chaîne possédant ces trois qualités est dite ergodique. Ainsi, pour n assez grand, la moyenne échantillonnale de la chaîne approche l'espérance selon la loi stationnaire π , au même titre que la célèbre loi des grands nombres.

Malheureusement, ce ne sont pas toutes les chaînes de Markov qui possèdent ces propriétés. À titre d'exemple, soient $\mathcal{X} = \{1, 2, 3\}$, et

$$K = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

On voit facilement que $\pi_p = (p, p, 1 - 2p)$, $0 \leq p \leq 1/2$, sont toutes stationnaires. Ici X_n converge en loi vers $\pi_{1/2}$ si $X_0 = 1$ ou $X_0 = 2$ et X_n converge en loi vers π_0 sinon. Le problème provient de la réductibilité de la chaîne.

Définition 0.2.2. Une chaîne de Markov est dite ϕ -irréductible s'il existe une mesure non-nulle ϕ telle que, quel que soit l'état initial de la chaîne, $X_0 = x$ et quel que soit $B \subseteq \mathcal{X}$ de ϕ -mesure positive, il existe $n \geq 1$ pour lequel $K^n(x, B) > 0$.

Le prochain résultat se trouve dans Meyn and Tweedie (1993), Chapitre 10.

Théorème 0.2.1. *Si une chaîne de Markov est ϕ -irréductible et possède une mesure de probabilité stationnaire π , alors celle-ci est unique.*

Une chaîne ϕ -irréductible possédant une (unique) mesure de probabilité stationnaire peut ne pas converger vers celle-ci. En effet, pour $\mathcal{X} = \{1, 2, 3\}$, et

$$K = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

on voit facilement que la chaîne est ϕ -irréductible, avec ϕ la mesure de dénombrement et $\pi = (1/3, 1/3, 1/3)$ est l'unique loi stationnaire. Cependant, si $X_0 = 1$, alors $P(X_n = 1) = 0$, si $n \not\equiv 0 \pmod{3}$ et donc X_n ne converge pas en loi vers π . Le problème provient de la périodicité de la chaîne.

Définition 0.2.3. *Une chaîne de Markov de loi stationnaire π est dite apériodique s'il n'existe pas d'entier $m \geq 2$ et une partition $\mathcal{X}_0, \dots, \mathcal{X}_{m-1}$ tels que $\pi(\mathcal{X}_1) > 0$ et*

$$x \in \mathcal{X}_i \Rightarrow K(x, \mathcal{X}_{i+1 \pmod{m}}) = 1.$$

L'apériodicité et la prochaine condition, plus forte que l'irréductibilité, sont suffisantes pour garantir la convergence de la chaîne vers la stationnarité et ce, quel que soit son point de départ.

Définition 0.2.4. *Une chaîne de Markov X_n ϕ -irréductible de loi stationnaire π est dite Harris-récurrente si pour tout $x \in \mathcal{X}$ et pour chaque $B \subseteq \mathcal{X}$, si B est de π -mesure positive, alors $P(\eta_B = \infty \mid X_0 = x) = 1$, où $\eta_B = \sum_{n \geq 1} \mathbb{1}_{X_n \in B}$.*

Le prochain résultat est le fameux théorème ergodique, voir Meyn and Tweedie (1993).

Théorème 0.2.2. *Si X_n est Harris-récurrente, de loi stationnaire π ,*

1. *alors la loi des grands nombres est valide,*
2. *si, de plus, la chaîne est apériodique, alors quelle que soit la distribution initiale μ , on a que X_n converge vers π en loi.*

0.2.3. Algorithme de Metropolis-Hastings

Soit $\pi(dx) = \pi(x)dx$ une densité cible sur \mathcal{X} . L'algorithme construit un noyau de transition K en s'assurant que la chaîne résultante soit ergodique avec π comme loi stationnaire.

Définition 0.2.5. Soit \mathcal{X} un espace d'états, un noyau de transition K est dit réversible par rapport à une mesure de probabilité π sur \mathcal{X} si

$$\forall A, B \subseteq \mathcal{X}, \int_A \pi(dx)K(x, B) = \int_B \pi(dx)K(x, A). \quad (0.2.1)$$

Notons que la réversibilité est une condition suffisante pour s'assurer de la stationnarité de la mesure d'intérêt π pour le noyau K , en effet

$$\int_{\mathcal{X}} \pi(dx)K(x, B) = \int_B \pi(dx)K(x, \mathcal{X}) = \pi(B).$$

La réversibilité est généralement plus facilement vérifiable que la stationnarité et donc beaucoup d'algorithmes MCMC exploitent cette condition suffisante. L'algorithme de Metropolis-Hastings en est un exemple. À chaque itération i , soit $X_i = x$ l'état actuel de la chaîne. On introduit deux quantités :

- $q(x, \cdot)$: une densité conditionnelle proposant un nouvel état,
- $\alpha(x, \cdot)$: représentant la probabilité d'acceptation du nouvel état (à préciser ci-bas).

Le noyau de la chaîne se décompose alors en deux parties, c'est-à-dire,

$$K(x, B) = \underbrace{\int_B q(x, y)\alpha(x, y)dy}_{(1)} + \underbrace{\left(1 - \int_{\mathcal{X}} q(x, y)\alpha(x, y)dy\right) \mathbb{1}_B(x)}_{(2)}$$

où

- (1) est la probabilité de quitter l'état actuel et se retrouver dans B : elle est absolument continue par rapport à la mesure de Lebesgue sur \mathcal{X} ,
- (2) est la probabilité de rester dans l'état actuel s'il est déjà dans B : c'est la partie discrète du noyau.

En appliquant la condition de réversibilité sur K ci-haut, on trouve que celle-ci sera satisfaite pourvu que

$$\int_A \pi(dx) \int_B q(x, y)\alpha(x, y)dy = \int_B \pi(dy) \int_A q(y, x)\alpha(y, x)dx$$

et que la probabilité d'acceptation définie par

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right\}, \quad (0.2.2)$$

nous donne l'égalité désirée. L'algorithme s'écrit formellement comme ceci

Algorithme 0.2.1. *À l'itération n , soit $X_n = x$ l'état actuel de la chaîne.*

1. On génère y selon $q(x, \cdot)$,

$$2. X_{n+1} = \begin{cases} y & \text{avec probabilité } \alpha(x, y), \\ x & \text{avec probabilité } 1 - \alpha(x, y), \end{cases}$$

où

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right\}.$$

Le théorème suivant nous assure de l'ergodicité de la chaîne, voir Roberts and Rosenthal (2006).

Théorème 0.2.3. *Si la chaîne de Metropolis-Hastings est π -irréductible, alors elle est Harris-récurrente.*

On note ici que la π -irréductibilité revient à celle du noyau instrumental $q(x, \cdot)$ et qu'une condition suffisante pour garantir l'apériodicité est que l'algorithme permette à certains états la possibilité de rester sur place. Ces conditions sont facilement vérifiables en général, expliquant d'une part la grande popularité de l'algorithme.

On retrouve notamment les versions suivantes :

– symétrique : lorsque le noyau instrumental $q(x, y) = q(y, x)$, la probabilité d'acceptation se simplifie

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\},$$

– indépendante : lorsque $q(x, y) = q(y)$, i.e. $q(x, \cdot)$ ne dépend pas de x ,

– marche aléatoire : lorsque $q(x, y) = q(|y - x|)$, comme par exemple $q(x, \cdot)$ suit une loi normale $N(x, \sigma^2)$ ou $q(x, \cdot)$ suit une loi uniforme $U(x - a, x + a)$.

0.2.4. Algorithme de Metropolis-Hastings à sauts réversibles

Dans certains contextes, notamment en statistique bayésienne, la dimension de la loi cible n'est pas connue à l'avance. Green (1995) propose une solution adaptant le noyau de Metropolis-Hastings sur un espace général constitué d'une

union disjointe de sous-espaces de différentes dimensions. Supposons que \mathcal{K} est l'ensemble indexant les dimensions, l'espace généralisé s'écrit alors

$$\mathcal{X} = \bigcup_{k \in \mathcal{K}} \{k\} \times \mathbb{R}^{n_k},$$

et la loi stationnaire s'écrit $\pi = \pi(x) = \pi(k, y)$. Plusieurs types de déplacements doivent alors être définis pour pouvoir explorer l'espace au complet. L'ensemble des déplacements est probabilisé et à chaque itération, l'un d'eux est proposé. Supposons qu'un déplacement de type $k \rightarrow k'$ propose de transformer $x = (k, y)$ en $x' = (k', y')$, où $y \in \mathbb{R}^{n_k}$ et $y' \in \mathbb{R}^{n_{k'}}$. La condition de réversibilité (0.2.1) sera alors satisfaite si pour chaque type de déplacement, on a

$$\int_A \pi(dx) \int_B q_{kk'}(x, dx') \alpha_{kk'}(x, x') = \int_B \pi(dx') \int_A q_{k'k}(x', dx) \alpha_{k'k}(x', x), \quad (0.2.3)$$

où $q_{kk'}(x, dx')$ décrit le noyau de transition correspondant au type de déplacement $k \rightarrow k'$ et de la destination x' . Tout comme l'algorithme standard, cette fois de manière informelle, la probabilité d'acceptation s'écrit

$$\alpha(x, x') = \min \left\{ 1, \frac{\pi(dx') q_{k'k}(x', dx)}{\pi(dx) q_{kk'}(x, dx')} \right\}.$$

Contrairement à l'algorithme standard, x et x' peuvent appartenir à des espaces de dimensions différentes, rendant plus délicate l'obtention de densités comparables, i.e. par rapport à une même mesure dominante, pour les lois conjointes $\pi(dx) q_{kk'}(x, dx')$ et $\pi(dx') q_{k'k}(x', dx)$. La difficulté se situe dans la recherche d'une mesure symétrique sur $\mathcal{X} \times \mathcal{X}$ qui domine les deux lois. L'auteur propose alors une telle mesure et obtient les densités désirées. En particulier, une probabilité d'acceptation $\alpha_{kk'}(x, x')$, validant l'égalité (0.2.3), est obtenue comme suit. Supposons, pour fixer les idées, que $n_k < n_{k'}$. Le passage de $x = (k, y)$ à $x' = (k', y')$ est effectué

1. en générant un vecteur u de longueur $n_{k'} - n_k$, selon une loi $g_{kk'}(\cdot)$,
2. en posant $y' = T(y, u)$, où T est un difféomorphisme sur $\mathbb{R}^{n_{k'}}$.

La probabilité d'acceptation devient alors

$$\alpha_{kk'}(x, x') = \min \left\{ 1, \frac{\pi(x') j_{k'k}(x')}{\pi(x) j_{kk'}(x) g_{kk'}(u)} \left| \frac{\partial y'}{\partial (y, u)} \right| \right\},$$

où $j_{kk'}(x)$ représente la probabilité de choisir le déplacement $k \rightarrow k'$.

À titre d'exemple, dans un contexte bayésien, supposons que nous soyons incertain de la dimension du paramètre θ d'un modèle d'intérêt. Pour simplifier, on suppose que le paramètre θ puisse être d'une seule ou de deux dimensions. La loi a posteriori est donc de la forme $\pi(k, \theta|x)$, où $k = 1, 2$ et x est le vecteur des données. Supposons que nous désirions simuler de la loi a posteriori à l'aide d'une chaîne de Markov. On a donc deux espaces : $\{1\} \times \mathbb{R}$ et $\{2\} \times \mathbb{R}^2$ et l'on veut construire une chaîne de Metropolis-Hastings pouvant se promener non seulement à l'intérieur d'un même espace, mais d'un espace à l'autre. Posons

$$\Theta = \{1\} \times \mathbb{R} \cup \{2\} \times \mathbb{R}^2,$$

et admettons que l'état actuel de la chaîne soit $(1, \theta)$. Supposons à présent que le déplacement de $\{1\} \times \mathbb{R}$ vers $\{2\} \times \mathbb{R}^2$ ait été choisi avec probabilité $j_{12}(1, \theta)$. On désire alors proposer un état $(2, \theta_1, \theta_2)$. L'approche de Green consiste d'abord à équilibrer de manière artificielle les dimensions des deux espaces en générant une variable aléatoire u selon une loi $g_{12}(\cdot)$. Ensuite, on transforme (θ, u) en (θ_1, θ_2) selon un certain difféomorphisme T qui est approprié pour le contexte. Dans le présent exemple, supposons que la transformation

$$T : (\theta, u) \mapsto (\theta_1, \theta_2) = (\theta - u, \theta + u)$$

soit pertinente. Sa transformation inverse est

$$T^{-1} : (\theta_1, \theta_2) \mapsto (\theta, u) = \left(\frac{\theta_1 + \theta_2}{2}, \frac{\theta_2 - \theta_1}{2} \right).$$

Selon la mesure que Green propose, la densité conjointe du déplacement de θ vers (θ_1, θ_2) est

$$\pi(1, \theta|x) j_{12}(1, \theta) g_{12}(u) \left| \frac{\partial(\theta, u)}{\partial(\theta_1, \theta_2)} \right|,$$

tandis que celle du déplacement inverse est

$$\pi(2, \theta_1, \theta_2|x) j_{21}(2, \theta_1, \theta_2).$$

La probabilité d'acceptation du déplacement est donc

$$\min \left\{ 1, \frac{\pi(2, \theta_1, \theta_2|x) j_{21}(2, \theta_1, \theta_2)}{\pi(1, \theta|x) j_{12}(1, \theta) g_{12}(u)} \cdot 2 \right\}.$$

Le lecteur est référé à Roberts and Rosenthal (2006) pour les propriétés ergodiques de l'algorithme décrit dans la cette section.

0.2.5. Algorithme hybride Metropolis-Gibbs

Supposons que $\mathcal{X} = \mathbb{R}^d$ et considérons la densité cible $\pi = \pi(x_1, \dots, x_d)$. Si l'on peut aisément simuler des observations des densités conditionnelles $\pi(x_i | x_{(-i)})$, $i = 1, \dots, d$, l'algorithme de Gibbs standard produit une chaîne ergodique convergant vers π , en simulant tour à tour de chacune de ses lois conditionnelles. Plus précisément,

Algorithme 0.2.2. À l'itération n , soit $X_n = x = (x_1, \dots, x_d)$, l'état actuel de la chaîne, pour $i = 1, \dots, d$, répéter

1. générer $x'_i \sim \pi(\cdot | x_{(-i)})$,
2. remplacer la $i^{\text{ième}}$ coordonnée de x par x'_i .

L'algorithme fut popularisé par Gelfand and Smith (1990) et le lecteur est référé à Robert and Casella (2004) pour plus de détails.

Cette méthode est intéressante pourvu que l'on puisse facilement simuler de toutes les densités conditionnelles $\pi(x_i | x_{(-i)})$, $i = 1, \dots, d$. Sinon, on a recours à un algorithme hybride Metropolis-Hastings dans Gibbs. En fait, l'algorithme précédent peut être vu comme un algorithme de Metropolis-Hastings avec probabilités d'acceptation uniformément égales à 1, voir par exemple Robert and Casella (2004). Dans l'algorithme hybride, on exploite cette représentation avec la modification suivante : les sous-itérations sont effectuées en utilisant des lois instrumentales $q_i(\cdot | x_{(-i)})$, $i = 1, \dots, d$, à la place des lois conditionnelles et les nouveaux états sont acceptés avec la probabilité d'acceptation de Metropolis-Hastings.

Algorithme 0.2.3. À l'itération n , soit $X_n = x = (x_1, \dots, x_d)$, l'état actuel de la chaîne, pour $i = 1, \dots, d$, répéter

1. générer $x'_i \sim q_i(\cdot | x_{(-i)})$,
2. poser x' le vecteur obtenu de x en remplaçant sa $i^{\text{ième}}$ coordonnée par x'_i ,
3. accepter $x = x'$ avec probabilité

$$\begin{aligned} \alpha(x, x') &= \min \left\{ 1, \frac{\pi(x') q_i(x_i | x_{(-i)})}{\pi(x) q_i(x'_i | x_{(-i)})} \right\}, \\ &= \min \left\{ 1, \frac{\pi(x'_i | x_{(-i)}) q_i(x_i | x_{(-i)})}{\pi(x_i | x_{(-i)}) q_i(x'_i | x_{(-i)})} \right\}. \end{aligned}$$

Le théorème suivant est démontré dans Roberts and Rosenthal (2006).

Théorème 0.2.4. *Si la chaîne hybride Metropolis-Gibbs est ϕ -irréductible et si quel que soit l'état initial x , avec probabilité 1, la chaîne se déplacera au moins une fois dans chacune des directions des coordonnées, alors elle est Harris-récurrente.*

0.3. PRINCIPALES CONTRIBUTIONS DE LA THÈSE

0.3.1. Estimation des copules de valeurs extrêmes

Soit \mathcal{D} l'espace des fonctions de dépendance. Dans le cas où les distributions marginales sont connues, l'estimateur classique de la fonction de dépendance $A \in \mathcal{D}$, proposé dans Capéraà, Fougères, and Genest (1997), est construit en transformant le couple bivarié (U, V) en une statistique Z , qu'ils nomment pseudo-observation. Les auteurs exploitent le fait que la distribution de Z fait intervenir la fonction de dépendance A à l'intérieur d'une équation différentielle. Cependant la statistique Z n'est pas exhaustive. Ainsi, en n'utilisant qu'une partie de l'information, les auteurs parviennent à de belles expressions asymptotiques pour leur estimateur. Nous n'avons pas suivi cette optique. Nous nous sommes plutôt intéressés à obtenir un estimateur qui soit plus performant pour des tailles échantillonales finies. Pour des échantillons de tailles finies, certains estimateurs existant doivent être corrigés afin de donner des courbes de dépendance en considérant par exemple l'enveloppe convexe et en ramenant l'estimateur sur la frontière lorsque ce dernier quitte le triangle. Ceci donne lieu à des estimations qui paraissent plus ou moins naturelles. L'approche bayésienne remédie à ce problème. En effet, il suffit de considérer des lois a priori se concentrant sur un sous-ensemble de courbe de dépendance se situant à l'intérieur du triangle.

Pour ce faire, nous construisons tout d'abord un sous-espace d'approximation $\Lambda_K, K \geq 3$ de dimension $2K$ pour les fonctions de dépendance. Les courbes de ce sous-espace sont paramétrisées par des noeuds d'interpolation θ . On considère alors

$$\Theta_K = \{(\theta_1, \dots, \theta_K) \in \mathbb{R}^{2K} : \theta_i = (t_i, A(t_i)), i = 1, \dots, K, 0 = t_1 < t_2 < \dots < t_K = 1, \\ \text{pour une fonction de dépendance strictement convexe } A\}.$$

et pour chaque $\theta \in \Theta_K$, on pose

$$\mathcal{D}_\theta = \{\varphi \in \mathcal{D} : \theta \text{ est interpolé par } \varphi\}.$$

Nous démontrons alors le résultat d'approximation suivant :

$$\sup_{A \in \mathcal{D}} \inf_{\theta \in \Theta_K} \sup_{\varphi \in \mathcal{D}_\theta} \|A - \varphi\|_\infty \leq \frac{1}{2(K-1)}.$$

Ainsi, afin d'approximer une courbe $A \in \mathcal{D}$, le choix de $\theta \in \Theta_K$ est critique, tandis que le choix particulier d'une courbe représentante $\varphi \in \mathcal{D}_\theta$ ne l'est pas. Afin de bien définir Λ_K , nous donnons une construction pour les courbes représentantes garantissant l'existence d'une densité pour la copule de valeurs extrêmes associée.

Nous désirons obtenir l'estimateur bayésien sous une perte quadratique. Nous travaillons sur une fonction de dépendance aléatoire $A \in \Lambda_K$. La distribution de A est donnée par une loi a priori. Lorsque les fonctions de répartition marginales F_1 et F_2 sont connues, la distribution de l'échantillon ne dépend que de A . La loi a priori pour A et l'échantillon mènent à la loi a posteriori. Une estimation \hat{A} est donnée par l'espérance selon la loi a posteriori. Cette dernière est évaluée au moyen d'un algorithme à sauts réversibles dont la forme générale est décrite à la section 0.2.4.

Pour les distributions bivariées de valeurs extrêmes, chacune des fonctions de répartition marginales F_1 et F_2 est caractérisée par trois paramètres. Lorsque ceux-ci sont inconnus, la loi a posteriori pour A dépend des lois a priori sur A , F_1 et F_2 , ainsi que de l'échantillon. Contrairement à certaines approches fréquentistes existantes, notre approche permet l'estimation simultanée de la fonction de dépendance A et des paramètres caractérisant les fonctions de répartition marginales.

Notre approche nous permet également de faire de la prévision concernant les maximums d'échantillons, ce que nous illustrons à l'aide d'un exemple en hydrologie.

0.3.2. Estimation des copules générales

Soit \mathcal{C} l'espace des copules et soit $\{(x_i, y_i) : i = 1, \dots, n\}$ un échantillon dont $(X_i, Y_i), i = 1, \dots, n$ possède une fonction de répartition conjointe F . On cherche

à estimer la copule $C \in \mathcal{C}$ par une copule \hat{C} qui dépend de l'échantillon. On dit que \hat{C} satisfait les contraintes de Deheuvels si

$$\hat{C} \left(\frac{i}{n}, \frac{j}{n} \right) = \frac{1}{n} \sum_{k=1}^n I(\text{rang}(x_k) \leq i, \text{rang}(y_k) \leq j), \text{ où } i, j = 1, \dots, n.$$

Historiquement Deheuvels (1984) est le premier à avoir traité l'estimation non paramétrique de copules. Il étudie le comportement asymptotique de la classe des copules satisfaisant la contrainte de Deheuvels. L'auteur exploite la représentation 0.1.1. En effet, remarquons que

$$\hat{C} \left(\frac{i}{n}, \frac{j}{n} \right) = \hat{F} \left(\hat{F}_X^{-1} \left(\frac{i}{n} \right), \hat{F}_Y^{-1} \left(\frac{j}{n} \right) \right), \text{ où } i, j = 1, \dots, n,$$

où \hat{F} est la fonction de répartition empirique et \hat{F}_X^{-1} et \hat{F}_Y^{-1} sont les inverses généralisés. En général, si \hat{F} est un estimateur lisse de F , alors une extension naturelle est donnée par

$$\hat{C}(u, v) = \hat{F} \left(\hat{F}_X^{-1}(u), \hat{F}_Y^{-1}(v) \right), \text{ pour chaque } (u, v) \in S. \text{ (Approche générale)}$$

Plusieurs auteurs ont considéré l'approche générale. Dans le second article, nous proposons une alternative bayésienne pour l'estimation non paramétrique des copules.

Pour ce faire, soit $m > 1$, nous considérons des lois a priori concentrées sur un sous-espace $\mathcal{A}_m \subset \mathcal{C}$ de dimension $(m-1)^2$. Les éléments de \mathcal{A}_m correspondent à des mélanges ayant m^2 composantes. Le sous-espace \mathcal{A}_m converge vers \mathcal{C} lorsque m tend vers l'infini. Cet espace d'approximation fut également étudié dans Li, Mikusiński, and Taylor (1998) et Kulpa (1999). Plus précisément, la construction de \mathcal{A}_m repose sur le choix d'une base formant une partition de l'unité. Les éléments de \mathcal{A}_m sont paramétrisés par une matrice doublement stochastique P . Le choix de la base est fixe tandis que P varie. Nous construisons également un modèle pour traiter le cas des fonctions de répartition marginales inconnues. Dans ce cas, la loi a posteriori dépend de la copule, des lois a priori sur les marges, ainsi que de l'échantillon. Contrairement à certaines approches fréquentistes existantes, notre approche permet l'estimation simultanée de la copule et des paramètres caractérisant les fonctions de répartition marginales. L'évaluation numérique de l'estimation de la copule est effectuée à l'aide d'une chaîne de Markov. Le noyau

de transition de cette chaîne correspond à un mélange de l'algorithme hybride Metropolis-Gibbs et d'un algorithme à sauts réversibles dont les formes générales apparaissent aux sections 0.2.5 et 0.2.4 respectivement.

Bibliographie

- Capéraà, P., Fougères, A.-L., Genest, C., 1997. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84, 567–577.
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Deheuvels, P., 1984. Probabilistic aspects of multivariate extremes. Reidel.
- Gelfand, P., Smith, A., 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Genest, C., Rivest, L.-P., 1989. A characterization of gumbel's family of extreme value distributions. *Statistics and Probability Letters* 8, 207–211.
- Green, P. J., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* 82, 711–732.
- Hastings, W., 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57, 97–109.
- Kulpa, T., 1999. On approximation of copulas. *International Journal of Mathematics and Mathematical Sciences* 22 (2), 259–269.
- Lambert, P., 2007. Archimedean copula estimation using bayesian splines smoothing techniques. *Computational Statistics & Data Analysis* In press.
- Li, X., Mikusiński, P., Taylor, M., 1998. Strong approximation of copulas. *Journal of Mathematical Analysis and Applications* 225 (2), 608–623.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21, 1087–1092.
- Meyn, S., Tweedie, R., 1993. *Markov Chains and stochastic stability*. Springer-Verlag.

- Nelsen, R. B., 2006. An Introduction to Copulas. Springer, New York.
- Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods. Springer.
- Roberts, G., Rosenthal, J., 2006. Harris recurrence of the metropolis-within-gibbs and trans-dimensional markov chains. *Annals of Applied Probability* 16, 2123–2139.
- Schweizer, B., Sklar, A., 2005. Probabilistic Metric Spaces. Dover.
- 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.

Chapitre 1

A BAYESIAN ESTIMATOR FOR THE DEPENDENCE FUNCTION OF A BIVARIATE EXTREME VALUE DISTRIBUTION

ABSTRACT

A bivariate distribution with continuous margins can be decomposed via a copula and its marginal distributions. In the case of extreme value distributions, the copula is characterized by a dependence function while each of the margins depends on three parameters. In this article the authors propose a Bayesian approach for the simultaneous estimation of the dependence function and the parameters defining the marginal distributions. A nonparametric model is constructed for the dependence function and a reversible jump MCMC algorithm is proposed for numerical evaluations of the Bayesian estimator. Comparisons are made with classical nonparametric estimators through an extensive simulation. The authors observe that the proposed estimator outperforms every competitor in terms of mean integrated squared error, especially for small sample sizes, a much desirable quality when working with extreme values. Finally, the authors end this article by illustrating the estimation method on a hydrological data set and we make predictions based on the predictive distribution using a reversible jump MCMC algorithm.

Key words and phrases : Bayesian estimator, Bivariate extreme value distribution, convex Hermite interpolation, copulas, dependence function, MCMC, Metropolis Hastings algorithm, prediction, reversible jumps.

AMS 2000 subject classifications : 62G32,62F30,65C60.

RÉSUMÉ

Toute distribution bivariée dont les marges sont continues peut être décomposée via sa copule et ses distributions marginales. Dans le cas des distributions de valeurs extrêmes, la copule est caractérisée par une fonction de dépendance tandis que chacune des marges est définie par trois paramètres. Dans cet article, les auteurs proposent une approche bayésienne pour l'estimation simultanée de la fonction de dépendance et des lois marginales. Un modèle nonparamétrique est construit pour la fonction de dépendance et un algorithme MCMC à sauts réversibles pour le calcul de l'estimateur bayésien est proposé. Des comparaisons

sont effectuées avec les estimateurs classiques. Des simulations semblent indiquer que notre estimateur est plus performant que les compétiteurs en terme de l'intégrale des erreurs quadratiques moyennes. Ceci est une qualité importante, surtout lorsque l'on travaille avec des valeurs extrêmes. Enfin, les auteurs illustrent l'utilité de l'approche en l'appliquant sur un jeu de données hydrologiques et des prévisions sont effectuées en utilisant un algorithme à sauts réversibles.

mots clés : Copules, distribution bivariée de valeurs extrêmes, estimateur bayésien, fonction de dépendance, MCMC, Metropolis-Hastings, prévisions, sauts réversibles.

1.1. INTRODUCTION

Multivariate extreme value theory is concerned with the distributions related to the componentwise maximums or minimums of a given random sample. The subject has received considerable attention in the past literature with applications ranging from finance to telecommunications and environmental studies. For example, recent research has been stimulated by the possibility of large losses on financial markets, resulting in measures of financial vulnerability such as “value at risk” : see Embrechts, Lindskog, and McNeil (2001) and Heffernan (2004). Extreme value analysis is also becoming a standard technique in risk management for insurance companies as large values of claims and other expenses are the main contributors to companies’ risk factor. Frees and Valdez (1998) present various analyses of this type. In environmental studies, earthquakes, volcanic eruptions, tsunamis, landslides, avalanches, windstorms and extreme sea levels are only but a few examples where extreme value theory is used in practice. In this paper we will present an example in hydrology, where weekly peak flows at hydro-stations were collected over long periods of time and maximums were recorded. Extreme value distributions are appropriate modeling tools for these maximums since they correspond to their distributions in the limiting case.

In general, a bivariate distribution $F_{X,Y}$ is completely characterized by its marginal distributions F_X, F_Y and its dependence function C , called the copula. More precisely, we have the representation $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$, for all x, y , where C is uniquely determined on $\text{Ran}(F_X) \times \text{Ran}(F_Y)$, see Nelsen (1999). In particular, the copula is unique if the random variables X and Y have continuous distributions.

Bivariate extreme value distributions arise as the limiting distributions of re-scaled componentwise maxima. It can be shown, Deheuvels (1984), that these are characterized by

- their marginal distributions, which are themselves univariate extreme value distributions

$$F_{\mu,\sigma,\xi}(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad 1 + \xi(x - \mu)/\sigma > 0, \quad (1.1.1)$$

where $-\infty < \mu, \xi < +\infty$, and $\sigma > 0$. These are respectively known as the Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) and Weibull ($\xi < 0$) distributions.

– a unique extreme value copula

$$C(u, v) = \exp \left\{ \log(uv) A \left(\frac{\log(u)}{\log(uv)} \right) \right\}, \quad \text{for all } u, v \in [0, 1], \quad (1.1.2)$$

where A satisfies the dependence function conditions, that is : a real-valued convex function with $\max(1 - t, t) \leq A(t) \leq 1$, for all $0 \leq t \leq 1$.

The particular scaling sequences used for weak convergence of the componentwise maxima affects only the location and scale parameters of the margins.

The function A is called the Pickands dependence function. The upper bound $A \equiv 1$ corresponds to the independence copula : $C(u, v) = uv$, for all $u, v \in [0, 1]$, while the lower bound $A(t) = \max\{1 - t, t\}$, for all $0 \leq t \leq 1$ corresponds to the comonotone copula : $C(u, v) = \min\{u, v\}$, for all $u, v \in [0, 1]$,

The existing literature presents various nonparametric estimators for the dependence function Pickands (1981), Tawn (1988), Deheuvels (1991), T. de Oliveira (1992), Capéraà, Fougères, and Genest (1997), Hall and Tajvidi (2000), Jiménez, Villa-Diharce, and Flores (2001), Heffernan (2004). The reader is referred to Abdous and Ghoudi (2005) for a complete review.

For finite samples, the proposed nonparametric estimators in the literature tend to either not meet themselves the constraints to be dependence functions, not look smooth, or not have good convergence rates, see Hall and Tajvidi (2000) or Beirlant, Goegebeur, Segers, and Teugels (2004). If these estimators are used without any further modifications, one can obtain poor results. In Abdous and Ghoudi (2005), the authors present several artifacts applied on the initial estimator to circumvent some of these problems, such as using convex hulls and applying constrained smoothing on the estimated curve.

Moreover, many authors have constructed estimators for the dependence function while considering the marginal distributions as known. Since we are mainly interested in the dependence structure, it is common practice to first estimate the marginal distributions, and then plug-in the estimates as the true margins thereafter. In Capéraà et al. (1997), the authors discuss the need to develop estimators that handle the realistic case of unknown marginals in a better way.

On the one hand, the classical estimators of the dependence function are developed for their asymptotic behavior. On the other hand, it is necessary to modify them whenever the sample size is finite, this is cumbersome. The aim of this paper is to propose an algorithm which satisfies the regularity conditions associated with the dependence function, while treating the estimation of the marginal distributions and the dependence function simultaneously, instead of relying on a plug-in method. Most importantly, the new approach should give a low mean integrated squared error (MISE). The main idea is to exploit the geometry of the problem. In an algebraic perspective, Klüppelberg and May (2006), have characterized all of the polynomials associated with dependence functions. Here, we are going to explore the space of dependence functions via convex interpolations within a Bayesian framework. Finally we will give an application of the method constructed in this article and we will make predictions.

1.2. THE MODEL

In the first part of this section, we propose a nonparametric model for the dependence structure in expression (1.1.2), assuming that the marginal distributions are known, and propose a Bayesian estimator for the dependence function A . In the second part, the known marginals assumption is relaxed, and a simultaneous estimation procedure for the Pickands dependence function and the marginal distributions is proposed.

In the following, we assume that the variable X is distributed according to a univariate extreme value distribution $F_{\mu_1, \sigma_1, \xi_1}$, Y is distributed according to $F_{\mu_2, \sigma_2, \xi_2}$, and the vector (X, Y) has a copula given by expression (1.1.2). Let $U = F_{\mu_1, \sigma_1, \xi_1}(X)$ and $V = F_{\mu_2, \sigma_2, \xi_2}(Y)$. It follows that U and V are uniformly distributed while the joint cumulative distribution function is also given by expression (1.1.2).

1.2.1. Known margins case

1.2.1.1. The approximation space

Let \mathcal{D} represent the space of dependence functions. Here \mathcal{D} is infinite dimensional, so in the same spirit as smoothing splines methodology used in nonparametric regression, we define a $2K$ -dimensional approximation subspace Λ_K , parameterized by vectors of interpolation points θ , which is then used to approximate the true dependence function. More precisely, let

$$\Theta_K = \{(\theta_1, \dots, \theta_K) \in \mathbb{R}^{2K} : \theta_i = (t_i, A(t_i)), i = 1, \dots, K, 0 = t_1 < t_2 < \dots < t_K = 1, \\ \text{for a strictly convex dependence function } A\}.$$

In the following, the abscissas $t_i, i = 1, \dots, K$, are called knots. For $\theta = (\theta_1, \dots, \theta_K) \in \Theta_K$, let A be a strictly convex dependence function such that $\theta_i = (t_i, A(t_i)), i = 1, \dots, K$, and let

$$\mathcal{D}_\theta = \{\varphi \in \mathcal{D} : \varphi(t_i) = A(t_i), i = 1, \dots, K\}.$$

We have the following approximation result, Thakur (1978).

$$\sup_{A \in \mathcal{D}} \inf_{\theta \in \Theta_K} \sup_{\varphi \in \mathcal{D}_\theta} \|A - \varphi\|_\infty \leq \frac{1}{2(K-1)}. \quad (1.2.1)$$

Consequently, the choice of $\theta \in \Theta_K$ is critical here, while the particular choice of a representative $\varphi \in \mathcal{D}_\theta$ is free. On the other hand, if $\varphi \in \mathcal{D}_\theta$ is sufficiently smooth, then (U, V) has a density. Therefore, for every \mathcal{D}_θ , we shall select a smooth representative. Such a construction is provided in the Appendix. Now Λ_K is now well defined.

1.2.1.2. The Bayesian estimator

It is very convenient to treat K as a parameter. The space of parameters is then given by

$$\Lambda = \bigcup_{K=L}^U \{K\} \times \Lambda_K,$$

where L and U are lower and upper bounds for K . Here, K is modeled as a discrete uniform distribution on $\{L, \dots, U\}$. Now, given K , the prior model on Λ_K is given

by specifying a uniform distribution over Θ_K . In this Bayesian framework, our prior distribution serves mainly as a tool rather than subjective beliefs regarding the dependence curves distribution.

In order to numerically evaluate the Bayesian estimate \widehat{A}_B for A , a Markov chain is constructed in the Appendix. The Bayesian estimate

$$\widehat{A}_B(w) = \int_{\Lambda} \varphi \pi(d\varphi | w), \quad (1.2.2)$$

(where $w = (u_1, v_1), \dots, (u_n, v_n)$ is the sample); is the average of the dependence curves generated by the Markov chain. A key tool for exploring the parameter space is a reversible jump algorithm as in Green (1995). The idea is essentially to use a proposal kernel allowing moves within a particular subspace of vectors of interpolation points having identical dimensions and jumps between subspaces of vectors of interpolation points having different dimensions.

In our model, the upper bound U could theoretically be made infinite, giving the nonparametric nature of our approach. However, we assume a finite support for K , so the lower and upper bounds for the number of knots should be investigated. First, we know by expression (1.2.1), that any curve in \mathcal{D} can be approximated arbitrarily well in the sup-norm provided K is large enough. On the other hand, many knots can evidently reduce flexibility in the chain, slowing down the convergence. Also, we should consider the amount of data available : if the sample size is small and there are many knots, then the prior may have too much influence over the posterior.

1.2.2. Unknown margins case

More realistically, if the parameters $\Xi_1 = (\mu_1, \phi_1, \xi_1)$ and $\Xi_2 = (\mu_2, \phi_2, \xi_2)$ are unknown, where $\phi_i = \log(\sigma_i)$, $i = 1, 2$, then the Bayesian estimator in expression (1.2.2) is extended as follows. As suggested in Coles (2001), we consider the individual prior distributions $\pi_{\Xi_i} = \pi_{\mu_i} \pi_{\phi_i} \pi_{\xi_i}$, $i = 1, 2$, where π_{μ_i} , π_{ϕ_i} and π_{ξ_i} , are Gaussian with zero mean and variances 10^4 , 10^4 and 100 respectively. The Bayesian estimator we suggest $(\widehat{\Xi}_1, \widehat{\Xi}_2, \widehat{A})$, is determined by the product prior $\pi = \pi_{\Xi_1} \pi_{\Xi_2} \pi_{\varphi}$, and the reversible jump algorithm described in the Appendix is

easily adapted for its computation. In particular, the parameter space $\Xi_1 \times \Xi_2$ of the marginal distributions is explored using a Metropolis-Hastings random walk.

1.3. SIMULATION EXPERIMENTS

In order to compare the performance of our estimator with classical competitors, an extensive simulation was carried out using artificial data. The margins used for the simulation were both standard Gumbel and as true dependence functions, we used the same parametric model as Capéraà et al. (1997), that is the asymmetric logistic family

$$A_{r,\alpha,\beta}(t) = 1 - \beta + (\beta - \alpha)t + \{\alpha^r t^r + \beta^r (1 - t)^r\}^{1/r}, \quad 0 \leq \alpha, \beta \leq 1, r \geq 1.$$

This family is known to be flexible enough to cover a wide range of dependence functions for bivariate extremes. We considered fifteen different levels of the parameters r, α and β , the corresponding dependence curves are plotted in Figure 1.1. Here, r is a dependence parameter with complete independence when $r = 1$, while α and β are the asymmetrical parameters with complete symmetry when $\alpha = \beta = 1$. For each level of the above parameters, 1000 data sets representing samples of sizes 10, 25 and 100 were generated. The aim of the simulation study was to evaluate the relative precision of the estimator in terms of mean integrated squared error (MISE), first in the case where we assume the marginal distributions are known, and then in the more realistic case where the margins are unknown. The classical competing estimators considered were the ones proposed by Capéraà, Fougères & Genest, by Hall & Tajvidi and by Deheuvels, since they are known to perform well in many situations, see Capéraà et al. (1997) and Hall and Tajvidi (2000). For these estimators, convex hulls were computed in order to meet the regularity conditions associated with dependence functions, and in the unknown margins case, maximum likelihood estimates were plugged-in. Concerning our estimator, the total number of iterations used in the MCMC procedure was 150 000, the initial state of the Markov chain comprises a set of 10 individual knots chosen on the symmetrical dependence curve $A_{2,1,1}$, and in the case of unknown margins, the maximum likelihood estimates were taken as the initial parameters for the margins. Tables 1.1 and 1.2 below show relative MISE

of the competitors as compared with our estimator in the known and unknown margins case respectively. The relative MISE is given by

$$\frac{\mathbb{E} \left\{ \int_0^1 (\widehat{A}_B(t) - A(t))^2 dt \right\}}{\mathbb{E} \left\{ \int_0^1 (\widehat{A}_c(t) - A(t))^2 dt \right\}}$$

where \widehat{A}_B is the Bayesian estimator and \widehat{A}_c is the competitor estimator.

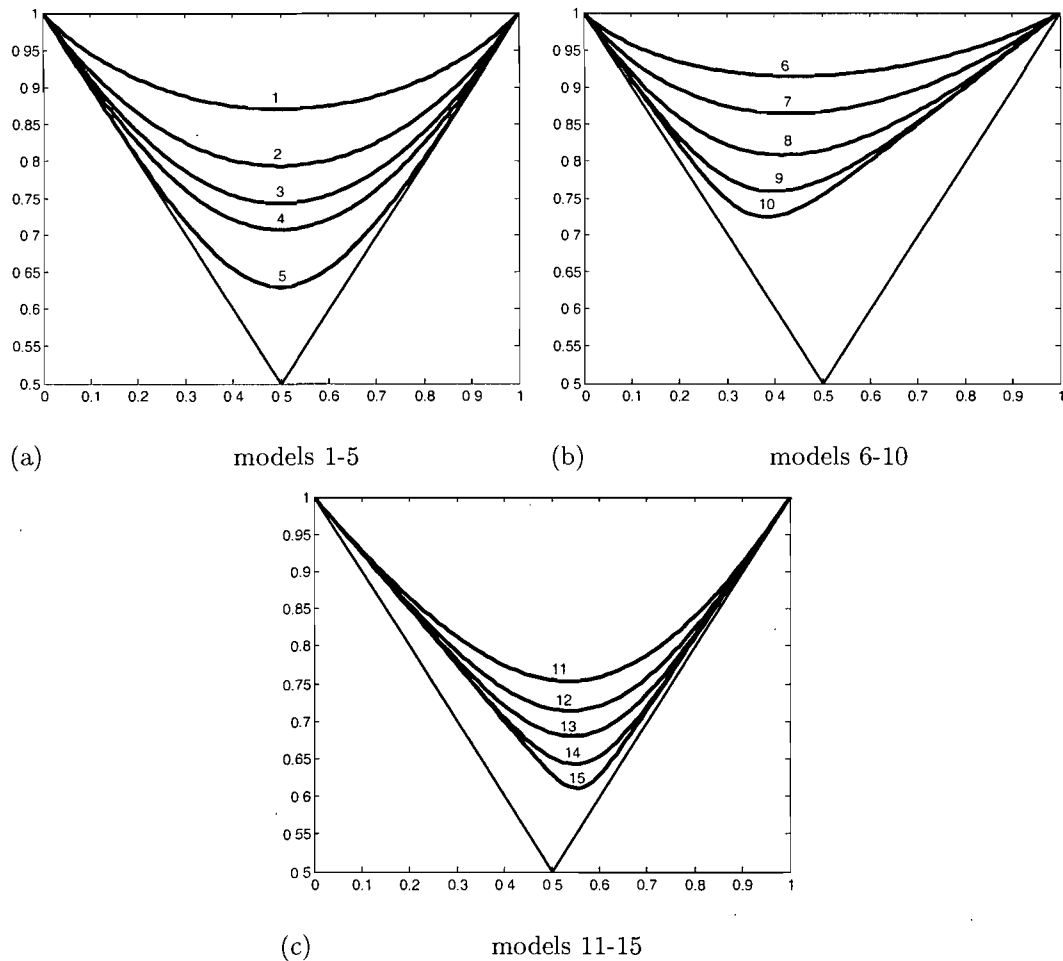


FIG. 1.1. The dependence curves used as the true models in the simulation.

In the Bayesian setup, the reduction in mean integrated squared error is important for small sample sizes. The reduction continues for moderate and large sample sizes in most cases, the size of the reduction depends on the model. From Tables 1.1 and 1.2, the Bayesian estimator is always better in models 1,6,7,8,9,10, it is better most of the time in models 2,3,11,12,13,14,15, while the competitors

TAB. 1.1. Relative MISE of the CFG estimator (C), Hall & Tajvidi's estimator (H) and Deheuvel's estimator (D), as compared with the Bayes estimator, in the known margins case. The simulation is based on 1000 experiments.

model	r, α, β	$n = 10$			$n = 25$			$n = 100$		
		C	H	D	C	H	D	C	H	D
1	1.25,1,1	0.27	0.16	0.20	0.25	0.18	0.19	0.39	0.31	0.29
2	1.5,1,1	0.24	0.18	0.18	0.56	0.45	0.36	1.45	1.10	0.92
3	1.75,1,1	0.74	0.63	0.48	1.36	1.10	0.74	2.23	1.72	1.14
4	2,1,1	1.50	1.39	0.81	2.21	1.84	1.06	3.24	2.31	1.21
5	3,1,1	6.04	6.46	1.60	6.02	5.06	1.40	5.19	4.39	1.15
6	1.25,0.9,0.5	0.56	0.32	0.41	0.55	0.33	0.38	0.39	0.30	0.30
7	1.5,0.9,0.5	0.27	0.17	0.22	0.27	0.21	0.22	0.42	0.34	0.32
8	2,0.9,0.5	0.20	0.16	0.17	0.38	0.30	0.28	0.68	0.53	0.47
9	3,0.9,0.5	0.35	0.29	0.26	0.51	0.42	0.37	0.66	0.50	0.43
10	5,0.9,0.5	0.44	0.34	0.33	0.51	0.41	0.36	0.56	0.42	0.38
11	2,0.75,0.95	0.39	0.38	0.31	0.68	0.59	0.46	1.13	0.90	0.69
12	2.5,0.75,0.95	0.74	0.72	0.52	0.96	0.88	0.63	1.18	1.08	1.08
13	3.25,0.75,0.95	1.02	0.94	0.61	0.96	0.86	0.56	1.04	0.84	0.56
14	5,0.75,0.95	0.98	0.97	0.62	0.76	0.61	0.43	0.86	0.63	0.44
15	10,0.75,0.95	0.65	0.63	0.40	0.46	0.31	0.23	0.87	0.62	0.45

are definitely better in models 4 and 5. Figure 1.2 shows estimates for 200 experiments with samples of sizes 25 generated from the asymmetric model $A_{1.5,0.9,0.5}$, under the known margins assumption, and Figure 1.3 shows the corresponding estimates in the unknown margins case. A remarkable feature illustrated in this figure is the curvature of the Bayesian estimator.

1.4. APPLICATION TO A HYDROLOGICAL DATA SET

Companies desiring to build new hydroelectrical structures are concerned with statistics regarding extremal events such as floods and storms since these greatly influence the design of new power sites. In particular, as described in Favre, El Adlouni, Perreault, Thiémondge, and Bobée (2004), much interest is vowed to

TAB. 1.2. Results in the unknown margins case. Simulation is based on the same data sets.

model	r, α, β	$n = 10$			$n = 25$			$n = 100$		
		C	H	D	C	H	D	C	H	D
1	1.25,1,1	0.31	0.24	0.24	0.30	0.25	0.25	0.40	0.35	0.35
2	1.5,1,1	0.12	0.17	0.17	0.37	0.39	0.39	1.26	1.18	1.18
3	1.75,1,1	0.25	0.48	0.48	0.84	0.98	0.98	2.01	1.86	1.86
4	2,1,1	0.45	1.12	1.11	1.52	1.62	1.62	3.01	2.44	2.44
5	3,1,1	0.89	3.92	3.72	4.37	4.82	4.80	6.07	4.35	4.35
6	1.25,0.9,0.5	0.89	0.51	0.51	0.86	0.57	0.57	0.57	0.45	0.45
7	1.5,0.9,0.5	0.32	0.26	0.26	0.32	0.28	0.28	0.48	0.43	0.43
8	2,0.9,0.5	0.15	0.18	0.18	0.31	0.33	0.33	0.72	0.62	0.62
9	3,0.9,0.5	0.21	0.32	0.32	0.47	0.50	0.50	0.74	0.70	0.70
10	5,0.9,0.5	0.30	0.46	0.47	0.56	0.61	0.61	0.65	0.65	0.65
11	2,0.75,0.95	0.20	0.30	0.30	0.53	0.57	0.57	1.14	1.07	1.07
12	2.5,0.75,0.95	0.34	0.67	0.67	0.82	0.93	0.93	1.19	0.94	0.62
13	3.25,0.75,0.95	0.45	1.05	1.04	0.99	1.11	1.11	1.02	1.03	1.03
14	5,0.75,0.95	0.52	1.29	1.28	0.99	1.09	1.09	0.76	0.79	0.79
15	10,0.75,0.95	0.51	1.31	1.30	0.66	0.73	0.73	1.32	1.44	1.44

annual peak flows of neighboring hydro-stations. Hydro-Québec is a public company that produces and distributes electricity throughout the Canadian province of Québec. The company is constantly planning new hydroelectrical works and their architecture depends on maximal flows. They have provided us with data from six of their hydro-stations and were interested in the dependency structure between these stations for classification and future predictions. More precisely, the data represents 42 annual maximums (from the years 1961 to 2002) of weekly peak flows (m^3/s), from stations located on the map in Figure 1.4. Of course weekly measurements are correlated and so exact mathematical assumptions underlying the limit theorems from classical extreme value theory are unverified. As it is mentioned in Coles (2001), the types of data to which extreme value models are applied rarely satisfy temporal independence. However, independence

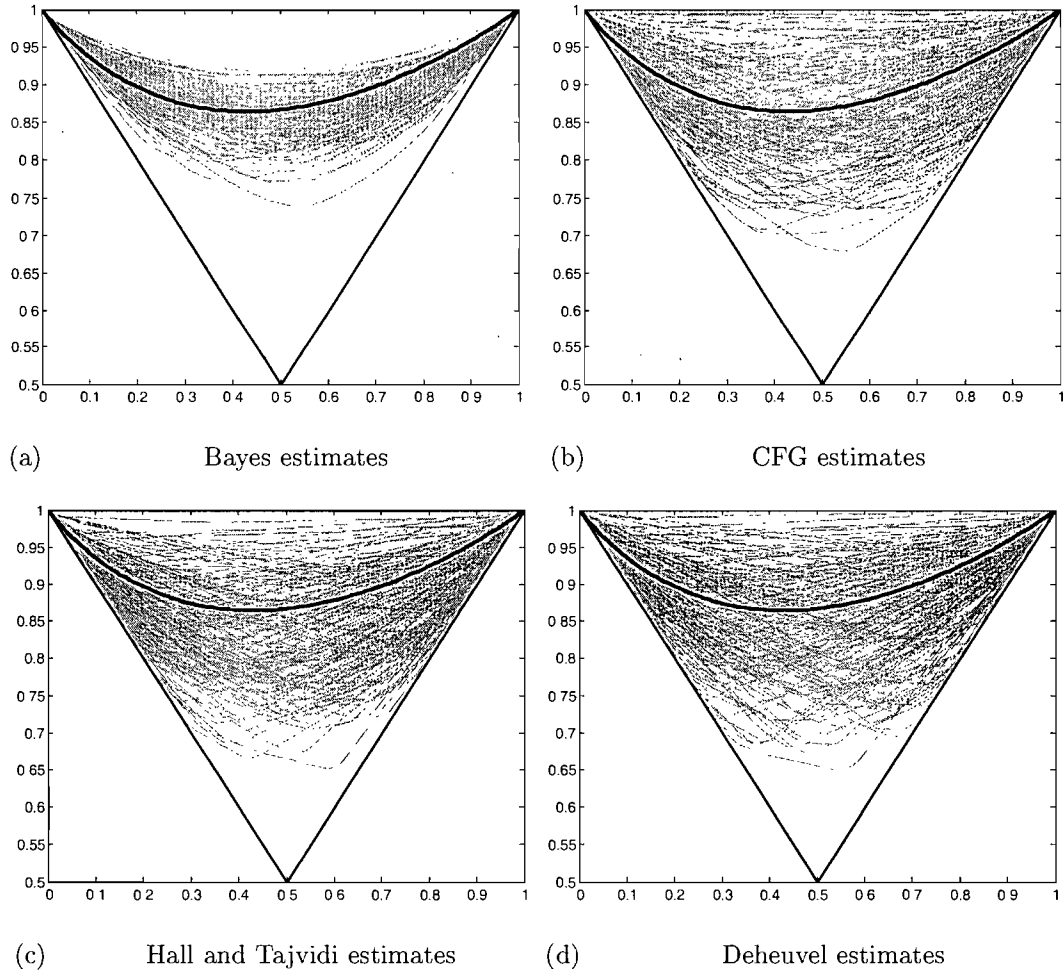


FIG. 1.2. Estimates for 200 experiments treating samples of sizes 25, each generated from the model $r = 1.5, \alpha = 0.9$ and $\beta = 0.5$, under known margins assumption. Light grey lines represent the estimates while thick solid lines represent the truth.

of the maximums remains a reasonable assumption implying similar asymptotic conclusions, see Falk, Hüsler, and Reiss (2004).

It is usually the case for environmental studies in general, that the number of years for which the maximums are observed is rather small. Also, in view of the increased presence of global climate changes, Hydro-Québec has decided to focus on recent data only. It has been observed, in recent studies, that the map seems to be divided into three main regions, north west, north east and south, each of which shares similar climatic and physiographic conditions. Thus, it should be expected that we find high dependency between stations in the same climatic

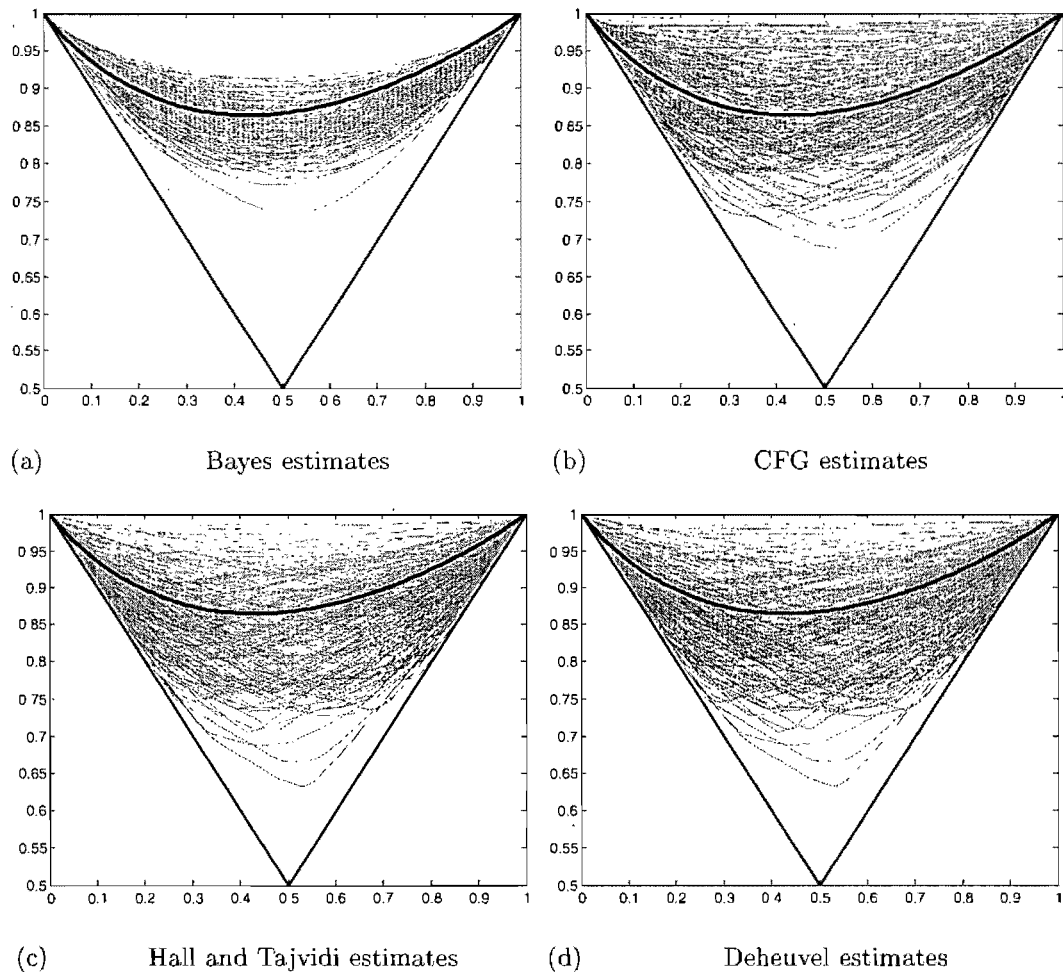


FIG. 1.3. Estimates for 200 experiments treating samples of sizes 25, each generated from the model $r = 1.5$, $\alpha = 0.9$ and $\beta = 0.5$, under unknown margins assumption. Light grey lines represent the estimates while thick solid lines represent the truth.

regions while low dependency between stations in different regions. We computed the Bayesian estimates for the dependence functions and compared them with the classical estimators considered in the simulation described in the previous section. We noted that the Deheuvels estimates were very similar to those of Hall & Tajvidi, and so we have decided not to include them here. The results were as anticipated : weak dependency was observed between stations from different climate zones and stronger dependency between stations located in same climate zones. Figure 1.5 shows the results we obtain in two cases, that is for stations LaGrande 4 and Manouanes, which are respectively in the north west and south

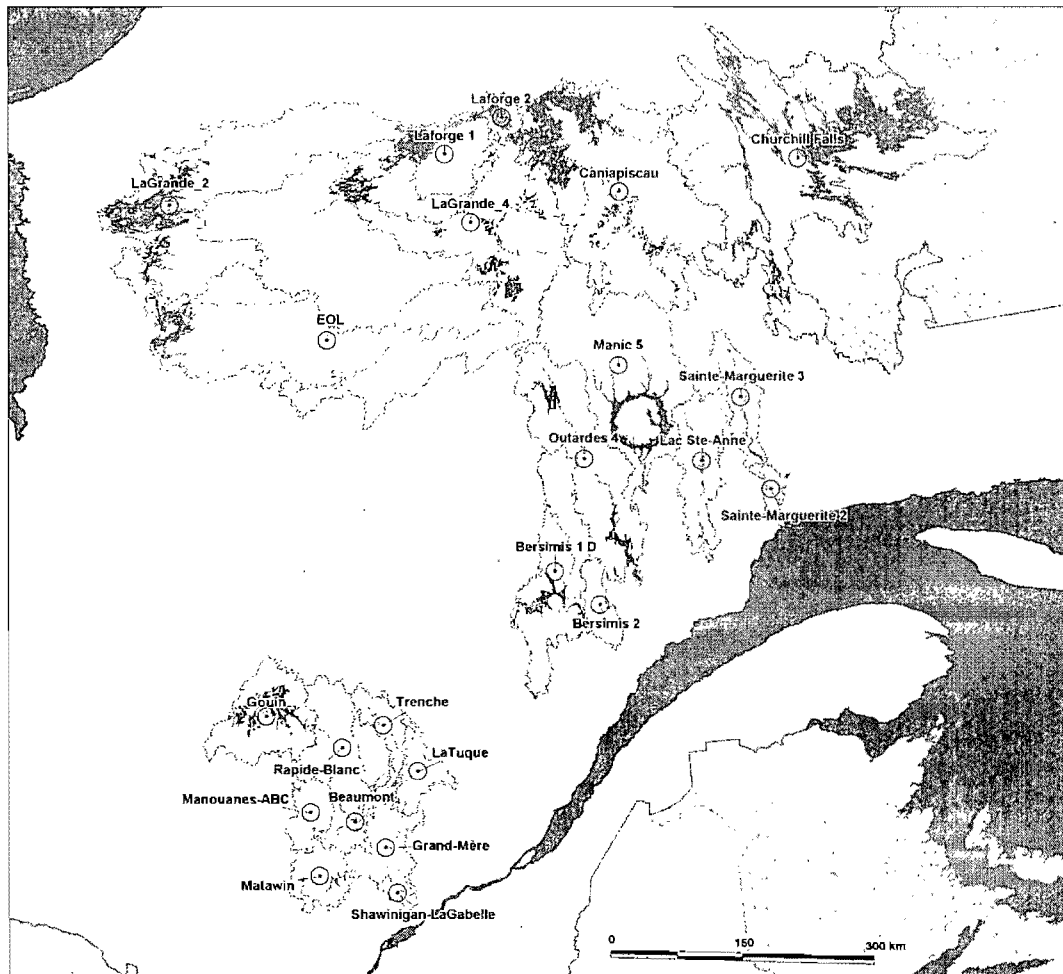
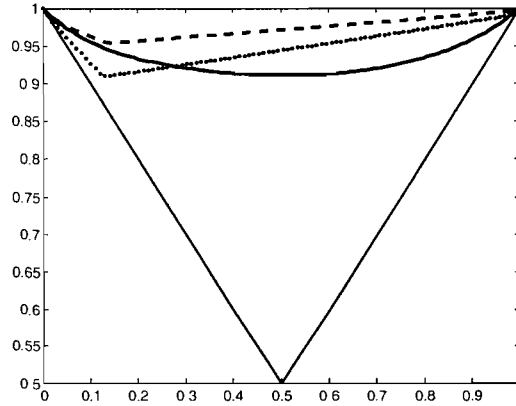


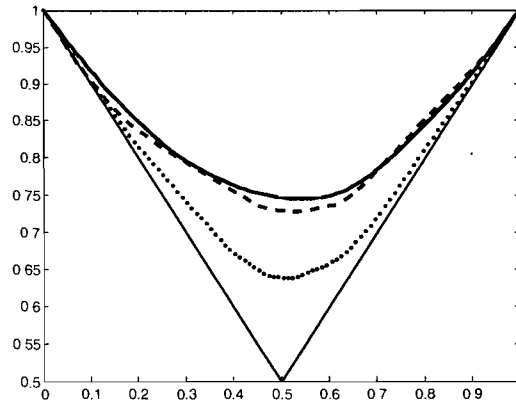
FIG. 1.4. Geographical locations of Quebec's main watersheds

regions, and for stations LaGrande 4 and EOL, which are both in the north west region. In the first case, the CFG and Hall estimates resemble each other in shape and magnitude, in particular, both of them are asymmetrical and exhibit a discontinuity in the first derivative. On the other hand, the Bayesian estimator seems symmetrical and shows slightly stronger dependence than the competitors. Interestingly, in the second case, the CFG and Bayesian estimator resemble each other while the Hall estimator seems to indicate much stronger dependence.

Finally, the Bayesian approach is particularly well suited for making predictions. This could be of interest for planning the allocation of resources in the future. For instance, Hydro-Québec wants to know what is going to happen at other stations if the measurements reach a certain level at a given station. In



(a) La Grande 4 vs Manouanes



(b) LaGrande 4 vs EOL

FIG. 1.5. Dependence function estimations for stations LaGrande 4 and Manouanes, and for stations LaGrande 4 and EOL. Solid line is Bayesian estimate, dashed line is CFG estimate and dotted line is Hall estimate.

other words, what are the conditional distributions? In a Bayesian terminology, the answer is given by the predictive distributions. Let $z = (x_1, y_1), \dots, (x_n, y_n)$, be the past yearly maximum flows from two stations A and B . We want to predict next year's maximum flow Y at station B , given the maximum flow X at station A has reached the level x . In order to do this, our algorithm for the unknown margins case is simply used to numerically evaluate the mean and 95% quantiles of the predictive distribution

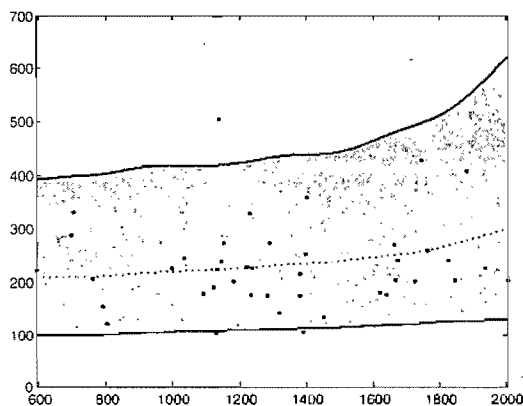
$$m(y | z, x) \propto \int_{\Xi_1 \times \Xi_2 \times \Lambda} f(z, x, y | \Xi_1, \Xi_2, \varphi) \pi(d\Xi_1, d\Xi_2, d\varphi),$$

where π is the prior and $f(\cdot | \Xi_1, \Xi_2, \varphi)$ is the joint density distribution of z, x, y , when the parameters are known. Figure 1.6, shows the results obtained for stations LaGrande 4 and Manouanes, and for stations LaGrande 4 and EOL, station A being the LaGrande 4 station in the general context just described. As anticipated, in both cases, the strength of dependency between stations is reflected in the shape of the predictive estimates. In the LaGrande 4 versus Manouanes case, the lower tail of the predictive distributions are not affected by the measurements at the LaGrande station while the upper tails are more sensible especially for large values of the measurements at the LaGrande station. In the case of the LaGrande versus the EOL station, there is clearly a dependency in the predictive estimates. This dependency is almost linear. The dispersion increases for the predictive distributions at the EOL station when the measurements at the Lagrande stations are large.

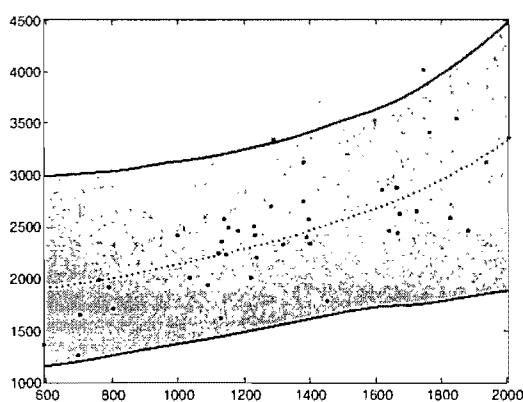
DISCUSSION AND PRACTICAL CONSIDERATIONS

A Bayesian approach for this problem turns out to be very appealing for many reasons. First, it enables simultaneous estimation of the dependence function and the parameters defining the margins in a natural way. Secondly, it produces nice smooth estimates and in most cases, reduces the MISE as compared with other asymptotic approaches, even if the sample size is 100. The approach exploits the geometry of the problem and allows us to make use of all the information given by the sample. For example, in the known margins case, the classical estimator CFG is based on transforming the bivariate couple (U, V) into a univariate statistic Z , which they call pseudo-observation. The authors exploit the fact that the distribution of Z involves the dependence function A through a nice differential equation. On the other hand, Z is not a sufficient statistic, so the authors have chosen to discard useful information in order to obtain tractable asymptotic expressions for their estimator.

One special feature of our algorithm is the way we interpolate points in order to obtain a convex and continuously differentiable curve. Notice that working with polynomial splines involves a quite complicated algorithm. It is known Passow and



(a) La Grande 4 vs Manouanes



(b) LaGrande 4 vs EOL

FIG. 1.6. 95% predictive bands for stations LaGrande 4 and Manouanes, and for stations LaGrande 4 and EOL. Dotted line is the predicted estimates.

Roulier (1977) that convex spline interpolation, with degree independent of the interpolated points, may be impossible. In fact, Costantini (1986) quote (p.210) “...it is shown in Passow and Roulier (1977) that, if convex interpolation is desired, then the data can force the degree to be very large, and this fact is inherent in the nature of convex spline interpolation.” Our approach circumvents this problem.

Finally, the method developed here opens the door to further work. Unfortunately, the method presents a certain weakness. In fact, the estimator shows some weaknesses when the true dependence curve A is near the boundary. This is due to our choice of prior model which does not favor such dependence curves. However, in practice the statistician is usually aware beforehand when such situations occur

and a different prior model should then be considered. In a theoretical perspective, it could be interesting to compare the asymptotic behavior of the estimator with the classical ones. Also, the general multivariate case remains to be further explored. Recently, Zang, Wells, and Liang (2007) have generalized the classical estimators discussed in this paper to the multivariate case. In our framework, a model using constrained hyper-surfaces might be envisaged but may be difficult to construct : the interpolation problem was already challenging in the bivariate case!

ACKNOWLEDGMENTS

We are particularly grateful to professors Christian Genest and Paul Gustafson, the Editor, for comments that have led to a much improved version of this paper. We also wish to thank the Associate Editor and the two referees for their useful suggestions, James Merleau, Luc Perreault and Frederic Guay from Hydro-Québec for providing us with the data and the map, and for showing great interest in our method and finally Louis-Alexandre Leclaire from DRAP polytechnique for many high performance computing advices. This research was supported in part by NSERC(Canada) and Institut des Sciences Mathématiques(Québec) grants.

APPENDIX

1.4.1

In this subsection, we give an explicit choice for a particular representative $\varphi \in \mathcal{D}_\theta$ such that the copula (1.1.2), possesses a density with respect to the two-dimensional Lebesgue measure. Sufficient smoothness conditions imposed on φ for the existence of such a density are the continuity of the first derivative along with the existence (except at finitely many points) and boundedness of the second derivative.

Consider $\theta \in \Theta_K$, $\theta_i = (t_i, A(t_i))$, $i = 1, \dots, K$, $0 = t_1 < t_2 < \dots < t_K = 1$, for a strictly convex dependence function A . We can concentrate our search for

φ having the form :

$$\varphi(t) = \sum_{i=1}^{K-1} \mathbf{1}_{[t_i, t_{i+1}]}(t) \left[c_{i_1} \psi_1 \left(k_{i_1} \left(\frac{t - t_i}{t_{i+1} - t_i} \right)^2 \right) + c_{i_2} \psi_2 \left(k_{i_2} \left(\frac{t - t_{i+1}}{t_{i+1} - t_i} \right)^2 \right) \right]$$

with $c_{i_1}, c_{i_2}, k_{i_1}, k_{i_2} \geq 0$ and ψ_1, ψ_2 nondecreasing, convex, twice continuously differentiable, with $\psi_1(0) = 0 = \psi_2(0)$. We must now find values of $c_{i_1}, c_{i_2}, k_{i_1}$ and k_{i_2} such that φ satisfies the above conditions. On every interval $[t_i, t_{i+1}]$, we must solve :

$$\begin{aligned} y_i &= c_{i_2} \psi_2(k_{i_2}) \\ y_{i+1} &= c_{i_1} \psi_1(k_{i_1}) \\ \alpha_i &= \frac{2c_{i_2} k_{i_2}}{t_{i+1} - t_i} \psi_2'(k_{i_2}) \\ \beta_i &= \frac{2c_{i_1} k_{i_1}}{t_{i+1} - t_i} \psi_1'(k_{i_1}) \end{aligned}$$

where $\alpha_i, \beta_i \geq 0, i = 1, \dots, K - 1$, are chosen so that φ is continuously differentiable over $[0, 1]$. Since the convexity and differentiability conditions are invariant with respect to the addition of straight lines, we can assume without losing any generality that $y_{i+1} = y_i$, for every $i = 1, \dots, K - 1$. Thus, using φ of this form, satisfying the conditions in order to obtain a density is equivalent to solving

$$c_{i_2} \psi_2(k_{i_2}) = c_{i_1} \psi_1(k_{i_1}) \quad (1.4.1)$$

with $c_{i_1} = \beta_i \frac{t_{i+1} - t_i}{2k_{i_1} \psi_1'(k_{i_1})}$ and $c_{i_2} = \alpha_i \frac{t_{i+1} - t_i}{2k_{i_2} \psi_2'(k_{i_2})}$. This equation has one degree of freedom. An interesting question at this stage is : what are the functions ψ_1 and ψ_2 , nondecreasing, convex, twice continuously differentiable, with $\psi_1(0) = 0 = \psi_2(0)$ that make equation (1.4.1) easily solvable for k_{i_1} and k_{i_2} ? Good candidates are $\psi_1(\cdot) = \exp(\cdot) - 1 = \psi_2(\cdot)$, which implies solving

$$\beta_i \frac{1 - \exp(-k_{i_1})}{k_{i_1}} = \alpha_i \frac{1 - \exp(-k_{i_2})}{k_{i_2}}.$$

In order for k_{i_1} and k_{i_2} to be nonnegative, we solve

$$\frac{1 - \exp(-k_{i_1})}{k_{i_1}} = \frac{\alpha_i \rho_i}{\alpha_i + \beta_i} \quad \text{and} \quad \frac{1 - \exp(-k_{i_2})}{k_{i_2}} = \frac{\beta_i \rho_i}{\alpha_i + \beta_i}, \quad (1.4.2)$$

where $\rho_i \in]0, 1[$ is arbitrary and can be chosen to control for numerical precision when solving (1.4.2) with either $\frac{\alpha_i}{\alpha_i + \beta_i}$ or $\frac{\beta_i}{\alpha_i + \beta_i}$ very small.

1.4.2

The algorithm described here generates a trans-dimensional Markov chain for computing the Bayesian estimator.

Let $\theta \in \Theta_K$, $\theta_i = (t_i, A(t_i))$, $i = 1, \dots, K$, $0 = t_1 < t_2 < \dots < t_K = 1$, for a strictly convex dependence function A . When $i = 2, \dots, K - 1$, the individual interpolation points θ_i are called interior interpolation points.

Let L and U be respectively the lower and upper bounds for the number of knots. Recall the parameter space is

$$\Lambda = \bigcup_{K=L}^U \{K\} \times \Lambda_K,$$

and each $\varphi \in \Lambda$ is constructed from $\theta \in \Theta_K$, for some K . Thus, the prior distribution is specified on

$$\Theta = \bigcup_{K=L}^U \{K\} \times \Theta_K.$$

Let $\pi_\theta(\cdot) = \pi_\theta(\cdot|K)\pi_K(\cdot)$ be the prior distribution for θ . Here, $\pi_K \propto \mathbf{1}_{\{L, \dots, U\}}$ and $\pi_\theta(\cdot|K) \propto \mathbf{1}_{\Theta_K}$. For every interior interpolation point θ_i , let $\theta_{(-i)}$ be the set of $K - 1$ interpolation points obtained after deleting θ_i . The interior interpolation point θ_i is represented by **C** in Figure 1.7. According to the prior model, the full conditional $\pi_{\theta_i}(\cdot|\theta_{(-i)})$ is uniformly distributed over the triangle **BFD** in Figure 1.7, that is

$$\pi_{\theta_i}(\cdot|\theta_{(-i)}) \sim \mathcal{U}(\Delta\mathbf{BFD}). \quad (1.4.3)$$

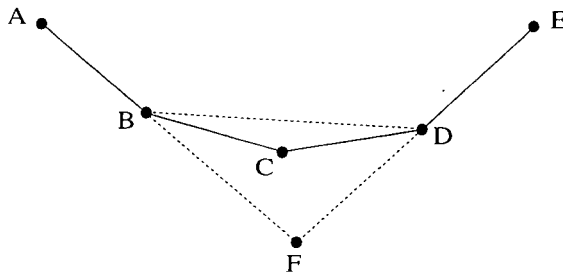


FIG. 1.7. Selection of the proposal knot

We consider three types of moves : 1. the displacement of an interior interpolation point, 2. the insertion of a new interior interpolation point and 3. the deletion

of an interior interpolation point; with probabilities of making such moves depending on the current number of knots. More precisely, given a current set of K knots, select move m with probability p_K^m , where $p_K^1 = 1/3$, p_K^2 and p_K^3 depend linearly on K with $p_L^3 = 0$, $p_U^2 = 0$ and $p_K^2 + p_K^3 = 2/3$.

Let $q_1(\cdot|\theta)$ be the density for proposing a new set of interpolation points to enter the Markov chain in the case of a displacement move, corresponding to the following : first, we select θ_i amongst $\{\theta_2, \dots, \theta_{K-1}\}$ using a discrete uniform distribution. Then we replace θ_i according to $\pi_{\theta_i}(\cdot|\theta_{(-i)})$ as in (1.4.3).

Let $q_2(\cdot|\theta)$ be the density for proposing a new set of interpolation points to enter the Markov chain in the case of an insertion move, corresponding to the following : first, we determine the location of the new knot by selecting $i \in \{1, \dots, K-1\}$, using a discrete uniform distribution (the new knot will be in the interval $[t_i, t_{i+1}]$). Then, we create the new interior interpolation point according to (1.4.3). This gives $K+1$ relabeled interpolation points.

Finally, let $w = (u_1, v_1), \dots, (u_n, v_n)$ be a sample and $L(\theta|w)$ be the likelihood. At each iteration, a transition is made according to the following :

1. if move 1 is selected :

1.1 generate θ^{proposal} from $q_1(\cdot|\theta^{\text{current}})$

1.2 take $\theta^{\text{new}} = \begin{cases} \theta^{\text{proposal}} & \text{with probability } \alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) \\ \theta^{\text{current}} & \text{with probability } 1 - \alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) \end{cases}$
 where $\alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) = \min \left\{ 1, \frac{L(\theta^{\text{proposal}}|w)}{L(\theta^{\text{current}}|w)} \right\}$

2. if move 2 is selected :

2.1 generate θ^{proposal} from $q_2(\cdot|\theta^{\text{current}})$

2.2 take $\theta^{\text{new}} = \begin{cases} \theta^{\text{proposal}} & \text{with probability } \alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) \\ \theta^{\text{current}} & \text{with probability } 1 - \alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) \end{cases}$
 where $\alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) = \min \left\{ 1, \frac{L(\theta^{\text{proposal}}|w)p_{K+1}^3}{L(\theta^{\text{current}}|w)p_K^2} \right\}$

3. if move 3 is selected :

3.1 choose a single interior interpolation point to delete,
 using a discrete
 uniform distribution

$$3.2 \text{ take } \theta^{\text{new}} = \begin{cases} \theta^{\text{proposal}} & \text{with probability } \alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) \\ \theta^{\text{current}} & \text{with probability } 1 - \alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) \end{cases}$$

where $\alpha(\theta^{\text{current}}, \theta^{\text{proposal}}) = \min \left\{ 1, \frac{L(\theta^{\text{proposal}}|w)p_{K-1}^2}{L(\theta^{\text{current}}|w)p_K^3} \right\}$

Bibliographie

- Abdous, B., Ghoudi, K., 2005. Nonparametric estimators of multivariate extreme dependence functions. *Nonparametric Statistics* 17, 915–935.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes*. Wiley.
- Capéraà, P., Fougères, A.-L., Genest, C., 1997. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84, 567–577.
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Costantini, P., 1986. On monotone and convex spline interpolation. *Mathematics of Computation* 46, 203–214.
- Deheuvels, P., 1984. *Probabilistic aspects of multivariate extremes*. Reidel.
- Deheuvels, P., 1991. On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters* 12, 429–439.
- Embrechts, P., Lindskog, F., McNeil, A., 2001. *Modelling dependence with copulas and applications to risk management*.
- Falk, M., Hüsler, J., Reiss, R.-D., 2004. *Laws of Small Numbers : Extremes and Rare Events*. Birkhäuser.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. *Water Resources Research* 40.
- Frees, W., Valdez, E., 1998. Understanding relationships using copulas. *North American Actuarial Journal* 2, 1–25.

- Green, P. J., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* 82, 711–732.
- Hall, P., Tajvidi, N., 2000. Distribution and dependence function estimation for bivariate extreme-value distributions. *Bernoulli* 6, 835–844.
- Heffernan, J. E., 2004. Nonparametric estimation of the bivariate extreme value distribution using a limiting conditional representation. *ExtremesSubmitted*.
- Jiménez, J., Villa-Diharce, E., Flores, M., 2001. Nonparametric estimation of the dependence function in bivariate extreme-value distributions. *Journal of Multivariate Analysis* 76, 159–191.
- Klüppelberg, C., May, A., 2006. Bivariate extreme value distributions based on polynomial dependence functions. *Mathematical Methods in the Applied Sciences* 29, 1467–1480.
- Nelsen, R. B., 1999. *An Introduction to Copulas*. Springer, New York.
- Passow, E., Roulier, J., 1977. Monotone and convex spline interpolation. *SIAM Journal of Numerical Analysis* 14, 904–909.
- Pickands, J., 1981. Multivariate extreme value distributions. In : *Bulletin of the International Statistical Institute, Proceedings of the 43rd Session, Buenos Aires*. pp. 859–878.
- T. de Oliveira, J., 1992. Intrinsic estimation of the dependance structure for bivariate extremes : a new and simpler approach. *Communications in Statistics-Theory and Methods* 21, 599–611.
- Tawn, J., 1988. Bivariate extreme value theory : models and estimation. *Biometrika* 75, 397–415.
- Thakur, L. S., 1978. Error analysis for convex separable programs : the piecewise linear approximation and the bounds on the optimal objective value. *SIAM J. Appl. Math.* 34 (4), 704–714.
- Zang, D., Wells, M., T., Liang, P., 2007. Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *Journal of Multivariate Analysis*In press.

Chapitre 2

BAYESIAN ESTIMATION OF A BIVARIATE COPULA FUNCTION

ABSTRACT

A bivariate distribution with continuous margins can be decomposed via a copula and its marginal distributions. On the space of copula functions, we construct a finite dimensional approximation subspace which is parameterized by a doubly stochastic matrix. The estimation is done in the space of doubly stochastic matrices also known as the Birkhoff polytope. A Bayesian approach is undertaken which allows for simultaneous estimation of the copula and the marginal distributions. Estimators based on various priors are analyzed via Markov chain Monte Carlo methodology. A rather extensive simulation experiment is carried out. In many cases, the results indicate that the estimators obtained from the Bayesian approach outperforms the standard kernel approach in terms of mean integrated squared error.

Some key words : Bayes, Birkhoff polytope, copula, doubly stochastic matrices, Gibbs sampler, Markov chain Monte Carlo, Metropolis-Hastings, nonparametric, partition of unity.

AMS 2000 subject classification : 65C05, 62G05, 62H20, 62C10.

RÉSUMÉ

Toute distribution bivariée dont les marges sont continues peut être décomposée via sa copule et ses distributions marginales. Nous construisons un sous-espace d'approximation de dimension finie, à l'intérieur de l'espace des copules. Les éléments de ce sous-espace sont paramétrisés par une matrice doublement stochastique. L'estimation est donc effectuée à l'intérieur de l'espace des matrices doublement stochastiques également connu sous le nom de polytope de Birkhoff. À l'aide du modèle, nous pouvons construire des estimateurs qui demeurent à l'intérieur de la classe des copules. Une approche bayésienne est considérée. Cette dernière a l'avantage de permettre l'estimation simultanée des lois marginales et de la copule lorsque l'intérêt porte sur la distribution bivariée. Les estimateurs de Bayes associés à plusieurs lois a priori sont étudiés à l'aide de méthodes Monte-Carlo par chaînes de Markov. Une simulation est effectuée en considérant

des modèles paramétriques connus pour la copule et les marges. Dans plusieurs cas, les résultats indiquent que certains estimateurs bayésiens considérés sont plus performants que les estimateurs classiques en terme de l'intégrale des erreurs quadratiques moyennes.

mots clés : Copules, estimateur bayésien, Gibbs, matrices doublement stochastiques, Metropolis-Hastings, partition de l'unité, polytope de Birkhoff.

2.1. INTRODUCTION

Copulas have received considerable attention over the last years because of their increasing use in multiple fields such as environmental studies, genetics, data networks and simulation. They are also currently one of the hot topics in quantitative finance and insurance, see for instance Cherubini, Luciano, and Vecchiato (2004).

Let $S = [0, 1] \times [0, 1]$. A copula is a real valued function $C: S \rightarrow [0, 1]$, that is fixed on the border of S and is monotone, that is :

boundary conditions

$$C(0, v) = 0 = C(u, 0), C(u, 1) = u, C(1, v) = v, \text{ for all } (u, v) \in S,$$

monotonicity

$$\begin{aligned} \Delta C(u_1, u_2, v_1, v_2) &= C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \\ &\geq 0, \text{ for all } u_1 \leq u_2 \text{ and } v_1 \leq v_2. \end{aligned}$$

In particular, copulas are Lipschitz continuous (with constant 1) and form an equicontinuous family. They are bounded by the so-called Fréchet-Hoeffding copulas, indeed,

$$\max(0, u + v - 1) \leq C(u, v) \leq \min(u, v), \text{ for all } (u, v) \in S.$$

Sklar's Theorem states that a bivariate distribution F is completely characterized by its marginal distributions F_X, F_Y and its copula C . More precisely, we have the representation $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$, for all x, y , where C is uniquely determined on $\text{Ran}(F_X) \times \text{Ran}(F_Y)$, see Nelsen (1999). In particular, the copula is unique if F_X and F_Y are continuous. We then have the following expression for the copula

$$C(u, v) = F(F_X^{-1}(u), F_Y^{-1}(v)), \text{ for all } (u, v) \in S. \quad (2.1.1)$$

Assume that $\{(x_i, y_i): i = 1, \dots, n\}$ is a sample and that each couple $(X_i, Y_i), i = 1, \dots, n$ has joint cumulative distribution function F . Consider the problem of

estimating the copula C by a copula \hat{C} where \hat{C} depends on the sample. Let us say that \hat{C} satisfies the Deheuvels constraint provided

$$\hat{C}\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{1}{n} \sum_{k=1}^n I(\text{rank}(x_k) \leq i, \text{rank}(y_k) \leq j), \text{ for } i, j = 1, \dots, n.$$

In Deheuvels (1979), the asymptotic behaviour of all of the copulas \hat{C} satisfying the Deheuvels constraint is described. Notice that this class of copulas is invariant under strictly increasing transformations. In other words, if f and g are two strictly increasing functions then Deheuvels' class of copulas based on the original sample and the one based on the sample $\{(f(x_i), g(y_i)) : i = 1, \dots, n\}$ are identical. Notice also that

$$\hat{C}\left(\frac{i}{n}, \frac{j}{n}\right) = \hat{F}\left(\hat{F}_X^{-1}\left(\frac{i}{n}\right), \hat{F}_Y^{-1}\left(\frac{j}{n}\right)\right), \text{ for } i, j = 1, \dots, n, \text{ (Deheuvels' approach)}$$

where \hat{F} is the empirical cumulative distribution function and \hat{F}_X^{-1} and \hat{F}_Y^{-1} are the generalized inverses. In general, if \hat{F} is a smooth estimator of F (\hat{F}_X and \hat{F}_Y are continuous say) then a straightforward extension becomes

$$\hat{C}(u, v) = \hat{F}\left(\hat{F}_X^{-1}(u), \hat{F}_Y^{-1}(v)\right), \text{ for all } (u, v) \in S. \text{ (general approach)}$$

On the frequentist side, nonparametric estimation of copulas many authors have considered this general approach. Asymptotic properties of gaussian kernels are given in Fermanian and Scaillet (2003). Local linear kernels are studied in Chen and Huang (2007). The reader is referred to Charpentier, Fermanian, and Scaillet (2006) for a recent review. On the Bayesian side, Sungur and Ng (2005) argue that researchers could benefit from the incorporation of prior knowledge not only on the marginal distributions but on the dependence structure as well.

Let \mathcal{C} be the space of all copulas. We want to develop a Bayesian alternative for the nonparametric estimation of $C \in \mathcal{C}$. In order to this we need to select a prior on \mathcal{C} but the dimension of \mathcal{C} is infinite. In our approach, for $m > 1$, the prior is concentrated on a $(m - 1)^2$ dimensional subspace $\mathcal{A}_m \subset \mathcal{C}$. The elements of \mathcal{A}_m are mixtures of m^2 components. As m gets larger, \mathcal{A}_m converges to \mathcal{C} . This approximation space has been studied in Li, Mikusiński, and Taylor (1998) and Kulpa (1999).

Essentially, \hat{C} is obtained in the following way. We work on C , a random copula in \mathcal{A}_m . The distribution of C depends on a prior. On the one hand, when the marginal distributions F_X and F_Y are known, the distribution of the sample depends only on C . The prior on C and the sample lead to the posterior. The estimate \hat{C} comes from the posterior mean which is evaluated using MCMC. In section 2.2, we make a choice for \mathcal{A}_m and we develop some priors. On the other hand, when the marginal distributions F_X and F_Y are unknown, we need to put priors on F_X and F_Y as well. Therefore, the posterior on C depends on the priors on C , F_X , F_Y and the sample. Here again, the dimension of the space of the marginal distributions is infinite. In section 2.3, we select a prior for the marginal distributions. A simulation study is conducted in section 2.4 in which we treat the cases of known and unknown marginal distributions separately. We compare various Bayesian estimators from our model with the general and Deheuvels' approach.

2.2. THE APPROXIMATION SPACE FOR THE COPULA FUNCTION

In section 2.2.1, for all $m > 1$, we construct a finite dimensional approximation subspace $\mathcal{A}_m \subset \mathcal{C}$. The construction of \mathcal{A}_m uses a basis which forms a partition of unity. The elements of \mathcal{A}_m are parameterized by a doubly stochastic matrix P . The choice of the basis is fixed while P varies. The representation is given in expression (2.2.3). Furthermore, we give upper bounds on

$$\sup_{C \in \mathcal{C}} \inf_{A \in \mathcal{A}_m} \|A - C\|_\infty.$$

Finally, in section 2.2.2 we describe ways to select priors over the set of doubly stochastic matrices.

2.2.1. Construction

A partition of unity is a set of functions $\varphi = \{\varphi_i\}_{i=1}^m$, defined over the unit interval $[0, 1]$, such that

1. $\varphi_i \geq 0$, $i = 1, \dots, m$,
2. $\int_0^1 \varphi_i = 1/m$, $i = 1, \dots, m$,

3. $\sum_{i=1}^m \varphi_i(u) = 1$, for all $u \in [0, 1]$.

We will subsequently consider characteristic functions

$$\varphi_1 = \mathbb{1}_{[0,1/m]}, \varphi_i = \mathbb{1}_{((i-1)/m, i/m]}, \quad i = 2, \dots, m \quad (2.2.1)$$

and Bernstein polynomials

$$\varphi_i = B_{i-1}^{m-1}, \quad i = 1, \dots, m, \quad (2.2.2)$$

where

$$B_i^m(u) = \binom{m}{i} u^i (1-u)^{m-i}, \quad \text{for all } u \in [0, 1].$$

See Li, Mikusiński, and Taylor (1998) for more examples. In the following, let $\Phi = (\Phi_1, \dots, \Phi_m)'$, where $\Phi_i(u) = \int_0^u \varphi_i(t) dt$, for all $u \in [0, 1]$, $i = 1, \dots, m$ and let

$$A_P(u, v) = m\Phi(u)'P\Phi(v), \quad \text{for all } (u, v) \in S, \quad (2.2.3)$$

where P is an $m \times m$ doubly stochastic matrix.

Lemma 2.2.1. *For every doubly stochastic matrix P , A_P is an absolutely continuous copula.*

PROOF. The boundary conditions are straightforward to verify. The monotonicity condition follows from

$$\begin{aligned} \Delta A_P(u_1, u_2, v_1, v_2) &= m\{\Phi(u_2) - \Phi(u_1)\}'P\{\Phi(v_2) - \Phi(v_1)\} \\ &\geq 0, \quad \text{for all } u_1 \leq u_2 \text{ and } v_1 \leq v_2. \end{aligned}$$

□

The approximation space is defined as

$$\mathcal{A}_m = \{A_P : P \text{ is a doubly stochastic matrix}\}.$$

The approximation power of \mathcal{A}_m is now discussed. Let $\mathcal{G}_m = \{(i/m, j/m) : i, j = 1, \dots, m\}$, be a uniformly spaced grid on the unit square S . For a given copula

C , let $R_C = (C(i/m, j/m))_{i,j=1}^m$ be the restriction of C on \mathcal{G}_m . Let

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix},$$

then $P_C = mDR_C D'$ is a doubly stochastic matrix. We have the following upper bounds for $\|A_{P_C} - C\|_\infty$, see Appendix 2.5.1 for a proof.

Theorem 2.2.1. *Let C be a copula and let $A = A_{P_C} \in \mathcal{A}_m$.*

a. *Using the characteristic basis we obtain that $R_A = R_C$ and $\|A - C\|_\infty \leq 2/m$.*

b. *Using the Bernstein polynomials we obtain that $\|A - C\|_\infty \leq 1/\sqrt{m}$.*

Notice that the bound given in part b. of Theorem 2.2.1 is suboptimal. We are giving it because it has a simple expression. In the Appendix, we derive a stronger result.

Corollary 2.2.1. *Assume that $\{(x_i, y_i) : i = 1, \dots, n\}$ is a sample. Let R be the $n \times n$ matrix given by*

$$r_{ij} = \frac{1}{n} \sum_{k=1}^n I(\text{rank}(x_k) \leq i, \text{rank}(y_k) \leq j), \text{ for } i, j = 1, \dots, n.$$

If we use the characteristic basis with $m = n$ then the copula $\hat{C}_{DEH} = n^2 \Phi' DRD' \Phi$ satisfies Deheuvels' constraint.

2.2.2. The prior

Let \mathcal{B} represent the set of doubly stochastic matrices of order $m > 1$. It forms a convex polytope also known as the Birkhoff polytope. Let

$$\mathcal{B}_0 = \{P - 1/m \mathbf{1} \mathbf{1}' : P \in \mathcal{B}\}$$

and $\mathcal{V} = \text{span}(\mathcal{B}_0)$. Consider the inner product $\langle V_1, V_2 \rangle = \text{tr}(V_1 V_2')$ on \mathcal{V} . Thus, \mathcal{V} is an $(m - 1)^2$ dimensional Hilbert space and an orthonormal basis is given by

$\{y_i y_j'\}_{i,j=1,\dots,m-1}$, with

$$y_i = \frac{1}{\sqrt{i(i+1)}} \underbrace{(1, \dots, 1)}_i, -i, 0, \dots, 0)', \quad i = 1, \dots, m-1.$$

Assume that μ is the Lebesgue measure on \mathcal{V} . It is well known that $0 < \mu(\mathcal{B}_0) < \infty$, for all $m > 1$. In fact, the exact value of $\mu(\mathcal{B}_0)$ is known for all $m \leq 10$, see Beck and Pixton (2003).

We now list the priors that we will subsequently consider. For the first and second one, we take the Lebesgue measure on \mathcal{B}_0 as a reference measure

$$\pi_1 = \frac{1}{\mu(\mathcal{B}_0)},$$

and,

$$\pi_2(\cdot) \propto \exp(\alpha \|\cdot\|), \quad \alpha > 0.$$

For the third and fourth priors, we take the $(m-1)^2 - 1$ Hausdorff measure μ_0 on $\partial\mathcal{B}_0$ as a reference measure and consider

$$\pi_3 = \frac{1}{\mu_0(\partial\mathcal{B}_0)},$$

and,

$$\pi_4(\cdot) \propto \exp(\alpha \|\cdot\|), \quad \alpha > 0.$$

We finally consider

$$\pi_5 \sim \mathcal{U}(S_m), \quad \text{where } S_m \text{ is the set of permutation matrices.}$$

By making use of the Birkhoff-von Neumann Theorem, doubly stochastic matrices can be decomposed via convex combinations of permutation matrices and this could also be used for priors over \mathcal{B} , see Sungur and Ng (2005). In fact, \mathcal{B} is the convex hull of the permutation matrices and these are precisely the extreme points (or vertices) of \mathcal{B} . Furthermore, every $m \times m$ doubly stochastic matrix P is a convex combination of at most $k = (m-1)^2 + 1$ permutation matrices, see Mirsky (1963). Thus, for $P \in \mathcal{B}$, $P = \sum_{i=1}^k \lambda_i \sigma_i$ for some weight vector $\lambda = (\lambda_1, \dots, \lambda_k)$ lying in the $(k-1)$ -simplex

$$\Lambda_k = \{(\lambda_1, \dots, \lambda_k) : 0 \leq \lambda_i, \sum_{i=1}^k \lambda_i = 1\},$$

and permutation matrices $\sigma_i \in S_m$, $i = 1, \dots, k$. A prior π_6 is described in the following. First, we assume $\lambda \in \Lambda_k$ is generated according to a Dirichlet distribution $\pi_\lambda \sim \text{Dir}(k, 1, \dots, 1)$. Then, after (lexicographical) ordering of S_m , we choose $\sigma = (\sigma_{i_1}, \dots, \sigma_{i_k})$ independently from λ using a uniform distribution over the set

$$\mathcal{O}_k = \{(\sigma_{i_1}, \dots, \sigma_{i_k}) \in S_m^k : 0 \leq i_1 < \dots < i_k \leq m! - 1\},$$

thus

$$\pi_6(\lambda, \sigma) = \frac{(k-1)!}{\binom{m!}{k}}, (\lambda, \sigma) \in \Lambda_k \times \mathcal{O}_k.$$

2.3. THE APPROXIMATION SPACE FOR THE MARGINAL DISTRIBUTIONS

Let \mathcal{F} represent the space of continuous univariate distribution functions. In section 2.3.1, for all $d \geq 1$, we construct a finite dimensional approximation subspace $\mathcal{F}_d \subset \mathcal{F}$ and give a bound on the approximation error. Subspace \mathcal{F}_d will serve as a model for the estimation of the individual marginal distributions and the selection of a prior is discussed in section 2.3.2.

2.3.1. Construction

Let $\theta \in \Theta$ be a finite dimensional vector, F_θ be a strictly increasing, absolutely continuous univariate distribution function. Let

$$\Gamma_d = \{0 \leq \gamma_0 < \gamma_1 < \dots < \gamma_{d-1} < \gamma_d \leq 1\}$$

be a set of knots and let S_{Γ_d} be a continuously differentiable monotone cubic spline. Assume that $S_{\Gamma_d}(\gamma_i) = i/d$, for $i = 0, \dots, d$. The construction of such splines may be found in Fritsch and Butland (1984). An element $F_{\Gamma_d, \theta}$ from \mathcal{F}_d is given by

$$F_{\Gamma_d, \theta} = S_{\Gamma_d} \circ F_\theta.$$

For any $F \in \mathcal{F}$, F_θ should be interpreted as a first approximation to F and S_{Γ_d} as a function used to correct the first approximation. The following result is proven in Appendix 2.5.1 :

Theorem 2.3.1. For $d \geq 1$, $\sup_{F \in \mathcal{F}} \sup_{\theta \in \Theta} \inf_{\Gamma_d} \|F - F_{\Gamma_d, \theta}\|_\infty \leq 1/d$.

2.3.2. The prior

In the above model, it is very convenient to treat d as a parameter. The knots spaces are then given by

$$\Gamma^{(j)} = \bigcup_{d=L}^U \{d\} \times \Gamma_d^{(j)}, \quad j = 1, 2,$$

where L and U are the lower and upper bounds for d and j is the margin index. The parameter space for both marginal distributions is

$$\mathcal{M} = \Gamma^{(1)} \times \Theta^{(1)} \times \Gamma^{(2)} \times \Theta^{(2)}.$$

We assume the same models for the marginal distributions and use independent and identical priors on $\Gamma^{(1)} \times \Theta^{(1)}$ and $\Gamma^{(2)} \times \Theta^{(2)}$. We take $F_{\mu, \sigma}$ as the cumulative distribution function of a normal $N(\mu, \sigma^2)$ and consider Jeffrey's priors $\pi_\mu = 1$ and $\pi_\sigma = 1/\sigma$. Finally, we take $\gamma_1, \dots, \gamma_{d-1}$ as the order statistics from $d-1$ independent uniform $U(0, 1)$ variates.

The general parameter space for our model in the unknown margins case is

$$\Pi = \mathcal{B} \times \mathcal{M}.$$

The prior model is extended over Π for simultaneous estimation of the copula and the marginal distributions by assuming independent priors on \mathcal{B} and \mathcal{M} .

2.4. SIMULATION EXPERIMENTS

In the experiment, we consider the independence copula and the following parametric families :

1. *Clayton family* : $C_\theta(u, v) = \{\max(0, u^{-\theta} + v^{-\theta} - 1)\}^{-1/\theta}$, $\theta \geq -1, \theta \neq 0$,
2. *Frank family* : $C_\theta(u, v) = -\frac{1}{\theta} \log \left\{ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right\}$, $\theta \neq 0$,
3. *Gumbel family* : $C_\theta(u, v) = \exp \left[- \{(-\log u)^\theta + (-\log v)^\theta\}^{1/\theta} \right]$, $\theta \geq 1$,

4. *Gaussian family* : $C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)), |\rho| \leq 1,$

where Φ_ρ is the standard bivariate Gaussian cdf with correlation coefficient ρ , and Φ is the univariate standard normal distribution.

These families are among popular ones for applications and represent different dependence structures. For instance, Gumbel's copula has a very heavy right tail, in fact it is an extreme value copula. On the other hand, Clayton's copula has a heavy left tail, while Frank's copula, has light and symmetrical tails. Finally, the Gaussian copula has asymmetric tails which are heavier than Frank's, although not as heavy as that of Gumbel's or Clayton's. For the Clayton, Frank and Gaussian families, values of the parameter away from 0 indicate departure from independence, while a parameter away from 1 indicates departure from independence for the Gumbel copula.

An extensive simulation experiment is carried out in two parts. In the first part, we consider the case of known marginal distributions and use bivariate data sampled from the copula models above. For each model, 1000 samples of sizes $n = 30$ and $n = 100$ are generated. In the second part of the experiment, the data sets are generated from the same copula families with various marginal distributions. For each model, 1000 samples of sizes $n = 100$ are generated. In the known margins case, the Bayesian estimators associated with $\pi_1, \pi_3, \pi_5, \pi_6$ and the maximum likelihood estimator from our model are considered. In the unknown margins case, we take priors π_2 and π_4 on \mathcal{B} and compute the corresponding Bayesian estimators described in section 2.3.2. In every case, the order of the doubly stochastic matrix in our model is given by $m = 6$. We use the Bernstein basis when π_5 is the prior and the characteristic basis in every other case. For π_5 , the evaluation of the estimator is exact. For the other cases, the evaluations are approximated using Markov chain Monte Carlo techniques.

We compare the performance of our estimators and Deheuvels' estimator given in corollary 2.2.1 with the general approach using gaussian kernels. The results we obtaine in the first and second part of the simulation are given in tables 2.1 and 2.2 respectively. The tables report the values of the relative mean integrated

squared error (RMISE)

$$\text{RMISE}(\widehat{C}) = \frac{\mathbb{E} \left\{ \int_0^1 \int_0^1 (\widehat{C}(u, v) - C(u, v))^2 dudv \right\}}{\mathbb{E} \left\{ \int_0^1 \int_0^1 (\widehat{C}_G(u, v) - C(u, v))^2 dudv \right\}},$$

where \widehat{C}_G is the gaussian kernel estimator.

We do not show the results for the Bayesian estimator corresponding to π_6 . It turns out that this prior is well suited for cases where the true underlying model is near independence. Indeed, doubly stochastic matrices near the center of the polytope have more representations in terms of convex combinations of permutation matrices than matrices near the faces of the polytope. Thus, considering this prior distribution puts much mass near the middle of the polytope, this has the effect of pulling the estimates towards the independence copula. For models further away from independence, the estimator behaves poorly.

A striking feature observed in table 2.1 is the improvement of the third Bayesian estimator when the sample size increases from $n = 30$ to $n = 100$. For $n = 100$, the third Bayesian estimator is always better than the general approach and Deheuvels' estimator. We can also observe how well the MLE performs. In table 2.2, the situation is different. None of the Bayesian estimators considered are systematically better than the others. The situation depends on the particular models and different priors should be explored. In contrast with the other estimators, our approach takes full advantage of the known marginal distributions for estimating the copula function.

2.5. APPENDIX

2.5.1. Proofs

Proof of Theorem 2.2.1.

- a. A direct evaluation shows that $R_A = R_C$. If two copulas C_1 and C_2 satisfy the constraint $R_{C_1} = R_{C_2}$ then $\|C_1 - C_2\|_\infty \leq 2/m$. It comes from the Lipschitz condition.

TAB. 2.1. Relative MISE of Bayes estimators associated to π_1, π_3, π_5 , the maximum likelihood estimator from our model and Deheuvels' estimator given in corollary 2.2.1.

Copula model	$n = 30$					$n = 100$				
	Bayes 1	Bayes 3	Bayes 5	MLE	DEH	Bayes 1	Bayes 3	Bayes 5	MLE	DEH
independence	0.29	2.12	0.95	0.85	2.08	0.80	0.56	1.78	0.27	2.04
Clayton(0.5)	0.96	1.45	0.74	0.50	1.91	1.30	0.77	1.10	0.39	1.86
Clayton(1)	2.29	1.10	0.70	0.41	1.81	2.03	0.65	0.77	0.36	1.47
Clayton(1.5)	3.62	0.99	0.77	0.44	1.63	2.89	0.59	0.77	0.49	1.27
Clayton(2)	4.95	0.96	0.92	0.53	1.50	3.45	0.51	0.84	0.42	1.05
Clayton(2.5)	6.35	0.91	1.22	0.60	1.36	3.98	0.44	1.15	0.39	0.88
Clayton(3)	7.19	0.91	1.49	0.61	1.20	4.32	0.42	1.59	0.35	0.77
Clayton(5)	9.87	0.75	2.58	0.53	0.79	4.67	0.35	3.06	0.31	0.38
Frank(1)	0.53	1.68	0.92	0.66	2.04	0.98	0.65	1.61	0.31	2.00
Frank(2)	1.12	1.34	0.92	0.41	1.94	1.48	0.81	1.45	0.40	1.92
Frank(3)	2.15	1.15	0.92	0.41	1.93	2.15	0.76	1.28	0.49	1.77
Frank(4)	3.29	0.98	0.88	0.61	1.81	2.85	0.62	0.82	0.51	1.53
Frank(5)	4.77	0.95	0.88	0.81	1.69	3.48	0.39	0.49	0.32	1.28
Frank(6)	6.06	0.93	0.98	0.77	1.54	4.20	0.34	0.50	0.29	1.13
Frank(7)	7.30	0.98	1.23	0.81	1.39	4.51	0.37	0.74	0.30	0.91
Gumbel(1.25)	0.94	1.38	0.80	0.51	1.92	1.29	0.76	1.19	0.50	1.86
Gumbel(1.5)	2.28	1.01	0.73	0.58	1.85	2.22	0.75	0.85	0.61	1.63
Gumbel(1.75)	3.72	0.80	0.72	0.72	1.75	2.98	0.51	0.62	0.45	1.36
Gumbel(2)	5.15	0.74	0.83	0.66	1.59	3.57	0.34	0.53	0.31	1.13
Gumbel(2.25)	6.56	0.82	1.03	0.72	1.43	3.91	0.33	0.65	0.30	0.90
Gumbel(2.5)	7.53	0.70	1.27	0.61	1.26	4.42	0.43	0.99	0.42	0.77
Gumbel(10)	10.57	0.22	4.16	0.21	0.13	3.92	0.28	4.85	0.27	0.04
Gaussian(0.1)	0.36	1.90	0.88	0.74	2.03	0.85	0.64	1.71	0.33	2.02
Gaussian(0.2)	0.56	1.76	0.83	0.69	2.04	1.03	0.72	1.50	0.33	1.98
Gaussian(0.3)	0.92	1.44	0.80	0.48	1.96	1.30	0.83	1.27	0.40	1.97
Gaussian(0.4)	1.47	1.17	0.74	0.40	1.94	1.63	0.79	1.00	0.41	1.80
Gaussian(0.5)	2.25	1.06	0.66	0.51	1.89	2.26	0.72	0.74	0.53	1.68
Gaussian(0.6)	3.61	1.01	0.66	0.69	1.78	2.90	0.57	0.46	0.45	1.41
Gaussian(0.75)	6.19	0.98	0.80	0.84	1.48	4.24	0.41	0.38	0.36	0.99

- b. First of all, it is well known that $m\Phi'D = (B_1^m, \dots, B_m^m)$. For any $(u, v) \in S$ consider X and Y two independent random variables, X has a binomial(m, u) distribution and Y has a binomial(m, v) distribution. Let $\theta = (u, v)$. We have

$$A(\theta) = E_\theta[C(X/m, Y/m)].$$

Therefore,

$$\begin{aligned}
\sup_{\theta \in S} |A(\theta) - C(\theta)| &= \sup_{\theta \in S} |\mathbb{E}_\theta[C(X/m, Y/m) - C(u, v)]|, \\
&\leq \sup_{\theta \in S} \mathbb{E}_\theta[|C(X/m, Y/m) - C(u, v)|], \\
&\leq \sup_{\theta \in S} \mathbb{E}_\theta[|X/m - u| + |Y/m - v|], \\
&= \frac{2}{m} \sup_{u \in (0,1)} \mathbb{E}_u[|X - mu|].
\end{aligned}$$

In Lemma 2 we give the value of $\sup_{u \in (0,1)} \mathbb{E}_u[|X - mu|]$. However, with the Schwarz inequality we have

$$\begin{aligned}
\sup_{u \in (0,1)} 2\mathbb{E}_u[|X - mu|]/m &\leq \sup_{u \in (0,1)} 2\sqrt{\text{Var}_u[X]}/m, \\
&= \frac{1}{\sqrt{m}}.
\end{aligned}$$

Lemma 2.5.1. *Consider X , a binomial(n, p) random variable. We have*

$$\sup_{0 \leq p \leq 1} \mathbb{E}[|X - np|] = \begin{cases} \frac{1}{B(1/2, (n+1)/2)} & \text{if } n \text{ is odd,} \\ \left(1 - (n+1)^{-2}\right)^{n/2} \left(1 + (n+1)^{-2}\right)^{\frac{1}{B(1/2, n/2)}} & \text{if } n \text{ is even.} \end{cases}$$

Proof of lemma 2. Let $g_n(p) = \mathbb{E}[|X - np|]$. We have

$$g_n(p) = \frac{2n!}{([\!np\!])(n-1-[\!np\!])!} p^{[\!np\!]+1} (1-p)^{n-[\!np\!]} \text{ for all } n \geq 1, p \in [0, 1],$$

where $\lfloor x \rfloor = \sup\{n : n \leq x, n \text{ is an integer}\}$ for all x . Therefore,

$$\begin{aligned}
\sup_{0 \leq p \leq 1} g_n(p) &= \max_{0 \leq k \leq n-1} \sup_{\{p : [\!np\!]=k\}} g_n(p), \\
&= \max_{0 \leq k \leq n-1} g_n\left(\frac{k+1}{n+1}\right).
\end{aligned}$$

First of all, $\sup_{0 \leq p \leq 1} g_1(p) = g_1(1/2) = 1/2 = 1/B(1/2, 1)$. Assume that $n > 1$.

Let $h_n(k) = g_n\left(\frac{k+2}{n+1}\right) / g_n\left(\frac{k+1}{n+1}\right)$, for $k = 0, \dots, n-2$. We have

$$h_n(k) = \frac{\left(1 + (k+1)^{-1}\right)^{k+2}}{\left(1 + (n-k-1)^{-1}\right)^{n-k}} \text{ and } h_n(k) = \frac{1}{h_n(n-2-k)} \text{ for } k = 0, \dots, n-2.$$

However,

$$\frac{d}{dt} \log\left(1 + \frac{1}{t}\right)^{t+1} = \log\left(1 + \frac{1}{t}\right) - \frac{1}{t} < 0 \text{ for all } t > 1.$$

This implies that h_n decreases on $\{0, \dots, n-2\}$. Therefore,

$$g_n\left(\frac{1}{n+1}\right) < \dots < g_n\left(\frac{(n+1)/2}{n+1}\right) > \dots > g_n\left(\frac{n}{n+1}\right) \quad \text{if } n \text{ is odd,}$$

$$g_n\left(\frac{1}{n+1}\right) < \dots < g_n\left(\frac{n/2}{n+1}\right) = g_n\left(\frac{n/2+1}{n+1}\right) > \dots > g_n\left(\frac{n}{n+1}\right) \quad \text{if } n \text{ is even.}$$

The final expression is obtained using the following identity :

$$n! = 2^n \Gamma\left(\frac{n}{2} + 1\right) \Gamma\left(\frac{n+1}{2}\right) / \Gamma\left(\frac{1}{2}\right) \quad \text{for all } n \geq 0.$$

Proof of Theorem 2.3.1. If F is a continuous univariate distribution function, then given any θ , we take $\gamma_i = F_\theta(q_i)$, $i = 0, \dots, d$, where $q_i = F^{-1}(i/d)$, $i = 0, \dots, d$ and F^{-1} is the generalized inverse of F . It follows that $F_{\Gamma_m, \theta}(q_i) = F(q_i)$, $i = 0, \dots, d$.

2.5.2. Algorithms

The following MCMC algorithms are used to generate copulas, the Bayes estimator is computed using the sampling mean of the chain.

1. *The Metropolis within Gibbs Algorithm.* The following algorithm describes the transition kernel for the Markov chain used to compute the Bayesian estimator associated to π_1 .

1. First, construct an orthonormal basis :

1. Consider the unit length m -dimensional vectors

$$y_i = \{(m-i)(m-i+1)\}^{-1/2} (0, \dots, m-i, -1, \dots, -1)', \quad i = 1, \dots, m-1.$$

2. The matrices $M_{ij} = y_i y_j'$, $i, j = 1, \dots, m-1$, forms an orthonormal basis inside the polytope.

2. At each iteration n , let $P^{(n)} = (1/m)\mathbf{1}\mathbf{1}' + \sum_{k,l=1}^{(m-1)^2} \alpha_{kl} y_k y_l'$, be the current doubly stochastic matrix, with combination vector $\alpha = (\alpha_{kl})$. Repeat for $i, j = 1, \dots, m-1$:

1. Select direction M_{ij} and compute segment $\Gamma_{ij} \subset \mathbb{R}$ as follows :

1. Find intervals $\Gamma_{ij}^{(pq)}$, for $\gamma_{ij}^{(pq)}$, such that

$$\gamma_{ij}^{(pq)} y_{ip} y_{lq} \geq -1/m - \sum_{k,l=1}^{(m-1)^2} \alpha_{kl} y_{kp} y_{lq}, \quad p, q = 1, \dots, m.$$

2. Take $\Gamma_{ij} = \bigcap_{p,q} \Gamma_{ij}^{(pq)}$.
2. Generate $\gamma_{ij} \sim \mathcal{U}(\Gamma_{ij})$, propose $\alpha'_{ij} = \alpha_{ij} + \gamma_{ij}$ and set α' as the new combination obtained by replacing α_{ij} by α'_{ij} .
3. Accept $\alpha = \alpha'$ with probability $\min \left\{ 1, \frac{L(\alpha'|x)}{L(\alpha|x)} \right\}$.

2. *The Random Walk Metropolis-Hastings algorithm.* This algorithm describes the transition kernel for the Markov chain used to compute the Bayesian estimator associated to π_6 .

Let i be the current iteration index, and let

$$\lambda^{(i)} = (\lambda_1^{(i)}, \dots, \lambda_k^{(i)}) \in \Lambda_k$$

and

$$\sigma^{(i)} = (\sigma_{i_1}^{(i)}, \dots, \sigma_{i_k}^{(i)}) \in \mathcal{O}_k,$$

be the current values of the weight vector and permutation matrices. Select randomly between two types of moves, with equal probability :

if move 1 is selected

1. Choose the number p of modifiable weight components, uniformly in $\{2, \dots, k\}$.
2. Choose $j_1 < \dots < j_p$ in $\{1, \dots, k\}$, by sampling without replacement.
3. Set $s = \sum_{l=1}^p \lambda_{j_l}^{(i)}$, generate $w \sim \text{Dir}(p, 1, \dots, 1)$ and set

$$\lambda_j^{(\text{prop})} = \begin{cases} sw_j & \text{if } j \in \{j_1, \dots, j_p\}, \\ \lambda_j^{(i)} & \text{otherwise.} \end{cases}$$

4. Accept $\lambda^{(\text{prop})}$ with probability $\min \left\{ 1, \frac{L(\lambda^{(\text{prop})}, \sigma^{(i)}|x)}{L(\lambda^{(i)}, \sigma^{(i)}|x)} \right\}$.

else move 2 is selected

1. Choose the number p of modifiable permutation matrices uniformly in $\{1, \dots, k\}$.
2. Choose $j_1 < \dots < j_p$ in $\{i_1, \dots, i_k\}$, by sampling without replacement.

3. Choose $l_1 < \dots < l_p$ in $\{0, \dots, m! - 1\} \setminus \{i_1, \dots, i_k\}$, by sampling without replacement, and exchange j_n for l_n , $n = 1, \dots, p$; this forms a new, possibly unordered, set of indices I^* .
4. Sort I^* into ascending order as $i_1^* < \dots < i_k^*$, and denote by $r(i_q^*)$, the rank of i_q^* in I^* , for $q = 1, \dots, k$.
5. Set $\lambda^{(\text{prop})} = (\lambda_{r(i_1^*)}^{(i)}, \dots, \lambda_{r(i_k^*)}^{(i)})$ and $\sigma^{(\text{prop})} = (\sigma_{i_1^*}, \dots, \sigma_{i_k^*})$.
6. Accept $(\lambda^{(\text{prop})}, \sigma^{(\text{prop})})$ with probability $\min \left\{ 1, \frac{L(\lambda^{(\text{prop})}, \sigma^{(\text{prop})} | x)}{L(\lambda^{(i)}, \sigma^{(i)} | x)} \right\}$.

3. *The full Markov chain Monte Carlo algorithm.* The following algorithm describes the transition kernel for the Markov chain used to compute the Bayesian estimator for the unknown marginal distributions model.

Let i be the current iteration index, let $P^{(i)}$ be the current doubly stochastic matrix, let

$$\Gamma_{d_j}^{(i)} = \{0 = \gamma_{0_j}^{(i)} < \gamma_{1_j}^{(i)} < \dots < \gamma_{d-1_j}^{(i)} < \gamma_{d_j}^{(i)} = 1\},$$

$\mu_j^{(i)}$ and $\sigma_j^{(i)}$, $j = 1, 2$, be the current knot vectors and parameters defining the marginal distributions. Select randomly between three types of moves, with equal probability :

if move 1 is selected

1. Propose a new doubly stochastic matrix $P^{(\text{prop})}$ according to algorithm 1.

2. Accept $P^{(\text{prop})}$ with probability $\min \left\{ 1, \frac{L(P^{(\text{prop})}, \Gamma_{d_1}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_2^{(i)}, \sigma_2^{(i)} | x)}{L(P^{(i)}, \Gamma_{d_1}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_2^{(i)}, \sigma_2^{(i)} | x)} \right\}$.

else if move 2 is selected

Select between two moves, with equal probability :

1. Propose new parameters :

- 1.1 Generate $\mu_1^{(\text{prop})} \sim N(\mu_1^{(i)}, \cdot)$ and $\sigma_1^{(\text{prop})}$ according to a log-normal distribution with parameters $(\sigma_1^{(i)}, \cdot)$.

- 2.2 Accept $(\mu_1^{(\text{prop})}, \sigma_1^{(\text{prop})})$ with probability

$$\min \left\{ 1, \frac{(1/\sigma_1^{(\text{prop})}) L(P^{(i)}, \Gamma_{d_1}^{(i)}, \mu_1^{(\text{prop})}, \sigma_1^{(\text{prop})}, \Gamma_{d_2}^{(i)}, \mu_2^{(i)}, \sigma_2^{(i)} | x)}{(1/\sigma_1^{(i)}) L(P^{(i)}, \Gamma_{d_1}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_2^{(i)}, \sigma_2^{(i)} | x)} \right\}.$$

2. Propose new knots :

2.1 Knot displacement

(1) Choose k uniformly in $\{1, \dots, d-1\}$ and propose

$$\Gamma_{d_1}^{(\text{prop})} = \{\gamma_{0_1}^{(i)} = 0 < \gamma_{1_1}^{(i)} < \dots < \gamma_{k-1_1}^{(i)} < \gamma_{k_1}^{(\text{prop})} < \gamma_{k+1_1}^{(i)} < \dots < \gamma_{d-1_1}^{(i)} < 1 = \gamma_{d_1}^{(i)}\},$$

$$\text{where } \gamma_{k_1}^{(\text{prop})} \sim \mathcal{U}[\gamma_{k-1_1}^{(i)}, \gamma_{k+1_1}^{(i)}].$$

(2) Accept $\Gamma_{d_1}^{(\text{prop})}$ with probability

$$\min \left\{ 1, \frac{L(P^{(i)}, \Gamma_{d_1}^{(\text{prop})}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)})|x)}{L(P^{(i)}, \Gamma_{d_1}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_2^{(i)}, \sigma_2^{(i)})|x)} \right\}.$$

2.2 Insert new knot

(1) Choose k uniformly in $\{1, \dots, d\}$, relabel $\gamma_{k+j_1}^{(i)*} = \gamma_{k+j-1_1}^{(i)}$, $j = 1, \dots, d$ and propose

$$\Gamma_{d+1_1}^{(\text{prop})} = \{\gamma_{0_1}^{(i)} = 0 < \gamma_{1_1}^{(i)} < \dots < \gamma_{k-1_1}^{(i)} < \gamma_{k_1}^{(\text{prop})} < \gamma_{k+1_1}^{(i)*} < \dots < \gamma_{d_1}^{(i)*} < 1 = \gamma_{d+1_1}^{(i)*}\},$$

$$\text{where } \gamma_{1_k}^{(\text{prop})} \sim \mathcal{U}[\gamma_{k-1_1}^{(i)}, \gamma_{k+1_1}^{(i)*}].$$

(2) Accept $\Gamma_{d+1_1}^{(\text{prop})}$ with probability

$$\min \left\{ 1, \frac{dL(P^{(i)}, \Gamma_{d+1_1}^{(\text{prop})}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)})|x)}{L(P^{(i)}, \Gamma_{d_1}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_2^{(i)}, \sigma_2^{(i)})|x)} \right\}.$$

2.3 Delete knot

(1) Choose k uniformly in $\{1, \dots, d-1\}$, relabel $\gamma_{k+j-1_1}^{(i)*} = \gamma_{k+j_1}^{(i)}$, $j = 1, \dots, d$ and propose

$$\Gamma_{d-1_1}^{(\text{prop})} = \{\gamma_{0_1}^{(i)} = 0 < \gamma_{1_1}^{(i)} < \dots < \gamma_{k-1_1}^{(i)} < \gamma_{k_1}^{(i)*} < \gamma_{k+1_1}^{(i)*} < \dots < \gamma_{d-2_1}^{(i)*} < 1 = \gamma_{d-1_1}^{(i)*}\}.$$

(2) Accept $\Gamma_{d-1_1}^{(\text{prop})}$ with probability

$$\min \left\{ 1, \frac{L(P^{(i)}, \Gamma_{d-1_1}^{(\text{prop})}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)})|x)}{(d-1)L(P^{(i)}, \Gamma_{d_1}^{(i)}, \mu_1^{(i)}, \sigma_1^{(i)}, \Gamma_{d_2}^{(i)}, \mu_2^{(i)}, \sigma_2^{(i)})|x)} \right\}.$$

else move 3 is selected

Propose new second marginal distribution.

TAB. 2.2. Relative MISE of Bayes estimators described in section 2.3.2 using π_2 and π_4 on \mathcal{B} and Deheuvels' estimator given in corollary 2.2.1. Here $n = 100$

Copula model	Marginal distributions											
	N(0,1)			Cauchy(0,1)			Weibull(2,1)			Beta(2,2)		
	Bayes 2	Bayes 4	DEH	Bayes 2	Bayes 4	DEH	Bayes 2	Bayes 4	DEH	Bayes 2	Bayes 4	DEH
independence	0.75	1.54	1.22	1.82	3.83	1.11	1.01	1.82	1.60	1.53	2.72	2.55
Clayton(0.5)	0.78	1.71	1.23	1.59	3.62	1.12	1.04	2.18	1.56	1.31	2.53	2.07
Clayton(1.0)	0.86	1.61	1.24	1.53	2.97	1.13	1.21	1.87	1.48	1.04	1.70	1.40
Clayton(1.5)	1.09	1.63	1.25	1.45	2.58	1.14	1.44	1.66	1.38	0.91	1.27	0.97
Clayton(2.0)	1.31	1.60	1.26	1.37	2.53	1.15	1.74	1.75	1.27	0.82	0.93	0.71
Clayton(2.5)	1.67	1.78	1.26	1.69	2.39	1.15	1.92	1.79	1.18	0.76	0.74	0.53
Clayton(3.0)	2.04	1.93	1.27	1.94	2.58	1.16	2.04	1.59	1.06	0.73	0.56	0.41
Clayton(3.5)	2.48	1.98	1.28	2.10	2.63	1.16	2.28	1.52	0.96	0.66	0.48	0.32
Clayton(4.0)	2.80	1.87	1.24	2.43	3.04	1.18	2.35	1.44	0.86	0.65	0.41	0.26
Clayton(4.5)	3.42	2.08	1.27	3.24	3.25	1.19	2.59	1.48	0.79	0.63	0.37	0.22
Clayton(5.0)	3.88	2.30	1.24	3.59	3.36	1.18	2.55	1.52	0.72	0.61	0.37	0.18
Frank(0.50)	0.75	1.56	1.23	1.83	5.58	1.12	0.98	2.02	1.59	1.55	2.83	2.53
Frank(1.00)	0.75	1.51	1.22	1.70	4.47	1.11	0.99	1.84	1.57	1.47	2.56	2.39
Frank(1.50)	0.78	1.51	1.22	1.94	4.54	1.12	0.98	1.88	1.55	1.39	2.57	2.21
Frank(2.00)	0.78	1.50	1.22	1.66	3.37	1.12	1.01	1.81	1.56	1.28	2.10	1.95
Frank(2.50)	0.83	1.43	1.24	1.67	2.91	1.12	1.04	1.76	1.54	1.18	1.67	1.69
Frank(3.00)	0.87	1.33	1.24	2.02	2.43	1.12	1.11	1.63	1.55	1.13	1.56	1.53
Frank(3.50)	0.97	1.22	1.25	1.82	2.08	1.13	1.20	1.45	1.51	1.08	1.31	1.33
Frank(4.00)	1.09	1.13	1.25	2.05	1.84	1.13	1.33	1.37	1.50	1.03	0.96	1.15
Frank(4.50)	1.28	1.05	1.25	2.21	1.63	1.14	1.49	1.33	1.49	1.01	0.84	0.99
Frank(5.00)	1.38	0.94	1.26	2.83	1.64	1.14	1.64	1.13	1.42	0.99	0.63	0.83
Gumbel(1.2)	0.79	1.54	1.22	1.64	4.04	1.12	1.01	2.16	1.59	1.41	2.50	2.28
Gumbel(1.4)	0.78	1.47	1.23	1.36	2.53	1.12	0.98	1.94	1.57	1.14	1.83	1.74
Gumbel(1.6)	0.91	1.07	1.23	1.34	1.76	1.13	1.01	1.49	1.53	0.97	1.11	1.29
Gumbel(1.8)	1.07	0.93	1.25	1.27	1.31	1.14	1.15	1.12	1.52	0.93	0.72	1.00
Gumbel(2.0)	1.23	0.71	1.26	1.28	1.08	1.15	1.33	0.95	1.47	0.84	0.44	0.76
Gumbel(2.2)	1.55	0.72	1.27	1.30	1.04	1.16	1.50	0.80	1.38	0.82	0.35	0.61
Gumbel(2.4)	1.82	0.78	1.28	1.41	0.97	1.16	1.72	0.87	1.34	0.78	0.30	0.49
Gumbel(2.6)	2.12	0.85	1.28	1.55	1.03	1.17	1.86	0.83	1.25	0.74	0.28	0.41
Gumbel(2.8)	2.51	1.00	1.29	2.18	1.14	1.17	1.97	0.86	1.14	0.70	0.27	0.33
Gumbel(3.0)	2.91	1.19	1.30	1.93	1.46	1.19	2.24	0.94	1.05	0.69	0.28	0.29
Gaussian(0.07)	0.76	1.61	1.22	1.70	5.69	1.11	0.97	2.03	1.57	1.55	3.13	2.56
Gaussian(0.15)	0.75	1.65	1.22	1.76	4.97	1.11	1.00	1.86	1.57	1.52	2.86	2.48
Gaussian(0.22)	0.75	1.60	1.22	1.78	4.61	1.12	1.01	2.08	1.60	1.41	2.73	2.34
Gaussian(0.30)	0.77	1.60	1.24	1.65	3.75	1.12	0.99	1.90	1.57	1.34	2.27	2.13
Gaussian(0.38)	0.77	1.45	1.24	1.49	3.16	1.12	1.00	1.97	1.59	1.27	2.28	1.96
Gaussian(0.45)	0.83	1.34	1.24	1.37	2.53	1.13	1.03	1.70	1.57	1.18	1.71	1.67
Gaussian(0.53)	0.89	1.19	1.24	1.50	2.19	1.13	1.13	1.62	1.56	1.11	1.36	1.43
Gaussian(0.60)	1.05	1.13	1.25	1.31	1.85	1.14	1.29	1.46	1.51	1.08	1.14	1.19
Gaussian(0.67)	1.35	1.01	1.28	1.66	1.55	1.15	1.57	1.28	1.47	1.00	0.75	0.90
Gaussian(0.75)	1.77	0.90	1.28	1.67	1.44	1.17	1.93	1.05	1.39	0.96	0.51	0.63

Bibliographie

- Beck, M., Pixton, D., 2003. The Ehrhart polynomial of the Birkhoff polytope. *Discrete & Computational Geometry*. 30 (4), 623–637.
- Charpentier, A., Fermanian, J.-D., Scaillet, O., 2006. The estimation of copulas : theory and practice. In : Rank, J. (Ed.), *Copulas : From theory to application in finance*.
- Chen, S., Huang, T., 2007. Nonparametric estimation of copula functions for dependence modeling.
- Cherubini, U., Luciano, E., Vecchiato, W., 2004. *Copula Methods in Finance*. Wiley.
- Deheuvels, P., 1979. La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. *Bulletin de l'Académie Royale de Belgique, Classes de Sciences* 65, 274–292.
- Fermanian, J.-D., Scaillet, O., 2003. Nonparametric estimation of copulas for time series. *Journal of Risk* 95, 25–54.
- Fritsch, F. N., Butland, J., 1984. A method for constructing local monotone piecewise cubic interpolants. *SIAM J. Sci. Statist. Comput.* 5 (2), 300–304.
- Kulpa, T., 1999. On approximation of copulas. *International Journal of Mathematics and Mathematical Sciences* 22 (2), 259–269.
- Li, X., Mikusiński, P., Taylor, M., 1998. Strong approximation of copulas. *Journal of Mathematical Analysis and Applications* 225 (2), 608–623.
- Mirsky, L., 1963. Results and problems in the theory of doubly stochastic matrices. *Probability Theory and Related Fields* 1, 319–334.
- Nelsen, R. B., 1999. *An Introduction to Copulas*. Springer, New York.

Sungur, E. A., Ng, P., 2005. A decomposition of copulas and its use. *Communications in Statistics- Theory and Methods* 34, 2269–2282.

CONCLUSION

Dans cette thèse, nous avons travaillé sur l'estimation bayésienne non paramétrique de copules. Dans un premier temps, nous avons traité le cas des copules de valeurs extrêmes et dans un second temps, celui des copules en général.

Dans le cadre des copules de valeurs extrêmes, nous avons modélisé la fonction de dépendance de manière non paramétrique. Le modèle a été construit à l'aide de noeuds d'interpolation et de fonctions exponentielles. L'évaluation numérique de l'estimateur de Bayes fut effectuée à l'aide d'une chaîne de Markov à sauts réversibles. Nous avons effectué des comparaisons avec les estimateurs classiques et avons observé que notre estimateur est plus performant que les compétiteurs en terme de l'intégrale des erreurs quadratiques moyennes. Ceci est une qualité importante, surtout lorsque l'on travaille avec des valeurs extrêmes. Enfin, notre approche a permis de faire des prévisions concernant les vitesses maximales de courants observées à certaines stations d'Hydro-Québec.

Une approche bayésienne non paramétrique pour l'estimation de copules s'est avérée intéressante sur plusieurs aspects. Premièrement, elle nous a permis d'exploiter la géométrie du problème et de construire des estimateurs demeurant à l'intérieur de la classe des copules. Deuxièmement, lorsque l'intérêt porte sur toute la distribution bivariée et non seulement sur la copule, notre approche a su apporter une solution simple à l'estimation simultanée des lois marginales et de la copule. Enfin, à l'aide de cette approche, nous avons construit des estimateurs lisses, permettant par le fait même l'estimation de la densité.

Les méthodes développées dans cette thèse ont ouvert la porte à de nouveaux travaux. Les estimateurs proposés possèdent certaines faiblesses lorsque la copule se situe proche d'une frontière. De nouvelles distributions a priori devraient être

étudiées. Sur une note plutôt théorique, il pourrait être intéressant de comparer le comportement asymptotique de nos estimateurs à celui des compétiteurs. Enfin, le cas multivarié reste à explorer : dans le premier article, la construction d'hyper-surfaces convexes, interpolantes et lisses peut être envisageable mais difficile. Dans le second article, la dimension du polytope de Birkhoff augmente rapidement, pouvant rendre la convergence des algorithmes MCMC beaucoup plus lente.