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 **:**

**MASTER en Mathématiques**

Option : **Statistique**

Par

**BACHIR Nadjoua**

Titre :

# Estimation de l'indice pour les lois à queue de type weibull

Membres du Comité d'Examen :

| | | | |
|---|---|---|---|
| Pr. | **Meraghni Djamel** | UMKB | Président |
| Pr. | **Brahimi Brahim** | UMKB | Encadreur |
| Dr. | **Berkane Hassiba** | UMKB | Examinateur |

Juin 2018

# Dédicace

Je dédie cet humble travail pour mon parents.

Et mes soeurs.

## REMERCIEMENTS

Je glorfie Allah le tout puissant de m'avoir donnée courage et patience qui m'ontpermis d'accomplir ce travail.

Le moment est venu pour moi d'exprimer ma plus grande gratitude envers tous celles et ceux qui m'ont aidé et encouragé dans l'accomplissement de cette mémoire dont le prix était l'obtention de diplôme de master.

Ainsi qu'il est d'usage, je tiens tout d'abord à exprimer toute ma reconnaissance à mon encadreur Monsieur, professeur, Brahimi Brahim, pour sa disponibilité, son écoute, ses conseils éclairés et ses encouragements.

Je tiens également à remercier chaleureusement Monsieur docteur, Ben Atia Fateh Pour son aide et son soutien.

Je profite aussi de l'occasion pour remercier tous les membres mon enseigenement, qu'ils sont enseignés moi de cinq année de université.

merci.

# Table des matières

**Remerciements** **ii**

**Table des matières** **iii**

**Introduction** **1**

**1 Théorie des valeurs extrêmes et fiabilité** **3**

1.1 Théorie des valeurs extrêmes . . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.1.1 Statistiques d'ordre . . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.1.2 Lois limites des valeurs extrêmes . . . . . . . . . . . . . . . . . . . 7

1.1.3 Loi max-stable . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

1.1.4 Fonctions à variations lentes . . . . . . . . . . . . . . . . . . . . . 9

1.1.5 Fonctions à variations régulières . . . . . . . . . . . . . . . . . . . 9

1.1.6 Condition de second ordre sur les fonctions à variations lentes . . . 11

1.1.7 Caractérisations Générales . . . . . . . . . . . . . . . . . . . . . . 11

1.1.8 Domaine d'attraction de Weibull . . . . . . . . . . . . . . . . . . . 12

1.2 Fiabilité . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

1.2.1 Durée de vie . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

1.2.2 Définition de la Fiabilité . . . . . . . . . . . . . . . . . . . . . . . 14

1.2.3 Fonction de fiabilité et fonction de défaillance . . . . . . . . . . . . 15

1.2.4 Indicateurs de fiabilité . . . . . . . . . . . . . . . . . . . . . . . . 17

1.2.5 Fiabilité des systèmes . . . . . . . . . . . . . . . . . . . . . . . . . 17
1.2.6 Utilisation de la loi de Weibull pour modéliser les données de fiabilité 19

**2 Estimation de l'indice de queue de Weibull 21**

2.1 Les lois à queues de type Weibull . . . . . . . . . . . . . . . . . . . . . . 21
2.2 Estimation de l'indice de queue . . . . . . . . . . . . . . . . . . . . . . . 22
2.2.1 Estimateur de type Hill . . . . . . . . . . . . . . . . . . . . . . . . 22
2.2.2 Utilisation de poids . . . . . . . . . . . . . . . . . . . . . . . . . . 27
2.2.3 Estimateur Zipf . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
2.2.4 Un estimateur débiaisé . . . . . . . . . . . . . . . . . . . . . . . . 29
2.2.5 Choix du nombre $k_n$ de statistiques d'ordre . . . . . . . . . . . . . . 30

**Conclusion 31**

**Bibliographie 32**

**Annexe B : Abréviations et Notations 33**

# Introduction

La modélisation des événements extrêmes (ouragan, tremblement de terre ou inondation, crues, crises ... nancières, krachs, chocs pétroliers) est aujourd'hui un champ de recherches particulièrement actif, notamment par l'importance de leurs impacts économiques et sociaux. En particulier, depuis quelques années, on note un intérêt croissant pour l'application de la Théorie des Valeurs Extrêmes (TVE) pour la modélisation de tels événements. La théorie des valeurs extrêmes permet d'évaluer les événements rares et les pertes associées à leur apparition. En d'autres termes lorsqu'une perte importante survient, cette théorie permet d'en évaluer l'ampleur. De plus cette théorie acquiert une importance particulière du fait qu'elle s'intéresse directement à la queue de la loi.

Nous parlerons également de la fiabilité, en particulier de la fiabilité des machines. Dans ce mémoire, l'étude sera seulement sur la loi de Weibull.

Notre mémoire est organisé comme suit :

**Chapitre 1 :** Divisez ce chapitre en deux parties, la première partie nous proposons la théorie des valeurs extrêmes. Cette théorie est basée sur la statistique d'ordre d'un échantillon de taille $n$, le comportement asymptotique du maximum. Cette approche est basé sur le théorème de Fisher et Tippett qui sert à modéliser les maximums preuvés en 1928, la preuve complète a été proposée par Gnedenko 1943.

La deuxième partie nous proposons la fiabilité. Dans cette section, nous examinons en particulier la fiabilité des machines. Où nous fournissons des lois de fiabilité en général, puis nous utilisons la loi de Webull pour modéliser les données de fiabilité.

**Chapitre 2 :** Dans ce chapitre, on s'intéresse à l'estimation de l'indice pour les lois à queue de type weibull. On a d'abord estimateur de type Hill et sa propriétés avec comparaison avec d'autres estimations. Ensuite nous utilisation de poids et estimateur Zipf. Enfin nous touchons un estimateur débiaisé.

# Chapitre 1

# Théorie des valeurs extrêmes et fiabilité

## 1.1 Théorie des valeurs extrêmes

La théorie des V.E a été développée pour l'estimation de la probabilité d'occurrence d'événements extrêmes. Elle permet d'extrapoler le comportement de la queue de distribution des données à partir des plus grandes observations.
Dans ce chapitre, nous parlons spécifiquement des lois à queue de type weibull.

### 1.1.1 Statistiques d'ordre

**Définition 1.1.1 (Statistique d'ordre)** *La statistique d'ordre d'une suite de $n$ variables aléatoires (v.a's) indépendantes et identiquement distribuées (i.i.d) $(X_1, ..., X_n)$ est le réangement croissant de cet échantillon elle est dénotée par $X_{1,n}, ..., X_{n,n}$ telle que*

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

*pour $1 \leq i \leq n$, la variable $X_{i,n}$ est appelée la $i$-ième statistique d'ordre (ou statistique d'ordre $i$). Elles sont nécessairement dépendantes en raison des relations d'inégalité entre* *elles.*

*La première statistique d'ordre, noté* $X_{1,n}$ *est le minimum de l'échantillon de taille* $n$

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

*La dernière statistique d'ordre noté* $X_{n,n}$ *est le maximum de l'échantillon de taille* $n$

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

*On obtient la correspondance entre minimum et maximum par la relation suivante :* $\min(X_1, ..., X_n) = -\max(-X_1, -X_2..., -X_n)$ *ainsi, tous les résultats que nous allons présenter pour les maxima pourront être transposés pour les minima.*

**Définition 1.1.2 (La distibution et la densité de la $i$-ème statistique d'ordre)** *La distribution de la* $i$*-ème statistique d'ordre*$(F_{i,n})$ *pour* $1 \leq i \leq n$ *est définit comme suit :*
*On pose* $Y_n = \sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}}$*(le nombre de répétitions de l'évenement* $X \leq x$ *en* $n$ *expériences indépendantes), alors la v.a* $Y_n$ *suit une loi binomial* $B(n, F(x))$*.Donc on a :*

$$P(Y_n = j) = C_n^j \left(F(x)\right)^j \left(1 - F(x)\right)^{n-j}.$$

*Alors,*

$$\begin{aligned}
F_{i,n}(x) &= P(X_{i,n} \leq x) \\
&= P(\textit{au moins } i \ X_k \textit{ sont plus petites que } x) \\
&= P\left(Y_n \geq i\right) \\
&= \sum_{r=i}^{n} P\left(Y_n = r\right) \\
&= \sum_{r=i}^{n} C_n^r \left(F(x)\right)^r \left(1 - F(x)\right)^{n-r} \quad , \quad -\infty < x < +\infty
\end{aligned}$$

*Sa densité est*

$$f_{i,n}(x) = \frac{n!}{(i-1)!(n-i)!} \left[F(x)\right]^{i-1} \left[1-F(x)\right]^{n-1} f(x) \quad , x \in \mathbb{R}$$

*où $f$ est la densité associée à $F$.*

**Proposition 1.1.1 (La distribution et la densité du $X_{1,n}$ et $X_{n,n}$)** *Soient $X_1, ...X_n$, $n$ v.a's i.i.d, de fonction de distribution commune $F$.*

– *La fonction de repartition de $X_{1,n}$ est*

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

– *La fonction de repartition de $X_{n,n}$ est*

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

– *La densite de $X_{1,n}$ est*

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

– La densite de $X_{n,n}$ est

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

**Définition 1.1.3 (Fonction de survie)** *On appelle fonctions de survie d'une v.a $X$ la fonction définie sur $\mathbb{R}$ par*

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

*elle est aussi appelée la queue de la distribution.*

**Définition 1.1.4 (Fonction de quantile)** *La fonction quantile associée à $F$ est l'inverse généralisé de $F$ c'est une fonction continue à gauche définie par :*

$$Q(t) = F^{\longleftarrow}(t) = \inf \left\{ x \in R : F(x) \geq t \right\}, \quad 0 < t < 1$$

**Définition 1.1.5 (La fonction de répartition empirique)** *La fonction de répartition empirique associée à un échantion $(X_1, ..., X_n)$ est définie par :*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}, \; x \in \mathbb{R}.$$

*avec*

$$F_n(X_{i,n}) = \frac{i}{n}, \text{ pour } i = 1, ..., n.$$

**Définition 1.1.6 (Fonction du quantile empirique)** *La fonction du quantile empirique d'un échantillon $(X_1, ..., X_n)$ est définie par*

$$Q_n(t) = F_n^{\longleftarrow}(t) = X_{i,n}, \frac{i-1}{n} < t \leq \frac{i}{n} \text{ pour } i = 1, ..., n.$$

### 1.1.2 Lois limites des valeurs extrêmes

Le théorème ci-dessous est fondamental en théorie des valeurs extrêmes puisqu'il précise les lois limites possibles pour le maximum renormalisé. Ce théorème est l'équivalent du théorème central limite pour les queues de distribution.

**Théor■me 1.1.1 (Fisher-Tippett(1928), Gnedenko(1943))** *Soit* $(X_i)_{i\geq 1}$ *une suite des v.a's, i.i.d. et s'il existe deux suites de constantes de normalisation* $(a_n)_{n\geq 1} > 0$ *et* $(b_n)_{n\geq 1} \in \mathbb{R}$ *et une loi non-dégénérée* $H$ *telles que pour tout* $x \in \mathbb{R}$

$$\lim_{n\to\infty} P\left[\frac{X_{n,n} - b_n}{a_n} \leq x\right] = \lim_{n\to\infty} F^n(a_n x + b_n) = H_\xi(x).$$

*où* $H_\xi$ *est la fonction de répartition de la loi des valeurs extrêmes(EVD). Alors à une translation et un changement d'échelle près. La lois limite peut prendre trois forme possible :*

$$\text{(Loi de Fréchet)} \quad \Phi_\xi(x) = \begin{cases} 0 & x \leq 0 \\ \exp\left[-x^{-\frac{1}{\xi}}\right] & x > 0 \end{cases} \qquad \text{et } \xi > 0$$

$$\text{(Loi de Weibull)} \quad \Psi_\xi(x) = \begin{cases} \exp\left[-(-x)^{-\frac{1}{\xi}}\right] & x < 0 \\ 1 & x \geq 0 \end{cases} \qquad \text{et } \xi < 0$$

$$\text{(Loi de Gumbel)} \quad H_0(x) = \Lambda(x) = \exp\left[-\exp(-x)\right]$$

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

**Définition 1.1.7 (Jenkinson-von Mises)** *Il est possible de rassembler les trois familles de lois de valeur extrême en une seule famille paramétrique* $(H_\xi(x), \xi \in \mathbb{R})$ *dite famille des lois des valeurs extrêmes généralisées notée GEV (Generalized Extreme Value). Elle est paramétrée par un seul nombre réel* $\xi \in \mathbb{R}$ *appelé indice des valeurs extrêmes notée EVI(Extreme Value Index) ou indice de queue, mais toujours à un facteur de changement d'échelle et de translation près. La fonction de distribution de la famille* $H$ *des valeurs* *extrêmes généralisée (GEV) est, pour* $\xi \in \mathbb{R}$,

$$H_\xi(x) = \begin{cases} \exp\left[-\left(1+\xi x\right)^{\frac{-1}{\xi}}\right] & si\ \xi \neq 0,\ \ 1+\xi x > 0. \\ \exp\left[-\exp\left(-x\right)\right] & si\ \xi = 0,\ \ x \in \mathbb{R}. \end{cases}$$

*Nous avons alors les correspondances suivantes*

$$\Phi_\xi(x) = H_\xi\left[\left(x-1\right)/\xi\right],\ x \in \mathbb{R}.$$
$$\Psi_\xi(x) = H_\xi\left[\left(x+1\right)/\xi\right],\ x \in \mathbb{R}.$$
$$\Lambda(x) = H_0(x).$$

*Pour les variables non centrées et non réduites, on peut écrire* $H_\xi(x)$ *sous une forme plus générale (appelée forme paramétrée de von Mises) dans laquelle on fait apparaître un paramètre de localisation* $\mu$ *et un paramètre d'échelle* $\sigma$

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left[-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right] & si\ \xi \neq 0,\ 1+\xi\dfrac{x-\mu}{\sigma} > 0. \\ \exp\left[-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right] & si\ \xi = 0,\ x \in \mathbb{R}. \end{cases},\ \mu \in \mathbb{R},\ \sigma \in \mathbb{R}_+^*$$

**Remarque 1.1.1** *On peut passer de l'une des trois lois à l'autre relativement facilement.Supposer* $X > 0$

$$X \sim \Phi_\alpha \Longleftrightarrow \ln\left(X^\alpha\right) \sim \Lambda \Longleftrightarrow -X^{-1} \sim \Psi_\alpha.$$

### 1.1.3 Loi max-stable

**Définition 1.1.8** *La loi* $H$ *est dite max-stable si pour tout* $n \geq 2$, $(X_1, ..., X_n)$ *étant des v.a's, i.i.d de loi* $H$, *il existe* $a_n > 0$ *et* $b_n \in \mathbb{R}$, *tels que* $\left(X_{n;n} - b_n\right)/a_n$ *suit la loi* $H$.
*Alors la loi limite des valeurs extrêmes est une loi max-stable.*

### 1.1.4 Fonctions à variations lentes

Les fonctions à variations lentes sont généralement dénotées par $L(.)$.

**Définition 1.1.9** *Une fonction mesurable positive $L$ est dite à variation lente à l'infini si :*

$$\lim_{x\to\infty}\frac{L(\lambda x)}{L(x)}=1,\ \forall\lambda>0.$$

**Corollaire 1.1.1 (La représentation de Karamata)** *$L(.)$ est une fonction à variation lente si et seulement si pour tout $x>0$ :*

$$L(x)=c(x)\exp\left\{\int_1^x\frac{\varepsilon(t)}{t}dt\right\},$$

*où $c:\mathbb{R}_+\longrightarrow\mathbb{R}_+$, $\varepsilon:\mathbb{R}_+\longrightarrow\mathbb{R}_+$ telles que :*

$$\lim_{x\to\infty}c(x)=c\in\,]0,+\infty[\ \ et\lim_{x\to\infty}\varepsilon(x)=0.$$

**Remarque 1.1.2** *Si la fonction $c$ est constante, on dit que $L$ est normalisée.*

### 1.1.5 Fonctions à variations régulières

**Définition 1.1.10** *Une fonction mesurable $U:\mathbb{R}^+\longrightarrow\mathbb{R}^+$ est à variations régulières d'indice $\alpha\in\mathbb{R}$ à l'infini (on notera par la suite $U\in\mathcal{RV}_\alpha$) si pour tout $\lambda>0$ :*

$$\lim_{x\to\infty}\frac{U(\lambda x)}{U(x)}=\lambda^\alpha.$$

*Si $\alpha=0$, on dit que la fonction $U$ est à variations lentes ($U\in\mathcal{RV}_0$).En remarquant que si $U$ est à variation régulière d'indice $\alpha$ alors $U(x)/x^\alpha$ est à variation lente.*

*Toute fonction à variation régulière d'indice $\alpha$ peut toujours s'écrire sous la forme :*

$$U(x)=x^\alpha L(x),\ L\in\mathcal{RV}_0.$$

**Remarque 1.1.3** *Soit $U$ une fonction à variation régulière d'indice $\alpha$. En utilisant le fait que $U(x) = x^{\alpha}L(x)$, on déduit facilement que pour tout $x > 0$ :*

$$U(x) = c(x)\exp\left\{\int_1^x t^{-1}\varepsilon(t)dt\right\}.$$

*où $c$ et $\varepsilon$ sont des fonctions positives telles que :*

$$\lim_{x\to\infty} c(x) = c \in \,]0, +\infty[\ \text{ et } \lim_{t\to\infty}\varepsilon(t) = \alpha.$$

**Théorème 1.1.2** *Supposant que pour une fonction $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$ mesurable, il existe une fonction $a > 0$ telle que pour tout $x > 0$ :*

$$\lim_{t\to\infty}\frac{f(xt) - f(t)}{a(t)} = W(x)\begin{cases} c\dfrac{x^{\alpha}-1}{\alpha} & si\ \alpha \neq 0. \\ c\log(x) & si\ \alpha = 0. \end{cases}$$

*Où la fonction $W$ n'est pas constante, $c \neq 0$. La fonction $a$ est appelée fonction auxiliaire de $f$. Si $f$ vérifé l'hypothèse du théorème ci–dessus avec $\alpha \neq 0$ alors $f \in \mathcal{RV}_{\alpha}$.*

**Théorème 1.1.3** *Soit $f$ et $g$ deux fonctions positives et continues. Soit $0 < a < b < \infty$ et soit $(a_n)$ et $(b_n)$ deux suites :*

$$\lim_{n\to\infty}\sup a_n = \infty,\ \lim_{n\to\infty}\sup\frac{a_n}{a_{n+1}} = 1.$$

$$\lim_{n\to\infty} b_n f(a_n x) = g(x),$$

*pour tout $x \in [a, b]$. Alors $f$ est à variation régulière.*

**Théorème 1.1.4** *Soit $f$ une fonction positive et monotone, $g$ une fonction $(a_n)$ et $(b_n)$ deux suites telles que :*

$$\lim_{n\to\infty}\sup b_n = \infty,\ \lim_{n\to\infty}\sup\frac{a_n}{a_{n+1}} = 1.$$

$$\lim_{n\to\infty} b_n f(a_n t) = g(t),$$

*pour $t$ est dense dans sous-intervalles de $[0,\infty[$. Alors $f$ est à variation régulière.*

**Corollaire 1.1.2** *Si $U \in \mathcal{RV}_\alpha$, alors :*

$$U(x) \longrightarrow \begin{cases} \infty & si\ \alpha > 0, \\ 0 & si\ \alpha < 0, \end{cases} \quad quand\ x \longrightarrow 1.$$

### 1.1.6 Condition de second ordre sur les fonctions à variations lentes

On fait l'hypothèse du second ordre habituelle sur la fonction à variations lentes,à savoir quand $x \to \infty$

$$\log\left(\frac{L(\lambda x)}{L(x)}\right) \sim b(x) K_\rho(\lambda), \tag{1.1}$$

où $\lambda > 0$, avec

$$K_\rho(\lambda) = \int_1^\lambda t^{\rho-1} dt.$$

Le paramètre de second ordre $\rho \leq 0$ contrôle la vitesse de convergence de $L(\lambda x)/L(x)$ vers 1. Plus $\rho$ est proche de 0, plus la vitesse de convergence est faible. On a toujours $b(x) \to 0$ quand $x \to \infty$ et $|b| \in \mathcal{RV}_\rho$.

Cette hypothèse est très souvent utilisée pour étudier le comportement asymptotique des estimateurs de l'indice des valeurs extrêmes.

### 1.1.7 Caractérisations Générales

**Définition 1.1.11 (Point terminal)** *On appelle point terminal de la fonction $F$, le réel $x_F$ définit par :*

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

**Définition 1.1.12 (Queue équivalente)** *Deux distributions $F$ et $H$ seront dites à queue* *de distribution équivalente (tail equivalent) si elle possède le même point terminal, ie.* $x_F = x_H$*, et si pour une certaine constante positive* $c > 0$

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x)}{\overline{H}(x)} = c.$$

**Définition 1.1.13 (Domaine d'Attraction)** *On dit qu'une distribution* $F$ *appartient au domaine d'attraction de* $H$*, et on note* $F \in D(H_\xi)$ *s'il existe des suites réelles* $(a_n)_{n\geq 1} > 0$ *et* $(b_n)_{n\geq 1} \in \mathbb{R}$ *telles que :*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = H_\xi(x).$$

**Définition 1.1.14 (Fonction queue)** *La fonction queue est définie par :*

$$U(t) = Q\left(1 - \frac{1}{t}\right) = \left(1/\overline{F}\right)^{\longleftarrow},\ 1 < t < \infty.$$

*où* $Q$ *est la fonction quantile associée à* $F$*.*

### 1.1.8 Domaine d'attraction de Weibull

Dans cette section, nous caractérisons le domaine d'attraction de $\Psi_\xi$ pour $\xi < 0$ Un fait important, bien que pas du tout évident, est que tout $F$ dans $D(\Psi_\xi)$ a un point terminal $x_F$ finite.

**Théorème 1.1.5** *La fonction de répartition* $F$ *appartient au domaine d'attraction de la loi de Weibull avec un indice de valeur extrême* $\xi < 0$ *si et seulement si* $x_F < +\infty$ *et* $1 - F^*$ *est une fonction à variations régulières d'indice* $1/\xi$ *avec*

$$F^*(x) = \begin{cases} 0 & si\ x \leq 0 \\ F\left(x_F - x^{-1}\right) & si\ x > 0 \end{cases}$$

*Les suites de normalisation* $(a_n)$ *et* $(b_n)$ *sont données par :*

$$a_n = x_F - U(n) = x_F - Q\left(1 - \frac{1}{n}\right), \; b_n = x_F.$$

**Corollaire 1.1.3** *On dit que* $F \in D(\Psi_\xi)$ *avec* $\xi < 0$ *si et seulement si* $x_F \leq +\infty$ *et*

$$F(x) = 1 - (x_F - x)^{-\frac{1}{\xi}} L\left((x_F - x)^{-1}\right), \; L(.) \in \mathcal{RV}_0.$$

**Corollaire 1.1.4** *(Condition de Von Mises)*

*Soit* $F$ *une fonction de répartition absolument continue avec une densité* $f$ *positive sur un certain intervalle* $[z, x_F[$

*si*

$$\lim_{x \uparrow x_F} \frac{(x_F - x) f(x)}{\bar{F}(x)} = -\frac{1}{\xi} > 0,$$

*alors* $F \in D(\Psi_\xi)$.

**Proposition 1.1.2** *(Propriétés de fermeture)*

*Soient* $F$ *et* $H$ *deux distributions de probabilité. Supposons que* $F \in D(H_1)$ *avec* $a_n > 0$ *et* $b_n$ *les constantes de normalisation ie,*

$$\lim_{n \longrightarrow \infty} F^n(a_n x + b_n) = H_1(x) \text{ alors } \lim_{n \longrightarrow \infty} H^n(a_n x + b_n) = H_1(ax + b),$$

*si et seulement si pour* $a > 0$ *et* $b \in \mathbb{R}$, $F$ *et* $H$ *sont à queue de distribution équivalente.et si*

$$H_1 = \Psi_\xi \text{ alors } b = 0 \text{ et } c = a^{\frac{1}{\xi}}.$$

## 1.2 Fiabilité

### 1.2.1 Durée de vie

#### Définition de durée de vie

Le terme de durée de vie est employé pour désigner le temps qui s'écoule jusqu'à la survenue d'un événement particulier. cette notion de durée de vie est observée dans plusieurs domaines tels que la médecine, la fiabilité, l'assurance, l'économie, la biologie et d'autres domaines. Dans cette étude, nous nous intéressons au domaine de la fiabilité.

### 1.2.2 Définition de la Fiabilité

(1) La fiabilité est l'une des composantes essentielles de la qualité d'un produit et elle est retenue en tant que critère fondamental pour leur élaboration. Elle est prise en considération dés le stade de la conception.

(2) La fiabilité est la caractéristique d'un dispositif exprimée par la probabilité que ce dispositif accomplisse une fonction requise dans des conditions d'utilisation et pour une période de temps déterminée.

**Remarque 1.2.1** – *Plus une machine est constituée d'un nombre important de composants plus la fiabilité de cette dernière à tendance à diminuer. Lorsque les composants sont trop nombreux ou trop complexes, il arrive fréquemment un moment où la maîtrise de la fiabilité n'est plus possible et l'hypothèse d'une défaillance très probable.*

– *Un ensemble composé de pièces de très haute qualité ne garantie pas nécessairement une grande fiabilité après assemblage (les interactions qui se produisent entre les composants diminuent la capacité de l'ensemble).*

– *De même, une grande fiabilité sous certaines conditions ne garantie pas forcément une grande fiabilité sous d'autres conditions.*

– *La meilleure connaissance de la fiabilité provient de l'analyse des défaillances lorsque* *les produits sont en service.*

– *La non-fiabilité d'un produit ou d'un bien augmente les coûts de l'après vente( application des garanties, frais judiciaires, etc...). Construire plus fiable augmente les coûts de conception et de production. Le coût total du produit prendra en compte ces deux tendances.*

### 1.2.3 Fonction de fiabilité et fonction de défaillance

On note par$T$ une variable aléatoire continue à valeurs positives et possédant une densité de probabilité $f$.

**Définition 1.2.1 (Fonction de défaillance)** *La fonction défaillance de $T$ est définie par*

$$F(t) = P(T \leq t) = \int_0^t f(t)dt \text{ pour } t \geq 0.$$

*$F(t)$ est la probabilité qu'un dispositif prélevé au hasard dans la population considérée ait une défaillance avant l'instant $t$.*

**Définition 1.2.2 (Fonction de fiabilité)** *La fonction de fiabilité est définie par*

$$R(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(x)dx = \text{pour } t \geq 0.$$

*$R(t)$ est la probabilité qu'un dispositif prélevé au hasard dans la population considérée n'ait pas de défaillance avant l'instant $t$.*

**Proposition 1.2.1** • *Comme $F(t)$ est une fonction croissante, la fiabilité $R(t)$ est une fonction décroissante.*

- $R(0) = 1$ *et* $\lim\limits_{t \to +\infty} R(t) = 0$.
- $P(t_1 \leq T \leq t_2) = R(t_1) - R(t_2)$
- $f(t) = -R'(t)$.
- $F(t) + R(t) = 1$.

**Estimation de $F(t)$ :**

**Méthode des rangs bruts :**

$$F(t_i) = \frac{n_i}{n}.$$

**Méthode des rangs moyens :**

$$F(t_i) = \frac{n_i}{n+1}.$$

**Méthode des rangs médians :**

$$F(t_i) = \frac{n_i - 0.3}{n_i + 0.4}.$$

Où,

$n_i$ : rang de la défaillance survenant à l'instant $t$.

**Définition 1.2.3** *Le taux d'avarie moyen dans un intervalle $[t, t+\Delta t]$ est défini par*

$$\frac{R\left(t\right) - R\left(t+\Delta t\right)}{R\left(t\right)} \times \frac{1}{\Delta t}.$$

**Définition 1.2.4 (Taux de défaillance instantané)** *L'écriture mathématique du taux de défaillance à l'instant t, noté $\lambda(t)$, défini sur $\mathbb{R}$ est le suivant :*

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{R(t) - R(t+\Delta t)}{R(t)} \times \frac{1}{\Delta t} = \frac{f(t)}{R(t)},$$

*et satisfait $\lambda(t) \geq 0$, $\forall t$ et*

$$\int_0^{\infty} \lambda(t) dt = \infty.$$

*On alors*

$$R\left(t\right) = \exp\left(-\int_0^t \lambda\left(x\right) dx.\right)$$

### 1.2.4 Indicateurs de fiabilité

**Taux de défaillance**

**Définition 1.2.5** *$\lambda$ représente le taux de défaillance ou le taux d'avarie. Il caractérise la vitesse de variation de la fiabilité au cours du temps. La durée de bon fonctionnement est égale à la durée totale en service moins la durée des défaillances.*

$$\lambda = \frac{\textit{Nombre total de défaillances pendant le service}}{\textit{Durée total de bon fonctionnement}}.$$

**Temps moyen de bon fonctionnement :**

**Définition 1.2.6** *Le MTBF (qui vient de l'anglais Mean Time Before Failure) représente la moyenne des temps de bon fonctionnement entre deux défaillances. En d'autres termes, Il correspond à l'espérance de la durée de vie $t$.*

$$MTBF = \mathbb{E}(T) = \int_0^{+\infty} R(t)\, dt.$$

### 1.2.5 Fiabilité des systèmes

On s'intéresse a un ensemble de $n$ composants ou sous-systèmes, montés en série ou en parallèle.

Soient $T_k$ la durée de vie du composant $k$, $R_k$ sa fiabilité et $\lambda_k$ son taux de panne.

La durée de vie $T$ du système est telle que

$$T > t \iff T_k > t \text{ pour tout } k \in \{1, 2, ..., n\},$$

(1) **En série**

$$
\begin{aligned}
R(t) &= P(T > t) \\
&= P(T_1 > t \cap T_2 > t \cap ... \cap T_n > t) \\
&= \prod_{k=1}^{k=n} R_k(t).
\end{aligned}
$$

Si les $n$ composants sont identiques avec une même fiabilité $R(t)$ la formule sera la suivante

$$R(t) = R(t)^n.$$

Si les taux de défaillances sont constants au cours du temps la fiabilité sera calculée selon la formule suivant :

$$R(t) = \prod_{k=1}^{n} e^{\lambda_k t},$$

avec

$$MTBF = \frac{1}{\lambda_1 + \lambda_2 + ... + \lambda_n},$$

si en plus, les composants sont identiques $\lambda_1 = \lambda_2 = ... = \lambda_n = \lambda$. Alors

$$R(t) = e^{-n\lambda t},\ MTBF = \frac{1}{n\lambda}.$$

Et $T = \min\{T_1, T_2, ...T_n\}$.

(2) **En parallèle**

$$R(t) = P(T > T) = 1 - \prod_{k=1}^{n} (1 - R_k(t)).$$

Si les "$n$" composants sont identiques $R = R_1 = R_2 = ... = R_n$ Alors

$$R(t) = 1 - (1 - R(t))^n.$$

On note $T = \max\{T_1, T_2, ..., T_n\}$.

### 1.2.6 Utilisation de la loi de Weibull pour modéliser les données de fiabilité

La loi de Weibull est la plus couramment utilisée pour la modélisation des données de fiabilité. Cette loi est simple à interpréter et a de nombreuses utilités dans l'analyse de fiabilité.

On retrouve alors, pour tout $t > \gamma$ la fiabilité, la défaillance, son taux et la densité pour la loi de Weibull.

**conclution 1.2.1** • *Fonction de fiabilité :*

$$R(t) = \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right].$$

• *Fonction de défaillance :*

$$F(t) = 1 - \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right].$$

• *Densité de probabilité :*

$$f_T(t \mid \gamma, \eta, \beta) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right].$$

• *Son taux de défaillance instantané :*

$$\lambda(t) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}.$$

*Où $\beta$, $\eta$ et $\gamma$ les paramètres de weibull*

$$\begin{cases} \beta : \text{ paramètre de forme } (\beta > 0). \\ \eta : \text{paramètre d'échelle et indique l'ordre de grandeur de la durée de vie moyenne } (\eta > 0). \\ \gamma : \text{paramètre de position ou paramètre de décalage } (-\infty < \gamma < +\infty). \end{cases}$$

**Relation entre les paramètres de la loi de Weibull et les fonctions de fiabilité**

- Si $\beta < 1$ le taux de défaillance est décroissant, nous avons donc des pannes de jeunesse.
- si $\beta = 1$ le taux de défaillance est constant.
- si $\beta > 1$ le taux de défaillance est croissant, panne de vieillesse ou maturité en mécanique.
- Si $\gamma = 0$ : on étudie la fiabilité dès la première utilisation de la machine.
- Si $\gamma > 0$ : on étudie la fiabilité un certain temps après la première utilisation de l'appareil.

**Modèle du maillon faible**

La loi de Weibull permet également de modéliser la distribution de données de durée de vie dans lesquelles de nombreux processus identiques et indépendants entraînent une défaillance, et où le premier processus qui atteint un stade critique détermine la durée avant défaillance. La théorie des valeurs extrêmes est un élément fondamental du modèle de "maillon faible", où de nombreux défauts peuvent être à l'origine de la défaillance finale. Etant donné que la loi de Weibull peut être dérivée de la loi des plus petites valeurs extrêmes d'un point de vue théorique, elle peut également fournir un modèle efficace pour les cas de type "maillon faible" comme les défaillances en lien avec les condensateurs, les roulements à billes, les relais et la résistance des matériaux. Cependant, si la variable recherchée admet des valeurs négatives, la loi des plus petites valeurs extrêmes est plus adaptée, car la loi de Weibull peut uniquement modéliser des valeurs positives en raison de sa borne inférieure de 0.

# Chapitre 2

# Estimation de l'indice de queue de Weibull

Nous nous intéressons ici à une famille particulière de lois : les lois a queues de type Weibull. Cette famille regroupe notamment les lois de Weibull, normales, gammas, etc... Ces lois sont indexées par un paramètre appelé l'indixe de queue. On présente ici un estimateur de cet indice ainsi que ses propriétés asymptotiques (consistance faible et normalité asymptotique).
L'indice des valeurs extrêmes ne fournit donc aucune information sur la vitesse de décroissance de la fonction de survie à l'intérieur de cette famille de loi. C'est l'indice de queue de Weibull qui nous donne cette information : une valeur de $\theta$ proche de zéro (resp. l'infini) correspond à une décroissance rapide (resp. lente) de la queue de distribution.

## 2.1 Les lois à queues de type Weibull

Les lois a queues de type Weibull sont définies par leur fonction de survie

$$1 - F(x) = \exp\left[-H(x)\right] \text{ avec } H^{\leftarrow}(t) = \inf\left\{x, H(x) \geq t\right\} = t^{\theta} L(t). \tag{2.1}$$

où $\theta > 0$ est appelé l'indixe de queue et $L(x)$ est une fonction à variation lente.

**Remarque 2.1.1** *En particulier, si $L(x)$ est constante, on retrouve les lois de Weibull classiques. Ce type de lois inclut également les lois normales ( $\theta = 1/2$) et les lois Gamma ( $\theta = 1$),pour lesquelles la fonction a variation lente s'écrit sous la forme*

$$L(x) = c_1 + c_2 \left( (\log x) / x \right) + O(1/x),$$

*où $c_1 > 0$ et $c_2$ sont deux constantes.*

## 2.2 Estimation de l'indice de queue

### 2.2.1 Estimateur de type Hill

Cette estimateur consiste d'utiliser le $k_n$ statistique d'ordre supérieur $X_{n-k_n+1} \leq ... \leq X_{n,n}$ où $k_n$ est une séquence d'entiers tels que $1 \leq k_n \leq n$. Notre estimateur appartient à cette famille. Il est défini par :

$$\hat{\theta}_n = \sum_{i=1}^{k_n-1} \left( \log \left( X_{n-i+1,n} \right) - \log \left( X_{n-k_n+1,n} \right) \right) \Big/ \sum_{i=1}^{k_n-1} \left( \log \log \left( n/i \right) - \log \log \left( n/k_n \right) \right), \quad (2.2)$$

**Proposition 2.2.1** *La formule (2.2) est similaire à l'expression de l'estimateur de Hill*

$$\frac{1}{\widehat{\xi}} = \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} \left( \log \left( X_{n-i,n} \right) - \log \left( X_{n-k_n+1,n} \right) \right) \quad (2.3)$$

*de l'indice de valeur extrême $1/\xi$ pour les distributions de type Pareto, i.e. tel que*

$$\frac{\overline{F}(tx)}{\overline{F}(x)} \longrightarrow t^{-\frac{1}{\xi}} \text{ lorsque } x \longrightarrow \infty \text{ pour tous } t > 0.$$

*Bien que les estimateurs (2.2) et (2.3) n'abordent pas le même problème, il semblerait qu'ils partagent des propriétés similaires.*

**Lemme 2.2.1** *La suite* $(k_n)$ *vérifie* $k_n \to \infty$ *et* $k_n/n \to 0$ *lorsque* $n \to \infty$.

*cette hypothèse assure que le nombre de statistiques d'ordre conservées* $k_n$ *est assez grand* $(k_n \to \infty)$ *pour obtenir des estimateurs stables, mais pas trop* $(n/k_n \to \infty)$ *pour que les observations utilisées restent dans la queue de distribution. Le choix de la suite* $(k_n)$ *est donc un compromis entre le biais et la variance de l'estimateur.*

**Théorème 2.2.1** *Si (2.1) vérifié et*

$$k_n \to \infty \text{ et } \frac{k_n}{n} \to 0, \tag{2.4}$$

*on a*

$$\hat{\theta}_n \xrightarrow{P} \theta \text{ lorsque } n \to \infty. \tag{2.5}$$

**Lemme 2.2.2** *Supposer* $k_n \longrightarrow 0$ *et* $k_n/n \longrightarrow 0$. *Alors*

*(i)* $E_{n-i+1,n}/\log(n/i) \xrightarrow{p} 1$, *Uniformément en* $i = 1, ..., k_n$

*(ii)* $k_n\left(E_{n-i+1,n} - \log(n/k_n)\right) \xrightarrow{d} \mathcal{N}(0,1)$.

**Preuve. (Théorème (2.2.1))** Pour des raisons de simplicité, dans cette preuve on notons $k$ pour $k_n$. Nous écrivons aussi $\hat{\theta}_n = \tau_n^{(2)}/\tau_n^{(1)}$ avec

$$\tau_n^{(1)} = \frac{1}{k-1}\sum_{i=1}^{k-}\left(\log\log(n/i) - \log\log(n/k)\right) = \frac{1}{k-1}\sum_{i=1}^{k-1}\log\left(1 + \frac{\log(k/i)}{\log(n/k)}\right).$$

L'inégalité bien connue $-x/2 \leq \log(1+x) - x \leq 0$, $x > 0$ Nous obtenons

$$-\frac{1}{2}\frac{1}{\log(n/k)} - \frac{1}{k-1}\sum_{i=1}^{k-1}\log\log(k/(i)) \leq \log(n/k)\,\tau_n^{(1)} - \frac{1}{k-1}\sum_{i=1}^{k-1}\log(k/i) \leq 0. \tag{2.6}$$

Maintenant quand $k \longrightarrow \infty$,

$$\frac{1}{k-1}\sum_{i=1}^{k-1}\log\log(k/i) \longrightarrow \int_0^1 \log\log(x)\,dx \text{ et } \frac{1}{k-1}\sum_{i=1}^{k-1}\log(k/i) \longrightarrow -\int_0^1 \log(x)\,dx = 1.$$

Alors,

$$\tau_n^{(1)} = \frac{1}{\log(n/k)} + O\left(\frac{1}{\log\log(n/k)}\right). \tag{2.7}$$

Et

$$\tau_n^{(2)} = \frac{1}{k-1}\sum_{i=1}^{k-1}\left(\log\left(X_{n-i+1,n}\right) - \log\left(X_{n-k_n+1,n}\right)\right).$$

Soit $E_{1,n}, ..., E_{n,n}$ une suite de v.a exponentielle. En dessous de (2.1) nous avons

$$\begin{aligned}
\tau_n^{(2)} &\overset{d}{=} \frac{1}{k_n-1}\sum_{i=1}^{k-1}\left(\log H^{\longleftarrow}\left(E_{n-i+1,n}\right) - \log H^{\longleftarrow}\left(E_{n-k+1,n}\right)\right) \\
&\overset{d}{=} \theta\frac{1}{k-1}\sum_{i=1}^{k-1}\log\left(\frac{E_{n-i+1,n}}{E_{n-k+1,n}}\right) + \frac{1}{k-1}\sum_{i=1}^{k-1}\log\left(\frac{L\left(E_{n-i+1,n}\right)}{L\left(E_{n-k+1,n}\right)}\right) \\
&=: \tau_n^{(3)} + \tau_n^{(4)}.
\end{aligned}$$

Pour simplifier $\tau_n^{(3)}$ Qui l'implique

$$\left\{E_{n-i+1,n}\right\}_{i=1,\dots,k-1} = \left\{E_{n-k+1,n} + \left(E_{n-i+1,n} - E_{n-k+1,n}\right)\right\}_{i=1,\dots,k-1} \overset{d}{=} \left\{E_{n-k+1,n} + E_{k-1,n}\right\}_{i=1,\dots,k-1}$$

où $\{E_{1,k-1}, ..., E_{k-1,k-1}\}$ sont des statistiques ordonnées indépendantes de $E_{n-k+1,n}$et généré par $k-1$ variables exponentielles standard indépendantes $\{E_1, ..., E_{k-1}\}$. Donc

$$\tau_n^{(3)} \overset{d}{=} \frac{1}{k-1}\sum_{i=1}^{k-1}\log\left(1 + \frac{E_{k-i,k-1}}{E_{n-k+1,n}}\right) \overset{d}{=} \frac{1}{k-1}\sum_{i=1}^{k-1}\log\left(1 + \frac{E_i}{E_{n-k+1,n}}\right),$$

et, de même que (2.6),

$$-\frac{1}{2}\frac{1}{E_{n-k+1,n}}\frac{1}{k-1}\sum_{i=1}^{k-1}E_i^2 \leq \tau_n^{(3)}E_{n-k+1,n} - \frac{1}{k-1}\sum_{i=1}^{k-1}E_i \leq 0.$$

Maintenant, depuis quand $k \longrightarrow \infty$

$$\frac{1}{k-1}\sum_{i=1}^{k-1}E_i^2 \xrightarrow{p.s} 2 \text{ et } \frac{1}{k-1}\sum_{i=1}^{k-1}E_i \xrightarrow{p.s} 1,$$

par la loi forte des grands nombres et lemme 2.2.2 (i), alors

$$\tau_n^{(3)} \stackrel{d}{=} \frac{1}{E_{n-k+1,n}} \frac{1}{k-1} \sum_{i=1}^{k-1} E_i + O_p\left(\frac{1}{\log^2(n/k)}\right)$$

et, en vue de (2.7)

$$\frac{\tau_n^{(3)}}{\tau_n^{(1)}} \stackrel{d}{=} \frac{\log(n/k)}{E_{n-k+1,n}} \frac{1}{k-1} \sum_{i=1}^{k-1} E_i + O_p\left(\frac{1}{\log^2(n/k)}\right) \xrightarrow{p} 1. \tag{2.8}$$

Avec lemme 2.2.2 (i) et la loi forte des grands nombres finalement $\tau_n^{(4)}$ peut être réécrit en utilisant la représentation Karamata de $L$,

$$\tau_n^{(4)} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log\left(\frac{c(E_{n-i+1,n})}{c(E_{n-k+1,n})}\right) + \frac{1}{k-1} \sum_{i=1}^{k-1} \int_{E_{n-k+1,n}}^{E_{n-i+1,n}} \frac{\varepsilon(t)}{t} dt =: \tau_n^{(5)} + \tau_n^{(6)}.$$

Ou $\tau_n^{(5)}$ converge à 0 en probabilité en vue de Lemme 2.2.2 (i) outre $\tau_n^{(6)}$ peut être contrôlé en introduisant la variable aléatoire

$$\zeta_n = \sup\{|\varepsilon(u)|, u > E_{n-k+1,n}\},$$

et remarquant que

$$\left|\tau_n^{(6)}\right| \leq \zeta_n \frac{1}{k-1} \sum_{i=1}^{k-1} \int_{E_{n-k+1,n}}^{E_{n-i+1,n}} \frac{1}{u} du = \zeta_n \tau_n^{(3)} = o_p\left(\tau_n^{(3)}\right),$$

avec lemme 2.2.2 (i). De (2.7) et (2.8) on obtient $\tau_n^{(4)}/\tau_n^{(1)}$ converge à 0 en probabilité et la conclusion suit avec (2.8)

$$\hat{\theta}_n = \tau_n^{(2)}/\tau_n^{(1)} \xrightarrow{p} \theta.$$

■

**Théor■me 2.2.2** *On suppose que les conditions (1.1) et (2.4) sont satisfaites. Si* $k_n^{1/2} b(\log(n/k_n)) \rightarrow$

0 *et* $k_n^{1/2}/\log(n/k_n) \to 0$ *alors*

$$k_n^{1/2}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \theta^2\right). \tag{2.9}$$

**Preuve.** voir([6]). ■

### Correction de biais

On montrons que l'estimateur $\hat{\theta}_n$ souffre d'un biais systématique ayant deux origines. Pour des lois de Weibull, c'est- a-dire lorsque la fonction $L$ est constante, l'espérance de $\hat{\theta}_n$ peut être calculée explicitement quels que soient $n$ et $k_n$ :

$$\mathbb{E}\left(\hat{\theta}_n\right) = k_n C_n^{k_n} \sum_{i=k_n}^{n} (-1)^{k_n - i} C_{n-k_n}^{i-k_n} \frac{\log(i)}{i-1} \Bigg/ \sum_{i=1}^{k_n-1} \left(\log\log(n/i) - \log\log(n/k_n)\right).$$

### Comparaison avec d'autres estimations

1- Broniatowski un autre estimateur d'indice de queue de Weibull est proposé. Il est défini par

$$\tilde{\theta}_n = \frac{1}{k_n} \sum_{i=1}^{k_n-1} \frac{\log(X_{n-i+1,n})}{\log(n/i)}.$$

Les deux estimateurs $\hat{\theta}_n$ et $\tilde{\theta}_n$ sont basés sur un principe similaire, basé sur le fonction du quantile définie par

$$q(t) = (1-F)^{\leftarrow}(t) = H^{\leftarrow}(\log(1/t)) = (\log(1/t))^{\theta} L(\log(1/t)).$$

2- **La comparaison avec estimateur Beirlant et al :**

Beirlant et al proposent d'utiliser la fonction de vie résiduelle moyenne $e(x) = \mathbb{E}(X - x \mid X > x)$ pour estimer le paramètre $\theta$.

$$\breve{\theta}_n = \frac{\log(n/k_n)}{X_{n-k_n+1,n}} \frac{1}{k_n-1} \sum_{i=1}^{k_n-1} \left\{X_{n-i+1,n} - X_{n-k_n+1,n}\right\}.$$

Et pour plus d'informations, voir [6].

### 2.2.2 Utilisation de poids

Nous nous focalisons sur la famille d'estimateurs Gardes and Girard (2008) de $\theta$ incorporant des poids dans l'estimateur $\hat{\theta}_n$ défini par (2.2). Les estimateurs obtenus sont donc des combinaisons linéaires de statistiques d'ordre c'est-à-dire des L-estimateurs. Plus précisément, on considère la famille d'estimateurs $\Theta_1 = \left\{\hat{\theta}_n(\zeta), \zeta = (\zeta_{1,n}, ..., \zeta_{k_n-1,n})\right\}$ avec

$$\hat{\theta}_n(\zeta) = \sum_{i=1}^{k_n-1} \zeta_{i,n} \left(\log\left(X_{n-i+1,n}\right) - \log\left(X_{n-k_n+1,n}\right)\right) \Big/ \sum_{i=1}^{k_n-1} \zeta_{i,n} \left(\log\log(n/i) - \log\log(n/k_n)\right) \tag{2.10}$$

où $\zeta_{i,n} = W(i/k_n) + \varepsilon_{i,n}$, $(\varepsilon_{i,n})$ $i = 1, ..., k_n - 1$ étant une suite non-aléatoire. La fonction déterministe $W(.)$ doit satisfaire les deux hypothèses de régularité ci dessous :

**(H.1)** la fonction$W(.)$ est définie et admet une dérivée continue sur l'intervalle $]0,1[$.

**(H.2)** Il existe $M > 0$, $0 < q < 1/2$ et $p < 1$ tels que pour tout $x \in ]0,1[$, $|W(x)| \leq Mx^{-q}$ et $\left|W'(x)\right| \leq Mx^{-p-q}$.

Ces conditions sont essentielles pour établir la normalité asymptotique des L-estimateurs. On pose :

$$\|\varepsilon\|_{n,\infty} = \max_{1,..,k_n-1} |\varepsilon_{i,n}|, \; \mu(W) = \int_0^1 W(x)\log(1/x)\,dx$$
$$\sigma^2(W) = \int_0^1 \int_0^1 W(x)\,W(y)\,\frac{\min(x,y) - xy}{xy} dxdy.$$

Nous rappelons à présent le résultat de normalité asymptotique des estimateurs appartenant à cette famille.

**Théor■me 2.2.3** *On se place sous le modèle (2.1) et on suppose que les conditions (2.4) à* ***(H.2)*** *sont satisfaites. Si* $k_n^{1/2} b \log(n/k_n) \to \varphi \in \mathbb{R}$ *et* $k_n^{1/2} \max\left\{1/\log(n), \|\varepsilon\|_{n,\infty}\right\} \to 0$

*lorsque* $n \to \infty$*, alors*

$$k_n^{1/2}\left(\hat{\theta}_n(\zeta) - \theta - b(\log(n/k_n))\right) \xrightarrow{d} \mathcal{N}\left(0, \theta^2\sigma^2(W)/\mu^2(W)\right).$$

Dans le cas où $\liminf \|\varepsilon\|_{n,\infty} \log(n) \leq 1$ avec $\varphi \neq 0$ le théorème (2.2.3) n'est valable que si $\rho > -1$ ce qui correspond à une vitesse de convergence lente dans la condition (1.1). Donnons à présent deux choix possibles pour les poids $\zeta_{i,n}$, $i = 1, ..., k_n - 1$.
En prenant $\zeta_{i,n} = 1$ pour tout $i = 1, ..., k_n - 1$( i.e $W(x) = 1$ pour tout $x \in ]0, 1[$ et $\varepsilon_{i,n} = 0$ pour tout $i = 1, ..., k_n - 1$), on retrouve l'estimateur $\hat{\theta}_n$(voir l'équation (2.2) proposé par Beirlant et al (1996).

### 2.2.3 Estimateur Zipf

En rappelant que les points $(\log\log(n/i), \log(X_{n-i+1,n}))$, $i = 1, ..., k_n - 1$ sont approximativement répartis sur une droite de pente $\theta$, il est possible d'estimer $\theta$ par l'estimateur des moindres carrés Il est montré dans (Gardes and Girard, 2008) que l'estimateur des moindres carrés appartient à la famille $\Theta_1$ avec les poids :

$$\zeta_{i,n} = \zeta_{i,n}^z = \log\log(n/i) - \frac{1}{k_n - 1}\sum_{i=1}^{k_n-1} \log\log(n/i) = W((i/k_n) + \varepsilon_{i,n}.$$

où $W(x) = -(\log(x) + 1)$ et uniformément en $i = 1, ..., k_n - 1$, $\varepsilon_{i,n} = O\left(\log^2(k_n)/\log(n)\right) + O(\log(k_n)/k_n)$ L'estimateur $\hat{\theta}_n(\zeta^z)$ ainsi obtenu est similaire à l'estimateur de l'indice des valeurs extrêmes de Zipf. Nous rappelons ci-dessous le résultat sur la normalité asymptotique de $\hat{\theta}_n(\zeta^z)$. C'est une conséquence directe du théorème (2.2.3).

**Corollaire 2.2.1** *On se place sous le modèle (2.1) et on suppose que les conditions (1.1) et (2.4) sont satisfaites. Si* $k_n^{1/2} b(\log(n/k_n)) \to 0$ *et* $k_n^{1/2} \log^2(k_n)/\log(n) \to 0$ *alors*

$$k_n^{1/2}\left(\hat{\theta}_n(\zeta^z) - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, 2\theta^2\right).$$

### 2.2.4 Un estimateur débiaisé

Considérons à présent un estimateur débiaisé de l'indice de queue de Weibull $\theta$. On définit les variables aléatoires

$$Z_j = j \log(n/j) \left(\log(X_{n-j+1,n}) - \log(X_{n-j,n})\right),\ j = 1, ..., k_n.$$

Le modèle suivant peut être établi

$$Z_j = \left(\theta + \left(\frac{\log(n/k_n)}{\log(n/j)}\right) b(\log(n/k_n))\right) f_j + o_p(b(\log(n/k_n))),\ j = 1, ..., k_n, \quad (2.11)$$

où $f_j,\ j = 1, ..., k_n$ sont des variables aléatoires indépendantes de loi exponentielle de paramètre 1 et le terme $o_p(b(\log(n/k_n)))$ ne dépend pas de $j$. On obtient à partir du modèle (2.11) l'approximation

$$Z_j \approx \theta + b(\log(n/k_n)) x_j + \eta_j,\ j = 1, ..., k_n, \quad (2.12)$$

où $\eta_j$ est un terme d'erreur aléatoire centré et $x_j = \log(n/k_n) / \log(n/j)$. En estimant les paramètres $\theta$ et $b(\log(n/k_n))$ du modèle de régression linéaire (2.12) par la méthode des moindres carrés ordinaires, on obtient un estimateur de $\theta$ débiaisé :

$$\hat{\theta}_n^D = \frac{1}{k_n} \sum_{j=1}^{k_n} Z_j - \frac{\hat{b}(\log(n/k_n))}{k_n} \sum_{j=1}^{k_n} x_j,$$

où

$$\hat{b}(\log(n/k_n)) = \sum_{j=1}^{k_n} \left(x_j - \frac{1}{k_n} \sum_{j=1}^{k_n} x_j\right) Z_j \Bigg/ \sum_{j=1}^{k_n} \left(x_j - \frac{1}{k_n} \sum_{j=1}^{k_n} x_j\right)^2$$

La normalité asymptotique de $\hat{\theta}_n^D$ est donnée par le résultat ci-dessous.

**Théor■me 2.2.4** *On se place sous le modèle (2.1) et on suppose que les conditions (1.1) et (2.4) sont satisfaites. Supposons de plus que la fonction $b(.)$ est telle que $x\,|b(x)| \to \infty$*

*lorsque* $x \to \infty$ *avec* $\left(k_n^{1/2}/\log(n/k_n)\right) b(\log(n/k_n)) \to \tilde{\varphi} \in \mathbb{R}$*Supposons enfin que, si* $\tilde{\varphi} = 0$, $\log^2(k_n)/\log(n/k_n) \to 0$ *et* $k_n^{1/2}/\log(n/k_n) \to \infty$*. On a :*

$$\frac{k_n^{1/2}}{\log(n/k_n)}\left(\hat{\theta}_n^D - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \theta^2\right).$$

### 2.2.5 Choix du nombre $k_n$ de statistiques d'ordre

En ne tenant pas compte du terme de biais dans le modèle de régression (2.11), on obtient un estimateur non débiaisé de $\theta$ défini par :

$$\frac{1}{k_n}\sum_{j=1}^{k_n} Z_j.$$

L'erreur moyenne quadratique asymptotique (AMSE) de cet estimateur est donnée par :

$$ASME(k_n) = \frac{\theta^2}{k_n} + \left(\frac{b(\log(n/k_n))}{k_n}\sum_{j=1}^{k_n}\frac{\log(n/k_n)}{\log(n/j)}\right)^2.$$

Un choix possible pour kn est alors de prendre $k_n^{opt} = \arg\min_{k_n} ASME(k_n)$. Nous pouvons estimer cette erreur par la quantité $\widehat{ASME}(k_n)$ obtenue en remplaçant $\theta$ et $b(\log(n/k_n))$ par les estimateurs $\hat{\theta}_n^D$ et $\hat{b}(\log(n/k_n))$ définis précédemment. Le nombre $k_n^{opt}$ est estimé par :

$$\hat{k}_n = \arg\min_{k_n}\widehat{ASME}(k_n).$$

# Conclusion

Après avoir dans le premier chapitre rappelé certaines des notions clés en théorie des valeurs extrêmes, et la fiabilité de machine avec utilisation de la loi de Weibull pour modéliser les données de fiabilité, nous nous sommes intéressés dans le deuxième chapitre au techniques pour l'estimation de l'indice de queue de type Weibull. Il y a d'autres estimations dont nous n'avons pas parlé.

Pour conclure nous signalons que notre thèse n'est qu'un point de départ pour mieux connaitre ce monde immense des valeurs extrêmes. Et d'examiner la fiabilité de tenter d'augmenter la vie ou d'estimer le moment de l'échec pour éviter ou minimiser les pertes.

# Bibliographie


[1] De Haan, L., & Ferreira, A. (2007). Extreme value theory : an introduction. Springer Science & Business Media.

[2] Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1989). Regular variation (Vol. 27). Cambridge university press.

[3] Resnick, S. I. (2013). Extreme values, regular variation and point processes. Springer.

[4] Gardes, L., & Girard, S. (2013). Estimation de quantiles extrêmes pour les lois à queue de type Weibull : une synthèse bibliographique. Journal de la Société Française de Statistique, 154(2), 98-118.

[5] Girard, S. (2004). A Hill type estimator of the Weibull tail-coefficient. Communications in Statistics-Theory and Methods, 33(2), 205-234.

[6] Girard, S. (2004). A Hill type estimator of the Weibull tail-coefficient. Communications in Statistics-Theory and Methods, 33(2), 205-234.

[7] Thomas, M. (2012). Fiabilité, maintenance prédictive et vibration des machines. PUQ.

[8] Benameur, S. (2010). Sur l'estimation de l'indice des valeurs extremes. Universite Mohamed Khider Biskra.

[9] La loi de Weibull dans l'analuse de fiabilité. https ://support.minitab.com/fr-fr/minitab/18/help-and.../weibull-distribution/.

# Annexe A : Abréviations et Notations

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

| | | |
|---|---|---|
| v.a | : | Variable alétoire. |
| i.i.d | : | Indépendantes et identiquement distribuées. |
| GEV | : | Generalized Extreme Value distribution. |
| EVI | : | Extreme Value Indixe. |
| $X$ | : | Population. |
| $(X_1, X_2..., X_n)$ | : | Échantillon de taille $n$ de v.a $X$. |
| $(X_{(1)}, X_{(2)}..., X_{(n)})$ | : | Statistiques d'ordre associées à $(X_1, X_2..., X_n)$. |
| $\mathcal{E}(\lambda)$ | : | Loi exponentielle de paramètre $\lambda$. |
| $F$ | : | Fonction de répartition. |
| $f$ | : | Densité de probabilité. |
| $\overline{F}$ | : | Fonction de survie. |
| $F_n$ | : | Fonction de répartition empirique. |
| $F^{\leftarrow}$ | : | Fonction des quantilles. |
| $F_n^{\leftarrow}$ | : | Fonction des quantilles empiriques. |
| $\overline{F},\ R$ | : | Fonction de survie ou Fonction de fiabilité. |

| | | |
|---|---|---|
| $L$ | : | Fonction à variationlente. |
| $\mathcal{RV}_\alpha$ | : | Fonction à variations régulières d'indice $\alpha$. |
| $\lambda$ | : | Taux de défaillance (ou d'avarie). |
| $AMSE$ | : | Erreur moyenne quadratique asymptotique. |
| $MTBF$ | : | Temps moyen de bon fonctionnement. |
| $\lambda$ | : | Taux de défaillance (ou d'avarie). |
| $Var\,(X)$ | : | Variance mathématique de $X$. |
| $\bar{X}_n$ | : | Moyenne empirique. |
| $\mathbf{1}_A$ | : | Fonction indicatrice de l'ensemble $A$. |
| $\xrightarrow{\mathcal{P}}$ | : | convergence en probabilité. |
| $\xrightarrow{d}$ | : | convarge en loi. |
| $\xrightarrow{p.s}$ | : | convarge en presque sure. |