**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**

Mémoire présenté en vue de l'obtention du

**DIPLÔME DE MASTER en MATHÉMATIQES**

Option : **Statistique**

Par

Cherifa Guettaf Temam

Thème

# Sur l'estimation des quantiles extrêmes sous censure aléatoire

**Membres du Comité d'Examen**

| | | | |
|---|---|---|---|
| Pr. | **Fatah Benatia** | UMKB | Président |
| Dr. | **Louiza Soltane** | UMKB | Encadreur |
| Dr. | **Amel CHINE** | UMKB | Examinateur |

**Juin 2023**

*Aucun dédicace ne saurait exprimer mon respect, pour le propriétaire de la biographie parfumée qui était l'éducateur, le mentor et le lien et a été crédité du succès de mon cher père,*

*À ma lune lumineuse mon héroïne et à mon premier professeur, c'était sa prière et sa satisfaction que ma chère mère m'apportait,*

*À mon frère qui m'a soutenu Moustapha,*

*À mes frères et mes soeurs,*

*À toute ma famille,*

*À tous mes enseignants.*

*Enfin, je ne terminerais pas sans citer toutes mes amies pour les excellents moments passés ensemble,*

# REMERCIEMENTS


*Je tiens à remercier ALLAH le tout-puissant qui m'a aidée et donnée la santé, la patience et le courage durant ces longues années d'études.*

*Je tiens remercié sincèrement mon encadreur, Dr. Louiza Soltane, qui s'est toujours montré l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi que pour l'inspiration, l'aide et le temps qu'il ont bien voulu me dédier.*

*Je remercie également tous les enseignants et employés du département de Mathématiques présidé par Monsieur Imad Eddine Lakhdari pendant mes années d'études., avec une mention spéciale à Messieurs les membres du jury "le Pr. Fatah Benatia et le Dr. Amel Chine" qui nous ont faites l'honneur de participer à notre soutenance.*

*J'adresse mes plus remerciements à toutes mes amies qui m'ont aidé et encouragé au cours de la réalisation ce mémoire.*

*Enfin, Je remercie ma famille à qui je n'ai jamais su dire toute l'affection que j'ai pour eux, mon père, ma mère, mes frères et ma sœur qui ont été et seront toujours présents à mes côtés, merci pour votre soutient et vos encouragements.*

*Merci toutes et à tous.*

G.T. Cherifa

# TABLE DES MATIÈRES

# TABLE DES FIGURES

# LISTE DES TABLEAUX

# INTRODUCTION

La théorie des valeurs Extrêmes (TVE), apparue entre 1920 et 1940, grâce à Fisher et Tippet [9], Gendenko et Gumbel [13], joue un rôle de plus en plus important dans le traitement de la modélisation des évènement rares. Elle voit le jour en premier fois en 1997 avec la sortie du livre Riess et Thomas [25]. Cette théorie, fondée sur des résultats de la théorie des probabilités, offre un cadre mathématique rigoureux pour l'estimation des probabilités d'événements rares. Il s'agit fondamentalement de modéliser un phénomène aléatoire, en s'intéressant principalement aux quantiles extrêmes et à la queue de distribution souvent modélisée par un indice appelé indice des valeurs extrêmes.

Une des caractéristiques des données de survie est l'existence d'observation incomplète. La censure font partie du processus générant de ce type de données. Notamment au cours de la dernière décennie, de nombreux auteurs ont commencé à prêter attention à l'estimation de l'indice de queue sous données incomplètes : sous censure aléatoire à droite on trouve Beirlant et al.[2] en 2007, ils ont introduit une méthode pour les estimateurs de Hill et de moment,... de plus, ils ont proposé les estimateurs des quantiles extrêmes et ont discuté leurs propriétés asymptotiques lorsque les données sont censurées pour un seuil déterministe et l'année suivante Einmahl et al. [8] ils ont adapté différents estimateurs de l'IVE au cas où les données sont censurées par un seuil aléatoire et ils ont proposé une méthode unifiée pour établir leur normalité asymptotique. Ndao et Dupuy [23] ont adressée l'estimation non paramétrique de l'IVE conditionnel et son quantile pour les distributions à queue lourde est abordée lorsque certaines informations sur les covariables sont disponibles et que les données sont censurées aléatoirement à droite.

L'objectif principal de ce mémoire est l'estimation des quantiles extrêmes sous données complètes ainsi sous des données censurées. Ce mémoire est composé de deux chapitres.

Dans le premier chapitre, on présente tout d'abord les principaux défnitions et concepts de base de l'TVE (variabl aléatoire, fonction de quantile, queue...). Ensuite, on présente des généralités sur la censure (analyse de survie, type de censure).

Dans le deuxième chapitre : stastiques des extrêmes sous censure aléatoire se compose de deux sections. Donc, on commence de présenter à l'estimation de l'IVE et quantiles extrêmes sans censure. Ensuite, on introduit l'estimation de l'IVE et quantiles extrêmes avec censure.

# CHAPITRE 1

# THÉORIE DES VALEURES EXTRÊMES ET CENSURE

Dans ce chapitre, on énonce des fondamentales sur la théorie des valeurs extrêmes (TVE), qui est utilisée pour la modélisation des évènements extrêmes. Pour des descriptions détaillaient sur la TVE, consulter les excellents bouquins comme Embrechtes et al. [7] et Cláudia et Maria [5] On discute également dans ce chapitre des notions de base sur la censure.

## 1.1 Généralités sur la TVE

### 1.1.1 Définitions de base

**Définition 1.1.1 (Variable aléatoire)**

*Une variable aléatoire (v.a) $X$ est une fonction de l'ensemble fondamental $\Omega$ à valeurs dans*

$\mathbb{R}$

$$X : \Omega \to \mathbb{R}. \tag{1.1}$$

**Définition 1.1.2 (Loi de probabilité)**

*On appelle loi de probabilité de $X$ notée $P_X$ l'application qui a toute partie $A$ de $\mathbb{R}$ associe :*

$$P_X(A) = P(X^{-1}(A)) = P(X \in A). \tag{1.2}$$

**Définition 1.1.3 (Fonctions de répartition et empirique)**

– *La fonction de répartition (fdr) de la v.a $X$ est définie par :*

$$F_X(x) = F(x) := P(X \leq x), \ \forall x \in \mathbb{R}. \tag{1.3}$$

*$F$ est appelée aussi fonction de distribution (fd) ou fonction de distribution cumulée (fdc).*

– *La fonction de répartition empirique, notée par $F_n$ et définie par :*

$$F_n(x) := \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}\left\{X_j \leq x\right\}, \ x \in \mathbb{R}. \tag{1.4}$$

*Où $\mathbb{I}_{\{X_j \leq x\}}$ est la fonction d'indicatrice de l'ensemble $A$. On appelle la statistique (1.4) l'estimateur de (1.3) .*

**Propriété 1.1.1 (Caractéristiques de $F(x)$)**

*$F$ est fonction de répartition si :*

- $0 \leq F(x) \leq 1.$
- *$F(x)$ fonction croissante et continue à droite.*
- $\lim_{x \to -\infty} F(x) = 0$ *et* $\lim_{x \to +\infty} F(x) = 1.$

**Définition 1.1.4 (Fonctions de quantile et quantile empirique)**

*La fonction du quantile d'une distribution de probabilités est l'inverse généralisé de la fonction de distribution $F$*

$$\tag{1.5}$$

*La fonction du quantile empirique d'un échantillon $X_1, ..., X_n$ est définie par :*

$$Q_n(p) = F_n^{-1}(p) := \inf \{x \in \mathbb{R} : F_n(x) \geq p\}, \quad 0 < p < 1. \tag{1.6}$$

**Définition 1.1.5 (Fonctions de queue et de queue empirique )**

– *La fonction de queue ou de survie, qu'on note par $\overline{F}(x)$ est définie sur $[0,1]$ par :*

$$\overline{F}(x) = 1 - F(x) := 1 - P(X \leq x) = P(X > x).$$

– *La fonction empirique de queue est :*

$$\overline{F}_n(x) := \frac{1}{n} \sum_{j=1}^{n} \mathbb{I} \{X_j > x\}, \quad \forall x \in \mathbb{R}. \tag{1.7}$$

**Propriété 1.1.2 (Caractéristiques de $S(x)$)**

*$\overline{F}(x)$ la fonction de survie si :*

- *$\overline{F}(x)$ fonction continue à droite, i.e $\overline{F}(x^+) = \overline{F}(x)$, pour $x \geq 0$.*
- *$\overline{F}(x)$ fonction décroissante telle que $\overline{F}(0) = 1$ et $\lim_{x \to \infty} \overline{F}(x) = 0$.*

### 1.1.2 Convergences des suites de variables aléatoire

Soient $X_1, ..., X_n$ une suites de v.a's étant une suite de fonctions de $\Omega$ dans $\mathbb{R}$, il existe diverses façons de définir la convergence de $X_1, ..., X_n$ dont certaines jouent un grand rôle en calcul des probabilités (voir Cláudia et Maria [5]). Avant de passer à la théorie asymptotique (ou limite) des extrêmes, on doit introduire avec les concepts suivants de convergence :

**Définition 1.1.6 (Convergence en distribution)**

*On dit qu'une suite de v.a's $X_1, ..., X_n$ converge en distribution vers une v.a $X$, on la note $X_n \xrightarrow{d} X$, si pour tout $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} F_n(x) = F(x).$$

*Ceci est également connu sous le nom de convergence faible.*

**Définition 1.1.7 (Convergence en probabilité)**

*La suite $X_1, ..., X_n$ converge en probabilité vers $X$, on la note $X_n \xrightarrow{P} X$, si pour tout $\epsilon > 0$,*

$$\lim_{n\to\infty} (\mid X_n - X \mid > \epsilon) = 0.$$

**Définition 1.1.8 (Convergence presque sûrement)**

*La suite $X_1, ..., X_n$ converge presque sûrement vers $X$, on la note $X_n \xrightarrow{p.s.} X$, si :*

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

*La convergence presque sûre est également appelée convergence forte.*

La convergence de $F_n$ vers $F$ est presque sûrement uniforme c'est-à-dire que :

$$\sup_x \{\mid F_n(x) - F(x) \mid\} \xrightarrow{p.s} 0, \qquad \text{quand } n \to \infty. \tag{1.8}$$

La convergence (1.8) est connue sous le nom de théorème de Glivenko-cantell. Il est l'un des résultats fondamentaux en statistiques non paramétriques. La preuve du résultat (1.8) peut-être trouvés dans tout manuel standard de la théorie des probabilités.

### 1.1.3 Statistique d'ordre

Soit $X_1, ..., X_n$, $n$ v.a's réelles iid de fdr $F$, on définit les statistiques d'ordre dans les suivantes

**Définition 1.1.9 (Statistique d'ordre)**

*la statistique d'ordre de l'échantillon $X_1, ..., X_n$ est le réarrangement croissant de $X_1, ..., X_n$, elle notait par $X_{1:n}, ..., X_{n:n}$, on écrit*

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

*En particulier, on note que :*

$$X_{1:n} = -\max(-X_1, ..., -X_n).$$

*$X_{1:n}$ elle représente la plus petite statistique d'ordre (où statistique du minimum) et $X_{n:n}$ est la plus grande statistique d'ordre (où statistique du maximum). $X_{j:n}$ elle représente le $j^{\text{ème}}$ statistique d'ordre (statistique d'ordre $j$) dans un échantillon de taille $n$.*

**Proposition 1.1.1 (Distributions du maximum et du minimum)**

*La distribution du maximum $X_{n:n}$ et du minimum $X_{1:n}$ sont donnés par les deux formules suivantes :*

$$F_{X_{1:n}}(x) = 1 - [1 - F(x)]^n \text{ et } F_{X_{n:n}}(x) = [F(x)]^n, \ \forall x \in \mathbb{R}. \tag{1.9}$$

On démontre que les deux expressions de (1.9)

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

et

$$\begin{aligned} F_{X_{1:n}}(x) &= P(X_{1:n} \leq x) = 1 - P(\min(X_j) \geq x) \\ &= 1 - P(\bigcap_{j=1}^{n} \{X_j \geq x\}) = 1 - \prod_{j=1}^{n} (1 - P\{X_j \geq x\}) \\ &= 1 - \left[\overline{F}(x)\right]^n. \end{aligned}$$

**Remarque 1.1.1** *La distance entre les statistiques d'ordre extrêmes, est appelée déviation extrême ou l'étendu de l'échantillon, notée $D_n$ donnée par :*

$$D_n = X_{n:n} - X_{1:n}.$$

### 1.1.4 Comportement asymptotique des extrêmes

Les définitions et les remarques énoncées dans cette partie sont issues de Cláudia et al. [5]. On remarque d'abord

$$\lim_{n\to\infty} F_{X_{n:n}}(x) = \lim_{n\to\infty} [F(x)]^n = \begin{cases} 1 & x \geq x^F \\ 0 & x < x^F. \end{cases} \tag{1.10}$$

On constate que la distribution asymptotique du maximum donne une loi dégénérée, une masse de Dirac en $x^F$, puisque pour certaines valeurs de $x$, la probabilité peut-être égale à 1 dans le cas où $x^F$ est fini et donc $X_{n:n}$ tend vers $x^F$ presque sûrement, avec $x^F \leq \infty$. On donne dans la suivante la définition du point terminal :

**Définition 1.1.10 (Point terminal)**

Le point terminal supérieur (ou droit) de $F$ est

$$x^F := \sup \{x \in \mathbb{R},\ F(x) \leq 1\}.$$

De façon équivalente, le point terminal inférieur (ou le point terminal gauche) d'une distribution $F$ est désigné par $x_F$ est

$$x_{_F} := \inf \left\{ x \in \mathbb{R},\ F(x) > 0 \right\}.$$

Ce fait (1.10) ne fournit pas assez d'informations, d'où l'idée d'utiliser une transformation afin d'obtenir des résultats plus exploitables pour les lois limites des maxima.On s'intéresse par conséquent à une loi non dégénérée pour le maximum, la TVE permet de donner une réponse à cette problématique. Les premiers résultats sur la caractérisation du comportement asymptotique des maxima $X_{n:n}$ convenablement normalisés ont été obtenus par Fisher et Tippet [9], Gnedenko [13]. Ce résultat est analogue au Théorème Central Limite (TCL), mais pour les phénomènes extrêmes. On a ici le théorème qui donne la loi du maximum.

**Théoreme 1.1.1 (Fisher-Tippet-Gnedenko-Mises-Jenkinson)**
*Sous certaines conditions de régularités sur la fonction $F$, s'il existe deux constantes de normalisation $(a_n)_{n\ \geq 1} > 0$ et $(b_n)_{n\ \geq 1}$ réelle telle que :*

$$\lim_{n\to\infty} P\left(\frac{X_{n:n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n\left(a_n x + b_n\right) = \mathcal{H}(x), \quad \forall x \in \mathbb{R}, \tag{1.11}$$

*la loi limite converge en distribution vers une loi non dégénérée $\mathcal{H}(x)$. Cette loi de la limite que l'un des trois types suivantes :*

$$\begin{aligned}
&\textit{Gumbel :}\quad \Lambda(x) &&= \exp(-e^{-x}), && x \in \mathbb{R}\ \textit{et}\ \alpha = 0.\\
&\textit{Fréchet :}\quad \Phi_\alpha(x) &&= \begin{cases} 0, & x \leq 0 \\ \exp\left(-x^{-\alpha}\right), & x > 0 \end{cases} && \alpha > 0.\\
&\textit{Weibull :}\quad \Psi_\alpha(x) &&= \begin{cases} \exp\left(-\left(-x^{-\alpha}\right)\right), & x \leq 0 \\ 0, & x > 0 \end{cases} && \alpha > 0.
\end{aligned} \tag{1.12}$$

*Ces trois distributions limites sont appelées les distributions de valeurs extrêmes standard.*

Une démonstration détaillée de ce théorème est donnée dans Embrechts et al. [7], page 152. Il est possible de rassembler les trois familles de lois en une seule famille paramétrique ($\mathcal{H}_\gamma(x)$, $\gamma \in \mathbb{R}$) de Von Mises-Jenkinson [[30], [18]], elle est dite famille des lois des valeurs extrêmes généralisées (GEV, Generalized Extreme Value distribution). Notamment, la distribution des valeurs extrêmes généralisées donnée par :

$$\mathcal{H}_\gamma(x) = \begin{cases} \exp\left\{-(1+\gamma)^{-1/\gamma}\right\} & \text{si } \gamma \neq 0, \quad 1+\gamma x > 0 \\ \exp\left\{-e^{(-x)}\right\} & \text{si } \gamma = 0. \end{cases} \tag{1.13}$$

Le paramètre $\alpha := 1/\gamma$, il représente l'indice des valeurs extrêmes (IVE) si $\gamma > 0$ ou indice de queue ou paramètre de forme.

### Distributions à queue lourde

Les distributions à queues lourdes sont liées à la TVE et permettent de modéliser beaucoup de phénomènes que l'on trouve dans différentes disciplines telles que la finance, l'hydrologie, la climatologie épidémiologie, ...etc. Ce type des distributions est défini ainsi :

**Définition 1.1.11 (Distribution à queue lourde)**
*Soit $X$ une v.a de fdr $F$, donc cette dernière elle est dite distribution à queue lourde, s'il existe un constant positif $\gamma$ qui représente l'indice de queue et prend la formule suivante :*

$$\overline{F}(x) \sim x^{-1/\gamma} l(x), \text{ pour } x \to \infty, \tag{1.14}$$

*où $l(x)$ la fonction à variation lente au voisinage de l'infini.*

Le type de distribution (1.14) satisfaite pour tout $x > 0$, les conditions suivantes :

**Définition 1.1.12 ( Conditions du premier et du second ordre)**

*(i)* $1 - F(x)$ *est une fonction à variation régulière d'indice* $-1/\gamma < 0$, *on a*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \qquad x > 0. \tag{1.15}$$

*(ii)* $\exists\ \beta \leq 0$ *et une fonction* $A \to 0$ *et ne change pas le signe au voisinage de l'infini, tel que*

$$\lim_{t\to\infty} \frac{1 - F(tx)/1 - F(t) - x^{-1/\gamma}}{A(t)} = x^{-1/\gamma} \frac{x^{\beta} - 1}{\gamma\beta}. \tag{1.16}$$

**Remarque 1.1.2**

*La condition du premier ordre en général n'est pas suffisante pour étudier les propriétés des estimateurs des paramètres de queue, en particulier la normalité asymptotique. Dans ce cas, une condition du second ordre des fonctions à variations régulières est nécessaire en spécifiant le taux de convergence dans* (1.15)*. Cette condition vient de de Haan et Ferreira [15] page* 48*.*

**Domaines d'attraction**

**Définition 1.1.13 (Domaines d'attraction)**

*Si* $F$ *vérifie le Théorème 1.1.1 alors on dit que* $F$ *appartient au Domaine D'attraction* ($\mathcal{DA}$) *du maximum de* $\mathcal{H}_\gamma$ *et on note* $F \in \mathcal{DA}(\mathcal{H}_\gamma)$, *s'il existe deux suites normalisantes* $a_n > 0$ *et* $b_n$ *fournis aussi dans le Théorème 1.1.1 de telle manière que*

$$\lim_{n\longrightarrow\infty} F^n \left(a_n x + b_n\right) = \mathcal{H}_\gamma(x). \tag{1.17}$$

Selon le signe de $\gamma$, on distingue trois cas de domaines d'attraction :

- Si $\gamma > 0$, on dit que $F \in \mathcal{DA}(\Phi_{1/\gamma})$, et $F$ a un point terminal à droite infinie ($x_F = +\infty$). Ce domaine d'attraction est celui des distributions à queues lourdes, c'est-à-dire qui ont une fonction de survie à décroissance polynomiale.

- Si $\gamma = 0$, on dit que $F \in \mathcal{DA}(\Lambda)$, le point terminal $x_F$ peut alors être fini ou non. Ce domaine d'attraction est celui des distributions à queues légères, c'est-à-dire qui ont une fonction de survie à décroissance exponentielle.
- Si $\gamma < 0$, on dit que $F \in \mathcal{DA}(\Psi_{1/\gamma})$, et $F$ a un point terminal à droite finie ($x_F < +\infty$). Ce domaine d'attraction est celui des fonctions de survie dont le support est borné supérieurement.

Le tableau suivant résume le classement de quelques lois usuelles selon leur appartenance à l'un des domaines d'attraction. (voir [7])

| Domaines d'attraction | Gumbel ($\gamma = 0$) | Fréchet ($\gamma > 0$) | Weibull ($\gamma < 0$) |
|---|---|---|---|
| Lois | Normal<br>Exponentielle<br>Log-normale<br>Gamma<br>Weibull | Pareto<br>Burr<br>Student<br>Log-gamma<br>Log-logistique | Uniforme<br>Beta<br>ReverseBurr |

TAB. 1.1 – Quelques loi usuelles classées en fonction de leur domaine d'attraction

## 1.2 Géneralité sur la censure

Une des caractéristiques des données de survie est l'existence d'observations incomplètes, en effet, les données sont souvent recueillies partiellement, notamment, à cause des processus de censure et troncature, les données censurées proviennent du fait qu'on n'a pas accès à toute l'information, au lieu d'observer des réalisations iid de durées $X$, on observe la réalisation diverses perturbations indépendantes ou non de l'événement étudie. Il intervient donc fréquemment lors de mesures qui portent sur les variables modélisant le temps écoulé entre deux événements : durée de vie d'un individu, durée entre le début d'une maladie et la guérison, durée d'un épisode de chômage, ... etc. (Source : thèse de doctorat de Soltane [28], mémoire de master de M'ziou [22] et Semmari [27]).

### 1.2.1 Notion de base en analyse de survie

On présente quelques définitions qui sont couramment utilisées dans l'analyse de survie.

**Définition 1.2.1 (de l'analyse de survie)**
*L'analyse de survie est le terme utilisé pour décrire l'analyse des données qu'ils sont sous forme de temps (times) allant de l'origine du temps (time origine) à la survenance d'un événement ou d'un point final spécifique (end point).*

Dans la recherche médicale, l'origine du temps est souvent la date d'enregistrement de l'élément (l'individu) dans une étude, comme les essais médicaux pour comparer deux types de médicaments ou plus (tester l'efficacité d'un médicament) si le point final est la mort du patient, les données résultantes sont les temps de survie, mais si le point final n'est pas la mort, alors les données résultantes sont appelées donner de temps sur événement.

**Remarque 1.2.1**
*On remarque qu'il existe des nombreux autres noms comme :*

- *Dans les statistiques médicales et épidémiologiques, il est connu sous le nom d'analyse de survie ou analyse de risque (Hazard Analysis).*
- *Dans les études d'ingénierie, elle est connue sous le nom d'analyse du temps d'échec (Failure Time Analysis).*
- *Dans les études de psychologie, il est connu comme analyse d'historique d'événement (Event History Analysis).*

**Exemple 1.2.1 (En biostatistique)** *:*

- *Survie d'un individu après l'apparition d'une tumeur.*
- *Âge de décès chez des patients atteints de diabète, ...*
- *Temps de rémission après opération chirurgicale.*

**Exemple 1.2.2 (D'autres domaines)** *:*

- *Fiabilité : durée entre deux pannes d'un matériel, d'un logiciel, ...*
- *Economie : durées des périodes de travail ou de chômage, temps avant faillite,...*
- *Assurances : durée de cotisation avant le premier remboursement.*

### Donnée de survie

**Définition 1.2.2** *Donnée de survie (Survival Data) est une expression utilisée pour décrire les données qui mesurent le temps jusqu'à un événement d'intérêt et que la variable résultante est le temps jusqu'à un événement d'intérêt, qui est connu par le temps de survie, qui est toujours une variable réelle des valeurs positives. Il y a trois conditions de base pour la durée de survie qui doivent être définies précisément qui sont la data d'origine, la date de point et la date des dernières nouvelles (l'événement), elles sont présentées dans la Figure 1.1, qu'on abordera en détail comme suit :*

**Date d'origine** L'origine de la durée étudiée. Elle peut être la date de naissance, le début d'une exposition à un facteur de risque, la date d'une opération chirurgicale, la date de début d'une maladie ou la date d'entrée dans l'étude. Chaque individu peut donc avoir une date d'origines différents, ce qui n'est pas important car c'est la durée qui nous intéresse.

**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.

**Date des dernières nouvelles** C'est la date à laquelle on a eu pour la dernière fois des nouvelles du sujet : cela peut être la date du décès ou la date de survenue de l'évènement étudié (guérison, première rechute, première apparition d'un évènement indésirable, ...), mais aussi la date de la dernière consultation si le sujet est perdu de vue ou n'a pas présenté l'évènement étudié (décès, guérison,...).

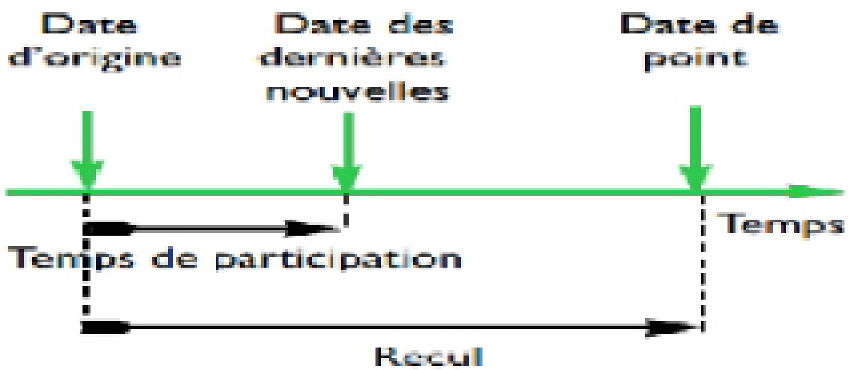


FIG. 1.1 – Schéma représentant les principales défnitions relatives à l'analyse de la durée de survie.

### 1.2.2 Données Censurées

Une caractéristique importante de l'analyse de la survie est la présence des données censurées. Cette caractéristique, source de difficultés, a nécessité le développement de techniques alternatives à l'inférence usuelle. La censure est le phénomène le plus couramment rencontré lors du recueil de données de survie.

**Définition 1.2.3** –

*Pour un individu $i$, considérons intervalle respectivement), Pour un individu donné $i$, on a*

**Définition 1.2.4** *Son temps de survie $S_i$.de fdr $F$.*

- *Son temps de censure $C_i$.de fdr $G$.*

- *La durée réellement observée $Z_i$ de fdr $H$.*

### 1.2.3 Type de censure

**Définition 1.2.5 (Variable de censure)**

La variable de censure $C$ est définie par la non-observation de l'événement étudié. Si au lieu d'observer $S$, on observe

**Définition 1.2.6 -** *$S > C$ est censure à dsroit.*

- *$S < C$ est censure à gauche.*

- *$C_1 < S < C_2$ est censure par intervalle.*

**1. Censure à droite** La variable d'intérêt est dite censurée à droite si 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. Un exemple typique est celui où l'événement considéré est le décès d'un patient malade et la durée d'observation est une durée totale d'hospitalisation. On trouve aussi ce genre de phénomène dans les études débilité quand la panne d'un appareil ou d'un composant électronique ne permet pas de continuer l'observation pour un autre appareil ou composant. On peut aussi trouver ces genres de phénomènes en hydrologie, en pluviométrie,... L'expérimentateur peut fixer une date de fin d'expérience et les observations pour les individus pour lesquels on n'a pas observé l'événement d'intérêt avant cette date seront censurées à droit.

**Exemple 1.2.3 (de censure droite)** *:*

L'étude porte sur la durée de survie $S$ de patients atteints d'une certaine maladie. Pour les patients perdus de vue au bout du temps $C$ alors qu'ils étaient encore vivants, $C$ censure $Z$ à droite puisque, pour eux, $Z$ est inconnue mais supérieur à $C$ , $Z > C$.

**2. Censure à gauche** Il y a censure à gauche lorsque l'individu 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. On sait uniquement que la variable d'intérêt est inférieure ou égale à une variable connue. Par exemple si on veut étudier en fiabilité un certain composant électronique qui est branché en parallèle avec un ou plusieurs autres composants : le système peut continuer à fonctionner, quoique de façon aberrante, jusqu'à ce que cette panne soit détectée (par exemple lors d'un contrôle ou en cas de l'arrêt du système). lors d'un contrôle ou en cas de l'arrêt du système). Ainsi donc, la durée observée pour ce composant est censurée à gauche. Dans la vie courante il y a plusieurs phénomènes qui présentent à la fois des données censurées à droite et à gauche.

**3. Censure double** La censure double (ou mixte) ce type de censure c'est un mélange entre les deux censures, la censure à droite et la censure à gauche, dans le même échantillon.

**4. Censure par intervalle** 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. On retrouve ce modèle en général dans des études de suivi médical où les patients sont contrôlés périodiquement, si un patient ne se présente pas à un ou plusieurs contrôles et se présente ensuite après que l'événement d'intérêt se soit produit. On a aussi ce genre de données qui sont censurées à droite ou, plus rarement, à gauche.Un avantage de ce type est qu'il permet de présenter les données censurées à droite ou à gauche par des intervalles du type $[a, +\infty[$ et $[0, a]$ respectivement.

Dans la littérature on retrouve les types suivants :

**Censure du type I : fixée** Soit $C$ une valeur fixée, au lieu d'observer les variables $S_1, ..., S_n$ qui nous intéressent, on observe Si que lorsque Si $C$, sinon on sait uniquement que $S_i > C$. On observe donc une variable $Z_i$ telle que :

$$Z_i = min(S_i, C), \qquad i = 1, ..., n$$

Ce mécanisme de censure est fréquemment rencontré dans les applications industrielles.

**Censure du type II : attente** L'expérimentateur fixe a priori le nombre d'événements à observer. La date définie d'expérience devient alors aléatoire, le nombre d'événements étant quant à lui, non aléatoire. Ce modèle est souvent utilisé dans les études fiabilité d'épidémiologie. Par exemple en épidémiologie on décide d'observer les durées de survie des $n$ patients jusqu'à ce que $k$ $(k = \overline{1,n})$, d'entre eux soient décédés et d'arrêter l'étude à ce moment-là. Soient $S_{i:n}$ et $Z_{i:n}$ les statistiques d'ordre des variables $S_i$ et $Z_i$. La date de censure est donc $S_{i:n}$ et on observe

$$Z_{i:n} = \begin{cases} S_{i:n} & \text{si } i \leq k. \\ S_{k:n} & \text{si } i \geq k. \end{cases}$$

**Censure du type III : aléatoire** Il existe $C$ une va's iid positive d'échantillon $C_1, ..., C_n$, on observe un couple de v.a's $(Z_i, \delta_i)$ avec

$$Z_i = min(S_i, C_i) \qquad \text{et} \qquad \delta_i = \mathbb{I}_{\{S_i \leq C_i\}},\ i = 1, ..., n \tag{1.18}$$

où $\delta_i$ d'indicateur de censure, qui détermine si $S_i$ a été censuré ou non :

$$\delta_i = \begin{cases} 1 & \text{où } Z_i = S_i \text{ (la durée d'intérêt est observée).} \\ 0 & \text{où } Z_i = C_i \text{ (elle est censurée).} \end{cases}$$

La censure aléatoire est la plus courante et la plus considérée en analyse de survie. Pour des détails complets sur les types de censure on réfère aux livres de ( Pierre [26], page 7). Dans ce travail, on s'intéresse uniquement au modèle de censure à droite du type aléatoire. Celui-ci correspond à un modèle fréquemment utilisé en pratique (voir aussi Soltane [28], page13).
Dans la Figure 1.2 on représente le suivi de trois patients. Le premier patient est décédé 4 semaines après le début du traitement : Il s'agit d'une observation non censurée. Le deuxième patient est vivant au terme des 14 semaines d'observation. L'information de ce patient n'est pas connue lorsque la constitution de la base de données est arrêtée (en $t = 10$). il est donc censuré. Quant au troisième patient, il a été perdu de vue à $t = 7$, donc il est censuré à $t = 7$.

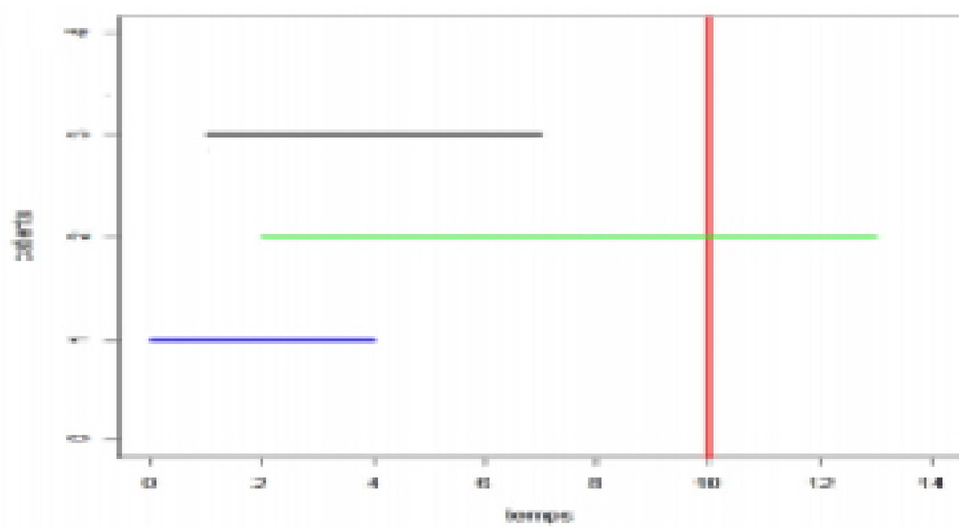


FIG. 1.2 – Schéma représentant les cas de censure aléatoire.

# CHAPITRE 2

# STATISTIQUES DES EXTRÊMES SANS ET SOUS CENSURE ALÉATOIRE

Pour la majorité des fdr $F$ la loi asymptotique du maximum $X_{n:n}$ est une loi des valeurs extrêmes qui étant indexée par le paramètre de queue $\gamma$. Il existe dans la littérature de la TVE de nombreux autres se sont intéressés à l'estimation de l'indice des valeurs extrêmes. Les méthodes les plus utilisées sont celles de Hill [16], de Pickands [24] et des moments. Ceci est de probablement au fait qu'elles sont les plus anciennes. Dans la réalité, quel que soit le phénomène étudié, on fait face toujours aux données incomplètes que ça soit tronqué ou censuré. Dans ce chapitre on s'intéresse à l'estimation du quantile extrême sans et avec censure.

## 2.1 Estimation de l'IVE et quantiles extrêmes sans censure

Dans cette section, on s'intéresse essentiellement à l'estimation de l'IVE et de quantiles extrêmes sur la base des données complètes.

### 2.1.1 Estimation de l'IVE

Les méthodes d'estimation des quantiles extrêmes sont basées sur des estimateurs des paramètres des lois GEV. Dans cette partie, nous faisons en particulier un rappel sur les estimateurs de l'indice des valeurs extrêmes les plus célèbres dans la littérature Hill, Pickands et moment

La difficulté consiste à choisir le nombre $k$ de statistiques d'ordre extrême à utiliser. Les résultats concernant les estimateurs de l'indice des valeurs extrêmes sont asymptotiques : ils sont obtenus lorsque

$$1 < k < n,\ k \to \infty \text{ et } k/n \to 0 \text{ quand } n \to \infty. \tag{2.1}$$

Comme en pratique, on ne dispose que d'un nombre d'observations $n$ fini, il s'agit de choisir $k$ de manière à ce que l'on dispose de suffisamment de matériel statistique tout en restant dans la queue de distribution $(k < n)$.

#### Estimateur de Hill

Cet estimateur qui a été présenté par Hill [16] en 1975 et l'estimateur le plus célèbre parmi tous les estimateurs de l'IVE et le plus utilisé. Il s'applique seulement dans le cas où l'indice de queue est positif $(\gamma > 0)$, est un distrubition appartenant au domain d'attraction de Fréchet.

**Définition 2.1.1 ( Estimateur de Hill)**

*Soit $X_1, ..., X_n$ une suite de v.a's iid de fdr $F$ appartenant au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $\gamma > 0$, et pour $k = k(n)$ une suite d'entier qui verifie les conditions (2.1). L'estimateur de Hill est définie par la statistique :*

$$\widehat{\gamma}_n^H = \frac{1}{k}\sum_{i=1}^{k} \log\left(X_{n-i+1:n}\right) - \log\left(X_{n-k:n}\right), \ 1 \leq k \leq n. \tag{2.2}$$

**Estimateur de Pickands**

L'estimateur de Pickands a été introduit en 1975 par Pickands dans [24], cette estimateure est plus génirale que l'estimateur de Hill et valable pour, $\gamma \in \mathbb{R}$. Il est défini de maniére suivant :

$$\widehat{\gamma}_n^p = \frac{1}{\log 2} \log\left(\frac{X_{n-k+1:n} - X_{n-2k_n+1:n}}{X_{n-2k_n+1:n} - X_{n-4k_n+1:n}}\right). \tag{2.3}$$

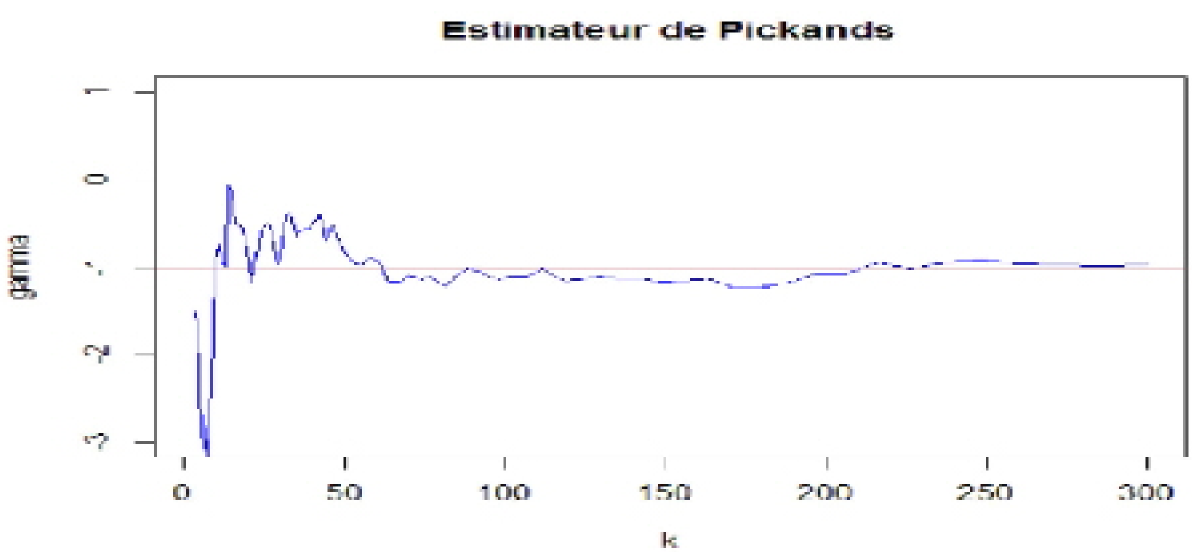


FIG. 2.1 – Représentation graphique del'estimateur de Pickands.

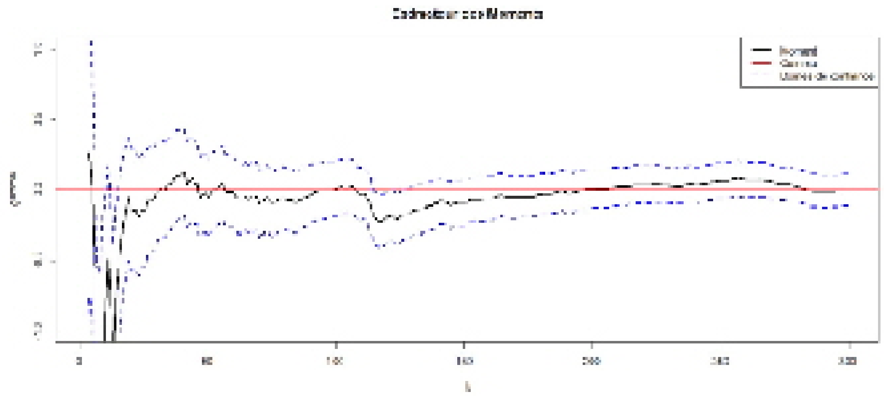

FIG. 2.2 – Estimateur de Moment pour l'EVI de la distribution de Gumbel ($\gamma = 0$) basé sur 300 échantillons de 3000 observations.

**Estimateur du moment**

Cet estimateur introduit par Dekkers et al (1989) [6] est une généralisation de l'estimateur de Hill, valable pour tout $\gamma \in \mathbb{R}$

**Définition 2.1.2 (Estimateur des moments)**

*Soit $F \in \mathcal{D}(H_\gamma)$, $k$ vérifie les conditions (1.10) et Pour $\gamma \in \mathbb{R}$, l'estimateur des moments est donnée par :*

$$\widehat{\gamma}_n^M = M_1 + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1}.$$

*Où*

$$M_n^j = \frac{1}{k}\sum_{i=1}^{k} \log\left(X_{n-i+1:n} - \log X_{n-k:n}\right)^r,\ j = 1,2 \tag{2.4}$$

*où $M_n^1$ est l'estimateur de Hill $\hat{\gamma}_n^H$.*

## 2.1.2 Estimation des quantiles extrêmes

On commençe par donner la définition du quantile d'ordre $(1-p)$ et du quantile extrême.

**Définition 2.1.3 (Quantile d'ordre $1-p$)**

*Soit $X_1, ...X_n$, $n$ v.a's iid de fonction de répartition commune $F$. On appelle quantile ou fractile d'ordre $1-p$ de la fonction de répartition $F$, le nombre $x_p$ défini par*

$$x_p := \overline{F}^{-1}(p) = inf\{x \in \mathbb{R} : \overline{F}(x) \leq p\} = Q\left(1-P\right), \quad avec\ p \in ]0,1[.$$

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

$$\overline{F}(x_p) = 1-p.$$

**Définition 2.1.4 (Quantile extrême)**

*Le quantile extrême d'ordre $1-p_n$ de la fonction de répartition $F$ est défini par :*

$$\widehat{x}_p := \overline{F}^{-1}(\widehat{p}),\ \ avec\ \widehat{p} \to 0,\ \ quand\ n \to \infty$$

*Dans tout ce qui suit, on suppose que $F \in \mathcal{D}(H_\gamma)$ pour certain $\gamma \in \mathbb{R}$. Afin d'estimer le quantile extrême, on introduit le résultat suivant.*

**Lemme 2.1.1**

*Si $\widehat{x}_p \to 0$ et $n\widehat{p} \to c$ (non nécessairement fini) quand $n \to \infty$, alors*

$$P(X_{n:n} < \widehat{x}_p) \to e^{-c}.$$

D'après le Lemma 2.1.1, on doit faire la distinction entre deux situations en fonction de $c$.

- Si $c = 1$ alors $P(X_{n:n} < \widehat{x}_p) = 0$. Dans un tel contexte, un estimateur classique de $\widehat{x}_p$ est le quantile empirique $X_{(n-[pn],n)}$.
- Si $c = 0$ alors $P(X_{n:n} < \widehat{x}_p) = 1$. La quantité $\widehat{x}_p$ est en dehors de l'échantillon. Par conséquent, on ne peut pas estimer le quantile de manière empirique. Pour résoudre ce

problème, on présente les méthodes suivantes.

Concernant l'étude du quantile extrêmes $x_p$, l'estimateur semi-paramétrique le plus fréquemment utilisé, associé à l'estimateur de Hill (2.2) a été proposé par Weissman (1978) [31].

**Définition 2.1.5 (Estimation des quantiles extrêmes)**

*Weissman a proposé un estimateur de la quantile extrême $Q(1-P)$ quand $\gamma > 0$, notée par $\widehat{\gamma}_n^H$*

$$\widehat{x}_p^H = \left(\frac{k}{np}\right)^{\widehat{\gamma}_n^H} X_{n-k:n.}$$

L'estimateur du quantile extrême, associé à l'estimateur de Pickands (2.3), est donnée par :

**Définition 2.1.6 (Estimation des quantiles extrêmes)**

*L'estimateur du quantile extrême, associé à l'estimateur de Pickands, est donnée par :*

$$\widehat{x}_p^p = X_{n-k+1:n} + \frac{\left(\frac{np/k}{(n+1)p}\right)^{-\hat{\gamma}^p} - 1}{1 - 2^{-\hat{\gamma}^p}} \left(X_{n-k+1:n} - X_{n-2k_n+1:n}\right).$$

Un estimateur pour des quantiles extrêmes sur la base de l'estimateur des moments est donné par :

**Définition 2.1.7 (Estimation des quantiles extrêmes)**

*Un estimateur pour les quantiles extrêmes $Q(1-P)$ sur la base de l'estimateur de moment*

$$\widehat{x}_p^M = X_{n-k:n} + \frac{\left(\frac{np}{k}\right)^{-\widehat{\gamma}_n^M} - 1}{\widehat{\gamma}_n^M} \times \frac{M_n^1 X_{n-k:n}}{\rho\left(\widehat{\gamma}_n^M\right)}.$$

*Où $M_n^1$ est définie par (2.4) :*

$$\rho(\gamma) = \begin{cases} 1 & \gamma \geq 0. \\ \frac{1}{1-\gamma} & \gamma < 0. \end{cases}$$

## 2.2 Estimation de l'IVE et quantiles extrêmes sous censure

### 2.2.1 Estimation de l'IVE

On s'intéresse dans cette partie à l'estimation de l'IVE sur la base des donnéés censurées aléatoirement à droite. Ce problème est trés récent dans la littérature, les premiers qui ont mentionné ce sujet sont Beirlant et al [2], Enimahl et al [8] et Ndao et Dupuy [23].

Soient $\overline{F}$ et $\overline{G}$ sont à queues lourdes avec les indices $-1/\gamma_1$ et $-1/\gamma_2$ où $\gamma_1 > 0$, $\gamma_2 > 0$ de l'échantillon $S_1, ..., S_n$ et $C_1, ...C_n$. Soit $\{(Z_i, \delta_i)\ ,\ 1 \leq i \leq n\}$ l'échantillon réellement observé défini par (1.18). Il est clair que les $Z_i's$ sont des variables indépendantes de loi $H$ liée à $F$ et $G$ par la relation $\overline{H} = \overline{F} \times \overline{G}$. L'IVE de la fdr $H$ de $Z$, existe et il est notée par $\gamma$ où $\gamma := \frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}$. Soit $x_F$, $x_G$ et $x_H$ les points terminaux du support de $F$, $G$ et $H$ respectivement. Enimahl et al [8] ont proposé une adaptation générale des estimateurs existants dans les trois cas suivants ($\gamma_1, \gamma_2 > 0$, $\gamma_1, \gamma_2 < 0$ et $\gamma_1, \gamma_2 = 0$).

Leur estimateur sont basés sur un estimateur standard de l'indice de queue (non adapter à la censure) divisé par proportion de données non censurées dans les plus grands $k$ de $Z$.

$$\widehat{\gamma}_1^{(\bullet,c)}(k) = \frac{\widehat{\gamma}^{(\bullet)}(k)}{\widehat{p}(k)}.$$

avec $\widehat{p} = \frac{1}{k}\sum_{i=1}^{k}\delta_{[n-i+1:n]}$, $\delta_{[j:n]}$ est le concomitant de la j-ème statististique d'ordre, c'est -à-dire, $\delta_{[j:n]} = \delta_i$ si $Z_{j:n} = Z_i, 1 \leq i \leq n$.

$\widehat{\gamma}^{\bullet}$ peut être n'importe quel estimateur non adapté à la censure, en particulier $\widehat{\gamma}_n^H$, $\widehat{\gamma}_n^M$, ..., et $\widehat{p}$ estime $p = \dfrac{\gamma_2}{\gamma_1 + \gamma_2}$, où $p$ représente la proportion des données observée dans la queue à droite de la distribution.

### 2.2.2 Estimateur des quantiles extrêmes

Le principal estimateur des quantile extrêmes $x_\varepsilon$ d'ordre $(1-\varepsilon)$ sous censure aléatoire disponible dans la littérature a été proposé par Beirlant et al [2] en 2007 [1] et par Einmahl et al [8] en 2008. Il est donné par la définition suivante.

**Définition 2.2.1 (Estimation du quantile extrême)** *:*

*L'estimation du quantile extrême sous censure aléatoire est défini par :*

$$\widehat{x}_\varepsilon^{(\bullet,c)} = Z_{n-k:n} + \widehat{a}^{(\bullet,c)} \frac{\left(\left(1-\overline{\widehat{F}}^{KM}(Z_{n-k,n})\right)\diagup\varepsilon\right)^{\widehat{\gamma}_1^{(\bullet,c)}} - 1}{\widehat{\gamma}_1^{(\bullet,c)}}.$$

*où*

$$\widehat{a}^{(\bullet,c)} = \frac{Z_{n-k:n} M_n^1 \left(\frac{1}{2}\left(1-\frac{\left(M_n^1\right)^2}{M_n^2}\right)^{-1}\right)}{\widehat{p}}.$$

*avec* $M_n^r, r = 1,2$ *est défini dans* 2.4 *et* $\overline{\widehat{F}}^{KM}$ *l'estimateur de Kaplan-Meier, pour* $t < Z_{n:n},$ *,* $i = 1, ..., n$ *est défini par*

$$\overline{\widehat{F}}^{KM}(t) = \prod_{Z_{i:n} \leq t} \left(1 - \frac{\delta_{[i:n]}}{n-i+1}\right).$$

*où* $Z_{1:n} \leq ... \leq Z_{n:n}$ *les statistiques d'ordre associées à* $Z_1, ..., Z_n$*.*

# CONCLUSION

Dans ce mémoire, on a présenté des généralités relatives à la TVE, en plus de la théorie de la censure qui tente d'estimer les quantiles extrêmes. Plusieurs méthodologies ont été trouvées dans le but d'estimer ces quantités, que ce soit pour des données complètes ou censurée.

Les points abordés dans ce mémoire fournit une bases pour d'autres pistes de recherche intéressantes qui méritent d'être étudiées de près.

Le point à envisagé est de chercher d'autre méthodes dans le but d'améliorer l'estimation des quantiles extrêmes, avec une illustration à travers une application sur une série de données simulées et réelles censurées.

Il arrive assez souvent qu'une série statistique, ayant des données extrêmes, présente des données tronquées. Il serait aussi intéressant de faire le même travail, par ce que ce cas mérite d'être considéré attentivement vu.

# BIBLIOGRAPHIE

[1] Beirlant, J, Goegebeur, Y, Segers, J, & Teugels, J. (2006). Statistics of Extrêmes : Theory and Applications. John Wiley.

[2] 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.

[3] Beirlant, J., and Guillou, A. (2001). Pareto index estimation under moderate right censoring. Scand. Actuar. J., 111 125.

[4] Bensmail, N. (2019). Estimation et test de l'indice des valeurs extrêmes en présence de censure. Mémoire de mastere, l'université de Biskra.

[5] Cláudia, N , and Maria, J. (2019) . Forecasting & assessing risk of individual electricity peaks : mathematics of planet earth, Springer, International Publishing.

[6] Dekkers, A. L., and De Haan, L. (1989). On the estimation of the extremevalue index and large quantile estimation. Ann. Statist., 17951832.

[7] Embrechts P.,Klüppelberg C., & Mikosh T. (1997). Modelling extremal events for Insurance and finance. Springer, Verlag Berlin Heidelberg.

[8] Einmahl, J. H., Fils-Villetard, A., & Guillou, A. (2008). Statistics of Extremes Under Random Censoring. Bernoulli, 14(1), 207227.

[9] Fisher R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Math. Proc. Cambridge Philos. Soc., 24(02), $180-190$.

[10] Fovea, Groupe. L'analyse de survie des études cliniques. Département Biométrie de Fovea. http ://www.Fovea-group.com/cro/pdf/fms8. Pdf.

[11] Gardes, L., Girard, S. A moving window approach for nonparametric estimation of the conditional tail index. Journal of Multivariate Analysis, 99 : $2368-2388$.

[12] Gharbi, D. (2020). Estimation des quantiles extrêmes. Mémoire de mastere, l'univèrsité de Biskra.

[13] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math., 423 453.

[14] Göran. B, (2012). Event history analysis with R. (Chapman & Hall/CRC The R Series).Taylor & francis group, LLC.

[15] de Haan, L. & Ferreira, A. (2006). Extreme value theory : an introduction. Springer-Verlag, New York.

[16] Hill, B. M.(1975). A simple general approach to inference about the tail of a distribution. Ann. Statist., 3(5), 11631174.

[17] Houhou, R. (2019). Statistiques des Extrêmes sous Données Censurés. Memoire de mastere, l'université de Biskra.

[18] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly J. R. Methodol.Soc., 81(348), $158-171$.

[19] Karboussa, M. (2020). Sur l'estimation des quantiles extrêmes conditionnels dans le cas des données censurées. Memoire de mastere. Université de Ourgla.

[20] Mason, D. M. (1982). Laws of large numbers for sums of extreme values. The Annals of Probability, $754 - 764$.

[21] Mokrani, A. (2020). Estimation de l'Indice des Valeurs Extrêmes sous Troncature. Memoire de mastere. Université de Biskra. Algerie.

[22] M'ziou, I. (2014). Estimation non paramétrique. Mémoire de master. Université de Biskra. Algerie.

[23] Ndao, P., Diop, A., & Dupuy, J. F. (2014). Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. Comput. Statist. Data Anal., 79 : 6379.

[24] Pickands, J. (1975). Statistical inference using Extreme order statistics. The Annals of Statistics, 3, $119 - 131$.

[25] Reiss, R.D, & Thomas, M. (2007). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields.Birkhäuser, Basel.

[26] Saint. P. P. (2012). Introduction à l'analyse des durées du survie.

[27] Semmari, M. (2020). Sur l'Analyse de Survie et Application. Mémoire de mastere. Université de Biskra. Algerie.

[28] Soltan, L. (2017). Analyse des Valeurs Extrême en Presence de Censure.Thèse de doctorat.Université de Biskra, Algerie.

[29] Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. J. Amer. Statis. Assoc.69 , 169173.

[30] Von Mises, R. (1936). La Distribution de la plus grande des $n$ valeurs. Selected papers, Amer. Math. Soc., $271 - 294$.

[31] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association, 73 (364), 812

[32] Worms, J. and Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. Extremes, 17 : 337358.

[33] Zerroukhi, H. (2019). Estimation de l'Indice des Valeurs Extrêmes sous Censure. Mémoire de mastere, l'université de Biskra.

# ANNEXE : ABRÉVIATIONS ET NOTATIONS

Les différentes abréviations et notation utilisées tout au long de cette thèse sont expliquées ci-dessous.

| | | |
|---|---|---|
| IVE | : | Indice des valeurs extrêmes. |
| TEV | : | Theorie des valeurs extremes. |
| $F_n$ | : | Fonction de répartition empirique. |
| $F^{-1}$ | : | Inverse généralisé de $F$. |
| iid | : | Indépendantes et identiquement distribué. |
| $\mathbb{1}_A$ | : | Fonction indicatrice de l'ensemble $A$. |
| v.a | : | variable aléatoire. |
| $\mathbb{R}$ | : | Ensemble de valeurs réelles. |
| $Q$ | : | Fonction de quantile. |
| $Q_n$ | : | Fonction de quantile empirique. |
| $\overline{F} = S$ | : | Fonction de queue ou survie. |
| $\overline{F}_n$ | : | Fonction empirique de queue. |
| $U$ | : | Fonction quantile de queue. |
| $U_n$ | : | Fonction empirique de queue. |

| | | |
|---|---|---|
| $\xrightarrow{P}$ | : | Convergence en probabilité. |
| $\xrightarrow{P.s}$ | : | Convergence presque surment. |
| $x_F$ | : | Point terminal inferieure. |
| $x^F$ | : | Point terminal. |
| $X_{1:n}$ | : | Minimum de $X_1, ..., X_n$. |
| $X_{n:n}$ | : | Maximum de $X_1, ..., X_n$. |
| $X_{j:n}$ | : | $j^{ème}$statistique d'ordre. |
| $X_1, ..., X_n$ | : | Une suite de $n$ variable aléatoire. |
| $\mathcal{H}_\gamma$ | : | Famille de la loi de valeurs extrêmes généralisé. |
| $\Lambda$ | : | Loi de Gumbel. |
| $\Phi$ | : | Loi de Fréchet. |
| $\Psi$ | : | Loi de Weibul. |
| $\mathcal{DA}\,(.)$ | : | Domain d'attraction. |
| $\delta_i$ | : | Indice indicateur de censure. |
| $\widehat{\gamma}_n^H$ | : | Estimateur de hill. |
| $\widehat{\gamma}_n^P$ | : | Estimateur de Pickand. |
| $\widehat{\gamma}_n^M$ | : | Estimateur de Moment. |
| $\widehat{\gamma}_n^{(c,h)}$ | : | Estimateur harmonique avec des données censurées. |

# الملخص

في الإحصاء ونظرية الاحتمالات، الدالة الكمية هي القيم التي تقسم مجموعة البيانات إلى فترات من نفس الإحتمال المتساوي. يلعب تقدير هذه الكمية دورا مهمًا في إدارة المخاطر والإحصاءات، البيئية والطب والموثوقية وغيرها من التطبيقات، لهذا نعطي أولاً التقدير الكلاسيكي وشبه البراميتري بناء على نظرية القيم المتطرفة الذي تم اقتراحه بواسطة تم اقتراحه بواسطة (Hill, Pickands et moment) تم اقتراحه بواسطة في إطار التقدير لبيانات مكتملة. ثانيا تم إجراء تجميع و تلخيص لتقدير ، (2008) Einmahl et al استناد على نتائج (1958) Kaplan-Meier و القيم المتطرفة في ظل وجود بيانات غير مكتملة

الكلمات المفتاحية: نظرية القيمة القصوى، تحليل البقاء، الرقابة العشوائية، التقدير التطرف الكمي

# Résumé

**En statistiques et en théorie des probabilités, les quantiles sont les valeurs qui divisent un jeu de données en intervalles de même probabilité égale. L'estimation de cette quantité joue un rôle important dans la gestion des risques, les statistiques environnementales, médecine, fiabilité et entre autres applications, pour cela on donne dans un premier temps, l'estimation classique et semi-paramétrique basée sur la TVE qui a été proposé par (Hill, Pickands et moment) dans le cadre de l'estimation sans censure. Dans un second temps, on effectue une synthèse sur l'estimation de Einmahl et al (2008) en se basant sur les résultats de Kaplan-Meier (1958) et la TVE dans le cadre de l'estimation en présence de censure.**

**Mots Clés: Théorie des valeurs extrêmes, analyse de survie, censure aléatoire, estimation des quantiles extrêmes.**

# Abstract

**In statistics and probability theory, quantiles are the values that divide a data set into intervals of the same equal probability. The estimation of this quantity plays an important role in risk management, environmental statistics, medicine, reliability, and among other applications, for this we first give the classic and semi-parametric estimation based on the TVE which was offered by (Hill, Pickands, and Moment) as part of the uncensored rating. In the second step, we perform a synthesis of the estimate of Einmahl et al (2008) based on the results of Kaplan-Meier (1958) and the TVE in the context of the estimate in the presence of censorship.**

Keywords: **Extreme value theory, survival analysis, random censoring, estimation of extreme quantiles.**