N° d'ordre :/**Année-/**

REPUBLIQUE ALGÉRIENNE DEMOCRATIQUE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ MOULOUD MAMMERI

TIZI-OUZOU

FACULTÉ DES SCIENCES

Département de mathématiques

Mémoire de master en mathématiques

Option : Probabilités-Statistiques.

Thème

Estimation des paramètres de la distribution de Pareto généralisée

Présenté par

M^{elle}**CHERIF Lina**

Devant le jury composé de

Mr Mamou Mohamed	M.A.A	Président
Mme Atil Lynda	M.C.A	Examinatrice
Mr Berkoun Youcef	Professeur	Encadreur

Soutenu octobre 2019

Je tiens à remercier Mr Berkoun. J'adresse aussi mes remerciements au président du jury Mr Mamou pour vos relectures attentives, vos remarques, vos conseils et pour l'intérêt que vous avez porté à mon travail. Ma gratitude va également à Madame Atil pour avoir accepté d'examiner mon mémoire, de m'avoir encouragée, soutenu et cru en moi durant tout mon cursus.

Mes remerciements vont également à monsieur Fellag et madame belkacem pour leurs remarques et les perspectives scientifiques.

Merci à tous les quatres de m'avoir tant apporté aussi bien humainement que scientifiquement. Merci pour nos dicussions si enrichissantes et d'avoir toujours été disponibles pour moi. J'espère avoir été digne de la confiance que vous m'avez accordé. J'ai beaucoup appris à vos côtés et je suis très honorée de vous avoir eu pour enseignants.

je tiens aussi à remercier Monsieur Hamadouche, Madame boualem, Madame Merabet et Monsieur Hamez pour l'excelence de leurs enseignenemnts qui m'ont été précieux et m'ont permis de progresser dans le domaine de la statistique.

A mes amies Nabiha, Baya, Sonia, Malha, Yasmine, Dahlia et Kamelia merci.

A mes soeurs Manel, Celia et anais pour m'avoir soutenu, merci.

A mes parents sans qui ce travail n'aurait pu se réaliser et à qui je dois tout. Merci pour votre patience, votre éducation, votre aide de tous les instants et votre soutien inconditionnel. Merci pour votre amour et pour l'exemple que vous êtes pour moi.

Table des matières

Introduction générale

Les événements extrêmes (tremblements de terre, inondations, accidents nucléaires, crises monétaires ou financières, krachs boursiers, émergence d'un nouveau phénomène endémique, etc.) sont par définition ceux qui causent le plus de dommages aux personnes, aux structures et aux infrastructures, raison pour laquelle ils dominent l'actualité quotidienne par leur caractère imprévisible.Ce sont des événements rares qui s'écartent fortement de la moyenne ou de la tendance habituelle et aux conséquences désastreuses . Divers domaines d'application nécessitent la prise en compte d'évènements rares. Or, dans de nombreuses applications, les données disponibles sont de trop courte durée pour pouvoir estimer empiriquement avec précision des valeurs aussi peu fréquentes. On doit donc, à partir de peu de données, construire un modèle nous permettant d'extrapoler et de prédire un événement sans commune mesure.

La théorie des valeurs extrêmes s'applique dans de nombreux domaines tels la fiabilité [13], la métallurgie [3]. et en astrophysique [9]. Elle s'intéresse également aux sciences de l'environnement, avec la modélisation de grands feux de forêts [1] ainsi que la climatologie [37] et la météorologie [8], .L'hydrologie [19], notamment suite aux travaux de Jules Emile Gumbel en 1954 [20] et son ouvrage [21], l'actuariat afin de se prémunir contre des sinistres ayant un grand impact [4] , [6] et la finance [15] , [16]. Pour d'autres exemples d'applications, se référer au livre de Reiss et Thomas [35]. Il existe également quelques articles de vulgarisation scientifique visant à introduire la problématique et les applications de la théorie des valeurs extrêmes. Citons notamment Matthews [29] qui donne une trés bonne vue d'ensemble.

La distribution de Pareto généralisée (GPD) a été largement utilisée dans le cadre

de la modélisation des événements extrêmes. Les résultats de la distribution de Pareto généralisée, lorsqu'elle est appliquée à des données réelles, dépendent en grande partie de l'estimation de ces paramètres. Il existe plusieurs méthodes dans la littérature pour estimer la distribution de Pareto généralisée ([7] , [10] , [24] , [34] ; [38] , [40] , [24]). Majoritairement, l'estimation est effectuée par la méthode du maximum de vraisemblance (MLE). Alternativement, la méthode des moments de probabilité pondérées (PWM) et la méthode des moments (MOM) sont souvent utilisés, surtout lorsque les échantillons sont de petite taille.

Ce mémoire est organisé en deux chapitres. Le premier chapitre est consacré à des rappels sur la loi limite du maximum, la loi des excés,définitions et estimaion du quantile.Dans le deuxième chapitre, nous présentons différentes méthodes d'estimation des paramètre de la disribution Pareto généralisée utilisées dans la littérature (méthode du maximum de vraisemblance (Maximum Likelihood Estimation, MLE), méthode des moments (Method Of Moments, MOM), méthode des moments de probabilité pondérés (Probability weighted moments, PWM), méthode des moments linéaires (L-moments) et méthode du maximum d'entropie) et nous présentons une exemple d'application qui a été traité par V.P. Singh et H.Guo [22] pour étudier la performance des estimateurs de la méthode du maximum de vraisemblance (MLE), méthode des moments (MOM), méthode des moments de probabilité pondérés (PWM) et méthode du maximum d'entropie .

Chapitre 1

Généralités

1.1 Introduction

La théorie des valeurs extrêmes est un domaine de la statistique qui s'est developpée dans les années trente avec le premier résultat de Fisher-Tippet [16]. Elle étudie l'occurence d'évenements extrêmes, i.e., d'évenements dont la probabilité d'apparition est relativement faible. Il est d'un grand intérêt de se prémunir contre les risques extrêmes, qu'ils résultent d'une crise financière, d'un accident nucléaire ou d'une catastrophe naturelle, compte tenu des répercussions humaines, économiques et financières que ces derniers peuvent avoir.
A l'inverse, on dispose d'un grand nombre d'observations pour les événements plus fréquents. La difficulté réside dans le fait de prédire des événements extrêmes à l'aide d'événements fréquents. On doit donc, à partir de peu de données, construire un modèle nous permettant d'extrapoler et de prédire un événement sans commune mesure.
La théorie des valeurs extrêmes fournit une base mathématique probabiliste rigoureuse sur laquelle il est possible de construire des modèles statistiques permettant de prévoir l'intensité et la fréquence de ces événements extrêmes.

1.2 Exemple illusratif

La problématique de ce sujet de mémoire peut être illustrée par l'exemple suivant de de Haan [12].

Dans la nuit du 31 janvier au 1er février 1953, de très fortes tempêtes traversant la mer du Nord balayèrent les côtes flamandes et néerlandaises d'ouest en est. Après que plusieurs digues eurent cédé, les provinces néerlandaises de la Hollande et de la Zélande furent particulièrement touchées. Les conséquences de ce raz-de-marée furent désastreuses. On dénombra plus de 2 500 morts, 47 000 habitations inondées et 10 000 détruites, 200 000 hectares de terres inondées, 30 000 têtes de bétail noyées, environ 9 pour 100 des fermes des Pays-Bas inondées et plus de 400 brèches dans les digues.

A la suite de cette tempête, le gouvernement néerlandais décida la mise en place du "plan Delta" pour se prémunir contre une nouvelle inondation. Il fallut construire un nouveau réseau de digues renforcées de plus de 500 km le long de la côte de la mer du Nord.

Un comité composé de nombreux scientifiques, parmi lesquels des statisticiens, se réunit afin d'étudier le phénomène et proposer ainsi des recommandations sur les hauteurs des digues. Il fallut prendre en compte des facteurs économiques (coût de construction, coût des inondations,...), des facteurs physiques (rôle du vent sur la marée,...), mais aussi les hauteurs de marées enregistrées lors des précédentes inondations. En 1953, les inondations avaient entraîné une montée des eaux à 3.85 mètres au-dessus du niveau de la mer, soit largement en-deçà des 4 mètres atteints le 1er novembre 1570, soit 382 ans auparavant.

Le but était de construire des digues assez grandes, de telle sorte qu'aucune vague ne les dépasse dans un horizon de 10 000 ans. Autrement dit, il convenait de déterminer quelle serait la hauteur de la plus grande vague dans un horizon de temps de 10 000 ans.

La difficulté résidait dans le fait que, pour calculer la hauteur des digues et donc la hauteur maximale d'une vague qui n'a jamais eu lieu, le comité d'experts devait se baser sur les informations des années précédentes ; or il n'y avait que très peu de données disponibles en particulier pour des événements de cette ampleur.

Le 4 octobre 1986 marqua l'achèvement du plan Delta aux Pays-Bas. Il s'agit du plus grand chantier de génie civil de tous les temps. Il permit de relier toutes les îles côtières de la province de Zélande par des digues.

Après calcul, le comité d'expert estima que ces digues devraient mesurer au moins 5 mètres, estimation fondée sur l'utilisation de techniques statistiques empruntées à la théorie des valeurs extrêmes qui constitue le sujet du présent mémoire.

1.3 La loi des valeurs extrêmes

Historiquement, l'étude de la loi de probabilité du maximum d'un échantillon de n variables aléatoires a été la première approche pour décrire les événements extrêmes. Fisher et Tippet [17] ont, les premiers, déduit de manière heuristique les lois limites possibles pour le maximum d'une suite de variables aléatoires indépendantes et de même loi, avant que Gnedenko [18] n'obtienne rigoureusement la convergence, dont la preuve fut simplifiée par de Haan [11]. Les travaux de Von Mises [42] et Jenkinson [26] ont permis de donner une forme unifiée à ce résultat. Les applications ont commencé suite aux travaux de Gumbel [20] , en particulier en hydrologie.

Plaçons nous dans un cadre statistique et considérons n variables aléatoires réelles $X_1, X_2, ..., X_n$ indépendantes et identiquement distribuées (i.i.d) de fonction de répartition F . On notera dans la suite l'échatillon ordonné :

$$X_{1,n} \leq X_{1,n} \leq ... \leq X_{n,n}$$

Deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes, le minimum de l'échantillon $Min = X_{1,n}$ et le Maximum $M_n = X_{n,n}$. On notera qu'il est très facile de passer de l'un à l'autre à l'aide de la relation :

$$Min = -max(-X_1, ..., -X_n)$$

Ainsi, la suite des développementS du premier chapitre se portera sur l'étude du comportement asymptotique du maximum uniquement.

1.3.1 Distribution Généralisée des valeurs extrêmes (GEV)

La théorie des valeurs extrêmes s'intéresse au comportement asymptotique du maximum de variable aléatoire indépendante et ideniquement distribuée, de répartition commun F.

Ces valeurs extrêmes se situent dans la queue de la distribution. Forcément, le comportement asymptotique du maximum va dépendre de la queue de distribution. On a $F_{M_n} = (F(x))^n$. la loi de F étan inconnue en pratique, le comportement de F^n sera encore

plus difficile à étudier.
On peut cependant remarquer que :

$$Lim_{n\to\infty}F_{M_n}(x) = \begin{cases} 0, & \text{si } x \geq x_F\,; \\ 1, & \text{si } x < x_F. \end{cases}$$

où x_F est le point terminal de la loi F, définit par :

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

Le théorème ci-dessous est fondamental en théorie des valeurs extrêmes car il établit la loi asymptotique du maximum M_n normalisé d'un échantillon.

Théorème 1.1. *[[17]]*
Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F. S'il existe deux suites normalisantes réelles $(a_n)_{n>1} > 0$ et $(b_n)_{n>1} \in \mathbb{R}$ et une loi non dégénérée H telle que :

$$\lim_{n\to\infty} P(\frac{M_n - b_n}{a_n}) = Lim_{n\to\infty}F^n(a_n x + b_n) = H(x),$$

alors H est l'une des trois lois limites :

$$\text{Fréchet :} \qquad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp -x^\alpha, & x > 0, \end{cases} \quad \alpha > 0$$

$$\text{Weibull :} \qquad \Psi_\alpha(x) = \begin{cases} \exp\left(-(-x)^\alpha\right), & x < 0, \\ 1, & x \geq 0, \end{cases} \quad \alpha < 0$$

$$\text{Gumbel :} \qquad \Lambda(x) = \Big\{\ \exp -(\exp(-x)), \quad x \in \mathbb{R}.$$

Illustrons le Théorème 1.1 à travers l'exemple d'une variable aléatoire X suivant une loi exponentielle de paramètre 1. Sa fonction de répartition est donnée par :

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

Le support de la loi étant $\mathbb{R}$ on a $x_F = \infty$ d'où M_n converge en probabilité vers 1. Effectuons la normalisation suivante :

$$\begin{aligned} P(\frac{M_n - \log(n)}{1} \leq x) &= P(M_n \leq x + log(n)) \\ &= (F(x + \log(n))^n \\ &= (1 - \exp(-x - log(n)))^n \\ &= (1 - \frac{\exp(-x)}{n})^n \\ &\simeq \exp(-\exp(-x)) \end{aligned}$$

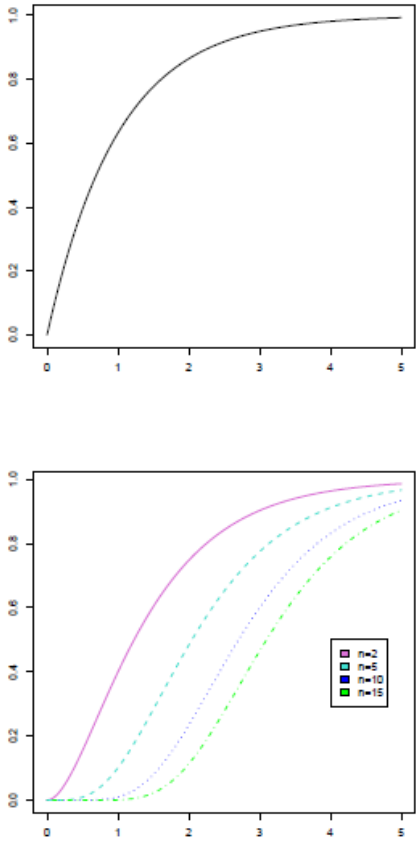

Figure 1.1 [28] : Convergence de la fonction de répartition de la loi exponentielle de moyenne 1 pour différentes puissances.

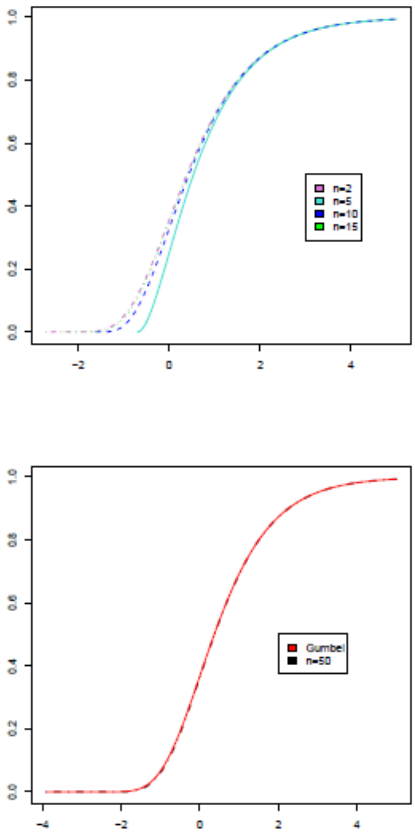

Figure 1.2 [28] : Convergence de la loi du maximum normalisé d'une loi exponentielle de moyenne 1 vers une loi de Gumbel.

Les trois distribuions peuvent être représentées par une distribution paramètrée définie comme suit :

Définition 1.1. *Représentation de Jenkinson von Mises de la distribution généralisée des valeurs extrêmes (GEV)*
la distribution généralisée des valeurs extrêmes $H_{k,\mu,\sigma}$ $,k \in \mathbb{R}$ $,\mu \in \mathbb{R}$ *et* $\sigma > 0$ *est donnée par :*

$$H_{k,\mu,\sigma}(x) = \begin{cases} \exp -(1+k\frac{x-\mu}{\sigma})^{-\frac{1}{k}}, & 1+k\frac{x-\mu}{\sigma} > 0, k \neq 0\,; \\ \exp\left(-\exp\left(-x\right)\right), & x \in \mathbb{R}, k = 0. \end{cases}$$

où k, μ *et* σ *sont respectivement les paramètres de forme, position et d'échelle.*

Remarques 1.1. *Le paramètre* k *fournit un indice sur le la loi limite tel que :*

– *Le cas* $k = \alpha^{-1} > 0$ *correspond à la loi de Fréchet* Φ_k

– *Le cas* $k = 0$ *correspond à la loi de Gumbel* Λ

– *Le cas* $k = -\alpha^{-1} < 0$ *correspond à la loi de Weibull* Ψ_k

Quelques densités ont été représentées pour k fixé sur la Figure 1.3.

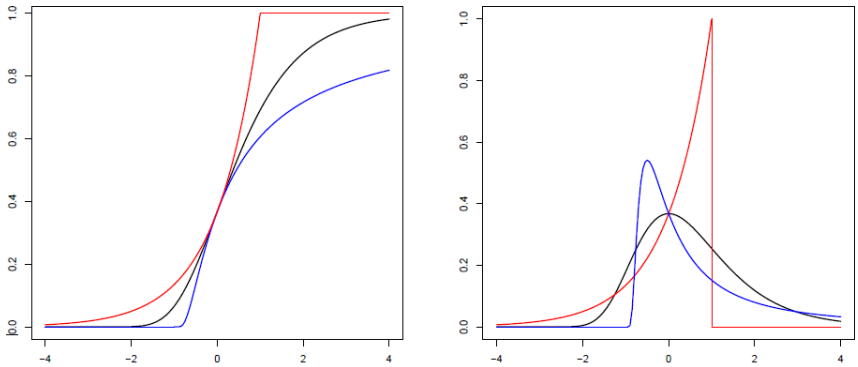

Figure 1.3[28] : A gauche : H_k . A droite : les densités associées à la loi des valeurs extrêmes (noir : k= 0, bleu : k= 1 et rouge : k= -1).

Définition 1.2. *Domaine d'attraction*
Soit X_n *une suite de variables aléatoires indépendante et identiquement distribuées. S'il existe deux suites de constantes de normalisation* a_n $(a_n > 0)$ *et* $b_n \in \mathbb{R}$ *telles que :*

$$\lim_{n\to\infty} P(\frac{X_n - b_n}{a_n}) = \lim_{n\to\infty} F^n(a_n x + b_n) = H(x)$$

pour toutes les valeurs de x *pour lesquelles* H *est continue.* F *est dite appartenir au domaine d'attraction de* H $(F \in D(H))$.

L'unification du comportement du maximum en une seule fonction de répartition facilite grandement l'étude du comportement du maximum. Cette loi dépend du seul paramètre de forme k appelé indice des valeurs extrêmes ou indice de queue. Comme on le verra tout au long de ce mémoire, k est le paramètre clé de toute la théorie des valeurs extrêmes. L'estimation de k nous fournira le comportement de la queue de distribution.
Rappelons la définition d'une distribution d'une queue lourde.

Dans la théorie des probabilités, les distributions à queue lourde sont les distributions qui ont des queues non exponentiellement bornées, i.e., qui ont des queues plus lourdes que celles des distribution exponentielles.

Définition 1.3. *Une distribution F d'une varibale aléatoire X est dite à queue lourde si la fonction génératrice de X, $M_X(t)$, est infinie pour tout $t > 0$:*

$$\int_{\mathbb{R}} \exp(tx) dF(x) = \infty, \ \forall t > 0$$

Ce qui implique que :

$$lim_{x\to\infty} \exp(tx) P[X > x] = \infty, \ \forall t > 0.$$

Remarques 1.2. *On dit qu'une fonction non négative est à queue lourde si elle est bornée par une fonction exponentielle décroissante. Parmis les distriutions à queue lourde on retrouve la distribution de Pareto, la distribution de Cauchy et la distribution de Weibull.*

Remarques 1.3. *Selon le signe de k, on distingue trois domaines d'attraction.*

- *si $k > 0$, on dit que F appartient au domaine d'attraction de Fréchet [[17]], que l'on notera $D(\Phi_k)$. Il contient les lois à queues lourdes ou lois de type Pareto. Ces lois ont un point terminal x_F infini.*
- *si $k = 0$, on dit que F est dans le domaine d'attraction de Gumbel, que l'on notera $D(\Lambda)$. Il contient les lois à queues légères, i.e., les lois à décroissance exponentielle.*
- *si $k < 0$, on dit que F appartient au domaine d'attraction de Weibull, que l'on notera $D(\Psi_k)$. Toutes les lois de ce domaine d'attraction ont un point terminal x_F fini.*

Un classement de nombreuses lois par domaine d'attraction est disponible dans Embrechts et al. [14]. On donne une liste non exhaustive de l'appartenance des lois à leur domaine d'attraction dans le Tableau 1.1.

Fréchet $k > 0$	Gumbel $k = 0$	Weibull $k < 0$
Pareto	Normale	Uniforme
Student	Exponentielle	Beta
Cauchy	Gamma	

Tableau 1.1 : Quelques lois et leurs domaines d'attractions

Dans ce qui va suivre on donnera les conditions pour que F appartienne à l'un des trois domaines d'attraction ainsi que ses suites de normalisation (an) et (bn). Ces conditions étant basées sur la notion de fonctions à variations régulières, on commencera par rappeler la définition de ces fonctions .

1.3.2 Caractérisation des domaines d'attraction

On définit les notions de fonctions à variations régulières et de fonctions à variations lentes qui nous seront utiles par la suite.

Définition 1.4. *Fonctions à variations régulières*
Une fonction $F : \mathbb{R}_+ \to \mathbb{R}_+$ mesurable au sens de Lebesgue, est à variations régulières si et seulement si, il existe un réel α tel que pour tout $t > 0$ on a :

$$Lim_{x\to\infty}\frac{F(tx)}{F(x)} = t^{\alpha}$$

On dit que la fonction F est à variations régulières d'indice $\alpha \in \mathbb{R}$ à l'infini, et on notera par la suite $F \in \mathcal{RV}$.

Définition 1.5. *Fonctions à variations lentes*
Une fonction $F : \mathbb{R}_+ \to \mathbb{R}_+$ mesurable au sens de Lebesgue, est à variations lentes si et seulement si, il existe un réel α tel que pour tout $t > 0$ on a :

$$Lim_{x\to\infty}\frac{F(tx)}{F(x)} = 1$$

Remarques 1.4.

- *Remarquons qu'une fonction à variations régulières à l'infini d'indice $\alpha = 0$, est une fonction à variations lentes à l'infini , et on notera par la suite $F \in \mathcal{RV}_l$.*
- *Toute fonction à variations régulières sera notée L(x) et tout fonction à variations lentes sera notée l(x) ;.*
- *On peut facilement montrer que toute fonction à variations régulières d'indice $\alpha \in \mathbb{R}$ peut s'écrire sous la forme :*

$$L(x) = l(x)x^{\alpha}.$$

Ce résultat montre que l'étude des fonctions à variations régulières à l'infini se ramène à l'étude des fonctions à variations lentes à l'infini.

A l'aide des definitions des fonctions à variations régulières et variations lentes à l'infini, on va pouvoir caractériser les différents domaines d'attraction. Sachant la distribution F, on voudrait connaître son domaine d'attraction et ses constantes de normalisation. On rappelle les conditions nécessaires et suffisantes sur une fonction de répartition pour qu'elle appartienne à un domaine d'attraction.

Domaine d'attraction de la loi de Fréchet

Théorème 1.2. *Une distribution F appartient au domaine de Fréchet si et seulement si :*

$x_F = \infty$ *et si la fonction de distribution* $F = 1 - x^{-\alpha} l(x)$ *(F est à variations régulières d'indice* $-\alpha$ *) i.e*

$$\forall x > 0 \ , \ Lim_{x \to \infty} \frac{1 - F(tx)}{1 - F(x)} = t^{-\alpha}$$

Les suites de normalisation (a_n) *et* (b_n) *sont données dans ce cas pour tout* $n > 0$ *par :*

$$a_n = 0 \ et \ b_n = \overleftarrow{F}(1 - \frac{1}{n})$$

Domaine d'attraction de Weibull

Théorème 1.3. *une distribuion F appartient au domaine de Weibull si et seulement :* $x_F < \infty$ *et* $F(x_F - \frac{1}{x}) = x^{\alpha} l(x)$

Les constantes de normalisation sont données par :

$$a_n = x_F \ et \ b_n = \overleftarrow{F}(1) - \overleftarrow{F}(1 - \frac{1}{n})$$

Domaine d'attraction de Gumbel

Théorème 1.4. *Une distribution F appartient au domaine d'attraction de la distribution de Gumbel si et seulement si :*

$$E(X/X > c) < \infty \ pour \ c < \overleftarrow{F}(1)$$

et

$$Lim_{t \to \overleftarrow{F(1)}} \frac{1 - F(t + XE(X - t/X > t))}{1 - F(t)} = \exp -x$$

Les constantes de normalisation sont données par :

$$a_n = \overleftarrow{F(1 - \frac{1}{n})} \ et \ b_n = E(X - a_n / X > t)$$

1.3.3 Méthode des maxima par blocs

Le théorème de Fisher-Tippet est la base de la théorie classique des valerus extrêmes, le théorème nous donne la seule possible loi limite du maximum convenablement normalisé (non dégénéré) d'un grand nombre d'observation i.i.d. De façon plus spécifique, un modèle des maxima par blocks est construit de la manière suivante :
Si on dispose de n observations $x_1, ..., x_n$, on commence par regrouper les données en k blocs de longueur l et on calcule le maximum sur chaque blocs(ou le minimum , selon l'objectif) :

$$mi = max(x_{(i-l)l+1}, ..., x_{il}) \; pour \; i \; \in \; 1, ..., k$$

Ces valeurs sont ensuite ajoustée à la loi GEV(i.e, on approche la loi de la variable aléatoire M_i par une loi GEV).Il faut alors trouver un bon compromis entre la taille des blocs l , qui doit être assez grande pour que l'approximation par la loi GEV soit réaliste, et le nombre de blocs k qui doit être assez grand pour avoir assez d'informations pour estimer les paramètres de la GEV.

Il est irréaliste de croire que seul le maximum de l'échantillon permet de modéliser le comportement des valeurs extrêmes. Les autres grandes valeurs de l'échantillon contiennent elles aussi de l'information sur la queue de la distribution. L'approche par dépassements de seuil est une alternative à la loi GEV dans la modélisation du comportement du maximum d'un échantillon se basant sur les "grandes valeurs" de l'échantillon.
Elle se base sur le résultat mathématique (que nous détaillerons ci-dessous), affirmant qu'il existe une équivalence entre la loi GEV et la loi des dépassements de seuil aussi appelée loi des excès.

1.4 La loi des excès

L'approche par dépassements de seuil, en anglais "Peaks-Over-Threshold approach" notée POT, repose sur l'utilisation des observations au dela d'un sueil.

1.4.1 Distribution Généralisée de Pareto (GPD)

Définition 1.6. *La distribution généralisée de Pareto (GPD)*
Une variable aléatoire X est dite avoir une distribution de Pareto généralisée $G_{k,\sigma}$ de paramètres k et σ (respectivement, paramètre de forme et d'échelle), si la fonction de distribution est donnée par :

$$G_{k,\sigma}(x) = \begin{cases} 1 - (1 + k\frac{x}{\sigma})^{-\frac{1}{k}}, & k \neq 0\,; \\ 1 - \exp(-\frac{x}{\sigma}), & k = 0. \end{cases}$$

Si $k \neq 0$ alors $G_{k,\sigma}(x)$ est définie sur $x \geq 0$ et $1 + \frac{kx}{\sigma} \geq 0$.
Si $k = 0$ alors $G_{k,\sigma}(x)$ est définie sur $x \geq 0$.

où $\sigma > 0$ (paramètre d'échelle) et $k \in \mathbb{R}$ (paramètre de forme).

La densité $g_{k,\sigma}$ de la $G_{k,\sigma}$ s'obtient en dérivant la fonction de répartition. Pour $k \neq 0$:

$$g_{k,\sigma}(x) = \begin{cases} \frac{1}{\sigma}(1 + k\frac{x}{\sigma})^{-\frac{1}{k}-1}, & si\ x > 0\ et\ 1 + k(\frac{x}{\sigma}) > 0\,; \\ 0, & sinon. \end{cases}$$

Les deux premiers moments de la GPD sont donnés par :

$$E(X) = \frac{\sigma}{1-k} \quad avec \quad k < 1 \qquad V(X) = \frac{\sigma^2}{(1-k)^2(1-2k)} \quad avec \quad k < \frac{1}{2}$$

Remarques 1.5.
- *pour $k = 0$, on retrouve la loi exponentielle au paramètre σ*
- *pour $k = 1$, on retrouve la loi uniforme $U[0, \sigma]$*
- *pour $k < 0$, on retrouve la loi de Pareto*

Les densités associées à la loi de Pareto généralisée pour k fixé sont présentées dans la figure 1.5

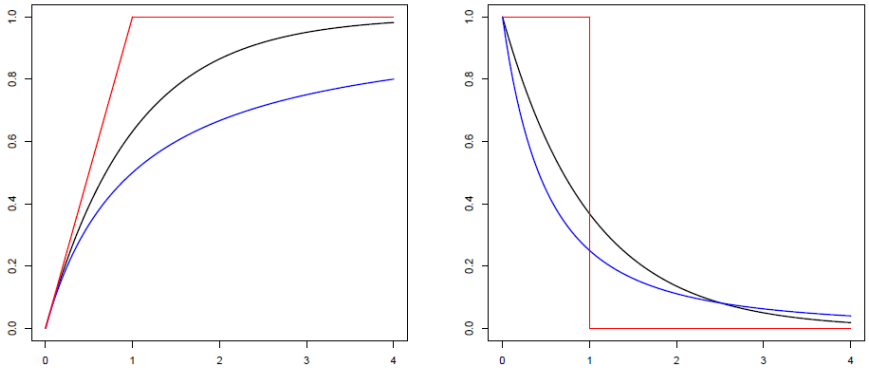

Figure 1.5 [28] : A gauche : $G_{k,\sigma=1}$. A droite : les densités associées à la loi de Pareto généralisée (noir : = 0, bleu : k= 1 et rouge : k= -1).

Remarques 1.6. *Voici quelques propriétés importantes de la distribution de la Pareto généralisée méritent d'être mentionnées :*

- *En comparant avec la distribution exponentielle, la GPD a une queue lourde pour $k < 0$ et une queue plus fine pour $k > 0$.*
- *Le taux de déffaillance $r(x) = \frac{f(x)}{1-F(x)}$ est donné par $r(x) = \frac{1}{\sigma - kx}$ et est monotone en x, si $k < 0$, alors r(x) décroit, si $k = 0$ r(x) est constant et si $k > 0$ alors r(x) croit*
- *Si une variable aléatoire X suit une Pareto généralisée, alors $P(X - \mu | X \geq \mu)$ est aussi une GPD*

Définition 1.7. *Fonction des excés*

Soit X une variable aléatoire , F sa fonction de distribution et x_F le point terminal. pour un $u < x_F$ fixé ,

$$F_u(x) = P(X - u \leq y | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \ x \geq 0$$

Soit $X_1, ...X_n$ un échantillon de variables aléatoires iid , F sa fonction de distribution telle que $F \in D(H)$.Soit u un seuil fixé (non aléatoire) et x_F un point terminal tel que $u < x_F$ fixé ; considérons les observations $x_1, ..., x_n$ dépassant le seuil u. On définit l'excés au-dessus du seuil u par $Y_1, ..., Y_{N_u}$,voir Figure 1.6 ci-dessous :

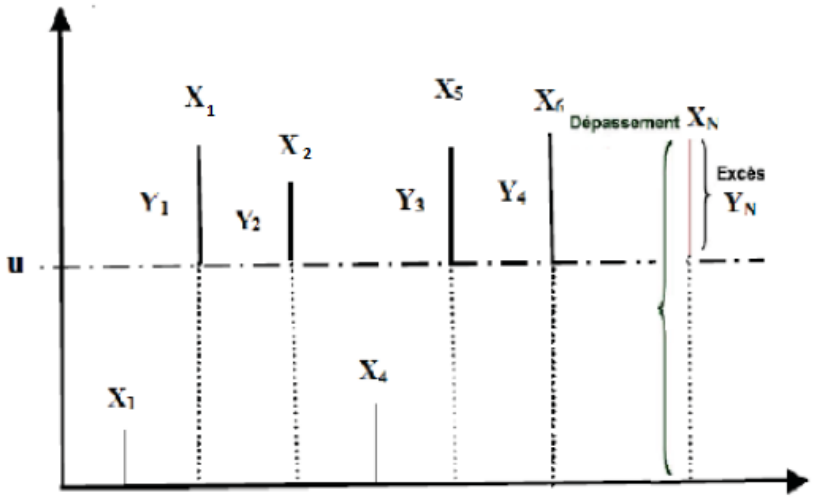

Figure 1.6 : Dépassement du seuil

On cherche un modèle paramètrique pour décrire la forme de la fonction de répartition F d'une variable aléatoire X au-dessus d'un niveau u élevé. D'aprés le théorème de Fisher et Tippet [[17]], il est naturel de supposer que pour un n assez "grand" :

$$F^n(x) \simeq \exp -(1 + k(\frac{x-\mu}{\sigma}))^{-\frac{1}{k}}$$

et donc :

$$nlogF(x) \simeq -(1 + k(\frac{x-\mu}{\sigma}))^{-\frac{1}{k}}.$$

Comme on s'intéresse à la queue de la distribution ($x \geq u$ avec u grand implique que $F(x) \approx 1$) on a :

$$logF(x) \simeq -1 - F(x)$$

et donc finalement

$$1 - F(x) \simeq \frac{1}{n}[1 + k(\frac{x-\mu}{\sigma})]^{-\frac{1}{k}}.$$

En pratique, tronquer au niveau u revient à s'intéresser à la loi conditionnelle

$$P[X \geq u + y / X \geq u]$$

avec $y \geq 0$.
En utilisant l'approximation précédente, on obtient alors :

$$P[X \geq u + y/X \geq u] \approx [1 + \frac{ky}{\tilde{\sigma}}]^{-\frac{1}{k}}$$

avec $\tilde{\sigma} = \sigma + k(u - \mu)$. Finalement on obtient pour une loi conditionnelle de $X - u$ (les dépassements du niveau u) sachant $X \geq u$

$$P[X - u \leq y/X \geq u] \approx 1 - [1 + \frac{ky}{\tilde{\sigma}}]^{-\frac{1}{k}}$$

Les travaux de Balkema et de Pickands [[33]] donnent un résultat très précis sur l'approximation de cette fonction lorsque le seuil u est proche du point terminal x_F.

Théorème 1.5. *Téorème de Pickands-Balkema-de Haan : [[33]]*
Soit $X_1, ..., X_n$ un echantillon de variables aléatoires i.i.d. suivant la loi F. appartenant au domaine d'attraction de $H_{k,\mu,\sigma}$, alors :

$$\lim_{n \uparrow x_F} \sup_{0 < x < x_F - u} |[F_u(X) - G_{k,\sigma}(x)]| = 0$$

ou $x_F \leq \infty$ est le point terminal de F et $G_{k,\sigma}(x)$ désigne la distribution généralisée de Pareto(GPD).

En résumé le Théorème 1.2 suggère donc d'approcher la loi des excès de l'échantillon par une distribution GPD lorsque le seuil u est choisi suffisament grand.

1.4.2 Estimation du quantile extrême par l'approche par dépassement du seuil

Définition 1.8. *Le quantile d'ordre $1 - \alpha$ de la fonction de répartition F est défini par :*

$$q(\alpha) = \overleftarrow{\overline{F}}(\alpha) = \inf\{x : \overline{F}(x) > 1 - \alpha \quad avec \quad \alpha \in]0, 1[\}$$

où $\overleftarrow{\overline{F}}$ est l'inverse généralisé de $\overline{F}$. Par convention $\inf\{\varnothing\} = \infty$. Notons que l'inverse généralisé d'une fonction coïncide avec l'inverse classique lorsque la fonction est srictement croissante.

Définition 1.9. *Soit une suite α_n, le quantile extrême d'ordre $1-\alpha_n$ de la fonction de répartition F est défini par :*

$$q(\alpha_n) = \overleftarrow{\overline{F}}(\alpha_n) \quad avec \quad \alpha_n \to 0 \quad quand \quad n \to \infty$$

En résumé, un quantile sera dit extrême si l'on remplace son ordre α par une suite α_n quand $n \to \infty$. Le fait que l'ordre $\alpha_n \to 0$ quand $n \to \infty$ indique que l'information la plus importante pour estimer des quantiles extrêmes est contenue dans la queue de distribution.

L'approche par la loi des excès est basée sur l'idée suivante. On a, d'après la définition 1.5, pour tout$x \geq 0$ la relation :

$$F_u(x) = P(Y \leq x/X > u) = P(X - u \leq x|X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}$$

Ou de manière equivalente :

$$\overline{F}_u(x) = 1 - F_u(x) = \frac{\overline{F}(x+u)}{\overline{F}(u)}$$

On obtien alors :

$$\overline{F}(x+u) = \overline{F}_u(x)\overline{F}(u)$$

Si on effectue le changement de variable y = u+x, alors l'approximation de la queue de distribution donne :

$$\overline{F}(y) = \overline{F}(u)\overline{F}_u(y-u) = \overline{F}(u)\overline{G}_{k,\sigma}(x-u)$$

Où $G_{k,\sigma}$ est la fonction de répartition de la loi de Pareto généralisée $G_{k,\sigma}$ qui nous est donnée dans le Théorème 1.1. On introduit alors la probabilité p que X dépasse le seuil u, $p = P(X > u) = \overline{F}(u)$, d'où :

$$\overline{F}(y) \approx p\overline{G}_{k,\sigma}(y - \overleftarrow{\overline{F}}(p)$$

avec $u = \overleftarrow{\overline{F}}(p)$, on obtient ainsi pour $k \in \mathbb{R}$:

$$\overline{F}(y) \approx p(1 + k(\frac{y - \overleftarrow{\overline{F}}(p)}{\sigma}))^{\frac{-1}{k}}$$

Le cas $k = 0$ peut être vu comme le cas limite $k \to 0$ dans l'équation précédente :

$$\overline{F}(y) \approx p \exp\left(-\frac{y - \overline{F}^{-1}(p)}{\sigma}\right)$$

On souhaite estimer des quantiles. Or, par définition du quantile, il nous faut inverser la fonction de l'équation précédente ce qui nous donne :

$$q(\alpha) = \overleftarrow{\overline{F}}(p) + \frac{\sigma}{k}((\frac{\alpha}{p})^{-k} - 1)$$

On fait tendre $k \to 0$, on obtient :

$$q(\alpha) = \overleftarrow{\overline{F}}(p) - \sigma\frac{\alpha}{p}$$

Définition 1.10. *L'estimateur du quantile extrême de la loi GPD est défini par :*

$$\widehat{q}(\alpha_n) = X_{n-N_u+1,n} + \frac{\widehat{\sigma}}{\widehat{k}}((\frac{N_u}{n\alpha_n})^{\widehat{k}} - 1)$$

où $\widehat{k}$ et $\widehat{\sigma}$ sont des estimateurs des paramètres de forme et d'échelle et N_u le nombre d'excès.

Différentes valeurs du seuil u donneront différents échantillons d'excès plus ou moins grands, ce qui influencera l'estimation des quantiles extrêmes.
Ce seuil doit être suffisament grand pour que l'on puisse appliquer le Théorème 1.2 mais pas trop car sinon on utilisera peu d'excès donc peu d'informations.

1.4.3 Choix du seuil

Comme le font remarquer Davison et Smith [[10]], on ne dispose pas d'outils théoriques permettant de choisir de manière optimale le seuil. Sa détermination reste donc empirique. Généralement, on le détermine graphiquement. En utilisant le ME-plot (Le Mean Excess Plot).
Le choix du seuil doit établir un compromis entre biais et variance. Concrétement, le seuil doit être suffisamment grand pour pouvoir utiliser les résultats asymptotiques mais pas trop élevé pour obtenir des estimations précises. Par contre le choix d'un seuil faible risque de déclarer abusivement des observations extrêmes, ce qui peut induire un biais

dans l'estimation et par conséquent, mal approximer la loi asymptotique. Dans ce sens, plusieurs méthodes de détection du seuil ont été proposées ; nous utilisons celle qui est la plus employée à savoir le graphe d'éxcés en moyenne (mean excess plot ou mean residual life plot (ME-plot)).

Le Mean Excess Plot (ME-plot) :

Le Mean Excess Plot est le graphe des points $(u, e_n(u)), X_{1:n} < u < X_{n:n}$ où $e_n(u)$ est définis par :

$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)I_{(X_i>u)}}{\sum_{i=1}^{n} I_{(X_i>u)}} = \frac{1}{N_u}\sum_{i=1}^{n}(X_i - u)I_{(X_i>u)}$$

c'est-à-dire la somme des excès au-dessus du seuil u divisé par le nombre N_u de données qui excèdent u. La sample mean excess function $e_n(u)$ est l'estimateur empirique de la fonction de la moyenne des excés

$$e(u) = E[(X - u)|X > u]$$

Comment interprète-t-on cette figure ? Il faut savoir tout d'abord que la fonction de la moyenne des excés de la GPD est donnée par :

$$e_u(x) = E(X - u|X > u) = \frac{ku + \sigma}{1 - k}$$

où $\sigma + ku > 0$. Ainsi, si le ME-plot semble avoir un comportement linéaire au-dessus d'une certaine valeur de u, cela signifie que les excès au-dessus de ce seuil suivent une GPD.

Dans l'exemple de la Figure 1.8 (pertes en valeur absolue d'un indice boursier), nous pouvons définir le seuil à 1

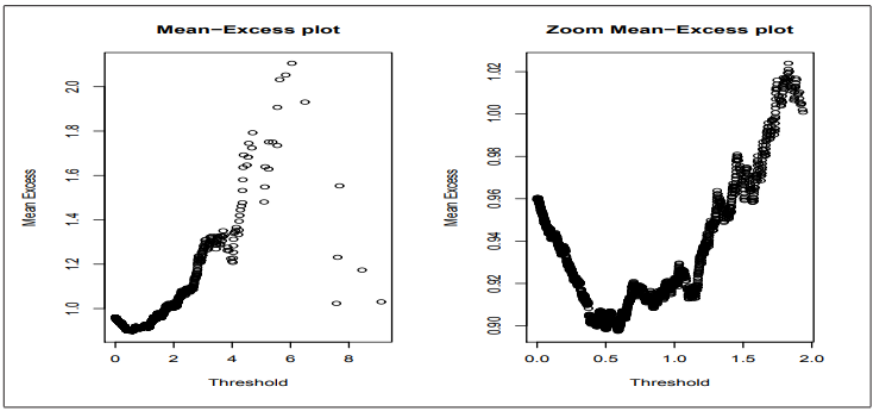

Figure 1.8 : Mean-Excess plots pour la queue des pertes

1.4.4 Méthode de dépassement de seuil

Dans la méthode des maxima par blocs, des blocs de taille identique sont constitués puis pn considère la distribution fermée par les maxima de chcune des blocs (une seule valeur maximale par bloc). Cette méthode implique donc généralement d'informations sur les événements extrêmes. Par exemple, certains blocs peuvent contenir plusieurs valeurs extrêmes intéressantes alors que d'autres peuvent ne pas en contenir. En pratique, cela signifie qu'il faut beaucoup de données pour pouvoir mettre en place la méthode des maxima par blocs, ce qui n'est généralement pas le cas. Une altérnative à la méthode des maxima par blocs consiste à conserver toutes les observations qui dépassent un niveau élevé puis à ajuster une loi appropriée à ces dépassements qui représente les événements "extrêmes". Cette méthode est généralement appelée méthode des dépassements de seuil (ou "Peak Over Threshold", POT).

Chapitre 2

Méthodes d'estimation des paramètres de la GPD

2.1 Introduction

L'estimation des paramètres d'un modèle statistique est une technique importante, sous-jacente dans l'analyse statistique. Bien que le statisticien peut effectuer certaines analyses intuitives, l'estimation nécessite une méthode spécifique. Dans cette section, nous présentons cinq méthodes d'estimation des paramètres de la loi de Pareto généralisée et nous en appliquons seulement quatre. à savoir la méthode basée sur le maximum de vraisemblance (MLE), la méthode des moments (MOM), la méthode des moments pondérés (PWM) et la méthode basée sur le principe du maximum d'entropie (POME).

2.2 Méthode du maximum de vraisemblance (MLE)

Nous allons appliquer la méthode du maximum de vraisemblance à la loi $GPD_{k,\sigma}$. Soit $X_1, X_2, ..., X_n$ n variables aléatoires de loi de Pareto généralisée. Rappelons d'abord l'expression de la fonction de densité d'une Pareto généralisée :

$$g(x) = \frac{1}{\sigma}\left(1 + \frac{kx}{\sigma}\right)^{\left(-\frac{1}{k}-1\right)}$$

Pour $x \in \mathbb{R}^+$ si $k > 0$, et $x \in [0\ ,\ \frac{\sigma}{k}]$,si $k < 0$.
Pour $k = 0$,

$$g(x) = \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right)$$

On en déduit alors la fonction de log-vraisemblance de la loi GPD pour :

$$k \neq 0 \quad , \quad l(k,\sigma) = \ln L(k,\sigma) = -n\ln(\sigma) - (1+\frac{1}{k})\sum_{i=1}^{n} \ln(1+\frac{kx_i}{\sigma})$$

Les estimateurs du maximum de vraisemblance de k et σ (lorsqu'ils existent) sont donnés par la résolution du système :

$$\begin{cases} \frac{\partial l(k,\sigma)}{\partial k} = 0, \\ \\ \frac{\partial l(k,\sigma)}{\partial \sigma} = 0, \end{cases}$$

On obtient alors :

$$\frac{\partial l(k,\sigma)}{\partial k} = \frac{1}{k^2}\sum_{i=1}^{n} \ln(1+k\frac{x_i}{\sigma}) - (\frac{1}{k}+1)\sum_{i=1}^{n} \frac{x_i}{\sigma + kx_i}$$

En annulant la dérivée , on aura :

$$\frac{\partial l(k,\sigma)}{\partial k} = 0 \quad \Rightarrow \quad \frac{1}{k^2}\sum_{i=1}^{n} \ln(1+k\frac{x_i}{\sigma}) - (\frac{1}{k}+1)\sum_{i=1}^{n} \frac{x_i}{\sigma + kx_i} = 0 \quad (*)$$

l'équation $(*)$ n'admet pas de solution explicites . En utilisant la reparamétrisation $\tau = \frac{k}{\sigma}$, on obtient :

$$l(k,\tau) = \ln L(k,\tau) = -n\ln(\frac{k}{\tau}) - (1+\frac{1}{k})\sum_{i=1}^{n} \ln(1+\frac{\tau}{x_i})$$

La dérivée de la fonction log-vraisemblance par rapport à k et τ sont donnés par :

$$\begin{cases} \frac{\partial l(k,\tau)}{\partial k} = -n\frac{1}{k} + \frac{1}{k^2}\sum_{i=1}^{n} \ln(1+\tau x_i) \\ \\ \frac{\partial l(k,\tau)}{\partial \tau} = n\frac{1}{\tau^2} - (1+k)\sum_{i=1}^{n} \frac{x_i}{1+\tau x_i}, \end{cases}$$

Les estimateurs du maximum de vraisemblance $(\hat{k}^{MLE}, \hat{\tau}^{MLE})$ sont donnés par la résolution du système :

$$\begin{cases} \frac{\partial l(k,\tau)}{\partial k} = 0 \\ \\ \frac{\partial l(k,\tau)}{\partial \tau} = 0, \end{cases}$$

En maximisant $l(k, \tau)$ par des méthodes numériques de manière itérative ,en utilisant des méthodes telles que l'algorithme de Newton-Raphson [[24]], pour autant que l'on dispose d'une valeur initiale τ_0 pas trop éloigné de τ . On obtiendra :

$$\hat{k}^{MLE} = \frac{1}{n} \sum_{i=1}^{n} \ln(1 - \hat{\tau}^{MLE} X_i)$$

et

$$\hat{\sigma}^{MLE} = \frac{\hat{k}^{MLE}}{\hat{\tau}^{MLE}}$$

Sous l'hypothèse $k > -1/2$, les estimateurs du maximum de vraisemblance sont asymptotiquement gaussiens et efficaces, avec une matrice variance-covariance :

$$\begin{bmatrix} 2\sigma^2(1-k) & \sigma(1-k) \\ \sigma(1-k) & (1+k)^2 \end{bmatrix}$$

([[40]] pour plus de détails). Les estimateurs obtenus sont cependant peu utilisés en pratique car ils sont peu performants sur des échantillons de petite taille ($n < 500$) (voir Davison, A. et Smith, R. [[10]]).

2.3 Méthode des moments (MOM)

Hosking et Wallis [[24]] ont été les premiers à proposer la méthode des moments pour estimer les deux paramètres de la GPD.Pour $k < \frac{1}{2}$, on sait que :

$$E(X) = \frac{\sigma}{1-k} \quad et \quad Var(X) = \frac{\sigma^2}{(1-k)^2(1-2k)}$$

On peut alors facilement exprimer les paramètres de la loi GPD k et σ en fonction de l'espérance et de la variance de X, soit :

$$k = \frac{1}{2}(1 - \frac{E(X)^2}{Var(X)}) \quad et \quad \sigma = \frac{E(X)}{2}(1 + \frac{E(X)^2}{Var(X)})$$

Ainsi en remplaçant $E(X)$ et $Var(X)$ par leurs moments empiriques :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad et \quad s^2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

On obtient les estimateurs des moments de k et σ (respectivement, k^{MOM} et σ^{MOM}) , soit :

$$\hat{k}^{MOM} = \frac{1}{2}(1 - \frac{\bar{X}^2}{s^2(X)}) \quad et \quad \hat{\sigma}^{MOM} = \frac{\bar{X}}{2}(1 + \frac{\bar{X}^2}{s^2(X)})$$

Ces estimateurs sont asymptotiquement gaussiens si $k < \frac{1}{4}$ .Leur matrice variance-covariance est donnée par :

$$\frac{(1+k)^2}{(1+2k)(1+3k)(1+4k)} \begin{bmatrix} 2\sigma^2(1+6k+12k^2) & \sigma(1+2k)(1+4k+12k^2) \\ \sigma(1+2k)(1+4k+12k^2) & (1+2k)^2(1+k+6k^2) \end{bmatrix}$$

(voir [24]) pour plus de détails).Leur domaine de validité est très limité (domaine d'existence [28]) et bien que ces estimateurs soient assez simples à calculer, la méthode implique l'élévation au carré des observations de l'échantillon, qui, dans le cas d'une distribution à queue lourde, peut augmenter les erreurs d'échantillonnage [32].

2.4 Méthode des moments de probabilité pondérée (PWM)

Les moments de probabilité pondérés(PWM) ont été introduits par Greenwood et al. [**?**]) comme outil d'estimation des paramètres des distributions. Les PWM de la variablé

aléatoire X sont définis comme suit :

$$M_{p,r,s} = E(X^p(F(X))^r(1-F(X))^s) = \int_\infty (x)^p F^r(x)(1-F(x))^s \; dF(x),$$

Pour p,r,s $\in \mathbb{N}$.

L'intégrale ci-dessus peut souvent être résolus analytiquement de façon à ce que les PWM puissent être exprimé en fonction des paramètres de la distribution. Les estimateurs des paramètres peuvent être obtenus en égalisant la forme analytique des PWM avec les moments emperiques ponderés [[34]].

En pratique, pour estimer les paramètres on peut p=1. La procédure d'estimation des moments de probabilités pondérés est la même que la méthode d'estimation classique, à savoir la méthode des moments. Cependant elle présente un avantage en évitant l'élévation au carré des observations, ce qui, dans certains cas peut attribuer un poids excessif à des observations extrêmes [[34]].

La définition des PWM donnée précédemment est valable pour toutes les valeurs de p,r et s tant que l'intégrale existe. Toutefois, généralement on considère les moments de probabilités pondérés pour p=1 et r et s comme des entiers non négatifs choisis aussi petits que possible. Hosking [[25]] définit les PWM α_r et β_r

$$\alpha_r = M_{1,0,r} = EX[1-F(X)]r \quad r = 0, 1, ...,$$

$$\beta_r = M_{1,r,0} = EX[F(X)]^r \quad r = 0, 1, ...,$$

En particulier ,$\alpha_0 = \beta_0 = E(X)$. Puisque r est un entier non negatif , α_r et β_r peuvent être facilement exprimés en fonction l'un et l'autre :

$$\alpha_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \beta_k$$

$$\beta_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \alpha_k$$

Où $\binom{r}{k}$ est le coefficient binomial.

Par exemple pour l'estimation de deux paramètres d'une distribution, la procédure traditionelle impliquerait la prise en compte de (α_0, α_1) ou (β_0, β_1). Le choix peut être fait arbitrairement parce que (α_0, α_1) peuvent être exprimés en fonction (β_0, β_1) (et vice-versa) et les deux peuvent aboutir aux même estimateurs [[34]].

Les quantités F^r et $(1 - F)^r$ peuvent être interprétées comme des poids affectés à l'échantillon ordonné. Si F est le poids correspondant à β_1, on pondère plus les grandes observations que les petites. Respectivement, si (1-F) est le poids correspondant à α_1, on pondère plus les peites observations que les grandes. Toutefois, lorsqu'il est utilisé conjointement avec la moyenne , α_0 ou β_0, il n'y a pas de différence entre l'utilisation de α_1 et β_1. Lors de l'estimation de deux paramètres d'une distribution , l'utilisation des moments (α_0, α_1) et (β_0, β_1) donnent un poids similaire à celui des petites et grandes observations. La conclusion ci-dessus n'est vraie que dans la mesure où les PWM avec des entiers non négatifs r et s,sont choisis aussi petits que possibles. Par exemple, l'utilisation de (α_0, α_2) et (β_0, β_2) ne conduirait pas aux même estimateurs [[34]].

Les moments de probabilités pondérés sont données par α_r :

$$\alpha_r = M_{1,0,r} = \frac{\sigma}{(r+1)(r+1-k)} \qquad k < 1$$

α_0 et α_1 qui sont donnés par :

$$\alpha_0 = M_{1,0,0} = E(X) = \frac{\sigma}{1-k}$$

et

$$\alpha_1 = M_{1,0,1} = E(X(1 - F(X))) = \frac{\sigma}{2(2-k)}$$

On peut alors facilement exprimer les paramètres k et σ de la loi GPD en fonction de α_0 et α_1 :

$$\sigma = \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1}, \qquad k = 2 - \frac{\alpha_0}{\alpha_0 - 2\alpha_1}$$

Les estimateurs des moments de probabilités pondérés $\hat{k}^{PWM}$ et $\hat{\sigma}^{PWM}$ sont obtenus en remplaçant par les estimateurs basés sur les moments empiriques pondérés qui sont donnés par :

$$\hat{\alpha}_r = n^{-1} \sum_{i=1}^{n} \frac{(n-i)(n-i-1)...(n-i-r+1)}{(n-1)(n-2)...(n-r)} X_{i:n}$$

On obtient les estimateurs :

$$\hat{k}^{PWM} = \frac{4\hat{\alpha}_1 - \hat{\alpha}_0}{2\hat{\alpha}_1 - \hat{\alpha}_0} \quad et \quad \hat{\sigma}^{PWM} = \frac{2\hat{\alpha}_0\hat{\alpha}_1}{\hat{\alpha}_0 - 2\hat{\alpha}_1}$$

Hosking et Wallis [[24]] ont montré que les estimateurs des moments pondérés sont asymptotiquement gaussiens pour $-1 < k < 1/2$ avec une matrice de variance-covariance :

$$\frac{1}{(1+2k)(3+2k)} \begin{bmatrix} \sigma^2(7+18k+11k^2+2k^3) & \sigma(2+k)(2+6k+7k^2+k^3) \\ \sigma(2+k)(2+6k+7k^2+2k^3) & (1+k)(2+k)^2(1+k+2k^2) \end{bmatrix}$$

et ont étudié leurs performances sur simulations. Pour des applications pratiques leur domaine de validité reste limité (domaine d'existence limité $-1 < k < 1/2$).

2.5 Méthode des L-moments

La méthode des L-moments introduite par Hosking [23] a connu une grande application en analyse fréquentielle des données hydrologiques. Les L-moments sont des combinaisons linéaires des statistiques d'ordre . Avant de présenter cette méthode, nous allons donner quelques définitions :

2.5.1 Les U-Statistiques

Les U-Statistiques sont une classe de statistiques présentant des propriétés particulièrement intéressantes. Introduites et développées par Hoeffding [22] et Arvesen [2] , elles permettent de construire des classes d'esimateurs sans biais.

Soit X une variable aléatoire d loi P_θ, $\theta \in \mathbb{R}$.

Définition 2.1. *On appelle degré d'un paramètre réel θ le nombre m tel que :*

$$m = \inf\{n \in \mathbb{N} | E[T(X_1, X_2, ..., X_n)] = \theta\}$$

où T est une statistique.
Le degré d'un paramètre est la plus petite taille d'échanillon assurant l'existance d'un estimateur sans biais de θ.

Définition 2.2. *On appelle noyau d'un paramètre réel θ de degré m la statistique T définie sur un échantillon de taille m qui soit sans biais pour θ :*

$$T : \mathbb{R}^m \to \Theta \text{ telle que } E(T) = \theta$$

$$\text{avec } : \ E(T(X_{i(1),...,X_{i(m)}})) = \theta \text{ et } (i(1), ..., i(m)) \in (1, ..., n).$$

Exemple 2.1. $\theta = E(x)$
m=1
$T(X_i) = X_i$
$E(T) = E(X)$

Soit θ un paramètre de degré m et de un noyau T.

Définition 2.3. *Le noyau symètrique est une statistique T^* définie sur $\mathbb{R}^m$ par :*

$$T^*(X_{i(1)}, ..., X_{i(m)}) = \sum_{\pi} T(X_{i(1)}, ..., X_{i(m)}) m!$$

où le symbole de $\sum_\pi$ désigne la sommation sur les m! permutations possibles de (i(1),...,i(m)).

Exemple 2.2. $\theta = V(x)$
$T^*(X_i, X_j) = \frac{X_i^2 + X_j^2 - 2X_iX_j}{2}$
$T^*(X_i, X_j) = \frac{X_i^2 - X_j^2}{2}$
$E(T) = V(X)$

Soit θ un paramètre de degré m, de noyau T, de noyau symètrique T^*

Définition 2.4. *On appelle U-statistique la quantité :*

$$U(X_1, ..., X_n) = \frac{\sum_{i(1)<,...,i(m)} T^*(X_{i(1)}, ..., X_m)}{\binom{m}{n}}$$

Exemple 2.3. *a)* $\theta = E(x)$
$m = 1;\ T(X_i, X_j) = \frac{(X_i - X_j)^2}{2};\ T^* = T$
$U = \sum \frac{X_i}{n} = \overline{X}$

2.5.2 Principe de la méthode des L-moments

Définition 2.5. *Soient $X_{1:n}, X_{2:n}, ..., X_{n:n}$ les statistiques d'ordre $1, 2, ..., n$ associées à un échantillon $X_1, ..., X_n$. Les L-moments de X sont définis par :*

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:n}) \qquad r = 1, 2, ...$$

où $E(X_{j:r}) = \frac{r!}{(j-1)!(r-j)!} \int_0^1 x(F)[F(x)]^{r-1}[1 - F(x)]^{n-r} dF(x)$

En remplaçant $E(X_{j:r})$ dans λ_r, on obtient :

$$\lambda_r = \int_0^1 x(F) P_{r-1}^*[F(x)] dF(x), \quad r = 1, 2, ...,$$

où

$$P_r^*[F(x)] = \sum_{k=0}^{r} p_{rk}^* [F(x)]^k \quad et \quad p_{rk}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}$$

$P_r^*[F(x)]$ étant la décomposition en série de polynômes de Legendre d'ordre r.On obtient alors :

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}\int_0^1 x(F)[F(x)]^{r-k-1}[1-F(x)]^k dF(x) \quad r \geq 2$$

Définition 2.6. *Soient $X_{1:n}, X_{2:n}, ..., X_{n:n}$ les statistiques d'ordre $1, 2, ..., n$ associées à un échantillon $X_1, ..., X_n$. Les quatre premiers L-moments sont donnés par :*

$$\lambda_1 = E(X) = \int x dF$$

$$\lambda_2 = E(X_{2:2} - X_{1:2}) = \int x(2F-1)dF$$

$$\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int x(6F^2 - 6F + 1)dF$$

$$\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X1:4) = \int x(20F^3 - 30F^2 + 12F - 1)dF$$

La distribution de probabilité peut être spécifiée par ces L-moment même si les moments conventionnels n'existent pas ; le contraire, cependant, n'est pas vrai.
On peut prouver que le premier L-moment est une caractéristique de localisation, le second étant une caractéristique de variabilité. Il est souhaitable de standardiser un L-moment d'ordre r λ_r,$r \geqslant 3$ pour qu'ils soient indépendants de l'unité de mesure de X. On définit alors les rapports des L-moments d'ordre r par :

$$\tau_r = \frac{\lambda_r}{\lambda_1} \quad r = 3, 4, ...$$

Une propriété des rapports des L-moments est qu'ils sont bornés ($|\tau_r < 1|$). λ_3 est une mesure de l'asymétrie et λ_4 est une mesure de l'aplatissement .
On peut aussi définir la fonction des L-moments [[5]] qui est l'analogie du coéfficient de variation dans le cas classique, i.e. appelé L-coéfficient de variation :

$$\tau_2 = \frac{\lambda_2}{\lambda_1}$$

L'estimation des paramètres par la méthode des L-moments est basée sur le même principe que celui des moments ordinaires. Les estimateurs sont la solution d'un système

d'équations obtenues en égalant les L-moments empiriques et les L-moments correspondants de la distribution.

Les L-moments sont généralement estimés par un échantillon aléatoire obtenu à partir d'une distribution inconnue. Puisque λ_r est la moyenne d'ordre r, il est naturel de l'estimer en utilisant les U-statistiques. Les L-moments empiriques défini par Hosking [[12]] comme des combinaisons linéaires des statistiques d'ordre sont donnés par :

$$l_r = \binom{n}{r}^{-1} \sum_{1<i_1<i_2<\ldots<i_n<n} \sum \ldots \sum \frac{1}{n} \sum_{k=0}^{n} (-1)^k X_{ir-k:n} \quad r = 1, 2, ..., n.$$

l_1 est la moyenne empirique et l_2 L-variance empirique. Naturellement on estime les rapports des L-moments par les rapports des L-moments empiriques qui sont donnés par :

$$t_r = \frac{l_r}{l_2}, \ r \geq 3$$

Les quatres premiers L-moments empiriques, s'ecrivent :

$$l_1 = \frac{1}{n} \sum_{i}^{n} X_{i:n}$$

$$l_2 = \frac{1}{2} \binom{n}{2}^{-1} \sum_{i>k} \sum (X_{i:n} - X_{k:n})$$

$$l_3 = \frac{1}{3} \binom{n}{3}^{-1} \sum_{i>k>j} \sum \sum (X_{i:n} - 2x_{k:n} + X_{j:n})$$

$$l_4 = \frac{1}{4} \binom{n}{4}^{-1} \sum_{i>k>j>l} \sum \sum \sum (X_{i:n} - 3X_{k:n} + 3X_{j:n} - X_{l:n})$$

Pour un échantillon $X_{1,n}, ..., X_{n:n}$, ordonné dans un ordre croissant, les estimateurs des L-moments peuvent être déduits de ceux des moments de probabilité pondérés définis par :

$$\lambda_1 = M_{1,0,0}$$
$$\lambda_2 = 2M_{1,1,0} - M_{1,0,0}$$
$$\lambda_3 = 6M_{1,2,0} - 5M_{1,1,0} + M_{1,0,0}$$
$$\lambda_4 = 20M_{1,3,0} - 3M_{1,2,0} + 12M_{1,1,0} - M_{1,0,0}$$

et les estimateurs des 4 premiers moments de probabilités pondérés sont donnés par :

$$\widehat{M_{1,0,0}} = \frac{1}{n}\sum_{i=1}^{n} X_{i:n} = \overline{X}$$
$$\widehat{M_{1,1,0}} = \frac{1}{n}\sum_{j=2}^{n} \frac{(j-1)}{(n-1)} X_{j:n}$$
$$\widehat{M_{1,2,0}} = \frac{1}{n}\sum_{j=3}^{n} \frac{(j-1)(j-2)}{(n-1)(n-2)} X_{j:n}$$
$$\widehat{M_{1,3,0}} = \frac{1}{n}\sum_{j=4}^{n} \frac{(j-1)(j-2)(j-3)}{(n-1)(n-2)(n-3)} X_{j:n}$$

2.5.3 Application de la méthode des L-moments à la GPD

Soit X une variable aléatoire qui suit une loi GPD dont les paramètre k et σ sont inconnus. Les estimateurs de ces paramètres sont la solution d'un système d'équations obtenues en égalant les L-moments empirique et les L-moments de la distribution.
$\lambda_i = l_i,\ i = 1, 2$

les 4 premiers moments d'une GPD :

$$\lambda_1 = \frac{\sigma}{k} - \beta(k-1)$$
$$\lambda_2 = \sigma(k-1)(1-2^{-k})$$
$$\lambda_3 = \sigma(k-1)(23^k + 32^{-k} - 1)$$
$$\lambda_4 = \sigma(k-1)(-54^{-k} + 103^{-k} - 62^{-k} + 1)$$

Les estimateurs des paramètres de la GPD sont donnés par :

$$\widehat{k}^{LM} = \frac{1-3t_3}{1+t_3} \quad et \;\; \widehat{\sigma}^{LM} = l_2(\widehat{k}^{LM}+1)(\widehat{k}^{LM}+2)$$

où $t_3 = \frac{1-k}{3+k}$

2.6 Méthode du maximum d'entropie (POME)

Le principe de maximum d'entropie (POME) renverse la présentation classique de la modélisation statistique au sens où il choisit en premier lieu les quantités statistiques que l'on juge essentielles pour résumer l'information apportée par un jeu de données. Le modèle, c'est-à-dire la loi de probabilité décrivant le phénomène aléatoire, n'apparaît qu'après et doit vérifier des contraintes mettant en jeu ces quantités statistiques essentielles. par exemple, dans certains domaines tels que l'hydrologie, il existe souvent des informations disponibles sur le phénomène à l'origine des données, qui peuvent être utilisées pour améliorer les performances de la procédure d'estimation. En fait, il est très courant que nous ne puissions pas exprimer notre connaissance de la distribution elle-même, mais que nous soyons tout à fait capables d'exprimer nos convictions sur certains aspects de la distribution, tels qu'un moment. conceptuellement, cette méthode est liée à la pensée bayésienne, dans le sens où nous essayons de surmonter l'inconnu en incorporant dans l'analyse toutes les informations disponibles. l'indisponibilité d'informations complètes est traitée ici comme un problème de maximisation sous contrainte.
La loi du maximum d'entropie est obtenue par maximisation d'une certaine fonctionnelle sur l'ensemble des lois pouvant servir de modèle. La fonctionnelle que nous avons choisie

est l'entropie de Shannon, car elle seule permet d'atteindre une loi qui possède la propriété de concentrer les lois empiriques dans son voisinage.

2.7 Méthode du maximum d'entropie (POME)

Shannon [10] a définit l'entropie comme une mesure numérique de l'incertitude,associée a une fonction de densité $f(x;\theta)$ de paramètre θ. La fonction d'entropie de shannon H(x) de X, ets donnée par :

$$H(x) \equiv H(f) = -\int_{-\infty}^{\infty} f(x;\theta) \ln f(x;\theta) dx, \ \int_{-\infty}^{\infty} f(x;\theta) dx = 1$$

H(x) est l'entropie de $f(x;\theta)$, elle est la moyenne de $-\ln f(x;\theta)$
Selon Jaynes [1], la fonction la moins biaisée de X est celle qui maximise l'entropie sous réserve d'une information donnée, ou qui satisfait le principe du maximum d'entropie (POME). Par conséquent, les paramètres de la distribution peuvent être obtenus en maximisant H(f). L'utilisation de ce principe pour générer la fonction la moins biaisée sur la base de données limitées et incomplètes a été discuté par plusieurs auteurs et a été appliqué dans plusieurs et divers domaines (voir l'étude faite par Fnigh et Fiorentino [12]). Jaynes [13] a donné un raisonement selon lequel la methode POME est le critère logique et rationnel pour le choix spécfique d'une fonction f(x) qui maximise H et satisfait l'information disponible qui est exprimée sous forme de contraintes. En d'autre termes, l'information disponible (la variance, la moyenne, l'asymetrie, la limite inférieure, la limite supérieure,...ect). la fonction dérivée par la méthode POME representerait le mieux X.Inversement, si l'on souhaite ajuster une fonction particulière à un échantillon de données, alors la méthode POME peut spécifier de façon unique les contrainte (ou les informations) nécessaires pour déduire cette fonction. Les paramètres de la fonction sont ensuite liés à ces contraintes. Une discussion sur la jusification mathématique est donnée dans Levine et tribus [14].
Etant donné m contraintes linéairement indépendantes C_i ,$i = 1, 2, ..., m$, définie par :

$$C_i = \int_b^a Y_i(x) f(x) dx, \quad i = 1, 2, ..., m$$

où les $Y_i(x)$ sont des fonctions dont les moyennes par rapport à f(x) sont spécifiées.
Le maximum de H sous les contraintes définié précédement, est donné par la distribution :

$$f(x) = \exp -a_0 - \sum_{i=1}^{m} a_i Y_i(x)$$

où $a_i, i = 1, 2, ..., m$, sont les multiplicateur de lagrange qui peuvent être déterminés par les contraintes et la distribution du maximum de H(f) sous les contraintes.
En insérant le maximum de H sous les contraintes dans la fonction d'entropie de shannon, on obtient l'entropie de f(x) en fonction des contraintes et des multiplicateurs de lagrange :

$$H(x) = a_0 + \sum_{i=1}^{m} a_i C_i$$

2.7.1 Méthode du maximum d'entropie appliquée à la distribution de la Pareto Généralisée

La maximisation de H établit la relation entre les contraintes et les multiplicateurs de Lagrange . Ainsi, pour dériver une méthode utilisant POME pour l'estimation des paramètres, trois étapes sont nécessaires :

- spécification des contraintes appropriées
- dérivation de la fonction entropique de la distribution
- dérivation des relations entre paramètres et contraintes

Une discussion mathématique complete de cette méthode peut être trouver dans [[41]], [[[28]], [[39]].

Spécification des contraintes

En prenant le logarithme de la densité de GPD, on obtient :

$$\ln f(x) = -\ln \sigma + \frac{1}{k} \ln(1 - k\frac{x}{\sigma}) - \ln(1 - k\frac{x}{\sigma})$$

En multipliant l'équation precédente par$[-f(x)]$ et en intégrant de 0 à ∞, on obtient l'entropie de la fonction $f(x)$:

$$H(x) = \ln \sigma \int_0^\infty f(x)dx - (\frac{1}{k} - 1) \int_0^\infty [1 - \frac{k}{\sigma}x]f(x)dx$$

Pour maximiser $H(x)$ dans l'équation précédente , les contraintes suivantes doivent satisfaire $H(x)$:

$$\int_0^\infty f(x)dx = 1$$
$$\int_0^\infty \ln[1-k\frac{x}{\sigma}]f(x)dx = E[\ln[1-k\frac{x}{\sigma}]]$$

Construction du multiplicateur zéro de Lagrange

La fonction de densité de la Pareto généralisée la moins biaisée et qui correspond à la méthode POME en satisfaisant les précédentes contraintes est donnée par :

$$f(x) = \exp[a_0 - a_1 \ln(1-k\frac{x}{\sigma})]$$

où a_0 *et* a_1 sont des multiplicateurs de Lagrange. En appliquant la premiere contrainte à l'équation précédente on obtient :

$$\exp(a_0) = \int_0^\infty \exp(-a_1 \ln[1-k\frac{x}{\sigma}])dx$$

ce qui nous donne :

$$\exp(a_0) = \frac{\sigma}{k}\frac{1}{1-a_1}$$

en prenant le logarithme on obtient le multiplicateur de Lagrange a_0 :

$$a_0 = \ln[\frac{\sigma}{k}\frac{1}{1-a_1}] = \ln\sigma - \ln k - \ln(1-a_1)$$

on peut aussi exprimer a_0 comme suit :

$$a_0 = \ln\int_0^\infty \exp(-a_1 \ln[1-k\frac{x}{\sigma}])dx$$

La relation entre les multiplicateurs de Lagrange et les contraintes

En dérivant la précédente équation par rapport à a_1 nous obtenons :

$$\begin{aligned}\frac{\partial a_0}{\partial a_1} &= -\frac{\int_0^\infty \exp(-a_1 \ln[1-k\frac{x}{\sigma}]) \ln[1-k\frac{x}{\sigma}]dx}{\int_0^\infty \exp[a_0 \ln 1 - k\frac{k}{\sigma}]dx} \\ &= -\int_0^\infty \exp a_0 - a_1 \ln[1-k\frac{x}{\sigma}] \ln[1-k\frac{x}{\sigma}]dx \\ &= -E[\ln[1-k\frac{x}{\sigma}]]\end{aligned}$$

De la même façon en prenant le logarithme et en dérivant l'equation $\exp(a_0) = \frac{\sigma}{k}\frac{1}{1-a_1}$ par rapport à a_1 on obtient :

$$\frac{\partial a_0}{\partial a_1} = \frac{1}{1-a_1}$$

En égalant les deux précédentes équations on aura :

$$\frac{1}{1-a_1} = -E[\ln[1-k\frac{x}{\sigma}]]$$

La distribution GD a deux paramètres, donc deux équations sont nécessaires. La deuxième équation est donnée par :

$$\frac{\partial^2 a_0}{\partial a_1^2} = Var[\ln(1-\frac{kx}{\sigma}]$$

En dérivant $\frac{1}{1-a_1}$ par rapport à a_1, on obtient :

$$\frac{\partial^2 a_0}{\partial a_1^2} = \frac{1}{(1-a_1)^2}$$

Des deux équations précédentes on obtient :

$$\frac{1}{(1-a_1)^2} = Var[\ln(1-\frac{kx}{\sigma}]$$

La relation entre les paramètres et les multiplicateurs de Lagrange

En insérant $a_0 = \ln[\frac{\sigma}{k}\frac{1}{1-a_1}]$ dans :

$$f(x) = \exp[a_0 - a_1 \ln(1-k\frac{x}{\sigma})]$$

on obtient :

$$f(x) = \frac{k(1-a_1)}{\sigma}(1-k\frac{x}{\sigma})^{a_1}$$

En égalant la fonction de densité de la Pareto généralisée avec la fonction obtenue on en déduit :

$$1 - a_1 = \frac{1}{k}$$

La relation entre les paramètres et les contraintes

Les paramètres k et σ de la distribution de Pareto généralisée sont liés aux multiplicateurs de Lagrange par l'équation :

$$a_0 = \ln[\frac{\sigma}{k}\frac{1}{1-a_1}]$$

et les multiplicateurs de Lagrange sont à leur tour liés au contraintes par les deux équations :

$$\frac{1}{1-a_1} = -E[\ln(1-\frac{kx}{\sigma})]$$
$$\frac{1}{(1-a_1)^2} = Var[\ln(1-\frac{kx}{\sigma})]$$

En éliminant les multiplticaeurs de Lagrange du système d'équations on aboutit à une relation entre les paramètres et les contraintes. Par conséquent, les équations d'estimation des paramètres sont données par :

$$k = -E[\ln(1-\frac{kx}{\sigma})]$$
$$k^2 = Var[\ln(1-\frac{kx}{\sigma})]$$

La distribution de l'entropie de la Pareto généralisée

L'entropie de la distribution de la Pareto généralisée est donnée par :

$$H(f) = \ln\sigma - \frac{(1-k}{k}E[1-\frac{k}{\sigma}x]$$

2.7.2 Avantages et Inconvénients des méthodes d'estimation

Afin d'aider le praticien dans les questions relatives à la méthode d'estimation optimale à utiliser dans le cadre des applications, (référence) ont donné dans la mesure du possible,

les avantages (ou les caractéristiques spéciales) et les inconvénients de plusieurs méthode, nous présetrons les avantages et les inconvéniant de quatre méthodes, à savoir la méthode MOM, la méthode MLE, la méthode PWM et la méthode L-moment, Certaines de ces méthodes n'ont jamais été comparées avec une étude de simulation et, par conséquent, nous ne pouvons pas faire de comparaisons entre ces méthodes .

méthode des moments Avantages – Très facile à utiliser .La méthode des moments est particulièrement utile quand une réponse rapide .

- Les estimateurs dépendent uniquement de : la moyenne et la variance de moyenne empirique.
- Les estimateurs sont asymptotiquement normaux ($k > -0,25$).

Inconvénients – L'existence des estimateurs est limitée à $k > -0,5$. Par conséquent, ce n'est pas une méthode recommandée quand il est y a une possibilité que la distribution soit une distribution à queue lourde.

- Les erreurs d'échantillonnage peuvent augmentées si l'échantillon contient des valeurs etrêmes.
- les estimateurs peuvent être inexistant (k¿0).

Méthode des moments de probabilité pondérées Avantages – Les estimateurs sont faciles à calculer .

- le l'existence des estimateurs est limitée à $k > -1$ .Cette restriction sur les valeurs de k permet des distributions plus lourdes.
- Pour un échantillon de grande taille, les estimateurs suivent asymptotiquement une distribution normale ($k > -0,5$).
- Les études de simulation montrent qu'ils se comportent particulièrement bien lorsque la taille de l'échantillon n'est pas grande et $-0,5 \leq k \leq 0$.

Inconvénients Les estimateurs peuvent être non existant ($k > 0$).

Méthode des moments linéaires Avantages – Les L-moments sont simplement des combinaisons linéaires de moments de probabilités pondérées, ils ont donc les mêmes propriétés des moments de probabilités pondérées.

- Comme les moments de probabilités pondérées, ils sont relativement simples à calculer et sont très populaires dans la littérature hydrologique.

Méthode du maimum de vraisemablance Avantages – C'est la méthode d'estimation des paramètres la plus efficace.

– Pour les grands échantillons, les estimateurs suivent asymptotiquement une distribution normale ($k < 0,5$).

Inconvénients – L'estimation des paramètres nécessite un algorithme numérique, tel que celui de Newton-Raphson.

– Les algorithmes utilisés pour calculer les estimaeurs du maximum de vraisemablance se heurtent souvent à des problèmes de convergence, même pour des échantillons de très grande taille.

2.8 Comparaison des méthodes d'estimation

Cet exemple d'application a été traité par V.P. Singh et H.Guo [[22]]. L'objectif est de comparer quatre méthodes à savoir la méthode basée sur le maximum de vraisemblance (MLE), la méthode des moments (MOM), la méthode des moments pondérés (PWM) et la méthode basée sur le principe du maximum d'entropie (POME).
Pour évaluer la performance de la méthode POME par rapport aux méthodes MOM, PWM et MLE. les auteurs ont procèdés à une comparaison par la méthode de Monte-Carlo (par simulation), en utilisant la racine de l'erreur quadratique moyenne standardisée (RMSE) et le biais standardisée .
Trois échantillons de tailles $n = 20$, $n = 50$ et $n = 100$ ont été générés avec une répétition de 1000.

2.8.1 Méthode de Monte-Carlo

La simulation de Monte-Carlo est une méthode d'estimation d'une quantité numérique qui utilise des nombres aléatoires. Stanislaw Ulam et John von Neumann l'appelèrent ainsi, en référence aux jeux de hasard dans les casinos, au cours du projet Manhattan qui produisit la première bombe atomique pendant la Seconde Guerre mondiale. La simulation d'un prêt bancaire par exemple n'est pas une simulation de Monte-Carlo car il s'agit de calculs exacts à partir du nombre de mensualités et du taux d'intérêt ; aucun phénomène

aléatoire n'intervient dans les calculs.
La simulation de Monte-Carlo présente le double avantage d'être simple d'utilisation et de pouvoir être appliquée à un très large éventail de problèmes. Elle est utilisée en finance, pour déterminer quand lever une option sur un bien financier ; en assurance, pour évaluer le montant d'une prime ; en biologie, pour étudier les dynamiques intra et intercellulaires ; en physique nucléaire, pour connaître la probabilité qu'une particule traverse un écran ; en télécommunications, pour déterminer la qualité de service ; ou de façon générale pour déterminer la fiabilité d'un système, sa disponibilité ou son temps moyen d'atteinte de la défaillance.
Pour cela, il faut cependant poser le problème, le modéliser de sorte que la quantité à rechercher s'exprime comme l'espérance d'une variable aléatoire X, notée $E(X)$. Une variable aléatoire est le résultat d'une expérience soumise au hasard ; son espérance est schématiquement ce que l'on s'attend à trouver en moyenne si l'on répète l'expérience un grand nombre de fois. Trouver cette modélisation est très probablement la tâche la plus complexe. Dans certains cas, la quantité à calculer est naturellement une espérance, par exemple la moyenne du nombre de clients dans une file d'attente quand les temps d'arrivée et de traitement sont aléatoires. Mais la méthode peut également être utilisée pour estimer une quantité purement déterministe, par exemple une surface ou une intégrale, en construisant artificiellement une variable aléatoire pour se ramener au calcul d'une moyenne.
Pour calculer les estimateurs, la méthode de simulation de Monte-Carlo est utile. C'est une méthode algorithmique visant à calculer une valeur numérique approchée en utilisant des techniques probabilistes. Les méthodes de Monte-Carlo sont particulièrement utilisées pour le calcul des intégrales, ces méthodes ont été inventées par Nicolas Metropolis en 1947.

Soit $g(x), x \in \mathbb{R}^m$ une fonction quelconque mesurable et $f(x)$ une fonction de densité de support $\chi \subset \mathbb{R}^m$
Nous voulons calculer une intégrale $I = \int_{\mathcal{A}\in\chi} g(x)f(x)dx$

1- Nous écrivons tout d'abord l'intégrale I sous forme d'espérance

$$\begin{aligned} I &= E^f(g(x)) \\ &= \int_{\mathcal{A}\in\chi} g(x)dx \end{aligned} \tag{2.1}$$

2- Simuler des variables aléatoires $g(x)$

Nous génèrons un certain nombre n de variables aléatoires $x_i, i = 1, \ldots, n$ iid de densité f

Nous estimons I par l'espérance empirique :

$$I = I_n = \frac{1}{n} \sum_{i=1}^{n} g(x_i) \tag{2.2}$$

La convergence de cette méthode est justifiée par la loi forte des grands nombres et la vitesse de la convergence est précisée par le théorème central limite.

2.8.2 Indices de performance :

Les auteurs ont évalués la performance de la méthode POME à l'aide des indices de performance suivants :

Biais standardisé :

$$BIAIS = \frac{E(\widehat{\theta}) - \theta}{\theta}$$

Racine de l'erreur quadratique moyenne standardisée :

$$RMSE = \frac{\sqrt{[E(\widehat{\theta} - \theta)^2]}}{|\theta|}$$

où :$\widehat{\theta}$ est un estimateur de θ (paramètre ou quantile) et

$$[E(\widehat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \widehat{\theta_i}$$

où N est le nombre d'échantillons de Monte Carlo (N = 1000 dans cette étude).

Coefficient de variantion :

$$CV = \frac{\mu}{\sigma}$$

où μ est la moyenne et σ l'écart type.

Le nombre d'échantillons de 1 000 n'est sans doute pas assez grand pour produire les vraies valeurs de BIAIS et RMSE, mais suffira pour comparer la performance des méthodes d'estimation.

2.8.3 BIAIS dans les estimations des paramètres

Les auteurs ont obtenu les résultats suivants :
Le biais du paramètre estimé par les quatre méthodes varie selon la taille de l'échantillon et le Coefficient de variation (CV) .
En général, le biais diminue avec l'augmentation de la taille de l'échantillon, mais sa variation avec le Coefficient de variation n'était pas consistante (une variation du CV va impliquer un changement de biais remarquable).
Pour toutes les tailles d'échantillons, la méthode du moment (MOM) produit le biais le plus élevé, tandis que les autres à l'inverse ont un biais plus petit, en particulier, pour $n \geq 50$. Pour $Cv \geq 2.0$ et $n \leq 20$, PWM est préférable aux autres(produit le biais le plus petit).
En conclusion, on peut dire que pour $cv > 1,5$ POME est la méthode préférable.

2.8.4 RMSE dans les estimations des paramètres

Les auteurs ont obtenu les résultats suivant :
En général, la RMSE diminue avec l'augmentation de la taille de l'échantillon pour un Cv donné.Les auteurs ont remarqués que la méthode PWM a le RMSE le plus petit, à l'inverse la méthode MOM a le RMSE la plus élevée.
La différence entre la valeur du RMSE de la méthode PWM et celle de la méthode POME, MLE et MOM (à l'exception des petits échantillons $n \leq 20$) était petite.
Pour $n \geq 50$ et $Cv \geq 1.5$, les quatre méthodes sont relativement analogues (produisent la même RMSE).

2.8.5 BIAIS dans l'estimation des quantiles

Les auteurs ont obtenu les résultats suivant :
Le biais d'une méthode donnée, en général, variait en fonction de la taille de l'échantillon, de la valeur de Cv, et l'ordre du quantile. Pour $P < 0,90$ et $n < 20$, le biais pour les méthodes POME, PWM et MLE sont relativement analogues (produisent le même biais). Pour $Cv > 2.0$ et $n \geq 50$, la méthode MLE produit le biais le moins important, et les méthodes POME et PWM sont analogues (produisent le même biais).

Pour $0,99 < P < 0,999$ et $n \geq 50$, les méthodes POME et PWM sont analogues (produisent le même biais).
Quand $n < 20$, les méthodes POME et MLE n'ont pas donné de bons résultats, en particulier pour $Cv > 3.0$.

2.8.6 RMSE dans l'estimation des quantiles

Les auteurs ont obtenu les résultats suivant :
La RMSE d'une méthode donnée variait considérablement selon la taille de l'échantillon (n), l'ordre du quantile et le coefficient de variation (Cv).
En général, les méthodes POME et MLE produisent la RMSE la plus élevé, en particulier pour $Cv > 1,5$ et la méthode des moments (MOM) le moins élevé, pour toutes les tailles d'échantillon et pour toutes les valeurs du Cv.
Par conséquent, la méthode des moments est préférable aux autres (produit le biais le plus petit).

2.8.7 Conclusion

Les auteurs ont tirés de cette étude les conclusions suivantes :
(1) En termes de biais, la méthode POME a les mêmes préférances que les méthodes PWM et MLE.
(2) La méthode POME est analogue aux méthodes MLE et PWM en termes de RMSE.
(3) Pour $P \leq 0,9$ et $n \leq 20$, la méthode POME produit le biais le plus petit dans l'estimation du quantile. Pour $0,9 \leq P \leq 0,99$ et $n \geq 50$, les méthodes POME et PWM sont relativement analogues en terme de biais.
(4) Pour $P \leq 0,99$, les trois méthodes POME, PWM et MLE le ont même RMSE. mais ont un RMSE relativement plus grand par rapport à la méthode MOM.

Conclusion générale

Dans ce mémoire, Nous avons d'abord rappelé les principes fondamentaux de la théorie des valeurs extrêmes pour des suites de variables aléatoires i.i.d, à savoir, la loi limite du maximum et la loi des excés. Nous nous sommes intéressés dans le deuxième chapitre à différentes méthodes d'estimation (la méthode du maximum de vraisemblance, la méthode des moments, la méthode des moments de probabilités pondérées, la méthode des moments linéaires et la métthode du maximum d'entropie). À titre illustratif en fin du deuxième chapitre, un exemple d'application sur des données simulées traiter par V.P. Singh et H.Guo (1992) .

Il est en effet impossible de recommander une approche spécifique plutôt qu'une autre parce que le choix de l'approche la plus appropriée dépend fortement du problème étudier et suite à ce travail plusieurs pistes de recherche nous semblent intéressantes. Conceptuellement, l'approche Bayésienne qui est peut-être les méthodes la plus adéquates et qui ets largement abordée dans la littéraure ([31] ,[31]). La rareté des données, qui est habituelle dans le cadre de la théorie des valeurs extrêmes, peut visiblement bénéficier de toute information supplémentaire qui peut être incluse dans le processus d'estimation.

Bibliographie

[1] Alvarado, E., S. D. e. P. S. Modeling large forest fires as extreme events. *Northwest Science 72* (1998), 66–75.

[2] Arvesen, J. N. Jackknifing u-statistics. *Ann. Math. Statist 40* (1960), 2076–2100.

[3] Beirlant, J., G. Y. S. J. T. J. d. W. D. e. F. C. *Statistics of Extremes : Theory and Applications*. Wiley Series in Probability and Statistics. 2004.

[4] Beirlant, J. et Teugels, J. Modelling large claims in non-life insurance. *Insurance : Mathematics and Economics 11*, (1) (1992), 17–29.

[5] Bílková, D. L-moments and tl-moments as an alternative tool of statistical data analysis. *Journal of Applied Mathematics and Physics* (2014), 919–929.

[6] Brodin, E. et Rootzén, H. Univariate and bivariate gpd methods for predicting extreme wind storm losses. *Insurance : Mathematics and Economics 44*, (3) (2009), 345–356.

[7] Castillo, E. et Hadi, A. "fitting the generalized pareto distribution to data". *Journal of the American Statistical Association* (1997), 92,1609–1620.

[8] Coles, S. et Tawn, J. A bayesian analysis of extreme rainfall data. *Applied Statistics* (1996), pages 463–478.

[9] Daouia, A., G. L. G. S. e. L. A. Kernel estimators of extreme level curves. *Test, 20*, (2) (2011), 311–333.

[10] Davison, A. et Smith, R. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B 52*, (3) (1990), 393–442.

[11] de Haan, L. Sample extremes : an elementary introduction. *Statistica Neerlandica 30*, (4) (1976), 161–172.

[12] de Haan, L. *Fighting the arch–enemy with mathematics*, vol. 44. 1990.

[13] Ditlevsen, O. "distribution arbitrariness in structural reliability". In *Conference on Structual Safety and Reliability (ICOSSAR'93)* (Balkema, Rotterdam, 1994), S. M. Ed :Schueller, G.I. and J. Yao, Eds., pp. 1241–1247.

[14] Embrechts, P., K. C. e. M. T. Modelling extremal events for insurance and finance. *Springer Verlag* (1997).

[15] Embrechts, P. *Extremes and integrated risk management*. 2000.

[16] Embrechts, P., R. S. e. S. G. Living on the edge. *Risk 11*, (1) (1998), 96–100.

[17] Fisher, R. et Tippet, L. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society 24* (1928), 180–190.

[18] Gnedenko, B. *Sur la distribution limite du terme maximum d'une série aléatoire*, vol. 44. 1943.

[19] Guillou, A. et Willems, P. Application de la théorie des valeurs extrêmes en hydrologie. *Revue de statistique appliquée 54*, (2) (2006), 5–32.

[20] Gumbel, E. *Statistical theory of extreme values and some practical applications : a series of lectures*, vol. 33 of *Numéro Applied Mathematics Series*. 1954.

[21] Gumbel, E. *Statistics of extremes*. 1958.

[22] Guo, V. P. S. Parameter estimation for 2-parameter generalized pareto distribution by pome. *Stochastic Hydrology and Hydraulics* (1992).

[23] Hoeffding, W. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics Volume 19, Number 3* (1948), 293–325.

[24] Hosking, J. et Wallis, J. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics 29*, (3) (1987), 339–349.

[25] Hosking, J., W. J. e. W. E. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics 27(3)* (1985), 251–261.

[26] Jenkinson, A. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society 81(348)* (1955), 158–171.

[27] Kotz, S. et Nadarajah, S. *Extreme value distributions : theory and applications*. 2000.

[28] Levine, R. D. et Tribus, M. *The Maximum Entropy Formalism*. PhD thesis, MIT Press, Cambridge, Massachusetts, USA., 1979.

[29] Matthews, R. Far out forecasting. *The New Scientist 2051* (October 12 1996), 36–40.

[30] Methni, J. E. *Contributions à l'estimation de quantiles extrêmes. Applications à des données environnementales.* PhD thesis, Mathématiques générales [math.GM]. Université de Grenoble, 2013.

[31] Mokrani, F., F. H. N. A. Robust bayesian inference of generalized pareto distribution. *Afrika Statistika* (2016).

[32] P. de Zea Bermudez, S. K. Parameter estimation of the generalized pareto distribution — parti. *Journal of Statistical Planning and Inference 140* (2009), 1353–1373.

[33] Pickands, J. Statistical inference using extreme order statistics. *The Annals of Statistics 3(1)* (1975), 119–131.

[34] Rasmussen, P. F. Generalized probability weighted moments : Application to the generalized pareto distribution. *Water Resources Research 37*, (6) (2001), 1745–1751.

[35] Reiss, R.-D. et Thomas, M. *Statistical analysis of extreme values : with applications to insurance, finance, hydrology and other fields.* 2001.

[36] Reiss, R.-D. *Approximate distributions of order statistics : with applications to nonparametric statistics.* 1989.

[37] Rootzén, H. et Tajvidi, N. Extreme value statistics and wind storm losses : a case study. *Scandinavian Actuarial Journal (1)* (2001), 70–94.

[38] Salvadori, G. Linear combinations of order statistics to estimate the position and scale parameters of generalized pareto law. *Stoch. Environ. Resear. and Risk. assess* (2002), 1–17.

[39] Singh, V. P., R. A. K. A new method of parameter estimation for hydrologie frequency analysis. *Hydrol. Sci. Technol. 2(3)* (1986), 33–40.

[40] Smith, R. Maximum likelihood estimation in a class of nonregular cases. *Biometrika 72*, 1 (1985), 67–90.

[41] Tribus, M. *Rational Descriptions, Decisions and Designs.* 1969.

[42] Von Mises, R. *La distribution de la plus grande de n valeurs.* Providence, RI, II(In Selected Papers). 1954.