

Université de Montréal

Estimation bayésienne d'une fonction de Pickands par des splines cubiques

par

Mohamed Gueye

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

Mémoire présenté à la Faculté des arts et des sciences en vue de l'obtention du grade de
Maître ès science (M.Sc.) en statistique

Juillet 2021

© Mohamed Gueye, 2021

Résumé

Le sujet de notre mémoire est l'intersection entre deux domaines : La théorie des valeurs extrêmes (TVE) et les copules. L'objet de la TVE est de trouver la loi limite du maximum d'un échantillon. Grâce aux résultats de la TVE, on peut modéliser les phénomènes extrêmes. Par ailleurs, il existe une variante bivariée de la TVE. La variante bivariée de la TVE utilise une famille de copules appelées copules de valeurs extrêmes pour tenir compte de la liaison entre les deux phénomènes extrêmes.

En dimension 2, toute copule de valeurs extrêmes dépend d'une fonction de Pickands. L'objet de notre mémoire est d'estimer la fonction de Pickands à partir de données. Nous avons trouvé un moyen de construire une fonction de Pickands grâce à des splines cubiques. À partir de cette construction, on obtient une famille élargie de fonctions de Pickands dans laquelle nous effectuons notre inférence statistique. Nous avons choisi l'approche bayésienne pour construire l'estimateur et les méthodes de MCMC pour les évaluations numériques. La méthode a été appliquée sur des données simulées et réelles.

Mots clés : Copules de valeurs extrêmes, Fonction de Pickands, MCMC

Abstract

The subject of our thesis is intersection between two fields: The Extreme Value Theory (EVT) and copulas. The object of EVT is to find the limit law of the maximum of a sample. Due to the results of EVT, we can model extreme phenomena. In addition, there is a bivariate variant of EVT. The bivariate variant of EVT uses a family of copulas called extreme value copulas to account for the connection between the two extreme events.

Any copula with extreme values depends on a Pickands function. The object of our thesis is to estimate the Pickands function from data. We have found a way to build a Pickands function using cubic splines. From this construction, we obtain an extended family of Pickands functions in which we perform our statistical inference. We chose the Bayesian approach to build the estimator and the MCMC methods for the estimates. The method was applied on simulated and real data.

Keywords : Extrem value copulas, Pickands Function, MCMC

Remerciements

Je tiens d'abord à remercier mon directeur de recherche François Perron qui m'a principalement aidé tout au long de la rédaction de mon mémoire. Au delà de la rédaction du mémoire, il a été indispensable tout au long de mon cursus de maîtrise à l'université de Montréal avec ses conseils, son soutien financier, sa documentation et son accompagnement lors des remplissages des dossiers administratifs. Vu sa renommée, ce n'est pas facile d'être aussi disponible pour un petit étudiant comme moi. J'ai beaucoup appris auprès de lui que ce soit sur la statistique que sur les autres aspects de la vie. Ce fut tout un plaisir d'être encadré par lui.

Je tiens aussi à remercier le corps professoral du département de Mathématique et Statistique pour la qualité de la formation et leur attitude humaine envers les étudiants. Je remercie aussi le personnel administratif pour leur disponibilité.

Je remercie particulièrement ma famille qui a été important tout au long de ma scolarité. Je rends hommage à mon père Modou GUEYE qui est mon modèle de persévérance et de qualité humaine. Mention spéciale à ma mère Maty Ba qui est à l'autre bout du monde mais très proche de moi de par son soutien émotionnel. Je remercie aussi ma tante Binette Ba qui m'a accueilli à ma venue au Canada et a grandement contribué à mon adaptation.

Table des matières

Résumé	i
Abstract	i
Remerciements	ii
Abréviations et symbols mathématiques	vi
Introduction	1
1 Théorie des valeurs extrêmes : Cas univarié	5
1.1 Lois de valeurs extrêmes et max-stables	6
1.2 Détermination du domaine d'attraction	17
2 Copules de valeurs extrêmes : Cas bivarié	23
2.1 Les copules	24
2.2 Copules de valeurs extrêmes	33
2.2.1 Principe	33
2.2.2 Familles paramétriques	37
2.2.3 Estimation	38
3 Estimation bayésienne et méthodes MCMC	44
3.1 Estimation bayésienne	45
3.2 Algorithme de Metropolis-Hastings	46
4 Estimation bayésienne d'une fonction de Pickands par des splines cubiques	52
4.1 Principe théorique de la méthode	54
4.1.1 La bijection	54
4.1.2 Les sous-espaces	55
4.1.3 Le modèle bayésien	61
4.1.4 L'approche numérique	62
4.2 Application de la méthode sur des données simulées	67
4.3 Application de la méthode sur des données réelles	73
Conclusion	78

Annexe A	Mesure spectrale	79
Annexe B	Démonstrations	81
Annexe C	Copule de valeurs extrêmes multivariés	92
Annexe D	Algorithme de simulation des nœuds	96

Table des figures

2.1	Résultats de la simulation des 4 sous-expériences	26
2.2	Simulation bornes des copules de Fréchet-Hoeffding	32
2.3	Exemple de fonctions de Pickands.	36
4.1	Exemples de fonctions $G_n(\cdot \theta)$ et $A_n(\cdot \theta)$ avec 5 noeuds.	57
4.2	Approximation de la fonction de pickands indépendante et comonotique avec 8 noeuds.	60
4.3	Illustration des fonctions de Pickands polygonales \overleftarrow{A} et \overrightarrow{A} pour approcher la fonction de Pickands $t \mapsto A(t) = [(1-t)^2 + t^2]^{1/2}$	64
4.4	Approximation de la fonction de Pickands $t \mapsto A(t) = [(1-t)^2 + t^2]^{1/2}$ avec n noeuds issus des fonctions Pickands polygonales pour $n = 4, 10, 20, 40, 50$	64
4.5	Fonction de Pickands des 15 modèles à simuler	68
4.6	Simulation de fonctions de Pickands à partir de la loi a priori avec 1000 itérations (courbes grises).	72
4.7	Estimation de la fonction de Pickands des 200 échantillons de taille 25 du modèle 2 ($r=1,5$; $\alpha = 1$; $\beta = 1$) avec les trois méthodes. La ligne noire est la vraie fonction de Pickands.	73
4.8	Nuages de points de catégories de compensations par paire	75
4.9	Fonctions de Pickands estimées (moyennes a posteriori) pour chaque paire	76

Liste des abréviations

MCMC	Monte Carlo par Chaînes de Markov
v.a.	Variable aléatoire
i.i.d.	Indépendantes et identiquement distribuées
GEV	Generalized Extrem Value
TVE	Théorie des Valeurs Extrêmes
$\mathcal{N}(\mu, \sigma^2)$	Loi normale de moyenne μ et de variance σ^2
$\exp(\lambda)$	Loi exponentielle de paramètre λ
MEMV	Méthode de l'estimateur du maximum de vraisemblance
$a \vee b$	Maximum entre a et b
$a \wedge b$	Minimum entre a et b

Introduction générale

En modélisation statistique, on cherche souvent la loi que suivent les données. Cependant, il est inimaginable de chercher la bonne loi parmi toutes les lois qui existent. Même les méthodes non paramétriques n'y parviennent pas. C'est pourquoi les statisticiens se restreignent à une famille de lois. Dans la pratique, on maîtrise souvent le phénomène étudié et donc la famille à laquelle appartient la loi que suivent les données.

Il existe une famille de lois appelée famille des lois de valeurs extrêmes qui permet de modéliser les phénomènes rares. Si X est une v.a. mesurant un phénomène rare, l'enjeu est de déterminer la probabilité que le phénomène étudié dépasse un seuil s c.à.d. $\mathbb{P}(X > s)$. Ce seuil est souvent très éloigné de zéro d'où le nom de valeur extrême. La bonne estimation de cette probabilité s'avère cruciale car l'occurrence de ces événements extrêmes peut avoir comme conséquence un crash boursier, la ruine d'un assureur, un tremblement de terre ou une quelconque catastrophe naturelle. Qu'est-ce qui nous amène à penser que la v.a. X suit une loi de valeurs extrêmes ? Les lois de valeurs extrêmes sont [max-stables](#) et sont les [domaines d'attraction](#) (concepts étudiés au chapitre 1). Gnedenko (1943) et Fisher (1928) sont les premiers à avoir déterminé la forme de la fonction de répartition des lois de valeurs extrêmes.

Cependant, il peut arriver qu'on veuille modéliser deux phénomènes rares simultanément. Soit X_1 et X_2 deux variables aléatoires mesurant deux phénomènes rares. Les v.a. X_1 et X_2 appartiennent au même espace probabilisé, donc il se peut qu'elles ne soient pas indépendantes. Alors, il est fondamental de tenir compte de la dépendance entre les v.a. X_1 et X_2 dans la modélisation, en trouvant la loi bivariée de (X_1, X_2) . Soit G la fonction de répartition de (X_1, X_2) . Si G_1 et G_2 sont les marginales de G alors il existe une fonction C sur $[0, 1] \times [0, 1]$ telle que $G(x_1, x_2) = C(G(x_1), G(x_2))$ (Sklar (1959)). La fonction C est appelée copule. Donc, pour déterminer G il suffit de connaître G_1 , G_2 et C .

Par extension G appartient à la famille des lois de valeurs extrêmes bivariées. De même, les distributions de valeurs extrêmes bivariées sont les [domaines d'attraction bivariés](#). Donc, les marginales G_1 et G_2 suivent une loi de valeurs extrêmes (univariée). On rappellera que la copule des lois de valeurs extrêmes bivariées est nécessairement de la forme :

$$C(u, v) = \exp \left\{ \log(uv) A \left(\frac{\log(u)}{\log(uv)} \right) \right\},$$

où A est la fonction de Pickands (voir Pickands (1981) et Deheuvels (1984)). La fonction de Pickands A caractérise la copule de valeurs extrêmes. La fonction de Pickands vérifie les conditions suivante :

(1) A est convexe

(2) $t \vee (1 - t) \leq A(t) \leq 1, t \in [0, 1]$

Les conditions (1) et (2) impliquent que $A(0) = A(1) = 1$. L'objet de notre étude est d'estimer la fonction Pickands à partir de données de la copule.

Dans Pickands (1981), l'auteur a construit le premier estimateur de la fonction de Pickands. Mais, cet estimateur ne vérifie pas toujours les conditions (1) et (2). Deheuvels (1984) a amélioré l'estimateur de Pickands de sorte qu'il soit égal à 1 sur les bords, mais l'estimateur de Deheuvels ne vérifie pas toujours les conditions (1) et (2). Tadjvidi proposent un estimateur qui vérifie toutes les conditions d'une fonction de Pickands (voir Hall and Tajvidi (2000)). Les estimateurs de Pickands, de Deheuvels et de Tadjvidi sont basés sur le même principe théorique. Capéraà et al. (1997) utilisent une autre approche pour proposer un estimateur qui empiriquement performe mieux que les estimateurs précédents même s'il ne vérifie pas toujours les conditions (1) et (2).

L'estimateur que nous proposons est bayésien. Guillote et Perron ont aussi proposé un estimateur bayésien basé sur la construction d'une fonction de Pickands à partir de polynômes (voir Guillotte and Perron (2008)). Notre estimateur bayésien est basé sur une autre approche. En fait, nous avons construit une fonction de Pickands à partir de splines cubiques. L'estimateur bayésien est donc construit à partir d'un modèle bayésien dont les paramètres sont des splines cubiques. L'évaluation s'effectue grâce à une méthode de Monte Carlo par Chaînes de Markov. En outre, notre estimateur bayésien vérifie toutes les conditions d'une fonction de Pickands.

Notre document est divisé en quatre chapitres :

- Dans le chapitre 1, nous présentons les lois de valeurs extrêmes univariées. Dans la foulée, nous introduisons la notion de domaine d'attraction et de loi max-stable qui sont nécessaires pour déterminer la forme de la fonction de répartition des lois de valeurs extrêmes univariées.

- Le chapitre 2 est une extension du chapitre 1, puisque nous y présentons les copules de valeurs extrêmes bivariées. En fait, dans le chapitre 2, nous introduisons la variante bivariée des lois de valeurs extrêmes. L'élément nouveau qui permet de passer des lois de valeurs extrêmes univariées aux lois de valeurs extrêmes bivariées est la copule de valeurs extrêmes. L'objet de notre étude étant l'estimation de la fonction de Pickands qui permet d'identifier toute copule de valeurs extrêmes, nous présenterons les propriétés et certaines méthodes d'estimation de la fonction de Pickands.
- Dans le chapitre 3, nous présentons les outils nécessaires pour bien comprendre le principe théorique et l'évaluation numérique de notre estimateur bayésien de la fonction de Pickands. Plus précisément, nous y présentons le principe théorique des estimateurs bayésiens et les méthodes MCMC pour l'évaluation numérique des estimateurs bayésiens.
- Le chapitre 4 est l'entrée en matière. Nous y présentons d'abord le principe théorique de notre méthode d'estimation bayésienne de la fonction de Pickands. Ensuite, nous évaluerons numériquement cette méthode sur des données simulées et enfin sur des données réelles.

Chapitre 1

Théorie des valeurs extrêmes : Cas univarié

Introduction

Dans ce chapitre nous introduisons les distributions de valeurs extrêmes univariées. Pour un échantillon X_1, \dots, X_n i.i.d. de loi continue de fonction de répartition F et de maximum $M_n = \max_{1 \leq i \leq n} X_i$, on cherche deux suites b_n et $a_n > 0$ tel que $(M_n - b_n)/a_n$ converge vers une loi non dégénérée de fonction de répartition G . On dit que la fonction de répartition F appartient au domaine d'attraction de G . L'ensemble des lois de fonction de répartition G correspond à l'ensemble des lois de valeurs extrêmes univariées. À partir de cette définition des lois de valeurs extrêmes, nous allons déterminer pas à pas la forme de la fonction de répartition des lois de valeurs extrêmes. Pour cela, nous devons définir les lois max-stables. Il se trouve que la classe des lois de valeurs extrêmes et la classe des

lois max-stables coïncident.

Une fois avoir défini et déterminé certaines propriétés des domaines d'attraction, est-il possible de déterminer le domaine d'attraction auquel appartient une loi quelconque si bien il existe? La meilleure réponse que nous avons trouvé sont des conditions suffisantes qui, lorsque satisfaites par la loi en question permettent de déterminer son domaine d'attraction.

1.1 Lois de valeurs extrêmes et max-stables

Soit $\{X_1, \dots, X_n\}$ un échantillon i.i.d. de loi continue dont la fonction de répartition est F et M_n est le maximum de l'échantillon (c.à.d. $M_n = \max_{1 \leq i \leq n} X_i$). A priori, on s'intéresse à la convergence en loi de M_n dont la fonction de répartition est :

$$\mathbb{P}(M_n \leq x) = F(x)^n$$

Posons $x^F = \sup \{x : F(x) < 1\}$: le supremum essentiel.

Si x^F est fini alors

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq x) = \begin{cases} 0 & \text{si } x < x^F, \\ 1 & \text{si } x \geq x^F. \end{cases}$$

Si $x^F = \infty$ alors

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq x) = 0 \text{ pour tout } x \in \mathbb{R}.$$

Lorsque x^F est fini alors M_n converge vers une loi dégénérée. Lorsque x^F est infini alors M_n diverge presque sûrement vers l'infini. Pour aller plus loin on cherche à déterminer

deux suites $a_n > 0$ et b_n telles que $(M_n - b_n)/a_n$ converge vers une loi non dégénérée.

Exemple 1.1 (*Distribution exponentielle*)

Supposons que X_1, \dots, X_n sont de loi exponentielle de paramètre 1.

$$F_{M_n}(t) = F^n(t) = \begin{cases} (1 - \exp(-t))^n & \text{si } t \geq 0, \\ 0 & \text{si } t < 0. \end{cases}$$

Posons $a_n = 1$ et $b_n = \log(n)$.

Si $x > -\log(n)$ alors

$$\begin{aligned} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \mathbb{P}(M_n - \log(n) \leq x) \\ &= F_{M_n}(x + \log(n)) \\ &= \left(1 - \frac{\exp(-x)}{n}\right)^n \\ &\xrightarrow[n \rightarrow \infty]{} \exp(-\exp(-x)) \equiv \Lambda(x) \text{ pour tout } x \in \mathbb{R}. \end{aligned}$$

Λ est la loi de Gumbel.

L'exemple 1.1 montre que parfois de telles suites existent. On dit que la loi $\exp(1)$ appartient au domaine d'attraction de la loi de Gumbel.

Définition 1.1

1. Soit G la fonction de répartition d'une loi non dégénérée. On dit que F appartient au domaine d'attraction de G s'il existe deux suites $a_n > 0$ et b_n telles que :

$$\lim_{n \rightarrow \infty} F(a_n x + b_n)^n = G(x), \text{ pour tout } x \in \mathbb{R}.$$

On le note $F \in D(G)$.

2. Si $D(G) \neq \emptyset$ alors on dit que G est une loi de valeurs extrêmes.

On s'intéresse aux propriétés et à la forme des lois de valeurs extrêmes. La première propriété très importante est que les lois de valeurs extrêmes sont les lois [max-stables](#).

Définition 1.2 Soit G une fonction de répartition. On dit que G est max-stable s'il existe deux suites $a_n > 0$ et b_n telles que :

$$G(a_n x + b_n)^n = G(x), \text{ pour tout } n \in \mathbb{N}^*, x \in \mathbb{R}.$$

Remarques :

- Soit Y_1, \dots, Y_n un échantillon i.i.d. de fonction de répartition G et Z_n le maximum de l'échantillon. Si G est max-stable avec les suites $a_n > 0$ et b_n alors la loi de $(Z_n - b_n)/a_n$ est la même que celle de Y_1 quelque soit n . En d'autres termes, le maximum normalisé de l'échantillon a la même loi que chacune des variables qui forme l'échantillon quelque soit la taille de l'échantillon, d'où le nom de loi max-stable.
- Si G appartient à la classe des distributions max-stables alors $G \in D(G)$. Il en découle que la classe des lois max-stables est incluse dans la classe des lois de valeurs extrêmes.

Exemple 1.2 L'exemple 1.1 nous a montré que la loi de Gumbel appartient à la classe des lois de valeurs extrêmes. Peut-on montrer qu'elle est max-stable ?

Posons $a_n = 1$ et $b_n = \log(n)$

$$\begin{aligned}
\Lambda(a_n x + b_n) &= \exp \{ - \exp \{ -a_n x - b_n \} \} \\
&= \exp \{ - \exp \{ -x + \log \left(\frac{1}{n} \right) \} \} \\
&= \exp \left\{ - \frac{\exp \{ -x \}}{n} \right\}, \text{ pour tout } x \in \mathbb{R}.
\end{aligned}$$

$$\Lambda(a_n x + b_n)^n = \exp \{ - \exp \{ -x \} \} = \Lambda(x), \text{ pour tout } n \in \mathbb{N}^*, x \in \mathbb{R}.$$

Alors, la loi de Gumbel Λ appartient à la classe des lois max-stables.

Théorème 1.1 *La classe des lois de valeurs extrêmes et la classe des lois max-stables coïncident.*

Démonstration :

On sait déjà que la classe des lois max-stables est incluse dans la classe des lois extrêmes.

Maintenant, démontrons la réciproque :

Soit G la fonction de répartition pour une loi de valeurs extrêmes telle que

$(Z_n - b_n) / a_n$ converge en loi vers G et Z_n est le maximum d'un échantillon Y_0, Y_1, \dots, Y_n

i.i.d dont la fonction de répartition est H .

Posons $\tilde{Z}_j^{(n)} = \max_{(j-1)n \leq i \leq jn} \{Y_i\}$ le maximum par blocs de l'échantillon.

On note que : $Z_{nk} = \max_{1 \leq j \leq k} \{ \tilde{Z}_j^{(n)} \}$

$$\Rightarrow \mathbb{P} \left(\frac{Z_{nk} - b_n}{a_n} \leq x \right) = \mathbb{P} \left(\frac{\tilde{Z}_1^{(n)} - b_n}{a_n} \leq x \right)^k \xrightarrow{n \rightarrow \infty} G^k(x) \quad (a).$$

La loi faible des maximums partiels nous montre que pour tout échantillon i.i.d T_1, \dots, T_k ,

il existe deux constantes $A_k > 0$ et B_k telles que : $\max_{1 \leq j \leq k} \{T_j\} = A_k T_1 + B_k$ (voir Krizmanic

(2014))

Donc, il existe des constantes $\alpha_k > 0$ et β_k tel que : $Z_{nk} = \alpha_k \tilde{Z}_1^{(n)} + \beta_k$.

Or $\tilde{Z}_1^{(n)} = Z_n$

Ce qui prouve que $\mathbb{P}\left(\frac{Z_{nk}-b_n}{a_n} \leq x\right) \xrightarrow{n \rightarrow \infty} G\left(\frac{x-\alpha_k}{\beta_k}\right)$ (b).

Les résultats (a) et (b) impliquent que $G^k(\alpha_k x + \beta_k) = G(x)$, pour tout $x \in \mathbb{R}, k \in \mathbb{N}^*$

Comme k est un entier arbitraire alors la relation est vrai pour tout entier k . D'où, la loi G est max-stable.

□

Le théorème 1.1 nous montre que toutes les lois de valeurs extrêmes sont max-stables et inversement. On peut dès lors déterminer la forme des lois de valeurs extrêmes puisqu'elles sont max-stables. C'est ainsi que Fisher (1928) et Gnedenko (1943) ont déterminé la forme des lois de valeurs extrêmes (max-stables).

Théorème 1.2 (Fisher et Tippett (1928) et Gnedenko (1943))

Soit $x, \mu \in \mathbb{R}, \sigma > 0$. Posons $y = (x - \mu) / \sigma$. Toute loi de valeurs extrêmes appartient à un des types de lois suivants :

1. Gumbel (Type I) : $\Lambda^{(\mu, \sigma)}(x) = \exp(-\exp(-y))$.

2. Fréchet (Type II) : $\phi_\alpha^{(\mu, \sigma)}(x) = \begin{cases} 0 & \text{si } y \leq 0, \\ \exp(-y^{-\alpha}) & \text{si } y > 0, \alpha > 0. \end{cases}$

3. Weibull (Type III) : $\psi_\alpha^{(\mu, \sigma)}(x) = \begin{cases} \exp(-(-y)^\alpha) & \text{si } y \leq 0, \alpha > 0, \\ 1 & \text{si } y > 0. \end{cases}$

Dans tous les cas μ est le paramètre de valeur centrale et σ le paramètre de dispersion.

Démonstration :

Supposons que $F \in D(G)$ tel que :

$$\lim_{n \rightarrow \infty} F(a_n x + b_n)^n = G(x), \text{ pour tout } x \in \mathbb{R}.$$

D'après le théorème 1.1, G est max-stable. Donc, il existe deux suites $c_k > 0$ et d_k telles que :

$$G(x)^k = G(c_k x + d_k), \text{ pour tout } k \in \mathbb{N}^*, x \in \mathbb{R}.$$

NB : Comme nous le verrons après, le comportement des suites $c_k > 0$ et d_k détermine lequel des trois types de distribution auquel G appartient.

En considérant les blocs on a :

$$\begin{aligned} G(x)^{nm} &= G(c_m x + d_m)^n = G(c_n (c_m x + d_m) + d_n) \\ &= G(c_n c_m x + (c_n d_m + d_n)) \\ &= G(c_{nm} x + d_{nm}). \end{aligned}$$

Comme G est bijective (strictement croissante et continue) sur $\{x : 0 < G(x) < 1\}$ on a :

$$c_{nm} = c_n c_m \text{ et } d_{nm} = c_n d_m + d_n, \text{ pour tout } n, m \in \mathbb{N}^*.$$

Dès lors $c_1 = c_1^2 \Rightarrow c_1 = 1$ et $d_1 = 0$. Pour la suite de la démonstration, nous devons trouver l'expression de $G(x)^r$ pour $r \in \mathbb{Q}$.

Supposons que $r = \frac{m}{n}$ où $n, m \in \mathbb{N}^*$. Par changement de variable on a :

$$G(x)^{1/n} = G\left(\frac{x - d_n}{c_n}\right).$$

$$\begin{aligned} G(x)^r &= G\left(\frac{x-d_n}{c_n}\right)^m = G\left(c_m \frac{x-d_n}{c_n} + d_m\right) \\ &= G\left(\frac{c_m}{c_n}x - \frac{c_m d_n}{c_n} + d_m\right). \end{aligned}$$

On a trois cas possibles :

- $c_k = 1$, pour tout $k \in \mathbb{N}^*$
- $d_k = 0$, pour tout $k \in \mathbb{N}^*$ et $G(0) = 0$
- $d_k = 0$, pour tout $k \in \mathbb{N}^*$ et $G(0) = 1$

Dans chaque cas on essaiera de trouver l'expression de $G(x)^u$ pour $u > 0$.

Cas où $c_k = 1$ pour tout $k \in \mathbb{N}^*$

Pour un nombre rationnel $r = m/n$ on a : $G(x)^r = G(x + \lambda_r)$ où $\lambda_r = d_m - d_n$.

Pour x tel que $0 < G(x) < 1$, $G(x)^r$ est strictement décroissante. Par bijectivité de G sur

$\{t : 0 < G(t) < 1\}$, la suite λ_r est strictement décroissante. Pour $r = m/n$ et $s = j/k$ on

a :

$$\lambda_{rs} = d_{mj} - d_{nk} = d_m + d_j - d_n - d_k = \lambda_r + \lambda_s.$$

Posons $\delta(u) = \inf_{0 < r \in \mathbb{Q} < u} \lambda_r$ pour $u \in \mathbb{R}$.

La fonction δ est décroissante puisque la suite λ_r est strictement décroissante en r . Pour

tout réel $u > 0$, il existe une suite de nombres rationnels q_n telle que $q_n \uparrow u$. Par continuité de la fonction $u \mapsto G(x)^u$, $G(x)^{q_n} \xrightarrow[n \rightarrow \infty]{} G(x)^u$.

Par autre, $G(x)^{q_n} = G(x + \lambda_{q_n})$ et $\lambda_{q_n} \xrightarrow[n \rightarrow \infty]{} \delta(u)$ (à cause de la décroissance de λ_r).
 $\Rightarrow G(x)^{q_n} \xrightarrow[n \rightarrow \infty]{} G(x + \delta(u))$.

Alors, $G(x)^u = G(x + \delta(u))$, pour tout $u > 0$, $x \in \mathbb{R}$.

$\Rightarrow \delta(uv) = \delta(u) + \delta(v)$ (équation de Hamel).

D'après Sahoo and Kannappan (2011), la solution de l'équation de Hamel est :

$\delta(u) = -c \log(u)$ où $c > 0$,

$G(x) = G(x - c \log(u))^{1/u}$.

En posant $u = \exp\{x/c\}$ on a :

$$G(x) = \exp\{\log(G(0)) \exp\{-x/c\}\}.$$

Ce qui prouve que si $c_k = 1$ pour tout $k \in \mathbb{N}^*$, G suit une loi de Gumbel.

Cas où $d_k = 0$ pour tout $k \in \mathbb{N}^*$ et $G(0) = 0$

Par analogie avec le cas précédent pour tout $r = m/n \in \mathbb{Q}^+$ on a :

$$G(x)^r = G(\beta_r x), \text{ avec } \beta_r = c_m/c_n.$$

De même, $\beta_{rs} = \beta_r \beta_s$ et β_r est strictement décroissante.

Posons, $\beta(u) = \inf_{0 < r \in \mathbb{Q} < u} \beta_r$ pour $u \in \mathbb{R}$,

Avec la même méthode que le cas précédent, on obtient :

$$G(x)^u = G(\beta(u)x), \text{ pour tout } u > 0, x \in \mathbb{R},$$

$\Rightarrow \beta(u + v) = \beta(u)\beta(v)$ (equation de Hamel).

La solution de l'équation de Hamel est : $\beta(u) = u^{-c}$ où $c > 0$.

Il en découle que :

$$G(x) = G(xu^{-c})^{1/u}, \text{ pour tout } u > 0.$$

En posant $u = x^{1/c}$ on obtient :

$$G(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ G(1)^{x^{-1/c}} & \text{si } x > 0. \end{cases}$$

Ce qui prouve que si $d_k = 0$, pour tout $k \in \mathbb{N}^*$ et $G(0) = 0$, G appartient aux distributions de type Fréchet.

Cas où $d_k = 0$ pour tout $k \in \mathbb{N}^*$ et $G(0) = 1$

On utilise le même principe que le cas précédent avec $\beta(u) = (-u)^c$ où $c > 0$.

$$\text{Finalement, } G(x) = \begin{cases} G(-1)^{(-x)^{1/c}} & \text{si } x \leq 0, \\ 1 & \text{si } x > 0. \end{cases}$$

Ce qui prouve que si $d_k = 0$, pour tout $k \in \mathbb{N}^*$ et $G(0) = 1$, G appartient aux distributions de type Weibull.

Ainsi s'achève la démonstration du théorème.

□

Table 1.1 – Paramètres de position et d'échelle des lois de valeurs extrêmes

Type de distribution	Moyenne	Variance	Mode
Gumbel (Λ)	$\gamma=0,5772\dots$	$\pi^2/6$	0
Fréchet (ϕ_α)	$\Gamma(1 - 1/\alpha)$ pour $\alpha > 1$	$\Gamma(1 - 2/\alpha) - \Gamma^2(1 - 1/\alpha)$ pour $\alpha > 2$	$(1 + 1/\alpha)^{-1/\alpha}$
Weibull (ψ_α)	$-\Gamma(1 + 1/\alpha)$	$\Gamma(1 + 2/\alpha) + \Gamma^2(1 + 1/\alpha)$	0 pour $0 \leq \alpha \leq 1$

γ est la constante d'Euler et Γ est la fonction gamma

Un cadre unificateur des lois de valeurs extrêmes est proposée. Ces lois s'appellent lois GEV (Generalized Extreme value). Les lois GEV résument les trois types de lois en une seule famille de lois qui dépend d'un paramètre ξ . Ce dernier détermine l'appartenance de la loi à l'une des trois types.

Définition 1.3 (*Loi GEV*)

Soit $\xi, x, \mu \in \mathbb{R}, \sigma > 0$. Posons $y = (x - \mu) / \sigma$. Soit G_ξ la fonction de répartition de la loi GEV d'indice ξ , de valeurs centrale μ et de dispersion σ . La fonction G_ξ s'écrit sous la forme :

$$G_\xi(x) = \begin{cases} \exp\left(-(1 + \xi y)^{-1/\xi}\right), & 1 + \xi y > 0, \text{ si } \xi \neq 0, \\ \exp(-\exp(-y)) , & y \in \mathbb{R}, \text{ si } \xi = 0. \end{cases}$$

Le paramètre ξ est l'Indice de Valeurs Extrêmes (IVE).

L'IVE est le paramètre le plus important car il caractérise la structure de G_ξ et facilite l'inférence statistique. De même, si $\xi > 0$, $\xi = 0$ ou $\xi < 0$ alors G appartient respectivement aux lois de Fréchet, Gumbel et Weibull.

Lemme 1.1 Soit G_ξ la fonction de répartition de la loi GEV :

1. Si $\xi = 0$ alors G_ξ est la fonction de répartition de la loi de Gumbel.
2. Si $\xi > 0$ alors $x \mapsto G_\xi((x - 1)/\xi)$ est la fonction de répartition de la loi de Fréchet d'indice $1/\xi$.
3. Si $\xi < 0$ alors $x \mapsto G_\xi(-(1 + x)/\xi)$ est la fonction de répartition de la loi de Fréchet d'indice $1/\xi$.

Les lois de Fréchet, Weibull et Gumbel se différencient aussi par la vitesse à laquelle leur densité décroît sur les ailes :

1. Les lois de Fréchet sont à aile lourde à droite, c'est-à-dire que leur densité décroît lentement sur l'aile droite. Cela a comme conséquence une différence de probabilités faible entre deux valeurs extrêmes plus grandes que la moyenne.
2. Les lois de Weibull sont à aile légère à gauche, c'est-à-dire que leur densité décroît très rapidement sur l'aile gauche. Cela a comme conséquence une différence de probabilités élevées entre deux valeurs extrêmes plus petites que la moyenne.
3. La loi de Gumbel a des ailes qui ne décroissent pas aussi lentement que les lois de Fréchet mais ne décroissent pas aussi rapidement que les lois de Weibull.

Dans plusieurs applications le choix du type d'aile dépend de la nature du phénomène étudié. En science environnementale, on suppose souvent que $\xi=0$. En finance, pour ne pas minimiser les risques on suppose que $\xi > 0$.

Table 1.2 – Vitesse de décroissance des distributions de valeurs extrêmes

x	$\mathbb{P}(X_\xi > x)$				$\mathbb{P}(X_\xi < -x)$		
	$\xi = 0$	$\xi = 0.5$	$\xi = 1$	$\xi = 2$	$\xi = -0.5$	$\xi = -1$	$\xi = -2$
10	4,54e-05	2,74e-02	8,69e-02	1,96e-01	2,32e-16	1,67e-05	1,03e-02
10^2	≈ 0.00	3,84e-04	9,85e-03	6,81e-02	≈ 0.00	1,37e-44	6,96e-07
10^3	≈ 0.00	3,98e-06	9,99e-04	2,21e-02	≈ 0.00	≈ 0.00	3,74e-20
10^3	≈ 0.00	4,00e-08	10,00e-05	7,05e-03	≈ 0.00	≈ 0.00	3,80e-62
10^4	≈ 0.00	4,00e-10	10,01e-06	2,23e-03	≈ 0.00	≈ 0.00	5,99e-195

Pour $\xi > 0$, la variation sur l'aile droite est plus lente au fur et à mesure que ξ augmente.

Pour $\xi < 0$, la variation sur l'aile gauche est plus rapide au fur et à mesure que ξ est proche de 0.

1.2 Détermination du domaine d'attraction

Soit F la fonction de répartition d'une variable aléatoire continue. Est-il possible de démontrer que F appartient à un domaine d'attraction ? Si oui, quel est ce domaine d'attraction ?

Malheureusement, il n'existe pas encore de méthode pour répondre à ces questions si F est quelconque. Néanmoins, sous certaines conditions liées à la notion de [variation régulière à l'infini](#), F admet un domaine d'attraction dont on connaît la forme et les constantes de normalisation $a_n > 0$ et b_n .

Définition 1.4 Soit $\delta > 0$ et L une fonction mesurable sur $[\delta, \infty)$ et $\alpha \in \mathbb{R}$. On dit que la

fonction L est à variation régulière à l'infini avec exposant α (noté $L \in \mathcal{VR}(\alpha)$) si :

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = x^\alpha, \text{ pour tout } x > 0.$$

En particulier, si $\alpha = 0$ on dit que L est à variation lente à l'infini (noté $L \in \mathcal{VL}$).

Théorème 1.3 (Bingham et al. (1989))

Soit $L \in \mathcal{VR}(\alpha)$.

Il existe $c \in \mathcal{VL}$ tel que : $L(x) = x^\alpha c(x)$.

Le théorème 1.3 permettra de démontrer le théorème 1.4.

Théorème 1.4

Soit $\alpha > 0$. Posons $x^F = \sup \{x : F(x) < 1\}$

1. Si $1 - F \in \mathcal{VR}(-\alpha)$ ($x^F = \infty$), alors $F \in D\left(\phi_\alpha^{(0,1)}\right)$ avec les constantes de normalisation $a_n = \inf \{x : 1 - F(x) \leq 1/n\}$ et $b_n = 0$;
2. Si $x^F < \infty$ et $1 - F^* \in \mathcal{VR}(-\alpha)$ où $F^*(x) = F\left(x^F - \frac{1}{x}\right)$, alors $F \in D\left(\psi_\alpha^{(0,1)}\right)$ avec les constantes de normalisation $a_n = \inf \{x : 1 - F\left(x^F - x\right) \leq 1/n\}$ et $b_n = x^F$;
3. S'il existe une fonction R tel que :

$$\lim_{t \rightarrow \infty} \frac{1 - F(t + xR(t))}{1 - F(t)} = e^{-x}, \text{ pour tout } x \in \mathbb{R}$$

alors $F \in D\left(\Lambda^{(0,1)}\right)$ avec les constantes de normalisation $a_n = R(b_n)$ et $b_n = \inf \{x : 1 - F(x) \leq 1/n\}$.

Démonstration :

1. Supposons que $1 - F \in \mathcal{VR}(-\alpha)$ et posons $a_n = \inf \{x : 1 - F(x) \leq 1/n\} \rightarrow \infty$ et

$$b_n = 0$$

D'après le théorème 1.3, il existe $c \in \mathcal{VL}$ tel que : $1 - F(x) = x^{-\alpha}c(x)$, pour tout $x >$

0

$$\begin{aligned} \lim_{n \rightarrow \infty} F(a_n x + b_n)^n &= \lim_{n \rightarrow \infty} (1 - (a_n x)^{-\alpha} c(a_n x))^n \\ &= \exp \left\{ - \lim_{n \rightarrow \infty} n (a_n x)^{-\alpha} c(a_n x) \right\} \\ &= \exp \left\{ -x^{-\alpha} \lim_{n \rightarrow \infty} n a_n^{-\alpha} c(a_n x) \right\} \\ &= \exp \left\{ -x^{-\alpha} \lim_{n \rightarrow \infty} n a_n^{-\alpha} c(a_n) \right\}, \text{ comme } c \in \mathcal{VL} \text{ et } a_n \rightarrow \infty \\ &= \exp \{-x^{-\alpha}\}, \text{ par définition de } a_n \end{aligned}$$

2. Supposons que $x^F < \infty$ et $1 - F^* \in \mathcal{VR}(-\alpha)$ où $F^*(x) = F(x^F - \frac{1}{x})$.

Posons $a_n = \inf \{x : 1 - F(x^F - x) \leq 1/n\} \rightarrow 0^+$ et $b_n = x^F$ et $x < 0$.

Il existe $c \in \mathcal{VL}$ tel que : $1 - F(x) = (x^F - x)^\alpha c(1/(x^F - x))$

$$\begin{aligned} \lim_{n \rightarrow \infty} F(a_n x + b_n)^n &= \lim_{n \rightarrow \infty} [1 - (-a_n x)^\alpha c(1/(-a_n x))]^n \\ &= \exp \left\{ - \lim_{n \rightarrow \infty} n (-a_n x)^\alpha c(1/(-a_n x)) \right\} \\ &= \exp \left\{ -(-x)^\alpha \lim_{n \rightarrow \infty} n a_n^\alpha c(1/(-a_n x)) \right\} \\ &= \exp \left\{ -(-x)^\alpha \lim_{n \rightarrow \infty} n a_n^\alpha c(1/a_n) \right\}, \text{ comme } c \in \mathcal{VL} \text{ et } a_n \rightarrow 0^+ \\ &= \exp \{-(-x)^\alpha\}, \text{ par définition de } a_n \end{aligned}$$

3. Supposons qu'il existe une fonction R tel que : $\lim_{t \rightarrow \infty} \frac{1 - F(t + xR(t))}{1 - F(t)} = e^{-x}$, , pour tout $x \in$

\mathbb{R}

Posons $a_n = R(b_n)$ et $b_n = \inf \{x : 1 - F(x) \leq 1/n\}$

$$\begin{aligned}\lim_{n \rightarrow \infty} F(a_n x + b_n)^n &= \lim_{n \rightarrow \infty} [1 - (1 - F(a_n x + b_n))]^n \\ &= \exp \left\{ - \lim_{n \rightarrow \infty} n (1 - F(a_n x + b_n)) \right\} \\ &= \exp \left\{ - \lim_{n \rightarrow \infty} n (1 - F(R(b_n)x + b_n)) \right\} \\ &= \exp \left\{ - \lim_{n \rightarrow \infty} n \exp\{-x\} (1 - F(b_n)) \right\} \\ &= \exp \{- \exp\{-x\}\}\end{aligned}$$

□

Exemple 1.3 Soit F la fonction de répartition de la loi de Pareto de paramètre $\alpha > 0$:

$$F(x) = \begin{cases} 1 - \frac{1}{x^\alpha} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

En fait, $\lim_{t \rightarrow \infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha} \Rightarrow F \in \mathcal{VR}(-\alpha)$.

Il en découle que la loi de Pareto de paramètre α appartient au domaine d'attraction de

Fréchet de paramètre α avec les constantes de normalisation $a_n = n^{1/\alpha}$ et $b_n = 0$.

Exemple 1.4 Soit F la fonction de répartition de la loi Bêta de paramètre $\alpha > 1$ et $\beta > 1$:

$$\begin{aligned}1 - F(x) &\propto \frac{\Gamma(\alpha+\beta)}{\beta\Gamma(\alpha)\Gamma(\beta)}(1-x)^\beta, \text{ pour tout } 0 < x < 1 \text{ et } x^F = 1 \\ \Rightarrow \lim_{t \rightarrow \infty} \frac{1-F(1-tx)}{1-F(1-t)} &= x^\beta\end{aligned}$$

Il en découle que la loi Bêta de paramètre α et β appartient au domaine d'attraction de

Fréchet de paramètre α avec les constantes de normalisation $a_n = \left[\frac{\Gamma(\alpha)\Gamma(\beta+1)}{n\Gamma(\alpha+\beta)} \right]^{1/\beta}$ et $b_n = 1$

Les lois suivantes appartiennent au domaine d'attraction de Fréchet :

- Pareto $Pa(\alpha)$: $F(x)$, pour $x > 1$ et $\alpha > 0$. On a $\xi = \frac{1}{\alpha}$;
- Pareto Généralisée $GP(\sigma, \alpha)$: $F(x) = 1 - \left(1 + \alpha \frac{x}{\sigma}\right)^{-\frac{1}{\alpha}}$, $x > 0$ et $\sigma, \alpha > 0$. On a $\xi = \alpha$;
- Burr (η, τ, λ) : $F(x) = 1 - \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda$, $x > 0$ et $\eta, \tau, \lambda > 0$. On a $\xi = \frac{1}{\lambda\tau}$;
- Fréchet (α) : $F(x) = \exp(-x^{-\alpha})$, pour $x > 0$ et $\alpha > 0$. On a $\xi = \frac{1}{\alpha}$;
- t -Student $t(\nu)$: On a $\xi = \frac{1}{\nu}$;
- Cauchy : $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x)$. On a $\xi = 1$;
- Log - Gamma : $F(x) = \int_1^x \frac{\lambda^\alpha}{\Gamma(\alpha)} (\log(t))^{\alpha-1} t^{-\lambda-1} dt$. On a $\xi = \frac{1}{\lambda}$.

Les lois suivantes appartiennent au domaine d'attraction de Weibull :

- Uniforme $\mathcal{U}(0, 1)$: $F(x) = x$, $0 < x < 1$. On a $\xi = -1$;
- Inverse Burr (η, τ, λ) : $F(x) = 1 - \left(\frac{\eta}{\eta + (-x)^{-\tau}}\right)^\lambda$, $x > 0$ et $\eta, \tau, \lambda > 0$. On a $\xi = \frac{-1}{\lambda\tau}$;
- Weibull maxima (α) : $F(x) = 1 - \exp(-x^{-\alpha})$, $x > 0$ et $\alpha > 0$. On a $\xi = \frac{-1}{\alpha}$.

Les lois suivantes appartiennent au domaine d'attraction de Gumbel :

- Exponentiel Exp(1) : $F(x) = 1 - e^{-x}$, $x > 0$;
- Weibull minima (τ, λ) : $F(x) = 1 - \exp(-\lambda x^\tau)$, $x > 0$ et $\tau, \lambda > 0$;
- Gumbel : $F(x) = \exp(-\exp(-x))$;

- *Normale* $\mathcal{N}(0, 1)$;
- *Lognormale* $(0, 1)$;
- *Gamma* (α, β) ;
- *Fréchet minima* (α) : $F(x) = 1 - \exp(-(-x)^{-\alpha})$, $x < 0$.

Chapitre 2

Copules de valeurs extrêmes : Cas bivarié

Introduction

Imaginons que nous disposons d'un vecteur aléatoire bivarié $X = (X_1, X_2)$ de fonction de répartition conjointe F . Soit F_1 et F_2 les fonctions de répartition marginales de (X_1, X_2) .

Soit $\{(X_{1,i}, X_{2,i})\}_{1 \leq i \leq n}$ un échantillon de X et posons $M_n = (\max_{1 \leq i \leq n} X_{i,1}, \max_{1 \leq i \leq n} X_{i,2}) = (M_{n,1}, M_{n,2})$. Supposons que $F_1 \in D(G_{\xi_1})$ et $F_2 \in D(G_{\xi_2})$ tel que $(M_{n,1} - b_{n,1})/a_{n,1}$ converge en loi vers G_{ξ_1} et $(M_{n,2} - b_{n,2})/a_{n,2}$ converge en loi vers G_{ξ_2} avec $a_{n,1} > 0$, $b_{n,1}$, $a_{n,2} > 0$ et $b_{n,2}$ des suites. Peut-on trouver une loi limite pour M_n normalisé?

Le vecteur M_n est bivarié, donc il va falloir définir un domaine d'attraction bivarié.

Définition 2.1 (*Domaine d'attraction bivarié*)

On dit que F appartient au domaine d'attraction bivarié G où G est la fonction de répartition d'une loi non dégénérée (notée $F \in D_2(G)$) s'il existe $a_n = (a_{n,1}, a_{n,2}) > 0$ avec $a_{n,1}, a_{n,2} > 0$ et $b_n = (b_{n,1}, b_{n,2})$ tel que : $((M_{n,1} - b_{n,1})/a_{n,1}, (M_{n,2} - b_{n,2})/a_{n,2})$ converge

vers G en loi.

Si $F \in D_2(G)$ alors il est clair que G_{ξ_1} et G_{ξ_2} sont les fonctions de répartition marginales de G . On connaît déjà la forme de G_{ξ_1} et G_{ξ_2} . Quelle est la forme de G ? Il manque un élément important pour lier G_{ξ_1} et G_{ξ_2} : la **copule**. La copule est une fonction à deux variables qui lie une fonction de répartition bivariée à ses fonctions de répartition marginales. La copule est une fonction de la fonction de répartition bivariée. La copule d'un domaine d'attraction bivariée a une forme particulière qu'on appelle copule de valeurs extrêmes. Une fois qu'on connaît la forme de la copule associée à G alors on connaît la forme de G . Dans ce chapitre, nous allons d'abord présenter la notion de copule, ensuite le cas particulier des copules de valeurs extrêmes et enfin certaines méthodes d'estimation des copules de valeurs extrêmes. Les copules mesurent la dépendance entre les variables. De même, plusieurs mesures de dépendances s'obtiennent à partir de la copule : le tau de Kendall, la corrélation de Spearman etc. Dans notre étude, on s'intéresse au cas bivarié, donc l'ensemble des définitions et exemples seront dans ce cadre.

2.1 Les copules

Pour caractériser la dépendance entre deux variables aléatoires, l'indicateur le plus communément utilisé est le coefficient de corrélation de Pearson :

$$\rho(X_1, X_2) = (\text{cov}(X_1, X_2)) / \sqrt{\text{Var}(X_1)\text{Var}(X_2)}.$$

Le coefficient de corrélation de Pearson est simple à interpréter car unidimensionnel et compris entre -1 et 1. Cependant, il ne décelle pas certaines formes de dépendances. Pour illustrer nos propos nous allons effectuer une expérience répétée à quatre reprises dans des conditions différentes (inspiré de Guillote et Guillote and Perron (2008)).

Dans l'expérience, nous disposons toujours d'un vecteur aléatoire bivarié $X = (X_1, X_2)$ dont la fonction de répartition est F et dans chaque sous-expérience nous simulons x sous des hypothèses différentes.

Sous-expérience 1 : $X_1, X_2 \sim \mathcal{N}(0, 9)$

X_1 et X_2 sont indépendantes

$$\rho(X_1, X_2) = 0,$$

$$F(x) = F_1(x_1)F_2(x_2) = \phi(x_1/3)\phi(x_2/3).$$

Sous-expérience 2 : $X_1 \sim \mathcal{N}(0, 9)$ et $X_2 = X_1 + \epsilon$ où $\epsilon \sim \mathcal{N}(0, 1)$ indépendante de X_1

$$\Rightarrow X \sim \mathcal{N}_2(O_2, \Sigma_2) \text{ avec } O_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ et } \Sigma_2 = \begin{pmatrix} 9 & 9 \\ 9 & 10 \end{pmatrix}$$

$$\Rightarrow \rho(X_1, X_2) = \frac{3}{\sqrt{10}} \approx 94,87\%$$

$$F(x) = \phi_2(\phi^{-1}(F_1(x_1)), \phi^{-1}(F_2(x_2)), \rho(X_1, X_2)) = \phi_2(\phi^{-1}(\phi(x_1/3)), \phi^{-1}(\phi(x_2/3)), \rho(X_1, X_2))$$

$$\text{Où } \phi_2(h, k, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^h \int_{-\infty}^k \exp\left\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\} dydx$$

Sous-expérience 3 : $X_1 \sim \mathcal{N}(0, 4)$ et $X_2 = X_1^2 + \epsilon$ où $\epsilon \sim \mathcal{N}(0, 1)$ indépendante de X_1

$$\Rightarrow \rho(X_1, X_2) = 0$$

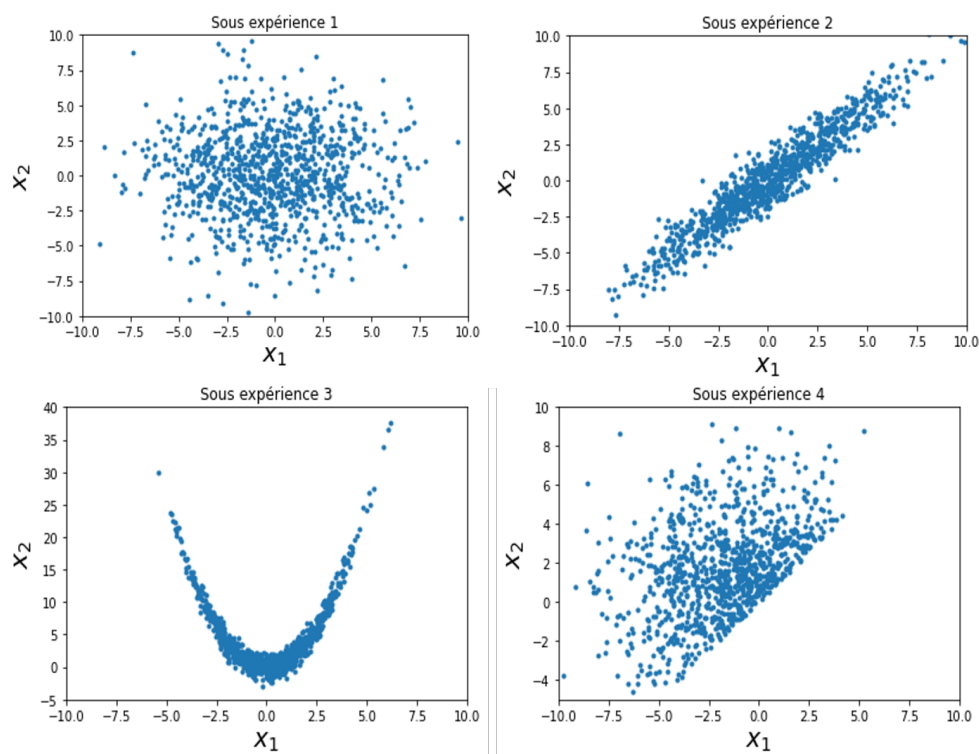
Sous-expérience 4 : $Z_1, Z_2 \sim \mathcal{N}(0, 9)$ et Z_1 et Z_2 sont indépendantes. On pose :

$$X_1 = \min \{Z_1, Z_2\} \text{ et } X_2 = \max \{Z_1, Z_2\}.$$

On a :

$$F(x) = 2 \min \{F_1(x_1), F_2(x_2)\} F_2(x_2) - \min \{F_1(x_1), F_2(x_2)\}^2.$$

Figure 2.1 – Résultats de la simulation des 4 sous-expériences



Dans la sous-expérience 1 et 2, la condition sur la normalité et la dépendance linéaire est satisfaite alors le coefficient de corrélation caractérise entièrement la relation entre X_1 et X_2 . La nullité du coefficient de corrélation à la sous-expérience 1 signifie que X_1 et X_2 sont indépendantes. Néanmoins, dans la sous-expérience 3 où la condition sur la normalité

et la dépendance linéaire n'est pas satisfaite, le coefficient de corrélation est nul alors que les deux variables aléatoires ne sont pas indépendantes. Cet exemple montre les limites du coefficient de corrélation et de la matrice de corrélation en général quand la condition sur la normalité et la dépendance linéaire n'est pas satisfaite. Les copules bivariées caractérisent la dépendance entre deux variables aléatoires dans un cadre général. La sous-expérience 4 permet d'introduire cette notion. Le plus important c'est d'observer la forme de la fonction de répartition conjointe F . On constate que $F(x_1, x_2)$ dépend des distributions marginales $F_1(x_1)$ et $F_2(x_2)$. Si on pose, $u = F_1(x_1)$ et $v = F_2(x_2)$ on a :

$$F(x_1, x_2) = 2 \min \{u, v\} v - \min \{u, v\}^2.$$

Posons $C(u, v) = 2 \min \{u, v\} v - \min \{u, v\}^2$.

La fonction C est la copule bivariée associée au vecteur aléatoire (x_1, x_2) . Selon le théorème de Sklar (1959), elle existe pour tout vecteur aléatoire bivarié.

Définition 2.2 Une copule (bivariée) est une fonction C définie sur le carré $S = [0, 1] \times [0, 1]$ satisfaisant :

- Les conditions de bords :

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

- La monotonie :

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

Avant de présenter le théorème de Sklar (1959), il faut définir la fonction de répartition inverse.

Définition 2.3 Soit $H : [-\infty, +\infty] \mapsto [0, 1]$ une fonction de répartition. La fonction de répartition inverse généralisée $H^{-1} : [0, 1] \mapsto [-\infty, +\infty]$ est définie par :

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

Cette fonction possède notamment les propriétés suivantes :

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

Théorème 2.1 (Sklar)

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 :

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

Pour la démonstration voir Carley and Taylor (2002).

Corollaire 2.1

La détermination de C est unique sur $\text{Im}(F_1) \times \text{Im}(F_2)$. Elle est définie par :

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

Corollaire 2.2

Si F_1 et F_2 sont continues alors la copule associée à F est unique.

¹ $\text{Im}(H)$ est l'image de H .

Une autre limite du coefficient de corrélation est qu'il n'est pas invariant par rapport aux transformations strictement croissantes non linéaires des marges. Cela signifie qu'un simple changement d'échelle (logarithmique par exemple) peut modifier le coefficient de corrélation. Le théorème suivant montre que la copule ne comporte pas cette limite dans le cas continu.

Théorème 2.2

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 la copule associée au vecteur aléatoire $(\phi_1(X_1), \phi_2(X_2))$.

Le théorème de Sklar est très puissant car il montre que pour tout vecteur aléatoire (X_1, X_2) il existe une copule C qui caractérise la dépendance entre X_1 et X_2 . Une question importante est comment la copule caractérise la dépendance.

La popularité du coefficient corrélation de Pearson réside dans sa facilité à être interprété puisqu'il est toujours compris entre -1 et 1 qui sont des cas limites. Les copules sont aussi uniformément bornées par les frontières de Fréchet-Hoeffding (voir Nelsen (2007) ou Lux et al. (2017)) :

Définition 2.4 *La copule C^+ donnée par : $C^+(u, v) = \min\{u, v\}$, pour tout $u, v \in S$ est appelée copule comonotone.*

La copule C^- donné par : $C^-(u, v) = \max\{0, u + v - 1\}$, pour tout $u, v \in S$ est appelée copule contre-monotone.

Lemme 2.1

Pour toute copule C , on a :

$$C^-(u, v) \leq C(u, v) \leq C^+(u, v), \text{ pour tout } u, v \in S.$$

NB : Les résultats suivants ne sont valables que dans le cas continu.

Définition 2.5 Soient X_1 et X_2 deux v.a. On a :

- Si (X_1, X_2) admet comme copule C^+ alors X_1 et X_2 sont comonotones.
- Si (X_1, X_2) admet comme copule C^- alors X_1 et X_2 sont contre-monotones.

Définition 2.6 Soient X_1 et X_2 deux v.a. continues, on a :

- Si X_1 et X_2 ont une dépendance positive alors

$$\mathbb{P}(X_1 > a_1, X_2 > a_2) \geq \mathbb{P}(X_1 > a_1) \mathbb{P}(X_2 > a_2), \text{ pour tout } (a_1, a_2) \in \mathbb{R}^2.$$

- Si X_1 et X_2 ont une dépendance négative alors

$$\mathbb{P}(X_1 > a_1, X_2 > a_2) \leq \mathbb{P}(X_1 > a_1) \mathbb{P}(X_2 > a_2), \text{ pour tout } (a_1, a_2) \in \mathbb{R}^2.$$

En d'autres termes, si X_1 et X_2 ont une dépendance positive alors elles deviennent souvent plus grandes ensembles ou plus petites ensembles. Nous avons l'inverse quand X_1 et X_2 ont une dépendance négative.

Lemme 2.2

Soient X_1 et X_2 deux variables aléatoires de fonction de répartition respectives F_1 et F_2 .

Les v.a X_1 et X_2 sont comonotones si et seulement si $X_2 = F_2^{-1}(F_1(X_1))$ presque sûrement.

NB : $X_2 = F_2^{-1}(F_1(X_1)) \Leftrightarrow X_1 = F_1^{-1}(F_2(X_2)).$

Deux v.a. comonotones ont une dépendance positive. En plus, on dit que la dépendance positive entre deux v.a. comonotones est parfaite. Mais, cela ne signifie pas forcément que le coefficient de corrélation entre deux variables comonotones est égal à 1.

Exemple 2.1 *Supposons que X_1 et X_2 sont comonotones :*

- Si $X_1, X_2 \sim \text{exp}(1)$ alors $X_2 = F_2^{-1}(F_1(X_1)) = X_1.$
- Si $X_1 \sim \text{Pareto}(\alpha, 1)$ et $X_2 \sim \text{exp}(1)$ alors $X_2 = F_2^{-1}(F_1(X_1)) = \alpha \log(1 + X_1).$

Dans le deuxième exemple, le coefficient corrélation entre X_1 et X_2 est différent de 1.

Lemme 2.3

Soient X_1 et X_2 deux variables aléatoires de fonction de répartition respectives F_1 et $F_2.$

Les v.a. X_1 et X_2 sont contre-monotones si $X_2 = F_2^{-1}(1 - F_1(X_1))$ presque sûrement.

NB : $X_2 = F_2^{-1}(1 - F_1(X_1)) \Leftrightarrow X_1 = F_1^{-1}(1 - F_2(X_2)).$

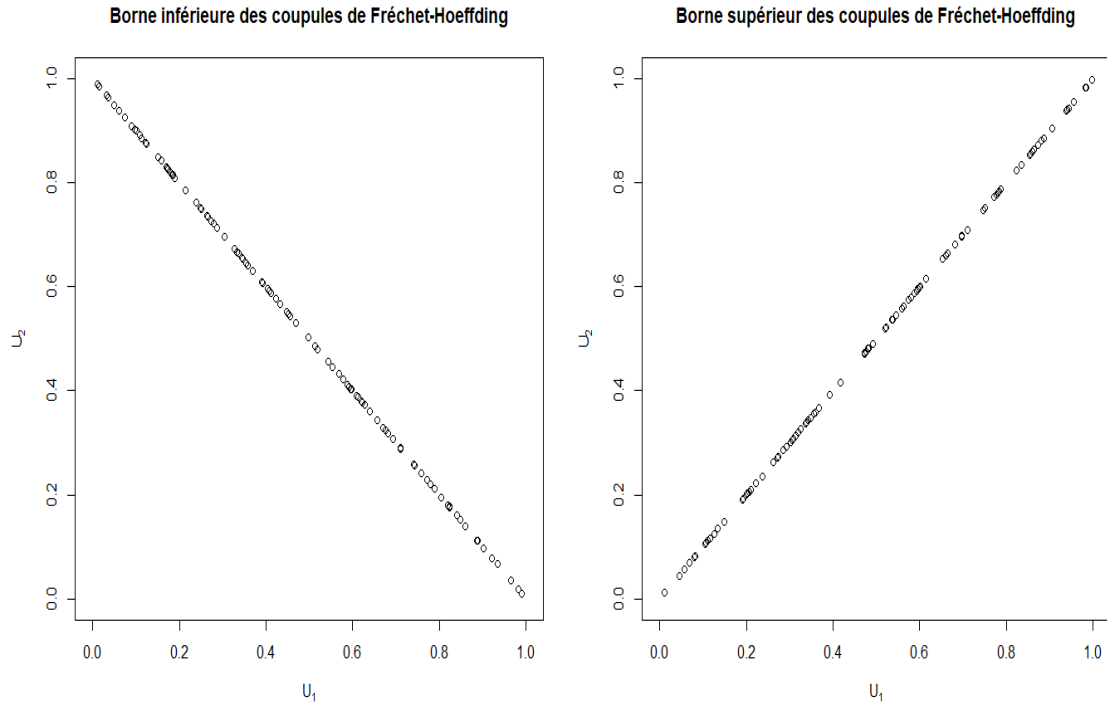
Deux v.a. contre-monotones ont une dépendance négatives. En plus, on dit que la dépendance négatives entre deux v.a. contre-monotones est parfaite. Mais, cela ne signifie pas forcément que le coefficient de corrélation entre deux variables comonotone est égal à -1.

Exemple 2.2 *Supposons que X_1 et X_2 sont contre-monotones :*

Si $X_1, X_2 \sim \text{exp}(1)$ alors $X_2 = F_2^{-1}(1 - F_1(X_1)) = -\log [1 - e^{-X_1}]$

La corrélation entre X_1 et X_2 est différente de -1.

Figure 2.2 – Simulation bornes des copules de Fréchet-Hoeffding



Au-delà des frontières de Fréchet-Hoeffding il y a la copule d'indépendance notée $C^\perp(u, v) = uv$. Les v.a. X_1 et X_2 sont indépendantes si et seulement si le vecteur aléatoire (X_1, X_2) admet C^\perp comme copule.

Les copules comonotone, contre-monotones et indépendantes sont des cas spéciaux. On peut interpréter une copule en vérifiant sa proximité à un de ces cas spéciaux. D'autre part, il existe des mesures de dépendance basées sur les copules comme le tau de Kendall, le rho de Spearman etc. Pour plus de détails sur ces mesures de dépendance se référer à Rodriguez (2007).

2.2 Copules de valeurs extrêmes

2.2.1 Principe

Considérons les vecteurs aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$ i.i.d de fonctions de répartition F associée à la copule C et $M_n = (M_n^{(X)}, M_n^{(Y)}) = (\max_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} Y_i)$. Supposons que $F \in D_2(G)$ avec les suites $(a_n^{(X)}, a_n^{(Y)}) > 0$ et $(b_n^{(X)}, b_n^{(Y)})$. Une copule est de valeur extrême si c'est la copule d'une loi de valeur extrêmes bivariée. Soit \tilde{C} la copule de valeurs extrêmes associée à G . Comme nous l'avons expliqué à l'introduction, on connaît la forme des marginales de G qui sont celles des distributions de valeurs extrêmes et nous cherchons à connaître la forme de \tilde{C} .

Lemme 2.4

$$\tilde{C}(u, v) = \lim_{n \rightarrow \infty} C(u^{1/n}, v^{1/n})^n, \text{ pour tout } (u, v) \in S \quad (2.1)$$

Démonstration :

Posons $M_n^* = \left((M_n^{(X)} - b_n^{(X)}) / a_n^{(X)}, (M_n^{(Y)} - b_n^{(Y)}) / a_n^{(Y)} \right)$, $M_n^* \xrightarrow[n \rightarrow \infty]{\text{loi}} G$.

Soit $F_{M_n^*}$ la fonction de répartition de M_n^* et $C_{M_n^*}$ la copule associée.

Il est clair que $\lim_{n \rightarrow \infty} F_{M_n^*}(x, y) = G(x, y)$, pour tout $(x, y) \in \mathbb{R}^2$

Or $C_{M_n^*}(u, v) = F_{M_n^*} \left(F_{M_n^*(X)}^{-1}(u), F_{M_n^*(Y)}^{-1}(v) \right)$, pour tout $(u, v) \in S$.

Alors, $\lim_{n \rightarrow \infty} C_{M_n^*}(u, v) = \lim_{n \rightarrow \infty} F_{M_n^*} \left(F_{M_n^*(X)}^{-1}(u), F_{M_n^*(Y)}^{-1}(v) \right) = G(G_X^{-1}(u), G_Y^{-1}(v)) = \tilde{C}(u, v)$

(a)

Maintenant prouvons que $C_{M_n^*}(u, v) = C(u^{1/n}, v^{1/n})^n$.

On sait que $F_{M_n^*} = F^n$ et $F(x, y) = C(F_X(x), F_Y(y))$

$$\Rightarrow C_{M_n^*}(u, v) = F(F_{M_n^*(X)}^{-1}(u), F_{M_n^*(Y)}^{-1}(v))^n = C\left(F_X(F_{M_n^*(X)}^{-1}(u)), F_Y(F_{M_n^*(Y)}^{-1}(v))\right)^n$$

Or $F_{M_n^*(X)}^{-1}(u) = F_X^{-1}(u^{1/n})$ et $F_{M_n^*(Y)}^{-1}(v) = F_Y^{-1}(v^{1/n})$

$$\Rightarrow C_{M_n^*}(u, v) = C(F_X(F_X^{-1}(u^{1/n}))^n, F_Y(F_Y^{-1}(v^{1/n}))^n)$$

Donc, $C_{M_n^*}(u, v) = C(u^{1/n}, v^{1/n})^n$ (b)

Finalement (a) et (b) implique que $\tilde{C}(u, v) = \lim_{n \rightarrow \infty} C(u^{1/n}, v^{1/n})^n$, pour tout $(u, v) \in S$.

□

Le lemme 2.4 montre qu'une copule est de valeurs extrêmes si et seulement si elle vérifie l'équation 2.1. En outre, les copules de valeurs extrêmes sont [max-stables](#).

Définition 2.7 Une copule bivariée \bar{C} est max-stable si :

$$\bar{C}(u^{1/n}, v^{1/n})^n = \bar{C}(u, v), \text{ pour tout } n \in \mathbb{N}^* \quad (2.2)$$

Il est clair que si \bar{C} est une copule max-stable alors $\bar{C} \in D_2(\bar{C})$. Alors, toute copule max-stable est de valeurs extrêmes. Le théorème suivant nous montre que la réciproque est vraie.

Théorème 2.3 Une copule bivariée est max-stable si et seulement si elle est de valeurs

extrêmes.

Démonstration : La démonstration de ce théorème s'effectue de la même manière que le théorème 1.1.

Grâce au lemme 2.4 et au théorème 2.3, Pickands (1981) et Deheuvels (1984) ont déterminé la forme d'une copule de valeurs extrêmes.

Théorème 2.4 (Pickands (1981) et Deheuvels (1984))

Toute copule de valeurs extrêmes \tilde{C} est de la forme :

$$\tilde{C}(u, v) = \exp \left\{ \log(uv) A \left(\frac{\log(u)}{\log(uv)} \right) \right\}, \text{ pour tout } u, v \in (0, 1),$$

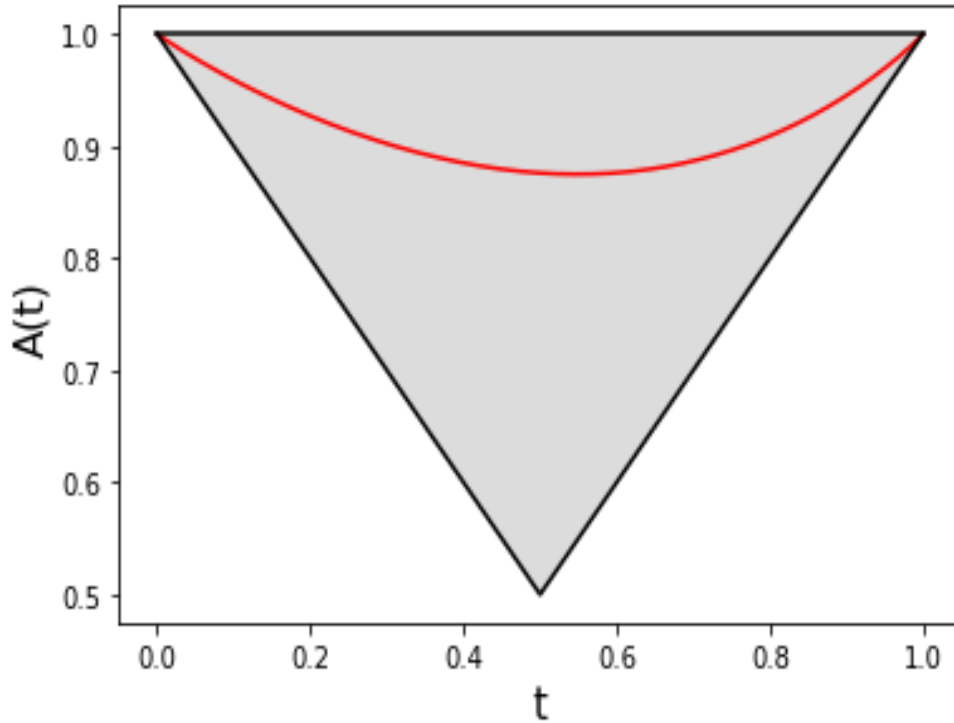
où A est une fonction convexe définie sur $[0, 1]$ et satisfaisant :

$$t \vee (1 - t) \leq A(t) \leq 1, \quad 0 \leq t \leq 1$$

On peut en déduire que $A(0) = A(1) = 1$.

La fonction A est souvent appelée fonction de Pickands. Elle est en réalité l'élément qui identifie toute copule de valeurs extrêmes. Pour toutes copules de valeurs extrêmes \tilde{C}_1 et \tilde{C}_2 de fonctions de Pickands respectives A_1 et A_2 , on a : $\tilde{C}_1 = \tilde{C}_2$ si et seulement si $A_1 = A_2$.

Figure 2.3 – Exemple de fonctions de Pickands.



L'ensemble des fonctions de Pickands possibles se trouve dans la partie grise. La courbe rouge est celle d'une fonction de Pickands intermédiaire. Les bars noirs montrent les bornes des fonctions de Pickands : $t \vee (1 - t) \leq A(t) \leq 1$.

La borne supérieure des fonctions de Pickands ($t \rightarrow 1$) est associée à la copule d'indépendance ($C(u, v) = uv$). La borne inférieure des fonctions de Pickands ($t \mapsto t \vee (1 - t)$) est associée à la copule comonotone ($C(u, v) = \min\{u, v\}$). En rappel, la copule comonotone caractérise la dépendance positive parfaite. En plus, les copules de valeurs extrêmes ne caractérisent que la dépendance positive parce que toute copule de valeurs extrêmes \tilde{C} vérifie : $\tilde{C}(u, v) \geq uv$ (car $A(t) \leq 1$).

2.2.2 Familles paramétriques

La classe des copules de valeurs extrêmes bivariées est de dimension infinie mais il existe plusieurs familles paramétriques spécifiées par la forme de leur fonction de Pickands.

Copules logistiques (ou de Gumbel-Hougaard)

L'ensemble des paramètres de cette famille est $\Theta = \{\theta \in \mathbb{R} : \theta \geq 1\}$. La famille est définie par l'ensemble des copules de la forme :

$$\tilde{C}_\theta(u, v) = \begin{cases} \exp \left\{ - \left[(-\log(u))^\theta + (-\log(v))^\theta \right]^{1/\theta} \right\} & \text{si } 1 \leq \theta < \infty, \\ \max \{u, v\} & \text{si } \theta = \infty. \end{cases}$$

La fonction de Pickands de cette famille est de la forme :

$$A(t) = \begin{cases} t [1 + (1 - t)^\theta]^{1/\theta} & \text{si } 1 \leq \theta < \infty, \\ \max \{t, 1 - t\} & \text{si } \theta = \infty. \end{cases}$$

Définissons les copules archimédiennes par :

$$C_\varphi(u, v) = \varphi^{\leftarrow}(\varphi(u) + \varphi(v)),$$

où $\varphi : [0, 1] \rightarrow [0, \infty[$ et $\varphi^{\leftarrow}(t) = \inf \{x \in [0, 1] : \varphi(x) \leq t\}$; la fonction φ est strictement décroissante, convexe et $\varphi(0) = 1$ voir Mc Neil and Nešlehová (2007).

En fait, toute copule archimédienne appartient au domaine d'attraction d'une copule logistique, plus précisément $C_\varphi \in \mathcal{D}_2(\tilde{C}_\theta)$ avec :

$$\theta = \lim_{s \downarrow 0} s\varphi'(1 - s)/\varphi(1 - s).$$

Copule d'Husler-Reiss

Les copules de cette famille sont les domaines d'attraction des copules gaussiennes (associées aux vecteurs gaussiens). La fonction de Pickands de ces copules sont de la forme :

$$A(t) = \begin{cases} \max\{t, 1-t\} & \text{si } \lambda = 0, \\ (1-t)\varphi\left(\lambda + \frac{1}{2\lambda} \log \frac{1-t}{t}\right) + t\varphi\left(\lambda + \frac{1}{2\lambda} \log \frac{t}{1-t}\right) & \text{si } 0 < \lambda < \infty, \\ 1 & \text{si } \lambda = \infty. \end{cases}$$

Où φ est fonction de répartition de la loi normal centrée réduite.

2.2.3 Estimation

Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un échantillon i.i.d. du vecteur aléatoire (X, Y) . Supposons que (X, Y) suit une distribution de fonction de répartition F dont la copule associée C est une copule de valeurs extrêmes. L'objet de ce document est de trouver une nouvelle méthode d'estimation de C . Avant de proposer notre méthode (au chapitre 4), nous allons présenter les méthodes populairement utilisées dans le domaine.

Il existe deux approches pour estimer les copules de valeurs extrêmes : l'approche paramétrique et l'approche non paramétrique.

Approche paramétrique

Dans cette approche, on suppose que la copule de valeurs extrêmes C appartient à une famille paramétrique $\{C_\theta : \theta \in \Theta\}$. Soit c_θ la densité de la copule C_θ :

$$c_\theta(u, v) = \frac{\partial^2}{\partial x \partial y} C_\theta(x, y) \Big|_{(x,y)=(u,v)}.$$

L'objet de l'approche paramétrique est d'estimer θ . La méthode la plus communément utilisée est l'estimateur du maximum de vraisemblance. Avant d'expliciter la fonction de vraisemblance, il faut gérer la distribution des marges (F_X et F_Y). Rappelons-nous que nous disposons seulement des données $(x_1, y_1), \dots, (x_n, y_n)$ et que la copule est une fonction des marges (u, v) avec $u = F_X(x)$ et $v = F_Y(y)$. Si F_X et F_Y sont connues, alors on dispose directement de l'échantillon $(u_1, v_1), \dots, (u_n, v_n)$, on peut donc utiliser la méthode de l'estimateur du maximum de vraisemblance (MEMV) aisément. Si F_X et F_Y sont inconnues, alors on peut soit les estimer avant d'appliquer la MEMV soit les intégrer dans la MEMV.

- **F_X et F_Y connues**

La fonction de vraisemblance pour l'échantillon $e_n = (x_1, y_1), \dots, (x_n, y_n)$ est donc définie par :

$$L(\theta; e_n) = \prod_{i=1}^n c_\theta(F_X(x_i), F_Y(y_i))$$

et $l(\theta; e_n) = \log(L(\theta; e_n)) = \sum_{i=1}^n \log(c_\theta(F_X(x_i), F_Y(y_i)))$

- **F_X et F_Y estimés au préalable**

Soient \hat{F}_X et \hat{F}_Y les estimations de F_X et F_Y . La fonction de vraisemblance pour l'échantillon $e_n = (x_1, y_1), \dots, (x_n, y_n)$ est donc définie par :

$$L(\theta; e_n) = \prod_{i=1}^n c_\theta(\hat{F}_X(x_i), \hat{F}_Y(y_i))$$

et $l(\theta; e_n) = \log(L(\theta; e_n)) = \sum_{i=1}^n \log(c_\theta(\hat{F}_X(x_i), \hat{F}_Y(y_i)))$

La manière la plus simple d'estimer F_X et F_Y est d'utiliser la fonction de répartition empirique : $\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq x\}}$ et $\hat{F}_Y(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{y_i \leq y\}}$

On peut estimer F_X et F_Y de manière paramétrique.

- F_X et F_Y intégrées dans la MEMV

À la différence de ce qui précède, ici on n'estime pas F_X et F_Y avant d'estimer la copule mais en même temps. Il s'agit de supposer que F_X et F_Y appartiennent respectivement aux familles paramétriques $\{F_X^{(\alpha)} : \alpha \in A\}$ et $\{F_Y^{(\beta)} : \beta \in B\}$. Notre modèle paramétrique ne dépend plus seulement de θ mais aussi α et β . Ainsi, la fonction de vraisemblance pour l'échantillon $e_n = (x_1, y_1), \dots, (x_n, y_n)$ est définie par :

$$L(\theta, \alpha, \beta; e_n) = \prod_{i=1}^n c_\theta \left(F_X^{(\alpha)}(x_i), F_Y^{(\beta)}(y_i) \right)$$

$$\text{et } l(\theta; e_n) = \log(L(\theta; e_n)) = \sum_{i=1}^n \log(c_\theta(F_X^{(\alpha)}(x_i), F_Y^{(\beta)}(y_i)))$$

Après avoir bien défini le modèle paramétrique, l'estimation du modèle reste un problème d'optimisation mais pas statistique. Il existe en effet plusieurs méthodes d'optimisation numérique qui permettent de résoudre le problème (pour plus de détail voir Kochenderfer (2019)).

Approche non-paramétrique

Supposons que nous disposons déjà des marges $(u_1, v_1), \dots, (u_n, v_n)$. Les méthodes non paramétriques permettent l'estimation de la fonction de Pickands A .

Posons $\xi(t) = \min \left\{ \frac{-\log(U)}{t}, \frac{-\log(V)}{1-t} \right\}$ où $0 \leq t \leq 1$ où (U, V) est le vecteur aléatoire dont la fonction de répartition est une copule C de fonction de Pickands A , donc $\xi(t)$ est une variable aléatoire. Par convention, $\xi(0) = \xi(1) = 1$ presque sûrement.

$$\text{En fait, } \mathbb{P}(\xi(t) > x) = \mathbb{P}(U < e^{-(1-t)x}, V < e^{-tx}) = C(e^{-(1-t)x}, e^{-tx}),$$

par le théorème 2.4, $\mathbb{P}(\xi(t) > x) = e^{-xA(t)}$ pour tout $x \in \mathbb{R}^+$.

En supposant que la v.a. $\xi(t)$ admet un espérance on a :

$$\int_0^\infty \mathbb{P}(\xi(t) > x) dx = \int_0^\infty e^{-xA(t)} dx.$$

Par suite, $\mathbb{E}[\xi(t)] = \frac{1}{A(t)}$.

Le premier estimateur de A est proposé par Pickands (1981), notons le $A_n^p(t)$. Son estimateur est basé sur les réalisations de ξ , c'est-à-dire $\widehat{\xi}_i(t) = \min \left\{ \frac{-\log(u_i)}{t}, \frac{-\log(v_i)}{1-t} \right\}$ et défini par :

$$\frac{1}{A_n^p(t)} = \frac{1}{n} \sum_{i=1}^n \widehat{\xi}_i(t).$$

L'inconvénient de cet estimateur est qu'il ne vérifie pas les contraintes imposées sur la famille des fonctions de Pickands ($t \vee (1-t) \leq A(t) \leq 1$ et $A(0) = A(1) = 1$). Mais, il renferme une importante propriété asymptotique démontré par Deheuvels (1984) :

Deheuvels (1984) a aussi amélioré l'estimateur de Pickands en proposant un estimateur qui vérifie la contrainte $A(0) = A(1) = 1$. Notons cet estimateur $A_n^{DHV}(t)$.

$$\frac{1}{A_n^{DHV}(t)} = \frac{1}{n} \sum_{i=1}^n \widehat{\xi}_i(t) - t \frac{1}{n} \sum_{i=1}^n \widehat{\xi}_i(1) - (1-t) \sum_{i=1}^n \widehat{\xi}_i(0) + 1$$

Toujours dans le but d'améliorer l'estimateur de Pickands, Hall et Hall and Tajvidi (2000) proposent un estimateur qui non seulement vérifie la contrainte sur les bords mais aussi la seconde contrainte ($t \vee (1-t) \leq A(t) \leq 1$). Notons $A_n^{TDJ}(t)$ l'estimateur de Hall et Tadjvidi.

Posons $\bar{\xi}_i(t) = \min \left\{ \frac{r_i}{t}, \frac{s_i}{1-t} \right\}$,

Avec $r_i = \frac{\log(u_i)}{\frac{1}{n} \sum_{i=1}^n \log(u_i)}$ et $s_i = \frac{\log(v_i)}{\frac{1}{n} \sum_{i=1}^n \log(v_i)}$,

$$\frac{1}{A_n^{TDJ}(t)} = \frac{1}{n} \sum_{i=1}^n \bar{\xi}_i(t).$$

Hall et Tadjvidi ont démontré que leur estimateur a la plus petite variance asymptotique parmi ceux présentés précédemment.

Par ailleurs, Capéraà et al. (1997) choisissent une autre approche. Ils définissent d'abord la variable aléatoire $Z = \log(U)/\log(UV)$. Nous avons :

$$P(Z \leq z) = z + z(1 - z) + \frac{A_+(z)}{A(z)}, 0 \leq z < 1,$$

où $A_+(z)$ est la dérivée à droite de A . Ils obtiennent ainsi une équation dont l'inconnu est A . La quantité $P(Z \leq z)$ est inconnu mais peut s'estimer par la fonction de répartition empirique $F_Z(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{z_i \leq z\}}$ où $z_i = \frac{\log(u_i)}{\log(u_i v_i)}$. Donc, en résolvant l'équation différentielle $F_Z(z) = z + z(1 - z) + \frac{A_+(z)}{A(z)}$ $0 \leq z < 1$, ils obtiennent l'estimateur A_n^{CFG} défini par :

$$\log(A_n^{CFG}(t)) = -\frac{1}{n} \sum_{i=1}^n \log(\hat{\xi}_i(t)) - t \sum_{i=1}^n \log(\hat{\xi}_i(1)) - (1 - t) \sum_{i=1}^n \log(\hat{\xi}_i(0)).$$

Capéraà et al. démontrent que cet estimateur est consistant et est asymptotiquement normal. De même, beaucoup de simulations indiquent qu'il est plus performant que l'estimateur de Pickands et ses variantes (Deheuvels, Hall et Tadjvidi). Toutefois, l'estimateur CFG ne vérifie pas toujours toutes les contraintes imposées par la famille des fonctions de Pickands.

Le fait que les estimateurs présentés précédemment (sauf l'estimateur de Tadjvidi) ne vérifient pas toujours toutes ces conditions d'une fonction de Pickands est problématique. C'est pourquoi Segers et Gudendorf (2011) ont élaboré une méthode qui permet de projeter n'importe quel estimateur d'une fonction de Pickands dans un sous-espace de fonctions sur $[0, 1]$ vérifiant toutes les conditions d'une fonction de Pickands.

Cependant, l'estimateur proposé par Gudendorf et Segers (2011) n'améliore pas forcément la qualité de l'estimateur intermédiaire (projeté). En effet, Gudendorf et Segers ont appliqué leur méthode sur des données simulées en projetant l'estimateur de Capéreaa et al. Ils ont aboutit à la conclusion que l'estimateur de Capéreaa et al. performe mieux que l'estimateur sans projection dans le cas où la vraie copule n'est pas très proche de la copule d'indépendance. La méthode que nous proposons ne dépend pas d'un autre estimateur et aboutit à une fonction de Pickands (c'est-à-dire qui vérifie toutes les contraintes imposées par la famille des fonctions de Pickands).

Chapitre 3

Estimation bayésienne et méthodes

MCMC

Introduction

La méthode que nous allons présenter dans le chapitre 4 rentre dans le cadre de l'inférence bayésienne. Dans ce chapitre nous allons premièrement présenter le principe de l'inférence bayésienne.

Cependant, l'évaluation numérique des estimateurs bayésiens se fera avec les méthodes MCMC. Donc, nous allons présenter en deuxième partie de ce chapitre les méthodes MCMC, plus particulièrement l'algorithme de Metropolis-Hastings.

Ainsi, ce chapitre présente les outils mathématiques et statistiques nécessaires pour comprendre notre méthode d'estimation dans le chapitre 4. Il n'est pas indispensable pour le lecteur qui a déjà compris ces méthodes.

3.1 Estimation bayésienne

Supposons qu'on a un modèle paramétrique $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$ où f_θ est la fonction de densité à valeurs dans \mathcal{K} et dépendant du paramètre θ et Θ est l'espace des paramètres.

On dispose d'un échantillon x_1, \dots, x_n i.i.d. généré par la loi de densité f_θ où θ est inconnu.

On cherche à estimer θ .

Dans l'approche classique, on considère que le paramètre θ est fixe et on cherche un élément de Θ qui sera l'estimation de θ . Dans l'approche bayésienne, on considère que le paramètre θ est aussi une v.a. de densité $\pi(\theta)$ et $\pi(\theta)$ est appelée la densité a priori. On définit la fonction de densité conditionnelle $x \mapsto f(x|\theta)$ par $f(x|\theta) = f_\theta(x)$ pour tout $\theta \in \Theta$. On peut en déduire la fonction de vraisemblance $f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$. Donc, on connaît la loi de l'échantillon conditionnellement au paramètre θ mais en statistique bayésienne on s'intéresse à l'inverse appelée loi a posteriori $\theta|x_1, \dots, x_n$. La densité de la loi a posteriori est noté $\pi(\theta|x_1, \dots, x_n)$.

D'après la formule de bayes,

$$\pi(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta)\pi(\theta)}{m(x_1, \dots, x_n)} \text{ où } m(x_1, \dots, x_n) = \int_{\Theta} f(x_1, \dots, x_n|\theta)\pi(\theta)d\theta.$$

Les estimateurs bayésiens sont basés sur la distribution a posteriori. Il en existe plusieurs mais le plus communément utilisé est l'espérance a posteriori défini par :

$$\mathbb{E}[\theta|x_1, \dots, x_n] = \int_{\Theta} \theta\pi(\theta|x_1, \dots, x_n)d\theta$$

Cependant, on ne connaît pas souvent l'espérance a posteriori de manière explicite à cause

de l'intégrale. Les méthodes de Monte Carlo pourrait permettre d'estimer l'espérance. Dans notre cas, les méthodes de Monte Carlo consistent d'abord à générer un échantillon i.i.d. $\theta_1, \dots, \theta_m$ dont la loi est $\pi(\theta|x_1, \dots, x_n)$ et enfin estimer l'espérance avec la moyenne de l'échantillon.

$$\hat{\theta}_m = \frac{1}{m} \sum_{i=1}^m \theta_i,$$

où $\hat{\theta}_m$ est l'estimateur Monte Carlo de $\mathbb{E}[\theta|x_1, \dots, x_n]$ avec un échantillon i.i.d. de taille m . Avec la loi des grands nombre on a :

$$\hat{\theta}_m \xrightarrow{m \rightarrow \infty} \mathbb{E}[\theta|x_1, \dots, x_n]$$

Cependant, il peut être difficile de simuler des réalisations de la loi de $\theta|x_1, \dots, x_n$. C'est pourquoi, on utilise les méthodes MCMC (Monte Carlo par Chaînes de Markov). Les méthodes MCMC construisent une chaîne de Markov irréductible dont la loi stationnaire est $\pi(\theta|x_1, \dots, x_n)$. Il existe plusieurs méthodes MCMC mais dans notre cas nous allons utiliser l'algorithme de Metropolis-Hastings. Bien que l'échantillon issu de l'algorithme de Metropolis-Hastings n'est pas i.i.d., la convergence de l'estimateur vers l'espérance est démontré par Tierney (1994).

3.2 Algorithme de Metropolis-Hastings

Définition 3.1 (*Chaîne de Markov*)

Un processus stochastique $\{X_n\}_{n \in \mathbb{N}}$ à valeurs dans χ muni d'une σ -algèbre $\mathcal{B}(\chi)$, est une

chaîne de Markov si la distribution de X_{n+1} conditionnellement à l'état actuel $X_n = x_n$ est indépendante de (X_0, \dots, X_{n-1}) :

$$P(X_{n+1} \in B | X_i = x_i, i \leq n) = P(X_{n+1} \in B | X_n = x_n), \text{ pour tout } x_0, \dots, x_n \in \chi$$

On dit que la chaîne de Markov est **homogène** si $P(X_{n+1} \in B | X_n = x) = P(X_{n'+1} \in B | X_{n'} = x) \forall n \neq n'$. Dans le cadre des méthodes MCMC on utilise les chaînes de Markov homogènes. L'homogénéité de la chaîne de Markov permet de considérer la fonction K de $\mathcal{B}(\chi) \times \chi$ vers $[0, 1]$ donnée : $K(x, B) = P(X_1 \in B | X_0 = x)$, pour tout $x \in \chi$. La fonction K est appelée **noyau de transition**. Le noyau de transition et la loi initiale caractérise la chaîne de Markov homogène.

Remarques :

- Pour tout $B \in \mathcal{B}(\chi)$, $K(\cdot, B)$ est une fonction mesurable sur χ
- Pour tout $x \in \chi$, $K(x, \cdot)$ est une mesure de probabilité sur $\mathcal{B}(\chi)$
- Pour tout $m \in \mathbb{N}^*, n \in \mathbb{N}, B \in \mathcal{B}(\chi)$ et $x \in \chi$, nous définissons le noyau de transition itéré $K^m(x, B) = P(X_m \in B | X_0 = x)$

Le noyau de transition permet de vérifier si la chaîne est irréductible et apériodique.

Définition 3.2 (*Chaîne irréductible*)

Une chaîne de Markov homogène $\{X_n\}_{n \in \mathbb{N}}$ à valeurs dans χ est irréductible si $\forall x \in \chi, B \in \mathcal{B}(\chi)$ et $B \neq \emptyset$ tel qu'il existe un entier non nul m tel que $K^m(B, x) > 0$.

Définition 3.3 (*Chaîne périodique*)

Soit $\{X_n\}_{n \in \mathbb{N}}$ chaîne de Markov homogène à valeurs dans χ . La chaîne est périodique s'il

existe un entier $m \geq 2$ et une suite S_0, S_1, \dots, S_{m-1} d'éléments de $\mathcal{B}(\chi)$ avec $S_i \cap S_j = \emptyset \forall i \neq j$ et $S_m = S_0$ satisfaisant pour tout $i \in \{1, \dots, m-1\}$:

$$P(X_1 \in S_{i+1}|x) = 1, \text{ pour tout } x \in S_i.$$

Une telle suite d'ensemble est appelée m -cycle. La période d'une chaîne est le plus grand entier m tel qu'un m -cycle existe. Une chaîne est apériodique si sa période $m = 1$.

Pour simuler les réalisations d'une distribution Π , les méthodes MCMC utilisent une chaîne de Markov dont la distribution stationnaire est Π . Soit Π la distribution cible à valeur dans χ munie d'une σ -algèbre $\mathcal{B}(\chi)$ et d'une densité π . Dans notre modèle bayésien cela correspond à $\pi(\theta|)$. La chaîne d'une méthode MCMC doit être irréductible et apériodique. En outre, une propriété importante pour la chaîne d'une méthode MCMC est la condition d'équilibre.

Définition 3.4 Soit K le noyau de transition d'une chaîne de markov $\{X_n\}_{n \in \mathbb{N}}$. Le noyau K vérifie la condition d'équilibre reliée à π si :

$$\int_A \int_B \pi(dx)K(x, dy) = \int_B \int_A \pi(dy)K(y, dx), \text{ pour tout } A \text{ et } B \text{ des éléments de } \mathcal{B}(\chi).$$

Ce qui est équivalent à :

$$K(y, x) \pi(y) = K(x, y) \pi(x), \forall x, y \in \chi \text{ presque partout}$$

Il existe plusieurs méthodes MCMC, la plus célèbre est l'algorithme de Metropolis-Hastings.

Définition 3.5 (Noyau de transition de Metropolis-Hastings)

Soit Π la loi cible à valeur dans χ munie d'une σ -algèbre $B(\chi)$ et d'une densité π et $Q(\cdot, \cdot)$

le noyau de transition d'une chaîne de Markov muni d'une densité $q(\cdot, \cdot)$ appelée **densité instrumentale**. On a,

$$Q(B, x) = \int_B q(y, x) dy, \text{ pour tout } B \in \mathcal{B}(X), x \in X$$

On définit la **probabilité d'acceptation Metropolis-Hastings** par :

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

Algorithme 1 : Metropolis-Hastings

1 Données : Densité cible π , densité instrumentale q , N la taille de l'échantillon et valeur initiale de la chaîne x_0

2 Début

3	Pour chaque $n \in \{1, \dots, N - 1\}$
4	<i>Générer l'observation</i> $y_n \sim q(\cdot, x_n)$;
5	<i>Calculer</i> $\alpha(y_n, x_n)$;
6	<i>Générer</i> $u \sim U(0, 1)$;
7	si $u \leq \alpha(y, x_n)$ alors
8	<i>accepter le candidat</i> y_n (c.à.d $x_{n+1} = y_n$);
9	sinon
10	<i>rejeter le candidat</i> y_n (c.à.d $x_{n+1} = x_n$);

11 Sortie : L'échantillon (x_0, x_1, \dots, x_N)

Le **noyau Metropolis-Hastings** est donc défini par :

$$P(B, x) = \int_B \alpha(y, x)q(y, x)dy + r(x)\mathbb{1}_{\{x \in B\}}$$

Où $r(x)$ est la probabilité que le candidat soit rejeté :

$$r(x) = 1 - \int_{\chi} \alpha(y, x)q(y, x)dy$$

Rappelons qu'il est suffisant que le noyau de transition P vérifie les trois propriétés suivantes : irréductible, apériodique et condition d'équilibre. La condition d'équilibre est garantie pour toute chaîne Metropolis-Hastings. Mais le fait que la chaîne soit apériodique et l'irréductible restent à vérifier.

Proposition 3.1 *Soit P un noyau de Metropolis-Hastings pour une densité cible π à valeurs dans χ et q la densité instrumentale.*

Le noyau de transition P satisfait la condition d'équilibre avec π .

Il existe une condition suffisante pour que la chaîne de Metropolis-Hastings soit irréductible et apériodique.

Proposition 3.2 *Soit P un noyau de Metropolis-Hastings pour une densité π à valeurs dans χ et q la densité instrumentale. Le noyau de transition P satisfait les trois propriétés (irréductible, apériodique et condition d'équilibre) si :*

$$\pi(x) > 0 \text{ et } \pi(y) > 0 \in q(y, x) > 0 \forall x, y \in \chi \tag{3.1}$$

Pour les preuves et les détails concernant les méthodes MCMC, se référer à Robert and Casella (2013).

Chapitre 4

Estimation bayésienne d'une fonction de Pickands par des splines cubiques

Introduction

Soit $(X_1, Y_1), (X_2, Y_2), \dots$ une suite de vecteurs aléatoires i.i.d. dans \mathbb{R}^2 . Notons par $X_{(n)}$ et $Y_{(n)}$ les statistiques pour les maximums sur X_1, \dots, X_n et Y_1, \dots, Y_n respectivement. S'il existe des suites de vecteurs $(a_{1n}, a_{2n}) \in (0, \infty)^2$, $n \geq 1$ et $(b_{1n}, b_{2n}) \in \mathbb{R}^2$, $n \geq 1$ telles que la suite $((X_{(n)} - b_{1n})/a_{1n}, (Y_{(n)} - b_{2n})/a_{2n})$ converge en loi vers une loi non dégénérée dont la fonction de répartition est F alors on dit que la loi de (X_1, Y_1) est dans le domaine d'attraction de la loi associée à la fonction de répartition F pour les lois de valeurs extrêmes. Les lois de valeurs extrêmes pour des vecteurs dans \mathbb{R}^2 ont des formes particulières. Les fonctions de répartitions marginales F_1 et F_2 sont celles de valeurs extrêmes sur \mathbb{R} . La copule qui fait le lien entre F_1 , F_2 et F est une copule de valeurs

extrêmes. Elle prend la forme

$$C(u, v) = \exp \{ \log(uv) A(\log(u)/\log(uv)) \}, \quad u, v \in (0, 1)$$

où A est une fonction appelée la fonction de Pickands. Une fonction $A: [0, 1] \mapsto \mathbb{R}$ est une fonction de Pickands si les conditions suivantes sont satisfaites :

1. La fonction A est convexe sur $[0, 1]$.
2. On a $t \vee (1 - t) \leq A(t) \leq 1$, $t \in [0, 1]$.

Le but de ce chapitre est l'estimation de la fonction A de façon bayésienne en concentrant les lois a priori sur des espaces vectoriels donnés par des splines.

Notons par \mathcal{A} l'espace des fonctions de Pickands. Afin de parvenir à notre objectif on passe par plusieurs étapes. Dans la première étape, on établit une bijection entre \mathcal{A} et \mathcal{G} , l'espace des fonctions de répartition G des variables aléatoires Y à valeurs dans $[0, 1]$ de sorte que $\mathbb{E}[Y] = 1/2$. Dans la deuxième étape, on concentre les lois a priori sur des sous-espaces \mathcal{A}_n de \mathcal{A} de sorte que \mathcal{A}_n puisse être paramétrisé par un paramètre de dimension n . Pour n fixé, on se pose la question, est-ce que \mathcal{A}_n est une bonne approximation de \mathcal{A} ? On essaie de donner une réponse à cette question. Dans la troisième étape, on propose de résoudre numériquement l'estimation de A en utilisant des chaînes de Markov. Comme \mathcal{A}_n est paramétrisé les chaînes évoluent dans l'espace paramétrique $\Theta_n \subset \mathbb{R}^n$ associé à \mathcal{A}_n .

4.1 Principe théorique de la méthode

4.1.1 La bijection

Pour $A \in \mathcal{A}$ on pose $A_+ : [0, 1) \mapsto \mathbb{R}$ la dérivée à droite de A . On introduit une transformation Ψ de sorte que

$$\Psi(A)(x) = \begin{cases} 0 & \text{si } x < 0, \\ (1 + A_+(x))/2 & \text{si } x \in [0, 1), \\ 1 & \text{si } x \geq 1 \end{cases} \quad x \in \mathbb{R}. \quad (4.1)$$

Pour $G \in \mathcal{G}$ on introduit une transformation Φ de sorte que

$$\Phi(G)(t) = (1 - t) + 2 \int_0^t G(x) dx \quad t \in [0, 1]. \quad (4.2)$$

On peut montrer que Ψ est à valeurs dans \mathcal{G} , Φ est à valeurs dans \mathcal{A} . De plus, $\Psi \circ \Phi$ est la fonction identité sur \mathcal{A} et $\Phi \circ \Psi$ est la fonction identité sur \mathcal{G} ce qui établit la bijection entre \mathcal{A} et \mathcal{G} .

Lemme 4.1 *On a les résultats suivants,*

1. *Si G est une fonction de répartition pour une variable aléatoire Y à valeurs dans $[0, 1]$ avec $\mathbb{E}[Y] = 1/2$ alors la fonction $A : [0, 1] \mapsto \mathbb{R}$ donnée par*

$$A(t) = (1 - t) + 2 \int_0^t G(\omega) d\omega, \quad t \in [0, 1] \quad (4.3)$$

est une fonction de Pickands.

2. Si A est une fonction de Pickands alors la fonction G satisfaisant

$$G(t) = \begin{cases} 0 & \text{si } t < 0 \\ (1 + A_+(t))/2 & \text{si } t \in [0, 1), \\ 1 & \text{si } t \geq 1 \end{cases} \quad (4.4)$$

est une fonction de répartition pour une variable aléatoire Y à valeurs dans $[0, 1]$

avec $\mathbb{E}[Y] = 1/2$.

La démonstration du lemme 4.1 se trouve en annexe B.1.

4.1.2 Les sous-espaces

La construction des sous-espaces exploite la fonction Φ donnée dans l'équation (4.2). Pour

$n > 1$ fixé, on pose

$$\Theta_n = \left\{ \theta \in [0, 1]^n : \sum_{k=1}^n \theta_k = n/2 \right\}.$$

Soit $\theta \in \Theta_n$, les éléments de θ_n sont appelés noeuds. Pour $\theta \in \Theta_n$, on note $\theta_{(1)}, \dots, \theta_{(n)}$

les valeurs de θ en ordre croissant. Pour des raisons techniques, on ajoute les noeuds

$$\theta_{(-1)} = \theta_{(0)} = 0,$$

$$\theta_{(n+1)} = \theta_{(n+2)} = 1.$$

Enfin on pose Δ , l'opérateur des différences finies, avec

$$\Delta\theta_{(k)} = \theta_{(k+1)} - \theta_{(k)}, \quad k = -1, \dots, n+1.$$

Pour $\theta \in \Theta_n$ on pose $G_n(1 | \theta) = 1$, $A_n(1 | \theta) = 1$ et

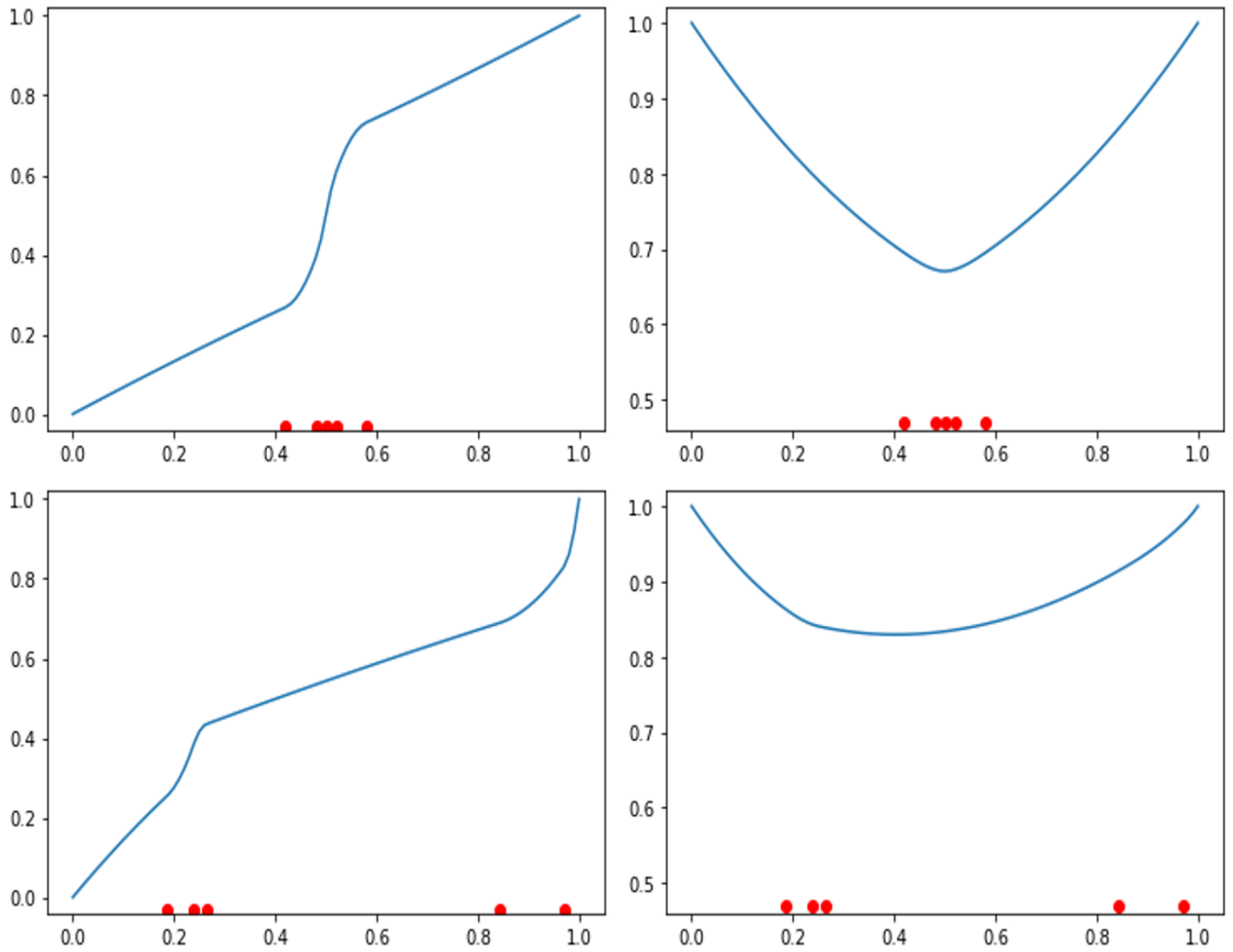
$$\begin{aligned} G_n(x | \theta) &= \frac{1}{n+2} \left[\ell + 1 + \frac{1}{\Delta\theta_{(\ell)}} \left\{ \frac{(x - \theta_{(\ell)})^2}{\Delta\theta_{(\ell)} + \Delta\theta_{(\ell+1)}} - \frac{(\theta_{(\ell+1)} - x)^2}{\Delta\theta_{(\ell-1)} + \Delta\theta_{(\ell)}} \right\} \right] \\ &\quad \text{si } \theta_{(\ell)} \leq x < \theta_{(\ell+1)}, \quad 0 \leq \ell \leq n. \\ A_n(t | \theta) &= (1-t) + 2 \int_0^t G_n(x | \theta) dx \\ &= (1-t) + \frac{2}{n+2} \left[(\ell+1)t - \sum_{k=0}^{\ell} \theta_{(k)} \right. \\ &\quad \left. + \frac{1}{3} \left\{ \frac{(\theta_{(\ell+1)} - t)^3}{\Delta\theta_{(\ell)}(\Delta\theta_{(\ell-1)} + \Delta\theta_{(\ell)})} + \frac{(t - \theta_{(\ell)})^3}{\Delta\theta_{(\ell)}(\Delta\theta_{(\ell)} + \Delta\theta_{(\ell+1)})} - \Delta\theta_{(\ell)} \right\} \right] \\ &\quad \text{si } \theta_{(\ell)} \leq t \leq \theta_{(\ell+1)}, \quad 0 \leq \ell \leq n. \end{aligned}$$

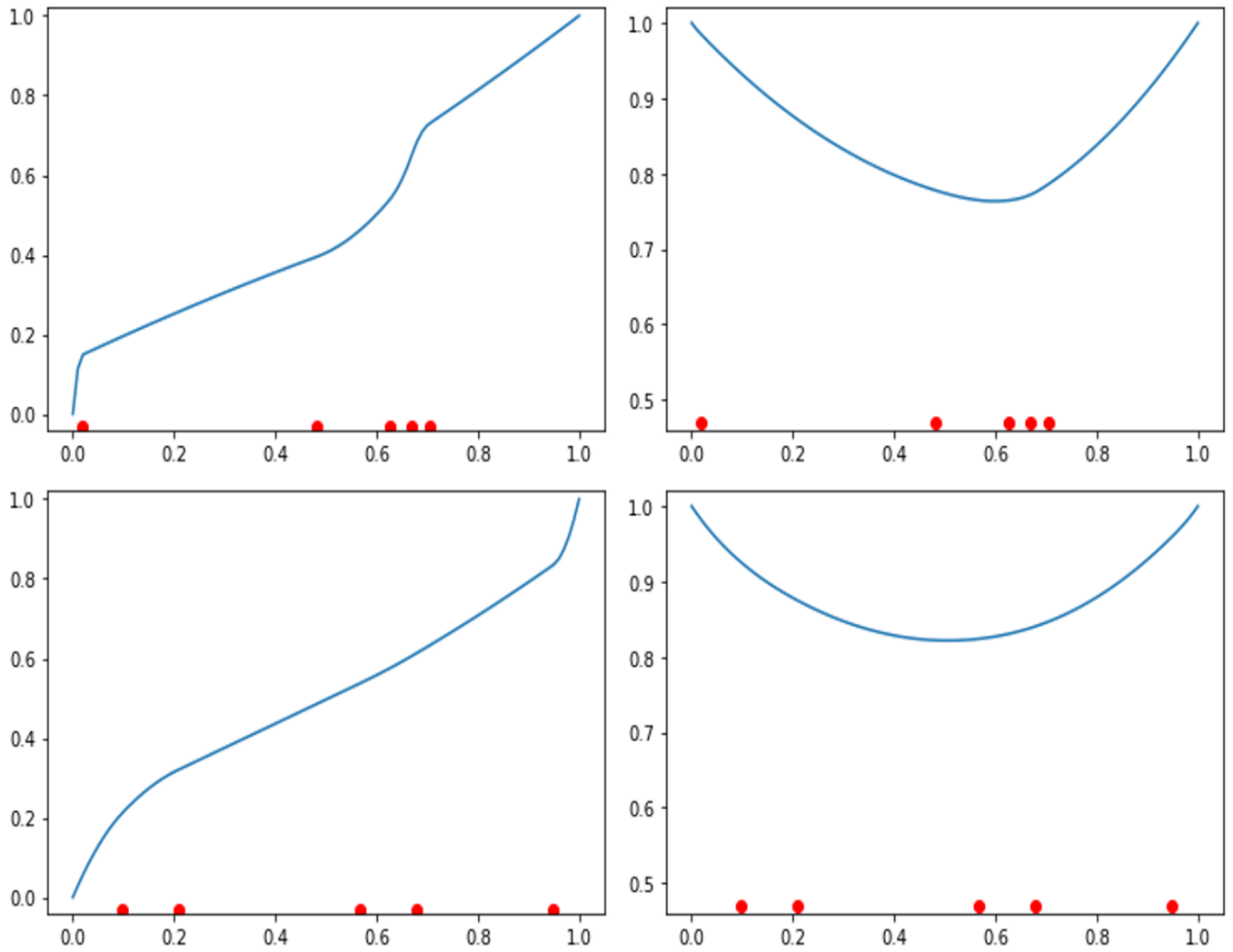
NB : Les fonctions $G_n(\cdot | \theta)$ et $A_n(\cdot | \theta)$ sont fonctions de spline. Plus précisément, la fonction $A_n(\cdot | \theta)$ est une fonction de splines cubiques.

Lemme 4.2 *Soit $\theta \in [0, 1]^n$. Les fonctions $G_n(\cdot | \theta)$ sont des fonctions de répartition pour des variables aléatoires Y sur $[0, 1]$ avec $\mathbb{E}[Y] = 1/2$ si et seulement si $\theta \in \Theta_n$.*

La démonstration du lemme 4.2 se trouve en annexe B.2. Dès que certains noeuds sont de multiplicité plus grande que 2, la fonction $G_n(\cdot | \Theta)$ devient discontinue. Une accumulation de noeuds dans une petit sous-intervalle fait en sorte que la fonction $A_n(\cdot | \theta)$ possède une forte courbure dans ce petit sous-intervalle.

Figure 4.1 – Exemples de fonctions $G_n(\cdot | \theta)$ et $A_n(\cdot | \theta)$ avec 5 noeuds.





Les points rouges sont les noeuds. les graphiques à gauche sont des fonctions de répartition appartenant à l'ensemble \mathcal{G}_5 et celle droite sont les fonctions de Pickands appartenant à l'ensemble \mathcal{A}_5 .

Exemple 4.1 Lorsque n est pair on approche la fonction de Pickands A , pour une copule indépendante, par la fonction $A_n(\cdot | \theta)$ avec des noeuds de multiplicité $n/2$ en 0 et en 1.

Ce qui donne,

$$\begin{aligned}A(t) &= 1, \quad t \in [0, 1] \\A_n(t | \theta) &= 1 - \frac{2}{n+2}t(1-t), \quad t \in [0, 1] \\ \|A_n(\cdot | \theta) - A\|_\infty &= 1/2(n+2)\end{aligned}$$

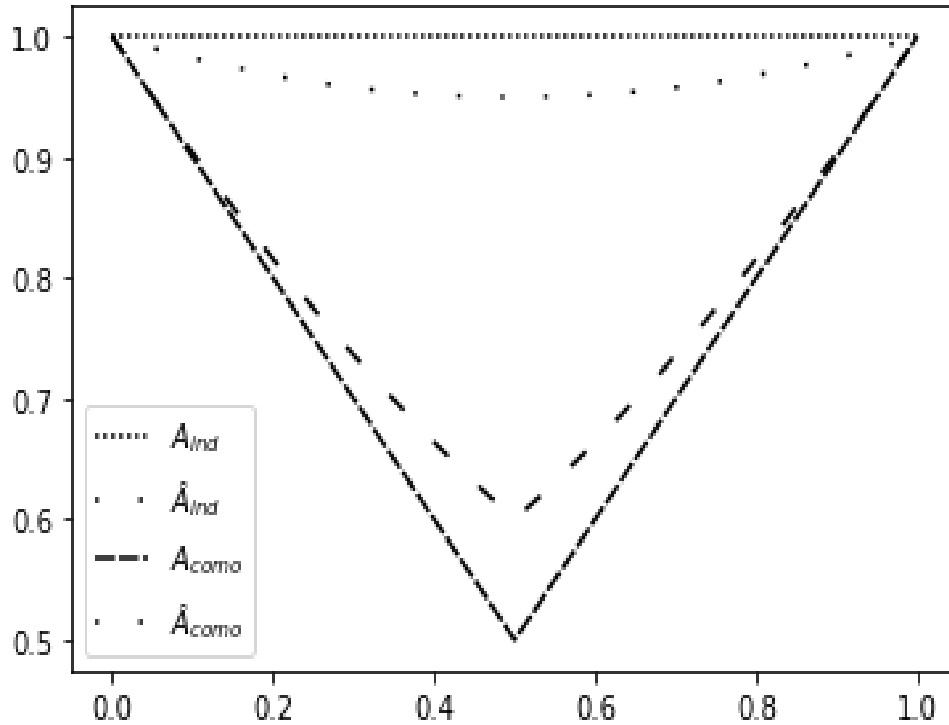
(voir figure 4.2)

Exemple 4.2 Pour $n > 1$ on approche la fonction de Pickands A , pour une copule comotone, par la fonction $A_n(\cdot | \theta)$ avec tous les noeuds en $1/2$. Ce qui donne,

$$\begin{aligned}A(t) &= t \vee (1-t), \quad t \in [0, 1] \\A_n(t | \theta) &= t \vee (1-t) + \frac{4}{n+2}(t \wedge (1-t))^2, \quad t \in [0, 1] \\ \|A_n(\cdot | \theta) - A\|_\infty &= 1/(n+2)\end{aligned}$$

Notez ici, que si $n > 2$ alors la fonction $A_n(\cdot | \theta)$ n'est pas dérivable en $1/2$. (voir figure 4.2)

Figure 4.2 – Approximation de la fonction de pickands indépendante et comonotique avec 8 noeuds.



Lemme 4.3 *Pour tout $A \in \mathcal{A}$, $n > 1$, il existe $\theta \in \Theta_n$ tel que*

$$|A(x) - A_n(x)| \leq \frac{8}{n+2}, \forall x \in [0, 1]$$

Le lemme 4.3 donne une information quantitative pour mesurer la qualité d'approximation du sous-espace \mathcal{A}_n pour approcher \mathcal{A} .

Dans la présentation du modèle bayésien et dans l'application de la méthode sur des données simulées et sur des données appliquées on travaillera directement sur les marges. Plus précisément, soit $(x_1, y_1), \dots, (x_m, y_m)$ les réalisations d'une loi bivariée de valeurs extrêmes dont la fonction de répartition est F . Soit F_X et F_Y les fonctions de répartition marginales

respectives de la première et de la deuxième composante et C la copule de F . En posant, $(u_i, v_i) = (F_X(x_i), F_Y(y_i))$ on obtient l'échantillon des marges $(u_1, v_1), \dots, (u_m, v_m)$. En fait, les couples $(u_1, v_1), \dots, (u_m, v_m)$ sont des réalisations de la copule C qui est aussi une fonction de répartition. Comme on veut estimer des copules, on peut travailler directement sur les données de marge $(u_1, v_1), \dots, (u_m, v_m)$. Il existe des packages du logiciel R qui permettent de simuler directement les réalisations de certaines familles de copule (voir Turlach et al. (2017)).

4.1.3 Le modèle bayésien

Soit $(u_1, v_1), \dots, (u_m, v_m)$ les réalisations d'une copule de valeurs extrêmes. On cherche à estimer la fonction de Pickands de la copule de valeurs extrêmes avec un modèle bayésien. On choisit comme espace des paramètres le sous-espace \mathcal{A}_n pour faciliter l'inférence. Comme tout élément de \mathcal{A}_n est lié à un et un seul élément de Θ_n alors on inférera dans Θ_n . Le nombre de noeuds est fixé à n .

La loi a priori qu'on choisit est concentrée sur Θ_n . L'espace Θ_n est un sous-ensemble convexe de \mathbb{R}^n de dimension $n - 1$. La loi a priori qu'on choisit est uniforme sur Θ_n , avec $n > 1$. Plus précisément, pour D de loi de Dirichlet(1, ..., 1), la loi a priori π satisfait

$$\pi(B) = \mathbb{P} \left(\frac{n}{2} D \in B \mid \max\{D_i : 1 \leq i \leq n\} \leq \frac{2}{n} \right), \quad B \subset \mathbb{R}^n, B \text{ mesurable} .$$

En effet, la loi de la v.a. $(n/2)D$ est uniforme sur $\Lambda_n = \left\{ d \in [0, n/2]^n \mid \sum_{i=1}^n d_i = n/2 \right\}$.

De même, l'ensemble Θ_n est un sous ensemble de Λ_n car on peut écrire $\Theta_n =$

$\left\{d \in \Lambda_n \mid \max\{d_i : 1 \leq i \leq n\} \leq 1\right\}$. Ce qui justifie que la loi a priori π est uniforme sur Θ_n . En plus, pour simuler uniformément dans Θ_n , on peut d'abord simuler uniformément des éléments de Λ_n et ensuite n'accepter que les éléments de Θ_n parmi les éléments de Λ_n simulés.

La fonction de vraisemblance de notre modèle est :

$$c((u_1, v_1), \dots, (u_m, v_m) | \theta) = \prod_{i=1}^m c((u_i, v_i) | \theta),$$

où $c(u_i, v_i | \theta) = \frac{\partial}{\partial x \partial y} C(u_i, v_i | \theta)$ et $C(u_i, v_i | \theta) = \exp \left\{ \log(u_i v_i) A_n \left(\frac{\log(u_i)}{\log(u_i v_i)} \mid \theta \right) \right\}$. Finalement, la densité de la loi a posteriori est tel que :

$$\pi(\theta | (u_1, v_1), \dots, (u_m, v_m)) \propto c((u_1, v_1), \dots, (u_m, v_m) | \theta) \pi(\theta).$$

4.1.4 L'approche numérique

Le but de ce chapitre est de trouver une méthode numérique à l'estimation de la fonction de Pickands A lorsqu'on dispose d'un échantillon. Pour y arriver on va générer une chaîne de Markov irréductible $\theta^0, \theta^1, \dots$ dans Θ_n de sorte que la loi stationnaire de la chaîne soit la loi a posteriori. Soit $K + 1$ les premières valeurs de la chaîne et $(x_1, y_1), \dots, (x_m, y_m)$ l'échantillon. On pose

$$\hat{A}(t) = \frac{1}{K + 1} \sum_{i=0}^K A_n(t | \theta^i), \quad t \in [0, 1].$$

Comme la chaîne est à valeurs dans Θ_n les fonctions $A_n(\cdot | \theta^i)$ seront à valeurs dans \mathcal{A}_n . Pour K fixé, il se pourrait que \widehat{A}_n ne soit pas dans \mathcal{A}_n . Cependant, l'ensemble \mathcal{A} est convexe ce qui nous assure que $\widehat{A}_n \in \mathcal{A}$.

Le choix de θ^0 est déterministe. De façon naïve, pour une valeur de n élevée, on peut penser qu'un bon choix de θ^0 va faire en sorte qu'on va approcher la loi stationnaire plus rapidement. L'idée est donc de partir avec θ^0 de sorte que $A_n(\cdot | \theta^0)$ soit proche d'un estimateur fréquentiste \widehat{A} de A qui est lui-même une fonction de Pickands. L'approximation d'une fonction de Pickands par une fonction dans \mathcal{A}_n , telle que présentée en Annexe, propose un choix de $\theta \in \Theta_n$.

Exemple 4.3 *Considérons la fonction de Pickands définie par : $A(t) = [(1 - t)^2 + t^2]^{1/2}$. La figure 4.3 illustre la détermination des noeuds avec les fonctions de Pickands polygonales de A . La figure 4.4 montre les approximations de la fonction Pickands avec différentes tailles de noeuds (issus des fonctions de Pickands polygonales).*

Figure 4.3 – Illustration des fonctions de Pickands polygonales \overleftarrow{A} et \overrightarrow{A} pour approcher la fonction de Pickands $t \mapsto A(t) = [(1-t)^2 + t^2]^{1/2}$.

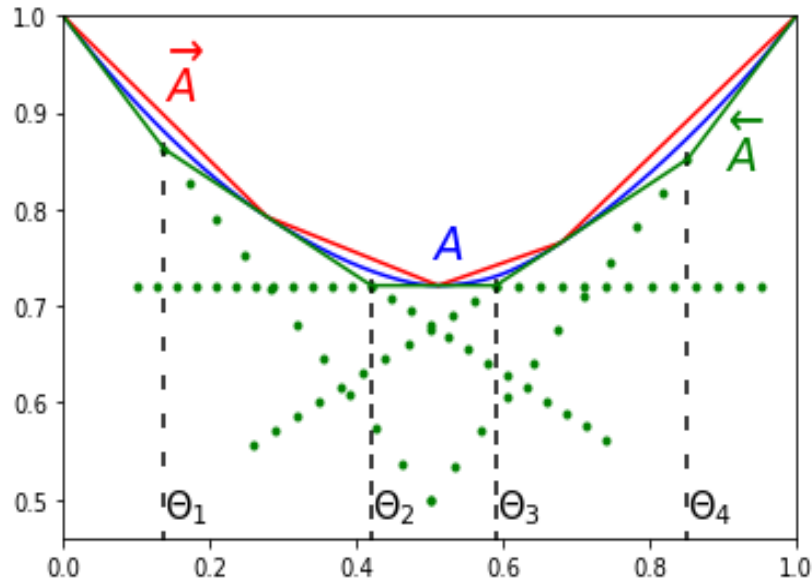
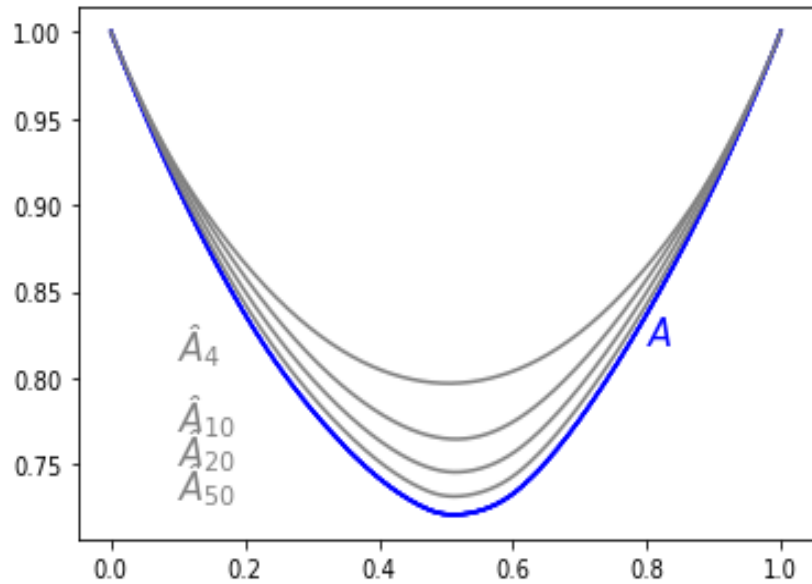


Figure 4.4 – Approximation de la fonction de Pickands $t \mapsto A(t) = [(1-t)^2 + t^2]^{1/2}$ avec n noeuds issus des fonctions Pickands polygonales pour $n = 4, 10, 20, 40, 50$.



Ainsi, de $\widehat{\widehat{A}}$ on obtient $\theta^0 \in \Theta_n$. Notez que ce ne sont pas tous les estimateurs fréquen-

tistes de A qui sont à valeurs dans \mathcal{A} . Cependant, Hall et Tajvidi en proposent un qui nous sert.

Le choix des transitions se fait en utilisant l'algorithme de Metropolis-Hastings. Pour $\theta \in \Theta_n$ on choisit deux indices i, j avec $i, j \in \{1, \dots, n\}$, $i \neq j$ et on déplace les noeuds θ_i, θ_j de façon symétrique par rapport à $1/2$ de sorte à rester sur $[0, 1]$. Les autres noeuds ne changent pas de position. On obtient ainsi une proposition $\theta' \in \Theta_n$. Cette proposition est acceptée si elle passe le test de Metropolis-Hastings. On choisit de faire ces transitions afin de rendre facile la vérification que la chaîne est irréductible. L'idée principale derrière la démonstration de l'irréductibilité repose sur le principe suivant. Pour tout choix de $\theta \in \Theta_n$, en un nombre fini de déplacements, on se rend au point $(1/2, \dots, 1/2)$.

Soit $n > 1$, $\theta, \theta' \in \Theta_n$. Le point θ' s'obtient par un déplacement $T_{ij}(\cdot \mid d)$, $i, j \in \{1, \dots, n\}$, $i \neq j$, si

$$\theta'_k = \begin{cases} \theta_i + d & \text{si } k = i \\ \theta_j - d & \text{si } k = j \\ \theta'_k = \theta_k & \text{sinon} \end{cases} \quad 0 < d \leq \min(1 - \theta_i, \theta_j)$$

Notez que les transformations $T_{ij}(\cdot \mid d)$ sont réversibles car $T_{ij}(\cdot \mid d) \circ T_{ji}(\cdot \mid d)$ est la transformation identité. Ainsi, pour montrer qu'on peut se déplacer entre les point θ et θ' en un nombre fini de pas revient à montrer qu'on peut se déplacer entre les points θ et $(1/2, \dots, 1/2)$ en un nombre fini de pas. Le reste de la démonstration se fait par induction sur $N(\theta)$ où N est la fonction qui donne le nombre de noeuds différents de $1/2$. En fait, pour $\theta \in \Theta_n$, $n > 1$,

1. Si $N(\theta) = 0$ alors on ne fait aucun pas et on est en $(1/2, \dots, 1/2)$
2. Si $N(\theta) \neq 0$ alors $N(\theta) > 1$ car $\theta_1 + \dots + \theta_n = n/2$.
3. Supposons, par hypothèse d'induction, que la proposition est vraie pour $N(\theta) = 0, 2, \dots, k, k < n$. Il existe i, j tels que $\theta_i < 1/2$ et $\theta_j > 1/2$. Posons

$$d = \min(1/2 - \theta_i, \theta_j - 1/2),$$

$$\theta' = T_{ij}(\theta \mid d).$$

On obtient $0 \leq N(\theta') < N(\theta)$.

Exemple 4.4 *On cherche une trajectoire pour se déplacer de $(1, 4, 7, 8)/10$ à $(2, 3, 5, 10)/10$.*

On obtient

$$\begin{aligned} (1, 4, 7, 8)/10 &\rightarrow (4, 4, 7, 5)/10, & (i, j) &= (1, 4), d = 3/10 \\ &\rightarrow (4, 5, 6, 5)/10, & (i, j) &= (2, 3), d = 1/10 \\ &\rightarrow (1/2, 1/2, 1/2, 1/2), & (i, j) &= (1, 3), d = 1/10 \end{aligned}$$

$$\begin{aligned} (2, 3, 5, 10)/10 &\rightarrow (5, 3, 5, 7)/10, & (i, j) &= (1, 4), d = 3/10 \\ &\rightarrow (1/2, 1/2, 1/2, 1/2), & (i, j) &= (2, 4), d = 2/10 \end{aligned}$$

$$\begin{aligned}
(1, 4, 7, 8)/10 &\rightarrow (4, 4, 7, 5)/10, & (i, j) &= (1, 4), & d &= 3/10 \\
&\rightarrow (4, 5, 6, 5)/10, & (i, j) &= (2, 3), & d &= 1/10 \\
&\rightarrow (1/2, 1/2, 1/2, 1/2), & (i, j) &= (1, 3), & d &= 1/10 \\
&\rightarrow (5, 3, 5, 7)/10, & (i, j) &= (4, 2), & d &= 2/10 \\
&\rightarrow (2, 3, 5, 10)/10, & (i, j) &= (4, 1), & d &= 3/10
\end{aligned}$$

une trajectoire de 5 pas.

4.2 Application de la méthode sur des données simulées

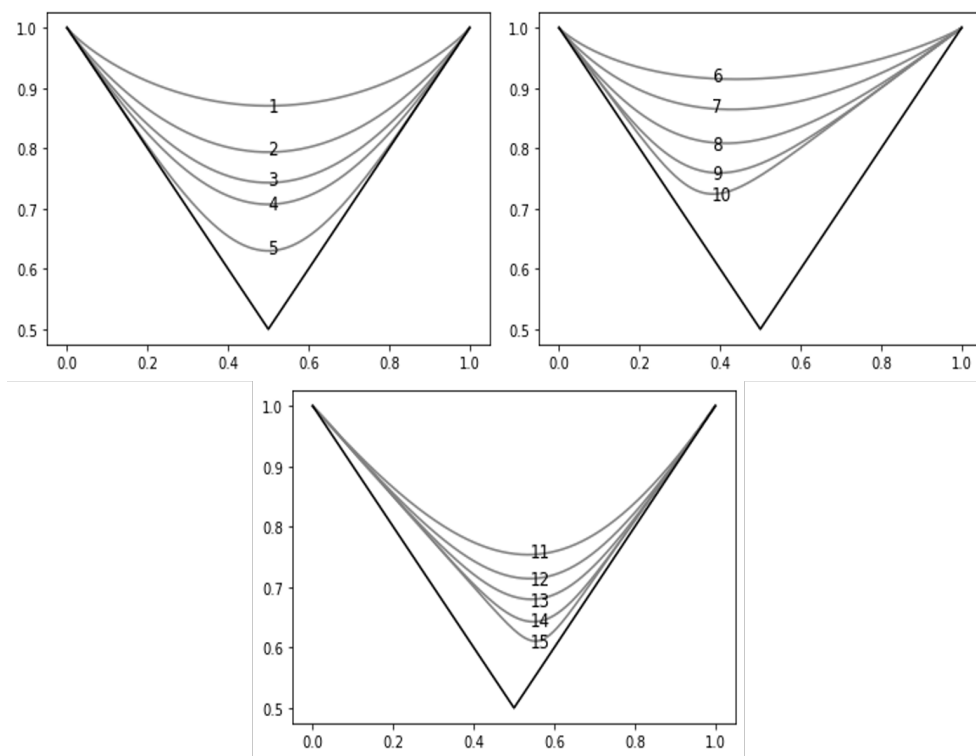
Nous avons appliqué notre méthode sur des données simulées pour évaluer sa performance. Comme dans Capéreaa, Fougere et Genest (1997) et Perron et Guillote (2009), les données simulées sont issues de la famille asymétrique logistique dont la fonction de Pickands est de la forme :

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

Nous avons utilisé le package R "SimCop" (voir Turlach et al. (2017)). Ce package simule directement les marges d'une loi de valeurs extrêmes. Nous avons défini 15 niveaux de paramètres r, α, β (voir Figure 3.4). Pour chaque niveau de paramètres, nous avons simulé 200 échantillons de taille 10, 25 et 100. Donc, on a appliqués notre méthode sur $15 \times 200 \times$

3 = 9000 échantillons. Dans un but comparatif, on a aussi calculé l'estimateur CFG et de Tadjvidi sur les 9000 échantillons. Pour chaque échantillon, nous avons appliqué notre méthode avec 20000 itérations et la taille des nœuds est 10.

Figure 4.5 – Fonction de Pickands des 15 modèles à simuler



Les 15 modèles représentent différents cas figures d'une fonction de Pickands (Voir Figure 4.5). Pour mesurer la performance de notre méthode par rapport aux estimateurs concurrents, nous avons estimé le RMISE (Relative Mean Integrated Squared Error) défini par :

$$RMISE \left(A, \hat{\mathcal{A}}_B^{(m)}, \hat{\mathcal{A}}_C^{(m)} \right) = \frac{E \left[\int_0^1 \{A_B^{(m)}(t) - A(t)\}^2 dt \right]}{E \left[\int_0^1 \{A_C^{(m)}(t) - A(t)\}^2 dt \right]}$$

Où A est la vraie fonction de Pickands, $A_B^{(m)}$ l'estimateur bayésien (notre méthode) et $A_C^{(m)}$ l'estimateur concurrent construit avec un échantillon de taille m .

On ne connaît pas la distribution des estimateurs A_B et A_C pour pouvoir calculer le RMISE de manière exact. Nous l'avons donc estimé.

Soit $M \in \{1, \dots, 15\}$, $C \in \{CFG, TDJ\}$ le concurrent, $m \in \{10, 25, 100\}$ la taille d'échantillon et $I = 200$ le nombre d'échantillons de taille M du modèle m de paramètres r, α, β .

L'estimateur de $RMISE \left(A_{r,\alpha,\beta}, \widehat{\mathcal{A}}_B^{(m)}, \widehat{\mathcal{A}}_C^{(m)} \right)$ est :

$$\widehat{RMISE} \left(A_{r,\alpha,\beta}, \widehat{\mathcal{A}}_B^{(m)}, \widehat{\mathcal{A}}_C^{(m)} \right) = \frac{\sum_{i=1}^I \int_0^1 \left\{ A_{B,i}^{(m)}(t) - A(t) \right\}^2 dt}{\sum_{i=1}^I \int_0^1 \left\{ A_{C,i}^{(m)}(t) - A(t) \right\}^2 dt}$$

Où $A_{B,i}^{(m)}$ ($A_{C,i}^{(m)}$) est l'estimation bayésienne (concurrente) de la fonction de Pickands de l'échantillon i de taille m du modèle M .

Table 4.1 – RMISE de notre estimateur bayésien par rapport à l’estimateur Capéraà et al (CFG) et de Tadjvidi (TDJ) suivant modèle et la taille de l’échantillon.

Modèle	r, α, β	$m = 10$		$m = 25$		$m = 100$	
		CFG	TDJ	CFG	TDJ	CFG	TDJ
1	1,25; 1; 1	0,07	0,07	0,14	0,06	0,27	0,03
2	1,5; 1; 1	0,05	0,05	0,11	0,06	0,47	0,07
3	1,75; 1; 1	0,42	0,44	0,75	0,46	1,57	0,28
4	2; 1; 1	0,57	1,35	1,6	1,22	2,58	0,64
5	3; 1; 1	2,93	11,74	6,47	8,81	5,37	2,74
6	1,25; 0,9; 0,5	0,21	0,2	0,48	0,17	0,65	0,07
7	1,5; 0,9; 0,5	0,06	0,07	0,18	0,07	0,28	0,03
8	2; 0,9; 0,5	0,03	0,04	0,09	0,05	0,33	0,04
9	3; 0,9; 0,5	0,16	0,16	0,27	0,17	0,57	0,10
10	5; 0,9; 0,5	0,24	0,34	0,46	0,26	0,61	0,11
11	2; 0,75; 0,95	0,24	0,27	0,50	0,25	0,89	0,17
12	2,5; 0,75; 0,95	0,78	0,78	1,20	0,77	1,47	0,32
13	3,25; 0,75; 0,95	1,06	1,76	2,01	1,54	1,86	0,49
14	5; 0,75; 0,95	1,66	3,01	1,63	2,59	1,61	1,45
15	10; 0,75; 0,95	1,64	3,58	1,99	2,51	1,49	0,61

On a 3 groupes de modèles :

- **Symétriques (modèle 1 à 5)** : L’estimateur bayésien performe mieux que ses concurrents pour les trois premiers modèles (1, 2, 3).

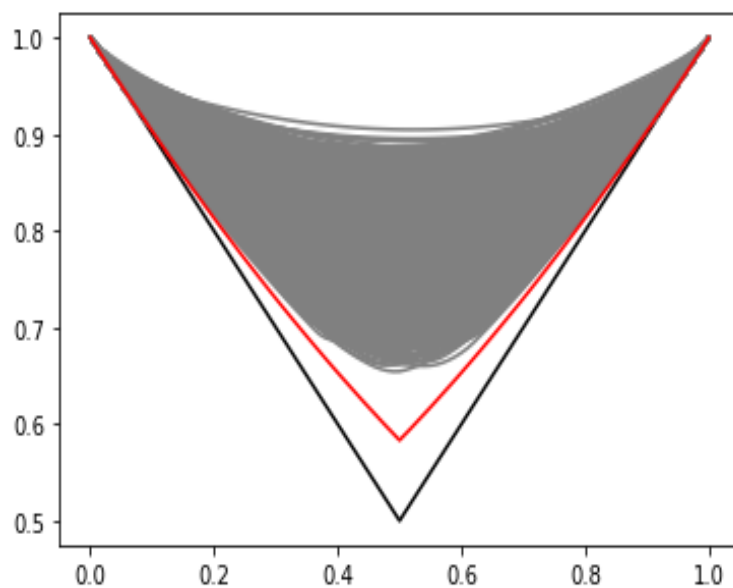
- **Asymétriques a gauche (modèle 6 à 10) :** L'estimateur bayésien performe mieux que ses concurrents pour toutes les modèles.
- **Asymétrique a droite (modèle 11 à 15) :** L'estimateur bayésien performe mieux que ses concurrents pour les deux premiers modèles (11, 12).

De manière globale, notre estimateur bayésien performe mieux en échantillon de petite taille. En fait, sa performance par rapport à l'estimateur CFG augmente au fur et à mesure que la taille de l'échantillon diminue. L'estimateur CFG a des propriétés asymptotiques puissantes comme nous l'avons expliqué dans le chapitre 2 mais il est désavantagé dans le cas où la taille de l'échantillon est petite. L'estimateur CFG avec un échantillon de petite taille a même tendance à ne pas donner une fonction de Pickands. La figure 3.6 (b) montre des fonctions qui ne sont clairement pas de Pickands : soit ils sortent des bornes ($t \vee (1 - t)$ et 1), soit ils ne sont pas convexes. Notre estimateur bayésien corrige cette anomalie liée à la taille de l'échantillon et donne toujours une fonction de Pickands (voir Figure 3.6 (a)).

Cependant notre estimateur bayésien sous performe par rapport à ses concurrents lorsque la vraie fonction de Pickands est proche de la borne $t \mapsto t \vee (1 - t)$ (modèles 4, 5, 13, 14, 15). Nous avons deux explications. La première explication se trouve dans l'exemple 4.2. L'exemple 4.2 nous a montré que si $A_n(\cdot | \theta)$ est la meilleure approximation dans \mathcal{A}_n de la fonction de Pickands comonotique alors $\|A_n(\cdot | \theta) - A_{comon}\|_\infty = 1/(n + 2)$ où A_{comon} est la copule comonotique. La deuxième explication se trouve dans la loi a priori. La figure 4.6 illustre des réalisations de la loi a priori. On constate que les fonctions de pickands générées par la loi a priori ont de la difficulté à s'approcher de la meilleure approximation de la fonction de Pickands comonotique dans \mathcal{A}_{10} . Cet aspect peut être un inconvénient

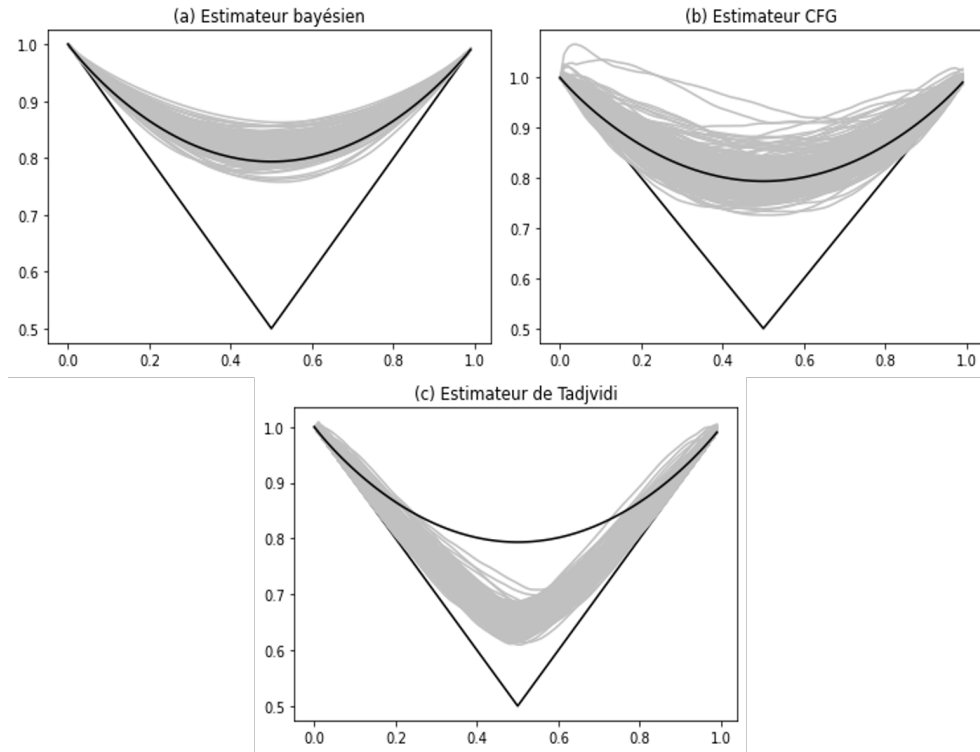
de notre méthode. Mais, n'oublions pas que nous avons choisi arbitrairement la loi a priori et la densité instrumentale. Il est possible que cette anomalie soit corrigée avec un bon choix de la loi a priori et de la densité instrumentale. La flexibilité et les perspectives d'amélioration rendent notre méthode avantageuse. En plus, dans la pratique le cas où la vraie fonction de Pickands est proche de la dépendance parfaite ($t \vee (1 - t)$) doit être extrêmement rare.

Figure 4.6 – Simulation de fonctions de Pickands à partir de la loi a priori avec 1000 itérations (courbes grises).



La courbe rouge est la meilleure approximation de la fonction de Pickands comonotique dans \mathcal{A}_{10} .

Figure 4.7 – Estimation de la fonction de Pickands des 200 échantillons de taille 25 du modèle 2 ($r=1,5$; $\alpha = 1$; $\beta = 1$) avec les trois méthodes. La ligne noire est la vraie fonction de Pickands.



4.3 Application de la méthode sur des données réelles

Les données sont fournies par Copenhagen Reinsurance Company (CRC) et concernent la compensation des pertes liées à 2167 incendies industrielles au Danemark entre 1980 et 1990. La compagnie d'assurance CRC a compensé les pertes. Les compensations se divisent en trois catégories de dommages : dommages liés au bâtiment (Building), dommages liés aux contenus (Contents) et dommages liés au profit (Profits). Donc, la base de données est composée de 2167 observations (les incendies) et 3 variables (compensations divisées

par catégories). On considère les données comme les maximums.

Les compensations sont des coûts pour l'assureur, c'est pourquoi il est crucial d'estimer la probabilité qu'elles dépassent un certain seuil. Dans McNeil (1997) et Resnick (1997), la distribution des compensations totales (Building+Contents+Profits) est modélisée par la théorie des valeurs extrêmes univariée. Le recours à une méthode univariée sous tend que les trois catégories de compensations sont indépendantes. Mais cette hypothèse n'est pas plausible : Guillotte et al. (2011) ont étudié la copule (dépendance) entre chaque paire de catégories de compensation en supposant que les copules sont de valeurs extrêmes. En outre, Guillotte et al. (2011) ont estimé les copules de valeurs extrêmes avec une méthode bayésienne. Nous avons pris la même direction mais en utilisant notre méthode bayésienne. Nous avons transformé les observations en marges avec la fonction de répartition empirique :

$$x_{ij} \longrightarrow z_{ij} = \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{\{x_{kj} \leq x_{ij}\}}$$

Où x_{ij} est la compensation de l'incendie $i \in \{1, \dots, N\}$ et de la catégorie $j \in \{1, 2, 3\}$ et $N = 2167$.

La fonction de Pickands «Building vs Contents» est légèrement asymétrique vers la gauche. La fonction de Pickands «Contents vs Profits» est asymétrique vers la droite et est la plus proche de la fonction de Pickands de dépendance parfaite ($A(t) = \max\{t, 1-t\}$). Celle de «Building vs Profits» est presque symétrique (voir figure 4.9). Toutes les trois fonctions de Pickands, ne sont pas proches de la fonction Pickands indépendante ($A(t) = 1$).

En résumé, il existe une dépendance deux à deux positive entre les trois catégories

de compensation. Mais, elle est plus aigüe entre celles liées au contenu et celles liées au profits. On peut en déduire qu'en modélisant la distribution des trois compensations indépendamment, on risque de sous-estimer la probabilité que la compensation totale dépasse un certain seuil. Ces résultats ressemblent à celui de Guillote et al. (2011).

Figure 4.8 – Nuages de points de catégories de compensations par paire

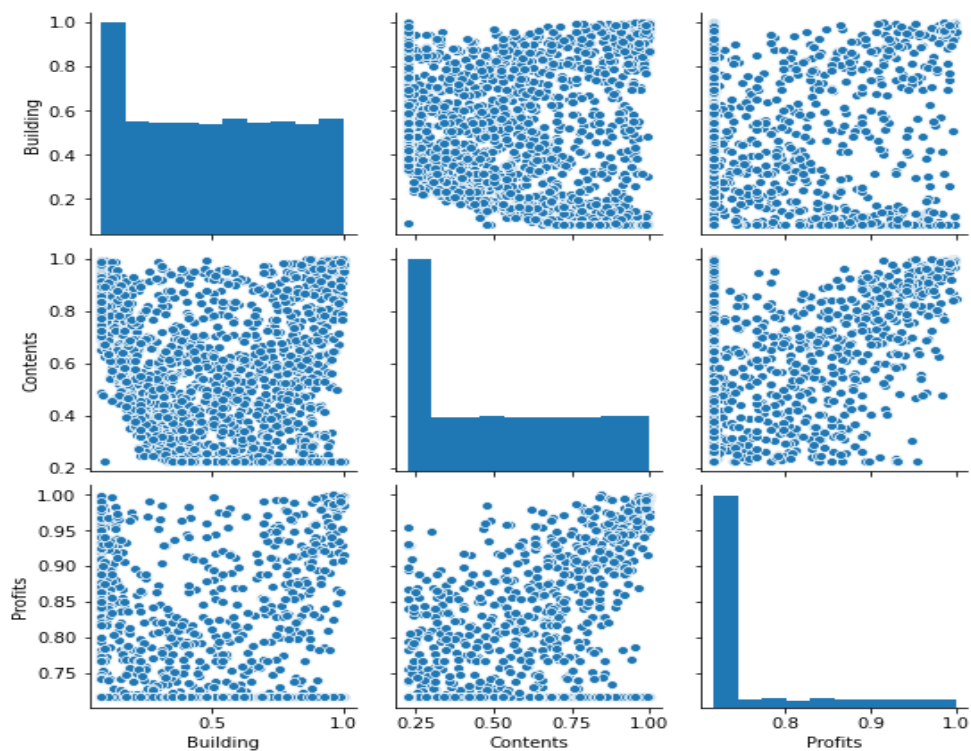
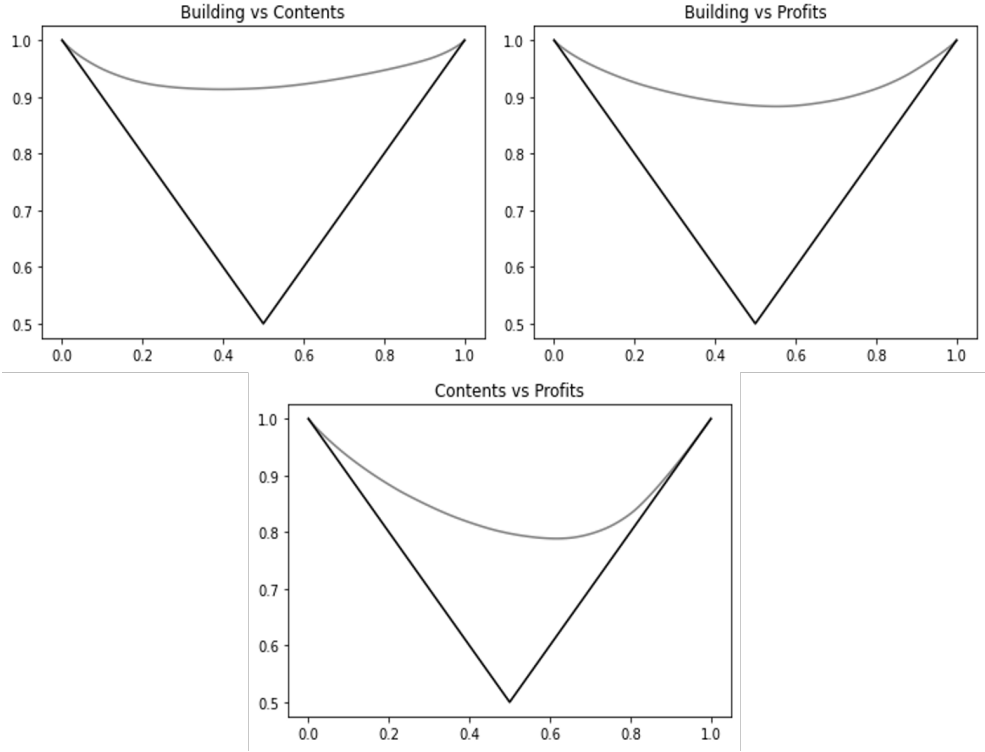


Figure 4.9 – Fonctions de Pickands estimées (moyennes a posteriori) pour chaque paire



Conclusion

Les résultats de la simulation ont montré que notre estimateur bayésien est un candidat crédible pour estimer la fonction de Pickands. Dans certains cas il est meilleur que l'estimateur de Capéràa et al et de Tadjvidi. Néanmoins, il performe moins que ses deux concurrents quand la vraie fonction de Pickands est proche de $t \mapsto t \vee (1 - t)$. Ce désavantage peut être liée à la loi a priori choisie. En rappel, on a choisi une loi a priori uniforme et qui fixe le nombre de nœuds. On peut améliorer les résultats en choisissant une loi a priori qui ne fixe pas le nombre de nœuds. Cela étant on doit trouver la méthode MCMC qui a comme loi stationnaire cette éventuelle loi. Une des pistes est d'utiliser la méthode des sauts réversibles (Voir Green (1995)).

Par ailleurs, la sous-performance relative de notre méthode dans certains cas (vrai fonction de Pickands proche de $t \mapsto t \vee (1 - t)$) peut être simplement liée au nombre de nœuds fixé trop petit (égale à 10). Il est certain qu'en choisissant un nombre de nœuds plus élevé, l'estimation sera plus précise. Mais, est-ce que l'augmentation du nombre nœuds améliorera assez les estimations ?

L'autre aspect important est le coût computationnel de notre méthode. L'algorithme d'évaluation numérique de notre méthode est programmable. En plus, notre estimateur bayésien

a performé sur les données simulées dans la majorité des cas de figures. Donc, il semble prêt à être utilisé dans la pratique. Cela étant, le praticien dispose d'une panoplie d'estimateurs dont les estimateurs de Tadjvidi et de Capéràa et al. Or, les estimateurs de Tadjvidi et CFG sont moins couteux computationnellement que notre estimateur bayésien. Toutefois, il est assez rare que la vraie fonction de Pickands soit proche de la fonction de dépendance parfaite ($t \mapsto t \vee (1 - t)$). C'est pourquoi, il est plus plausible que l'estimateur bayésien soit meilleur que ses concurrents surtout quand le nombre de nœuds est élevé. Le praticien est amené à arbitrer entre le coût computationnel et la précision des estimations.

Finalement, les perspectives de recherche relative à l'estimation bayésienne des fonctions de Pickands sont nombreuses. On peut trouver une nouvelle manière de construire une fonction de Pickands et mettre en œuvre un estimateur bayésien basé sur cette construction. Également, on peut décider d'estimer de manière bayésienne un modèle déjà existant ou d'améliorer un estimateur bayésien comme le nôtre.

Annexe A

Mesure spectrale

Rappelons certaines définitions,

Définition A.1 Une fonction $A: [0, 1] \mapsto \mathbb{R}$ est une fonction de Pickands si

1. A est convexe,
2. $t \vee (1 - t) \leq A(t) \leq 1$, $t \in [0, 1]$.

Définition A.2 Une mesure positive Λ sur $([0, 1], \sigma([0, 1]))$ est une mesure spectrale si

1. $\int_{[0,1]} \omega \Lambda(d\omega) = 1$,
2. $\int_{[0,1]} (1 - \omega) \Lambda(d\omega) = 1$.

Notons par A_+ la dérivée à droite de A sur $[0, 1)$. On a le lien suivant entre les fonctions de Pickands et les mesures spectrales.

Lemme A.1 *Si Λ est une mesure spectrale alors la fonction $A: [0, 1] \mapsto \mathbb{R}$ donnée par*

$$A(t) = \int_{[0,1]} \{[t(1-\omega)] \vee [(1-t)\omega]\} \Lambda(d\omega), \quad t \in [0, 1] \quad (\text{A.1})$$

est une fonction de Pickands.

Si A est une fonction de Pickands alors la mesure Λ satisfaisant

$$\Lambda([0, t]) = \begin{cases} 1 + A_+(t) & \text{si } t \in [0, 1), \\ 2 & \text{si } t = 1 \end{cases}$$

est une mesure spectrale et cette mesure spectrale permet de retrouver A à partir de l'équation (A.1).

Le lemme suivant vient donner une autre représentation de la fonction spectrale proche du Lemme [A.1](#).

Annexe B

Démonstrations

Dans cette partie de l'annexe nous énonçons et démontrons les lemmes utiles pour notre méthode.

1. **Lemme B.1** *On a les résultats suivants,*

1. *Si G est une fonction de répartition pour une variable aléatoire Y à valeurs dans $[0, 1]$ avec $\mathbb{E}[Y] = 1/2$ alors la fonction $A: [0, 1] \mapsto \mathbb{R}$ donnée par*

$$A(t) = (1 - t) + 2 \int_0^t G(\omega) d\omega, \quad t \in [0, 1] \tag{B.1}$$

est une fonction de Pickands.

2. Si A est une fonction de Pickands alors la fonction G satisfaisant

$$G(t) = \begin{cases} 0 & \text{si } t < 0 \\ (1 + A_+(t))/2 & \text{si } t \in [0, 1), \\ 1 & \text{si } t \geq 1 \end{cases} \quad (\text{B.2})$$

est une fonction de répartition pour une variable aléatoire Y à valeurs dans $[0, 1]$ avec $\mathbb{E}[Y] = 1/2$.

3. Si A est une fonction de Pickands et les équations (A.1) et (B.1) sont satisfaites alors

$$\Lambda([0, t]) = 2G(t), \quad t \in [0, 1].$$

Démonstration

1. Soit G une fonction de répartition pour une variable aléatoire Y à valeurs dans $[0, 1]$ avec $\mathbb{E}[Y] = 1/2$ et

$$A(t) = (1 - t) + 2 \int_0^t G(\omega) d\omega, \quad t \in [0, 1].$$

La fonction A est convexe sur $[0, 1]$, $A(0) = 1$ et $A(1) = 1$ car $\mathbb{E}[Y] = 1/2$. De plus

$$\begin{aligned}
A(t) - (1 - t) &= 2 \int_0^t G(\omega) d\omega, \quad t \in [0, 1], \\
&\geq 0, \quad t \in [0, 1], \\
A(t) - t &= (1 - 2t) + 2 \int_0^t G(\omega) d\omega, \quad t \in [0, 1], \\
&= 1 - 2 \int_0^t \{1 - G(\omega)\} d\omega, \quad t \in [0, 1], \\
&\geq 1 - 2 \int_0^1 \{1 - G(\omega)\} d\omega \\
&= 0
\end{aligned}$$

ce qui montre que $(1 - t) \vee t \leq A(t) \leq 1$ pour tout $t \in [0, 1]$.

2. Soit A une fonction de Pickands. Posons

$$\begin{aligned}
B(t) &= A(t) - (1 - t), \quad t \in [0, 1] \\
B^+(t) &= A_+(t) + 1, \quad t \in [0, 1]
\end{aligned}$$

La fonction $B: [0, 1] \mapsto \mathbb{R}$ est continue, croissante et convexe sur $[0, 1]$, $B(0) = 1$, $B(1) = 1$. Cette fonction est aussi absolument continue et on peut écrire (voir Rockafellar (2015)),

$$\begin{aligned}
B(t) &= B(0) + \int_0^t B_+(\omega) d\omega, \quad t \in [0, 1] \\
&= \int_0^t \{A_+(\omega) + 1\} d\omega, \quad t \in [0, 1] \\
&= 2 \int_0^t G(\omega) d\omega, \quad t \in [0, 1]
\end{aligned}$$

avec G la fonction donnée à l'équation (B.2). Par la convexité de la fonction A la fonction G est croissante, continue à droite. De plus $0 \leq G(t) \leq 1$, $t \in [0, 1]$ car $t \vee (1 - t) \leq A(t)$, $t \in [0, 1]$. Ceci montre que G est une fonction de répartition pour une variable aléatoire Y à valeurs dans $[0, 1]$. Enfin,

$$\mathbb{E}[Y] = \int_0^1 G(\omega) d\omega = B(1)/2 = 1/2.$$

3. Si la fonction de Pickands A est obtenue à partir de la mesure spectrale Λ selon l'équation (A.1) alors

$$\begin{aligned} A(t) - (1 - t) &= \int_{[0,1]} [\{[t(1 - \omega)] \vee [(1 - t)\omega]\} - (1 - t)\omega] \Lambda(d\omega) \\ &= \int_{[0,1]} [\{t - \omega\} \vee 0] \Lambda(d\omega) \\ &= \int_{[0,t]} \int_{[\omega,t]} du \Lambda(d\omega) \\ &= \int_{[0,t]} \int_{[0,u]} \Lambda(d\omega) du \\ &= \int_0^t \Lambda([0, u]) du \end{aligned}$$

Si A s'obtient de l'équation (B.1) alors on obtient que G et la fonction $t \mapsto \Lambda([0, t])$ sont égales presque partout sur $[0, 1]$. Par la continuité à droite et le fait que $G(1) = 1$ et $\Lambda([0, 1]) = 2$ on obtient que $2G(t) = \Lambda([0, t])$ pour tout $t \in [0, 1]$.

Notez que dans la partie 1. on utilise le fait que $\int_0^1 G(\omega) d\omega = 1/2$ pour montrer que

$A(1) = 1$. En fait,

$$\begin{aligned}
\int_{[0,1]} G(\omega) d\omega &= \int_{[0,1]} \int_{[0,\omega]} G(dt) d\omega \\
&= \int_{[0,1]} \int_{[t,1]} d\omega G(dt) \\
&= \int_{[0,1]} (1-t) G(dt) \\
&= \mathbb{P}(Y \in [0, 1]) - \mathbb{E}[Y].
\end{aligned}$$

□

Exemple B.1 Dans la famille des fonctions de Pickands les cas extrêmes sont

1. La copule indépendante, $A(t) = 1$ pour tout $t \in [0, 1]$ ce qui, dans le Lemme [B.1](#), correspond à Y est de loi Bernoulli(1/2). On choisit la loi de Y qui maximise $\text{Var}(Y)$ sous la contrainte $\mathbb{P}(Y \in [0, 1]) = 1$, $\mathbb{E}[Y] = 1/2$.
2. La copule comonotone, $A(t) = t \vee (1 - t)$ pour tout $t \in [0, 1]$ ce qui, dans le Lemme [B.1](#), correspond à $Y = 1/2$ avec probabilité 1. On choisit la loi de Y qui minimise $\text{Var}(Y)$ sous la contrainte $\mathbb{P}(Y \in [0, 1]) = 1$, $\mathbb{E}[Y] = 1/2$.

Soit $\Theta_n = \{\theta \in [0, 1]^n : \theta_1 + \dots + \theta_n = n/2\}$, $\theta \in \Theta_n$, $n > 1$. Considérons les fonctions $G_n(\cdot | \theta)$ et $A_n(\cdot | \theta)$ définies précédemment. C'est-à-dire, pour $\theta \in \Theta_n$ on pose

$G_n(1 \mid \theta) = 1$ et

$$\begin{aligned}
G_n(x \mid \theta) &= \frac{1}{n+2} \left[\ell + 1 + \frac{1}{\Delta\theta_{(\ell)}} \left\{ \frac{(x - \theta_{(\ell)})^2}{\Delta\theta_{(\ell)} + \Delta\theta_{(\ell+1)}} - \frac{(\theta_{(\ell+1)} - x)^2}{\Delta\theta_{(\ell-1)} + \Delta\theta_{(\ell)}} \right\} \right] \\
&\quad \text{si } \theta_{(\ell)} \leq x < \theta_{(\ell+1)}, 0 \leq \ell \leq n. \\
A_n(t \mid \theta) &= (1-t) + 2 \int_0^t G_n(x \mid \theta) dx \\
&= (1-t) + \frac{2}{n+2} \left[(\ell+1)t - \sum_{k=0}^{\ell} \theta_{(k)} \right. \\
&\quad \left. + \frac{1}{3} \left\{ \frac{(\theta_{(\ell+1)} - t)^3}{\Delta\theta_{(\ell)}(\Delta\theta_{(\ell-1)} + \Delta\theta_{(\ell)})} + \frac{(t - \theta_{(\ell)})^3}{\Delta\theta_{(\ell)}(\Delta\theta_{(\ell)} + \Delta\theta_{(\ell+1)})} - \Delta\theta_{(\ell)} \right\} \right] \\
&\quad \text{si } \theta_{(\ell)} \leq t < \theta_{(\ell+1)}, 0 \leq \ell \leq n.
\end{aligned}$$

Lemme B.2 Soit $\theta \in [0, 1]^n$. Les fonctions $G_n(\cdot \mid \theta)$ sont des fonctions de répartition pour des variables aléatoires Y sur $[0, 1]$ avec $\mathbb{E}[Y] = 1/2$ si et seulement si $\theta \in \Theta_n$.

Démonstration On décompose la fonction $G_n(\cdot \mid \theta)$ dans une moyenne de fonctions.

Soit $x \in [0, 1]$.

$$\begin{aligned}
G_n(x \mid \theta) &= \frac{1}{n+2} \sum_{k=0}^{n+1} G_{nk}(x \mid \theta) \\
G_{nk}(x \mid \theta) &= \begin{cases} 0 & \text{si } x \in (-\infty, \theta_{(k-1)}) \\ \frac{(x - \theta_{(k-1)})^2}{\Delta\theta_{(k-1)}(\Delta\theta_{(k-1)} + \Delta\theta_{(k)})} & \text{si } x \in [\theta_{(k-1)}, \theta_{(k)}) \\ \left\{ 1 - \frac{(\theta_{(k+1)} - x)^2}{\Delta\theta_{(k)}(\Delta\theta_{(k-1)} + \Delta\theta_{(k)})} \right\} & \text{si } x \in [\theta_{(k)}, \theta_{(k+1)}) \\ 1 & \text{si } x \in [\theta_{(k+1)}, +\infty) \end{cases}
\end{aligned}$$

Suite à des calculs élémentaires, on vérifie que les fonctions $G_k(\cdot \mid \theta)$, sont des fonctions de répartition pour $k = 0, \dots, n+1$. Ceci montre que $G(\cdot \mid \theta)$ est aussi une fonction de

répartition pour une variable aléatoire Y à valeurs dans $[0, 1]$. Comme la variable aléatoire Y est positive on trouve

$$\begin{aligned}
 \mathbb{E}_\theta[Y] &= \int_0^\infty \mathbb{P}_\theta(Y > x) dx \\
 &= \int_0^1 \mathbb{P}_\theta(Y > x) dx \\
 &= 1 - \int_0^1 \mathbb{P}_\theta(Y \leq x) dx \\
 &= 1 - A_n(1 | \theta)/2
 \end{aligned}$$

On obtient $\mathbb{E}_\theta[Y] = 1/2$ si et seulement si $A(1 | \theta) = 1$. Un calcul direct d'intégrale donne la fonction $A_n(\cdot | \theta)$ donnée plus haut. En évaluant on trouve

$$A_n(1 | \theta) = \frac{2}{n+2}[(n+1) - \sum_{k=0}^n \theta_{(k)}] = \frac{2}{n+2}[(n+1) - \sum_{k=1}^n \theta_k]$$

ce qui montre que $\mathbb{E}[Y] = 1/2$ si et seulement si $\theta_1 + \dots + \theta_n = 1$.

□

Lemme B.3 *Pour tout $A \in \mathcal{A}$, $n > 1$, il existe $\theta \in \Theta_n$ tel que*

$$|A(x) - A_n(x)| \leq \frac{8}{n+2}, \forall x \in [0, 1]$$

Démonstration. La démonstration consiste à borner la fonction A par des fonctions de Pickands \vec{A} et \overleftarrow{A} qui sont linéaires par morceaux. Plus précisément, on veut obtenir,

1. $\overleftarrow{A} \leq A(x) \leq \overrightarrow{A}(x)$ pour tout $x \in [0, 1]$
2. $\overrightarrow{A}(x) - \overleftarrow{A}(x) \leq 1/n$, pour tout $x \in [0, 1]$
3. $|A(x) - \overleftarrow{A}(x)| \leq 6/(n+2)$ pour tout $x \in [0, 1]$

Pour $k = 0, \dots, n$ on considère L_k la droite de pente $(2k/n) - 1$ tangente à la fonction

A. En fait, pour tout $k \in \{0, \dots, n\}$ il existe un point $p_k \in [0, 1]$ tel que

$$A(p_k) - \left(\frac{2k}{n} - 1\right) p_k \leq A(x) - \left(\frac{2k}{n} - 1\right) x \quad \forall x \in [0, 1]$$

et

$$L_k(x) = \left(\frac{2k}{n} - 1\right) (x - p_k) + A(p_k), \quad x \in [0, 1]$$

On pose

$$\overleftarrow{A}(x) = \max(L_0(x), \dots, L_n(x)), \quad x \in [0, 1]$$

$$\overrightarrow{A}(x) = A(x) \quad \text{si} \quad \overleftarrow{A}(x) = A(x), \quad x \in [0, 1]$$

Si $x \in [0, 1]$ et $\overleftarrow{A}(x) < A(x)$ alors on pose

$$\begin{aligned} u &= \sup\{t \in [0, x]: A(t) = \overleftarrow{A}(t)\} \\ v &= \inf\{t \in [x, 1]: A(t) = \overleftarrow{A}(t)\} \\ \overrightarrow{A}(x) &= \frac{(v-x)A(u) + (x-u)A(v)}{v-u} \\ k &= \min\{j \in \{1, \dots, n\}: \left(\frac{2j}{n} - 1\right) (x - v) + A(v) \leq A(x)\}. \end{aligned}$$

On obtient $L_{k-1}(u) = A(u)$ et $L_k(v) = A(v)$. De plus,

$$\left(\frac{2(k-1)}{n} - 1\right) \leq \frac{A(v) - A(u)}{v - u} \leq \left(\frac{2k}{n} - 1\right)$$

Ainsi, lorsque $x \in [0, 1]$ et $\overleftarrow{A}(x) < A(x)$, on a

$$\begin{aligned} L_{k-1}(x) &< \overleftarrow{A}(x) \\ &< A(x) \\ &\leq \overrightarrow{A}(x), \quad \text{convexité} \\ &= \frac{(v-x)A(u) + (x-u)A(v)}{v-u} \\ &= \left(\frac{A(v) - A(u)}{v-u}\right)(x-u) + A(u) \\ &\leq \left(\frac{2k}{n} - 1\right)(x-u) + A(u) \end{aligned}$$

De la même façon,

$$\begin{aligned} L_k(x) &< \overleftarrow{A}(x) \\ &< A(x) \\ &\leq \overrightarrow{A}(x), \quad \text{convexité} \\ &= \frac{(v-x)A(u) + (x-u)A(v)}{v-u} \\ &= \left(\frac{A(v) - A(u)}{v-u}\right)(x-v) + A(v) \\ &\leq \left(\frac{2(k-1)}{n} - 1\right)(x-v) + A(v) \end{aligned}$$

et

$$\vec{A}(x) - \overleftarrow{A}(x) \leq \frac{2}{n} \min((x - u), (v - x)) \leq 1/n.$$

La fonction \overleftarrow{A} est une fonction de Pickands car

1. \overleftarrow{A} est convexe car les fonctions L_0, \dots, L_n sont convexes.
2. $\overleftarrow{A}(t) \geq \max(L_0(t), L_n(t)) = \max(1 - t, t)$, $t \in [0, 1]$.
3. $1 = \max(t, 1 - t) \Big|_{t=0} \leq \overleftarrow{A}(0) \leq A(0) = 1$.
4. $1 = \max(t, 1 - t) \Big|_{t=1} \leq \overleftarrow{A}(1) \leq A(1) = 1$.

Pour $k = 1, \dots, n$ on pose θ_k la solution en $\in [0, 1]$ de l'équation

$$L_{k-1}(t) = L_k(t).$$

On obtient,

$$\begin{aligned} \overleftarrow{A}(t) - (1 - t) &= \begin{cases} 0 & \text{si } t \in [0, \theta_1) \\ 2[\sum_{k=0}^{\ell-1} k(\theta_{k+1} - \theta_k) + \ell(t - \theta_\ell)]/n & \text{si } \theta_\ell \leq t < \theta_{\ell+1}, \ell = 1, \dots, n \\ 1 & \text{si } t = 1 \end{cases} \\ &= \begin{cases} 0 & \text{si } t \in [0, \theta_1) \\ 2[\sum_{k=1}^{\ell} (t - \theta_k)]/n & \text{si } \theta_\ell \leq t < \theta_{\ell+1}, \ell = 1, \dots, n \\ 1 & \text{si } t = 1 \end{cases} \end{aligned}$$

Soit $\ell \in \{0, \dots, n\}$ tel que $\theta_\ell < 1$, $\theta_{\ell+1} = 1$. Par continuité de \overleftarrow{A} on obtient

$$\begin{aligned}
1 &= 2\left[\sum_{k=1}^{\ell}(1 - \theta_k)\right]/n \\
&= 2\left[\sum_{k=1}^n(1 - \theta_k)\right]/n, \quad (\ell < n \Rightarrow \theta_{\ell+1} = \dots = \theta_n = 1). \\
\sum_{k=1}^n \theta_k &= n/2.
\end{aligned}$$

On choisit $A_n \in \mathcal{A}_n$ basé sur $\theta_1, \dots, \theta_k$ tirés de \overleftarrow{A} . Pour $\theta_\ell \leq t < \theta_{\ell+1}$ on obtient

$$\begin{aligned}
A_n(t) - \overleftarrow{A}(t) &= 2\left(\frac{\ell+1}{n+2} - \frac{\ell}{n}\right)t - 2\left(\frac{1}{n+2} - \frac{1}{n}\right)\sum_{k=0}^{\ell}\theta_k \\
&\quad + \frac{2}{3(n+2)}\left\{\frac{(\theta_{\ell+1}-t)^3}{\Delta\theta_\ell(\Delta\theta_{\ell-1} + \Delta\theta_\ell)} + \frac{(t-\theta_\ell)^3}{\Delta\theta_\ell(\Delta\theta_\ell + \Delta\theta_{\ell+1})} - \Delta\theta_\ell\right\} \\
|A_n(t) - \overleftarrow{A}(t)| &\leq 2\frac{|n-2\ell|}{n(n+2)}t + 4\frac{1}{n(n+2)}\sum_{k=0}^{\ell}\theta_k \\
&\quad + \frac{2}{3(n+2)}\left\{\frac{(a_{\ell+1}-t)^3}{\Delta\theta_\ell(\Delta\theta_{\ell-1} + \Delta\theta_\ell)} + \frac{(t-\theta_\ell)^3}{\Delta\theta_\ell(\Delta\theta_\ell + \Delta\theta_{\ell+1})} + \Delta\theta_\ell\right\} \\
&\leq \frac{2}{n+2} + \frac{2}{n+2} + \frac{2}{n+2}
\end{aligned}$$

□

Annexe C

Copule de valeurs extrêmes multivariés

Dans notre mémoire on s'est intéressé aux copules bivariées. Mais quand on parle de copule il s'agit des copules multivariées. Donc, la copule bivariée est un cas particulier.

Définition C.1 (*Copules de dimension $d \geq 2$*)

Une copule de dimension d , $C : [0, 1]^d \rightarrow [0, 1]$ est une fonction de répartition dont les marginales uniformes sur $[0, 1]$.

Le théorème de Sklar s'applique aussi en dimension d .

Théorème C.1 (*Sklar (1959)*)

Soit F la fonction de répartition conjointe d'une variable aléatoire de dimension $d \geq 2$.

Soient F_1, \dots, F_d les fonctions de répartition marginale de F .

Il existe une copule C de dimension d tel que :

$$F(x_1, \dots, x_d) = C(F(x_1), \dots, F(x_d))$$

Et inversement,

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$$

Le domaine d'attraction bivarié peut s'étendre en dimension $d \geq 2$. Soit F la fonction de répartition conjointe d'une variable X aléatoire continue de dimension d . Soient X_1, \dots, X_n un échantillon i.i.d suivant la loi de F .

Définition C.2 (Domaine d'attraction de dimension d)

Soit G une fonction de répartition de dimension d . On dit F appartient au domaine d'attraction de G de dimension d (noté $F \in D_d(G)$) si et seulement s'il existe 2d suites $a_n = (a_n^{(1)}, \dots, a_n^{(d)}) > 0$ et $b_n = (b_n^{(1)}, \dots, b_n^{(d)})$ tel que :

$$\lim_{n \rightarrow \infty} F(a_n^{(1)}x_1 + b_n^{(1)}, \dots, a_n^{(d)}x_d + b_n^{(d)})^n = G(x_1, \dots, x_d)$$

Ce qui est équivalent à : $\frac{M_n - b_n}{a_n}$ converge en loi vers G ,

$$\text{Où } M_n = \left(\max_{1 \leq i \leq n} \{X_1^{(i)}\}, \dots, \max_{1 \leq i \leq n} \{X_d^{(i)}\} \right).$$

Supposons que $F \in D_d(G)$. G est une **distribution de valeurs extrêmes multivariée**. Si \tilde{C} est la copule associée à G alors \tilde{C} est une **copule de valeurs extrêmes**. Il est claire que les marginales de G suivent une distribution de valeurs extrêmes. Il reste à connaître la forme de la copule de valeurs extrêmes.

Lemme C.1 La copule de valeurs extrêmes \tilde{C} vérifie :

$$\lim_{n \rightarrow \infty} C(u_1^{1/n}, \dots, u_d^{1/n})^n = \tilde{C}(u_1, \dots, u_d), \forall (u_1, \dots, u_d) \in [0, 1]^d$$

Définition C.3 Une copule \tilde{C} est max-stable si et seulement si :

$$\tilde{C}\left(u_1^{1/n}, \dots, u_d^{1/n}\right)^n = \tilde{C}(u_1, \dots, u_d), \forall n \in \mathbb{N}^* \text{ et } \forall (u_1, \dots, u_d) \in [0, 1]^d$$

Théorème C.2 Une copule est de valeurs extrêmes si et seulement si elle est max-stable.

La coïncidence entre la classe de copules max-stables et la classe des copules de valeurs extrêmes, a permis Pickands (1981) et Deheuvels (1984) de trouver la forme de la copule de valeurs extrêmes.

Théorème C.3 (Pickands (1981) et Deheuvels (1984))

\tilde{C} est une copule de valeurs extrêmes si et seulement s'il existe une mesure borélienne finie

H sur $\Delta_{d-1} = \{(x_1, \dots, x_d) \in [0, 1]^d : x_1 + \dots + x_d = 1\}$, appelée **mesure spectrale** tel que :

$$\tilde{C}(u_1, \dots, u_d) = \exp \{-\ell(-\log u_1, \dots, -\log u_d)\}, \forall (u_1, \dots, u_d) \in [0, 1]^d$$

où $\ell(x_1, \dots, x_d) = \int_{\Delta_{d-1}} \max\{w_1 x_1, \dots, w_d x_d\} dH(w_1, \dots, w_d)$

ℓ est convexe, homogène (c-à-d que $\ell(cx_1, \dots, cx_d) = c\ell(x_1, \dots, x_d)$ pour $c > 0$) et vérifie

$$\max\{x_1, \dots, x_d\} \leq \ell(x_1, \dots, x_d) \leq x_1 + \dots + x_d.$$

La fonction de Pickands se définit aussi dans le cas multivarié :

$$\tilde{C}(u_1, \dots, u_d) = \exp \left\{ \left(\sum_{i=1}^d \log u_i \right) A \left(\frac{\log u_1}{\sum_{j=1}^d \log u_j}, \dots, \frac{\log u_d}{\sum_{j=1}^d \log u_j} \right) \right\}$$

où $\ell(x_1, \dots, x_d) = (x_1 + \dots + x_d)A(w_1, \dots, w_d)$, $w_j = \frac{x_j}{x_1 + \dots + x_d}$

A est la fonction de pickands en dimension d . Comme dans le cas bivarié, A est convexe

et satisfait $\max\{w_1, \dots, w_d\} \leq A(w_1, \dots, w_d) \leq 1$.

Annexe D

Algorithme de simulation des nœuds

Algorithme 2 : Simulation par MCMC des nœuds

1 Données : $e_K = \{(u_1, v_1), \dots, (u_K, v_K)\}$ échantillon de la copule , N le nombre d'itération de la chaîne, n_0 le nombre de nœuds, $\mathcal{L}(\cdot|e_K)$ la fonction de vraisemblance et $\theta^{(0)}$ la valeur initiale de la chaîne

2 Début

3 **Pour chaque** $p \in \{1, \dots, N\}$

4 Choisir de manière uniforme $i \neq j$ tel que i et j appartiennent à $\{1, \dots, n_0\}$;

5 Calculer $\bar{d}_{ij} = \min \{1 - \theta_i^{(p)}, \theta_j^{(p)}\}$;

6 Simuler $\alpha \sim U(0, 1)$;

7 Calculer $\theta'^{(p)} = (\theta_1'^{(p)}, \dots, \theta_{n_0}'^{(p)})$ tel que

$$\theta_k'^{(p)} = \theta_k^{(p)} \forall k \in \{1, \dots, n\} \setminus \{i, j\}, \theta_i'^{(p)} = \theta_i^{(p)} + \alpha \bar{d}_{ij} \text{ et}$$

$$\theta_j'^{(p)} = \theta_j^{(p)} - \alpha \bar{d}_{ij};$$

8
9
10 Calculer $\alpha^{(p)} = \min \left\{ 1, \frac{\mathcal{L}(\theta'^{(p)})}{\mathcal{L}(\theta^{(p)})} \right\}$;
11 Générer $u \sim U(0, 1)$;
12 **si** $u \leq \alpha^{(p)}$ **alors**
13 | accepter le candidat $\theta'^{(p)}$ (c.à.d $\theta^{(p+1)} = \theta'^{(p)}$);
14 **sinon**
15 | rejeter le candidat $\theta'^{(p)}$ (c.à.d $\theta^{(p+1)} = \theta^{(p)}$);

16 Sortie : Échantillon Monte Carlo des nœuds $(\theta^{(1)}, \dots, \theta^{(N)})$

Bibliographie

Bingham, N. H., Goldie, C. M., Teugels, J. L., and Teugels, J. (1989). *Regular variation*.

Number 27. Cambridge university press.

Capéraà, P., Fougere, A., and Genest, C. (1997). A stochastic ordering based on a decom-

position of kendall's in distributions with given marginals and moment problems, eds.

v. benes and j. stepan.

Carley, H. and Taylor, M. (2002). A new proof of sklar's theorem. In *Distributions with*

given marginals and statistical modelling, pages 29–34. Springer.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In *Statistical extremes*

and applications, pages 117–130. Springer.

Fisher, Ronald Aylmer, T. L. H. C. (1928). Limiting forms of the frequency distribution

of the largest or smallest member of a sample. In *Mathematical proceedings of the*

Cambridge philosophical society, volume 24, pages 180–190. Cambridge University Press.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire.

Annals of mathematics, pages 423–453.

- Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4) :711–732.
- Gudendorf (2011). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of multivariate analysis*, 102(1) :37–47.
- Guillotte, S. and Perron, F. (2008). A bayesian estimator for the dependence function of a bivariate extreme-value distribution. *Canadian Journal of Statistics*, 36(3) :383–396.
- Guillotte, S., Perron, F., and Segers, J. (2011). Non-parametric bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73(3) :377–406.
- Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, pages 835–844.
- Kochenderfer (2019). *Algorithms for optimization*. Mit Press.
- Krizmanic, D. (2014). Weak convergence of partial maxima processes in the m 1 topology. *Extremes*, 17(3) :447–465.
- Lux, T., Papapantoleon, A., et al. (2017). Improved fréchet–hoeffding bounds on d -copulas and applications in model-free finance. *Annals of Applied Probability*, 27(6) :3633–3671.
- Mc Neil, A. J. and Nešlehová, J. (2007). Multivariate archimedean copulas, d. 1.
- McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin : The Journal of the IAA*, 27(1) :117–137.

- Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.
- Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981*.
- Resnick, S. I. (1997). Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin : The Journal of the IAA*, 27(1) :139–151.
- Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.
- Rockafellar, R. T. (2015). *Convex analysis*. Princeton university press.
- Rodriguez, J. C. (2007). Measuring financial contagion : A copula approach. *Journal of empirical finance*, 14(3) :401–423.
- Sahoo, P. K. and Kannappan, P. (2011). *Introduction to functional equations*. CRC Press.
- Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8 :229–231.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728.
- Turlach, B. A., Tajvidi, N., Turlach, M. B. A., and CopApprox, G. (2017). Package ‘simcop’.