

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

**RAJAH SARA**

Titre :

# Estimation Bayésienne de Mesures de Risque Extrêmes : Théorie et Applications

Membres du Comité d'Examen :

Pr. **MERAGHNI Djamel** UMKB Président

Dr. **BERKANE Hassiba** UMKB Encadreur

Dr. **SOLTANE Louiza** UMKB Examineur

Juin 2021

## *Dédicace*

*Pour m'avoir donné la courge, la volonté, la patience et la sagesse pour accomplir ce travail, et surtout pour sa bien vaillance, car sans lui je n'aurai rien pu faire et grâce a lui je suis devenu celle que je suis aujourd'hui et j'ai eu tout ce que j'ai, alors je dois avant tout dire*

*<< الحمد و الشكر لله >> merci ALLAH.*

*Mon père, en signe d'amour, de reconnaissance et de gratitude pour tous les soutiens et les sacrifices dont il a fait preuve à mon égard.*

*À la fleur de ma vie, Ma mère, ma raison d'être et de vivre, la lanterne qui éclaire mon chemin.*

*À mes frères: Ferhat, Yazid, Hicham*

*Et spécialement mon cher frère Mourad pour son suivi continu tout au long des années d'étude*

*Ames sœurs: Imane et Rahma*

*Toute la famille: Rajah*

*Mon encadreur: Berkane Hassiba*

*Et à tous ceux qui m'ont Soutenu et encour*

## *Remerciements*

*Je remercie le DIEU le tout puissant d'avoir convenus bonté infinie, courage, la force et la patience de terminer ce modeste travail.*

*Je remercie également mes parents, mes frères et sœurs et pour la soute moral et matériel.*

*Après, je voudrais faire une note profonde pour remercier le directeur du message Dr. Hassiba Berkane, à l'Université Mohamed Khidr. Biskra, pour la confiance qu'elle m'a témoignée lors des travaux, pour son soutien et ses conseils, m'a accompagnée au quotidien dans la préparation de ce travail avec le plus grand sérieux, en allouant le temps nécessaire aux discussions. Sa persévérance au quotidien a été un capital important pour accomplir ce travail d'un point de vue scientifique et éthique.*

*Nous voudrions remercier également Pr. MERAGHNI Djamel et Dr. SOLTANE Louiza, membres de jury, de nous avoir fait l'honneur d'accepter de jurer ce travail.*

*Je tiens vivement à exprimer mes remerciements à mes enseignants département de Mathématiques qui m'ont enseigné.*

*Je remercie tout particulièrement le Dr Hisham Ali et le professeur Bedjaoui. Dr, à l'Université Mohamed Khidr à département d'hydrologie Ils me soutiennent avec les données de mon mémoire*

*Je tiens à remercier tous mes chères amies Surtout chahla pour tous les fous rires qu'on a partagés. Je tiens enfin à remercier tous ceux qui nous ont aidées de près ou de loin pour réaliser.*

# Table des matières

<b>Table des matières</b>	<b>2</b>
<b>Liste des figures</b>	<b>4</b>
<b>Liste des tableaux</b>	<b>5</b>
<b>Introduction</b>	<b>1</b>
<b>1 Théorie des Valeurs Extrêmes, Mesures de Risque</b>	<b>3</b>
<b>1.1 Théorie des Valeurs Extrêmes</b> . . . . .	3
<b>1.1.1 Statistique d'ordre</b> . . . . .	3
<b>1.1.2 Estimation des Valeurs Extrêmes</b> . . . . .	10
<b>1.2 Mesures de risque</b> . . . . .	12
<b>1.2.1 Région de risque</b> . . . . .	13
<b>1.2.2 Quelques mesures de risque</b> . . . . .	13
<b>1.2.3 Estimation par méthode du maximum de vraisemblance</b> . . . . .	16
<b>2 Approche bayésienne sur les paramètres des lois des valeurs extrêmes</b>	<b>18</b>
<b>2.1 Définitions</b> . . . . .	18
<b>2.2 Principe de la statistique bayésienne</b> . . . . .	20
<b>2.3 Distribution a priori</b> . . . . .	20
<b>2.4 Distribution a posteriori</b> . . . . .	23

<b>2.5 Estimation ponctuelle</b> . . . . .	23
<b>2.5.1 Fonction de perte</b> . . . . .	23
<b>2.5.2 Risque</b> . . . . .	25
<b>2.6 Estimateur bayésien</b> . . . . .	25
<b>2.7 Méthodes usuelles de calcul Bayésien</b> . . . . .	27
<b>2.7.1 Les méthodes de Monte Carlo</b> . . . . .	28
<b>2.7.2 Méthodes de Monte Carlo par chaînes de Markov (MCMC)</b> . . . . .	28
<b>2.8 Inférence bayésienne sur les valeurs extrêmes</b> . . . . .	30
<b>2.8.1 Inférence bayésienne sur les paramètres d'un modèle GEV</b> . . . . .	31
<b>2.8.2 Inférence bayésienne sur les paramètres d'un modèle POT</b> . . . . .	33
<b>3 Simulation et résultats numériques</b>	<b>37</b>
<b>3.1 Application Numérique</b> . . . . .	37
<b>3.1.1 L'approche GEV</b> . . . . .	37
<b>3.1.2 Mesures de risque</b> . . . . .	43
<b>Conclusion</b>	<b>46</b>
<b>Bibliographie</b>	<b>47</b>
<b>Annexe A : Abréviations et Notations</b>	<b>48</b>

# Table des figures

3.1 Niveaux maximaux annuels et de rendement pour Oued Abioud.	39
3.2 Diagnostique graphique de Modèle.	41
3.3 Réalisations des paramètres de la GEV.	42
3.4 Rendements journalière de la compagnie d'assurance AXA.	43
3.5 Test graphique de la normalite de série des rendements par QQplot et his- togramm.	44

# Liste des tableaux

2.1 Lois a priori conjuguées pour quelques familles exponentielles usuelles.	21
3.1 Statistiques descriptives des données	39
3.2 Statistiques descriptives des blocs	40
3.3 Intervalle de confiance entre les deux estimations	40
3.4 Estimations ML des paramètres du modèle GEV pour GEV.	40
3.5 Estimations Bayesian des paramètres du modèle GEV .	41
3.6 Intervalle de confiance pour les estimateurs (GEV).	41
3.7 Estimation des paramètres des valeurs extrêmes par la méthode de MV et "Bayésien" pour La modèle GEV.	42
3.8 Statistique de série des rendements	44
3.9 Comparaison de la VaR et de la CVaR en termes de rendement aux niveau de confiance différents.	44

# Introduction

Ces dernières années, la théorie des valeurs extrêmes a reçu une grande attention pour ses applications dans la modélisation d'événements extrêmes (par exemple, tremblements de terre, inondations, crises financières, chocs pétroliers, etc.). Parce que les événements extrêmes prélèvent un lourd tribut, l'étude des indicateurs de risque et la théorie des valeurs extrêmes sont devenue un domaine d'étude actif.

La théorie classique des valeurs extrêmes est apparue entre 1920 et 1940, due à *M. Fréchet*, *R.A. Fisher* 1927 et *L. H. C. Tippett* 1928, *E. J. Gumbel* 1958 et *B. V. Gnedenko* 1943. Lorsqu'on modélise le maximum d'un ensemble de variables aléatoires. Alors, sous certaines conditions, la limite de distribution ne peut appartenir qu'à l'une des trois lois classes suivantes : *Weibull* (à support borné), *Gumbel* et *Fréchet* (à support non borné).

Une autre méthode pour modéliser les valeurs extrêmes est la méthode de dépassement du seuil (*POT*), cette dernière technique a été développée par *Pickand.J* 1975, et *Davison A. C et Smith R. L* 1990, *Smith* 1987, *Coles and Tawn* 1994...

On peut utiliser différentes méthodes pour estimer les paramètres des lois extrêmes, comme la méthode de maximum de vraisemblance (Coles 2001, Smith, 1987), Ces méthodes ne sont pas les seuls techniques pour estimer les paramètres des lois extrêmes. Une des méthodologies très importante pour faire de l'inférence statistique paramétrique est l'analyse bayésienne (Berger (1985) et Robert(1992, 2001)). Au lieu d'utiliser les données pour estimer un paramètre directement, on suppose que le paramètre suit une loi a priori et on utilise la *loi a posteriori* pour l'estimation. L'approche *bayésienne* offre une autre métho-



dologie d'estimation. Cette méthode est basé sur le traitement des paramètres inconnus comme des variables aléatoires et obtenir par la suite des estimateurs appropriés en utilisant les méthodes MONTE CARLO par CHAINES DE MARKOV (*MCMC*) (Coles et Powell 1996, Coles et tawn 1996, Smith 1999, Smith et Goodman 2000, Coles et Pericchi 2001), en utilisant la méthode traditionnelle de *Metropolis - Haskings* (Haskings, 1970; Metropolis et al, 1953). Nous avons structuré notre mémoire comme suit :

Le premier chapitre fondements partie premier est théorique des valeurs extrêmes est un rappel de différents aspects probabiliste des valeurs extrêmes uni-variées, à savoir les deux approches *GEV* proposée par Jenkinson (1955) et *GPD* (*GPD*) introduite par Pickands (1975), avec une présentation de la méthode d'estimation des paramètres méthode de maximum de vraisemblance. Et partie deuxième est l'exposé de la théorie de mesure de risque telles que, la variance, Value at-Risk (*VaR*), Tail Value-at-Risk (*TVaR*) cette dernière est connue par Conditional Tail Expectation (*CTE*) lorsque les risques sont continus, l'expected shortfall (*ES*) et les mesures basées sur la distorsion.

Le deuxième chapitre Approche bayésienne sur les paramètres des lois des valeurs extrêmes qui est réservé à l'exposition d'une approche bayésienne d'estimation des lois des valeurs extrêmes à savoir la méthode de *Metropolis - Hastings*.

Aux dernier chapitre présente deux modèles, le premier modèle étudiant le phénomène des crues dans les bassins versants secs à semi-arides de l'oued Abiud dans la région de Biskra. À cette fin, les données de crues de 1950 à 2011 de la station du barrage de *Foum El Gherza*. Ils ont été examinés par la méthode d'estimation de masse annuelle (*GEV*) et à l'aide de deux méthodes d'estimation *MV* et *MH*. Le second modèle sert à compiler les différentes méthodes d'estimation d'une mesure de risque particulière : la valeur à risque. Nous avons choisi l'une des méthodes d'estimation non paramétriques, la simulation historique de l'estimation de la *VaR* et de la *CVaR* sur des données réelles collectées en finance.

Tous les résultats numériques et graphiques sont produits à l'aide de *R*.

# Chapitre 1

## Théorie des Valeurs Extrêmes, Mesures de Risque

### 1.1 Théorie des Valeurs Extrêmes

#### 1.1.1 Statistique d'ordre

Soient  $X_1, X_2, \dots, X_n$ ,  $n$  v.a (*iid*) définies sur l'espace  $(\Omega, \mathfrak{B})$ , de fonction de distribution commune  $F$ . En rangeant ces v.a's par ordre croissant, on obtient ce que l'on appelle les statistiques d'ordre associées à l'échantillon  $(X_1, X_2, \dots, X_n)$  et que l'on note par :

$$X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$$

Pour  $1 \leq k \leq n$ , la v.a  $X_{k,n}$  est appelée la  $k$ -*ième* statistique d'ordre. En effet, les v. a. r. minimum et maximum du  $n$ -échantillon iid correspondent mieux à l'idée que l'on se fait d'une valeur extrême :

$$X_{1,n} = m_n = \min(X_1, X_2, \dots, X_n); \quad X_{n,n} = M_n = \max(X_1, X_2, \dots, X_n).$$

En fait, on peut se limiter à l'étude du maximum car

$$m_n = -\max(-X_1, -X_2, \dots, -X_n).$$

$F$  étant une fonction de répartition, l'hypothèse *i.i.d* entraîne

$$\mathbb{P}(M_n \leq z) = (F(z))^n.$$

### Loi des valeurs extrêmes

**Définition 1.1.1** *On dit que deux variables aléatoires réelles  $X$  et  $Y$  des lois respectives  $F$  et  $H$  sont de même type, s'il existe des constantes réelles  $a > 0$  et  $b \in \mathbb{R}$  tels que :*

$$Y \stackrel{\mathcal{L}}{=} aX + b \text{ i.e. } F(ax + b) = H(x).$$

Fisher et Tippett (1928) ont démontré l'existence des suites de normalisation  $a_n > 0$  et  $b_n \in \mathbb{R}$  et une loi non-dégénérée  $H$  telle que :

$$\mathbb{P}\left\{\frac{X_{n,n} - b_n}{a_n}\right\} = (F(a_n x + b_n))^n \rightarrow H(x).$$

Le théorème suivant donne une caractérisation de la distribution limite du maximum  $X_{n,n}$ .

**théorème 1.1.1** (Fisher-Tippett-Gnedenko) *S'il existe deux suites de constantes de normalisation  $a_n > 0$  et  $b_n \in \mathbb{R}$  et une loi non-dégénérée de loi  $H$  telle que  $\frac{X_{n,n} - b_n}{a_n} \xrightarrow{\mathcal{L}} H$ , alors  $H$  appartient à une des trois distributions standard des valeurs extrêmes suivantes :*

$$1. \text{ Fréchet : } \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} \quad \alpha > 0$$

$$2. \textit{ Weibull} : \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & x \leq 0 \\ 0, & x > 0 \end{cases} \quad \alpha < 0$$

$$3. \textit{ Gumbel} : \Lambda_0 = \{\exp(-\exp(-x)), \quad x \in \mathbb{R}\}$$

Ce théorème montre que la loi limite des extrêmes a toujours la même forme. Les trois formules précédentes peuvent être combinées en une seule paramétrisation :

$$H_\varepsilon(x) = \begin{cases} \exp(-(1 + \varepsilon x)^{-1/\varepsilon}), & \varepsilon = 0, 1 + \varepsilon x > 0 \\ \exp(-\exp(-x)), & \varepsilon \neq 0, -\infty \leq x \leq +\infty \end{cases} \quad (1.1)$$

Où  $H$  est une fonction de répartition non-dégénérée et est un paramètre qui contrôle la lourdeur de la queue de loi appelé indice des valeurs extrêmes. Cette loi est appelée loi de valeurs extrêmes généralisée (Generalized Extrême Value) que l'on note  $GEV$ . La forme la plus générale de la  $GEV$  est :

$$H_{\zeta, \mu, \sigma}(x) = \exp\left(-\left(1 + \zeta \frac{x - \mu}{\sigma}\right)^{-1/\zeta}\right), \quad \zeta \neq 0, \quad 1 + \zeta \frac{x - \mu}{\sigma} > 0 \quad (1.2)$$

$\mu$  et  $\sigma$  sont respectivement les paramètres de localisation et de dispersion.

### Domaines d'attraction

**Définition 1.1.2** *On dit qu'une distribution  $F$  appartient au domaine d'attraction de  $H$ , et on note  $F \in \mathcal{D}(H)$  si la distribution du maximum renormalisée converge vers  $H$ . Autrement dit, s'il existe des constantes réelles  $a_n > 0$  et  $b_n$  tels que :*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_\zeta(x) = \exp(-(1 + \zeta x)^{-1/\zeta})$$

Pour tout  $x$  avec  $1 + \zeta x > 0$ .

**Définition 1.1.3** *Le point terminal  $x_F$  et l'inverse généralisé  $F(y)$  d'une fonction  $F$  est définie par :*

$$x_F = \sup \{x, F(x) \leq 1\}. \quad F^{\leftarrow}(y) = \inf \{x \in \mathbb{R}, F(x) \geq y\}.$$

**Fonction à variations régulières** Une fonction  $U$  positive est à variations régulières d'indice  $\delta \in \mathbb{R}$  à l'infini, et on note :  $U(\cdot) \in RV_\delta$ , si :

$$\lim_{x \rightarrow \infty} \frac{U(\lambda x)}{U(x)} = \lambda^\delta, \quad \forall \lambda > 0$$

Dans le cas  $\delta = 0$  on dit que la fonction  $U(\cdot)$  est à variations lentes, et on note :  $U \in RV_0$ .

Toute fonction à variations régulières d'indice  $\delta \in \mathbb{R}$ , peut s'écrire :

$$U(x) = x^\delta L(x), \quad L(x) \in RV_0$$

**Lemme 1.1.1** Resnick (1987) *Soit  $U$  est une fonction à variations régulières d'indice, alors :*

$$\lim_{x \rightarrow \infty} \sup_{\lambda \in [a, b]} \left| \frac{U(\lambda x)}{U(x)} - \lambda^\delta \right| = 0, \quad \text{pour tout } 0 < a < b$$

**Lemme 1.1.2** Resnick (1987) *Si  $U$  est à variations régulières d'indice  $\delta > 0$ , alors  $U^{\leftarrow}(x)$  est à variations régulières d'indice  $1/\delta$ . Si  $U$  est à variations régulières d'indice  $\delta < 0$ , alors  $U^{\leftarrow}(x)$  est à variations régulières d'indice  $-1/\delta$ .*

**théorème 1.1.2** (Représentation de Karamata) *Toutes fonction à variations lentes  $L(\cdot)$  s'écrivent sous la forme :*

$$L(x) = c(x) \exp \left\{ \int_1^x \frac{\Delta(u)}{u} du \right\} \quad c(x) \longrightarrow c > 0 \quad \text{et } \Delta(x) \longrightarrow 0 \quad \text{lorsque } x \longrightarrow \infty.$$

*Cette représentation des fonctions à variations lentes est connue sous le nom de représentation de Karamata Bingham et al. (1987). Si la fonction  $c(\cdot)$  est constante, la fonction*

$L(\cdot)$  est dite normalisée.

### Caractérisation des domaines d'attraction

**Domaine d'attraction de Fréchet** Ce domaine d'attraction regroupe la majorité des distributions à queue lourde comme par exemple la loi de Cauchy, la loi de Pareto, Log-Gamma, et Student, etc...

**théorème 1.1.3** Soit  $F \in \mathcal{D}(\phi_\zeta)$  avec  $\zeta > 0$  si et seulement si  $x_F = +\infty$  et  $1 - F$  est une fonction à variations régulière d'indice  $-1/\zeta$  (i.e.  $1 - F(x) = x^{-1/\zeta}l(x)$ , où  $l$  est une fonction à variations lentes). Dans ce cas, un choix possible pour les suites  $a_n$  et  $b_n$  est  $a_n = F^{\leftarrow}(1 - \frac{1}{n})$  et  $b_n = 0$ .

**Domaine d'attraction de Weibull** Ce domaine d'attraction contient la majorité des fonctions de répartition dont le point terminal est fini (loi uniforme, etc...).

**théorème 1.1.4** Soit  $F \in \mathcal{D}(\Psi_\zeta)$  avec  $\zeta < 0$  si et seulement si  $x_F < +\infty$  et  $1 - F^*$  est une fonction à variations régulière d'indice  $1/\zeta$  (i.e.  $1 - F(x) = (x_F - x)^{1/\zeta}[l(x_F - x)^{-1}]$ )

avec :

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

Dans ce cas, un choix possible pour les suites  $a_n$  et  $b_n$  est :

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

**Domaine d'attraction de Gumbel** Ce domaine d'attraction regroupe la majorité des distributions à queue fine, par exemple loi normale, exponentielle, gamma, lognormale, etc...

**théorème 1.1.5** Soit  $F \in \mathcal{D}(\Lambda_\zeta)$  si et seulement si il existe  $t < x_F \leq +\infty$  tel que :

$$\bar{F}(x) = c(x) \exp \left\{ - \int_t^x \frac{1}{a(u)} du \right\}, \quad t < x < x_F$$

où  $\lim_{x \rightarrow x_F} c(x) = c > 0$  et  $a(\cdot)$  est une fonction positive et dérivable de dérivée  $a'(\cdot)$  telle que :  $\lim_{x \rightarrow x_F} a'(x) = 0$ , dans ce cas, un choix possible pour les suites  $(a_n)$  et  $(b_n) \forall n > 0$  est :

$$a_n = q(1/n) \quad b_n = \frac{1}{\bar{F}(a_n)} \int_{a_n}^{x_F} \bar{F}(s) ds.$$

**Exemple 1.1.1** La fonction de répartition de la loi exponentielle de paramètre  $\lambda > 0$  est :

$$F(x) = 1 - \exp(-\lambda x), \quad 0 \leq x \leq \infty.$$

On pose  $a_n = \frac{1}{\lambda}$  et  $b_n = \frac{1}{\lambda} \ln(n)$

$$F^n(a_n x + b_n) = (1 - \exp(-\frac{1}{\lambda} \lambda x - \lambda \frac{1}{\lambda} \ln(n)))^n = (1 - \frac{\exp(-x)}{n})^n \rightarrow \exp(-\exp(-x)) = \Lambda(x).$$

D'où le maximum normalisé de la loi exponentielle converge vers la loi de *Gumbel*, et  $F(x)$  appartient au domaine d'attraction de *Gumbel*.

### Loi des excès, approche POT

Soit  $u$  un réel suffisamment élevé appelé seuil. On définit les excès au-delà du seuil  $u$  comme l'ensemble des variables aléatoires conditionnelles  $\{Y_j\} = \{X_j - u \mid X_j > u\}$ . La loi conditionnelle des excès  $F_u$  par rapport au seuil  $u$  est :

$$F_u(y) = \Pr(X - u \leq y \mid X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y \geq 0$$

Ou de manière équivalente :

$$\bar{F}_u(y) = \Pr(X - u > y \mid X > u) = 1 - F_u(y) = \frac{\bar{F}(u + y)}{\bar{F}(u)}, \quad y \geq 0$$

**théorème 1.1.6** Une fonction de répartition  $F$  appartient au domaine d'attraction maximale de  $H_\zeta$ , si et seulement si, il existe une fonction positive  $\beta(u)$  telle que :

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x \leq x_F - u} |F_u(y) - G_{\zeta, \beta(u)}(y)| = 0.$$

où  $F_u(y)$  est la fonction de répartition conditionnelle des excès au delà d'un seuil  $u$ ,  $x_F$  est le point terminal de  $F$  et  $G_{\zeta, \beta(u)}(y)$  est la GPD donnée par :

$$G_{\zeta, \beta(u)}(y) = \begin{cases} 1 - (1 + \zeta \frac{y}{\beta(u)})^{-1/\zeta} & \zeta \neq 0 \\ 1 - \exp(-\frac{y}{\beta(u)}) & \zeta = 0 \end{cases}$$

Où  $y \geq 0$  pour  $\zeta \geq 0$  et  $y \leq 0 \leq -\frac{\beta(u)}{\zeta}$  pour  $\zeta < 0$ . On conclut de ce théorème, que si  $F$  vérifie le théorème de *Fisher* et *Tippet*, alors il existe une fonction  $\beta(\cdot)$  positive et un réel tel que la loi des excès  $F_u$  peut être uniformément approchée par une *distribution de Pareto généralisée (GPD)* notée  $G_{\zeta, \beta(u)}$ . Ainsi, l'indice des valeurs extrêmes donnée par le théorème de *Fisher* et *Tippet* est le même que celui de la loi des excès.

**Exemple 1.1.2** Pour le cas de la loi de Pareto, la fonction de répartition est :

$$F(x) = 1 - cx^{-\alpha}, \quad \text{avec } c > 0 \text{ et } \alpha > 0,$$

On prend  $\beta_u = ub$ , avec  $b > 0$  et  $y \geq 0$ , on obtient :

$$F_u(y) = \frac{F(u + uby) - F(u)}{1 - F(u)} = \frac{cu^{-\alpha} - c(u + uby)^{-\alpha}}{cu^{-\alpha}} = 1 - (1 + by)^{-\alpha}.$$



D'où, pour  $\zeta = 1/\alpha$  et  $b = \zeta$  la loi limites est la loi *GPD* de paramètre  $\zeta$ .

## 1.1.2 Estimation des Valeurs Extrêmes

### Estimation de l'indice des queues

**Estimateur du maximum de vraisemblance** L'estimateur du maximum de vraisemblance est construit à partir des observations des maxima, il s'agit d'estimer l'indice des valeurs extrêmes ainsi que les deux suites normalisantes  $a_n$  et  $b_n$ .

Soit un échantillon de  $k$  maxima  $Z_1, \dots, Z_k$  *i.i.d.*, suivant une loi *GEV*. La fonction de *log-vraisemblance* est donnée par :

$$L(Z, \zeta, b_n, a_n) = \begin{cases} k \log a_n - \left(\frac{1}{\zeta} + 1\right) \sum_{i=1}^k \log\left(1 + \zeta \frac{Z_i - b_n}{a_n}\right) - \sum_{i=1}^k \left(1 + \zeta \frac{Z_i - b_n}{a_n}\right)^{-1/\zeta} & \zeta \neq 0 \\ k \log a_n - \sum_{i=1}^k \exp\left(\frac{Z_i - b_n}{a_n}\right) - \sum_{i=1}^k \left(\frac{Z_i - b_n}{a_n}\right) & \zeta = 0 \end{cases}$$

*Smith (1985)* a démontré les propriétés de consistance et la normalité asymptotique de cet estimateur lorsque  $\zeta > 1/2$  et  $m \rightarrow \infty$  :

$$\sqrt{m} \left( \begin{pmatrix} \hat{a}_n \\ \hat{\zeta} \\ \hat{b}_n \end{pmatrix} - (a_n, \zeta, b_n) \right) \rightarrow \mathcal{N}(0, I^{-1})$$

Où  $I$  est la matrice *d'information de Fisher* estimée par sa version empirique :

$$I(\Theta) = -\mathbb{E}\left(\frac{\partial^2 L(Z; \Theta)}{\partial \Theta^2}\right)$$

$L(Z; \Theta)$  est la fonction de *log-vraisemblance* associée à la loi de la variable aléatoire  $Z$ , paramétrée par un ensemble de paramètres  $\Theta$ .

**Estimation de quantile extrême**

**Définition 1.1.4** On appelle quantile d'ordre  $p$ , le nombre  $X_p$  est définie par :

$$X_p = \inf \{x, F(x) \geq p\}, \quad p \in [0, 1]$$

La fonction quantile de queue définie par :

$$U(t) = F^{\leftarrow} \left( 1 - \frac{1}{t} \right), \quad t \in ]1, +\infty[.$$

Où  $F^{\leftarrow}$  est l'inverse généralisé de  $F$ .

**Définition 1.1.5** Le quantile extrême d'ordre  $n$  est définie par :

$$q_{\alpha_n} = \bar{F}^{\leftarrow}(\alpha_n) = \inf \{x : F(x) \leq \alpha_n\} \quad \text{avec} \quad \lim_{n \rightarrow \infty} \alpha_n = 0$$

$Q_n$  Peut être exprimée comme une fonction simple des statistiques d'ordre relatives à l'échantillon  $(X_1, X_2, \dots, X_n)$  et nous avons :

$$Q_n(s) = \begin{cases} X_{i,n} & \frac{i-1}{n} < s \leq \frac{i}{n} \\ X_{[np]+1,n} & 0 < s \leq 1 \end{cases}$$

Notons que pour  $0 < p < 1$ ,  $X_{[np]+1,n}$  est le quantile d'échantillon de l'ordre  $p$ .

**L'approche par la loi des valeurs extrêmes** D'après le théorème de *Fisher-Tippett*,

on a :

$$Pr = \left\{ \frac{X_{n,n} - b_n}{a_n} \leq x \right\} = (F(a_n x + b_n))^n \simeq H_\zeta(x) \quad (1.3)$$

Cela nous donne

$$Pr = \{X_{n,n} \leq z\} \approx H_\zeta\left(\frac{z - b_n}{a_n}\right) = H_{\zeta, a_n, b_n}$$

Donc la loi de  $X_{n,n}$  peut être approchée par  $H_{\zeta, a_n, b_n}$ . L'équation [1.3](#) est équivalent à

$$\lim_{n \rightarrow \infty} n \log F(a_n x + b_n) = \lim_{n \rightarrow \infty} n \log \left( 1 - \bar{F}(a_n x + b_n) \right) = \log H_{\zeta}(x)$$

Lorsque  $n \rightarrow \infty$  on a  $\bar{F}(a_n x + b_n) \rightarrow 0$ , donc le développement limité de  $\log(1 - \bar{F}(a_n x + b_n))$  au premier ordre est :

$$\bar{F}(a_n x + b_n) \simeq -\frac{1}{n} \log H_{\zeta}(x)$$

Cela implique que

$$\bar{F}(x) \simeq -\frac{1}{n} \log H_{\zeta, a_n, b_n}$$

D'après l'équation, une approximation de fonction de survie en queue est :

$$\bar{F}(x) \simeq \frac{1}{n} \left( 1 + \zeta \left( \frac{x - b_n}{a_n} \right)^{-1/\zeta} \right)$$

d'où on peut approcher le quantile  $q(\alpha)$  par :

$$q(\alpha) \simeq b_n + \frac{a_n}{\zeta} \left( \left( \frac{1}{n\alpha} \right)^{\zeta} - 1 \right)$$

**Définition 1.1.6** *L'estimateur du quantile extrême de la loi GEV est défini par :*

$$\hat{q}_n(\alpha_n) \simeq \hat{b}_n + \frac{\hat{a}_n}{\hat{\zeta}} \left( \left( \frac{1}{n\alpha} \right)^{\hat{\zeta}} - 1 \right)$$

où  $(\hat{a}, \hat{b}, \hat{\zeta})$  sont respectivement des estimateurs des paramètres  $(a_n, b_n, \zeta)$ .

## 1.2 Mesures de risque

**Définition 1.2.1** *Une mesure de risque univarié est une fonction  $\rho$  associant à chaque variable aléatoire du montant de perte  $X$  une valeur réelle  $\rho(X)$ .*

### 1.2.1 Région de risque

**Définition 1.2.2** On définit la région de risque acceptable pour la mesure  $\varrho$  par :

$$G = \{X, \varrho(X) < 0\}$$

**Définition 1.2.3** Une mesure de risque  $\varrho$  contient un chargement de sécurité si pour tout risque  $X$ , on a :

$$\varrho(X) \geq E(X)$$

A partir de la définition de la mesure de risque, on peut considérer toute fonction définie sur l'espace des variables aléatoires, qui prend ses valeurs dans  $\mathbb{R}$  comme mesure de risque.

### 1.2.2 Quelques mesures de risque

#### La variance

La variance est une mesure de dispersion d'une variable aléatoire autour de sa moyenne.

Si  $X$  est une variable aléatoire de carré intégrable, sa variance est définie par :

$$Var(X) = \mathbb{E} [(X - \mathbb{E}(X))^2].$$

#### La Value-at-Risk

L'une de plus populaires mesures de risque est la Value-at-Risk (VaR), appelée également "Valeur à Risque".

**Définition 1.2.4** On appelle Value-at-Risk de niveau  $\alpha \in (0, 1)$  le quantile de niveau  $\alpha$ .

$$\varrho[X] = VaR_\alpha[X] = x_\alpha \quad \text{où} \quad P(X \leq x_\alpha) = \alpha$$

Ou encore :

$$\begin{aligned} VaR_\alpha[X] &= F^{-1}(\alpha) = \inf\{x : P[X > x] \leq 1 - \alpha\} \\ &= \inf\{x : P(X \leq x) \geq \alpha\} = \sup\{x : F(x) < \alpha\} = Q(\alpha) \end{aligned}$$

i.e, la  $VaR$  est le seuil minimum excédé par  $X$  avec la probabilité tout au plus  $1 - \alpha$ .

La  $VaR$  a les propriétés suivantes :

- $VaR_\alpha[X] \leq \max[X]$  ,  $\forall \alpha \in (0, 1)$
- La  $VaR$  est monotone :  $VaR_\alpha[X] \leq VaR_\alpha[Y]$  si  $X \leq Y$ .
- La  $VaR$  est translation invariant,  $VaR_\alpha[X + c] = VaR_\alpha[X] + c$ .
- La  $VaR$  est positive homogène  $VaR_\alpha[cX] = c \times VaR_\alpha[X]$ .
- La  $VaR$  est comonotone additive : Si les risques  $X_1, X_2, \dots, X_n$  sont comonotons, alors :

$$VaR_\alpha[X_1 + X_2 + \dots + X_n] = VaR_\alpha[X_1] + VaR_\alpha[X_2] + \dots + VaR_\alpha[X_n].$$

Les deux dernières propriétés découlent également de la relation générale suivante.

**Proposition 1.2.1** *La  $VaR$  d'une v.a.  $Y = g(X)$ , où  $g(\cdot)$  est une fonction non-décroissante d'une v.a.  $X$ , peut être calculée comme suivant :*

$$VaR_\alpha[X] = g(VaR_\alpha[X])$$

## La Tail Value-at-Risk

**Définition 1.2.5** *Tail Value-at-Risk de niveau associée au risque  $X$ , est donnée par :*

$$TVaR_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_s(X) ds$$

Notons que  $TVaR_0(X) = \int_0^1 VaR_s[X] ds = E[X]$  et comme

$$TVaR_\alpha(X) = \frac{1}{1-\alpha} \left( E[X] - \int_\alpha^1 VaR_s(X) ds \right)$$

On en déduit que la  $TVaR$  est une fonction croissante en  $\alpha$ . De plus

$$TVaR_\alpha(X) \geq TVaR_0(X) = E[X]$$

Donc la Tail-VaR contient toujours un chargement de sécurité.

**Lemme 1.2.1** Soient le risque  $X$  et le niveau de perte  $x$  tels que  $F_X(x) > 0$ . Quel que soit l'événement aléatoire  $E$  tel que  $P[E] = F_X(x)$ , on a :

$$E[X_j | E] \leq E[X_j | X > x]$$

Ce lemme garanti que la  $TVaR$  est sous-additive lorsque les risques sont continus.

### Conditional Tail Expectation

**Définition 1.2.6** La Conditional-Tail-Expectation (CTE) d'une v.a.  $X \sim F(x)$  au niveau de probabilité  $\alpha$ , notée  $CTE_\alpha[X]$  comme l'espérance de  $X$  conditionnellement aux valeurs supérieures à  $VaR_\alpha(X)$

$$CTE_\alpha[X] = E[X | X > VaR_\alpha[X]]$$

La  $CTE$  et la  $TVaR$  coïncident pour des risques dont la fonction de répartition est continue, i.e.

$$CTE_\alpha(X) = TVaR_\alpha(X), \alpha \in (0, 1)$$

**Expected shortfall**

**Définition 1.2.7** *L'expected shortfall au niveau  $\alpha$ , notée  $ES_\alpha[X]$ , est*

$$ES_\alpha[X] = E[(X - VaR_\alpha[X])_+].$$

**Proposition 1.2.2** *Pour tout  $\alpha \in [0, 1]$ , les identités suivantes sont valides :*

$$TVaR_\alpha(X) = VaR_\alpha(X) + \frac{1}{1-\alpha} ES_\alpha(X)$$

$$CTE_\alpha(X) = VaR_\alpha(X) + \frac{1}{\bar{F}_X(VaR_\alpha(X))} ES_\alpha(X)$$

$$CVaR_\alpha(X) = \frac{ES_\alpha(X)}{\bar{F}_X(VaR_\alpha(X))}$$

**Preuve.** Voir *Denuit (2005)* ■

**1.2.3 Estimation par méthode du maximum de vraisemblance**

Le modèle *GEV* non stationnaire a été introduit dans le cas d'une tendance au niveau du paramètre de position par *SCARF (1992)*, qui a présenté la méthode du *E.M.V* pour l'estimation des paramètres du modèle *GEV* à quatre paramètres, qui ne sont pas des constantes mais dépendent du temps ou d'autre covariables.

**Définition 1.2.8** *Soit  $X = (X_1, X_2, \dots, X_n)$ , le vecteur des observations de la variable  $X$  et  $\theta$  le vecteur des paramètres à estimer. Dans notre cas :*

\*  $\theta = (\mu, \sigma, \zeta)$  pour le modèle stationnaire *GEV*<sub>0</sub>.

\*  $\theta = (\mu_0, \mu_1, \sigma, \zeta)$  pour le modèle *GEV*<sub>1</sub> (Le cas où le paramètre de position est une fonction linéaire de la covariable :  $\mu(t) = \mu_0 + \mu_1 t$ ).

\*  $\theta = (\mu_0, \mu_1, \mu_2, \sigma, \zeta)$  pour le modèle *GEV*<sub>2</sub> (Le cas d'une dépendance quadratique :  $\mu(t) = \mu + \mu_1 t + \mu_2 t^2$ ).

La méthode du maximum de vraisemblance s'écrit d'une manière générale :

$$\begin{aligned}
 L_n(X \setminus \mu_t, \sigma_t, \zeta_t) &= \prod_{t=1}^{n_1} \frac{1}{\sigma_t} \exp \left\{ - \left[ 1 - \zeta_t \left( \frac{X_t - \mu_t}{\sigma_t} \right) \right]^{\frac{-1}{\zeta_t}} \right\} \\
 &* \left[ 1 - \zeta_t \left( \frac{X_t - \mu_t}{\sigma_t} \right) \right]^{-\left(1 - \frac{1}{\zeta_t}\right)} \\
 &* \prod_{t=n_1+1}^n \frac{1}{\sigma_t} \exp \left\{ - \left( \frac{X_t - \mu_t}{\sigma_t} \right) \right\} * \exp \left\{ - \exp - \left[ \left( \frac{X_t - \mu_t}{\sigma_t} \right) \right] \right\}
 \end{aligned}$$

$n_1$  est le nombre d'observations telles que  $\zeta_t \neq 0$  pour les modèles étudiés dans ce travail, la non stationnarité est liée au paramètre de position. Dans ce cas, et pour tous les modèles considérées,  $\sigma_t = \sigma$  et  $\zeta_t = \zeta$  sont constants. Lorsque  $\zeta_t \neq 0$  et  $n_1 = n$ , la fonction logvraisemblance est :

$$\begin{aligned}
 L_n(X, \mu_t, \sigma, \zeta) &= -n \log(\sigma) - \sum_{t=1}^n \left[ 1 - \zeta \left( \frac{X_t - \mu_t}{\sigma} \right) \right]^{\frac{-1}{\zeta}} \\
 &\sum_{t=1}^n \left( 1 - \frac{1}{\zeta} \right) \log \left[ 1 - \zeta \left( \frac{X_t - \mu_t}{\sigma} \right) \right]
 \end{aligned}$$

Les estimateurs du maximum de vraisemblance des paramètre  $(\mu_0, \mu_1, \sigma, \zeta)$  sont des solutions du système d'équation :

$$\begin{aligned}
 \sum_{t=1}^n \left( \frac{1 - \zeta - z_t^1}{z_t} \right) &= 0, \quad \sum_{t=1}^n t \left( \frac{1 - \zeta - z_t^1}{z_t} \right) = 0 \\
 -n + \sum_{t=1}^n \left[ \frac{1 - \zeta - z_t^{\frac{1}{\zeta}}}{z_t} \left( \frac{X_t - \mu_t}{\sigma} \right) \right] &= 0 \\
 \sum_{t=1}^n \left\{ \ln(z_t) \left[ 1 - \zeta - z_t^{\frac{1}{\zeta}} \right] + \frac{1 - \zeta - z_t^{\frac{1}{\zeta}}}{z_t} \zeta \left( \frac{X_t - \mu_t}{\sigma} \right) \right\} &= 0, \quad \text{où } z_t = \left[ 1 - \frac{\zeta}{\sigma} (X_t - \mu_t) \right].
 \end{aligned}$$

La résolution de ce système peut se faire d'une manière numérique en utilisant la méthode de *Newton-Raphson*. De la même façon, on obtient un système équivalent pour le modèle  $GEV_2$ , avec une cinquième équation correspondant au paramètre  $\mu_2$ .



# Chapitre 2

## Approche bayésienne sur les paramètres des lois des valeurs extrêmes

L'approche bayésienne peut être présentée comme une généralisation de l'approche classique. Les paramètres ne sont plus des valeurs inconnues fixées, mais des variables aléatoires. Nous considérons que l'objet principal de la statistique est une inférence au sujet de la loi générant ces observations ; afin, soit d'analyser un phénomène passé, soit de prévoir un événement futur, dans ce chapitre, nous allons donner quelques notions de bases qui concernent la statistique bayésienne sur valeur extrême.

### 2.1 Définitions

**Définition 2.1.1** *Approche statistique classique* : Dans l'approche fréquentiste, les données observées sont considérées comme des observations de variables aléatoires, dans cette approche, on écrit :

$$L(\theta|x) = f(x|\theta).$$

On considérant  $L(\theta|x)$  comme une fonction de  $\theta$ . En estimation on cherche la valeur  $\hat{\theta}_{MV}$  qui maximise  $L(\theta|x)$ , c'est l'estimation au sens du maximum de vraisemblance :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} \{L(\theta|x)\}.$$

**Définition 2.1.2 Approche statistique bayésienne :** Dans l'approche bayésienne, l'idée de base consiste à traiter le paramètre inconnu  $\theta$  comme une variable aléatoire admet une densité de probabilité  $\pi(\theta)$  qui s'appelle densité a priori. L'objectif est donc d'utiliser cette information sachant que l'information contenue dans les observations  $x$  est contenue dans  $f(x|\theta)$  et l'information a priori sur  $\theta$  dans  $\pi(\theta)$ , on peut utiliser la formule de Bayes pour combiner ces deux types d'information en définissant la densité a posteriori par :

$$\pi(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta},$$

Qui contiendra toutes les informations sur  $\theta$ . On appelle modèle statistique bayésien sur  $(\Omega, \mathcal{F}, \mathcal{P})$  la donnée d'un modèle statistique paramétré  $f(\theta)$ , et d'une loi a priori sur les paramètres  $\pi(\theta)$ .

**Définition 2.1.3 Lien entre les deux approches :** Considérant que l'absence d'information a priori  $\pi(\theta)$  justifiait que la fonction de vraisemblance  $L(x|\theta)$  est équivalente à la densité a posteriori  $\pi(\theta/x)$  lorsqu'on choisit une loi a priori uniforme définie sur  $[0, 1]$  pour  $\pi(\theta)$  c'est-à-dire  $\pi(\theta/x) \propto L(x|\theta)$ . L'approche bayésienne est celle passant par la loi a posteriori. En effet travaillant conditionnellement aux observations, cette approche suit d'une manière cohérente l'idée d'inversion des paramètres aux observations, tout en restant principe de vraisemblance. On peut dire que l'approche statistique fréquentiste où toute l'inférence est basée sur la vraisemblance.

## 2.2 Principe de la statistique bayésienne

Le principe de l'analyse statistique bayésienne est de modéliser des paramètres inconnus de la distribution d'échantillonnage à travers une structure probabiliste. Elle vise à exploiter le plus efficacement possible l'information apportée par  $X$  sur le paramètre  $\theta$ , pour ensuite construire des procédures d'inférence sur ce paramètre.

**Théorème de Bayes** La règle de Bayes est la conséquence mathématique directe du théorème des probabilités conditionnelles cas continu.

Soit  $\Theta \subseteq \mathbb{R}^p$  espace paramétrique et  $\pi(\theta)$  la densité de probabilité de  $\theta$ . La densité conditionnelle  $f(x|\theta)$  considérée comme fonction de  $\theta$  lorsque  $x$  est observé est appelée vraisemblance de  $\theta$  on la note  $L(\theta|x)$  est la densité de probabilité a posteriori est donnée par :

$$\pi(\theta|x) = \frac{L(\theta|x)\pi(\theta)}{m(x)} \quad \text{où} \quad m(x) = \int_{\Theta} L(\theta|x)\pi(\theta)d\theta$$

Est la densité marginale de  $x$ . on a aussi :

$\pi(\theta x)$	$\propto$	$L(\theta x)$	$\pi(\theta)$
densité a posteriori		vraisemblance	densité a priori

La loi a posteriori représente l'actualisation de l'information a priori  $\pi(\theta)$  en vue de l'information contenue dans les observations  $f(x|\theta)$ .

## 2.3 Distribution a priori

### Le problème de choix de l'a priori

Le choix de l'apriori est une étape fondamentale dans l'analyse bayésienne. En théorie cette loi doit représenter notre information a priori sur le paramètre. En pratique cette information est rarement fournie sous la forme probabiliste.

Les différents choix possible peuvent être motivés par différents points de vue :

- Choix basé sur des expériences du passé ou sur une intuition du statisticien, dite "*distribution a priori informative*".
- Choix basé sur la faisabilité de calcul, dite "*distribution a priori non informative*".

### Distribution a priori informative

**Lois a priori conjuguées** Une famille  $F$  de lois sur  $\Theta$  est dite conjuguée ou fermée par échantillonnage si pour tout  $\pi(\theta) \in F$ , la loi a posteriori  $\pi(\theta|X)$  appartient également à  $F$ . *Raiiffa et Schlaifer* (1961), qui sont l'origine des lois conjuguées. Lors que les observations  $x$  modifient la loi a priori  $\pi(\theta)$  en la loi a posteriori  $\pi(\theta|X)$ , l'information apportée par  $x$  sur  $\theta$  est limitée, donc elle ne doit pas conduire à une remise en cause de la forme de la loi  $\pi(\theta)$ . Mais elle peut seulement changer ses paramètres en fonction de  $x$ .

Le tableau ci-dessus représente quelques lois a priori conjuguées naturelles pour quelques familles exponentielles usuelles.

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
<i>Binomiale</i> $B(n, \theta)$	<i>Beta</i> $Beta(\alpha, \beta)$	<i>Beta</i> $Beta(\alpha + x, \beta + n - x)$
<i>Normale</i> $N(\theta, \sigma^2)$	$N(\mu, \tau^2)$	$N(\varphi(\sigma^2\mu + \tau^2x), \varphi\sigma^2\tau^2)$ avec $\varphi^{-1} = \sigma^2 + \tau^2$
<i>Poisson</i> $P(\theta)$	<i>Gamma</i> $G(\alpha, \beta)$	<i>Gamma</i> $G(\alpha + x, \beta + 1)$
<i>Gamma</i> $G(v, \theta)$	<i>gamma</i> $G(\alpha, \beta)$	<i>gamma</i> $G(\alpha + v, \beta + x)$

TAB. 2.1 – Lois a priori conjuguées pour quelques familles exponentielles usuelles.

### Distribution a priori non informative

De nombreux travaux, initiés par *Bayes* (1763), *Laplace* (1825), ont tenté de construire des a priori avec lesquels la subjectivité du statisticien n'interfère pas. *Jeffreys* (1946), propose une approche intrinsèque fournit une des meilleures technique pour obtenir une loi a priori non informative.

**Lois a priori de Jeffreys** Les lois a priori non informative de *Jeffreys* sont fondées sur *l'information de Fisher* donnée par :

$$I(\theta) = E_{\theta} \left[ \left( \frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right]$$

Dans le cas unidimensionnel.

Sous certaines conditions, cette information est aussi égale à

$$I(\theta) = -E_{\theta} \left[ \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right]$$

La loi a priori de *Jeffreys* est :

$$\pi(\theta) \propto I^{-1}(\theta)$$

Le choix d'une loi a priori dépendant de *l'information de Fisher* se justifie par le fait que  $I(\theta)$  est largement accepter comme une indicateur de la quantité d'information apportée par le modèle ou l'observation sur  $(\theta)$ .

## 2.4 Distribution a posteriori

C'est la loi conditionnelle de  $\theta$  sachant  $x$  sa densité est notée  $\pi(\theta|x)$ . En vertu de la formule de Bayes, on a :

$$\pi(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta} = \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} = \frac{h(\theta, x)}{m(x)}$$

Où encore, on peut écrire :

$$\pi(\theta|x) \propto \prod_{i=1}^n f(x_i|\theta)(p)$$

Où  $\propto$  signifie proportionnelle au produit de la vraisemblance par la densité a priori.

## 2.5 Estimation ponctuelle

### 2.5.1 Fonction de perte

**Définition 2.5.1** Soit  $\hat{\theta}$ , un estimateur de  $\theta$ . On appelle "fonction de perte" et on note  $l(\hat{\theta}, \theta)$ , tout fonction satisfaisant :

$$\begin{cases} l(\hat{\theta}, \theta) \geq 0 & \text{pour tout valeur d'estimateur } \hat{\theta} \text{ et } \forall (\theta, \hat{\theta}) \in \Theta \subset \mathbb{R}^p. \\ l(\hat{\theta}, \theta) = 0 & \text{si } \hat{\theta} = \theta. \end{cases}$$

Cette fonction mesure la perte occasionnée lorsqu'on estimer  $\theta$  par  $\hat{\theta}$ .

Nous représentons quelques fonctions de perte :

$$\begin{aligned} l_1(\hat{\theta}, \theta) &= a_1(\hat{\theta} - \theta)I_{\{\theta \leq \hat{\theta}\}}(t) + a_2(\theta - \hat{\theta})I_{\{\theta > \hat{\theta}\}}(t). \\ l_2(\hat{\theta}, \theta) &= \varphi(\theta) \left| \hat{\theta} - \theta \right|^r \quad \text{avec } \varphi(\theta) \geq 0 \text{ et } r > 0 \end{aligned}$$

**Remarque 2.5.1** Si  $\varphi(\theta) = 1$ , et  $r = 2$ ,  $l_2(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$  est l'erreur quadratique, qui pénalise trop fortement les grandes erreurs. En effet, on a :

$$\begin{aligned} E[l_2(\theta, \hat{\theta})] &= E[(\theta - \hat{\theta})^2] \\ &= E[\theta^2] - 2\hat{\theta}E[\theta] + \hat{\theta}^2 \\ &= (\hat{\theta} - E[\theta])^2 + \text{var}[\theta], \end{aligned}$$

Qui est minimisé lorsque

$$\hat{\theta} = E[\theta].$$

Si  $a_1 = a_2$ ,  $l_1(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$  est erreur absolue.

### Fonction de perte intrinsèque

On souhaite mesurer l'écart entre le vrai modèle  $M \equiv \{p(x|\theta, \lambda), x \in \mathcal{X}, \theta \in \Theta, \lambda \in \Lambda\}$  et sa plus proche approximation dans l'ensemble des modèles  $M_{\hat{\theta}} \equiv \{p(x|\hat{\theta}, \hat{\lambda}), x \in \mathcal{X}, \hat{\lambda} \in \Lambda\}$ . C'est précisément ce que réalise la divergence intrinsèque, c'est-à-dire :

$$\delta(M_{\hat{\theta}}, p_{x|\theta, \lambda}) = \delta(\hat{\theta}, (\theta, \lambda)) = \min_{\hat{\lambda} \in \Lambda} \delta(p_{x|\hat{\theta}, \hat{\lambda}}, p_{x|\theta, \lambda})$$

Cette dernière fait partie des fonctions de perte intrinsèques, introduites par *Robert* (1996).

Remarquons que dans le cas simple où il n'y pas le paramètre  $\lambda$ , la fonction de perte intrinsèque est simplement  $\delta(\hat{\theta}, \theta)$ , est l'espérance postérieurs de la perte intrinsèque s'écrit sans l'intégration sur  $\Lambda$  :

$$d(\hat{\theta}|x) = E_{\theta|x}(\delta(\hat{\theta}, \theta)) = \int_{\Theta} \delta(\hat{\theta}, \theta) \pi(\theta|x) d\theta$$

## 2.5.2 Risque

L'approche fréquentiste restreint l'espace d'estimation en préférant la classe des estimateurs sans biais dans la quelle il existe des estimateurs de risque uniformément minimal, l'école Bayésienne ne perd pas en généralité en définissant un risque a posteriori. L'idée est intégrer sur l'espace des paramètres pour pallier cette difficulté.

**Risque a posteriori** Si en données la loi a posteriori sur le paramètre et la fonction de perte, le risque a posteriori est défini par :

$$R(\pi, \hat{\theta}|x) = E^\pi[l(\theta, \hat{\theta})|x] = \int_{\Theta} l(\theta, \hat{\theta})\pi(\theta|x)d\theta$$

On générale dire que  $\hat{\theta}_1$  est meilleur que  $\hat{\theta}_2$  au sens du *risque de Bayes*, si :

$$R(\hat{\theta}_1) < R(\hat{\theta}_2).$$

## 2.6 Estimateur bayésien

Soit  $X = (X_1, \dots, X_n)$  un  $n$ -échantillon de variables aléatoire *i.i.d* de densité  $f(x|\theta)$ , où  $\theta$  est un paramètre inconnu à estimer. On dit que le  $\theta$  est une valeur d'une variable aléatoire aquilon assigne une loi a priori  $\pi(\theta)$ , puis on déterminer la distribution a posteriori  $\pi(\theta|x)$  à l'aide du théorème de *Bayes*.

### Estimation intrinsèque

**Définition 2.6.1** *L'estimateur de Bayes intrinsèque est la fonction des données qui minimise l'espérance postérieure de la perte intrinsèque*

$$\hat{\theta}(x) = \arg \min_{\hat{\theta} \in \Theta} d(\hat{\theta}|x) = \arg \min_{\hat{\theta} \in \Theta} \int_{\Theta} \delta(\hat{\theta}, \theta)\pi(\theta|x)d\theta.$$



On suppose dans cette section que le paramètre  $\theta$  est réel.

**Proposition 2.6.1** *Si*

$$l(\hat{\theta}, \theta) = \begin{cases} a & \text{si } \hat{\theta} \neq \theta \quad a > 0 \\ 0 & \text{si } \hat{\theta} = \theta, \end{cases}$$

*l'estimateur de Bayes correspond au mode de la distribution a posteriori  $\pi(\theta|x)$ ; est défini par :*

$$\pi(\hat{\theta}|x) = \sup_{\theta \in \Theta} \pi(\theta|x)$$

Si

$$l(\hat{\theta}, \theta) = a_1(\hat{\theta} - \theta)\chi_{\{\theta \leq \hat{\theta}\}}(t) + a_2(\theta - \hat{\theta})\chi_{\{\theta > \hat{\theta}\}}(t),$$

*l'estimateur de Bayes  $\hat{\theta}$  est le quantile  $\frac{a_1}{a_1+a_2}$  de la distribution a posteriori  $\pi(\theta|x)$ ; est défini par :*

$$P(\Theta < \hat{\theta}|x) = \frac{a_1}{a_1 + a_2}$$

Lorsque  $a_1 = a_2$ , l'estimateur de Bayes est la *médiane a posteriori*  $\pi(\theta|x)$ ; est définie par :

$$\int_{-\infty}^{\hat{\theta}} \pi(\theta|x) d\theta = \int_{\hat{\theta}}^{+\infty} \pi(\theta|x) d\theta$$

Si  $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ , l'estimateur de Bayes est la *moyenne a posteriori*  $\pi(\theta|x)$ ; est définie par :

$$\hat{\theta}_B(x) = E[\theta|x] = \int_{\Theta} \theta \pi(\theta|x) d\theta = \frac{\int_{\Theta} \theta f(x|\theta) \pi(\theta) d\theta}{\int_{\Theta} f(x|\theta) \pi(\theta) d\theta}.$$

**Remarque 2.6.1** *Il faut signaler que le mode de la loi a posteriori est parfois retenu comme estimation bayésienne de  $\theta$  si l'on suspecte que la vraisemblance comporte des*

*maximum local.* En ce cas, on a :

$$\hat{\theta}_B(x) = \arg \max_{\theta \in \Theta} \pi(\theta|x)$$

*La moyenne et la médiane sont généralement de meilleurs estimateurs pour  $\theta$  que le mode.*

### **Propriétés de l'estimateur de Bayes**

- L'estimateur de Bayes est admissible.
- L'estimateur de Bayes est biaisé.
- Sous certaines hypothèses de régularité le plus souvent satisfaites en pratique, on a les deux propriétés :
  - L'estimateur de Bayes est convergent en probabilité (quand la taille de l'échantillon  $n \rightarrow +\infty$ )
  - La loi a posteriori peut être asymptotiquement (c-à-d pour de grandes valeurs de  $n$ ) approximée par une loi normale  $N(E[\theta|x], Var[\theta|x])$ . Cette dernière propriété est particulièrement utile pour construire des intervalles de confiance a posteriori.

## **2.7 Méthodes usuelles de calcul Bayésien**

La simplicité ultime de l'approche Bayésienne est que, pour une fonction de coût et une loi a priori données, l'estimation Bayésienne associée à une observation  $x$  est la décision  $d$  minimisant le coût a posteriori, dans la pratique, minimiser un tel coût peut être difficile car le calcul explicite de la loi a posteriori peut être impossible des fois.

### 2.7.1 Les méthodes de Monte Carlo

Dans un problème statistique, l'approximation de l'intégrale

$$\int_{\Theta} g(\theta) f(x|\theta) \pi(\theta) d\theta$$

doit tirer avantage de la nature particulière, à savoir de fait que  $\pi$  soit densité de probabilité ou plutôt, que  $f(x|\theta)\pi(\theta)$  soit proportionnel à densité. Une conséquence naturelle de cette perspective est d'utiliser la méthode de *Monte Carlo*, introduite par *Metropolis* et *Ulam* (1949) et von *Neumann* (1951).

### 2.7.2 Méthodes de Monte Carlo par chaînes de Markov (MCMC)

Les méthodes de Monte Carlo par chaîne de Markov (MCMC) génèrent une suite de variables aléatoires  $(\theta^1, \dots, \theta^n, \dots)$  et, hormis la première à laquelle on donne une valeur arbitraire, chacune d'entre elles dépend uniquement de celle qui la précède les calculs sont ensuite poursuivis en appliquant à cette séquence une liste des grands nombres pour les chaînes markoviennes ergodiques de forme identique.

#### Algorithme Hasting-Metropolis

Pour  $\theta^0$  est une valeur initiale, on définit par récurrence les valeurs de  $\theta^t$ .

A l'étape  $t$ , à partir de  $\theta^{(t-1)}$ ,  $\theta^t$  est construit en tirant un  $\acute{\theta}$  à l'aide d'une distribution de probabilité instrumentale :  $\acute{\theta} \sim q(\cdot | \theta^{(t-1)})$ .  $\theta^t$  est alors donné par :

$$\theta^t = \begin{cases} \acute{\theta} & \text{avec une probabilité } \alpha(\acute{\theta}, \theta^{(t-1)}) \\ \theta^{(t-1)} & \text{avec une probabilité } 1 - \alpha(\acute{\theta}, \theta^{(t-1)}) \end{cases}$$

où

$$\alpha(\hat{\theta}, \theta^{(t-1)}) = \min \left( \frac{\pi(\hat{\theta}) q(\theta^{(t-1)} | \hat{\theta})}{\pi(\theta^{(t-1)}) q(\hat{\theta} | \theta^{(t-1)})}, 1 \right).$$

La loi densité  $\pi(\theta)$  est souvent appelée *loi cible* ou *loi objet*, tandis que la loi de densité dite loi de proposition  $q(\cdot | \theta)$ . Une propriété stupéfiante de cet algorithme est d'autoriser un nombre infini de lois de proposition produisant toute une *chaîne de Markov* convergeant vers la loi d'intérêt.

### L'échantillonnage de Gibbs

Le second groupe de méthodes de *Monte Carlo par chaînes de Markov* (MCMC) est encore appelé *échantillonnage de Gibbs*. Il ne demande pas de mettre en place une fonction d'exploration de l'espace des états de la nature. De plus, algorithme de Gibbs pour l'estimation et construction de modèle par conditionnement probabilité.

De fait, les méthodes de *Gibbs* utilisent plus complètement que ne le font les méthodes de *Hasting-Metropolis*, les structures conditionnelles des modèles.

Pour  $(\theta_1, \dots, \theta_p)$ , on, veut simuler  $\pi(\theta)$  à partir de  $\pi_i(\theta_i | \theta_{(-i)}) = \pi_i(\theta_i | \theta_j, j \neq i)$  pour tout  $i$ . On initialise avec  $\theta^{(0)}$  et à l'instant  $t$ , à écrit :

$$\left\{ \begin{array}{l} \left( \theta_1^{(t)} | \theta^{(t-1)} \right) \sim \pi_1 \left( \theta_1^{(t)} | \theta_{(-1)}^{(t-1)} \right) \\ \left( \theta_2^{(t)} | \theta^{(t-1)}, \theta_1^{(t)} \right) \sim \pi_2 \left( \theta_2^{(t)} | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)} \right) \\ \cdot \\ \cdot \\ \left( \theta_p^{(t)} | \theta^{(t-1)}, \theta_{(-p)}^{(t)} \right) \sim \pi_p \left( \theta_p^{(t)} | \theta_{(-p)}^{(t-1)} \right) \end{array} \right.$$

Dans le cas où une telle loi  $\pi$  existe,  $\theta^{(t)}$  issu de cet algorithme est une *chaînes de Markov ergodique de loi stationnaire*  $\pi$ .

## 2.8 Inférence bayésienne sur les valeurs extrêmes

Il y a un certain nombre de raisons pour lesquelles une analyse bayésienne de valeur extrême des données pourraient être souhaitables. D'abord et avant tout, en raison de la rareté des données, la possibilité d'inclure d'autres sources d'information par une distribution préalable a un attrait évident, deuxièmement le résultat d'une analyse bayésienne la distribution a posteriori fournit une inférence plus complète que l'analyse du maximum de vraisemblance correspondant.

En particulier, de puis l'objectif d'une analyse des valeurs extrêmes est généralement une estimation de la probabilité que les événements futurs atteignent des niveaux extrêmes, l'expression à travers des distributions prédictives est naturelle.

On pose, un modèle adapté pour le maximum annuel  $z$  d'un processus est  $z \sim GEV(\mu, \sigma, \zeta)$ . estimation de  $\theta = (\mu, \sigma, \zeta)$  cela peut être fait sur la base des notes quotidiennes précédentes  $X = (x_1, \dots, x_n)$  en utilisant bayésienne, le résultat est une distribution postérieure  $f(\theta | x)$ .

Une félicitation préalable dans une analyse de valeur extrême est plus raisonnablement atteint en termes de quantiles extrêmes d'un processus, plutôt que les paramètres du modèle de valeur extrême eux-mêmes. Sous réserve d'auto-cohérence, un la distribution a priori sur un ensemble de trois quantiles peut toujours être transformée en une distribution a priori sur  $(\mu, \sigma, \zeta)$ . Les algorithmes *MCMC* standard peuvent alors être appliqué pour obtenir des réalisations à partir de la distribution postérieure correspondante qui, à son tour, peut être utilisé pour estimer la distribution prédictive de, disons, la distribution annuelle maximale à travers :

$$\Pr \{Z \leq z | x_1, \dots, x_n\} \simeq \frac{1}{s} \sum_{i=1}^s \Pr \{Z \leq z | \theta_i\}.$$

### 2.8.1 Inférence bayésienne sur les paramètres d'un modèle GEV

Le processus stochastique a temps discret est divisé en  $k$  blocs générant une série de maxima  $Z_1, \dots, Z_k$ . Puisque les populations sous-jacentes sont indépendantes (hypothèse i.i.d), ces maxima le sont aussi et, pourvu que la taille des blocs soit assez grande, on peut considérer qu'ils sont identiquement distribués selon le modèle *GEV*. Les modèle GEV avec les paramètres  $\rho_0 = \sigma^{-1} > 0$ ,  $\beta_0 = -\rho_0\zeta_0$  s'écrit :

$$GEV : \quad \Pr(Z \leq z \mid \beta_0, \mu, \rho_0) = \exp[-(1 - \beta_0(z - \mu))^{\rho_0/\beta_0}]. \quad (2.1)$$

#### La distribution conjointe a posteriori

Le modèle *GEV* [2.1](#) est donc caractérisé par le paramètre  $\theta = (\beta_0, \mu, \rho_0)$  et la densité de probabilité correspondante s'écrit

$$[z \mid \theta] = \rho_0(1 - \beta_0(z - \mu))^{\rho_0/\beta_0 - 1} G(z \mid \theta)$$

où  $\rho_0 \in \mathbb{R}_0^+$ ,  $\beta_0 \in \mathbb{R}_0$ ,  $\mu \in \mathbb{R}$  et  $\beta_0(z - \mu) < 1$ .

L'hypothèse i.i.d entraîne la vraisemblance d'un  $k$  échantillon de maxima :

$$[z_1, \dots, z_k \mid \theta] = \rho_0^k \prod_{i=1}^k \{(1 - \beta_0(z_i - \mu))^{\rho_0/\beta_0 - 1} G(z_i \mid \theta)\}.$$

Pour le Prior, on postulera l'indépendance des composantes du vecteur  $\theta$ .

$$\beta_0 \perp \mu \perp \rho_0$$

Et un Prior non informatif simple a la forme suivante :

$$[\theta] = [\beta_0][\mu][\rho_0] \propto \frac{1}{\rho_0}$$

Pour le construire, nous avons pris  $\beta_0$  et  $\mu$  uniformes sur un domaine assez grand. Pour le paramètre d'échelle  $\rho_0 > 0$ , le Prior habituel est une distribution gamma dont les paramètres tendent vers zéro.

$$[\rho_0 \mid a, b] \propto \rho_0^{a-1} \exp(-b\rho_0) \xrightarrow{a,b \rightarrow 0} \frac{1}{\rho_0}.$$

L'application de la règle de *Bayes* nous donne le posteriori non normalisé

$$[\theta \mid z_1, \dots, z_k] \propto \rho_0^{k-1} \prod_{i=1}^k \{(1 - \beta_0(z_i - \mu))^{\rho_0/\beta_0 - 1} G(z_i \mid \theta)\}.$$

La normalisation par calcul intégral n'est pas possible et aucune conditionnelle n'est standard. L'inférence peut se faire via un algorithme de *Metropolis-Hastings*

### Algorithme MH séquentiel appliqué au modèle GEV

Il sera commode de poser :

$$f(\beta_0, \mu, \rho_0) = \rho_0^{k-1} \prod_{i=1}^k \{[1 - \beta_0(z_i - \mu)]^{\rho_0/\beta_0 - 1} G(z_i \mid \theta)\}.$$

Puisque  $\rho_0 > 0$ , l'algorithme *MH* est plus facile à mettre en œuvre avec le changement de paramètre suivant :

$$\phi = \ln \rho_0 \Leftrightarrow \rho_0 = e^\phi$$

La transformation logarithmique donne :

$$\ln f(\beta_0, \mu, \rho_0) = (k-1)\phi + \left(\frac{e^\phi}{\beta_0} - 1\right) \sum_{i=1}^k \ln[1 - \beta_0(z_i - \mu)] - \sum_{i=1}^k [1 - \beta_0(z_i - \mu)]^{e^\phi/\beta_0}.$$

Où

$$\beta_0(z_i - \mu) < 1; \quad i = 1, \dots, k$$

## 2.8.2 Inférence bayésienne sur les paramètres d'un modèle POT

Soit un processus stochastique à temps discret  $\{X_t\}$  et soit un seuil  $c > 0$  fixé. On s'intéresse aux dépassements du seuil. Leur nombre est distribué selon une loi de Poisson. Pourvu que le seuil soit choisi assez haut, les dépassements sont i.i.d selon une loi de Pareto généralisée. Avec le nouveau paramétrage, la densité au dépassement  $y_t$  s'écrit :

$$[y_t | \eta, \beta] = \rho (1 - \beta y_t)^{\rho/\beta - 1} \quad (2.2)$$

Et vérifie les conditions :

$$\beta \in \mathbb{R}, \quad \rho > 0, \quad \forall t : 1 - \beta y_t > 0 \quad (2.3)$$

Le modèle *POT* avec les paramètres  $\rho = \eta^{-1} > 0$ ,  $\beta = -\rho\zeta$  s'écrit :

$$POT : \quad \Pr(Z \leq z | \beta, \lambda, \rho) = \exp[-\lambda(1 - \beta(z - c))^{\rho/\beta}].$$

### Distribution conjointe a posteriori et inférence

Le respect de la condition [2.3](#) fait que la densité [2.2](#) peut se mettre sous la forme suivante :

$$[y_t | \eta, \beta] = \rho \exp \left( (\rho - \beta) \frac{1}{\beta} \ln(1 - \beta y_t) \right)$$

Le seuil  $c$  étant convenablement fixé, la vraisemblance d'un  $k$  échantillon i.i.d de dépassements s'écrit :

$$[y_t | \rho, \beta] = \rho^k \exp \{ (\rho - \beta) S(\beta) \}$$

Où on a posé :

$$S(\beta) = \frac{1}{\beta} \sum_{t=1}^k \ln(1 - \beta y_t); \quad \beta < \frac{1}{\max(y_t)} \equiv b \quad (2.4)$$

Soit  $k$  v.a discrète nombre de dépassements sur une fenêtre de longueur  $L$ . Les hypothèses



d'indépendance permettent d'écrire la vraisemblance totale :

$$[Y_1 = y_1, \dots, Y_k = y_k, K = k] = [K = k] \prod_{t=1}^k [Y_t = y_t]$$

C'est-à-dire

$$[y_1, \dots, y_k, k \mid \lambda, \rho, \beta, L] = \exp \frac{(-\lambda L)^k}{k!} \rho^k \exp \{(\rho - \beta) S(\beta)\}.$$

Bien que le seuil  $c$  n'apparaisse pas dans le conditionnement, il ne faut pas oublier que le paramètre de Poisson,  $\lambda$ , qui représente la valeur attendue du nombre de dépassements sur une fenêtre unitaire, dépend du seuil  $c$ .

Dans la plupart des applications pratiques, on postule l'indépendance des paramètres :  $\beta \perp \lambda \perp \rho$ . Un Prior non informatif a la forme suivante puisque le paramètre  $\beta$  est alors uniforme sur  $\mathbb{R}_0$  :

$$[\rho, \beta, \lambda] \propto \frac{1}{\lambda \rho}$$

La loi conjointe a posteriori suit par la règle de *Bayes* :

$$[\lambda, \rho, \beta \mid y_1, \dots, y_k, k, L] \propto \lambda^{k-1} \exp(-\lambda L) \rho^{k-1} \exp(\rho S(\beta)) \exp(-\beta S(\beta))$$

De cette loi conjointe on peut tirer quelques propriétés intéressantes.

- Pour le paramètre de Poisson,  $\lambda$ , on reconnaît la signature fonctionnelle d'une distribution gamma de paramètre de forme  $k$  (nombre de dépassements observés) et de paramètre d'échelle  $L$  (longueur de la fenêtre). Ce paramètre est indépendant des paramètres  $\beta$  et  $\rho$  et la loi marginale est donc confondue avec la conditionnelle complète :

$$\lambda \mid k, L \sim \text{dgamma}(\lambda \mid k, L)$$

- Pour le paramètre d'échelle,  $\rho$ , on reconnaît la signature fonctionnelle d'une distribution

gamma de paramètre de forme  $k$  et de paramètre d'échelle  $-S(\beta)$  2.4 :

$$\rho \mid y_1, \dots, y_k, k, \beta \sim \text{dgamma}(\rho \mid k, -S(\beta))$$

Cette conditionnelle complète est, de fait, indépendante de  $\lambda$  mais pas de  $\beta$ . Cette propriété est caractéristique du modèle *POT*.

- La conditionnelle complète de  $\beta$  s'écrit :

$$[\beta \mid y_1, \dots, y_k, k] \propto \exp((\rho - \beta)S(\beta))$$

La dépendance de  $\beta$  avec  $\rho$  doit être notée car le paramètre  $\beta$  n'est pas un paramètre adimensionnel puisque sa dimension est proportionnelle à celle de  $\rho$  par le changement de paramètre  $\beta = \rho/\zeta$  où généralement le paramètre de forme  $\zeta$  est considéré comme adimensionnel.

- Enfin, un résultat complémentaire utile est la forme de la distribution marginale de  $\beta$  obtenue par intégration de la loi conjointe  $[\rho, \beta \mid X, , k, c, \tau]$  soit :

$$[\beta \mid y_1, \dots, y_k, k, L] \propto \frac{\exp(-\beta S(\beta))}{(S(\beta))^k}$$

Compte tenu de ces propriétés, l'inférence bayésienne peut être entreprise de diverses façons.

On peut réaliser une simulation directe de la distribution a posteriori via I enchaînement :

$$[\beta \mid X, , k, c, \tau] \rightarrow \text{dgamma}(\rho \mid k, -S(\beta, X, c)) \rightarrow \text{dgamma}(\lambda \mid k, L)$$

Seule la simulation initiale de  $\beta$  dont la densité n'est pas de forme analytique connue peut être délicate. On peut réaliser un échantillonnage de Gibbs puisque les conditionnelles complètes sont connues.

**Exemple 2.8.1** On a le modèle GEV :

$$G(x | \mu, \rho_0, \beta_0) = \exp \left[ -(1 - \beta_0(z - \mu))^{\frac{\rho_0}{\beta_0}} \right]$$

est caractérisé par le paramètre  $\theta = (\beta_0, \mu, \rho_0)$  est sa densité de probabilité s'écrit

$$[x | \theta] = \rho_0(1 - \beta_0(z - \mu))^{\frac{\rho_0}{\beta_0} - 1} G(x | \theta)$$

La densité de maxima de vraisemblance d'un  $k$  échantillon :

$$[x_1, \dots, x_k | \theta] = \rho_0^k \prod_{i=1}^k [1 - \beta_0(x_i - \mu)]^{\frac{\rho_0}{\beta_0} - 1} G(x | \theta)$$

Pour la loi a priori, On postulera l'indépendance des composantes du vecteur  $\theta(\beta_0 \perp \mu \perp \rho_0)$  telle que :

$$\beta_0 \sim U_{[a,b]} \quad \mu \sim U_{[a,b]} \quad \rho_0 \sim \text{Gamma}(a, b)$$

Dont les paramètres tendent vers zéro.

$$[\rho_0 | a, b] \propto \rho_0^{a-1} \exp(-b\rho_0) \longmapsto \frac{1}{\rho_0}$$

Donc : la loi a priori s'écrit de la forme suivante :

$$[\theta] = [\beta_0][\mu][\rho_0] \propto \frac{1}{\rho_0}$$

La loi a posteriori non normalisé est donnée par :

$$[\theta | x_1, \dots, x_k] \propto \rho_0^{k-1} \prod_{i=1}^k [1 - \beta_0(x_i - \mu)]^{\frac{\rho_0}{\beta_0} - 1} G(x_i | \theta).$$

# Chapitre 3

## Simulation et résultats numériques

Dans ce chapitre, nous examinerons deux modèles. Le premier modèle qui étudie le phénomène des crues dans les bassins versants secs à semi-arides de l'oued Abiud dans la région de Biskra. Ils ont été examinés par la méthode d'estimation de la masse annuelle (*GEV*) et à l'aide de deux méthodes d'estimation *MV* et *MH*. Le deuxième module étudie les techniques de mesure des risques d'assurance.

### 3.1 Application Numérique

#### 3.1.1 L'approche *GEV*

Les principaux objectifs de cette partie sont :

1. Statistiques descriptives des données.
2. Estimation des paramètres.
3. Anticiper les retours futurs en plus de la période de retour et du niveau de retour.

L'expression de modèle *GEV* dépend du paramétrage choisi :

$$GEV(\mu, \sigma, \zeta) \longleftrightarrow GEV(\mu, \rho = \sigma^{-1}, \beta = -\zeta\sigma^{-1})$$

On utilisés le critère d'information Bayésien. Dans les applications suivantes, l'estimation ponctuelle de leurs paramètres est réalises avec le paramétrage initial. Comme nous l'avons dit, le second paramétrage facilite l'inférence bayésienne : algorithme *MH* pour *GEV*.

### Méthodes d'estimation et outils

Deux méthodes principales : Maximum de vraisemblance et Méthode des moments.

### Inférence bayésienne sur les paramètres d'un modèle GEV

Les séries *POT* ont approximativement une distribution de la famille de Pareto généralisée (*GPD*).

La fonction de répartition de la loi *GEV* s'écrit comme suit :

$$F(x|\mu, \sigma, \zeta) = \exp\left(-\left\{1 + \zeta \left(\frac{x - \mu}{\sigma}\right)\right\}^{-\frac{1}{\zeta}}\right)$$

Avec :

$\mu$  : Paramètre de position ou de *localisation*.

$\sigma$  : Paramètre *d'échelle* ( $\sigma > 0$ ).

$\zeta$  : Paramètre de *forme*.

Soit, en termes de *période de retour* :

$$1 - \frac{1}{T_x(x)} = \exp\left[-\frac{1}{T_x(x)}\right].$$

### Zone d'étude et données

#### Présentation de la zone d'étude

L'oued *Abiod* s'étale du massif des Aurès (monts Chélia) jusqu'aux gorges de *Foum el Gherza*. Le bassin versant est caractérisé par sa dissymétrie, il englobe une partie montagneuse au Nord à plus de 2000 *m* et une autre basse au Sud (295*m* à El Habel). S'étendant

sur une superficie d'environ  $1300 \text{ Km}^2$  circonscrite par un périmètre de  $200 \text{ Km}$ . La longueur de oued Abiod est de  $85 \text{ km}$ , formé par la réunion des torrents descendants des raides pentes de Chelia ( $2326 \text{ m}$ ) et Ichemoul ( $2100 \text{ m}$ ). L'indice de compacité de Gravélius de  $1.59$  reflète la forme allongée de bassin versant de La vallée de oued qui est principalement composé de roches sédimentaires. L'oued Abiod est caractérisé par deux régions climatiques : les Aurès où la moyenne des précipitations est de  $450 \text{ mm/an}$ , et les plaines sahariennes avec une moyenne de précipitation de  $150 \text{ mm/an}$ . Le climat du bassin versant est un climat aride à semi-aride.

### Description des données

On commence par l'analyse descriptive des données :

$n$	$Min$	$1st \text{ Qu}$	$Median$	$Mean$	$3rd \text{ Qu}$	$Max$
123	185.0	259.5	440.6	568.5	660.1	2152.0

TAB. 3.1 – Statistiques descriptives des données

On remarque que les données est entre  $185.0 \text{ (Hm3)}$  et  $2152.0 \text{ (Hm3)}$  mais la moyenne est  $440.6 \text{ (Hm3)}$ . Dans ce cas la moyenne n'est pas représentative. Donc on parle des extrêmes.

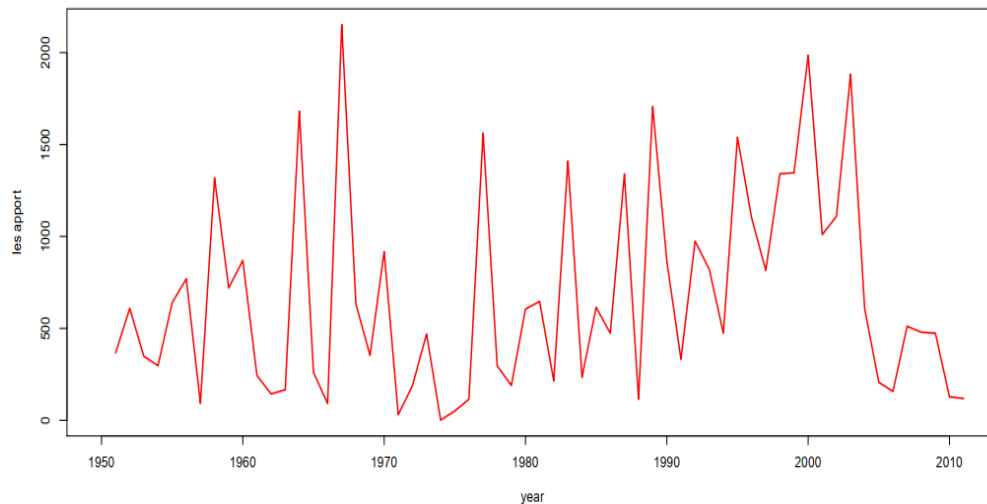


FIG. 3.1 – Niveaux maximaux annuels et de rendement pour Oued Abiod.

### Estimation du modèle GEV

La distribution des valeurs extrêmes généralisée modélise le comportement du maximum d'un échantillon.

Pour déterminer la loi et estimer ces paramètres, on utilise la méthode du "*maximum par blocs*" qui consiste à construire un échantillon de maximum à partir d'un échantillon de données en formant des blocs de même dimension. A partir 123 observation on a formé 62 blocs chaque bloc contient un seul maximum. La statistique descriptive du nouveau tableau est

<i>n</i>	<i>Min</i>	<i>1st Qu</i>	<i>Median</i>	<i>Mean</i>	<i>3rd Qu</i>	<i>Max</i>
62	1.11	214.28	511.53	675.70	974.55	2151.98

TAB. 3.2 – Statistiques descriptives des blocs

On remarque que les données est entre  $-Inf (Hm3)$  et  $2152.0(Hm3)$  mais la moyenne est  $675.70 (Hm3)$ . Dans ce cas la moyenne n'est pas représentative. Donc on parle des extrêmes.

### Intervalle de confiance

<i>Methode</i>	<i>Moy</i>	<i>IC</i>
<i>MLE</i>	529.75	[526.570, 532.952]
<i>Bayesian</i>		1596.1

TAB. 3.3 – Intervalle de confiance entre les deux estimations

### Estimation des paramètres on utilisons la méthode MLE

Les résultats selon le langage R sont comme suit :

	<i>Méthode MLE</i>	
	<i>Estimâtes</i>	<i>Std.err</i>
<i>Location</i>	1980.987	2.293
<i>Scale</i>	16.684	1.758
<i>shape</i>	-0.488	0.077

TAB. 3.4 – Estimations ML des parametres du modele GEV pour GEV.

### Estimation des paramètres on utilisons la méthode Bayesian

Les résultats selon le langage R sont comme suit :

	<i>Méthode Bayesian</i>
<i>Location</i>	
<i>log.Scale</i>	<i>0.4861972</i>
<i>shape</i>	<i>0.4216843</i>

TAB. 3.5 – Estimations Bayesian des parametres du modele GEV .

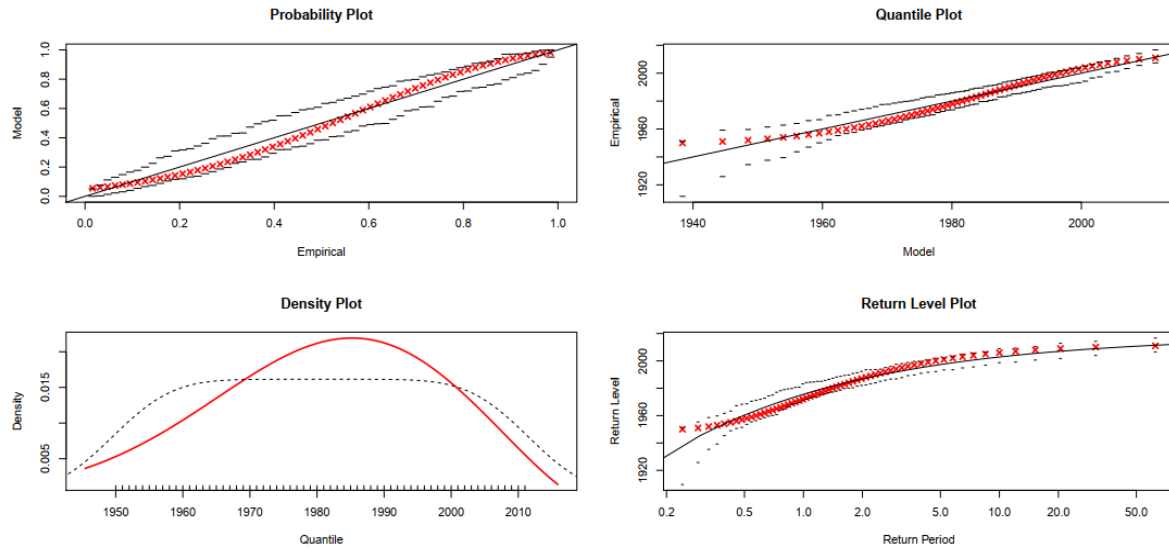


FIG. 3.2 – Diagnostique graphique de Modèle.

### Intervalles de confiance

	$IC_{2.5\%}$	<i>Posterior Mean</i>	$IC_{97.5\%}$
<i>Location</i>	1970.0422255	1972.6760341	1974.9800587
<i>scale</i>	15.3370542	19.1886126	23.9802443
<i>shape</i>	-0.4915251	-0.3408912	-0.1745042

TAB. 3.6 – Intervalle de confiance pour les estimateurs (GEV).

Remarquera que  $\zeta$  n'est pas significativement différent de zéro (cas de Gumbel) et que l'estimation ponctuelle de la valeur en pourcentage du niveau d'eau est la moyenne de la distribution marginale postérieure.

### Période de retour et niveau de retour

Dans cette partie on va donner des prévisions à court terme et à long terme pour la période de retour et le niveau de retour associé.



Niveau de retour (quantile)  $z_p$  correspondant à une période de retour  $1/p$  :

$$z_p = G_{\zeta}^{-1}(1 - p) = \begin{cases} \mu - \left[1 - \{-\log(1 - p)\}^{-\zeta}\right] \frac{\sigma}{\zeta} & \text{si } \zeta \neq 0 \\ \mu - \sigma \log(-\log(1 - p)) & \text{si } \zeta = 0 \end{cases}$$

Les résultats obtenus sont les suivant :

<i>ans</i>	<i>MLE</i>			<i>Bayésien</i>		
	<i>level</i>	<i>Mean</i>	<i>IC</i>	<i>level</i>	<i>Mean</i>	<i>IC</i>
5	1997.081	1997.081	[1992.987, 2001.175]	1993.312	1994.588	[1991.193, 1998.504]
10	2003.024	2003.024	[1999.188, 2006.860]	1999.717	2000.511	[1996.837, 2005.011]
50	2010.930	2010.930	[2006.875, 2014.985]	2008.728	2008.361	[2004.436, 2015.159]
100	2012.904	2012.904	[2008.462, 2017.347]	2011.123	2010.329	[2006.318, 2018.220]

TAB. 3.7 – Estimation des paramètres des valeurs extrêmes par la méthode de MV et "Bayésien" pour La modèle GEV.

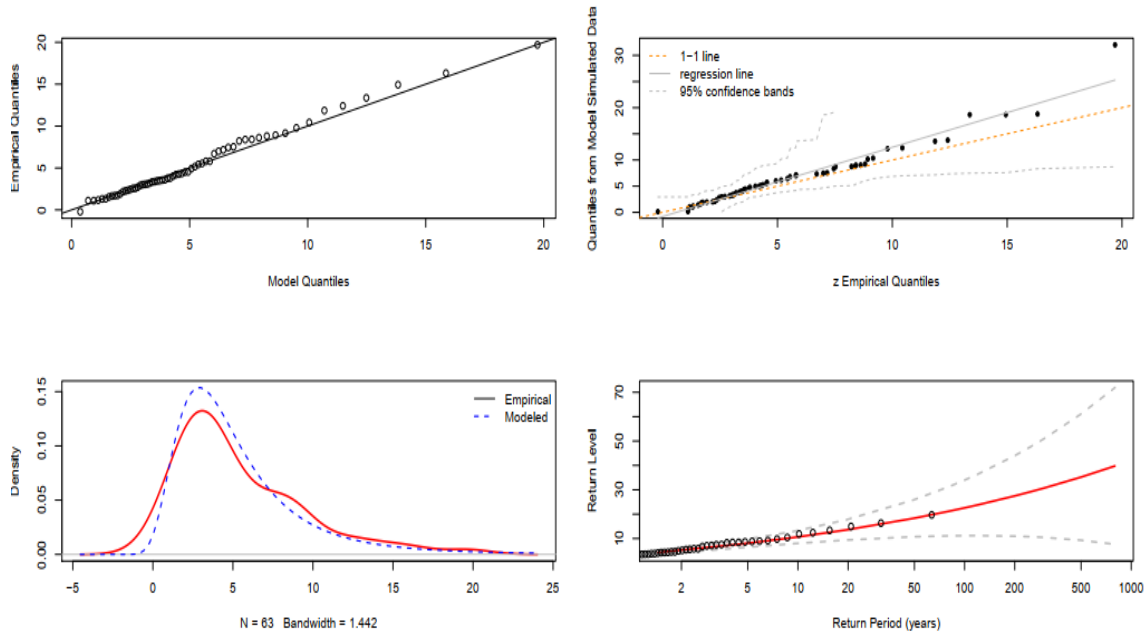


FIG. 3.3 – Réalisations des paramètres de la GEV.

### 3.1.2 Mesures de risque

#### *Présentation de la compagnie d'assurance AXA*

Axa est un groupe international français spécialisé dans l'assurance, et dans la gestion d'actifs depuis 1994. Il est le numéro deux de l'assurance dans le monde en termes de chiffre d'affaires.

#### **Données :**

Nous avons téléchargé les cours de bourse sur Internet au format *TXT*.

(URL : <http://www.abcbourse.com> ). Les cours quotidiens sont observés sur la période du 25/10/2007 au 23/10/2009. À partir de ces valeurs, on calcule le rendement associé :

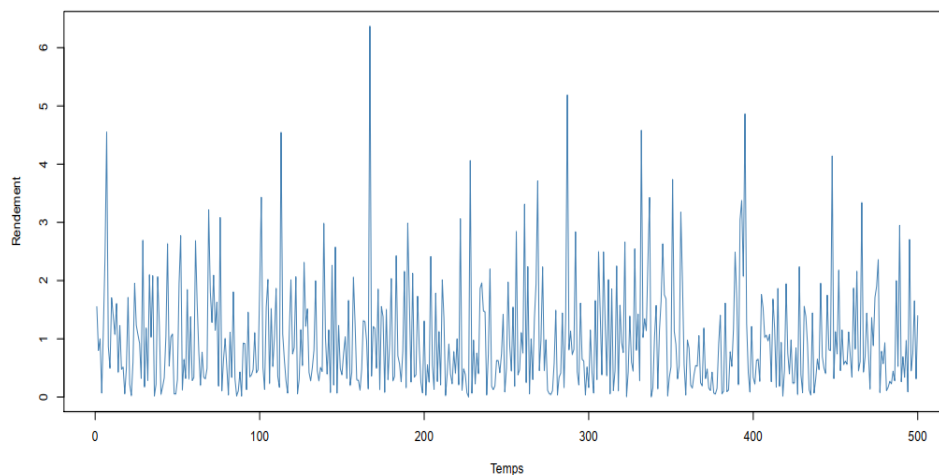


FIG. 3.4 – Rendements journalière de la compagnie d'assurance AXA.

L'objectif principal de cette étude est d'appliquer la simulation historique pour le calcul de la  $VaR$  et  $CVaR$ . En assurance la loi des pertes est définie sur  $[0, \infty]$  et représente des pertes associées a des réclamations.

#### **La statistique des rendements d'AXA :**

Le tableau ci-dessus donne la statistique des rendements d'AXA  $n = 510$  :

Graphique des enregistrements de sous-échantillon le tracé QQ sont utilisés pour déterminer si les retours Le cours de l'action AXA satisfait à l'hypothèse de normalité.

	<i>Rendements d'AXA</i>
<i>Min</i>	0
<i>1st Qu</i>	-0.022472856
<i>Median</i>	0.00878483
<i>Mean</i>	-0.000557258
<i>3rd Qu</i>	-0.000969462
<i>Max</i>	-0.001112553

TAB. 3.8 – Statistique de série des rendements

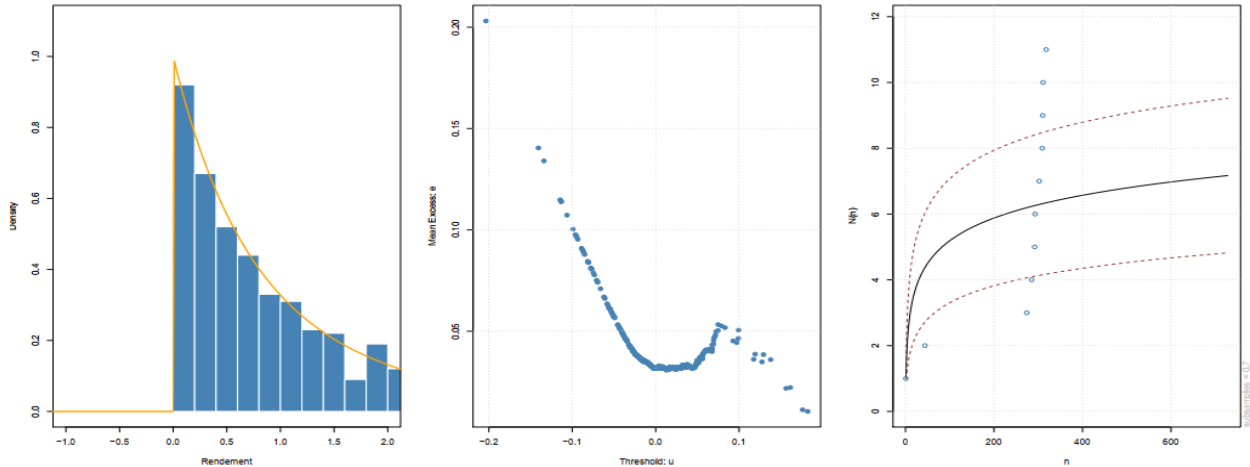


FIG. 3.5 – Test graphique de la normalité de série des rendements par QQplot et histogramme.

### Calcul de la VaR et la CVaR historique

Le tableau suivant donne les valeurs de la *VaR* et *CVaR* au niveau de confiance entre 1% et 10% :

$\alpha$	<i>VaR</i>	<i>CVaR</i>
1%	-0.1377922	-0.1730751
2.5%	-0.09715574	-0.1389327
5%	-0.06758266	-0.1067554
10%	-0.04795439	-0.08088358

TAB. 3.9 – Comparaison de la VaR et de la CVaR en termes de rendement aux niveaux de confiance différents.

## Comparaison entre la VaR et la CVaR

La *VaR* est imposé par les régulateurs comme une mesure plus pertinente, tandis que de nombreux statisticiens recommandent la surveillance de la *VaR* comme mesure alternative au risque. Ainsi, nous résumons les caractéristiques spécifiques de chaque échelle selon le modèle étudié et à partir du tableau comme suit :

Mathématiquement, la *CVaR* est supérieure ou égale à la *VaR*, par exemple pour :  $\alpha = 1\%$  :

$$\widehat{VaR}_\alpha = -0.1377922 \leq \widehat{CVaR}_\alpha = -0.1730751.$$

Mais la *CVaR* est une mesure fixe alors que la *VaR* ne l'est pas. Nous concluons donc ce qui suit :

### *VaR*

- Une représentation simple et pratique du risque de perte.
- La plupart des investisseurs utilisent une mesure standard du risque de perte.
- Jauge de risque de baisse.
- Elle s'applique aux distributions non linéaires.
- Ne mesurez pas les pertes au-delà de la valeur à risque.

### *CVaR*

- Représentation adéquate et simple du risque de perte.
- L'échelle de perte recommandée est supérieure à la valeur à risque.
- La meilleure mesure du risque de baisse.
- Elle s'applique aux distributions non linéaires.
- Il n'est pas facile d'appliquer des méthodes de backtesting aux tests de validation *CvaR*.
- Plus prudent que la *VaR* : mesure le risque qui dépasse la *VaR*.

Nombreux outils développés dans des packages *R* : *ismev*, *extRemes*, *evd*, *fExtremes*, *Xtremes*, *EXTREMES*, *POT*, *OpVaR*, *cvar*, *Risk*.

# Conclusion

Ce mémoire cherche à étudier l'utilité de l'approche de Bayes pour déduire des modèles de valeurs extrêmes, deux classes (*GEV*) et (*GPD*) de distributions de valeurs extrêmes sont utilisées pour trouver la distribution limite correcte pour les maxima et les minima. L'approche *ML* est généralement utilisée, pendant que l'approche *Bayesian* permet tant pour l'incertitude des paramètres et l'aspect aléatoire de paramètre des futures observations d'être incluses dans le procédé d'évaluation. Dans le cadre Bayesian, deux différentes façons de construire les distributions préalables sont examinées : Des distributions presque plates pour les paramètres du modèle et l'incorporation des informations des sites dans le voisinage de celui qui est considéré par la distribution des différences de quantile. Le cadre Bayésien offre des avantages substantiels pour l'analyse de valeurs extrêmes dans les deux cas. Il est prouvé que cette sélection d'échelle est avantageuse pour des variables naturellement limitées et améliore la capture du caractère "relatif" des extrêmes. Enfin, nous avons analysé certaines des métriques de risque les plus couramment utilisées telles que la *VaR*, la *TVaR* et *CVaR*.

# Bibliographie

- [1] **Betrlant.J. Geogebur.** Staistics of Extremes, Jhon Wiley and Sons, England, 2004.
- [2] **Bezaf. A et Nikeghbalt. A.** la théorie des valeurs extrêmes et la gestion des risques, ENSAE, France, 2000.
- [3] **Castillo.E et autres.** Extreme Value and related models with Applications in Engineering and sciences, Wiley inter science, 2004.
- [4] **Christian.P.R.** Le choix bayésien, deuxième édition, Springer, France, 2007.
- [5] **Coles. S.G et Powell E.A.** Bayesian methods in extreme value modelling : a review and new developments, International Statistical Review 64, 119-36, 1996.
- [6] **Coles, S.** An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics, 2001.
- [7] **Denuit, M., Dhaene, J., Govaerts, M. and Kaas, R..** Actuarial Theory for Dependent Risks : Measures, Orders and Models. Jhon Wiley & Sons, Ltd, 2005.
- [8] **Gençayr.R et autres,** Extreme value theory and value-at-risk : Relative performance in emerging markets, International Journal of Forecasting 20, p 287-303, 2004.
- [9] **Norman Fenton, Martin Neil,** Risk Assessment and Decision Analysis with Bayesian Networks. Chapman & Hall/CRC, p 637 / 661, 2019.
- [10] **Jean Jaque.B et Eric.P.** Pratique de Calcule Bayésien. Springer, Paris 2010.
- [11] **Smith.R.L et Goodman. D.J.** Bayesian risk analysis, risk books, London, 235-251, 2000.

# Annexe A : Abréviations et Notations

## Abréviations :

<i>GEV</i>	Distribution des Valeurs Extrêmes Généralisée.
<i>GPD</i>	Distribution de Paréto Généralisée.
<i>POT</i>	Peaks-Over-Threshold.
<i>MCMC</i>	Monté Carlo par Chaine de Markov.
<i>MH</i>	Métropolis-Hasting.
<i>VaR</i>	Valeur à Risk.
<i>ES</i>	Expected shortfull.
<i>CVaR</i>	Valeur en risque conditionnel.

## Notations

$\propto$	Proportionnel à.
$\pi(x), \pi(\theta x)$	A priori, A posteriori.
$f(\theta x)$	Loi de probabilité, indexée par le paramètre $\theta$ .
$\gamma$	Paramètre de forme.
$\beta$	Paramètre de position.
$\lambda$	Paramètre d'échelle.
$\varrho$	Mesure de risque.
$\zeta$	Indice des valeurs extrêmes.

## Résumé

Les problèmes des valeurs extrêmes sont caractérisés par la pénurie des données et par l'exigence de modélisation où les Ce fait devient un vrai dilemme quand on aborde l'inférence d'une manière Bayésienne: la valeur des renseignements antérieurs est probablement important, mais il n'est pas du tout certain qu'on peut bien appliquer ces renseignements aux valeurs extrêmes. Dans ce mémoire, nous avons analysé certaines des mesures de risque les plus couramment utilisées telles que la VaR, la TVaR et la CvaR, ainsi que nous passons en revue la littérature qui relie les thèmes de l'analyse Bayésienne et de l'analyse des valeurs extrêmes, et utilisons des avancées récentes en outils informatiques Bayésiens pour juger la valeur d'analyse Bayésien des valeurs extrêmes en trois situations différents: premièrement, où un expert est disponible pour fournir des renseignements antérieurs; deuxièmement, où 'maximum de vraisemblance' échoue; et troisièmement, où on utilise des renseignements en espace sur des variables liées pour formuler un 'empirique Prior'.

Mots clés: Bayésienne, Valeur à risque, Méthode de Monte Carlo, Valeurs Extrêmes, MCMC.

## Abstract

The extreme value problems are characterized by the scarcity of data and by the requirement of modeling where the This fact becomes a real dilemma when approaching inference in a Bayesian way: the value of prior information is probably important, but it is not at all certain that this information can be applied well to extreme values. In this memory, we have analyzed some of the more commonly used risk measures such as VaR, TVaR and CvaR, as well as we review the literature that links the themes of Bayesian analysis and analysis of extreme values, and use recent advances in Bayesian computing tools to judge the Bayesian analysis value of extreme values in three different situations: first, where an expert is available to provide prior information; second, where 'maximum likelihood' fails; and third, using spatial information on related variables to formulate an 'empirical Prior'.

Key words: Bayesian, Value-at-Risk, Monte Carlo method, Extreme Value Theory, MCMC.

## ملخص

تتميز مشاكل القيم القصوى بندرة البيانات ومتطلبات النمذجة حيث تصبح هذه المشكلة معضلة حقيقية عند الاقتراب من الاستدلال بطريقة بايزية, ربما تكون قيمة المعلومات السابقة مهمة, لكن ليس مؤكد على الإطلاق أن هذه المعلومات يمكن تطبيقها بشكل جيد على القيم القصوى. في هذه المذكرة قمنا بتحليل بعض مقاييس المخاطر الأكثر استخدامًا مثل القيم المعرضة للمخاطر. قيمة الذيل المعرضة للخطر... بالإضافة إلى مراجعة الأدبيات التي تربط موضوعات التحليل البايزي وتحليل القيم المتطرفة, واستخدام التطورات الحديثة في حوسبة أدوات للحكم على قيمة تحليل بايزي للقيم المتطرفة في ثلاث حالات مختلفة: أولاً, عندما يكون الخبير متاحاً لتقييم المعلومات المسبقة؛ ثانياً, حيث تفشل "الاحتمالية القصوى"; وثالثاً, استخدام المعلومات المكائنية حول المتغيرات ذات الصلة لصياغة "السابقة التجريبية".

الكلمات المفتاحية: بايزي, القيمة المعرضة للخطر, طريقة مونت كارلو, القيم المتطرفة.