

N° d'ordre : 14/2017-D/MT

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DES SCIENCES ET TECHNOLOGIE DE HOUARI BOUMEDIENNE  
FACULTÉ DES MATHÉMATIQUES



THÈSE  
PRÉSENTÉE POUR L'OBTENTION DU GRADE DE DOCTEUR EN  
SCIENCES

EN MATHÉMATIQUES  
SPÉCIALITÉ : PROBABILITÉS & STATISTIQUE

Par : Abdelkader AMERAOUI

Sujet

Étude d'un modèle des valeurs extrêmes en  
présence de censure aléatoire à droite

Soutenue publiquement le 16/07/2017 à 10h00, devant le jury composé de :

Mme. Zohra GUESSOUM	Maître de Conférence A,	à l'USTHB / FMT	Présidente.
Mr. Kamal BOUKHETALA	Professeur,	à l'USTHB / FMT	Directeur de thèse.
Mr. Jean François DUPUY	Professeur,	à l'INSA (Rennes-France)	Co-directeur de thèse.
Mr. Abdelouahab LATRECHE	Professeur,	à l'E.N.S.S.E.A	Examineur.
Mr. Mohamed Riad REMITA	Professeur,	à l'Univ. d'Annaba	Examineur.
Mme. Ourida SADKI	Professeur,	à l'USTHB / FMT	Examinatrice.


# DÉDICACE



*À la mémoire de mon Père, et À ma chère Maman,  
À ma petite et grande famille,  
À mes amis et à tous ceux qui me sont chers.*



# REMERCIEMENTS

 Je tiens tout d'abord à exprimer ma profonde reconnaissance à mon directeur de thèse le Professeur Kamal BOUKHETALA et mon co-directeur de thèse le Professeur Jean François DUPUY (INSA de Rennes). Je leur adresse mes plus sincères remerciements pour la confiance qu'ils m'ont accordé, pour leurs disponibilité et leurs précieux conseils, et d'avoir permis à ce travail de voir le jour. Merci infiniment.

Je voudrais exprimer ma gratitude à Madame Zohra GUESSOUM (MCA, à l'USTHB) pour l'intérêt qu'elle a bien voulu porter à ce travail et avoir acceptée de présider mon jury.


Je suis très honoré que les professeurs Pr. Ourida SADKI (USTHB), Pr Abdelouahab LATRECHE (l'E.N.S.S.E.A) et Pr. Mohamed Riad REMITA (l'Univ. d'Annaba) aient accepté d'examiner ma thèse.

Toute ma reconnaissance aux programmes CNRS-DFRSU (Projet SESA, N° EDC25895), Projet CNEPRU (Modélisation Risque Actuariel, Assurance, Finance (MSAAF) - N° C00L03UN160420150004) qui ont contribué financièrement à ma thèse. Je remercie également l'institut Nationale de Sciences Appliquées (INSA) de Rennes ainsi que le service administratif de l'IRMAR - (UMR 6625).

Il ne serait pas possible de travailler à l'université sans eux : Un grand merci à mes enseignants d'hier, et collègues d'aujourd'hui, du département probabilités et statistiques de l'USTHB. Je tiens aussi à remercier tout le personnel administratif de la faculté des mathématiques.

Merci à mes amis : Arsalane, Souad, Meriem, Abderrahim, Amel, et toute l'équipe de recherche autour de Pr. BOUKHETALA, d'avoir passer de bons moments avec moi. Je dédie une mention très spéciale et toute particulière à ma "*femme*" surtout pour sa patience, puis à ma famille et à mes amis les plus proches.

# RÉSUMÉ

a présente thèse apporte quelques contributions à l'inférence statistique sur l'indice des valeurs extrêmes de distributions à queue épaisse, en présence d'information censurée aléatoirement à droite. Elle s'étale sur quatre chapitres, auxquels une introduction générale et une conclusion s'ajoutent. Le premier chapitre est consacré à l'introduction de la théorie des valeurs extrêmes univariée (TVEU), en rappelant des notions, des définitions ainsi que des concepts fondamentaux dans la dite théorie. Nous explorons dans un second chapitre, un ensemble de contributions statistiques des plus marquantes, liées à l'estimation de l'indice des valeurs extrêmes (EVI) noté  $\gamma$  ou encore l'indice de queue de distribution noté  $\alpha = 1/\gamma$ ; Une revue de la littérature mettant l'accent essentiellement sur les modèles d'inférence semi-paramétrique dans le cas d'un échantillon *i.i.d* complet est considérée, à l'instar des travaux de Hill (1975), Pickands (1975), Dekkers *et al.* (1989) ou de Beirlant *et al.* (1999). Une deuxième partie de ce chapitre focalise la lumière sur quelques estimateurs semi-paramétriques (qui font objet de références dans la littérature), adaptés au cas de données censurées aléatoirement à droite : Einmahl *et al.* (2008), Gomes and Neves (2011), ou encore Worms and Worms (2014). Un troisième chapitre rappelle un concept statistique robuste en la théorie bayésienne. Nous décrivons dans ce chapitre, les notions essentielles à la construction et la mise en œuvre d'un modèle bayésien. Nous présentons à la fin de ce chapitre, une contribution bayésienne à l'estimation de l'indice de queue  $\alpha$ , dans le cas d'un échantillon *i.i.d* complet, pour une distribution à queue lourde  $\alpha > 0$ , à l'idée de Ameraoui *et al.* (2016). Le dernier chapitre est dédié essentiellement à nos modestes contributions : nous mettons le point sur l'analyse de deux modèles dans la SVEU en présence de données (supposées être *i.i.d*) censurées aléatoirement à droite

- La première contribution est présentée sous la forme d'un article publié dans la revue *Computational Statistics & Data Analysis*, intitulé "Bayesian estimation of the tail index of a heavy tailed distribution under random censoring", Volume 104C (2016), pages 148–168 . ( <http://dx.doi.org/10.1016/j.csda.2016.06.009>).
- La deuxième contribution porte sur l'estimation d'un modèle composite présenté lors de la 4<sup>th</sup> Tunisian Society of Financial Studies Conference (Sousse, Tunis) du 08 au 10 décembre 2016, intitulée "Modeling Algerian cars claims using LPC model under right random censoring data". La contribution est acceptée pour publication au Journal of Business and Economics (ISSN 2155-7950), USA.

**Mots clés** : Théorie des valeurs extrêmes, censure aléatoire, estimation, inférence bayésienne, indice de queue, distribution à queue lourde, consistance de distributions a posteriori.

# ABSTRACT

Abstract

# ABRÉVIATIONS & NOTATIONS

Abréviations

# Table des matières

Dédicace	i
Remerciements	ii
Résumé	iii
Abstract	iv
Abréviations & Notations	v
Table des matières	vi
Table des figures	viii
Liste des tableaux	ix
<b>I INTRODUCTION À LA THÉORIE DES VALEURS EXTRÊMES</b>	<b>5</b>
I.1 Notions de statistiques d'ordre . . . . .	6
I.1.1 Distributions de statistiques d'ordre . . . . .	8
I.1.2 Existence des moments d'une statistique d'ordre . . . . .	9
I.1.3 Comportement limite des maxima . . . . .	9
I.2 Fondements de la théorie des valeurs extrêmes . . . . .	10
I.2.1 Caractérisation des max-domaines d'attraction . . . . .	17
I.2.2 Caractérisation du max-domaine d'attraction de Fréchet . . . . .	18
I.2.3 Caractérisation du max-domaine d'attraction de Weibull . . . . .	18
I.2.4 Caractérisation du max-domaine d'attraction de Gumbel . . . . .	18
I.3 Fluctuation des maxima . . . . .	19
I.3.1 Approche des maxima par bloc (ou GEV) . . . . .	19
I.3.2 Approche des excès au-delà d'un seuil (ou POT) . . . . .	20
I.3.3 Approche des moments pondérés . . . . .	21
I.3.4 Approche moyenne des excès au-delà d'un seuil . . . . .	22
<b>II INFÉRENCE STATISTIQUE FONDAMENTALE DANS LA THÉORIE DES VALEURS EXTRÊMES</b>	<b>27</b>
II.1 Modèles semi-paramétriques de premier, second et d'ordre élevé . . . . .	28
II.2 Estimation semi-paramétrique dans le cas d'un échantillon i.i.d complet . . . . .	28
II.2.1 Estimation de l'indice des valeurs extrêmes . . . . .	29
II.2.2 Estimateurs généralisés de l'indice des valeurs extrêmes . . . . .	34
II.2.3 Réduction du biais dans l'estimation de l'indice des valeurs extrêmes $\gamma$ . . . . .	36

II.2.4	Estimation du quantile extrême pour les distribution à queue lourde . . . . .	37
II.3	Estimation semi-paramétrique dans le cas d'un échantillon i.i.d en présence de censure aléatoire . . . . .	39
II.3.1	Présentation du cadre d'estimation . . . . .	39
II.3.2	Estimation de l'indice des valeurs extrêmes en présence de censure aléatoire à droite . . . . .	41
<b>III</b>	<b>NOTIONS DE STATISTIQUE BAYÉSIENNE</b> . . . . .	<b>47</b>
III.1	Approche bayésienne . . . . .	48
III.1.1	Le paradigme bayésien . . . . .	48
III.1.2	Fondement de la construction bayésienne . . . . .	49
III.2	Principes d'inférence bayésienne . . . . .	50
III.2.1	Fonction de perte et risque bayésien . . . . .	50
III.2.2	Estimateurs bayésiens ponctuels . . . . .	51
III.2.3	Exemple du contrôle de qualité . . . . .	51
III.3	Quel choix a priori? . . . . .	53
III.3.1	Distribution a priori informative . . . . .	53
III.3.2	Distribution a priori non-informative . . . . .	54
III.3.3	Loi a priori au sens de Jeffrey . . . . .	56
III.3.4	La loi de Laplace . . . . .	57
III.3.5	Loi a priori de Zellner . . . . .	57
III.4	Modèle bayésien hiérarchique . . . . .	58
III.5	Consistance de la distribution a posteriori . . . . .	58
III.5.1	Normalité asymptotique de la distribution a posteriori . . . . .	60
III.6	Estimation bayésienne de l'indice de queue $\alpha > 0$ . . . . .	61
III.6.1	Approche bayésienne informative . . . . .	62
III.6.2	Approche bayésienne non-informative . . . . .	63
III.6.3	Comparaison entre les estimateurs bayésiens de $\alpha$ . . . . .	65
III.6.4	Estimation de $\alpha$ par une approche MCMC . . . . .	66
<b>IV</b>	<b>QUELQUES CONTRIBUTIONS À L'ESTIMATION DE L'INDICE DE QUEUE <math>\alpha</math> POUR UNE DISTRIBUTION A QUEUE LOURDE EN PRÉSENCE DE CENSURE ALÉATOIRE À DROITE</b> . . . . .	<b>70</b>
IV.1	PREMIÈRE PARTIE : BAYESIAN ESTIMATION OF THE TAIL INDEX OF A HEAVY TAILED DISTRIBUTION UNDER RANDOM CENSORING . . . . .	70
IV.2	DEUXIÈME PARTIE : MODELING ALGERIAN CARS CLAIMS USING LPC MODEL UNDER RANDOMLY RIGHT-CENSORED DATA . . . . .	92
IV.2.1	Lognormal-Pareto composite model . . . . .	92
IV.2.2	LPC parameters estimation under random censoring . . . . .	93
IV.2.3	An application to Algerian cars claims data . . . . .	97
IV.2.4	Conclusion and perspectives . . . . .	98
	<b>Conclusion générale</b> . . . . .	<b>100</b>

# Table des figures

I.1	<b>À gauche :</b> Fonction de répartition du maximum $H_n(x)$ , <b>à droite :</b> Fonction de répartition du minimum $L_n(x)$ pour $n = 10$ (en gras) , 20, 50, 100, 1000 et 10000 respectivement pour une loi exponentielle $\mathcal{E}(1)$ (ligne 1), Log-logistique standard (ligne 2) et une loi uniforme $\mathcal{U}_{[0,1]}$ (ligne 3) . . . . .	11
I.2	<b>À gauche :</b> variation de la distribution $EVD_M(\gamma)$ en fonction de $\gamma$ , <b>à droite :</b> sa densité de probabilité $h_\gamma(x)$ – ligne en bleu $\gamma > 0$ , en rouge $\gamma = 0$ , en vert $\gamma < 0$ . . . . .	13
I.3	<b>À gauche :</b> variation de la distribution $EVD_m(\gamma)$ en fonction de $\gamma$ , <b>à droite :</b> sa densité de probabilité $l_\gamma(x)$ – ligne en bleu $\gamma > 0$ , en rouge $\gamma = 0$ , en vert $\gamma < 0$ . . . . .	14
I.4	Fonction moyenne des excès empirique (noir) <i>v.s</i> théorique (rouge) de la loi $\mathcal{U}_{[0,1]}$ . . . . .	23
I.5	Fonction moyenne des excès empirique (noir) <i>v.s</i> théorique (rouge) de la loi $\mathcal{E}(1.5)$ . . . . .	24
I.6	Fonction moyenne des excès empirique (noir) <i>v.s</i> théorique (rouge) de la loi de Pareto $\mathcal{P}(0.5)$ . . . . .	25
I.7	Fonction moyenne des excès empirique relative aux données danish fire . . . . .	26
II.1	Comportement empirique de quelques estimateurs de $\gamma$ (en rouge l'estimateur de Hill, en vert l'estimateur de Pickand, en bleu l'estimateur de Dekker et en noir l'estimateur de Hill généralisé) - <b>ligne 1 :</b> Distribution Pareto(2), <b>ligne 2 :</b> Distribution exponentielle standard, <b>ligne 3 :</b> Distribution uniforme sur $[0,1]$ . . . . .	35
II.2	Comportement empirique des estimateurs en présence de censure - ligne 1 : une proportion de non-censure de 90%, ligne 2 : une proportion de non-censure de 75% et ligne 3 : une proportion de non-censure de 50%. . . . .	46
III.1	Effet des hyper-paramètres $a$ et $b$ sur la loi a posteriori . . . . .	55
III.2	Histogramme des réalisations de l'algorithme MH . . . . .	68
IV.1	<b>Line 1 :</b> QQ-plot of $m$ estimations of the scale parameter $\beta$ , <b>Line 2 :</b> QQ-plot of $m$ estimations of the shape parameter $\theta$ under simulated LPC(1.5, 10) right censored by : $\mathcal{E}(0.01)$ -left plot-, $\mathcal{E}(0.1)$ -right plot- . . . . .	96
IV.2	Estimated density for $\beta$ estimations, for $X \sim LPC(1.5, 10)$ right censored by : $Y \sim \mathcal{E}(0.01)$ -left plot-, $Y \sim \mathcal{E}(0.1)$ -right plot- . . . . .	97
IV.3	Log Quantile-Quantile plot of Algerian claims data for LPC (Black), Pareto (Green) and Lognormal (Pink) distribution. . . . .	98

# Liste des tableaux

I.1	Estimation du modèle POT (ou GPD) en fonction de $u$ , pour une loi $\mathcal{U}_{[0,1]}$ . . . . .	23
I.2	Estimation du modèle POT (ou GPD) en fonction de $u$ , pour une loi $\mathcal{E}(1.5)$ . . . . .	24
I.3	Estimation du modèle POT (ou GPD) en fonction de $u$ , pour une loi $\mathcal{P}(0.5)$ . . . . .	25
I.4	Estimation du modèle POT (ou GPD) en fonction de $u$ , pour les données Danish fire . . . . .	26
III.1	Exemples de lois conjuguées usuelles . . . . .	53
III.2	Estimation bayésienne de l'indice de queue $\alpha$ relativement à une distribution de Fréchet : $F(x) = \exp(-x^{-\alpha})$ , et une distribution de Pareto : $\bar{F}(x) = x^{-\alpha}$ . . . . .	67
IV.1	LPC( $\beta = 1.5, \theta = 5$ ) parameters estimation right censored by $\mathcal{E}(\lambda)$ distribution . . . . .	95
IV.2	LPC( $\beta = 1.5, \theta = 10$ ) parameters estimation right censored by $\mathcal{E}(\lambda)$ distribution . . . . .	95
IV.3	Estimated Parameter and Goodness-of-fit for Algerian data claims . . . . .	98

# INTRODUCTION GÉNÉRALE



## Généralités

La théorie statistique moderne est une des disciplines s'inspirant d'un raisonnement logique, conçu sur des bases mathématiques dont le traitement nécessite l'introduction d'un outil informatique de plus en plus puissant. Ce traitement de la donnée (qu'elle soit une donnée unitaire, agrégée, ou structurée selon des méta-données) est lié forcément au modèle mathématique et aux hypothèses (ou contraintes) supposées.

Lors de l'analyse statistique des données, un des constats qu'on peut observer est l'apparition de valeurs aberrantes par rapport au comportement moyen. Cette "éruption" peut-être due à une mauvaise mesure du phénomène, ou parfois (même le plus souvent) une réalisation naturelle du caractère étudié, engendrant ainsi des variations de très grandes amplitudes. La prise en charge d'une telle situation est devenue plus que nécessaire, par la conception d'outils statistiques performants et capable d'expliquer les observations à forte intensité.

Une nouvelle branche de la statistique, au nom de la théorie des valeurs extrêmes (TVE), a vu le jour, afin d'expliquer, de modéliser et de décrire la survenance d'évènements dits rares ou extrêmes, lorsqu'il s'agit de valeurs beaucoup plus grandes ou plus petites que celles observées habituellement. Ces évènements rares sont caractérisés par :

- Une faible probabilité d'occurrence mais relativement significative (l'importance des queues d'une distribution)
- Une longue période de retour d'un tel évènement
- Une forte intensité de l'évènement lorsqu'il se réalise

Ainsi, une des questions fondamentales dans la TVE, est de savoir comment peut-on évaluer la probabilité d'un évènement rare ? ou encore, comment mesure-t-on le risque associé à ces évènements extrêmes ?

Dans la nature, les évènements rares ou extrêmes sont souvent liés aux catastrophes naturelles (ouragans, tremblements de terre, crues ou inondations, ...) et leurs analyse apporte une aide à la décision à grande portée socio-économique, parfois à l'échelle des politiques d'états même.

Contrairement à la théorie statistique "classique", qui s'appuie en grande partie sur une analyse des valeurs centrales de l'échantillon (en donnant peu d'importance aux queues de la distribution), la théorie des valeurs extrêmes s'articule autour d'une caractérisation des queues, susceptible de pouvoir contenir des valeurs extrêmes. L'inférence dans la TVE est ainsi développée afin d'estimer au mieux la probabilité d'occurrence de tels évènements, ou encore leurs amplitudes et leurs périodes de retour, même si le nombre de fois où on observe ces évènements est petit. Cette théorie fut initiée par Fisher et Tippett en 1928, puis

développée par les apports de Gnedenko en 1943, en énonçant le comportement probable des extremums. On en trouve récemment dans la littérature, quelques nouvelles adaptations du théorème fondateur de la TVE (voir [Embrechts \*et al.\* \(1997\)](#)), ce qui montrent le développement autour de cette théorie, en terme de résultats théoriques, mais aussi en terme d'adaptation de la dite théorie dans différents domaines scientifiques :

1. **Climatologie** : en analyse et modélisation des vagues de froids et des canicules ([Shen \*et al.\* \(2010\)](#)), en observation et veille climatique ([Naveau \*et al.\* \(2005\)](#)), ou encore en analyse des précipitations ([Katz \(1999\)](#)).
2. **Hydrologie** : le dimensionnement des projets de canalisation hydrologique dans les zones urbaines, est lié aux phénomènes des crues extrêmes et aux séquences de pluies torrentielles exceptionnelles (voir à titre d'exemple le papier de [Blanchet \*et al.\* \(2016\)](#), ou [Katz \*et al.\* \(2002\)](#)).
3. **Assurances et finance** où la TVE intervient comme un outil performant dans la modélisation des risques extrêmes : forte fluctuation des cours boursiers impactée par une crise financière (crise des subprimes aux EU 2008), sinistralité corporelle en assurance automobile, actes terroristes contre un complexe industriel (Tiguentourine - Algérie, 2013). Voir aussi [Asimit \*et al.\* \(2016\)](#), [Lia \(2016\)](#).

## Présentation des travaux de thèse

L'objectif principal de cette thèse est d'apporter une approche statistique d'estimation ponctuelle de l'indice des queues, et l'application d'une telle contribution dans un contexte, où on suppose que les données sont censurées aléatoirement à droite. Notre contribution majeure est une modeste innovation dans l'estimation bayésienne de l'indice d'une distribution à queue lourde en présence d'information censurée aléatoirement à droite. Le développement de cette thèse est décrit comme suit :

### Chapitre I : " Introduction à la théorie des valeurs extrêmes "

Nous explorons dans ce chapitre, les généralités sur la TVE univariée permettant d'introduire les notions et les concepts nécessaires dans la suite de cette thèse : Nous rappelons tout d'abord le fondement de la TVE, en énonçant le théorème de Tippet-Fisher-Gnedenko, et une variante de ce théorème permettant de distinguer le max-domaine d'attraction d'un extremum selon la nature de la queue de distribution. Ces deux théorèmes sont énoncés par une paramétrisation d'une loi limite non-dégénérée, dépendante du dit indice des valeurs extrêmes (noté  $\gamma$ ), ou encore indice de queue de la distribution (noté  $\alpha$ ). Nous citons aussi, quelques approches statistiques d'estimation de l'indice  $\gamma$ , ainsi que l'application de la TVE à l'estimation de la notion des quantiles extrêmes.

### Chapitre II : " Inférence statistique fondamentale dans la théorie des valeurs extrêmes "

Ce chapitre est une revue de la littérature statistique, relative à l'estimation de l'indice des valeurs extrêmes  $\gamma$ , dans un contexte d'échantillon *i.i.d* complet, puis lorsque la présence de censure est considérée.

Une première partie est dédiée à l'estimation semi-paramétrique dans le cas d'un échantillon  $X_1, X_2, \dots, X_n$  supposées *i.i.d* et complet (sans censure de données). Nous rappelons des estimateurs (des plus répondus dans la littérature), ainsi que leurs propriétés asymptotiques : estimateur de [Hill \(1975\)](#), estimateur de [Pickands \(1975\)](#), estimateur des moments ([Dekkers \*et al.\*, 1989](#)), estimateur des excès au-delà d'un seuil aléatoire ([Davison , 1984](#)), ainsi que des estimateurs généralisés de l'indice des valeurs extrêmes ([Csörgő \*et al.\* \(1985\)](#), [Groenboom \*et al.\* \(2003\)](#), [Beirlant \*et al.\* \(1996\)](#) et [Fraga \*et al.\* \(2009\)](#)). Nous discutons aussi le problème de la réduction du biais dans l'estimation de l'indice des valeurs extrêmes  $\gamma$ , en citant les travaux de [Peng en 1998](#), [Beirlant \*et al.\* \(1999\)](#), [Gomes \*et al.\* \(2002\)](#), puis ceux de [Caeiro \*et al.\* \(2005\)](#). Une application actuarielle des différentes notions de la TVE est présentée à la fin de cette partie (Cette partie issue d'un exposé oral lors des Journées Internationales de Statistiques Théoriques et Appliquées JISTA'13 (USTHB, 2013),

intitulé "Estimateur à biais réduit de la prime de réassurance pour une distribution de sinistralité extrême").

Une deuxième partie de ce chapitre est réservée à l'estimation semi-paramétrique dans le cas d'un échantillon *i.i.d* en présence de censure aléatoire à droite : L'idée initiale fut introduite par Reiss et Thomas en 1997 (voir [66], section 6.1), quand ils ont proposé un estimateur de  $\gamma > 0$ , mais sans énoncer de résultats asymptotiques sur l'estimateur proposé. Récemment, Beirlant *et al.* (2007) proposa un estimateur généralisé de  $\gamma \in \mathbb{R}$ , ainsi que l'estimation du quantile extrême en présence de censure aléatoire à droite. Ces contributions ont été fondatrices de travaux récents à l'instar d'Einmahl *et al.* (2008), Gomes and Neves (2011), ou tout récemment Worms and Worms (2014) qui proposèrent un nouveau type d'estimateur non-paramétrique de  $\gamma$  pour une distribution à queue lourde. Cette deuxième partie du chapitre II, a fait l'objet d'un exposé oral lors du colloque international MSS'14 (USTHB, 2014) intitulé "Optimal approach on the heavy-tailed distributions mean estimation with right random censoring".

### Chapitre III : "Notions de la statistique bayésienne"

Une présentation non-exhaustive, et un rappel des définitions, et des notions basiques de la théorie bayésienne, représentant le fondement de notre contribution objet de cette thèse : Nous décrivons dans ce chapitre, quelques notions essentielles à l'analyse bayésienne, de sa construction à la mise en œuvre pratique d'un tel modèle. Nous commençons par rappeler tout d'abord, les fondements théoriques des modèles bayésiens, permettant d'introduire dans un second temps, l'ensemble des techniques bayésiennes en étalant le mécanisme d'inférence sur un paramètre du modèle statistique classique où on prétend avoir une connaissance a priori. La troisième partie de ce chapitre sera consacrée au choix discutable de la loi a priori du paramètre qu'on supposera, et l'idée derrière la construction d'une telle loi en fonction de la structure du modèle statistique. À la fin de ce chapitre, nous présentons une application des concepts bayésiens à l'estimation de l'indice de queue  $\alpha$  relatif à une distribution appartenant au max-domaine d'attraction de Fréchet pour un échantillon de données *i.i.d* complet (Une contribution inspirée du cas traité dans [2]).

### Chapitre IV : "Quelques contributions à l'estimation de l'indice de queue $\alpha$ pour une distribution à queue lourde en présence de censure aléatoire à droite"

Nous présentons dans ce chapitre, l'ensemble de nos contributions de recherche développée pendant l'élaboration de cette thèse :

La première contribution est présentée sous la forme d'un article publié dans la revue *Computational Statistics & Data Analysis*, intitulé "Bayesian estimation of the tail index of a heavy tailed distribution under random censoring", volume 104C, page 148–168 (<http://dx.doi.org/10.1016/j.csda.2016.06.009>), dans laquelle nous présentons un nouveau type d'estimateurs bayésiens de l'indice de queue  $\alpha$  relatif à une distribution à queue lourde ( $\alpha > 0$ ). Pour différents types de lois a priori, ces estimateurs sont construits sous un modèle semi-paramétrique, et en considérant un échantillon de données *i.i.d* censurées aléatoirement à droite. Nous établirons la convergence asymptotique de la loi a posteriori de  $\alpha$  vers une distribution gaussienne. Les propriétés empiriques de chacun des estimateurs proposés seront discutées à l'aide d'une étude par simulation et une application sur des données réelles.

Une deuxième contribution porte sur l'utilisation d'un modèle hybride (ou composite) dit Lognormal-Pareto Composite model (LPC), issu d'un mélange d'une loi lognormale et une loi de Pareto de type II. Nous rappellerons tout d'abord, dans le cas d'un échantillon complet, la construction d'une procédure d'inférence statistique par l'approche du maximum de vraisemblance, puis nous étendrons ce travail dans le cas où les données sont censurées aléatoirement à droite. La partie que nous présentons a été présentée lors de la 4<sup>th</sup> Tunisian Society of Financial Studies Conference (Sousse, Tunis) du 08 au 10 décembre 2016, intitulée "Modeling Algerian cars claims using LPC model under right random censoring data", et acceptée pour publication dans la revue "Journal of

*Business and Economics*” (ISSN 2155-7950), USA.

# CHAPITRE I

## INTRODUCTION À LA THÉORIE DES VALEURS EXTRÊMES



“

*... En toutes choses les extrêmes sont rares, les choses moyennes sont donc très communes ...* ”

Platon, *Apologie de Socrate*, 428 Av. JC - 347 Av. JC



a modélisation probabiliste a émergé comme l'une des disciplines mathématiques les plus importantes et la plus répondeuse dans les sciences appliquées au cours des cinquante dernières années. Ces techniques sont devenues largement utilisées dans différentes sciences et disciplines formant ainsi un champ étendu d'application.

À l'instar des modèles probabilistes les plus célèbres dans la littérature mathématique moderne, la modélisation des événements ou phénomènes extrêmes (qu'ils soient naturels : les ouragans, les tremblements de terre, les inondations, de nature économique : tels que les crises financières, les crashes boursiers, ou même de nature technologique tel que l'occupation d'un réseau internet...) est considérée aujourd'hui comme un champ de recherches particulièrement actif, notamment par l'impact socio-économiques d'occurrence de tels événements dans notre vie. En particulier, et depuis quelques années, une nouvelle discipline s'intéressant à l'analyse des événements rares a vu le jour, en tentant de mettre le point sur leur aspect imprévisible ; c'est la Théorie des Valeurs Extrêmes (TVE) qui a pour objectif la modélisation de tels événements.

Dans le domaine hydrologique par exemple, où la prévision des crues est particulièrement importante, la théorie des valeurs extrêmes apporte des indicateurs très précis permettant le bon dimensionnement des projets hydrauliques.

En météorologie, où l'étude de la vitesse des rafales de vent, permet d'évaluer le degré de résistance des matériaux face à la pression exercée par le vent. On peut citer un autre exemple, qui serait l'estimation de l'intensité des vagues des canicules : une analyse de ces vagues de chaleurs extrêmes permet aux autorités

publiques la mise en place des moyens sanitaires nécessaires pour éviter une sur-mortalité des personnes vulnérables.

En assurance dont l'une des préoccupations majeure est la prise en compte des sinistres extrêmes dans les modèles de tarification, afin de mieux estimer la juste prime que doit payer l'assuré.

Ainsi, la TVE cherche à fournir une réponse aux deux questions suivantes :

- Comment évalue-t-on la probabilité d'occurrence d'un événement rare ?
- Comment mesurer le risque associé à ces événements extrêmes ?

La notion d'événements rares ou extrêmes a été définie pour la première fois par le mathématicien Britannique A. Gumbel (1891-1966) :

*Un évènement rare est un évènement dont la probabilité d'occurrence est trop "petite".*

Cette notion de rareté est caractérisée par :

- Une probabilité d'occurrence significative (l'importance des queues de la distribution)
- Une période de retour d'un tel évènement assez longue
- Une grande intensité de l'évènement quand il se réalise

Le développement asymptotique de la théorie des valeurs extrêmes fut en parallèle du théorème centrale limite (TCL). En effet, les deux théories portent quelques éléments de ressemblance :

Soit  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon de variables aléatoires supposées indépendantes et identiquement distribuées. Si le TCL concerne la recherche du comportement asymptotique d'une somme  $S_n = X_1 + X_2 + \dots + X_n$  lorsque  $n \rightarrow \infty$ , la théorie des valeurs extrêmes quand à elle, cherche à dégager une loi limite de statistiques extrêmes  $M = \max(X_1, X_2, \dots, X_n)$  et  $m = \min(X_1, X_2, \dots, X_n)$  quand  $n \rightarrow \infty$ . Le point de départ d'un développement de la théorie des valeurs extrêmes fut la construction d'une distribution limite non-dégénérée de la dite statistique d'ordre.

Dans ce contexte, il est important de citer quelques ouvrages des plus appréciés par la communauté s'intéressant à la question des valeurs extrêmes, que nous recommandons vivement aux lecteurs :

- **Extreme Values Theory : and Introduction** (2006) par de Haan L. et Ferreira A. [40].
- **Extreme Values, Regular Variation, and Point Processes** (2008) par Sidney I. Resnick [67].
- **An Introduction to Statistical Modeling of Extreme Values** (2001) par Stuart Coles [18].
- **Statistical Analysis of Extreme value : Application to hydrology, insurance an finance** (Third Edition, 2007) par Reiss R. D et Thomas M. [66]
- **Statistics of Extremes - Theory and Applications** (2004) par Beirlant J., Goegebeur Y., Segers J. et Teugels J. [8]
- **Modelling Extremal Events for Insurance and Finance** (en deux éditions, 1997 et 2011) par Embrechts P., Kläuppelberg C. et Mikosch T. [30]

## I.1 Notions de statistiques d'ordre

La notion de statistiques d'ordre est considérée comme un instrument fondamental dans la théorie des valeurs extrêmes, car elle permet de décrire (notamment par la première et la dernière statistique) les queues de la distribution. Nous résumons dans la présente section, quelques définitions, propriétés et résultats relatifs a la notion de statistiques d'ordre. Nous recommandons pour plus de détails, le livre A First Course in Order Statistics (Classics in Applied Mathematics, 2008) par Arnold B.C. et Balakrishnan N., [3].

Considérons maintenant, un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  formé par une expérience aléatoire, auquel on associe une variable aléatoire  $X$ , de fonction de répartition  $F$  (qu'on suppose dans un cadre restreint absolument continue et admettant une densité de probabilité notée  $f$ ).

Soit un  $n$ -échantillon de variables aléatoires  $X_1, X_2, \dots, X_n$  supposées *i.i.d* et de même loi que  $X$ .

Pour chaque  $\omega \in \Omega$ , on a une valeur observée  $X_1(\omega), \dots, X_n(\omega)$  de ce  $n$ -échantillon. Ordonnons cette suite de valeurs par ordre croissant  $X_{k_1(\omega),n}(\omega) \leq X_{k_2(\omega),n}(\omega) \leq \dots \leq X_{k_n(\omega),n}(\omega)$ , où  $k_i(\omega)$  est le

numéro (ou rang) de la  $i^{\text{ième}}$  plus petite valeur. On définit ainsi,  $X_{i,n}(\omega) = X_{k_i(\omega),n}(\omega)$ , est la  $i^{\text{ième}}$  valeur rangée où le vecteur  $(X_{1,n}, \dots, X_{n,n})$  représente l'échantillon rangé.

**Définition I.1.1.** *La statistique d'ordre  $k$  d'un échantillon de  $n$  variables aléatoires  $X_1, X_2, \dots, X_n$  est la  $k^{\text{ième}}$  plus petite valeur des  $X_i$ , qu'on note par  $X_{k,n}$  ( $\forall k = 1$  à  $n$ ).*

Par la définition I.1.1, nous construisons l'échantillon aléatoire ordonné

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

Notons que par convention, le premier indice  $k$  désigne le rang de la  $k$  plus petite valeur de l'échantillon tandis que le second indice désigne la taille de l'échantillon  $n$ .

**Exemple I.1.1.** *Deux cas importants de statistiques d'ordre sont le minimum et le maximum.*

–  $X_{1,n}$  désigne le minimum des valeurs présent par les  $X_i$ 's :

$$X_{1,n} = \min(X_1, X_2, \dots, X_n)$$

–  $X_{n,n}$  désigne le maximum des valeurs présent par les  $X_i$ 's :

$$X_{n,n} = \max(X_1, X_2, \dots, X_n)$$

–  $X_{n,n} - X_{1,n}$  désigne l'étendue empirique de l'échantillon aléatoire  $X$ .

–  $X_{[\frac{n}{2}]+1,n}$  est l'estimateur empirique de la médiane de  $X$ .

Dans la suite, nous nous intéressons uniquement au deux premiers cas de statistiques d'ordre (le minimum  $X_{1,n}$  et le maximum  $X_{n,n}$ ).

**Définition I.1.2.** *(Statistiques d'ordre et fonction de répartition empirique)*

La fonction de répartition empirique d'un échantillon  $X_1, X_2, \dots, X_n$  est donnée par

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}, \forall x \in \mathbb{R} \\ &= \begin{cases} 0 & \text{si } x < X_{1,n} \\ \frac{i}{n} & \text{si } X_{i,n} \leq x < X_{i+1,n} \\ 1 & \text{si } x \geq X_{n,n} \end{cases} \end{aligned}$$

**Proposition I.1.1.**  $F_n(x)$  est un estimateur sans biais, converge presque surement vers  $F(x)$ , i.e :

$$\mathbb{E}[F_n(x)] = F(x) \text{ et } \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{ps} 0$$

Le résultat de convergence uniforme a été annoncé par le théorème de Gilvenko-Chantelli (1998). Plus de détails sur la démonstration de convergence peuvent être trouvés dans les ouvrages standards de probabilité (voir par exemple [16]).

**Définition I.1.3.** *(Statistiques d'ordre et fonction quantile empirique)*

Étant donné un échantillon  $X_1, X_2, \dots, X_n$  de fonction de répartition empirique  $F_n(\cdot)$ , la fonction quantile empirique associée à  $F_n$  est définie par

$$\begin{aligned} F_n^{-1} &: [0, 1] \longrightarrow \mathbb{R} \\ p &\longmapsto F_n^{-1}(p) = \inf \left\{ t : \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, t](X_i)} \geq p \right\} \end{aligned}$$

Autrement dit, pour  $m = 1, 2, \dots, n$

$$F_n^{-1}(p) = X_{m,n}, \forall p \in \left] \frac{m-1}{n}, \frac{m}{n} \right]$$

### I.1.1 Distributions de statistiques d'ordre

Nous donnons dans ce paragraphe, quelques résultats théoriques sur les lois exactes des principales statistiques d'ordre.

**Proposition I.1.2.** .

– Densité conjointe de  $X_{1,n}, \dots, X_{n,n}$

$$f_{X_{1,n}, \dots, X_{n,n}}(x_1, \dots, x_n) = n! \prod_{i=1}^n f(x_i), \text{ pour } -\infty < x_1 < x_2 < \dots < x_n < \infty$$

– Densité marginale d'une  $k^e$  statistique d'ordre,  $k = 1, \dots, n$

$$f_{X_{k,n}}(x) = \frac{n!}{(k-1)!(n-k)!} \{F(x)\}^{k-1} \{1-F(x)\}^{n-k} f(x) \text{ pour } -\infty < x < \infty$$

– Fonction de répartition d'une  $k^e$  statistique d'ordre,  $k = 1, \dots, n$

$$F_{X_{k,n}}(x) = \sum_{j=k}^n \binom{j}{n} F(x)^j (1-F(x))^{n-j}$$

Remarquons que la fonction de répartition d'une  $k^e$  statistique d'ordre est donnée par la queue d'une distribution binomiale de paramètre  $n$  et  $F(x)$ . Ainsi, et à partir de la proposition I.1.2, nous déduisons facilement les distributions exactes des maximas :

– La fonction de répartition de la variable aléatoire  $X_{1,n} = \min(X_1, X_2, \dots, X_n)$  est donnée par

$$F_{X_{1,n}}(x) = 1 - [1 - F(x)]^n,$$

sa densité de probabilité est

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

– La fonction de répartition de la variable aléatoire  $X_{n,n} = \max(X_1, X_2, \dots, X_n)$  est donnée par

$$F_{X_{n,n}}(x) = [F(x)]^n,$$

sa densité de probabilité est

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

**Proposition I.1.3.** (Transformation quantile)

Soit  $(U_1, \dots, U_n)$  un  $n$ -échantillon aléatoire issu d'une loi uniforme  $U_{[0,1]}$  et  $(U_{1,n}, \dots, U_{n,n})$  l'échantillon des statistiques d'ordre correspondant.

– Pour toute fonction de répartition  $F$ , on a

$$F(X_i) \rightsquigarrow U_{[0,1]} \text{ et } X_{i,n} \stackrel{d}{=} F^{\leftarrow}(U_{i,n}), \text{ pour tout } i = 1, \dots, n$$

où  $F^{\leftarrow}(t) = \inf\{x \in \mathbb{R}, F(x) \geq t\}$ .

– Si  $F$  est continue, alors

$$F(X_{i,n}) \stackrel{d}{=} U_{i,n}, \text{ pour tout } i = 1, \dots, n$$

### I.1.2 Existence des moments d'une statistique d'ordre

Soit  $X_1, X_2, \dots, X_n$  un échantillon issu de la loi  $F$ , continue, et de densité notée  $f$ . Nous montrons dans cette partie l'existence des moments des statistiques d'ordre.

**Proposition I.1.4.** (*Existence*)

Supposons que la loi  $F$  est telle qu'il existe un moment d'ordre  $k$ , i.e  $\mathbb{E}[|X|^k] = \int |x|^k dF(x) < \infty$ . Alors il existe un moment d'ordre  $k$  pour toutes les statistiques d'ordre, i.e :

$$\mathbb{E}[|X_{i,n}|^k] < \infty, \text{ pour tout } i \in 1, 2, \dots, n].$$

*Démonstration.* Nous définissant le moment d'ordre  $k$  de la  $i^e$  statistique d'ordre par :

$$\mathbb{E}[|X_{i,n}|^k] = \int_{-\infty}^{\infty} |x|^k f_{X_{i,n}}(x) dx$$

Comme

$$f_{X_{i,n}}(x) = \frac{n!}{(i-1)!(n-i)!} f(x) F(x)^i (1-F(x))^{n-i}$$

$$\begin{aligned} \mathbb{E}[|X_{i,n}|^k] &= \int_{-\infty}^{\infty} |x|^k \frac{n!}{(i-1)!(n-i)!} f(x) F(x)^i (1-F(x))^{n-i} dx \\ &\leq \int_{-\infty}^{\infty} |x|^k \frac{n!}{(i-1)!(n-i)!} f(x) dx \\ &= \frac{n!}{(i-1)!(n-i)!} \mathbb{E}[|X|^k] \end{aligned}$$

□

### I.1.3 Comportement limite des maximas

Notons par  $H_n(x)$  (resp.  $L_n(x)$ ) la fonction de répartition de la  $n^{\text{ième}}$  statistique d'ordre  $X_{n,n}$  (resp. de la  $1^{\text{ère}}$  statistique d'ordre  $X_{1,n}$ ) donnée par :

$$H_n(x) = \mathbb{P}[X_n, n \leq x] = [F(x)]^n$$

respectivement par

$$L_n(x) = \mathbb{P}[X_1, n \leq x] = 1 - [1 - F(x)]^n$$

Théoriquement, lorsque  $n \rightarrow \infty$ , on a :

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

et

$$\lim_{n \rightarrow \infty} L_n(x) = \begin{cases} 0 & \text{si } F(x) = 0 \\ 1 & \text{si } F(x) > 0 \end{cases}$$

Ainsi, la distribution de  $X_{n,n}$  (resp.  $X_{1,n}$ ) devrait nous fournir des informations sur les deux queues de la distribution. Cependant, la limite de cette distribution, lorsque  $n$  tend vers l'infini est dégénérée, i.e : le maximum (resp. le minimum) aura tendance à converger vers une valeur (peut aussi correspondre à l'infini) dite point terminal ou d'extrémité supérieur(e) (resp. inférieur(e)) de  $F$  -Il s'agit d'une traduction du terme anglais upper end point (resp. lower end point)- :

**Définition I.1.4.** (*Extrémité d'une distribution*)

Un point terminal supérieur (ou à droite)  $x_D$  d'une fonction de répartition  $F$  est défini par :

$$\begin{aligned} x_D &:= \sup \{x \in \mathbb{R}, F(x) < 1\} \\ &:= \inf \{x \in \mathbb{R}, F(x) = 1\} \end{aligned}$$

Un point terminal inférieur (ou à gauche)  $x_G$  d'une fonction de répartition  $F$  est défini par :

$$\begin{aligned} x_G &:= \sup \{x \in \mathbb{R}, F(x) = 0\} \\ &:= \inf \{x \in \mathbb{R}, F(x) > 0\} \end{aligned}$$

Cette notion de point terminal supérieur (resp. inférieur) se confond avec  $X_{n,n}$  (resp.  $X_{1,n}$ ) lorsque  $n \rightarrow \infty$ , et :

$$X_{n,n} \xrightarrow{d} x_D \text{ et } X_{1,n} \xrightarrow{d} x_G$$

Ainsi, l'objectif de la théorie des valeurs extrêmes consiste à rechercher une distribution limite non-dégénérée du maximum (resp. du minimum) à travers une transformation linéaire telle que, pour tout  $x \in \mathbb{R}$ ,  $\exists a_n \in \mathbb{R}, b_n > 0$  respectivement  $\exists c_n \in \mathbb{R}, d_n > 0$  :

$$\lim_{n \rightarrow \infty} H_n(a_n + b_n x) = \lim_{n \rightarrow \infty} [F(a_n + b_n x)]^n = H(x), \forall x \quad (\text{I.1})$$

et

$$\lim_{n \rightarrow \infty} L_n(c_n + d_n x) = \lim_{n \rightarrow \infty} 1 - [1 - F(c_n + d_n x)]^n = L(x), \forall x \quad (\text{I.2})$$

Sur la figure (I.1.3), on constate que la fonction  $H_n(x)$  se déplace à droite et change de forme quand  $n$  augmente. Inversement, la fonction  $L_n(x)$  translate à gauche même si cette translation ne s'opère pas de la même manière d'une loi à une autre. Ainsi, on peut imaginer qu'une translation  $a_n$  (resp.  $c_n$ ) et un changement d'échelle  $b_n$  (resp.  $d_n$ ) dépendants tout les deux de  $n$ , permettra d'avoir une distribution limite non-dégénérée avec la même forme que  $H(x)$  (resp.  $L(x)$ ). En d'autres termes, cette recherche d'une distribution limite du maximum (resp. du minimum) ce traduit par :

- La détermination des règles qui permettent la construction de suites  $a_n$  et  $b_n$  (resp.  $c_n$  et  $d_n$ ) sous lesquelles les équations (I.1) et (I.2) sont vérifiées
- L'identification de la ou les distributions  $H(x)$  (resp.  $L(x)$ ) possibles

La réponse à la problématique précédente est apportée par les deux théorèmes fondamentaux, énoncés par Fisher-Tippet(1928)- Gnedenko(1943) puis celui énoncé par Embrechts, Klüppelberg et Mikosch (1997).

## I.2 Fondements de la théorie des valeurs extrêmes

**Théorème I.2.1** (Distribution limite du maximum). *Sous des conditions générales sur  $F$ , il existe deux suites  $a_n, b_n$  et un paramètre réel  $\gamma$  tels que :*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{X_{n,n} - a_n}{b_n} \leq x \right) = H_\gamma(x)$$

avec,

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

où  $y_+ = \max(y, 0)$ .

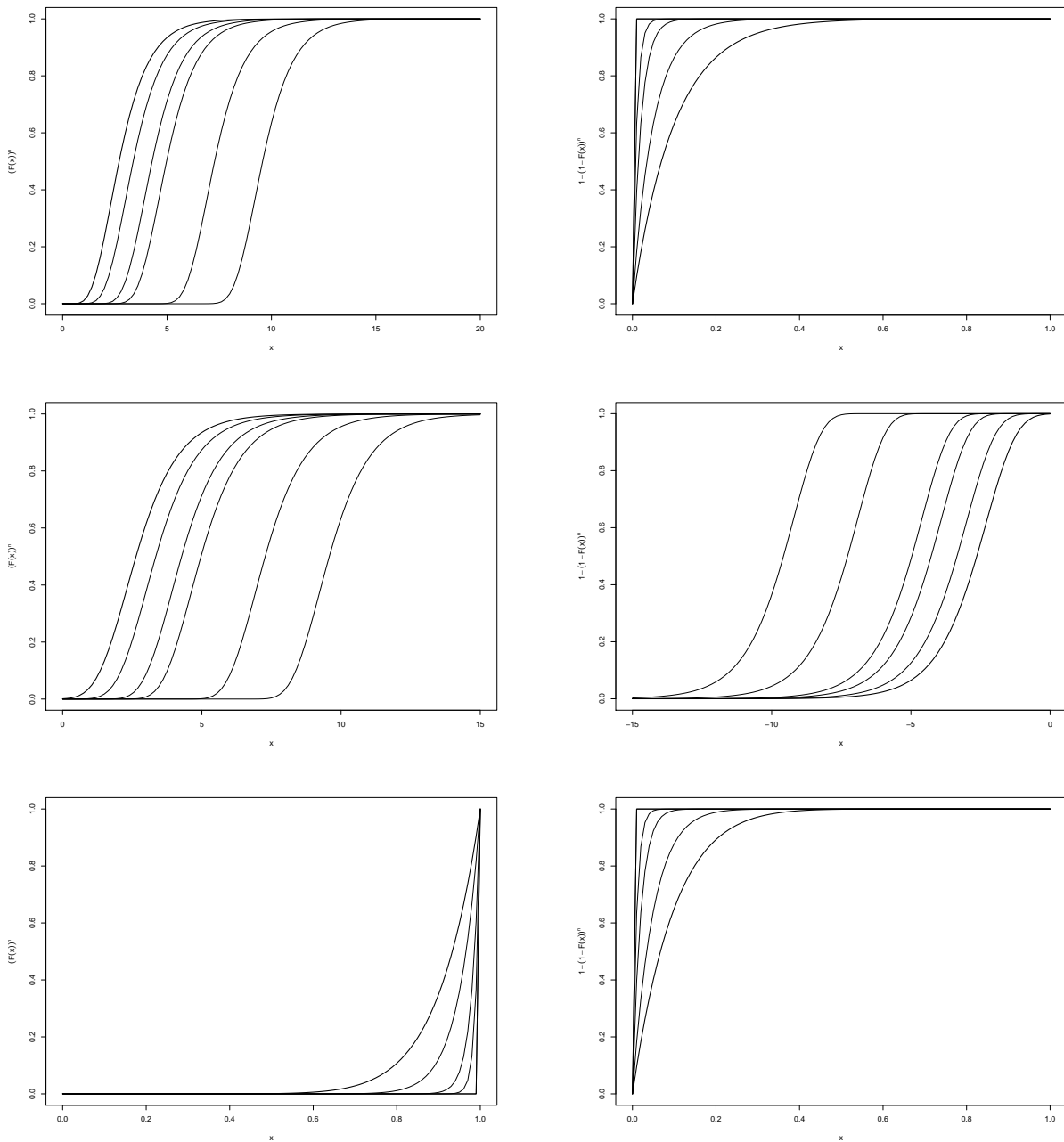


FIGURE I.1 – **À gauche** : Fonction de répartition du maximum  $H_n(x)$ , **à droite** : Fonction de répartition du minimum  $L_n(x)$  pour  $n = 10$  (en gras), 20, 50, 100, 1000 et 10000 respectivement pour une loi exponentielle  $\mathcal{E}(1)$  (ligne 1), Log-logistique standard (ligne 2) et une loi uniforme  $\mathcal{U}_{[0,1]}$  (ligne 3)

**Vocabulaire :**

- $H_\gamma$  est dite la loi des valeurs extrêmes maximum (notée aussi  $EV D_M(\gamma)$ ),
- $\gamma$  est dit indice des valeurs extrêmes
- $a_n$  et  $b_n$  sont des paramètres (ou suites) de normalisation.

*Démonstration.*

**Définition I.2.1.** On appelle fonction quantile généralisée d'une fonction  $F$ , la fonction définie, pour chaque  $t > 1$ , par

$$U(t) := F^{\leftarrow} \left( 1 - \frac{1}{t} \right)$$

où  $F^{\leftarrow}(s) = \inf\{x \in \mathbb{R}, F(x) > s\}$

On suppose que l'équation (I.1) est vérifiée, et en appliquant le théorème 1.1.2 dans deHaan et Ferriera [40, page 5] on a :

$$\lim_{t \rightarrow \infty} \frac{U(tx) - a(t)}{b(t)} = D(x) \quad (\text{I.3})$$

Pour tout  $x > 0$ , point de continuité de  $D(x) = F^{\leftarrow}(e^{-1/x})$ . Ainsi, on a :

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{b(t)} = D(x) - D(1) := E(x) \quad (\text{I.4})$$

Pour tout  $x > 0$ , il existe au moins un  $y > 0$  tel que

$$E(xy) = E(x)A(y) + E(y) \quad (\text{I.5})$$

où  $A(y) = \lim_{t \rightarrow \infty} \frac{b(ty)}{b(t)}$ . Posons  $s = \log(x)$ ,  $z = \log(y)$ , ( $x, y \neq 1$ ) et  $G(x) = E(e^x)$ . L'équation (I.5) se réécrit

$$G(z + s) = G(s)A(e^z) + G(z)$$

ou encore (comme  $G(0) = 0$ )

$$\frac{G(z + s) - G(z)}{s} = \frac{G(s) - G(0)}{s} A(e^z)$$

Certainement, il existe un point  $z$  où la fonction  $G$  est différentiable (comme  $G$  est monotone), et donc

$$G'(z) = G'(0)A(e^z)$$

On notons  $Q(z) = \frac{G(z)}{G'(0)}$ , comme  $G$  ne peut être constante de faite que  $H$  est non-dégénérée,  $Q(0) = 0$ ,  $Q'(0) = 1$  et  $Q'(z) = A(e^z)$ , on obtient

$$Q(z + s) - Q(z) = Q(s)Q'(z) \quad (\text{I.6})$$

En écrivant deux fois l'équation (I.6) en intervenant les variables  $z$  et  $s$  et en substituant l'une de l'autre on aura

$$Q(z) \frac{Q'(s) - 1}{s} = \frac{Q(s)}{s} (Q'(z) - 1)$$

Ainsi, lorsque  $s \rightarrow 0$ , on obtient

$$Q(z)Q''(0) = Q'(z) - 1$$

Par différentiation de l'équation précédente, on a

$$Q''(0)Q'(z) = Q''(z)$$

et

$$(\log Q')'(z) = Q''(0) := \gamma \in \mathbb{R}$$

Pour tout  $z$  (comme  $Q'(0) = 1$ )

$$Q'(z) = e^{z\gamma} \Rightarrow Q(z) = \int_0^z e^{u\gamma} du$$

$$G(z) = G'(0) \frac{e^{z\gamma} - 1}{\gamma}$$

d'où

$$D(x) = D(1) + G'(0) \frac{x^\gamma - 1}{\gamma},$$

et

$$D^{\leftarrow}(y) = \left(1 + \gamma \frac{y - D(1)}{G'(0)}\right)^{1/\gamma}$$

Finalement, le résultat du théorème [I.2.1](#) est démontré du fait que  $D^{\leftarrow}(x) = \frac{-1}{\log H(x)}$  □

### Interprétation graphique du théorème [I.2.1](#)

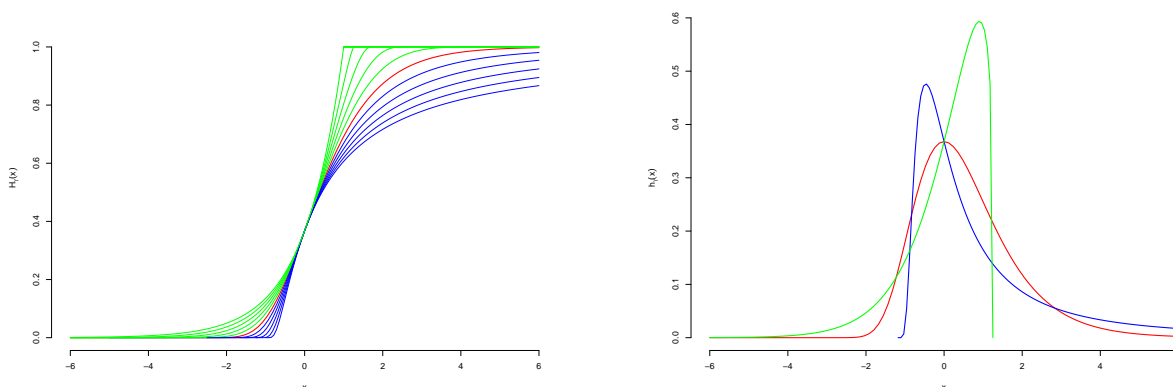


FIGURE I.2 – **À gauche** : variation de la distribution  $EVD_M(\gamma)$  en fonction de  $\gamma$ , **à droite** : sa densité de probabilité  $h_\gamma(x)$  – ligne en bleu  $\gamma > 0$ , en rouge  $\gamma = 0$ , en vert  $\gamma < 0$

À gauche de la figure [I.2](#), le tracé de la distribution  $EVD_M(\gamma)$  en fonction de différentes valeurs de  $\gamma$  illustre bien le fait que la forme de la fonction  $H_\gamma(x)$  tend vers la forme d'une fonction  $H_0()$  lorsque  $\gamma \rightarrow 0$ .

À droite de la figure [I.2](#), nous représentons la fonction densité de probabilité d'une loi  $EVD_M(\gamma)$  donnée par :

$$h_\gamma(x) = \begin{cases} (1 + \gamma x)^{-\left(\frac{\gamma+1}{\gamma}\right)} \exp\left(-(1 + \gamma x)^{-1/\gamma}\right) & \text{si } \gamma \neq 0 \text{ et } 1 + \gamma x > 0 \\ \exp(-x) \exp(-e^{-x}) & \text{si } \gamma = 0 \text{ et } x \in \mathbb{R} \end{cases}$$

On distingue essentiellement trois allures particulières selon la valeur de  $\gamma$  :

- $\gamma > 0$ , la fonction  $h_\gamma(x)$  est fortement asymétrique avec une queue importante à droite.
- $\gamma = 0$ ,  $h_\gamma(x)$  est faiblement asymétrique avec une queue fine à droite.
- $\gamma < 0$ ,  $h_\gamma(x)$  possède un point terminal supérieur fini.

**Théorème I.2.2** (Distribution limite du minimum). *Sous des conditions générales sur  $F$ , il existe deux suites  $c_n, d_n$  et un paramètre réel  $\gamma$  tels que :*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{X_{1,n} - c_n}{d_n} \leq x \right) = L_\gamma(x)$$

avec,

$$L_\gamma(x) = \begin{cases} 1 - \exp(-(1 - \gamma x)_+^{-1/\gamma}) & \text{si } \gamma \neq 0 \\ 1 - \exp(-e^x) & \text{si } \gamma = 0 \end{cases}$$

où  $y_+ = \max(y, 0)$ .

$L_\gamma$  est dite distribution des valeurs extrêmes minimum ( $EVD_m(\gamma)$ ), et  $c_n, d_n$  sont les paramètres de normalisation.

**Proposition I.2.1.** *Si une variable aléatoire  $X \sim EVD_m(\gamma)$  et  $Y = -X$ , alors :*

1.  $Y \sim EVD_M(\gamma)$ .
2.  $F_X(x) = 1 - F_Y(-x)$

**Remarque I.2.1.** *La réciproque reste aussi valable. Ainsi, nous nous intéresserons dans la suite de ce chapitre uniquement au comportement asymptotique du maximum. Celui du minimum s'obtient par la proposition I.2.1 (Aussi pour l'interprétation graphique du théorème I.2.2 à partir de la figure I.3)*

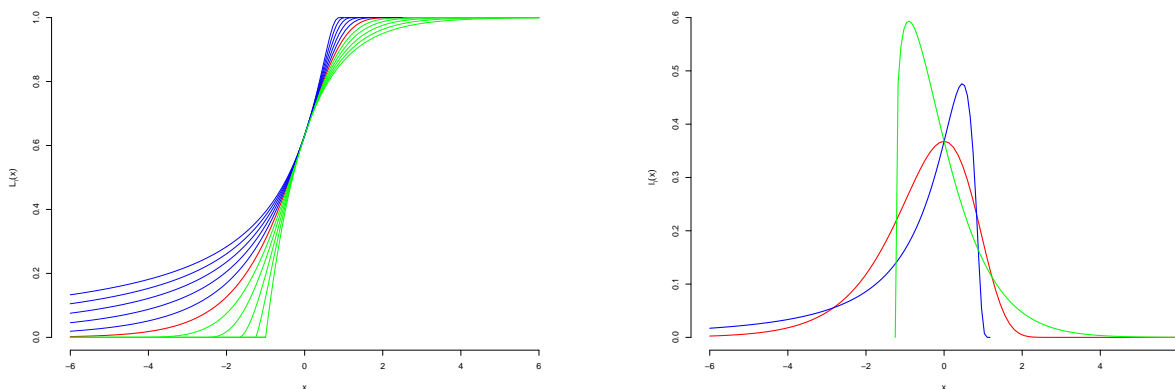


FIGURE I.3 – **À gauche** : variation de la distribution  $EVD_m(\gamma)$  en fonction de  $\gamma$ , **à droite** : sa densité de probabilité  $l_\gamma(x)$  – ligne en bleu  $\gamma > 0$ , en rouge  $\gamma = 0$ , en vert  $\gamma < 0$

Dans la littérature, une adaptation du théorème I.2.1 a été proposée sous une forme plus explicite et en liaison avec l'interprétation graphique selon les valeurs de  $\gamma$ .

**Théorème I.2.3** (Embrechts, Klüppelberg et Mikosch (1997)). *Supposons  $n$  variables aléatoires  $X_1, \dots, X_n$  indépendantes identiquement distribuées, de même loi  $F$  (supposée absolument continue). S'il existe des constantes  $a_n$  et  $b_n$  et une distribution limite non dégénérée  $G$  telles que*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{X_{n,n} - a_n}{b_n} \leq x \right) = G(x), \forall x \in \mathbb{R}$$

alors  $G$  appartient forcément à l'une de ces trois distributions

(I) *Fréchet type*

$$\Phi_\beta(x) = \begin{cases} \exp(-x^{-\beta}) & x \geq 0, \beta > 0 \\ 0 & \text{sinon} \end{cases}$$

(II) *Weibull type*

$$\Psi_\beta(x) = \begin{cases} \exp(-(-x)^\beta) & x < 0, \beta > 0 \\ 1 & \text{si } \geq 0 \end{cases}$$

(III) *Gumbel type*

$$\Lambda(x) = \exp(-e^{-x}), x \in \mathbb{R}$$

*Démonstration.* Supposons qu'il existe deux suites  $a_n \in \mathbb{R}$  et  $b_n > 0$ , et une distribution non dégénérée  $G$  telles que l'équation (I.1) est satisfaite quand  $n \rightarrow \infty$ . Pour tout  $t \in \mathbb{R}_+$  et  $[t]$  sa partie entière on a

$$F^{[nt]}(a_{[nt]} + b_{[nt]}x) \rightarrow G(x)$$

et

$$F^{[nt]}(a_n + b_n x) = (F^n(a_n + b_n x))^{\frac{[nt]}{n}} \rightarrow G^t(x)$$

Afin d'assurer une convergence uniforme telle qu'énonce le théorème I.2.3, on suppose l'existence de  $\alpha(t) > 0$  et  $\beta(t) \in \mathbb{R}$  telles qu'en appliquant le théorème de Skorohod (voir Resnick [66], page 6) pour tout  $t > 0$

$$\lim_{n \rightarrow \infty} \frac{b_n}{b_{[nt]}} = \alpha(t), \quad \lim_{n \rightarrow \infty} \frac{a_n - a_{[nt]}}{b_{[nt]}} = \beta(t) \quad (\text{I.7})$$

et

$$G^t(x) = G(\alpha(t)x + \beta(t)) \quad (\text{I.8})$$

Pour tout  $t > 0, s > 0$ , on a d'une part

$$G^{ts}(x) = G(\alpha(ts)x + \beta(ts))$$

et d'autre part

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

Comme  $G$  est supposée être non-dégénérée, on conclut par identification

$$\alpha(ts) = \alpha(t)\alpha(s) \quad (\text{I.9})$$

et

$$\beta(ts) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s) \quad (\text{I.10})$$

L'équation (I.9) est dite équation fonctionnelle de Hamel (ou Cauchy puissance). Sa seule solution mesurable, non négative à la forme  $\alpha(t) = t^{-\theta}, \theta \in \mathbb{R}$ . Ainsi, nous considérons trois cas possibles, (a) :  $\theta = 0$ , (b) :  $\theta > 0$  et (c) :  $\theta < 0$ .

**Cas (a) :** ( $\theta = 0$ )

Dans ce  $\alpha = 1$  et l'équation (I.10) devient

$$\beta(ts) = \beta(t)\beta(s),$$

une simple variante de l'équation fonctionnelle de Cauchy dont la solution est

$$\beta(t) = -c \log(t), t > 0 \text{ et } c \in \mathbb{R}$$

l'équation (I.6) s'écrit ainsi

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

comme  $G(x < 1)$  pour tout  $x$ , et particulièrement pour  $x = 0$  et  $t > 0$

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

Finalement, et on posons le changement de variables  $u = -c \log(t)$ ,  $u \in \mathbb{R}$  et  $G(0) = \exp(-e^{-p}) \in (0, 1)$ , on obtient

$$\begin{aligned} G(u) &= \exp(-e^{-p} e^{-u/c}) \\ &= \Lambda(p + c^{-1}u) \end{aligned}$$

**Cas (b) :** ( $\theta > 0$ )

de l'équation (I.10)

$$\alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s)$$

d'où

$$\frac{\beta(s)}{1 - \alpha(s)} = \frac{\beta(t)}{1 - \alpha(t)}$$

la fonction  $\beta(\cdot)(1 - \alpha(\cdot))^{-1}$  semble être constante. ce qui nous permet d'écrire pour  $t \neq 1$   $\beta(t) = c(1 - t^{-\theta})$ , et l'équation (I.6) devient

$$G^t(x) = G(t^{-\theta(x-c)+c}) \Rightarrow G^t(x+c) = G(t^{-\theta}x+c)$$

Posons  $H(x) + G(x+c)$ , où  $H$  est une fonction non dégénérée satisfaisante la relation

$$H^t(x) = H(t^{-\theta}x)$$

Pour  $x = 1$ ,  $H(1) \in (0, 1)$  et on posons  $\alpha = \frac{1}{\theta}$ ,  $H(1) = \exp(-p^{-\alpha})$ ,  $u = t^{-1/\alpha}$ , on obtient

$$\begin{aligned} H(u) &= \exp(-(pu)^{-\alpha}) \\ &= \Phi_{\alpha}(pu) \end{aligned}$$

**Cas (c) :** ( $\theta < 0$ )

Ce dernier cas se traite de la même façon que le cas (b) précédent, avec  $\alpha(t) = t^{\theta}$

□

Ce théorème présente un intérêt important, car si l'ensemble des distributions est 'grand', l'ensemble des distributions de valeurs extrêmes est lui très petit. (Stuart Coles, 2001) commente ce résultat et dit dans [18] : "The remarkable feature of this result is that the three types of extreme value distributions are the only possible limits for the distribution of the  $\frac{X_{n,n} - a_n}{b_n}$ , regardless of the distribution  $F$  for the population. It is in this sense that the theorem provides an extreme value analog of the central limit theorem "

Néanmoins, ce théorème n'est valable et utile que si les suites  $a_n$  et  $b_n$  existent! Cela revient à donner une réponse à la problématique suivante : Existe-t-il un critère que doit satisfaire une distribution quelconque  $F$  pour que la loi du maximum d'une suite de variables aléatoires i.i.d.  $X_1, X_2, \dots, X_n$  issues de  $F$  converge vers  $H$ ?

## I.2.1 Caractérisation des max-domaines d'attraction

**Définition I.2.2** (Famille de lois max-stables). Une famille paramétrique de fonctions de répartition,  $F_\theta = \{F(x, \theta); \theta \in \Theta\}$ , est dite max-stable (où stable par rapport à la loi du maximum) si est seulement si la loi du maximum reste dans la même famille de loi  $F_\theta$ , i.e :

$$[F(x, \theta)]^n = F(x, \theta(n)),$$

où  $\theta(n)$  est un paramètre dépendant de  $n$ .

**Exemple I.2.1** (Max-stabilité de la loi de Weibull). Soit  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon aléatoire issu de la loi de Weibull généralisée :

$$F_X(x, \mu, \sigma, \gamma) = \exp \left[ - \left( -\frac{x - \mu}{\sigma} \right)^\gamma \right]$$

Il est évident que la loi du maximum est une loi de Weibull de paramètres  $\mu, \sigma n^{-1/\gamma}$  et  $\gamma$ .

**Définition I.2.3.** (Fonction à variation régulière)

1. Une fonction  $h$ , positive et mesurable sur  $[0, +\infty[$  est dite à variation régulière d'indice  $\rho \in \mathbb{R}$  au voisinage de l'infini (notée  $h \in \mathcal{R}_\rho$ ) ssi :

$$\lim_{t \rightarrow \infty} \frac{h(tx)}{h(t)} = x^\rho, \forall x > 0.$$

Si  $\rho = 0$ , la fonction  $h$  est dite à variation lente à l'infini.

2.  $h$  est une fonction à variation régulière en zéro d'indice  $\rho \in \mathbb{R}$  (on note  $h \in \mathcal{R}_\rho^0$ ) ssi  $h(1/x) \in \mathcal{R}_{-\rho}$ .
3.  $h$  est dite fonction régulière en un point  $a > 0$  (on note  $h \in \mathcal{R}_\rho^a$ ) ssi la fonction  $h(a - \frac{1}{x}) \in \mathcal{R}_\rho$ .

**Propriété I.2.1.**  $h \in \mathcal{R}_\rho$  si et seulement si  $h(x) = x^\rho \ell(x)$ , où  $\ell \in \mathcal{R}_0$ .

**Théorème I.2.4** (Représentation de Karamata). Si  $h \in \mathcal{R}_\rho$ , pour quelque  $\rho \in \mathbb{R}$ , alors pour tout  $x \geq A$  on a :

$$h(x) = c(x) \exp \left( \int_A^x \frac{r(t)}{t} dt \right),$$

où  $A > 0$ ,  $c$  et  $r$  sont deux fonction mesurables avec :

$c(x) \rightarrow c_0 \in [0, +\infty[$  et  $r(x) \rightarrow \rho$  lorsque  $x \rightarrow \infty$ .

**Définition I.2.4.** (Domaine d'attraction d'une distribution)

On dit qu'une distribution  $F$  appartient au Max-domaine d'attraction de  $G$ , -et on note  $F \in MDA(G)$ - si l'équation (I.1) est satisfaite au moins pour une paire de suites  $a_n$  et  $b_n > 0$  (la distribution du maximum normalisée converge vers  $G$ ).

D'une manière similaire, s'il existe  $c_n$  et  $d_n > 0$  pour lesquelles  $F$  satisfait l'équation (I.2), alors  $F$  est dite dans le Min-domaine d'attraction de  $G$ .

**Exemple I.2.2.** Pour la distribution de Pareto  $F(x) = 1 - cx^{-1/\theta}$  où  $c > 0, \theta > 0$

En posant  $b_n = (cn)^\theta$  et  $a_n = 0$ , on a :

$$\begin{aligned} \lim_{n \rightarrow \infty} [F(b_n x + a_n)]^n &= \lim_{n \rightarrow \infty} \left[ 1 - \frac{x^{-1/\theta}}{n} \right]^n \\ &= \exp(-x^{-1/\theta}), x > 0 \text{ et } \theta > 0 \end{aligned}$$

On dit donc que  $F$  appartient au domaine d'attraction de Fréchet et on note  $F \in MDA(\Phi_{1/\theta})$

**Théorème I.2.5** (Castillo et al. (1989), Galambos (1987)). *Une condition nécessaire et suffisante pour qu'une f.d.r  $F$  soit dans l'un des trois Max-domaines d'attraction  $H_\gamma$ , est que :*

$$\lim_{\varepsilon \rightarrow 0} \frac{F^{-1}(1 - \varepsilon) - F^{-1}(1 - 2\varepsilon)}{F^{-1}(1 - 2\varepsilon) - F^{-1}(1 - 4\varepsilon)} = 2^\gamma,$$

où  $\gamma$  est le paramètre de forme associé à la distribution limite  $EVD_M$ .

Le résultat du théorème (I.2.5) implique que :

- Si  $\gamma > 0$  alors  $F \in MDA(\Phi)$  (Fréchet)
- Si  $\gamma = 0$  alors  $F \in MDA(\Lambda)$  (Gumbel)
- Si  $\gamma < 0$  alors  $F \in MDA(\Psi)$  (Weibull)

## I.2.2 Caractérisation du max-domaine d'attraction de Fréchet

**Théorème I.2.6** (Gnedenko 1943).  *$F \in MDA(\Phi_{1/\gamma})$  si et seulement si  $1 - F \in \mathcal{R}_{-1/\gamma}$ . Dans ce cas,  $b_n = F^{-1}(1 - \frac{1}{n})$  et  $a_n = 0$ .*

Ce théorème formulé par Gnedenko permet de caractériser très simplement les distributions  $F \in MDA(\Phi_{1/\gamma})$ . En effet, elles doivent vérifier pour  $\gamma > 0$  et  $x > 0$  :

$$\lim_{n \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}$$

## I.2.3 Caractérisation du max-domaine d'attraction de Weibull

**Théorème I.2.7.**  *$F \in MDA(\Psi_{1/\gamma})$  si et seulement si  $1 - F(x_F - \frac{1}{x}) \in \mathcal{R}_{-1/\gamma}$  et  $x_F < \infty$ . Dans ce cas,  $a_n = x_F$  et  $b_n = x_F - F^{-1}(1 - \frac{1}{n})$ .*

**Exemple I.2.3.** *Considérons la distribution uniforme standard. Comme  $x_F = 1$  et*

$$\lim_{t \rightarrow \infty} \frac{1 - F(1 - \frac{1}{tx})}{1 - F(1 - \frac{1}{t})} = \lim_{t \rightarrow \infty} \frac{\frac{1}{tx}}{\frac{1}{t}} = x^{-1}$$

alors, la loi uniforme standard est dans max-domaine d'attraction de Weibull ( $\mathcal{U}_{[0,1]} \in MDA(\Psi)$ )

## I.2.4 Caractérisation du max-domaine d'attraction de Gumbel

**Théorème I.2.8.** *La caractérisation du MDA de Gumbel est difficile à énoncer du fait qu'il n'existe pas de condition nécessaire et suffisante relativement simple. Mais si la fonction  $F$  est de classe  $\mathcal{C}^2$ , une condition suffisante relativement simple à vérifier est la suivante :*

$$\lim_{x \rightarrow \infty} \frac{(1 - F(x))\partial^2 F}{(\partial F)^2} = -1$$

Les coefficients de normalisation sont ainsi donnés par :  $a_n = F^{-1}(1 - \frac{1}{n})$  et  $b_n = h(a_n)$  où  $h(x) = \frac{1 - F(x)}{f(x)}$ .

**Exemple I.2.4.** Prenons l'exemple de la distribution normale standard avec :  $\lim_{x \rightarrow \infty} \Phi(x) = 1$

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\partial^2 \Phi = -x f(x)$$

En utilisant l'approximation de Abramowitz et Stegun donnée pour  $x \gg 1$ , par :

$$\Phi(x) = 1 - \frac{f(x)}{x} \left( 1 - \frac{1}{x^2} + \frac{1 \cdot 3}{x^4} - \frac{1 \cdot 3 \cdot 5}{x^6} + \cdots + \frac{1 \cdot 3 \cdots (2n-1)}{x^{2n}} \right) + R_n(x),$$

où  $R_n(x) = (-1)^{n+1} 1 \cdot 3 \cdots (2n+1) \int_x^\infty \frac{f(y)}{y^{2n+2}} dy$  la condition nécessaire est suffisante dans le théorème I.2.8 est vérifiée. D'où  $\Phi \in MDA(\Lambda)$ .

### I.3 Fluctuation des maxima

**Définition I.3.1** (Distributions des extrêmes généralisées). :

1. On appelle loi des valeurs extrêmes généralisée (notée  $GEV(\mu, \sigma, \gamma)$ ), la loi définie par une fonction de répartition  $G_{EV}$  donnée par :

$$G_{EV}(x) = \begin{cases} \exp\left(-\left[1 + \gamma \left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right) & \text{pour } \gamma \neq 0, 1 + \gamma \left(\frac{x-\mu}{\sigma}\right) > 0 \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & \text{pour } \gamma = 0 \end{cases}$$

2. On appelle loi de Pareto généralisée notée  $GPD(\mu, \sigma, \gamma)$ , la loi définie par une fonction de répartition  $G_P$  donnée par :

$$G_P(x) = \begin{cases} 1 - \left[1 + \gamma \left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma} & \text{pour } \gamma \neq 0, 1 + \gamma \left(\frac{x-\mu}{\sigma}\right) > 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{pour } \gamma = 0 \end{cases}$$

$\mu \in \mathbb{R}$  est le paramètre de position (ou de localisation),  $\sigma > 0$  est le coefficient d'échelle et  $\gamma \in \mathbb{R}$  est dit coefficient de forme.

#### I.3.1 Approche des maxima par bloc (ou GEV)

L'étude des extrêmes passe par l'analyse du maximum d'un échantillon de taille  $n$  donnée. Soit  $X_1, X_2, \dots, X_n$  un échantillon de variables aléatoires i.i.d., représentant des données observées journalièrement durant un an ( $n = 365$ ), et  $X_{n,n} = \max(X_1, X_2, \dots, X_n)$  est la plus grande valeur de l'échantillon.

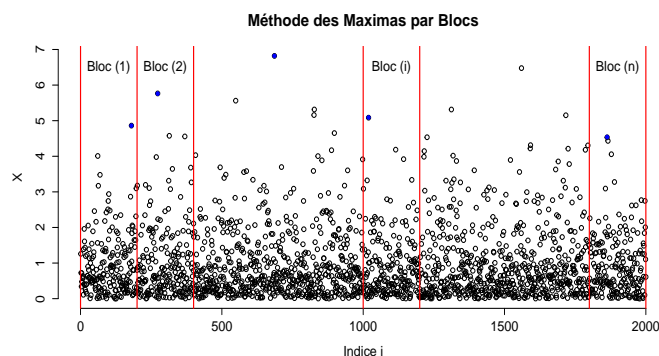
Le théorème de Gnedenko affirme que la loi de  $X_{n,n}$  est donnée par  $H_\gamma(x)$  précédemment définie. Cette loi appartient à la famille des distributions Généralisées des Valeurs Extrêmes (GEV).

On peut ainsi construire un échantillon des maxima en s'intéressant aux valeurs records annuels  $X_{n,n}^{(1)}, X_{n,n}^{(2)}, \dots, X_{n,n}^{(k)}$  enregistrées sur  $k$  années (Analyse en blocs) dont la loi commune est une GEV.

Différentes méthodes existent pour estimer les paramètres d'une GEV à partir d'un échantillon iid des maxima  $(m_1, m_2, \dots, m_k)$ . La méthode la plus répandue est la méthode du MV.

La log-vraisemblance est donnée par :

$$\begin{aligned} \log \mathcal{L}(\tilde{m}, \mu, \sigma, \gamma) &= -k \log(\sigma) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^k \log \left( 1 + \gamma \left(\frac{m_i - \mu}{\sigma}\right)^{-1/\gamma} \right) \\ &\quad - \sum_{i=1}^k \left( 1 + \gamma \left(\frac{m_i - \mu}{\sigma}\right)^{-1/\gamma} \right) \end{aligned}$$



Avec les contraintes  $1 + \gamma \frac{m_i - \mu}{\sigma} > 0, \forall i = 1 \dots k$ .

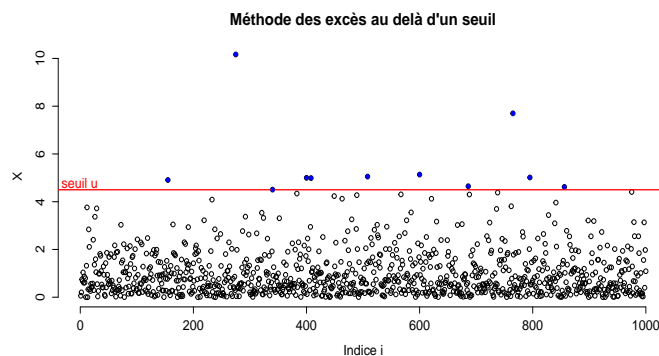
La résolution analytique du système précédent n'existe pas et nécessite donc, l'introduction d'approches numériques de résolution telles que l'algorithme de Newton Raphson.

Le choix du nombre de blocs dans la méthode GEV est généralement arbitraire, et on perd généralement de l'information sur les événements extrêmes qui par nature sont déjà peu observés. Par exemple, plusieurs événements extrêmes intéressants peuvent être dans un même bloc alors qu'un autre bloc ne contient pas d'événements "extrêmes" !

En pratique, cela signifie qu'il faut beaucoup de données pour pouvoir mettre en place la méthode des maxima par bloc (typiquement quelques décennies si on fait des blocs annuels), ce qui n'est généralement pas le cas en pratique.

### I.3.2 Approche des excès au-delà d'un seuil (ou POT)

Une alternative à la méthode des maxima par bloc consiste à conserver toutes les observations qui dépassent un niveau élevé puis les ajuster par une loi de probabilité aux dépassements de seuil. Il s'agit de la méthode des dépassements de seuil ("Peak Over Threshold" (POT)).



Soit  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon aléatoire issu d'une loi  $F$ . Pour un seuil  $u > 0$ , on définit l'échantillon des excès au-delà de  $u$  par  $Z_1, Z_2, \dots, Z_{N_u}$  où  $N_u$  désigne le nombre de variables  $X_i$  dépassant le seuil  $u$ .

D'après le théorème de Fisher-Tippet, il est naturel de supposer pour  $n$  assez grand que :

$$F^n(x) \approx \exp\left(-\left[1 + \gamma \frac{x - \mu}{\sigma}\right]^{-1/\gamma}\right)$$

$$\Rightarrow n \log(F(x)) \approx -[1 + \gamma \frac{x - \mu}{\sigma}]^{-1/\gamma}$$

Or, à la queue de la distribution (pour  $x > u$ ,  $u$  assez grand  $F(x) \approx 1$ ) on a  $\log(F(x)) = -(1 - F(x))$ , d'où l'approximation :  $P(X \geq x) = 1 - F(x) \approx \frac{1}{n}[1 + \gamma \frac{x - \mu}{\sigma}]^{-1/\gamma}$

En effet, le but est de donner la loi conditionnelle sachant que la variable  $X$  dépasse le seuil  $u$  donnée par

$$\begin{aligned} P(X \geq y + u | X > u) &= \frac{1 - F(y + u)}{1 - F(u)} \text{ pour } y > 0 \\ &\approx \frac{n^{-1}[1 + \gamma \frac{y + u - \mu}{\sigma}]^{-1/\gamma}}{n^{-1}[1 + \gamma \frac{u - \mu}{\sigma}]^{-1/\gamma}} \\ &\approx \left[1 + \frac{\gamma y}{\sigma(u)}\right]^{-1/\gamma} \end{aligned}$$

où  $\sigma(u) = \sigma + \gamma(u - \mu)$ .

$$\mathbb{P}(X \geq y + u | X > u) \approx 1 - \left[1 + \frac{\gamma y}{\sigma(u)}\right]^{-1/\gamma}$$

**Théorème I.3.1** (Pickands, Balkema and de Haan (1975)). *Pour un seuil  $u > 0$  assez grand, si  $F$  appartient à l'un des trois domaines d'attraction de la loi des valeurs extrêmes (Fréchet, Gumbel ou Weibull), alors il existe une fonction  $\sigma(u)$  strictement positive et un réel  $\gamma$  tels que*

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - G_P^{(0, \sigma(u), \gamma)}(y)| = 0$$

où  $G_P^{(0, \sigma(u), \gamma)}$  est la fonction de répartition de la loi de Pareto Généralisée et  $F_u$  est la fonction de répartition des excès au delà du seuil  $u$ .

D'après le théorème précédent, l'échantillon des excès au-delà d'un seuil  $u$  est distribué selon une loi GPD de paramètres  $\mu = 0$ ,  $\sigma(u) = \sigma + \gamma(u - \mu)$  et  $\gamma > 0$ .

Finalement, le modèle POT précédent s'identifie à partir d'un échantillon de dépassement  $z_1, z_2, \dots, z_{N_u}$ , en estimant les paramètres  $\gamma$  et  $\sigma(u)$ . Avec les notations précédente, la log-vraisemblance est donnée par

$$\log \mathcal{L}(\tilde{z}, \sigma(u), \gamma) = -N_u \log \sigma(u) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \log \left(1 + \gamma \frac{z_i}{\sigma(u)}\right)$$

Les estimateurs du MV,  $\hat{\gamma}$  et  $\hat{\sigma}(u)$  sont ainsi obtenues par la résolution numérique du système d'équation suivant :

$$\begin{cases} \frac{\partial}{\partial \gamma} \log \mathcal{L}(\tilde{z}, \sigma(u), \gamma) = 0 \\ \frac{\partial}{\partial \sigma(u)} \log \mathcal{L}(\tilde{z}, \sigma(u), \gamma) = 0 \end{cases}$$

### I.3.3 Approche des moments pondérés

**Définition I.3.2.** *On appelle moment pondéré d'ordre  $r$  d'une v.a issue de la loi GPD(0,  $\sigma$ ,  $\gamma$ ) (notons par  $G_P^{(\sigma, \gamma)}$  sa fonction de répartition), la quantité donnée par*

$$\begin{aligned} \omega_r &:= \mathbb{E}[X(\bar{G}_P(X))^r] \\ &= \frac{\sigma}{(r+1)(r+1-\gamma)}, \end{aligned}$$

où  $\bar{G}_P = 1 - G_P^{(\sigma, \gamma)}$ .

En exprimant  $\omega_r$  pour  $r = 0$  et  $r = 1$ , on peut écrire :

$$\begin{cases} \sigma = \frac{2\omega_0\omega_1}{\omega_0 - 2\omega_1} \\ \gamma = 2 - \frac{\omega_0}{\omega_0 - 2\omega_1} \end{cases}$$

En remplaçant  $\omega_0$  et  $\omega_1$  par l'estimateur empirique de  $\omega_r$  donnée par :

$$\hat{\omega}_r = \frac{1}{n_u} \sum_{i=1}^{n_u} z_i [F_{n_u}(z_i)]^r, \text{ pour } r = 0, 1$$

où  $F_n$  désigne la *fdr* empirique de  $z_1, z_2, \dots, z_{n_u}$ , on obtient :

$$\begin{cases} \hat{\sigma} = \frac{2\hat{\omega}_0\hat{\omega}_1}{\hat{\omega}_0 - 2\hat{\omega}_1} \\ \hat{\gamma} = 2 - \frac{\hat{\omega}_0}{\hat{\omega}_0 - 2\hat{\omega}_1} \end{cases}$$

### I.3.4 Approche moyenne des excès au-delà d'un seuil

L'approche moyenne des excès au-delà d'un seuil vient en appui et en complément à l'approche POT qui repose essentiellement sur la détermination du seuil  $u$ , suffisamment élevé pour assurer une bonne qualité d'ajustement du modèle GPD. En effet :

- Une sélection d'un seuil trop élevé, garantit certes l'hypothèse sous laquelle la méthode POT a été fondée mais réduit le nombre d'observations extrêmes.
- En choisissant un seuil trop bas, cela risque de fausser l'hypothèse fondatrice du théorème précédent.

**Définition I.3.3.** 1. On appelle fonction moyenne des excès (*Mean excess function*) au-delà d'un seuil  $u > 0$ , la quantité donnée par

$$e(u) = \mathbb{E}(Z - u | Z > u)$$

2. Soit  $z_1, z_2, \dots, z_n$  un  $n$ -échantillon *i.i.d* issu d'une v.a  $Z$  (abs. continue). On appelle fonction moyenne des excès empirique (*Empirical mean excess function*) au-delà d'un seuil  $u > 0$ , la quantité donnée par

$$\hat{e}(u) = \frac{1}{N_u} \sum_{i=1}^n (z_i - u)^+$$

où  $y^+ = \max(y, 0)$  et  $N_u$  désigne le nombre d'observation  $z_i$  dépassant le seuil  $u$ .

Comme  $Z - u | Z > u$  est distribuée selon une loi GPD de paramètres  $\sigma + \gamma(u - \mu)$  et  $\gamma \in \mathbb{R}$ , on a :

$$e(u) = \frac{\sigma + \gamma(u - \mu)}{1 - \gamma} = a + bu$$

où  $a = (\sigma - \mu\gamma)/(1 - \gamma)$  et  $b = \frac{\gamma}{1 - \gamma}$ .

Ainsi, si le graphe de la fonction  $\hat{e}(u)$  semble avoir un comportement linéaire au-delà d'une certaine valeur  $u$ , cela signifie que les excès suivent une loi GPD.

## Application à la sélection du seuil

A travers quelques exemples de mise en œuvre sur données simulées et réelles, nous présentons un moyen d'interpréter le comportement du graphe de la fonction moyenne des excès empirique au-delà d'un seuil  $u$  (Mean excess function *mef* plot), mais aussi de sélection du seuil.

**Exemple I.3.1.** Sous  $R$ , nous simulons un échantillon de taille 10000 de la loi  $\mathcal{U}_{[0,1]}$  :

```
> x<-runif(10000)
```

dont on représente dans la figure I.3.1, le graphe de la fonction moyenne des excès empirique relative à l'échantillon simulé

```
> meplot(x, labels=title(xlab="seuil u",ylab="Moyenne des Excès"))
```

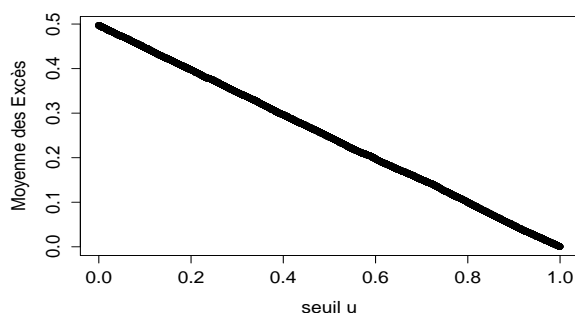


FIGURE I.4 – Fonction moyenne des excès empirique (noir) *v.s* théorique (rouge) de la loi  $\mathcal{U}_{[0,1]}$

**Interprétation :** théoriquement, la fonction moyenne des excès au-delà d'un seuil  $u \in ]0, 1]$  pour une *v.a*  $X \sim \mathcal{U}_{[0,1]}$  est donnée par

$$e_1(u) = \mathbb{E}(X - u | X > u) = \frac{1}{2} - \frac{u}{2},$$

car  $f_{X|X>u}(x) = \frac{1}{1-u} \mathbb{1}_{\{x>u\}}$ . Nous constatons que l'allure du tracé de la fonction moyenne des excès empirique  $\hat{e}(u)$  (en noir) coïncide parfaitement avec le comportement décroissant linéaire décrit par la fonction  $e_1$ . Le tableau I.1 donne l'ajustement à un modèle POT pour un choix raisonnable du seuil  $u$  entre 0.7 et 0.90 avec un pas de 0.05. Nous suggérant une valeur de seuil  $u$  égale à 0.80, qui correspond à une racine carrée de l'erreur standard (*se*) la plus petite, relativement à l'estimation des paramètres  $\sigma$  et  $\gamma$ .

$u$	$N_u$	$\hat{\gamma} (\sqrt{se_{\hat{\gamma}}})$	$\hat{\sigma} (\sqrt{se_{\hat{\sigma}}})$
0.70	2991	-0.9808 (0.0017)	0.2942 (0.0017)
0.75	2483	-0.9818 (0.0015)	0.2454 (0.0015)
0.80	1985	-0.9788 (0.0014)	0.1957 (0.0014)
0.85	1478	-0.9537 (0.0015)	0.1430 (0.0016)
0.90	986	-0.9689 (0.0019)	0.0968 (0.0020)

TABLE I.1 – Estimation du modèle POT (ou GPD) en fonction de  $u$ , pour une loi  $\mathcal{U}_{[0,1]}$

**Exemple I.3.2.** Exemple sur données simulés de la loi exponentielle  $\mathcal{E}(\lambda)$  :

Procédons de la même manière que dans l'exemple précédent, à la simulation d'un échantillon de taille

```

n = 10000, suivant la loi exponentielle de paramètre  $\lambda = 1.5$ 
> x<-rexp(10000,rate=1.5)
puis nous traçons le graphe de la fonction moyenne des excès relative à l'échantillon simulé
> meplot(x,labels=title(xlab="seuil u",ylab="Moyenne des Excès"))

```

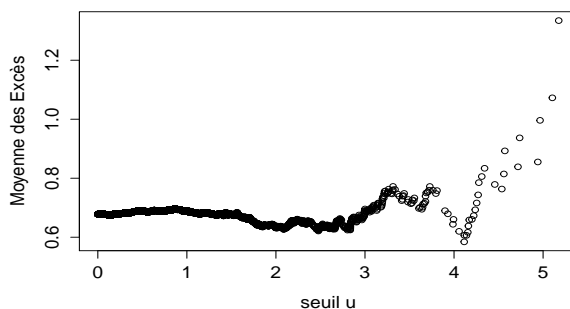


FIGURE I.5 – Fonction moyenne des excès empirique (noir) *v.s* théorique (rouge) de la loi  $\mathcal{E}(1.5)$

**Interprétation :** Théoriquement, pour une loi exponentielle de paramètre  $\lambda$ , la fonction moyenne des excès  $e(u)$  au-delà d'un seuil  $u > 0$ , reste constante et égale à  $1/\lambda$ . Sur la figure I.5, on distingue deux phases de variation. Une première phase  $u \in ]0, 1.5]$  où la fonction moyenne des excès empirique  $\hat{e}(u)$  suit la tendance attendue autour de  $\lambda^{-1} = \frac{2}{3}$ . La deuxième phase pour  $u > 1.5$ , se distingue quant à elle, par un comportement plus volatil et instable. Nous confrontons ces phases en estimant le modèle POT pour des valeurs de  $u = 1.0, 1.25, 1.5, 1.75$  et  $2.0$

$u$	$N_u$	$\hat{\gamma} (\sqrt{se_{\hat{\gamma}}})$	$\hat{\sigma} (\sqrt{se_{\hat{\sigma}}})$
1.00	3757	0.0091 (0.1257)	0.9943 (0.1503)
1.25	2963	0.0247 (0.1362)	0.9660 (0.1587)
1.50	2293	0.0306 (0.1462)	0.9630 (0.1697)
1.75	1770	0.0294 (0.1558)	0.9722 (0.1816)
2.00	1381	0.0317 (0.1657)	0.9718 (0.1836)

TABLE I.2 – Estimation du modèle POT (ou GPD) en fonction de  $u$ , pour une loi  $\mathcal{E}(1.5)$

Même si un seuil  $u = 1$  correspond au meilleur ajustement du modèle POT, mais il semble être trop petit vu le nombre d'observations qui lui sont inférieures (soit moins de deux tiers de l'échantillon total). un compromis serait de prendre un seuil de juste meilleur ( $u = 1.5$ ) qui garanti un bon ajustement à la loi GPD, et aussi un nombre raisonnable d'excès (soit moins du quart).

**Exemple I.3.3.** Exemple sur données simulés de la loi de Pareto  $\mathcal{P}(\gamma)$  :

Simulons un échantillon de taille 10000 de la loi de Pareto  $\mathcal{P}(0.5)$ , puis nous traçons le graphe de la fonction moyenne des excès relative à l'échantillon obtenu :

```

> x<-(1-runif(10000))^(0.5)
> meplot(x,labels=title(xlab="seuil u",ylab="Moyenne des Excès"))

```

**Interprétation :** Théoriquement, pour une loi de Pareto, de paramètre  $\gamma > 0$ , la fonction moyenne des excès  $e(u)$  au-delà d'un seuil  $u > 0$ , est donnée par

$$e_3(u) = \mathbb{E}(X - u | X > u) = \frac{u}{1 - \gamma} - u,$$

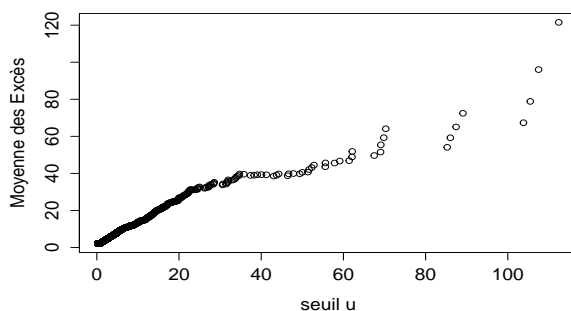


FIGURE I.6 – Fonction moyenne des excès empirique (noir) *v.s* théorique (rouge) de la loi de Pareto  $\mathcal{P}(0.5)$

$$\text{car } F_{X|X>u}(x) = 1 - \left(\frac{x}{u}\right)^{-1/\gamma} \mathbb{1}_{\{x>u\}}$$

Ainsi pour notre exemple, le comportement typique attendu de la fonction moyenne des excès au-delà du seuil  $u$  devrait être une droite d'équation  $e(u) = u$ . Le tracé de la fonction moyenne des excès empirique dans la figure I.6 fait ressortir une allure linéaire, pour des valeurs de  $u$  comprises entre 1 et 20. Le tableau I.3 résume l'ajustement d'un modèle POT pour différentes valeurs de  $u$

$u$	$N_u$	$\hat{\gamma} (\sqrt{se_{\hat{\gamma}}})$	$\hat{\sigma} (\sqrt{se_{\hat{\sigma}}})$
6.0	278	0.5789 (0.3111)	2.9557 (0.5663)
8.0	158	0.5687 (0.3609)	4.1327 (0.7760)
10.0	97	0.4445 (0.3876)	6.5560 (1.0733)
12.0	71	0.3865 (0.4025)	8.2449 (1.2692)
15.0	50	0.4196 (0.4558)	9.0930 (1.4906)

TABLE I.3 – Estimation du modèle POT (ou GPD) en fonction de  $u$ , pour une loi  $\mathcal{P}(0.5)$

Un choix de compromis du seuil  $u$  serait une valeur autour de 8 combinant à la fois un nombre assez grands d'excès et une valeur de seuil la plus grande possible (on constate aussi que cette valeur de seuil permet la meilleure estimation du paramètre  $\gamma$  cible égale à 0.5).

**Exemple I.3.4.** *Exemple sur un jeu données réelles - Danish fire insurance, disponible dans le package R `evir` :*

*Ces données décrivent les montants de sinistres relatifs à une garantie incendie & catastrophes naturelles d'une compagnie Danoise, couvrant la période du 03 janvier 1980 au 31 décembre 1990. Les montants sont exprimés en couronne danoise (DKK), dont voici un brève descriptif*

<i>Min.</i>	<i>1st Qu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rd Qu.</i>	<i>Max.</i>
1.000	1.321	1.778	3.385	2.967	263.30

*Traçons le graphe de la fonction moyenne des excès relatif aux données danish fire dont on ignore la nature de la loi théorique*

```
> data(danish)
> meplot(danish, labels=title(xlab="seuil u", ylab="Moyenne des Excès"))
```

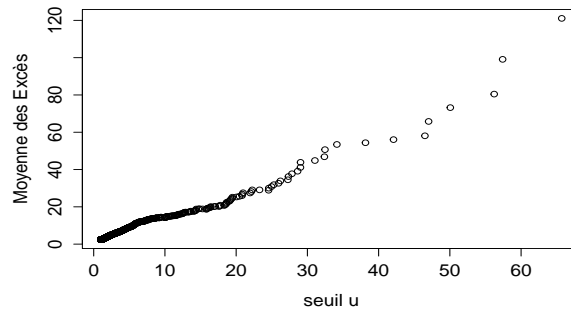


FIGURE I.7 – Fonction moyenne des excès empirique relative aux données danish fire

**Interprétation :** L'allure du tracé de la fonction moyenne des excès empirique  $\hat{e}(u)$  dans la figure I.7, semble avoir une tendance linéaire pour des valeurs de seuil comprise entre 0 et 8, puis une croissance presque parabolique pour  $u > 8$ . Ce comportement est proche de celui observé dans le cas d'une distribution de type Pareto (voir l'exemple I.3.3). Le tableau I.4 résume l'ajustement de la sinistralité au delà d'un seuil  $u > 0$  par un modèle GPD, pour différentes valeurs de  $u$  :

$u$	$N_u$	$\hat{\gamma} (\sqrt{se_{\hat{\gamma}}})$	$\hat{\sigma} (\sqrt{se_{\hat{\sigma}}})$
5.0	254	0.6320 (0.1117)	3.8074 (0.4637)
6.0	186	0.4701 (0.1069)	5.8484 (0.7310)
8.0	131	0.4087 (0.1104)	7.6225 (1.0383)
10.0	109	0.4968 (0.1362)	6.9745 (1.1131)
12.0	85	0.5210 (0.1553)	7.5516 (1.3659)

TABLE I.4 – Estimation du modèle POT (ou GPD) en fonction de  $u$ , pour les données Danish fire

Un choix du seuil  $u$  autour de 6 semble être plus judicieux, car il combine à la fois une valeur de seuil la plus grande possible tout en respectant une erreur standard d'estimation tolérable.

# CHAPITRE II

## INFÉRENCE STATISTIQUE FONDAMENTALE DANS LA THÉORIE DES VALEURS EXTRÊMES



“

... *La mesure d'une erreur est en même temps la mesure de la vérité correspondante ...*

”

---

Claude de Saint-Martin , *Esprit des choses* , 1743 - 1803

“

... *Dans toute statistique, l'inexactitude du nombre est compensée par la précision des décimales. ...*

”

---

Alfred Sauvy, *Bulletin Rouge-Brique*, 1898 - 1990

**D**ans ce chapitre, nous tentons d'explorer brièvement les différents apports d'inférence statistique dans la théorie des valeurs extrêmes uni-variée. Une revue de la littérature mettant l'accent essentiellement sur les modèles d'inférence semi-paramétrique dans le cas d'échantillon complet, ainsi que dans le cas de présence de censure aléatoire à droite.

La représentation standard de la distribution des valeurs extrêmes  $H_\gamma$ , montre l'importance du paramètre dit indice des valeurs extrêmes  $\gamma$ , ou respectivement l'argument dit indice de queue défini par  $\alpha = 1/\gamma$  - conditionnant ainsi le max-domaine d'attraction de la distribution -

De même, une question cruciale est l'estimation de la probabilité d'occurrence d'un évènement rare ou extrême. Cela revient aussi à l'estimation de la période de retour d'un tel évènement. D'une manière générale, la réponse à une telle question est fournie par l'estimation des quantiles extrêmes.

## II.1 Modèles semi-paramétriques de premier, second et d'ordre élevé

Nous considérons une fonction  $F$  supposée absolument continue, et  $U$  la fonction quantile de queue telle que définit dans [I.2.1](#). Une condition nécessaire et suffisante pour que  $F$  soit dans le max-domaine de  $H_\gamma$ , est celle proposée par de Haan (1984), comme suit :

$$F \in MDA(H_\gamma) \Leftrightarrow \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \varphi_\gamma(x) := \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{si } \gamma \neq 0 \\ \ln x & \text{si } \gamma = 0 \end{cases} \quad (\text{II.1})$$

pour tout  $x > 0$ ; où  $a$  est une fonction mesurable positive. La condition dans [II.1](#) est connue sous le nom de la condition du premier ordre. À titre d'exemple, si  $F \in MDA(\Phi_\gamma)$ ; on peut choisir  $a(t) = \gamma U(t)$  pour avoir  $\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma$ . Cela signifie que  $U$  (resp.  $\bar{F}$ ) soit une fonction à variation régulière d'indice  $\gamma > 0$  (resp. d'indice  $-1/\gamma$ ).

Toutefois, une condition de premier ordre n'est pas suffisante en générale afin d'établir les propriétés asymptotiques des estimateurs qu'on peut dégager à partir d'un tel modèle semi-paramétrique (en particulier la normalité asymptotique). En pratique, nous nécessitons une condition du second ordre (ou d'ordre plus élevé) pour spécifier la vitesse de convergence dans ([II.1](#)). Plusieurs écritures de la dite condition du second ordre existent dans la littérature, l'exprimant soit en fonction de  $\bar{F}$ ; de  $U$  ou même en  $\ln U$ , mais la commune est celle donnée par :

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \begin{cases} \frac{1}{\rho} \left( \frac{x^\rho - 1}{\rho} - \ln x \right) & \text{si } \gamma = 0, \rho \neq 0 \\ \frac{1}{\gamma} \left( x^\gamma \ln x - \frac{x^\gamma - 1}{\gamma} \right) & \text{si } \gamma \neq 0, \rho = 0 \\ \ln^2(x/2) & \text{si } \gamma = \rho = 0 \\ \frac{1}{\rho} \left( \frac{x^{\rho+\gamma} - 1}{\rho+\gamma} - \frac{x^\gamma - 1}{\gamma} \right) & \text{sinon} \end{cases} \quad (\text{II.2})$$

pour tout  $x > 0$ ; où  $\rho \leq 0$  est le paramètre du second ordre, et  $A$  une fonction à signe constant tendant vers 0 lorsque  $t \rightarrow \infty$ , telle que  $|A| \in \mathcal{R}_\rho$ . de même; et afin d'avoir la vitesse de convergence dans ([II.2](#)), nous aurons besoin d'un modèle semi-paramétrique d'ordre trois ou voir plus. Pour  $\gamma > 0$  et en choisissons  $a(t) = \gamma U(t)$ , l'équation ([II.2](#)) peut s'écrire sous la forme

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^{\rho\gamma} - 1}{\rho\gamma} \quad (\text{II.3})$$

## II.2 Estimation semi-paramétrique dans le cas d'un échantillon i.i.d complet

Comme mentionné dans la section précédente, sous un modèle semi-paramétrique, on n'a pas besoin d'ajuster un modèle paramétrique (basé sur un paramètre de localisation, d'échelle et de forme); mais juste supposer que la fonction  $F$  est dans le max-domaine d'attraction avec un indice  $\gamma$  approprié. En suite, la construction de l'estimateur se base exclusivement sur les  $k$ -plus grandes observations de l'échantillon supposé i.i.d  $X_1, X_2, \dots, X_n$ ; où  $k$  est une suite intermédiaire (théorique dont on discutera la mise en œuvre pratique à la fin de ce chapitre); telle que  $k_n \rightarrow \infty, k_n \rightarrow o(n)$  quand  $n \rightarrow \infty$ . Nous présentons dans cette section, une revue des estimateurs les plus répandus dans la littérature, ainsi que leurs propriétés asymptotiques.

## II.2.1 Estimation de l'indice des valeurs extrêmes

### Estimateur de Hill (1975)

Parmi les estimateurs les plus répandus, on peut citer l'estimateur introduit par Hill (1975), et défini comme suit :

**Définition II.2.1.** Soit  $X_{1,n}, \dots, X_{n,n}$  les statistiques d'ordre associées à l'échantillon  $X_1, X_2, \dots, X_n$  issu de  $F$ . On appelle estimateur de Hill, la quantité donnée par :

$$\hat{\gamma}_H(k) := \frac{1}{k} \sum_{i=1}^k \log(X_{n-i+1,n}) - \log(X_{n-k,n}) \quad (\text{II.4})$$

Cet estimateur se construit de différentes façons, une de ces approches est celle qui consiste à définir la queue d'une distribution de type Pareto par la dite condition de premier ordre :

$$\frac{1 - F(xu)}{1 - F(u)} \rightarrow x^{-1/\gamma} \text{ quand } u \rightarrow \infty, x > 1$$

En effet, cette définition découle du fait que  $F$  soit une fonction à variation régulière.

Soit  $Z_j = \frac{X_i}{u} | X_i > u$ , l'échantillon des excès relatifs au-delà d'un seuil  $u > 0$  (On note  $N_u$  le nombre des excès au delà de  $u$ ). La log-vraisemblance issue de l'échantillon des  $Z_i$  est donnée par :

$$\log \mathcal{L}(Z_1, \dots, Z_{N_u}; \gamma) = -N_u \log \gamma - \left(1 + \frac{1}{\gamma}\right) \sum_{j=1}^{N_u} \log(Z_j)$$

L'estimateur du maximum de vraisemblance est donné par :

$$\hat{\gamma} = \frac{1}{N_u} \sum_{j=1}^{N_u} \log(Z_j)$$

En choisissant pour niveau d'excès  $u$  la  $(n-k)$ <sup>ème</sup> statistique d'ordre  $X_{n-k,n}$ , on retrouve l'expression de l'estimateur de Hill défini dans (II.4).

### Estimateur de Pickands (1975)

Introduit par Pickands (1975), Cet estimateur est considéré comme une généralisation de l'estimateur de Hill. Il a l'avantage d'être utilisé pour l'estimation de l'indice des valeurs extrêmes dans le cas des trois types de domaine d'attraction.

**Définition II.2.2.** Soit  $X_1, X_2, \dots, X_n$  un échantillon de v.a issues de  $F$  et  $X_{1,n}, \dots, X_{n,n}$  l'échantillon des statistiques d'ordre associées. On appelle estimateur de Pickands la quantité donnée par :

$$\hat{\gamma}_P(k) := \frac{1}{\log(2)} \log \left( \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right), \quad (\text{II.5})$$

pour  $k = 1, \dots, \lfloor \frac{n}{4} \rfloor$ .

La construction du dit estimateur, repose sur la notion de classe de Hall introduite dans Hall (1982). Une distribution de probabilité  $F$  est dite de classe de Hall d'ordre 1 si :

$$F(t) = 1 - ct^{-1/\gamma}$$

On pose  $u(t) = F^{-1}(1 - 1/t) = Kt^\gamma$ , avec  $K = c^{-1}$  et on définit le rapport de l'espacement entre le quantile extrême d'ordre  $(1 - k/n)$  et celui d'ordre  $(1 - 2k/n)$ , à l'espacement entre le quantile extrême d'ordre  $(1 - 2k/n)$  et d'ordre  $(1 - 4k/n)$ , par

$$\Delta(k) := \frac{u(n/k) - u(n/2k)}{u(n/2k) - u(n/4k)}.$$

En remplaçant  $u(n/k)$  par son estimateur empirique  $X_{n-k,n}$  dans  $\Delta(k)$ , on obtient :

$$\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \sim 2^{\hat{\gamma}}$$

### Estimateur des moments (1989)

Dekkers *et al.* (1989) considèrent le problème d'estimation de l'indice des valeurs extrêmes  $\gamma$  pour toute valeur dans  $\mathbb{R}$ . Ils proposèrent ainsi une généralisation de l'estimateur de Hill, en introduisant un estimateur par la méthode des moments.

**Définition II.2.3.** L'estimateur des moments de  $\gamma \in \mathbb{R}$ , basé sur le  $j^{\text{ème}}$  moment  $M_j(u)$  des Log-excès relatifs ( $\log \frac{X}{u} | X > u$ ), est donné par

$$\hat{\gamma}_M(k) := M_1(X_{n-k,n}) + 1 - \frac{1}{2} \left[ 1 - \frac{(M_1(X_{n-k,n}))^2}{M_2(X_{n-k,n})} \right], \quad (\text{II.6})$$

où  $M_j(X_{n-k,n}) = \frac{1}{k} \sum_{i=1}^k [\log(X_{n-i+1,n}) - \log(X_{n-k,n})]^j$ , pour  $j = 1, 2$ .

Pour plus de détail sur la construction ainsi que les propriétés asymptotiques de cet estimateur, nous recommandons aux lecteurs de lire le papier [25].

### Normalité asymptotique des estimateurs classiques

**Théorème II.2.1.** Soit  $k = k(n)$  une suite d'entiers telle que  $k(n) \rightarrow \infty$  et  $k(n)/n \rightarrow 0$  quand  $n \rightarrow \infty$ . Supposons qu'il existe une fonction  $A$  telle que  $\lim_{k \rightarrow \infty} \sqrt{k} A(\frac{n}{k}) = \lambda \in \mathbb{R}$ , pour laquelle l'équation (II.2) est vérifiée, Alors :

1.  $\sqrt{k}(\hat{\gamma}_H(k) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1 - \rho\gamma}, \gamma^2\right)$
2.  $\sqrt{k}(\hat{\gamma}_P(k) - \gamma) \xrightarrow{d} \mathcal{N}\left(\lambda \frac{2^{\rho\gamma} - 1}{(2^\gamma - 1)\rho\gamma \log 2} (2^{(\rho+1)\gamma} - 1), \frac{\gamma^2(2^{2\gamma+1} + 1)}{((2^\gamma - 1) \log 2)^2}\right)$
3.  $\sqrt{k}(\hat{\gamma}_M(k) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1 - \rho\gamma} \times \left(1 + \frac{\rho}{1 - \rho\gamma}\right), 1 + \gamma^2\right)$

**Remarque II.2.1.** Nous constatons à travers le comportement asymptotique décrit dans le théorème précédent que :

- si  $\rho = 0$ , alors les trois estimateurs ont le même biais  $\lambda/\sqrt{k}$ .
- $\gamma_H(k)$  à la plus petite variance  $\gamma^2$ .
- L'erreur quadratique moyenne asymptotique (AMSE) est respectivement donnée pour  $\gamma_H(k)$ ,  $\gamma_P(k)$  et  $\gamma_M(k)$ , par

1.  $AMSE(\gamma_H(k)) = k^{-1} \left( \gamma^2 + \left( \frac{\lambda}{1 - \rho\gamma} \right)^2 \right)$ .
2.  $AMSE(\gamma_P(k)) = k^{-1} \left( \frac{\gamma^2(2^{2\gamma+1} + 1)}{((2^\gamma - 1) \log 2)^2} + \left( \lambda \frac{2^{\rho\gamma} - 1}{(2^\gamma - 1)\rho\gamma \log 2} (2^{(\rho+1)\gamma} - 1) \right)^2 \right)$ .
3.  $AMSE(\gamma_M(k)) = k^{-1} \left( 1 + \gamma^2 + \left( \frac{\lambda}{1 - \rho\gamma} \times \left( 1 + \frac{\rho}{1 - \rho\gamma} \right) \right)^2 \right)$ .

*Démonstration.* Soit  $Y_1, Y_2, \dots, Y_n$  un échantillon i.i.d issu d'une loi commune donnée par sa fonction de répartition  $G(x) = 1 - x^{-1}, x \geq 1$ , et  $Y_{1,n}, Y_{2,n}, \dots, Y_{n,n}$  l'échantillon associé des statistiques d'ordre. Notons que si  $X_1, X_2, \dots, X_n$  sont issues d'une loi  $F(x)$ , et  $U(t) = F^{\leftarrow}(1 - \frac{1}{t})$  est la fonction quantile extrême correspondante, alors  $X_{i,n} \stackrel{d}{=} U(Y_{i,n})$  pour tout  $i = 1$  à  $n$ . Afin de montrer le résultat dans le théorème II.2.1, nous aurons besoin d'énoncer le Lemme 3.1 de [Dekkers et de Haan](#) (1993)

**Lemme II.2.1.** *Soit une suite d'entiers  $k = k(n)$  telle que  $0 < k(n) < n$ . Lorsque  $k(n) \rightarrow \infty$  et  $k(n)/n \rightarrow 0$  quand  $n \rightarrow \infty$ , on a*

1.  $\frac{Y_{n-k,n}}{(n/k)} \xrightarrow{\mathbb{P}} 1$ ,

2. Le vecteur aléatoire  $(P_n, Q_n)$  donné par

$$P_n =: \sqrt{k} \left\{ \frac{1}{k} \sum_{i=1}^k \log Y_{n-i+1,n} - \log Y_{n-k,n} - 1 \right\}$$

et

$$Q_n =: \sqrt{k} \left\{ \frac{1}{k} \sum_{i=1}^k (\log Y_{n-i+1,n} - \log Y_{n-k,n})^2 - 2 \right\}$$

est asymptotiquement gaussien de moyenne nulle, et de matrice variance-covariance  $\Sigma = \begin{pmatrix} 1 & 4 \\ 4 & 20 \end{pmatrix}$

**Preuve de 1.** À partir de l'équation du second ordre dans (II.3), on a

$$\log U(tx) - \log U(t) = \gamma \log x + A(t) \frac{x^{\rho\gamma} - 1}{\rho\gamma} (1 + o_{\mathbb{P}}(1)), \text{ quand } t \rightarrow \infty$$

En combinant ce résultat avec le Lemme II.2.1, on obtient

$$\begin{aligned} \hat{\gamma}^H(k) &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \log U \left( \frac{Y_{n-i+1,n}}{Y_{n-k,n}} Y_{n-k,n} \right) - \log U(Y_{n-k,n}) \\ &= \gamma + \frac{1}{k} \sum_{i=1}^k \left\{ \gamma \left( \log \frac{Y_{n-i+1,n}}{Y_{n-k,n}} - 1 \right) + A(Y_{n-k,n}) \frac{(Y_{n-i+1,n}/Y_{n-k,n})^{\rho\gamma} - 1}{\rho\gamma} \right\} + o_{\mathbb{P}} \left( A \left( \frac{n}{k} \right) \right) \\ &= \gamma + \gamma \frac{P_n}{\sqrt{k}} + d_1 A(Y_{n-k,n}) + o_{\mathbb{P}} \left( A \left( \frac{n}{k} \right) \right) \end{aligned}$$

$$\text{où } d_1 = \int_1^\infty \frac{x^{\rho\gamma} - 1}{\rho\gamma} \frac{dx}{x^2} = \frac{1}{1 - \rho\gamma}. \text{ (c.q.f.d)}$$

**Preuve de 2.** Partons de la relation (4.8) dans [de Haan et al.](#) (1993), on a

$$\sqrt{k} \left( \frac{k}{n} Y_{n-[kx],n} - x^{-1} \right)$$

converge dans  $D(0, \infty]$  vers  $x^{-2}\mathbb{B}(x)$ , où  $\mathbb{B}$  est un mouvement brownien. La convergence de  $\frac{k}{n}Y_{n-[kx],n} \rightarrow x^{-1}$  en probabilité quand  $n \rightarrow \infty$ , combinée à l'équation (II.3)

$$\lim_{n \rightarrow \infty} \frac{\frac{U(Y_{n-[kx],n})}{U(n/k)} - \left(\frac{k}{n}Y_{n-[kx],n}\right)^\gamma}{A(n/k)} = x^{-\gamma} \frac{x^{-\rho\gamma} - 1}{\rho\gamma} \quad (\text{II.7})$$

uniformément et localement dans  $D(0, \infty]$ . Cependant, l'équation II.7 nous permet d'écrire

$$\lim_{t \rightarrow \infty} \sqrt{k} \left( \frac{U(Y_{n-[kx],n})}{U(n/k)} - x^\gamma \right) = \lim_{t \rightarrow \infty} \sqrt{k} \left( \frac{U(Y_{n-[kx],n})}{U(n/k)} - \left(\frac{k}{n}Y_{n-[kx],n}\right)^\gamma \right) + \sqrt{k} \left( \left(\frac{k}{n}Y_{n-[kx],n}\right)^\gamma - x^{-\gamma} \right)$$

qui converge uniformément dans  $D(0, \infty]$  vers

$$\lambda x^{-\gamma} \frac{x^{-\rho\gamma} - 1}{\rho\gamma} + \gamma x^{-1-\gamma} \mathbb{B}(x).$$

Ainsi,

$$\sqrt{k} \left( \frac{X_{n-[kx],n}}{U(n/k)} - x^\gamma \right) \rightarrow \lambda x^{-\gamma} \frac{x^{-\rho\gamma} - 1}{\rho\gamma} + \gamma x^{-1-\gamma} \mathbb{B}(x). \quad (\text{II.8})$$

uniformément dans  $D(0, \infty]$ . À partir de l'équation (II.8), on a

$$\sqrt{k} \left\{ \frac{X_{n-[k/4],n} - X_{n-[k/2],n}}{U(n/k)} - 2^\gamma(2^\gamma - 1) \right\} \xrightarrow{d} \gamma 4^{\gamma+1} \mathbb{B}(1/4) + \lambda 4^\gamma \frac{4^{\rho\gamma} - 1}{\rho\gamma} - \gamma 2^{\gamma+1} \mathbb{B}(1/2) - \lambda 2^\gamma \frac{2^{\rho\gamma} - 1}{\rho\gamma}$$

et

$$\sqrt{k} \left\{ \frac{X_{n-[k/2],n} - X_{n-k,n}}{U(n/k)} - (2^\gamma - 1) \right\} \xrightarrow{d} \gamma 2^{\gamma+1} \mathbb{B}(1/2) + \lambda 2^\gamma \frac{2^{\rho\gamma} - 1}{\rho\gamma} - \gamma \mathbb{B}(1).$$

D'où

$$\begin{aligned} & \sqrt{k} \left\{ \frac{X_{n-[k/4],n} - X_{n-[k/2],n}}{X_{n-[k/2],n} - X_{n-k,n}} - 2^\gamma \right\} \\ &= \sqrt{k} \left\{ \frac{X_{n-[k/4],n} - X_{n-[k/2],n}}{U(n/k)} \frac{U(n/k)}{X_{n-[k/2],n} - X_{n-k,n}} - 2^\gamma \right\} \\ &= \sqrt{k} \left\{ \left[ \frac{X_{n-[k/4],n} - X_{n-[k/2],n}}{U(n/k)} - 2^\gamma(2^\gamma - 1) \right] \frac{U(n/k)}{X_{n-[k/2],n} - X_{n-k,n}} \right. \\ &\quad \left. + 2^\gamma(2^\gamma - 1) \left[ \frac{U(n/k)}{X_{n-[k/2],n} - X_{n-k,n}} - \frac{1}{2^\gamma - 1} \right] \right\} \\ &\xrightarrow{d} \frac{1}{2^\gamma - 1} \left\{ \gamma 4^{\gamma+1} \mathbb{B}(1/4) + \lambda 4^\gamma \frac{4^{\rho\gamma} - 1}{\rho\gamma} - \gamma 2^{\gamma+1} \mathbb{B}(1/2) - \lambda 2^\gamma \frac{2^{\rho\gamma} - 1}{\rho\gamma} \right\} \\ &\quad - \frac{2^\gamma}{2^\gamma - 1} \left\{ \gamma 2^{\gamma+1} \mathbb{B}(1/2) + \lambda 2^\gamma \frac{2^{\rho\gamma} - 1}{\rho\gamma} - \gamma \mathbb{B}(1) \right\} \\ &= \frac{2^\gamma}{2^\gamma - 1} \left\{ \mathbb{B}(1) - (2 + 2^{\gamma+1}) \mathbb{B}(1/2) + 2^{\gamma+2} \mathbb{B}(1/4) \right\} \\ &\quad + \frac{\lambda 2^\gamma}{2^\gamma - 1} \frac{2^{\rho\gamma} - 1}{\rho\gamma} (2^{\gamma+\rho\gamma} - 1) \\ &\stackrel{d}{=} \mathcal{N} \left( \frac{2^\gamma}{2^\gamma - 1} \frac{2^{\rho\gamma} - 1}{\rho\gamma} (2^{\gamma+\rho\gamma} - 1), \frac{\gamma 2^{2\gamma}}{(2^\gamma - 1)^2} (1 + 2^{2\gamma+1}) \right) \end{aligned}$$

**Preuve de 3.** À partir de l'équation du second ordre dans (II.3), on a

$$[\log U(tx) - \log U(t)]^2 = \gamma^2 \log^2 x + 2\gamma \log x A(t) \frac{x^{\rho\gamma} - 1}{\rho\gamma} + o_{\mathbb{P}}(A(t)) \text{ quand } t \rightarrow \infty$$

En reprenant la même démarche que dans la **preuve de 1.** mais par rapport à  $(\hat{\gamma}^H(k))^2$  et l'équation précédente, on obtient

$$(\hat{\gamma}^H(k))^2 \stackrel{d}{=} \gamma^2 + 2\gamma^2 \frac{P_n}{\sqrt{k}} + 2\gamma d_1 A(Y_{n-k,n}) + o_{\mathbb{P}}\left(A\left(\frac{n}{k}\right)\right)$$

et de manière similaire,

$$\begin{aligned} M_2(X_{n-k,n}) &= \frac{1}{k} \sum_{i=1}^k [\log(X_{n-i+1,n}) - \log(X_{n-k,n})]^2 \\ &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \left\{ \gamma^2 \left( \log^2 \frac{Y_{n-i+1,n}}{Y_{n-k,n}} \right) + 2\gamma \left( \log \frac{Y_{n-i+1,n}}{Y_{n-k,n}} \right) A(Y_{n-k,n}) \frac{\left( \frac{Y_{n-i+1,n}}{Y_{n-k,n}} \right)^{\rho\gamma} - 1}{\rho\gamma} \right\} \\ &\quad + o_{\mathbb{P}}\left(A\left(\frac{n}{k}\right)\right) \\ &= 2\gamma^2 + \gamma^2 \frac{Q_n}{\sqrt{k}} + d_2 A(Y_{n-k,n}) + o_{\mathbb{P}}\left(A\left(\frac{n}{k}\right)\right) \end{aligned}$$

où  $d_2 = \int_1^{\infty} \frac{x^{\rho\gamma-1}}{\rho\gamma} \log x \frac{dx}{x^2} = \frac{2\gamma(2-\rho\gamma)}{(1-\rho\gamma)^2}$ . Finalement, on en déduit par définition de l'estimateur des moments

$$\hat{\gamma}^M(k) \stackrel{d}{=} \gamma + \frac{Q_n}{2\sqrt{k}} + (\gamma - 2) \frac{P_n}{\sqrt{k}} \left( \frac{d_2}{2\gamma^2} + \frac{\gamma - 2}{\gamma} d_1 \right) A(Y_{n-k,n}) + o_{\mathbb{P}}\left(A\left(\frac{n}{k}\right)\right)$$

(c.q.f.d)

□

### Estimateur des excès au-delà d'un seuil aléatoire (1984)

**Davison** (1984) construira son estimateur (noté communément PORT-ML, pour Pick Over Random Threshold, Maximum Likelihood), basé sur les excès au-delà de la statistique d'ordre  $X_{n-k,n}$ . Soit  $D_{i,k} := X_{n-k+i,n} - X_{n-k,n}$ , pour  $1 \leq i \leq k$ , l'échantillon des  $k$  excès au-delà du seuil  $X_{n-k,n}$ . Selon le théorème de Pickands, Balkema et deHaan (1975) (voir I.3.1), la loi des  $D_{i,k}$  est approximée par la loi d'une  $i^{\text{ème}}$  statistique d'ordre associée à un échantillon  $k$  v.a issues d'une loi de Pareto généralisée ayant pour fonction de répartition  $G(y) = 1 - (1 + \alpha y)^{-1/\gamma}$ ,  $\alpha, \gamma \in \mathbb{R}$  et  $y > 0$ .

La solution explicite associée à l'équation d'estimation au sens du maximum de vraisemblance est donnée par

$$\begin{cases} \gamma = \frac{1}{k} \sum_{i=1}^k \log(1 + \alpha D_{i,k}) \\ \frac{k}{\alpha} = \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^k \frac{D_{i,k}}{1 + \gamma D_{i,k}} \end{cases}$$

## II.2.2 Estimateurs généralisés de l'indice des valeurs extrêmes

### Estimateurs à noyau (1985)

Une classe générale de l'estimateur de Hill a été proposé par [Csörgö et al.](#) (1985), estimant l'indice des valeurs extrêmes  $\gamma > 0$  par

$$\hat{\gamma}^{\mathcal{K}}(k) := \frac{\sum_{i=1}^n \mathcal{K}(i/k) [\log X_{n-i+1,n} - \log X_{n-k,n}]}{\sum_{i=1}^n \mathcal{K}(i/k)}$$

où  $\mathcal{K}(\cdot)$  est un noyau non-décroissant défini sur  $\mathbb{R}^+$  tel que  $\int_0^\infty \mathcal{K}(u) du = 1$ .

Par identification, l'estimateur de Hill s'obtient comme un cas particulier de  $\hat{\gamma}^{\mathcal{K}}(k)$ , où  $\mathcal{K}(u) = \mathbb{1}_{]0,1]}(u)$ .

Notons qu'une classe plus générale des estimateurs à noyau, fut proposée par [Groenboom et al.](#) (2003) pour  $\gamma \in \mathbb{R}$ .

Ce type d'estimateurs peut être obtenu aussi par une régression Quantile d'une distribution de Pareto type I par Quantile empirique. Elle consiste à une approximation naïve de la pente de l'extrémité droite du Q-Q plot (voir [Beirlant et al.](#) (1996) et [Oliveira et al.](#) (2006)).

### Estimateur de Hill généralisé (1996)

[Beirlant et al.](#) (1996) proposèrent une forme généralisée de l'estimateur de Hill, s'inspirant d'une analyse de la pente du graphe Quantile-Quantile (QQplot). Valide pour toute valeur de  $\gamma \in \mathbb{R}$ , il est donné par

$$\hat{\gamma}^{GH}(k) := \hat{\gamma}^H(k) + \frac{1}{k} \sum_{i=1}^k [\log \hat{\gamma}^H(i) - \log \hat{\gamma}^H(k)].$$

Les performances asymptotiques et empiriques de  $\hat{\gamma}^{GH}(k)$  sont décrites dans [6].

### Estimateur de Pickands généralisé (1996)

La grande variance asymptotique de l'estimateur de Pickands (une variance asymptotique minimale supérieure à 12 atteinte au voisinage de  $\gamma = -0.5$ ), a motivé une re-formulation fonctionnelle de  $\hat{\gamma}^P(k)$  sous la forme

$$\hat{\gamma}^{GP}(k, \theta) := \frac{1}{\log(\theta)} \log \left( \frac{X_{n-[\theta^2 k],n} - X_{n-[\theta k],n}}{X_{n-[\theta k],n} - X_{n-[k],n}} \right), \text{ pour } 0 < \theta < 1.$$

On reconnaît facilement le cas particulier  $\theta = \frac{1}{2}$ ,  $\hat{\gamma}^{GP}(k, \theta)$  correspond à l'estimateur de Pickands  $\hat{\gamma}^P([k/4])$ . Plusieurs variantes de  $\hat{\gamma}^{GP}(k, \theta)$ , existent dans la littérature, même si celle proposé dans le papier de [Sergers](#) (2005) semble être la plus détaillée en terme de propriétés empiriques et asymptotiques.

### Estimateurs des moments mixés (2009)

[Fraga et al.](#) (2009) proposèrent un généralisation de l'estimateurs des moments  $\hat{\gamma}_M(k)$ , en définissant un nouveau type d'estimateur asymptotiquement gaussien donné par

$$\hat{\gamma}^{GM}(k) := \frac{\hat{\varphi}_n(k) - 1}{1 + 2 \min\{\hat{\varphi}_n(k), 0\}},$$

où  $\hat{\varphi}_n(k) := \frac{\hat{\gamma}^H(k) - L(k)}{(L(k))^2}$  et  $L(k) = \frac{1}{k} \sum_{i=1}^k \left( 1 - \frac{X_{n-k,n}}{X_{n-i+1,n}} \right)$ .

L'estimateur proposé fait intervenir à la fois, une certaine combinaison entre les excès relatifs ainsi que leurs logarithmes Il semble être une alternative pour l'estimation de l'indice des valeurs extrêmes pour  $\gamma \in \mathbb{R}$ .

**Remarque II.2.2.** Les différents estimateurs de l'indice des valeurs extrêmes  $\gamma$ , proposés dans cette section, ont tous une propriété commune : Quand le nombre des  $k$  plus grandes statistique d'ordre intervenant dans l'expression des estimateurs est petit, la variance de l'estimateur sera grande. de l'autre côté, un choix de  $k$  assez grand, induira un biais important dans l'estimation. Ainsi, un choix raisonnable du nombre d'excès  $k$  est un compromis entre réduction de biais et minimisation de la variance des estimateurs. La figure II.1 décrit le comportement empirique en terme de biais (première colonne) et

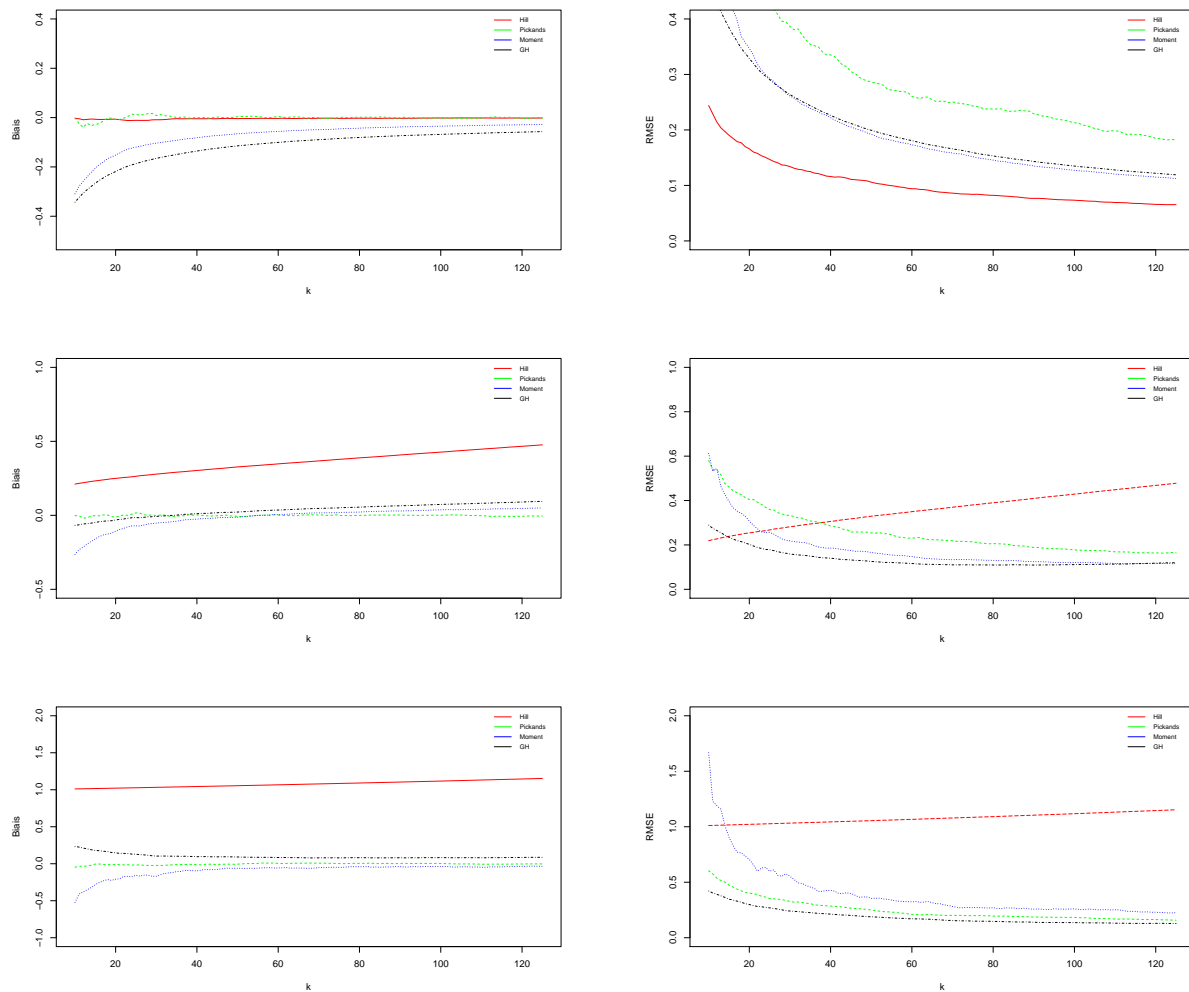


FIGURE II.1 – Comportement empirique de quelques estimateurs de  $\gamma$  (en rouge l'estimateur de Hill, en vert l'estimateur de Pickand, en bleu l'estimateur de Dekker et en noir l'estimateur de Hill généralisé) - **ligne 1** : Distribution Pareto(2), **ligne 2** : Distribution exponentielle standard, **ligne 3** : Distribution uniforme sur  $[0,1]$ .

de RMSE (deuxième colonne), des quatre estimateurs  $\hat{\gamma}^H, \hat{\gamma}^M, \hat{\gamma}^P$  et  $\hat{\gamma}^{GH}$ , à travers trois (03) exemples de distribution (chacune appartient à l'un des 03 max-domaine d'attraction cités dans le chapitre 1) : chaque exemple est issu de 1000 répliquions d'échantillons de taille  $n = 1000$ . Le premier exemple (ligne 1) représente le comportement des estimateurs sus-cités de  $\gamma$  en fonction du nombre d'excès  $k$ , relative-

ment à 1000 échantillon issus d'une loi de Pareto standard de paramètre  $\gamma = 2$  ( $F(x) = 1 - x^{-\frac{1}{\gamma}, x > 1}$ ). Nous constatons que l'estimateur de Hill  $\hat{\gamma}^H$  est meilleur que les trois autres estimateurs proposés en terme de biais et de RMSE. Le deuxième exemple (ligne 2) est relatif à l'estimation de l'indice  $\gamma$  sous une loi d'échantillonnage exponentielle ( $F(x) = 1 - e^{-x}$ , appartenant au max-domaine d'attraction de Gumbel où  $\gamma = 0$ ) : Nous remarquons que l'estimateur  $\hat{\gamma}^H$  ne s'adapte plus à ce type de situation (cela s'explique par le fait qu'il est conçu exclusivement pour des distributions à queue lourde  $\gamma > 0$ ), et nous relevons plutôt l'adéquation d'estimateurs de type  $\hat{\gamma}^M$  et  $\hat{\gamma}^{GH}$  qui semble avoir de meilleures performances en terme de biais et de RMSE. Le dernier exemple (ligne 3) relève le comportement des quatre estimateurs en fonction du nombre d'excès  $k$  relativement à une loi d'échantillonnage uniforme standard ( $F(x) = x\mathbb{1}_{\{0 < x < 1\}} + \mathbb{1}_{\{x \geq 1\}}$ , appartenant au max-domaine d'attraction de Weibull  $\gamma < 0$ ) : nous constatons à travers cet exemple, que l'estimateur de Pickands  $\hat{\gamma}^P$  s'adapte (en terme de biais) mieux à une telle situation lorsque la distribution admet un point terminal supérieur (distribution distinguée par un indice des valeurs extrêmes  $\gamma < 0$ ), même si l'estimateur  $\hat{\gamma}^{GH}$  a de meilleures performances en terme de RMSE.

### II.2.3 Réduction du biais dans l'estimation de l'indice des valeurs extrêmes

$\gamma$

Un des constats marquants, relevé à travers les résultats asymptotiques des estimateurs de l'indice des valeurs extrêmes  $\gamma$ , est la présence d'une composante de biais qui peut s'avérer non-négligeable et néfaste. Une réduction de cette composante est ainsi plus qu'appréciable et nécessaire. L'un des pionniers dans ce domaine fut Peng en 1998, qui proposa pour la première fois l'idée d'un estimateur asymptotiquement sans biais de l'indice des valeurs extrêmes. L'idée a été reprise par Beirlant *et al.* (1999), puis par Gomes *et al.* (2002), où l'estimation se base sur un échantillon de Log-excès relatifs ainsi que le Log-espacements entre une sous-suite extrême de statistiques d'ordre issues d'une distribution de Pareto. Les auteurs de ces travaux ont proposé des estimateurs (dits estimateurs de second ordre à biais réduit) de l'indice des valeurs extrêmes, avec une variance asymptotique plus grande ou égale à  $\gamma^2(1 - \frac{1}{\rho})$  pour  $\rho < 0$ . Cette réduction de biais impacte négativement la variance asymptotique des estimateurs proposés (voir une variance infinie quand  $\rho$  tend vers 0).

Il faudra attendre les travaux de Caeiro *et al.* (2005), pour palier à la contrainte d'augmentation de la variance asymptotique, lorsqu'on réduit le biais des estimateurs. L'estimateur proposé est dit asymptotiquement, sans biais et à variance minimale égale à  $\gamma^2$ . Son expression est obtenue en supposant que la fonction quantile extrême  $U(t)$  appartient à la classe de Hall d'ordre deux (voir [41]), i.e :

$$U(t) = Ct^\gamma(1 + \frac{\gamma\beta t^\rho}{\rho} + o(t^\rho)) \quad (\text{II.9})$$

satisfait la condition du second ordre donnée par (II.2), où  $\gamma > 0, \beta \neq 0$ , et  $\rho \in \mathbb{R}_*$ . L'estimateur à biais réduit proposé par Caeiro (2005) est donné par :

$$\bar{H}(k) = H(k)(1 - \frac{\beta(k) (\frac{n}{k})^{\rho(k)}}{1 - \rho(k)}) \quad (\text{II.10})$$

L'expression de l'estimateur  $\bar{H}(k)$  requiert la connaissance des paramètres  $\rho$  et  $\beta$  du modèle dans (II.9). Fraga Alves (2003) propose une estimation de ces paramètres par

$$\hat{\rho}_n^{(\tau)}(k) = - \left| \frac{3(T_n^{(\tau)}(k) - 1)}{T_n^\tau(k) - 3} \right|$$

où

$$T_n^\tau(k) = \frac{(M_n^{(1)}(k))^\tau - (M_n^{(2)}(k)/2)^{\tau/2}}{(M_n^{(2)}(k)/2)^{\tau/2} - (M_n^{(3)}(k)/6)^{\tau/3}}$$

pour tout  $\tau$  dans  $\mathbb{R}$ ,  $T_n^\tau(k)$  converge vers  $3(1-\rho)/(3-\rho)$ , et

$$\hat{\beta}_{\hat{\rho}}(k) = \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{\nu_n^{(1-\hat{\rho})}(k)N_n^{(1)}(k) - N_n^{(1-\hat{\rho})}(k)}{\nu_n^{(1-\hat{\rho})}(k)N_n^{(1-\hat{\rho})}(k) - N_n^{(1-2\hat{\rho})}(k)}$$

où  $\nu_n^{(\alpha)}(k) = \frac{1}{k} \sum_{i=1}^k \left(\frac{i}{k}\right)^{\alpha-1}$  and  $N_n^{(\alpha)}(k) = \frac{1}{k} \sum_{i=1}^k i(\log X_{n-i+1,n} - \log X_{n-i,n})\left(\frac{i}{k}\right)^{\alpha-1}$ .

La normalité asymptotique de l'estimateur dans (II.10) est établi par [Caeiro et al. \(2005\)](#), pour une suite d'entiers  $k = k(n)$ ,  $k(n) \rightarrow \infty$  et  $k(n)/n \rightarrow 0$  quand  $n \rightarrow \infty$ . Si la condition du second ordre dans (II.2) est satisfaite avec  $\lim_{k \rightarrow \infty} \sqrt{k}A\left(\frac{n}{k}\right) = \lambda \in \mathbb{R}$ , alors

$$\sqrt{k} \left( \bar{H}_{\hat{\rho}, \hat{\beta}}(k) - \gamma \right) \stackrel{d}{=} \gamma Z_k + O_P(\sqrt{k}A(n/k)).$$

## II.2.4 Estimation du quantile extrême pour les distribution à queue lourde

Dans la théorie des valeurs extrêmes univariée, il existe une multitude d'approches portant sur l'estimation des quantiles extrêmes. Deux méthodes paramétriques se dégagent intuitivement des approches classiques citées dans le chapitre 1 : celle des maxima par blocs, qui modélise la distribution des extrêmes par la loi des valeurs extrêmes généralisée (GEV), ainsi que la méthode des excès au-delà d'un seuil modélisée par une distribution de Pareto généralisée (GPD). Une des alternatives d'estimation des plus répondues dans la littérature, est une approche semi-paramétrique proposée par Weissman (1978).

### Quantile extrême des maxima par blocs

L'estimation du quantile extrême par l'approche des maxima par blocs se caractérise par 03 étapes suivantes :

Etape 01 : Cette approche consiste à diviser un échantillon  $X_1, \dots, X_N$  en  $k$  blocs de même taille  $n$  (on suppose  $N = nk$ ), le  $j^{\text{ième}}$  bloc est défini par  $X_{(j-1)n+1}, \dots, X_{(j-1)n+n}$ , de tel sort que la taille des blocs  $n$  et le nombre de blocs  $k$  soient suffisamment grands.

Etape 02 : On calcule le maximum  $M_j$  du bloc  $j$  défini par :

$$M_j = \max \{X_{(j-1)n+1}, \dots, X_{(j-1)n+n}\}, j = 1, \dots, k$$

On suppose que la taille de chaque bloc est grande, alors la condition asymptotique du théorème de Tippet-Fisher-Gnedenko est vérifiée et l'échantillon des maxima obtenu  $M_1, \dots, M_k$  est issu de la loi GEV.

Etape 03 : L'estimateur du quantile extrême d'un niveau  $p$ , par la méthode des maxima par blocs est ainsi :

$$Q_k^{BM}(p) = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\gamma}} \left[ (-\ln(1-p))^{-\hat{\gamma}} - 1 \right]$$

### Quantile extrême par l'approche POT

L'approche des excès au-delà d'un seuil  $u$  consiste à estimer le quantile extrême selon les 04 étapes suivantes :

Etape 01 : Pour un échantillon  $X_1, X_2, \dots, X_n$ , on choisit un certain seuil élevé  $u$ . Notons  $N_u$  le nombre d'observations  $X_i$  dépassant le seuil  $u$ . On définit ainsi les *v.a* des excès,  $Y_j = X_i - u | X_i > u$

Etape 02 : Soit  $Y_1, Y_2, \dots, Y_{N_u}$  un échantillon des excès au-dessus du seuil  $u$ , de distribution conditionnelle

$$F_u(y) = \mathbb{P}[X - u \leq y | X > u] = \frac{F(y + u) - F(u)}{1 - F(u)}, x > 0$$

D'après le théorème de de Haan-Balkema, on a l'approximation de  $F_u(x)$  par :

$$F_u(y) \approx G_{(0, \sigma, \gamma)}(y) \Rightarrow \bar{F}(y) \approx \bar{G}_{(0, \sigma, \gamma)}(y), y > 0, u \rightarrow \infty$$

Etape 03 : La fonction  $F(y)$  peut s'écrire sous la forme :

$$F(x) = (1 - F(u))F_u(y) + F(u)$$

En remplaçant  $y$  par  $x = y + u$ , la fonction de distribution conditionnelle  $F_u$  par la distribution de Pareto généralisée et  $F(u)$  par l'estimateur empirique  $1 - \frac{N_u}{n}$ , nous obtenons la fonction de distribution  $F$

$$F(x) = 1 - \frac{N_u}{n} \left( 1 + \gamma \frac{x - u}{\sigma} \right)^{-1/\gamma}$$

Etape 04 : Finalement, on inverse la fonction de répartition de  $F$ , l'estimateur du quantile extrême d'un niveau  $p$  est donnée par :

$$Q_u^{POT}(p) = u + \frac{\hat{\sigma}}{\hat{\gamma}} \left[ \left( \frac{n}{N_u}(p) \right)^{-\hat{\gamma}} - 1 \right]$$

Où  $\hat{\sigma}$  et  $\hat{\gamma}$  sont les estimateurs des paramètres de la loi GPD centrée.

### Quantile extrême de Weissman (1978)

Cette approche repose sur une caractéristique des fonction de survie appartenant au Max-domaine d'attraction de Fréchet ( $\gamma > 0$ ) :

$$\bar{F}(x) = x^{-1/\gamma} \mathcal{L}(x)$$

Où  $\mathcal{L}$  est une fonction à variation lente. On peut écrire alors :

$$\frac{\bar{F}(x)}{\bar{F}(X_{n-k,n})} = \frac{\bar{\mathcal{L}}(x)}{\bar{\mathcal{L}}(X_{n-k,n})} \left( \frac{x}{X_{n-k,n}} \right)^{-1/\gamma}$$

En considérant que le rapport des fonctions à variation lente est proche de 1, on a :

$$\bar{F}(x) \sim \bar{F}(X_{n-k,n}) \left( \frac{x}{X_{n-k,n}} \right)^{-1/\gamma}$$

Pour en déduire l'estimateur de Weissman d'un niveau  $p$  donnée par :

$$Q_k^W(p) = X_{n-k,n} \left( \frac{n}{k} p \right)^{-\hat{\gamma}_H(k)}$$

La normalité asymptotique de l'estimateur de Weissman est établi par Peng (1998) sous les mêmes conditions que celle permettant d'établir la normalité asymptotique de l'estimateur de Hill.

### Application à l'estimation de la prime de réassurance

Dans la théorie moderne d'actuariat, une exigence typique serait une tarification crédible, adéquate et juste, assurant ainsi l'équilibre de la branche. Cette exigence ne peut être soulevée qu'à travers la connaissance du risque à assurer et la mise en œuvre d'outils performants d'évaluation de ce risque notamment en présence d'une probabilité significative d'occurrence d'évènement extrême lié directement ou indirectement au risque assuré.

Dans la littérature actuarielle, plusieurs principes de tarification sont proposés telle que le principe de la moyenne, principe de la variance, principe de la Value at Risk (VaR) ou du Tail Value at Risk (TVaR), mais le plus répandu reste celui proposé par Wang en 1996 qui se base sur une transformation proportionnelle de la fonction de survie. Ce principe définit une prime de réassurance dite à risque proportionnel (notée  $\Pi_R$ ), d'un risque ré-assurable  $X$  (supposé une variable aléatoire absolument continue, et d'une distribution de probabilité  $F(x)$  sur un support de  $x > 0$ ), comme une combinaison entre la fonction de survie  $S(x) = 1 - F(x)$ , d'un paramètre  $r \geq 1$  appelé indice d'aversion au risque ainsi que du niveau de rétention  $R > 0$ . Cela peut être formulé mathématiquement par une transformation :

$$\begin{aligned} T_r : [0, 1] &\longmapsto [0, +\infty[ \\ S(x) &\longrightarrow S(x)^{\frac{1}{r}} \end{aligned}$$

et la prime de réassurance associée à un niveau de rétention  $R$  est donnée par :

$$\Pi_r(R) = \int_R^\infty S(x)^{\frac{1}{r}} dx. \quad (\text{II.11})$$

L'estimation de la prime de réassurance donnée dans (II.11) repose essentiellement sur l'estimation de la fonction de survie, obtenue en fonction de l'expression des quantiles extrêmes par  $S(x) = Q^\bullet(1 - x)$ , où  $Q^\bullet(1 - x)$  peut être l'un des trois estimateurs proposés précédemment. Par exemple, l'expression de l'estimateur de la prime de réassurance dans (II.11), relative à un seuil de rétention  $R = X_{n-k,n}$  peut être obtenue en injectant un estimateur de la fonction de survie relatif au quantile extrême de Weissman par :

$$\hat{\Pi}_r(k) := \left(\frac{k}{n}\right)^{(1/r)} \frac{rH(k)}{1 - rH(k)} X_{n-k,n} \quad (\text{II.12})$$

La normalité asymptotique de l'estimateur dans (II.12) a été établie par Necir *et al.* (2009) pour un indice d'aversion au risque  $r \geq \frac{1}{\gamma}$  et un niveau de rétention optimal  $R_{opt} = F^{-1}(1 - \frac{k}{n})$ , où  $k = k(n) \rightarrow \infty$ , est une séquence d'entiers vérifiant  $\frac{k}{n} \rightarrow 0$  et  $\sqrt{k}A(k/n) \rightarrow 0$  quand  $n \rightarrow \infty$ .

## II.3 Estimation semi-paramétrique dans le cas d'un échantillon i.i.d en présence de censure aléatoire

### II.3.1 Présentation du cadre d'estimation

La notion de censure d'information en sciences appliquées (Analyse de survie, fiabilité ou encore en assurances) a connue un intérêt majeur porté par la communauté statistique, afin de palier à un éventuel manque d'information et la mise en œuvre du modèle approprié.

En analyse de survie par exemple, une caractéristique des données, est l'existence d'observations incomplètes, en effet, les données sont souvent recueillies partiellement, à cause des processus de fin d'étude ou d'observations. Les données censurées, proviennent du fait qu'on n'a plus accès à toute l'information : par exemple au lieu d'observer des réalisations indépendantes et identiquement distribuées (i.i.d.) de

durées  $T$ , on observe la réalisation d'une variable  $Z$  soumise à diverses perturbations, indépendantes ou non du phénomène étudié.

La censure est le phénomène le plus fréquent dans les modèles des données de survie. En pratique, nous distinguons deux cas de censure :

**Définition II.3.1** (Censure à droite). *Une observation statistique est dite censurée à droite si l'individu sur lequel porte l'observation d'une variable aléatoire  $X$ , n'a pas subi l'évènement jusqu'à sa réalisation entière.*

**Exemple II.3.1.** *Un médecin observant la durée de guérison d'un panel composé de 100 patients, constate que 20 entre eux ne se sont pas guéris après la durée de l'étude. Ainsi, la durée de guérison des 20 patients est supérieure à celle observée.*

**Définition II.3.2** (Censure à gauche). *Un observation est dite censurée à gauche quand l'individu a déjà subi l'évènement avant que cet individu soit observé. On sait seulement que la date de l'évènement est inférieure à une certaine date connue.*

**Exemple II.3.2.** *Afin de mesurer la fluidité d'une ligne aérienne pendant la période des fêtes, une compagnie s'intéresse à l'observation des durées d'entre l'enregistrement et l'embarquement d'un client. Certaines observations sont impactées par des mouvements de grève perturbant ainsi l'observation du phénomène étudié. par conséquent, la durée réelle entre l'enregistrement et l'embarquement d'un client est inférieure à la durée observée en présence de perturbations.*

Dans le cas d'une censure à droite, on peut distinguer l'un des trois cas de figures suivants :

**Censure de type I** (censure fixée) : Soit  $C$  une constante, et au lieu d'observer les variables  $(X_1, X_2, \dots, X_n)$  qui nous intéressent, on observe pour chaque individu  $i$  la variable :

$$Z_i = \begin{cases} X_i & \text{si } X_i < C \\ C & \text{sinon} \end{cases}$$

L'échantillon des observations  $Z_1, Z_2, \dots, Z_n$  est dit censuré de type I.

**Censure de type II** (attente) : Dans ce type on observe l'avènement d'un phénomène aléatoire sur  $n$  individus, on arrête l'étude au moment où  $k$  individu réalisent l'évènement. Ainsi on observe  $Z_1, Z_2, \dots, Z_k$  pour les  $k$  premiers individus, et  $Z_k$  pour le reste des  $n - k$  individus restants.

**Censure de type III** (censure aléatoire) : Soit  $(C_1, C_2, \dots, C_n)$  des variables aléatoires i.i.d. Dans ce type de censure, chaque individu  $i$  peut être censuré d'une manière différente de celle opérée sur les autres individus. c'est-à-dire, au lieu d'observer des couples de variables  $(X_i, C_i)$  on observe un échantillon  $(Z_i, \delta_i)$  donné par :

$$Z_i = \begin{cases} X_i & \text{si } X_i \leq C_i \\ C_i & \text{sinon} \end{cases}$$

et

$$\delta_i = \begin{cases} 1 & \text{si } X_i \leq C_i \\ 0 & \text{sinon} \end{cases}$$

Le troisième type de censure est le plus fréquent dans la pratique. on peut citer quelques exemples en analyse de durée, relatifs à une censure aléatoire à droite

1. **la perte de vue** : le patient quitte l'étude en cours et on ne le revoit plus.
2. **l'arrêt ou le changement du traitement** : les effets secondaires ou l'inefficacité du traitement peuvent entraîner un changement ou un arrêt du traitement. Ces patients sont exclus de l'étude.

3. **la fin de l'étude** : l'étude se termine alors que certains patients sont toujours vivants (ils n'ont pas subi l'événement - censure la plus courante).

**Remarque II.3.1.** Une hypothèse naturelle est de considérer que la variable d'intérêt est indépendante de la variable de censure. Du point de vue mathématique, cette hypothèse est indispensable pour la construction des modèles qu'on va présenter par la suite.

Considérons une variable d'intérêt  $X$  définie par sa densité de probabilité  $f$  et sa fonction de survie  $S$ , censurée par une variable  $C$  ayant pour densité de probabilité  $g$  et une fonction de survie  $G$ . Pour un modèle de censure aléatoire à droite, on suppose qu'on observe  $n$  réplifications *i.i.d* du couple  $(Z = \min(X, C), \delta = \mathbf{1}_{\{X \leq C\}})$ .

**Proposition II.3.1.** La fonction de vraisemblance relative à l'échantillon de couple observé  $(\tilde{z}, \tilde{\delta}) = \{(z_i, \delta_i), \text{ pour } i = 1 \text{ à } n\}$ , est donnée par

$$\mathcal{L}(\tilde{z}, \tilde{\delta}) = \prod_{i=1}^n [f(z_i)G(z_i)]^{\delta_i} [g(z_i)S(z_i)]^{1-\delta_i}.$$

**Définition II.3.3** (Estimateur de Kaplan-Meier (1958)). Soit  $(Z_i, \delta_i)$ , pour  $i = 1$  à  $n$  un échantillon de données censurées aléatoirement à droite. L'estimateur de Kaplan-Meier de la fonction de survie  $S(t)$  est donné par

$$\hat{S}(t) = \prod_{i: Z_{i,n} \leq t} \left( \frac{n-i}{n-i+1} \right)^{\delta_i},$$

et en tout point de continuité de  $S$ ,  $t_0 \in [0, \tau]$  et  $S(\tau^-) > 0$ , on a

$$\sqrt{n} \left( \hat{S}(t_0) - S(t_0) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left( 0, V^2(t_0) \right),$$

où  $V^2(t_0) = -S^2(t_0) \int_0^{t_0} \frac{S(du)}{S^2(u)\bar{G}(u)}$  et  $\bar{G}$  est la fonction de survie de la variable de censure  $Y$ .

**Définition II.3.4** (Estimateur de Nilson-Aalen (1972)). Soit  $(Z_i, \delta_i)$ , pour  $i = 1$  à  $n$  un échantillon de données censurées aléatoirement à droite. L'estimateur de la fonction de risque cumulé  $\Lambda(t) = \int_0^t \frac{f(u)}{S(u)} du$  est donné par

$$\hat{\Lambda}(t) = \sum_{i: Z_i \leq t} \frac{d_i}{N_i},$$

où  $Y_i$  représente le nombre d'individus à risque juste avant  $Z_i$  et  $d_i$  représente le nombre de décès en  $Z_i$ .

### II.3.2 Estimation de l'indice des valeurs extrêmes en présence de censure aléatoire à droite

La problématique d'inférence statistique en présence de données censurées dans la théorie des valeurs extrêmes, est considérée comme un récente question de recherche en regard du peu de travaux portant sur la dite problématique. La difficulté est essentiellement liée au fait qu'une partie des données supposées décrire une queue de la distribution<sup>1</sup> (dont le nombre est petit), sont incomplètes, et risque ainsi de biaiser l'information visant à exprimer le comportement des extrêmes.

1. Il est intéressant de regarder l'effet de la censure dans la queue à droite (resp. à gauche) quand on analyse le comportement du maximum (resp. du minimum)

Les premiers à avoir introduit l'idée d'inférence statistique dans la théorie des valeurs extrêmes en présence de censure aléatoire, ont été Reiss et Thomas en 1997 (voir [66], section 6.1), quand ils ont proposé un estimateur de l'IVE  $\gamma > 0$  mais sans donner de résultats asymptotiques de l'estimateur proposé. Récemment, Beirlant *et al.* (2007) proposa un estimateur généralisé de  $\gamma \in \mathbb{R}$ , ainsi que l'estimation du quantile extrême en présence de censure aléatoire à droite. Cette construction est considérée comme le début d'une série de travaux d'analyse asymptotique des différents estimateurs utilisant l'échantillon des observations dépassant un seuil déterministe sous l'hypothèse d'un modèle de second ordre (donné dans II.2). Depuis, plusieurs autres contributions autour de l'inférence en théorie des valeurs extrêmes en présence de données censurées on vu le jour, à l'instar d'Einmahl *et al.* (2008), Gomes and Neves (2001), ou tous récemment Worms and Worms (2014) qui proposèrent un nouveau type d'estimateur non-paramétrique de  $\gamma$  pour une distribution à queue lourde.

**Description du contexte** Considérons un échantillon de *v.a*  $X_1, X_2, \dots, X_n$  supposées i.i.d et issues d'une loi commune  $F \in MDA(H_{\gamma_1})$ . Le modèle de censure aléatoire à droite consiste à considérer l'existence d'un échantillon de censure  $Y_1, Y_2, \dots, Y_n$ , où les  $Y_i$  sont supposées être indépendantes identiquement distribuées selon la même loi  $G \in MDA(H_{\gamma_2})$ , et qu'on observe

$$Z_i = \min(X_i, Y_i) \text{ et } \delta_i = \mathbf{1}_{\{X_i \leq Y_i\}} \text{ pour tout } i = 1 \text{ à } n.$$

La variable indicatrice  $\delta_i$  désigne si l'observation  $Z_i$  est censurée ou pas. Notons par  $\gamma$ , l'indice des valeurs extrêmes associé à la variable  $Z$  (on suppose ainsi que les  $Z_i$  sont issues d'une loi commune  $H \in MDA(H_\gamma)$ ). L'objectif principal consiste à apporter quelques éléments d'inférence statistique sur la queue droite de la distribution inconnue de  $X$  dans les trois cas suivants

$$\left\{ \begin{array}{l} \text{Cas 01 : } \gamma_1 > 0 \text{ et } \gamma_2 > 0 \\ \text{Cas 02 : } \gamma_1 < 0, \gamma_2 < 0 \text{ et } x_F = x_G \text{ (} F \text{ et } G \text{ admettent le même point terminal)} \\ \text{Cas 03 : } \gamma_1 = \gamma_2 = 0 \text{ et } x_F = x_G = +\infty \end{array} \right.$$

Dans chacun des trois cas précédant, Einmahl *et al.* (2008) ont montré que  $\gamma = \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2}$ , avec la notation  $\gamma = 0$  pour le cas 03. De plus, définissant la proportion de non-censure par

$$\begin{aligned} p \equiv p(z) &:= \mathbb{P}(X \leq Y | Z = z) \\ &= \mathbb{P}(\delta = 1 | Z = z) \\ &= \frac{f(z)(1 - G(z))}{h(z)} \\ p(z) &\rightarrow \frac{\gamma_2}{\gamma_1 + \gamma_2}, \text{ quand } z \rightarrow \infty \end{aligned}$$

Ce qui motive l'estimation de  $\gamma_1$  par le rapport de  $\gamma$  par  $p(z)$ .

**Exemple II.3.3.** Soit une variable aléatoire  $X$  distribuée selon une loi de Pareto généralisée avec  $F_X(x) = 1 - (1 + \gamma_1 x)^{-1/\gamma_1}$ ,  $\gamma_1 > 0$ , censurée par une variable aléatoire  $Y$  distribuée selon la même loi, avec  $G_Y(x) = 1 - (1 + \gamma_2 x)^{-1/\gamma_2}$ ,  $\gamma_2 > 0$ , on a :

$$\begin{aligned} 1 - H_Z(z) &= \mathbb{P}(\min(X, Y) \geq z) \\ &= (1 + \gamma_1 z)^{-1/\gamma_1} (1 + \gamma_2 z)^{-1/\gamma_2} \end{aligned}$$

et  $f_Z(z) = (1 + \gamma_1 x)^{-1/\gamma_1 - 1} (1 + \gamma_2 x)^{-1/\gamma_2 - 1} \times \left(1 + \frac{1 + \gamma_1 z}{1 + \gamma_2 z}\right)$ . d'où

$$\begin{aligned} p \equiv p(z) &:= \mathbb{P}(X \leq Y | Z = z) \\ &= \frac{1}{1 + \frac{1 + \gamma_1 z}{1 + \gamma_2 z}} \\ p(z) &\rightarrow \frac{\gamma_2}{\gamma_1 + \gamma_2}, \text{ quand } z \rightarrow \infty. \end{aligned}$$

Einmahl *et al.* (2008) proposèrent un estimateur consistant de  $p$ , basé sur les  $k$  plus grandes statistiques d'ordre, et donné par

$$\hat{p} = \hat{p}_k := \frac{1}{k} \sum_{i=1}^k \delta_{[n-i+1]},$$

où  $\delta_{[j]}$  est la valeur de  $\delta$  concomitante à la  $j^{\text{ème}}$  statistique d'ordre  $Z_{j,n}$ .

Il est important de noter, que toute estimation de l'indice des valeurs extrêmes  $\gamma_1$ , basée sur l'échantillon  $Z_1, Z_2, \dots, Z_n$  (sans tenir compte de la censure), converge vers  $\gamma$  l'indice des valeurs extrêmes de  $Z$  et non pas celui de la variable d'intérêt  $X$ . Par conséquent, la procédure d'estimation doit être adapter à la présence de censure. Einmahl *et al.* (2008) proposa une estimation de  $\gamma_1$  au sens des estimateurs probabilité inverse pondérée (Inverse Probability Weighted), qui consiste à estimer l'indice des valeurs extrêmes relatif à la variable  $X$  par le rapport entre l'estimateur de l'indice des valeurs extrêmes de  $Z$  (sur la base de  $k$  plus grandes valeurs de  $Z$ ) et l'estimateur de la proportion de non-censure dans la queue droite de distribution, i.e :

$$\hat{\gamma}_c^\bullet(k) = \frac{\hat{\gamma}^\bullet(k)}{\hat{p}_k}, \quad (\text{II.13})$$

où  $\hat{\gamma}^\bullet(k)$  peut être l'un des estimateurs présentés dans la sous-section II.2.1 (en particulier  $\hat{\gamma}^H(k)$ ,  $\hat{\gamma}^P(k)$  ou  $\hat{\gamma}^M(k)$ ). Le théorème suivant énoncé dans [29] décrit la normalité asymptotique de  $\hat{p}$  vers  $\frac{\gamma_2}{\gamma_1 + \gamma_2}$  mais aussi celle de  $\hat{\gamma}_c^\bullet(k)$  vers  $\gamma_1$  relativement à la convergence de  $\hat{\gamma}^\bullet(k)$  vers  $\gamma$ .

**Théorème II.3.1.** *Pour  $n \rightarrow \infty$ , et sous les hypothèses :*

(H1) :  $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow \lambda \in \mathbb{R}$

(H2) :  $\frac{1}{\sqrt{k}} \sum_{i=1}^k [p(H^{\leftarrow}(1 - \frac{i}{n})) - p] \rightarrow \alpha \in \mathbb{R}$

(H3) : pour  $1 - \frac{k}{n} < t < 1$  et  $|t - s| \leq c\frac{\sqrt{k}}{n}$ ,  $s < 1, c > 0$ , on a :

$$\sqrt{k} \sup_{t,s} |p(H^{\leftarrow}(t)) - p(H^{\leftarrow}(s))| \rightarrow 0$$

on a :

$$\sqrt{k}(\hat{\gamma}_c^\bullet(k) - \gamma_1) \xrightarrow{d} \mathcal{N}\left(\frac{1}{p}(\lambda b_0 - \gamma_1 \alpha), \frac{\sigma^2 + \gamma_1^2 p(1-p)}{p^2}\right)$$

où  $\lambda b_0$  (respectivement  $\sigma^2$ ) représente la moyenne (respectivement la variance) de  $\sqrt{k}(\hat{\gamma}^\bullet(k) - \gamma)$ .

*Démonstration.* Notons que  $\sqrt{k}(\hat{\gamma}_c^\bullet(k) - \gamma_1) = \sqrt{k}\left(\frac{\hat{\gamma}^\bullet(k)}{\hat{p}} - \gamma_1\right)$ . Ainsi, une décomposition naturelle serait de réécrire le second membre en faisant intervenir un terme en  $\gamma$ , ie :

$$\begin{aligned} \sqrt{k}(\hat{\gamma}_c^\bullet(k) - \gamma_1) &= \frac{1}{\hat{p}} \sqrt{k}(\hat{\gamma}^\bullet(k) - \gamma) + \frac{1}{\hat{p}} \sqrt{k}(\gamma - \gamma_1 \hat{p}) \\ &= \frac{1}{\hat{p}} \sqrt{k}(\hat{\gamma}^\bullet(k) - \gamma) + \frac{\gamma_1}{\hat{p}} \sqrt{k} \left( \frac{\gamma_2}{\gamma_1 + \gamma_2} - \hat{p} \right) \end{aligned}$$

d'où

$$\sqrt{k}(\hat{\gamma}_c^\bullet(k) - \gamma_1) = \frac{1}{\hat{p}}\sqrt{k}(\hat{\gamma}^\bullet(k) - \gamma) - \frac{\gamma_1}{\hat{p}}\sqrt{k}(\hat{p} - p) \quad (\text{II.14})$$

le comportement asymptotique du terme  $\sqrt{k}(\hat{\gamma}^\bullet(k) - \gamma)$  est obtenu par le théorème II.2.1 quand  $\hat{\gamma}^\bullet(k)$  est choisi parmi les trois estimateurs classiques de  $\gamma$ . Nous focalisons ainsi sur le second terme dans la décomposition II.14 en considérons la construction suivante :

Soit  $Z$  et  $U$  deux *v.a* indépendantes issues respectivement d'une loi  $H$ , et d'une loi  $\mathcal{U}_{[0,1]}$ . définissant

$$\delta = \begin{cases} 1 & \text{si } U \leq p(Z) \\ 0 & \text{si } U > p(Z) \end{cases} \quad \text{et } \tilde{\delta} = \begin{cases} 1 & \text{si } U \leq p \\ 0 & \text{si } U > p \end{cases}$$

Il est aussi clair que  $Z$  et  $\tilde{\delta}$  sont indépendantes et

$$\mathbb{P}(|\delta - \tilde{\delta}| = 1 | Z = z) = |p - p(z)|$$

On dupliquant d'une manière indépendante cette construction  $n$  fois, chaque couple  $(Z_i, \delta_i)$  ainsi construit aura la même distribution que le couple initial obtenu par l'échantillon. On peut écrire donc

$$\hat{p} = \frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{U_{[n-i+1, n]} \leq p(Z_{n-i+1, n})\}},$$

où  $U_{[n-i+1, n]}$  est la valeur de  $U_j$  concomitante à  $Z_{n-i+1, n}$ , et de manière similaire

$$\tilde{p} = \frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{U_{[n-i+1, n]} \leq p\}}.$$

Par construction, il est clair que  $U_{1, n}, U_{2, n}, \dots, U_{n, n}$  sont indépendantes de  $Z$ . Ainsi, nous décomposons le terme  $\hat{p} - p$  sous la forme

$$\sqrt{k}(\hat{p} - p) = \sqrt{k}(\hat{p} - \tilde{p}) + \sqrt{k}(\tilde{p} - p).$$

comme  $\tilde{p} \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{U_i \leq p\}}$ , alors

$$\sqrt{k}(\tilde{p} - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)). \quad (\text{II.15})$$

Intéressons nous maintenant au terme  $\sqrt{k}(\hat{p} - \tilde{p})$ , qui semble être le terme biais

$$\begin{aligned} \sqrt{k}(\hat{p} - \tilde{p}) &\stackrel{d}{=} \frac{1}{\sqrt{k}} \sum_{i=1}^k [\mathbb{1}_{\{U_i \leq p(Z_{n-i+1, n})\}} - \mathbb{1}_{\{U_i \leq p\}}] \\ &\stackrel{d}{=} \frac{1}{\sqrt{k}} \sum_{i=1}^k [\mathbb{1}_{\{U_i \leq p(Z_{n-i+1, n})\}} - \mathbb{1}_{\{U_i \leq p(H^{\leftarrow(1-\frac{i}{n})})\}}] \\ &\quad + \frac{1}{\sqrt{k}} \sum_{i=1}^k [\mathbb{1}_{\{U_i \leq p(H^{\leftarrow(1-\frac{i}{n})})\}} - \mathbb{1}_{\{U_i \leq p\}}] \\ &=: T_{1, k} + T_{2, k} \end{aligned}$$

D'après le corollaire 1 de [Chow et Teicher](#) (1997, page 357) et sous les hypothèses **(H2)** et **(H3)**,  $T_{2, k}$  converge en probabilité vers  $\alpha$ . L'objectif maintenant est de montrer que  $T_{1, k} \xrightarrow{\mathbb{P}} 0$ .

Notons par  $V_i = H(Z_i)$ ,  $i = 1$  à  $n$  (les  $V_i$ 's sont des variables aléatoires i.i.d distribuées selo une loi uniforme sur  $[0,1]$ ) et par  $r(t) = p(H^{\leftarrow}(t))$ , on a :

$$T_{1,k} = \frac{1}{\sqrt{k}} \sum_{i=1}^k \left[ \mathbb{1}_{\{U_i \leq r(V_{n-i+1,n})\}} - \mathbb{1}_{\{U_i \leq r(1-\frac{i}{n})\}} \right]$$

et

$$\begin{aligned} |T_{1,k}| &\leq \frac{1}{\sqrt{k}} \sum_{i=1}^k \left| \mathbb{1}_{\{U_i \leq r(V_{n-i+1,n})\}} - \mathbb{1}_{\{U_i \leq r(1-\frac{i}{n})\}} \right| \\ &\stackrel{d}{=} \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{1}_{\{U_i \leq |r(V_{n-i+1,n}) - r(1-\frac{i}{n})|\}} \\ &\leq \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{1}_{\{U_i \leq \frac{\eta}{\sqrt{k}}\}}, \text{ pour } \eta > 0. \end{aligned}$$

Par le théorème 1 dans [Chow et Teicher](#) (page 356), et pour  $\eta > 0$  arbitrairement petit,  $T_{1,k} \xrightarrow{\mathbb{P}} 0$ .

En combinons [II.14](#) et [II.15](#), on obtient

$$\sqrt{k}(\hat{\gamma}_c^\bullet(k) - \gamma_1) = \frac{1}{\hat{p}} \sqrt{k}(\hat{\gamma}^\bullet(k) - \gamma) - \frac{\gamma_1}{\hat{p}} \sqrt{k}(\hat{p} - p) - \frac{\gamma_1 \alpha}{\hat{p}} + o_{\mathbb{P}}(1).$$

les deux termes en  $\sqrt{k}$  sont indépendant, car le premier (respectivement le second) est construit à partir de l'échantillon des  $Z_i$ 's (respectivement des  $U_i$ 's). Finalement, et sous les hypothèses **(H1)**-**(H3)**, le résultat dans le théorème [II.3.1](#) est vérifié.  $\square$

**Remarque II.3.2.** Nous notons qu'un autre type d'estimateur de l'indice des valeurs extrêmes, fut introduit par [Worms and Worms \(2014\)](#), donné par :

$$\hat{\gamma}_c^W(k) = \frac{1}{n(1 - \hat{F}_n(Z_{n-k,n}))} \sum_{i=1}^k \frac{1}{1 - \hat{G}_n(Z_{n-i+1}^-)} i \log \left( \frac{Z_{n-i+1,n}}{Z_{n-i,n}} \right), \quad (\text{II.16})$$

où  $\hat{F}_n$  et  $\hat{G}_n$  représentent respectivement les estimateurs de Kaplan-Meier de  $F$  et  $G$ . Nous illustrons par un exemple comparatif, les qualités et les défauts empirique des estimateurs proposés par : Einmahl dans [II.13](#) respectivement à une adaptation de l'estimateur de Hill (qu'on note  $\hat{\gamma}_c^H$ ), de Pickand (qu'on note  $\hat{\gamma}_c^P$ ), de Dekker (qu'on note  $\hat{\gamma}_c^M$ ) ainsi que celui proposé dans [\(II.16\)](#) par Worms (qu'on note  $\hat{\gamma}_c^W$ ).

La figure [II.2](#) illustre le comportement empirique des estimateurs  $\hat{\gamma}_c^H$ ,  $\hat{\gamma}_c^P$ ,  $\hat{\gamma}_c^M$  et  $\hat{\gamma}_c^W$ , à travers trois jeux de simulation (1000 réplifications d'échantillons de taille  $n = 500$ ) : chaque échantillon est issu d'une censure d'une loi de Pareto standard  $\gamma_1 = 2$  par une loi de Pareto de paramètre respectif  $\gamma_2 = 18, 6$  et  $2$ , ce qui nous permet d'espérer une proportion de non-censure  $p$  respective de 90%, 75%, 50%, soit une proportion de censure de 10%, 25%, 50% (cela découle du faite que  $p = \frac{\gamma_2}{\gamma_1 + \gamma_2}$ , dans le cas d'échantillon issu d'une distribution de Pareto censurée par un échantillon d'une loi de Pareto). Nous constatons que les deux estimateurs  $\hat{\gamma}_c^H$  et  $\hat{\gamma}_c^W$  semble être de meilleure qualité en terme de biais et RMSE que  $\hat{\gamma}_c^P$  et  $\hat{\gamma}_c^M$  dans le cas d'une censure à 10% (ligne 1), même si la qualité de ces deux estimateurs ce dégrade lorsque la proportion de censure augmente (notamment le biais significatif de l'estimateur de Worms présenté dans le graphe de la ligne 3, en fonction du nombre d'excès  $k$ ). L'adéquation de l'estimateur de Hill (par rapport aux deux estimateurs  $\hat{\gamma}_c^P$  et  $\hat{\gamma}_c^M$ ) adapté au cas de données censurées s'explique forcément par le constat soulevé dans la remarque [II.2.2](#), relevant la régularité de l'estimateur de Hill dans le cas d'échantillon complet et une loi d'échantillonnage appartenant au max-domaine d'attraction de Fréchet).

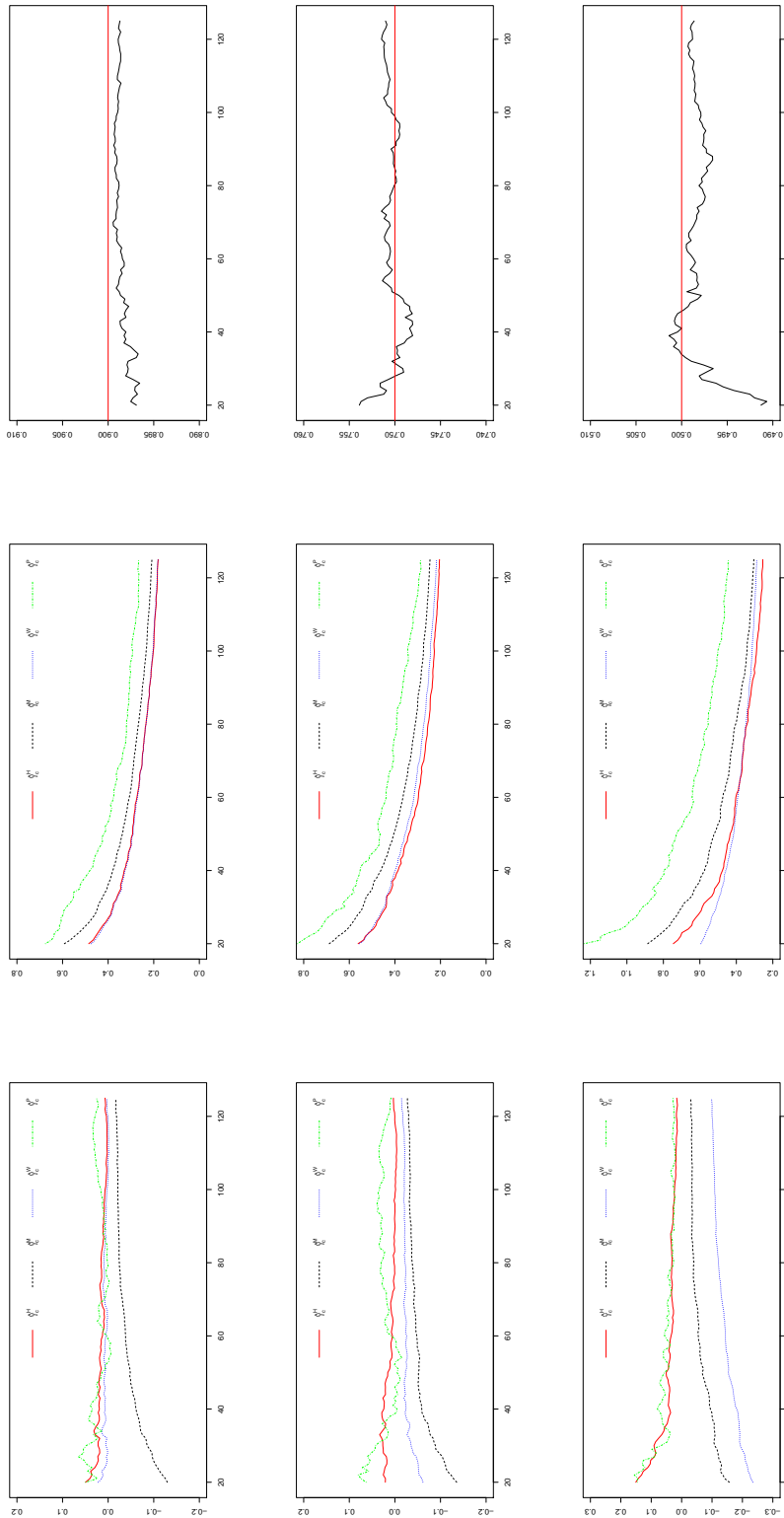


FIGURE II.2 – Comportement empirique des estimateurs en présence de censure - ligne 1 : une proportion de non-censure de 90%, ligne 2 : une proportion de non-censure de 75% et ligne 3 : une proportion de non-censure de 50%.

# CHAPITRE III

## NOTIONS DE STATISTIQUE BAYÉSIENNE



“

... Nous ne connaissons a priori des choses que ce que nous y mettons nous-mêmes ...

”

---

Emmanuel Kant, *Critique de la raison pure*, 1724 - 1804

La théorie bayésienne est considérée comme une robuste machine d'inférence statistique, portée par un modèle combinant à la fois une expertise (ou connaissance avant la réalisation de l'expérience aléatoire) a priori, et l'observation du phénomène aléatoire afin d'en déduire un jugement a posteriori. Ainsi, un modèle bayésien suppose la connaissance d'un jugement a priori sur le ou les paramètres du modèle statistique, en les supposant aléatoires et distribués selon une dite loi a priori, du ou des paramètres.

Nous décrivons dans ce chapitre, quelques notions essentielles à l'analyse bayésienne, de sa construction à la mise en œuvre pratique d'un tel modèle. Nous commençons par rappeler tout d'abord, les fondements théoriques des modèles bayésiens, permettant d'introduire dans un second temps, l'ensemble des techniques bayésiennes en étalant le mécanisme d'inférence sur un paramètre du modèle statistique classique où on prétend avoir une connaissance a priori. La troisième partie de ce chapitre sera consacrée au choix discutable de la loi a priori du paramètre qu'on supposera, et l'idée derrière la construction d'une telle loi en fonction de la structure du modèle statistique. Nous terminons ce chapitre par un apport bayésien à l'estimation de l'indice de queue pour une distribution dans le max-domaine d'attraction de Fréchet ( $\alpha > 0$ ).

Pour une description détaillée des concepts et principes de la théorie bayésienne, nous recommandons aux lecteurs l'ouvrage de [Christian Robert](#) (2006).

### III.1 Approche bayésienne

Dans une approche d'inférence statistique fréquentielle, on considère que le paramètre du modèle  $\theta$  est fixe, appartenant à un espace des paramètres  $\Theta$ , qu'on l'estime sur la base d'un échantillon observé. L'idée de base derrière un concept bayésien, consiste à traiter le paramètre inconnu  $\theta$  comme une variable aléatoire admettant une densité de probabilité  $\pi(\theta)$  qu'on appelle loi a priori.

L'inférence bayésienne est fondée ainsi sur une combinaison judicieuse entre une connaissance a priori sur le paramètre du modèle  $\theta$  et l'information apportée par les observations, dont l'objet est de construire un nouveau modèle bayésien fonction d'une loi sur  $\theta$  dite a posteriori.

On se place dans un espace  $(\Omega, \mathcal{B}, \mathbb{P}_\theta, \theta \in \Theta)$ , où  $\Omega$  représente l'espace d'états et  $\Theta$  celui des paramètres. L'objectif d'une analyse statistique est d'associer au mieux un modèle paramétrique  $\mathbb{P}_\theta$  à une situation observée. Le concept bayésien suppose de plus, l'existence d'une loi  $\pi$ , telle que l'espace des paramètres muni de la loi  $\pi$  forme un espace probabilisé  $(\Theta, \mathcal{A}, \pi)$ .

**Définition III.1.1** (Modèle dominé). *Un modèle statistique est dit dominé s'il existe une mesure commune dominante  $\mu$ , c'est-à-dire pour tout  $\theta$ ,  $\mathbb{P}_\theta$  admet une densité par rapport à  $\mu$  :*

$$f(X|\theta) = \frac{d\mathbb{P}_\theta}{d\mu}$$

$f(X|\theta)$  appelée vraisemblance du modèle ou loi conditionnelle de  $X$  sachant  $\theta$ , est vue comme une fonction de  $\theta$  une fois qu'on a observé un tirage de  $X$ .

**Définition III.1.2** (Modèle bayésien [68]). *Un modèle statistique bayésien est la double donnée d'un modèle paramétrique  $\{f_\theta(x), \theta \in \Theta\}$  et d'une loi de probabilité, de densité  $\pi$  dite loi a priori qui est la marginale de la variable aléatoire  $\theta$ .*

*i.e :*

$$x \sim f(x | \theta), \quad \theta \sim \pi(\theta)$$

#### III.1.1 Le paradigme bayésien

Étant donné un modèle paramétrique relatif à la loi conditionnelle d'une v.a  $X$  sachant  $\theta \in \Theta$  (notée  $X \sim f(x | \theta)$  et l'espace  $\Theta$  est à dimension finie). L'analyse statistique bayésienne vise à exploiter le plus efficacement possible l'information apportée par  $x$  sur le paramètre  $\theta$  pour ensuite construire des procédures d'inférence sur  $\theta$ . Bien que  $x$  ne soit qu'une réalisation observée d'une loi gouvernée par  $\theta$ . L'information fournie par l'observation  $x$  est contenue dans la densité  $f(x|\theta)$  que l'on représente classiquement sous la forme inversée de vraisemblance :

$$\ell(\theta|x) = f(x|\theta)$$

Pour traduire que la fonction  $\ell(\theta|x)$  dépend d'un paramètre  $\theta$  (inconnu), pour une valeur observée  $x$ .

Le paradigme bayésien stipule que, le paramètre du modèle  $\theta$  n'est plus considéré comme une constante inconnue et déterministe mais comme une variable aléatoire. L'incertitude sur le paramètre  $\theta$  dans un modèle bayésien peut être décrite par une distribution de probabilité  $\pi$  fonction de  $\theta$ , ce qui revient à supposer que  $\theta$  est distribué suivant  $\pi(\theta)$  avant que  $X$  ne soit généré selon une loi d'échantillonnage  $f(x|\theta)$ . La loi a posteriori est déterminée à l'aide de la formule de *Bayes*

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

où  $\int_{\Theta} f(x|\theta)\pi(\theta)d\theta$  représente la loi marginale de  $X$ . Ainsi on peut écrire

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \quad (\text{III.1})$$

Le problème qui s'impose lorsqu'on construit un tel modèle, réside dans le choix de la loi a priori qui est perçue comme une difficulté majeure des approches bayésiennes en tentant d'interpréter le rôle de l'information a priori disponible car elle est rarement assez précise pour conduire à la détermination d'une seule et unique loi. Il existe néanmoins deux façons permettant de construire la loi a priori, l'une en fonction des observations, dites lois conjuguées ou informatives, et une deuxième à faible contenu informatif, dites loi-non informatives. Dans la suite de ce chapitre, une section sera consacrée à la description des deux alternatives relatives au choix d'une loi a priori.

### III.1.2 Fondement de la construction bayésienne

Une observation  $x$  peut être considérée comme étant générée par la loi conditionnelle de  $X$  sachant  $\theta$ . La fonction de vraisemblance est donnée par :

$$\mathcal{L}(\theta, x) = \begin{cases} P(X = x | \theta) & \text{si le modèle est discret} \\ f(x | \theta) & \text{si le modèle est continu} \end{cases}$$

Si  $\pi$  est la loi a priori de  $\theta$  alors, la loi du couple  $(X, \theta)$  est donné par :

$$\pi(x, \theta) \propto \begin{cases} P(X = x | \theta)\pi(\theta) & \text{si le modèle est discret} \\ f(x | \theta)\pi(\theta) & \text{si le modèle est continu} \end{cases}$$

et la loi marginale de  $X$  est obtenue de la façon suivante :

- pour un modèle discret :  $P(X = x) = \sum_{\theta \in \Theta} P(X = x | \theta)\pi(\theta)$
- pour un modèle continu :  $f_X(x) = \int_{\Theta} f(x | \theta)\pi(\theta)d\theta$

Les deux cas précédents peuvent ainsi être résumés sous la forme générale :

$$\mathcal{L}_X(x) = \int \mathcal{L}(\theta, x)\pi(\theta)d\theta$$

Si on s'intéresse à la loi conditionnelle de  $\theta$  sachant  $X = x$  (appelée loi a posteriori), cela revient à déterminer la relation entre  $\theta$  et  $X$  par la formule de Bayes :

$$P(B | A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A | B)P(B)}{P(A)}.$$

La loi a posteriori est déterminée donc par :

- Pour un modèle discret :  $\pi(\theta|x) = \frac{P(X=x|\theta)\pi(\theta)}{P(X=x)} = \frac{P(X=x|\theta)\pi(\theta)}{\sum_{\theta \in \Theta} P(X=x|\theta)\pi(\theta)}$
- Pour un modèle continu :  $\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f_X(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}$

On écrit d'une manière générale, la densité a posteriori ou vraisemblance a posteriori du paramètre  $\theta$  par :

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

Comme, la marginale  $\mathcal{L}_X(x)$  ne dépend pas du  $\theta$ , on peut aussi écrire :

$$\pi(\theta | x) \propto \mathcal{L}(\theta, x)\pi(\theta) \quad (\text{III.2})$$

Pour un échantillon d'observations  $x_1, x_2, \dots, x_n$  supposées être indépendantes et issues de  $X$ , la distribution a posteriori  $\pi(\theta | x)$  représente alors l'information dont on dispose sur  $\theta$ , après observation de  $X$ . C'est un compromis entre l'information a priori (donnée par  $\pi(\theta)$ ) et l'information tirée de l'échantillon (donnée par  $\mathcal{L}(\theta, x_1, \dots, x_n)$ ).

## III.2 Principes d'inférence bayésienne

### III.2.1 Fonction de perte et risque bayésien

**Définition III.2.1** (Fonction de perte). Soit  $t = T(x_1, \dots, x_n)$  un estimateur de  $\theta$  supposé appartenir à un ensemble de décision  $\mathcal{D}$ . On appelle "fonction de perte" (notée  $l(t, \theta)$ ), toute fonction mesurable de  $\mathcal{D} \times \Theta$  à valeurs positives satisfaisant :

- $l(t, \theta) \geq 0$  pour toute valeur d'estimateur  $t \neq \theta$  et  $\forall \theta \in \Theta$ .
- $l(t, \theta) = 0$  si  $t = \theta$ .

Cette fonction mesure la perte engendrée par l'estimation de  $\theta$  par  $t$ . L'exemple suivant donne quelques fonction de perte des plus usuelles.

**Exemple III.2.1** (Fonctions de perte usuelles). :

- $l_1(t, \theta) = c_1(t - \theta)1_{\{\theta \leq t\}}(t) + c_2(\theta - t)1_{\{\theta > t\}}(t)$ .
- $l_2(t, \theta) = \varphi(\theta)|t - \theta|^r$  avec  $\varphi(\theta) \geq 0$  et  $r > 0$ .
- $l_3(t, \theta) = \begin{cases} A & \text{si } |t - \theta| > \varepsilon \\ 0 & \text{si } |t - \theta| \leq \varepsilon \end{cases} \quad \forall \varepsilon > 0, \quad A > 0$

En particulier, si :

- $\varphi(\theta) = 1$  et  $r=2$ ,  $l_2(t, \theta) = (t - \theta)^2$ .  $l_2(t, \theta)$  est dite fonction de perte quadratique.
- $c_1 = c_2$ ,  $l_1(t, \theta) = |t - \theta|$  est dite fonction de perte absolue.

**Définition III.2.2** (Risque fréquentiste). Soit  $t$  un estimateur de  $\theta$  et  $l(t, \theta)$  la fonction de perte associée. On appelle risque fréquentiste, la quantité notée  $R_t(\theta)$  et donnée par :

$$R_t(\theta) = \mathbb{E}_X[l(t, \theta)] = \int_{\Omega} l(t, \theta) d\mathbb{P}_{\theta}$$

**Définition III.2.3** (Risque intégré). Si  $\pi$  désigne la loi a priori de  $\theta$ , alors on appelle risque intégré, la quantité notée  $r(t)$  et donnée par :

$$r(t) = \int_{\Theta} R_t(\theta) d\pi(\theta)$$

**Définition III.2.4** (Risque bayésien). Soit  $t$  un estimateur de  $\theta$  et  $l(t, \theta)$  la fonction de perte associée et  $\pi$  la loi a priori de  $\theta$ . On appelle risque bayésien moyen, la quantité notée  $R^\pi(t)$  et donnée par :

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

**Définition III.2.5** (Décision préférable et admissible). Soit  $R^\pi(t)$  la fonction de risque bayésien moyen relative à l'estimation d'un paramètre  $\theta \in \Theta$ .

1. Une règle de décision  $d_1 \in \mathcal{D}$  est dite préférable au sens bayésienne (ou meilleure), qu'une autre règle  $d_2 \in \mathcal{D}$  (et on note  $d_1 \succcurlyeq d_2$ ), si et seulement si  $R^\pi(d_1) \leq R^\pi(d_2)$ .
2. Une règle de décision  $d_1 \in \mathcal{D}$  est dite admissible au sens bayésienne s'il n'existe aucune autre règle qui lui soit strictement préférable. i.e :

$$\forall d \in \mathcal{D}, d_1 \succcurlyeq d.$$

3. Une règle de décision  $d_B$  est dite bayésienne si elle minimise la perte moyenne a posteriori, i.e :

$$\mathbb{E}_{\theta|X=x}(l(\theta, d_B)) = \min_{t \in \mathcal{D}} R^\pi(t)$$

### III.2.2 Estimateurs bayésiens ponctuels

On considère un modèle bayésien donné par  $X \sim f(x | \theta)$ ,  $\theta \sim \pi(\theta)$ , à partir duquel on construit la loi a posteriori  $\pi(\theta | x)$  par l'équation (III.2).

**Définition III.2.6** (Estimateurs bayésiens). *Un estimateur  $\hat{\theta}$  associé à une fonction de perte  $l(t, \theta)$  et à une loi a priori  $\pi(\theta)$ , est dit estimateur bayésien de  $\theta$ , si et seulement s'il minimise la fonction de risque bayésien moyen, i.e*

$$\hat{\theta} = \arg \min_{t \in \mathcal{D}} R^\pi(t)$$

La proposition suivante permet d'exprimer quelques formes d'estimateurs bayésiens.

**Proposition III.2.1.** *Soit  $\theta \in \Theta$  est un paramètre unidimensionnel à estimer, lorsque :*

**Cas 01** *la fonction de perte est uniforme,  $l(t, \theta) = 1 - \mathbf{1}_{\{|t-\theta| < \delta\}}$ ,  $\delta > 0$*

*l'estimateur bayésien relatif à la fonction perte  $l(t, \theta)$  lorsque  $\delta \rightarrow 0$ , correspond au mode de la distribution a posteriori  $\pi(\theta|x)$ . Il est dit estimateur du maximum a posteriori, noté  $\hat{\theta}_{MAP}$  et donné par*

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

**Cas 02** *la fonction de perte est absolue,  $l(t, \theta) = c_1(t - \theta)\mathbf{1}_{\{\theta \leq t\}}(t) + c_2(\theta - t)\mathbf{1}_{\{\theta > t\}}(t)$ ,  $c_1 > 0, c_2 > 0$ . L'estimateur bayésien  $\hat{\theta}_q$ , correspond au quantile d'ordre  $\frac{c_1}{c_1 + c_2}$  de la distribution a posteriori, i.e*

$$P(\theta < \hat{\theta}_q | x) = \frac{c_1}{c_1 + c_2}.$$

*Dans le cas où  $c_1 = c_2$ , l'estimateur bayésien correspond à la médiane de la loi a posteriori.*

**Cas 03** *la fonction de perte est quadratique,  $l(t, \theta) = (t - \theta)^2$ , qui correspond à une fonction de perte quadratique.*

*l'estimateur bayésien coïncide avec la moyenne à posteriori de  $\theta$ . Il est dit estimateur de la moyenne a posteriori (Mean posterior estimator), noté  $\hat{\theta}_{MPE}$  et donné par :*

$$\hat{\theta}_{MPE} = \mathbb{E}[\theta|x] = \int_{\Theta} \theta \pi(\theta|x) d\theta$$

Pour plus de détails sur les résultats présentés dans la proposition III.2.1, nous recommandons de voir [68] page 85.

### III.2.3 Exemple du contrôle de qualité

Soit un échantillon de variables aléatoires  $X_1, X_2, \dots, X_n$  supposées être i.i.d distribuées selon une loi de Bernoulli  $B(p)$ , où chaque  $X_i$  désigne l'état de défaillance de la pièce  $i$ , i.e

$$X_i = \begin{cases} 1 & \text{si la } i^{\text{e}} \text{ pièce est défectueuse} \\ 0 & \text{sinon} \end{cases}$$

Supposons aussi, qu'un expert de fabrication estime qu'il y a peu de chance qu'une pièce soit défectueuse. Cela se traduit par un choix naturel et intuitif d'une loi a priori de  $p$  à support dans  $[0, 1]$ , qui soit concentrée sur les petites valeurs de  $p$ . Ainsi, une distribution Beta,  $\mathcal{B}(a, b)$  avec  $a > 0$  petit devant  $b$

semble être a priori une loi candidate judicieuse. On cherche a présent, à déterminer la loi a posteriori de  $p$  par l'équation (III.2), relativement à un échantillon observé  $x_1, x_2, \dots, x_n$  et une loi a priori :

$$\pi(p) = \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} \mathbb{1}_{[0,1]}(p)$$

où  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ . Rappelons que l'espérance et la variance d'une loi  $\mathcal{B}_{[0,1]}(a, b)$  sont :

$$\mathbb{E}(p) = \frac{a}{a+b} \quad \text{et} \quad \text{Var}(p) = \frac{ab}{(a+b)^2(a+b+1)}$$

Le choix des hyperparamètres  $a$  et  $b$  est conditionné par une connaissance a priori du paramètre  $p$ , en terme de valeur moyenne et en variabilité.

Nous commençons tous d'abord, par l'écriture de la vraisemblance comme suit :

$$\mathcal{L}(p, x_1, \dots, x_n) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}.$$

Ainsi, la loi marginale de  $X$  est donnée par :

$$\begin{aligned} \mathcal{L}_X(x_1, \dots, x_n) &= \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \int_0^1 \mathcal{L}(p, x_1, \dots, x_n) \pi(p) dp \\ &= \int_0^1 p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} \mathbb{1}_{[0,1]}(p) dp \\ &= \frac{B(\sum_{i=1}^n x_i + a, n - \sum_{i=1}^n x_i + b)}{B(a, b)} \times \int_0^1 \frac{p^{\sum_{i=1}^n x_i + a - 1} (1-p)^{n - \sum_{i=1}^n x_i + b - 1}}{B(\sum_{i=1}^n x_i + a, n - \sum_{i=1}^n x_i + b)} dp \end{aligned}$$

Ce qui nous permet d'obtenir finalement

$$\mathcal{L}_X(x_1, \dots, x_n) = \frac{B(\sum_{i=1}^n x_i + a, n - \sum_{i=1}^n x_i + b)}{B(a, b)}$$

La loi a posteriori est ainsi donnée par :

$$\begin{aligned} \pi(p|x_1, x_2, \dots, x_n) &= \frac{\mathcal{L}(p, x_1, \dots, x_n) \pi(p)}{\mathcal{L}_X(x_1, \dots, x_n)} \\ &= \frac{B(a, b)}{B(\sum_{i=1}^n x_i + a, n - \sum_{i=1}^n x_i + b)} \times \frac{1}{B(a, b)} p^{\sum_{i=1}^n x_i + a - 1} (1-p)^{n - \sum_{i=1}^n x_i + b - 1} \\ &= \frac{1}{B(\sum_{i=1}^n x_i + a, n - \sum_{i=1}^n x_i + b)} \times p^{\sum_{i=1}^n x_i + a - 1} (1-p)^{n - \sum_{i=1}^n x_i + b - 1} \end{aligned}$$

On reconnaît la densité d'une loi Beta de paramètres  $\sum_{i=1}^n x_i + a$  et  $n - \sum_{i=1}^n x_i + b$ . Ainsi, l'estimateur bayésien de  $p$  associée a une fonction de perte quadratique, est donné par :

$$\hat{p}_{MPE} = \frac{\sum_{i=1}^n x_i + a}{n + a + b}$$

Pour une fonction de perte uniforme, l'estimateur bayésien de  $p$  est donné par :

$$\hat{p}_{MAP} = \frac{\sum_{i=1}^n x_i + a - 1}{n + a + b - 2}$$

**Remarque III.2.1.** On constate que la loi a priori et la loi a posteriori résultante sont de la même famille de loi Beta. Nous décrivants dans la section suivante ce type d'a priori particulière dite informatives.

### III.3 Quel choix a priori ?

Le choix d'une distribution a priori des paramètres dans un modèle bayésien est crucial afin d'apporter un maximum d'éléments en inférence statistique. Ce choix touche d'une façon directe la distribution a posteriori résultante de l'équation (III.2).

En théorie, il existe deux modes de pensées qui sont mis à l'évidence par rapport a un choix a priori. Le premier est “*subjectif*” : Il repose sur l'information disponible et l'expertise sur le paramètre obtenue et des opinions collectées. Cette expertise est exprimée par une loi de probabilité dite “*distribution a priori non-informative*”.

Cependant, la théorie bayésienne s'applique aussi dans le cas où on ne dispose pas d'une expertise ou connaissance a priori. Ce mode de pensée est plutôt “*objectif*” qui assigne une loi de probabilité a priori informative au sens qu'elle soit de la même famille que celle qu'on s'attend à avoir a posteriori.

#### III.3.1 Distribution a priori informative

La modélisation a priori informative est le point le plus délicat de l'analyse bayésienne. Il existe plusieurs procédés pour obtenir des lois informatives, même si nous décrivons dans cette partie, une façon des plus simples permettant de construire un a priori informatif relativement à la définition des familles naturelles conjuguées.

**Définition III.3.1** (Familles naturelles conjuguées). *Une famille  $\mathcal{F}$  de lois sur  $\Theta$  est dite conjuguée si, pour tout loi a priori  $\pi[\theta]$  appartenant à cette famille  $\mathcal{F}$ , la loi a posteriori  $\pi(\theta|x)$  appartient également à  $\mathcal{F}$ .*

**Remarque III.3.1.** *En pratique, on cherche à identifier le noyau d'une loi conjuguée par la formule (III.2), c'est-à-dire, si la forme fonctionnelle (par rapport aux paramètres du modèle fréquentiel) de la vraisemblance  $f(x|\theta)$  forme un noyau d'une loi de probabilité, alors le choix d'une loi a priori de la même forme permet la construction d'une loi a posteriori naturellement de la même forme (appartenant à la même famille).*

**Exemple III.3.1.** *Le tableau suivant regroupe quelques exemples de lois conjuguées, où la loi conditionnelle  $\pi(\theta|x)$  s'obtient proportionnellement au produit de la loi a priori  $\pi(\theta)$  et la vraisemblance  $f(x|\theta)$ .*

Vraisemblance $f(x \theta)$	Loi a priori $\pi(\theta)$	Loi a posteriori $\pi(\theta x)$
$\mathcal{N}(\theta, 1)$	$\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(x + \mu/\tau^2, [1 + 1/\tau^2]^{-1})$
$\mathcal{N}(0, \sigma^2)$	IGAMMA( $\alpha, \beta$ ) <sup>1</sup>	IGAMMA( $\alpha + 1/2, \beta + x^2/2$ )
$\mathcal{B}\text{in}(n, \theta)$	$\mathcal{B}_{[0,1]}(\alpha, \beta)$	$\mathcal{B}_{[0,1]}(\alpha + x, n + \beta - x)$
$\mathcal{P}(\lambda)$	GAMMA( $\alpha, \beta$ )	GAMMA( $\alpha + x, \beta + 1$ )

TABLE III.1 – Exemples de lois conjuguées usuelles

La distribution IGAMMA<sup>1</sup> est la loi conjuguée naturelle d'une loi normale lorsque le paramètre de position est connu.

1. Distribution dite inverse gamma de densité  $\frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{1+\alpha} \times \exp\left(-\frac{\beta}{\sigma^2}\right)$

**Exemple III.3.2.**

- Considérons une variable aléatoire  $X$  distribuée selon une loi de Pareto (type II) de paramètres  $(\theta, a)$ , définie par sa densité de probabilité :

$$f(x|\theta, a) = \frac{\theta a^\theta}{x^{\theta+1}} \mathbb{1}_{[a, +\infty[}(x)$$

En supposant  $a$  connu, la forme fonctionnelle par rapport à  $\theta$  de la vraisemblance est  $f(x|\theta) \propto \theta \exp[\theta \log(a/x)]$ , où on reconnaît ainsi le noyau d'une loi gamma (de paramètre 2 et  $\log(x/a)$ ). En prenant une loi a priori de type gamma, la loi a posteriori le sera aussi. La loi gamma est donc la famille conjuguée naturelle de la loi de Pareto.

- Un deuxième exemple plus simple est de considérer une loi binomiale négative de paramètre  $(n, p)$  :

$$\mathbb{P}(X = x | p) = C_{n+x-1}^x p^x (1-p)^n, \quad 0 < p < 1, \quad x \in \mathbb{N}$$

Clairement, une loi naturelle conjuguée sera une loi Beta puisque :

$$\mathbb{P}(X = x|p) \propto p^x (1-p)^n$$

L'approche conjuguée reste la solution la plus standard dans le cadre informatif. Une famille naturelle conjuguée est un concept qui simplifie considérablement le calcul des distributions a posteriori, c'est donc la raison du développement de ce type de lois a priori. Il faut noter que seuls les modèles à structure exponentielle admettent la construction d'une famille conjuguée.

**III.3.2 Distribution a priori non-informative**

Le concept non-informatif des lois a priori est vu comme une alternative dans le cas où la vraisemblance n'admet pas une structure exponentielle, permettant la construction d'une loi conjuguée naturelle. Le principe de base dans la conception d'une distribution non-informative est de faire une analyse bayésienne lorsqu'il y a absence de l'information a priori sur le paramètre d'intérêt, ou dans le cas où c'est difficile de traduire en terme de loi a priori l'information disponible sur les paramètres par une loi de probabilité.

Le choix d'une loi a priori non-informative conduit souvent à la spécification d'une mesure dite loi impropre et non pas d'une probabilité. L'approche est ainsi dite "objectif".

**Définition III.3.2** (Lois impropres). Une loi impropre est une loi qui est définie sur l'espace  $\Theta$  de densité  $\pi$  non intégrable

$$\int_{\Theta} \pi(\theta) d\theta = +\infty$$

Le choix de ce type de loi n'a qu'un intérêt calculatoire et s'interprète difficilement. Nous verrons par la suite que la construction de lois non-informatives peut conduire à des lois a priori impropres.

**Exemple III.3.3.** Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire i.i.d issu d'une loi exponentielle de paramètre  $\lambda > 0$ . Considérons que la loi a priori de  $\lambda$  est une loi impropre  $\pi(\lambda) = \mathbb{1}_{]0, +\infty[}(\lambda)$  (la loi uniforme sur  $\mathbb{R}^+$ ). Ainsi, la loi a posteriori est donnée par :

$$\pi(\lambda | x) = \frac{\lambda \exp(-\lambda x)}{\int_0^{+\infty} \lambda \exp(-\lambda x) d\lambda}$$

Le dénominateur est égal à  $\Gamma(2)/x^2$ , et la loi a posteriori  $\pi(\lambda | x) = \frac{x^2}{\Gamma(2)} \lambda \exp(-\lambda x)$ , représente une loi GAMMA(2, x).

**Exemple III.3.4.** On veut connaître la probabilité  $\theta$  d'avoir une face après le jet d'une pièce de monnaie. Nous imaginons avoir deux choix a priori :

**Cas 01 :** Toutes les valeurs de  $\theta$  sont équiprobables (a priori non-informatif) :  $\pi(\theta) = 1, \forall \theta \in [0, 1]$ .

**Cas 02 :** Il y a 75% de chances que  $\theta$  soit entre 60% et 80% (a priori subjectif), i.e :

$$\pi(\theta) = \frac{3}{4} \mathbb{1}_{[0.60, 0.80]}(\theta) + \frac{1}{4} \mathbb{1}_{[0, 0.60] \cup [0.80, 1]}(\theta)$$

Après avoir procédé à 10 lancers successifs, on obtient l'échantillon  $\{F, F, F, P, F, F, P, P, F, F\}$ , soit 7 faces et 3 piles. La probabilité d'obtenir 07 faces s'exprime en fonction de  $\theta$  par :

$$\mathbb{P}(X = 7) = 120\theta^7(1 - \theta)^3$$

Ainsi, la distribution a posteriori  $\pi(\theta|X = 7)$  est proportionnelle à

$$\theta^7(1 - \theta)^3\pi(\theta).$$

La distribution a posteriori du paramètre sera proche de la vraisemblance dans le cas 01, lorsque l'a priori est non-informatif ou si l'on dispose de beaucoup de données. Par contre il sera proche de la distribution a priori informative dans le second cas ou si l'on dispose de peu de données.

En pratique :

- Avec une taille d'échantillon usuelle, on s'attend à ce que  $\pi(\theta|y) \approx f(y|\theta)$ .
- On choisira un a priori  $\pi(\theta)$  non-informative si on souhaite que son impact sur la distribution a posteriori  $\pi(\theta|x)$  soit minimal. Graphiquement,  $\pi(\theta)$  sera plus plate que la vraisemblance  $f(x|\theta)$ , et donc dominée par celle-ci.
- Un a priori informatif n'est pas dominé par la vraisemblance. Son impact sur  $\pi(\theta|y)$  est plus important. Dans l'exemple précédant, on choisit d'un a priori  $\pi(\theta)$  informative de type Beta (car la forme fonctionnelle de la vraisemblance s'écrit sous un noyau d'une loi Beta), la loi a posteriori prendra la forme

$$\pi(\theta|X = 7) \propto \theta^{7+a-1}(1 - \theta)^{3+b-1},$$

$a$  et  $b$  désignent les hyper-paramètres de la loi a priori.

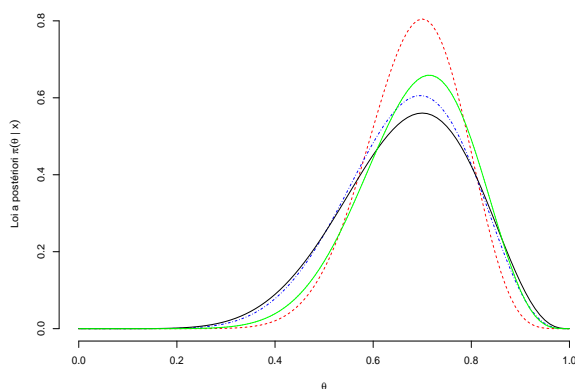


FIGURE III.1 – Effet des hyper-paramètres  $a$  et  $b$  sur la loi a posteriori

Dans la figure III.1, nous visualisons l'effet d'un choix des hyper-paramètres  $a$  et  $b$  de la loi a priori, sur la loi a posteriori  $\pi(\theta|X = 7)$ . La courbe en noire représente le choix d'une priori uniforme

(cas 01), celle en bleu (respectivement en vert et en rouge) est relative à un a priori  $\mathcal{B}_{[0,1]}(1, 0.5)$  (respectivement  $\mathcal{B}_{[0,1]}(3, 1)$  et  $\mathcal{B}_{[0,1]}(7, 3)$ ). Le constat graphique fait ressortir une allure de plus en plus concentrée de la a posteriori lorsque la valeur de  $a$  augmente. On soupçonne alors que la taille d'échantillon et l'information apportée par les observations n'a pas pu atténuer l'effet de la connaissance a priori.

### III.3.3 Loi a priori au sens de Jeffrey

Jeffreys (1961) proposa une technique permettant la construction de lois a priori non-informative, en utilisant la quantité d'information de Fisher notée  $I(\theta)$ . L'argument pourrait être le suivant :  $I(\theta)$  représente une mesure de la quantité d'information sur  $\theta$  contenue dans l'observation  $x$ . Plus  $I(\theta)$  est grande, plus l'observation apporte de l'information. Il sera naturel donc de favoriser (au sens rendre plus probable suivant  $\pi(\theta)$ ), les valeurs de  $\theta$  pour lesquelles la quantité d'information de Fisher est la plus grande (ce qui minimise l'influence de la loi a priori au profit des observations).

**Définition III.3.3** (Mesure a priori de Jeffrey). *Étant donné un échantillon de v.a  $X_1, X_2, \dots, X_n$ , issues d'une loi  $f(x | \theta)$ , où  $\theta$  est un paramètre réel, on appelle loi a priori non-informative au sens de Jeffrey, la loi (éventuellement impropre) donnée par :*

$$\pi^{(J)}(\theta) \propto [I(\theta)]^{1/2}$$

où  $I(\theta) = \mathbb{E}(-\frac{\partial^2}{\partial \theta^2} \log(f(X | \theta)))$  désigne la quantité d'information de Fisher apportée par  $x$  sur  $\theta$ .

**Remarque III.3.2.** *Dans le cas où,  $\theta$  est un paramètre vectoriel (de dimension  $p$  finie), et  $I(\theta)$  la matrice d'information de Fisher définie par :*

$$I(\theta) = (I_{ij}(\theta))_{i,j \in \{1, \dots, p\}} = \left( -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(X | \theta)) \right] \right)_{i,j \in \{1, \dots, p\}}$$

la mesure a priori au sens de Jeffrey s'écrit :

$$\pi^{(J)}(\theta) \propto (\det(I(\theta)))^{1/2}$$

#### Exemple III.3.5.

- Soit une v.a  $X \sim N(\mu, \sigma^2)$  de moyenne  $\mu$  et de variance  $\sigma^2$  supposées être des paramètres inconnus, et

$$f_X(x | \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}$$

Le paramètre vectoriel de la loi de  $X$  est  $\theta = (\mu, \sigma^2)$ . Nous déduisons facilement, la matrice d'information de Fisher  $I(\theta)$ , égale à :

$$\begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}$$

Ainsi, la loi a priori au sens de Jeffrey s'écrit :

$$\pi^{(J)}(\mu, \sigma) \propto \frac{1}{\sigma^2}$$

On remarque que la loi a priori obtenue est invariante par translation et homothétie. Ce constat représente d'une manière générale, une caractéristique principale des loi a priori au sens de Jeffrey.

- Étant donné une v.a  $Y$  distribuée selon une loi de Bernoulli de paramètre  $\theta \in [0, 1]$ .  
On a :  $I(\theta) = \frac{1}{\theta(1-\theta)} \mathbb{1}_{[0,1]}(\theta)$ , d'où :

$$\pi^{(J)}(\theta) \propto (\theta(1-\theta))^{-1/2} \mathbb{1}_{]0,1[}(\theta)$$

Ainsi, la loi a priori de  $\theta$  au sens de Jeffrey est une  $\mathcal{B}_{[0,1]}(1/2, 1/2)$ .

### III.3.4 La loi de Laplace

**Définition III.3.4.** La loi de Laplace (1774) correspond à un choix d'une loi a priori donnée par :

$$\pi(\theta) \propto \mathbb{1}_{\{\theta \in \Theta\}}$$

En fonction de l'espace des paramètres  $\Theta$ , nous retrouvons alternativement une loi uniforme sur  $\Theta$  ou une loi impropre. Le choix de la loi de Laplace peut sembler naturel car aucune valeur de paramètre n'est a priori favorisée par rapport à une autre. Or, l'inconvénient de cette loi impropre est qu'elle soit non-invariante par reparamétrisation bijective : En effet, si la reparamétrage  $\eta = g(\theta)$  est considérée (où  $g$  étant une fonction bijective)  $\theta$ , on obtient alors (par changement de variable) :

$$\pi(\theta) \propto 1 \implies \tilde{\pi}(\eta) = \left| \frac{d}{d\eta} g^{-1}(\eta) \right|$$

où  $\tilde{\pi}$  désigne la loi a priori correspondante à  $\eta = g(\theta)$ . Bien qu'aucune information a priori ne soit disponible sur  $\eta$ , puisqu'aucune information a priori n'est disponible sur  $\theta$ , le choix de loi a priori sur  $\eta$  n'est donc plus (en général) la loi de Laplace, et le choix de la loi a priori semble donc dépendre de la formulation même du problème.

**Exemple III.3.6.** Si  $p$  désigne la proportion de naissances d'une certaine catégorie d'espèce animale (que nous considérons distribuée selon une loi uniforme sur  $[0, 1]$ ) alors, le paramètre de rapport des chances  $\varrho(p) = \frac{p}{1-p}$  suit une loi a priori de densité  $1/(1+\varrho)^2$ , qui est donc non uniforme.

### III.3.5 Loi a priori de Zellner

**Définition III.3.5** (Zellner (1971)). Étant donné  $n$ -échantillon de variables aléatoires  $X_1, X_2, \dots, X_n$  supposées être indépendantes et identiquement distribuées selon une loi commune  $f(x|\theta)$  où  $\theta \in \Theta$ , on appelle loi a priori non-informative au sens de Zellner (1971), la loi impropre donnée par :

$$\pi^{(Z)}(\theta) \propto \exp(\mathbb{E}_X(\log f(x|\theta)))$$

La loi a priori est ainsi définie, de façon à maximiser l'information moyenne apportée par les données, par rapport à celle dans la loi a priori (voir [80], page 03).

**Exemple III.3.7.** Soit une v.a  $X \sim \text{Pareto}(\theta, 1)$ , donnée par sa fonction de répartition  $F(x) = 1 - x^{-\theta}$ . Le modèle bayésien s'écrit comme :

$$f(x|\theta) = \theta x^{-(\theta+1)}, \theta > 0.$$

Ainsi, la loi a priori au sens de Zellner est donnée par :

$$\pi^{(Z)}(\theta) \propto \theta \exp\left(-\frac{1}{\theta}\right)$$

### III.4 Modèle bayésien hiérarchique

Un modèle bayésien hiérarchique consiste à analyser l'information en faisant intervenir des niveaux successifs et conditionnels, en distinguant entre apports structurels et apports proprement subjectifs. La logique bayésienne repose essentiellement sur un modèle d'incertitude sur la loi a priori, en introduisant une loi sur les paramètres du modèle fréquentiel. Souvent, la loi du premier niveau est considérée conjuguée (par souci de simplification) tout en injectant autant de fois qu'il est nécessaire dans le premier modèle, des niveaux supérieurs (sur les hyper-paramètres du modèle a priori) afin de corriger éventuellement cette erreur de jugement a priori.

**Définition III.4.1.** *On appelle modèle bayésien hiérarchique, la donnée d'un modèle statistique bayésien  $(f(x|\theta), \pi(\theta))$ , où la loi a priori  $\pi(\theta)$  est décomposée en distributions conditionnelles :*

$$\pi_1(\theta|\theta_1), \pi_2(\theta_1|\theta_2), \dots, \pi_d(\theta_{d-1}|\theta_d)$$

*et en distribution marginale  $\pi_{d+1}(\theta_d)$ . Les paramètres  $\theta_i$  sont dits hyper-paramètres du  $i^e$  niveau hiérarchique de la loi a priori.*

**Remarque III.4.1.** *Notons qu'un modèle bayésien hiérarchique s'écrit comme :*

$$x \sim f(x|\theta), \quad \theta \sim \pi_1(\theta|\theta_1), \quad \theta_1 \sim \pi_2(\theta_1|\theta_2), \dots, \theta_{d-1} \sim \pi_d(\theta_{d-1}|\theta_d), \quad \theta_d \sim \pi_{d+1}(\theta_d),$$

Ainsi, nous pouvons considérer que le modèle bayésien usuel n'est rien qu'un cas particulier, écrit sous la forme

$$x \sim f(x|\theta), \quad \theta \sim \pi(\theta)$$

où

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_d} \pi_1(\theta|\theta_1) \pi_2(\theta_1|\theta_2) \dots \pi_d(\theta_{d-1}|\theta_d) \pi_{d+1}(\theta_d) d\theta_1 \dots d\theta_d$$

Dans la littérature moderne, les travaux de Gerfland et Smith (1990) sont considérés comme étant les fondements des techniques bayésiennes d'échantillonnage utilisant une structuration hiérarchique. Leur principe consiste à évaluer  $\pi(\theta|x)$  de la manière suivante :

$$\pi(\theta|x) = \int_{\Theta_1} \pi_1(\theta|x, \theta_1) \pi_2(\theta_1) d\theta_1$$

si les lois  $\pi_1(\theta|x, \theta_1)$ ,  $\pi_2(\theta_1)$  sont connues, générer  $\theta$  suivant  $\pi(\theta|x)$  revient en fait à générer  $\theta_1$  suivant  $\pi_2(\theta_1)$  puis  $\theta$  suivant  $\pi_1(\theta|x, \theta_1)$ .

### III.5 Consistance de la distribution a posteriori

L'inférence bayésienne sur d'un paramètre  $\theta$  issu du modèle  $f(x|\theta)$ , consiste essentiellement à choisir une connaissance a priori adéquate sur le paramètre  $\theta$ . Sous certaines conditions de régularité, la distribution a posteriori peut être approximée par une loi normale de moyenne l'estimateur du maximum de vraisemblance, et de variance l'inverse de la quantité d'information de Fisher (ou la matrice d'information de Fisher dans le cas où  $\theta$  est multidimensionnel). Si une plus grande précision est nécessaire, on peut alors utiliser un développement asymptotique de la loi postérieure. Une logique intuitive derrière la normalité de la distribution a posteriori est décrite ci-dessous.

Nous partons tout d'abord d'une question cruciale, qui est "Comment l'inférence bayésienne influencée par un choix a priori particulier, dépend-elle de l'importance relative à la quantité d'informations dans

les données?”. Cette quantité d’information (dans le cas unidimensionnel) peut être mesurée pour un échantillon i.i.d par

$$\hat{I}_n = -\frac{\partial^2 \log f(x|\theta)}{\partial^2 \theta} \Big|_{\theta=\hat{\theta}},$$

où  $\hat{\theta}$  est l’estimateur du maximum de vraisemblance de  $\theta$ .

Soit  $X_1, X_2, \dots, X_n$ , un échantillon de variables aléatoires supposées être i.i.d issue chacune d’une loi commune  $f(x|\theta_0)$ . L’approche bayésienne consiste à analyser cet échantillon conjointement avec un jugement a priori  $\pi(\theta)$  afin de nous rapprocher au mieux de  $\theta_0$ . L’idéal serait d’avoir un a priori capable de corriger toute déviation dans l’observation, en la rendant a posteriori, de plus en plus concentrée autour de  $\theta_0$ , quand la taille de l’échantillon augmente. Cette propriété asymptotique est connue sous le nom de ”Consistance de la distribution a posteriori” en  $\theta_0$ .

**Définition III.5.1.**<sup>1</sup> Une suite de mesures de probabilité aléatoire  $\mu_n = \pi(\cdot|X_1, X_2, \dots, X_n)$  sur  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$  est dite faiblement consistante si, pour un voisinage  $U$  de  $P_{\theta_0}$ , et  $\mu_n(U) \rightarrow 1$  quand  $n \rightarrow \infty$ ,  $P_{\theta_0}^{(n)}$ -presque surement, i.e :

$$\pi(\theta : d(\theta, \theta_0) < \epsilon | X_1, \dots, X_n) \rightarrow 1 \text{ in } P_{\theta_0}^{(n)} - \text{prob},$$

où  $P_{\theta_0}^{(n)}$  est la distribution des observations indexée par la vraie valeur  $\theta_0$ .

**Exemple III.5.1.** Étant donné  $X_1, X_2, \dots, X_n$  supposées être indépendantes et identiquement distribuées selon une loi de Bernoulli avec  $\mathbb{P}_\theta(X = 1) = \theta$ . Considérons une loi  $\mathcal{B}_{[0,1]}(\alpha, \beta)$  comme un a priori de  $\theta$ . La loi a posteriori est ainsi distribuée selon une loi  $\mathcal{B}_{[0,1]}(\alpha + \sum_{i=1}^n X_i, n + \beta - \sum_{i=1}^n X_i)$ , et :

$$\mathbb{E}(\theta | X_1, X_2, \dots, X_n) = \frac{\alpha + \sum_{i=1}^n X_i}{n + \alpha + \beta}, \quad \text{Var}(\theta | X_1, X_2, \dots, X_n) = \frac{(\alpha + \sum_{i=1}^n X_i)(n + \beta - \sum_{i=1}^n X_i)}{(n + \alpha + \beta)^2(n + \alpha + \beta + 1)}.$$

Par la loi des grands nombre,  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow \theta_0$   $P_{\theta_0}$ -presque surement. Ainsi  $\mathbb{E}(\theta | X_1, X_2, \dots, X_n) \rightarrow \theta_0$  et  $\text{Var}(\theta | X_1, X_2, \dots, X_n) \rightarrow 0$   $P_{\theta_0}$ -presque surement. D’où :

$$\begin{aligned} \mathbb{P}(\theta \notin \{\theta_0 - \epsilon, \theta_0 + \epsilon\} | X_1, X_2, \dots, X_n) &= \mathbb{P}(|\theta - \theta_0| > \epsilon | X_1, X_2, \dots, X_n) \\ &\leq \frac{\mathbb{E}[(\theta - \theta_0)^2 | X_1, X_2, \dots, X_n]}{\epsilon^2} \\ &= \frac{\text{Var}(\theta | X_1, X_2, \dots, X_n) + [\mathbb{E}(\theta | X_1, X_2, \dots, X_n) - \theta_0]^2}{\epsilon^2} \\ &\rightarrow 0 \text{ } P_{\theta_0} - \text{presque surement quand } n \rightarrow \infty. \end{aligned}$$

Une conséquence importante de cette consistance a posteriori (en distribution) réside dans la robustesse de l’inférence bayésien par rapport au choix de l’a priori.

Notons aussi, que la convergence de la distribution a posteriori vers  $P_{\theta_0}$  découle de l’analyse du concept de vitesse de convergence :

**Définition III.5.2.** La distribution a posteriori converge avec un taux  $\epsilon_n \rightarrow 0$ , pour une distance  $d$  (usuellement la distance de Hellinger) sur  $\theta_0$  si

$$\mathbb{E}_{\theta_0} [\pi(\theta : d(\theta, \theta_0) \leq \epsilon_n | X_1, \dots, X_n)] \rightarrow 1 \text{ when } n \rightarrow \infty$$

Pour plus de résultats sur le taux de convergence de la distribution a posteriori, nous recommandons de voir l’ouvrage [33].

1. See for instance Doob (1949), Schwartz (1965), Ghosh (2009)

### III.5.1 Normalité asymptotique de la distribution a posteriori

Les méthodes d'inférence bayésienne sont essentiellement basées sur une approximation gaussienne de la distribution a posteriori de  $\theta$ . Quand la taille de l'échantillon  $n$  augmente, la distribution a posteriori approche la normalité sous certaines conditions de régularité et se concentre au voisinage du mode a posteriori  $\tilde{\theta}$ . Supposons que la première dérivée partielle de  $\log \pi(\theta|X)$  s'annule en  $\tilde{\theta}$ , et définissons

$$\tilde{I}_n = -\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \Big|_{\theta=\tilde{\theta}}.$$

Un développement de Taylor d'ordre 2 au voisinage de  $\tilde{\theta}$  nous permet d'écrire :

$$\begin{aligned} \log \pi(\theta|X_1, X_2, \dots, X_n) &= \log \pi(\tilde{\theta}|X_1, X_2, \dots, X_n) - \frac{1}{2}(\theta - \tilde{\theta})^2 \times \left( -\frac{\partial^2 \log \pi(\theta|X_1, X_2, \dots, X_n)}{\partial \theta^2} \Big|_{\theta=\tilde{\theta}} \right) \\ &= \log \pi(\tilde{\theta}|X_1, X_2, \dots, X_n) - \frac{1}{2}(\theta - \tilde{\theta})^2 \times \tilde{I}_n \end{aligned}$$

Ainsi,

$$\begin{aligned} \pi(\theta|X_1, X_2, \dots, X_n) &= \pi(\tilde{\theta}|X_1, X_2, \dots, X_n) \exp \left( -\frac{1}{2}(\theta - \tilde{\theta})^2 \times \tilde{I}_n \right) \\ &\propto \exp \left( -\frac{1}{2}(\theta - \tilde{\theta})^2 \times \tilde{I}_n \right), \end{aligned}$$

Ce qui signifie que la loi a posteriori est proportionnelle à une densité gaussienne de moyenne  $\tilde{\theta}$  et de variance  $\times \tilde{I}_n^{-1}$ .

Mais, un résultat des plus importants sur la distribution a posteriori asymptotique, demeure le théorème de Bernstein, Von Mises, donnant la convergence dans l'espace  $L_1$  :

**Théorème III.5.1** (Bernstein (1917), Von Mises (1931)). *Sous certaines conditions de régularité, et pour  $n \rightarrow \infty$  on a*

$$\left\| \pi(\cdot|X_1, \dots, X_n) - \mathcal{N} \left( \hat{\theta}, \frac{\mathcal{I}_{\theta_0}^{-1}}{n} \right) (\cdot) \right\| \rightarrow 0 \text{ sous } P_{\theta_0}^{(n)},$$

où  $\mathcal{I}_{\theta_0}$  est la matrice d'information de Fisher,  $\hat{\theta}$  est un estimateur efficace de  $\theta_0$  et  $\|\cdot\|$  représente la distance de variation totale.

Sous les conditions suivantes :

H3 pour tout  $\theta \in \Theta$ , un support compact ne dépendant pas du paramètre.

H4 Une fonction  $\ell(\theta, \cdot)$  trois fois dérivable par rapport à  $\theta$  dans un voisinage  $U = (\theta_0 - \delta, \theta_0 + \delta)$  de  $\theta_0$  et  $\sup_{\theta \in U} \ell^{(3)}(\theta, \cdot) \leq M(\cdot) < \infty$

H5  $\mathbb{E}_{\theta_0} \left[ \frac{\partial}{\partial \theta} \ell(\theta_0, \cdot) \right] = 0$  et  $\mathbb{E}_{\theta_0} \left[ \frac{\partial}{\partial \theta} \ell(\theta_0, \cdot) \right]^2 = \mathbb{E}_{\theta_0} \left[ -\frac{\partial^2}{\partial \theta^2} \ell(\theta_0, \cdot) \right] < \infty$

H6 Pour tout  $\xi > 0$ ,  $\exists \epsilon > 0$  et une sous-suite  $k$  de  $n$  ( $k$  suffisamment grand) telle que :

$$\sup_{|\theta - \hat{\theta}| > \xi} \frac{1}{k} \left[ L_k(\theta) - L_k(\hat{\theta}) \right] < -\epsilon,$$

où

$$L_k(\theta) = \sum_{i=1}^k \log f(x_i|\theta).$$

le corollaire suivant découle du théorème précédent

**Corollaire III.5.1.** *Pour  $\hat{\theta}$  une solution consistante forte de l'équation  $\frac{\partial}{\partial \theta} L_k(\theta) = 0$ , et pour tout a priori  $\pi(\theta)$  continue et positive en  $\theta_0$  -la vraie valeur du paramètre- quand  $k \rightarrow \infty$  on a :*

$$\sqrt{k} (\theta_0 - \hat{\theta}) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_k^{-1}(\theta_0)),$$

où  $\mathcal{I}_k^{-1}(\theta_0) = \frac{1}{k} \mathbb{E} \left[ -\frac{\partial^2}{\partial \theta^2} \log f(X|\theta)|_{\theta=\theta_0} \right]^{-1}$  est l'inverse de la quantité d'information de Fisher.

**Théorème III.5.2.** <sup>2</sup> *En plus des conditions du corollaire III.5.1, si on suppose que la loi a priori  $\pi(\theta)$  admet un moment d'ordre 1 fini et  $\hat{\theta}^\bullet$  est un estimateur bayésien de la moyenne a posteriori, alors  $\sqrt{k}(\hat{\theta}^\bullet(k) - \hat{\theta}(k)) \rightarrow 0$ , avec  $\pi(\theta_0) = 1$  quand  $k \rightarrow \infty$ , et :*

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}} |w| \pi^W(w|X_1, X_2, \dots, X_n) - \frac{\sqrt{\mathcal{I}_k(\theta_0)}}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} w^2 \mathcal{I}_k(\theta_0) \right] |dw \rightarrow 0.$$

**Remarque III.5.1.** *le résultat dans le théorème III.5.2 implique que :*

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}} |w| \pi^W(w|X_1, X_2, \dots, X_n) dw \rightarrow \int_{\mathbb{R}} w \frac{\sqrt{\mathcal{I}_k(\theta_0)}}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} w^2 \mathcal{I}_k(\theta_0) \right] |dw = 0.$$

Comme  $\hat{\theta}^\bullet = \mathbb{E}(\theta|X_1, X_2, \dots, X_n) = \mathbb{E}(\hat{\theta} + \frac{w}{\sqrt{k}}|X_1, X_2, \dots, X_n)$ , alors :

$$\sqrt{k}(\hat{\theta}^\bullet(k) - \hat{\theta}(k)) = \int_{\mathbb{R}} w \pi^W(w|X_1, X_2, \dots, X_n) dw = 0 \text{ quand } k \rightarrow \infty.$$

### III.6 Estimation bayésienne de l'indice de queue $\alpha > 0$

En guise d'application des concepts bayésiens, nous tentons d'apporter une réponse bayésienne à une question des plus cruciales en théorie des valeurs extrêmes. À l'instar des travaux examinant la possibilité d'appliquer des approches bayésiennes (Coles and Powell (1996), Stephenson and Tawn (2004), Diebolt *et al.* (2005), de Zea Bermudez *et al.* (2010), Cabras and Castellanos (2011), do Nascimento *et al.* (2012), So and Chan (2014), ou aussi l'ouvrage de Beirlant *et al.* (2004, Chapitre 11)), nous présentons une alternative bayésienne d'estimation d'indice de queue  $\alpha > 0$ , en vue de reprendre la même philosophie de construction adoptée dans le cadre fréquentiel, en introduisant à la place des estimateurs du maximum de vraisemblance, des estimateurs bayésiens au sens maximum a posteriori (MAP) et moyenne a posteriori (MPE), en partant d'une loi a posteriori construite selon que l'a priori soit informative ou non. L'efficacité empirique des estimateurs proposés sera examinée par simulation. L'estimation d'indice de queue nécessite la sélection d'un seuil approprié, pour la construction d'un échantillon d'excès relatifs. Une procédure numérique de type MCMC sera mise en œuvre afin de permettre la construction d'estimateurs a posteriori.

Considérons un échantillon de *v.a*  $X_1, X_2, \dots, X_n$  supposées i.i.d, issues d'une loi commune  $F$ , dans le Max-domaine d'attraction de Fréchet ( $F \in MDA(\Phi_\alpha)$ ), i.e :

$$F \in MDA(\Phi_\alpha) \Leftrightarrow \lim_{t \rightarrow +\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}$$

2. voir Ghosh *et al.* (2006)

En fixant un seuil  $u > 0$  assez grand, notons  $n_u$  le nombre de variable  $X_i$  supérieure au seuil  $u$  (i.e  $N_u = \mathbb{1}_{\{X_i > u\}}$ ) et définissant la variable d'excès relatif  $Z = \{\frac{X}{u} | X > u\}$ . Ainsi, nous construisons un échantillon d'excès  $\tilde{Z} = Z_1, Z_2, \dots, Z_{n_u}$  distribuées selon une loi :

$$F_Z(z) = 1 - z^{-\alpha}, \quad \alpha > 0, z > 1$$

Ainsi, la fonction de vraisemblance relative à un échantillon observé de  $\tilde{z} = z_1, z_2, \dots, z_{n_u}$  est donnée par :

$$\begin{aligned} \mathcal{L}_u(\tilde{z}, \alpha) &= \prod_{i=1}^{n_u} \alpha z_i^{-(\alpha+1)} \\ &\propto \alpha^{n_u} \exp\left(-\alpha \sum_{i=1}^{n_u} \log z_i\right) \end{aligned}$$

### III.6.1 Approche bayésienne informative

L'écriture fonctionnelle de la vraisemblance  $\mathcal{L}_u(\tilde{z}, \alpha)$  par rapport à  $\alpha$ , de visualiser un noyau d'une loi Gamma. Ainsi, en optant pour une loi a priori conjuguée naturelle  $\pi(\alpha)$  la loi GAMMA( $a, b$ ), la loi a posteriori  $\pi(\alpha|z)$  sera de la même famille et s'obtient par :

$$\pi(\alpha|z) = \frac{\mathcal{L}_u(\tilde{z}, \alpha)\pi(\alpha)}{\mathcal{L}_u(\tilde{z})} = \frac{\mathcal{L}_u(\tilde{z}, \alpha)\pi(\alpha)}{\int_0^{+\infty} \mathcal{L}_u(\tilde{z}, \alpha)\pi(\alpha)d\alpha}.$$

Calculons tout d'abord, le dénominateur représentant la vraisemblance marginale :

$$\begin{aligned} \mathcal{L}_u(\tilde{z}) &= \int_0^{+\infty} \prod_{i=1}^{n_u} \alpha z_i^{-(\alpha+1)} \frac{b^a}{\Gamma(a)} (\alpha)^{a-1} e^{-b\alpha} d\alpha \\ &= \frac{b^a}{\Gamma(a)} \exp\left(-\sum_{i=1}^{n_u} \log(z_i)\right) \int_0^{+\infty} \alpha^{n_u+a-1} \exp\left(-\alpha \sum_{i=1}^{n_u} \log(z_i) + b\alpha\right) d\alpha \\ &= \frac{b^a}{\Gamma(a)} \exp\left(-\sum_{i=1}^{n_u} \log(z_i)\right) \frac{\Gamma(n_u + a)}{(\sum_{i=1}^{n_u} \log(z_i) + b)^{n_u+a}} \end{aligned}$$

Ainsi, la loi a posteriori est donnée par :

$$\pi(\alpha|\tilde{z}) = \frac{(\sum_{i=1}^{n_u} \log(z_i) + b)^{n_u+a}}{\Gamma(n_u + a)} \exp\left(-\alpha \sum_{i=1}^{n_u} \log(z_i) + b\alpha\right) \alpha^{n_u+a-1}.$$

Finalement, on en déduit qu'a posteriori

$$\alpha|\tilde{z} \sim \text{GAMMA}\left(n_u + a, \sum_{i=1}^{n_u} \log(z_i) + b\right).$$

L'estimateur moyenne a posteriori de  $\alpha$  s'écrit ainsi :

$$\begin{aligned} \hat{\alpha}_{MPE}^{(c)} &:= \mathbb{E}(\alpha|\tilde{z}) \\ &= \frac{a + n_u}{b + \sum_{i=1}^{n_u} \log(z_i)}, \end{aligned}$$

De même, l'estimateur du maximum a posteriori, s'obtient comme :

$$\begin{aligned}\hat{\alpha}_{MAP}^{(c)} &:= \arg \max_{\alpha} \{\pi(\alpha|\tilde{z})\} \\ &= \frac{a + n_u - 1}{b + \sum_{i=1}^{n_u} \log(z_i)}.\end{aligned}$$

**Définition III.6.1.** En remplaçant le seuil  $u$  par la statistique d'ordre  $n - k$ , les estimateurs bayésiens MPE et MAP de  $\alpha$ , relative a une loi a priori informative GAMMA( $a, b$ ), sont respectivement donnés par :

$$\begin{aligned}\hat{\alpha}_{MPE}^{(c)}(k) &= \frac{a + k}{b + \sum_{i=1}^k \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)} \\ \hat{\alpha}_{MAP}^{(c)}(k) &= \frac{a + k - 1}{b + \sum_{i=1}^k \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)}\end{aligned}$$

### III.6.2 Approche bayésienne non-informative

Une deuxième alternative serait de considérer une loi a priori non-informative au sens de Jeffrey et au sens de Zellner (Maximum Data Information prior).

#### Méthode du Jeffery's

Rappelons que la loi a priori impropre de  $\alpha$  au sens de Jeffrey s'écrit :

$$\pi^{(J)}(\alpha) \propto [\det(I_1(\alpha))]^{1/2}$$

où  $I_1(\alpha) = \mathbb{E}\left(-\frac{\partial^2}{\partial \alpha^2} \log(f(z|\alpha))\right)$  désigne la quantité d'information de Fisher apportée par une observation  $z$  sur  $\alpha$ .

Dans ce cas, on obtient :

$$\begin{aligned}\pi^{(J)}(\alpha) &\propto \left[ \mathbb{E} \left( -\frac{\partial^2}{\partial \alpha^2} \log(\alpha z^{-(\alpha+1)}) \right) \right]^{1/2} \\ &\propto \frac{1}{\alpha}\end{aligned}$$

Afin de déterminer la loi a posteriori :

$$\pi(\alpha|z) = \frac{\mathcal{L}_u(\tilde{z}, \alpha) \pi^{(J)}(\alpha)}{\mathcal{L}_u(\tilde{z})} = \frac{\mathcal{L}_u(\tilde{z}, \alpha) \pi^{(J)}(\alpha)}{\int_0^{+\infty} \mathcal{L}_u(\tilde{z}, \alpha) \pi^{(J)}(\alpha) d\alpha},$$

nous commençons par calculer la vraisemblance marginale :

$$\begin{aligned}\mathcal{L}_u(\tilde{z}) &= \int_0^{+\infty} \mathcal{L}_u(\tilde{z}, \alpha) \pi^{(J)}(\alpha) d\alpha \\ &= (z_i)^{-n_u} \prod_{i=1}^{n_u} \frac{\Gamma(n_u)}{(\sum_{i=1}^{n_u} \log(z_i))^{n_u}}\end{aligned}$$

d'où :

$$\pi(\alpha|z) = \frac{(\sum_{i=1}^{n_u} \log(z_i))^{n_u}}{\Gamma(n_u)} \alpha^{n_u-1} \exp\left(-\alpha \sum_{i=1}^{n_u} \log(z_i)\right)$$

qui représente la densité d'une loi  $\text{GAMMA}(n_u, \sum_{i=1}^{n_u} \log(z_i))$ .

L'estimateur moyenne a posteriori de  $\alpha$  au sens de Jeffrey, est donné par :

$$\hat{\alpha}_{MPE}^{(J)} = \frac{n_u}{\sum_{i=1}^{n_u} \log(z_i)}$$

L'estimateur maximum a posteriori de  $\alpha$  au sens de Jeffrey, s'obtient en maximisant la loi a posteriori, d'où :

$$\hat{\alpha}_{MAP}^{(J)} = \frac{n_u - 1}{\sum_{i=1}^{n_u} \log(z_i)}$$

**Définition III.6.2.** *En remplaçant le seuil  $u$  par la statistique d'ordre  $n - k$ , les estimateurs bayésiens MPE et MAP de  $\alpha$ , relatifs a une loi a priori non-informative au sens de Jeffrey, sont respectivement donnés par :*

$$\begin{aligned} \hat{\alpha}_{MPE}^{(J)}(k) &= \frac{k}{b + \sum_{i=1}^k \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)} \\ \hat{\alpha}_{MAP}^{(J)}(k) &= \frac{k - 1}{b + \sum_{i=1}^k \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)} \end{aligned}$$

**Remarque III.6.1.** *Nous constatons que l'estimateur moyenne a posteriori au sens de Jeffrey  $\hat{\alpha}_{MPE}^{(J)}(k)$ , correspond exactement à l'inverse de l'estimateur de Hill (estimateur de l'indice des valeurs extrêmes  $\gamma = \frac{1}{\alpha}$ ).*

### Approche Maximum Data Information (MDI)

Étant donné un modèle bayésien  $Z|\alpha \sim z^{-(\alpha+1)}$  et  $\alpha \sim \pi(\alpha)$ , la loi a priori non-informative au sens de Zellner est donné par :

$$\pi^{(Z)}(\alpha) \propto \exp(\mathbb{E}_Z(\log f(z|\alpha)))$$

où  $f(z|\alpha)$  représente la vraisemblance pour une observation  $z$ , ce qui nous permet d'écrire

$$\begin{aligned} \pi^{(Z)}(\alpha) &\propto \exp\left(\mathbb{E}_Z\left[\log \alpha z^{-(\alpha+1)}\right]\right) \\ &\propto \alpha \exp\left(-\frac{1}{\alpha}\right) \end{aligned}$$

et la loi a posteriori

$$\begin{aligned} \pi(\alpha|z) &\propto \mathcal{L}_u(\tilde{z}, \alpha) \pi^{(Z)}(\alpha) \\ &\propto \alpha^{n_u+1} \exp\left(-\frac{1}{\alpha} - \alpha \sum_{i=1}^{n_u} \log z_i\right) \end{aligned}$$

l'estimateur maximum a posteriori de  $\alpha$  est solution de l'équation

$$\frac{n_u + 1}{\alpha} + \frac{1}{\alpha^2} = \sum_{i=1}^{n_u} \log(z_i).$$

En multipliant les deux membres de l'équation précédente par  $\alpha^2$ , l'estimateur de maximum a posteriori (MAP) de  $\alpha$ , au sens de Zellner est :

$$\hat{\alpha}_{MAP}^{(Z)} = \left( \frac{B + \sqrt{B^2 + 4A}}{2A} \right)$$

où  $A = \sum_{i=1}^{n_u} \log(z_i)$  et  $B = n_u + 1$ . En remplaçant  $u$  par  $X_{n-k,n}$ , l'estimateur du maximum a posteriori au sens de Zellner correspond à

$$\hat{\alpha}_{MAP}^{(Z)}(k) = \frac{k+1}{2} \left( 1 + \sqrt{1 + 4 \frac{\sum_{i=1}^k \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)}{(k+1)^2}} \right) \left( \sum_{i=1}^k \log(X_{n-i+1,n}) - \log(X_{n-k,n}) \right)^{-1}.$$

**Remarque III.6.2.** Contrairement à l'estimateur maximum à posteriori (MAP), nous notons qu'il n'existe pas de forme explicite de l'estimateur moyenne a posteriori (MPE) qui s'écrit en fonction Gamma incomplète qu'on évalue numériquement.

### III.6.3 Comparaison entre les estimateurs bayésiens de $\alpha$

À présent, nous tentons à travers une procédure de ré-échantillonnage, d'établir une comparaison entre les différents estimateurs bayésiens ( $\hat{\alpha}_{MPE}^{(c)}$ ,  $\hat{\alpha}_{MAP}^{(c)}$ ,  $\hat{\alpha}_{MPE}^{(J)}$ ,  $\hat{\alpha}_{MAP}^{(J)}$  et  $\hat{\alpha}_{MAP}^{(Z)}$ ) selon le critère de minimisation de l'erreur quadratique moyenne.

L'expression de chacun des estimateurs bayésiens proposés est fonction du nombre d'excès  $k$ . Ce choix reste subjectif mais crucial : en pratique, si la vraie valeur de  $\alpha$  est connue, chaque estimateur est obtenu en un niveau  $k/n$  dit "niveau de fraction optimal" au sens

$$k_{opt} = \arg \min_k \mathbb{E} [(\hat{\alpha}^\bullet(k) - \alpha)^2]$$

où  $\hat{\alpha}^\bullet(k)$  est l'un des cinq estimateurs bayésiens. Si on considère  $m$  échantillons à répliqués, l'erreur moyenne quadratique est estimée par

$$\frac{1}{m} \sum_{i=1}^m (\hat{\alpha}^\bullet(k) - \alpha)^2$$

**Procédure de simulation :**

1. Pour une valeur de  $\alpha = \alpha_0$ , on simule un échantillon de taille  $n$  suivant une loi satisfaisante l'équation (II.1), c'est à dire, une loi à queue lourde.
2. Variant la valeur de  $k$  entre  $k_{min}$  et  $k_{max}$ , pour laquelle on calcul l'ensemble des cinq estimateurs proposés de  $\alpha$ .
3. on réplique l'étape 1 - 2  $m$  fois afin d'obtenir pour chaque estimateur,  $m$  valeurs d'estimations (pour tout  $k = k_{min}, \dots, k_{max}$ ). Calculer pour chaque estimateur, la valeur de

$$k_{opt} = \arg \min_k \frac{1}{m} \sum_{i=1}^m [\hat{\alpha}_i^\bullet(k) - \alpha_0]^2$$

puis en déduire le biais empirique d'estimation  $\sum_{i=1}^m (\hat{\alpha}_i^\bullet(k) - \alpha_0)$ , et la racine carré de l'erreur moyenne quadratique (qu'on note RMSE) donnée par

$$\frac{1}{m} \sqrt{\sum_{i=1}^m (\hat{\alpha}_i^\bullet(k_{opt}) - \alpha_0)^2}$$

la table III.6.3, dresse les différentes performances empiriques des cinq estimateurs présentés précédemment, sur un jeu de réplifications correspondant à  $m = 1000$ , d'échantillons simulés selon une loi de Fréchet, puis dans un deuxième cas selon une loi de Pareto (les deux distributions sont dans le max-domaine d'attraction de Fréchet), tout en variant la taille des échantillons simulés afin de détecter tout éventuel effet de taille ( $n = 100, 200, 500$  et  $1000$ ).

Une lecture de la table, nous permet de constater tout d'abord un effet de taille lié à la diminution de l'erreur moyenne quadratique (RMSE) lorsque la taille de l'échantillon  $n$  augmente. D'autre part, la valeur d'indice de queue  $\alpha$  influence la vitesse de convergence des estimateurs, car un choix de  $\alpha = 2$  semble réduire plus rapidement l'erreur quadratique que dans le cas où  $\alpha = 1$  (un effet probablement lié à l'existence du moment d'ordre 2, pour une distribution à queue lourde). En terme du choix d'estimateur bayésien performant, il est noté dans la plus part des 12 cas illustrés dans la table III.6.3, que l'estimateur  $\hat{\alpha}_{MAP}^{(Z)}$  renvoie les meilleurs performances en terme de minimisation de l'erreur moyenne quadratique.

### III.6.4 Estimation de $\alpha$ par une approche MCMC

L'estimation bayésienne est basée sur la construction d'une connaissance a posteriori dont l'expression est parfois compliquée, ou même difficile à mettre sous une forme analytique (voir inexploitable en terme de recherche d'estimateurs de type MAP ou MPE).

Une alternative permettant de palier à cette complexité, se traduit par l'introduction d'approches numériques dites "MCMC<sup>10</sup>". Ces procédures permettent la construction d'un noyau de transition (en fonction d'une distribution d'intérêt donnée), telle que les suites d'échantillons simulées par la procédure MCMC forment une chaîne de Markov convergente vers la distribution stationnaire désirée.

Nous présentons dans cette partie (à titre d'exemple uniquement), la mise en œuvre d'une telle approche numérique quant à l'estimation de l'indice des valeurs extrêmes  $\alpha$  par l'algorithme de Metropolis-Hastings [56] (1953) et [43] (1970) (car en travail dans un cadre d'estimation unidimensionnel). Nous suggérons aux lecteurs de voir l'ouvrage de Robert *et al.* (2011) pour plus de détails.

#### Algorithme de Metropolis-Hastings

Considérons que la distribution a posteriori de  $\alpha$  est une loi  $Gamma(a, b)$ . La procédure itérative de Metropolis-Hastings consiste, pour toute distribution a posteriori  $\pi(\alpha | z)$ , en la génération successive d'une suite de valeurs  $\alpha_i$  construite en deux temps comme suit :

1. Choisir un candidat  $\alpha^*$  tiré aléatoirement selon une distribution de probabilité  $q(\alpha^* | \alpha_i)$  (dite loi de proposition ou noyau de transition), éventuellement dépendante de  $\alpha_i$ .
2. Le candidat  $\alpha^*$  est accepté avec une probabilité :

$$\rho(\alpha_i, \alpha^*) = \min \left( 1, \frac{\pi(\alpha^* | x)q(\alpha_i | \alpha^*)}{\pi(\alpha_i | x)q(\alpha^* | \alpha_i)} \right)$$

En pratique, on tire à chaque itération, une réalisation  $u$  suivant une loi  $\mathcal{U}_{[0,1]}$

$$\alpha_{i+1} = \begin{cases} \alpha^* & \text{si } u \leq \rho \\ \alpha_i & \text{sinon} \end{cases}$$

10. MCMC : Markov chaine Monte Carlo

Distribution	Taille d'éch. $n$		$\hat{\alpha}_{MAP}^{(Z)}$	$\hat{\alpha}_{MPE}^{(c)}$	$\hat{\alpha}_{MAP}^{(c)}$	$\hat{\alpha}_{MPE}^{(J)}$	$\hat{\alpha}_{MAP}^{(J)}$
Fréchet ( $\alpha=1$ )	$n=100$	$k_{opt}/n$	<b>0.2300</b>	0.3400	0.3100	0.2600	0.2800
		biais( $\hat{\alpha}$ )	-0.0263	0.0092	0.1316	0.0569	0.1293
		RMSE( $\hat{\alpha}$ )	<b>0.0314</b>	0.0509	0.0704	0.0485	0.0672
	$n=200$	$k_{opt}/n$	<b>0.2350</b>	0.3450	0.4300	0.2550	0.3050
		biais( $\hat{\alpha}$ )	0.0266	0.0671	0.1063	0.0637	0.0863
		RMSE( $\hat{\alpha}$ )	<b>0.0196</b>	0.0261	0.0374	0.0258	0.0296
	$n=500$	$k_{opt}/n$	<b>0.2284</b>	0.3425	0.3950	0.2345	0.3440
		biais( $\hat{\alpha}$ )	-0.0335	-0.0391	-0.0531	-0.0497	-0.0579
		RMSE( $\hat{\alpha}$ )	<b>0.0085</b>	0.0088	0.0108	0.0103	0.0116
Fréchet ( $\alpha=2$ )	$n=100$	$k_{opt}/n$	<b>0.2400</b>	0.3210	0.3910	0.2500	0.3310
		biais( $\hat{\alpha}$ )	0.0136	0.0360	0.0600	0.0374	0.0567
		RMSE( $\hat{\alpha}$ )	<b>0.0083</b>	0.0110	0.0134	0.0127	0.0147
	$n=200$	$k_{opt}/n$	<b>0.2450</b>	0.3920	0.3150	0.2900	0.3050
		biais( $\hat{\alpha}$ )	0.0205	0.0355	0.0433	0.0316	0.0510
		RMSE( $\hat{\alpha}$ )	<b>0.0048</b>	0.0064	0.0070	0.0069	0.0079
	$n=500$	$k_{opt}/n$	<b>0.2300</b>	0.2980	0.3100	0.2500	0.2980
		biais( $\hat{\alpha}$ )	-0.0087	-0.0095	-0.0106	-0.0107	-0.0122
		RMSE( $\hat{\alpha}$ )	<b>0.0032</b>	0.0035	0.0041	0.0038	0.0043
Pareto ( $\alpha=1$ )	$n=200$	$k_{opt}/n$	0.2800	0.3850	0.4050	<b>0.2500</b>	0.34250
		biais( $\hat{\alpha}$ )	0.1174	0.1307	0.8621	0.0429	0.0106
		RMSE( $\hat{\alpha}$ )	0.0655	0.0735	0.0883	<b>0.0483</b>	0.0504
	$n=500$	$k_{opt}/n$	0.2500	0.2980	0.3280	<b>0.2400</b>	0.3200
		biais( $\hat{\alpha}$ )	0.0182	0.0255	0.0111	0.0056	-0.0019
		RMSE( $\hat{\alpha}$ )	0.0088	0.0080	0.0078	<b>0.0073</b>	0.0082
	$n=1000$	$k_{opt}/n$	<b>0.2500</b>	0.2990	0.3010	0.2600	0.3220
		bais( $\hat{\alpha}$ )	-0.0027	0.0163	0.0070	0.0055	-0.0042
		RMSE( $\hat{\alpha}$ )	<b>0.0034</b>	0.0042	0.0042	0.0038	0.0038
Pareto ( $\alpha=2$ )	$n=200$	$k_{opt}/n$	<b>0.2350</b>	0.4500	0.3100	0.2700	0.2850
		bais( $\hat{\alpha}$ )	0.0023	0.0308	0.0894	<b>0.0007</b>	0.0147
		RMSE( $\hat{\alpha}$ )	<b>0.0043</b>	0.0045	0.0051	0.0053	0.0058
	$n=500$	$k_{opt}/n$	<b>0.2380</b>	0.3220	0.3460	0.2800	0.4020
		bais( $\hat{\alpha}$ )	0.0359	0.0421	0.0207	0.0020	0.0078
		RMSE( $\hat{\alpha}$ )	<b>0.0308</b>	0.0570	0.0352	0.0321	0.0356
	$n=1000$	$k_{opt}/n$	0.2500	0.3010	0.3550	<b>0.2430</b>	0.3440
		bais( $\hat{\alpha}$ )	-0.0012	-0.0003	0.0026	-0.0019	0.0024
		RMSE( $\hat{\alpha}$ )	0.0009	0.0009	0.0010	<b>0.0008</b>	0.0011

TABLE III.2 – Estimation bayésienne de l'indice de queue  $\alpha$  relativement à une distribution de Fréchet :  $F(x) = \exp(-x^{-\alpha})$ , et une distribution de Pareto :  $\bar{F}(x) = x^{-\alpha}$

Le choix initial de  $\alpha^{(0)}$  doit avoir une probabilité positive

$$\mathbb{P}(\alpha^{(0)} | z) > 0$$

L'algorithme initial de Metropolis dans [56], correspondait seulement au cas des noyaux de transition symétrique (i.e  $q(x | y) = q(y | x)$ ) et dans ce cas, le ratio  $\rho$  prend la forme simple

$$\rho(\alpha_i, \alpha^*) = \min \left( 1, \frac{\pi(\alpha^* | z)}{\pi(\alpha_i | z)} \right)$$

**Exemple d'application** On simule un échantillon de taille  $n = 1000$ , issu d'une loi de Fréchet standard (de paramètre  $\alpha = 1$ ), et on fixe un seuil  $u$  correspondant à l'observation de  $n - k^{\text{ème}}$  rang. Supposons que la loi a priori de  $\alpha$  est non-informative au sens de Jeffrey, i.e :  $\pi(\alpha) \propto \frac{1}{\alpha}$ . La loi a posteriori de  $\alpha$  est ainsi une loi du GAMMA  $\left(n_u, \sum_{j=1}^{n_u} \log(z_j)\right)$ , où  $n_u$  représente le nombre d'observations  $x_i > u$  et  $z_j = \frac{x_i}{u} | x_i > u$ .

Le programme suivant permet d'estimer a posteriori le paramètre  $\alpha$ , on simulant une séquence (de taille  $N = 10000$ ) de la loi cible GAMMA par l'algorithme de Metropolis-Hastings (MH)<sup>11</sup> :

```
> library("evd", lib.loc="~/R/win-library/3.3")
> x=rfréchet(1000)
> N=10000;
> burn.in=500;
> u <- 10;
> a=length(x[x>u]);
> b=sum(log(x[x>u]/u));
> d <- rnorm(N, 0, 1)
> u1 <- runif(N);
> y<- vector("numeric",N)
> y0=0;
> y[1]=y0;
> for(i in 2:N) {
+   y[i] <- y[i-1] + d[i] # valeur proposée ou candidate
+   v0 = dgamma(y[i-1], shape=a, scale=(1/b))
+   v1 = dgamma(y[i], shape=a, scale=(1/b))
+   if(v0*u1[i] > v1) {y[i] <- y[i-1]} # condition de rejet
+ }
> return(y)
```

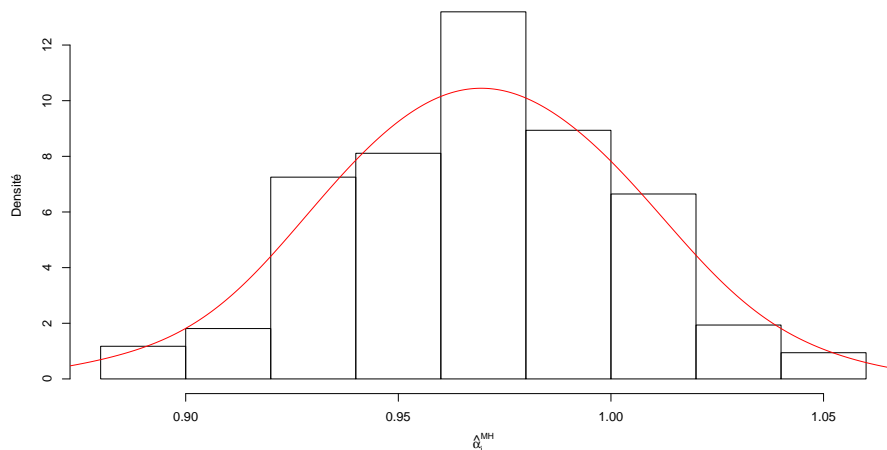


FIGURE III.2 – Histogramme des réalisations de l'algorithme MH

11. On choisissant un noyau de transition  $q(x|y)$  gaussienne, symétrique.

La figure III.2 représente l'histogramme des 9500 réalisations (soit un burn-in à partir de la 501<sup>e</sup> réalisation de l'algorithme de MH), issues de l'algorithme de MH précédent, dont la loi cible est une  $\text{GAMMA}(n_u, \sum_{j=1}^{n_u} \log(z_j))$ .

Un test statistique de type Kolmogorov-Smirnov permet de confirmer l'adéquation de la suite simulée par l'algorithme MH, à la loi cible

```
> ks.test(y[501:10000], "pgamma", a, 1/b)
One-sample Kolmogorov-Smirnov test
data: y[501:10000]
D = 0.01370, p-value = 0.20578
alternative hypothesis: two-sided
```

Une estimation de  $\alpha$  par l'algorithme MH est donnée par


$$\hat{\alpha}^{MH} = \frac{1}{N - m} \sum_{i=m+1}^N \alpha^{(i)} = 0.9696029$$

où  $m$  désigne le "burn-in", et  $\alpha^{(i)}, i = \{m + 1 \dots N\}$  représente la valeur de la chaîne de Markov obtenue à partir de l'algorithme MH.

# CHAPITRE IV

## QUELQUES CONTRIBUTIONS À L'ESTIMATION DE L'INDICE DE QUEUE $\alpha$ POUR UNE DISTRIBUTION A QUEUE LOURDE EN PRÉSENCE DE CENSURE ALÉATOIRE À DROITE

### IV.1 Première Partie : Bayesian estimation of the tail index of a heavy tailed dis- tribution under random censoring

ous présentons dans cette partie, un nouveau type d'estimateurs bayésiens, de l'indice de queue  $\alpha$  pour une distribution à queue lourde ( $\alpha > 0$ ), et différents types de lois a priori. Ces estimateurs sont construits sous un modèle semi-paramétrique, et en considérant un échantillon de données *i.i.d* censurées aléatoirement à droite. Nous établirons la convergence asymptotique de la loi a posteriori de  $\alpha$  vers une distribution gaussienne. Les propriétés empiriques de chacun des estimateurs proposés seront discutée à l'aide d'une étude par simulation et une application sur un jeu de données réelles.



# Bayesian estimation of the tail index of a heavy tailed distribution under random censoring



Abdelkader Ameraoui<sup>a,\*</sup>, Kamal Boukhetala<sup>a</sup>, Jean-François Dupuy<sup>b</sup>

<sup>a</sup> Faculty of Mathematics, PoBox 32, Al alia Bab ezzouar, USTHB - Algiers, Algeria

<sup>b</sup> IRMAR-INSA de Rennes, 20 Avenue des Buttes de Coësmes, 35708 Rennes cedex 7, France

## ARTICLE INFO

### Article history:

Received 31 July 2015  
Received in revised form 20 May 2016  
Accepted 17 June 2016  
Available online 11 July 2016

### Keywords:

Extreme value modeling  
Random censoring  
Maximum a posteriori estimator  
Mean posterior estimator  
Asymptotic normality of posterior  
Simulations

## ABSTRACT

Bayesian estimation of the tail index of a heavy-tailed distribution is addressed when data are randomly right-censored. Maximum a posteriori and mean posterior estimators are constructed for various prior distributions of the tail index. Convergence of the posterior distribution of the tail index to a Gaussian distribution is established. Finite-sample properties of the proposed estimators are investigated via simulations. Tail index estimation requires selecting an appropriate threshold for constructing relative excesses. A Monte Carlo procedure is proposed for tackling this issue. Finally, the proposed estimators are illustrated on a medical dataset.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Tail index estimation is one of the most important issues in extreme value theory. The tail index measures the thickness of the tail of a probability distribution function and thus plays a crucial role for evaluating the risk of occurrence of extreme events. In particular, estimation of the tail index constitutes usually a first step in an extreme value analysis. A vast literature has been dedicated to this topic. Recent overviews can be found in the monographs (Beirlant et al., 2004; Coles, 2001; de Haan and Ferreira, 2006).

Let  $F$  be the cumulative distribution function (cdf) of some non-negative random variable  $Y$ . We assume that  $F$  is heavy-tailed, that is, there exists a constant  $\alpha > 0$  such that

$$1 - F(x) = x^{-\alpha} \ell(x), \quad (1)$$

where  $\ell$  is a slowly varying function at infinity:

$$\lim_{x \rightarrow \infty} \frac{\ell(tx)}{\ell(x)} = 1 \quad \text{for all } t > 0.$$

If (1) holds, we have:

$$\lim_{x \rightarrow \infty} \frac{1 - F(tx)}{1 - F(x)} = t^{-\alpha} \quad \text{for all } t > 0$$

\* Corresponding author.

E-mail addresses: [aameraoui@usthb.dz](mailto:aameraoui@usthb.dz) (A. Ameraoui), [kboukhetala@usthb.dz](mailto:kboukhetala@usthb.dz) (K. Boukhetala), [Jean-Francois.Dupuy@insa-rennes.fr](mailto:Jean-Francois.Dupuy@insa-rennes.fr) (J.-F. Dupuy).

and we say that  $\bar{F} = 1 - F$  is regularly varying at infinity with tail index  $\alpha$ , which we denote by  $\bar{F} \in \mathcal{R}_{-\alpha}$ . The positive number  $\gamma := \alpha^{-1}$  is called the extreme value index (EVI) of  $F$ . The conditions above amount to assuming that the distribution function  $F$  is in the max-domain of attraction of a Fréchet distribution. Such distribution functions are useful in practice for investigating phenomena where exceptional values have a significant occurrence frequency. Examples include the number of claims in insurance (Embrechts et al., 1997), transmission times in telecommunications (Resnick, 2007), log-returns of price speculation (Embrechts et al., 1997).

Several estimators have been proposed for the tail index  $\alpha$ , or equivalently, for the EVI  $\gamma$ , such as Pickands estimator (Pickands, 1975), Dekkers et al. (or moment) estimator (Dekkers et al., 1989) and Hill estimator (Hill, 1975), which is the most popular estimator of  $\gamma$  in model (1). Let  $X_1, X_2, \dots, X_n$  be independent and identically distributed (i.i.d. thereafter) random variables with common cdf  $F$ . Let  $k \in \{2, \dots, n\}$  and  $X_{n,1} \leq X_{n,2} \leq \dots \leq X_{n,n}$  be the order statistics of the sample  $X_1, X_2, \dots, X_n$ . Hill estimator is defined as

$$H(k) := \frac{1}{k} \sum_{i=1}^k \log(X_{n,n-i+1}) - \log(X_{n,n-k}). \tag{2}$$

Consistency of Hill estimator was proved in Mason (1982) under the regular variation condition (1). Its asymptotic normality was further established under an additional condition known as the second-order regular variation condition (see de Haan and Ferreira, 2006).

In this paper, we address the estimation of the tail index  $\alpha$  when observations  $X_1, X_2, \dots, X_n$  are randomly right-censored. Censoring commonly occurs in the analysis of event time data. For example,  $X$  may represent the duration until the occurrence of some event of interest, such as death of a patient. If a patient is still alive or has dropped out of the study for some reason when the data are collected, the variable of interest  $X$  is not available. An appropriate way to model this situation is to introduce a random variable  $Y$  (called a censoring random variable) such that observations consist of pairs  $(Z_i, \delta_i)$ ,  $1 \leq i \leq n$  where  $Z_i = \min(X_i, Y_i)$ ,  $\delta_i = \mathbb{1}_{\{X_i \leq Y_i\}}$  and  $\mathbb{1}$  is the indicator function. Estimation of the EVI with censored data was considered in Beirlant et al. (2007), Brahimi et al. (2013), Einmahl et al. (2008) and Gomes and Neves (2011). For example, Beirlant et al. (2007) proposed to estimate  $\gamma$  by the following modified version of Hill estimator:

$$H^C(k) := \frac{\sum_{i=1}^k \log(Z_{n,n-i+1}) - \log(Z_{n,n-k})}{\sum_{i=1}^k \delta_{[n-i+1]}}, \tag{3}$$

where  $Z_{n,1} \leq Z_{n,2} \leq \dots \leq Z_{n,n}$  are the order statistics of the censored sample  $Z_1, Z_2, \dots, Z_n$  and  $\delta_{[n-i+1]}$  is the concomitant value of  $\delta$  associated with  $Z_{n,n-i+1}$ , that is,  $\delta_j = \delta_{[n-i+1]}$  if  $Z_j = Z_{n,n-i+1}$ . More generally, Einmahl et al. (2008) proposed to estimate  $\gamma$  by

$$\hat{\gamma}_Z^{(C)}(k) = \frac{\hat{\gamma}_Z(k)}{\hat{p}}, \tag{4}$$

where  $\hat{\gamma}_Z(k)$  is any of the classical EVI estimators calculated on the censored observations  $Z_1, \dots, Z_n$  and  $\hat{p} = \frac{1}{k} \sum_{i=1}^k \delta_{[n-i+1]}$  is the proportion of uncensored values in the  $k$  largest observations of  $Z$ . Obviously, the estimator (4) coincides with the adapted Hill estimator (3) when  $\hat{\gamma}_Z(k)$  is Hill estimator (2). In this paper, we adopt a completely different approach and investigate Bayesian estimation of the tail index  $\alpha := \gamma^{-1}$ . Bayesian estimation will allow us to incorporate *a priori* knowledge about the data.

Bayesian estimation provides an alternative to frequentist methods. In the context of extreme values analysis without censoring, Bayesian estimators have been investigated in Cabras and Castellanos (2011), Coles and Powell (1996), Diebolt et al. (2005), do Nascimento et al. (2012), Stephenson and Tawn (2004) and de Zea Bermudez and Kotz (2010). See also Beirlant et al. (2004) (Chapter 11). Applications include operational risk modeling (Ergashev et al., 2013), hydrology (Liang et al., 2011) and market indices modeling (So and Chan, 2014). But to the best of our knowledge, no Bayesian estimator of the tail index  $\alpha$  has been proposed when censoring is present. The present work intends to fill in this gap.

We construct several Bayesian estimators of  $\alpha$  in model (1). Bayesian estimation requires specifying a prior distribution for the unknown parameter. We investigate various priors (Jeffreys, maximal data information), leading to several maximum a posteriori and mean posterior estimators of  $\alpha$ . Convergence of the posterior distribution of  $\alpha$  to a Gaussian distribution is established. Finite-sample performance of the proposed estimators is assessed via simulations. Tail index estimation requires choosing an appropriate threshold for defining excesses. We propose a randomization method for tackling this issue. This procedure is evaluated by simulation. Finally, we illustrate our methodology on a set of AIDS survival data.

The paper is organized as follows. In Section 2, we construct our estimators and consider asymptotic results. Proofs are deferred to an Appendix. Section 3 reports the results of a comprehensive simulation study. The randomization procedure for threshold selection is proposed and assessed by simulations in Section 4. Application to AIDS data is carried out in Section 5. A discussion and some perspectives are given in Section 6.

## 2. Bayesian estimation of the tail index

We first introduce some notations. Let  $X_1, X_2, \dots, X_n$  be  $n$  i.i.d. copies of a non-negative random variable  $X$  with cdf  $F$ . The probability density function of  $X$  is denoted by  $f$ . We assume that  $F$  is heavy-tailed with tail index  $\alpha$ , i.e.,  $\bar{F} \in \mathcal{R}_{-\alpha}$ . Let  $Y_1, Y_2, \dots, Y_n$  be an i.i.d. sample of the non-negative censoring random variable  $Y$  with cdf  $G$  and density function  $g$ . We assume that  $G$  is also heavy-tailed, with positive tail index  $\beta$ , i.e.,  $\bar{G} \in \mathcal{R}_{-\beta}$ . We assume that  $X$  and  $Y$  are independent and that we observe the  $n$  independent pairs

$$(Z_i, \delta_i), \quad 1 \leq i \leq n,$$

where  $Z_i = \min(X_i, Y_i)$  and  $\delta_i = \mathbb{1}_{\{X_i \leq Y_i\}}$ . Let  $Z_{n,1} \leq Z_{n,2} \leq \dots \leq Z_{n,n}$  be the order statistics of the sample  $(Z_1, Z_2, \dots, Z_n)$  and  $\delta_{[n-i+1]}$  be the concomitant value of  $\delta$  associated with  $Z_{n,n-i+1}$ . Let  $H$  be the cdf of  $Z$ . Note that  $H$  is also heavy-tailed and  $\bar{H} \in \mathcal{R}_{-(\alpha+\beta)}$ , by independence of  $X$  and  $Y$ .

Let  $E_{j,u} := \frac{Z_{n,n-N_u+j}}{u}$ , given  $Z_{n,n-N_u+j} > u$ , be the  $j$ th relative excess over a threshold  $u$  and  $N_u$  denote the number of such excesses. For  $j = 1, \dots, N_u$ ,  $E_{j,u}$  satisfies  $\mathbb{P}(E_{j,u} > t) \rightarrow t^{-\alpha}$  for  $t > 1$ , as  $u \rightarrow +\infty$ . We can obtain the partial likelihood of  $\alpha$  based on the sample  $(E, \Delta) = (e_{j,u}, \delta_{[n-N_u+j]})_{j=1, \dots, N_u}$  (see Andersen et al., 1993; Beirlant et al., 2007):

$$\begin{aligned} \mathcal{L}_u(E, \Delta | \alpha) &= \prod_{j=1}^{N_u} [\alpha e_{j,u}^{-(\alpha+1)}]^{\delta_{[n-N_u+j]}} [e_{j,u}^{-\alpha}]^{1-\delta_{[n-N_u+j]}} \\ &= \alpha^{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]}} \left( \prod_{j=1}^{N_u} e_{j,u} \right)^{-\alpha} \left( \prod_{j=1}^{N_u} e_{j,u}^{-\delta_{[n-N_u+j]}} \right). \end{aligned} \quad (5)$$

In Bayesian framework, the initial set of beliefs about unknown parameters is represented by a probability density function called ‘‘prior density’’. The prior distribution is updated by using information contained in the data, yielding the posterior density of the parameters. In our setting, we provide the unknown tail index  $\alpha$  with a prior density  $\pi(\alpha)$ . Then, using Bayes theorem, we obtain the posterior density  $\pi(\alpha | E, \Delta) = \mathcal{L}_u(E, \Delta | \alpha) \pi(\alpha) / \int_{\Lambda} \mathcal{L}_u(E, \Delta | \alpha) \pi(\alpha) d\alpha$  of  $\alpha$ , where  $\Lambda$  is the support of the distribution of  $\alpha$ . The posterior distribution of  $\alpha$  is proportional to the product of the partial likelihood (5) and the prior, namely:  $\pi(\alpha | E, \Delta) \propto \mathcal{L}_u(E, \Delta | \alpha) \pi(\alpha)$ . Choosing the prior density is a central issue in Bayesian estimation. When information available for prior elicitation is minimal, one can use objective (or non-informative) priors, such as Jeffreys prior (Jeffreys, 1961) and the maximal data information (MDI) prior introduced by Zellner (1971) (see Zellner, 1971 and Zellner, 1996). On the other hand, when prior choice can be based on experts opinion, subjective prior distributions can be used, such as conjugate priors (conjugate priors are such that prior and posterior distributions are in the same family). In Sections 2.1 and 2.2, we investigate MDI and Jeffreys priors for the estimation of the tail index  $\alpha$  under random censoring.

### 2.1. Maximal data information (MDI) prior

Zellner (1971) defined the MDI prior so as to maximize the average information in the data density relative to that in the prior. If  $\mathcal{L}_u(E, \Delta | \alpha)$  is the likelihood of a single observation  $(E, \Delta)$ , then the MDI prior for  $\alpha$  is defined as  $\pi(\alpha) \propto \exp[\mathbb{E}(\log \mathcal{L}_u(E, \Delta | \alpha))]$  (see Beirlant et al., 2004, Chapter 11.3). As  $u \rightarrow \infty$ , the distribution of log-relative excesses over  $u$  is approximately exponentially distributed with parameter  $\alpha$ , and calculations yield the following MDI prior for  $\alpha$ :

$$\pi(\alpha) \propto \alpha \exp\left[-\frac{1}{\alpha}\right]. \quad (6)$$

Using (5) and (6), the posterior density of  $\alpha$  based on MDI prior is given by:

$$\pi(\alpha | E, \Delta) \propto \alpha^{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]}} \exp\left[-\alpha \sum_{j=1}^{N_u} \log e_{j,u}\right] \alpha \exp\left[-\frac{1}{\alpha}\right],$$

and letting  $a_u := \sum_{j=1}^{N_u} \log e_{j,u}$  and  $b_u := \sum_{j=1}^{N_u} \delta_{[n-N_u+j]}$ , the log-posterior density  $\ell(\cdot | E, \Delta)$  of  $\alpha$  satisfies:

$$\ell(\alpha | E, \Delta) \propto (1 + b_u) \log(\alpha) - a_u \alpha - \frac{1}{\alpha}.$$

From this, we can define the maximum a posteriori (or MAP) tail index estimator as:

$$\hat{\alpha}_{MAP}^{(MDI)} = \arg \max_{\alpha} \ell(\alpha | E, \Delta) := \frac{B + \sqrt{B^2 + 4A}}{2A}, \quad (7)$$

where  $A = a_u$  and  $B = 1 + b_u$ . Letting  $u = Z_{n,n-k}$  in (7), we obtain the following formal definition of our MAP estimator of  $\alpha$  under MDI prior:

**Definition 2.1.** Let  $k := k_n$  be a sequence such that  $k \rightarrow \infty$  and  $\frac{k}{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Then the MAP estimator of the tail index  $\alpha$  under MDI prior is given by:

$$\hat{\alpha}_{MAP}^{(MDI)} := \frac{B + \sqrt{B^2 + 4A}}{2A}, \tag{8}$$

where  $A = \sum_{i=1}^k \log \left( \frac{Z_{n,n-k+i}}{Z_{n,n-k}} \right)$  and  $B = 1 + \sum_{i=1}^k \delta_{[n-k+i]}$ .

**Remark 2.1.** In order to simplify notations, the dependency of  $\hat{\alpha}_{MAP}^{(MDI)}$  on  $k$  and  $n$  is not made explicit. This convention will also be adopted for other estimators defined below.

### 2.2. Jeffreys prior

A second kind of non-informative prior is Jeffreys prior [Jeffreys \(1961\)](#), which has an interesting invariant re-parametrization property (see also [Beirlant et al., 2004](#), Chapter 11, for an application of this prior in an extreme value setting). This prior is proportional to the square root of Fisher’s information. If  $\mathcal{L}_u(E, \Delta|\alpha)$  is the likelihood of a single observation  $(E, \Delta)$ , Jeffreys prior can be written as:

$$\begin{aligned} \pi(\alpha) &\propto \left[ -\mathbb{E} \left( \frac{\partial^2}{\partial \alpha^2} \log \mathcal{L}_u(E, \Delta|\alpha) \right) \right]^{\frac{1}{2}} \\ &\propto \frac{1}{\alpha}. \end{aligned} \tag{9}$$

Using (5) and (9), the posterior density of  $\alpha$  based on Jeffreys prior is given by:

$$\pi(\alpha|E, \Delta) \propto \alpha^{\left( \sum_{j=1}^{N_u} \delta_{[n-N_u+j]} \right) - 1} \exp \left[ -\alpha \sum_{j=1}^{N_u} \log e_{j,u} \right],$$

which coincides (up to some normalizing constants) with the probability density function of the distribution

$$\text{Gamma} \left( \sum_{j=1}^{N_u} \delta_{[n-N_u+j]}, \sum_{j=1}^{N_u} \log e_{j,u} \right).$$

Based on this posterior, we construct two classical Bayesian estimators of  $\alpha$ , namely the mean posterior estimator (MPE) and MAP estimator.

- **mean posterior estimator (MPE):** this estimator is defined as the mean of the posterior distribution of  $\alpha$ :

$$\hat{\alpha}_{MPE}^{(J)} := \frac{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]}}{\sum_{j=1}^{N_u} \log e_{j,u}}. \tag{10}$$

Letting  $u = Z_{n,n-k}$  in (10), we obtain the following formal definition of our MPE estimator of the tail index  $\alpha$ :

**Definition 2.2.** Let  $k := k_n$  be a sequence such that  $k \rightarrow \infty$  and  $\frac{k}{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Then the MPE estimator of the tail index  $\alpha$  under Jeffreys prior is given by:

$$\hat{\alpha}_{MPE}^{(J)} := \frac{\sum_{i=1}^k \delta_{[n-i+1]}}{\sum_{i=1}^k \log \left( \frac{Z_{n,n-i+1}}{Z_{n,n-k}} \right)}. \tag{11}$$

We note that  $(\hat{\alpha}_{MPE}^{(J)})^{-1}$  coincides with the estimator of the EVI  $\gamma$  proposed by [Einmahl et al. \(2008\)](#) when censoring occurs.

- **maximum a posteriori estimator (MAP):** this estimator is defined as the mode of the posterior distribution of  $\alpha$ :

$$\hat{\alpha}_{MAP}^{(J)} := \arg \max_{\alpha} \pi(\alpha|E, \Delta) = \frac{\sum_{j=1}^{N_u} \delta_{[n-N_u+j]} - 1}{\sum_{j=1}^{N_u} \log e_{j,u}}. \tag{12}$$

Letting  $u = Z_{n,n-k}$  in (12), we obtain the following formal definition of our MAP estimator of the tail index  $\alpha$  under Jeffreys prior:

**Definition 2.3.** Let  $k := k_n$  be a sequence such that  $k \rightarrow \infty$  and  $\frac{k}{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Then the MAP estimator of the tail index  $\alpha$  under Jeffreys prior is given by:

$$\hat{\alpha}_{MAP}^{(J)} := \frac{\sum_{i=1}^k \delta_{[n-i+1]} - 1}{\sum_{i=1}^k \log \left( \frac{Z_{n,n-i+1}}{Z_{n,n-k}} \right)}. \quad (13)$$

**Remark 2.2.** In Bayesian estimation, conjugate priors are such that the prior and posterior distributions belong to the same family of cdf. Such priors are particularly useful from a computational point of view. It can easily be shown that the Gamma distribution is a conjugate prior for estimating the tail index  $\alpha$ . However, unreported simulations suggest no improvement over Jeffreys and MDI priors. Thus, in what follows, we do not consider this Gamma prior.

### 2.3. Asymptotic normality of the posterior distribution

In this section, we present our main theoretical result, *i.e.* convergence of the posterior distribution of  $\alpha$  to a Gaussian distribution. Proof proceeds along the same lines as proof of Theorem 4.2 in Ghosh et al. (2006). Some additional notations are needed. The open interval  $\mathcal{A} = (a_1, a_2) \subset \mathbb{R}$ , where  $0 < a_1 < a_2 < \infty$ , will denote the parameter space for  $\alpha$ . In what follows, we fix  $\alpha_0$  and  $\beta_0$  and regard them as the true values of  $\alpha$  and  $\beta$  respectively, in the sense that all probability statements will be made under these values. Expectation calculated under  $\alpha_0$  and  $\beta_0$  will be denoted by  $\mathbb{E}$ . We also denote by  $k := k_n$  a sequence such that  $k \rightarrow \infty$  and  $\frac{k}{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Finally,

$$L_k(\alpha) = \log \alpha \sum_{i=1}^k \delta_{[n-k+i]} - \alpha \sum_{i=1}^k \log e_{[n-k+i]} - \sum_{i=1}^k \delta_{[n-k+i]} \log e_{[n-k+i]}$$

denotes the log-likelihood  $\log \mathcal{L}_U(E, \Delta | \alpha)$  calculated with the threshold  $u = Z_{n,n-k}$ . We assume the following regularity conditions:

- (C1) The support of the density of  $(Z, \delta)$  is the same for all  $\alpha, \beta$ .
- (C2)  $L_1(\alpha)$  is thrice differentiable with respect to  $\alpha$  and if  $U = (\alpha_0 - \delta, \alpha_0 + \delta)$  is a neighborhood of  $\alpha_0$ , we have  $\sup_{\alpha \in U} |\partial^3 L_1(\alpha) / \partial \alpha^3| < h(Z, \delta)$ , where  $\mathbb{E}h(Z, \delta) < \infty$ .
- (C3) Let  $\delta > 0$ . There exists  $\epsilon > 0$  such that for all sufficiently large  $k$ ,

$$\sup_{|\alpha - \alpha_0| > \delta} \frac{1}{k} [L_k(\alpha) - L_k(\alpha_0)] < -\epsilon. \quad (14)$$

**Remark 2.3.** It is easily seen that conditions (C1) and (C2) are verified. To see that (C3) holds, consider a Taylor expansion of  $L_k(\alpha)$  at  $\alpha_0$ :

$$L_k(\alpha) = L_k(\alpha_0) + (\alpha - \alpha_0) \frac{\partial L_k(\alpha_0)}{\partial \alpha} + \frac{(\alpha - \alpha_0)^2}{2} \frac{\partial^2 L_k(\tilde{\alpha})}{\partial \alpha^2},$$

where  $\tilde{\alpha}$  belongs to the line segment between  $\alpha$  and  $\alpha_0$ . Then

$$\begin{aligned} \frac{1}{k} [L_k(\alpha) - L_k(\alpha_0)] &= \left( \frac{(\alpha - \alpha_0)}{\alpha_0} \frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]} - (\alpha - \alpha_0) \frac{1}{k} \sum_{i=1}^k \log e_{[n-k+i]} \right) + \left( -\frac{(\alpha - \alpha_0)^2}{2\tilde{\alpha}^2} \frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]} \right) \\ &:= T_1 + T_2. \end{aligned}$$

By Einmahl et al. (2008),  $\hat{p} = \frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]}$  converges in probability to  $p := \frac{\alpha_0}{\alpha_0 + \beta_0}$ . Thus there exists a subsequence of  $k$  along which  $\hat{p}$  converges almost surely (a.s.) to  $p$ . From Deheuvels et al. (1988),  $\frac{1}{k} \sum_{i=1}^k \log \left( \frac{Z_{n,n-k+i}}{Z_{n,n-k}} \right)$  converges a.s. to the EVI  $\frac{1}{\alpha_0 + \beta_0}$  of  $Z$  as  $n \rightarrow \infty$  and thus, also converges to  $\frac{1}{\alpha_0 + \beta_0}$  a.s. along the subsequence. It follows that  $T_1$  converges a.s. to 0 along the aforementioned subsequence and for every  $\epsilon > 0$ , we have that  $T_1 < \epsilon$  for all sufficiently large  $k$ . Next, let  $\xi > 0$  be arbitrarily small. For  $k$  sufficiently large,  $\frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]} > p - \xi$  and one easily shows that for every  $\alpha$  such that  $|\alpha - \alpha_0| > \delta$ , and  $k$  sufficiently large:

$$T_2 < -\frac{\delta^2(p - \xi)}{2a_2^2}.$$

Finally, (14) follows by letting  $\epsilon = \frac{\delta^2(p - \xi)}{4a_2^2}$ .

In what follows, in order to avoid proliferation of notations, we will use the same notation  $k := k_n$  for the subsequence along which a.s. convergence of  $\hat{p}$  holds. We are now in position to state our result:

**Theorem 2.1.** *Let  $\pi(\cdot)$  be a prior density on  $\alpha$ , continuous and positive at  $\alpha_0$ . Let  $\hat{\alpha}_\bullet$  be any of the estimators  $\hat{\alpha}_{MAP}^{(MDI)}$ ,  $\hat{\alpha}_{MPE}^{(I)}$  and  $\hat{\alpha}_{MAP}^{(J)}$ . Then, under assumptions (C1), (C2), (C3), there exists a subsequence  $k := k_n$  of  $n$  such that*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \left| \pi_k(w|E, \Delta) - \sqrt{\frac{p}{2\pi\alpha_0^2}} \exp\left[-\frac{1}{2}w^2 \frac{p}{\alpha_0^2}\right] \right| dw = 0 \quad \text{a.s.,}$$

where  $\pi_k(\cdot|E, \Delta)$  is the posterior density of  $W_k = \sqrt{k}(\alpha - \hat{\alpha}_\bullet)$  given  $(E, \Delta)$ .

Proof is given in [Appendix A](#).

### 3. Simulation study

In this section, we assess, via simulations, finite-sample performance of estimators (8), (11) and (13). We also provide comparisons with a recently proposed estimator of the tail index of a heavy-tailed distribution under random censoring ([Worms and Worms, 2014](#)).

*Study design.* The simulation design is as follows. Let  $X$  and  $Y$  be independent random variables with cdf  $F$  and  $G$  respectively.  $X$  is the variable of interest and  $Y$  is the censoring random variable. We assume that  $F$  and  $G$  are heavy-tailed with tail index  $\alpha > 0$  and  $\beta > 0$  respectively. As  $u \rightarrow \infty$ ,  $p_u := \mathbb{P}(\delta = 1|Z > u)$  tends to  $\frac{\alpha}{\alpha + \beta}$ . Thus, for a given  $\alpha$ , a relevant choice of  $\beta$  allows us to simulate data with an approximate proportion  $p_u$  of non-censored  $X_i$ s among observations with  $Z_i > u$ . Our simulation procedure is as follows:

1. We simulate a sample of  $n$  independent copies of  $(Z, \delta)$ , where  $Z = \min(X, Y)$ ,  $\delta = \mathbb{1}_{\{X \leq Y\}}$  and the target censoring proportion in the right tail of  $X$  is  $1 - p_u$ . For a given  $\alpha$ , letting  $\beta = \frac{\alpha}{9}, \frac{\alpha}{3}, \alpha$  and  $3\alpha$  should allow the censoring percentage in the right tail of  $X$  to be approximately 10%, 25%, 50% and 75% respectively.
2. For each  $\beta$ , we compute the proposed estimates, by incrementing the threshold (or fraction level)  $\frac{k}{n}$  from  $\frac{k_{\min}}{n}$  to  $\frac{k_{\max}}{n}$ . For notational simplicity, we use the following notations:
  - $\hat{\alpha}_1(k)$  is the MPE estimator (11) in [Definition 2.2](#),
  - $\hat{\alpha}_2(k)$  is the MAP estimator (13) in [Definition 2.3](#),
  - $\hat{\alpha}_3(k)$  is the MAP estimator (8) in [Definition 2.1](#),
  - $\hat{\alpha}_4(k)$  is the estimator proposed by [Worms and Worms \(2014\)](#), defined as:

$$\hat{\alpha}_4(k) = \left( \frac{1}{n(1 - \hat{F}_n(Z_{n-k,n}))} \sum_{i=1}^k \frac{1}{1 - \hat{G}_n(Z_{n-i+1}^-)} i \log \left( \frac{Z_{n-i+1,n}}{Z_{n-i,n}} \right) \right)^{-1},$$

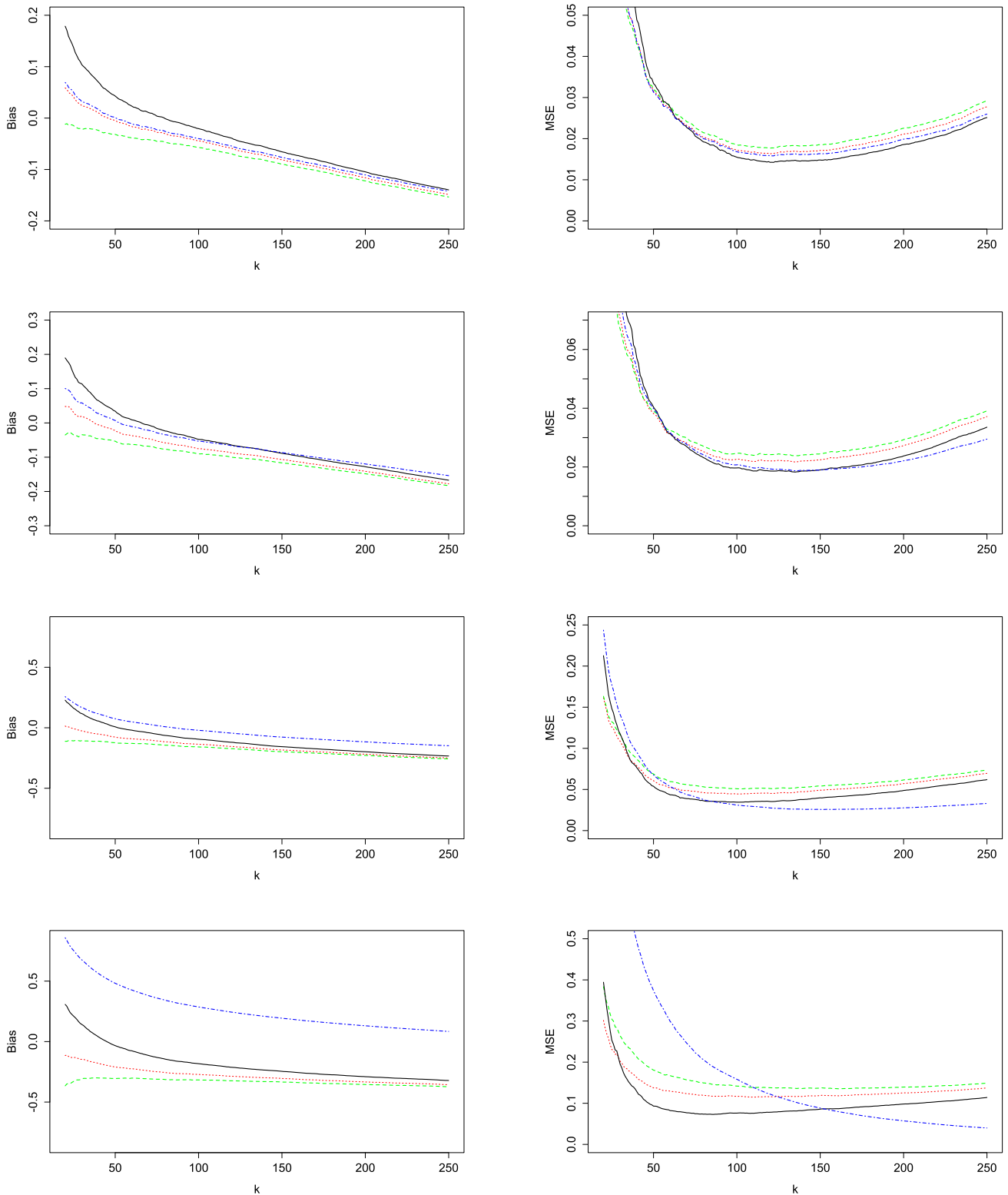
where  $\hat{F}_n$  and  $\hat{G}_n$  denote Kaplan–Meier estimators of  $F$  and  $G$  respectively.

3. Steps 1–2 are repeated  $m$  times, so that we obtain  $m$  realizations of each  $\hat{\alpha}_\ell(k)$ , for each  $\beta$ .
4. For each  $\beta$ , we calculate the average value and MSE (Mean Square Error) of each  $\hat{\alpha}_\ell(k)$  (for  $k = k_{\min}, \dots, k_{\max}$  and  $\ell = 1, 2, 3, 4$ ) over the  $m$  estimates.

We consider the following three simulation settings:

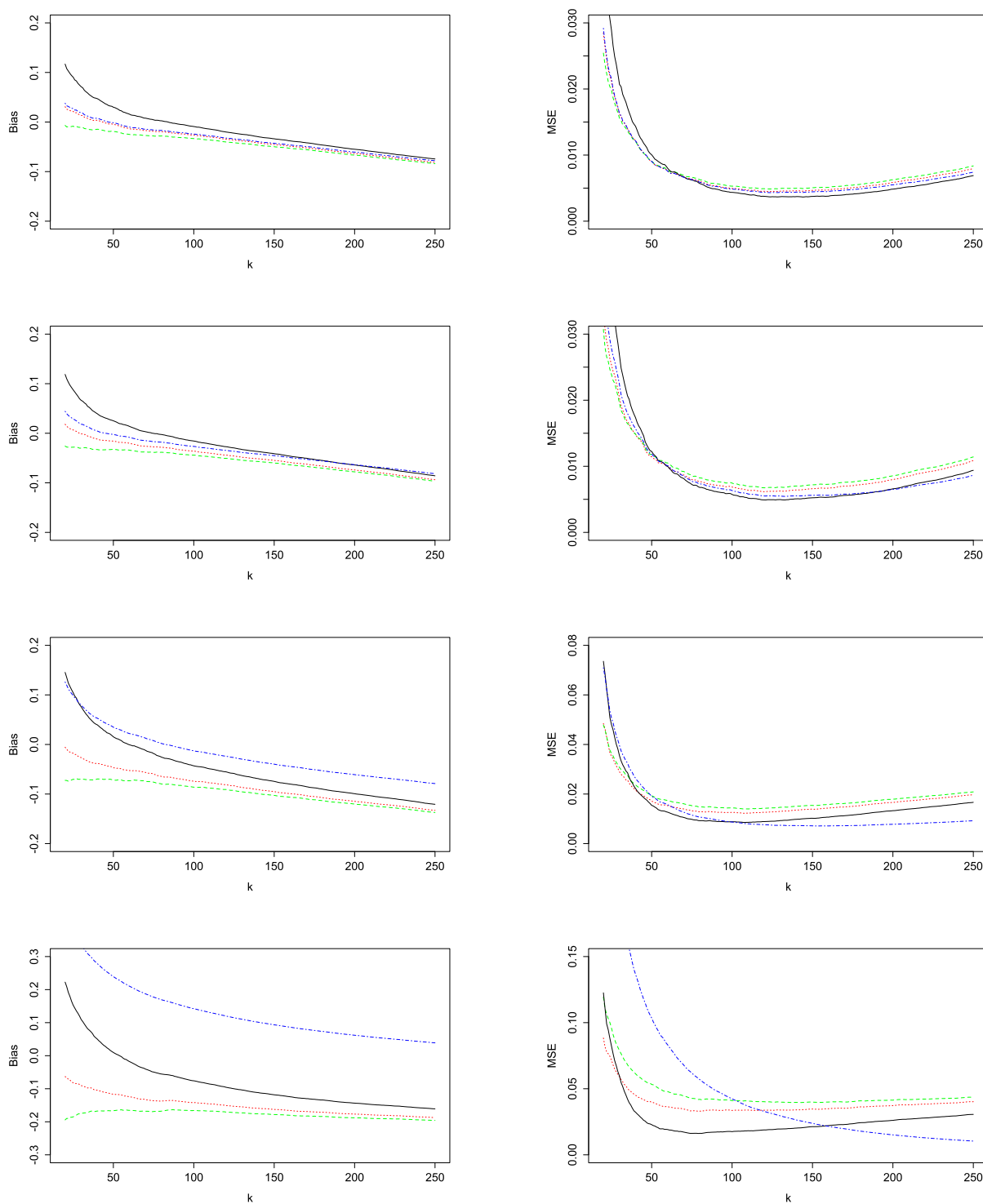
- $X$  and  $Y$  are both distributed as Fréchet random variables, with cdf  $F_\alpha(x) = \exp(-x^{-\alpha})$  and  $G_\beta(x) = \exp(-x^{-\beta})$  respectively,
- $X$  and  $Y$  follow Generalized Pareto distributions (GPD) with cdf  $F_{\mu,\sigma,\alpha}(x) = 1 - (1 + \frac{x-\mu}{\sigma\alpha})^{-\alpha}$  and  $G_{\mu,\sigma,\beta}(x) = 1 - (1 + \frac{x-\mu}{\sigma\beta})^{-\beta}$  respectively,
- $X$  and  $Y$  are both distributed as standard log-logistic random variables with cdf  $F_\alpha(x) = \frac{1}{1+x^{-\alpha}}$  and  $G_\beta(x) = \frac{1}{1+x^{-\beta}}$  respectively (the log-logistic distribution is commonly used in health science to model survival data).

*Results for Fréchet distribution.* We consider the following values for  $\alpha$ :  $\alpha = 1.25$  and  $\alpha = \frac{2}{3}$ . Simulations are conducted using the statistical software R ([R Development Core Team, 2008](#)). Results are provided for a sample of size  $n = 1000$  and  $m = 1000$  simulated samples, with  $k_{\min} = 20$  and  $k_{\max} = 250$ . For each  $\ell = 1, \dots, 4$ , we plot the empirical bias and MSE of  $\hat{\alpha}_\ell$  as functions of  $k$  for  $\beta = \frac{\alpha}{9}, \frac{\alpha}{3}, \alpha$  and  $3\alpha$  ([Fig. 1](#):  $\alpha = 1.25$  and [Fig. 2](#):  $\alpha = \frac{2}{3}$ ). In [Tables 1](#) and [2](#), we report the averaged (over the  $m$  simulated samples) value and empirical MSE of each  $\hat{\alpha}_\ell$  at the optimal fraction level  $k_{opt} = \arg \min_k \text{MSE}[\hat{\alpha}_\ell(k)]$ . For each  $\hat{\alpha}_\ell$ , we also report the averaged proportion  $\bar{p}(k_{opt}) = \frac{1}{k_{opt}} \sum_{i=1}^{k_{opt}} \delta_{[n-i+1]}$  of uncensored data among observations  $i$  such that  $Z_i > Z_{n-k_{opt},n}$  ([Table 1](#):  $\alpha = 1.25$  and [Table 2](#):  $\alpha = \frac{2}{3}$ ). For every  $\beta$ , the average value of  $1 - \bar{p}(k_{opt})$  is close to the target censoring proportion in the right tail of  $X$  (namely, 0.1, 0.25, 0.5, 0.75). From these results, it appears, as expected,



**Fig. 1.** Empirical bias (left) and MSE (right) of  $\hat{\alpha}_1$  (red dotted),  $\hat{\alpha}_2$  (green dashed),  $\hat{\alpha}_3$  (black line),  $\hat{\alpha}_4$  (blue dotdash) for Fréchet distribution with  $\alpha = 1.25$  and  $\beta = \frac{5}{36}$  (top row),  $\beta = \frac{5}{12}$  (second row),  $\beta = \frac{5}{4}$  (third row) and  $\beta = \frac{15}{4}$  (bottom row). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that the bias and MSE of all estimators decrease when censoring in the right tail of  $X$  decreases. The MAP estimator under MDI prior  $\hat{\alpha}_3$  outperforms  $\hat{\alpha}_1$  and  $\hat{\alpha}_2$  in terms of bias and MSE for almost every  $k$  and might be regarded as the best among all three estimators. When  $\alpha \geq \beta$  (weak to moderate censoring in the tail of  $X$ ),  $\hat{\alpha}_3$  performs better than or similarly to Worms and Worms (2014) estimator. When  $\alpha < \beta$  (strong censoring in the tail of  $X$ ), Worms and Worms (2014) estimator seems to perform better (but needs to select a much higher  $k$ ).



**Fig. 2.** Empirical bias (left) and MSE (right) of  $\hat{\alpha}_1$  (red dotted),  $\hat{\alpha}_2$  (green dashed),  $\hat{\alpha}_3$  (black line),  $\hat{\alpha}_4$  (blue dotdash) for Fréchet distribution with  $\alpha = \frac{2}{3}$  and  $\beta = \frac{2}{27}$  (top row),  $\beta = \frac{2}{9}$  (second row),  $\beta = \frac{2}{3}$  (third row) and  $\beta = 2$  (bottom row). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

*Results for the GPD and log–logistic distribution.* Results for the GPD and log–logistic distribution are given in [Appendix B](#). Conclusions are similar to conclusions obtained for Fréchet distribution.

#### 4. Randomizing the fraction level

In the previous section, the  $(n - k)$ th order statistic  $Z_{n,n-k}$  was used as threshold in the excesses  $e_{j,u}$  and  $k$  (or equivalently, the fraction level  $k/n$ ) was chosen so as to minimize the MSE. Obviously, this procedure cannot be applied in real data analysis

**Table 1**

Optimal results for Fréchet model with  $\alpha = 1.25$ . For each  $\beta$ , the estimator with smallest empirical bias and MSE is indicated in bold.

		$k_{opt}$	$\bar{p}(k_{opt})$	Estimator	MSE
$\beta = \frac{5}{36}$	$\hat{\alpha}_1$	136	0.9268	1.1796	0.0155
	$\hat{\alpha}_2$	136	0.9268	1.1703	0.0168
	$\hat{\alpha}_3$	117	0.9263	<b>1.1968</b>	<b>0.0135</b>
	$\hat{\alpha}_4$	136	0.9268	1.1841	0.0148
$\beta = \frac{5}{12}$	$\hat{\alpha}_1$	135	0.7843	1.1545	0.0209
	$\hat{\alpha}_2$	135	0.7843	1.1436	0.0230
	$\hat{\alpha}_3$	109	0.7835	<b>1.1774</b>	<b>0.0176</b>
	$\hat{\alpha}_4$	132	0.7841	1.1746	0.0184
$\beta = \frac{5}{4}$	$\hat{\alpha}_1$	105	0.5013	1.1105	0.0429
	$\hat{\alpha}_2$	105	0.5013	1.0893	0.0488
	$\hat{\alpha}_3$	96	0.5005	1.1601	0.0305
	$\hat{\alpha}_4$	157	0.4992	<b>1.1627</b>	<b>0.0295</b>
$\beta = \frac{15}{4}$	$\hat{\alpha}_1$	108	0.2157	0.9734	0.1159
	$\hat{\alpha}_2$	108	0.2157	0.9214	0.1374
	$\hat{\alpha}_3$	77	0.2170	1.1142	0.0717
	$\hat{\alpha}_4$	250	0.2174	<b>1.3339</b>	<b>0.0399</b>

**Table 2**

Optimal results for Fréchet model with  $\alpha = \frac{2}{3}$ . For each  $\beta$ , the estimator with smallest empirical bias and MSE is indicated in bold.

		$k_{opt}$	$\bar{p}(k_{opt})$	Estimator	MSE
$\beta = \frac{2}{27}$	$\hat{\alpha}_1$	127	0.9270	0.6295	0.0044
	$\hat{\alpha}_2$	127	0.9270	0.6242	0.0048
	$\hat{\alpha}_3$	118	0.9265	<b>0.6433</b>	<b>0.0035</b>
	$\hat{\alpha}_4$	127	0.9270	0.6319	0.0045
$\beta = \frac{2}{9}$	$\hat{\alpha}_1$	121	0.7843	0.6197	0.0059
	$\hat{\alpha}_2$	121	0.7843	0.6135	0.0065
	$\hat{\alpha}_3$	121	0.7843	<b>0.6358</b>	<b>0.0048</b>
	$\hat{\alpha}_4$	144	0.7848	0.6222	0.0053
$\beta = \frac{2}{3}$	$\hat{\alpha}_1$	105	0.4988	0.5916	0.0118
	$\hat{\alpha}_2$	105	0.4988	0.5809	0.0135
	$\hat{\alpha}_3$	87	0.4993	0.6212	0.0082
	$\hat{\alpha}_4$	144	0.4990	<b>0.6222</b>	<b>0.0073</b>
$\beta = 2$	$\hat{\alpha}_1$	113	0.2178	0.5202	0.0326
	$\hat{\alpha}_2$	113	0.2178	0.4990	0.0390
	$\hat{\alpha}_3$	82	0.2188	0.5789	0.0180
	$\hat{\alpha}_4$	250	0.2186	<b>0.7055</b>	<b>0.0151</b>

since the true tail index  $\alpha$  used to calculate the MSE is unknown. Thus, in this section, we develop an alternative approach for tackling this threshold issue.

We propose to consider the fraction level  $t := k/n$  as a random quantity. As  $k/n$  lies in  $]0, 1[$ , we suggest to use a Beta distribution for  $t$  ( $t \sim \mathcal{Beta}(c, d)$ ), with probability density function:

$$\pi_u(t) = \frac{1}{B(c, d)} \frac{(t - t_{\min})^{c-1} (t_{\max} - t)^{d-1}}{(t_{\max} - t_{\min})^{c+d-1}} \mathbb{1}_{\{t_{\min} \leq t \leq t_{\max}\}},$$

where  $B(c, d) := \int_0^1 s^{c-1} (1-s)^{d-1} ds$  ( $c > 0, d > 0$ ) is the Beta function and  $t_{\min}$  and  $t_{\max}$  are suitable bounds for the fraction level. Then we propose the following Monte Carlo approach for inferring on  $\alpha$  from a sample of observations  $(Z_i, \delta_i)_{i=1, \dots, n}$ .

First, we simulate  $x \sim \mathcal{Beta}(c, d)$  and we calculate  $t_x = (t_{\max} - t_{\min})x + t_{\min}$ . Then we obtain  $k_x = [nt_x]$  (where  $[\cdot]$  denotes the integer part) and we calculate the estimate  $\hat{\alpha}_\ell$  with  $k = k_x$ . This procedure is repeated  $N$  times, which yields  $N$  Monte Carlo realizations  $\hat{\alpha}_\ell^{(1)}, \dots, \hat{\alpha}_\ell^{(N)}$  of  $\hat{\alpha}_\ell$ . Finally,  $\alpha$  can be estimated from these  $N$  realizations (by taking their empirical mean, median or mode, for example).

We illustrate this procedure in a short simulation study. Samples of size  $n = 1000$  of  $(Z, \delta)$  are obtained by simulating  $X$  and  $Y$  from Pareto distributions with tail index  $\alpha = 2$  and  $\beta = \frac{\alpha}{9}, \frac{\alpha}{3}, \alpha, 3\alpha$ , so that the censoring percentage in the right tail of  $X$  is approximately 10%, 25%, 50% and 75% respectively. We take  $t_{\min} = 0.01$  and  $t_{\max} = 0.4$ . Under these values,  $k$  is allowed to range from 10 to 400, with a higher concentration of values around  $k = 300$ . We also take  $(c, d) = (4, 1.5)$  (a procedure for choosing  $c$  and  $d$  will be discussed in Section 5). Finally, we let  $N = 10\,000$ .

For each  $\beta = \frac{\alpha}{9}, \frac{\alpha}{3}, \alpha, 3\alpha$  and  $l \in \{1, 2, 3\}$ , we obtain the empirical mean, median and mode and RMSE of the  $N$  Monte Carlo estimates  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  (see Table 3) and we plot their histogram and estimated density (see Fig. 3). In Table 3, we

**Table 3**  
Randomization of the fraction level: empirical mode, median and mean and RMSE of the Monte Carlo estimates  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  of the tail index  $\alpha = 2$ .

		$(c, d) = (4, 1.5)$			$(c, d) = (1, 1)$		
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
$\beta = 6$	Mode	2.037	<b>1.983</b>	2.048	2.107	2.061	2.145
	Median	2.074	<b>2.015</b>	2.085	2.120	2.083	2.169
	Mean	2.052	<b>2.014</b>	2.096	2.083	2.035	2.161
	RMSE	0.074	<b>0.072</b>	0.082	0.160	0.190	0.124
$\beta = 2$	Mode	2.014	<b>1.992</b>	2.027	2.073	2.070	2.109
	Median	2.024	<b>2.008</b>	2.034	2.069	2.051	2.103
	Mean	<b>2.002</b>	1.988	2.027	2.089	2.064	2.120
	RMSE	0.061	<b>0.059</b>	0.065	0.076	0.083	0.092
$\beta = \frac{2}{3}$	Mode	<b>1.985</b>	1.975	2.033	2.034	2.021	2.088
	Median	2.026	<b>2.015</b>	2.044	2.026	2.027	2.047
	Mean	2.038	2.027	2.054	2.021	<b>2.013</b>	2.045
	RMSE	0.043	<b>0.042</b>	0.045	0.065	0.069	0.068
$\beta = \frac{2}{9}$	Mode	1.973	1.938	<b>1.997</b>	2.043	2.041	2.048
	Median	2.044	<b>1.995</b>	2.051	2.066	2.044	2.053
	Mean	2.007	<b>1.998</b>	2.027	2.067	2.033	2.047
	RMSE	0.035	<b>0.034</b>	0.036	0.056	0.051	0.061

also provide results for the case  $(c, d) = (1, 1)$ , where the Beta distribution reduces to the uniform distribution on  $[0, 1]$ . In this case, no prior information is available on the fraction level.

From Table 3, the empirical mean, median and mode of the  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  provide satisfactory approximations of  $\alpha$  for every  $\beta$  and  $\ell = 1, 2, 3$ . Best results (in terms of accuracy and RMSE) are almost always obtained under a Beta distribution (for each  $\beta$ , the most accurate values of the empirical mean, median and mode and the smallest RMSE are indicated in bold). From Fig. 3, histograms are rather symmetric and concentrated around the true tail index value  $\alpha = 2$  (represented by the vertical dotted line) in every case. In that sense, the empirical mean and median of the  $\hat{\alpha}_\ell^{(j)}$  might be used to estimate  $\alpha$ . Overall, combining the proposed Bayesian estimators with the randomization tool described above seems to provide a relevant approach for estimating  $\alpha$ . This approach is now illustrated on real data.

### 5. An application to AIDS survival data

In this section, we illustrate our methodology on a set of AIDS survival data. The dataset contains  $n = 2843$  patients diagnosed with AIDS in Australia before 1 July 1991. The source of the data is Dr P.J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research (Venables and Ripley, 2002). The data are available in the R package “MASS” (R Development Core Team, 2008). Information on each patient includes gender, date of diagnosis, age at diagnosis, date of death or end of observation and an indicator which equals 1 if the patient died and 0 otherwise. 1761 patients died. The other survival times are right-censored.

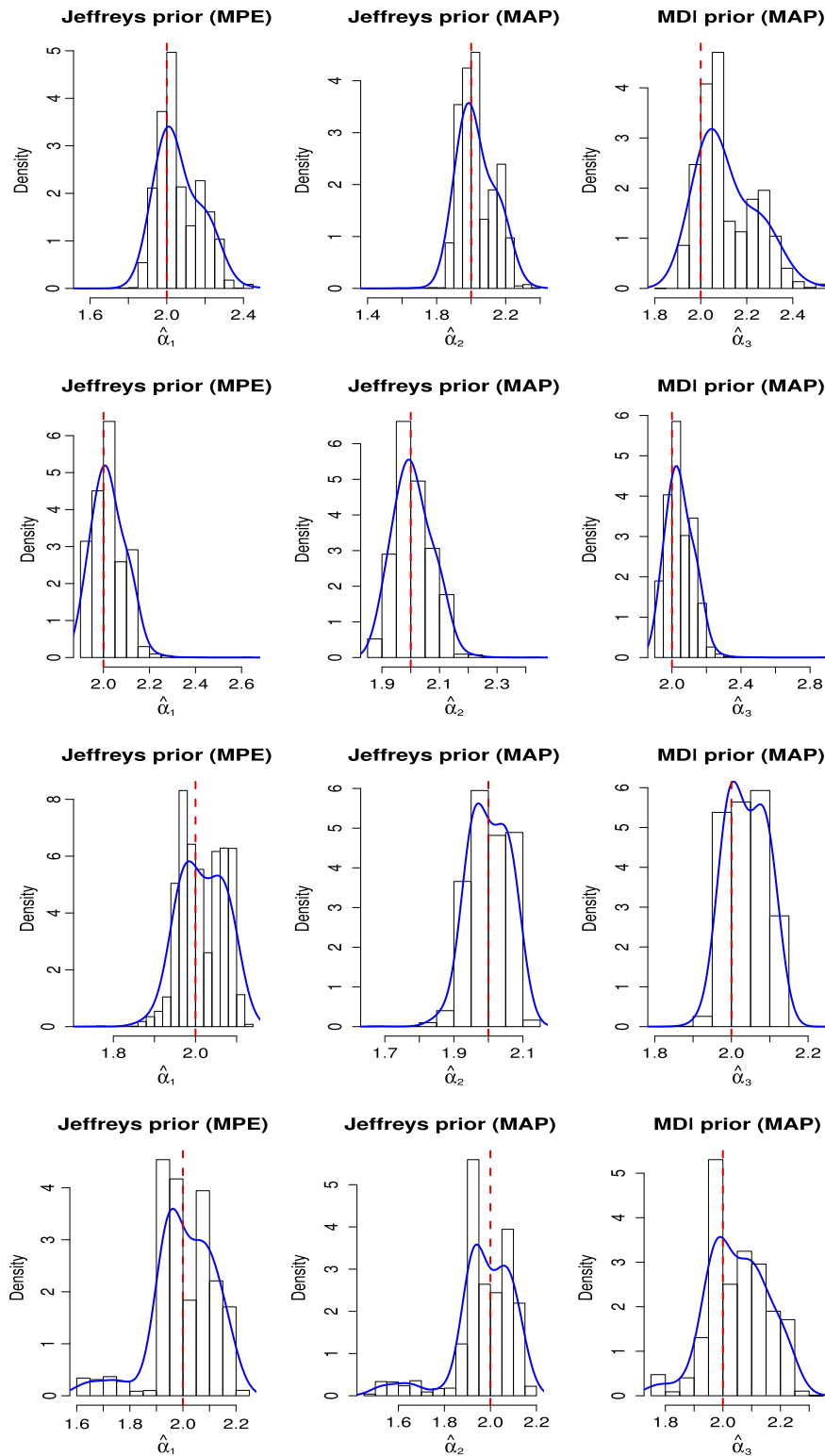
#### 5.1. Investigation of the heavy-tail assumption

First, we investigate the hypothesis that these data arise from a heavy-tailed distribution. The model described in Section 2 assumes that the cdf of both survival and censoring times are heavy-tailed, which implies that the observed time  $Z$  is also heavy-tailed. In order to check these assumptions on the AIDS survival data, we apply Kolmogorov–Smirnov (KS) and Berk–Jones (BJ) goodness-of-fit tests for heavy-tailed distributions (described, for example, by Koning and Peng, 2008). KS and BJ test statistics are defined as:

$$KS(k, \hat{\alpha}) = \sup_{r>1} \sqrt{k} |1 - G_k(r) - r^{\hat{\alpha}}| \quad \text{and} \quad BJ(k, \hat{\alpha}) = \sup_{r>1} k2K(G_k(r), 1 - r^{\hat{\alpha}}),$$

where  $1 - G_k(r) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{Z_i > rZ_{n-k,n}\}}$  and  $K(x, y) = x \ln \frac{x}{y} + (1 - x) \ln \frac{(1-x)}{(1-y)}$ , with  $0 \ln(0/p) = 0$ . We calculate KS and BJ statistics (with  $k$  ranging from 20 to 600 with increment 1) on the subset of uncensored (respectively censored) individuals and on the whole dataset. Results are provided in Figs. 4, 5, 6 respectively. From Koning and Peng (2008), the limiting critical value with level 0.95 for KS test is 1.076. This value is represented as a red horizontal line on the top graphs of Figs. 4–6.

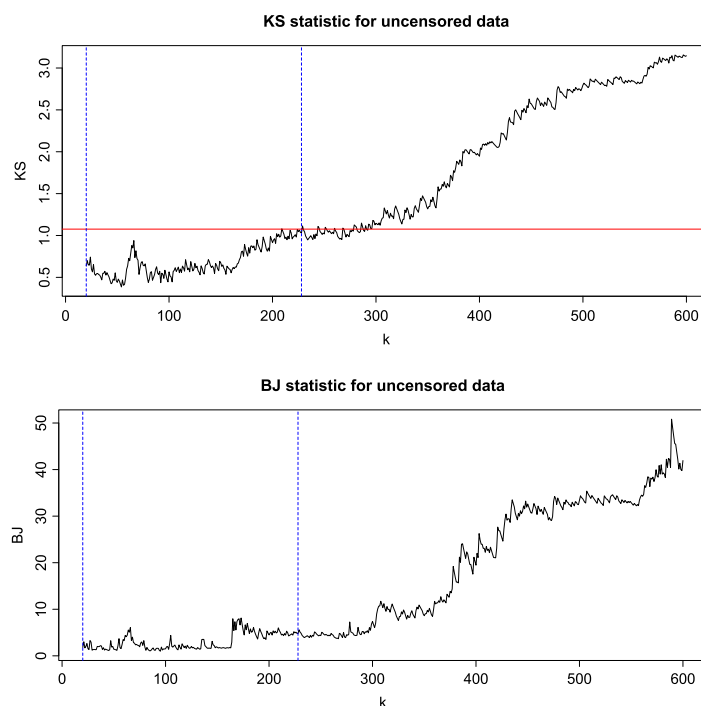
From Fig. 4, the hypothesis of a heavy-tailed distribution for  $X$  is accepted by KS test when  $k$  ranges over the integer interval  $\{20, \dots, 218\}$ . We will see in Section 5.2 that this interval includes the range  $\{70, \dots, 170\}$  of values of  $k$  which should be retained to estimate the tail index and extreme quantiles (this range will be called “stable phase” thereafter). The maximum of BJ statistic for  $k$  ranging from 20 to 218 is equal to 8.130, with mean value approximately equal to 3.219 (and standard deviation equal to 1.786).



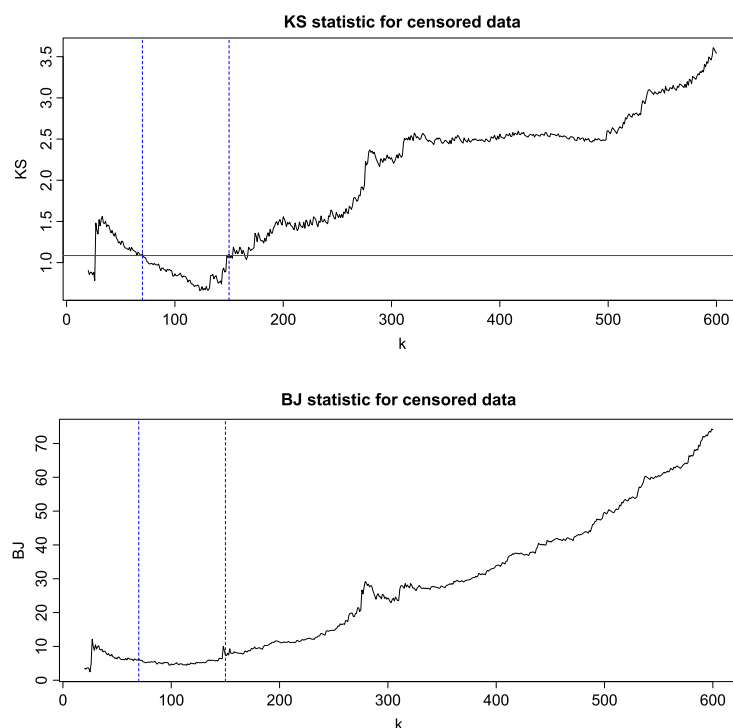
**Fig. 3.** Randomization of the fraction level: histograms of the Monte Carlo estimates  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  of the tail index  $\alpha = 2$  for  $\beta = 6$  (top row),  $\beta = 2$  (second row),  $\beta = \frac{2}{3}$  (third row),  $\beta = \frac{2}{9}$  (bottom row).

From Fig. 5, the hypothesis of a heavy-tailed distribution for the censoring variable  $Y$  is accepted by KS test for  $k$  ranging from 70 to 150, which coincides approximately with the aforementioned stable phase. The maximum of BJ statistic in this phase is 9.991 with mean around 6.163 (and standard deviation equal to 1.544).

From Fig. 6, the hypothesis of a heavy-tailed distribution for  $Z$  is accepted by KS test for  $k$  ranging from 74 to 250, which includes the stable phase mentioned above. The maximum of BJ statistic for  $k$  ranging from 74 to 200 (say, which is closer



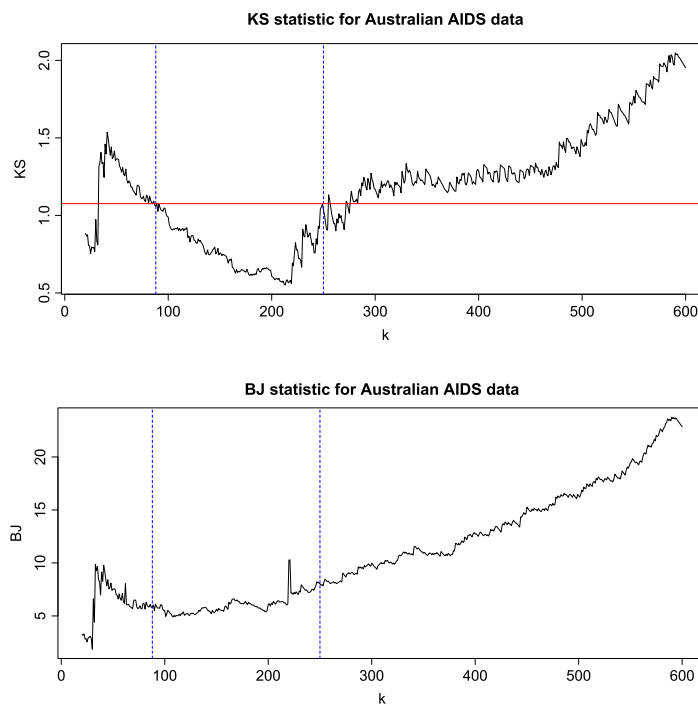
**Fig. 4.** KS and BJ test statistics (Australian AIDS uncensored data). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** KS and BJ test statistics (Australian AIDS censored data). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to the right-endpoint of the stable phase than  $k = 250$ ) is 6.609 with mean around 5.740 (and standard deviation equal to 0.467).

All these results support the assumption that the survival time  $X$ , censoring variable  $Y$  and observed variable  $Z = \min(X, Y)$  are heavy-tailed. Indeed, the hypothesis of a heavy-tailed distribution is accepted by KS test when  $k$  varies within the stable phase identified in Section 5.2 (this stable phase contains the set of  $k$ -values which should be retained for estimating the tail index and extreme quantiles. Therefore, this range of  $k$ -values might also be considered as the relevant one for performing tests of a heavy-tailed distribution). Moreover, the maximum and mean values of BJ statistic over the



**Fig. 6.** KS and BJ test statistics (Australian AIDS data, whole dataset). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

stable phase stay below or close to the 0.95-level critical value for the BJ test provided by [Koning and Peng \(2008\)](#). Hence, both KS and BJ tests seem to agree that the heavy-tailed assumption holds for the AIDS survival dataset.

5.2. Tail index and extreme quantile estimation

[Einmahl et al. \(2008\)](#) analyzed these AIDS survival data without distinction of sex or age. Our objective here is to estimate the tail index  $\alpha$  of  $F$  by using Bayes estimates, but since age and sex are important prognostic factors, we take them into account by considering sub-populations of the whole data set. For illustrative purpose, we consider three sub-populations: male individuals (without distinction of age), male individuals aged 30–35 and male individuals aged 35–40. These sub-populations are denoted by  $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$  respectively.

We calculate the MPE and MAP estimators under Jeffreys prior ( $\hat{\alpha}_1$  and  $\hat{\alpha}_2$  respectively) and the MAP estimator under MDI prior ( $\hat{\alpha}_3$ ), on  $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ . To choose the fraction level  $k/n$ , we first rely on a method proposed by [Einmahl et al. \(2008\)](#). Then we use the randomizing approach described in Section 4. [Einmahl et al. \(2008\)](#) propose to plot the proportion  $\hat{p}(k) = \frac{1}{k} \sum_{i=1}^n \delta_i \mathbb{1}_{\{Z_i > Z_{n,n-k}\}}$  of non-censored observations in the  $k$  largest  $Z_i$ 's as a function of  $k$ .

Consider the first row of [Fig. 7](#). Three distinct phases can be distinguished on the left graph (graph of  $k \mapsto \hat{p}(k)$ ). In phase (A), the behavior of  $k \mapsto \hat{p}(k)$  is somewhat erratic. Then, there is a stable part (phase (B), where  $k$  ranges approximately from 70 to 170 and the average  $\hat{p}(k)$  is approximately 0.267) and an increasing part (phase (C)). [Einmahl et al. \(2008\)](#) suggest to estimate  $\alpha$  based on the  $k$ -values of phase (B). Indeed, the estimate of  $\alpha$  should be quite stable within this range of  $k$ . On the second graph (first row, right column), we plot the estimates  $\hat{\alpha}_1(k), \hat{\alpha}_2(k)$  and  $\hat{\alpha}_3(k)$  as functions of  $k$ .

We clearly distinguish the same three phases (A), (B) and (C). All three estimates appear to be close to each other and quite stable in phase (B). The average value of  $\hat{\alpha}_\ell(k)$  on phase (B) is approximately 1.27 (for  $\ell = 1, 2, 3$ ). Based on [Einmahl et al. \(2008\)](#) empirical methodology, one may thus retain 1.27 as an estimate of  $\alpha$  in the sub-population  $\mathcal{S}_1$ . Using the same procedure, one may estimate  $\alpha$  by 1.45 (respectively 1.32) in the sub-population  $\mathcal{S}_2$  (respectively  $\mathcal{S}_3$ ).

In extreme value analysis, tail index estimation is often only a first step and the final objective is to estimate extreme quantiles  $Q_\varepsilon$  of upper order  $\varepsilon$  ( $0 < \varepsilon < 1$ ) of  $F$ . In the right-censored framework, [Beirlant et al. \(2007\)](#) proposed the following Weissman-type estimator of extreme quantiles:

$$\hat{Q}_\varepsilon(k) = Z_{n,n-k} \left( \frac{1 - \hat{F}_n(Z_{n,n-k})}{\varepsilon} \right)^{\frac{1}{\hat{\alpha}}}, \tag{15}$$

where  $1 - \hat{F}_n(Z_{n,n-k}) = \prod_{i=1}^n \left[ 1 - \frac{\delta_i \mathbb{1}_{\{Z_i < Z_{n,n-k}\}}}{n-i+1} \right]$  is Kaplan–Meier estimator of  $1 - F$  (see [Andersen et al. \(1993\)](#)).

[Fig. 8](#) represents the functions  $k \mapsto \hat{Q}_{0.001}(k)$  obtained in each sub-population  $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$  by using the corresponding estimate of  $\alpha$  (i.e., 1.27 in  $\mathcal{S}_1$ , 1.45 in  $\mathcal{S}_2$ , 1.32 in  $\mathcal{S}_3$ ). We propose to estimate the extreme quantile  $Q_{0.001}$  of the survival time

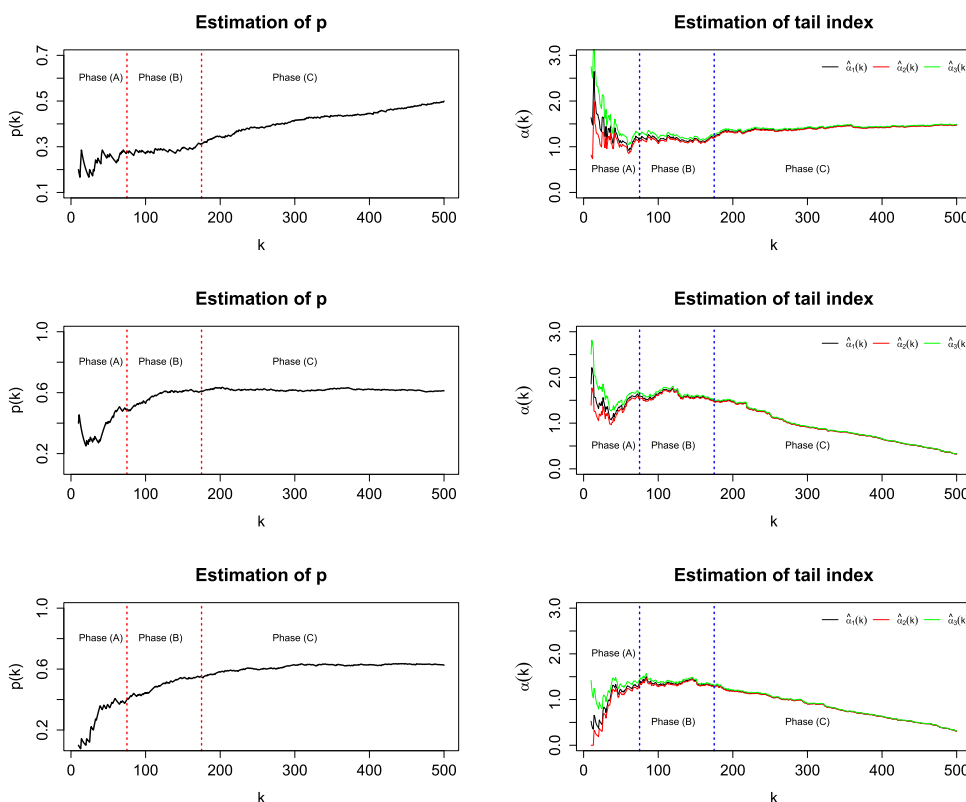


Fig. 7. Proportion of uncensored  $Z_i$ 's among the  $k$  largest observations and tail index estimation for Australian AIDS survival data ( $\delta_1$ : first row,  $\delta_2$ : second row,  $\delta_3$ : third row).

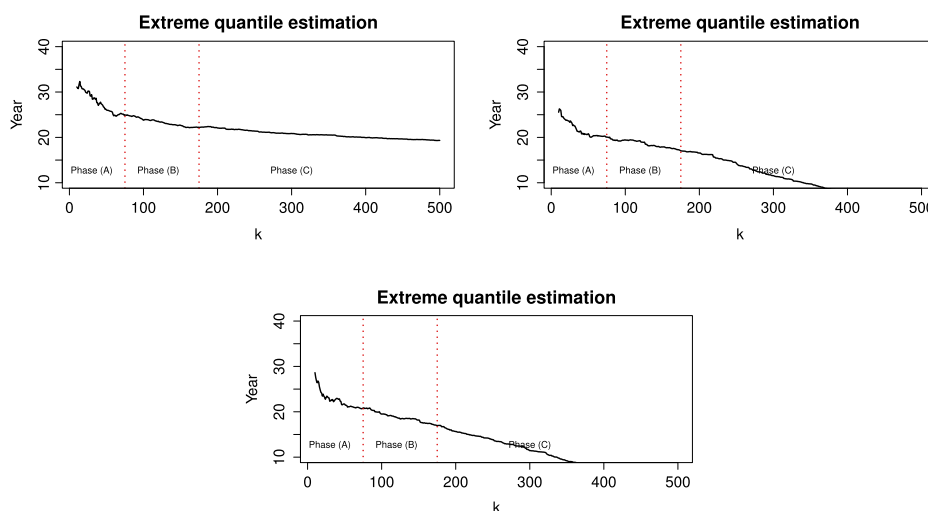


Fig. 8. Extreme quantile estimation for Australian AIDS survival data, with  $\varepsilon = 0.001$  (first row: sub-populations  $\delta_1$  (left) and  $\delta_2$  (right), second row: sub-population  $\delta_3$ ).

distribution by the averaged value of  $\hat{Q}_\varepsilon(k)$ , where  $k$  ranges within phase (B). We find that 0.1% of males (without distinction of age) can exceptionally survive AIDS for 22.68 years, 0.1% of males aged 30–35 can survive AIDS for 17.31 years and 0.1% of males aged 35–40 years can survive AIDS for 16.70 years.

We now apply our randomization procedure (for illustrative purpose, we consider the sub-population  $\delta_1$ ). We take  $t_{\min} = 0.01$  and  $t_{\max} = 0.4$ , so that  $k$  ranges from 27 to 1100 (this range includes and goes far beyond the stable phase (B)). We consider several values for the parameters  $c, d > 0$  of the Beta distribution used to randomize the fraction level, namely  $(c, d) \in \mathcal{S} := \{1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7\} \times \{1, 1.5, 2, 2.5, 3, 3.5, 4\}$ . For every  $(c, d) \in \mathcal{S}$ , we apply the proposed randomization procedure with  $N = 10\,000$  and we calculate the standard deviation (Std.) of the resulting  $N$  Monte Carlo estimates  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$ . Then, we retain the value  $(c^*, d^*)$  such that the standard deviation is minimum. Finally, we propose to infer on  $\alpha$  from the  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  obtained with  $(c, d) = (c^*, d^*)$ .

**Table 4**

Standard deviations of the Monte Carlo estimates  $\hat{\alpha}_1^{(j)}, j = 1, \dots, N$  for various values of  $(c, d)$ . The minimum standard deviations are indicated in bold.

		<i>d</i>						
		1.0	1.5	2.0	2.5	3.0	3.5	4.0
<i>c</i>	1.0	0.375	0.415	0.448	0.478	0.498	0.501	0.514
	1.5	0.259	0.302	0.343	0.378	0.404	0.425	0.451
	2.0	0.194	0.224	0.240	0.268	0.298	0.337	0.368
	2.5	0.163	0.182	0.195	0.207	0.229	0.254	0.272
	3.0	0.142	0.146	0.152	0.168	0.175	0.191	0.220
	3.5	0.131	0.136	0.141	0.128	0.134	0.141	0.159
	4.0	0.122	0.129	0.130	0.123	0.119	0.115	0.127
	5.0	0.119	0.121	0.118	0.115	0.112	0.109	0.109
	6.0	0.114	0.115	0.110	0.109	<b>0.107</b>	<b>0.104</b>	0.110
7.0	0.117	0.116	0.118	0.110	0.110	<b>0.105</b>	0.111	

**Table 5**

Summary statistics for the Monte Carlo estimates  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  obtained with  $(c, d) = (6, 3)$  (see (\*)),  $(c, d) = (6, 3.5)$  (see (\*\*)) and  $(c, d) = (7, 3.5)$  (see (\*\*\*)).

		Std.	Min.	1st Qu.	Median	Mean	Mode	3rd Qu.	Max.
(*)	$\hat{\alpha}_1$	0.107	1.137	1.314	1.417	1.415	1.418	1.516	2.033
	$\hat{\alpha}_2$	0.106	1.136	1.310	1.414	1.411	1.420	1.511	2.010
	$\hat{\alpha}_3$	0.108	1.141	1.321	1.422	1.421	1.420	1.523	2.068
(**)	$\hat{\alpha}_1$	0.104	1.127	1.342	1.436	1.431	1.508	1.528	2.464
	$\hat{\alpha}_2$	0.103	1.125	1.336	1.433	1.427	1.507	1.524	2.434
	$\hat{\alpha}_3$	0.104	1.130	1.349	1.441	1.438	1.513	1.535	2.506
(***)	$\hat{\alpha}_1$	0.105	1.143	1.323	1.419	1.420	1.428	1.518	1.775
	$\hat{\alpha}_2$	0.105	1.144	1.320	1.417	1.418	1.425	1.515	1.728
	$\hat{\alpha}_3$	0.106	1.150	1.329	1.425	1.427	1.432	1.526	1.777

Table 4 reports standard deviations for  $\hat{\alpha}_1$ . The minimum is achieved when  $(c, d) = (6, 3.5)$  (Std. = 0.104). Standard deviations for the pairs  $(6, 3)$  (Std. = 0.107) and  $(7, 3.5)$  (Std. = 0.105) are close to the minimum, thus, we also retain these values for making inference on  $\alpha$ . The minimum standard deviations for  $\hat{\alpha}_2$  and  $\hat{\alpha}_3$  are achieved for the same values of  $(c, d)$ . The corresponding tables are omitted for conciseness.

We can now infer on  $\alpha$  from the  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  obtained with  $(c^*, d^*)$  (by taking their empirical mean, median or mode, for example). We report some summary statistics for  $\hat{\alpha}_1, \hat{\alpha}_2$  and  $\hat{\alpha}_3$  when  $(c^*, d^*) = (6, 3), (6, 3.5), (7, 3.5)$  (see Table 5). We also plot the corresponding histograms of the  $\hat{\alpha}_\ell^{(j)}, j = 1, \dots, N$  (see Fig. 9).

From Fig. 9, histograms are quite symmetric and show limited spread. Thus, it seems reasonable to estimate  $\alpha$  by any of the usual measures of central tendency. Moreover, from Table 5, the empirical means and medians of the  $\hat{\alpha}_\ell (\ell = 1, 2, 3)$  are close to each other, for every  $(c, d) = (6, 3), (6, 3.5), (7, 3.5)$ . From these results, one may retain 1.4 as a reasonable and consensual (over the  $\hat{\alpha}_\ell, \ell = 1, 2, 3$ ) approximation of  $\alpha$ . The same procedure can be applied to sub-populations  $\mathcal{S}_2$  and  $\mathcal{S}_3$ .

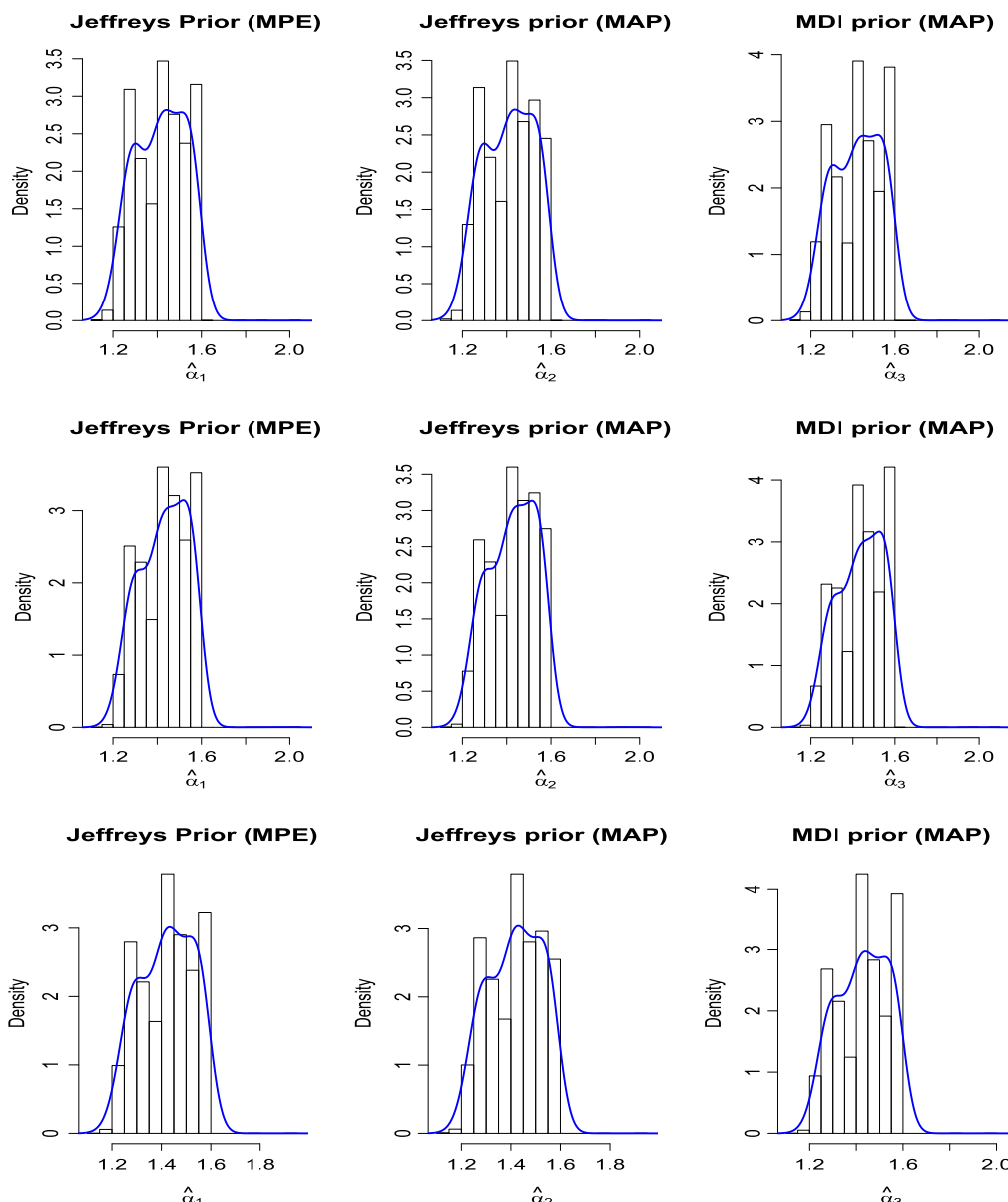
Letting  $\hat{\alpha} = 1.4$  in (15), we obtain the estimated extreme quantiles  $k \mapsto \hat{Q}_{0.001}(k)$  (see Fig. 10). Again, we estimate  $Q_{0.001}$  by the averaged value of the  $\hat{Q}_\varepsilon(k)$ , where  $k$  ranges within phase (B). We conclude that 0.1% of male can exceptionally survive AIDS for 19.71 years.

## 6. Conclusion and perspectives

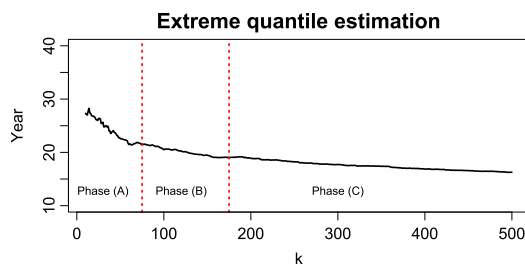
In this paper, we address the estimation of the tail index of a heavy-tailed distribution when data are randomly right-censored. We constructed two kinds of Bayesian estimators, namely mean posterior and maximum a posteriori estimators, for various prior distributions. Our simulation results indicate good performance of these estimators. We also proposed an original procedure for selecting the threshold (or fraction level  $k/n$ ) required to estimate the tail index. The whole procedure was applied to a set of real data and provided coherent results.

Now, several issues still deserve attention. First, we mentioned that estimation of the tail index constitutes usually a first step in an extreme value analysis. Extreme quantile estimation often comes as a second step. In the real-data example, we incorporated our Bayesian estimators of the tail index in a Weissman-type estimator of extreme quantiles. A more detailed investigation of the performances of this new estimator is needed. This is a subject for our future research.

In practice, it often arises that some covariate information  $W$  is available to the investigator and the distribution of  $X$  depends on  $W$ . In this case, interest turns to estimation of the conditional tail index  $\alpha(w)$ . This is a difficult problem which we overcame, in our real-data analysis, by considering sub-populations defined by gender and age categories. However, a more advanced approach should be developed to tackle this issue. This also constitutes a topic for our future research.



**Fig. 9.** Randomization of the fraction level: histograms of the  $\hat{\alpha}_\ell$  ( $\ell = 1, 2, 3$ ) with  $(c^*, d^*) = (6, 3)$  (top row),  $(c^*, d^*) = (6, 3.5)$  (second row),  $(c^*, d^*) = (7, 3.5)$  (bottom row).



**Fig. 10.** Extreme quantile estimation for Australian AIDS survival data, with  $\varepsilon = 0.001$  (sub-population  $\mathcal{S}_1$ ).

**Acknowledgments**

Authors thank the Associate Editor and anonymous reviewers for their valuable comments and suggestions that helped improve successive versions of the paper.

Authors acknowledge financial support from the CNEPRU project “Modélisation - Simulation Actuarielle de Risques (MSAR)” (project number B00220120012, funded by Agence Nationale de Développement de la Recherche Universitaire, Algeria) and from the project SESA (project number EDC25895, funded by Centre National de la Recherche

Scientifique – CNRS, France – and Direction de la Post-Graduation et de la Recherche-Formation – DPGRF, Algeria). Part of this work was made while A. Ameraoui was visiting INSA of Rennes and while J.-F. Dupuy was visiting the Université des Sciences et de la Technologie Houari Boumediene, Algiers.

**Appendix A. Proof of Theorem 2.1**

The posterior density of  $W_k$  can be written as

$$\pi_k(w|E, \Delta) := C_k^{-1} \pi(\hat{\alpha}_\bullet + k^{-1/2}w) \exp [L_k(\hat{\alpha}_\bullet + k^{-1/2}w) - L_k(\hat{\alpha}_\bullet)],$$

where  $C_k = \int_{\mathbb{R}} \pi(\hat{\alpha}_\bullet + k^{-1/2}w) \exp [L_k(\hat{\alpha}_\bullet + k^{-1/2}w) - L_k(\hat{\alpha}_\bullet)] dw$ . Let

$$g_k(w) = \pi(\hat{\alpha}_\bullet + k^{-1/2}w) \exp [L_k(\hat{\alpha}_\bullet + k^{-1/2}w) - L_k(\hat{\alpha}_\bullet)] - \pi(\alpha_0) \exp \left[ -\frac{1}{2} w^2 \frac{p}{\alpha_0^2} \right].$$

Then

$$\begin{aligned} I_k &:= \int_{\mathbb{R}} \left| \pi_k(w|E, \Delta) - \sqrt{\frac{p}{2\pi\alpha_0^2}} \exp \left[ -\frac{1}{2} w^2 \frac{p}{\alpha_0^2} \right] \right| dw \\ &= \int_{\mathbb{R}} \left| C_k^{-1} g_k(w) + \left( C_k^{-1} \pi(\alpha_0) - \sqrt{\frac{p}{2\pi\alpha_0^2}} \right) \exp \left[ -\frac{1}{2} w^2 \frac{p}{\alpha_0^2} \right] \right| dw \\ &\leq C_k^{-1} \int_{\mathbb{R}} |g_k(w)| dw + \int_{\mathbb{R}} \left| C_k^{-1} \pi(\alpha_0) - \sqrt{\frac{p}{2\pi\alpha_0^2}} \right| \exp \left[ -\frac{1}{2} w^2 \frac{p}{\alpha_0^2} \right] dw. \end{aligned} \tag{16}$$

We first prove that  $\int_{\mathbb{R}} |g_k(w)| dw \rightarrow 0$  as  $n \rightarrow \infty$  (and thus, as  $k \rightarrow \infty$ ). Let  $\mathbb{R} = R_1 \cup R_2$ , where  $R_1 = \{w : |w| > k^{1/2}\delta\}$  and  $R_2 = \{w : |w| < k^{1/2}\delta\}$  for some  $\delta > 0$ . We show that  $\int_{R_j} |g_k(w)| dw \rightarrow 0$  as  $n \rightarrow \infty$ , for  $j = 1, 2$ .

**Proof for  $j = 1$ .** Note first that

$$\begin{aligned} \int_{R_1} |g_k(w)| dw &\leq \int_{R_1} \pi(\hat{\alpha}_\bullet + k^{-1/2}w) \exp [L_k(\hat{\alpha}_\bullet + k^{-1/2}w) - L_k(\hat{\alpha}_\bullet)] dw + \int_{R_1} \pi(\alpha_0) \exp \left[ -\frac{1}{2} w^2 \frac{p}{\alpha_0^2} \right] dw \\ &\leq \sqrt{k} \exp(-k\epsilon) + \pi(\alpha_0) \int_{R_1} \exp \left[ -\frac{1}{2} w^2 \frac{p}{\alpha_0^2} \right] dw, \end{aligned}$$

where the second inequality follows from (14). The term  $\int_{R_1} \exp \left[ -\frac{1}{2} w^2 \frac{p}{\alpha_0^2} \right] dw$  is proportional to  $\Psi(-k^{1/2}\delta)$ , where  $\Psi$  is

the cdf of the normal distribution  $\mathcal{N}(0, \frac{\alpha_0^2}{p})$ . Therefore, this term converges to 0 as  $n \rightarrow \infty$  and finally,  $\int_{R_1} |g_k(w)| dw \rightarrow 0$  as  $n \rightarrow \infty$ .

**Proof for  $j = 2$ .** Expanding  $L_k$  in a Taylor series, we have

$$L_k(\hat{\alpha}_\bullet + k^{-1/2}w) - L_k(\hat{\alpha}_\bullet) = \frac{w}{\sqrt{k}} \frac{\partial L_k(\hat{\alpha}_\bullet)}{\partial \alpha} + \frac{1}{2} \frac{w^2}{k} \frac{\partial^2 L_k(\hat{\alpha}_\bullet)}{\partial \alpha^2} + R_n(w),$$

where  $\frac{\partial^2 L_k(\alpha)}{\partial \alpha^2} = -\frac{1}{\alpha^2} \sum_{i=1}^k \delta_{[n-k+i]}$ ,  $R_n(w) = \frac{1}{6} \left( \frac{w}{\sqrt{k}} \right)^3 \frac{\partial^3 L_k(\tilde{\alpha})}{\partial \alpha^3}$  and  $\tilde{\alpha}$  lies between  $\hat{\alpha}_\bullet$  and  $\hat{\alpha}_\bullet + k^{-1/2}w$ . First, we note that  $\frac{w}{\sqrt{k}} \frac{\partial L_k(\hat{\alpha}_\bullet)}{\partial \alpha}$  is equal to 0 or converges to 0 as  $n \rightarrow \infty$ . To see this, consider first the case where  $\hat{\alpha}_\bullet = \hat{\alpha}_{MPE}^{(j)}$ . Note that  $\hat{\alpha}_{MPE}^{(j)}$  coincides with the maximum likelihood estimator of  $\alpha$ . Thus we have  $\frac{w}{\sqrt{k}} \frac{\partial L_k(\hat{\alpha}_{MPE}^{(j)})}{\partial \alpha} = 0$ . Now, consider the case where  $\hat{\alpha}_\bullet = \hat{\alpha}_{MAP}^{(j)} = (\sum_{i=1}^k \delta_{[n-k+i]} - 1) / \sum_{i=1}^k \log e_{[n-k+i]}$ . Then

$$\begin{aligned} \frac{\partial L_k(\hat{\alpha}_{MAP}^{(j)})}{\partial \alpha} &= \frac{1}{\hat{\alpha}_{MAP}^{(j)}} \sum_{i=1}^k \delta_{[n-k+i]} - \sum_{i=1}^k \log e_{[n-k+i]} \\ &= \frac{\sum_{i=1}^k \log e_{[n-k+i]}}{\sum_{i=1}^k \delta_{[n-k+i]} - 1} \left[ \sum_{i=1}^k \delta_{[n-k+i]} - 1 \right] + \frac{\sum_{i=1}^k \log e_{[n-k+i]}}{\sum_{i=1}^k \delta_{[n-k+i]} - 1} - \sum_{i=1}^k \log e_{[n-k+i]} \\ &= \frac{\frac{1}{k} \sum_{i=1}^k \log e_{[n-k+i]}}{\frac{1}{k} \sum_{i=1}^k \delta_{[n-k+i]} - \frac{1}{k}}, \end{aligned}$$

**Table 6**  
Optimal results for GPD with  $\mu = 2, \sigma = 4, \alpha = 1.25$ .

		$k_{opt}$	$\bar{p}(k_{opt})$	Estimator	MSE
$\beta = \frac{5}{36}$	$\hat{\alpha}_1$	97	0.8880	1.1592	0.0216
	$\hat{\alpha}_2$	97	0.8880	1.1458	0.0239
	$\hat{\alpha}_3$	99	0.8876	<b>1.1817</b>	<b>0.0179</b>
	$\hat{\alpha}_4$	103	0.8874	1.1631	0.0204
$\beta = \frac{5}{12}$	$\hat{\alpha}_1$	82	0.7267	1.1315	0.0337
	$\hat{\alpha}_2$	83	0.7267	1.1125	0.0381
	$\hat{\alpha}_3$	77	0.7275	<b>1.1686</b>	<b>0.0269</b>
	$\hat{\alpha}_4$	77	0.7275	1.1634	0.0277
$\beta = \frac{5}{4}$	$\hat{\alpha}_1$	71	0.4974	1.0648	0.0643
	$\hat{\alpha}_2$	109	0.4982	1.0130	0.0746
	$\hat{\alpha}_3$	71	0.4974	1.1306	0.0467
	$\hat{\alpha}_4$	109	0.4982	<b>1.1416</b>	<b>0.0354</b>
$\beta = \frac{15}{4}$	$\hat{\alpha}_1$	98	0.3598	0.9919	0.0929
	$\hat{\alpha}_2$	98	0.3598	0.9638	0.1077
	$\hat{\alpha}_3$	78	0.3500	1.1088	0.0625
	$\hat{\alpha}_4$	163	0.3728	<b>1.1719</b>	<b>0.0261</b>

**Table 7**  
Optimal results for log–logistic distribution with  $\alpha = 0.75$ .

		$k_{opt}$	$\bar{p}(k_{opt})$	Estimator	MSE
$\beta = \frac{3}{36}$	$\hat{\alpha}_1$	80	0.9347	0.6881	0.0099
	$\hat{\alpha}_2$	79	0.9348	0.6890	0.0110
	$\hat{\alpha}_3$	76	0.9347	<b>0.7102</b>	<b>0.0079</b>
	$\hat{\alpha}_4$	80	0.9347	0.6916	0.0096
$\beta = \frac{3}{12}$	$\hat{\alpha}_1$	68	0.7947	0.6790	0.0132
	$\hat{\alpha}_2$	68	0.7947	0.6665	0.0149
	$\hat{\alpha}_3$	72	0.7946	<b>0.7033</b>	<b>0.0099</b>
	$\hat{\alpha}_4$	72	0.7946	0.6899	0.0114
$\beta = \frac{3}{4}$	$\hat{\alpha}_1$	52	0.5009	0.6466	0.0254
	$\hat{\alpha}_2$	67	0.5018	0.6121	0.0303
	$\hat{\alpha}_3$	55	0.5010	<b>0.6997</b>	0.0163
	$\hat{\alpha}_4$	81	0.5008	0.6973	<b>0.0133</b>
$\beta = \frac{9}{4}$	$\hat{\alpha}_1$	60	0.2047	0.5347	0.0691
	$\hat{\alpha}_2$	73	0.2038	0.4860	0.0873
	$\hat{\alpha}_3$	53	0.2049	0.6582	0.0301
	$\hat{\alpha}_4$	227	0.2158	<b>0.7103</b>	<b>0.0167</b>

which converges a.s. to  $\frac{1}{\alpha_0}$  and thus,  $\frac{w}{\sqrt{k}} \frac{\partial L_k(\hat{\alpha}_{MAP}^{(j)})}{\partial \alpha}$  converges to 0. Finally, consider the case where  $\hat{\alpha}_\bullet = \hat{\alpha}_{MAP}^{(MDI)}$ . Recall that  $\hat{\alpha}_{MAP}^{(MDI)}$  maximizes the log-posterior density  $\ell(\alpha|E, \Delta)$  (see Section 2.1) and note that  $\ell(\alpha|E, \Delta) \propto L_k(\alpha) + \log \alpha - \frac{1}{\alpha}$ . It follows that

$$0 = \frac{\partial \ell(\hat{\alpha}_{MAP}^{(MDI)}|E, \Delta)}{\partial \alpha} = \frac{\partial L_k(\hat{\alpha}_{MAP}^{(MDI)})}{\partial \alpha} + \frac{1}{\hat{\alpha}_{MAP}^{(MDI)}} + \frac{1}{(\hat{\alpha}_{MAP}^{(MDI)})^2}$$

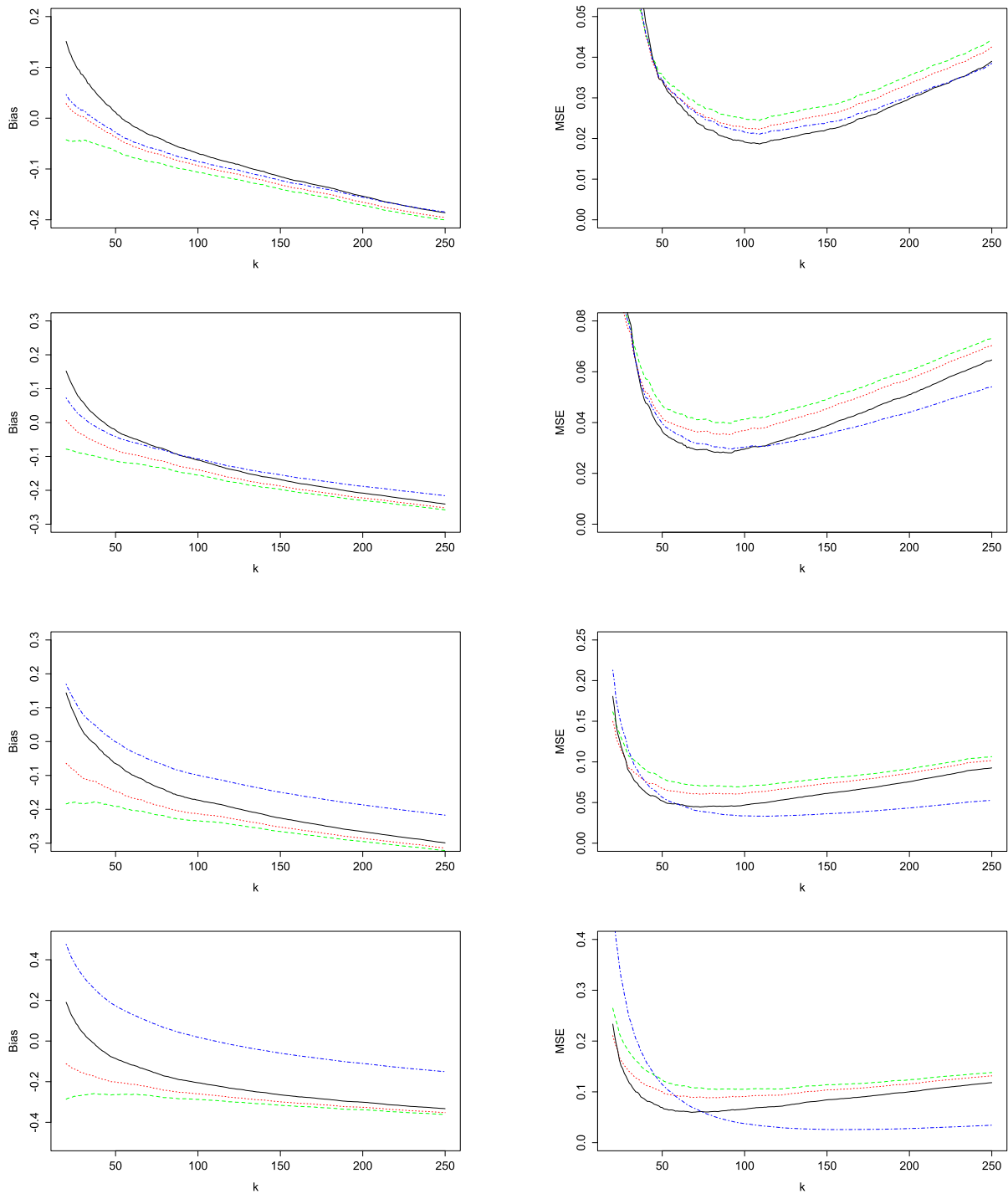
and thus

$$\frac{w}{\sqrt{k}} \frac{\partial L_k(\hat{\alpha}_{MAP}^{(MDI)})}{\partial \alpha} = -\frac{w}{\sqrt{k}} \frac{1}{\hat{\alpha}_{MAP}^{(MDI)}} - \frac{w}{\sqrt{k}} \frac{1}{(\hat{\alpha}_{MAP}^{(MDI)})^2}.$$

This converges to 0 since  $\hat{\alpha}_{MAP}^{(MDI)}$  converges a.s. to  $\alpha_0$ . Now, by a.s. convergence of  $\hat{\alpha}_\bullet$  to  $\alpha_0$ , we get that  $\frac{1}{2} \frac{w^2}{k} \frac{\partial^2 L_k(\hat{\alpha}_\bullet)}{\partial \alpha^2}$  converges to  $-\frac{1}{2} \frac{w^2}{\alpha_0^2}$ . Moreover, for every  $w, R_n(w) \rightarrow 0$  as  $n \rightarrow \infty$ . To see this, note that for every  $w$  and large enough  $n$ ,

$$|R_n(w)| \leq \frac{1}{6} \frac{|w|^3}{\sqrt{k}} \frac{1}{k} \sum_{i=1}^k h(Z_i, \delta_i),$$

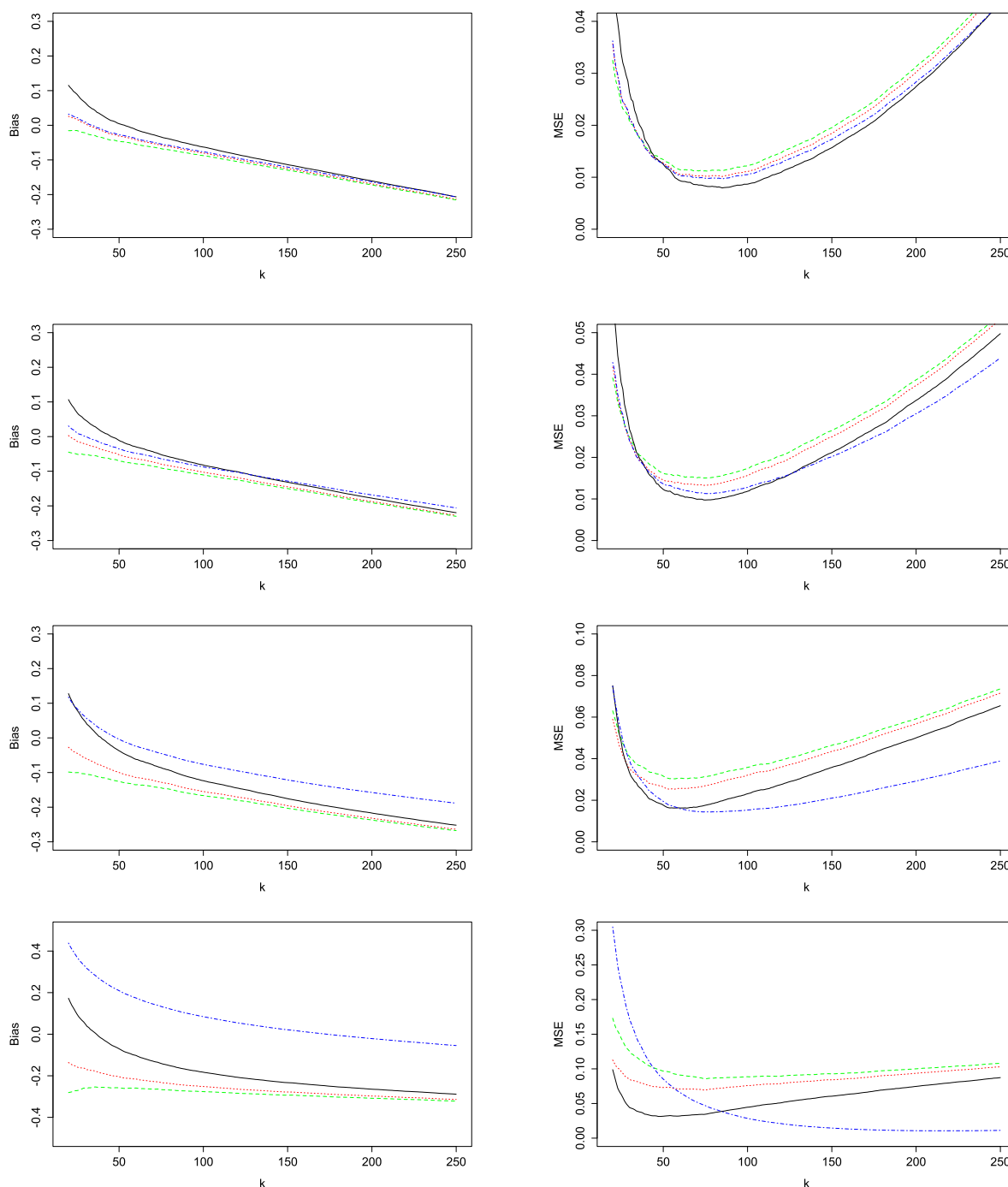
by condition (C2). As  $n \rightarrow \infty$ , the right-hand side of this inequality tends to 0, by condition (C2). Thus, as  $n \rightarrow \infty, L_k(\hat{\alpha}_\bullet + k^{-1/2}w) - L_k(\hat{\alpha}_\bullet)$  converges to  $-\frac{1}{2}w^2 \frac{p}{\alpha_0^2}$  and finally,  $g_k(w) \rightarrow 0$ . Now, we prove that for large enough  $n, |g_k|$  is dominated by an integrable function on the set  $R_2 = \{w : |w| < k^{1/2}\delta\}$ . Let  $\xi > 0$  be an arbitrarily small constant and  $n$  be



**Fig. 11.** Empirical bias (left) and MSE (right) of  $\hat{\alpha}_1$  (red dotted),  $\hat{\alpha}_2$  (green dashed),  $\hat{\alpha}_3$  (black line),  $\hat{\alpha}_4$  (blue dotdash) for GPD with  $\mu = 2$ ,  $\sigma = 4$ ,  $\alpha = 1.25$  and  $\beta = \frac{5}{36}$  (top row),  $\beta = \frac{5}{12}$  (second row),  $\beta = \frac{5}{4}$  (third row) and  $\beta = \frac{15}{4}$  (bottom row). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sufficiently large (say,  $n \geq n_\xi$ ) so that  $\frac{1}{k} \sum_{i=1}^k h(Z_i, \delta_i) / \mathbb{E}h(Z, \delta) < 1 + \xi$  and  $(p/\alpha_0^2) / (-\frac{1}{k} \frac{\partial^2 L_k(\hat{\alpha}_\bullet)}{\partial \alpha^2}) < 1 + \xi$ . Take  $\delta$  such that  $\delta < \frac{3}{2} \frac{p}{\alpha_0^2} \frac{1}{(1+\xi)^2 \mathbb{E}h(Z, \delta)}$ . Then, for any  $w \in R_2$  and  $n \geq n_\xi$ , we have

$$\begin{aligned}
 |R_n(w)| &\leq \frac{1}{6} w^2 \delta \frac{1}{k} \sum_{i=1}^k h(Z_i, \delta_i) \\
 &< -\frac{1}{4} \frac{w^2}{k} \frac{\partial^2 L_k(\hat{\alpha}_\bullet)}{\partial \alpha^2}.
 \end{aligned}$$



**Fig. 12.** Empirical bias (left) and MSE (right) of  $\hat{\alpha}_1$  (red dotted),  $\hat{\alpha}_2$  (green dashed),  $\hat{\alpha}_3$  (black line),  $\hat{\alpha}_4$  (blue dotdash) for log–logistic distribution with  $\alpha = 0.75$  and  $\beta = \frac{3}{36}$  (top row),  $\beta = \frac{3}{12}$  (second row),  $\beta = \frac{3}{4}$  (third row) and  $\beta = \frac{9}{4}$  (bottom row). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Thus, for  $n$  large enough, we have

$$\begin{aligned} \exp [L_k(\hat{\alpha}_\bullet + k^{-1/2}w) - L_k(\hat{\alpha}_\bullet)] &< C \exp \left[ \frac{1}{4} \frac{w^2}{k} \frac{\partial^2 L_k(\hat{\alpha}_\bullet)}{\partial \alpha^2} \right] \\ &< C \exp \left[ -\frac{p}{8\alpha_0^2} w^2 \right], \end{aligned}$$

for some positive constant  $C$  and it easily follows that  $|g_k(w)|$  is dominated by an integrable function on  $R_2$ . Finally, it follows from the dominated convergence theorem that  $\int_{R_2} |g_k(w)| dw$  tends to 0 as  $n \rightarrow \infty$ .

Now, we consider the second term on the right-hand side of (16). Note first that

$$\left| C_k - \pi(\alpha_0) \sqrt{\frac{2\pi\alpha_0^2}{p}} \right| \leq \int_{\mathbb{R}} |g_k(w)| dw.$$

This implies that  $C_k^{-1}\pi(\alpha_0) \rightarrow \sqrt{\frac{p}{2\pi\alpha_0^2}}$  as  $n \rightarrow \infty$  and thus,  $\int_{\mathbb{R}} \left| C_k^{-1}\pi(\alpha_0) - \sqrt{\frac{p}{2\pi\alpha_0^2}} \right| \exp[-\frac{1}{2}w^2 \frac{p}{\alpha_0^2}] dw \rightarrow 0$  as  $n \rightarrow \infty$ . Finally,  $I_k$  tends to 0 as  $n \rightarrow \infty$ , which concludes the proof.  $\square$

## Appendix B. Results for the GPD and log–logistic distribution

We provide similar results as in Tables 1 and 2 and Figs. 1 and 2, for the GPD with  $\mu = 2$ ,  $\sigma = 4$ ,  $\alpha = 1.25$  (Table 6 and Fig. 11) and log–logistic distribution with  $\alpha = 0.75$  (Table 7 and Fig. 12). Results are obtained for a sample of size  $n = 1000$  and  $m = 1000$  simulated samples.

## References

- Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N., 1993. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D., Ferro, C., 2004. *Statistics of Extremes*. John Wiley & Sons Ltd., Chichester.
- Beirlant, J., Guillou, A., Dierckx, G., Fils-Villetard, A., 2007. Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* 10 (3), 151–174.
- Brahimi, B., Meraghni, D., Necir, A., 2013. Gaussian approximation to the extreme value index estimator of a heavy-tailed distribution under random censoring. *Math. Methods Statist.* 24 (4), 266–279.
- Cabras, S., Castellanos, M.E., 2011. A Bayesian approach for estimating extreme quantiles under a semiparametric mixture model. *Astin Bull.* 41 (1), 87–106.
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London.
- Coles, S., Powell, E., 1996. Bayesian methods in extreme value modelling: a review and new developments. *Internat. Statist. Rev.* 64 (1), 119–136.
- de Haan, L., Ferreira, A., 2006. *Extreme Value Theory*. Springer, New York.
- Deheuvels, P., Haeusler, E., Mason, M., 1988. Almost sure convergence of the Hill estimator. *Math. Proc. Camb. Phil. Soc.* 104 (2), 371–381.
- Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* 17 (4), 1833–1855.
- de Zea Bermudez, P., Kotz, S., 2010. Parameter estimation of the generalized Pareto distribution-Part II. *J. Statist. Plann. Inference* 140 (6), 1374–1388.
- Diebolt, J., El-Aroui, M., Garrido, M., Girard, S., 2005. Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling. *Extremes* 8 (1–2), 57–78.
- do Nascimento, F.F., Gamerman, D., Lopes, H.F., 2012. A semiparametric Bayesian approach to extreme value estimation. *Stat. Comput.* 22, 661–675.
- Einmahl, J.H.J., Fils-Villetard, A., Guillou, A., 2008. Statistics of extremes under random censoring. *Bernoulli* 14 (1), 207–227.
- Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events*. Springer-Verlag, Berlin.
- Ergashev, B., Mittnik, S., Sekeris, E., 2013. A Bayesian approach to extreme value estimation in operational risk modeling. *J. Oper. Risk* 8 (4), 55–81.
- Ghosh, J.K., Delampady, M., Samanta, T., 2006. *An Introduction to Bayesian Analysis*. Springer, New York.
- Gomes, M.I., Neves, M.M., 2011. Estimation of the extreme value index for randomly censored data. *Biom. Lett.* 48 (1), 1–22.
- Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3 (5), 1163–1174.
- Jeffreys, H., 1961. *Theory of Probability*. Oxford University Press.
- Koning, A.J., Peng, L., 2008. Goodness-of-fit tests for a heavy tailed distribution. *J. Statist. Plann. Inference* 138 (12), 3960–3981.
- Liang, Z.M., Li, B.Q., Yu, Z.B., Chang, W.J., 2011. Application of Bayesian approach to hydrological frequency analysis. *Sci. China Technol. Sci.* 54 (5), 1183–1192.
- Mason, D.M., 1982. Laws of large numbers for sums of extreme values. *Ann. Probab.* 10 (3), 754–764.
- Pickands, J., 1975. Statistical inference using extreme order statistics. *Ann. Statist.* 3 (1), 119–131.
- R Development Core Team, 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org>.
- Resnick, S.I., 2007. *Heavy-Tail Phenomena. Probabilistic and Statistical Modeling*. Springer, New York.
- So, M.K.P., Chan, R.K.S., 2014. Bayesian analysis of tail asymmetry based on a threshold extreme value model. *Comput. Statist. Data Anal.* 71, 568–587.
- Stephenson, A., Tawn, J., 2004. Bayesian inference for extremes: accounting for the three extremal types. *Extremes* 7 (4), 291–307.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Springer, New York.
- Worms, J., Worms, R., 2014. New estimators of the extreme value index under random right censoring, for heavy tailed distributions. *Extremes* 17 (2), 337–358.
- Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons Ltd., New York.
- Zellner, A., 1996. Models, prior information, and Bayesian analysis. *J. Econometrics* 75 (1), 51–68.

## Further reading

- Beirlant, J., Guillou, A., Toulemonde, G., 2010. Peaks-over-threshold modeling under random censoring. *Comm. Statist. Theory Methods* 39 (7), 1158–1179.
- van der Vaart, A.W., 1998. *Asymptotic Statistics*. Cambridge University Press, Cambridge.

## IV.2 Deuxième Partie :

### Modeling Algerian cars claims using LPC model under randomly right-censored data

Recently in actuarial cars insurance studies, various kind of models were introduced by mixing two or more probability density functions (pdf) to describe a complex behaviour of cars amount claims (typically a lognormal distribution for moderate cars sinister, and a Pareto type for corporals cars costs). The well formulated model was given by [Cooray et al. \(2005\)](#) in [19], called the two parameters lognormal-Pareto composite model. Many works recently discussed the statistical inference in the LPC model as [Scollnick \(2007\)](#), [Pigeon et al. \(2011\)](#), [Nadarajah et al. \(2012\)](#), or the more general composite Stoppa model proposed by [Calderín-Ojeda et al. \(2015\)](#). Over this literature, the statistical inference is based essentially on the maximum likelihood (ML) estimation approach. Other inference alternatives include construction of a Bayesian estimation procedure as in [20] or Bayesian inference for stable Paretian distribution under right random censoring proposed by [Ameraoui et al. \(2016\)](#) in [2].

In this paper, we suggest an adaptation of the LPC model when randomly right-censored data are considered. Both simulation and real data validation are examined to evaluate the proposed estimation procedure. The paper is constructed as follows : First, we describe in section IV.2.1, the LPC model, and give the basic characteristics of a LPC distribution. In section IV.2.2, we construct our maximum likelihood estimators under random right-censoring. We also report some results of a comprehensive simulation study. Finally, an application to Algerian cars claims data is carried out in section IV.2.3. A discussion and some perspectives are given in section IV.2.4.

#### IV.2.1 Lognormal-Pareto composite model

[Klugman et al. \(2004,2012\)](#) expressed a composite expression of two probability density functions ( $f_1$  and  $f_2$ , supposed absolutely continuous functions, and  $F_1, F_2$  are their corresponding cumulative density functions (cdf)) as :

$$f(x) = \begin{cases} a_1 f_1^*(x) & \text{if } -\infty < x < \theta \\ a_2 f_2^*(x) & \text{if } \theta \leq x < +\infty, \end{cases} \quad (\text{IV.1})$$

where  $f_1^*(x)$  (respectively  $f_2^*(x)$ ) is given by  $\frac{f_1(x)}{\int_{-\infty}^{\theta} f_1(x)dx}$  (respectively  $\frac{f_2(x)}{\int_{\theta}^{+\infty} f_2(x)dx}$ ). The weights  $a_1$  and  $a_2$  serve as normalisation coefficients of the pdf  $f$ .

Hence, any construction of a composite model is conditioned by the definition of  $a_1$  and  $a_2$ , necessary with summation equals to 1. [Nadarajah et al. \(2012\)](#) consider some criteria for a composite model in (IV.1). The first condition is to consider that  $f$  is continuous in any  $x \in ]-\infty, +\infty[$ , which holds if and only if  $f(\theta^-) = f(\theta^+)$ . This means that the left piecewise of  $f$  is joined to the tail, at the threshold  $\theta$ . The second criterion is smoothness of the pdf  $f$ , i.e it is necessary to ensure the differentiability of  $f$  at  $\theta$ . A major precondition as cited above, is the mixing weight criteria, which can be written as :

$$a_1 = \frac{1}{1 + \phi}, \quad a_2 = \frac{\phi}{1 + \phi},$$

where  $\phi = \frac{f_1(x)[1 - F_2(x)]}{f_2(x)F_1(x)} > 0$ .

[Cooray et al. \(2005\)](#) propose a composite model combining a lognormal pdf with a Pareto type II pdf,

as

$$f(x) = \begin{cases} \frac{\beta\theta^\beta}{(1 + \Phi(k))x^{\beta+1}} \exp \left[ -0.5 \left( (\beta/k) \ln(x/\theta) \right)^2 \right] & \text{if } 0 < x \leq \theta \\ \frac{\beta\theta^\beta}{(1 + \Phi(k))x^{\beta+1}} & \text{if } \theta < x < \infty \end{cases}, \quad (\text{IV.2})$$

where  $\Phi(\cdot)$  is the cumulative distribution function (cdf) of the standard normal distribution, and  $k$  is a positive constant solution of the equation  $\exp(-k^2) = 2\pi k^2$  -numerical resolution give  $k = 0.372238898$ -.  $\theta$  (respectively  $\beta$ ) is naturally positive, representing the scale (respectively the shape) parameter of the LPC distribution. Thus, we denote any random variable with LPC pdf by  $X \sim LPC(\beta, \theta)$ .

The survival function  $S(x) = 1 - F(x)$ , where  $F(x)$  is the cdf of a random variable  $X \sim LPC(\beta, \theta)$ , and the quantile function  $Q(u)$  are respectively given by

$$S(x) = \begin{cases} 1 - \frac{1}{(1 + \Phi(k))} \Phi \left( (\beta/k) \ln(x/\theta) + k \right) & \text{if } 0 < x \leq \theta \\ \frac{1}{(1 + \Phi(k))} (\theta/x)^\beta & \text{if } \theta < x < \infty \end{cases}, \quad (\text{IV.3})$$

and

$$Q(u) = \begin{cases} \theta \exp \left[ (k/\beta) \left( \Phi^{-1}((1 + \Phi(k))u) - k \right) \right] & \text{if } 0 < u \leq u_0 \\ \theta [(1 - u)(1 + \Phi(k))]^{-1/\beta} & \text{if } u_0 < u < 1 \end{cases}, \quad (\text{IV.4})$$

where  $u_0 = \Phi(k)/(1 + \Phi(k))$ .

The  $m^{th}$  integer moment,  $\mathbb{E}[X^m]$  of the LPC distribution exists only for  $m < \beta$  and is given by

$$\mathbb{E}[X^m] = \frac{1}{1 + \Phi(k)} \left[ \Phi(k - km/\beta) \exp \left\{ \frac{1}{2} \left( \frac{k}{\beta} \right)^2 (m^2 - 2m\beta) \right\} + \frac{\beta}{\beta - m} \right]$$

## IV.2.2 LPC parameters estimation under random censoring

In this paper, we address the estimation of the two LPC distribution parameters when observations  $X_1, X_2, \dots, X_n$  are randomly right-censored. Censoring commonly occurs in the analysis of event time or incomplete data observation. For example,  $X$  may represent the amount of claims relating to an insurance policy, and the amount of loss can not be definitively evaluated when the data are collected, the variable of interest  $X$  is not completely available (or partially observed). An appropriate way to model this situation is to introduce a random variable  $Y$  (called a censoring random variable) such that observations consist of pairs  $(Z_i, \delta_i), 1 \leq i \leq n$  where  $Z_i = \min(X_i, Y_i)$ ,  $\delta_i = \mathbb{1}_{\{X_i \leq Y_i\}}$  and  $\mathbb{1}$  is the indicator function. Let consider now, a variable of interest  $X$  with a  $LPC(\beta, \theta)$  pdf  $f$  as in (IV.2) and a survival function  $S$  given by (IV.3), and a censoring random variable  $Y$  with pdf  $g$  and survival function  $\bar{G}$  (supposed independent of  $\beta$  and  $\theta$ ). We assume also that  $X$  and  $Y$  are independent and that we observe only the  $n$  independent pairs  $(Z_i, \delta_i), 1 \leq i \leq n$ . The likelihood function from the observed sample  $\chi = (Z_i, \delta_i), 1 \leq i \leq n$  is given by

$$\mathcal{L}(\chi, \beta, \theta) = \prod_{i=1}^n (f(z_i) \cdot \bar{G}(z_i))^{\delta_i} \times (g(z_i) \cdot S(z_i))^{1-\delta_i}$$

We keep in mind that we want to bring the inference on the two LPC distribution parameters  $\beta$  and  $\theta$ . This results in the following partial likelihood function

$$\begin{aligned}\mathcal{L}^*(\chi, \beta, \theta) &= \prod_{i=1}^n (f(z_i))^{\delta_i} \times (S(z_i))^{1-\delta_i} \\ &= \prod_{i=1}^n \left( \frac{\beta \theta^\beta}{(1 + \Phi(k)) z_i^{\beta+1}} \exp \left[ -0.5 \mathbb{1}_{\{z_i \leq \theta\}} ((\beta/k) \ln(z_i/\theta))^2 \right] \right)^{\delta_i} \times \left( \frac{1}{1 + \Phi(k)} \left( \frac{\theta}{z_i} \right)^\beta \right)^{(1-\delta_i) \mathbb{1}_{\{z_i > \theta\}}} \\ &\quad \times \left[ 1 - \frac{1}{(1 + \Phi(k))} \Phi((\beta/k) \ln(z_i/\theta) + k) \right]^{(1-\delta_i) \mathbb{1}_{\{z_i \leq \theta\}}}\end{aligned}$$

and the corresponding partial log-likelihood function is

$$\begin{aligned}\ell(\chi, \beta, \theta) &= \log \mathcal{L}^*(\chi, \beta, \theta) \\ &= \sum_{i=1}^n \delta_i \left[ \log \beta + \beta \log \theta - \log(1 + \Phi(k)) - (\beta + 1) \log(z_i) - \mathbb{1}_{\{z_i \leq \theta\}} \frac{\beta^2}{2k^2} \log^2 \left( \frac{z_i}{\theta} \right) \right] \\ &\quad + \sum_{i=1}^n \mathbb{1}_{\{z_i \leq \theta\}} (1 - \delta_i) \log \left( 1 - \frac{1}{(1 + \Phi(k))} \Phi((\beta/k) \ln(z_i/\theta) + k) \right) \\ &\quad + \sum_{i=1}^n \mathbb{1}_{\{z_i > \theta\}} (1 - \delta_i) \left[ \beta \log \left( \frac{\theta}{z_i} \right) - \log(1 + \Phi(k)) \right]\end{aligned}$$

The maximum likelihood estimators of  $\beta$  and  $\theta$  are the simultaneous solution of

$$\begin{cases} \frac{\partial}{\partial \beta} \ell(\chi, \beta, \theta) = 0 \\ \frac{\partial}{\partial \theta} \ell(\chi, \beta, \theta) = 0 \end{cases} \quad (\text{IV.5})$$

We can easily see that there is no closed form expression of the maximum likelihood estimates of  $(\beta, \theta)$ . So, they must be obtained numerically using for example a quasi-Newton algorithm. The following subsection will be dedicated to the estimation procedure using Monte-Carlo replication.

## Simulation study

In this subsection, we assess, via simulations, finite-sample performance of estimators obtained as numerical solution of the system (IV.5), using the R-package `maxLik` (see for instance <http://www.maxlik.org/>).

*Study design.* The simulation design is as follows. Let  $X$  and  $Y$  be independent random variables with cdf  $F_X \sim LPC(\beta, \theta)$  and  $G_Y \sim \mathcal{E}(\lambda)$  respectively.  $X$  is the variable of interest and  $Y$  is the censoring random variable. Thus, for a given  $\beta, \theta$ , a relevant choice of  $\lambda$  allows us to simulate data with an approximate proportion  $\hat{p} := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq Y_i\}}$  of non-censored  $X_i$ s. Our simulation procedure is as follows :

1. We simulate a sample of  $n$  independent copies of  $(Z, \delta)$ , where  $Z = \min(X, Y)$ ,  $\delta = \mathbb{1}_{\{X \leq Y\}}$  and the target censoring proportion of  $X$  is  $1 - \hat{p}$ . For a given  $\beta$  and  $\theta$ , letting  $\lambda = 0.01, 0.02, 0.03, 0.05$  and  $0.1$ , should allow various censoring percentage of  $X$ .
2. For each  $\lambda$ , we compute the maximum likelihood estimators of  $\beta, \theta$
3. Steps 1-2 are repeated  $m$  times, so that we obtain  $m$  realisations pairs of  $(\hat{\beta}, \hat{\theta})$ , for each  $\lambda$ .
4. For each  $\lambda$ , we calculate the average value and MSE (Mean Square Error) of  $(\hat{\beta}, \hat{\theta})$  over the  $m$  estimates.

*Simulation results discussion.* Simulations are conducted using the statistical software R [65]. Results are provided for a sample of size  $n = 1000$  and  $m = 1000$  simulated samples. Let first denote  $\forall j = 1$  to  $m$ ,  $\hat{p}^{(j)} := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i^{(j)} \leq Y_i^{(j)}\}}$ , representing the proportion of uncensored data over the  $j^{th}$  simulated sample. In tables IV.1 and IV.2, we report the averaged (over the  $m$  simulated samples) value and empirical MSE of each  $\hat{\beta}$  and  $\hat{\theta}$ . We also report the averaged proportion of uncensored data among observations  $i$  such that  $X_i < Y_i$  given by  $\bar{p} = \frac{1}{m} \sum_{j=1}^m \hat{p}^{(j)}$ . From these results, it appears, as expected, that the bias and MSE of all estimators decrease when censoring in the right tail of  $X$  decreases (one can see also that the proportion of censoring decreases when the value of the exponential parameter  $\lambda$  increases). We also propose a bootstrapped confidence interval ( $IC_B$ ) at level 95% for each LPC parameter estimator of size  $B = 10000$  (see Algorithm 1 in appendix).

$\lambda$	0.01	0.02	0.03	0.05	1
$\bar{p}$	0.9152	0.8461	0.7872	0.6870	0.5053
$\beta$	1.5013	1.5021	1.4979	1.5040	1.5060
(MSE( $\beta$ ))	(0.0019)	(0.0021)	(0.0023)	(0.0026)	(0.0037)
$IC_B(95\%)$ for $\beta$	$1.509 \pm 0.0871$	$1.5016 \pm 0.1043$	$1.5002 \pm 0.1621$	$1.5041 \pm 0.1833$	$1.5032 \pm 0.2204$
$\theta$	5.0028	4.9975	5.0097	5.0015	5.0014
(MSE( $\theta$ ))	(0.0050)	(0.0056)	(0.0061)	(0.0064)	(0.0077)
$IC_B(95\%)$ for $\theta$	$5.0009 \pm 0.1795$	$5.001 \pm 0.2301$	$5.0013 \pm 0.2617$	$5.0041 \pm 0.3411$	$5.0012 \pm 0.3702$

TABLE IV.1 – LPC( $\beta = 1.5, \theta = 5$ ) parameters estimation right censored by  $\mathcal{E}(\lambda)$  distribution

$\lambda$	0.01	0.02	0.03	0.05	1
$\bar{p}$	0.8464	0.7338	0.6447	0.5048	0.2932
$\hat{\beta}$	1.5012	1.5016	1.5020	1.5050	1.5061
(MSE( $\beta$ ))	(0.0021)	(0.0026)	(0.0031)	(0.0037)	(0.0057)
$IC_B(95\%)$ for $\beta$	$1.5046 \pm 0.1480$	$1.5036 \pm 0.2033$	$1.5089 \pm 0.2817$	$1.5092 \pm 0.3041$	$1.5100 \pm 0.3166$
$\hat{\theta}$	10.0051	10.0059	10.0061	9.9915	10.0041
(MSE( $\theta$ ))	(0.0219)	(0.0295)	(0.0260)	(0.0315)	(0.0483)
$IC_B(95\%)$ for $\theta$	$10.003 \pm 0.2837$	$10.0067 \pm 0.3109$	$10.0028 \pm 0.2980$	$10.0017 \pm 0.3428$	$10.0039 \pm 0.4006$

TABLE IV.2 – LPC( $\beta = 1.5, \theta = 10$ ) parameters estimation right censored by  $\mathcal{E}(\lambda)$  distribution

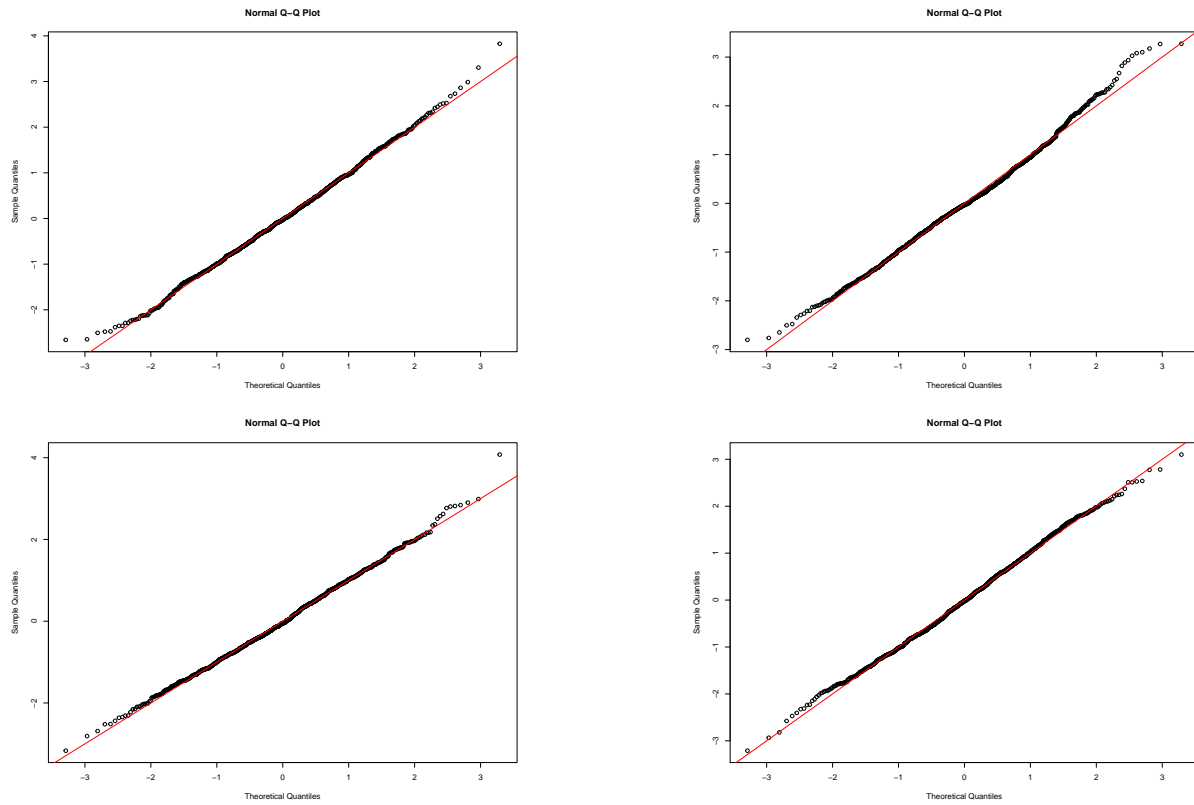


FIGURE IV.1 – **Line 1** : QQ-plot of  $m$  estimations of the scale parameter  $\beta$ , **Line 2** : QQ-plot of  $m$  estimations of the shape parameter  $\theta$  under simulated LPC(1.5, 10) right censored by :  $\mathcal{E}(0.01)$  -left plot-,  $\mathcal{E}(0.1)$  -right plot-

Figure IV.1 shows the parameter estimations empirical (reduced and centred values) quantile to standard normal quantile plot, to investigate for each estimator, the normality behaviour over the  $m$  replications. One can see that almost estimations values of both  $\beta$  and  $\theta$  are well aligned with the first bisecting line. This behaviour is also supported by figure IV.2, where we present the estimated kernel density of the sample  $(\hat{\beta}_j, j = 1 \text{ to } m)$ , which are clearly concentrated around the target value of 1.5

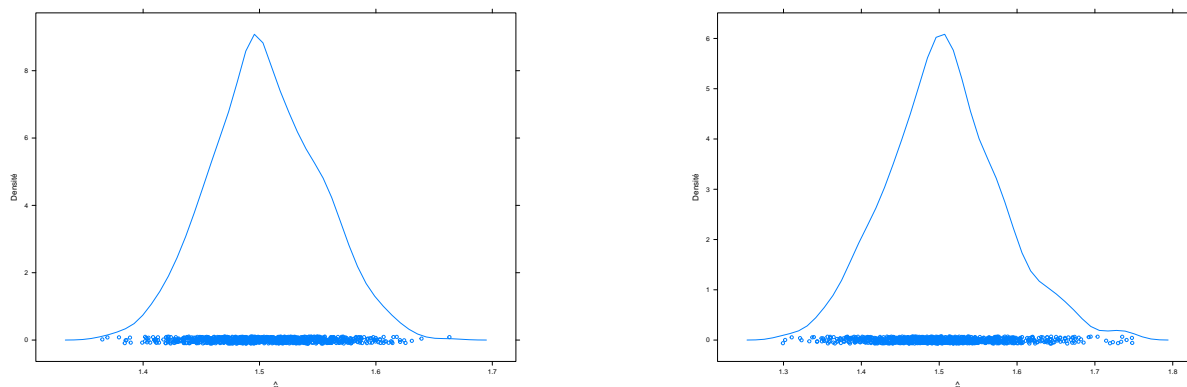


FIGURE IV.2 – Estimated density for  $\beta$  estimations, for  $X \sim LPC(1.5, 10)$  right censored by :  $Y \sim \mathcal{E}(0.01)$  -left plot-,  $Y \sim \mathcal{E}(0.1)$  -right plot-

### IV.2.3 An application to Algerian cars claims data

In this section, we illustrate our methodology on a set of Algerian cars claims data. The dataset contains  $n = 1167$  car insurance claims in Algeria (expressed in  $10^6$  Algerian dinars DZD) between 2007 and 2012. The data is a subset arising from the Project "Étude de Revalorisation du Tarif de la RC Automobile" directed by the actuarial office A.R.M.S.S.M (Algiers) for the Algerian Insurances Union (U.A.R). According to the Algerian premium pricing schema, the subset represents cars claims amount, of type "particular cars without trailer" for "Affair" used, with "3 to 7" horsepower and an indicator which equals 1 if the claim is entirely settled and 0 otherwise. 997 claims are close settled. The other sinister amount are right-censored. Some summary of the data are : minimum claims 0.02354 (which correspond to 2,354.00DZD), the mean claim 2.22300 (222,300.00DZD) and a maximum claim of 57.72000 (5,772,000.00DZD).

Our objective here is to estimate the LPC models parameters using ML estimates, and compares this fit to Lognormal (Adjusted using ML estimation for right censored data in R package STAND - see <https://CRAN.R-project.org/package=STAND>) and Pareto (Adjusted using ML estimation procedure adapted to right censored data) fit. The estimation results are given in Table IV.3. Hence, it's useful to express the fit quality to the data using the following goodness-of-fit measures (see for instance D'Agostino *et al.* (1986)) : For any data sample  $x_1, x_2, \dots, x_n$ , with corresponding ordered data  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  and by denoting  $\hat{F}$  the cdf of the fitted model, we define

- **Kolmogorov-Smirnov (KS) test statistic** :  $D = \max(D^-, D^+)$ , where

$$D^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - \hat{F}(x_{(i)}) \right\}, \quad D^- = \max_{1 \leq i \leq n} \left\{ \hat{F}(x_{(i)}) - \frac{i-1}{n} \right\}$$

- **Anderson-Darling (AD) test statistic** :

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \log(\hat{F}(x_{(i)})) + (2n-2i+1) \log(\hat{F}(x_{(i)}))$$

Note that smaller values of the above measures indicate better adjustment of the model to the data.

Distribution	Parameter	Estimate	Std. Error	AIC	KS statistic ( $D$ )	AD statistics ( $A^2$ )
LPC model	$\beta$	1.4551	0.0409	1991.044	0.0315	24.2853
	$\theta$	2.0897	0.0295			
Pareto	$\alpha$	1.8652	0.0547	4867.346	0.1509	432.5689
	$x_{min}$	0.0235				
Lognormal	$\mu$	0.9265	0.0191	4324.124	0.1768	104.9364
	$\sigma$	0.6417	0.0139			

TABLE IV.3 – Estimated Parameter and Goodness-of-fit for Algerian data claims

Results from Table IV.3 suggest that LPC model provide the better adequate fit to the data. Moreover, we illustrate in figure IV.3 the natural logarithm of Quantile-Quantile plot of LPC, Lognormal and Pareto distribution to the Algerian cars claim. One can clearly see that the LPC model is more adequate to claims data than Pareto or Lognormal distribution. All these findings confirm results from AIC (Akaike Information Criterion),  $D$  statistics and  $A^2$  statistics values in table IV.3.

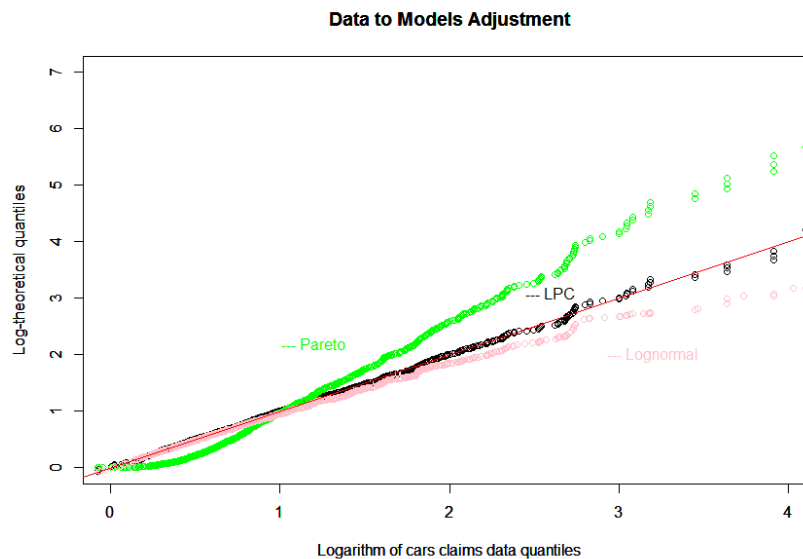


FIGURE IV.3 – Log Quantile-Quantile plot of Algerian claims data for LPC (Black), Pareto (Green) and Lognormal (Pink) distribution.

#### IV.2.4 Conclusion and perspectives

In this paper, we address the identification problem of the LPC model to real actuarial data, when randomly right-censoring is considered. We constructed the corresponding ML estimators of the two LPC distribution. Our simulation results indicate good performance of these estimators. We also proposed an original procedure applied to a set of real data and provided coherent results.

Now, several issues still deserve attention. First, we mentioned that estimation of the LPC parameters constitutes usually a first step in an actuarial analysis. Premium estimation often comes as a second step. In the real-data example, we incorporated our fitted LPC model in the so-called Wang's premium

principle. More investigations of the LPC model performances and applications are needed. This is a subject for our future research.

In practice, it often arises that some covariate information  $W$  is available to the distribution of an interest variable  $X$  depends on  $W$ . In this case, interest turns to estimation of the LPC parameters considering the co-variables in  $W$ . This is a difficult problem which we overcame, in our real-data analysis, by considering sub-populations defined by gender, the used and horsepower of the car. However, a more advanced approach should be developed to tackle this issue. This also constitutes a topic for our future research.

## Acknowledgements

Authors thank the Conference Chair and anonymous reviewers for agreeing to review this paper and their future suggestions that helped improve successive versions of the paper.

Authors acknowledge financial support from the CNEPRU project "Modélisation Risque Actuariel, Assurance, Finance (MSAAF)" (project number C00L03UN160420150004, funded by Agence Nationale de Développement de la Recherche Universitaire, Algeria) and from the project SESA (project number EDC25895, funded by Centre National de la Recherche Scientifique - CNRS, France - and Direction de la Post-Graduation et de la Recherche-Formation - DPGRF, Algeria).

## Appendix

Let  $\Theta = (\theta_1, \theta_2, \dots, \theta_m)$  be a vector of  $m$  any LPC parameter estimations, obtained over the replication procedure in subsection IV.2.2. Construction of the parameter bootstrapped confidence interval at level  $1 - \alpha$  ( $IC_B$ ), is obtained as follows

---

**Algorithm 1** Computing bootstrapped confidence interval at level  $1 - \alpha$

---

**Require:**  $\Theta$  : as vector,  $B$  : integer

$\triangleright B$  represents the number of bootstrap

**procedure** ICBOOT( $\alpha$ )

$\triangleright 1 - \alpha$  is the confidence level

**for**  $b = 1$  to  $B$  **do**

    Drawn randomly from  $\Theta$ , with replacement, a sample  $\theta_1^b, \theta_2^b, \dots, \theta_m^b$  of size  $m$ .

    Compute  $\bar{\theta}_b = \frac{1}{m} \sum_{i=1}^m \theta_i^b$

**end for**

Compute  $\bar{\theta} = \frac{1}{B} \sum_{b=1}^B \bar{\theta}_b$

Compute  $\hat{\sigma}^2 = \frac{1}{B-1} \sum_{b=1}^B [\bar{\theta}_b - \bar{\theta}]^2$

**return**  $(\bar{\theta}, t_{\alpha/2, n-1} \times \sqrt{\hat{\sigma}^2})$

$\triangleright$  where  $t_{\alpha/2, n-1}$  is the  $(1 - \frac{\alpha}{2})$ -Student quantile

**end procedure**

---

# CONCLUSION GÉNÉRALE

À travers cette thèse, nous avons exploré d'une manière non-exhaustive, quelques éléments d'inférence statistique non-paramétrique, semi-paramétrique et bayésien dans la théorie des valeurs extrêmes univariée (TVEU) :

Nous avons présenté tout d'abord, les fondements et les concepts élémentaires introduits dans la littérature de la TVEU, en étalant ces notions en deux chapitres; un premier chapitre dans lequel nous avons énoncé le théorème Tippet-Fisher-Gnedenko, fondateur de la théorie des valeurs extrêmes et une variante de ce théorème, permettant de distinguer le max-domaine d'attraction d'un extremum selon la nature de la distribution d'échantillonnage. Des approches statistiques d'estimation de l'indice  $\gamma$ , basées sur l'échantillon des  $k$ -plus grandes valeurs, ainsi que l'application de la TVE à l'estimation de la notion des quantiles extrêmes, ont été mise au point par des exemples d'application. Un second chapitre relevant la question d'inférence semi-paramétrique dans la TVEU (essentiellement), et dans lequel nous avons aussi rappelé l'ensemble des contributions majeures autour de cette question. Nous considérons que l'estimation de l'indice des valeurs extrêmes, est la première pierre dans toute analyse statistique de données, s'appuyant sur le théorie des valeurs extrêmes, dans un contexte d'information complète ou censurée. Nous avons exploré aussi le comportement des estimateurs proposés à travers des exemples sur données simulées.

La deuxième partie de cette thèse porte sur l'introduction du paradigme bayésien en statistique : un rappel sur les mécanismes d'élaboration d'un modèle bayésien de sa conception a priori, sa construction à la mise au point d'une décision a posteriori. Nous avons aussi soulevé du point de vue théorique, le comportement gaussien de la distribution a posteriori établi asymptotiquement sous certaines hypothèses sur le modèle bayésien supposé et l'estimateur conçu a posteriori. La mise en œuvre pratique de ces concepts bayésiens a été illustrée par un exemple d'estimation de l'indice de queue  $\alpha > 0$  relatif à une distribution à queue lourde et un échantillon supposé être *i.i.d* et complet.

Nos contributions dans cette thèse ont été rapportées dans un dernier chapitre, dans lequel nous avons proposé deux modèles statistiques, conçus pour des distributions à queue lourde, où l'inférence statistique est portée sur un échantillon de données *i.i.d* en présence de censure aléatoire à droite :

- Le premier modèle introduit, porté sur la question d'estimation bayésienne de l'indice  $\alpha > 0$ , d'une distribution à queue lourde en présence de censure aléatoire à droite. Le modèle bayésien (conçu par rapport à l'échantillon des excès relatifs au delà d'un seuil égal à la statistique d'ordre  $(n - k)$ , a été introduit selon trois types de loi a priori (02 lois impropres et une loi conjuguée), à partir duquel nous avons proposé cinq estimateurs bayésiens. Le comportement empirique des cinq estimateurs proposés a été comparé à l'estimateur introduit par [Worms and Worms](#) dans [77] : l'étude par simulation proposé montre l'efficacité (au sens minimisation de l'erreur standard) de l'estimateur MAP construit sous une loi a priori de Zellner, mais conditionnellement à la connaissance d'un seuil de fraction  $k/n$  "optimal". La normalité asymptotique de la distribution a posteriori a été énoncée par le théorème 2.1 dans [2]. Nous avons conclu cette contribution par une application du modèle proposé, sur des données réelles issues du Package MASS disponible dans le langage R.

- La deuxième contribution proposée dans cette thèse, porte sur l'utilisation d'un modèle composite ou hybride, dit *Lognormal-Pareto Composite model* (LPC), issu d'un mélange d'une loi lognormale et une loi de Pareto (de type II). Nous avons proposé une adaptation du modèle LPC au cas d'un échantillon *i.i.d* en présence de censure aléatoire à droite. Afin de tester la validité des procédures d'identification du modèle proposé, nous avons discuté l'adéquation du modèle LPC, sur un jeu de données simulées puis en l'appliquant des données financières, tout en testant statistiquement, la validité d'un tel ajustement à ce modèle.

Ainsi, nous envisageons dans un futur travail, d'étendre les méthodologies présentées dans cette thèse à des modèles plus généraux, en intégrant tout d'abord la possibilité d'extrapoler les résultats obtenus au cas de la TVE multidimensionnelle, ou encore par l'introduction des covariables afin d'observer l'influence d'une telle dépendance, non seulement sur l'estimation de l'indice de queue  $\alpha$ , mais aussi l'effet sur l'estimation des quantiles extrêmes. Ces extensions trouvent leurs intérêts dans différents domaines d'application, notamment pour la conception d'un modèle de tarification actuariel en assurance, où la mesure des multi-risques extrêmes est plus que nécessaire afin d'élaborer un juste tarif. D'autre part, nous envisagerons aussi de développer le modèle composite présenté dans le quatrième chapitre, en établissant les propriétés asymptotiques des estimateurs proposés et d'étendre le modèle à d'autres formes de composition au sens de [Calderín-Ojeda et al. \(2015\)](#).

# Bibliographie

- [1] Andersen P.K., Borgan, O., Gill, R.D. and Keiding N. (1993). Statistical Models Based on Counting Processes. Springer-Verlag, New York
- [2] Amaraoui, A., Boukhetala, K. and Dupuy, J.F., (2016). Bayesian estimation of the tail index of a heavy tailed distribution under random censoring. *Computational Statistics & Data Analysis* 104C, 148-168. <http://dx.doi.org/10.1016/j.csda.2016.06.009>
- [3] Arnold B.C., Balakrishnan N. and Nagaraja H.N. (2008). A First Course in Order Statistics (Classics in Applied Mathematics). SIAM-Society for Industrial and Applied Mathematics. ISBN : 978-0898716481.
- [4] Asimit A.V. and Li J. (2016). Extremes for coherent risk measures. *Insurance : Mathematics and Economics*, 71, p. 332-341. <http://dx.doi.org/10.1016/j.insmatheco.2016.10.003>
- [5] Beirlant, J., Vynckier, P., and Teugels, J. (1996a). Excess functions and estimation of the extreme-value index. *Bernoulli* (2), p. 293-318.
- [6] Beirlant, J., Teugels, J., and Vynckier, P. (1996b). Tail index estimation, Pareto quantile plots regression. *Journ. of Americ. Statisti.* (91), p. 1659-1667.
- [7] Beirlant, J., Dierckx, G., Goegebeur, Y and Matthys, G. (1999). Tail index estimation and exponential regression model. *Journ. of Computational Appl. Math.* (108), p. 99-116.
- [8] Beirlant J., Goegebeur Y., Segers J., Teugels J., De Waal D., Ferro C. (2004). Statistics of Extremes. John Wiley & Sons Ltd, Chichester. ISBN : 9780471976479. <http://dx.doi.org/10.1002/0470012382>
- [9] Beirlant J., Guillou A., Dierckx G., Fils-Villetard A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* 10(3), p. 151-174.
- [10] Beirlant J., Guillou A., Toulemonde G. (2010). Peaks-over-threshold modeling under random censoring. *Communications in Statistics-Theory and Methods* 39(7), p. 1158-1179.
- [11] Blanchet J., Ceresetti D., Molini G., Creutin J.D. (2016). A regional GEV scale-invariant framework for Intensity-Duration-Frequency analysis. *Journal of Hydrology*, 540, p. 82-95. <http://dx.doi.org/10.1016/j.jhydrol.2016.06.007>
- [12] Brahim B., Meraghi D. and Necir A. (2013). Gaussian approximation to the extreme value index estimator of a heavy-tailed distribution under random censoring. *Mathematical Methods of Statistics* 24(4), p. 266-279.

- 
- [13] Cabras, S. and Castellanos, M.E. (2011). A Bayesian approach for estimating extreme quantiles under a semiparametric mixture model. *Astin Bulletin* **41**(1), p. 87-106.
- [14] Caeiro F., Gomes M.I. and Pestana D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3**(2), p. 113-136.
- [15] Calderín-Ojeda, E. and Kwok, C.F. (2015). Modeling claims data with composite Stoppa models. *Scandinavian Actuarial Journal* 2015(0), 1-20. <http://dx.doi.org/10.1080/03461238.2015.1034763>
- [16] Chow, Y.S. and Teicher, H. (1997). Probability theory. Independence, Interchangeability, Martingales (3rd. Edition). *Springer*, New York.
- [17] Coles, S. and Powel, E. (1996). Bayesian methods in extreme value modelling : a review and new developments. *International Statistical Review* **64**(1), p. 119-136.
- [18] Coles, S.G. (2001). An introduction to statistical modeling of extreme values. Springer-Verlag, London.
- [19] Cooray, K. and Ananda, M.A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal* 5(1), 321-334. <http://www.tandfonline.com/doi/abs/10.1080/03461230510009763>
- [20] Cooray, K. and Cheng, C.I. (2015). Bayesian estimators of the lognormal-Pareto composite distribution. *Scandinavian Actuarial Journal* 2015(6), 500-515. <http://dx.doi.org/10.1080/03461238.2013.853368>
- [21] Csörgő, S., Deheuvels, P., and Mason, D.M. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* (**13**), p. 1050-1077.
- [22] D'Agostino, R.B. and Stephens, M.A. (1986). Goodness-of-fit techniques. New York : Marcel Dekker. ISBN :0-824-77487-6. <http://dx.doi.org/10.1002/qre.4680030121>
- [23] Davison, A.C. (1984). Modelling excess over high thresholds. *Statistical Extremes and Applications (J. Tiago de Oliveira, Ed.)*, p. 461-482, D. Reidel, Dordrecht.
- [24] Deheuvels, P., Haeusler, E. and Mason, M. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society* **104**(2), p. 371-381.
- [25] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L., (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* **17**(4), p. 1833-1855.
- [26] Dekker, A.L.M and de Haan, L. (1993). Optimal choice of fraction in extreme-value estimation. *Journal of Multivariate Analysis* (**47**), p. 173-195.
- [27] Dewan, I. and Prakasa, Rao B.L.S. (1997). Remarks on the strong law of large numbers for a triangular array of associated random variables. *Metrika* **45**(3), p. 225-234.
- [28] Diebolt, J., El-Aroui, M., Garrido, M. and Girard S. (2005). Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling. *Extremes* **8**(1-2), p. 57-78.
- [29] Einmahl, J.H.J., Fils-Villetard, A. and Guillou A. (2008). Statistics of extremes under random censoring. *Bernoulli* **14**(1), p. 207-227.
- [30] Embrechts, P., Kluppelberg, C. and Mikosch T. (1997). Modelling Extremal Events. Springer-Verlag, Berlin.

- 
- [31] Ergashev, B., Mittnik, S. and Sekeris E. (2013). A Bayesian approach to extreme value estimation in operational risk modeling. *Journal of Operational Risk* **8**(4), p. 55-81.
- [32] Fraga Alves M.I., Gomes M.I., de Haan L. and Neves C. (2009). Mixed moment estimator and location invariant alternatives. *Extremes* **12**(2), p. 149-185.
- [33] Ghosh, J.K., Delampady, M. and Samanta, T. (2006). An Introduction to Bayesian Analysis. Springer, New York.
- [34] Gomes, M.I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of second order parameter in statistics of extremes. *Extremes* **5**(4), p. 387-414.
- [35] Gomes, M.I., Neves, M.M. (2011). Estimation of the extreme value index for randomly censored data. *Biometrical Letters* **48**(1), p. 1-22.
- [36] Gomes M.I., Oliveira O., (2003). Censoring estimators of a positive tail index. *Statistics & Probability Letters* **65**(3), p. 147-159.
- [37] Groenboom, P., Lopuha, H.P. and de Wolf, P.P., (2003). Kernel-type estimators for the extreme-value index. *Ann. Statisti.* (**31**), p. 1956-1996.
- [38] de Haan, L., (1984). Slow variation and characterization of domains of attraction. *Statistical Extremes and Applications (J. Tiago de Oliveira, Ed.)*, p. 31-48, D. Reidel, Dordrecht.
- [39] de Haan L. and Rootzén H. (1993). On the estimation of high quantiles. *J. Statist. Plann. & Infer.* **35**, p. 1-13.
- [40] de Haan L., Ferreira A., (2006). Extreme Value Theory. Springer, New York.
- [41] Hall, P., (1982). On Some simple estimates of an exponent of regular variation. *J. Ro. Statist. Soc. Ser. B*, (**44**) p. 37-42.
- [42] Häusler, E. and Teugels, J.L., (1985). On asymptotic normality of the Hills estimator for the exponent of regular variation. *Ann. Statisti.* (**13**), p. 743-756.
- [43] Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* (**57**), p. 97-109.
- [44] Hill B.M., (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**(5), p. 1163-1174.
- [45] Howlader H., Hossain A., (2002) Bayesian survival estimation of Pareto distribution of the second kind based on failure-censored data. *Computational Statistics & Data Analysis*, **38**, p. 301-314.
- [46] Jeffreys H., (1961). Theory of Probability. *Oxford University Press*.
- [47] Katz RW, (1999). Extreme value theory for precipitation : sensitivity analysis for climate change. *Advances in Water Resources*,**23**(2), p. 133-139. [http://dx.doi.org/10.1016/S0309-1708\(99\)00017-2](http://dx.doi.org/10.1016/S0309-1708(99)00017-2)
- [48] Katz R.W, Parlange M.B., and Naveau P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*,**25**(8), p. 1287-1304. [http://dx.doi.org/10.1016/S0309-1708\(02\)00056-8](http://dx.doi.org/10.1016/S0309-1708(02)00056-8)
- [49] Kleinbaum D. G. and Klein M. (2012). Survival Analysis, Third Edition . Springer-Verlag New York. ISBN : 978-1-4419-6646-9. <http://dx.doi.org/10.1007/978-1-4419-6646-9>

- 
- [50] Klugman, S.A., Panjer, H.H. and Willmot, G.E., (2012). Loss models : from data to decisions (3rd Ed.). New York, John Wiley & Sons Ltd. ISBN : 9780470187814. <http://dx.doi.org/10.1002/9780470391341>
- [51] Koning A.J., Peng L., (2008). Goodness-of-fit tests for a heavy tailed distribution. *Journal of Statistical Planning and Inference* 138(12), 3960-3981. <http://dx.doi.org/10.1016/j.jspi.2008.02.013>
- [52] Lia Y., Tangb N., and Jiange X. (2016) Bayesian approaches for analyzing earthquake catastrophic risk *Insurance : Mathematics and Economics*,**68**, p. 110-119. <http://dx.doi.org/10.1016/j.insmatheco.2016.02.004>
- [53] Liang Z.M., Li B.Q., Yu Z.B., Chang W.J., (2011). Application of Bayesian approach to hydrological frequency analysis. *Science China Technological Sciences* **54**(5), p. 1183-1192.
- [54] Martinussen T. and Scheike T.H. (2006). *Dynamic Regression Models for Survival Data* . Springer-Verlag New York. ISBN : 978-0-387-33960-3. <http://dx.doi.org/10.1007/0-387-33960-4>
- [55] Mason D.M., (1982). Laws of large numbers for sums of extreme values. *Annals of Probability* **10**(3), p. 754-764.
- [56] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., and Teller E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* (**21**), p. 1087-1092.
- [57] Nadarajah, S. and Bakar, S.A.A, (2012). A new composite model for the Danish fire insurance data. *Scandinavian Actuarial Journal* 2012(1), 1-8. <http://dx.doi.org/10.1080/03461238.2012.695748>
- [58] Naveau P., Nogaj M., Ammann C., Yiou P., Cooley D., and Jomelli V. (2005). Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience*,**337**(10), p. 1013-1022
- [59] do Nascimento F.F., Gamerman D., Lopes H.F., (2012). A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing* **22**, p. 661-675.
- [60] Necir A. and Meraghni D. (2009). Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts. *Insurance Math. Econom.* (**45**), p. 49-58.
- [61] Oliveira, O., Gomes, M.I. and Fraga Alves, M.I., (2006). Improvements in the estimation of heavy tail. *RevStat.* (**4**), p. 81-109.
- [62] Peng, L., (1998). Asymptotically unbiased estimators for the extreme value index. *Statisti. and Probab. Lett.* **38**(2), p. 107-105.
- [63] Pickands J., (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics* **3**(1), p. 119-131.
- [64] Pigeon, M. and Denuit, M., (2011). Composite lognormal-Pareto model with random threshold. *Scandinavian Actuarial Journal* 2011(10), 177-192. <http://dx.doi.org/10.1080/03461231003690754>
- [65] R Development Core Team, (2008). *R : A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [66] Reiss R.D., Thomas M., (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields – 3rd edition*, *Birkhäuser, Basel – Boston*.

- 
- [67] Resnick S.I., (2007). Heavy-Tail Phenomena. Probabilistic and Statistical Modeling. Springer, New York.
- [68] Robert, C. (2006). Le choix bayésien : Principes et pratique (3rd. Edition). *Springer-Verlag*, Paris. ISBN 978-2-287-25173-3.
- [69] Robert, C., and Casella, G. (2011). Méthodes de Monte-Carlo avec R (Pratique R series). *Springer-Verlag*, Paris. ISBN 978-2-8178-0180-3.
- [70] Scollnick, D.P.M., (2007). On composite lognormal-Pareto model. *Scandinavian Actuarial Journal* 2007(7), 20-33. <http://dx.doi.org/10.1080/03461230601110447>
- [71] Shen L., Mickley L. J., and Gilleland E. (2010) Impact of increasing heat waves on U.S. ozone episodes in the 2050s : Results from a multimodel analysis using extreme value theory. *Geophys. Res. Lett.*, **43**, p. 4017-4025, [doi:10.1002/2016GL068432](https://doi.org/10.1002/2016GL068432).
- [72] Sergers, J., (2005). Generalized Pickands estimators for the extreme-value index. *Journal Statist. Plann. Infer.* (**28**), p. 381-396.
- [73] So M.K.P., Chan R.K.S., (2014). Bayesian analysis of tail asymmetry based on a threshold extreme value model. *Computational Statistics and Data Analysis* **71**, p. 568-587.
- [74] Stephenson A., Tawn J., (2004). Bayesian inference for extremes : accounting for the three extremal types. *Extremes* **7**(4), p. 291-307.
- [75] van der Vaart A.W., (1998). Asymptotic Statistics. Cambridge University Press, Cambridge.
- [76] Venables W.N., Ripley B.D., (2002). Modern Applied Statistics with S. Springer, New York.
- [77] Worms J., Worms R., (2014). New estimators of the extreme value index under random right censoring, for heavy tailed distributions. *Extremes* **17**(2), p. 337-358.
- [78] de Zea Bermudez P., Kotz S., (2010). Parameter estimation of the generalized Pareto distribution-Part I. *Journal of Statistical Planning and Inference* **140**(6), p. 1353-1373.
- [79] de Zea Bermudez P., Kotz S., (2010). Parameter estimation of the generalized Pareto distribution-Part II. *Journal of Statistical Planning and Inference* **140**(6), p. 1374-1388.
- [80] Zellner A., (1971). An Introduction to Bayesian Inference in Econometrics. John Wiley & Sons Ltd, New York.
- [81] Zellner A., (1996). Models, prior information, and Bayesian analysis. *Journal of Econometrics* **75**(1), p. 51-68.

