

DIPLÔME D'INGÉNIEUR GÉNÉRALISTE
PROJET PRATIQUE D'OPTIONS MATHÉMATIQUES APPLIQUÉES

Théorie des valeurs extrêmes avec applications environnementales ou en finance

Romane TÉZÉ
Elyas BENYAMINA
Zihao GUO

Tuteur :
MATHIEU Ribatet
mathieu.ribatet@ec-nantes.fr

May 25, 2023

CONTENTS

| | | |
|------------|---|-----------|
| I | Introduction | 2 |
| II | Outils mathématiques | 4 |
| 1 | Modélisation paramétrique | 4 |
| III | Théorie Classique des Valeurs Extremes et Modélisation | 11 |
| 2 | Modèles asymptotiques | 11 |
| 3 | Estimation des paramètres de la distribution GEV | 17 |
| IV | Applications | 22 |
| 4 | Introduction | 22 |
| V | Perspectives et conclusions | 23 |
| 5 | Perspectives | 23 |
| VI | Annexe | 25 |
| A | Code de programmation | 25 |

ABSTRACT

La théorie des valeurs extrêmes, en tant que discipline statistique, est unique dans la mesure où elle développe des techniques et des modèles pour décrire des situations inhabituelles plutôt qu'habituelles. En tant qu'étude abstraite des phénomènes aléatoires, le sujet remonte au début du XXe siècle. Ce n'est que dans les années 1950 que des méthodes de modélisation des phénomènes physiques réels ont été sérieusement proposées. Cette théorie de la valeur limite a fourni un cadre pour l'estimation des valeurs limites attendues en utilisant des données historiques.

KEYWORDS Modélisation statistique, théorie des valeurs extrêmes

Part I.

Introduction

L'objectif de ce document est de rendre compte des connaissances acquises et comprises lors de l'étude du livre *An Introduction to Statistical Modeling of Extremes Values*, de Stuart Coles, publié en 2001 [1]. La structure de ce rapport suit la structure du livre jusqu'au chapitre 3 environ. Nous nous sommes permis d'ignorer certaines sections qui nous semblaient d'un intérêt moindre. De même, nous avons étayé d'autres de détails, d'exemples ou encore de démonstrations supplémentaires, pouvant être issus de nouvelles sources.

L'origine de la théorie des valeurs extrêmes remonte aux années 1920, lorsque l'anglais Leonard Tippett, qui travaille alors dans l'industrie textile, réalise que la résistance d'une pièce textile dépend de la résistance de ses plus faibles fibres. Avec l'aide de Ronald Fisher, il établit l'existence de trois limites asymptotiques pour décrire la distribution des valeurs extrêmes, dans le cas où les variables aléatoires sont indépendantes.

Dans les années 1940, le mathématicien russe Boris Gnedenko développe et complète cette théorie. Mais c'est Gumbel, en 1958, qui pose les bases solides de la théorie des valeurs extrêmes en unifiant toutes les approches précédentes. Il est aujourd'hui considéré comme le père de la théorie des valeurs extrêmes. Depuis lors, et dû aux très nombreuses possibilités d'applications de la théorie des valeurs extrêmes, celle-ci s'est grandement développée. Le livre de Stuart Coles, publié en 2001, va ainsi plus loin que la théorie classique, et s'intéresse également au cas des variables aléatoires dépendantes, non-stationnaires, ou encore aux extrêmes multivariés.

La théorie des valeurs extrêmes a pour objectif de décrire le comportement statistique du maximum ou du minimum des occurrences d'un événement. C'est-à-dire que si l'on représente cet événement par une variable aléatoire X_i , dont on possède n réalisations indépendantes et identiquement distribuées et que l'on note:

$$M_n = \max(X_1, \dots, X_n) \quad (1)$$

Alors l'objectif de la théorie des valeurs extrêmes est de modéliser le comportement de M_n , c'est-à-dire le comportement du maximum sur n réalisations de l'événement. Dans la plupart des cas, la distribution des variables aléatoires X_i est inconnue. On utilise alors des arguments asymptotiques, et pour des valeurs suffisamment grandes de n , la théorie classique des valeurs extrêmes nous donne des distributions limites pour M_n . Cette distribution peut ensuite être utilisée pour obtenir la probabilité que l'événement considéré prenne une valeur encore plus grande (ou plus petite) que ce qui a jamais été observé, impliquant ainsi l'hypothèse que les événements modélisés ont un comportement statistique suffisamment "régulier" pour pouvoir être extrapolés. Malgré ce défaut, la théorie classique des valeurs extrêmes reste adaptée et largement utilisée.

La méthode d'estimation des paramètres du modèle de M_n choisie par Stuart Coles est celle du maximum de vraisemblance, pour ses propriétés très pratiques. Il insiste également sur l'importance de la mesure d'incertitude sur les paramètres. En effet, la théorie des valeurs extrêmes contient un grand nombre de sources d'incertitude. Par exemple,

comme dit précédemment, les paramètres estimés vont être utilisés pour l'extrapolation à des niveaux non observés, et l'incertitude sur les paramètres estimés aura un impact important sur l'incertitude des niveaux extrapolés.

Les applications de la théorie des valeurs extrêmes sont très nombreuses. En finance, elle est utilisée pour modéliser la *Value at risk* [2]. Il s'agit d'un indicateur pour mesurer le risque financier d'une société ou encore d'un portefeuille. La *Value at risk* estime le potentiel de perte et la probabilité que cela se produise. On cherche donc à modéliser, à l'aide des valeurs extrêmes, la distribution du risque maximal et les probabilités associées. En environnement, la théorie des valeurs extrêmes est par exemple utilisée pour modéliser les niveaux maximaux que peut atteindre la mer à certains endroits. La distribution obtenue peut ainsi être utilisée dans le dimensionnement d'un mur de protection.

Dans la suite de ce rapport, nous détaillerons donc les outils mathématiques qui ont présenté un caractère nouveau à nos yeux, puis nous restituerons les connaissances acquises sur la Théorie Classique des Valeurs Extrêmes.

Part II.

Outils mathématiques

1 MODÉLISATION PARAMÉTRIQUE

1.1 Introduction

L'objectif de la modélisation statistique est de pouvoir estimer les paramètres de la loi de probabilité dont sont issues les données. On suppose donc qu'on connaît une famille de loi dans laquelle la loi qui distribue les données appartient.

Cette famille peut être déterminée à partir de considérations physiques, à partir de certaines techniques empiriques ou bien encore être déterminée asymptotiquement, via des lois limites (à l'aide du Théorème Centrale Limite par exemple). Dans la suite, on suppose que l'on dispose de n réalisations indépendantes x_1, \dots, x_n d'une variable aléatoire X dont la densité, continue, appartient à une famille de loi de paramètre θ . C'est ce paramètre que l'on cherche à estimer dans la suite, dont la vraie valeur sera notée θ_0 .

1.2 Estimateur du Maximum de Vraisemblance

La **fonction de vraisemblance** de la variable aléatoire X , de densité $f(x, \theta)$ est définie par :

$$L(x, \theta) = L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (2)$$

La vraisemblance $L(x_1, \dots, x_n; \theta)$ correspond à la probabilité que la variable aléatoire X prenne effectivement l'ensemble des valeurs : x_1, \dots, x_n . La vraisemblance change lorsque le paramètre inconnu à estimer, θ , prend des valeurs différentes. Si la vraisemblance est faible, pour une valeur de θ donnée, alors la probabilité d'observer les données à disposition est faible, et le paramètre θ testé peut être remis en cause.

Il faut donc choisir le θ qui donne la valeur maximale de la vraisemblance $L(x_1, \dots, x_n; \theta)$ dans la plage de valeurs du paramètre inconnu θ , comme estimation du paramètre inconnu. θ_0 est l'estimateur du maximum de vraisemblance du paramètre inconnu θ_0 .

On utilisera souvent la fonction log-vraisemblance pour faciliter les calculs :

$$\ln(L(x_1, \dots, x_n; \theta)) = \sum_{i=1}^n \ln(f(x_i; \theta))$$

1.3 Comportement asymptotique de l'estimateur du Maximum de Vraisemblance

L'estimateur du maximum de vraisemblance présente certaines propriétés qui le rendent très pratique à utiliser. Notamment, il possède des propriétés asymptotiques particulièrement agréables.

THÉORÈME On note x_1, \dots, x_n des réalisations indépendantes d'une variable aléatoire X dont la densité appartient au modèle statistique F de paramètre θ de dimension d . On note $l(\cdot)$ la log-likelihood et $\hat{\theta}_0$ l'estimateur du maximum de vraisemblance de θ_0 . Sous certaines conditions de régularité, et pour n suffisamment grand,

$$\hat{\theta}_0 \sim \mathcal{N}(\theta_0, I_E(\theta_0)^{-1})$$

où $I_E(\theta)$ est la *matrice d'information de Fisher* et mesure la courbure de la surface de la log-likelihood au voisinage de θ .

$$I_E(\theta) = \begin{pmatrix} e_{1,1}(\theta) & \cdot & \cdot & e_{1,d}(\theta) \\ \cdot & \cdot & e_{i,j}(\theta) & \cdot \\ \cdot & e_{j,i}(\theta) & \cdot & \cdot \\ e_{d,1}(\theta) & \cdot & \cdot & e_{d,d}(\theta) \end{pmatrix}$$

avec

$$e_{i,j} = \mathbb{E}\left(-\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta)\right),$$

l'information de Fisher. La valeur de θ_0 étant inconnue, on utilise en pratique la *matrice observée d'information de Fisher* $I_O(\theta)$ définie par :

$$I_O(\theta) = \begin{pmatrix} -\frac{\partial^2}{\partial \theta_1^2} l(\theta) & \cdot & \cdot & -\frac{\partial^2}{\partial \theta_1 \partial \theta_d} l(\theta) \\ \cdot & \cdot & -\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta) & \cdot \\ \cdot & -\frac{\partial^2}{\partial \theta_j \partial \theta_i} l(\theta) & \cdot & \cdot \\ -\frac{\partial^2}{\partial \theta_d \partial \theta_1} l(\theta) & \cdot & \cdot & -\frac{\partial^2}{\partial \theta_d^2} l(\theta) \end{pmatrix}$$

et évaluée en $\theta = \hat{\theta}_0$.

Ainsi, pour chaque estimateur de chaque composante θ_i de θ_0 , on en déduit :

$$\hat{\theta}_i \sim \mathcal{N}(\theta_i, \psi_{i,i})$$

avec $\psi_{i,j}$ les coefficients de l'inverse de la matrice (observée) d'information de Fisher.

Ce théorème permet ensuite d'obtenir des intervalles de confiance pour chaque composante de l'estimateur du maximum de vraisemblance en utilisant les quantiles de la loi normale.

DÉMONSTRATION [3] Commençons par définir la log-likelihood normalisée, c'est-à-dire :

$$L_n(\theta) = \frac{1}{n} \log(f_X(X, \theta))$$

On a donc :

$$L'_n(\theta) = \frac{\partial}{\partial \theta} \left(\frac{1}{n} \log(f_X(X, \theta)) \right)$$

$$L''_n(\theta) = \frac{\partial^2}{\partial \theta^2} \left(\frac{1}{n} \log(f_X(X, \theta)) \right)$$

Or, $\hat{\theta}_0$ est par définition la valeur de θ qui maximise la log-likelihood, soit $L'_n(\hat{\theta}_0) = 0$.

En appliquant le théorème des accroissements finis à la fonction L'_n , et en considérant un θ_n compris entre θ_0 et $\hat{\theta}_0$, on obtient

$$L'_n(\hat{\theta}_0) = L'_n(\theta_0) + L''_n(\theta_n)(\hat{\theta}_0 - \theta_0).$$

Soit

$$\hat{\theta}_0 - \theta_0 = -\frac{L'_n(\theta_0)}{L''_n(\theta_n)}$$

car $L'_n(\hat{\theta}_0) = 0$.

Multiplions par \sqrt{n} et étudions $\sqrt{n}L'_n(\theta_0)$.

$$\begin{aligned} \sqrt{n}L'_n(\theta_0) &= \sqrt{n} \left(\frac{1}{n} \left(\frac{\partial}{\partial \theta} \log(f_X(X, \theta_0)) \right) \right) \\ &= \sqrt{n} \left(\frac{1}{n} \left(\frac{\partial}{\partial \theta} \log \left(\prod f_X(x_i, \theta_0) \right) \right) \right) \\ &= \sqrt{n} \left(\frac{1}{n} \sum \left(\frac{\partial}{\partial \theta} \log(f_X(x_i, \theta_0)) \right) \right) \\ &= \sqrt{n} \left(\frac{1}{n} \sum \left(\frac{\partial}{\partial \theta} \log(f_X(x_i, \theta_0)) \right) - \mathbb{E} \left(\frac{\partial}{\partial \theta} \log f_X(x_1, \theta_0) \right) \right) \end{aligned}$$

car $\mathbb{E} \left(\frac{\partial}{\partial \theta} \log f_X(x_1, \theta_0) \right) = 0$. Cela vient du fait que la valeur attendue de la fonction score (dérivée de la log-likelihood) vaut zéro (propriété ici admise).

En appliquant ensuite le Théorème Central Limite sur la dernière ligne (les x_i étant i.i.d), on en déduit :

$$\sqrt{n}L'_n(\theta_0) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, \mathbb{V} \left(\frac{\partial}{\partial \theta} \log(f_X(x_1, \theta_0)) \right))$$

Or on sait que la variance de la dérivée de la log-likelihood (fonction score) vaut :

$$\mathbb{V} \left(\frac{\partial}{\partial \theta} \log(f_X(x_1, \theta_0)) \right) = \mathbb{E} \left(\left(\frac{\partial}{\partial \theta} \log(f_X(x_1, \theta_0)) \right)^2 \right) - \left(\mathbb{E} \left(\frac{\partial}{\partial \theta} \log(f_X(x_1, \theta_0)) \right) \right)^2.$$

Or le dernier terme vaut 0, comme expliqué plus haut, d'où

$$\mathbb{V} \left(\frac{\partial}{\partial \theta} \log(f_X(x_1, \theta_0)) \right) = I(\theta_0)$$

où $I(\theta_0)$ désigne l'information de Fisher de θ_0 (pour une observation).

Etudions à présent $L''_n(\theta_n)$. D'abord,

$$\begin{aligned}
L_n''(\theta) &= \frac{1}{n} \left(\frac{\partial^2}{\partial \theta^2} \log(f_X(X, \theta)) \right) \\
&= \frac{1}{n} \left(\frac{\partial^2}{\partial \theta^2} \log\left(\prod f_X(x_i, \theta)\right) \right) \\
&= \frac{1}{n} \sum \left(\frac{\partial^2}{\partial \theta^2} \log(f_X(x_i, \theta)) \right) \\
&\xrightarrow{p} \mathbb{E} \left(\frac{\partial^2}{\partial \theta^2} \log(f_X(x_1, \theta)) \right)
\end{aligned}$$

d'après la loi faible des grands nombres. Or nous avons pris θ_n dans $(\hat{\theta}_0, \theta_0)$ et on suppose que $\hat{\theta}_0 \xrightarrow{p} \theta_0$. On peut donc écrire

$$L_n''(\theta_n) \xrightarrow{p} \mathbb{E} \left(\frac{\partial^2}{\partial \theta^2} \log(f_X(x_1, \theta_0)) \right) = -I(\theta_0)$$

A présent, on utilise le théorème de Slutsky et on obtient

$$\sqrt{n}(\hat{\theta}_0 - \theta_0) \sim -\frac{1}{I(\theta_0)} \mathcal{N}(0, I(\theta_0))$$

c'est-à-dire

$$\sqrt{n}(\hat{\theta}_0 - \theta_0) \sim \mathcal{N}(0, \frac{1}{I(\theta_0)}).$$

Soit

$$\hat{\theta}_0 \sim \mathcal{N}(\theta_0, \frac{1}{nI(\theta_0)}) = \mathcal{N}(\theta_0, I_E(\theta_0)^{-1}),$$

car l'information de Fisher pour n observations est bien égale à n fois l'information de Fisher pour une information. On obtient ainsi le résultat souhaité.

Un autre avantage de l'estimateur du maximum de vraisemblance est sa stabilité par composition avec une fonction.

THÉORÈME Soit $\hat{\theta}_0$ un estimateur du maximum de vraisemblance de θ_0 . Soit $\phi = g(\theta)$ une fonction scalaire de θ . Alors l'estimateur du maximum de vraisemblance de $\phi_0 = g(\theta_0)$ est $\hat{\phi}_0 = g(\hat{\theta}_0)$.

Les propriétés de normalité asymptotiques sont aussi valables pour $\hat{\phi}_0$.

THÉORÈME (DELTA-MÉTHODE) Soit $\hat{\theta}_0$ un estimateur du maximum de vraisemblance de θ , de matrice de variance-covariance V_θ . Alors si $\phi = g(\theta)$ est une fonction scalaire, alors $\hat{\phi}_0 = g(\hat{\theta}_0)$, l'estimateur du maximum de vraisemblance de $g(\theta_0)$, satisfait

$$\hat{\phi}_0 \sim \mathcal{N}(\phi_0, V_\phi)$$

avec

$$V_\phi = \Delta \phi^T V_\theta \Delta \phi$$

et

$$\Delta \phi = \left[\frac{\partial \phi}{\partial \theta_1}, \dots, \frac{\partial \phi}{\partial \theta_d} \right]^T$$

évalué en $\theta = \hat{\theta}_0$.

1.4 Déviance

La fonction de déviance permet également de quantifier l'incertitude sur l'estimateur du maximum de vraisemblance. Elle est définie par :

$$D(\theta) = 2(l(\hat{\theta}_0) - l(\theta)).$$

Plus l'estimateur du maximum de vraisemblance θ_0 se rapproche du vrai paramètre du modèle, plus leurs log-likelihood respectives seront proches, et plus la déviance sera petite. On cherche donc à obtenir une déviance faible. Autrement dit, si la déviance est trop grande, il y a peu de chance que l'on ait réussi à correctement approcher θ_0 par $\hat{\theta}_0$. On peut donc définir des régions de confiance de la forme

$$C = \{\theta : D(\theta) \leq c\}$$

Idéalement, si on parvient à déterminer c tel qu'on puisse contrôler la probabilité que θ_0 appartienne à C , on pourrait obtenir des intervalles de confiance de niveau $1 - \alpha$. En pratique, cela nécessite de calculer la log-likelihood pour tout θ et donc de connaître la distribution réelle des données, ce qui n'est généralement pas le cas. On peut utiliser le théorème suivant :

THÉORÈME Soit (x_1, \dots, x_n) des réalisations indépendantes d'une même variable aléatoire X . Soit $\hat{\theta}_0$ l'estimateur du maximum de vraisemblance du vrai paramètre θ_0 de la loi de X . Pour n grand, et sous des conditions de régularité,

$$D(\theta_0) \sim \chi_d^2$$

où d est la dimension de θ_0 .

On peut donc choisir la constance c selon les quantiles de la loi χ_d^2 pour obtenir des intervalles de confiance de niveau souhaité.

1.5 Profile Likelihood Function

Supposons θ de dimension d . Il existe une autre méthode pour trouver un estimateur de θ . On appelle **profile log-likelihood** la quantité suivante :

$$l_p(\theta_i) = \max_{\theta_{-i}} l(\theta_i, \theta_{-i}),$$

où θ_i désigne la i ème composante de θ et θ_{-i} désigne le vecteur des composantes privé de la i ème.

Ainsi, en fixant une valeur pour θ_i , on calcule les valeurs des autres paramètres qui maximisent la log-likelihood avec le θ_i choisi. En répétant cela pour différentes valeurs de θ_i , on peut obtenir le graphique du profil de la fonction log-likelihood selon θ_i , et choisir pour θ_i la valeur qui maximise la vraisemblance.

MÉTHODE POUR TROUVER LE MAXIMUM DE VRAISEMBLANCE EN UTILISANT LA PROFILE LOG-LIKELIHOOD Considérons un problème de statistique où le paramètre à estimer θ (de dimension d) est composé de paramètres d'intérêt α et de paramètres de nuisance β , de sorte que $\theta = (\alpha, \beta)$.

On considère tout d'abord les variables aléatoires x_i pour $i \in 1, \dots, n$ indépendantes identiquement distribuées suivant toutes la même loi dépendante de $\theta = (\alpha, \beta)$. On suppose que cette loi est à densité et admet pour densité, $x \mapsto f_\theta(x) = f_{(\alpha, \beta)}(x)$.

La log-vraisemblance s'écrit alors : $\mathcal{L}_n(\alpha, \beta) = \sum_{i=1}^n \log f(x_i; \alpha, \beta)$.
On fixe dans un premier temps les paramètres d'intérêt α , et on maximise, $\hat{\beta}_\alpha = \arg \max_\beta \mathcal{L}_n(\alpha, \beta)$, avec $\mathcal{L}_n(\alpha, \beta)$ qui dépend de α .

On maximise ensuite $\hat{\beta}_\alpha$ selon α , et l'on note $\hat{\alpha}_n$, un maximiseur de $\hat{\beta}_\alpha$.

On admet que le l'estimateur du maximum de vraisemblance est alors $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_{\hat{\alpha}_n})$
Considérons l'exemple suivant :

DISTRIBUTION DE WEIBULL On considère les variables aléatoires x_i pour $i \in 1, \dots, n$ indépendantes identiquement distribuées suivant toutes la même loi de densité

$$f(x; \alpha, \beta) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} \exp(-(x/\beta)^\alpha).$$

En fixant α , on a :

$$\begin{aligned} \hat{\beta}_\alpha &= \arg \max_\beta \mathcal{L}_\alpha(X, \beta) \\ &= \arg \max_\beta \sum_{i=1}^n \left(\log \alpha + (\alpha - 1) \log x_i - \alpha \log \beta - \left(\frac{x_i}{\beta} \right)^\alpha \right) \\ &= \arg \max_\beta \sum_{i=1}^n \left(-\alpha \log \beta - \left(\frac{x_i}{\beta} \right)^\alpha \right) = \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha \right)^{1/\alpha} \end{aligned}$$

Puis, en maximisant $\mathcal{L}_\alpha(X, \alpha)$ selon α , on trouve

$$\hat{\alpha}_n = \arg \max_\alpha \sum_{i=1}^n \left(\log \alpha + (\alpha - 1) \log x_i - \alpha \log \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha \right)^{1/\alpha} - \left(\frac{x_i}{\left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha \right)^{1/\alpha}} \right)^\alpha \right)$$

Et enfin, on a l'estimateur du maximum de vraisemblance de β qui est, $\hat{\beta}_n = \left(\frac{1}{n} \sum_{i=1}^n x_i^{\hat{\alpha}_n} \right)^{1/\hat{\alpha}_n}$

Finalement, on obtient $\hat{\theta}_0 = (\hat{\alpha}_n, \hat{\beta}_n)$, une estimation par maximum de vraisemblance via la méthode du *profile log-likelihood* pour la distribution de Weibull.

DÉVIANCE STATISTIQUE ET THÉORÈME DE WILKS On reprend les notations précédentes, et on note $\theta_0 = (\alpha_0, \beta_0)$, les vrais paramètres, dont l'estimateur du maximum de vraisemblance par la méthode de la fonction *Profile Likelihood* sont notés $\hat{\theta}_0 = (\hat{\alpha}, \hat{\beta})$. La déviance s'exprime alors : $D_p(\alpha_0) = 2(l(\hat{\theta}_0) - l(\alpha_0))$ qui converge en distribution vers la loi χ_k^2 , où k est la dimension du vecteur des paramètres d'intérêt α_0 .

Ainsi, pour des grandes valeurs de n , on considérera $D_p(\alpha)$ suit une loi du χ^2 à k degrés de libertés.

L'une des applications possibles de cette formulation de la déviance est la sélection de modèle, via le *likelihood ratio test*. La déviance statistique est définie par $D = 2(l(M_1) - l(M_0))$, où M_1 et M_0 sont deux modèles, M_0 étant inclus dans M_1 .

APPLICATION AU CALCUL D'INTERVALLE DE CONFIANCE On peut se servir du résultat précédent pour trouver des intervalles de confiance asymptotique et tester des hypothèses. Par exemple, on considère les variables aléatoires gaussiennes indépendantes et identiquement distribuées X_i de loi $N(m, \sigma^2)$.

La log-vraisemblance s'écrit alors

$$l_n(m, \sigma^2) = l_n(\sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - m)^2 - \frac{n}{2} \log(2\pi\sigma^2).$$

En fixant m , on trouve après calculs :

$$\hat{\sigma}^2(m) = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \left(1 + \frac{n(\bar{X} - m)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

On peut ensuite écrire :

$$l_n(m, \hat{\sigma}^2(m)) = -\frac{n}{2} - \frac{n}{2} \log \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \left(1 + \frac{n(\bar{X} - m)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \right)$$

Cette quantité est maximale pour $m = \bar{X}$ et on a alors :

$$l_n(\hat{m}, \hat{\sigma}^2(\hat{m})) = -\frac{n}{2} - \frac{n}{2} \log \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)$$

La déviance statistique s'écrit alors :

$$D_p(m) = 2 \left(l_n(\hat{m}, \hat{\sigma}^2(\hat{m})) - l_n(m, \hat{\sigma}^2(m)) \right) = n \log \left(1 + \frac{t_n^2(m)}{n-1} \right)$$

où $t_n^2(m) = n(n-1)(\bar{X} - m)^2 / \sum_{i=1}^n (X_i - \bar{X})^2$ et l'on obtient l'intervalle de confiance asymptotique sur la moyenne à 95% suivant

$$\left\{ m; n \log \left(1 + \frac{t_n^2(m)}{n-1} \right) \leq \chi_{1(0.95)}^2 \right\}$$

Part III.

Théorie Classique des Valeurs Extremes et Modélisation

2 MODÈLES ASYMPTOTIQUES

2.1 Distribution des valeurs extrêmes

1. Soit X_1, X_2, \dots, X_n n réalisations indépendantes d'une même variable aléatoire, dont on note la fonction de répartition F , et pour $n \in \mathbb{N}^*$, soit $M_n = \max(X_1, X_2, \dots, X_n)$, et $m_n = \min(X_1, X_2, \dots, X_n)$ les maximums et minimums des n réalisations.

$$P(M_n \leq z) = P(X_1 \leq z, \dots, X_n \leq z) = F^n(z), z \in \mathbb{R}$$

$$P(m_n \leq z) = 1 - P(m_n \geq z) = 1 - [1 - F(z)]^n, z \in \mathbb{R}$$

- M_n et m_n sont des variables aléatoires dont les distributions respectives sont entièrement déterminées par la donnée de F .
 - Cependant, généralement, F est inconnue. L'idée derrière la théorie des valeurs extrêmes est donc de directement approximer la distribution F^n .
2. $A = \{z : 0 < F(z) < 1\}$, $z^* = \sup(A)$ et $z_* = \inf(A)$. A est le support de la distribution F . z^* et z_* sont respectivement les extrémités supérieure et inférieure du support de la distribution F . Ainsi, on peut remarquer directement que $F^n \xrightarrow[n \rightarrow \infty]{} 0$ pour tout $z < z_*$, ce qui peut être problématique lors qu'on essaye d'approximer F^n . Pour résoudre ce problème, on travaille sur la variable aléatoire $M_n^* = \frac{M_n - b_n}{a_n}$ où $a_n > 0$ et b_n sont des suites de constantes, qui stabilisent l'échelle et le centrage de M_n . On admettra ce résultat.
 3. Soit X_1, X_2, \dots, X_n n réalisations indépendantes d'une même variable aléatoire. Si il existe une suite de constantes $a_n > 0$ et b_n telle que

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq z\right) = G(z), z \in \mathbb{R},$$

où G est une fonction de répartition non dégénérée, alors G doit appartenir à l'une des trois familles suivantes :

a) Type I :

$$G_1(z) = \exp\left\{-\exp\left(-\frac{z-b}{a}\right)\right\}, -\infty < z < +\infty$$

est appelée distribution de **Gumbel**, dont la densité est $g_1(z) = \frac{1}{a} \exp\left(-\frac{z-b}{a}\right) G_1(z)$

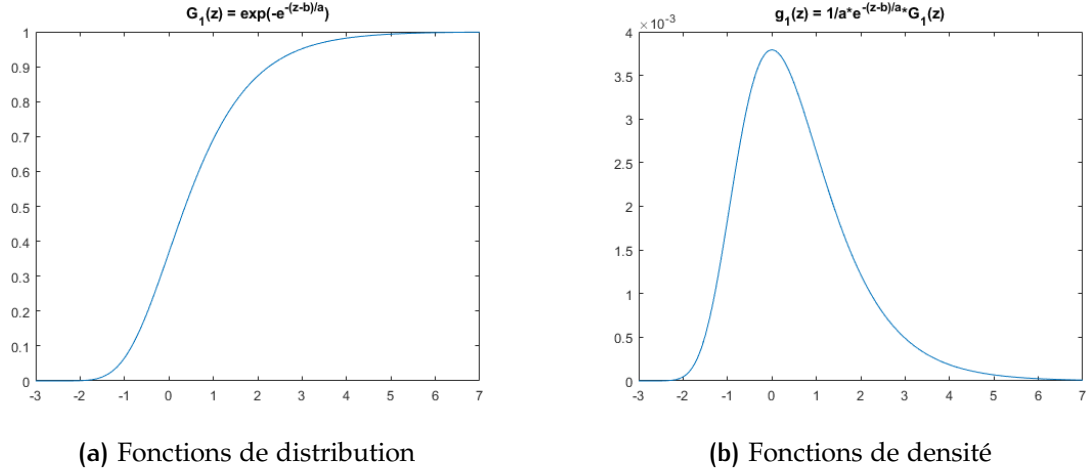


Figure 1: Distribution de Gumbel

b) Type II :

$$G_2(z) = \begin{cases} 0 & \text{si } z \leq b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\} & \text{sinon.} \end{cases}$$

est appelée la distribution de **Fréchet**, dont la densité est $g_2(z) = \frac{\alpha}{a} \left(\frac{z-b}{a}\right)^{-(1+\alpha)} G_2(z)$

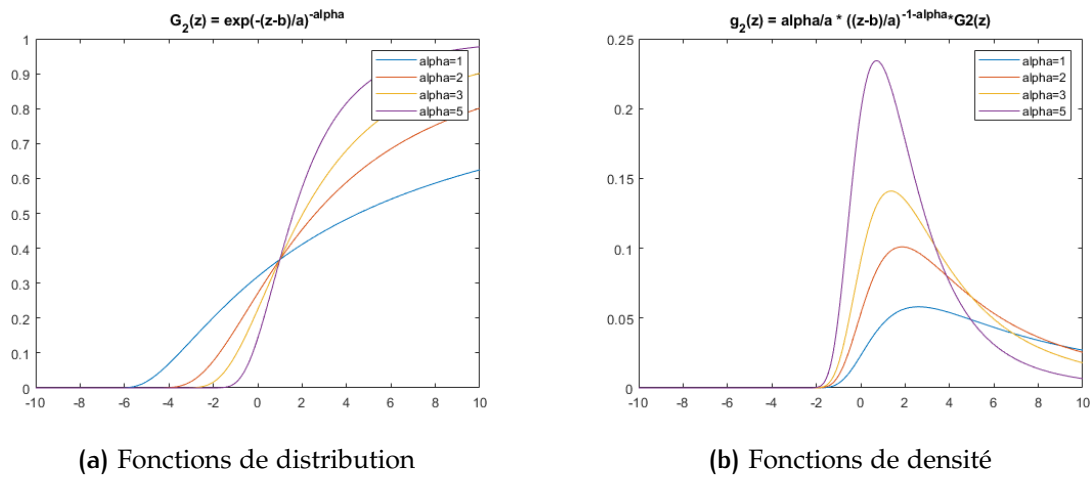


Figure 2: Distribution de Fréchet

c) Type III :

$$G_3(z) = \begin{cases} \exp\left\{-\left(-\left(\frac{z-b}{a}\right)^\alpha\right)\right\} & \text{si } z < b, \\ 1 & \text{sinon.} \end{cases}$$

est appelée la distribution de **Weibull**, dont la fonction de densité est $g_3(z) = \frac{\alpha}{a} \left(\frac{z-b}{a}\right)^{\alpha-1} G_3(z)$.

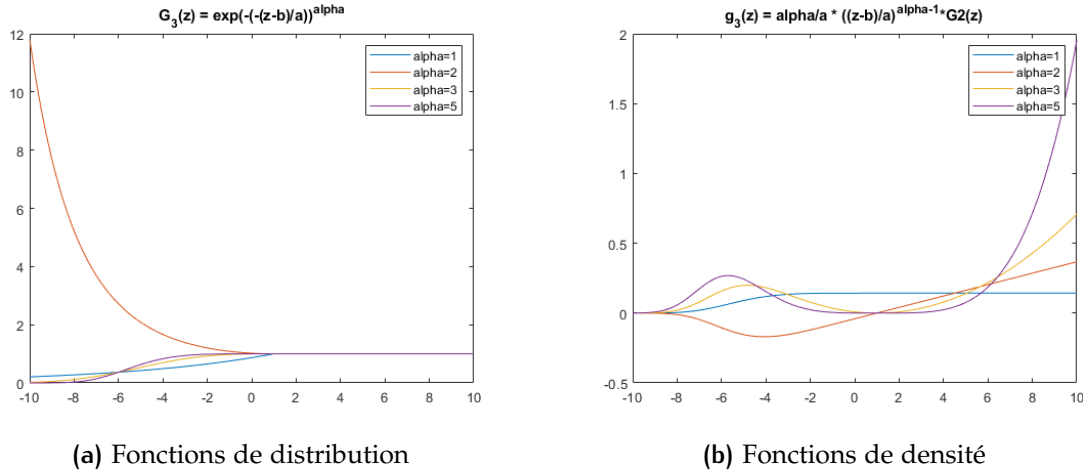


Figure 3: Distribution de Weibull

- Ce théorème, appelé *Extremal Types Theorem* stipule que lorsque M_n subit une transformation linéaire et que la variable normalisée correspondante converge vers une certaine distribution non dégénérée, alors cette distribution limite doit appartenir à l'un des trois types mentionnés ci-dessus, quelle que soit la forme de F . D'un point de vue mathématique, ces trois modèles sont interchangeables, de sorte qu'en certaines occasions, un type particulier peut être arbitrairement supposé pour des raisons de commodité.
4. Ces trois distributions représentent chacune trois comportements de valeurs extrêmes différents, mais peuvent être regroupées sous une forme unique :

$$G(z) = \exp\left\{-\left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-1/\xi}\right\} \quad (3)$$

avec $\{1 + \xi(z - \mu)/\sigma\} > 0$ et avec $\mu, \xi \in \mathbb{R}, \sigma > 0$. Cette distribution est la distribution **Généralisée des Valeurs Extrêmes**, plus rapidement appelée GEV, et dont ξ est le paramètre de forme.

- Lorsque $\xi = 0$, la distribution indique une distribution de type I.
- Lorsque $\xi > 0, \alpha = 1/\xi$ et la distribution indique une distribution de type II.
- Lorsque $\xi < 0, \alpha = -1/\xi$, la distribution indique une distribution de type III.

Il peut être montré que M_n suit la même loi, avec les mêmes paramètres, que M_n^* , à un décalage de centrage et de d'échelle près.

Ainsi, dans la théorie classique des valeurs extrêmes, on cherchera à modéliser le maximum des observations par une distribution GEV, dont on cherchera à estimer au mieux les paramètres.

5. Ce théorème s'adapte également dans le cas où on s'intéresse au minimum m_n d'une suite de n réalisations de variables aléatoires :

THÉORÈME S'il existe des suites de constantes $a_n > 0$ et b_n telles que

$$\Pr\{(m_n - b_n)/a_n \leq z\} \rightarrow \tilde{G}(z), n \rightarrow \infty \quad (4)$$

pour une fonction de distribution \tilde{G} non dégénérée, alors \tilde{G} est un membre de la famille GEV des distributions pour les minimums :

$$\tilde{G}(z) = 1 - \exp\{-[1 - \xi(\frac{z - \tilde{\mu}}{\sigma})]^{-1/\xi}\} \quad (5)$$

définie sur $z : 1 - \xi(z - \tilde{\mu})\sigma > 0$, avec $\tilde{\mu} = -\mu$ et où $-\infty < \mu < \infty, \tilde{\sigma} > 0$ et $-\infty < \xi < \infty$.

DÉMONSTRATION On considère $m_n = \min(X_1, \dots, X_n)$ où les X_i sont n réalisations indépendantes d'une même variable aléatoire. On pose $Y_i = -X_i$. Chercher le minimum des X_i revient donc à chercher le maximum des Y_i , et on a donc $\min(X_1, \dots, X_n) = -\max(Y_1, \dots, Y_n) = -M_n$. D'où

$$\begin{aligned} P(m_n \leq z) &= P(-M_n \leq z) \\ &= P(M_n \geq -z) \\ &= 1 - P(M_n \leq -z) \\ &= 1 - \exp(-(1 + \xi \frac{-z - \mu}{\sigma})^{-1/\xi}) \\ &= 1 - \exp(-(1 - \xi \frac{z - (-\mu)}{\sigma})^{-1/\xi}) \end{aligned}$$

qui est la forme recherchée.

6. Pour une fonction de distribution donnée $G(z)$, on dit que la fonction de distribution $G(z)$ est max-stable s'il existe des suites de constantes $\{\alpha_n > 0\}, \{\beta_n\}$ telles que $G^n(\alpha_n z + \beta_n) = G(z)$. G^n étant la distribution du maximum des maximums, ce théorème stipule que la distribution du maximum parmi des maximums d'observations suit la même loi, à paramètres d'échelle et de centrage près.

THÉORÈME Une distribution est max-stable si et seulement si c'est une distribution GEV.

2.2 Démonstration du théorème *Extremal Types*

L'objectif de cette démonstration est de prouver que s'il existe des suites de constantes $a_n > 0$ et b_n telles que $P(\frac{M_n - b_n}{a_n} \leq z) \rightarrow G(z)$ alors G est de l'un des trois types de la famille GEV.

La propriété de max-stabilité, selon laquelle une fonction de répartition max-stable est nécessairement de l'un des trois types de la famille GEV, est primordiale à la démonstration. En effet, si on montre que G est max-stable, alors on peut en déduire qu'elle appartient à la famille GEV.

Avant de démontrer la propriété de max-stabilité, il faut vérifier que G est max-stable. Pour cela, on se réfère à la démonstration donnée dans le chapitre 3.

On a donc G max-stable. Montrons la propriété de max-stabilité ([4]).

LEMME DE KINTCHIN On considère $\{F_n\}_{n \in \mathbb{N}}$ une suite de fonctions de répartition. Soit des fonctions de répartition non dégénérées H et H^* , des suites $a_n > 0$, b_n , $\alpha_n > 0$ et β_n . On a

$$F_n(a_n x + b_n) \rightarrow H$$

$$F_n(\alpha_n x + \beta_n) \rightarrow H^*$$

si et seulement si il existe $a > 0$ et b réel tels que

$$\lim_{n \rightarrow \infty} \frac{\alpha_n}{a_n} = a$$

et

$$\lim_{n \rightarrow \infty} \frac{\beta_n - b_n}{a_n} = b.$$

Dans ce cas, on a

$$H^*(x) = H(ax + b)$$

On dit que H^* et H sont de **même type**.

LEMME 2 Une fonction de répartition non-dégénérée G a un domaine d'attraction non vide, c'est-à-dire qu'il existe des fonctions de répartition F et deux suites $a_n > 0$ et b_n telles que $P(\frac{M_n - b_n}{a_n} \leq z) \xrightarrow[n \rightarrow \infty]{} G(z)$, si et seulement si pour tout $s > 0$, il existe $A(s) > 0$ et $B(s)$ réel tels que

$$G^s(A(s)x + B(s)) = G(x)$$

pour tout x .

LEMME 3 Soit u une fonction monotone de \mathbb{R} vers \mathbb{R}_+ qui satisfait $u(t+s) = u(t) + u(s)$ pour tous réels s et t . Alors $u(t) = 0$ ou $u(t) = \exp(\rho t)$ avec ρ réel.

DÉMONSTRATION : EXTREMAL TYPES Par hypothèse, G a un domaine d'attraction non vide, donc d'après le **Lemme 2**, pour $0 < G(z) < 1$, on a :

$$G^s(A(s)z + B(s)) = G(z)$$

$$\iff -s \log(G(A(s)z + B(s))) = -\log(G(z))$$

$$\iff -\log(-s \log(G(A(s)z + B(s)))) = -\log(-\log(G(z)))$$

$$\iff -\log(-\log(G(A(s)z + B(s)))) - \log(s) = -\log(-\log(G(z)))$$

On pose $x = -\log(-\log(G(z)))$ et on considère U , l'inverse de la fonction qui à z associe $-\log(-\log(G(z)))$. On a donc $U(x) = z$, et on peut récrire l'expression précédente sous la forme :

$$-\log(-\log(G(A(s)U(x) + B(s)))) - \log(s) = x.$$

On en déduit :

$$\begin{aligned} x + \log(s) &= -\log(-\log(G(A(s)U(x) + B(s)))) \\ \Leftrightarrow U(x + \log(s)) &= A(s)U(x) + B(s) \\ \Leftrightarrow \frac{U(x + \log(s)) - B(s)}{A(s)} &= U(x) \end{aligned}$$

En soustrayant cette dernière équation obtenue, à la même évaluée en $x = 0$, on obtient :

$$\frac{U(x + \log(s)) - U(\log(s))}{A(s)} = U(x) - U(0).$$

On pose maintenant $A_1(y) = A(e^y)$, $\tilde{U}(x) = U(x) - U(0)$ et $y = \log(s)$. L'équation précédente devient :

$$\tilde{U}(x + y) - \tilde{U}(y) = \tilde{U}(x)A_1(x).$$

Si on échange les rôles de x et y dans l'équation précédent et qu'on soustrait les deux résultats, on a alors :

$$\tilde{U}(x)(1 - A_1(y)) = \tilde{U}(y)(1 - A_1(x)).$$

Etudions les différents cas possibles :

- Si $A_1(x) = 1$ pour tout x . Alors on a $\tilde{U}(x + y) = \tilde{U}(x) + \tilde{U}(y)$. On applique le **Lemme 3** à la fonction $\exp(\tilde{U}(x))$, qui est monotone et strictement positive. On obtient donc $\exp(\tilde{U}(x)) = \exp(\rho x)$ avec ρ réel positif, et donc $\tilde{U}(x) = \rho x$. D'où

$$\begin{aligned} U(x) &= \rho x + U(0) \\ \Leftrightarrow x &= -\log(-\log(G(\rho x + U(0)))) \\ \Leftrightarrow \exp(-\exp(-x)) &= G(\rho x + U(0)) \\ \Leftrightarrow \exp(-\exp(-\frac{y - U(0)}{\rho})) &= G(y) \end{aligned}$$

G est donc une distribution de Gumbel (type I) et appartient bien à la famille des distributions GEV.

- Si $A_1(x) \neq 1$ pour au moins un x . Nous admettons alors qu'on peut prouver dans ce cas que G est de type II ou III, et donc qu'elle appartient bel et bien à la famille de distribution GEV.

[Note : difficultés à comprendre et maîtriser la preuve à ce moment-là]

2.3 Max-stabilité de la loi GEV

PROPRIÉTÉ DE MAX-STABILITÉ On considère les *blocks maximas* notés Z_1, \dots, Z_n iid, qui suivent une loi GEV. Alors d'après la propriété de max-stabilité, pour tout n , $M_n = \max(Z_1, \dots, Z_n)$ suit également une loi GEV. On a de plus les résultats suivants :

1. Weibull de paramètre α , alors M_n a même loi que $n^{-1/\alpha}Z$;
2. Gumbel, alors M_n a même loi que $Z + \log(n)$;
3. Fréchet de paramètre α , alors M_n a même loi que $n^{1/\alpha}Z$.

2.4 Loi de Pareto

DÉFINITION On définit la loi généralisée de Pareto de paramètres γ et β par la fonction de répartition suivante :

$$F_{\gamma,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}} & \text{pour } 1 + \frac{\gamma x}{\beta} > 0 \text{ si } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{pour } x > 0 \text{ si } \gamma = 0 \end{cases}$$

Cette loi nous sera utile ultérieurement.

3 ESTIMATION DES PARAMÈTRES DE LA DISTRIBUTION GEV

Comme expliqué précédemment, un modèle GEV est ajusté sur les maximums de blocs de données de taille égale. La taille de ces blocs de données ne doit pas être choisie au hasard, et doit prendre en compte le problème considéré. D'une part, s'il y a trop peu de données par bloc, le cadre asymptotique nécessaire pour supposer une distribution GEV n'est plus valable. D'autre part, si les blocs contiennent beaucoup de données, alors ils sont en nombre réduits, et on possède ainsi moins de maxima pour ajuster les paramètres de la distribution.

Dans la suite, on notera les *block maximas* Z_1, \dots, Z_m . On supposera que ce sont des variables indépendantes (c'est le cas si les données à l'intérieur de chaque blocs le sont ; et on considérera que c'est aussi le cas si elles ne le sont pas). L'estimateur du maximum de vraisemblance sera utilisé pour estimer les paramètres de la distribution GEV. On suppose que l'estimateur du maximum de vraisemblance est suffisamment régulier pour avoir la propriété de normalité asymptotique. En réalité, ce n'est pas le cas pour toutes les valeurs du paramètre ξ . Smith (1985) a donné des détails à ce sujet, mais nous ne l'avons pas étudié plus avant.

3.1 Estimation par maximum de vraisemblance

Calculons la vraisemblance des variables Z_1, \dots, Z_m , indépendantes, et suivant une distribution GEV de mêmes paramètres (μ, σ, ξ) , avec $\xi \neq 0$. Z_i admet donc pour fonction de répartition :

$$G_{Z_i}(z) = \exp\left(-\left(1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right).$$

En dérivant, on obtient l'expression de densité suivante :

$$\begin{aligned} f(z) &= \frac{\partial G_{Z_i}(z)}{\partial z} \\ &= \frac{1}{\sigma} \left(1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}-1} \exp\left(-\left(1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right) \end{aligned}$$

On a donc :

$$L(\mu, \sigma, \xi) = \frac{1}{\sigma^m} \prod_{i=1}^m \left(1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi} - 1} \prod_{i=1}^m \exp\left(-\left(1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right)$$

En passant à la log-likelihood, on obtient donc :

$$l(\mu, \sigma, \xi) = -m \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \log\left(1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right) - \sum_{i=1}^m \left(1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}$$

Dans le cas $\xi = 0$, il faut utiliser la limite $\xi \rightarrow 0$ qui conduit à la famille des lois de Gumbel, dont la fonction de répartition vaut :

$$G_{Z_i}(z) = \exp\left(-\exp\left(-\left(\frac{z - \mu}{\sigma}\right)\right)\right).$$

La densité de Z_i vaut alors :

$$\begin{aligned} f(z) &= \frac{\partial G_{Z_i}(z)}{\partial z} \\ &= \frac{1}{\sigma} \exp\left(-\left(\frac{z - \mu}{\sigma}\right)\right) \exp\left(-\exp\left(-\left(\frac{z - \mu}{\sigma}\right)\right)\right) \end{aligned}$$

Et on en déduit la log-likelihood suivante :

$$l(\mu, \sigma, \xi) = -m \log(\sigma) - \sum_{i=1}^m \frac{z_i - \mu}{\sigma} - \sum_{i=1}^m \exp\left(-\left(\frac{z_i - \mu}{\sigma}\right)\right)$$

Il faut ensuite chercher les valeurs de paramètres maximisant la log-vraisemblance. Il est possible d'obtenir des solutions numériques. Nous calculerons, lorsque nous réaliserons l'application, des estimateurs par maximum de vraisemblance. Nous devons alors penser à :

- Imposer une condition sur les paramètres dans le cas $\xi \neq 0$, afin que le logarithme prenne toujours ses valeurs dans \mathbb{R}_+ ;
- Utiliser la bonne expression de la log-vraisemblance selon la valeur de ξ .

Ayant considéré le modèle régulier, on a la normalité asymptotique des estimateurs de μ , σ et ξ . On peut donc facilement en déduire des intervalles de confiance, par exemple.

3.2 Return Levels

On appelle niveau de retour (ou *Return Level*) z_p le quantile d'ordre $1 - p$ d'une distribution GEV, c'est-à-dire tel que $G(z_p) = 1 - p$. C'est une quantité d'intérêt dans le cadre de la théorie des valeurs extrêmes, puisque l'on cherche à savoir avec quelle probabilité des événements encore plus extrêmes que ceux observés peuvent avoir lieu. Ainsi, la probabilité d'excéder la valeur z_p pour le phénomène étudié vaut p . On peut aussi dire qu'on

observe un phénomène au moins aussi extrême que z_p toutes les $\frac{1}{p}$ années. On obtient du *Return level* de niveau p en inversant l'expression de la distribution GEV, c'est-à-dire :

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - (-\log(1 - p))^{-\xi}) & \text{si } \xi \neq 0 \\ \mu - \sigma \log(-\log(1 - p)) & \text{sinon.} \end{cases}$$

Pour estimer le niveau de retour d'ordre p , on peut évaluer z_p , dont l'expression a été donnée plus haut, en $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ estimés par maximum de vraisemblance. Il est possible de calculer la variance de \hat{z}_p grâce à la delta-méthode multivariée et la normalité asymptotique des estimateurs du maximum de vraisemblance.

3.3 Profile Likelihood

Il est également tout à fait possible d'estimer les paramètres de la distribution GEV (et même le niveau de retour d'ordre p par reparamétrisation) via la méthode de la fonction Profile Likelihood. Il sera intéressant, en réalisant l'application, de réaliser les deux méthodes et de comparer leurs résultats, notamment au niveau des intervalles de confiance.

3.4 Validation du modèle

Après avoir estimé les paramètres d'une façon ou d'une autre, il faut à minima vérifier l'adéquation du modèle muni de ces paramètres et des données observées. On peut utiliser les méthodes graphiques suivantes :

- *Probability plot:*

En considérant les maximums ordonnés $z_{(1)} \leq \dots \leq z_{(m)}$, on trace

$$\left\{ \frac{i}{m+1}, \hat{G}(z_{(i)}); i = 1, \dots, m \right\}$$

où \hat{G} est la distribution GEV évaluée avec les paramètres estimés, et où le premier terme est l'expression de la fonction de répartition empirique. Si le modèle est cohérent avec les données observées, on devrait obtenir des points proches de la première bissectrice.

- *QQ plot:*

On trace

$$\left\{ \hat{G}^{-1}\left(\frac{i}{m+1}\right), z_{(i)}; i = 1, \dots, m \right\}.$$

Si on obtient des points près de la première bissectrice, le modèle estimé est cohérent avec les données observées.

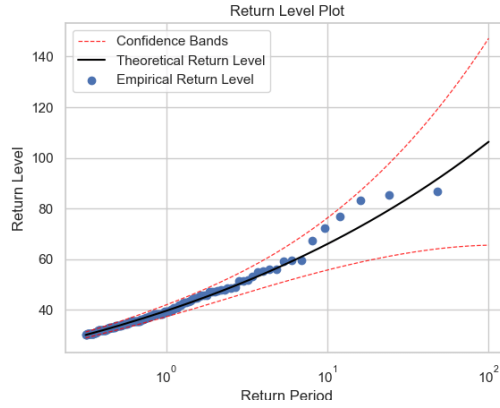
- *Return Level plot:*

Tracer

$$\{\log(-\log(1 - p)), \hat{z}_p, 0 < p < 1\}$$

où \hat{z}_p est l'estimateur du maximum de vraisemblance de z_p , permet d'observer les estimations pour les grands niveaux de retours, normalement compressés dans les queues de distribution.

Enfin, ajouter au *Return Level plot* les estimations empiriques des niveaux de retour permet de vérifier ou non la validité du modèle.



(a) Exemple de *Return Level plot*

Ces méthodes graphiques nous permettront de valider (ou non) notre propre modèle d'application.

3.5 Résultats asymptotiques

Nous donnons ici une suite de résultats qui nous seront utiles plus tard. Certains ont déjà été évoqués précédemment.

3.5.1 Convergences

THÉORÈME Soit (X_n) une suite de variables iid de fonction de répartition F . On note $M_n = \max(X_1, \dots, X_n)$. Si il existe deux suites réelles (a_n) (avec $a_n > 0$ pour tout n) et (b_n) et l'existence d'une loi non dégénérée de loi de répartition G telle qu'on a la convergence en loi suivante :

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

alors G est une fonction de répartition d'une loi GEV.

En notant, pour $k \in 1, \dots, n$, $M_n^{(k)} = X_{i_1}, \dots, X_{i_k}$, où l'on réordonne (pour n fixé) les X_i par ordre décroissant, on peut étendre ce résultat aux processus stationnaires.

DÉFINITION (CONDITION $D(u_n)$) On note $I_{j,k}(u_n)$ l'évènement $I_{j,k}(u_n) = \bigcap_{j \leq i \leq k} \{X_i \leq u_n\}$. Pour tous évènements A de $I_{1,\ell}(u_n)$ et B de $I_{\ell+s,n}(u_n)$ tels que $1 \leq \ell \leq n-s$, on a

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \alpha(n, s),$$

où $\alpha(n, s)$ tend vers 0 en $+\infty$ pour une suite d'entiers positifs $s_n = o(n)$.

Cette définition caractérise le fait que les événements $\{\max_{k \in I_1} X_k \leq u_n\}$ et $\{\max_{k \in I_2} X_k \leq u_n\}$ deviennent indépendants pour n grand si les ensembles d'indices $I_i \subset \{1, \dots, n\}$ sont séparés par une distance relativement faible par rapport à n .

THÉORÈME DE LEADBETTER Soit X un processus stationnaire. On note $M_n = \max(X_1, \dots, X_n)$ et pour $k \in 1, \dots, n$, $M_n^{(k)} = X_{i_1}, \dots, X_{i_k}$, où l'on réordonne (pour n fixé) les X_i par ordre décroissant. Si il existe deux suites réelles (a_n) (avec $a_n > 0$ pour tout n) et (b_n) et l'existence d'une loi non dégénérée de fonction de répartition G telle qu'on a la convergence en loi suivante :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq z \right) = G(z)$$

et que de plus la condition $D(u_n)$ est vérifiée pour tout seuil $u_n = a_n z + b$ avec $G(x) > 0$ alors G est une fonction de répartition d'une loi GEV.

DÉFINITION (DOMAINE D'ATTRACTION) En reprenant les notations précédentes, et le théorème dans le cadre de variables iid, on définit le domaine d'attraction d'une fonction G comme étant l'ensemble des fonctions de répartition F tels qu'il existe deux suites (a_n) et (b_n) vérifiant les hypothèses du théorème précédent.

Un des objectifs est alors de mettre en évidence le lien entre F et les paramètres des distributions GEV vers lesquelles elle converge éventuellement.

STABILITÉ PAR MAXIMUM Si les X_i suivent une loi GEV de paramètres $0, 1, \xi$ (de fonction de répartition F , alors on a pour tout $n \in \mathbb{N}$ $\mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) = F(x)$ et M_n suit la même distribution GEV.

3.6 Exemple : Niveau d'eau à Venise

Nous résumons ici l'exemple donné dans le livre de référence. Les données que nous avons sont les 10 niveaux d'eau les plus hauts relevés chaque année entre 1931 et 1981 à Venise (à l'exception de l'année 1935 où nous disposons uniquement des 6 observations les plus grandes). En fonction du nombre de paramètres que l'on choisit d'étudier, et en faisant le choix de modéliser les r plus grandes valeurs annuelles par une loi GEV, on trouve les estimateurs du maximum de vraisemblance adéquats. On remarque que plus le choix de r est grand, plus la loi GEV trouvée est pertinente (erreur moindre).

Part IV.

Applications

4 INTRODUCTION

Les valeurs extrêmes surviennent rarement, et elles peuvent de plus être difficiles à observer pour plusieurs raisons. En effet, les capteurs peuvent ne pas être adaptés à la venue d'événements extrêmes, ou alors celles-ci peuvent causer des dégâts mettant à mal la récolte des données. On peut imaginer dans le cadre de l'observation de phénomènes naturels, le fait que la venue d'un cyclone endommage certains outils de mesure par exemple.

La collecte des données est donc très importante en pratique, et des données manquantes, erronées, ou biaisées peuvent fausser leur étude.

En pratique, il semble important de disposer de plusieurs sources de données corrélées, éventuellement non numériques (signalements humains), pour vérifier les données présentes d'une part, et pour pouvoir tenter d'interpréter les résultats obtenus via l'utilisation des méthodes étudiées.

Il existe plusieurs manières de traiter les phénomènes extrêmes et les approches varient. En premier lieu, deux options s'offrent en ce qui concerne la récolte (ou l'utilisation) des données.

- Données sous forme de maximums sur une période temporelle donnée. Par exemple, on récolte chaque jour les 4 (nombre choisi arbitrairement) plus grandes pressions artérielles d'un patient
- Données dépassant un seuil donné. On récolte ici, les données que l'on considère extrêmes lorsqu'elles dépassent un seuil choisi à l'avance.

Dans le premier cas, on a vu précédemment que si l'on suppose que les données récoltées suivent des lois iid, alors la probabilité d'obtenir un échantillon de maximums donné converge en loi (sous certaines conditions) vers une distribution des valeurs extrêmes.

Dans le second cas, on a un résultat similaire, en fixant un seuil u à l'avance, et en considérant les lois restrictions des lois X_i iid à leurs valeurs supérieures au seuil u : la probabilité d'obtenir un échantillon de la loi ainsi obtenue converge vers une loi de Pareto généralisée.

Nous étudierons plus formellement ce dernier résultat, et nous verrons qu'il existe un lien entre les deux approches qui se traduit dans certains cas par une équivalence entre les deux résultats de convergence précédent.

Dans ces deux cas, il faut suffisamment de données pour mener notre étude. Ensuite, selon le type de données que l'on a, on modélise les observations par une loi des valeurs extrêmes ou de Pareto.

Enfin, il reste à valider ou non notre modélisation.

Part V.

Perspectives et conclusions

5 PERSPECTIVES

Notre objectif immédiat est de réaliser une application de la théorie des valeurs extrêmes, c'est-à-dire d'ajuster un modèle GEV sur un jeu de données dont les valeurs extrêmes nous intéressent. On pourra alors en déduire les probabilités des événements rares, via les niveaux de retours notamment.

Nous souhaiterions pouvoir réaliser une application en biologie. En effet, la théorie des valeurs extrêmes peut être utilisée pour étudier les effets secondaires des médicaments, les mutations pendant le processus d'évolution, les évolutions épidémiques ou encore le comportement neurologique.

Nous avons notamment lu, lors de nos recherches, des papiers sur :

- Une application en neuroscience, dans laquelle la théorie des valeurs extrêmes est appliquée à l'étude de l'activité cérébrale intense. Le modèle univarié de la théorie classique des valeurs extrêmes peut être appliqué ici, mais il nous serait également possible d'utiliser la méthode du *Peak over the Threshold* évoquée rapidement précédemment et développée dans le Chapitre 4 et la théorie des valeurs extrêmes multivariée, développée dans le Chapitre 8. La principale difficulté réside ici dans l'obtention et le traitement des données ;
- Une application plus classique en biologie, qui consiste par exemple à étudier les pressions artérielles extrêmes ;
- Une application en épidémiologie, qui consiste à étudier les valeurs extrêmes du nombre de cas d'une maladie pour estimer la probabilité d'un départ épidémique.

Nous ne réaliserons pas toutes les applications citées ci-dessus ; nous nous concentrerons sur l'une d'entre elle selon les données trouvées.

En ce qui concerne le choix du langage utilisé pour notre cas d'étude, il est utile de noter que Python et R disposent de packages (respectivement *pyextremes* et *extRemes*) permettant d'appliquer directement la théorie des valeurs extrêmes via des fonctions d'ores et déjà implémentées. Ces deux packages ont d'ailleurs été inspirés par le livre de S. Coles [1].

RÉFÉRENCES

- [1] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.
- [2] Wikipedia. Value at risk, 2004. [Online; accessed 26-Jan-2023].
- [3] Gregory Gundersen. Asymptotic normality of max likelihood estimator. [Online; accessed 26-Jan-2023].
- [4] Richard L. Smith and Ishay Weissman. Extreme values. [Online; accessed 26-Jan-2023].

Part VI.

Annexe

A CODE DE PROGRAMMATION

A.1 Fonctions de distribution non dégénérées

```
1 %Perturber les graines d'un générateur de nombres aléatoires
2 rng('shuffle');
3 low = 0; high = 100;
4
5 %distribution de Gumbel
6 z = linspace(-10,10,100);
7 b = randi([low, high]);
8 a = randi([low, high]);
9
10 X = (z-b)/a;
11 G1 = exp(-exp(-x));
12 g1 = (1./a).*exp(-x).*G1;
13 figure()
14 plot(x,G1);
15 title('G_1(z) = exp(-e^{-(z-b)/a})');
16 figure()
17 plot(x,g1);
18 title('g_1(z) = 1/a*e^{-(z-b)/a}*G_{1}(z)')
```

```
1 %Perturber les graines d'un générateur de nombres aléatoires
2 rng('shuffle');
3 low = -10; high = 10;
4
5 %distribution de Gumbel
6 b = randi([low, high]);
7 a = randi([low, high]);
8
9 %distribution de Frechet
10 alpha = [1,2,3,5];
11 z = linspace(-10,10,1000);
12 x = (z-b)/a;
13
14 figure()
15 for i = 1:length(alpha)
16     G2 = zeros(size(z));
```

```
17     index = z > b;  
18     G2(index) = exp(-(x(index).^(-alpha(i))));  
19     plot(z, G2)  
20     hold on  
21 end  
22 legend('alpha=1','alpha=2','alpha=3','alpha=5')  
23 title('G_2(z) = exp(-(z-b)/a)^{-alpha}')
```

```
24  
25 figure()  
26 for i = 1:length(alpha)  
27     g2 = (alpha(i)/a) * x.^(-1-alpha(i)).*G2;  
28     plot(z, g2)  
29     hold on  
30 end  
31 legend('alpha=1','alpha=2','alpha=3','alpha=5')  
32 title('g_2(z) = alpha/a * ((z-b)/a)^{-1-alpha}*G2(z)')
```

```
1 %Perturber les graines d'un générateur de nombres aléatoires  
2 rng('shuffle');  
3 low = 0; high = 10;  
4  
5 %distribution de Gumbel  
6 b = randi([low, high]);  
7 a = randi([low, high]);  
8  
9 %distribution de Weibul  
10 alpha = [1,2,3,5];  
11 z = linspace(0,10,1000);  
12 x = (z-b)/a;  
13  
14 figure()  
15 for i = 1:length(alpha)  
16     G3 = ones(size(z));  
17     index = z < b;  
18     G3(index) = exp(-(-x(index).^(alpha(i))));  
19     plot(z, G3)  
20     hold on  
21 end  
22 legend('alpha=1','alpha=2','alpha=3','alpha=5')  
23 title('G_3(z) = exp(-(-(z-b)/a))^{alpha}')
```

```
24  
25 figure()  
26 for i = 1:length(alpha)  
27     g3 = (alpha(i)/a) * x.^(alpha(i)-1).*G3;
```

```
28     plot(z, g3)
29     hold on
30 end
31 legend('alpha=1','alpha=2','alpha=3','alpha=5')
32 title('g_3(z) = alpha/a * ((z-b)/a)^{alpha-1}*G3(z)')
```