

## Chapitre 3

# Modélisation des extrêmes hydrologiques

L'objectif de ce chapitre est de modéliser les extrêmes hydrologiques. De nombreux résultats et outils portant sur la modélisation des extrêmes existent dans la littérature probabiliste et statistique. Cependant, les séries hydrologiques sont courtes et les estimateurs proposés dans la littérature statistique sont, pour la plupart, entachés d'erreurs d'estimation rédhibitoires compte tenu des exigences pratiques hydrologiques. Ce chapitre propose donc une nouvelle méthode de modélisation des extrêmes, comparable en terme d'erreur à celle du maximum de vraisemblance, mais présentant l'avantage d'être semi-paramétrique (dans le sens où l'on ne fait d'hypothèses que sur l'appartenance à une classe de modèles et non pas sur un modèle particulier). Ces deux méthodes, appliquées à de nombreuses séries pluviométriques, permettent de conclure que les queues de distribution des séries de pluie ont un comportement algébrique et que le paramètre statistique caractérisant la décroissance algébrique des queues de distribution est un invariant spatial et temporel.

### 3.1 Introduction

L'abondance croissante d'informations sur la magnitude de nombreux événements naturels extrêmes récents et sur leurs conséquences dramatiques porte à croire que la fréquence de ceux-ci a augmenté brusquement. En fait, il se peut que seules leur relation et leur description soient devenues plus complètes. L'accroissement de la vulnérabilité des sites, du fait de leur utilisation de plus en plus intense, conduit aussi le plus grand nombre à penser que l'intensité des phénomènes naturels est en augmentation. La prévention contre les phénomènes extrêmes, au moins dans le domaine de l'hydrologie, est encore fondée sur des modèles de prédiction indiscutés. La non conformité d'événements extrêmes aux prédictions de ces modèles standard pousse des personnes même averties à croire que les conditions naturelles de génération de ces phénomènes exceptionnels ont subi des modifications. Ainsi l'accroissement de l'information sur les catastrophes naturelles, une plus grande vulnérabilité à ces catastrophes, les écarts entre les observations et les prédictions des modèles utilisés font croire à un dérèglement de la nature.

Pour l'hydrologie opérationnelle il s'agit de proposer un outil de description et de

prédiction, dégagé d'hypothèses trop arbitraires (souvent adoptées pour le confort qu'elles procurent), et rendant plausibles des intensités de phénomènes, intensités qui, dans le cadre des modèles actuels, auraient des probabilités infinitésimales.

La théorie des valeurs extrêmes fournit un cadre probabiliste visant à construire un modèle asymptotique pour les variables aléatoires extrêmes. Plus précisément, elle établit les conditions sous lesquelles les extrêmes d'un échantillon aléatoire  $(X_1, \dots, X_n)$  convergent vers une distribution limite non dégénérée quand la taille de l'échantillon  $n$  tend vers l'infini. Les lois rencontrées dans la littérature hydrologique sont le plus souvent de type exponentiel (loi exponentielle, normale, Weibull ...). La loi normale a connu un développement considérable dans les applications hydrologiques, car elle a la propriété d'être invariante par addition. Elle assure ainsi à l'hydrologue une invariance dans la nature de la répartition des erreurs; mieux encore, elle garantit, grâce au Théorème de la Limite Centrale, que la distribution de la somme d'erreurs de distributions quelconques converge vers la loi Normale :

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{n \rightarrow \infty} N(0, 1)$$

ce qui n'est vérifié que si la queue de la distribution  $F$  des variables aléatoires  $(X_1, \dots, X_n)$  n'est pas trop lourde. Cette condition peut ne pas être vérifiée lorsque l'analyse statistique porte sur les  $k^{\text{ième}}$  extrêmes de l'échantillon  $(X_1, \dots, X_n)$  (au sens des  $k$  plus grandes observations), et l'on peut se demander plus généralement s'il existe des constantes  $a_n$  et  $b_n$  telles que le rapport :

$$\frac{X_1 + \dots + X_n - b_n}{a_n}$$

tende en loi vers une distribution non dégénérée.

Dans ce chapitre, nous présentons dans ses grandes lignes la théorie des valeurs extrêmes en *section 3.3*, en détaillant le domaine d'attraction de la loi de Fréchet (l'ouvrage de Galambos [36] est la référence en ce domaine). Puis nous présentons en *section 3.4* l'application de cette théorie probabiliste en statistique. Nous passons en revue les principaux outils statistiques (en nous référant à l'ouvrage de Embrechts [30]) avant de développer un outil graphique original de caractérisation et d'estimation de l'indice de décroissance algébrique. Nous comparons, sur des données simulées, la méthode proposée avec des méthodes paramétriques (maximum de vraisemblance ou moments) et des méthodes semi-paramétriques. En *section 3.5*, on estime le paramètre de décroissance algébrique de quelques longues séries pluviométriques à différents pas de temps (à pas journalier en une station de Dédougou (Burkina-Faso), à pas horaire sur une série d'Alabama, et à pas égal à la minute pour une station de l'Ile de la Réunion), ainsi qu'en divers sites (232 longues séries annuelles à travers le monde). Les résultats obtenus sont ensuite confrontés aux estimations découlant du traitement brut de séries de basculements d'auge en *section 3.6*, en adaptant au préalable les estimateurs à ces séries.

## 3.2 Notations

Dans tout ce chapitre, on adoptera les notations et définitions suivantes :

**Notation 10** *Considérant une suite de  $N$  vap (variables aléatoires positives)  $(X_i)_{i=1\dots N}$  indépendantes et identiquement distribuées (iid) de **densité**  $f$ , on notera :*

$$(X_i)_{i=1\dots N} \stackrel{iid}{\sim} f$$

et  $M_N$  le maximum de ce  $N$ -échantillon

$$M_N = \max_{i=1\dots N} (X_i)$$

**Hypothèse :** Cette densité  $f$ , à support dans  $\mathbb{R}_+$ , sera supposée ici intégrable contre  $t^2$  et de primitive intégrable :  $\int_1^{+\infty} t^2 f(t) dt < \infty$  et  $\int_1^{+\infty} \int_s^{+\infty} f(t) . ds . dt < \infty$

**Définition 11** *Sous cette hypothèse, on définit sur  $\mathbb{R}_+^*$  les fonctions suivantes :*

$$\text{fonction de survie : } G(s) = \int_0^{+\infty} f(t) . 1_{t>s} dt = \int_s^{+\infty} f(t) dt$$

$$\text{fonction de répartition : } F(s) = \int_0^{+\infty} f(t) . 1_{t\leq s} dt = \int_0^s f(t) dt$$

$$\text{cumul de dépassement } H : H(s) = \int_0^{+\infty} t . f(t) . 1_{t>s} dt = sG(s) + \int_s^{+\infty} G(t) dt$$

**Notation 12** *La notation  $[x]$  désignera la partie entière de  $x$ .*

**Notation 13** *La notation  $1_E$  désignera la fonction indicatrice de  $E$  :*

$$1_E(x) = \begin{cases} 1 & \text{si } x \in E \\ 0 & \text{si } x \notin E \end{cases}$$

### 3.3 Rappels sur la théorie des valeurs extrêmes

La modélisation des extrêmes en hydrologie constitue un problème complexe car plus les données sont irrégulières, plus l'estimation de leurs maxima est entachée d'erreurs. On sait que la qualité des outils statistiques classiques d'estimation et de prévision se dégrade quand la variabilité et/ou la dépendance au sein de la série étudiée augmente. Il est donc nécessaire de renforcer le socle théorique pour réduire et quantifier cette erreur. A partir d'une série d'observations  $(X_1, \dots, X_N)$ , l'hydrologue cherche à quantifier, de façon relative, le "caractère extrême" de l'une d'entre elles. Par exemple, la cote atteinte par la crue de 1910 à Paris dépasse de 2 mètres celle de 1876 ; dans quelle mesure peut-on dire que cette crue est exceptionnelle ? On est donc amené à s'intéresser à la distribution des données, ou tout au moins à quelques caractéristiques d'intérêt de cette distribution.

Selon Galambos [36], la théorie des valeurs extrêmes semble de prime abord contradictoire : en effet elle vise à "comprendre la régularité du comportement extrême". Mais sous certaines hypothèses, elle fournit un cadre de modélisation simple que nous présentons dans cette section.

### 3.3.1 Loi des extrêmes

On rappelle dans cette section les principaux résultats de la théorie des extrêmes (Galambos [36]). On cherche à savoir s'il est possible de caractériser, à une normalisation près, les comportements des extrêmes d'un échantillon d'une loi quelconque. On s'intéresse à la distribution asymptotique du maximum  $M_N$  de  $N$  *vap*  $(X_i)_{i=1\dots N}$  indépendantes et identiquement distribuées, que l'on notera de façon générique  $X$ . Soit  $F$  la fonction de répartition de  $X$ . On sait que la fonction de répartition  $F_N$  de  $M_N$  s'écrit, pour tout  $x > 0$  :

$$F_N(x) = F^N(x)$$

ce qui indique que, lorsque  $N$  tend vers l'infini, le maximum  $M_N$  tend vers la borne à droite  $x_F$  du support de la fonction de répartition  $F$  ( $x_F$  est éventuellement infini)<sup>1</sup> :

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

Dans la théorie des valeurs extrêmes, on cherche à trouver, pour le maximum  $M_N$  convenablement normalisé, quelles sont les lois limites possibles quand  $N$  tend vers l'infini.

**Définition 14** *Les lois de Fréchet, Weibull et Gumbel sont respectivement définies par les fonctions de répartition (voir figure 3.1) :*

$$\begin{aligned} \phi_\alpha(x) &= \begin{cases} 0 & \text{si } x \leq 0 \\ \exp[-x^{-\alpha}] & \text{si } x > 0 \end{cases} \text{ avec } \alpha > 0 \\ \psi_\alpha(x) &= \begin{cases} \exp[-(-x)^{-\alpha}] & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \text{ avec } \alpha > 0 \\ \Lambda(x) &= \exp[-e^{-x}] \text{ pour } x \in \mathbb{R} \end{aligned}$$

**Définition 15** *On dit qu'une distribution  $f$  appartient au domaine d'attraction maximal d'une loi extrême  $H$  s'il existe des suites de réels  $c_N$  et  $d_N$  telles que*

$$\frac{M_N - d_N}{c_N} \xrightarrow[N \rightarrow \infty]{L} H$$

avec  $M_N = \max_{1 \leq i \leq N} X_i$  et  $X_i \stackrel{iid}{\sim} f$

**Théorème 16** (de Fisher-Tippet). *Lois limites pour les maxima : Soit  $(X_N)_N$  un échantillon de *vap* de loi  $f$ . S'il existe des suites de réels  $c_N$  et  $d_N$  et une loi extrême  $H$  telle que  $f$  appartienne au domaine d'attraction maximal de  $H$ , alors  $H$  est l'une des trois distributions Fréchet, Weibull ou Gumbel.*

---

<sup>1</sup>En effet,

$$\forall x < x_F, \quad F_N(x) \xrightarrow[N \rightarrow \infty]{} 0$$

et dans le cas où  $x_F < \infty$ ,

$$\forall x \geq x_F, \quad F_N(x) = 1$$

donc  $M_N$  tends vers  $x_F$  presque sûrement, avec  $x_F \leq \infty$ .

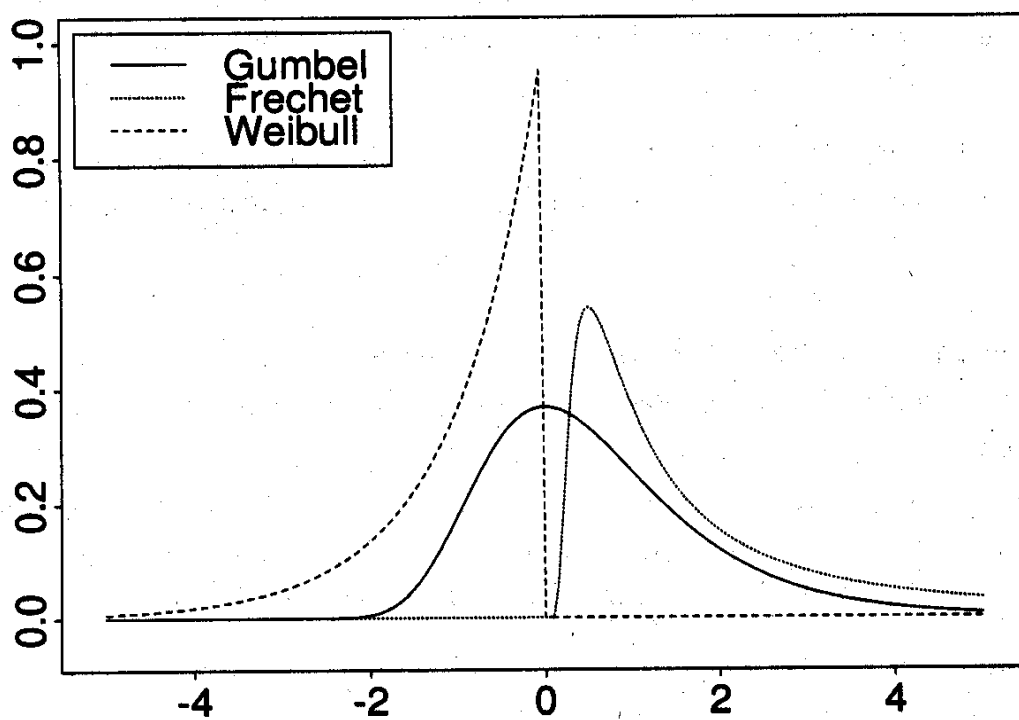


Figure 3.1: Densité de probabilité des lois de Gumbel, Fréchet et Weibull (Embrechts et al. [38]).

Mais il est possible de synthétiser ces trois lois extrêmes en une seule famille :

**Définition 17** (*Représentation de Jenkinson et Von Mises*). On définit la loi généralisée des valeurs extrêmes (loi *GVE*) par :

$$H_\xi(x) = \begin{cases} \exp \left[ - (1 + \xi x)^{-1/\xi} \right] & \text{si } \xi > 0 \\ \exp [-\exp(-x)] & \text{si } \xi = 0 \end{cases} \quad \text{avec } 1 + \xi x > 0$$

On montre que les cas  $\xi > 0$ ,  $\xi = 0$  et  $\xi < 0$  correspondent respectivement aux distributions Fréchet  $\phi_{\frac{1}{\xi}}$ , Gumbel  $\Lambda$  ou Weibull  $\psi_{\frac{-1}{\xi}}$ .

La loi *GVE* fournit donc un moyen efficace d'unification des trois lois extrêmes. Son introduction a été motivée par des considérations pratiques, car on peut estimer le paramètre  $\xi$  (par le maximum de vraisemblance par exemple), et faire ainsi de l'inférence paramétrique dans un cadre non paramétrique.

### 3.3.2 Le domaine d'attraction de la loi de Fréchet

Réciproquement, étant donnée une loi  $H$  appartenant à l'un des trois types Fréchet, Gumbel ou Weibull, quelles conditions imposer à la loi  $F$  pour que le maximum  $M_N$  converge vers  $H$  ? Différentes caractérisations des trois domaines d'attraction (Fréchet, Gumbel, Weibull) sont proposées (voir [36]) mais dans ce travail, on ne détaillera que le domaine d'attraction de la loi de Fréchet (en raison de son potentiel d'applications dans le domaine de l'hydrologie stochastique). Dans cette section, on commence par définir quelques familles de lois avant de présenter quelques conditions suffisantes d'appartenance au domaine d'attraction de Fréchet.

#### 1- Fonction moyenne des dépassements et lois algébriques

**Définition 18** Soit  $X$  une vap telle que  $EX < \infty$ . La fonction moyenne des dépassements (*FMD*) est définie par

$$\begin{aligned} e : [0, \infty[ &\longrightarrow \mathbb{R} \\ s &\longmapsto e(s) = E(X - s | X > s) \end{aligned}$$

Cette fonction quantifie en quelque sorte la vitesse de décroissance de la queue de la distribution et donc l'irrégularité de la variable aléatoire.

On peut utiliser aussi la fonction moyenne des dépassements relatifs  $h$  (*FMDR*) définie pour tout  $s > 0$  par :

$$h(s) = \frac{e(s)}{s}$$

Rappelons deux propositions qui établissent une relation bi-univoque entre la *FMD*  $e$  et la fonction de survie  $G$ , garantissant ainsi une équivalence entre l'étude de ces deux fonctions.

**Proposition 19** Pour tout  $s \geq 0$ ,  $e(s) = \frac{H(s)}{G(s)} - s$  <sup>(2)</sup>

---

<sup>2</sup>En effet, pour toute vap  $Y$  telle que  $E(Y) < \infty$ ,  $E(Y) = \int_0^\infty G_Y(y) dy$

**Proposition 20**  $\forall s \geq 0 \quad G(s) = e(0) \cdot \exp \left\{ - \int_0^s \frac{du}{e(u)} \right\} \quad (3)$

Le tableau suivant fournit les *FMD* de quelques distributions standards.

Pour une loi exponentielle  $Exp(\lambda)$ , la *FMD* est une fonction constante (égale à  $\frac{1}{\lambda}$ ). Dès qu'une distribution attribue plus de poids que la loi exponentielle aux observations au delà d'un certain seuil  $s_0$ , sa *FMD* est croissante sur le domaine  $[s_0, +\infty[$ . Elle est d'autant plus croissante que ce poids est lourd. Les lois de type Pareto ont typiquement une *FMD* qui converge vers 0.

**Définition 21** On dit qu'une loi est de type algébrique de coefficient  $q > 1$  quand sa fonction de survie au delà de 1 s'écrit :

$$\forall s \geq 1 \quad G(s) = r(s) \cdot s^{-q}$$

avec  $r$  vérifiant :

$$\begin{cases} (C1) \ r(s) \xrightarrow{s \rightarrow \infty} K \text{ et } r \text{ croissante sur } [1, +\infty[ \\ (C2) \ \forall s \geq 1 \quad r'(s) \leq As^{-\theta} \text{ avec } A > 0 \text{ et } \theta > 1 \end{cases}$$

La *FMDR* d'une loi de type algébrique (voir *section 3.3.2*) est décroissante sur  $[1, +\infty[$ <sup>4</sup>.

De plus, pour une loi de type algébrique de paramètre  $q$ , la *FMDR*  $h$  tend vers  $\frac{1}{q-1}$  à l'infini. On a même équivalence sous certaines conditions (pour la démonstration, on pourra se reporter à l'*Annexe B*) :

**Proposition 22** Une distribution est de type algébrique si sa *FMDR* vérifie les deux conditions suivantes :

$$\begin{cases} (C3) \ \forall s \geq 1 \quad 0 \leq q - 1 - \frac{1}{h(s)} \leq Cs^{-\nu} \text{ avec } C > 0 \text{ et } \nu > 0 \\ (C4) \ \forall s \geq 1 \quad 0 \leq -h'(s) \leq Ds^{-\theta'} \text{ avec } D > 0 \text{ et } \theta' > 1 \end{cases}$$

---

Comme :  $\forall x > 0 \quad G_{X-a|X>a}(x) = \frac{P(X > x+a)}{P(X > a)} = \frac{G(x+a)}{G(a)}$

$\forall a \geq 0 \quad e(a) = \int_0^{\infty} G_{X-a|X>a}(x) dx$  ce qui nous conduit au résultat.

<sup>3</sup>En effet :  $\forall a \geq 0 \quad \int_0^a \frac{du}{e(u)} = \int_0^a \frac{G(u)}{\int_u^{\infty} G(v) dv} du = \left[ -\log \left( \int_u^{\infty} G(v) dv \right) \right]_0^a$

Donc  $\forall a \geq 0 \quad \int_0^a \frac{du}{e(u)} = \log e(0) - \log [e(a)G(a)]$

<sup>4</sup>En effet :  $\forall s' \geq s \geq 1 \quad h(s) - h(s') = \int_s^{s'} \frac{G(t)}{s^{1-q} \cdot r(s)} dt \geq \int_s^{s'} \frac{G(t)}{K} dt \geq 0$

Nom	Fonction de survie $G(x)$	FMD $e(s)$	$\lim_{s \rightarrow \infty} h(s)$
Uniforme	$(1-x) \cdot 1_{0 < x < 1} + 1_{x < 0}$	$\frac{1}{2}(1-s)$	0
Benktander II $\begin{cases} \alpha > 0 \\ 0 < \beta < 1 \end{cases}$	$x^{-(1-\beta)} \cdot \exp\left(-\frac{\alpha}{\beta}x^\beta\right)$	$\frac{1}{\alpha}s^{1-\beta}$	0
Weibull $\begin{cases} \lambda > 0 \\ 0 < \tau < 1 \end{cases}$	$\exp(-\lambda x^\tau)$	$\frac{s^{1-\tau}}{\lambda\tau} [1 + o(1)]$	0
Exponentielle $\lambda > 0$	$\exp(-\lambda x)$	$\frac{1}{\lambda}$	0
Gamma $m, \lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty u^{m-1} \exp(-\lambda u) du$	$\frac{1}{\lambda} + \frac{m-1}{\alpha\lambda^2} [1 + o(1)]$	0
Logistique	$\frac{2}{1+\exp(x)}$	$(1+e^s) \log(1+e^{-s})$	$\infty$
lognormale $\begin{cases} \mu \in R \\ \sigma > 0 \end{cases}$	$\int_x^\infty \frac{1}{\sqrt{2\pi}\sigma u} \exp\left(-\frac{1}{2\sigma^2[\log u - \mu]^2}\right) du$	$\frac{s}{\log s} [1 + o(1)]$	0
Benktander I $\alpha, \beta > 0$	$\gamma x^{-(1+\alpha+\beta \log x)} \cdot (\alpha + 2\beta \log x)$	$\frac{s}{\alpha+2\beta \log s}$	0
Pareto $\alpha > 1$	$x^{-\alpha}$	$\frac{s}{\alpha-1}$	$\frac{1}{\alpha-1}$
Burr $\beta, \lambda, \tau > 0$	$\left(\frac{\beta}{\beta+x^\tau}\right)^\lambda$	$\frac{(s^\tau+\beta)^{\frac{1}{\tau}}}{\lambda\tau-1} [1 + o(1)]$	$\frac{1}{\lambda\tau-1}$
Pareto gén. $\alpha, \beta > 0$	$\left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}}$	$\frac{\xi s}{1-\xi} [1 + o(1)]$	$\frac{\xi}{1-\xi}$
Loggamma $m, \lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty (\log u)^{m-1} \cdot u^{-\lambda-1} du$	$\frac{s}{\lambda-1} [1 + o(1)]$	$\frac{1}{\lambda-1}$
Log-logistique $\alpha > 1$	$1/(1+\beta x^\alpha)$	$\frac{s}{\alpha-1} [1 + o(1)]$	$\frac{1}{\alpha-1}$
Fréchet $\alpha > 1$	$1 - \exp(-x^{-\alpha})$	$\frac{s}{\alpha-1} [1 + o(1)]$	$\frac{1}{\alpha-1}$

Tableau 3.1: FMD pour quelques lois.

## 2- Propriétés

On montre (Galambos, 78 [36] et David, 81 [24] p.260) que le domaine d'attraction de la loi de Fréchet est constitué entre autres :

- des lois  $f$  satisfaisant à la condition dite de Von Mises :

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{G(x)} = \alpha > 0$$

(où  $G$  désigne la fonction de survie).

- des lois dites de type Pareto, c'est-à-dire dont la fonction de survie décroît algébriquement au voisinage de l'infini<sup>5</sup> :

$$G(x) = l(x) \cdot x^{-\alpha} \text{ avec } l \text{ à variations lentes ie : } \frac{l(\lambda y)}{l(y)} \xrightarrow{y \rightarrow \infty} 1 \quad \forall \lambda > 0$$

Le schéma 3.2 détaille quelques conditions suffisantes d'appartenance au domaine d'attraction de la loi de Fréchet.

## 3- Exemples de lois appartenant au domaine de Fréchet

Les exemples les plus simples de lois appartenant au domaine de Fréchet sont les lois de Pareto stricte et généralisée.

**Définition 23** On dit que  $X$  suit une loi de Pareto stricte de paramètre  $\alpha$   $Pa(\alpha)$  si et seulement si sa fonction de survie s'écrit :

$$\forall x \geq 1 \quad G(x) = x^{-\alpha}$$

On vérifie facilement que cette loi satisfait la condition de Von Mises. Elle est simple d'utilisation car elle se ramène à la loi exponentielle par transformation logarithmique<sup>6</sup>.

Le support de la loi de Pareto étant trop particulier ( $[1, +\infty[$ ), on doit introduire des paramètres supplémentaires pour modifier la forme de cette distribution. La prise en compte d'un paramètre de localisation  $\nu$  peut se faire en considérant les données translatées  $(X_i - \nu)_{1 \leq i \leq N}$ .

**Définition 24** On introduit un paramètre d'aplatissement  $\beta$  dans la loi définie précédemment pour définir la Pareto généralisée à deux paramètres  $PG(\xi, \beta)$  :

$$G_{\xi, \beta}(x) = \begin{cases} \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{si } \xi = 0 \end{cases} \text{ avec } \begin{cases} x \geq 0 & \text{si } \xi \geq 0 \\ 0 \leq x \leq -\beta/\xi & \text{si } \xi < 0 \end{cases}$$

<sup>5</sup>Un développement de Taylor au voisinage de l'infini de la fonction de survie de Fréchet permet d'avoir une visualisation rapide de cette décroissance algébrique :

$$1 - \Phi_{\alpha}(x) = 1 - \exp(-x^{-\alpha}) \stackrel{x \rightarrow \infty}{\sim} x^{-\alpha}$$

<sup>6</sup>Si  $X \sim Pa(\alpha)$  alors  $Y = \log X \sim Exp(\alpha)$ . En effet :  $\forall y \geq 0 \quad P(Y > y) = P(X > e^y) = e^{-\alpha y}$  car  $e^y \geq 1$