

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES



Thèse présentée en vue de l'obtention du diplôme :

de doctorat en mathématiques

Option : Statistiques

Par

KOUIDER Mohammed Ridha

Titre :

**Estimation de l'indice de queue d'une distribution à
queue lourde sous censure aléatoire**

Membres du jury :

Pr. Djebrane Yahia Président Université de Biskra

Pr. Fatah Benatia Directeur de thèse Université de Biskra

Pr. Abdellah Sayah Examineur Université de Biskra

Pr. El Amir Djefal Examineur Université de Batna

2023.

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
MOHAMED KHIDER UNIVERSITY, BISKRA
FACULTY of EXACT SCIENCES, SIENCE of NATURE and LIFE
DEPARTMENT of MATHEMATICS



A thesis submitted for the fulfillment of the requirements of :

The Doctorate Degree in Mathematics

Option : **Statistics**

By

KOUIDER Mohammed Ridha

Title :

**Estimation of the tail index of a heavy tail
distribution under random censorship**

Members of the jury :

Pr. Djebrane Yahia	President	University of	Biskra
Pr. Fatah Benatia	Director of thesis	University of	Biskra
Pr. Abdellah Sayah	Examiner	University of	Biskra
Pr. El Amir Djefal	Examiner	University of	Batna

2023.

Dédicace

À Ma grande famille

Mon père, ma mère, mes soeurs, mes frères

&

Ma petite famille.
Ma femme et mes enfants

Remerciements

Je glorifie Allah le tout puissant de m'avoir donnée courage et patience qui m'ont permis d'accomplir ce travail .

Le travail présenté dans notre thèse a été réalisé au Laboratoire de Mathématiques Appliquées à la Faculté des Sciences Exactes et des Sciences Naturelles et de la Vie - Mathématiques - de Université Mohamed Khider, Biskra, Algérie, sous la direction du Professeur **Fatah Benatia** .

Le moment est venu pour moi d'exprimer ma grande gratitude à tous ceux qui m'ont aidé et encouragé dans la réalisation de cette thèse, dont le prix était un doctorat.

Je tiens à exprimer mes remerciements à mon directeur de thèse, le professeur **Fatah Benatia** pour son accompagnement, sa patience, ses conseils avisés et sa présence, qui ont contribué à la réalisation de cette thèse. Je le remercie également pour la grande confiance qu'il m'a accordée tout au long de la période de rédaction de cette thèse. Son encadrement a contribué à développer mes compétences et mon goût pour la recherche.

Mon enseignement, je me souviendrai toujours de nos discussions, de vos conseils et surtout de votre gentillesse. J'applaudis également sa flexibilité et son ouverture d'esprit, qui m'ont laissé une grande marge de liberté pour mener ce travail de recherche.

Je tiens également à témoigner toute ma reconnaissance au cher professeur **Djabrane Yahia** , pour l'honneur qui m'a été fait en faisant partie de ce jury. Aussi, le professeur **Mokhtari Zouhir** pour tout le soutien moral et les beaux conseils au début de ma carrière de doctorant.

Je tiens également à remercier chaleureusement au professeurs inspirant **Brahimi Brahimi** pour son humilité et son humanité que l'on retrouve rarement chez un chercheur comme lui. Il me donnait de précieux conseils pour m'aider à avancer

Je tiens, également à exprimer toute ma reconnaissance au le merveilleux professeur **Necir Abdelhakim** pour sa sagesse et sa patience. Il a également un grand crédit pour ma formation et mon éducation au début de ma carrière universitaire. Merci cher professeur.

J'exprime ma gratitude au Professeur **Abdellah Sayah** , Professeur à l'Université Biskra, pour l'honneur qu'il me fait en prenant la charge de rapporteur et en siégeant à ce jury.

Mes remerciements s'adressent aussi au Professeur **El Amir Djefal** , Professeur à l'Université de Batna, pour nos échanges au cours de cette thèse mais aussi pour l'hommage qu'il me fait en participant à ce jury.

Un grand merci à mes amis et camarades doctorants, en particulier **Nesrine Idiou** , pour leur soutien lors des moments difficiles.

Enfin, je dédie cette thèse à ma famille qui, depuis de nombreuses années, a su m'encourager dans mes choix, et sans qui je ne serais pas devenu ce que je suis aujourd'hui. Pardonnez à ceux que j'ai oublié de mentionner ici, car je leur dois peut-être aussi beaucoup.

Abstract

In Extreme statistics, we mainly deal with the estimation of extreme parameters. In addition, with all applications of extreme value theory (EVT), the estimation of the index of extreme values (EVI) or the tail index γ is of primary importance and forms the basis for the estimation of all other parameters of extreme events. And among the most important are the generalized Hill estimator $\hat{\gamma}_{k,n}^{GH}$ and the maximum likelihood estimator $\hat{\gamma}^{ML}$. Because $\hat{\gamma}_{k,n}^{GH}$ is proper by the fact that it has a good asymptotic limit bias with $\gamma \in \mathbb{R}$ and $\hat{\gamma}^{ML}$ consistent with lower bias and the variance where $\gamma > (-1/2)$. Most research in EVT focuses on heavy-tailed distributions where $\gamma > 0$. This is the area we are going to focus on. We will now assume the case of random censorship, and will devote our attention to the estimation of EVI under random censorship. Where apart from a recent paper by Einmahl et al [11] (2008), to our knowledge, all that is required is to adapt the classic estimator of EVI under the data censored between them the $\hat{\gamma}^{ML}$ where $\gamma > (-1/2)$. In this work, we will introduce a numerical method in order to estimate the EVI under random censorship. And we will see that it is not possible to adapt the $\hat{\gamma}^{ML}$ under random censorship by the definition given by Einmahl et al [11] (2008). And this is what he indicated Kouider et al [28] (2021) and this is the crux of our subject.

Key words: Order statistics; Distribution of generalized extreme values; Generalized Pareto distribution; Heavy tail distributions; Generalized Hill estimator, Maximum likelihood estimator; KIB-Estimator; Random censorship; Kaplan-Meier estimator; Asymptotic normality.

Résumé

Dans le traitement et l'analyse des statistiques extrême, on se focalise essentiellement sur l'estimation des paramètres de l'indice de queue. Ce qui nous amène à l'estimation de la distribution des extrêmes (EVT), qui repose donc sur l'estimation de l'indice des valeurs extrêmes (EVI) ou l'indice de queue γ , qui est donc d'une importance primordiale et constitue la base de l'estimation de tous les autres paramètres liés aux événements extrêmes. Parmi les plus importants on peut citer l'estimateur de Hill généralisé $\hat{\gamma}_{k,n}^{GH}$ et l'estimateur du maximum de vraisemblance $\hat{\gamma}^{ML}$, car $\hat{\gamma}_{k,n}^{GH}$ est asymptotiquement sans biais pour $\gamma \in \mathbb{R}$, on sait aussi que $\hat{\gamma}^{ML}$ est un estimateur consistant avec un faible biais. La plupart des recherches en EVT se concentrent sur les distributions à queues lourdes où le EVI $\gamma > 0$. C'est le domaine sur lequel nous allons nous concentrer. Nous supposons le cas de la censure aléatoire (c.a), et consacrerons notre attention à l'estimation de EVI sous c.a. Notons qu'en dehors un article récent d'Einmahl et al [11] (2008), il n'y a, à notre connaissance, que l'estimateur classique adapter de EVI sous c.a, $\hat{\gamma}^{ML}$ où $\gamma > \frac{-1}{2}$. Dans notre travail, nous allons introduire une méthode numérique afin d'estimer EVI sous c.a, nous verrons qu'il n'est pas possible d'adapter $\hat{\gamma}^{ML}$ sous c.a par la définition donnée par [11]. Et c'est ce qui est indiqué dans [28] (2023) ceci représente le nœud de notre sujet.

Mots clés: Statistique d'ordre; Distribution des valeurs extrêmes généralisées; Distribution de Pareto généralisée; Distribution à queue lourde; Estimateur de Hill généralisée, Estimateur du maximum de vraisemblance; KIB-estimateur; Censure aléatoire; Estimateur de Kaplan-Meier; Normalité asymptotique.

Abréviations et Notations

Les différents symboles et abréviations utilisés dans cette thèse sont:

$F_n(\cdot)$: La fonction de répartition empirique.

\bar{F} : La fonction de queue de F

$\mathbf{1}_{(X_i \leq x)}$: Fonction indicatrice

$S(\cdot)$: Fonction de la survie

$Q_n(\cdot)$: Fonction des quantiles empirique.

(Ω, A, P) : Espace probabilisé

$N(\mu; \sigma^2)$: La loi normale

$\xrightarrow{p.s.}$: La converge presque sûrement

$n \rightarrow +\infty$: n tend vers l'infinie.

$\xrightarrow[n \rightarrow \infty]$: La limite quand n tend vers l'infinie

\xrightarrow{d} : Convergence en loi(en distribution)

\xrightarrow{p} : Convergence en probabilité

$X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$: La statistique d'ordre associée à l'échantillon X_1, X_2, \dots, X_n .

$X_{k,n}$: Est la $k^{\text{ème}}$ statistique d'ordre pour $1 \leq k \leq n$.

$X_{1,n} := \min(X_i)$:

$X_{n,n} := \max(X_i)$:

$X_{1,n}; X_{n,n}$: La statistique d'ordre extrême

$F_{X_{n,n}}$: La distribution de $X_{n,n}$

$X_{n,n} \stackrel{d}{=} a_n x + b_n$ v.a X non dégénérée est max-stable

\widehat{F}_n : C'est l'estimateur de la fonction de répartition empirique F_n

\overline{F}_n^{KM} : C'est l'estimateur de Kaplan-Meier

$U \in GRV_2(\gamma; \rho; a_0(t); A_0(t); \tilde{\alpha})$, On dit U est à variation régulière du deuxième ordre généralisée

$\widehat{\gamma}_{k,n}^{GH}$: L'estimateur de Hill généralisée.

$\widehat{\gamma}^{ML}$: L'estimateur de maximum de vraisemblance

$\widehat{\gamma}_{k,n}^H$:L'estimateur de Hill.

$\widehat{\gamma}^{KIB}$: L'estimateur de KIB(KIB-estimateur)

$\widehat{\gamma}_1^{(c, \cdot)}$: C'est estimateurs adaptive de l'indice des valeurs extrêmes sous les données censurées .

$\widehat{\gamma}_1^{(c, GH)}$: L'estimateur adaptive de Hill généralisée sous les données censurées

$\widehat{\gamma}_1^{(c, ML)}$: L'estimateur adaptive de maximum de vraisemblance sous les données censurées

$\widehat{\gamma}_1^{(c, KIB)}$: L'estimateur adaptive de KIB sous les données censurées

$F_t(x)$: Une probabilité conditionnelle ou la distribution conditionnelle .

$E_t(x)$: La fonction d'excès moyen ou l'espérance conditionnelle.

$F \in D(\Phi_\gamma)$: On dit que F appartient au domaine d'attraction de Fréchet.

$F \in D(\Psi_\gamma)$: On dit que F appartient au domaine d'attraction de Weibull.

$F \in D(\Lambda_\gamma)$: On dit que F appartient au domaine d'attraction de Gumbel.

$F \in D(H_\gamma)$: On dit que F appartient au domaine d'attraction des distributions des valeurs extrêmes.

$U \in RV_{\gamma, \rho}$ ou $U \in 2RV_{\gamma, \rho}U$: On dit U est à variation régulière du deuxième ordre

U : C'est une fonction quantile de queue.

$f \in RV_0$: On dit que f est à variation régulière d'indice 0. Ou on dit que f est à variation lente.

RV_γ : L'ensemble des fonctions à variations régulières d'indice γ .

$f \in RV_\gamma$:On dit f est à variation régulière d'indice γ .

x_F : Le point terminal de F .

Nous avons également utilisé des codes courts pour certaines des phrases présentées ci-

dessous:

v.a : La variable aléatoire.

c.a : censure aléatoire.

f.d.c : la fonction de distribution cumulative.

i.i.d : indépendantes et identiquement distribuées.

TCL: Théorème Central Limite.

MBM: Méthode de Bisection Modifiée.

Nous avons également utilisé des acronymes pour des phrases courantes en anglais:

CDF: The cumulative distribution function.

PDF: The probability density function.

ML-estimator: Maximum Likelihood estimator.

MLE : Maximum Likelihood Estimation.

POT : Peaks over Threshold.

GPD : Generalized Pareto Distribution.

EVT : Extremes Value Theory.

EVD: Extremes values Distribution.

GEV: General Extreme Value.

GEVD: The Generalized Extreme Value Distribution.

MBA : Modified Bisection Algorithm.

Table des matières

Dédicace	i
Remerciements	iii
Abstract	iv
Résumé	v
Abréviations et Notations	vi
Table des matières	xi
Liste des tableaux	xii
Liste des figures	xiii
1 Statistique d'ordre	4
1.1 Définitions	4
1.1.1 La fonction de répartition	7
1.1.2 La fonction de répartition empirique	8
1.1.3 La fonction de survie	9
1.1.4 La fonction de quantile	9
1.1.5 La fonction des quantiles empirique	10
1.2 Caractéristiques de bases	10

1.2.1	Inégalité de Markov	11
1.2.2	Inégalité de Bienaymé-Chebychev	11
1.2.3	Lois des grands nombres	12
1.2.4	Théorème central limite	13
1.2.5	Statistique d'ordre	15
1.2.6	Distributions des statistiques d'ordres extrêmes	16
2	Statistiques d'indice de queue	21
2.1	Lois des valeurs extrêmes	21
2.1.1	Théorème de Fisher et Tippett	21
2.1.2	Théorème de Gnedenko	22
2.2	Distribution des Valeurs Extrêmes Généralisées(La Distribution de GEV)	23
2.2.1	Définition la distribution de GEV	23
2.2.2	Fonctions à variation régulière de premier ordre	25
2.2.3	Fonctions à variation régulière du deuxième ordre	27
2.2.4	Domaines d'attraction	28
2.3	Distribution de Pareto généralisée (GPD)	32
2.3.1	Définition de la GPD	33
2.3.2	Distribution des excès	35
2.4	Estimation d'indice des valeurs extrêmes	35
2.4.1	Estimateur de maximum de vraisemblance($\gamma > (-1/2)$)	35
2.4.2	Estimateur de Hill ($\gamma > 0$)	44
2.4.3	Estimateur de Hill Généralisé ($\gamma > 0$)	46
3	Statistiques d'Indice de queue sous données censurés	51
3.1	Données censurés	51
3.1.1	Définitions	52
3.1.2	Estimateur de Kaplan-Meier	55

3.2	Estimation d'indice des valeurs extrêmes sous données censurés	63
3.2.1	Estimateur de la proportion d'observations non censurées	64
3.2.2	Estimateurs adaptatifs d'indice de queue sous données censurées	68
4	Algorithme de bisection modifié pour estimer l'indice de queue	71
4.1	Algorithme de bisection modifié pour multi-racine	71
4.1.1	Méthode de bisection modifiée (MBM)	71
4.1.2	Algorithme de bisection modifié (MBA)	77
4.1.3	MBA pour multi-racine	79
4.2	MBA pour estimer l'indice de queue sous données complètes($\gamma > -1$)	82
4.3	MBA pour estimer l'indice de queue sous données censurés	87
5	Appendix	95
5.1	Appendix A	95
5.1.1	Propriétés des estimateurs du maximum de vraisemblance	95
5.1.2	MBM pour multi-racines avec le R programme	104
5.1.3	MBA pour multi-racines (4.1.3) avec le R	105
5.2	Appendix B	106
5.2.1	Preuve des théorèmes (4.1) , (4.5) et (4.6)	106
5.2.2	Application MBA pour estimer l'indice de queue sous données complet (4.2)	113
5.2.3	Application MBA pour estimer l'indice de queue sous données censurées	116
	Conclusion	122
	Bibliographie	126

Liste des tableaux

2.1	Quelques loi usuelles classées en fonction de leur domaine d'attraction . . .	32
3.1	Estimation de la survie par l'estimateur de Kaplan-Meier	62
5.1	Exemples de fonctions de perte	103
5.2	Pertes maximales mensuelles à partir de données de Danish pour un échantillon aléatoire de $n = 73$	114
5.3	Échantillon $(Z_n; n \geq 1)$ de taille 15 généré avec la GPD sous censure. . . .	116
5.4	Présentant des résultats obtenues de l'application la méthode d'estimation présentée en (5.3)	117

Liste des figures

1.1	La signification géométrique de PDF de la statistique d'ordre k	17
1.2	Exemple de PDF avec la statistique d'ordre k	18
2.1	La fonction de répartition de la loi EVD	24
2.2	La densité de la distribution EVD	25
2.3	La PDF pour GPD avec les paramètres $(0, 1)$, $(1, 1)$, $(2, 1)$	34
3.1	Présentant un exemple pratique de données censure à gauche et censure à droite suivant.	55
3.2	Estimateur de Kaplan-Meier (ligne continue)	63
4.1	Illustration de la MBM avec cas 1.	73
4.2	Illustration de la MBM avec cas 2.	74
4.3	Illustration de la MBA (dont la racine est x_2).	79
5.1	MBM pour la multi-racines avec le programme R.	104
5.2	MBA pour multi-racine avec le programme R.	106
5.3	Procédure de MBA pour estimer l'indice de queue sous données complètes	115
5.4	Procédure MBA pour estimer l'indice de queue sous données censurées . .	120

Introduction Générale

La théorie des valeurs extrêmes (EVT : Extremes Value Theory) est une branche de la statistique dont le but est l'étude des événements rares ; c'est-à-dire des événements dont la probabilité d'apparition est très faible. Les valeurs extrêmes sont des événements rares, puisque il s'agit des valeurs beaucoup plus grandes ou plus petites que celles observées habituellement. Cette théorie repose principalement sur des distributions limites des extrêmes et leurs domaines d'attraction.

Une distribution des valeurs extrêmes généralisée (GEVD) particulièrement utile est la distribution de Pareto généralisée (GPD), qui a été introduite par Pickands [38] (1975) pour modéliser l'excès par rapport au seuil au-delà d'un maxima. Nous savons que le phénomène de queue lourde et le GPD sont liés par le célèbre théorème de Pickands-Balkema-de Haan (Balkema et De Haan [2] (1974) et Pickands [38] (1975)) qui stipule que, pour une distribution arbitraire dont le maximum d'échantillon tend vers une distribution non dégénérée après une normalisation appropriée, la fonction de distribution de ses dépassements par rapport à un seuil converge vers la GPD. En particulier, elle est largement utilisée pour modéliser les piques au-delà d'un certain seuil important, une procédure communément appelée méthode de piques au-delà d'un seuil t (POT) dans la littérature sur la théorie des valeurs extrêmes (EVT). De plus, l'application de la GPD à des ensembles de données réelles n'est cependant pas une tâche triviale.

Dans la littérature il existe beaucoup de méthodes d'estimation des paramètres de la GPD. Parmi ces méthodes, nous citons la méthode du maximum de vraisemblance. Grimshaw

[15] (1993) a publié un algorithme de calcul de l'estimation du maximum de vraisemblance (MLE) des paramètres de la GPD. Il s'est également appuyé sur de nombreux calculs et conditions pour mettre son algorithme en place. Ainsi, en 2019, [26] a présenté un algorithme permettant l'estimation par la méthode de la MLE des paramètres de la GPD sous certaines conditions. Nous verrons dans ce travail l'algorithme en présence de données censurées.

Dans le domaine des statistiques des extrêmes, l'hypothèse la plus courante sur n'importe quel ensemble de données univariées est de les considérer comme un échantillon complet d'observations indépendantes et identiquement distribuées, à partir d'une fonction de distribution inconnue. Cependant, dans l'analyse des données sur la durée de vie, les observations sont généralement censurées de manière aléatoire. L'estimation de l'indice de queue joue un rôle très important dans EVT aussi bien pour les données complètes que pour les données censurées avec une distribution à queue lourde.

Notre thèse est organisée en cinq chapitres. Dans le premier chapitre, nous introduisons certaines propriétés de base ainsi que certains concepts dont dépendent les propriétés et les résultats, que nous présenterons et utiliserons dans les chapitres suivants. Après avoir défini ces concepts et différents outils, dans la première partie de cette présentation, nous nous intéressons au modèle des valeurs extrêmes.

Nous commençons, dans le chapitre 2, par présenter la progression historique des travaux connexes à la théorie des valeurs extrêmes (Fisher et Tippett [13] (1928) et Gnedenko [14] (1943)) qui nous emène à définir la GEVD. Nous avons également introduit le concept de GPD, la distribution à queue lourde et leur relation avec la GEVD. Tout cela dans le but d'introduire les estimateurs de l'indice de queue: $\hat{\gamma}^{ML}$ l'estimateur du maximum de vraisemblance et $\hat{\gamma}_{k,n}^{GH}$ l'estimateur de Hill généralisé, où l'indice de queue est strictement positif. Les résultats asymptotiques obtenus pour ces estimateurs sont aussi présent selon les travaux de Drees et al [10] (2004) et de Haan et Ferreira [18] (2006) et l'article de Beirlant et al [4] (2005) respectivement.

Dans le chapitre 3, nous avons brièvement présenté quelques concepts liés à la statistique de données censurée ainsi que des définitions sur l'analyse des données sur la durée de vie, et une la définition de l'estimateur de Kaplan-Meier [24] (1958). Le travail de Stute [36] (1995) qui traite de la façon d'adapter le théorème central limite avec les données censurées. Les estimateurs de l'indice de queue avec les données censurées, en utilisant la définition donnée par Einmahl et al [11] (2008) sont cités et qui montrerons que l'on peut adapter toutes les estimations classiques de l'indice de queue à leur définition donnée à $\hat{\gamma}^{ML}$ et $\hat{\gamma}_{k,n}^{GH}$, on conclut ce chapitre en présentant leur résultats asymptotiques en présence de censure.

Dans le chapitre 4, nous proposons l'estimateur non paramétrique par la méthode du maximum de vraisemblance de la GPD [26] (2019) pour estimer l'indice de queue avec des données complètes, une adaptation de ce travail est le résultat [28] (2023) qui mène à $\hat{\gamma}_1^{(c,ML)}$ l'estimateur du maximum de vraisemblance adapté pour les données censurées qui est lui même équivalent à $\hat{\gamma}^{ML}$, à l'aide de l'identification de Einmahl et al [11] (2008) sous une certaine condition.

L'aspects numérique dans le chapitre 5 décrit les algorithmes avec le programme R développés dans [27] (2019) sur laquelle il s'est appuyé pour trouver $\hat{\gamma}^{ML}$ avec des données complètes et en présence de censure. Aussi, nous donnons des preuves des théorèmes présentés dans le chapitre précédent. De plus nous avons appliqués la méthode du MLE pour la GPD sans censure et en présence de censure, avec le modèle de la valeur extrême discuté dans le chapitre 4.

Chapitre 1

Statistique d'ordre

1.1 Définitions

Introduisons d'abord la terminologie utilisée dans cette thèse. Les unités sur lesquelles nous mesurons les données, telles que les personnes, les voitures,...etc, sont appelées **observations**. Ces observations sont représentées par le symbole grec ω . L'ensemble de toutes les unités est appelé **population**, est représenté par Ω . On se réfère à $\omega \in \Omega$ une sélection d'observations $\omega_1, \dots, \omega_n$.

Parfois, le concept de population n'est pas applicable. Par exemple, imaginons que nous mesurons la température dans Wilaya de Biskra toutes les heures. L'échantillon sera alors la série chronologique des températures dans une fenêtre temporelle spécifique, disons de juin à septembre 2022. Il n'y a pas de communauté ici au sens d'unités de contrôle. Mais supposons maintenant que nous mesurons les températures dans plusieurs villes algériennes différentes. Ainsi, toutes les villes constituent la population, et l'échantillon est n'importe quel sous-ensemble de villes.

Si nous avons spécifié la population d'intérêt pour une question de recherche spécifique, nous pouvons penser à ce qui est intéressant dans nos observations. Une particularité de ces observations peut être recueillie dans une **variable** statistique X . Toutes les informa-

tions qui nous intéressent peuvent être saisies dans une telle variable. Par exemple, si nos observations se réfèrent à des êtres humains "masculin ou féminin", X peut décrire l'état matrimonial, le sexe, l'âge ou tout autre élément pouvant se rapporter à une personne. Bien sûr, nous pouvons nous intéresser à de nombreuses caractéristiques différentes, chacune d'entre elles collectées dans une variable différente. Chaque observation ω prend une valeur particulière pour X . La définition formelle d'une variable est

$$X : \Omega \rightarrow S$$
$$w \rightarrow x$$

Exemple 1.1 *Une variable X qui fait référence à l'âge peut prendre n'importe quelle valeur entre 1 et 95. Chaque personne ω se voit attribuer une valeur x qui représente l'âge de cette personne.*

Les variables qualitatives sont les variables qui prennent des valeurs x qui ne peuvent pas être ordonnées de manière logique ou naturelle. Par exemple: La couleur des yeux, Le type de transport utilisé pour se rendre au travail, sont toutes des variables qualitatives. Il n'y a pas non plus de raison de lister les yeux bleus avant les yeux marrons (ou vice versa) et cela n'a aucun sens de lister les bus avant les trains (ou vice versa).

Les variables quantitatives représentent des quantités mesurables. Les valeurs que peuvent prendre ces variables peuvent être ordonnées de manière logique et naturelle. Par exemple: taille de chaussures, prix des maisons, poids d'une personne.

Il est courant d'attribuer des nombres aux variables qualitatives à des fins pratiques dans les analyses de données. Par exemple, si l'on considère la variable « sexe », alors chaque observation peut prendre soit la « valeur » masculine ou féminine. Nous pouvons décider d'attribuer 1 à la femme et 0 à l'homme (ou vice versa) et utiliser ces nombres au lieu des catégories d'origine. Cependant. Il n'y a pas d'ordre logique et naturel sur la façon d'organiser masculin et féminin, et ainsi, la variable sexe reste une variable qualitative,

même après avoir utilisé des nombres pour coder les valeurs que X peut prendre.

Les variables discrètes sont des variables qui ne peuvent prendre qu'un nombre fini de valeurs. Toutes les variables qualitatives sont discrètes, comme la couleur de l'œil ou la région d'un pays. Mais aussi les variables quantitatives peuvent être discrètes : la pointure des chaussures ou le nombre de semestres étudiés seraient discrets car le nombre de valeurs que peuvent prendre ces variables est limité.

Les variables qui peuvent prendre un nombre infini de valeurs sont appelées **variables continues**. Parfois, on dit que les variables continues sont des variables « mesurées plutôt que comptées ». Il s'agit d'une définition plutôt informelle qui aide à comprendre la différence entre les variables discrètes et continues. Le point crucial est que les variables continues peuvent, en théorie, prendre un nombre infini de valeurs ; par exemple, la taille d'une personne peut être enregistrée à 170cm . Cependant, la hauteur réelle sur le ruban à mesurer pourrait être de $170,3\text{cm}$, ce qui a été arrondi à 170cm . Si l'on disposait d'un meilleur instrument de mesure, on aurait peut-être obtenu $170,342\text{cm}$. Mais la taille réelle de cette personne est un nombre avec un nombre indéfini de décimales tel que $170,342975328\dots\text{cm}$. Peu importe ce que nous rapportons ou obtenons finalement, une variable qui peut prendre un nombre infini de valeurs est définie comme une variable continue.

La fréquence relative de la i -ème classe sont définis comme $f_i = n_i/n$ avec n_i est défini comme le nombre d'unités ω_i et n la taille de l'échantillon. Par conséquent, il existe un moyen de résumer et de visualiser la distribution (fréquence) des variables qui est la fonction de distribution cumulative empirique.

Une expérience aléatoire nécessite que l'expérience puisse être répétée un certain nombre de fois dans le même ensemble de conditions, et son résultat n'est connu qu'après la fin de l'expérience. Un exemple simple et classique d'une expérience aléatoire est le lancer d'une pièce de monnaie ou le roulement d'un dé. Lorsque vous lancez une pièce, on ne sait pas quel sera le résultat, pile ou face, jusqu'à ce que la pièce soit lancée. L'expérience peut

être répétée et différents résultats peuvent être observés à chaque répétition. Un résultat possible d'une expérience aléatoire est appelé **un événement simple** (ou événement élémentaire) et noté ω_i . L'ensemble de tous les résultats possibles est appelé **l'espace d'échantillonnage** et est noté Ω . Les sous-ensembles de Ω sont appelés **événements** et sont désignés par des lettres majuscules telles que A, B, C, \dots etc

Il existe un lien étroit entre f_i la fréquence relative et la probabilité d'un événement. Maintenant on peut compter la fréquence absolue n_i qui est égal au nombre de fois qu'un événement A_i avec n répétitions de l'expérience comme $f(A_i) = n_i/n$. Si l'on suppose que l'expérience est répétée un grand nombre de fois (mathématiquement, cela signifierait que n tend vers l'infini) et les conditions expérimentales restent les mêmes (au moins approximativement) sur toutes les répétitions, alors la fréquence relative $f(A_i)$ converge vers une valeur limite pour A_i . Cette valeur limite est interprétée comme la probabilité de A et notée $P(A_i)$. Une définition différente de la probabilité a été donnée par Pierre-Simon Laplace (1749-1827), qui se réalisent si le nombre d'événements simples possibles est fini et tous les résultats sont également probables. La proportion $P(A) = \frac{|A|}{|\Omega|}$ est appelée la probabilité de Laplace ou la distribution de probabilité uniforme, avec $|A|$ est le nombre cardinal de A , et $|\Omega|$ est le nombre cardinal de Ω .

Les statistiques et l'exploration de données concernent les données. Comment lier les espaces d'échantillonnage et les événements aux données ? Le lien est assuré par la notion de variable aléatoire. Nous pouvons donc considérer X comme **une variable aléatoire** qui recueille les résultats possibles d'une expérience aléatoire et capture l'incertitude qui leur est associée. Les variables aléatoires nous aident à voir les données collectées comme le résultat d'une expérience aléatoire.

1.1.1 La fonction de répartition

Définition 1.1 *La fonction de répartition ou la fonction de distribution cumulative (f.d.c) (The cumulative distribution function: CDF), d'une variable aléatoire (v.a) réelle X est la*

fonction F définie pour tout réel x à $[0; 1]$, par suivante:

$$F(x) := P(X \leq x) = P(]-\infty; x]) \quad (1.1.1)$$

Parfois, nous écrivons le f.d.c comme F_X au lieu de F et noton $X \sim F$. Cette fonction est caractéristique de la loi de probabilité de la v.a X . Elle est toujours croissante, continue à droite, avec $F(-\infty) = 0$ et $F(+\infty) = 1$. De plus f.d.c F est définie, pour tout nombre réel $x \in \mathbb{R}$, par:

$$F(x) := \int_{-\infty}^x f(u) du \quad (1.1.2)$$

où $f(x)$ appelée fonction de densité ou densité de probabilité (the probability density function: PDF). Les fonctions de densités sont les fonctions essentiellement positives et intégrables sur \mathbb{R} d'intégrale égale 1.

1.1.2 La fonction de répartition empirique

Soit un échantillon X_1, X_2, \dots, X_n , de v.a indépendantes et identiquement distribuées (i.i.d), avec f.d.c F . On définit aussi la fonction de répartition empirique F_n de l'échantillon X_1, X_2, \dots, X_n pour tout $x \in \mathbb{R}$ par :

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(X_i \leq x)}, \quad (1.1.3)$$

où $\mathbf{1}_{(X_i \leq x)}$ la fonction indicatrice, qui est définie par :

$$\mathbf{1}_{(X_i \leq x)} := \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si } X_i > x \end{cases} \quad (1.1.4)$$

1.1.3 La fonction de survie

L'analyse des durées de vie est un sujet important en biologie et en médecine en particulier mais aussi dans l'analyse de fiabilité avec des applications d'ingénierie. De telles données sont souvent très distribuées de manière non normale, de sorte que l'utilisation de modèles linéaires standard est problématique. Les données sur la durée de vie sont souvent censurées: on ne connaît pas la durée de vie exacte, mais seulement qu'elle est supérieure à une valeur donnée.

Définition 1.2 *Pour t fixé, la fonction de survie S « Survival function », d'une v.a positive et continue X dite "durée de survie", est définie par:*

$$S(t) := \bar{F}(t) = P(X > t) = 1 - F(t) \quad (1.1.5)$$

avec une fonction de densité d'évènement de survie définie par :

$$s(t) = \frac{dS(t)}{dt} := -f(t) \quad (1.1.7)$$

1.1.4 La fonction de quantile

Les quantiles sont une généralisation de l'idée de la médiane. **La médiane** est la valeur qui divise les données en deux parties égales. De même, un quantile partitionne les données en d'autres proportions. Par exemple, un quantile à 25 % divise les données en deux parties de sorte qu'au moins 25 % des valeurs soient inférieures ou égales au quantile et au moins 75 % des valeurs soient supérieures ou égales au quantile. En général, soit s un nombre compris entre 0 et 1. Le $s \times 100$ % quantile, noté x_s est défini comme la valeur qui divise les données en proportions de $s \times 100$ % et $(1 - s) \times 100$ % telle qu'au moins $s \times 100$ % des valeurs sont inférieures ou égales au quantile et au moins $(1 - s) \times 100$ % des valeurs sont supérieures ou égales au quantile.

Définition 1.3 Soit X une v.a avec f.d.c F . La fonction de quantile $Q(s)$ est définie pour tout $0 < s < 1$ par,

$$Q(s) := F^{-1}(s) := \inf \{t : F(t) \geq s\}, \quad (1.1.8)$$

où F^{-1} représente l'inverse généralisée de F . Avec F est continue et monotone alors F est bijective donc F^{-1} existe.

On appelle $F^{-1}(1/4)$ le premier quartile, $F^{-1}(1/2)$ the median (or second quartile), and $F^{-1}(3/4)$ le troisième quartile.

1.1.5 La fonction des quantiles empirique

Soit un échantillon X_1, X_2, \dots, X_n , de v.a's i.i.d. La fonction des quantiles empirique $Q_n(s)$ d'un échantillon X_1, X_2, \dots, X_n est définie pour tout $0 < s < 1$ par :

$$Q_n(s) := \inf \{x \in \mathbb{R} : F_n(x) \geq s\}, \quad (1.1.9)$$

Où $F_n(x)$ définie on (1.1.3) est la distribution empirique. On dit que l'estimateur de $Q_n(s)$ est $Q(s)$.

1.2 Caractéristiques de bases

Certaines autres caractéristiques des variables aléatoires sont caractérisées par les concepts d'espérance et de variance.

L'espérance d'une variable aléatoire continue X , ayant la fonction de densité de probabilité $f(x)$ avec $\int |x| f(x) dx < \infty$, est défini comme

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

et pour une variable aléatoire discrète X

$$E(X) = \sum_{i=1}^n x_i p_i$$

L'espérance de X , c'est-à-dire $E(X)$, est généralement notée $\mu = E(X)$ et se rapporte à la moyenne arithmétique de la distribution de la population. Il reflète la tendance centrale de la population.

La variance d'une variable aléatoire X est définie comme

$$Var(X) = E(X - E(x))^2 = E(X^2) - (E(X))^2$$

La variance de la variable aléatoire continue X est

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

1.2.1 Inégalité de Markov

l'inégalité de Markov applique beaucoup avec la theories des la probabilités pour une variable aléatoire réelle à valeurs positives soit supérieure ou égale à une constante positive. Cette inégalité a été nommée ainsi en l'honneur d'Andreï Markov

Définition 1.4 Soit X une v.a réelle et positive définie sur un espace probabilisé (Ω, A, P) .

Alors

$$\forall \varepsilon > 0, P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon} \tag{1.2.1}$$

1.2.2 Inégalité de Bienaymé-Chebychev

Inégalité de Bienaymé-Tchebychev indique comment une variable dévie de sa moyenne. Ce résultat nécessitant la connaissance de peu de propriétés de l'espérance et la variance. La démonstration est grâce à l'inégalité de Markov(1.2.1).

Définition 1.5 Soit X une v.a avec d'espérance $\mu := E(X)$ et de variance σ^2 . Alors

$$\forall \varepsilon > 0, P(|X - E(X)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \quad (1.2.2)$$

Cela équivaut à

$$\forall \varepsilon > 0, P(|X - E(X)| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

Si nous ne connaissons pas la distribution d'une variable aléatoire X , nous pouvons toujours faire des déclarations sur la probabilité que X prenne des valeurs dans un certain intervalle (qui doit être symétrique autour de l'espérance μ) si la moyenne μ et la variance σ^2 de X sont connus.

1.2.3 Lois des grands nombres

Les lois des grands nombres stipulent que si nous prenons un échantillon aléatoire dans une grande série statistique. Notons que plus la taille de l'échantillon est grande, plus les propriétés statistiques de l'échantillon sont proches des propriétés statistiques de la série. Absolument, cela signifie que $\bar{X}_n = \sum_{i=1}^n X_i$, calculée sur les valeurs de l'échantillon, converge vers l'espérance μ existant et soient finis lorsque la taille de l'échantillon tend vers l'infini. Plusieurs théories des probabilités expriment cette loi. Consiste en deux types, la loi faible des grands nombres approuve une convergence en probabilité noton (\xrightarrow{P}), tandis que la loi forte des grands nombres met en évidence une convergence presque sûre ($\xrightarrow{P-s}$). Les lois des grands ne s'appliquent pas pour des lois de probabilité sans espérance, comme la loi de Cauchy.

Théorème 1.1 (Khintchine) Soit un échantillon X_1, X_2, \dots, X_n de v.a's i.i.d suivant la même loi provenant d'une v.a. X tel que $\mu := E(X) < \infty$. Alors \bar{X}_n convergence en probabilité vers μ . Signifie que pour tout $\varepsilon > 0$:

$$\lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0 \quad (1.2.3)$$

D'après (1.1) on dit que \bar{X}_n est un estimateur convergent de $\mu := E(X) < \infty$.

Preuve. Suppose que $\sigma < \infty$. Ce n'est pas nécessaire mais cela simplifie la preuve. En utilisant l'inégalité de Chebyshev (1.2.2), avec $Var(\bar{X}_n) = \sigma^2/n$ où $\sigma^2 = Var(X_1)$ qui tend à 0 comme $n \rightarrow \infty$. signifie que la distribution de \bar{X}_n devient plus concentré autour μ quand n est plus grand. ■

Théorème 1.2 *Considérons un échantillon X_1, X_2, \dots, X_n de v.a.'s i.i.d suivant la même loi provenant d'une v.a. X tel que $\mu := E(X) < \infty$. Alors \bar{X}_n convergence presque sûrement vers $E(X)$. C'est-à-dire que :*

$$P\left(\lim_{n \rightarrow +\infty} \bar{X}_n = E(X)\right) = 1 \quad (1.2.4)$$

La loi des grands nombres dit que la distribution de \bar{X}_n s'accumule près μ . Ce n'est pas suffisant pour nous aider à approximer les déclarations de probabilité sur \bar{X}_n . Pour cela, nous avons besoin du théorème central limite.

1.2.4 Théorème central limite

Le théorème central limite (aussi appelé théorème limite central) basé sur la convergence en loi de la somme d'une suite de variables aléatoires i.i.d tend vers la loi normale. Énoncés de probabilité sur $\bar{X}_n = \sum_{i=1}^n X_i$ une approximation à l'aide d'une distribution normale. Ce sont les déclarations de probabilité que nous approchons, pas la variable aléatoire elle-même.

Théorème de Glivenko-Cantelli

Avant de parler sur le théorème central limite nous l'ençons le théorème de Glivenko-Cantelli. En théorie des probabilités, le théorème de Glivinko-Cantelli, connu sous le nom de "théorie fondamentale de la statistique". Il montre dans quelle mesure la loi

de probabilité peut être révélée en connaissant un (grand) échantillon de ladite loi de probabilité.

Théorème 1.3 (Glivenko-Cantelli (1933)) *la fonction de répartition empirique $F_n(x)$ converge presque sûrement uniforme vers la fonction de répartition $F(x)$, c'est-à-dire quand $n \rightarrow +\infty$ on a,*

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p.s.} 0 \quad (1.2.5)$$

Donc d'après Glivenko-Cantelli (1933), nous avons que $F_n(x)$ (1.1.3) converge presque sûrement vers $F(x)$ (1.1.1). Alors on dit que, $F_n(x)$ est un estimateur de $F(x)$.

Théorème 1.4 (TCL) *Soit un échantillon X_1, X_2, \dots, X_n , de v.a's i.i.d suivant la même fonction de répartition F et définies sur le même espace de probabilité (Ω, A, P) , avec d'espérance μ et de variance σ^2 existent et soient finis où $\sigma \neq 0$. Considérons la somme suivant:*

$$S_n = X_1 + X_2 + \dots + X_n \quad (1.2.6)$$

quand $n \rightarrow +\infty$ on a,

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0; 1) \quad (1.2.8)$$

Autrement dit,

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \phi(z) = \int_{-\infty}^z \frac{1}{2\pi} e^{-x^2/2} dx \quad (1.2.9)$$

Définition 1.6 *Soient a_n et b_n deux suites strictement positive*

$$a_n = o(b_n) : \iff \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$$

On écrit $a_n = o(1)$ si $\lim_{n \rightarrow \infty} a_n = 0$. Et

$$a_n = O(b_n) : \iff \lim_{n \rightarrow \infty} \frac{a_n}{b_n} \leq M$$

pour $M > 0$. On écrit $a_n = O(1)$ si $\exists M < \infty$ telque $|a_n| \leq M, \forall n \geq 1$.

D'après (1.1.4) on a que $\mathbf{1}_{(X_i \leq x)}$ est une variable aléatoire de Bernoulli avec $p = F(x)$. Par conséquent, la v.a $nF_n(x)$ où $F_n(x)$ définie on (1.1.3) est distribuée selon une loi binomial, avec moyenne $nF(x)$ et variance $nF(x)(1-F(x))$. Donc par le théorème limite central (1.4) on a $\sqrt{n}(F_n(x) - F(x))$ converge en loi vers une loi normale $N(0; F(x)(1 - F(x)))$ pour un x fixé. La convergence joue un rôle central dans la théorie moderne des probabilités. Pour plus de détails, voir Grimmett et Stirzaker (1982), Karr (1993) et Billingsley (1979). La théorie de la convergence avancée est expliquée en détail dans van der Vaart et Wellner (1996) et van der Vaart (1998).

1.2.5 Statistique d'ordre

Les valeurs ordonnées d'un échantillon d'observations sont appelées les statistiques d'ordre de l'échantillon, et les plus petites et les plus grandes sont appelées les extrêmes. Les statistiques d'ordre et les extrêmes sont parmi les fonctions les plus importantes d'un ensemble de variables aléatoires que nous étudions en probabilités et en statistiques. Il y a un intérêt naturel à étudier les hauts et les bas d'une séquence, et les autres statistiques d'ordre aident à comprendre la concentration de probabilité dans une distribution, ou de manière équivalente, la diversité de la population représentée par la distribution. Les statistiques d'ordre sont également utiles dans l'inférence statistique, où les estimations des paramètres sont souvent basées sur certaines fonctions appropriées des statistiques d'ordre.

Définition 1.7 *Soit un échantillon X_1, X_2, \dots, X_n de v.a's i.i.d suivant la même f.d.c F . La statistique d'ordre de l'échantillon précédent notées $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ est rangeant les v.a's X_1, X_2, \dots, X_n par ordre croissant.*

En particulier, la variable aléatoire $X_{k,n}$ est la $k^{\text{ème}}$ statistique d'ordre avec $1 \leq k \leq n$. Il est appelé le concomitant de la statistique d'ordre k , Le terme concomitant de statistiques

d'ordre a été introduit pour la première fois par David H. A. (1973) “Concomitants of Order Statistics”. Ils peuvent être exprimés comme,

$$\begin{aligned}
 X_{1,n} &:= \min(X_1, X_2, \dots, X_n) \\
 X_{2,n} &:= \min(\{X_1, X_2, \dots, X_n\} - \{X_1\}) \\
 &\vdots \\
 X_{k,n} &:= \min(\{X_1, X_2, \dots, X_n\} - \{X_1, X_2, \dots, X_{k-1}\}) \\
 X_{n,n} &:= \min(\{X_1, X_2, \dots, X_n\} - \{X_1, X_2, \dots, X_{n-1}\})
 \end{aligned}$$

Définition 1.8 *Les statistiques d'ordres extrêmes sont $X_{1,n}$ et $X_{n,n}$. Sont définies pour tout $i = 1, \dots, n$ par:*

$$X_{1,n} := \min(X_i) \tag{1.3.1}$$

et

$$X_{n,n} := \max(X_i) \tag{1.3.2}$$

Les statistiques sont utilisées pour tirer des conclusions sur la population à l'aide de modèles de probabilité. Les statistiques d'échantillon telles que le minimum, le premier quartile, la médiane, le troisième quartile et le maximum, appelées résumé à cinq chiffres, sont des exemples de statistiques d'ordre. Les observations minimales et maximales sont importantes dans la théorie des valeurs extrêmes.

1.2.6 Distributions des statistiques d'ordres extrêmes

Un lien important à comprendre est la connexion que les statistiques d'ordre ont avec la CDF empirique, une fonction d'une immense importance théorique et méthodologique à la fois en probabilité et en statistique. Auparavant, nous avons trouvé que la v.a $nF_n(x)$ où $F_n(x)$ définie on (1.1.3) est suit une loi binomial de paramètre (n, p) avec $p = F(X_i)$. Remarquons, que l'on a $nF_n(x) \geq k$ si et seulement si $X_{k,n} \leq x$. Il vient donc :

$$(X_{k,n} \leq x) \Leftrightarrow \left(F_n(x) \geq \frac{k}{n} \right) := \left(\sum_{i=1}^n \mathbf{1}_{(X_i \leq x)} \geq k \right) \quad (1.3.3)$$

On déduit que :

$$F_{X_{k,n}}(x) := p(X_{k,n} \leq x) = p \left(F_n(x) \geq \frac{k}{n} \right) := \sum_{i=k}^n C_n^i F(x)^i (1 - F(x))^{n-i} \quad (1.3.4)$$

où $C_n^i = \frac{n!}{i!(n-i)!}$. La PDF de la statistique d'ordre k ($f_{X_{k,n}}(x)$) s'exprime comme suit:

$$f_{X_{k,n}}(x) := \frac{n!}{(k-1)!(n-k)!} f(x) F(x)^{k-1} (1 - F(x))^{n-k} \quad (1.3.5)$$

où $F(x)$ et $f(x)$ represent la f.d.c et la densité de X_1, X_2, \dots, X_n v.a's i.i.d respectivement.

La présentation de cette formule est détaillée dans l'ouvrage de Reiss et Thomas [34] .

Donc, exactement 1 observation correspond exactement à la statistique d'ordre k avec probabilité f_X . Ainsi, le produit des 3 termes nous donne une idée de la signification géométrique de l'équation pour PDF de la statistique d'ordre k . Mais d'où vient le terme factoriel ? La figure (1.1) ci-dessus vient de montrer l'une des nombreuses commandes. Il peut y avoir de nombreuses combinaisons de ce type. Le nombre total de ces combinaisons est indiqué comme suit :

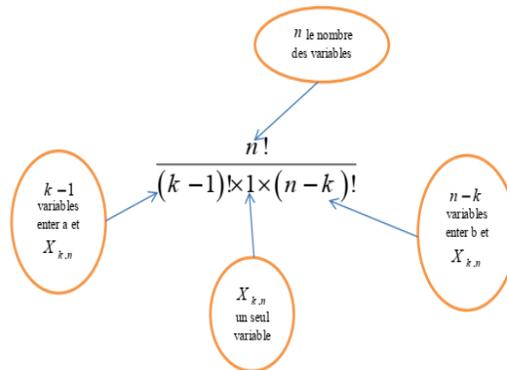


Figure 1.1: La signification géométrique de PDF de la statistique d'ordre k

Ainsi, le produit de tous ces termes nous donne la distribution générale de la statistique d'ordre k . La figure (1.2) suivant montre un exemple de PDF avec la statistique d'ordre k obtenue à partir d'un échantillonnage aléatoire :

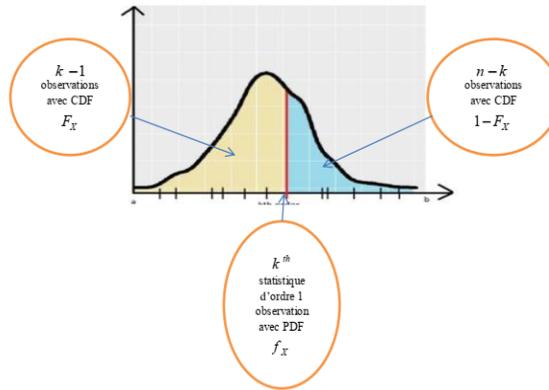


Figure 1.2: Exemple de PDF avec la statistique d'ordre k

Ainsi, la PDF des variables aléatoires f_X est définie entre l'intervalle $[a, b]$. La statistique d'ordre k pour un échantillon aléatoire est indiquée par la ligne rouge. Les autres réalisations variables (pour l'échantillon aléatoire) sont représentées par les petites lignes noires sur l'axe des abscisses.

Il y a exactement $(k - 1)$ observations de variables aléatoires qui tombent dans la région jaune du graphique (la région entre a & une statistique d'ordre k). La probabilité qu'une observation particulière tombe dans cette région est donnée par la CDF des variables aléatoires F_X . Mais nous sommes conscients que $(k - 1)$ observations sont tombées dans la région, ce qui nous donne le terme (par indépendance) F_X^{k-1} .

Il y a exactement $(n - k)$ observations de variables aléatoires qui se situent dans la région bleue du graphique (la région entre la statistique d'ordre k & b). La probabilité qu'une observation particulière tombe dans cette région est donnée par $1 - F_X$. Mais nous sommes conscients que $(n - k)$ observations sont tombées dans la région, ce qui nous donne le terme (par indépendance) $1 - F_X^{n-k}$.

Donc par l'inégalité (1.3.4) c'est facile de obtenir la suivant:

1. La distribution de la statistique d'ordre extrême $X_{n,n}$ (1.3.2) donnée par :

$$\begin{cases} F_{X_{n,n}}(x) := [F(x)]^n \\ f_{X_{n,n}}(x) := nf(x)[F(x)]^{n-1} \end{cases} \quad (1.3.6)$$

2. La distribution de la statistique d'ordre extrême $X_{1,n}$ (1.3.1) donnée par :

$$\begin{cases} F_{X_{1,n}}(x) := 1 - [1 - F(x)]^n \\ f_{X_{1,n}}(x) := nf(x)[1 - F(x)]^{n-1} \end{cases} \quad (1.3.7)$$

Autrement, on peut définir la fonction des quantiles empirique (1.1.3) par rapport la statistique d'ordre par:

$$Q_n(s) := X_{i,n} \text{ où } \frac{i-1}{n} \leq s \leq \frac{i}{n}, \quad (1.3.8)$$

Pour trouver le quantile théorique x_s avec $s \in [0; 1]$ il faut de chercher la solution de l'équation $F(x) = s$. lui même pour determine le quantile empirique $X_{i,n}$ il faut de chercher la solution de l'équation $F_n(X_{i,n}) = s$.

Proposition 1.1 *Soit $s \in [0; 1]$. Considèrent une f.d.c F est continue et qu'il existe une seule solution x_s à l'équation $F(x) = s$ Soit k_n une suite d'entiers telle que $1 \leq k_n \leq n$ et $(k_n/n) \rightarrow s$ quand $n \rightarrow +\infty$. Alors, $X_{k_n,n}$ converge presque sûrement vers x_s .*

Bien sûr, on peut facilement étudier le minimum au lieu du maximum. De toute évidence, les résultats de l'un peuvent être transférés immédiatement à l'autre grâce à la relation pour tout $i = 1, \dots, n$:

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

Les statistiques d'ordre et la théorie associée ont de nombreuses applications intéressantes et importantes en statistique, dans la modélisation de phénomènes empiriques, par exemple les caractéristiques climatiques, et dans la théorie des probabilités elle-même. Pour

en savoir plus sur la vaste littérature sur les applications des statistiques d'ordre, nous recommandons, parmi de nombreuses possibilités, Lehmann (1975), Galambos (1978), Shorack et Wellner (1986), David (1980), Reiss (1989), Martynov (1992), Falk, Hüsler et Reiss (1994), Coles (2001), Embrechts, Klüppelberg et Mikosch (2008).

Nous nous occupons maintenant principalement du problème de la distribution du maximum d'un échantillon. Pour cela, nous étudierons à la fois les limites possibles et les différentes manières de décrire les ensembles de distributions sur lesquels convergent les suites d'extrémités vers ces limites. C'est ce que nous aborderons dans le prochain chapitre.

Chapitre 2

Statistiques d'indice de queue

2.1 Lois des valeurs extrêmes

On trouve que $F_{X_{n,n}}$ la distribution de $X_{n,n}$ (1.3.6) converge vers 1 si $F(X) = 1$ et converge vers 0 si $F(X) < 1$. On dit que la limite de cette distribution conduit à une loi dégénérée !. On s'intéresse à trouver la distribution du maximum d'un échantillon de v.a's i.i.d. Comment déterminer alors une loi non dégénérée.

2.1.1 Théorème de Fisher et Tippett

Le résultat principal du théorème de la valeur extrême (1.3.6) est le théorème de Fisher et Tippett [13] (1928) (2.1), qui a permis de définir la distribution des extrêmes. Mais, avant de lancer cette théorème en présent la définition le lois max-stable. Donc Une v.a X non dégénérée est max-stable si elle satisfait pour deux suites normalisation $a_n \geq 0$ et $b_n \in \mathbb{R}$ pour tout $n \geq 1$ on a $X_{n,n} \stackrel{d}{=} a_n x + b_n$ où $X_{n,n}$ définie en (1.3.2). Les lois max-stables coïncide avec la classe de toutes les lois limites possibles pour les lois des maxima de v.a's i.i.d, correctement normalisées.

Définition 2.1 Soit un échantillon X_1, X_2, \dots, X_n de v.a's i.i.d suivant la même f.d.c G et la statistique d'ordre extrême $X_{n,n}$ (1.3.2). On dit que G est max-stable si pour tout

$n \geq 1$ il existe deux suites $a_n \geq 0$ et $b_n \in \mathbb{R}$ telque $\frac{X_{n,n} - b_n}{a_n}$ suit la loi G .

En particulier que $\frac{X_{n,n} - b_n}{a_n}$ converge en loi vers une limite non triviale, c'est-à-dire vers une v.a non constante alors la limite est une loi max-stable. De plus la suite $((a_n, b_n), n \geq 1)$ n'est pas unique. Maintenant, nous pouvons énoncer le théorème de Fisher-Tippett [13] (1928) qui donne le clés pour caractériser la distribution des valeurs extrêmes.

Théorème 2.1 (Fisher-Tippett, 1928) Soit X_1, X_2, \dots, X_n une suite de v.a's i.i.d suivant la même loi F . Supposons qu'il existe deux suites réelle $a_n \geq 0$ et $b_n \in \mathbb{R}$ alors,

$$\forall n \geq 1, P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow +\infty} F^n(a_n x + b_n) := H(X), \forall x \in \mathbb{R} \quad (2.1.1)$$

où H est loi max-stable.

On dit que H est une fonction de distribution non dégénérée, appelée lois des valeurs extrêmes. Notons que le théorème (2.1) coïncide avec TCL (1.4) en ce qui concerne les maxima (Analogie avec la suite a_n joue le rôle de $\sigma\sqrt{n}$ et la suite b_n joue le rôle de $n\mu$ et $X_{n,n}$ joue le rôle de S_n). Interprétons avec le théorème (2.1) que la suit a_n (resp b_n) comme un paramètre d'échelle (resp un paramètre de position). A partir de la formule (2.1.1) de théorème (2.1) on trouve qu'il équivalente,

$$\lim_{n \rightarrow +\infty} n \log(F(a_n x + b_n)) := \log(H(X)), 0 < H(X) < 1. \quad (2.1.2)$$

2.1.2 Théorème de Gnedenko

En théorie des valeurs extrêmes le théorème (2.2) de Gnedenko [14] (1943) relatif à la distribution asymptotique des statistiques d'ordre extrêmes (1.3.2). Après le théorème (2.1) en peut converger en loi la distribution des valeurs extrêmes (EVD: Extremes values

Distribution) H (2.1.1) vers 3 types de loi : la loi de Gumbel, la loi de Fréchet, ou la loi de Weibull. Voir le théorème suivant.

Théorème 2.2 (Gnedenko, 1943) *Sous certaines conditions de régularité sur f.d.c F . il existe un paramètre $\gamma \in \mathbb{R}$ telles que la fonction de répartition H (2.1.1) de la loi des valeurs extrêmes peut prendre trois distributions des valeurs extrêmes (EVD) standard de type I et type II et type III (respectivement) comme suivante :*

$$\textit{Distribution de Gumbel: } \Lambda(x) := \exp(-\exp(-x)) \text{ si } x \in \mathbb{R} \text{ et } \gamma = 0 \quad (2.1.3)$$

$$\textit{Distribution de Fréchet : } \Phi_\gamma(x) := \begin{cases} \exp\left(-x^{-\frac{1}{\gamma}}\right) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \text{ et } \gamma > 0 \quad (2.1.4)$$

$$\textit{Distribution de Weibull : } \Psi_\gamma(x) := \begin{cases} 1 & \text{si } x > 0 \\ \exp\left(-(-x)^{-\frac{1}{\gamma}}\right) & \text{si } x \leq 0 \end{cases} \text{ et } \gamma < 0 \quad (2.1.5)$$

Pour la démonstration de théorème (2.2) voir [33].

2.2 Distribution des Valeurs Extrêmes Généralisées(La Distribution de GEV)

2.2.1 Définition la distribution de GEV

Grâce aux travaux de von Mises [39] (1936) et de Jankinson [22] (1955). On peut rassembler les trois familles de distributions des valeurs extrêmes en une seule famille paramétrique noté par H_γ pour tout $\gamma \in \mathbb{R}$, mais toujours à un facteur de changement d'échelle et de position près, c'est appelé distribution des valeurs extrêmes généralisées (distribution GEV: General Extreme Value). La distribution généralisée des valeurs extrêmes (The generalized extreme value distribution :GEVD) est une famille de distributions qui sont généralement utilisées pour modéliser les maxima de longues séquences de

variables aléatoires.

Théorème 2.3 *La fonction de distribution H_γ de la famille des distributions GEV, elle est paramétrée par une seule paramètre $\gamma \in \mathbb{R}$. Est donnée par:*

$$H_\gamma(x) := \begin{cases} \exp\left(- (1 + \gamma x)^{\frac{-1}{\gamma}}\right) & \text{si } \gamma \neq 0, \text{ telque } 1 + \gamma x > 0 \\ \exp(-\exp(-x)) & \text{si } \gamma = 0, \text{ pour tout } x \in \mathbb{R} \end{cases} \quad (2.2.1)$$

où γ s'appelle indice des valeurs extrêmes ou indice de queue.

Remarque 2.1 *D'après le théorème (2.2) on peut réécrire les trois types qui définissent les distributions GEV pour tout $x \in \mathbb{R}$ comme suit:*

$$\begin{cases} \text{Distribution de Gumbel: } \Lambda(x) := H_0(x) & \text{si } \gamma = 0 \\ \text{Distribution de Fréchet: } \Phi_\gamma(x) := H_\gamma\left(\frac{1}{\gamma}(x-1)\right) & \text{si } \gamma > 0 \\ \text{Distribution de Weibull: } \Psi_\gamma(x) := H_\gamma\left(\frac{1}{\gamma}(-x-1)\right) & \text{si } \gamma < 0 \end{cases}$$

Vient ensuite, la Fig (2.1) le graphe de ces distributions et la Fig (2.2) le graphe de leurs intensités pour des valeurs données de γ . Les courbes en rouge, en noir foncé et en bleu correspondent respectivement aux lois EVD dans le domaine d'attraction de Weibull, de Gumbel et de Fréchet d'indices $\gamma = -1; \gamma = 0; \gamma = 1$.

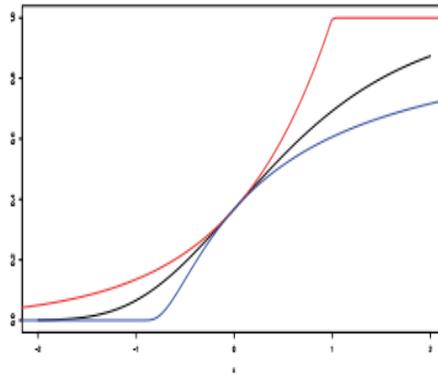


Figure 2.1: La fonction de répartition de la loi EVD

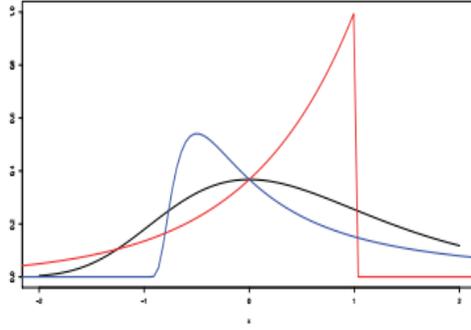


Figure 2.2: La densité de la distribution EVD

2.2.2 Fonctions à variation régulière de premier ordre

Le concept de variation régulière du premier ordre est utilisé pour montrer la consistance des estimateurs d'indice des valeurs extrêmes.

Théorème 2.4 *Considérons une fonction f mesurable et positive définie sur $[0; +\infty[$. On dit f est à variation régulière d'indice γ si et seulement si:*

$$\forall t \geq 0, \lim_{x \rightarrow +\infty} \frac{f(tx)}{f(x)} := t^\gamma. \quad (2.2.2)$$

et on note $f \in RV_\gamma$. Où RV_γ l'ensemble des fonctions à variations régulières d'indice γ .

On peut bien définir que si $f(x) \in RV_\gamma$ (2.4), alors $f(1/x) \in RV_{-\gamma}$ pour tout $x > 0$. De plus si $f(x) \in RV_\gamma$ alors $f^{-1}(x) \in RV_{1/\gamma}$ avec $\gamma > 0$ et $f^{-1}(x) \in RV_{-1/\gamma}$ avec $\gamma < 0$ (voir [30] et [33])

Définition 2.2 *Si $\gamma = 0$ on note $f \in RV_0$ ou f est à variation régulière d'indice 0. On dit que f est à variation lente, C'est-à-dire*

$$\forall t \geq 0, \lim_{x \rightarrow +\infty} \frac{f(tx)}{f(x)} := 1. \quad (2.2.3)$$

et on note $f \in SV$ (Slowly Varying). Telle que f est toujours mesurable.

Corollaire 2.1 Soit $f \in RV_\gamma$ avec $\gamma \neq 0$. Alors il existe une fonction $\ell \in RV_0$, telle que :

$$f(x) = x^\gamma \ell(x) \quad (2.2.4)$$

Proposition 2.1 Soit $\ell \in RV_0$. Alors pour tout $\gamma > 0$, nous avons

$$\lim_{x \rightarrow +\infty} x^\gamma \ell(x) = \infty \quad \text{et} \quad \lim_{x \rightarrow +\infty} x^{-\gamma} \ell(x) = 0, \quad \lim_{x \rightarrow +\infty} \frac{\log(\ell(x))}{\log(x)} = 0$$

Donc a partir de ce proposition(2.1) et corollaire (2.1), quand $x \rightarrow \infty$, $f(x) = \infty$ si $\gamma > 0$ et $f(x) = 0$ si $\gamma < 0$

Théorème 2.5 Soit une fonction $f \in RV_\gamma$. On dit que a est une fonction auxiliaire de f si et seulement si:

$$\forall x > 0, \lim_{t \rightarrow +\infty} \frac{f(tx) - f(t)}{a(t)} = \Psi(x) := \begin{cases} c \frac{x^\gamma - 1}{\gamma} & \text{si } \gamma \neq 0 \\ c \log(x) & \text{si } \gamma = 0 \end{cases}. \quad (2.2.5)$$

avec $c \neq 0$ et la fonction Ψ n'est pas constante. Telle que a est une fonction mesurable et positive.

La preuve de théorème (2.5) découle avec la démonstration de théorème (2.3). Pour l'instant voir le référence [18].

Théorème 2.6 (Théorème de Karamata) Si $f \in RV_\gamma$, il exist deux fonctions mesurables $\eta : \mathbb{R}^+ \rightarrow \mathbb{R}$ et $\varepsilon : \mathbb{R}^+ \rightarrow \mathbb{R}$ avec

$$\lim_{t \rightarrow +\infty} \varepsilon(t) = \varepsilon_0 \quad \text{et} \quad \lim_{t \rightarrow +\infty} \eta(t) = \gamma \quad (2.2.6)$$

et

$$\forall t_0 > 0, f(t) = \varepsilon(t) \exp \left(\int_{t_0}^t \frac{\eta(s)}{s} ds \right) \quad \text{telle que } t > t_0 \quad (2.2.7)$$

Intuitivement, si (2.2.7) exist et les deux fonctions η et ε sont vérifiées (2.2.6) alors $f \in RV_\gamma$. Pour la preuve le théorème de Karamata voir ([18], la démonstration du Theorem B.1.6, Appendix B)

Lemme 2.1 (Resnick,1987) Soit $f \in RV_\gamma$ et pour tout $0 < a < b$. Alors

$$\lim_{x \rightarrow +\infty} \sup_{t \in [a,b]} \left| \frac{f(tx)}{f(x)} - t^{-\gamma} \right| = 0$$

Le lemme (2.1) montre que la convergence de deux fonctions à variation régulière d'indice γ est localement uniforme lorsque $x \rightarrow +\infty$.

Théorème 2.7 (Borne de Potter) Soit $f \in RV_\gamma$, alors pour tout les deux constants $\epsilon > 0$ et $C > 1$ il existe x_0 telque

$$\forall y \geq x \geq x_0, C^{-1} \left(\frac{y}{x} \right)^{\gamma-\epsilon} \leq \frac{f(y)}{f(x)} \leq C \left(\frac{y}{x} \right)^{\gamma+\epsilon} \quad (2.2.8)$$

2.2.3 Fonctions à variation régulière du deuxième ordre

La variation régulière du deuxième ordre est utilisé pour preuve la normalité asymptotique des estimateurs d'indice des valeurs extrêmes.

Définition 2.3 (Fonction quantile de queue) On dit U est une fonction quantile de queue. S'elle définie par :

$$U(x) := Q \left(1 - \frac{1}{x} \right) = F^{-1} \left(1 - \frac{1}{x} \right) \quad (2.2.9)$$

où F^{-1} est l'inverse généralisé de F .

Théorème 2.8 (de Haan and Stadtmüller (1996), [19]) On dit que la fonction U (2.2.9) est à variation régulière du deuxième ordre si pour une fonction positive a et une fonction positive ou négative A ,

$$\forall x > 0, \lim_{t \rightarrow +\infty} \frac{U(tx) - U(t)}{a(t)} - h_\gamma(x) = A(t) \psi(x) \quad (2.2.10)$$

Telle que $|A| \in RV_\rho$ avec $\lim_{t \rightarrow +\infty} A(t) = 0$ et la fonction $h_\gamma \in RV_\gamma$ vérifie:

$$h_\gamma(x) := \int_1^x t^{\gamma-1} dt, \quad (2.2.11)$$

Et la fonction $\psi \in RV_\rho$ est donnée par :

$$\psi(x) := c_1 \int_1^x s^{\gamma-1} \left(\int_1^s u^{\rho-1} du \right) ds + c_2 \int_1^x s^{\gamma+\rho-1} ds \quad (2.2.12)$$

où $\rho \leq 0$ et $(c_1, c_2) \in \mathbb{R}^2$ sont des constants. On note $U \in RV_{\gamma,\rho}$ ou $U \in 2RV_{\gamma,\rho}$.

Notons que les fonctions a et A comme fonctions auxiliaires du premier et du deuxième ordre.(respectivement)

Définition 2.4 Soit une fonction $U \in 2RV_{\gamma,\rho}$ avec les paramètres $\gamma > 0$ et $\rho \leq 0$ (respectivement) comme des paramètres du premier et du deuxième ordre . S'il existe une fonction auxiliaire $|A| \in RV_\rho$ avec $\lim_{t \rightarrow +\infty} A(t) = 0$ qui ne change pas le signe au voisinage de ∞ telles que,

$$\forall x > 0, \lim_{t \rightarrow +\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \quad (2.2.13)$$

on dit que U satisfait la condition du deuxième ordre

2.2.4 Domaines d'attraction

Définition 2.5 Soit X_1, X_2, \dots, X_n une suite de v.a's i.i.d suivant la même f.d.c F . Le point terminal de F (1.1.1) est défini par :

$$x_F = \sup \{x \in \mathbb{R}, F(x) \leq 1\} \leq \infty \quad (2.2.14)$$

Il est clair que, $X_{n,n}$ (1.3.2) converge en probabilité vers x_F quand $n \rightarrow +\infty$. De plus la distribution de $X_{n,n}$ (1.3.6) converge vers 1 si $x \geq x_F$ et converge vers 0 si $x < x_F$. Sous

l'inégalité (2.1.2) s'ensuit que $F(a_n x + b_n) \rightarrow 1$ pour chaque x . Donc on trouve que

$$\lim_{n \rightarrow +\infty} \frac{n \log(F(a_n x + b_n))}{1 - F(a_n x + b_n)} := 1$$

Alors l'inégalité (2.1.2) est équivalente à

$$\lim_{n \rightarrow +\infty} n(1 - F(a_n x + b_n)) := -\log(H(X)), 0 < H(X) < 1. \quad (2.2.15)$$

où H est une lois max-stables pour deux suites de normalisation $a_n \geq 0$ et $b_n \in \mathbb{R}$ pour tout $n \geq 1$. En particulier on peut réécrire la formule (2.2.15) comme suit:

$$\forall x > 0, \lim_{t \rightarrow +\infty} t(1 - F(a(t)x + b(t))) := -\log(H(X)), 0 < H(X) < 1.$$

avec $a(t) = a_{[t]}$ et $b(t) = b_{[t]}$ où $[t]$ la partie entière de t . Déjà vu avec le théorème (2.1) de Fisher et Tippett (1928), le théorème (2.2) de Gnedenko(1943) et le théorème (2.3) que $H = H_\gamma$ où H_γ est la distribution des valeurs extrêmes où γ s'appelle indice des valeurs extrêmes. Une manière courante suggérée dans la littérature pour étudier le comportement de la queue consiste à utiliser le logarithme de l'ensemble des données d'origine afin de réduire la variabilité de l'échantillon. Dans ce cas on dit que F appartient au domaine d'attraction des distributions des valeurs extrêmes. Notation: $F \in D(H_\gamma)$.

Théorème 2.9 *Soit $\gamma \in \mathbb{R}$, on dit que $F \in D(H_\gamma)$ s'il existe une fonction mesurable $a(t)$ et pour tout $1 + \gamma x > 0$ avec $x \in \mathbb{R}$ on a,*

$$\lim_{t \rightarrow x_F} \frac{1 - F(a(t)x + t)}{1 - F(t)} := \begin{cases} (1 + \gamma x)^{\frac{-1}{\gamma}} & \text{si } \gamma \neq 0, \\ \exp(-x) & \text{si } \gamma = 0. \end{cases} \quad (2.2.16)$$

Pour la preuve de ce théorème voir ([18], la démonstration du Theorem 1.1.6, page 11). Le théorème suivant énonce une condition suffisante pour déterminer qu'une fonction de répartition $F \in D(H_\gamma)$, cette condition est basé sur le travail de Von Mises (1936).

Théorème 2.10 (*condition de Von Mises*) Soit une fonction de répartition F . Supposons que la fonction de densité $f(x)$ associé est positive pour tout $x < x_F$ avec $f'(x)$ existe. Si

$$\lim_{t \rightarrow x_F} \left(\frac{\overline{F}(t)}{f(t)} \right)' := \gamma \quad (2.2.17)$$

alors on dit que $F \in D(H_\gamma)$, telle que $\gamma \in \mathbb{R}$ où x_F (2.2.14) est le point terminal de F .

Pour la démonstration de ce théorème voir ([18], la preuve du Theorem 1.1.8, page 15).

Par consequence, soit une f.d.c $F \in D(H_\gamma)$ avec $\gamma \in \mathbb{R}$. Selon le signe de γ trois domaines d'attraction sont possible:

1. Si $\gamma > 0$ on dit que F appartient au domaine d'attraction de Fréchet. Notation: $F \in D(\Phi_\gamma)$. Ce domaine d'attraction est bien définit par étudié les caractéristique des distributions à queues lourdes. C'est-à-dire qui ont une fonction de survie à décroissance polynomiale. Comme les lois: de Cauchy, de Pareto, du Chi-deux, de Student, de Fréchet,...etc.
2. Si $\gamma < 0$ on dit que F appartient au domaine d'attraction de Weibull. Notation: $F \in D(\Psi_\gamma)$. Ce domaine d'attraction est définit des fonctions de survie dont le qu'elle est borné supérieurement. Comme les lois: de Uniforme, de Beta,...etc.
3. Si $\gamma = 0$ on dit que F appartient au domaine d'attraction de Gumbel. Notation: $F \in D(\Lambda_\gamma)$. Ce domaine d'attraction est bien définit des distributions à queues légères. C'est-à-dire les fonctions de survie à décroissance exponentielle. Comme les lois: de Normale, de Exponentielle, de Log-normale, de Gamma,...etc.

Proposition 2.2 (*Critère condition de Von Mises*) Soit une fonction de répartition F avec f une fonction de densité associé. Alors

1. Supposons $x_F = \infty$ et la fonction densité f existes, on dit que $F \in D(\Phi_\gamma)$. Si

$$\lim_{t \rightarrow \infty} \frac{\overline{F}(t)}{t f(t)} := \gamma > 0, \quad (2.2.18)$$

2. Supposons $x_F < \infty$ et la fonction densité f existes, on dit que $F \in D(\Psi_\gamma)$. Si

$$\forall x < x_F, \lim_{t \rightarrow x_F} \frac{\bar{F}(t)}{(x_F - t) f(t)} := -\gamma > 0, \quad (2.2.19)$$

Voir (cf., e.g., [18], le théorème 1.1.11 et le théorème 1.1.13, page 17). Notons que les fonctions de répartitions $F \in D(\Psi_\gamma)$ nécessitent un point terminal(2.2.14). C'est-à-dire qu'il n'y a pas de borne supérieure. C'est ce qui rend ces distributions appartenant au domaine d'attraction de Weibull inutiles pour modéliser beaucoup phénomènes. D'autre part, avec les distrinution appartenant au domaine d'attraction de Gumbel ou Fréchet x_F n'est pas affecté.

Théorème 2.11 *Soit $\gamma \in \mathbb{R}$ indice des valeurs extrêmes et F une fonction de répartition avec l'une fonction $\ell \in RV_0$ (2.2). Alors*

1. On dit que $F \in D(\Phi_\gamma)$ avec $\gamma > 0$ et $x_F = \infty$ si et seulement si

$$\bar{F}(x) := x^{-\gamma} \ell(x) \quad (2.2.20)$$

2. On dit que $F \in D(\Psi_\gamma)$ avec $\gamma > 0$ et $x_F < \infty$ si et seulement si

$$\bar{F}\left(x_F - \frac{1}{x}\right) := x^{-\gamma} \ell(x) \quad (2.2.21)$$

3. On dit que $F \in D(\Lambda_\gamma)$ avec $\gamma = 0$ et $x_F = \infty$ si et seulement si, F satisfait le théorème de Karamata (2.6). C'est-à-dire, si $\lim_{x \rightarrow x_F} \varepsilon(x) = \varepsilon_0 > 0$ avec $\lim_{x \rightarrow x_F} \eta'(x) = 0$ et

$$\forall x_0 > 0, \bar{F}(x) = \varepsilon(x) \exp\left(-\int_{x_0}^x \frac{1}{\eta(s)} ds\right) \text{ telle que } x_F > x > x_0 \quad (2.2.22)$$

Pour une présentation exhaustive du domaine d'attraction de Gumbel voir [18]. De plus, on trouve que la suite $(a_n^{-1} X_{n,n}, n \geq 1)$ converge en loi vers la loi de Fréchet . Si $F \in D(\Phi_\gamma)$ et en prend $a_n = U(n) = F^{-1}(1 - 1/n)$ où $X_{n,n}$ est définit en (1.3.2) et U définit en

(2.2.9). Si $F \in D(\Psi_\gamma)$ en prenant $a_n = x_F - U(n) = x_F - F^{-1}(1 - 1/n)$, la suite $(a_n^{-1}X_{n,n} - x_F, n \geq 1)$ converge en loi vers la loi de Weibull. D'autre part si $F \in D(\Lambda_\gamma)$ en prenant les suites de normalisation $a_n = x_F - F^{-1}(1 - 1/n)$ et $b_n = (1/\bar{F}(x)) \int_{a_n}^{x_F} \bar{F}(s) ds$, la suite $(a_n^{-1}(X_{n,n} - b_n), n \geq 1)$ converge en loi vers la loi de Gumbel. Le tableau suivant résume le classement de quelques lois usuelles selon leur appartenance à l'un des domaines d'attraction.

Domaines d'attraction	Fréchet $\gamma > 0$	Gumbel $\gamma = 0$	Weibull $\gamma < 0$
Lois	Pareto	Exponentielle	Uniforme
	Log-gamma	Normale	Beta
	Student	Log-normale	
	Burr	Gamma	
	Chi-deux		

Table 2.1: Quelques loi usuelles classées en fonction de leur domaine d'attraction

2.3 Distribution de Pareto généralisée (GPD)

La distribution de Pareto généralisée (GPD) joue un rôle central dans la modélisation des phénomènes de queue lourde dans diverses applications en finance, risque opérationnel, assurance et études environnementales.

Définition 2.6 (*Loi de Pareto*) On dit qu'une v.a réelle positive suit la lois de Pareto (*distribution à queue lourde*) si $\bar{F} \in RV_{-\gamma}$ pour $\gamma > 0$. C'est-à-dire,

$$\forall x \geq 0, \lim_{t \rightarrow +\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} := x^{-\gamma}. \quad (2.3.1)$$

où γ est appelé l'indice de Pareto (l'indice de queue) de la fonction de répartition F avec $F \in D(H_\gamma)$.

Définition 2.7 (*loi de Pareto de première espase*) On dit que la fonction de répartition

F est une loi de Pareto de première espèce si elle satisfait pour tout $x > x_0$:

$$F(x) = 1 - \left(\frac{x_0}{x}\right)^\gamma \quad (2.3.2)$$

2.3.1 Définition de la GPD

Le comportement de grandes observations qui dépassent un certain seuil est décrit par la distribution de Pareto généralisée (GPD : Generalized Pareto Distribution) dont la présentation est la suivante:

Définition 2.8 La fonction de distribution cumulative de Pareto généralisée est généralement exprimé comme une distribution à deux paramètres $\gamma \in \mathbb{R}$ et $\sigma > 0$, par

$$F_{\gamma,\sigma}(x) := \begin{cases} 1 - \left(1 + \frac{\gamma}{\sigma}x\right)^{\frac{-1}{\gamma}} & \text{si } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \text{si } \gamma = 0 \end{cases} \quad (2.3.3)$$

pour tout $x > 0$, si $\gamma > 0$ et $0 < x < -\sigma/\gamma$ si $\gamma < 0$, et pour tout $x > 0$, avec $\gamma = 0$. Où σ s'appelle paramètre d'échelle et γ s'appelle paramètre de shape

La GPD est une distribution de probabilité à deux paramètres. Si $\gamma > 0$ nous avons une version reparamétrée de la distribution de Pareto habituelle (2.6). Si $\gamma < 0$ nous avons une distribution de Pareto de type II. Si $\gamma = 0$ donne la distribution exponentielle. En particulier avec $\sigma = 1$, la GPD (2.3.3) est défini comme une GPD standard. De plus on peut étendre la famille de GPD en ajoutant un paramètre de localisation u la GPD (2.3.3) connu comme $F_{\gamma,\sigma}(x - u)$. Soit x_F (2.2.14) le point terminal de la fonction de distribution F et considérons $F_t(x)$ (2.3.6) la distribution des excès de X au-dessus du seuil t . Alors c'est bien connu le théorèmes de Balkema et de Haan [2] et Pickands [38] montrent que si $F \in D(H_\gamma)$ alors il existe une fonction de normalisation $\sigma(t)$ positive, tel que pour tout x

$$\lim_{t \rightarrow x_F} \sup_{0 < x < x_F - t} |F_t(x) - F_{\gamma,\sigma(t)}(x)| = 0, \quad (2.3.4)$$

où $F_{\gamma,\sigma(t)}(x)$ est la GPD (2.3.3). Ce théorème suggère que, pour des seuils t tel que $X > t$ la fonction de distribution des excès $F_t(x)$ peut être approchée par la GPD (2.3.3). Par conséquent, la GPD (2.3.3) joue un rôle très important dans les méthodes statistiques utilisées pour estimer les distributions des plus grandes observations.

Proposition 2.3 *Si X une v.a suit la GPD (2.3.3). Alors la distribution des excès de $X - t$ avec $X > t$ est suit la GPD comme $F_{\gamma,\sigma(t)}(x - t)$ avec même paramètre γ et la fonction d'excès moyen (2.3.7) est défini par,*

$$E_t(x) = \frac{\sigma + \gamma t}{1 - \gamma}, \text{ telque } \gamma < 1 \quad (2.3.5)$$

La GPD (2.3.3) est le seul « stables », c'est-à-dire le seul pour lequel la distribution conditionnelle d'un dépassement est une transformation d'échelle de la distribution d'origine, et peuvent être utilisés pour estimer l'indice de valeur extrême. La GPD (2.3.3) a bonnes propriétés (pour plus de détails, voir, par exemple, Embrechts, Klüppelberg et Mikosch, 1997, Section 3.4, [12], et Reiss et Thomas, Section 1.4, [34]).

Dans la Fig.(2.3), nous présentons la fonction de densité GPD avec les paramètres $\sigma = 1$ et $\gamma = 0, \gamma = 1$ et $\gamma = 2$ sélectionnés côte à côte

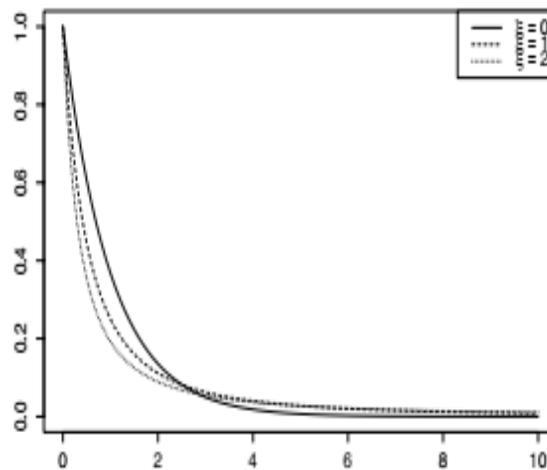


Figure 2.3: La PDF pour GPD avec les paramètres $(0, 1)$, $(1, 1)$, $(2, 1)$

2.3.2 Distribution des excès

Soit X_1, X_2, \dots, X_n une suite de v.a's i.i.d suivant la même fonction de répartition F , de point terminal x_F (2.2.14). On dit que Y_i pour $i = 1, \dots, n$ sont une v.a des excès au-dessus du seuil t si données sont par $Y_i = X_i - t$ qui présentent le modèle de piques au-delà du seuil t (POT : Peaks over Threshold t). La distribution des excès de X au-dessus du seuil t est donné par une probabilité conditionnelle pour tout $X > t$ comme suivant

$$F_t(x) = P(X - t \leq x | X > t) = \frac{F(t+x) - F(t)}{1 - F(t)}. \quad (2.3.6)$$

avec $1 - F(t) > 0, t < x_F$ et $x > 0$. Une autre fonction qui joue un rôle important dans le modèle de piques au-dessus du seuil t est fonction d'excès moyen qui est l'espérance conditionnelle de piques au-dessus du seuil t . La fonction d'excès moyen est définie comme

$$E_t(x) = E(X - t \leq x | X > t). \quad (2.3.7)$$

2.4 Estimation d'indice des valeurs extrêmes

2.4.1 Estimateur de maximum de vraisemblance ($\gamma > (-1/2)$)

Soit X une v.a's de fonction de répartition F avec $F \in D(H_\gamma)$. considérons $X - t$ une v.a des excès au-dessus du seuil t pour tout $X > t$ qui définit le modèle de piques au-delà du seuil t . D'après la relation (2.2.16), la distribution conditionnelle de $(X - t)/a(t)$ pour $X > t$ est donnée par,

$$\lim_{t \rightarrow x_F} P\left(\frac{X - t}{a(t)} > x | X > t\right) := F_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{\frac{-1}{\gamma}} & \text{si } \gamma \neq 0, \\ 1 - \exp(-x) & \text{si } \gamma = 0. \end{cases} \quad (2.4.1)$$

où F_γ la GPD standard avec a une fonction positive et croissante. Cette relation indique que les plus grandes observations déposant $t (X > t)$ suivent approximativement une GPD

standard. Cela suggère que si nous appliquons la procédure du maximum de vraisemblance aux plus grandes observations en utilisant la GPD standard pour donner un estimateur de γ indice des valeurs extrêmes, appelé l'estimateur de maximum de vraisemblance (ML-estimator:Maximum Likelihood estimator). Considérons $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ la statistique d'ordre de l'échantillon X_1, X_2, \dots, X_n de v.a's i.i.d suit la même distribution F avec $F \in D(H_\gamma)$. Précédent, on trouve que pour estimer l'indice de queue il faut générer un échantillon $X_{n-k,n}, X_{n-k+1,n}, \dots, X_{n,n}$ pour $X > X_{n-k,n}$ ou un échantillon $C_0, C_1, C_2 \dots, C_k$ où

$$C_0 = X_{n-k,n}, C_1 = X_{n-k+1,n} - X_{n-k,n}, \dots, C_k = X_{n,n} - X_{n-k,n}, \quad (2.4.2)$$

où dans l'asymptotique on pose que $k = k_n$ est une séquence intermédiaire, c'est-à-dire, $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$ quand $n \rightarrow \infty$ et donc $X_{n-k,n} \rightarrow x_F$ où x_F le point terminal de F (2.2.14). Telle que la distribution de l'échantillon C_i pour $i = 1, \dots, k$ avec $C_0 = X_{n-k,n}$ de v.a's i.i.d est la distribution conditionnelle $F_t(x)$, qu'elle définit en la relation (2.3.6). Puis trouvons que d'après la relation (2.4.1) $F_t(x)$ est approximé par $F_{\gamma,\sigma}$ est la GPD (2.3.3) donc en appliquant le maximum de vraisemblance avec la GPD (2.3.3) pour estimer γ indice des valeurs extrêmes, cette méthode appelle l'approche de la vraisemblance conditionnelle. Notons que la GPD (2.3.3) il a deux paramètres qui sont γ l'indice de queue et σ paramètre d'échelle, alors le maximum de vraisemblance de la GPD définit l'estimateur de γ et l'estimateur de σ . Donc l'estimateur de γ et l'estimateur de σ sont obtenus en maximisant par rapport à γ et σ la fonction du maximum de vraisemblance avec la GPD suivante:

$$\ell(C_i; (\gamma, \sigma)) = \prod_{i=1}^k f_{\gamma,\sigma}(C_i), \quad (2.4.3)$$

avec $C_i = X_{n-i+1,n} - X_{n-k,n}$ pour tout $i = 1, \dots, k$ où $f_{\gamma,\sigma}(C_i)$ fonction de densité la GPD associée est donnée par :

$$f_{\gamma,\sigma}(x) := \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\gamma}{\sigma}x\right)^{\frac{-1}{\gamma}-1} & \text{si } \gamma \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right) & \text{si } \gamma = 0 \end{cases} \quad (2.4.4)$$

pour tout $x > 0$, si $\gamma > 0$ et $0 < x < -\sigma/\gamma$ si $\gamma < 0$, et pour tout $x > 0$, avec $\gamma = 0$. Donc le logarithme de maximum de vraisemblance avec la GPD du (2.4.3) est défini par:

$$\log(\ell(C_i; (\gamma, \sigma))) = \sum_{i=1}^k \left(\log\left(\frac{1}{\sigma}\right) - \left(1 + \frac{1}{\gamma}\right) \log\left(1 + \frac{\gamma}{\sigma}C_i\right) \right) \quad (2.4.5)$$

où $\sigma > 0$ pour $\gamma > 0$ et $\sigma > -\gamma C_{k,k}$ pour $\gamma \leq 0$ avec $C_{k,k} = \max(C_i)$ pour tout $i = 1, \dots, k$. On remarque que si $\gamma < -1$ il n'y a pas d'estimation du maximum de vraisemblance puisque, pour tout $\gamma < -1$ on trouve que $\log(\ell(C_i)) \rightarrow \infty$ quand $(\sigma/\gamma) \rightarrow C_{k:k}^+$. Donc nous concentrerons sur l'espace $\sigma > 0$ et $\gamma \geq -\frac{1}{2}$ parce que l'estimateur du maximum de vraisemblance se comporte de façon irrégulière si $\gamma \leq -\frac{1}{2}$. Les équations de vraisemblance sont données en termes de dérivées partielles du (2.4.5) par rapport rapport à γ (σ respect) par

$$\begin{aligned} \frac{\partial \log(\ell(C_i; (\gamma, \sigma)))}{\partial \gamma} &= \sum_{i=1}^k \left(\frac{1}{\gamma^2} \log\left(1 + \frac{\gamma}{\sigma}C_i\right) - \left(1 + \frac{1}{\gamma}\right) \frac{\frac{C_i}{\sigma}}{1 + \frac{\gamma}{\sigma}C_i} \right) \\ \frac{\partial \log(\ell(C_i; (\gamma, \sigma)))}{\partial \sigma} &= \sum_{i=1}^k \left(-\frac{1}{\sigma} + \left(1 + \frac{1}{\gamma}\right) \frac{\frac{\gamma}{\sigma^2}C_i}{1 + \frac{\gamma}{\sigma}C_i} \right) \end{aligned} \quad (2.4.6)$$

Et pour $\gamma = 0$ ces termes doivent être interprétés par le développement de Taylor du système (2.4.6), comme suivant

$$\begin{aligned} \frac{\partial \log(\ell(C_i; (\gamma, \sigma)))}{\partial \gamma} \Big|_{\gamma=0} &= \frac{1}{2} \sum_{i=1}^k \left(\frac{C_i}{\sigma}\right)^2 - \sum_{i=1}^k \frac{C_i}{\sigma} \\ \frac{\partial \log(\ell(C_i; (\gamma, \sigma)))}{\partial \sigma} \Big|_{\gamma=0} &= -\sum_{i=1}^k \frac{1}{\sigma} + \sum_{i=1}^k \left(\frac{C_i}{\sigma}\right)^2 \end{aligned}$$

À partir (2.4.6) les équations de vraisemblance résultantes en termes de $C_i = X_{n-i+1,n} - X_{n-k,n}$ pour $\gamma \neq 0$ sont donnant par

$$\begin{cases} \sum_{i=1}^k \left[\log \left(1 + \frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n}) \right) - (\gamma + 1) \frac{\frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n})}{1 + \frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n})} \right] = 0 \\ \sum_{i=1}^k \left(1 + \frac{1}{\gamma} \right) \frac{\frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n})}{1 + \frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n})} = k \end{cases} \quad (2.4.7)$$

Puis, on peut simplifier le resultat (2.4.7) comme,

$$\begin{cases} \frac{1}{k} \sum_{i=1}^k \log \left(1 + \frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n}) \right) = \gamma \\ \frac{1}{k} \sum_{i=1}^k \frac{1}{1 + \frac{\gamma}{\sigma} (X_{n-i+1,n} - X_{n-k,n})} = \frac{1}{\gamma + 1} \end{cases} \quad (2.4.8)$$

Nous donnons maintenant la proposition suivant pour prouver le théorème (2.12). Cette proposition dépend de ce qu'il a fait Drees et al [10] pour $\gamma > \frac{-1}{2}$. Avant de lance cette proposition nous définitions un processus gaussien avec de moyenne et de covariance connue $X_n(s)$ pour tout $\varepsilon > 0$ suivant

$$X_n(s) := s^{-\gamma-1} W_n(s) - W_n(1) + \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \Psi_{\gamma,\rho} \left(\frac{1}{s} \right) + o_p(1) s^{-\gamma-1/2-\varepsilon} \quad (2.4.9)$$

où $W_n(s)$ est une séquence de motions browniennes avec $s \in [0; 1]$ avec une fonction positive ou négative \tilde{A} et $\Psi_{\gamma,\rho}(x)$ est donnée en relation (2.2.12) avec $c_2 = 0$. C'est-à-dire :

$$\Psi_{\gamma,\rho}(x) := c_1 \int_1^x s^{\gamma-1} \left(\int_1^s u^{\rho-1} du \right) ds \quad (2.4.10)$$

Proposition 2.4 *Soit X_1, X_2, \dots, X_n une suite de v.a.'s i.i.d suivant la même loi F . Supposons que F satisfait la condition du deuxième ordre du théorème (2.8) pour $\gamma > \frac{-1}{2}$. Et pour $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$ et une fonction auxiliaire $|\tilde{A}| \in RV_\rho$ avec*

$\lim_{n \rightarrow +\infty} \tilde{A}\left(\frac{n}{k}\right) < \infty$. Toute solution $(\tilde{\gamma}, \tilde{\sigma})$ de le système d'équations (2.4.8) satisfaisante que

$$\left| \begin{array}{c} \tilde{\gamma} \\ \tilde{\sigma} \end{array} - \gamma \right| = O_p\left(\frac{1}{\sqrt{k}}\right) \quad (2.4.11)$$

avec γ est le paramètre correct. Alors quand $n \rightarrow \infty$ le système d'équations (2.4.8) admet l'approximation suivante:

$$\begin{aligned} \sqrt{k}(\tilde{\gamma} - \gamma) - \frac{(\gamma + 1)^2}{\gamma} \int_0^1 (s^\gamma - (2\gamma + 1)s^{2\gamma}) X_n(s) ds &= o_p(1) \\ \sqrt{k}(\tilde{\sigma} - 1) - \frac{\gamma + 1}{\gamma} \int_0^1 ((\gamma + 1)(2\gamma + 1)s^{2\gamma} - s^\gamma) X_n(s) ds &= o_p(1) \end{aligned} \quad (2.4.12)$$

où $s \in [0; 1]$ et $X_n(s)$ est défini en (2.4.9).

Pour la preuve de cette proposition voir ([18], la démonstration de la proposition 3.4.4, p 96).

Théorème 2.12 Soit X_1, X_2, \dots, X_n une suite de v.a's i.i.d suivant la même distribution F . Supposons que F satisfait la condition du deuxième ordre du théorème (2.8) pour $\gamma > \frac{-1}{2}$ avec $k \rightarrow \infty$ et $k/n \rightarrow 0$ as $n \rightarrow \infty$ et une fonction auxiliaire $|A| \in RV_\rho$ avec $\lim_{n \rightarrow +\infty} A\left(\frac{n}{k}\right) = \lambda < \infty$. Alors, le système d'équations (2.4.8) de vraisemblance a une suite de solutions $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ qui satisfait

$$\sqrt{k}(\hat{\gamma}^{ML} - \gamma) \xrightarrow{d} \mathfrak{N}(\lambda\mu_a, \sigma_a^2) \quad (2.4.13)$$

où

$$\mu_a := \begin{cases} \frac{\gamma + 1}{(1 - \rho)(1 + \gamma - \rho)} & \text{si } \rho < 0 \\ 1 & \text{si } \rho = 0 \end{cases} \quad \text{avec } \sigma_a^2 := (\gamma + 1)^2$$

Et

$$\sqrt{k} \left(\frac{\hat{\sigma}^{ML}}{a\left(\frac{n}{k}\right)} - 1 \right) \xrightarrow{d} N(\lambda\mu_b, \sigma_b^2) \quad (2.4.14)$$

où

$$\mu_b := \begin{cases} \frac{-\rho}{(1-\rho)(1+\gamma-\rho)} & \text{si } \rho < 0 \\ 0 & \text{si } \rho = 0 \end{cases} \quad \text{avec } \sigma_b^2 := 1 + (\gamma + 1)^2$$

où \aleph est la loi normal. Avec

$$\text{Cov} \left(\hat{\gamma}^{ML} - \gamma, \frac{\hat{\sigma}^{ML}}{a\left(\frac{n}{k}\right)} - 1 \right) := -\gamma - 1 \quad (2.4.15)$$

Preuve. D'après la proposition (2.4), en remplaçant le processus gaussien $X_n(s)$ défini en (2.4.9) dans la relation (2.4.12). La séquence de solution $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ du système d'équations (2.4.8) quand $n \rightarrow \infty$ satisfaisant:

$$\begin{aligned} & \sqrt{k} (\hat{\gamma}^{ML} - \gamma) - \frac{(\gamma + 1)^2}{\gamma} \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \int_0^1 (s^\gamma - (2\gamma + 1) s^{2\gamma}) \Psi_{\gamma, \rho} \left(\frac{1}{s} \right) ds \\ & \xrightarrow{d} \frac{(\gamma + 1)^2}{\gamma} \int_0^1 (s^\gamma - (2\gamma + 1) s^{2\gamma}) (s^{-\gamma-1} W_n(s) - W_n(1)) ds \end{aligned}$$

Sous la relation (2.4.1) et la relation (2.2.16) on a: $\sigma \rightarrow \tilde{a}(t)$ quand $n \rightarrow \infty$ donc $(\sigma/\tilde{a}(t)) \rightarrow 1$. Alors quand $n \rightarrow \infty$ on trouve,

$$\begin{aligned} & \sqrt{k} \left(\frac{\hat{\sigma}^{ML}}{\tilde{a}\left(\frac{n}{k}\right)} - 1 \right) - \frac{\gamma + 1}{\gamma} \sqrt{k} \tilde{A} \left(\frac{n}{k} \right) \int_0^1 ((\gamma + 1)(2\gamma + 1) s^{2\gamma} - s^\gamma) \Psi_{\gamma, \rho} \left(\frac{1}{s} \right) ds \\ & \xrightarrow{d} \frac{\gamma + 1}{\gamma} \int_0^1 ((\gamma + 1)(2\gamma + 1) s^{2\gamma} - s^\gamma) (s^{-\gamma-1} W_n(s) - W_n(1)) ds \end{aligned}$$

Il existe A une fonction positive ou négative où $|A| \in RV_\rho$ avec $\lim_{n \rightarrow +\infty} A\left(\frac{n}{k}\right) = \lambda$ et telle que quand $t \rightarrow \infty$ on a

$$\tilde{A}(t) \sim A(t) \left(\mathbf{1}_{\{\rho < 0\}} \frac{1}{\rho} + \mathbf{1}_{\{\rho = 0\}} \right)$$

De plus, il existe une fonction positive dit première fonction auxiliaire telle que quand

$t \rightarrow \infty$ on a

$$\left(\frac{\tilde{a}(t)}{a(t)} - 1 \right) \sim \beta \times A(t) \quad \text{où } \beta = \begin{cases} \frac{-1}{\rho} \text{ si } \mathbf{1}_{\{\rho < 0\}} \\ \frac{-1}{\gamma} \text{ si } \mathbf{1}_{\{\gamma \neq \rho = 0\}} \\ 0 \text{ si } \mathbf{1}_{\{\gamma = \rho = 0\}} \end{cases} \quad (2.4.16)$$

Alors, les deux relations (2.2.16) et (2.2.16) être définit respectivement comme

$$\begin{aligned} & \sqrt{k} \left(\hat{\gamma}^{ML} - \gamma \right) - \frac{(\gamma + 1)^2}{\gamma} \left(\mathbf{1}_{\{\rho < 0\}} \frac{1}{\rho} + \mathbf{1}_{\{\rho = 0\}} \right) \sqrt{k} A \left(\frac{n}{k} \right) \\ & \times \int_0^1 (s^\gamma - (2\gamma + 1) s^{2\gamma}) \Psi_{\gamma, \rho} \left(\frac{1}{s} \right) ds \\ & \xrightarrow{d} \frac{(\gamma + 1)^2}{\gamma} \int_0^1 (s^\gamma - (2\gamma + 1) s^{2\gamma}) (s^{-\gamma-1} W_n(s) - W_n(1)) ds \end{aligned}$$

et

$$\begin{aligned} & \sqrt{k} \left(\frac{\hat{\sigma}^{ML}}{a \left(\frac{n}{k} \right)} - 1 \right) - \beta \times \sqrt{k} A \left(\frac{n}{k} \right) - \frac{\gamma + 1}{\gamma} \left(\mathbf{1}_{\{\rho < 0\}} \frac{1}{\rho} + \mathbf{1}_{\{\rho = 0\}} \right) \sqrt{k} A \left(\frac{n}{k} \right) \\ & \times \int_0^1 ((\gamma + 1) (2\gamma + 1) s^{2\gamma} - s^\gamma) \Psi_{\gamma, \rho} \left(\frac{1}{s} \right) ds \\ & \xrightarrow{d} \frac{\gamma + 1}{\gamma} \int_0^1 ((\gamma + 1) (2\gamma + 1) s^{2\gamma} - s^\gamma) (s^{-\gamma-1} W_n(s) - W_n(1)) ds \end{aligned}$$

Avec $\lim_{n \rightarrow +\infty} A \left(\frac{n}{k} \right) = \lambda$ nous savons $\mu_a^* = \lambda \mu_a$ le terme de biais de $\sqrt{k} (\hat{\gamma}^{ML} - \gamma)$ et $\mu_b^* = \lambda \mu_b$ le terme de biais de $\sqrt{k} \left(\frac{\hat{\sigma}^{ML}}{a \left(\frac{n}{k} \right)} - 1 \right)$. Tellque μ_a et μ_b sont données respectivement par:

$$\begin{aligned} \mu_a &= \frac{(\gamma + 1)^2}{\gamma} \left(\mathbf{1}_{\{\rho < 0\}} \frac{1}{\rho} + \mathbf{1}_{\{\rho = 0\}} \right) \times \int_0^1 (s^\gamma - (2\gamma + 1) s^{2\gamma}) \Psi_{\gamma, \rho} \left(\frac{1}{s} \right) ds \quad (2.4.17) \\ \mu_b &= \left(\beta - \frac{\gamma + 1}{\gamma} \left(\mathbf{1}_{\{\rho < 0\}} \frac{1}{\rho} + \mathbf{1}_{\{\rho = 0\}} \right) \right) \times \int_0^1 ((\gamma + 1) (2\gamma + 1) s^{2\gamma} - s^\gamma) \Psi_{\gamma, \rho} \left(\frac{1}{s} \right) ds \end{aligned}$$

où β est défini en (2.4.16) et $\Psi_{\gamma,\rho}$ défini en (2.4.10). Ci-dessous, nous connaissons les deux expressions suivantes qui représentent $\tilde{\mu}_a$ et $\tilde{\mu}_b$ les termes de biais de $\sqrt{k}(\hat{\gamma}^{ML} - \gamma)$ et $\sqrt{k}\left(\frac{\hat{\sigma}^{ML}}{a(\frac{n}{k})} - 1\right)$ selon Drees et al [10], obtenir respectivement:

$$\begin{aligned}\tilde{\mu}_a &= \lambda \frac{(\gamma + 1)^2}{\gamma} \times \int_0^1 (s^\gamma - (2\gamma + 1) s^{2\gamma}) \Psi_{\gamma,\rho} \left(\frac{1}{s}\right) ds \\ \tilde{\mu}_b &= \lambda \frac{\gamma + 1}{\gamma} \times \int_0^1 ((\gamma + 1)(2\gamma + 1) s^{2\gamma} - s^\gamma) \Psi_{\gamma,\rho} \left(\frac{1}{s}\right) ds\end{aligned}\quad (2.4.18)$$

où $\lambda = \lim_{n \rightarrow +\infty} A\left(\frac{n}{k}\right)$. Et selon le théorème(2.2.16) de Haan and Stadtmüller [19] et la relation (2.2.20) pour $|A| \in RV_\rho$ avec $\lim_{t \rightarrow +\infty} A(t) = 0$. La fonction $\Psi_{\gamma,\rho}$ vérifie

$$\Psi_{\gamma,\rho}(x) := \begin{cases} \frac{x^{\gamma+\rho}-1}{\gamma+\rho} & \text{avec } \rho \neq \gamma \neq 0 \text{ où } \rho < 0 \\ \frac{-1}{\gamma} x^{-\gamma} \log x & \text{avec } \gamma \neq \rho = 0 \\ (\log x)^2 & \text{avec } \rho = \gamma = 0 \end{cases}\quad (2.4.19)$$

Conformément à Drees et al [10] avec $\Psi_{\gamma,\rho}(x)$ donnée dans la relation (2.4.19) on sait que:

$$\tilde{\mu}_a = \begin{cases} \frac{\rho(\gamma+1)}{(1-\rho)(\gamma-\rho+1)} & \text{si } \rho < 0 \\ 1 & \text{si } \gamma \neq \rho = 0 \\ 2 & \text{si } \gamma = \rho = 0 \end{cases} \quad \text{et} \quad \tilde{\mu}_b = \begin{cases} \frac{-\rho+(1-\rho)(1+\gamma)}{(1-\rho)(\gamma-\rho+1)} & \text{si } \rho < 0 \\ \frac{1}{\gamma} & \text{si } \gamma \neq \rho = 0 \\ 0 & \text{si } \gamma = \rho = 0 \end{cases}\quad (2.4.20)$$

En nous référant aux deux relations (2.4.17) et (2.4.18), nous trouvons

$$\mu_a^* = \left(\mathbf{1}_{\{\rho < 0\}} \frac{1}{\rho} + \mathbf{1}_{\{\rho = 0\}} \right) \tilde{\mu}_a \quad \text{et} \quad \mu_b^* = \lambda \beta - \left(\mathbf{1}_{\{\rho < 0\}} \frac{1}{\rho} + \mathbf{1}_{\{\rho = 0\}} \right) \tilde{\mu}_b\quad (2.4.21)$$

où $\beta = \left(-\frac{1}{\rho} \times \mathbf{1}_{\{\rho < 0\}} - \frac{1}{\gamma} \times \mathbf{1}_{\{\gamma \neq \rho = 0\}} + 0 \times \mathbf{1}_{\{\gamma = \rho = 0\}} \right)$. De plus, de Haan et Ferreira [18] ils vont encore plus loin en donnant une autre définition de la fonction $\Psi_{\gamma,\rho}$ qui vérifie la

relation (2.2.20). Et présenté comme suit:

$$\Psi_{\gamma,\rho}(x) := \begin{cases} \frac{x^{\gamma+\rho}-1}{\gamma+\rho} & \text{avec } \rho + \gamma \neq 0 \text{ où } \rho < 0 \\ \log x & \text{avec } \rho + \gamma = 0 \text{ où } \rho < 0 \\ \frac{1}{\gamma} x^\gamma \log x & \text{avec } \gamma \neq \rho = 0 \\ \frac{1}{2} (\log x)^2 & \text{avec } \rho = \gamma = 0 \end{cases} \quad (2.4.22)$$

utilisant la définition de $\Psi_{\gamma,\rho}$ donnée dans la relation (2.4.22) on trouve que le résultat suit :

$$\mu_a = \begin{cases} \frac{(\gamma+1)}{(1-\rho)(\gamma-\rho+1)} & \text{si } \rho < 0 \\ 1 & \text{si } \rho = 0 \end{cases} \quad \text{et } \mu_b = \begin{cases} \frac{-\rho}{(1-\rho)(\gamma-\rho+1)} & \text{si } \rho < 0 \\ 0 & \text{si } \rho = 0 \end{cases} \quad (2.4.23)$$

Pour calculer la variance de $\sqrt{k}(\hat{\gamma}^{ML} - \gamma)$, Nous posons la variable aléatoire suivant

$$Y(s) = \frac{(\gamma+1)^2}{\gamma} \int_0^1 (s^\gamma - (2\gamma+1)s^{2\gamma}) (s^{-\gamma-1}W_n(s) - W_n(1)) ds$$

Ensuite, nous montrons que

$$Var \left(\int_0^1 Y(s) ds \right) = \int_0^1 \int_0^1 E(Y(s)Y(t)) dt = (\gamma+1)^2.$$

Pour calculer la variance de $\sqrt{k} \left(\frac{\hat{\sigma}^{ML}}{a(\frac{n}{k})} - 1 \right)$, nous définissons la v.a suivant :

$$Z(s) = \frac{\gamma+1}{\gamma} \int_0^1 ((\gamma+1)(2\gamma+1)s^{2\gamma} - s^\gamma) (s^{-\gamma-1}W_n(s) - W_n(1)) ds$$

Puis, on trouve que

$$\text{Var} \left(\int_0^1 Z(s) ds \right) = \int_0^1 \int_0^1 E(Z(s)Z(t)) dt = 1 + (\gamma + 1)^2.$$

De même, nous trouvons la covariance asymptotique de $\sqrt{k}(\hat{\gamma}^{ML} - \gamma)$ et $\sqrt{k} \left(\frac{\hat{\sigma}^{ML}}{a(\frac{n}{k})} - 1 \right)$ est donné par

$$\text{Cov} \left(\int_0^1 Y(s) ds, \int_0^1 Z(s) ds \right) = \int_0^1 \int_0^1 E(Y(s)Z(t)) dt = -1 - \gamma.$$

■

Généralement, l'estimation est préférée par le maximum de vraisemblance car elle a un estimateur consistant avec un biais plus faibles. L'estimateur du maximum de vraisemblance est un estimateur de variance consistant et il est utilisé pour remplacer la variance asymptotique des paramètres inconnus.

2.4.2 Estimateur de Hill ($\gamma > 0$)

Cet estimateur a été défini par Hill [20] pour estimer d'une manière semi-paramétrique le paramètre γ , l'indice de queue pour $\gamma > 0$. Pour construire cet estimateur, Hill utilise la méthode du maximum de vraisemblance sur l'ensemble de k plus grandes observations d'un échantillon comme suit:

Soit X_1, X_2, \dots, X_n une suite de v.a.'s i.i.d suivant la même distribution F . Supposons que $F \in D(\Phi_{-1/\gamma})$ appartient au domaine d'attraction de Fréchet c'est-à-dire F vérifie la condition de Von Mises (2.2.18) et (2.2.20) où $\gamma > 0$, avec \bar{F} la fonction de survie vérifiant pour tout $x \geq t$ on a :

$$\forall x \geq 0, \lim_{t \rightarrow +\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} := x^{-1/\gamma} \text{ où } t \in \mathbb{R}.$$

grâce au théorème (2.4) $\bar{F} \in RV_{-1/\gamma}$ avec $\gamma > 0$. Ce qui est équivalente à: pour $\gamma > 0$ et $F(x) < 1$ que :

$$\forall x \geq 0, \lim_{t \rightarrow +\infty} \frac{1}{\bar{F}(t)} \int_t^{\infty} \frac{\bar{F}(x)}{x} dx := \gamma \text{ où } t \in \mathbb{R}.$$

Par intégration par partie on trouve que :

$$\forall x \geq 0, \lim_{t \rightarrow +\infty} \frac{1}{\bar{F}(t)} \int_t^{\infty} (\log(x) - \log(t)) dF(x) := \gamma \text{ où } t \in \mathbb{R}.$$

Considérons la statistique d'ordre $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ de l'échantillon X_1, X_2, \dots, X_n . On génère $X_{n-k,n}, X_{n-k+1,n}, \dots, X_{n,n}$ un échantillon des excès de X au-dessus du seuil $X_{n-k,n}$. Afin de développer un estimateur de Hill, on remplace t par $X_{n-k,n}$ et F par la distribution empirique F_n , on trouve

$$\hat{\gamma}_{k,n}^H := \frac{1}{\bar{F}_n(X_{n-k,n})} \int_{X_{n-k,n}}^{\infty} (\log(x) - \log(X_{n-k,n})) dF_n(x)$$

Définition 2.9 Soit X_1, X_2, \dots, X_n une suite de v.a's i.i.d suivant la même loi F avec $F \in D(\Phi_{-1/\gamma})$. Pour $1 < k < n$ où $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$ et $F_n(x)$ définie en (1.1.3) et $\bar{F}_n(X_{n-k,n}) := \frac{n}{k}$ donc on trouve que l'estimateur de Hill est défini par

$$\hat{\gamma}_{k,n}^H := \frac{1}{k} \sum_{i=1}^k \log(X_{n-i+1,n}) - \log(X_{n-k,n}) \quad (2.4.24)$$

De nombreux travaux théoriques ont été consacrés à l'étude des propriétés de l'estimateur de Hill. Le théorème suivant définit la consistance de l'estimateur de Hill. Mason [31] a démontré la consistance faible et Deheuvels et al [8] ont établi la consistance forte.

Théorème 2.13 Soit X_1, X_2, \dots, X_n une suite de v.a's i.i.d suivant la même loi F . Supposons $k_n = k$ une suite d'entiers où $1 < k < n$ avec $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$. Alors

1. $\widehat{\gamma}_{k,n}^H$ converge en probabilité vers γ (consistance faible)
2. Si $k/\log(\log(n)) \rightarrow \infty$ quand $n \rightarrow \infty$ alors $\widehat{\gamma}_{k,n}^H$ converge presque sûrement vers γ (consistance forte)

La normalité asymptotique de l'estimateur de Hill (1975) est due entre autres à Davis et Resnick [6] et à L. de Haan et Ferreira, [18] (voir théorème 3.2.5 page 74).

Théorème 2.14 Soit X_1, X_2, \dots, X_n une suite de v.a.'s i.i.d suivant la même loi F . Supposons que F satisfait la condition du deuxième ordre du théorème (2.2.13) pour $\gamma > 0$. Et pour $k_n = k$ une suite d'entiers où $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$ avec $1 < k < n$ et une fonction auxiliaire $|A| \in RV_\rho$ avec $A\left(\frac{n}{k}\right) \rightarrow \lambda$ quand $n \rightarrow \infty$. Alors

$$\sqrt{k} (\widehat{\gamma}_{k,n}^H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right) \quad (2.4.25)$$

Il est clair que si $\lambda = 0$ l'estimateur de Hill devient sans biais. L'estimateur de Hill n'est valable que pour les distributions qui n'appartiennent pas au domaine d'attraction de Fréchet où quand l'indice de queue est négatif. Il est à noter que l'estimateur de Hill nécessite toujours des hypothèses d'indépendance. Pour ce cas Beirlant et al [4] (2005) ont donné une généralisation de l'estimateur de Hill. Il est valide pour tout domaine d'attraction. Mais dans notre travail nous présentons seulement le cas où la distribution appartient au domaine d'attraction de Fréchet dans la sous-section suivante:

2.4.3 Estimateur de Hill Généralisé ($\gamma > 0$)

Soit la fonction quantile de queue de F donnée en (2.2.9), avec $\overline{F} \in RV_{-1/\gamma}$. Il a été montré par de Haan [16] (1970) que là où $\gamma > 0$ équivalent à $U(x) = x^\gamma L(x)$ où $L \in RV_0$ (2.2.3) et stipule que $U \in RV_\gamma$. Pour ce modèle à variation régulière, Hill [20] (1975) propose un estimateur de l'indice de queue γ , noté $\widehat{\gamma}_{k,n}^H$, d'une manière semi-paramétrique où $\gamma > 0$. Et ce que nous savons dans la partie (2.4.2). Et dans l'introduction de partie (2.4.2),

nous tirons la résultat suivant quand $t \rightarrow +\infty$:

$$\forall x \geq 0, \frac{1}{\bar{F}(t)} E(\log(X) - \log(t) / X > t) \rightarrow \gamma \text{ où } t \in \mathbb{R}.$$

Par la loi forte des grands nombres on a $\frac{1}{n} \sum_{i=1}^n 1_{X_i > t} \xrightarrow{p.s} \bar{F}(t)$, donc $\frac{1}{n} \sum_{i=1}^n (\log(X_i) - \log(t)) 1_{X_i > t} \xrightarrow{p.s} E(\log(X) - \log(t) / X > t)$. En remplaçant à nouveau t par $X_{n-k+1,n}$ on trouve

$$\frac{1}{\bar{F}_n(t)} E(\log(X_i) - \log(t) / X > t) = \hat{\gamma}_{k,n}^H,$$

qui définit un estimateur empirique pour la moyenne de la fonction de durée de vie résiduelle. En supposant $U(\infty) > 0$. La relation asymptotique suivante, équivalente à (2.1.1) coïncide avec la relation (2.2.5) dans le théorème (2.5) à condition qu'il existe une fonction positive a_0 telle que pour $t > 0$,

$$\forall x > 0, \lim_{t \rightarrow +\infty} \frac{U(tx) - U(t)}{a_0(t)} := \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{si } \gamma \neq 0 \\ \log(x) & \text{si } \gamma = 0 \end{cases}.$$

et donc :

$$\forall x > 0, \lim_{t \rightarrow +\infty} \frac{\log U(tx) - \log U(t)}{a_0(t) / U(t)} := \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{si } \gamma \neq 0 \\ \log(x) & \text{si } \gamma = 0 \end{cases}.$$

En se basant sur cette limite Beirlant et al [4] (2005), ont proposés un estimateur de γ plus généralisé pour tout $\gamma \in \mathbb{R}$, noté $\hat{\gamma}_{k,n}^{GH}$ basé sur la pente graphiques des quantiles généralisés(Generalized Quantile Plots). Ils ont indiqués que γ peut être estimé de manière cohérente en utilisant la relation $U(x) = x^\gamma L(x)$. On inspectant le nuage de points avec les coordonnées $\left(\log\left(\frac{n+1}{i}\right); \log U\tilde{F}\left(\frac{n+1}{i}\right) \right)$ pour tous $i = 1, \dots, n$. Ils constatent pour $U\tilde{F}\left(\frac{n+1}{i}\right)$ où $U\tilde{F}(t) = U(t)\tilde{F}(t) = t^\gamma L(t)$ avec $L \in RV_0$ donc, $U\tilde{F}(t) \in RV_\gamma$, implique $\log\left(U\tilde{F}(t)\right) \sim \gamma \log t$ quand $t \rightarrow \infty$ où

$$\tilde{F}(t) = E \circ \log U = E(\log X - \log U(t) / X > U(t))$$

Et comme substitut empirique de $U\tilde{F}_{i,n} := U\tilde{F}\left(\frac{n+1}{i}\right)$ pour $i = 1, \dots, n$,

$$U\tilde{F}_{i,n} := X_{n-i,n} \left(\frac{1}{i} \sum_{j=1}^i \log(X_{n-j+1,n}) - \log(X_{n-i,n}) \right) = X_{n-i,n} \hat{\gamma}_{i,n}^H \quad (2.4.26)$$

Puisque $\gamma > 0$, si $i = k$ on trouve que $U\tilde{F}_{k,n} = X_{n-k,n} \hat{\gamma}^H$, de même, si on remplace $X_{n-i+1,n}$ par $U\tilde{F}_{i,n}$ et $X_{n-k,n}$ par $U\tilde{F}_{k+1,n}$ dans la relation (2.4.24) nous trouvons :

$$\hat{\gamma}_{k,n}^{GH} := \frac{1}{k} \sum_{i=1}^k \log(U\tilde{F}_{i,n}) - \log U\tilde{F}_{k+1,n} = \hat{\gamma}_{k,n}^H + \frac{1}{k} \sum_{i=1}^k \log \hat{\gamma}_{i,n}^H - \frac{1}{k} \log \hat{\gamma}_{k,n}^H \quad (2.4.27)$$

qui est l'estimateur de Hill généralisé présenté par Beirlant et al [4] (2005).

Sous les hypothèses du théorème (2.8) de variation régulière généralisée du second ordre décrit par de Haan and Stadtmüller [19] (1996), supposons qu'il existe une fonction positive a_0 et une seconde fonction éventuellement positive $A_0(t)$ avec $A_0(t) \rightarrow 0$ quand $t \rightarrow \infty$ telle que la limite

$$\forall x > 0, \lim_{t \rightarrow +\infty} \frac{1}{A_0(t)} \left(\frac{U(tx) - U(t)}{a_0(t)} - h_\gamma(x) \right) = \psi(x) \quad (2.4.28)$$

où la fonction $h_\gamma(x) = \int_1^x t^{\gamma-1} dt$. Il s'ensuit qu'il existe une constante $c \in \mathbb{R}$ et un paramètre de second ordre $\rho \leq 0$ pour les quels la fonction a_0 satisfait,

$$\forall x > 0, \lim_{t \rightarrow +\infty} \frac{1}{A_0(t)} \left(\frac{a_0(tx)}{a_0(t)} - x^\gamma \right) = cx^\gamma h_\rho(x) \quad (2.4.29)$$

tel que $|A_0| \in RV_\rho$. La fonction ψ qui apparaît en (2.4.28) pour $\tilde{\alpha} \in \mathbb{R}$ admet la représentation suivante:

$$\psi(x) = cx^\gamma h_\rho(x) + \tilde{\alpha} h_{\gamma+\rho}(x)$$

On se limite au cas où $\rho < 0$, lorsque $c = 0$ et pour un choix approprié de la fonction a_0 conduit à une simplification de la fonction ψ qui apparaît dans (2.4.28) peut être réécrit

pour $\tilde{\alpha} \in \mathbb{R}^*$ par :

$$\psi(x) = \tilde{\alpha} h_{\gamma+\rho}(x) \quad (2.4.30)$$

Nous utilisons la notation suivante $U \in GRV_2(\gamma; \rho; a_0(t); A_0(t); \tilde{\alpha})$, qui symbolise U est à variation régulière du deuxième ordre généralisée (satisfaisante (2.4.28)-(2.4.30) pour $c = 0$). Cependant, que l'estimateur de Hill n'est utile que lorsque $\gamma > 0$.

À partir de Einmahl et al [11] (2008), on obtient les représentations suivantes de $U_H \in GRV_2(\gamma; \rho; a_0(t); A_0(t); \tilde{\alpha})$ avec les constantes $\tilde{\alpha} \in \mathbb{R}^*$, $\tilde{\beta} > 0$ et $\beta = \frac{1}{\tilde{\beta}} \lim_{t \rightarrow \infty} (U(t) - a_0(t) / \gamma)$ avec $\beta \in \mathbb{R}$ et pour $\gamma > 0$ et $-\rho > 0$:

$$\frac{U(t)}{\tilde{\beta}} = t^\gamma \left[\frac{1}{\gamma} + \left(\frac{\tilde{\alpha}}{\gamma + \rho} A_0(t) (1 + o(1)) \right) \right] \mathbf{1}_{\{-\rho \neq \gamma\}} + \beta \mathbf{1}_{\{-\rho > \gamma\}} + \ell(t) \mathbf{1}_{\{-\rho = \gamma\}} \quad (2.4.31)$$

où $\ell(t) = \alpha + \int_1^x (\tilde{\alpha} + o(1)) \frac{\tilde{\ell}(t)}{t} dt + o(\tilde{\ell}(t))$ pour une constante $\alpha \in \mathbb{R}$ et $\tilde{\ell} \in RV_0$ (2.2.3).

Grâce à Beirlant et al [4] (2005) nous utilisons la notation suivante, similaire à celle utilisée dans Einmahl et al [11] (2008) comme,

$$t^\gamma b(t) = \frac{\tilde{\alpha} \rho (\rho + \gamma (1 - \rho))}{(\gamma + \rho) (1 - \rho)} A_0(t) t^\gamma \mathbf{1}_{\{-\rho \neq \gamma, \beta = 0\}} - \frac{\gamma^3}{1 + \gamma} (\beta \mathbf{1}_{\{-\rho > \gamma, \beta \neq 0\}} + \ell(t) \mathbf{1}_{\{-\rho = \gamma\}}) \quad (2.4.32)$$

On suppose ici que $k = k_n$ est une séquence intermédiaire, c'est-à-dire, pour $1 < k < n$ où $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$.

Pour déterminer la variance asymptotique et le biais d'estimateur $\hat{\gamma}_{k,n}^{GH}$ Beirlant et al [4] (2005) utilisant les représentations asymptotiques suivantes pour $\log U \tilde{F}_{i,n}$ avec $i = 1, \dots, k$ qu'il définit en (2.4.26) par

$$\log U \tilde{F}_{i,n} = \gamma \log U_{k+1,n}^{-1} + \log \tilde{\beta} \left(\frac{k}{i} \right)^\gamma + \left\{ \sqrt{k} \frac{T_{k,n} \left(\frac{i}{k} \right)}{i} + \tilde{\rho}^{-1} b \left(\frac{n}{k} \right) \left(\frac{i}{k} \right)^{|\tilde{\rho}|} \right\} (1 + o(1))$$

pour $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$, est la statistique d'ordre de l'échantillon U_i de lois Uniform pour $i = 1, \dots, n$. Où b est défini en (2.4.32) avec $\tilde{\rho} = -\gamma \mathbf{1}_{\{\gamma + \rho > 0; \beta \neq 0\}} + \rho \mathbf{1}_{\{\gamma + \rho > 0; \beta = 0\}}$.

Et $\{T_{k,n}(t), t \in [0; 1]\}$ est le processus stochastique représenté asymptotiquement par le processus gaussien suivant :

$$T_{k,n}(t) = W^0(t) - \gamma t W(1) := \int_0^t \frac{W(s)}{s} ds + (\gamma - 1) W(t) - \gamma t W(1)$$

avec

$$Cov(W^0(s); W^0(t)) := s(1 + \gamma^2 + \gamma \log(t/s)) \text{ pour } 0 \leq s \leq t \leq 1$$

Théorème 2.15 *Supposant que $\sqrt{kb} \left(\frac{n}{k}\right) \rightarrow \tilde{\lambda} \in \mathbb{R}$. Alors pour $\gamma \geq 0$*

$$\sqrt{k} (\hat{\gamma}_{k,n}^{GH} - \gamma) \rightarrow N \left(\frac{\tilde{\lambda}}{1 - \tilde{\rho}}, 1 + \gamma^2 \right)$$

$$\text{où } \tilde{\rho} = -\gamma \mathbf{1}_{\{\gamma + \rho > 0; \beta \neq 0\}} + \rho \mathbf{1}_{\{\gamma + \rho > 0; \beta = 0\}}$$

Il est clair que si $\tilde{\lambda} = 0$ l'estimateur de Hill généralisé deviendra sans biais. L'estimateur de Hill généralisé où l'indice de queue est positif, est valable pour la distribution qui appartient au domaine d'attraction de Fréchet. De plus, $\hat{\gamma}_{k,n}^{GH}$ est valable avec les valeurs positives mais pour les observations négatives, il faut déplacer les observations vers des valeurs positives. Mais cette opération a son influence sur les estimations. Et prendre des logarithmes souvent introduit un biais supplémentaire.

Mais Beirlant et al [4] (2005) ont donnés un biais réduit pour de $\hat{\gamma}_{k,n}^{GH}$. Pour cela en chapitre suivant on étudie l'estimation de l'indice des valeurs extrêmes lorsque les données sont soumises à une censure aléatoire. Et nous essayons d'adapter les deux estimateurs

$\hat{\gamma}_{k,n}^{GH}$ et $\hat{\gamma}^{ML}$ pour les données censurées

Chapitre 3

Statistiques d'Indice de queue sous données censurés

3.1 Données censurés

Intuitivement, quand un ensemble de données contient des observations dans une marge restreinte des valeurs, mais autrement non mesuré, s'appelle un ensemble de données censuré. Ces données censurées se trouvent fréquemment dans l'analyse de survie qui étudie fondamentalement le temps qui s'écoule entre le début d'une observation et le moment d'un événement. Donc, soit la variable d'intérêt X et la variable de censure Y définie par la non-observation de l'événement étudié. Si l'on observe Y , et non X , on trouve les trois cas suivant $X > Y$, $X < Y$ et $Y_1 < X < Y_2$ respectivement. On dit qu'il y a censure à droite, censure à gauche et censure par intervalle respectivement. Les modèles de censures se rencontrent dans de nombreuses situations : bien connus et fréquemment utilisés en biostatistique, ils trouvent également leurs applications en assurance ou même en finance. La censure à droite ou censure de type I, est l'exemple le plus fréquent d'observation incomplète en analyse de survie, et a largement été décrit dans la littérature (Anderson et al (1993) [1]). En mathématique, la censure à droite signifie que l'on observe seulement

la variable $Z = X \wedge Y$ et on observe également l'indicateur de censure, $\delta_{X \leq Y}$ qui nous informe si notre variable observée correspond bien à la variable d'intérêt ou à la variable de censure.

3.1.1 Définitions

Définition 3.1 *Dans cette définition, nous avons rassemblé quelques définitions et concepts fréquemment utilisés dans l'analyse de survie que nous présentons comme suit:*

Date d'origine: Elle correspond à l'origine de la durée étudiée. Comme la date de naissance.

Date des dernières nouvelles: C'est la date la plus récente où des informations sur un sujet ont été recueillies.

Date de point: C'est la date au-delà de laquelle on arrêtera l'étude et on ne tiendra plus compte des informations sur les sujets.

Durée de survie: Elle correspond au temps de survie lorsque les décès sont observés avant la date de point, qui est donc la durée entre la date d'origine et la survenue de l'évènement d'intérêt, c'est-à-dire du décès.

Donc, pour obtenir une série statistique liée aux données censurées, trois séries statistiques doivent être disponibles comme suit:

- Le temps de survie: un échantillon X_i de v.a's i.i.d suivant la même loi F .
- Le temps de censure: un échantillon Y_i de v.a's i.i.d suivant la même loi G .
- La durée réellement observée: un échantillon Z_i de v.a's i.i.d suivant la même loi H .

Noté que le temps de survie X c'est la variable d'intérêt parceque: Si l'évènement se produit, X est réalisée, par contre si X (l'individu étant perdu de vue) n'est pas réalisée, l'évènement ne se produit pas c'est Z qui est réalisée. Alors on trouve les types suivant:

Censure à droite(Right censoring): X est dite censurée à droite si X_i (l'individu) concerné n'a aucune information sur sa dernière observation. Ainsi, en présence de censure à droite les variables d'intérêt ne sont pas toutes observées.

Censure à gauche (Left censoring): X est dite censure à gauche lorsque X_i a déjà subi l'événement avant qu'il soit observé. On sait uniquement que la variable d'intérêt est inférieure ou égale à une variable connue

Censure double ou mixte: On dit qu'on a une censure double ou mixte si on a des données censurées à droite et des données censurées à gauche dans le même échantillon.

Censure par intervalle (Interval censoring): Dans ce cas, comme son nom l'indique, on observe à la fois une borne inférieure et une borne supérieure de la variable d'intérêt

Définition 3.2 (*censure aléatoire de type I*) Soit (Z_i, δ_i) un couple de v a's i.i.d suivant la même loi H pour $i = 1, \dots, n$. On dit que Z_i est une censure à droite s'il existe un échantillon X_i de v a's i.i.d suivant la même loi F et un deuxième échantillon Y_i de v a's i.i.d suivant la même loi G pour tout $i = 1, \dots, n$ telles

$$Z_i = \min(X_i, Y_i) \text{ et } \delta_i := \begin{cases} 1 & \text{si } X_i \leq Y_i \\ 0 & \text{si } X_i > Y_i \end{cases} \quad (3.1.1)$$

avec X_i et Y_i indépendantes

Cela signifie qu'on observe le véritable temps de survie X s'il est inférieur au temps de censure Y , dans ce cas la donnée n'est pas censurée si $\delta_i = 1$, par contre si $\delta_i = 0$ la donnée est dite censurée à droite. C'est-à-dire, au lieu d'observer X on observe une valeur Y avec pour seule information le fait que X soit supérieur à Y .

Définition 3.3 (*Censure de type I : (fixée)*) Considérons un échantillon X_i de v a's i.i.d suivant la même loi F . et soit Y une valeur fixée. On dit qu'un couple (Z_i, δ_i) de v a's i.i.d suivant la même loi H est une censure de type I s'il est donné pour tout $i = 1, \dots, n$ par

$$Z_i = \min(X_i, Y) \text{ et } \delta_i := \begin{cases} 1 & \text{si } X_i \leq Y \\ 0 & \text{si } X_i > Y \end{cases} \quad (3.1.2)$$

Définition 3.4 (*Censure à gauche*) Soit un échantillon X_i de v.a's i.i.d suivant la même loi F et un deuxième échantillon Y_i de v.a's i.i.d suivant la même loi G . On dit que le couple (Z_i, δ_i) de v a's i.i.d suivant la même loi H est censure à gauche si

$$Z_i = \sup(X_i, Y_i) \text{ et } \delta_i := \begin{cases} 1 & \text{si } X_i > Y_i \\ 0 & \text{si } X_i \leq Y_i \end{cases} \quad (3.1.3)$$

pour tout $i = 1, \dots, n$ avec X_i et Y_i sont indépendantes

C'est-à-dire que, pour $\delta_i = 1$ la donnée est observée et pour $\delta_i = 0$ la donnée est dite censurée à gauche. C'est-à-dire, au lieu d'observer X on observe Y avec pour seule information le fait que X soit inférieur à Y .

Proposition 3.1 Les deux type de censure peuvent être combinés. On dispose alors de deux temps de censure Y_1 et Y_2 , l'un à gauche et l'autre à droite, avec $Y_1 < Y_2$. Donc on observe le triplet $(Z_i, (\delta_1)_i, (\delta_2)_i)$ avec

$$\begin{cases} Z_i = Y_1 & \text{si} & X_i \leq Y_1 \\ Z_i = X_i & \text{si} & Y_1 \leq X_i \leq Y_2 \\ Z_i = Y_2 & \text{si} & X_i \geq Y_2 \end{cases}$$

Définition 3.5 (*Censure double et mixte*) Si on a des données censurées à droite et des données censurées à gauche dans le même échantillon. On dit qu'on a une censure double ou mixte.

Exemple 3.1 On cherche à étudier l'apprentissage de la marche chez les jeunes enfants. Pour cela, on décide de suivre tous les enfants d'une ville moyenne entre 10 à 16 mois. La durée d'intérêt X est donc ici l'âge d'acquisition de la marche (ou l'intervalle de temps entre la naissance et les premiers pas). Sur l'échantillon examiné Y , trois cas se produisent : soit l'enfant sait déjà marcher avant 10 mois (censure à gauche), soit il apprend à

marcher entre 10 et 16 mois (données complètes), soit il ne sait toujours pas marcher à 16 mois (censure à droite) ou qu'il est perdu de vue entre 10 et 16 mois.

La figure (3.1) suivante montre ce que nous avons abordé précédemment à propos des données censurées

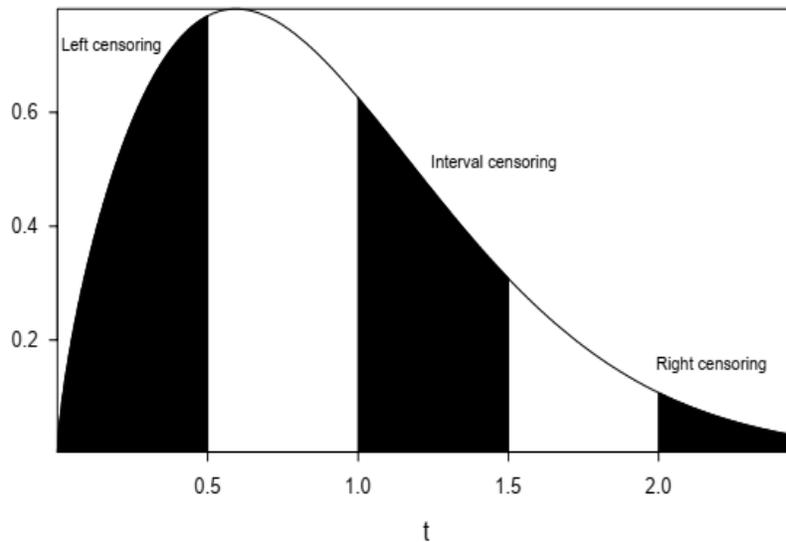


Figure 3.1: Présentant un exemple pratique de données censure à gauche et censure à droite suivant.

Illustration de la probabilité pour les observations censurées.

3.1.2 Estimateur de Kaplan-Meier

Comme nous l'avons mentionné plus tôt avec les données censurées à droite, la variable d'intérêt X n'est plus la variable observée, mais la durée sous-jacente. Plus précisément, soit la v.a d'intérêt X avec \bar{F} sa fonction de survie qui définit en relation (1.2.8) et la v.a de temps censure Y suit une distribution G et \bar{G} sa fonction de survie. Considérons la v.a $Z = X \wedge Y$ le temps observée, avec l'indicateur de censure, $\delta_{X \leq Y}$ qui définit une v.a de censure à droite et suit une distribution H et \bar{H} sa fonction de survie. On suppose que la variable d'intérêt X et la variable censurée Y sont indépendantes, donc on a $\bar{H} = \bar{F}\bar{G}$.

Actuellement, dans cette partie on va estimer la fonction de survie \bar{F} (1.2.8) avec des données censurées à droite. C'est-à-dire, on va définir \bar{F}_n la survie empirique des données observées X_i pour tout $i = 1, \dots, n$ ou même par les données non censurées $Z_i = X_i$ avec $\delta_i = 1$ où Z_i et δ_i sont définis en (3.1.1). Cette fonction peut être estimée grâce à plusieurs méthodes non paramétriques. L'estimateur le plus intéressant de la fonction de survie (1.2.8) sous censure est l'estimateur de Kaplan-Meier [24] (1958) appelé aussi un estimateur de produit (P-L estimateur) car il s'obtient comme un produit. \bar{F}_n^{KM} l'estimateur de Kaplan-Meier est obtenue de la façon suivante:

En effet, la fonction de survie est estimée sous les données censurées par

$$\bar{F}_n(x) := \frac{\sum_{i=1}^n \mathbf{1}_{\{X_i > x, \delta_i = 1\}}}{\sum_{i=1}^n \mathbf{1}_{\{\delta_i = 1\}}}, \quad (3.1.4)$$

On observe un l'échantillon (décès et censure) à temps (les temps d'événements) distincts $Z_{0,k}, Z_{1,k}, \dots, Z_{k,k}$ avec $Z_{0,k} = 0$ sont les durées de vie distinctes parmi $Z_{i,n}$ pour ($i = 1, \dots, n$) ordonnés par ordre croissant. Pour chaque temps $Z_{i,k}$ ($i = 1, \dots, k$) où $1 \leq k \leq n$ on connaît

r_i : le nombre de décès ($\delta_i = 1$) dans l'intervalle $I_{i,k} =]Z_{i-1,k}, Z_{i,k}]$.

m_i : le nombre de censures ($\delta_i = 0$) dans $I_{i,k}$.

n_i : le nombre de sujet présents dans l'échantillon à temps $Z_{i-1,k}$ (ni décédés, ni censurés jusque là, i.e. pour lesquels $X > Z_{i-1,k}$).

Supposons que $m_0 = r_0 = 0$ et $n_1 = n$. Alors $n_i := n - \sum_{j=1}^i (r_{j-1} + m_{j-1})$ dans l'intervalle

$I_{i,k}$. De plus, on trouve que $n = \sum_{j=1}^i (r_{j-1} + m_{j-1})$. On notons q_i la probabilité de décéder ($\delta_i = 1$) dans l'intervalle $I_{i,k}$ sachant qu'on est vivant en $Z_{i-1,k}$. Donc la probabilité empirique de chaque q_i est donné par

$$\hat{q}_i := P(X \leq Z_{i,k} | X > Z_{i-1,k}) := \frac{r_i}{n_i - m_i}, \quad (3.1.5)$$

Par conséquent, la probabilité de rester vivant dans l'intervalle $I_{i,k}$ sachant qu'on ait vivant en $Z_{i-1,k}$ est donnée par ,

$$P(X > Z_{i,k} | X > Z_{i-1,k}) := 1 - \hat{q}_i := 1 - \frac{r_i}{n_i - m_i} \quad (3.1.6)$$

Et la probabilité empirique de survivre après un temps $Z_{i,n}$ c'est être en vie juste avant $Z_{i,n}$ et ne pas mourir au temps $Z_{i,n}$, s'obtient par:

$$\bar{F}_n(Z_{i,n}) := P(X > Z_{i,n}) := \prod_{k=1}^i P(X > Z_{k,n} | X > Z_{k-1,n}) := \prod_{k=1}^i (1 - \hat{q}_i) \quad (3.1.7)$$

Pour $\sup |Z_{i-1,k} - Z_{i,k}| \rightarrow 0$ quand $k \rightarrow \infty$ (avec $n \rightarrow \infty$). On peut dire que dans l'intervalle $]Z_{i-1,k}, Z_{i,k}]$ avec $|Z_{i-1,k} - Z_{i,k}| \simeq 0$ on peut remarquer que $m_i = 0$ et r_i avec n_i existent. Alors l'estimateur de Kaplan-Meier est donné par

$$\bar{F}_n^{KM}(x) \rightarrow \prod_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \left(1 - \frac{r_i}{n_i} \right) \quad (3.1.8)$$

La formule (3.1.8) de Kaplan-Meier implique

$$\bar{F}_n^{KM}(x) = \exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log \left(1 - \frac{r_i}{n_i} \right) \right] \quad (3.1.9)$$

On peut approximer $\bar{F}_n^{KM}(x)$ pour une valeur c fixée et une fonction $\varphi(x) = e^{\log(1-x)}$ avec $\varphi'(x) = \frac{-1}{1-x} e^{\log(1-x)}$ par:

$$\bar{F}_n^{KM}(x) \approx \exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log(1-c) \right] - \sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \frac{1}{1-c} \exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log(1-c) \right] \left(\frac{r_i}{n_i} - c \right)$$

Implique que

$$\left\{ \begin{array}{l} E \left[\overline{F}_n^{KM}(x) \right] \approx \exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log(1-c) \right] - \sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \left(\frac{1}{1-c} \exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log(1-c) \right] \right) \left(E \left(\frac{r_i}{n_i} \right) - c \right) \\ Var \left(\overline{F}_n^{KM}(x) \right) \approx \left(\exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log(1-c) \right] \right)^2 \sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \left(\frac{1}{1-c} \right)^2 Var \left(\frac{r_i}{n_i} \right) \end{array} \right. \quad (3.1.10)$$

Avec $r_i := \sum \delta_{\{X_i \leq Y_i, \delta_i=1\}}$ sur n_i . Alors r_i suit la loi binomial avec $E(r_i) := n_i p_i$ et $var(r_i) := n_i p_i (1 - p_i)$. Sous la relation (3.1.10) avec $c := p_i$ en trouve:

$$\left\{ \begin{array}{l} E \left[\overline{F}_n^{KM}(x) \right] \approx \exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log(1-p_i) \right] \\ Var \left(\overline{F}_n^{KM}(x) \right) \approx \left(\exp \left[\sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \log(1-p_i) \right] \right)^2 \sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \frac{p_i}{n_i(1-p_i)} \end{array} \right.$$

en définissant $p_i := \hat{p}_i := (r_i/n_i)$. On trouve que

$$\left\{ \begin{array}{l} \hat{E} \left[\overline{F}_n^{KM}(x) \right] \approx \overline{F}_n^{KM}(x) \geq 0 \\ \widehat{Var} \left(\overline{F}_n^{KM}(x) \right) \approx \left(\overline{F}_n^{KM}(x) \right)^2 \sum_{\substack{k=1, \dots, i \\ Z_{i,n} \leq x}} \frac{r_i}{n_i(n_i - r_i)} \end{array} \right. \quad (3.1.11)$$

Ces estimateurs de $E \left[\overline{F}_n^{KM}(x) \right]$ et $Var \left(\overline{F}_n^{KM}(x) \right)$ sont donnés par Greenwood (1926).

On a l'indentité suivante :

$$\left(1 - \frac{r_i}{n_i} \right) = \left(1 - \frac{1}{n_i} \right) \left(1 - \frac{1}{n_i - 1} \right) \cdots \left(1 - \frac{1}{n_i - r_i + 1} \right)$$

Donc on peut récrit la fonction de répartition, de la survie empirique en relation (3.1.8)

comme

$$\overline{F}_n^{KM}(x) := \prod_{i=1}^n \left(1 - \frac{1}{n_i - r_i + 1} \right)^{(\delta_i=1)\mathbf{1}\{Z_{i,n} \leq x\}} \quad (3.1.12)$$

Se implique que on prend l'intervalle $]Z_{i-1,n}, Z_{i,n}]$ où $Z_{i,n}$ sont les temps d'événements (décès et censure) ordonnés par ordre croissant avec $Z_{0,n} = 0$ alors pour tout $i = 1, \dots, n$.

De plus, si on désigne par $\delta_{[i,n]}$ le concomitant de $Z_{i,n}$ ou les indicateurs de censure (c'est-à-dire, $\delta_{[i,n]} = \delta_j$ si $Z_{i,n} = Z_j$ avec $j = 1, \dots, n$). On obtient alors l'estimateur de Kaplan-Meier pour tout $Z_{i,n} \leq x$ par:

$$\begin{aligned} \overline{F}_n^{KM}(x) &:= \prod_{i=1}^n \left(\frac{n-i}{n-i+1} \right)^{\delta_{[i,n]}\mathbf{1}\{Z_{i,n} \leq x\}} = \prod_{i=1}^n \left(1 - \frac{\delta_{[i,n]}}{n-i+1} \right)^{\mathbf{1}\{Z_{i,n} \leq x\}} \\ &:= \prod_{i=1}^n \left(1 - \frac{\delta_{[i,n]}\mathbf{1}\{Z_{i,n} \leq x\}}{n-i+1} \right) \end{aligned} \quad (3.1.13)$$

où $\delta_{[i,n]} = 0$ en cas de censure en $Z_{i,n}$ et $\delta_{[i,n]} = 1$ en cas de décès en $Z_{i,n}$.

De plus, soit (Z_i, δ_i) un échantillon de censure aléatoire (c.a) de type I réellement observé comme définit en (3.2) et soit $(Z_{(i,n)}, \delta_{(i,n)})$ les statistiques d'ordre associées pour tout $i = 1, \dots, n$. L'estimateur Kaplan-Meier de la fonction de répartition empirique en présence de censure \widehat{F}_n^{KM} est défini par :

$$\widehat{F}_n^{KM}(x) := 1 - \overline{F}_n^{KM}(x) := 1 - \prod_{i=1}^n \left(1 - \frac{\delta_{[i,n]}}{n-i+1} \right)^{\mathbf{1}\{Z_{i,n} \leq x\}} := \sum_{i=1}^n D_{i,n} \mathbf{1}_{Z_{i,n} \leq x} \quad (3.1.14)$$

où les poids $D_{i,n}$ pour la i -valeur censurée sont donnés par :

$$D_{i,n} := \frac{\delta_{[i,n]}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[i,n]}} \quad (3.1.15)$$

avec $\widehat{F}_n^{KM}\{x_i\} = \widehat{F}_n^{KM}(x_i) - \widehat{F}_n^{KM}(x_{i-})$. Nous avons l'indentité suivante :

$$\frac{n-i}{n} \prod_{j=1}^i \left(\frac{n-j+1}{n-j} \right) := 1$$

Alors on trouve que les poids $D_{i,n}$ (3.1.15) pour la i -valeur censurée sont réécrit par

$$D_{i,n} := \frac{\delta_{[i,n]}}{n} \prod_{j=1}^{i-1} \left(1 + \frac{1}{n-j}\right)^{1-\delta_{[i,n]}} \quad (3.1.16)$$

Ce qui implique qu'on peut réécrire l'estimation de la fonction de répartition en présence de censure par

$$\widehat{F}_n^{KM}(x) := \sum_{i=1}^n D_{i,n} \mathbf{1}_{Z_{i:n} \leq x} = \frac{1}{n} \sum_{i=1}^n W_{i,n} \mathbf{1}_{Z_{i:n} \leq x} := \frac{1}{n} \sum_{i=1}^n \delta_{[i,n]} \prod_{i=1}^n \left(1 + \frac{1}{n-i}\right)^{1-\delta_{[i,n]}} \mathbf{1}_{Z_{i:n} \leq x} \quad (3.1.17)$$

où $W_{i,n}$ sont les poids pour la i -valeur censurées, sont données par

$$W_{i,n} := \delta_{[i,n]} \prod_{i=1}^n \left(1 + \frac{1}{n-i}\right)^{1-\delta_{[i,n]}} \quad (3.1.18)$$

Noté que si tout $\delta_{[i,n]} = 1$ on trouve que $W_{i,n} = 1$ avec $Z_{i,n} = X_{i,n}$ et $\widehat{F}_n(x) := \frac{1}{n} \sum \mathbf{1}_{X_{i:n} \leq x} = F_n(x)$. C'est à dire que si tout $\delta_{[i,n]} = 1$ la fonction \widehat{F}_n devient la fonction de répartition empirique F_n . Stute, dans les années 95 s'intéresse dans un cadre très général à l'estimation de la fonction de répartition en présence de censure $\int \varphi d\widehat{F}_n$ où φ est présenté toutes les fonctions intégrables. Dans ce cas on prend $\varphi(x) := \mathbf{1}_{Z_{i:n} \leq x} := \delta_{Z_{i:n}}(x)$ où $\delta_{Z_{i:n}}(x)$ c'est la masse de dirac. D'après (3.1.17) et grâce au théorème 1.1 de Stute [36] pour une distribution H continue pour tout $x < x_H$ où $x_H = \infty$ le point terminal de H où (Z_i, δ_i) un couple de c.a de type I (Définition (3.2)) i.i.d suivant la même loi H et avec condition que

$$\int_{-\infty}^{+\infty} A(x) dH^1(x) < \infty, \text{ où } A(x) := (\varphi(x) \eta_0(x))^2 \quad (3.1.19)$$

et

$$\int_{-\infty}^{+\infty} (\varphi(x))^2 \sqrt{C(x)} dF(x) < \infty \text{ où } C(x) := \int_{-\infty}^x \frac{dH^0(s)}{(1-H(s))^2} \quad (3.1.20)$$

Il existe U_1, U_2, \dots, U_n une suite de v.a's i.i.d avec $\bar{U} = 0$ (moyenne arithmétique égale a zéro) avec variance $\sigma^2 = var(U_1)$, telleque :

$$\hat{F}_n(x) := \sum_{i=1}^n W_{i,n} \mathbf{1}_{Z_{i:n} \leq x} = \frac{1}{n} \sum_{i=1}^n U_i + o_p(\sqrt{n}) \quad (3.1.21)$$

où

$$U_i = \delta_{[i,n]} \eta_0(Z_i) \mathbf{1}_{Z_{i:n} \leq x} + (1 - \delta_{[i,n]}) \eta_1(Z_i) - \eta_2(Z_i) \quad (3.1.22)$$

avec la fonction φ est la masse de dirac et d'après Stute 1995 qui a donné $\eta_0(Z_i)$ et $\eta_1(Z_i)$, $\eta_2(Z_i)$ respectivement par :

$$\eta_0(x) := \exp\left(\int_{-\infty}^x \frac{dH^0(z)}{1-H(z)}\right), \eta_1(x) := \frac{1}{1-H(x)} \int_x^{+\infty} \sqrt{A(s)} dH^1(s) \quad (3.1.23)$$

et

$$\eta_2(x) := [1-H(x)] \int_x^{+\infty} \int_t^{+\infty} d\eta_1(s) dC(t) \quad (3.1.24)$$

où $A(s)$ et $C(t)$ sont définies par deux relations (3.1.19) (3.1.20) respectivement. Avec

$$\begin{cases} H^1(s) := P(Z \leq s; \delta_i = 1) := \int_{-\infty}^s \bar{G}(s) dF(s), \\ H^0(s) := P(Z \leq s; \delta_i = 0) := \int_{-\infty}^s \bar{F}(s) dG(s), \end{cases} \quad (3.1.25)$$

et leur version empirique sont respectivement

$$H_n^1(s) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Z_i \leq s, \delta_i = 1\}}, H_n^0(s) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Z_i \leq s, \delta_i = 0\}}, \text{ et } H_n(s) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Z_i \leq s\}}, \quad (3.1.26)$$

On remarque que si tout les $\delta_{[i,n]} = 1$ on trouve $\eta_0(Z_i) = 1$ avec $\eta_2(x) = 0$ et $U_i := \mathbf{1}_{Z_{i:n} \leq x}$.

Théorème 3.1 Soit (Z_i, δ_i) un couple de c a's de type I i.i.d suivant la même loi H pour

$i = 1, \dots, n$ où $Z_i = \min(X_i, Y_i)$ et $\delta_{X_i \leq Y_i}$ est l'indicateur de censure. Tel que X_i est un échantillon de v.a's i.i.d suivant la même loi F et Y_i un deuxième échantillon de v.a's i.i.d suivant la même loi G pour tout $i = 1, \dots, n$ avec X_i et Y_i sont indépendants. Alors sous la condition (3.1.19) et (3.1.20) on obtient

$$\sqrt{n} \left(\widehat{F}_n - F \right) \rightarrow \mathfrak{N} \left(0; \sigma^2 \right) \quad (3.1.27)$$

où

$$\sigma^2 = Var \{ \delta \eta_0(Z) \mathbf{1}_{Z \leq x} + (1 - \delta) \eta_1(Z) - \eta_2(Z) \} \quad (3.1.28)$$

avec $x_H = \infty$ où x_H le point terminal de H et $\eta_0(Z), \eta_1(Z), \eta_2(Z)$ sont définis en (3.1.23) et (3.1.24) respectivement.

Ci-dessous, nous fournissons un exemple simple de la façon de calculer l'estimation de la survie par l'estimateur de Kaplan-Meier.

Exemple 3.2 On observe les durées de séjour de 10 épisodes de chômage exprimés en mois (* si censurées): 1; 2; 4*; 5; 7*; 8; 9; 10*; 11; 13*. Donc les couple (Z_i, δ_i) sont: $\{(1, 1); (2, 1); (4, 0); (5, 1); (7, 0); (8, 1); (9, 1); (10, 0); (11, 1); (13, 0)\}$. Nous programmons cet exemple à l'aide d'un programme illustré comme suit:

```
x=c(1,2,4,5,7,8,9,10,11,13),d=c(1,1,0,1,0,1,1,0,1,0),library(survival),
s=survfit(Surv(x,d) ~1), plot(s)
```

Pour dériver les résultats numériques de l'estimation de la fonction de survie, nous utilisons "summary(s)". Donc, $\widehat{S}_n(t)$ indiqué dans le tableau (3.1) suivant:

t	$[1, 2[$	$[2, 5[$	$[5, 8[$	$[8, 9[$	$[9, 11[$	$[11, [$
$\widehat{S}_n(t)$	0.900	0.800	0.686	0.549	0.411	0.206

Table 3.1: Estimation de la survie par l'estimateur de Kaplan-Meier

La figure (3.2) illustré les résultats obtenus de l'estimateur de Kaplan-Meier en fonction de survie et la fonction de survie (lignes simulées) et bornes de confiance (lignes entières).

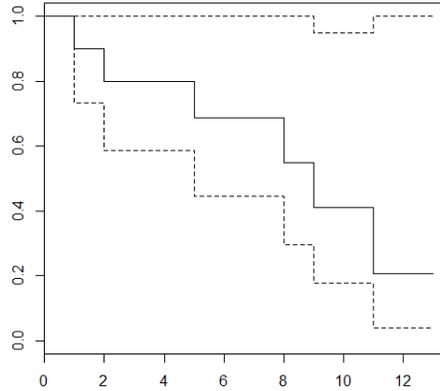


Figure 3.2: Estimateur de Kaplan-Meier (ligne continue)

3.2 Estimation d'indice des valeurs extrêmes sous données censurés

Soit $(X_n, n \geq 1)$ une suite de v.a's i.i.d de distribution F , $F \in D(H_{\gamma_1})$. La fonction de queue F est à variation régulière d'indice $(-1/\gamma_1)$ i.e. $\bar{F} \in RV_{-1/\gamma_1}$ et du théorème (2.4) \bar{F} est aussi à variation régulière :

$$\forall x \geq 0, \lim_{t \rightarrow +\infty} \frac{1 - F(tx)}{1 - F(t)} := x^{-1/\gamma_1} \text{ où } t \in \mathbb{R}.$$

Soit aussi $(Y_n, n \geq 1)$ une suite de v.a i.i.d de distribution G où $G \in D(H_{\gamma_2})$. La fonction de queue de G est variation régulière d'indice $(-1/\gamma_2)$ i.e. $\bar{G} \in RV_{-1/\gamma_2}$ ce implique aussi que \bar{G} est à variation régulière :

$$\forall x \geq 0, \lim_{t \rightarrow +\infty} \frac{1 - G(tx)}{1 - G(t)} := x^{-1/\gamma_2} \text{ où } t \in \mathbb{R}.$$

Nous observons seulement les évènements définies par le couple (Z_i, δ_i) avec $Z_i = \min(X_i, Y_i)$ et l'indicateur $\delta_{X_i \leq Y_i}$ pour tout $i = 1, \dots, n$ où les $(Z_n)_{n \geq 1}$ est une suite de v.a's i.i.d avec la même distribution H où $H \in D(H_\gamma)$. C'est-à-dire, on observe le véritable temps de survie que s'il est inférieur à Y_i qui d'après le définition (3.2) en present une suite de c.a de type I. Alors on a $(1 - H) = (1 - F) \times (1 - G)$ et $\bar{H} := P(X > x, \delta_i = 1) := P(x < Z < Y)$.

Alors on trouve que

$$\forall x \geq 0, \lim_{t \rightarrow +\infty} \frac{1 - H(tx)}{1 - H(t)} = \lim_{t \rightarrow +\infty} \frac{1 - F(tx)}{1 - F(t)} \times \lim_{t \rightarrow +\infty} \frac{1 - G(tx)}{1 - G(t)} := x^{-(1/\gamma_1 + 1/\gamma_2)} \text{ où } t \in \mathbb{R}.$$

et d'après le théorème (2.4) implique que H est une fonction à variation régulière d'indice $-(1/\gamma_1 + 1/\gamma_2)$ notons d'indice $1/\gamma := \left(\frac{\gamma_2 + \gamma_1}{\gamma_1 \gamma_2}\right)$.

D'autre parte par le corollaire (2.1) en déduit $\bar{H} \in RV_{-1/\gamma}$ par

$$\begin{aligned} H(x) &= P(\min(X, Z) < x) := 1 - (1 - F(x)) \times (1 - G(x)) \\ &= 1 - x^{-1/\gamma_1} \times x^{-1/\gamma_2} = 1 - x^{-1/\gamma} \\ \bar{H}(x) &:= x^{-1/\gamma} \end{aligned}$$

3.2.1 Estimateur de la proportion d'observations non censurées

Soit $x_H = \sup\{x \in \mathbb{R}, H(x) \leq 1\}$ (resp., x_F et x_G) le point terminal de support associé a la fonction H (resp., F et G). Dans la suite, on suppose que F et G dans l'un des trois cas suivants :

$$\left\{ \begin{array}{ll} \text{Cas 1 : } \gamma_1 > 0, \gamma_2 > 0, & \text{dans ce cas } \gamma := \left(\frac{\gamma_1 \gamma_2}{\gamma_2 + \gamma_1}\right) \\ \text{Cas 2 : } \gamma_1 < 0, \gamma_2 < 0, x_F = x_G, & \text{dans ce cas } \gamma := \left(\frac{\gamma_1 \gamma_2}{\gamma_2 + \gamma_1}\right) \\ \text{Cas 3 : } \gamma_1 = \gamma_2 = 0, x_F = x_G = \infty & \text{dans ce cas } \gamma := \left(\frac{\gamma_1 \gamma_2}{\gamma_2 + \gamma_1}\right) = 0 \end{array} \right. \quad (3.2.1)$$

D'une part, soit $f(x)$, $g(x)$ et $h(x)$ les fonctions de densité associées aux les distributions F , G et H respectivement et $H^1(x)$ est défini en (3.1.25) on a la proportion d'observations non censurées définie par:

$$\begin{aligned} p &:= P(X \leq Y | Z = x) := \frac{(1 - G(x)) f(x)}{h(x)} := \frac{dH^1(x)}{dH(x)} & (3.2.2) \\ &:= \frac{-\frac{1}{\gamma_1} x^{-1/\gamma_1 - 1} \times x^{-1/\gamma_2}}{-\left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right) x^{-(1/\gamma_1 + 1/\gamma_2) - 1}} := \frac{\frac{1}{\gamma_1}}{\left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right)} \\ p &:= \frac{\gamma}{\gamma_1} \text{ où } \gamma_1 := \frac{\gamma}{p} \end{aligned}$$

En conséquence, le quotient l'estimateur $\hat{\gamma}^{(\cdot)}$ de γ sur l'estimateur $\hat{\gamma}_1^{(\cdot)}$ de γ_1 définit un estimateur \hat{p} de p . Il s'en suivra que \hat{p} estime $\frac{\gamma_2}{(\gamma_1 + \gamma_2)}$ et $\hat{\gamma}^{(\cdot)}$ divisé par \hat{p} estime γ_1 .

Considérons $Z_{1,n} \leq Z_{2,n} \leq \dots \leq Z_{n,n}$ la statistique d'ordre de l'échantillon Z_1, Z_2, \dots, Z_n . Soit $\delta_{[i,n]}$ le concomitant de $Z_{i,n}$ (c'est-à-dire, $\delta_{[i,n]} = \delta_j$ si $Z_{i,n} = Z_j$ où $j = 1, \dots, n$). Soit $Z_{n-k,n}, Z_{n-k+1,n}, \dots, Z_{n,n}$ pour $X > X_{n-k,n}$ les k plus grandes valeur de X . En effet, si la censure n'est pas prise en compte tous les estimateurs d'indice des valeurs extrêmes basé sur les $(Z_n, n \geq 1)$ sont évidemment consistants et tous convergent vers γ l'indice des valeurs de H et non γ_1 l'indice des valeurs extrêmes de F . C'est le contraire de ce que nous allons trouver dans notre cas, avec censure,. Donc, nous devons adapter tous es estimateurs d'indice des valeurs extrêmes à la censure. Nous diviserons tous ces estimateurs par la proportion d'observations non censurées dans les l'échantillon $Z_{n-k,n}, Z_{n-k+1,n}, \dots, Z_{n,n}$ comme suivant:

$$\hat{\gamma}_1^{(c,\cdot)} := \frac{\hat{\gamma}^{(\cdot)}}{\hat{p}} \text{ où } \hat{p} := \frac{1}{k} \sum_{i=1}^k \delta_{n-i+1;n} \quad (3.2.3)$$

où $\hat{\gamma}^{(\cdot)}$ pourrait être tout estimateur non adapté à la censure, en particulier, $\hat{\gamma}^{ML}$, $\hat{\gamma}_H$. On peut définir \hat{p} a partir le version empirique de $\bar{H}_n^1(x)$ et $\bar{H}_n(x)$ en (3.1.25) de relation

précédent comme suivant

$$\widehat{p} := \frac{d\overline{H}_n^{-1}(x)}{d\overline{H}_n(x)} := \frac{1}{k} \sum_{i=1}^k \delta_{n-i+1;n}$$

avec $\delta_{1,n}, \delta_{2,n}, \dots, \delta_{n,n}$ étant ($\delta_n, n \geq 1$) les correspondant à Z_1, Z_2, \dots, Z_n respectivement. Sous la définition de $H^1(x)$ et $H^0(x)$ on a les deux relations (3.1.25) et (3.1.26) respectivement, on définit $p(x)$ par :

$$p(x) := \frac{d\overline{H}^{-1}(x)}{d\overline{H}(x)} := \frac{(1 - G(x)) f(x)}{(1 - G(x)) f(x) + (1 - F(x)) g(x)} \quad (3.2.4)$$

où $f(x)$ et $g(x)$ les fonctions de densité associées aux distributions F, G respectivement avec F et G satisfait l'un des trois cas associés à (3.2.1) et qui vérifient $\lim_{x \rightarrow x_H} p(x) := p$. Mais nous intéressons aux deux premier cas dans (3.2.1). Car pour cas 3: $\gamma_1 = \gamma_2 = \gamma = 0$ on a $\sqrt{k}(\widehat{p} - p) := \frac{1}{\gamma_1} \sqrt{k}(\gamma_1 \widehat{p} - \gamma) = 0$ puisque par (3.2.2) on a $p := \frac{\gamma}{\gamma_1}$. Cela signifie que $\widehat{p} \xrightarrow{p} p > 0$. Intéressons-nous maintenant en détail au terme $\sqrt{k}(\widehat{p} - p)$ pour les cas 1 et 2.

Théorème 3.2 *Supposons $k_n = k$ une suite d'entiers où $1 < k < n$ avec $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{k}} \sum_{i=1}^k \left[p \left(H^{-1} \left(1 - \frac{i}{n} \right) \right) - p \right] \rightarrow \alpha_1 \in \mathbb{R} \quad (3.2.5)$$

et

$$\sqrt{k} \sup_{\{1-k/n \leq t < 1, |t-s| \leq C\sqrt{k}/n, s < 1\}} |p(H^{-1}(t)) - p(H^{-1}(s))| \rightarrow 0 \text{ pour tout } C > 0 \quad (3.2.6)$$

avec H^{-1} représente l'inverse généralisée de H . Alors

$$\sqrt{k}(\widehat{p} - p) \rightarrow N(\alpha_1, p(1-p)) \quad (3.2.7)$$

où α_1 désigne le biais de $\sqrt{k}(\hat{p} - p)$.

Preuve. Considérer la construction suivante. Soit Z une v.a de distribution H . Et soit U une v.a de loi uniforme(0, 1) indépendante de Z . Définissant :

$$\delta := \begin{cases} 1, & U \leq p(Z) \\ 0, & U > p(Z) \end{cases} \quad \text{et } \tilde{\delta} := \begin{cases} 1, & U \leq p \\ 0, & U > p \end{cases}$$

Nous répétons indépendamment cette construction n fois. Nous montrons que les paires (Z_i, δ_i) pour $i = 1, \dots, n$ ont la même distribution que les paires initiales $(Z_n, \delta_n)_{n \geq 1}$. Donc nous travaillons avec les nouvelles paires (Z_i, δ_i) pour $i = 1, \dots, n$. En outre, on a que Z et $\tilde{\delta}$ sont indépendantes et vérifient :

$$P\left(\left|\delta - \tilde{\delta}\right| = 1 \mid Z = x\right) = |p - p(x)|$$

Considérons $Z_{1,n} \leq Z_{2,n} \leq \dots \leq Z_{n,n}$ la statistique d'ordre de l'échantillon Z_1, Z_2, \dots, Z_n . et la statistique d'ordre de U_1, U_2, \dots, U_n est donnée par $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$, telle que Z_i et U_i sont indépendants pour $i = 1, \dots, n$. Nous pouvons écrire \hat{p} comme suit

$$\hat{p} := \frac{1}{k} \sum_{i=1}^k \delta_{n-i+1;n} := \frac{1}{k} \sum_{i=1}^k 1_{\{U_{n-i+1} \leq p(Z_{n-i+1})\}} \quad (3.2.8)$$

et même \tilde{p} par

$$\tilde{p} := \frac{1}{k} \sum_{i=1}^k \tilde{\delta}_{n-i+1;n} := \frac{1}{k} \sum_{i=1}^k 1_{\{U_{n-i+1} \leq p\}} \quad (3.2.9)$$

Alors quand $k \rightarrow \infty$ on trouve:

$$\sqrt{k}(\hat{p} - p) - \sqrt{k}(\hat{p} - \tilde{p}) \xrightarrow{d} \sqrt{k}(\tilde{p} - p)$$

Puis sous (3.2.9) on trouve que \tilde{p} est distribuée selon une loi binomial (k, p) avec moyenne

kp et variance $kp(1-p)$, alors par le théorème limite central (1.4) on a

$$\sqrt{k}(\tilde{p} - p) \xrightarrow{d} N(0; p(1-p))$$

Pour $\sqrt{k}(\hat{p} - \tilde{p})$ le terme du biais de $\sqrt{k}(\hat{p} - p)$ peut être réécrit en la décomposition suivante :

$$\begin{aligned} \sqrt{k}(\hat{p} - \tilde{p}) &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \left[1_{\{U_i \leq p(Z_{n-i+1})\}} - 1_{\{U_i \leq p\}} \right] \\ &= \frac{1}{k} \sum_{i=1}^k \left[1_{\{U_i \leq p(Z_{n-i+1})\}} - 1_{\{U_i \leq p(H^{-1}(1-\frac{i}{n}))\}} \right] + \frac{1}{k} \sum_{i=1}^k \left[1_{\{U_i \leq p(H^{-1}(1-\frac{i}{n}))\}} - 1_{\{U_i \leq p\}} \right] \\ &= S_{1:k} + S_{2:k} \end{aligned}$$

Einmahl et al [11] (2008) prouve sous les hypothèses (3.2.5) et (3.2.6) que $S_{2:k} \rightarrow \alpha_1$ et $S_{1:k} \rightarrow 0$ respectivement ■

3.2.2 Estimateurs adaptatifs d'indice de queue sous données censurées

Einmahl et al [11] (2008) ont proposés différents estimateurs de l'indice des valeurs extrêmes sous les données censurées . Une façon de procéder est décrite dans la relation (3.2.3).

Alors, à partir (3.2.2) et (3.2.3) et le théorème (3.2). On considère la décomposition suivante

$$\sqrt{k} \left(\hat{\gamma}_1^{(c,\cdot)} - \gamma_1 \right) := \frac{1}{p} \sqrt{k} (\hat{\gamma}^{(\cdot)} - \gamma) - \gamma_1 \sqrt{k} (\tilde{p} - p) - \frac{\gamma_1}{p} \alpha_1 + o_p(1) \quad (3.2.10)$$

Le comportement asymptotique de $\sqrt{k}(\hat{\gamma}^{(\cdot)} - \gamma)$ est bien connu puisque cet estimateur est basé sur le Z -échantillon. C'est-à-dire sur la situation non censurée. Mais dans notre

travail on considère le cas $\gamma > 0$ c'est à dire le cas 1 (3.2.1) pour cela nous choisissons les deux estimateur pour les distributions à queue lourde et l'estimateur de Hill généralisé qui est présenté par Beirlant et al [4] (2005) ainsi que l'estimateur du maximum de vraisemblance (ML-estimateur) de Drees et al [10] (2004) de plus, ML-estimateur est qualifié pour les trois cas défini en (3.2.1) comme similaire; avec la relation (3.2.10) on trouve que :

$$\sqrt{k} \left(\hat{\gamma}_1^{(c,ML)} - \gamma_1 \right) := \frac{1}{p} \sqrt{k} (\hat{\gamma}^{ML} - \gamma) - \gamma_1 \sqrt{k} (\tilde{p} - p) - \frac{\gamma_1}{p} \alpha_1 + o_p(1)$$

et

$$\sqrt{k} \left(\hat{\gamma}_1^{(c,GH)} - \gamma_1 \right) := \frac{1}{p} \sqrt{k} (\hat{\gamma}_{k,n}^{GH} - \gamma) - \gamma_1 \sqrt{k} (\tilde{p} - p) - \frac{\gamma_1}{p} \alpha_1 + o_p(1)$$

On suppose ici que $k = k_n$ est une séquence intermédiaire, c'est-à-dire, pour $1 < k < n$ où $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$.

Théorème 3.3 *Supposons que toutes les conditions du théorèmes (2.12) et (2.15) sont satisfaites, et pour α_1 une constante définie en (3.2.5) et $\rho < 0$. Alors pour $n \rightarrow \infty$ nous avons*

$$\sqrt{k} \left(\hat{\gamma}_1^{(c,ML)} - \gamma_1 \right) \xrightarrow{d} N \left(\frac{1}{p} \lambda \frac{(\gamma + 1)}{(1 - \rho)(1 + \gamma - \rho)} - \frac{\gamma_1}{p} \alpha_1; \frac{\gamma(2 + \gamma_1) + 1}{p^2} \right) \quad (3.2.11)$$

où μ_a est défini au théorèmes (2.12) pour $\gamma > \frac{-1}{2}$. Même pour $\gamma > 0$, quand $n \rightarrow \infty$ nous avons

$$\sqrt{k} \left(\hat{\gamma}_1^{(c,GH)} - \gamma_1 \right) \xrightarrow{d} N \left(\frac{1}{p} \frac{\tilde{\lambda}}{1 - \tilde{\rho}} - \frac{\gamma_1}{p} \alpha_1; \frac{1 + \gamma_1^2 p}{p^2} \right) \quad (3.2.12)$$

où $\tilde{\rho} = -\gamma \mathbf{1}_{\{\gamma + \rho > 0; \beta \neq 0\}} + \rho \mathbf{1}_{\{\gamma + \rho > 0; \beta = 0\}}$.

Nous mentionnons que le biais asymptotique de l'estimateur du maximum de vraisemblance (ML-estimateur) découle facilement des calculs directs que nous présentons dans la sous-section (2.4.1). Il est basé sur les travaux de Drees et al [10] (2004) et De Haan et Ferriera [18].

Toutefois, Einmahl et al [11] (2008) ont proposés un autre biais asymptotique pour ML-estimateur. Ils utilisent une condition de second ordre à partir des relations (2.4.28)-(2.4.30).

Ils ont prouvés que

$$\sqrt{k} \left(\hat{\gamma}_1^{(c,ML)} - \gamma_1 \right) \xrightarrow{d} N \left(\frac{1}{p} \left(\frac{\lambda \tilde{\alpha} \rho (\gamma + 1)}{(1 - \rho)(1 + \gamma - \rho)} - \gamma_1 \alpha_1 \right); \frac{\gamma (2 + \gamma_1) + 1}{p^2} \right) \quad (3.2.13)$$

où $\lim_{n \rightarrow +\infty} A\left(\frac{n}{k}\right) = \lambda$ et $\tilde{\alpha}$ présente dans (2.4.30).

Einmahl et al [11] (2008) ont adaptés tous les estimateurs classiques de l'indice de queue sous données censurées; parais eux l'estimateur du maximum de vraisemblance $\hat{\gamma}^{ML}$ décrit dans la section (2.4.1) est considéré comme un estimateur consistant de l'indice de queue basé sur l'échantillon complet, et divisé par la proportion d'observations non censurées dans la queue , voir la relation (3.2.3). Par contre Kouider et al [28] (2023) ont constatés qu'il n'était pas possible d'adapter l'estimateur du maximum de vraisemblance sous données censurées $\hat{\gamma}_1^{(c,ML)}$ par relation (3.2.3), sauf sous certaines conditions que nous expliquerons dans le chapitre suivant.

Chapitre 4

Algorithme de bisection modifié pour estimer l'indice de queue

4.1 Algorithme de bisection modifié pour multi-racine

Le processus de recherche de la racine de la fonction consiste à trouver la valeur de x pour laquelle $f(x) = 0$. Si la fonction est égale à zéro, x est la racine de la fonction. L'algorithme de recherche de racine le plus simple est la méthode de la bisection, également appelée méthode de réduction de moitié de l'intervalle. Ça marche quand $f(x)$ est une fonction continue et elle nécessite une connaissance préalable de deux suppositions initiales α et β tel que $f(\alpha)$ et $f(\beta)$ ont des signes opposés, puis trouvez le milieu de $[\alpha; \beta]$ et après on décide on $[\alpha; (\alpha + \beta) / 2]$ ou $[(\alpha + \beta) / 2; \beta]$ répéter jusqu'à ce que l'intervalle soit suffisamment petit. Cependant, la méthode de bisection est fiable, mais elle converge lentement, gagnant un peu en précision avec l'une ou l'autre itération.

4.1.1 Méthode de bisection modifiée (MBM)

Dans cette partie nous avons intéressé à la méthode numérique qui est la méthode de bisection modifiée (MBM: The Modified Bisection Method). Elle est comme une méthode

d'interpolation linéaire est également connue sous le nom de méthode de regula falsi (fausse position) voir [23] , mais avec une autre manière et plus de précision. Par le théorème des valeurs intermédiaires implique qu'un nombre x^* existe dans $[\alpha; \beta]$ avec $f(x^*) = 0$ si $f(\alpha) \cdot f(\beta) < 0$ ce qui signifie que l'un d'eux est au-dessus de l'axe des x et l'autre sous l'axe des x .

Donc, soit $f(x)$ une fonction continue et définie sur l'intervalle $[\alpha; \beta]$ avec $f(\alpha) \cdot f(\beta) < 0$. Alors, nous fixons $\alpha = \alpha_1$ et $\beta = \beta_1$. Pour un entier $s \geq 1$, on peut trouver l'équation de la droite à partir des points $(\alpha_s; f(\alpha_s))$ et $(\beta_s; f(\beta_s))$ comme $y = ax + b$ où

$$a = \frac{f(\beta_s) - f(\alpha_s)}{\beta_s - \alpha_s} \text{ ou } a = \frac{f(\alpha_s) - f(\beta_s)}{\alpha_s - \beta_s}$$

et

$$b = f(\beta_s) - a\beta_s \text{ ou } b = f(\alpha_s) - a\alpha_s$$

D'où le x -l'intersection de la droite est en un point $(x_s; 0)$ avec $x_s = \frac{-b}{a}$ c'est

$$x_s = \beta_s - f(\beta_s) \frac{\beta_s - \alpha_s}{f(\beta_s) - f(\alpha_s)} \quad (4.1.1)$$

ou

$$x_s = \alpha_s - f(\alpha_s) \frac{\beta_s - \alpha_s}{f(\beta_s) - f(\alpha_s)} \quad (4.1.2)$$

C'est clair que

$$x_1 = \alpha - f(\alpha) \frac{\beta - \alpha}{f(\beta) - f(\alpha)} \quad (4.1.3)$$

ou

$$x_1 = \beta - f(\beta) \frac{\beta - \alpha}{f(\beta) - f(\alpha)} \quad (4.1.4)$$

Par le théorème des valeurs intermédiaires lorsque f est une fonction continue et monotone sur $[\alpha; \beta]$, nous avons pour $s \geq 1$, $f(\alpha) < f(x_s) < f(\beta)$ ou $f(\alpha) > f(x_s) > f(\beta)$ on peut alors définir deux cas.

Cas1: si $f(\alpha)$ et $f(x_1)$ ont des signes opposés ceci implique que $f(\beta)$ et $f(x_1)$ ont le même signe donc le travail reste sur l'intervalle $[\alpha; x_1]$ et pour tous $x_s \in [\alpha; x_1]$ tel que défini dans la relation (4.1.2) on a $f(\alpha) \cdot f(x_s) < 0$ alors x_s dans (4.1.2) devient défini pour $x_0 = \beta$ par

$$x_{s+1} = \alpha - f(\alpha) \frac{x_s - \alpha}{f(x_s) - f(\alpha)} \quad (4.1.5)$$

Enfin, nous choisissons le nouveau intervalle pour la prochaine itération $[\alpha; x_{s+1}^*] = [\alpha; x_{s+1}]$ avec $s = s+1$. Nous restons dans le même processus jusqu'à ce que nous trouvions une solution ou une valeur approchée, voir la figure (4.1) suivante.:

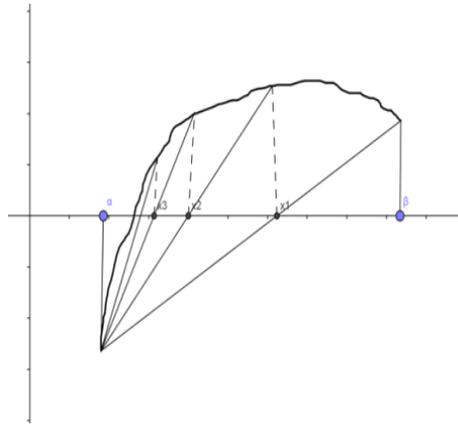


Figure 4.1: Illustration de la MBM avec cas 1.

Cas2: si $f(\beta)$ et $f(x_1)$ ont des signes opposés implique que $f(\alpha)$ et $f(x_1)$ ont le même signe donc le travail reste sur l'intervalle $[x_1; \beta]$. On trouve dans ce cas $x_s \in [x_1; \beta]$. Et pour $x_0 = \alpha$ on a

$$x_{s+1} = \beta - f(\beta) \frac{x_s - \beta}{f(x_s) - f(\beta)} \quad (4.1.6)$$

Et nous choisissons le nouveau sous-intervalle pour la prochaine itération $[x_{s+1}^*; \beta] = [x_{s+1}; \beta]$.

On peut dire que cette méthode est efficace avec des fonctions qui ont les représentations montrées dans la figure (4.1) précédente et la figure (4.2) suivante, telles que les fonctions concaves. Et ce qui distingue cette méthode, c'est qu'elle vérifie le théorème des valeurs

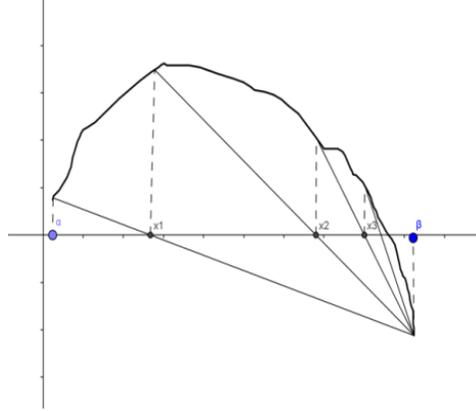


Figure 4.2: Illustration de la MBM avec cas 2.

intermédiaires pour la première étape pour déterminer la valeur du nombre tel qu'il est donné dans les deux formules (4.1.3) et (4.1.4) selon chacun des deux cas indiqués ci-dessus respectivement. Donc les étapes pour appliquer la méthode de bisection modifiée pour trouver la racine de l'équation $f(x) = 0$ est comme suivant:

Étape 1: Choisir α et β tels que $\alpha < \beta$ avec $f(\alpha) \cdot f(\beta) < 0$ $i = 1, \dots, d$, avec une tolérance $\zeta \approx 1 \times 10^{-7}$.

Étape 2: Calculer,

$$x_1 = \beta - f(\beta) \frac{\beta - \alpha}{f(\beta) - f(\alpha)} \text{ ou } x_1 = \alpha - f(\alpha) \frac{\beta - \alpha}{f(\beta) - f(\alpha)}$$

Étape 3: si $f(\alpha) \cdot f(x_1) < 0$, pour $n > 1$, composer

$$x_n = \alpha - f(\alpha) \frac{x_{n-1} - \alpha}{f(x_{n-1}) - f(\alpha)} \quad (4.1.7)$$

sinon

$$x_n = \beta - f(\beta) \frac{x_{n-1} - \beta}{f(x_{n-1}) - f(\beta)} \quad (4.1.8)$$

Étape 4: Si $|x_n - x^*| < 1 \times 10^{-7}$, alors stopper le programme (i.e $x_n = x^*$) donc la racine est x_n . Sinon, si $f(\alpha) \cdot f(x_1) < 0$ alors $\alpha = \alpha$ et $x_n^* = x_{n-1}$ sinon $\beta = \beta$ et $x_n^* = x_{n-1}$. Et mettre $n = n + 1$, et aller à l'étape 3.

Ensuite, nous présentons des théorèmes pour la borne d'erreur d'un MBM, mais avant d'aller à ce théorème rappelons les mêmes définitions. On dit qu'une fonction f est définie sur l'intervalle I de \mathbb{R} admet un maximum en $\theta \in I$ si pour tout $x \in I$ on a $f(x) \leq f(\theta)$ et on dit que $f(\theta)$ est le maximum global de f . De plus, si l'intervalle I a une borne, ou une norme, on dit que f admet un maximum local, ou un maximum relatif en $\theta \in I$ s'il y a un voisinage V de θ telle que pour tout $x \in V$, on a $f(x) \leq f(\theta)$. Bien sûr, en changeant les inégalités de sens on peut définir un minimum global et un minimum local. Un extremum est un maximum ou un minimum. La recherche de l'extremum est liée au calcul différentiel. Ainsi si f est défini sur intervalle $I \in \mathbb{R}$ et si f est différentiable en θ , alors, $f'(\theta) = 0$. L'étude du signe de f' sur le voisinage de θ permet souvent de conclure à l'existence d'un maximum ou d'un minimum.

Théorème 4.1 *Soit f une fonction continue et définie sur $[\alpha; \beta]$ telle que $f(\alpha) \cdot f(\beta) < 0$. Une nouvelle bisection modifiée génère une séquence $\{x_n\}_{n=1}^{\infty}$ se rapprochant d'un zéro x^* de f et pour*

$$\tau = \frac{f(\theta_2)}{f(\theta_1)} \quad (4.1.9)$$

où $f(\theta_2)$ est un maximum de f en $[\alpha; \beta]$ et $f(\theta_1)$ est un minimum de f en $[\alpha; \beta]$ respectivement. On a :

Cas 1 : Pour $f(\alpha) \cdot f(x_1) < 0$

$$|x_n - x^*| < \left| \frac{f(\alpha)}{f(\theta_1)} \right|^n \frac{\beta - \alpha}{|\tau - 1|^n} \text{ où } x_1 \text{ défini dans (4.1.3).} \quad (4.1.10)$$

Cas 2: Pour $f(\beta) \cdot f(x_1) < 0$

$$|x_n - x^*| < \left| \frac{f(\beta)}{f(\theta_2)} \right|^n \frac{\beta - \alpha}{\left| \frac{1}{\tau} - 1 \right|^n} \text{ où } x_1 \text{ défini dans (4.1.4).} \quad (4.1.11)$$

Théorème 4.2 *Soit f une fonction continue définie sur $[\alpha; \beta]$ telle que $f(\alpha) \cdot f(\beta) < 0$. Avec $f(\alpha) \leq f(x) \leq f(\beta)$ comme un extremum global pour tous $x \in [\alpha; \beta]$. Le théorème*

précédent devient pour,

$$\tau = \frac{f(\beta)}{f(\alpha)} \quad (4.1.12)$$

on a,

$$\left\{ \begin{array}{l} \text{Cas1: Pour } f(\alpha) \cdot f(x_1) < 0 : |x_n - x^*| < \frac{\beta - \alpha}{|\tau - 1|^n} \text{ où } x_1 \text{ défini dans (4.1.3).} \\ \text{Cas2: Pour } f(\beta) \cdot f(x_1) < 0 : |x_n - x^*| < \frac{\beta - \alpha}{|\frac{1}{\tau} - 1|^n} \text{ où } x_1 \text{ défini dans (4.1.4).} \end{array} \right. \quad (4.1.13)$$

Nous avons intéressons au théorème (4.2), On a $f(\alpha) \cdot f(\beta) < 0$ implique que le nombre qui est définit en (4.1.12) $\tau < 0$. Donc, si $\tau = -1$ avec les deux cas du théorème (4.2) on a :

$$|x_n - x^*| < \frac{\beta - \alpha}{2^n}. \quad (4.1.14)$$

qui est la borne d'erreur de la méthode de bisection, donc un MBM lié à la valeur du nombre réel $\tau \neq 0$. Toutefois, avec le théorème (4.2) pour $\tau < 0$ nous trouvons que

$$\left\{ \begin{array}{l} \text{Cas1: Pour } f(\alpha) \cdot f(x_1) < 0 : \frac{\beta - \alpha}{2^n} < \frac{\beta - \alpha}{|\tau - 1|^n} \quad \text{si } \tau \in [-1; 0[\text{ où } x_1 \text{ défini dans (4.1.3).} \\ \text{Cas2: Pour } f(\beta) \cdot f(x_1) < 0 : \frac{\beta - \alpha}{2^n} < \frac{\beta - \alpha}{|\frac{1}{\tau} - 1|^n} \quad \text{si } \tau \in]-\infty; -1[\text{ où } x_1 \text{ défini dans (4.1.4).} \end{array} \right.$$

Alors si les deux cas précédents sont satisfait, la MBM est de convergence lente que la méthode de la bisection et c'est ce que nous devons calculer la valeur de τ avant de travailler avec le MBM. De plus, si $\tau < -1$ avec le premier cas et si $-1 < \tau < 0$ avec le deuxième cas qui sont données dans le théorème (4.2) respectivement, la MBM a une meilleure convergence que la méthode de Bisection. Donc le choix du nombre τ a un rôle important dans la crédibilité du MBM car si nous avons $|\tau - 1| \leq 2$ avec cas1 ou $|\frac{1}{\tau} - 1| \leq 2$ avec cas2 comme décrit au théorème (4.1) et théorème (4.2) respectivement, la MBM devien inutile et la méthode bisection est meilleure qu'une MBM dans cette situation. Contrairement à cette situation, le MBM est meilleur et plus rapide que la méthode de la bisection si nous trouvons que $|\tau - 1| > 2$ avec cas1 ou $|\frac{1}{\tau} - 1| > 2$ avec cas2. Et un autre point important

est le choix de l'intervalle $[\alpha; \beta]$ augmente avec $f(\alpha) \cdot f(\beta) < 0$ et la puissance de la MBM consiste à sélectionner des images $f(\alpha)$ et $f(\beta)$ comme un extremum global et ceci est indiqué par le théorème (4.2) que si f est continue et monotone sur $[\alpha; \beta]$.

4.1.2 Algorithme de bisection modifié (MBA)

Tanakan [37] dans son article a proposé un algorithme nommé Algorithme de Bisection Modifié (MBA: Modified Bisection Algorithm). Qui est basé sur la méthode de la bisection et la valeur de x_s comme nous l'avons défini dans (4.1.7) et (4.1.8) pour résoudre des équations non linéaires qui ont une racine, pour plus de détails, voir Tanakan [37]. Donc, dans cette partie nous utilisons deux méthodes numérique qui sont la méthode de Bisection et la MBM comme selon ce qui suit :

Soit $f(x)$ une fonction continue et définie sur $[\alpha; \beta]$ où $f(\alpha) \cdot f(\beta) < 0$. Notons $\alpha_0 = \alpha$ et $\beta_0 = \beta$. Pour un entier $n \geq 1$, par la méthode de la Bisection, on a $x_{n-1}^* = (\alpha_{n-1} + \beta_{n-1})/2$. Ensuite, nous considérons un nouveau sous-intervalle $(\alpha_{n-1}^*, \beta_{n-1}^*)$ par

$$(\alpha_{n-1}^*, \beta_{n-1}^*) := \begin{cases} (\alpha_{n-1}, x_{n-1}^*), & \text{si } f(\alpha) \cdot f(x_{n-1}^*) < 0 \\ (x_{n-1}^*, \beta_{n-1}), & \text{si } f(x_{n-1}^*) \cdot f(\beta) < 0 \end{cases}$$

Ensuite, par la méthode de la Bisection Modifié dans un nouveau sous-intervalle $(\alpha_{n-1}^*, \beta_{n-1}^*)$ nous avons

$$x_n = \alpha_{n-1}^* - f(\alpha_{n-1}^*) \frac{\beta_{n-1}^* - \alpha_{n-1}^*}{f(\beta_{n-1}^*) - f(\alpha_{n-1}^*)} \text{ ou } x_n = \beta_{n-1}^* - f(\beta_{n-1}^*) \frac{\beta_{n-1}^* - \alpha_{n-1}^*}{f(\beta_{n-1}^*) - f(\alpha_{n-1}^*)}$$

Enfin, nous choisissons le nouveau sous-intervalle pour la prochaine itération comme suit:

$$(\alpha_n^*, \beta_n^*) := \begin{cases} (\alpha_{n-1}^*, x_n), & \text{si } f(\alpha) \cdot f(x_n) < 0 \\ (x_n, \beta_{n-1}^*), & \text{si } f(x_n) \cdot f(\beta) < 0 \end{cases} \quad (4.1.15)$$

Le processus se poursuit jusqu'à ce que l'intervalle soit suffisamment petit ou que la so-

lution approximative soit suffisamment proche de la solution exacte x^* . En bref, Tanakan [37] a innové algorithme de bisection modifié pour le racine de f comme suivant:

Étape 1: Pour une tolérance de l'ordre de $\zeta \approx 10^{-7}$; et $\alpha_0^* = \alpha$, $\beta_0^* = \beta$ où $f(\alpha) \cdot f(\beta) < 0$.

Étape 2: Pour $n \geq 1$, calculer $x_{n-1}^* = (\alpha_{n-1} + \beta_{n-1}) / 2$.

Étape 3: Calculer un sous-intervalle $(\alpha_{n-1}^*, \beta_{n-1}^*)$ par

$$(\alpha_{n-1}^*, \beta_{n-1}^*) = \begin{cases} (\alpha, x_{n-1}^*), & \text{si } f(\alpha) \cdot f(x_{n-1}^*) < 0 \\ (x_{n-1}^*, \beta), & \text{si } f(x_{n-1}^*) \cdot f(\beta) < 0 \end{cases}$$

Étape 4: Calculer

$$x_n = \beta_{n-1}^* - f(\beta_{n-1}^*) \frac{\beta_{n-1}^* - \alpha_{n-1}^*}{f(\beta_{n-1}^*) - f(\alpha_{n-1}^*)} \text{ ou } x_n = \alpha_{n-1}^* - f(\alpha_{n-1}^*) \frac{\beta_{n-1}^* - \alpha_{n-1}^*}{f(\beta_{n-1}^*) - f(\alpha_{n-1}^*)}$$

Étape 5: Si $|f(x_n)| < 1 \times 10^{-7}$, alors stops le programme (i.e $x_n = x_{n-1}^*$) donc le racine est x_n . Sinon

$$(\alpha_n^*, \beta_n^*) = \begin{cases} (\alpha_{n-1}^*, x_n), & \text{si } f(\alpha_{n-1}^*) \cdot f(x_n) < 0 \\ (x_n, \beta_{n-1}^*), & \text{si } f(x_n) \cdot f(\beta_{n-1}^*) < 0 \end{cases}$$

et mettre $n = n + 1$, et aller a l'étape 3.

De l'algorithme de la méthode de bisection modifiée, nous avons le théorème suivant.

Théorème 4.3 *Soit f une fonction continue définie sur $[\alpha; \beta]$ telle que $f(\alpha) \cdot f(\beta) < 0$. L'algorithme de bisection modifié qui satisfait l'équation (4.1.15) génère une séquence $\{x_n\}_{n=1}^{\infty}$ se rapprochant d'un zéro x^* de f et pour le quel on a :*

$$|x_n - x^*| < \frac{\beta - \alpha}{2^{n+1}}. \quad (4.1.16)$$

Pour la démonstration du théorème (4.3) voir Tanakan [37].

On peut donc dire que l'algorithme de bisection modifiée est devenue valable pour toutes les fonctions, quel que soit leur comportement sur l'intervalle $[\alpha; \beta]$, avec la nécessité de

remplir la condition du théorème des valeurs intermédiaires à chaque étape, voir la figure (4.3) suivante.:

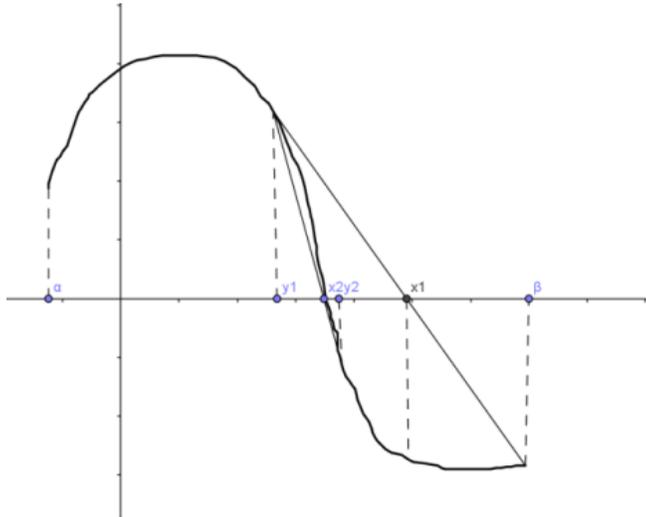


Figure 4.3: Illustration de la MBA (dont la racine est x_2).

4.1.3 MBA pour multi-racine

L'algorithme de bisection modifié proposé par Tanakan [37] pour résoudre une équation non linéaire comme $\psi(\theta) = 0$. Est beaucoup plus efficace que la méthode de la bisection. De plus, elle est plus rapide que la méthode de Newton, et ne compte pas la dérivée d'une fonction au point de référence, ce qui n'est pas toujours facile. Dans la pratique, la solution initiale est vraiment importante pour la méthode de Newton. Mais certaines solutions initiales la méthode de Newton peuvent faire diverger. De plus Kouider [26] a développé cette méthode numérique et a innové algorithme de bisection modifié pour les multi-racines que nous proposons comme suivant:

Étape 1: Divisez l'intervalle $[\alpha; \beta]$ en sous intervalles par $h = (\beta - \alpha)/d$ pour $d \in \mathbb{N}$ est le nombre d'intervalles. Par exemple $d = 10$

Étape 2: Pour $i = 1, \dots, d$, une tolérance de l'ordre de $\zeta \approx 1 \times 10^{-7}$; et $\alpha_0^* = \alpha_i, \beta_0^* = \beta_{i+1}$ où $f(\alpha_i) \cdot f(\beta_{i+1}) < 0$.

Étape 3: Pour $n \geq 1$, calculer

$$x_{n-1}^* = \frac{(\alpha_{n-1} + \beta_{n-1})}{2}. \quad (4.1.17)$$

Étape 4: Calculer un sous-intervalle $(\alpha_{n-1}^*, \beta_{n-1}^*)$ par

$$(\alpha_{n-1}^*, \beta_{n-1}^*) = \begin{cases} (\alpha, x_{n-1}^*), & \text{si } f(\alpha) \cdot f(x_{n-1}^*) < 0 \\ (x_{n-1}^*, \beta), & \text{si } f(x_{n-1}^*) \cdot f(\beta) < 0 \end{cases}$$

Étape 5: Calculer,

$$x_n = \beta_{n-1}^* - f(\beta_{n-1}^*) \frac{\beta_{n-1}^* - \alpha_{n-1}^*}{f(\beta_{n-1}^*) - f(\alpha_{n-1}^*)} \text{ ou } x_n = \alpha_{n-1}^* - f(\alpha_{n-1}^*) \frac{\beta_{n-1}^* - \alpha_{n-1}^*}{f(\beta_{n-1}^*) - f(\alpha_{n-1}^*)}$$

Étape 6: Si $|x_n - x_{n-1}^*| < 1 \times 10^{-7}$, alors stops le programme (i.e $x_n = x_{n-1}^*$) donc le racine est x_n . Sinon

$$(\alpha_n^*, \beta_n^*) = \begin{cases} (\alpha_{n-1}^*, x_n), & \text{si } f(\alpha_{n-1}^*) \cdot f(x_n) < 0 \\ (x_n, \beta_{n-1}^*), & \text{si } f(x_n) \cdot f(\beta_{n-1}^*) < 0 \end{cases}$$

et mettre $n = n + 1$, et aller a l'étape 3.

Pour travailler avec MBA pour les multi-racines, il faut calculer x_{n-1}^* par (4.1.17) et x_n par (4.1.15) respectivement. Mais on peut inverser le processus en calculant x_n puis un x_{n-1}^* .

Théorème 4.4 Soit f une fonction continue et définie sur $[\alpha_i; \beta_{i+1}] \subset [\alpha; \beta]$ avec $f(\alpha_i) \cdot f(\beta_{i+1}) < 0$ existant à chaque itération. MBA pour les multi-racines génère les séquences $\{x_{n-1}^*\}_{n=1}^\infty$ et $\{x_n\}_{n=1}^\infty$ se rapprochant d'un zéro x^* de f avec

$$|x_n - x_{n-1}^*| < \frac{3(\beta - \alpha)}{4 \cdot 2^n}. \quad (4.1.18)$$

où $i = 1, \dots, d$ et d le nombre de sous-intervalles de $[\alpha; \beta]$

Preuve. Il est facile de trouver que les sous intervalles $[\alpha_i; \beta_{i+1}] \subset [\alpha; \beta]$ vérifient

$$(\beta_{i+1} - \alpha_i) \leq \frac{(\beta - \alpha)}{2}$$

Après le résultat théorique (4.1.14) pour la méthode de Bisection nous avons :

$$|x_{n-1}^* - x^*| < \frac{(\beta - \alpha)}{2^{n+1}}. \quad (4.1.19)$$

et pour la MBA avec (4.1.16) on a aussi :

$$|x_n - x^*| < \frac{(\beta - \alpha)}{2^{n+2}}. \quad (4.1.20)$$

On peut montrer que : $|x_n - x_{n-1}^*| := |x_n - x^* - (x_{n-1}^* - x^*)|$. Et il est facile de prouver ce théorème en utilisant une induction mathématique. ■

Par conséquent, pour appliquer la MBM, il faut connaître le comportement de la fonction sur l'intervalle $[\alpha; \beta]$ donné, et il faut diviser les l'intervalles $[\alpha; \beta]$ en plusieurs intervalles $[\alpha_i; \beta_{i+1}]$ qui sont contenues dans $[\alpha; \beta]$, et plus les $[\alpha_i; \beta_{i+1}]$ sont petits, meilleur est le travail de cette façon. Les fonctions concaves est très compatible avec la MBM. Et le reste des fonctions alors. C'est ce que nous allons discuter avec MBA pour les multi-racines. Pour se débarrasser de la divergence MBM, nous introduisons la méthode Bisection avec la condition que le théorème des valeurs intermédiaires soit toujours satisfait pour chaque itération de MBA pour les multi-racines donc nous pouvons dire qu'il est valable pour toute fonction concave et non concave qu'elle admet une multi-racine. De plus, nous définitions la MBM pour multi-racines voir appendix (5.1.2).

4.2 MBA pour estimer l'indice de queue sous données complètes ($\gamma > -1$)

L'estimation du maximum de vraisemblance de la GPD $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ est la valeur qui maximise la fonction de log-vraisemblance définie en (2.4.5). L'estimations du maximum de vraisemblance de la GPD doit être dérivé numériquement car il n'y a pas de simplification évidente du système d'équations (2.4.8) de vraisemblance non linéaire. Il existe de nombreuses techniques numériques pour calculer $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ qui ont été proposées dans de nombreux articles dont les ouvrages importants comme Hosking et Wallis [21] pour $-1/2 < \gamma < 1/2$ et Grimshaw [15] pour $\gamma \leq 1$ et Kouider [26] pour $\gamma \geq -1$.

Nous avons proposés un algorithme, pour estimer (γ, σ) mais avec des condition meilleurs. Nous utilisons cet algorithme pour estimer l'indice de queue lorsque $\gamma \geq -1$. Pour cela, nous introduisons l'approche du maximum de vraisemblance de l'indice des valeurs extrêmes définit en section (2.4.1) et qui fournit la log-vraisemblance de la GPD en (2.4.5). De plus, nous indiquons que l'estimation du maximum de vraisemblance de la GPD n'existe pas pour $\gamma < -1$. Et pour $\gamma = -1$ la log-vraisemblance de la GPD (2.4.5) donne ainsi pour $\sigma > C_{k:k}$:

$$\log(\ell(C_i; (-1, \sigma))) = \sum_{i=1}^k \log\left(\frac{1}{\sigma}\right) \quad (4.2.1)$$

Par conséquent, noté $\hat{\gamma}^{ML} = -1$, $\hat{\sigma}^{ML} = C_{k:k}$. Alors $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ est donnée par le maximum local si $\log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML}))) > M$ et est donnée par le maximum limite si $\log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML}))) < M$ où $M := -\sum_{i=1}^k \log(C_{k:k})$. Si aucun maximum local n'est trouvé, alors il n'y a pas d'estimation du maximum de vraisemblance de la GPD et nous recommandons d'utiliser les estimateurs alternatifs donnés par Hosking et Wallis [21]. Tandis que pour obtenir un maximum fini de la log-vraisemblance GPD (2.4.5), la contrainte $\gamma \geq -1$ doit être imposé. Par conséquent, le calcul de $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ est optimal sur l'espace $A = \{\sigma > 0, \gamma > 0\} \cup \{-1 \leq \gamma < 0, \sigma/\gamma < -C_{k:k}\}$.

La log-vraisemblance de la GPD (2.4.5) est donnée en termes de dérivées partielles comme

définit en (2.4.8) pour $\theta = -(\gamma/\sigma)$ par,

$$\begin{cases} \frac{1}{k} \sum_{i=1}^k \log(1 - \theta(X_{n-i+1,n} - X_{n-k,n})) = \gamma \\ \frac{1}{k} \sum_{i=1}^k \frac{1}{1 - \theta(X_{n-i+1,n} - X_{n-k,n})} = \frac{1}{\gamma + 1} \end{cases} \quad (4.2.2)$$

À partir du système d'équations (4.2.2), nous définissons la fonction suivante :

$$\psi(\theta) = \left(\frac{1}{k} \sum_{i=1}^k \log(1 - \theta C_i) + 1 \right) \cdot \left(\frac{1}{k} \sum_{i=1}^k \frac{1}{1 - \theta C_i} \right) - 1. \quad (4.2.3)$$

pour $C_i = X_{n-i+1,n} - X_{n-k,n}$. Telle que cette fonction est définie en $\mathcal{B} = \{\theta < 1/C_{k:k}, \theta \neq 0\}$ où $C_{k,k} = \max(C_i)$ pour tout $i = 1, \dots, k$. Par conséquent, pour estimer (γ, σ) par le maximum de vraisemblance de la GPD ou applique la procédure suivante :

1. Trouver la racine $\hat{\theta}$ qui détermine par chercher la solution de l'équation $\psi(\theta) = 0$ pour $\theta \in]-\infty; 0[\cup]0; 1/C_{k:k}[$ où,

$$\psi(\theta) = \left(\frac{1}{k} \sum_{i=1}^k \log(1 - \theta C_i) + 1 \right) \cdot \left(\frac{1}{k} \sum_{i=1}^k (1 - \theta C_i)^{-1} \right) - 1$$

2. Calculer $\hat{\gamma}^{ML}$ par

$$\hat{\gamma}^{ML} = \frac{1}{k} \sum_{i=1}^k \log(1 - \hat{\theta} C_i). \quad (4.2.4)$$

3. Pour $\hat{\theta} = -(\hat{\gamma}^{ML}/\hat{\sigma}^{ML})$, alors $\hat{\sigma}^{ML} = -(\hat{\gamma}^{ML}/\hat{\theta})$.

Le théorème suivant énonce les propriétés de ψ qui sont définies en (4.2.3) et qui sont utiles dans la formulation d'un algorithme pour déterminer ses zéros.

Théorème 4.5 *Considérons la fonction $\psi(\theta)$ donnée en (4.2.3) définie sur l'espace \mathcal{B} .*

Alors

1. $\lim_{\theta \rightarrow 0} \psi_k(\theta) = 0$.
2. $\lim_{\theta \rightarrow 1/C_{k:k}^-} \psi(\theta) = -\infty$.
3. $\psi(\theta) < 0$ pour tous $\theta < \theta_* := \frac{2(C_1 - \bar{C})}{(C_1)^2}$ où $\bar{C} = \frac{1}{k} \sum_{i=1}^k C_i$, et $C_1 = \min(C_{i,k})$ pour $C_{1,k} < \dots < C_{k,k}$ la statistique d'ordre associé.

La fonction de vraisemblance est toujours positive (puisqu'il s'agit de la densité conjointe de l'échantillon) mais la fonction de log-vraisemblance est généralement négative (étant le log d'un nombre inférieur à 1). Ici, le log de vraisemblance est globalement concave et a un maximum unique à $\hat{\theta}^{ML}$. Par conséquent, la fonction score est positive à gauche du maximum, passe par zéro au maximum et devient négative à droite du maximum. Intuitivement, la précision de $\hat{\theta}^{ML}$ dépend de la courbure de la fonction log-vraisemblance près de $\hat{\theta}^{ML}$. Si le log-vraisemblance est très courbé ou « raide » autour $\hat{\theta}^{ML}$ alors θ sera précisément estimé. Dans ce cas, on dit qu'on a beaucoup d'informations sur θ . D'autre part, si la log-vraisemblance n'est pas courbée ou "plate" près de $\hat{\theta}^{ML}$ alors θ ne sera pas estimé avec précision.

En général, cependant, le maximum peut ne pas être unique. La fonction log-vraisemblance peut avoir de multiples maxima locaux. Mais il peut être décidé en prenant l'estimation du maximum de la log-vraisemblance. En conséquence, nous disons que nous n'avons pas beaucoup d'informations sur θ . Sans oublier que la courbure de la log-vraisemblance est mesurée par sa dérivée seconde.

Alors pour déterminer la racine de $\psi(\theta) = 0$, nous utilisons l'algorithme développé par [27] appelé MBA pour les multi-racines. MBA pour les multi-racines est basé sur le théorème des valeurs intermédiaires qui marche pour une fonction continue sur l'intervalle $[\alpha; \beta]$ avec la condition initiale $f(\alpha) \cdot f(\beta) < 0$ existant à chaque itération.

Puisque la borne supérieure, $1/C_{k:k} \notin \mathcal{B}$, MBA pour les multi-racines que nous représenterons dans le paragraphe suivant utilisera $1/C_{k:k} - \varepsilon$ et $\theta_* - \varepsilon$ pour $\varepsilon > 0$ comme la borne supérieure et la borne inférieure respectivement où $\theta_* = 2(C_1 - \bar{C}) / (C_1)^2$. Nous Posons

$\varepsilon = 10^{-n}/\overline{C}$ où $n \in \mathbb{N}$, donc \mathcal{B} sera défini par $\mathcal{B} = \{\theta \in [\theta_L; -\varepsilon] \cup [\varepsilon; \theta_U]\}$ où $\theta_U = 1/C_{k:k} - \varepsilon$ et $\theta_L = \theta_* - \varepsilon$.

Les estimateurs du maximum de vraisemblance ont un estimateur consistant de la variance et il l'a utilisé pour remplacer la variance asymptotique des paramètres inconnus. La matrice d'information de Fisher donne la variance-covariance asymptotique de l'estimateur du maximum de vraisemblance, qui peut être calculée par

$$\begin{pmatrix} Var(\widehat{\gamma}^{ML}) & Cov(\widehat{\gamma}^{ML}, \widehat{\sigma}^{ML}) \\ Cov(\widehat{\gamma}^{ML}, \widehat{\sigma}^{ML}) & Var(\widehat{\sigma}^{ML}) \end{pmatrix} := \begin{pmatrix} \frac{\partial^2 \log(\ell(C_i; (\widehat{\gamma}^{ML}, \widehat{\sigma}^{ML})))}{\partial^2 \gamma} & \frac{\partial^2 \log(\ell(C_i; (\widehat{\gamma}^{ML}, \widehat{\sigma}^{ML})))}{\partial \gamma \partial \sigma} \\ \frac{\partial^2 \log(\ell(C_i; (\widehat{\gamma}^{ML}, \widehat{\sigma}^{ML})))}{\partial \gamma \partial \sigma} & \frac{\partial^2 \log(\ell(C_i; (\widehat{\gamma}^{ML}, \widehat{\sigma}^{ML})))}{\partial^2 \sigma} \end{pmatrix}^{-1}$$

Les dérivées du deuxième ordre de la log-vraisemblance sont calculant où $\theta = -(\gamma/\sigma)$ par

$$\begin{aligned} \frac{\partial^2 \log(\ell(C_i; (\gamma, \sigma)))}{\partial^2 \gamma} &:= \frac{1}{\gamma^3} \sum_{i=1}^k \left[(1 + \gamma) \left(1 - \frac{1}{1 - \theta C_i}\right)^2 - 2 \log(1 - \theta C_i) + 2 \left(1 - \frac{1}{1 - \theta C_i}\right) \right], \\ \frac{\partial^2 \log(\ell(C_i; (\gamma, \sigma)))}{\partial^2 \sigma} &:= \left(\frac{\theta}{\gamma}\right)^2 \sum_{i=1}^k \left[1 - \left(1 + \frac{1}{\gamma}\right) \left(2 \left(1 - \frac{1}{1 - \theta C_i}\right) - \left(1 - \frac{1}{1 - \theta C_i}\right)^2\right) \right], \\ \frac{\partial^2 \log(\ell(C_i; (\gamma, \sigma)))}{\partial \gamma \partial \sigma} &:= \frac{\theta}{\gamma^2} \sum_{i=1}^k \left[\left(1 + \frac{1}{\gamma}\right) \left(1 - \frac{1}{1 - \theta C_i}\right)^2 - \left(1 - \frac{1}{1 - \theta C_i}\right) \right]. \end{aligned}$$

Alors, nous la construction des intervalles de confiance :

$$\widehat{\gamma}^{ML} - t_{\kappa/2} \sqrt{Var(\widehat{\gamma}^{ML})} \text{ et } \widehat{\sigma}^{ML} - t_{\kappa/2} \sqrt{Var(\widehat{\sigma}^{ML})}$$

avec $100(1 - \kappa)\%$ comme suit de confiance.

Ensuit, Nous préférons un algorithme qui calcule $(\widehat{\gamma}^{ML}, \widehat{\sigma}^{ML})$ de la façon suivante:

1. Choisissons un ε , par exemple $\varepsilon = 10^{-8}/\overline{C}$, où $\overline{C} = \frac{1}{k} \sum_{i=1}^k C_i$

2. Calculer θ_L la borne inférieure (resp θ_U la borne supérieure) par

$$\theta_L = \frac{2(C_{1,k} - \bar{C})}{(C_{1,k})^2}; \theta_U = \frac{1}{C_{k:k}} - \varepsilon.$$

où $C_{1,k} = \min(C_{i,k})$ et $C_{k,k} = \min(C_{i,k})$ pour $C_{1,k} < \dots < C_{k,k}$ la statistique d'ordre associée.

3. Pour trouver la racine de $\psi(\theta) = 0$ où ψ est défini en relation (4.2.3) sur les deux intervalles $[\theta_L; -\varepsilon]$ et $[\varepsilon; \theta_U]$ utilisé MBA pour les multi-racines avec chaque intervalle. Notons $\theta_s^{(0)}$ où s est le nombre de racine.

4. Pour chaque valeur de $\theta_s^{(0)}$ obtenu de l'étape précédente calculer la valeur de γ_s donné par $\gamma_s = \frac{1}{k} \sum_{i=1}^k \log(1 - \theta_s^{(0)} C_i)$. Et calculer la valeur de $\sigma_s = -(\gamma_s / \theta_s^{(0)})$.

5. Soit γ_s et σ_s désignent les résultats de l'étape précédente qui appartiennent à l'espace A et on trouve qui un maximum local existe, Alors l'estimateur du maximum de vraisemblance des paramètres de la GPD est $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ ont un maximum $\log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML})))$ définit en (2.4.5).

6. Pour $(\hat{\gamma}^{ML}, \hat{\sigma}^{ML})$ calculer tous les estimateurs des dérivées partielles du second ordre: $\frac{\partial^2 \log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML})))}{\partial^2 \gamma}$ et $\frac{\partial^2 \log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML})))}{\partial^2 \sigma}$, $\frac{\partial^2 \log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML})))}{\partial \gamma \partial \sigma}$ et le déterminant jacobien des dérivées du deuxième ordre: $\frac{\partial^2 \log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML})))}{\partial^2 \gamma} \times \frac{\partial^2 \log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML})))}{\partial^2 \sigma} - \left(\frac{\partial^2 \log(\ell(C_i; (\hat{\gamma}^{ML}, \hat{\sigma}^{ML})))}{\partial \gamma \partial \sigma} \right)^2$

7. Calculer les intervalles de confiance de (γ, σ) , en utilisant $\hat{\gamma} \pm t_{\kappa/2} \sqrt{Var(\hat{\gamma})}$ et $\hat{\sigma} \pm t_{\kappa/2} \sqrt{Var(\hat{\sigma})}$ respectivement avec une seuil de confiance $100(1 - \kappa)\%$ où $t_{\kappa/2}$ désigne le quantile de la distribution normale standard est symétrique par rapport à 0.

8. L'algorithme est terminé.

Nous rappelons que cette méthode est une méthode numérique basée sur un algorithme afin d'estimer l'indice des valeurs extrêmes par maximum de vraisemblance de la GPD par lequel on obtient les caractéristiques liées à toute estimation comme la normalité asymptotique et l'intervalle de confiance. Sans oublier que les estimateurs du maximum

de vraisemblance sont des estimateurs consistant de la variance.

4.3 MBA pour estimer l'indice de queue sous données censurés

Dans cette sous section on va présenter le travail [26] adapté sous les données censurées. Il est basé sur l'estimation du maximum de vraisemblance de la GPD de $(\hat{\gamma}_1^{(c,ML)}, \hat{\sigma}_1^{(c,ML)})$ en présence de censure, sont les valeurs qui maximisent la fonction de log-vraisemblance de la GPD avec les données censurées. Il existe de nombreuses techniques numériques pour calculer $(\hat{\gamma}_1^{(c,ML)}, \hat{\sigma}_1^{(c,ML)})$ proposés par Pham et al [32] pour $\gamma_1 \leq 1$ et Kouider et al [28] pour $\gamma_1 \geq -1$.

Kouider et al [28] ont adaptés la méthode MLE de GPD pour estimer (γ_1, σ_1) mais en mieux condition sur les données censurées. Nous utilisons cette méthode pour estimer l'indice de queue en présence de censure pour $\gamma_1 \geq -1$. Pour cela, nous introduisons l'approche du maximum de vraisemblance en présence de censure de l'indice des valeurs extrêmes appeler l'estimateur de maximum de vraisemblance adapter (ML-estimator adapt : Maximum Likelihood estimator adapt) qui utilise les deux paramètres (γ_1, σ_1) de la GPD avec censure aléatoire de type I.

Soit X_1, \dots, X_n de v.a's i.i.d suit la même distribution F avec $F \in D(H_{\gamma_1})$ censurés à droite par un échantillon Y_1, \dots, Y_n de v.a's i.i.d suit la même distribution G avec $G \in D(H_{\gamma_2})$ où X_i et Y_i sont indépendantes. Nous observons $Z_i = \min(X_i, Y_i)$ et on observe également l'indicateur de censure, $\delta_{X_i \leq Y_i}$ pour tout $i = 1, \dots, n$ où les $(Z_n)_{n \geq 1}$ sont une suite de v.a's i.i.d avec la même distribution H où $H \in D(H_\gamma)$ où $\gamma := \gamma_1 \gamma_2 / \gamma_2 + \gamma_1$. Considérons $Z_{1,n} \leq \dots \leq Z_{n,n}$ la statistique d'ordre de l'échantillon Z_1, \dots, Z_n . Soit $\delta_{[i,n]}$ le concomitant de $Z_{i,n}$ ou les indicateurs de censure (c'est-à-dire, $\delta_{[i,n]} = \delta_j$ si $Z_{i,n} = Z_j$ avec $j = 1, \dots, n$). Considérons un couple (Z_j, δ_j) pour tout $j = 1, \dots, n$. Pour estimer γ_1 en utilisant l'approche de la vraisemblance conditionnelle en présence de censure, l'approche

de la vraisemblance conditionnelle selon le travaux de Balkema et de Haan [2] et Pickands [38], indique que sous (2.3.4) qu'on peut approximer la GPD comme distribution des excès pour $C_j = Z_j - t$ pour $Z_j > t$ pour tout $1 \leq j \leq k$ quand $t \rightarrow x_H$ où x_H le point terminal de H avec $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$. Grâce à Andersen et al [1], le maximum de vraisemblance avec les données censurées consiste en maximisant par rapport à γ_1 et σ_1 la fonction du maximum de vraisemblance adapté avec la GPD qui est donné par:

$$\ell(C_j; (\gamma_1, \sigma_1); \delta_j) = \prod_{j=1}^k (f_{\gamma_1, \sigma_1}(C_j))^{\delta_j} (1 - F_{\gamma_1, \sigma_1}(C_j))^{1-\delta_j}, \quad (4.3.1)$$

avec $C_j = Z_j - Z_{n-k,n}$ pour tout $j = 1, \dots, k$ où F_{γ_1, σ_1} est la GPD (2.3.3) et $f_{\gamma, \sigma}(C_j)$ la fonction de densité associée (2.4.4).

Alors l'estimation du maximum de vraisemblance de γ_1 et σ_1 avec des données censurées est obtenue par la méthode du maximum l'équation de log-vraisemblance du (4.3.1) suivant:

$$\log(\ell(C_j; (\gamma_1, \sigma_1); \delta_j)) = \sum_{j=1}^k \left(\delta_j \log\left(\frac{1}{\sigma_1}\right) - \left(\delta_j + \frac{1}{\gamma_1}\right) \log\left(1 + \frac{\gamma_1}{\sigma_1} C_j\right) \right) \quad (4.3.2)$$

Alors, si $\gamma_1 < -1$ on trouve

$$\lim_{\substack{\sigma_1 \rightarrow C_{k:k}^+ \\ \gamma_1}} \log(\ell(C_j; (\gamma_1, \sigma_1); \delta_j)) := \begin{cases} +\infty & \text{si } \delta_j = 1 \\ -\infty & \text{si } \delta_j = 0 \end{cases}$$

donc pour tout $\gamma < -1$, il n'y a pas d'estimation du maximum de vraisemblance avec données censurées. Les équations de log-vraisemblance du (4.3.2) sont définit en termes de dérivées partielles par rapport à γ_1 (σ_1 respect) pour $\gamma_1 \neq 0$ par

$$\begin{aligned} \frac{\partial \log(\ell(C_j; (\gamma_1, \sigma_1); \delta_j))}{\partial \gamma_1} &= \sum_{j=1}^k \left(\frac{1}{\gamma_1} \right)^2 \log\left(1 + \frac{\gamma_1}{\sigma_1} C_j\right) - \left(\delta_j + \frac{1}{\gamma_1}\right) \frac{\frac{C_j}{\sigma_1}}{1 + \frac{\gamma_1}{\sigma_1} C_j} \quad (4.3.3) \\ \frac{\partial \log(\ell(C_j; (\gamma_1, \sigma_1); \delta_j))}{\partial \sigma_1} &= \sum_{j=1}^k -\frac{\delta_j}{\sigma_1} + \left(\delta_j + \frac{1}{\gamma_1}\right) \frac{\frac{\gamma_1}{\sigma_1^2} C_j}{1 + \frac{\gamma_1}{\sigma_1} C_j} \end{aligned}$$

Et pour $\gamma_1 = 0$ ces termes doivent être interprétés par le développement de Taylor du système (4.3.3), comme suivant:

$$\begin{aligned} \frac{\partial \log(\ell(C_j; (\gamma_1, \sigma_1); \delta_j))}{\partial \gamma_1} \Big|_{\gamma_1=0} &= \frac{1}{2} \sum_{j=1}^k \left(\frac{C_j}{\sigma_1} \right)^2 - \sum_{j=1}^k \frac{C_j}{\sigma_1} \\ \frac{\partial \log(\ell(C_j; (\gamma_1, \sigma_1); \delta_j))}{\partial \sigma_1} \Big|_{\gamma_1=0} &= - \sum_{j=1}^k \frac{1}{\sigma_1} + \sum_{j=1}^k \left(\frac{C_j}{\sigma_1} \right)^2 \end{aligned} \quad (4.3.4)$$

À partir de (4.3.3) les équations de vraisemblance pour $\gamma_1 \neq 0$ sont données par :

$$\left\{ \begin{array}{l} \sum_{j=1}^k \left[\log \left(1 + \frac{\gamma_1 C_j}{\sigma_1} \right) - (\gamma_1 \delta_j + 1) \frac{\frac{\gamma_1 C_j}{\sigma_1}}{1 + \frac{\gamma_1 C_j}{\sigma_1}} \right] = 0 \\ \sum_{j=1}^k (\gamma_1 \delta_j + 1) \frac{\frac{\gamma_1 C_j}{\sigma_1}}{1 + \frac{\gamma_1 C_j}{\sigma_1}} = \gamma_1 \sum_{j=1}^k \delta_j \end{array} \right. = 0 \quad (4.3.5)$$

où $r = \sum_{j=1}^k \delta_j$. Puis, on peut simplifier le résultat (4.3.5) par :

$$\left\{ \begin{array}{l} \gamma_1 = \frac{1}{r} \sum_{j=1}^k \log \left(1 + \frac{\gamma_1 C_j}{\sigma_1} \right) \\ \frac{1}{k} \sum_{j=1}^k (\gamma_1 \delta_j + 1) \frac{1}{1 + \frac{\gamma_1 C_j}{\sigma_1}} = 1 \end{array} \right. \quad (4.3.6)$$

Soit l'échantillon C_1, C_2, \dots, C_k de v.a's i.i.d telles que $k = r + m$ où $r := \sum_{j=1}^k \mathbf{1}_{\{\delta_j=1\}}$ représente le nombre des observation et m représente le nombre des données censurées dans l'échantillon C_i , c'est à dire le nombre où $\delta_j = 0$. Considérons $C_{1,k} \leq \dots \leq C_{k,k}$ la statistique d'ordre de l'échantillon C_1, \dots, C_k . Soit $\delta_{1,k} \leq \dots \leq \delta_{k,k}$ la statistique d'ordre des indicateurs de censure associée à C_1, \dots, C_k . Donc la relation (4.3.2) est réécrit comme

suit

$$\log(\ell(C_j; (\gamma_1, \sigma_1))) = r \log\left(\frac{1}{\sigma_1}\right) - \sum_{j=1}^r \log\left(1 + \frac{\gamma_1}{\sigma_1} C_{j,r}\right) - \frac{1}{\gamma_1} \sum_{j=1}^k \log\left(1 + \frac{\gamma}{\sigma} C_{j,k}\right) \quad (4.3.7)$$

où $\sigma_1 > 0$ pour $\gamma_1 > 0$ et $\sigma_1 > -\gamma_1 C_{k,k}$ pour $\gamma_1 \leq 0$ avec $C_{k,k} = \max(C_j)$ pour tout $j = 1, \dots, k$. Nous indiquons que l'estimation du maximum de vraisemblance de la GPD n'existe pas pour $\gamma < -1$. Et pour $\gamma_1 = -1$ pour $\sigma_1 > C_{k:k}$ on obtient :

$$M := \log(\ell(C_j; (-1, \sigma_1))) = r \log\left(\frac{1}{\sigma_1}\right) + \sum_{j=1}^m \log\left(1 - \frac{1}{\sigma_1} C_{j,m}\right) \quad (4.3.8)$$

pour $m = n - r$. Alors, $\hat{\gamma}_1^{(c,ML)} = -1$, $\hat{\sigma}_1^{(c,ML)} = C_{k:k}$. De plus, si $C_{k:k}$ n'est pas observé i.e $C_{k:k} := C_{m:m}$ on trouve que M n'existe pas. Donc

$$M_1 \approx r \log\left(\frac{1}{\sigma_1}\right) + m \log\left(1 - \frac{\bar{C}'}{\sigma_1}\right) \text{ où } \bar{C}' = \frac{1}{m} \sum_{j=1}^m C_{j,m} \quad (4.3.9)$$

Par contre si $C_{k:k}$ est observé i.e. $C_{k:k} := C_{r,r}$ on trouve que M existe.

Par conséquent, la contrainte $\gamma_1 \geq -1$ doit être imposé pour l'existence de $(\hat{\gamma}_1^{(c,ML)}, \hat{\sigma}_1^{(c,ML)})$ qui est déterminé par l'un des trois cas suivant :

1- Si $C_{k:k}$ n'est pas observé: alors $(\hat{\gamma}_1^{(c,ML)}, \hat{\sigma}_1^{(c,ML)})$ est défini par le maximum local si $\tilde{M} > M_1|_{\hat{\sigma}_1^{(c,ML)}=C_{k:k}}$ et il est donnée par le maximum limite si $\tilde{M} < M_1|_{\hat{\sigma}_1^{(c,ML)}=C_{k:k}}$ où $\tilde{M} := \log\left(\ell\left(C_j; \left(\hat{\gamma}_1^{(c,ML)}, \hat{\sigma}_1^{(c,ML)}\right)\right)\right)$.

2- Si $C_{k:k}$ est observé: $(\hat{\gamma}_1^{(c,ML)}, \hat{\sigma}_1^{(c,ML)})$ est défini par le maximum local si $\tilde{M} > M|_{\hat{\sigma}_1^{(c,ML)}=C_{k:k}}$ et il est donnée par le maximum limite si $\tilde{M} < M|_{\hat{\sigma}_1^{(c,ML)}=C_{k:k}}$.

3- Si aucun maximum local n'est trouvé, alors il n'y a pas d'estimation du maximum de vraisemblance GPD sous censure aléatoire. Dans cette situation, nous pouvons utiliser l'estimateur adaptatif en utilisant les estimateurs alternatifs donnés par Hosking et Wallis [21] sans censure, pour obtenir un maximum fini de la log-vraisemblance GPD (4.3.7) sous données censurées.

De plus, le calcul de $(\widehat{\gamma}_1^{(c,ML)}, \widehat{\sigma}_1^{(c,ML)})$ est optimal sur l'espace:

$$A = \{\sigma_1 > 0, \gamma_1 > 0\} \cup \{-1 \leq \gamma_1 < 0, \sigma_1/\gamma_1 < -C_{k,k}\}. \quad (4.3.10)$$

Donc pour déterminer $(\widehat{\gamma}_1^{(c,ML)}, \widehat{\sigma}_1^{(c,ML)})$ il y a deux valeurs de (γ_1, σ_1) qui doivent être calculer en même temps. Le premier, est le maximum local de la log-vraisemblance sur l'espace A . Et le deuxième est la limite de A avec $\gamma_1 = -1$.

À partir du système d'équations (4.3.6), nous définissons la fonction suivante :

$$\widetilde{\psi}(\theta_1) = \frac{1}{k} \sum_{j=1}^k \left[\left(\delta_j \left(\frac{1}{r} \sum_{j=1}^k \log(1 - \theta_1 C_j) \right) + 1 \right) \cdot \left(\frac{1}{1 - \theta_1 C_j} \right) \right] - 1 = 0. \text{ où } \theta_1 = -\frac{\gamma_1}{\sigma_1}$$

Telleque $\widetilde{\psi}(\theta_1)$ est définit sur $\widetilde{\mathcal{B}} = \{\theta_1 < 1/C_{k:k}, \theta_1 \neq 0\}$ où $C_{k,k} = \max(C_{j,k})$ pour tout $i = 1, \dots, k$.

Par conséquent, pour estimer (γ_1, σ_1) le maximum de vraisemblance de la GPD avec des données censurées doit être utilisé selon dans la procédure suivante :

1. Trouver la racine $\widehat{\theta}_1^{c,ML}$ de l'équation $\widetilde{\psi}(\theta_1) = 0$ pour $\theta_1 \in]-\infty; 0[\cup]0; 1/C_{k:k}[$ où,

$$\widetilde{\psi}(\theta_1) = \left(\frac{1}{r} \sum_{j=1}^k \log(1 - \theta_1 C_{j,k}) \right) \left(\frac{1}{k} \sum_{j=1}^r \left(\frac{1}{1 - \theta_1 C_{j,r}} \right) \right) + \left(\frac{1}{k} \sum_{j=1}^k \left(\frac{1}{1 - \theta_1 C_{j,k}} \right) \right) - 1 \quad (4.3.11)$$

2. Calculer $\widehat{\gamma}_1^{c,ML}$ par

$$\widehat{\gamma}_1^{c,ML} = \frac{1}{r} \sum_{j=1}^k \log(1 - \widehat{\theta}_1^{c,ML} C_j). \quad (4.3.12)$$

3. Pour $\widehat{\theta}_1^{c,ML} = -\left(\widehat{\gamma}_1^{c,ML}/\widehat{\sigma}_1^{c,ML}\right)$, alors $\widehat{\sigma}_1^{c,ML} = -\left(\widehat{\gamma}_1^{c,ML}/\widehat{\theta}_1^{c,ML}\right)$.

En remarque que si tous les $\delta_{[i,n]} = 1$ on trouve $C_j = X_j - t$. De plus on ne peut pas définir $\widehat{\gamma}_1^{c,ML}$ du (4.3.12) comme un estimateur du maximum de vraisemblance adapter parce que

la racine de $\widetilde{\psi}(\theta_1)$ est déférente de $\psi(\theta)$, donc on trouve que $\widehat{\theta}_1^{c,ML} \neq \theta$ ou $\theta = -(\gamma/\sigma)$.

Par contre si $\widehat{\theta}_1^{c,ML} = \theta$ en trouve que $\widehat{\gamma}_1^{c,ML}$ du (4.3.12) définit bien un estimateur adapter.

Mais dans ce cas l'estimateur du maximum de vraisemblance adapter n'est pas définit pour

cela on va remplacer $(\widehat{\gamma}_1^{c,ML}, \widehat{\sigma}_1^{c,ML}, \widehat{\theta}_1^{c,ML})$ par $(\widehat{\gamma}_1^{c,KIB}, \widehat{\sigma}_1^{c,KIB}, \widehat{\theta}_1^{c,KIB})$ respectivement.

On peut donc reformuler la procédure pour estimer (γ_1, σ_1) par le maximum de vraisemblance de la GPD sous les données censurées. Il est décrit dans les étapes suivantes:

1. Résoudre l'équation $\psi(\widehat{\theta}_1^{c,KIB}) = 0$ qui donnée (4.3.11) où $\theta_1 \in]-\infty; 0[\cup]0; 1/C_{k:k}[$.
2. Calculer $\widehat{\gamma}_1^{c,KIB}$ par la formule :

$$\widehat{\gamma}_1^{c,KIB} = \frac{1}{r} \sum_{j=1}^k \log(1 - \widehat{\theta}_1^{c,KIB} C_j). \quad (4.3.13)$$

3. Pour $\widehat{\theta}_1^{c,KIB} = -(\widehat{\gamma}_1^{c,KIB} / \widehat{\sigma}_1^{c,KIB})$, alors $\widehat{\sigma}_1^{c,KIB} = -(\widehat{\gamma}_1^{c,KIB} / \widehat{\theta}_1^{c,KIB})$.

Le théorème suivant présente quelques propriétés utiles pour pouvoir appliquer la méthode numérique vu dans la sub-section (4.1.3) pour déterminer les racines de la fonction $\widetilde{\psi}$ défini en (4.3.11)

Théorème 4.6 *Soit la fonction $\widetilde{\psi}(\theta_1)$ donne en (4.3.11) défini sur l'espace $\widetilde{\mathcal{B}}$. Alors*

1. $\lim_{\theta_1 \rightarrow 0} \widetilde{\psi}(\theta_1) = 0$.
2. $\lim_{\theta_1 \rightarrow 1/C_{k:k}^-} \widetilde{\psi}(\theta_1) = -\infty$.
3. $\widetilde{\psi}(\theta_1) < 0$ pour tous $\theta_1 < (\theta_1)_* := \frac{2(C_{1,k} - \overline{C})}{(C_{1,k})^2}$. où $\overline{C} = \frac{1}{k} \sum_{j=1}^k C_j$, et $C_1 = \min(C_{j,k})$ pour $C_{1,k} < \dots < C_{k,k}$ la statistique d'ordre associée puis, $1/C_{k:k} \notin \widetilde{\mathcal{B}}$

La méthode numérique que nous utilisons, utilise $1/C_{k:k} - \varepsilon$ et $(\theta_1)_* - \varepsilon$ pour $\varepsilon > 0$ comme borne supérieure et inférieure respectivement où $(\theta_1)_* = 2(C_{1,k} - \overline{C}) / (C_{1,k})^2$. Posons $\varepsilon = 10^{-n} / \overline{C}$ où $n \in \mathbb{N}$, donc $\widetilde{\mathcal{B}}$ sera défini par $\widetilde{\mathcal{B}} = \{\theta_1 \in [(\theta_1)_L; -\varepsilon] \cup [\varepsilon; (\theta_1)_U]\}$ où $(\theta_1)_U = 1/C_{k:k} - \varepsilon$ et $(\theta_1)_L = (\theta_1)_* - \varepsilon$.

Alors pour déterminer la racine de $\widetilde{\psi}(\theta_1) = 0$, nous utilisons l'algorithme développé par [27] appelé algorithme de bisection modifié pour les multi-racines présenté en (4.1.3). La matrice d'information de Fisher donne la variance-covariance asymptotique des estima-

teurs du maximum de vraisemblance $(\widehat{\gamma}_1^{c,KIB}, \widehat{\sigma}_1^{c,KIB})$ en présence de censure, par :

$$\begin{pmatrix} Var(\widehat{\gamma}_1^{c,KIB}) & Cov(\widehat{\gamma}_1^{c,KIB}, \widehat{\sigma}_1^{c,KIB}) \\ Cov(\widehat{\gamma}_1^{c,KIB}, \widehat{\sigma}_1^{c,KIB}) & Var(\widehat{\sigma}_1^{c,KIB}) \end{pmatrix} := \begin{pmatrix} \frac{\partial^2 \widetilde{M}}{\partial^2 \gamma} & \frac{\partial^2 \widetilde{M}}{\partial \gamma \partial \sigma} \\ \frac{\partial^2 \widetilde{M}}{\partial \gamma \partial \sigma} & \frac{\partial^2 \widetilde{M}}{\partial^2 \sigma} \end{pmatrix}^{-1}$$

où $\widetilde{M} := \log(\ell(C_j; (\widehat{\gamma}_1^{c,KIB}, \widehat{\sigma}_1^{c,KIB})))$. Les dérivées du deuxième ordre de la log-vraisemblance en présence de censure sont calculant avec $\theta_1 = -(\gamma_1/\sigma_1)$ par

$$\begin{cases} \frac{\partial^2 \log(\ell(C_j; (\gamma_1, \sigma_1), \delta_j))}{\partial^2 \gamma_1} := \frac{1}{\gamma_1^3} \sum_{j=1}^k \left[(1 + \gamma_1 \delta_j) \left(1 - \frac{1}{1 - \theta_1 C_j}\right)^2 - 2 \log(1 - \theta_1 C_j) + 2 \left(1 - \frac{1}{1 - \theta_1 C_j}\right) \right], \\ \frac{\partial^2 \log(\ell(C_j; (\gamma_1, \sigma_1), \delta_j))}{\partial^2 \sigma_1} := \left(\frac{\theta_1}{\gamma_1}\right)^2 \sum_{j=1}^k \left[\delta_j - \left(\delta_j + \frac{1}{\gamma_1}\right) \left(2 \left(1 - \frac{1}{1 - \theta_1 C_j}\right) - \left(1 - \frac{1}{1 - \theta_1 C_j}\right)^2\right) \right], \\ \frac{\partial^2 \log(\ell(C_j; (\gamma_1, \sigma_1), \delta_j))}{\partial \gamma_1 \partial \sigma_1} := \frac{\theta_1}{\gamma_1^2} \sum_{j=1}^k \left[\left(\delta_j + \frac{1}{\gamma_1}\right) \left(1 - \frac{1}{1 - \theta_1 C_j}\right)^2 - \delta_j \left(1 - \frac{1}{1 - \theta_1 C_j}\right) \right]. \end{cases}$$

L'intervalle de confiance au seuil $95(1 - \kappa)\%$ par γ_1 et σ_1 est :

$$\widehat{\gamma}_1^{c,KIB} \pm t_{\kappa/2} \sqrt{Var(\widehat{\gamma}_1^{c,KIB})} \text{ et } \widehat{\sigma}_1^{c,KIB} \pm t_{\kappa/2} \sqrt{Var(\widehat{\sigma}_1^{c,KIB})}$$

L'algorithme suivant est préférable au précédent pour calculer $(\widehat{\gamma}_1^{c,KIB}, \widehat{\sigma}_1^{c,KIB})$:

1. Choisir une valeur pour ε , proche de zéro, par exemple $\varepsilon = 10^{-8}/\overline{C}$, où $\overline{C} = \frac{1}{k} \sum_{i=1}^k C_i$
2. Calculer $(\theta_1)_L$ la borne inférieure (resp $(\theta_1)_U$ la borne supérieure) par

$$(\theta_1)_L = \frac{2(C_{1,k} - \overline{C})}{(C_{1,k})^2}; (\theta_1)_U = \frac{1}{C_{k:k}} - \varepsilon.$$

où $C_{1,k} = \min(C_{j,k})$ et $C_{k,k} = \min(C_{j,k})$ pour $C_{1,k} < \dots < C_{k,k}$ la statistique d'ordre associée.

3. Pour trouver la racine de $\widetilde{\psi}(\theta_1) = 0$ où ψ est défini en (4.3.11) sur les deux intervalles $[(\theta_1)_L; -\varepsilon]$ et $[\varepsilon; (\theta_1)_U]$ on doit utiliser la MBA pour les multi-racines donnée dans le paragraphe (4.1.3) pour chaque intervalle. Notons $(\theta_1)_s^{(0)}$ où s est le nombre de racine.

4. Pour chaque valeur de $(\theta_1)_s^{(0)}$ obtenue de l'étape précédente, calculer $(\gamma_1)_s$ par $(\gamma_1)_s = \frac{1}{r} \sum_{j=1}^k \log\left(1 - (\theta_1)_s^{(0)} C_j\right)$ où $r := \sum_{j=1}^k \mathbf{1}_{\{\delta_j=1\}}$, et calculer $(\sigma_1)_s$ par $(\sigma_1)_s = -\frac{(\gamma_1)_s}{(\theta_1)_s^{(0)}}$.

5. Soit γ_s et σ_s les résultats de l'étape précédente qui appartiennent à l'espace A et si un maximum local existe, alors l'estimateur du maximum de vraisemblance des paramètres de la GPD est $\left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)$ dont le maximum de $\log\left(\ell\left(C_j; \left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)\right)\right)$ est défini en (4.3.7).

6. Pour $\left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)$ calculer tous les estimateurs des dérivées partielles du second ordre:

$$\frac{\partial^2 \log\left(\ell\left(C_j; \left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)\right)\right)}{\partial^2 \widehat{\gamma}_1^{(c,KIB)}} \text{ et } \frac{\partial^2 \log\left(\ell\left(C_j; \left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)\right)\right)}{\partial^2 \widehat{\sigma}_1^{(c,KIB)}}, \frac{\partial^2 \log\left(\ell\left(C_j; \left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)\right)\right)}{\partial \widehat{\gamma}_1^{(c,KIB)} \partial \widehat{\sigma}_1^{(c,KIB)}} \text{ et}$$

$$\text{le jacobien des dérivées du deuxième ordre: } \frac{\partial^2 \log\left(\ell\left(C_j; \left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)\right)\right)}{\partial^2 \widehat{\gamma}_1^{(c,KIB)}} \times \frac{\partial^2 \log\left(\ell\left(C_j; \left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)\right)\right)}{\partial^2 \widehat{\sigma}_1^{(c,KIB)}} -$$

$$\left(\frac{\partial^2 \log\left(\ell\left(C_j; \left(\widehat{\gamma}_1^{(c,KIB)}, \widehat{\sigma}_1^{(c,KIB)}\right)\right)\right)}{\partial \widehat{\gamma}_1^{(c,KIB)} \partial \widehat{\sigma}_1^{(c,KIB)}}\right)^2$$

7. Calculer à $100(1 - \kappa)\%$ les intervalles de confiance pour (γ_1, σ_1) , en utilisant $\widehat{\gamma}_1^{(c,KIB)} -$
 $t_{\kappa/2} \sqrt{Var\left(\widehat{\gamma}_1^{(c,KIB)}\right)}$ et $\widehat{\sigma}_1^{(c,KIB)} -$
 $t_{\kappa/2} \sqrt{Var\left(\widehat{\sigma}_1^{(c,KIB)}\right)}$ respectivement où $t_{\kappa/2}$ désigne le quantile de la distribution normale standard est symétrique par rapport à 0

8. L'algorithme est terminé.

Nous présentons l'algorithme pour estimer l'indice de queue en présence de censure qui se base sur l'algorithme donnée par Kouider et al [28] avec un calcul plus rapide. En déduit que si $\theta_1 := \theta$ dans la définition de $\widehat{\gamma}_1^{(c,KIB)}$ (4.3.13) alors il est défini dans (4.3.12) l'estimateur du maximum de vraisemblance adapter en présence de censure comme nous défini dans la relation générale présentée en (3.2.3) où $\widehat{\gamma}^{ML}$ est défini en (4.2.4) et c'est ce que Einmahl et al [11] a évoqué. De plus, sous la relation (3.2.13) de non censurée on prend $p = 1$ avec $\gamma_1 = \gamma$ et $\alpha_1 = 0$. Et pour $\theta_1 := \theta$ on trouve que $\widehat{\gamma}^{KIB} = \widehat{\gamma}^{ML}$. Aussi Kouider et al [29] présentons la normalité asymptotique du KIB-estimateur $\widehat{\gamma}^{KIB}$ basé sur l'échantillon Z_1, \dots, Z_n par :

$$\sqrt{k} \left(\widehat{\gamma}^{KIB} - \gamma\right) \xrightarrow{d} N\left(\frac{\lambda \widetilde{\alpha}(\gamma + 1)}{(1 - \rho)(1 + \gamma - \rho)}; (\gamma + 1)^2\right). \quad (4.3.14)$$

où $\widetilde{\alpha}$ est dans (2.4.30). Par contre, si $\theta_1 \neq \theta$ on trouve que $\widehat{\gamma}_1^{(c,KIB)}$ n'est pas adapter pour les données censurées et la relation (3.2.3) n'est plus fiable.

Chapitre 5

Appendix

5.1 Appendix A

5.1.1 Propriétés des estimateurs du maximum de vraisemblance

La méthode la plus courante pour estimer les paramètres d'un modèle paramétrique est la méthode du maximum de vraisemblance. Soit X_1, X_2, \dots, X_n des v.a.'s i.i.d de même distribution $F(x; \theta)$ avec PDF $f(x; \theta)$. La fonction de vraisemblance est définie en (2.4.3) par :

$$\ell(X_i; \theta) = \prod_{i=1}^n f(X_i; \theta), \quad (5.A.1)$$

Parfois, nous utilisons la fonction log-vraisemblance est définie par $\log(\ell(X_i; \theta))$. L'estimateur du maximum de vraisemblance MLE, noté $\hat{\theta}^{ML}$ est la valeur de θ qui maximise (5.A.1).

Le maximum de La fonction log-vraisemblance se produit au même endroit que le maximum de $\ell(\theta)$, donc maximiser la log-vraisemblance conduit à la même réponse que maximiser la vraisemblance. Souvent, il est plus facile de travailler avec la log-vraisemblance. Sous certaines conditions du modèle, l'estimateur du maximum de vraisemblance $\hat{\theta}^{ML}$ possède de nombreuses propriétés qui permettent un choix attrayant d'estimateurs. Les principales propriétés du MLE sont :

1. Le MLE est **cohérent** : $\hat{\theta}^{ML} \xrightarrow{P} \theta$ où θ désigne la vraie valeur.
2. Le MLE est **équivariant** : si $\hat{\theta}^{ML}$ est le MLE de θ alors $g(\hat{\theta}^{ML})$ est le MLE de $g(\theta)$.
3. Le MLE est asymptotiquement Normal.
4. Le MLE est **asymptotiquement optimal** ou **efficace** : à peu près, cela signifie que parmi tous les estimateurs, la MLE a la plus petite variance, du moins pour les grands échantillons.
5. La MLE est approximativement l'estimateur de Bayes.

Ce sont essentiellement des conditions de régularité sur la PDF $f(x; \theta)$. Sauf indication contraire, nous supposons tacitement que ces conditions sont remplies.

Cohérence de MLE

La cohérence signifie que le MLE converge en probabilité vers la vraie valeur. Pour continuer, nous avons besoin d'une définition. Si f et g sont des PDF's, définissez la distance de Kullback-Leibler entre f et g comme

$$D(f, g) = \int f(x) \log \left(\frac{f(x)}{g(x)} \right) dx \quad (5.A.2)$$

On peut montrer que $D(f, g) \geq 0$ et $D(f, f) = 0$. Notons que, Ce n'est pas une distance au sens formel car $D(f, g)$ n'est pas symétrique. Pour toute θ, λ Nous écrivons $D(\theta, \lambda)$ au lieu de $D(f(x; \theta), f(x; \lambda))$ avec $\theta \neq \lambda$. Cela signifie que différentes valeurs du paramètre correspondent à différentes distributions. Nous supposons désormais que le modèle est **identifiable**. Soit θ_* indique la vraie valeur de θ Maximiser $\log(\ell(X_i; \theta))$ où $\ell(X_i; \theta)$ est donnée dans (5.A.1) équivaut à maximiser

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log \left(\frac{f(X_i; \theta)}{f(X_i; \theta_*)} \right) \quad (5.A.3)$$

Cela suit depuis $M_n(\theta) = n^{-1} (\log(\ell(X_i; \theta)) - \log(\ell(X_i; \theta_*)))$. Par la loi des grands nombres, $M_n(\theta)$ converge vers

$$E_{\theta_*} \left(\log \left(\frac{f(X_i; \theta)}{f(X_i; \theta_*)} \right) \right) = -D(\theta_*, \theta)$$

Ainsi, $M_n(\theta) \approx -D(\theta_*, \theta)$ qui est maximisée à θ_* comme $-D(\theta_*, \theta) = 0$ et $-D(\theta_*, \theta) < 0$ pour $\theta \neq \theta_*$. Par conséquent, nous nous attendons à ce que le maximiseur ait tendance à θ_* . Pour le prouver formellement, il nous faut plus que $M_n(\theta) \xrightarrow{P} -D(\theta_*, \theta)$. Nous avons besoin que cette convergence soit uniforme sur θ . Il faut aussi s'assurer que la fonction $D(\theta_*, \theta)$ se comporte bien. Soit $M(\theta) = -D(\theta_*, \theta)$ et supposons $\sup_{\theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0$ et que, pour chaque $\varepsilon > 0$, $\sup_{\theta: |\theta - \theta_*| < \varepsilon} M(\theta) < M(\theta_*)$. Alors $\hat{\theta}^{ML} \xrightarrow{P} \theta_*$.

Équivariance du MLE

Théorème 5.1 *Soit $g(\theta)$ être une fonction de θ . Et soit $\hat{\theta}^{ML}$ être le MLE de θ . Alors, $g(\hat{\theta}^{ML})$ est le MLE de $g(\theta)$.*

Preuve. Soit $h = g^{-1}$ désigne l'inverse de g . Alors $\hat{\theta}^{ML} = h(g(\hat{\theta}^{ML}))$. Pour toute $g(\theta)$: $\ell(X_i; h(g(\theta))) = \prod_{i=1}^n f(X_i; \theta) = \ell(X_i; \theta)$ où $\theta = h(g(\theta))$. Par conséquent, pour tout $g(\theta)$, $\ell(X_i; \theta) < \ell(X_i; \hat{\theta}^{ML}) = \ell(X_i; h(g(\hat{\theta}^{ML})))$. ■

Exemple 5.1 *Soit $X \sim N(\theta, 1)$. Le MLE pour θ est $\hat{\theta}^{ML} = \bar{X}_n$. Laisser $g(\theta) = e^\theta$. Ensuite, le MLE pour $g(\theta)$ est $g(\hat{\theta}^{ML}) = e^{\bar{X}_n}$.*

Normalité asymptotique

Il s'avère que la distribution de $\hat{\theta}^{ML}$ est approximativement Normal et nous pouvons calculer analytiquement sa variance approximative. Pour explorer cela, nous avons d'abord besoin de quelques définitions.

Définition 5.1 *La fonction de score est définie par*

$$s(x; \theta) := \frac{\partial \log(f(x; \theta))}{\partial \theta} \quad (5.A.4)$$

L'information de Fisher est définie par

$$I_n(\theta) := \sum_{i=1}^n \text{Var}(s(X_i; \theta)) := \sum_{i=1}^n (E(s^2(X_i; \theta)) - [E(s(X_i; \theta))]^2) \quad (5.A.5)$$

Pour $n = 1$ nous écrivons parfois $I(\theta)$ au lieu de $I_1(\theta)$. On peut montrer que $E(s(x; \theta)) = 0$. Il s'ensuit alors que $\text{Var}(s(x; \theta)) = E(s^2(x; \theta))$. L'information de Fisher est vérifiée pour les familles de distribution qui sont régulières (Familles exponentielles) telles qu'elle

1. Pour tous θ , support $\{x : f(x; \theta) > 0\}$ est le même.
2. $\frac{\partial \log(f(x; \theta))}{\partial \theta}$ doit exister et être fini pour tout x et θ .
3. $E_\theta |W(X)| < \infty$ pour tous les θ , alors $(\frac{\partial}{\partial \theta})^k E_\theta W(X) = \int W(x) (\frac{\partial}{\partial \theta})^k f(x; \theta) dx$.

En effet, une simplification supplémentaire de $I_n(\theta)$ est donnée dans le résultat suivant :

Théorème 5.2

1. On a $I_n(\theta) = nI(\theta)$ avec :

$$I(\theta) := -E\left(\frac{\partial^2 \log(f(x; \theta))}{\partial^2 \theta}\right) = -\int \left(\frac{\partial^2 \log(f(x; \theta))}{\partial^2 \theta}\right) f(x, \theta) dx \quad (5.A.6)$$

(Normalité asymptotique de la MLE) Avec une régularité appropriée et si les conditions précédentes sont vérifiées, alors :

2. $\text{Var}(\hat{\theta}^{ML}) = \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}$ et on a pour $n \rightarrow \infty$,

$$\frac{(\hat{\theta}^{ML} - \theta)}{\sqrt{\text{Var}(\hat{\theta}^{ML})}} \xrightarrow{d} N(0, 1) \quad (5.A.7)$$

$$\begin{aligned}
 3. \text{Var}(\widehat{\theta}^{ML}) &= \frac{1}{I_n(\widehat{\theta}^{ML})} = \frac{1}{nI(\widehat{\theta}^{ML})} \text{ alors} \\
 &\frac{(\widehat{\theta}^{ML} - \theta)}{\sqrt{\text{Var}(\widehat{\theta}^{ML})}} \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty \tag{5.A.8}
 \end{aligned}$$

Preuve. Avant de procéder à la démonstration de ce théorème, il est nécessaire de prouver que l'espérance de la fonction de score de (5.A.4) est $E(s(x; \theta)) = 0$. On sait d'abord que la PDF est vérifie la condition suivante : $\int f(x, \theta) dx = 1$, donc

$$0 = \int \frac{\partial}{\partial \theta} f(x; \theta) dx = \int \frac{\frac{\partial}{\partial \theta} f(x; \theta)}{f(x; \theta)} f(x; \theta) dx = \int \frac{\partial \log(f(x; \theta))}{\partial \theta} f(x; \theta) dx$$

Revenant à la relation (5.A.4), où $E(s(x; \theta)) = 0$. (1) est simple à prouver, de la même manière que nous avons utilisés dans la preuve de $E(s(x; \theta)) = 0$,

$$0 = \frac{\partial}{\partial \theta} \int \frac{\partial \log(f(x; \theta))}{\partial \theta} f(x; \theta) dx = \int \frac{\partial}{\partial \theta} \left(\frac{\partial \log(f(x; \theta))}{\partial \theta} f(x; \theta) dx \right)$$

Ensuite

$$\int \left(\frac{\partial}{\partial \theta} \right)^2 \log(f(x; \theta)) \cdot f(x; \theta) dx + \int \frac{\partial \log(f(x; \theta))}{\partial \theta} \cdot \frac{\partial}{\partial \theta} f(x; \theta) = 0$$

Alors

$$\begin{aligned}
 \int \left(\frac{\partial}{\partial \theta} \right)^2 \log(f(x; \theta)) \cdot f(x; \theta) dx + \int \frac{\partial \log(f(x; \theta))}{\partial \theta} \cdot \frac{\frac{\partial}{\partial \theta} f(x; \theta)}{f(x; \theta)} f(x; \theta) &= 0 \\
 \int \left(\frac{\partial^2 \log(f(x; \theta))}{\partial^2 \theta} \right) f(x; \theta) dx + \int \left(\frac{\partial \log(f(x; \theta))}{\partial \theta} \right)^2 f(x; \theta) &= 0
 \end{aligned}$$

Nous savons que $E(s(x; \theta)) = 0$, donc on peut écrire (5.A.5) dans la forme

$$I_n(\theta) = E(s^2(X_i; \theta)) = \int \left(\frac{\partial \log(f(X_i; \theta))}{\partial \theta} \right)^2 f(X_i; \theta).$$

Puisque l'échantillon X_1, X_2, \dots, X_n de v.a's i.i.d suit la même distribution F , alors $I_i(\theta) = I_1(\theta)$ pour tous $i = 1, 2, \dots, n$ donc on a $I_n(\theta) = nI_1(\theta) = nI(\theta)$. Et à partir de là, la dernière égalité équivaut à ce qui suit :

$$I(\theta) = - \int \left(\frac{\partial^2 \log(f(x; \theta))}{\partial^2 \theta} \right) f(x, \theta) dx = -E \left(\frac{\partial^2 \log(f(x; \theta))}{\partial^2 \theta} \right)$$

Preuve de la relation (5.A.7), soit $L(x; \theta) = \log(f(x; \theta))$. Alors

$$L'(x; \hat{\theta}^{ML}) \approx L'(x; \theta) + (\hat{\theta}^{ML} - \theta) L''(x; \theta) \quad (5.A.9)$$

On a $L'(x; \hat{\theta}^{ML}) = 0$. Réorganisez l'équation ci-dessus pour obtenir que $(\hat{\theta}^{ML} - \theta) = -L'(x; \theta) / L''(x; \theta)$ où, en d'autres termes,

$$\sqrt{n} (\hat{\theta}^{ML} - \theta) = \frac{\frac{1}{\sqrt{n}} L'(x; \theta)}{-\frac{1}{n} L''(x; \theta)} \quad (5.A.10)$$

Soit $S_i = \partial \log f(X_i; \theta) / \partial \theta$. Rappel que $E(S_i) = 0$ et aussi $Var(S_i) = I(\theta)$. Ainsi

$$\frac{1}{\sqrt{n}} L'(X_i; \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i := \sqrt{n} \bar{S} = \sqrt{n} (\bar{S} - 0)$$

Par le théorème centrale limite, on trouve que $\sqrt{n} (\bar{S} - 0) \xrightarrow{d} N(0, I(\theta))$. Soit $R_i = -\partial^2 \log f(X_i; \theta) / \partial \theta^2$. Alors $E(R_i) = I(\theta)$ et

$$\frac{1}{n} L''(X_i; \theta) = \frac{1}{n} \sum_{i=1}^n R_i := \bar{R}_i.$$

et on a $\overline{R_i} \xrightarrow{P} I(\theta)$ par la loi des grands nombres. Appliquer le théorème de Slutsky (qui dit que si $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} c$ alors $X_n Y_n \xrightarrow{P} cX$) pour conclure que :

$$\sqrt{n}(\widehat{\theta}^{ML} - \theta) \xrightarrow{d} \frac{N(0, I(\theta))}{I(\theta)} = N\left(0, \frac{1}{I(\theta)}\right)$$

En admettant que $I(\theta)$ est une fonction continue de θ , il s'ensuit que $I_n(\widehat{\theta}^{ML}) \xrightarrow{P} I(\theta)$.

Donc

$$\frac{(\widehat{\theta}^{ML} - \theta)}{\sqrt{I_n(\widehat{\theta}^{ML})}} = \frac{\sqrt{n}(\widehat{\theta}^{ML} - \theta)}{\sqrt{I(\widehat{\theta}^{ML})}} = \left(\frac{\sqrt{n}(\widehat{\theta}^{ML} - \theta)}{\sqrt{I(\theta)}} \right) \sqrt{\frac{I(\theta)}{\widehat{\theta}^{ML}}}$$

Le premier terme tend en distribution vers $N(0, 1)$. Le second terme tend en probabilité vers 1. De manière informelle, le théorème (5.2) dit que la distribution de la MLE peut être approchée avec $N\left(\theta, 1/I_n(\widehat{\theta}^{ML})\right)$. De ce fait on peut construire un intervalle de confiance (asymptotique). ■

Théorème 5.3 Soit T_n l'intervalle de confiance dit asymptotique tel que :

$$T_n = \left[\widehat{\theta}^{ML} - t_{\alpha/2} \widehat{Var}(\widehat{\theta}^{ML}); \widehat{\theta}^{ML} + t_{\alpha/2} \widehat{Var}(\widehat{\theta}^{ML}) \right] \quad (5.A.11)$$

où $t_{\alpha/2}$ est le $\alpha/2$ th-quantile à partir du haut de $N(0, 1)$ la distribution normale standard. Alors $P(\theta \in T_n) = 1 - \alpha$ quand $n \rightarrow \infty$.

Preuve. Soit T désignant une variable aléatoire normale standard. Alors,

$$\begin{aligned} P(\theta \in T_n) &= P\left(\widehat{\theta}^{ML} - t_{\alpha/2} \widehat{Var}(\widehat{\theta}^{ML}) \leq \theta \leq \widehat{\theta}^{ML} + t_{\alpha/2} \widehat{Var}(\widehat{\theta}^{ML})\right) \\ &= P\left(-t_{\alpha/2} \leq \frac{\widehat{\theta}^{ML} - \theta}{\widehat{Var}(\widehat{\theta}^{ML})} \leq t_{\alpha/2}\right) \rightarrow P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) \\ &= 1 - \alpha \end{aligned}$$

■

Par exemple pour $\alpha = 0.05$, on a $t_{\alpha/2} = 1.96 \approx 2$. Donc

$$T_n = \left[\widehat{\theta}^{ML} - 2\widehat{Var}(\widehat{\theta}^{ML}); \widehat{\theta}^{ML} + 2\widehat{Var}(\widehat{\theta}^{ML}) \right] \quad (5.A.12)$$

est un intervalle de confiance approximatif un seuil de confiance de 95%. De plus, la normalité asymptotique de la MLE indique que la distribution des paramètres ajustés aux données est normale centrée sur les vrais paramètres, avec une matrice de covariance qui dépend de la taille de l'échantillon.

Optimalité

Considérons deux estimateurs Z_n et E_n et supposons que $\sqrt{n}(Z_n - \theta) \xrightarrow{d} N(0, var(z))$ et $\sqrt{n}(E_n - \theta) \xrightarrow{d} N(0, var(e))$. On définit l'efficacité relative asymptotique de E par rapport à Z par :

$$ARE(E, Z) = var(z) / var(e). \quad (5.A.13)$$

Théorème 5.4 Si $\widehat{\theta}^{ML}$ est le MLE et $\widehat{\theta}^{(\cdot)}$ un autre estimateur alors

$$ARE(\widehat{\theta}^{(\cdot)}, \widehat{\theta}^{ML}) \leq 1 \quad (5.A.14)$$

Ainsi, la MLE a la plus petite variance (asymptotique) et on dit que le MLE est efficace ou asymptotiquement optimal le qu'elle dépend de n .

Ce résultat repose sur le fait que le modèle supposé est correct. Si le modèle est erroné, la MLE n'est plus optimale. En fait, il existe de nombreuses autres façons de générer des estimateurs. Comment choisir un parmi eux ? La réponse se trouve dans la théorie de la décision qui est une théorie formelle de comparaison des procédures statistiques. Soit $\widehat{\theta}^{(\cdot)}$ un estimateur de θ . Dans le langage de la théorie de la décision, un estimateur est parfois appelé règle de décision et les valeurs possibles de la règle de décision sont appelées actions. On mesurera l'écart entre $\widehat{\theta}^{(\cdot)}$ et θ en utilisant une fonction de perte

$L(\theta, \hat{\theta}^{(\cdot)})$. Voici quelques exemples de fonctions de perte, Pour les modèles paramétriques

$.L(\theta, \hat{\theta}^{(\cdot)})$	nom
$(\theta - \hat{\theta}^{(\cdot)})^2$	Perte d'erreur quadratique
$ \theta - \hat{\theta}^{(\cdot)} $	Perte d'erreur absolue
$ \theta - \hat{\theta}^{(\cdot)} ^p$	Perte L_p
0 si $\theta = \hat{\theta}^{(\cdot)}$ et 1 si $\theta \neq \hat{\theta}^{(\cdot)}$	Zéro-Un Perte
$\int \log \left(\frac{f(x, \theta)}{f(x, \hat{\theta}^{(\cdot)})} \right) f(x, \theta) dx$	Perte de Kullback-Leibler.

Table 5.1: Exemples de fonctions de perte

qui satisfont des conditions de régularité faible, l'estimateur du maximum de vraisemblance est approximativement minimax. Considérez la perte d'erreur au carré qui est le biais au carré plus la variance. Dans les modèles paramétriques avec de grands échantillons, on peut montrer que le terme de variance domine le biais, de sorte que le risque de MLE est à peu près égal à la variance

$$R(\theta, \hat{\theta}^{ML}) = Var(\hat{\theta}^{ML}) - bias^2 \approx Var(\hat{\theta}^{ML}) \quad (5.A.15)$$

Typiquement, le biais carré est de ordre de $O(n^{-2})$ tandis que la variance est de ordre de $O(n^{-1})$. Comme nous l'avons vu au la théorie (5.2), la variance de la MLE est d'environ $Var(\hat{\theta}^{ML}) \approx (1/n) I(\theta)$ où $I(\theta)$ est l'information de Fisher (5.A.5). Ainsi, $nR(\theta, \hat{\theta}^{ML}) \approx I(\theta)$. Pour tout autre estimateur $\hat{\theta}^{(\cdot)}$ on peut montrer que pour n grand, $R(\theta, \hat{\theta}^{(\cdot)}) \geq R(\theta, \hat{\theta}^{ML})$. Plus précisément,

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\theta - \hat{\theta}^{(\cdot)}| < \varepsilon} nR(\hat{\theta}^{(\cdot)}, \hat{\theta}^{ML}) \geq \frac{1}{I(\theta)} \quad (5.A.16)$$

Cela signifie que, dans un sens d'échantillon local et large, le MLE est minimax. Dans la plupart des modèles paramétriques, avec de grands échantillons, le MLE est approximativement minimax et Bayessienne.

5.1.2 MBM pour multi-racines avec le R programme

Le programme R est une suite intégrée d'installations logicielles pour la manipulation de données, le calcul et l'affichage graphique. Il s'est développé rapidement et a été étendu par une grande collection de packages, de nombreuses personnes utilisent R comme un système de calcul en statistiques.

En pratique, nous avons obtenues plusieurs résultats importants de l'application de cette nouvelle méthode numérique pour la recherche de racines d'équation non linéaire appelée MBM pour multi-racines. Le point important est le choix de l'intervalle en prenant les points suivants, $f(\alpha) \cdot f(\beta) < 0$ avec $f(\alpha)$ et $f(\beta)$ comme un extrémum global et, plus important encore, chaque fois que les grandes différences $|\tau - 1|$ avec cas1 ou $|\frac{1}{\tau} - 1|$ avec cas2 pour multi-racines devient meilleur.

La méthode MBM pour multi-racines avec le logiciel R est présenté comme suit:

```

ridhaRoot<-function(f,a,b,num=10,eps=1e-07)
{h=abs(b-a)/num
i=0
j=0
a1=b1=0
while(i<=num)
{ a1=a+i*h
  b1=a1+h
  if(f(a1)==0){print(a1)
               print(f(a1))}
  else if (f(b1)==0){print(b1)
                    print(f(b1))}
  else if (f(a1)*f(b1)<0){
  repeat{ if(abs(b1-a1)<eps) break
          m=(b1-a1)/(f(b1)-f(a1))
          d=b1-f(b1)*m
          if(f(a1)*f(d)<0) b1<-d else a1<-d }
  print(j+1)
  j=j+1
  print(d)
  print(f(d))
  }
  i=i+1
  }
  if(j==0)
  print("finding root is fail")
  else print("finding root is successful")
  }

```

Figure 5.1: MBM pour la multi-racines avec le programme R.

Exemple 5.2 : Nous allons chercher les racines de $f(x) = x^3 - 2x + 2$ sur l'intervalle $[-3; 3]$

```
f <- function(x){x^3-2*x+2}
ridhaRoot(f,-3,3)
#####
> f <- function(x){x^3-2*x+2}
> ridhaRoot(f,-3,3)
[1] 1
[1] -1.769292
[1] 0
[1] "finding root is successful"
#####
```

R Archive Network (CRAN) qui se trouve à : <http://cran.r-project.org/>. Il existe plusieurs sites ou serveurs dans le monde où le logiciel peut être téléchargé, accessible à l'adresse : <http://cran.r-project.org/mirrors.html>. La version R pour Windows est la plus utilisée, donc si vous utilisez les systèmes d'exploitation Linux ou Mac OS X, suivez les instructions sur le site Web du CRAN.

5.1.3 MBA pour multi-racines (4.1.3) avec le R

L'algorithme de bisection modifié est basé sur deux méthodes : méthode de bisection et bisection modifiée méthode (MBM). La différence entre MBA et la méthode de regula falsi (fausse position) réside dans le fait qu'elle s'appuie sur une méthode de Bisection pour déterminer le nouveau l'intervalle $(\alpha_{n-1}^*, \beta_{n-1}^*)$. Et nous trouvons que La MBM diverge avec les fonction à courbe non lisse comme nous donne en figure (4.3), par contre la méthode de regula falsi est très active. Pour cela nous utilisons la méthode de Bisection dans La MBM pour définir MBA pour multi-racine.

Sur le plan pratique, nous avons programmé MBA pour multi-racines (4.1.3) à l'aide de la version du logiciel R i386 3.3.3. comme suit:

```

KouiderRoot<-function(f,a,b,num=10,eps=1e-07)
{h=abs(b-a)/num
i=0
j=0
a1=b1=d=0
while(i<=num)
{ a1=a+i*h
b1=a1+h
if(f(a1)==0){print(a1)
print(f(a1))}
if (f(b1)==0){print(b1)
print(f(b1))}
else if (f(a1)*f(b1)<0) {
repeat{
c=(a1+b1)/2
if (f(a1)*f(c)<0){
m= (a1-c)/(f(a1)-f(c))
d=c-f(c)*m
if(f(a1)*f(d)>0){
if(f(c)*f(d)<0)
a1<-c
b1<-d }
else b1<-d
}
}
}
}
else
{ m= (b1-c)/(f(b1)-f(c))
d=c-f(c)*m
if(f(b1)*f(d)>0){
if(f(c)*f(d)<0)
a1<-c
b1<-d }
else a1<-d
}
}
if(abs(c-d)<eps)
break
}
print(j+1)
j=j+1
print(d)
}
i=i+1
}
if(j==0)
print("finding root is fail")
else print("finding root is successful")
}

```

Figure 5.2: MBA pour multi-racine avec le programme R.

5.2 Appendix B

5.2.1 Preuve des théorèmes (4.1) , (4.5) et (4.6)

Preuve du théorème (4.1)

Soit f est une fonction continue sur $[\alpha; \beta]$ où $f(\alpha)$ et $f(\beta)$ sont de signe opposé c'est-à-dire $f(\alpha) \cdot f(\beta) < 0$ et à partir de (4.1.3) ou (4.1.4) nous avons $\alpha < x_1 < \beta$. Ensuite, formons le MBM, nous obtenons deux cas :

Cas1 : $f(\alpha) f(x_1) < 0$ c'est-à-dire $f(x_1)$ et $f(\beta)$ ont de même signe. Alors, pour tout $n \geq 0$ nous avons $\alpha < x^* < x_n < x_1 < \beta$ donc on a $x_n - x^* < x_n - \alpha < x_1 - \alpha < \beta - \alpha$.

Alors on peut écrire les termes x_n en (4.1.3) avec $x_0 = \beta$ par

$$x_1 = \alpha - f(\alpha) \frac{x_0 - \alpha}{f(x_0) - f(\alpha)} \quad (5.B.1)$$

$$x_2 = \alpha - f(\alpha) \frac{x_1 - \alpha}{f(x_1) - f(\alpha)}$$

\vdots

$$x_n = \alpha - f(\alpha) \frac{x_{n-1} - \alpha}{f(x_{n-1}) - f(\alpha)}$$

par récurrence on peut écrire x_n comme

$$x_n = \alpha - f(\alpha)^n \frac{\beta - \alpha}{\prod_{i=0}^{n-1} (f(x_i) - f(\alpha))} \quad (5.B.2)$$

alors pour tout $n \geq 0$ à partir de (5.B.2) on a

$$x_n - \alpha = -f(\alpha)^n \frac{\beta - \alpha}{\prod_{i=0}^{n-1} (f(x_i) - f(\alpha))} \quad (5.B.3)$$

Supposer que $f(\theta_1) \leq f(x_i) < f(\alpha) \leq f(\theta_2)$ pour tout $x_i \in [\alpha; x_1]$ et $(\theta_1; \theta_2) \in [\alpha; x_1]^2$ avec $f(x_i) \leq 0$ et $f(\alpha) > 0$ pour $i = \{0, 1, \dots, n-1\}$ nous avons $f(x_i) - f(\alpha) < f(\theta_2) - f(\theta_1)$. Donc, on a

$$\frac{1}{f(x_i) - f(\alpha)} < \frac{1}{f(\theta_2) - f(\theta_1)}; \quad (5.B.4)$$

où $f(\theta_2) - f(\theta_1) > 0$ et $f(x_i) - f(\alpha) < 0$.

De plus, supposer que $f(\theta_1) \leq f(\alpha) < f(x_i) \leq f(\theta_2)$ avec $f(x_i) \geq 0$ et $f(\alpha) < 0$. Alors

nous avons $f(\alpha) - f(x_i) < f(\theta_2) - f(\theta_1)$ et

$$\frac{1}{f(x_i) - f(\alpha)} > \frac{-1}{f(\theta_2) - f(\theta_1)}; \quad (5.B.5)$$

où $f(\theta_2) - f(\theta_1) > 0$ et $f(\alpha) - f(x_i) < 0$. Donc avec (5.B.4) et (5.B.5) nous avons que

$$\left| \frac{1}{f(x_i) - f(\alpha)} \right| < \frac{1}{f(\theta_2) - f(\theta_1)} \quad (5.B.6)$$

Alors, d'après (5.B.3) et (5.B.6) on a

$$|x_n - \alpha| = |f(\alpha)|^n \frac{\beta - \alpha}{\left| \prod_{i=0}^{n-1} (f(x_i) - f(\alpha)) \right|} < |f(\alpha)|^n \frac{\beta - \alpha}{\prod_{i=0}^{n-1} (f(\theta_2) - f(\theta_1))}$$

Puis, pour $|x_n - x^*| < |x_n - \alpha|$ et $f(\theta_2) - f(\theta_1) = |f(\theta_2) - f(\theta_1)|$ nous aurons

$$|x_n - x^*| < |f(\alpha)|^n \frac{\beta - \alpha}{\prod_{i=0}^{n-1} |f(\theta_2) - f(\theta_1)|} \quad (5.B.7)$$

Efin on obtenir :

$$|x_n - x^*| < \left| \frac{f(\alpha)}{f(\theta_2) - f(\theta_1)} \right|^n (\beta - \alpha)$$

Pour $\tau = \frac{f(\theta_2)}{f(\theta_1)}$, on a :

$$|x_n - x^*| < \left| \frac{f(\alpha)}{f(\theta_1)} \right|^n \frac{\beta - \alpha}{|\tau - 1|^n} \quad (5.B.8)$$

et si $f(\theta_1) := f(\alpha)$ on a $\tau = \frac{f(\theta_2)}{f(\alpha)}$

$$|x_n - x^*| < \frac{\beta - \alpha}{|\tau - 1|^n} \quad (5.B.9)$$

Cas2 : $f(\beta)f(x_1) < 0$ c'est-à-dire $f(x_1)$ et $f(\alpha)$ ont le même signe. Alors, pour tout $n \geq 0$ nous avons $\alpha < x_1 < x_n < x^* < \beta$ donc on a $x_n - x^* < x_n - \beta < x_1 - \beta < \beta - \alpha$. Alors on peut écrire les termes x_n en (4.1.4) avec $x_0 = \alpha$ par récurrence on peut écrire x_n par

$$x_n = \beta - f(\beta)^n \frac{\beta - \alpha}{\prod_{i=0}^{n-1} (f(\beta) - f(x_i))} \quad (5.B.10)$$

alors pour tout $n \geq 0$ à partir de (5.B.10) nous avons

$$x_n - \beta = -f(\beta)^n \frac{\beta - \alpha}{\prod_{i=0}^{n-1} (f(\beta) - f(x_i))} \quad (5.B.11)$$

Supposer que $f(\theta_1) \leq f(x_i) < f(\beta) \leq f(\theta_2)$ pour tous $x_i \in [\alpha; x_1]$ et $(\theta_1; \theta_2) \in [\alpha; x_1]^2$ avec $f(x_i) \leq 0$ et $f(\beta) > 0$ pour $i = \{0, 1, \dots, n-1\}$ nous avons $f(x_i) - f(\beta) < f(\theta_2) - f(\theta_1)$. Donc, on a

$$\frac{1}{f(\beta) - f(x_i)} > \frac{-1}{f(\theta_2) - f(\theta_1)}; \quad (5.B.12)$$

où $f(\theta_2) - f(\theta_1) > 0$ et $f(x_i) - f(\beta) < 0$.

De plus, supposer que $f(\theta_1) \leq f(\beta) < f(x_i) \leq f(\theta_2)$ avec $f(x_i) \geq 0$ et $f(\beta) < 0$. Alors on a $f(\beta) - f(x_i) < f(\theta_2) - f(\theta_1)$ et

$$\frac{1}{f(\beta) - f(x_i)} < \frac{1}{f(\theta_2) - f(\theta_1)}; \quad (5.B.13)$$

où $f(\theta_2) - f(\theta_1) > 0$ et $f(\alpha) - f(x_i) < 0$. Donc avec (5.B.12) et (5.B.13) nous avons :

$$\left| \frac{1}{f(\beta) - f(x_i)} \right| < \frac{1}{f(\theta_2) - f(\theta_1)} \quad (5.B.14)$$

Et, d'après (5.B.11) et (5.B.14) on a :

$$|x_n - \beta| < |f(\beta)|^n \frac{\beta - \alpha}{\prod_{i=0}^{n-1} (f(\theta_2) - f(\theta_1))}$$

Puis, pour $|x_n - x^*| < |x_n - \beta|$ et $f(\theta_2) - f(\theta_1) = |f(\theta_2) - f(\theta_1)|$ on obtienir :

$$|x_n - x^*| < \left| \frac{f(\beta)}{f(\theta_2) - f(\theta_1)} \right|^n (\beta - \alpha) \quad (5.B.15)$$

Mettons $\tau = \frac{f(\theta_2)}{f(\theta_1)}$, donc

$$|x_n - x^*| < \left| \frac{f(\beta)}{f(\theta_2)} \right|^n \frac{\beta - \alpha}{\left| \frac{1}{\tau} - 1 \right|^n} \quad (5.B.16)$$

et si $f(\theta_2) := f(\beta)$ on a $\tau = \frac{f(\beta)}{f(\theta_1)}$

$$|x_n - x^*| < \frac{\beta - \alpha}{\left| \frac{1}{\tau} - 1 \right|^n} \quad (5.B.17)$$

et si on prend $f(\theta_1) := f(\alpha)$ avec $f(\theta_2) := f(\beta)$ en (5.B.8) et (5.B.16) on trouve que le preuve de théorème (4.1) est justifiat.

Pour voir combien d'itérations seront nécessaires, pour le **cas1** $f(\alpha) f(x_1) < 0$ supposons que nous voulons avoir

$$|x_n - x^*| < \eta_1$$

qui est satisfait pour :

$$\frac{\beta - \alpha}{|\tau - 1|^n} < \eta_1$$

En prenant des logarithmes des deux côtés, on obtenir :

$$n < \frac{\log \left(\frac{\beta - \alpha}{\eta_1} \right)}{\log |\tau - 1|} \quad (5.B.18)$$

Et pour le **cas2** où $f(\beta) f(x_1) < 0$ on trouve :

$$n < \frac{\log\left(\frac{\beta-\alpha}{\eta_1}\right)}{\log\left|\frac{1}{\tau} - 1\right|} \quad (5.B.19)$$

Si le $|\tau - 1|$ pour le cas1 ou $\left|\frac{1}{\tau} - 1\right|$ pour le cas2 est trop grande, la limite de l'erreur deviendra trop petite et le nombre d'itérations plus faible.

Preuve de théorème (4.5)

Les résultats (1) et (2) sont simples.

La preuve du résultat (3) : la fonction (4.2.3) est représenté par :

$$\psi(\theta) = \left(\frac{1}{k} \sum_{i=1}^k \log(1 - \theta C_{i,k}) + 1\right) \left(\frac{1}{k} \sum_{i=1}^k (1 - \theta C_{i,k})^{-1}\right) - 1 \quad (5.B.20)$$

En utilisant l'inégalité de Jensen $\frac{1}{k} \sum_{i=1}^k \log(1 - \theta C_{i,k}) \leq \log(1 - \theta \bar{C})$ pour $\theta \in \mathcal{B}$ et puisque $C_{1,k} < C_{i,k}$ pour tous $i = 1, \dots, k$, avec $\theta < 0$ nous avons :

$$\frac{1}{k} \sum_{i=1}^k \left(\frac{1}{1 - \theta C_{i,k}}\right) < \left(\frac{1}{1 - \theta C_{1,k}}\right) \quad (5.B.21)$$

Pour $\theta < 0$, il s'ensuit que :

$$\psi(\theta) \leq (\log(1 - \theta \bar{C}) + 1) (1 - \theta C_{1,k})^{-1} - 1 < 0 \quad (5.B.23)$$

Puis, pour $\theta < 0$ on trouve que (5.B.23) est réalisée et conduit à :

$$\log(1 - \theta \bar{C}) + 1 < (1 - \theta C_{1,k}) \quad (5.B.24)$$

Alors, $1 - \theta \bar{C} < \exp(-\theta C_{1,k})$. D'où $1 - \theta \bar{C} < 1 + (-\theta C_{1,k}) + \frac{1}{2}(-\theta C_{1,k})^2$.

Et supposant que $\theta < \theta_L$ nous obtenons

$$\theta_L = \frac{2(C_{1,k} - \bar{C})}{(C_{1,k})^2} \quad (5.B.25)$$

Ceci permet d'obtenir la preuve du théorème (4.5).

Preuve du théorème (4.6)

Les résultats (1) et (2) sont simples. Pour la preuve du résultat (3), on peut réécrire la fonction (4.3.11) ainsi :

$$\psi(\theta_1) = \left(\frac{1}{r} \sum_{j=1}^k \log(1 - \theta_1 C_{j,k}) + 1 \right) \left(\frac{1}{k} \sum_{j=1}^r \left(\frac{1}{1 - \theta_1 C_{j,r}} \right) \right) + \left(\frac{1}{k} \sum_{j=1}^m \left(\frac{1}{1 - \theta_1 C_{j,m}} \right) \right) - 1 \quad (5.B.26)$$

En utilisant l'inégalité de Jensen $\frac{1}{r} \sum_{j=1}^k \log(1 - \theta_1 C_{j,k}) \leq \frac{k}{r} \log(1 - \theta_1 \bar{C})$ pour $\theta_1 \in \mathcal{B}$ puisque $C_{1,r} < C_{j,r}$ et $C_{1,m} < C_{j,m}$ pour tous $j = 1, \dots, k$ respectivement, et avec $\theta_1 < 0$ on a :

$$\frac{1}{k} \sum_{j=1}^r \left(\frac{1}{1 - \theta_1 C_{j,r}} \right) < \frac{r}{k} \left(\frac{1}{1 - \theta_1 C_{1,r}} \right) \quad \text{et} \quad \frac{1}{k} \sum_{j=1}^m \left(\frac{1}{1 - \theta_1 C_{j,m}} \right) < \frac{m}{k} \left(\frac{1}{1 - \theta_1 C_{1,m}} \right) \quad (5.B.27)$$

Pour $\theta_1 < 0$ il s'ensuit que

$$\psi(\theta_1) \leq \left(\frac{k}{r} \log(1 - \theta_1 \bar{C}) + 1 \right) \left(\frac{r}{k} \left(\frac{1}{1 - \theta_1 C_{1,r}} \right) \right) + \frac{m}{k} \left(\frac{1}{1 - \theta_1 C_{1,m}} \right) - 1$$

alors

$$\psi(\theta_1) \leq \log(1 - \theta_1 \bar{C}) \left(\frac{1}{1 - \theta_1 C_{1,r}} \right) + \frac{r}{k} \left(\frac{1}{1 - \theta_1 C_{1,r}} \right) + \frac{m}{k} \left(\frac{1}{1 - \theta_1 C_{1,m}} \right) - 1 < 0 \quad (5.B.28)$$

Puis, pour $\theta_1 < 0$ si $C_{1,r} < C_{1,m}$ ($\min(C_{1,m}; C_{1,r})$) on a $\left(\frac{1}{1 - \theta_1 C_{1,m}} \right) < \left(\frac{1}{1 - \theta_1 C_{1,r}} \right)$ on

trouve que (5.B.28) est implique que :

$$\psi(\theta_1) \leq \log(1 - \theta_1 \bar{C}) \left(\frac{1}{1 - \theta_1 C_{1,r}} \right) + \left(\frac{1}{1 - \theta_1 C_{1,r}} \right) - 1 < 0 \quad (5.B.29)$$

Alors $\log(1 - \theta_1 \bar{C}) + 1 < (1 - \theta_1 C_{1,r})$, donc $1 - \theta_1 \bar{C} < \exp(-\theta_1 C_{1,r})$. D'où $1 - \theta_1 \bar{C} < 1 + (-\theta_1 C_{1,r}) + \frac{1}{2}(-\theta_1 C_{1,r})^2$ et supposons que $\theta_1 < (\theta_1)_L$.

Ensuite, nous obtenons $(\theta_1)_L = \frac{2(C_{1,r} - \bar{C})}{(C_{1,r})^2}$. Et si $C_{1,m} < C_{1,r}$ ($\min(C_{1,m}; C_{1,r})$) nous avons $\left(\frac{1}{1 - \theta_1 C_{1,r}} \right) < \left(\frac{1}{1 - \theta_1 C_{1,m}} \right)$ on trouve $(\theta_1)_L = \frac{2(C_{1,m} - \bar{C})}{(C_{1,m})^2}$.

Cependant, nous savons que $C_{1,k} := \min(C_j; \delta_j)$ pour tous les $j = 1, \dots, k$. Cela implique que, si $C_{1,r} := C_{1,k}$ est observé ou si $C_{1,m} := C_{1,k}$ est censuré on trouve:

$$(\theta_1)_L = \frac{2(C_{1,k} - \bar{C})}{(C_{1,k})^2} \quad (5.B.30)$$

Ceci permet d'obtenir la preuve de théorème (4.6).

5.2.2 Application MBA pour estimer l'indice de queue sous données complet (4.2)

Pour un exemple réel d'application de modèle de valeur extrême. Considérons les données de Danish (Danish Data), ces données décrivent de grandes réclamations d'assurance incendie au Danemark du jeudi 3 janvier 1980 au lundi 31 décembre 1980 . Un vecteur numérique contenant 2167 observations. Ils offrent de nombreuses possibilités de modélisation et ont été utilisés par de nombreux chercheurs pour illustrer leurs méthodes, voir McNeil (1997) et Resnick (1997). Ces données peuvent être trouvées dans le package `evir` du logiciel R (Ihaka et Gentleman citent R).

Dans cet exemple, nous nous intéressons aux limites de confiance pour les 73 pertes maximales mensuelles au cours de la période mentionnée à partir des 2167 observations données,

répertoriées dans un ordre croissant voir tableau(5.2)

4.450262	4.500000	4.625000	4.657070	4.821601	4.867987	5.469679
5.600000	6.773920	6.798457	7.425743	7.613470	7.643979	7.863696
8.678881	10.011123	10.072303	10.500000	10.820452	10.998350	11.123471
11.595547	11.890606	12.376238	12.523191	12.631813	13.348165	14.013548
14.239000	14.293194	14.394581	14.678899	15.284653	16.441659	17.068467
17.569546	17.739274	17.743491	18.322083	18.424135	18.646484	19.070278
19.162304	19.400000	20.049941	20.826733	20.863675	20.969856	22.258226
23.283859	24.970273	25.288376	25.953860	26.214641	27.338066	27.829314
29.026037	29.037106	31.055901	32.387807	32.467532	34.141547	38.154392
42.091448	46.500000	47.019521	50.065531	56.225426	57.410636	65.707491
144.657591	152.413209	263.250366				

Table 5.2: Pertes maximales mensuelles à partir de données de Danish pour un échantillon aléatoire de $n = 73$

Pour commencer la recherche les racines de $\psi(\theta)$, on calcule les bornes $(\varepsilon, \theta_U, \theta_L)$ pour $k = 50$. Alors $\varepsilon = 2.817172 \times 10^{-9}$ et autres limites données par $\theta_L = -3.018889 \times 10^{-1}$, et $\theta_U = 3.798661 \times 10^{-3}$.

L'algorithme recherchera deux racines sur $[\varepsilon; \theta_U]$.et $[\theta_L; -\varepsilon]$ de $\psi(\theta)$. En utilisant les rendements de notre programme, nous avons trouvé les estimations du maximum de vraisemblance GPD pour un dépassement du seuil $k > 50$ les pertes maximales mensuelles pendant la période mentionnée à partir de 73 des observations données pour les données de Danish et qui sont $\hat{\gamma} = 0.08586831$ and $\hat{\sigma} = 32.31397400$.

La matrice de variance-covariance asymptotique des paramètres (γ, σ) est donnée par l'inverse de la matrice d'information de Fisher. Ensuite, leur correspondant 95% les intervalles de confiance sont $-0.1249036 < \gamma < 0.2966402$ et $21.08807 < \sigma < 43.53988$.

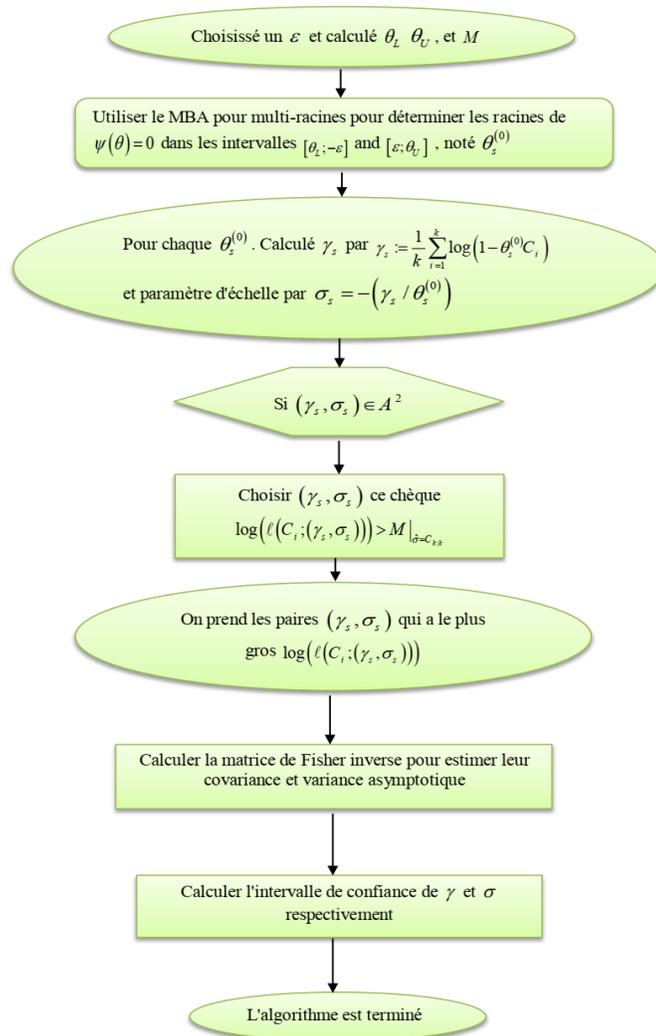


Figure 5.3: Procédure de MBA pour estimer l'indice de queue sous données complètes

5.2.3 Application MBA pour estimer l'indice de queue sous données censurées

Une étude de simulation est réalisée en utilisant la distribution de Pareto généralisée sous données censurées, nous générons un échantillon $(X_i; i = 1, \dots, 15)$ de taille 15 qui suit la distribution GPD avec des paramètres $\gamma_1 = 0.5$ et $\sigma_1 = 25$ et de plus un deuxième échantillon $(Y_i; i = 1, \dots, 15)$ avec la même taille que le premier et avec des paramètres $\gamma_2 = 0.8$ et $\sigma_2 = 20$. Les 15 valeurs sont classées à partir du tableau (5.2) suivant:

36.330	4.903	0.078	70.547	5.377
13.065	16.349	4.321	11.689	6.322
0.226	6.552	22.814	6.004	58.550

Table 5.3: Échantillon $(Z_n; n \geq 1)$ de taille 15 généré avec la GPD sous censure.

Nous avons $r = 9$ données observées (les données n'ont pas été censurées à différents moments 36.330, 4.903, 5.377, 13.065, 4.321, 6.322, 0.226, 22.814 et 58.550 respectivement) et $m = 6$ données censurées avec $Z_{15,15}$ est censuré.

Lorsque les données ne sont pas censurées, nous utilisons MBA pour estimer l'indice de queue qui représenté dans la figure (5.3) précédente (basées sur des données complètes), telles que $r = n = 15$ et $m = 0$ (dans ce cas, l'algorithme est identique au MLE classique proposé par Kouider [26]).

Dans cette étude, nous notons que nous avons considéré l'échantillon $(Z_n; n \geq 1)$ qui nous générons où $Z_i = \min(X_i; Y_i)$ pour $i = 1, \dots, n$ comme un échantillon des excès pour $Z_j = E_j - t$ pour $E_j > t$ avec $j = 1, \dots, 15$. Notons $Z_j = C_j$. Nous donnons aussi un algorithme pour estimer l'indice de queue sous données censurées et après nous définissons une méthode d'estimation pour estimer les paramètres de la GPD sous données censurées. L'indice des valeurs extrêmes basé sur $(Z_n; n \geq 1)$ ou $\gamma := \left(\frac{\gamma_1 \gamma_2}{\gamma_2 + \gamma_1} \right) = 0.30769$ et avec $\theta = -(\gamma/\sigma)$. Pour estimer (γ, σ) par le maximum de vraisemblance de la GPD sera utilisé dans la procédure suivante :

1. Trouver la racine $\hat{\theta}^{ML}$ qui est la solution de l'équation $\psi(\theta) = 0$ pour $\theta \in]\theta_L; -\varepsilon[\cup]\varepsilon; \theta_U[$ où,

$$\psi(\theta) = \left(\frac{1}{15} \sum_{j=1}^{15} \log(1 - \theta C_j) + 1 \right) \cdot \left(\frac{1}{15} \sum_{j=1}^{15} (1 - \theta C_j)^{-1} \right) - 1$$

2. Calculer $\hat{\gamma}^{ML}$ par :

$$\hat{\gamma}^{ML} = \frac{1}{15} \sum_{j=1}^{15} \log(1 - \hat{\theta}^{ML} C_j).$$

3. Pour $\hat{\theta}^{ML} = -(\hat{\gamma}^{ML}/\hat{\sigma}^{ML})$, calculer $\hat{\sigma}^{ML}$ par $\hat{\sigma}^{ML} = -(\hat{\gamma}^{ML}/\hat{\theta}^{ML})$.

Puis, grâce à théorème (4.6) en calculant les bornes $\varepsilon = 3.42 \times 10^{-9}$ et $\theta_L = -6.81 \times 10^2$, et $\theta_U = 1.42 \times 10^{-2}$. Et appliquer MBA pour multi-racines en figure (5.2) sur les intervalles $[\varepsilon; \theta_U]$. et $[\theta_L; -\varepsilon]$ de $\psi(\theta)$, on trouve $\theta_1^{(0)} = -0.024207$ et $\theta_2^{(0)} = 0.013466$ aussi les valeurs correspondantes à $\theta_1^{(0)}$ et $\theta_2^{(0)}$ du tableau (5.4) illustre ces résultats:

$\theta_1^{(0)} = -0.024207$	$\gamma_1^* = 0.3049$	$\sigma_1^* = 12.5956$	$\log(\ell(C_j; (\gamma_1^*, \sigma_1^*))) = -57.5737$
$\theta_2^{(0)} = 0.013466$	$\gamma_2^* = -0.44494$	$\sigma_2^* = 33.04180$	$\log(\ell(C_j; (\gamma_2^*, \sigma_2^*))) = -60.79249$

Table 5.4: Présentant des résultats obtenus de l'application la méthode d'estimation présentée en (5.3)

Maintenant nous avons $\gamma_1^* > 0$ et $\sigma_1^* > 0$ et $\sigma_2^*/\gamma_2^* < -70.457$ pour $C_{15,15} := 70.457$ sont le maximum local de la log-vraisemblance GPD sur A . Le maximum limite pour $\hat{\gamma} = -1$ et $\hat{\sigma} = 70.457$ qui est donné en (4.2.1) ou en (4.3.9) avec $m = 0$ est égale à : $\log(\ell(C_j; (-1, 70.457))) := -63.84419$. Donc si on considère $\log(\ell(C_j; (\gamma_1^*, \sigma_1^*))) > \log(\ell(C_j; (\gamma_2^*, \sigma_2^*)))$ les estimateurs du maximum de vraisemblance GPD pour les données non censurées sont $\hat{\gamma}^{ML} = 0.3049$ et $\hat{\sigma}^{ML} = 12.5956$ et $\theta^{ML} = -0.024207$.

Pour estimer (γ_1, σ_1) dans le cas de données censurées par le maximum de vraisemblance de la GPD on doit être utilisé la procédure suivante :

1. Trouver la racine $\hat{\theta}_1^{c,KIB}$ qui est la solution de l'équation $\tilde{\psi}(\theta_1) = 0$ pour $\theta_1 \in$

$](\theta_1)_L; -\varepsilon[\cup]\varepsilon; (\theta_1)_U[$ où,

$$\tilde{\psi}(\theta_1) = \left(\frac{1}{9} \sum_{j=1}^{15} \log(1 - \theta_1 C_{j,15}) \right) \left(\frac{1}{15} \sum_{j=1}^9 (1 - \theta_1 C_{j,9})^{-1} \right) + \frac{1}{15} \sum_{j=1}^{15} (1 - \theta_1 C_{j,15})^{-1} - 1$$

2. Calculer $\hat{\gamma}_1^{c,KIB}$ par :

$$\hat{\gamma}_1^{c,KIB} = \frac{1}{15} \sum_{j=1}^{15} \log(1 - \hat{\theta}_1^{c,KIB} C_j).$$

3. Pour $\hat{\theta}_1^{c,KIB} = - \left(\hat{\gamma}_1^{c,KIB} / \hat{\sigma}_1^{c,KIB} \right)$, calculer $\hat{\sigma}_1^{c,KIB}$ par $\hat{\sigma}_1^{c,KIB} = - \left(\hat{\gamma}_1^{c,KIB} / \hat{\theta}_1^{c,KIB} \right)$.

La méthode d'estimation présentée en (5.4) utilise les bornes $\varepsilon = 3.42 \times 10^{-9}$ et $(\theta_1)_L = -6.81 \times 10^2$, et $(\theta_1)_U = 1.42 \times 10^{-2}$, elle recherche deux racines de $\tilde{\psi}(\theta_1)$ sur les intervalles $[\varepsilon; (\theta_1)_U]$.et $[(\theta_1)_L; -\varepsilon]$ respectivement.

Sur l'intervalle $[(\theta_1)_L; -\varepsilon]$ en employant l'algorithme bisection modifié pour les multi-racines comme nous l'avons défini dans la figure (5.2) qui converge vers $\hat{\theta}_1^{(c,KIB)} = -0.0241$ donc la racine $\hat{\theta}_1^{(c,KIB)}$ correspond à l'estimation $\hat{\gamma}_1^{(c,KIB)} = 0.506$ et $\hat{\sigma}_1^{(c,KIB)} = 21.014$ et par (4.3.7) on trouve que $\log \left(\ell \left(C_j; \left(\hat{\gamma}_1^{(c,KIB)}, \hat{\sigma}_1^{(c,KIB)} \right) \right) \right) = -39.108$ et $\hat{\gamma}_1^{(c,KIB)} > 0$ et $\hat{\sigma}_1^{(c,KIB)} > 0$ le maximum local de la log-vraisemblance GPD sous censure sur A , et comme $C_{n,n}$ est le temps de défaillance, nous obtenons le maximum limite pour $\hat{\gamma}_1^{(c,KIB)} = -1$ et $\hat{\sigma}_1^{(c,KIB)} = 70.457$ donné en (4.3.9) avec $m = 6$ par $\log \left(\ell \left(C_j; (-1, 70.457) \right) \right) := -40.136$. Mais sur l'intervalle $[\varepsilon; (\theta_1)_U]$ aucune racine de $\tilde{\psi}(\theta_1)$ est existe. Donc les estimateurs du maximum de vraisemblance GPD pour les données censurées sont $\hat{\gamma}_1^{(c,KIB)} = 0.506$ et $\hat{\sigma}_1^{(c,KIB)} = 21.014$ et $\hat{\theta}_1^{(c,KIB)} = -0.0241$.

Par conséquent, leurs intervalles de confiance correspondants à 95% pour γ_1 et σ_1 sont $-1.08 < \gamma_1 < 2.09$ avec $-2.08 < \sigma_1 < 44.09$.

D'une autre côté $\hat{\gamma}_1^{(c,KIB)} = 0.506$ et l'estimateur ML adapté (3.2.3) sont :

$$\hat{\gamma}_1^{(c,ML)} = \frac{\hat{\gamma}^{(ML)}}{\hat{p}} = 0.50817 \text{ où } \hat{\gamma}^{ML} = 0.3049 \text{ et } \hat{p} = \frac{9}{15}$$

de plus, on obtient $\frac{\widehat{\sigma}^{ML}}{\widehat{\sigma}_1^{(c, KTB)}} = 8.99909 \simeq 9$ qui est le nombre d'observation, dans cette étude de simulation.

Nous pouvons réécrire les dérivées du deuxième ordre de la log-vraisemblance en présence de censure avec $\theta_1 = -(\gamma_1/\sigma_1)$ et $\widetilde{M}(C_j) := \log(\ell(C_j; (\gamma, \sigma)))$ par:

1.
$$\frac{\partial^2 \widetilde{M}(C_j)}{\partial^2 \gamma_1} := \frac{1}{\gamma_1^3} \left(\sum_{j=1}^r \left[(1 + \gamma_1) (1 - (1 - \theta_1 C_{j:r})^{-1})^2 - 2 \log(1 - \theta_1 C_{j:r}) + 2 (1 - (1 - \theta_1 C_{j:r})^{-1}) \right] \right. \\ \left. + \sum_{j=1}^m \left[(1 - (1 - \theta_1 C_{j:m})^{-1})^2 - 2 \log(1 - \theta_1 C_{j:m}) + 2 (1 - (1 - \theta_1 C_{j:m})^{-1}) \right] \right),$$
2.
$$\frac{\partial^2 \widetilde{M}(C_j)}{\partial^2 \sigma_1} := \left(\frac{\theta_1}{\gamma_1} \right)^2 \left(\sum_{j=1}^r \left[1 - \left(1 + \frac{1}{\gamma_1} \right) \left(2 (1 - (1 - \theta_1 C_{j:r})^{-1}) - (1 - (1 - \theta_1 C_{j:r})^{-1})^2 \right) \right] \right. \\ \left. - \frac{1}{\gamma_1} \sum_{j=1}^m \left(2 (1 - (1 - \theta_1 C_{j:m})^{-1}) - (1 - (1 - \theta_1 C_{j:m})^{-1})^2 \right) \right),$$
3.
$$\frac{\partial^2 \widetilde{M}(C_j)}{\partial \gamma_1 \partial \sigma_1} := \frac{\theta_1}{\gamma_1^2} \left(\sum_{j=1}^r \left[\left(1 + \frac{1}{\gamma_1} \right) (1 - (1 - \theta_1 C_{j:r})^{-1})^2 - (1 - (1 - \theta_1 C_{j:r})^{-1}) \right] \right. \\ \left. + \frac{1}{\gamma_1} \sum_{j=1}^m (1 - (1 - \theta_1 C_{j:m})^{-1})^2 \right).$$

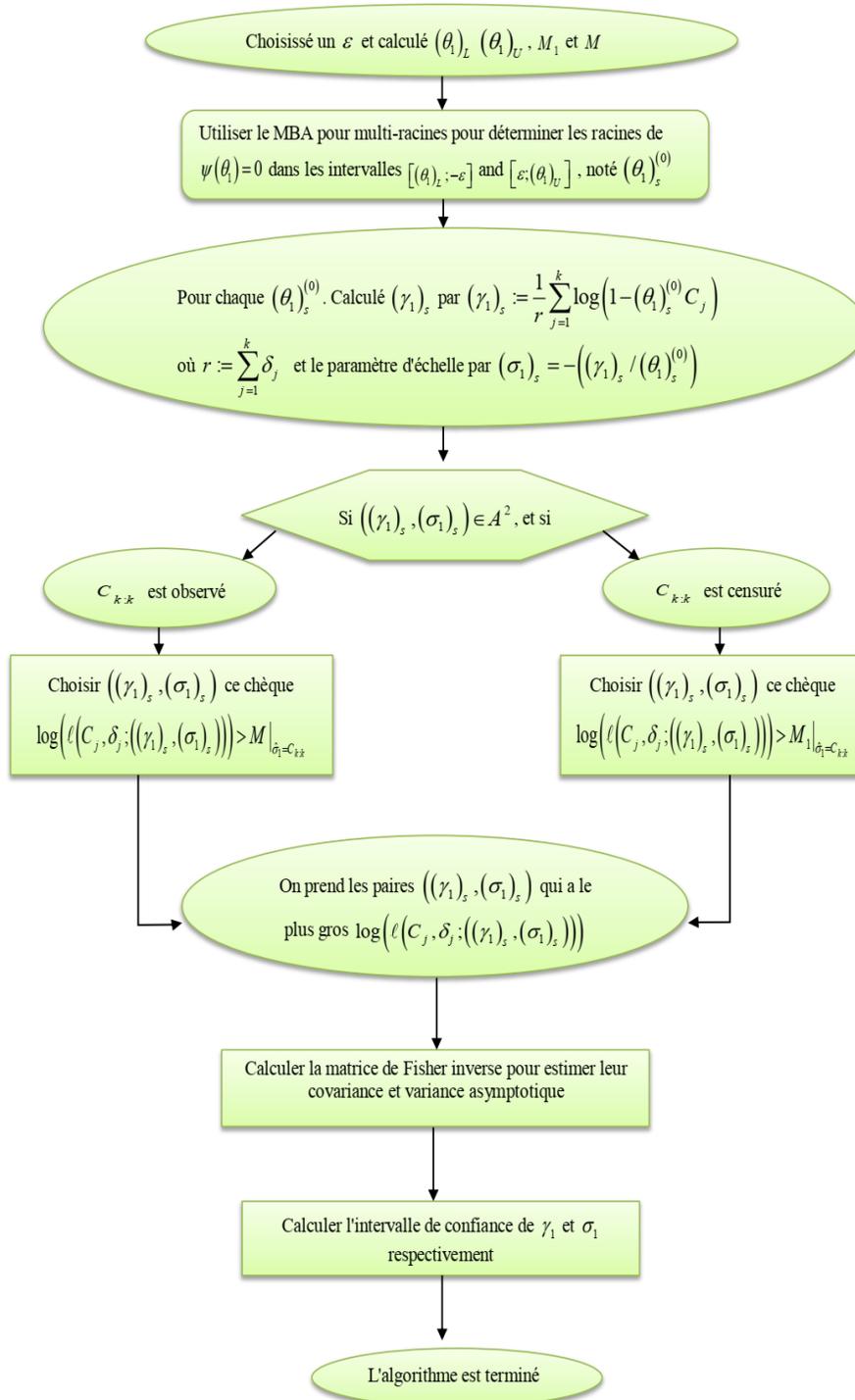


Figure 5.4: Procédure MBA pour estimer l'indice de queue sous données censurées

Conclusion

L'intérêt principal de notre travail repose sur l'étude des estimateurs de l'indice de queue γ (dans le cas strictement positif) pour une variable aléatoire positive et une censure aléatoire positive et particulièrement dans le cas 1, les distributions à queue lourde et la loi de la GPD. Comme les lois à queue lourde appartiennent évidemment au domaine d'attraction de Fréchet (i.e. avec un indice de queue $\gamma > 0$).

L'indice de queue lourde γ donne une information sur la vitesse de décroissance de la fonction de survie à l'intérieur de cette famille de loi: Une valeur de γ proche de zéro (resp. l'infini) correspond à une décroissance rapide (resp. lente) de la queue de distribution. La GPD a récemment émergé comme une distribution influente dans la modélisation des distributions de données à queue lourde dans diverses applications en finance, risque opérationnel, assurance et études environnementales....etc.

En conclusion ce travail, a enrichi la méthode du maximum de vraisemblance pour l'estimation des paramètres de la GPD, grâce à la présentation d'un algorithme efficace qui s'appuie sur une très excellente méthode numérique et avec une erreur minimale car il s'appuie dans toutes ses étapes sur le théorème des valeurs intermédiaires. Nous avons également présenté un nouvel estimateur de l'indice de queue basé sur la méthode du maximum de vraisemblance adapté en présence de données censurées présenté dans [28] (2023) noté: $\hat{\gamma}_1^{(c,KIB)}$ KIB-estimateur. De plus, nous avons également fait adapter cet estimateur au cas des données complètes, noté $\hat{\gamma}^{KIB}$.qui vérifie les conditions de la définition dans Einmahl et al [11] (2008) si $\theta_1 := \theta$ (Dans la définition de $\hat{\gamma}_1^{(c,KIB)}$ (4.3.13) on a défini $\hat{\gamma}_1^{(c,ML)}$

(4.3.12)), sachant que p est la proportion d'observations non censurées estimer par \hat{p} , Une application numérique détaillée à la fin de ce travail vient pour confirmer l'efficacité de cet estimateur.

Concluding notes

We were interested in the tail index estimators γ if we have a positive random variable or a positive random censor. Knowing that, the tail index is strictly positive. Because we have concentrated in this work on heavy-tailed distributions and rather the laws of GPD. Heavy-tailed laws obviously belong to the Fréchet domain of attraction (i.e. with a tail index $\gamma > 0$).

The heavy tail index γ gives information on the rate of decrease of the survival function within this family of law: A value of γ close to zero (resp. infinity) corresponds to a rapid (resp. slow) decrease in the tail of the distribution. GPD has recently emerged as an influential distribution in modeling heavy-tailed datasets in various applications in finance, operational risk, insurance, and environmental studies.

In conclusion of this work, we can say that we have enriched the maximum likelihood method for MLE with the parameters of the GPD. This is done by presenting an efficient algorithm which relies on a very excellent numerical method and is not subject to error as it relies at all its stages on the theorem of intermediate values. We also learned a new tail index estimator based on the maximum likelihood method. He presented by Kouider et al [28] (2023) and he: $\hat{\gamma}^{KIB}$ KIB-estimator. In addition, we also had this estimator adapted with the censored data, we mentioned by $\hat{\gamma}_1^{(c,KIB)}$. Then we find that $\hat{\gamma}_1^{(c,KIB)}$ achieve the definition of Einmahl et al [11] (2008) if $\theta_1 := \theta$ (In the definition of $\hat{\gamma}_1^{(c,KIB)}$ (4.3.13) so it is defined $\hat{\gamma}_1^{(c,ML)}$ (4.3.12)), knowing that p the proportion of uncensored observations is already estimated. Thus, we also supported this result with a numerical application.

Bibliographie

- [1] Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N. (1993). Statistical Models Based on counting Processes. New York: Springer.
- [2] Balkema, August A., and Laurens De Haan. (1974). Residual life time at great age. The Annals of probability, 792-804.
- [3] Beirlant, J. , Vynckier, P., and Teugels, J. (1996). Tail index estimation, pareto quantile plots regression diagnostics. Journal of the American Statistical Association. 91(436) : 1659-1667.
- [4] Beirlant J, Dierckx G, Guillou A. (2005). Estimation of the extreme-value index and generalized quantile plots. Bernoulli 11 949–970
- [5] Beirlant J, Guillou A., Dierckx G, Fils-Villetard A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. Extremes, 10(3),151 174.
- [6] Davis R, Resnick S. (1984). Tail estimates motivated by extreme value theory. The Annals of Statistics, 1467 1487.
- [7] Davison Anthony C. (1984) .Modelling excesses over high thresholds, with an application. Statistical extremes and applications. Springer Netherlands, 461-482.
- [8] Deheuvels P, Haeusler E, Mason D M. (1988). Almost sure convergence of the Hill estimator. Math. Proc. Cambridge Philos. Soc, 104 (02) : 371 381.

- [9] Dekkers A. L., de Haan L. (1989). On the estimation of the extreme value index and large quantile estimation. *Ann. Statist*, 1795-1832.
- [10] Drees, Holger, Ana Ferreira, and Laurens De Haan. (2004). On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability*, 1179-1201.
- [11] Einmahl, J.H.J., Fils-Villetard, A., Guillou. (2008) . *Statistics of extremes under random censoring ; Bernoulli*
- [12] Embrechts, P., and C. Klüppelberg., T. Mikosch. (1997). *Modelling Extremal Events for Insurance and Finance. Applications of Mathematics-Stochastic Modelling and Applied Probability, No. 33.* New York: Springer.
- [13] Fisher, Ronald Aylmer, and Leonard Henry Caleb Tippett. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society.* Vol. 24. No. 02. Cambridge University Press.
- [14] Gnedenko, Boris. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of mathematics*, 423-453.
- [15] Grimshaw, Scott D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* 35.2 : 185-191.
- [16] de Haan,L. (1970). On regular variation and its application to the weak convergence of sample extremes. *Mathematical Centre Tracts* 32.
- [17] de Haan,L. (1976). *Sample extremes : an Introduction.* Springer-Verlag, New York.
- [18] de Haan,L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* Springer.
- [19] de Haan,L and U. Stadtmüller. (1996). Generalized regular variation of second order. *J. Australian Math. Soc. (Series A)* 61: 381-395.

- [20] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist*, 3(5) : 1163-1174.
- [21] Hosking, Jonathan R.M., and James R. Wallis. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29.3 : 339-349.
- [22] Jankinson, A. F.(1955). The Frequency Distribution of the Annual Maximum (or Minimum) of the Meteorological Elements Quarterly. *Journal of the Royal Meteorological Society* 81, 185-171.
- [23] Jay, Villanueva. (2012). Approximation of equations with a hand-held calculator. Florida Memorial University Miami , [http:// archives.math.utk.edu / ICTCM / VOL23 / C012 / paper](http://archives.math.utk.edu/ICTCM/VOL23/C012/paper).
- [24] Kaplan, E.L, Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association* and, 53 :457-481.
- [25] Kouider Mohammed Ridha. (2012). Estimation de moyenne des distributions à queues lourdes. Mémoire de Master, de l'université Mohamed Khider, Biskra.
- [26] Kouider Mohammed Ridha. (2019). On Maximum Likelihood Estimates for the Shape Parameter of the Generalized Pareto Distribution. *Science Journal of Applied Mathematics and Statistics*. Vol. 7, No. 5, pp. 89-94.
- [27] Kouider Mohammed Ridha.(2019). Modified Bisection Algorithm for Multiple Roots of Nonlinear Equation With the R Software. <http://dx.doi.org/10.2139/SSRN.3451155>.
- [28] Kouider M, R., Idiou, Nesrine., Benatia Fatah. (2023). Modified Bisection Algorithm In Estimating The Extreme Value Index Under Random Censoring. *TWMS J. App. and Eng. Math*. V.13, N.1, pp. 1408-1422.

- [29] Kouider M.R, Idiou N , Benatia F. (2023). Adaptive Estimators of the General Pareto Distribution Parameters under Random Censorship and Application. *Journal of Science and Arts.*, No 2, pp. 395-412.
- [30] N.H.Bingham.(1987). Regular Variation. *Encyclopedia of mathematics and its applications*, 27.
- [31] Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *Ann. Proba*, 754 764.
- [32] Pham, M. H., Tsokos, C., & Choi, B.-J. (2018). Maximum likelihood estimation for the generalized Pareto distribution and goodness-of-fit test with censored data. *Journal of Modern Applied Statistical Methods*, 17(2).
- [33] Resnick, Sidney I. (1987). *Extreme values, regular variation, and point processes*. Springer, New York.
- [34] Reiss, Rolf-Dieter, Michael Thomas, and R. D. Reiss. (2007). *Statistical analysis of extreme values*. Vol. 2. Basel: Birkhäuser.
- [35] Smith, Richard L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72.1: 67-90.
- [36] Stute Winfried. (1995). The Central limit Theorem Under Random Censorship. *The Annal of statistics* Vol. 23. No.2, 422-439
- [37] Tanakan, S. (2013). A New Algorithm of Modified Bisection Method for Nonlinear Equation. *Applied Mathematical Sciences* 7.123: 6107-6114.
- [38] Pickands III, James. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 119-131.
- [39] Von Mises, R. (1936) . La distribution de la plus grande de n valeurs, *Rev. Math. Union Interbalk*. 1, 141-160.