

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique
UNIVERSITÉ KASDI MERBAH OUARGLA
FACULTÉ DE MATHÉMATIQUES ET SCIENCES DE LA
MATIÈRE
DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté En Vue De L'obtention Du

DIPLÔME DE MASTER

EN MATHÉMATIQUES

Option : Probabilité et Statistique

Par

Rezzage habla Manal

Intitulé

Modélisation statistique des valeurs extrêmes
par la distribution GEV

Membres du jury

MANSOUL Brahim	M. A. A	UKMO	Président
BOURDJI Hind	M. C. B	UKMO	Examineur
ARBIA Hanane	M. C. B	UKMO	Rapporteur

2023-2024

التشكر

بسم الله والصلاة والسلام على رسول الله
الحمد لله الذي بنعمته تتم الصالحات
أحمد الله تعالى أن أعطاني الشجاعة
والصبر اللذان احتجت لهما لإنجاز هذا
العمل.

هذا العمل رغم كونه ليس سهلا لكنه
ليس فرديا أبداً، لقد ساعدني الكثير
من الأشخاص وساهم كل على طريقته في
إنجازه:

مشرفتي الأستاذة "عربية حنان" ،
أشكرك على منحي الثقة والحرية في
العمل وأشكرك على تواجذك الدائم
وحصرك على متابعة تقدمي المستمر رغم
ظروفك الصعبة هذا العام لكنك عملتي
بجد معي فلك مني كل التقدير
والاحترام .

أستاذي " منسول ابراهيم " ، أشكرك
على صفاتك الانسانية العالية و كل
الوقت الذي قدمته لي و علمك الذي
منحتني اياه طوال دراستي في هذا
التخصص، أشكرك أيضا خالص الشكر على
حضورك اليوم ويشرفني مشاركتك كرئيس
في لجنة المناقشة لمذكرتي.

الأستاذة " بن ابراهيم " و الأستاذة
" برجي هند " أشكركم على كل ما

قدمتم لي من علمكم ووقتكم الثمين
ولن أنسى فضلكم طوال حياتي وشكرا
جزيلا على حضوركم اليوم لمناقشة
مذكرتي .

سأبقي دائما معي هذه الذكرى التي
تجمعني بكم فأنتم قدوتي المثلى.
أشكر أيضا شركة سوناطراك على
استقبالهم الرائع لي وأتقدم قبل هذا
بخالص امتنان ل التي كانت سببا في
حصولي على فرصة التدريب العملي هناك
وهي الغالية "نسيمة عرباوي" كما لا
أنسى جميع من ساعدوني هناك ابتداءا
من مؤطري الرائع المهندس "سليمان
زلماط" الذي منحني الكثير من وقته
وجهدده و ألهمني بنصائحه القيمة و
كان أول من شجعني على دمج الاحصاء
بالصيانة والموثوقية وقام بتوفير
البيانات التي اعتمدت عليها في هذا
العمل فله كل الشكر والتقدير.
شكرا جزيلا أيضا للسيد "محمد سعادي"
و "توفيق ركال" و العزيزة "فاطمة"
و جميع من ساعدوني هناك في قسم
الصيانة .

أشكر زملاء الدراسة على الأيام
الرائعة و أخص بالشكر زميلي "علي
دباخ"

الذي قدم لي يد العون .
أشكر جميع الأساتذة اللذين أحبوني و

وقفوا الى جانبي طوال مسيرتي
الجامعية وقبلها , أساتذتي في
الثانوية وفي الاعدادية ومعلمتي
ومعلمي في الابتدائية
شكرا جزيلا لكم جميعا
لن انساكم أبدا .

الاهداء

أهدي تخرجي وفرحتي التي انتظرتها طوال حياتي إلى من مهدت لي طريق العلم وكانت الداعم الأول لي لتحقيق طموحاتي إلى من ضحت بالكثير والكثير في سبيل وصولي إلى هنا

وغمرتني بالحب والحنان ولم تنساني بدعائها في كل سجد
إلى التي لا أجد كلمات تعبر عن قيمتها في حياتي ومدى
امتناني لها

" جنتي و حبيبتي الغالية أمي "

أهدي عملي المتواضع أيضا إلى الذي صنع لي قاعدة سليمة
في الرياضيات منذ صغري وحرص على تعليمي جيدا
إلى من يكفيني فخرا أنه يثق بي أين ما ذهبت ويشجعني على
طموحاتي الكبيرة
" أبي العزيز "

إلى من لهن سري وبهن سعادتي إلى حبيباتي أخواتي رفيقات
دربي العزيزات على قلبي " زينة " و "كوثر" وأخي الغالي
سندي و فخري وكنزي " أبو بكر الصديق "

أهدي لكم فرحتي اليوم
فلولا تشجيعكم ودعمكم لما وثقت بنفسي
أهدي نجاحي أيضا إلى أقرب شخص لقلبي و منبع الحنان
والمحبة في حياتي إلى من لا تصف الكلمات حبي له "جدي
الحبيب" رحمه الله و نعمه في الفردوس الأعلى
و إلى جدتي وكل أحوالي وخالاتي و إلى جميع أفراد عائلتي
الأعزاء على قلبي

لن أنسى طبعاً من كان لي أبا حنوناً و سندا و قدوة
ساعدني في تجاوز أصعب العقبات في حياتي و أنار عقلي
للعلم والبحث

وشجعني على تحقيق أهدافي و الإيمان بها
"أستاذي الرائع عطوي محمد"

أهدي نجاحي أيضا إلى معلمتي "جبالله رشيدة" التي قرأت
وكتبت معها أول حروفي وصنعت معها شخصيتي و مبادئني و لن
أنسى فضلها طوال عمري
و إلى من جمعتنني بها أحلى الصدف وكانت لي صديقة وفية
وأختا طيبة

"حبيبتي عرباوي شيماء"

أهدي هذا العمل وفرحتي لك و إلى كل صديقاتي
و جميع زملاء الدراسة المقربين منذ الابتدائية إلى

الجامعة

رغم أنه الإهداء الأخير لكنه لا يقل أهمية عن ما سبقه
إلى من أحبوني في الله وساعدوني ووقفوا بجانبني في أصعب

أيام الجامعة

"شعيب عريس"

"أسامة"

Table des matières

Table des matières	i
Table des figures	iii
Notations et abriviations	v
Introduction	1
1 La théorie des valeurs extrêmes	3
1.1 Notions fondamentales sur les probabilités	3
1.1.1 Variable aléatoire	3
1.1.2 Fonction de répartition	4
1.1.3 Fonction de densité	4
1.1.4 Fonction de survie	4
1.2 Notions fondamentales sur les valeurs extrêmes	5
1.2.1 Statistique d'ordre	5
1.2.2 Fonction de répartition empirique	7
1.2.3 Inverse généralisée	7
1.2.4 Fonction des quantiles empiriques	8
1.2.5 Point terminal	8
1.2.6 Fonction à variation réguliers	9
1.2.7 Lois à queue lourde	9
1.2.8 Les quantiles	9
1.2.9 Fonction quantile de queue	10
1.2.10 Quantile extrême	10
1.2.11 Quantile empirique	10
1.3 La loi des valeurs extrêmes	11
1.3.1 La distribution GEV	14
1.3.2 Domaines d'attractions	16

2	Estimation des paramètres de loi GEV	18
2.1	Estimation paramétrique	18
2.1.1	Méthode de maximum de vraisemblance	19
2.1.2	Méthode de moments	21
2.1.3	Méthode de L-moments	24
2.2	Estimation semi- paramétrique	27
2.2.1	Estimation de Hill	27
2.2.2	Estimation de Pickands	31
2.2.3	Estimateur du moment	34
2.2.4	Le choix du nombre k	35
2.3	Estimation des quantiles extrêmes	37
2.4	Estimation de période de retour	38
3	APPLICATION	39
3.1	Description et explication des données :	39
3.1.1	Explique la sonde de corrosion :	39
3.2	Analyse statistique des données :	57
3.3	Test de normalité :	59
3.3.1	Interprétation :	59
3.4	Modélisation de la distribution GEV :	59
3.5	Estimation :	62
3.5.1	Estimation paramétrique :	62
3.5.2	Estimation semi-paramétrique :	64
3.6	Estimation des quantiles extrêmes(niveau de retour) :	65
3.6.1	Méthode de maximum vraisemblance :	65
3.6.2	Méthode des moments :	66
3.6.3	Méthode des L-moments :	68
3.7	Simulation	70
3.7.1	Généraliser et simuler un échantillon de données pour une distribution GEV :	70
3.7.2	Estimation des paramètres et des quantiles extremes pour la loi GEV :	72
3.7.3	Conclusion :	75
	Conclusion générale	76
	Bibliographie	77

Table des figures

1.1	Densité de la distribution Fréchet	12
1.2	Densité de la distridution Weibull	13
1.3	Fonction de répartition de la distribution GEV	15
1.4	Fonction densité de la distribution GEV	16
2.1	Représentation graphique de l'estimation de Hill	31
2.2	Représentation graphique de l'estimateur de Pickands	33
2.3	Représentation graphique de l'estimateur de Moment	35
2.4	Diagramme schématique de la méthode des blocs maxima	36
2.5	Le graphe de $(k, \hat{\xi}_n)$ pour les différents estimateur de ξ	37
3.1	Une illustration de la corrosion et comment la mesurer	40
3.2	Tableau des données 1	41
3.3	Tableau des données 2	42
3.4	Tableau des données 3	43
3.5	Tableau des données 4	44
3.6	Tableau des données 5	45
3.7	Tableau des données 6	46
3.8	Tableau des données 7	47
3.9	Tableau des données 8	48
3.10	Tableau des données 9	49
3.11	Tableau des données 10	50
3.12	Tableau des données 11	51
3.13	Tableau des données 12	52
3.14	Tableau des données 13	53
3.15	Tableau des données 14	54
3.16	Tableau des données 15	55
3.17	Tableau des données 16	56
3.18	Représentation graphique des données à l'aide d'Excel	57
3.19	Correspondance les quantiles des données avec les quantiles de GEV	61
3.20	Ajuster la fonction de densité des données à la fonction de densité GEV	61

3.21	Histogramme avec distribution de Weibull ajustée	64
3.22	Niveau de retour par MLE	66
3.23	Niveau de retour par ME	67
3.24	Niveau de retour par LME	69
3.25	Données simulées de la distribution GEV	71

Notations et abriviations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

$v.a$:	variable aléatoire.
F :	Fonction de répartition.
\bar{F} :	Fonction de survie.
F^{\rightarrow} :	Inverse généralisée de F .
$i.i.d$:	Indépendente et identiquement distribuées.
$Card$:	Cardinale.
x_F :	Point terminale .
TCL :	Théorème Centrale Limite.
H_{ξ} :	Distribution des valeur extrême.
GEV :	Distribution des valeurs extrêmes généralisés.
PDF :	Probability Density Function (La densité).
PWM_j :	Moments de probabilité pondérés β_j .
$MDA(H\xi)$:	Domaine d'attraction maximale.
R_{V0} :	Les fonction à variation lente.
$\hat{\xi}_{((k(n)),n)}^{(H)}$:	Estimateur de Hill.
$\hat{\xi}_{(k(n),n)}^{(p)}$:	Estimateur de Pickands.
$\hat{\xi}_{(k(n),n)}^{(M)}$:	Estimateur de Moment.
$:=$:	Egalité en définition .
\xrightarrow{p} :	convergence en probabilité .
$\xrightarrow{p.s}$:	Convergence presque sur .
$N(0, v(\xi))$:	La loi normal centré et de variance $v(\xi)$.
GPD :	Distribution de Pareto généralisée.
MLE :	Méthode de maximum vraisemblance.
ME :	Méthode des moments.
LME :	Méthode des L-moments.

Introduction

Ces dernières années, la théorie de la valeur extrême (TVE) ou "Extreme Value Theory" (EVT en anglais) a reçu beaucoup d'attention aussi bien sur le plan théorique que sur le plan pratique. Cela est dû aux domaines d'application vastes et diversifiés, par exemple Hydrologie, météorologie, biologie, ingénierie, fiabilité, et la gestion de l'environnement, Finance, assurances, sciences sociales, etc...

La TVE est une branche de la théorie des probabilités et des statistiques mathématiques qui se concentre sur l'analyse et l'inférence des événements extrêmes, c'est-à-dire les événements avec très faible probabilité d'occurrence. Les événements extrêmes revêtent une grande importance dans presque tous les domaines de la science et de la technologie, car elles peuvent être désastreuses et être une source d'énormes pertes.

Cette théorie a été développée dans le contexte d'observations indépendantes les auteurs Fisher et Tippett (1928) [22], montre-le ci-dessous dans certaines conditions, les seules distributions limites des extrêmes sont les lois de Fréchet, Gumbel et Weibull.

Ceci nous permet de classer la plupart des lois en trois domaines d'attracteurs où chaque domaine est déterminé par des caractérisations sur les fonctions de répartition. Von Mises (1954) [34] puis Jenkinson (1955) [28] ont rassemblé les distributions de ces trois domaines en un écriture (voir Embrechts et al (1997) [19], de Haan et Ferreira (2006) [18]). On trouve deux modèles des valeurs extrêmes : loi des valeurs extrêmes généralisée (*GEV* : Generalized Extreme Value) et loi de Pareto généralisée (*GPD* : Generalized Pareto Distribution). Divers travaux ont été consacrés à l'estimation de l'indice des extrêmes dont l'objectif revient à construire des estimateurs et étudier leurs propriétés; citons Hill (1975) [26] et Pickands (1975) [32]. La plupart des estimateurs reposent sur l'utilisation de la statistique d'ordre. Les modèles des valeurs extrêmes sont appliquées à une grande variété de problèmes tels l'environnement (vitesse du vent, extrêmes pluviométriques et de températures,...), la pluviométrie, la finance et l'assurance (Mesure du risque, Valeur à risque VaR, modèle de volatilité stochastique), hydrologie pour calculer la probabilité que la hauteur d'eau d'une rivière dépasse un certain seuil, communication, la physique, ...

Les méthodes statistiques les plus avancées pour les événements extrêmes seront étudiées à la fois du côté théorique et du côté des applications. Pour ce dernier, le logiciel d'analyse statistique R sera utilisé.

Dans ce travail nous intéressons à la modélisation statistique des valeurs extrêmes en utilisant la distribution GEV . L'objectif principal est la prévision, qui consiste à calculer les niveaux de retour des événements rares qui présentent un risque élevé et entraînent d'énormes pertes.

Ce mémoire s'organise en 3 chapitres :

chapitre 01 : on présente tout d'abord quelques notions fondamentales dans la probabilités et statistiques et quelques définitions de certains outils nécessaires dans la théorie des valeurs extrêmes, dont le plus important est la définition des statistiques d'ordre, Théorème de Fisher-Tippett (1928), Fonction de répartition de la distribution GEV , Fonction de densité de la distribution GEV , Le quantile $Q(p)$ de la distribution GEV , Domaines d'attractions...

chapitre 02 : Nous discutons des méthodes d'estimation des paramètres de distribution GEV qui est divisé en deux types : Estimation paramétrique, dans laquelle nous avons fait référence à des méthodes communes et différentes (ME, MLE, LME), et estimation semi-paramétrique, dans laquelle spécialisés uniquement dans l'estimation du paramètre de forme de la distribution GEV , dans laquelle nous avons également abordé différentes méthodes (Hill, Pickands, Moment). Nous n'avons pas oublié de mentionner les propriétés de l'estimateur. Il a ensuite fallu aborder la méthode de choix du nombre de blocs (k), que ce soit par méthode graphique ou analytique. Ensuite, nous avons abordé le sujet le plus important, qui est l'estimation des quantiles extrêmes et du niveau de retour, qui est la prévision que nous cherchons à atteindre.

chapitre 03 : Ce chapitre constitue la partie la plus importante de notre recherche en raison de ses bénéfices directs sur le terrain. Il représente l'aspect appliqué qui s'appuie sur des données réelles. Nous avons analysé ces données et réalisé un test statistique (dans notre étude nous nous sommes appuyés sur le test de Shapiro), qui détermine si nos données suivent une distribution normale ou ne la suivent pas. Nous l'avons ensuite ajusté à la distribution GEV et estimé ses paramètres à l'aide des méthodes décrites précédemment, puis estimé les quantiles extrêmes et les niveaux de retour, sur la base des résultats de l'estimation.

Chapitre 1

La théorie des valeurs extrêmes

La théorie des valeurs extrêmes communément est une vaste théorie dont le but est d'étudier les événements rares c'est-à-dire, les événements dont la probabilité d'apparition est faible. Autrement dit elle essaie d'amener des éléments de réponses aux intempéries, aux inondations, aux catastrophes naturelles, aux problèmes financiers, etc, en prédisant leurs occurrences dans les années à venir. En d'autres termes on veut estimer des petites probabilités ou des quantités dont la probabilité d'observation est très faible c'est-à-dire proche de zéro. Ces quantités sont appelées quantiles extrêmes car l'ordre de ces quantiles tend vers zéro lorsque la taille de l'échantillon, n tend vers l'infini.

1.1 Notions fondamentales sur les probabilités

Il est important de mentionner quelques concepts généraux de base et fondamentaux en probabilités et statistiques.

1.1.1 Variable aléatoire

La variable aléatoire est la notion principale de la théorie des probabilités et de la statistique.

Définition 1.1.1 *Soit (Ω, F, P) un espace de probabilité. Une variable aléatoire réelle est une application X dans \mathbb{R} :*

$$X : w \in \Omega \rightarrow X(w) \in \mathbb{R}$$

Lorsque l'ensemble des réalisations possibles de la variable aléatoire réelle X est un fini ou dénombrable, on dit que la variable aléatoire réelle X est discrète. Sinon, on dit qu'elle est continue.

1.1.2 Fonction de répartition

En théorie des probabilités ou en statistiques la fonction de répartition d'une variable aléatoire caractérise la loi de probabilité de cette variable aléatoire.

Définition 1.1.2 *La fonction de répartition d'une v.a X est l'application F définie de \mathbb{R} dans $[0; 1]$ par :*

$$F(t) = P(X \leq t)$$

F est aussi appelée fonction de distribution ou fonction de distribution cumulée.

1.1.3 Fonction de densité

En théorie des probabilités, une variable aléatoire à densité est une variable aléatoire réelle, scalaire ou vectorielle, pour laquelle la probabilité d'appartenance à un domaine se calcule à l'aide d'une intégrale sur ce domaine.

Définition 1.1.3 *On appelle densité de probabilité toute application telle que :*

1. Pour tout $x \in \mathbb{R}$ $f(x) \geq 0$.
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$ (en supposant que $\int_{-\infty}^{+\infty} f(x)dx$ existe).

Ainsi la relation entre la fonction de répartition F et la fonction de densité de probabilité f est la suivante :

$$F(t) = P(X \leq t) = \int_{-\infty}^t f(x)dx.$$

1.1.4 Fonction de survie

La fonction de survie d'une v.a est définie par :

Définition 1.1.4 *La fonction de survie est la probabilité pour que l'évènement d'intérêt T intervienne après un délai supérieur à t , autrement dit l'évènement d'intérêt T ne survienne pas avant la date t .*

La fonction de survie, aussi appelée queue de distribution qu'on note par $\bar{F}(t)$, définie sur \mathbb{R}_+ par :

$$\bar{F}(t) = 1 - F(t) = P(X > t)$$

1.2 Notions fondamentales sur les valeurs extrêmes

Nous allons maintenant discuter des concepts primaires et de base spécifiquement liés au sujet de notre recherche.

1.2.1 Statistique d'ordre

Définition 1.2.1 La statistique d'ordre de l'échantillon (X_1, \dots, X_n) est le réarrangement croissant, notée par $(X_{1,n}, \dots, X_{n,n})$ de sorte que $X_{1,n} < X_{2,n} < \dots < X_{n,n}$.

En particulier, la variable aléatoire $X_{i,n}$ où $i = 1, \dots, n$ désigne la $i^{\text{ème}}$ statistiques d'ordre.

Les statistiques d'ordre extrêmes sont définies comme des termes du maximum et du minimum de l'échantillon (X_1, \dots, X_n) .

La variable $X_{1,n}$ est la plus petite statistique d'ordre et $X_{n,n}$ est la plus grande statistique d'ordre, elles sont définies par :

$$X_{1,n} = \min(X_1, X_2, \dots, X_n),$$

$$X_{n,n} = \max(X_1, X_2, \dots, X_n).$$

Lois des statistiques d'ordre

Nous présentons les lois qui apparaissent dans l'étude des échantillons ordonnés.

Corollaire 1.2.1 Soit (X_1, \dots, X_n) une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) d'une densité commune f et d'une fonction de distribution F , alors on a les lois de $X_{i,n}$:

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

La fonction de densité est :

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

où $f(x)$ est la densité de probabilité de X_i et F sa fonction de répartition associée.

En utilisant la propriété d'indépendance des variables aléatoires X_1, \dots, X_n , on obtient :

La fonction de répartition de $X_{1,n}$

$$F_{1,n}(x) = P(X_{1,n} \leq x) = 1 - [1 - F(x)]^n.$$

La fonction de densité de $X_{1,n}$

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

La fonction de répartition de $X_{n,n}$

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

La fonction de densité de $X_{n,n}$

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

Preuve Les expressions de $F_{1,n}$ et $F_{n,n}$ peuvent s'obtenir comme suivant :

$$\{X_{1,n} \geq x\} \Leftrightarrow \{\min(X_1, X_2, \dots, X_n) \geq x\}$$

$$\Leftrightarrow \bigcap_{i=1}^n \{X_i \geq x\}$$

Et

$$\{X_{n,n} \geq x\} \Leftrightarrow \{\max(X_1, X_2, \dots, X_n) \leq x\}$$

$$\Leftrightarrow \bigcap_{i=1}^n \{X_i \leq x\}$$

En utilisant la propriété d'indépendance des variables aléatoires X_1, X_2, \dots, X_n , on déduit que :

$$\begin{aligned} F_{1,n}(x) &= P\{X_{1,n} \leq x\} \\ &= 1 - P\{X_{1,n} \geq x\} \\ &= 1 - P\{\bigcap_{i=1}^n \{X_i \geq x\}\} \\ &= 1 - \prod_{i=1}^n P\{X_i \geq x\} \\ &= 1 - \prod_{i=1}^n [1 - P\{X_i \leq x\}] \\ &= 1 - [1 - F(x)]^n. \end{aligned}$$

Et on déduit la densité :

$$f_{1,n}(x) = \frac{dF_{1,n}(x)}{dx} = nf(x)[1 - F(x)]^{n-1}$$

Et

$$\begin{aligned} F_{n,n}(x) &= P\{X_{n,n} \leq x\} \\ &= P\{\bigcap_{i=1}^n \{X_i \leq x\}\} \\ &= \prod_{i=1}^n P\{X_i \leq x\} \\ &= [F(x)]^n \end{aligned}$$

Et on déduit la densité

$$f_{n,n}(x) = \frac{dF_{n,n}(x)}{dx} = nf(x) [F(x)]^{n-1}$$

■

1.2.2 Fonction de répartition empirique

En statistiques, une fonction de répartition empirique est une fonction de répartition qui attribue la probabilité $1/n$ à chacun des n nombres dans un échantillon.

Définition 1.2.2 *La fonction de répartition empirique associée à l'échantillon (X_1, X_2, \dots, X_n) , notée F_n est évaluée à l'aide des statistiques d'ordre et définie par :*

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum 1 \{X_k \leq x\} \\ &= \frac{1}{n} \text{card} \{k \in \{1, \dots, n\} : X_k \leq x\} \end{aligned}$$

$$F_n(x) = \begin{cases} 0 & \text{si } x < X_{1,n} \\ \frac{1}{n} & \text{si } X_{1,n} \leq x < X_{2,n} \\ \frac{2}{n} & \text{si } X_{2,n} \leq x < X_{3,n} \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & \text{si } x \geq X_{n,n} \end{cases}$$

1.2.3 Inverse généralisée

L'inverse généralisé reste défini même lorsque F n'est pas bijective soit parce que cette fonction est discontinue soit parce qu'elle est constante sur des intervalles d'intérieur non vide.

Définition 1.2.3 Soit (X_1, X_2, \dots, X_n) un échantillon issu d'une fonction de répartition commune F , telle que :

$$F(x) = P(X_i < x) ; i = 1, \dots, n.$$

La fonction inverse généralisée de F définie par :

$$Q(p) = F^{\rightarrow}(p) = \inf \{x, F(x) \geq p\}, \quad 0 < p < 1$$

telle que Q est une fonction des quantiles.

Remarque 1.2.1 L' inverse généralisé F^{\rightarrow} coïncide avec l' inverse F^{-1} lorsque F est strictement croissante et continue.

1.2.4 Fonction des quantiles empiriques

La fonction quantile empirique de la fonction de distribution empirique F_n est la fonction inverse généralisée de F_n .

Définition 1.2.4 La fonction quantile empirique de la fonction de distribution empirique F_n définie par :

$$Q_n(t) = \inf \{x, F_n(x) \geq t\}, \quad 0 < t < 1$$

$$Q_n(t) = \begin{cases} X_{i,n} & \text{si } \frac{i-1}{n} \leq t \leq \frac{i}{n} \\ X_{n,n} & \text{si } \frac{i-1}{n} \leq t \leq \frac{i}{n} \end{cases}$$

1.2.5 Point terminal

Définition 1.2.5 On note par x_F le point extrême supérieur de la distribution F i.e la plus grande valeur possible pour $X_{k,n}$, elle peut prendre la valeur $+\infty$.

$$x_F = \sup \{x \in \mathbb{R} : F(x) < 1\} \leq \infty.$$

1.2.6 Fonction à variation réguliers

Une fonction positive U est à variation réguliers d'indice $\delta \in \mathbb{R}$ à l'infini, et on note $U(\cdot) \in R_{V\delta}$ si pour tout $\lambda > 0$,

$$\lim_{x \rightarrow \infty} \frac{U(\lambda x)}{U(x)} = \lambda^\delta.$$

Dans le cas on dit que la fonction $U(\cdot)$ est à variation lentes, et on note $U \in R_{V0}$.

Toute fonction à variation réguliers d'indice $\delta \in \mathbb{R}$, peut s'écrire sous forme :

$$U(x) = x^\delta L(x).$$

telle que $L(x) \in R_{V0}$.

1.2.7 Lois à queue lourde

Les distributions à queues lourdes sont liées à la théorie des valeurs extrêmes et permettent de modéliser beaucoup de phénomènes que l'on trouve dans différentes disciplines. Les définitions les plus simples basées sur la comparaison avec la distribution normale ont été proposées pour ces distributions.

Définition 1.2.6 On dit que la fonction F est à queue lourde si pour $a > 0, b > 0$

$$\bar{F}(x) > ae^{-b}$$

où $\bar{F} = (1 - F)$ est la fonction de survie.

Corollaire 1.2.2 Si le Kurtosis d'une distribution F vérifie :

$Kurtosis = \frac{\mu_4}{\mu_2^2} > 3$, on peut dire que cette distribution est à queue lourde, μ_4 et μ_2 sont les moments centrés d'ordres respectifs 4 et 2 de X .

Ce qui est équivalent à dire qu'une distribution est à queue lourde si et seulement si son coefficient d'asymétrie est supérieur à celui de la distribution normale.

1.2.8 Les quantiles

Les quantiles d'une variable aléatoire discrète (entière) ou continue (réelle) sont les valeurs que prend cette variable pour des valeurs de probabilité $p(0 < p < 1)$.

On les appelle aussi fractiles, elles sont les valeurs réciproques de la fonction de répartition de la loi de probabilité considérée.

Définition 1.2.7 On appelle quantile d'ordre p , le nombre x_p défini par :

$$x_p = \inf \{x \in \mathbb{R} : F(x) \geq p\} \quad \text{avec } p \in [0, 1].$$

Remarque 1.2.2 Si F est strictement croissante et continue, alors x_p est l'unique nombre réel tel que :

$$F(x_p) = p.$$

1.2.9 Fonction quantile de queue

La fonction quantile de queue est définie par :

Définition 1.2.8 On appelle fonction quantile de queue la fonction :

$$U(t) = F^{-1}\left(1 - \frac{1}{t}\right),$$

avec $1 < t < \infty$ et F^{-1} étant l'inverse généralisée de F .

1.2.10 Quantile extrême

Définition 1.2.9 On appelle quantile extrême le quantile d'ordre $(1 - p)$, défini par :

$$\begin{aligned} x_{1-p} &= \inf \{x \in \mathbb{R} : F(x) \geq 1 - p\} \\ &= F^{-1}(1 - p) \end{aligned}$$

Où p est un nombre proche de zéro.

1.2.11 Quantile empirique

Définition 1.2.10 Soit (X_1, \dots, X_n) un échantillon issu d'une loi F , et $\{X_{(1,n)}, \dots, X_{(n,n)}\}$ l'échantillon ordonné. Soit $p \in]0, 1[$, on appelle quantile empirique d'ordre p , la variable aléatoire notée $\hat{X}_{(np,n)}$ définie par :

$$\hat{X}_{(np,p)} = \begin{cases} \frac{X_{(np,n)} + X_{(np+1,n)}}{2} & \text{si } np \in \mathbb{N} \\ X_{([np]+1,n)} & \text{sinon} \end{cases}$$

où $[np]$ désigne la partie entière de np .

En particulier $X_{([n/2]+1,n)}$ est la médiane empirique.

1.3 La loi des valeurs extrêmes

En probabilité et statistique, la loi d'extrêmes généralisée est une famille de lois de probabilité continues qui servent à représenter des phénomènes de valeurs extrêmes (minimum ou maximum). Elle comprend la loi de Gumbel, la loi de Fréchet et la loi de Weibull, respectivement lois d'extrêmes de type I, II et III.

Théorème 1.3.1 ("Fisher-Tippett (1928)") Soit $X_{n,n}$ le maximum d'un échantillon aléatoire X_1, \dots, X_n de fonction de répartition F . S'il existe deux suites réelles $a_n > 0$ et $b_n \in \mathbb{R}$, appelées constantes de normalisation et une loi non-dégénérée de loi H , telles que :

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_\xi(x) \quad , \quad x \in \mathbb{R}$$

a_n est un paramètre de position.

b_n est un paramètre d'échelle.

ξ est un paramètre de forme.

H appartient à une loi des trois distributions standard des valeurs extrêmes suivant :

1) Distribution de probabilité de Fréchet :

Le mathématicien français Maurice René Fréchet a introduit une distribution considérée comme loi des valeurs extrêmes. Les travaux de Fréchet ont inspiré les recherches de *FisherTippett (1928)*.

Définition 1.3.1 Si $\xi > 0$, on dit que F appartient au domaine d'attraction de Fréchet. Il contient les lois dont la fonction de survie est à décroissance polynomiale comme une fonction puissance, i.e. les lois à queues lourdes. On dit qu'une variable aléatoire X est de distribution de Fréchet de paramètres $\mu, \sigma > 0, \xi > 0$ si sa fonction de distribution cumulative est :

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

La densité de probabilité de la distribution de Fréchet s'écrit :

$$f(x) = \begin{cases} \xi(-x)^{-\xi-1} \exp(-x^{-\xi}) & \text{si } x > 0, \xi > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

et la fonction de survie est :

$$\bar{F}(x) = 1 - F(x) = \begin{cases} 1 - [\exp(-x^{-\xi})] & \text{si } x > 0, \xi > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

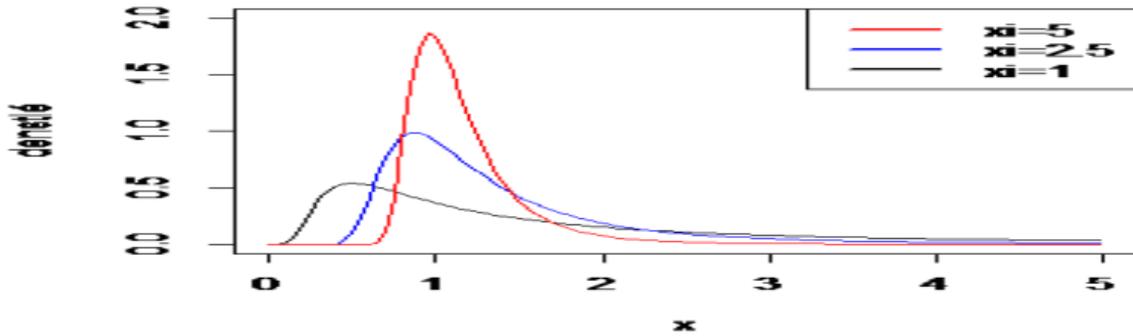


FIG. 1.1 – Densité de la distribution Fréchet

2) Distribution de probabilité de Weibull :

Le mathématicien suédois *Waloddi Weibull (1951)* a introduit la distribution de Weibull, cette dernière est couramment utilisée dans l'analyse de survie et en fiabilité des systèmes.

Définition 1.3.2 Si $\xi < 0$, on dit que F appartient au domaine d'attraction de Weibull. Il contient des lois à queues légères. On dit qu'une variable aléatoire X est de distribution de Weibull de paramètres $\mu, \sigma > 0$, avec $\xi < 0$, si sa fonction de distribution cumulative est :

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

La fonction de densité de probabilité est donnée par la formule suivante :

$$f(x) = \begin{cases} \xi(-x)^{\xi-1} \exp[-((-x)^\xi)] & \text{si } x < 0, \xi < 0 \\ 0 & \text{si } x \geq 0 \end{cases}$$

La fonction de survie de Weibull est :

$$\bar{F}(x) = 1 - F(x) = \begin{cases} 1 - [\exp(-(-x)^\xi)] & \text{si } x < 0, \xi < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$$

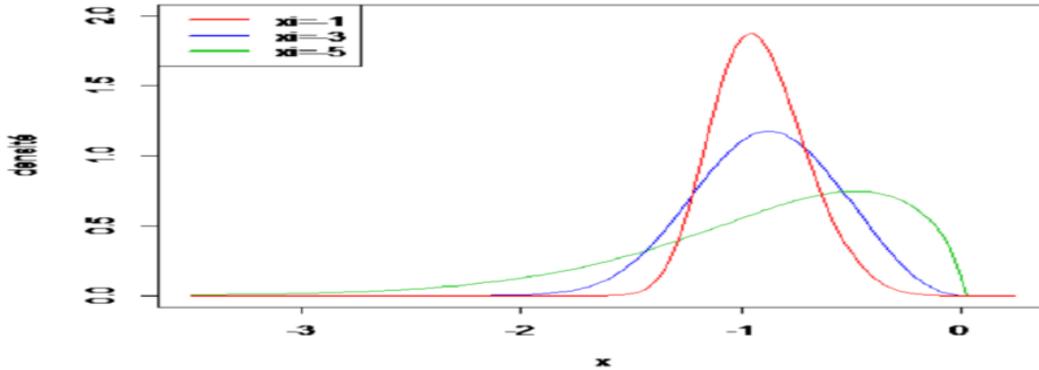


FIG. 1.2 – Densité de la distribution Weibull

3) Distribution de probabilité de Gumbel :

La distribution Gumbel a été introduite par le mathématicien Américain *Emil Julius Gumbel* qui l'a développée comme loi de valeurs extrêmes. La distribution de Gumbel est l'une des distributions les plus populaires dans de nombreux domaines grâce à sa flexibilité.

Définition 1.3.3 Si $\xi = 0$, on dit que F appartient au domaine d'attraction de Gumbel. Il contient les lois dont la fonction de survie est à décroissance exponentielle, i.e il n'y a pas de queue de distribution. On dit qu'une variable aléatoire X est de distribution de Gumbel de paramètres μ et $\sigma > 0$, si seulement si la fonction de distribution cumulative s'écrit :

$$F(x) = \exp(-\exp(-x)), \text{ si } x \in \mathbb{R}.$$

La fonction de densité de probabilité est donnée par la formule suivante :

$$f(x) = \exp(-x) \exp(-\exp(-x)) , \text{ si } x \in \mathbb{R}.$$

La fonction de survie est :

$$\bar{F}(x) = 1 - F(x) = 1 - [\exp(-\exp(-x))], \text{ si } x \in \mathbb{R}.$$

Le théorème de *FisherTippett* montre que la loi limite des extrêmes à toujours la même forme.

Les trois formules précédentes peuvent être combinées en une seule paramétrisation :

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{\frac{-1}{\xi}}) & \xi \neq 0, 1 + \xi x > 0 \\ \exp(-\exp(-x)) & \xi = 0, -\infty \leq x \leq +\infty \end{cases}$$

Où H est une fonction de répartition non-dégénérée et ξ est un paramètre qui contrôle la lourdeur de la queue de la loi, appelée indice des valeurs extrêmes (ou indice de queue), cette loi est appelée loi de valeurs extrêmes généralisée (*GEV*).

1.3.1 La distribution GEV

Pour faciliter le travail avec ces trois distributions limites, *Jenkinson-Von Mises* a donné une représentation, obtenue en introduisant les paramètres de localisation et de dispersion dans la paramétrisation des distributions extrêmes. On appelle cette distribution la distribution généralisée des valeurs extrêmes (*Generalized Extreme Value GEV*), qui a la forme suivante :

$$H_{\mu,\sigma,\xi}(x) = \exp \left[- \left\{ 1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right\}^{\frac{-1}{\xi}} \right], \quad \xi \neq 0, 1 + \xi \frac{x - \mu}{\sigma} > 0.$$

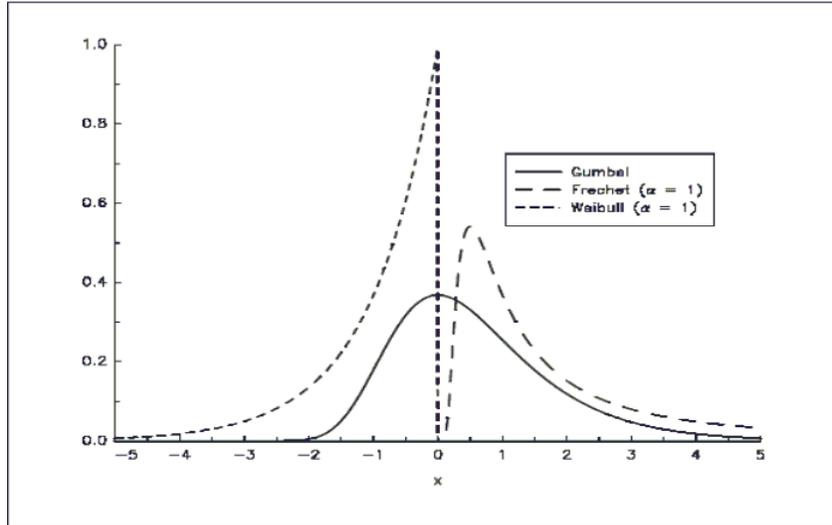


FIG. 1.3 – Fonction de répartition de la distribution GEV

Remarque 1.3.1 *On peut facilement montrer que la fonction de densité de la distribution GEV correspondant à :*

$$h_{\mu,\sigma,\xi}(x) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right)^{\frac{1+\xi}{\xi}} H_{\mu,\sigma,\xi}(x) \right] & \text{si } \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\left(\frac{x-\mu}{\sigma}\right) - \exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right) & \text{si } \xi = 0 \end{cases}$$

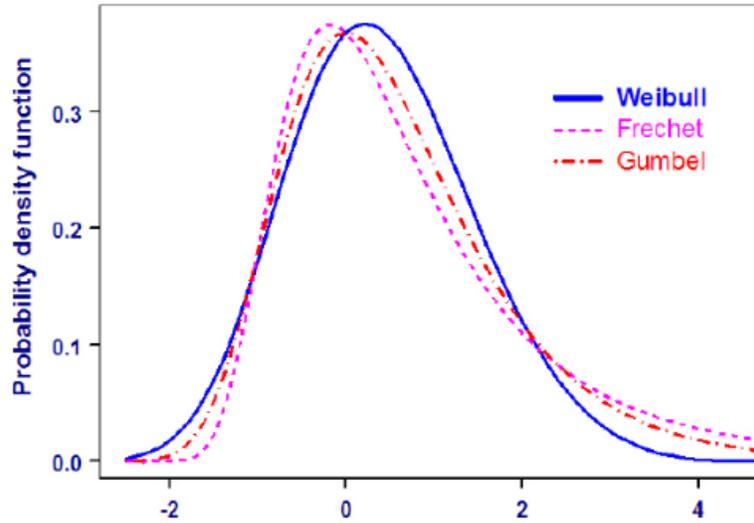


FIG. 1.4 – Fonction densité de la distribution GEV

Remarque 1.3.2 *Le quantile $Q(p)$ de la distribution est donné par la formule suivante :*

$$Q(p) = H_{\mu, \sigma, \xi}^{-1}(x) = \begin{cases} \mu - \sigma \xi^{-1} [1 - (-\log(p))^{-\xi}] & \text{si } \xi \neq 0 \\ \mu - \sigma \log(-\log(p)) & \text{si } \xi = 0 \end{cases}$$

1.3.2 Domaines d'attractions

Le domaine d'attraction regroupe la plupart des distributions ayant un point terminal fini.

Domaine d'attraction de Fréchet

Une fonction de répartition F appartient au domaine d'attraction de Fréchet si et seulement si elle s'écrit sous la forme :

$$F(x) = 1 - x^{\frac{-1}{\xi}} L(x).$$

telle que $L(\cdot) \in R_{V0}$.

Dans ce cas les suites de normalisation a_n et b_n sont données pour tout $n > 0$ par :

$$a_n = \overline{F}^{\rightarrow}\left(\frac{1}{n}\right).$$

et

$$b_n = 0.$$

Domaine d'attraction de Wiebull

Une fonction de répartition F appartient au domaine d'attraction de Wiebull si et seulement si son point terminal x_F est fini et si la fonction répartition $F_*(.)$ définie par :

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

appartient au domaine d'attraction de Fréchet .

Les suites de normalisation a_n et b_n sont données par :

$$a_n = x_F - \overline{F}^{-1}\left(\frac{1}{n}\right).$$

et

$$b_n = x_F.$$

Domaine d'attraction de Gumbel

Une fonction de répartition F appartient au domaine d'attraction de Gumbel si et seulement si il existe $t < x_F \leq \infty$, telle que :

$$\overline{F}(x) = c(x) \exp \left\{ - \int_t^x \frac{1}{a(U)} dU, t < x \leq x_F \right\}$$

Où $c(x) \rightarrow c > 0$ lorsque $x \rightarrow x_F$, et $a(.)$ est une fonction positive et dérivable, de dérivée $a'(.) \rightarrow 0$ lorsque $x \rightarrow x_F$.

Dans ce cas, un choix possible pour les suites a_n et b_n sont données pour tout $n > 0$ par :

$$a_n = Q\left(\frac{1}{n}\right),$$

et

$$b_n = \frac{1}{\overline{F}(a_n)} \int_{a_n}^{x_F} \overline{F}(s) ds.$$

Chapitre 2

Estimation des paramètres de loi GEV

Il existe diverses méthodes qui ont été proposées pour estimer les paramètres des lois *GEV*. Parmi eux on trouve la méthode basée sur le maximum de vraisemblance (Smith1985[33]), la méthode des moments pondérés (Greenwood et al 1979[23]), la méthode des moments (Christopeit 1994[8]), la méthode des L-moments (Hosking1990[24]).

Il existe également des approches semi-paramétriques pour l'estimation de l'indice de queue. Les plus utilisées en pratique sont l'estimateur de Pickands (Pickands1975[32]), l'estimateur de Hill (Hill 1975[26]) pour le cas des lois de type Fréchet uniquement et l'estimateur de Dekkers-Einmahl-de Hann (Dekkers et al 1989[11]). Cependant, suivant le type de problématique que l'on a, on va avoir besoin non seulement d'un estimateur de paramètre de forme, mais aussi d'un estimateur de paramètre d'échelle et estimateur de paramètre de location. C'est typiquement le cas par exemple si l'on s'intéresse à l'estimation d'un quantile extrême.

2.1 Estimation paramétrique

L'estimation paramétrique qui considère que les modèles sont connues avec des paramètres inconnus. L'ensemble des valeurs possibles pour les paramètres, appelé espace paramétrique sera noté Θ lequel est inclus dans \mathbb{R}^d où d est la dimension du paramètre θ .

La loi de la variable étudiée est supposée appartenir à une famille de lois pouvant être caractérisée par une forme fonctionnelle connu.

Nous pouvons estimer les paramètres de la *GEV* par trois méthodes d'estimation sont envisageables : estimation par le Maximum de Vraisemblance (*MLE*), estimation par les Moments (*ME*), estimation par les L-Moments (*LME*).

2.1.1 Méthode de maximum de vraisemblance

L'estimation par maximum de vraisemblance donne des résultats asymptotiques efficaces, les estimateurs obtenus convergent sous certaines conditions vers les vrais valeurs des paramètres.

Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes identiquement distribuées (*iid*). On appelle fonction de vraisemblance pour l'échantillon (X_1, X_2, \dots, X_n) la fonction du paramètre θ :

$$L(\theta, X_1, X_2, \dots, X_n) = \begin{cases} \prod_{i=1}^n P(X, \theta) & (\text{cas discrète}) \\ \prod_{i=1}^n f(X, \theta) & (\text{cas continue}) \end{cases}$$

L'estimation de maximum de vraisemblance de θ est la valeur $\hat{\theta}$ qui rend maximale la fonction de vraisemblance $L(\theta, X_1, X_2, \dots, X_n)$.

Dans la plus part des cas , la fonction de vraisemblance s'exprime comme un produit , donc $\hat{\theta}$ sera en générale calculé par maximisation la log-vraisemblance.

Quand $\theta = (\theta_1, \theta_2, \dots, \theta_d)$ et que toutes les dérivées partielle ci _dessous existe, $\hat{\theta}$ est la solution système d'équations :

$$\begin{cases} \frac{\partial}{\partial \theta_j} \log L(\theta, X_1, X_2, \dots, X_n) = 0, \forall j \in (1, 2, \dots) \\ \frac{\partial^2}{\partial \theta_j^2} \log L(\theta, X_1, X_2, \dots, X_n) \leq 0, \forall j \in (1, 2, \dots) \end{cases}$$

Pour la distribution GEV :

la fonction de log-vraisemblance est donnée par :

Pour $\xi \neq 0$:

$$\log L(\mu, \sigma, \xi) = -n \log(\sigma) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \log\left(1 + \xi \frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^n \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{\frac{-1}{\xi}}$$

et Pour $\xi = 0$:

$$\log L(\mu, \sigma) = -n \log(\sigma) - \sum_{i=1}^n \frac{x_i - \mu}{\sigma} - \sum_{i=1}^n \exp\left[-\left(\frac{x_i - \mu}{\sigma}\right)\right]$$

En prenant la dérivée de $\log L(\mu, \sigma, \xi)$ par rapport à δ, μ et ξ ; on obtient le système d'équations suivant :

Pour $\xi \neq 0$:

$$\frac{\partial \log L(\mu, \sigma, \xi)}{\partial \mu} = \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \frac{\xi}{\sigma + \xi(x_i - \mu)} - \frac{1}{\sigma} \sum_{i=1}^n \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{\frac{1}{\xi} - 1} = 0$$

$$\frac{\partial \log L(\mu, \sigma, \xi)}{\partial \sigma} = -\frac{n}{\sigma} + \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \frac{\xi(x_i - \mu)}{\sigma^2 + \sigma\xi(x_i - \mu)} - \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right) \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{\frac{1}{\xi} - 1} = 0$$

$$\frac{\partial \log L(\mu, \sigma, \xi)}{\partial \xi} = \frac{1}{\xi^2} \sum_{i=1}^n \log\left(1 + \xi \frac{x_i - \mu}{\sigma}\right) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \frac{(x_i - \mu)}{\sigma + \xi(x_i - \mu)}$$

$$+ \sum_{i=1}^n \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{\frac{1}{\xi} - 1} \left\{ \sum_{i=1}^n \frac{(x_i - \mu)}{\sigma\xi + \xi^2\sigma(x_i - \mu)} - \frac{1}{\xi^2} \log\left(1 + \xi \frac{x_i - \mu}{\sigma}\right) \right\} = 0$$

et Pour $\xi = 0$:

$$\frac{\partial \log L(\mu, \sigma)}{\partial \mu} = 0 = n - \left[\sum \exp\left(-\left(\frac{x_i - \mu}{\sigma}\right)\right) \right]$$

$$\frac{\partial \log L(\mu, \sigma)}{\partial \sigma} = 0 = n + \sum \frac{x_i - \mu}{\sigma} \left[\exp\left(-\left(\frac{x_i - \mu}{\sigma}\right)\right) - 1 \right]$$

Le vecteur $\hat{\theta}(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ est solution du système suivant :

$$\hat{\mu} = \begin{cases} \frac{\partial \log L(\mu, \sigma, \xi)}{\partial \mu} = 0 \\ \frac{\partial^2 \log L(\mu, \sigma, \xi)}{\partial \mu^2} \leq 0 \end{cases}$$

$$\hat{\sigma} = \begin{cases} \frac{\partial \log L(\mu, \sigma, \xi)}{\partial \sigma} = 0 \\ \frac{\partial^2 \log L(\mu, \sigma, \xi)}{\partial \sigma^2} \leq 0 \end{cases}$$

$$\hat{\xi} = \begin{cases} \frac{\partial \log L(\mu, \sigma, \xi)}{\partial \xi} = 0 \\ \frac{\partial^2 \log L(\mu, \sigma, \xi)}{\partial \xi^2} \leq 0 \end{cases}$$

La résolution de ce système est relativement difficile et n'admet pas généralement de solutions explicites, dans ce cas on fait appel à des méthodes d'optimisation numérique.

2.1.2 Méthode de moments

les moments de population (moyenne, variance, asymétrie, etc.) sont assimilés à leurs moments d'échantillon correspondants, et les paramètres de la distribution sont résolus algébriquement.

La distribution des valeurs extrêmes généralisées (*GEV*) est une famille de distributions de probabilité continues. La fonction de densité de probabilité de la distribution *GEV* est définie comme :

Pour $\xi \neq 0$:

$$h(x, \mu, \sigma, \xi) = \frac{1}{\sigma} \left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi} - 1} \exp \left[- \left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{\frac{-1}{\xi}} \right]$$

Pour $\xi = 0$:

$$h(x, \mu, \sigma, 0) = \frac{1}{\sigma} \exp\left(-\left(\frac{x - \mu}{\sigma}\right)\right) \exp\left(-\exp\left(-\left(\frac{x - \mu}{\sigma}\right)\right)\right)$$

tq :

μ est le paramètre de localisation.

σ est le paramètre d'échelle.

ξ est le paramètre de forme.

Pour $\xi \neq 0$:

La moyenne $E(X)$ et la variance $Var(X)$ de la distribution *GEV* sont données par :

$$E(X) = \mu + \frac{\sigma}{\xi} (\Gamma(1 - \xi) - 1)$$

$$Var(X) = \frac{\sigma^2}{\xi^2} [\Gamma(1 - 2\xi) - (\Gamma(1 - \xi))^2]$$

tq : $\Gamma(\cdot)$ est la fonction gamma.

La moyenne et la variance de la distribution *GEV* dépendent des paramètres μ , σ et ξ .

Preuve 1/ Moyenne :

Pour trouver la moyenne $E(X)$, nous intégrons le PDF sur toute la plage de x :

$$E(X) = \int_{-\infty}^{+\infty} x h(x) dx$$

En branchant l'expression pour $h(x)$, nous obtenons :

$$E(X) = \int_{-\infty}^{+\infty} x \left(\frac{1}{\sigma} \left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi} - 1} \exp \left[- \left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{\frac{-1}{\xi}} \right] \right) dx$$

cette intégrale peut être assez complexe et peut ne pas avoir de solution de forme fermée. Par conséquent, des méthodes numériques sont souvent utilisées pour

approximer l'intégrale.

Une approche courante consiste à utiliser des techniques d'intégration numérique telles que la règle de Simpson, la quadrature gaussienne ou l'intégration de Monte

Carlo pour approximer l'intégrale sur toute la plage de x . Ces méthodes consistent à diviser la plage d'intégration en intervalles plus petits et à rapprocher l'intégrale dans

chaque intervalle.

Une fois l'intégrale approchée, la valeur résultante donne la moyenne ($E(X)$) de la distribution *GEV*.

Bien que les étapes mathématiques de l'intégration numérique impliquent des techniques informatiques plutôt que des manipulations symboliques, elles fournissent des

approximations précises de la moyenne de la distribution *GEV*.

2/ Variance :

$$Var(X) = \int_{-\infty}^{+\infty} x^2 h(x)$$

En branchant l'expression pour $h(x)$, nous obtenons :

$$Var(X) = \int_{-\infty}^{+\infty} x^2 \left(\frac{1}{\sigma} \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi} - 1} \exp \left[- \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{\frac{-1}{\xi}} \right] \right) dx$$

Semblable au calcul de la moyenne, cette intégrale peut être assez complexe et peut ne pas avoir de solution de forme fermée. Par conséquent, des méthodes numériques sont souvent utilisées pour approximer l'intégrale. ■

Résoudre l'ensemble des équations :

Résoudre l'ensemble des équations impliquant la moyenne et la variance de la distribution *GEV* pour les paramètres μ , σ et ξ , vous avez généralement besoin de méthodes numériques car les équations impliquent des fonctions spéciales comme la fonction gamma. Une approche courante consiste à utiliser des techniques d'optimisation numérique pour trouver les valeurs des paramètres qui minimisent la différence entre la moyenne et la variance observées et la moyenne et la variance théoriques de la distribution *GEV*. Voici un aperçu général de l'approche :

1. Définir les équations pour la moyenne et la variance de la distribution *GEV* en termes de μ, σ et ξ .
2. Choisissez une première estimation pour les valeurs de μ, σ et ξ .

3. Utilisez un algorithme d'optimisation numérique (tel que la méthode Newton-Raphson, l'algorithme de Levenberg-Marquardt ou autres) pour minimiser la différence entre la moyenne et la variance observées et la moyenne et la variance théoriques de la distribution *GEV*.
4. Répétez jusqu'à ce que la convergence soit atteinte, c'est-à-dire jusqu'à ce que la différence entre les moments observés et théoriques soit suffisamment petite.

Pour $\xi = 0$:

$$E(X) = \mu + \sigma.\gamma$$

où γ est la constante d'Euler-Mascheroni, approximativement égale à 0,5772.

$$Var(X) = \frac{\pi^2}{6}.\sigma^2$$

Preuve 1/ Moyenne :

Pour trouver la moyenne $E(X)$, nous intégrons le PDF sur toute la plage de x :

$$E(X) = \int_{-\infty}^{+\infty} xh(x)dx$$

En branchant l'expression pour $h(x)$, nous obtenons :

$$E(X) = \int_{-\infty}^{+\infty} x\left(\frac{1}{\sigma} \exp\left(\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)\right)dx$$

Cette intégrale n'a pas de solution de forme fermée. Cependant, la moyenne $E(X)$ peut être calculée numériquement ou approchée à l'aide de méthodes telles que

l'intégration numérique.

2/ Variance :

$$Var(X) = \int_{-\infty}^{+\infty} x^2h(x)$$

En branchant l'expression pour $h(x)$, nous obtenons :

$$Var(X) = \int_{-\infty}^{+\infty} x^2\left(\frac{1}{\sigma} \exp\left(\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)\right)dx$$

Semblable à la moyenne, cette intégrale n'a pas de solution fermée. Cependant, la variance $Var(X)$ peut être calculé numériquement ou approximé à l'aide de méthodes

numériques. ■

Résoudre un ensemble de deux équations :

Nous pouvons maintenant résoudre cet ensemble simplifié d'équations pour μ et σ .

À partir de l'équation moyenne :

$$\mu = E(X) - \sigma \cdot \gamma$$

Et

$$\sigma^2 = Var(X) \frac{6}{\pi^2}$$

$$\hat{\sigma} = \sqrt{Var(X) \frac{6}{\pi^2}}$$

Remplacez δ dans l'équation moyenne pour trouver μ :

$$\hat{\mu} = E(X) - \sqrt{Var(X) \frac{6}{\pi^2}} \cdot \gamma$$

2.1.3 Méthode de L-moments

Les L-moments sont des combinaisons linéaires des statistiques d'ordres. Ont été introduit et extrait de (Silitto 1951[6]) et revu de manière exhaustive par (Hosking 1990[18]).

Les L-moments sont analogues aux conventionnels moments mais L-moment ont certains avantages notamment : leur existence, leur unicité et leur robustesse.

L-moment ont trouvés de nombreuses applications dans des domaines de la recherche appliquées tq : le génie civil, la météorologie et hydrolique...

Représentation de L-moment sous forme de moment de probabilité pondéré

Les moments pondérés en fonction des probabilités (Probability-weighted moment, noté *PWM*) est une généralisation des moments habituels d'une distribution de probabilité, ont été introduits par (Greenwood et al 1979[23]).

Les moments de probabilité pondérés sont définis par :

$$M_{i,j,k} = E [X^i F^j (1 - F)^k], i, j, k \in \mathbb{R}$$

Nous pouvons récrire en fonction de quantile, alors :

$$M_{i,j,k} = \int_0^1 Q(u)^i u^j (1 - u)^k du$$

Les deux formes suivants de PWM sont particulièrement simples et utiles :

$$\begin{cases} \alpha_k = M_{1,0,k} = E[X(1-F(X))^k] & k \text{ est entier positif.} \\ \beta_j = M_{1,j,0} = E[XF^j(X)] & j \text{ est entier positif.} \end{cases}$$

Donc :

$$\begin{cases} \alpha_k = \int_0^1 Q(u)(1-u)^k du \\ \beta_j = \int_0^1 Q(u)^j u^j du \end{cases}$$

Les α_k et β_j ont été utilisées comme base de méthode de L-moment pour l'estimation des paramètres de distribution de probabilité car : il existe des relations simples entre les moments de probabilité pondérés α_k , β_j et les paramètres de distribution de probabilité. Donc : il est possible d'exprimer les paramètres de loi de probabilité comme des fonction PWM . D'autre part, pour calculer les L-moments.

On écrit : L-moment comme des combinaisons linéaires de PWM :

$$\lambda_r = (-1)^{r-1} \sum_{k=0}^{r-1} P_{r,k}^* \alpha_k$$

Donc :

$$\lambda_1 = \alpha_0 = \beta_0$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$$

Pour la loi de valeur extrême généralisé (GEV) :

Soit X v.a suit la loi GEV des paramètres μ, σ, ξ .

Pour trouver les L-moments on utilisons la forme de PWM j ,

On a :

$$\beta_j = \frac{1}{j+1} \left[\mu - \frac{\sigma}{\xi} \left(1 + \frac{1}{(j+1)^\xi} \Gamma(1-\xi) \right) \right], \xi > 1$$

Nous calculons β_j , $j = 0, \dots, 3$ et remplacent dans λ_r pour trouver :

$$\lambda_1 = \mu - \frac{\sigma}{\xi}(1 + \Gamma(1 - \xi))$$

$$\lambda_2 = \frac{\sigma}{\xi}(1 - 2^\xi)\Gamma(1 - \xi)$$

$$\lambda_3 = \frac{\sigma}{\xi}(3 * 2^{-\xi} - 2 * 3^{-\xi} - 1)\Gamma(1 - \xi)$$

$$\lambda_4 = \frac{\sigma}{\xi}(1 - 6 * 2^{-\xi} + 10 * 3^{-\xi} - 5 * 4^{-\xi})\Gamma(1 - \xi)$$

Alors :

par L-moments des paramètres du modèle *GEV* lorsque $-\frac{1}{2} < \xi < \frac{1}{2}$ sont donnés par :

$$\hat{\xi} \approx 7.8590c + 2.9554c^2$$

$$\hat{\sigma} = \frac{l_2 \hat{\xi}}{(1 - 2^{-\hat{\xi}})(\Gamma(1 - \hat{\xi}))}$$

$$\hat{\mu} = l_1 + \frac{\hat{\delta}(1 + \Gamma(1 - \hat{\xi}))}{\hat{\xi}}$$

où

$$c = \frac{2}{3 + \tau_3} - \frac{\log 2}{\log 3}$$

et l_1, l_2 sont respectivement les estimateurs des L-moments d'ordre 1 et 2 et du rapport des L-moments .

Ces estimateurs peuvent être définis à partir des moments de probabilité pondérés :

$$\beta_r = E(XH^r(x))$$

Un estimateurs sans biais de $\beta_r (r > 0)$ est :

$$b_0 = \frac{1}{n} \sum_{i=1}^n X_i$$

$$b_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-3)} X_i$$

Et les estimateurs des quatre premiers L-moments sont :

$$l_1 = b_0$$

$$l_2 = 2b_1 - b_0$$

$$l_3 = 6b_2 - 6b_1 + b_0$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$$

En divisant les L-moments d'ordre supérieure à 2 par la mesure de dispersion l_2 , on obtient les estimateurs des rapports des L-moments $\tau_r = l_r/l_2$ avec $r = 3, 4$.

En pratique :

Le coefficient de l'asymétrie τ_3 peut s'écrire d'une manière directe de la façon suivante (L-skwenes) est :

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = \sum_{i=1}^n \frac{c_i x_i + \frac{\bar{X}}{n}}{l_2}$$

où

$$c_i = 6 \frac{(i-1)(i-2)}{n(n-1)(n-2)} - 6 \frac{i-1}{n(n-1)}$$

et \bar{X} est la moyenne empirique de l'échantillon.

2.2 Estimation semi- paramétrique

Des procédures statistiques semi-paramétrique appropriées à cette situation, n'exigent pas la connaissance entière de la distribution, mais seulement, sur les queues de distribution.

Les estimateurs classiques sont basés sur les plus grandes statistiques d'ordre $X_{n-k,n} \leq \dots \leq X_{n,n}$ où k est une suite intermédiaire d'entiers liés à la taille de l'échantillon n de la façon suivante :

$$k = k(n) \rightarrow \infty \quad \text{quand } n \rightarrow \infty$$

2.2.1 Estimation de Hill

Cet estimateur a été introduit par (Hill 1975[26]) de forme semi paramétrique, il est basé sur les statistiques d'ordre qui n'est utilisable que pour les distributions appartenant au domaine de Fréchet, l'estimateur de Hill est probablement l'estimateur le plus étudié dans la littérature.

Un grand nombre de travaux a été consacré à l'étude de la normalité asymptotique de l'estimateur de Hill, (Davis et Resnick 1984[14]) et (Haeusler et Teugels

1985[27]) et pour les propriétés de l'estimateur, (Mason1982[31]), (Deheuvels et al 1988[15]).

Définition 2.2.1 Soit $(k(n))_{n \geq 1}$ une suite d'entiers avec $1 < k(n) < n$, l'estimateur de Hill est donné par la forme empirique suivante :

$$\widehat{\xi}_{(k(n),n)}^{(H)} = \frac{1}{k} \sum_{i=1}^k \log X_{n-i+1,n} - \log X_{n-k+1,n}$$

Preuve On peut justifier la démonstration par la caractérisation :

$$\overline{F}(x) = x^{-\frac{1}{\xi}} l(x)$$

avec l une fonction à variation lente.

Nous avons deux parties : $x^{-\frac{1}{\xi}}$ la partie paramétrique et $l(x)$ la partie non paramétrique.

Pour $t > 0$;

$$\lim_{u \rightarrow \infty} \frac{l(tu)}{l(u)} = 1$$

Avec $t > 0$;

$$\lim_{u \rightarrow \infty} \frac{\overline{F}(tu)}{\overline{F}(u)} = t^{-\frac{1}{\xi}} \left(\lim_{u \rightarrow \infty} \frac{l(tu)}{l(u)} \right) = t^{-\frac{1}{\xi}}$$

On déduit l'approximation :

$$\overline{F}(tu) \approx t^{-\frac{1}{\xi}} \overline{F}(u)$$

Dans l'approximation nous remplaçons tu par x on trouve

$$\overline{F}(x) = \overline{F}(u) \left(\frac{x}{u} \right)^{-\frac{1}{\xi}}$$

La densité $f(x)$ est comme suit :

$$f(x) = \frac{1}{\xi} \frac{\overline{F}(u)}{u} \left(\frac{x}{u} \right)^{-\frac{1}{\xi} - 1}$$

L'expression de la fonction de vraisemblance est :

$$L(x_1, x_2, \dots, x_k; \xi) = \prod_{i=1}^k f(x_i)$$

Maximiser la vraisemblance L revient à résoudre le système suivant :

$$\begin{cases} \frac{\partial \log L(x_1, x_2, \dots, x_k; \xi)}{\partial \xi} = 0 \\ \frac{\partial^2 \log L(x_1, x_2, \dots, x_k; \xi)}{\partial \xi^2} < 0 \end{cases}$$

en effet

$$\begin{aligned} \log L(x_1, x_2, \dots, x_k; \xi) &= \log(\prod_{i=1}^k f(x_i)) \\ &= \sum_{i=1}^k \log f(x_i) \\ &= \sum_{i=1}^k \log\left(\frac{1}{\xi} \frac{\bar{F}(u)}{u} \left(\frac{x_i}{u}\right)^{\frac{1}{\xi}-1}\right) \\ &= \sum_{i=1}^k \left[\log \bar{F}(u) - \log(\xi u) - \left(\frac{1}{\xi} + 1\right) \log\left(\frac{x_i}{u}\right) \right] \end{aligned}$$

Résoudre le système revient à résoudre l'équation suivant :

$$\begin{aligned} \sum_{i=1}^k \left[\frac{-u}{\xi u} + \frac{1}{\xi^2} \log\left(\frac{x_i}{u}\right) \right] &= 0 \\ -k\xi + \sum_{i=1}^k \log \frac{x_i}{u} &= 0 \end{aligned}$$

La solution de cette équation est de la forme :

$$\hat{\xi}^{(H)} = \frac{1}{k} \sum_{i=1}^k \log \frac{x_i}{u} = \frac{1}{k} \sum_{i=1}^k [\log x_i - \log u]$$

On remplace x_i et u par leurs équivalents dans la statistique d'ordre on trouve la formule de l'estimateur de Hill :

$$\hat{\xi}_{(k(n),n)}^{(H)} = \frac{1}{k} \sum_{i=1}^k [\log X_{n-i+1,n} - \log X_{n-k+1,n}]$$

■

Propriétés de l'estimateur de Hill

1 / Convergence en probabilité

Théorème 2.2.1 Soit X_1, \dots, X_n un échantillon de variables aléatoires i.i.d de fonction de réparation $F \in MDA(H_\xi)$, $\xi > 0$.

Soit $(k(n))_{n>1}$ une suite d'entiers telle que $1 < k(n) < n$, si

$$\frac{k(n)}{n} \rightarrow 0 \quad \text{et} \quad k(n) \rightarrow \infty$$

alors :

$$\widehat{\xi}_{((k(n)),n)}^{(H)} \xrightarrow{p} \xi \quad , \quad n \rightarrow \infty$$

2 / Convergence forte (convergence presque sûre) Si

$$\frac{k(n)}{n} \rightarrow 0 \quad \text{et} \quad \frac{k(n)}{\log \log(n)} \rightarrow \infty \quad , \quad n \rightarrow \infty$$

alors :

$$\widehat{\xi}_{(k(n),n)}^{(H)} \xrightarrow{ps} \xi \quad , \quad n \rightarrow \infty$$

3 / Normalité asymptotique

Théorème 2.2.2 [26] Soit $(k(n))_{n>1}$ une suite d'entiers telle que $1 < k(n) < n$, si $\frac{k(n)}{n} \rightarrow 0$ et $k(n) \rightarrow \infty$, si la condition de l'ordre est satisfaite avec

$$k(n)A\left(\frac{n}{k(n)}\right) \rightarrow 0 \quad \text{quand} \quad n \rightarrow \infty$$

alors :

$$k(n)^{\frac{1}{2}}(\widehat{\xi}_{(k(n),n)}^{(H)} - \xi) \rightarrow N(0, \xi^2)$$

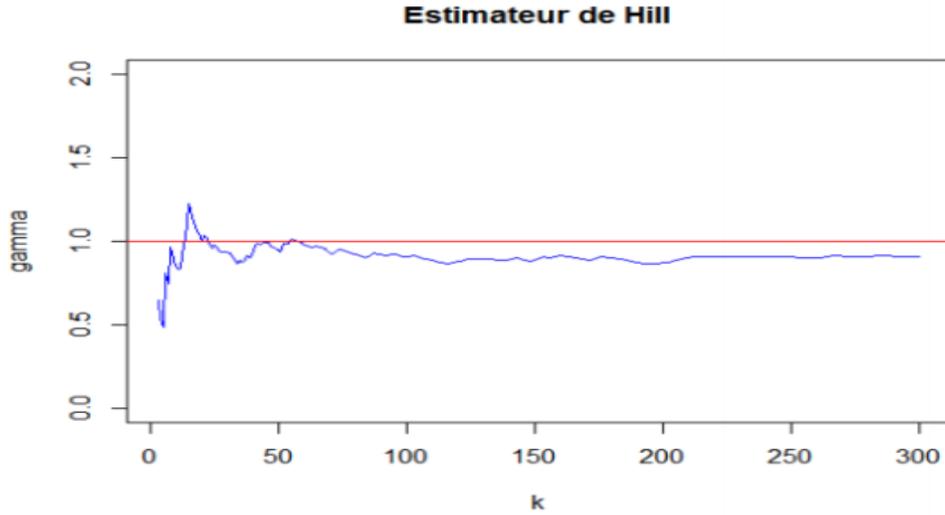


FIG. 2.1 – Représentation graphique de l’estimation de Hill .

2.2.2 Estimation de Pickands

L’estimateur de Pickands est construit en utilisant les statistiques d’ordres, il a l’avantage d’être valable quel que soit le domaine d’attraction de la distribution et par conséquent du domaine de définition de l’indice des valeurs extrêmes.

Cet estimateur a été introduit par (Pickands et al 1975[32]) et revisité (Drees 1996[17]), (Drees and Kaufmann 1998[16]) d’une part, (Dekkers et al 1989[10]) d’autre part, ils ont établi sa consistance faible ainsi que sa normalité asymptotique.

Définition 2.2.2 *On suppose que $\{X_i, i = 1, \dots, n\}$ une suite de variables aléatoires indépendantes de loi F appartenant à l’un des domaines d’attractions.*

Soit $(k(n))_{n \geq 1}$ une suite d’entiers et $1 < kn < n$ un estimateur de Pickands est défini par :

$$\widehat{\xi}_{(k(n),n)}^{(p)} = \frac{1}{\log 2} \log \left(\frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}} \right)$$

Propriétés de l’estimateur de Pickands

1 / Convergence en probabilité

Théorème 2.2.3 [32] Soit X_1, \dots, X_n un échantillon de variables aléatoires i.i.d de fonction de répartition $F \in MDA(H_\xi)$, si

$$n \rightarrow \infty, k(n) \rightarrow \infty \text{ et } \frac{k(n)}{n} \rightarrow \infty$$

alors

$$\widehat{\xi}_{(k(n),n)}^{(p)} \xrightarrow{p} \xi$$

2 / Convergence forte (convergence presque sûre)

Théorème 2.2.4 [32] Si

$$\frac{k}{n} \rightarrow 0 \text{ et } \frac{k(n)}{\log \log(n)} \rightarrow \infty, n \rightarrow \infty$$

alors :

$$\widehat{\xi}_{(k(n),n)}^{(p)} \xrightarrow{p.s} \xi, n \rightarrow \infty$$

3 / Normalité asymptotique

Théorème 2.2.5 [32] Sous les conditions additionnelles sur la suite $k(n)$ et la fonction de répartition F on a :

$$k(\widehat{\xi}_{(k(n),n)}^{(p)} - \xi) \rightarrow N(0, v(\xi))$$

avec

$$k(n)A\left(\frac{n}{k(n)}\right) \rightarrow 0 \text{ quand } n \rightarrow \infty$$

où

$$v(\xi) = \frac{\xi^2(2^{2\xi+1} + 1)}{(2(2^\xi - 1) \log 2)^2}$$

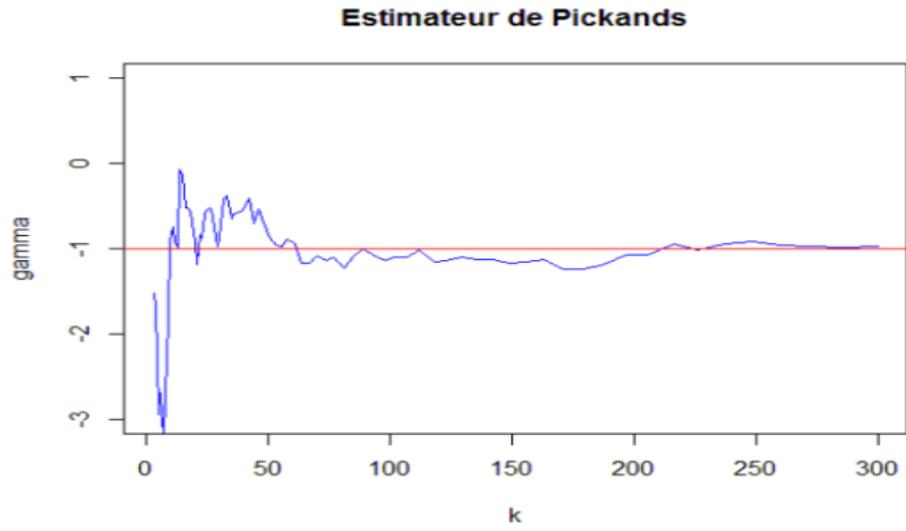


FIG. 2.2 – Représentation graphique de l'estimateur de Pickands .

2.2.3 Estimateur du moment

Cet estimateur introduit par (Dekkers et al 1989[11]) est une généralisation de l'estimateur de Hill, valable pour tout $\xi \in \mathbb{R}$:

$$\widehat{\xi}_{(k(n),n)}^{(M)} = \widehat{\xi}_{(k(n),n)}^{(H)} + 1 - \frac{1}{2} \left(1 - \frac{\widehat{\xi}_{(k(n),n)}^{(H)}}{M_n^{(2)}}\right)^{-1}$$

où

$$M_n^{(r)} := \frac{1}{k} \sum_{i=0}^k (\log X_{n-i+1,n} - \log X_{n-k+1,n})^r; r = 1, 2.$$

Théorème 2.2.6 [11] (*Propriétés asymptotiques de $\widehat{\xi}_{(k(n),n)}^{(M)}$*)

Supposons que $F \in D(H_\xi)$, $\xi \in \mathbb{R}$, $k \rightarrow \infty$ et $\frac{k}{n} \rightarrow 0$ quand $n \rightarrow \infty$:

– **Consistance faible**

$$\widehat{\xi}_{(k(n),n)}^{(M)} \xrightarrow{p} \xi, \text{ quand } n \rightarrow \infty.$$

– **Consistance forte**

$$\widehat{\xi}_{(k(n),n)}^{(M)} \xrightarrow{p.s} \xi, \text{ quand } n \rightarrow \infty.$$

– **Normalité asymptotique** pour $\xi \geq 0$:

$$\sqrt{k}(\widehat{\xi}_{(k(n),n)}^{(M)} - \xi) \xrightarrow{d} N(0, \eta^2), \text{ quand } n \rightarrow \infty$$

tq :

$$\eta^2 = \begin{cases} 1 + \xi^2 & , \xi \geq 0 \\ (1 - \xi^2)(1 - 2\xi)(4 - 8\frac{1-2\xi}{1-3\xi} + \frac{(5-11\xi)(1-2\xi)}{(1-3\xi)(1-4\xi)}) & , \xi < 0 \end{cases}$$

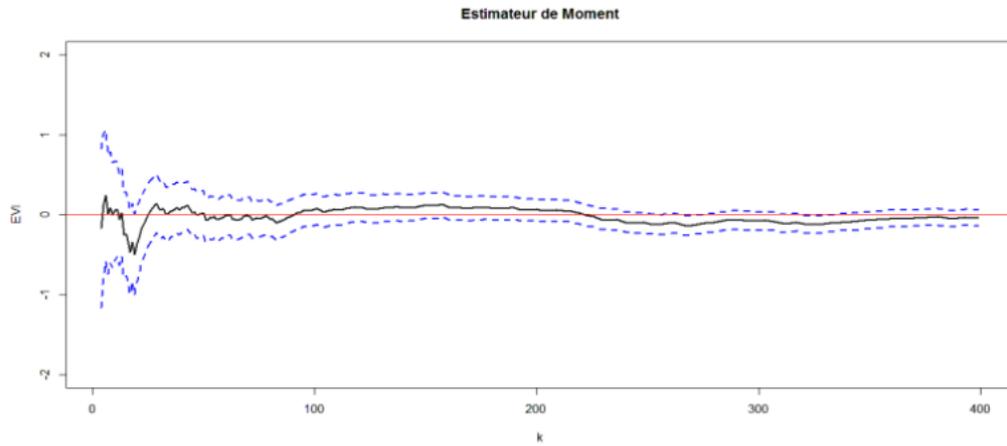


FIG. 2.3 – Représentation graphique de l'estimateur de Moment .

2.2.4 Le choix du nombre k

La difficulté consiste à choisir le nombre k de statistiques d'ordre extrême à utiliser. Les résultats concernant les estimateurs de l'indice des valeurs extrêmes sont asymptotiques : ils sont obtenus lorsque $k \rightarrow \infty$ et $\frac{k}{n} \rightarrow 0$. Comme en pratique, on ne dispose que d'un nombre d'observations n fini, il s'agit de choisir k de manière à ce que l'on dispose de suffisamment de matériel statistique tout en restant dans la queue de distribution ($k < n$).

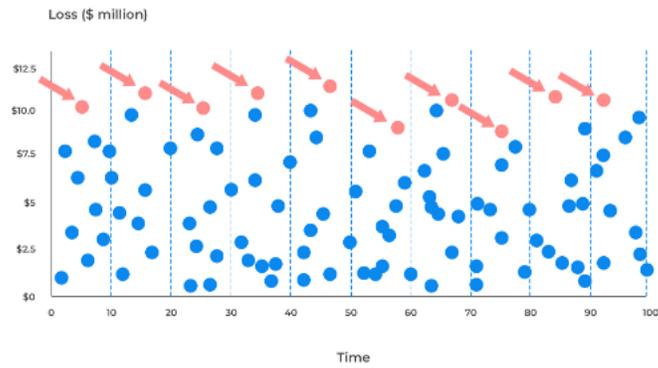


FIG. 2.4 – Diagramme schématique de la méthode des blocs maxima

Méthode graphique

C'est une méthode la plus simple pour la détermination de k . Elle consiste à tracer le graphe $(k, \widehat{\xi}_{(k(n),n)})$ et de prendre la valeur où $(k, \widehat{\xi}_{(k(n),n)})$ devient horizontal. Cet estimateur est valable seulement dans le domaine d'attraction de Fréchet c'est -à- dire si $\xi > 0$.

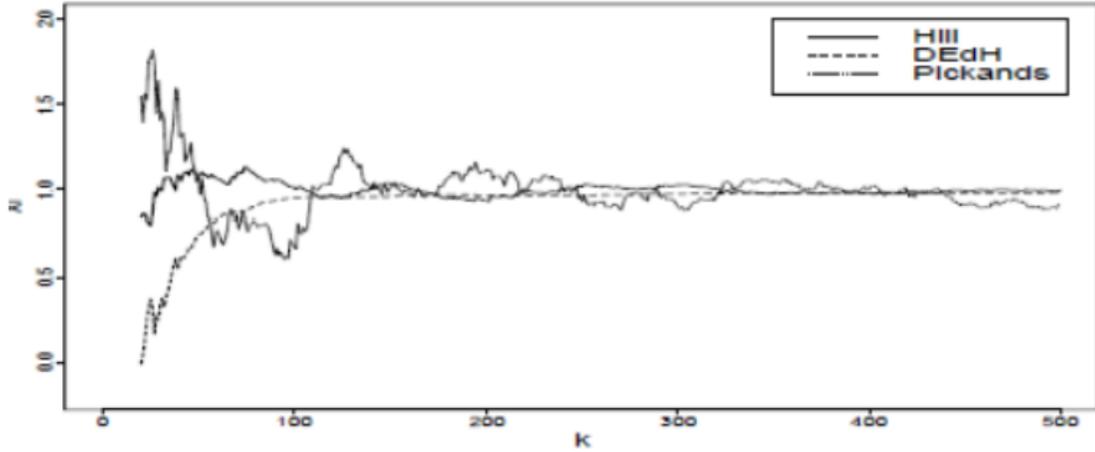


FIG. 2.5 – Le graphe de $(k, \hat{\xi}_n)$ pour les différents estimateur de ξ

Méthode analytique

Il est nécessaire pour donner une précision à l'estimateur $\hat{\xi}_{(k(n),n)}$, de calculer l'erreur quadratique moyenne (MSE), elle est en fonction de k :

$$\begin{aligned} MSE(\hat{\xi}_{(k(n),n)}) &= E(\hat{\xi}_{(k(n),n)} - \xi)^2 \\ &= \text{Biais}^2(\hat{\xi}_{(k(n),n)}) + \text{Var}(\hat{\xi}_{(k(n),n)}) \end{aligned}$$

Le choix optimal de k , correspond à minimiser la MSE .

2.3 Estimation des quantiles extrêmes

Estimation des quantiles extrêmes joue un rôle important dans le contexte de la gestion des risques où il est crucial d'évaluer de manière adéquate le risque d'une grande perte qui se produit très rarement.

On considère X_1, X_2, \dots, X_n la réalisation de n variables réelles indépendantes et de fonction de répartition commune F supposée continue.

Pour $0 < p < 1$, le quantile d'ordre $(1 - p)$ de la fonction de distribution F noté x_p est défini comme étant la solution de l'équation :

$$1 - F(x) = p$$

On définit les quantiles extrêmes par :

$$x_{1-p} = F^{\leftarrow}(1-p) = Q(1-p) \quad \text{quand } p \rightarrow 0$$

On veut estimer le quantile d'ordre p où p est strictement inférieur à $1/n$.

L'estimateur de quantile extrême obtenu par la méthode des blocks s'écrit sous la forme :

$$\hat{x}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left[1 - (-\log(1-p))^{-\hat{\xi}} \right] & , \quad \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log(-\log(1-p)) & , \quad \hat{\xi} = 0 \end{cases}$$

où $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ et sont estimateurs des paramètres de la loi *GEV*.

Supposons maintenant que nous disposions d'un quantile x_p et que nous désirions estimer la période de retour T , associée à la probabilité d'excès au delà de x_p . Cette période s'interprète comme une moyenne d'unités de temps séparant un événement de grandeur donnée x_p d'un second événement d'une grandeur égale ou supérieure.

2.4 Estimation de période de retour

Nous définissons le niveau de retour comme la valeur x_T telle que nous espérons détecter en moyenne un seul dépassement de cette quantité au bout de T périodes c'est -à -dire :

$$\begin{aligned} E(\sum_{i=1}^T 1_{x_i > x_T}) &= 1 \\ \Leftrightarrow P(X_i > x_T) &= \frac{1}{T} \\ \Leftrightarrow 1 - F(x_T) &= \frac{1}{T} \end{aligned}$$

L'estimateur d'un niveau de retour d'ordre T revient à l'estimation d'un quantile extrême d'ordre $p = 1 - \frac{1}{T}$.

Un estimateur d'ordre \hat{p}^{GEV} du quantile x_p pour la loi des *GEV* est donné par :

$$\hat{p}^{GEV} = \begin{cases} \exp\left(-\left(1 + \hat{\xi}\left(\frac{x_p - \hat{\mu}}{\hat{\sigma}}\right)\right)^{-\frac{1}{\hat{\xi}}}\right) & si \quad \hat{\xi} \neq 0 \\ \exp\left(-\exp\left(-\left(\frac{x_p - \hat{\mu}}{\hat{\sigma}}\right)\right)\right) & si \quad \hat{\xi} = 0 \end{cases}$$

d'où :

$$\hat{T}^{GEV} = \begin{cases} \left(\left(1 - \exp\left(-\left(1 + \hat{\xi}\left(\frac{x_p - \hat{\mu}}{\hat{\sigma}}\right)\right)^{-\frac{1}{\hat{\xi}}}\right)\right)^{-1} & si \quad \hat{\xi} \neq 0 \\ \left(\left(1 - \exp\left(-\exp\left(-\left(\frac{x_p - \hat{\mu}}{\hat{\sigma}}\right)\right)\right)\right)^{-1} & si \quad \hat{\xi} = 0 \end{cases}$$

$$\hat{T}^{GEV} = \frac{1}{1 - H_{\hat{\xi}}(x_p)} \quad \forall \hat{\xi} \in \mathbb{R}$$

Chapitre 3

APPLICATION

Dans ce chapitre, nous appliquerons tous les concepts théoriques liés à l'estimation des paramètres et des quantités que nous avons évoqués précédemment dans le chapitre précédent afin d'obtenir des résultats précis et utiles dans la réalité.

3.1 Description et explication des données :

Nous donnerons une idée générale et brève des données sur lesquelles nous sommes appuyés dans notre étude appliquée.

Ces données représentent les valeurs de corrosion dans les oléoducs appartenant au puits RB66, qui ont été mesurées par la sonde chaque semaine tout au long de la période 2001 à 2009.

Ce sont de vraies valeurs que nous avons reprises de la :

Activité Amont

Division Production

Direction Régionale de Rhourde el Baguel

Division maintenance

Service méthode

Voyons ensuite un petit aperçu de la corrosion et des avantages que les mesures continues par sonde apportent pour éviter les risques.

La corrosion étant considérée comme l'un des plus gros problèmes de production des grandes entreprises , la réalisation d'une étude statistique à visée prévisionnelle permettra d'éviter de nombreuses pertes matérielles , que ce soit en termes de coûts d'équipements et de canalisations ou en termes d'arrêt temporaire de production dans les pires cas.

3.1.1 Explique la sonde de corrosion :



FIG. 3.1 – Une illustration de la corrosion et comment la mesurer

Une sonde de corrosion est un type de dispositif de surveillance de la réponse électrochimique utilisé pour enregistrer des données en direct sur le taux de corrosion dans un environnement chimique donné.

Les sondes sont généralement placées en contact direct avec des objets métalliques ou des structures sujettes à la corrosion.

Ils sont souvent utilisés pour extraire des lectures de résistance électrique dans les environnements de traitement pétrolier, chimique et industriel.

Les sondes de corrosion sont utilisées pour les activités de maintenance proactive destinées à prévenir les pannes catastrophiques des matériaux et des équipements.

De telles défaillances peuvent entraîner des coûts importants si elles ne sont pas corrigées.

Les données sont résumés dans les tableaux suivants :

1	<u>Date Probe Read</u>	<u>Corrosion Rate (mm/yr)</u>
2	9-29-01	0.0409
3	10-6-01	0.0444
4	10-13-01	0.1011
5	10-20-01	0.0688
6	10-27-01	0.1195
7	11-3-01	0.0790
8	11-10-01	0.0225
9	11-17-01	0.0449
10	11-24-01	0.0695
11	12-1-01	0.0518
12	12-8-01	0.0000
13	12-15-01	0.0310
14	12-22-01	0.0258
15	12-29-01	0.0000
16	1-5-02	0.0000
17	1-12-02	0.0172
18	1-19-02	0.0257
19	1-26-02	0.0439
20	2-2-02	0.0000
21	2-9-02	0.0511
22	2-16-02	0.0381
23	2-23-02	0.0304
24	3-2-02	0.0000
25	3-9-02	-0.0217

FIG. 3.2 – Tableau des données 1

26	3-16-02	-0.3524
27	3-23-02	0.0000
28	3-30-02	0.7682
29	4-6-02	-0.1009
30	4-13-02	-0.0253
31	4-20-02	0.0379
32	4-27-02	0.1134
33	5-4-02	0.0000
34	5-11-02	0.0000
35	5-18-02	0.0000
36	5-25-02	-0.2013
37	6-1-02	-0.2376
38	6-8-02	0.1697
39	6-15-02	-0.7197
40	6-22-02	-0.1659
41	6-29-02	0.0208
42	7-6-02	-0.0557
43	7-13-02	-0.0277
44	7-20-02	0.3413
45	7-27-02	0.0226
46	8-3-02	0.0350
47	8-10-02	0.0000
48	8-17-02	0.0448
49	8-24-02	0.0261
50	8-31-02	0.0446

FIG. 3.3 – Tableau des données 2

51	9-7-02	-0.0260
52	9-14-02	0.0467
53	9-21-02	0.0194
54	9-28-02	0.0000
55	10-5-02	0.0442
56	10-12-02	0.0220
57	10-19-02	0.0257
58	10-26-02	0.0220
59	11-2-02	0.0438
60	11-9-02	0.0191
61	11-16-02	0.0457
62	11-23-02	0.0673
63	11-30-02	0.0431
64	12-7-02	0.1070
65	12-14-02	0.2033
66	12-21-02	0.0000
67	12-28-02	0.1216
68	1-4-03	0.0869
69	1-11-03	0.1152
70	1-18-03	0.1249
71	1-25-03	0.1475
72	2-1-03	0.2049
73	2-8-03	0.2681
74	2-15-03	0.1505
75	2-22-03	0.1302

FIG. 3.4 – Tableau des données 3

76	3-1-03	0.0806
77	3-8-03	0.0427
78	3-15-03	0.0547
79	3-22-03	0.0477
80	3-29-03	0.0422
81	4-5-03	0.0505
82	4-12-03	0.0315
83	4-19-03	0.0358
84	4-26-03	0.0536
85	5-3-03	0.0208
86	5-10-03	0.0207
87	5-17-03	0.0000
88	5-24-03	0.0000
89	5-31-03	0.0000
90	6-7-03	0.0355
91	6-14-03	0.0177
92	6-21-03	0.1077
93	6-28-03	0.0349
94	7-5-03	0.0000
95	7-12-03	0.0348
96	7-19-03	0.0304
97	7-26-03	0.1607
98	8-2-03	0.1195
99	8-9-03	0.0893
100	8-16-03	0.0395

FIG. 3.5 – Tableau des données 4

101	8-23-03	0.0354
102	8-30-03	0.0940
103	9-6-03	0.0000
104	9-13-03	0.0130
105	9-20-03	0.0000
106	9-27-03	0.0000
107	10-4-03	0.0389
108	10-11-03	0.0000
109	10-18-03	0.3231
110	10-25-03	0.0711
111	11-1-03	0.1128
112	11-8-03	0.0639
113	11-15-03	0.0448
114	11-22-03	0.0123
115	11-29-03	0.0317
116	12-6-03	0.0000
117	12-13-03	0.0123
118	12-20-03	0.0000
119	12-27-03	-0.0184
120	1-3-04	0.0158
121	1-10-04	0.0138
122	1-17-04	0.0000
123	1-24-04	0.0158
124	1-31-04	0.0785
125	2-7-04	0.0384

FIG. 3.6 – Tableau des données 5

126	2-14-04	-0.0273
127	2-21-04	0.0448
128	2-28-04	0.0000
129	3-6-04	0.0149
130	3-13-04	0.0297
131	3-20-04	0.0148
132	3-27-04	0.0148
133	4-3-04	0.0295
134	4-10-04	0.0295
135	4-17-04	-0.0295
136	4-24-04	0.0295
137	5-1-04	0.0129
138	5-8-04	-0.1205
139	5-15-04	0.0928
140	5-22-04	0.0851
141	5-29-04	0.0677
142	6-5-04	0.0577
143	6-12-04	0.0717
144	6-19-04	0.0143
145	6-26-04	-0.0125
146	7-3-04	0.0428
147	7-10-04	0.0000
148	7-17-04	0.0331
149	7-24-04	-0.0142
150	7-31-04	0.0142

FIG. 3.7 – Tableau des données 6

152	8-14-04	0.0142
153	8-21-04	0.0000
154	8-28-04	0.0142
155	9-4-04	0.0141
156	9-11-04	0.0000
157	9-18-04	0.0000
158	9-25-04	0.0282
159	10-2-04	0.0141
160	10-9-04	-0.0141
161	10-16-04	0.0000
162	10-23-04	0.0000
163	10-30-04	0.0141
164	11-6-04	-0.0164
165	11-13-04	-0.1661
166	11-20-04	0.2491
167	11-27-04	-0.0141
168	12-4-04	-0.0424
169	12-11-04	0.0142
170	12-18-04	0.0000
171	12-25-04	0.0564
172	1-1-05	0.0281
173	1-8-05	0.0000
174	1-15-05	-0.0422
175	1-22-05	0.0141
176	1-29-05	0.0000

FIG. 3.8 – Tableau des données 7

177	2-5-05	0.0141
178	2-12-05	0.0000
179	2-19-05	-0.0846
180	2-26-05	-0.0853
181	3-5-05	0.1855
182	3-12-05	0.0000
183	3-19-05	0.0140
184	3-26-05	0.0000
185	4-2-05	0.0140
186	4-9-05	0.0279
187	4-16-05	0.0139
188	4-23-05	0.0121
189	4-30-05	0.0000
190	5-7-05	0.0061
191	5-14-05	-0.5409
192	5-21-05	0.0098
193	5-28-05	-0.1074
194	6-4-05	0.0372
195	6-11-05	0.1148
196	6-18-05	0.0140
197	6-25-05	-0.0109
198	7-2-05	0.0276
199	7-9-05	0.0161
200	7-16-05	-0.8957
201	7-23-05	0.2531

FIG. 3.9 – Tableau des données 8

202	7-30-05	0.0498
203	8-6-05	-0.2520
204	8-13-05	0.8243
205	8-20-05	0.0695
206	8-27-05	0.0604
207	9-3-05	-0.0414
208	9-10-05	0.1100
209	9-17-05	0.0000
210	9-24-05	-0.4371
211	10-1-05	0.0840
212	10-8-05	-0.4236
213	10-15-05	0.2724
214	10-22-05	0.3213
215	10-29-05	-0.6518
216	11-5-05	1.2227
217	11-12-05	-0.0642
218	11-19-05	-0.3071
219	11-26-05	0.1840
220	12-3-05	0.4795
221	12-10-05	0.0267
222	12-17-05	0.0000
223	12-24-05	0.0000
224	12-31-05	-0.0134
225	1-7-06	-0.0268
226	1-14-06	0.0134

FIG. 3.10 – Tableau des données 9

227	1-21-06	0.0134
228	1-28-06	-0.0402
229	2-4-06	0.0134
230	2-11-06	-0.0087
231	2-18-06	-0.0235
232	2-25-06	0.0403
233	3-4-06	-0.0313
234	3-11-06	0.0000
235	3-18-06	-0.0289
236	3-25-06	0.0671
237	4-1-06	0.0000
238	4-8-06	-0.0402
239	4-15-06	0.0134
240	4-22-06	0.0000
241	4-29-06	-0.0157
242	5-6-06	-0.0118
243	5-13-06	-0.0404
244	5-20-06	-0.0542
245	5-27-06	-0.0272
246	6-3-06	-0.0273
247	6-10-06	0.1625
248	6-17-06	-3.5620
249	6-24-06	2.0308
250	7-1-06	-0.1334
251	7-8-06	-0.4814

FIG. 3.11 – Tableau des données 10

252	7-15-06	0.0428
253	7-22-06	-2.0702
254	7-29-06	1.7418
255	8-5-06	-0.0155
256	8-12-06	-0.0544
257	8-19-06	-0.8650
258	8-26-06	1.0402
259	9-2-06	0.0134
260	9-9-06	-0.0117
261	9-16-06	-0.0234
262	9-23-06	0.1861
263	9-30-06	-0.0551
264	10-7-06	-0.0989
265	10-14-06	0.4333
266	10-21-06	-0.0152
267	10-28-06	0.0152
268	11-4-06	0.0000
269	11-11-06	0.0130
270	11-18-06	0.0000
271	11-25-06	-0.0130
272	12-2-06	0.0101
273	12-9-06	0.0000
274	12-16-06	0.0113
275	12-23-06	0.0101
276	12-30-06	-0.0181

FIG. 3.12 – Tableau des données 11

277	1-6-07	-0.0182
278	1-13-07	0.0000
279	1-20-07	0.0091
280	1-27-07	0.0000
281	2-3-07	0.0000
282	2-10-07	0.0000
283	2-17-07	0.0000
284	2-24-07	0.0130
285	3-3-07	0.0000
286	3-10-07	0.0000
287	3-17-07	0.0129
288	3-24-07	-0.0129
289	3-31-07	0.0000
290	4-7-07	-0.0130
291	4-14-07	-0.0781
292	4-21-07	0.0114
293	4-28-07	0.0910
294	5-5-07	0.0000
295	5-12-07	0.0258
296	5-19-07	0.0129
297	5-26-07	0.0000
298	6-2-07	-0.0129
299	6-9-07	0.0258
300	6-16-07	0.0000
301	6-23-07	-0.0129

FIG. 3.13 – Tableau des données 12

302	6-30-07	0.0000
303	7-7-07	0.0000
304	7-14-07	0.0000
305	7-21-07	0.0129
306	7-28-07	0.0000
307	8-4-07	0.0129
308	8-11-07	-0.0129
309	8-18-07	0.0257
310	8-25-07	-0.0338
311	9-1-07	0.0000
312	9-8-07	-0.0100
313	9-15-07	0.0000
314	9-22-07	-0.0151
315	9-29-07	0.0751
316	10-6-07	0.0000
317	10-13-07	-0.0128
318	10-20-07	0.0257
319	10-27-07	0.0000
320	11-3-07	-0.0128
321	11-10-07	0.0000
322	11-17-07	0.0000
323	11-24-07	0.0150
324	12-1-07	0.0221
325	12-8-07	0.0193
326	12-15-07	0.0221

FIG. 3.14 – Tableau des données 13

327	12-22-07	0.0221
328	12-29-07	0.0127
329	1-5-08	0.0000
330	1-12-08	0.0000
331	1-19-08	0.0000
332	1-26-08	0.0127
333	2-2-08	-0.0127
334	2-9-08	0.0127
335	2-16-08	-0.0127
336	2-23-08	0.0127
337	3-1-08	0.0000
338	3-8-08	-0.0127
339	3-15-08	0.0000
340	3-22-08	0.0127
341	3-29-08	0.0000
342	4-5-08	0.0111
343	4-12-08	0.0295
344	4-19-08	-0.0507
345	4-26-08	-0.0381
346	5-3-08	0.0381
347	5-10-08	-0.0127
348	5-17-08	0.0253
349	5-24-08	-0.0127
350	5-31-08	0.0000
351	6-7-08	-0.2179

FIG. 3.15 – Tableau des données 14

352	6-14-08	-0.0303
353	6-21-08	0.2403
354	6-28-08	0.0445
355	7-5-08	0.0127
356	7-12-08	0.0253
357	7-19-08	-0.0128
358	7-26-08	0.0128
359	8-2-08	0.0000
360	8-9-08	0.0000
361	8-16-08	0.0000
362	8-23-08	0.0000
363	8-30-08	0.0000
364	9-6-08	-0.0128
365	9-13-08	0.0111
366	9-20-08	-0.0148
367	9-27-08	0.0000
368	10-4-08	-0.0127
369	10-11-08	0.0253
370	10-18-08	0.0000
371	10-25-08	-0.0128
372	11-1-08	0.0000
373	11-8-08	0.0128
374	11-15-08	-0.0128
375	11-22-08	0.0000
376	11-29-08	-0.0380

FIG. 3.16 – Tableau des données 15

377	12-6-08	0.0254
378	12-13-08	0.0000
379	12-20-08	0.0127
380	12-27-08	-0.0380
381	1-3-09	0.0254
382	1-10-09	-0.0148
383	1-17-09	0.0222
384	1-24-09	0.0000
385	1-31-09	-0.0127
386	2-7-09	0.0253
387	2-14-09	0.0504
388	2-21-09	0.0000
389	2-28-09	0.0251
390	3-7-09	0.0125
391	3-14-09	-0.0219
392	3-21-09	-0.0147

FIG. 3.17 – Tableau des données 16

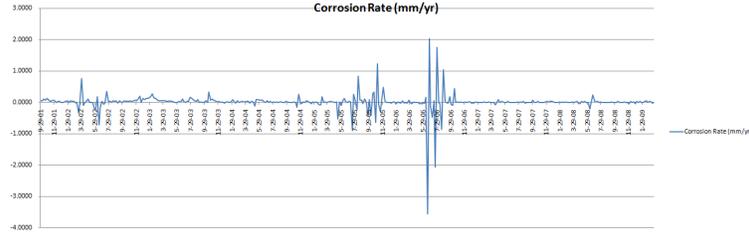


FIG. 3.18 – Représentation graphique des données à l'aide d'Excel

3.2 Analyse statistique des données :

```

>install.packages("extRemes")
>library(extRemes)
>data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
>summary(data)
>library(moments)
>skew <- skewness(data)
>print(skew)
>kurt <- kurtosis(data)
>print(kurt)
>mean_value <- mean(data)
>sd_value <- sd(data)
>print(paste("Mean (moyenne) :", mean_value))
>print(paste("Standard Deviation (écart type) :", sd_value))

```

	Les données de corrosion
Péroïde	29/09/2001 – 21/03/2009
Nombre d'observations	391
Moyenne	0.01134
Ecart type	0.29412
Max	2.0308
Min	-3.562
Q1	0
Médiane	0.0123
Q3	0.03545
Asymétrie(Skewness)	-4.005661
Aplatissement(Kurtosis)	73.37052

TAB. 3.1 – Analyse statistique des données.

La kurtosis est supérieur à 3 donc la distribution à queue lourde.

3.3 Test de normalité :

```
>data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
>shapiro.test(data)
data : data
p-value = 2.2e-16
```

Grâce à ce test, nous pouvons déterminer si nos données suivent ou non la loi normale.

3.3.1 Interprétation :

Si p-value est inférieure à $\alpha = 0.05$ on rejette l'hypothèse nulle H_0 (les données suivent la loi normale) et on accepte l'hypothèse H_1 (les données ne suivent pas la loi normale).

Il est clair que nos données ne suivent pas la loi normale car *p - value* = 2.2×10^{-16} est inférieure à $\alpha = 0.05$.

3.4 Modélisation de la distribution GEV :

```
>install.packages("extRemes")
>library(extRemes)
>install.packages("fitdistrplus")
>library(fitdistrplus)
>library(moments)
> boxplot(data, main = "Boxplot of Data")
> q1 <- quantile(data, 0.25)
> q3 <- quantile(data, 0.75)
> iqr <- q3 - q1
> upper_bound <- q3 + 1.5 * iqr
> lower_bound <- q1 - 1.5 * iqr
> cleaned_data <- data[data >= lower_bound & data <= upper_bound]
> gev_fit_cleaned <- fevd(cleaned_data, type = "GEV", method = "MLE")
> summary(gev_fit_cleaned)
fevd(x = cleaned_data, type = "GEV", method = "MLE")
[1] "Estimation Method used : MLE"
Negative Log-Likelihood Value : -717.5538
Estimated parameters :
location                    scale                    shape
0.001107761 0.023452458 -0.150905337
Standard Error Estimates :
```

```
location          scale          shape
0.0014520121 0.0009946831 0.0342675940
AIC = -1429.108
BIC = -1417.831
> plot(gev_fit_cleaned)
```

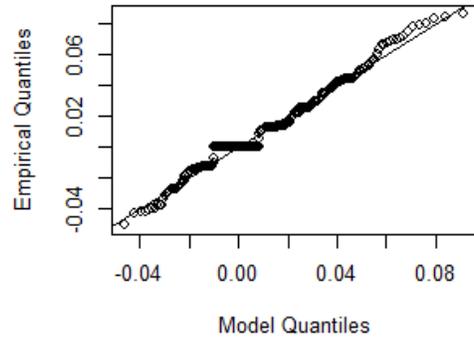


FIG. 3.19 – Correspondance les quantiles des données avec les quantiles de GEV

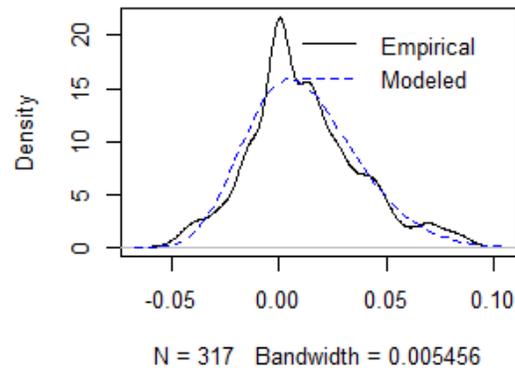


FIG. 3.20 – Ajuster la fonction de densité des données à la fonction de densité GEV

3.5 Estimation :

3.5.1 Estimation paramétrique :

Méthode de maximum vraisemblance :

```
>install.packages("extRemes")
>library(extRemes)
>summary(gev_fit_mle)
Paramètres Estimation par la méthode MLE
 $\hat{\mu}$           -0.1117854
 $\hat{\sigma}$        0.4974721
 $\hat{\xi}$           -0.2263024
```

Méthode des moments :

```
>library(moments)
> data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
>sample_mean <- mean(data)
>sample_variance <- var(data)
>sample_skewness <- skewness(data)
>location_param <- sample_mean - 0.5772 * sample_variance / sample_mean
>scale_param <- sqrt(sample_variance)
>shape_param <- 0.147 * (sample_skewness + 0.845)
>cat("Estimated Location Parameter (mu) :", location_param, "\n")
>cat("Estimated Scale Parameter (sigma) :", scale_param, "\n")
>cat("Estimated Shape Parameter (xi) :", shape_param, "\n")
Paramètres Estimation par la méthode ME
 $\hat{\mu}$           -4.390694
 $\hat{\sigma}$        0.2941283
 $\hat{\xi}$           -0.4646172
```

Méthode des L-moments :

```
>install.packages("extRemes")
>library(extRemes)
>data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
>library(moments)
>lmom_params <- fevd(data, method = "Lmoments")
>lmom_params
```

Paramètres	Estimation par la méthode LME
$\hat{\mu}$	-0.03396561
$\hat{\sigma}$	0.14094370
$\hat{\xi}$	-0.33266425

Conclusion :

Sur la base de ce que nous avons obtenu des résultats d'estimation des paramètres de distribution GEV avec différentes méthodes statistiques que nous avons comparés, notamment en ce qui concerne le paramètre de forme, qui avait la même valeur négative avec les trois méthodes d'estimation, ce qui nous fait conclure que nos données suivent non seulement la distribution GEV, mais suivent spécifiquement la distribution Weibull.

Cela est logique étant donné que nos données relèvent du domaine de la mécanique, de la fiabilité et de la maintenance, car la distribution de Weibull y est courante.

Ce code dans R confirme la fiabilité de nos conclusions :

```
>library(MASS)
>data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
>min_value <- min(data)
>data_positive <- data + abs(min_value) + 0.01
>fit_weibull <- fitdistr(data_positive, "weibull")
>hist(data_positive, prob = TRUE, main = "Histogram with Fitted Weibull
Distribution")
>curve(dweibull(x, shape = fit_weibull$estimate[1], scale = fit_weibull$estimate[2]),
col = "blue", lwd = 2, add = TRUE, yaxt = "n")
```

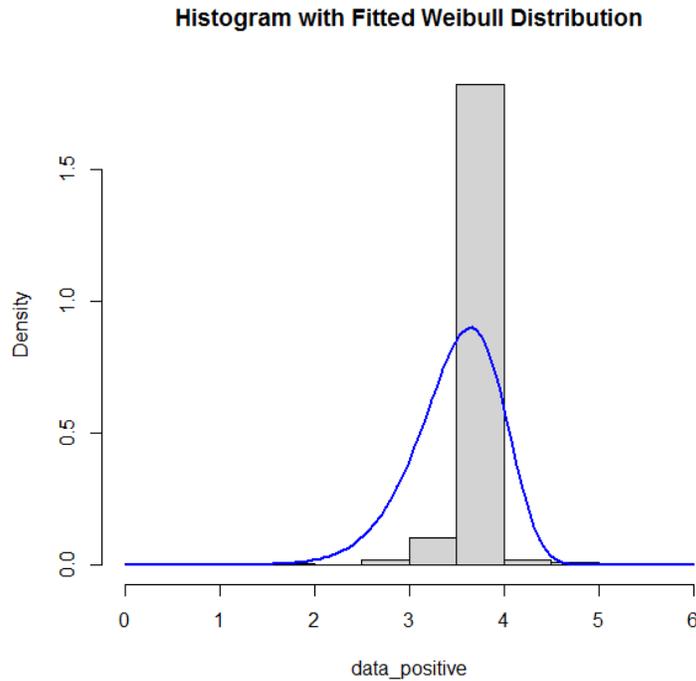


FIG. 3.21 – Histogramme avec distribution de Weibull ajustée

3.5.2 Estimation semi-paramétrique :

Estimateur de Pickands :

```

> install.packages("extRemes")
> library(extRemes)
> data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
> threshold <- -1.01
> exceedance_probs <- (n - (1 : n - threshold)) / (n + 0.4)
> log_ratio <- log(exceedance_probs[-n] / exceedance_probs[-1])
> lm_fit <- lm(log_ratio ~ sorted_data[-n])
> shape_param <- coef(lm_fit)[2]
> print(paste("Threshold :", threshold))
[1] "Threshold : -1.01"
> print(paste("Shape Parameter :", shape_param))
[1] "Shape Parameter : -0.21814515995686"

```

Estimateur de moments :

```
> moments_estimator <- function(data) {
+ valid_data <- data[data > 0]
+ if (length(valid_data) < 3) {
+ return(NA)
+ }
+ n <- length(valid_data)
+ s <- sum(log(valid_data))
+ shape_parameter <- exp(s/n) + (-0.2181452 - exp(s/n))
+ return(shape_parameter)
+ }
> shape_param <- moments_estimator(data)
> print(shape_param)
[1] -0.2181452
```

3.6 Estimation des quantiles extrêmes(niveau de retour) :

3.6.1 Méthode de maximum vraisemblance :

```
> install.packages("extRemes")
> library(extRemes)
> library(evd)
> data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
> return.level(mle_params, return.period = c(2, 5, 10, 25, 50, 100), plot =
TRUE)
fevd(x = data, method = "MLE")
get(paste("return.level.fevd.", newcl, sep = ""))(x = x,
return.period = return.period, plot = TRUE)
GEV model fitted to data
Data are assumed to be stationary
[1] "Return Levels for period units in years"
 2-year level  5-year level  10-year level  25-year level
0.06318783  0.52094266   0.76545765   1.02056455
50-year level  100-year level
1.17742313   1.31028679
> return_levels <- return.level(mle_params, return.period = return_periods)
> plot(return_periods, return_levels, type = "b", xlab = "Return Period",
ylab = "Return Level", col = "blue")
```

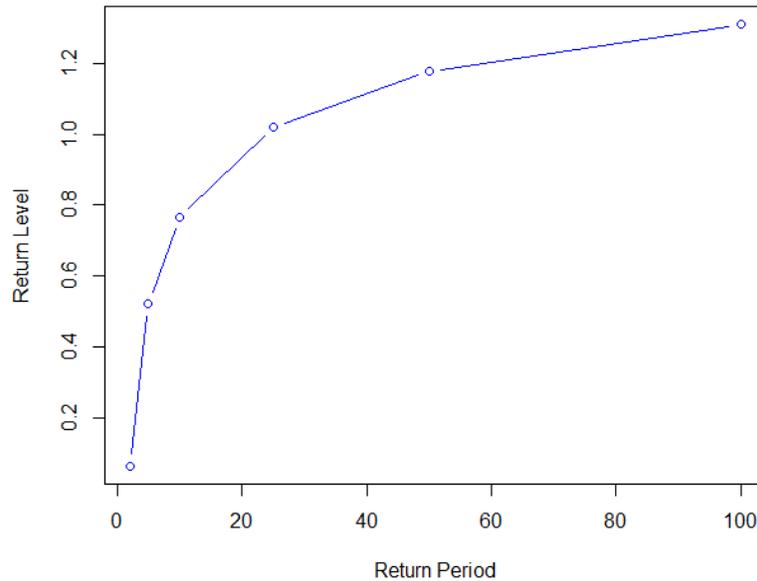


FIG. 3.22 – Niveau de retour par MLE

3.6.2 Méthode des moments :

```

> install.packages("extRemes")
> library(extRemes)
> library(moments)
> data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
> sample_mean <- mean(data)
> sample_variance <- var(data)
> sample_skewness <- skewness(data)
> location_param <- sample_mean - 0.5772 * sample_variance / sample_mean
> scale_param <- sqrt(sample_variance)
> shape_param <- 0.147 * (sample_skewness + 0.845)
> cat("Estimated Location Parameter (mu) :", location_param, "\n")
Estimated Location Parameter (mu) : -4.390694
> cat("Estimated Scale Parameter (sigma) :", scale_param, "\n")
Estimated Scale Parameter (sigma) : 0.2941283
> cat("Estimated Shape Parameter (xi) :", shape_param, "\n")
Estimated Shape Parameter (xi) : -0.4646172
> return_periods <- c(2, 5, 10, 25, 50, 100)
> p <- 1 - 1/return_periods

```

```
> return_levels <- location_param - (scale_param / shape_param) * (1 -  
(-log(1 - p))(-shape_param))  
> print(return_levels)  
[1] -4.291571 -4.547346 -4.690320 -4.847396 -4.950751 -5.044693  
plot(return_periods, return_levels, type = "b", xlab = "Return Period", ylab =  
"Return Level", col = "blue")
```

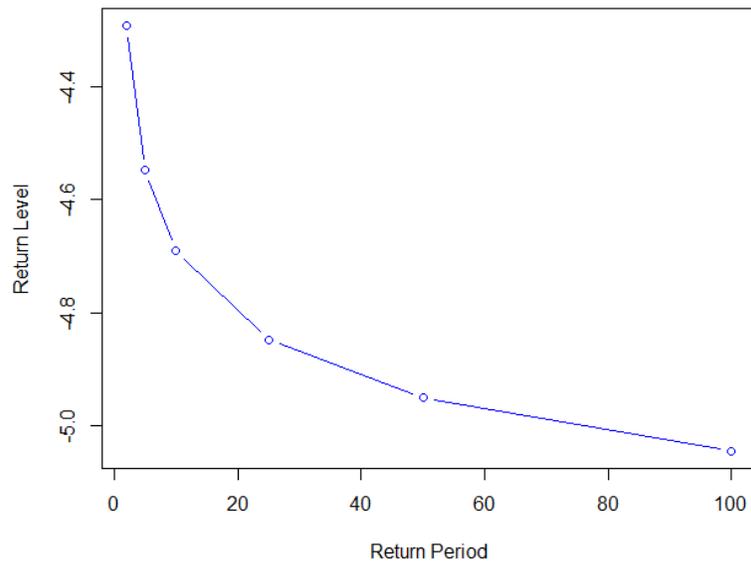


FIG. 3.23 – Niveau de retour par ME

3.6.3 Méthode des L-moments :

```
> install.packages("extRemes")
> library(extRemes)
> library(evd)
> data <- c(0.0409, 0.0444, 0.1011, 0.0688, 0.1195, 0.0790, 0.0225, ... )
> lmom_params <- fevd(data, method = "Lmoments")
> return.level(lmom_params, return.period = c(2, 5, 10, 25, 50, 100), plot =
TRUE)
fevd(x = data, method = "Lmoments")
get(paste("return.level.fevd.", newcl, sep = ""))(x = x,
return.period = return.period, plot = TRUE)
GEV model fitted to data
Data are assumed to be stationary
[1] "Return Levels for period units in years"
2-year level    5-year level    10-year level    25-year level
0.01466706     0.13247691     0.18930592     0.24352053
50-year level   100-year lev
0.27402218     0.29800176
> return_levels <- return.level(lmom_params, return.period = return_periods)
> plot(return_periods, return_levels, type = "b", xlab = "Return Period",
ylab = "Return Level", col = "blue")
```

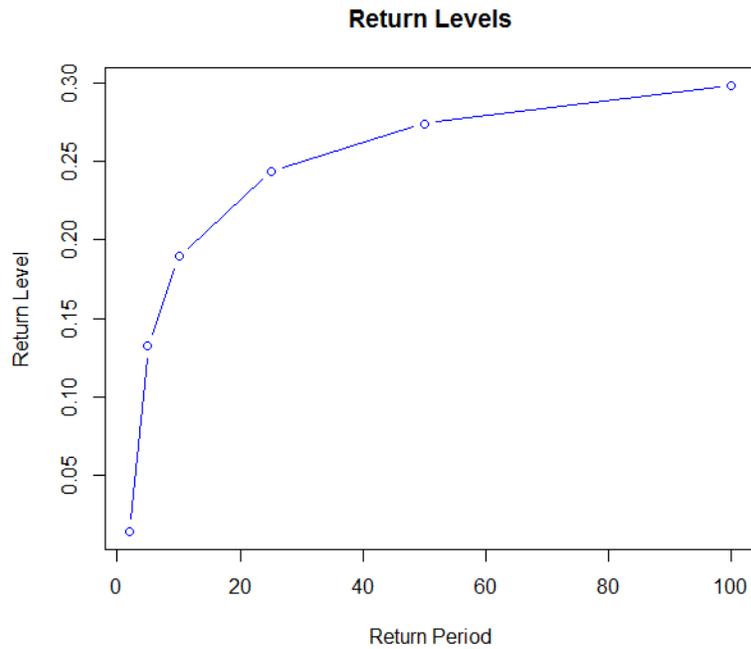


FIG. 3.24 – Niveau de retour par LME

Conclusion 3.6.1 *L'estimation du niveau de retour par les trois méthodes d'estimation MLE, ME et LME permet de pédiere les risques majeurs que la corrosion peut entraîner pour les entreprises, Surtout Sonatrach.*

3.7 Simulation

Dans cette étude de simulation, nous explorons et comprenons le comportement des valeurs extrêmes en utilisant la distribution GEV. Nous générons des données synthétiques pour modéliser ces événements et analysons leurs propriétés à l'aide du framework GEV. Nos objectifs incluent la génération de données synthétiques, l'estimation des paramètres de la distribution GEV, l'analyse des caractéristiques des valeurs extrêmes, et la validation du modèle en le comparant à des données réelles. Cette étude vise à approfondir notre compréhension de la théorie des valeurs extrêmes et à appliquer la distribution GEV dans divers contextes où la prévision et la gestion des risques extrêmes sont cruciales.

3.7.1 Généraliser et simuler un échantillon de données pour une distribution GEV :

```
> install.packages("extRemes")
> library(extRemes)
> install.packages("ismev")
> library(ismev)
> set.seed(123)
> n <- 1000
> shape <- 0.5
> scale <- 1
> location <- 0
> rgev_alternative <- function(n, loc, scale, shape) {
+ u <- runif(n)
+ x <- loc + (scale / shape) * ((-log(u))(-shape) - 1)
+ return(x)
+ }
> simulated_data <- rgev_alternative(n, loc = location, scale = scale, shape
= shape)
> hist(simulated_data, breaks = 30, freq = FALSE, main = "Simulated Data
from GEV Distribution", xlab = "Value")
```

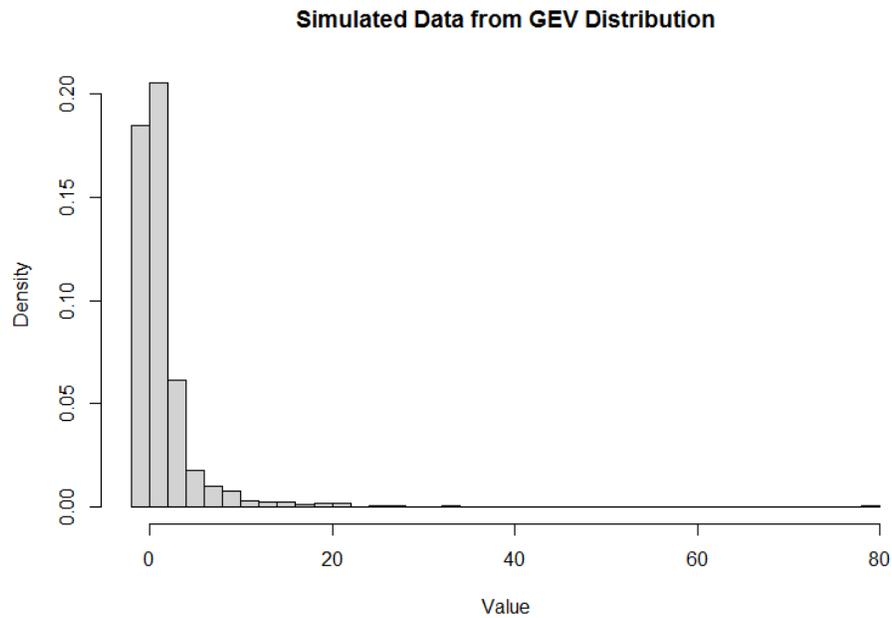


FIG. 3.25 – Données simulées de la distribution GEV

```

> fit <- gev.fit(simulated_data)
$conv
[1] 0
$nllh
[1] 1847.745
$mle
[1] -0.004158118  0.990181445  0.484675991
$se
[1] 0.03552985  0.03382240  0.03013757
> print(fit)
$trans
[1] FALSE
$model
$model[[1]]
NULL
$model[[2]]
NULL
$model[[3]]
NULL
$link

```

```

[1] "c(identity, identity, identity)"
$conv
[1] 0
$nullh
[1] 1847.745
$mle
[1] -0.004158118 0.990181445 0.484675991
$cov
  [,1] [,2] [,3]
[1,] 0.0012623700 0.0008710331 -0.0002433405
[2,] 0.0008710331 0.0011439551 0.0001582125
[3,] -0.0002433405 0.0001582125 0.0009082730
$se
[1] 0.03552985 0.03382240 0.03013757
$vals
  [,1] [,2] [,3]
[1,] -0.004158118 0.9901814 0.484676
[2,] -0.004158118 0.9901814 0.484676
[3,] -0.004158118 0.9901814 0.484676...
attr("class")
[1] "gev.fit"

```

3.7.2 Estimation des paramètres et des quantiles extrêmes pour la loi GEV :

```

> install.packages("lmomco")
> library(lmomco)
> install.packages("isnev")
> library(isnev)
> library(evd)
> library(MASS)
> estimate_gev_parameters <- function(sample_data) {
+ mu_moments <- mean(sample_data)
+ sigma_moments <- sd(sample_data)
+ xi_moments <- (mean(sample_data) - median(sample_data)) / sd(sample_data)
+ fit_output <- capture.output(fit_mle <- gev.fit(sample_data)$mle)
+ mu_mle <- fit_mle[1];
+ sigma_mle <- fit_mle[2];
+ xi_mle <- fit_mle[3];
+ lmoments <- lmoms(sample_data)
+ mu_lm <- lmoments$lambda[1] # First L-moment (mean)

```

```

+ sigma_lm <- sqrt(lmoments$lambda[2]) # Second L-moment (variance)
+ xi_lm <- (lmoments$lambda[3] / lmoments$lambda[2]) * 1.5 # Third
L-moment ratio (shape parameter)
+ results <- list(
+ Method_of_Moments = list(mu = mu_moments, sigma = sigma_moments,
xi = xi_moments),
+ Maximum_Likelihood_Estimation = list(mu = mu_mle, sigma = sigma_mle,
xi = xi_mle),
+ L_moments_Estimation = list(mu = mu_lm, sigma = sigma_lm, xi =
xi_lm)
+ )
+ return(results)
+ }
> sample_data <- rgev(n = 1000, loc = 0, scale = 1, shape = 0.5)
> estimated_params <- estimate_gev_parameters(sample_data)
> print(estimated_params)
$Method_of_Moments
$Method_of_Moments$mu
[1] 1.34973
$Method_of_Moments$sigma
[1] 3.150646
$Method_of_Moments$xi
[1] 0.2865109
$Maximum_Likelihood_Estimation
$Maximum_Likelihood_Estimation$mu
[1] 0.03635258
$Maximum_Likelihood_Estimation$sigma
[1] 1.005116
$Maximum_Likelihood_Estimation$xi
[1] 0.451668
$L_moments_Estimation
$L_moments_Estimation$mu
[1] 1.34973
$L_moments_Estimation$sigma
[1] 1.121475
$L_moments_Estimation$xi
[1] 0.7013185
> return_levels <- sapply(estimated_params, function(params) {
+ periods <- c(2, 5, 10, 25)
+ result <- numeric(length(periods))
+ for (j in 1:length(periods)) {

```

```
+ pt <- 1 - (1 / periods[j])
+ for (i in 1 :3) {
+ mu <- params[[1]]
+ sigma <- params[[2]]
+ xi <- params[[3]]
+ if (xi != 0) {
+ result[j] <- mu - (sigma / xi) * (1 - (-log(1 - pt))^-xi)
+ } else {
+ result[j] <- sigma * log(-log(1 - pt))
+ }
+ }
+ return(result)
+ })
> print(return_levels)
```

Method_of_Moments
2.5672919
-0.0518915
-0.9876320
-1.7801151
Maximum_Likelihood_Estimation
0.4369872
-0.3940551
-0.6621429
-0.8765393
L_moments_Estimation
1.8184244
0.8959630
0.6415696
0.4550232

3.7.3 Conclusion :

Sur la base des résultats de ce code, en particulier des résultats du paramètre de forme estimé par les trois méthodes, qui ont été positifs et compris dans la plage de 0,5. Ce qui nous amène à conclure que la distribution exacte des données est la distribution de Fréchet, qui se caractérise par un queue lourd.

Quant aux résultats du niveau de retour par les trois méthodes, ils sont logiques car ils se situent dans les valeurs des données, et c'est très bien car la précision de l'estimation des niveaux de retour signifie une prévision précise de la possibilité du retour de phénomènes extrêmes et dangereux événements, ce qui contribue à éviter les catastrophes et les pertes.

Conclusion générale

Au cours de ce travail nous avons divisé le travail en deux parties, la partie théorique et la partie théorique appliqué. La partie théorique dans laquelle nous sommes passés à la présentation fondements théoriques des valeurs extrêmes dans les distributions univariées modulées phénomènes extrêmes. Nous l'avons présenté selon l'approche (*GEV*) puis estimation des paramètres de cette distribution, des quantiles extrêmes, de la période de retour et du niveau de retour dans cette approche. La deuxième partie est une application de la première partie basée sur des données réelles "la corrosion des oléoducs". Dans ce contexte, nous avons atteint l'objectif de notre application, qui est d'estimer la période de retour et le niveau de retour.

Notre travail s'est arrêté là , mais ses perspectives sont bien plus longues que cela, car la corrosion est considérée comme l'un des plus gros problèmes de production pour les grandes entreprises. Notre objectif est de couvrir tous les types de puits, nous devons donc analyser les données de corrosion de chaque puits, comprendre les facteurs de forçage naturels et les propriétés chimiques et définir d'abord un modèle théorique, puis nous le développerons en un modèle d'intelligence artificielle capable de mesurer avec précision prédire quand le tuyau se corrodera jusqu'au point d'expiration. Cela nous permettra d'éviter de nombreux problèmes de pertes matérielles importantes, tant en termes de coûts qu'en termes d'arrêt temporaire de la production dans les pires cas.

Bibliographie

- [1] ABBA, M., (2019). Sur la théorie de l'estimation paramétrique, Mémoire master, Université Mohamed Khider, Biskra.
- [2] Balkema, A. A., De Haan, L., (1974). Residual life time at great age. The Annals of probability, 792–804.
- [3] Beirlant, J., Matthys, G. (2003). Estimating the extreme value index and high quantiles.
- [4] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., (2006). Statistics of extremes : theory and applications. John Wiley & Sons.
- [5] Brahimi, B. (2010). Statistics of Bivariate Extreme Values. Thèse de doctorat, Université de Biskra.
- [6] Benelmir.I.,(2010). Les L-moments : Application en hydrologie.magister en mathématique.Univ-biskra.
- [7] Coles, S., Bawa, J., Trenner, L., Dorazio, P., (2001). An introduction to statistical modeling of extreme values. Vol. 208. Springer.
- [8] Christopeit, N. (1994). Method of moments for semiparametric location-scale models. Journal of Statistical Planning and Inference, 38(2), 167-185.
- [9] Danielsson, J., Jansen, D. W., De vries, C. G., (1996). The method of moments ratio estimator for the tail shape parameter. Communications in Statistics-Theory and Methods 25 (4), 711–720.
- [10] Dekkers, A. L., De Haan, L., et al., (1989). On the estimation of the extreme-value index and large quantile estimation. The Annals of Statistics 17 (4), 1795–1832.
- [11] Dekkers, A. L., Einmahl, J. H., De Haan, L., (1989). A moment estimator for the index of an extreme-value distribution. The Annals of Statistics, 1833–1855.
- [12] Dekkers, A.L.M. et de Haan, L. (1989). On the Estimation of the Extreme-Value Index and Large Quantile Estimation. Annals of Statistics 17, 1795-183.
- [13] Drees, H., Ferreira, A. and de Haan, L. (2002). On maximum likelihood estimation of the extreme value index. To appear in Ann. Appl. Probab.

- [14] Davis, R. and Resnick, S.I. (1984). Tail Estimates Motivated by Extreme Value Theory. *Annals of Statistics* 12, 1467-1487.
- [15] Deheuvels, P., Hausler, E. and Mason, D.M. (1988). Almost Sure Convergence of the Hill Estimator. *Mathematical Proceedings of the Cambridge Philosophical Society* 104, 371-381.
- [16] Drees, H. and Kaufmann, E. (1998). Selection of the Optimal Sample Fraction in Univariate Extreme Value Estimation. *Stochastic Processes and their Applications* 75, 149-195.
- [17] Drees, H.(1996). Refined Pickands Estimator of the Extreme Value Index. *Annals of Statistics* 23, 2059-2080.
- [18] de Haan, L. and Ferreira, A., 2006. *Extreme Values Theory : An introduction*. Springer.
- [19] Embrechts,P., Kluppelberg, C., Mikosh, T., (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Verlag, Berlin, Heidelberg, New-York, London,Paris, Tokyo, Hong Kong, Barcelona, Budapest.
- [20] El Adlouni, S.-E., Bobée, B., Ouarda, T. B., (2007). Caractérisation des distributions à queue lourde pour l'analyse des crues. No. R929. INRS, Centre Eau, Terre et Environnement.
- [21] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, 2nd edition.Wiley, New York.
- [22] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample.*Proceedings of the Cambridge Philosophical Society* 24, 180-190.
- [23] Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments : Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049-1054.
- [24] Hosking.J.R.M., (1990). L-moments : analysis and estimation of distributions using linear combinations of statistics,*Journal of the Royal Statistical Society*.105 124.
- [25] Hosking.J.R.M., Wallis.J.R., (1995). A comparison of unbiased and plotting-position estimators of L-moments.T.J.watson research centre .New York.2019 2025.
- [26] Hill, B. (1975). A Simple Approach to Inference About the Tail of a Distribution. *Annals of Statistics* 3, 1163-1174.
- [27] Hausler, E. and Teugels, J.L. (1985). On Asymptotic Normality of Hill's estimator for the Exponent of Regular Variation. *Annals of Statistics* 13, 743-756.

- [28] Jenkinson, A. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(3 48) :158–171. 14.
- [29] Jean-marie., (1991). Méthode des moments de probabilité pondérés : application à la loi jenkinson. University Montpellier.67 84.
- [30] Meraghni, D., (2008). Modelling distribution tails, Thèse de Doctorat, Université de Biskra,Algérie.
- [31] Mason, D.M. (1982). Laws of Large Numbers for Sums of Extreme Values. *Annals of Probability* 10, 756-764.
- [32] Pickands III, J., et al., (1975). Statistical inference using extreme order statistics. *the Annals of Statistics* 3 (1), 119–131.
- [33] Smith, R., (1985), Maximum likelihood estimation in a class of non regular cases, *Biometrika*,72, 6790.
- [34] Von Mises, R. (1954). La distribution de la plus grande de n valeurs. *American Mathematical Society*, Providence, RI, II(In Selected Papers) :271–294. 14, 26.
- [35] Yun, S. (2002). On a Generalized Pickands Estimator of the Extreme Value Index. *Journal of Statistical Planning and Inference* 102, 389-409.

Résumé

Notre recherche se concentre sur la modélisation statistique des valeurs extrêmes en utilisant la distribution (GEV). L'application pratique de cette recherche implique l'analyse de données réelles sur les valeurs de corrosion dans les oléoducs du puits RB66 à Sonatrach. Les données sont ajustées à la distribution GEV, et les estimations des paramètres permettent de calculer les niveaux de retour. Cela fournit des prévisions précises qui aident à atténuer les risques de corrosion des oléoducs, évitant ainsi les arrêts de production.

Mots clés: valeurs extrêmes, GEV, niveaux de retour, corrosion.

Abstract

Our research focuses on the statistical modeling of extreme values using the (GEV) distribution. The practical application of this research involves the analysis of real data on corrosion values in the oil pipelines of the RB66 well in Sonatrach. The data are fitted to the GEV distribution, and the parameter estimates are used to calculate the return levels. This provides accurate forecasts that help mitigate pipeline corrosion risks, preventing production shutdowns.

Key words: extreme values, GEV, return levels, corrosion.

ملخص

يركز بحثنا على النمذجة الاحصائية للقيم المتطرفة باستخدام توزيع GEV. يتضمن التطبيق العملي لهذا البحث تحليل بيانات حقيقية لقيم التآكل في أنابيب النفط ببئر RB66 في سوناطراك، تم تجهيز البيانات لتوزيع GEV، ويتم استخدام تقديرات المعلمات لحساب مستويات العائد ويوفر ذلك تنبؤات دقيقة تساعد في التخفيف من مخاطر تآكل خطوط الأنابيب و بالتالي تجنب إيقاف الإنتاج. الكلمات المفتاحية: القيم المتطرفة، GEV، مستويات العودة، التآكل.