



HAL
open science

Concentration results on extreme value theory

Maud Thomas

► **To cite this version:**

Maud Thomas. Concentration results on extreme value theory . Statistics [math.ST]. Univeristé Paris Diderot Paris 7, 2015. English. NNT: . tel-01177197

HAL Id: tel-01177197

<https://theses.hal.science/tel-01177197>

Submitted on 16 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS DIDEROT (PARIS 7)
SORBONNE PARIS CITE

ÉCOLE DOCTORALE DE SCIENCES MATHÉMATIQUES
DE PARIS CENTRE

Laboratoire de Modèles Aléatoires et de Probabilités - CNRS UMR 7599

THÈSE DE DOCTORAT

Discipline : Mathématiques Appliquées

Présentée par

Maud THOMAS

RÉSULTATS DE CONCENTRATION
EN THÉORIE DES VALEURS EXTRÊMES

CONCENTRATION RESULTS
IN EXTREME VALUE THEORY

Sous la direction de Stéphane BOUCHERON

Rapporteurs :

M. Patrice BERTAIL Université Paris Ouest Nanterre La Défense
M. Holger ROOTZÉN Chalmers University of Technology, Suède

Soutenue publiquement le **2 juillet 2015** devant le jury composé de :

M. Sylvain	ARLOT	CNRS - ENS Paris	Examineur
M. Patrice	BERTAIL	Université Paris Ouest Nanterre La Défense	Rapporteur
M. Stéphane	BOUCHERON	Université Paris Diderot	Directeur
M. Ismael	CASTILLO	CNRS - Université Paris Diderot	Examineur
Mme Fabienne	COMTE	Université Paris Descartes	Examinatrice
Mme Magalie	FROMONT-RENOIR	Université Rennes 2	Examinatrice
M. Pascal	MASSART	Université Paris Sud	Examineur

*À Mamie,
à qui, un jour, son professeur avait donné
100 problèmes de mathématiques,
qu'elle avait faits tous justes.*

Remerciements

Et voilà, nous y sommes ! La dernière étape de ce manuscrit, le moment de remercier (ou du moins essayer) toutes les personnes qui ont, d'une façon ou d'une autre, contribué à cette thèse.

Mes premiers remerciements vont à mon directeur Stéphane Boucheron pour sa disponibilité, pour son soutien à toute épreuve, pour sa patience, pour m'avoir transmis sa passion pour la recherche, et enfin pour m'avoir aidé à faire concentrer les extrêmes.

Je tiens aussi à remercier Patrice Bertail pour avoir accepté de rapporter cette thèse et de représenter le monde des extrêmes dans mon jury.

I would like also to sincerely thank Holger Rootzén for accepting to review my thesis (all the more that the introduction is in French) and I deeply regret that he cannot attend my PhD defense. I think this is also the moment to thank him for the great week I spent in Gothenburg, even though it was pouring rain. I am looking forward to come live in Sweden next year and work with him.

Merci à Sylvain Arlot, Ismael Castillo, Fabienne Comte, Magalie Fromont-Renoir et Pascal Massart pour avoir accepté de faire partie de mon jury. Sylvain, je me souviens encore du jour où je présentais un papier de Pascal et Stéphane afin de valider le cours d'Apprentissage lors de mon Master 2 à Orsay, c'est un grand plaisir de pouvoir te présenter mes travaux de thèse. Ismael, merci pour ces discussions autour d'un café. Magalie, merci pour cette journée à Rennes, pour ton accueil, pour ce partage de ton expérience, pour tes encouragements. Pascal, un grand merci pour m'avoir fait découvrir cette grande théorie qu'est la concentration.

Elisabeth Gassiat et Pascal Massart, cette thèse vous doit beaucoup. Merci pour votre écoute et vos conseils lors des mes études à Orsay. Merci de m'avoir présenté Stéphane, et pour n'avoir jamais été très loin (30min de RER ce n'est pas bien loin...).

Un des travaux de cette thèse est le résultat d'une collaboration avec des épidémiologistes Fabrice Carrat, Magali Lemaitre, Cécile Viboud, Hans Wackernagel, et Mark Wilson. Merci à vous, ce fut une expérience très enrichissante.

Merci à Aurélie Fisher, Sylvain Delattre, Gabrielle Viennet et Raphaël Lefevre dont j'ai assuré les travaux dirigés lors de mon monitorat, j'ai beaucoup apprécié ces années d'enseignement à vos côtés.

Un grand merci au personnel administratif du LPMA, de l'U.F.R. et l'I.E.D. pour leur grande efficacité et leur sympathie : Florence Deschamps, Valérie Juvé, Nathalie Bergame, Pascal Chiettini, Sylvie Le Houëzec et Gaëlle Le Camus.

Ces années furent également l'occasion de rencontres agréables et enrichissantes. Je

tiens donc à remercier l'ensemble du LPMA pour son accueil. Merci à Francis Comets, Mathilde Mougeot, Dominique Picard, Marie Théret, Vianney Perchet, Justin Salez, Mathieu Merle, Sandrine Péch , Nathalie Akakpo, et   ceux qui sont partis St phane Goutte et Camille Male. Je tiens  galement   remercier toutes les personnes que j'ai crois  lors de s minaires ou de conf rences. Merci   Yohann de Castro, Cl ment Dombry, Anne-Laure Foug res, Philippe Naveau, Ivan Nourdin, Patricia Reynaud-Bouret, Mathieu Ribatet, Gilles Stoltz et Olivier Wintenberger, et tous ceux que j'ai pu oublier au moment o  j' cris ces remerciements.

Un merci tout particulier pour Anne Sabourin qui a su  tre pour moi une "grande soeur adoptive" dans ce monde tout particulier qu'est celui des extr mes.

Merci   Guillaume Cebron, Marc-Antoine Giuliani et Bastien Mallein pour avoir  t  les meilleurs co-organisateurs du Groupe de Travail des Th sards du LPMA.

Merci   tous les doctorants du LPMA, Pierre, Jiatsu, Huy, et ceux qui sont d j  grands Oriane, Christophe, Thomas, Nicolas pour ces d jeuners et pauses-caf s qui permettent de d compresser. Je tiens   remercier tout particuli rement tous ceux qui ont  t  bien plus que des coll gues, merci pour votre soutien, pour toutes ces soir es et tous ces d lires : Adrien qui voudrait encore  tre un enfant, Arturo qui me fait danser la salsa, Aser qui donne une touche br silienne   notre langue fran aise, Cl ment qui d tient le record de l'assiette de l gumes la plus grande, Guillaume qui ne se nourrit pas tout   fait comme nous, Lorick l'homme qui  teint les bougies plus vite que son ombre, Marc-Antoine qui fait des chroniques culturelles humoristiques, Noufel le berger de tous les biquets et de toutes les biquettes, Sophie qui est toujours attentive au lieu o  l'on se trouve et en quelle compagnie, Thomas le petit dernier arriv , Vu Lan qui apporte son propre kebab aux soir es, et enfin last but not least ma petite "soeur de th se" Anna, ma concentrateuse, ma mangeuse de chocolat, merci pour ces deux ann es de face- -face.

Merci   la loi des $5(+1)$, merci d' tre toujours l  malgr  les distances, le temps qui passe.

Merci   tout mon petit monde de la Zumba, l'AC Team et ses folies (en particulier Sophie et Vicky),   notre instructor Adrien, Kalidou mon PowerRanger bleu et Pierre   qui j'ai appris que tout  tait relatif. Merci   mes troyennes "Virginettes un jour, Virginette toujours" et en particulier   mon bin me. Merci   ma copine Virginie et toute sa famille pour ces week-ends, repas, r veillons pass s   vos c t s.

Merci   mes grands fr res bacots de me supporter depuis tant d'ann es. Une chose est s re "C'est fort Bois-Le-Roi" #cfblr.

Enfin, cette th se n'aurait pas  t  la m me sans le soutien et les pr cieux conseils de ma famille.

Pour terminer ces remerciements je vous propose un petit jeu : pour faire des maths, il faut souvent un petit coup de baguette magique. Saurez-vous le retrouver dans la bibliographie ?

Résumé

La théorie statistique univariée des valeurs extrêmes repose essentiellement sur des théorèmes asymptotiques rendant délicate l'élaboration de procédures d'estimation adaptatives, voire d'inégalités oracle. Le but de cette thèse est de montrer comment le développement d'inégalités de concentration en théorie des valeurs extrêmes permet d'établir des résultats de type oracle.

Le premier chapitre rappelle les résultats principaux concernant les statistiques d'ordre, la théorie univariée des valeurs extrêmes et la théorie de la concentration. Les trois prochains chapitres réunissent les trois articles écrits au cours de cette thèse. Les chapitres 2 et 3 présentent des travaux en collaboration avec Stéphane Boucheron : le premier propose des inégalités de concentration pour les statistiques d'ordre d'un échantillon de variables aléatoires ; le deuxième s'intéresse au développement d'une version adaptative de l'estimateur de Hill, après avoir établi des bornes non-asymptotiques pour cet estimateur.

Le chapitre 4 est un travail en collaboration avec des épidémiologistes. Il consiste en une application de la théorie des valeurs extrêmes à l'exemple particulier des épidémies de grippe.

Abstract

Since univariate extreme value theory is essentially based on asymptotic theorems, the elaboration of adaptive estimation procedures or oracle inequalities is challenging. The purpose of this thesis is to show how the development of concentration inequalities in extreme value theory allows us to derive oracle type results.

The first chapter gathers the principal results on order statistics, extreme value theory and concentration theory. Chapters 2 and 3 represent a joint work with Stéphane Boucheron: the first chapter presents concentration inequalities for order statistics of a sample of random variables; the second chapter deals with the development of an adaptive version of the Hill estimator based on non asymptotic bounds of this estimator.

Chapter 4 summarises the work done in collaboration with epidemiologists on the application of extreme value theory to the example of influenza epidemics.

Table des matières

1	Introduction	11
1	Contexte et objectifs de la thèse	11
2	Statistiques d'ordre	13
2.1	Lois des statistiques d'ordre	14
2.2	Exemples	15
2.3	Une représentation pour les statistiques d'ordre	16
2.4	Comportement asymptotique	17
3	Théorie des valeurs extrêmes	20
3.1	Lois limites des excès et domaines d'attraction	21
3.2	Estimation de l'indice de valeurs extrêmes γ	28
3.3	Méthode <i>Peaks over threshold</i> et méthode des maxima par blocs	38
4	Inégalités de concentration	41
4.1	Premières inégalités	42
4.2	Majoration de la variance	45
4.3	Inégalité exponentielle d'Efron-Stein	52
4.4	Méthode entropique	52
4.5	Une inégalité maximale	56
2	Concentration inequalities for order statistics	57
1	Introduction	57
2	Order statistics and spacings	59
3	Asymptotic assessment	63
4	Order statistics of Gaussian samples	64
5	Appendix	70
5.1	Proof of Proposition 2.12	70
5.2	Proof of Proposition 2.16	72
3	Tail index estimation	75
1	Introduction	76
2	Background, notations and tools	78
2.1	The Hill estimator as a smooth tail statistics	78
2.2	Frameworks	80
2.3	Lepski's method and adaptive tail index estimation	81
2.4	Lower bound	84

2.5	Talagrand's concentration phenomenon for products of exponential distributions	85
3	Main results	87
3.1	Bounding the variance of the Hill estimator	87
3.2	Concentration inequalities for the Hill estimators	88
3.3	Adaptive Hill estimation	90
4	Proofs	93
4.1	Proof of Proposition 3.2	93
4.2	Proof of Proposition 3.11	93
4.3	Proof of Theorem 3.13	94
4.4	Proof of Theorem 3.17	98
4.5	Proof of Corollary 3.19	100
4.6	Proof of Theorem 3.20	101
4.7	Proof of Corollary 3.23	103
5	Simulations	104
6	Appendix	107
6.1	Calibration of the preliminary selection rule	107
6.2	Proof of Corollary 3.10	108
6.3	Proof of Abelian Proposition 3.12	109
6.4	Proof of Inequality 3.10	112
6.5	Revisiting the lower bound on adaptive estimation error	112
4	An application of Extreme Value Theory	117
1	Background	118
2	Methods	119
2.1	Extreme Value Theory (EVT) - Block Maxima Method	119
2.2	Data sources	119
2.3	Application of EVT	120
3	Results	121
3.1	Descriptive summary	121
3.2	EVT results	122
4	Discussion	123
A	Outils mathématiques	127
1	Inverses généralisées	127
2	Théorème de convergence des types	127
3	Notion de variation régulière	128
B	Méthode de Lepski	131

Chapitre 1

Introduction

1 Contexte et objectifs de la thèse

Dans la nuit du 31 janvier au 1^{er} février 1953, une tempête submergea plusieurs centaines de kilomètres de digues le long de la côte néerlandaise. L'inondation tua 1836 personnes et détruisit 50 000 habitations. A la suite de cette catastrophe, il fut décidé de construire une digue dont la hauteur devait assurer qu'il n'y ait pas plus d'une inondation tous les 10 000 ans. Les données disponibles ne couvrant qu'une période de 100 ans, comment déterminer une hauteur de digue qui ne soit dépassée qu'une fois tous les 10 000 ans ?

En santé publique, une épidémie grippale exceptionnellement sévère peut nécessiter l'hospitalisation de nombreux patients ou la mise à disposition d'un grand nombre d'antiviraux. Comment estimer la probabilité qu'une épidémie plus sévère que toutes les précédentes se produise l'année prochaine ?

Pour le statisticien, déterminer la hauteur de la digue, c'est estimer un quantile. Prendre comme référence la plus haute vague reviendrait à considérer que le pire s'est déjà produit. Dans le cas de la grippe, estimer la probabilité d'une épidémie par sa fréquence observée reviendrait à considérer qu'il ne pourrait y avoir d'épidémie plus sévère que celles observées dans le passé.

Il existe bien d'autres domaines où des questions similaires peuvent se poser : en finance (crack boursier, VaR *Value at Risk*), en climatologie (pluies torrentielles, tempêtes de vents violents) ou encore en assurance (incendies, accidents d'avion).

Née dans la période de l'entre-deux-guerres des travaux de Fisher & Tippett (1928); Fréchet (1927); Gnedenko (1943); Gumbel (1958), la théorie des valeurs extrêmes propose un cadre statistique pour traiter ces deux problèmes : estimer un quantile extrême et estimer la probabilité d'occurrence d'un événement qui n'a pas (encore) été observé.

Intuitivement, les valeurs extrêmes peuvent être vues comme les plus grandes observations d'un échantillon ou comme les observations dépassant un certain seuil. On peut donc s'attendre à ce que leur comportement soit lié à l'épaisseur de la queue de la loi de l'échantillon. Le résultat fondamental, prouvé par Fisher & Tippett (1928) et complété par Gnedenko (1943), identifie les lois limites du maximum de n variables aléatoires. Puis,

en 1970, grâce à l’extension de la notion de variation régulière, de Haan (1970) a établi une condition nécessaire et suffisante de convergence.

La famille de lois limites est indexée par un paramètre de forme γ dont la valeur reflète l’épaisseur de la queue de la loi. Les deux premiers estimateurs de ce paramètre ont été proposés en 1975 par Hill (1975) et Pickands (1975). Puis d’autres estimateurs ont été suggérés, comme l’estimateur du maximum de vraisemblance ou l’estimateur des moments (de Haan & Ferreira, 2006). Drees (1998b) a montré que les estimateurs de γ sont, en général, des fonctions régulières des grandes statistiques d’ordre.

Les statisticiens sont alors confrontés à un dilemme biais-variance : choisir le nombre de grandes statistiques d’ordre à partir desquelles estimer γ . Un nombre trop grand engendre un biais important alors qu’un nombre trop petit donne une grande variance. Le nombre optimal est dicté par des paramètres inconnus du statisticien, et l’élaboration de procédures d’estimation adaptatives, voire d’inégalités oracle, reste délicate.

Depuis les années 1990, des auteurs se sont intéressés à cette question du choix du nombre de statistiques d’ordre, et ont proposé des procédures adaptatives fondées sur des techniques de bootstrap (Beirlant et al., 1996; Danielsson et al., 2001; Hall, 1990), de moyennisation (Resnick & Stărică, 1997) ou de tests successifs (Grama & Spokoiny, 2008).

Au début des années 1990, Lepski (1990, 1991) a développé une méthode générale pour définir des procédures adaptatives de sélection de la largeur de la fenêtre dans les problèmes d’estimation de densité. Drees & Kaufmann (1998) ont été les premiers à utiliser cette méthode en proposant une procédure séquentielle pour sélectionner le nombre de statistiques d’ordre à prendre en compte dans l’estimateur de Hill. Récemment, Carpentier & Kim (2014a) ont construit un estimateur adaptatif à partir de probabilités d’événements plutôt qu’à partir de statistiques d’ordre et ont montré qu’il était possible d’en déduire une version adaptative à l’aide de la méthode de Lepski. L’application de la méthode de Lepski suppose de pouvoir disposer des bornes non asymptotiques pour la queue de probabilité de la loi des estimateurs étudiés, alors que la théorie des valeurs extrêmes repose essentiellement sur le comportement asymptotique des statistiques d’ordre (Reiss, 1989).

Afin de développer des inégalités non asymptotiques pour la queue du processus empirique, représentée par les plus grandes statistiques d’ordre, le recours aux outils de la théorie de la concentration de la mesure, comme la méthode entropique (Boucheron et al., 2013; Ledoux, 2001) initiée par Ledoux au début des années 1990, mérite d’être envisagé. En effet, l’usage averti de l’inégalité de Talagrand permet d’établir des résultats de type oracle dans des situations techniquement désespérées¹. En particulier, la méthode entropique a permis d’établir des versions tendues de l’inégalité de Talagrand pour les suprema de processus empiriques, qui ont été à l’origine de développements importants en théorie de l’apprentissage (Koltchinskii, 2008) et de la sélection de modèle (Massart, 2006).

Suivant cette idée, le premier objectif de cette thèse est d’établir des bornes non asymptotiques de la variance et de la queue de probabilité des statistiques d’ordre d’un

1. "We have been lucky enough to discover that concentration inequalities were indeed the probabilistic tools we needed to develop a nonasymptotic theory for model selection" (Massart, 2006)

échantillon de variables aléatoires. Dans l'article *Concentration inequalities for order statistics* (Boucheron & Thomas, 2012) (Chapitre 2), nous montrons que les outils de la théorie de la concentration, associés à la représentation de Rényi pour les statistiques d'ordre d'un échantillon exponentiel, permettent d'établir des bornes tendues pour la variance et la queue de probabilité des statistiques d'ordre.

En considérant l'estimateur de Hill comme une fonction régulière des k plus grandes statistiques d'ordre, le second objectif de cette thèse est d'obtenir des inégalités de concentration tendues pour cet estimateur, puis d'en déduire une version adaptative grâce à la méthode de Lepski. Nous avons été essentiellement inspirés par le travail de Carpentier & Kim (2014a). Les auteurs proposent une approche basée sur la méthode de Lepski pour sélectionner un estimateur proche de l'estimateur optimal parmi une famille d'estimateurs pour lesquels des inégalités de concentration s'obtiennent facilement. Dans l'article *Tail index estimation, concentration and adaptivity* (Boucheron & Thomas, 2015) (Chapitre 3), sous l'hypothèse que la fonction de répartition sous-jacente appartient au max-domaine d'attraction d'une loi de valeurs extrêmes généralisée, grâce à la représentation de Rényi et à la représentation de Karamata, nous représentons l'estimateur de Hill comme une fonction régulière de variables aléatoires exponentielles indépendantes, conditionnellement à la $(k + 1)^{\text{e}}$ statistique d'ordre d'un échantillon exponentiel. Nous sommes alors en mesure d'appliquer les résultats de concentration de Bobkov & Ledoux (1997) pour des fonctions régulières de vecteurs exponentiels, afin d'établir des bornes non asymptotiques de la variance et de la queue de probabilité de l'estimateur de Hill. Puis, ces inégalités de concentration nous permettent de construire une version adaptative de l'estimateur de Hill, en proposant une procédure du choix du nombre de statistiques d'ordre à prendre en compte grâce à la méthode de Lepski.

En parallèle de ces deux travaux, en collaboration avec des épidémiologistes, nous avons appliqué la théorie des valeurs extrêmes aux données de mortalité et de morbidité de la grippe en France. Nous avons utilisé la méthode classique des maxima par blocs. Le but de cette approche est de montrer dans un cadre statistique simple comment la théorie des valeurs extrêmes peut fournir un outil de prédiction d'épidémies particulièrement sévères dans les prochaines années. Ce travail est présenté dans le chapitre 4.

Les parties suivantes de ce chapitre sont consacrées aux propriétés des statistiques d'ordre, aux éléments de la théorie des valeurs extrêmes et aux inégalités de concentration. Les chapitres suivants sont organisés autour des travaux effectués pendant cette thèse. En annexe figurent des résultats techniques utilisés dans ce manuscrit et le principe général de la méthode de Lepski.

2 Statistiques d'ordre

La théorie des valeurs extrêmes est directement liée à celle des statistiques d'ordre. Ce chapitre rassemble les définitions et les résultats dont nous aurons besoin. Nous faisons référence aux ouvrages (de Haan & Ferreira, 2006; Reiss, 1989; Smirnov, 1952).

2.1 Lois des statistiques d'ordre

Soit $\mathbf{X}^n = (X_1, \dots, X_n)$ un n -échantillon de variables aléatoires réelles indépendantes et identiquement distribuées (i.i.d.).

Définition 1.1 (STATISTIQUES D'ORDRE). *Pour $1 \leq k \leq n$, la statistique d'ordre de rang k , notée $X_{k,n}$ (ou $X_{(k)}$ s'il n'y a pas d'ambiguïté), est la k^e plus grande valeur de \mathbf{X}^n .*

Soulignons que, selon cette définition, les statistiques d'ordre sont rangées par ordre décroissant $X_{1,n} \geq \dots \geq X_{n,n}$.

Soit F la fonction de répartition commune des variables X_1, \dots, X_n . Nous supposons dans la suite que F est continue et définie sur un intervalle I de \mathbb{R} , ce qui garantit que les statistiques d'ordre sont presque sûrement distinctes deux à deux.

La loi de chaque statistique d'ordre est connue explicitement.

Proposition 1.2 (LOI DES STATISTIQUES D'ORDRE). *Pour tout $1 \leq k \leq n$, la fonction de répartition $F_{(k)}$ de $X_{(k)}$ a pour expression*

$$F_{(k)}(t) = \mathbb{P}\{X_{(k)} \leq t\} = \sum_{i=n-k+1}^n \binom{n}{i} F^i(t)(1-F(t))^{n-i}, \quad t \in I.$$

De plus, si F admet une densité f par rapport à la mesure de Lebesgue, alors $X_{(k)}$ admet une densité $f_{(k)}$ dont l'expression est donnée par

$$f_{(k)}(x) = (n-k+1) \binom{n}{n-k+1} f(x) F^{n-k}(x) (1-F(x))^{k-1}, \quad x \in I.$$

Les statistiques d'ordre ne sont évidemment pas indépendantes entre elles, mais il est possible de calculer leur loi jointe.

Proposition 1.3 (LOI JOINTE DES STATISTIQUES D'ORDRE). *Si F admet une densité par rapport à la mesure de Lebesgue, la densité jointe $f_{1:n}$ de $(X_{(1)}, \dots, X_{(n)})$ est donnée par*

$$f_{1:n}(x_1, \dots, x_n) = n! \prod_{i=1}^n f(x_i) \mathbb{1}_{\{x_1 < \dots < x_n\}}, \quad (x_1, \dots, x_n) \in I^n.$$

Proposition 1.4 (EXISTENCE DES MOMENTS). *Soit $g: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction mesurable. Si $\mathbb{E}[|g(X_1, \dots, X_n)|] < \infty$, alors $\mathbb{E}[|g(X_{(1)}, \dots, X_{(n)})|] < \infty$.*

Le moment d'ordre q de la k^e statistique d'ordre a pour expression

$$\mathbb{E}[X_{(k)}^q] = (n-k+1) \binom{n}{n-k+1} \int_I x^q f(x) F^{n-k}(x) (1-F(x))^{k-1} dx.$$

On voit qu'en général, le calcul des moments d'ordre q nécessite des intégrations numériques. Par exemple, le moment d'ordre 2 du maximum de n variables aléatoires gaussiennes centrées réduites vaut

$$\mathbb{E}[X_{(1)}^2] = \frac{n}{(2\pi)^n} \int_{\mathbb{R}} x^2 e^{-x^2/2} \left(\int_{-\infty}^x e^{y^2/2} dy \right)^{n-1} dx.$$

Définition 1.5 (ESPACEMENTS ENTRE STATISTIQUES D'ORDRE). *On définit les espacements $(\Delta_1^X, \dots, \Delta_n^X)$ (ou $(\Delta_1, \dots, \Delta_n)$ s'il n'y a pas d'ambiguïté) entre les statistiques d'ordre $X_{(1)} \geq \dots \geq X_{(n)}$ par*

$$(\Delta_1^X, \dots, \Delta_{n-1}^X, \Delta_n^X) = (X_{(1)} - X_{(2)}, \dots, X_{(n-1)} - X_{(n)}, X_{(n)}).$$

2.2 Exemples

Nous noterons $X \stackrel{d}{=} Y$ si les deux variables aléatoires X et Y sont distribuées selon la même loi.

Variabes aléatoires uniformes

Soit V_1, V_2, \dots, V_n n variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$ et $V_{(1)} \geq \dots \geq V_{(n)}$ les statistiques d'ordre associées.

Proposition 1.6. *Pour tout $1 \leq k \leq n$, la k^e statistique d'ordre $V_{(k)}$ est distribuée selon une loi beta de Dirichlet de paramètres $n - k + 1$ et k dont la densité est donnée par*

$$x \mapsto \frac{\Gamma(n+1)}{\Gamma(n-k+1)\Gamma(k)} x^{n-k} (1-x)^{k-1} \mathbb{1}_{[0,1]}(x),$$

où Γ est la fonction gamma d'Euler définie par $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

De plus,

$$\mathbb{E}[V_{(k)}] = 1 - \frac{k}{n+1} \quad \text{et} \quad \text{Var}[V_{(k)}] = \frac{k}{(n+1)(n+2)} \left(1 - \frac{k}{n+1}\right).$$

Proposition 1.7. *Les espacements $\Delta_1^V, \dots, \Delta_n^V$ sont des variables aléatoires échangeables, c'est-à-dire que pour toute permutation σ de $\{1, \dots, n\}$,*

$$(\Delta_{\sigma(1)}^V, \dots, \Delta_{\sigma(n)}^V) \stackrel{d}{=} (\Delta_1^V, \dots, \Delta_n^V).$$

Ainsi, pour tout $1 \leq k \leq n$,

$$\Delta_k^V \stackrel{d}{=} V_{(n)},$$

et

$$\mathbb{E}[\Delta_k^V] = \frac{1}{n+1} \quad \text{et} \quad \text{Var}[\Delta_k^V] = \frac{n}{(n+1)^2(n+2)}.$$

Variabes aléatoires exponentielles

Soit Y_1, \dots, Y_n n variables aléatoires i.i.d. de loi exponentielle de paramètre 1, et $Y_{(1)} \geq \dots \geq Y_{(n)}$ les statistiques d'ordre associées.

La représentation de Rényi affirme que les statistiques d'ordre d'un échantillon exponentiel de paramètre 1 sont distribuées comme des sommes partielles de variables aléatoires exponentielles indépendantes.

Théorème 1.8 (REPRÉSENTATION DE RÉNYI (Rényi, 1953)).

$$(Y_{(n)}, \dots, Y_{(k)}, \dots, Y_{(1)}) \stackrel{d}{=} \left(\frac{E_n}{n}, \dots, \sum_{i=k}^n \frac{E_i}{i}, \dots, \sum_{i=1}^n \frac{E_i}{i} \right)$$

où E_1, \dots, E_n sont des variables aléatoires exponentielles indépendantes de paramètre 1.

Les espacements sont donc distribués comme des variables aléatoires exponentielles de paramètre 1 renormalisées :

$$\Delta_k^Y = Y_{(k)} - Y_{(k+1)} \stackrel{d}{=} \frac{E_k}{k}.$$

Pour tout $1 \leq k \leq n$, $\mathbb{E}[Y_{(k)}] = \sum_{i=k}^n 1/i$, d'où

$$\ln\left(\frac{n}{k}\right) + \frac{1}{n} \leq \mathbb{E}[Y_{(k)}] \leq \ln\left(\frac{n}{k}\right) + \frac{1}{k},$$

et $\text{Var}[Y_{(k)}] = \sum_{i=k}^n 1/i^2$, d'où

$$\frac{1}{k} + \frac{1}{n^2} \leq \text{Var}[Y_{(k)}] \leq \frac{1}{k} + \frac{1}{k^2}. \quad (1.1)$$

2.3 Une représentation pour les statistiques d'ordre

Définition 1.9 (FONCTION DE SURVIE). *La fonction de survie associée à F est la fonction $\bar{F} = 1 - F$.*

Définition 1.10 (INVERSE GÉNÉRALISÉE). *Soit g est une fonction monotone d'un intervalle fini ou infini $]a, b[$ dans un intervalle fini ou infini $]c, d[$. Son inverse généralisée à gauche $g^{\leftarrow} :]c, d[\rightarrow]a, b[$ est définie par $g^{\leftarrow}(y) = \inf\{x : a < x < b, g(x) \geq y\}$.*

Pour plus de détails, le lecteur peut se référer à la partie 1 de l'annexe A.

Définition 1.11 (U -TRANSFORMÉE). *La U -transformée de F est la fonction définie pour tout $t > 0$ par*

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

$U(t)$ est donc le quantile d'ordre $1 - 1/t$ de F .

Contrairement aux statistiques d'ordre d'un échantillon exponentiel, les statistiques d'ordre d'un échantillon quelconque ne sont pas en général distribuées comme des sommes de variables aléatoires indépendantes. Néanmoins, il est possible de représenter les statistiques d'ordre d'un échantillon quelconque comme une transformation monotone des statistiques d'ordre d'un échantillon exponentiel.

Dans toute la suite, $Y_{(1)} \geq \dots \geq Y_{(n)}$ désigneront les statistiques d'ordre de n variables aléatoires indépendantes exponentielles de paramètre 1, et E_1, \dots, E_n des variables aléatoires exponentielles de paramètre 1 indépendantes entre elles et indépendantes des $Y_{(i)}$, $i = 1, \dots, n$.

Proposition 1.12 (REPRÉSENTATION DES STATISTIQUES D'ORDRE).

$$(X_{(1)} \dots, X_{(n)}) \stackrel{d}{=} (U(e^{Y_{(1)}}), \dots, U(e^{Y_{(n)}})) .$$

Démonstration. Soit X une variable aléatoire de fonction de répartition F et Y une variable aléatoire exponentielle de paramètre 1. Alors $X \stackrel{d}{=} U(e^Y)$. De plus, la fonction $U \circ \exp$ est croissante, donc l'ordre est conservé. \square

Il est également possible de représenter les statistiques d'ordre d'un échantillon quelconque à l'aide des statistiques d'ordre d'un échantillon uniforme par la transformation quantile F^{\leftarrow} .

$$(X_{(1)} \dots, X_{(n)}) \stackrel{d}{=} (F^{\leftarrow}(V_{(1)}), \dots, F^{\leftarrow}(V_{(n)})) .$$

Afin d'établir des inégalités de concentration pour les statistiques d'ordre, nous utiliserons la proposition 1.12, car les sommes de variables aléatoires indépendantes se prêtent bien à l'application des inégalités de concentration.

2.4 Comportement asymptotique

Pour une suite de fonctions de répartition F_n et une fonction de répartition F , nous noterons $F_n \xrightarrow{d} F$ si

$$F_n(x) \xrightarrow[n \rightarrow \infty]{} F(x)$$

en tout point de continuité x de F , et $X_n \xrightarrow{d} X$ si X_n tend en loi vers X .

Plusieurs travaux (Fisher & Tippett, 1928; Fréchet, 1927; von Mises, 1936) sont consacrés à la recherche des lois limites des plus grandes (ou des plus petites) statistiques d'ordre. Gnedenko (1943) a démontré que la loi limite du maximum d'une suite de variables aléatoires i.i.d. (si cette limite existe) appartient à une famille de lois à un seul paramètre. Puis, Smirnov (1952) a étudié, plus généralement, le comportement asymptotique des statistiques d'ordre.

Le comportement asymptotique des statistiques d'ordre dépend de leur nature : central, extrême ou intermédiaire.

Définition 1.13 (NATURE DES STATISTIQUES D'ORDRE). *Soit (k_n) une suite d'indices dépendant de n . On distingue alors trois cas suivant le comportement de k_n lorsque $n \rightarrow \infty$*

- a) *si $k_n \rightarrow \infty$ et $k_n/n \rightarrow q \in]0, 1[$, on dit que $X_{k_n, n}$ est une statistique d'ordre central*
- b) *si k_n ou $n - k_n$ est fixé, on dit que $X_{k_n, n}$ est une statistique d'ordre extrême*
- c) *si $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$, on dit que $X_{k_n, n}$ est une statistique d'ordre intermédiaire.*

Définition 1.14. *Lorsque $k_n \xrightarrow[n \rightarrow \infty]{} \infty$ et $k_n/n \xrightarrow[n \rightarrow \infty]{} 0$, on dit que la suite (k_n) est intermédiaire.*

Statistiques d'ordre central (Reiss, 1989)

Laplace (1818) a prouvé la normalité asymptotique de la médiane empirique d'un échantillon en calculant sa densité et en démontrant la convergence presque sûre des

densités. Smirnov (1935) a utilisé la même méthode pour obtenir la normalité asymptotique des statistiques d'ordre central en général. Puis, d'autres approches ont réduit le problème à une application du théorème central limite (Cramér, 1946; Rényi, 1953; Sen, 1968; Smirnov, 1952; van der Vaart, 1961).

Les statistiques d'ordre central peuvent être utilisées comme estimateurs des quantiles de la loi F , d'où l'intérêt de l'étude de leur comportement asymptotique.

Smirnov a montré que, pour tout réel $q \in]0, 1[$, l'existence de deux suites $a_n > 0$ et b_n et d'une fonction limite τ , dépendant de q , telles que

$$\sqrt{n} \frac{F(a_n x + b_n) - q}{\sqrt{q(1-q)}} \longrightarrow \tau(x)$$

était équivalente à la relation asymptotique

$$F_{(k_n)}(a_n x + b_n) \xrightarrow{d} \Phi(\tau(x)),$$

où Φ est la fonction de répartition de la loi $\mathcal{N}(0, 1)$. Pour assurer l'unicité de la limite, il faut imposer une condition du second ordre sur la suite (k_n) , $\sqrt{n}(k_n/n - q) \xrightarrow[n \rightarrow \infty]{} 0$.

Cette équivalence signifie que $X_{k_n, n}$ converge en loi après renormalisation si la fonction de répartition sous-jacente F converge après renormalisation vers une fonction τ . Ainsi, la connaissance de toutes les fonctions limites possibles τ permet de connaître toutes les lois limites $\Phi \circ \tau$ de $X_{k_n, n}$. Par le théorème de convergence des types (Théorème A.4), pour tout n , il existe deux suites $\alpha_n > 0$ et β_n telles que les fonctions τ possibles sont solutions de l'équation fonctionnelle

$$\tau(x) = \sqrt{n} \tau(\alpha_n x + \beta_n), \alpha_n > 0, \beta_n \in \mathbb{R}.$$

Théorème 1.15 (Reiss (1989); Smirnov (1952)). *Soit (k_n) une suite telle que $\sqrt{n}(k_n/n - q) \xrightarrow[n \rightarrow \infty]{} 0$ pour un certain $0 < q < 1$. Les lois limites non dégénérées possibles pour une suite $X_{k_n, n}$ de statistiques d'ordre central ont pour fonction de répartition*

$$\begin{aligned} H_{1,\alpha}(x) &= \begin{cases} 0 & \text{si } x < 0 \\ \Phi(x^\alpha) & \text{si } x \geq 0, \end{cases} \\ H_{2,\alpha}(x) &= \begin{cases} \Phi(-(-x)^\alpha) & \text{si } x < 0 \\ 1 & \text{si } x \geq 0, \end{cases} \\ H_{3,\alpha,\sigma}(x) &= H_{1,\alpha}(x/\sigma) \mathbb{1}_{[0,\infty[}(x) + H_{2,\alpha}(x/\sigma) \mathbb{1}_{]-\infty,0]}(x), \\ H_4(x) &= \frac{1}{2} \mathbb{1}_{]-1,1]}(x) + \mathbb{1}_{[1,\infty[}(x) \end{aligned}$$

où α, σ sont des paramètres strictement positifs.

$H_{1,\alpha}$ et $H_{2,\alpha}$ sont continues à droite avec un saut d'amplitude $1/2$ en 0 , $H_{3,\alpha,\sigma}$ est continue et H_4 vaut 0 sur $] - \infty, -1[$, $1/2$ sur $[-1, 1[$ et 1 sur $[1, \infty[$.

Statistiques d'ordre extrême (Arnold et al., 2008; de Haan & Ferreira, 2006; Reiss, 1989)

Fisher & Tippett (1928) ont établi une liste complète de toutes les lois limites possibles pour le maximum d'un échantillon. Gnedenko (1943) a obtenu des conditions nécessaires et suffisantes pour la convergence du maximum vers ces lois limites. Puis, de Haan (1970) a spécifié la caractérisation de Gnedenko.

S'il existe deux suites $a_n > 0$ et b_n telles que $(X_{1,n} - b_n)/a_n$ converge vers une loi non dégénérée alors la fonction de répartition de cette loi limite, définie sur \mathbb{R} , est nécessairement du type

$$G_\gamma(x) = \begin{cases} \exp\left(-\left(1 + \gamma x\right)_+^{-1/\gamma}\right) & \text{pour } \gamma \neq 0 \\ \exp(-e^{-x}) & \text{pour } \gamma = 0 \end{cases} \quad (1.2)$$

où γ est un paramètre réel.

Nous traiterons plus en détail le comportement asymptotique du maximum dans la partie 3.

Smirnov (1952) a étendu ce résultat aux statistiques d'ordre $X_{k,n}$ avec k fixé.

Théorème 1.16 (LOI ASYMPTOTIQUE DES AUTRES STATISTIQUES D'ORDRE EXTRÊME). *Il existe des suites $a_n > 0$ et b_n telles que $F^n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{d} G_\gamma$ si, et seulement si, pour tout $k > 1$,*

$$F_{(k)}(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{d} G_\gamma(x) \sum_{i=0}^{k-1} \frac{(-\ln G_\gamma(x))^i}{i!}.$$

Les constantes a_n et b_n sont les mêmes que celles intervenant dans la convergence du maximum.

Statistiques d'ordre intermédiaire (de Haan & Ferreira, 2006; Smirnov, 1952)

Chibisov (1964) a énoncé, pour des suites (k_n) particulières, des conditions nécessaires et suffisantes sur F pour assurer la convergence de ces statistiques d'ordre intermédiaire et a déterminé toutes les loi limites possibles. Balkema & de Haan (1978a,b) ont montré que pour une classe de fonctions de répartition dense dans l'ensemble des fonctions de répartition, $X_{k_n,n}$ peut avoir n'importe quelle loi limite après renormalisation si k_n est bien choisie. Néanmoins, Smirnov (1967) a démontré que les lois limites possibles non dégénérées sont gaussiennes sauf pour des suites (k_n) ayant un comportement asymptotique bien précis.

Définition 1.17 (CONDITION DE VON MISES). *Si la U -transformée de F est deux fois différentiable et $U'(t) > 0$ pour tout t dans un voisinage de l'infini, alors on dit que U satisfait la condition de von Mises si*

$$\lim_{t \rightarrow \infty} \frac{tU''(t)}{U'(t)} = \gamma - 1.$$

Théorème 1.18. *Supposons que F satisfait la condition de von Mises pour un $\gamma \in \mathbb{R}$. Alors pour toute suite intermédiaire (k_n) ,*

$$\sqrt{k_n} \frac{X_{k_n+1,n} - U\left(\frac{n}{k_n}\right)}{\frac{n}{k_n} U'\left(\frac{n}{k_n}\right)} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

3 Théorie des valeurs extrêmes

Le problème de la statistique des extrêmes peut se formuler ainsi : on dispose de n observations X_1, \dots, X_n i.i.d. de fonction de répartition (inconnue) F . Ces observations sont ré-ordonnées par ordre décroissant $X_{1,n} \geq \dots \geq X_{n,n}$ (ou $X_{(1)} \geq \dots \geq X_{(n)}$). On cherche à répondre à deux questions :

- i) pour $x > X_{(1)}$, voire $x \gg X_{(1)}$, quelle est la probabilité $1 - F(x) := \bar{F}(x)$ qu'une nouvelle observation dépasse x ?
- ii) comment estimer un quantile d'ordre $1 - 1/(xn)$ avec $x \gg 1$ lorsque l'on dispose seulement de n observations ?

Les deux approches les plus classiques sont les suivantes.

Approche non paramétrique

- i) Estimer F par sa version empirique

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}.$$

- ii) Estimer le quantile à l'aide de la fonction quantile empirique $\widehat{F}_n^{\leftarrow}(p)$ qui vaut $X_{(i)}$ pour $p \in](n-i)/n, (n-i+1)/n[$.

Malheureusement,

- i) en estimant F par \widehat{F}_n , on estime la probabilité de dépasser $x (> X_{(1)})$ par 0.
- ii) en estimant $F^{\leftarrow}(1 - \alpha)$ par $X_{(1)}$ pour $\alpha < 1/n$, on affirme que le pire s'est déjà produit.

Approche paramétrique

On suppose que F appartient à un modèle paramétrique $(F_\theta, \theta \in \Theta)$ et on estime θ à l'aide des données X_1, \dots, X_n . Cependant, un bon ajustement sur l'échantillon ne garantit pas une bonne modélisation en dehors de l'intervalle $[X_{(n)}, X_{(1)}]$. De plus, l'inférence dans le modèle peut être très sensible aux valeurs aberrantes, et lorsque l'on s'intéresse aux valeurs extrêmes, il est très délicat d'identifier une valeur aberrante.

Pour contourner ces difficultés, il existe deux points de vue différents en théorie des valeurs extrêmes. La première approche étudie les excès par rapport à un seuil τ préalablement choisi. Nous verrons que la fonction de répartition empirique des excès au-dessus de τ approche la loi conditionnelle d'une variable aléatoire X distribuée selon F

sachant que $X > \tau$. La deuxième approche adopte le point de vue de l'étude des lois limites du maximum d'un échantillon. Ces deux approches se traduiront par des problèmes de type semi-paramétrique dans le domaine des lois à queue épaisse.

Nous avons choisi d'introduire la théorie des valeurs extrêmes à travers l'approche par les excès, car les problèmes de valeurs extrêmes se traduisent plus naturellement en termes de dépassement de seuil : une valeur est dite extrême si elle dépasse un certain seuil fixé.

3.1 Lois limites des excès et domaines d'attraction

Le but de cette partie est de présenter les lois limites possibles des excès d'une variable aléatoire rapport à un seuil élevé et de caractériser leur domaine d'attraction. Nous verrons qu'il existe une équivalence avec les lois limites des maxima et leur domaine d'attraction.

Lois limites des excès et lois de Pareto généralisées

Pour toute fonction de répartition F , on notera $x^* = \sup\{x : F(x) < 1\}$ son extrémité droite.

Soit X une variable aléatoire de fonction de répartition F . Pour un seuil $\tau < x^*$ fixé, l'excès de X au-dessus de τ est la variable aléatoire $Y_\tau = X - \tau$, définie sur l'événement $\{X > \tau\}$.

Définition 1.19 (LOI D'EXCÈS). *Soit $\tau < x^*$. La loi d'excès de X au-dessus du seuil τ est la loi de Y_τ , c'est-à-dire la loi dont la fonction de survie \bar{F}_τ , est définie sur $]0, \infty[$ par*

$$\bar{F}_\tau(x) = \frac{\bar{F}(\tau + x)}{\bar{F}(\tau)}.$$

Afin d'obtenir des limites en loi non dégénérées, on s'autorise une renormalisation affine des excès, en considérant la variable $(Y_\tau - \mu_\tau)/\sigma_\tau$. S'il existe une fonction de répartition non dégénérée Q telle que

$$\bar{F}_\tau(\mu_\tau + x\sigma_\tau) \xrightarrow[\tau \rightarrow x^*]{d} \bar{Q}(x), \quad (1.3)$$

on dit que F appartient au domaine d'attraction des excès de Q . Grâce au théorème de convergence des types de Khintchine (Théorème A.4), (1.3) implique que Q est stable par seuillage, c'est-à-dire que, pour tout τ tel que $0 < Q(\tau) < 1$, il existe deux constantes $a_\tau > 0$ et b_τ telles qu'en tout point de continuité x de Q ,

$$Q_\tau(a_\tau x + b_\tau) = Q(x)$$

où Q_τ est la fonction de répartition de la loi d'excès associée à Q . Réciproquement, si Q est stable par seuillage, elle appartient à son propre domaine d'attraction des excès (en prenant $\mu_\tau = a_\tau$ et $\sigma_\tau = b_\tau$ dans (1.3)). Donc, la famille des lois limites possibles dans (1.3) coïncide avec la famille de lois stables par seuillage.

La loi exponentielle, les lois de Pareto et la loi uniforme sur $[0, 1]$ sont stables par seuillage.

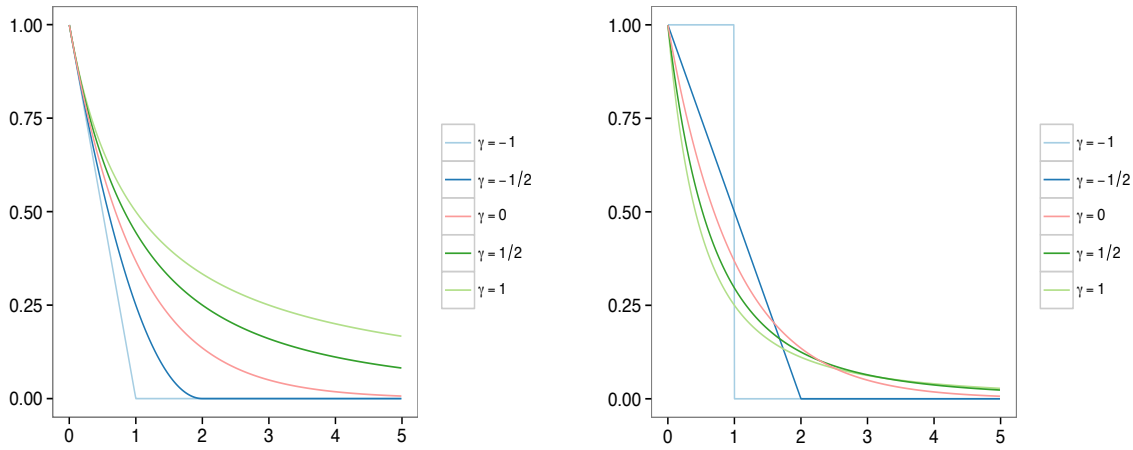


FIGURE 1.1. Fonctions de survie et densités des lois de Pareto généralisées. Les courbes en bleu correspondent à un paramètre de forme $\gamma > 0$, la courbe en rouge à $\gamma = 0$ et les courbes en vert à $\gamma < 0$.

Théorème 1.20 (CARACTÉRISATION DES LOIS LIMITES DES EXCÈS (Balkema & de Haan, 1974)). *S'il existe une fonction de répartition F telle que*

$$\bar{F}_\tau(\mu_\tau + x\sigma_\tau) \xrightarrow[\tau \rightarrow x^*]{d} \bar{Q}(x),$$

alors \bar{Q} est du type

$$\bar{Q}(x) = (1 + \gamma x)_+^{-1/\gamma}, \quad x \geq 0. \quad (1.4)$$

Le paramètre de forme $\gamma \in \mathbb{R}$ est appelé indice de valeurs extrêmes.

Pour $\gamma = 0$, le membre de droite s'interprète comme sa limite lorsque γ tend vers 0, c'est-à-dire e^{-x} .

Ainsi, la famille des lois stables par seuillage est la famille des lois de Pareto généralisées, dont la fonction de répartition est définie par

$$\bar{Q}_{\mu, \sigma, \gamma}(x) = \begin{cases} (1 + \gamma \frac{x-\mu}{\sigma})_+^{-1/\gamma} & \text{pour } x > \mu \\ 1 & \text{pour } x \leq \mu. \end{cases}$$

La figure 1.1 représente les fonctions de survie et les densités des lois de Pareto généralisées de paramètre de forme $\gamma = -1, -1/2, 0, 1/2$ et 1 .

Lien avec les lois limites des maxima

Le théorème de Balkema & de Haan (Théorème 1.20) est une reformulation du résultat fondamental de Fisher & Tippett (1928), complété par Gnedenko (1943), sur les limites possibles du maximum renormalisé d'une suite de variables aléatoires i.i.d.

Théorème 1.21 (Fisher & Tippett (1928); Gnedenko (1943)). *S'il existe une fonction de répartition non dégénérée G et deux constantes $\alpha_n > 0$ et β_n telles que*

$$F^n(\alpha_n x + \beta_n) \xrightarrow[n \rightarrow \infty]{d} G(x) \quad (1.5)$$

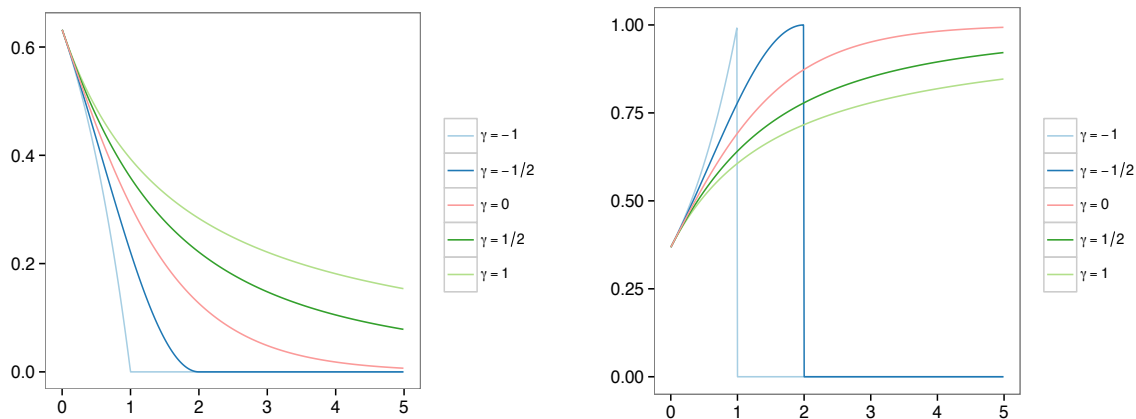


FIGURE 1.2. Fonctions de survie et densités des lois de valeurs extrêmes généralisées. Les courbes en bleu correspondent à un paramètre de forme $\gamma > 0$, la courbe en rouge à $\gamma = 0$ et les courbes en vert à $\gamma < 0$.

alors G est du type

$$G_\gamma(x) = \exp\left(-\left(1 + \gamma x\right)_+^{-1/\gamma}\right), \quad x \in \mathbb{R} \quad (1.6)$$

où le membre de droite s'interprète, pour $\gamma = 0$, comme sa limite lorsque γ vers 0, soit $G_0(x) = \exp(-e^{-x})$.

Si F vérifie une condition du type (1.5), nous dirons que F appartient au max-domaine d'attraction de G_γ .

Comme pour l'étude des lois d'excès, le théorème de convergence des types de Khintchine (Théorème A.4) implique que G est max-stable, c'est-à-dire que pour tout n , il existe des constantes $a_n > 0$ et b_n telles que

$$G^n(a_n x + b_n) = G(x).$$

Si G est max-stable, alors elle appartient à son propre max-domaine d'attraction.

Ainsi, la famille des lois limites possibles pour le maximum de n variables i.i.d. est la famille de lois de valeurs extrêmes généralisées, dont la fonction de répartition est définie sur \mathbb{R} par

$$G_{\mu,\sigma,\gamma}(x) = \exp\left(-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)_+^{-1/\gamma}\right).$$

La figure 1.2 représente les fonctions de répartition et les densités des lois de valeurs extrêmes généralisées pour $\gamma = -1, -1/2, 0, 1/2$ et 1.

Domaines d'attraction

Nous connaissons maintenant toutes les lois limites possibles des lois des excès et du maximum de n variables aléatoires i.i.d. Il est naturel de s'intéresser aux conditions sur la fonction de répartition F de X_1, \dots, X_n pour qu'elle appartienne à un max-domaine d'attraction.

Pour cela, nous aurons besoin des notions de variation régulière et de variation régulière étendue, dont nous rappelons les définitions. Pour plus de détails, nous renvoyons à l'annexe A partie 3.

Définition 1.22 (VARIATION RÉGULIÈRE). Une fonction $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ mesurable, strictement positive au voisinage de l'infini, est dite à variation régulière (à l'infini) d'indice α , ce que l'on note $f \in \text{RV}_\alpha$, si pour tout $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha.$$

Si $\alpha = 0$, on dit que f est à variation lente.

Si $f \in \text{RV}_\alpha$, alors il existe une fonction L à variation lente telle que $f(x) = x^\alpha L(x)$.

La définition suivante propose une généralisation de la variation régulière.

Définition 1.23 (VARIATION RÉGULIÈRE ÉTENDUE). Une fonction $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ mesurable est dite à variation régulière étendue d'indice α , ce que l'on note $f \in \text{ERV}_\alpha$, s'il existe une fonction auxiliaire positive a telle que, pour tout $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{f(tx) - f(t)}{a(t)} = \frac{x^\alpha - 1}{\alpha}.$$

Remarque 1.24. – Une fonction à variation régulière étendue d'indice $\alpha > 0$ est à variation régulière avec le même indice et on peut choisir pour fonction auxiliaire $a(t) = \alpha U(t)$.

– La fonction $[\ln t]$ est à variation lente mais n'est pas à variation régulière étendue.

La condition suivante dite de von Mises, que nous avons déjà vue dans la partie 2, est une condition suffisante d'appartenance à un max-domaine d'attraction, qui sera souvent utilisée dans la suite.

Théorème 1.25 (CONDITION DE VON MISES (de Haan & Ferreira, 2006)). Sous l'hypothèse que $F''(x)$ existe et que $F'(x) > 0$ pour tout x dans un voisinage de x^* , si

$$\lim_{t \rightarrow x^*} \left(\frac{\bar{F}}{F'} \right)'(t) = \gamma \tag{1.7}$$

alors F appartient au max-domaine d'attraction de G_γ .

Corollaire 1.26. La condition (1.7) est équivalente à la condition suivante

$$\lim_{t \rightarrow \infty} \frac{tU''(t)}{U'(t)} = \gamma - 1,$$

ce qui implique

$$\lim_{t \rightarrow \infty} \frac{U'(tx)}{U'(t)} = x^{\gamma-1} \text{ localement uniformément sur }]0, \infty[,$$

soit $U' \in \text{RV}_{\gamma-1}$. Et enfin,

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{tU'(t)} = \frac{x^\gamma - 1}{\gamma},$$

soit $U \in \text{ERV}_\gamma$ avec $a(t) = tU'(t)$.

Ainsi, F appartient au max-domaine d'attraction d'une loi de valeurs extrêmes généralisée G_γ , si U' est à variation régulière d'indice $\gamma - 1$, ou de façon équivalente, si U est à variation régulière étendue d'indice γ .

Pour $\gamma \neq 0$, la condition de von Mises peut s'exprimer de façon simple :

Théorème 1.27. *Lorsque $\gamma > 0$, si F' existe et si*

$$\lim_{t \rightarrow \infty} \frac{tF'(t)}{\overline{F}(t)} = \frac{1}{\gamma} \quad (1.8)$$

alors F appartient au max-domaine d'attraction de G_γ . La condition (1.8) est équivalente à

$$\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(t)} = \gamma$$

ce qui implique $U \in \text{RV}_\gamma$.

On pose $U(\infty) = \lim_{t \rightarrow x^*} U(t)$ lorsqu'elle existe.

Théorème 1.28. *Lorsque $\gamma < 0$, si $F'(x)$ existe pour $x < x^*$ et si*

$$\lim_{t \rightarrow x^*} \frac{(x^* - t)F'(t)}{\overline{F}(t)} = -\frac{1}{\gamma} \quad (1.9)$$

alors F appartient au max-domaine d'attraction de G_γ . La condition (1.9) est équivalente à

$$\lim_{t \rightarrow x^*} \frac{tU'(t)}{U(\infty) - U(t)} = -\gamma$$

ce qui implique $U(\infty) - U \in \text{RV}_\gamma$.

Le théorème suivant montre que la variation régulière étendue est une condition nécessaire et suffisante d'appartenance au max-domaine d'attraction d'une loi de valeurs extrêmes généralisée. Ce résultat, démontré par de Haan (1970), a constitué une grande avancée dans la théorie des valeurs extrêmes.

Théorème 1.29 (de Haan & Ferreira (2006)). *F appartient au max-domaine d'attraction de G_γ si et seulement s'il existe une fonction auxiliaire positive a telle que pour tout $x > 0$*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \quad (1.10)$$

soit $U \in \text{ERV}_\gamma$. Pour $\gamma = 0$, le membre de droite s'interprète comme sa limite lorsque $\gamma \rightarrow 0$.

Remarque 1.30. Pour $\gamma > 0$, on peut choisir comme fonction auxiliaire γU et pour $\gamma < 0$, $-\gamma(U(\infty) - U)$.

Corollaire 1.31. *La condition (1.10) est équivalente à l'existence d'une fonction f strictement positive telle que*

$$\lim_{t \rightarrow x^*} \frac{\overline{F}(t + xf(t))}{\overline{F}(t)} = (1 + \gamma x)_+^{-1/\gamma}.$$

Remarque 1.32. La preuve de ce corollaire repose sur le lemme de Vervaat (Lemme A.3).

Ce corollaire montre que toutes ces conditions assurent également l'appartenance aux domaines d'attraction des excès. Ainsi, il y a identité entre les domaines d'attraction des excès et les max-domaines d'attraction. Dans toute la suite, nous noterons simplement $F \in \text{DA}(\gamma)$ pour signifier indifféremment que F appartient au domaine d'attraction des excès de Q_γ ou au max-domaine d'attraction de G_γ .

Comme le suggèrent les figures 1.1 et 1.2, on distingue trois domaines d'attraction différents, en fonction du signe de γ . Afin de les décrire, nous utilisons une nouvelle paramétrisation.

- Pour $\gamma > 0$, x^* est infini. Lorsque x tend vers l'infini, $\overline{G}_\gamma(x)$ est équivalent à $\gamma^{-1/\gamma}x^{-1/\gamma}$. Les moments d'ordre supérieur à $1/\gamma$ n'existent pas. Si X est une variable aléatoire de fonction de répartition G_γ , alors la variable $\gamma X + 1$ suit une loi de Fréchet de paramètre $1/\gamma$ dont la fonction de répartition a pour expression $\exp(-x^{-1/\gamma})\mathbb{1}_{]0,\infty[}(x)$. On parle du *domaine de Fréchet* : domaine des lois à queue épaisse. Les lois de Cauchy, Pareto, log-gamma et Student appartiennent au domaine de Fréchet.
- Pour $\gamma = 0$, x^* est infini. $\overline{G}_0(x)$ est équivalent à e^{-x} lorsque x tend vers l'infini, et tous les moments existent. La loi de fonction de répartition G_0 est appelée loi de Gumbel, on parle donc du *domaine de Gumbel* : domaine des lois à queue fine. Les lois gamma, normale, log-normale et exponentielle appartiennent au domaine de Gumbel.
- Pour $\gamma < 0$, $x^* = -\frac{1}{\gamma} < \infty$. $\overline{G}_\gamma(1/\gamma - x)$ est équivalent à $(-\gamma x)^{-1/\gamma}$, lorsque x tend vers 0, et tous les moments existent. Si X est une variable aléatoire de fonction de répartition G_γ , alors la variable $-(\gamma X + 1)$ suit une loi dont la fonction de répartition a pour expression $\exp(-(-x)^{-1/\gamma})\mathbb{1}_{]-\infty,0[}(x) + \mathbb{1}_{[0,\infty[}(x)$. On parle du *domaine de Weibull* : domaine des lois à queue finie. Les lois uniforme et beta appartiennent au domaine de Weibull.

En théorie des valeurs extrêmes, il est possible d'énoncer un théorème de convergence des moments.

Théorème 1.33 (CONVERGENCE DES MOMENTS (de Haan & Ferreira, 2006)). *Soit X une variable aléatoire de fonction de répartition $F \in \text{DA}(\gamma)$, $\gamma \in \mathbb{R}$, et q un entier tel que $0 < q < 1/\gamma_+$, où $\gamma_+ = \max(0, \gamma)$. Si $\mathbb{E}|X|^q < \infty$, alors*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\frac{X_{1,n} - \beta_n}{\alpha_n} \right)^q \right] = \int_{-\infty}^{\infty} x^q dG_\gamma(x).$$

où les suites α_n et β_n sont les mêmes que celles qui apparaissent dans (1.5).

Exemple 1.34 (LOI NORMALE). La fonction de répartition Φ de la loi normale $\mathcal{N}(0, 1)$ appartient au domaine de Gumbel. En effet, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} F^n(\alpha_n x + \beta_n) = \exp(-e^{-x}) \quad (1.11)$$

avec $\beta_n = \sqrt{2 \ln n - \ln \ln n - \ln(4\pi)}$ et $\alpha_n = 1/\beta_n$. En passant au logarithme et en utilisant que $-\ln x$ est équivalent à $1 - x$ pour $x \rightarrow 1$, l'équation (1.11) est équivalente à

$$\lim_{n \rightarrow \infty} n(1 - F(\alpha_n x + \beta_n)) = e^{-x}.$$

Puisque $\beta_n/\sqrt{2 \ln n} \xrightarrow[n \rightarrow \infty]{} 1$, $\ln \beta_n - \frac{1}{2} \ln \ln n - \frac{1}{2} \ln 2 \xrightarrow[n \rightarrow \infty]{} 0$, et

$$\frac{\beta_n^2}{2} + \ln \beta_n - \ln n + \frac{1}{2} \ln(2\pi) \xrightarrow[n \rightarrow \infty]{} 0.$$

Ainsi, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} -\frac{d}{dx} n(1 - F(\alpha_n x + \beta_n)) &= \frac{n}{\beta_n \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x}{\beta_n} + \beta_n\right)^2\right) \\ &= \exp\left\{-\left(\frac{\beta_n^2}{2} + \ln \beta_n - \ln n + \frac{1}{2} \ln(2\pi)\right)\right\} e^{-x^2/(2\beta_n^2)} e^{-x} \\ &\rightarrow e^{-x}. \end{aligned}$$

D'où, par convergence dominée,

$$\begin{aligned} n(1 - F(\alpha_n x + \beta_n)) &= \frac{n}{\beta_n \sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{1}{2} \left(\frac{u}{\beta_n} + \beta_n\right)^2\right) du \\ &= \exp\left\{-\left(\frac{\beta_n^2}{2} + \ln \beta_n - \ln n + \frac{1}{2} \ln(2\pi)\right)\right\} \int_x^\infty e^{-u^2/(2\beta_n^2)} e^{-u} du \\ &\rightarrow e^{-x}. \end{aligned}$$

De plus, si $X_{1,n}$ est le maximum de n variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$ alors, d'après le théorème 1.33, $\text{Var}[X_{1,n}] = O(1/\ln n)$.

Domaine d'attraction de Fréchet

Dans ce paragraphe, nous énonçons certaines propriétés du domaine de Fréchet dont nous aurons besoin dans le chapitre 3. La condition nécessaire et suffisante d'appartenance au domaine de Fréchet se traduit par une propriété de variation régulière de la fonction de survie \bar{F} ou de la fonction U .

Théorème 1.35. $F \in \text{DA}(\gamma)$, $\gamma > 0$, si et seulement si l'une de ces deux assertions équivalentes suivantes est vérifiée

1. $\bar{F} \in \text{RV}_{-1/\gamma}$
2. $U \in \text{RV}_\gamma$.

Ce résultat repose sur le fait qu'une fonction à variation régulière étendue d'indice strictement positif est à variation régulière avec le même indice et que l'on peut choisir pour fonction auxiliaire $a(t) = \gamma U(t)$.

Remarque 1.36. Pour l'appartenance au domaine de Weibull, il existe également une condition nécessaire et suffisante d'appartenance qui se traduit par une propriété de variation régulière, mais d'une façon un peu différente : $F \in \text{DA}(\gamma)$, $\gamma < 0$, si et seulement si $\overline{F}(x^* - \cdot) \in \text{RV}_{-1/\gamma}$, ce qui est encore équivalent à $U(\infty) - U \in \text{RV}_\gamma$ où $U(\infty) = \lim_{t \rightarrow \infty} U(t)$.

En revanche, pour le domaine de Gumbel, il n'existe pas de traduction simple en termes de variation régulière sur \overline{F} ou U .

La représentation de Karamata est une représentation générale pour les fonctions à variation régulière.

Théorème 1.37 (REPRÉSENTATION DE KARAMATA (de Haan & Ferreira, 2006; Resnick, 2007)). *Une fonction $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ est à variation régulière d'indice α , si et seulement s'il existe deux fonctions mesurables $c: \mathbb{R}_+ \rightarrow \mathbb{R}$ et $\eta: \mathbb{R}_+ \rightarrow \mathbb{R}$ vérifiant*

$$\lim_{t \rightarrow \infty} c(t) = c > 0 \quad \text{et} \quad \lim_{t \rightarrow \infty} \eta(t) = 0$$

et $t_0 \geq 0$ tels que

$$f(t) = c(t)t^\alpha \exp \left\{ \int_{t_0}^t \frac{\eta(t)}{t} dt \right\}.$$

Remarque 1.38. – Sous la condition de von Mises, la représentation de Karamata prend une forme plus simple. La fonction c peut être choisie constante et

$$\eta(t) = \frac{tU'(t)}{U(t)} - \gamma.$$

Cette fonction est appelée la fonction de von Mises.

– Le choix de t_0 est arbitraire, il suffit de changer les valeurs des expressions des fonctions c et η .

3.2 Estimation de l'indice de valeurs extrêmes γ

Dans le domaine des lois à queue épaisse, la statistique de la théorie des valeurs extrêmes se traduit par un problème d'estimation semi-paramétrique. En effet, si F appartient au domaine de Fréchet, alors sa fonction de survie (respectivement sa U -transformée) est de la forme $t^{-1/\gamma}L(t)$ (respectivement $t^\gamma L(t)$). Le paramètre d'intérêt est $\gamma > 0$ et la fonction L , fonction à variation lente, est un paramètre de nuisance. Dans le paragraphe suivant, nous présentons quelques estimateurs classiques pour l'indice de valeurs extrêmes γ .

Si l'on observe un échantillon $\mathbf{X}_n = (X_1, \dots, X_n)$ de fonction de répartition F inconnue appartenant à un domaine d'attraction $\text{DA}(\gamma)$, $\gamma \in \mathbb{R}$. Le paramètre γ reflète le comportement de la queue de F : il paraît donc judicieux de l'estimer comme une fonction de la queue du processus empirique, représentée par les plus grandes statistiques d'ordre.

Durant les 50 dernières années, différents estimateurs de γ ont été proposés. Ces estimateurs supposent que l'on ait préalablement déterminé le nombre k de statistiques d'ordre à prendre en compte : choisir k revient à réaliser un compromis biais-variance, le

k optimal étant celui qui minimise le risque quadratique asymptotique. Or, nous allons voir que ce risque dépend de paramètres inconnus. Depuis les années 1990, plusieurs auteurs (Carpentier & Kim, 2014a; Danielsson et al., 2001; Drees & Kaufmann, 1998; Grama & Spokoiny, 2008; Hall, 1990) ont proposé des procédures adaptatives pour choisir k de façon optimale.

Estimateur de Hill ($\gamma > 0$)

Introduit en 1975 par Hill (1975), l'estimateur de Hill, défini pour $\gamma > 0$, est probablement l'estimateur le plus étudié dans la littérature.

Définition 1.39 (ESTIMATEUR DE HILL). *L'estimateur de Hill, noté $\hat{\gamma}_H(k)$, construit à partir des k plus grandes statistiques d'ordre, est défini par*

$$\hat{\gamma}_H(k) = \frac{1}{k} \sum_{i=1}^k \ln \frac{X_{(i)}}{X_{(k+1)}}.$$

Exemple 1.40 (LOI DE PARETO). Soit ξ_1, \dots, ξ_k k variables aléatoires de loi de Pareto de paramètre de localisation 1 et de paramètre de forme γ , c'est-à-dire

$$\mathbb{P}\{\xi_i \leq x\} = (1 - x^{-1/\gamma}) \mathbb{1}_{[1, \infty[}(x).$$

Alors l'estimateur du maximum de vraisemblance de γ est

$$\frac{1}{k} \sum_{i=1}^k \ln \xi_i.$$

Ainsi, utiliser l'estimateur de Hill revient à considérer que les k variables $X_{(1)}/X_{(k+1)}, \dots, X_{(k)}/X_{(k+1)}$, sont distribuées selon un k -échantillon de loi de Pareto de paramètres 1 et γ .

Pour toute variable aléatoire X de fonction de répartition F , on définit pour tout $\tau > 0$

$$e(\tau) = \mathbb{E}[X - \tau \mid X > \tau].$$

$e(\tau)$ correspond à l'espérance sous la loi d'excès de X au-dessus du seuil τ . Elle peut être estimée à partir de données X_1, \dots, X_n par sa version empirique

$$\hat{e}_n(\tau) = \frac{\sum_{i=1}^n X_i \mathbb{1}_{\{X_i > \tau\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > \tau\}}} - \tau.$$

En prenant $\tau = X_{(k+1)}$, on peut réécrire

$$\hat{e}_n(\tau) = \frac{1}{k} \sum_{j=1}^k X_{(j)} - X_{(k+1)}.$$

Ainsi, l'estimateur de Hill est égal à $\hat{e}_n(\ln X_{(k+1)})$ calculée à partir des transformations logarithmiques des données $\ln X_1, \dots, \ln X_n$.

Proposition 1.41. *Si les X_i suivent une loi de Pareto de paramètres 1 et γ , alors $k\widehat{\gamma}_H(k)/\gamma$ est distribué selon une loi gamma de paramètres k et 1.*

Démonstration. La preuve de ce résultat repose sur la représentation des statistiques d'ordre comme transformation monotone des statistiques d'ordre d'un échantillon exponentiel (Proposition 1.12). \square

Sous la condition de von Mises (Condition (1.7)), il est possible d'établir deux autres représentations de l'estimateur de Hill à partir de la représentation de Karamata (Théorème 1.37). Ces nouvelles représentations permettront de proposer une autre démonstration de la consistance et de la normalité de l'estimateur de Hill, puis dans la partie 4 et dans le chapitre 3, nous verrons qu'elles offrent la possibilité d'obtenir facilement des bornes non asymptotiques sur les fluctuations de l'estimateur de Hill autour de son espérance.

Proposition 1.42 (REPRÉSENTATIONS POUR L'ESTIMATEUR DE HILL). *Si $F \in \text{DA}(\gamma)$, $\gamma > 0$,*

$$\widehat{\gamma}_H(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \ln \frac{c(e^{Y_{(i)}})}{c(e^{Y_{(k+1)}})} + \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{u+Y_{(k+1)}})) du \quad (1.12)$$

et

$$\widehat{\gamma}_H(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k i \ln \frac{c(e^{Y_{(i)}})}{c(e^{Y_{(i+1)}})} + \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{\frac{u}{i}+Y_{(i+1)}})) du. \quad (1.13)$$

Si de plus F satisfait la condition de von Mises,

$$\widehat{\gamma}_H(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{u+Y_{(k+1)}})) du \quad (1.14)$$

et

$$\widehat{\gamma}_H(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{\frac{u}{i}+Y_{(i+1)}})) du \quad (1.15)$$

où E_1, \dots, E_k sont des variables exponentielles de paramètre 1 indépendantes et $Y_{(i+1)}$ et $Y_{(k+1)}$ désignent respectivement la $(i+1)^e$ et la $(k+1)^e$ plus grandes statistiques d'ordre d'un n -échantillon exponentiel de paramètre 1.

La représentation (1.14) met en évidence le fait que, sous la condition de von Mises (Condition 1.7), $\widehat{\gamma}_H(k)$ est distribué comme un mélange de sommes de variables aléatoires i.i.d., conditionnellement à $Y_{(k+1)}$, et la représentation (1.15) permet de représenter simultanément les estimateurs $(\widehat{\gamma}_H)_{k < n}$ sur un même espace de probabilité où les variables aléatoires $E_i, 1 \leq i \leq n$ sont toutes définies.

Démonstration. Comme $F \in \text{DA}(\gamma)$, avec $\gamma > 0$, d'après la représentation de Karamata (Théorème 1.37),

$$U(t) = c(t)t^\gamma \exp \int_1^t \frac{\eta(v)}{v} dv$$

avec $\lim_{t \rightarrow \infty} c(t) = c > 0$ et $\lim_{t \rightarrow \infty} \eta(t) = 0$.

Ainsi,

$$\ln \frac{X_{(i)}}{X_{(k+1)}} \stackrel{d}{=} \ln \frac{c(e^{Y_{(i)}})}{c(e^{Y_{(k+1)}})} + \gamma(Y_{(i)} - Y_{(k+1)}) + \int_{Y_{(k+1)}}^{Y_{(i)}} \eta(e^v) dv$$

et

$$\widehat{\gamma}_H(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \ln \frac{c(e^{Y_{(i)}})}{c(e^{Y_{(k+1)}})} + \gamma \frac{1}{k} \sum_{i=1}^k (Y_{(i)} - Y_{(k+1)}) + \frac{1}{k} \sum_{i=1}^k \int_{Y_{(k+1)}}^{Y_{(i)}} \eta(e^v) dv.$$

Dans cette écriture, on peut oublier que les statistiques d'ordre $Y_{(i)}$ sont triées, et simplement se souvenir que les écarts $Y_{(i)} - Y_{(k+1)}$ sont distribués comme les statistiques d'ordre d'un k -échantillon de la loi exponentielle et qu'ils sont indépendants de $Y_{(k+1)}$. Si on introduit k variables exponentielles indépendantes E_1, \dots, E_k et indépendantes de $Y_{(k+1)}$, alors

$$\widehat{\gamma}_H(k) \stackrel{d}{=} \ln \frac{c(e^{Y_{(i)}})}{c(e^{Y_{(k+1)}})} + \frac{\gamma}{k} \sum_{i=1}^k E_i + \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} \eta(e^{Y_{(k+1)}+v}) dv.$$

(1.13) se démontre exactement de la même façon et sous la condition de von Mises, la fonction c est constante, donc le premier terme est nul. \square

Proposition 1.43 (Consistance de l'estimateur de Hill (Mason, 1982)). *Si $F \in \text{DA}(\gamma)$, $\gamma > 0$ et si (k_n) est une suite intermédiaire, alors*

$$\widehat{\gamma}_H(k_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \gamma. \quad (1.16)$$

Démonstration. Partons de la représentation (1.12) de $\widehat{\gamma}_H(k)$,

$$\widehat{\gamma}_H(k) \stackrel{d}{=} \underbrace{\frac{1}{k} \sum_{i=1}^k \ln \frac{c(e^{Y_{(i)}})}{c(e^{Y_{(k+1)}})}}_{\text{(I)}} + \underbrace{\frac{\gamma}{k} \sum_{i=1}^k E_i}_{\text{(II)}} + \underbrace{\frac{1}{k} \sum_{i=1}^k \int_0^{E_i} \eta(e^{v+Y_{(k+1)}}) dv}_{\text{(III)}}.$$

Le terme (II) converge en probabilité vers γ , d'après la loi des grands nombres.

Les termes de (I), (III) tendent en probabilité vers 0 parce que $Y_{(k_n)}$ converge en probabilité vers l'infini. Pour tout $\delta > 0$, il existe $t(\delta)$ tel que pour tout $t > t(\delta)$, $1/(1+\delta) \leq \frac{c(t)}{c} \leq 1+\delta$. Si $Y_{(k+1)} > \ln t(\delta)$,

$$\left| \ln \frac{c(e^{Y_{(i)}})}{c(e^{Y_{(k+1)}})} \right| \leq 2 \ln(1+\delta) \leq 2\delta.$$

D'où, pour tout $\delta > 0$, $\mathbb{P}\{|\text{(I)}| \leq 2\delta\} \geq \mathbb{P}\{Y_{(k+1)} > \ln t(\delta)\} \xrightarrow[n \rightarrow \infty]{} 1$.

Pour tout $\delta > 0$, il existe $t'(\delta)$ tel que $\forall t > t'(\delta)$, $|\eta(t)| < \delta$. D'où, $\mathbb{P}\{|\text{(III)}| \leq \delta \text{(II)}\} \geq \mathbb{P}\{Y_{(k+1)} > \ln t(\delta)\} \xrightarrow[n \rightarrow \infty]{} 1$. Ainsi, (III) tend en probabilité vers 0, car (II) est borné. \square

Dans son article, Mason (1982) a démontré que la proposition 1.43 admettait une réciproque : $F \in \text{DA}(\gamma)$ si et seulement si pour toute une suite intermédiaire (k_n) (1.16) est vérifiée.

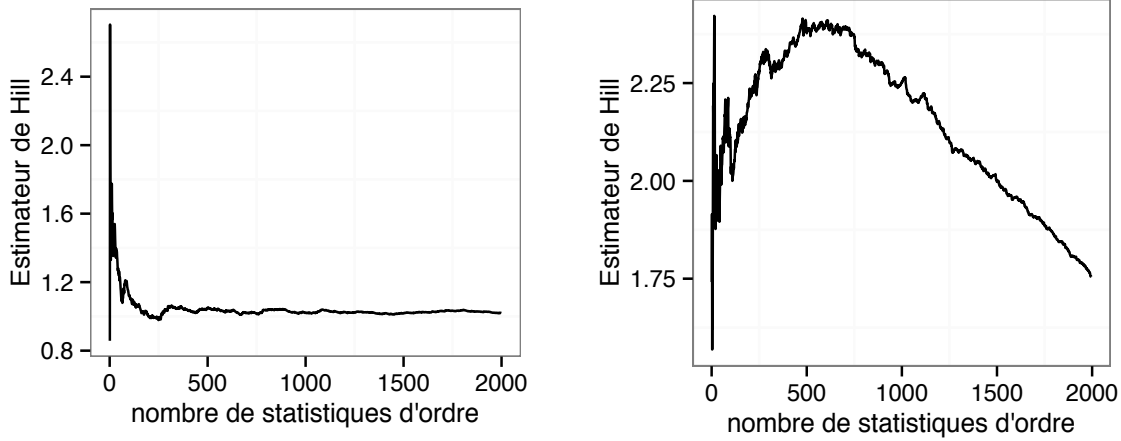


FIGURE 1.3. Hill Plots à partir d'un nombre de statistiques d'ordre allant de 1 à 2000 pour 10 000 observations de la loi Fréchet de paramètre 1 (Figure de Gauche) et pour 10 000 observations d'une loi α -stable symétrique ($\alpha = 1.7$) (Figure de Droite)

La proposition 1.43 assure la consistance de l'estimateur de Hill, mais ne donne pas d'indication sur la vitesse de convergence. Or, la convergence du rapport $L(tx)/L(t)$ vers 1 lorsque $t \rightarrow \infty$ peut être très lente, ce qui retentira sur la vitesse de convergence de $\hat{\gamma}_h(k)$ vers γ .

Ce phénomène est illustré par l'exemple suivant (Resnick, 2007). Si l'on trace le graphe construit à partir des points $\{(k, \hat{\gamma}_H(k)), 1 \leq k \leq n-1\}$, appelé Hill Plot, il semble naturel d'estimer γ à partir d'une région stable du graphe, ce qui n'est pas toujours évident. La figure de gauche représente un Hill Plot construit à partir d'un nombre de statistiques d'ordre allant de 1 à 2000 pour 10 000 observations d'une loi de Fréchet de paramètre 1, et la figure de droite, un Hill Plot construit à partir d'un nombre de statistiques d'ordre allant de 1 à 2000 pour 10 000 observations d'une loi α -stable symétrique ($\alpha = 1.7$). Le premier graphe indique correctement la valeur 1, en revanche, le deuxième est plus difficile à lire.

Pour mieux comprendre ce qu'il se passe, étudions les fluctuations de l'estimateur de Hill. La représentation (1.12) permet de remarquer que les termes contribuant au biais sont $\frac{1}{k} \sum_{i=1}^k \ln \frac{c(e^{Y^{(i)}})}{c(e^{Y^{(k+1)}})}$ et $\frac{1}{k} \sum_{i=1}^k \int_0^{E_i} \eta(e^{v+Y^{(k+1)}}) dv$. Sous la condition de von Mises, le biais s'écrit

$$\mathbb{E} \hat{\gamma}_H(k) - \gamma = \mathbb{E} \left[\int_0^E \eta(e^{Y^{(k+1)}+v}) dv \right] = \mathbb{E} \left[\int_1^\infty \frac{\eta(e^{Y^{(k+1)}+v})}{v^2} dv \right].$$

Ainsi, on définit la fonction b sur $]1, +\infty[$ par

$$b(t) = \int_0^\infty e^{-v} \eta(te^v) dv = \int_1^\infty \frac{\eta(tv)}{v^2} dv = t \int_t^\infty \frac{\eta(v)}{v^2} dv. \quad (1.17)$$

La quantité $b(t)$ désigne le biais conditionnel de $\hat{\gamma}_H(k)$ sachant que $\bar{F}(X_{(k+1)}) = 1/t$. Autrement dit, $\mathbb{E} [\hat{\gamma}_H(k) | Y_{(k+1)}] - \gamma = b(e^{Y_{(k+1)}})$. La deuxième expression permet de voir que b est différentiable (même si η n'est différentiable nulle part) et

$$b'(t) = \frac{b(t) - \eta(t)}{t}.$$

Soit $\bar{\eta}(t) = \sup_{s \geq t} |\eta(s)|$, et $\bar{\eta}(t) \xrightarrow[t \rightarrow \infty]{} 0$, car $\eta(t) \xrightarrow[t \rightarrow \infty]{} 0$. On supposera que $\bar{\eta}$ est bornée. Nous attirons l'attention du lecteur sur le fait que cette notation n'est pas à confondre avec la notation pour la fonction de survie \bar{F} .

Théorème 1.44. *Si $F \in \text{DA}(\gamma)$, $\gamma > 0$ satisfait la condition de von Mises (Condition 1.7) et si (k_n) est une suite intermédiaire, alors*

$$\sqrt{k_n} (\hat{\gamma}_H(k_n) - \mathbb{E} \hat{\gamma}_H(k_n)) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

Ce théorème suggère que la variance de $\hat{\gamma}_H(k_n)$ est de l'ordre de γ^2/k_n .

Démonstration. Si F satisfait la condition de von Mises (1.7), alors d'après la représentation (1.14)

$$\begin{aligned} & \sqrt{k_n} (\hat{\gamma}_H(k_n) - \mathbb{E} \hat{\gamma}_H(k_n)) \\ & \stackrel{d}{=} \sqrt{k_n} (\hat{\gamma}_H(k_n) - \gamma - b(e^{Y(k_n+1)}) + \mathbb{E} [\hat{\gamma}_H(k_n) | Y(k_n+1)] - \mathbb{E} \hat{\gamma}_H(k_n)) \\ & \stackrel{d}{=} \underbrace{\frac{\gamma}{\sqrt{k_n}} \sum_{i=1}^{k_n} (E_i - 1)}_{(I)} + \underbrace{\sqrt{k_n} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \int_0^{E_i} \eta(e^{Y(k_n+1)+v}) dv - b(e^{Y(k_n+1)}) \right)}_{(II)} \\ & \quad + \underbrace{\sqrt{k_n} (\mathbb{E} [\hat{\gamma}_H(k_n) | Y(k_n+1)] - \mathbb{E} \hat{\gamma}_H(k_n))}_{(III)}. \end{aligned}$$

(I) converge en loi vers une loi gaussienne centrée de variance γ^2 par le théorème central limite. (III) tend en probabilité vers 0 grâce à l'inégalité de Chebyshev et à l'inégalité de Poincaré pour les vecteurs exponentiels (Théorème 1.75) que nous verrons dans la partie 4. Puis on montre que (II) converge en moyenne quadratique, et donc en probabilité, vers 0. Comme

$$(II) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \int_0^\infty (\mathbb{1}_{E_i > v} - e^{-v}) \eta(e^{Y(k_n+1)+v}) dv,$$

on a

$$\begin{aligned} \mathbb{E} [(II)^2] &= \mathbb{E} [\mathbb{E} [(II)^2 | Y(k_n+1)]] \\ &= \mathbb{E} \left[\frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{E} \left[\left(\int_0^\infty (\mathbb{1}_{E_i > v} - e^{-v}) \eta(e^{Y(k_n+1)+v}) dv \right)^2 \mid Y(k_n+1) \right] \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\left(\int_0^\infty (\mathbb{1}_{E > v} - e^{-v}) \eta(e^{Y(k_n+1)+v}) dv \right)^2 \mid Y(k_n+1) \right] \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\left(\int_0^\infty (1+v^2) (\mathbb{1}_{E > v} - e^{-v}) \eta(e^{Y(k_n+1)+v}) \frac{dv}{1+v^2} \right)^2 \mid Y(k_n+1) \right] \right] \\ &\leq \mathbb{E} \left[\mathbb{E} \left[\left(\int_0^\infty (1+v^2) (\mathbb{1}_{E > v} - e^{-v})^2 dv \right) \left(\int_0^\infty \eta(e^{Y(k_n+1)+v})^2 \frac{dv}{1+v^2} \right) \mid Y(k_n+1) \right] \right] \\ &= \mathbb{E} \left[\left(\int_0^\infty (1+v^2) (\mathbb{1}_{E > v} - e^{-v})^2 dv \right) \mathbb{E} \left[\int_0^\infty \eta(e^{Y(k_n+1)+v})^2 \frac{dv}{1+v^2} \right] \right] \\ &\leq 4 \mathbb{E} \left[(\bar{\eta}(e^{Y(k_n+1)}))^2 \right] \end{aligned}$$

où la première majoration est obtenue grâce à l'inégalité de Cauchy-Schwarz et la dernière majoration vient de

$$\begin{aligned}
& \int_0^\infty \int_0^\infty (1+v^2) (\mathbb{1}_{z>v} - e^{-v})^2 e^{-z} dz dv \\
&= \int_0^\infty \left(\int_0^z (1+v^2) (1-e^{-v})^2 dv + \int_z^\infty (1+v^2) e^{-2v} dv \right) e^{-z} dz \\
&\leq \int_0^\infty \left(z + \frac{z^3}{3} + \int_0^\infty (1+v^2) e^{-2v} dv \right) e^{-z} dz \\
&= \int_0^\infty \left(z + \frac{z^3}{3} + \frac{1}{2} + \frac{1}{4} \right) e^{-z} dz \\
&= 1 + 2 + \frac{1}{2} + \frac{1}{4} \\
&\leq 4.
\end{aligned}$$

Ainsi, la variable aléatoire

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \int_0^\infty (\mathbb{1}_{E_i>v} - e^{-v}) \eta(e^{Y_{(k_n+1)+v}}) dv$$

tend en moyenne quadratique vers 0. En effet la fonction $\bar{\eta}$ tend vers 0 en l'infinie et $Y_{(k_n+1)}$ tend presque sûrement vers ∞ , lorsque n tend vers l'infini, on peut donc conclure par convergence dominée. \square

Une autre preuve de ce théorème se trouve dans (Resnick, 2007), en partant de la convergence de la queue du processus empirique vers un mouvement brownien.

Nous avons montré un résultat de normalité asymptotique pour l'estimateur de Hill où le terme de recentrage est son espérance. Nous aimerions le remplacer par γ . Pour cela, nous avons besoin d'une condition du second d'ordre afin de contrôler la vitesse de convergence du biais.

Corollaire 1.45. *Si $F \in \text{DA}(\gamma)$, $\gamma > 0$ satisfait la condition de von Mises, et si la suite intermédiaire (k_n) vérifie une condition du second ordre*

$$\lim_{n \rightarrow \infty} \sqrt{k_n} b \left(\frac{n}{k_n} \right) = \lambda$$

alors

$$\sqrt{k_n} (\hat{\gamma}_H(k_n) - \gamma) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\lambda, \gamma^2).$$

Remarque 1.46. λ est un terme de biais dépendant de la suite (k_n) .

D'autres conditions du second ordre ont été proposées. En effet d'après Segers (2002), le comportement du biais de l'estimateur de Hill est lié aux propriétés de variation régulière de \bar{F} et de U . Il existe deux sortes d'hypothèses. La première suppose que F satisfait une condition de variation régulière du second ordre. Sous la condition de von Mises, cela revient à faire des hypothèses sur le comportement asymptotique de $\ln(x^{-\gamma}U(tx)/U(t)) = \int_t^{tx} (\eta(s)/s) ds$: par exemple, qu'il existe une fonction $A \in \text{RV}_\rho$, $\rho < 0$ telle que

$$\left| \frac{\int_t^{tx} (\eta(s)/s) ds}{A(t)} \right| \xrightarrow[t \rightarrow \infty]{} \frac{x^\rho - 1}{\rho}.$$

Etant donné que cela implique que, pour toute suite intermédiaire (k_n) , le rapport $|\mathbb{E}\widehat{\gamma}_H(k_n) - \gamma|/A(n/k_n)$ tend vers une limite finie $c/(1 + |\rho|)$, l'estimation du paramètre du second ordre devient alors un objectif naturel. La deuxième suppose que F satisfait une condition d'enveloppe. La O -variation régulière fournit alors un cadre pour décrire les vitesses de convergence (Segers, 2002).

Estimateur de Pickands ($\gamma \in \mathbb{R}$)

Les preuves des paragraphes suivants peuvent être trouvées dans (de Haan & Ferreira, 2006).

L'intérêt de l'estimateur de Pickands (1975) est d'être défini et consistant pour $\gamma \in \mathbb{R}$.

Définition 1.47 (ESTIMATEUR DE PICKANDS). *L'estimateur de Pickands est défini par :*

$$\widehat{\gamma}_P(k) = \frac{1}{\ln 2} \ln \frac{X_{(k+1)} - X_{(2k+1)}}{X_{(2k+1)} - X_{(4k+1)}}.$$

Cet estimateur est consistant pour toute suite intermédiaire (k_n) et asymptotiquement normal sous des conditions du second ordre.

Proposition 1.48 (CONSISTANCE). *Si $F \in \text{DA}(\gamma)$, $\gamma \in \mathbb{R}$, et si (k_n) est une suite intermédiaire, alors*

$$\widehat{\gamma}_P(k_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \gamma.$$

La preuve de cette proposition est basée sur la représentation de Rényi, et sur la normalité asymptotique du vecteur

$$\sqrt{2k_n} \left(\frac{1}{2} \frac{Y_{(k_n+1)}}{Y_{(2k_n+1)}} - 1, \frac{\sqrt{2}}{2} \frac{Y_{(2k_n+1)}}{Y_{(4k_n+1)}} - \sqrt{2} \right)$$

lorsque (k_n) est une suite intermédiaire.

Remarque 1.49. Au cours de la preuve de ce résultat, il est montré que

$$\frac{X_{(k_n+1)} - X_{(2k_n+1)}}{X_{(2k_n+1)} - X_{(4k_n+1)}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{4^\gamma - 1}{2^\gamma - 1} - 1 = 2^\gamma,$$

ce qui explique la normalisation par $\ln 2$ dans l'expression de $\widehat{\gamma}_P(k)$. Par ailleurs, $(2^\gamma - 1)/\gamma$ et $(4^\gamma - 1)/\gamma$ sont respectivement la médiane et le troisième quartile de la loi de Pareto généralisée Q_γ . Ainsi, l'estimateur de Pickands estime γ via les quantiles de la loi limite tandis que l'estimateur de Hill estime γ via un moment de la loi limite.

Théorème 1.50 (NORMALITÉ ASYMPTOTIQUE). *Si $F \in \text{DA}(\gamma)$, $\gamma \in \mathbb{R}$, satisfait une condition de variation régulière du second ordre d'indices γ et $\rho \leq 0$ avec pour fonction auxiliaire A (Théorème A.14), et si (k_n) est une suite intermédiaire telle que*

$$\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda < \infty$$

alors

$$\sqrt{k} (\widehat{\gamma}_P(k_n) - \gamma) \xrightarrow{d} \mathcal{N}(\lambda b_{\gamma, \rho}, v_\gamma)$$

où $b_{\gamma, \rho}$ est un terme de biais dépendant à la fois de γ et du paramètre du second ordre ρ et v_γ est un terme de variance qui ne dépend que de γ .

de Haan & Ferreira (2006) proposent deux preuves de ce résultat : la première utilise la représentation de Rényi et les inégalités de Potter pour la variation régulière du second ordre (Proposition A.15) ; la deuxième utilise la convergence de la queue du processus empirique vers un mouvement brownien.

Estimateur du maximum de vraisemblance ($\gamma > -1/2$)

La famille des fonctions de répartition appartenant à un domaine d'attraction donné ne peut pas être paramétrée par un nombre fini de paramètres ; il n'existe donc pas d'estimateur paramétrique du maximum de vraisemblance de γ . Cependant, l'appartenance à un domaine d'attraction pour un certain $\gamma \in \mathbb{R}$ est équivalente à la convergence

$$\bar{F}_\tau(\mu_\tau + \sigma_\tau x) \xrightarrow{\tau \rightarrow \infty} (1 + \gamma x)^{-1/\gamma}.$$

Cela suggère que les observations dépassant un seuil élevé sont approximativement distribuées selon une loi de Pareto généralisée Q_γ . Comme la famille des lois limites est paramétrée selon un seul paramètre γ , il est possible d'appliquer la procédure du maximum de vraisemblance aux plus grandes observations sous un modèle de Pareto généralisé.

Ainsi, l'estimateur du maximum de vraisemblance $\hat{\gamma}_{\text{MLE}}(k)$ maximise une vraisemblance approchée qui est calculée en considérant que les excès $(X_{(k)} - X_{(k+1)}, \dots, X_{(1)} - X_{(k+1)})$ au-dessus de $X_{(k+1)}$ constituent un k -échantillon de loi Q_γ .

Lorsque $\gamma \leq -1/2$, $\hat{\gamma}_{\text{MLE}}(k)$ n'est pas défini, car la vraisemblance approchée n'a pas de maximum. Nous nous placerons donc dans le cas où $\gamma > -1/2$.

Si F satisfait une condition de variation régulière du second ordre d'indices γ et $\rho \leq 0$, alors $\hat{\gamma}_{\text{MLE}}(k)$ est asymptotiquement normal avec un biais asymptotique dépendant de γ et du paramètre du second d'ordre ρ et une variance asymptotique ne dépendant que de γ .

Estimateurs des moments ($\gamma \in \mathbb{R}$)

Le but de l'estimateur des moments, introduit par Dekkers et al. (1989), est de proposer un estimateur mimant l'estimateur de Hill pour $\gamma \leq 0$. Le problème avec l'estimateur de Hill lorsque $\gamma \leq 0$ est que la limite de U en l'infini peut être négative. Dans ce cas, les X_i peuvent prendre des valeurs négatives et leur logarithme n'existe pas. Nous nous plaçons, dans ce paragraphe, dans le cas où $U(\infty) = \lim_{t \rightarrow \infty} U(t) > 0$. Il faut toutefois savoir que cette hypothèse peut influencer le comportement de l'estimateur.

Définition 1.51 (ESTIMATEUR DES MOMENTS). *On définit les moments empiriques suivants*

$$\begin{aligned} M_n^{(1)} &= \frac{1}{k} \sum_{i=1}^k \ln \frac{X_{(i)}}{X_{(k+1)}} \\ M_n^{(2)} &= \frac{1}{k} \sum_{i=1}^k \left(\ln \frac{X_{(i)}}{X_{(k+1)}} \right)^2 \end{aligned}$$

et l'estimateur des moments par

$$\hat{\gamma}_M(k) = M_n^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1}.$$

Cet estimateur est consistant et asymptotiquement normal sous une condition de variation régulière du second ordre d'indices γ et ρ avec un terme de biais qui dépend de γ et du paramètre du second ordre ρ et un terme de variance qui ne dépend que de γ .

Remarque 1.52. La méthode des moments fournit également un estimateur pour la fonction auxiliaire a qui témoigne de $U \in \text{ERV}_\gamma$ (Définition 1.23). Ainsi, cet estimateur peut être utilisé pour construire des estimateurs de grands quantiles. Le quantile d'ordre $1 - k/(nx)$ peut être estimé par :

$$\hat{U} \left(\frac{n}{k}x \right) = \hat{U} \left(\frac{n}{k} \right) + \hat{a} \left(\frac{n}{k} \right) \left(\frac{x^{\hat{\gamma}_M(k)} - 1}{\hat{\gamma}_M(k)} \right).$$

Comment choisir le nombre de statistiques d'ordre

Nous venons de voir des exemples d'estimateurs de l'indice de valeurs extrêmes γ . Tous ces estimateurs sont des fonctionnelles des plus grandes statistiques d'ordre. De nombreux auteurs se sont intéressés au choix optimal du nombre de statistiques d'ordre à prendre en compte dans leur construction. Ce nombre optimal k_n^* est celui qui réalise le compromis biais-variance, c'est-à-dire celui qui minimise le risque quadratique asymptotique. La loi des observations n'étant pas connue, elle est approchée par sa limite. Tous les estimateurs que nous venons de voir sont asymptotiquement normaux sous des conditions du second ordre avec un terme de biais asymptotique qui dépend du paramètre du second ordre ρ . Ainsi, sous ces conditions, le choix de k_n^* dépend de ρ .

Par exemple, Hall & Welsh (1985) ont montré que pour l'estimateur de Hill, k_n^* est équivalent à

$$\left(-\frac{C^{-2\rho}(1-\rho)^2}{2D^2\rho^3} \right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}$$

si F satisfait

$$\bar{F}(x) = Cx^{-1/\gamma} (1 + Dx^{\rho/\gamma} + o(x^{\rho/\gamma})),$$

où $C > 0$ et $D \neq 0$ sont des constantes et $\rho < 0$ est un paramètre du second ordre.

Cependant, comme ρ , C , et D sont inconnus, ce résultat ne peut pas être utilisé dans la pratique.

Ainsi, depuis 40 ans, des auteurs Beirlant et al. (1996); Carpentier & Kim (2014a); Danielsson et al. (2001); Drees & Kaufmann (1998); Grama & Spokoiny (2008); Hall (1990) se sont intéressés au choix de k selon des procédures adaptatives, c'est-à-dire des stratégies de choix ne dépendant que de quantités connues du statisticien et permettant d'estimer γ presque aussi efficacement que si l'on connaissait le paramètre du second ordre. Dans le chapitre 3, nous proposons une version adaptative de l'estimateur de Hill en utilisant la méthode de Lepski. En effet, dans les années 1990, Lepski a introduit une méthode générale pour la sélection de modèle permettant de construire des procédures

adaptatives et d'établir des inégalités de type oracle dans des contextes inférentiels allant de l'estimation de densité à la classification. Nous renvoyons à l'annexe B pour plus de détails.

3.3 Méthode *Peaks over threshold* et méthode des maxima par blocs

Dans cette partie, nous présentons deux techniques couramment utilisées dans la pratique. La première est basée sur la convergence de la loi des excès vers une loi de Pareto généralisée, tandis que la deuxième est fondée sur la convergence du maximum.

Méthode *Peaks over threshold*

On suppose que l'on dispose de n variables aléatoires i.i.d. X_1, \dots, X_n et que l'on a choisi un seuil τ (choix qui n'est pas évident). On note alors

$$\begin{aligned} t_1 &= \inf\{j \geq 1 : X_j > \tau\} \\ t_2 &= \inf\{j > t_1 : X_j > \tau\} \\ &\vdots \\ t_r &= \inf\{j > t_{r-1} : X_j > \tau\} \end{aligned}$$

La suite $(X_{t_r}, r \geq 1)$ est la suite des X_i ayant dépassé τ .

Le principe général de la méthode POT (*Peaks over threshold*) (illustré à la figure 1.4) repose sur l'idée suivante : la caractérisation de l'appartenance au domaine de Fréchet $\bar{F} \in \text{RV}_{-1/\gamma}$ suggère que la loi d'excès de F peut être approchée par une loi de Pareto de paramètre de forme γ . On suppose donc que F possède une queue de type Pareto à droite, c'est-à-dire qu'il existe $c > 0$ tel que

$$\bar{F}(x) = cx^{-1/\gamma}, \quad x > \tau$$

Alors,

$$\mathbb{P}\{X_{t_1} > x\} = \frac{\bar{F}(x)}{\bar{F}(\tau)} = \left(\frac{x}{\tau}\right)^{-1/\gamma},$$

soit pour $y > 1$,

$$\mathbb{P}\left\{\frac{X_{t_1}}{\tau} > y\right\} = y^{-1/\gamma}.$$

Ainsi, la suite de variables aléatoires

$$\left\{\frac{X_{t_r}}{\tau}, r \geq 1\right\}$$

est un échantillon de loi de Pareto de paramètre de forme γ et de localisation 1. L'estimateur de Hill est alors le maximum de vraisemblance de γ , lorsque l'on a choisi $\tau = X_{(k+1)}$.

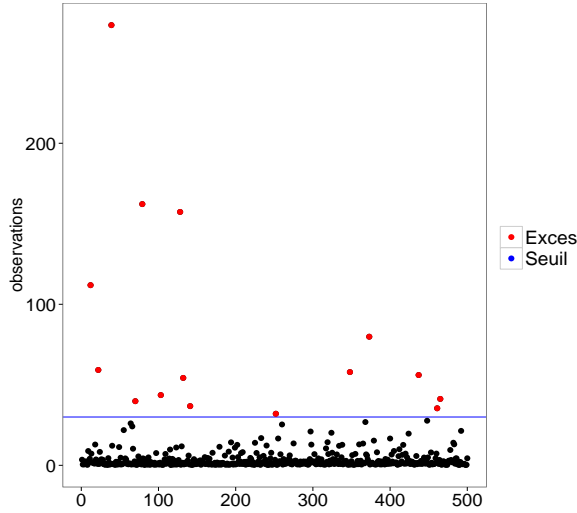


FIGURE 1.4. Illustration de la méthode POT

Méthode des maxima par blocs

La méthode des maxima par blocs (Coles, 2001) consiste à construire une suite de maxima i.i.d., afin de pouvoir ajuster une loi de valeurs extrêmes généralisée.

Supposons que nous observons X_1, X_2, \dots i.i.d. que nous avons rangées dans m blocs de même taille n .

$$\underbrace{X_1, \dots, X_n}_{\text{Bloc 1}} \mid \underbrace{X_{n+1}, \dots, X_{2n}}_{\text{Bloc 2}} \mid \dots \mid \underbrace{X_{(j-1)n+1}, \dots, X_{jn}}_{\text{Bloc } j} \mid \dots \mid \underbrace{X_{(m-1)n+1}, \dots, X_{mn}}_{\text{Bloc } m}$$

Pour chaque bloc j , on note $M_{n,j}$ le maximum de $X_{(j-1)n+1}, \dots, X_{jn}$. On obtient ainsi une suite de maxima i.i.d. $M_{n,1}, \dots, M_{n,m}$.

Par exemple, dans le cas des épidémies de grippe, les données peuvent représenter le nombre de cas de grippe par semaine. Nous disposons alors d'un échantillon de maxima hebdomadaires.

Une loi de valeurs extrêmes généralisée $G_{(\mu, \sigma, \gamma)}$ est alors ajustée en estimant les trois paramètres (μ, σ, γ) , par exemple par maximum de vraisemblance. La méthode des maxima par blocs est basée sur l'hypothèse que les observations sont distribuées selon une fonction de répartition appartenant à un domaine d'attraction d'une loi de valeurs extrêmes. Ainsi, le maximum d'un nombre suffisamment grand d'observations suit approximativement une loi de valeurs extrêmes généralisée. Cependant, les propriétés du maximum de vraisemblance reposent sur l'hypothèse que le maximum de chaque bloc est exactement une loi de valeurs extrêmes généralisée. Dans de nombreuses situations, cette hypothèse n'est pas satisfaisante, Dombry (2013) justifie l'utilisation de la méthode du maximum de vraisemblance pour la méthode des maxima par blocs sous l'hypothèse que les données sont distribuées selon une loi appartenant à un domaine d'attraction.

Les quantiles extrêmes z_p d'ordre $1 - p$ sont déterminés en inversant l'équation $G_{(\mu, \sigma, \gamma)}(z_p) = 1 - p$, soit

$$z_p = \begin{cases} \mu - \frac{\sigma}{\gamma} [1 - \{-\ln(1-p)\}^{-\gamma}] & \text{pour } \gamma \neq 0 \\ \mu - \sigma \ln\{-\ln(1-p)\} & \text{pour } \gamma = 0. \end{cases} \quad (1.18)$$

Le quantile z_p est appelé niveau de retour et p la période de retour : on s'attend à ce que le niveau z_p soit dépassé en moyenne une fois toutes les $1/p$ années, ou plus précisément, à ce que z_p soit dépassé par le maximum d'une année donnée avec probabilité p .

En posant $y_p = -\ln(1-p)$, l'équation (1.18) devient

$$z_p = \begin{cases} \mu - \frac{\sigma}{\gamma} (1 - y_p^{-\gamma}) & \text{pour } \gamma \neq 0 \\ \mu - \sigma \ln y_p & \text{pour } \gamma = 0. \end{cases}$$

Ainsi, si l'on trace z_p en fonction de y_p en échelle logarithmique, le graphe est linéaire si $\gamma = 0$, concave si $\gamma < 0$ avec une asymptote en $\mu - \sigma/\gamma$ lorsque $p \rightarrow 0$ et convexe si $\gamma > 0$ et n'a pas de borne finie. Ce graphe est souvent appelé « ReturnLevel Plot ».

Généralisation au cas m -dépendant (Leadbetter et al., 1983)

La méthode des maxima par blocs telle que nous venons de la présenter n'est valable que pour une série d'observations i.i.d., mais nous aurons besoin d'une généralisation au cas m -dépendant lorsque nous l'appliquerons à l'épidémiologie.

Définition 1.53. *Une suite de variables aléatoires (X_1, \dots, X_n) est m -dépendante si X_i et X_j sont indépendants dès que $|i - j| > m$.*

Théorème 1.54 (Leadbetter et al. (1983, Théorème 3.3.3)). *Soit $(X_i)_i$ une suite de variables aléatoires m -dépendante et $\alpha_n > 0$ et β_n deux suites telles que $\mathbb{P}\{(X_{1,n} - \beta_n)/\alpha_n \leq x\}$ converge vers une fonction de répartition non dégénérée G . Alors G est une loi de valeurs extrêmes généralisée.*

Nous connaissons le type de la limite en loi du maximum d'une suite de variables m -dépendantes, mais à quelles conditions cette limite existe-t-elle ? Nous nous intéressons donc à la convergence des probabilités de la forme $\mathbb{P}\{X_{1,n} \leq u_n\}$ où $u_n = x/\alpha_n + \beta_n$.

Théorème 1.55 (Leadbetter et al. (1983, Théorème 3.4.1)). *Soit $(X_i)_i$ une suite m -dépendante telle qu'il existe une suite de réels (u_n) telle que*

$$\limsup_{n \rightarrow \infty} n \sum_{j=2}^{[n/k]} \mathbb{P}\{X_1 > u_n, X_j > u_n\} \rightarrow 0 \text{ quand } k \rightarrow \infty.$$

Alors $\mathbb{P}\{(X_{1,n} - \beta_n)/\alpha_n \leq x\}$ converge vers une fonction de répartition non dégénérée.

Remarque 1.56. Ces résultats se généralisent aux suites de variables aléatoires stationnaires satisfaisant des conditions de mélange.

Application à l'épidémiologie de la grippe

Dans le chapitre 4, nous appliquons le modèle suivant afin de prédire la probabilité d'épidémies de grippe exceptionnellement meurtrières.

Supposons que nous disposons d'une série d'observations i.i.d. X_1, X_2, \dots que nous avons regroupées en m blocs de longueur n .

Soit ℓ un entier tel que $n \geq 2\ell$. Pour chaque bloc j , nous définissons

$$\begin{aligned} S_1^{(j)} &= X_{(j-1)n+1} + \dots + X_{(j-1)n+\ell} \\ S_2^{(j)} &= X_{(j-1)n+2} + \dots + X_{(j-1)n+\ell+1} \\ &\vdots \\ S_i^{(j)} &= X_{(j-1)n+i} + \dots + X_{(j-1)n+\ell+i-1} \\ &\vdots \\ S_{n-\ell+1}^{(j)} &= X_{jn-\ell+1} + \dots + X_{jn}. \end{aligned}$$

Les $S^{(j)}$ sont des sommes cumulées permettant de rendre compte de la taille d'une épidémie de grippe. Nous nous intéressons au comportement asymptotique de

$$M_n^{(j)} = \max(S_1^{(j)}, \dots, S_{n-\ell+1}^{(j)}).$$

Si les X_i sont identiquement distribuées, les $S^{(j)}$ le sont également mais ne sont pas indépendantes.

Proposition 1.57. *Pour $1 \leq j \leq m$ fixé, $(S_i^{(j)})_i$ est $(\ell-1)$ -dépendante, où ℓ est un entier supérieur à 2 et tel que $2\ell \leq n$.*

D'après le paragraphe précédent, s'il existe une suite u_n telle que

$$\limsup_{n \rightarrow \infty} n \sum_{i=2}^{\lfloor n/k \rfloor} \mathbb{P}\{S_1^{(j)} > u_n, S_i^{(j)} > u_n\} \rightarrow 0 \text{ quand } k \rightarrow \infty.$$

alors $M_n^{(j)}$ converge vers une loi de valeurs extrêmes généralisée.

4 Inégalités de concentration

Les inégalités de concentration permettent de quantifier les fluctuations d'une fonction de plusieurs variables aléatoires indépendantes en majorant la probabilité que cette fonction s'écarte de son espérance ou de sa médiane de plus qu'un certain réel. Elles ont été motivées par un grand nombre de domaines tels que la statistique, la théorie de l'apprentissage, la mécanique statistique, la théorie des matrices aléatoires ou encore la théorie de l'information.

Les premières inégalités de concentration ont été établies pour des sommes de variables aléatoires indépendantes (Bennett, 1962; Bernstein, 1946; Chernoff, 1953; Craig, 1933; Hoeffding, 1963; Okamoto, 1958; Uspensky, 1937). Des outils plus puissants pour des

fonctions plus générales de variables aléatoires n'ont pas été introduits avant l'apparition des martingales dans les années 1970 (Maurey, 1979; Milman & Schechtman, 1986). Puis, des inégalités de Sobolev logarithmiques pour les vecteurs gaussiens ont été établies à partir de 1975 par Gross, Cirelson, Ibragomiv et Sudakov. A la fin des années 1980, Talagrand s'intéresse à la concentration gaussienne et aux inégalités isopérimétriques (Ledoux & Talagrand, 1991; Talagrand, 1987, 1988). Puis, au milieu des années 1990, Talagrand (1993, 1994, 1995, 1996a,b) fournit un nouveau cadre à ce problème et offre de nouvelles perspectives de recherche qui reposent sur l'idée qu'une variable qui dépend de façon régulière de plusieurs variables aléatoires indépendantes vérifie des bornes de type Chernoff².

Dans cette partie, nous commençons par introduire, à partir de l'inégalité de Markov, la méthode de Cramér-Chernoff qui permet d'obtenir des inégalités de concentration à l'aide de la fonction génératrice des moments. Puis, nous présentons l'inégalité d'Efron-Stein, majoration simple, mais puissante, de la variance d'une fonction générale de plusieurs variables indépendantes. Enfin, nous expliquons le principe général de la méthode entropique, initiée par Ledoux à la fin des années 1990 (Ledoux, 1994, 1997).

4.1 Premières inégalités

L'inégalité de Markov affirme que si Y est une variable aléatoire positive alors pour tout $t > 0$,

$$\mathbb{P}\{Y > t\} \leq \frac{\mathbb{E}Y}{t}.$$

Plusieurs inégalités simples découlent de cette inégalité. De façon générale, si ϕ est une fonction croissante positive définie sur un intervalle $I \subset \mathbb{R}$ alors pour toute variable aléatoire Z à valeurs dans I et pour tout $t \in I$ tel que $\phi(t) > 0$,

$$\mathbb{P}\{Z > t\} \leq \mathbb{P}\{\phi(Z) > \phi(t)\} \leq \frac{\mathbb{E}\phi(Z)}{\phi(t)}.$$

Nous retrouvons ainsi, par exemple, l'inégalité de Chebyshev (en appliquant l'inégalité précédente avec $\phi(t) = t^2$ et $Z = |X - \mathbb{E}X|$ où X est une variable aléatoire quelconque)

$$\mathbb{P}\{|X - \mathbb{E}X| > t\} \leq \frac{\text{Var}[X]}{t^2}.$$

Méthode de Cramér-Chernoff

En prenant $\phi(t) = e^{\lambda t}$ avec $\lambda > 0$, nous obtenons une inégalité de Markov exponentielle, pour tout $t > 0$

$$\mathbb{P}\{Z > t\} = \mathbb{P}\{e^{\lambda Z} > e^{\lambda t}\} \leq \frac{\mathbb{E}e^{\lambda Z}}{e^{\lambda t}}.$$

Soit $F_Z: \lambda \in \mathbb{R} \mapsto \mathbb{E}[e^{\lambda Z}]$ la fonction génératrice des moments d'une variable aléatoire Z et notons $L_Z: \lambda \in \mathbb{R}_+ \mapsto \ln \mathbb{E}[e^{\lambda Z}]$. Ainsi, pour tout $\lambda > 0$,

$$\mathbb{P}\{Z > t\} \leq \exp\{-(\lambda t - L_Z(\lambda))\}.$$

2. 'a random variable that smoothly depends on the influence of many independent random variables satisfies Chernoff type bounds' (Talagrand, 1995).

Puis, en optimisant en λ ,

$$\mathbb{P}\{Z > t\} \leq \exp \left\{ - \sup_{\lambda \geq 0} (\lambda t - L_Z(\lambda)) \right\}.$$

En notant $L_Z^*(t) = \sup_{\lambda \geq 0} (\lambda t - L_Z(\lambda))$, la transformée de Cramér de L_Z , on peut énoncer le résultat suivant

Théorème 1.58 (INÉGALITÉ DE CHERNOFF). *Pour tout $t > 0$,*

$$\mathbb{P}\{Z > t\} \leq \exp \{-L_Z^*(t)\}.$$

Exemple 1.59 (VARIABLES ALÉATOIRES GAUSSIENNES). Soit Z une variable aléatoire gaussienne centrée et de variance σ^2 . Alors, pour tout $\lambda \in \mathbb{R}$,

$$L_Z(\lambda) = \ln \left(\int_{\mathbb{R}} e^{\lambda x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx \right) = \ln \left(e^{\lambda^2 \sigma^2 / 2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-(x-2\sigma^2\lambda)/2\sigma^2} dx \right) = \frac{\lambda^2 \sigma^2}{2},$$

et

$$L_Z^*(t) = \frac{t^2}{2\sigma^2}.$$

L'inégalité de Chernoff implique alors que, pour tout $t > 0$,

$$\mathbb{P}\{Z \geq t\} \leq e^{-t^2/2\sigma^2}.$$

De façon plus générale, la méthode de Cramér-Chernoff peut être décomposée en deux étapes :

1. Majorer L_Z par une fonction g de λ pour tout λ .
2. Optimiser $\lambda t - g(\lambda)$ en λ pour obtenir une minoration de la transformée de Cramér.

Nous allons voir deux exemples d'application : les distributions sous-gaussiennes et sous-gamma.

Distributions sous-gaussiennes

Beaucoup de lois de probabilité possèdent des queues de probabilité décroissant au moins aussi rapidement que la loi de Gauss, ce qui peut être formalisé par la définition suivante.

Définition 1.60 (VARIABLE ALÉATOIRE SOUS-GAUSSIENNE). *Une variable aléatoire centrée Z est dite sous-gaussienne, de facteur de variance $v > 0$, si pour tout $\lambda \in \mathbb{R}$,*

$$L_Z(\lambda) \leq \frac{\lambda^2 v}{2}.$$

Les distributions sous-gaussiennes sont donc celles dont la fonction génératrice des moments est majorée par celle des variables aléatoires gaussiennes.

Il est facile de voir que si Z est une variable sous-gaussienne de facteur de variance v alors

$$\exp \{ - (\lambda t - L_Z(\lambda)) \} \leq \exp \left\{ - \left(\lambda t - \frac{\lambda^2 v}{2} \right) \right\},$$

et donc par la méthode de Chernoff,

$$\mathbb{P}\{Z \geq t\} \leq e^{-t^2/2v}.$$

Distributions sous-gamma

Définition 1.61 (VARIABLE ALÉATOIRE SOUS-GAMMA). Une variable aléatoire Z est dite sous-gamma à droite, de facteur de variance $v > 0$ et de paramètre d'échelle $c > 0$, si pour tout $0 < \lambda < 1/c$,

$$L_Z(\lambda) \leq \frac{v\lambda^2}{2(1-c\lambda)}.$$

Z est dite sous-gamma à gauche, de facteur de variance v et de paramètre d'échelle c si $-Z$ est sous-gamma à droite de facteur de variance v et de paramètre d'échelle c .

Exemple 1.62 (LOI GAMMA). Soit Y une variable aléatoire distribuée selon une loi gamma de paramètres a et b . Alors $X = Y - \mathbb{E}Y$ est sous-gamma à droite de paramètre de variance $v = ab^2$ et de paramètre d'échelle $c = b$. En effet, Y a pour densité

$$f_Y(x) = \frac{x^{a-1}e^{-x/b}}{\Gamma(a)b^a}, \quad x \geq 0$$

et $\mathbb{E}Y = ab$.

D'où, pour tout $0 < \lambda < 1/b$

$$\mathbb{E}[e^{\lambda X}] = \int_0^\lambda e^{\lambda(x-ab)} f_Y(x) dx = \exp(-\lambda ab - a \ln(1 - \lambda b)).$$

En remarquant que pour tout $0 < u < 1$,

$$-\ln(1-u) - u \leq \frac{u^2}{2(1-u)},$$

on a

$$L_X(\lambda) \leq \frac{ab^2\lambda^2}{2(1-b\lambda)}.$$

X est donc bien sous-gamma à droite de facteur de variance $v = ab^2$ et de paramètre d'échelle $c = b$.

Exemple 1.63. En calculant puis en majorant sa transformée de Laplace, on vérifie que la k^{e} statistique d'ordre d'un échantillon exponentiel de paramètre 1 est une variable aléatoire sous-gamma à droite de paramètre d'échelle 1 et de paramètre de variance $\frac{1}{k}(1 + \frac{1}{k})$.

Appliquons la méthode de Cramér-Chernoff à Z , variable aléatoire sous-gamma à droite. Par définition, nous connaissons déjà une borne pour L_Z , pour tout $0 < \lambda < 1/c$,

$$L_Z(\lambda) \leq \frac{v\lambda^2}{2(1-c\lambda)}.$$

Pour tout $t > 0$,

$$\begin{aligned} L_Z^*(t) &= \sup_{0 < \lambda < 1/c} (\lambda t - L_Z(\lambda)) \\ &\geq \sup_{0 < \lambda < 1/c} \left(\lambda t - \frac{v\lambda^2}{2(1-c\lambda)} \right) \\ &= \frac{v}{c} h_1 \left(\frac{ct}{v^2} \right) \end{aligned}$$

où $h_1(u) = 1 + u - \sqrt{1 + 2u}$, $u > 0$.

Ainsi, par la méthode de Chernoff, pour tout $t > 0$

$$\mathbb{P}\{Z > t\} \leq \exp(-L_Z^*(t)) \leq \exp\left(-\frac{v}{c}h_1\left(\frac{ct}{v^2}\right)\right).$$

Cette borne peut être réécrite sous la forme

$$\mathbb{P}\left\{Z > \sqrt{2vt} + ct\right\} \leq \exp(-t)$$

en vérifiant que $h_1^{-1}(u) = \sqrt{2vu} + cu$.

Bernstein (1946) a démontré que les sommes de variables aléatoires indépendantes satisfaisant une condition de moments réalisaient cette borne. Le théorème suivant peut être trouvée dans Birgé & Massart (1998).

Théorème 1.64 (INÉGALITÉ DE BERNSTEIN). *Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes telles qu'il existe des réels $v, c > 0$ tels que*

$$\begin{aligned} & - \sum_{i=1}^n \mathbb{E}[X_i^2] \leq v \\ & - \sum_{i=1}^n \mathbb{E}[(X_i)_+^q] \leq \frac{q!}{2}vc^{q-2} \quad \forall q \geq 3. \end{aligned}$$

Si $S = \sum_{i=1}^n (X_i - \mathbb{E}X_i)$ alors pour tout $0 < \lambda < 1/c$

$$L_S(\lambda) \leq \frac{v\lambda^2}{2(1-c\lambda)}$$

et pour tout $t > 0$,

$$L_S^*(t) \geq \frac{v}{c^2}h_1\left(\frac{ct}{v}\right).$$

En particulier, pour tout $t > 0$,

$$\mathbb{P}\{S > \sqrt{2vt} + ct\} \leq e^{-t}.$$

Ainsi, nous dirons qu'une variable aléatoire Z satisfait une inégalité de type Bernstein de facteur de variance v et de paramètre d'échelle c si pour tout $t > 0$,

$$\mathbb{P}\left\{Z > \sqrt{2vt} + ct\right\} \leq \exp(-t),$$

ou

$$\mathbb{P}\{Z > t\} \leq \exp\left(-\frac{t^2}{2(v+ct)}\right).$$

4.2 Majoration de la variance

Inégalité d'Efron-Stein

L'inégalité d'Efron-Stein est une façon simple mais puissante de majorer la variance d'une fonction générale g de plusieurs variables aléatoires indépendantes X_1, \dots, X_n . Elle propose une borne en fonction des fluctuations locales de la fonction g .

Elle a d'abord été prouvée en supposant g symétrique par Efron & Stein (1981) et dans le cas général par Steele (1986).

Pour tout échantillon \mathbf{X}^n , notons $\mathbf{X}^{(i)}$, $1 \leq i \leq n$, le sous-échantillon où l'on a supprimé X_i , c'est-à-dire, $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$; et $\mathbb{E}^{(i)}$ l'espérance conditionnelle sachant $\mathbf{X}^{(i)}$.

Soit $Z = g(\mathbf{X}^n)$ où $g: \mathbb{R}^n \rightarrow \mathbb{R}$ mesurable.

Théorème 1.65 (INÉGALITÉ D'EFRON-STEIN (Boucheron et al., 2013)). *Si Z est de carré intégrable, alors*

1.

$$\text{Var}[Z] \leq \sum_{i=1}^n \mathbb{E} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right] \stackrel{\text{def}}{=} v.$$

2. Si X'_1, \dots, X'_n sont des copies indépendantes de X_1, \dots, X_n et si

$$Z'_i = g(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n),$$

alors

$$v = \frac{1}{2} \sum_{i=1}^n \mathbb{E} [(Z - Z'_i)^2] = \sum_{i=1}^n \mathbb{E} [(Z - Z'_i)_+^2] = \sum_{i=1}^n \mathbb{E} [(Z - Z'_i)_-^2]$$

où $x_+ = \max(x, 0)$ et $x_- = \max(-x, 0)$ sont les parties positive et négative de x .

3.

$$v = \inf_{Z_i} \sum_{i=1}^n \mathbb{E} [(Z - Z_i)^2]$$

où la borne inférieure est prise sur l'ensemble des variables Z_i qui sont $\mathbf{X}^{(i)}$ -mesurables et de carré intégrable, $i = 1, \dots, n$.

$V := \sum_{i=1}^n (Z - \mathbb{E}^{(i)} Z)^2$ est appelée l'estimée Efron-Stein de la variance de Z . Ainsi, $v = \mathbb{E}V$.

Remarque 1.66. L'inégalité d'Efron-Stein reste valide pour des variables aléatoires X_1, \dots, X_n indépendantes, mais non nécessairement identiquement distribuées.

Remarque 1.67. Il existe au moins deux preuves de cette inégalité : la première utilise un argument de martingales et la deuxième repose sur la formule de dualité de la variance. La deuxième preuve permet de faire le lien avec d'autres résultats comme l'inégalité de Sobolev logarithmique modifiée que nous verrons un peu plus loin.

Exemple 1.68 (SOMME DES VARIABLES ALÉATOIRES INDÉPENDANTES). Soit $Z = \sum_{i=1}^n X_i$ où X_1, \dots, X_n sont des variables aléatoires indépendantes (non nécessairement identiquement distribuées). Alors $\text{Var}[Z] = \sum_{i=1}^n \text{Var}[X_i]$.

Soit X'_1, \dots, X'_n des copies indépendantes de X_1, \dots, X_n et définissons $Z'_i = \sum_{j \neq i} X_j + X'_i$. Alors l'inégalité d'Efron-Stein donne

$$\text{Var}[Z] \leq \sum_{i=1}^n \mathbb{E} \left[(X_i - X'_i)_+^2 \right] = \sum_{i=1}^n \text{Var}[X_i].$$

Ainsi, pour une somme de variables aléatoires indépendantes, l'inégalité d'Efron-Stein est une égalité. Elle est donc tendue en ce sens.

L'inégalité Efron-Stein a été motivée par l'étude de l'estimée jackknife de la variance (Miller, 1974; Shao & Wu, 1989). Lorsqu'un paramètre θ de F est estimé par une fonction $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ des données X_1, \dots, X_n , la qualité de $\hat{\theta}$ peut être mesurée par son biais $\mathbb{E}\hat{\theta} - \theta$ et par sa variance $\text{Var}\hat{\theta}$. La fonction F n'étant pas connue, le biais et la variance doivent également être estimés. L'estimateur jackknife du biais est de la forme

$$\frac{n-1}{n} \sum_{i=1}^n \hat{\theta}_i - \hat{\theta}$$

où $\hat{\theta}_i$ est une fonction bien définie de $\mathbf{X}^{(i)}$, et l'estimateur jackknife de la variance est de la forme

$$\sum_{i=1}^n (\hat{\theta}_i - \hat{\theta})^2.$$

L'inégalité d'Efron-Stein affirme que le biais de l'estimateur jackknife de la variance est toujours positif.

Dans cette thèse, nous appliquerons l'inégalité d'Efron-Stein sous la forme

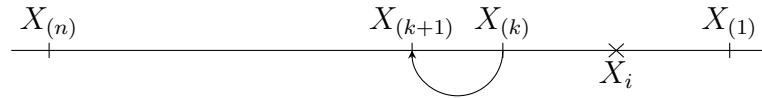
$$\text{Var}[Z] \leq \mathbb{E} \left[\sum_{i=1}^n (Z - Z_i)^2 \right]$$

où $Z_i = g_i(\mathbf{X}^{(i)})$ avec $g_i: \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ une fonction mesurable arbitraire. L'idée est de choisir Z_i assez proche de Z pour que les différences $Z - Z_i$ ne soient pas trop grandes et obtenir ainsi des bornes correctes de la variance de Z .

Exemple 1.69 (STATISTIQUES D'ORDRE ET ESPACEMENTS (Proposition 2.3)). Soit k fixé et Z la k^{e} statistique d'ordre de l'échantillon \mathbf{X}^n .

Pour $k \leq n/2$, on définit Z_i comme la k^{e} statistique d'ordre du sous-échantillon $\mathbf{X}^{(i)}$: si $X_i < X_{(k)}$ alors les k premières statistiques d'ordre ne changent pas et $X_{(k)}$ reste inchangée ;

En revanche, si $X_i \geq X_{(k)}$ alors la k^{e} statistique d'ordre du sous-échantillon $\mathbf{X}^{(i)}$ est égale à $X_{(k+1)}$.



$$Z_i = \begin{cases} X_{(k+1)} & \text{si } X_i \geq X_{(k)} \\ Z & \text{sinon.} \end{cases}$$

D'où, d'après l'inégalité d'Efron-Stein, pour tout $k \leq n/2$

$$\text{Var}[Z] \leq k \mathbb{E} [(X_{(k)} - X_{(k+1)})^2] = k \mathbb{E} [\Delta_k^2]. \quad (1.19)$$

Pour $k > n/2$, on définit Z_i comme la $(k-1)^{\text{e}}$ statistique d'ordre du sous-échantillon $\mathbf{X}^{(i)}$, et on obtient de la même façon

$$\text{Var}[Z] \leq (n-k+1) \mathbb{E} [(X_{(k-1)} - X_{(k)})^2]. \quad (1.20)$$

Nous pouvons évaluer la qualité cette inégalité sur les deux exemples des variables exponentielles et des variables uniformes.

1. Variables uniformes

D'une part, d'après la proposition 1.6, pour $k \leq n/2$

$$\text{Var} [V_{(k)}] = \left(1 - \frac{k}{n+1}\right)$$

et d'autre part, d'après le proposition 1.7,

$$k\mathbb{E} \left[(V_{(k)} - V_{(k+1)})^2 \right] = k\mathbb{E} [V_{(n)}^2] = \frac{2k}{(n+1)(n+2)}.$$

2. Variables exponentielles

D'après la représentation de Rényi (Théorème 1.8) et l'équation (1.1), nous avons d'une part,

$$\frac{1}{k} + \frac{1}{n^2} \leq \text{Var} [Y_{(k)}] \leq \frac{1}{k} + \frac{1}{k^2},$$

et d'autre part la borne de l'équation (1.19) vaut

$$k\mathbb{E} [(Y_{(k)} - Y_{(k+1)})^2] = k\mathbb{E} \left[\left(\frac{E_k}{k} \right)^2 \right] = \frac{2}{k}.$$

Sur ces deux exemples, la borne obtenue par l'inégalité d'Efron-Stein est du bon ordre de grandeur à un facteur plus petit que 4 près.

Nous pouvons également appliquer l'inégalité d'Efron-Stein à l'estimateur de Hill (Définition 1.39), vu comme une fonctionnelle des $(k+1)$ plus grandes statistiques d'ordre de l'échantillon.

Exemple 1.70 (VARIANCE DE L'ESTIMATEUR DE HILL). Soit $\hat{\gamma}_H(k)$ l'estimateur de Hill calculé à partir des $(k+1)$ plus grandes statistiques d'ordre, et $\hat{v}_H(k)$ l'estimée Efron-Stein de la variance de $\hat{\gamma}_H(k)$, alors

$$\hat{v}_H(k) = \frac{M_n^{(2)}}{k} + (k+1) \left(\ln \frac{X_{(k+1)}}{X_{(k+2)}} \right)^2 - 2 \ln \frac{X_{(k+1)}}{X_{(k+2)}} M_n^{(1)},$$

où $M_n^{(1)}$ et $M_n^{(2)}$ ont été définis à la définition 1.51, et

$$\text{Var} [\hat{\gamma}_H(k)] \leq \mathbb{E} \hat{v}_H(k).$$

Démonstration. Pour chaque $i \in \{1, \dots, n\}$, définissons Z_i comme l'estimateur de Hill calculé à partir des $(k+1)$ plus grandes statistiques d'ordre du sous-échantillon $\mathbf{X}^{(i)}$, donc si $X_i = X_{(\ell+1)}$,

$$Z_i = \begin{cases} Z & \text{si } \ell > k+1 \\ \frac{1}{k} \sum_{j < k+1} \ln \frac{X_{(j)}}{X_{(k+2)}} & \text{si } \ell = k+1 \\ \frac{1}{k} \sum_{\substack{j \neq \ell \\ j \leq k+1}} \ln \frac{X_{(j)}}{X_{(k+2)}} + \frac{1}{k} \ln \frac{X_{(k+1)}}{X_{(k+2)}} & \text{si } \ell < k+1 \end{cases}$$

et

$$Z - Z_i = \frac{1}{k} \ln \frac{X_{(\ell)}}{X_{(k+1)}} + \ln \frac{X_{(k+2)}}{X_{(k+1)}} \text{ pour } \ell \geq k+1.$$

D'où,

$$\begin{aligned}
\widehat{v}_H(k) &= \sum_{i=1}^n (Z - Z_i)^2 \\
&= \sum_{\ell=1}^{k+1} \left(\frac{1}{k} \ln \frac{X_{(\ell)}}{X_{(k+1)}} + \ln \frac{X_{(k+2)}}{X_{(k+1)}} \right)^2 \\
&= \sum_{\ell=1}^{k+1} \frac{1}{k^2} \left(\ln \frac{X_{(\ell)}}{X_{(k+1)}} \right)^2 + (k+1) \left(\ln \frac{X_{(k+1)}}{X_{(k+2)}} \right)^2 - 2 \ln \frac{X_{(k+1)}}{X_{(k+2)}} \sum_{\ell=1}^{k+1} \frac{1}{k} \left(\ln \frac{X_{(\ell)}}{X_{(k+1)}} \right) \\
&= \frac{M_n^{(2)}}{k} + (k+1) \left(\ln \frac{X_{(k+1)}}{X_{(k+2)}} \right)^2 - 2 \ln \frac{X_{(k+1)}}{X_{(k+2)}} M_n^{(1)}.
\end{aligned}$$

Ainsi, par l'inégalité d'Efron-Stein (Théorème 1.65)

$$\text{Var}[\widehat{\gamma}_H(k)] \leq \frac{1}{k} \mathbb{E} [M_n^{(2)}] + (k+1) \mathbb{E} \left[\left(\ln \frac{X_{(k+1)}}{X_{(k+2)}} \right)^2 \right] - 2 \mathbb{E} \left[\ln \frac{X_{(k+1)}}{X_{(k+2)}} M_n^{(1)} \right].$$

□

Inégalités de Poincaré

Nous présentons ici deux inégalités de Poincaré : la première est valide pour les vecteurs gaussiens et la seconde pour les vecteurs exponentiels.

Théorème 1.71 (INÉGALITÉ DE POINCARÉ (CAS GAUSSIEN)). *Soit $g: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continûment différentiable et \mathbf{X}^n un vecteur de n variables aléatoires gaussiennes centrées réduites indépendantes. Alors*

$$\text{Var}[g(\mathbf{X}^n)] \leq \mathbb{E} [\|\nabla g(\mathbf{X}^n)\|^2].$$

où $\|\cdot\|$ désigne la norme euclidienne.

Remarque 1.72. Cette inégalité peut être prouvée grâce à l'inégalité d'Efron-Stein (Ané et al., 2000).

Corollaire 1.73. *Si on suppose de plus que g est L -lipschitzienne, alors*

$$\text{Var}[g(\mathbf{X}^n)] \leq L^2.$$

Exemple 1.74 (APPLICATION AUX STATISTIQUES D'ORDRE D'UN ÉCHANTILLON GAUSSIEN). Soit X_1, \dots, X_n n variables aléatoires gaussiennes centrées réduites indépendantes.

La fonction qui à n variables associe la k^e plus grande est 1-lipschitzienne, donc d'après le corollaire précédent, pour tout $1 \leq k \leq n$

$$\text{Var} [X_{(k)}] \leq 1.$$

Or nous avons vu à la partie précédente (Exemple 1.34) que la variance du maximum de n variables aléatoires gaussiennes centrées réduites indépendantes était de l'ordre de $1/\ln n$. Dans la partie 4 du chapitre 2, nous verrons comment obtenir des bornes du bon ordre de grandeur pour des échantillons gaussiens.

Théorème 1.75 (INÉGALITÉ DE POINCARÉ (CAS EXPONENTIEL) (BOBKOV & LE-DOUX, 1997)). Soit $\mathbf{Y}^n = (Y_1, \dots, Y_n)$ un vecteur de n variables aléatoires exponentielles de paramètre 1 indépendantes, et g une fonction continûment différentiable. Alors

$$\text{Var}[g(\mathbf{Y}^n)] \leq 4\mathbb{E} [\|\nabla g(\mathbf{Y}^n)\|^2] .$$

Pour l'exemple suivant, nous aurons besoin de la propriété suivante.

Définition 1.76 (TAUX DE HASARD). Si F admet une densité f par rapport à la mesure de Lebesgue, son taux de hasard h est défini par $h = f/\bar{F}$.

Proposition 1.77. Si la U -transformée de F est différentiable, et si on note h le taux de hasard associé, alors

$$(U \circ \exp)' = \frac{1}{h(U \circ \exp)} .$$

Exemple 1.78 (APPLICATION AUX STATISTIQUES D'ORDRE (Corollaire 3.6)). D'après la proposition 1.12,

$$\text{Var}[X_{(k)}] = \text{Var}[U(e^{Y_{(k)}})] = \text{Var} \left[U \left(e^{\sum_{i=k}^n \frac{E_i}{i}} \right) \right] .$$

En supposant que la fonction U est différentiable, nous pouvons appliquer l'inégalité de Poincaré pour les variables exponentielles avec $g(E_k, \dots, E_n) = U \left(e^{\sum_{i=k}^n \frac{E_i}{i}} \right)$.

Pour $k \leq i \leq n$,

$$\begin{aligned} \partial_i g &= \frac{1}{i} \frac{1}{h \left(U \left(e^{\sum_{i=k}^n \frac{E_i}{i}} \right) \right)} \\ &\stackrel{d}{=} \frac{1}{i} \frac{1}{h(X_{(k)})} \end{aligned}$$

et

$$\|\nabla g\|^2 = \sum_{i=k}^n \frac{1}{i^2} \frac{1}{h(X_{(k)})^2} \leq \frac{1}{k} \left(1 + \frac{1}{k} \right) \frac{1}{h(X_{(k)})^2} .$$

Nous obtenons donc

$$\text{Var}[X_{(k)}] \leq \frac{4}{k} \left(1 + \frac{1}{k} \right) \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right] . \quad (1.21)$$

Si F a un taux de hasard croissant, le facteur 4 peut être remplacé par un facteur 2 (Proposition 2.10).

Dans le cas gaussien, cette inégalité permet d'obtenir des bornes tendues de la variance des statistiques d'ordre d'un échantillon gaussien (Proposition 2.14).

Nous pouvons également utiliser l'inégalité de Poincaré pour les variables aléatoires exponentielles afin d'obtenir une borne de la majoration de la variance de l'estimateur de Hill : d'après la représentation (1.14), celui-ci peut être vu comme une fonction régulière de E_1, \dots, E_k , conditionnellement à $Y_{(k+1)}$.

Exemple 1.79 (BORNES POUR LA VARIANCE DE L'ESTIMATEUR DE HILL (Proposition 3.11)). Si $F \in \text{DA}(\gamma)$, $\gamma > 0$, et satisfait la condition de von Mises (Condition 1.7).

$$-\frac{2\gamma}{k} \mathbb{E} [\bar{\eta}(e^{Y_{(k+1)}})] \leq \text{Var}[\hat{\gamma}_H(k)] - \frac{\gamma^2}{k} \leq \frac{2\gamma}{k} \mathbb{E} [\bar{\eta}(e^{Y_{(k+1)}})] + \frac{5}{k} \mathbb{E} [\bar{\eta}(e^{Y_{(k+1)}})^2].$$

En effet, posons $Z = k\hat{\gamma}_H(k)$. Par la relation de Pythagore,

$$\text{Var}(Z) = \mathbb{E} [\text{Var}(Z | Y_{(k+1)})] + \text{Var}(\mathbb{E}[Z | Y_{(k+1)}]).$$

La représentation (1.14) affirme que, conditionnellement à $Y_{(k+1)}$, Z est distribuée comme une somme de variables aléatoires exponentielles indépendantes de paramètre 1 E_1, \dots, E_k .

$$\begin{aligned} \text{Var}(Z | Y_{(k+1)} = y) &= k \text{Var}\left(\gamma E + \int_0^E \eta(e^{y+u}) du\right) \\ &= k\gamma^2 + 2k\gamma \text{Cov}\left(E, \int_0^E \eta(e^{y+u}) du\right) + k \text{Var}\left(\int_0^E \eta(e^{y+u}) du\right) \\ &\leq k\gamma^2 + 2k\gamma \bar{\eta}(e^y) + k(\bar{\eta}(e^y))^2, \end{aligned}$$

où nous avons appliqué l'inégalité de Cauchy-Schwarz et $\text{Var}(\int_0^E \eta(e^{y+u}) du) \leq \bar{\eta}(e^y)^2$.

En prenant l'espérance par rapport à $Y_{(k+1)}$, nous obtenons

$$\mathbb{E} [\text{Var}(Z | Y_{(k+1)})] \leq k\gamma^2 + 2k\gamma \mathbb{E} [\bar{\eta}(e^{Y_{(k+1)}})] + k \mathbb{E} [\bar{\eta}(e^{Y_{(k+1)}})^2].$$

Le second terme peut s'écrire sous la forme suivante :

$$\mathbb{E}[Z | Y_{(k+1)}] - \mathbb{E}Z = k \int_0^\infty (\eta(e^{u+Y_{(k+1)}}) - \mathbb{E}[\eta(e^{u+Y_{(k+1)}})]) e^{-u} du.$$

Comme $Y_{(k+1)}$ est une fonction de variables aléatoires exponentielles indépendantes ($Y_{(k+1)} = \sum_{i=k+1}^n E_i/i$), la variance de $\mathbb{E}[Z | Y_{(k+1)}]$ peut être majorée par l'inégalité de Poincaré pour les vecteurs exponentiels

$$\text{Var}(\mathbb{E}[Z | Y_{(k+1)}]) \leq 4k \mathbb{E} [\bar{\eta}(e^{Y_{(k+1)}})^2].$$

Afin d'établir une borne inférieure, observons que

$$\text{Var}(Z) \geq \mathbb{E} [\text{Var}(Z | Y_{(k+1)})].$$

En appliquant de nouveau l'inégalité de Cauchy-Schwarz,

$$\begin{aligned} \text{Var}(Z | Y_{(k+1)} = y) &\geq k\gamma^2 - 2k\gamma \left| \text{Cov}\left(E, \int_0^E \eta(e^{y+u}) du\right) \right| \\ &\geq k\gamma^2 - 2k\gamma \left(\text{Var}\left(\int_0^E \eta(e^{y+u}) du\right) \right)^{1/2} \\ &\geq k\gamma^2 - 2k\gamma \bar{\eta}(e^y). \end{aligned}$$

4.3 Inégalité exponentielle d'Efron-Stein

A partir de l'inégalité d'Efron-Stein, il est possible d'obtenir des inégalités exponentielles reliant le logarithme de la transformée de Laplace d'une variable aléatoire et la transformée de Laplace de l'estimée Efron-Stein de sa variance.

Théorème 1.80 (INÉGALITÉ EXPONENTIELLE D'EFRON-STEIN). *Soit V l'estimée Efron-Stein de la variance de la variable aléatoire $Z = g(X_1, \dots, X_n)$, où X_1, \dots, X_n sont des variables aléatoires indépendantes, alors Z vérifie une inégalité exponentielle d'Efron-Stein, c'est-à-dire que pour tout $\theta, \lambda > 0$ tels que $\lambda\theta < 1$ et $\mathbb{E}[e^{\lambda V/\theta}] < \infty$,*

$$\ln \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)}] \leq \frac{\lambda\theta}{1 - \lambda\theta} \ln \mathbb{E}[e^{\lambda V/\theta}].$$

Cherchons à appliquer cette inégalité à $Z = Y_{(k)}$ la k^e statistique d'ordre d'un échantillon exponentiel. Alors $V = k\Delta_k^2 \stackrel{d}{=} \frac{E^2}{k}$ où E est une variable aléatoire exponentielle de paramètre 1. Or, E^2 n'admet pas de moment exponentiel, donc $\ln \mathbb{E} [e^{\lambda V/\theta}] = \infty$. L'inégalité exponentielle d'Efron-Stein n'apporte donc aucune information, ce qui est dû au fait que la démonstration de cette inégalité utilise une inégalité de découplage trop grossière basée sur la formule variationnelle de l'entropie. Nous allons voir comment cette difficulté peut être contournée grâce à la méthode entropique et à l'association négative.

4.4 Méthode entropique

La méthode entropique permet d'établir des inégalités exponentielles pour la queue de probabilité de fonctions générales de plusieurs variables aléatoires indépendantes. Elle peut être décomposée en deux étapes : la première étape consiste à prouver une inégalité de Sobolev logarithmique et la deuxième étape à résoudre une inéquation différentielle obtenue par l'argument de Herbst. Le point de départ de la méthode entropique est la sous-additivité de l'entropie. Les principes clés de la méthode reposent sur la preuve d'inégalités de concentration gaussiennes à partir d'inégalités de Sobolev logarithmiques. Ledoux (1997) a montré que ces idées pouvaient être utilisées afin d'établir des inégalités de concentration exponentielles à la Talagrand pour des processus empiriques. Par la suite, en suivant ces mêmes idées, Boucheron et al. (2013); Bousquet (2002); Klein (2002); Klein & Rio (2005); Massart (2000); Rio (2001) ont obtenu des inégalités de concentration tendues pour des processus empiriques et les fonctions auto-bornées.

Entropie

Définition 1.81 (ENTROPIE). *On définit, et on note $\text{Ent}[X]$, l'entropie d'une variable aléatoire X positive et intégrable par*

$$\text{Ent}[X] = \mathbb{E}[X \ln X] - \mathbb{E}X \ln \mathbb{E}X.$$

Proposition 1.82 (SOUS-ADDITIVITÉ DE L'ENTROPIE). *Soit $g: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction mesurable. Posons $Z = g(\mathbf{X}^n)$, alors*

$$\text{Ent}[Z] \leq \mathbb{E} \left[\sum_{i=1}^n \text{Ent}^{(i)}[Z] \right]$$

où $\text{Ent}^{(i)} = \mathbb{E}^{(i)}[Z \ln Z] - \mathbb{E}^{(i)} Z \ln \mathbb{E}^{(i)} Z$.

L'idée principale de la méthode entropique est d'appliquer la sous-additivité de l'entropie à la variable aléatoire positive $T = e^{\lambda Z}$ où λ est un nombre réel, ainsi

$$\mathbb{E}[T \ln T] - \mathbb{E}T \ln \mathbb{E}T \leq \sum_{i=1}^n \mathbb{E} \left[\mathbb{E}^{(i)}[T \ln T] - \mathbb{E}^{(i)}T \ln \mathbb{E}^{(i)}T \right]. \quad (1.22)$$

En renormalisant par $\mathbb{E}e^{\lambda Z}$, le membre de gauche de cette inégalité devient

$$\frac{\text{Ent} [e^{\lambda Z}]}{\mathbb{E}e^{\lambda Z}} = \lambda L'_{Z-\mathbb{E}Z}(\lambda) - L_{Z-\mathbb{E}Z}(\lambda).$$

Ainsi, par résolution d'une inéquation différentielle, nous obtenons une borne pour la queue de probabilité de Z via l'inégalité de Chernoff, ce qui est le principe de l'argument de Herbst. Néanmoins, pour cela, nous avons besoin de borner le membre de droite, ce qui sera réalisé par des inégalités de Sobolev logarithmiques.

Argument de Herbst

Proposition 1.83 (ARGUMENT DE HERBST). *Soit Z une variable aléatoire intégrable telle que, pour un certain $v > 0$,*

$$\frac{\text{Ent} [e^{\lambda Z}]}{\mathbb{E}e^{\lambda Z}} \leq \frac{\lambda^2 v}{2} \text{ pour tout } \lambda > 0. \quad (1.23)$$

Alors, pour tout $\lambda > 0$,

$$\ln \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \leq \frac{\lambda^2 v}{2}.$$

Plus généralement, l'argument de Herbst consiste en la résolution de l'inéquation différentielle provenant de la sous-additivité de l'entropie et des inégalités de Sobolev logarithmiques. La proposition 1.83 montre que si Z satisfait (1.23) alors Z est une variable aléatoire sous-gaussienne.

Inégalité de Sobolev logarithmique modifiée

Nous présentons ici une inégalité simple qui permet de réécrire la sous-additivité de l'entropie sous une forme plus malléable. Comme pour l'inégalité d'Efron-Stein, définissons $Z = g(\mathbf{X}^n)$, et $Z_i = g_i(\mathbf{X}^{(i)})$ où $g_i: \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ une fonction mesurable arbitraire.

Théorème 1.84 (INÉGALITÉ DE SOBOLEV LOGARITHMIQUE MODIFIÉE (MASSART, 2000; WU, 2000)). *Soit τ la fonction définie sur \mathbb{R} par $\tau(x) = e^x - x - 1$, alors pour tout $\lambda \in \mathbb{R}$,*

$$\text{Ent} [e^{\lambda Z}] = \lambda \mathbb{E} [Z e^{\lambda Z}] - \mathbb{E} [e^{\lambda Z}] \ln \mathbb{E} [e^{\lambda Z}] \leq \sum_{i=1}^n \mathbb{E} [e^{\lambda Z} \tau(-\lambda(Z - Z_i))].$$

Remarque 1.85. La preuve de cette inégalité consiste à majorer de chaque terme du membre de droite de l'inégalité (1.22) en s'appuyant sur la formule variationnelle de l'entropie.

Exemple 1.86 (APPLICATION AUX STATISTIQUES D'ORDRE (Proposition 2.3)). Pour appliquer cette inégalité aux statistiques d'ordre, nous procédons comme pour l'inégalité d'Efron-Stein. Pour $1 \leq k \leq n/2$, Z désigne $X_{(k)}$ et Z_i la k^{e} statistique d'ordre du sous-échantillon $\mathbf{X}^{(i)}$.

Alors, pour tout $\lambda \in \mathbb{R}$,

$$\text{Ent} [e^{\lambda X_{(k)}}] \leq k \mathbb{E} [e^{\lambda X_{(k+1)}} \psi(\lambda \Delta_k)]$$

où ψ est la fonction croissante définie par $\psi(x) = e^x \tau(-x)$.

Afin de pouvoir appliquer l'argument de Herbst, il est nécessaire de découpler les variables $X_{(k+1)}$ et Δ_k , ce qui est possible grâce à l'association négative entre ces deux variables lorsque le taux de hasard de la loi sous-jacente est croissant.

Une famille finie de variables aléatoires (Z_1, \dots, Z_n) est dite négativement associée si pour tous sous-ensembles A_1 et A_2 de $\{1, \dots, n\}$,

$$\text{Cov} [g_1(Z_i, i \in A_1) g_2(Z_i, i \in A_2)] \leq 0$$

où g_1 et g_2 sont des fonctions croissantes coordonnée par coordonnée. Cette définition a été introduite par Alam & Saxena (1981) et Joag-Dev & Proschan (1983).

Proposition 1.87. *Si F a un taux de hasard croissant, alors Δ_k et $X_{(k+1)}$ sont négativement associés : pour toutes fonctions croissantes g_1 et g_2*

$$\mathbb{E} [g_1(X_{(k+1)}) g_2(\Delta_k)] \leq \mathbb{E} [g_1(X_{(k+1)})] \mathbb{E} [g_2(\Delta_k)] .$$

Une preuve de ce résultat se trouve dans le chapitre 2 (Proposition 2.9).

Remarque 1.88. Les lois exponentielle, normale et uniforme, ou plus généralement toutes les lois dont la queue décroît plus vite que celle de la loi exponentielle ont un taux de hasard croissant. Toutes les lois log-concaves, c'est-à-dire les lois dont la densité f est log-concave, admettent un taux de hasard croissant. En revanche, il existe des lois dont le taux de hasard est croissant mais qui ne sont pas log-concaves, par l'exemple la loi définie comme l'image en miroir de la loi de Pareto de paramètres $x_{\min} > 0$ et $\gamma > 0$ dont la fonction de répartition est définie sur $] -\infty, -x_{\min}[$ par $x \mapsto (x/x_{\min})^{-1/\gamma}$.

Une inégalité pour les variables aléatoires exponentielles

Théorème 1.89 (INÉGALITÉ DE TYPE BERNSTEIN POUR LES VARIABLES EXPONENTIELLES (Bobkov & Ledoux, 1997)). *Soit g une fonction différentiable de \mathbb{R}^n avec $\max_i |\partial_i g| < \infty$, et $Z = g(E_1, \dots, E_n)$ où E_1, \dots, E_n sont des variables aléatoires exponentielles indépendantes de paramètre 1. Soit $c < 1$, alors pour tout λ tel que $0 \leq \lambda \max_i |\partial_i g| \leq c$,*

$$\text{Ent} [e^{\lambda(Z - \mathbb{E}Z)}] \leq \frac{2\lambda^2}{1 - c} \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)} \|\nabla g\|^2] .$$

Soit v le supremum essentiel de $\|\nabla g\|^2$, alors pour tout $s \geq 0$,

$$\mathbb{P} \left\{ |Z - \mathbb{E}Z| \geq \sqrt{8vs} + \max_i |\partial_i g| s \right\} \leq 2e^{-s} .$$

Une preuve de ce résultat peut facilement être reconstituée à partir des pages 104-105 de (Ledoux, 2001). Cette preuve est basée sur la méthode entropique : Bobkov & Ledoux (1997) commencent par établir la première inégalité du théorème puis montrent la deuxième à l'aide de l'argument de Herbst et de la borne de Chernoff.

Appliquons ce résultat aux statistiques d'ordre vues comme des fonctions générales de variables aléatoires exponentielles.

Exemple 1.90 (APPLICATION AUX STATISTIQUES D'ORDRE (Corollaire 3.10)). Supposons que F admette un taux de hasard croissant h .

Posons $X_{(k)} \stackrel{d}{=} (U \circ \exp) (\sum_{i=k}^n E_i/i) = g(E_k, \dots, E_n)$, alors pour tout $i \leq k \leq n$,

$$\begin{aligned} \partial_i g &= \frac{1}{i} \frac{1}{h((U \circ \exp) (\sum_{i=k}^n E_i/i))} \\ &= \frac{1}{i} \frac{1}{h(X_{(k)})}. \end{aligned}$$

D'où,

$$|\partial_i g| \leq \frac{1}{i} \sup_x \frac{1}{h(x)}, \text{ pour } i \geq k$$

et

$$\begin{aligned} \|\nabla g\|^2 &= \sum_{i=k}^n \frac{1}{i^2} \frac{1}{(h(X_{(k)}))^2} \\ &\leq \frac{1}{k} \left(1 + \frac{1}{k}\right) \frac{1}{(h(X_{(k)}))^2}. \end{aligned}$$

Soit $c < 1$, alors pour tout $\lambda, 0 \leq \lambda \leq c(k \inf_x h(x))$,

$$\ln \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)}] \leq \frac{\lambda^2}{2(1-c)} \frac{4}{k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right].$$

En effet, d'après la première inégalité de 3.9,

$$\begin{aligned} \text{Ent} [e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}] &\leq \frac{2\lambda^2}{1-c} \mathbb{E} [e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})} \|\nabla f\|^2] \\ &= \frac{4\lambda^2}{2(1-c)} \frac{1}{k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}}{h(X_{(k)})^2} \right] \\ &\leq \frac{4\lambda^2}{2(1-c)} \frac{1}{k} \left(1 + \frac{1}{k}\right) \mathbb{E} [e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}] \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right] \end{aligned}$$

où la dernière inégalité est obtenue par l'inégalité d'association négative de Chebychev.

D'où,

$$\frac{d \frac{1}{\lambda} \ln \mathbb{E} [e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}]}{d\lambda} = \frac{\text{Ent} [e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}]}{\lambda^2 \mathbb{E} [e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}]} \leq \frac{1}{2(1-c)} \frac{4}{k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right],$$

et

$$\ln \mathbb{E} [e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})}] \leq \frac{\lambda^2}{2(1-c)} \frac{4}{k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right].$$

Ainsi,

$$\mathbb{P} \left\{ |X_{(k)} - \mathbb{E}X_{(k)}| \geq \sqrt{\frac{8}{k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right]^s + \frac{s}{k \inf_x h(x)}} \right\} \leq 2e^{-s}.$$

Autrement dit, $X_{(k)}$ est sous-gamma à gauche et à droite de facteur de variance $\frac{4}{k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right]$ et de facteur d'échelle $\frac{1}{k \inf_x h(x)}$. Notons que le facteur de variance est égale à la borne (1.21) obtenue pour la variance de $X_{(k)}$.

4.5 Une inégalité maximale

Théorème 1.91. (Boucheron et al., 2013) Soit Z_1, \dots, Z_n des variables aléatoires telles que pour tout λ appartenant à un certain intervalle $(0, b)$ et tout $i = 1, \dots, N$

$$L_{Z_i}(\lambda) \leq L(\lambda)$$

où L est une fonction convexe et continûment différentiable sur $[0, b)$ telle que $L(0) = L'(0) = 0$. Alors

$$\mathbb{E} \max_{i=1, \dots, N} Z_i \leq L^{*-1}(\ln N).$$

La preuve de cette inégalité maximale est basée sur un argument utilisé par Pisier (1983) pour contrôler l'espérance du supremum de variables appartenant à un certain espace d'Orlicz.

On peut appliquer cette inégalité aux variables sous-gaussiennes et sous-gamma :

1. Variables sous-gaussiennes

Si les Z_i sont sous-gaussiennes de facteur de variance v , alors :

$$\mathbb{E} \max_{i=1, \dots, N} Z_i \leq \sqrt{2v \ln N}.$$

2. Variables sous-gamma

Si les Z_i sont sous-gamma de facteur de variance v et de facteur d'échelle c , alors :

$$\mathbb{E} \max_{i=1, \dots, N} Z_i \leq \sqrt{2v \ln N} + c \ln N.$$

Dans le chapitre 3 (Théorème 3.13), nous utilisons cette dernière inégalité pour des variables sous-gamma afin de majorer l'espérance du supremum du processus $(\sqrt{k}(\widehat{\gamma}_H(k) - \mathbb{E}\widehat{\gamma}_H(k)))_k$. Nous montrons également que ce processus est sous-gamma à droite et à gauche par un argument de type chainage (Talagrand, 2005). En effet, le théorème 1.91 est souvent utilisé dans des arguments dits de chainage. Ces arguments permettent d'obtenir des bornes tendues pour les suprema de processus empiriques. Dans cette thèse, nous utilisons la technique de « peeling » dont l'idée est de découper le processus empirique en sous-familles de variables aléatoires ayant des propriétés similaires puis d'appliquer, par exemple, le théorème 1.91.

Chapitre 2

Concentration inequalities for order statistics

Ce chapitre a fait l'objet d'un article publié dans Electronic Communications in Probability (Boucheron & Thomas, 2012).

Contents

1	Introduction	57
2	Order statistics and spacings	59
3	Asymptotic assessment	63
4	Order statistics of Gaussian samples	64
5	Appendix	70
5.1	Proof of Proposition 2.12	70
5.2	Proof of Proposition 2.16	72

This chapter describes non-asymptotic variance and tail bounds for order statistics of samples of independent identically distributed random variables. When the sampling distribution belongs to a maximum domain of attraction, these bounds are checked to be asymptotically tight. When the sampling distribution has a non-decreasing hazard rate, we derive an exponential Efron-Stein inequality for order statistics, that is an inequality connecting the logarithmic moment generating function of order statistics with exponential moments of Efron-Stein (jackknife) estimates of variance. This connection is used to derive variance and tail bounds for order statistics of Gaussian samples that are not within the scope of the Gaussian concentration inequality. Proofs are elementary and combine Rényi's representation of order statistics with the entropy approach to concentration of measure popularized by M. Ledoux.

1 Introduction

The purpose of this chapter is to develop non-asymptotic variance and tail bounds for order statistics. In the sequel, X_1, \dots, X_n are independent random variables, distributed

according to a certain probability distribution F , and $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ denote the corresponding order statistics (the non-increasing rearrangement of X_1, \dots, X_n). In Extreme Value Theory (EVT), the Fisher-Tippett-Gnedenko Theorem characterizes the asymptotic behavior of the maximum $X_{(1)}$ (de Haan & Ferreira, 2006) while asymptotics of the median $X_{(\lfloor n/2 \rfloor)}$ and other central order statistics are well documented (van der Vaart, 1998). Although the distribution function of order statistics is explicitly known, simple variance or tail bounds for order statistics do not seem to be well documented when sample size is kept fixed.

The search for variance and tail bounds for order statistics is driven by the desire to understand some aspects of the concentration of measure phenomenon (Ledoux, 2001; Massart, 2006). Concentration of measure theory tells us that a function of many independent random variables that does not depend too much on any of them is almost constant. The best known results in that field are the Poincaré and Gross logarithmic Sobolev inequalities and the Tsirelson-Ibragimov-Sudakov tail bounds for functions of Gaussian vectors. If X_1, \dots, X_n are independent standard Gaussian random variables, and $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is L -Lipschitz, then $Z = g(X_1, \dots, X_n)$ satisfies $\text{Var}(Z) \leq L^2$, $\ln \mathbb{E}[\exp(\lambda(Z - \mathbb{E}Z))] \leq \lambda^2 L^2 / 2$ and $\mathbb{P}\{Z - \mathbb{E}Z \geq t\} \leq \exp(-t^2 / (2L^2))$. If we apply these bounds to $Z = X_{(k)}$, the Lipschitz constant is (almost surely) $L = 1$, so Poincaré inequality allows us to establish $\text{Var}(X_{(k)}) \leq 1$. This upper bound is far from being satisfactory : it is well-known that $\text{Var}(X_{(1)}) = O(1/\ln n)$ and $\text{Var}(X_{(\lfloor n/2 \rfloor)}) = O(1/n)$ (de Haan & Ferreira, 2006; van der Vaart, 1998). Naive use of off-the-shelf concentration bounds does not work when handling order statistics. This situation is not uncommon : the analysis of the largest eigenvalue of random matrices from the Gaussian Unitary Ensemble (GUE) (Ledoux, 2006) provides a setting where the derivation of sharp concentration inequalities requires an ingenious combination of concentration inequalities and special representations.

Even so, our purpose is to show that the tools and methods of concentration of measure theory are relevant to the analysis of order statistics. To address this, the Efron-Stein inequalities are our main tools. They assert that, on average, the jackknife estimate(s) of the variance of functions of independent random variables are upper bounds. Extensions allow us to derive exponential bounds (see Theorem 2.1). We refer to (Miller, 1974; Shao & Wu, 1989) and references therein for an account of the early interplay between jackknife estimates, order statistics, EVT and statistical inference. When properly combined with Rényi's representation for order statistics (see Theorem 2.5), the so-called entropy method (Ledoux, 2001) allows us to recover sharp variance and tail bounds. Proofs are elementary and parallel the approach followed by Ledoux (2006) in a much more sophisticated setting. Ledoux builds on the determinantal structure of the joint density of the eigenvalues of random matrices from the GUE to upper bound tail bounds by sums of Gaussian integrals that can be handled by concentration arguments. In the sequel, we build on Rényi's representation of order statistics : $X_{(1)}, \dots, X_{(n)}$ can be represented as the monotone image of the order statistics of a sample of the standard exponential distribution which turn out to be distributed as partial sums of independent random variables.

In Section 2, we derive simple relations between the variance or the entropy of or-

der statistics $X_{(k)}$ and moments of spacings $\Delta_k = X_{(k)} - X_{(k+1)}$. When the sampling distribution has a non-decreasing hazard rate (a condition that is satisfied by Gaussian, exponential, Gumbel, logistic distributions, . . .) we are able to build on these connections, obtaining Theorem 2.10 that may be considered as an exponential Efron-Stein inequality for order statistics. In Section 3, using the framework of EVT, these connections are checked to be asymptotically tight.

In Section 4, using explicit bounds on the Gaussian hazard rate, we derive Bernstein-like inequalities for the maximum and the median of a sample of independent standard Gaussian random variables with correct variance and scale factors (Proposition 2.19). We provide non-asymptotic variance bounds for order statistics of Gaussian samples with the right order of magnitude in Propositions 2.14, and 2.16.

2 Order statistics and spacings

Efron-Stein inequalities (Efron & Stein, 1981) allow us to derive upper bounds on the variance of functions of independent random variables. The next version can be found in (Boucheron et al., 2004, p. 221)

Theorem 2.1. (EFRON-STEIN INEQUALITIES.) *Let $g: \mathbb{R}^n \rightarrow \mathbb{R}$ be measurable, and let $Z = g(X_1, \dots, X_n)$. Let $Z_i = g_i(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ where $g_i: \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ is an arbitrary measurable function. Suppose Z is square-integrable. Then*

$$\text{Var}[Z] \leq \sum_{i=1}^n \mathbb{E} [(Z - Z_i)^2] .$$

The quantity $\sum_{i=1}^n (Z - Z_i)^2$ is called a jackknife estimate of variance. Efron-Stein inequalities form a special case of a more general collection of inequalities that encompasses the so-called modified logarithmic Sobolev inequalities (Ledoux, 2001; Massart, 2000) or (Massart, 2006, p. 157).

Henceforth, the entropy of a non-negative random variable X is defined by $\text{Ent}[X] = \mathbb{E}[X \ln X] - \mathbb{E}X \ln \mathbb{E}X$.

Theorem 2.2. (MODIFIED LOGARITHMIC SOBOLEV INEQUALITY.) *Let $\tau(x) = e^x - x - 1$. Then for any $\lambda \in \mathbb{R}$,*

$$\text{Ent} [e^{\lambda Z}] = \lambda \mathbb{E} [Z e^{\lambda Z}] - \mathbb{E} [e^{\lambda Z}] \ln \mathbb{E} [e^{\lambda Z}] \leq \sum_{i=1}^n \mathbb{E} [e^{\lambda Z} \tau(-\lambda(Z - Z_i))] .$$

Theorems 2.1 and 2.2 provide a transparent connection between moments of order statistics and moments of spacings.

Henceforth, let $\psi: \mathbb{R} \rightarrow \mathbb{R}_+$ be defined by $\psi(x) = e^x \tau(-x) = 1 + (x - 1)e^x$.

Proposition 2.3. (ORDER STATISTICS AND SPACINGS.) *For all $1 \leq k \leq n/2$,*

$$\text{Var}[X_{(k)}] \leq k \mathbb{E} [(X_{(k)} - X_{(k+1)})^2]$$

and for all $\lambda \in \mathbb{R}$,

$$\text{Ent} [e^{\lambda X_{(k)}}] \leq k \mathbb{E} [e^{\lambda X_{(k+1)}} \psi(\lambda(X_{(k)} - X_{(k+1)}))] .$$

For all $n/2 < k \leq n$,

$$\text{Var}[X_{(k)}] \leq (n - k + 1) \mathbb{E} [(X_{(k-1)} - X_{(k)})^2]$$

and for all $\lambda \in \mathbb{R}$,

$$\text{Ent} [e^{\lambda X_{(k)}}] \leq (n - k + 1) \mathbb{E} [e^{\lambda X_{(k)}} \tau(\lambda(X_{(k-1)} - X_{(k)}))] .$$

Proof of Proposition 2.3. Let $Z = X_{(k)}$ and for $k \leq n/2$ define Z_i as the rank k statistic from subsample $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, that is, $Z_i = X_{(k+1)}$ if $X_i \geq X_{(k)}$ and $Z_i = Z$ otherwise. Apply Theorem 2.1.

For $k > n/2$, define Z_i as the rank $k - 1$ statistic from $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, that is $Z_i = X_{(k-1)}$ if $X_i \leq X_{(k)}$ and $Z_i = Z$ otherwise. Apply Theorem 2.1 again.

For $k \leq n/2$, define Z and Z_i as before, apply Theorem 2.2 :

$$\begin{aligned} \text{Ent} [e^{\lambda X_{(k)}}] &\leq k \mathbb{E} [e^{\lambda X_{(k)}} \tau(-\lambda(X_{(k)} - X_{(k+1)}))] \\ &= k \mathbb{E} [e^{\lambda X_{(k+1)}} e^{\lambda(X_{(k)} - X_{(k+1)})} \tau(-\lambda(X_{(k)} - X_{(k+1)}))] \\ &= k \mathbb{E} [e^{\lambda X_{(k+1)}} \psi(\lambda(X_{(k)} - X_{(k+1)}))] . \end{aligned}$$

The proof of the last statement proceeds by the same argument. \square

In the sequel, we focus on the setting $1 \leq k \leq n/2$: the setting $k > n/2$ can be treated in a similar way.

Proposition 2.3 can be fruitfully complemented by Rényi's representation of order statistics (de Haan & Ferreira, 2006, and references therein).

In the sequel, if f is a monotone function from (a, b) (where a and b may be infinite) to (c, d) , its generalized inverse $g^\leftarrow : (c, d) \rightarrow (a, b)$ is defined by $g^\leftarrow(y) = \inf\{x : a < x < b, g(x) \geq y\}$ (de Haan & Ferreira, 2006, for properties of this transformation).

Definition 2.4. *The U - transform of a distribution function F is defined as a non-decreasing function on $(1, \infty)$ by $U = (1/(1 - F))^\leftarrow$, $U(t) = \inf\{x : F(x) \geq 1 - 1/t\} = F^\leftarrow(1 - 1/t)$.*

Rényi's representation asserts that the order statistics of a sample of independent standard exponentially distributed random variables are distributed as partials sums of independent rescaled exponentially distributed random variables.

Theorem 2.5. (RÉNYI'S REPRESENTATION) *Let $X_{(1)} \geq \dots \geq X_{(n)}$ be the order statistics of a sample from distribution F , let $U = (1/(1 - F))^\leftarrow$, let $Y_{(1)} \geq Y_{(2)} \geq \dots \geq Y_{(n)}$ be the order statistics of an independent sample of the standard exponential distribution, then*

$$(Y_{(n)}, \dots, Y_{(i)}, \dots, Y_{(1)}) \stackrel{d}{=} \left(\frac{E_n}{n}, \dots, \sum_{k=i}^n \frac{E_k}{k}, \dots, \sum_{k=1}^n \frac{E_k}{k} \right)$$

where E_1, \dots, E_n are independent and identically distributed standard exponential random variables, and $(X_{(n)}, \dots, X_{(1)}) \stackrel{d}{=} (U \circ \exp(Y_{(n)}), \dots, U \circ \exp(Y_{(1)}))$.

We may readily test the tightness of Proposition 2.3. By Theorem 2.5, $Y_{(k)} = \frac{E_n}{n} + \dots + \frac{E_k}{k}$ and $\text{Var}[Y_{(k)}] = \sum_{i=k}^n \frac{1}{i^2}$. Hence, for any sequence $(k_n)_n$ with $\lim_n k_n = \infty$, and $\limsup k_n/n < 1$, $\lim_{n \rightarrow \infty} k_n \text{Var}[Y_{(k_n)}] = 1$, while by Proposition 2.3, $\text{Var}[Y_{(k)}] \leq k \mathbb{E}[(E_k/k)^2] = \frac{2}{k}$.

The next condition makes combining Proposition 2.3 and Theorem 2.5 easy.

Definition 2.6. (HAZARD RATE.) *The hazard rate of an absolutely continuous probability distribution with distribution function F is : $h = f/\bar{F}$ where f and $\bar{F} = 1 - F$ are respectively the density and the survival function associated with F .*

From elementary calculations, letting $U = (1/(1 - F))^\leftarrow$, we get $(U \circ \exp)' = 1/h(U \circ \exp)$.

Proposition 2.7. *Let F be an absolutely continuous distribution function with hazard rate h , let $U = (1/(1 - F))^\leftarrow$. Then, h is non-decreasing if and only if $U \circ \exp$ is concave.*

Proposition 2.8. *If the hazard rate h is non-decreasing, then for all $t > 0$ and $x > 0$,*

$$U(\exp(t + x)) - U(\exp(t)) \leq \frac{x}{h(U(\exp(t)))}.$$

Proof. If h is non-decreasing, then $U \circ \exp$ is concave. Therefore, for all $t > 0$ and $x > 0$ then

$$\begin{aligned} U(\exp(t + x)) - U(\exp(t)) &\leq (U \circ \exp)'(t)x \\ &= \frac{x}{h(U(\exp(t)))}. \end{aligned}$$

□

Moreover, assuming that the hazard rate is non-decreasing warrants negative association between spacings and related order statistics.

Proposition 2.9 (Negative association). *If F has non-decreasing hazard rate, then the k th spacing $\Delta_k = X_{(k)} - X_{(k+1)}$ and $X_{(k+1)}$ are negatively associated : for any pair of non-decreasing functions g_1 and g_2 ,*

$$\mathbb{E}[g_1(X_{(k+1)})g_2(\Delta_k)] \leq \mathbb{E}[g_1(X_{(k+1)})]\mathbb{E}[g_2(\Delta_k)].$$

Proof of Proposition 2.9. Let $Y_{(n)} \leq \dots \leq Y_{(1)}$ be the order statistics of a standard exponential sample. Let $E_k = k(Y_{(k)} - Y_{(k+1)})$ be the rescaled k th spacing of the exponential sample. By Theorem 2.5, E_k is standard exponentially distributed and independent of $Y_{(k+1)}$. Let g_1 and g_2 be two non-decreasing functions.

$$\begin{aligned} \mathbb{E}[g_1(X_{(k+1)})g_2(\Delta_k)] &= \mathbb{E}[g_1(U(e^{Y_{(k+1)}}))g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}}))] \\ &= \mathbb{E}[\mathbb{E}[g_1(U(e^{Y_{(k+1)}}))g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}]] \\ &= \mathbb{E}[g_1(U(e^{Y_{(k+1)}}))\mathbb{E}[g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}]]. \end{aligned}$$

The function $g_1 \circ U \circ \exp$ is non-decreasing. Almost surely, as the conditional distribution of $Y_{(k)}$ with respect to $Y_{(k+1)}$ is the exponential distribution with scale parameter $\frac{1}{k}$,

$$\mathbb{E} [g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}] = \int_0^\infty e^{-x} g_2(U(e^{\frac{x}{k}+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) dx .$$

As F has a non-decreasing hazard rate, $U(\exp(x/k + y)) - U(\exp(y)) = \int_0^{x/k} (U \circ \exp)'(y + z) dz$ is non-increasing with respect to y .

Hence, $\mathbb{E} [g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}]$ is a non-increasing function of $Y_{(k+1)}$.

Hence, by Chebyshev's association inequality,

$$\begin{aligned} & \mathbb{E}[g_1(X_{(k+1)})g_2(\Delta_k)] \\ & \leq \mathbb{E} [g_1(U(e^{Y_{(k+1)}}))] \mathbb{E} [\mathbb{E} [g_2(U(e^{E_k/k+Y_{(k+1)}}) - U(e^{Y_{(k+1)}})) \mid Y_{(k+1)}]] \\ & = \mathbb{E} [g_1(X_{(k+1)})] \mathbb{E} [g_2(\Delta_k)] . \end{aligned}$$

□

Negative association between order statistics and spacings allows us to establish our main result.

Theorem 2.10 (Exponential Efron-Stein inequality). *Let X_1, \dots, X_n be independently distributed according to F , let $X_{(1)} \geq \dots \geq X_{(n)}$ be the order statistics and let $\Delta_k = X_{(k)} - X_{(k+1)}$ be the k th spacing.*

Let $V_k = k\Delta_k^2$ denote the Efron-Stein estimate of the variance of $X_{(k)}$ (for $k = 1, \dots, n/2$).

If F has a non-decreasing hazard rate h , then for $1 \leq k \leq n/2$,

$$\text{Var} [X_{(k)}] \leq \mathbb{E}V_k \leq \frac{2}{k} \mathbb{E} \left[\left(\frac{1}{h(X_{(k+1)})} \right)^2 \right] .$$

For $\lambda \geq 0$, and $1 \leq k \leq n/2$,

$$\ln \mathbb{E} e^{\lambda(X_{(k)} - \mathbb{E}X_{(k)})} \leq \lambda \frac{k}{2} \mathbb{E} [\Delta_k (e^{\lambda\Delta_k} - 1)] = \lambda \frac{k}{2} \mathbb{E} \left[\sqrt{\frac{V_k}{k}} \left(e^{\lambda\sqrt{V_k/k}} - 1 \right) \right] . \quad (2.1)$$

Inequality (2.1) may be considered as an exponential Efron-Stein inequality for order statistics : it connects the logarithmic moment generating function of the k th order statistic with the exponential moments of the square root of the Efron-Stein estimate of variance $k\Delta_k^2$. This connection provides correct bounds for the standard exponential distribution whereas the exponential Efron-Stein inequality described in (Boucheron et al., 2003) does not. This comes from the fact that negative association between spacing and order statistics dispenses us from the general decoupling argument used in (Boucheron et al., 2003). It is then possible to carry out the Herbst's argument in an effortless way.

Proof of Theorem 2.10. Let $Y_{(k)}, Y_{(k+1)}$ denote the k th and $(k+1)$ th order statistics of a standard exponential sample of size n . By Proposition 2.3, using Rényi's representation (Theorem 2.5), and Proposition 2.7, for $k \leq n/2$,

$$\begin{aligned} \text{Var}[X_{(k)}] &\leq k \left[(U(e^{Y_{(k)}}) - U(e^{Y_{(k+1)}}))^2 \right] \\ &\leq k \mathbb{E} \left[(U(e^{Y_{(k+1)}}) e^{Y_{(k)} - Y_{(k+1)}} - U(e^{Y_{(k+1)}}))^2 \right] \\ &\leq k \mathbb{E} \left[((U \circ \exp)')^2(Y_{(k+1)}) (Y_{(k)} - Y_{(k+1)})^2 \right] \\ &\leq \frac{2}{k} \mathbb{E} \left[\left(\frac{1}{h(X_{(k+1)})} \right)^2 \right], \end{aligned}$$

as by Theorem 2.5, $Y_{(k)} - Y_{(k+1)}$ is independent of $Y_{(k+1)}$ and exponentially distributed with scale parameter $1/k$.

By Propositions 2.3 and 2.7, as ψ , defined by $\psi(x) = 1 + (x-1)e^x$ is non-decreasing

$$\begin{aligned} \text{Ent} [e^{\lambda X_{(k)}}] &\leq k \mathbb{E} [e^{\lambda X_{(k+1)}} \psi(\lambda \Delta_k)] \\ &\leq k \mathbb{E} [e^{\lambda X_{(k+1)}}] \times \mathbb{E} [\psi(\lambda \Delta_k)] \\ &\leq k \mathbb{E} [e^{\lambda X_{(k)}}] \times \mathbb{E} [\psi(\lambda \Delta_k)]. \end{aligned}$$

Multiplying both sides by $\exp(-\lambda \mathbb{E} X_{(k)})$,

$$\text{Ent} [e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})}] \leq k \mathbb{E} [e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})}] \times \mathbb{E} [\psi(\lambda \Delta_k)].$$

Let $G(\lambda) = \mathbb{E} e^{\lambda \Delta_k}$. Obviously, $G(0) = 1$, and as $\Delta_k \geq 0$, G and its derivatives are increasing on $[0, \infty)$,

$$\mathbb{E} [\psi(\lambda \Delta_k)] = 1 - G(\lambda) + \lambda G'(\lambda) = \int_0^\lambda s G''(s) ds \leq G''(\lambda) \frac{\lambda^2}{2}.$$

Hence, for $\lambda \geq 0$,

$$\frac{\text{Ent} [e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})}]}{\lambda^2 \mathbb{E} [e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})}]} = \frac{d \frac{1}{\lambda} \ln \mathbb{E} e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})}}{d\lambda} \leq \frac{k}{2} \frac{dG'}{d\lambda}.$$

Integrating both sides, using the fact that $\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \ln \mathbb{E} e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})} = 0$,

$$\frac{1}{\lambda} \ln \mathbb{E} e^{\lambda(X_{(k)} - \mathbb{E} X_{(k)})} \leq \frac{k}{2} (G'(\lambda) - G'(0)) = \frac{k}{2} \mathbb{E} [\Delta_k (e^{\lambda \Delta_k} - 1)].$$

□

3 Asymptotic assessment

Assessing the quality of the variance bounds from Proposition 2.3 in full generality is not easy. However, Extreme Value Theory (EVT) provides us with a framework where the Efron-Stein estimates of variance for maxima are asymptotically of the right order of magnitude.

Definition 2.11 (Maximum domain of attraction). *The distribution function F belongs to a maximum domain of attraction with tail index $\gamma \in \mathbb{R}$ ($F \in \text{DA}(\gamma)$), if and only if there exists a non-negative auxiliary function a on $[1, \infty)$ such that for $x \in [0, \infty)$ (if $\gamma > 0$), $x \in [0, -1/\gamma)$ (if $\gamma < 0$), $x \in \mathbb{R}$ (if $\gamma = 0$)*

$$\lim_n \mathbb{P} \left\{ \frac{\max(X_1, \dots, X_n) - F^{\leftarrow}(1 - 1/n)}{a(n)} \leq x \right\} = \exp(-(1 + \gamma x)^{-1/\gamma}).$$

If $\gamma = 0$, $(1 + \gamma x)^{-1/\gamma}$ should read as $\exp(-x)$.

If $F \in \text{DA}(\gamma)$ and has a finite variance ($\gamma < 1/2$), the variance of $(\max(X_1, \dots, X_n) - F^{\leftarrow}(1 - 1/n))/a(n)$ converges to the variance of the limiting extreme value distribution (de Haan & Ferreira, 2006).

Membership in a maximum domain of attraction is characterized by the *extended regular variation* property of $U = (1/(1 - F))^{\leftarrow} : F \in \text{DA}(\gamma)$ with auxiliary function a if and only if for all $x > 0$

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},$$

where the right-hand-side should read as $\ln x$ when $\gamma = 0$ (de Haan & Ferreira, 2006).

Using Theorem 2.1.1 from (de Haan & Ferreira, 2006) and Theorem 1.33, and performing simple calculations, we readily obtain that the asymptotic ratio between the Efron-Stein upper bound and the variance of $X_{(1)}$ converges toward a limit that depends only on γ (for $\gamma = 0$ this limit is $12/\pi^2 \approx 1.21$).

Proposition 2.12. *Assume $X_{(1)} \geq \dots \geq X_{(n)}$ are the order statistics of an independent sample distributed according to F , where $F \in \text{DA}(\gamma)$, $\gamma < 1/2$, with auxiliary function a . Then*

$$\lim_n \frac{\mathbb{E}[(X_{(1)} - X_{(2)})^2]}{a(n)^2} = \frac{2\Gamma(1-2\gamma)}{(1-\gamma)(1-2\gamma)} \quad \text{while} \quad \lim_n \frac{\text{Var}(X_{(1)})}{a(n)^2} = \frac{1}{\gamma^2} (\Gamma(1-2\gamma) - \Gamma(1-\gamma)^2).$$

For $\gamma = 0$, the last expression should read as $\pi^2/6$.

When the tail index γ is negative, the asymptotic ratio degrades as $\gamma \rightarrow -\infty$, it scales as -4γ .

4 Order statistics of Gaussian samples

We now turn to the Gaussian setting and establish Bernstein inequalities for order statistics of absolute values of independent standard Gaussian random variables.

A real-valued random variable X is said to be *sub-gamma on the right tail with variance factor v and scale parameter c* if

$$\ln \mathbb{E} e^{\lambda(X - \mathbb{E}X)} \leq \frac{\lambda^2 v}{2(1 - c\lambda)} \quad \text{for every } \lambda \quad \text{such that} \quad 0 < \lambda < 1/c.$$

Such a random variable satisfies a so-called Bernstein inequality : for $t > 0$, $\mathbb{P} \{X \geq \mathbb{E}X + \sqrt{2vt} + ct\} \leq \exp(-t)$. A real-valued random variable X is said to be *sub-gamma on the left tail with variance factor v and scale parameter c* , if $-X$ is sub-gamma

on the right tail with variance factor v and scale parameter c . A Gamma random variable with shape parameter p and scale parameter c (expectation pc and variance pc^2) is sub-gamma on the right tail with variance factor pc^2 and scale factor c while it is sub-gamma on the left-tail with variance factor pc^2 and scale factor 0. The Gumbel distribution (with distribution function $\exp(-\exp(-x))$) is sub-gamma on the right-tail with variance factor $\pi^2/6$ and scale factor 1, and sub-gamma on the left-tail with scale factor 0 (note that this statement is not sharp, see Lemma 2.15 below).

Order statistics of Gaussian samples provide an interesting playground for assessing Theorem 2.10. Let Φ and ϕ denote respectively the standard Gaussian distribution function and density. Throughout this section, let $\tilde{U}: (1, \infty) \rightarrow [0, \infty)$ be defined by $\tilde{U}(t) = \Phi^\leftarrow(1 - 1/(2t))$, $\tilde{U}(t)$ is the $1 - 1/t$ quantile of the distribution of the absolute value of a standard Gaussian random variable, or the $1 - 1/(2t)$ quantile of the standard Gaussian distribution.

Proposition 2.13. *Absolute values of Gaussian random variables have a non-decreasing hazard rate :*

i) $\tilde{U} \circ \exp$ is concave ;

ii) For $y > 0$, $\phi(\tilde{U}(\exp(y)))/\bar{\Phi}(\tilde{U}(\exp(y))) \geq \sqrt{\kappa_1(y + \ln 2)}$ where $\kappa_1 \geq 1/2$.

iii) For $t \geq 3$,

$$\sqrt{2 \ln(2t) - \ln \ln(2t) - \ln(4\pi)} \leq \tilde{U}(t) \leq \sqrt{2 \ln(2t) - \ln \ln(2t) - \ln \pi}.$$

Proof of Proposition 2.13. i) As $(\tilde{U} \circ \exp)'(t) = \bar{\Phi}(\tilde{U}(e^t))/\phi(\tilde{U}(e^t))$, it suffices to check that the standard Gaussian distribution has a non-decreasing hazard rate on $[0, \infty)$. Let $h = \phi/\bar{\Phi}$, by elementary calculations, for $x > 0$, $h'(x) = (\phi(x) - x\bar{\Phi}(x))\phi(x)/\bar{\Phi}^2(x) \geq 0$ where the last inequality is a well known fact.

ii) For $\kappa_1 = 1/2$, for $p \in (0, 1/2]$, the fact that $p\sqrt{\kappa_1 \ln 1/p} \leq \phi \circ \Phi^\leftarrow(p)$ follows from $\phi(x) - x\bar{\Phi}(x) \geq 0$ for $x > 0$. Hence,

$$\frac{\bar{\Phi}(\Phi^\leftarrow(1 - e^{-y}/2))}{\phi(\Phi^\leftarrow(1 - e^{-y}/2))} = \frac{e^{-y}/2}{\phi(\Phi^\leftarrow(e^{-y}/2))} \leq \frac{1}{\sqrt{\kappa_1(\ln 2 + y)}}.$$

iii) The first inequality can be deduced from $\phi \circ \Phi^\leftarrow(p) \leq p\sqrt{2 \ln 1/p}$, for $p \in (0, 1/2)$ (Tillich & Zémor, 2001), the second from $p\sqrt{\kappa_1 \ln 1/p} \leq \phi \circ \Phi^\leftarrow(p)$. □

The next proposition shows that when used in a proper way, Efron-Stein inequalities may provide seamless bounds on extreme, intermediate and central order statistics of Gaussian samples.

Proposition 2.14. *Let $n \geq 3$, let $X_{(1)} \geq \dots \geq X_{(n)}$ be the order statistics of absolute values of a standard Gaussian sample,*

$$\text{For } 1 \leq k \leq n/2, \quad \text{Var}[X_{(k)}] \leq \frac{1}{k \ln 2 \ln \frac{2n}{k} - \ln(1 + \frac{4}{k} \ln \ln \frac{2n}{k})} \cdot 8.$$

By Theorem 1.33, $\lim_n 2 \ln n \operatorname{Var}[X_{(1)}] = \pi^2/6$, while the above described upper bound on $\operatorname{Var}[X_{(1)}]$ is equivalent to $(8/\ln 2)/\ln n$. If $\lim_n k_n = \infty$ and $\lim_n k_n/n = 0$, by Theorem 1.18, $\lim_n k(\tilde{U}(n/k))^2 \operatorname{Var}[X_{(k)}] = 1$. For the asymptotically normal median of absolute values, $\lim_n (4\phi(\tilde{U}(2))^2 n) \operatorname{Var}[X_{(n/2)}] = 1$ (van der Vaart, 1998). Again, the bound in Proposition 2.14 has the correct order of magnitude.

Lemma 2.15. *Let $Y_{(k)}$ be the k th order statistic of a sample of n independent standard exponential random variables, let $\ln 2 < z < \ln(n/k)$, then*

$$\mathbb{P} \{Y_{(k+1)} \leq \ln(n/k) - z\} \leq \exp\left(-\frac{k(e^z-1)}{4}\right).$$

Proof of Lemma 2.15.

$$\mathbb{P} \{Y_{(k+1)} \leq \ln(n/k) - z\} = \sum_{j=0}^k \binom{n}{j} \left(1 - \frac{ke^z}{n}\right)^{n-j} \left(\frac{ke^z}{n}\right)^j \leq \exp\left(-\frac{k(e^z-1)^2}{2e^z}\right)$$

since the right-hand-side of the first line is the probability that a binomial random variable with parameters n and $\frac{ke^z}{n}$ is less than k , which is sub-gamma on the left-tail with variance factor less than ke^z and scale factor 0. \square

Proof of Proposition 2.14. By Propositions 2.10 and 2.13, letting $\kappa_1 = 1/2$

$$\begin{aligned} \operatorname{Var}(X_{(k)}) &\leq \frac{2}{k} \mathbb{E} \left[\frac{2}{\ln 2 + Y_{(k+1)}} \right] \\ &\leq \frac{1}{\ln 2} \frac{4}{k} \mathbb{P} \{Y_{(k+1)} \leq \ln(n/k) - z\} + \frac{4}{k} \frac{1}{\ln \frac{n}{k} - z + \ln 2} \\ &\leq \frac{4}{k \ln 2} \frac{1}{\ln \frac{2n}{k}} + \frac{4}{k} \frac{1}{\ln \frac{2n}{k} - \ln(1 + \frac{4}{k} \ln \ln \frac{2n}{k})}, \end{aligned}$$

where we used Lemma 2.15 with $z = \ln(1 + \frac{4}{k} \ln \ln \frac{2n}{k})$. \square

Our next goal is to establish that the order statistics of absolute values of independent standard Gaussian random variables are sub-gamma on the right-tail with variance factor close to the Efron-Stein estimates of variance derived in Proposition 2.13 and scale factor not larger than the square root of the Efron-Stein estimate of variance.

Before describing the consequences of Theorem 2.10, it is interesting to look at what can be obtained from Rényi's representation and exponential inequalities for sums of gamma distributed random variables.

Proposition 2.16. *Let $X_{(1)}$ be the maximum of the absolute values of n independent standard Gaussian random variables, and let $\tilde{U}(s) = \Phi^{\leftarrow}(1 - 1/(2s))$ for $s \geq 1$. For $t > 0$,*

$$\mathbb{P} \left\{ X_{(1)} - \mathbb{E}X_{(1)} \geq t/(3\tilde{U}(n)) + \sqrt{t/\tilde{U}(n)} + \delta_n \right\} \leq \exp(-t),$$

where $\delta_n > 0$ and $\lim_n (\tilde{U}(n))^3 \delta_n = \frac{\pi^2}{12}$.

This inequality looks like what we are looking for : $\tilde{U}(n)(X_{(1)} - \mathbb{E}X_{(1)})$ converges in distribution, but also in quadratic mean, or even according to the Orlicz norm defined by $x \mapsto \exp(|x|) - 1$, toward a centered Gumbel distribution. As the Gumbel distribution is sub-gamma on the right tail with variance factor $\pi^2/6$ and scale factor 1, we expect $X_{(1)}$ to satisfy a Bernstein inequality with variance factor of order $1/\tilde{U}(n)^2$ and scale factor $1/\tilde{U}(n)$. Up to the shift δ_n , this is the content of the proposition.

Note that the shift is asymptotically negligible with respect to typical fluctuations. The next proposition shows that Theorem 2.10 captures the correct order of growth for the right-tail of Gaussian maxima even though the constants are not sharp enough to make it competitive with Proposition 2.16

Proposition 2.17. *For n such that the solution v_n of the equation $16/x + \ln(1 + 2/x + 4 \ln(4/x)) = \ln(2n)$ is smaller than 1, for all $0 \leq \lambda < \frac{1}{\sqrt{v_n}}$,*

$$\ln \mathbb{E} e^{\lambda(X_{(1)} - \mathbb{E}X_{(1)})} \leq \frac{v_n \lambda^2}{2(1 - \sqrt{v_n} \lambda)}.$$

For all $t > 0$,

$$\mathbb{P} \left\{ X_{(1)} - \mathbb{E}X_{(1)} > \sqrt{v_n}(t + \sqrt{2t}) \right\} \leq e^{-t}.$$

Proof of Proposition 2.17. By Proposition 2.10,

$$\ln \mathbb{E} e^{\lambda(X_{(1)} - \mathbb{E}X_{(1)})} \leq \frac{\lambda}{2} \mathbb{E} [\Delta (e^{\lambda \Delta} - 1)]$$

where $\Delta = X_{(1)} - X_{(2)} \stackrel{d}{=} U(2e^{Y_{(2)} + E_1}) - U(2e^{Y_{(2)}})$, with E_1 a standard exponentially distributed random variable and independent of $Y_{(2)}$ which is distributed like the 2nd largest order statistics of a standard exponential sample.

On the one hand, the conditional expectation

$$\mathbb{E} \left[(U(2e^{E_1 + Y_{(2)}}) - U(2e^{Y_{(2)}})) \left(e^{\lambda(U(2e^{E_1 + Y_{(2)}}) - U(2e^{Y_{(2)}}))} - 1 \right) \mid Y_{(2)} \right]$$

is a non-increasing function of $Y_{(2)}$. The maximum is achieved for $Y_{(2)} = 0$, and is equal to :

$$2 \int_0^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} x (e^{\lambda x} - 1) dx \leq 2\lambda e^{\frac{\lambda^2}{2}}.$$

On the other hand, by Proposition 2.13,

$$U(2e^{E_1 + Y_{(2)}}) - U(2e^{Y_{(2)}}) \leq \frac{\sqrt{2}E_1}{\sqrt{(\ln 2 + Y_{(2)})}}.$$

For $0 \leq \mu < 1/2$,

$$\int_0^\infty \mu x (e^{\mu x} - 1) e^{-x} dx = \frac{\mu^2(2 - \mu)}{(1 - \mu)^2} \leq \frac{2\mu^2}{1 - 2\mu}.$$

Hence,

$$\lambda \mathbb{E} \left[\left(U(2e^{E_1+Y_{(2)}}) - U(2e^{Y_{(2)}}) \right) \left(e^{\lambda(U(2e^{E_1+Y_{(2)}}) - U(2e^{Y_{(2)}}))} - 1 \right) \mid Y_{(2)} \right] \leq \frac{4\lambda^2}{\ln 2 + Y_{(2)}} \frac{1}{1 - \frac{2\sqrt{2}\lambda}{\sqrt{\ln 2 + Y_{(2)}}}}.$$

Letting $\tau = \ln n - \ln(1 + 2\lambda^2 + 4\ln(4/v_n))$,

$$\ln \mathbb{E} \left[e^{\lambda(X_{(1)} - \mathbb{E}X_{(1)})} \right] \leq \underbrace{\lambda^2 e^{\lambda^2/2} \mathbb{P}\{Y_{(2)} \leq \tau\}}_{:=I} + \underbrace{\frac{4\lambda^2}{\ln 2 + \tau} \frac{1}{1 - \frac{2\sqrt{2}\lambda}{\sqrt{\ln 2 + \tau}}}}_{:=II}.$$

By Lemma 2.15, (I) $\leq \frac{v_n \lambda^2}{4}$.

As $\lambda \leq 1/\sqrt{v_n}$ and by assumption on v_n , $\ln 2 + \tau \geq 16/v_n$ and (II) $\leq \frac{v_n \lambda^2}{4(1 - \sqrt{v_n \lambda})}$. \square

Remark 2.18. We may notice that the v_n is the order of $1/\ln(2n)$, and that is the order of the Efron-Stein estimate of the variance of $X_{(1)}$.

We may also use Theorem 2.10 to provide a Bernstein inequality for the median of absolute values of a standard Gaussian sample. We assume $n/2$ is an integer.

Proposition 2.19. *Let $v_n = 8/(n \ln 2)$.*

For all $0 \leq \lambda < n/(2\sqrt{v_n})$,

$$\ln \mathbb{E} e^{\lambda(X_{(n/2)} - \mathbb{E}X_{(n/2)})} \leq \frac{v_n \lambda^2}{2(1 - 2\lambda\sqrt{v_n/n})}.$$

For all $t > 0$,

$$\mathbb{P} \left\{ X_{(n/2)} - \mathbb{E}X_{(n/2)} > \sqrt{2v_n t} + 2t\sqrt{v_n/n} \right\} \leq e^{-t}.$$

Proof of Proposition 2.19. By Proposition 2.10,

$$\ln \mathbb{E} e^{\lambda(X_{(n/2)} - \mathbb{E}X_{(n/2)})} \leq \frac{n}{4} \lambda \mathbb{E} \left[\Delta_{n/2} (e^{\lambda \Delta_{n/2}} - 1) \right]$$

where $\Delta_{n/2} = X_{(n/2)} - X_{(n/2+1)} \stackrel{d}{=} U(2e^{E_{n/2}/(n/2) + Y_{(n/2+1)}}) - U(e^{Y_{(n/2+1)}})$ where $E_{n/2}$ is standard exponentially distributed and independent of $Y_{(n/2+1)}$.

By Proposition 2.13,

$$\Delta_{n/2} \leq \frac{\sqrt{2}\lambda E_{n/2}}{(n/2)\sqrt{\ln 2 + Y_{(n/2+1)}}} \leq \frac{\sqrt{2}E_{n/2}}{(n/2)\sqrt{\ln 2}} = \sqrt{\frac{v_n}{n}} E_{n/2}.$$

Reasoning as in the proof of Proposition 2.17,

$$\ln \mathbb{E} e^{\lambda(X_{(n/2)} - \mathbb{E}X_{(n/2)})} \leq \frac{v_n \lambda^2}{2(1 - 2\lambda\sqrt{v_n/n})}.$$

\square

Remark 2.20. Once again, v_n is the order of $1/n$, and that is the order of the Efron-Stein estimate of the variance of $X_{(n/2)}$.

As the hazard rate $\phi(x)/\bar{\Phi}(x)$ of the Gaussian distribution tends to 0 as x tends to $-\infty$, the preceding approach does not work when dealing with order statistics of Gaussian samples.

Nevertheless, Proposition 2.14 paves the way to simple bounds on the variance of maxima of Gaussian samples.

Proposition 2.21. *Let X_1, \dots, X_n be n independent and identically distributed standard Gaussian random variables, let $X_{(1)} \geq \dots \geq X_{(n)}$ be the order statistics.*

For all $n \geq 11$,

$$\text{Var}[X_{(1)}] \leq \frac{8/\ln 2}{\ln(n/2) - \ln(1 + 4 \ln \ln(n/2))} + 2^{-n} + \exp\left(-\frac{n}{8}\right) + \frac{4\pi}{n}.$$

Proof of Proposition 2.21. We may generate n independent standard Gaussian random variables in two steps : first generate n independent random signs $(\varepsilon_1, \dots, \varepsilon_n : \mathbb{P}\{\varepsilon_i = 1\} = 1 - \mathbb{P}\{\varepsilon_i = -1\} = 1/2)$, then generate absolute values of independent standard Gaussian random variables (V_1, \dots, V_n) , the resulting sample (X_1, \dots, X_n) is obtained as $X_i = \varepsilon_i V_i$. Let N be the number of positive random signs.

$$\text{Var}(X_{(1)}) = \underbrace{\mathbb{E} [\text{Var} (X_{(1)} | \sigma(N))]}_{\text{I}} + \underbrace{\text{Var} (\mathbb{E} [X_{(1)} | \sigma(N)])}_{\text{II}}.$$

Conditionally on $N = m$, if $m \geq 1$, $X_{(1)}$ is distributed as the maximum of a sample of m independent absolute values of standard Gaussian random variables. If $m = 0$, $X_{(1)}$ is negative, its conditional variance is equal to the variance of the minimum of a sample of size n .

Hence, letting $V_{(k)}^m$ denote the k th order statistic of a sample of m independent absolute values of standard Gaussian random variables. Let w_m denote the upper bound on the variance of $V_{(1)}^m$ from Proposition 2.14. Note that w_m is non-increasing, and that, by Poincaré's inequality, $\text{Var}(V_{(k)}^n) \leq 1$.

(II) \ll (I), as

$$\begin{aligned} \text{(I)} &= \sum_{m=1}^n \binom{n}{m} 2^{-n} \text{Var} (V_{(1)}^m) + \binom{n}{0} 2^{-n} \text{Var} (V_{(n)}^n) \\ &= \sum_{m=1}^n \binom{n}{m} 2^{-n} w_m + 2^{-n} \\ &\leq \sum_{m=1}^{n/4} \binom{n}{m} 2^{-n} + w_{n/4} + 2^{-n} \\ &\leq \exp(-n/8) + w_{n/4} + 2^{-n}. \end{aligned}$$

Let H_N be the random harmonic number $H_N = \sum_{i=1}^N 1/i$,

$$\begin{aligned}
(\text{II}) &= \mathbb{E}_{N,N'} \left[\left(\mathbb{E}[X_{(1)} | N] - \mathbb{E}[X_{(1)} | N'] \right)_+^2 \right] \\
&= \mathbb{E}_{N,N'} \left[\left(\mathbb{E} \left[\tilde{U} \left(\exp \left(\sum_{i=1}^N \frac{E_i}{i} \right) \right) - \tilde{U} \left(\exp \left(\sum_{i=1}^{N'} \frac{E_i}{i} \right) \right) \right] \right)_+^2 \right] \\
&\leq \mathbb{E}_{N,N'} \left[\left(1/h(\tilde{U}(\exp(\sum_{i=1}^{N'} \frac{E_i}{i}))) (H_N - H_{N'}) \right)_+^2 \right] \\
&\leq \frac{\pi}{2} \mathbb{E}_{N,N'} \left[(H_N - H_{N'})_+^2 \right] \\
&= \frac{\pi}{2} \text{Var}(H_N).
\end{aligned}$$

Now, as $N = \sum_{i=1}^n (1 + \varepsilon_i)/2$, letting $Z = H_N$ and $Z_i = \sum_{j=1}^{N-(1+\varepsilon_i)/2} 1/j$, by Efron-Stein inequality, $\text{Var} Z \leq \mathbb{E}[0 \wedge 1/N]$.

Finally, using Hoeffding's inequality (Boucheron et al., 2004) in a crude way leads to $\mathbb{E}[0 \wedge 1/N] \leq \exp(-n/8) + 4/n \leq 8/n$. We may conclude by (II) $\leq (4\pi)/n$. \square

Remark 2.22. Trading simplicity for tightness, sharper bounds on II could be derived and show that $\text{II} = O(1/(n \ln n))$.

5 Appendix

5.1 Proof of Proposition 2.12

Lemma 2.23. For $\gamma < 1/2$, $\mathbb{E} [e^{2\gamma(Y_{(2)} - \ln n)}] = \Gamma(2(1 - \gamma))$.

Lemma 2.24 (POTTER TYPE BOUNDS). For all $\varepsilon, \delta > 0$, there exists $t_0(\varepsilon, \delta) > 0$ such that for all $t, tx \geq t_0$,

$$\begin{aligned}
\left| \left(\frac{U(tx) - U(t)}{a(t)} \right)^2 - \left(\frac{x^\gamma - 1}{\gamma} \right)^2 \right| &\leq 3\varepsilon^2 x^{2\gamma} (\max(x^{-\delta}, x^\delta))^2 + 2\varepsilon x^\gamma \left(\frac{x^\gamma - 1}{\gamma} \right) \max(x^{-\delta}, x^\delta) \\
\left| \left(\frac{a(tx)}{a(t)} \right)^2 - x^{2\gamma} \right| &\leq \varepsilon x^{2\gamma} \max(x^{-\delta}, x^\delta).
\end{aligned}$$

Proof of Proposition 2.12. By Rényi's representation (Theorem 2.5),

$$X_{(1)} - X_{(2)} \stackrel{d}{=} U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})$$

where E is an exponentially distributed random variable independent of $Y_{(2)}$, distributed as the second order statistic of a standard exponential sample.

We prove that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\frac{U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})}{a(n)} \right)^2 - \left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 e^{2\gamma(Y_{(2)} - \ln n)} \right] = 0.$$

Indeed, this implies that

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\frac{X_{(1)} - X_{(2)}}{a(n)} \right)^2 \right] &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\frac{U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})}{a(n)} \right)^2 \right] \\
&= \mathbb{E} \left[\left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 \right] \lim_{n \rightarrow \infty} \mathbb{E} [e^{2\gamma(Y_{(2)} - \ln n)}] \\
&= \frac{2\Gamma(2(1 - \gamma))}{(1 - \gamma)(1 - 2\gamma)}.
\end{aligned}$$

Write

$$\begin{aligned}
&\left(\frac{U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})}{a(n)} \right)^2 - \left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 e^{2\gamma(Y_{(2)} - \ln n)} \\
&= \left(\frac{U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})}{a(e^{Y_{(2)}})} \right)^2 \left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 - \left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 e^{2\gamma(Y_{(2)} - \ln n)} \\
&= \underbrace{\left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 \left(\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 - e^{2\gamma(Y_{(2)} - \ln n)} \right)}_{:=\text{(I)}} \\
&\quad + \underbrace{\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 \left(\left(\frac{U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})}{a(e^{Y_{(2)}})} \right)^2 - \left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 \right)}_{:=\text{(II)}}.
\end{aligned}$$

Let $\varepsilon, \delta > 0$ and t_0 as in Lemma 2.24, and let $n \geq t_0$, then, on the one hand,

$$\begin{aligned}
\mathbb{E}[\text{(I)}] &= \mathbb{E} \left[\left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 \right] \\
&\quad \times \left(\mathbb{E} \left[\left(\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 - e^{2\gamma(Y_{(2)} - \ln n)} \right) \mathbb{1}_{\{Y_{(2)} > t_0\}} \right] + \mathbb{E} \left[\left(\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 - e^{2\gamma(Y_{(2)} - \ln n)} \right) \mathbb{1}_{\{Y_{(2)} \leq t_0\}} \right] \right)
\end{aligned}$$

The first term in the sum tends to 0 using Lemma 2.24 and the second tends to 0 thanks to the uniform convergence of $a(tx)/a(t)$ to x^γ on the compact $[0, t_0]$. Thus, $\mathbb{E}[\text{(I)}]$ tends to 0, as n tends to ∞ .

On the other hand,

$$\begin{aligned}
\mathbb{E}[\text{(II)}] &= \mathbb{E} \left[\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 \left(\left(\frac{U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})}{a(e^{Y_{(2)}})} \right)^2 - \left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 \right) \mathbb{1}_{\{Y_{(2)} > t_0\}} \right] \\
&\quad + \mathbb{E} \left[\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 \left(\left(\frac{U(e^{Y_{(2)}} e^E) - U(e^{Y_{(2)}})}{a(e^{Y_{(2)}})} \right)^2 - \left(\frac{e^{\gamma E} - 1}{\gamma} \right)^2 \right) \mathbb{1}_{\{Y_{(2)} \leq t_0\}} \right].
\end{aligned}$$

Using Lemma 2.24, the first term is bounded by $\mathbb{E}[A(E, \varepsilon, \delta)] \mathbb{E} \left[\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 \right]$ where $A(E, \varepsilon, \delta)$ is a factor depending only on ε, δ and E such that $\mathbb{E}[A(E, \varepsilon, \delta)] = o(\varepsilon)$; then $\mathbb{E} \left[\left(\frac{a(e^{Y_{(2)}})}{a(n)} \right)^2 \right]$ is a convergent, thus bounded sequence. The first term tends to 0. For the

second term, use the uniform convergence of $(U(tx) - U(t))/a(t)$ on the compact $[0, t_0]$ and again the fact that $\mathbb{E} \left[\left(\frac{a(e^{Y_{(1)}})}{a(n)} \right)^2 \right]$ is a convergent, thus bounded sequence. The second term tends also to 0. Thus, $\mathbb{E}[(\text{II})]$ tends to 0, as n tends to ∞ . \square

5.2 Proof of Proposition 2.16

Let $Y_{(1)}$ denote the maximum of a sample of independent standard exponentially distributed random variables, so that $X_{(1)} \sim \tilde{U}(e^{Y_{(1)}})$.

Thanks to the concavity of $\tilde{U} \circ \exp$, and the fact that $\bar{\Phi}(x)/\phi(x) \leq 1/x$ for $x > 0$, $\tilde{U}(\exp(Y_{(1)})) - \tilde{U}(\exp(H_n)) \leq (Y_{(1)} - H_n)/(\tilde{U}(\exp(H_n)))$. Now, $Y_{(1)}$ satisfies a Bernstein inequality with variance factor not larger than $\text{Var}(Y_{(1)}) \leq \pi^2/6 \leq 2$ and scale factor not larger than 1, so

$$\mathbb{P}\{Y_{(1)} - H_n \geq t\tilde{U}(\exp(H_n))\} \leq \exp\left(-\frac{t^2\tilde{U}(\exp(H_n))^2}{2(2 + t\tilde{U}(\exp(H_n))/3)}\right).$$

Agreeing on $\delta_n = \tilde{U}(e^{H_n}) - \mathbb{E}\tilde{U}(e^{Y_{(1)}})$, using the monotonicity of $x^2/(2 + x/3)$ and $\ln n \leq H_n$, we obtain the first part of the proposition.

The second-order extended regular variation condition satisfied by the standard Gaussian distribution (de Haan & Ferreira, 2006, Exercise 2.9, p. 61) allows us to assert

$$\tilde{U}(n)^3(\tilde{U}(e^x n) - \tilde{U}(n) - \frac{x}{\tilde{U}(n)}) \rightarrow -\frac{x^2}{2} - x. \quad (2.2)$$

This suggests that

$$\tilde{U}(\exp(H_n))^3 \left(\mathbb{E}\tilde{U}(e^{Y_{(1)}-H_n}e^{H_n}) - \tilde{U}(e^{H_n}) - \frac{\mathbb{E}Y_{(1)} - H_n}{\tilde{U}(\exp(H_n))} \right) \rightarrow -\frac{\pi^2}{12},$$

or that the order of magnitude of $\tilde{U}(e^{H_n}) - \mathbb{E}X_{(1)}$ is $O(1/\tilde{U}(n)^3)$ which is small with respect to $1/\tilde{U}(n)$.

The random variable $Z_n = Y_{(1)} - H_n$ converges in distribution towards a recentred Gumbel distribution, denoted G_0 , and as $Y_{(1)}$ is positive, $Z_n \geq -H_n$.

In (2.2), the convergence is locally uniform and $\tilde{U}(n)$ can be replaced by $\tilde{U}(e^{H_n})$, as e^{H_n} is an increasing function of n , and tends to ∞ , as n tends to ∞ .

We prove that

$$\mathbb{E} \left[\tilde{U}(e^{H_n})^3 \left(\tilde{U}(e^{H_n}e^{X_n}) - \tilde{U}(e^{H_n}) - \frac{X_n}{\tilde{U}(e^{H_n})} \right) + \frac{X_n^2}{2} + X_n \right] \xrightarrow{n \rightarrow \infty} 0.$$

Indeed, this implies that

$$\lim_{n \rightarrow \infty} \tilde{U}(e^{H_n})^3 \left(\mathbb{E} \left[\tilde{U}(e^{H_n}e^{X_n}) \right] - \tilde{U}(e^{H_n}) \right) = \lim_{n \rightarrow \infty} -\frac{1}{2} \mathbb{E} [X_n^2].$$

The right-hand side member is equal to $-\frac{1}{2} \mathbb{E} [G_0^2] = -\frac{\pi^2}{12}$ thanks to Theorem 1.33.

Let $A_n = \left| \tilde{U}(e^{H_n})^3 \left(\tilde{U}(e^{H_n}e^{X_n}) - \tilde{U}(e^{H_n}) - \frac{X_n}{\tilde{U}(e^{H_n})} \right) + \frac{X_n^2}{2} + X_n \right|$, and

$$\mathbb{E} [A_n] = \underbrace{\mathbb{E} [A_n \mathbb{1}_{\{X_n \geq -H_n/2\}}]}_{:=\text{(I)}} + \underbrace{\mathbb{E} [A_n \mathbb{1}_{\{X_n \leq -H_n/2\}}]}_{:=\text{(II)}}.$$

On the one hand, Theorem B.3.10 from (de Haan & Ferreira, 2006) asserts that for all $\delta, \varepsilon > 0$, for n, x such that $\min(e^{H_n}, e^{H_n} e^x) \geq t_0(\varepsilon, \delta)$,

$$\left| \tilde{U}(e^{H_n})^3 \left(\tilde{U}(e^{H_n} e^x) - \tilde{U}(e^{H_n}) - \frac{x}{\tilde{U}(e^{H_n})} \right) + \frac{x^2}{2} + x \right| \leq \varepsilon e^{\delta|x|}.$$

Then, for n such that $e^{H_n/2} \geq t_0$, (I) $\leq \varepsilon \mathbb{E} [e^{\delta|X_n|}]$. It can be showed by straightforward calculations that $\mathbb{E} [e^{\delta|X_n|}]$ is convergent, thus upper-bounded by C , a constant independent of n . Thus, (I) $\leq \varepsilon C$ which tends to 0 by letting ε tends to 0.

On the other hand, by straightforward arguments A_n is upper-bounded by a term equivalent to $\kappa(\ln n)^2$ on the event $\{X_n \leq -H_n/2\}$, so that (II) is upper-bounded by a term equivalent to

$$\begin{aligned} \kappa(\ln n)^2 \mathbb{P} \{X_n \leq -H_n/2\} &= \kappa(\ln n)^2 \mathbb{P} \{Y_{(1)} \leq H_n/2\} \\ &\leq \kappa(\ln n)^2 e^{-(\ln n)^2/16} \longrightarrow 0. \end{aligned}$$

Therefore, we have proved that $\mathbb{E} [A_n] \longrightarrow 0$.

Chapitre 3

Tail index estimation, concentration and adaptivity

Ce chapitre a fait l'objet d'un article soumis dans Electronic Journal of Statistics (Boucheron & Thomas, 2015).

Contents

1	Introduction	76
2	Background, notations and tools	78
2.1	The Hill estimator as a smooth tail statistics	78
2.2	Frameworks	80
2.3	Lepski's method and adaptive tail index estimation	81
2.4	Lower bound	84
2.5	Talagrand's concentration phenomenon for products of exponential distributions	85
3	Main results	87
3.1	Bounding the variance of the Hill estimator	87
3.2	Concentration inequalities for the Hill estimators	88
3.3	Adaptive Hill estimation	90
4	Proofs	93
4.1	Proof of Proposition 3.2	93
4.2	Proof of Proposition 3.11	93
4.3	Proof of Theorem 3.13	94
4.4	Proof of Theorem 3.17	98
4.5	Proof of Corollary 3.19	100
4.6	Proof of Theorem 3.20	101
4.7	Proof of Corollary 3.23	103
5	Simulations	104
6	Appendix	107
6.1	Calibration of the preliminary selection rule	107
6.2	Proof of Corollary 3.10	108

6.3	Proof of Abelian Proposition 3.12	109
6.4	Proof of Inequality 3.10	112
6.5	Revisiting the lower bound on adaptive estimation error	112

This chapter presents an adaptive version of the Hill estimator based on Lespki’s model selection method. This simple data-driven index selection method is shown to satisfy an oracle inequality and is checked to achieve the lower bound recently derived by Carpentier and Kim. In order to establish the oracle inequality, we derive non-asymptotic variance bounds and concentration inequalities for Hill estimators. These concentration inequalities are derived from Talagrand’s concentration inequality for smooth functions of independent exponentially distributed random variables combined with three tools of Extreme Value Theory : the quantile transform, Karamata’s representation of slowly varying functions, and Rényi’s characterisation of the order statistics of exponential samples. The performance of this computationally and conceptually simple method is illustrated using Monte-Carlo simulations.

1 Introduction

The basic questions faced by Extreme Value Analysis consist in estimating the probability of exceeding a threshold that is larger than the sample maximum and estimating a quantile of an order that is larger than 1 minus the reciprocal of the sample size. In words, they consist in making inferences on regions that lie outside the support of the empirical distribution. In order to face these challenges in a sensible framework, Extreme Value Theory (EVT) assumes that the sampling distribution F satisfies a regularity condition. Indeed, in heavy-tail analysis, the tail function $\bar{F} = 1 - F$ is supposed to be regularly varying that is, $\lim_{\tau \rightarrow \infty} \bar{F}(\tau x)/\bar{F}(\tau)$ exists for all $x > 0$. This amounts to assume the existence of some $\gamma > 0$ such that the limit is $x^{-1/\gamma}$ for all x . In other words, if we define the *excess distribution above the threshold* τ by its survival function : $x \mapsto \bar{F}_\tau(x) = \bar{F}(x)/\bar{F}(\tau)$ for $x \geq \tau$, then \bar{F} is regularly varying if and only if \bar{F}_τ converges weakly towards a Pareto distribution. The sampling distribution F is then said to belong to the *max-domain of attraction* of a Fréchet distribution with index $\gamma > 0$ (abbreviated in $F \in \text{MDA}(\gamma)$) and γ is called the *extreme value index*.

The main impediment to large exceedance and large quantile estimation problems alluded above turns out to be the estimation of the extreme value index. Since the inception of Extreme Value Analysis, many estimators have been defined, analysed and implemented into software. Hill (1975) introduced a simple, yet remarkable, collection of estimators : for $k < n$,

$$\hat{\gamma}_H(k) = \frac{1}{k} \sum_{i=1}^k \ln \frac{X_{(i)}}{X_{(k+1)}} = \frac{1}{k} \sum_{i=1}^k i \ln \frac{X_{(i)}}{X_{(i+1)}} \text{ledoux\&talagrand1991}$$

where $X_{(1)} \geq \dots \geq X_{(n)}$ are the *order statistics* of the sample X_1, \dots, X_n (the non-increasing rearrangement of the sample).

An integer sequence (k_n) is said to be *intermediate* if $\lim_{n \rightarrow \infty} k_n = \infty$ while $\lim_{n \rightarrow \infty} k_n/n = 0$. It is well known that F belongs to $\text{MDA}(\gamma)$ for some $\gamma > 0$ if and only if, for all intermediate sequences (k_n) , $\hat{\gamma}_H(k_n)$ converges in probability towards γ (de Haan & Ferreira, 2006; Mason, 1982). Under mildly stronger conditions, it can be shown that $\sqrt{k_n}(\hat{\gamma}_H(k_n) - \mathbb{E}\hat{\gamma}_H(k_n))$ is asymptotically Gaussian with variance γ^2 . This suggests that, in order to minimise the quadratic risk $\mathbb{E}[(\hat{\gamma}_H(k_n) - \gamma)^2]$ or the absolute risk $\mathbb{E}|\hat{\gamma}_H(k_n) - \gamma|$, an appropriate choice for k_n has to be made. If k_n is too large, the Hill estimator $\hat{\gamma}_H(k_n)$ suffers a large bias and, if k_n is too small, $\hat{\gamma}_H(k_n)$ suffers erratic fluctuations.

As all estimators of the extreme value index face this dilemma (See Beirlant et al., 2004; de Haan & Ferreira, 2006; Resnick, 2007, and references therein), during the last three decades, a variety of data-driven selection methods for k_n has been proposed in the literature (See Hall & Weissman (1997), Hall & Welsh (1985), Danielsson et al. (2001), Draisma et al. (1999), Drees & Kaufmann (1998), Drees et al. (2000), Grama & Spokoiny (2008), Carpentier & Kim (2014a) to name a few). A related but distinct problem is considered by Carpentier & Kim (2014b) : constructing uniform and adaptive confidence intervals for the extreme value index.

The rationale for investigating adaptive Hill estimation stems from computational simplicity and variance optimality of properly chosen Hill estimators (Beirlant et al., 2006).

The hallmark of our approach is to combine techniques of Extreme Value Theory with tools from concentration of measure theory. As, up to our knowledge, the impact of the concentration of measure phenomenon in Extreme Value Theory has received little attention, we comment and motivate the use of concentration arguments. Talagrand's concentration phenomenon for products of exponential distributions is one instance of a general phenomenon : concentration of measure in product spaces (Ledoux, 2001; Ledoux & Talagrand, 1991). The phenomenon may be summarised in a simple quote : functions of independent random variables that do not depend too much on any of them are almost constant (Talagrand, 1996b).

The concentration approach helps to split the investigation in two steps : the first step consists in bounding the fluctuations of the random variable under concern around its median or its expectation, while the second step focuses on the expectation. This approach has seriously simplified the investigation of suprema of empirical processes and made the life of many statisticians easier (Koltchinskii, 2008; Massart, 2006; Talagrand, 1996a, 2005). To point out the potential uses of concentration inequalities in the field of Extreme Value Theory is one purpose of this paper. In statistics, concentration inequalities have proved very useful when dealing with estimator selection and adaptivity issues : sharp, non-asymptotic tail bounds can be combined with simple union bounds in order to obtain uniform guarantees of the risk of collection of estimators. Using concentration inequalities to investigate adaptive choice of the number of order statistics to be used in tail index estimation is a natural thing to do.

In the present setting, tail index estimators are functions of independent random va-

riables. Talagrand’s quote raises a first question : in which way are these tail functionals smooth functions of independent random variables ? We do not attempt here to revisit the asymptotic approach described by (Drees, 1998b) which equates smoothness with Hadamard differentiability. Our approach is non-asymptotic and our conception of smoothness somewhat circular, smooth functionals are these functionals for which we can obtain good concentration inequalities.

In this paper, we combine Talagrand’s concentration inequality for smooth functions of independent exponentially distributed random variables (Theorem 3.9) with three traditional tools of EVT : the quantile transform, Karamata’s representation for slowly varying functions, and Rényi’s characterisation of the joint distribution of order statistics of exponential samples. This allows us to establish concentration inequalities for the Hill process $(\sqrt{k}(\widehat{\gamma}_H(k) - \mathbb{E}\widehat{\gamma}_H(k)))_k$ (Theorems 3.13 and 3.17 and Corollary 3.19) in Section 3.2.

In Section 3.3, we build on these concentration inequalities to analyse the performance of a variant of Lepski’s rule defined in Sections 2.3 and 3.3 : Theorem 3.20 describes an oracle inequality and Corollary 3.23 assesses the performance of this simple selection rule under the assumption that

$$|\overline{F}(x)x^{1/\gamma} - C| \leq C'x^{\rho/\gamma}$$

for some unknown $\rho < 0$ and $C, C' > 0$. Note that this condition is less demanding than the so-called second order condition that has often been assumed when looking for adaptive tail index estimators (notable exceptions being (Carpentier & Kim, 2014a) and (Grama & Spokoiny, 2008)). It reveals that the performance of Hill estimators selected by Lepski’s method matches known lower bounds (See Section 2.4), that is, they suffer the loss of efficiency that is inherent to this problem, but not more.

Proofs are given in Section 4. Finally, in Section 5, we examine the performance of this resulting adaptive Hill estimator for finite sample sizes using Monte-Carlo simulations.

2 Background, notations and tools

2.1 The Hill estimator as a smooth tail statistics

The quantile function F^{\leftarrow} is the generalised inverse of the distribution function F . The *tail quantile function* of F is a non-decreasing function defined on $(1, \infty)$ by $U = (1/(1 - F))^{\leftarrow}$, or

$$U(t) = \inf\{x : F(x) \geq 1 - 1/t\} = F^{\leftarrow}(1 - 1/t).$$

In this text, we use a variation of the quantile transform that fits EVT : if E is exponentially distributed, then $U(\exp(E))$ is distributed according to F . Moreover, by the same argument, the order statistics $X_{(1)} \geq \dots \geq X_{(n)}$ are distributed as a monotone transformation of the order statistics $Y_{(1)} \geq \dots \geq Y_{(n)}$ of a sample of n independent standard exponential random variables,

$$(X_{(1)}, \dots, X_{(n)}) \stackrel{d}{=} (U(e^{Y_{(1)}}), \dots, U(e^{Y_{(n)}})) .$$

Thanks to Rényi's representation for order statistics of exponential samples, agreeing on $Y_{(n+1)} = 0$, the rescaled exponential spacings $Y_{(1)} - Y_{(2)}, \dots, i(Y_{(i)} - Y_{(i+1)}), \dots, (n-1)(Y_{(n-1)} - Y_{(n)}), nY_{(n)}$ are independent and exponentially distributed.

The quantile transform and Rényi's representation are complemented by Karamata's representation for slowly varying functions. Recall that a function L is *slowly varying at infinity* if for all $x > 0$, $\lim_{t \rightarrow \infty} L(tx)/L(t) = x^0 = 1$. The von Mises condition specifies the form of Karamata's representation (See Resnick, 2007, Corollary 2.1) of the slowly varying component of $U(t)$ ($t^{-\gamma}U(t)$).

Definition 3.1 (VON MISES CONDITION). *A distribution function F belonging to $MDA(\gamma)$, $\gamma > 0$, satisfies the von Mises condition if there exist a constant $t_0 \geq 1$, a measurable function η on $(1, \infty)$ and a constant $c = U(t_0)t_0^{-\gamma}$ such that for $t \geq t_0$*

$$U(t) = ct^\gamma \exp\left(\int_{t_0}^t \frac{\eta(s)}{s} ds\right)$$

with $\lim_{s \rightarrow \infty} \eta(s) = 0$. The function η is called the von Mises function.

In the sequel, we assume that the sampling distribution $F \in MDA(\gamma)$, $\gamma > 0$, satisfies the von Mises condition with $t_0 = 1$, von Mises function η and define the non-increasing function $\bar{\eta}$ from $[1, \infty)$ to $[0, \infty)$ by $\bar{\eta}(t) = \sup_{s \geq t} |\eta(s)|$.

Combining the quantile transformation, Rényi's and Karamata's representations, it is straightforward that under the von Mises condition, the sequence of Hill estimators is distributed as a function of the largest order statistics of a standard exponential sample.

Proposition 3.2. *The vector of Hill estimators $(\hat{\gamma}_H(k))_{k < n}$ is distributed as the random vector*

$$\left(\frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{\frac{u}{i} + Y_{(i+1)}})) du\right)_{k < n} \quad (3.1)$$

where (E_1, \dots, E_n) are independent standard exponential random variables while, for $i \leq n$, $Y_{(i)} = \sum_{j=i}^n E_j/j$ is distributed like the i th order statistic of an n -sample of the exponential distribution.

For a fixed $k < n$, a second distributional representation is available,

$$\hat{\gamma}_H(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{u + Y_{(k+1)}})) du \quad (3.2)$$

where (E_1, \dots, E_k) and $Y_{(k+1)}$ are defined as in Proposition 3.2.

This second, simpler, distributional representation stresses the fact that, conditionally on $Y_{(k+1)}$, $\hat{\gamma}_H(k)$ is distributed as a mixture of sums of independent random variables approximately distributed as exponential random variables with scale γ . This distributional identity suggests that the variance of $\hat{\gamma}_H(k)$ scales like γ^2/k , an intuition that is corroborated by analysis, see Section 3.1.

The bias of $\hat{\gamma}_H(k)$ is connected with the von Mises function η by the next formula

$$\mathbb{E}\hat{\gamma}_H(k) - \gamma = \mathbb{E}\left[\int_0^\infty e^{-v} \eta(e^{Y_{(k+1)} + v}) dv\right] = \mathbb{E}\left[\int_1^\infty \frac{\eta(e^{Y_{(k+1)} + v})}{v^2} dv\right].$$

Henceforth, let b be defined for $t > 1$ by

$$b(t) = \int_1^\infty \frac{\eta(tv)}{v^2} dv = t \int_t^\infty \frac{\eta(v)}{v^2} dv. \quad (3.3)$$

The quantity $b(t)$ is the bias of the Hill estimator $\hat{\gamma}_H(k)$ given $\bar{F}(X_{(k+1)}) = 1/t$. The second expression for b shows that b is differentiable with respect to t (even though η might be nowhere differentiable), and that

$$b'(t) = \frac{b(t) - \eta(t)}{t}.$$

The von Mises function governs both the rate of convergence of $U(tx)/U(t)$ towards x^γ , or equivalently of $\bar{F}(tx)/\bar{F}(t)$ towards $x^{-1/\gamma}$, and the rate of convergence of $|\mathbb{E}\hat{\gamma}_H(k) - \gamma|$ towards 0.

In the sequel, we work on the probability space where the independent standard exponential random variables $E_i, 1 \leq i \leq n$ are all defined, and therefore consider the Hill estimators defined by Representation (3.1).

2.2 Frameworks

The difficulty in extreme value index estimation stems from the fact that, for any collection of estimators, for any intermediate sequence (k_n) , and for any $\gamma > 0$, there is a distribution function $F \in \text{MDA}(\gamma)$ such that the bias $|\mathbb{E}\hat{\gamma}_H(k_n) - \gamma|$ decays at an arbitrarily slow rate. This has led authors to put conditions on the rate of convergence of $U(tx)/U(t)$ towards x^γ as t tends to infinity while $x > 0$, or equivalently on the rate of convergence of $\bar{F}(tx)/\bar{F}(t)$ towards $x^{-1/\gamma}$. These conditions have then to be translated into conditions on the rate of decay of the bias of estimators. As we focus on Hill estimators, the connection between the rate of convergence of $U(tx)/U(t)$ towards x^γ and the rate of decay of the bias is transparent and well-understood (Segers, 2002) : the theory of O -regular variation provides an adequate setting for describing both rates of convergence (Bingham et al., 1987). In words, if a positive function g defined over $[1, \infty)$ is such that for some $\alpha \in \mathbb{R}$, for all $\Lambda > 1$, $\limsup_t \sup_{x \in [1, \Lambda]} g(tx)/g(t) < \infty$, g is said to have *bounded increase*. If g has bounded increase, the class $O\Pi_g$ is the class of measurable functions f on some interval $[a, \infty), a > 0$, such that as $t \rightarrow \infty$, $f(tx) - f(t) = O(g(t))$ for all $x \geq 1$.

For example, the analysis carried out by Carpentier & Kim (2014a) rests on the condition that, if $F \in \text{MDA}(\gamma)$, for some $C > 0$, $D \neq 0$ and $\rho < 0$,

$$\left| \frac{\bar{F}(x)}{x^{-1/\gamma}} - C \right| \leq Dx^{\rho/\gamma}. \quad (3.4)$$

This condition implies that $\ln(t^{-\gamma}U(t)) \in O\Pi_g$ with $g(t) = t^\rho$ (Segers, 2002, p. 473). Thus under the von Mises condition, Condition (3.4) implies that the function $\int_t^\infty (\eta(s)/s) ds$ belongs to $O\Pi_g$ with $g(t) = t^\rho$. Moreover, the Abelian and Tauberian Theorems from (Segers, 2002) assert that $\int_t^\infty (\eta(s)/s) ds \in O\Pi_g$ if and only if $|\mathbb{E}\hat{\gamma}_H(k_n) - \gamma| = O(g(n/k_n))$ for any intermediate sequence (k_n) .

In this text, we are ready to assume that if $F \in \text{MDA}(\gamma)$, then for some $C > 0$ and $\rho < 0$,

$$|\mathbb{E}\widehat{\gamma}_H(k_n) - \gamma| \leq C \left(\frac{n}{k_n} \right)^\rho.$$

However, we do not want to assume that U (or equivalently \overline{F}) satisfies a so-called second-order regular variation property ($t \mapsto |x^{-\gamma}U(tx)/U(t) - 1|$ is asymptotically equivalent to a ρ -regularly varying function where $\rho < 0$). By (Segers, 2002), this would be equivalent to assuming that $t \mapsto |b(t)|$ is ρ -regularly varying.

Indeed, assuming as in (Hall & Welsh, 1985) and several subsequent papers that F satisfies

$$\overline{F}(x) = Cx^{-1/\gamma} (1 + Dx^{\rho/\gamma} + o(x^{\rho/\gamma})) \quad (3.5)$$

where $C > 0, D \neq 0$ are constants and $\rho < 0$, or equivalently (Csörgő, Deheuvels, & Mason, 1985; Drees & Kaufmann, 1998) that U satisfies

$$U(t) = C^\gamma t^\gamma (1 + \gamma DC^\rho t^\rho + o(t^\rho))$$

makes the problem of extreme value index estimation easier (but not easy). These conditions entail that, for any intermediate sequence (k_n) , the ratio $|\mathbb{E}[\widehat{\gamma}_H(k_n) - \gamma]|/(n/k_n)^\rho$ converges towards a finite limit as n tends to ∞ (Beirlant et al., 2004; de Haan & Ferreira, 2006; Segers, 2002), this makes the estimation of the second-order parameter a very natural intermediate objective (See for example Drees & Kaufmann, 1998).

2.3 Lepski's method and adaptive tail index estimation

The necessity of developing data-driven index selection methods is illustrated in Figure 3.1, which displays the estimated standardised root mean squared error (RMSE) of Hill estimators

$$\mathbb{E} \left[\left(\frac{\widehat{\gamma}_H(k)}{\gamma} - 1 \right)^2 \right]^{1/2}$$

as a function of k for four related sampling distributions which all satisfy the second-order condition (3.5) with different values of the second-order parameters.

Under this second-order condition (3.5), Hall & Welsh proved that the asymptotic mean squared error of the Hill estimator is minimal for sequences $(k_n^*)_n$ satisfying

$$k_n^* \sim K(C, D, \rho) n^{2|\rho|/(1+2|\rho|)}$$

with $K(C, D, \rho) = (C^{2|\rho|}(1 + |\rho|)^2/2D^2|\rho|^3)^{1/(1+2|\rho|)}$. Since, $C > 0, D \neq 0$ and the second-order parameter $\rho < 0$ are usually unknown, many authors have been interested in the construction of data-driven selection procedure for k_n under conditions such as (3.5), a great deal of ingenuity has been dedicated to the estimation of the second-order parameters and to the use of such estimates when estimating first order parameters.

As we do not want to assume a second-order condition such as (3.5), we resort to Lepski's method which is a general attempt to balance bias and variance.

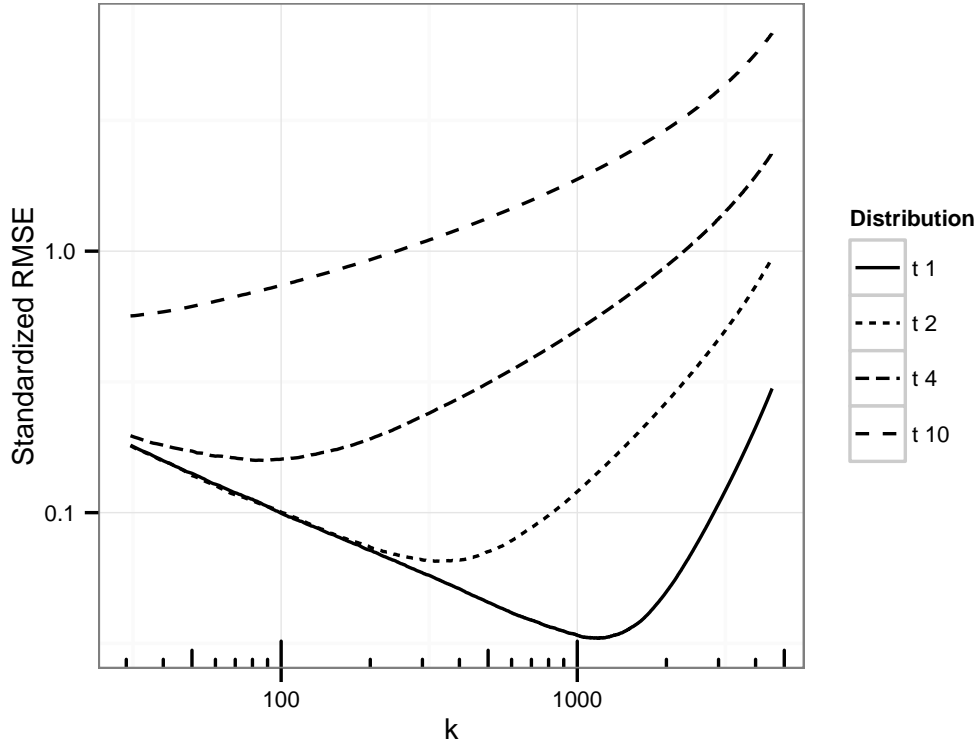


FIGURE 3.1. Estimated standardised RMSE as a function of k for samples of size 10000 from Student's distributions with different degrees of freedom $\nu = 1, 2, 4, 10$. All four distributions satisfy condition (3.5) with $|\rho| = 2/\nu$. The increasing parts of the lines reflect the values of ρ . RMSE is estimated by averaging over 5000 Monte-Carlo simulations.

Since its introduction (Lepski, 1991), this general method for model selection that has been proved to achieve adaptivity and provide one with oracle inequalities in a variety of inferential contexts ranging from density estimation to inverse problems and classification (Lepski, 1990, 1991, 1992; Lepski & Tsybakov, 2000). Very readable introductions to Lepski's method and its connections with penalised contrast methods can be found in (Birgé, 2001; Mathé, 2006). In Extreme Value Theory, we are aware of three papers that explicitly rely on this methodology : (Drees & Kaufmann, 1998), (Grama & Spokoiny, 2008) and (Carpentier & Kim, 2014a).

The selection rule analysed in the present paper (see Section 3.3 for a precise definition) is a variant of the preliminary selection rule introduced in (Drees & Kaufmann, 1998)

$$\bar{\kappa}_n(r_n) = \min \left\{ k \in \{2, \dots, n\} : \max_{2 \leq i \leq k} \sqrt{i} |\hat{\gamma}_H(i) - \hat{\gamma}_H(k)| > r_n \right\} \quad (3.6)$$

where $(r_n)_n$ is a sequence of thresholds such that $\sqrt{\ln \ln n} = o(r_n)$ and $r_n = o(\sqrt{n})$, and $\hat{\gamma}_H(i)$ is the Hill estimator computed from the $(i+1)$ largest order statistics. The definition of this "stopping time" is motivated by Lemma 1 from (Drees & Kaufmann, 1998) which asserts that, under the von Mises condition,

$$\max_{2 \leq i \leq k_n} \sqrt{i} |\hat{\gamma}_H(i) - \mathbb{E}[\hat{\gamma}_H(i)]| = O_P \left(\sqrt{\log \log n} \right).$$

In words, this selection rule almost picks out the largest index k such that, for all i smaller than k , $\hat{\gamma}_H(k)$ differs from $\hat{\gamma}_H(i)$ by a quantity that is not much larger than the typical fluctuations of $\hat{\gamma}_H(i)$. This index selection rule can be performed graphically by interpreting an alternative Hill plot as shown on Figure 3.2 (See Drees et al., 2000; Resnick, 2007, for a discussion on the merits of alt-Hill plots).

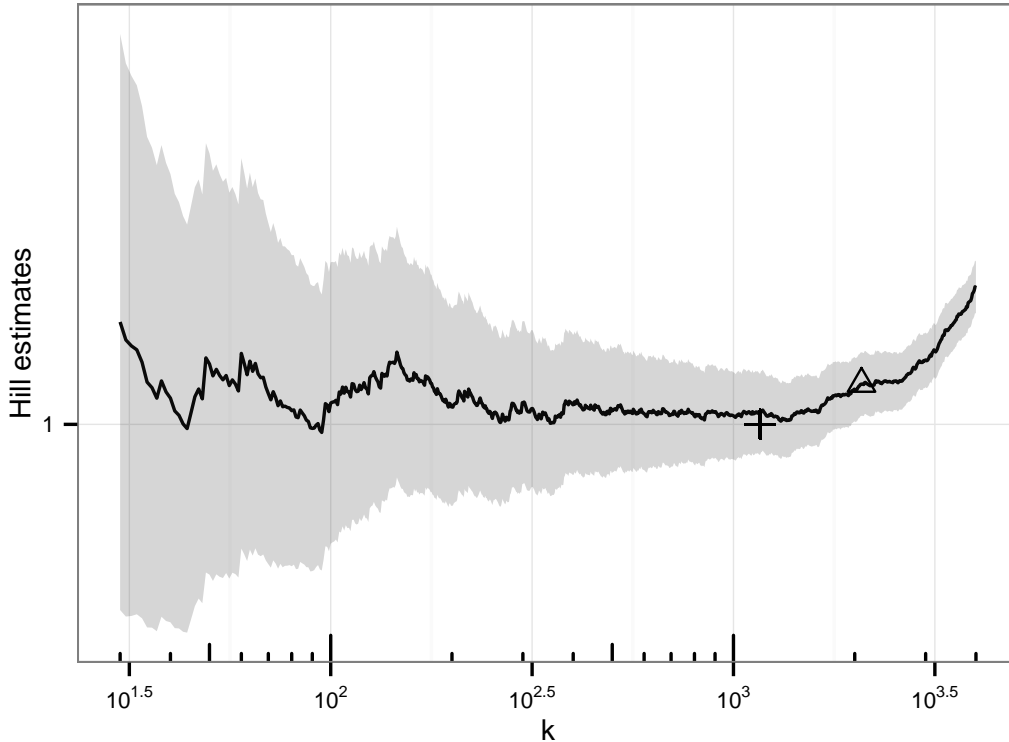


FIGURE 3.2. Lepski's method illustrated on a alt-Hill plot. The plain line describes the sequence of Hill estimates as a function of index k computed on a pseudo-random sample of size $n = 10000$ from Student distribution with 1 degree of freedom (Cauchy distribution). Hill estimators are computed from the positive order statistics. The grey ribbon around the plain line provides a graphic illustration of Lepski's method. For a given value of i , the width of the ribbon is $2r_n\hat{\gamma}_H(i)/\sqrt{i}$. A point $(k, \hat{\gamma}_H(k))$ on the plain line corresponds to an eligible index if the horizontal segment between this point and the vertical axis lies inside the ribbon that is, if for all $i, 30 \leq i < k$, $|\hat{\gamma}_H(k) - \hat{\gamma}_H(i)| \leq r_n\hat{\gamma}_H(i)/\sqrt{i}$. If r_n were replaced by an appropriate quantile of the Gaussian distribution, the grey ribbon would just represent the confidence tube that is usually added on Hill plots. The triangle represents the selected index with $r_n = \sqrt{2.1 \ln \ln n}$. The cross represents the oracle index estimated from Monte-Carlo simulations, see Table 3.2.

The goal of Drees & Kaufmann (1998) was not to investigate the performance of the preliminary selection rule defined in Display (3.6) but to design a selection rule $\hat{\kappa}_n(r_n)$, based on $\bar{\kappa}_n(r_n)$, that would, under second-order conditions, asymptotically mimic the optimal selection rule k_n^* .

Our goal, as in (Carpentier & Kim, 2014a; Grama & Spokoiny, 2008), is to derive non-asymptotic risk bounds without making a second order assumption. In both papers, the rationale for working with some special collection of estimators seems to be the ability

to derive non-asymptotic deviation inequalities for $\widehat{\gamma}_H(k)$ either from exponential inequalities for log-likelihood ratio statistics or from simple binomial tail inequalities such as Bernstein's inequality (See Boucheron et al., 2013, Section 2.8).

In models satisfying Condition (3.5), the estimators from (Grama & Spokoiny, 2008) achieve the optimal rate up to a $\log(n)$ factor. Carpentier & Kim (2014a) prove that the risk of their data-driven estimator decays at the optimal rate $n^{|\rho|/(1+2|\rho|)}$ up to a factor $r_n^{2|\rho|/(1+2|\rho|)} = (\ln \ln n)^{|\rho|/(1+2|\rho|)}$ in models satisfying Condition (3.4).

We aim at achieving optimal risk bounds under Condition (3.4), using a simple estimation method requiring almost no calibration effort and based on mainstream extreme value index estimators. Before describing the keystone of our approach in Section 2.5, we recall the recent lower risk bound for adaptive extreme value index estimation.

2.4 Lower bound

One of the key results in (Carpentier & Kim, 2014a) is a lower bound on the accuracy of adaptive tail index estimation. This lower bound reveals that, just as for estimating a density at a point (Lepski, 1991, 1992), or point estimation in Sobolev spaces (Tsybakov, 1998), as far as tail index estimation is concerned, adaptivity has a price. Using Fano's Lemma, and a Bayesian game that extends cleanly in frameworks of (Grama & Spokoiny, 2008) and (Novak, 2014), Carpentier & Kim were able to prove the next minimax lower bound.

Theorem 3.3. *Let $\rho_0 < -1$, and $0 \leq v \leq e/(1+2e)$. Then, for any tail index estimator $\widehat{\gamma}_H$ and any sample size n such that $M = \lfloor \ln n \rfloor > e/v$, there exists a probability distribution P such that*

- i) $P \in \text{MDA}(\gamma)$ with $\gamma > 0$,
- ii) P meets the von Mises condition with von Mises function η satisfying

$$\bar{\eta}(t) \leq \gamma t^\rho$$

for some $\rho_0 \leq \rho < 0$,

iii)

$$P \left\{ |\widehat{\gamma}_H - \gamma| \geq \frac{C_\rho}{4} \gamma \left(\frac{v \ln \ln n}{n} \right)^{|\rho|/(1+2|\rho|)} \right\} \geq \frac{1}{1+2e}$$

and

$$\mathbb{E}_P \left[\frac{|\widehat{\gamma}_H - \gamma|}{\gamma} \right] \geq \frac{C_\rho}{4(1+2e)} \left(\frac{v \ln \ln n}{n} \right)^{|\rho|/(1+2|\rho|)},$$

with $C_\rho = 1 - \exp\left(-\frac{1}{2(1+2|\rho|)^2}\right)$.

Using Birgé's Lemma instead of Fano's Lemma, we provide a simpler, shorter proof of this theorem (Appendix 6.5).

The lower rate of convergence provided by Theorem 3.3 is another incentive to revisit the preliminary tail index estimator from (Drees & Kaufmann, 1998), but, instead of

using a sequence $(r_n)_n$ of order larger than $\sqrt{\ln \ln n}$ in order to calibrate pairwise tests and ultimately to design estimators of the second-order parameter (if there are any), it is worth investigating a minimal sequence where r_n is of order $\sqrt{\ln \ln n}$, and check whether the corresponding adaptive estimator achieves the Carpentier-Kim lower bound (Theorem 3.3).

In this paper, we focus on r_n of the order $\sqrt{\ln \ln n}$. The rationale for imposing r_n of the order $\sqrt{\ln \ln n}$ can be understood by the fact that, even if the sampling distribution is a pure Pareto distribution with shape parameter γ ($\bar{F}(x) = (x/\tau)^{-1/\gamma}$ for $x \geq \tau > 0$), if

$$\limsup r_n / (\gamma \sqrt{2 \ln \ln n}) < 1,$$

the preliminary selection rule will, with high probability, select a small value of k and thus pick out a suboptimal estimator. This can be justified using results from (Darling & Erdős, 1956) (See Appendix 6.1 for details).

Such an endeavour requires sharp probabilistic tools. They are the topic of the next section.

2.5 Talagrand's concentration phenomenon for products of exponential distributions

Deriving authentic concentration inequalities for Hill estimators is not straightforward. Fortunately, the construction of such inequalities turns out to be possible thanks to general functional inequalities that hold for functions of independent exponentially distributed random variables. We recall these inequalities (Proposition 3.4 and Theorem 3.9) that have been largely overlooked in statistics. A thorough and readable presentation of these inequalities can be found in (Ledoux, 2001). We start by the easiest result, a variance bound that pertains to the family of Poincaré inequalities.

Proposition 3.4 (Poincaré inequality for exponentials, (Bobkov & Ledoux, 1997)). *If g is a differentiable function over \mathbb{R}^n , and $Z = g(E_1, \dots, E_n)$ where E_1, \dots, E_n are independent standard exponential random variables, then*

$$\text{Var}(Z) \leq 4\mathbb{E} [\|\nabla g\|^2] .$$

Remark 3.5. The constant 4 can not be improved.

The next corollary is stated in order to point the relevance of this Poincaré inequality to the analysis of general order statistics and their functionals. Recall that the *hazard rate* of an absolutely continuous probability distribution with distribution F is : $h = f/\bar{F}$ where f and $\bar{F} = 1 - F$ are the density and the survival function associated with F , respectively.

Corollary 3.6. *Assume the distribution of X has a positive density, then the k th order statistic $X_{(k)}$ satisfies*

$$\text{Var}(X_{(k)}) \leq C \sum_{i=k}^n \frac{1}{i^2} \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right] \leq \frac{C}{k} \left(1 + \frac{1}{k} \right) \mathbb{E} \left[\frac{1}{h(X_{(k)})^2} \right]$$

where C can be chosen as 4.

Remark 3.7. By Smirnov's Lemma (de Haan & Ferreira, 2006), C can not be smaller than 1. If the distribution of X has a non-decreasing hazard rate, the factor of 4 can be improved into a factor 2 (Boucheron & Thomas, 2012).

Bobkov & Ledoux (1997); Maurey (1991); Talagrand (1991) show that smooth functions of independent exponential random variables satisfy Bernstein type concentration inequalities. The next result is extracted from the derivation of Talagrand's concentration phenomenon for product of exponential random variables in (Bobkov & Ledoux, 1997).

The definition of sub-gamma random variables will be used in the formulation of the theorem and in many arguments.

Definition 3.8. A real-valued centered random variable X is said to be sub-gamma on the right tail with variance factor v and scale parameter c if

$$\ln \mathbb{E} e^{\lambda X} \leq \frac{\lambda^2 v}{2(1 - c\lambda)} \text{ for every } \lambda \text{ such that } 0 < \lambda < 1/c.$$

We denote the collection of such random variables by $\Gamma_+(v, c)$. Similarly, X is said to be sub-gamma on the left tail with variance factor v and scale parameter c if $-X$ is sub-gamma on the right tail with variance factor v and tail parameter c . We denote the collection of such random variables by $\Gamma_-(v, c)$.

If $X - \mathbb{E}X \in \Gamma_+(v, c)$, then for all $\delta \in (0, 1)$, then with probability larger than $1 - \delta$,

$$X \leq \mathbb{E}X + \sqrt{2v \ln(1/\delta)} + c \ln(1/\delta).$$

The entropy of a non-negative random variable X is defined by $\text{Ent}[X] = \mathbb{E}[X \log X] - \mathbb{E}X \log \mathbb{E}X$.

Theorem 3.9. If g is a differentiable function on \mathbb{R}^n with $\max_i |\partial_i g| < \infty$, and $Z = g(E_1, \dots, E_n)$ where E_1, \dots, E_n are n independent standard exponential random variables. Let $c < 1$, then for all λ such that $0 \leq \lambda \max_i |\partial_i g| \leq c$,

$$\text{Ent} [e^{\lambda(Z - \mathbb{E}Z)}] \leq \frac{2\lambda^2}{1 - c} \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)} \|\nabla g\|^2].$$

Let v be the essential supremum of $\|\nabla g\|^2$, then Z is sub-gamma on both tails with variance factor $4v$ and scale factor $\max_i |\partial_i g|$.

Again, we illustrate the relevance of these versatile tools on the analysis of general order statistics. This general theorem implies that if the sampling distribution has non-decreasing hazard rate, then the order statistics $X_{(k)}$ satisfy Bernstein type inequalities (see Boucheron et al., 2013, Section 2.8) with variance factor $4/k \mathbb{E} [1/h(X_{(k)})^2]$ (the Poincaré estimate of variance), and scale parameter $(\sup_x 1/h(x))/k$. Starting back from the Efron-Stein-Steele inequality, the authors derived a somewhat sharper inequality (Boucheron & Thomas, 2012).

Corollary 3.10. *Assume the distribution function F has non-decreasing hazard rate h that is, $U \circ \exp$ is C^1 and concave. Let $Z = g(E_1, \dots, E_n) = (U \circ \exp)(\sum_{i=k}^n E_i/i)$ be distributed as the k th order statistic of a sample distributed according to F .*

Then Z is sub-gamma on both tails with variance factor $4/k(1 + 1/k)\mathbb{E}[1/h(Z)^2]$ and scale factor $1/(k \inf_x h(x))$.

This corollary describes in which way central, intermediate and extreme order statistics can be portrayed as smooth functions of independent exponential random variables. This possibility should not be taken for granted as it is non trivial to capture in a non-asymptotic way the tail behaviour of maxima of independent Gaussians (Boucheron & Thomas, 2012; Chatterjee, 2014; Ledoux, 2001). In the next section, we show in which way the Hill estimator can fit into this picture.

3 Main results

In this section, the sampling distribution F is assumed to belong to $\text{MDA}(\gamma)$ with $\gamma > 0$ and to satisfy the von Mises condition (Definition 3.1) with von Mises function η .

3.1 Bounding the variance of the Hill estimator

It is well known that, under the von Mises condition, if (k_n) is an intermediate sequence, the sequence $\sqrt{k_n}(\hat{\gamma}_H(k_n) - \mathbb{E}\hat{\gamma}_H(k_n))$ converges in distribution towards $\mathcal{N}(0, \gamma^2)$ suggesting that the variance of $\hat{\gamma}_H(k_n)$ scales like γ^2/k_n (See Beirlant et al., 2004; de Haan & Ferreira, 2006; Geluk et al., 1997; Resnick, 2007).

In this subsection, we will use the representation (3.2) :

$$\hat{\gamma}_H(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{u+Y_{(k+1)}})) du$$

Proposition 3.11 provides us with a handy upper bound on $\text{Var}[\hat{\gamma}_H(k)] - \gamma^2/k$ using the von Mises function.

Proposition 3.11. *Let $\hat{\gamma}_H(k)$ be the Hill estimator computed from the $(k + 1)$ largest order statistics of an n -sample from F . Then,*

$$-\frac{2\gamma}{k}\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})] \leq \text{Var}[\hat{\gamma}_H(k)] - \frac{\gamma^2}{k} \leq \frac{2\gamma}{k}\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})] + \frac{5}{k}\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})^2].$$

The next Abelian result might help in appreciating these variance bounds.

Proposition 3.12. *Assuming that η is ρ -regular varying with $\rho < 0$, then for any intermediate sequence (k_n) ,*

$$\lim_{n \rightarrow \infty} \frac{k_n \text{Var}(\hat{\gamma}_H(k_n)) - \gamma^2}{\eta(n/k_n)} = \frac{2\gamma}{(1 - \rho)^2}.$$

We may now move to genuine concentration inequalities for the Hill estimator.

3.2 Concentration inequalities for the Hill estimators

The exponential representation (3.1) suggests that the Hill estimator $\hat{\gamma}_H(k)$ should be approximately distributed according to a $\text{gamma}(k, \gamma)$ distribution where k is the shape parameter and γ the scale parameter. We expect the Hill estimators to satisfy Bernstein type concentration inequalities that is, to be sub-gamma on both tails with variance factors connected to the tail index γ and to the von Mises function. Representation (3.1) actually suggests more. Following (Drees & Kaufmann, 1998), we actually expect the sequence $(\sqrt{k}(\hat{\gamma}_H(k) - \mathbb{E}\hat{\gamma}_H(k)))_k$ to behave like normalized partial sums of independent square integrable random variables, that is, we believe $\max_{2 \leq k \leq n} \sqrt{k}(\hat{\gamma}_H(k) - \mathbb{E}\hat{\gamma}_H(k))$ to scale like $\sqrt{\ln \ln n}$ and to be sub-gamma on both tails. The purpose of this section is to meet these expectations in a non-asymptotic way.

Proofs use the Markov property of order statistics : conditionally on the $(J + 1)$ th order statistic, the first largest J order statistics are distributed as the order statistics of a sample of size J of the excess distribution. They consist of appropriate invocations of Talagrand's concentration inequality (Theorem 3.9). However, this theorem generally requires a uniform bound on the gradient of the relevant function. When Hill estimators are analysed as functions of independent exponential random variables, the partial derivatives depend on the points at which the von Mises function is evaluated. In order to get interesting bounds, it is worth conditioning on an intermediate order statistic.

Throughout this subsection, let ℓ be an integer larger than $\sqrt{\ln \log_2 n}$ and J an integer not larger than n . As we use the exponential representation of order statistics, besides Hill estimators, the random variables that appear in the main statements are order statistics of exponential samples, $Y_{(k)}$ will denote the k th order statistic of a standard exponential sample of size n (we agree on $Y_{(n+1)} = 0$).

The first theorem provides an exponential refinement of the variance bound stated in Proposition 3.11. However, as announced, there is a price to pay, statements hold conditionally on some order statistic.

In the sequel, let

$$\xi_n = c_1 \sqrt{\ln \log_2 n} + c'_1,$$

where c_1 may be chosen not larger than 4 and c'_1 not larger than 16.

Theorem 3.13. *Let T be a shorthand for $\exp(Y_{(J+1)})$. For some $k, \ell \leq k \leq J$, let*

$$Z^a = \max_{\ell \leq i \leq k} \sqrt{i} \left| \hat{\gamma}_H(i) - \mathbb{E} [\hat{\gamma}_H(i) \mid Y_{(k+1)}] \right|.$$

Then, conditionally on T ,

i) For $\ell \leq i \leq k$, the random variable

$$\sqrt{i} (\hat{\gamma}_H(i) - \mathbb{E}[\hat{\gamma}_H(i) \mid T])$$

is sub-gamma on both tails with variance factor $4(\gamma + 2\bar{\eta}(T))^2$ and scale factor $(\gamma + \bar{\eta}(T)) / \sqrt{i}$.

ii) The random variable Z^a is sub-gamma on both tails with variance factor $4(\gamma + 2\bar{\eta}(T))^2$ and scale factor $(\gamma + 2\bar{\eta}(T))/\sqrt{\ell}$ and

$$\mathbb{E}[Z^a | T] \leq \xi_n(\gamma + 2\bar{\eta}(T)) . \quad (3.7)$$

Remark 3.14. If F is a pure Pareto distribution with shape parameter $\gamma > 0$, then $k\hat{\gamma}_H(k)/\gamma$ is distributed according to a gamma distribution with shape parameter k and scale parameter 1. Tight and well-known tail bounds for gamma distributed random variables assert that

$$\mathbb{P}\left\{|\hat{\gamma}_H(k) - \mathbb{E}[\hat{\gamma}_H(k)]| \geq \frac{\gamma}{\sqrt{k}} \left(\sqrt{2 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{k}} \right)\right\} \leq 2\delta .$$

Remark 3.15. Statement ii) reads as

$$\mathbb{P}\left\{|Z^a - \mathbb{E}[Z^a | T]| \geq (\gamma + 2\bar{\eta}(T)) \left(\sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\ell} \right)\right\} \leq \delta$$

where $0 < \delta < 1/2$. Combining Statements ii) and iii), we also get

$$\mathbb{P}\left\{Z^a \geq (\gamma + 2\bar{\eta}(T)) \left(\xi_n + \sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\ell} \right)\right\} \leq \delta .$$

Remark 3.16. The reader may wonder whether resorting to the exponential representation and usual Chernoff bounding would not provide a simpler argument. The straightforward approach leads to the following conditional bound on the logarithmic moment generating function,

$$\begin{aligned} & \ln \mathbb{E} \left[\exp \left(\lambda \left(\hat{\gamma}_H(k) - \mathbb{E}[\hat{\gamma}_H(k) | Y_{(k+1)}] \right) \right) \mid Y_{(k+1)} \right] \\ & \leq \frac{(\gamma + \bar{\eta}(e^{Y_{(k+1)}}))^2}{2k(1 - \lambda(\gamma + \bar{\eta}(e^{Y_{(k+1)}})))} + \lambda(\bar{\eta}(e^{Y_{(k+1)}}) - b(e^{Y_{(k+1)}})) . \end{aligned}$$

A similar statement holds for the lower tail. This leads to exponential bounds for deviation of the Hill estimator above $\mathbb{E}[\hat{\gamma}_H(k) | Y_{(k+1)}] + \bar{\eta}(e^{Y_{(k+1)}}) - b(e^{Y_{(k+1)}})$ that is, to control deviations of the Hill estimator above its expectation plus a term that may be of the order of magnitude of the bias.

Attempts to rewrite $\hat{\gamma}_H(k) - \mathbb{E}[\hat{\gamma}_H(k) | Y_{(k+1)}]$ as a sum of martingale increments $\mathbb{E}[\hat{\gamma}_H(k) | Y_{(i)}] - \mathbb{E}[\hat{\gamma}_H(k) | Y_{(i+1)}]$ for $1 \leq i \leq k$ and to exhibit an exponential supermartingale met the same impediments.

At the expense of inflating the variance factor, Theorem 3.9 provides a genuine (conditional) concentration inequality for Hill estimators. As we will deal with values of k for which bias exceeds the typical order of magnitudes of fluctuations, this is relevant to our purpose.

The next theorem is concerned with the fluctuations of the conditional bias of Hill estimators. It states that the fluctuations of conditional expectations $|\mathbb{E}[\hat{\gamma}_H(i) | Y_{(k+1)}]|$ are small and even negligible with respect to the fluctuations of $\hat{\gamma}_H(i)$. Let J such that $\ell \leq k \leq J \leq n$, and again $T = \exp(Y_{(J+1)})$.

Theorem 3.17. For all $1 \leq i \leq k$, let the random variable Z^b be defined by

$$Z^b = \max_{\ell \leq i \leq k} |\mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}] - \mathbb{E}[\widehat{\gamma}_H(i)]| .$$

Then, conditionally on T ,

i) For all $1 \leq i \leq k$,

$$\mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}] - \mathbb{E}[\widehat{\gamma}_H(i)]$$

is sub-gamma on both tails with variance factor at most $16\bar{\eta}(T)^2/k$ and scale factor $2\bar{\eta}(T)/k$.

ii) Z^b is sub-gamma on both tails with variance factor $16\bar{\eta}(T)^2/k$ and scale factor $2\bar{\eta}(T)/k$, and

$$\mathbb{E}Z^b \leq 10\sqrt{\frac{\mathbb{E}[\bar{\eta}(T)^2]}{k}} .$$

Remark 3.18. Statement ii) in Theorem 3.17 can be summarised by the following inequality. For any $0 < \delta < 1/2$,

$$\mathbb{P} \left\{ Z^b \geq \frac{2\bar{\eta}(1)}{\sqrt{k}} \left(5 + \sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{k}} \right) \right\} \leq \delta . \quad (3.8)$$

Combining Theorem 3.13, Theorem 3.17 leads to a non-asymptotic perspective on Lemma 1 from (Drees & Kaufmann, 1998).

Corollary 3.19. Let k be such that $\ell \leq k \leq n$ and let $T = \exp(Y_{(k+1)})$, then

$$\mathbb{E} \left[\max_{\ell \leq i \leq k} \sqrt{i} |\widehat{\gamma}_H(i) - \mathbb{E}\widehat{\gamma}_H(i)| \right] \leq \xi_n(\gamma + 2\mathbb{E}\bar{\eta}(T)) + 10\sqrt{\mathbb{E}[\bar{\eta}(T)^2]} .$$

For $0 < \delta < 1/3$, with probability larger than $1 - 3\delta$,

$$\begin{aligned} & \max_{\ell \leq i \leq k} \sqrt{i} |\widehat{\gamma}_H(i) - \mathbb{E}\widehat{\gamma}_H(i)| \\ & \leq \xi_n(\gamma + 2\bar{\eta}(n/k')) + 10\sqrt{\mathbb{E}[\bar{\eta}(T)^2]} \\ & \quad + 2(\gamma + 2\bar{\eta}(n/k') + 2\bar{\eta}(1))\sqrt{2 \ln(2/\delta)} \\ & \quad + \left(\frac{(\gamma + 2\bar{\eta}(n/k'))}{\sqrt{\ell}} + \frac{2\bar{\eta}(1)}{\sqrt{k}} \right) \ln(2/\delta) \end{aligned}$$

where $k' = k + \frac{\ln(1/\delta)}{2}$.

3.3 Adaptive Hill estimation

We are now able to characterise the performance of the variant of the selection rule defined by (3.6) (Drees & Kaufmann, 1998) with $r_n = c_2\sqrt{\ln \ln n}$ where c_2 is a constant not smaller than $\sqrt{2}$.

The deterministic sequence of indices $(\tilde{k}_n(r_n))$ is defined by :

$$\tilde{k}_n(r_n) = \min \left\{ k \in \{c_3 \ln n, \dots, n\} : |\mathbb{E}[\widehat{\gamma}_H(k) - \gamma]| > \frac{\gamma r_n}{\sqrt{k}} \right\} - 1 .$$

and the sequence $(\tilde{k}_n(1))_n$ is defined by

$$\tilde{k}_n(1) = \min \left\{ k \in \{c_3 \ln n, \dots, n\} : |\mathbb{E}[\hat{\gamma}_H(k) - \gamma]| > \frac{\gamma}{\sqrt{k}} \right\} - 1.$$

The deterministic sequences $(\tilde{k}_n(1))_n$ and $(\tilde{k}_n(r_n))_n$ achieve specific balances between bias and variance. In full generality, it is difficult to precisely connect $(\tilde{k}_n(1))_n$ and $(\tilde{k}_n(r_n))_n$ with the oracle sequence $(k_n^*)_n$. We call these two sequences the pivotal sequences.

Let $0 < \delta < 1/2$. The index \hat{k}_n is selected according to the following rule :

$$\hat{k}_n = \min \left\{ k \in \{c_3 \ln n, \dots, n\} \text{ and } \exists i \in \{\ell, \dots, k\}, |\hat{\gamma}_H(i) - \hat{\gamma}_H(k)| > \frac{\hat{\gamma}_H(i)r_n(\delta)}{\sqrt{i}} \right\} - 1$$

where c_3 is a constant larger than 60 and $r_n(\delta) = 8\sqrt{2 \ln((2/\delta) \log_2 n)}$. The tail index estimator is $\hat{\gamma}_H(\hat{k}_n)$.

Note that

$$\tilde{k}_n(r_n) = \min \left\{ k \in \{c_3 \ln n, \dots, n\} : \max_{c_3 \ln n \leq i \leq k} \sqrt{i} \frac{|\mathbb{E}[\hat{\gamma}_H(i) - \gamma]|}{\gamma} > r_n \right\} - 1.$$

while

$$\tilde{k}_n(1) = \min \left\{ k \in \{c_3 \ln n, \dots, n\} : \max_{c_3 \ln n \leq i \leq k} \sqrt{i} \frac{|\mathbb{E}[\hat{\gamma}_H(i) - \gamma]|}{\gamma} > 1 \right\} - 1.$$

Thus, \hat{k}_n is an empirical version of $\tilde{k}_n(r_n)$. In the sequel, \tilde{k}_n stands for $\tilde{k}_n(r_n)$. If the context is not clear, we specify $\tilde{k}_n(1)$ or $\tilde{k}_n(r_n)$.

As tail adaptivity has a price (see Theorem 3.3), the ratio between the risk of the data-driven estimator $\hat{\gamma}_H(\hat{k}_n)$ and the risk of the pivotal index $\hat{\gamma}_H(\tilde{k}_n(1))$ cannot be upper bounded by a constant factor, let alone by a factor close to 1. This is why in the next theorem, we compare the risk of $\hat{\gamma}_H(\hat{k}_n)$ with the risk of $\hat{\gamma}_H(\tilde{k}_n)$.

Recall that

$$\xi_n = c_1 \sqrt{\ln \log_2 n} + c'_1 \quad \text{with } c_1 \leq 4 \text{ and } c'_1 \leq 16.$$

The next theorem describes a non-asymptotic risk bound for $\hat{\gamma}_H(\hat{k}_n)$. As \tilde{k}_n is usually not an oracle index, this theorem is not an oracle inequality in the usual sense, but a comparison between the risk of a data-driven index selection method and the risk of a pivotal index that achieves a certain balance between bias and variance.

Theorem 3.20. *Assume the sampling distribution $F \in \text{MDA}(\gamma)$, $\gamma > 0$ satisfies the von Mises condition with von Mises function η , and $\bar{\eta}(t) = \sup_{s \geq t} |\eta(s)|$.*

Assume that n is large enough so that

- i) $\bar{\eta}(1) < \xi_n \gamma$,
- ii) $\bar{\eta}\left(\frac{n}{\tilde{k}_n + \frac{\ln(1/\delta)}{2}}\right) < \gamma/4$,

With probability larger than $1 - 3\delta$,

$$\left| \gamma - \hat{\gamma}_H(\hat{k}_n) \right| \leq \left| \gamma - \hat{\gamma}_H(\tilde{k}_n) \right| \left(1 + \frac{r_n(\delta)}{\sqrt{\tilde{k}_n}} \right) + \frac{r_n(\delta)}{\sqrt{\tilde{k}_n}} \gamma.$$

With probability larger than $1 - 4\delta$,

$$\left| \widehat{\gamma}_H(\widehat{k}_n) - \gamma \right| \leq \frac{2r_n(\delta)}{\sqrt{\widehat{k}_n}} \gamma (1 + \alpha(\delta, n)), \quad (3.9)$$

where

$$\begin{aligned} \alpha(\delta, n) \leq & \frac{3}{16\sqrt{2}} \sqrt{\frac{\ln(2/\delta)}{\ln(2/\delta \log_2 n)}} + \frac{3}{16\sqrt{2c_3 \ln n}} \frac{\ln(2/\delta)}{\sqrt{\ln(2/\delta \log_2 n)}} \\ & + \frac{3\sqrt{\ln(2/\delta)}}{2\sqrt{2c_3 \ln n}} + \frac{3 \ln(2/\delta)}{2\sqrt{c_3 \ln n}}. \end{aligned}$$

Remark 3.21. For $0 < \delta < 1/2$,

$$\alpha(\delta, n) = o(1) \quad \text{as } n \rightarrow \infty.$$

Remark 3.22. If we assume that the bias b is ρ -regularly varying (or equivalently that the von Mises function η is regularly varying), then elaborating on Proposition 1 from (Drees & Kaufmann, 1998), the oracle index sequence $(k_n^*)_n$ and the sequence $(\widetilde{k}_n(1))_n$ are connected by

$$\lim_n \frac{\widetilde{k}_n(1)}{k_n^*} = (2|\rho|)^{1/(1+2|\rho|)}$$

and their quadratic risk are related by

$$\lim_n \frac{\mathbb{E}[(\gamma - \widehat{\gamma}_H(\widetilde{k}_n(1)))^2]}{\mathbb{E}[(\gamma - \widehat{\gamma}_H(k_n^*))^2]} = \frac{2}{2|\rho| + 1} (2|\rho|)^{2|\rho|/(1+2|\rho|)}.$$

Indeed, under the second order assumption, the two pivotal sequences $(\widetilde{k}_n(1))_n$ and $(\widetilde{k}_n(r_n))_n$ are connected.

Thus if the bias is ρ -regularly varying, Theorem 3.20 provides us with a connection between the performance of the simple selection rule and the performance of the (asymptotically) optimal choice.

Recall that one the main aims of this paper is to derive performance guarantees for the data-driven index selection method \widehat{k}_n without resorting to second order assumptions, this is without assuming that the von Mises function is regularly varying. The next corollary upper bounds the risk of the preliminary estimator when we just have an upper bound on the bias.

Corollary 3.23. *Assume that for some $C > 0$ and $\rho < 0$, for all n, k ,*

$$\left| \gamma - \mathbb{E}\widehat{\gamma}_H(k) \right| \leq C \left(\frac{n}{k} \right)^\rho,$$

then, there exists a constant $\kappa_{\delta, \rho}$ depending on δ and ρ such that, with probability larger than $1 - 4\delta$,

$$\left| \widehat{\gamma}_H(\widehat{k}_n) - \gamma \right| \leq \kappa_{\delta, \rho} \left(\frac{\gamma^2 \ln((2/\delta) \ln n)}{n} \right)^{|\rho|/(1+2|\rho|)} (1 + \alpha(\delta, n))$$

where $\alpha(\delta, n)$ is defined in Theorem 3.20.

Under the assumption that the bias of the Hill estimators is upper bounded by a power function, the performance of the data-driven estimator $\widehat{\gamma}_H(\widehat{k}_n)$ meets the information-theoretic lower bound of Theorem 3.3.

4 Proofs

4.1 Proof of Proposition 3.2

This proposition is a straightforward consequence of Rényi's representation of order statistics of standard exponential samples.

As F belongs to $\text{MDA}(\gamma)$ and meets the von Mises condition, there exists a function η on $(1, \infty)$ with $\lim_{x \rightarrow \infty} \eta(x) = 0$ such that

$$U(x) = cx^\gamma \exp\left(\int_1^x \frac{\eta(s)}{s} ds\right),$$

and

$$U(e^y) = c \exp\left(\int_0^y (\gamma + \eta(e^u)) du\right).$$

Then,

$$\begin{aligned} \widehat{\gamma}_H(k) &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k i \frac{\log U(e^{Y^{(i)}})}{\log U(e^{Y^{(i+1)}})} \\ &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k i \int_{Y^{(i+1)}}^{Y^{(i)}} (\gamma + \eta(e^u)) du \\ &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k i \int_0^{E_i/i} (\gamma + \eta(e^{u+Y^{(i+1)}})) du \\ &\stackrel{d}{=} \frac{1}{k} \sum_{i=1}^k \int_0^{E_i} (\gamma + \eta(e^{\frac{u}{i} + Y^{(i+1)}})) du. \end{aligned}$$

4.2 Proof of Proposition 3.11

Let $Z = k\widehat{\gamma}_H(k)$. By the Pythagorean relation,

$$\text{Var}(Z) = \mathbb{E}[\text{Var}(Z | Y_{(k+1)})] + \text{Var}(\mathbb{E}[Z | Y_{(k+1)}]).$$

Representation (3.2) asserts that, conditionally on $Y_{(k+1)}$, Z is distributed as a sum of independent, exponentially distributed random variables. Let E be an exponentially distributed random variable.

$$\begin{aligned} \text{Var}(Z | Y_{(k+1)} = y) &= k \text{Var}\left(\gamma E + \int_0^E \eta(e^{u+y}) du\right) \\ &= k\gamma^2 + 2k\gamma \text{Cov}\left(E, \int_0^E \eta(e^{u+y}) du\right) + \text{Var}\left(\int_0^E \eta(e^{u+y}) du\right) \\ &\leq k\gamma^2 + 2k\gamma\bar{\eta}(e^y) + k(\bar{\eta}(e^y))^2, \end{aligned}$$

where we have used the Cauchy-Schwarz inequality and $\text{Var}(\int_0^E \eta(e^{y+u}) du) \leq \bar{\eta}(e^y)^2$. Taking expectation with respect to $Y_{(k+1)}$ leads to

$$\mathbb{E}[\text{Var}(Z | Y_{(k+1)})] \leq k\gamma^2 + 2k\gamma\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})] + k\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})^2].$$

The last term in the Pythagorean decomposition is also handled using elementary arguments.

$$\mathbb{E}[Z \mid Y_{(k+1)}] - \mathbb{E}Z = k \int_0^\infty e^{-u} (\eta(e^{u+Y_{(k+1)}}) - \mathbb{E}[\eta(e^{u+Y_{(k+1)}})]) du.$$

As $Y_{(k+1)}$ is a function of independent exponential random variables ($Y_{(k+1)} = \sum_{i=k+1}^n E_i/i$), the variance of $\mathbb{E}[Z \mid Y_{(k+1)}]$ may be upper bounded using Poincaré inequality (Proposition 3.4)

$$\text{Var}(\mathbb{E}[Z \mid Y_{(k+1)}]) \leq 4k\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})^2].$$

In order to derive the lower bound, we first observe that

$$\text{Var}(Z) \geq \mathbb{E}[\text{Var}(Z \mid Y_{(k+1)})].$$

Now, using Cauchy-Schwarz inequality again

$$\begin{aligned} \text{Var}(Z \mid Y_{(k+1)} = y) &\geq k\gamma^2 - 2k\gamma \left| \text{Cov}\left(E, \int_0^E \eta(e^{u+y}) du\right) \right| \\ &\geq k\gamma^2 - 2k\gamma \left(\text{Var}\left(\int_0^E \eta(e^{u+y}) du\right) \right)^{1/2} \\ &\geq k\gamma^2 - 2k\gamma\bar{\eta}(e^y). \end{aligned}$$

4.3 Proof of Theorem 3.13

In the proofs of Theorem 3.13, we will use the next maximal inequality (See Boucheron et al., 2013, Corollary 2.6).

Proposition 3.24. *Let Z_1, \dots, Z_N be real-valued random variables belonging to $\Gamma_+(v, c)$. Then*

$$\mathbb{E} \left[\max_{i=1, \dots, N} Z_i \right] \leq \sqrt{2v \log N} + c \log N.$$

Proofs follow a common pattern. In order to check that some random variable is sub-gamma, we rely on its representation as a function of independent exponential variables and compute partial derivatives, derive convenient upper bounds on the squared Euclidean norm and the supremum norm of the gradient and then invoke Theorem 3.9.

At some point, we will use the next corollary of Theorem 3.9.

Corollary 3.25. *If g is an almost everywhere differentiable function on \mathbb{R} with uniformly bounded derivative g' , then $g(Y_{(k+1)})$ is sub-gamma with variance factor $4\|g'\|_\infty^2/k$ and scale factor $\|g'\|_\infty/k$.*

Proof of Theorem 3.13. We start from the exponential representation of Hill estimators (Proposition 3.2) and represent all $\hat{\gamma}_H(i)$ as functions of independent random variables $E_1, \dots, E_k, Y_{(k+1)}$ where the $E_j, 1 \leq j \leq k$ are standard exponentially distributed and $Y_{(k+1)}$ is distributed like the $(k+1)$ th largest order statistic of an n -sample of the standard exponential distribution.

$$\begin{aligned}
i\widehat{\gamma}_H(i) &= \sum_{j=1}^i \int_0^{E_j} \left(\gamma + \eta(e^{\frac{u}{j} + Y_{(j+1)}}) \right) du \\
&= \sum_{j=1}^i \int_0^{E_j} \left(\gamma + \eta(e^{\frac{u}{j} + \sum_{m=j+1}^k \frac{E_m}{m} + Y_{(k+1)}}) \right) du.
\end{aligned}$$

Let i' be such that $0 \leq i' < i$, agree on $\widehat{\gamma}_H(0) = 0$. Then,

$$i\widehat{\gamma}_H(i) - i'\widehat{\gamma}_H(i') = \sum_{j=i'+1}^i \int_0^{E_j} \left(\gamma + \eta(e^{\frac{u}{j} + \sum_{m=j+1}^k \frac{E_m}{m} + Y_{(k+1)}}) \right) du.$$

Letting

$$g(x_{i'+1}, \dots, x_k) = \sum_{j=i'+1}^i \int_0^{x_j} \left(\gamma + \eta(e^{\frac{u}{j} + \sum_{m=j+1}^k \frac{x_m}{m} + Y_{(k+1)}}) \right) du,$$

a few lines of computations lead to

$$\frac{\partial g}{\partial x_j} = \begin{cases} 0 & \text{for } j \leq i' \\ \gamma + \frac{1}{j} \sum_{m=i'+1}^j \eta(e^{Y_{(m)}}) & \text{for } i' < j \leq i \\ \frac{1}{j} \sum_{m=i'+1}^i \eta(e^{Y_{(m)}}) & \text{for } i < j \leq k \end{cases}$$

which entails that

$$\left| \frac{\partial g}{\partial x_j} \right| \leq \begin{cases} \gamma + \bar{\eta}(e^{Y_{(j+1)}}) & \text{for } i' < j \leq i \\ \frac{i-i'}{j} \bar{\eta}(e^{Y_{(j+1)}}) & \text{for } i < j \leq k. \end{cases}$$

Recalling that $T = \exp(Y_{(j+1)})$, this can be summarised by

$$\left| \frac{\partial g}{\partial x_j} \right| \leq \gamma + \bar{\eta}(T)$$

and

$$\sum_{j=1}^k \left| \frac{\partial g}{\partial x_j} \right|^2 \leq (i - i') (\gamma + 2\bar{\eta}(T))^2.$$

Theorem 3.9 now allows us to establish that, conditionally on T , the centered version of $i\widehat{\gamma}_H(i) - i'\widehat{\gamma}_H(i')$ is sub-gamma on both tails with variance factor $4|i - i'|(\gamma + 2\bar{\eta}(T))^2$ and scale factor $(\gamma + \bar{\eta}(T))$. Using Theorem 3.9 conditionally on T , and choosing $i' = 0$,

$$\mathbb{P} \left\{ |\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) | T]| \geq \frac{\gamma + 2\bar{\eta}(T)}{\sqrt{i}} \left(\sqrt{8s} + \frac{s}{\sqrt{i}} \right) | T \right\} \leq 2e^{-s}.$$

Taking expectation on both sides, this implies that

$$\mathbb{P} \left\{ |\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) | T]| \geq \frac{\gamma + 2\bar{\eta}(T)}{\sqrt{i}} \left(\sqrt{8s} + \frac{s}{\sqrt{i}} \right) \right\} \leq 2e^{-s}.$$

The proof of the upper bound on $\mathbb{E}[Z \mid T]$ in Statement ii) from Theorem 3.13 relies on standard chaining techniques from the theory of empirical processes and uses repeatedly the concentration Theorem 3.9 for smooth functions of independent exponential random variables and the maximal inequality for sub-gamma random variables (Proposition 3.24).

Recall that

$$Z^a = \max_{\ell \leq i \leq k} \sqrt{i} \left| \widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) \mid Y_{(k+1)}] \right|.$$

As it is commonplace in the analysis of normalized empirical processes (See Giné & Koltchinskii, 2006; Massart, 2006; van de Geer, 2000, and references therein), we peel the index set over which the maximum is computed.

Let $\mathcal{L}_n = \{\lfloor \log_2(\ell) \rfloor, \dots, \lfloor \log_2(k) \rfloor\}$. For all $j \in \mathcal{L}_n$, let $\mathcal{S}_j = \{\ell \vee 2^j, \dots, k \wedge 2^{j+1} - 1\}$ and define Z_j^a as

$$Z_j^a = \max_{i \in \mathcal{S}_j} \sqrt{i} \left| \widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) \mid Y_{(k+1)}] \right|.$$

Then,

$$\begin{aligned} \mathbb{E}[Z^a \mid Y_{(k+1)}] &= \mathbb{E}[\max_{j \in \mathcal{L}_n} Z_j^a \mid Y_{(k+1)}] \\ &\leq \mathbb{E}[\max_{j \in \mathcal{L}_n} (Z_j^a - \mathbb{E}[Z_j^a \mid Y_{(k+1)}]) \mid Y_{(k+1)}] + \max_{j \in \mathcal{L}_n} \mathbb{E}[Z_j^a \mid Y_{(k+1)}]. \end{aligned}$$

We now derive upper bounds on both summands by resorting to the maximum inequality for sub-gamma random variables (Proposition 3.24). We first bound $\mathbb{E}[Z_j^a \mid Y_{(k+1)}]$ for $j \in \mathcal{L}_n$.

Fix $j \in \mathcal{L}_n$,

$$\max_{i \in \mathcal{S}_j} \sqrt{i} \left| \widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) \mid Y_{(k+1)}] \right| \leq \frac{1}{2^{j/2}} \max_{i \in \mathcal{S}_j} i \left| \widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) \mid Y_{(k+1)}] \right|.$$

In order to alleviate notation, let $W(i) = i (\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) \mid Y_{(k+1)}])$ for $i \in \mathcal{S}_j$. For $i \in \mathcal{S}_j$, let

$$i = 2^j + \sum_{m=1}^j b_m 2^{j-m} \quad \text{where } b_m \in \{0, 1\}$$

be the binary expansion of i . Then, for $h \in \{0, \dots, j\}$, let $\pi_h(i)$ be defined by

$$\pi_h(i) = 2^j + \sum_{m=1}^h b_m 2^{j-m}$$

so that $\pi_j(i) = i$, $\pi_0(i) = 2^j$ and $0 \leq \pi_{h+1}(i) - \pi_h(i) \leq 2^{j-h-1}$.

Using that $W(\pi_0(i))$ does not depend on i and that $\mathbb{E}[W(\pi_0(i)) | Y_{(k+1)}] = 0$,

$$\begin{aligned}
& \mathbb{E} \left[\max_{i \in \mathcal{S}_j} i (\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]) \mid Y_{(k+1)} \right] \\
&= \mathbb{E} \left[\max_{i \in \mathcal{S}_j} W(i) \mid Y_{(k+1)} \right] \\
&= \mathbb{E} \left[\max_{i \in \mathcal{S}_j} W(\pi_j(i)) - W(\pi_0(i)) \mid Y_{(k+1)} \right] \\
&= \mathbb{E} \left[\max_{i \in \mathcal{S}_j} \sum_{h=0}^{j-1} (W(\pi_{h+1}(i)) - W(\pi_h(i))) \mid Y_{(k+1)} \right] \\
&\leq \sum_{h=0}^{j-1} \mathbb{E} \left[\max_{i \in \mathcal{S}_j} (W(\pi_{h+1}(i)) - W(\pi_h(i))) \mid Y_{(k+1)} \right].
\end{aligned}$$

Now for each $h \in \{0, \dots, j-1\}$, the maximum is taken over 2^h random variables which are sub-gamma with variance factor $4 \times 2^{j-h-1}(\gamma + 2\bar{\eta}(T))^2$ and scale factor $(\gamma + \bar{\eta}(T))$. By Proposition 3.24, since $i \in \mathcal{S}_j$,

$$\begin{aligned}
& \mathbb{E} \left[\max_{i \in \mathcal{S}_j} i (\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]) \mid Y_{(k+1)} \right] \\
&\leq \sum_{h=0}^{j-1} \left((\gamma + 2\bar{\eta}(T)) \sqrt{8h2^{j-h-1} \ln 2} + (\gamma + \bar{\eta}(T)) h \ln 2 \right) \\
&\leq 8(\gamma + 2\bar{\eta}(T))2^{j/2}.
\end{aligned}$$

So that for all $j \in \mathcal{L}_n$,

$$\mathbb{E} \left[\max_{i \in \mathcal{S}_j} \sqrt{i} (\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]) \mid Y_{(k+1)} \right] \leq 8(\gamma + 2\bar{\eta}(T))$$

and

$$\mathbb{E}[Z_j^a | Y_{(k+1)}] \leq 16(\gamma + 2\bar{\eta}(T)).$$

In order to prove Statement ii), we check that for each $j \in \mathcal{L}_n$, Z_j is sub-gamma on the right-tail with variance factor at most $4(\gamma + 2\bar{\eta}(T))^2$ and scale factor not larger than $(\gamma + 2\bar{\eta}(T)) / \sqrt{\ell}$. Under the von Mises Condition (3.1), the sampling distribution is absolutely continuous with respect to Lebesgue measure. For almost every sample, the maximum defining Z_j is attained at a single index $i \in \mathcal{S}_j$. Starting again from the exponential representation, and repeating the computation of partial derivatives, we obtain the desired bounds.

By Proposition 3.24,

$$\begin{aligned}
\mathbb{E} \left[\max_{j \in \mathcal{L}_n} (Z_j^a - \mathbb{E}[Z_j^a | Y_{(k+1)}]) \mid Y_{(k+1)} \right] &\leq \left(\sqrt{8 \ln |\mathcal{L}_n|} + \frac{\ln |\mathcal{L}_n|}{\sqrt{\ell}} \right) (\gamma + 2\bar{\eta}(T)) \\
&\leq 6(\sqrt{\ln |\mathcal{L}_n|} (\gamma + 2\bar{\eta}(T))).
\end{aligned}$$

Combining the different bounds leads to Inequality (3.7). \square

4.4 Proof of Theorem 3.17

Adopting again the exponential representation, $\mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]$ is an $Y_{(k+1)}$ -measurable random variable, say $g(Y_{(k+1)})$.

As a function g of $Y_{(k+1)} = y$, the conditional expectation $\mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]$ reads as

$$g(y) = \frac{1}{i} \sum_{j=1}^i \int_{[0, \infty)} \dots \int_{[0, \infty)} e^{-\sum_{m=j}^k u_m} \eta\left(e^{y + \sum_{m=j}^k \frac{u_m}{m}}\right) du_j \dots du_k.$$

Its derivative with respect to y is readily computed, and after integration by parts and handling a telescoping sum, it reads as

$$\begin{aligned} g'(y) &= \frac{1}{i} \sum_{j=1}^i \int_{[0, \infty)} \dots \int_{[0, \infty)} e^{-\sum_{m=j}^k u_m} \left(\eta\left(e^{y + \sum_{m=j}^k \frac{u_m}{m}}\right) \right. \\ &\quad \left. - \eta\left(e^{y + \sum_{m=j+1}^k \frac{u_m}{m}}\right) \right) du_j \dots du_k. \end{aligned}$$

A conservative upper-bound on $|g'(y)|$ is $2\bar{\eta}(e^{Y_{(k+1)}})$ which is upper bounded by $2\bar{\eta}(T)$. The statement of the proposition then follows from Proposition 3.25.

A byproduct of the proof is the next variance bound, for $i \leq k$,

$$\text{Var}\left(\mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}] | T\right) \leq \frac{16}{k} \mathbb{E}\left[\bar{\eta}\left(e^{Y_{(k+1)}}\right)^2 | T\right].$$

In the rest of the proof, Δ_i denotes the spacing $Y_{(i)} - Y_{(k+1)}$, \mathbb{E}_{Δ_i} the expectation with respect to Δ_i , $Y'_{(k+1)}$ an independent copy of $Y_{(k+1)}$, and \mathbb{E}' the expectation with respect to $Y'_{(k+1)}$. We will also use the next lemma.

Lemma 3.26. *Let X be a non-negative random variable, and $a, b \in [0, \infty]$, then*

$$\mathbb{E}\left[e^X \left| \int_{e^{X+a}}^{e^{X+b}} \frac{\eta(v)}{v^2} dv \right| \right] \leq \bar{\eta}(e^{a \wedge b}) |e^{-a} - e^{-b}|.$$

Let

$$Z^b = \max_{\ell \leq i \leq k} |\mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}] - \mathbb{E}[\widehat{\gamma}_H(i)]|$$

and recall that $b(t) = t \int_t^\infty \eta(v)/v^2 dv$. Then

$$\begin{aligned} &|\mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}] - \mathbb{E}[\widehat{\gamma}_H(i)]| \\ &= |\mathbb{E}[b(e^{Y_{(i+1)}}) | Y_{(k+1)}] - \mathbb{E}[b(e^{Y_{(i+1)}})]| \\ &= \left| e^{Y_{(k+1)}} \mathbb{E}_{\Delta_i} \left[e^{\Delta_i} \int_{e^{\Delta_i + Y_{(k+1)}}}^\infty \frac{\eta(v)}{v^2} dv \right] \right. \\ &\quad \left. - \mathbb{E}' \left[e^{Y'_{(k+1)}} \mathbb{E}_{\Delta_i} \left[e^{\Delta_i} \int_{e^{\Delta_i + Y'_{(k+1)}}}^\infty \frac{\eta(v)}{v^2} dv \right] \right] \right| \\ &\leq \left| \mathbb{E}' \left[\left(e^{Y_{(k+1)}} - e^{Y'_{(k+1)}} \right) \mathbb{E}_{\Delta_i} \left[e^{\Delta_i} \int_{e^{\Delta_i + Y_{(k+1)}}}^\infty \frac{\eta(v)}{v^2} dv \right] \right] \right| \\ &\quad + \left| \mathbb{E}' \left[e^{Y'_{(k+1)}} \mathbb{E}_{\Delta_i} \left[e^{\Delta_i} \int_{e^{\Delta_i + Y_{(k+1)}}}^{e^{\Delta_i + Y'_{(k+1)}}} \frac{\eta(v)}{v^2} dv \right] \right] \right|. \end{aligned}$$

The expectation of Z is thus upper bounded by the following sum

$$\begin{aligned} \mathbb{E}Z^b &\leq \underbrace{\mathbb{E} \left[\max_{\ell \leq i \leq k} \left| \mathbb{E}' \left[\left(e^{Y_{(k+1)}} - e^{Y'_{(k+1)}} \right) \mathbb{E}_{\Delta_i} \left[e^{\Delta_i} \int_{e^{\Delta_i + Y_{(k+1)}}}^{\infty} \frac{\eta(v)}{v^2} dv \right] \right] \right| \right]}_{:=\text{(I)}} \\ &\quad + \underbrace{\mathbb{E} \left[\max_{\ell \leq i \leq k} \left| \mathbb{E}' \left[e^{Y'_{(k+1)}} \mathbb{E}_{\Delta_i} \left[e^{\Delta_i} \int_{e^{\Delta_i + Y_{(k+1)}}}^{e^{\Delta_i + Y'_{(k+1)}}} \frac{\eta(v)}{v^2} dv \right] \right] \right| \right]}_{:=\text{(II)}}. \end{aligned}$$

In the sequel, we use twice the next upper bound which holds for $k \geq 5$

$$\mathbb{E} \left[\left(1 - e^{Y'_{(k+1)} - Y_{(k+1)}} \right)^2 \right] \leq \frac{4}{k} \frac{1}{1 - 4/k}. \quad (3.10)$$

Note that the right-hand side is less than $5/k$ for $k \geq 30$.

By Lemma 3.26,

$$\begin{aligned} \text{(I)} &\leq \mathbb{E} \left[\max_{\ell \leq i \leq k} \mathbb{E}' \left[\left| e^{Y_{(k+1)}} - e^{Y'_{(k+1)}} \right| \mathbb{E}_{\Delta_i} \left[\left| e^{\Delta_i} \int_{e^{\Delta_i + Y_{(k+1)}}}^{\infty} \frac{\eta(v)}{v^2} dv \right| \right] \right] \right] \\ &\leq \mathbb{E} \left[\max_{\ell \leq i \leq k} \mathbb{E}' \left[\left| e^{Y_{(k+1)}} - e^{Y'_{(k+1)}} \right| \times \bar{\eta} \left(e^{Y_{(k+1)}} \right) e^{-Y_{(k+1)}} \right] \right] \\ &= \mathbb{E} \left[\mathbb{E}' \left[\left| e^{Y_{(k+1)}} - e^{Y'_{(k+1)}} \right| \times \bar{\eta} \left(e^{Y_{(k+1)}} \right) e^{-Y_{(k+1)}} \right] \right] \\ &= \mathbb{E} \left[\bar{\eta} \left(e^{Y_{(k+1)}} \right) \left| 1 - e^{Y'_{(k+1)} - Y_{(k+1)}} \right| \right] \\ &\leq \sqrt{\mathbb{E} \left[\bar{\eta} \left(e^{Y_{(k+1)}} \right)^2 \right]} \sqrt{\mathbb{E} \left[\left| 1 - e^{Y'_{(k+1)} - Y_{(k+1)}} \right|^2 \right]} \end{aligned}$$

and

$$\begin{aligned} \text{(II)} &\leq \mathbb{E} \left[\max_{\ell \leq i \leq k} \mathbb{E}' \left[\left| e^{Y'_{(k+1)}} \mathbb{E}_{\Delta_i} \left[\left| e^{\Delta_i} \int_{e^{\Delta_i + Y_{(k+1)}}}^{e^{\Delta_i + Y'_{(k+1)}}} \frac{\eta(v)}{v^2} dv \right| \right] \right| \right] \right] \\ &\leq \mathbb{E} \left[\max_{\ell \leq i \leq k} \mathbb{E}' \left[\left| e^{Y'_{(k+1)}} \bar{\eta} \left(e^{Y_{(k+1)} \wedge Y'_{(k+1)}} \right) \left| e^{-Y_{(k+1)}} - e^{-Y'_{(k+1)}} \right| \right] \right] \right] \\ &= \mathbb{E} \left[\left| e^{Y'_{(k+1)}} \bar{\eta} \left(e^{Y_{(k+1)} \wedge Y'_{(k+1)}} \right) \left| e^{-Y_{(k+1)}} - e^{-Y'_{(k+1)}} \right| \right] \right] \\ &= \mathbb{E} \left[\bar{\eta} \left(e^{Y_{(k+1)} \wedge Y'_{(k+1)}} \right) \left| e^{Y'_{(k+1)} - Y_{(k+1)}} - 1 \right| \right] \\ &\leq \sqrt{\mathbb{E} \left[\bar{\eta} \left(e^{Y_{(k+1)}} \right)^2 \right]} \sqrt{\mathbb{E} \left[\left| 1 - e^{Y'_{(k+1)} - Y_{(k+1)}} \right|^2 \right]}. \end{aligned}$$

Finally, thanks to Inequality (3.10),

$$\mathbb{E}Z^b \leq \frac{2C \sqrt{\mathbb{E} \left[\bar{\eta} \left(e^{Y_{(k+1)} \wedge Y'_{(k+1)}} \right)^2 \right]}}{\sqrt{k}}$$

where the constant C can be chosen not larger than 5 as we assume that $k \geq 30$.

4.5 Proof of Corollary 3.19

We first check that, with probability larger than $1 - \delta$, $Y_{(k+1)} \geq n/k'$ where $k' = k + \ln(\delta)/2$. Starting from the proof of Proposition 4.3 from (Boucheron & Thomas, 2012) and recalling that $\operatorname{arsinh}(x) = \ln(x + \sqrt{1 + x^2})$, a few lines of computation yields that, with probability larger than $1 - \delta$, for $\ln 2 \leq z \leq \ln\left(\frac{n}{k}\right)$,

$$\mathbb{P}\left\{\exp(Y_{(k+1)}) \geq \frac{n}{k}e^{-z}\right\} \geq 1 - \exp\left(-\frac{k(e^z - 1)^2}{2e^z}\right).$$

Solving

$$\delta = \exp\left(-\frac{k(e^z - 1)^2}{2e^z}\right)$$

leads to

$$z = 2 \operatorname{arsinh}\left(\sqrt{\frac{\ln(1/\delta)}{2k}}\right) = 2 \ln\left(\sqrt{\frac{\ln(1/\delta)}{2k}} + \sqrt{1 + \frac{\ln(1/\delta)}{2k}}\right),$$

so that

$$e^z \leq 1 + \frac{\ln(1/\delta)}{2k}.$$

Now,

$$\begin{aligned} & \max_{\ell \leq i \leq k} \sqrt{i} |\widehat{\gamma}_H(i) - \mathbb{E}\widehat{\gamma}_H(i)| \\ & \leq \max_{\ell \leq i \leq k} \sqrt{i} |\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]| + \max_{\ell \leq i \leq k} \sqrt{i} |\mathbb{E}[\widehat{\gamma}_H(i)] - \mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]|. \end{aligned}$$

By Theorem 3.13, with probability $1 - 2\delta$,

$$\max_{\ell \leq i \leq k} \sqrt{i} |\widehat{\gamma}_H(i) - \mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]| \leq (\gamma + 2\bar{\eta}(n/k')) \left(\xi_n + \sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{\ell}} \right).$$

Then, by Theorem 3.17, with probability larger than $1 - \delta$,

$$\max_{\ell \leq i \leq k} \sqrt{i} |\mathbb{E}[\widehat{\gamma}_H(i)] - \mathbb{E}[\widehat{\gamma}_H(i) | Y_{(k+1)}]| \leq 4\sqrt{\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})^2]} + 2\bar{\eta}(1) \left(2\sqrt{2 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{k}} \right).$$

Combining all tail bounds leads to

$$\begin{aligned} & \max_{\ell \leq i \leq k} \sqrt{i} |\widehat{\gamma}_H(i) - \mathbb{E}\widehat{\gamma}_H(i)| \\ & \leq (\gamma + 2\bar{\eta}(n/k')) \left(\xi_n + \sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{\ell}} \right) \\ & \quad + 4\sqrt{\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})^2]} + 2\bar{\eta}(1) \left(2\sqrt{2 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{k}} \right) \\ & \leq \xi_n(\gamma + 2\bar{\eta}(n/k')) + 4\sqrt{\mathbb{E}[\bar{\eta}(e^{Y_{(k+1)}})^2]} + 2(\gamma + 2\bar{\eta}(n/k') + 2\bar{\eta}(1))\sqrt{2 \ln(2/\delta)} \\ & \quad + \left(\frac{(\gamma + 2\bar{\eta}(n/k'))}{\sqrt{\ell}} + \frac{\bar{\eta}(1)}{\sqrt{k}} \right) \ln(2/\delta) \end{aligned}$$

with probability larger than $1 - 3\delta$.

4.6 Proof of Theorem 3.20

Throughout this proof, let

$$\begin{aligned} T_n &= \exp\left(Y_{(\tilde{k}_n+1)}\right) \\ \xi_n &= c_1 \sqrt{\ln \log_2 n} + c'_1 \quad \text{with } c_1 \leq 4 \text{ and } c'_1 \leq 16, \\ z_\delta &= \xi_n + \sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{c_3 \ln n}} \\ y_\delta &= 5 + \sqrt{2 \ln(8/\delta)} + \frac{\ln(2/\delta)}{\sqrt{c_3 \ln n}}. \end{aligned}$$

Let us consider the event $E_1 \cap E_2 \cap E_3$ as

$$\begin{aligned} E_1 &= \left\{ c_3 \ln n \leq i \leq \tilde{k}_n, \sqrt{i} |\hat{\gamma}_H(i) - \mathbb{E}[\hat{\gamma}_H(i) | T_n]| \leq (\gamma + 2\bar{\eta}(T_n)) z_\delta \right\}, \\ E_2 &= \left\{ c_3 \ln n \leq i \leq k, \sqrt{k} |\mathbb{E}[\hat{\gamma}_H(i) - \hat{\gamma}_H(k) | T_n] - \mathbb{E}[\hat{\gamma}_H(i) - \hat{\gamma}_H(k)]| \leq 8\bar{\eta}(1) y_\delta \right\}, \\ E_3 &= \left\{ T_n \geq n / \left(\tilde{k}_n + \ln(1/\delta)/2 \right) \right\}. \end{aligned}$$

$E_1 \cap E_2 \cap E_3$ has probability at least $1 - 3\delta$. Indeed, by Theorem 3.13, $\mathbb{P}(E_1) \geq 1 - \delta$, by Theorem 3.17, $\mathbb{P}(E_2) \geq 1 - \delta$ and thanks to the beginning of the proof of Corollary 3.19, $\mathbb{P}(E_3) \geq 1 - \delta$.

Clearly, $E_1 = E_1(\delta, n)$, $E_2 = E_2(\delta, n)$, and $E_3 = E_3(\delta, n)$. However, since δ and n are fixed, we simply denote them E_1 , E_2 and E_3 to alleviate the notations.

We check that under $E_1 \cap E_2 \cap E_3$, the selected index is not smaller than \tilde{k}_n . This amounts to check that for all $k \leq \tilde{k}_n - 1$, and for all $i \in \{[c_3 \ln n], \dots, k\}$,

$$\sqrt{i} |\hat{\gamma}_H(i) - \hat{\gamma}_H(k)| \leq r_n(\delta) \hat{\gamma}_H(i).$$

Now, under $E_1 \cap E_3$, $\bar{\eta}(T_n) \leq \gamma/4$ and conditionally on T_n , the process $(|\hat{\gamma}_H(i) - \mathbb{E}[\hat{\gamma}_H(i) | T_n]|)_i$ can be controlled : for all $i \in \{[c_3 \ln n], \dots, k\}$

$$|\hat{\gamma}_H(i) - \mathbb{E}[\hat{\gamma}_H(i) | T_n]| \leq \frac{(\gamma + 2\bar{\eta}(T_n))}{\sqrt{i}} z_\delta \leq \frac{3\gamma}{2\sqrt{i}} z_\delta. \quad (3.11)$$

For all $i \in \{[c_3 \ln n], \dots, \tilde{k}_n\}$,

$$\begin{aligned} \gamma - \hat{\gamma}_H(i) &\leq |\gamma - \mathbb{E}[\hat{\gamma}_H(i) | T_n]| + |\hat{\gamma}_H(i) - \mathbb{E}[\hat{\gamma}_H(i) | T_n]| \\ &\leq \bar{\eta}(T_n) + \frac{3\gamma}{2\sqrt{i}} z_\delta \\ &\leq \frac{\gamma}{4} + \frac{3\gamma}{2\sqrt{i}} z_\delta \end{aligned}$$

so that

$$\frac{\hat{\gamma}_H(i)}{\gamma} \geq \frac{3}{4} \left(1 - \frac{2z_\delta}{\sqrt{i}} \right).$$

Meanwhile, for all $k \leq \tilde{k}_n - 1$ and for all $i \in \{[c_3 \ln n], \dots, k\}$,

$$\begin{aligned} &|\hat{\gamma}_H(i) - \hat{\gamma}_H(k)| \\ &\leq \underbrace{|\hat{\gamma}_H(i) - \mathbb{E}[\hat{\gamma}_H(i) | T_n]|}_{(i)} + \underbrace{|\mathbb{E}[\hat{\gamma}_H(i) - \hat{\gamma}_H(k) | T_n]|}_{(ii)} + \underbrace{|\hat{\gamma}_H(k) - \mathbb{E}[\hat{\gamma}_H(k) | T_n]|}_{(iii)}. \end{aligned}$$

Using again (3.11), under $E_1 \cap E_3$,

$$(I) + (III) \leq \frac{3\gamma}{2} z_\delta \left(\frac{1}{\sqrt{i}} + \frac{1}{\sqrt{k}} \right) \leq \frac{3\gamma}{\sqrt{i}} z_\delta.$$

Now, under E_2 , thanks to Assumption i) in the theorem statement,

$$\begin{aligned} (II) &\leq |\mathbb{E}[\widehat{\gamma}_H(i) - \widehat{\gamma}_H(k) \mid T_n] - \mathbb{E}[\widehat{\gamma}_H(i) - \widehat{\gamma}_H(k)]| + |\mathbb{E}[\widehat{\gamma}_H(i) - \gamma]| + |\mathbb{E}[\gamma - \widehat{\gamma}_H(k)]| \\ &\leq \frac{2\bar{\eta}(1)}{\sqrt{k}} y_\delta + \gamma r_n \left(\frac{1}{\sqrt{i}} + \frac{1}{\sqrt{k}} \right) \\ &\leq \frac{2\xi_n \gamma}{\sqrt{k}} y_\delta + \frac{2r_n \gamma}{\sqrt{i}}. \end{aligned}$$

Plugging upper bounds on (I), (II) and (III), it comes that under E_δ , for all $k \leq \tilde{k}_n - 1$ and for all $i \in \{\lceil c_3 \ln n \rceil, \dots, k\}$,

$$\begin{aligned} \sqrt{i} \frac{|\widehat{\gamma}_H(i) - \widehat{\gamma}_H(k)|}{\gamma} &\leq 3z_\delta + 2\xi_n y_\delta + 2r_n \\ &\leq 3 \left(\xi_n + \sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{c_3 \ln n}} \right) + 2\xi_n y_\delta + 2r_n \\ &\leq \xi_n (3 + 2y_\delta) + 2r_n + 3\sqrt{8 \ln(2/\delta)} + \frac{3 \ln(2/\delta)}{\sqrt{c_3 \ln n}}. \end{aligned}$$

In order to warrant that under $E_1 \cap E_2 \cap E_3$, for all $k \leq \tilde{k}_n - 1$ and for all i such that $c_3 \ln n \leq i \leq k$, $\sqrt{i} |\widehat{\gamma}_H(i) - \widehat{\gamma}_H(k)| \leq r_n(\delta) \widehat{\gamma}_H(i)$, it is enough to have

$$r_n(\delta) \left(1 - \frac{2z_\delta}{\sqrt{c_3 \ln n}} \right) \leq \frac{4}{3} \left(\xi_n (3 + 8y_\delta) + 2r_n + 3\sqrt{8 \ln(2/\delta)} + \frac{3 \ln(2/\delta)}{\sqrt{c_3 \ln n}} \right)$$

which holds by definition of $r_n(\delta)$.

We now check that the risk of $\widehat{\gamma}_H(\widehat{k}_n)$ is not much larger than the risk of $\widehat{\gamma}_H(\tilde{k}_n)$.

$$\begin{aligned} \left| \gamma - \widehat{\gamma}_H(\widehat{k}_n) \right| &\leq \left| \gamma - \widehat{\gamma}_H(\tilde{k}_n) \right| + \left| \widehat{\gamma}_H(\tilde{k}_n) - \widehat{\gamma}_H(\widehat{k}_n) \right| \\ &\leq \left| \gamma - \widehat{\gamma}_H(\tilde{k}_n) \right| + \frac{r_n(\delta) \widehat{\gamma}_H(\tilde{k}_n)}{\sqrt{\widehat{k}_n}} \\ &\leq \left| \gamma - \widehat{\gamma}_H(\tilde{k}_n) \right| \left(1 + \frac{r_n(\delta)}{\sqrt{\widehat{k}_n}} \right) + \frac{r_n(\delta) \gamma}{\sqrt{\widehat{k}_n}}. \end{aligned}$$

Therefore, with probability larger than $1 - 3\delta$,

$$\left| \gamma - \widehat{\gamma}_H(\widehat{k}_n) \right| \leq \left| \gamma - \widehat{\gamma}_H(\tilde{k}_n) \right| \left(1 + \frac{r_n(\delta)}{\sqrt{\widehat{k}_n}} \right) + \frac{r_n(\delta) \gamma}{\sqrt{\widehat{k}_n}}. \quad (3.12)$$

Now, consider the event $E_1 \cap E_2 \cap E_3 \cap E_4$ with E_1 , E_2 and, E_3 defined as in the beginning of the proof and

$$E_4 = \left\{ \sqrt{\widehat{k}_n} \left| \widehat{\gamma}_H(\widehat{k}_n) - \mathbb{E}[\widehat{\gamma}_H(\widehat{k}_n) \mid T_n] \right| \leq (\gamma + 2\bar{\eta}(T_n)) \left(\sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{\widehat{k}_n}} \right) \right\}.$$

Since, $\mathbb{P}(E_4) \geq 1 - \delta$ thanks to Statement i) from Theorem 3.13, the event $E_1 \cap E_2 \cap E_3 \cap E_4$ has probability at least $1 - 4\delta$.

Then, by definition of \tilde{k}_n , $|\gamma - \mathbb{E}\hat{\gamma}_H(\tilde{k}_n)| \leq \gamma r_n / \sqrt{\tilde{k}_n}$, under $E_3 \cap E_4$,

$$\begin{aligned} \left| \hat{\gamma}_H(\tilde{k}_n) - \gamma \right| &\leq |\gamma - \mathbb{E}\hat{\gamma}_H(\tilde{k}_n)| + |\hat{\gamma}_H(\tilde{k}_n) - \mathbb{E}\hat{\gamma}_H(\tilde{k}_n)| \\ &\leq \frac{1}{\sqrt{\tilde{k}_n}} \left(\gamma r_n + (\gamma + 2\bar{\eta}(T_n)) \left(\sqrt{8 \ln(2/\delta)} + \frac{\ln(2/\delta)}{\sqrt{\tilde{k}_n}} \right) \right) \\ &\leq \frac{\gamma}{\sqrt{\tilde{k}_n}} \left(r_n + \frac{3}{2} \sqrt{8 \ln(2/\delta)} + \frac{3 \ln(2/\delta)}{2\sqrt{\tilde{k}_n}} \right). \end{aligned}$$

Therefore, plugging this bound into (3.12), with probability larger than $1 - 4\delta$,

$$\left| \hat{\gamma}_H(\hat{k}_n) - \gamma \right| \leq \frac{\gamma}{\sqrt{\tilde{k}_n}} \left(\left(r_n + \frac{3}{2} \sqrt{8 \ln(2/\delta)} + \frac{3 \ln(2/\delta)}{2\sqrt{\tilde{k}_n}} \right) \left(1 + \frac{r_n(\delta)}{\sqrt{\tilde{k}_n}} \right) + r_n(\delta) \right).$$

4.7 Proof of Corollary 3.23

If, for some $C > 0$ and $\rho < 0$,

$$\left| \gamma - \mathbb{E}\hat{\gamma}_H(k) \right| \leq C \left(\frac{n}{k} \right)^\rho,$$

then

$$\frac{\gamma r_n}{\sqrt{\tilde{k}_n + 1}} \leq C \left(\frac{n}{\tilde{k}_n + 1} \right)^\rho,$$

that is,

$$\sqrt{\tilde{k}_n + 1} \geq \left(\frac{\gamma r_n}{C} \right)^{1/(1+2|\rho|)} n^{|\rho|/(1+2|\rho|)}.$$

Thus, there exists a constant c such that

$$\sqrt{\tilde{k}_n} \geq \left(\frac{\gamma r_n}{c} \right)^{1/(1+2|\rho|)} n^{|\rho|/(1+2|\rho|)}.$$

Starting from Equation (3.9) of Theorem 3.20, with probability $1 - 4\delta$,

$$\left| \hat{\gamma}_H(\hat{k}_n) - \gamma \right| \leq 16\gamma \sqrt{\frac{2 \ln(2/\delta \ln n)}{\tilde{k}_n}} (1 + \alpha(\delta, n)),$$

and, there exists a constant $\kappa_{\delta, \rho}$, depending on δ and ρ , such that

$$\sqrt{\frac{\ln(2/\delta \ln n)}{\tilde{k}_n}} \leq \kappa_{\delta, \rho} \gamma^{-1/(1+2|\rho|)} \left(\frac{\ln(2/\delta \ln n)}{n} \right)^{|\rho|/(1+2|\rho|)}.$$

Hence, with probability larger than $1 - 4\delta$,

$$\left| \hat{\gamma}_H(\hat{k}_n) - \gamma \right| \leq \kappa_{\delta, \rho} \left(\frac{\gamma^2 \ln(2/\delta \ln n)}{n} \right)^{|\rho|/(1+2|\rho|)} (1 + \alpha(\delta, n)).$$

5 Simulations

Risk bounds like Theorem 3.20 and Corollary 3.23 are conservative. For all practical purposes, they are just meant to be reassuring guidelines. In this numerical section, we intend to shed some light on the following issues :

1. Is there a reasonable way to calibrate the threshold $r_n(\delta)$ used in the definition of \widehat{k}_n ? How does the method perform if we choose $r_n(\delta)$ close to $\sqrt{2 \ln \ln(n)}$?
2. How large is the ratio between the risk of $\widehat{\gamma}_H(\widehat{k}_n)$ and the risk of $\widehat{\gamma}_H(k_n^*)$ for moderate sample sizes?

The finite-sample performance of the data-driven index selection method described and analysed in Section 3.3 has been assessed by Monte-Carlo simulations. Computations have been carried out in R using packages `ggplot2` (Wickham, 2009), `knitr`, `foreach`, `iterators`, `xtable` and `dplyr` (See Wickham, 2014, for a modern account of the R environment). To get into the details, we investigated the performance of index selection methods on samples of sizes 10000, 20000 and 100000 from the collection of distributions listed in Table 3.1. The list comprises the following distributions.

- i) Fréchet distributions $F_\gamma(x) = \exp(x^{-1/\gamma})$ for $x > 0$ and $\gamma \in \{1, 0.5, 0.2\}$.
- ii) Student distributions t_ν with $\nu \in \{1, 2, 4, 10\}$ degrees of freedom.
- iii) log-gamma distribution with density proportional to $(\ln(x))^{2-1}x^{-3-1}$, which means $\gamma = 1/3$ and $\rho = 0$.
- iv) The Lévy distribution with density $\sqrt{1/(2\pi)} \frac{e^{-\frac{1}{2x}}}{x^{3/2}}$, $\gamma = 2$ and $\rho = -1$ (this is the distribution of $1/X^2$ when $X \sim \mathcal{N}(0, 1)$).
- v) The H distribution is defined by $\gamma = 1/2$ and von Mises function equal to $\eta(s) = (2/s) \ln 1/s$. This distribution satisfies the second-order regular variation condition with $\rho = -1$ but does not fit into Model (3.5).
- vi) Two Pareto change point distributions with distribution functions

$$\overline{F}(x) = x^{-1/\gamma'} \mathbb{1}_{\{1 \leq x \leq \tau\}} + \tau^{-1/\gamma'} (x/\tau)^{-1/\gamma'} \mathbb{1}_{\{x \geq \tau\}}$$

and $\gamma \in \{1.5, 1.25\}$, $\gamma' = 1$, and thresholds τ adjusted in such a way that they respectively correspond to quantiles of order $1 - 1/15$ and $1 - 1/25$.

Fréchet, Student, log-gamma distributions were used as benchmarks by (Drees & Kaufmann, 1998), (Danielsson et al., 2001) and (Carpentier & Kim, 2014a).

Table 3.1, which is complemented by Figure 3.3, describes the difficulty of tail index estimation from samples of the different distributions. Monte-Carlo estimates of the standardised root mean square error (RMSE) of Hill estimators

$$\mathbb{E} [(\widehat{\gamma}_H(k)/\gamma - 1)^2]^{1/2}$$

are represented as functions of the number of order statistics k for samples of size 10000 from the sampling distributions. All curves exhibit a common pattern : for small values of k , the RMSE is dominated by the variance term and scales like $1/\sqrt{k}$. Above a threshold

that depends on the sampling distribution but that is not completely characterised by the second-order regular variation index, the RMSE grows at a rate that may reflect the second-order regular variation property of the distribution. Not too surprisingly, the three Fréchet distributions exhibit the same risk profile. The three curves are almost undistinguishable. The Student distributions illustrate the impact of the second-order parameter on the difficulty of the index selection problem. For sample size $n = 10000$, the optimal index for t_{10} is smaller than 30, it is smaller than the usual recommendations and for such moderate sample sizes seems as hard to handle as the log-gamma distribution which usually fits in the Horror Hill Plot gallery. The 1/2-stable Lévy distribution and the H -distribution behave very differently. Even though they both have second-order parameter ρ equal to -1 , the H distribution seems almost as challenging as the t_4 distribution while the Lévy distribution looks much easier than the Fréchet distributions. The Pareto change point distributions exhibit an abrupt transition.

TABLE 3.1. Estimated oracle index k_n^* and standardised RMSE $\mathbb{E}[(\gamma - \hat{\gamma}_H(k_n^*))^2]^{1/2}/\gamma$ for benchmark distributions. Estimates were computed from 5000 replicated experiments on samples of size 10000.

d.f.	γ	ρ	k_n^*	RMSE
$F_{0.2}$	0.2	1.0	1132	3.7e-02
$F_{0.5}$	0.5	1.0	1145	3.6e-02
F_1	1.0	1.0	1155	3.6e-02
t_1	1.0	2.0	1161	3.3e-02
t_2	0.5	1.0	341	6.5e-02
t_4	0.2	0.5	77	1.6e-01
t_{10}	0.1	0.2	15	5.3e-01
H	0.5	1.0	130	1.1e-01
log-gamma	0.3	0.0	213	1.6e-01
Stable	2.0	1.0	3172	2.0e-02
Pcp	1.5	0.3	943	3.3e-02
Pcp (bis)	1.2	0.2	593	4.2e-02

Index $\hat{k}_n(r_n)$ was computed according to the following rule

$$\hat{k}_n(r_n) = \min \left\{ k : 30 \leq k \leq n \text{ and } \exists i \in \{30, \dots, k\}, |\hat{\gamma}_H(i) - \hat{\gamma}_H(k)| > \frac{r_n \hat{\gamma}_H(i)}{\sqrt{i}} \right\} - 1 \quad (3.13)$$

with $r_n = \sqrt{c \ln \ln n}$ where $c = 2.1$ unless otherwise specified.

The Fréchet, Student, H and stable distributions all fit into the framework considered by (Drees & Kaufmann, 1998). They provide a favorable ground for comparing the performance of the optimal index selection method described by Drees & Kaufmann (1998) which attempts to take advantage of the second-order regular variation property and the performance of the simple selection rule described in this paper.

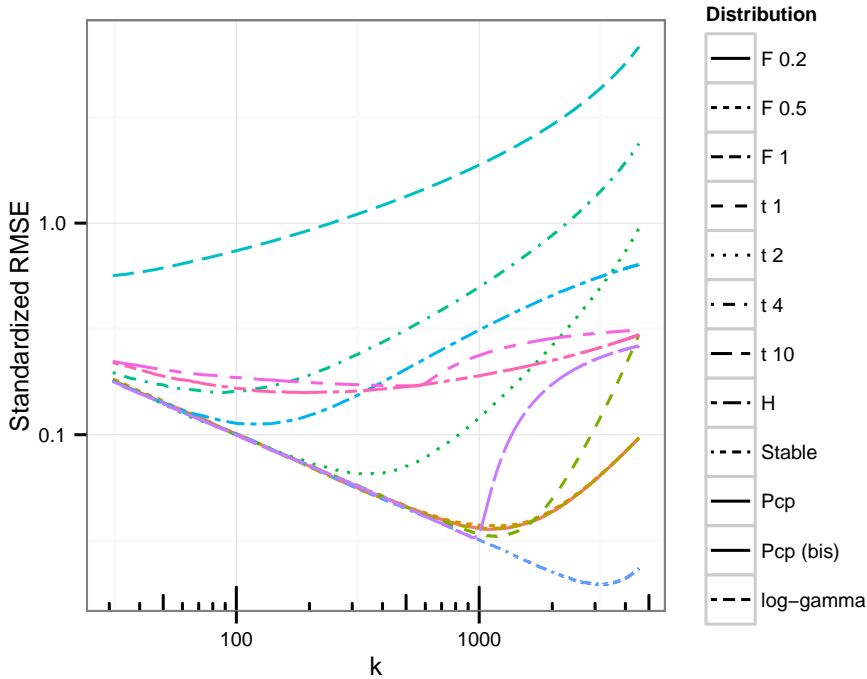


FIGURE 3.3. Monte-Carlo estimates of the standardised root mean square error (RMSE) of Hill estimators as a function of the number of order statistics k for samples of size 10000 from the sampling distributions.

Index $\widehat{\gamma}_H(\widehat{k}_n^{\text{DK}})$ was computed following the recommendations from Theorem 1 and discussion in (Drees & Kaufmann, 1998) :

$$\widehat{k}_n^{\text{DK}} = (2|\widehat{\rho}| + 1)^{-1/|\widehat{\rho}|} (2\widehat{\rho}\widehat{\gamma}_H)^{1/(1+2|\widehat{\rho}|)} \left(\frac{\widehat{k}_n(r_n^\zeta)}{(\widehat{k}_n(r_n))^\zeta} \right)^{1/(1-\zeta)} \quad (3.14)$$

where $\widehat{\rho}$ should belong to a consistent family of estimators of ρ (under a second-order regular variation assumption), $\widehat{\gamma}_H$ should be a preliminary estimator of γ such as $\widehat{\gamma}_H(\sqrt{n})$, $\zeta = .7$, and $r_n = 2n^{1/4}$. Following the advice from (Drees & Kaufmann, 1998), we replaced $|\widehat{\rho}|$ by 1. Note that the method for computing $\widehat{k}_n^{\text{DK}}$ depends on a variety of tunable parameters.

Comparison between performances of $\widehat{\gamma}_H(\widehat{k}_n(r_n))$ and $\widehat{\gamma}_H(\widehat{k}_n^{\text{DK}})$ are reported in Tables 3.2 and 3.3. For each distribution from Table 3.1, for sample sizes $n = 10000, 20000$, and 1000000 , 5000 experiments were replicated. As pointed out in (Drees & Kaufmann, 1998), on the sampling distributions that satisfy a second-order regular variation property, carefully tuned $\widehat{k}_n^{\text{DK}}$ is able to take advantage of it. Despite its computational and conceptual simplicity, despite the fact that it is almost parameter free, the estimator $\widehat{\gamma}_H(\widehat{k}_n(r_n))$ only suffers a moderate loss with respect to the oracle. When $|\rho| = 1$, the observed ratios are of the same order as $(2 \ln \ln n)^{1/3} \approx 1.65$. Moreover, where $\widehat{\gamma}_H(\widehat{k}_n^{\text{DK}})$ behaves erratically when facing Pareto change point distributions, $\widehat{\gamma}_H(\widehat{k}_n(r_n))$ behaves consistently.

Figure 3.4 concisely describes the behaviour of the two index selection methods on samples from the Pareto change point distribution with parameters $\gamma = 1.5, \gamma' = 1$ and

TABLE 3.2. Ratios between median selected indexes $\widehat{k}_n(r_n)$ (Lepski), $\widehat{k}_n^{\text{DK}}$ (Drees-Kaufmann) and estimated oracle index k_n^* .

d.f.	γ	$\widehat{k}_n^{\text{DK}}/k_n^*$			$\widehat{k}_n(r_n)/k_n^*$		
		$n = 1000$	20000	10000	1000	2000	10000
$F_{0.2}$	0.2	0.61	0.67	0.94	2.94	2.97	3.47
$F_{0.5}$	0.5	1.12	1.18	1.45	2.90	2.87	2.91
F_1	1	1.76	2.05	2.32	2.90	3.10	2.93
t_1	1	1.33	1.55	1.98	2.03	2.16	2.16
t_2	0.5	1.00	0.99	0.91	3.05	3.06	2.96
t_4	0.25	1.27	1.28	1.18	5.62	5.50	5.30
t_{10}	0.1	2.00	1.54	2.28	13.87	10.92	14.12
H	0.5	0.41	0.35	0.30	5.14	4.97	4.96
Stable	2	0.97	0.95	1.04	1.43	1.41	1.55
Pcp	1.5	1.85	0.45	0.15	1.32	1.21	1.10
Pcp (bis)	1.25	3.29	3.03	2.45	1.83	1.50	1.22
log-gamma	0.33	5.13	7.71	12.41	10.50	12.99	12.40

threshold τ corresponding to the $1 - 1/15$ quantile. The plain line represents the standardised RMSE of Hill estimators as a function of selected index. This figure contains the superposition of two density plots corresponding to $\widehat{k}_n^{\text{DK}}$ and $\widehat{k}(r_n)$. The density plots were generated from 5000 points with coordinates $(\widehat{k}(r_n), |\widehat{\gamma}_H(\widehat{k}(r_n))/\gamma - 1|)$ and 5000 points with coordinates $(\widehat{k}_n^{\text{DK}}, |\widehat{\gamma}_H(\widehat{k}_n^{\text{DK}})/\gamma - 1|)$. The contoured and well-concentrated density plot corresponds to the performance of $\widehat{\gamma}_H(\widehat{k}_n)$. The diffuse tiled density plot corresponds to the performance of $\widehat{k}_n^{\text{DK}}$. Facing Pareto change point samples, the two selection methods behave differently. Lepski's rule detects correctly an abrupt change at some point and selects an index slightly above that point. As the conditional bias varies sharply around the change point, this slight over estimation of the correct index still results in a significant loss as far as RMSE is concerned. The Drees-Kaufmann rule, fed with an a priori estimate of the second-order parameter, picks out a much smaller index, and suffers a larger excess risk.

6 Appendix

6.1 Calibration of the preliminary selection rule

Darling & Erdős (1956) establish (among other things) that letting Z_n denote $\sup_{k \leq n} \sum_{i=1}^k (E_i - 1)/\sqrt{k}$, where E_i 's are independent exponentially distributed random variables, the sequence $\sqrt{2 \ln \ln n} \left(Z_n - \sqrt{2 \ln \ln n - \ln \ln \ln(n)} \right)$ converges in distribution towards a non degenerate distribution which is closely related to the Gumbel distribution. In other words, asymptotically, Z_n behaves almost like the maximum of $\ln n$ independent standard Gaussian random variables.

TABLE 3.3. Ratios between median RMSE of and median optimal RMSE.

d.f.	γ	RMSE($\widehat{\gamma}_H(\widehat{k}_n^{\text{DK}})$)/RMSE($\widehat{\gamma}_H(k_n^*)$)			RMSE($\widehat{\gamma}_H(\widehat{k}_n(r_n))$)/RMSE($\widehat{\gamma}_H(k_n^*)$)		
		$n = 1000$	20000	10000	1000	2000	10000
$F_{0.2}$	0.2	1.12	1.12	1.02	2.06	2.26	2.69
$F_{0.5}$	0.5	1.03	1.03	1.14	2.12	2.23	2.70
F_1	1	1.22	1.31	1.59	2.07	2.23	2.64
t_1	1	1.26	1.34	1.74	2.31	2.39	3.11
t_2	0.5	1.11	1.08	1.05	2.06	2.09	2.20
t_4	0.25	1.10	1.07	1.04	1.85	1.81	1.84
t_{10}	0.1	1.10	1.09	1.08	1.76	1.72	1.64
H	0.5	1.28	1.37	1.48	2.15	2.18	2.12
Stable	2	1.01	0.99	0.98	1.99	2.52	3.60
Pcp	1.5	4.25	1.66	2.52	2.50	2.68	3.63
Pcp (bis)	1.25	3.38	4.47	7.45	2.43	2.56	3.10
log-gamma	0.33	1.23	1.28	1.39	1.45	1.43	1.37

6.2 Proof of Corollary 3.10

Let $Z = f(E_1, \dots, E_n) = (U \circ \exp) \left(\sum_{i=1}^k E_i/i \right)$. Then,

$$|\partial_i f| \leq \frac{1}{i} \sup_x \frac{1}{h(x)}, \quad \text{for } i \geq k$$

and

$$\|\nabla f\|^2 = \sum_{i=k}^n \frac{1}{i^2} \frac{1}{(h \circ f)^2}.$$

Let $c < 1$, then for all $\lambda, 0 \leq \lambda \leq c(k \inf_x h(x))$,

$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{4/k(1 + 1/k) \mathbb{E}[1/h(Z)^2] \lambda^2}{2(1 - c)}.$$

Now, start from the first statement in Theorem 3.9,

$$\begin{aligned} \text{Ent} [e^{\lambda(Z - \mathbb{E}Z)}] &\leq \frac{2\lambda^2}{1 - c} \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)} \|\nabla f\|^2] \\ &= \frac{4\lambda^2}{2(1 - c)k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{e^{\lambda(Z - \mathbb{E}Z)}}{h(Z)^2} \right] \\ &\leq \frac{4\lambda^2}{2(1 - c)k} \left(1 + \frac{1}{k}\right) \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)}] \mathbb{E} \left[\frac{1}{h(Z)^2} \right] \end{aligned}$$

where the last inequality follows from Chebychev negative association inequality. Hence,

$$\frac{d \log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)}}{d\lambda} = \frac{\text{Ent} [e^{\lambda(Z - \mathbb{E}Z)}]}{\lambda^2 \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)}]} \leq \frac{1}{2(1 - c)k} \left(1 + \frac{1}{k}\right) \mathbb{E} \left[\frac{1}{h(Z)^2} \right].$$

This differential inequality is readily solved and leads to the corollary.

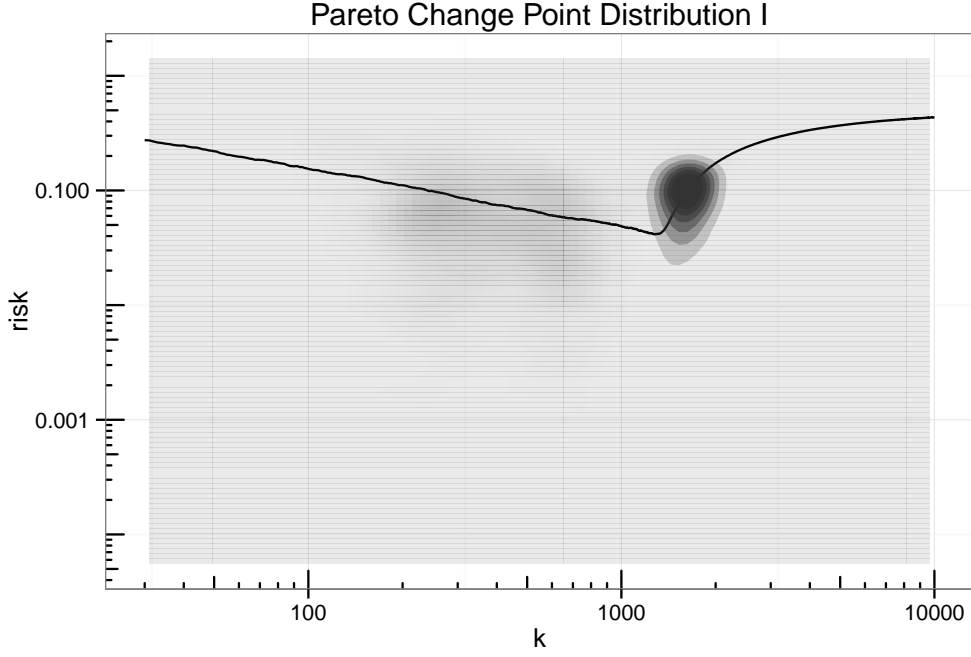


FIGURE 3.4. Risk plot for samples of size 10000 from the Pareto change point distribution with parameters $\gamma = 1.5, \gamma' = 1$ and threshold τ corresponding to the $1 - 1/15$ quantile. The concentrated density plot corresponds to points $(\hat{k}(r_n), |\hat{\gamma}_H(\hat{k}(r_n))/\gamma - 1|)$.

6.3 Proof of Abelian Proposition 3.12

The proof proceeds by classical arguments. In the sequel, we use the almost sure representation argument. Without loss of generality, we assume that all the random variables live on the same probability space, and that for any intermediate sequence (k_n) , $\sqrt{k_n}(Y_{(k_n+1)} - \ln(n/k_n))$ converges almost surely towards a standard Gaussian random variable. Complemented with dominated convergence arguments, the next lemma will be the key element of the proof.

Lemma 3.27. *Let $\eta \in RV_\rho, \rho \leq 0$ and $Y_{(k_n+1)}$ be the $(k_n + 1)$ th largest order statistic of a standard exponential sample, then for any intermediate sequence (k_n) and for all $u > 0$,*

$$\lim_{n \rightarrow \infty} \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} = e^{\rho u} \text{ p.s.}$$

Démonstration. Note that

$$\frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} = \frac{\eta((n/k_n)e^{u+Y_{(k_n+1)} - \log(n/k_n)})}{\eta(n/k_n)}.$$

Then, the result follows since $Y_{(k_n+1)} - \log(n/k_n) \xrightarrow{\text{p.s.}} 0$ and the convergence $\eta(tx)/\eta(t) \rightarrow x^\rho$ is locally uniform on $(0, \infty)$. \square

In order to secure dominated convergence arguments, we will use Drees's improvement of Potter's inequality (See de Haan & Ferreira, 2006, page 369). For every $\epsilon, \delta > 0$, there exists $t_0 = t_0(\epsilon, \delta)$ such that for $t, tx \geq t_0$,

$$|\eta(tx)/\eta(t) - x^\rho| \leq x^\rho \epsilon \max(x^\delta, x^{-\delta}). \quad (3.15)$$

To prove Proposition 3.12, we start from Representation (3.2) :

$$\widehat{\gamma}_H(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \int_0^{E_i} (\gamma + \eta(e^{u+Y_{(k_n+1)}})) du.$$

By the Pythagorean relation,

$$\text{Var}(\widehat{\gamma}_H(k_n)) = \text{Var}(\mathbb{E}[\widehat{\gamma}_H(k_n) | Y_{(k_n+1)}]) + \mathbb{E}[\text{Var}(\widehat{\gamma}_H(k_n) | Y_{(k_n+1)})],$$

so that

$$\begin{aligned} & \frac{k_n \text{Var}(\widehat{\gamma}_H(k_n)) - \gamma^2}{\eta(n/k_n)} \\ &= \frac{k_n \text{Var}(\mathbb{E}[\widehat{\gamma}_H(k_n) | Y_{(k_n+1)}])}{\eta(n/k_n)} + k_n \mathbb{E} \left[\frac{\text{Var}(\widehat{\gamma}_H(k_n) | Y_{(k_n+1)}) - \frac{\gamma^2}{k_n}}{\eta(n/k_n)} \right]. \end{aligned}$$

The second summand can be further decomposed using (3.2),

$$\begin{aligned} & \frac{k_n \text{Var}(\widehat{\gamma}_H(k_n)) - \gamma^2}{\eta(n/k_n)} \\ &= \underbrace{\frac{k_n \text{Var}(\mathbb{E}[\widehat{\gamma}_H(k_n) | Y_{(k_n+1)}])}{\eta(n/k_n)}}_{(I)} + \underbrace{\eta\left(\frac{n}{k_n}\right) \mathbb{E} \left[\text{Var} \left[\int_0^E \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} du \mid Y_{(k_n+1)} \right] \right]}_{(II)} \\ & \quad + 2\gamma \underbrace{\mathbb{E} \left[\text{Cov} \left[E, \int_0^E \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} du \mid Y_{(k_n+1)} \right] \right]}_{(III)}. \end{aligned}$$

We check that (I) and (II) tend to 0 and then that (III) converges towards a finite limit.

Fix $\epsilon, \delta > 0$ and define $M = \sup\{\eta(t), t \leq t_0\}$.

Let A_n denote the event $\{Y_{(k_n+1)} > \ln t_0(\epsilon, \delta)\}$. For n such that $\ln(n/k_n) \leq 2 \ln t_0$, as $Y_{(k_n+1)}$ sub-gamma with variance factor $1/k_n$,

$$\mathbb{P}\{A_n^c\} \leq \exp(-k_n(\ln(n/k_n))^2/8).$$

We first check that (II) tends to 0. Let n be such that $n/k_n \geq t_0$ and W_n denote the random variable $Y_{(k_n+1)} - \ln(n/k_n)$. Note that for $0 \leq \lambda \leq k_n/2$,

$$\mathbb{E}e^{\lambda|W_n|} \leq 2e^{\frac{\lambda^2}{k_n}}.$$

Using Jensen's inequality and Fubini's Theorem,

$$\begin{aligned} \mathbb{E} \left[\text{Var} \left[\int_0^E \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} du \mid Y_{(k_n+1)} \right] \right] &\leq \mathbb{E} \left[\mathbb{E} \left[E \int_0^E \left(\frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} \right)^2 du \mid Y_{(k_n+1)} \right] \right] \\ &= \int_0^\infty e^{-v} v \int_0^v \mathbb{E} \left[\left(\frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} \right)^2 \right] dudv \\ &= \int_0^\infty e^{-v} v \int_0^v \mathbb{E} \left[\left(\frac{\eta(e^{u+W_n} n/k_n)}{\eta(n/k_n)} \right)^2 \right] dudv \end{aligned}$$

We now apply Potter's inequality (3.15) on the event A_n with $t = n/k_n > t_0$ and $tx = e^{u+Y_{(k_n+1)}} > t_0, u > 0$:

$$\begin{aligned} & \mathbb{E} \left[\text{Var} \left[\int_0^E \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} du \mid Y_{(k_n+1)} \right] \right] \\ & \leq \int_0^\infty e^{-v} v \int_0^v \mathbb{E} \left[\mathbb{1}_{A_n} e^{2\rho(u+W_n)} (1 + \epsilon e^{\delta(u+|W_n|)})^2 + \mathbb{1}_{A_n^c} \frac{M^2}{\eta(n/k_n)^2} \right] dudv \\ & \leq \int_0^\infty e^{-v} v \int_0^v \mathbb{E} [e^{2\rho W_n} 2 (1 + \epsilon^2 e^{2\delta(u+|W_n|)})] dudv + \frac{2M^2}{\eta(n/k_n)^2} \mathbb{E} \mathbb{1}_{A_n^c}. \end{aligned}$$

The first summand has a finite limit thanks to Lemma 3.27. The second summand converges to 0 as $\mathbb{E} \mathbb{1}_{A_n^c}$ tends to 0 exponentially fast while $1/\eta(n/k_n)^2$ tends to infinity algebraically fast.

Bounds on (i) are easily obtained, using Jensen's Inequality and Poincaré Inequality.

$$\begin{aligned} \frac{k_n \text{Var} (\mathbb{E}[\widehat{\gamma}_H(k_n) \mid Y_{(k_n+1)}])}{\eta(n/k_n)} &= \frac{k_n \text{Var} (\int_0^\infty \eta(e^{u+Y_{(k_n+1)}}) e^{-u} du)}{\eta(n/k_n)} \\ &\leq 4\eta(n/k_n) \mathbb{E} \left[\left(\int_0^\infty \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} e^{-u} du \right)^2 \right] \\ &\leq 4\eta(n/k_n) \mathbb{E} \left[\int_0^\infty \left(\frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} \right)^2 e^{-u} du \right]. \end{aligned}$$

Using the line of arguments as for handling the limit of (II), we establish that (I) converges to 0.

We now check that (III) converges towards a finite limit. Note that

$$\mathbb{E} \left[\text{Cov} \left[E, \int_0^E \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} du \mid Y_{(k_n+1)} \right] \right] = \mathbb{E} \left[(E-1) \int_0^E \frac{\eta(e^{u+Y_{(k_n+1)}})}{\eta(n/k_n)} du \right].$$

By Lemma 3.27, for almost every $u > 0$,

$$(E-1) \frac{\eta(e^{u+W_n} n/k_n)}{\eta(n/k_n)} \xrightarrow{n \rightarrow \infty} (E-1) e^{\rho u},$$

and

$$|E-1| \int_0^E \left| \frac{\eta(e^{u+W_n} n/k_n)}{\eta(n/k_n)} \right| du \leq |E-1| \int_0^E e^{\rho(u+W_n)} (1 + \epsilon e^{\delta(u+|W_n|)}) du + \mathbb{1}_{A_n^c} E |E-1| \frac{M}{|\eta(n/k_n)|}.$$

The first term is finite as the integral of a continuous function on a compact.

Thus,

$$(E-1) \int_0^E \frac{\eta(e^{u+W_n} n/k_n)}{\eta(n/k_n)} du \rightarrow_n (E-1) \int_0^E e^{\rho u} du = (E-1) \frac{e^{\rho E} - 1}{\rho}.$$

The expected value of the last random variable is $1/(1-\rho)^2$.

We check that for sufficiently large n ,

$$\begin{aligned}
& \mathbb{E} \left[|E - 1| \int_0^E \frac{|\eta(e^{u+W_n} n/k_n)|}{|\eta(n/k_n)|} du \right] \\
& \leq \mathbb{E} \left[|E - 1| \int_0^E e^{\rho(u+W_n)} (1 + \epsilon e^{\delta(u+|W_n|)}) + \mathbb{1}_{A_n^c} |E - 1| \frac{M}{|\eta(n/k_n)|} du \right] \\
& \leq \mathbb{E} \left[e^{\rho W_n} \left(2 + \frac{\epsilon}{\delta(1-\delta)^2} e^{\delta|W_n|} \right) \right] + \frac{M}{|\eta(n/k_n)|} \mathbb{E} \mathbb{1}_{A_n^c} \\
& \leq 4e^{\frac{\rho^2}{k_n}} + \frac{2\epsilon}{\delta(1-\delta)^2} e^{\frac{(\delta-\rho)^2}{k_n}} + \frac{M}{|\eta(n/k_n)|} \mathbb{E} \mathbb{1}_{A_n^c}.
\end{aligned}$$

We now way conclude by dominated convergence that

$$(III) \xrightarrow{n \rightarrow \infty} \frac{2\gamma}{(1-\rho)^2}.$$

6.4 Proof of Inequality 3.10

For $k \geq 5$,

$$\mathbb{E} \left[\left(1 - e^{Y'_{(k+1)} - Y_{(k+1)}} \right)^2 \right] \leq \frac{4}{k} \frac{1}{1 - 4/k}.$$

As $Y'_{(k+1)} - Y_{(k+1)}$ is symmetric, $\mathbb{E}[Y'_{(k+1)} - Y_{(k+1)}] = 0$, and applying Jensen's inequality and elementary calculus leads to

$$\begin{aligned}
& \mathbb{E} \left[\left(1 - e^{Y'_{(k+1)} - Y_{(k+1)}} \right)^2 \right] \\
& = \mathbb{E} \left[e^{2(Y'_{(k+1)} - Y_{(k+1)})} - 1 \right] - 2\mathbb{E} \left[e^{(Y'_{(k+1)} - Y_{(k+1)})} - 1 \right] \\
& \leq \mathbb{E} \left[e^{2(Y'_{(k+1)} - Y_{(k+1)})} - 1 \right] \\
& = \prod_{i=k+1}^n \frac{1}{1 - 4i^{-2}} - 1 \\
& \leq \frac{1}{1 - 4 \sum_{i=k+1}^n i^{-2}} - 1 \\
& \leq \frac{1}{1 - 4/k} - 1 \\
& \leq \frac{4}{k} \frac{1}{1 - 4/k}.
\end{aligned}$$

6.5 Revisiting the lower bound on adaptive estimation error

Lower bounds on tail index estimation error (Carpentier & Kim, 2014a; Drees, 1998a, 2001; Novak, 2014) are usually constructed by defining sequences of local models around a pure Pareto distribution with shape parameter γ_0 . When deriving lower bounds for the estimation error under constraints like $\bar{\eta}$ is regularly varying, the elements of the local model for sample size n may be defined by

$$U_{n,h}(t) = t^{\gamma_0 + d_n h(0)} \exp \int_1^t d_n \frac{h(c_n/s) - h(0)}{s} ds$$

where h is square integrable over $[0, 1]$, $d_n \rightarrow 0$, $nd_n^2/c_n \rightarrow 1$ (Drees, 2001). The sequences d_n and c_n are chosen in such a way that $d_n |h(c_n/s) - h(0)| = |\bar{\eta}(s)|$ satisfies the required constraint. If the local alternatives are Pareto change point distributions as in (Novak, 2014) and (Carpentier & Kim, 2014a), $h(x) = \mathbb{1}_{\{x \leq 1\}}$, $c_n = \tau_n^{1/\gamma_0}$. Drees (2001) explores a richer collection of local alternatives in order to fit into the theory of weak convergence of local experiments.

In order to explore adaptivity as in (Carpentier & Kim, 2014a), it is necessary to handle simultaneously a collection of sequences $(d_n, c_n)_n$ corresponding to different rates of decay of the von Mises function. The difficulty of estimation is connected with the difficulty of distinguishing alternatives with different tail indexes, that is, with the hardness of a multiple hypotheses testing problem. In order to lower bound the testing error, Carpentier & Kim chose to use Fano's Lemma (Cover & Thomas, 1991, See). Using Fano's Lemma requires bounding the Kullback-Leibler divergence between the different local alternatives which is not as easy as bounding the divergence between a Pareto change point distribution and a pure Pareto distribution.

The next lemma is from (Birgé, 2005). It can be used in the derivation of risk lower bounds instead of the classical Fano Lemma. Just as Fano's Lemma, it states a lower bound on the error in multiple hypothesis testing. But as it only requires computing the Kullback-Leibler divergence to the localisation center, in the present setting, it significantly alleviates computations and makes the proof more concise and more transparent.

Lemma 3.28. (*Birgé-Fano*) *Let P_0, \dots, P_N be a collection of probability distributions on some space, and let A_0, \dots, A_M be a collection of pairwise disjoint events, then the following holds*

$$\min_i P_i\{A_i\} \leq \frac{2e}{1+2e} \vee \frac{\frac{1}{M} \sum_{i=1}^M \mathcal{K}(P_i, P_0)}{\ln(M+1)}.$$

In order to take advantage of Lemma 3.28, we use the Bayesian game designed in (Carpentier & Kim, 2014a).

Theorem 3.29. *Let $\gamma > 0$, $\rho < -1$, and $0 \leq v \leq e/(1+2e)$. Then, for any tail index estimator $\hat{\gamma}_H$ and any sample size n such that $M = \lfloor \ln n \rfloor > e/v$, there exists a collection $(P_i)_{i \leq M}$ of probability distributions such that*

- i) $P_i \in \text{MDA}(\gamma_i)$ with $\gamma_i > \gamma$,
- ii) P_i meets the von Mises condition with von Mises function η_i satisfying

$$\bar{\eta}_i(t) \leq \gamma t^{\rho_i}$$

where $\rho_i = \rho + i/M < 0$,

iii)

$$\max_{i \leq M} P_i^{\otimes n} \left\{ |\hat{\gamma}_H - \gamma_i| \geq \frac{C_\rho}{4} \gamma_i \left(\frac{v \ln \ln n}{n} \right)^{|\rho_i|/(1+2|\rho_i|)} \right\} \geq \frac{1}{1+2e}$$

and

$$\max_{i \leq M} \mathbb{E}_{P_i^{\otimes n}} \left[\frac{|\hat{\gamma}_H - \gamma_i|}{\gamma_i} \right] \geq \frac{C_\rho}{4(1+2e)} \left(\frac{v \ln \ln n}{n} \right)^{|\rho|/(1+2|\rho|)},$$

with $C_\rho = 1 - \exp\left(-\frac{1}{2(1+2|\rho|^2)}\right)$.

Proof of Theorem 3.29. Choose v so that $0 \leq v \leq 2e/(1+2e)$. The number of alternative hypotheses M is chosen in such a way that $\ln(n/(v \ln M)) \leq M$. If $\lfloor \ln n \rfloor \geq e/v$, $M = \lfloor \ln n \rfloor$ will do.

The center of localisation P_0 is the pure Pareto distribution with shape parameter $\gamma > 0$ ($P_0\{(\tau, \infty)\} = \tau^{-1/\gamma}$). The local alternatives P_1, \dots, P_M are Pareto change point distributions. Each P_i is defined by a breakpoint $\tau_i > 1$ and an ultimate Pareto index γ_i . If F_i denotes the distribution function of P_i ,

$$\bar{F}_i(x) = x^{-1/\gamma} \mathbb{1}_{\{1 \leq x \leq \tau_i\}} + \tau_i^{-1/\gamma} (x/\tau_i)^{-1/\gamma_i} \mathbb{1}_{\{x \geq \tau_i\}}.$$

Karamata's representation of $(1/\bar{F}_i)^{\leftarrow}$ is

$$U_i(t) = t^{\gamma_i} \exp\left(\int_1^t \frac{\eta_i(s)}{s} ds\right)$$

with $\eta_i(s) = (\gamma - \gamma_i) \mathbb{1}_{\{s \leq \tau_i^{1/\gamma}\}}$.

The Kullback-Leibler divergence between P_i and P_0 is readily calculated,

$$\mathcal{K}(P_i, P_0) = \bar{F}_i(\tau_i) \left(\frac{\gamma_i}{\gamma} - 1 - \ln \frac{\gamma_i}{\gamma} \right) = \tau_i^{-1/\gamma} \left(\frac{\gamma_i}{\gamma} - 1 - \ln \frac{\gamma_i}{\gamma} \right).$$

If $\gamma_i > \gamma$, the next upper bound holds,

$$\mathcal{K}(P_i, P_0) \leq \frac{\tau_i^{-1/\gamma}}{2} \left(\frac{\gamma_i}{\gamma} - 1 \right)^2.$$

The breakpoints and tail indices are chosen in such a way that all upper bounds are equal (namely $n\tau_i^{-1/\gamma}(\gamma_i/\gamma - 1)^2$ does not depend on i),

$$\begin{aligned} \tau_i &= (n/(v \ln M))^{\gamma/(1+2|\rho_i|)} \\ \gamma_i &= \gamma + \gamma (n/(v \ln M))^{\rho_i/(1+2|\rho_i|)}, \end{aligned}$$

so that $\mathcal{K}(P_i^{\otimes n}, P_0^{\otimes n}) = n\mathcal{K}(P_i, P_0) \leq v \ln M$ for all $1 \leq i \leq M$.

Note that for all $t > 1$,

$$|\eta_i(t)| = |\gamma - \gamma_i| \mathbb{1}_{\{t \leq \tau_i^{1/\gamma}\}} \leq \gamma \tau_i^{\rho_i/\gamma} \mathbb{1}_{\{t \leq \tau_i^{1/\gamma}\}} \leq \gamma t^{\rho_i}$$

the upper bound being achieved at $t = \tau_i$.

Now, let $\hat{\gamma}_H$ be any tail index estimator. Define region A_i , as the set of samples such that γ_i minimises $|\hat{\gamma}_H - \gamma_j|$ for $1 \leq j \leq M$. Then, if the event A_i is not realised,

$$|\hat{\gamma}_H - \gamma_i| \geq \frac{1}{2} \min_{1 \leq j \leq M, j \neq i} |\gamma_j - \gamma_i|.$$

By Birgé's Lemma,

$$\max_{i \leq M} \mathbb{P}_i^{\otimes n} \left\{ |\hat{\gamma}_H - \gamma_i| \geq \frac{1}{2} \min_{1 \leq j \leq M, j \neq i} |\gamma_j - \gamma_i| \right\} \geq \frac{1}{1+2e}.$$

In order to make the whole construction useful, it remains to choose the “second-order parameters” ρ_i 's (the true second-order parameter of each P_i is infinite!). We will need an upper bound on γ_i/γ (but we already have $\gamma_i/\gamma \leq 2$), as well as a lower bound on $|\gamma_j - \gamma_i|/\gamma$ for $j \neq i$ that scales like $(n/\ln \ln n)^{\rho_i/(1+2|\rho_i|)}$.

Following Carpentier & Kim (2014a), we finally choose ρ_i as $\rho_i = \rho + i/M$ for $1 \leq i \leq M$. Then, for $j < i$, using that $\ln(n/(v \ln M)) \leq M$ and $\rho_i - \rho_j = (i - j)/M$,

$$\begin{aligned}
\frac{|\gamma_j - \gamma_i|}{\gamma_i} &\geq \frac{|\gamma_j - \gamma_i|}{2\gamma} \\
&\geq \frac{1}{2} \left(\frac{n}{v \ln M} \right)^{\rho_i/(1+2|\rho_i|)} \left| 1 - \left(\frac{n}{v \ln M} \right)^{\rho_j/(1+2|\rho_j|) - \rho_i/(1+2|\rho_i|)} \right| \\
&\geq \frac{1}{2} \left(\frac{n}{v \ln M} \right)^{\rho_i/(1+2|\rho_i|)} \left[1 - \exp \left(\frac{i - j}{M(1 + 2|\rho_i|)(1 + 2|\rho_j|)} \ln \left(\frac{n}{v \ln M} \right) \right) \right] \\
&\geq \frac{1}{2} \left(\frac{n}{v \ln M} \right)^{\rho_i/(1+2|\rho_i|)} \left[1 - \exp \left(\frac{i - j}{M(1 + 2|\rho_i|)(1 + 2|\rho_j|)} \right) \right] \\
&\geq \frac{C_\rho}{2} \left(\frac{n}{v \ln M} \right)^{\rho_i/(1+2|\rho_i|)}
\end{aligned}$$

where C_ρ may be chosen as $1 - \exp \left(-\frac{1}{2(1+2|\rho|^2)} \right)$. □

Chapitre 4

Predicting unusual but expected events : an application of Extreme Value Theory to influenza-associated mortality and morbidity

Contents

1	Background	118
2	Methods	119
2.1	Extreme Value Theory (EVT) - Block Maxima Method	119
2.2	Data sources	119
2.3	Application of EVT	120
3	Results	121
3.1	Descriptive summary	121
3.2	EVT results	122
4	Discussion	123

Extreme Value Theory (EVT) has been applied to studies in earth sciences, genetics and finance to predict extreme events from distribution of their maxima. We applied this methodology to estimate unusual influenza mortality and morbidity.

We compiled weekly rates of Pneumonia and Influenza (P&I) deaths over the period 1970-2011 and Influenza-like-illness (ILI) over the period 1984-2011. The sum of weekly P&I mortality and weekly ILI over a moving window of eight consecutive weeks was used to create cumulative standardized maximum rates. Fifty-two weeks maxima were assumed to follow a Generalized Extreme Value distribution, fitted using the maximum likelihood method. We derived the return levels, i.e. the expected maxima which should be exceeded on average once over different (return) periods. We also estimated probabilities to exceed some rates values during one-year periods.

The predicted P&I cumulative standardized mortality rate return level was estimated at 23 per 100,000 (95% confidence interval = 11-34) for a 30-year return period, meaning

that this rate should be exceeded once during the next 30 years or that there is approximately 3% risk each year (1/30) that this rate will exceed 23 per 100,000. Similarly the predicted cumulative ILI rate 30-year return level was estimated at 7,371 per 100,000, which corresponds to 91% (66%-117%) the observed maximum. Each year the probability of a cumulative P&I mortality rate over 64 per 100,000, i.e. twice the observed maximum over the last 40 years, is 0.23%.

This method can be applied to other topics in epidemiology thus contributing to public health planning for extreme events.

1 Background

A central question for resource planning in public health is to predict the likelihood that exceptional or extreme events will occur in the not too distant future (Bresee & Hayden, 2013). Such events might be, for example, an unusually severe community epidemic, a major heat wave, or an accidental toxic exposure. Statistical approaches focused on extreme values have shown promising results to forecast unusual events in earth sciences, genetics and finance, but to our knowledge they have not yet been applied to public health data. For instance, the Extreme Value Theory (EVT) was developed in the 1920s (Beirlant et al., 2004; Coles, 2001) and has been used to predict the occurrence of events as varied as droughts and flooding (Katz et al., 2002) or financial crashes (Embrechts et al., 1997). To our knowledge, there is a single application of EVT in public health (Guillou & Kratz, 2013), but in this case EVT was used to detect outliers (here seen as extreme events) in time series surveillance data, rather than estimate the risk of future extreme events.

The main goal of EVT is to assess, from a series of observations, the probability of events that are more extreme than those previously recorded. For example, 40% of the Netherlands is below the sea level and has to be protected against the sea by dikes. The height of dikes can be calculated from storm data collected for around 100 years using EVT, so that the risk of flooding would be less than one every 10,000 years. Similarly, the likelihood of epidemics of unusual sizes could be determined by applying EVT to past epidemic observations, which could then help planning resources for mitigating the burden of these epidemics. The aim of this paper is to present how EVT can be applied in epidemiology and to illustrate with a working example how the expected maximal seasonal influenza morbidity or mortality burden can be estimated. We chose to concentrate on influenza due to the availability of long-term mortality and morbidity records in many countries. We provide forecasts of expected maximal influenza-associated mortality and influenza-like-illnesses for periods up to 50 years based on past observations.

2 Methods

2.1 Extreme Value Theory (EVT) - Block Maxima Method

Based on EVT (Coles, 2001), the limit distribution of the maximum of n random variables belongs to a three-parameter Generalized Extreme Value (GEV) family, and has a distribution function of the form :

$$G(z) = \exp \left\{ - \left[1 + \gamma \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\gamma} \right\}$$

where μ is the location parameter, σ the scale parameter and γ the shape parameter. Three classes of GEV distributions are defined by on the value of γ :

- The Fréchet class ($\gamma > 0$), representing distributions with heavy tail.
- The Gumbel class ($\gamma = 0$), representing distributions with lighter tail.
- The Weibull class ($\gamma < 0$), representing distributions with finite tail.

A classical method for modelling the extremes of a stationary time series is the method of block maxima, in which consecutive observations are grouped into non-overlapping blocks of length n , generating a series of m block maxima, $M_{n,1} \dots, M_{n,m}$, say, to which the GEV distribution can be fitted for some large value of n . The usual approach is to consider blocks of length one year, thus yielding annual maxima (Coles, 2001). Once a GEV distribution is fitted to n empirical observations, it becomes possible to (1) estimate the probability of an event that has not been observed yet, e.g the probability to exceed a certain threshold larger than the largest observation, and (2) estimate an extreme quantile, e.g a quantile of order of $1 - 1/(xn)$ where x is much larger than 1. This is an extreme quantile because only n observations are included in the block. The first estimate is simply given by the distribution function of the GEV. The $(1 - p)$ -quantile z_p of a GEV distribution is called the return level associated with the return period $t_p (= 1/p)$. For blocks of one year, the level z_p is expected to be exceeded on average once every t_p years. More precisely, z_p is exceeded by the annual maximum in any particular year with probability p . The level z_p can be expressed in terms of the GEV parameters :

$$z_p = \begin{cases} \mu - \frac{\sigma}{\gamma} [1 - \{-\log(1 - p)\}^{-\gamma}] & \text{for } \gamma \neq 0 \\ \mu - \sigma \log\{-\log(1 - p)\} & \text{for } \gamma = 0. \end{cases}$$

2.2 Data sources

Mortality and population data

Pneumonia and Influenza (P&I) mortality data provide us a specific indicator of influenza mortality (Simonsen et al., 1997). We obtained weekly number of P&I deaths in France from July 1970 to June 2011 from death certificates collected by the Centre d'épidémiologie sur les causes médicales de décès (CépiDc, 2014). We used the following codes from the International Classification of Diseases (ICD) : 480-493 (1970 -1971, 7th revision), 470-474 (1972 -1978 ICD-8th revision), 480-487 (1979 - 1999 ICD-9th revision)

and J09-J18 (2000-2011, ICD-10th revision). Mortality data were broken down by 9 age groups (0-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, ≥ 75). Population data obtained from regular censuses were used to calculate weekly age-specific P&I mortality rates (INSEE, 2014). We then calculated weekly age-standardized death rates using the 2011 French population structure as reference.

Morbidity data

Influenza-like-illnesses (ILI) has been shown to be a good proxy for influenza incidence in France and elsewhere (Denoué et al., 2007; Goldstein et al., 2011). Weekly ILI incidence was obtained from November 1984 to June 2011 from the French Sentinelles system, a nationwide network of general practitioners (sen, 2014; Falchi et al., 2011). Approximately 1,500 practicing physicians in France voluntarily participate in disease surveillance through this system. The ILI case definition consisted of a combination of fever $> 39^{\circ}\text{C}$ myalgia, and respiratory symptoms.

Definition of cumulative rates of P&I Mortality and ILI

We defined the cumulative rates of P&I mortality (cPI) and ILI ($cILI$) as the sum of weekly P&I mortality and weekly ILI over eight consecutive weeks using a moving time window, through the entire time series. The eight-week period coincides with the length of a typical influenza epidemic, which is estimated to last approximately 2 months on average (Bresee & Hayden, 2013; Denoué et al., 2007).

2.3 Application of EVT

In our application, $M_{n,1}, \dots, M_{n,m}$ stand for either the maxima of n cPI or n $cILI$ observations within a respiratory year (from July to June to encompass annual influenza epidemics). We obtained 42 annual maxima for the cPI time series spanning 1970-2011 and 28 maxima for the $cILI$ time series spanning 1984-2011, denoted $cPIM$ and $cILIM$, respectively. These maxima are akin to the “size” of an influenza epidemic as they represent the (maximum) cumulative burden of an 8-week period.

Assuming that $cPIM$ and $cILIM$ were distributed according to a GEV, we estimated the GEV parameters and their 95% confidence intervals by the maximum-likelihood method. The fit of GEV models was assessed by quantile-quantile plots (QQ plots). Return level plots with 95% confidence intervals were produced for return periods up to 50 years. Then, return levels were estimated with 95% confidence intervals (CI) using a profile likelihood method, at three particular return periods : 20, 30 and 50 for P&I data, and 10, 20 and 30 years for ILI data (Coles, 2001). Finally, we computed the probabilities to exceed 1.0, 1.2, 1.5 or 2.0 times the largest annual maxima of $cPIM$ or $cILI$ ever observed. We used the *evd* and *extRemes* packages in R v 3.0.2 for all calculations (Coles, 2001).

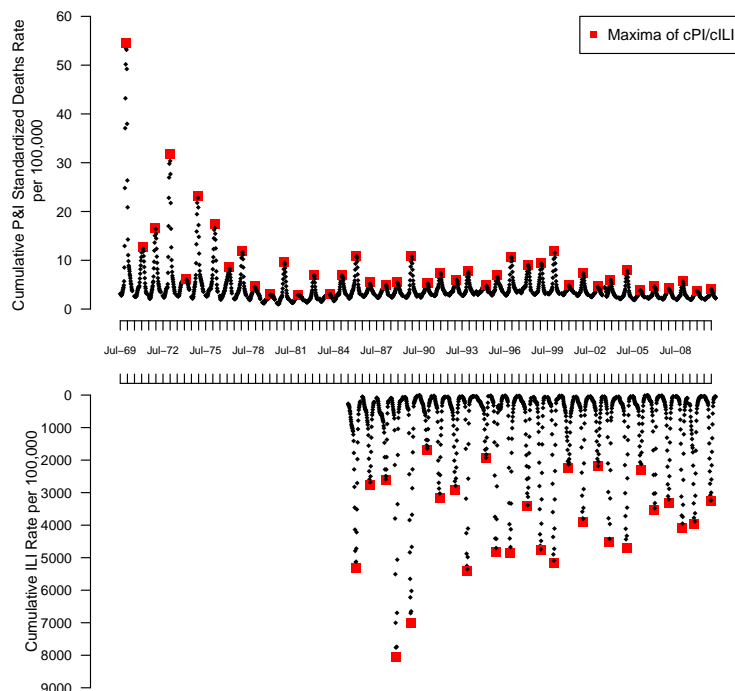


FIGURE 4.1. Maxima of Age-standardized cumulative P&I mortality rates (1969-2007) during transmission season months in France and ILI rates (1984-2011) in France. cPI and $cILI$ rates correspond to black symbols and the annual maxima of cPI and $cILI$ to the red symbols.

Outcomes	Nb of obs	Min	Max	Mean	Median	Interquartile
Mortality ($cPIM$)	42	3	32	8	7	5
Morbidity ($cILIM$)	28	1,681	8,062	4,007	3,893	1,992

TABLE 4.1. Distributions of the annual maxima of cumulative Pneumonia and Influenza mortality ($cPIM$) and cumulative Influenza-Like Illness ($cILIM$) in France

3 Results

3.1 Descriptive summary

Between 1970 and 2011, low $cPIM$ mortality rates (<4 per 100,000) were reported in four respiratory years (1979-80, 1981-82, 1983-84, and the pandemic year 2009-10, Figure 4.1). In contrast, four respiratory years (1971-72, 1972-73, 1974-75, 1975-76) were associated with high $cPIM$ death rates (>15 per 100,000, Table 4.1). For morbidity during the period between 1984 and 2011, $cILIM$ rates $<2,000$ per 100,000 were reported in two respiratory years (1990-91, 1993-94), while three years (1984-85, 1988-89, 1989-90) were associated with $cILIM$ rates $>6,000$ per 100,000 (Table 4.1, Figure 4.1).

Influenza Measure	Location parameter μ	Scale parameter σ	Shape parameter γ
$cPIM$	5.61 [4.80 ; 6.59]	2.49 [1.83 ; 3.45]	0.37 [0.12 ; 0.74]
$cILIM$	3,295 [2,779 ; 3,882]	1,227 [915 ; 1,779]	-0.01 [-0.30 ; 0.41]

TABLE 4.2. Parameter estimates and 95% confidence intervals (CI) of the Generalized Extreme Value distributions for cumulative Pneumonia and Influenza mortality ($cPIM$), 1970-2011, and cumulative Influenza-Like Illness ($cILIM$), 1984-2011, France

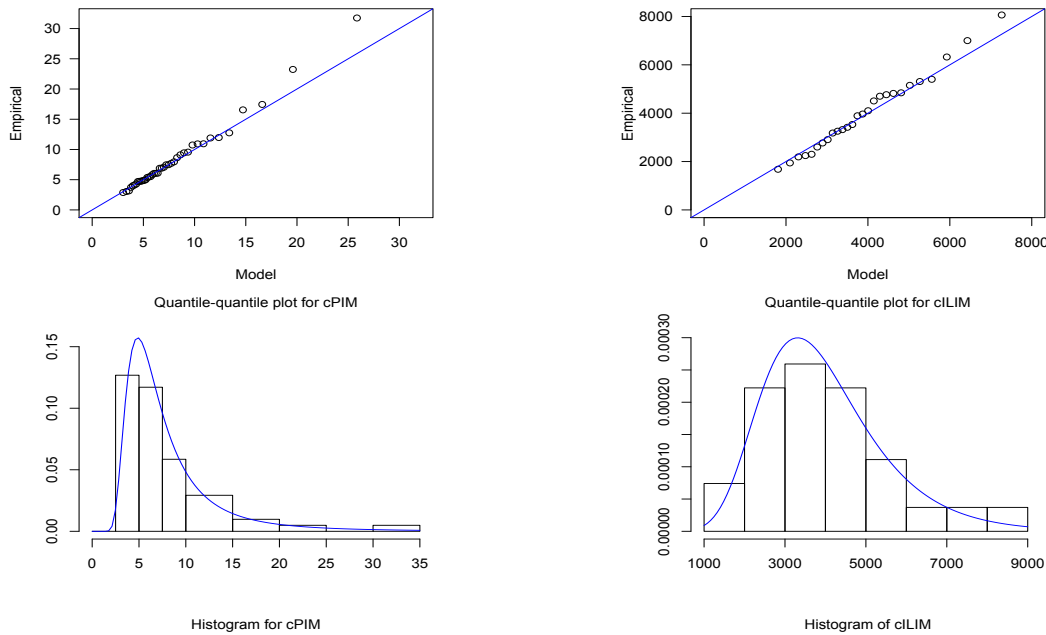


FIGURE 4.2. Quantile-Quantile (QQ) plots and histograms for (right) the annual maxima of cumulative P&I mortality and (left) the annual maxima of cumulative ILI morbidity rates.

The blue curve corresponding for the density of the fitted GEV.

3.2 EVT results

Maximum-likelihood estimates of distribution parameters for the maxima or mortality and morbidity are given in Table 4.2. The shape parameter estimate was positive for $cPIM$, with 95% CI including only positive values, suggesting a heavy tail distribution for mortality. In contrast, for $cILIM$, the shape parameter was close to 0 with a symmetric 95% CI, suggesting a less heavy tail for morbidity data. QQ-plots and histograms (Figure 4.2) showed that the fit of the GEV distributions was correct, *i.e.* the observed and predicted quantiles agreed overall, although the two most extreme observed maxima were higher than those predicted from the fitted distributions in both morbidity and mortality analyses.

The return level plots are depicted in Figure 4.3 and selected values presented in Table 4.3. To illustrate how these plots might be interpreted, Figure 4.3 (Left) suggests that over the next 30 years (the return period), the $cPIM$ should exceed 23 per 100,000 (the return

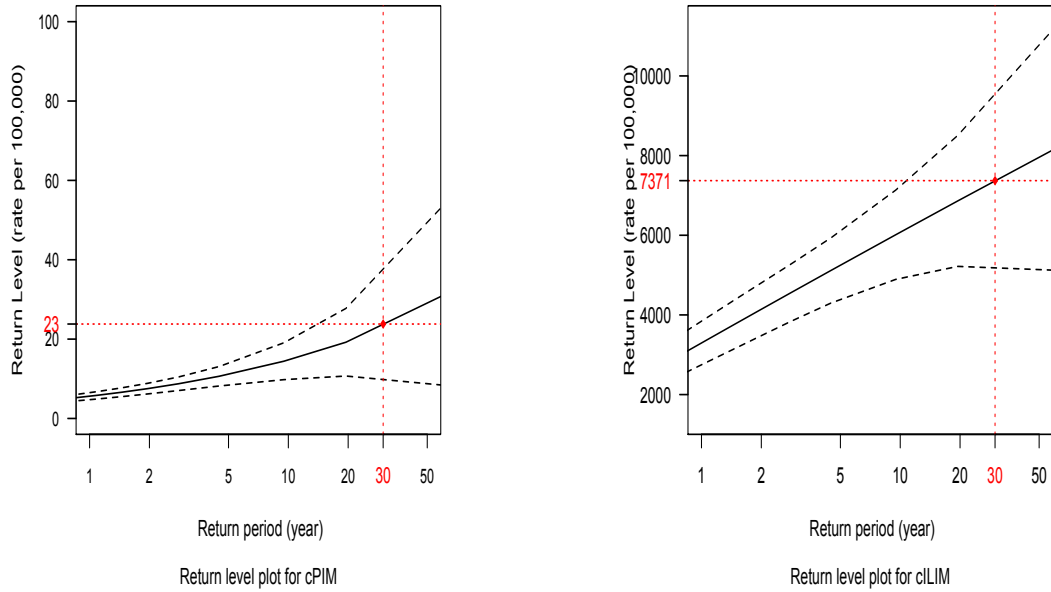


FIGURE 4.3. Return plots for (left) the annual maxima of cumulative P&I mortality, and (right) the annual maxima of cumulative ILI morbidity rates.

level) once. Another way to interpret the plot is to say that there is approximately a 3% chance (1/30) each year that the *cPIM* will exceed 23 per 100,000. Regarding *cILIM* (Figure 4.3 (Right)), the 30-year return level estimate was 7,371 per 100,000 (95% CI : 5,294-9,447), which was in the same order magnitude as the maximum *cILIM* observed over the last 25 years. Based on 2011 French census data, this corresponds to 14,530 deaths and 4,656,651 ILIs.

<i>cPIM</i> Return Period (years)	20 years	30 years	50 years
<i>cPIM</i> Return Level per 100,000 [95% CI]	19 [11 ; 28]	23 [11 ; 34]	28 [10 ; 45]
<i>cILIM</i> Return Period (years)	10 years	20 years	30 years
<i>cILIM</i> Return Level per 100,000 [95% CI]	6,035 [4,900 ; 7,170]	6,886 [5,222 ; 8551]	7,371 [5,294 ; 9,447]

TABLE 4.3. Return level estimates per 100,000 and 95% confidence intervals for *cPIM* and *cILIM*

The probabilities of exceeding some threshold value during a year (defined as multiples of the largest *cPIM* or *cILIM* observed) are summarized in Table 4.4. As an example, there is a 0.49% chance that the *cPIM* of a given year will exceed 48 per 100,000, and a 0.05% chance that the *cILIM* will exceed 12,093 per 100,000 each year. Similarly, there is a 5% chance that the *cPIM* and *cILIM* will exceed 19 and 6,882 per 100,000, respectively. This represents 0.6 and 0.8 times the maximum *cPIM* and *cILIM* observed each year.

4 Discussion

Using a simple illustrative example, we demonstrated the applicability of EVT to epidemiologic data. A GEV distribution was fitted to yearly maxima and was used to

	Times	0.6	1.0	1.2	1.5	2.0
Multiples of max(<i>cPIM</i>)		19	32	38	48	64
	Probabilities	5%	1.41%	0.88%	0.49%	0.23%
	Times	0.8	1.0	1.2	1.5	2
Multiples of max(<i>cILIM</i>)		6,882	8,062	9,674	12,093	16,124
	Probabilities	5%	1.82%	0.45%	0.05%	0.001%

TABLE 4.4. Annual probabilities to exceed 0.6, 1.0, 1.2, 1.5 or 2 times the largest *cPIM* observed and to exceed 0.8, 1.0, 1.2, 1.5 or 2 times the largest *cILIM* observed

calculate estimates of return levels and of risks of exceeding a defined threshold value over given time periods.

In our analysis, we assumed stationarity of P&I rates and ILI rates. This was clearly not the case as the data exhibited a yearly seasonal pattern, and we observed a declining trend in both *cPIM* and *cILIM* over the study period ($p < 0.0001$, $p = 0.0002$ respectively for test to zero of coefficient for time in a regression model). However, the model fits were correct and it is likely that our estimates were only slightly affected by departures from these assumptions. Methods for dealing with non-stationary distributions of maxima have been described in EVT. For example, it might be useful to consider a cyclical GEV model, that is a GEV model with time-varying location and scale parameters (Coles, 2001; Katz et al., 2002). Yet, this method requires the estimation of at least two more parameters than the model presented in this paper : this might produce large confidence intervals due to the small numbers of observations available. To improve forecasting with relatively few annual observations, one might leverage other available information that could represent covariates (temperature, humidity levels, vaccination, etc.) (Katz et al., 2002). It would also be interesting to consider a multivariate GEV model, that is to combine patterns from several neighboring countries with good surveillance data such as United Kingdom, Germany, and Spain (Coles & Tawn, 2005). Further, it may also be interesting to apply a similar approach to deaths in excess of seasonal background mortality, a metric traditionally used in influenza research, which may more specifically capture the burden of influenza every winter. While such refinements might have improved the accuracy of our estimates, they are beyond the scope of this study as the choice of the specific approach would depend on the intended use of the forecasts.

Return level estimates should be helpful in planning resource needs, much like the statistical rationale for building dikes in the Netherlands. If one assumes that antivirals, vaccines or face-masks stockpiles should be amassed for influenza epidemics, they can easily be dimensioned using estimates of ILI return levels based on an annual risk of exceeding an a priori defined *cILIM* value. For example, one could recommend sizing stockpiles based on a value that might be exceeded once in the next ten years – in our case 6,035 ILI cases per 100,000. Another example of EVT use for planning would have been to derive the return levels estimates for hospitalizations due to influenza. However the data were only available since 2002 for France, which did not allow us to conduct a

similar analysis.

We believe that extreme value theory and methods could be applied to other topics in epidemiology or in other areas of public health. For example, extreme values methods can be applied to hospital admissions to improve health care use during winter bed crises. More specifically, hospitals could forecast expected patient inflows over a given period and evaluate the surge capacity of the institution during this period. Several indicators may be worth forecasting, including the number of incoming and outgoing patients in emergency rooms or intensive care units, the number of patients waiting for a hospital bed, or the occupancy rate per service. Other examples of potential applications include anticipating the impact of heat waves.

Annexe A

Outils mathématiques

1 Inverses généralisées

Cette partie est basée sur (de Haan & Ferreira, 2006).

Définition A.1 (INVERSE GÉNÉRALISÉE (Annexe 1)). *Soit g est une fonction monotone d'un intervalle fini ou infini $]a, b[$ dans un intervalle fini ou infini $]c, d[$. Son inverse généralisée $g^{\leftarrow} :]c, d[\rightarrow]a, b[$ est définie par $g^{\leftarrow}(y) = \inf\{x : a < x < b, g(x) \geq y\}$.*

Lemme A.2. *Soit f_n une suite de fonctions croissantes et g une fonction croissante. Si pour tout x appartenant à un certain interval $]a, b[$ point de continuité de g ,*

$$\lim_{n \rightarrow \infty} f_n(x) = g(x),$$

alors

$$\lim_{n \rightarrow \infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x)$$

en tout point de continuité $x \in]g(a), g(b)[$ de g^{\leftarrow} .

Lemme A.3 (Vervaat (1971)). *Soit g une fonction continue et f_n des fonctions croissantes définie sur un intervalle $[a, b]$, et f définie sur $[a, b]$ avec $f' > 0$. Si $\delta_n > 0$ est une suite telle que $\delta_n \xrightarrow[n \rightarrow \infty]{} 0$ et*

$$\lim_{n \rightarrow \infty} \frac{f_n(x) - f(x)}{\delta_n} = g(x)$$

uniformément sur $[a, b]$. Alors

$$\lim_{n \rightarrow \infty} \frac{f_n^{\leftarrow}(x) - f^{\leftarrow}(x)}{\delta_n} = -(f^{\leftarrow})'(x)g(f^{\leftarrow}(x))$$

uniformément sur $[g(a), g(b)]$.

2 Théorème de convergence des types

Une preuve de ce théorème se trouve dans (Leadbetter et al., 1983).

Théorème A.4 (Khintchine). *Soit (F_n) une suite de fonctions de répartition et G une fonction de répartition non-dégénérée. S'il existe des constantes $a_n > 0$ et b_n telles que*

$$F_n(a_n x + b_n) \xrightarrow{d} G(x),$$

alors il existe une fonction de répartition G_ et des constantes $\alpha_n > 0$ et β_n telles que*

$$F_n(\alpha_n x + \beta_n) \xrightarrow{d} G_*(x)$$

si et seulement si

$$a_n^{-1} \alpha_n \xrightarrow{n \rightarrow \infty} a \quad \text{et} \quad a_n^{-1} (\beta_n - b_n) \xrightarrow{n \rightarrow \infty} b$$

où $a > 0$ et b sont deux constantes, et

$$G_*(x) = G(ax + b).$$

3 Notion de variation régulière

Cette partie est basée sur le livre de (Bingham et al., 1987), l'annexe B de (de Haan & Ferreira, 2006) et le chapitre 2 de (Resnick, 2007).

Les fonctions de survie et leurs inverses rencontrées dans l'étude des extrêmes possèdent des propriétés de variation régulière ou de variation régulière étendue. La théorie de la variation régulière est une branche de l'analyse réelle (Bingham et al., 1987).

La théorie classique de la variation régulière, dite de Karamata, part de la notion suivante.

Définition A.5 (VARIATION RÉGULIÈRE). *Une fonction f mesurable sur \mathbb{R}_+ est dite à variation régulière (au voisinage de l'infini) si*

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)}$$

existe pour un ensemble de valeurs de $x > 0$ de mesure de Lebesgue positive.

Théorème A.6. *Si une fonction f est à variation régulière (au voisinage de ∞) alors il existe un réel α tel que, pour tout $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha,$$

On dit alors que f est à variation régulière d'indice α , ce que l'on note $f \in \text{RV}_\alpha$. et si $\alpha = 0$, f est dite à variation lente.

Remarque A.7. Les fonctions auxiliaires utilisées pour la renormalisation dans les théorèmes 1.20 et 1.21 sont des fonctions à variation régulière.

La propriété qui convient le mieux à l'étude de la max-stabilité est la notion introduite par de Haan dans sa thèse vers 1970.

Définition A.8 (VARIATION RÉGULIÈRE ÉTENDUE). Une fonction $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ mesurable est dite à variation régulière étendue d'indice α , ce que l'on note $f \in \text{ERV}_\alpha$ s'il existe une fonction auxiliaire positive a telle que

$$\lim_{t \rightarrow \infty} \frac{f(tx) - f(t)}{a(t)} = \frac{x^\alpha - 1}{\alpha}.$$

La représentation de Karamata est une représentation générale pour les fonctions à variation régulière.

Théorème A.9 (REPRÉSENTATION DE KARAMATA (Resnick, 2007)). Une fonction $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ est à variation régulière d'indice α , si et seulement s'il existe deux fonctions mesurables $c: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ et $\eta: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ telles que, pour tout $t > 0$,

$$f(t) = c(t)t^\alpha \exp \left\{ \int_0^t \frac{\eta(t)}{t} dt \right\}. \quad (\text{A.1})$$

avec

$$\lim_{t \rightarrow \infty} c(t) = c > 0 \quad \text{et} \quad \lim_{t \rightarrow \infty} \eta(t) = 0.$$

De plus, si $f \in \text{RV}_\alpha$ satisfait la condition de von Mises, alors la fonction c dans la représentation (A.1) est constante.

Remark A.10. La fonction $[\log t]$ est à variation lente, mais pas à variation régulière étendue.

Théorème A.11. (UNIFORMITÉ DE LA CONVERGENCE SUR LES COMPACTS.) Si f est à variation régulière d'indice γ , alors pour tout $0 < a < b < \infty$,

$$\lim_{t \rightarrow \infty} \sup_{x \in [a, b]} \left| \frac{f(tx)}{f(t)} - x^\gamma \right| = 0.$$

On peut trouver une preuve lisible de ce théorème dans (Bingham et al., 1987) ou dans (de Haan & Ferreira, 2006). On peut démontrer la convergence localement uniforme à moindre frais si on suppose les fonctions croissantes (Resnick, 2007, Chapitre 2).

Théorème A.12. Si $I \subset \mathbb{R}$ est compact et $(f_n)_n$ est une suite de fonctions continues croissantes sur I qui converge simplement vers une fonction continue f , alors la convergence est uniforme sur I .

Exemple A.13. Dans le contexte du Théorème A.12, montrer que si de plus $\lim_{t \rightarrow \infty} f(t) < \infty$, alors la convergence est uniforme sur les intervalles de la forme $[x, \infty)$.

Théorème A.14 (CONDITION DE VARIATION RÉGULIÈRE AU SECOND ORDRE). Si pour une fonction mesurable f et deux fonctions positives a et A la limite

$$\lim_{t \rightarrow \infty} \frac{\frac{f(tx) - f(t)}{a(t)} - \frac{x^\alpha - 1}{\alpha}}{A(t)}$$

existe pour tout $x > 0$, et qu'elle n'est pas un multiple de $\frac{x^\alpha - 1}{\alpha}$, alors il existe des constantes c_1, c_2 et un paramètre $\rho \leq 0$ telles que pour tout $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{\frac{f(tx) - f(t)}{a(t)} - \frac{x^\alpha - 1}{\alpha}}{A(t)} = c_1 \int_1^x s^{\alpha-1} \int_1^s u^{\rho-1} du ds + c_2 \int_1^x s^{\alpha+\rho-1} ds.$$

De plus, pour $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{\frac{a(tx)}{a(t)} - x^\alpha}{A(t)} = c_1 x^\alpha \frac{x^\rho}{\rho}$$

and

$$\lim_{t \rightarrow \infty} \frac{A(tx)}{A(t)} = x^\rho.$$

On dit alors que f satisfait une condition de variation régulière du second ordre d'indices α et ρ .

Proposition A.15 (INÉGALITÉS DE POTTER). Soit $f \in \text{RV}_\alpha$, alors pour tous $\varepsilon, \delta > 0$, il existe $t_0 = t_0(\varepsilon, \delta)$ tel que pour tous $t \geq t_0$, $tx \geq t_0$

1. (Potter, 1942)

$$(1 - \varepsilon)x^\alpha \min(x^\delta, x^{-\delta}) < \frac{f(tx)}{f(t)} < (1 + \varepsilon)x^\alpha \max(x^\delta, x^{-\delta})$$

2. (Drees, 1998b)

$$\left| \frac{f(tx)}{f(t)} - x^\alpha \right| \leq \varepsilon \max(x^{\alpha+\delta}, x^{\alpha-\delta})$$

3. (Drees, 1998b)

$$\left| \frac{f(tx) - f(t)}{a_0(t)} - \frac{x^\alpha - 1}{\alpha} \right| \leq \varepsilon x^\alpha \max(x^\delta, x^{-\delta})$$

avec

$$a_0(t) = \begin{cases} \alpha f(t) & \gamma > 0 \\ -\alpha(f(\infty) - f(t)) & \gamma < 0 \\ f(t) - t^{-1} \int_0^t f(s) dx & \gamma = 0. \end{cases}$$

Annexe B

Méthode de Lepski

Dans cette annexe, nous présentons l'idée générale de la méthode dite de Lepski (1990), dans le cadre de l'estimation de densité, pour plus de détails le lecteur peut se référer à l'introduction de la thèse (section 1.5.2) de Chichignoud (2010). Dans sa version originale, la méthode de Lepski permet de sélectionner la fenêtre dans les problèmes d'estimation de densité de façon adaptative, c'est-à-dire uniquement à partir des quantités connues du statisticien et presque aussi efficacement que si l'on connaissait la régularité de la densité cible, parmi une grille finie. Puis, Lepski (1991) développe une procédure plus générale pour n'importe quel estimateur d'une fonction.

On suppose donnée une collection $(\widehat{s}_b)_{b \in \mathcal{B}}$ d'estimateurs de la densité $s \in L^2(\mathbb{R})$ dépendant d'un paramètre b , inconnu du statisticien. Par exemple, b est l'indice du modèle pour les estimateurs par projecteur et la fenêtre pour les estimateurs à noyaux. Le risque quadratique d'un estimateur \widehat{s}_b est donné par :

$$\mathbb{E} [|\widehat{s}_b - s|^2] = \mathbb{E} [|\widehat{s}_b - \mathbb{E}\widehat{s}_b|^2] + |\mathbb{E}\widehat{s}_b - s|^2$$

Le premier terme correspond à un terme de variance et le second à terme de biais. On suppose également construit un estimateur \widehat{s}_{b^*} où b^* est l'indice qui permet d'équilibrer le compromis biais-variance, c'est-à-dire qui minimise le risque quadratique. Le plus souvent, cet estimateur dépend de quantités inconnues du statisticien comme par exemple la régularité de la densité s .

La procédure de Lepski commence par la construction d'une grille, souvent géométrique, sur l'ensemble \mathcal{B} , soit

$$b_k = \theta^k, k = k_m, \dots, k_M,$$

où $\theta > 1$, $k_m \in \mathbb{N}$ et $k_M \in \mathbb{N}$ sont des paramètres à calibrer. L'estimateur adaptatif est alors à choisir dans la famille $\{\widehat{s}_{b_k}, k_m \leq k \leq k_M\}$.

Le but de l'adaptation est de choisir \widehat{b} , fonction des données, de telle sorte que l'estimateur $\widehat{s}_{\widehat{b}}$ imite \widehat{s}_{b^*} , c'est-à-dire que son risque doit être aussi proche que possible que celui de b^* .

La règle de sélection générale peut être présentée de la façon suivante :

$$\widehat{k} = \inf\{k = k_m, \dots, k_M : \text{pour tout } k' > k, |\widehat{s}_{b_k} - \widehat{s}_{b_{k'}}| \leq C\sigma(b_{k'})\}$$

où $\sigma(b_k)$ est une borne de l'écart-type de \widehat{s}_{b_k} . En général, la variance de notre estimateur est croissante en k tandis que le biais est décroissant. Cette règle peut s'interpréter comme une procédure de tests multiples : pour chaque k compris entre k_m et k_M , on teste si l'écart avec tous les estimateurs $\widehat{s}_{b_{k'}}$ est de l'ordre de l'écart-type de \widehat{s}_{b_k} . On garde le plus petit indice k qui passe tous les tests.

Bibliographie

- Réseau sentinelles. Available from : <http://websenti.u707.fr/sentiweb/>, March 2014.
- K. Alam & K. M. L. Saxena. Positive dependence in multivariate distributions. *Communications in Statistics*, 10 :1183–1196, 1981.
- C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, & G. Schef-fer. *Sur les inégalités de Sobolev logarithmiques. (On logarithmic Sobolev inequalities. With a preface of Dominique Bakry and Michel Ledoux)*. Panoramas et Synthèses. 10. Paris : Société Mathématique de France. xiii, 217 p., 2000.
- B. C. Arnold, N. Balakrishnan, & H. N. Nagaraja. *A first course in order statistics*, volume 54 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.
- A. A. Balkema & L. de Haan. Residual life time at great age. *The Annals of Probability*, 2(5) :792–804, 1974.
- A. A. Balkema & L. de Haan. Limit distributions for order statistics I. *Theory of Probability and its Applications*, 23 :77–92, 1978a.
- A. A. Balkema & L. de Haan. Limit distributions for order statistics II. *Theory of Probability and its Applications*, 23 :341–358, 1978b.
- J. Beirlant, P. Vynckier, & J. L. Teugels. Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association*, 91(436) : 1659–1667, 1996.
- J. Beirlant, Y. Goegebeur, J. Segers, & J. Teugels. *Statistics of extremes. Theory and applications*. Wiley, 2004.
- J. Beirlant, C. Bouquiaux, & B. Werker. Semiparametric lower bounds for tail index estimation. *Journal of Statistical Planning and Inference*, 136(3) :705–729, 2006.
- G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57 :33–45, 1962.
- S. Bernstein. *The Theory of Probabilities*. Gastehizdat Publishing House, Moscow, 1946.

- N. Bingham, C. Goldie, & J. Teugels. *Regular variation*. Cambridge University Press, 1987.
- L. Birgé. An alternative point of view on Lepski's method. In *State of the art in probability and statistics (Leiden, 1999)*, volume 36 of *IMS Lecture Notes Monograph Series*, pages 113–133. Institute of Mathematical Statistics Math. Statist., Beachwood, OH, 2001.
- L. Birgé. A new lower bound for multiple hypothesis testing. *IEEE Trans. Inform. Theory*, 51 :1611–1615, 2005.
- L. Birgé & P. Massart. Minimum constraint estimators on sieves : exponential bounds and rates of convergence. *Bernoulli*, 4 :329–375, 1998.
- S. Bobkov & M. Ledoux. Poincaré's inequality and talagrand's concentration phenomenon for the exponential distribution. *Probability Theory and Related Fields*, 107 :383–400, 1997.
- S. Boucheron & M. Thomas. Concentration inequalities for order statistics. *Electronic Communications in Probability*, 17 :1–12, 2012.
- S. Boucheron & M. Thomas. Tail index estimation, concentration and adaptivity. *Submitted at Electronic Journal of Statistics*, 2015.
- S. Boucheron, G. Lugosi, & P. Massart. Concentration inequalities using the entropy method. *Annals of Probability*, 31(3) :1583–1614, 2003.
- S. Boucheron, O. Bousquet, & G. Lugosi. *Machine Learning Summer School 2003*, volume 3176 of *LNAI*, chapter Concentration inequalities, pages 169–207. Springer-Verlag, 2004.
- S. Boucheron, G. Lugosi, & P. Massart. *Concentration inequalities*. Oxford University Press, 2013.
- O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus de l'Académie des Sciences de Paris*, 334 :495–500, 2002.
- J. Bresee & F. Hayden. Epidemic influenza-responding to the expected but unpredictable. *The New England Journal of Medicine*, 368(7) :589–92, 2013.
- A. Carpentier & A. Kim. Adaptive and minimax optimal estimation of the tail coefficient. *Preprint available at <http://arxiv.org/pdf/1309.2585.pdf>*, 2014a.
- A. Carpentier & A. Kim. Adaptive confidence intervals for the tail coefficient in a wide second order class of pareto models. *Preprint available at <http://arxiv.org/pdf/1312.2968v2.pdf>*, 2014b.
- CépiDc. Centre d'épidémiologie sur les causes médicales de décès. *Available from : <http://www.cepidec.vesinet.insmer.fr/>*, March 2014.

- S. Chatterjee. *Superconcentration and related topics*. Springer, Cham, 2014.
- H. Chernoff. A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23 :493–507, 1953.
- D. Chibisov. On limit distributions for order statistics. *Theory of Probability and its Applications*, 9 :142–148, 1964.
- M. Chichignoud. *Performances statistiques d'estimateurs non-linéaires*. PhD thesis, Université de Provence, 2010.
- S. Coles. *An introduction to statistical modeling of extremes values*. Springer-Verlag, London, 2001.
- S. Coles & J. Tawn. Seasonal effects of extreme surges. *Stochastic Environmental Research and Risk Assessment*, 19 :417–427, 2005.
- T. Cover & J. Thomas. *Elements of Information Theory*. John Wiley, New York, 1991.
- C. Craig. On the Tchebychef inequality of Bernstein. *Annals of Mathematical Statistics*, 4 :94–102, 1933.
- H. Cramér. *Mathematical methods of statistics*. Princeton University Press, 1946.
- S. Csörgő, P. Deheuvels, & D. Mason. Kernel estimates of the tail index of a distribution. *Annals of Statistics*, 13(3) :1050–1077, 1985.
- J. Danielsson, L. de Haan, L. Peng, & C. G. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76 :226–248, 2001.
- D. Darling & P. Erdős. A limit theorem for the maximum of normalized sums of independent random variables. *Duke Mathematical Journal*, 23 :143–155, 1956.
- L. de Haan. On regular variation and its applications to the weak convergence of sample extremes. *Amsterdam, Mathematical Centre Tracts*, 32, 1970.
- L. de Haan & A. Ferreira. *Extreme value theory*. Springer-Verlag, 2006.
- A. L. M. Dekkers, J. H. J. Einmahl, & L. de Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17(4) :1833–1855, 1989.
- L. Denoued, C. Turbelin, S. Ansart, A. Valleron, F. A., & F. Carrat. Predicting pneumonia and influenza mortality data. *PLoS ONE*, 2(5) :4–64, 2007.
- C. Dombry. Maximum likelihood estimators for the extreme value index based on the block maxima method. *Arxiv preprint arXiv :1301.2611v1*, 2013.
- G. Draisma, L. de Haan, L. Peng, & T. Pereira. A bootstrap-based method to achieve optimally in estimating the extreme value index. *Extremes*, 2 :367–404, 1999.

- H. Drees. Optimal rates of convergence for estimates of the extreme value index. *Annals of Statistics*, 26(1) :434–448, 1998a.
- H. Drees. On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25 (1) :187–210, 1998b.
- H. Drees. Minimax risk bounds in extreme value theory. *Annals of Statistics*, 29(1) : 266–294, 2001.
- H. Drees & E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and Applications*, 75 :149 :172, 1998.
- H. Drees, L. De Haan, & S. Resnick. How to make a Hill plot. *Annals of Statistics*, 28 (1) :254–274, 2000.
- B. Efron & C. Stein. The jackknife estimate of variance. *Annals of Statistics*, 9(3) : 586–596, 1981.
- P. Embrechts, C. Klüpperberg, & T. Mikosch. *Modelling extremal events for insurance and finance*. Springer Verlag, Berlin, 1997.
- A. Falchi, C. Turbelin, L. Andreoletti, C. Arena, B. T., & I. Bonmarin. Nationwide surveillance of 18 respiratory viruses in patients with influenza-like illnesses : a pilot feasibility study in the french sentinel network. *Journal of Medical Virology*, 83(8) : 1451–1457, 2011.
- R. Fisher & L. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24 : 180–190, 1928.
- M. Fréchet. Sur la loi de probabilité de l'écart du maximum. *Annales de la Société Polonaise de Mathématique*, 6 :93–116, 1927.
- J. Geluk, L. de Haan, S. Resnick, & C. Stărică. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and Applications*, 69(2) :139–159, 1997.
- E. Giné & V. Koltchinskii. Concentration inequalities and asymptotic results for ratio type empirical processes. *Annals of Probability*, 34(3) :1143–1216, 2006.
- B. Gnedenko. Sur la distribution limite du terme maximal d'une série aléatoire. *Annals of Mathematics*, 44(3) :423–453, 1943.
- E. Goldstein, S. Cobey, S. Takahashi, J. Miller, & M. Lipsitch. Predicting the epidemic sizes of influenza A/H1N1, A/H2N3 and B : a statistical method. *PLoS Medicine*, 8 (7) :1001–1051, 2011.
- I. Grama & V. Spokoiny. Statistics of extremes by oracle estimation. *Annals of Statistics*, 36(4) :1619–1648, 2008.

- A. Guillou & Y. Kratz, M. Le Strat. An extreme value theory approach for the early detection of time clusters with application to the surveillance of salmonella. *Statistics Applications*, 2013.
- E. Gumbel. *Statistics of extremes*. Columbia University Press. New York, 1958.
- P. Hall. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32 :177–203, 1990.
- P. Hall & I. Weissman. On the estimation of extreme tail probabilities. *Annals of Statistics*, 25(3) :1311–1326, 1997.
- P. Hall & A. Welsh. Adaptive estimates of parameters of regular variation. *Annals of Statistics*, 13(1) :331–341, 1985.
- B. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3 :1163–1174, 1975.
- W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(13-30), 1963.
- INSEE. Institut national de la statistique et des études économiques. Available from : http://www.insee.fr/fr/ffc/docs_ffc/elp_reg_dep.htm, March 2014.
- K. Joag-Dev & F. Proschan. Negative association of random variables with applications. *Annals of Statistics*, 11 :286–253, 1983.
- R. Katz, M. Parlange, & P. Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25 :1287–1304, 2002.
- T. Klein. Une inégalité de concentration à gauche pour les processus empiriques. *Comptes Rendus de l'Académie des Sciences de Paris*, 334(6) :501–504, 2002.
- T. Klein & E. Rio. Concentration around the mean for maxima of empirical processes. *Annals of Probability*, 33(3) :1060–1077, 2005.
- V. Koltchinskii. *Oracle inequalities in empirical risk minimization and sparse recovery problems*. Ecole d'été de Probabilités de Saint-Flour 2008. Lecture Notes in Mathematics. Springer, 2008.
- P. Laplace. Deuxième supplément à la théorie analytique des probabilités. *Paris : Courcier*, 1818.
- M. Leadbetter, G. Lindgren, & H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin, 1983.

- M. Ledoux. A simple analytic proof of an inequality by P. Buser. *Proceedings of the American Mathematical Society*, 3 :951–959, 1994.
- M. Ledoux. On Talagrand’s deviation inequalities for product measures. *ESAIM Probability and Statistics*, 1 :63–87, 1997.
- M. Ledoux. *The concentration of measure phenomenon*. American Mathematical Society, 2001.
- M. Ledoux. *A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices*, volume 1982. Séminaire de probabilités XXXVII, Lecture Notes in Mathematics, Springer, Berlin, 2006.
- M. Ledoux & M. Talagrand. *Probability in Banach spaces*. Springer, 1991.
- O. Lepski. A problem of adaptive estimation in Gaussian white noise. *Teoriya Veroyatnostei i ee Primeneniya*, 35(3) :459–470, 1990.
- O. Lepski. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teoriya Veroyatnostei i ee Primeneniya*, 36(4) :645–659, 1991.
- O. Lepski. Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teoriya Veroyatnostei i ee Primeneniya*, 37(3) :468–481, 1992.
- O. Lepski & A. Tsybakov. Asymptotic exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, 117(1) :17–48, 2000.
- D. Mason. Laws of large numbers for sums of extreme values. *Annals of Probability*, 10 :754–764, 1982.
- P. Massart. About the constants in Talagrand’s concentration inequality. *Annals of Probability*, 28 :863–885, 2000.
- P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, 2006.
- P. Mathé. The Lepski principle revisited. *Inverse Problems*, 22(3) :L11–L15, 2006.
- B. Maurey. Construction de suites symétriques. *Comptes Rendus de l’Académie des Sciences de Paris Série I Mathématiques*, 288 :679–681, 1979.
- B. Maurey. Some deviation inequalities. *Geometric and Functional Analysis*, 1(2) :188–197, 1991.
- R. Miller. The jackknife : a review. *Biometrika*, 61 :1–15, 1974.
- V. Milman & G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*. Springer-Verlag, New York, 1986.

- S. Novak. Lower bounds to the accuracy of inference on heavy tails. *Bernoulli*, 20(2) : 979–989, 2014.
- M. Okamoto. Some inequalities relating to the partial sum of binomial probabilities. *Annals of the Institute of Statistical Mathematics*, 10 :29–35, 1958.
- J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3 : 119–131, 1975.
- G. Pisier. Some applications of the metric entropy condition to harmonic analysis. In *In Banach spaces, Harmonic analysis and Probability*, volume 995 of *Lecture Notes in Mathematics*, pages 123–159. University of Connecticut 1980-81. Springer, Berlin, Heidelberg, 1983.
- H. Potter. The mean value of a Dirichlet series II. *Proceedings of the London Mathematical Society*, 47 :1–19, 1942.
- R. Reiss. *Approximate distributions of order statistics. With applications to nonparametric statistics*. Springer Series in Statistics. Springer-Verlag. New York, 1989.
- A. Rényi. On the theory of order statistics. *Acta Mathematica Scientia*, 4 :191–227, 1953.
- S. Resnick. *Heavy-tail phenomena : probabilistic and statistical modeling*, volume 10. Springer Verlag, 2007.
- S. Resnick & C. Stărică. Smoothing the hill estimator. *Advances in Applied Probability*, 29 :271–293, 1997.
- E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probability Theory and Related Fields*, 119 :163–175, 2001.
- J. Segers. Abelian and Tauberian theorems on the bias of the hill estimator. *Scandinavian Journal of Statistics*, 29 :461–483, 2002.
- P. Sen. Asymptotic normality of sample quantiles of m -dependent sequences. *Annals of Mathematical Statistics*, 39 :1724–1730, 1968.
- J. Shao & C. Wu. A general theory for jackknife variance estimation. *The Annals of Statistics*, 17 :1176–1197, 1989.
- L. Simonsen, M. Clarke, G. Williamson, D. Stroup, N. Arden, & L. Schonberger. The impact of influenza epidemics on mortality : introducing a severity index. *American Journal of Public Health*, 87(12) :1944–1950, 1997.
- N. Smirnov. Über die verteilung des allgemeinen gliedes in der variationsreihe. *Metron*, 12 :59–81, 1935.
- N. Smirnov. Limit distributions for the terms of a variational series. *American Mathematical Society Translations*, 1952(67), 1952.

- N. Smirnov. Some remarks on limits laws for order statistics. *Theory of Probability and its Applications*, 12 :196–205, 1967.
- J. M. Steele. An Efron-Stein inequality for nonsymmetric statistics. *The Annals of Statistics*, 14 :753–758, 1986.
- M. Talagrand. Regularity of Gaussian processes. *Acta Mathematica Scientia*, 159 :99–149, 1987.
- M. Talagrand. An isoperimetric inequality on the cube and Khintchine-Kahane inequalities in product spaces. *Proceedings of the American Mathematical Society*, 104 :905–909, 1988.
- M. Talagrand. A new isoperimetric inequality and the concentration of measure phenomenon. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Mathematics*, pages 94–124. Springer, Berlin, 1991.
- M. Talagrand. Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and margulis’ graph connectivity theorem. *Geometric And Functional Analysis*, 3 :296–314, 1993.
- M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Annals of Probability*, 22 :28–76, 1994.
- M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l’I.H.E.S.*, 81 :73–205, 1995.
- M. Talagrand. New concentration inequalities in product spaces. *Inventiones Mathematicae*, 126 :505–563, 1996a.
- M. Talagrand. A new look at independence. *Annals of Probability*, 24 :1–34, 1996b.
- M. Talagrand. *The generic chaining*. Springer, New York, 2005.
- J.-P. Tillich & G. Zémor. Discrete isoperimetric inequalities and the probability of a decoding error. *Combinatorics, Probability and Computing*, 9(05) :465–479, 2001.
- A. B. Tsybakov. Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Annals of Statistics*, 26(6) :2420–2469, 1998.
- J. Uspensky. *Introduction to Mathematical Probability*. McGraw-Hill, New-York, 1937.
- S. A. van de Geer. *Applications of empirical process theory*. Cambridge University Press, 2000.
- A. van der Vaart. *Asymptotic statistics*. Cambridge University Press, 1998.
- H. van der Vaart. A simple derivation of the limiting distribution function of a sample quantile with increasing sample size. *Statistica Neerlandica*, 15 :239–242, 1961.

- W. Vervaat. Functional limit theorems for processes with positive drift and their inverses. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 2 :245–253, 1971.
- R. von Mises. La distribution de la plus grande de n valeurs. *Review Mathématique Union Interbalcanique*, 1 :141–160, 1936.
- H. Wickham. *ggplot2 : elegant graphics for data analysis*. Springer Science & Business Media, 2009.
- H. Wickham. *Advanced R*. Chapman & Hall/CRC, 2014.
- L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probability Theory and Related Fields*, 118(3) :427–438, 2000.