

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

الجمهورية الجزائرية الديمقراطية الشعبية

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

وزارة التعليم العالي و البحث العلمي

UNIVERSITE -SAAD DAHLAB- BLIDA 1

جامعة سعد دحلب - البليدة -



Faculté des Sciences

كلية العلوم

Département de Mathématiques

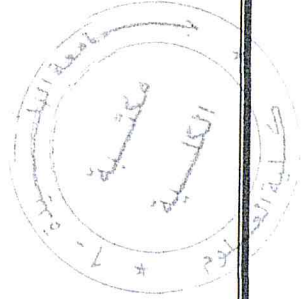
قسم الرياضيات

MEMOIRE DE PROJET DE FIN D'ETUDES

مذكرة نهاية التدرج

Pour l'Obtention du diplôme
DE MASTER

Option : *Modélisation Stochastique et Statistique*



Thème : Copules Bivariées extrêmes-Application aux données
météorologiques

Présenté par :

SAADI SABRINA

devant le jury composé

Mr. O.TAMI

Mr. Rassoul Abdelaziz

Mr. R.FRIHI

President

Examineur

Encadreur

Promotion -2016/ 2017-

MA-510-46-2

REMERCIEMENT

En préambule à ce mémoire nous remercions **ALLAH** qui nous a aide et donné la patience et le courage durant ces longues années d'étude.

Nous souhaitons adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.

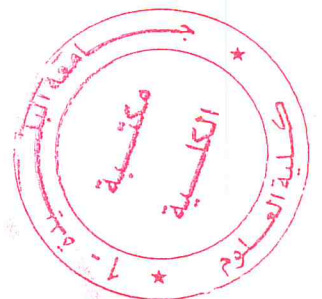
Ces remerciements vont tout d'abord au corps professoral et administratif de la Faculté des Sciences, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Nous tenant à remercier sincèrement Monsieur **REDHOUANE FRIHI**, qui en tant que Directeur de mémoire, s'est toujours montrés à l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi que pour l'inspiration, l'aide et le temps qu'il a bien voulu nous consacrer et sans qui ce mémoire n'aurait jamais vu le jour.

Nous remercions les membres de jury pour avoir accepté d'évaluer notre travail.

Nous n'oublions pas nos parents pour leur contribution, leur soutien et leur patience.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragés au cours de la réalisation de ce mémoire.
Merci à tous et à toutes.



Dédicace

Ce travail modeste est dédié :

A mes parents, aucun hommage ne pourrait être à la hauteur de l'amour
dont ils ne cessent de me combler. Que dieu leur procure
Bonne santé et longue vie.

Mes adorables sœurs : Fatima, Ratiba, Fatiha et Nacera

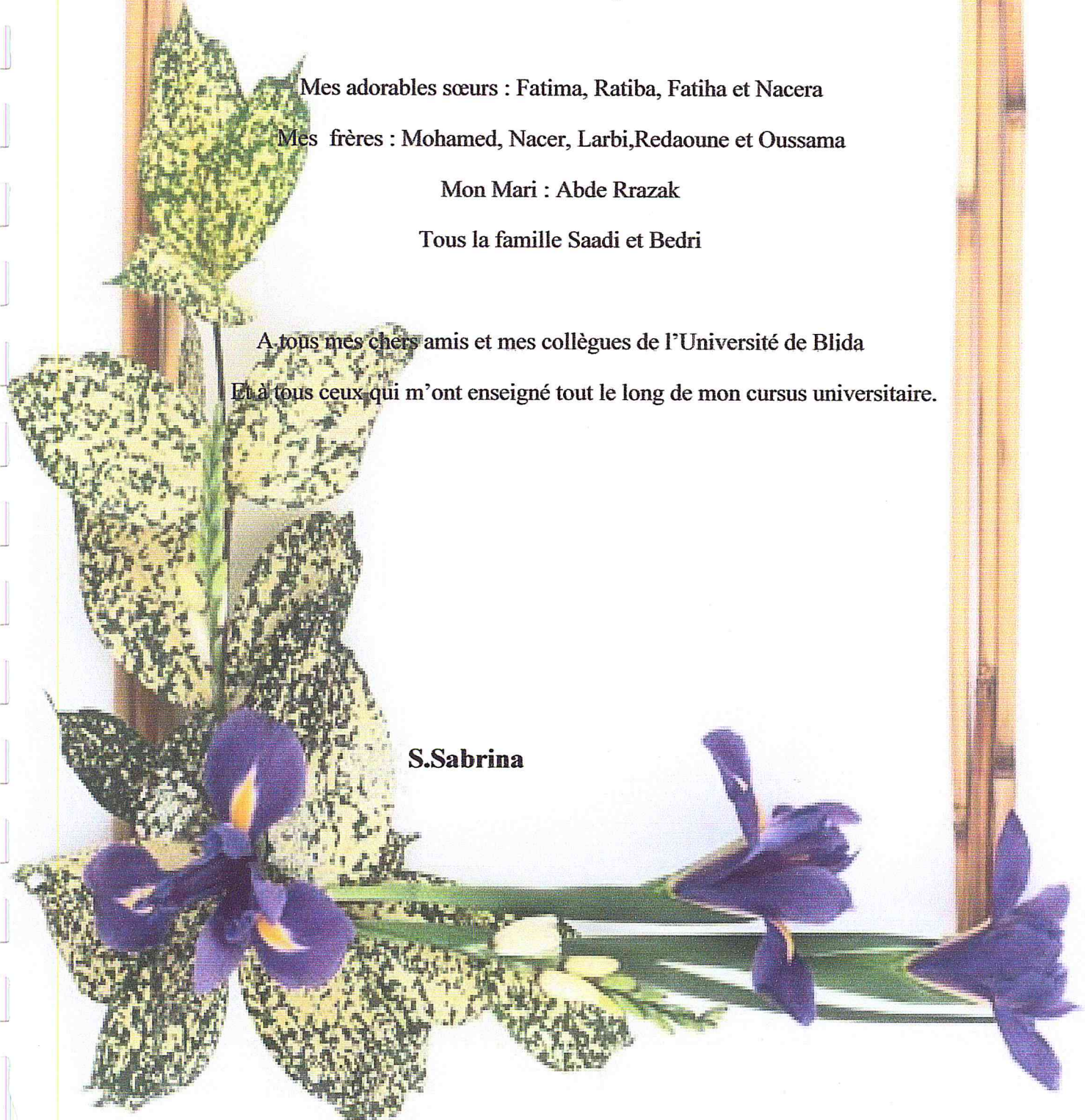
Mes frères : Mohamed, Nacer, Larbi, Redaoune et Oussama

Mon Mari : Abde Rrazak

Tous la famille Saadi et Bedri

A tous mes chers amis et mes collègues de l'Université de Blida
Et à tous ceux qui m'ont enseigné tout le long de mon cursus universitaire.

S.Sabrina



ملخص

يعتبر الارتباط ذات القيم المتغيرين القسوى مهم لنمذجة التبعية بين عدة متغيرات عشوائية تم نمذجته من طرف سكلار 1959 لحساب الاحتمالات. وقد نم طرحه من طرف موريس فريشي لقد اصبح الارتباط اداة مهم لإدراك عدة مجالات تطبيقية كالأرصاد الجوية والهيدرولوجية والمالية. يعود السبب الرئيسي للاهتمام بيه لأنه يعطي عبارات بسيطة لتعبير على الارتباط الذي يجمع هامشية قانون متعددة الأبعاد وبالضبط ثنائي الأبعاد.

في هذه المذكرة قمنا بتعريف الارتباط ذات القيم المتغيرين القسوى و ذكر خصائصه و مختلف أنواعه كما قمنا بتقديم مختلف طرق التقدير الحدودية و الغير حدودية و قمنا في النهاية بتطبيق الارتباط في مجال الأرصاد الجوية.

Résumé

Les copules des valeurs extrêmes bivariées représentent un outil innovant pour modéliser la structure de dépendance de plusieurs variables aléatoires. Introduit par Sklar 1959 pour résoudre un problème de probabilité énoncé par Maurice Fréchet, les copules deviennent essentielles à l'appréhension de nombreux domaines d'application tels que la météorologie, l'hydrologie, les sciences actuarielles et la finance.

Leur grand intérêt est qu'elles fournissent des expressions relativement simples des structures des dépendances liant les marginales d'une loi multidimensionnelle. Plus précisément, pour le cas bidimensionnel.

Dans ce mémoire nous avons commencé par les définitions et les propriétés liées aux copules ainsi que les modèles paramétriques des copules. Ensuite nous présentons les différentes méthodes d'estimation : paramétriques et non-paramétriques et nous terminons par une application des copules en météorologie.

Abstract

The copula of the extreme bivariate values are an innovative tool for the modelisation of the dependence of many random variables introduced by Sklar 1959 for the resolution of a probability problem enunciated by Maurice Fréchet the copulas becomes necessary for the apprehension of lot of areas like metrology, hydrology, actuarial sciences and finance.

The big interest is that they provide simple expression of dependence bound of the marginals of multidimensional law precisely two-dimensional case.

In this memoir we have started with the definitions and properties related to copula models and parametric copula. Then we present the different estimation methods: parametric and non- parametric and we end with an application of copula in meteorology.

Table des matières

Dédicace	i
Remerciement	ii
Table des matières	iii
Liste des figures	v
Liste des tableaux	vi
Introduction	1
I Théorème des valeurs extrêmes	
1.1 Les lois de probabilité	4
1.2 Convergence	7
1.2.1 Convergence en loi	7
1.2.2 Convergence en probabilité	7
1.2.3 Convergence presque sure	7
1.2.4 Convergence dans L^p	7
1.3 Théorème limités	8
1.3.1 lois des grands nombres.....	8
1.3.1.1 loi faible des grands nombres 1.....	8
1.3.1.2 loi faible des grands nombres 2.....	8
1.3.1.3 loi forte des grands nombre.....	8
1.4 Théorème central limité	9
1.4.1 loi forte des grands nombres	9
1.5 Théorème des valeurs extrêmes	9
1.5.1 Les statistique d'ordre	10
1.6 lois des valeurs extrêmes	11
1.7 Loi des valeurs extrêmes généralisées	15

1.7.1	Estimation du quantile des valeurs extrêmes.....	15
1.8	La loi des excès.....	16
1.9	La loi de Pareto Généralisée.....	17
1.9.1	Estimation du quantile de loi de Pareto généralisée.....	19
1.10	Quantile extrême (niveau de retour).....	20
II Théorie des copules		
2.1	Définitions	19
2.1.1	Fonction de densité conjointe d'une copule de dimension d	22
2.1.2	Propriétés des copules	22
2.1.3	(Ordre partiel).....	23
2.2	Mesures de dépendance.....	24
2.2.1	Le coefficient de corrélation linéaire de Pearson.....	24
2.2.2	Tau de Kendall et rho de Spearman	25
2.3	Dépendance de queue.....	26
2.3.1	Coefficients de dépendance de queue.....	26
2.4	Exemple de copules usuelles.....	28
2.4.1	Copules d'indépendance.....	28
2.4.2	Copule de survie.....	28
2.4.3	Copule Gaussienne	29
2.4.4	Copule de Student.....	29
2.4.5	Copule archimédienn.....	30
2.5	Famille des copules archimédienne	31
2.5.1	Copule de Gumbel.....	31
2.5.2	Copule de Clayton	32
2.5.3	Copule de Frank.....	32
2.6	Estimation des copules	33
2.6.1	Estimation non paramétrique	33
2.6.1.1	Copule empirique.....	34
2.6.1.2	Cas des copules bivariées	34

2.6.1.3 Procédure de GENEST et RIVEST	35
2.6.2 Estimation paramétrique.....	35
2.6.1.1 Méthode des moments.....	36
2.6.1.2 Méthode du maximum de vraisemblance (MLE).....	36
3 Copules des valeurs extrêmes	
3.1 Caractérisation d'une copule des valeurs extrêmes bivariées.....	38
3.2 Familles paramétriques usuelles des copules des valeurs extrêmes.....	40
IV Application : Application aux données météorologiques	
I. Simulation.....	46
1. Copule de Gumbel	46
2. Comparaison des méthodes d'estimation des paramètres des copules.....	48
II. Application	48
4.1 Introduction.....	48
4.2 Description de l'échantillon.....	48
1. Graphes des densités de probabilités	49
1. Ajustement des données météorologiques univariées.....	52
3. Estimation des paramètres de la copule de Gumbel.....	53
4. Probabilité de dépassement (de survie)	53
Conclusion.....	54
Annexe	55
Bibliographie	57

Liste des figures

1.1	représenté A gauche : H_γ .A droite : les densités associées à la loi des valeurs extrêmes (noire : $\gamma = 0$, bleu : $\gamma = 1$ et rouge : $\gamma = -1$).....	12
1.2	La fonction de répartition de la loi $GPD_{\gamma,1}$ $GPD_{\xi,1}$	18
1.3	La densité de la loi $GPD_{\gamma,1}$	19
4.1	Fonction de Densité de Gumbel	46
4.2	Fonction de répartition de Gumbel	46
4.3	Densité de copule de Clayton	47
4.4	Fonction de répartition de Clayton.....	47
4.5	Ajustement de la vitesse du vent par loi des valeurs extrêmes.....	52
4.6	Ajustement de Humidité par la loi des valeurs extrêmes	52
4.7	Ajustement de la précipitation par la loi des valeurs extrêmes.....	53

Liste des tableaux

1.1 Tableau Lois de probabilités discrètes.....4

1.2 Le tableau suivant résume les lois de probabilités continues.....5

Tableau 2: Estimation du paramètre de la copule de Gumbel48

4.1 Tableau représente la statistique descriptive.....48

4.2 Tableau Paramètres de la copule de Gumbel.....53

4.3 Tableau Probabilité de dépassement53

Introduction générale

La statistique des valeurs extrêmes est surtout utilisée en finance, hydrologie, assurance, environnement, internet, là où il est important d'estimer les probabilités «improbables».

Les copules constituent un outil statistique qui présente de nombreux avantages, tant pour les statisticiens que pour les financiers. Outre une grande souplesse dans la mise en œuvre de l'analyse multivariée, les copules autorisent une sélection plus étendue des distributions conjointes des séries financières. Les fonctions copules permettent une représentation moins naïve de la dépendance statistique en finance fondée sur la mesure traditionnelle de corrélation qui présente des limites dans l'étude de l'interdépendance entre deux variables (cf. Embrechts et al. (1999)). En outre, elles autorisent des distributions de probabilités jointes moins restrictives, prenant mieux en compte certains faits en finance (*leptokurticité, asymétrie, dépendance des queues*). Elles permettent la construction de distributions multidimensionnelles assez générales et ce, indépendamment des lois des marginales qui peuvent avoir des lois différentes et quelconques. Par conséquent, elles permettent de s'affranchir de certaines hypothèses peu réalistes faites dans les études empiriques.

Par ailleurs, l'approche par les copules a beaucoup contribué dans l'analyse, dans la modélisation statistique multivariée. En effet, la théorie des copules permet une décomposition de la loi multidimensionnelle en ses marginales univariées et en une fonction de dépendance, rendant possibles des extensions naturelles de certains résultats obtenus dans le cas univarié au cas multivarié. Les distributions multidimensionnelles ainsi obtenues sont davantage en adéquation avec la réalité surtout dans l'utilisation financière des statistiques. Par souci de simplicité et du fait que la théorie multivariée est une extension du cas bivarié nous nous limiterons à la théorie bivariée des couples

L'objectif de ce mémoire est :

Ce mémoire s'articule autour de quatre chapitres. Dans le chapitre 1, nous allons exposer brièvement un aperçu sur les lois des probabilités, ainsi les lois asymptotique des échantillons indépendantes de même loi. Nous allons donner la théorie des valeurs extrêmes, ainsi que les théorèmes fondamentaux de cette théorie.

Dans le deuxième chapitre nous présentons la notion des copules et ses propriétés, nous présentons quelques familles de copules paramétriques les plus utilisées en pratique et leurs propriétés fondamentales comme les copules elliptiques et les copules Archimédiennes.

Et de fournir différentes méthodes d'estimation des copules à savoir l'approche paramétrique et non-paramétrique.

La première approche consiste à estimer le paramètre de la copule par l'une des méthodes suivantes : méthode des moments, méthode du maximum de vraisemblance.

Dans la seconde approche on utilise les méthodes suivantes : la méthode de la copule empirique, procédure de Genest et Rivest.

Le chapitre trois est réservé aux copules des valeurs extrêmes bivariées nous allons donner les familles paramétriques usuelles des copules des valeurs extrêmes.

Dans le dernier chapitre nous avons essayé d'exprimer la dépendance des variables météorologique (Précipitation, Température, Humidité, vitesse de vent, Inseillement, Evaporation) par les copules des valeurs extrêmes bivariées exploitant les données météorologiques de l'ONM (office nationale de méthodologique).

Nous clôturons ce travail par une conclusion générale.

Chapitre I

Théorie des valeurs extrêmes

Dans ce chapitre, nous rappelons d'abord quelques notions principales et des théorèmes de convergences avant d'établir le lien entre ces résultats et la théorie des valeurs extrêmes à travers le théorème de Fisher-Tippett un des fondements de cette théorie.

1.1 Les lois de probabilité

Les lois de probabilités sont des objets mathématiques qui permettent aux statisticiens de fabriquer des modèles pour décrire des phénomènes où le hasard intervient. Une loi de probabilité est une distribution théorique de fréquences.

Il existe deux types de lois, les lois discrètes et les lois continues.

Le tableau suivant résume les lois de probabilité discrète.

Tableau -1.1 -Lois de probabilités discrètes

Notation de loi de probabilité	Densité de probabilité	$E(X)$	$Var(X)$
Bernoulli $X \sim B(1, p)$	$P(X = 0) = q$ $P(X = 1) = p$ avec $p + q = 1$ $P(X = x) = p^x(1 - p)^{1-x}$ $x = 0, 1$	p	pq
Binomiale $X \sim B(n, p)$	$P(X = k) = C_n^k p^k q^{n-k}$ avec $p + q = 1$	np	npq
Poisson $X \sim P(\lambda)$	$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$	λ	λ
Géométrique $X \sim G(p)$	$p(1 - p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Le tableau 1.2 suivant résume quelques lois de probabilités continues

Notation de loi de probabilité	Densité de probabilité	$E(X)$	$Var(X)$
Normale $X \sim N(\mu, \sigma^2)$	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ pour $-\infty < x < +\infty$	μ	σ^2
Gamma $X \sim G(\alpha, \lambda)$	$f_X(x) = \frac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Weibull $X \sim W(\lambda, \beta)$	$f_X(x) = \lambda \beta x^{\beta-1} \exp(-\lambda x^\beta)$ pour $x > 0, \lambda > 0$ et $\beta > 0$	$\frac{\Gamma(1 + \beta^{-1})}{\lambda^{\frac{1}{\beta}}}$	$\frac{\Gamma(1 + 2\beta^{-1}) - \Gamma^2(1 + \beta^{-1})}{\lambda^{\frac{2}{\beta}}}$
Bêta $X \sim Be(\alpha, \beta)$	$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$
Lognormale $X \sim LN(\mu, \sigma^2)$	$f_X(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}$ pour $x > 0$ ou $\mu \in \mathbb{R}$ et $\sigma > 0$	$e^{\mu + \frac{1}{2}\sigma^2}$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$
Uniforme $X \sim U_{[a,b]}$	$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Student $X \sim T_n$	$f(t) = c(n) \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$ et $c(n) = \int_{-\infty}^{+\infty} f(x) dx$	$0, si n > 2$	$\frac{n}{n-2}, si n > 2$
Pareto	$f(x) = a \frac{x_0^a}{x^{a+1}} \text{ pour } x \geq x_0, x_0 > 0$	$\frac{ax_0}{a-1}$ pour $a > 1$	$\frac{ax_0^2}{(a-1)^2(a-2)}$ pour $a > 2$
Exponentielle $X \sim \exp(1, \lambda)$	$f_X(x) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Cauchy	$f(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2}, a \in \mathbb{R}$	non définie	non définie

Remarque 1

-Pour la distribution normale, si $\mu = 0$ et $\sigma = 1$.

$X \sim N(0,1)$ est appelée normale centre réduite.

-Pour la distribution Gamma on a :

La fonction Γ est définie par :

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \text{ pour } \alpha > 0.$$

Nous avons trois cas particuliers de la loi Gamma, à savoir :

1-Si α est un entier naturel, alors la loi est dite d'Erlang.

2-Si $\alpha = \frac{n}{2}$, ou $n \in \{1,2, \dots\}$ et $= \frac{1}{2}$, alors la distribution Gamma porte aussi le nom de distribution de Khi-deux(ou Khi-carré) de n degrés de liberté et on noté : $X \sim \chi_n^2$.

3-Si $\alpha = 1$, alors la fonction densité f_X devient : $f_X(x) = \lambda e^{-\lambda x}$, pour $x > 0$.

La distribution est dite Exponentielle de paramètre $\lambda > 0$, on écrit : $X \sim \exp(\lambda)$.

D'espérance : $E(x) = \frac{1}{\lambda}$ et de variance : $Var(X) = \frac{1}{\lambda^2}$.

-Pour la distribution de Weibull, nous remarquons que si $\beta = 1$, la distribution de Weibull généralise la distribution exponentielle.

-Pour la distribution Béta, nous avons un cas particulier.

Si $\alpha = \beta = 1$, alors on a : $X \sim U_{[a,b]}$ si $f_X(x) = \frac{1}{b-a}$ pour $a < x < b$

Avec : $E(X) = \frac{a+b}{2}$ et $Var(X) = \frac{(b-a)^2}{12}$.

-Pour la distribution de Cauchy, nous remarquons que tous les moments divergent.

-Il est important de citer que la loi de Pareto et la loi de Cauchy n'est pas de la famille exponentielle.

1.2 Convergence

Soit (X_n) une suite de v.a. réelles définies sur un espace probabilisé (Ω, \mathcal{B}, P) . On désigne par (F_n) la suite de fonctions de répartition correspondantes. Soit X une v.a. réelle définie sur le même espace (sauf pour la convergence en loi), de fonction de répartition F .

1.2.1 Convergence en loi

La suite (X_n) converge en loi vers X et l'on écrit $X_n \xrightarrow{L} X$

- Si $F_n(x) \rightarrow F(x)$ en tout point de continuité de F .

- Ou si pour toute fonction h continue bornée de \mathbb{R} dans \mathbb{R} , $\lim_n E(h(X_n)) = E(h(X))$.

1.2.2 Convergence en probabilité

La suite (X_n) converge en probabilité vers X et l'on écrit $X_n \xrightarrow{P} X$ si pour tout $\varepsilon > 0$, on a :

$$\lim_n P(\{|X_n - X| > \varepsilon\}) = 0$$

1.2.3 Convergence presque sûre

La suite (X_n) converge presque sûrement vers X et l'on écrit $X_n \xrightarrow{p.s} X$

$$P(\{\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)\}) = 1$$

1.2.4 Convergence dans L^p

La suite (X_n) converge dans L^p vers X et l'on écrit $X_n \xrightarrow{L^p} X$

$$\lim_{n \rightarrow +\infty} E(|X_n - X|^p) = 0$$

1.3 Théorèmes limites

1.3.1 Lois des grands nombres

En statistiques, la loi des grands nombres indique que lorsque l'on fait un tirage aléatoire dans une série de grande taille, plus on augmente la taille de l'échantillon, plus les caractéristiques statistiques de l'échantillon se rapprochent des caractéristiques statistiques de la population. On distingue loi faible et loi forte en fonction du mode de convergence, en probabilité et presque sûre, respectivement.

1.3.1.1 loi faible des grands nombres 1

Soit (X_n) une suite de v.a. indépendantes, de même loi ayant une espérance $E(X)$ finie.

On a alors :

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p} E(X)$$

1.3.1.2 loi faible des grands nombres 2

Soit (X_n) une suite de v.a. indépendantes, ayant une même espérance $E(X)$ finie et même variance σ^2 finie (carré intégrable). On a alors :

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p} E(X)$$

1.3.1.3 loi forte des grands nombres

Soit (X_n) une suite de v.a. indépendantes, de même loi ayant une espérance $E(X)$ finie. On a alors :

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p.s} E(X)$$

1.4 Théorème central limite

1.4.1 loi forte des grands nombres

Soit (X_n) une suite de v.a. indépendantes, de même loi ayant une espérance $E(X)$ finie et une variance $\sigma^2 < +\infty$ finie. On a alors :

$$\frac{X_1 + \dots + X_n - nE(X)}{\sigma\sqrt{n}} \xrightarrow{L} N(0,1)$$

1.5 Théorème des valeurs extrêmes

La théorie des valeurs extrêmes a pour but d'étudier et de caractériser le comportement des valeurs extrêmes d'un échantillon de variables aléatoires. On souhaite estimer des petites probabilités ou des quantités dont la probabilité d'observation est très faible, c'est-à-dire proche de zéro. Ces quantités sont appelées quantiles, on parle de quantile extrême lorsque l'ordre du quantile (probabilité d'observation) converge vers zéro quand la taille de l'échantillon tend vers l'infini.

Ces quantiles extrêmes se situent dans les queues de distributions des lois de probabilité.

Ainsi, à l'inverse de l'approche statistique classique qui s'intéresse au comportement moyen et à la variabilité des phénomènes autour de la moyenne, on s'intéresse ici au comportement des queues de distributions.

Un événement extrême est un événement qui a une faible probabilité de se produire mais qui, lorsqu'il se produit, prend de très petites ou de très grandes valeurs et a un grand impact. On notera la différence avec un événement rare qui, par définition, est un événement dont la probabilité d'occurrence est faible. Le fait qu'un événement soit rare n'implique pas qu'il soit extrême, il est dépourvu de la notion de quantifiabilité (petites ou grandes valeurs). A l'inverse, tout événement extrême est rare au sens où il a une faible probabilité de se produire.

1.5.1 Les statistiques d'ordre

Soient n variables aléatoires $X = (X_1, \dots, X_n)$. Rangeons ces variables aléatoires par "ordre croissant de grandeur". Pour cela, nous introduisons la notation $X_{i:n}$ avec

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n-1:n} \leq X_{n:n}$$

$X_{i:n}$ est donc la i -ième statistique d'ordre (ou statistique d'ordre i) dans un échantillon de taille n . Soient n observations x_1, \dots, x_n . La valeur observée de $X_{i:n}$ est notée $x_{i:n}$ et nous avons

$$x_{i:n} = e_i^T \cdot \text{sort}(x)$$

où sort est la fonction de tri ascendant et x le vecteur des observations x_1, \dots, x_n et e_i^T est le $i^{\text{ème}}$ vecteur unitaire de la base canonique de \mathfrak{R}^n .

Deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes. Ce sont les statistiques d'ordre extrême qui correspondent à la plus petite statistique d'ordre $X_{1:n}$ (ou statistique du minimum)

$$X_{1:n} = \min(X_1, \dots, X_n)$$

et à la plus grande statistique d'ordre $X_{n:n}$ (ou statistique du maximum)

$$X_{n:n} = \max(X_1, \dots, X_n)$$

L'écart $W = X_{n:n} - X_{1:n}$ est appelée la déviation extrême.

Nous pouvons vérifier la relation : $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$.

1.6 Lois des valeurs extrêmes

L'étude de la loi de probabilité du maximum d'un échantillon de n variables a été la première approche pour décrire les événements extrêmes. Fisher et Tippet [en 1928 ont les premiers, déduit de manière heuristique les lois limites possibles pour le maximum d'une suite de variables aléatoires indépendantes et de même loi, avant que Gnedenko en 1943 n'obtienne rigoureusement la convergence, dont la preuve fut simplifiée par de Haan en 1976.

Le théorème ci-dessous est fondamental en théorie des valeurs extrêmes car il établit la loi asymptotique du maximum $X_{n:n}$ convenablement normalisé d'un échantillon.

Théorème 1.6.1 (Fisher-Tippet)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F . S'il existe deux suites normalisantes réelles $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ et $M_n = \max(X_1, \dots, X_n)$ et $M_n = X_{n:n}$ et une loi non dégénérée H_γ telle que :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_\gamma(x)$$

Avec H appartenant à la classe de fonction suivantes:

$$H_\gamma(x) = \begin{cases} e^{-(1+\gamma x)^{-\frac{1}{\gamma}}}, & \text{pour } x > 0, \text{ si } \gamma \neq 0 \\ e^{-e^{-x}}, & \text{pour } x \in \mathbb{R}, \text{ si } \gamma = 0 \end{cases}$$

H_γ est appelée fonction de répartition de la loi des valeurs extrêmes.

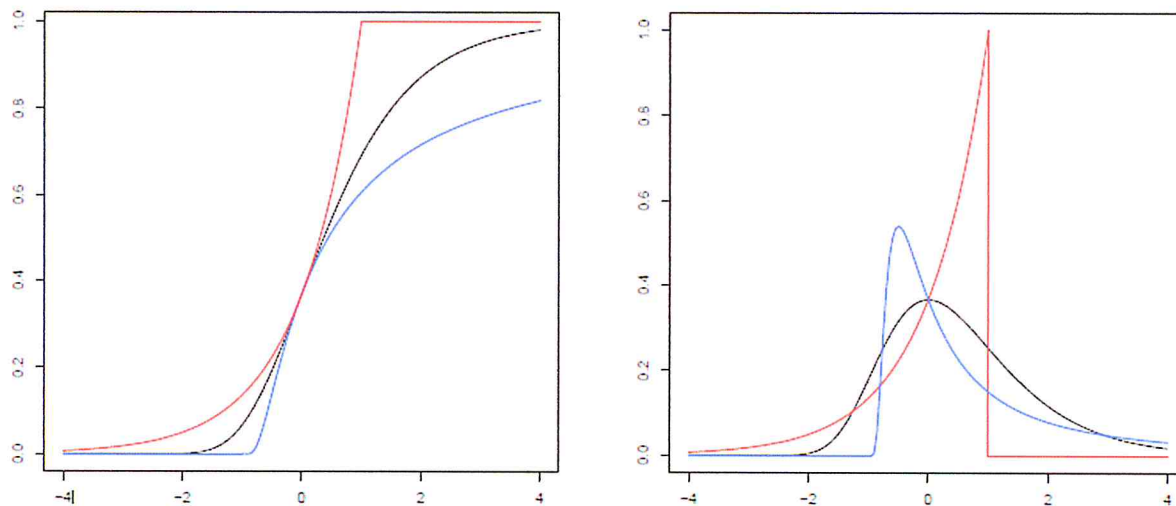
Cette loi dépend du seul paramètre de forme γ appelé indice des valeurs extrêmes ou indice de queue. γ est le paramètre clé de toute la théorie des valeurs extrêmes.

L'estimation de γ nous fournira le comportement de la queue de distribution. En effet, selon son signe, on distingue trois domaines d'attraction dont quelques densités ont été représentées pour γ fixé.

- Si $\gamma > 0$, on dit que F appartient au domaine d'attraction de Fréchet, que l'on notera $D(\text{Fréchet})$. Il contient les lois dont la fonction de survie est à décroissance,

polynomiale, *i.e.* les lois à queues lourdes ou lois de type Pareto. Les lois de ce domaine ont un point terminal x_F infini.

- Si $\gamma < 0$, on dit que F appartient au domaine d'attraction de Weibull, que l'on notera $D(\text{Weibull})$. Toutes les lois de ce domaine d'attraction ont un point terminal x_F fini.
- Si $\gamma = 0$, on dit que F est dans le domaine d'attraction de Gumbel, que l'on notera $D(\text{Gumbel})$. Il contient les lois dont la fonction de survie est à décroissance exponentielle, *i.e.* les lois à queues légères.



1.1 FIGURE : A gauche : H_γ .A droite : les densités associées à la loi des valeurs extrêmes (noire : $\gamma = 0$, bleu : $\gamma = 1$ et rouge : $\gamma = -1$).

Donc les lois usuelles peuvent être classées selon leur domain d'attraction comme dans le tableau suivant:

Fréchet ($\gamma > 0$)	Gumbel ($\gamma = 0$)	Weibull ($\gamma < 0$)
Cauchy Student Pareto	Normale Log-normale Gamma Exponentielle Weibull	Uniforme Béta

Exemples 1: Loi uniforme

On suppose que la loi X_1 est la loi uniforme sur $[0, \theta]$, $\theta > 0$. la fonction de répartition de la loi est $F(x) = \frac{x}{\theta}$, pour $x \in [0, \theta]$.

La suite $(M_n)_{n \geq 1}$ converge *p.s* vers θ .

On note F_n la fonction de répartition de $n \left(\frac{M_n - 1}{\theta} \right) =$

$$= \frac{M_n - \frac{\theta}{n}}{\frac{\theta}{n}}$$

avec $a_n = \frac{\theta}{n}$ et $b_n = \theta$, comme $M_n < \theta$, on a $F_n(x) = 1$.

Si $x \geq 0$, considérons la cas $x < 0$.

$$\begin{aligned} F_n(x) &= \mathbb{P} \left(M_n \leq \theta + \theta \frac{x}{n} \right) = \left(\mathbb{P} \left(X_1 \leq \theta + \theta \frac{x}{n} \right) \right)^n \\ &= \left(1 + \frac{x}{n} \right)^n \end{aligned}$$

Donc $\lim_{n \rightarrow +\infty} F_n(x) = e^x$ pour $x < 0$, donc le domaine d'attraction de la loi uniforme est le domaine de Weibull.

Exemple 2: Loi exponentielle

Dans le cas d'une exponentielle de paramètre $\lambda > 0$, $F(X) = 1 - e^{-\lambda x}$, d'où l'on déduit que :

$$P(\lambda X_{(n)} - \ln(n) \leq x) = P\left(X_1 \leq \frac{x + \ln(n)}{\lambda}\right)^n = \left(1 - \frac{e^{-x}}{n}\right)^n \text{ est donc :}$$

$$\lim_{n \rightarrow +\infty} P(\lambda X_{(n)} - \ln(n) \leq x) = \exp(-\exp(-x))$$

Ce qui montre que le domaine d'attraction de la loi exponentielle est le domaine de Gumbel.

Exemple 3 : loi de Cauchy

On suppose que X_1 la loi de Cauchy (de paramètre $a = 1$), donc la fonction de densité est :

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Comme le support de densité est non borné, il est clair que la suite $(M_n)_{n \geq 1}$ diverge.

On note F_n la fonction de répartition de $\frac{\pi M_n}{n}$, on a :

$$F_n(x) = P\left(M_n \leq \frac{nx}{\pi}\right) = \left(P\left(X_1 \leq \frac{nx}{\pi}\right)\right)^n = \left(1 - \int_{\frac{nx}{\pi}}^{\infty} \frac{1}{\pi(1+y^2)} dy\right)^n$$

Po $x > 0$, on a : $\int_{\frac{nx}{\pi}}^{\infty} \frac{1}{\pi(1+y^2)} dy = \frac{1}{nx} + o((nx)^{-3})$

On a alors pour $x > 0$, $F_n(x) = \frac{1}{nx} + o((nx)^{-3})$

On déduit que $\lim_{n \rightarrow \infty} F_n(x) = e^{-\frac{1}{x}}$, pour $x > 0$.

Donc la loi de Cauchy appartient au domaine d'attraction de Fréchet.

1.7 Loi des valeurs extrêmes généralisées

En probabilité et en statistique, la loi d'extremum généralisée (en anglais: generalized extreme value ou GEV) est une famille de loi de probabilité continues qui servent à représenter des phénomènes de valeurs extrêmes.

L'approche par dépassements de seuil est une alternative à la loi GEV dans la modélisation du comportement du maximum d'un échantillon se basant sur les "grandes valeurs" de l'échantillon.

Elle se base sur le résultat mathématique, affirmant qu'il existe une équivalence entre la loi GEV et la loi des dépassements de seuil aussi appelée loi des excès.

La fonction de répartition GEV a la forme suivante:

$$H_{\gamma, \mu, \sigma}(x) = \begin{cases} e^{-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}}, & \text{pour } x > 0, \text{ si } \gamma \neq 0 \\ e^{-e^{-\frac{x - \mu}{\sigma}}}, & \text{pour } x \in \mathbb{R}, \text{ si } \gamma = 0 \end{cases}$$

Telle que : μ est le paramètre de position, σ est le paramètre de dispersion et γ est l'indice de queue.

1.7.1 Estimation du quantile des valeurs extrêmes

Il est possible d'estimer un quantile extrêmes en utilisant le résultat précédent sur la GEV.

Supposons que $(X_i)_{i=1, \dots, n}$ un échantillon de variables aléatoires i.i.d. de la loi F , on sépare ces variables en k blocs de taille r , si r est assez grand, alors on fait l'hypothèse que la distribution de M_r est une distribution GEV de paramètres $\hat{\mu}, \hat{\sigma}$ et $\hat{\gamma}$.

Or la distribution de M_r est F^r car les X_i sont indépendantes, alors on a l'approximation :

$$F^r(x) = H_{\hat{\mu}, \hat{\sigma}, \hat{\gamma}}(x)$$

$$F^r(x) = \begin{cases} e^{-\left(1 + \hat{\gamma} \frac{x - \hat{\mu}}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}}} & \text{pour } x > 0, \text{ si } \gamma \neq 0 \\ e^{-e^{-\frac{x - \hat{\mu}}{\hat{\sigma}}}} & \text{si } \gamma = 0 \end{cases}$$

On cherche q telle que $p = P[X > q] = 1 - F(q)$.

Pour un quantile p élevé il suffit de prendre le cas $\gamma \geq 0$, cette condition borne inférieurement la fonction de répartition.

Dans le cas $\gamma > 0$, il faut résoudre l'équation :

$$p^r = e^{-\left(1 + \gamma \frac{\hat{q} - \hat{\mu}}{\hat{\sigma}}\right)^{\frac{1}{\gamma}}}$$

ce qui donne finalement

$$\hat{q} = \hat{\mu} + \frac{\hat{\sigma}}{\gamma} \left((-r \ln(p))^{-\gamma} - 1 \right)$$

Ce fournit une estimation d'un quantile extrême.

Dans le cas $\gamma = 0$, on a :

$$p^r = e^{-e^{\frac{\hat{q} - \hat{\mu}}{\hat{\sigma}}}}$$

Donc on trouve

$$\hat{q} = \hat{\mu} + \hat{\sigma}(-r \ln(p))$$

Il existe une autre approche de la théorie des valeurs extrêmes permettant d'estimer des quantiles est loi de Pareto généralisée.

1.8 La loi des excès

La nouvelle approche de la théorie des valeurs extrêmes appelée *POT* consiste à utiliser les observations qui dépassent un certain seuil, plus particulièrement les différences entre ces observations et le seuil, appelées excès. Il est clair que cette méthode nécessite la détermination d'un seuil ni trop faible pour ne pas prendre en considération des valeurs non extrêmes, ni trop élevé pour avoir suffisamment d'observations. Notons le seuil par μ .

Définition 1.8.1

Soit X une variable aléatoire de fonction de répartition F et de point terminal x_F . Pour tout $\mu < x_F$, la fonction

$$F_\mu(x) = P[X - \mu \leq x | X > \mu], x \geq 0$$

est appelée fonction de répartition des excès au dessus du seuil μ .

Définition 1.8.1

Soit X une variable aléatoire de fonction de répartition F et de point terminal x_F . Pour tout $\mu < x_F$, la fonction

$$F_\mu(x) = P[X - \mu \leq x | X > \mu], x \geq 0$$

est appelée fonction de répartition des excès au dessus du seuil μ .

Remarque 2 Par définition des probabilités conditionnelles, F_μ peut être également définie par :

$$F_\mu(x) = \begin{cases} \frac{F(\mu+x) - F(\mu)}{1 - F(\mu)} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Notons $Y = X - \mu$ pour $X > \mu$ et pour n v.a. observées X_1, \dots, X_n nous pouvons écrire $Y_j = X_j - \mu$ telle que i est l'indice du j -me excès et $j = 1, \dots, N_\mu$.

Le *Mean Excess Plot* appelé aussi le *Mean Residual life Plot* est un outil spécifique pour retenir le seuil performant.

De plus, nous approchons la loi des excès (Y_1, \dots, Y_{N_μ}) par une loi de Pareto Généralisée $GPD_{\gamma, \sigma}$ (*Generalized Pareto Distribution*) que nous présentons ci-dessous.

1.9 La loi de Pareto Généralisée

Le théorème de Pickands est très utile lorsqu'on travaille avec des observations qui dépassent un seuil fixé puisqu'il assure que la loi des excès peut être approchée par une loi de Pareto généralisée.

Définition 1.9.1

Soient $\sigma(\mu)$ une fonction strictement positive et $\gamma \in \mathbb{R}$. La loi de Pareto généralisée a pour fonction de répartition $G_{\gamma, \sigma}$:

$$G_{\gamma, \sigma(\mu)}(x) = \begin{cases} 1 - \left(1 + \gamma \left(\frac{x}{\sigma(\mu)}\right)\right)^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma(\mu)}\right) & \text{si } \gamma = 0 \end{cases}$$

où $x \geq 0$ si $\gamma \geq 0$ et $0 \leq x \leq -\frac{\sigma(\mu)}{\gamma}$ si $\gamma < 0$.

Théorème 1.9.2 Pickands

Si F appartient à l'un des trois domaines d'attraction de la loi des valeurs extrêmes (Fréchet, Gumbel ou Weibull), alors il existe une fonction $\sigma(\mu)$ strictement positive et un réel γ tels que

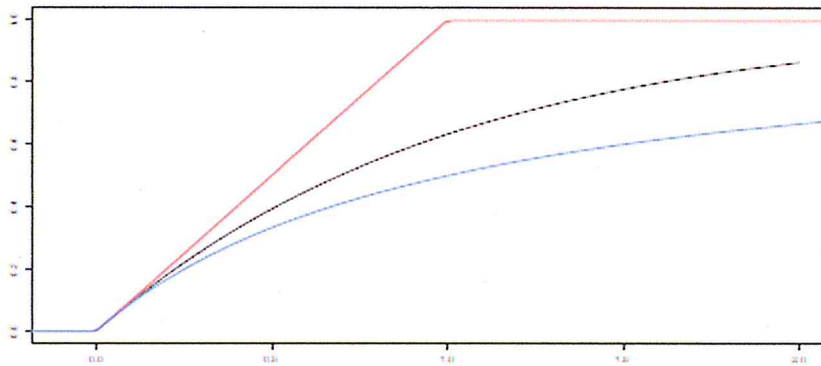
$$\lim_{\mu \uparrow x_F} \sup_{0 \leq x \leq x_F - \mu} |F_\mu(x) - G_{\gamma, \sigma(\mu)}(x)| = 0$$

où $G_{\gamma, \sigma(\mu)}$ est la fonction de répartition de la loi de Pareto Généralisée et F_μ est la fonction de répartition des excès au delà du seuil μ .

Ainsi, pour u grand, la loi des excès est approchée par une loi Pareto généralisée.

$$F_\mu \approx G_{\gamma, \sigma(\mu)}$$

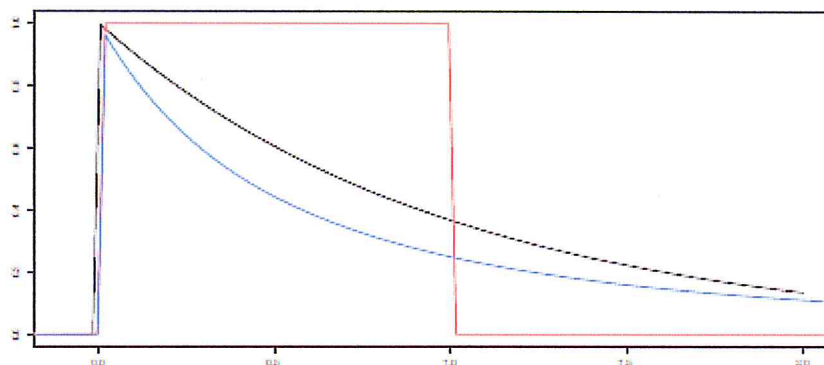
Voici une représentation de la $GPD_{\gamma, \sigma}$ pour certaines valeurs de γ



1.2 FIGURE – La fonction de répartition de la loi $GPD_{\gamma,1}$ $GPD_{\xi,1}$.

Les courbes rouge, noire foncée et bleue correspondent respectivement aux lois de Pareto généralisée d'indices $\gamma = -1, \gamma = 0$ et $\gamma = 1$ avec $\sigma = 1$ fixe.

Nous remarquons que plus les valeurs extrêmes sont extrêmes, plus le paramètre γ est élevé, ça se voit très bien dans le graphe de la fonction de répartition

1.3 FIGURE – La densité de la loi $GPD_{\gamma,1}$ 

1.9.1 Estimation du quantile de loi de Pareto généralisée

Supposons que on a un seuil μ convenable, donc on approche la quantité $\frac{F(x + \mu) - F(\mu)}{1 - F(x)}$

par $G_{\hat{\gamma}, \hat{\sigma}}(x)$, où $\hat{\gamma}$ et $\hat{\sigma}$ sont estimés sur ces données.

En posant $y = x + \mu$, alors

$$\frac{F(y) - F(\mu)}{1 - F(\mu)} = G_{\hat{\gamma}, \hat{\sigma}}(y - \mu)$$

Notre objectif est de trouver y telle que $F(y) = p$, avec p fixé, lorsque $\gamma > 0$ relation

$$y = \mu + G_{\hat{\gamma}, \hat{\sigma}}^{-1}\left(\frac{p - F(\mu)}{1 - F(\mu)}\right)$$

$$\Rightarrow y = \mu + \frac{\hat{\sigma}}{\hat{\gamma}} \left(\left(1 - \frac{p - F(\mu)}{1 - F(\mu)}\right)^{-\hat{\gamma}} - 1 \right)$$

Pour estimer y , en estimant $1 - F(\mu)$ par $\frac{N_\mu}{n}$ qui est la proportion d'observation du seuil μ :

$$y = \mu + \frac{\hat{\sigma}}{\hat{\gamma}} \left(\left(\frac{n}{N_\mu} (1 - p) \right)^{-\hat{\gamma}} - 1 \right)$$

De même lorsque $\gamma = 0$.

$$y = \mu + G_{\hat{\gamma}, \hat{\sigma}}^{-1} \left(\frac{p - F(\mu)}{1 - F(\mu)} \right)$$

ce qui donne :

$$y = \mu - \hat{\sigma} \ln \left(\frac{n}{N_\mu} (1 - p) \right)$$

N_μ est le nombre de observations au-dessus du seuil μ .

$\hat{\gamma}$ et $\hat{\sigma}$ sont les estimateurs des paramètres de la loi GPD.

1.10 Quantile extrême (niveau de retour)

Dans le cas de GEV et GPD on peut obtenir ainsi un estimateur de la période de retour.

L'estimation d'une période de retour associée à une probabilité qui représente la probabilité d'excès au-delà d'un niveau de retour, cette période s'interprète comme une moyenne du temps ou du nombre d'années séparant un événement de grandeur donnée d'un second événement d'une grandeur égale ou supérieure.

On donne les deux tableaux suivants qui résume les formules de temps de retour dans la cas de GEV et GPD.

Pour GEV

$\xi \neq 0$	$\xi = 0$
$\mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}]$	$\mu - \sigma \log\{-\log(1 - p)\}$

Pour GPD

$\xi \neq 0$	$\xi = 0$
$\mu - \frac{\sigma}{\xi} (1 - (1 - p)^\xi)$	$\mu - \sigma (\log(1 - p))$

Chapitre II

Théorie des copules

Remarque 1 : Il suffit de définir C sur $[0,1]$.

Ce théorème est fondamentale dans la théorie des copules, établi par Sklar en 1959, il met la relation de la loi jointe d'un couple aléatoire et ses marginales.

Théorème 2.1 (Sklar 1959) : Si H est une fonction de répartition conjointe avec des marges F_1, F_2, \dots, F_d .

Alors, il existe une copule C telle que pour tout $x_1, x_2, \dots, x_d \in \mathbb{R}$:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (1.2)$$

Si F_1, \dots, F_d , sont continues, alors C est unique, sinon C est déterminée *image* $F_1 \times \dots \times \text{image} F_d$. Réciproquement si C est une copule et F_1, F_2, \dots, F_d sont des fonctions de répartition, alors la fonction H , définie par l'équation (1.2) est une de répartition conjointe avec les marges F_1, F_2, \dots, F_d .

Commentaires

- La loi d'un vecteur aléatoire (X_1, X_2, \dots, X_d) , c'est des d fonctions de répartition, qui capturent les marginales, et une copule qui mesure la dépendance.
- En pratique, on connaît «bien» les marginales et mal la dépendance.
- Difficulté : choisir la copule qui capture le mieux les structures de dépendance entre les données.

Du théorème de Sklar, nous pouvons tirer le corollaire suivant qui donne la relation entre la loi conjointe et copule.

Corollaire 2.1 (Sklar) : Si H est une fonction de répartition conjointe avec les marges F_1, \dots, F_d , alors $F_1^{-1}, \dots, F_d^{-1}$ sont les fonctions inverses de F_1, \dots, F_d . Avec :

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

Alors pour tout $(u_1, \dots, u_d) \in \text{domaine } C$:

$$C(u_1, \dots, u_d) = H(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \quad (1.3)$$

Le théorème de Sklar implique que pour une distribution conjointe multivariée et continue, les marges et la structure de dépendance peuvent être dissociées de façon unique. De plus, la structure de dépendance est représentée par la copule C .

Théorème 2.2 (Bornes de Fréchet-Hoeffding) : $\forall (u_1, \dots, u_d) \in \text{domaine de } C$ nous avons :

$$W(u_1, \dots, u_d) = \max(u_1 + \dots + u_d - d, 0) \leq C(u_1, \dots, u_d) \leq \min(u_1, \dots, u_d) = M(u_1, \dots, u_d) \quad (1.4)$$

Le théorème de Fréchet-Hoeffding implique qu'une copule (mesure de dépendance) est bornée, elle se trouve à l'intérieur d'un minimum et d'un maximum.

Pour les distributions conjointes bivariées, les bornes inférieures et supérieures de Fréchet-Hoeffding sont des copules. W et M sont les copules des vecteurs aléatoires $(U, 1-U)$ et (U, V) respectivement loi uniforme sur $[0, 1]$.

Dans ce cas, W représente la parfaite dépendance négative et M la parfaite dépendance positive. Notons que pour les copules de dimensions $n > 2$, W n'est pas une copule.

Les théorèmes (1) et (2) ont de nombreuses implications pratiques. Vu que les mesures obtenues par les copules sont bornées, cela implique que ces mesures existent toujours.

Les copules sont de plus invariantes sous les transformations monotones croissantes de leurs marges.

Corollaire 2.2 (inversion de Sklar) : Soient H la fonction de répartition conjointe et continue, C est copule, F, G deux fonction de répartition marginales, supposons que F et G sont continues, $\forall u, v \in [0, 1] \times [0, 1]$ on a :

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \quad (1.5)$$

Le théorème de Sklar peut être utilisé pour construire des fonction copules à partir des distribution bidimensionnelles. Considérons par exemple les distributions bivariées suivantes :

Exemple 1 (la distribution bivariée de Galambos)

Soit H_θ une fonction de distribution donnée par :

$$H_\theta(x, y) = \begin{cases} \exp\left\{-\left[(x+y) - (x^{-\theta} + y^{-\theta})^{\frac{-1}{\theta}}\right]\right\} & \text{si } x \geq 0, y \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (1.6)$$

Pour tout réel $\theta \geq 1$.

Cette fonction de distribution admet pour marginales

$$F(x) = e^{-x}, G(y) = e^{-y} \quad \forall x, y \in \mathbb{R}^+$$

Les fonctions inverses sont $F^{-1}(u) = -\ln(u), G^{-1}(v) = -\ln(v)$ respectivement, pour $u, v \in [0, 1]$. on en déduit que la copule associée :

$$C_\theta(u, v) = uv \exp\left\{-\left[(-\ln u)^{-\theta} + (-\ln v)^{-\theta}\right]^{\frac{-1}{\theta}}\right\} \quad (1.7)$$

2.1.1 Fonction de densité conjointe d'une copule de dimension d

Si nous avons une distribution multivariée, absolument continue, elle admet alors une densité conjointe:

$$f(x_1, x_2, \dots, x_d) = c(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i) \quad (1.8)$$

C la fonction de densité de la copule est donnée par la dérivée partielle suivante :

$$c(u_1, u_2, \dots, u_d) = \frac{\partial^d C(u_1, u_2, \dots, u_d)}{\partial u_1 \partial u_2 \dots \partial u_d}.$$

2.1.2 Propriétés des copules

Les résultats suivants donnent les propriétés et les théorèmes les plus importants d'une copule bivariée.

Théorème 1.1 (Continuité) : 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 :

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1| \quad (1.9)$$

Théorème 1.2 (Différentiabilité) : Soit C une copule bivariée, $\forall u_1, u_2 \in [0, 1]$

- 1 les dérivés partielles $\frac{\partial C(u_1, u_2)}{\partial u_j}$ existent p.s et $0 \leq \frac{\partial C(u_1, u_2)}{\partial u_j} \leq 1, \forall j = 1, 2$

2 les fonctions $u \mapsto \frac{\partial C(u,v)}{\partial v}$ et $v \mapsto \frac{\partial C(u,v)}{\partial u}$ sont bien définies et décroissantes sur $[0,1]$ p.s .

Exemple 2 : La dérivé partielle de la copule $W(u,v) = \max(u+v-1, 0)$ existe si $u+v-1 \neq 0$ c'est-à-dire $v = 1-u$, et on a $\max(u+v-1, 0) = u+v-1 \geq 0$, ce qui donne :

$$\frac{\partial C}{\partial u} = \frac{\partial C}{\partial v} = \begin{cases} 0 & \text{si } v < 1-u \\ 1 & \text{si } v > 1-u \end{cases} \quad (1.10)$$

Définition 2.1.3 (Ordre partiel)

Soient C_1 et C_2 deux copules. On dit que C_1 est plus petite que C_2 (ou C_2 plus grand que C_1) et on note $C_1 \prec C_2, C_2 \succ C_1$, si :

$$C_1(u,v) \leq C_2(u,v) \quad \forall u,v \in [0,1]$$

Cette relation d'ordre est partielle car on ne peut pas comparer toutes les copules entre elles. Néanmoins, nous avons toujours

$$C^- \prec C \prec C^+$$

Nous en déduisons qu'une structure de dépendance positive est une fonction copule C qui vérifie l'inégalité suivante :

$$C^\perp \prec C \prec C^+$$

Avec

$$C^- = (u_1, \dots, u_d) = \max\left(\sum_{i=1}^d u_i - d + 1, 0\right)$$

$$C^+ (u_1, \dots, u_d) = \min(u_1, \dots, u_d)$$

$$C^\perp (u_1, \dots, u_d) = \prod_{i=1}^d u_i$$

De même, une structure de dépendance négative est une fonction copule C qui vérifie l'inégalité suivante :

$$C^- \prec C \prec C^\perp$$

Cependant, comme cette relation d'ordre est partielle, il existe des fonctions copules telles que C n'est pas plus petite que C^\perp et C n'est pas plus grande que C^\perp . Il existe donc des fonctions copules qui n'ont pas de structure de dépendance positive, ni de structure de dépendance de dépendance négative.

2.2 Mesures de dépendance

Il existe en statistique un certain nombre de grandeurs proposées par leurs auteurs comme des mesures de la dépendance entre deux variables aléatoires.

2.2.1 Le coefficient de corrélation linéaire de Pearson

Définition 1: Soient X et Y deux variables aléatoires ayant des variances finies.

Le coefficient de corrélation linéaire de Pearson définit par :

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Où $\text{Cov}(X, Y) = E(X.Y) - E(X)E(Y)$ est la covariance entre X et Y .

$r(X, Y)$ est la mesure la plus connue. Ce coefficient est une mesure très imparfaite de la dépendance. Il ne peut se mettre sous une forme ne faisant intervenir que la copule, les marginales restent présentes dans l'expression : ce coefficient n'est donc pas une mesure de dépendance. Il n'intègre que la composante linéaire de la dépendance, au sens où $r(X, Y) = \pm 1$ si et seulement si $Y = aX + b$. Au surplus, pour les marginales fixées, toutes les valeurs comprises entre -1 et $+1$ ne sont pas atteignables.

- X et Y indépendantes $\Rightarrow X$ et Y non corrélées
- X et Y indépendantes $\Rightarrow r(X, Y) = 0$ mais la réciproque est fautive sauf dans le cas où les variables sont gaussiennes car la dépendance est alors entièrement caractérisée par le coefficient de corrélation.

Le contre-exemple le plus connu dans la littérature est le suivant :

Soit $X \sim N(0, 1)$ et $Y = X^2$, alors $\text{Cov}(X, Y) = E(X^3) = 0$.

2.2.2 Tau de Kendall et rho de Spearman : Le tau de Kendall et le rho de Spearman sont deux mesures d'association bivariées qui mesurent la concordance. Elles donnent une mesure de la corrélation entre les rangs des observations, à la différence du coefficient de corrélation linéaire qui lui apprécie la corrélation entre les valeurs des observations. Elles offrent par ailleurs l'avantage d'exprimer simplement en fonction de la copule associée au couple de variables aléatoires.

Les définitions de ces deux coefficients sont intimement liées à la notion de concordance.

a-Notion de la concordance :

Définition 2.3 : 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 :**

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

Plus généralement soit $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ un échantillon de n observations d'un couple (X, Y) . Il existe $C_n^2 = \frac{n!}{2!(n-2)!}$ paires de distributions de couples $(x_i, y_i), (x_j, y_j)$ qui sont soit concordantes, soit discordantes.

Définition 2.4 : La fonction de concordance entre les deux couple (X_1, Y_1) et (X_2, Y_2) est définie par :

$$Q = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

b- Le tau de Kendall

Définition 2.5 : Soient (X_1, Y_1) et (X_2, Y_2) deux vecteurs aléatoires continus i.i.d de fonction de répartition conjointe H et de fonction marginales F (pour X_1 et X_2) et G (pour Y_1 et Y_2). Le tau de Kendall noté τ est défini par :

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Le tau de Kendall possède les propriétés suivantes :

- Le tau de Kendall est symétrique, c'est-à-dire : $\tau(X, Y) = \tau(Y, X)$.
- $-1 \leq \tau \leq +1$

- Si X et Y sont indépendantes alors $\tau = 0$.
- Si a et b sont des fonctions strictement croissantes, alors $\tau(a(X), b(Y)) = \tau(X, Y)$.
- $\tau = 0$ X et Y ne sont pas forcément indépendantes.

c-Rho de Spearman

Définition 2.6 : Soient (X_1, Y_1) et (X_2, Y_2) et (X_3, Y_3) trois vecteurs aléatoires indépendants de même lois H . Le coefficient de corrélation de Spearman est défini par :

$$\rho_s = 3([P(X_1 - X_2)(Y_1 - Y_3) > 0] - [P((X_1 - X_2)(Y_1 - Y_3) < 0)])$$

Les propriétés du ρ_s sont les suivants :

- $-1 \leq \rho_s \leq +1$
- Si X et Y sont indépendantes alors $\rho_s = 0$
- Si a et b sont des fonctions strictement croissantes, alors

$$\rho_s(a(X), b(Y)) = \rho_s(X, Y)$$

Le rho de Spearman s'écrit aussi en fonction du coefficient ρ de corrélation linéaire de Pearson : $\rho_s(X, Y) = \rho(F_x(X), F_y(Y))$ où F_x et F_y sont les fonctions de répartition respectives de X et Y .

2.3 Dépendance de queue

Le concept de dépendance de queue fournit une description de la dépendance au niveau des queues de distribution, très intéressante pour étudier la survenance simultanée de valeurs extrêmes. C'est une mesure locale contrairement au tau de Kendall et au rho de Spearman qui mesurent la dépendance sur l'ensemble de la distribution.

2.3.1 Coefficients de dépendance de queue

Le coefficient de dépendance de queue inférieure ou *lower tail dependence coefficient* de deux variables aléatoires X et Y , de fonctions de répartition respectives F_x et F_y est défini par :

$$\lambda_L(X, Y) = \lim_{\alpha \rightarrow 0^+} \Pr[X \leq F_x^{-1}(\alpha) | Y \leq F_y^{-1}(\alpha)] \quad (\text{si cette limite existe}).$$

Le coefficient de dépendance de queue supérieure (*upper tail dependence coefficient*) de deux variables aléatoires X et Y , de fonctions de répartition respectives F_X et F_Y , est défini par :

$$\lambda_U(X, Y) = \lim_{\alpha \rightarrow 1^-} \Pr \left[X > F_X^{-1}(\alpha) \mid Y > F_Y^{-1}(\alpha) \right] \quad (\text{si cette limite existe}).$$

On dit que X et Y sont asymptotiquement dépendantes au niveau supérieur de la queue de distribution si $\lambda_U \in [0, 1]$ et asymptotiquement indépendantes au niveau supérieur de la queue de distribution si $\lambda_U = 0$.

La notion de copule de survie est très utile pour l'étude de la dépendance de queue. La définition en est la suivante.

Définition 2.7 (copule de survie) : Soit $\tilde{C}(u_1, \dots, u_d)$ la fonction définie par

$$\tilde{C}(u_1, \dots, u_d) = \bar{C}(1 - u_1, \dots, 1 - u_d) \quad \text{où} \quad \bar{C}(u_1, \dots, u_d) = \Pr [U_1 > u_1, \dots, U_d > u_d].$$

Alors $\tilde{C}(u_1, \dots, u_d)$ est appelée copule de survie (survival copula) de la copule C .

Définition 2.8 : Soient X et Y deux variables aléatoires de copule C , alors on a

$$\lambda_L(X, Y) = \lim_{u \rightarrow 0^+} \frac{C(u, v)}{u} \quad \text{et} \quad \lambda_U(u, v) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, v)}{1 - u}$$

On montre facilement que le coefficient de dépendance de queue inférieure (resp. supérieure) de C est le coefficient de dépendance de queue supérieure (resp. inférieure) de \tilde{C} . Autrement dit, à partir d'une copule donnée, il est possible de créer une autre copule présentant une structure de dépendance de queue inversée.

Exemple 3 : Soit la copule de Gumbel donnée par :

$$C_\theta(u, v) = \exp \left(- \left[(-\ln u)^\theta + (\ln v)^\theta \right]^{\frac{1}{\theta}} \right), \theta \geq 0$$

Alors

$$\frac{1 - 2u + C(u, v)}{1 - u} = \frac{1 - 2u + \exp(2^{\frac{1}{\theta}} \ln u)}{1 - u} = \frac{1 - 2u + u^{2^{\frac{1}{\theta}}}}{1 - \theta}$$

Par conséquent

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

La famille de Gumbel possède pour $\theta > 1$, une dépendance au niveau supérieur de la queue de distribution.

2.4 Exemple de copules usuelles

2.4.1 Copules d'indépendance

Soient X, Y deux v.a continues, et H la fonction de répartition jointe dont les marginales sont F et G .

Définition 2.9 : Si X et Y sont indépendantes, alors la copule associée est le produit de ses marginales comme suit :

$$C_{X,Y}(x, y) = F(x)G(y).$$

2.4.2 Copule de survie

Cette copule est très intéressante, car dans la majorité des applications on s'intéresse à la durée de vie des individus dans une certaine population. Dans le cas univarié la probabilité de survie est définie par :

$$P(X > x) = 1 - F(x)$$

telle que F représente la fonction de répartition de X .

De même pour le cas bivarié, si H est la fonction de répartition jointe associée au couple aléatoire (X, Y) , et soit $C_{X,Y}$ sa copule de la fonction jointe de survie est

$\bar{H}(x, y) = P(X > x, Y > y)$ et les marginales sont \bar{F} et \bar{G} , alors:

$$\begin{aligned} \bar{H} &= P(X > x, Y > y) \\ &= 1 - P(X \leq x, Y \leq y) \\ &= 1 - F(x) - G(y) + H(x, y) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + G_{X,Y}(1 - \bar{F}(x), 1 - \bar{G}(y)) \end{aligned}$$

Donc la copule de survie \hat{C} de X, Y est

$$\hat{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y)) = \bar{F}(x) + \bar{G}(y) - 1 + C_{X,Y}(1 - \bar{F}(x), 1 - \bar{G}(y)).$$

2.4.3 Copule Gaussienne

La copule gaussienne de deux variables aléatoires X, Y est définie par ::

$$C_{\rho}^{\phi}(u, v) = \phi_{\rho}(\phi^{-1}(u), \phi^{-1}(v)).$$

telle que ϕ_{ρ} est la fonction de répartition jointe de la loi normale bivariée, de coefficient de corrélation linéaire $\rho \in [-1, 1]$ et ϕ est la fonction de la loi normale standard, Par suite :

$$\phi_{\rho}(\phi^{-1}(u), \phi^{-1}(v)) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{2\rho_{X,Y}st - s^2 - t^2}{1-\rho_{X,Y}^2}\right) ds dt.$$

Cette copule est paramétrée par le coefficient de corrélation linéaire ρ .

$-C_{\rho}^{\phi}(u, v) \rightarrow W(u, v)$ quand $\rho \rightarrow -1$ et $-C_{\rho}^{\phi}(u, v) \rightarrow M(u, v)$ quand $\rho \rightarrow 1$.

2.4.4 Copule de Student

Dans le cas univarié, la fonction de répartition de Student est définie par :

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

Telle que Γ est la fonction d'Euler.

Dans le cas bivarié, soit $\rho \in [-1, 1]$, alors la fonction de répartition bivariée est

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

Définition 2.10 : La copule de Student est une copule paramétrique, paramétrée par le coefficient de corrélation linéaire ρ et de degré de liberté ν . Cette copule est définie par

$$C'_{\rho,\nu}(u, \nu) = t_{\rho,\nu}(t_v^{-1}(u), t_v^{-1}(\nu)) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(\nu)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 + t^2 - 2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds dt$$

Sa densité est définie par

$$c'_{\rho,\nu}(u, \nu) = \rho^{\frac{1}{2}} \frac{\Gamma\left(\frac{\nu+2}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)^2} \left(\frac{\left(1 + S_1^2 + S_2^2 - 2\rho S_1 S_2\right)^{-\left(\frac{\nu-2}{2}\right)}}{\nu(1-\rho^2)} \right) \left(\prod_{j=1}^2 (1 + S_j^2)^{-\left(\frac{\nu+2}{2}\right)} \right)$$

Où $S_1 = t_v^{-1}(u), S_2 = t_v^{-1}(\nu)$.

Remarque 2.1.2

- (a) La copule gaussienne et la copule de Student appartiennent à la famille de copules elliptique.
- (b) Si le degré de liberté $\nu \rightarrow \infty$, alors la copule de Student converge vers la copule gaussienne et dans ce cas très difficile de différencier entre ces deux copules.

2.4.5 Copule Archimédienne

Cette famille de copule a été nommé par Ling (1965), mais il a été reconnu par Schweizer et Sklar (1961) dans l'étude de t-norme. Avant d'être introduit dans la finance et aussi dans d'autres domaines. Mais avant de définir cette famille de copules, on a besoin de présenter quelque définitions, et propriétés utiles pour définir les copules archimédiennes.

Définition 2.11 : Soit $\varphi: [0,1] \rightarrow \mathbb{R}^{+*}$, continue, décroissante et convexe, telle que $\varphi(1) = 0$, alors φ est dite générateur la pseudo-inverse de φ est définie par :

$$\varphi^{-1}(u) = \begin{cases} \varphi^{-1}(u) & \text{si } 0 \leq u \leq \varphi(0) \\ 0 & \text{si } \varphi(0) \leq u \leq +\infty \end{cases}$$

Définition 2.12 : Une copule est dite Archimédienne si elle s'écrit sur la forme :

$$C^A(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)).$$

Propriété 2.1.1

Cette copule est

a) **Symétrique**

$$C^A(u, v) = C^A(v, u), \forall (u, v) \in [0, 1]^2.$$

b) **Associative**

c)

$$C^A(C^A(u, v), z) = C^A(u, C^A(v, z)), \forall (u, v, z) \in [0, 1]^3.$$

d) **De contour convexe**

$$\{(u, v) \in [0, 1]^2 : \varphi(u) + \varphi(v) = \varphi(k)\}, k > 0.$$

e) **De densité**

$$c^A(u, v) = \frac{-\varphi''(C(v, u))\varphi'(u)\varphi'(v)}{(\varphi'(C(v, u)))^3}.$$

f) Si c est une constante strictement positive, alors $c\varphi$ un générateur de la copule C^A .

Exemple 5

(a) Pour $\varphi(t) = -\ln(t)$ comme générateur et de pseudo-inverse $\varphi^{-1}(t) = \exp(-t)$.

On génère la copule C par (2.11) :

$$C^A(u, v) = \exp(-[(-\ln(u)) + (-\ln(v))]) = uv = \Pi(u, v).$$

Π est une copule archimédienne stricte.

(b) Soit $\varphi(t) = 1-t$ pour $t \in [0, 1]$. Donc $\varphi^{-1}(t) = 1-t$ pour $t \in [0, 1]$ et 0 pour $t > 1$, alors $\varphi^{-1}(t) = \max(1-t, 0)$. D'après (2.11) on a :

$$C^A(u, v) = \max(u + v - 1, 0) = W(u, v)$$

donc la copule $W(u, v)$ est Archimédienne.

2.5 Famille des copules archimédienne

Les copules les plus utilisés de cette famille sont les copule de Gumbel, de Clayton et de Frank, définies dans ce qui suit .

2.5.1 Copule de Gumbel

En prenant comme générateur $\varphi(t) = (-\ln(t))^a$ 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 :

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

Les cas limites suivant la valeur du paramètre sont donc :

$$\lim_{a \rightarrow \infty} C_a = C^M.$$

$$\lim_{a \rightarrow 1} C_a = C^\pi.$$

2.5.2 Copule de Clayton

Pour $\varphi(t) = a \left(t^{-\frac{1}{a}} - 1 \right)$ avec $a > 0$.Nous obtenons ainsi la copule de Clayton

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

Les cas limites suivant la valeur du paramètre sont donc :

$$\lim_{a \rightarrow \infty} C_a = C^W.$$

$$\lim_{a \rightarrow 0} C_a = C^\Pi.$$

2.5.3 Copule de Frank

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

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

Et de densité

$$c^{Fr}(u, v) = \frac{(a-1) \ln a^{u+v}}{(a-1 + (a^u - 1)(a^v - 1))^2} .$$

Les cas limites sont donc les suivants :

$$\lim_{a \rightarrow -\infty} C_a = C^{W} .$$

$$\lim_{a \rightarrow 0} C_a = C^{\Pi} .$$

$$\lim_{a \rightarrow \infty} C_a = C^M .$$

2.6 Estimation des copules

L'estimation en statistique est une opération qui calcule la valeur inconnue d'un paramètre θ du modèle statistique (Ω, P) à partir d'un échantillon observé (X_1, \dots, X_n) , on prend comme exemple la loi normale de paramètres μ et σ^2 ($N(\mu, \sigma^2)$), telle que l'estimation de

μ est la moyenne $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et l'estimation de σ^2 est la variance $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Il existe beaucoup de méthodes d'estimation telle que la méthode du maximum de vraisemblance (méthode classique), la méthode des moments et procédure de Genest et Rivest, copule empirique .

On a expliqué les méthodes d'estimation des paramètres d'une copule à savoir les méthodes paramétriques et non paramétriques.

2.6.1 Estimation non paramétrique

On va expliquer les méthodes d'estimation des paramètres d'une copule avant on donne la notion de fonction de répartition empirique et copule empirique et procédure de Genest et Rivest.

Rappel : Fonction de répartition empirique

Définition 1 : Soit (X_1, \dots, X_n) un échantillon de taille n de loi F . La fonction de répartition empirique est donc définie par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(X_i \leq x)}$$

Cet estimateur conduit à un estimateur non paramétrique naturel d'une copule.

2.6.1.1 Copule empirique

La copule empirique a été introduite et d'abord étudié par Deheuvels (1979)[9], qui l'appelle fonction de dépendance empirique. Dans le cas bivarié, nous donnons une définition.

Définition 2 (Copule empirique) : Soit $\{(x_i^1, \dots, x_i^n)\}_{i=1}^T$ un échantillon d'un vecteur aléatoire X de dimension n , la fonction de répartition empirique est donnée par :

$$F_n(x_1, \dots, x_n) = \frac{1}{T} \sum_{i=1}^T \mathbb{I}_{(x_i^1 < x^1, \dots, x_i^n < x^n)}$$

Où \mathbb{I} est la fonction indicatrice.

La notion de copule empirique a été introduite pour la première fois par Deheuvels (1979), qui est connue sous le nom « Fonction de dépendance empirique ».

2.6.1.2 Cas des copules bivariées

Définition 1 :

Soit $\{(x_k, y_k)\}_{k=1}^n$ un échantillon de taille n d'un couple de variables aléatoires. La copule empirique est la fonction \hat{C} définie par :

$$\hat{C}\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{Nombre des paires } (x, y) \text{ dans l'échantillon tels que } x \leq x_{(i)} \text{ et } y \leq y_{(j)}}{n}$$

Où $x_{(i)}$ et $y_{(j)}$ Représentant les statistiques d'ordre associées à l'échantillon.

La fonction densité empirique de la copule C notée \hat{c} est donnée par :

$$\hat{c}\left(\frac{i}{n}, \frac{j}{n}\right) = \begin{cases} \frac{1}{n} & \text{si } (x_{(i)}, y_{(j)}) \text{ est un élément de l'échantillon;} \\ 0 & \text{sinon} \end{cases}$$

Elle est parfois appelée « empirical copula frequency » ; il existe une relation entre \hat{C} et \hat{c} donnée par :

$$\hat{C}\left(\frac{i}{n}, \frac{j}{n}\right) = \sum_{p=1}^i \sum_{q=1}^j \hat{c}\left(\frac{p}{n}, \frac{q}{n}\right)$$

Et

$$\hat{c}\left(\frac{i}{n}, \frac{j}{n}\right) = \hat{C}\left(\frac{i}{n}, \frac{j}{n}\right) - \hat{C}\left(\frac{i-1}{n}, \frac{j}{n}\right) - \hat{C}\left(\frac{i}{n}, \frac{j-1}{n}\right) + \hat{C}\left(\frac{i-1}{n}, \frac{j-1}{n}\right)$$

2.6.1.3 Procédure de GENEST et RIVEST :

La méthode suivante a été proposée par Christian Genest et Louis-Paul Rivest (1993), pour sélectionner une copule appartenant à la famille Archimédienne. Cette méthodologie se base directement sur le générateur de la copule.

Soit X un vecteur de n variables aléatoires, C est la copule associée au générateur φ et K la fonction définie par :

$$K(u) = P\{C(u, v) \leq u\}$$

Nous pouvons calculer pour une copule Archimédienne la quantité suivante

$$k(u) = u - \frac{\varphi(u)}{\varphi'(u)}$$

Un estimateur non paramétrique de K est alors donné par :

$$\hat{K}(u) = \frac{1}{T} \sum_{t=1}^T 1_{\vartheta_t \leq u}$$

Où

$$\vartheta_t = \frac{1}{T-1} \sum_{j=1}^T 1_{\{x_j^1 < x_t^1, x_j^2 < x_t^2\}}$$

Remarque : La fonction K est liée au tau de Kendall du fait que :

$$\tau = 4 \int_0^1 (1 - K(u)) du - 1$$

2.6.2 Estimation paramétrique

On se place ici dans le cas où la distribution conjointe dépend d'un paramètre, que l'on cherche à estimer.

Dans cette partie, on va expliquer les méthodes d'estimation telle que la méthode des moments et la méthode du maximum de vraisemblance.

2.6.1.1 Méthode des moments

Cette méthode est notamment utilisée pour les mesures de dépendance, l'estimateur des moments de la mesure de dépendance considérée est alors simplement obtenu en égalant l'expression paramétrique (analytique) de la mesure avec un estimateur non paramétrique de cette même mesure.

Exemple 6

Pour la copule de Gumbel de paramètre θ , on a

$$\tau = 1 - \frac{1}{\theta}$$

Nous en déduisons que :

$$\theta = \frac{1}{1 - \tau}$$

Si nous avons une estimation $\hat{\tau}$ du tau de Kendall, nous pouvons obtenir une estimation du paramètre de la copule en posant :

$$\hat{\theta} = \frac{1}{1 - \hat{\tau}}$$

Dans le cas général, on a l'estimation non paramétrique du tau de Kendall est donné par :

$$\hat{\tau} = \frac{c - d}{c + d}$$

Où c et d sont respectivement le nombre de paire disjointes concordantes et discordantes.

2.6.1.2 Méthode du maximum de vraisemblance (MLE)

Soit l'échantillon $\{x_1^t, \dots, x_n^t\}_{t=1}^T$ de fonction de répartition multivariée $H(x, \theta)$ qui dépend du vecteur des paramètres $\theta \in \Theta \subset \mathbb{R}^p$ le théorème de Sklar nous donne :

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i)$$

$$c(u_1, \dots, u_n) = \frac{\partial C(u_1, \dots, u_n)}{\partial u_1, \dots, \partial u_n}$$

Désigne la densité de la copule d-dimensionnelle $C(u_1, \dots, u_n, \theta)$ et f_i sont les densités des fonctions de répartition marginales F_i .

La fonction log-vraisemblance $L(x, \theta) = \sum_{t=1}^T \ln f(x_1^t, \dots, x_n^t)$ de l'échantillon $\{x_1^t, \dots, x_n^t\}_{t=1}^T$ est donnée par :

$$L(x, \theta) = \sum_{t=1}^T \ln c(F_1(x_1^t, \theta), \dots, F_n(x_n^t, \theta)) + \sum_{t=1}^T \ln f_1(x_1^t, \theta) + \dots + \sum_{t=1}^T \ln f_n(x_n^t, \theta)$$

L'estimateur de maximum de vraisemblance de est donnée par

$$\hat{\theta}_{MLE} = \arg \max L(x, \theta)$$

On peut montrer dans les conditions de régularité que l'estimateur existe qu'il est consistant, asymptotiquement efficace et vérifie la propriété de la normalité asymptotique. Nous avons :

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta_0) \rightarrow \mathcal{N}(0, \mathfrak{I}^{-1}(\theta_0))$$

Avec $\mathfrak{I}(\theta_0)$ est l'information usuelle de Fisher et θ_0 la vraie valeur.

Chapitre III

Copules des valeurs extrêmes

Dans le chapitre 1, nous avons présenté la théorie des valeurs extrêmes et dans le chapitre 2, nous avons étudié les copules bivariées, dans ce chapitre nous allons présenter les copules bivariées des valeurs extrêmes.

3. Copules des valeurs extrêmes bivariées

Pour parachever la théorie des valeurs extrêmes nous développons l'étude des copules des valeurs extrêmes bivariées en établissant un lien entre les copules et les distributions valeurs extrêmes. Comme pour la théorie des distributions nous donnons les grandes familles usuelles de copules bivariées.

3.1. Caractérisation d'une copule des valeurs extrêmes bivariées

Rappel : une fonction C est une copule des valeurs extrêmes si seulement si :

$$\forall t > 0, \forall u, v \in I, C^t \left(u^{\frac{1}{t}}, v^{\frac{1}{t}} \right) = C(u, v).$$

Le terme "*copule des valeurs extrêmes*" suggère l'existence d'un lien entre ces copules et une distribution des valeurs extrêmes. Considérons une distribution des valeurs extrêmes H de distributions marginales respectives F et G . Le résultat suivant établit que la copule associée à H par sa représentation canonique $H(x, y) = C(F(x), G(y))$ est une copule des valeurs extrêmes.

Théorème 1 : Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ une suite de copules aléatoires i.i.d. de distribution commune H alors il existe des suites de constantes réelles $a_n, c_n > 0$ et $b_n, d_n \in \mathbb{R}$ telles que :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{M_{X,n} - b_n}{a_n} \leq x, \frac{M_{Y,n} - d_n}{c_n} \leq y \right) = G(x, y)$$

où G est une distribution non dégénérée si et seulement si les marginales de G sont des lois de valeurs extrêmes univariées et si copule C qui lui est associée satisfait l'équation :

$$\forall t > 0, \forall u, v \in I, C^t \left(u^{\frac{1}{t}}, v^{\frac{1}{t}} \right) = C(u, v).$$

En utilisant la relation: $\forall u, v \in I, C(u, v) = H(F^{-1}(u), G^{-1}(v))$ on établit l'analogie du théorème de représentation de Pickands qui permet de caractériser toute copule des valeurs extrêmes bivariées au moyen de la fonction de dépendance d'une distribution des valeurs extrêmes dont elle est sous-jacente.

Théorème : Pour toute copule des valeurs extrêmes bivariées C , il existe une fonction

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

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

De plus A vérifie : $\max(t, 1-t) < A(t) < 1, \forall t \in [0,1]$.

Comme pour la distribution des valeurs extrêmes, A est dite générateur en ou fonction de dépendance de la copule.

Exemple1 : Retrouvons la fonction de dépendance associée à la copule de Galambos

$$C_{\theta}(u, v) = uv \exp \left\{ - \left[(-\log u)^{-\theta} + (-\log v)^{-\theta} \right]^{\frac{-1}{\theta}} \right\}, \theta \geq 1 \text{ on a :}$$

$$\exp \left[-(\log u + \log v) A \left(\frac{\log u}{\log u + \log v} \right) \right] = uv \exp \left\{ - \left[(-\log u)^{-\theta} + (-\log v)^{-\theta} \right]^{\frac{-1}{\theta}} \right\}$$

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

$$\Rightarrow (\log u + \log v) A \left(\frac{\log u}{\log u + \log v} \right) = (\log u + \log v) + \left[(-\log u)^{-\theta} + (-\log v)^{-\theta} \right]^{\frac{-1}{\theta}}$$

$$= (\log u + \log v) \left[1 + \left[\left(\frac{-\log u}{\log u + \log v} \right)^{-\theta} + \left(\frac{-\log v}{\log u + \log v} \right)^{-\theta} \right]^{\frac{-1}{\theta}} \right]$$

En posant $t = \frac{-\log u}{\log u + \log v}$ il vient que : $\frac{-\log v}{\log u + \log v} = 1 - t$ et par suite on retrouve la fonction

de dépendance $A_\theta(t) = 1 - \left[t^{-\theta} + (1-t)^{-\theta} \right]^{\frac{-1}{\theta}}$

3.2. Familles paramétriques usuelles des copules des valeurs extrêmes

Comme pour les distributions des valeurs extrêmes, il existe essentiellement deux grandes familles de modèles paramétriques usuels de copules des valeurs extrêmes bivariées :

- Le modèle mixte (Tawn 1988) et le modèle logistique (Gumbel 1960a).
- Les autres modèles proviennent généralement d'une extension symétrique ou asymétrique de ces deux modèles.

Nous donnons ici quelques unes d'entre elles en adoptant les notations :

$$\tilde{u} = -\ln u, \tilde{v} = -\ln v.$$

Famille, modèle	Copule $C_\theta(u, v)$	Générateur $A_\theta(t)$
Indépendante	$C(u, v) = \prod(u, v) = uv$	$A(t) = 1$
Gumbel1 ou logistique	$C_\theta(u, v) = \exp \left\{ -(\tilde{u}^\theta + \tilde{v}^{-\theta})^{\frac{1}{\theta}} \right\}$	$A_\theta(t) = \left[t^\theta + (1-t)^\theta \right]^{\frac{1}{\theta}}$
Gumbel2	$C_\theta(u, v) = uv \exp \left\{ \theta \frac{\tilde{u}\tilde{v}}{\tilde{u} + \tilde{v}} \right\}$	$A_\theta(t) = t^2 - \theta t + 1$
Galambos	$C_\theta(u, v) = uv \exp \left\{ -(\tilde{u}^{-\theta} + \tilde{v}^{-\theta})^{\frac{1}{\theta}} \right\}$	$A_\theta(t) = 1 - \left[t^{-\theta} + (1-t)^{-\theta} \right]^{\frac{-1}{\theta}}$
Husler-Reiss	$C_\theta(u, v) = \exp \left\{ -\tilde{v} \Phi \left[\frac{1}{\theta} + \frac{1}{2} \theta \log \left(\frac{\tilde{v}}{\tilde{u}} \right) \right] - \tilde{u} \Phi \left[\frac{1}{\theta} + \frac{1}{2} \theta \log \left(\frac{\tilde{u}}{\tilde{v}} \right) \right] \right\}$	$A_\theta(t) = t \Phi \left[\frac{1}{\theta} + \frac{1}{2} \theta \log \left(\frac{t}{1-t} \right) \right] + (1-t) \Phi \left[\frac{1}{\theta} - \frac{1}{2} \theta \log \left(\frac{t}{1-t} \right) \right]$

<p>Logistique de Joe</p>	$C_{\theta,\delta}(u, v) = \exp \left\{ - \left[\tilde{u}^\theta + \tilde{v}^\theta - (\tilde{u}^{-\theta\delta} + \tilde{v}^{-\theta\delta})^{\frac{-1}{\delta}} \right]^{\frac{-1}{\theta}} \right\}, \delta > 0, \theta \geq 1$ $A_{\theta,\delta}(t) = \left[t^\theta + (1-t)^\theta - (t^{-\theta\delta} + (1-t)^{-\theta\delta})^{\frac{1}{\delta}} \right]^{\frac{1}{\theta}}, t > 0$
<p>Galambos</p>	$C_{\theta,\delta}(u, v) = \exp \left\{ - \left[\tilde{u}^\theta + \tilde{v}^\theta - (\tilde{u}^{-\theta\delta} + \tilde{v}^{-\theta\delta})^{\frac{-1}{\delta}} \right]^{\frac{1}{\theta}} \right\}, \delta > 0, \theta \geq 1$ $A_{\theta,\delta}(t) = 1 - \left[t^{-\theta} + (1-t)^{-\theta} - (t^{\theta\delta} + (1-t)^{\theta\delta})^{\frac{-1}{\delta}} \right]^{\frac{-1}{\theta}}$
<p>Marshal-Olkin</p>	$C_{\alpha,\beta}(u, v) = u^{1-\alpha} v^{1-\beta} \min(u^\alpha, v^\beta) = \begin{cases} uv^{1-\beta} & \text{si } u^\alpha < v^\beta \\ u^{1-\alpha}v & \text{si } u^\alpha > v^\beta \end{cases}$ $A_{\alpha,\beta}(t) = \max\{1 - \alpha t, 1 - \beta(1 - t)\}, \alpha \leq 1, \beta \geq 0$
<p>Tawn</p>	$C_{\theta,\delta,\lambda}(u, v) = uv \exp \left\{ - (1-\delta) + (\theta - \delta)u + \left[(\theta u)^\lambda + (\delta v)^\lambda \right]^{\frac{1}{\lambda}} \right\}$ $A_{\theta,\delta,\lambda}(t) = (1-\delta) + (\delta - \theta)t + [(\theta t)^\lambda + \delta(1-t)^\lambda]^{\frac{1}{\lambda}}$

Propriétés

La distribution de (X_1, Y_1) étant H et celle de (X_2, Y_2) étant Π (car les variables X_2 et Y_2 sont indépendantes) alors d'après ce qui précède on a le théorème suivant.

Théorème

Soit (X, Y) un couple de variables continues de copule, la version population du rho de Spearman pour X et Y est donnée par :

$$\rho_{XY} = \rho_C = 3Q(C, \Pi) = 12 \iint_{I^2} uv dC(u, v) - 3 = 12 \iint_{I^2} C(u, v) dudv - 3$$

Exemple 2

Pour la copule de Marshall-Olkin :

$$C_{\alpha, \beta}(u, v) = \begin{cases} u^{1-\alpha} v & \text{si } u^\alpha \geq v^\beta \\ uv^{1-\beta} & \text{si } u^\alpha \leq v^\beta \end{cases}$$

On établit que :

$$\rho_{\alpha, \beta} = \frac{1}{4} \left(\frac{\alpha\beta}{2\alpha - \alpha\beta + 2\beta} \right)$$

Chapitre IV

Application aux météorologies

Introduction

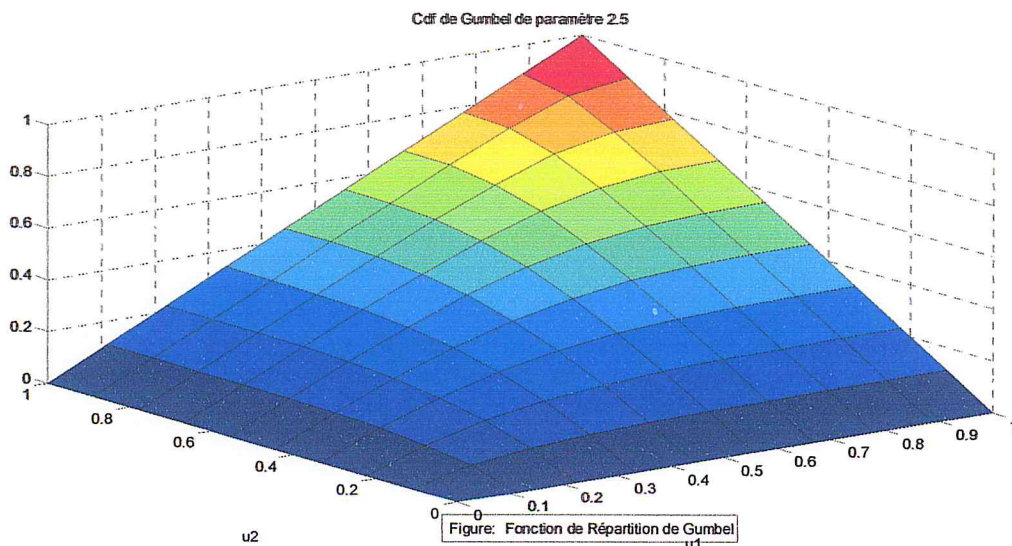
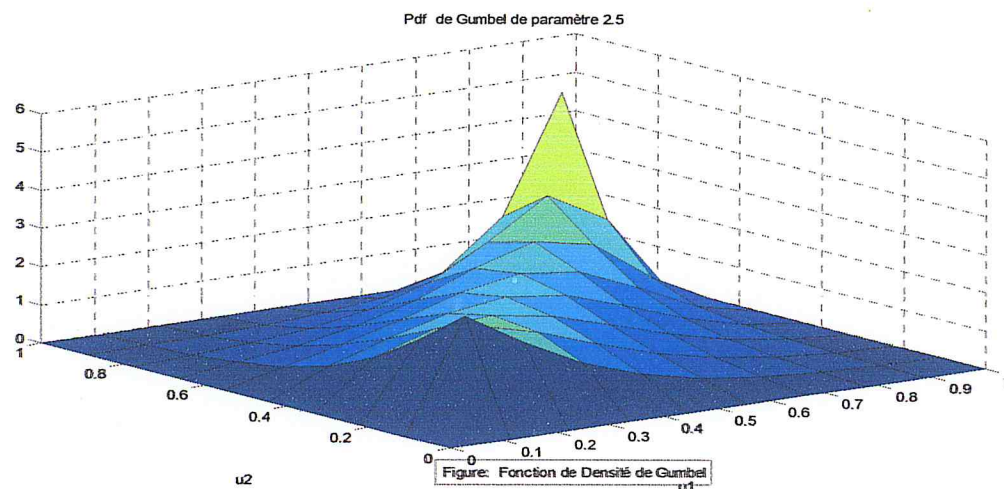
Nous clôturons, notre travail par deux études, la première consiste à comparer entre des méthodes d'estimation des paramètres des copules extrêmes et la seconde est l'ajustement des données météorologiques par des copules extrêmes et précisément la copule extrême de Gumbel.

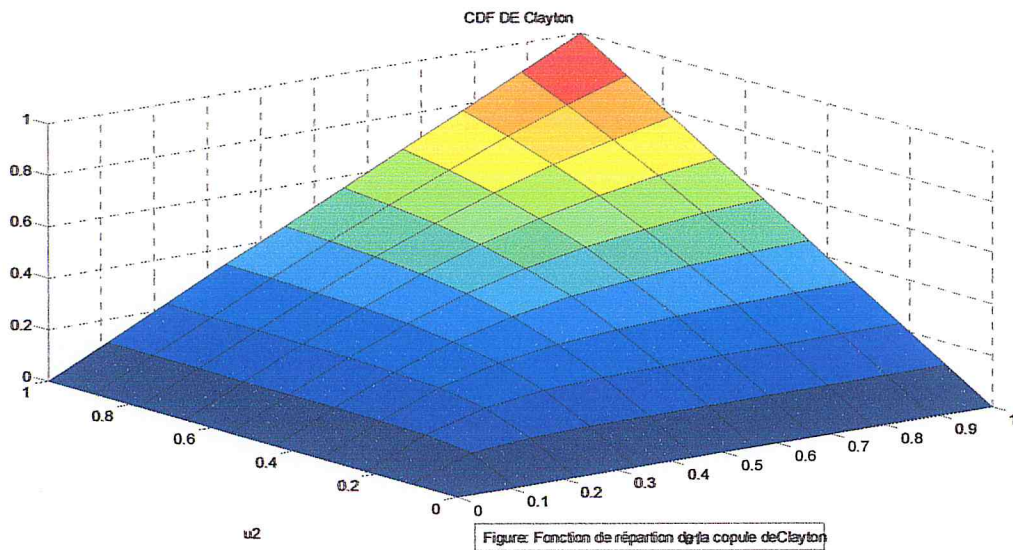
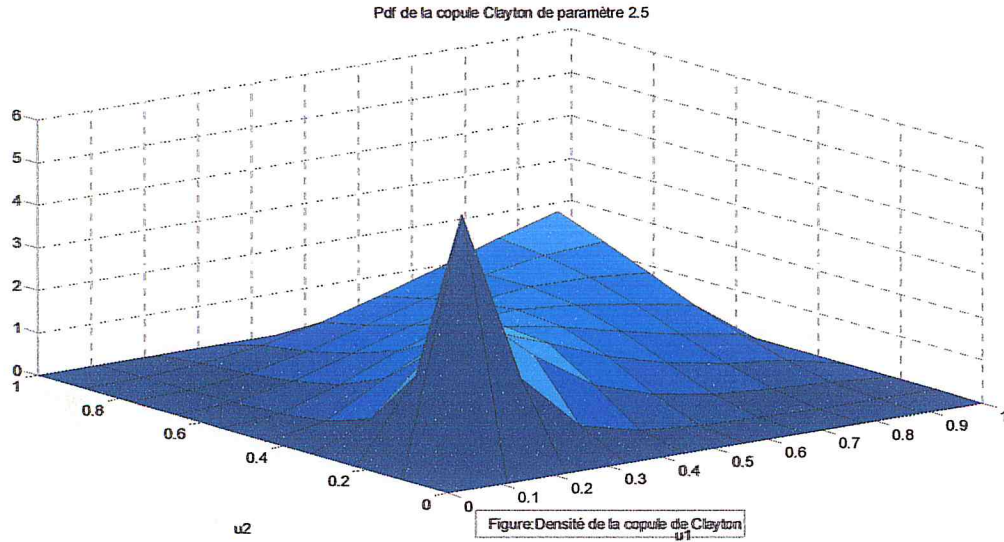
I. Simulation :

1-Copule de Gumbel :

Nous avons simulé 100 échantillon de taille 100.

Les graphes représentent la densité et la fonction de répartition d'une copule de Gumbel et de Clayton simulée de paramètre $\theta = 2.5$





Le tableau ci-dessous, nous avons calculé quelques valeurs à l'aide de la copule de Gumbel et Clayton et Clayton à savoir probabilité $C(u, v)$ en (u, v) , le quantile en q et la probabilité conditionnelle $C(U < 0.3/V = 0.5)$

Gumbel		
Probabilité	Quantile	Probabilité conditionnelle
$C(0.3, 0.5)$	$C(q, 0.5)$	$c(U < 0.3/V = 0.5)$
0.2906142	0.3000175	0.1025705
Clayton		
$c(0.3, 0.5)$	$c(q, 0.5)$	$c(U < 0.3/V = 0.5)$

2. Comparaison des méthodes d'estimation des paramètres des copules

Nous avons simulé 100 échantillons de taille 100 de la copule de Gumbel de paramètre $\theta = 2.5$ et nous avons estimé le paramètre θ par les méthodes du maximum de vraisemblance, la méthode des Moindres carrées et la méthode du coefficient de Kendall.

Tableau 2: Estimation du paramètre de la copule de Gumbel

	Maximum de vraisemblance(ML)	Moindre carrée(LS)	Coefficient de Kendall
$\bar{\theta}$	2.078463	2.928536	2.002582
σ_{θ}	0.1887564	2.005658	0.1873831
UIS	2.130783	3.484476	2.054522
UIF	2.026142	2.372596	1.950642

II. Application

4.1 Introduction

L'office national de la météorologie (ONM) crée par ordonnance N° 75-25 du 29 avril 1975, Les données météorologiques utilisées dans notre étude est de l'ONM de la zone de Dellys de Boumerdes qui représentent les mesures mensuelles de l'ÉVApourisation, le temps d'insolation (ensoleillement), la température, l'humidité, la précipitation et la vitesse de vent.

4.2 Description de l'échantillon

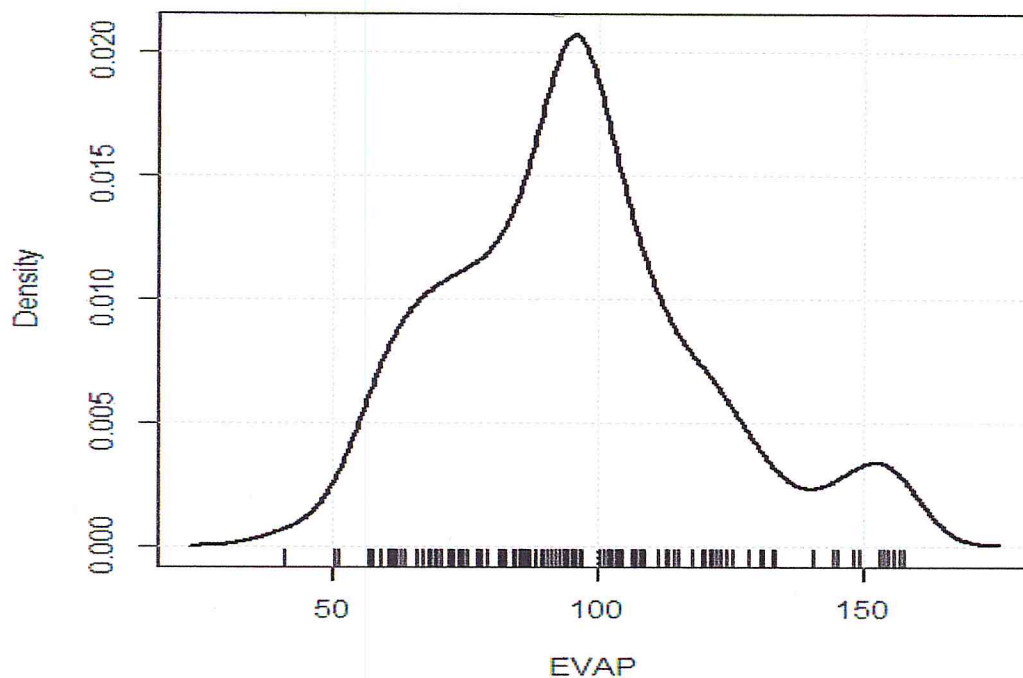
Le tableau ci-dessous donne une analyse préliminaire de la statistique descriptive des paramètres météorologiques.

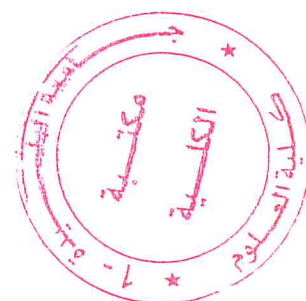
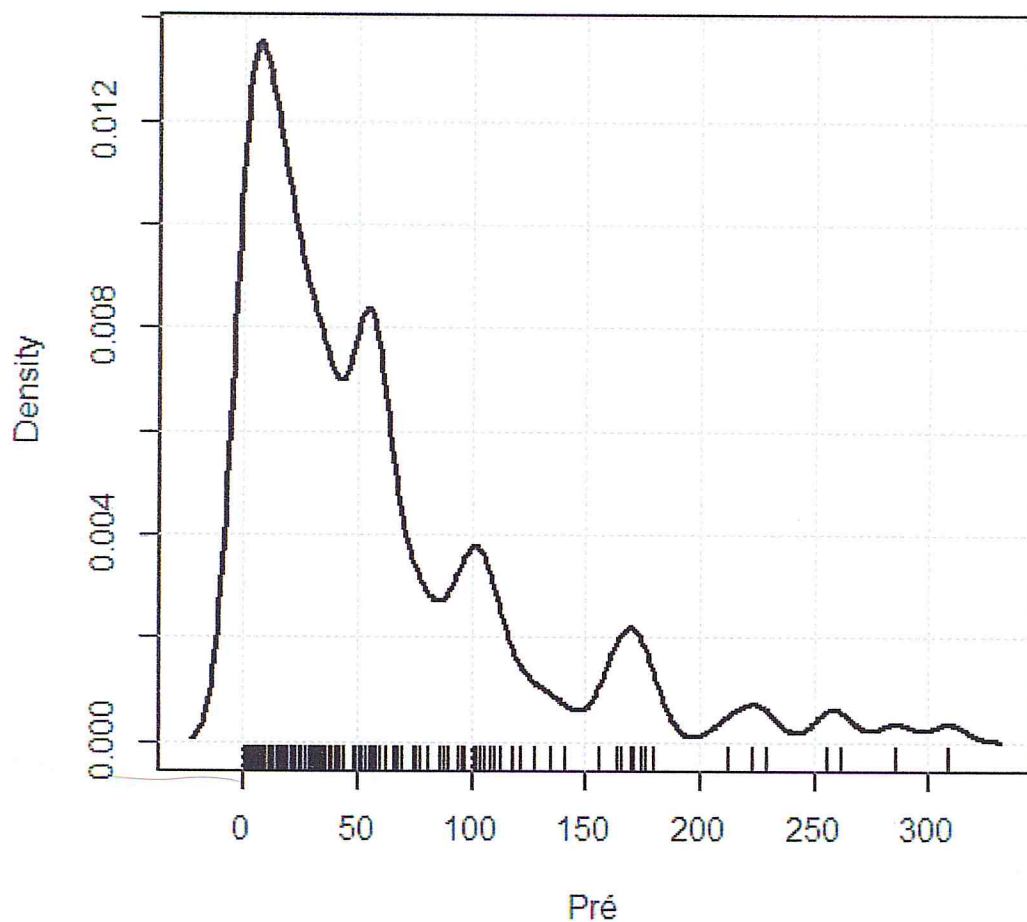
4.1 tableau represente la statistique descriptive

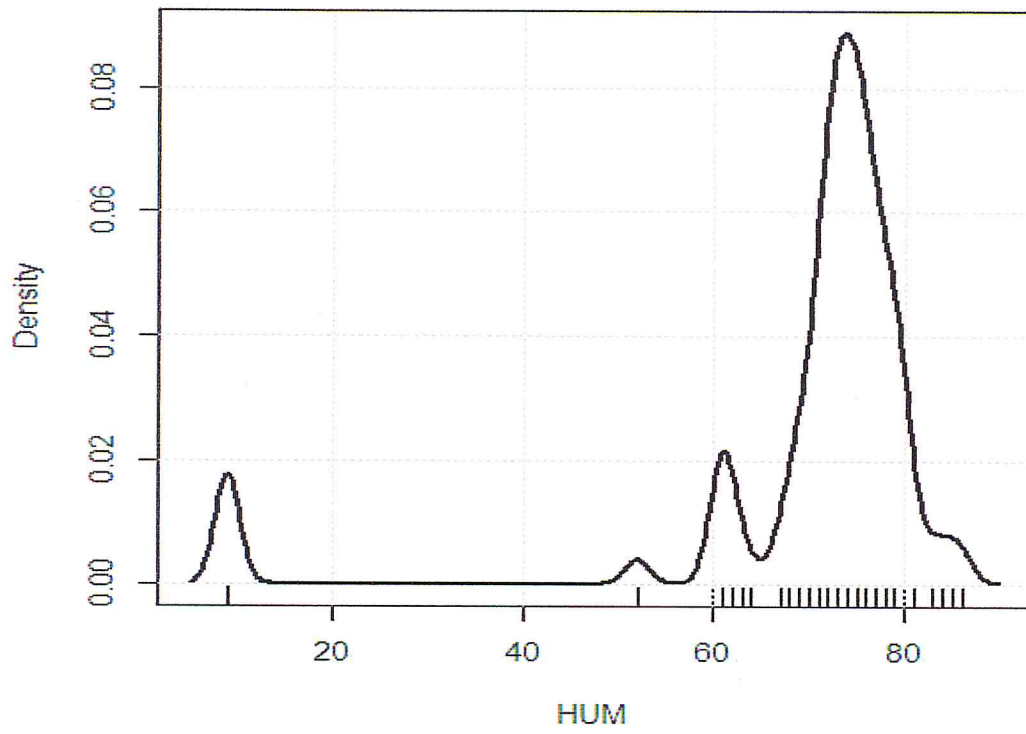
N=156	Moy	Var	Sd	Vr min	Vr max	kurtosis	skewness
EVAP	95.522436	593.792907	24.3678677	41.20	157.50	0.0581861 1	0.4884272
INS	69.602564	3485.36503	16.0767030	120.2	354.8	0.0517489	-0.3178398
TEM	230.867308	18.7634876	59.0369815	10.10	27.60	-1.004119	0.2413399
HUM	57.775641	258.46038	62.4956106	9.0	86.0	8.562229	-3.013754
Pré	18.222436	3905.70134	4.3316923	0.00	308.70	3.046117	1.733056
vent	3.098718	0.64038544	0.8002409	1.600	5.900	0.9689927	0.7938512

1. Graphes des densités de probabilités :

Les graphes sous dessus représentent les densités de probabilités des paramètres météorologiques.







1. Ajustement des données météorologiques univariées:

Nous avons ajusté la précipitation, l'Humidité et la vitesse du vent par la loi des valeurs extrêmes à savoir la loi de Gumbel. Nous remarquons que que la loi de Gumbel ajuste bien ces données météorologiques. Nous confirmé cet ajustement par le test de Kolmogorov-Smirnov.

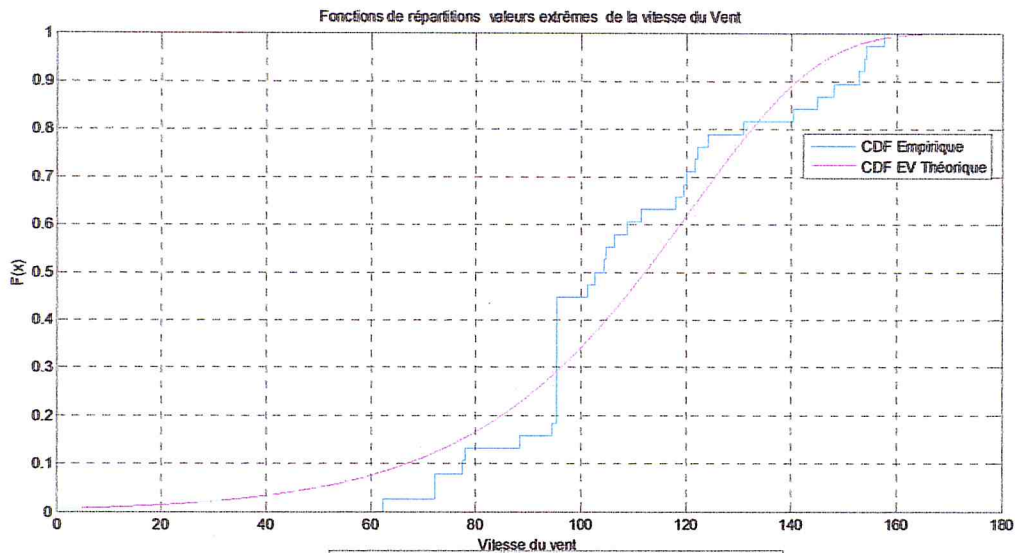


Figure: Ajustement de la vitesse du Vent par loi des valeurs extrêmes

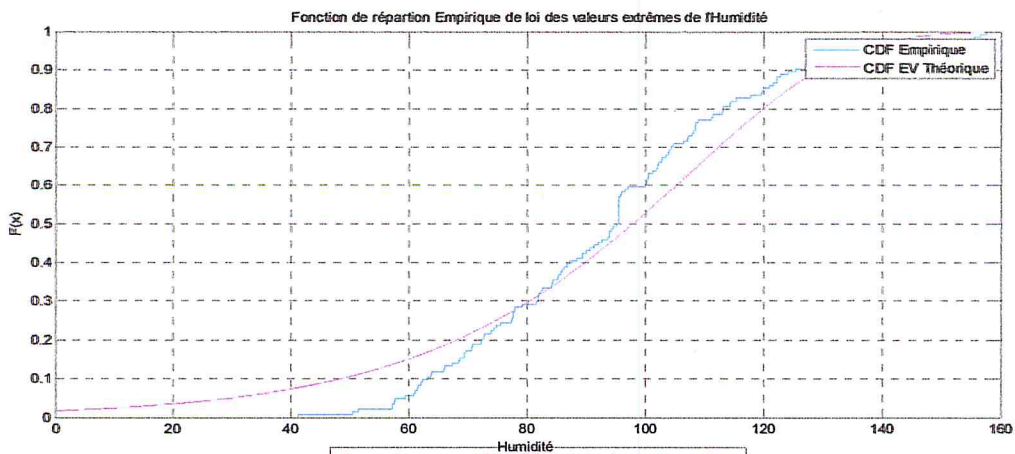
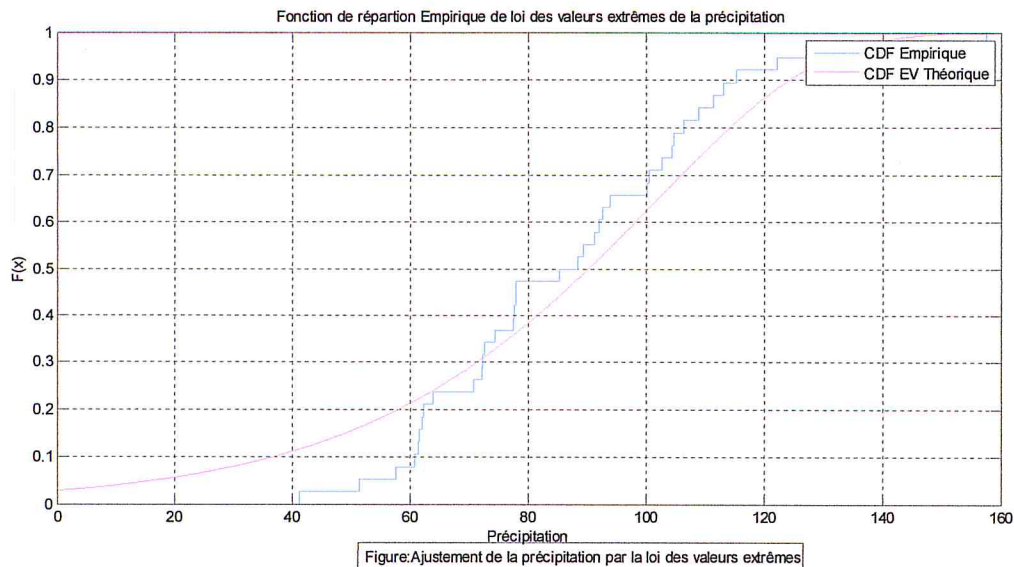


Figure: Ajustement de l'Humidité par la loi des valeurs extrêmes



3. Estimation des paramètres de la copule de Gumbel :

Nous avons ajusté les données météorologiques (Précipitation, Humidité, Vitesse du vent) par la copule de Gumbel. Les résultats sont dans le tableau suivant :

Tableau 4.2 : Paramètres de la copule de Gumbel

P : Précipitation , Vitesse du vent, H :Humidité

	(H,P)	(H,V)	(P,V)
θ	1.013	1	0.886
Taux de Kendal	1	1	1
Coefficient de corrélation	0.00000135	0.00000135	0.013

4. Probabilité de dépassement (de survie) :

Nous calculé quelques probabilités de survie données dans le tableau ci-dessous

Tableau 4.3 : Probabilité de dépassement

$P(X>a, Y>b)$	
Proba($P>80, H>80$)	0.0664
Proba($P>80, V>30$)	0.3201
Proba($H>80, V>30$)	0.0895

Conclusion générale

Conclusion

Les copules des valeurs extrêmes constituent un outil statistique permettant de modéliser la dépendance entre des variables aléatoires. La fonction copule relie en émet la densité jointe aux densités marginales et contient ainsi toute l'information sur la structure de dépendance du modèle.

Nous avons montré par simulation que les méthodes d'estimation du paramètre de la copule des valeurs maximales de Gumbel à savoir la méthode du maximum du vraisemblance (ML), la méthode des moindres carrées (LS) et la méthode du coefficient de Kendall estiment bien les paramètres de la copule de Gumbel et sont presque similaire.

Nous avons vérifié par le biais du test de Kolmogorov- Smirnov et l'ajustement graphique que les données de Précipitations , Humidité et Vitesse de vent sont bien ajustés par la loi de Gumbel.

Nous avons aussi ajusté les couples (Précipitation, Humidité), (Précipitation, vitesse du vent) et (Humidité, Vitesse de vent) par la copule de Gumbel.

Nous avons estimé les paramètres de ces copules et le coefficient de corrélation et le tau de Kendall ainsi que les probabilités de dépassement (de survie).

Parmi les perspectives de ce travail est le calcul de la probabilité des risques extrêmes multiple qui permet d'estimer le quantile extrême multiple (Valeur a risque VaR) et la probabilité de ruine multibranches des sociétés d'assurance.

Annexe

H_γ est appelée fonction de répartition de la loi des valeurs extrêmes.

γ appelé indice des valeurs extrêmes ou indice de queue.

γ est le paramètre clé de toute la théorie des valeurs extrêmes.

μ position.

σ dispersion .

F fonction de répartition.

N_μ est le nombre de observations au-dessus du seuil μ .

$\hat{\gamma}$ et $\hat{\sigma}$ sont les estimateurs des paramètres de la loi GPD.

X, Y variables aléatoires.

U, V variables aléatoires uniformes.

C copule.

c densité de la copule.

$C_{X,Y}$ copule bivariée

M copule min

W copule max

Π copule produit

C_ρ^\emptyset copule gaussienne

$C_{\rho,v}^t$ copule student

C^A copule Archimédienne

C_α^{Gum} copule de Gumbel

C_{α}^{Fra} copule de Frank

C_{α}^{Cla} copule de Clayton

C^* copule de valeur extrêmes

\tilde{C} copule empirique

F_1, \dots, F_d fonction de répartition univariées

F^{-1} inverse généralisé de F

\emptyset fonction de la loi normale standard

\emptyset_{ρ} fonction de la loi normale multivariée

$L(\theta)$ fonction de maximum de vraisemblance

λ_L dépendance de queue inférieur

λ_U dépendance de queue supérieur

τ tau de Kendall

ρ rho de Spearman

H fonction de répartition jointe

Bibliographie :

- [1] Abdous, B ,K.Ghoudi et A .Khoudraji, « Nom-parametric estimation of the limit dependence function of multivariate extremes », *Extremes*,3,245-268,(1998).
- [2] A.Guillon,P.Willems. Thèse, Application de la théorie des valeurs extrêmes en Hydrologie ,2006.
- [3] Alexandre Lekina .Estimation non- paramétrique des quantiles extrêmes conditionnels .Thèse, 2010.
- [4] Beirlant,J.,Goegebeur,Y.,Segers,J.,Teugels,J. « Statistics of extremes : Theory and Applications, Wiley Series in Probabilty and Statistics » .JohnWiley and Snons Ltd., Chichester (2004).
- [5] Gnedenko. B, Sur la distribution limite du terme maximum d'une série aléatoire. Cours.
- [6] J Beirlant, P Vynckier, and J.M Teugels. Excess functions and estimation of the extrême value index. *Bernoulli*, 2 :293–318, 1996.
- [7] Jürgen ANG ST, Eléments de la théorie des probabilités et de statistique .Notes de cours ,2014/2015.
- [8] Slimani. Mémoire de fin d'étude en Master Intitulé : Estimation des paramètres d'une copule. Université Mohamed Khider Biskra 2011-2012.

