

Théorie des valeurs extrêmes  
N. Kazi-Tani  
M2 PSA  
Semestre S9

# Table des matières

<b>1</b>	<b>Comportement asymptotiques des sommes</b>	<b>4</b>
<b>2</b>	<b>Comportement asymptotique du maximum</b>	<b>6</b>
<b>3</b>	<b>Distribution asymptotiques des excès</b>	<b>12</b>
<b>4</b>	<b>Applications statistiques</b>	<b>13</b>
I	Estimation de $\xi$ . . . . .	13

# Introduction

L'un des buts de la théorie des valeurs extrêmes est d'extrapoler une probabilité de survie à partir des données collectées (par exemple les sinistres et leurs coûts pour les assurances/réassurances).

On peut par exemple s'intéresser à une suite de variables aléatoires indépendantes et identiquement distribuées  $(X_k)_{k \geq 1}$  et se poser la question de la loi suivie par la suite  $(M_n = \max\{X_k, 1 \leq k \leq n\})_{n \geq 1}$  pour  $n$  assez grand.

On peut par exemple se poser les questions suivantes :

1. Quelle est la vitesse minimale d'un 100 mètres ?
2. Quelle est l'estimation des sinistres extrêmes en assurance ou en réassurance ?
3. Quels sont les mouvements extrêmes des prix des actifs sur un marché financier ?
4. À quelle hauteur doit-on placer des digues pour contrer des inondations décennale, centennale... ?
5. Quel estimation des temps de services d'un réseau télécom ?

Toutes ces questions peuvent se traiter comme l'étude de queues de distributions, ce qui fait intervenir la théorie des valeurs extrêmes.

# Bibliographie

[1] Sidney I. Resnick. Extreme Values, Regular Variation and Point Processes. Springer New York, 1987.

# Chapitre 1

## Comportement asymptotiques des sommes

L'idée de ce chapitre est de donner une version généralisée du théorème de la limite centrée. On va d'abord introduire la notion de stabilité pour la somme puis la notion de domaine d'attraction pour enfin terminer par le théorème de la limite centrée généralisé.

Commençons donc par introduire la notion de stabilité pour une somme.

**Définition 1.1** (stabilité pour la somme). Une variable aléatoire  $X$  (resp. une mesure de probabilité) est dite *stable pour la somme* si pour tout couple  $(c_1, c_2) \in \mathbb{R}^2$  et toutes copies  $X_1, X_2$  iid de  $X$ , il existe  $a(c_1, c_2) > 0$  et  $b(c_1, c_2) \in \mathbb{R}$ , tels que

$$P_{c_1 X_1 + c_2 X_2} = P_{a(c_1, c_2) X + b(c_1, c_2)}.$$

En particulier, si  $X$  est stable pour la somme et si  $X_1, \dots, X_n$  est une suite iid de même loi que  $X$ , alors en posant  $S_n = \sum_{1 \leq k \leq n} X_k$  il existe  $a_n > 0$  et  $b_n \in \mathbb{R}$  tels que

$$P_{S_n} = P_{a_n X + b_n}$$

et par renormalisation

$$P_{\frac{S_n - b_n}{a_n}} = P_X.$$

**Remarque 1.2.**

1. Une loi stable pour la somme est une limite possible pour  $S_n$  correctement renormalisée;
2. Toutes les lois ne vérifient pas bien entendu cette propriété (par exemple la loi de Poisson).

**Proposition 1.3.** Les lois stables pour la somme sont les seules limites possible en cas de convergence de  $S_n$  (renormalisée).

On obtient le théorème de la limite centrée généralisée.

**Théorème 1.4.** Soit  $(X_k)_{k \geq 1}$  une suite de variables iid. S'il existe  $a_n > 0$  et  $b_n \in \mathbb{R}$  tels que

$$\frac{S_n - b_n}{a_n} \xrightarrow{\mathcal{L}} Z$$

où  $Z$  est non dégénérée, alors  $Z$  est une loi stable pour la somme et on a pour tout  $t \in \mathbb{R}$ ,

$$\Phi_Z(t) = E\left(e^{i\langle t, Z \rangle}\right) = \exp\left(it\gamma - c|t|^\alpha (1 + i\beta \operatorname{sgn}(t)z(\alpha, t))\right),$$

où :

1.  $z(\alpha, t) = \tan\left(\frac{\pi\alpha}{1}\right) \mathbf{1}_{\alpha=1} - \frac{2}{\pi} \log|t| \mathbf{1}_{\alpha=1}$ ;
2.  $\gamma \in \mathbb{R}$  (ce qui correspond à la moyenne);
3.  $c > 0$  (correspondant à l'oscillation);
4.  $\alpha \in ]0, 2[$  (gère la queue de distribution ainsi que le coefficient de normalisation);
5.  $\beta \in [-1, 1]$  (correspondant à la symétrie de la loi);

6.  $\text{sgn}$  est la fonction signe ie  $\text{sgn}(t) = \mathbf{1}_{t>0}(t) - \mathbf{1}_{t<0}(t)$ .

7.  $a_n \propto n^{\frac{1}{\alpha}}$ .

**Remarque 1.5.**

1. Le cas  $\alpha = 2$  correspond à  $\Phi_Z(t) = \exp(it\gamma - \frac{1}{2}t^2(2c))$  et donc  $P_Z = \mathcal{N}(\gamma, 2c)$ ;
2. Le cas  $\alpha = 1$  correspond à la loi de Cauchy;
3. On dit que ces lois sont  $\alpha$ -stables;
4. Si  $Z$  est  $\alpha$ -stable et  $E(Z^2) < +\infty$  ssi  $\alpha = 2$ .

On introduit maintenant la notion de domaine d'attraction.

**Définition 1.6** (domaine d'attraction pour les lois  $\alpha$ -stables). On dit qu'une variable aléatoire  $X$  de loi  $F$  (ici  $F$  désigne sa fonction de répartition) *appartient au domaine d'attraction* d'une distribution  $\alpha$ -stable  $G_\alpha$  (encore une fois une fonction de répartition) s'il existe deux suites  $(a_n)_{n \geq 1} \in (\mathbb{R}^+)^{\mathbb{N}}$  et  $(b_n)_{n \geq 1} \in \mathbb{R}^{\mathbb{N}}$  telles que

$$\frac{X_1 + X_2 + \dots + X_n - b_n}{a_n} \xrightarrow{\mathcal{L}} G_\alpha,$$

où  $X_1, \dots, X_n$  iid de loi  $X$ .

Dans ces conditions, on écrira  $F \in \text{DA}_\Sigma(G_\alpha)$  (ou encore  $X \in \text{DA}_\Sigma(G_\alpha)$ ).

**Remarque 1.7.** Toute variable aléatoire  $X$  avec  $E(X^2) < \infty$  vérifie  $X \in \text{DA}_\Sigma(G_2)$  (c'est le théorème usuel de la limite centrée).

On introduit la notion de fonction à variation lente.

**Définition 1.8.** Une fonction  $L : \mathbb{R} \rightarrow \mathbb{R}$  est à *variation lente* sur un voisinage de  $+\infty$  si

$$\lim_{x \rightarrow \infty} \frac{L(ax)}{L(x)} = 1, \forall a > 0.$$

**Exemple 1.9.** La fonction  $\log$  est à variation lente.

On peut alors caractériser les domaines d'attraction pour la somme.

**Proposition 1.10.**

1.  $F \in \text{DA}_\Sigma(G_2)$  ssi

$$\lim_{x \rightarrow \infty} \frac{x^2 \int_{[-x, x]^c} dF(y)}{\int_{[-x, x]^c} y^2 dF(y)} = 0.$$

2.  $F \in \text{DA}_\Sigma(G_\alpha)$ ,  $0 < \alpha < 2$ , ssi

$$F(-x) = \frac{c_1 + o(1)}{x^\alpha} L(x), \text{ quand } x \rightarrow \infty$$

et

$$\bar{F}(x) = \frac{c_2 + o(1)}{x^\alpha} L(x)$$

où  $\bar{F} = 1 - F$ ,  $L$  est à variation lente et  $c_1, c_2 \geq 0, c_1 + c_2 > 0$ .

## Chapitre 2

# Comportement asymptotique du maximum

Dans toute la suite  $(X_k)_{k \geq 1}$  désigne une suite de va iid de loi  $G$  et on définit

$$M_n = \max(X_1, \dots, X_n), n \geq 1.$$

Comme dans le chapitre précédent, on se pose les deux questions suivantes :

1. S'il existe deux suites  $(a_n), (b_n)$  telles que  $\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H$ , comment peut-on caractériser  $H$  ?
2. Comment calculer  $a_n$  et  $b_n$  ?

On notera par  $x_F$  le réel définit par

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

On a la première proposition concernant la suite  $(M_n)$ .

**Proposition 2.1.** *Il y a convergence presque sûre de la suite  $(M_n)_{n \geq 1}$  vers le réel  $x_F$ .*

*Démonstration.* On sait que pour tout  $x \in \mathbb{R}$ ,

$$P(M_n \leq x) = F^n(x)$$

et ce par indépendance, donc

$$\lim_{n \rightarrow +\infty} P(M_n \leq x) = \begin{cases} 0 & \text{si } F(x) < 1 \\ 1 & \text{si } F(x) = 1 \end{cases}$$

Ainsi  $M_n \xrightarrow{\mathcal{L}} G$  ou  $G(x) = \mathbf{1}_{x \geq x_F}(x)$ . Par conséquent  $M_n \xrightarrow{\mathcal{L}} x_F$  donc  $M_n \xrightarrow{P} x_F$  et puisque la suite  $(M_n)$  est presque sûrement croissante (ie.  $M_n \leq M_{n+1}$  ps) alors  $M_n \xrightarrow{p.s.} x_F$ .  $\square$

On cherche  $H$  tel que  $\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H$  et donc en tout point de continuité de  $H$  on a

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n}\right) = H(x)$$

ce qui est équivalent à

$$\lim_{n \rightarrow \infty} P(M_n \leq a_n x + b_n) = H(x).$$

On peut alors se poser la question de la convergence de  $P(M_n \leq a_n x + b_n)$  vers  $K \in [0, 1]$  et plus particulièrement quand  $0 < K < 1$ . Pour ce faire on a le lemme d'approximation de Poisson.

**Lemme 2.2** (lemme d'approximation de Poisson). *Soit  $\tau \in [0, \infty]$  et soit  $(u_n)_{n \geq 1}$  une suite de réels. Alors il y a équivalence entre :*

1.  $\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau$ ;
2.  $\lim_{n \rightarrow \infty} P(M_n \leq u_n) = e^{-\tau}$ .

*Démonstration.* La preuve se fait en scindant les cas entre  $\tau < \infty$  et  $\tau = \infty$ . Supposons que  $\tau < \infty$  pour commencer.

On a

$$\begin{aligned} P(M_n \leq u_n) &= F(u_n)^n \\ &= (1 - \bar{F}(u_n))^n \\ &= \left(1 - \frac{n\bar{F}(u_n)}{n}\right) \rightarrow e^{-\tau} \end{aligned}$$

Réciproquement, on commence par montrer que  $\bar{F}(u_n) \rightarrow 0$ . Supposons qu'il existe une sous-suite  $(u_{n_k})$  de  $(u_n)$  telle que  $(F(u_{n_k}))$  ne converge pas vers 0. Alors,

$$P(M_{n_k} \leq u_{n_k}) = \left(1 - \bar{F}(u_{n_k})\right)^{n_k} \rightarrow 0$$

ce qui est une contradiction avec 2). Ainsi  $\bar{F}(u_n) \rightarrow 0$ .

De plus en prenant le logarithme dans 2), on a que  $n \log F(u_n) = n \log(1 - \bar{F}(u_n)) \rightarrow -\tau$  et en utilisant que  $\log(1 - x) \sim_0 -x$ , on a le résultat.

Le cas  $\tau = \infty$  est laissé en exercice. □

**Proposition 2.3** (CTT, Convergence to Type Theorem). Soient  $Z, \tilde{Z}, (Y_n)_{n \geq 1}$  des va non dégénérées et soient  $(a_n), (\tilde{a}_n), (b_n), (\tilde{b}_n)$  des suites telles que

$$\frac{Y_n - b_n}{a_n} \xrightarrow{\mathcal{L}} Z \text{ et } \frac{Y_n - \tilde{b}_n}{\tilde{a}_n} \xrightarrow{\mathcal{L}} \tilde{Z}.$$

Alors il existe  $a > 0$  et  $b \in \mathbb{R}$  tels que

$$\lim_{n \rightarrow \infty} \frac{a_n}{\tilde{a}_n} = a \text{ et } \lim_{n \rightarrow \infty} \tilde{b}_n - b_n = b$$

et

$$P_Z = P_{a\tilde{Z}+b}$$

et donc  $Z$  et  $\tilde{Z}$  sont de même type.

De la même manière que l'on a défini la loi stables pour la somme, on peut définir les lois stables pour le maximum.

**Définition 2.4** (lois max-stables). Une va  $X$  est dite *max-stable* si pour tout  $n \geq 1$ , il existe  $a_n > 0$  et  $b_n \in \mathbb{R}$  tels que

$$P_{\max(X_1, \dots, X_n)} = P_{a_n X + b_n}$$

où  $X_i$  désigne une copie iid de  $X$ .

**Théorème 2.5** (Fischer-Tippett-Gnedenko). Soit  $(X_n)_{n \geq 1}$  une suite de va iid réelles. S'il existe une suite  $(a_n)_{n \geq 1}, a_n > 0$  et une suite  $(b_n)$  telles que

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H \tag{2.1}$$

où  $H$  est la fonction de répartition d'une distribution non dégénérée, alors  $H$  est du type (à transformation linéaire près) d'une des distributions suivantes :

1. *Fréchet* :  $\Phi_\alpha(x) = \exp(-x^{-\alpha})\mathbf{1}_{(x \geq 0)}(x)$ ,  $\alpha > 0$ ;
2. *Weibull* :  $\Psi_\alpha(x) = \exp(-(-x)^\alpha)\mathbf{1}_{(x < 0)}(x) + \mathbf{1}_{(x \geq 0)}(x)$ ,  $\alpha \geq 0$ ;
3. *Gumbel* :  $\Lambda(x) = \exp(-\exp(-x))$ ,  $x \in \mathbb{R}$ .

*Démonstration.* Soit  $t > 0$ , alors d'après l'équation 2.1,

$$P(M_{\lfloor nt \rfloor} \leq a_{\lfloor nt \rfloor} X + b_{\lfloor nt \rfloor}) = F(a_{\lfloor nt \rfloor} X + b_{\lfloor nt \rfloor}) \xrightarrow{n \rightarrow \infty} H(x)$$

et

$$F(a_n X + b_n)^{\lfloor nt \rfloor} = (F(a_n X + b_n))^{\frac{\lfloor nt \rfloor}{n}} \xrightarrow{n \rightarrow \infty} H(x)^t.$$

Ainsi d'après la Proposition 2.3 il existe  $\alpha(t) > 0$  et  $\beta(t) \in \mathbb{R}$  (tels que  $t \mapsto \alpha(t)$  et  $t \mapsto \beta(t)$  mesurables comme limite simples de fonctions mesurables) telles que

$$H(x)^t = H(\alpha(t)x + \beta(t)).$$

Par conséquent pour  $s > 0$ , on a

$$\begin{aligned} H(x)^{st} &= H(\alpha(st)x + \beta(st)) \\ &= (H(x)^s)^t \\ &= H(\alpha(t)(\alpha(s)x + \beta(s)) + \beta(t)) \end{aligned}$$

et par identification, on a

$$\begin{cases} \alpha(st) = \alpha(s)\alpha(t) \\ \beta(st) = \alpha(t)\beta(s) + \beta(t) \end{cases}, \forall t, s > 0.$$

Cette équation de Hamel n'a comme solution mesurable que les fonctions du type

$$t \mapsto \alpha(t) = e^{-\theta}, t \in \mathbb{R}, \theta \in \mathbb{R}$$

On distingue les cas  $\theta < 0$ ,  $\theta = 0$  et  $\theta > 0$ .

Si  $\theta = 0$ , alors  $\alpha \equiv 1$  et  $\beta(st) = \beta(t)\beta(s)$ , par suite il existe  $c \in \mathbb{R}$  tel que  $\beta(t) = -c \log(t)$ .

On a obtenu alors,

$$H(x)^t = H(x - c \log t)$$

avec  $c \neq 0$  car  $H$  est non dégénérée.

Supposons que  $H(x_0) = 1$  pour un certain  $x_0 \in \mathbb{R}$ , alors on a

$$H(x_0 - c \log(t)) = 1, \forall t > 0$$

ce qui n'est pas possible car  $H$  est non dégénérée. Par suite  $H(x) < 1, \forall x \in \mathbb{R}$ .

En prenant  $x = 0$  dans l'équation fonctionnelle de  $H$ , on obtient

$$H(0)^t = H(-c \log t)$$

et par changement de variable  $u = -c \log t$ , en prenant  $p \in \mathbb{R}$  défini par  $e^{-e^p} = H(0)$ , obtient au final

$$H(u) = \exp(-e^{-\frac{u}{c}-p}) = \Lambda\left(\frac{u}{c} - p\right).$$

Par suite la loi de  $H$  est de même type que la loi de Gumbell. □

**Remarque 2.6.** Les lois  $\Phi_\alpha$ ,  $\Psi_\alpha$  et  $\Lambda$  sont appelées *lois des valeurs extrêmes standards*.

**Remarque 2.7.** Si  $(X_k)_{k \geq 1}$  est une suite de va iid de loi de Fréchet, alors

$$\begin{aligned} P(M_n \leq n^{\frac{1}{\alpha}}) &= \Phi_\alpha(n^{-\frac{1}{\alpha}}x)^n \\ &= \exp(-(n^{-\frac{1}{\alpha}}x)^{-\alpha}) = \Phi_\alpha(x). \end{aligned}$$

Donc la loi de Fréchet est bien max-stable.

On introduit maintenant comme dans le chapitre précédent, la notion de domaine d'attraction, mais cette fois pour le maximum.

**Définition 2.8.** Une loi de probabilité  $F$  appartient au domaine d'attraction du maximum d'une loi de valeurs extrêmes  $H$ , que l'on note  $F \in \text{MDA}(H)$ , s'il existe  $a_n > 0$  et  $b_n \in \mathbb{R}$  tels que

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H.$$

**Remarque 2.9.** Il est possible que l'on ne puisse pas trouver de normalisation telle que la loi  $H$  ne soit pas dégénérée.

On donne une proposition sur le comportement de la fonction de survie.



**Proposition 2.10** (Leadbetter, Lindgren, Rootzén). *Il existe une suite  $(u_n)_n$  tel que  $n\bar{F}(u_n) \xrightarrow{n \rightarrow \infty} \tau$  si et seulement si  $\lim_{x \rightarrow x_F} \frac{\bar{F}(x)}{\bar{F}(x^-)} = 1$ .*

**Remarque 2.11.** La loi de Poisson ne vérifie pas cette propriété.

On introduit maintenant une loi de probabilité qui permet de généraliser le triptyque Fréchet, Weibull, Gumbel.

**Définition 2.12** (loi des valeurs extrêmes généralisées). On définit la *loi des valeurs extrêmes généralisée* comme

$$H_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}}) & , \text{ si } \gamma \neq 0 \\ e^{-e^{-x}} & , \text{ si } \gamma = 0 \end{cases}$$

si  $1 + \gamma x > 0$ .

**Remarque 2.13.** Le support de cette loi change en fonction de  $\gamma$ .

On va pouvoir donner des équivalences de domaines d'attraction du maximum.

**Proposition 2.14.** *Soit  $F$  une fonction de répartition d'une mesure de probabilité.*

1.  $F \in \text{MDA}(\Phi_\alpha) \iff F \in \text{MDA}(H_\gamma)$  avec  $\alpha = \frac{1}{\gamma} > 0$ ;
2.  $F \in \text{MDA}(\Lambda) \iff F \in \text{MDA}(H_0)$ ;
3.  $F \in \text{MDA}(\Psi_\alpha) \iff F \in \text{MDA}(H_\gamma)$  avec  $\alpha = -\frac{1}{\gamma} > 0$ ;

*Démonstration.* Exercice assez simple. □

On introduit la notion de fonction à variation régulière afin de pouvoir déterminer dans la suite de chapitre les coefficients de normalisation.

**Définition 2.15** (fonction à variation régulière). On dit qu'une fonction  $h : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$  mesurable est à *variation régulière en  $+\infty$  d'indice  $\rho \in \mathbb{R}$*  si pour tout  $x > 0$ ,

$$\frac{h(tx)}{h(x)} \xrightarrow{t \rightarrow \infty} x^\rho.$$

On parle de variation lente quand  $\rho = 0$ . On note usuellement  $h \in \text{RV}_\rho$ .

On continue avec la notion d'inverse généralisé d'une fonction (qui sert pour les quantiles).

**Définition 2.16** (inverse généralisé). Soit  $h$  une fonction croissante. On note et définit

$$\bar{h}_-(u) = h^{-1}(u) = \inf \{x \in \mathbb{R}; h(x) \geq u\}$$

l'*inverse généralisée continue à gauche* de  $h$ . De la même manière on peut définir l'*inverse généralisée à droite* que l'on note  $\bar{h}_+$ .

**Remarque 2.17.** Si  $F$  est une fonction de répartition  $\bar{F}_-$  est la fonction quantile associée à  $F$ .

En utilisant ces fonctions quantiles, on va pouvoir déterminer les coefficients de normalisation pour les lois max-stables.

**Proposition 2.18.**  $F \in \text{MDA}(\Phi_\alpha)$  si et seulement si  $\bar{F} = 1 - F \in \text{RV}_{-\alpha}$ .

Dans ce cas

$$\frac{1}{a_n} M_n \xrightarrow{\mathcal{L}} \Phi_\alpha$$

avec  $a_n = F^{-1}(1 - \frac{1}{n})$ .

**Remarque 2.19.** La loi de Pareto est définie par

$$\bar{F}(x) = \begin{cases} \frac{1}{x^\alpha} & , x > 1 \\ 1 & , x \leq 1 \end{cases} \in \text{RV}_{-\alpha}.$$

En particulier  $F(x) = 1 - \frac{1}{x^\alpha}$  et  $q_F(u) = \frac{1}{(1-u)^{\frac{1}{\alpha}}}$  et on a

$$a_n = n^{-\frac{1}{\alpha}}.$$

*Démonstration.* Pour la condition nécessaire voir [1, pp 55-57] .  
On remarque que si

$$\bar{F}(a_n) = 1 - F(F^{-1}(1 - \frac{1}{n})) \sim_{n \rightarrow \infty} \frac{1}{n}.$$

Pour  $x > 0$ ,

$$\begin{aligned} n\bar{F}(a_n x) &= n \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} \bar{F}(a_n) \\ &\sim_{n \rightarrow \infty} \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} \\ &\longrightarrow_{n \rightarrow \infty} x^{-\alpha}. \end{aligned}$$

D'après le lemme d'approximation de Poisson on a alors

$$P(M_n \leq a_n x) \xrightarrow[n \rightarrow \infty]{} \exp(-(x)^{-\alpha}) = \Phi_\alpha(x).$$

Pour  $x < 0$ , on a

$$P(M_n \leq a_n x) = F(a_n x)^n$$

or,  $F(a_n x) \leq F(0)$  et  $F(0) \neq 1$  (car sinon  $\bar{F} \notin \text{RV}_{-\alpha}$ ), ainsi

$$F(a_n x) < 1$$

et donc

$$\lim_{n \rightarrow \infty} P(M_n \leq a_n x) = 0.$$

□

**Remarque 2.20.** On a  $\Psi_\alpha(-\frac{1}{x}) = \Phi_\alpha(x)$ ,  $x > 0$ .

**Proposition 2.21.**  $F \in \text{MDA}(\Psi_\alpha)$  si et seulement si  $x_F < +\infty$  et  $\bar{F}(x_F - \frac{1}{\bullet}) \in \text{RV}_{-\alpha}$ . Dans ce cas,

$$\frac{M_n - x_F}{a_n} \xrightarrow{\mathcal{L}} \Psi_\alpha$$

avec

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

*Démonstration.* On prouve que  $\bar{F}(x_F - \frac{1}{x}) \in \text{RV}_{-\alpha}$  implique  $F \in \text{MDA}(\Psi_\alpha)$ . On définit

$$F_*(x) = F(x_F - \frac{1}{x}) \mathbf{1}_{(x>0)}(x)$$

et d'après la proposition précédente  $F_* \in \text{MDA}(\Phi_\alpha)$  avec  $a_n^* = F_*^{-1}(1 - \frac{1}{n})$ .

On a pour tout  $x \in \mathbb{R}_+^*$ ,

$$\begin{aligned} F_*(a_n^* x) &\xrightarrow[n \rightarrow \infty]{} \Phi_\alpha(x) \\ F(x_F - \frac{1}{a_n^* x})^n &\xrightarrow[n \rightarrow \infty]{} \Phi_\alpha(x) \\ F(x_F + \frac{1}{a_n^*} y)^n &\xrightarrow[n \rightarrow \infty]{} \Phi_\alpha(-\frac{1}{y}) = \Psi_\alpha(y) \quad (y = -\frac{1}{x}) \end{aligned}$$

Par conséquent,

$$P(a_n^*(M_n - x_F) \leq y) \xrightarrow[n \rightarrow \infty]{} \Psi_\alpha(y), \quad \forall y < 0.$$

De plus,

$$\begin{aligned} a_n^* &= \inf\{u \in \mathbb{R}; F_*(u) \geq 1 - \frac{1}{n}\} \\ &= \inf\{u \in \mathbb{R}; F(x_F - \frac{1}{u}) \geq 1 - \frac{1}{n}\} = (\sup\{y \in \mathbb{R}; F(x_F - y) \geq 1 - \frac{1}{n}\})^{-1} \\ &= (x_F - \inf\{z; F(z) \geq 1 - \frac{1}{n}\})^{-1} \\ &= (x_F - F^{-1}(1 - \frac{1}{n}))^{-1} \\ &= \frac{1}{a_n}. \end{aligned}$$

□

**Exemple 2.22.** Si  $F$  est la loi uniforme sur  $[0, 1]$ , on a  $x_F = 1$  et on a

$$\frac{M_n - 1}{\frac{1}{n}} \xrightarrow{\mathcal{L}} \Psi_1.$$

## Chapitre 3

# Distribution asymptotiques des excès

**Définition 3.1** (loi de Pareto généralisée). La fonction de répartition

$$G_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-\frac{1}{\gamma}} & , \text{ si } \gamma \neq 0 \\ 1 - e^{-x} & , \text{ si } \gamma = 0 \end{cases}$$

pour  $x \in D(\gamma)$ , avec

$$D(\gamma) = \begin{cases} 0 \leq x < +\infty & , \text{ si } \gamma \geq 0 \\ 0 \leq x \leq -\frac{1}{\gamma} & , \text{ si } \gamma < 0 \end{cases}$$

est appelée *loi de Pareto généralisée*.

**Notation** (Fonction de répartition des excès (Excess distribution function)). Soit  $X$  une variable aléatoire de loi  $F$ . On note :

1.  $F_u(x) = P(X - u \leq x / X > u), x \geq 0$ ;
2.  $G_{\gamma, \nu, \beta}(x) = G_\gamma(\frac{x - \nu}{\beta}), x \in D(\gamma, \nu, \beta)$ ;
3.  $G_{\gamma, 0, \beta}(x) = G_{\alpha, \beta}(x) = 1 - (1 + \frac{\gamma}{\beta}x)^{-\frac{1}{\gamma}}, x \in D(\gamma, \beta)$ .

**Théorème 3.2** (Gnedenko). *Les assertions suivantes sont équivalentes :*

1.  $F \in \text{MDA}(H_\gamma)$ ;
2. *Il existe une fonction  $u \mapsto \beta(u)$  mesurable strictement positive telle que*

$$\lim_{u \rightarrow X_F} \sup_{x \in ]0, X_F - u[} |F_u(x) - G_{\gamma, \beta(u)}(x)| = 0.$$

**Remarque 3.3.** La loi de Pareto généralisée est la seule loi stable par troncature.

# Chapitre 4

## Applications statistiques

**Définition 4.1** (indice des valeurs extrêmes). On appelle l'indice  $\xi$  dans la loi généralisée des valeurs extrêmes l'indice des valeurs extrêmes.

### I- Estimation de $\xi$

**Théorème 4.2.** Soit  $(X_1, \dots, X_n)$  une suite de va iid de loi  $F \in \text{MDA}(H_\xi)$ . Si  $k : \mathbb{N} \rightarrow \mathbb{N}$  est une application croissante (avec  $\lim_{n \rightarrow \infty} k(n) = +\infty$  évidemment) et si  $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = 0$ , alors l'estimateur de Pickands défini par

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

converge en probabilité vers  $\xi$ .

#### Remarque 4.3.

1. L'estimateur de Pickands est biaisé;
2. Si on suppose  $\lim_{n \rightarrow \infty} \frac{k(n)}{\log \log(n)} = +\infty$ , alors  $\widehat{\xi}_{k(n)} \xrightarrow{p.s.} \xi$ .

**Proposition 4.4.** Soit  $p \in ]0, 1[$ . On suppose que  $F$  est continue et admet une seule solution à l'équation  $F(x_p) = p$ . Soit  $k(n)$  une suite d'entiers telle que  $1 \leq k(n) \leq n$  avec  $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = p$  (par exemple  $k(n) = \lfloor np \rfloor + 1$ ), alors

$$X_{k(n)} \xrightarrow{P} x_p \quad \text{et} \quad X_{k(n)} \xrightarrow{p.s.} x_p$$

où  $x_p$  est le quantile théorique d'indice  $p$ .

**Remarque 4.5.** C'est une version analogue à la loi forte/faible des grands nombres mais cette fois pour les quantiles.

*Démonstration.* Soit  $S_n(x)$  le nombre de franchissement de  $x$  par les  $X_i$ , c-à-d

$$S_n(x) = \sum_{1 \leq i \leq n} \mathbf{1}_{(X_i \leq x)}.$$

On remarque que

$$(X_{k(n)} \leq x) = (S_n(x) \geq k(n)).$$

De plus,

$$\frac{S_n(x)}{k(n)} = \frac{S_n(x)}{n} \frac{n}{k(n)} \xrightarrow{n \rightarrow \infty} \frac{F(x)}{p} \text{ ps}$$

et donc la convergence est encore vrai en probabilité.

Ainsi,

$$\begin{aligned} P(X_{k(n)} \leq x) &= P\left(\frac{S_n(x)}{k(n)} \geq 1\right) \\ &\xrightarrow{n \rightarrow \infty} \begin{cases} 0 & , \text{ si } F(x) < p \\ 1 & , \text{ si } F(x) \geq p \end{cases} \\ &\xrightarrow{n \rightarrow \infty} \begin{cases} 0 & , \text{ si } x < x_p \\ 1 & , \text{ si } x \geq x_p \end{cases} \quad (\text{unicité } F(x_p) = p) \end{aligned}$$

Par conséquent,

$$X_{k(n)} \xrightarrow{\mathcal{L}} x_p$$

et puisque la limite  $x_p$  est constante alors

$$X_{k(n)} \xrightarrow{P} x_p.$$

□