



Sur la dépendance des queues de distributions

Mohalilou Aleiyouka

► To cite this version:

Mohalilou Aleiyouka. Sur la dépendance des queues de distributions. Physique mathématique [math-ph]. Normandie Université, 2018. Français. NNT : 2018NORMLH28 . tel-02082090

HAL Id: tel-02082090

<https://theses.hal.science/tel-02082090>

Submitted on 28 Mar 2019

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

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



Normandie Université

THESE

Pour obtenir le diplôme de doctorat

Spécialité : Mathématiques

Préparée au sein de l'Université du Havre

SUR LA DÉPENDANCE DES QUEUES DE DISTRIBUTIONS

Présentée et soutenue par

Mohalilou ALEIYOUKA

**Thèse soutenue publiquement le 27 septembre 2018
devant le jury composé de**

Madame Ghislaine GAYRAUD	Professeure à l'Université de Technologie de Compiègne	Rapporteuse
Monsieur Mounir MESBAH	Professeur à l'Université Sorbonne Université (Paris 6)	Rapporteur
Monsieur Joseph NGATCHOU WANDJI	Professeur à l'Université de Lorraine	Examineur
Monsieur Sergio ALVAREZ-ANDRADE	Maître de Conférences HDR à l'Université de Technologie de Compiègne	Examineur
Monsieur Alexandre BERRED	Professeur à l'Université du Havre	Directeur de thèse

Thèse dirigée par Alexandre BERRED, laboratoire LMAH

TABLE DES MATIÈRES

Table des figures	VII
Liste des tableaux	XI
1 Copules	9
1 Notion de copule et ses propriétés	9
1.1 Introduction	9
1.2 Définition d'une copule	10
1.3 Exemple	11
1.4 Définition	11
1.5 Propriétés des copules	12
1.6 Corollaire	12
1.7 Exemple	13
1.8 Les copules usuelles	14
2 Les copules paramétriques	18
2.1 La copule Elliptique	18
2.2 Les copules Archimédiennes	25
2.3 La copule des valeurs extrêmes	32
2 Dépendance de queue de distribution	35

1	Introduction	35
2	Notion de la concordance	36
2.1	Définition	36
3	Le tau de Kendall	36
3.1	Définition	36
3.2	Théorème	37
4	Le rho de Spearman	37
4.1	Définition	37
4.2	Théorème	37
5	Dépendance de queue de distribution	37
5.1	Définition	38
5.2	Proposition	38
5.3	Exemple	39
5.4	Exemple	40
5.5	Exemple	40
6	Dépendance de queue de copules Archimédiennes	41
7	Dépendance de queue de copules elliptiques	43
8	Estimation de la dépendance de queue de distribution	45
8.1	Estimation paramétrique	45
8.2	Estimation non-paramétrique	47
3	La Dépendance de queue pour la loi hyperbolique généralisée	51
1	Introduction	51
1.1	Loi inverse gaussienne	51
1.2	Loi inverse gaussienne généralisée	52
2	Loi hyperbolique généralisée	55

2.1	Paramétrisations	57
3	Théorème	63
3.1	Démonstration	63
4	Coefficient de dépendance extrême	69
1	Introduction	69
1.1	Théorème de Fisher et Tippett	69
2	Mesures de dépendance extrême	73
2.1	Lien entre $\bar{\chi}$ et l'indice des valeurs extrêmes γ	80
3	Coefficient de dépendance extrême	81
3.1	Définition	81
5	Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues	83
1	Introduction	83
2	Les transactions	84
2.1	Les transactions de mise à jour	84
2.2	Les transactions "Utilisateur"	84
2.3	Transactions non-temps réel	85
2.4	Propriétés des transactions	85
3	Les modèles de transactions étendus	86
4	Les transactions imbriquées (emboîtées)	86
4.1	Définition	87
5	Protocoles d'ordonnancement	87
5.1	Protocole Rate Monotonic (RM)	87
5.2	Protocole Deadline Monotonic (DM)	88
5.3	Earliest Dealine First (EDF)	88

5.4	Adaptive Earliest Deadline (AED)/ Hierarchical Earliest Deadline (HED)	88
5.5	Protocole Adaptive Earliest Virtual Deadline (AEVD)	89
5.6	Protocole Generalized Earliest Deadline First (GEDF)	90
6	Étude stochastique dans les SGBDTR	92
6.1	Étude stochastique	94
7	Simulations	95
7.1	Influence du poids de <i>SPriority</i> sur les performances du système en fonction de la charge système sous des transactions plates et imbriquées.	96
7.2	Influence du nombre de données temps réel sur la performance du système.	102
7.3	Influence de la capacité informatique	106
8	Conclusion	112
A annexe A		115
1	La dérivé partielle d'une copule	115
2	La fonction de densité conditionnelle d'une loi de Student	116
B annexe B		117
1	Fonction de Bessel modifiée du troisième type	117
Bibliographie		123

TABLE DES FIGURES

1.1	Copule d'indépendance Π	15
1.2	Copule Maximale M	16
1.3	Copule Minimale W	17
1.4	Graphique de la copule gaussienne et son contour avec $\rho = 0.6$	20
1.5	Graphique de la copule gaussienne et son contour avec $\rho = -0.8$	20
1.6	Densité de la copule gaussienne et son contour avec $\rho = 0.6$	21
1.7	Densité de la copule gaussienne et son contour avec $\rho = -0.8$	21
1.8	Graphique de la copule de Student $\rho = 0.8$ et $\nu = 4$	22
1.9	Graphique de la copule de Student $\rho = -0.6$ et $\nu = 4$	23
1.10	Densité de la copule de Student et son contour avec $\rho = 0.8$ et $\nu = 4$	24
1.11	Densité de la copule de Student et son contour avec $\rho = -0.6$ et $\nu = 4$	24
1.12	Graphique de la copule de Frank	26
1.13	Contour de copule Frank	26
1.14	Densité de la copule Frank	27
1.15	Contour de la copule Frank	27
1.16	Graphique de la copule de Clayton	28
1.17	Contour de la copule de Clayton	29

1.18	Densité de la copule de Clayton	29
1.19	Contour de la copule Clayton	30
1.20	Graphique de la copule de Gumbel	31
1.21	Contour de la copule de Gumbel	32
1.22	Densité de la copule Gumbel	33
1.23	Contour de la copule Gumbel	33
2.1	Coefficient de la dépendance de queue de copule de gaussienne	43
2.2	Coefficient de la dépendance de queue de copule de Student .	44
3.1	La densité de la loi inverse gaussienne avec $\mu = 1$ et $\lambda = 0.2, 0.5, 2$.	52
3.2	La densité de la loi inverse gaussienne avec $\mu = 3, 4, 8$ et $\lambda = 1.5$.	53
3.3	La densité de la loi inverse gaussienne généralisée avec $\delta = 1, 2, 4$, $\gamma = 4$ et $\lambda = -0.5$.	54
3.4	La densité de la loi inverse gaussienne généralisée avec $\delta = 1$, $\gamma = 2, 3, 5$ et $\lambda = 2$.	54
3.5	La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\delta = 1$, $\beta = 2$ et $\lambda = 2$.	57
3.6	La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\delta = 1$, $\lambda = 2$ et $\alpha = 5$.	57
3.7	La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\beta = 0$, $\lambda = 1$ et $\alpha = 1$.	58
3.8	La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\beta = 0$, $\lambda = 1$ et $\alpha = 1$.	58
3.9	La densité de la loi hyperbolique généralisée avec $\mu = 0$, $\delta = 1$, $\beta = 0$ et $\alpha = 1$.	59

4.1	Fonction de répartition des lois de valeurs extrêmes, avec $\gamma = 1$ pour la loi de Fréchet, $\gamma = 0$ pour la loi de Gumbel et $\gamma = -1$ pour la loi de Weibull.	71
4.2	Densité des lois de valeurs extrêmes, avec $\gamma = 1$ pour la loi de Fréchet, $\gamma = 0$ pour la loi de Gumbel et $\gamma = -1$ pour la loi de Weibull.	72
4.3	χ avec $n = 500$ observations	76
4.4	χ avec $n = 2000$ observations	77
4.5	La fonction $\bar{\chi}$ avec $n = 500$ observations	78
4.6	$\bar{\chi}$ avec $n = 2000$ observations	79
5.1	Influence du poids de <i>Spriority</i> en fonction de la charge du système.	97
5.2	Influence de la structure des transactions "Utilisateur" sur la performance du système.	99
5.3	<i>SRatio</i> et sa densité bêta équivalente sous $\text{GEDF}_{a=0.5}$, <i>Quantum</i> = 30, TD = 20%.	103
5.4	Influence du nombre de données temporelles sur <i>SRatio</i>	104
5.5	Influence du nombre de données en temps réel sur <i>SRatio</i> _{user}	108
5.6	<i>SRatio</i> et sa densité bêta équivalente sous $\text{GEDF}_{a=0.3}$, <i>Quantum</i> = 30, TD = 10%.	109
5.7	Influence de la capacité informatique sur <i>SRatio</i> _{user}	110
5.8	<i>SRatio</i> et sa densité bêta équivalente avec différentes capacités d'exécution sous EDF, TD = 20%.	112
B.1	La fonction de Bessel du troisième type.	118
B.2	La fonction de Bessel de première espèce.	119
B.3	La fonction de Bessel de deuxième espèce.	120

LISTE DES TABLEAUX

1.1	Quelque copules de valeurs extrêmes	34
2.1	Dépendance de queue pour différentes copules	45
3.1	La relation entre le quatre paramètres de la distribution hyperbolique généralisée à des valeurs fixes de μ et δ	60
4.1	Quelques exemples de lois classées selon leur domaine d'attraction.	73
5.1	Caractéristiques du système et les paramètres de simulation utilisés.	99
5.2	Caractéristiques du système et les paramètres de simulation utilisés.	100
5.3	Caractéristiques du système et les paramètres de simulation utilisés.	101
5.4	Estimation et résultats du test de K.S. sous EDF et $\text{GEDF}_{a=0.5}$	102
5.5	Estimation et résultats du test de K.S. avec $\text{GEDF}_{a=0.5}$ lorsque $\text{RT} = 10\%$	105
5.6	Estimation et résultats du test K.S sous $\text{GEDF}_{a=0.5}$, lorsque $\text{RT} = 10\%$	113

LISTE DES TABLEAUX

À la mémoire de mon Père.

Remerciements

C'est avec une immense gratitude que je tiens à exprimer ma reconnaissance envers les personnes qui m'ont encouragées et soutenues tout au long de ma thèse.

Je tiens à exprimer mes vifs remerciements au Professeur *Alexandre BERRED*, mon directeur de thèse pour son encadrement continu, pour ses remarques constructives qu'il m'a fournies ainsi que pour ses précieux conseils durant toute la période de ma thèse. Je le remercie également pour la confiance et la grande liberté d'idées et de travail qu'il m'a accordées. En dehors de ses apports scientifiques, je n'oublierai pas aussi de le remercier pour ses qualités humaines, son hospitalité et son soutien permanent qui m'ont permis de mener à bien ma thèse de doctorat.

Toute ma gratitude s'adresse à Monsieur *Mounir MESBAH*, Professeur à l'Université de Pierre et Marrie Curie (Paris 6), d'avoir accepté de lire mes travaux et de rédiger un rapport sur ma thèse et pour toutes ses remarques minutieuses sur manuscrit.

Ma gratitude et mes sincères remerciements vont à Madame *Ghislaine GAYRAUD*, Professeure à l'Université de Technologie de Compiègne, pour sa gentillesse, son soutien et ses directives, et d'avoir accepté sans hésitation d'évaluer et de rapporter mes travaux. Elle n'a pas redouté de faire un voyage long et sans doute fatigant depuis Compiègne.

Je remercie infiniment Monsieur *Joseph Ngatchou-Wandji*, Professeur à l'Université de Lorraine, qui a bien voulu prendre le temps, malgré ses multiples occupations, d'avoir accepté d'examiner mes travaux et pour son rapport et ses remarques précises et de m'avoir fait l'honneur de présider le jury de ma soutenance, ainsi que Monsieur *Sergio Alvarez-Andrade*, Maître de Conférence-HDR à l'Université de Technologie de Compiègne, d'avoir accepté de faire partie de ce jury, je le remercie vivement pour le temps qu'il a consacré à la lecture de ce manuscrit.

Je voudrais exprimer mes sincères remerciements à tous mes amis ici et mes collègues thésards, pour la bonne ambiance au laboratoire et leur

amitié. Merci également aux enseignants ainsi qu'aux équipes pédagogiques qui m'ont accompagné et intégré avec beaucoup de gentillesse. Je remercie particulièrement Monsieur *Abdel-Jelil MEYINE*, Ingénieur en Mathématiques Appliquées à l'Assurance et Finance Professeur de Mathématiques à l'Académie de Versailles, il a toujours été à mes côtés dans les bons moments comme dans des moments difficiles, mais aussi pour la confiance qu'il n'a jamais cessé de témoigner à mon égard.

Enfin sans vouloir m'étendre sur ma vie en dehors du labo, je tiens à remercier mes amis *Fatimetou KHABANE*, doctorante au laboratoire EDEHN, *Ghassem LEMEIMY*, doctorant au laboratoire LMAH, *Moussa WALED* et *Yahya MOUGDER* pour m'aider à préparer mon pot de thèse. Ainsi que ma famille pour n'avoir jamais douté et m'avoir tant encouragée. Un immense merci à *Meima MOHAMED* pour l'amour qu'elle me donne chaque jour, merci pour ton soutien, ta tendresse et ton affection qui me comblent au quotidien. Tu es vraiment la perle rare de ma vie !

À vous tous, je dis merci !

Introduction générale

De nombreux événements rares mais qui furent qualifiés de première importance ont marqué la fin du $XX^{\text{ème}}$ siècle. En 1953, une forte tempête avait frappé les Pays-Bas, suite à quoi, la mer avait passé par-dessus de plusieurs digues, provoquant ainsi l'inondation de la région et de nombreux dégâts humains et matériels y ont été constatés.

Un comité a été mis en place pour faire l'état des lieux, analyser le phénomène et en tirer des recommandations pour prédire de prochains épisodes de cette catastrophe naturelle afin d'agir en conséquence. S'agissant notamment des hauteurs de digues. Ce comité devait tenir compte de beaucoup de facteurs, non seulement économiques (coût de construction, coût des inondations, $\dots$), mais aussi physiques (prévisions météorologiques, rôle du vent sur la marée, $\dots$). S'y ajoutait des données enregistrées par des balises marines concernant la hauteur des marées. En effet, il était plus judicieux de considérer les surcotes, c'est-à-dire la différence entre la hauteur réelle et la hauteur prévue de la marée, que les hauteurs des marées elles mêmes.

Ce fut à l'époque l'une des "tempêtes du siècle". Pourtant, nul doute que le "Lothar", le "Martin" survenus à la fin des années 90 et ayant touché la France en pleine figure furent des "tempêtes du millénaire".

Bien que ces événements aient de très faibles probabilités de survenue, c'est-à-dire se trouvant dans les queues de distributions. Ils apparaissent pourtant en général dans des contextes physiques nombreux et variés, en particulier les catastrophes naturelles : en hydrologie (crues décennales ou centennales et hauteur des barrages et digues susceptibles de les contenir [Aldama et Ramirez \(2002\)](#), tempêtes occasionnant d'importants dommages matériels et environnementaux [Bengtsson et Nilsson \(2007\)](#), dans les grands incendies [Alvarado *et al.* \(1998\)](#), dans les tremblements de terre [Al-Abbasi et Fahmi \(1985\)](#), [Lavenda et Cipollone \(2001\)](#), dans les risques financiers (les krachs boursiers, les crises financières), [Bouchaud et Potters \(2003\)](#)), dans les records sportifs [Eichner *et al.* \(2006\)](#), [Einmahl et Magnus \(2008\)](#) ; Mais aussi dans l'étude de la résistance d'un matériau fibreux [Daniels \(1945\)](#), $\dots$,

etc.

Une branche des statistiques essaie de trouver une solution en cas de survenue de ces phénomènes, il s'agit de la théorie des valeurs extrêmes. Notamment en Macro-économie et dans le but de prévoir les crises financières et économiques, de prédire certains événements à partir d'une étude des valeurs extrêmes d'une série, alors qu'en réalité, les risques faibles, à modérés, s'avèrent souvent bien connus et assez facilement "maîtrisables". Il est donc capital de disposer d'un certain nombre d'outils permettant d'assurer au long terme la pérennité de ces structures, la continuité des activités qui y sont liées ou encore de s'assurer de la solidité et solvabilité d'un acteur économique face à de tels risques.

L'étude des valeurs extrêmes revient à l'étude des queues de distributions des fonctions, ou de façon équivalente, à l'analyse de la plus grande observation d'un échantillon. En ce sens, nous pouvons considérer la théorie des valeurs extrêmes comme la contrepartie de la théorie statistique classique, qui est principalement basée sur l'étude de la moyenne d'un échantillon plutôt que des observations extrêmes.

L'analyse des valeurs extrêmes repose principalement sur les distributions limites des extrêmes et leurs domaines d'attraction. Ces distributions apparaissent comme les seules distributions limites possibles du maximum d'un échantillon de variables aléatoires indépendantes et identiquement distribuées (i.i.d).

Les événements extrêmes exigent une modélisation plus robuste et une analyse plus poussée. Simplement puisque les pertes qui en résultent sont le plus souvent significatives et peuvent ruiner, à titre d'exemple, les fonds d'investissement les plus rentables. L'étude de la dépendance est donc primordiale pour identifier et/ou construire le portefeuille qui soit le moins vulnérable possible aux pertes.

La dépendance entre les rendements des actifs financiers a significativement augmenté durant ces dernières années. La fameuse crise des Subprimes (crise financière de 2007) en constitue un témoignage redoutable et illustre cette forte dépendance entre les processus financiers, les actifs, les structures, surtout en

période de récession. "When America sneezes, the whole world catches cold".

La modélisation de la dépendance entre plusieurs marchés financiers peut se focaliser soit sur la corrélation positive comme négative entre les actifs, soit sur d'autres moyens plus efficaces, qui déterminent la dépendance des queues de distribution. On peut citer, à titre d'exemple, la théorie des copules.

Cette théorie constitue un moyen efficace pour la modélisation de la dépendance entre différents marchés financiers et afin de mieux comprendre leurs comportements conjoints.

Organisation de la thèse

Cette thèse s'organise en cinq chapitres. Dans un premier temps, on va présenter tout d'abord et de manière succincte les principales définitions et outils de la théorie des copules. Il sera montré que, à l'exception de certaines lois pathologiques, on peut regrouper les copules usuelles en deux groupes appelés les copules elliptiques et les copules archimédiennes. Nous exposerons les critères pour qu'une copule appartienne à l'un de ces groupes. Nous présenterons ensuite les méthodes classiques permettant de décrire le coefficient de dépendance de queue en termes de copule.

Ensuite et au second chapitre, nous donnerons une définition précise du concept de dépendance de queue, en énonçant le fameux théorème de Joe ainsi qu'en présentant quelques propriétés et résultats. Suite à quoi, nous allons mener une étude du coefficient de dépendance de queue pour les trois approches : paramétrique, non paramétrique et semi-paramétrique. Le 3^{ème} chapitre de cette thèse sera intitulé le coefficient de dépendance de queue d'une loi hyperbolique généralisée. Cette partie contient une contribution qui a été inspirée des travaux de [Fung et Seneta \(2010\)](#), [Banachewicz et Van der Vaart \(2008\)](#) et [Schluter and Fischer \(2012\)](#).

Le 4^{ème} chapitre de cette thèse intitulé mesure de dépendance extrême, on va définir le coefficient de dépendance extrême, puis on va exposer quelques propriétés et résultats du coefficient de dépendance extrême dans le cas où les variables aléatoires suivent une loi de Fréchet unitaire.

Dans le 5^{ème} chapitre de cette thèse, nous présenterons un travail indépendant réalisé avec M. Kaddes. Il s'agit d'un système de gestion de bases de données temps réel (SGBDTR), dans un premier temps nous allons réaliser des simulations pour pouvoir déterminer les valeurs optimales à assigner aux paramètres du protocole d'ordonnancement Generalized Earliest Deadline First (GEDF) sous différentes conditions système. Ensuite, nous allons définir un modèle probabiliste pour décrire le comportement des taux de succès des transactions "Utilisateur" sous GEDF. Puis nous allons analyser les performances du système, i.e., les taux de succès, utilisant des transactions emboîtées, sous le protocole GEDF. De la même manière que pour les transactions plates, nous allons mener des études stochastiques pour décrire le comportement des taux de succès des transactions ("Utilisateur "). Et en fin on va pouvoir faire une comparaison entre les transactions plates et emboîtées sous les protocoles d'ordonnancement EDF (Earliest Deadline First) et GEDF qui s'intéresse plus particulièrement à l'approximation de taux de réussite.

COPULES

1. Notion de copule et ses propriétés

1.1. Introduction

Abe Sklar était à l'origine du concept de copules qu'il introduisait, en 1959, comme solution d'un problème de probabilité énoncé par Maurice Fréchet dans le cadre des espaces métriques aléatoires. En revanche, ce concept a connu peu d'utilisation en statistique pendant un certain nombre d'années.

Quelques travaux, entre autres, méritent d'être soulignés dans ce sens :

- Les travaux de Kimeldorf et Sampson sur la dépendance (1975)
- Les recherches de Paul Deheuvels (fin des années 70).

L'étude systématique des copules a débuté au milieu des années 1980 avec Christian Genest et son équipe. L'article fondateur est [Genest et Mackay \(1986\)](#). De nombreux développements statistiques ont été menés par C. Genest et ses coauteurs depuis lors (R.J. MacKay, L.-P. Rivest, P. Capéraà, A.-L. Fougères, K. Ghoudi, B. Rémillard).

Les copules permettent d'introduire et de caractériser des formes très flexibles de dépendance entre différentes variables aléatoires. Elles tiennent considération, de façon meilleure de certains faits en finance (Asymétrie, Dépendance des queues...).

Une copule s'exprime comme une fonction des distributions marginales des variables aléatoires, permettant ainsi de faire le lien entre ces distributions marginales et la distribution bi-variée. Autrement dit, on parle d'une distribution bi-variée dont les distributions marginales sont uniformes.

Les copules furent appelées "fonction de dépendance" par [Deheuvels \(1979\)](#). Elles sont devenues, quelques années plus-tard, un outil incontournable dans la modélisation de la structure de dépendance entre deux ou plusieurs variables aléatoires.

Le développement d'applications des copules n'a cessé d'avoir lieu, depuis les années 90, et ce, dans une multitude de domaines. A titre d'exemple, en Finance [Embrechts *et al.* \(1999\)](#), en Analyse de survie [Lakhal-Chaieb *et al.* \(2008\)](#), en Hydrologie [Favre *et al.* \(2004\)](#), [De Michele *et al.* \(2002\)](#).

Dans un premier lieu, on va proposer une idée générale du concept de la fonction de répartition bi-variée. Ensuite, on va donner une définition et les rapports fondamentaux nécessaires. Pour avoir des informations supplémentaires sur les copules voir [Nelsen \(2006\)](#), [Joe \(1997\)](#) et [Genest et Favre \(2007\)](#).

1.2. Définition d'une copule

Une copule $C : [0, 1]^2 \rightarrow [0, 1]$ est une fonction de répartition dont les lois marginales sont uniforme sur $[0, 1]$.

De manière équivalente, on a la caractérisation suivante

Théorème 1. *Une copule $C : [0, 1]^2 \rightarrow [0, 1]$ est une fonction qui vérifie les trois conditions suivantes*

1. $C(u_1, 0) = C(0, u_2) = 0$ pour tout $u_1, u_2 \in [0, 1]$,
2. $C(u_1, 1) = u_1$ et $C(1, u_2) = u_2$ pour tout $u_1, u_2 \in [0, 1]$,

3. C est une fonction 2-croissante, i.e. pour tout $u_1 \leq u_2$, $v_1 \leq v_2$ et $u_1 \leq v_1$, $u_2 \leq v_2$
- $$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0.$$

1.3. Exemple

Soit $u, v \in I = [0, 1]$, la fonction $M(u, v) = \min(u, v)$

définit une copule. En effet :

- $\forall u, v \in [0, 1], \min(u, 0) = \min(0, u) = 0 \Rightarrow M$ vérifie le point (1) du théorème 1.
- $\forall u, v \in [0, 1], \min(u, 1) = \min(1, u) = u \Rightarrow M$ vérifie le point (2) du théorème 1.
- De même, $\forall u_1, u_2, v_1, v_2 \in [0, 1]$ avec $u_1 \leq u_2$ et $v_1 \leq v_2$ $\min(u_2, v_2) \geq \min(u_2, v_1)$ et $\min(u_1, v_2) \geq \min(u_1, v_1) \Rightarrow M$ vérifie le point (3) du théorème 1.

Si C est la copule associée à un vecteur aléatoire (X_1, X_2) , alors C couple les fonctions de répartition, au sens où

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = C(\mathbb{P}(X_1 \leq x_1), \mathbb{P}(X_2 \leq x_2))$$

Théorème 2. Soit F une fonction de distribution bidimensionnelle dont les marges sont F_1 et F_2 . Alors F admet une représentation copule : $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$. La copule C est unique si les marges sont continues.

Il est aussi possible de coupler les fonctions de survie, au sens où il existe une copule C telle que $\mathbb{P}(X > x, Y > y) = \hat{C}(\mathbb{P}(X > x), \mathbb{P}(Y > y))$. On montre aisément que $\hat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$.

1.4. Définition

La copule de survie $\hat{C}$ associée à la copule C est la copule définie par

$$\hat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

À partir d'une fonction copule C , nous pouvons construire trois autres copules :

- $\bar{C}(u_1, u_2) = 1 - u_1 - u_2 + C(u_1, u_2)$
- $\tilde{C}(u_1, u_2) = u_1 + u_2 - C(u_1, u_2)$
- $C^*(u_1, u_2) = 1 - C(1 - u_1, 1 - u_2)$.

1.5. Propriétés des copules

Théorème 3. Soit C une copule $\forall v \in [0, 1]$, la dérivée partielle $\frac{\partial C}{\partial u}$ existe presque sûre pour tout $u \in [0, 1]$, si la dérivée existe on a :

$$0 \leq \frac{\partial}{\partial u} C(u, v) \leq 1$$

De même pour tout $u \in [0, 1]$, la dérivée partielle $\frac{\partial C}{\partial v}$ existe pour presque tout $v \in [0, 1]$, si la dérivée existe, on a :

$$0 \leq \frac{\partial}{\partial v} C(u, v) \leq 1$$

De plus, les fonctions $u \mapsto \frac{\partial}{\partial v} C(u, v)$ et $v \mapsto \frac{\partial}{\partial u} C(u, v)$ sont définies et croissantes presque partout sur $[0, 1]$.

Théorème 4. Soit C une copule bivariée, pour tout $u_1, u_2, v_1, v_2 \in [0, 1]$ avec $u_1 \leq u_2$ et $v_1 \leq v_2$, on a :

$$|C(u_1, v_1) - C(u_2, v_2)| \leq |u_2 - u_1| + |v_2 - v_1|, \text{ (cf [Nelsen \(2006\)](#))}.$$

1.6. Corollaire

Soit F une fonction de répartition bi-dimensionnelle de fonctions de répartition marginales continues F_1, F_2 . Alors, la fonction copule C associée à F est donnée par :

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right), \text{ pour tout } u_1, u_2 \in I.$$

1.7. Exemple

La fonction de répartition de la distribution logistique bi-variée de Gumbel est donnée par : $F(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}$ qui est définie sur $\mathbb{R}^2$. Nous pouvons montrer que les fonctions de répartition marginales sont

$$F(x_1) = (1 + e^{-x_1})^{-1}$$

et

$$F(x_2) = (1 + e^{-x_2})^{-1}.$$

Les fonctions quantiles sont donc

$$F_1^{-1}(u_1) = \log(u_1) - \log(1 - u_1)$$

et

$$F_2^{-1}(u_2) = \log(u_2) - \log(1 - u_2).$$

Nous en déduisons que la fonction copule est

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)) = \left(1 + \frac{1 - u_1}{u_1} + \frac{1 - u_2}{u_2}\right)^{-1} = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}.$$

Cette fonction copule est appelée la fonction copule logistique de Gumbel.

Les fonctions copules $\hat{C}$, $\tilde{C}$ et C^* sont appelées, respectivement, la fonction copule de survie (survival copula), la fonction copule duale (dual copula) et la co-copule (co-copula).

Soient X_1 et X_2 deux variables aléatoires. Les interprétations probabilistes sont les suivantes :

$$P[X_1 \leq x_1, X_2 \leq x_2] = C(F_1(x_1), F_2(x_2))$$

$$P[X_1 \geq x_1, X_2 \geq x_2] = \bar{C}(\bar{F}_1(x_1), \bar{F}_2(x_2))$$

$$P[X_1 \leq x_1 \cup X_2 \leq x_2] = \tilde{C}(F_1(x_1), F_2(x_2))$$

$$P[X_1 \geq x_1 \cup X_2 \geq x_2] = C^*(\bar{F}_1(x_1), \bar{F}_2(x_2))$$

avec $\bar{F}(x) = 1 - F(x)$ qui désigne la fonction de survie.

1.8. Les copules usuelles

Parmi les copules usuelles, nous citons dans cette section les copules usuelles suivantes : La Copule d'indépendance, la Copule Maximale, la Copule Minimale.

1.8.1. La Copule d'indépendance

La copule d'indépendance (on parlera aussi de copule produit) est définie par

$$\Pi(u_1, u_2) = u_1 \cdot u_2$$

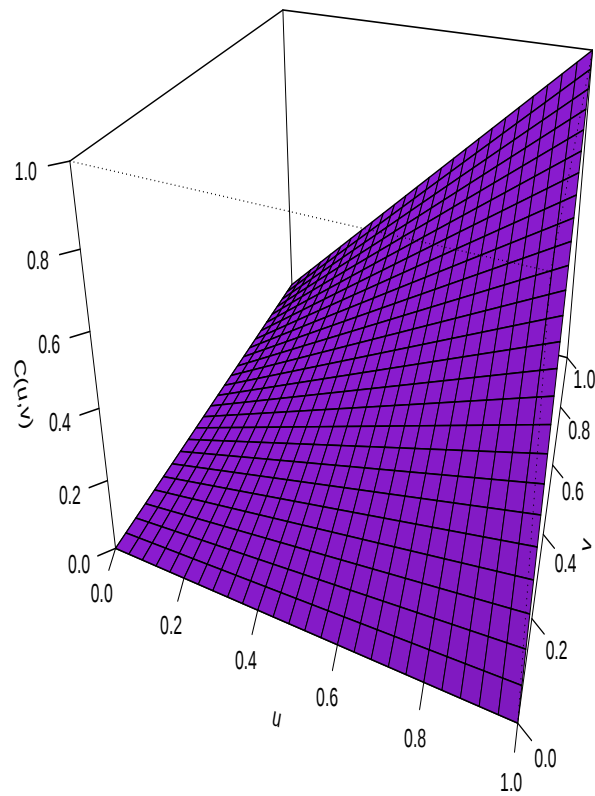


Figure 1.1. *Copule d'indépendance Π*

1.8.2. La Copule maximale

La copule maximale (on parlera aussi de copule comonotone, de dépendance positive, born sup Fréchet) est donc définie comme suit :

$$M(u_1, u_2) = \min\{u_1, u_2\}$$

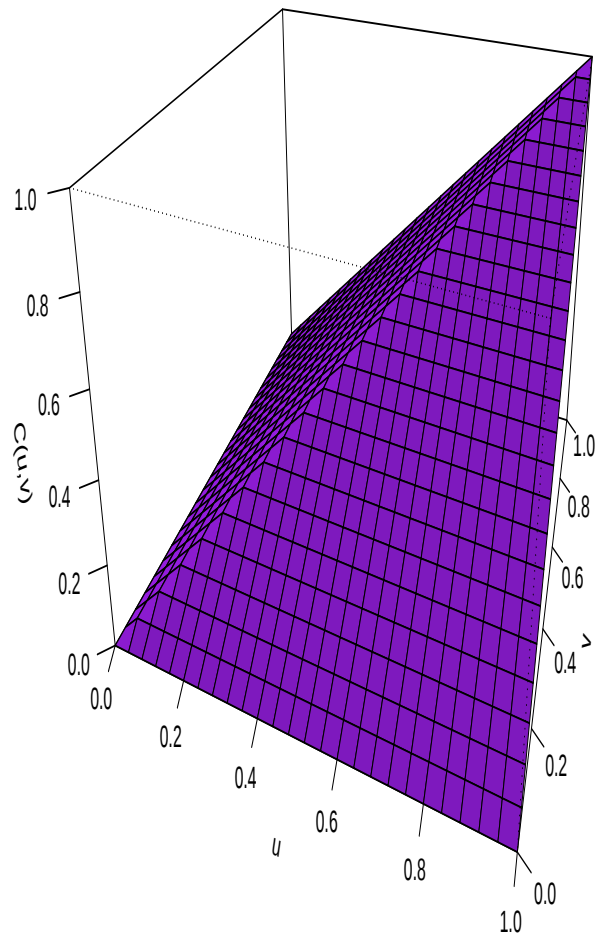


Figure 1.2. *Copule Maximale M*

1.8.3. La Copule minimale

La copule minimale (Copule antimonotone ou de dépendance négative).
 Cette copule se définit donc ainsi :

$$W(u_1, u_2) = \max\{0, u_1 + u_2 - 1\}$$

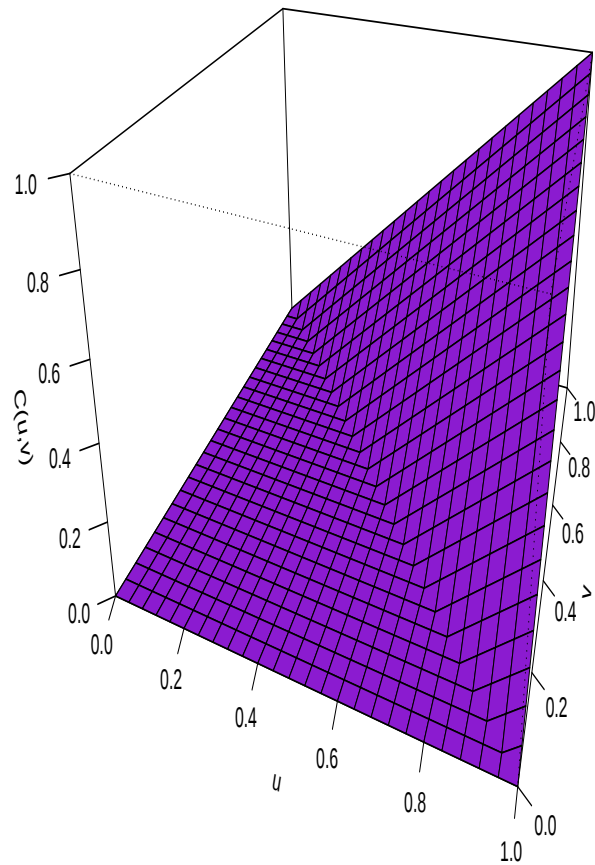


Figure 1.3. *Copule Minimale W*

Théorème 5 (Bornes de Fréchet-Hoeffding). *Soit G une fonction de répartition conjointe d'un vecteur aléatoire (X_1, X_2) de fonctions de répartition marginales F_1 et F_2 . Pour toute copule C associée à G on a*

$$\forall u_1, u_2 \in I, \quad W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2).$$

2. Les copules paramétriques

2.1. La copule Elliptique

Les copules elliptiques sont les copules associées aux distributions elliptiques, dans ce qui suit, nous donnons quelques définitions de la distribution elliptique bi-variée ainsi que les deux exemples les plus connus de cette famille, la copule Gaussienne et la copule de Student.

2.1.1. Définition

Un couple de variables aléatoires $X = (X_1, X_2)$ est de distribution elliptique s'il admet la représentation suivante :

$$X = \mu + R\Sigma^{-1/2}U$$

Où

- μ est la moyenne, si elle existe,
- U un vecteur aléatoire uniformément distribuée dans la sphère unité de $\mathbb{R}^2$,
- Σ une matrice de variance-covariance,
- R un vecteur aléatoire positive indépendante de U .

La fonction de densité d'une distribution elliptique est défini pour tout $x \in \mathbb{R}^2$, par

$$f(x) = |\Sigma|^{-1/2} \cdot g((x - \mu)^T \Sigma^{-1}(x - \mu))$$

tel que g est un générateur défini de $\mathbb{R}^+ \rightarrow \mathbb{R}^+$.

2.1.2. La Copule Gaussienne

Une copule gaussienne est une mesure de dépendance entre deux variables aléatoires. Cette copule est définie à partir de famille des loi gaussienne.

2.1.2.1. Définition Soient $\rho \in]-1, 1[$ le coefficient de corrélation, Φ la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$ et Φ_ρ la distribution normale bi-variée de matrice de variance covariance Σ (matrice de corrélation dans ce cas) associé à ρ , la copule gaussienne C_ρ est définie par

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

Où

$$\Phi_\rho(u, v) = \int_{-\infty}^u \int_{-\infty}^v \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 + t^2 - 2\rho st}{2(1-\rho^2)}\right) ds dt$$

Φ^{-1} est l'inverse de la fonction de répartition de la loi normale centrée réduite.

2.1.2.2. Remarques

- Le cas où $\rho = 0$ correspond au cas d'indépendance et la copule $C_0 = \Pi$.
- Si $\rho = -1$, alors $C_{-1} = W$.
- Si $\rho = 1$, alors $C_1 = M$.

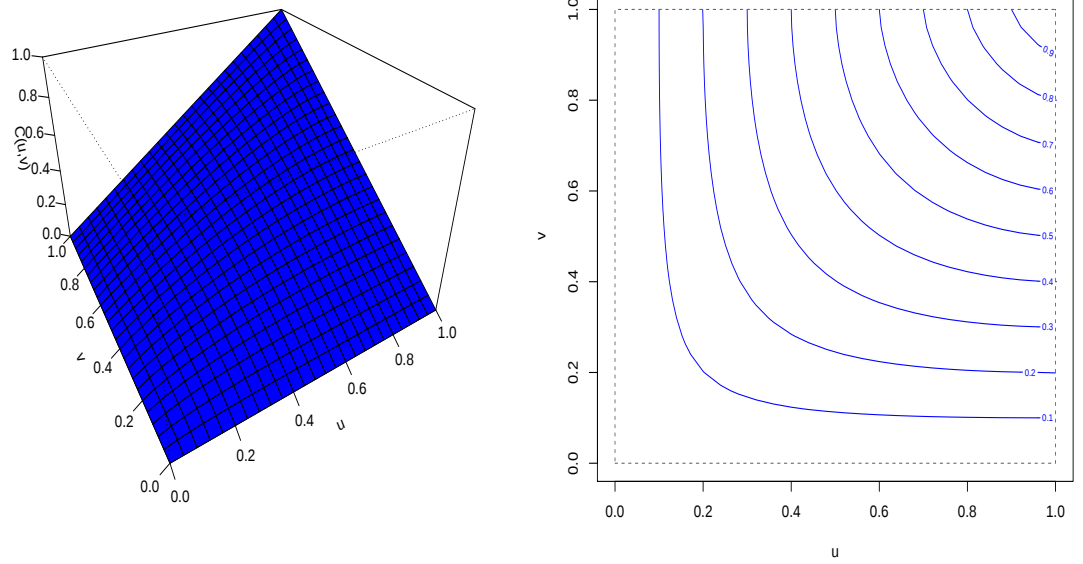


Figure 1.4. Graphique de la copule gaussienne et son contour avec $\rho = 0.6$

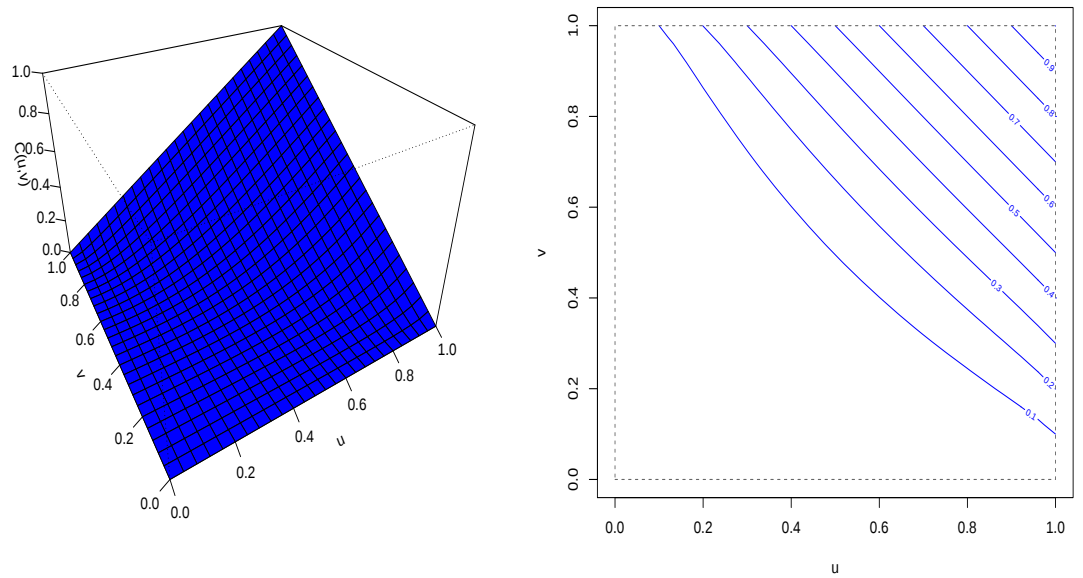


Figure 1.5. Graphique de la copule gaussienne et son contour avec $\rho = -0.8$

Théorème 6. *La densité d'une copule gaussienne s'écrit sous la forme*

$$c_\rho = \frac{1}{\sqrt{(1-\rho^2)}} \exp \left(-\frac{\Phi^{-1}(u)^2 + \Phi^{-1}(v)^2 - 2\rho\Phi^{-1}(u)\Phi^{-1}(v)}{2(1-\rho^2)} + \frac{\Phi^{-1}(u)^2 + \Phi^{-1}(v)^2}{2} \right).$$

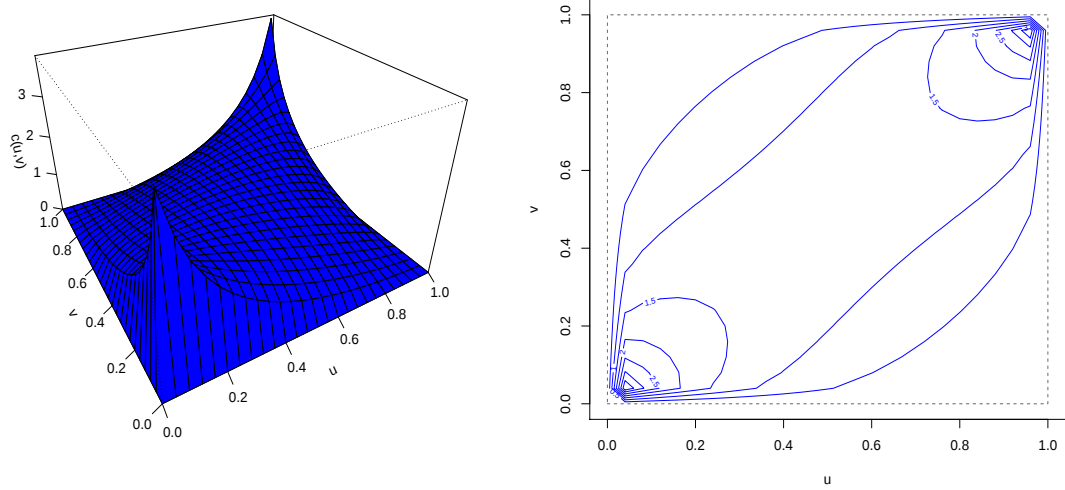


Figure 1.6. *Densité de la copule gaussienne et son contour avec $\rho = 0.6$*

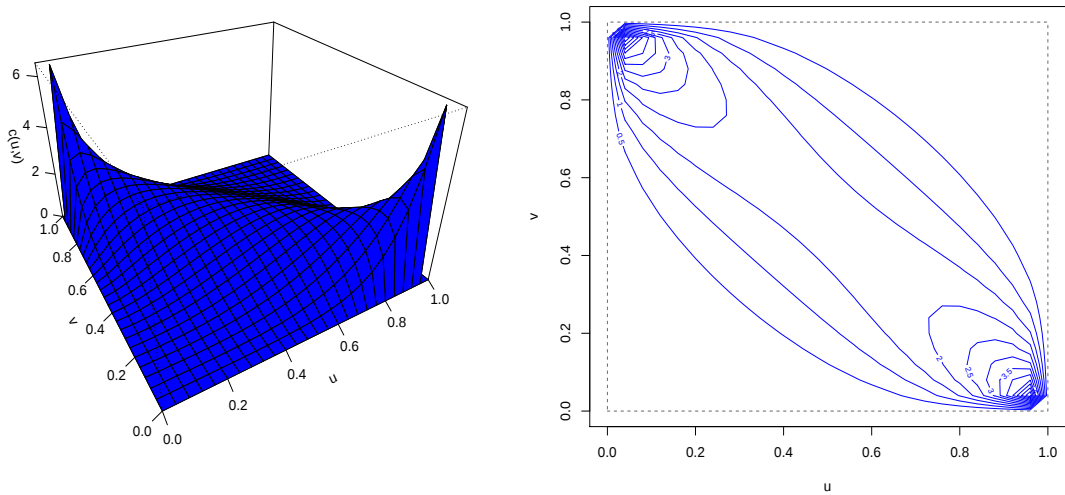


Figure 1.7. *Densité de la copule gaussienne et son contour avec $\rho = -0.8$*

2.1.3. La Copule de Student

Cette copule est construite de la même manière que la copule gaussienne. La copule de Student est la fonction de dépendance associée à la distribution de Student.

2.1.3.1. Définition La copule de Student est une copule paramétrée par le coefficient de corrélation ρ et de degré de liberté ν . Cette copule est donnée par

$$C_{\rho,\nu}(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{\Gamma(\frac{\nu+2}{2})}{2\pi\Gamma(\frac{\nu}{2})\sqrt{1-\rho^2}} \times \left(1 + \frac{s^2 + t^2 - 2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds dt.$$

Où t_{ν} est la fonction de répartition de la loi de Student à ν degrés de libertés, i.e.

$$t_{\nu}(x) = \int_{-\infty}^x \frac{\Gamma(\frac{\nu+2}{2})}{\sqrt{\rho\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt$$

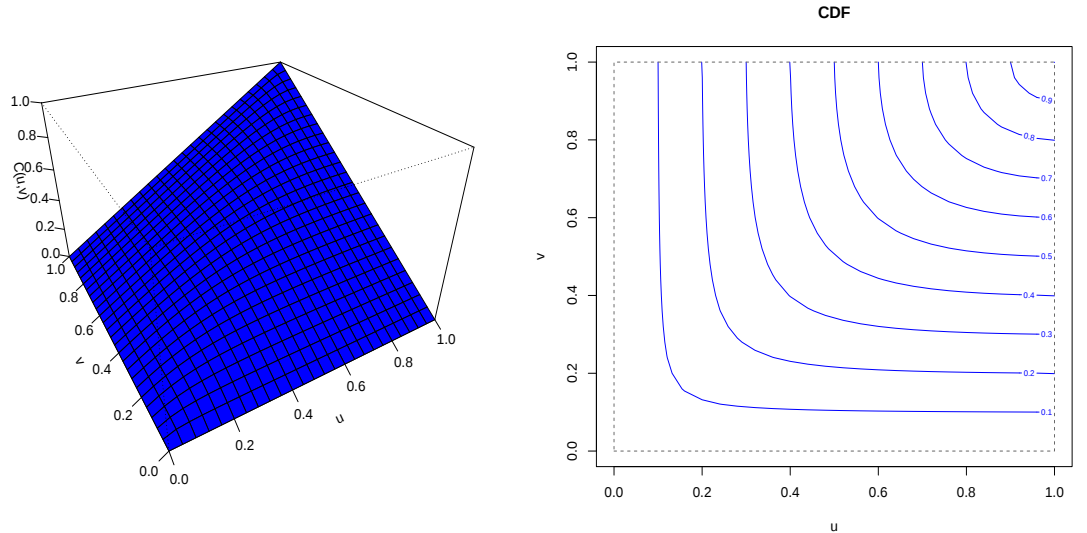


Figure 1.8. Graphique de la copule de Student $\rho = 0.8$ et $\nu = 4$

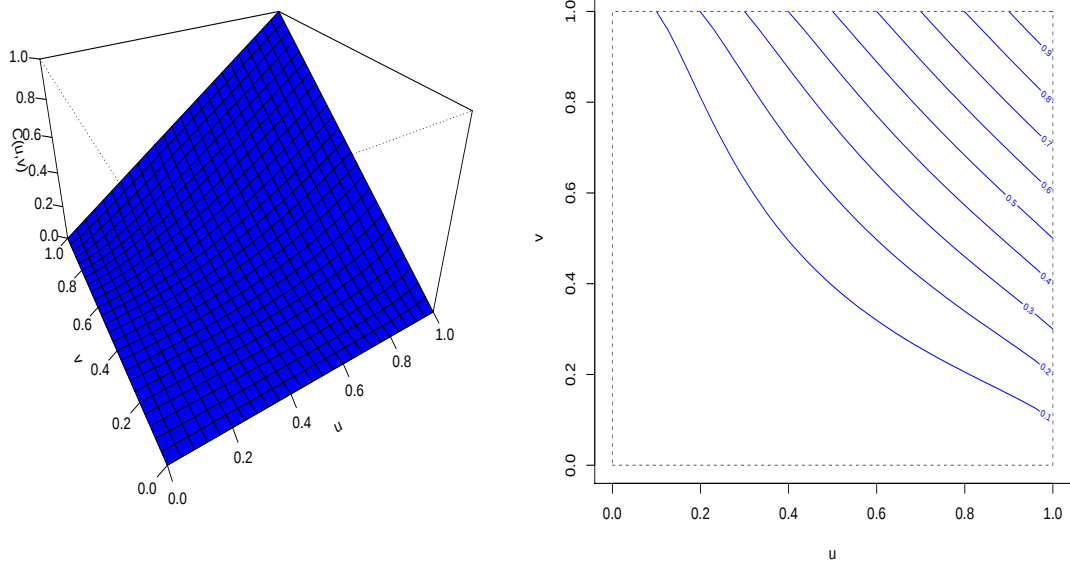


Figure 1.9. Graphique de la copule de Student $\rho = -0.6$ et $\nu = 4$

2.1.3.2. Remarques Si le degré de liberté $\nu \rightarrow +\infty$, alors la copule de Student converge vers la copule gaussienne.

Théorème 7. La densité d'une copule Student est donnée par

$$c_{\rho,\nu}(u, v) = \frac{1}{\sqrt{\rho}} \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{1+\nu}{2})^2} \times \frac{\left(1 + \frac{t_1^2 + t_2^2 - 2\rho t_1 t_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}}{\prod_{i=1}^2 \left(1 + \frac{t_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}}.$$

avec $t_1 = t_\nu^{-1}(u)$ et $t_2 = t_\nu^{-1}(v)$.

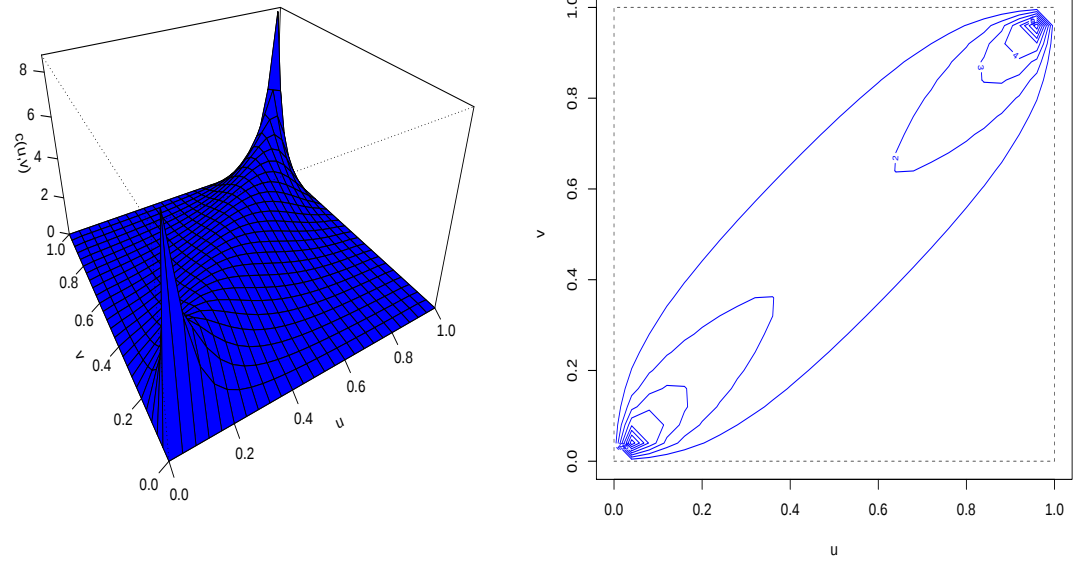


Figure 1.10. Densité de la copule de Student et son contour avec $\rho = 0.8$ et $\nu = 4$

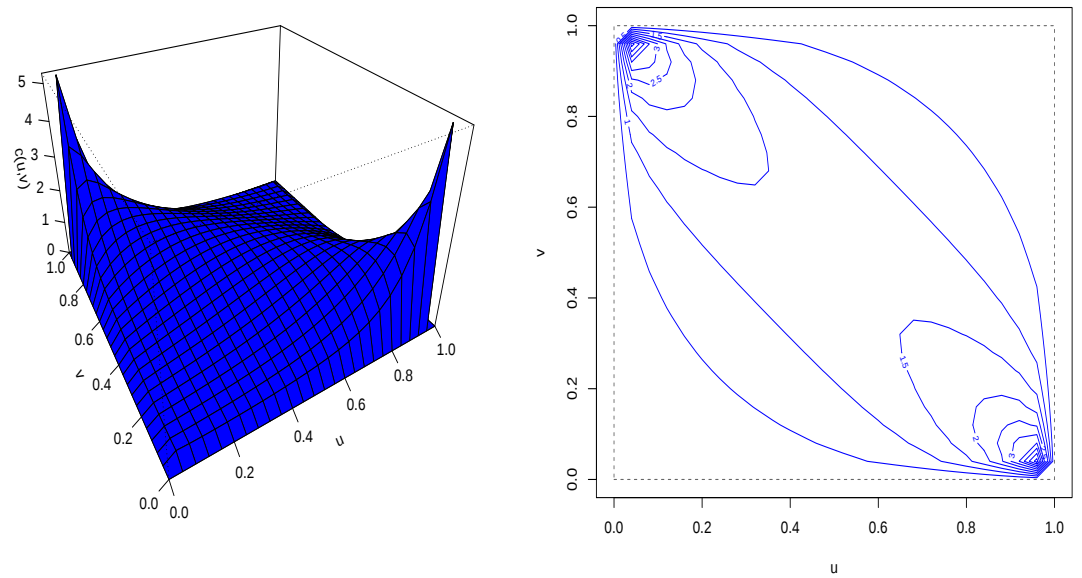


Figure 1.11. Densité de la copule de Student et son contour avec $\rho = -0.6$ et $\nu = 4$

2.2. Les copules Archimédiennes

Les copules archimédiennes ont l'avantage de décrire des structures de dépendance diverses dont notamment les dépendances asymétriques, où les coefficients de queue inférieure et de queue supérieure diffèrent. Les copules les plus utilisés de cette famille sont les copule de Frank, de Clayton et de Gumbel, définies dans ce qui suit

2.2.1. Définition

Soit ϕ une fonction décroissante convexe sur $(0, 1] \rightarrow [0, \infty]$ telle que $\phi(1) = 0$ et $\phi(0) = +\infty$. On appellera copule Archimédienne stricte de générateur ϕ la copule définie par

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), \quad u, v \in [0, 1].$$

2.2.2. Copule de Frank

Si $\phi(x) = -\log\left(\frac{e^{-ax} - 1}{e^{-a} - 1}\right)$ avec $a \neq 0$. On obtient ainsi la copule de Frank qui est donc définie comme suit :

$$C_a(u, v) = -\frac{1}{a} \log\left(1 + \frac{(e^{-au} - 1)(e^{-av} - 1)}{e^{-a} - 1}\right)$$

2.2.2.1. Remarques Les cas limites sont donc les suivants :

- $\lim_{a \rightarrow +\infty} C_a = M$
- $\lim_{a \rightarrow 0} C_a = \Pi$
- $\lim_{a \rightarrow -\infty} C_a = W$

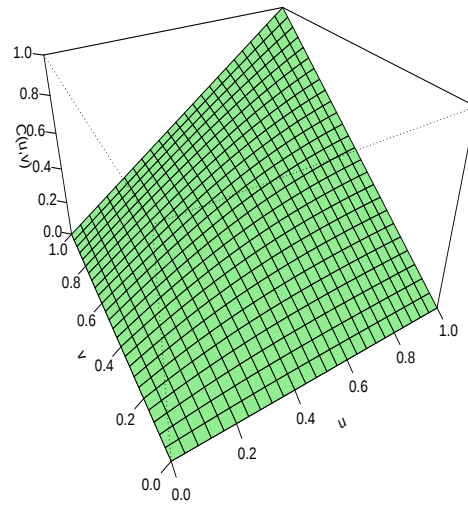


Figure 1.12. Graphique de la copule de Frank

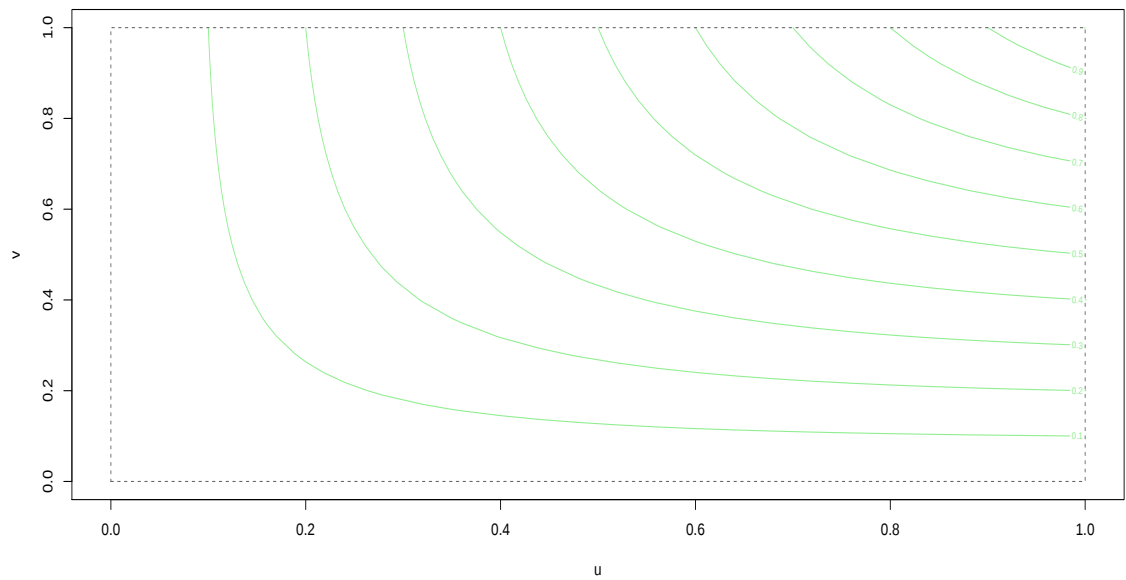


Figure 1.13. Contour de copule Frank

Théorème 8. *La fonction de densité d'une copule de Frank est définie pour tout $a \in \mathbb{R}^*$, par*

$$c_a(u, v) = \frac{a - 1}{\log a^{u+v}} (a - 1 + (a^u - 1)(a^v - 1))^2$$

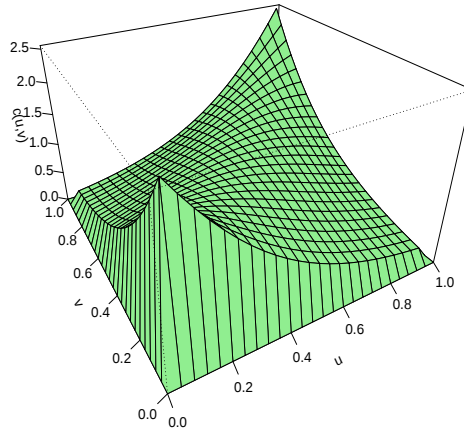


Figure 1.14. *Densité de la copule Frank*

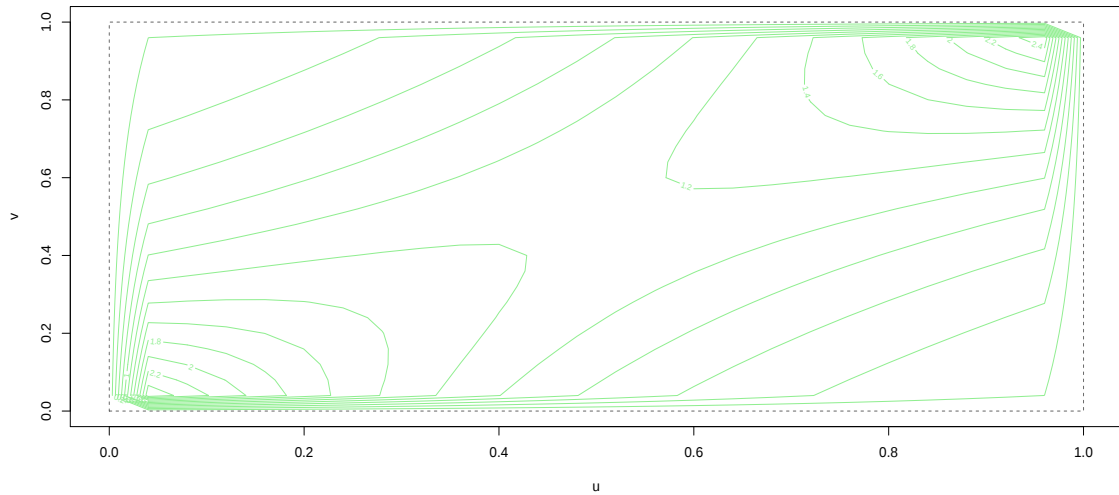


Figure 1.15. *Contour de la copule Frank*

2.2.3. Copule de Clayton

Si $\phi(t) = a \left(t^{-\frac{1}{a}} - 1 \right)$ avec $a > 0$. On obtient ainsi la copule de Clayton qui est donc définie comme suit :

$$C_a(u, v) = \left(u^{-\frac{1}{a}} + v^{-\frac{1}{a}} - 1 \right)^{-a}$$

2.2.3.1. Remarques Les cas limites sont donc les suivants :

- $\lim_{a \rightarrow -\infty} C_a = W$
- $\lim_{a \rightarrow 0} C_a = \Pi$

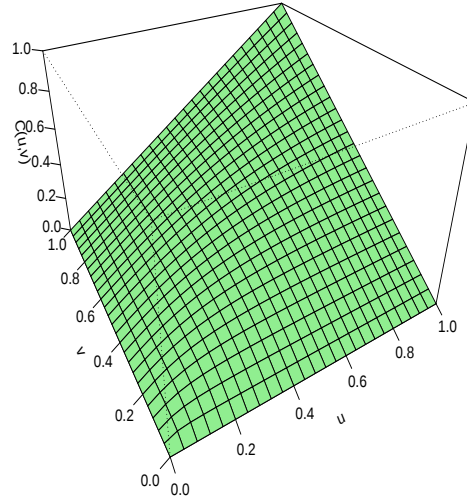


Figure 1.16. Graphique de la copule de Clayton

Théorème 9. La fonction de densité d'une copule de Clayton est définie pour tout $a \in \mathbb{R}_+^*$, par

$$c_a(u, v) = (a + 1)(u.v)^{-a-1} \left(u^{-a} + v^{-a} - 1 \right)^{-\frac{1}{a}-2}$$

1. Copules

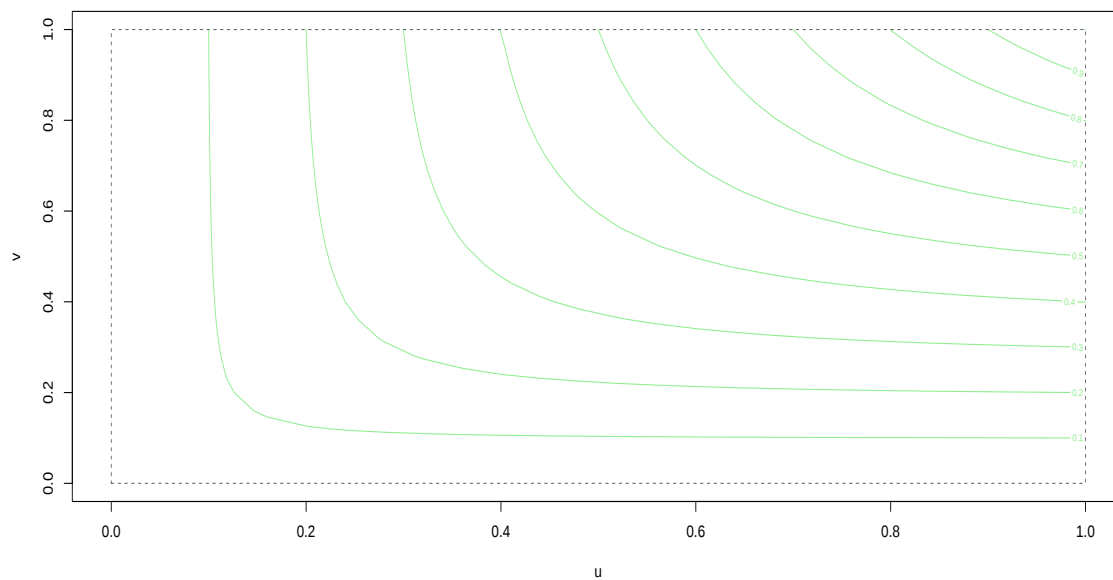


Figure 1.17. *Contour de la copule de Clayton*

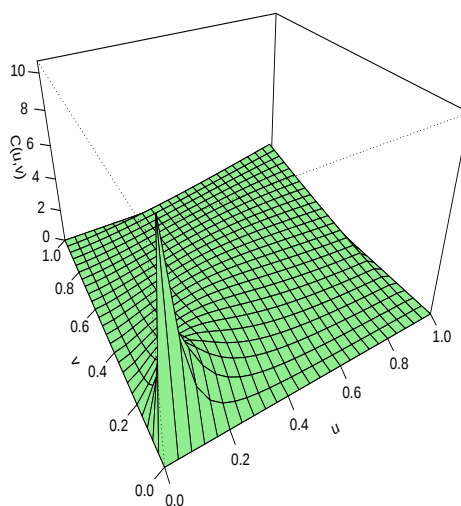


Figure 1.18. *Densité de la copule de Clayton*

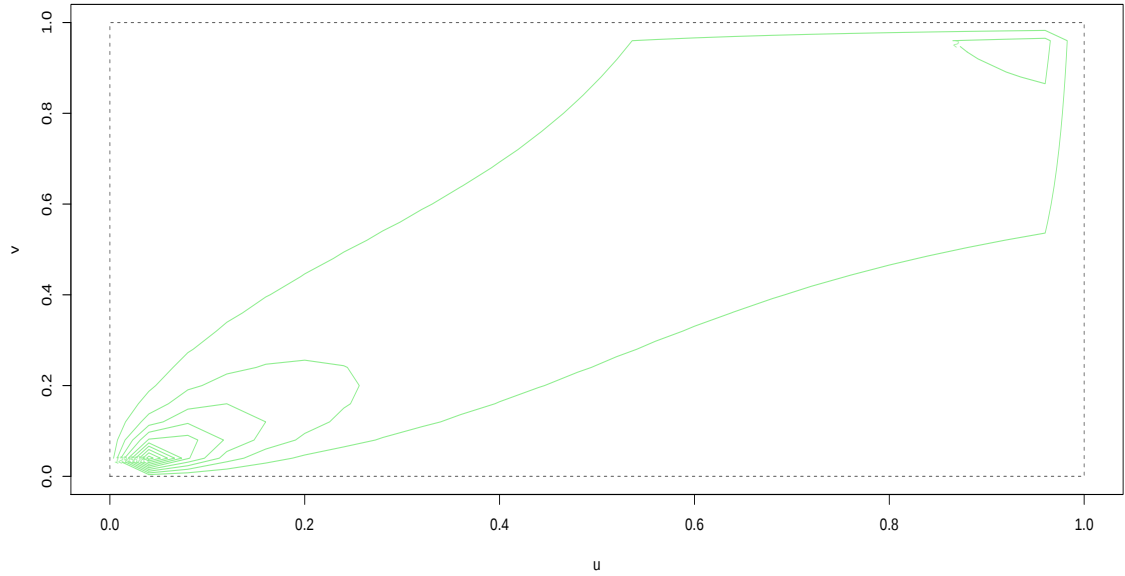


Figure 1.19. *Contour de la copule Clayton*

2.2.4. Copule de Gumbel

En prenant comme générateur $\phi(x) = (-\log(t))^a$ avec $a > 1$. La fonction ainsi définie satisfait à les conditions du théorème sur les copules Archimédiennes, ce qui permet générer la copule de Gumbel en prenant :

$$C_a(u, v) = \exp \left(-(-\log(u))^a + (-\log(v))^a \right)^{\frac{1}{a}}$$

2.2.4.1. Remarques Les cas limites sont donc les suivants :

- $\lim_{a \rightarrow +\infty} C_a = M$
- $\lim_{a \rightarrow 1} C_a = \Pi$

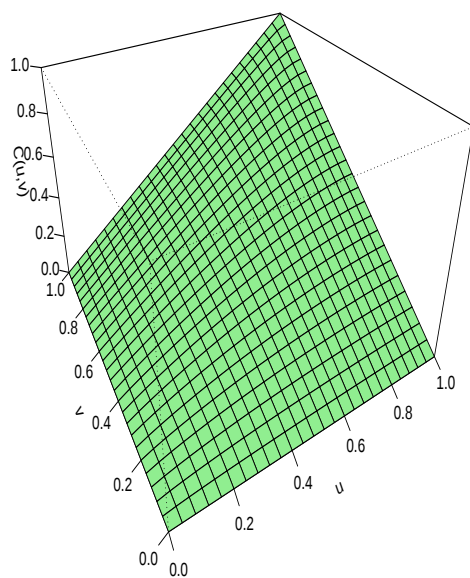


Figure 1.20. *Graphique de la copule de Gumbel*

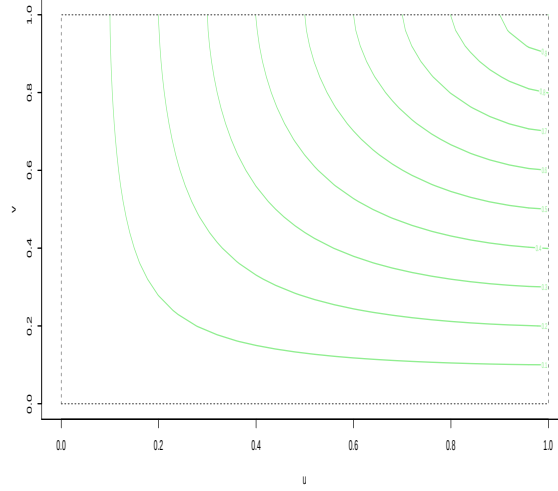


Figure 1.21. Contour de la copule de Gumbel

Théorème 10. La fonction de densité d'une copule de Gumbel est définie pour tout $a > 1$, par

$$c_a(u, v) = C_a(u, v) [\phi_a(u) + \phi_a(v)]^{\frac{1}{a}-2} \left[a - 1 + \left(\phi_a(u) + \phi_a(v) \right)^{\frac{1}{a}} \right] \frac{\phi_{a-1}(u) \phi_{a-1}(v)}{uv}$$

où

$$\phi_a(t) = (-\log(t))^a \text{ comme fonction génératrice avec } a \geq 1 \text{ et } t \in [0, 1].$$

2.3. La copule des valeurs extrêmes

2.3.1. Définition

On appelle copule de valeurs extrêmes toute copule C vérifiant

$$C(u^t, v^t) = C^t(u, v), \quad \forall t > 0 \quad u, v \in [0, 1]$$

Théorème 11. Pour toute copule des valeurs extrêmes bi-varées C il existe

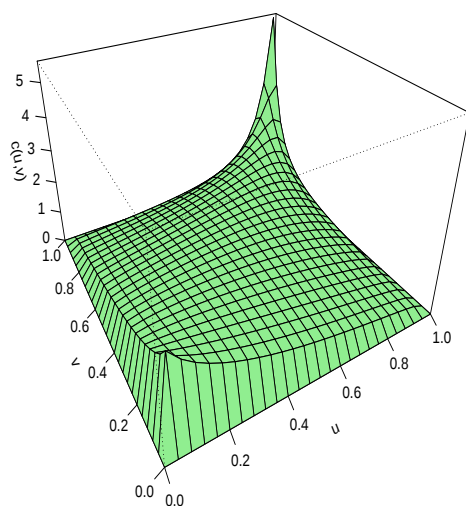


Figure 1.22. *Densité de la copule Gumbel*

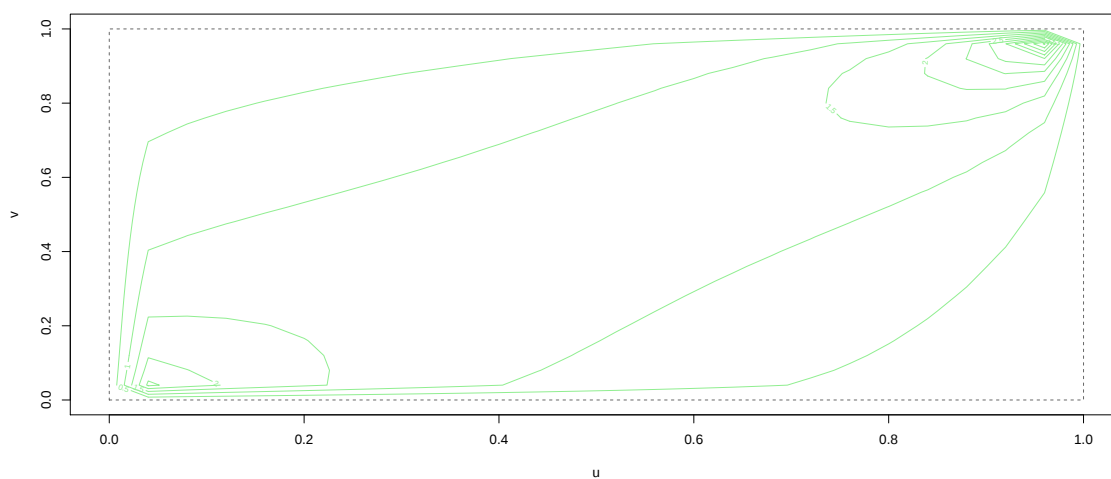


Figure 1.23. *Contour de la copule Gumbel*

une fonction convexe A définie de $[0, 1]$ dans $[\frac{1}{2}, 1]$ telle que

$$C(u, v) = \exp \left\{ [\log(u) + \log(v)] A \left(\frac{\log(u)}{\log(u) + \log(v)} \right) \right\}.$$

Où $\max\{t, 1-t\} \leq A(t) \leq 1$. (cf [Salvadori et al. \(2002\)](#))

Considérons la copule de Galambos :

$$C_\theta(u, v) = uv \exp(-[(-\log(u))^\theta + (-\log(v))^\theta]^{-\frac{1}{\theta}}), \quad \theta \geq 0$$

On a :

$$C_\theta(u^t, v^t) = u^t v^t \exp(-t[(-\log(u))^\theta + (-\log(v))^\theta]^{-\frac{1}{\theta}}) = C_\theta^t(u, v), \quad \forall t > 0$$

Donc est une copule bi-variée de valeurs extrêmes.

Famille	C_α	$A_\alpha(t)$
Gumbel1	$\exp(-(u_1 + \log v_1^\alpha)^{\frac{1}{\alpha}}), \alpha \geq 1$	$(t^\alpha + (1-t^\alpha))^{\frac{1}{\alpha}}$
Gumbel2	$uv \exp(\alpha \frac{u_1 v_1}{u_1 + v_1}), \alpha \geq 0$	$t^2 - \alpha t + 1$
Indépendance	uv	1
Galambos	$uv \exp(-(u_1^{-\alpha} + v_1^{-\alpha})^{\frac{1}{\alpha}})$	$1 - (t^\alpha + (1-t^\alpha))^{\frac{1}{\alpha}}$

Table 1.1. Quelques copules de valeurs extrêmes

Avec $u_1 = \log(u), v_1 = \log(v)$.

2.3.1.1. Remarque La copule de Gumbel est la seule copule qui appartient à la famille de copules archimédiennes et à la famille de copules de valeurs extrêmes.

CHAPITRE

DÉPENDANCE DE QUEUE DE DISTRIBUTION

1. Introduction

Une des premières mesures de dépendance introduite dans l'étude des statistiques est probablement le coefficient de corrélation de Pearson, une mesure de dépendance linéaire entre deux variables aléatoires quantitatives. Par contre, lorsque nous travaillons avec des variables aléatoires continues non normales, l'utilisation de notions basées sur la linéarité n'est pas toujours appropriée.

De plus, contrairement au coefficient de corrélation de Pearson, certaines mesures sont invariantes sous transformations strictement croissantes et valent 1 pour la borne supérieure de Fréchet et -1 pour la borne inférieure de Fréchet.

Le coefficient de corrélation de Pearson ne mesure que la dépendance linéaire. L'utilisation de cette mesure est alors naturelle dans le cas de distributions elliptiques bidimensionnelles comme la distribution normale et la distribution t . Mais quand la distribution conjointe de deux variables aléa-

toires ne suit pas une loi elliptique, l'utilisation du coefficient de corrélation de Pearson n'est pas adéquate et mène à des conclusions erronées. Plusieurs mesures alternatives peuvent alors être utilisées [Nelsen \(2006\)](#) notamment le tau de Kendall et le rho de Spearman [Kendall et Stuart \(1963\)](#), [Kruskal \(1958\)](#), [Hollander et Wolfe \(1973\)](#) et [Lehmann \(1973\)](#).

2. Notion de la concordance

2.1. Définition

Notons (x_1, y_1) et (x_2, y_2) deux observations d'un vecteur aléatoire continu (X, Y) , alors (x_1, y_1) et (x_2, y_2) sont dites : concordantes si :

$$(x_1 - x_2)(y_1 - y_2) > 0 \Leftrightarrow (x_1 < x_2 \text{ et } y_1 < y_2) \text{ ou } (x_1 > x_2 \text{ et } y_1 > y_2).$$

Discordantes si :

$$(x_1 - x_2)(y_1 - y_2) < 0 \Leftrightarrow (x_1 < x_2 \text{ et } y_1 > y_2) \text{ ou } (x_1 > x_2 \text{ et } y_1 < y_2).$$

3. Le tau de Kendall

3.1. Définition

Soient (X_1, X_2) et (Y_1, Y_2) deux paires indépendantes de variables aléatoires, ayant pour fonctions de répartition marginales communes F (pour X_1 et X_2) et G (pour Y_1 et Y_2). Le tau de Kendall se définit comme suit :

$$\tau = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]. \quad (2.1)$$

Autrement dit, il s'agit de la différence entre la probabilité de concordance et la probabilité de non concordance. Son expression en terme de copule est la suivante :

3.2. Théorème

Soit (X, Y) deux variables aléatoires, ayant pour fonction de répartition jointe H . Soit C la copule associée à la fonction de répartition jointe H . Alors le tau de Kendall est défini par :

$$\tau = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1$$

4. Le rho de Spearman

La valeur de ce coefficient dénote par ρ_S , est équivalente au coefficient de corrélation de Pearson. Il a été développée par [Spearman \(1904\)](#). La définition de cette mesure est la suivante :

4.1. Définition

Soient (X_1, Y_1) , (X_2, Y_2) et (X_3, Y_3) trois paires indépendantes de variables aléatoires de même fonction de répartition jointe H . Le coefficient de corrélation de Spearman se définit comme suit :

$$\rho_S = 3\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0]. \quad (2.2)$$

4.2. Théorème

Soit (X, Y) deux variables aléatoires continues de copule C . Le rho de Spearman de (X, Y) est défini par :

$$\rho_S = 12 \int_{I^2} uv dC(u, v) - 3$$

5. Dépendance de queue de distribution

Le coefficient de dépendance de queue mesure la probabilité conditionnelle que des valeurs extrêmes se réalisent simultanément au niveau des queues

de distribution, d'un point de vue financier la probabilité que l'actif X subit une perte plus grande que X_q , associée au quantile q tendant vers zéro, conditionnée à la réalisation d'une perte de l'actif Y plus grande que Y_q associée au même quantile q , Joe (1997). Cette notion de dépendance de queue est l'approche la plus adéquate pour estimer la survenue des risques simultanés au niveau des queues de distribution et elle permet de modéliser l'occurrence concomitante des valeurs extrêmes.

5.1. Définition

Le coefficient de dépendance de queue supérieure de deux variables aléatoires continues X et Y ayant pour fonction de distribution respectives F_X et F_Y , est défini par la limite (si elle existe) :

$$\lambda_U = \lim_{u \rightarrow 1} \mathbb{P}(X > F_X^{(-1)}(u) | Y > F_Y^{(-1)}(u))$$

où $F^{-1}(t) = \inf\{x | F(x) \geq t\}$

et

$$\lambda_L = \lim_{u \rightarrow 0} \mathbb{P}(X \leq F_X^{(-1)}(u) | Y \leq F_Y^{(-1)}(u))$$

Donc λ_U représente la probabilité qu'une variable aléatoire dépasse une valeur u , sachant que l'autre variable est également au-dessus de u , dans la limite de u proche de 1.

Et de la même manière, on définit le coefficient de dépendance de queue inférieure dans la limite des quantiles extrêmes proche de 0. On dit que X et Y sont asymptotiquement dépendantes au niveau supérieur de la queue de distribution si $u \in [0, 1]$ et asymptotiquement indépendantes au niveau supérieur de la queue de distribution si $u = 0$. On peut exprimer λ_U et λ_L en terme de la copule C .

5.2. Proposition

Étant données deux variables aléatoires X et Y de copule C , on a :

$$\lambda_L = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}$$

$$\lambda_U = \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u} = \lim_{u \rightarrow 1} \frac{\bar{C}(u, u)}{1 - u}$$

En effet, par définition, on a :

$$\lambda_L = \lim_{u \rightarrow 0} \mathbb{P}(X \leq F_X^{-1}(u) | Y \leq F_Y^{-1}(u)) = \lim_{u \rightarrow 0} \frac{\mathbb{P}(X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(u))}{\mathbb{P}(Y \leq F_Y^{-1}(u))} = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}$$

Ainsi comme pour les copules ce coefficient est invariant par rapport à des transformations croissantes. Cela nous permet d'envisager une autre définition possible pour deux variables aléatoires uniformes : $\mathbb{P}(V \leq v | U = u) = \frac{\partial}{\partial u} C(u, v)$. D'où : $\lambda_U = \lim_{u \rightarrow 1} (\mathbb{P}(V > u | U = u) + \mathbb{P}(U > u | V = u))$

Si la copule est symétrique

$$\lambda_U = 2 \lim_{u \rightarrow 1} (\mathbb{P}(V > u | U = u))$$

De même, on peut écrire :

$$\begin{aligned} \lambda_U &= \lim_{u \rightarrow 1} \mathbb{P}(X > F_X^{-1}(u) | Y > F_Y^{-1}(u)) \\ &= \lim_{u \rightarrow 1} \frac{\mathbb{P}(X > F_X^{-1}(u), Y > F_Y^{-1}(u))}{\mathbb{P}(Y > F_Y^{-1}(u))} \\ &= \lim_{u \rightarrow 1} \frac{1 - \mathbb{P}(X \leq F_X^{-1}(u) \text{ ou } Y \leq F_Y^{-1}(u))}{1 - \mathbb{P}(Y \leq F_Y^{-1}(u))} \\ &= \lim_{u \rightarrow 1} \frac{1 - (\mathbb{P}(X \leq F_X^{-1}(u) + \mathbb{P}(Y \leq F_Y^{-1}(u) - \mathbb{P}(X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(u)))}{1 - \mathbb{P}(Y \leq F_Y^{-1}(u))} \\ &= \lim_{u \rightarrow 1} \frac{1 - (u + u - C(u, u))}{u} \end{aligned}$$

5.3. Exemple

On considère la copule d'indépendance donnée par :

$$\Pi(u, v) = uv,$$

nous avons

$$\begin{aligned} \lambda_U &= \lim_{u \rightarrow 1} \frac{1 - 2u + u^2}{1 - u} \\ &= \lim_{u \rightarrow 1} (1 - u) \\ &= 0 \end{aligned}$$

et

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0} \frac{u^2}{u} \\ &= \lim_{u \rightarrow 0} (u) \\ &= 0\end{aligned}$$

5.4. Exemple

Soit la copule minimale, définit par :

$$W(u, v) = \min(u, v).$$

On obtient

$$\begin{aligned}\lambda_U &= \lim_{u \rightarrow 1} \frac{1 - 2u + \min(u, u)}{1 - u} \\ &= \lim_{u \rightarrow 1} \frac{1 - 2u + u}{1 - u} \\ &= 1\end{aligned}$$

et

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0} \frac{\min(u, u)}{u} \\ &= \lim_{u \rightarrow 0} 1 \\ &= 1\end{aligned}$$

5.5. Exemple

Soit la copule maximale, donnée par :

$$M = \max(u + v - 1, 0),$$

Le coefficient de dépendance de queue supérieure et inférieure sont nuls

$$\begin{aligned}\lambda_U &= \lim_{u \rightarrow 1} \frac{1 - 2u + \max(2u - 1, 0)}{1 - u} \\ &= 0\end{aligned}$$

et

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0} \frac{\max(2u - 1, 0)}{u} \\ &= 0\end{aligned}$$

Dans le cas des copules $C^\perp$, C^+ et C^- , nous avons $\lambda_L = \lambda_U$. Néanmoins, cela n'est pas toujours le cas. Nous pouvons avoir une copule qui a une dépendance de queue supérieure sans posséder une dépendance de queue inférieure.

6. Dépendance de queue de copules Archimédiennes

6.0.1. Exemple

Considérons la copule de Gumbel. Nous avons

$$C_\theta(u, v) = \exp \left(- \left((-\log(u))^\theta + (-\log(v))^\theta \right)^{\frac{1}{\theta}} \right)$$

Nous en déduisons que

$$\begin{aligned}\lambda_U &= \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \\ &= \lim_{u \rightarrow 1} \frac{1 - 2u + u^{2^{1/\theta}}}{1 - u} \\ &= \lim_{u \rightarrow 1} \frac{-2 + 2^{1/\theta} u^{2^{1/\theta} - 1}}{1 - u} \\ &= 2 - 2^{1/\theta}\end{aligned}$$

et

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0} \frac{u^{2^{1/\theta}}}{u} \\ &= 0\end{aligned}$$

La copule de Gumbel possède donc une dépendance de queue supérieure mais pas une dépendance de queue inférieure.

6.0.2. Exemple

Soit la copule de Clayton, définie par :

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$$

Nous en déduisons que

$$\begin{aligned}\lambda_U &= \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \\ &= \lim_{u \rightarrow 1} \frac{1 - 2u + (2u^{-\theta} - 1)^{-1/\theta}}{1 - u} \\ &= \lim_{u \rightarrow 1} \frac{-2 + (2u^{-\theta} - 1)^{-1/\theta-1} u^{-\theta-1}}{1 - u} \\ &= \lim_{u \rightarrow 1} \frac{-2 + 2(2 - u^{\theta-1/\theta-1})}{-1} \\ &= 0\end{aligned}$$

et

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0} \frac{C(u, u)}{u} &= \lim_{u \rightarrow 0} \frac{(2u^\theta - 1)^{-1/\theta}}{u} \\ &= \lim_{u \rightarrow 0} (2 - u^\theta)^{-1/\theta} \\ &= 2^{-1/\theta}\end{aligned}$$

7. Dépendance de queue de copules elliptiques

7.0.3. Exemple

Pour évaluer la dépendance de queue de la copule gaussienne, on se donne deux variables X_1, X_2 suivant une loi gaussienne bi-variée, de corrélation ρ . On montre facilement que la loi de X_2 conditionnellement à $X_1 = x$ est la loi gaussienne $N(\rho x, 1 - \rho^2)$, Ainsi, pour $\rho < 1$.

$$\lambda_U = \lim_{u \rightarrow 1} \mathbb{P}(V > u | U > u) = \lim_{x \rightarrow \infty} \mathbb{P}(\Phi^{-1}(V) > x | \Phi^{-1}(U) > x) = \lim_{x \rightarrow \infty} \mathbb{P}(X_2 > x | X_1 > x)$$

et comme $X_2 | X_1 = x \sim \mathcal{N}(\rho x, 1 - \rho^2)$, alors

$$\begin{aligned} \lambda_U &= 2 \lim_{x \rightarrow \infty} \bar{\Phi} \left(\frac{\sqrt{x - x\rho}}{\sqrt{1 + \rho}} \right) \\ &= 2 \lim_{x \rightarrow \infty} \bar{\Phi} \left(\frac{x\sqrt{1 - \rho}}{\sqrt{1 + \rho}} \right) \end{aligned}$$

pour $\rho < 1$, $\lambda_U = 0$. La copule gaussienne n'admet pas une dépendance de queue supérieure. Et de même manière on trouve $\lambda_L = 0$. Ainsi, la copule gaussienne est asymptotiquement indépendante pour chacune des queues de distribution.

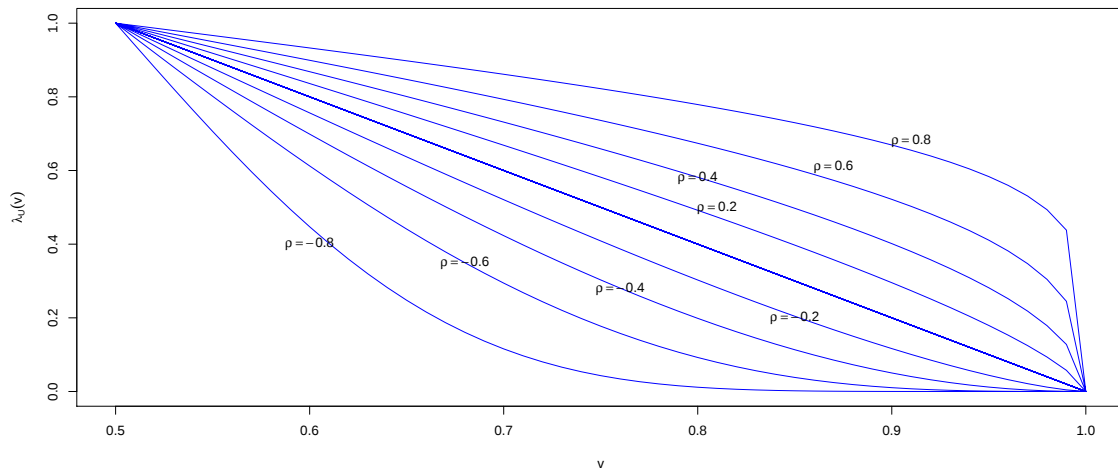


Figure 2.1. Coefficient de la dépendance de queue de copule de gaussienne

7.0.4. Exemple

Soient X_1, X_2 deux variables aléatoires de lois de Student bi-variée à ν degrés de liberté et de corrélation ρ et comme $X_2|X_1 \sim t_{\nu+1}$, alors

$$\mathbb{E}(X_2|X_1 = x) = \rho x \text{ et } \text{Var}(X_2|X_1 = x) = \frac{\nu + x^2}{1 + \nu}(1 - \rho)$$

Donc

$$\begin{aligned} \lambda_U &= \lim_{x \rightarrow \infty} \mathbb{P}(X_2 > x | X_1 > x) \\ &= \lim_{x \rightarrow \infty} 2\bar{t}_{\nu+1} \left(\left(\frac{(1 + \nu)}{\nu + x^2} \right)^{1/2} \frac{x + \rho x}{\sqrt{1 - \rho^2}} \right) \\ &= \lim_{x \rightarrow \infty} 2\bar{t}_{\nu+1} \left(\left(\frac{(1 + \nu)}{\nu/x^2 + 1} \right)^{1/2} \frac{\sqrt{1 - \rho}}{\sqrt{1 + \rho}} \right) \\ &= 2\bar{t}_{\nu+1} \left(\left(\frac{(1 + \nu)(1 - \rho)}{(1 + \rho)} \right)^{1/2} \right) \end{aligned}$$

où $t_{\nu+1}$ la fonction de répartition de la loi de Student à $\nu + 1$ degrés de liberté.

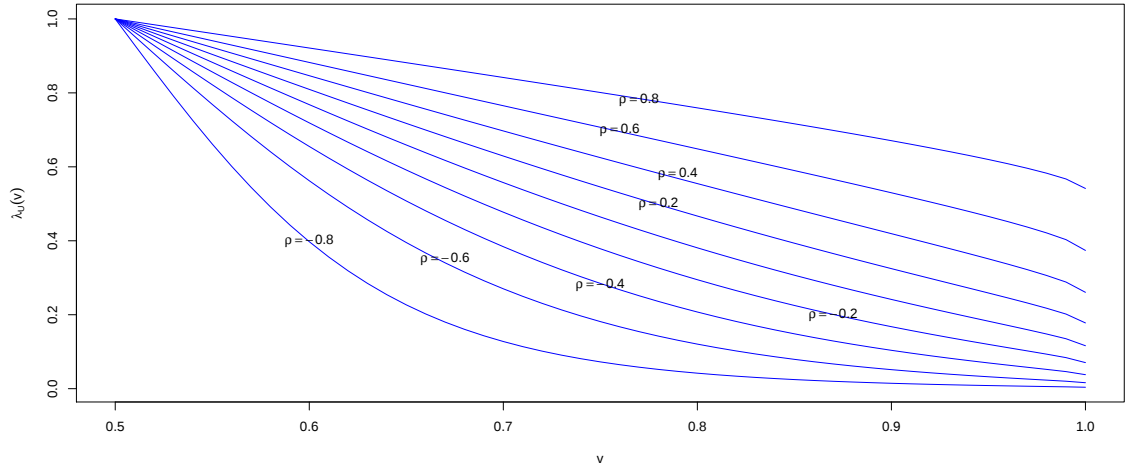


Figure 2.2. Coefficient de la dépendance de queue de copule de Student

On peut constater que le coefficient de la dépendance de queue supérieure est croissante pour ρ et décroissante pour ν et aussi tend vers zéro si le degré de liberté tend vers l'infini avec $\rho < 1$.

Le tableau suivant résume le coefficient de la dépendance de queue supérieure et inférieure pour différentes copules.

Copule	λ_L	λ_U
Gumbel	0	$2 - 2^{\frac{1}{\theta}}$
Clayton	$2^{-\frac{1}{\theta}}$	0
Cot-copule	$2^{-\frac{1}{\theta}}$	$2 - 2^{\frac{1}{\theta}}$
Frank	0	0
Student	$2\bar{t}_{\nu+1} \left(\frac{\sqrt{(\nu+1)(1-\rho)}}{\sqrt{(\rho+1)}} \right)$	$2\bar{t}_{\nu+1} \left(\frac{\sqrt{(\nu+1)(1-\rho)}}{\sqrt{(\rho+1)}} \right)$
Gaussienne	0	0

Table 2.1. *Dépendance de queue pour différentes copules*

8. Estimation de la dépendance de queue de distribution

Il existe plusieurs méthodes pour estimer un coefficient de dépendance de queue, dont celle de l'approche paramétrique qui se base sur l'hypothèse que d'une part la structure de dépendance entre les variables est représentée par une copule archimédienne et d'autre part que la limite du théorème de [Juri et Wüthrich \(2002\)](#) permet d'estimer la dépendance de queue. En plus de l'approche paramétrique, il y a la méthode non-paramétrique.

Plusieurs études ont essayé de présenter des estimateurs non paramétriques, dont certaines sont basées sur la transformation des variables aléatoires de départ en variables aléatoires de Fréchet [Poon et al. \(2004\)](#), tandis que d'autres se basent sur l'estimation empirique de la copule de queue [Schmidt et Stadtmüller \(2006\)](#).

8.1. Estimation paramétrique

[Frahm et al. \(2005\)](#) ont publié des approches paramétriques basées sur

le degré d'information sur les distributions des variables. Ces auteurs ont présenté une estimation en utilisant une distribution spécifique, une copule spécifique et une classe de copule. Parmi ces trois études la dernière relative aux copules archimédiennes est basée sur le résultat de [Juri et Wüthrich \(2002\)](#). Elle est la plus générale.

Dans le cadre de la dernière méthode utilisant une classe de distribution, les auteurs ont considéré la classe des distributions elliptiques. Mais cette méthode est limitée du fait que le cadre des copules elliptiques ne tient que dans le cas où les queues supérieure et inférieure sont égales.

Néanmoins l'étude empirique menée par [Malevergne et Sornette \(2004\)](#) montre un comportement similaire entre les deux queues positives et négatives. Cet argument vient en appui à l'étude de la dépendance de queue sous l'hypothèse des distributions elliptiques.

L'estimation de la dépendance de queue sous l'hypothèse de copule spécifique a été introduite par [Longin et Solnik \(2001\)](#), [Malevergne et Sornette \(2002\)](#) et [Patton \(2006\)](#). La limite de cette méthode est l'obligation de choisir les paramètres de la copule avant de s'assurer que les variables présentent ou non une dépendance de queue.

Pour présenter l'approche paramétrique de la dépendance de queue, on va aborder celle qui correspond à un cadre plus général, parce-que cette approche utilise toute une classe de copule, à savoir la classe de copule archimédienne. La limite vers laquelle tend la copule archimédienne établie par [Juri et Wüthrich \(2002\)](#) permet d'estimer la dépendance de queue.

On va considérer la fonction suivante :

$$F_t = \frac{C(\min(x, t))}{C(t, t)}, \quad 0 \leq x \leq 1, t \in [0, 1]$$

La copule associée à cette fonction est

$$C_t(u, v) = \frac{C(F_t^{-1}(u), F_t^{-1}(v))}{C(t, t)}$$

Cette copule tend vers une copule de Clayton quand u tend vers zéro ainsi le coefficient de dépendance de queue inférieure peut s'exprimer à l'aide du

paramètre de la copule

$$\lim_{t \rightarrow 0} C_t(u, v) = C_{Cl}(u, v, \theta), \quad \theta > 0,$$

$$\lambda_L = 2^{-1/\theta}.$$

8.2. Estimation non-paramétrique

8.2.1. Estimateur de Frahm *et al.* (2005)

En 2005, Frahm *et al.* proposent un estimateur du coefficient de dépendance de queue. Cet estimateur est obtenu à l'aide d'une copule empirique $\hat{C}_n$. La copule empirique est définie par la relation suivante entre les fonctions de distributions empiriques $\hat{F}_n$, $\hat{G}_n$ et $\hat{H}_n$

$$\hat{F}_n(x, y) = (\hat{G}_n(x), \hat{H}_n(y)).$$

Si $C(u, u)$ est différentiable pour tout $u \in (1 - \epsilon, 1)$, avec $\epsilon > 0$, alors

$$2 - \lambda_U = \lim_{u \rightarrow 1} \frac{1 - C(u, u)}{1 - u} = \lim_{u \rightarrow 1} \frac{dC(u, u)}{du} = \lim_{u \rightarrow 1} \frac{\log C(u, u)}{\log(u)}$$

Le log-estimateur est basé sur l'équation (10), cet estimateur est une généralisation de l'estimateur paramétrique présenté par [Coles *et al.* \(1999\)](#).

L'estimateur proposé est :

$$\hat{\lambda}_U^{(LOG)} = 2 - \frac{1 - \hat{C}_m((m - k)/m, (m - k)/m)}{1 - (m - k)/m}, \quad 0 < k < m.$$

$\hat{C}_m$ est la copule empirique (voir [Deheuvels \(1979\)](#)), pour plus détails).

8.2.2. Estimateur de Joe *et al.* (1992)

Parmi les estimateurs les plus répandus, on trouve l'estimateur de [Joe \(1997\)](#) défini de la manière suivante :

$$\hat{\lambda}_U^{(SEC)} = 2 - \frac{1 - \hat{C}_n((n - k)/n, (n - k)/n)}{1 - (n - k)/n}, \quad 0 < k \leq n.$$

La normalité asymptotique et la consistance forte de cet estimateur ont été étudiées dans [Schmidt et Stadtmüller \(2006\)](#). Dans leurs articles, Joe *et al.* ont aussi présenté l'estimateur non paramétrique du coefficient de dépendance de queue supérieure et inférieure en fonction de la copule de queue.

8.2.3. Choix de k

Les deux estimateurs $\hat{\lambda}_U^{(SEC)}$ et $\hat{\lambda}_U^{(LOG)}$ sont définis à l'aide d'un seuil k . Le choix de k est basée sur la propriété de l'homogénéité de la copule de queue (voir théorème 1, [Schmidt et Stadtmüller \(2006\)](#)). Il correspond à la recherche d'un équilibre entre le biais et la variance. Pour plus d'informations voir l'algorithme heuristique "plateau-finding" présenté dans [Frahm *et al.* \(2005\)](#).

8.2.4. Estimateur de Capéreaa *et al.* (1997)

Capéreaa *et al.*, ont proposé un estimateur non paramétrique du coefficient de dépendance de queue supérieure. On considère donc $((U_1, V_1), \dots, (U_n, V_n))$ un échantillon de variables aléatoires obtenu à partir de la copule C . Cet estimateur repose sur l'hypothèse que la fonction de copule empirique se rapproche d'une copule de valeurs extrêmes et définir

$$\hat{\lambda}_U^{(CFG)} = 2 - 2 \exp \left(\frac{1}{n} \sum_{i=1}^n \log \left(\frac{\sqrt{\log \frac{1}{u_i} \log \frac{1}{v_i}}}{\log \frac{1}{\max(u_i, v_i)^2}} \right) \right),$$

Où n la taille de l'échantillon, $u_i = F_X(x_i)$ et $v_i = F_Y(y_i)$, $x_i, y_i \in \mathbb{R}$.

[Frahm *et al.* \(2005\)](#), montrent que cette hypothèse est pas forte et $\hat{\lambda}_U^{(CFG)}$ fonctionne bien même lorsque la copule ne fait pas partie de la classe d'une copule de valeurs extrêmes.

8.2.5. Estimateur de Dobric et Schmidt. (2005)

On note : $C(u, u) = \lambda_L u + o(u)$ pour $0 \leq u \leq 1$ tel que $(o(u))/u \rightarrow 0$ pour $u \rightarrow 0$, si la limite existe. Alors l'estimation de la dépendance de queue inférieure peut s'écrire de la manière suivante :

$$\hat{\lambda}_L^{(1)} = 2 - \frac{\log C_n(1 - \frac{k}{n}, 1 - \frac{k}{n})}{\log(1 - \frac{k}{n})},$$

où $k \in \mathbb{N}$ doit être déterminée en fonction de n . Cet estimateur basé sur l'approximation de la dérivée de la dépendance de queue inférieure λ_L par la pente d'une sécante appropriée.

Un autre estimateur

$$\hat{\lambda}_L^{(2)} = 2 - \frac{1 - C_n(1 - \frac{k}{n}, 1 - \frac{k}{n})}{1 - \frac{n-k}{n}}.$$

Nous avons appliqué la méthode des moindres carrés pour le modèle de régression.

$$\hat{C}_n(\frac{i}{n}, \frac{i}{n}) = \lambda_L \frac{i}{n} + \epsilon_i, \quad i = 1, \dots, k.$$

avec ϵ_i le terme d'erreur.

Le troisième estimateur est basé sur l'idée que la copule $C(u, v)$ peut être approchée par un mélange d'une Copule comonotone et une copule indépendante pour $u = v$, i.e

$$C_n(u, u) = \lambda u + (1 - \lambda)u^2$$

Donc l'estimateur de λ_L est donné par :

$$\hat{\lambda}_L^{(3)} = 2 - \frac{\sum_{i=1}^k \left(\hat{C}_n(\frac{i}{n}, \frac{i}{n}) - (\frac{i}{n})^2 \right) \left((\frac{i}{n}) - (\frac{i}{n})^2 \right)}{\sum_{i=1}^k \left((\frac{i}{n}) - (\frac{i}{n})^2 \right)^2}.$$

LA DÉPENDANCE DE QUEUE POUR LA LOI HYPERBOLIQUE GÉNÉRALISÉE

1. Introduction

Avant de définir une loi hyperbolique généralisée, il faut d'abord rappeler quelques notions élémentaires sur la loi inverse gaussienne et la loi inverse gaussienne généralisée.

1.1. Loi inverse gaussienne

1.1.1. Définition

Une variable aléatoire X suit une loi inverse gaussienne, $X \sim IG(x; \lambda, \mu)$

La densité de probabilité de la loi inverse gaussienne est donnée par :

$$f(x; \mu, \lambda) = \begin{cases} \left(\frac{\lambda}{2\pi x^3} \right)^{1/2} \exp \frac{-\lambda(x - \mu)^2}{2\mu^2 x} & \text{pour } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

où $\mu > 0$ est son espérance et $\lambda > 0$ est un paramètre de forme.

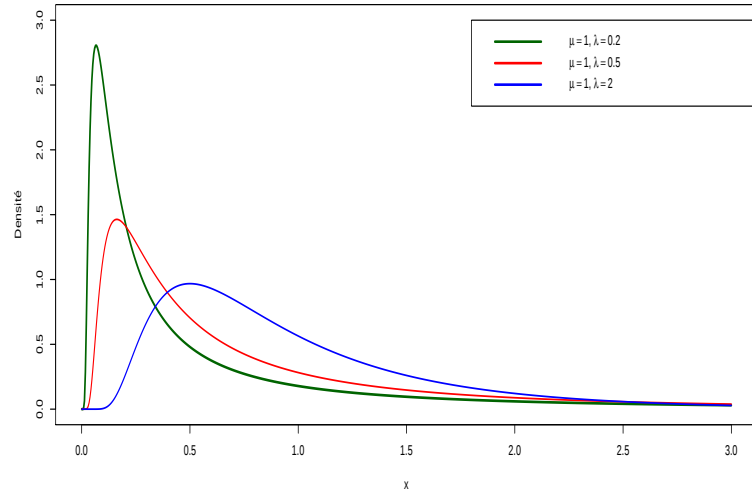


Figure 3.1. La densité de la loi inverse gaussienne avec $\mu = 1$ et $\lambda = 0.2, 0.5, 2$.

1.1.2. Remarque

Lorsque λ tend vers l'infini, la loi inverse-gaussienne se comporte comme une loi normale (cf [Wald \(1947\)](#)).

1.2. Loi inverse gaussienne généralisée

1.2.1. Définition

Une variable aléatoire X suit une loi inverse gaussienne généralisée, $X \sim GIG(\lambda, \delta, \gamma)$

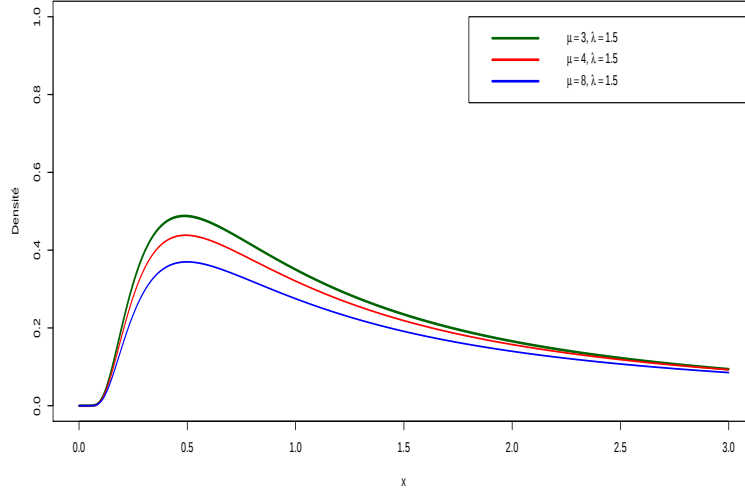


Figure 3.2. La densité de la loi inverse gaussienne avec $\mu = 3, 4, 8$ et $\lambda = 1.5$.

La densité de probabilité de la loi inverse gaussienne généralisée est donnée par :

$$f(x) = \begin{cases} \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\delta\gamma)} x^{\lambda-1} e^{-\frac{1}{2}(\gamma^2 x + \frac{\delta^2}{x})} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

où κ_λ est la fonction de Bessel modifiée de troisième espèce et de paramètre λ , et les paramètres vérifient :

$$\begin{cases} \delta \geq 0, \gamma > 0 & \text{si } \lambda > 0, \\ \delta > 0, \gamma > 0 & \text{si } \lambda = 0, \\ \delta > 0, \gamma \geq 0 & \text{si } \lambda < 0. \end{cases}$$

Elle intervient dans la modélisation de nombreux phénomènes réels, notamment pour la modélisation de données concernant des temps d'attente [Jorgensen \(1982\)](#), des phénomènes extrêmes en hydrologie [Chebana et al. \(2010\)](#), l'activité neuronale [Iyengar et Liao \(1997\)](#). Pour plus de détails sur l'étude des propriétés statistiques de cette loi, on peut citer par exemple [Jorgensen \(1982\)](#), [Perreault et al. \(1999a\)](#), [Perreault et al. \(1999b\)](#).

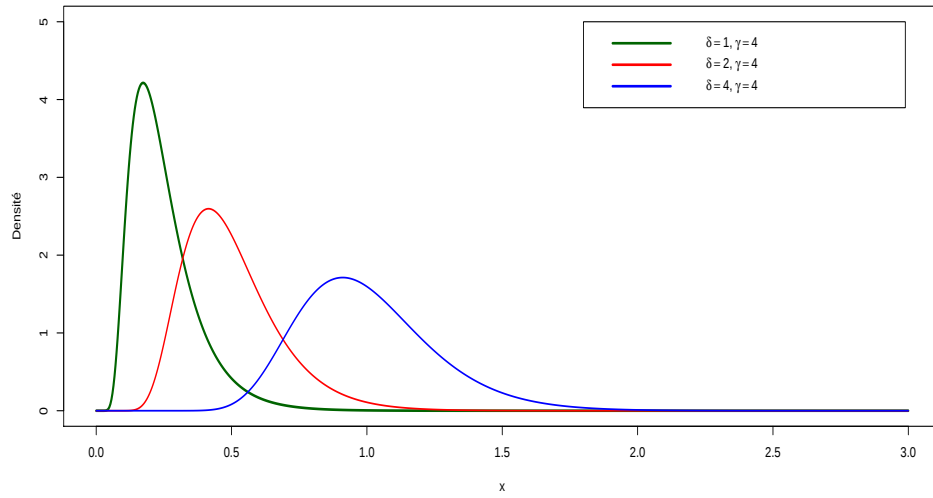


Figure 3.3. La densité de la loi inverse gaussienne généralisée avec $\delta = 1, 2, 4$, $\gamma = 4$ et $\lambda = -0.5$.

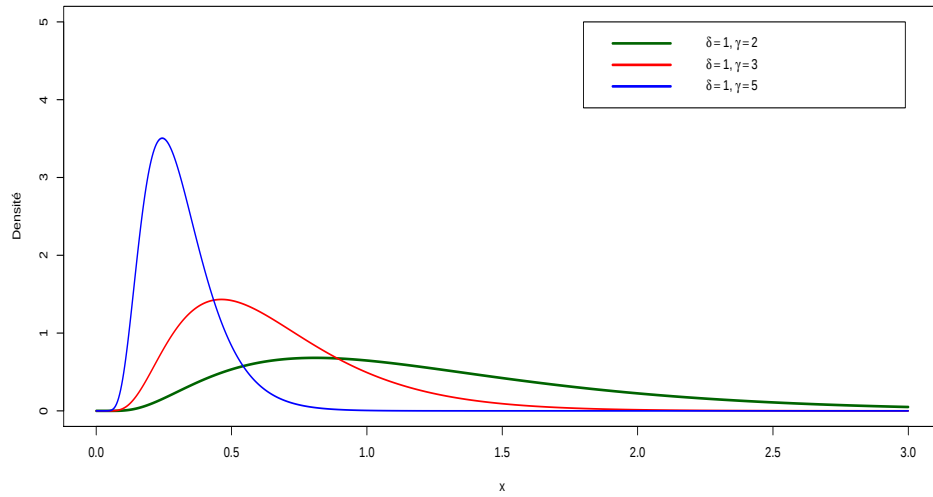


Figure 3.4. La densité de la loi inverse gaussienne généralisée avec $\delta = 1$, $\gamma = 2, 3, 5$ et $\lambda = 2$.

1.2.2. Remarque

Les cas Particuliers de cette distribution sont la loi inverse-gaussienne ($\lambda = -\frac{1}{2}$) la loi hyperbolique positive ($\gamma = 1$), la loi inverse- χ^2 de δ^2 degrés de liberté ($\lambda = \frac{\delta^2}{2}, \gamma = 0$) et la loi gamma ($\delta = 0, \lambda > 0, \gamma > 0$).

2. Loi hyperbolique généralisée

La distribution hyperbolique généralisée a été introduite dans [Barndorff-Nielsen \(1997\)](#), et développée par de nombreux auteurs, voir [Schlueter et Fischer \(2012\)](#), [McNeil *et al.* \(2005\)](#), [Schmidt \(2002\)](#), [Barndorff-Nielsen and Christian \(1997\)](#), en particulier en relation avec plusieurs applications dans la finance, voir [Prause \(1999\)](#), [Eberlein et Prause \(2002\)](#). Cette famille contient et généralise de nombreuses distributions familières telles que la loi de Student, Gaussienne, variance-Gamma, Cauchy et autres.

Connue pour être remarquablement adaptée à l'étude des séries de données observées en finance, la loi hyperbolique généralisée fut utilisée avec succès, aussi bien en temps discret [Guegan et Zhang \(2009\)](#) qu'en temps continu, où elle est, le plus souvent utilisée pour la construction de processus exponentiels de Lévy [Eberlein et Prause \(2002\)](#).

Introduite par Barndorff-Nielsen suivant une approche assez particulière "mélange moyenne-variance", la distribution hyperbolique généralisée, utilisée suivant cette technique, consiste à dire qu'une variable aléatoire X est issue d'un mélange moyenne-variance, si sa loi de distribution peut être décrite par $x = \mu + \beta\sigma^2 + \sigma\epsilon$ où ϵ et σ sont des variables aléatoires indépendantes. On rappelle que dans un mélange-variance standard, ϵ est de loi normale et σ est la distribution "mélangeante" qui doit avoir $\mathbb{R}^+$ comme support. Dans ce cas, nous avons un "mélange" entre une loi normale et une loi Gaussienne inverse généralisée (GIG) où la moyenne et la variance de la loi normale sont définies à partir d'une variable aléatoire W de distribution *GIG*. L'équation (3.1) illustre formellement le lien entre les paramètres de la loi normale et

ceux de la GIG voir [Prause \(1999\)](#).

$$GH(x; \lambda, \alpha, \beta, \delta, \mu) = \int_0^\infty \mathcal{N}(x; \mu + \beta, w) GIG(w; \lambda, \delta, \sqrt{\alpha^2 - \beta^2}) dw \quad (3.1)$$

Cette formulation est en fait la définition d'une distribution conditionnelle (cf [Billingsley \(1986\)](#)).

$$\begin{aligned} X|W = w &\sim \mathcal{N}(x; \mu + \beta, w) \\ W &= GIG(w; \lambda, \delta, \sqrt{\alpha^2 - \beta^2}) \quad \text{donc on a :} \\ X &\sim GH(x; \lambda, \alpha, \beta, \delta, \mu). \end{aligned}$$

2.0.3. définition

La densité f_X de la loi hyperbolique généralisée de paramètres $\lambda, \alpha, \theta, \delta, \gamma$ est donnée par :

$$f_X(x) = \ell_\lambda(\alpha, \theta, \delta) \cdot k_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) \cdot \left(\sqrt{\delta^2 + (x-\mu)^2}\right)^{\lambda-\frac{1}{2}} \cdot e^{\theta(x-\mu)},$$

où ℓ_λ est défini par :

$$\ell_\lambda(\alpha, \theta, \delta) = \frac{(\alpha^2 - \theta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi} \alpha^{\lambda-\frac{1}{2}} \delta^\lambda k_\lambda(\delta\sqrt{\alpha^2 - \beta^2})}.$$

Les paramètres $\lambda, \mu \in \mathbb{R}$, $\alpha, \delta \geq 0$, $|\beta| \leq \alpha$ doivent satisfaire les conditions suivantes :

- $|\beta| < \alpha$ et $\delta > 0$ si $\lambda = 0$
- $|\beta| < \alpha$ et $\delta \geq 0$ si $\lambda > 0$
- $|\beta| = \alpha$ et $\delta > 0$ si $\lambda < 0$

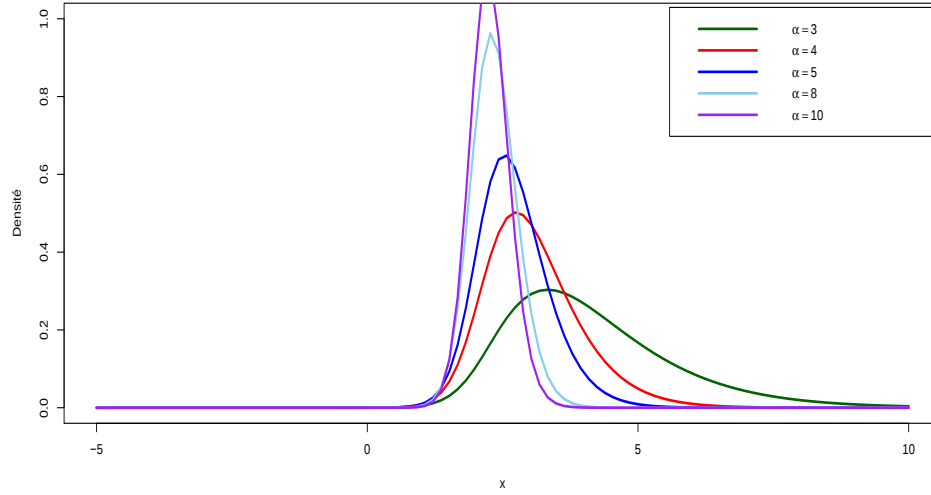


Figure 3.5. La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\delta = 1$, $\beta = 2$ et $\lambda = 2$.

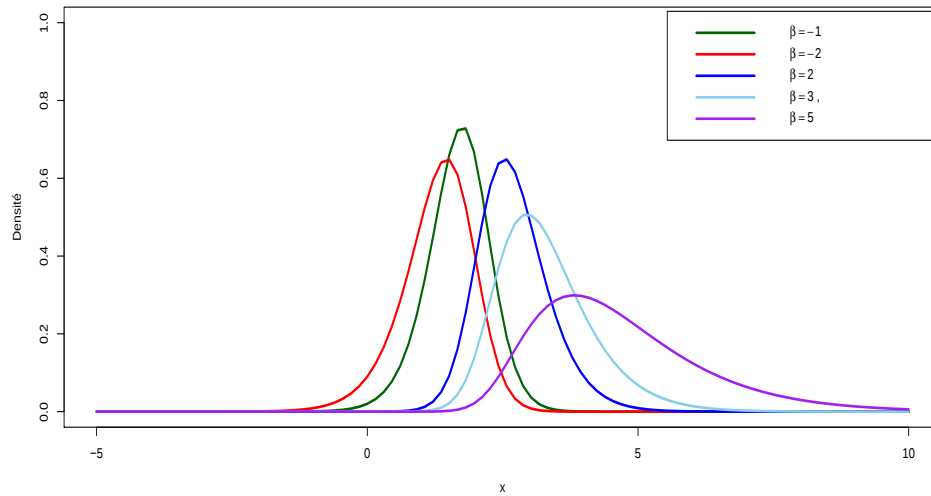


Figure 3.6. La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\delta = 1$, $\lambda = 2$ et $\alpha = 5$.

2.1. Paramétrisations

Différentes paramétrisations de la distribution hyperbolique généralisée ont été proposées dans la littérature.

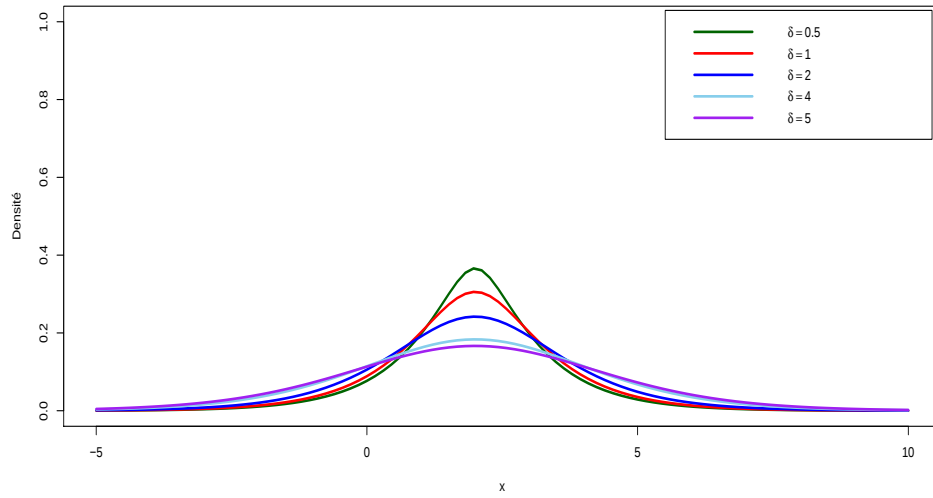


Figure 3.7. La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\beta = 0$, $\lambda = 1$ et $\alpha = 1$.

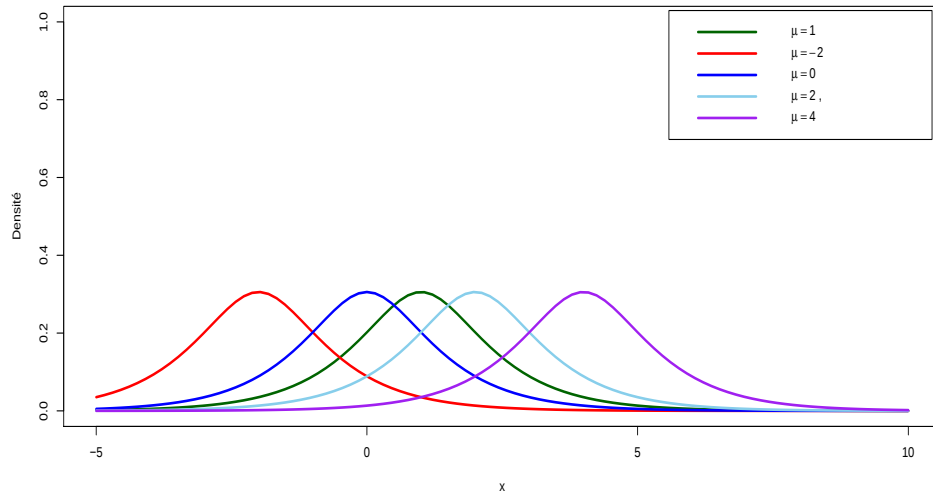


Figure 3.8. La densité de la loi hyperbolique généralisée avec $\mu = 2$, $\beta = 0$, $\lambda = 1$ et $\alpha = 1$.

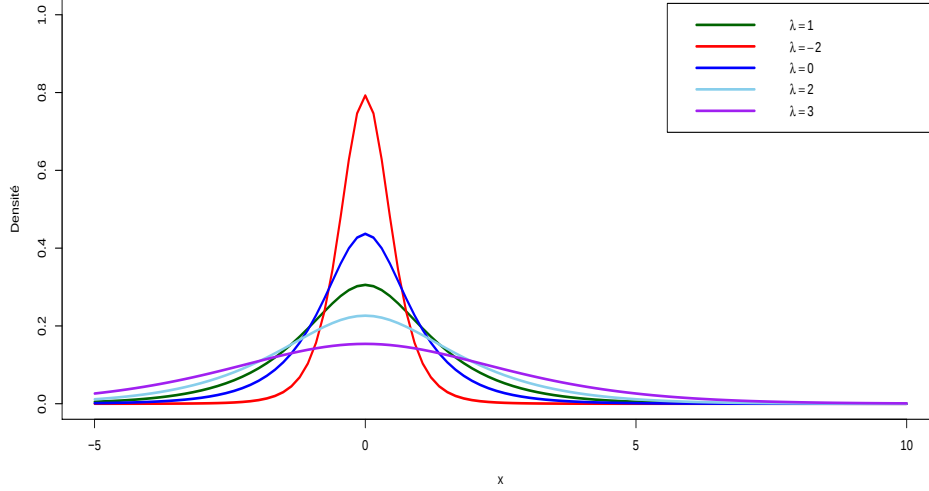


Figure 3.9. La densité de la loi hyperbolique généralisée avec $\mu = 0$, $\delta = 1$, $\beta = 0$ et $\alpha = 1$.

- 1^{re} paramétrisation : $\mu, \delta, \alpha, \beta, \lambda$
- 2^e paramétrisation : $\mu, \delta, \zeta, \varrho, \lambda$
- 3^e paramétrisation : $\mu, \delta, \xi, \chi, \lambda$
- 4^e paramétrisation : $\mu, \delta, \bar{\alpha}, \bar{\beta}, \lambda$.

Parfois, il est utile de reparamétriser la fonction de densité de la loi hyperbolique généralisée en termes de paramètres λ , τ , ζ , δ et μ , tel que $\tau = \beta/\gamma$ et $\zeta = \delta\gamma$.

En utilisant cette paramétrisation, la fonction de densité de la loi hyperbolique généralisée peut s'écrire comme suit :

$$f(x; \nu, \beta, \delta, \mu) = \frac{\sqrt{\zeta}}{\sqrt{2\pi\delta}K_\lambda(\zeta)} \cdot \frac{\sqrt{\zeta}K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\tau^2}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\right)}{\left(\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}/\sqrt{1+\tau^2}\right)^{\frac{1}{2}-\lambda}} \cdot e^{\tau\zeta\frac{x-\mu}{\delta}}, \quad x \in \mathbb{R}.$$

Table 3.1. La relation entre le quatre paramètres de la distribution hyperbolique généralisée à des valeurs fixes de μ et δ .

	À : paramétrisation			
De : paramétrisation	1	2	3	4
	$\mu, \delta, \alpha, \beta, \lambda$	$\mu, \delta, \zeta, \varrho, \lambda$	$\mu, \delta, \xi, \chi, \lambda$	$\mu, \delta, \bar{\alpha}, \bar{\beta}, \lambda$
1. $\mu, \delta, \alpha, \beta, \lambda$ $(\alpha > 0, \delta > 0)$		$\zeta = \frac{1}{\delta\sqrt{\alpha^2 - \beta^2}}$ $\varrho = \beta/\alpha$	$\xi = \frac{1}{\sqrt{1+\delta\sqrt{\alpha^2 - \beta^2}}}$ $\chi = \frac{1}{\alpha\sqrt{1+\delta\sqrt{\alpha^2 - \beta^2}}}$	$\bar{\alpha} = \delta\alpha$ $\bar{\beta} = \delta\beta$
2. $\mu, \delta, \zeta, \varrho, \lambda$ $(\zeta > 0, \delta > 0)$	$\alpha = \frac{\zeta}{\delta\sqrt{1-\varrho^2}}$ $\beta = \varrho\alpha$		$\xi = \frac{1}{\sqrt{1+\zeta}}$ $\chi = \zeta\varrho$	$\bar{\alpha} = \frac{\zeta}{\sqrt{1-\varrho^2}}$ $\bar{\beta} = \varrho\alpha$
3. $\mu, \delta, \xi, \chi, \lambda$ $(\xi > 0, \delta > 0)$	$\alpha = \frac{1-\xi^2}{\delta\xi\sqrt{\xi^2 - \chi^2}}$ $\beta = \alpha\frac{\xi}{\chi}$	$\zeta = \frac{1}{\xi^2 - 1}$ $\varrho = \frac{\xi}{\chi}$		$\bar{\alpha} = \frac{1-\xi^2}{\xi\sqrt{\xi^2 - \chi^2}}$ $\bar{\beta} = \bar{\alpha}\frac{\xi}{\chi}$
4. $\mu, \delta, \bar{\alpha}, \bar{\beta}, \lambda$ $(\bar{\alpha} > 0, \delta > 0)$	$\alpha = \frac{\bar{\alpha}}{\delta}$ $\beta = \frac{\bar{\beta}}{\delta}$	$\zeta = \frac{1}{\delta\sqrt{\bar{\alpha}^2 - \bar{\beta}^2}}$ $\varrho = \frac{\bar{\beta}}{\bar{\alpha}}$	$\xi = \frac{1}{\sqrt{1+\delta\sqrt{\bar{\alpha}^2 - \bar{\beta}^2}}}$ $\chi = \frac{\bar{\beta}}{\bar{\alpha}\sqrt{1+\delta\sqrt{\bar{\alpha}^2 - \bar{\beta}^2}}}$	

2.1.1. Remarque

Les cas Particuliers de cette distribution sont :

□ $\lambda = 1$: loi hyperbolique

$$f_H(x; \alpha, \beta, \delta, \mu) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \cdot e^{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)}$$

□ $\lambda = -\frac{1}{2}$: Inverse Gaussienne Normale NIG(mélange d'une normale et d'une IG)

$$f_{NIG}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 + (x-\mu)^2}} \cdot e^{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x-\mu)}$$

□ $\lambda = -\frac{\nu}{2}, \alpha = \beta = 0, \delta = \sqrt{\nu}$: Loi de Student de ν degrés de liberté

$$f(x; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

□ $\delta = 0$: Loi de Variance-Gamma

$$f_H(x; \alpha, \beta, \mu, \lambda) = \frac{(\sqrt{\alpha^2 - \beta^2})^{2\lambda} |x - \mu|^{\lambda-1/2} K_{\lambda-1/2}(\alpha|x - \mu|)}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-1/2}} \cdot e^{\beta(x-\mu)}$$

□ $|\beta| \rightarrow \alpha, \lambda < 0$ et $\delta^2 = -2\lambda = \nu$: loi de Student asymétrique

$$f(x; \nu, \beta, \delta, \mu) = \frac{\delta^\nu K_{(\nu+1)/2}(|\beta|\sqrt{\delta^2 + (x-\mu)^2})}{\sqrt{\pi}2^{(\nu-1)/2} \Gamma(\frac{\nu}{2}) (|\beta|\sqrt{\delta^2 + (x-\mu)^2})^{(\nu+1)/2}} \cdot e^{\beta(x-\mu)}$$

Le but de ce chapitre est d'étudier le coefficient de la dépendance de queue supérieure dans le cas de la distribution hyperbolique généralisée. Plus précisément, une variable aléatoire bi-variée $X = (X_1, X_2)$ suit une loi hyperbolique généralisée $GH(\lambda, \alpha, \beta, \delta, \gamma)$ si

$$X_1 = \mu_1 + \theta_1 W_1 + \sqrt{W_1} Z_1, \quad X_2 = \mu_2 + \theta_2 W_2 + \sqrt{W_2} Z_2, \quad (3.2)$$

où $Z = (Z_1, Z_2) \sim N_2(0, \Sigma)$, est une loi normale bi-variée d'espérance 0 et de matrice de corrélation Σ et $W = (W_1, W_2) \sim GIG(\lambda, \delta, \gamma)$, est la loi inverse gaussienne généralisée avec la fonction de densité

$$f_W(x) = \begin{cases} \frac{(\frac{\gamma}{\delta})^\lambda}{2k_\lambda(\delta\gamma)} x^{\lambda-1} \exp(-\frac{1}{2}(\gamma^2 x + \frac{\delta^2}{x})) & \text{if } x > 0, \\ 0 & \text{sinon.} \end{cases}$$

où $\gamma = \sqrt{\alpha^2 - \theta^2}$, $\lambda \in \mathbb{R}, \delta > 0$ et $\gamma \geq 0$, et K_λ la fonction de Bessel modifiée du troisième type avec l'indice $\lambda \in \mathbb{R}$, $K_\lambda(x) = \frac{1}{2} \int_0^\infty \exp(-\frac{x}{2}(y + \frac{1}{y})) y^{\lambda-1} du$.

La fonction de densité de la loi $GH(\lambda, \alpha, \beta, \delta, \mu)$ est donnée par

$$f_X(x) = \frac{(\alpha^2 - \theta^2)^{\frac{\lambda}{2}} k_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2 + (x-\mu)^2}) \exp(\theta(x-\mu))}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}} k_\lambda(\delta\sqrt{\alpha^2 - \beta^2}) \delta^\lambda (\sqrt{\delta^2 + (x-\mu)^2})^{-\lambda+\frac{1}{2}}},$$

où K_λ est la fonction de Bessel modifiée du troisième type (voir l'annexe), donnée par

$$K_\lambda(x) = \frac{1}{2} \int_0^\infty \exp\left(-\frac{x}{2}\left(y + \frac{1}{y}\right)\right) y^{\lambda-1} dy, \quad \text{pour } x > 0, \lambda \in \mathbb{R}, \quad (3.3)$$

voir par exemple, [Erdelyi et al. \(1954\)](#), [Jorgensen \(1982\)](#). [Banachewicz et Van der Vaart \(2008\)](#) ont donné des résultats généraux pour la dépendance de queue de la loi "skewed grouped t -distributions" dans le cas où ses degrés de liberté sont différents, *i.e.* pour une variable aléatoire $X = (X_1, X_2)$, on peut écrire

$$X_1 = \frac{\theta_1}{G_\mu^{-1}(U)/\mu} + \frac{Z_1}{\sqrt{G_\mu^{-1}(U)/\mu}}, \quad X_2 = \frac{\theta_2}{G_\nu^{-1}(U)/\nu} + \frac{Z_2}{\sqrt{G_\nu^{-1}(U)/\nu}}, \quad (3.4)$$

où $Z = (Z_1, Z_2)$ est une normale bi-variée avec moyenne 0 et de matrice de corrélation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, avec coefficient de corrélation ρ et Z indépendamment distribuée de la loi uniforme U . Le paramètre $\theta = (\theta_1, \theta_2)' \in \mathbb{R}^2$ contrôle l'asymétrie de la distribution et τ, ν sont des constantes positives. La fonction $G_\eta^{-1}(\cdot)$ est la fonction de distribution inverse de $\Gamma(\frac{\eta}{2}, \frac{\eta}{2})$ avec $\eta > 0$.

[Fung et Seneta \(2010\)](#) ont étudié un cas particulier de la loi "skewed grouped t -distributions", voir [Banachewicz et Van der Vaart \(2008\)](#), quand $\eta = \nu = \tau$ dans (3.4), qui donne le cas de la loi "skew t -distribution", introduite dans [Demarta and McNeil \(2005\)](#). Ce modèle peut être défini pour $X = (X_1, X_2)$ par

$$X_1 = \frac{\theta_1}{G_\tau^{-1}(U)} + \frac{Z_1}{\sqrt{G_\tau^{-1}(U)}}, \quad X_2 = \frac{\theta_2}{G_\nu^{-1}(U)} + \frac{Z_2}{\sqrt{G_\nu^{-1}(U)}},$$

où Z_1, Z_2, Σ, τ et ν sont défini précédemment.

Le coefficient de la dépendance de queue inférieure pour une vecteur aléatoire bi-variée $X = (X_1, X_2)$, avec des fonctions de répartition F_1 et F_2 est défini par $\lambda_L = \lim_{u \rightarrow 0} \mathbb{P}(X_1 \leq F_1^{-1}(u) | X_2 \leq F_2^{-1}(u))$ si la limite existe, où F_1^{-1} et F_2^{-1} sont des fonctions de répartition de la loi inverse gaussienne généralisée de F_1 et F_2 , voir [Joe \(1997\)](#).

Le coefficient de la dépendance de queue supérieure d'un vecteur aléatoire X peut être défini de la même manière que

$$\lambda_U = \lim_{u \rightarrow 1} \mathbb{P} \left(X_1 \geq F_1^{(-1)}(u) | X_2 \geq F_2^{(-1)}(u) \right) \text{ ou de manière équivalente}$$

$$\lambda_U = \lim_{x \rightarrow +\infty} \mathbb{P} \left(X_1 \geq F_1^{-1}(F_2(x)) | X_2 \geq x \right). \quad (3.5)$$

Le but de cet chapitre est de fournir des résultats analytiques du coefficient de dépendance de la queue du modèle hyperbolique généralisé. Nous allons considérer la dépendance de la queue supérieure pour X dans le modèle (3.2), en utilisant une méthode dérivée d'une forme équivalente de (3.5) pour λ_U :

$$\lambda_U = \lim_{x \rightarrow +\infty} \mathbb{P} \left(X_1 \geq F_1^{-1}(F_2(x)) | X_2 = x \right) + \lim_{x \rightarrow +\infty} \mathbb{P} \left(X_2 \geq F_2^{-1}(F_1(x)) | X_1 = x \right),$$

voir [Nelsen \(1999\)](#).

3. Théorème

Le coefficient de la dépendance queue supérieur du vecteur (X_1, X_2) défini dans (3.2) est donné par :

1) Si $\theta_1 = \theta_2 = 0$, alors

$$\lambda_U = 1 - \int_0^\infty g_{\lambda, \gamma, \delta}(t) dt,$$

où

$$g_{\lambda, \gamma, \delta}(t) = \frac{\gamma^{\lambda-2}}{\lambda \delta^\lambda k_\lambda(\delta \lambda)} e^{-\frac{1}{2}(\gamma t)^2}$$

2) Si $\theta_1, \theta_2 > 0$, alors

$$\lambda_U = 1,$$

3) Si $\theta_1 = \theta_2 = \theta < 0$, alors

$$\lambda_U = 1.$$

3.1. Démonstration

Soit $\theta_1 = \theta_2 = 0$. Pour $x \rightarrow a$, on note $f(x) \sim g(x)$ si $\frac{f(x)}{g(x)} \rightarrow 1$ et $f(x) \propto g(x)$ s'il existe $c > 0$, tel que $g(x) = cf(x)$. On considère $h(x) = F_1^{-1} \circ F_2(x)$, $g(x) = \sqrt{Q_{\lambda_1, \delta_1, \gamma_1}^{-1} \circ Q_{\lambda_2, \delta_2, \gamma_2}(x^2)}$ et $S^2 = Q_{\lambda_2, \delta_2, \gamma_2}^{-1}(U)$.

où $Q_{\lambda,\gamma,\delta}$ est la fonction de répartition de la loi inverse gaussienne généralisée et U la loi uniforme. L'équation (3.2) peut être décrite comme suit :

$$X_1 = \mu_1 + \frac{\theta_1}{(g(S))^2} + \frac{Z_1}{g(S)}, \quad X_2 = \mu_2 + \frac{\theta_2}{S^2} + \frac{Z_2}{S},$$

où μ_1, μ_2, Z_1 et Z_2 sont définis précédemment.

Le coefficient de la dépendance de queue supérieure peut être réécrit :

$$\begin{aligned} \lambda_U &= \lim_{x \rightarrow \infty} \frac{\mathbb{P}((X_1 \geq h(x), X_2 \geq x))}{\mathbb{P}(X_2 \geq x)}, \\ &= \lim_{x \rightarrow \infty} \frac{\int_{h(x)}^{\infty} \int_x^{\infty} f_{X_1, X_2}(s, t) ds dt}{\mathbb{P}(X_2 \geq x)}, \end{aligned}$$

Nous utilisons la règle de l'Hopital pour obtenir

$$\begin{aligned} \lambda_U &= \lim_{x \rightarrow \infty} \frac{-\int_{h(x)}^{\infty} f_{X_1, X_2}(s, x) ds - h'(x) \int_x^{\infty} f_{X_1, X_2}(h(x), t) dt}{-f_2(x)}, \\ &= \lim_{x \rightarrow \infty} \left(\mathbb{P}(X_1 \geq h(x) | X_2 = x) + h'(x) \frac{f_1(x)}{f_2(x)} \mathbb{P}(X_2 \geq x | X_1 = h(x)) \right). \end{aligned} \quad (3.6)$$

La fonction de répartition Q_{W_2} de la loi inverse gaussienne généralisée est donné par

$$Q_{W_2}(x) = \int_{\infty}^x \frac{(\frac{\gamma}{\delta})^{\lambda}}{2k_{\lambda}(\delta\lambda)} t^{\lambda-1} \exp(-\frac{1}{2}(\gamma^2 t - \frac{\delta^2}{t})) dt.$$

Lorsque $x \rightarrow \infty$ on a la fonction de répartition Q_{W_2} qui vérifie,

$$1 - Q_{W_2}(x) \sim c_{\lambda,\gamma,\delta} \frac{x^{\lambda}}{\lambda} \exp(-\frac{1}{2}\gamma^2 x), \text{ où } c_{\lambda,\gamma,\delta} = \frac{(\frac{\gamma}{\delta})^{\lambda}}{2k_{\lambda}(\delta\lambda)}. \text{ Alors}$$

$$Q_{W_2}(x) = 1 - c_{\gamma,\delta} \frac{x^{\lambda}}{\lambda} \cdot \exp(-\frac{1}{2}\gamma^2 x).$$

On utilise le lemme 3.1, voir [Banachewicz et Van der Vaart \(2008\)](#), on inverse la dernière équation et on obtient, quand $u \rightarrow 1$,

$$Q_{W_2}^{-1}(u) = -\frac{\log(1-u)}{\gamma^2/2}.$$

Alors pour $x \rightarrow \infty$

$$g(x) = m_{\gamma_2, \delta_2} x^{2\lambda_2} \exp\left(-\frac{1}{2}\gamma_2^2 x^2\right), \quad \text{où} \quad m_{\lambda_2, \gamma_2, \delta_2} = 2 \frac{c_{\lambda_2, \gamma_2, \delta_2}}{\lambda_2 \gamma_2^2}.$$

La densité f_1 de la loi hyperbolique généralisée de paramètres $\lambda, \alpha, \theta, \delta, \gamma$ est donnée par

$$f_1(x) = \ell_\lambda(\alpha, \theta, \delta) k_{\lambda-\frac{1}{2}}(\alpha \sqrt{\delta^2 + (x-\mu)^2}) (\sqrt{\delta^2 + (x-\mu)^2})^{\lambda-\frac{1}{2}} \exp(\theta(x-\mu)),$$

où $\ell_\lambda(\alpha, \theta, \delta) = \frac{(\alpha^2 - \theta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi} \alpha^{\lambda-\frac{1}{2}} \delta^\lambda k_\lambda(\delta \sqrt{\alpha^2 - \beta^2})}$. Lorsque $x \rightarrow \infty$, on a la fonction de répartition F_1 qui vérifie,

$$\begin{aligned} 1 - F_1(x) &\sim \int_x^\infty \ell_\lambda(\alpha, \theta, \delta) t^{\lambda-1/2} k_{\lambda-\frac{1}{2}}(\alpha \sqrt{\delta^2 + (t-\mu)^2}) \exp(\theta(t-\mu)) dt, \\ &\sim d_\lambda x^\lambda \exp(-(\alpha - \theta)x), \end{aligned}$$

où $d_\lambda = \frac{\pi}{2\lambda\alpha} \ell_\lambda(\alpha, \theta, \delta)$. En utilisant le même lemme que ci-dessus, pour $u \rightarrow 1$, la fonction inverse F_1 est donnée par

$$F_1^{-1}(u) = -\frac{\log(1-u)}{\alpha - \theta}.$$

La fonction $h(x) = F_1^{-1} \circ F_2(x)$ est strictement croissante et vérifie, quand $x \rightarrow \infty$,

$$h(x) \sim (\alpha - \theta)x.$$

La dérivée de h est égale à $h' = \frac{f_2}{f_1 \circ h}$, et donc, lorsque $x \rightarrow \infty$,

$$\frac{f_1(x)}{f_2(x)} h'(x) = \frac{f_1(x)}{f_1 \circ h(x)} \sim (\alpha - \theta)^{1-\lambda} \exp(-(\alpha - \theta)x(1 - (\alpha - \theta))).$$

On constate que cette quantité tend vers zéro pour $\alpha < 1 + \theta$. Par conséquent, le deuxième terme dans l'équation (3.6) tend vers zéro. Il nous reste à calculer le premier terme de l'équation (3.6).

La densité conditionnelle de S^2 sachant X_2 satisfait, selon le théorème de Bayes,

$$\begin{aligned} p_{S^2|X_2=x_2}(t) &\propto p_{X_2|S^2=t}(x_2) p_{S^2}(t), \\ &\propto t^{(\lambda-1/2)-1} \exp\left(-\frac{1}{2}\left(\gamma^2 t + \frac{\delta^2 + x^2}{t}\right)\right). \end{aligned}$$

Alors $S^2|X_2 = x_2 \sim GIG(\lambda - 1/2, \gamma, \sqrt{\delta^2 + x^2})$.

Maintenant, nous calculons le deuxième terme de (3.6).

$$\begin{aligned} \mathbb{P}(X_1 \geq x_1 | X_2 = x_2) &= \int_0^\infty \mathbb{P}(X_1 \geq x_1 | X_2 = x_2, s^2 = t) f_{S^2|X_2=x_2}(t) dt \\ &= \int_0^\infty \mathbb{P}\left(\frac{1}{g(\sqrt{t})}(\rho\sqrt{t}x_2 + (1 - \rho^2)W) \geq x_1\right) g_{\lambda', \gamma, \delta'}(t) dt, \end{aligned}$$

où $\lambda' = \lambda - 1/2$, $\delta' = \sqrt{\delta^2 + x^2}$ et $W \sim \mathcal{N}(0, 1)$. Nous pouvons ré-exprimer la partie droite de (3.6) comme $\lim_{x \rightarrow \infty} \int_0^\infty \mathbb{P}((1 - \rho^2)W \geq f(x)g(t) - \rho\sqrt{t}x) g_{\lambda', \gamma, \delta'}(t) dt$.

En utilisant le théorème de convergence dominé, on obtient

$$\lambda_U = 1 - \int_0^\infty g_{\lambda', \gamma, \delta'}(t) dt.$$

Soit $\theta_1, \theta_2 > 0$ ou $\theta_1, \theta_2 < 0$, pour notre développement quantitatif, nous avons besoin de la fonction de Bessel modifiée du troisième type, voir [Paolella \(2007\)](#), avec $\lambda \in \mathbb{R}$:

$$K_\lambda(x) = \sqrt{\frac{\pi}{2x}} \exp(-x) \left(1 + o\left(\frac{1}{x}\right)\right), \quad (3.7)$$

d'après l'équation (3.3), nous avons

$$\frac{K_{\lambda+1}(x)}{K_\lambda(x)} = 1 + \frac{8\lambda + 4}{8x} + \frac{32\lambda^2 - 8}{(8x)^2} + o\left(\frac{1}{x^3}\right), \quad (3.8)$$

$$\frac{K_{\lambda+1}(x)K_{\lambda-1}(x)}{(K_\lambda(x))^2} = 1 + \frac{1}{x} + \frac{64 - 256\lambda^2}{(8x)^3} + o\left(\frac{1}{x^4}\right). \quad (3.9)$$

La densité de X_1 est définie par

$$f_{X_1}(x_1) = \frac{k_{\lambda-\frac{1}{2}}(\delta^2 + (x_1 - \mu_1)^2, \alpha^2)}{\sqrt{2\pi}k_\lambda(\delta^2, \alpha^2 - \theta_1^2)} e^{\theta_1(x_1 - \mu_1)}, \quad (3.10)$$

où

$$K_\lambda(\delta^2, \alpha^2 - \beta^2) = \int_0^\infty x^{\lambda-1} e^{-\frac{1}{2}\left((\alpha^2 - \beta^2)x + \frac{\delta^2}{x}\right)} dx.$$

La densité conditionnelle de $X_{2,1} = X_2|X_1 = x_1$ est égal à

$$\begin{aligned} f_{X_2|X_1}(x_2|x_1) &= \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)} \\ &= \frac{k_{\lambda-1}(\delta^2 + a(x_1, x_2), \alpha^2)}{\sqrt{2\pi}k_{\lambda-\frac{1}{2}}(\delta^2 + (x_1 - \mu_1)^2, \alpha^2)} e^{\theta_2(x_2 - \mu_2)}, \end{aligned}$$

où

$$a(x_1, x_2) = \frac{1}{(1 - \rho^2)} \left((x_1 - \mu_1)^2 + 2(x_1 - \mu_1)(x_2 - \mu_2) + (x_2 - \mu_2)^2 \right).$$

On constate que $X_{2.1}$ suit une loi hyperbolique généralisée. On sait que la distribution conditionnelle de $X_{2.1}|W = w$ est normale $N(\mu_2 + \theta_2 w, w)$, où $W \sim GIG\left(\lambda - \frac{1}{2}, \delta^2 + (x_1 + \mu_1)^2, \alpha^2 + \theta_2^2\right)$.

Dans cette partie on va utiliser l'inégalité de Tchebychev, Nous allons donc calculer les deux premiers moments. On a

$$\begin{aligned} \mathbb{E}(X_{2.1}) &= \mu + \theta_2 \mathbb{E}(W) \\ &= \mu + \theta_2 \frac{\delta^2 + (x_1 + \mu_1)^2}{\alpha^2 + \theta_2^2} \frac{K_{\lambda + \frac{1}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))}{K_{\lambda - \frac{1}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))}, \end{aligned}$$

On peut facilement calculer la variance de $X_{2.1}$ à partir de la formule de décomposition de la variance et obtenir

$$\begin{aligned} Var(X_{2.1}) &= \theta_2^2 \left(\frac{\delta^2 + (x_1 + \mu_1)^2}{\alpha^2 + \theta_2^2} \right)^2 \left[\frac{K_{\lambda + \frac{1}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))}{K_{\lambda - \frac{1}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))} \right]^2 \\ &\quad \times \left(\frac{K_{\lambda + \frac{3}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))}{K_{\lambda - \frac{1}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))} - 1 \right) \\ &\quad + \theta_2 \frac{\delta^2 + (x_1 + \mu_1)^2}{\alpha^2 + \theta_2^2} \frac{K_{\lambda + \frac{1}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))}{K_{\lambda - \frac{1}{2}}((\alpha^2 + \theta_2^2)(\delta^2 + (x_1 + \mu_1)^2))}, \end{aligned}$$

Par l'inégalité de Tchebychev, pour tout positif ϵ ,

$$\begin{aligned} \mathbb{P}(|X_{2.1} - \mathbb{E}(X_{2.1})| \geq \epsilon |x_1|) &\leq \frac{Var(X_{2.1})}{\epsilon^2 |x_1|^2} \\ &= \frac{1}{\epsilon^2 |x_1|^2} \left[(\delta^2 + (x_1 - \mu_1)^2) \left(\frac{\alpha^2 + 2\theta_2^2}{(\alpha^2 + \theta_2^2)^3} \right) \right. \\ &\quad \left. + \lambda \left(\frac{\alpha^2 + 3\theta_2^4}{(\alpha^2 + \theta_2^2)^4} \right) + o\left(\frac{1}{|x_1|}\right) \right] = o\left(\frac{1}{|x_1|}\right). \end{aligned}$$

On combine la variance de $X_{2.1}$ avec (3.8) et (3.9), lorsque $x_1 \rightarrow -\infty$. On trouve

$$\mathbb{P}(\mathbb{E}(X_{2.1}) - \epsilon |x_1| \leq X_{2.1} \leq \mathbb{E}(X_{2.1}) + \epsilon |x_1|) \geq 1 - o\left(\frac{1}{|x_1|}\right), \text{ as } x_1 \rightarrow -\infty,$$

d'où $\lim_{x_1 \rightarrow -\infty} \mathbb{P}(X_2 \leq F_2^{-1}(F_1(y)) | X_1 = x_1) = 0$ si

$$\begin{aligned} F_2^{-1}(F_1(x_1)) &< \mathbb{E}(X_{2.1}) - \epsilon |x_1| \\ &= \left(\mu_2 - \frac{\theta_2}{(\alpha^2 + 2\theta_2^2)^2} + \epsilon \right) x_1 + \lambda \frac{\theta_2}{(\alpha^2 + 2\theta_2^2)^2} + o\left(\frac{1}{|x_1|}\right), \end{aligned}$$

On utilise l'espérance mathématique de la variable aléatoire $X_{2,1}$ et (3.8), quand $x_1 \rightarrow -\infty$, pour montrer

$$F_1(x_1) < F_2 \left(\left(\mu_2 - \frac{\theta_2}{(\alpha^2 + 2\theta_2^2)^2} + \epsilon \right) x_1 + \lambda \frac{\theta_2}{(\alpha^2 + 2\theta_2^2)^2} + o \left(\frac{1}{|x_1|} \right) \right). \quad (3.11)$$

Alors pour $i = 1, 2$,

$$\begin{aligned} F_i(x_1) &= \mathbb{P}(X_i \leq x_1) \\ &= \frac{x^{\lambda - \frac{3}{2}} e^{(\theta_i^2 + \alpha^2 + \theta_i)x_1}}{2K_\lambda(\delta^2, \alpha^2 - \theta_1^2)(\theta_i^2 + \alpha^2)^{(\lambda - 1/2)/2}(\theta_i^2 + \alpha^2 + \theta_i)} \left(1 + o \left(\frac{1}{|x_1|} \right) \right). \end{aligned} \quad (3.12)$$

En combinant (3.12) et (3.11), nous avons

$$(\theta_1^2 + \alpha^2 + \theta_1) > (\theta_2^2 + \alpha^2 + \theta_2) \left(\mu_2 - \frac{\theta_2}{\sqrt{\theta_1^2 + b^2}} + \epsilon \right).$$

Comme $\epsilon > 0$ est arbitraire, nous obtenons

$$\frac{\theta_1^2 + \alpha^2 + \theta_1}{\theta_2^2 + \alpha^2 + \theta_2} - \left(\mu_2 - \frac{\theta_2}{(\alpha^2 + 2\theta_2^2)^2} \right) > 0. \text{ si } \theta_1 = \theta_2 = \theta \in \mathbb{R}^*, \text{ alors}$$

$$1 - \mu_2 + \frac{\theta}{(\alpha^2 + 2\theta^2)^2} > 0. \text{ Donc}$$

$$\lim_{x_1 \rightarrow +\infty} \mathbb{P}(X_2 \geq F_2^{-1}(F_1(y)) | X_1 = x_1) = 1. \text{ De même, on peut montrer que}$$

$$\lim_{x_1 \rightarrow +\infty} \mathbb{P}(X_1 \geq F_1^{-1}(F_2(y)) | X_2 = x_1) = 0 \text{ et en fin on trouve } \lambda_U = 1.$$

COEFFICIENT DE DÉPENDANCE EXTRÉMALE

1. Introduction

L'étude des valeurs extrêmes d'un processus passe naturellement par l'analyse du maximum d'un échantillon de taille n donnée : $M_n = \max(X_1, \dots, X_n)$ où $X_1, \dots, X_n$ est un échantillon i.i.d. de loi F .

Si on note $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$, alors $M_n \rightarrow x_F$ presque sûrement si $n \rightarrow \infty$.

D'autre part, la distribution de M_n est connue exactement :

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

1.1. Théorème de Fisher et Tippet

S'il existe des suites de réels $(a_n)_{n \in \mathbb{N}} > 0$, $(b_n)_{n \in \mathbb{N}}$ et une distribution non dégénérée G telles que :

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x)$$

Alors G est dans le domaine d'attraction de l'une des trois lois suivantes :

- Gumbel

$$G(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}$$

- Fréchet

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

- Weibull

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

Ces trois familles possibles de lois sont appelées les distributions des valeurs extrêmes. Il est possible de les écrire sous forme d'une seule fonction de répartition, la distribution des valeurs extrêmes généralisée (distribution GEV), dont la forme standard est :

$$G_\gamma(x) = \begin{cases} \exp\left(-(1 + \gamma x)^{\frac{-1}{\gamma}}\right) & \text{si } \gamma \neq 0, \forall x \text{ tel que } 1 + \gamma x > 0, \\ \exp(-\exp(-x)) & \text{si } \gamma = 0, \forall x \in \mathbb{R}. \end{cases}$$

Il est ainsi clair que le comportement de la queue de distribution dépend d'un unique paramètre γ , appelé indice des valeurs extrêmes. Le signe de ce Paramètre définit le domaine d'attraction auquel appartient le maximum des observations M_n .

En effet :

- $\gamma > 0$ correspond au domaine d'attraction de Fréchet. Toutes les lois appartenant au domaine d'attraction de Fréchet sont appelées lois à queues lourdes. On a alors $x_F = \infty$ et F le type de Pareto, *i.e.* $1 - F(x) = x^{-\frac{1}{\gamma}} \ell_F(x)$ où ℓ_F est une fonction à variations lentes à l'infini.

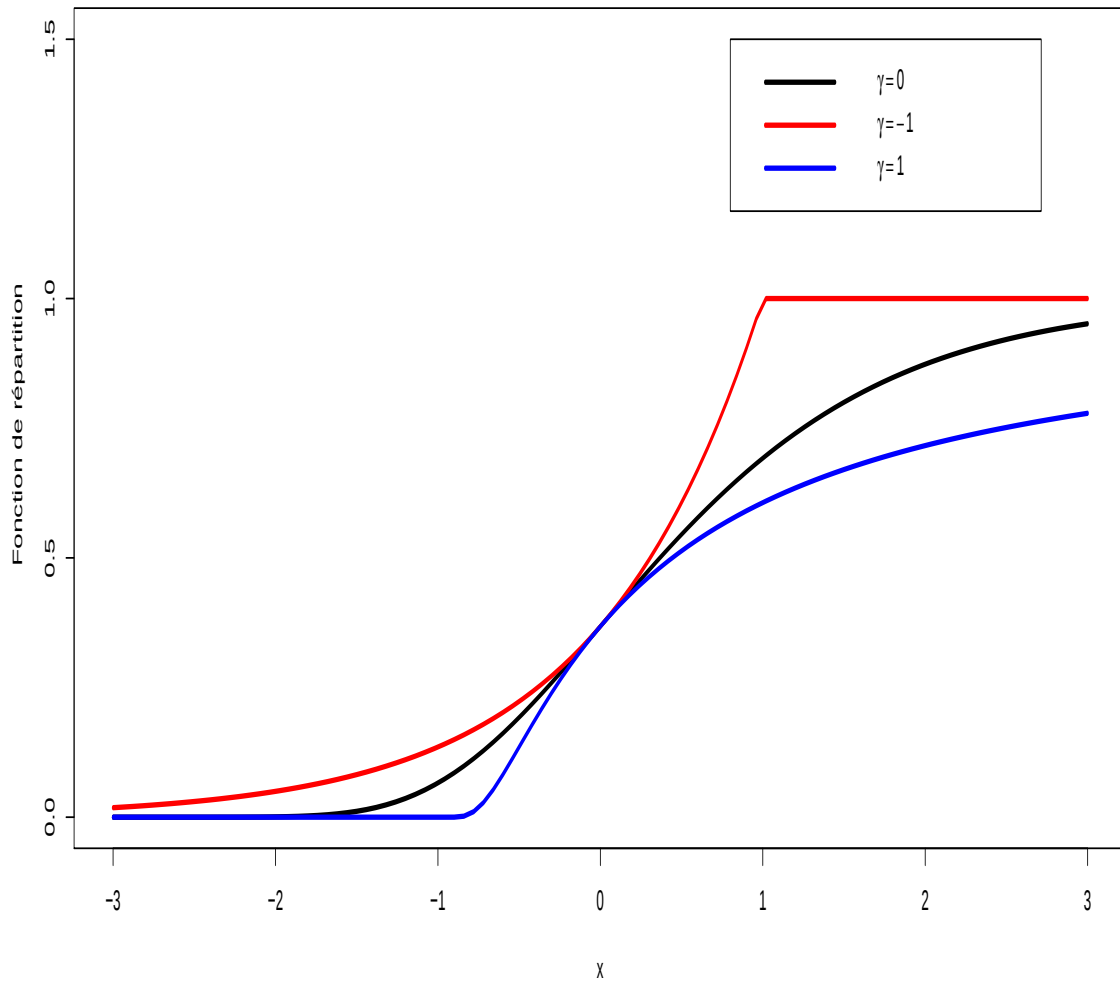


Figure 4.1. *Fonction de répartition des lois de valeurs extrêmes, avec $\gamma = 1$ pour la loi de Fréchet, $\gamma = 0$ pour la loi de Gumbel et $\gamma = -1$ pour la loi de Weibull.*

- $\gamma = 0$ correspond au domaine d'attraction de Gumbel. Le point terminal x_F peut alors être fini ou non, et les queues des lois appartiennent à ce domaine décroissent de manière exponentielle.
- $\gamma < 0$ correspond au domaine d'attraction de Weibull, caractérisé par

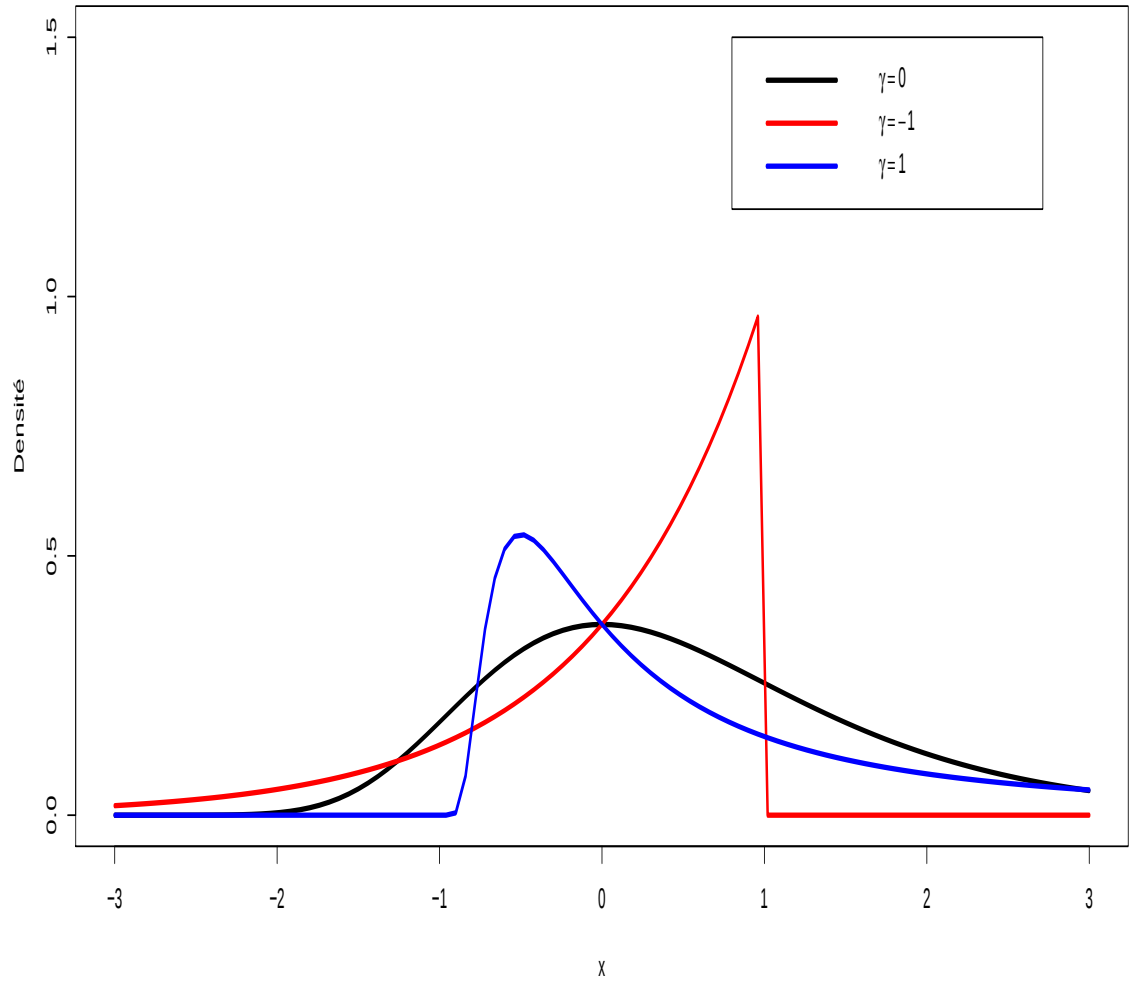


Figure 4.2. Densité des lois de valeurs extrêmes, avec $\gamma = 1$ pour la loi de Fréchet, $\gamma = 0$ pour la loi de Gumbel et $\gamma = -1$ pour la loi de Weibull.

$x_F < \infty$ et $1 - F(x) = (x_F - x)^{-\frac{1}{\gamma}} \ell_F((x_F - x)^{-1})$ où ℓ_F est une fonction

à variations lentes à l'infini.¹

Domaine d'attraction	Gumbel $\gamma = 0$	Weibull $\gamma < 0$	Fréchet $\gamma > 0$
Loi	Normale Exponentielle Lognormale Gamma Weibull	Uniforme Bêta	Cauchy Pareto Student Burr

Table 4.1. Quelques exemples de lois classées selon leur domaine d'attraction.

2. Mesures de dépendance extrême

L'analyse de valeurs extrêmes est fréquemment appliquée pour modéliser des données environnementales, pour lesquelles le phénomène de dépendance est souvent intrinsèque. La dépendance intervient, par exemple, quand différents processus que l'on étudie ont un comportement stochastique qui est lié, tels que les conditions météorologiques communes à titre d'exemple. La dépendance peut aussi surgir quand un seul processus est étudié sur différentes stations spatiales ou bien en termes de son évolution temporelle.

Le coefficient de dépendance de queue supérieure de deux variables aléatoires continues X et Y ayant pour fonction de distribution respectives F_X et F_Y , est défini par la limite (si elle existe) :

$$\lambda_U = \lim_{u \rightarrow 1} \mathbb{P}(X > F_X^{-1}(u) | Y > F_Y^{-1}(u))$$

où $F^{-1}(t) = \inf\{x | F(x) \geq t\}$

et le coefficient de dépendance de queue inférieure est défini par la limite (si

1. Une fonction $\ell(\cdot)$ positive est à variations régulières d'indice $\delta \in \mathbb{R}$ à l'infini ($\ell(\cdot) \in \mathcal{RV}_\delta$), si pour tout $t > 0$, $\lim_{x \rightarrow \infty} \frac{\ell(tx)}{\ell(x)} = t^\delta$.

Si $\delta = 0$, on dit que la fonction $\ell(\cdot)$ est à variations lentes ($\ell(\cdot) \in \mathcal{RV}_0$).

elle existe) :

$$\lambda_L = \lim_{u \rightarrow 0} \mathbb{P}(X \leq F_X^{-1}(u) | Y \leq F_Y^{-1}(u)).$$

Cette mesure a été étudiée par [Ledford *et al.* \(1996\)](#), [Coles *et al.* \(1999\)](#) et [Poon *et al.* \(2004\)](#) et [Aleiyouka, Berred et Ahsanullah \(2017\)](#), parmi d'autres. Deux variables sont dites asymptotiquement dépendantes quand $\lambda_L > 0$ et asymptotiquement indépendantes quand $\lambda_L = 0$.

Clairement, λ_L est une mesure du degré de dépendance dans les extrêmes. Cette mesure de dépendance asymptotique n'est pas sans inconvénients, cependant. Par exemple, pour deux variables aléatoires normales avec le coefficient de corrélation $\rho = 1$, on a $\lambda_U = \lambda_L = 0$, indépendamment de la valeur de u . En outre, les variables sont dépendantes, pourvu que $\rho = 0$. En faisant face à cette difficulté, [Ledford *et al.* \(1996\)](#) et [Coles *et al.* \(1999\)](#) ont proposé un complémentaire statistique, qui mesure l'indépendance asymptotique.

Notre objectif est l'étude des mesures de dépendance extrême pour des variables aléatoires bi-variées (X, Y) . On suppose que les distributions marginales de X et Y sont identiques, une mesure de dépendance est

$$\chi = \lim_{u \rightarrow 1} \chi(u) = 2 - \lim_{u \rightarrow 1} \frac{\log C(u, u)}{\log(u)}$$

Tel que

$$\chi(u) = 2 - \frac{\log \mathbb{P}(U < u, V < u)}{\log \mathbb{P}(U < u)}$$

La fonction χ peut, non seulement fournir la limite, mais aussi, elle peut être interprétée comme une mesure de dépendance. Plus particulièrement, le signe de $\chi(u)$ peut déterminer si les variables sont positivement ou négativement associées au niveau u de quantile, avec des limites

$$2 - \frac{\log(2u - 1)}{\log(u)} \leq \chi(u) \leq 1,$$

où la borne inférieure est interprétée comme $-\infty$ pour $u < 1/2$, et 0 pour $u = 1$.

La mesure de dépendance extrême, χ est la probabilité qu'une variable soit extrême sachant que l'autre l'est. Dans le cas de $\chi = 0$, on dit que les variables sont asymptotiquement indépendantes. L'importance de cette classe a été reconnue dès [Geffroy \(1958/59\)](#), [Sibuya \(1960\)](#), [Tiago de Oliveira \(1962/63\)](#) et [Mardia \(1964\)](#).

De plus, l'analyse empirique de données réelles mène souvent à des évaluations de $\chi = 0$. Malgré cela, la méthodologie standard pour des extrêmes bi-variés est basée sur des distributions pour lesquelles $\chi > 0$, ou le cas spécial d'indépendance exacte, pour lequel $\chi = 0$, suggérant une limitation dans l'applicabilité de méthodes standard.

Appliquant des modèles pour lesquelles $\chi > 0$ aux données asymptotiquement indépendantes mène à la sur-estimation des probabilités d'événements extrêmes communs, puisqu'il y a une supposition mal-placée que les événements marginaux les plus extrêmes puissent arriver simultanément.

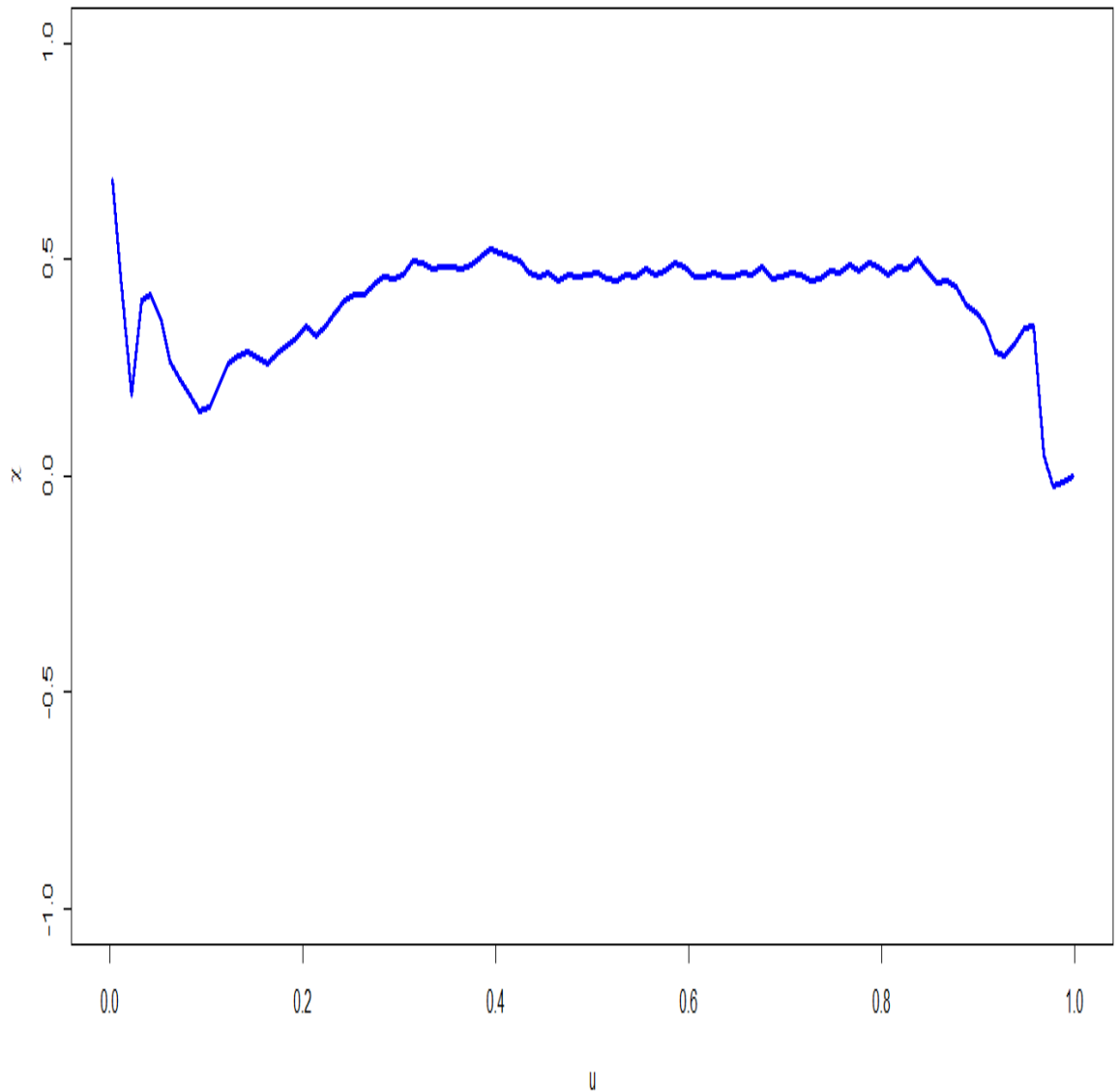


Figure 4.3. χ avec $n = 500$ observations

De récentes recherches ont mis en évidence l'importance de la classe de distributions asymptotiquement indépendantes dans la modélisation de valeurs extrêmes multivariées. Par définition, $\chi = 0$ dans cette classe, si la mesure χ ne peut pas fournir des informations sur la force relative de dépendance pour de tels modèles.

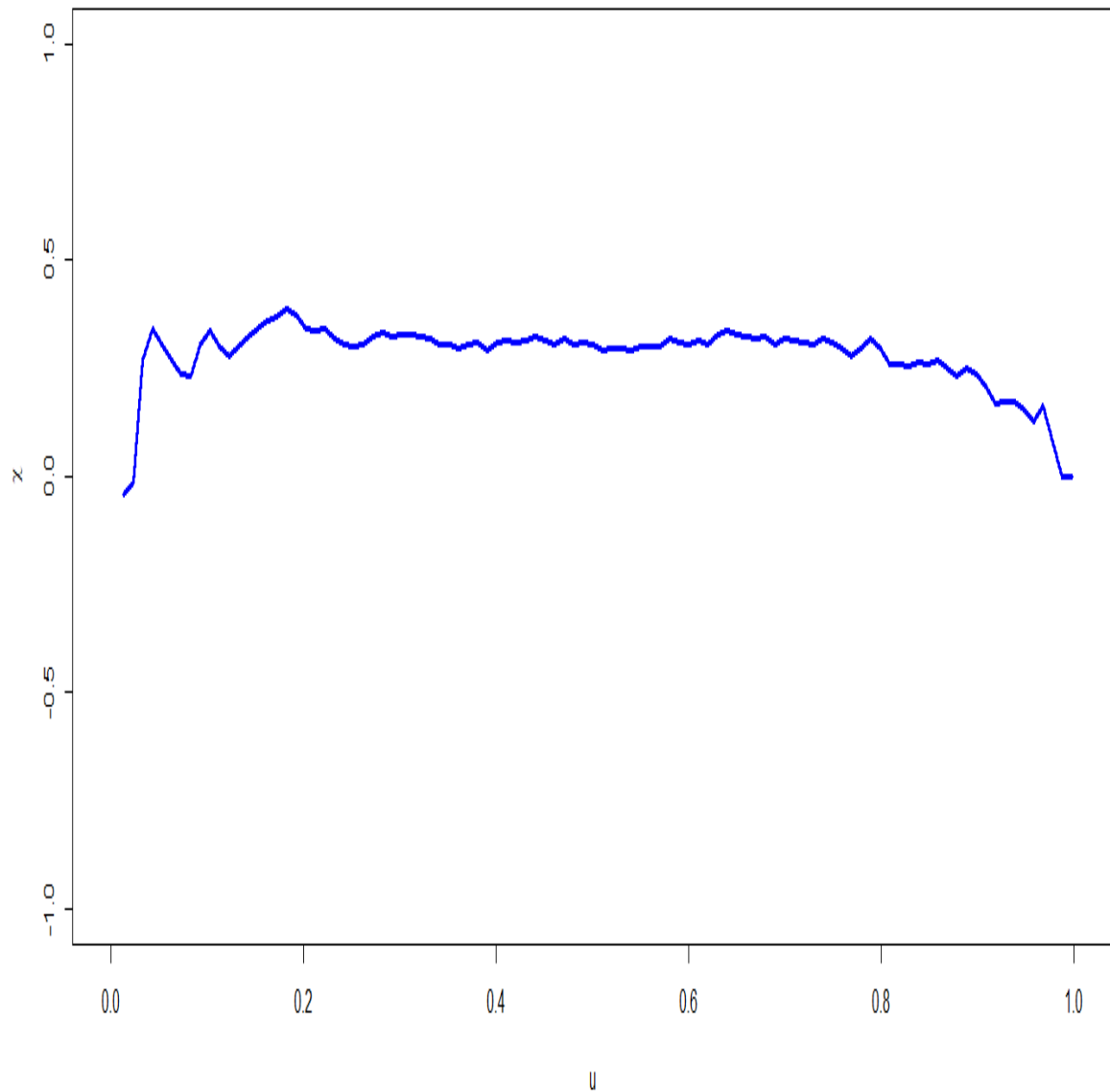


Figure 4.4. χ avec $n = 2000$ observations

On peut surmonter ces limitations en définissant une deuxième mesure de dépendance. En effet

$$\bar{\chi} = \lim_{u \rightarrow 1} \frac{2 \log \mathbb{P}(U > u)}{\log \mathbb{P}(U > u, V > u)} - 1 = \lim_{u \rightarrow 1} \frac{2 \log (1 - u)}{\log \bar{C}(u, u)} - 1$$

Tel que

$$\bar{\chi}(u) = \frac{2 \log \mathbb{P}(U > u)}{\log \mathbb{P}(U > u, V > u)} - 1$$

avec $-1 < \bar{\chi}(u) \leq 1$ pour tout $0 \leq u \leq 1$ et $\bar{C}(u, u) = 1 - 2u + C(u, u)$.

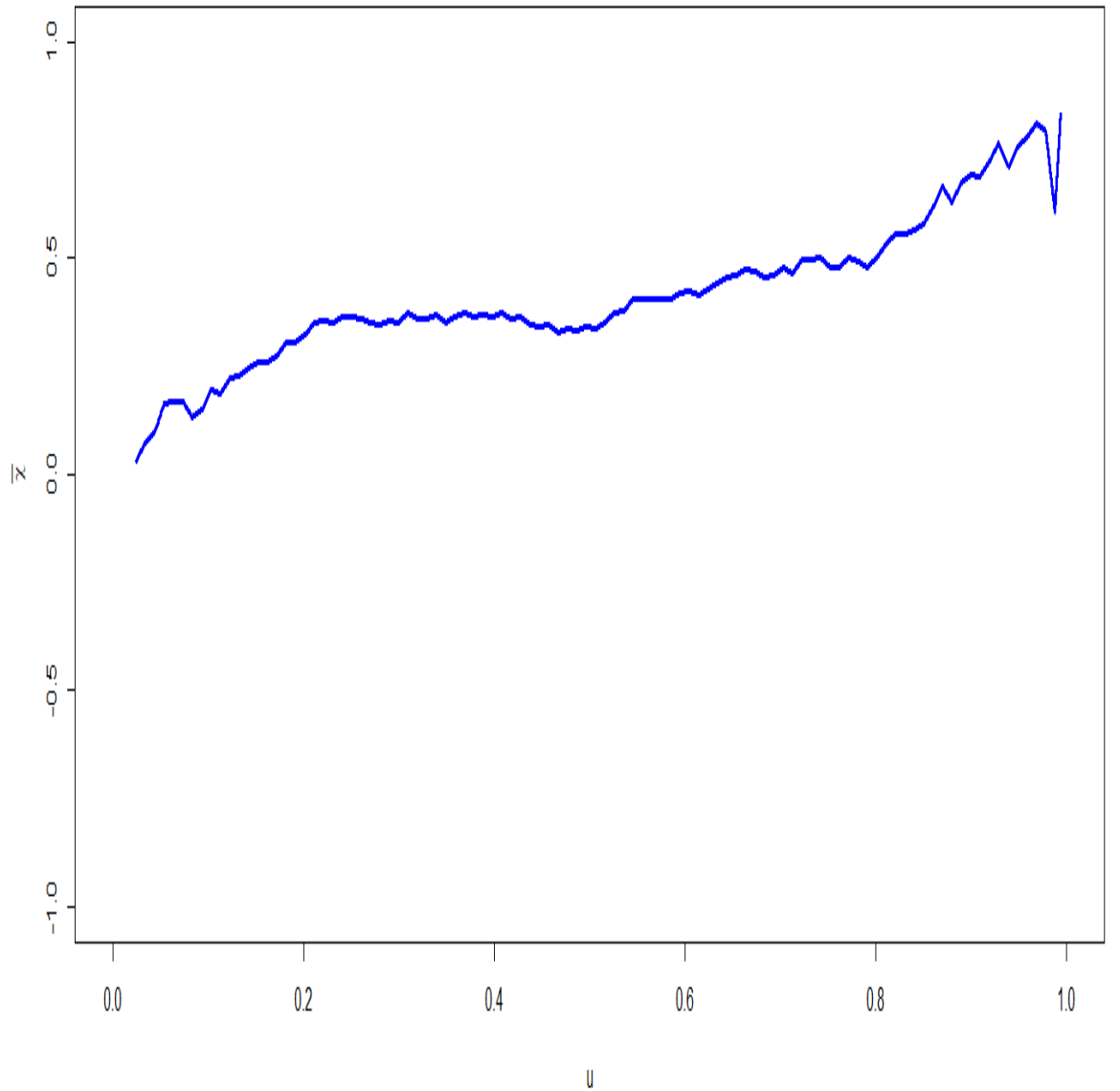


Figure 4.5. La fonction $\bar{\chi}$ avec $n = 500$ observations

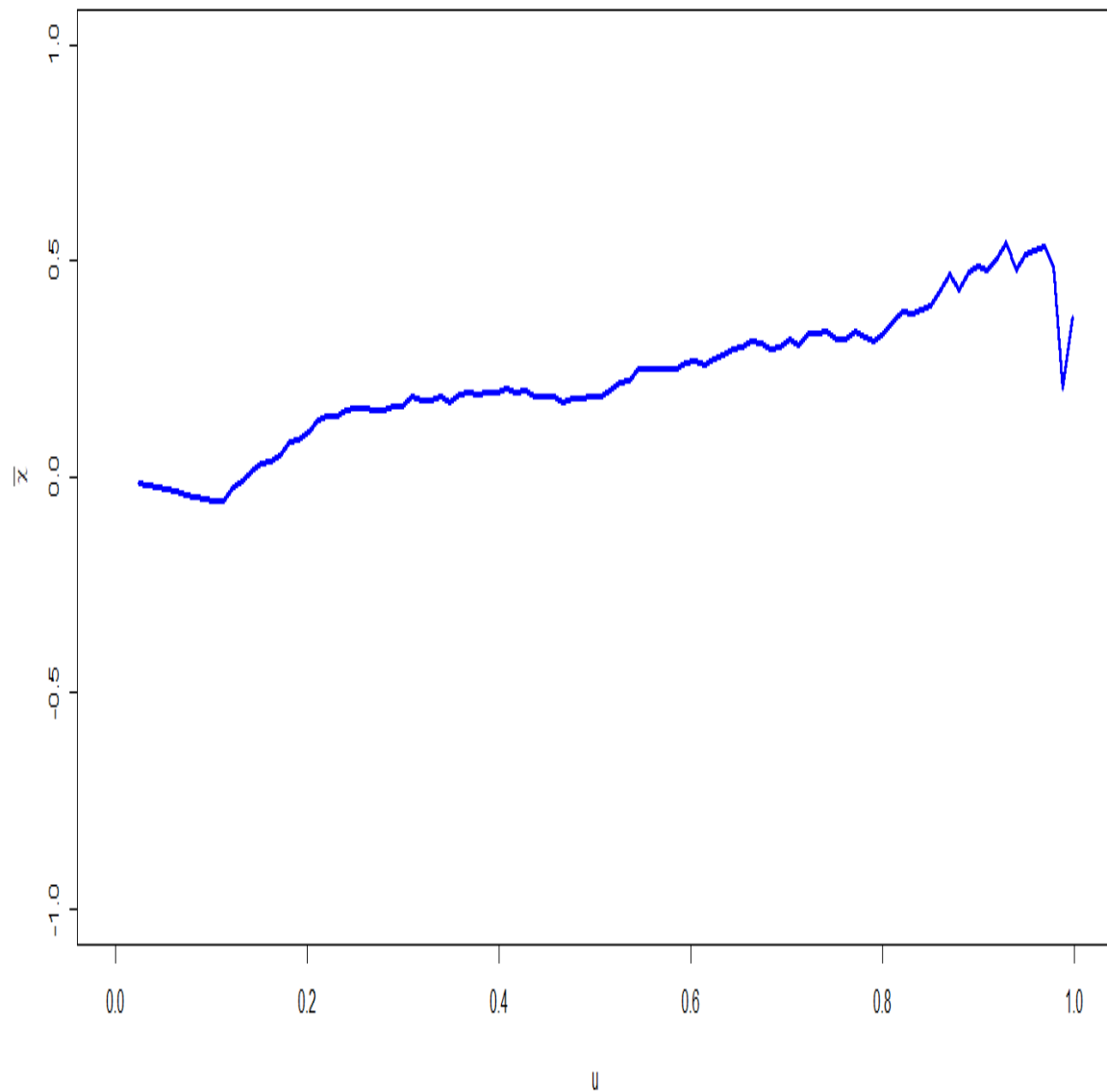


Figure 4.6. $\bar{\chi}$ avec $n = 2000$ observations

Les deux mesures de dépendance $(\chi, \bar{\chi})$ fournissent ensemble toute l'information nécessaire pour caractériser la forme et le degré de dépendance extrême. Pour des variables asymptotiquement dépendantes $\chi = 1$ avec le degré de dépendance donné par $\chi > 0$. Et pour des variables asymptotiquement indépendantes $\chi = 0$ avec le degré de dépendance donnée par le

coefficient χ . Il est important de tester en premier si $\chi = 1$ avant de tirer des conclusions de la dépendance asymptotique basée sur les évaluations de χ .

2.1. Lien entre $\bar{\chi}$ et l'indice des valeurs extrêmes γ

[Ledford *et al.* \(1996\)](#), [Ledford et Tawn \(1997\)](#), ont proposé la modélisation suivante pour la loi jointe d'un couple de marginales Fréchet standards

$$\begin{aligned} \mathbb{P}(X > x, Y > y) &\sim \ell(x, y) x^{-c_1} y^{-c_2} \\ &\quad x \rightarrow \infty, \\ &\quad y \rightarrow \infty \end{aligned}$$

avec ℓ fonction à variation lente. $c_1 > 0$, $c_2 > 0$, $c_1 + c_2 \geq 1$. Le coefficient $\gamma = \frac{1}{c_1 + c_2}$ est appelé l'indice des valeurs extrêmes.

Prenant $x = y = z$, on peut écrire :

$$\begin{aligned} \mathbb{P}(X > z, Y > z) &\sim \ell(z) z^{-\frac{1}{\gamma}} \\ &\quad z \rightarrow \infty \end{aligned}$$

et puisque $\mathbb{P}(X > z) = 1 - \exp(-\frac{1}{z}) \sim \frac{1}{z}$ pour z grand, on a aussi

$$\begin{aligned} \mathbb{P}(X > z, Y > z) &\sim \ell(z) [\mathbb{P}(X > z)]^{\frac{1}{\gamma}} \\ &\quad z \rightarrow \infty \end{aligned}$$

où $\ell(\cdot)$ est une fonction à variation lente et $0 < \gamma \leq 1$.

Soit $C(\cdot, \cdot)$ la copule associée à H , où $H(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(F(X) \leq F(x), F(Y) \leq F(y)) = C(F(x), F(y))$. On a

$$\begin{aligned} \bar{C}(u, u) &\sim \ell((1-u)^{-1}) (1-u)^{-\frac{1}{\gamma}} \\ &\quad u \rightarrow 1 \end{aligned}$$

$$\bar{\chi} \sim \frac{2 \log(1-u)}{\log \ell((1-u)^{-1}) + \frac{1}{\gamma} \log(1-u)} - 1 \rightarrow 2\gamma - 1 \quad u \rightarrow 1$$

On a donc,

$$\bar{\chi} = 2\gamma - 1$$

Remarques :

1. Si $\gamma = 1$ et $\ell(z) \rightarrow c$, $0 < c \leq 1$ alors

a) $\chi = c$. En effet, $\chi = \lim_{u \rightarrow 1} \frac{\bar{C}(u, u)}{1 - u} = \lim_{u \rightarrow 1} \ell((1 - u)^{-1}) = c$.

b) $\bar{\chi} = 1$

2. Si $\gamma = \frac{1}{2}$, $\bar{\chi} = 0$ donc indépendance asymptotique.

Pour étudier le comportement de dépendance de queue de vecteurs aléatoires, nous avons non seulement l'intérêt dans les cas où le vecteur aléatoire est asymptotiquement dépendant, mais aussi dans le cas où l'indépendance asymptotique intervient. [Ledford et al. \(1996\)](#) ont proposé le modèle suivant pour un vecteur aléatoire bi-varié $(X_1, X_2)^T$, où X_1 et X_2 sont des Fréchet unitaires avec une fonction de distribution $F_i(x) = e^{-\frac{1}{x}}$, $x > 0$, $i = 1, 2$.

3. Coefficient de dépendance extrême

3.1. Définition

Soit X un vecteur aléatoire bi-varié avec la fonction de répartition bi-dimensionnelle F et dont les marges sont F_1 et F_2 . De plus, soit $E_{\min} := \min\{e^{\frac{1}{F_1(X_1)}}, e^{\frac{1}{F_2(X_2)}}\}$ et $E_{\max} := \max\{e^{\frac{1}{F_1(X_1)}}, e^{\frac{1}{F_2(X_2)}}\}$.

Le coefficient de dépendance extrême inférieur de X est défini par

$$\xi = 2 - \lim_{u \rightarrow 1} \mathbb{P}(E_{\max} < u | E_{\min} < u),$$

tandis que le coefficient de dépendance extrême supérieur de X est défini par

$$\bar{\xi} = \lim_{u \rightarrow 1} \mathbb{P}(E_{\max} > u | E_{\min} > u) - 1,$$

si les limites existent.

On note :

$$\begin{aligned}
 \mathbb{P}(E_{\max} \leq t | E_{\min} \leq t) &= \frac{\mathbb{P}(E_{\max} \leq t, E_{\min} \leq t)}{\mathbb{P}(E_{\min} \leq t)} \\
 &= \frac{\mathbb{P}\left(e^{-\frac{1}{F_1(X_1)}} \leq t, e^{\frac{1}{F_2(X_2)}} \leq t\right)}{1 - \mathbb{P}\left(e^{-\frac{1}{F_1(X_1)}} > t, e^{\frac{1}{F_2(X_2)}} > t\right)} \\
 &= \frac{\mathbb{P}\left(F_1(X_1) \leq \frac{1}{\log(\frac{1}{t})}, F_2(X_2) \leq \frac{1}{\log(\frac{1}{t})}\right)}{1 - \mathbb{P}\left(F_1(X_1) > \frac{1}{\log(\frac{1}{t})}, F_2(X_2) > \frac{1}{\log(\frac{1}{t})}\right)}
 \end{aligned}$$

La probabilité $\mathbb{P}(E_{\min} > t | E_{\max} > t)$ se calcule d'une manière similaire

$$\begin{aligned}
 \mathbb{P}(E_{\min} > t | E_{\max} > t) &= \frac{\mathbb{P}(E_{\min} > t, E_{\max} > t)}{\mathbb{P}(E_{\max} > t)} \\
 &= \frac{\mathbb{P}\left(F_1(X_1) > \frac{1}{\log(\frac{1}{t})}, F_2(X_2) > \frac{1}{\log(\frac{1}{t})}\right)}{1 - \mathbb{P}\left(F_1(X_1) \leq \frac{1}{\log(\frac{1}{t})}, F_2(X_2) \leq \frac{1}{\log(\frac{1}{t})}\right)}
 \end{aligned}$$

De même pour le cas supérieur/inférieur. Alors

$$\xi = \lim_{t \rightarrow 1} \frac{\mathbb{P}\left(F_1(X_1) \leq \frac{1}{\log(\frac{1}{t})}, F_2(X_2) \leq \frac{1}{\log(\frac{1}{t})}\right)}{1 - \mathbb{P}\left(F_1(X_1) > \frac{1}{\log(\frac{1}{t})}, F_2(X_2) > \frac{1}{\log(\frac{1}{t})}\right)} = \lim_{t \rightarrow 1} \frac{\log C(t, t)}{1 - \bar{C}(1 - t, 1 - t)}$$

et

$$\bar{\xi} = \lim_{t \rightarrow 1} \frac{\mathbb{P}(F_1(X_1) > t, F_2(X_2) > t)}{1 - \mathbb{P}(F_1(X_1) \leq t, F_2(X_2) \leq t)} = \lim_{t \rightarrow 1} \frac{\log \bar{C}(1 - t, 1 - t)}{1 - \log C(t, t)}$$

où C est la copule de X et $\bar{C}$ est la copule de survie de C . Sachant que X suit une loi de Fréchet unitaire i.e $\mathbb{P}(X \leq \log(x)) = \log(\mathbb{P}(X \leq x))$ alors

$$\xi = \lim_{t \rightarrow 0} \frac{\log C(t, t)}{1 - \log \bar{C}(1 - t, 1 - t)}$$

et

$$\bar{\xi} = \lim_{t \rightarrow 1} \frac{\mathbb{P}(F_1(X_1) > t, F_2(X_2) > t)}{1 - \mathbb{P}(F_1(X_1) \leq t, F_2(X_2) \leq t)} = \lim_{t \rightarrow 1} \frac{\log \bar{C}(1 - t, 1 - t)}{1 - \log C(t, t)}.$$

ÉTUDE STOCHASTIQUE DANS LE SGBD TEMPS RÉEL POUR COMPARER DES TRANSACTIONS PLATES ET ÉTENDUES

1. Introduction

Un Système de Gestion de Bases de Données (SGBD) Temps Réel, a pour objectif principal de respecter, non seulement les contraintes logiques, mais aussi celles temporelles. En effet, chaque transaction doit avoir lieu en son temps imparti, i.e respecter son échéance. Dans la pratique, cet objectif double est parfois délicat à atteindre.

Parmi ces difficultés, se trouvent, entre autres, la liaison assez forte des problèmes posés à l'incapacité de ces systèmes, par rapport à la prédiction des temps d'accès à l'ensemble des données de la base, et aux différents périphériques d'une part, d'autre part, aux causes liées à la complexité du contexte distribué.

Dans ce chapitre, nous allons effectuer une étude probabiliste et statistique des différents aspects de transactions temps réel, à titre d'exemple : probabilité d'accès aux données en mémoire centrale, probabilité de pannes de réseau et/ou de site dans une architecture distribuée ou parallèle.

Le but étant de proposer des modèles probabilistes qui facilitent de maîtriser le comportement des transactions temps réel, afin d'optimiser les performances des SGBD Temps Réel.

2. Les transactions

Le concept de transaction a été introduit dans les SGBD vers les années soixante-dix. Une transaction représente une unité logique de traitement et d'exécution. Elle est constituée par des opérations de lecture et d'écriture. En SGBDTR, les transactions peuvent être des transaction non-temps réel (classiques) ou des transactions temps réel. Ces dernières sont souvent répertoriées, selon leurs fonctionnalités, en transactions de mise à jour (transactions système) et transactions "Utilisateur". Les transactions peuvent être aussi classées selon leur périodicité en transactions périodiques, sporadiques et apériodiques.

2.1. Les transactions de mise à jour

Les transactions de mise à jour ont pour tâche le rafraîchissement de la base de données afin de refléter correctement l'environnement. Elles sont généralement déclenchées de façon périodique pour mettre à jour régulièrement les données acquises depuis des capteurs, mais elles peuvent être déclenchées sur demande.

2.2. Les transactions "Utilisateur"

Les transactions "Utilisateur" effectuent des opérations de lecture/écriture sur les données non temps réel et des opérations de lecture sur les données

temps réel [Ramamritham *et al.*, 1996a]. Les temps d'arrivée et les tailles de ces transactions sont aléatoires, les charges qu'elles induisent sont imprévisibles.

2.3. Transactions non-temps réel

Les transactions non-temps réel (classiques) sont les transactions que nous trouvons dans les SGBD classiques. Aucune contrainte de temps n'est imposée sur l'exécution de ces transactions et sur les données qu'elles manipulent.

2.4. Propriétés des transactions

Une transaction se compose d'une suite de requêtes dépendantes à la base qui se doivent de vérifier les propriétés d'atomicité, de cohérence, d'isolation et de durabilité, résumées par l'acronyme ACID. On cite, d'une brève manière, les propriétés ACID suivantes :

2.4.1. Atomicité

Une transaction suit la loi de tout ou rien. C'est à dire qu'elle doit effectuer toutes ses mises à jour ou ne rien faire du tout. Le système doit pouvoir annuler toutes les modifications qu'elles aient engagées en cas d'échec.

L'atomicité est fortement limitée par un certain nombre de circonstances matérielles et/ou logicielles, à savoir bugs de programme, pannes du système ou du matériel. On parle plus généralement de tout événement susceptible d'interrompre une transaction en cours.

2.4.2. Cohérence

Une transaction doit se faire de manière à passer la base de données d'un état cohérent à un autre. En cas d'échec, la base doit pouvoir restaurer l'état cohérent initial. La cohérence de la base peut être affecté, voire violée, par un programme erroné ou même par un conflit d'accès concurrent entre transactions.

2.4.3. Isolation

Une fois que la transaction se valide, et uniquement dans ce cas, les résultats de celle-ci peuvent être visibles aux autres transactions, et ceci, pour éviter les interférences avec les autres transactions. Les accès concurrents peuvent nuire au principe de l'isolation.

2.4.4. Durabilité

Dès qu'une transaction se valide, le système doit pouvoir sauvegarder ses modifications pour les prendre en compte en cas de panne. Notamment lors des pannes disques qui constituent l'un des problèmes les plus fréquents.

3. Les modèles de transactions étendus

Comme nous les avons cités, les techniques implémentées dans les SGBD permettant d'assurer l'atomicité des transactions sont aujourd'hui assez présentes, bien connues et opérationnelles pour les transactions en gestion.

Cependant, les techniques de gestion de transactions atomiques demeurent trop limitées dans le cas de longues transactions que l'on rencontre en conception (CAO, CFAO), en production de documents (PAO), et plus généralement dans les domaines techniques nécessitant des objets complexes.

De plus, différents modèles de transactions ont été proposés, cherchant à permettre une plus grande flexibilité que "le tout ou rien". Ces modèles étendent plus ou moins les propriétés ACID mentionnées ci-dessus.

4. Les transactions imbriquées (emboîtées)

Le modèle des transactions imbriquées est dû à [Moss \(1985\)](#). Il donne extension au modèle plat, ayant un seul niveau, pour aboutir à une structure de transaction à niveaux multiples. Chaque transaction peut être décomposée en sous-transactions, et cela se fait de manière récursive.

4.1. Définition

Chaque transaction emboîtée est composée de plusieurs sous-transactions qui sont à leur tour composées de sous-transactions et/ou d'opérations, c'est-à-dire commençant par BEGIN et se terminant par COMMIT. une transaction

5. Protocoles d'ordonnancement

Plusieurs protocoles d'ordonnancement ont été proposés dans la littérature pour répondre aux différentes contraintes temporelles imposées dans les applications temps réel. Ceux-ci peuvent être caractérisés différemment : statiques ou dynamiques, préemptifs ou non, à ressource unique ou multiple, ayant un seul ou plusieurs type(s) de transaction, centralisés ou distribués... Dans la suite, nous allons exposer les protocoles les plus classiques et ceux sur lesquels se basent nos travaux.

5.1. Protocole Rate Monotonic (RM)

Les priorités de ce protocole sont constantes et inversement proportionnelles aux périodes des transactions [Liu et Layland \(1973\)](#). C'est-à-dire, que la priorité la plus forte est attribuée à la transaction possédant la plus petite période.

Les conflits entre plusieurs transactions de même priorité sont résolus de façon aléatoire. Les auteurs ont montré que l'algorithme RM est optimal pour des tâches périodiques, synchrones, à départ simultané, échéance sur requête (période = échéance), et indépendantes avec un ordonnanceur préemptif. L'emploi de RM dans les SGBDTR n'est pas bénéfique pour la simple raison que les transactions "Utilisateur" sont apériodiques et leurs tailles et échéances ne sont pas connues à l'avance.

5.2. Protocole Deadline Monotonic (DM)

Le protocole DM est présenté dans [Leung et Whitehead \(1982\)](#). De même que RM, DM est un protocole d'ordonnancement statique, c'est-à-dire, les priorités sont constantes. Cependant, la priorité est ici inversement proportionnelle au délai critique, i.e., la transaction la plus prioritaire est la transaction qui possède le plus petit délai critique.

Les transactions doivent être exécutées simultanément, et avoir des délais critiques inférieurs à leurs périodes. Les conflits entre les transactions de même priorité sont résolus aléatoirement.

DM présente une performance similaire que celle de RM pour les tâches à échéance sur requête et donne par contre de meilleures performances lorsque les délais critiques des tâches sont inférieurs à leurs périodes.

5.3. Earliest Dealine First (EDF)

EDF est un protocole dynamique introduit dans [Liu et Layland \(1973\)](#). Il assigne la priorité la plus élevée à la transaction ayant l'échéance la plus proche. Il s'agit de l'algorithme le plus utilisé en raison de sa simplicité et de sa performance pour les transactions à départs simultanés et à départs différés.

Néanmoins, il est sous entendu que les performances d'EDF se dégradent rapidement lorsque le système devient chargé. Ainsi des protocoles comme AED, AVED et GEDF ont été proposés pour améliorer ses performances. Dans cette thèse, nous nous penchons particulièrement sur GEDF que nous présentons en détail dans la section 1.3.6.

5.4. Adaptive Earliest Deadline (AED)/ Hierarchical Earliest Deadline (HED)

Dans [Haritsa *etal* \(1991\)](#), [Han *et al.* \(2012\)](#), les auteurs ont proposé le protocole AED, qui est basé sur EDF, afin d'améliorer les performances des

systèmes lors de surcharges. Dans AED, à leurs arrivée dans le système, les transactions sont divisées en deux groupes : (Hit) et (Miss). Une transaction est introduite dans une structure en file ordonnée si sa position est inférieure à la capacité de groupe (Hit), elle est alors assignée à ce groupe ; elle est assignée au groupe (Miss) le cas contraire.

En effet, Dans le groupe (Hit), les priorités sont assignées selon EDF, alors que dans le groupe (Miss) les priorités sont assignées aléatoirement. Les transactions du groupe (Miss) ne sont exécutées que lorsque toutes les transactions du groupe (Hit) sont terminées. Les auteurs ont proposé aussi le protocole Hierarchical Earliest Deadline HED, une version étendue du protocole AED.

Ce protocole maintient une liste dynamique de (buckets) où chaque (bucket) est caractérisé par deux attributs : *MinValue* et *MaxValue*. Lorsqu'une nouvelle transaction T arrive dans le système, elle est affectée au (bucket) qui satisfait la contrainte suivante : $MinValue < ValeurT \leq MaxValue$. De même que dans AED, les transactions de chaque (bucket) sont divisées à leur tour en deux groupes (Hit) et (Miss).

5.5. Protocole Adaptive Earliest Virtual Deadline (AEVD)

AEVD est proposé pour permettre un partage équitable des ressources par les transactions [Pang et al. \(1992\)](#). Il combine à la fois l'échéance et l'instant d'arrivée. De même que AED, les transactions sont divisées en deux groupes (Hit) et (Miss). Cependant, ce protocole utilise les échéances virtuelles pour ordonnancer les transactions du groupe (Hit) au lieu des échéances dans AED. L'échéance virtuelle est calculée selon la formule suivante : $V_T = (D_T - Clock) * PF_T + Clock$ où D_T représente l'échéance de la transaction et PF_T (pace factor) sa valeur, comprise entre 0 et 1. À chaque fois que l'échéance virtuelle expire, une nouvelle échéance est attribuée à la transaction.

5.6. Protocole Generalized Earliest Deadline First (GEDF)

GEDF est un protocole d'ordonnancement dynamique proposé dans [Semghouni et al. \(2007\)](#) pour combler les insuffisances d'EDF et améliorer les performances du système. Pour cela, il propose la prise en compte de l'importance de transaction (*Spriority*) lors de l'ordonnancement de transactions, en plus de l'échéance. La priorité est calculée par la formule suivante :

$$Priority(T) = (1 - a) \times Deadline(T) + a \times SPriority(T) \quad (5.1)$$

où $0 \leq a \leq 1$. a représente le poids de *Spriority* dans la formule.

Dans GEDF, les transactions qui ont une échéance stricte non critique (firm) sont prises en considération. Une transaction à échéance stricte non critique est écartée (abandonnée ou annulée) par le système si elle n'arrive pas à terminer avant son échéance. À la différence des protocoles précédents, GEDF classe les transactions en deux groupes : les transactions "Utilisateur" et les transactions de "mise à jour".

- Transactions de mise à jour : elles sont des transactions système périodiques dont le but est de maintenir la cohérence du système avec son environnement. La périodicité d'une transaction de mise à jour dépend de la donnée à laquelle elle est rattachée. Ces travaux adoptent l'approche more-less, proposée par ([Xiong et Ramamritham 2004](#)), afin de minimiser la charge soumise au système, tout en garantissant la fraîcheur des données.
- Transactions "Utilisateur" : ce sont des transactions apériodiques. Elles accèdent aux données classiques en mode lecture et écriture et aux données temps réel en mode lecture seulement.

Dans GEDF, à chaque transaction est rattaché le paramètre priorité système (*Spriority*). Ce paramètre permet de définir le rang de la transaction par rapport aux autres. Il tient compte de l'importance des transactions lors de l'ordonnancement. En effet, l'importance est utilisée par GEDF, voir l'équation (5.2), en plus de son échéance, pour le calcul de la priorité de la transaction. La valeur de *Spriority* est comprise dans l'intervalle $[0, MaxValue]$.

La transaction est d'autant plus importante que sa *Spriority* est proche de 0. Cet intervalle est divisé en deux sous-intervalles $[0, N]$ et $]N, MaxValue]$, où N est une valeur définie par l'administrateur système : l'intervalle $[0, N]$ contient les valeurs de *Spriority* des transactions de mise à jour. Nous émettons l'hypothèse que les transactions de mise à jour sont plus importantes parce qu'elles maintiennent la cohérence temporelle des données et permettent de refléter plus fidèlement l'environnement (Xiong et Ramamritham 2004). L'intervalle $]N, MaxValue]$ contient les valeurs de *Spriority* des transactions "Utilisateur". Les valeurs de *Spriority* sont calculées par les formules (5.2) et (5.3).

- Transactions de mise à jour : soient *MaxPeriode*, la plus longue période de mise à jour et *Periode_T* la période de la transaction T . Les transactions avec les périodes les plus courtes sont considérées comme les plus importantes. Elles manipulent des données plus sensibles (pression, température dans une centrale nucléaire) qui doivent être mises à jour fréquemment. La valeur de *Spriority* est définie comme suit :

$$SPriority_{update} = N \times \frac{Periode_T}{MaxPeriode} \quad (5.2)$$

où *Periode_T* est la période de la transaction T , et *MaxPeriode* est la plus longue période définie.

- Transactions "Utilisateur" : la valeur de *Spriority* pour les transactions "Utilisateur" est attribuée selon le nombre et le type (écriture (write) et lecture (read)) des opérations qu'elles contiennent. Soit T_{user} une transaction "Utilisateur", *Spriority* est définie comme suit :

$$SPriority_{T_{user}} = MaxValue - \gamma \times Weight_{T_{user}} - (1 - \gamma) \times DBAValue \quad (5.3)$$

où

1. $\text{Weight}_{T_{\text{user}}}$ est le poids de la transaction "Utilisateur" donné par la formule suivante :

$$\text{Weight}_{T_{\text{user}}} = \left(\sum_{i=1}^n \text{Wread}_{\text{TD}} + \sum_{j=1}^m \text{Wwrite}_{\text{TD}} - \sum_{j=1}^l \text{Wread}_{\text{NTD}} \right) \quad (5.4)$$

avec

Wread_{TD} , $\text{Wwrite}_{\text{NTD}}$ et $\text{Wread}_{\text{NTD}}$ représentent les poids assignés respectivement aux opérations de lecture des données temps réel, aux opérations d'écriture des données non temps réel et aux opérations de lecture des données non temps réel.

n , m , et l représentent respectivement le nombre des opérations (Read_{TD} / $\text{Write}_{\text{NTD}}$ / Read_{NTD}) dans chaque transaction.

2. $\gamma \in [0, 1]$ est la valeur assignée au poids de chaque transaction (voir Formule (5.4)).
3. $\text{DBA}_{\text{Value}}$ est une variable aléatoire entre 0 et $(\text{MaxValue} - N)$.

Dans [Semghouni et al. \(2007\)](#), les auteurs ont montré que GEDF donne de meilleurs résultats comparé à EDF lorsque le système commence à être chargé. Dans un premier temps, nous approfondissons ces travaux afin d'améliorer les performances du système en proposant les valeurs optimales à assigner aux différents paramètres sous différentes conditions. Ensuite, nous étudions les performances de GEDF en utilisant les transactions emboîtées.

6. Étude stochastique dans les SGBDTR

Les études probabilistes des SGBD se sont beaucoup plus intéressées aux données incertaines. En effet, la gestion de telles données fut étudiée dans la littérature des SGBD traditionnels depuis la fin des années 80 [Cavallo and Pittarelli \(1987\)](#), [Barbara et al. \(1990\)](#).

Cependant, l'intérêt manifesté aux données incertaines n'a cessé d'augmenter avec le développement des nouvelles technologies qui ont été à l'origine du déploiement et de la sauvegarde de gros flux de données de façon continue.

On constate assez fréquemment, que ces données contiennent des erreurs ou sont incomplètes ou incompatibles. Par exemple, des facteurs exogènes peuvent générer des erreurs (en provenance des entrées des capteurs, suite à des bruits ou des transmissions sans fils...). Dans d'autres cas, les données peuvent être sous forme de moyennes et de distributions, simplement pour des raisons de confidentialité de celles-ci.

Ainsi, les données incertaines ont tendance à nécessiter une représentation par des valeurs à distribution probabilistes plutôt que par des valeurs déterministes [Aggarwal, Yu \(2009\)](#).

Dans [Green et Tannen \(2006\)](#), une BDD probabiliste est définie comme des espaces probabilisés finis dont les éléments représentent toutes les instances possibles respectant ainsi un schéma bien défini. Elle peut être représentée par la paire (X, p) où X est un ensemble fini d'instances conforme à un schéma donné et $p(I)$ la probabilité associée à chaque $I \in X$. Si $p(\cdot)$ représente le vecteur de probabilité de toutes les instances de X alors $\sum_{I \in X} p(I) = 1$.

L'adoption des données incertaines a induit la définition de nouvelles techniques de représentation, de modélisation, d'intégration, de requêtes, de jointure et d'indexation des données [Aggarwal \(2009\)](#). Dans [Semghouni et al. \(2008\)](#), et par le biais des simulations à la méthode de Monte-Carlo, l'auteur a pu extraire un modèle probabiliste qui exprime le comportement des transactions. Ce modèle permet, plus particulièrement, de prédire des taux de succès de transactions en fonction de la charge du système, offrant ainsi des outils formels pour la gestion efficace de celui-ci.

Ces travaux s'effectuent sur deux plate-formes, dont une première qui ne peut contenir que des données classiques et une seconde dans laquelle peuvent s'y ajouter des données temps réel. Ils utilisent des protocoles de contrôle de concurrence optimistes et pessimistes.

Ce travail s'inscrit donc, dans la continuité et l'approfondissement de ce

sujet. Pour mesurer les performances du système, nous considérons le taux de succès des transactions et le taux de succès des transactions "Utilisateur".

- Le taux de succès des transactions, noté $SRatio$, est donné par la formule suivante :

$$SRatio = \frac{Commit_T}{Submitted_T}$$

où $Commit_T$ désigne le nombre de transactions validées en respectant leur échéance et $Submitted_T$ le nombre de transactions qui sont soumises au système.

- Le taux de succès des transactions "Utilisateur" est défini par

$$SRatio_{User} = \frac{Commit_{User}}{Submitted_{User}}$$

6.1. Étude stochastique

Nous nous intéressons dans ce travail à l'étude de la distribution des fréquences des taux de succès des transactions "Utilisateur" $SRatio_{user}$.

Nous rappelons que nous nous basons principalement sur les travaux de notre équipe qui ont permis de déduire que « La distribution des fréquences du taux de succès des transactions "Utilisateur" peut être approximée par une variable aléatoire de la loi Bêta. Ce résultat est validé par le test statistique de Kolmogorov-Smirnov (K.S.) » [Semghouni et al. \(2008\)](#).

Pour notre étude, nous procédons à l'exécution d'un ensemble d'expériences où chacune d'elles regroupe un cycle de 1000 simulations pour un état de système donné. Une fois ce cycle terminé, nous récupérons les taux de succès constatés via la méthode des moments, nous estimons les valeurs des paramètres p et q de la loi Bêta, dénotés par $\hat{p}_{user}$ et $\hat{q}_{user}$.

Nous vérifions ensuite la correspondance de la distribution obtenue avec la loi Bêta à travers le test statistique de Kolmogorov-Smirnov. Ainsi, chaque expérience nous permet de dégager un couple de valeurs de $\hat{p}_{user}$ et $\hat{q}_{user}$. Sachant que les distributions des fréquences de $SRatio$ Utilisateur peuvent être approximées toujours par la loi Bêta, quelles que soit les valeurs des

paramètres des simulations (protocole d'ordonnancement, distribution des tailles des transactions, taille des transactions, taille de la base, ...), nous nous intéressons au comportement de $\hat{p}_{user}$ et $\hat{q}_{user}$ en fonction de la charge.

On rappelle qu'une variable aléatoire X suit une distribution bêta avec les paramètres p et q si sa densité de probabilité est donnée par :

$$f(x, p, q) = \frac{x^{p-1}(1-x)^{q-1}}{\mathcal{B}(p, q)}, x \in [0, 1],$$

avec

$$\mathcal{B}(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx.$$

En utilisant la méthode des moments, nous estimons les paramètres p et q par :

$$\hat{p}_{user} = \bar{x} \left[\frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right] \quad \text{et} \quad \hat{q}_{user} = (1-\bar{x}) \left[\frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right]$$

$$\text{tel que } \bar{x} = \sum_{i=1}^n xi \text{ et } s^2 = \sum_{i=1}^n (xi - \bar{x})^2.$$

7. Simulations

Afin de démontrer l'influence de la structure des transactions "Utilisateur" sur le taux de succès, on analyse le taux de succès avec la transaction plate et imbriquée. Ensuite, nous nous concentrons sur l'avantage de la politique d'ordonnancement GEDF. Pour ce faire, nous étudions l'influence du poids a sur les performances du système *SPriority*.

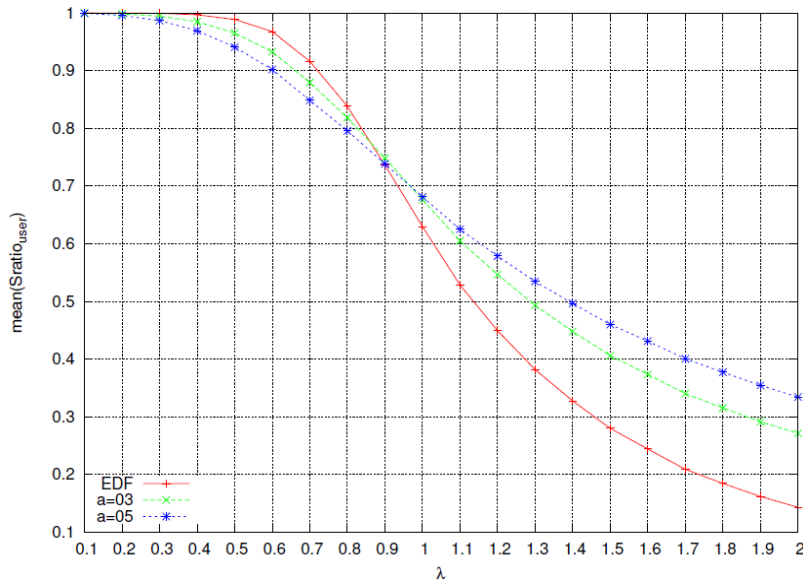
Nous étudions également l'influence de la composition de la base de données sur les performances du système. Et enfin, nous analysons l'influence de la capacité de calculs sur le taux de succès. Par ailleurs, nous donnons une approximation raisonnable de la transaction "Utilisateur" de taux de succès de la distribution de fréquence. A cet effet, nous effectuons des simulations de Monte Carlo qui nous permettent de comprendre le comportement du taux de succès des transactions. Et selon les valeurs des paramètres du système donnés dans les tableaux 5.1, 5.2 et 5.3, nous effectuons chaque expérience 1000 fois afin d'obtenir un échantillon de 1000 valeurs.

7.1. Influence du poids de *SPriority* sur les performances du système en fonction de la charge système sous des transactions plates et imbriquées.

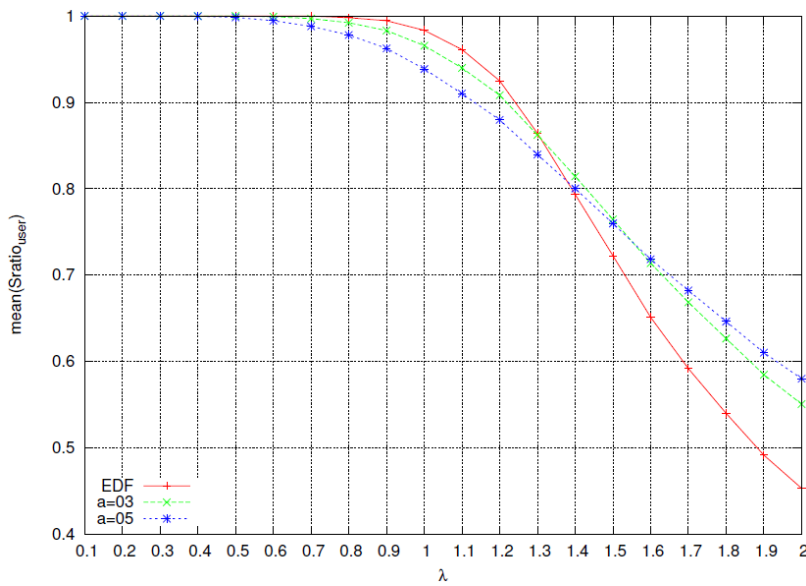
Pour montrer l'influence du poids *SPriority*, nous avons affecté différentes valeurs au paramètre a sous différentes charges de travail système. Nous notons que la valeur du paramètre γ est fixé à $4/5$ afin de minimiser l'influence du *DBA*, i.e, l'interaction de l'administrateur de la BDD, dans la formule (5.3) *SPriority*.

Nous avons utilisé des transactions plates dans le premier ensemble de simulations et des transactions imbriquées dans le deuxième. L'effectif d'opérations par transaction est inclus dans $[1,51]$ et suit une distribution de Poisson. Pour les transactions imbriquées, nous nous sommes limités à 4, pour le nombre de sous-transactions.

5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues



(a) Transactions plate



(b) Transactions imbriquée

Figure 5.1. Influence du poids de Spriority en fonction de la charge du système.

La figure 5.1 montre que le système donne de meilleures performances sous EDF au début, indépendamment de la structure de la transaction. Cette situation est inversée avec une charge croissante.

Nous notons que le point d'inflexion est atteint rapidement avec des transactions plates, quoi que nous avons utilisé le même nombre de transactions plates et imbriquées et que les transactions ont le même nombre d'opérations.

En effet, les figures 5.1(a) et 5.1(b) montrent que le point d'inflexion est atteint quand $\lambda \in]0.8, 1.0[$ (respectivement $\lambda \in [1.2, 1.4]$) avec des transactions plates (respectivement des transactions imbriquées). Cela peut s'expliquer par le fait que les transactions plates sont entièrement redémarrées lorsque des conflits de données surviennent. Toutefois, seules les sous-transactions sont redémarrées lors de l'utilisation de la transaction imbriquée.

En outre, les performances du système, c'est-à-dire $SRatio$ et $SRatio_{U_{ser}}$ obtenues avec des transactions imbriquées, sont supérieures à celles obtenues avec des transactions plates selon différentes valeurs de $Spriority$, comme le montrent les figures 5.2(a) et 5.2(b). Cependant, cette amélioration de la performance est plus importante avec EDF que $GEDF_{a=0,3}$ et $GEDF_{a=0,5}$.

En effet, l'amélioration de $SRatio$ en utilisant des transactions imbriquées atteint 17% sous EDF, 13% sous $GEDF_{a=0,3}$, et 11% sous $GEDF_{a=0,5}$. De la même manière, l'amélioration de $SRatio_{U_{ser}}$ est plus importante avec EDF.

Ce $SRatio$ est proche de zéro quand $\lambda \in [0.1, 0.5]$ alors qu'il augmente rapidement pour atteindre 48% quand $\lambda = 1.3$. Avec $GEDF_{a=0,5}$ (respectivement $GEDF_{a=0,3}$) cette amélioration de $SRatio_{U_{ser}}$ ne peut dépasser 30% (respectivement 36%).

La figure 5.3 montre que la distribution de fréquence du taux de succès pourrait raisonnablement être approchée par une probabilité Bêta¹. Selon le komolgorov-Smirnov (K.S.)². Le K.S. Résultats des tests liés à la figure 5.3 sont présentés dans le tableau 5.4 .

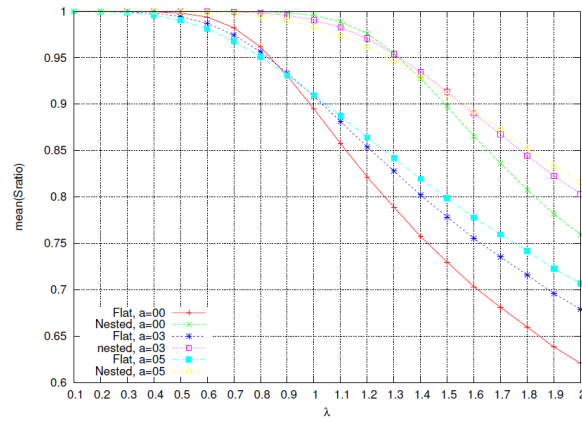
1. La loi Bêta est une loi de probabilité continue définie sur l'intervalle $[0, 1]$, elle dépend de deux paramètres p et q .

2. Le test de Kolmogorov-Smirnov est un test d'hypothèse utilisé pour déterminer si un échantillon suit bien une loi donnée connue par sa fonction de répartition continue, ou bien si deux échantillons suivent une même loi.

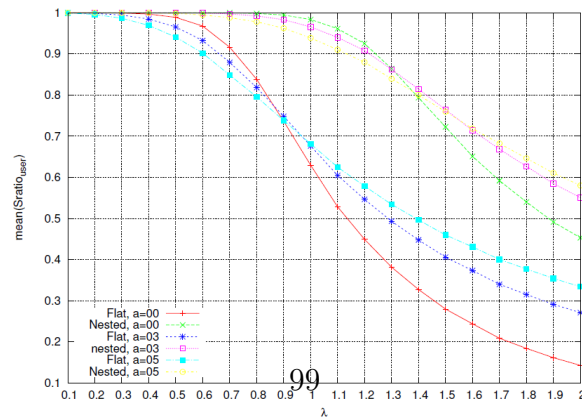
5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues

Table 5.1. *Caractéristiques du système et les paramètres de simulation utilisés.*

Caractéristiques du système		
Notation	Signification	Valeurs
λ	Taux d'arrivée des transactions "Utilisateur".	0.1 to 2.0.
Time	Durée d'une expérience.	1000 cycles CPU.
DBSize	Nombre de données dans la BD.	300, 400, 500, 600, 700.
TD-size	Nombre des données temps réel dans la BD.	$15\% \times \text{DBSize}$
Min_avi	avi minimal et maximal.	Min_avi = 5 cycles CPU,
Max_avi		Max_avi = 100 cycles CPU.



(a) $Sratio$



(b) $Sratio_{User}$

Figure 5.2. *Influence de la structure des transactions "Utilisateur" sur la performance du système.*

Table 5.2. *Caractéristiques du système et les paramètres de simulation utilisés.*

Caractéristiques du système		
Notation	Définition	Valeurs
User_ <i>SInerva</i>	Taille des transactions "Utilisateur".	[1, 51] des opérations combinées générées par la distribution de Poisson.
NStr	Nombre de sous-transaction.	NStr = 4.
Update_ <i>size</i>	Nombre d'opérations dans la transaction de mise à jour.	1 opération d'écriture.
D_ <i>UpT</i>	Échéance des transactions de mise à jour (approche more-less).	D_ <i>UpT</i> = 1/3 × avi.
P_ <i>UpT</i>	Période des transactions de mise à jour (approche more-less).	P_ <i>UpT</i> = 1/3 × avi.
<i>SPriority</i>	Intervalle de <i>SPriority</i> .	<i>SPriority</i> _{U_{pdate}} ∈ [0, 16], <i>SPriority</i> _{U_{ser}} ∈]16, 80].
γ	Initialisation de γ.	γ = 0.8.

Table 5.3. *Caractéristiques du système et les paramètres de simulation utilisés.*

Caractéristiques du système		
Notation	Définition	Valeurs
Quantum	Capacité d'exécution pendant un cycle CPU.	20, 30 Tâches/cycle.
Task	Action atomique.	Une opération de lecture ou d'écriture.
ReadTime	Consommation d'une opération de lecture.	1 quantum.
WriteTime	Consommation d'une opération d'écriture.	2 quantum.
a	Initialisation du paramètre a dans la formule (1) en utilisant GEDF.	$a = 0.3, 0.5$.
SP	Politique d'ordonnancement.	"EDF", "GEDF".
CC	Protocole de contrôle de concurrence.	"2PL-HP".

7.2. Influence du nombre de données temps réel sur la performance du système.

Dans cette section, on va analyser l'influence de la composition de la base de données en particulier, ainsi que l'influence du volume de données temps réel dans celle-ci, sur la performance du système. Dans ce sens, nous utilisons trois bases de données de même taille et nous faisons varier le taux de données temporelles,

Table 5.4. Estimation et résultats du test de K.S. sous EDF et GEDF_{a=0.5}

a	λ	Estimateurs			Résultats du test k.S	
Transactions plate						
		$\hat{p}_{user}$	$\hat{q}_{user}$		Distance	p -value
EDF(a=0)	0.8	33.62323	9.565422	0.0118	0.999	
	1.1	40.56774	46.63446	0.0175	0.9187	
	2.0	54.44303	385.4587	0.0156	0.9666	
Transactions imbriquée						
EDF(a=0.5)	0.8	144.2348	5.733316	0.0173	0.9244	
	1.2	84.90213	16.76632	0.022	0.7156	
	2.0	120.0873	104.0783	0.0249	0.5602	

Les données temporelles représentent respectivement 10%, 15% et 20% du volume total de données dans les deux bases de données. Le nombre de transactions de mise à jour est en moyenne de 1700 transactions (respectivement 2500, 3300) lorsque les données temporelles représentent 10% (respectivement 15%, 20%) de la taille de la base de données. Nous constatons que le nombre de transactions "Utilisateur" est lié à λ et est indépendant de la taille de la BDD et du pourcentage de données temporelles.

5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues

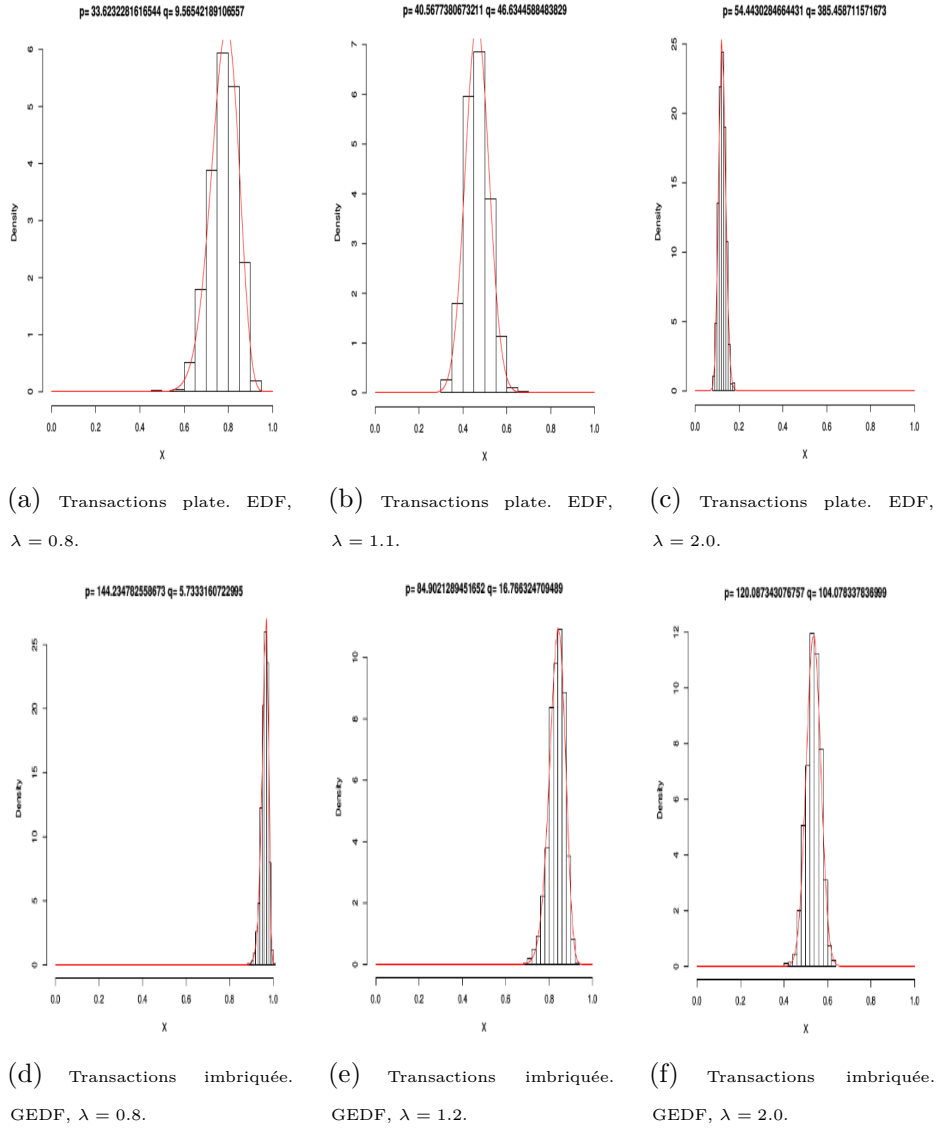
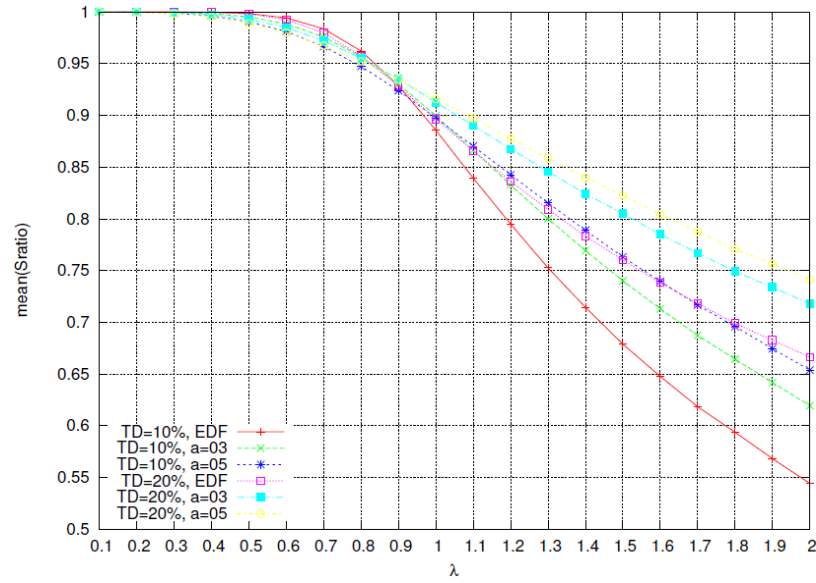
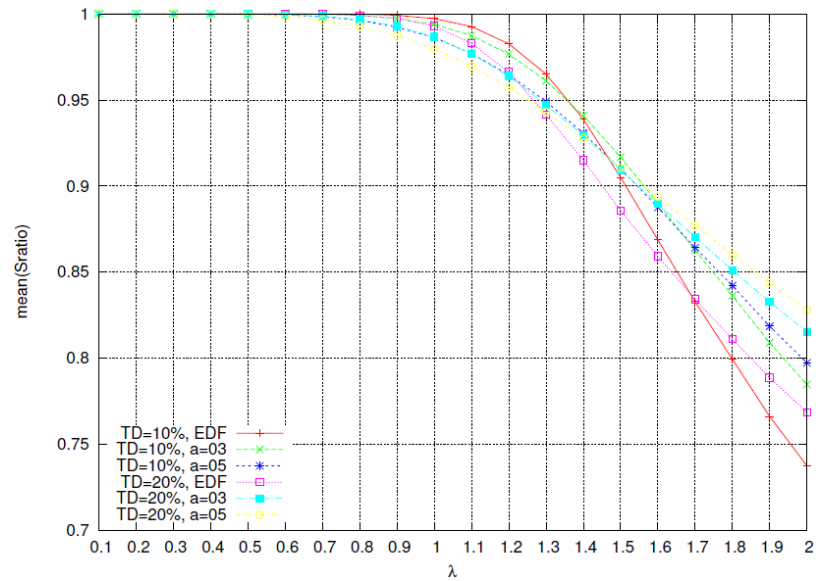


Figure 5.3. *SRatio et sa densité bêta équivalente sous $GEDF_{a=0.5}$, $Quantum = 30$, $TD = 20\%$.*



(a) Transactions plate



(b) Transactions imbriquée

Figure 5.4. Influence du nombre de données temporelles sur $SRatio$.

Comme le montrent les figures 5.4(a) et 5.4(b), quelle que soit la structure de transaction (plate ou imbriquée), le point d'inflexion est d'autant plus rapidement atteint que le taux de données temporelles augmente.

5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues

Cependant, le point d'inflexion est obtenu plus rapidement avec des transactions plates qu'avec des transactions imbriquées alors que l'impact des données temporelles est le plus important avec les transactions imbriquées. En effet, avec des transactions plates (respectivement avec Actions), le point d'inflexion est obtenu lorsque $\lambda \in]0.7, 0.9[$ (respectivement $\lambda \in]1.1, 1.3[$) lorsque le taux de données temporelles représente 20% de la taille de la base de données.

En revanche, lorsque le taux de données en temps réel représente 10% de la taille de la base de données, le point d'inflexion est obtenu lorsque $\lambda \in]0.8, 1.0[$ (respectivement $\lambda \in]1.3, 1.5[$). Cela s'explique par le fait que le nombre de transactions de mise à jour augmente avec le nombre de données en temps réel, afin de les maintenir à jour et donc le système devient surchargé plus tôt.

Les figures 5.5(a) et 5.5(b) montrent que l'utilisation de $Sratio$ diminue lorsque le pourcentage de données en temps réel croît indépendamment de la structure de transaction. Cela s'explique par le fait que le nombre de transactions temps réel affecte la capacité de calcul dédiée aux transactions "Utilisateur".

Table 5.5. Estimation et résultats du test de K.S. avec $GEDF_a = 0.5$ lorsque $RT = 10\%$

a	λ	Estimateurs		Results du test K.S	
		$\hat{p}_{user}$	$\hat{q}_{user}$	Distance	p -value
EDF($a=0.3$)	0.6	148.7153	6.942471	0.0123	0.998
	1.1	110.5015	56.21234	0.0211	0.7625
	1.6	165.6935	237.526	0.0181	0.897

Toutefois, la perte de performance, ie, la diminution de $SRatio_{User}$, est moins significative Avec $GEDF_{a=0.5}$. par exemple. Lorsque nous utilisons des transactions EDF $SRatio_{User}$ diminue de 6, 5% (respectivement 13%) lorsque le pourcentage de données en temps réel augmente 15% (respectivement 20%) à 10%. Cependant, sous $GEDF_{a=0.5}$, l'utilisateur $SRatio$ diminue de 4% (respectivement 9, 6%) lorsque le pourcentage des données en temps réel augmente de 15% (respectivement 20%) à 10%. De plus, l'influence du temps

réel les données sont plus importantes dans les transactions imbriquées.

En fait, lorsque le pourcentage de temps réel diminue de 20% à 10% $SRatio$ augmente de 16% dans le cadre d'EDF et de 10% sous $GEDF_{a=0.5}$. On peut aussi noter que les distributions de fréquence de $SRatio_{U_{ser}}$ peuvent être raisonnablement approximées par la densité de probabilité bêta, indépendamment des valeurs de RTW et de poids de $Spriority$ (Voir la figure 5.6). Le test Kolmogorov-Smirnov (K.S.) Les résultats sont présentés dans les tableaux 5.5 et 5.4.

7.3. Influence de la capacité informatique

La capacité de calcul est l'un des facteurs les plus influents sur le taux de réussite des transactions. Par conséquent, nous analysons dans cette section l'influence de la capacité de calcul sur la performance du système avec différentes structures de transactions et le pourcentage de données en temps réel sous EDF et GEDF.

À cette fin, nous avons répété des expériences antérieures avec diverses capacités de déplacement ($Quantum = 10, 20$ et 30). On peut noter qu'il est évident que lorsque l'exécution de la capacité augmente, les performances des systèmes augmentent à leur tour. Par conséquent, nous nous sommes intéressés à combiner l'influence de la capacité de calcul avec les paramètres précédents.

Dans les sections précédentes, nous avons montré que $SRatio_{U_{ser}}$ obtenu avec des transactions imbriquées surpasse celles obtenues avec des transactions plates indépendamment de $Spriority$. Dans ce chapitre, nous avons réalisé des expériences identiques avec des capacités d'exécution différentes.

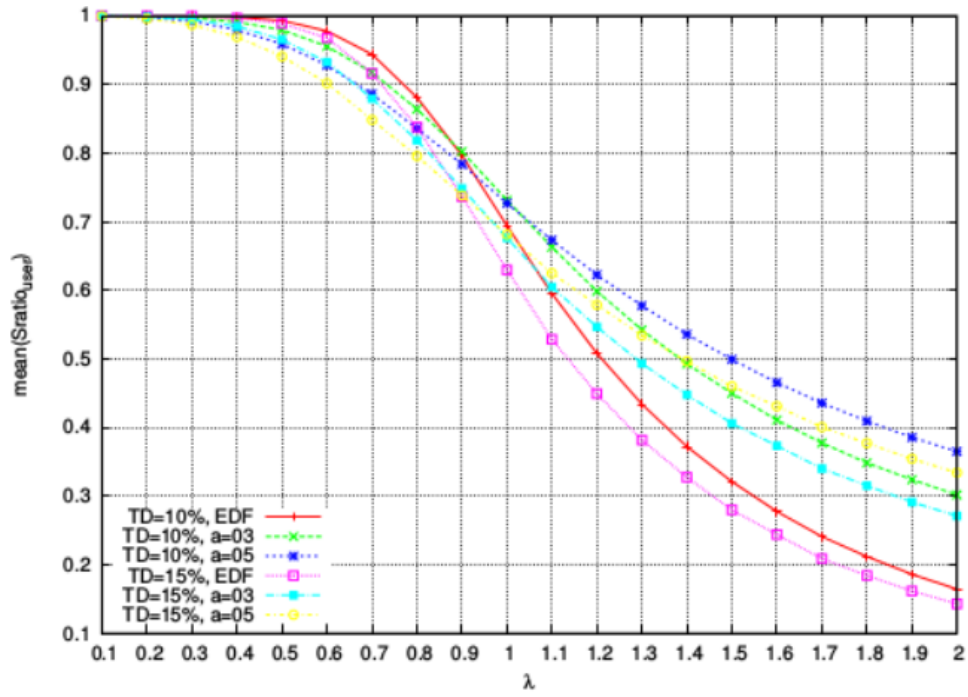
Les résultats obtenus, comme le montre la figure 5.7, confirment que le $SRatio_{U_{ser}}$ obtenu avec des transactions imbriquées surpasse celles obtenues avec des transactions plates sous des capacités de calcul différentes. De plus, on peut noter que l'influence de la structure est plus importante sous EDF que sous GEDF quelles que soient les performances des outils informatiques.

Cependant, cette influence diminue avec le pourcentage de données en

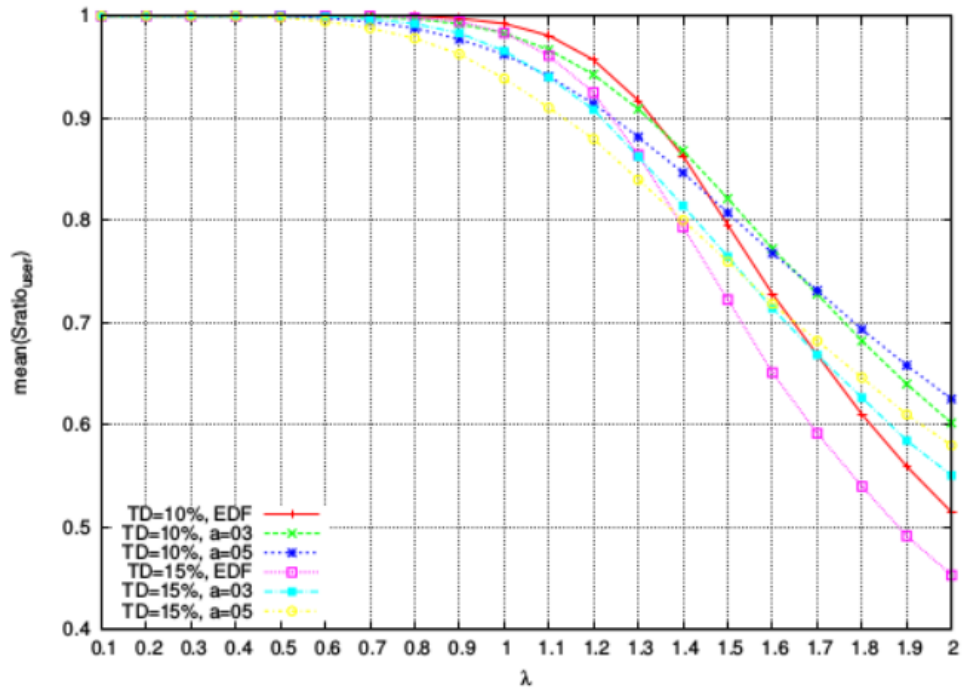
5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues

temps réel. Par exemple. Avec EDF l'amélioration atteint 17% (respectivement 11%) avec $\text{GEDF}_{a=0.5}$ lorsque le $Quantum = 30$ (respectivement $Quantum = 10$) l'amélioration ne dépasse pas 5% avec EDF et 3% avec $\text{GEDF}_{a=0.5}$.

Afin d'améliorer notre analyse, nous avons fait varier le pourcentage de données en temps réel avec différentes capacités de calcul.



(a) Transactions plate



(b) Transactions imbriquée

Figure 5.5. Influence du nombre de données en temps réel sur $SRatio_{user}$.

5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues

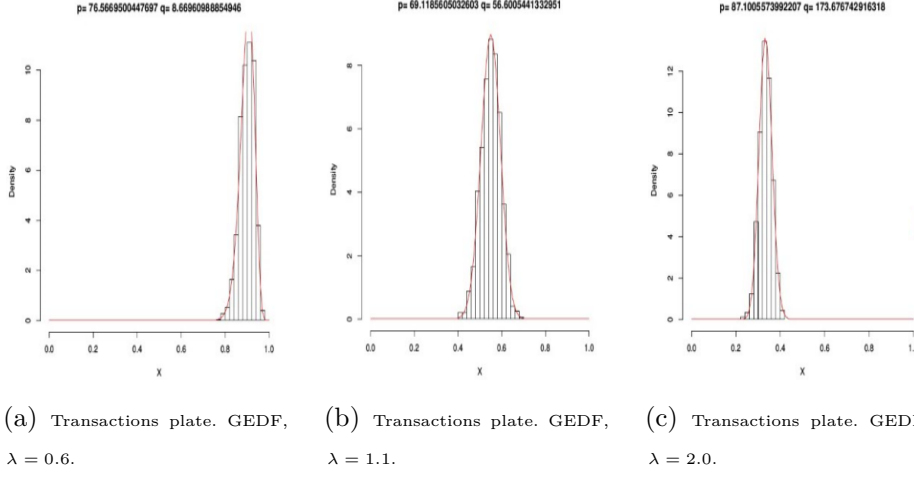
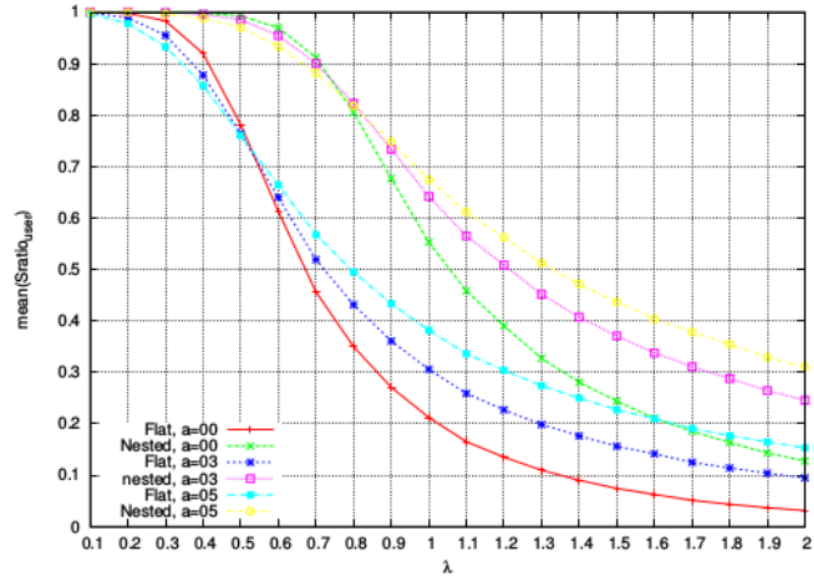
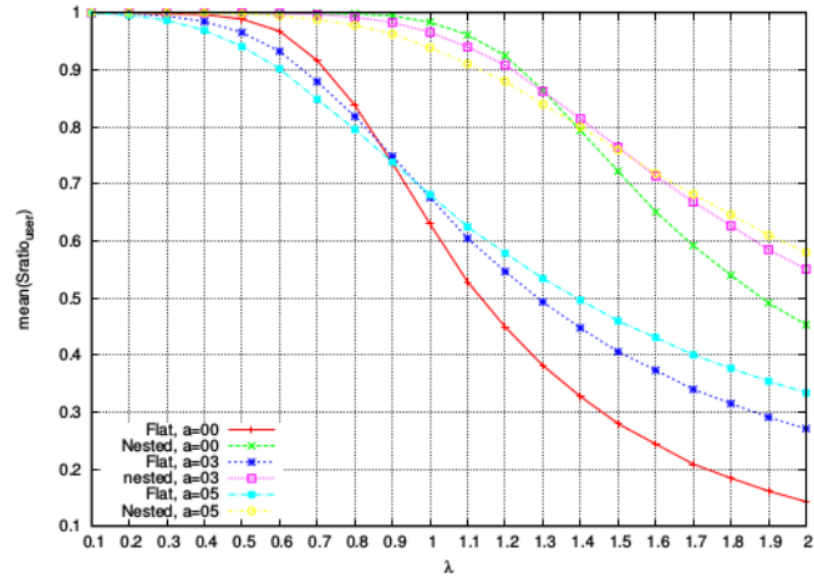


Figure 5.6. $SRatio$ et sa densité bêta équivalente sous $GEDF_{a=0.3}$, $Quantum = 30$, $TD = 10\%$.

Les résultats obtenus confirment que lorsque le pourcentage de données en temps réel diminue, le $SRatio_{user}$ augmente avec différentes capacités de calcul. De plus, l'influence de la composition de la base de données est moins importante lorsque la capacité de calcul diminue. Par exemple. Lorsque le pourcentage de données en temps réel diminue de 20% à 15% l'amélioration de $SRatio_{user}$ atteint 12% lorsque le $Quantum = 20$ mais il ne dépasse pas 7% lorsque le $Quantum = 30$ sous EDF. Cette amélioration de $SRatio_{user}$ avec $GEDF_{a=0.5}$: atteint 9% lorsque le $Quantum = 20$ mais ne dépasse pas 5% lorsque le $Quantum = 30$.

(a) $Quantum = 20$ (b) $Quantum = 30$ **Figure 5.7.** Influence de la capacité informatique sur $SRatio_{user}$.

La figure 5.8 montre que la distribution de fréquence du taux de succès de transaction "Utilisateur" peut être raisonnablement approchée par une fonction de densité de probabilité bêta, selon le test statistique Kolmogorov-

5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues

Smirnov (k.S.), indépendamment des capacités d'exécution. Les résultats sont donnés dans le tableau [5.6](#).

8. Conclusion

Le protocole GEDF est une politique d'ordonnancement multi-paramètres basée sur la politique EDF bien connue. Ses paramètres doivent être réglés en tenant compte des caractéristiques des systèmes.

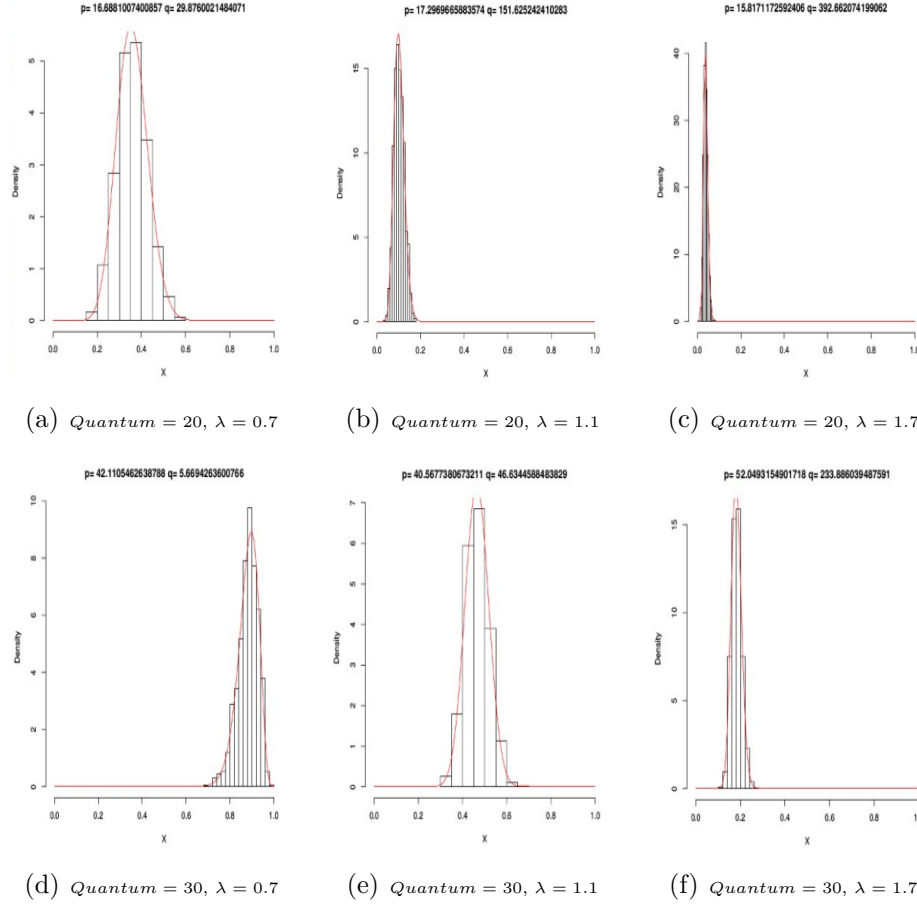


Figure 5.8. *SRatio et sa densité bêta équivalente avec différentes capacités d'exécution sous EDF, TD = 20%.*

Les simulations effectuées ont montré que l'utilisation de transactions imbriquées améliore les performances du système. Les performances des transactions imbriquées sont moins sensibles que pour les transactions à plat. De plus, GEDF est également moins sensible que EDF quelle que soit la

5. Étude stochastique dans le SGBD Temps Réel pour comparer des transactions plates et étendues

Table 5.6. *Estimation et résultats du test K.S sous $GEDF_{a=0.5}$, lorsque $RT = 10\%$.*

$quantum$	λ	Estimateurs		résultats du test K.S	
		$\hat{p}_{user}$	$\hat{q}_{user}$	Distance	p -value
20	0.7	16.6881	29.876	0.0155	0.97
	1.2	17.29697	151.6252	0.018	0.9002
	1.7	15.81712	392.6621	0.0131	0.995
30	0.7	42.11055	5.669426	0.0205	0.7919
	1.2	40.56774	46.63446	0.0175	0.9187
	1.7	52.04932	233.886	0.0176	0.9156

structure des transactions, la composition de la base de données et la capacité d'exécution. S'y ajoute le fait que nous avons montré que la répartition en fréquence du taux de succès pour les transactions "Utilisateur" peut être approchée par une distribution bêta indépendamment du protocole de programmation utilisé, de la structure des transactions, de la composition de la base de données et de la capacité d'exécution. Ainsi, nous pouvons prédire le taux de réussite du système sans effectuer de simulations qui nécessitent souvent un temps de calcul important.



1. La dérivé partielle d'une copule

$$\mathbb{P}(V \leq v | U = u) = \frac{\partial}{\partial u} C(u, v)$$

$$\begin{aligned} \mathbb{P}(V \leq v | U = u) &= \mathbb{P}(V \leq v | u \leq U \leq u + \Delta u) \\ &= \lim_{\Delta u \rightarrow 0} \frac{\mathbb{P}(V \leq v, u \leq U \leq u + \Delta u)}{\mathbb{P}(u \leq U \leq u + \Delta u)} \end{aligned}$$

Comme U une variable aléatoire appartient de $[0,1]$ on a $\mathbb{P}(u \leq U \leq u + \Delta u) = \Delta u$ et on déduire

$$\begin{aligned} \mathbb{P}(V \leq v | U = u) &= \lim_{\Delta u \rightarrow 0} \frac{\mathbb{P}(V \leq v, u \leq U \leq u + \Delta u)}{\Delta u} \\ &= \lim_{\Delta u \rightarrow 0} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u} \\ &= \frac{\partial}{\partial u} C(u, v). \end{aligned}$$

2. La fonction de densité conditionnelle d'une loi de Student

Soit X et Y deux variables aléatoires de fonctions de répartitions respectives F_X et F_Y

$$\begin{aligned}
 f(Y|X=x) &= \frac{f(Y=y, X=x)}{f(X=x)} \\
 &= \frac{\frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)(1-r^2)^{1/2}} \left(\frac{\nu(1-r^2)+x^2-2rxy+y^2}{\nu(1-r^2)} \right)^{-\frac{\nu+2}{2}}}{\frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{1/2}} \frac{1}{(1+x^2/\nu)^{\frac{\nu+1}{2}}}} \\
 &= \frac{\frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})(\nu\pi)(1-r^2)^{1/2}} \left(\frac{\nu}{\nu+x^2} \right)^{1/2}}{\left(\frac{\nu(1-r^2)+x^2+(y-rx)^2-(rx)^2}{(\nu+x^2)(1-r^2)} \right)^{\frac{(\nu+2)}{2}}} \\
 &= \frac{\frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})(\nu+1)\pi(1-r^2)^{1/2}}}{\left(1 + \frac{(y-rx)^2}{(\nu+x^2)(1-r^2)} \right)^{\frac{(\nu+2)}{2}} \left(\frac{\nu+1}{\nu+x^2} \right)^{-1/2}} \\
 &= \frac{\Gamma(\frac{(\nu+1)+1}{2})}{\Gamma(\frac{\nu+1}{2})((\nu+1)\pi)^{1/2}} \frac{\left(\frac{\nu+1}{(\nu+x^2)(1-r^2)} \right)^{1/2}}{\left(1 + \frac{(y-rx)^2}{(\nu+x^2)(1-r^2)} \frac{\nu+1}{\nu+1} \right)^{\frac{(\nu+1)+1}{2}}} \\
 &= \frac{\Gamma(\frac{(\nu+1)+1}{2})}{\Gamma(\frac{\nu+1}{2})((\nu+1)\pi)^{1/2}} \left(1 + \frac{\left(\frac{y-\mu}{\sigma} \right)^2}{\nu+1} \right)^{-\frac{(\nu+1)+1}{2}} \cdot \frac{1}{\sigma}
 \end{aligned}$$

avec $\mu = rx$ et $\sigma = \frac{(\nu+x^2)(1-r^2)}{1+\nu}$.

Donc $Y|X=x$ suit une loi de student de paramètres μ , σ et à $\nu+1$ degrés de liberté.



ANNEXE B

1. Fonction de Bessel modifiée du troisième type

Soit $K_\lambda(x)$ la fonction de Bessel modifiée du troisième type d'ordre λ . celle-ci est la solution à l'équation différentielle suivante :

$$x^2 \frac{d^2 w}{dx^2} + x \frac{dw}{dx} - (x^2 + y^2)w = 0$$

où x est un nombre complexe.

La fonction de Bessel modifiée du troisième type avec l'indice $\lambda \in \mathbb{R}$ peut être définie par la formule suivante :

$$K_\lambda(x) = \frac{1}{2} \int_0^\infty \exp\left(-\frac{x}{2}\left(y + \frac{1}{y}\right)\right) y^{\lambda-1} du, \quad \text{pour } x > 0, \lambda \in \mathbb{R}.$$

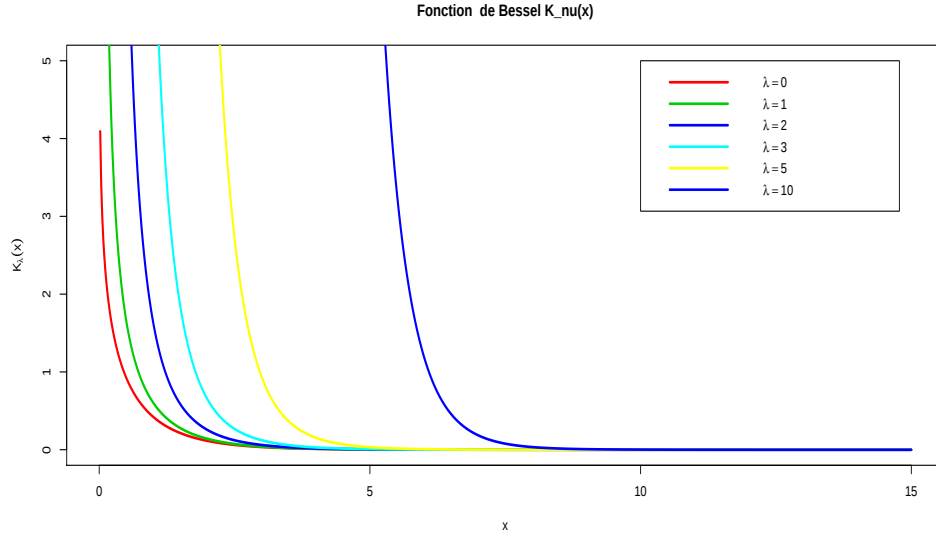


Figure B.1. La fonction de Bessel du troisième type.

La fonction de Bessel modifiée a les propriétés suivantes :

1.0.0.1. Théorème

$$K_\lambda(x) = \sqrt{\frac{\pi}{2x}} \exp(-x) \left(1 + o\left(\frac{1}{x}\right)\right), \quad x > 0$$

$$K_\lambda(x) = K_{-\lambda}(x)$$

$$K_{\lambda+1}(x) = \frac{2\lambda}{x} K_\lambda(x) + K_{\lambda-1}(x)$$

$$K_\lambda(x) = -\frac{\lambda}{x} K_\lambda(x) - K_{\lambda-1}(x).$$

1.0.0.2. Théorème Pour tout $\lambda = n + \frac{1}{2}, n = 0, 1, 2, \dots$, on a :

$$K_{n+\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}} e^{-x} \left\{ 1 + \sum_{i=1}^n \frac{(n+i)!}{(n-i)!i!} (2x)^{-i} \right\}$$

$$K_{-\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}} e^{-x}$$

$$K_{\frac{3}{2}}(x) = \sqrt{\frac{\pi}{2x}} e^{-x} \left(1 + \frac{1}{x}\right).$$

1.0.0.3. Théorème Pour tout $t > 0$ et $\lambda \geq 0$ on a :

$$\frac{2\lambda}{x} K_\lambda(x) + K_{\lambda-1}(x) = \int_0^\infty \frac{g_\lambda(t)}{x^2 + t} dt$$

où

$$g_\lambda(t) = \frac{2}{\pi^2 t J_\lambda^2(\sqrt{t}) + Y_\lambda^2(\sqrt{t})} > 0$$

tel que J_λ est la fonction de Bessel de première d'ordre λ .

$$J_\lambda(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin(u) - \lambda u) du - \frac{\sin(\lambda\pi)}{\pi} \int_0^\infty e^{x \sinh(u) - \lambda u} du$$

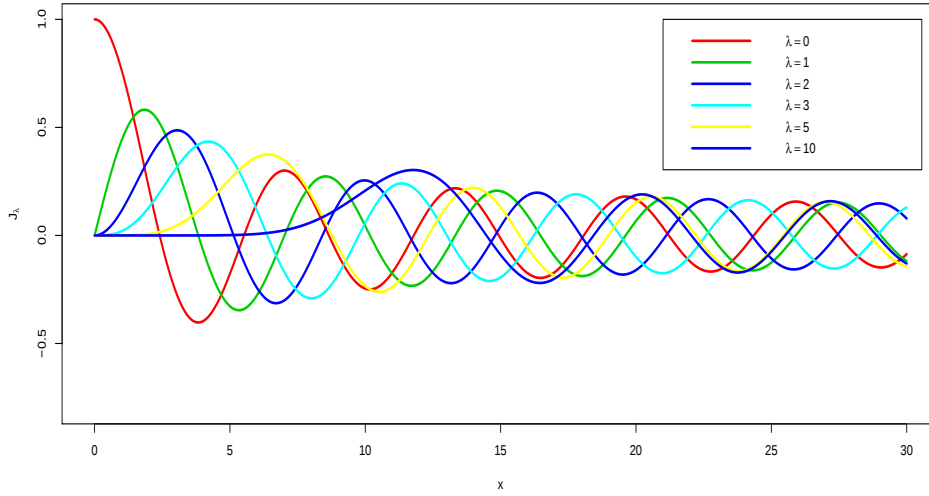


Figure B.2. La fonction de Bessel de première espèce.

et Y_λ est la fonction de Bessel de deuxième espèce d'ordre λ , voir [Grosswald \(1976\)](#).

$$Y_\lambda(x) = \frac{1}{\pi} \int_0^\pi \sin(x \sin(u) - \lambda u) du - \frac{1}{\pi} \int_0^\infty \left[e^{\lambda u} + e^{-\lambda u} \cos(\lambda\pi) \right] e^{x \sinh(u)} du.$$

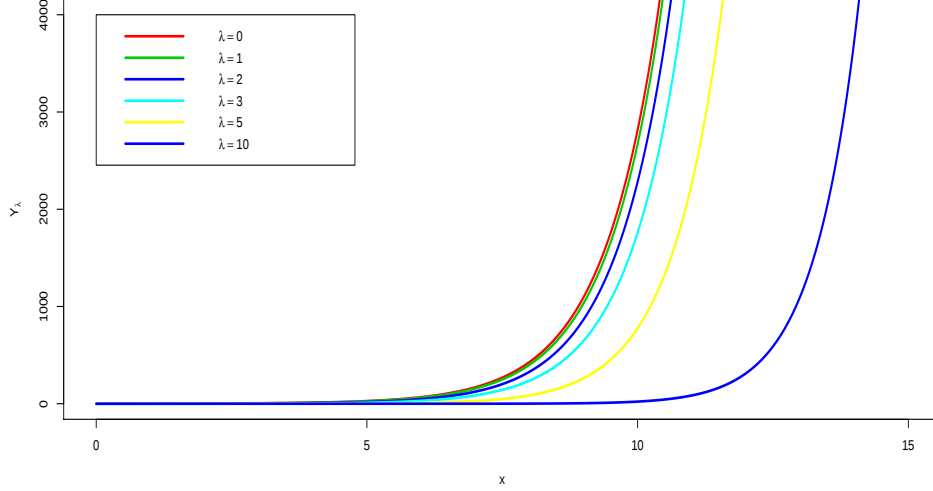


Figure B.3. La fonction de Bessel de deuxième espèce.

1.0.0.4. Théorème

$K_\lambda(x) \sim \Gamma(\lambda)2^{\lambda-1}x^{-\lambda}$, lorsque $x \rightarrow 0$ si $\lambda > 0$

où $\Gamma(\lambda)$ est la fonction gamma standard définit comme suit :

$$\Gamma(\lambda) = \int_0^\infty e^{-y} y^{\lambda-1} dy$$

$K_0(x) \sim -\log(x)$, quand $x \rightarrow 0$

$$\frac{K_\lambda(x)}{K_{\lambda+\epsilon}(x)} \rightarrow 1, \text{ lorsque } x \rightarrow \infty, \quad \epsilon > 0 \quad \text{et} \quad \lambda > 0.$$

1.0.0.5. Théorème

$$K'_0(x) = -K_1(x)$$

$$K'_\lambda(x) = -\frac{1}{2} (K_{\lambda+1}(x) + K_{\lambda-1}(x)) = -\frac{\lambda}{x} K_\lambda(x) + K_{\lambda-1}(x)$$

$$K'_\lambda(x) = \frac{\lambda}{x} K_\lambda(x) + K_{\lambda-1}(x) K_{\lambda+1}(x).$$

Plus avoir plus d'informations voir [Abramowitz et Stegun \(1968\)](#) et [Barndor-Nielsen and Bloesild \(1981\)](#).

BIBLIOGRAPHIE

- Abramowitz, M. and Stegun, I. A. (1968). Handbook of Mathematical Functions. NewYork : Dover Publ.
- Aggarwal, C. (2009). An introduction to uncertain data algorithms and applications. In Aggarwal, C. C., editor, Managing and Mining Uncertain Data, volume 35 of Advances in Database Systems, pages 1-8. Springer US.
- Aggarwal, C. C. and Yu, P. S. (2009). A survey of uncertain data algorithms and applications. IEEE Trans. Knowl. Data Eng., 21(5), 609-623.
- Al-Abbasi, J. N., Fahmi and K. J. (1985). Estimating maximum magnitude earthquakes in iraq using extreme value statistics. Geophysical journal of the Royal Astronomi- cal Society, 82, 535-548.
- Aldama, A. A. and Ramirez A. I. (2002). Dam design flood estimation based on bivariate extreme value distributions. in the extremes of the extremes : extraordinary floods. Wallingford, IAHS Press. 1, 257-262.
- Aleiyouka, M., Berred, A. and Ahsanullah, M. (2017). Tail dependence coefficient of generalized hyperbolic distribution. Journal of Statistical Theory and Applications, 16, 375-381.
- Alvarado, E., Sandberg, D. V. and Pickford, S. G. (1998). Modeling large forest fires as extreme events. Northwest Science, 1, 66-75.
- Banachewicz, K. and van der Vaart, A. (2008). Tail dependence of skewed grouped t -distributions. *Statistics and Probability Letters*. 78, 2388-2399.
- Barbara, D., Garcia-Molina, H., and Porter, D. (1990). A probabilistic relational data model. In Bancilhon, F., Thanos, C., and Tsichritzis, D., editors,

- Advances in Database Technology, EDBT '90, volume 416 of Lecture Notes in Computer Science, pages 60-74. Springer Berlin Heidelberg.
- Barndorff-Nielsen, O. E. and Christian, H. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Probability Theory and Related Fields, Springer*. 38(4), 309-311.
- Barndorff-Nielsen, O. E. (1997). Exponentially decreasing distributions for the logarithm of particle size, *Royal Society of London. Proceedings A. Mathematical, Physical and Engineering Sciences*. 353, 401-419.
- Barndor-Nielsen, O. E. and Bløesild, P. (1981). Hyperbolic distributions and ramifications : Contributions to theory and application. In C. Taillie, G. Patil, and B. Baldessari (Eds.), *Statistical Distributions in Scientific Work*, Volume 4, pp. 19-44. Dordrecht : Reidel.
- Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes Theory and Applications*, Wiley, Chichester.
- Bengtsson, A. and Nilsson, C. (2007). Extreme value modelling of storm damage in swedish forests. *Nat. Hazards Earth Syst. Sci*, 7, 515-521.
- Billingsley, P. (1986). *Probability and measure*, Wiley New York.
- Bortot, P. and Tawn, J. A. (1998). Models for the extremes of Markov chains, *Biometrika* 85, 851-867.
- Bouchaud, J. -P. and Potters, M. (2003). *Theory of financial risk and derivative pricing : from statistical physics to risk management*. Cambridge University Press, Cambridge, 2nd edition.
- Cavallo, R. and Pittarelli, M. (1987). The theory of probabilistic databases. In *Proceedings of the 13th International Conference on Very Large Data Bases, VLDB '87*, pages 71-81, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Caperra, P., Fougères, A. L. and Genest C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*. 84, 567-577.

BIBLIOGRAPHIE

- Chebana, F., El Adlouni, S. and Bobée, B. (2010). Mixed estimation methods for Halphen distributions with applications in extreme hydrologic events. *Stoch. Environ. Res. Risk Assess.* 24, 359-376.
- Coles, S. G., Heffernan, J. E. and Tawn, J. A. (1999). *Extremes*, 2(4), 339-365.
- Daniels, H.E. (1945). The statistical theory of the strength of bundles and threads. *Proc. Royal Soc.* 405-435.
- Datta A., Mukherjee S., Konana P., Viguier I.R. and Bajaj A. (1996). Multiclass transaction scheduling and overload management in firm real-time database systems. *Information Systems*, 21, no. 1, 29-54.
- Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas, *International Statistical Review.* 73, 111-129.
- De Haan, L. and Ferreira, A. (2006). *Extreme Value Theory : An Introduction.* Berlin : Springer.
- Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non-paramétrique d'indépendance. Académie royale de Belgique Bulletin de la classe des sciences. 5e séries 65, 274-292.
- De Michele, C. and Salvadori, G. (2002). A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research*, 108, 1-11.
- Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989). A moment estimator for the index of an extreme value distribution, *Ann. Statist.* 17, 1833-1855.
- Demarta, S., McNeil, A. (2005). The t copula and related copulas. *Int. Stat. Rev.*, 73 (1), 111-129.
- Dobrić, J. and Schmid, F. (2005). Nonparametric estimation of the lower tail dependence λ_L in bivariate copulas, *Journal of Applied Statistics*, 32 :4, 387-407.

- Eberlein, E. and Prause, K. (2002). *The generalized hyperbolic model : Financial derivatives and risk measures*, In *Mathematical Finance-Bachelier Congress 2000*, H. German, D. Madan, S. Pliska and T. Vorst (Eds). 245-267.
- Eichner, J. F., Kantelhardt, J. W., Bunde, A. and Havlin, S. (2006). Extreme value statistics in records with long-term persistence. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 73, 016130.
- Einmahl, J. H. J. and Magnus, J. R. (2008). Records in athletics through extreme value theory. *Journal of the American Statistical Association*, 103, 1382-1391.
- Embrechts, P., McNeil, A. J., Straumann, D. (1999). Correlation : Pitfalls and Alternatives. *Risk Magazine*, 12, 69-71.
- Embrechts, P., Lindskog F. and McNeil, A. (2001). Modelling Dependence with Copulas and Applications to Risk Management. ETH Zurich Departement of mathematics working paper.
- Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management : properties and pitfalls. In "Risk Management : Value at Risk and Beyond" (M.A.H. Dempster, Ed.), Cambridge University Press, Cambridge. 176-223.
- Erdelyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F. G. (1954). *Bateman Manuscript Project : Tables of Integral Transform*, McGraw-Hill, New York.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N. and Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40, 1-12.
- Fisher, R. A., Tippett and L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc Camb Philos Soc* 24 :180-190.

- Frahm, G., Junker, M. and Schmidt. (2005). Estimating the tail dependence coefficient : properties and pitfalls. *Insurance : Mathematics & Economics*, 37(1), 80–100.
- Fung, T. and Seneta, E. (2010). Tail dependence for two skew t distributions, *Statistics and Probability Letters*. **80** 784-791.
- Geffroy, J. (1958/59). Contributions à la theorie des valeurs extrême, Publ. Inst. Statist. Univ. Paris 7/8, 37-185.
- Genest, C. and Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4), 347-368.
- Genest, C. and MacKay, R. J. (1986). The joy of copulas : Bivariate distributions with uniform marginals. *The American Statistician*, 40, 280-283.
- Green, T. J. and Tannen, V. (2006). Models for incomplete and probabilistic information. In *Proceedings of the 2006 international conference on Current Trends in Database Technology, EDBT'06*, pages 278-296, Berlin, Heidelberg. Springer-Verlag.
- Grosswald, E. (1976). The Student t -distribution of any degree of freedom is infinitely divisible. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete* 36, 103-109.
- Guegan, D. and Zhang, J. (2009). Pricing Bivariate Option under GARCH-GH Model with Dynamic Copula : Application for Chinese Market. *The European Journal of finance*, 15(7-8), 777-795.
- Han, S., Lam, K. y., Wang, J., Son S. H. and Mok, A. K. (2012). Adaptive co-scheduling for periodic application and update transactions in real-time database systems. *J Syst Softw*, 85, no. 8, 1729-1743.
- Haritsa J. R., Livny M., and Carey M. J. (1991). Earliest Dead-line Scheduling for Real-Time Database Systems. In *12th IEEE Real-Time Systems Symposium*. 232-243.

- Hollander, M. and Wolfe, D. A. (1973). Nonparametric Statistical Methods. John Wiley & Sons, New York.
- Hsu, W. W., Smith, A. J., and Young, H. C. (2001). Characteristics of production database workloads and the tpc benchmarks. IBM Systems Journal, 40(3) :781-802.
- Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. Adv. Appl. Probab., 34, 587-608.
- Iyengar, S. and Liao, Q. (1997). Modeling neural activity using the generalized inverse Gaussian distribution. Biol. Cybern. 77, 289-295.
- Joe, H., Smith, R. L., and Weissman, I. (1992). Bivariate threshold models for extremes, J. R. Statist. Soc. B54, 171-183.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.
- Jondeau, E. (2008). Asymmetry in tail dependence in equity portfolios. *Computational Statistics & Data Analysis*.
- Jorgensen, B. (1982). Statistical properties of the generalized inverse Gaussian distribution for log-linear and stochastic volatility models, *Lecture Notes in Statistics*.
- Juri, A., Wüthrich, M. V. (2002). Copula convergence theorems for tail events. Insurance : Mathematics and Economics, 24, 139-148.
- Kaddes, M., Amanton, L., Berred, A., Sadeg, B. and Abdouli, M. (2013). Enhancement of Generalized Earliest Deadline First Policy. In Proceedings of the 15th ICEIS. vol. 1, 183-190.
- Kaddes, M., Amanton, L., Berred, A., Sadeg, B., Abdouli, M. and Bouaziz, R. (2014). A Stochastic Analysis of Nested Transactions Success Ratio in RTDBS. In ESM'2014, October 22-24, FEUP - University of Porto, Porto, Portugal. 108-115.

BIBLIOGRAPHIE

- Kaddes, M., Amanton, L., Abdouli, M., Aleiyouka, M. and Berred, M. (2015). Flat vs Nested Firm Real-time Transactions : an Analysis of their Success Ratio, 29th European Simulation and Modelling Conference-ESM'2015, October 2015, Leicester, United Kingdom.
- Kendall, M. and Stuart, A. (1963). The Advanced Theory of Statistics, Volume 1, Hodder Education.
- Kluppelberg, C. and Rasmussen, M. G. (2013). Outcrossings of safe regions by generalized hyperbolic processes. *Statistics and Probability Letters*. 83, 2197-2204.
- Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53, 814-861.
- Lakhal-Chaieb, M. L., Rivest, L. P. and Abdous, B. (2008). Estimating survival and association in a semicompeting risks model. *Biometrics*, 64, 180-188.
- Lavenda, B. H. and Cipollone, E. (2001). Extreme value statistics and thermodynamics of earthquakes : large earthquakes. *Annali di Geofisica*, 43, 469-484.
- Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*. 83, 169-187.
- Ledford, A. W. & Tawn, J. A. (1997). Modelling dependence within joint tail regions, *J. R. Statist. Soc. B*, **59**, 475-499.
- Lehmann, E. L. (1975). *Nonparametrics : Statistical Methods Based on Ranks*. San Francisco : Holden-Day.
- Leung, J. Y. T. and Whitehead, J. (1982). On the complexity of fixed-priority scheduling of periodic, real-time tasks. *Performance Evaluation*, 2(4) : 237-250.
- Liu, C. L. and Layland, J. W. (1973). Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment. *J. ACM*, 20(1), 46-61.

- Longin, F. M. and Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance* 56, 649-676.
- Malevergne, Y. and Sornette, D. (2002). Minimising extremes. *Risk*, 15(11), 129-34.
- Malevergne, Y. and Sornette, D. (2004). How to account for extreme co-movements between individual stock and the market. *Journal of Risk*, 6(30), 71-116.
- Malevergne, Y. and Sornette, D. (2005). *Extreme Financial Risks*. Springer-Verlag, London.
- Mardia, K. V. (1964). Asymptotic independence of bivariate extremes, *Calcutta Statist. Assoc. Bull.* 13, 172-178.
- McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management : Concepts, Techniques and Tools*, Princeton University Press.
- Moss, J. E. (1985). *Nested Transactions : An Approach to Reliable Distributed Computing*. Massachusetts Institute of Technology, Cambridge, MA, USA.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Verlag, New York.
- Nelsen, R. B. (1999). *An Introduction to Copulas, Lecture Notes in Statistics, vol. 139*.
- Pang, H., Livny M., and Carey, M. J. (1992). Transaction Scheduling in Multiclass Real-Time Database Systems. In *IEEE Real-Time Systems Symposium*. 23-34.
- Paolella, M. (2007). *Intermediate Probability : A Computational Approach*, Wiley, Chichester.
- Patton, J. A. (2006). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics*, 21(2), 147-173.

- Perreault, L., Bobée, B. and Rasmussen, P. F. (1999a). Halphen distribution system. I : Mathematical and statistical properties. J. Hydrol. Eng. 4, 189-199.
- Perreault, L., Bobée, B. and Rasmussen, P. F. (1999b). Halphen distribution system. II : Parameter and quantile estimation. J. Hydrol. Eng. 4, 200-208.
- Poon, S. H., Rockinger, M. and Tawn, J. (2004). Extreme-value dependence in financial markets : Diagnostics, models and financial implications. Review of Financial Studies 17, 581-610.
- Prause, K. (1999). *The Generalized Hyperbolic Model : Estimation Financial Derivatives and Risk Measures, PhD diss., University of Freiburg.*
- Ramamritham, K., Han, S., Chen, D., Xiong, M., Lam, K. Y., and Mok, A. K. (2014). Schedulability Analysis of Deferrable Scheduling Algorithms for Maintaining Real-Time Data Freshness. IEEE Transactions on Computers, 63, no. 4, 979-994.
- Resnick, S. I. (1987). Extreme Values, Point Processes and Regular Variation, Springer-Verlag, New York.
- Salvadori, De Michele, C. G., Kottegoda, N. T. and Rosso, R. (2007). Extremes in Nature : An Approach Using Copulas, Water Science and technology Library, vol. 56 Spring, Dordrecht(NT).
- Schlueter, S. and Fischer, M. (2012). The weak tail dependence coefficient of the elliptical generalized hyperbolic distribution, *Extremes*. **15(2)** 159-174.
- Schluter, S. and Fischer, M. (2012). A tail quantile approximation formula for the student t and the symmetric generalized hyperbolic distribution. FAU Discussion Papers in Economics.
- Schmidt, R. (2002). Tail dependence for elliptically contoured distributions, *Mathematical Methods of Operations Research*. **55** 301-327.
- Schmidt, R., Stadtmüller, U. (2006). Non parametric estimation of tail dependence. The Scandinavian Journal of Statistics. 33, 307-335.

- Semghouni, S., Amanton, L., Sadeg B., and Berred, A. (2007). On new scheduling policy for the improvement of firm RT-DBSs performances. *Data Knowl Eng*, 63, no. 2, 414-432.
- Semghouni, S., Sadeg, B., Berred, A., and Amanton, L. (2008). The Behavior of Transactions Success Ratio in Firm Real-Time Database Systems. *Journal of Advances in Computer Science and Engineering*, 2(2), 133-163. 30 pages.
- Sibuya, M. (1960). Bivariate extreme statistics, *Ann. Inst. Statist. Math.* 11, 195-210.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*. 15 : 72-101.
- Tawn, J. A. (1988). Bivariate extreme value theory : models and estimation, *Biometrika* 75, 397-415.
- Thomasian, A. and Ryu, I. K. (1991). Performance analysis of two-phase locking. *IEEE Trans. Softw. Eng.*, 17(5) : 386-402.
- Tiago de Oliveira, J. (1962/63). Structure theory of bivariate extremes, extensions, *Est. Mat., Estat. e. Econ.* 7, 165-195.
- Wald, A. (1947). *Sequential 1947 Analysis*. Wiley, New York.
- Xiong, M. and Ramamritham, K. (2004). Deriving Deadlines and Periods for Real-Time Update Transactions. *IEEE Trans Computers*, 53, no. 5, 567-583.