

Théorie des valeurs extrêmes

Laurent Gardes

Université de Strasbourg

Chapitre 1

Introduction

Alors que la statistique "classique" s'intéresse principalement à la partie dite centrale de la loi modélisant au mieux le phénomène considéré (calcul de l'espérance, la médiane, la variance, utilisation du théorème central limite, etc.), nous souhaitons ici étudier les "grandes" valeurs c'est-à-dire la queue de distribution de la loi. La théorie des valeurs extrêmes propose un cadre théorique solide pour l'étude de ces valeurs dites extrêmes. La difficulté principale réside dans l'application concrète de la théorie (estimation, inférence, etc.) puisque par nature, un évènement extrême est très peu observé. On parle d'évènement rare.

Enfin, la théorie des valeurs extrêmes est évidemment utile dans d'autres champs d'applications comme en finance, pour l'étude des "log-return" qui sont définis par $r_t := \log(p_t/p_{t-1})$ où p_t est la cotation en bourse à la date t d'une action donnée. Elle est également utilisée pour estimer (si elle existe) la durée maximale pour la vie d'un être humain.

Avant de présenter en détail cette théorie des valeurs extrêmes, le paragraphe suivant est consacré à la notion de quantile.

1.1 Introduction à la notion de quantile

Nous commençons par donner la définition de l'inverse généralisée d'une fonction croissante et continue à droite.

Définition 1 Soit Φ une fonction croissante et continue à droite sur \mathbb{R} . L'inverse généralisée de Φ est définie par :

$$\Phi^{\leftarrow}(y) := \inf\{x \mid \Phi(x) \geq y\},$$

avec la convention $\inf\{\emptyset\} = +\infty$.

L'inverse généralisée vérifie les propriétés utiles suivantes :

Proposition 1 Soit Φ une fonction croissante et continue à droite. L'inverse généralisée Φ^{\leftarrow} est une fonction croissante et continue à gauche. On a de plus les propriétés ci-dessous :

- i) $\Phi(\Phi^{\leftarrow}(y)) \geq y$,
- ii) $\Phi^{\leftarrow}(y) \leq x \Leftrightarrow y \leq \Phi(x)$,
- iii) $x < \Phi^{\leftarrow}(y) \Leftrightarrow y > \Phi(x)$.

Cette Proposition nous permet de démontrer facilement un résultat très utile tant d'un point de vu théorique que pratique pour simuler des réalisations de variables aléatoires.

Lemme 1 (Lemme de transformation des quantiles) Soit $U : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto ([0, 1], \mathcal{B}[0, 1])$ une variable aléatoire de loi uniforme sur $[0, 1]$. Soit $F : \mathbb{R} \mapsto [0, 1]$ une fonction de répartition. Alors la variable aléatoire $X := F^{\leftarrow} \circ U : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une variable aléatoire de fonction de répartition F .

Nous pouvons à présent donner la définition d'un quantile.

Définition 2 Soit F une fonction de répartition. Le quantile d'ordre α de la fonction de répartition F est défini pour $\alpha \in [0, 1]$ par

$$F^{\leftarrow}(\alpha) = \inf\{x | F(x) \geq \alpha\}.$$

En théorie des valeurs extrêmes, on s'intéressera donc à des quantiles dont l'ordre α est proche de 1. Comme l'on préfère manipuler des quantités qui sont proches de 0 (plutôt que de 1), on appellera par abus de langage quantile d'ordre α la quantité :

$$q(\alpha) = F^{\leftarrow}(1 - \alpha) = \inf\{x | \bar{F}(x) \leq \alpha\},$$

où $\bar{F} = 1 - F$ est appelée fonction de survie. Pour estimer $q(\alpha)$, l'idée la plus naturelle (sans faire d'hypothèse supplémentaire) est d'estimer F par la fonction de répartition empirique définie par :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\}.$$

On montre alors facilement qu'un estimateur de $q(\alpha)$ est donné par :

$$\hat{F}_n^{\leftarrow}(1 - \alpha) = \sum_{i=1}^n X_{n-i+1, n} \mathbb{I}\left\{\alpha \in \left[\frac{i-1}{n}, \frac{i}{n}\right]\right\} = X_{n - \lfloor n\alpha \rfloor, n},$$

où $\lfloor x \rfloor$ est le plus grand entier inférieur ou égal à x et $X_{1, n} \leq \dots \leq X_{n, n}$ est l'échantillon ordonné (la variable aléatoire $X_{i, n}$, $i \in \{1, \dots, n\}$ est une statistique d'ordre).

On remarque que l'utilisation de la fonction de répartition empirique n'est pas conseillée lorsque l'on s'intéresse à l'estimation de quantile d'ordre α proche de 0. En effet, dès que $\alpha < 1/n$, $\hat{F}_n^{\leftarrow}(1 - \alpha) = X_{n, n}$ et donc l'estimateur du

quantile est dans ce cas borné au maximum de l'échantillon. La zone $[X_{n,n}, \infty)$ est appelée la queue de distribution de F et un quantile appartenant à la queue de distribution est appelé un quantile extrême. Nous classons ci-dessous plus précisément les différents types de quantiles en fonction de leur ordre α_n (qui peut donc dépendre de la taille de l'échantillon).

Définition 3 On classe les quantiles en trois catégories selon leur ordre α_n :

- i) On dira d'un quantile qu'il est **classique** si $n\alpha_n \rightarrow \infty$.
- ii) On dira d'un quantile qu'il est **intermédiaire** si $n\alpha_n \rightarrow c \in [1, \infty[$.
- iii) On dira d'un quantile qu'il est **extrême** si $n\alpha_n \rightarrow c \in [0, 1[$.

Noter que les quantiles "classiques" et "intermédiaires" peuvent être estimés à l'aide de la fonction de survie empirique.

1.2 Quantile et assurance

La notion de quantile est très largement utilisée en actuariat pour quantifier le risque d'un portefeuille d'assurances. La mesure de risque la plus connue est la "Value at Risk (VaR)" définie ci-dessous.

Définition 4 Soit X une variable aléatoire réelle de fonction de répartition F et soit $\alpha \in [0, 1]$. La Value at Risk de X au niveau α est donnée par :

$$VaR_\alpha(X) := F^{\leftarrow}(1 - \alpha).$$

La Value at Risk de niveau α n'est donc rien d'autre que le quantile d'ordre α que l'on avait noté $q(\alpha)$ dans le paragraphe précédent. Généralement, X est une variable aléatoire modélisant un risque pour la compagnie d'assurance (par exemple, le niveau d'eau d'une rivière, etc.) et la Value at Risk représente donc un seuil du risque ayant une probabilité α d'être dépassé. Cependant, bien que très utilisée, la Value at Risk ne donne aucune information sur ce qui se passe lorsque ce seuil est dépassé. On peut par exemple avoir deux variables aléatoires X et Y ayant une même valeur de la Value at Risk au niveau α mais présentant des queues de distribution très différentes. Pour remédier à ce problème, une autre mesure de risque a été proposée : la "Tail Value at Risk (TVaR)".

Définition 5 Soit X une variable aléatoire réelle de fonction de répartition F et soit $\alpha \in [0, 1]$. La Tail Value at Risk de X au niveau α est donnée par :

$$TVaR_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_s(X) ds, \quad (1.1)$$

si cette intégrale existe.

La Tail Value at Risk de niveau α est donc la moyenne des Value at Risk de niveaux inférieurs à α . Ceci permet d'avoir de l'information sur la queue de distribution au delà du niveau α . Le résultat ci-dessous donne une autre écriture de la Tail Value at Risk dans le cas d'une fonction de répartition dérivable presque-partout.

Proposition 2 *Soit X une variable aléatoire intégrable, de fonction de répartition F continue et dérivable presque-partout. On a :*

$$TVaR_\alpha(X) = \frac{1}{\alpha} \mathbb{E}(X \mathbb{I}\{X \geq VaR_\alpha(X)\}) = \mathbb{E}(X|X \geq VaR_\alpha(X)).$$

Ce résultat a motivé la définition d'une autre mesure de risque : la "Conditional Tail Expectation" définie pour une variable aléatoire X quelconque par

$$CTE_\alpha(X) := \mathbb{E}(X|X \geq VaR_\alpha(X)).$$

Cette mesure donne donc la valeur moyenne du risque X sachant que l'on a dépassé le seuil $VaR_\alpha(X)$. La Proposition 2 assure que pour une variable aléatoire admettant une densité par rapport à la mesure de Lebesgue, la Conditional Tail Expectation coïncide avec la Tail Value at Risk.

En conclusion, les mesures de risque en assurance sont intimement liées à la notion de quantile. Lorsque le niveau de risque α est proche de zéro, l'estimation des mesures de risque passe donc par la théorie des valeurs extrêmes.

Chapitre 2

Théorie des valeurs extrêmes

2.1 Convergence en loi du maximum d'un échantillon

Soient X_1, \dots, X_n un échantillon de n variables aléatoires indépendantes et identiquement distribuées de fonction de répartition commune F . En statistique "classique", on s'intéresse essentiellement à la variable aléatoire $S_n := X_1 + \dots + X_n$. La loi faible des grands nombres assure que si X_1 est intégrable alors $S_n/n \xrightarrow{P} \mathbb{E}(X_1)$. De plus, si X_1 est de carré intégrable, le Théorème de la Limite Centrale assure que

$$\left(\frac{n}{\text{Var}(X_1)} \right)^{1/2} \left[\frac{S_n}{n} - \mathbb{E}(X_1) \right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

La somme S_n permet donc d'obtenir des résultats sur la partie centrale de la distribution. Dans notre cas, on s'intéresse à la queue de distribution et c'est donc tout naturellement que l'on souhaite étudier le comportement du maximum de l'échantillon. Ce maximum est noté dans la suite $X_{n,n} = \max(X_1, \dots, X_n)$. On rappelle que la fonction de répartition de $X_{n,n}$ est donnée par :

$$\mathbb{P}(X_{n,n} \leq x) = F^n(x).$$

Si on note $x_F := q(0) = \inf\{x; F(x) = 1\}$, on montre que $X_{n,n} \xrightarrow{P} x_F$ lorsque $n \rightarrow \infty$. Pour obtenir une convergence en loi, il faut donc, à l'instar du TCL, normaliser le maximum.

Définition 6 Soit H une fonction de répartition non-dégénérée (c'est-à-dire une fonction de répartition qui n'est pas associée à une variable constante presque sûrement). On dit que F **appartient au domaine d'attraction de H** (et on note $F \in \mathcal{DA}(H)$) s'il existe deux suites $a_n > 0$ et b_n telles que en tout point de continuité x de H ,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x) \Leftrightarrow \frac{X_{n,n} - b_n}{a_n} \xrightarrow{\mathcal{L}} Y,$$

où la variable aléatoire Y admet H pour fonction de répartition.

La plupart des fonctions de répartition continues appartient à un domaine d'attraction. Il est à noter cependant que la loi de Poisson n'appartient à aucun domaine d'attraction bien que cette loi soit parfois appelé "loi des événements rares". Nous nous intéressons à présent à la question de l'unicité du domaine d'attraction. Commençons par définir la notion de fonctions de répartition de même type.

Définition 7 Soient H et G deux fonctions de répartition non-dégénérées. On dit que H et G sont de même type s'il existe $a > 0$ et b tels que pour tout $x \in \mathbb{R}$, $H(x) = G(ax + b)$.

En d'autres termes, si $X \sim H$ et $Y \sim G$ alors, si H et G sont de même type, $Y = aX + b$ presque-sûrement.

Proposition 3 Soient H et G deux fonctions de répartition non-dégénérées. Si $F \in \mathcal{DA}(H)$ (c'est-à-dire s'il existe deux suites $a_n > 0$ et b_n telles que $F^n(a_n x + b_n) \rightarrow H(x)$) et si $F \in \mathcal{DA}(G)$ (c'est-à-dire s'il existe deux suites $u_n > 0$ et v_n telles que $F^n(u_n x + v_n) \rightarrow G(x)$), alors, les fonctions de répartition H et G sont de même type ($G(x) = H(ax + b)$). De plus, $u_n/a_n \rightarrow a > 0$ et $(v_n - b_n)/a_n \rightarrow b$.

Cette proposition montre que le domaine d'attraction d'une fonction de répartition F est unique à un changement des paramètres d'échelle et de position près.

Nous donnons à présent le résultat fondamental de la théorie des valeurs extrêmes. Ce résultat est à rapprocher du théorème de la limite centrale.

Théorème 1 Si $F \in \mathcal{DA}(G)$ où G est non dégénérée, alors G est de même type que la fonction de répartition définie pour tout x tel que $1 + \gamma x > 0$ par

$$H_\gamma(x) = \begin{cases} \exp[-(1 + \gamma x)^{-1/\gamma}] & \text{si } \gamma \neq 0, \\ \exp(-e^{-x}) & \text{si } \gamma = 0. \end{cases}$$

La loi limite du maximum dépend donc du seul paramètre γ appelé l'indice des valeurs extrêmes. Selon le signe de γ , on définit trois types de domaines d'attraction : domaine d'attraction de Fréchet lorsque $\gamma > 0$; de Weibull lorsque $\gamma < 0$ et de Gumbel lorsque $\gamma = 0$. Le tableau suivant regroupe quelques lois usuelles classées en fonction de leur domaine d'attraction.

Gumbel	Fréchet	Weibull
Normale	Pareto	Uniforme
Exponentielle	Log-gamma	Beta
Log-normale	Student	
Gamma		
Weibull		

On rappelle que la loi de Weibull de paramètre $\beta > 0$ admet pour fonction de répartition $F(x) = 1 - \exp(-x^\beta)$ pour $x > 0$. La fonction de répartition de la loi de Pareto de paramètres $\gamma > 0$ et $a > 0$ est $F(x) = 1 - (x/a)^{-1/\gamma}$ pour $x > a$.

Le paramètre γ est une paramètre de forme. Il contrôle la forme de la queue de distribution. Nous verrons un peu plus loin que

- 1) Si $\gamma > 0$, la fonction de survie $1 - F(x)$ converge vers 0 lorsque $x \rightarrow x_F^* = +\infty$ à une vitesse polynomiale (proportionnelle à une puissance de x). Ce domaine d'attraction regroupe les distributions *heavy-tailed* (à queue "lourde").
- 2) Si $\gamma = 0$, la fonction de survie $1 - F(x)$ converge vers 0 lorsque $x \rightarrow x_F^* = +\infty$ à une vitesse exponentielle. On parle ici de *light-tailed distributions* (à queue légère)
- 3) Si $\gamma < 0$, la fonction de survie $1 - F(x)$ converge vers 0 lorsque $x \rightarrow x_F^* < +\infty$ à une vitesse polynomiale (proportionnelle à une puissance de x). Ce domaine d'attraction regroupe la plupart des distributions ayant un point terminal fini.

Il est donc intuitivement clair que si deux fonctions de survie sont asymptotiquement proportionnelles, elles appartiennent au même domaine d'attraction avec le même indice des valeurs extrêmes.

Définition 8 Soient F et G deux fonctions de répartition. On dit que F et G ont des queues de distribution proportionnelles si elles ont le même point terminal x^* et si

$$\lim_{x \rightarrow x^*} \frac{\bar{F}(x)}{\bar{G}(x)} = c \in]0, \infty[.$$

Proposition 4 Si F et G ont des queues de distribution proportionnelles alors si $F \in \mathcal{DA}(H_\gamma)$, il en est de même pour G .

2.2 Caractérisation des domaines d'attraction

2.2.1 Quelques résultats sur les fonctions à variations régulières

Définition 9 Une fonction mesurable $U : \mathbb{R} \mapsto [0, \infty[$ est dite à variations régulières d'indice $\alpha \in \mathbb{R}$ à l'infini si pour tout $t > 0$,

$$\lim_{x \rightarrow \infty} \frac{U(tx)}{U(x)} = t^\alpha.$$

On notera dans la suite \mathcal{RV}_α l'ensemble des fonctions à variations régulières d'indice α . Une fonction à variations régulières d'indice $\alpha \in \mathbb{R}$ se comporte asymptotiquement comme la fonction $x \mapsto x^\alpha$.

Définition 10 Si une fonction $L : \mathbb{R} \mapsto [0, \infty[$ est à variations régulières d'indice 0 ($L \in \mathcal{RV}_0$), on dit que L est à variations lentes.

Le résultat ci-dessous fournit une représentation des fonctions à variations régulières.

Proposition 5 (Représentation de Karamata) Toute fonction $U \in \mathcal{RV}_\alpha$ avec $\alpha \in \mathbb{R}$ s'écrit sous la forme

$$U(x) = c(x) \exp \left\{ \int_{x_0}^x \frac{\Delta(t)}{t} dt \right\},$$

où $x_0 \in \mathbb{R}$, c et Δ sont des fonctions mesurables avec $c(x) \rightarrow c > 0$ et $\Delta(x) \rightarrow \alpha$ lorsque x converge vers l'infini.

Définition 11 Si la fonction c est constante, on dit que U est une fonction à variations régulières normalisée.

Nous donnons à présent quelques propriétés utiles sur les fonctions à variations régulières. Le premier résultat assure que la convergence simple de la définition 9 est en fait une convergence localement uniforme (c'est-à-dire uniforme sur les intervalles fermés de \mathbb{R}).

Proposition 6 (Convergence uniforme locale) Si $U \in \mathcal{RV}_\alpha$, $\alpha \in \mathbb{R}$ alors pour tout intervalle $I = [a, b] \subset \mathbb{R}$, $0 < a < b < \infty$, on a :

$$\sup_{t \in I} \left| \frac{U(tx)}{U(x)} - t^\alpha \right| \rightarrow 0 \text{ lorsque } x \rightarrow \infty.$$

Si $\alpha < 0$, le résultat de convergence uniforme est vrai pour des intervalles I de la forme $[a, \infty[$, $a > 0$. Si $\alpha > 0$ et si U est bornée sur les intervalles $]0, b]$, $b > 0$, alors la convergence uniforme est vraie sur les intervalles $]0, b]$, $b > 0$.

Une propriété intéressante des fonctions à variations régulières est la conservation des équivalents à l'infini.

Proposition 7 Soit $U \in \mathcal{RV}_\alpha$, $\alpha \in \mathbb{R}$ et soient (u_n) et (v_n) deux suites positives telles que $u_n \rightarrow \infty$ et $v_n \rightarrow \infty$ lorsque n tend vers l'infini. Alors si $u_n \sim v_n$, on a $U(u_n) \sim U(v_n)$.

Le prochain résultat s'intéresse au comportement à l'infini d'une fonction à variations régulières (ou rapides).

Proposition 8 Si $U \in \mathcal{RV}_\alpha$, $\alpha \in [-\infty, +\infty]$, alors,

$$\lim_{x \rightarrow \infty} \frac{\log U(x)}{\log(x)} = \alpha.$$

Ainsi,

$$\lim_{x \rightarrow \infty} U(x) = \begin{cases} 0 & \text{si } \alpha < 0, \\ +\infty & \text{si } \alpha > 0. \end{cases}$$

On s'intéresse à présent à l'inverse généralisée d'une fonction à variations régulières croissante.

Proposition 9 *Si $U \in \mathcal{RV}_\alpha$ avec $\alpha \geq 0$ est une fonction croissante telle que $U(x) \rightarrow \infty$ lorsque $x \rightarrow \infty$ alors l'inverse généralisée de U est à variations régulières d'indice $1/\alpha$ ($U^\leftarrow \in \mathcal{RV}_{1/\alpha}$).*

Pour finir ce paragraphe, nous nous intéressons à la dérivée d'une fonction à variations régulières normalisée.

Proposition 10 *Si U est une fonction à variations régulières normalisée d'indice $\alpha \in \mathbb{R}$ et si U est dérivable de dérivée u alors*

$$\lim_{x \rightarrow \infty} \frac{xu(x)}{U(x)} = \alpha. \quad (2.1)$$

Si de plus $\alpha \neq 0$, $|u| \in \mathcal{RV}_{\alpha-1}$.

2.2.2 Domaine d'attraction de Fréchet

Le résultat ci-dessous assure que toute fonction appartenant au domaine d'attraction de Fréchet est une fonction à variations régulières.

Théorème 2 *Une fonction de répartition F appartient au domaine d'attraction $\Psi_\gamma^{(F)}$ (de Fréchet, avec un indice des valeurs extrêmes $\gamma > 0$) si et seulement si la fonction de survie $\bar{F} \in \mathcal{RV}_{-1/\gamma}$. Des suites possibles de normalisation (a_n) et (b_n) sont données par $a_n = \bar{F}^\leftarrow(1/n)$ et $b_n = 0$.*

Autrement dit, une fonction de répartition F appartenant au domaine d'attraction de Fréchet s'écrit sous la forme :

$$F(x) = 1 - x^{-1/\gamma}L(x), \quad L \in \mathcal{RV}_0. \quad (2.2)$$

2.2.3 Domaine d'attraction de Weibull

Le résultat suivant montre que l'on passe du domaine d'attraction de Fréchet à celui de Weibull par un simple changement de variable dans la fonction de répartition.

Théorème 3 *Une fonction de répartition F appartient au domaine d'attraction de $\Psi_\gamma^{(W)}$ (Weibull, avec un indice des valeurs extrêmes $\gamma < 0$) si et seulement si son point terminal x_F est fini et si la fonction de répartition F^* définie par*

$$F^*(x) = \begin{cases} 0 & \text{si } x < 0 \\ F(x_F - 1/x) & \text{si } x \geq 0, \end{cases}$$

appartient au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $-\gamma > 0$. Des suites possibles de normalisation (a_n) et (b_n) sont données par $a_n = x_F - \bar{F}^\leftarrow(1/n)$ et $b_n = x_F$.

Ainsi, si une variable aléatoire X appartient au domaine d'attraction de Weibull avec un indice $\gamma < 0$ alors la variable aléatoire $Y = 1/(x_F^* - X)$ appartient au domaine d'attraction de Fréchet avec un indice $-\gamma > 0$.

2.2.4 Domaine d'attraction de Gumbel

La caractérisation des fonctions de répartition du domaine d'attraction de Gumbel est plus complexe.

Théorème 4 Une fonction de répartition F appartient au domaine d'attraction de $\Psi^{(G)}$ (Gumbel) si et seulement si il existe $z < x_F \leq \infty$ tel que

$$\bar{F}(x) = c(x) \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\}, \quad z < x < x_F, \quad (2.3)$$

où $c(x) \rightarrow c > 0$ lorsque $x \rightarrow x_F$ et a est une fonction positive et dérivable de dérivée a' telle que $a'(x) \rightarrow 0$ lorsque $x \rightarrow x_F$. Des suites possibles de normalisation (a_n) et (b_n) sont données par $b_n = \bar{F}^{\leftarrow}(1/n)$ et $a_n = a(b_n)$.

Le domaine d'attraction de Gumbel regroupe une grande diversité de lois comptant parmi elles la plupart des lois usuelles (loi normale, exponentielle, gamma, log-normale). Cette famille étant difficile à étudier dans toute sa généralité, de nombreux auteurs se sont concentrés sur une sous-famille : les lois à queue de type Weibull.

Définition 12 On dit qu'une loi est à queue de type Weibull si sa fonction de répartition est de la forme

$$\bar{F}(x) = \exp \{-V^{\leftarrow}(x)\}, \quad (2.4)$$

où $V \in \mathcal{RV}_\theta$, $\theta > 0$ étant appelé l'indice de queue de Weibull.

Cette famille de lois contient par exemple les lois normale, Gamma, exponentielle, etc. Par contre, la loi log-normale qui appartient au domaine d'attraction de Gumbel n'est pas une loi à queue de type Weibull.

2.2.5 Loi de Pareto Généralisée - Théorème de Pickands

Nous donnons tout d'abord la définition de la loi de Pareto Généralisée.

Définition 13 La loi de Pareto Généralisée de paramètres $\gamma \in \mathbb{R}$ et $\sigma > 0$ est définie par sa fonction de répartition donnée par :

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

pour $x \in \{t \in \mathbb{R}; 1 + \gamma t / \sigma > 0\} \cap [0, \infty[$.

Le résultat suivant permet de caractériser l'ensemble des trois domaines d'attraction.

Proposition 11 La fonction de répartition F de point terminal x_F appartient au domaine d'attraction de H_γ si et seulement si il existe une fonction positive a telle que

$$\lim_{u \uparrow x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = 1 - G_{\gamma, 1}(x).$$

Nous donnons à présent le résultat important de ce paragraphe assurant que la loi des excès au dessus d'un seuil peut-être approchée par une loi de Pareto Généralisée. Donnons dans un premier temps la définition de la loi des excès.

Définition 14 Soit X une variable aléatoire de fonction de répartition F et de point terminal x_F . Pour tout $u < x_F$ la fonction de répartition des excès de X au dessus du seuil u est définie pour tout $x \in [0, x_F - u[$ par :

$$F_u(x) := \mathbb{P}(X - u \leq x | X > u).$$

De même, si X est intégrable, la fonction excès moyen (mean excess function) de X est donnée par :

$$e_u(X) := \mathbb{E}(X - u | X > u).$$

Nous pouvons à présent énoncer le théorème de Pickands. C'est en fait une conséquence directe de la Proposition 11.

Théorème 5 La fonction de répartition F de point terminal x_F appartient au domaine d'attraction de H_γ si et seulement si il existe une fonction positive a telle que

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\gamma, a(u)}(x)| = 0.$$

Chapitre 3

Statistiques d'ordres

Lorsqu'on dispose d'un échantillon X_1, \dots, X_n de variables aléatoires indépendantes et de même fonction de répartition F , les statistiques d'ordres sont les observations ordonnées c'est-à-dire les variables aléatoires $X_{1,n} \leq \dots \leq X_{n,n}$. Evidemment, les statistiques d'ordre ne sont ni indépendantes ni de même loi. L'objectif de ce chapitre est de donner quelques résultats sur la loi de la statistique $X_{k,n}$ pour $k \in \{1, \dots, n\}$.

3.1 Distribution des statistiques d'ordre

Proposition 12 *La fonction de répartition de la statistique d'ordre $X_{k,n}$ est donnée pour tout $x \in \mathbb{R}$ par*

$$F_{k,n}(x) := \mathbb{P}(X_{k,n} \leq x) = \sum_{r=k}^n C_n^r F^r(x) (1 - F(x))^{n-r}.$$

Dans le cas où la fonction de répartition F est continue et dérivable presque-partout de dérivée f , les statistiques d'ordres ont une loi qui admet une densité par rapport à la mesure de Lebesgue. La densité de $X_{k,n}$ est donnée par

$$f_{k,n}(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x) [1 - F(x)]^{n-k}.$$

Comme autre conséquence directe de ce résultat, on montre facilement que si U_1, \dots, U_n sont des variables aléatoires indépendantes de loi uniforme alors $U_{k,n}$ suit une loi Beta de paramètres k et $n - k + 1$. On rappelle que la densité d'une loi Beta de paramètres $a > 0$ et $b > 0$ est

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{I}_{[0,1]}(x).$$

En particulier, pour tout $s \in \mathbb{N}$,

$$\mathbb{E}(U_{k,n}^s) = \frac{n!}{(n+s)!} \frac{(k+s-1)!}{(k-1)!}.$$

Terminons ce paragraphe en donnant la densité du vecteur aléatoire des statistiques d'ordre.

Proposition 13 *Si la fonction de répartition F est continue et dérivable presque-partout de dérivée f , la densité du vecteur $(X_{1,n}, \dots, X_{n,n})$ est*

$$f_n(x_1, \dots, x_n) = n! \prod_{i=1}^n f(x_i) \mathbb{I}\{x_1 \leq \dots \leq x_n\}.$$

Le résultat suivant est un résultat essentiel sur les statistiques d'ordre d'un échantillon uniforme ou exponentiel. Ce résultat sera utilisé à plusieurs reprises dans la suite.

Théorème 6 (Représentation de Rényi) *On note U_1, \dots, U_n un échantillon de loi uniforme, E_1, \dots, E_{n+1} et F_1, \dots, F_n deux échantillons de loi exponentielle. On a les deux égalités en lois suivantes :*

$$\{U_{j,n}, j = 1, \dots, n\} \stackrel{\mathcal{L}}{=} \left\{ \frac{T_j}{T_{n+1}}, j = 1, \dots, n \right\}$$

où $T_j = E_1 + \dots + E_j$ et

$$\{E_{j,n}, j = 1, \dots, n\} \stackrel{\mathcal{L}}{=} \left\{ \sum_{r=1}^j \frac{F_r}{n-r+1}, j = 1, \dots, n \right\}.$$

Chapitre 4

Estimation de quantiles extrêmes

La question qui nous intéresse dans ce chapitre est la suivante : Soient X_1, \dots, X_n des variables aléatoires indépendantes et de même fonction de répartition F . Pour une suite (α_n) qui converge vers 0 lorsque n converge vers l'infini, on souhaite estimer le quantile extrême d'ordre α_n c'est-à-dire la quantité $q(\alpha_n) := \bar{F}^{\leftarrow}(\alpha_n)$. Comme nous l'avons déjà mentionné dans l'introduction, la valeur $q(\alpha_n)$ est située dans la queue de distribution de F et ne peut pas être généralement estimée par l'estimateur empirique classique \hat{F}_n .

Dans ce chapitre nous allons dans un premier temps proposer des approximations possibles de la quantité $q(\alpha)$ lorsque $\alpha \rightarrow 0$. Dans un second temps, nous présenterons différentes méthodes d'estimation.

4.1 Approximation d'un quantile extrême

Il existe essentiellement deux méthodes permettant de proposer une valeur approchée pour $q(\alpha)$ avec α proche de 0 : celle utilisant la loi des valeurs extrêmes et celle utilisant la loi de Pareto Généralisée. On considèrera également le cas la loi sous-jacente des observations est dans le domaine d'attraction de Fréchet.

4.1.1 Loi des valeurs extrêmes

Le théorème fondamental de la théorie des valeurs extrêmes assure que si une fonction de répartition F appartient à un domaine d'attraction alors il existe deux suites $(a_k) > 0$ et (b_k) telles que

$$\lim_{k \rightarrow \infty} F^k(a_k x + b_k) = H_\gamma(x) = \begin{cases} \exp[-(1 + \gamma x)_+^{-1/\gamma}] & \text{si } \gamma \neq 0, \\ \exp(-e^{-x}) & \text{si } \gamma = 0. \end{cases}$$

Cette convergence est équivalente à la convergence

$$\lim_{k \rightarrow \infty} k \bar{F}(a_k x + b_k) = -\log H_\gamma(x).$$

Ainsi, en posant $y = a_k x + b_k$, on a pour k assez grand (ou de manière équivalente pour y proche du point terminal de F puisque $a_k x + b_k \rightarrow x_F$ lorsque $k \rightarrow \infty$) l'approximation

$$\bar{F}(y) \approx -\frac{1}{k} \log H_\gamma \left(\frac{y - b_k}{a_k} \right) = \frac{1}{k} \left(1 + \gamma \frac{y - b_k}{a_k} \right)_+^{-1/\gamma}.$$

Noter que lorsque $\gamma = 0$, l'expression ci-dessus a un sens par passage à la limite. Pour obtenir une approximation de $q(\alpha)$ avec α proche de 0, il suffit simplement d'inverser l'expression ci-dessus ce qui donne

$$q(\alpha) \approx b_k + \frac{a_k}{\gamma} [(k\alpha)^{-\gamma} - 1]. \quad (4.1)$$

4.1.2 Loi de Pareto Généralisée

Cette méthode d'approximation se base sur le théorème de Pickands qui annonce que pour un seuil u proche du point terminal x_F de F , la fonction de répartition des excès définie pour $x > 0$ par

$$F_u(x) := \mathbb{P}(X - u \leq x | X > u) = 1 - \frac{\bar{F}(x + u)}{\bar{F}(u)},$$

peut être approchée par la loi de Pareto Généralisée de fonction de répartition

$$G_{\gamma, a(u)}(x) := 1 + \log H_\gamma \left(\frac{x}{a(u)} \right),$$

où a est une fonction positive. En posant $y = x + u$, on a donc pour y proche du point terminal l'approximation

$$\bar{F}(y) \approx -\bar{F}(u) \log H_\gamma \left(\frac{y - u}{a(u)} \right).$$

En inversant cette expression, on en déduit que pour α proche de 0,

$$q(\alpha) \approx u + \frac{a(u)}{\gamma} \left[\left(\frac{\alpha}{\bar{F}(u)} \right)^{-\gamma} - 1 \right]. \quad (4.2)$$

On peut remarquer que cette approximation est de la même forme que l'approximation (4.1) en prenant $k = 1/\bar{F}(u) \rightarrow \infty$ lorsque $u \rightarrow x_F$, $b_k = u$ et $a_k = a(u)$. La différence entre ces deux approximations réside donc dans la méthode que l'on utilisera pour estimer les paramètres inconnus (b_k et a_k pour (4.1) et $\bar{F}(u)$ et $a(u)$ pour (4.2)).

4.1.3 Cas particulier du domaine d'attraction de Fréchet

Nous nous intéressons ici à l'estimation de quantiles extrêmes pour des lois dont on sait qu'elles appartiennent au domaine d'attraction de Fréchet. On rappelle que dans ce cas la fonction de répartition est de la forme

$$\bar{F}(x) = x^{-1/\gamma} L(x),$$

avec $\gamma > 0$ et L une fonction à variations lentes. En d'autre terme, \bar{F} est une fonction à variations régulières d'indice $-1/\gamma$. On sait également (voir Proposition 9) que le quantile associé est tel que

$$q(\alpha) = \alpha^{-\gamma} \ell(\alpha^{-1}),$$

où ℓ est une autre fonction à variations lentes. Pour estimer $q(\alpha)$ il suffit donc d'estimer γ et la fonction ℓ . Malheureusement, la fonction à variations lentes est très difficile à estimer en pratique. Pour se passer de cette estimation, on utilise l'astuce suivante : pour $t > 1$,

$$\frac{q(\alpha)}{q(t\alpha)} = t^\gamma \frac{\ell(\alpha^{-1})}{\ell(t^{-1}\alpha^{-1})}.$$

Ainsi, pour α proche de 0, en utilisant le fait que ℓ est une fonction à variations lentes, on obtient l'approximation

$$q(\alpha) \approx q(t\alpha)t^\gamma. \quad (4.3)$$

Pour estimer $q(\alpha)$, il suffira donc d'estimer γ et $q(t\alpha)$. L'estimation de $q(\alpha)$ requiert donc l'estimation de $q(t\alpha)$ et il semble donc que le problème soit simplement déplacé. L'idée sera de prendre t suffisamment grand de telle sorte que $t\alpha$ soit un ordre de quantile classique que l'on pourra estimer par la fonction de répartition empirique.

4.2 Estimation d'un quantile extrême

En utilisant l'une des trois approximations présentées dans le paragraphe précédent, il s'agit ici d'estimer le quantile extrême $q(\alpha_n)$ à partir d'un échantillon X_1, \dots, X_n .

4.2.1 Estimation en utilisant l'approximation EVD

Remarquons que cette méthode est basée sur l'approximation

$$\mathbb{P}[\max(X_1, \dots, X_k) \leq x] \approx H_\gamma \left(\frac{x - b_k}{a_k} \right),$$

qui est d'autant meilleure que k est grand. Les paramètres a_k , b_k et γ doivent être estimés à l'aide d'un échantillon de maxima de k variables aléatoires indépendantes de fonction de répartition F . Pour obtenir un tel échantillon, il

faut utiliser la méthode dite des "maxima par blocs". On choisit une valeur $k \in \{1, \dots, n\}$ assez grande, on divise l'échantillon X_1, \dots, X_n en $m = \lfloor n/k \rfloor$ sous-échantillons de taille k et on récupère ainsi m maxima M_1, \dots, M_m que l'on pourra utiliser pour estimer a_k, b_k et γ . Noter que ces m maxima sont bien *i.i.d.* mais que leur loi commune n'est pas exactement une loi EVD.

Pour l'estimation, on peut par exemple utiliser la méthode du maximum de vraisemblance en faisant comme si la loi commune des maxima était une loi des valeurs extrêmes de paramètres a_k, b_k et γ . Cette méthode requiert un grand nombre d'observations (pour avoir un nombre m de sous-échantillons conséquent). Elle est souvent utilisée pour étudier des données pluviométriques en prenant pour bloc les observations sur une journée (ou sur un mois, une année). On ne traitera pas en détail dans ce document les méthodes d'estimation pour cette approximation.

4.2.2 Estimation en utilisant l'approximation GPD

On rappelle que pour un seuil u proche du point terminal de F , on peut approcher $q(\alpha_n)$ avec $\alpha_n \rightarrow 0$ par

$$u + \frac{a(u)}{\gamma} \left[\left(\frac{\alpha_n}{\bar{F}(u)} \right)^{-\gamma} - 1 \right].$$

Pour estimer $q(\alpha_n)$ il faut donc dans un premier temps choisir un seuil u suffisamment grand et ensuite estimer les paramètres γ et $a(u)$ de la loi de Pareto Généralisée suivie approximativement par les excès au dessus du seuil u .

En pratique, on choisit tout d'abord u de telle sorte que $\bar{F}(u) = k_n/n$ (*i.e* $u = q(k_n/n)$) où $k_n \in \{1, \dots, n\}$ est une suite intermédiaire c'est-à-dire telle que

$$k_n \rightarrow \infty \text{ et } k_n/n \rightarrow 0.$$

La fonction de survie \bar{F} étant inconnue, une telle valeur de u est elle aussi inconnue mais on peut l'estimer facilement car k_n/n est un ordre de quantile classique. On prend donc pour seuil la statistique

$$\hat{u}_n := \hat{\bar{F}}_n^{\leftarrow}(k_n/n) = X_{n-k_n, n}.$$

Pour estimer les paramètres γ et $a(k_n) := a(q(k_n/n))$ de la loi de Pareto Généralisée, on utilise donc les excès au dessus de \hat{u}_n c'est-à-dire les k_n variables aléatoires

$$Y_1 := X_{n-k_n+1, n} - X_{n-k_n, n}, \dots, Y_{k_n} := X_{n, n} - X_{n-k_n, n}.$$

D'après le théorème de Pickands, Y_1, \dots, Y_{k_n} suivent approximativement une loi de Pareto Généralisée de paramètres γ et $a(u)$. Pour proposer des estimateurs de ces paramètres, on raisonnera comme si les variables Y_1, \dots, Y_{k_n} étaient indépendantes (ce qui n'est évidemment pas le cas) et de même loi de Pareto

Généralisée (ce qui n'est pas non plus le cas). En notant $\hat{\gamma}(k_n)$ et $\hat{a}(k_n)$ des estimateurs de γ et $a(k_n)$ obtenus à partir de l'échantillon des excès, on estimera donc $q(\alpha_n)$ par

$$\hat{q}_n(\alpha_n) := X_{n-k_n+1,n} + \frac{\hat{a}(k_n)}{\hat{\gamma}(k_n)} \left[\left(\frac{n\alpha_n}{k_n} \right)^{-\hat{\gamma}(k_n)} - 1 \right].$$

Dans la suite de ce paragraphe, on s'intéresse à la question suivante : à partir de variables aléatoires Z_1, \dots, Z_k indépendantes et de même loi GPD de paramètres $\gamma \in \mathbb{R}$ et $\theta > 0$, estimer γ et θ . Pour ce faire, on recense essentiellement trois méthodes :

- 1) la méthode du maximum de vraisemblance,
- 2) la méthode des moments pondérés,
- 3) la méthode des percentiles.

Dans ce document, on ne présentera pas la méthode du maximum de vraisemblance qui ne fonctionne pas dans tous les cas et fournit souvent des estimateurs non explicites.

Méthode des moments pondérés

Cette méthode présente l'avantage de donner de bons résultats pour des échantillons de petites tailles. Par contre, comme nous le verrons, elle n'est utilisable que si l'indice des valeurs extrêmes γ est inférieur (strictement) à 1. On définit dans un premier temps la notion de moments pondérés.

Définition 15 *Soit Z une variable aléatoire de fonction de répartition F . Si Z est intégrable, le moment pondéré d'ordres $r \in \mathbb{N}$ et $s \in \mathbb{N}$ de Z est*

$$\text{WM}_Z(r, s) := \mathbb{E}[ZF^r(Z)(1 - F(Z))^s].$$

Dans la suite, on prendra $r = 0$. Le résultat suivant donne la valeur du moment pondéré d'une loi GPD de paramètres γ et θ .

Théorème 7 *Soit Z une variable aléatoire de loi de Pareto Généralisé de paramètres γ et θ . Si $\gamma < 1$, on a pour tout $s \in \mathbb{N}$*

$$\text{WM}_Z(0, s) = \frac{\theta}{(s+1)(s+1-\gamma)}.$$

En utilisant deux moments pondérés d'ordres s différents, on peut écrire γ et θ en fonction des ces moments.

Proposition 14 *Pour $s \in \mathbb{N}$, en posant $\nu_s = \text{WM}_Z(0, s)$, on a*

$$\theta = \frac{2\nu_0\nu_1}{\nu_0 - 2\nu_1} \text{ et } \gamma = \frac{4\nu_1 - \nu_0}{2\nu_1 - \nu_0}. \quad (4.4)$$

De la même façon que pour la méthode des moments classique, on propose d'estimer γ et θ en remplaçant dans (4.4) les moments ν_0 et ν_1 par des estimateurs convenables. Il s'agit donc à présent d'estimer le moment pondéré

$$\nu_s = \mathbb{E} [Z(1 - G_{\gamma,\theta}(Z))^s],$$

où Z est une variable aléatoire de loi de Pareto Généralisée de paramètres γ et θ . On dispose pour ce faire d'un échantillon Z_1, \dots, Z_k de k copies indépendantes de Z . La première idée consiste à remplacer l'espérance par l'espérance empirique. On obtient

$$\tilde{\nu}_s := \frac{1}{k} \sum_{i=1}^k Z_i [1 - G_{\gamma,\theta}(Z_i)]^s = \frac{1}{k} \sum_{i=1}^k Z_{i,k} [1 - G_{\gamma,\theta}(Z_{i,k})]^s.$$

Evidemment la variable aléatoire $\tilde{\nu}_s$ n'est pas un estimateur puisque la fonction de répartition $G_{\gamma,\theta}$ reste inconnue. Pour palier ce problème, on propose de remplacer dans l'expression de $\tilde{\nu}_s$ la variable aléatoire $1 - G_{\gamma,\theta}(Z_{i,k})$ par son espérance. Pour calculer cette espérance, on utilise le fait que pour tout $i = 1, \dots, k$, si U_1, \dots, U_k sont k variables aléatoires indépendantes de loi uniforme sur $[0, 1]$ alors $G_{\gamma,\theta}(Z_i) \stackrel{\mathcal{L}}{=} U_i$. Les variables aléatoires étant indépendantes, le résultat reste vrai pour le vecteur c'est-à-dire

$$\{G_{\gamma,\theta}(Z_i), i = 1, \dots, k\} \stackrel{\mathcal{L}}{=} \{U_i, i = 1, \dots, k\}.$$

Une fonction de répartition étant croissante, on en déduit que

$$\{1 - G_{\gamma,\theta}(Z_{i,k}), i = 1, \dots, k\} \stackrel{\mathcal{L}}{=} \{1 - U_{i,k}, i = 1, \dots, k\} \stackrel{\mathcal{L}}{=} \{U_{k-i+1,k}, i = 1, \dots, k\}.$$

En utilisant le fait que

$$\mathbb{E}(U_{k-i+1,k}^s) = \frac{(k-i+s)!}{(k-i)!} \frac{k!}{(k+s)!},$$

on estime le moment pondéré ν_s par

$$\hat{\nu}_s := \frac{1}{k} \sum_{i=1}^k Z_{i,k} \frac{(k-i+s)!}{(k-i)!} \frac{k!}{(k+s)!}.$$

Méthode des percentiles

L'avantage de cette méthode par rapport à la méthode des moments pondérés est qu'elle est utilisable quelque soit la valeur de l'indice γ . On rappelle que l'on dispose d'un échantillon Z_1, \dots, Z_k de variables aléatoires indépendantes de loi de Pareto Généralisée de paramètres $\gamma \in \mathbb{R}$ et $\theta > 0$. On a déjà montré que

$$\{\bar{F}(Z_{i,k}), i = 1, \dots, k\} \stackrel{\mathcal{L}}{=} \{U_{k-i+1,k}, i = 1, \dots, k\},$$

où $F := G_{\gamma, \theta}$ est la fonction de répartition d'une loi GPD. La méthode des percentiles est basée sur l'approximation

$$\bar{F}(Z_{i,k}) \approx \mathbb{E}[\bar{F}(Z_{i,k})] = \mathbb{E}[U_{k-i+1,k}] = 1 - \frac{i}{k+1}.$$

Pickands en 1975 propose d'estimer les paramètres d'une loi de Pareto Généralisée en résolvant le système d'équations obtenu en prenant dans l'approximation ci-dessus $i = k/2$ et $j = 3k/4$ (on suppose que k est divisible par 4). Le système s'écrit

$$\begin{cases} \bar{F}(Z_{k/2,k}) = (1 + \gamma Z_{k/2,k}/\theta)^{-1/\gamma} & \approx 1 - k/[2(k+1)] \approx 1/2 \\ \bar{F}(Z_{3k/4,k}) = (1 + \gamma Z_{3k/4,k}/\theta)^{-1/\gamma} & \approx 1/4. \end{cases}$$

pour une valeur de k assez grande. On propose finalement d'estimer γ et θ par les statistiques

$$\hat{\gamma} := \frac{1}{\log(2)} \log \frac{Z_{3k/4,k} - Z_{k/2,k}}{Z_{k/2,k}} \text{ et } \hat{\theta} := \frac{\hat{\gamma} Z_{k/2,k}^2}{Z_{3k/4,k} - 2Z_{k/2,k}}.$$

Remarque – L'estimateur de γ ci-dessus sera en fait calculé en prenant les excès Y_1, \dots, Y_{k_n} définies au début de ce paragraphe à la place des variables aléatoires Z_1, \dots, Z_k (même si les excès ne sont ni indépendants ni de loi GPD). En rappelant que

$$Y_i = Y_{i,k_n} = X_{n-k_n+i,n} - X_{n-k_n,n},$$

l'estimateur de γ s'écrit

$$\hat{\gamma} = \frac{1}{\log(2)} \log \frac{X_{n-k_n/4,n} - X_{n-k_n/2,n}}{X_{n-k_n/2,n} - X_{n-k_n,n}}.$$

Cet estimateur est connu dans la littérature sous le nom d'estimateur de Pickands. On peut montrer que sous certaines hypothèses, l'estimateur de Pickands est un estimateur consistant de γ .

4.2.3 Estimation pour le domaine d'attraction de Fréchet

On se place ici sous l'hypothèse que les variables aléatoires X_1, \dots, X_n sont indépendantes et de même fonction de répartition F appartenant au domaine d'attraction de Fréchet. On rappelle que dans ce cas, le quantile extrême $q(\alpha_n)$ avec $\alpha_n \rightarrow 0$ est tel que, pour $t > 1$,

$$q(\alpha_n) \approx q(t\alpha_n)t^\gamma.$$

Pour estimer $q(\alpha_n)$ à partir de cette approximation, il faut donc dans un premier temps choisir une valeur de t de telle sorte que $t\alpha_n$ soit un ordre de quantile classique. On prend en pratique $t = t_n = k_n/(n\alpha_n)$ où (k_n) est une suite intermédiaire. On a ainsi $q(t\alpha_n) = q(k_n/n)$ que l'on peut estimer en inversant

la fonction de répartition empirique par la statistique d'ordre $X_{n-k_n, n}$. Il reste donc ensuite à proposer un estimateur de γ . En notant $\hat{\gamma}_n$ un tel estimateur, on aura ainsi estimé $q(\alpha_n)$ par

$$\hat{q}(\alpha_n) = X_{n-k_n, n} \left(\frac{k_n}{n\alpha_n} \right)^{\hat{\gamma}_n}.$$

Cet estimateur est connu sous le nom d'estimateur de Weissman. Il existe de nombreux estimateurs de γ , le plus connu (et utilisé) étant l'estimateur de Hill que nous présentons ci-après.

Estimateur de Hill

Nous donnons pour commencer la définition de cet estimateur.

Définition 16 Soient X_1, \dots, X_n des variables aléatoires indépendantes et de même fonction de répartition F appartenant au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $\gamma > 0$. L'estimateur de Hill de γ est

$$\hat{\gamma}_n^{(H)}(k_n) := \frac{1}{k_n} \sum_{i=1}^{k_n} \log(X_{n-i+1, n}) - \log(X_{n-k_n, n}),$$

où $1 \leq k_n \leq n$ est une valeur à choisir par l'utilisateur.

Remarque – Le choix de k_n (c'est-à-dire du nombre d'observations que l'on considère) est crucial. Si k_n est trop grand, alors l'approximation par une loi de Pareto sera mauvaise et l'estimateur de Hill aura un biais important. A l'inverse, si k_n est trop petit, on disposera de très peu d'observations pour le calcul de l'estimateur qui aura alors une variance importante. Le bon choix de k_n est donc celui de meilleur compromis biais/variance. En théorie, on verra que pour obtenir la consistance de l'estimateur de Hill, il faut prendre pour k_n une suite intermédiaire : $k_n \rightarrow \infty$ (k_n assez grand) et $k_n/n \rightarrow 0$ (k_n pas trop grand).

Pour choisir cette suite k_n en pratique, une méthode graphique est souvent utilisée. Elle consiste à tracer le "Hill Horror Plot" c'est-à-dire le graphique des points $\{(k_n, \hat{\gamma}_n^{(H)}(k_n)), k_n = 1, \dots, n\}$ (voir exemple ci-dessous pour un échantillon de taille $n = 200$ issu de la valeur absolue d'une loi de Cauchy standard).

On remarque sur la Figure 4.1 que le graphe est très chaotique pour k_n petit (entre 0 et 20), qu'il y a ensuite une zone de stabilité et enfin que pour $k_n > 120$, l'estimateur de Hill devient de plus en plus biaisé. Le meilleur choix de k_n se situe dans la zone de stabilité du "Hill Horror Plot".

Comportement asymptotique de l'estimateur de Hill

Nous donnons dans ce paragraphe un résultat sur la normalité asymptotique de l'estimateur de Hill défini dans le paragraphe précédent. Nous commençons par donner une décomposition de l'estimateur de Hill. Comme d'habitude, on

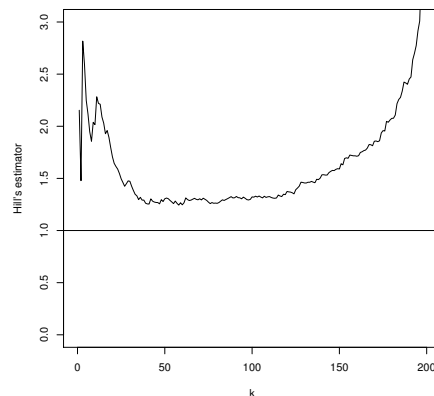


FIGURE 4.1 – Estimateur de Hill en fonction de la suite (k_n) pour un échantillon de taille $n = 200$.

utilisera les notations $\{E_i, i \geq 1\}$ pour des variables aléatoires indépendantes de loi exponentielle et $\{U_i, i \geq 1\}$ pour des variables aléatoires indépendantes de loi uniforme.

Proposition 15 *Si X_1, \dots, X_n sont des variables aléatoires indépendantes dont la loi commune appartient au domaine d'attraction de Fréchet (et donc telle que le quantile d'ordre α s'écrit $q(\alpha) = \alpha^{-\gamma} \ell(\alpha^{-1})$) alors*

$$\hat{\gamma}_n^{(H)}(k_n) \stackrel{\mathcal{L}}{=} \gamma A(k_n) + R(k_n),$$

où

$$A(k_n) \stackrel{\mathcal{L}}{=} \frac{1}{k_n} \sum_{i=1}^{k_n} E_i \text{ et } R(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{\ell(U_{i,n}^{-1})}{\ell(U_{k_n+1,n}^{-1})}.$$

A partir de cette décomposition, il est facile de voir en utilisant le Théorème Central Limite que dans le cas d'une loi de Pareto (*i.e.* lorsque la fonction à variations lentes ℓ est constante)

$$\sqrt{n} \left(\hat{\gamma}_n^{(H)}(n) - \gamma \right) \stackrel{\mathcal{L}}{\rightarrow} \mathcal{N}(0, \gamma^2).$$

Remarquer que dans ce cas on a pris $k_n = n$. Si la fonction à variations lentes n'est pas constante, il reste à montrer que $(k_n)^{1/2} R(k_n) \xrightarrow{\mathbb{P}} 0$ lorsque $n \rightarrow \infty$ pour montrer que

$$\sqrt{k_n} \left(\hat{\gamma}_n^{(H)}(k_n) - \gamma \right) \stackrel{\mathcal{L}}{\rightarrow} \mathcal{N}(0, \gamma^2).$$

En effet,

$$\begin{aligned} \sqrt{k_n} \left(\hat{\gamma}_n^{(H)}(k_n) - \gamma \right) &\stackrel{\mathcal{L}}{=} \gamma \sqrt{k_n} [A(k_n) - 1] + \sqrt{k_n} R(k_n) \\ &\stackrel{\mathcal{L}}{=} \gamma \sqrt{k_n} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} E_i - 1 \right) + \sqrt{k_n} R(k_n). \end{aligned}$$

Le premier terme convergeant en loi vers une loi normale (d'après le TCL), on en déduit directement le résultat. Pour contrôler le terme $R(k_n)$ (et donc la fonction à variations lentes ℓ) on introduit les conditions supplémentaires suivantes.

(H.1) La fonction à variations lentes ℓ est normalisée *i.e.* il existe x_0 tel que

$$\ell(x) = c \exp \left\{ \int_{x_0}^x \frac{\Delta(t)}{t} dt \right\},$$

avec $c > 0$ et $\Delta(t) \rightarrow 0$ lorsque $t \rightarrow \infty$.

Il est facile de remarquer que sous l'hypothèse (H.1), la fonction ℓ est dérivable avec

$$\Delta(t) = \frac{t\ell'(t)}{\ell(t)}.$$

(H.2) La fonction $|\Delta|$ est à variations régulières d'indice $\rho < 0$.

Le paramètre ρ est appelée le paramètre du second ordre.

(H.3) La fonction $|\Delta|$ est asymptotiquement décroissante.

Sous ces hypothèses supplémentaires, on obtient le résultat suivant.

Théorème 8 *Si X_1, \dots, X_n sont des variables aléatoires indépendantes dont la loi commune appartient au domaine d'attraction de Fréchet, sous les hypothèses (H.1), (H.2) et (H.3), si $k_n \rightarrow \infty$, $n/k_n \rightarrow \infty$ et $k_n^{1/2} \Delta(n/k_n) \rightarrow 0$ lorsque $n \rightarrow \infty$ alors*

$$\sqrt{k_n} \left(\hat{\gamma}_n^{(H)}(k_n) - \gamma \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2).$$

Comportement asymptotique de l'estimateur de Weissman

On rappelle que pour toute suite $\alpha_n \rightarrow 0$, l'estimateur de Weissman du quantile $q(\alpha_n)$ est donné par

$$\hat{q}_n(\alpha_n) = X_{n-k_n, n} \left(\frac{k_n}{n\alpha_n} \right)^{\hat{\gamma}_n^{(H)}(k_n)},$$

en utilisant l'estimateur de Hill comme estimateur de $\gamma > 0$. On a le résultat suivant.

Théorème 9 Soient X_1, \dots, X_n des variables aléatoires indépendantes dont la loi commune appartient au domaine d'attraction de Fréchet. Sous les hypothèses (H.1), (H.2) et (H.3), pour toutes suites k_n et α_n telles que $k_n \rightarrow \infty$, $n/k_n \rightarrow \infty$, $k_n^{1/2} \Delta(n/k_n) \rightarrow 0$ et $n\alpha_n/k_n \rightarrow 0$, on a

$$\frac{k_n^{1/2}}{\log[k_n/(n\alpha_n)]} \log \left(\frac{\hat{q}_n(\alpha_n)}{q(\alpha_n)} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2).$$

Chapitre 5

Introduction à la théorie des copules

Lorsque l'on souhaite connaître la loi d'un vecteur aléatoire $X \in \mathbb{R}^p$ où $X := (X^{(1)}, \dots, X^{(p)})^t$, la connaissance des p lois marginales F_1, \dots, F_p , où $F_i(x_i) = \mathbb{P}(X^{(i)} \leq x_i)$ n'est évidemment pas suffisante. Afin de caractériser la fonction de répartition du vecteur X définie pour $(x_1, \dots, x_p)^t \in \mathbb{R}^p$ par

$$G(x_1, \dots, x_p) := \mathbb{P}(X^{(1)} \leq x_1; \dots; X^{(p)} \leq x_p),$$

il faut également connaître la structure de dépendance reliant les marginales $F_i(\cdot) = G(+\infty, \dots, +\infty, \cdot, +\infty, \dots, +\infty)$. Cette dépendance peut (dans certains cas) être modélisée de manière unique par une fonction $C : [0, 1]^p \mapsto [0, 1]$ appelée copule et ceci indépendamment des lois marginales.

5.1 Définition d'une copule

Définition 17 Une copule $C : [0, 1]^p \mapsto [0, 1]$ (de dimension p) est une fonction de répartition multivariée dont les lois marginales sont uniformes sur $[0, 1]$. Autrement dit, pour toute copule C , il existe des variables aléatoires $U^{(1)}, \dots, U^{(p)}$ définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et uniformément distribuées sur $[0, 1]$ telles que pour tout $(u_1, \dots, u_p)^t \in [0, 1]^p$,

$$C(u_1, \dots, u_p) = \mathbb{P}(U^{(1)} \leq u_1, \dots, U^{(p)} \leq u_p).$$

Une copule peut également être définie sans utiliser de modèle probabiliste.

Proposition 16 Une copule $C : [0, 1]^p \mapsto [0, 1]$ est une fonction vérifiant les propriétés suivantes : soit $(u_1, \dots, u_p)^t \in [0, 1]^p$,

- i) $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ pour tout $i = 1, \dots, p$,
- ii) $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_p) = 0$ pour tout $i = 1, \dots, p$,
- iii) C est une fonction p -croissante.

Les deux premières propriétés reflètent le fait que les marginales sont de loi uniforme sur $[0, 1]$. La troisième propriété assure que C est une fonction de répartition multivariée (avec notamment des fonctions de répartition marginales croissantes). Nous donnons ci-dessous la définition exacte de la notion de p -croissance.

Définition 18 Une fonction $h : \mathbb{R}^p \mapsto \mathbb{R}$ est dite p -croissante si pour tout hyper-rectangle $\mathcal{R}_p := [x_1^{(1)}, x_2^{(1)}] \times \dots \times [x_1^{(p)}, x_2^{(p)}]$ le h -volume de \mathcal{R}_p noté $\mathcal{V}_h(\mathcal{R}_p)$ est positif. On rappelle que

$$\mathcal{V}_h(\mathcal{R}_p) := \sum_{j_1=1}^2 \dots \sum_{j_p=1}^2 (-1)^{j_1+\dots+j_p} h(x_{j_1}^{(1)}, \dots, x_{j_p}^{(p)}).$$

Afin de mieux comprendre cette notion de h -volume, revenons à l'interprétation probabiliste de la copule dans le cas simple $p = 2$. Soit $(U^{(1)}, U^{(2)})$ un vecteur aléatoire de fonction de répartition la copule C . Le C -volume de \mathcal{R}_2 est la probabilité que $(U^{(1)}, U^{(2)})$ appartienne au rectangle \mathcal{R}_2 . Autrement dit,

$$\begin{aligned} \mathcal{V}_C(\mathcal{R}_2) &= \mathbb{P}\left(\{x_1^{(1)} \leq U^{(1)} \leq x_2^{(1)}\} \cap \{x_1^{(2)} \leq U^{(2)} \leq x_2^{(2)}\}\right) \\ &= \mathbb{P}(U^{(1)} \leq x_2^{(1)}; U^{(2)} \leq x_2^{(2)}) - \mathbb{P}(U^{(1)} \leq x_2^{(1)}; U^{(2)} \leq x_1^{(2)}) \\ &\quad - \mathbb{P}(U^{(1)} \leq x_1^{(1)}; U^{(2)} \leq x_2^{(2)}) + \mathbb{P}(U^{(1)} \leq x_1^{(1)}; U^{(2)} \leq x_1^{(2)}) \\ &= C(x_2^{(1)}, x_2^{(2)}) - C(x_2^{(1)}, x_1^{(2)}) - C(x_1^{(1)}, x_2^{(2)}) + C(x_1^{(1)}, x_1^{(2)}), \end{aligned}$$

ce qui correspond bien à la définition précédente.

En tant que fonction de répartition multivariée, une copule peut être caractérisée par sa densité (si elle existe) par rapport à la mesure de Lebesgue. Lorsqu'elle existe, la densité d'une copule est donnée pour $(u_1, \dots, u_p) \in [0, 1]^p$ par

$$c(u_1, \dots, u_p) = \frac{\partial^p}{\partial_1 \dots \partial_p} C(u_1, \dots, u_p),$$

où ∂/∂_i désigne la dérivée selon la i -ème composante. Si de plus les fonctions de répartition marginales F_i sont continues (de densité f_i), la densité $g : \mathbb{R}^p \mapsto [0, \infty)$ du vecteur aléatoire $X = (X^{(1)}, \dots, X^{(p)})^t$ est donnée par

$$g(x_1, \dots, x_p) = c(F_1(x_1), \dots, F_p(x_p)) \prod_{i=1}^p f_i(x_i).$$

On en déduit alors une autre expression de la densité de la copule C :

$$c(u_1, \dots, u_p) = g(F_1^{\leftarrow}(u_1), \dots, F_p^{\leftarrow}(u_p)) \left/ \prod_{i=1}^p f_i(F_i^{\leftarrow}(u_i)) \right. .$$

5.2 Théorème de Sklar

Nous donnons à présent le résultat essentiel de la théorie des copules. Ce théorème, établi par Sklar en 1959, montre qu'une copule permet de caractériser complètement la dépendance dans un vecteur aléatoire.

Théorème 10 *i) Si $C : [0, 1]^p \mapsto [0, 1]$ est une copule de dimension p et si F_1, \dots, F_p sont des fonctions de répartition alors, pour tout $(x_1, \dots, x_p) \in \mathbb{R}^p$,*

$$G(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p)), \quad (5.1)$$

est une fonction de répartition multivariée dont les marges sont distribuées selon les fonctions de répartition F_i .

ii) Réciproquement, si $G : \mathbb{R}^p \mapsto [0, 1]$ est une fonction de répartition multivariée de lois marginales F_1, \dots, F_p alors il existe une copule $C : [0, 1]^p \mapsto [0, 1]$ vérifiant (5.1). Cette copule n'est pas forcément unique sauf si F_1, \dots, F_p sont continues et alors

$$C(u_1, \dots, u_p) = G(F_1^{\leftarrow}(u_1), \dots, F_p^{\leftarrow}(u_p)).$$

La démonstration de ce résultat est essentiellement basée sur un résultat important permettant de faire le lien entre une variable aléatoire X de loi quelconque et une variable aléatoire de loi uniforme. Dans la suite, pour une fonction de répartition F , on notera

$$F(z_-) = \lim_{x \uparrow z} F(x).$$

De plus, pour tout $v \in [0, 1]$, on introduit la fonction

$$F_v(z) := F(z_-) + v(F(z) - F(z_-)).$$

Remarquons que F_v n'est en général pas une fonction de répartition (car pas continue à droite) et que $F_v = F$ en tout point de continuité de F .

Proposition 17 *Soit X une variable aléatoire réelle de fonction de répartition F et soit V une variable aléatoire indépendante de X et de loi uniforme sur $[0, 1]$. La variable aléatoire $U := F_V(X)$ suit une loi uniforme sur $[0, 1]$ et de plus, $X = F^{\leftarrow}(U)$ presque-sûrement.*

Le théorème de Sklar est donc essentiel pour comprendre l'utilité des copules. Le point *i)* assure qu'à partir de lois marginales données, on peut construire une loi multivariée dont la dépendance sera gérée par une copule C . Le point *ii)* quant à lui montre que pour une loi multivariée donnée, on peut isoler la copule qui gère la dépendance et ceci indépendamment des marginales. Il faut bien avoir conscience que cette copule n'est pas unique (et n'a donc que peu d'intérêt) pour une loi multivariée discrète par exemple.

Les copules peuvent également être utilisées pour simuler une réalisation d'un vecteur aléatoire $X = (X^{(1)}, \dots, X^{(p)})^t$ de fonction de répartition G (de densité

g) dont les lois marginales F_1, \dots, F_p sont continues.

- Notons C la copule associée à X , la première étape consiste à simuler une réalisation d'un vecteur (U_1, \dots, U_p) de fonction de répartition C . Pour ce faire, on génère une réalisation u_1 d'une variable aléatoire uniforme U_1 . On génère ensuite u_2 selon la fonction de répartition conditionnelle

$$\begin{aligned} C_{2|1}(u) &:= \mathbb{P}(U_2 \leq u | U_1 = u_1) = \int_0^u \frac{\partial^2}{\partial_1 \partial_2} C(u_1, t, 1, \dots, 1) dt \\ &= \frac{\partial}{\partial_1} C(u_1, u, 1, \dots, 1). \end{aligned}$$

Autrement dit, u_2 est une réalisation de la variable aléatoire $C_{2|1}^{\leftarrow}(V_2)$ où V_2 est une variable aléatoire de loi uniforme indépendante de U_1 . On génère ensuite u_3 selon la fonction de répartition

$$\begin{aligned} C_{3|1,2}(u) &:= \mathbb{P}(U_3 \leq u | U_1 = u_1, U_2 = u_2) \\ &= \int_0^u \left[\frac{\partial^3}{\partial_1 \partial_2 \partial_3} C(u_1, u_2, t, 1, \dots, 1) \Big/ \frac{\partial^2}{\partial_1 \partial_2} C(u_1, u_2, 1, \dots, 1) \right] dt \\ &= \frac{\partial^2}{\partial_1 \partial_2} C(u_1, u_2, u, 1, \dots, 1) \Big/ \frac{\partial^2}{\partial_1 \partial_2} C(u_1, u_2, 1, \dots, 1). \end{aligned}$$

et ainsi de suite jusqu'à u_p .

- La deuxième et dernière étape consiste à calculer le vecteur

$$x = (F_1^{\leftarrow}(u_1), \dots, F_p^{\leftarrow}(u_p)),$$

qui sera notre réalisation du vecteur aléatoire X .

Nous terminons ce paragraphe en donnant un résultat d'invariance sur les copules.

Proposition 18 *Soit $X = (X^{(1)}, \dots, X^{(p)})^t$ un vecteur aléatoire dont les lois marginales F_1, \dots, F_p sont supposées continues. On note C la copule associée à X . Soient T_1, \dots, T_p des fonctions strictement croissantes. La copule associée au vecteur aléatoire $X^* := (T_1(X^{(1)}), \dots, T_p(X^{(p)}))^t$ est C .*

Ce résultat montre encore une fois que la copule modélise la dépendance indépendamment des marginales.

5.3 Quelques exemples de copules

Pour simplifier, on se restreint au cas $p = 2$. On note F_1 et F_2 les fonctions de répartition marginales que l'on supposera continues. La fonction de répartition bivariée sera notée G .

5.3.1 Copule d'indépendance

C'est la copule qui caractérise l'indépendance des marginales. Elle est définie pour $(u_1, u_2) \in [0, 1]^2$ par $\Pi(u_1, u_2) = u_1 u_2$. Il est facile de voir que la loi bivariée construite avec cette copule est telle que :

$$\begin{aligned} G(x_1, x_2) &= \mathbb{P}(X^{(1)} \leq x_1, X^{(2)} \leq x_2) = \Pi(F_1(x_1), F_2(x_2)) = F_1(x_1)F_2(x_2) \\ &= \mathbb{P}(X^{(1)} \leq x_1)\mathbb{P}(X^{(2)} \leq x_2), \end{aligned}$$

montrant ainsi que $X^{(1)}$ et $X^{(2)}$ sont indépendantes.

5.3.2 Les bornes de Fréchet-Hoeffding

En utilisant le résultat classique $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$, il est facile de voir que pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$\max\{F_1(x_1) + F_2(x_2) - 1; 0\} \leq G(x_1, x_2) \leq \min\{F_1(x_1), F_2(x_2)\}.$$

Les marginales étant continues, on peut réécrire ces inégalités en terme de copules. En effet, d'après le point *ii*) du Théorème de Sklar, en notant C la copule associée à G , on a $C(u_1, u_2) = G(F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2))$ et ainsi, pour tout $(u_1, u_2) \in [0, 1]^2$,

$$C^-(u_1, u_2) := \max\{u_1 + u_2 - 1; 0\} \leq C(u_1, u_2) \leq C^+(u_1, u_2) := \min\{u_1; u_2\}.$$

La copule C^+ est appelée la copule comonotone. On montre facilement que C^+ est la copule associée au vecteur aléatoire $X := (X^{(1)}, X^{(2)})^t$ où $X^{(1)}$ est une fonction strictement croissante de $X^{(2)}$. La copule C^- est la copule anti-comonotone. Elle est associée au vecteur aléatoire $X := (X^{(1)}, X^{(2)})^t$ où $X^{(1)}$ est une fonction strictement décroissante de $X^{(2)}$. Ces bornes existent également dans le cas $p > 2$. Elles sont données par

$$\begin{aligned} C^-(u_1, \dots, u_p) &:= \max\{u_1 + \dots + u_p - (p-1); 0\} \\ \text{et } C^+(u_1, \dots, u_p) &:= \min\{u_1; \dots; u_p\}. \end{aligned}$$

Il faut prendre garde au fait que C^+ est une copule pour tout p mais que C^- est une copule uniquement si $p = 2$.

5.3.3 Famille de Farlie-Gumble-Morgenstern

C'est une famille paramétrique de copules définies par

$$C_\theta^{(\text{FGM})}(u_1, u_2) = u_1 u_2 + \theta u_1 u_2 (1 - u_1)(1 - u_2),$$

où le paramètre θ appartient à l'intervalle $[-1, 1]$. Il est à noter que cette copule est symétrique (*i.e.* $C_\theta^{(\text{FGM})}(u_1, u_2) = C_\theta^{(\text{FGM})}(u_2, u_1)$) et que $C_0^{(\text{FGM})} = \Pi$.

5.3.4 Copule Gaussienne

C'est une des copules les plus utilisées dans la pratique. Elle est définie de la façon suivante. Soit $\Sigma \in \mathbb{R}^{2 \times 2}$ une matrice de la forme

$$\Sigma = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix},$$

avec $r \in [-1, 1]$. On note $\Phi_2(\cdot|\Sigma)$ la fonction de répartition de la loi normale bivariable centrée et de matrice de variance Σ et Φ_1 la fonction de répartition d'une loi normale centrée et réduite. La copule Gaussienne est donnée pour tout $(u_1, u_2) \in [0, 1]^2$ par :

$$C^{(G)}(u_1, u_2) = \Phi_2(\Phi_1^{\leftarrow}(u_1), \Phi_1^{\leftarrow}(u_2)|\Sigma).$$

Evidemment, si les marginales sont telles que $F_1 = F_2 = \Phi_1$, la fonction de répartition bivariable G associée à la copule $C^{(G)}$ est égale à $\Phi_2(\cdot|\Sigma)$. La copule Gaussienne permet donc de construire des vecteurs aléatoires dont la dépendance entre les marginales est de la même forme que pour un vecteur Gaussien (mais avec des marginales pas nécessairement Gaussiennes).

5.3.5 Copules Archimédiennes

Cette famille de copules est très utilisée en pratique. Elle est définie à partir d'une fonction génératrice $\varphi : [0, 1] \mapsto [0, \infty[$ que l'on suppose décroissante, convexe et telle que $\varphi(1) = 0$. La copule Archimédienne de générateur φ est définie par :

$$C^{(A)}(u_1, u_2) := \begin{cases} \varphi^{\leftarrow}(\varphi(u_1) + \varphi(u_2)) & \text{si } \varphi(u_1) + \varphi(u_2) \leq \varphi(0), \\ 0 & \text{sinon.} \end{cases}$$

On peut montrer (exercice difficile) que la fonction $C^{(A)}$ est bien une copule. Il existe différents choix pour la fonction φ dont les principaux sont données dans le tableau ci-dessous.

Fonction φ	Paramètre(s)	Nom de la copule
$\varphi(t) = (t^{-\theta} - 1)/\theta$	$\theta \in [-1, \infty[$	Clayton
$\varphi(t) = (-\log(t))^\theta$	$\theta \in [1, \infty[$	Gumble
$\varphi(t) = -\log[(e^{-\theta t} - 1)/(e^{-\theta} - 1)]$	$\theta \in \mathbb{R}$	Frank

5.3.6 Copule min-max

La copule min-max est définie pour $(u_1, u_2) \in [0, 1]^2$ et $n \in \mathbb{N}^*$ par

$$C^{\text{mM}}(u_1, u_2) = u_2 - \max\{u_2^{1/n} - 1 + (1 - u_1)^{1/n}; 0\}.$$

Cette copule (qui n'est pas symétrique) est en fait la copule associée au couple $(X_{1,n}, X_{n,n})$ où $X_{1,n} \leq \dots \leq X_{n,n}$ sont les statistiques d'ordre associées aux variables X_1, \dots, X_n que l'on suppose indépendantes et de même fonction de répartition continue F . En effet, soit $(x_1, x_2) \in \mathbb{R}^2$,

$$G(x_1, x_2) := \mathbb{P}(X_{1,n} \leq x_1; X_{n,n} \leq x_2) = \mathbb{P}(X_{n,n} \leq x_2) - \mathbb{P}(X_{1,n} > x_1; X_{n,n} \leq x_2).$$

Ainsi,

$$G(x_1, x_2) = \begin{cases} F^n(x_2) - [F(x_2) - F(x_1)]^n & \text{si } x_1 \leq x_2, \\ F^n(x_2) & \text{si } x_1 > x_2 \end{cases}.$$

D'après le point *ii*) du Théorème de Sklar, on sait que la copule C associée au couple $(X_{1,n}, X_{n,n})$ est donnée par $C(u_1, u_2) = G(F_{1,n}^{\leftarrow}(u_1), F_{n,n}^{\leftarrow}(u_2))$, où $F_{1,n}$ et $F_{n,n}$ sont les fonctions de répartition de $X_{1,n}$ et $X_{n,n}$. On a donc $F_{1,n}(u_1) = 1 - (1 - F(u_1))^n$ et $F_{n,n}(u_2) = F^n(u_2)$ et ainsi,

$$\begin{aligned} C(u_1, u_2) &= G(F^{\leftarrow}(1 - (1 - u_1)^{1/n}), F^{\leftarrow}(u_2^{1/n})) \\ &= \begin{cases} u_2 - [u_2^{1/n} - 1 + (1 - u_1)^{1/n}]^n & \text{si } 1 - (1 - u_1)^{1/n} \leq u_2^{1/n}, \\ u_2 & \text{sinon,} \end{cases} \end{aligned}$$

qui est égale à la copule C^{mM} .

5.3.7 Copule extrême

Dans le cadre de ce cours sur la théorie des valeurs extrêmes, nous nous sommes intéressés pour l'instant au cas des variables aléatoires univariées. Nous allons ici introduire la notion d'extrême pour des vecteurs aléatoires. Soit $X = (X^{(1)}, \dots, X^{(p)})^t$ un vecteur aléatoire de fonction de répartition G . Notons X_1, \dots, X_n des copies indépendantes du vecteur aléatoire X et posons $M_n := (M_n^{(1)}, \dots, M_n^{(p)})^t$ avec, pour $j = 1, \dots, p$,

$$M_n^{(j)} = \max\{X_1^{(j)}; \dots; X_n^{(j)}\}.$$

Comme dans le cas univarié, on définit la notion de domaine d'attraction pour la fonction de répartition G .

Définition 19 Soit $H : \mathbb{R}^p \mapsto [0, 1]$ une fonction de répartition multivariée. On dit que G appartient au domaine d'attraction de H s'il existe des suites $a_n = (a_n^{(1)}, \dots, a_n^{(p)})^t$ et $b_n = (b_n^{(1)}, \dots, b_n^{(p)})^t$ (avec $a_n^{(j)} > 0$ pour tout $j = 1, \dots, p$) telles que pour tout $(x_1, \dots, x_p) \in \mathbb{R}^p$,

$$\lim_{n \rightarrow \infty} G^n(b_n^{(1)} + a_n^{(1)}x_1, \dots, b_n^{(p)} + a_n^{(p)}x_p) = H(x_1, \dots, x_p), \quad (5.2)$$

ou, de manière équivalente telles que

$$\left(\frac{M_n^{(1)} - b_n^{(1)}}{a_n^{(1)}}, \dots, \frac{M_n^{(p)} - b_n^{(p)}}{a_n^{(p)}} \right) \xrightarrow{\mathcal{L}} Y,$$

où $Y \in \mathbb{R}^p$ admet H pour fonction de répartition.

La fonction de répartition H est la fonction de répartition de la loi des valeurs extrêmes multivariée. D'après le Théorème 1, les marginales de H vérifient pour tout $j = 1, \dots, p$

$$\begin{aligned} H(+\infty, \dots, +\infty, x_j, +\infty, \dots, +\infty) &= H_{\gamma_j}(x_j) \\ &= \begin{cases} \exp[-(1 + \gamma_j x_j)_+^{-1/\gamma_j}] & \text{si } \gamma_j \neq 0, \\ \exp(-e^{-x_j}) & \text{si } \gamma_j = 0, \end{cases} \end{aligned}$$

où $(\gamma_1, \dots, \gamma_p) \in \mathbb{R}^p$ sont les p indices des valeurs extrêmes marginaux. Evidemment, ceci ne suffit pas pour caractériser la loi des valeurs extrêmes multivariée H . Il faut connaître la structure de dépendance et donc la copule associée. Pour ce faire on définit ci-après la notion de copules extrêmes et max-stable.

Définition 20 *On dit qu'une copule C est extrême si elle est associée à une fonction de répartition H satisfaisant (5.2). On dit qu'une copule C est max-stable si pour tout $n \in \mathbb{N}^*$ et $(u_1, \dots, u_p) \in [0, 1]^p$,*

$$C^n(u_1^{1/n}, \dots, u_p^{1/n}) = C(u_1, \dots, u_p).$$

Le résultat suivant (démontré en Annexe) montre qu'une copule extrême est une copule max-stable et inversement.

Proposition 19 *Une copule C est max-stable si et seulement si elle est extrême.*

En résumé, pour caractériser complètement la loi des valeurs extrêmes multivariée, il suffit donc de donner une caractérisation des copules max-stables. Elle est donnée par le théorème ci-dessous que l'on admettra.

Théorème 11 *Une copule $C : [0, 1]^p \mapsto [0, 1]$ est max-stable si et seulement si pour tout $(x_1, \dots, x_p) \in \mathbb{R}^p$ il existe une copule \tilde{C} telle que la limite ci-dessous existe :*

$$\lim_{t \downarrow 0} t^{-1}(1 - \tilde{C}(1 - tx_1, \dots, 1 - tx_p)) = \ell_{\tilde{C}}(x_1, \dots, x_p),$$

avec pour tout $(u_1, \dots, u_p) \in [0, 1]^p$

$$C(u_1, \dots, u_p) = \exp(-\ell_{\tilde{C}}(-\log u_1, \dots, -\log u_p)). \quad (5.3)$$

La fonction $\ell_{\tilde{C}}$ est appelée la fonction de dépendance stable. Il est facile de voir qu'une copule définie par (5.3) est max-stable mais la réciproque est plus difficile à démontrer. On admet aussi le résultat ci-dessous.

Proposition 20 *Si une fonction de répartition G de copule associée C_G appartient au domaine d'attraction de H (de copule max-stable associée C) alors C sera de la forme (5.3) avec $\tilde{C} = C_G$.*

Nous terminons ce paragraphe en donnant un exemple de copule max-stable : la copule de Gumbel définie par

$$C(u_1, \dots, u_p) = \exp \left\{ - \left[(-\log u_1)^\theta + \dots + (-\log u_p)^\theta \right]^{1/\theta} \right\},$$

où $\theta \in [1, \infty[$. Il est facile de voir que cette copule est max-stable et qu'elle admet pour fonction de dépendance stable

$$\ell_{\tilde{C}}(x_1, \dots, x_p) = (x_1^\theta + \dots + x_p^{1/\theta})^{1/\theta},$$

avec pour copule \tilde{C} la copule de Gumbel. Notons qu'en prenant $\theta = 1$, la copule de Gumbel est la copule d'indépendance Π . On dit alors qu'il y a indépendance asymptotique.

5.4 Mesures de dépendance et copules

Nous venons de montrer que la dépendance au sein d'un vecteur aléatoire est modélisée par une copule. Il est donc naturel de définir des mesures de dépendance basées sur ces copules. Nous nous intéresserons dans un premier temps aux mesures dites de concordance-discordance dans le cas $p = 2$.

Définition 21 Soient $X_1 = (X_1^{(1)}, X_1^{(2)})^t$ et $X_2 = (X_2^{(1)}, X_2^{(2)})^t$ deux vecteurs aléatoires indépendants de fonctions de répartition respectives G_1 et G_2 (on notera C_1 et C_2 les copules associées). La mesure de concordance-discordance entre les couples X_1 et X_2 est définie par :

$$\mathcal{Q}(X_1, X_2) := \mathbb{P}[(X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) > 0] - \mathbb{P}[(X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) < 0].$$

La quantité $\mathbb{P}[(X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) > 0]$ est la probabilité que les vecteurs X_1 et X_2 soient concordants c'est-à-dire que leurs composantes évoluent dans le même sens (si la première composante augmente (ou diminue) alors la seconde augmente (ou diminue) également). A l'inverse, le nombre $\mathbb{P}[(X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) < 0]$ est la probabilité que les vecteurs soient discordants. Comme le montre le résultat ci-dessous, la mesure $\mathcal{Q}(X_1, X_2)$ dépend uniquement des copules C_1 et C_2 .

Proposition 21 Soient $X_1 = (X_1^{(1)}, X_1^{(2)})^t$ et $X_2 = (X_2^{(1)}, X_2^{(2)})^t$ deux vecteurs aléatoires indépendants de fonctions de répartition respectives G_1 et G_2 (on notera C_1 et C_2 les copules associées). Si les vecteurs X_1 et X_2 admettent les mêmes lois marginales que l'on suppose continues alors,

$$\mathcal{Q}(X_1, X_2) = 4 \int_0^1 \int_0^1 C_2(u, v) dC_1(u, v) - 1. \quad (5.4)$$

Rappelons que si G est la fonction de répartition d'un vecteur aléatoire $X \in \mathbb{R}^p$ défini sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ alors, pour toute fonction f telle que $f(X)$ est intégrable,

$$\mathbb{E}(f(X)) = \int_{\Omega} f(X) d\mathbb{P} = \int_{\mathbb{R}^p} f(x) \mathbb{P}_X(dx) = \int_{\mathbb{R}^p} f(x) dF(x).$$

Ainsi, pour un vecteur aléatoire (U, V) de fonction de répartition C_1 ,

$$\mathcal{Q}(X_1, X_2) = 4\mathbb{E}(C_2(U, V)) - 1.$$

Enfin, bien qu'en apparence l'ordre entre X_1 et X_2 semble important dans (5.4), il faut bien se rappeler que $\mathcal{Q}(X_1, X_2) = \mathcal{Q}(X_2, X_1)$.

Nous allons à présent considérer deux cas particuliers de mesures de concordance-discordance permettant de mesurer la dépendance entre deux variables aléatoires.

5.4.1 Tau de Kendall

Définition 22 Soit $X = (X^{(1)}, X^{(2)})^t$ un vecteur aléatoire de copule associée C , les marges étant supposés continues. Le Tau de Kendall de X noté $\tau(X)$, est la mesure de concordance-discordance entre deux copies indépendantes X_1 et X_2 de X . Autrement dit,

$$\tau(X) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \in [-1, 1].$$

Le Tau de Kendall mesure donc la dépendance entre $X^{(1)}$ et $X^{(2)}$. Si la valeur du Tau de Kendall est proche de 0, la probabilité que deux copies indépendantes de X soient concordantes sera approximativement égale à la probabilité qu'elles soient discordantes. Donc, intuitivement, une valeur du Tau de Kendall proche de zéro signifie que $X^{(1)}$ et $X^{(2)}$ sont proche de l'indépendance. Il est effectivement facile de vérifier que si $C(u, v) = uv$ alors $\tau(X) = 0$. Inversement une valeur de $\tau(X)$ proche de 1 (resp. -1) signifie que les deux copies ont une forte probabilité d'être concordantes (resp. discordantes) et donc que $X^{(2)}$ est approximativement une fonction croissante (resp. décroissante) de $X^{(1)}$.

Nous nous intéressons à présent à l'estimation du Tau de Kendall. Considérons des copies indépendantes X_1, \dots, X_n d'un vecteur aléatoire X . L'estimateur naturel de $\tau(X)$ consiste à compter parmi les $C_n^2 = n(n-1)/2$ paires (X_i, X_j) disponibles, le nombre de paires concordantes et le nombre de paires discordantes. L'estimateur s'écrit :

$$\hat{\tau}_n(X) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\mathbb{I}_{\{(X_i^{(1)} - X_j^{(1)})(X_i^{(2)} - X_j^{(2)}) > 0\}} - \mathbb{I}_{\{(X_i^{(1)} - X_j^{(1)})(X_i^{(2)} - X_j^{(2)}) < 0\}} \right).$$

5.4.2 Rho de Spearman

Pour un vecteur aléatoire $X = (X^{(1)}, X^{(2)})^t$, le Rho de Spearman est, à une constante de normalisation près, la mesure de concordance-discordance entre X et un vecteur aléatoire Y de fonction de répartition $G(x_1, x_2) = F_1(x_1)F_2(x_2)$ où F_1 et F_2 sont les fonctions de répartition marginales de X (autrement dit, les marges de Y sont indépendantes). Plus précisément, on a la définition suivante.

Définition 23 Soit $X = (X^{(1)}, X^{(2)})^t$ un vecteur aléatoire de copule associée C , les marges étant supposés continues. Le Rho de Spearman de X noté $\rho(X)$ est défini par

$$\rho(X) = 3 \times \left(4 \int_0^1 \int_0^1 uv dC(u, v) - 1 \right) \in [-1, 1]$$

Ainsi, $\rho(X)$ est 3 fois la mesure de discordance entre le vecteur X de copule C et un vecteur Y de copule la copule d'indépendance Π . Le facteur 3 assure que $\rho(X)$ est dans l'intervalle $[-1, 1]$. Pour s'en rendre compte, il suffit de remarquer que si (U, V) est un couple aléatoire de fonction de répartition C ,

$$\rho(X) = 12\mathbb{E}(UV) - 3 = \frac{\mathbb{E}(UV) - 1/4}{1/12} = \frac{\text{Cov}(U, V)}{(\text{Var}(U)\text{Var}(V))^{1/2}}, \quad (5.5)$$

car, U et V étant de loi uniforme sur $[0, 1]$, $\mathbb{E}(U) = \mathbb{E}(V) = 1/2$ et $\text{Var}(U) = \text{Var}(V) = 1/12$. Donc, $\rho(X)$ est simplement le coefficient de corrélation linéaire entre U et V qui appartient bien à $[-1, 1]$.

Soient X_1, \dots, X_n des copies indépendantes du vecteur X . Dans la littérature, le Rho de Spearman est très fréquemment estimé par

$$\hat{\rho}_n(X) = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R(X_i^{(1)}) - R(X_i^{(2)}))^2,$$

où, pour un ensemble $\{y_1, \dots, y_n\}$ de réels, on désigne pour tout $i = 1, \dots, n$ par $R(y_i)$ le rang (la position) de y_i dans l'ensemble ordonné $\{y_{1,n} \leq \dots, \leq y_{n,n}\}$.

L'expression (5.5) permet de justifier l'expression de cet estimateur. L'idée consiste à estimer le coefficient de corrélation entre U et V (où (U, V) admet C pour fonction de répartition). Pour estimer ce coefficient, il faudrait dans l'idéal disposer de réalisations du vecteur aléatoire (U, V) ce qui n'est évidemment pas le cas. Cependant, d'après le Théorème de Sklar, on sait qu'en posant $U_i := F_1(X_i^{(1)})$ et $V_i := F_2(X_i^{(2)})$, $(U_1, V_1), \dots, (U_n, V_n)$ sont des vecteurs aléatoires indépendants de fonction de répartition C . Les marginales F_1 et F_2 étant inconnues, on les remplace par les versions empiriques définies pour tout $x \in \mathbb{R}$ par :

$$\hat{F}_{n,1}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i^{(1)} \leq x\}} \quad \text{et} \quad \hat{F}_{n,2}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i^{(2)} \leq x\}}.$$

Il est alors facile de remarquer que pour tout $i = 1, \dots, n$,

$$\left(\hat{U}_{n,i}, \hat{V}_{n,i}\right) := \left(\hat{F}_{n,1}(X_i^{(1)}), \hat{F}_{n,2}(X_i^{(2)})\right) = \left(\frac{R(X_i^{(1)})}{n}, \frac{R(X_i^{(2)})}{n}\right).$$

L'estimateur du Rho de Spearman est basé sur ces statistiques.

Proposition 22 *L'estimateur $\hat{\rho}_n(X)$ est le coefficient de corrélation linéaire empirique liant les séries $\hat{U} = (\hat{U}_{n,1}, \dots, \hat{U}_{n,n})$ et $\hat{V} = (\hat{V}_{n,1}, \dots, \hat{V}_{n,n})$.*

5.4.3 Mesure de la dépendance extrême

En théorie des valeurs extrêmes, on s'intéresse aux grandes observations c'est-à-dire à la queue de distribution. Disposant d'un vecteur aléatoire, il peut être utile d'avoir une information quantitative sur la dépendance entre les grandes observations des composantes du vecteur. Pour mesurer ce type de dépendance, on va utiliser la notion de copule de survie.

Définition 24 *Soit $C : [0, 1]^p \mapsto [0, 1]$ une copule. La copule de survie \bar{C} associée à C est définie pour $(u_1, \dots, u_p) \in [0, 1]^p$ par*

$$\bar{C}(u_1, \dots, u_p) = \mathbb{P}(U_1 > 1 - u_1, \dots, U_p > 1 - u_p),$$

où le vecteur aléatoire $(U_1, \dots, U_p)^t$ admet C pour fonction de répartition.

Une autre façon de définir la copule de survie est de dire que si un vecteur aléatoire $(U_1, \dots, U_p)^t$ admet C pour fonction de répartition, la copule de survie \bar{C} est la fonction de répartition de $\bar{U} := (1 - U_1, \dots, 1 - U_p)^t$ qui définit bien une copule puisque les marges de \bar{U} sont de loi uniforme sur $[0, 1]$.

Il est bien sûr possible d'écrire la copule de survie \bar{C} en fonction de C mais l'expression devient relativement complexe pour $p > 2$. Dans le cas $p = 2$, il est facile de remarquer que $\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$. La dépendance extrême peut être mesurée à l'aide de la fonction de dépendance extrême.

Définition 25 *La fonction de dépendance extrême associée à une copule C est définie pour $(x_1, \dots, x_p) \in \mathbb{R}^p$ par la limite suivante (si elle existe) :*

$$\Lambda_C(x_1, \dots, x_p) := \lim_{t \downarrow 0} \frac{1}{t} \bar{C}(tx_1, \dots, tx_p).$$

Il existe bien sûr un lien entre la fonction de dépendance Λ_C et la fonction de dépendance stable ℓ_C associée à la copule C . Dans le cas simple $p = 2$, on remarque que $\Lambda_C(x_1, x_2) = x_1 + x_2 - \ell_C(x_1, x_2)$. En pratique, pour mesurer la dépendance extrême, on se restreint souvent au calcul du coefficient de dépendance extrême.

Définition 26 *Le coefficient de dépendance extrémal $\lambda_C \in [0, 1]$ d'une copule C est définie par $\lambda_C = \Lambda_C(1, \dots, 1)$.*

Remarquons que si $X = (X^{(1)}, \dots, X^{(p)})^t$ est un vecteur aléatoire de copule C ,

$$\lambda_C = \lim_{\alpha \downarrow 0} \mathbb{P} \left(X^{(i)} > F_i^{\leftarrow}(1 - \alpha), i \neq j | X^{(j)} > F_j^{\leftarrow}(1 - \alpha) \right).$$

Cette quantité ne dépend pas du choix de j . L'égalité précédente montre que le coefficient extrémal est en fait la probabilité que les composantes dépassent un seuil tendant vers leur point terminal sachant qu'une des composantes dépassent un seuil (lui aussi convergeant vers le point terminal correspondant). Ainsi, une valeur de λ_C égale à 0 signifie que les valeurs extrêmes des composantes sont indépendantes (on parle d'indépendance asymptotique).

Pour terminer, calculons la valeur du coefficient de dépendance extrémal pour deux copules classiques dans le cas bivarié. Si C est la copule de Gumbel définie pour $\theta > 1$ par $C(u_1, u_2) = \exp\{-[(-\log u_1)^\theta + (-\log u_2)^\theta]^{1/\theta}\}$, on montre que $\lambda_C = 2 - 2^{1/\theta}$. Ainsi, si $\theta = 1$, il y a indépendance asymptotique ce qui était attendu puisque dans ce cas $C = \Pi$.

En prenant pour C la copule de Farlie-Gumble-Morgenstern définie pour $\theta \in [-1, 1]$ par $C(u_1, u_2) = u_1 u_2 + \theta u_1 u_2 (1 - u_1)(1 - u_2)$ on a $\lambda_C = 0$. Ainsi, si un vecteur $X = (X^{(1)}, X^{(2)})^t$ est associé à la copule de Farlie-Gumble-Morgenstern, les variables $X^{(1)}$ et $X^{(2)}$ ne sont pas indépendantes (sauf si $\theta = 0$) et pourtant leurs valeurs extrêmes peuvent être considérées comme indépendantes.

5.5 Estimation non paramétrique d'une copule

Au vu des nombreuses utilisation des copules, il paraît important de pouvoir estimer la copule C associée à un vecteur aléatoire X lorsque l'on dispose de n copies indépendantes X_1, \dots, X_n de X . Pour ce faire, on se base sur le résultat annonçant que

$$C(u_1, \dots, u_p) = \mathbb{P}(F_1(X^{(1)}) \leq u_1, \dots, F_p(X^{(p)}) \leq u_p),$$

où F_1, \dots, F_p sont les fonctions de répartition marginales que l'on suppose continues. En remplaçant G, F_1, \dots, F_p par les estimateurs empiriques classiques, l'estimateur non paramétrique de la copule C est donné par :

$$\hat{C}_n(u_1, \dots, u_p) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{R(X_i^{(1)}) \leq nu_1, \dots, R(X_i^{(p)}) \leq nu_p\}}.$$