



HAL
open science

Régression quantile extrême : une approche par couplage et distance de Wasserstein.

Benjamin Bobbia

► **To cite this version:**

Benjamin Bobbia. Régression quantile extrême : une approche par couplage et distance de Wasserstein.. Statistiques [math.ST]. Université Bourgogne Franche-Comté, 2020. Français. NNT : 2020UBFCD043 . tel-03156028

HAL Id: tel-03156028

<https://theses.hal.science/tel-03156028>

Submitted on 2 Mar 2021

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.

Regression quantile extrême : une approche par couplage et distance de Wasserstein.

Thèse de doctorat de
l'Université Bourgogne Franche-Comté

École doctorale CARNOT-PASTEUR

présentée et soutenue publiquement le 11 Décembre 2020
en vue de l'obtention du grade de

Docteur de l'Université Bourgogne Franche-Comté
(mention Mathématiques Appliquées et Applications des Mathématiques)

par

Benjamin BOBBIA

Composition du jury :

<i>Rapporteurs :</i>	M. Philippe BERTHET M. Olivier WINTENBERGER	Université Paul Sabatier Toulouse Sorbonne Université
<i>Examineurs :</i>	Mme. Anne-Laure FOUGÈRES	Université Claude Bernard Lyon 1
<i>Directeurs de thèse :</i>	Mme. Camelia GOGA M. Clément DOMBRY M. Davit VARRON	Université Bourgogne Franche-Comté Université Bourgogne Franche-Comté Université Bourgogne Franche-Comté

(Lm^B)

Remerciements

Je tiens à remercier chaleureusement Anne-Laure, Camelia, Philippe et Olivier pour avoir accepté de donner de leur temps en cette période troublée pour faire partie de mon jury et d'avoir pris le temps d'évaluer mon travail. Merci à vous pour les conseils que vous m'avez donnés lors de nos rencontres.

Je remercie énormément mes directeurs pour ces trois ans, pour tout ce qu'ils m'ont enseigné ainsi que pour leur soutien indéfectible. Merci d'avoir été si présents (jusqu'à venir me chercher le matin de la soutenance), d'avoir su me guider, m'encourager et même me secouer quand il a été nécessaire. Merci aussi d'avoir investi dans les litres d'encre rouge nécessaire aux corrections de ce que je vous donnais à relire.

Merci également aux membres exceptionnels qui ont partagé le bureau 401 avec moi, Marine, Thao, Julien, Colin, Youssef, Francesco et qui y ont fait régner une ambiance de folie. Merci plus généralement à tous les camarades doctorants, Loris, Lucie, Mehdi, Yoël, Raphaël, Isabelle, Valentin, Cécile, Charles, Mathieu, Audrey (et j'en oublie certainement) pour tous les super moments conviviaux au labo. Parmi eux, je remercie particulièrement Thao et son énergie, sa bonne humeur permanente et sa soif insatiable de discuter. Je remercie également tout particulièrement Marine et Lucie pour leur support administratif et logistique sans faille, et sans qui je me serais perdu dans les papiers et rendu nombre de choses en retard, vous avez été géniales. Je remercie également ceux présents depuis plus longtemps, Loris pour les soirées sur ta terrasse, Guillaume pour m'avoir si généreusement hébergé pour ma venue à Caen lors de mon premier séminaire loin de Besançon. Merci aussi à Nico, avec qui j'ai partagé de nombreuses heures de travail acharné durant le temps de midi. Bon, on va aussi nommer Alexis parce que ça lui fera plaisir d'être ici. Encore plus généralement, je tiens à remercier tout les membres du laboratoire pour m'avoir accueilli et m'avoir permis de passer des magnifiques années en ces murs.

Je remercie également tous les copains, François, Mariana, Alex, Gat, et tous les autres pour les encouragements et les heures de rires en parallèle de la fac. Un merci particulier à Gat, grande amie et colocataire, pour ses super Coxinhas de soutenance.

Un grand merci également à toute ma famille, à mes parents (les 4) qui m'ont soutenu depuis le début et ont participé grandement à tout ce qui m'arrive aujourd'hui. Merci à Claude, Marie, Sacha et Emy pour ces mois de confinement à Corre, c'était quand même vachement bien.

Un merci aussi à mes autres professeurs, François, Marc, Franck(s), Kevin et Selma, qui m'ont également énormément appris pendant mes années d'études.

Bien sûr, la meilleure pour la fin, je remercie celle sans laquelle mes suites de Cauchy ne convergeraient pas, celle qui partage ma vie, celle qui a su supporter mes sautes d'humeur quand "ça marchait pas", celle qui m'a encouragé tous les jours, jusqu'à recouvrir la maison de mots d'encouragement le matin de la soutenance, Merci petit Rachon ♡.

Pour finir, merci à tous ceux venus me soutenir le jour J.

Merci à tous.

Table des matières

1	Introduction	7
I	Le terrain de jeux	13
2	Fondamentaux de la théorie des valeurs extrêmes	15
2.1	Comportement des extrêmes	16
2.2	Estimation des paramètres	25
2.3	Estimation de quantiles extrêmes	28
2.4	Extrêmes Hétéroscédastiques	30
3	Boîte à outils	33
3.1	Convergence faible	33
3.2	Processus empiriques	39
3.3	Bootstrap et processus empiriques	42
3.4	Variations régulières	44
4	Couplage optimal	49
4.1	Couplage optimal et distance de Wasserstein	49
4.2	Applications aux statistiques	53
II	Contributions	55
5	Extreme quantile regression in proportional tail framework	57
5.1	Introduction and main results	57
5.1.1	Framework	57
5.1.2	The proportional tail model	58
5.1.3	Extreme quantile regression	61
5.2	A coupling approach	63
5.3	Proofs	66
5.3.1	Proof of Theorem 7.1	66

5.3.2	Proof of theorem 5.2	74
5.3.3	Proof of theorem 5.3	77
5.4	Appendix	79
6	The coupling method in extreme value theory	81
6.1	Introduction	81
6.2	Coupling and sampling	83
6.2.1	Background on coupling and Wasserstein spaces	83
6.2.2	Wasserstein distance between empirical distributions	85
6.2.3	Wasserstein distance between sample statistics	86
6.3	Wasserstein distance estimates in extreme value theory	86
6.3.1	Background on univariate extreme value theory	86
6.3.2	Analysis of the Peak-over-Threshold method	88
6.3.3	Extensions	92
6.4	Proofs	96
6.4.1	Proofs related to Section 6.2	96
6.4.2	Proofs related to Section 6.3	99
6.5	Supplementary material	104
7	Analysis of proportional tail model with optimal coupling.	109
7.1	Analysis of a proportional tail model	109
7.1.1	The proportional tail model	109
7.1.2	Parameter estimation	110
7.1.3	Model validation	111
7.1.4	Coupling arguments for the proof	114
7.1.5	Approximation of empirical measures	115
7.1.6	Bootstrap	116
7.2	Simulation study	116
7.2.1	Level control	116
7.2.2	Power study	118
7.3	Proofs	122
7.3.1	Proofs related to section 7.1.4	122
7.3.2	Proofs related to section 7.1.2	128
7.3.3	Idea of the proof of Conjecture 7.1	130
7.3.4	Bootstrap and proof of Theorem 7.3	134
7.4	Appendix	136
7.4.1	About proportional tail model	136
7.4.2	About maximal coupling	138
	Perspectives	140
	Bibliographie	141

Un peu de contexte

Dans le contexte actuel de globalisation, la notion de crise semble être un thème récurrent. Ces crises, qu'elles soient d'origine naturelle (tsunami, inondation, tremblement de terre...) ou humaine (chute d'un indice boursier, accident nucléaire...) ont toutes en commun un impact important sur des populations. Il est donc naturel de vouloir comprendre et prévoir de tels phénomènes. Ces événements ont pour autre point commun d'être rares, au sens de survenir avec une faible probabilité, et par conséquent sont peu observés. La modélisation de tels événements est ainsi rendue difficile par la faible fréquence, voire l'absence, d'observation.

Ce dernier point rend leur étude complexe car, là où les statistiques liées à la loi des grands nombres et au théorème central limite arrivent à tirer de l'information des nombreuses observations autour de la moyenne et ainsi arrivent à rendre compte des événements courants, elles échouent à comprendre le comportement des événements rares et éloignés de la moyenne. Il paraît alors nécessaire de développer une nouvelle approche pour mieux comprendre ces événements, rares et intenses, que l'on appelle extrêmes.

Cette théorie des valeurs extrêmes se concentre, non pas sur l'étude de la moyenne empirique, mais sur l'étude du maximum d'un échantillon. Similairement à la théorie normale avec son théorème central limite, elle met en évidence une loi, ou plutôt une famille de lois qui permettent d'approcher le comportement des extrêmes. Ces lois, qui modélisent les queues de distribution, permettent l'extrapolation au-delà des données observées, et ainsi d'entrevoir les valeurs extrêmes, ainsi que leurs fréquences, sans la nécessité d'en avoir été témoin. Certains attribuent les prémices

de cette théorie à N. Bernoulli au début du XVIII^{ème} siècle [62] mais elle ne sera formalisée que deux siècles plus tard par Fisher et Tippett [41], Gnedenko [45] ou encore Gumbel [49]. De nos jours, la théorie des valeurs extrêmes connaît de nombreuses applications. Bien que, suivant les travaux de Gumbel [48], les extrêmes soient historiquement très liés à l'hydrologie ([47], [21], [67]) ils sont également mis en application dans d'autres domaines comme en climatologie [76], en météorologie [44], actuariat [73] ou en finance [39]. Ce dernier point est aussi une conséquence de l'investigation du lien entre extrêmes et séries temporelles [74]. La théorie des valeurs extrêmes ne s'applique pas uniquement dans le cadre de la prédiction de

catastrophes, Einmahl et Magnus se sont intéressés aux limites des records en athlétisme [36] alors que Einmahl *et al.* ont cherché à explorer les limites de la longévité humaine [37]. On pourra se référer à la monographie [72] pour plus d'exemples.

Parmi toutes ces applications, certaines se concentrent sur l'estimation de quantiles [79]. C'est-à-dire si on se donne un niveau de risque α , par exemple $\alpha = 1\%$, quelle serait la valeur qui correspondrait à une fréquence de dépassement de 1% ? Lorsque le niveau de risque est petit, plus petit que l'inverse du nombre de données, ce quantile est dit extrême. C'est dans ce contexte que s'inscrit le présent manuscrit.

Principaux objectifs

Un des enjeux de cette théorie est l'estimation de quantiles extrêmes. C'est-à-dire, considérons une variable $Y \in \mathbb{R}$ pouvant représenter le niveau d'un cours d'eau, le prix d'une action, la pression dans un réacteur, et dépendant de plusieurs covariables représentées par $X \in \mathbb{R}^d$ qui peuvent être autant de propriété topologique du terrain, ou de détérioration de matériel. Notre objectif est d'estimer, pour une configuration des covariables $x \in \mathbb{R}^d$, le quantile conditionnel q qui est défini par

$$\mathbb{P}(Y > q | X = x) = \alpha$$

avec α très petit (tendant vers 0).

Dans le présent manuscrit, nous étudions un modèle permettant d'estimer cette quantité q . Ce modèle est appelé *modèle des queues proportionnelles* et revient essentiellement à faire deux hypothèses sur F et F_x qui sont respectivement les fonctions de répartition de Y et de Y sachant $X = x$.

La première hypothèse est celle qui donne son nom au modèle.

Hypothèse 1 :

$$\lim_{y \rightarrow \infty} \frac{1 - F_x(y)}{1 - F(y)} = \sigma(x) \quad \text{uniformément en } x \in \mathbb{R}^d.$$

La seconde est une hypothèse de variation régulière provenant directement du cadre d'étude de la théorie des valeurs extrêmes.

Hypothèse 2 :

$$\lim_{t \rightarrow \infty} \frac{F^{\leftarrow} \left(1 - \frac{1}{tu}\right)}{F^{\leftarrow} \left(1 - \frac{1}{u}\right)} = u^\gamma, \quad u > 0,$$

avec F^{\leftarrow} l'inverse généralisée de F et $\gamma > 0$.

Les travaux présentés ici concernent l'étude d'estimateurs de γ , σ , appelé fonction *skedasis*, et également de sa version intégrée. Ensemble, ces estimations nous permettent de construire et d'étudier un estimateur du quantile extrême conditionnel.

Ces estimateurs permettent également de fournir un test de validation pour notre modèle, dont le niveau et la puissance seront étudiés empiriquement par simulation.

Une approche par couplage

La méthode que nous utilisons pour de nombreuses preuves est une approche par couplage. On appelle couplage entre deux mesures de probabilité P et Q tout couple (X_P, X_Q) tel que $X_P \sim P$ et $X_Q \sim Q$.

La définition du modèle des queues proportionnelles nous amène à considérer un modèle limite. C'est un modèle "idéal", qui satisfait les Hypothèses 1 et 2 pour tout $y > 0$ et $t > 0$, et non pas uniquement asymptotiquement. L'idée du couplage réside dans le fait que le modèle limite est plus simple à étudier et, si les deux modèles sont assez proches, ils pourraient avoir un comportement similaire. Cette méthodologie consistera alors en l'établissement d'un couplage entre la loi P de (X, Y) et la loi P^* du modèle limite. Ainsi l'étude du modèle des queues proportionnelles sera fait en deux étapes :

1. Les résultats sont d'abord établis pour le modèle limite à l'aide de méthodes liées au processus empirique. Nous pouvons alors établir, par exemple, la normalité asymptotique des estimateurs pour le modèle limite en exhibant des classes de Donsker.
2. Les propriétés que l'on souhaite établir pour le modèle des queues proportionnelles sont alors transférées du modèle limite à notre modèle en construisant un couplage satisfaisant de "bonnes propriétés de proximité". Cette notion, floue à ce moment de lecture, de "bonne proximité" est un des sujets importants abordé en parcourant ces pages.

Ainsi la notion de distance entre P et P^* est centrale dans cette approche. Dans un premier temps, nous avons créé explicitement un couplage $(X, Y) \sim P$ (satisfaisant les hypothèses du modèle des queues proportionnelles) et $(X^*, Y^*) \sim P^*$ proche au sens où les quantités $\mathbb{P}(X \neq X^*)$ et $\sup_{y>1} |\mathbb{P}(Y > y) - \mathbb{P}(Y^* > y)|$ convergent vers 0 de façon contrôlée. Avec cette construction nous avons pu montrer la convergence de l'estimateur d'un quantile extrême. Cette approche est très liée aux processus empiriques. En effet, la proximité du couplage nous permet de construire deux processus empiriques (un sur chaque échantillon) qui seront eux-mêmes assez proches pour converger vers la même limite. Ceci a cependant été fait au prix d'une hypothèse légèrement plus restrictive que des hypothèses plus classiques.

C'est pourquoi, dans un second temps, nous avons porté notre intérêt sur le couplage optimal. Cette notion provient de l'aspect probabiliste de la théorie du transport optimal, un couplage optimal étant un couplage minimisant la distance L^p entre les variables. En d'autres termes, cela revient à considérer la distance de Wasserstein entre P et P^* . Les hypothèses du modèle des queues proportionnelles permettent de déduire des bornes pour ce type de distance.

Dans le cadre de cette approche par couplage, nous avons étudié des propriétés générales des distances de Wasserstein, et particulièrement des distances entre mesures empiriques. Ces propriétés sont réparties entre les chapitres 6 et le chapitre 7, ce dernier contient les résultats pour la distance entre mesures empiriques

pondérées conditionnées par les données utiles dans l'étude du bootstrap. Bien que nous appliquions ces résultats dans le contexte des extrêmes, ce sont des résultats généraux qui ne sont pas cloisonnés à cette théorie. Ceci laisse espérer que cette approche soit applicable dans d'autres champs des statistiques.

Plan détaillé de la thèse

Ce manuscrit s'organise en six chapitres répartis en deux parties. La première partie contient le contexte détaillé des extrêmes dans lequel évolue cette thèse ainsi que les outils utiles à sa rédaction. Dans la seconde partie sont présentées les contributions.

- Le premier chapitre présente la théorie des valeurs extrêmes univariées réelles. Il contient l'étude de la distribution du maximum d'un échantillon ainsi que celle des excès au-dessus d'un seuil et les caractérisations des distributions limites et leurs domaines d'attractions. Ensuite, deux estimateurs de l'indice des valeurs extrêmes sont introduits, l'estimateur de Hill et l'estimateur des moments pondérés (Probability Weighted Moments). Naturellement, étant nécessaires à la construction des précédents estimateurs, les statistiques d'ordre sont introduites et également utilisées pour définir des estimateurs des quantiles, l'estimateur de Weissmann et l'estimateur GPD (pour Generalized Pareto Distribution). Ce chapitre se termine sur une présentation du modèle des *extrêmes hétéroscédastiques*, modèle qui a servi d'exemple pour définir le modèle des queues proportionnelles.
- Le second chapitre contient les outils que nous utilisons pour l'étude du modèle limite. Nous commençons par rappeler divers moyens d'établir des convergences faibles de suites de mesures et détaillons la méthodologie pour faire fi des problèmes de mesurabilité inhérents aux processus empiriques. Ces derniers tiennent une place importante dans ce chapitre, il est possible d'y retrouver les principaux ingrédients servant à établir leurs convergences. C'est dans ce contexte que sont abordées les classes de fonctions de Donsker ainsi que celles de Vapnik-Červonenkis. Il est aussi question de processus empiriques sur des échantillons bootstrap ainsi que de convergences faibles conditionnelles. Ce chapitre fait également mention de la notion de variation régulière au premier et second ordre, notion qui est intimement liée aux extrêmes. Nous y présentons principalement les bornes de Potter et de Drees qui fournissent des bornes pour les queues de distribution.
- Le troisième chapitre concerne le transport optimal. On y définit espaces et distances de Wasserstein. Certaines propriétés topologiques de ces espaces ainsi que des exemples et cas particuliers utiles à ce manuscrit y sont également recensés. Nous présentons également la notion de convergence dans ces espaces ainsi que la forme duale de la distance de Wasserstein. La seconde moitié de ce chapitre est dédiée aux liens entre cette métrique et les statistiques ainsi qu'à son lien avec les mesures empiriques.

- Le premier chapitre de la seconde partie concerne l'étude du modèle des queues proportionnelles à l'aide d'un couplage explicite et une approche processus empiriques. Ce chapitre correspond à un article publié dans *Georgian Mathematical Journal*. Le premier résultat concerne l'estimation par l'approche *Peaks Over Threshold* de γ ainsi que de la version intégrée de σ . Ces deux quantités sont estimées respectivement à l'aide de l'estimateur de Hill et par la fonction de répartition empirique des concomitants, et nous établissons la normalité jointe des deux estimateurs ainsi que leur indépendance asymptotique. Ce résultat est établi pour un seuil quelconque, ce au prix d'une hypothèse plus contraignante que celles pour l'estimateur de Hill classique. Cependant, dans ce modèle, l'estimation des quantiles nécessite d'estimer directement σ , nous proposons alors un estimateur à noyau construit sur un tableau triangulaire. Ce qui nous permet d'estimer les quantiles à l'aide d'un estimateur plug-in façonné à partir d'un estimateur de Weissman. Nous démontrons alors que ces deux estimateurs sont asymptotiquement normaux pour un seuil choisi déterministe.
- Le pénultième chapitre se concentre sur la méthode de couplage optimal et son application en théorie univariée des valeurs extrêmes. Le résultat le plus important de ce chapitre établit que la distance de Wasserstein entre les lois de mesures empiriques est la même que celle entre les distributions qui ont généré les échantillons. De plus, nous établissons, dans le contexte des extrêmes, que la distance de Wasserstein entre la loi des excès et une distribution de Pareto peut être explicitement calculée et converge vers 0. Nous illustrons ensuite cette approche en montrant qu'elle permet d'établir des résultats de convergence d'estimateurs comme l'estimateur de Hill, l'estimateur des moments pondérés ou encore celui de Weissman. Cette méthode nous permet également de traiter le cas de la présence de biais pour l'estimateur de Hill et fonctionne également avec l'approche *Block Maxima* des extrêmes. Ce chapitre à un article soumis pour publication.
- Le dernier chapitre retourne au modèle des queues proportionnelles, mais traité avec le couplage optimal. Il correspond à un travail en cours au moment où ces lignes sont écrites. Nous montrons qu'il est de nouveau possible d'exprimer la distance en Wasserstein entre la loi du modèle et la loi limite explicitement. Ce qui nous permet d'établir la normalité jointe des estimateurs de γ et de la version intégrée de σ sous des conditions standard, mais également d'étendre ce résultat à la présence de biais. Nous proposons également une version bootstrap de ce théorème. Avec l'objectif de proposer une méthode de validation du modèle, nous proposons de tester si γ est constant. Nous étudions alors une statistique de test qui compare les valeurs de l'estimateur de γ à ce même estimateur mais construit sur des sous-échantillons. Nous établissons ensuite, sous l'hypothèse de constance de γ , la convergence en loi de cette statistique vers le même pont brownien que les deux résultats précédents. Ainsi, même si cette loi limite reste inconnue car elle dépend des paramètres du modèle, nous pouvons estimer le

niveau et la puissance du test empiriquement. Nous proposons alors une étude de ce test par simulations.

Listes des publications

- **Extreme quantile regression in proportional tail framework** : accepté dans *Georgian Mathematical Journal*.
- **The coupling method in extreme value theory** : soumis, en révision mineure pour *Bernoulli Journal*, arXiv :1912.03155.
- **Analysis of proportional tail model with optimal coupling** : article en préparation.

Première partie
Le terrain de jeux

Fondamentaux de la théorie des valeurs extrêmes

La théorie des valeurs extrêmes naît en 1928 lorsque Léonard Tippett, alors chercheur pour la British Cotton Industry Research Association, cherche à améliorer la solidité de tresses de coton. Le chercheur britannique s'aperçoit alors que la solidité de la tresse est contrôlée par celle de la plus forte des fibres. Les statistiques classiques fournissent alors peu de réponses car elle se focalisent sur le comportement de la moyenne, ou de la somme, d'un échantillon et non de son maximum.

Le but devient, pour un échantillon (X_1, \dots, X_n) de réels, d'étudier le comportement asymptotique de $M_n := \max(X_1, \dots, X_n)$. À l'instar de la théorie normale, Tippett cherche deux suites d'entiers $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ et une loi limite G telles que $a_n M_n + b_n$ convergerait en loi vers la loi G . Fisher et Tippett résolvent cette question et mettent en lumière trois régimes possibles pour le maximum [41]. Des résultats similaires sont obtenus par Gnedenko [45] en 1943. Gumbel formalise ensuite cette théorie dans son livre *Statistics of extremes* [49].

Des évènements tels que le raz de marée qui frappa les Pays Bas en 1953 ont demandé une autre approche des extrêmes. En effet, dans l'espoir de construire une digue pour protéger les côtes des vagues, on s'intéresse non seulement au maximum possible mais également à la probabilité de dépasser un certain seuil, ici la hauteur de la digue. Il s'agit alors de l'estimation de quantiles élevés, même de quantiles à des niveaux non observés. Les pionniers dans cette approche sont Balkema et De Haan [3], qui étudient la convergence de la loi des excès vers une distribution de Pareto, ainsi que Pickands [68].

Dans la pratique, les deux différents points de vue vont conduire à deux manières bien distinctes de construire des estimateurs. Intuitivement, pour étudier le comportement du maximum d'un échantillon ou d'excès au-delà d'un seuil élevé, les observations proches de la moyenne (et a fortiori les plus petites) semblent donner peu d'informations. C'est pourquoi les estimations sont faites à l'aide de sous-échantillons d'évènements extrêmes. Les deux précédentes approches fournissent chacune une méthode de sélection d'échantillon extrémal. Le point de vue de Tippett et Fisher conduit à subdiviser l'échantillon en k blocs de même effectif, et dans chaque bloc on choisit la plus grande observation. Cette approche est appelée Block Maxima (BM). Quant à celui de Balkema et de Haan, le sous-échantillon est construit en choisissant les observations qui sont supérieures à un certain seuil donné, ou alors les k plus grandes observations, ce qui revient à se donner un seuil aléatoire égal

à la $(k + 1)$ -ème plus grande observation. Cette approche est appelée Peak over Threshold (PoT).

Ces différentes approches et méthodes, sont détaillées dans les monographies de de Haan et Ferreira [22] et Beirlant et al [4]. Aujourd’hui la théorie des valeurs extrêmes a trouvé sa place dans de nombreux domaines comme l’hydrologie, la météorologie, l’assurance, la finance et la géologie pour n’en citer que quelques uns.

2.1 Comportement des extrêmes

Domaines d’attraction du maximum

Soient (X_1, \dots, X_n) n copies indépendantes et identiquement distribuées (i.i.d) d’une variable aléatoire réelle X définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ ayant pour fonction de répartition F . Ici, on munit \mathbb{R} de sa tribu Borélienne $\mathcal{B}(\mathbb{R})$.

La compréhension du comportement asymptotique de M_n vient de l’étude de la fonction F^n , qui n’est autre que la fonction de répartition du maximum. En effet, le caractère i.i.d des X_i nous fournit, pour tout $x \in \mathbb{R}$,

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

Une première remarque sur le comportement de M_n est que, considérant $x_F := \sup\{x | F(x) < 1\}$ la borne supérieure du support de F ,

$$M_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} x_F,$$

où $\xrightarrow{\mathbb{P}}$ désigne la convergence en probabilité. Cette convergence ne fournit pas d’informations satisfaisantes sur la limite de M_n . Cette limite en probabilité, sans information sur la vitesse permet, par exemple, d’estimer x_F , mais pas de déduire un intervalle de confiance. L’analyse plus minutieuse de la limite du maximum passe par la notion de domaine d’attraction.

Définition 2.1. Soit G la fonction de répartition d’une variable aléatoire Z . On dit que F appartient au domaine d’attraction de G , ce que l’on note $F \in D(G)$, s’il existe deux suites de réels $(a_n)_{n \in \mathbb{N}} > 0$ et $(b_n)_{n \in \mathbb{N}}$ telles qu’en tout point x de continuité de G on ait,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x).$$

On a ainsi l’équivalence,

$$F \in D(G) \Leftrightarrow \frac{M_n - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} Z,$$

où \xrightarrow{d} désigne la convergence en loi.

Remarque 2.1. On définit ici le domaine d’attraction du maximum d’un échantillon. Mais en s’appuyant sur la relation $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$, on définit également le domaine d’attraction pour le minimum. Tout ce qui suit, énoncé en terme de maximum, est également vrai pour le minimum.

Tippett, Fischer et Gnedenko ont montré que les fonctions G , qui satisfont la Définition 2.1, ne peuvent pas être quelconques. Par la suite, Jenkinson en fournit une paramétrisation [56].

Théorème 2.1 (Tippett et Fischer (1928), Gnedenko (1943)). *Supposons $F \in D(G)$ avec G non dégénérée. Il existe alors trois constantes réelles a , b et γ telles que $G(x) = G_\gamma(ax + b)$ pour tout $x \in \mathbb{R}$ avec*

$$G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \quad 1 + \gamma x > 0.$$

Lorsque $\gamma = 0$, le côté droit de l'équation précédente est à interpréter comme $\exp(e^{-x})$.

Les fonctions G_γ sont appelées distribution GEV (pour Generalized Extreme Value). Le paramètre γ est appelé indice des valeurs extrêmes. Cet indice est de première importance car il contient les informations sur la queue de distribution de F . On peut classifier les distributions GEV en fonction du signe de γ en trois classes qui ont chacune leurs propriétés propres :

- Cas $\gamma > 0$: Le domaine d'attraction de G_γ avec γ strictement positif est appelé domaine d'attraction de Fréchet. Une distribution F dans ce domaine possède une borne supérieure de son support infinie et une queue à décroissance polynomiale. La distribution GEV associée à ce domaine est du même type que la loi de Fréchet défini par,

$$\Phi_\gamma(x) = \exp\left(-x^{-1/\gamma}\right) \mathbf{1}_{x \geq 0}, \quad x \in \mathbb{R}.$$

C'est-à-dire que les deux distributions ne diffèrent que d'un paramètre de position et d'échelle, ainsi $F \in D(G_\gamma)$ si et seulement si $F \in D(\Phi_\gamma)$. On peut caractériser l'appartenance à ce domaine d'attraction grâce au comportement asymptotique de F , plus précisément en observant ses variations.

Théorème 2.2. *Une fonction de répartition F appartient au domaine d'attraction d'une distribution GEV G_γ avec $\gamma > 0$ si et seulement si x_F est infini et si pour tout $x \in \mathbb{R}$*

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

Dans ce cas, on dit que la queue de distribution $\bar{F} := 1 - F$ est à variation régulière d'indice $-1/\gamma$, voir Section 3.4 pour plus de détails sur les fonctions à variations régulières.

Parmi les distributions appartenant au domaine d'attraction de Fréchet on notera par exemple la distribution de Fréchet, de Student, de Fisher et la loi de Pareto (ainsi que la distribution de Pareto généralisée). Ces deux dernières sont de grande importance en extrême car elles sont directement en lien avec la distribution des excès. Les distributions dont la fonction de répartition appartient au domaine d'attraction de Fréchet sont également appelées distributions à queues lourdes.

- Cas $\gamma = 0$: Le domaine d'attraction de G_γ avec γ nul est appelé domaine d'attraction de Gumbel. Une telle distribution peut avoir une borne supérieure de son support finie ou infinie, dans les deux cas la décroissance de la queue de distribution sera exponentielle. La distribution GEV est du même type que la distribution de Gumbel

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

De même que pour le domaine d'attraction de Fréchet, on peut caractériser l'appartenance au domaine de Gumbel avec les variations asymptotiques de F .

Théorème 2.3. *Une fonction de répartition F appartient au domaine d'attraction d'une distribution GEV G_γ avec $\gamma = 0$ si et seulement si, pour tout $x \in \mathbb{R}$,*

$$\lim_{t \rightarrow x_F} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x},$$

avec f une fonction bien adaptée.

Nous pouvons faire deux remarques importantes sur cette caractérisation. Premièrement, cette caractérisation ne dépend pas de la finitude de x_F car en effet, lorsque l'indice des valeurs extrêmes est nul, ce point peut être fini comme infini. Deuxièmement, le terme "adapté" du précédent théorème peut paraître flou, il est à comprendre comme "une fonction qui permet à la convergence d'avoir lieu", ce qui n'est pas une explication beaucoup plus précise. Concrètement, on sait que si la convergence du Théorème 2.3 a lieu pour un certain f alors elle a également lieu pour la fonction

$$t \mapsto \tilde{f} := \frac{\int_t^{x_F} (1 - F(s)) ds}{1 - F(t)}.$$

Cette fonction est bien définie car dans ces conditions, la quantité $\int_t^{x_F} (1 - F(s)) ds$ est finie. On peut ainsi toujours choisir $f = \tilde{f}$ pour tester l'appartenance au domaine d'attraction de Gumbel avec ce théorème.

Les lois logistiques, normale, gamma ou encore exponentielles appartiennent au domaine d'attraction de Gumbel

- Cas $\gamma < 0$: Le domaine d'attraction de G_γ avec γ strictement négatif est appelé domaine d'attraction de Weibull. Une distribution appartenant à ce domaine a une borne supérieure du support finie et une queue à décroissance polynomiale. Le domaine d'attraction de la distribution GEV associée est le même que celui d'un loi de Weibull négative, défini par

$$\Psi_\gamma(x) = \exp\left(-(-x)^{-1/\gamma}\right) \mathbf{1}_{x \leq 0} + \mathbf{1}_{x > 0}, \quad x \in \mathbb{R}.$$

On caractérise également ce domaine d'attraction avec le théorème suivant :

Théorème 2.4. Une fonction de répartition F appartient au domaine d'attraction d'une distribution GEV G_γ avec $\gamma < 0$ si et seulement si x_F est fini et si pour tout $x \in \mathbb{R}$,

$$\lim_{t \rightarrow 0} \frac{1 - F(x_F - tx)}{1 - F(x_F - t)} = x^{-1/\gamma}.$$

Parmi les lois dans le domaine d'attraction de Weibull on retrouve par exemple les lois uniformes ou les lois beta.

Ces résultats permettent de caractériser chacun des domaines d'attractions possible, cependant ils présupposent la connaissance du signe de γ car les ratios à étudier ne sont pas les mêmes dans les trois cas. De plus, ces caractérisations permettent d'établir que $F \in D(G_\gamma)$ pour un certain γ mais ne fournissent pas les suites $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ de la définition 2.1, indispensables à l'élaboration d'un théorème limite pour le maximum de l'échantillon. Pour passer outre ce problème, il est nécessaire d'introduire les fonctions quantiles.

Définition 2.2. Soit F une fonction de répartition.

1. On appelle fonction quantile de F , que l'on note F^{\leftarrow} , la fonction définie pour tout $t \in [0, 1]$ par

$$F^{\leftarrow}(t) := \inf\{x | F(x) \geq t\}.$$

2. On appelle fonction quantile de queue de F , la fonction définie pour tout $t \geq 1$ par

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

La fonction quantile est également appelée inverse généralisée ou encore pseudo-inverse. Dans le cas où F est continue et strictement monotone la fonction quantile coïncide avec l'inverse, mais ce n'est pas vrai en toute généralité. Plus précisément, pour $t \in [0, 1]$ on a $F \circ F^{\leftarrow}(t) \geq t$ avec égalité si et seulement si t est dans l'image de F . On a alors que l'égalité est fautive uniquement lorsque F n'est pas continue en $F^{\leftarrow}(p)$. Pour $x \in \mathbb{R}$, on a $F^{\leftarrow} \circ F(x) \leq x$, il n'y a pas égalité si et seulement si x est sur la partie intérieure, ou sur l'extrémité droite, d'un plat de F . La fonction quantile de queue nous permet une nouvelle fois de caractériser l'appartenance à un domaine d'attraction, mais sans pré-requis sur la connaissance du signe de γ .

Théorème 2.5. Soient F une fonction de répartition de fonction quantile de queue U et $\gamma \in \mathbb{R}$. Les assertions suivantes sont équivalentes.

1. Il existe deux suites de constantes réelles $(a_n)_{n \in \mathbb{N}} > 0$ et $(b_n)_{n \in \mathbb{N}}$ telles que

$$\lim_{t \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x), \quad 1 + \gamma x > 0.$$

2. Il existe une fonction a , strictement positive, telle que pour tout $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}. \quad (2.1)$$

Lorsque $\gamma = 0$ la partie droite de l'équation est à comprendre comme $\log(x)$.

De plus, le premier point est vrai avec $b_n := U(n)$ et $a_n := a(n)$. Il est également possible de fournir une expression explicite pour la fonction a

$$a : t \geq 0 \mapsto \begin{cases} \gamma U(t) & \text{si } \gamma > 0 \\ \gamma(U(t) - U(\infty)) & \text{si } \gamma = 0 \\ \frac{\int_{U(t)}^{x^*} (1-F(s)) ds}{1-F(t)} & \text{si } \gamma < 0. \end{cases}$$

La condition (2.1) du point 2 est appelée condition des valeurs extrêmes du premier ordre, ou simplement condition du premier ordre s'il n'y a pas d'ambiguïté. Par la suite, c'est cette condition qui caractérisera l'appartenance à un domaine d'attraction. Si le signe de l'indice des valeurs extrêmes est connu, ou si on s'y restreint, on peut simplifier la condition du premier ordre.

Théorème 2.6. 1. Si $\gamma > 0$, alors la condition du premier ordre (2.1) est équivalente à

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad x > 0. \quad (2.2)$$

2. Si $\gamma < 0$, alors la condition du premier ordre est équivalente à $U(\infty) := \lim_{t \rightarrow \infty} U(t) < \infty$ et

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

Loi des excès

Une autre approche des extrêmes d'un phénomène est de regarder le comportement au-dessus d'un seuil, généralement élevé, que l'on appelle u . C'est-à-dire que l'on s'intéresse à une quantité de la forme $\mathbb{P}(X > u + x | X > u)$, il s'agit de la probabilité de dépasser le seuil u de la quantité x sachant que l'on est déjà au-dessus du seuil. En termes de loi, nous étudions donc $\mathcal{L}(X - u | X > u)$. Les assertions du théorème 2.5 sont également équivalentes à l'existence d'une fonction a telle que

$$\lim_{t \rightarrow x_F} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0,$$

avec la fonction f qui peut être choisie comme $t \mapsto a(1/(1 - F(t)))$. Ce qui nous donne déjà le candidat pour la loi limite de tels dépassements de seuil. En effet on peut écrire

$$\frac{1 - F(u + xf(u))}{1 - F(u)} = \frac{\mathbb{P}(u + xf(u))}{\mathbb{P}(X > u)} = \mathbb{P}\left(\frac{X - u}{f(u)} > x | X > u\right).$$

Il apparaît alors que la limite dans ce point 4 joue un rôle essentiel dans la compréhension du comportement des excès (renormalisés). La loi ayant pour fonction de répartition

$$H_\gamma(x) := 1 - (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0$$

est appelée loi de Pareto généralisée (ou GPD pour Generalized Pareto Distribution). Si $\gamma = 0$ cette fonction de répartition est à comprendre comme $H_0(x) = 1 - \exp(-x)$.

Théorème 2.7. [Balkema et de Haan (1974), Pickands (1975)] Soient F une fonction de répartition de fonction quantile de queue U et $\gamma \in \mathbb{R}$. Les assertions suivantes sont équivalentes.

1. Il existe deux suites de constantes réelles $(a_n)_{n \in \mathbb{N}} > 0$ et $(b_n)_{n \in \mathbb{N}}$ telles que

$$\lim_{t \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x), \quad 1 + \gamma x > 0.$$

2. Il existe une fonction strictement positive f telle que

$$\lim_{u \rightarrow x_F} \mathbb{P} \left(\frac{X - u}{f(u)} > x \mid X > u \right) = 1 - H_\gamma(x), \quad 1 + \gamma x > 0.$$

Ce théorème établit l'équivalence entre l'appartenance à un domaine d'attraction, c'est-à-dire que le maximum suit asymptotiquement une GEV, et le fait que les comportement des excès est approché par une distribution de Pareto généralisée. Les deux approches, point de vue du maximum ou des excès, sont donc équivalentes. En outre, elles sont toutes deux équivalentes à la condition du premier ordre. Cette équivalence s'arrête au premier ordre, nous verrons plus tard, que si on considère les vitesses dans les convergences des points 1 et 3 du théorème 2.7, l'équivalence entre les approches est perdue.

Au-delà de l'équivalence entre les approches, on remarque que les limites, que ce soit la GPD, la GEV ou même la limite dans la condition du premier ordre, sont toutes paramétrées par le même nombre γ . L'indice des valeurs extrêmes caractérisant également les queues de distribution, ce paramètre apparaît comme étant central dans cette théorie. Il sera alors de première importance de savoir l'estimer.

Dans le cas où l'on saurait que γ est strictement positif, on peut simplifier, à l'image de la condition du premier ordre, l'expression du comportement des excès. En effet, le théorème 2.2 donne

$$\frac{1 - F(tu)}{1 - F(u)} = \mathbb{P}(u^{-1}X > x \mid X > u) \xrightarrow{u \rightarrow \infty} x^{-1/\gamma}.$$

Dans le cas $\gamma > 0$ la distribution GPD comme loi limite des excès est remplacée par une loi Pareto. Il s'agit de la loi ayant pour fonction de répartition $x \mapsto x^{-1/\gamma}$, cette loi est appelée plus précisément, loi de Pareto d'indice $1/\gamma$. Dans ce cas, on peut énoncer une version simplifiée, adaptée au cas Fréchet, du théorème précédent.

Théorème 2.8. Soient F une fonction de répartition de fonction quantile de queue U et $\gamma > 0$. Les assertions suivantes sont équivalentes.

1. Il existe deux suites de constantes réelles $(a_n)_{n \in \mathbb{N}} > 0$ et $(b_n)_{n \in \mathbb{N}}$ telles que

$$\lim_{t \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp(-x^{-1/\gamma}) \mathbf{1}_{x \geq 0}, \quad x \in \mathbb{R}.$$

2. Pour tout $x > 0$,

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

3. Pour tout $x > 0$,

$$\lim_{u \rightarrow \infty} \mathbb{P}(u^{-1}X > x \mid X > u) = x^{-1/\gamma}.$$

Comportement des statistiques d'ordre

Précédemment, nous nous sommes intéressés au maximum et aux éléments au-dessus d'un seuil. Ce seuil peut être déterministe ou aléatoire. Dans la pratique, la plupart des résultats se concentrent sur le cas du seuil aléatoire. Plus précisément, le seuil choisit est le $(k + 1)$ ème plus grand élément de l'échantillon observé. Il est, dans ce contexte, utile d'introduire la notion de statistique d'ordre.

Définition 2.3. Soient (X_1, \dots, X_n) un échantillon et k un entier compris entre 1 et n . La k ème statistique d'ordre est la k -ième plus petite valeur de l'échantillon. On la note $X_{k:n}$.

On a donc, avec ces notations,

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{k:n} \leq \dots \leq X_{n:n}$$

et $X_{n:n} = \max(X_1, \dots, X_n)$. Il est possible de retrouver une étude détaillée des statistiques d'ordre dans l'ouvrage de David et Nagaraja qui leur est consacré [19]. On peut par exemple y retrouver l'expression de la densité de la k -ième statistique d'ordre d'un échantillon i.i.d de densité f et de fonction de répartition F . Cette densité est donnée par

$$f_{X_{k:n}}(x) := \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} (1-F(x))^{n-k} f(x).$$

Dans le présent manuscrit nous nous intéressons plus particulièrement au comportement de certaines statistiques d'ordre. Celle de la forme $X_{n-k:n}$ où $k := k(n)$ est une suite qui dépend de n telle que

$$k \xrightarrow[n \rightarrow \infty]{} \infty \quad \text{et} \quad \frac{k}{n} \xrightarrow[n \rightarrow \infty]{} 0.$$

Une telle suite est dite intermédiaire. Dans la pratique, le seuil choisi pour étudier les estimateurs construits à partir de sous-échantillons de valeurs extrêmes est $X_{n-k:n}$. Il est alors naturel de commencer à étudier ces statistiques d'ordre. La normalité des statistiques d'ordre ne peut être établie sous la condition du premier ordre. Dans un premier temps nous pouvons considérer le comportement de ces statistiques d'ordre sous les conditions dites de Von Mises. Conditions, qui sont des conditions suffisantes d'appartenance à un domaine d'attraction.

Théorème 2.9. Soit F une fonction de répartition telle que F'' existe et que F' soit strictement positive dans un voisinage (à gauche) de x_F . Alors, si

$$\lim_{t \rightarrow x_F} \left(\frac{1-F}{F'} \right)'(t) = \gamma,$$

alors F appartient au domaine d'attraction de G_γ .

Ces conditions sont bien plus fortes et restrictives que la condition du premier ordre, et ne sont pas équivalentes à l'appartenance aux domaines d'attraction. Mais elles permettent d'établir le résultat suivant :

Théorème 2.10. Soient F satisfaisant aux conditions du Théorème 2.9 et $k = k(n)$ une suite intermédiaire, alors

$$\sqrt{k} \frac{X_{n-k:n} - U\left(\frac{n}{k}\right)}{\frac{n}{k} U'\left(\frac{n}{k}\right)} \xrightarrow[n \rightarrow \infty]{d} N,$$

avec $N \sim \mathcal{N}(0, 1)$ la loi normale centrée réduite.

Par exemple, si on spécifie ce théorème avec des variables aléatoires de Pareto d'indice 1. On obtient alors

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

Si nous voulons étendre les résultats précédents à un cadre plus général, la condition du premier ordre n'est plus suffisante. Cette condition peut être interprétée comme un développement asymptotique de U au premier ordre en l'infini, pour de nombreux résultats il est indispensable de pousser ce développement un ordre plus loin. C'est ce qu'on appelle la condition du second ordre.

Définition 2.4. La distribution F , ou de manière équivalente la fonction quantile de queue U , est dite satisfaisant la condition du second ordre s'il existe une fonction strictement positive a , une fonction A ne changeant pas de signe et tendant vers 0 en l'infini telles que

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H(x), \quad x > 0,$$

avec H une fonction qui n'est pas un multiple de $(x^\gamma - 1)/\gamma$.

Les fonctions a et A sont appelées respectivement fonctions auxiliaires du premier et second ordre. Comme A tend vers 0, il est clair que la condition du second ordre implique l'appartenance au domaine d'attraction de G_γ . Ainsi, cette fonction A peut être interprétée comme la vitesse de convergence vers 0 dans la condition du premier ordre. Par définition, la fonction H ne peut pas être identiquement nulle, mais il est possible d'aller plus loin dans la caractérisation de cette limite.

Théorème 2.11. Soit F une fonction satisfaisant la condition du second ordre de la Définition 2.4. Alors il existe une fonction a positive et une fonction A de signe constant tendant vers 0 ainsi qu'un réel $\rho \leq 0$ tels que

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \Psi_{\gamma, \rho}(x) := \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds, \quad x > 0. \quad (2.3)$$

De plus,

$$\lim_{r \rightarrow \infty} \frac{A(tx)}{A(t)} = x^\rho.$$

Les fonctions a et A ne sont pas nécessairement les mêmes que celles de la Définition 2.4.

Nous pouvons réécrire la fonction limite comme

$$\Psi_{\gamma,\rho}(x) = \frac{1}{\rho} \left(\frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right), \quad x > 0.$$

Les cas avec γ ou ρ nul sont à interpréter comme suit,

$$\begin{cases} \frac{1}{\gamma} \left(x^\gamma \log(x) - \frac{x^\gamma - 1}{\gamma} \right) & \text{si } \rho = 0 \neq \gamma \\ \frac{1}{\rho} \left(\frac{x^\rho - 1}{\rho} - \log(x) \right) & \text{si } \rho \neq 0 = \gamma \\ \frac{1}{2} \log(x)^2 & \text{si } \rho = 0 = \gamma. \end{cases}$$

Dorénavant, la condition des valeurs extrêmes du second ordre sera toujours à entendre au sens de l'Équation (2.3). Le paramètre ρ est alors appelé paramètre du second ordre.

Dans le cas où F est une distribution à queue lourde, on peut énoncer une condition du second ordre simplifiée.

Théorème 2.12. *Soit F une fonction de répartition d'indice des valeurs extrêmes $\gamma > 0$. S'il existe une fonction A_0 ne changeant pas de signe telle que*

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A_0(t)} = K(x), \quad x > 0,$$

avec K non identiquement nulle, alors il existe une fonction A ne changeant pas de signe, convergant vers 0, possiblement différente de A_0 et un réel $\rho \leq 0$ tels que

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad x > 0. \quad (2.4)$$

Remarque 2.2 (Exemple 2.3.11 dans [22]). *Lorsque γ est strictement positif il est commun de travailler plutôt avec la condition (2.4) qu'avec la condition (2.3) cependant les deux conditions ne sont pas équivalentes. Par exemple, pour $U(t) = t^\gamma + 1 + t^{\gamma+\rho}$ avec $\gamma > 0$ et $\gamma + \rho < 0$, la condition (2.4) est satisfaite avec $A(t) \sim t^{-\gamma}$ alors que la condition (2.3) est satisfaite pour $A(t) \sim t^\rho = o(t^{-\gamma})$. Ce qui signifie que l'approximation par une GPD est meilleur que l'approximation Pareto.*

Ainsi, sous cette condition du second ordre, il est possible de rétablir la normalité asymptotique des statistiques d'ordre.

Théorème 2.13. *Soit F une fonction de répartition satisfaisant la condition du second ordre (2.3) avec $\gamma \in \mathbb{R}$ et $\rho \leq 0$. Alors, pour $k = k(n)$ une suite intermédiaire telle que $\lim_{n \rightarrow \infty} \sqrt{k}A(n/k)$ existe et est finie, on a*

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

La question de l'équivalence entre les méthodes BM et PoT au second ordre, établie au premier ordre par le Théorème 2.7, a été bien clarifiée en 2003 par Drees et

al. [31]. Commençons par préciser l'équivalence entre les deux approches au premier ordre. La première chose à remarquer, est que les expressions 2.1 et 2.3 sont plus adaptées à l'approche Peak over Threshold. Dans l'approche Block Maxima, il est plus commode de considérer ces équations avec la fonction U remplacée par V , définie par

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

Ainsi les deux expressions de la condition du premier ordre sont

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a_{PoT}(t)} = \frac{x^\gamma - 1}{\gamma}, \quad x > 0$$

et

$$\lim_{t \rightarrow \infty} \frac{V(tx) - V(t)}{a_{BM}(t)} = \frac{x^\gamma - 1}{\gamma}, \quad x > 0.$$

Ces deux conditions sont équivalentes au sens où, si l'une est vraie alors l'autre est vraie également, et dans ce cas $a_{PoT} = a_{BM}$.

La question est alors celle de l'équivalence entre

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_{PoT}(t)} - \frac{x^\gamma - 1}{\gamma}}{A_{PoT}(t)} = \Psi_{\gamma, \rho_{PoT}}(x), \quad x > 0$$

et

$$\lim_{t \rightarrow \infty} \frac{\frac{V(tx) - V(t)}{a_{BM}(t)} - \frac{x^\gamma - 1}{\gamma}}{A_{BM}(t)} = \Psi_{\gamma, \rho_{BM}}(x) \quad x > 0.$$

Si la condition du second ordre est vraie avec U alors, sous réserve que V soit une fonction à variations régulières étendues au deuxième ordre (notion détaillée en section 3.4), la condition du second ordre en V est aussi vraie avec A_{BM} et a_{BM} obtenus par transformation explicite de A_{PoT} et a_{PoT} et $\rho_{BM} = \max(\rho_{PoT}, -1)$. La réciproque est également vraie.

2.2 Estimation des paramètres

Étant une quantité de grande importance, il y a de nombreux estimateurs de l'indice des valeurs extrêmes. Parmi tous, le plus célèbre est certainement l'estimateur de Hill. Cet estimateur, introduit par Hill en 1975 [50], est défini avec les k plus grandes statistiques d'ordre comme suit

$$\hat{\gamma}_k^H := \frac{1}{k} \sum_{i=1}^k \log \left(\frac{X_{n-i+1:n}}{X_{n-k:n}} \right).$$

Cet estimateur est valable uniquement pour estimer $\gamma > 0$, mais a l'avantage de s'exprimer simplement.

Déjà sous la condition du premier ordre, il est possible d'établir la consistance de cet estimateur.

Théorème 2.14. *Supposons $F \in D(G_\gamma)$ avec $\gamma > 0$. Soit $k = k(n)$ une suite intermédiaire. Alors*

$$\hat{\gamma}_k^H \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \gamma.$$

De plus, si nous ajoutons une condition supplémentaire à la suite k , une réciproque à ce théorème est également vraie. En effet, si k est une suite intermédiaire telle que $k(n+1)/k(n) \rightarrow 1$ et $\hat{\gamma}_k^H \xrightarrow{\mathbb{P}} \gamma > 0$, alors $F \in D(G_\gamma)$.

Un tel résultat de convergence de probabilité ne permet pas d'obtenir d'intervalles de confiance et ainsi d'être pratiquement utilisable. La nécessité de condition de second ordre apparaît de nouveau.

Théorème 2.15. *Supposons que F satisfasse la condition du second ordre (2.3) avec $\gamma > 0$ et $\rho \leq 0$, ou de manière équivalente (2.4). Soit k une suite intermédiaire telle que $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$, alors*

$$\sqrt{k}(\hat{\gamma}_k^H - \gamma) \xrightarrow[k \rightarrow \infty]{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

Dans ce théorème, λ dans le terme de biais est fini, et la fonction A , qui est la vitesse dans la condition du second ordre 2.4 est de signe constant et tend vers 0 en l'infini. On peut également noter que cet estimateur est sans biais lorsque $\sqrt{k}A(n/k) \rightarrow 0$. C'est également le cas lorsque la convergence de $U(tx)/U(t)$ vers x^γ est plus rapide que toute puissance.

Théorème 2.16. *Soit F telle que pour tout $x > 0$ et tout $\beta > 0$ on a*

$$\lim_{t \rightarrow \infty} t^\beta \left(\frac{U(tx)}{U(t)} - x^\gamma \right) = 0$$

alors $\sqrt{k}(\hat{\gamma}_k^H - \gamma) \xrightarrow{d} \mathcal{N}(0, \gamma^2)$.

Nous avons pu voir que l'estimateur de Hill ne peut que estimer γ s'il est strictement positif, il est donc nécessaire de construire un estimateur pour estimer γ négatif ou nul. C'est pour estimer un tel indice des valeurs extrêmes qu'en 1985, Hosking et al. [53] ainsi que Hosking et Wallis [52] en 1987, introduisent et utilisent des estimateurs des moments pondérés (PWM pour Probability Weighted Moment). Cette méthode consiste à considérer les moments pondérés d'une distribution de Pareto généralisée à laquelle on ajoute un facteur d'échelle $\sigma > 0$. C'est-à-dire de la forme

$$H_{\gamma,\sigma}(x) := 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma}, \quad 1 + \gamma \frac{x}{\sigma} > 0.$$

Les moments pondérés d'une variable aléatoire X avec pour fonction de répartition F , introduits originellement par Greenwood et al [46] en 1979, sont les quantités

$$M_{p,r,s} := \mathbb{E}(X^p F(x)^r (1 - F(x))^s)$$

pour p, r et s des entiers. Pour l'estimation de γ et σ on s'intéresse particulièrement à deux moments pondérés de $H_{\gamma,\sigma}$, précisément

$$M_{1,0,0} = \frac{\sigma}{1-\gamma} \quad \text{et} \quad M_{1,0,1} = \frac{\sigma}{2(2-\gamma)}.$$

Pour s'assurer de l'existence de ces moments, nous considérons que $F \in D(\gamma)$ avec $\gamma < 1$. Ce qui n'est pas une restriction en soi car le but ici est d'estimer un γ

négatif ou nul. Les expressions de $M_{1,0,0}$ et $M_{1,0,1}$ nous permettent alors d'exprimer les paramètres de la GPD comme

$$\gamma = 2 - \frac{M_{1,0,0}}{M_{1,0,0} - 2M_{1,0,1}} \quad \text{et} \quad \sigma = \frac{2M_{1,0,0}M_{1,0,1}}{M_{1,0,0} - 2M_{1,0,1}}.$$

Ainsi, l'estimation de γ (jointe à celle de σ) relève de l'estimation des moments pondérés. On estime les moments $M_{1,0,s}$, pour $s \in \mathbb{R}$ par

$$\hat{M}_{k,s} := \frac{1}{k} \sum_{i=1}^k \left(1 - \frac{i}{k}\right)^s (X_{n-i+1:n} - X_{n-K:n}).$$

Dans cette notation pour l'estimateur, les indices p et r de la définition de moments pondérés n'apparaissent plus. En effet, ils sont considérés comme fixes, valant 1 et 0, cela allège la notation. Dans le cadre de l'estimation de γ et σ , nous nous intéressons uniquement aux estimateurs $\hat{M}_{k,0}$ et $\hat{M}_{k,1}$. L'étude du comportement de tels estimateurs est détaillée par Hosking dans sa monographie [54]. Ainsi, les estimateurs par moments pondérés sont construits comme suit

$$\left(\hat{\gamma}_k^{PWM}, \hat{\sigma}_k^{PWM}\right) := \left(2 - \frac{\hat{M}_{k,0}}{\hat{M}_{k,0} - 2\hat{M}_{k,1}}, \frac{2\hat{M}_{k,0}\hat{M}_{k,1}}{\hat{M}_{k,0} - 2\hat{M}_{k,1}}\right).$$

On peut alors établir la convergence jointe de ces deux estimateurs.

Théorème 2.17. *Supposons que F ait des excès exactement GPD avec $\gamma < 1/2$, alors, pour $k = k(n)$ une suite intermédiaire on a,*

$$\sqrt{k} \left(\hat{\gamma}_k^{PWM} - \gamma, \hat{\sigma}_k^{PWM} - \sigma\right) \xrightarrow[k \rightarrow \infty]{d} \mathcal{N}(0, V)$$

avec

$$V := \begin{pmatrix} \frac{(1-\gamma)(2-\gamma)^2(1-\gamma+2\gamma^2)}{(1-2\gamma)(3-2\gamma)} & \frac{(2-\gamma)(-2+6\gamma-7\gamma^2+2\gamma^3)}{(1-2\gamma)(3-2\gamma)} \\ \frac{(2-\gamma)(-2+6\gamma-7\gamma^2+2\gamma^3)}{(1-2\gamma)(3-2\gamma)} & \frac{(31-94\gamma+102\gamma^2-126\gamma^3+144\gamma^4-80\gamma^5+16\gamma^6)}{(1-2\gamma)(3-2\gamma)} \end{pmatrix}.$$

Ce théorème est valable pour les distributions ayant des excès exactement Pareto, mais nous pouvons tout de même établir un théorème limite pour ces estimateurs sous les conditions des valeurs extrêmes du premier et second ordre. Nous allons nous restreindre pour ce théorème au cas où $\sigma = 1$, ce qui n'est pas à proprement parler une restriction, car nous retrouvons ainsi l'expression de H_γ comme loi limite des excès.

Théorème 2.18. *Supposons $F \in D(\gamma)$ avec $\gamma < 1$ et soit $k = k(n)$ une suite intermédiaire. Alors,*

$$\hat{\gamma}_k^{PWM} \xrightarrow[k \rightarrow \infty]{\mathbb{P}} \gamma \quad \text{et} \quad \frac{\hat{\sigma}_k^{PWM}}{a(n/k)} \xrightarrow[k \rightarrow \infty]{\mathbb{P}} 1.$$

De plus, si F satisfait la condition du second ordre 2.3 avec $\gamma < 1/2$ et que k est telle que $\sqrt{k}A(n/k) \rightarrow \lambda$ fini, alors

$$\sqrt{k} \left(\hat{\gamma}_k^{PWM} - \gamma, \frac{\hat{\sigma}_k^{PWM}}{a(n/k)} - 1\right) \xrightarrow[k \rightarrow \infty]{d} \mathcal{N}(B, V)$$

avec

$$B := \begin{cases} \frac{\lambda}{(1-\lambda-\rho)(2-\lambda-\rho)}((1-\lambda)(2-\lambda), -\rho) & \text{si } \rho < 0 \\ (\lambda, 0) & \text{si } \rho = 0 \neq \gamma \\ \left(\lambda, -\frac{\lambda}{2}\right) & \text{si } \rho = 0 = \gamma \end{cases}$$

et V la matrice définie dans le Théorème 2.17.

Ces deux précédents estimateurs sont ceux utilisés dans mes travaux de thèse, cependant il en existe d'autres. Par exemple, l'estimateur de Pickands [68] valable pour tout $\gamma \in \mathbb{R}$, l'estimateur du maximum de vraisemblance [30] valable pour $\gamma > -1/2$ ou encore l'estimateur des moments [25] valable pour tout $\gamma \in \mathbb{R}$. Le Théorème 2.17 établit que lorsque les excès sont renormalisés, $\hat{\sigma}_k^{PWM}$ est un estimateur asymptotiquement normal de $a(n/k)$.

2.3 Estimation de quantiles extrêmes

Comme mentionné dans l'introduction, un des enjeux de la théorie des valeurs extrêmes est l'estimation de quantiles élevés. On appelle quantile d'ordre $1 - \alpha$, avec $\alpha \in [0, 1]$ la quantité

$$q(\alpha) := F^{\leftarrow}(1 - \alpha).$$

Un quantile est dit extrême, non seulement si α est petit, mais si $\alpha = \alpha(n) := \alpha_n$ dépend de la taille de l'échantillon et tend vers 0. Ainsi, reformulé en termes de fonction quantile de queue, notre intérêt est l'estimation de

$$q(\alpha_n) = U\left(\frac{1}{\alpha_n}\right)$$

avec $\alpha_n \rightarrow 0$.

On a déjà donné une réponse pour une forme particulière de α_n . En effet, pour $\alpha_n = k/n$ avec k une suite intermédiaire, on a $q(\alpha_n) = U(n/k)$. Dans ce cas, si F satisfait la condition du second ordre alors $X_{n-k:n}$ est un estimateur asymptotiquement normal du quantile d'après le Théorème 2.13. Plus généralement, on peut estimer le quantile d'ordre α_n par $\mathbb{F}_n^{\leftarrow}(1 - \alpha_n)$, avec \mathbb{F}_n la fonction de répartition empirique, définie par

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

Cependant, lorsque $\alpha_n < 1/n$ ce type d'estimateur estimera toujours le quantile par la plus grande statistique d'ordre, ie le maximum. Il apparaît alors nécessaire d'extrapoler la queue de distribution pour pouvoir estimer convenablement un tel quantile extrême. La construction des estimateurs présentés ici vient de l'hypothèse du premier ordre via l'heuristique suivante. On sait que si $F \in D(G_\gamma)$, pour t grand, on a

$$\frac{U(tx) - U(t)}{a(t)} \approx \frac{x^\gamma - 1}{\gamma}.$$

Ainsi, en prenant $t = n/k$ et $x = k/n\alpha_n$, la précédente approximation devient

$$U\left(\frac{1}{\alpha_n}\right) \approx U\left(\frac{n}{k}\right) + a\left(\frac{n}{k}\right) \frac{\left(\frac{k}{n\alpha_n}\right)^\gamma - 1}{\gamma}.$$

Nous voyons ici apparaître naturellement, de la condition du premier ordre, un estimateur de $q(\alpha_n) = U(1/\alpha_n)$. En effet, nous avons déjà vu que $X_{n-k:n}$ est un estimateur asymptotiquement normal de $U(n/k)$. Nous connaissons aussi plusieurs estimateurs pour γ . En conséquence, pour obtenir un estimateur de $q(\alpha_n)$ il ne manque plus qu'un estimateur $\hat{a}(n/k)$ de $a(n/k)$. On peut, par exemple, choisir $\hat{\sigma}_k^{PMM}$ qui est donné par le Théorème 2.17. On considère alors l'estimateur, appelé estimateur GPD,

$$\hat{q}^{GPD}(\alpha_n) = X_{n-k:n} + \hat{a}\left(\frac{n}{k}\right) \frac{\left(\frac{k}{n\alpha_n}\right)^{\hat{\gamma}_n} - 1}{\hat{\gamma}_n}$$

avec $\hat{a}(n/k)$ un estimateur de $a(n/k)$ et $\hat{\gamma}_n$ un estimateur de γ .

Théorème 2.19. *Supposons*

1. F satisfait la condition du second ordre avec $\rho < 0$ (ou $\rho = 0$ si $\gamma < 0$)
2. $k = k(n)$ est une suite intermédiaire telle que $\sqrt{k}A(n/k) \rightarrow 0 \in \mathbb{R}$
3. $n\alpha_n = o(k)$ et $\log(n\alpha_n) = o(\sqrt{k})$ lorsque $n \rightarrow \infty$
4. $\hat{\gamma}_n$ et $\hat{a}(n/k)$ sont des estimateurs tels que, sous les conditions précédentes,

$$\sqrt{K} \left(\hat{\gamma}_n - \gamma, \frac{\hat{a}(n/k)}{a(n/k)} - 1, \frac{X_{n-k:n} - U(n/k)}{a(n/k)} \right) \xrightarrow[k \rightarrow \infty]{d} (\Gamma, \Lambda, N)$$

avec (Γ, Λ, N) un vecteur de loi normale avec espérance et matrice de covariance connues.

On a alors, lorsque $n \rightarrow \infty$

$$\frac{\sqrt{k}}{v_n} \left(\hat{q}^{GPD}(\alpha_n) - q(\alpha_n) \right) \xrightarrow{d} \Gamma + \min(0, \gamma)^2 N - \min(0, \gamma)\Lambda - \frac{\lambda \min(0, \gamma)}{\min(0, \gamma) + \rho},$$

avec

$$v_n := a(n/k) \int_1^{k/(n\alpha_n)} s^{\gamma-1} \log(s) ds.$$

Les hypothèses de ce théorème paraissent assez classiques au vu des énoncés des théorèmes précédents, à l'exception de l'hypothèse 3). Cette hypothèse est propre à l'extrapolation de la queue de distribution. La partie $n\alpha_n = o(k)$ signifie que l'on estime bien un quantile extrême alors que $\log(n\alpha_n) = o(\sqrt{k})$ impose une limite à l'extrapolation. C'est-à-dire si α_n décroît trop vite vers 0 et que le quantile sera en conséquence trop loin du maximum observé, alors on aura besoin de trop extrapoler la queue de distribution pour avoir un estimateur efficace.

Lorsque γ est strictement positif la condition du premier ordre est simplifiée, il est alors possible d'en déduire une autre approximation du quantile. C'est cette considération qui a poussé Weissman à proposer un autre estimateur [86]. L'heuristique précédente appliquée avec la condition 2.2 nous donne

$$U\left(\frac{1}{\alpha_n}\right) \approx U\left(\frac{n}{k}\right) \left(\frac{k}{n\alpha_n}\right)^\gamma.$$

De cette façon, seule l'estimation de γ est nécessaire pour estimer un quantile dans le régime des queues lourdes. Étant donné un estimateur $\hat{\gamma}_n$ de γ , on peut estimer le quantile d'ordre $1 - \alpha_n$ par

$$\hat{q}^W(\alpha_n) := X_{n-k:n} \left(\frac{k}{n\alpha_n} \right)^{\hat{\gamma}_n}.$$

Théorème 2.20. *Supposons $F \in D(G_\gamma)$ avec $\gamma >$ telle que*

1. *F satisfait la condition du second ordre 2.4 avec $\rho < 0$*
2. *$k = k(n)$ est une suite intermédiaire telle que $\sqrt{k}A(n/k) \rightarrow \lambda$ lorsque $n \rightarrow \infty$*
3. *$n\alpha_n = o(k)$ et $\log(n\alpha_n) = o(\sqrt{k})$ lorsque $n \rightarrow \infty$*
4. *$\hat{\gamma}_n$ est un estimateur de γ tel que, sous les conditions précédentes,*

$$\sqrt{k}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \Gamma$$

avec Γ une loi normale d'espérance et variance connues (pouvant dépendre de γ et ρ).

On a alors, lorsque $n \rightarrow \infty$,

$$\frac{\sqrt{k}}{\log\left(\frac{k}{n\alpha_n}\right)} \left(\frac{\hat{q}^W(\alpha_n)}{q(\alpha_n)} - 1 \right) \xrightarrow{d} \Gamma.$$

Par exemple, comme l'estimateur de Hill satisfait les hypothèses de ce théorème, on retrouve l'estimateur de Weissman :

$$X_{n-k:n} \left(\frac{k}{n\alpha_n} \right)^{\hat{\gamma}_k^H}$$

qui est asymptotiquement normal. Bien que cet estimateur de Weissman est plus simple à définir et à étudier que son homologue GPD, il possède le désavantage d'être moins performant. En effet, le rapport des deux vitesses de convergence (Weissman/GPD) est de l'ordre de $(n\alpha_n/k)^\gamma$ qui tend vers 0 sous les hypothèse du théorème 2.20. Ainsi, lorsque γ est strictement positif, l'estimateur GPD converge plus rapidement que l'estimateur de Weissman.

2.4 Extrêmes Hétéroscédastiques

Une hypothèse de travail classique et confortable est de considérer les échantillons que l'on étudie sont i.i.d. Mais en pratique cette hypothèse peut être mise en défaut, soit par une structure de dépendance sur les variables, soit parce que les variables peuvent avoir des lois différentes. Le premier cas a été étudié par exemple par Hsing [55] ou Drees et Rootzèn [33]. C'est dans le second contexte, c'est à dire celui d'échantillons indépendants mais d'observations tirées selon des lois différentes

qu'en 2016 Einmahl, de Haan et Zhou [35] proposent le modèle des extrêmes hétéroscédastiques pour estimer l'indice des valeurs extrêmes.

Ils considèrent n variables aléatoires $(X_1^{(n)}, \dots, X_n^{(n)})$, respectivement de fonction de répartition $F_{n,1}, \dots, F_{n,n}$ qui partagent toutes la même borne supérieure du support x_F . L'hypothèse principale de ce modèle est l'existence d'une fonction F , partageant la même borne supérieure du support x_F et une fonction c positive définie sur $[0, 1]$, telles que

$$\lim_{x \rightarrow x_F} \frac{1 - F_{n,i}(x)}{1 - F(x)} = c\left(\frac{i}{n}\right), \quad (2.5)$$

uniformément et $n \in \mathbb{N}$ et $1 \leq i \leq n$. Cette fonction c , qui peut être choisie comme d'intégrale 1, est appelée fonction skedasis. Cette fonction skedasis renseigne sur la tendance des extrêmes, plus exactement sur la tendance de leur fréquence d'apparition. La seconde hypothèse du modèle, celle qui concerne plus directement les extrêmes, est de se placer dans le contexte des distributions à queues lourdes et de supposer la condition du premier ordre (2.2) pour F . Ainsi $\gamma > 0$, x_F est infini et l'ensemble des fonctions de répartition $F_{n,i}$ appartiennent au domaine d'attraction d'une même GEV d'indice des valeurs extrêmes γ . En conséquence, la convergence

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

est également vraie en remplaçant U par $U_{n,i}(t) := F_{n,i}^{\leftarrow}(1 - 1/t)$.

Pour étudier ce modèle, il est naturel de s'intéresser à l'estimation des paramètres γ et c . Cependant, comme la version intégrée de c ,

$$C(s) := \int_0^s c(u)du, \quad s \in [0, 1],$$

est plus facile à estimer, on préfère estimer cette fonction plutôt que c . L'estimation de cette fonction skedasis intégrée est néanmoins intéressante car elle permet de construire des tests sur C et par conséquent c et ainsi de reconnaître une tendance des extrêmes. Les estimateurs de γ et C sont construits avec les statistiques d'ordre $X_{1:n}^{(n)}, \dots, X_{n:n}^{(n)}$ et sont définis par

$$\hat{\gamma}_k^H := \frac{1}{k} \sum_{i=1}^k \log \left(\frac{X_{n-i+1:n}^{(n)}}{X_{n-k:n}^{(n)}} \right)$$

pour estimer γ , il s'agit de l'estimateur de Hill classique et par

$$\hat{C}_n(s) := \frac{1}{k} \sum_{i=1}^{\lfloor ns \rfloor} \mathbb{1}_{X_i^{(n)} > X_{n-k:n}^{(n)}},$$

la fonction de répartition empirique pour estimer C , où $\lfloor ns \rfloor$ désigne la partie entière de ns .

Pour établir la normalité asymptotique de ces deux estimateurs, il est nécessaire d'imposer d'autres conditions au modèle. La première est l'existence de deux fonctions A_1 et A_2 , tendant toutes deux vers 0 à l'infini telles que

$$\sup_{n \in \mathbb{N}} \max_{1 \leq i \leq n} \left| \frac{1 - F_{n,i}(x)}{1 - F(x)} - c\left(\frac{i}{n}\right) \right| = O\left(A_1\left(\frac{1}{1 - F(x)}\right)\right) \quad (2.6)$$

et

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A_2(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \quad (2.7)$$

avec $\rho < 0$.

La seconde condition concerne les variations de la fonction skedasis

$$\lim_{n \rightarrow \infty} \sqrt{k} \sup_{|u-v| \leq 1/n} |c(u) - c(v)| = 0. \quad (2.8)$$

Sous ces conditions, il est possible de déduire le théorème suivant :

Théorème 2.21. *Supposons (2.6), (2.7) et (2.8). Soit $k = k(n)$ une suite intermédiaire telle que $\sqrt{k}A_1(n/k) \rightarrow 0$ et $\sqrt{k}A_2(n/k) \rightarrow 0$ alors, sous une construction de Skorokhod, on a*

$$\sup_{0 \leq s \leq 1} \left| \sqrt{k}(\hat{C}_n(s) - C(s)) - B(C(s)) \right| \xrightarrow[n \rightarrow \infty]{} 0 \quad p.s$$

et

$$\sqrt{k}(\hat{\gamma}_k^H - \gamma) \xrightarrow[n \rightarrow \infty]{} \gamma N_0 \quad p.s,$$

avec N une variable aléatoire normale centrée réduite et B un pont brownien standard. De plus, B et N sont indépendants.

Des détails sur la représentation de Skorokhod sont donnés en section 3.1.

Remarque 2.3. *Si l'on souhaite s'affranchir de cette construction, ce théorème peut être interprété comme la convergence en loi de $\sqrt{k}(\hat{C}_n(s) - C(s))$ vers B dans $(\mathcal{L}^\infty(\mathbb{R}), \|\cdot\|_\infty)$ où B est un C -pont Brownien ainsi que la convergence en loi de $\sqrt{k}(\hat{\gamma}_k^H - \gamma)$ vers N avec $N \sim \mathcal{N}(0, \gamma^2)$ avec N et B indépendants.*

Ce modèle permet également l'estimation des quantiles de $F_{n,i}$ à partir de ceux de F . En effet, pour $\alpha \in [0, 1]$, on peut déduire de l'hypothèse (2.5) la relation

$$\alpha = 1 - F_{n,i} \left(U_{n,i} \left(\frac{1}{\alpha} \right) \right) \approx c \left(\frac{i}{n} \right) \left(1 - F \left(U_{n,i} \left(\frac{1}{\alpha} \right) \right) \right).$$

On peut alors en déduire une approximation du quantile de $F_{n,i}$ au niveau $1 - \alpha$,

$$U_{n,i} \left(\frac{1}{\alpha} \right) \approx U \left(\frac{c(i/n)}{\alpha} \right).$$

Il est alors possible d'estimer ce quantile par

$$X_{n-k:n} \left(\frac{k\hat{c}(i/n)}{n\alpha} \right)^{\hat{\gamma}_k^H}.$$

Cette estimation relève de l'estimation de c , qui n'est pas précisée dans ce modèle. En revanche Einmahl et al. préconise l'utilisation d'un estimateur à noyau [57] pour estimer cette fonction skedasis.

Les résultats précédents, ainsi que d'autres liés à l'étude des extrêmes, sont établis à l'aide de la théorie des processus empiriques. Plus particulièrement à la convergence faible de ces processus construits à l'aide d'un échantillon i.i.d et pouvant être indexés par un vaste choix d'ensembles ou de fonctions. Historiquement, le premier résultat concerne un processus empirique particulier, qui n'est autre que la fonction de répartition empirique sur \mathbb{R} centrée et normalisée à l'échelle du théorème central limite. En 1952, Donsker établit alors que, lorsque la taille de l'échantillon croît vers l'infini, ce processus converge vers un pont Brownien [28]. Ce résultat avait été conjecturé par Doob en 1949 [29] suite à une remarque de Bachelier. Aujourd'hui nous savons établir des résultats similaires pour bien d'autres classes de processus empiriques. Les résultats classiques sont recensés dans les ouvrages [61] ou [82].

3.1 Convergence faible

Avant de nous intéresser au comportement asymptotique de processus empiriques, intéressons nous d'abord à la notion de convergence en loi dans un espace métrique.

Soit (E, d) un espace métrique, on munit cet espace de sa tribu Borélienne \mathcal{E} . Soient $(\mu_n)_{n \in \mathbb{N}}$ et μ des mesures de probabilité sur (E, \mathcal{E}) . La convergence en loi se définit de la façon suivante.

Définition 3.1. On dit que μ_n converge faiblement vers μ , ce que l'on note $P_n \xrightarrow{d} P$, lorsque

$$\int_E f d\mu_n \xrightarrow{n \rightarrow \infty} \int_E f d\mu,$$

pour tout $f \in C_b(E)$, où $C_b(E)$ désigne l'ensemble des fonctions continues et bornées sur E .

Soient $(Z_n)_{n \in \mathbb{N}}$ et Z des variables aléatoires à valeurs dans E . On dit que Z_n converge en loi (ou en distribution) vers Z , lorsque pour tout $f \in C_b(E)$

$$\mathbb{E}(f(Z_n)) \xrightarrow{n \rightarrow \infty} \mathbb{E}(f(Z)),$$

ce qui correspond à la convergence faible de leurs distributions. On le note $Z_n \xrightarrow{d} Z$. Il existe plusieurs caractérisations de la convergence faible, certaines sont rassemblées dans le célèbre théorème Portmanteau.

Théorème 3.1 (Théorème 2.1 dans [9]). *Soient $(\mu_n)_{n \in \mathbb{N}}$ et μ des mesures de probabilité sur un espace séparable (E, \mathcal{E}) . Les assertions suivantes sont équivalentes*

1. $\mu_n \xrightarrow{d} \mu$.
2. $\mu_n(f) \rightarrow \mu(f)$ pour toute fonction f bornée et uniformément continue.
3. $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$ pour tout fermé F .
4. $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G)$ pour tout ouvert G .
5. $\mu_n(A) \rightarrow \mu(A)$ pour tout ensemble de μ -continuité A .

Dans un souci de concision, on notera $\mu(f)$ pour désigner $\int_E f d\mu$. Un ensemble de μ -continuité est un ensemble avec une frontière de μ -mesure nulle.

L'exemple le plus célèbre de convergence en distribution est le théorème central limite démontré par Laplace en 1809. Dans ce manuscrit nous utiliserons l'une de ses nombreuses extensions, le théorème de Lindeberg-Levy. Ce théorème concerne les tableaux triangulaires, c'est-à-dire les échantillons dont les lois peuvent changer avec n . Ce phénomène peut naturellement se produire dans le contexte des extrêmes car l'attention est portée sur des échantillons d'observations conditionnées au dépassement d'un seuil y_n choisi.

Théorème 3.2 (Lindeberg-Levy, Théorème 27.2 dans [8]). *Pour tout $n \geq 1$, soit $(Z_1^{(n)}, \dots, Z_{r_n}^{(n)})$ un échantillon indépendant (de loi pouvant changer avec n) tel que pour tout $1 \leq k \leq r_n$,*

$$\mathbb{E}(Z_k^{(n)}) = 0 \quad \text{et} \quad \sigma_{n,k} := \mathbb{E}\left(\left(Z_k^{(n)}\right)^2\right) < \infty.$$

Si pour tout $\varepsilon > 0$

$$\lim_{k \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|Z_k^{(n)}| \geq \varepsilon s_n} \left(Z_k^{(n)}\right)^2 dP = 0, \quad (3.1)$$

avec $s_n^2 = \sum_{k=1}^{r_n} \sigma_{n,k}^2$, alors

$$\frac{1}{s_n} \sum_{k=1}^{r_n} Z_k^{(n)} \xrightarrow[n \rightarrow \infty]{d} N,$$

avec $N \sim \mathcal{N}(0, 1)$.

La condition de Lindeberg (3.1) peut être remplacée par la condition dite de Lyapounov, sur les moments de $Z_1^{(n)}$ (Théorème 27.3 dans [8]).

Problèmes de mesurabilité

Dans la théorie des processus empiriques, il est courant d'être amené à considérer des objets pouvant être non mesurables. C'est typiquement le cas d'une application de $(\Omega, \mathcal{A}, \mathbb{P})$ vers un espace métrique (E, d) (non séparable) qui rend les boules mesurables mais pas tout les boréliens.

Exemple 3.1 (Exercice 3 page 48 dans [82]). Soit $Z : [0, 1] \rightarrow D([0, 1])$ défini par $Z(s) = \mathbf{1}_{[s, 1]}$. L'ensemble $D([0, 1])$ désigne l'ensemble des fonctions continues à droite et limite à gauche (Càdlàg) que l'on munit de la norme uniforme $\|\cdot\|_\infty$. Si l'on munit $[0, 1]$ de sa tribu Borélienne alors l'application Z n'est pas Borélienne. Car tout sous-ensembles $S \subset [0, 1]$ est l'image réciproque par Z de l'ensemble ouvert $\bigcup_{s \in S} B_s$ avec B_s la boule ouverte de centre $\mathbf{1}_{[s, 1]}$ et de rayon $1/2$. Si Z était Borélienne alors tous les sous-ensembles de $[0, 1]$ seraient Boréliens, ce qui n'est pas le cas.

Cet exemple met en exergue le fait que le processus empirique uniforme, c'est-à-dire l'application $t \mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i < t\}} - t$ pour $t \in [0, 1]$ and $(X_i)_{i=1}^n$ i.i.d de loi uniforme sur $[0, 1]$, peut ne pas être Borel mesurable dans $(D([0, 1]), \|\cdot\|_\infty)$. Il est donc nécessaire de d'étendre la notion de convergence faible pour de tels objets. Pour ce faire, nous utiliserons l'approche d'Hoffmann-Jørgensen [51].

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit Z une application de Ω dans $\bar{\mathbb{R}} := [-\infty, +\infty]$, on appelle espérance extérieure (ou intégrale extérieure) de Z la quantité

$$\mathbb{E}^*(Z) := \inf \{ \mathbb{E}(U) : U \geq Z, U \text{ mesurable telle que } \mathbb{E}(U) \text{ existe} \}.$$

On définit également la probabilité extérieure d'un sous-ensemble B de Ω par

$$\mathbb{P}^*(B) := \inf \{ \mathbb{P}(A) : B \subset A, A \in \mathcal{A} \}.$$

Ces notions coïncident avec l'espérance et la probabilité dans le cas où Z et B sont mesurables.

Remarque 3.1. Pour toute application $Z : \Omega \mapsto \bar{\mathbb{R}}$, $\inf \{ U \geq Z, U \text{ mesurable} \}$ est toujours atteint pour un certain Z^* mesurable. Ainsi, si $\mathbb{E}(Z^*)$ existe on a $\mathbb{E}^*(Z) = \mathbb{E}(Z^*)$. De plus, cet élément est unique au sens de l'égalité presque sûre.

On définit également l'espérance intérieure et la probabilité intérieure par

$$\mathbb{E}_*(Z) := \sup \{ \mathbb{E}(U) : U \leq Z, U \text{ mesurable telle que } \mathbb{E}(U) \text{ existe} \}, \text{ et}$$

$$\mathbb{P}_*(B) := \sup \{ \mathbb{P}(A) : A \subset B, A \in \mathcal{A} \}.$$

On a alors les identités suivantes $\mathbb{E}_*(Z) = -\mathbb{E}^*(-Z)$ et $\mathbb{P}_*(B) = 1 - \mathbb{P}^*(\Omega - B)$.

Ce concept d'espérance extérieure permet de généraliser la notion d'espérance à des applications non mesurables tout en conservant certaines propriétés de l'espérance. Par exemple, pour Z et Y deux applications de Ω dans $\bar{\mathbb{R}}$ on a $|\mathbb{E}^*(Z) - \mathbb{E}^*(Y)| \leq \mathbb{E}^*(|Z - Y|)$. Cependant il est important de rester prudent car certaines propriétés sont perdues. Ainsi l'espérance extérieure perd son caractère linéaire, mais reste sous-additive, c'est-à-dire que l'on a toujours $\mathbb{E}^*(Z + Y) \leq \mathbb{E}^*(Z) + \mathbb{E}^*(Y)$ avec égalité si l'une des deux applications est mesurable. Parmi les résultats courants qui se

trouvent modifiés pour l'espérance extérieure, on peut également citer le théorème de Fubini. Il reste toujours, pour une application Z définie sur $\Omega_1 \times \Omega_2$, l'inégalité $\mathbb{E}_1^*(\mathbb{E}_2^*(Z)) \leq \mathbb{E}^*(Z)$ mais l'égalité est perdue tant que Z n'est pas mesurable.

Armé de l'espérance extérieure, il est maintenant possible de définir la convergence faible pour des objets non mesurables.

Définition 3.2. Soit $(\Omega_n, \mathcal{A}_n, P_n)_{n \geq 1}$ une suite d'espaces probabilisés et $Z_n : \Omega_n \mapsto E$ des applications à valeurs dans un certain espace métrique E (les applications Z_n ne sont pas nécessairement Boréliennes). La suite Z_n converge faiblement vers une mesure Borélienne L si

$$\mathbb{E}^*(f(Z_n)) \xrightarrow{n \rightarrow \infty} \int f dL, \quad \text{pour tout } f \in C_b(E).$$

On le note $Z_n \xrightarrow{d} L$.

Comme l'espérance extérieure dépend de l'espace probabilisé, les \mathbb{E}^* de la définition précédente dépendent de n . Cependant on peut supposer, sans perdre de généralité, que tout les Z_n sont définis sur le même ensemble $\Omega := \prod_{n \geq 1} \Omega_n$ muni de la tribu produit. Ainsi, la définition de \mathbb{E}^* ne dépend plus de n . On entend alors $\mathbb{E}^*(f(Z_n))$ au sens de $\mathbb{E}^*(f(Z_n) \circ \mathfrak{p}_n)$ avec \mathfrak{p}_n la projection coordonnée de Ω dans Ω_n . Ceci ne change en rien la définition de la convergence faible car les projections coordonnées sont des applications dite parfaites. C'est-à-dire, pour une application T définie sur Ω , $(T \circ \mathfrak{p}_n)^* = T^* \circ \mathfrak{p}_n$. Pour plus de précisions sur ces applications, on peut se référer à [82] page 10.

Remarque 3.2. Comme $-f \in C_b(E)$ dès que $f \in C_b(E)$, lorsque $Z_n \xrightarrow{d} L$ la relation $\mathbb{E}_*(Z) = -\mathbb{E}^*(-Z)$ nous fournit

$$\mathbb{E}_*(f(Z_n)) - \mathbb{E}^*(f(Z_n)) \xrightarrow{n \rightarrow \infty} 0.$$

Une suite vérifiant cette dernière propriété est dite asymptotiquement mesurable.

Théorème 3.3 (Théorème Portmanteau, Théorème 7.6 dans [61]). *Les assertions suivantes sont équivalentes*

1. $Z_n \xrightarrow{d} L$.
2. $\liminf_{n \rightarrow \infty} \mathbb{P}_*(Z_n \in G) \geq L(G)$ pour tout ouvert G .
3. $\limsup_{n \rightarrow \infty} \mathbb{P}^*(Z_n \in F) \leq L(F)$ pour tout fermé F .
4. $\liminf_{n \rightarrow \infty} \mathbb{E}_*(f(Z_n)) \geq \int f dL$ pour toute fonction f minorée et semi-continue inférieurement.
5. $\limsup_{n \rightarrow \infty} \mathbb{E}^*(f(Z_n)) \leq \int f dL$ pour toute fonction f majorée et semi-continue supérieurement.
6. $\lim_{n \rightarrow \infty} \mathbb{P}^*(Z_n \in B) = \lim_{n \rightarrow \infty} \mathbb{P}_*(Z_n \in B) = L(B)$ pour tout Borélien B de L -continuité.

7. $\liminf_{n \rightarrow \infty} \mathbb{E}_*(f(Z_n)) \geq \int f dL$ pour toute fonction f Lipschitz, bornée et positive.

De plus, si L admet un support séparable les assertions précédentes sont également équivalentes à

$$\sup_{f \in BL_1} \left| \mathbb{E}^*(f(Z_n)) - \int f dL \right| \xrightarrow{n \rightarrow \infty} 0$$

L'ensemble BL_1 est l'ensemble des fonctions 1-Lipschitz bornées par 1.

On sait que la convergence en loi de variables aléatoires est impliquée par la convergence presque sûre et que la réciproque est fautive. Cependant il est toujours possible de représenter une convergence faible avec une convergence presque sûre sur un espace ad hoc.

Théorème 3.4 (Théorème de représentation de Dudley, Théorème 1.10.4 dans [82]). Soient $(Z_n)_{n \geq 1}$ une suite d'applications définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace métrique (E, d) et Z une application borélienne. Si $Z_n \xrightarrow{d} Z$ alors il existe un espace probabilisé $(\Omega', \mathcal{A}', \mathbb{P}')$ et des applications (Z'_n) et Z' définies sur Ω' et à valeurs dans E telles que

1. "**convergence presque uniforme :**" Pour tout $\varepsilon > 0$, il existe $A \in \mathcal{A}'$ tel que $\mathbb{P}'(A) \geq 1 - \varepsilon$ et $\sup_A d(Z'_n, Z') \rightarrow 0$
2. "**égalités en loi :**" $\mathbb{E}^*(f(Z'_n)) = \mathbb{E}^*(f(Z_n))$ pour toute fonction f bornée et pour tout n , relation également vraie pour Z et Z' .

Ainsi si on a $Z_n \xrightarrow{d} Z$ pour des variables aléatoires (i.e les Z_n sont mesurables) alors on a $Z'_n \xrightarrow{p.s} Z'$ avec $\mathcal{L}(Z_n) = \mathcal{L}(Z'_n)$ et $\mathcal{L}(Z) = \mathcal{L}(Z')$ mais Z_n et Z'_n ne sont pas définies sur le même espace probabilisé. C'est cette construction qui est mentionnée dans le Théorème 2.21 pour énoncer un résultat de convergence presque sûre.

Remarque 3.3. En l'absence de Borel mesurabilité, la notion de convergence presque sûre perd toute sa force, et n'est pas équivalente à la convergence presque uniforme (contre-exemple 1.9.4 page 54 dans [82]).

Stabilité de la convergence faible

Maintenant que la notion de convergence faible est définie, nous pouvons nous poser la question de sa stabilité. Il vient directement de la définition 3.1 la stabilité de la convergence au travers d'applications continues.

Théorème 3.5 (continuous mapping, Théorème 1.3.6 dans [82]). Soient $(Z_n)_{n \in \mathbb{N}}$ et Z des variables aléatoires à valeurs dans E et g une application continue de E dans un espace métrique H . Si $Z_n \xrightarrow{d} Z$ alors $g(Z_n) \xrightarrow{d} g(Z)$.

Les hypothèses de ce théorème peuvent être affaiblies. En effet, pour établir le résultat il suffit que Z soit presque sûrement un point de continuité de g . De plus, il est possible d'étendre ce théorème en considérant une suite de fonctions convergent convenablement vers une telle fonction g .

Théorème 3.6 (extended continuous mapping, Théorème 1.1.11 dans [82]). *Soient $E_n \subset E$, $g_n : E_n \rightarrow H$ et $g : E_0 \subset E \rightarrow H$ tels que, si $z_n \rightarrow z$ avec $z_n \in E_n$ pour tout $n > 1$ et $z \in E_0$ alors $g_n(z_n) \rightarrow g(z)$. Soient Z_n des variables aléatoires prenant respectivement leurs valeurs dans E_n et Z une variable aléatoire Borélienne à support séparable prenant ses valeurs dans E_0 . Alors, si $Z_n \xrightarrow{d} Z$ on a $g_n(Z_n) \xrightarrow{d} g(Z)$*

Les Théorèmes 3.5 et 3.6 sont également valables pour les convergences en probabilité et presque sûre. De plus ces théorèmes sont valables pour des Z_n non mesurables, auquel cas la convergence faible est à entendre avec la définition par espérance extérieure.

Théorème 3.7 (Delta méthode, Théorème 3.1 dans [81]). *Soient $\varphi : E_\varphi \subset \mathbb{R}^k \mapsto \mathbb{R}^m$ une application différentiable en un point θ de \mathbb{R}^k , $(Z_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans E_φ et Z une variable aléatoire à valeurs dans \mathbb{R}^k . Si*

$$r_n(Z_n - \theta) \xrightarrow[n \rightarrow \infty]{d} Z$$

pour une suite de nombres $r_n \rightarrow \infty$, alors

$$r_n(\varphi(Z_n) - \varphi(\theta)) \xrightarrow[n \rightarrow \infty]{d} \varphi'_\theta(Z).$$

φ'_θ désigne la matrice Jacobienne de φ en θ .

Ce résultat est, dans la pratique, utile pour établir des normalités asymptotiques d'estimateurs. Cependant, la théorie des processus empiriques est par nature de dimension infinie. C'est pourquoi la delta méthode doit être étendue à ce cadre. Il existe plusieurs notions de dérivation dans les espaces de Banach, celle qui permet de généraliser la delta méthode est la dérivation au sens de Hadamard.

Définition 3.3. Soit φ une application définie sur E_φ un sous-ensemble d'un espace normé E et à valeurs dans un espace normé H et considérons θ un élément de E . φ est dite Hadamard différentiable en θ s'il existe une application linéaire φ'_θ de E dans H telle que, pour tout $h \in E$ et toute suite $h_t \rightarrow h$,

$$\left\| \frac{\varphi(\theta + th_t) - \varphi(\theta)}{t} - \varphi'_\theta(h) \right\|_H \xrightarrow[t \rightarrow 0]{} 0.$$

La convergence étant prise lorsque t décroît vers 0. Si cette propriété ne peut pas être vérifiée pour tout h dans E mais seulement sur un sous-ensemble E_0 , alors φ'_θ existe uniquement sur ce sous-ensemble et φ est dite Hadamard différentiable tangentiellement à E_0 .

Lorsque l'objectif est d'établir une convergence faible, l'ensemble E_0 correspondra au support de la loi limite.

Théorème 3.8 (Delta méthode fonctionnelle, Théorème 20.8 dans [81]). *Soient E et H deux espaces normés et $\varphi : E_\varphi \subset E \mapsto H$ une application Hadamard différentiable en θ tangentiellement à E_0 . Considérons également $(Z_n)_{n \geq 1}$ des variables aléatoires*

à valeurs dans E_φ et Z une variable aléatoire à valeurs dans E_0 , telles que pour une certaine suite de nombres r_n tendant vers l'infini, $r_n(Z_n - \theta) \xrightarrow{d} Z$. Alors

$$r_n(\varphi(Z_n) - \varphi(\theta)) \xrightarrow{d} \varphi'_\theta(Z).$$

De plus, si φ est Hadamard différentiable sur E tout entier alors

$$r_n(\varphi(Z_n) - \varphi(\theta)) = \varphi'_\theta(r_n(Z_n - \theta)) + o_P(1).$$

3.2 Processus empiriques

Dans cette section, nous nous concentrons sur la convergence (faible) d'un type particulier de processus, les processus empiriques. Dans toute cette section nous considérons (X_1, \dots, X_n) un échantillon i.i.d dans un espace \mathcal{X} muni d'une tribu \mathcal{A} , de distribution commune P . On appelle mesure empirique associée à l'échantillon (X_1, \dots, X_n) la mesure

$$\mathbb{P}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i},$$

où δ_x désigne la masse de Dirac en x .

Soit \mathcal{F} une collection d'applications de \mathcal{X} dans \mathbb{R} , toutes de carré P -intégrable. On appelle processus empirique indexé par \mathcal{F} l'application

$$f \mapsto \mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - Pf).$$

Pour $f \in \mathcal{F}$ fixé, on sait déjà que

$$\mathbb{G}_n f \xrightarrow{d} \mathcal{N}(0, P(f - Pf)^2).$$

On souhaite ici trouver des conditions pour établir un résultat similaire mais uniforme en $f \in \mathcal{F}$. Pour montrer un tel résultat nous supposons que pour tout $x \in \mathcal{X}$

$$\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty.$$

Dans un tel cas \mathbb{G}_n peut être interprété comme un "élément aléatoire" de $\ell^\infty(\mathcal{F})$, c'est-à-dire une application de Ω vers $\ell^\infty(\mathcal{F})$. Le résultat auquel on cherche à aboutir peut être exprimé en une condition sur la classe \mathcal{F} .

Définition 3.4. Une classe de fonctions \mathcal{F} est appelée classe de Donsker (ou P -Donsker) si

$$\mathbb{G}_n \xrightarrow{d} \mathbb{G}, \quad \text{dans } \ell^\infty(\mathcal{F}),$$

avec \mathbb{G} une mesure borélienne séparable dans $\ell^\infty(\mathcal{F})$.

Comme c'est le cas pour le processus empirique on confondra communément le processus \mathbb{G} avec sa loi.

Le théorème centrale limite multivarié nous donne le seul candidat possible en

tant que limite \mathbb{G} . Il s'agit d'un processus Gaussien de moyenne 0 et de fonction de covariance donnée par

$$\mathbb{E}(\mathbb{G}f_1\mathbb{G}f_2) = Pf_1f_2 - Pf_1Pf_2.$$

Ce processus est appelé le P -pont Brownien.

Exemple 3.2. *En 1952, Donsker établit que la classe $\{\mathbf{1}_{]-\infty, x]}, x \in \mathbb{R}\}$ est Donsker et que le processus empirique uniforme indexé par cette classe converge vers le pont Brownien standard (i.e de covariance $\text{cov}(B_s, B_t) = \min(s, t) - st$ pour $0 \leq s, t \leq 1$).*

La propriété de Donsker dépend de la "taille" de la famille \mathcal{F} . Cette taille est évaluée en termes de contrôle de nombres de recouvrement par des boules ou des "crochets". Les détails sont rassemblés dans les sections 2.2 et 2.5 de [82]. Dans le présent manuscrit, nous nous intéressons à un certain type de classes de fonctions, les classes VC . Bien que le caractère VC d'une classe ne soit pas équivalent au caractère Donsker, il permet d'exhiber des classes de Donsker (sous réserve de finitude de certains moments) et est souvent plus facile à manipuler.

Les classes de Vapnik-Červonenkis (VC)

Soient \mathcal{C} un ensemble de sous-ensembles de \mathcal{X} et $\{x_1, \dots, x_n\}$ un ensemble de n points de \mathcal{X} . On dit que \mathcal{C} prélève un sous-ensemble $E \subset \{x_1, \dots, x_n\}$ s'il existe $C \in \mathcal{C}$ tel que $E = C \cap \{x_1, \dots, x_n\}$. On dit que \mathcal{C} pulvérise $\{x_1, \dots, x_n\}$ si \mathcal{C} peut prélever chacun des 2^n sous-ensembles de $\{x_1, \dots, x_n\}$. On appelle dimension VC (ou indice VC) de la classe \mathcal{C} , noté $V(\mathcal{C})$, le plus petit entier n tel que \mathcal{C} ne pulvérise aucun ensemble de taille n .

Définition 3.5. Une collection \mathcal{C} d'ensembles mesurables est appelée classe de Vapnik-Červonenkis d'ensembles (ou classe VC d'ensembles) lorsque sa dimension VC est finie.

Lorsqu'il n'y a pas d'ambiguïté on parle seulement de classe VC au lieu de classe VC d'ensembles.

Les ensembles de la forme $]-\infty, x]$ pour $x \in \mathbb{R}$ sont des classes VC . En effet cette classe ne peut pas pulvériser un ensemble de la forme $\{x_1, x_2\}$ avec $x_1 \leq x_2$ car il est impossible de prélever $\{x_2\}$. Sa dimension VC est alors égale à 2. De même, les quadrants de \mathbb{R}^d forment également un classe VC .

Remarque 3.4. *Le caractère VC d'une classe d'ensembles est stable avec les opérations ensemblistes usuelles. Par exemple, la classe des unions d'ensembles de deux classes VC est une classe VC d'ensembles. C'est également le cas pour l'intersection, le passage au complémentaire, l'image injective ou encore l'image réciproque. Une liste plus exhaustive est fournie par le Lemme 2.6.17 dans [82].*

Nous avons jusqu'ici considéré des classes VC constituées d'ensembles. Pour pouvoir établir des théorèmes de Donsker (Convergence faible de processus empiriques vers un mouvement Brownien) il sera aussi nécessaire d'établir le caractère VC de classes de fonctions.

Soit $f : \mathcal{X} \rightarrow \mathbb{R}$. On appelle sous-graphe de la fonction f le sous-ensemble de $\mathcal{X} \times \mathbb{R}$ défini par

$$\{(x, t) : t < f(x)\}.$$

Définition 3.6. Une collection \mathcal{F} de fonctions de \mathcal{X} dans \mathbb{R} est appelée classe VC -sous-graphe si l'ensemble des sous-graphes des fonctions de \mathcal{F} forme une classe VC d'ensembles de $\mathcal{X} \times \mathbb{R}$.

Il vient alors directement que les indicatrices d'ensembles VC forment des ensembles VC -sous-graphes. De plus, à l'image des classes VC d'ensembles, les classes VC -sous-graphes sont stables par certaines opérations sur les fonctions, comme par exemple l'addition, la multiplication ou la composition. D'autres propriétés de stabilité des classes VC sont recensées dans le Lemme 2.6.18 dans [82].

Ces classes VC nous fournissent un moyen de vérifier si une classe est Donsker, mais il est nécessaire de rajouter une hypothèse sur les moments.

Définition 3.7. On appelle enveloppe de \mathcal{F} la fonction

$$F : x \in \mathcal{X} \mapsto \sup_{f \in \mathcal{F}} |f(x)|.$$

Dans cette définition F n'est pas nécessairement un élément de \mathcal{F} et n'est pas nécessairement mesurable.

Théorème 3.9. Soit \mathcal{F} une classe VC -sous-graphe de fonctions mesurables d'enveloppe F . Alors \mathcal{F} est P -Donsker pour toute mesure de probabilité P telle que $P^*F^2 < \infty$.

Mesure empirique dans le contexte des extrêmes

Dans le cadre des extrêmes, nous serons amenés à considérer des mesures empiriques construites sur un sous-échantillon sélectionné à partir d'une condition sur les observations. Concrètement, dans l'approche *Peaks over Threshold* nous nous intéressons à la mesure empirique construite sur le sous-échantillon $\{Y_i, Y_i > Y_{n-k:n}\}$ pour un certain k . Dans le présent manuscrit, nous nous intéressons en particulier à l'estimation de quantiles conditionnels. Ainsi, considérons deux variables aléatoires X à valeurs dans \mathcal{X} et Y à valeurs dans \mathbb{R} définies sur des espaces probabilisés. Considérons également $((X_1, Y_1), \dots, (X_n, Y_n))$ n copies i.i.d générées avec la loi de (X, Y) . Les estimateurs que nous souhaitons étudier sont construits à partir des (X_i, Y_i) pour lesquels $Y_i > Y_{n-k:n}$. Ce qui revient à considérer la mesure empirique suivante

$$\Pi_{n,k} := \frac{1}{k} \sum_{\substack{i=1 \\ Y_i > Y_{n-k:n}}}^n \delta_{(X_i, Y_i)}.$$

Proposition 3.1. Soit $y > 1$ et $(\tilde{Y}_1, \dots, \tilde{Y}_k)$ un échantillon i.i.d de loi $\mathcal{L}((X, Y)|Y > y)$ et considérons la mesure empirique

$$\tilde{\Pi}_{k,y} := \frac{1}{k} \sum_{i=1}^k \delta_{(\tilde{X}_i, \tilde{Y}_i)}.$$

Alors, si la fonction de répartition de Y est continue, on a

$$\mathcal{L}(\Pi_{n,k}|Y_{n-k:n} = y) = \mathcal{L}(\tilde{\Pi}_{k,y}).$$

La continuité de F , la fonction de répartition de Y , est nécessaire pour établir ce résultat. Bien que ce soit le cas dans le contexte du présent manuscrit il est possible de s'en affranchir. En effet, le problème lorsque F n'est pas continue est la potentielle présence d'ex-æquo pour les Y_i ce qui induit que le sous-échantillon $((X_i, Y_i) | Y_i > Y_{n-k:n})_{i=1}^n$ peut ne pas être de taille k . Il faut donc réordonner l'échantillon $(X_i, Y_i)_{i=1}^n$ pour casser les égalités. Ce qui peut être fait par le procédé suivant :

1. Considérons un échantillon (y_1, \dots, y_n) généré avec la loi de Y .
2. On génère un échantillon $(\eta_i)_{i=1}^n$ i.i.d uniformément distribué sur $[0, 1]$.
3. Pour tout $i \in \{1, \dots, n\}$ on pose $U_i(y_i) = F^-(y_i) + \eta_i \mathbb{P}(Y = y_i)$. En notant $F^-(y_i)$ la limite à gauche de F au point y_i .
4. On considère alors le vecteur (R_1, \dots, R_n) qui sont les rangs de $(U_1(y_1), \dots, U_n(y_n))$ qui est presque sûrement sans égalité. On redéfinit alors $\Pi_{n,k}$ comme

$$\frac{1}{k} \sum_{\substack{i=1 \\ R_i > R_{n-k:n}}}^n \delta_{(X_i, Y_i)}.$$

On a alors, pour $(Y_i)_{i=1}^n$ i.i.d, que $(U_i)_{i=1}^n := (U_i(Y_i))_{i=1}^n$ est i.i.d, uniformément distribué sur $[0, 1]$ et tel que pour tout $i \in \{1, \dots, n\}$, $F^{\leftarrow}(U_i) = Y_i$ presque sûrement. De cette façon les échantillons $(Y_i)_{i=1}^n$ et $(U_i)_{i=1}^n$ sont presque sûrement ordonnés de la même manière. Ainsi, la Proposition 3.1 peut être réécrite en remplaçant Y par une variable uniforme U et en comprenant $y > 1$ comme $F^{\leftarrow}(1 - 1/u)$ avec $u \in [0, 1]$. On obtient ainsi le même résultat en s'affranchissant de la continuité de F .

3.3 Bootstrap et processus empiriques

On s'intéresse ici aux processus empiriques construits sur des échantillons bootstrap eux-même construits à partir d'un échantillon (X_1, \dots, X_n) . Soient $(\xi_{1,n}, \dots, \xi_{n,n})$ des variables aléatoires réelles que nous appellerons poids. On définit la mesure empirique pondérée (pondérée par les poids $(\xi_{1,n}, \dots, \xi_{n,n})$) par

$$\mathbb{P}_{\xi,n} := \frac{1}{n} \sum_{i=1}^n \xi_{i,n} \delta_{X_i}.$$

Si on choisit les poids i.i.d alors $\mathbb{P}_{\xi,n}$ n'est plus une mesure de probabilité. Pour éviter cet écueil, on suppose que $\sum_{i=1}^n \xi_{i,n} = n$ presque sûrement, les poids ne sont alors pas indépendants. Nous portons notre attention sur des poids positifs et échangeables, c'est-à-dire que $(\xi_{1,n}, \dots, \xi_{n,n}) \stackrel{d}{=} (\xi_{\sigma(1),n}, \dots, \xi_{\sigma(n),n})$ pour toute permutation σ de $\{1, \dots, n\}$.

Remarque 3.5. *Le cadre classique du bootstrap discuté par Efron [34] où l'échantillon bootstrap est construit par tirage avec remise dans l'échantillon observé correspond au cas où le vecteur des poids $(\xi_{1,n}, \dots, \xi_{n,n})$ suit la loi multinomiale $\mathcal{M}(n, \vec{a})$ avec $\vec{a} := (1/n, \dots, 1/n)$. Ce bootstrap empirique est investigué dans le contexte des extrêmes par Resnick [75] dans sa section 6.4.*

On définit ensuite le processus empirique bootstrap par

$$\mathbb{G}_{\xi,n} := \sqrt{n} \left(\mathbb{P}_{\xi,n} - \bar{\xi}_n \mathbb{P}_n \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_{i,n} - \bar{\xi}_n) \delta_{X_i},$$

avec $\bar{\xi}_n$ désignant la moyenne des $\xi_{i,n}$. Comme dans le cas du processus empirique classique, on cherche alors à obtenir la convergence faible de $\mathbb{G}_{\xi,n}$ vers un processus Gaussien dans $\ell^\infty(\mathcal{F})$ pour une certaine classe de fonctions \mathcal{F} . Ce processus empirique bootstrap dispose de deux sources d'aléas, l'échantillon et les poids. Formellement, nous avons déjà mentionné que tous les X_i pouvaient être considérés comme définis sur un même espace (Ω, \mathcal{A}, P) (quitte à considérer l'espace, la tribu et la mesure produit), ainsi nous pouvons de façon similaire considérer que pour chaque $n \geq 1$, les vecteurs $(\xi_{1,n}, \dots, \xi_{n,n})$ sont tous définis sur un même espace probabilisé $(\mathcal{Z}, \mathcal{G}, Q)$. De cette façon, le processus empirique bootstrap est défini sur $(\Omega \times \mathcal{Z}, \mathcal{A} \times \mathcal{G}, P \otimes Q)$. Le but étant d'obtenir des tests de validation, ces processus seront couramment étudiés à échantillon fixé et les poids seront la seule source d'aléa. Nous étudions donc la convergence faible de $\mathbb{G}_{\xi,n}$ conditionnellement aux données. On note $\mathbb{E}_\xi(\mathbb{G}_{\xi,n})$ l'espérance en ξ avec (X_1, \dots, X_n) fixé.

Définition 3.8. On dit que la suite $\mathbb{G}_{\xi,n}$ converge faiblement conditionnellement vers \mathbb{G} si $\mathbb{G}_{\xi,n}$ est asymptotiquement mesurable et si

$$\sup_{h \in BL_1} |\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n})) - \mathbb{E}(h(\mathbb{G}))| \xrightarrow{\mathbb{P}^*} 0.$$

Pour établir ce type de convergence, il n'est pas suffisant d'établir que la classe de fonctions \mathcal{F} soit Donsker. Il est nécessaire d'ajouter des conditions tant sur les poids que sur la classe de fonctions. On suppose alors que

$$\sup_{n \geq 1} \|\xi_{1,n} - \bar{\xi}_n\|_{2,1} < \infty, \quad (3.2)$$

$$\frac{1}{\sqrt{n}} \mathbb{E} \left(\max_{1 \leq i \leq n} |\xi_{i,n} - \bar{\xi}_n| \right) \rightarrow 0, \quad (3.3)$$

$$\frac{1}{n} \sum_{i=1}^n (\xi_{i,n} - \bar{\xi}_n)^2 \xrightarrow{\mathbb{P}} c^2 > 0. \quad (3.4)$$

La "norme" dans la condition (7.7) est définie par

$$\|\xi\|_{2,1} := \int_0^\infty \sqrt{\mathbb{P}(|\xi| > x)} dx,$$

ce n'est pas rigoureusement une norme mais nous pouvons l'utiliser pour définir une topologie plus faible que celle induite par la norme $L^2(P)$. Quant à la condition sur la classe \mathcal{F} , il est question de mesurabilité.

Définition 3.9. Une classe \mathcal{F} de fonctions mesurables définies sur (Ω, \mathcal{A}, P) est dite mesurable (ou P -mesurable) si l'application

$$(X_1, \dots, X_n) \mapsto \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n e_i f(X_i) \right|$$

est mesurable dans la tribu complétée de $(\Omega^n, \mathcal{A}^n, P^n)$ pour tout n et tout $(e_1, \dots, e_n) \in \mathbb{R}^n$.

Si \mathcal{F} contient un sous-espace \mathcal{G} dénombrable tel que pour toute fonction $f \in \mathcal{F}$ il existe une suite $(g_n)_{n>1} \subset \mathcal{G}$ telle que g_n converge simplement vers f , alors \mathcal{F} est P -mesurable pour tout P . Ceci s'applique aux familles d'indicatrices d'intervalles de \mathbb{R} ou de quadrants de \mathbb{R}^d qui sont les principales familles que nous allons considérer dans ce manuscrit.

Théorème 3.10 (Théorème 3.6.13 dans [82]). *Soit \mathcal{F} une classe de Donsker de fonctions mesurables telle que l'ensemble*

$$\mathcal{F}_\delta := \left\{ f - g : f, g \in \mathcal{F}, P(f - g - P(f - g))^2 < \delta^2 \right\}$$

est mesurable pour tout $\delta > 0$. Supposons que pour tout $n \geq 1$, les poids $(\xi_{1,n}, \dots, \xi_{n,n})$ sont positifs, échangeables, indépendants de (X_1, \dots, X_n) et tels que les conditions (3.2), (3.3) et (3.4) soient satisfaites. Alors

$$\sup_{h \in BL_1} |\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n})) - \mathbb{E}(h(c\mathbb{G}))| \xrightarrow{\mathbb{P}^*} 0 \text{ et } \mathbb{G}_{\xi,n} \text{ est asymptotiquement mesurable.}$$

Ici, \mathbb{G} désigne la limite du processus empirique \mathbb{G}_n indexé par \mathcal{F} .

Dans le cas du bootstrap d'Efron, les poids $(\xi_{1,n}, \dots, \xi_{n,n})$ sont multinomiaux de paramètres n et $(1/n, \dots, 1/n)$. Les conditions (3.2), (3.3) et (3.4) sont alors vérifiées pour $c = 1$. Ainsi pour toute classe \mathcal{F} de fonctions mesurables d'enveloppe finie on a \mathcal{F} est Donsker si et seulement si $\mathbb{G}_{\xi,n}$ converge faiblement conditionnellement vers \mathbb{G} .

3.4 Variations régulières

L'étude des extrêmes est intrinsèquement liée à la considération de fonctions à variations régulières, leur analyse est donc fondamentale dans la compréhension de phénomènes extrêmes.

En effet, que ce soit par l'intermédiaire de la fonction U dans la condition du premier ordre (2.2) ou de la fonction auxiliaire du second ordre A , les fonctions à variations régulières apparaissent de manière récurrente dans les caractérisations liées au comportement des extrêmes. Plus précisément, la condition du premier ordre (2.2) établit l'équivalence entre " $F \in D_\gamma$ " et " U est à variation régulière d'indice γ au premier ordre". De plus les conditions de variations régulières sont utilisées pour étudier le biais des estimateurs.

Cette section recense les définitions et résultats liés aux variations régulières utilisées dans ce manuscrit. Les résultats ainsi présentés sont essentiellement des bornes qui permettent de contrôler les vitesses de convergence dans les différentes expressions des conditions du premier et second ordre. Pour une revue plus détaillée du sujet, on peut se référer à la monographie de Bergham et al. [10] ou à l'appendice B dans [22].

Variations régulières

Définition 3.10. Soit f une fonction Lebesgue mesurable, à valeurs réelles et strictement positive dans un voisinage de l'infini. Alors f est dite à variations régulières

à l'infini s'il existe $\alpha \in \mathbb{R}$ tel que

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha, \quad x > 0. \quad (3.5)$$

On le note $f \in RV_\alpha$.

Les variations régulières considérées ici sont toutes en l'infini, ainsi le terme "à l'infini" de la définition précédente est sous-entendu. De plus, α est appelé indice des variations régulières, on parle ainsi de variations régulières d'indice α . Les fonctions à variations régulières d'indice 0 sont appelées fonctions à variations lentes.

Dès que la limite de $f(tx)/f(t)$ est finie alors les variations régulières sont le seul comportement possible pour les fonctions satisfaisant le préambule de la Définition 3.10.

Théorème 3.11. *Soit f une fonction réelle mesurable au sens de Lebesgue et strictement positive sur un voisinage de l'infini telle que*

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)}$$

existe et est fini pour tout x appartenant à un ensemble de mesure de Lebesgue strictement positive. Alors, il existe $\alpha \in \mathbb{R}$ tel que $f \in RV_\alpha$.

Sans ajouter d'hypothèses à la définition de variations régulières, il est possible de donner plus de détails sur la convergence de $f(tx)/f(t)$. Déjà, la convergence (3.5) est uniforme sur les compacts de \mathbb{R}_+ .

Théorème 3.12. *Soit $f \in RV_\alpha$, alors la convergence (3.5) est vraie uniformément en $x \in [a, b]$ pour tout $0 < a < b < \infty$.*

Les possibles cas de convergence non uniforme sont donc en 0 et en l'infini, cependant il est possible de contrôler la vitesse de convergence en l'infini grâce aux bornes de Potter [69].

Proposition 3.2 (Potter (1942)). *Soient $f \in RV_\alpha$ et $\delta_1, \delta_2 > 0$. Alors il existe t_0 (dépendant de δ_1 et δ_2) tel que pour tout $t > t_0$, $tx > t_0$, on a*

$$(1 - \delta_1)x^\alpha \min(x^{\delta_2}, x^{-\delta_2}) < \frac{f(tx)}{f(t)} < (1 + \delta_1)x^\alpha \max(x^{\delta_2}, x^{-\delta_2}).$$

On peut remarquer que la réciproque est également vraie. Ainsi si une fonction satisfait la propriété précédente alors elle est à variations régulières d'indice α . On utilise parfois ce résultat sous une autre forme, plus compacte, due à H. Drees [30] qui énonce que $|f(tx)/f(t) - x^\alpha| \leq \varepsilon \max(x^{\alpha+\delta}, x^{\alpha-\delta})$, pour tout $\varepsilon, \delta > 0$ et t assez grand.

Variations régulières étendues

Pour étendre la notion de variations régulières il est possible de faire le parallèle avec le taux de variation

$$\frac{f(x+h) - f(x)}{h}.$$

Ordinairement, on étudie ce taux à x fixé avec h tendant vers 0, mais ici on s'intéresse plutôt au comportement de ce taux lorsque x tend vers l'infini à h fixé. Une généralisation de la définition 3.10 serait alors l'existence d'une fonction a , strictement positive telle que

$$\lim_{t \rightarrow \infty} \frac{f(tx) - f(t)}{a(t)} \quad \text{existe et est finie pour tout } x > 0.$$

On sait alors que si une telle limite existe dans \mathbb{R} alors elle est nécessairement de la forme

$$c \frac{x^\gamma - 1}{\gamma}, \quad x > 0 \text{ avec } c \neq 0 \text{ et } \gamma \in \mathbb{R},$$

ou $c \log(x)$, qui correspond au cas $\gamma = 0$. De plus, il est toujours possible de choisir une fonction $a \in RV_\gamma$.

Définition 3.11. Soit f une fonction réelle mesurable. La fonction f est dite à variations régulières étendues s'il existe une fonction $a : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ et un réel $\gamma \in \mathbb{R}$ tels que pour tout $x > 0$ on a

$$\lim_{t \rightarrow \infty} \frac{f(tx) - f(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}. \quad (3.6)$$

On le note $f \in ERV$ ou $f \in ERV_\gamma$. La fonction a est appelée fonction auxiliaire de f .

Lorsque $\gamma = 0$ la partie droite de l'équation est à comprendre comme $\log(x)$. Parfois, ces fonctions à variations régulières étendues d'indice 0 sont dites appartenant à la classe Π , ce qu'on note $f \in \Pi$ ou $f \in \Pi(a)$.

Ainsi la classe Π correspond aux fonctions à variations régulières étendues d'indice 0. En ce sens, la classe Π étend celle des fonctions à variations lentes. Comme dans le cadre des variations régulières non étendues, il est possible de donner plus d'informations sur les convergences précédentes.

Théorème 3.13. Soit $f \in \Pi$, la convergence (3.6) a lieu uniformément en $x \in [a, b]$ pour tout $0 < a < b < \infty$.

À l'image du théorème de Drees pour les fonctions à variations régulières, on dispose d'un contrôle sur la distance avec la limite dans le cadre des fonctions dans ERV .

Théorème 3.14 (Drees (1998)). Soit $f \in ERV_\gamma$. Pour tout ε et $\delta > 0$ il existe t_0 tel que pour tout t , $tx > t_0$ on a

$$\left| \frac{f(tx) - f(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma} \right| \leq \varepsilon x^\gamma \max(x^\delta, x^{-\delta})$$

avec

$$a_0(t) := \begin{cases} \gamma f(t), & \gamma > 0 \\ -\gamma(f(\infty) - f(t)), & \gamma < 0 \\ f(t) - t^{-1} \int_0^t f(s) ds, & \gamma = 0. \end{cases}$$

Variations régulières au second ordre

Dans cette section, il est question de quantifier les vitesses de convergence de la définition 3.11. C'est-à-dire de s'intéresser à des limites de la forme

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

Stadtmüller et de Haan [23] montrent que si, pour des fonctions a et A strictement positives, une telle limite existe, est finie et n'est pas un multiple de $(x^\gamma - 1)/\gamma$ alors elle est nécessairement de la forme

$$c \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds$$

avec $c \in \mathbb{R}$ et $\rho \leq 0$. De plus $A \in RV_\rho$ et

$$\lim_{t \rightarrow \infty} \frac{\frac{a(tx)}{a(t)} - x^\gamma}{A(t)} = cx^\gamma \frac{x^\gamma - 1}{\gamma}.$$

Définition 3.12. Une fonction f est dite à variations régulières au second ordre (ce qui est noté $f \in 2ERV$ ou $f \in 2ERV_{\gamma,\rho}$) s'il existe deux fonctions, a strictement positive et A de signe constant, telles que pour tout $x > 0$

$$\lim_{t \rightarrow \infty} \frac{\frac{f(tx) - f(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds := \psi_{\gamma,\rho}(x),$$

avec $\rho \leq 0$.

On appelle ρ paramètre du second ordre.

Cette fois encore, il est possible de borner la distance entre le ratio et la limite dans un voisinage de l'infini.

Théorème 3.15 (Drees (1998)). *Soit $f \in 2ERV_{\gamma,\rho}$. Pour tout $\varepsilon, \delta > 0$ il existe t_0 , dépendant de ε et δ , tel que pour tout $t, tx > t_0$,*

$$\left| \lim_{t \rightarrow \infty} \frac{\frac{f(tx) - f(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma}}{A_0(t)} - \psi_{\gamma,\rho} \right| \leq \varepsilon \max(x^{\gamma+\rho+\delta}, x^{\gamma+\rho+\delta}),$$

avec les fonctions a_0 et A_0 adéquatement choisies.

De plus, $A_0 \in RV_\rho$ et

$$\left| \frac{\frac{a_0(tx)}{a_0(t)} - x^\gamma}{A_0(t)} - x^\gamma \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon \max(x^{\gamma+\rho+\delta}, x^{\gamma+\rho+\delta}).$$

Les fonctions a_0 et A_0 sont respectivement équivalentes à a et A . Elles ont des expressions explicites, dépendantes de γ, ρ et f , que l'on peut trouver dans [22] Theorem B.3.10 page 392-393.

Il existe une forme simple de variations régulières au second ordre qui relève de la définition classique des variations régulières et non de celles étendues. Soit $f \in RV_\gamma$

pour un certain $\gamma \in \mathbb{R}$. Supposons qu'il existe une fonction A , de signe constant, telle que la limite

$$\lim_{t \rightarrow \infty} \frac{\frac{f(tx)}{f(t)} - x^\gamma}{A(t)}, \quad x > 0$$

existe et n'est pas constante. Alors l'application $t \mapsto t^{-\gamma}f(t)$ est à variations régulières étendues. On a alors

$$\lim_{t \rightarrow \infty} \frac{\frac{f(tx)}{f(t)} - x^\gamma}{A(t)} = \frac{x^\rho - 1}{\rho}, \quad x > 0$$

avec $\rho \leq 0$. On dit alors que f est à variations régulières au second ordre. Ce que l'on note $f \in 2RV$ ou $f \in 2RV_{\gamma,\rho}$.

Couplage optimal

Nombre de résultats de théorie des valeurs extrêmes, notamment ceux présentés en section 2, reposent sur la théorie des processus empiriques. Un des travaux présentés dans cette thèse a été de proposer une nouvelle approche basée sur une méthode de couplage, outil important en théorie des probabilités ([64, 80]), qui a connu un essor important lors des dernières décennies. Ces arguments de couplage sont formalisés à l'aide de la notion de distance de Wasserstein, notion introduite et étudiée, par exemple, dans les livres de Villani [83] ou Ambrosio et al [1]. Ces ouvrages traitent de la notion de distance de Wasserstein tant dans le contexte de mesure de probabilité que dans le cadre général du transport optimal introduit par Monge en 1781 [2] ensuite développé par Kantorovich [58, 59]. L'article de Panaretos et Zemel [66] et les références qu'il contient permettent de donner une vue d'ensemble de l'utilisation de ce formalisme dans le contexte des statistiques.

Dans cette section, nous considérerons un espace métrique (\mathcal{X}, d) que l'on munit de sa tribu Borélienne $\mathcal{B}(\mathcal{X})$. On notera également $\mathcal{M}(\mathcal{X})$ l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

4.1 Couplage optimal et distance de Wasserstein

Principales définitions et propriétés

Un couplage entre deux mesures de probabilité $P_1, P_2 \in \mathcal{M}(\mathcal{X})$ est un couple de variables aléatoires X_1, X_2 , à valeurs dans \mathcal{X} et défini sur un espace probabilisé commun, tel que $X_1 \sim P_1$ et $X_2 \sim P_2$.

Définition 4.1. On définit la distance de Wasserstein d'ordre $p \in [1, +\infty]$ entre deux mesures P_1 et $P_2 \in \mathcal{M}(\mathcal{X})$ par

$$W_p(P_1, P_2) := \inf \left\{ \|d(X_1, X_2)\|_{L^p} : X_1 \sim P_1, X_2 \sim P_2 \right\}.$$

Lorsque (\mathcal{X}, d) est complet et séparable, il existe un couplage qui atteint l'infimum de la définition de la distance de Wasserstein. C'est-à-dire il existe (X_1, X_2) tel que $W_p(P_1, P_2) = \|d(X_1, X_2)\|_{L^p}$. Un tel couplage est appelé couplage optimal. Dans certains cas, il est possible d'expliciter cette distance de Wasserstein. Cependant ce calcul explicite est impossible dans une majorité des cas.

Exemple 4.1. • Pour x et $y \in \mathcal{X}$, $W_p(\delta_x, \delta_y) = d(x, y)$. Dans ce cas, la distance ne dépend pas de p , mais ce n'est pas la règle en général.

- Considérons P_1 et $P_2 \in \mathcal{M}(\mathbb{R})$ avec F_1 et F_2 leurs fonctions de répartition respectives. Alors, pour $p < \infty$

$$W_p(P_1, P_2) = \left(\int_0^1 |F_1^{\leftarrow}(u) - F_2^{\leftarrow}(u)|^p du \right)^{1/p},$$

et si $p = \infty$,

$$W_\infty(P_1, P_2) = \sup_{0 < u < 1} |F_1^{\leftarrow}(u) - F_2^{\leftarrow}(u)|.$$

La distance de Wasserstein définie en 4.1 n'est pas à proprement parler une distance sur l'espace $\mathcal{M}(\mathcal{X})$. En effet, elle peut prendre la valeur ∞ il est donc nécessaire de restreindre l'espace pour avoir une distance bien définie.

Définition 4.2. On appelle espace de Wasserstein d'ordre p le sous espace de $\mathcal{M}(\mathcal{X})$

$$\mathcal{W}_p(\mathcal{X}) := \left\{ P \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} d(x, x_0)^p P(dx) < \infty \right\}$$

avec x_0 un point arbitraire dans \mathcal{X} . L'espace ainsi défini ne dépend pas du choix de l'origine x_0 .

La distance W_p est finie sur \mathcal{W}_p , elle y est donc bien une distance. Ce qui fait alors de $(\mathcal{W}_p(\mathcal{X}), W_p)$ un espace métrique. Cet espace peut être vu comme l'ensemble des probabilités avec un moment d'ordre p fini.

Exemple 4.2. • Soit $\mathcal{X} = \mathbb{R}^d$ muni de sa norme euclidienne. Alors, lorsque $p < \infty$,

$$\mathcal{W}_p(\mathbb{R}^d) = \left\{ P \in \mathcal{M}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|x\|^p P(dx) < \infty \right\}.$$

Lorsque $p = \infty$ alors $\mathcal{W}_p(\mathbb{R}^d)$ est l'ensemble des mesures à support borné.

- On a vu que $(\mathcal{W}_p(\mathcal{X}), W_p)$ est un espace métrique, ainsi il est possible de considérer l'espace de Wasserstein $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$. On appelle cet espace l'espace de Wasserstein au second ordre. Un élément $\mathcal{P} \in \mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$ peut être vu comme la distribution d'une mesure de probabilité aléatoire Π sur \mathcal{X} . La condition d'intégrabilité pour l'appartenance à l'espace de Wasserstein de second ordre se traduit alors comme

$$\mathbb{E} \left(\int_{\mathcal{X}} d(x, x_0)^p \Pi(dx) \right) < \infty.$$

On construit alors naturellement une distance de Wasserstein (au second ordre), que l'on note $W_p^{(2)}$, entre \mathcal{P}_1 et $\mathcal{P}_2 \in \mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$ défini par

$$W_p^{(2)}(\mathcal{P}_1, \mathcal{P}_2) := \inf \left\{ \|W_p(\Pi_1, \Pi_2)\|_{L^p} : \Pi_1 \sim \mathcal{P}_1, \Pi_2 \sim \mathcal{P}_2 \right\}$$

pour $p \in [1, \infty]$.

- Dans certains cas, il peut être utile de munir l'espace \mathcal{X} de la distance d définie par $d(x, y) = \mathbb{1}_{x \neq y}$, $x, y \in \mathcal{X}$. L'espace (\mathcal{X}, d) n'est alors plus séparable. On ne peut donc pas garantir l'existence d'un couplage optimal. Néanmoins, d'autres formes de couplage nous permettent d'obtenir des informations sur la distance de Wasserstein alors définie par

$$W_p(P_1, P_2) = \inf \{ \mathbb{P}(X_1 \neq X_2) : X_i \sim P_i, i = 1, 2 \}.$$

Le couplage qui nous sera utile ici est un couplage maximal défini par rapport à la distance en variation totale. On appelle distance en variation totale entre deux mesures de probabilité P_1 et P_2 sur \mathcal{X} la distance définie par

$$\|P_1 - P_2\|_{TV} := \sup_{B \text{ borélien de } \mathcal{X}} |P_1(B) - P_2(B)|.$$

Cette distance en variation totale vérifie une propriété de couplage optimal.

Proposition 4.1. Soient P_1 et P_2 deux mesures de probabilité sur \mathcal{X} . Il existe un couplage (X_1, X_2) entre P_1 et P_2 tel que

$$\|P_1 - P_2\|_{TV} = \mathbb{P}(X_1 \neq X_2).$$

À l'image des espaces L^p les espaces de Wasserstein sont également emboîtés. Si $p < q$ alors $W_p \leq W_q$ ainsi on a l'inclusion $\mathcal{W}_q(\mathcal{X}) \subset \mathcal{W}_p(\mathcal{X})$. La distance W_1 est la plus souple de toutes alors que W_∞ est la plus restrictive.

Propriétés topologiques

Avant d'aller plus avant dans la description des distances et espaces de Wasserstein, il est important de s'intéresser à la caractérisation de la convergence dans ces espaces.

Définition 4.3. Soit $(P_n)_{n \in \mathbb{N}}$ une suite dans $\mathcal{W}_p(\mathcal{X})$ et $P \in \mathcal{W}_p(\mathcal{X})$. On dit que P_n converge vers P dans $\mathcal{W}_p(\mathcal{X})$ (ou au sens de Wasserstein) si $W_p(P_n, P) \rightarrow 0$

Le théorème suivant caractérise la convergence au sens de Wasserstein.

Théorème 4.1. Soient (\mathcal{X}, d) un espace Polonais et $p \in [1, \infty[$. Soit $(P_n)_{n \in \mathbb{N}}$ une suite dans $\mathcal{W}_p(\mathcal{X})$ et $P \in \mathcal{W}_p(\mathcal{X})$. Alors, pour un $x_0 \in \mathcal{X}$ (donc tous), les assertions suivantes sont équivalentes.

1. P_n converge vers P dans $\mathcal{W}_p(\mathcal{X})$.
2. $P_n \xrightarrow{d} P$ et $\int_{\mathcal{X}} d(x, x_0)^p P_n(dx) \xrightarrow{n \rightarrow \infty} \int_{\mathcal{X}} d(x, x_0)^p P(dx)$.
3. Pour toute fonction continue φ telle que $|\varphi(x)| \leq C(1 + d(x, x_0)^p)$, $C \in \mathbb{R}$, on a $\int_{\mathcal{X}} \varphi dP_n \rightarrow \int_{\mathcal{X}} \varphi dP$.

Ainsi la convergence vers 0 de la distance de Wasserstein implique la convergence des moments mais également la convergence faible au sens de la section 3.1. Par conséquent, pour montrer la convergence en distribution de variables aléatoires, il suffit d'établir que la distance de Wasserstein entre les distributions tend vers 0. Il est également possible de comparer le comportement asymptotique de suites de mesures.

Proposition 4.2. *Soient (\mathcal{X}, d) un espace Polonais, $p \in [1, \infty[$, $(P_n)_{n \in \mathbb{N}}$ et $(P_n^*)_{n \in \mathbb{N}}$ deux suites dans $\mathcal{W}_p(\mathcal{X})$ et $P \in \mathcal{W}_p(\mathcal{X})$. Alors si P_n^* converge faiblement vers P et $W_p(P_n, P_n^*) \rightarrow 0$ alors P_n converge faiblement vers P .*

Un autre avantage à munir $\mathcal{W}_p(\mathcal{X})$ de la distance W_p est qu'il peut alors hériter de certaines propriétés de (\mathcal{X}, d) .

Théorème 4.2. • *Si (\mathcal{X}, d) est un espace complet (resp. complet séparable) alors $(\mathcal{W}_p(\mathcal{X}), W_p)$, pour $p \in [1, \infty[$, est également complet (resp. complet séparable). De plus, toute mesure de probabilité sur \mathcal{X} peut être approchée par une suite de mesures de probabilité à support fini.*

• *Si (\mathcal{X}, d) est compact alors il en est de même pour $(\mathcal{W}_p(\mathcal{X}), W_p)$.*

Bien que l'espace de Wasserstein hérite de l'espace sous-jacent sa compacité, ce n'est pas le cas pour la compacité locale.

Dualité de Kantorovich-Rubinstein

La distance de Wasserstein peut être caractérisée en termes de couplage optimal.

$$W_p^p(P, Q) = \inf \left\{ \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p \mu(dx, dy) : \mu \in \Gamma(P, Q) \right\}$$

avec $\Gamma(P, Q)$ l'ensemble des mesures sur $\mathcal{X} \times \mathcal{X}$ qui admettent P et Q comme lois marginales. Ici nous nous intéressons à une autre caractérisation, une définition duale grâce à un supremum sur un ensemble de fonctions tests. Considérons l'ensemble $\Omega_p(d)$ défini par

$$\Omega_p(d) := \{(\phi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{X}) : \phi(x) - \psi(y) \leq d^p(x, y), x, y \in \mathcal{X}\}$$

avec $C_b(\mathcal{X})$ l'ensemble des fonctions continues et bornées sur \mathcal{X} à valeurs dans \mathbb{R} .

Théorème 4.3 (dualité de Kantorovich). *Soient (\mathcal{X}, d) un espace Polonais et $p \in [1, \infty[$. Soient P et Q deux mesures de probabilité sur \mathcal{X} , on a alors*

$$W_p^p(P, Q) = \sup_{(\phi, \psi) \in \Omega_p(d)} \left(\int_{\mathcal{X}} \phi dQ - \int_{\mathcal{X}} \psi dP \right)$$

Cette expression peut être étendue à un cadre plus général (détaillé dans le théorème 5.10 dans [83]). Le cas de la distance de Kantorovich-Rubinstein W_1 est singulier vis-à-vis de cette dualité.

Théorème 4.4 (Dualité de Kantorovich-Rubinstein). *Soient P et $Q \in \mathcal{W}_p(\mathcal{X})$, on a alors*

$$W_1(P, Q) = \sup_{\|\varphi\|_{Lip} \leq 1} \left\{ \int_{\mathcal{X}} \varphi dP - \int_{\mathcal{X}} \varphi dQ \right\}$$

où $\|\varphi\|_{Lip} = \sup_{x, y} \frac{|f(x) - f(y)|}{|x - y|}$ avec φ une fonction Lipschitz.

4.2 Applications aux statistiques

L'utilisation principale que nous avons de cette approche par couplage est l'étude du comportement asymptotique de divers estimateurs. Considérons un estimateur $\hat{\theta}_n$ d'un certain paramètre θ . L'utilisation direct de cette méthodologie pour montrer la normalité asymptotique de $\hat{\theta}_n$ est d'établir, pour un certain $p \in [1, \infty[$, la convergence vers 0 de $W_p(P_{\sqrt{n}(\hat{\theta}_n - \theta)}, P_N)$ avec P_N une loi normale. Dans certain cas il est même possible de donner une borne explicite pour la distance de Wasserstein, avec la méthode de Stein [77] dans le cas du théorème centrale limite par exemple.

Proposition 4.3. *Soient P et $P^* \in \mathcal{W}_p(\mathcal{X})$ et f une application Lipschitz. On a alors, pour $p \in [1, \infty[$,*

$$W_p(P \circ f^{-1}, P^* \circ f^{-1}) \leq \|f\|_{Lip} W_p(P, P^*).$$

La proposition 4.2 suggère, pour établir la convergence en loi d'un estimateur, de coupler l'échantillon d'étude avec un autre échantillon qui sera plus facile à étudier. Ce couplage doit être réalisé de sorte à ce que la distance en Wasserstein entre les lois des estimateurs tende vers 0. Il faut alors être capable de contrôler la distance en Wasserstein entre deux suites de mesures empiriques. Une des contributions de ce manuscrit concerne la distance de Wasserstein entre deux échantillons et l'utilisation de cette méthodologie pour l'étude d'estimateurs.

Considérons deux échantillons i.i.d (X_1, \dots, X_n) et (X_1^*, \dots, X_n^*) générés respectivement à partir des loi P et $P^* \in \mathcal{W}_p(\mathcal{X})$. Nommons Π_n et Π_n^* les mesures empiriques construites sur ces échantillons, à savoir

$$\Pi_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \quad \text{et} \quad \Pi_n^* := \frac{1}{n} \sum_{i=1}^n \delta_{X_i^*}.$$

Ces mesures étant aléatoires, on notera leurs loi P_{Π_n} et $P_{\Pi_n^*}$ qui sont des éléments de $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$.

Remarque 4.1. *Les deux mesures Π_n et Π_n^* sont boréliennes. Cela résulte de la continuité de l'application*

$$\begin{aligned} \varphi : \quad \mathcal{X}^n &\rightarrow \mathcal{W}_p(\mathcal{X}) \\ (x_1, \dots, x_n) &\rightarrow \frac{1}{n} \sum_{i=1}^n \delta_{x_i}. \end{aligned}$$

En effet, pour tous (x_1, \dots, x_n) et $(y_1, \dots, y_n) \in \mathcal{X}$, on peut écrire

$$W_p\left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{n} \sum_{i=1}^n \delta_{y_i}\right) = \min_{\sigma \in \mathfrak{S}(n)} \frac{1}{n} \sum_{i=1}^n d(x_i, y_{\sigma(i)}),$$

avec $\mathfrak{S}(s)$ l'ensemble des permutations de $\{1, \dots, s\}$. Il s'ensuit le fait que φ est Lipschitz.

Théorème 4.5 (Bobbia, Dombry, Varron). *Si (\mathcal{X}, d) est complet et séparable et $p \in [1, \infty[$ alors*

$$W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) = W_p(P, P^*).$$

De nombreux estimateurs peuvent être construits comme image de mesures empiriques. De plus si cette image est une image par une application Lipschitz, la proposition précédente nous incite à considérer les distances de Wasserstein entre mesures empiriques. Ce résultat peut être étendu au cas des mesures empiriques pondérées, mesures qui interviennent dans le cadre du bootstrap par exemple. Soit (ξ_1, \dots, ξ_n) des poids positifs tels que $\sum \xi_i = n$. Soient

$$\Pi_{n,\xi} := \frac{1}{n} \sum_{i=1}^n \xi_i \delta_{X_i} \quad \text{et} \quad \Pi_{n,\xi}^* := \frac{1}{n} \sum_{i=1}^n \xi_i \delta_{X_i^*}$$

les mesures empiriques pondérées.

Théorème 4.6 (Bobbia, Dombry, Varron). *Si (\mathcal{X}, d) est complet et séparable, $p \in [1, \infty[$ et $\sum \xi_i = n$. Si $(\xi_i)_{i=1}^n$ sont indépendants de $(X_i)_{i=1}^n$ et $(X_i^*)_{i=1}^n$ alors*

$$W_p^{(2)} \left(P_{\Pi_{n,\xi}}, P_{\Pi_{n,\xi}^*} \right) = W_p(P, P^*).$$

On a également

$$W_p^{(2)} \left(P_{\Pi_{n,\xi}|(X_1, \dots, X_n)}, P_{\Pi_{n,\xi}^*|(X_1^*, \dots, X_n^*)} \right) = W_p(P, P^*).$$

Dans tous ces cas, la distance entre les mesures empiriques est contrôlée par celle entre les lois qui ont générées les échantillons. Ainsi le contrôle de la distance de Wasserstein entre les lois des estimateurs repose alors entièrement sur le contrôle des lois sous-jacentes.

Deuxième partie

Contributions

Extreme quantile regression in proportional tail framework

Résumé

The model of heteroscedastic extremes initially introduced by Einmahl et al. (JRSSB, 2016) describes the evolution of a non stationary sequence whose extremes evolve over time. We revisit this model and adapt it into a general extreme quantile regression framework. We provide estimates for the extreme value index and the integrated skedasis function and prove their joint asymptotic normality. Our results are quite similar to those developed for heteroscedastic extremes but with a different proof approach emphasizing coupling arguments. We also propose a pointwise estimator of the skedasis function and a Weissman estimator of conditional extreme quantiles and prove the asymptotic normality of both estimators.

5.1 Introduction and main results

5.1.1 Framework

One of the main goals of extreme value theory is to propose estimators of extreme quantiles : given an i.i.d. sample Y_1, \dots, Y_n with distribution F , one wants to estimate the quantile of order $1 - \alpha_n$ defined as $q(\alpha_n) := F^{\leftarrow}(1 - \alpha_n)$, with $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$ and

$$F^{\leftarrow}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}, \quad u \in (0, 1)$$

denotes the quantile function. The extreme regime corresponds to the case when $\alpha_n < 1/n$ in which case extrapolation beyond the sample maximum is needed. Considering an application in hydrology, these mathematical problems correspond to the following situation : given a record over $n = 50$ years of the level of a river, can we estimate the 100-year return level? The answer to this question is provided by the univariate extreme value theory and we refer to the monographs by Coles [17], Beirlant et al. [5] or de Haan and Ferreira [22] for a general background.

In many situations, auxiliary information is available and represented by a covariate X taking values in \mathbb{R}^d and, given $x \in \mathbb{R}^d$, one wants to estimate $q(\alpha_n|x)$, the conditional $(1 - \alpha_n)$ -quantile of Y with respect to some given values of the covariate $X = x$. This is an extreme quantile regression problem. Recent advances in extreme

quantile regression include the works by Chernozhukov [15], El Methni et al. [38] or Daouia et al. [18].

In this paper we develop the proportional tail framework for extreme quantile regression. It is an adaptation of the heteroscedastic extremes developed by Einmahl et al. [35], where the authors propose a model for the extremes of independent but non stationary observations whose distribution evolves over time, a model which can be viewed as a regression framework with time as covariate and deterministic design with uniformly distributed observation times $1/n, 2/n, \dots, 1$. In our setting, the covariate X takes values in \mathbb{R}^d and is random with arbitrary distribution. The main assumption, directly adapted from Einmahl et al. [35], is the so called proportional tail assumption formulated in Equation (7.1) and stating that the conditional tail function of Y given $X = x$ is asymptotically proportional to the unconditional tail. The proportionality factor is given by the so called skedasis function $\sigma(x)$ that accounts for the dependency of the extremes of Y with respect to the covariate X . Furthermore, as it is standard in extreme value theory, the unconditional distribution of Y is assumed to be regularly varying. Together with the proportional tail assumption, this implies that all the conditional distributions are regularly varying with the same extreme value index. Hence the proportional tail framework appears suitable for modeling covariate dependent extremes where the extreme value index is constant but the scale parameter depends on the covariate X in a non parametric way related to the skedasis function $\sigma(x)$. Note that this framework is also considered by Gardes [43] for the purpose of estimation of the extreme value index.

Our main results are presented in the following subsections. Section 5.1.2 considers the estimation of the extreme value index and integrated skedasis function in the proportional tail model and our results of asymptotic normality are similar to those in Einmahl et al. [25] but with a different proof emphasizing coupling arguments. Section 5.1.3 considers pointwise estimation of the skedasis function and conditional extreme quantile estimation with Weissman estimators and state their asymptotic normality. Section 5.2 develops some coupling arguments used in the proofs of the main theorems, proofs gathered in Section 5.3. Finally, an appendix states a technical lemma and its proof.

5.1.2 The proportional tail model

Let (X, Y) be a generic random couple taking values in $\mathbb{R}^d \times \mathbb{R}$. Define the conditional cumulative distribution function of Y given $X = x$ by

$$F_x(y) := \mathbb{P}(Y \leq y | X = x), \quad y \in \mathbb{R}, \quad x \in \mathbb{R}^d.$$

The main assumption of *the proportional tail model* is

$$\lim_{y \rightarrow \infty} \frac{1 - F_x(y)}{1 - F^0(y)} = \sigma(x) \quad \text{uniformly in } x \in \mathbb{R}^d, \quad (5.1)$$

where F^0 is some baseline distribution function and where σ is the so-called skedasis function following the terminology introduced in [35]. By integration, the unconditional distribution F of Y satisfies

$$\lim_{y \rightarrow \infty} \frac{1 - F(y)}{1 - F^0(y)} = \int_{\mathbb{R}^d} \sigma(x) \mathbf{P}_X(dx).$$

We can hence suppose without loss of generality that $F = F^0$ and that $\int \sigma d\mathbf{P}_X = 1$. We also make the assumption that F is of $1/\gamma$ -regular variation

$$1 - F(y) = y^{-1/\gamma} \ell(y), \quad y \in \mathbb{R},$$

with ℓ slowly varying at infinity. Together with the proportional tail condition (7.1) with $F = F^0$, this implies that F_x is also of $1/\gamma$ -regular variation for each $x \in \mathbb{R}^d$, namely This is a strong consequence of the model assumptions. In this model, the extremes are driven by two parameters : the common extreme value index $\gamma > 0$ and the skedasis function $\sigma(\cdot)$. Following [35], we consider the usual ratio estimator (see, e.g., [65, p. 198]) for γ and we propose a non-parametric estimator of the integrated (or cumulative) skedasis function

$$C(x) := \int_{\{u \leq x\}} \sigma(u) \mathbf{P}_X(du), \quad x \in \mathbb{R}^d,$$

where $u \leq x$ stands for the componentwise comparison of vectors. Note that - putting aside the case where X is discrete - the function C is easier to estimate than σ , in the same way that a cumulative distribution function is easier to estimate than a density function. Estimation of C is useful to derive tests while estimation of σ will be considered later on for the purpose of extreme quantile estimation.

Let $(X_i, Y_i)_{1 \leq i \leq n}$ be i.i.d copies of (X, Y) . The estimators are built with observations (X_i, Y_i) for which Y_i exceeds a high threshold \mathbf{y}_n . Note that, in this article, $(\mathbf{y}_n)_{n \in \mathbb{N}}$ can be deterministic or data driven. For the purpose of asymptotics, \mathbf{y}_n depends on the sample size $n \geq 1$ in a way such that

$$\mathbf{y}_n \rightarrow \infty \quad \text{and} \quad N_n \rightarrow \infty \quad \text{in probability,}$$

with $N_n := \sum_{i=1}^n \mathbf{1}_{\{Y_i > \mathbf{y}_n\}}$ the (possibly random) number of exceedances. The extreme value index $\gamma > 0$ is estimated by the ratio estimator

$$\hat{\gamma}_n := \frac{1}{N_n} \sum_{i=1}^n \log \left(\frac{Y_i}{\mathbf{y}_n} \right) \mathbf{1}_{\{Y_i > \mathbf{y}_n\}}.$$

The integrated skedasis function C can be estimated by the following empirical pseudo distribution function

$$\hat{C}_n(x) := \frac{1}{N_n} \sum_{i=1}^n \mathbf{1}_{\{Y_i > \mathbf{y}_n, X_i \leq x\}}, \quad x \in \mathbb{R}^d.$$

When Y is continuous and $\mathbf{y}_n := Y_{n-k_n:n}$ is the $(k_n + 1)$ -th highest order statistic, then $N_n = k$ and $\hat{\gamma}_n$ coincides with the usual Hill estimator.

Our first result addresses the joint asymptotic normality of $\hat{\gamma}_n$ and \hat{C}_n , namely

$$v_n \begin{pmatrix} \hat{C}_n(\cdot) - C(\cdot) \\ \hat{\gamma}_n - \gamma \end{pmatrix} \xrightarrow{\mathcal{L}} \mathbb{W}, \quad (5.2)$$

where \mathbb{W} is a Gaussian Borel probability measure on $L^\infty(\mathbb{R}^d) \times \mathbb{R}$, and $v_n \rightarrow \infty$ is a deterministic rate. To prove the asymptotic normality, the threshold \mathbf{y}_n must scale suitably with respect to the rates of convergence in the proportional tail and

domain of attraction conditions. More precisely, we assume the existence of a positive function A converging to zero and such that, as $y \rightarrow \infty$,

$$\sup_{x \in \mathbb{R}^d} \left| \frac{\bar{F}_x(y)}{\sigma(x)\bar{F}(y)} - 1 \right| = O \left(A \left(\frac{1}{\bar{F}(y)} \right) \right), \text{ and} \quad (5.3)$$

$$\sup_{z > \frac{1}{2}} \left| \frac{\bar{F}(zy)}{z^{-1/\gamma}\bar{F}(y)} - 1 \right| = O \left(A \left(\frac{1}{\bar{F}(y)} \right) \right), \quad (5.4)$$

with $\bar{F}(y) := 1 - F(y)$ and $\bar{F}_x(y) := 1 - F_x(y)$. Our main result can then be stated as follows. At the reading of the present article, the reader shall probably notice that the domain $\{z > 1/2\}$ in (5.4) can be replaced by any domain $\{z > c\}$ for some $c \in]0, 1[$.

Theorem 5.1. *Assume that assumptions (5.3) and (5.4) hold and that $\mathbf{y}_n/y_n \rightarrow 1$ in probability for some deterministic sequence y_n such that $p_n := \bar{F}(y_n)$ satisfies*

$$p_n \rightarrow 0, \quad np_n \rightarrow \infty \quad \text{and} \quad \sqrt{np_n}^{1+\varepsilon} A(1/p_n) \rightarrow 0 \text{ for some } \varepsilon > 0.$$

Then, the asymptotic normality (7.6) holds with

$$v_n := \sqrt{np_n} \quad \text{and} \quad \mathbb{W} \stackrel{\mathcal{L}}{=} \begin{pmatrix} B \\ N \end{pmatrix},$$

with B a C -Brownian bridge on \mathbb{R}^d and N a centered Gaussian random variable with variance γ^2 and independent of B .

By C -Brownian bridge, we here mean a centered Gaussian process on \mathbb{R}^d with covariance function

$$\text{cov}(B(x), B(x')) := \int_{\mathbb{R}^d} \mathbb{1}_{]-\infty, x]} \mathbb{1}_{]-\infty, x']} dC - C(x)C(x').$$

Remark : Theorem 7.1 extends Theorem 2.1 of Einmhal *et al.* [35] in two directions : first, it states that their estimators and theoretical results have natural counterparts in the framework of proportional tails. We also could go past their univariate dependency $i/n \rightarrow \sigma(i/n)$ to a multivariate dependency $x \rightarrow \sigma(x)$, $x \in \mathbb{R}^d$. Second, it shows that general data-driven thresholds can be used. Those extensions come at the price of a slightly more stringent condition upon the bias control. Indeed, their condition $\sqrt{k_n}A(n/k_n) \rightarrow 0$ corresponds to our condition $\sqrt{np_n}^{1+\varepsilon} A(1/p_n) \rightarrow 0$ with $\varepsilon = 0$. We believe that this loss is small in regard to the gain on the practical side : the threshold \mathbf{y}_n in $(\hat{\gamma}_n, \hat{C}_n)$ can be data-driven. Take for example $\mathbf{y}_n := Y_{n-k_n:n}$, which is equivalent in probability to $y_n := F^{\leftarrow}(1 - k_n/n)$ is $k_n \rightarrow \infty$. As a consequence, Theorem 7.1 holds for this choice of \mathbf{y}_n if

$$k_n \rightarrow \infty, \quad \frac{k_n}{n} \rightarrow 0, \quad \text{and} \quad \sqrt{k_n}^{1+\varepsilon} A \left(\frac{n}{k_n} \right) \rightarrow 0.$$

An example where (5.3) and (5.4) hold :

The reader might wonder if a model imposing (5.3) and (5.4) is not too restrictive for modeling. First, note that condition (5.4) has been well studied as the second order condition holding uniformly over intervals (see, e.g., [22, p. 383, Section

B.3],[42],[32]). A generic example of regression model where (5.3) and (5.4) hold is as follows : take a c.d.f H fulfilling the second order heavy tail condition (5.4) on any domain $\{z > c\}$. Then assume that the laws of $Y | X = x$ obey a location scale model in the sense that

$$F_x(y) = H\left(\frac{y - \mu(x)}{\Delta(x)}\right),$$

for some functions $\mu(\cdot)$ and $\Delta(\cdot)$ that are uniformly bounded on \mathbb{R}^d . Then, since $1 - \Delta(x)\mu(x)/y \rightarrow 1$ uniformly in x as $y \rightarrow \infty$, condition (5.4) entails

$$\sup_{x \in \mathbb{R}^d} \left| \frac{\bar{F}_x(y)}{\Delta(x)^{1/\gamma} \bar{H}(y)} - 1 \right| = O(A(1/\bar{H}(y))), \text{ as } y \rightarrow \infty.$$

Integrating in x gives $\bar{H}(y) = \theta \bar{F}(y)$ as $y \rightarrow \infty$ for some $\theta > 0$, which yields (5.3) with the choice of $\sigma(\cdot) := \theta \Delta(\cdot)^{1/\gamma}$.

5.1.3 Extreme quantile regression

In this subsection, we restrict ourselves to the case where \mathbf{y}_n is deterministic i.e. $\mathbf{y}_n = y_n$ according to the notations of Theorem 7.1. We now address the estimation of extreme conditional quantiles in the proportional tail model, namely

$$q(\alpha_n | x) := F_x^{\leftarrow}(1 - \alpha_n),$$

for some $x \in \mathbb{R}^d$ that will be fixed once for all in this section, and for a sequence $\alpha_n = O(1/n)$. To that aim, we shall borrow the heuristics behind the Weissman estimator [86], for which we here write a short reminder. It is known that $F \in D(G_\gamma)$ is equivalent to

$$\lim_{t \rightarrow \infty} \frac{U(tz)}{U(t)} = z^\gamma, \quad \text{for each } z > 0,$$

with $U(t) = F^{\leftarrow}(1 - 1/t)$, $t > 1$. Recall that $p_n = \bar{F}(y_n)$. Since U is of γ -regular variation, the unconditional quantile $q(\alpha_n) := F^{\leftarrow}(1 - \alpha_n)$ is approximated by

$$q(\alpha_n) = U(1/p_n) \frac{U(1/\alpha_n)}{U(1/p_n)} \approx y_n \left(\frac{p_n}{\alpha_n} \right)^\gamma,$$

leading to the Weissman-type quantile estimator

$$\hat{q}(\alpha_n) := y_n \left(\frac{\hat{p}_n}{\alpha_n} \right)^{\hat{\gamma}_n}.$$

Where $\hat{p}_n := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Y_i > y_n\}}$ is the empirical counterpart of p_n .

Now going back to quantile regression in the proportional tail model, it is readily verified that assumption (7.1) implies

$$q(\alpha_n | x) \sim q\left(\frac{\alpha_n}{\sigma(x)}\right) \text{ as } n \rightarrow \infty.$$

This immediately leads to the plug-in estimator

$$\hat{q}(\alpha_n|x) := \hat{q}\left(\frac{\alpha_n}{\hat{\sigma}_n(x)}\right) = y_n \left(\frac{\hat{p}_n \hat{\sigma}_n(x)}{\alpha_n}\right)^{\hat{\gamma}_n}$$

where $\hat{\sigma}_n(x)$ denotes a consistent estimator of $\sigma(x)$.

In the following, we propose a kernel estimator of $\sigma(x)$ and prove its asymptotic normality before deriving the asymptotic normality of the extreme conditional quantile estimator $\hat{q}(\alpha_n|x)$. The proportional tail assumption (7.1) implies

$$\sigma(x) = \lim_{n \rightarrow \infty} \frac{\bar{F}_x(y_n)}{\bar{F}(y_n)}.$$

We propose the simplest kernel estimator with bandwidth $h_n > 0$

$$\frac{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}} \mathbb{1}_{\{Y_i>y_n\}}}{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}}}$$

as an estimator of $\bar{F}_x(y_n)$, while the denominator is estimated by \hat{p}_n . Combining the two estimators yields

$$\hat{\sigma}_n(x) := n \frac{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}} \mathbb{1}_{\{Y_i>y_n\}}}{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}} \sum_{i=1}^n \mathbb{1}_{\{Y_i>y_n\}}}.$$

Our next result states the asymptotic normality of $\hat{\sigma}_n(x)$. The more general case of a random threshold is left for future works.

Theorem 5.2. *Take the notations of Theorem 7.1, and let $h_n \rightarrow 0$ be deterministic and positive. Assume that*

$$np_n h_n^d \rightarrow \infty, \quad \sqrt{np_n h_n^d} A(1/p_n) \rightarrow 0.$$

Assume that the law of X is continuous on a neighborhood of x . Also assume that σ is continuous and positive on a neighborhood of $x \in \mathbb{R}^d$, and that some version f of the density of X also shares those properties. Then, under assumption (5.3), we have

$$\sqrt{np_n h_n^d} \left(\hat{\sigma}_n(x) - \sigma(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma(x)}{f(x)} \right).$$

The asymptotic normality of the extreme quantile estimate $\hat{q}(\alpha_n|x)$ is deduced from the asymptotic normality of $\hat{\gamma}_n$ and $\hat{\sigma}_n(x)$ stated respectively in Theorem 7.1 and 5.2. This is stated in our next theorem, which has to be seen as the counterpart of [22, p.138, Theorem 4.3.8] for conditional extreme quantiles. Also see [65, p. 170, Theorem 9.8] for a similar result when $\log(p_n/\alpha_n) \rightarrow d \in \mathbb{R}$.

Theorem 5.3. *Under assumptions of Theorems 7.1 and 5.2, if $\sqrt{h_n^d} \log(p_n/\alpha_n) \rightarrow \infty$ we have*

$$\frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \log \left(\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \gamma^2 \right).$$

The condition $\sqrt{h_n^d} \log(p_n/\alpha_n)$ requires the bandwidth to be of larger order than $1/\log(p_n/\alpha_n)$ so that the error in the estimation of $\sigma(x)$ is negligible. As a consequence of Theorem 5.3, the consistency

$$\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)} \xrightarrow{\mathbb{P}} 1.$$

That condition seems to state a limit for the extrapolation : α_n cannot be too small or one might lose consistency.

5.2 A coupling approach

We will first prove Theorem 7.1 when \mathbf{y}_n is deterministic (i.e $\mathbf{y}_n \equiv y_n$). In this case N_n is binomial (n, p_n) . Moreover $N_n/np_n \rightarrow 1$ in probability since $np_n \rightarrow \infty$. A simple calculus shows that (7.1) entails, for each A Borel and $t \geq 1$:

$$\mathbb{P}\left(\frac{Y}{\mathbf{y}} \geq t, X \in A \mid Y \geq \mathbf{y}\right) \longrightarrow \int_t^\infty \int_A \mathbf{y}^{-1/\gamma} \sigma(x) d\mathbf{y} \mathbf{P}_X(dx), \text{ as } \mathbf{y} \rightarrow \infty, \quad (5.5)$$

defining a "limit model" for $(X, Y/\mathbf{y})$, the law

$$\mathbf{P}^* := \sigma(x) \mathbf{P}_X \otimes \text{Pareto}(1/\gamma)$$

with independent marginals. Fix $n \geq 1$. Using the heuristic of (5.5), we shall build an explicit coupling between $(X, Y/y_n)$ and the limit model \mathbf{P}^* . Define the conditional tail quantile function as $U_x(t) := F_x^{\leftarrow}(1 - 1/t)$ and recall that the total variation distance between two Borel probability measures on \mathbb{R}^d is defined as

$$\|P_1 - P_2\| := \sup_{B \text{ Borel}} |P_1(B) - P_2(B)|.$$

This distance is closely related with the notion of optimal coupling detailed in [64]. The following fundamental result is due to Dobrushin [26].

Lemme 5.1 (Dobrushin, 1970). *For two probability measures P_1 and P_2 defined on the same measurable space, there exist two random variables (V_1, V_2) on a probability set $(\Omega, \mathcal{A}, \mathbb{P})$ such that*

$$V_1 \sim P_1, \quad V_2 \sim P_2 \quad \text{and} \quad \|P_1 - P_2\| = \mathbb{P}(V_1 \neq V_2).$$

This lemma will be a crucial tool of our coupling construction, which is described as follows.

Coupling construction : Fix $n \geq 1$. Let $(E_{i,n})_{1 \leq i \leq n}$ be i.i.d. Bernoulli random variables with $\mathbb{P}(E_{i,n} = 1) = \bar{F}(y_n)$ and $(Z_i)_{1 \leq i \leq n}$ i.i.d. with distribution Pareto(1) and independent from $(E_{i,n})_{1 \leq i \leq n}$. For each $1 \leq i \leq n$ construct $(\tilde{X}_{i,n}, \tilde{Y}_{i,n}, X_{i,n}^*, Y_{i,n}^*)$ as follows.

- If $E_{i,n} = 1$, then

- ▷ Take $\tilde{X}_{i,n} \sim P_{X|Y>y_n}$, $X_{i,n}^* \sim \sigma(x)\mathbf{P}_X(dx)$ on the same probability space, satisfying $\mathbb{P}(\tilde{X}_{i,n} \neq X_{i,n}^*) = \|\mathbf{P}_{X|Y>y_n} - \sigma(x)\mathbf{P}_X(dx)\|$. Their existence is guaranteed by Lemma 5.1.
- ▷ Set $\tilde{Y}_{i,n} := U_{\tilde{X}_{i,n}}(\frac{Z_i}{F_{\tilde{X}_{i,n}}(y_n)})$, $Y_{i,n}^* := y_n Z_i^\gamma$.

► If $E_{i,n} = 0$, then

- ▷ Randomly generate $(\tilde{X}_{i,n}, \tilde{Y}_{i,n}) \sim \mathbf{P}_{(X,Y)|Y \leq y_n}$.
- ▷ Randomly generate $(X_{i,n}^*, Y_{i,n}^*/y_n) \sim \sigma(x)\mathbf{P}_X(dx) \otimes \text{Pareto}(1/\gamma)$.

The following proposition states the properties of our coupling construction, which will play an essential role in our proof of Theorem 7.1.

Proposition 5.1. *For each $n \geq 1$, the coupling $(\tilde{X}_{i,n}, \tilde{Y}_{i,n}, X_{i,n}^*, Y_{i,n}^*)_{1 \leq i \leq n}$ has the following properties :*

1. $(\tilde{X}_{i,n}, \tilde{Y}_{i,n})_{1 \leq i \leq n}$ has the same law as $(X_i, Y_i)_{1 \leq i \leq n}$.
2. $(X_{i,n}^*, Y_{i,n}^*/y_n) \rightsquigarrow \mathbf{P}^*$.
3. $(X_{i,n}^*, Y_{i,n}^*)_{1 \leq i \leq n}$ and $(E_{i,n})_{1 \leq i \leq n}$ are independent. Moreover, $(Y_{i,n}^*)_{1 \leq i \leq n}$ are i.i.d, and independent from $(\tilde{X}_{i,n}, X_{i,n}^*)$.
4. There exists $M > 0$ such that

$$\max_{\substack{1 \leq i \leq n, \\ E_{i,n}=1}} \left| \frac{Y_{i,n}^*}{\tilde{Y}_{i,n}} - 1 \right| \leq MA(1/p_n) \quad \text{and} \quad (5.6)$$

$$\mathbb{P}(\tilde{X}_{1,n} \neq X_{1,n}^* | E_{1,n} = 1) \leq MA(1/p_n), \quad (5.7)$$

where A is given by assumptions (5.3) and (5.4).

Démonstration. To prove Point 1, it is sufficient to see that

$$\mathcal{L}((\tilde{X}_{1,n}, \tilde{Y}_{1,n}) | E_{1,n} = 1) = \mathcal{L}((X, Y) | Y > y_n).$$

Since $U_x(z/(1 - F_x(y_n))) \leq y$ if and only if $1 - (1 - F_x(y_n))/z \leq F_x(y)$ we have, for $y \geq y_n$:

$$\begin{aligned} & \int_1^\infty \mathbf{1}_{\{U_x(z/(1-F_x(y_n))) \leq y\}} \frac{dz}{z^2} \\ &= \int_1^\infty \mathbf{1}_{\{1-(1-F_x(y_n))/z \leq F_x(y)\}} \frac{dz}{z^2} \\ &= \int_{F_x(y_n)}^1 \mathbf{1}_{\{t \leq F_x(y)\}} \frac{dt}{1 - F_x(y_n)} \\ &= \int_{F_x(y_n)}^{F_x(y)} \frac{dt}{1 - F_x(y_n)} = \frac{F_x(y) - F_x(y_n)}{1 - F_x(y_n)}, \end{aligned}$$

with the second equality given by the change of variable $t = 1 - (1 - F_x(y_n))/z$. We can deduce from this computation that, for a Borel set B and $y \geq y_n$,

$$\begin{aligned}
& \mathbb{P} \left(\tilde{X}_{1,n} \in B, U_{\tilde{X}_{1,n}} \left(\frac{Z}{1 - F_{\tilde{X}_{1,n}}(y_n)} \right) \leq y \middle| E_{1,n} = 1 \right) \\
&= \int_{x \in B} \int_1^\infty \mathbf{1}_{\{U_x(z/(1-F_x(y_n))) \leq y\}} \frac{dz}{z^2} dP_{X|Y>y_n}(x) \\
&= \int_{x \in B} \frac{F_x(y) - F_x(y_n)}{1 - F_x(y_n)} dP_{X|Y>y_n}(x) \\
&= \int_{x \in B} \mathbb{P}(Y \leq y | Y > y_n, X = x) dP_{X|Y>y_n}(x) \\
&= \mathbb{P}(X \in B, Y \leq y | Y > y_n).
\end{aligned}$$

This proves Point 1. Points 2 and 3 are immediate.

Point 4 will be proved with the two following lemmas.

Lemma 5.2. *Under conditions (5.3) and (5.4), we have*

$$\sup_{z \geq 1/2} \sup_{x \in \mathbb{R}^d} \left| \frac{1}{z^\gamma y} U_x \left(\frac{z}{\bar{F}_x(y)} \right) - 1 \right| = O \left(A \left(\frac{1}{\bar{F}(y)} \right) \right), \text{ as } y \rightarrow \infty.$$

Démonstration. According to assumptions (5.3) and (5.4), there exists a constant M such that

$$\begin{aligned}
\left| \frac{\bar{F}_x(y)}{\sigma(x)\bar{F}(y)} - 1 \right| &\leq MA \left(\frac{1}{\bar{F}(y)} \right), \text{ uniformly in } x \in \mathbb{R}^d, \text{ and} \\
\left| \frac{\bar{F}(zy)}{z^{-1/\gamma}\bar{F}(y)} - 1 \right| &\leq MA \left(\frac{1}{\bar{F}(y)} \right), \text{ uniformly in } z \geq 1/2. \tag{5.8}
\end{aligned}$$

From the definition of U_x we have

$$\begin{aligned}
U_x \left(\frac{Z}{\bar{F}_x(y)} \right) &= F_x^{\leftarrow} \left(1 - \frac{\bar{F}_x(y)}{z} \right) \\
&= \inf \left\{ w \in \mathbb{R} : F_x(w) \geq 1 - \frac{\bar{F}_x(y)}{z} \right\} \\
&= \inf \left\{ w \in \mathbb{R} : z \frac{\bar{F}_x(w)}{\bar{F}_x(y)} \leq 1 \right\}.
\end{aligned}$$

Hence for any $w^- < w^+$ one has :

$$z \frac{\bar{F}_x(w^+)}{\bar{F}_x(y)} < 1 < z \frac{\bar{F}_x(w^-)}{\bar{F}_x(y)} \Rightarrow U_x \left(\frac{z}{\bar{F}_x(y)} \right) \in [w^-, w^+]. \tag{5.9}$$

Now write $\epsilon(y) := MA(1/\bar{F}(y))$ and choose $w^\pm := z^\gamma y(1 \pm 4\gamma\epsilon(y))$, so that one can write

$$\begin{aligned}
z \frac{\bar{F}_x(w^-)}{\bar{F}_x(y)} &= z \frac{\sigma(x)\bar{F}(w^-)(1-\epsilon(y))}{\sigma(x)\bar{F}(y)(1+\epsilon(y))} \\
&\geq z \frac{1-\epsilon(y)}{1+\epsilon(y)} \frac{1}{\bar{F}(y)} \bar{F}(z^\gamma y(1-4\gamma\epsilon(y))) \\
&\geq z \frac{1-\epsilon(y)}{1+\epsilon(y)} \frac{1}{\bar{F}(y)} \bar{F}(y)(1-\epsilon(y)) (z^\gamma(1-4\gamma\epsilon(y)))^{-1/\gamma}, \text{ by (5.8)} \\
&\geq \frac{(1-\epsilon(y))^2}{1+\epsilon(y)} (1-4\gamma\epsilon(y))^{-1/\gamma}.
\end{aligned}$$

A similar computation gives

$$z \frac{\bar{F}_x(w^+)}{\bar{F}_x(y)} \leq \frac{(1 + \epsilon(y))^2}{1 - \epsilon(y)} (1 + 4\gamma\epsilon(y))^{-1/\gamma}.$$

As a consequence the condition before " \Rightarrow " in (5.9) holds if

$$4\gamma \geq \frac{1}{\epsilon(y)} \max \left\{ 1 - \left(\frac{(1 - \epsilon(y))^2}{1 + \epsilon(y)} \right)^\gamma ; \left(\frac{(1 + \epsilon(y))^2}{1 - \epsilon(y)} \right)^\gamma - 1 \right\}.$$

But a Taylor expansion of the right hand side shows that it is $3\gamma + o(1)$ as $y \rightarrow \infty$. This concludes the proof of Lemma 5.2. \square

Applying Lemma 5.2 with $z := Z_i$ and $y := y_n$ gives

$$\max_{i: E_{i,n}=1} \left| \frac{Y_{i,n}^*}{\tilde{Y}_{i,n}} - 1 \right| = O(A(1/p_n)).$$

Now by construction of $(\tilde{X}_{1,n}, X_{1,n}^*)$ when $E_{1,n} = 1$, we see that (5.7) is a consequence of the following lemma.

Lemma 5.3. *Under conditions (5.3) and (5.4), we have*

$$\|P_{X|Y>y} - \sigma(x)\mathbf{P}_X(dx)\| = O\left(A\left(\frac{1}{\bar{F}(y)}\right)\right), \text{ as } y \rightarrow \infty.$$

Démonstration. For $B \in \mathbb{R}^d$, we have

$$\begin{aligned} & |P(X \in B|Y > y) - \int_B \sigma(x)\mathbf{P}_X(dx)| \\ &= \left| \frac{\int_B \bar{F}_x(y)P_X(dx)}{\bar{F}(y)} - \int_B \sigma(x)\mathbf{P}_X(dx) \right| \\ &\leq \int_B \left| \frac{\bar{F}_x(y)}{\bar{F}(y)} - \sigma(x) \right| \mathbf{P}_X(dx) \\ &= O\left(A\left(\frac{1}{\bar{F}(y)}\right)\right), \text{ by (5.3).} \end{aligned}$$

\square

This proves (5.7) and hence concludes the proof of Proposition 5.1. \square

5.3 Proofs

5.3.1 Proof of Theorem 7.1

Change of notation : Since, for each n the law of $(\tilde{X}_{i,n}, \tilde{Y}_{i,n})_{i=1,\dots,n}$ is $\mathbf{P}_{X,Y}^{\otimes n}$, we shall confound them with $(X_i, Y_i)_{i=1,\dots,n}$ to unburden notations.

Proof when $y_n = y_n$ is deterministic

Fix $0 < \varepsilon < \frac{1}{2}$ and $0 < \beta < \varepsilon/(2\gamma)$. We consider the empirical process defined for every $x \in \mathbb{R}^d$ and $y \geq 1/2$ as

$$\mathbb{G}_n(x, y) := \sqrt{np_n}(\mathbb{F}_n(x, y) - \mathbb{F}(x, y)), \text{ with}$$

$$\mathbb{F}_n(x, y) := \frac{1}{N_n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{Y_i/y_n > y\}} E_{i,n}, \text{ and}$$

$$\mathbb{F}(x, y) := C(x)V_\gamma(y) = Q([\!-\infty, x] \times]y, +\infty[),$$

where $V_\gamma(y) := y^{-1/\gamma}$ for $y \geq 1$ and $V_\gamma(y) := 1$ otherwise.

Note that neither \mathbb{F} nor any realisation of \mathbb{F}_n is a cumulative distribution function in the strict sense, since they are decreasing in y . Their roles should however be seen as the same as for c.d.f. Now denote by $L^{\infty, \beta}(\mathbb{R}^d \times [1/2, \infty[)$ the (closed) subspace of $L^\infty(\mathbb{R}^d \times [1/2, \infty[)$ of all f satisfying

$$\|f\|_{\infty, \beta} := \sup_{x \in \mathbb{R}^d, y \geq 1/2} |y^\beta f(x, y)| < \infty,$$

$$f(\infty, y) := \lim_{\min\{x_1, \dots, x_d\} \rightarrow \infty} f(x, y) \text{ exists for each } y \geq 1,$$

$$\{y \mapsto f(\infty, y)\} \text{ is Càdlàg (see e.g. [9] p. 121).}$$

Simple arguments show that \mathbb{G}_n takes values in $L^{\infty, \beta}(\mathbb{R}^d \times [1/2, \infty[)$.

First note that $\widehat{C}_n - C$ and $\widehat{\gamma}_n - \gamma$ are images of \mathbb{G}_n by the following map φ .

$$\begin{aligned} \varphi : L^{\infty, \beta}(\mathbb{R}^d \times [1/2, \infty[) &\rightarrow L^\infty(\mathbb{R}^d) \times \mathbb{R} \\ f &\mapsto (\{x \mapsto f(x, 1)\}, \int_1^\infty y^{-1} f(\infty, y) dy). \end{aligned}$$

and remark that φ is continuous since $\beta > 0$. By the continuous mapping theorem, we hence see that Theorem 7.1 will be a consequence of

$$\mathbb{G}_n \xrightarrow{\mathcal{L}} \mathbf{W} \text{ in } L^{\infty, \beta}(\mathbb{R}^d \times [1/2, \infty[), \quad (5.10)$$

where \mathbf{W} is the centered Gaussian process with covariance function

$$\text{cov}(\mathbf{W}(x_1, y_1), \mathbf{W}(x_2, y_2)) = C(x_1 \wedge x_2)V_\gamma(y_1) \wedge V_\gamma(y_2) - C(x_1)C(x_2)V_\gamma(y_1)V_\gamma(y_2),$$

and where $x_1 \wedge x_2$ is understood componentwise.

The proof is divided into two steps. In *step 1* we prove (5.10) for the counterpart of \mathbb{G}_n that is built on the Q sample $(X_{i,n}^*, Y_{i,n}^*)_{1 \leq i \leq n}$. Our proof relies on standard argument from empirical processes. In *step 2* we use the coupling properties of Proposition (5.1) to deduce (5.10) for the original sample $(X_i, Y_i)_{1 \leq i \leq n}$.

Step 1 : Define

$$\mathbb{F}_n^*(x, y) := \frac{1}{N_n} \sum_{i=1}^n \mathbb{1}_{\{X_i^* \leq x\}} \mathbb{1}_{\{Y_{i,n}^*/y_n > y\}} E_{i,n} \quad x \in \mathbb{R}^d, \quad y \geq 1/2.$$

The following proposition is a Donsker theorem in weighted topology for $\mathbb{G}_n^* := \sqrt{np_n}(\mathbb{F}_n^* - \mathbb{F})$.

Proposition 5.2. *If (5.3) and (5.4) hold, then*

$$\mathbb{G}_n^* \xrightarrow{\mathcal{L}} \mathbf{W}, \text{ in } L^{\infty, \beta}(\mathbb{R}^d \times [1/4, \infty[).$$

Démonstration. Write $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise.

Since $(X_{i,n}^*, Y_{i,n}^*)_{1 \leq i \leq n}$ is independent of $(E_{i,n})_{1 \leq i \leq n}$, Lemma 5.6 entails the following equality in laws

$$\sum_{i=1}^n \delta \left(X_{i,n}^*, \frac{Y_{i,n}^*}{y_n} \right) E_{i,n} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{\nu(n)} \delta \left(X_{i,n}^*, \frac{Y_{i,n}^*}{y_n} \right),$$

where $\nu(n) \sim \mathcal{B}(n, p_n)$ is independent of $(X_{i,n}^*, Y_{i,n}^*)_{1 \leq i \leq n}$.

Since $(X_{i,n}^*, Y_{i,n}^*/y_n) \rightsquigarrow Q$ and since $\nu(n) \xrightarrow{\mathbb{P}} \infty$, $\nu(n)/np_n \xrightarrow{\mathbb{P}} 1$ and $\nu(n)$ independent of $(X_{i,n}^*, Y_{i,n}^*)_{1 \leq i \leq n}$, we see that $\mathbb{G}_n \xrightarrow{\mathcal{L}} \mathbf{W}$ will be a consequence of

$$\sqrt{k} \left(\frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{U_i \leq \cdot, V_i > \cdot\}} - \mathbb{F}(\cdot, \cdot) \right) \xrightarrow{\mathcal{L}}_{k \rightarrow \infty} \mathbf{W} \text{ in } L^{\infty, \beta}(\mathbb{R}^d \times [1/4, \infty[),$$

where the (U_i, V_i) are i.i.d. with distribution Q . Now consider the following class of functions on $\mathbb{R}^d \times [1/4, \infty[$

$$\mathcal{F}_\beta := \left\{ f_{x,y} : (u, v) \mapsto y^\beta \mathbb{1}_{(-\infty, x]}(u) \mathbb{1}_{]y, \infty[}(v), x \in \mathbb{R}^d, y \geq 1/4 \right\}.$$

Using the isometry :

$$\begin{aligned} L^{\infty, \beta}(\mathbb{R}^d \times [1/4, \infty[) &\rightarrow L^\infty(\mathcal{F}_\beta) \\ g &\mapsto \{\Psi : f_{x,y} \mapsto g(x, y)\}, \end{aligned}$$

it is enough to prove that the abstract empirical process indexed by \mathcal{F}_β converges weakly to the Q -Brownian bridge indexed by \mathcal{F}_β . In other words, we need to verify that \mathcal{F}_β is Q -Donsker. This property can be deduced from two remarks :

1. \mathcal{F}_β is a VC-subgraph class of function (see, e.g, Van der Vaart and Wellner [82], p.141). To see this, note that

$$\mathcal{F}_\beta \subset \left\{ f_{x,s,z} : (u, v) \mapsto z \mathbb{1}_{(-\infty, x]}(u) \mathbb{1}_{]s, \infty[}(v), x \in \mathbb{R}^d, s \in [1/4, \infty[, z \in \mathbb{R} \right\}$$

which is a VC-subgraph class : the subgraph of each of its members is an hypercube of \mathbb{R}^{d+2} .

2. \mathcal{F}_β has a square integrable envelope F . This is proved by noting that for fixed $(u, v) \in \mathbb{R}^d \times [1/4, \infty[$.

$$F^2(u, v) = \sup_{x \in \mathbb{R}^d, y \geq 1/4} y^{2\beta} \mathbb{1}_{[0, x]}(u) \mathbb{1}_{]y, \infty[}(v) = v^{2\beta}$$

as a consequence F^2 is Q -integrable since $\beta < (2\gamma)^{-1}$.

This concludes the proof of Proposition 5.2. □

Step 2 : We show here that the two empirical processes \mathbb{G}_n and \mathbb{G}_n^* must have the same weak limit, by proving the next proposition.

Proposition 5.3. *Under Assumptions (5.3) and (5.4), we have*

$$\sup_{x \in \mathbb{R}^d, y \geq 1/2} y^\beta \sqrt{np_n} |\mathbb{F}_n^*(x, y) - \mathbb{F}_n(x, y)| = o_{\mathbb{P}}(1).$$

Démonstration. Adding and subtracting

$$\mathbb{F}_n^\sharp(x, y) := \frac{1}{N_n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{Y_{i,n}^*/y_n > y\}} E_{i,n}$$

in $|\mathbb{F}_n(x, y) - \mathbb{F}_n^*(x, y)|$, the triangle inequality entails, almost surely

$$\begin{aligned} & |\mathbb{F}_n(x, y) - \mathbb{F}_n^*(x, y)| \\ &= |\mathbb{F}_n(x, y) - \mathbb{F}_n^\sharp(x, y) + \mathbb{F}_n^\sharp(x, y) - \mathbb{F}_n^*(x, y)| \\ &\leq \frac{1}{N_n} \sum_{i=1}^n |\mathbb{1}_{\{X_i \leq x\}} - \mathbb{1}_{\{X_{i,n}^* \leq x\}}| \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}} E_{i,n} \\ &\quad + \frac{1}{N_n} \sum_{i=1}^n |\mathbb{1}_{\left\{\frac{Y_i}{y_n} > y\right\}} - \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}}| \mathbb{1}_{\{X_i \leq x\}} E_{i,n} \\ &\leq \frac{1}{N_n} \sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}} E_{i,n} + \frac{1}{N_n} \sum_{i=1}^n |\mathbb{1}_{\left\{\frac{Y_i}{y_n} > y\right\}} - \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}}| E_{i,n}. \end{aligned}$$

Let us first focus on the first term. Notice that

$$\begin{aligned} & \sup_{x \in \mathbb{R}^d, y \geq 1/2} \frac{y^\beta \sqrt{np_n}}{N_n} \sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}} E_{i,n} \\ &= \sup_{y \geq 1/2} \frac{y^\beta \sqrt{np_n}}{N_n} \sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}} E_{i,n} \\ &\leq \sup_{y \geq 1/2} \frac{y^\beta \sqrt{np_n}}{N_n} \left(\max_{i=1, \dots, n} \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}} E_{i,n} \right) \sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} E_{i,n}. \end{aligned}$$

Now notice that

$$\begin{aligned} \sup_{y \geq 1/2} \max_{i=1, \dots, n} y^\beta \mathbb{1}_{\left\{\frac{Y_{i,n}^*}{y_n} > y\right\}} E_{i,n} &= \max_{i=1, \dots, n} \sup_{y \geq 1/2} y^\beta \mathbb{1}_{[1, Y_{i,n}^*/y_n]}(y) E_{i,n} \\ &= \max_{i=1, \dots, n} \left(\frac{Y_{i,n}^*}{y_n} \right)^\beta E_{i,n}. \end{aligned}$$

By independence between $E_{i,n}$ and $Y_{i,n}^*/y_n$, Lemma 5.6 in the Appendix gives

$$\max_{i=1, \dots, n} \left(\frac{Y_{i,n}^*}{y_n} \right)^\beta E_{i,n} \stackrel{\mathcal{L}}{=} \max_{i=1, \dots, \nu(n)} \left(\frac{Y_{i,n}^*}{y_n} \right)^\beta$$

where $Y_{i,n}^*/y_n$ in the right hand side have a Pareto($1/\gamma$) distribution, whence

$$\max_{i=1,\dots,n} \left(\frac{Y_{i,n}^*}{y_n} \right)^\beta E_{i,n} = O_{\mathbb{P}}(\nu(n)^{\beta\gamma}) = O_{\mathbb{P}}((np_n)^{\beta\gamma}). \quad (5.11)$$

Moreover, writing $A_n := A(1/p_n)$ one has

$$\mathbb{E} \left(\sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} E_{i,n} \right) = np_n A_n,$$

which entails

$$\frac{1}{np_n A_n} \sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} E_{i,n} = O_{\mathbb{P}}(1). \quad (5.12)$$

As a consequence

$$\begin{aligned} & \frac{\sqrt{np_n}}{N_n} \max_{i=1,\dots,n} \left(\frac{Y_{i,n}^*}{y_n} \right)^\beta E_{i,n} \left(\sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} E_{i,n} \right) \\ &= \frac{np_n}{N_n} \max_{i=1,\dots,n} \left(\frac{Y_{i,n}^*}{y_n} \right)^\beta E_{i,n} \left(\frac{1}{np_n A_n} \sum_{i=1}^n \mathbb{1}_{\{X_i \neq X_{i,n}^*\}} E_{i,n} \right) \sqrt{np_n} A_n \\ &= O_{\mathbb{P}}(1) O_{\mathbb{P}}((np_n)^{\beta\gamma}) O_{\mathbb{P}}(1) \sqrt{np_n} A_n, \text{ by (5.11) and (5.12)} \\ &= o_{\mathbb{P}}(1), \text{ by assumption of Theorem 7.1, and since } \beta\gamma < \frac{\varepsilon}{2}. \end{aligned}$$

Let us now focus on the convergence

$$\sup_{x \in \mathbb{R}^d, y \geq 1/2} y^\beta \sqrt{np_n} \frac{1}{N_n} \sum_{i=1}^n \left| \mathbb{1}_{\left\{ \frac{Y_i}{y_n} > y \right\}} - \mathbb{1}_{\left\{ \frac{Y_{i,n}^*}{y_n} > y \right\}} \right| E_{i,n} \xrightarrow{\mathbb{P}} 0.$$

We deduce from Proposition 5.1 that, almost surely, writing $\epsilon_n := MA_n$:

$$(1 - \epsilon_n) \frac{Y_i}{y_n} E_{i,n} \leq \frac{Y_{i,n}^*}{y_n} E_{i,n} \leq (1 + \epsilon_n) \frac{Y_i}{y_n} E_{i,n}.$$

Which entails, almost surely, for all $y \geq 1$:

$$E_{i,n} \mathbb{1}_{\left\{ \frac{Y_{i,n}^*}{y_n} \geq (1 + \epsilon_n)y \right\}} \leq E_{i,n} \mathbb{1}_{\left\{ \frac{Y_i}{y_n} \geq y \right\}} \leq E_{i,n} \mathbb{1}_{\left\{ \frac{Y_i}{y_n} \geq (1 - \epsilon_n)y \right\}},$$

implying

$$\left| \mathbb{1}_{\left\{ \frac{Y_i}{y_n} > y \right\}} - \mathbb{1}_{\left\{ \frac{Y_{i,n}^*}{y_n} > y \right\}} \right| E_{i,n} \leq \left| \mathbb{1}_{\left\{ \frac{Y_{i,n}^*}{y_n} > (1 - \epsilon_n)y \right\}} - \mathbb{1}_{\left\{ \frac{Y_{i,n}^*}{y_n} > (1 + \epsilon_n)y \right\}} \right| E_{i,n}.$$

This entails

$$\begin{aligned} & \sup_{x \in \mathbb{R}^d, y \geq 1/2} y^\beta \sqrt{np_n} \frac{1}{N_n} \sum_{i=1}^n \left| \mathbb{1}_{\left\{ \frac{Y_i}{y_n} > y \right\}} - \mathbb{1}_{\left\{ \frac{Y_{i,n}^*}{y_n} > y \right\}} \right| E_{i,n} \\ & \leq \sup_{x \in \mathbb{R}^d, y \geq 1/2} y^\beta \sqrt{np_n} \left| \mathbb{F}_n^*(\infty, (1 - \epsilon_n)y) - \mathbb{F}_n^*(\infty, (1 + \epsilon_n)y) \right|. \end{aligned}$$

Consequently we have, adding and subtracting expectations :

$$\begin{aligned} & \sup_{x \in \mathbb{R}^d, y \geq 1/2} y^\beta \sqrt{np_n} \frac{1}{N_n} \sum_{i=1}^n \left| \mathbf{1}_{\{\frac{Y_i}{y_n} > y\}} - \mathbf{1}_{\{\frac{Y_{i,n}^*}{y_n} > y\}} \right| E_{i,n} \\ & \leq \sup_{x \in \mathbb{R}^d, y \geq 1/2} y^\beta \left| \tilde{\mathbb{G}}_n^*((1 - \epsilon_n)y) - \tilde{\mathbb{G}}_n^*((1 + \epsilon_n)y) \right| \end{aligned} \quad (5.13)$$

$$+ \sqrt{np_n} \sup_{y \geq 1/2} y^\beta (V_\gamma((1 - \epsilon_n)y) - V_\gamma((1 + \epsilon_n)y)), \quad (5.14)$$

where we wrote $\tilde{\mathbb{G}}_n^*(y) := \mathbb{G}_n^*(\infty, y)$.

We will first prove that (5.14) converges to 0. For, $y \geq 1$ we can bound

$$\begin{aligned} & y^\beta (V_\gamma((1 - \epsilon_n)y) - V_\gamma((1 + \epsilon_n)y)) \\ & \leq y^\beta |1 - ((1 + \epsilon_n)y)^{-1/\gamma}| \mathbf{1}_{\{(1 - \epsilon_n)y < 1\}} \\ & \quad + y^\beta |((1 - \epsilon_n)y)^{-1/\gamma} - ((1 + \epsilon_n)y)^{-1/\gamma}| \mathbf{1}_{\{(1 - \epsilon_n)y \geq 1\}}. \end{aligned} \quad (5.15)$$

In the first term of the right hand side we can write, since $(1 - \epsilon_n)y < 1$,

$$\begin{aligned} & y^\beta |1 - ((1 + A_n)y)^{-1/\gamma}| \mathbf{1}_{\{(1 - \epsilon_n)y < 1\}} \\ & \leq y^{\beta-1/\gamma} |y^{1/\gamma} - (1 + A_n)^{-1/\gamma}| \mathbf{1}_{\{(1 - \epsilon_n)y < 1\}} \\ & \leq y^{\beta-1/\gamma} |(1 - \epsilon_n)^{-1/\gamma} - (1 + A_n)^{-1/\gamma}| \mathbf{1}_{\{(1 - \epsilon_n)y < 1\}} \\ & \leq 4\gamma^{-1}\epsilon_n, \text{ since } \beta - 1/\gamma < 0. \end{aligned}$$

The second term of (5.15) is bounded by similar arguments, from where :

$$\begin{aligned} & \sqrt{np_n} \sup_{x \in \mathbb{R}^d, y \geq 1/2} y^\beta |V_\gamma((1 - \epsilon_n)y) - V_\gamma((1 + \epsilon_n)y)| \\ & \leq 8\gamma^{-1}M\sqrt{np_n}A_n, \end{aligned}$$

which converges in probability to 0 by assumptions of Theorem 7.1.

We will now prove that (5.13) converges to zero in probability. By Proposition 5.2, the continuous mapping theorem together with the Portmanteau theorem entail :

$$\begin{aligned} & \forall \varepsilon > 0, \forall \rho > 0, \overline{\lim} \mathbb{P} \left(\sup_{y \geq 1/2, \delta < \rho} y^\beta |\tilde{\mathbb{G}}_n^*((1 - \delta)y) - \tilde{\mathbb{G}}_n^*((1 + \delta)y)| \geq \varepsilon \right) \\ & \leq \mathbb{P} \left(\sup_{y \geq 1/2, \delta < \rho} y^\beta |\tilde{\mathbf{W}}((1 - \delta)y) - \tilde{\mathbf{W}}((1 + \delta)y)| \geq \varepsilon \right), \end{aligned}$$

where $\tilde{\mathbf{W}}(y) := \mathbf{W}(\infty, y)$ is the centered Gaussian process with covariance function

$$\text{cov}(\tilde{\mathbf{W}}(y_1), \tilde{\mathbf{W}}(y_2)) := V_\gamma(y_1) \wedge V_\gamma(y_2) - V_\gamma(y_1)V_\gamma(y_2), \quad (y_1, y_2) \in [1/4, \infty]^2.$$

With Proposition 5.2 together with the continuous mapping theorem, we see that the proof of Proposition 5.3 will be concluded if we establish the following lemma.

Lemme 5.4. *We have*

$$\sup_{y \geq 1/2, \delta < \rho} y^\beta |\tilde{\mathbf{W}}((1 - \delta)y) - \tilde{\mathbf{W}}((1 + \delta)y)| \xrightarrow[\rho \rightarrow 0]{\mathbb{P}} 0.$$

Démonstration. Let \mathbb{B}_0 be the standard Brownian bridge with \mathbb{B}_0 identically zero on $[1, \infty[$. $\tilde{\mathbf{W}}$ has the same law as $\{y \mapsto \mathbb{B}_0(y^{-1/\gamma})\}$ (see [78], p. 99), from where

$$\begin{aligned} & \sup_{y \geq 1/2, \delta < \rho} y^\beta |\tilde{\mathbf{W}}((1-\delta)y) - \tilde{\mathbf{W}}((1+\delta)y)| \\ & \stackrel{\mathcal{L}}{=} \sup_{y \geq 1/2, \delta < \rho} y^\beta \left| \mathbb{B}_0(((1-\delta)y)^{-1/\gamma}) - \mathbb{B}_0(((1+\delta)y)^{-1/\gamma}) \right| \\ & \leq \sup_{0 \leq y \leq 2, \delta < \rho} y^{-\beta\gamma} \left| \mathbb{B}_0((1-\delta)^{-1/\gamma}y) - \mathbb{B}_0((1+\delta)^{-1/\gamma}y) \right|, \text{ almost surely.} \end{aligned}$$

Since $\beta\gamma < 1/2$ the process \mathbb{B}_0 is a.s.- $\beta\gamma$ -Hölder continuous on $[0, +\infty[$. Consequently, for an a.s finite random variable H one has with probability one :

$$\begin{aligned} & \sup_{0 \leq y \leq 2, \delta < \rho} y^{-\beta\gamma} |\mathbb{B}_0((1-\delta)^{-1/\gamma}y) - \mathbb{B}_0((1+\delta)^{-1/\gamma}y)| \\ & \leq \sup_{0 \leq y \leq 2} y^{-\beta\gamma} |(1-\rho)^{-1/\gamma} - (1+\rho)^{-1/\gamma}|^{\beta\gamma} y^{\beta\gamma} H \\ & = |2(1-\rho)^{-1/\gamma} - 2(1+\rho)^{-1/\gamma}|^{\beta\gamma} H \\ & = (4 \frac{\rho}{\gamma})^{\beta\gamma} H. \end{aligned}$$

□

The preceding lemma concludes the proof of Proposition 5.3, which, combined with Proposition (5.2), proves (5.10). This concludes the proof of Theorem 7.1 when $\mathbf{y}_n \equiv y_n$. □

Proof of Theorem 7.1 in the general case.

We now drop the assumption $\mathbf{y}_n \equiv y_n$ and we relax it to $\frac{\mathbf{y}_n}{y_n} \xrightarrow{\mathbb{P}} 1$ to achieve the proof of Theorem 7.1 in its full generality. We shall use the results of §5.3.1. Define

$$\check{\mathbb{F}}_n(x, y) := \frac{1}{\sum_{i=1}^n \mathbb{1}_{\{Y_i > y_n\}}} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \mathbb{1}_{\{Y_i/y_n > y\}} \text{ and}$$

$$\check{\mathbb{G}}_n(x, y) := \sqrt{np_n} \left(\check{\mathbb{F}}_n(x, y) - \mathbb{F}(x, y) \right).$$

Now write $u_n := \frac{\mathbf{y}_n}{y_n}$. From §5.3.1, we know that

$$\left(\check{\mathbb{G}}_n, u_n \right) \xrightarrow{\mathcal{L}} (\mathbf{W}, 1) \text{ in } \mathbf{D} \times]0, +\infty[, \text{ where } \mathbf{D} := L^{\infty, \beta}(\mathbb{R}^d \times [1/2, \infty[).$$

Moreover, as pointed out in Lemma 5.4, \mathbf{W} almost surely belongs to

$$\mathbf{D}_0 = \left\{ \varphi \in L^{\infty, \beta}(\mathbb{R}^d \times [1/2, \infty[), \sup_{x \in \mathbb{R}^d, y, y' > 1/2} \frac{|\varphi(x, y) - \varphi(x, y')|}{|y - y'|^{\beta\gamma}} < \infty \right\}.$$

Consider the followings maps $(g_n)_{n \in \mathbb{N}}$ and g from \mathbf{D} to $L^{\infty, \beta}(\mathbb{R}^d \times [1, \infty[)$

$$g_n : (\varphi, u) \mapsto \sqrt{np_n} \left(\frac{\mathbb{F}(\cdot, u.) + \frac{1}{\sqrt{np_n}} \varphi(\cdot, u.)}{\mathbb{F}(\infty, u) + \frac{1}{\sqrt{np_n}} \varphi(\infty, u)} - \mathbb{F}(\cdot, \cdot) \right), \text{ and}$$

$$g : (\varphi, u) \mapsto u^{1/\gamma} \left(\varphi(\cdot, u) - \varphi(\infty, u) \mathbb{F}(\cdot, \cdot) \right).$$

Notice that $\mathbb{G}_n = g_n(\mathbb{G}_n^\vee, u_n)$ and $g(\mathbf{W}, 1) = \mathbf{W}$. The achievement of the proof of Theorem 7.1 hence boils down to making use of the extended continuous mapping theorem (see, e.g Theorem 1.11.1 p. 67 in [82]) which is applicable to the sequence (g_n, \mathbb{G}_n) provided that we establish the following lemma

Lemme 5.5. *For any sequence φ_n of elements of \mathbf{D} that converges to some $\varphi \in \mathbf{D}_0$, and for any sequence $u_n \rightarrow 1$ one has $g_n(\varphi_n, u_n) \rightarrow g(\varphi, 1)$ in $L^{\infty, \beta}(\mathbb{R}^d \times [1, \infty[)$. Here, convergence in $L^{\infty, \beta}(\mathbb{R}^d \times [1, \infty[)$ is understood as with respect to $\|\cdot\|$, the restriction of $\|\cdot\|_{\infty, \beta}$ to $\mathbb{R}^d \times [1, \infty[$.*

Démonstration. For fixed $(x, y) \in \mathbb{R}^d \times [1/2, \infty[$ and $n \geq 1$, we have, writing $t_n := (np_n)^{-1/2}$:

$$\begin{aligned} & |g_n(\varphi_n, u_n)(x, y) - g(\varphi, 1)(x, y)| \\ &= \left| \frac{1}{t_n} \left(\frac{\mathbb{F}(x, u_n y) + t_n \varphi_n(x, u_n y)}{\mathbb{F}(\infty, u_n) + t_n \varphi_n(\infty, u_n)} - \mathbb{F}(x, y) \right) - \left(\varphi(x, y) - \varphi(\infty, 1) \mathbb{F}(x, y) \right) \right|. \end{aligned}$$

Now elementary algebra using $\mathbb{F}(x, y u_n)/\mathbb{F}(\infty, u_n) = \mathbb{F}(x, y)$ shows that

$$\begin{aligned} & \frac{\mathbb{F}(x, u_n y) + t_n \varphi_n(x, u_n y)}{\mathbb{F}(\infty, u_n) + t_n \varphi_n(\infty, u_n)} - \mathbb{F}(x, y) \\ &= \mathbb{F}(x, y) \left(\frac{1 + t_n \frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, u_n y)}}{1 + t_n \frac{\varphi_n(\infty, u_n)}{\mathbb{F}(\infty, u_n)}} - 1 \right) \\ &= \mathbb{F}(x, y) \left(\left(1 + t_n \frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, u_n y)} \right) \left(1 - t_n u_n^{1/\gamma} \varphi_n(\infty, u_n) (1 + \epsilon_n) \right) (1 + \theta_n(x, y)) - 1 \right) \\ &= \mathbb{F}(x, y) \left(t_n \left(\frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, u_n y)} - u_n^{1/\gamma} \varphi_n(\infty, u_n) \right) + R_n(x, y) \right), \end{aligned}$$

with $\epsilon_n \rightarrow 0$ a sequence of real numbers not depending upon x and y , and with

$$R_n(x, y) := t_n u_n^{1/\gamma} \varphi_n(\infty, u_n) \epsilon_n + (t_n u_n^{1/\gamma})^2 \varphi_n(\infty, u_n) \frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, y)} (1 + \epsilon_n).$$

This implies that

$$\|g_n(\varphi_n, u_n) - g(\varphi, 1)\| \leq B_{1,n} + B_{2,n} + B_{3,n} + B_{4,n},$$

where the four terms $B_{1,n}, \dots, B_{4,n}$ are detailed below and will be proved to converge to zero as $n \rightarrow \infty$.

First term

$$\begin{aligned}
B_{1,n} &:= \|u_n^{1/\gamma} \varphi_n(\cdot, u_n) - \varphi(\cdot, \cdot)\| \\
&\leq \|u_n^{1/\gamma} \varphi_n(\cdot, u_n) - \varphi_n(\cdot, u_n)\| + \|\varphi_n(\cdot, u_n) - \varphi(\cdot, \cdot)\| \\
&= |u_n^{1/\gamma} - 1| \|\varphi_n(\cdot, u_n)\| + \|\varphi_n(\cdot, u_n) - \varphi(\cdot, \cdot)\| \\
&\leq |u_n^{1/\gamma} - 1| \|\varphi_n(\cdot, u_n)\| + \|\varphi_n(\cdot, u_n) - \varphi(\cdot, u_n)\| \\
&\quad + \|\varphi(\cdot, u_n) - \varphi(\cdot, \cdot)\| \\
&\leq |u_n^{1/\gamma} - 1| \|\varphi_n(\cdot, u_n)\| + u_n^{-\beta} \|\varphi_n(x, y) - \varphi(x, y)\|_{\infty, \beta} \\
&\quad + H_\varphi |u_n - u|^{\beta\gamma},
\end{aligned}$$

where $H_\varphi := \sup\{|y - y'|^{-\beta\gamma} |\varphi(x, y) - \varphi(x, y')|, x \in \mathbb{R}^d, y, y' \geq 1/2\}$ is finite since $\varphi \in \mathbf{D}_0$. The first two terms converge to 0 since $u_n \rightarrow 1$ and $\varphi_n \rightarrow \varphi$ in \mathbf{D} . The third term converges to zero since H_φ is finite.

Second term

$$\begin{aligned}
B_{2,n} &:= \|(u_n^{1/\gamma} \varphi_n(\infty, u_n) - \varphi(\infty, 1))\mathbb{F}\| \\
&\leq (|u_n^{1/\gamma} \varphi_n(\infty, u_n) - \varphi_n(\infty, u_n)| + |\varphi_n(\infty, u_n) - \varphi(\infty, 1)|) \|\mathbb{F}\|.
\end{aligned}$$

But $\|\mathbb{F}\|$ is finite since $\beta\gamma < \varepsilon < 1/2$, from where $B_{2,n} \rightarrow 0$ by similar arguments as those used for $B_{1,n}$.

Third term

$$B_{3,n} := \|u_n^{1/\gamma} \varphi_n(\infty, u_n) \epsilon_n \mathbb{F}\| \leq |u_n^{1/\gamma} \varphi_n(\infty, u_n)| \times |\epsilon_n| \times \|\mathbb{F}\|.$$

Since $\|\mathbb{F}\|$ is finite, since $|u_n^{1/\gamma} \varphi_n(\infty, u_n)|$ is a converging sequence, and since $|\epsilon_n| \rightarrow 0$, we deduce that $B_{3,n} \rightarrow 0$.

Fourth term

$$\begin{aligned}
B_{4,n} &:= (1 + |\epsilon_n|) \left\| (t_n u_n^{1/\gamma})^2 \varphi_n(\infty, u_n) \varphi_n(\cdot, u_n) \right\| \\
&\leq (1 + |\epsilon_n|) \left| (t_n u_n^{1/\gamma})^2 \varphi_n(\infty, u_n) \right| \times \|\varphi_n(\cdot, u_n)\|.
\end{aligned}$$

Since $\varphi_n \rightarrow \varphi$ in $L^{\infty, \beta}(\mathbb{R}^d \times [1/2, \infty[)$, the same arguments as for $B_{3,n}$ entail the convergence to zero of $B_{4,n}$. \square

5.3.2 Proof of theorem 5.2

Let $x \in \mathbb{R}^d$, which will be kept fixed in all this section. To prove the asymptotic normality of $\hat{\sigma}_n(x)$ we first establish the asymptotic normality of the numerator and the denominator separately. Note that we don't need to study their joint asymptotic normality, because only the numerator will rule the asymptotic normality of $\hat{\sigma}_n(x)$, as its rate of convergence is the slowest.

Proposition 5.4. *Assume that $(p_n)_{n \geq 1}$ and $(h_n)_{n \geq 1}$ both converge to 0 and satisfy $np_n h_n^d \rightarrow 0$. We have*

$$\frac{1}{\sqrt{np_n h_n^d}} \sum_{i=1}^n \frac{\mathbb{1}_{\{|X_i-x| \leq h_n, Y_i > \mathbf{y}_n\}} - \mathbb{P}(|X_i-x| \leq h_n, Y_i > \mathbf{y}_n)}{\sqrt{\sigma(x)f(x)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \quad (5.16)$$

$$\frac{1}{\sqrt{nh_n^d}} \sum_{i=1}^n \frac{\mathbb{1}_{\{|X_i-x| \leq h_n\}} - \mathbb{P}(|X_i-x| \leq h_n)}{\sqrt{f(x)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \text{ and} \quad (5.17)$$

$$\frac{1}{\sqrt{np_n}} \sum_{i=1}^n (\mathbb{1}_{\{Y_i > \mathbf{y}_n\}} - p_n) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \quad (5.18)$$

Démonstration. Note that (5.18) is the central limit theorem for binomial(n, p_n) sequences with $p_n \rightarrow 0$ and $np_n \rightarrow \infty$, while (5.17) is the well known pointwise asymptotic normality of the Parzen-Rosenblatt density estimator. The proof of (5.16) is a straightforward use of the Lindeberg-Levy Theorem (see, e.g [8], Theorem 27.2 p. 359). First define

$$Z_{i,n} := \frac{\mathbb{1}_{\{|X_i-x| \leq h_n, Y_i > \mathbf{y}_n\}} - \mathbb{P}(|X_i-x| \leq h_n, Y_i > \mathbf{y}_n)}{\sqrt{np_n h_n^d} \sqrt{\sigma(x)f(x)}}$$

and remark that $\mathbb{E}(Z_{i,n}) = 0$. Moreover we can write,

$$\begin{aligned} \mathbb{E} \left(\mathbb{1}_{\{|X_i-x| \leq h_n, Y_i > \mathbf{y}_n\}} \right) &= \int_{B(x,h)} \mathbb{P}(Y_i > \mathbf{y}_n | X_i = z) P_X(dz) \\ &\approx \int_{B(x,h)} \sigma(z) p_n P_X(dz) \quad (a) \\ &\approx \sigma(x) f(x) p_n h_n^d, \quad (b) \end{aligned}$$

where (a) is a consequence of the uniformity in assumption (5.3), while equivalence (b) holds by our assumptions upon the regularity of both f and σ in Theorem 5.2. We conclude that $\sup\{|n \text{Var}(Z_{i,n}) - 1|, i = 1, \dots, n\} \rightarrow 0$. Note that we can invoke the Lindeberg-Levy Theorem if for all $\varepsilon > 0$, we have

$$\sum_{i=1}^n \int_{\{|Z_{i,n}| > \varepsilon\}} Z_{i,n}^2 P_X(dx) \rightarrow 0.$$

This convergence holds since the set $\{Z_{i,n} > \varepsilon\}$ can be rewritten

$$\left\{ \left| \mathbb{1}_{\{|X_i-x| \leq h_n, Y_i > \mathbf{y}_n\}} - \mathbb{P}(|X_i-x| \leq h_n, Y_i > \mathbf{y}_n) \right| \geq \varepsilon \sqrt{\sigma(x)f(x)} \sqrt{np_n h_n^d} \right\},$$

which is empty when n is large enough, since $np_n h_n^d \rightarrow \infty$. This proves (5.16). \square

Now, writing

$$\hat{\sigma}_n(x) = \frac{n}{\sum_{i=1}^n \mathbb{1}_{\{Y_i > \mathbf{y}_n\}}} \times \frac{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i| < h_n\}} \mathbb{1}_{\{Y_i > \mathbf{y}_n\}}}{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i| < h_n\}}},$$

we have

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}} \sum_{i=1}^n Z_{i,n}^\#} \times \frac{\frac{\mathbb{P}(|X-x| \leq h_n, Y > \mathbf{y}_n)}{p_n h_n^d} + \sqrt{\frac{f(x)\sigma(x)}{np_n h_n^d}} \sum_{i=1}^n Z_{i,n}}{\frac{\mathbb{P}(|X-x| \leq h_n)}{h_n^d} + \sqrt{\frac{f(x)}{nh_n^d}} \sum_{i=1}^n \tilde{Z}_{i,n}}, \text{ where}$$

$$\tilde{Z}_{i,n} := \frac{\mathbb{1}_{\{|X_i-x| \leq h_n\}} - \mathbb{P}(|X_i-x| \leq h_n)}{\sqrt{f(x)} \sqrt{nh_n^d}}, \text{ and}$$

$$Z_{i,n}^\# := \frac{\mathbb{1}_{\{Y_i > \mathbf{y}_n\}} - p_n}{\sqrt{np_n}}.$$

Now write

$$\sigma_{h_n}(x) := \frac{\mathbb{P}(|X-x| \leq h_n, Y > \mathbf{y}_n)}{p_n h_n^d f(x)}.$$

Since f is continuous and bounded away from zero on a neighbourhood of x we have

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}} \sum_{i=1}^n Z_{i,n}^\#} \frac{\sigma_{h_n}(x) f(x) (1 + \varepsilon_{n,1}) + \sqrt{\frac{f(x)\sigma(x)}{np_n h_n^d}} \sum_{i=1}^n Z_{i,n}}{f(x) (1 + \varepsilon_{n,2}) + \sqrt{\frac{f(x)}{nh_n^d}} \sum_{i=1}^n \tilde{Z}_{i,n}},$$

with $|\varepsilon_{n,1}| \vee |\varepsilon_{n,2}| \rightarrow 0$. Now a Taylor expansion of the denominator gives

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}} \sum_{i=1}^n Z_{i,n}^\#} \left(\sigma_{h_n}(x) + \sqrt{\frac{\sigma(x)}{np_n h_n^d f(x)}} \sum_{i=1}^n Z_{i,n} \right) \\ \times \left(1 - \sqrt{\frac{1}{nh_n^d f(x)}} \sum_{i=1}^n \tilde{Z}_{i,n} + o_{\mathbb{P}} \left(\sqrt{\frac{1}{nh_n^d f(x)}} \right) \right).$$

By similar arguments, remarking that $(nh_n^d)^{-1} = o((np_n h_n^d)^{-1})$ we have, by (5.16) and (5.17) :

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}} \sum_{i=1}^n Z_{i,n}^\#} \left(\sigma_{h_n}(x) + \sqrt{\frac{\sigma(x)}{np_n h_n^d f(x)}} \sum_{i=1}^n Z_{i,n} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{np_n h_n^d}} \right) \right).$$

Moreover, with one more Taylor expansion of the denominator, we have, by (5.18),

$$\hat{\sigma}_n(x) = \sigma_{h_n}(x) + \sqrt{\frac{\sigma(x)}{np_n h_n^d f(x)}} \sum_{i=1}^n Z_{i,n} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{np_n h_n^d}} \right),$$

which entails

$$\sqrt{np_n h_n^d} (\hat{\sigma}_n(x) - \sigma_{h_n}(x)) = \sqrt{\frac{\sigma(x)}{f(x)}} \sum_{i=1}^n Z_{i,n} + o_{\mathbb{P}}(1).$$

The asymptotic normality of $\sum_{i=1}^n Z_{i,n}$ gives

$$\sqrt{np_n h_n^d}(\hat{\sigma}_n(x) - \sigma_{h_n}(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma(x)}{f(x)}\right).$$

The proof is achieved by noticing that assumption (5.3) entails

$$\begin{aligned} \sqrt{np_n h_n^d}|\sigma_{h_n}(x) - \sigma(x)| &= \sqrt{np_n h_n^d} \left| \frac{\mathbb{P}(|X - x| \leq h_n, Y > \mathbf{y}_n)}{f(x)h_n^d \mathbb{P}(Y > \mathbf{y}_n)} - \sigma(x) \right| \\ &= \sqrt{np_n h_n^d} \left| \frac{\mathbb{P}(Y > \mathbf{y}_n | X \in B(x, h_n))}{\mathbb{P}(Y > \mathbf{y}_n)} - \sigma(x) \right| \\ &= O\left(\sqrt{np_n h_n^d} A(1/p_n)\right) \rightarrow 0. \end{aligned}$$

5.3.3 Proof of theorem 5.3

For sake of clarity, we first express conditions (5.3) and (5.4) in terms of the tail quantile function U : we have, uniformly in x ,

$$\left| \frac{U_x(1/\alpha_n)}{U(\sigma(x)/\alpha_n)} - 1 \right| = O(A_n) \quad \text{and} \quad \left| \frac{U(1/\alpha_n)}{xU(x^{-1/\gamma}/\alpha_n)} - 1 \right| = O(A_n),$$

where $A_n := A(1/p_n)$. Start the proof by splitting the quantity of interest into four parts,

$$\begin{aligned} \log\left(\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)}\right) &= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)} \left(\frac{\hat{p}_n \hat{\sigma}_n(x)}{\alpha_n}\right)^{\hat{\gamma}_n}\right) \\ &= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)} \left(\frac{p_n \hat{\sigma}_n(x)}{\alpha_n}\right)^{\hat{\gamma}_n} \left(\frac{\hat{p}_n}{p_n}\right)^{\hat{\gamma}_n}\right) \\ &= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)}\right) + \hat{\gamma}_n \log\left(\frac{p_n}{\alpha_n}\right) + \hat{\gamma}_n \log(\hat{\sigma}_n(x)) + \hat{\gamma}_n \log\left(\frac{\hat{p}_n}{p_n}\right) \\ &= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)} \left(\frac{p_n}{\alpha_n}\right)^\gamma\right) + (\hat{\gamma}_n - \gamma) \log\left(\frac{p_n}{\alpha_n}\right) \\ &\quad + \hat{\gamma}_n \log(\hat{\sigma}_n(x)) + \hat{\gamma}_n \log\left(\frac{\hat{p}_n}{p_n}\right). \end{aligned}$$

Moreover we can see that

$$\begin{aligned} \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)} \left(\frac{p_n}{\alpha_n}\right)^\gamma\right) &= \log\left(\frac{U(1/p_n)}{U_x(1/\alpha_n)} \left(\frac{p_n}{\alpha_n}\right)^\gamma\right) \\ &= \log\left(\frac{U(1/p_n)}{U(1/\alpha_n)} \left(\frac{p_n}{\alpha_n}\right)^\gamma\right) + \log\left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)}\right) \end{aligned}$$

Then, we write

$$\frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \log\left(\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)}\right) = Q_{1,n} + Q_{2,n} + Q_{3,n} + Q_{4,n}, \quad \text{with}$$

$$\begin{aligned}
Q_{1,n} &:= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \log \left(\frac{U(1/p_n)}{U(1/\alpha_n)} \left(\frac{p_n}{\alpha_n} \right)^\gamma \right), \\
Q_{2,n} &:= \sqrt{np_n}(\hat{\gamma}_n - \gamma), \\
Q_{3,n} &:= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \left(\hat{\gamma}_n \log(\hat{\sigma}_n(x)) + \log \left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)} \right) \right), \\
Q_{4,n} &:= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \hat{\gamma}_n \log \left(\frac{\hat{p}_n}{p_n} \right).
\end{aligned}$$

First, condition (5.4) entails

$$\begin{aligned}
Q_{1,n} &\sim \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \left(\frac{U(1/p_n)}{U(1/\alpha_n)} \left(\frac{p_n}{\alpha_n} \right)^\gamma - 1 \right) \\
&\sim \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \left(\frac{U((\alpha_n/p_n)^{\alpha\gamma}/\alpha_n)}{U(1/\alpha_n)} \left(\frac{p_n}{\alpha_n} \right)^\gamma - 1 \right) \\
&= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} O(A_n).
\end{aligned}$$

Since $\alpha_n = o(p_n)$, we see that $\log(p_n/\alpha_n)^{-1} \rightarrow 0$ together with $\sqrt{np_n}A_n \rightarrow 0$ entails that $Q_{1,n} \rightarrow 0$.

Second, we know by Theorem 7.1 that $Q_{2,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2)$.

Now $Q_{3,n}$ is studied remarking that

$$\log \left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)} \right) = \log \left(\frac{U(\sigma(x)/\alpha_n)}{U_x(1/\alpha_n)} \right) + \log \left(\frac{U(1/\alpha_n)}{\sigma(x)^{-\gamma} U(\sigma(x)/\alpha_n)} \right) - \gamma \log(\sigma(x)).$$

Together with (5.3) and (5.4), one has

$$\log \left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)} \right) = O(A_n) - \gamma \log(\sigma(x)).$$

Consequently

$$Q_{3,n} = \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} O(A_n) + \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \left(\hat{\gamma}_n \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x)) \right).$$

Hence, the asymptotic behavior of $Q_{3,n}$ is ruled by that of $\hat{\gamma}_n \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x))$, which we split into

$$(\hat{\gamma}_n - \gamma) \log(\hat{\sigma}_n(x)) + \gamma \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x)).$$

Now Theorem 7.1 entails that

$$\frac{\log(\hat{\sigma}_n(x))}{\log(p_n/\alpha_n)} \sqrt{np_n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathbb{P}} 0.$$

Moreover Theorem 5.2 together with the delta-method show that

$$\begin{aligned}
&\frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} (\gamma \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x))) \\
&= \frac{\sqrt{np_n h_n^d}}{\sqrt{h_n^d} \log(p_n/\alpha_n)} (\gamma \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x))) \xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

Finally, using the notation introduced in the proof of Theorem 5.2, we have

$$\begin{aligned} \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \log\left(\frac{\hat{p}_n}{p_n}\right) &= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \log\left(1 + \frac{1}{\sqrt{np_n}} \sum_{i=1}^n Z_{i,n}^\sharp\right) \\ &\sim \frac{1}{\log(p_n/\alpha_n)} \sum_{i=1}^n Z_{i,n}^\sharp + o_{\mathbb{P}}\left(\frac{1}{\log(p_n/\alpha_n)}\right) \\ &\xrightarrow{\mathbb{P}} 0, \end{aligned}$$

which proves that $Q_{4,n} \xrightarrow{\mathbb{P}} 0$ since $\hat{\gamma}_n \xrightarrow{\mathbb{P}} \gamma$.

5.4 Appendix

Lemme 5.6. *For fixed $n \geq 1$, let $(Y_i)_{1 \leq i \leq n}$ be a sequence of i.i.d. random variables taking values in $(\mathfrak{X}, \mathcal{X})$. Let be $E = (E_i)_{1 \leq i \leq n}$ a n -uple of independent Bernoulli random variables independent from Y_i . Write*

$$\nu(k) := \sum_{i=1}^k E_i, \quad k \leq n.$$

Then we have

$$\sum_{i=1}^n \delta_{Y_i} E_i \stackrel{\mathcal{L}}{=} \sum_{i=1}^{\nu(n)} \delta_{Y_i}, \quad (5.19)$$

where the equality in law is understood as on the sigma algebra spanned by all Borel positive functions on $(\mathfrak{X}, \mathcal{X})$. Moreover if the (Y_i) are almost surely positive, then

$$\max_{i=1, \dots, n} Y_i E_i \stackrel{\mathcal{L}}{=} \max_{i=1, \dots, \nu(n)} Y_i. \quad (5.20)$$

Démonstration. Note that (5.19) is exactly Khinchin's equality (see [65, p. 307, (14.6)]). We shall now prove (5.20). $e \in \{0, 1\}^n$, and let g be real measurable and positive function. Since the variables $(Y_i)_{1 \leq i \leq n}$ are i.i.d and independent from E we have, for any given permutation σ of $\llbracket 1, n \rrbracket$,

$$(Y_1, \dots, Y_n) \stackrel{\mathcal{L}}{=} (Y_1, \dots, Y_n)|_{E=e} \stackrel{\mathcal{L}}{=} (Y_{\sigma(1)}, \dots, Y_{\sigma(n)})|_{E=e}$$

by exchangeability. Now define σ by

$$\sigma(k) := \begin{cases} \sum_{j=1}^i e_j & \text{if } e_i = 1 \\ n - \sum_{j=1}^i (1 - e_j) & \text{if } e_i = 0 \end{cases} \quad 1 \leq i \leq n.$$

Write $s(e) := \sum_{i=1}^n s(e_i)$ for the total number of ones in (e_1, \dots, e_n) . By construction, the indices i for which $e_i = 1$ are mapped injectively to the set of first indices $\llbracket 1, s(e) \rrbracket$, while those for which $e_i = 0$ are injectively mapped into $\llbracket s(e) + 1, n \rrbracket$. Since e has fixed and non random coordinates, we have

$$(Y_1 e_1, \dots, Y_n e_n)|_{E=e} \stackrel{\mathcal{L}}{=} (Y_{\sigma(1)} e_1, \dots, Y_{\sigma(n)} e_n)|_{E=e}.$$

Hence

$$\begin{aligned}
\max_{i=1,\dots,n} Y_i e_i \mid_{E=e} &\stackrel{\mathcal{L}}{=} \max_{i=1,\dots,n} Y_i e_i \\
&\stackrel{\mathcal{L}}{=} \max_{i=1,\dots,n} Y_{\sigma(i)} e_{\sigma(i)} \\
&\stackrel{\mathcal{L}}{=} \max_{i=1,\dots,s(e)} Y_{\sigma(i)} \quad (\text{a}) \\
&\stackrel{\mathcal{L}}{=} \max_{i=1,\dots,s(e)} Y_i \quad (\text{b}) \\
&\stackrel{\mathcal{L}}{=} \max_{i=1,\dots,s(e)} Y_i \mid_{E=e} \\
&\stackrel{\mathcal{L}}{=} \max_{i=1,\dots,s(E)} Y_i \mid_{E=e} ,
\end{aligned}$$

where (a) holds because $e_{\sigma(i)} = 0$ for $i > s(e)$ by construction and the Y_i are a.s. positive, while (b) is obtained by noticing that $F_e(Y_{\sigma(1)}, \dots, Y_{\sigma(n)}) \stackrel{\mathcal{L}}{=} F_e(Y_1, \dots, Y_n)$ with

$$F_e : (y_1, \dots, y_n) \mapsto \max_{i=1,\dots,s(e)} y_i.$$

Unconditioning upon E gives (5.20) \square .

\square

The coupling method in extreme value theory

Résumé

A coupling method is developed for univariate extreme value theory, providing an alternative to the use of the tail empirical/quantile processes. We emphasize the Peak-over-Threshold approach that approximates the distribution above high threshold by the Generalized Pareto Distribution (GPD) and compare the empirical distribution of exceedances to the empirical distribution associated with the limit GPD model. Sharp bounds for their Wasserstein distance in the second order Wasserstein space are provided. As an application, we recover standard results on the asymptotic behavior of the Hill estimator, the Weissman extreme quantile estimator or the probability weighted moment estimators, shedding some new light on the theory.

6.1 Introduction

The purpose of extreme value theory (EVT) is to make statistical inference on the tail region of a probability distribution, e.g. assess rare event probabilities or high order quantiles. In the extreme regime of interest, only a limited number of observations are usually available and classical non parametric methods break down due to the lack of relevant data. To circumvent this issue, extreme value theory assumes some kind of regularity in the tail of the distribution that makes extrapolation possible and allows to exploit moderately high observations for inference in the extreme regime. In the univariate setting, this results in semi-parametric methods where the distribution tail is approximated by a parametric model such as the generalized extreme value (GEV) or generalized Pareto (GP) distribution. Therefore, EVT is essentially an asymptotic theory where the statistical analysis focuses on k extreme observations in a sample of size n , for an intermediate sequence $k = k(n)$ satisfying $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$.

In the Peak-over-Threshold (PoT) method, the k largest order statistics are considered ; suitably rescaled, they follow approximately the GP distribution and the probability weighted moment (PWM) or maximum likelihood (ML) are classically used to estimate the tail parameters and extreme quantiles (see e.g. [30], [22] Section 2.3). In the block maxima (BM) method, the data sample is divided into k blocks of size $m = \lfloor n/k \rfloor$ and the suitably rescaled block maxima follow approximately a GEV distribution whose parameters can again be estimated with PWM or ML method

[40, 13, 14, 27]. In both cases, one has to deal with the fundamental misspecification inherent with EVT : the extreme sample follows only approximately the limiting model (GEV or GP). This framework also raises the difficult practical issue of the choice of the effective sample size k : larger values of k provide a smaller estimation variance, but also a poorer approximation by the limit model leading to a higher estimation bias. There is a rich literature on methods for threshold selection and bias reduction but we do not address these issues in the present paper. For a general background on extreme value statistics, the reader should refer to the monographs by [5] or [22].

Many results in EVT rely on the theory of empirical processes and use the tail empirical process or the tail quantile process. The purpose of the present paper is to develop an alternative approach emphasizing coupling arguments between the effective extreme sample and an ideal sample from the limiting model. Using the formalism of Wasserstein spaces, we obtain quantitative results measuring the quality of the approximation of the pre-limit model by the limiting model. As a first illustration, we give (in the unbiased case) a simple and elegant proof of the asymptotic normality of the celebrated Hill estimator [50] together with a non asymptotic upper bound for the Wasserstein distance to Gaussianity, see Corollary 6.3.

The philosophy of the coupling method in EVT is that, in the limit GP or GEV model, the behavior of the statistic of interest is often much simpler to analyze. The corresponding result in the PoT or BM framework taking into account the misspecification inherent to EVT is often much more difficult to prove and usually requires involved empirical process theory with weighted norms. We present here how the coupling method can be used to carry over the result from the limit model to the pre-limit model using Wasserstein distance estimates, which works quite smoothly under Lipschitz assumptions. To be more specific, the asymptotic normality of the Hill estimator for Pareto random variables is a straightforward application of the central limit theorem. The PWM based estimators in the GEV and GP models were considered by [53] and [52] respectively and their analysis relies on the general theory of L -estimation for i.i.d. samples. Using the coupling method, these results can be carried over to the PoT or BM framework quite easily, offering an original and conceptually important point of view.

Let us mention that the last decades have seen a fast development of the use of coupling methods, Wasserstein spaces and optimal transport in statistics. The reader should refer to the excellent review by [66] and the reference therein. In this paper, we see how these insightful and powerful notions apply successfully in EVT. Our approach share similarities with the results in total variation obtained by [70, 71] under stronger assumption of absolute continuity of the underlying distribution. In a multivariate setting and with a very different perspective, optimal transport for heavy tailed multivariate distribution is considered by [24].

The structure of the present paper is as follows. In Section 6.2, we introduce the required background on coupling theory and Wasserstein spaces. Then, in a general framework not related to EVT, we state in Theorem 6.1 the equality of the Wasserstein distance (in the second order Wasserstein space) between the distributions of the empirical measures of two i.i.d. samples with the same size and the Wasserstein distance (in the first order Wasserstein space) of the underlying distributions

generating the samples. The proof uses elementary coupling arguments together with the more subtle Kantorovitch duality [83, Chapter 5]. We believe this result is quite powerful and may find application in many different situations. In Section 6.3, we see how it applies to EVT, mostly in the PoT framework, and introduce some background on EVT. In Section 3.2, for the sake of clarity, we first consider the heavy-tailed case and provide a sharp estimate between the empirical measure of observations above high-threshold and the empirical measure associated with i.i.d. Pareto observations (Theorem 6.2). Statistical consequences for the Hill estimator and Weissman extreme quantile estimates are discussed in Corollaries 6.3 and 6.4. Further extensions are considered in Section 3.3 : the asymptotic regime where bias occurs, the generalization to all domain of attractions and some results in the BM framework. All the proofs are gathered in Section 4.

We finally introduce some notations used throughout the paper. Given two real-valued sequences $(u_n)_{n \geq 1}$ and $(v_n)_{n \geq 1}$, the notation $u_n = o(v_n)$ (respectively $u_n = O(v_n)$) means that $u_n = \varepsilon_n v_n$ for some sequence $(\varepsilon_n)_{n \geq 1}$ converging to 0 (resp. bounded). Similarly, given two sequences of random variables $(U_n)_{n \geq 1}$ and $(V_n)_{n \geq 1}$, the notation $U_n = o_P(V_n)$ (resp. $U_n = O_P(V_n)$) means that $U_n = \varepsilon_n V_n$ for some sequence of random variables $(\varepsilon_n)_{n \geq 1}$ converging to 0 in probability (resp. bounded in probability). The notations $\stackrel{d}{=}$ and \xrightarrow{d} stand respectively for equality in distribution and convergence in distribution. All random variable are defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and for $p \in [1, \infty]$, we denote by $\|\cdot\|_{L^p}$ the usual norm on $L^p(\Omega, \mathcal{F}, \mathbb{P})$.

6.2 Coupling and sampling

We introduce some background on coupling theory and Wasserstein spaces as well as general results about coupling and sampling. More precisely, we evaluate the Wasserstein distance between the empirical distributions of two independent and identically distributed (i.i.d.) samples and between statistics thereof.

In the following, we consider a metric space (\mathcal{X}, d) endowed with its Borel σ -algebra $\mathcal{B}(\mathcal{X})$. We denote by $\mathcal{M}(\mathcal{X})$ the set of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and by δ_x the Dirac mass at $x \in \mathcal{X}$.

6.2.1 Background on coupling and Wasserstein spaces

The coupling method has long been an important tool in probability theory, see e.g. the monographs by [64] and [80]. A coupling between two probability measures $P_1, P_2 \in \mathcal{M}(\mathcal{X})$ is a pair of \mathcal{X} -valued random variables X_1, X_2 defined on a common probability space and such that $X_1 \sim P_1$ and $X_2 \sim P_2$ where the notation $X \sim P$ means that the random variable X has distribution P . This notion of coupling is crucial in the definition of the Wasserstein distance which is a powerful tool in statistics ; it is for instance central in the analysis of the bootstrap by [7]. We provide below some basic properties of Wasserstein spaces, more details are to be found in [83].

The Wasserstein distance of order $p \in [1, \infty]$ between two probability measures

$P_1, P_2 \in \mathcal{M}(\mathcal{X})$ is defined by

$$W_p(P_1, P_2) = \inf \left\{ \|d(X_1, X_2)\|_{L^p} : X_i \sim P_i, i = 1, 2 \right\}.$$

When (\mathcal{X}, d) is complete and separable, the infimum in this definition is achieved and there exist optimal couplings, i.e. couplings (X_1, X_2) such that $W_p(P_1, P_2) = \|d(X_1, X_2)\|_{L^p}$. The Wasserstein space of order p is defined as the set of probability measures

$$\mathcal{W}_p(\mathcal{X}) = \left\{ P \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} d(x_0, x)^p P(dx) < \infty \right\},$$

where $x_0 \in \mathcal{X}$ denotes some origin, i.e. some fixed point whose choice is irrelevant. Note that $(\mathcal{W}_p(\mathcal{X}), W_p)$ is itself a metric space, which is complete and separable if both $p \in [1, \infty)$ and (\mathcal{X}, d) is itself complete and separable. Also note that $W_p(P_n, P) \rightarrow 0$ is equivalent to the convergence in distribution $P_n \xrightarrow{d} P$ together with the convergence of moments

$$\int_{\mathcal{X}} d(x_0, x)^p P_n(dx) \rightarrow \int_{\mathcal{X}} d(x_0, x)^p P(dx) \quad \text{for some (all) } x_0 \in \mathcal{X}.$$

Example 6.1. When $\mathcal{X} = \mathbb{R}^d$ is endowed with the Euclidean norm, the Wasserstein space of order $p < \infty$ simply consists of all probability measures with finite moment of order p , i.e.

$$\mathcal{W}_p(\mathbb{R}^d) = \left\{ P \in \mathcal{M}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|x\|^p P(dx) < \infty \right\}.$$

For $p = \infty$, $\mathcal{W}_\infty(\mathbb{R}^d)$ is the set of probability measures with bounded support. In general, the Wasserstein distance between probability measures on \mathbb{R}^d cannot be computed explicitly, except in the unidimensional case. When $d = 1$, optimal couplings are provided by the probability integral transform. For $P \in \mathcal{M}(\mathbb{R})$, the random variable $X = F^\leftarrow(U)$ has distribution P , where F^\leftarrow denotes the quantile function of P and U a random variable with uniform distribution on $[0, 1]$. Using obvious notations, $X_1 = F_1^\leftarrow(U)$ and $X_2 = F_2^\leftarrow(U)$ is hence a coupling between $P_1, P_2 \in \mathcal{M}(\mathbb{R})$ and it turns out that this coupling is optimal so that

$$W_p(P_1, P_2) = \left(\int_0^1 |F_1^\leftarrow(u) - F_2^\leftarrow(u)|^p du \right)^{1/p}, \quad \text{if } p \in [1, \infty),$$

and $W_\infty(P_1, P_2) = \sup_{u \in (0,1)} |F_1^\leftarrow(u) - F_2^\leftarrow(u)|$.

Example 6.2. Given a metric space (\mathcal{X}, d) , the Wasserstein space $(\mathcal{W}_p(\mathcal{X}), W_p)$ is also a metric space so that we can iterate the construction and consider the second order Wasserstein space $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$. An element $\mathcal{P} \in \mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$ can be seen as the distribution of a random measure Π on (\mathcal{X}, d) satisfying the integrability condition $\mathbb{E}[\int_{\mathcal{X}} d(x_0, x)^p \Pi(dx)] < \infty$. In the next section, we will consider the empirical distribution of random samples so that random measures and second order Wasserstein spaces will naturally arise. The Wasserstein distance on $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$ is denoted by $W_p^{(2)}$ and defined by

$$W_p^{(2)}(\mathcal{P}_1, \mathcal{P}_2) = \inf \left\{ \|W_p(\Pi_1, \Pi_2)\|_{L^p} : \Pi_i \sim \mathcal{P}_i, i = 1, 2 \right\}.$$

Note that [63] have considered second order Wasserstein spaces when discussing Wasserstein barycenters for random measures.

6.2.2 Wasserstein distance between empirical distributions

Consider two i.i.d. samples X_1, \dots, X_n and X_1^*, \dots, X_n^* , with size $n \geq 1$, taking values in (\mathcal{X}, d) and with distribution P and P^* respectively. The corresponding empirical distributions are defined by

$$\Pi_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \quad \text{and} \quad \Pi_n^* = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^*}.$$

These are random measures on (\mathcal{X}, d) satisfying

$$\mathbb{E} \left[\int_{\mathcal{X}} d(x_0, x)^p \Pi_n(dx) \right] = \int_{\mathcal{X}} d(x_0, x)^p P(dx).$$

and similarly with Π_n^* and P^* . The expectation is finite if and only if $P \in \mathcal{W}_p(\mathcal{X})$ and, according to Example 6.2, the distribution of Π_n , denoted by P_{Π_n} , is then an element of the second order Wasserstein space $\mathcal{W}_p(\mathcal{W}_p(\mathcal{X}))$. The following Theorem states that the Wasserstein distance between the distributions of the empirical measures Π_n and Π_n^* is equal to the Wasserstein distance between the distributions P and P^* that have generated the samples.

Theorem 6.1. *Assume (\mathcal{X}, d) is complete and separable. Let $p \in [1, \infty)$ and $P, P^* \in \mathcal{W}_p(\mathcal{X})$. Then,*

$$W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) = W_p(P, P^*). \quad (6.1)$$

Remarque 6.1. *As will be stated in the proof, the inequality*

$$W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) \leq W_p(P, P^*)$$

always holds true, even with $p = \infty$, and relies on elementary coupling arguments. The assumptions $p < \infty$ and (\mathcal{X}, d) complete and separable are required for the converse inequality whose proof uses the Kantorovitch duality from optimal transport, see [83, Chapter 5].

Remarque 6.2. *Theorem 6.1 can be generalized to samples with random size or weighted samples with random weights. Consider $(X_i)_{i \geq 1}$ and $(X_i^*)_{i \geq 1}$ sequences of i.i.d. random variables with distribution P and P^* respectively. Let $(w_i)_{i \geq 1}$ be a random sequence of non-negative weights summing up to 1 and independent of the sequences of observations $(X_i)_{i \geq 1}$ and $(X_i^*)_{i \geq 1}$. The weighted empirical distributions are defined by*

$$\Pi = \sum_{i \geq 1} w_i \delta_{X_i} \quad \text{and} \quad \Pi^* = \sum_{i \geq 1} w_i \delta_{X_i^*}$$

and are random elements in $\mathcal{W}_p(\mathcal{X})$. Then, Theorem 6.1 still holds and the proof is readily adapted. The case of a sample with random size N corresponds to $w_i = 1/N$ for $i \leq N$ and $w_i = 0$ otherwise. Multinomial weights naturally appear when considering bootstrap procedures.

6.2.3 Wasserstein distance between sample statistics

In many cases, one is interested in statistics $S_n = S(X_1, \dots, X_n)$ and $S_n^* = S(X_1^*, \dots, X_n^*)$. Due to the exchangeability of i.i.d. samples, it is natural to consider symmetric statistics that can be written as a functional of the empirical distributions, that is $S_n = S(\Pi_n)$ and $S_n^* = S(\Pi_n^*)$ – the same letter S is used without risk of confusion. The following corollary of Theorem 6.1 provides a simple upper bound under a natural Lipschitz condition.

Corollary 6.1. *Let $p \in [1, \infty]$ and $P, P^* \in \mathcal{W}_p(\mathcal{X})$. Let (\mathcal{Y}, ρ) be a metric space and $S : \mathcal{W}_p(\mathcal{X}) \rightarrow \mathcal{Y}$ be a Lipschitz functional with Lipschitz constant $\text{Lip}(S)$. Then, the statistics $S_n = S(\Pi_n)$ and $S_n^* = S(\Pi_n^*)$ satisfy*

$$W_p\left(P_{S_n}, P_{S_n^*}\right) \leq \text{Lip}(S)W_p(P, P^*),$$

where the Wasserstein distance in the left-hand side is taken on the space $\mathcal{W}_p(\mathcal{Y})$ and $\text{Lip}(S) = \sup_{\pi_1 \neq \pi_2} \rho(S(\pi_1), S(\pi_2)) / W_p(\pi_1, \pi_2)$.

In extreme value theory, important examples of Lipschitz statistics are the moments and probability weighted moments, as stated in the next proposition.

Proposition 6.1. *i) Let $p \in [1, \infty]$ and $q \in [1, p]$ finite. For arbitrary fixed point $x_0 \in \mathcal{X}$, consider the moment functional defined, for $\pi \in \mathcal{W}_p(\mathcal{X})$, by*

$$S(\pi) = \left(\int_{\mathcal{X}} d(x_0, x)^q \pi(dx) \right)^{1/q}.$$

Then, $S : \mathcal{W}_p(\mathcal{X}) \rightarrow \mathbb{R}$ is Lipschitz with $\text{Lip}(S) = 1$.

ii) Let $p \in [1, \infty]$, $q \in [1, p]$ finite and $r, s \geq 0$. Consider the probability weighted moment functional defined, for $\pi \in \mathcal{W}_p(\mathbb{R})$, by

$$S(\pi) = \left(\int_0^1 F_{\pi}^{\leftarrow}(u)^q u^r (1-u)^s du \right)^{1/q}$$

with F_{π}^{\leftarrow} the quantile function of π . Then, $S : \mathcal{W}_p(\mathbb{R}) \rightarrow \mathbb{R}$ is Lipschitz with $\text{Lip}(S) \leq 1$.

6.3 Wasserstein distance estimates in extreme value theory

6.3.1 Background on univariate extreme value theory

Two main approaches exist in univariate extreme value theory : the Block Maxima (BM) method and the Peaks-over-Threshold (PoT) method. The Fisher-Tipett-Gnedenko theorem [41, 45] is the basis of the BM method and states that, if a distribution F satisfies

$$F^n(a_n \cdot + b_n) \xrightarrow{d} G(\cdot) \tag{6.2}$$

for some non-degenerate limit distribution G and normalizing sequences $a_n > 0$ and $b_n \in \mathbb{R}$, then, up to location and scale, the limit distribution G is necessarily equal to an extreme value distribution G_γ defined by

$$G_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma}), \quad 1 + \gamma x > 0. \quad (6.3)$$

The parameter $\gamma \in \mathbb{R}$ is called the extreme value index. The case $\gamma = 0$ corresponds to the limit as $\gamma \rightarrow 0$ and G_0 is the Gumbel distribution $G_0(x) = \exp(-\exp(-x))$. The link with block maxima is that Equation (6.2) is equivalent to the convergence of the rescaled maximum

$$a_n^{-1} \left(\max_{1 \leq i \leq n} X_i - b_n \right) \xrightarrow{d} G \quad \text{as } n \rightarrow \infty,$$

with X_1, \dots, X_n i.i.d. random variables with common distribution F . When Equation (6.2) holds, we say that F belongs to the max-domain of attraction of G , noted $F \in \mathcal{D}(G)$.

The characterization of the domain of attraction of G_γ is due to [45] in the case $\gamma \neq 0$ and to [20] in the case $\gamma = 0$: $F \in \mathcal{D}(G_\gamma)$ if and only if the tail quantile function $U(t) = F^{\leftarrow}(1 - 1/t)$, $t > 1$, satisfies the first order condition

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \quad x > 0, \quad (6.4)$$

for some positive normalizing function a .

The Peaks-over-Threshold method focuses on exceedances over a high threshold, that is observations of $X \sim F$ satisfying $X > u$ for u closed to the right endpoint $x^* = \sup\{x : F(x) > 1\}$. The Balkema-de Haan-Pickands Theorem [3, 68] states that the first order condition is also equivalent to the convergence

$$\lim_{u \rightarrow x^*} \frac{1 - F(u + f(u)x)}{1 - F(u)} = (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0, \quad (6.5)$$

for some positive function f . A possible choice is $f(u) = a(1/(1 - F(u)))$. In terms of exceedances, this writes

$$P_{\frac{X-u}{f(u)} | X > u} \xrightarrow{d} H_\gamma \quad \text{as } u \rightarrow x^*, \quad (6.6)$$

where H_γ is the Generalized Pareto (GP) distribution

$$H_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0. \quad (6.7)$$

In the framework of regular variation theory, the study of rates of convergence usually relies on second-order regular variation, see [22, Chapter B.3]. The so-called second-order condition reads

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \Psi_{\gamma, \rho}(x), \quad x > 0, \quad (6.8)$$

where $\rho \leq 0$ is the second order parameter, the normalizing function A is regularly varying at infinity with index ρ , eventually negative or positive and such that $\lim_{t \rightarrow \infty} A(t) = 0$ and

$$\Psi_{\gamma, \rho}(x) = \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds. \quad (6.9)$$

6.3.2 Analysis of the Peak-over-Threshold method

We analyze the PoT method and, for the sake of clarity, we consider first the case $\gamma > 0$. One can then take $a(t) = \gamma U(t)$ so that the first order condition (6.4) simplifies into standard regular variation

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad x > 0. \quad (6.10)$$

Equation (6.6) stating the convergence in distribution of normalized exceedances to the GP distribution is equivalent to

$$P_{u^{-1}X|X>u} \xrightarrow{d} P_\alpha, \quad \text{as } u \rightarrow \infty, \quad (6.11)$$

with $P_\alpha(x) = 1 - x^{-\alpha}$, $x > 1$, the Pareto distribution with index $\alpha = 1/\gamma > 0$. It is then standard to work under the simpler second order condition

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad x > 0. \quad (6.12)$$

Note that conditions (6.8) and (6.12) are not equivalent as the following simple example shows : for $U(t) = t^\gamma + 1 + t^{\gamma+\rho}$ with $\gamma > 0$ and $\gamma + \rho < 0$, condition (6.12) holds with $A(t) \sim t^{-\gamma}$ while (6.8) holds with $A(t) \sim t^\rho = o(t^{-\gamma})$ for a suitable choice of a , meaning that the GPD approximation is better than the Pareto one.

Pareto approximation of exceedance above high threshold

We consider the rate of the convergence (6.11) in the Wasserstein space and compare the exceedance distribution $P_{u^{-1}X|X>u}$ to the Pareto distribution P_α in $\mathcal{W}([1, \infty))$. Because the Pareto distribution has finite moments of order $p < \alpha$ only, we introduce the logarithmic distance

$$d(x, x') = |\log(x) - \log(x')|, \quad x, x' \in [1, \infty). \quad (6.13)$$

We will see below that this distance is also convenient for analyzing the behavior of the Hill estimator and the Weissman quantile estimator.

Let $t_0 > 1$ be such that $U(t_0) > 0$. For $t \geq t_0$, we define

$$A_p(t) = \begin{cases} \left(\int_1^\infty \left| \log \frac{U(zt)}{z^\gamma U(t)} \right|^p \frac{dz}{z^2} \right)^{1/p} & \text{if } p \in [1, \infty), \\ \sup_{z>1} \left| \log \frac{U(zt)}{z^\gamma U(t)} \right| & \text{if } p = \infty. \end{cases} \quad (6.14)$$

Without loss of generality, we can assume that the random variable $X \sim F$ is given by $X = U(Z)$ where $U(t) = F^{\leftarrow}(1 - 1/t)$ is the tail quantile function of X and Z follows a standard unit Pareto distribution.

Proposition 6.2. *Let $p \in [1, \infty]$ and consider the Wasserstein space $\mathcal{W}_p([1, \infty))$ with underlying distance (6.13).*

i) For $t \geq t_0$,

$$W_p(P_{U(t)^{-1}X|Z>t}, P_\alpha) = A_p(t).$$

ii) If $p < \infty$ and $F \in D(G_\gamma)$ with $\gamma > 0$, then the function A_p is bounded on $[t_0, \infty)$ and such that $\lim_{t \rightarrow \infty} A_p(t) = 0$.

iii) If the second-order condition (6.12) holds, then

$$\lim_{t \rightarrow \infty} \frac{A_p(t)}{A(t)} = c_p(\rho) := \begin{cases} \left(\int_1^\infty \left| \frac{z^\rho - 1}{\rho} \right|^p \frac{dz}{z^2} \right)^{1/p} & \text{if } p \in [1, \infty), \\ 1/|\rho| & \text{if } p = \infty \text{ and } \rho < 0, \\ +\infty & \text{if } p = \infty \text{ and } \rho = 0. \end{cases}$$

Remarque 6.3. We discuss the link between X and Z . Conditioning with respect to X is more natural but conditioning w.r.t. Z is mathematically much more convenient. When F is continuous on \mathbb{R} , the two coincide since we have $\{X > U(t)\} = \{Z > t\}$. When F is not continuous at $U(t)$, the conditioning event writes $\{X > U(t)\} = \{Z > 1/(1 - F(U(t)^-))\}$ with $F(x^-)$ the left limit of F at x . Then the conditioning event $\{Z > t\}$ is not measurable with respect to the σ -field generated by X but it can be recovered introducing an extra-randomness. Let W be uniform on $[0, 1]$ and independent of X and consider $V = F(X) + (F(X) - F(X^-))W$ and $Z = 1/(1 - V)$. Then V has a uniform distribution on $[0, 1]$ and satisfies $X = F^{\leftarrow}(V)$. It follows that Z has a standard unit Pareto distribution and satisfies $X = U(Z)$ so that the conditioning event $\{Z > t\}$ can be written in terms of X and an auxiliary random variable W .

Approximation of the empirical distribution of exceedances

Peaks-over-Threshold inference uses the fact that the GP distribution is a good approximation for the distribution of exceedances above high threshold and all the statistics of interest are built using only observations above high threshold. It is customary to use a random threshold equal to the order statistic of order $n - k$ so that the exceedances are the k top order statistics. In a first approach, a statistical procedure is often studied on the limiting model itself, that is assuming the exceedances are exactly Pareto distributed. This amounts to neglecting the misspecification inherent to extreme value theory. The error made in this approximation can be quantified by the Wasserstein distance between the empirical distributions.

Let X_1, \dots, X_n be an i.i.d. sample with distribution $F \in D(G_\gamma)$, $\gamma > 0$ and denote by $X_{1,n} \leq \dots \leq X_{n,n}$ the order statistics. Define the empirical distributions of exceedances above threshold $X_{n+1-k,n}$ by

$$\Pi_{n,k} = \frac{1}{k} \sum_{i=1}^k \delta_{X_{n+1-i,n}/X_{n-k,n}}.$$

We compare this empirical distribution to $\Pi_k^* = \frac{1}{k} \sum_{i=1}^k \delta_{X_i^*}$, where the sample X_1^*, \dots, X_k^* is i.i.d. with Pareto distribution P_α .

Without loss of generality, we can assume that $X_i = U(Z_i)$ with Z_1, \dots, Z_n i.i.d. random variables with standard unit Pareto distribution with order statistics $Z_{1,n} \leq \dots \leq Z_{n,n}$. We denote by $P_{\Pi_{n,k}|Z_{n-k,n}=t}$ the conditional distribution of $\Pi_{n,k}$ given $Z_{n-k,n} = t$ and compare it to $P_{\Pi_k^*}$ in the second order Wasserstein space $\mathcal{W}_p(\mathcal{W}_p([1, \infty)))$ where $[1, \infty)$ is equipped with the logarithmic distance.

Theorem 6.2. *Let $p \in [1, \infty)$.*

i) Let $t_0 \geq 1$ be such that $U(t_0) > 0$, then

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\Pi_k^*}\right) = A_p(t), \quad t \geq t_0,$$

with A_p defined in Equation (6.14).

ii) If $F \in D(G_\gamma)$ and $k = k(n)$ is an intermediate sequence, that is $k(n) \rightarrow \infty$ and $k(n)/n \rightarrow 0$ as $n \rightarrow \infty$, then

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_k^*}\right) \rightarrow 0 \quad \text{in probability.}$$

iii) If the second order condition (6.12) holds, then

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_k^*}\right) = c_p(\rho)A(n/k)(1 + o_P(1)),$$

with $c_p(\rho)$ defined in Proposition 6.2.

Remarque 6.4. *Following Remark 6.3, the conditioning with respect to $X_{n-k,n}$ is more natural, since $X_{n-k,n}$ is observed while $Z_{n-k,n}$ is not. Note that $X_{n-k,n} = U(t)$ is equivalent to $1/(1 - F(U(t)^-)) \leq Z_{n-k,n} \leq 1/(1 - F(U(t)))$. In the case F is continuous at $U(t)$, this conditioning event is equal to $Z_{n-k,n} = t$. When F is not continuous at $U(t)$, the conditioning event $Z_{n-k,n} = t$ can be expressed in terms of $X_{n-k,n}$ and an auxiliary random variable W .*

When X is positive and bounded away from 0, integration with respect to the threshold $Z_{n-k,n}$ provides the following upper bound for the Wasserstein distance between $P_{\Pi_{n,k}}$ (unconditional distribution) and $P_{\Pi_k^*}$. We denote by $\beta_{p,q}$ the density of the Beta distribution with parameter (p, q) .

Corollary 6.2. *If X is positive and bounded away from 0, then*

$$W_p^{(2)}\left(P_{\Pi_{n,k}}, P_{\Pi_k^*}\right) \leq \int_1^\infty A_p(t)\beta_{n-k,k+1}\left(1 - \frac{1}{t}\right) \frac{dt}{t^2}.$$

If furthermore $F \in D(G_\gamma)$ and $k = k(n)$ is an intermediate sequence, then

$$W_p^{(2)}\left(P_{\Pi_{n,k}}, P_{\Pi_k^*}\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The assumption $X > 0$ is not as restrictive as it may seem and is required because the exceedances $X_{n+1-i,n}/X_{n-k,n}$ are ill defined when the threshold $X_{n-k,n}$ is non positive. However, it is always possible to apply Corollary 6.2 to the censored variables $X'_i = \max(X_i, 1)$. We have $X_{n-k,n} > 1$ with high probability for large n and then the censoring does not affect the exceedances over threshold $X_{n-k,n}$, showing that the effect of censoring is negligible in the limit.

The Hill estimator and Weissman quantile estimates

Using the previous results, we analyze the Hill estimator [50]

$$\hat{\gamma}_{n,k} = \frac{1}{k} \sum_{i=1}^k \log(X_{n+1-i,n}/X_{n-k,n})$$

and compare it to the corresponding estimator in the limit Pareto model

$$\hat{\gamma}_k^* = \frac{1}{k} \sum_{i=1}^k \log X_i^*.$$

The random variables $\log(X_i^*)$ being i.i.d. with mean γ and variance γ^2 , the central limit theorem implies $\sqrt{k}(\hat{\gamma}_k^* - \gamma) \xrightarrow{d} \mathcal{N}(0, \gamma^2)$ as $n \rightarrow \infty$. Using our bound on $\mathcal{W}_p^{(2)}(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_k^*})$, we deduce the asymptotic normality of the Hill estimator $\hat{\gamma}_{n,k}$ and we even obtain a non asymptotic quantitative estimate for the Wasserstein distance to the Gaussian distribution.

Corollary 6.3. *Let $1 \leq p < \infty$ and $t_0 \geq 1$ such that $U(t_0) > 0$. For $1 \leq k \leq n$ and $t \geq t_0$,*

$$W_p\left(P_{\sqrt{k}(\hat{\gamma}_{n,k}-\gamma)|Z_{n-k,n}=t}, \mathcal{N}(0, \gamma^2)\right) \leq \sqrt{k}A_p(t) + \frac{(4 + 3\sqrt{2/\pi})\gamma}{\sqrt{k}}.$$

Corollary 6.3 implies that the Hill estimator is asymptotically normal and asymptotically independent of the threshold $X_{n-k,n}$ as soon as $k \rightarrow \infty$ and $\sqrt{k}A_p(Z_{n-k,n}) \rightarrow 0$ in probability. Under the second order condition (6.12) and with Proposition 6.2 at hand we retrieve the classical condition $\sqrt{k}A(n/k) \rightarrow 0$ for asymptotic normality without bias.

For $\alpha \in (0, 1)$, let $q(\alpha) = F^{\leftarrow}(1 - \alpha)$ be the quantile of order $1 - \alpha$. In order to estimate an extreme quantile $q(\alpha_n)$ with $\alpha_n = O(1/n)$, [86] proposed to extrapolate from the intermediate quantile $q(k/n)$ thanks to regular variation :

$$q(\alpha_n) = q(k/n) \frac{U(1/\alpha_n)}{U(n/k)} \approx q(k/n) \left(\frac{n}{k\alpha_n}\right)^\gamma.$$

This leads to Weissman extreme quantile estimate

$$\hat{q}(\alpha_n) = X_{n-k,n} \left(\frac{k}{n\alpha_n}\right)^{\hat{\gamma}_{n,k}}.$$

Corollary 6.4. *Let $1 \leq p < \infty$. Assume the second order condition (6.12) holds with $\rho < 0$. Let $k = k(n)$ be an intermediate sequence and $\alpha_n = o(k/n)$. Then, as $n \rightarrow \infty$,*

$$W_p\left(P_{v_n^{-1} \log \frac{\hat{q}(\alpha_n)}{q(\alpha_n)}|Z_{n-k,n}}, \mathcal{N}(0, \gamma^2)\right) = O_P\left(\sqrt{k}A(n/k) + 1/\sqrt{k}\right)$$

where $v_n = \log(k/(n\alpha_n))/\sqrt{k}$.

When $\sqrt{k}A(n/k) \rightarrow 0$, we deduce the asymptotic normality

$$v_n^{-1} \log \frac{\hat{q}(\alpha_n)}{q(\alpha_n)} \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

Setting $z_{1-u/2}$ the Gaussian quantile of order $1 - u/2$, we obtain

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(q(\alpha_n) / \hat{q}(\alpha_n) \in \left[e^{-z_{1-u/2} \hat{\gamma}_{n,k} v_n}, e^{z_{1-u/2} \hat{\gamma}_{n,k} v_n} \right] \right) = 1 - u.$$

This confidence interval for $q(\alpha_n)$ with asymptotic level $1 - u$ is valid for any sequence $\alpha_n = o(k/n)$ but it is accurate only if $v_n \rightarrow 0$, or equivalently $\log(n\alpha_n)/\sqrt{k} \rightarrow 0$. We retrieve the fact that the Weissman extreme quantile estimator is consistent if and only if $\log(n\alpha_n)/\sqrt{k} \rightarrow 0$. This excludes arbitrary small value of α_n and provides a limit for extrapolation. When $\log(n\alpha_n)/\sqrt{k} \rightarrow 0$, the delta-method can be used to recover the standard asymptotic normality

$$v_n^{-1} \left(\frac{\hat{q}(\alpha_n)}{q(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2),$$

see Theorem 4.3.8 and Corollary 4.3.9 in [22].

6.3.3 Extensions

Dealing with the bias

The asymptotic normality of the Hill estimator is classically considered under the second order condition (6.12) with the intermediate sequence $k = k(n)$ satisfying $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$. The next theorem extends the previous results to cover this case as well. We define the empirical measure

$$\Pi_{k,t}^* = \frac{1}{k} \sum_{i=1}^k \delta_{X_i^* \left(1 + A(t) \frac{(X_i^*)^{\rho/\gamma - 1}}{\rho} \right)}, \quad k \geq 1, t \geq 1,$$

with X_1^*, \dots, X_k^* i.i.d. with Pareto distribution P_α . The empirical measure $\Pi_{n,k}$ is the same as in Theorem 6.2.

Theorem 6.3. *Assume F satisfies the second order condition (6.12). Let $p \in [1, \infty)$ and $t_0 \geq 1$ such that $U(t_0) > 0$. For $t \geq t_0$,*

$$\begin{aligned} & W_p^{(2)} \left(P_{\Pi_{n,k} | Z_{n-k,n}=t}, P_{\Pi_{k,t}^*} \right) \\ & \leq \int_1^\infty \left| \log \frac{U(zt)}{z^\gamma U(t)} - \log \left(1 + A(t) \frac{z^\rho - 1}{\rho} \right) \right|^p \frac{dz}{z^2} = o(A(t)). \end{aligned}$$

If $k = k(n)$ is an intermediate sequence, then

$$W_p \left(P_{\sqrt{k}(\hat{\gamma}_{n,k} - \gamma) | Z_{n-k,n}}, \mathcal{N}(\sqrt{k}A(n/k)b(\rho), \gamma^2) \right) = O_P(1/\sqrt{k}) + o_P(\sqrt{k}A(n/k))$$

with bias given by $b(\rho) = \int_1^\infty \frac{z^\rho - 1}{\rho} \frac{dz}{z^2}$.

Under the condition $\sqrt{k}A(n/k) \rightarrow \lambda$, the asymptotic normality of the Hill estimator follows straightforwardly :

$$\sqrt{k}(\hat{\gamma}_{n,k} - \gamma) \xrightarrow{d} \mathcal{N}(\lambda b(\rho), \gamma^2),$$

with $\hat{\gamma}_{n,k}$ asymptotically independent of the threshold $X_{n-k,n}$.

Analysis of the case $\gamma < 1$ and probability weighted moments

We consider the theory for all domain of attractions, i.e. without the simplifying assumption $\gamma > 0$. Let X_1, \dots, X_n be an i.i.d. sample with distribution $F \in D(G_\gamma)$, $\gamma \in \mathbb{R}$ and X_1^*, \dots, X_k^* be i.i.d. with GP distribution H_γ . In view of Equation (6.6), we compare the empirical distributions

$$\Pi_{n,k} = \sum_{i=1}^k \delta_{\tilde{X}_{n+1-i,n}} \quad \text{and} \quad \Pi_k^* = \sum_{i=1}^k \delta_{X_i^*}$$

where $\tilde{X}_{n+1-i,n} = (X_{n+1-i,n} - X_{n-k,n})/f(X_{n-k,n})$. We focus on the case $\gamma < 1$ to ensure that the GP distribution H_γ defined by Equation (6.7) has a finite mean. Then $H_\gamma \in \mathcal{W}_p([0, \infty))$ for all $p \in [1, 1/\gamma_+)$ with $\gamma_+ = \max(0, \gamma)$ and the convention $1/0 = \infty$. When focusing on the PWM estimator of the extreme value index, the restriction $\gamma < 1$ is sensible because it is required for consistency. We define the function

$$A'_p(t) = \left(\int_1^\infty \left| \frac{U(zt) - U(t)}{a(t)} - \frac{z^\gamma - 1}{\gamma} \right|^p \frac{dz}{z^2} \right)^{1/p}, \quad t \geq 1.$$

Theorem 6.4. *Assume $F \in D(G_\gamma)$ with $\gamma < 1$ and let $p \in [1, 1/\gamma_+)$.*

i) (convergence to the GP distribution) In the Wasserstein space $\mathcal{W}_p([0, \infty))$,

$$W_p(P_{(X-U(t))/a(t)|Z>t}, H_\gamma) = A'_p(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

Under the second order condition (6.8),

$$A'_p(t) \sim c'_p(\gamma, \rho) A(t), \quad \text{as } t \rightarrow \infty,$$

$$\text{with } c'_p(\gamma, \rho) = \left(\int_1^\infty \Psi_{\gamma, \rho}(x)^p x^{-2} dx \right)^{1/p}.$$

ii) (PoT method) In the second order Wasserstein space $\mathcal{W}_p^{(2)}([0, \infty))$,

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\Pi_k^*}\right) = A'_p(t).$$

If $k = k(n)$ is an intermediate sequence and the second order condition (6.12) holds, then

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_k^*}\right) = c'_p(\gamma, \rho) A(n/k)(1 + o_P(1)).$$

Following [52] (see also [5] Chapter 5.3), the PWM estimators

$$\hat{M}_{k,n,s} = \frac{1}{k} \sum_{i=1}^k \left(1 - \frac{i}{k}\right)^s (X_{n+1-i,n} - X_{n-k,n}), \quad s = 0, 1,$$

are used to estimate the GP parameters after a suitable transformation. We define also the non-observed and rescaled statistic

$$\tilde{M}_{k,n,s} = \frac{\hat{M}_{k,n,s}}{f(X_{n-k,n})} = \frac{1}{k} \sum_{i=1}^k \left(1 - \frac{i}{k}\right)^s \tilde{X}_{n+1-i,n}$$

that we compare with the PWM estimators of the limit GP model

$$\hat{M}_{k,s}^* = \frac{1}{k} \sum_{i=1}^k \left(1 - \frac{i}{k}\right)^s X_{i,k}^*, \quad s = 0, 1,$$

with $X_{i,k}^* \leq \dots \leq X_{k,k}^*$ the order statistics of the GP sample. The estimation of γ relies on the following relations : for $H \sim H_{\gamma,\sigma}$ having a GP distribution with shape $\gamma < 1$ and scale $\sigma > 0$, it holds

$$m_0 := \mathbb{E}(H) = \frac{\sigma}{(1-\gamma)} \quad \text{and} \quad m_1 := \mathbb{E}(H(1 - F_H(H))) = \frac{\sigma}{2(2-\gamma)}$$

or equivalently

$$\gamma = 2 - \frac{m_0}{m_0 - 2m_1} \quad \text{and} \quad \sigma = \frac{2m_0m_1}{m_0 - 2m_1}.$$

This suggests the estimators

$$(\hat{\gamma}_{n,k}^{PWM}, \hat{\sigma}_{n,k}^{PWM}) = g(\hat{M}_{k,n,0}, \hat{M}_{k,n,1}) \quad (6.15)$$

where $g(x, y) = (2 - x/(x - 2y), 2xy/(x - 2y))$.

Corollary 6.5. *In the Wasserstein space $\mathcal{W}_p(\mathbb{R}^2)$,*

$$W_p\left(P_{(\tilde{M}_{k,n,s})_{s=0,1}|Z_{n-k,n}=t}, P_{(\hat{M}_{k,s}^*)_{s=0,1}}\right) \leq A'_p(t).$$

As a consequence, if $F \in D(G_\gamma)$ with $\gamma < 1/2$ satisfies the second order condition (6.8) and $k = k(n)$ is an intermediate sequence such that $\sqrt{k}A(n/k) \rightarrow 0$, then

$$\sqrt{k}\left(\tilde{M}_{k,n,0} - m_0, \tilde{M}_{k,n,1} - m_1\right) \xrightarrow{d} \mathcal{N}(0, \Gamma)$$

and

$$\left(\sqrt{k}(\hat{\gamma}_{n,k}^{PWM} - \gamma), \sqrt{k}\left(\frac{\hat{\sigma}_{n,k}^{PWM}}{f(X_{n-k,n})} - 1\right)\right) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

as $n \rightarrow \infty$, with the covariance matrices Γ and Σ respectively given by [54, Equation (5.3) p.28] and [22, Theorem 3.6.1].

Analysis of the block maxima method

We discuss shortly the block maxima method to demonstrate that our results for the PoT method can be extended to the BM case as well. We assume $F \in D(G_\gamma)$ so that the block maxima converge in distribution, that is Equation (6.2) holds with GEV limit distribution G_γ given by (6.3). We introduce the function $V(t) = F^{\leftarrow}(e^{-1/t})$, $t > 0$, so that $X = V(Z)$ has distribution F if Z has a unit Fréchet distribution. It is convenient since, for $m \geq 1$, the distribution of $V(mZ)$ is equal to the distribution F^m of the the block maxima with size m . Note that the functions U and V are asymptotically equivalent so that the first order condition (6.4) can be formulated equivalently with U replaced by V . However, when it comes to second

order condition, the conditions for U and V are not equivalent, see [31] and the discussion in the appendix there. The second-order condition for V reads

$$\lim_{t \rightarrow \infty} \frac{\frac{V(tx) - V(t) - \frac{x^\gamma - 1}{\gamma}}{a(t)}}{A(t)} = \Psi_{\gamma, \rho}(x), \quad x > 0, \quad (6.16)$$

with A regularly varying with index $\rho \leq 0$ and $\Psi_{\gamma, \rho}$ defined in Equation (6.9). Consider X_1, \dots, X_n an i.i.d sample with distribution F and, for $m \geq 1$ and $k = \lfloor n/m \rfloor$, the block maxima

$$M_i := \max_{(i-1)m+1 \leq j \leq im} (X_j - b_m)/a_m, \quad 1 \leq i \leq k.$$

Consider also M_1^*, \dots, M_k^* and i.i.d sample with GEV distribution G_γ . We compare the empirical measures

$$\Pi_{n,m} := \frac{1}{k} \sum_{i=1}^k \delta_{(M_i - b_m)/a_m} \quad \text{and} \quad \Pi_k^* := \frac{1}{k} \sum_{i=1}^k \delta_{X_i^*}$$

in the second order Wasserstein space $\mathcal{W}_p^{(2)}([0, \infty))$ and define

$$A_p''(t) = \left(\int_0^\infty \left| \frac{V(tz) - V(t)}{a(t)} - \frac{z^\gamma - 1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \right)^{1/p}, \quad t > 0. \quad (6.17)$$

We focus in the following theorem on the case $\gamma < 1$ that is useful for the analysis of probability weighted estimators. The case $\gamma > 0$ could easily be analysed with the logarithmic distance similarly as in Theorem 6.2 .

Theorem 6.5. *Let $\gamma < 1$ and $p \in [1, 1/\gamma_+)$.*

i) In the Wasserstein space $\mathcal{W}_p([0, \infty))$,

$$W_p \left(P_{(M_1 - b_m)/a_m}, G_\gamma \right) = A_p''(m), \quad m \geq 1.$$

Consequently, in the second order Wasserstein space $\mathcal{W}_p^{(2)}([0, \infty))$,

$$W_p^{(2)} \left(P_{\Pi_{n,m}}, P_{\Pi_k^*} \right) = A_p''(m), \quad 1 \leq m \leq n.$$

ii) If F has a finite left endpoint and satisfies the first order condition (6.4) and if $m = m(n) \rightarrow \infty$, then

$$W_p^{(2)} \left(P_{\Pi_{n,m}}, P_{\Pi_k^*} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

iii) If furthermore the second order condition (6.16) is satisfied, then

$$W_p^{(2)} \left(P_{\Pi_{n,m}}, P_{\Pi_k^*} \right) = c_p''(\gamma, \rho) A(m) (1 + o_P(1)), \quad \text{as } n \rightarrow \infty,$$

$$\text{with } c_p''(\gamma, \rho) = \left(\int_0^\infty \Psi_{\gamma, \rho}(z)^p e^{-1/z} \frac{dz}{z^2} \right)^{1/p}.$$

Probability weighted moments in the BM method can then be analyzed similarly as in the PoT method with similar results as in Corollary 6.5.

6.4 Proofs

6.4.1 Proofs related to Section 6.2

Proof of Theorem 6.1. The case $W_p(P, P^*) = 0$ is trivial because we have then $P = P^*$ so that the two samples have the same distribution $P_{\Pi_n} = P_{\Pi_n^*}$ and $W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) = 0$. We assume in the sequel that $W_p(P, P^*) > 0$.

Proof of the upper bound $W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) \leq W_p(P, P^*)$.

For the sake of generality, we consider $p \in [1, \infty]$ and do not assume the metric space (\mathcal{X}, d) to be complete and separable. Let $\varepsilon > 0$ and (\tilde{X}, \tilde{X}^*) be a coupling between P and P^* such that

$$\|d(\tilde{X}, \tilde{X}^*)\|_{L^p} \leq (1 + \varepsilon)W_p(P, P^*).$$

Taking i.i.d copies $(\tilde{X}_1, \tilde{X}_1^*), \dots, (\tilde{X}_n, \tilde{X}_n^*)$ of (\tilde{X}, \tilde{X}^*) , we obtain the coupling

$$\tilde{\Pi}_n = \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{X}_i} \stackrel{d}{=} \Pi_n \quad \text{and} \quad \tilde{\Pi}_n^* = \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{X}_i^*} \stackrel{d}{=} \Pi_n^*.$$

For fixed ω , define $(Z_\omega, Z_\omega^*) = (\tilde{X}_\iota(\omega), \tilde{X}_\iota^*(\omega))$ with ι uniformly distributed on $\{1, \dots, n\}$. Clearly $Z_\omega \sim \tilde{\Pi}_n(\omega)$ and $Z_\omega^* \sim \tilde{\Pi}_n^*(\omega)$ so that

$$W_p(\tilde{\Pi}_n(\omega), \tilde{\Pi}_n^*(\omega)) \leq \|d(Z_\omega, Z_\omega^*)\|_{L^p}. \quad (6.18)$$

In the case $1 \leq p < \infty$, Equation (6.18) yields

$$W_p(\tilde{\Pi}_n(\omega), \tilde{\Pi}_n^*(\omega)) \leq \|d(Z_\omega, Z_\omega^*)\|_{L^p} = \left(\frac{1}{n} \sum_{i=1}^n d(\tilde{X}_i(\omega), \tilde{X}_i^*(\omega))^p \right)^{1/p}$$

and we deduce

$$\begin{aligned} W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) &\leq \left[\mathbb{E}(W_p(\tilde{\Pi}_n, \tilde{\Pi}_n^*)^p) \right]^{1/p} \\ &\leq \left[\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n d(\tilde{X}_i, \tilde{X}_i^*)^p \right) \right]^{1/p} \\ &= \left[\mathbb{E}(d(\tilde{X}, \tilde{X}^*)^p) \right]^{1/p} \\ &\leq (1 + \varepsilon)W_p(P, P^*). \end{aligned}$$

In the case $p = +\infty$, Equation (6.18) yields similarly

$$W_\infty(\tilde{\Pi}_n(\omega), \tilde{\Pi}_n^*(\omega)) \leq \|d(Z_\omega, Z_\omega^*)\|_{L^\infty} = \max_{1 \leq i \leq n} d(\tilde{X}_i(\omega), \tilde{X}_i^*(\omega))$$

and we get

$$\begin{aligned} W_\infty(P_{\Pi_n}, P_{\Pi_n^*}) &\leq \|W_\infty(\tilde{\Pi}_n, \tilde{\Pi}_n^*)\|_{L^\infty} \\ &\leq \left\| \max_{1 \leq i \leq n} d(\tilde{X}_i, \tilde{X}_i^*) \right\|_{L^\infty} \\ &\leq \|d(\tilde{X}, \tilde{X}^*)\|_{L^\infty} \\ &\leq (1 + \varepsilon)W_\infty(P, P^*). \end{aligned}$$

In both cases, we obtain the announced upper bound by letting $\varepsilon \rightarrow 0$.

Proof of the lower bound $W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) \geq W_p(P, P^*)$.

We assume now the metric space (\mathcal{X}, d) to be complete and separable and $p \in [1, \infty)$. We first consider the case $p = 1$. By the Kantorovich-Rubinstein duality [60, 83, Remark 6.5],

$$W_1(P, P^*) = \sup \int_{\mathcal{X}} \varphi d(P - P^*)$$

with the infimum taken over all 1-Lipschitz functions $\varphi : \mathcal{X} \mapsto \mathbb{R}$. Hence, for all $\varepsilon > 0$, there exists a 1-Lipschitz function $\varphi : \mathcal{X} \mapsto \mathbb{R}$ such that

$$\int_{\mathcal{X}} \varphi d(P - P^*) \geq (1 - \varepsilon)W_1(P, P^*). \quad (6.19)$$

Define the map $\varphi^\sharp : \mathcal{W}_1(\mathcal{X}) \mapsto \mathbb{R}$ by $\varphi^\sharp(\pi) = \int_{\mathcal{X}} \varphi d\pi$. The integral is well defined because φ is Lipschitz and $\pi \in \mathcal{W}_1(\mathcal{X})$. We prove below that φ^\sharp is 1-Lipschitz : for $\pi_1, \pi_2 \in \mathcal{W}_1(\mathcal{X})$ and (Z_1, Z_2) a coupling between π_1 and π_2 ,

$$\varphi^\sharp(\pi_1) - \varphi^\sharp(\pi_2) = \int_{\mathcal{X}} \varphi d(\pi_1 - \pi_2) = \mathbb{E}[\varphi(Z_1) - \varphi(Z_2)]$$

whence, since φ is 1-Lipschitz,

$$|\varphi^\sharp(\pi_1) - \varphi^\sharp(\pi_2)| \leq \mathbb{E}[|\varphi(Z_1) - \varphi(Z_2)|] \leq \mathbb{E}[d(Z_1, Z_2)].$$

Taking the infimum over all couplings (Z_1, Z_2) , we get

$$|\varphi^\sharp(\pi_1) - \varphi^\sharp(\pi_2)| \leq W_1(\pi_1, \pi_2),$$

proving that φ^\sharp is 1-Lipschitz. Using the Kantorovich-Rubinstein duality again, we deduce

$$W_1^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) \geq \int_{\mathcal{W}_1(\mathcal{X})} \varphi^\sharp d(P_{\Pi_n} - P_{\Pi_n^*}), \quad (6.20)$$

where the right hand side equals

$$\begin{aligned} \int_{\mathcal{W}_1(\mathcal{X})} \varphi^\sharp d(P_{\Pi_n} - P_{\Pi_n^*}) &= \mathbb{E}[\varphi^\sharp(\Pi_n) - \varphi^\sharp(\Pi_n^*)] \\ &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \frac{1}{n} \sum_{i=1}^n \varphi(X_i^*)\right] \\ &= \mathbb{E}[\varphi(X) - \varphi(X^*)] \\ &= \int_{\mathcal{X}} \varphi d(P - P^*). \end{aligned} \quad (6.21)$$

Equations (6.19), (6.20) and (6.21) together entail

$$W_1^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}) \geq (1 - \varepsilon)W_1(P, P^*).$$

We obtain the announced lower bound by letting $\varepsilon \rightarrow 0$.

We now consider the case $p \in (1, +\infty)$. By the Kantorovich duality [83, Theorem 5.10(i)], for all $\varepsilon > 0$ there exist functions $\varphi \in L^1(\mathcal{X}, P)$ and $\psi \in L^1(\mathcal{X}, P^*)$ such that

$$\varphi(x) - \psi(y) \leq d^p(x, y), \quad x, y \in \mathcal{X},$$

and

$$\int_{\mathcal{X}} \varphi \, dP - \int_{\mathcal{X}} \psi \, dP^* \geq (1 - \varepsilon)W_p^p(P, P^*).$$

Define the functions $\varphi^\sharp, \psi^\sharp : \mathcal{W}_p(\mathcal{X}) \rightarrow \mathbb{R}$ by $\varphi^\sharp(\pi) = \int_{\mathcal{X}} \varphi \, d\pi$ and $\psi^\sharp(\pi) = \int_{\mathcal{X}} \psi \, d\pi$. We have $\varphi^\sharp \in L^1(\mathcal{W}(\mathcal{X}), P_{\Pi_n})$ and

$$\int_{\mathcal{W}(\mathcal{X})} \varphi^\sharp \, dP_{\Pi_n} = \mathbb{E}[\varphi^\sharp(\Pi_n)] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \varphi(X_i)\right] = \mathbb{E}[\varphi(X)] = \int_{\mathcal{X}} \varphi \, dP.$$

Similarly, $\psi^\sharp \in L^1(\mathcal{W}(\mathcal{X}), P_{\Pi_n^*})$. Furthermore, for $\pi_1, \pi_2 \in \mathcal{W}_p(\mathcal{X})$ and (Z_1, Z_2) a coupling between π_1 and π_2 , we have

$$\varphi^\sharp(\pi_1) - \psi^\sharp(\pi_2) = \mathbb{E}[\varphi(Z_1) - \psi(Z_2)] \leq \mathbb{E}[d^p(Z_1, Z_2)],$$

and, taking the infimum over all couplings,

$$\varphi^\sharp(\pi_1) - \psi^\sharp(\pi_2) \leq W_p^p(\pi_1, \pi_2).$$

We deduce, by the Kantorovich duality,

$$\begin{aligned} W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*})^p &\geq \int_{\mathcal{W}_p(\mathcal{X})} \varphi^\sharp \, dP_{\Pi_n} - \int_{\mathcal{W}_p(\mathcal{X})} \psi^\sharp \, dP_{\Pi_n^*} \\ &= \int_{\mathcal{X}} \varphi \, dP - \int_{\mathcal{X}} \psi \, dP^* \\ &\geq (1 - \varepsilon)W_p^p(P, P^*). \end{aligned}$$

Letting $\varepsilon \rightarrow 0$, the announced lower bound follows. \square

Proof of Corollary 6.1. Let $(\tilde{\Pi}_n, \tilde{\Pi}_n^*)$ be a coupling between P_{Π_n} and $P_{\Pi_n^*}$. Since $(S(\tilde{\Pi}_n), S(\tilde{\Pi}_n^*))$ is a coupling between P_{S_n} and $P_{S_n^*}$,

$$\begin{aligned} W_p(P_{S_n}, P_{S_n^*}) &\leq \mathbb{E}[\rho(S(\tilde{\Pi}_n), S(\tilde{\Pi}_n^*))^p]^{1/p} \\ &\leq \text{Lip}(S)\mathbb{E}[W_p(\tilde{\Pi}_n, \tilde{\Pi}_n^*)^p]^{1/p}. \end{aligned}$$

Taking the infimum in the right hand side over all couplings between P_{Π_n} and $P_{\Pi_n^*}$, we get

$$W_p(P_{S_n}, P_{S_n^*}) \leq \text{Lip}(S)W_p^{(2)}(P_{\Pi_n}, P_{\Pi_n^*}).$$

Theorem 6.1 finally entails $W_p(P_{S_n}, P_{S_n^*}) \leq \text{Lip}(S)W_p(P, P^*)$. \square

Proof of Proposition 6.1. Proof of point i) : let π_1 and $\pi_2 \in \mathcal{W}_p(\mathcal{X})$ and (X_1, X_2) be a coupling between π_1 and π_2 . By the triangle inequality,

$$\begin{aligned} |S(\pi_1) - S(\pi_2)| &= \left| \|d(x_0, X_1)\|_{L^q} - \|d(x_0, X_2)\|_{L^q} \right| \\ &\leq \|d(x_0, X_1) - d(x_0, X_2)\|_{L^q} \\ &\leq \|d(X_1, X_2)\|_{L^q}. \end{aligned}$$

Taking the infimum in the right hand side, we get

$$|S(\pi_1) - S(\pi_2)| \leq W_q(\pi_1, \pi_2) \leq W_p(\pi_1, \pi_2).$$

Moreover, for $\pi_1 = \delta_{x_0}$ and $\pi_2 = \delta_x$ with $x \neq x_0$, we have $S(\pi_1) = 0$, $S(\pi_2) = d(x_0, x) > 0$ and $W_p(\pi_1, \pi_2) = d(x_0, x)$. Hence $|S(\pi_1) - S(\pi_2)| = W_p(\pi_1, \pi_2)$ and $\text{Lip}(S) = 1$.

Proof of point ii) : Let π_1 and $\pi_2 \in \mathcal{W}_p(\mathcal{X})$. By the triangle inequality for the L^q -norm, we have

$$|S(\pi_1) - S(\pi_2)| \leq \left(\int_0^1 |F_{\pi_1}^{\leftarrow}(u) - F_{\pi_2}^{\leftarrow}(u)|^q (1-u)^s u^r du \right)^{1/q}.$$

For all $u \in [0, 1]$, we have $(1-u)^s u^r \leq 1$. We deduce, using the explicit expression of the Wasserstein distance on the real line (see Example 6.1),

$$\begin{aligned} |S(\pi_1) - S(\pi_2)| &\leq \left(\int_0^1 |F_{\pi_1}^{\leftarrow}(u) - F_{\pi_2}^{\leftarrow}(u)|^q du \right)^{1/q} \\ &\leq W_q(\pi_1, \pi_2) \leq W_p(\pi_1, \pi_2). \end{aligned}$$

□

6.4.2 Proofs related to Section 6.3

Proof of Proposition 6.2. Proof of i). We begin with the proof of Equation (6.14) by giving the expression of the Wasserstein distance on $\mathcal{X} = [1, \infty)$ with logarithmic distance d given by (6.13). Because the logarithm is an isometry between $([1, \infty), d)$ and $([0, \infty), d_2)$ with d_2 the usual Euclidean distance, the Wasserstein distance between probability measures P_1 and P_2 on (\mathcal{X}, d) is equal to the Wasserstein distance between the image measures $P_1 \circ \log^{-1}$ and $P_2 \circ \log^{-1}$ on (\mathbb{R}, d_2) . Using the explicit form of the Wasserstein distance on \mathbb{R} recalled in Example 6.1, we deduce

$$W_p(P_1, P_2) = \left(\int_0^1 |\log F_1^{\leftarrow}(u) - \log F_2^{\leftarrow}(u)|^p du \right)^{1/p},$$

with F_1^{\leftarrow} and F_2^{\leftarrow} the quantile functions of P_1 and P_2 respectively. Introducing the change of variable $z = 1/(1-u)$ and the tail quantile functions $U_i(t) = F_i^{\leftarrow}(1-1/t)$, $i = 1, 2$, we get

$$W_p(P_1, P_2) = \left(\int_1^\infty \left| \log \frac{U_1(z)}{U_2(z)} \right|^p \frac{dz}{z^2} \right)^{1/p}. \quad (6.22)$$

Similarly, in the case $p = \infty$, we obtain

$$W_\infty(P_1, P_2) = \sup_{u \in (0,1)} |\log F_1^{\leftarrow}(u) - \log F_2^{\leftarrow}(u)| = \sup_{z > 1} \left| \log \frac{U_1(z)}{U_2(z)} \right|.$$

Specializing these formulas when $P_1 = P_{U(t)^{-1}X|Z>t}$ and $P_2 = P_\alpha$, we obtain Equation (6.14) because

$$U_1(z) = \frac{U(tz)}{U(t)} \quad \text{and} \quad U_2(z) = z^\gamma.$$

To check the first equality, one can use the fact that, given $Z > t$, Z has the same distribution as tZ so that $X/U(t) = U(Z)/U(t)$ has the same distribution as $U(tZ)/U(t)$.

Proof of ii). We next consider the properties of A_p when $F \in D(G_\gamma)$, or equivalently, under the first order condition (6.10). Since the function U is regularly varying at infinity with order $\gamma > 0$, Potter's bounds [22, Proposition B.1.9 p. 366] imply that, for all $\delta > 0$, there exists $t_1 > 1$ such that

$$(1 - \delta)z^{-\delta} \leq \frac{U(tz)}{z^\gamma U(t)} \leq (1 + \delta)z^\delta \quad \text{for all } t \geq t_1 \text{ and } z > 1. \quad (6.23)$$

This implies that, for $p \in [1, \infty)$, $A_p(t)$ defined by Equation (6.14) is bounded for $t \geq t_1$. On the other hand, the monotonicity of U implies that, for $t \in [t_0, t_1]$, $A_p^p(t) \leq \int_1^\infty \max\left(\left|\log \frac{U(zt_1)}{z^\gamma U(t_0)}\right|^p, \left|\log \frac{U(zt_0)}{z^\gamma U(t_1)}\right|^p\right) \frac{dz}{z^2}$, whence A_p is bounded on $[t_0, \infty)$. Furthermore, the first order condition (6.10) implies

$$\lim_{t \rightarrow \infty} \left| \log \frac{U(zt)}{z^\gamma U(t)} \right|^p = 0 \quad \text{for all } z > 1.$$

Applying the dominated convergence theorem, with the dominating function provided by Potter's bound (6.23), we deduce $\lim_{t \rightarrow \infty} A_p(t) = 0$.

Proof of iii). We finally consider the asymptotic behavior of A_p under the second order condition (6.12). With the auxiliary function $f(z) = \log(U(z)/z^\gamma)$, $z > 1$, the ratio $A_p(t)/A(t)$ can be rewritten as

$$\frac{A_p(t)}{A(t)} = \begin{cases} \left(\int_1^\infty \left| \frac{f(tz) - f(t)}{A(t)} \right|^p \frac{dz}{z^2} \right)^{1/p} & \text{for } 1 \leq p < \infty, \\ \sup_{z > 1} \left| \frac{f(tz) - f(t)}{A(t)} \right| & \text{for } p = \infty. \end{cases} \quad (6.24)$$

From the second order condition (6.12), the function f satisfies, as $t \rightarrow \infty$,

$$\frac{f(tz) - f(t)}{A(t)} = \frac{\log\left(\frac{U(tz)}{z^\gamma U(t)}\right)}{A(t)} \sim \frac{\frac{U(tz)}{z^\gamma U(t)} - 1}{A(t)} \longrightarrow \frac{z^\rho - 1}{\rho}. \quad (6.25)$$

In the terminology of [22, Definition B.2.3], the function f is of extended regular variations at infinity with index $\rho \leq 0$. Equations (6.24) and (6.25) together suggest

$$\frac{A_p(t)}{A(t)} \longrightarrow \begin{cases} \left(\int_1^\infty \left| \frac{z^\rho - 1}{\rho} \right|^p \frac{dz}{z^2} \right)^{1/p} & \text{for } 1 \leq p < \infty, \\ \sup_{z > 1} \left| \frac{z^\rho - 1}{\rho} \right| & \text{for } p = \infty. \end{cases} \quad (6.26)$$

To justify the limits in Equation (6.26), we use Theorem B.2.18 p.383 in [22] : possibly replacing A by an asymptotically equivalent function, we can assume that for arbitrary $\varepsilon, \delta > 0$, there exists t_0 such that

$$\left| \frac{f(tz) - f(t)}{A(t)} - \frac{z^\rho - 1}{\rho} \right| \leq \varepsilon z^{\rho + \delta}, \quad \text{for all } t \geq t_0, z > 1. \quad (6.27)$$

In the case $p \in [1, \infty)$, the limit (6.26) is a consequence of the dominated convergence Theorem where the pointwise convergence is given by Equation (6.25) and the dominating function by Equation (6.27).

In the case $p = \infty$ and $\rho < 0$, the right hand side in Equation (6.27) can be made arbitrary small uniformly in $z > 1$ by choosing ε and δ small enough (note that

$z^{\delta+\rho} < 1$ for $\delta + \rho < 0$ and $z > 1$). This is enough to justify the limit (6.26) and the value of the limit is $1/|\rho|$.

In the case $p = \infty$ and $\rho = 0$, Equation (6.27) implies

$$\frac{f(tz) - f(t)}{A(t)} \geq \log(z) - \varepsilon z^\delta$$

For arbitrary $M > 0$, one can choose $z > 1$ large enough and $\varepsilon, \delta > 0$ small enough so that the right hand side is larger than M for $t \geq t_0$. Then, for $t \geq t_0$,

$$\frac{A_\infty(t)}{A(t)} = \sup_{z>1} \left| \frac{f(tz) - f(t)}{A(t)} \right| \geq M,$$

proving $\lim_{t \rightarrow \infty} A_\infty(t)/A(t) = +\infty$. \square

Proof of Theorem 6.2. Proof of i). For a random variable Z with standard unit Pareto distribution, the conditional distribution of Z given $Z > t$ is equal to the distribution of tZ . We deduce that the conditional distribution of the k top order statistics $(Z_{n-k+1,n}, \dots, Z_{n,n})$ given $Z_{n-k,n} = t$ is equal to the distribution of $(t\tilde{Z}_{1,k}, \dots, t\tilde{Z}_{k,k})$ with $\tilde{Z}_{1,k} \leq \dots \leq \tilde{Z}_{k,k}$ the order statistics of an i.i.d. sample $\tilde{Z}_1, \dots, \tilde{Z}_k$ with standard unit Pareto margins. As a consequence, the conditional distribution of $\Pi_{n,k}$ given $Z_{n-k,n} = t$ is equal to the distribution of

$$\tilde{\Pi}_k = \frac{1}{k} \sum_{i=1}^k \delta_{\tilde{X}_i} \quad \text{with } \tilde{X}_i = U(t\tilde{Z}_i)/U(t).$$

Theorem 6.1 implies

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\tilde{\Pi}_k^*}\right) = W_p(P_{\tilde{X}}, P_\alpha) = A_p(t).$$

Proof of ii). For an intermediate sequence $k = k(n)$, $Z_{n-k,n}$ converges to ∞ in probability as $n \rightarrow \infty$. By Proposition 6.2, A_p has limit 0 at infinity and it follows

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_k^*}\right) = A_p(Z_{n-k,n}) \rightarrow 0 \quad \text{in probability.}$$

Proof of iii). By Proposition 6.2 point *iii*), the second order condition implies $A_p(t) \sim c_p(\rho)A(t)$ as $t \rightarrow \infty$. On the other hand, for an intermediate sequence $k = k(n)$, $Z_{n-k,n} = (n/k)(1 + o_P(1))$ as $n \rightarrow \infty$. Combining the two results, we deduce

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_k^*}\right) = A_p(Z_{n-k,n}) = c_p(\rho)A(n/k)(1 + o_P(1)).$$

\square

Proof of Corollary 6.2. When X is positive and bounded away from 0, the equality $W_p^{(2)}(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\Pi_k^*}) = A_p(t)$ holds for all $t \geq 1$. Integrating with respect to the distribution of $Z_{n-k,n}$ and using the convexity of the Wasserstein distance [83, Theorem 4.8], we deduce

$$W_p^{(2)}\left(P_{\Pi_{n,k}}, P_{\Pi_k^*}\right) \leq \mathbb{E}[A_p(Z_{n-k,n})] = \int_1^\infty A_p(t)\beta_{n-k,k+1} \left(1 - \frac{1}{t}\right) \frac{dt}{t^2}.$$

The last equality holds since $Z_{n-k,n}$ has density $\beta_{n-k,k+1}(1-1/t)/t^2$. For an intermediate sequence $k = k(n)$, $Z_{n-k,n} \rightarrow \infty$ in probability and Proposition 6.2 *ii*) implies that $A_p(Z_{n-k,n}) \rightarrow 0$ in probability. Since $A_p(t)$ is bounded, $\mathbb{E}[A_p(Z_{n-k,n})] \rightarrow 0$ whence the Wasserstein distance converges to 0. \square

Proof of Corollary 6.3. By the triangular inequality,

$$\begin{aligned} & W_p\left(P_{\sqrt{k}(\hat{\gamma}_{n,k}^{Hill}-\gamma)}|Z_{n-k,n}=t, \mathcal{N}(0, \gamma^2)\right) \\ & \leq W_p\left(P_{\sqrt{k}(\hat{\gamma}_{n,k}-\gamma)}|Z_{n-k,n}=t, P_{\sqrt{k}(\hat{\gamma}_k^*-\gamma)}\right) + W_p\left(P_{\sqrt{k}(\hat{\gamma}_k^*-\gamma)}, \mathcal{N}(0, \gamma^2)\right). \end{aligned}$$

The first term is estimated by Corollary 6.1 with $\hat{\gamma}_{n,k} = S(\Pi_{n,k})$, $\hat{\gamma}_k^* = S(\Pi_k^*)$ and $S : \mathcal{W}_p([1, \infty)) \rightarrow \mathbb{R}$ given by $\varphi(\pi) = \int_1^{+\infty} \log(x)\pi(dx)$. Proposition 6.1 with $x_0 = 1$ and $d(x, x_0) = \log(x)$ implies that S is Lipschitz with $\text{Lip}(S) = 1$ and together with Theorem 6.2 this yields

$$W_p\left(P_{\hat{\gamma}_{n,k}|Z_{n-k,n}=t}, P_{\hat{\gamma}_k^*}\right) \leq \text{Lip}(S)W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\Pi_k^*}\right) = A_p(t).$$

A change of variable entails

$$W_p\left(P_{\sqrt{k}(\hat{\gamma}_{n,k}-\gamma)}|Z_{n-k,n}=t}, P_{\sqrt{k}(\hat{\gamma}_k^*-\gamma)}\right) \leq \sqrt{k}A_p(t).$$

The second term is handled using the central limit theorem in Wasserstein distance derived from Stein's method, see [77]. We write

$$\sqrt{k}(\hat{\gamma}_k^* - \gamma) = \frac{\gamma}{\sqrt{k}} \sum_{i=1}^k \tilde{X}_i \quad \text{with } \tilde{X}_i = (\log(X_i^*) - \gamma)/\gamma.$$

An application of [77, Theorem 3.2] yields

$$\begin{aligned} W_p\left(P_{k^{-1/2} \sum_{i=1}^k \tilde{X}_i}, \mathcal{N}(0, 1)\right) & \leq k^{-3/2} \sum_{i=1}^k \mathbb{E}|\tilde{X}_i|^3 + \sqrt{2/\pi}k^{-1} \left(\sum_{i=1}^k \mathbb{E}\tilde{X}_i^4\right)^{1/2} \\ & = (4 + 3\sqrt{2/\pi})\frac{1}{\sqrt{k}}, \end{aligned}$$

where the constants $\mathbb{E}|\tilde{X}_i|^3 = 4$ and $\mathbb{E}\tilde{X}_i^4 = 9$ are easily computed since \tilde{X}_i is a centered standard exponential random variable. With a change of variable, we deduce

$$W_p\left(P_{\sqrt{k}(\hat{\gamma}_k^*-\gamma)}, \mathcal{N}(0, \gamma^2)\right) \leq (4 + 3\sqrt{2/\pi})\frac{\gamma}{\sqrt{k}}. \quad (6.28)$$

\square

Proof of Corollary 6.4. We write

$$v_n^{-1} \log \frac{\hat{q}(\alpha_n)}{q(\alpha_n)} = I + II + III$$

with

$$\begin{aligned} I &= \sqrt{k}(\hat{\gamma}_{n,k} - \gamma), \\ II &= v_n^{-1} \log \frac{X_{n-k,n}}{U(n/k)}, \\ III &= v_n^{-1} \log \frac{(k/(n\alpha_n))^\gamma U(n/k)}{U(1/\alpha_n)}. \end{aligned}$$

The first term does not depend on $Z_{n-k,n}$ while the second and third terms are measurable with respect to $Z_{n-k,n}$ and can be treated as constant, yielding

$$W_p \left(P_{v_n^{-1} \log \frac{\hat{q}(\alpha_n)}{q(\alpha_n)} | Z_{n-k}}, \mathcal{N}(0, \gamma^2) \right) \leq W_p \left(P_{\sqrt{k}(\hat{\gamma}_k^* - \gamma)}, \mathcal{N}(0, \gamma^2) \right) + II + III.$$

The asymptotic behavior of the first term is given by Corollary 6.3 together with Proposition 6.2 *iii*) :

$$W_p \left(P_{\sqrt{k}(\hat{\gamma}_{n,k} - \gamma) | Z_{n-k,n}}, \mathcal{N}(0, \gamma^2) \right) = O_P \left(\sqrt{k}A(n/k) + 1/\sqrt{k} \right).$$

The second term can be handled seeing that $X_{n-k,n} = U(Z_{n-k,n})$ with $Z_{n-k,n} = \frac{n}{k}(1 + o_P(1))$. This implies

$$\begin{aligned} II &= v_n^{-1} \log \frac{U(Z_{n-k,n})}{U(n/k)} \\ &= v_n^{-1} \left(\frac{U(Z_{n-k,n})}{U(n/k)} - 1 \right) (1 + o_P(1)) \\ &= v_n^{-1} \left((Z_{n-k,n}k/n)^\gamma O_P(A(n/k)) \right) (1 + o_P(1)) \\ &= o_P \left(\sqrt{k}A(n/k) \right), \end{aligned}$$

where the third line relies on the second order condition (6.12).

The third term is deterministic and satisfies

$$\begin{aligned} III &= v_n^{-1} \left(\frac{(k/(n\alpha_n))^\gamma U(k/n)}{U(1/\alpha_n)} - 1 \right) (1 + o(1)) \\ &= v_n^{-1} \left(A(n/k) \frac{(k/(n\alpha_n))^\rho - 1}{\rho} \right) (1 + o(1)) \\ &= o_P \left(\sqrt{k}A(n/k) \right), \end{aligned}$$

where the second equality relies on [22, Theorem 2.3.9.] and on the fact that $k/(n\alpha_n) \rightarrow \infty$ and $\rho < 0$. \square

Proof of Corollary 6.5 . First note that $\hat{M}_{k,n,s}$ and $\hat{M}_{k,s}^*$ for $s = 0, 1$ are the weighted probability moments associated with the empirical measures $\Pi_{n,k}$ and Π_k^* respectively. According to Proposition 6.1 *ii*), the weighted probability moment functional S (with $q = 1$, $r = 0$ and $s = 0, 1$) is 1-Lipschitz so that Corollary 6.1 implies

$$W_p \left(P_{(\hat{M}_{k,n,s})_{s=0,1} | Z_{n-k,n}=t}, P_{(\hat{M}_{k,s}^*)_{s=0,1}} \right) \leq W_p^{(2)} \left(P_{\Pi_{n,k} | Z_{n-k,n}=t}, P_{\Pi_k^*} \right).$$

Under the assumption $\sqrt{k}A(n/k) \rightarrow 0$, Theorem 6.4 ii) entails

$$W_p \left(P_{\sqrt{k}(\hat{M}_{k,n,s}-m_s)_{s=0,1}|Z_{n-k,n}}, P_{\sqrt{k}(\hat{M}_{k,s}^*-m_s)_{s=0,1}} \right) = O_P \left(\sqrt{k}A(n/k) \right) \rightarrow 0.$$

[54, Equation 5.3] states the joint asymptotic normality of $\hat{M}_{k,0}^*$ and $\hat{M}_{k,1}^*$:

$$\sqrt{k} \left(\hat{M}_{k,0}^* - m_0, \hat{M}_{k,1}^* - m_1 \right) \xrightarrow{d} \mathcal{N}(0, \Gamma) \quad \text{as } k \rightarrow \infty.$$

Then Lemma 1 in Supplement A implies the joint asymptotic normality of $\hat{M}_{k,n,0}$ and $\hat{M}_{k,n,1}$:

$$\sqrt{k} \left(\hat{M}_{k,n,0} - m_0, \hat{M}_{k,n,1} - m_1 \right) \xrightarrow{d} \mathcal{N}(0, \Gamma) \quad \text{as } k \rightarrow \infty.$$

In view of Equation (6.15), the asymptotic normality of $(\hat{\gamma}_{n,k}^{PWM}, \hat{\sigma}_{n,k}^{PWM})$ follows by the delta method, see e.g. van der Vaart [81], Theorem 3.1 p.26). Since g is infinitely differentiable on $\mathbb{R}^2 \setminus \{(x, y), x = 2y\}$ and $(\gamma, 1)$ does belong to this set, we can apply the delta method and deduce

$$\begin{aligned} & \sqrt{k} \left((\hat{\gamma}_{n,k}^{PWM}, \hat{\sigma}_{n,k}^{PWM}) - (\gamma, 1) \right) \\ &= \sqrt{k} \left(g(\hat{M}_{k,n,0}, \hat{M}_{k,n,1}) - g(m_0, m_1) \right) \xrightarrow{d} dg_{(\gamma,1)}(N), \end{aligned}$$

with $dg_{(\gamma,1)}$ the derivative of f at $(\gamma, 1)$ and $N \sim \mathcal{N}(0, \Sigma)$. We conclude the proof by noting that $dg_{(\gamma,1)}(N) \stackrel{d}{=} \mathcal{N}(0, \Gamma)$. \square

Acknowledgements

The current manuscript benefited a lot from constructive comments and fruitful suggestions after presentations of the work at an early stage of the project. In particular, Laurens de Haan suggested Proposition 6.2 *iii*) and its proof, Ana Ferreira suggested to consider the asymptotic regime with bias $\sqrt{k}A(n/k) \rightarrow \lambda$ (section 3.3.1), Chen Zhou suggested to consider the block maxima method (section 3.3.3) and Stéphane Girard suggested to consider Weissman extreme quantile estimate (Corollary 6.4). We gratefully acknowledge their contributions to the paper.

6.5 Supplementary material

Proof of Theorem 6.3. Similar arguments as in the proof of Theorem 6.2 imply

$$W_p^{(2)} \left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\Pi_{k,t}^*} \right) = W_p \left(P_{U(t)^{-1}X|Z>t}, P_{X^*(1+A(t)\frac{(X^*)^\rho/\gamma-1}{\rho})} \right)$$

where $X^* \sim P_\alpha$. Introducing a random variable Z with standard unit Pareto distribution and the coupling

$$\frac{U(tZ)}{U(t)} \sim P_{U(t)^{-1}X|Z>t} \quad \text{and} \quad Z^\gamma \left(1 + A(t) \frac{Z^\rho - 1}{\rho} \right) \sim P_{X^*(1+A(t)\frac{(X^*)^\rho/\gamma-1}{\rho})},$$

we deduce, similarly as in Equation (6.22),

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\Pi_{k,t}^*}\right) \leq \left(\int_1^\infty \left|\log \frac{U(zt)}{z^\gamma U(t)} - \log\left(1 + A(t)\frac{z^\rho - 1}{\rho}\right)\right|^p \frac{dz}{z^2}\right)^{1/p}.$$

Equation (6.27) with $f(z) = U(z)/z^\gamma$ implies

$$\left|\log \frac{U(zt)}{z^\gamma U(t)} - A(t)\frac{z^\rho - 1}{\rho}\right| \leq \varepsilon |A(t)|z^{\rho+\delta},$$

and, with $x - 2x^2 \leq \log(1+x) \leq x$ for $x \geq -1/2$, we deduce

$$\left|\log \frac{U(zt)}{z^\gamma U(t)} - \log\left(1 + A(t)\frac{z^\rho - 1}{\rho}\right)\right| \leq \varepsilon |A(t)|z^{\rho+\delta} + 2A(t)^2 \left|\frac{z^\rho - 1}{\rho}\right|^2.$$

Taking the L^p -norm and $\varepsilon > 0$ being arbitrary, we see that

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\Pi_{k,t}^*}\right) = o(A(t)) \quad \text{as } t \rightarrow \infty.$$

With the same notations as in the proof of Corollary 6.3, we compare

$$\hat{\gamma}_{n,k} = \int_1^\infty \log(x)\Pi_{n,k}(dx) \quad \text{and} \quad \hat{\gamma}_{k,t}^* = \int_1^\infty \log(x)\Pi_{k,t}^*(dx).$$

By Corollary 6.1,

$$W_p\left(P_{\sqrt{k}(\hat{\gamma}_{n,k}-\gamma)|Z_{n-k,n}}, P_{\sqrt{k}(\hat{\gamma}_{k,t}^*-\gamma)}\right) \leq \sqrt{k}W_p^{(2)}(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_{k,t}^*}).$$

On the other hand,

$$\sqrt{k}(\hat{\gamma}_{k,t}^* - \gamma) = \sqrt{k}(\hat{\gamma}_k^* - \gamma) + \frac{1}{\sqrt{k}} \sum_{i=1}^k \log\left(1 + A(t)\frac{Z_i^\rho - 1}{\rho}\right).$$

The first term is compared to $\mathcal{N}(0, \gamma^2)$ thanks to Equation (6.28) :

$$W_p\left(P_{\sqrt{k}(\hat{\gamma}_k^* - \gamma)}, \mathcal{N}(0, \gamma^2)\right) \leq \frac{4 + 3\sqrt{2/\pi}}{\sqrt{k}}.$$

The second term is compared to $\sqrt{k}A(t)b(\rho)$. Using the triangle inequality and $x - 2x^2 \leq \log(1+x) \leq x$ for $x \geq -1/2$, we get

$$\begin{aligned} & \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \log\left(1 + A(t)\frac{Z_i^\rho - 1}{\rho}\right) - \sqrt{k}A(t)b(\rho) \right\|_{L^p} \\ & \leq \sqrt{k}A(t) \left\| \frac{1}{k} \sum_{i=1}^k \left(\frac{Z_i^\rho - 1}{\rho} - b(\rho)\right) \right\|_{L^p} + \sqrt{k}A(t)^2 \left\| \frac{1}{k} \sum_{i=1}^k \left|\frac{Z_i^\rho - 1}{\rho}\right|^2 \right\|_{L^p} \\ & = o(\sqrt{k}A(t)). \end{aligned}$$

The last line is a consequence of the law of large numbers in L^p norm. Combining the different estimates, we get

$$W_p\left(P_{\sqrt{k}(\hat{\gamma}_{n,k}-\gamma)|Z_{n-k,n}=t}, \mathcal{N}(\sqrt{k}A(t)b(\rho), \gamma^2)\right) = O(1/\sqrt{k}) + o(\sqrt{k}A(t))$$

and the asymptotic normality of the Hill estimator with bias $\lambda b(\rho)$ follows if $k \rightarrow \infty$ and $\sqrt{k}A(n/k) \rightarrow \lambda$. □

Proof of Theorem 6.4. Proof of i). The structure of the proof is the same as the proof of Proposition 6.2. The tail quantile function of $P_{(X-U(t))/a(t)|T>t}$ is given by $(U(tz) - U(t))/a(t)$, $z > 1$ and the tail quantile function of the GP distribution by $(z^\gamma - 1)/\gamma$. The explicit expression of the Wasserstein distance given in Example 6.1 yields

$$W_p\left(P_{(X-U(t))/a(t)|T>t}, H_\gamma\right) = A'_p(t).$$

The convergence of A'_p to 0 is a consequence of the dominated convergence Theorem with the domination condition provided by Drees's Theorem (Theorem B.2.18 p.383 in [22]) : for all $\varepsilon, \delta > 0$, there exists $t' > 0$ such that

$$\left|\frac{U(tz) - U(t)}{a(t)} - \frac{z^\gamma - 1}{\gamma}\right| \leq \varepsilon z^{\gamma+\delta}, \quad z > 1, t > t',$$

where $z^{(\gamma+\delta)p}/z^2$ is integrable on $[1, +\infty)$ for $\gamma p < 1$ and δ small enough.

Under the second order condition (6.8),

$$\frac{A'_p(t)}{A(t)} = \left(\int_1^\infty \left|\frac{\frac{U(zt)-U(t)}{a(t)} - \frac{z^\gamma-1}{\gamma}}{A(t)}\right|^p \frac{dz}{z^2}\right)^{1/p} \longrightarrow \left(\int_1^\infty \Psi_{\gamma,\rho}(z)^p z^{-2} dz\right)^{1/p}.$$

The convergence is proved using the dominated convergence theorem with dominating function provided by [22], Theorem B.3.10 p.392 : for $\varepsilon, \delta > 0$, there exists $t' > 0$ such that

$$\left|\frac{\frac{U(zt)-U(t)}{a(t)} - \frac{z^\gamma-1}{\gamma}}{A(t)} - \Psi_{\gamma,\rho}(z)\right| \leq \varepsilon z^{\gamma+\rho+\delta}, \quad x > 1, t > t',$$

and the function $z^{(\gamma+\rho+\delta)p}/z^2$ is integrable on $[1, +\infty)$ for $(\gamma + \rho)p < 1$ and δ small enough.

Proof of ii). The conditional distribution of the k top order statistics $(Z_{n-k+1,n}, \dots, Z_{n,n})$ given $Z_{n-k,n} = t$ is equal to the distribution of $(t\tilde{Z}_{1,k}, \dots, t\tilde{Z}_{k,k})$ with $\tilde{Z}_{1,k} \leq \dots \leq \tilde{Z}_{k,k}$ the order statistics of an i.i.d. sample $\tilde{Z}_1, \dots, \tilde{Z}_k$ with standard unit Pareto margins. We deduce that $P_{\Pi_{n,k}|Z_{n-k,n}=t}$ is equal to the distribution of

$$\tilde{\Pi}_k = \frac{1}{k} \sum_{i=1}^k \delta_{\tilde{X}_i} \quad \text{with } \tilde{X}_i = (U(t\tilde{Z}_i) - U(t))/a(t).$$

Theorem 6.1 implies

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}=t}, P_{\tilde{\Pi}_k^*}\right) \leq W_p(P_{\tilde{X}}, H_\gamma) = A'_p(t)$$

because $P_{\tilde{X}} = P_{(X-U(t))/a(t)|T>t}$.

Under the second order condition (6.8), the first point of this theorem yields $A'_p(t) \sim c'_p(\gamma, \rho)A(t)$. Since $Z_{n-k,n} = n/k(1 + o_P(1))$ and A'_p is regularly varying, we deduce

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Z_{n-k,n}}, P_{\Pi_k^*}\right) = A'_p(Z_{n-k,n}) = c'_p(\gamma, \rho)A(n/k)(1 + o_P(1)).$$

□

Proof of Theorem 6.5. The proof is similar to the proof of Proposition 6.2.

Proof of i). The explicit form of the Wasserstein distance on \mathbb{R} recalled in Example 6.1 implies

$$W_p(P_1, P_2) = \left(\int_0^1 |F_1^{\leftarrow}(u) - F_2^{\leftarrow}(u)|^p \, du \right)^{1/p}$$

for all probability measures P_1, P_2 on \mathbb{R} with quantile functions $F_1^{\leftarrow}, F_2^{\leftarrow}$ respectively. Introducing the change of variable $z = -1/\log(u)$ and the functions $V_i(t) = F_i^{\leftarrow}(e^{-1/t})$, $i = 1, 2$, we get

$$W_p(P_1, P_2) = \left(\int_0^1 |V_1(z) - V_2(z)|^p e^{-1/z} \frac{dz}{z^2} \right)^{1/p}.$$

When $P_1 = P_{(M_1 - b_m)/a_m}$ is the distribution of the normalized block maxima and $P_2 = G_\gamma$ is the GEV distribution, we get $V_1(z) = (V(mz) - V(m))/a(m)$ and $V_2(m) = (z^\gamma - 1)/\gamma$, yielding

$$\begin{aligned} W_p(P_{(M_1 - b_m)/a_m}, G_\gamma) &= \left(\int_0^\infty \left| \frac{V(mz) - V(m)}{a(m)} - \frac{z^\gamma - 1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \right)^{1/p} \\ &= A_p''(m), \quad m \geq 1. \end{aligned}$$

Theorem 6.1 then implies, in the second order Wasserstein space $\mathcal{W}_p^{(2)}([0, \infty))$,

$$W_p^{(2)}(P_{\Pi_{n,m}}, P_{\Pi_k^*}) = A_p''(m), \quad 1 \leq m \leq n.$$

Proof of ii). We prove that if $F \in \mathcal{D}(G_\gamma)$ with $\gamma < 1$ and $p \in [1, 1/\gamma_+)$, then $A_p''(m) \rightarrow 0$ as $m \rightarrow \infty$. By Drees Theorem (see e.g. [22], Theorem B.2.18 p.383), the first order condition implies that for all $\varepsilon, \delta > 0$, there exists m_0 such that

$$\left| \frac{V(mz) - V(m)}{a(m)} - \frac{z^\gamma - 1}{\gamma} \right| \leq \varepsilon \max(z^{\gamma+\delta}, z^{\gamma-\delta}), \quad m > m_0, mz > m_0.$$

We deduce that

$$\begin{aligned} & \int_{m_0/m}^\infty \left| \frac{V(mz) - V(m)}{a(m)} - \frac{z^\gamma - 1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \\ & \leq \varepsilon^p \int_0^\infty \max(z^{(\gamma+\delta)p}, z^{(\gamma-\delta)p}) e^{-1/z} \frac{dz}{z^2}, \end{aligned}$$

where the last integral is finite for δ small enough because $\gamma p < 1$. On the other hand, the contribution of the interval $[0, m/m_0]$ to the integral defining $A_p''(m)$ is upper bounded as follows. We use the assumption that F has a finite left end point so that V is bounded by M on $[0, 1]$, whence

$$\begin{aligned} & \int_0^{m_0/m} \left| \frac{V(mz) - V(m)}{a(m)} - \frac{z^\gamma - 1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \\ & \leq 2^{p-1} \int_0^{m_0/m} \left(\frac{2M}{a(m)} \right)^p e^{-1/z} \frac{dz}{z^2} + 2^{p-1} \int_0^{m_0/m} \left| \frac{z^\gamma - 1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \\ & = \frac{2^{p-1}(2M)^p}{a(m)^p} e^{-m/m_0} + 2^{p-1} \int_0^{m_0/m} \left| \frac{z^\gamma - 1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2}. \end{aligned}$$

Combining these estimates, we get

$$A_p''(m)^p = O(\varepsilon^p) + O(a(m)^{-p}e^{-m/m_0}) + o(1).$$

Since a is regularly varying with index γ , $a(m)^pe^{-m/m_0} \rightarrow 0$ as $m \rightarrow \infty$. Letting $m \rightarrow \infty$ and $\varepsilon \rightarrow 0$, we obtain the convergence $A_p''(m) \rightarrow 0$.

Proof of iii). We have

$$\left(\frac{A_p''(m)}{A(m)}\right)^p = \int_0^\infty \left| \frac{\frac{V(mz)-V(m)}{a(m)} - \frac{z^\gamma-1}{\gamma}}{A(m)} \right|^p e^{-1/z} \frac{dz}{z^2},$$

and the second order condition (6.16) suggests the limit

$$\left(\frac{A_p''(m)}{A(m)}\right)^p \longrightarrow \int_0^\infty \Psi_{\gamma,\rho}(z)^p e^{-1/z} \frac{dz}{z^2} = c_p''(\gamma, \rho)^p.$$

The limit is justified as follows. Drees theorem (see [22], Theorem B.3.10 p.392) implies that for $\varepsilon, \delta > 0$, there is $m_0 \geq 1$ such that

$$\begin{aligned} & \left| \left(\int_{m_0/m}^\infty \left| \frac{V(mz)-V(m)}{a(m)} - \frac{z^\gamma-1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \right)^{1/p} - \left(\int_{m_0/m}^\infty \Psi_{\gamma,\rho}(z)^p e^{-1/z} \frac{dz}{z^2} \right)^{1/p} \right| \\ & \leq \varepsilon \left(\int_0^\infty \max(z^{p(\gamma+\rho+\delta)}, z^{p(\gamma+\rho-\delta)}) e^{-1/z} \frac{dz}{z^2} \right)^{1/p} \end{aligned}$$

where the last integral is finite for δ small enough because $\gamma p < 1$ and $\rho \leq 0$. On the other hand, the contribution of the interval $[0, m_0/m]$ to the integral vanishes as $m \rightarrow \infty$ because

$$\begin{aligned} & \int_0^{m_0/m} \left| \frac{V(mz)-V(m)}{a(m)} - \frac{z^\gamma-1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \\ & \leq 2^{p-1} \left(\frac{2M}{a(m)A(m)} \right)^p e^{-m/m_0} + \frac{2^{p-1}}{A(m)^p} \int_0^{m_0/m} \left| \frac{z^\gamma-1}{\gamma} \right|^p e^{-1/z} \frac{dz}{z^2} \rightarrow 0. \end{aligned}$$

We deduce that $A_p''(m) \sim c_p''(\gamma, \rho)A(m)$ as $m \rightarrow \infty$. \square

Lemme 6.1. *Let (\mathcal{X}, d) a polish space. Let $(P_n)_{n \in \mathbb{N}}$ and $(P_n^*)_{n \in \mathbb{N}}$ be two sequences in $\mathcal{W}(\mathcal{X})$ and $P \in \mathcal{W}(\mathcal{X})$. If P_n^* converge weakly to P and $W_p(P_n, P_n^*) \rightarrow 0$ then P_n converge weakly to P .*

Démonstration. Let f be a real valued bounded Lipschitz function. Let $\varepsilon > 0$ and consider (X_n, X_n^*) a coupling between P_n and P_n^* such that

$$\|d(X_n, X_n^*)\|_{L^p} \leq (1 + \varepsilon)W_p(P_n, P_n^*).$$

Since $W_p(P_n, P_n^*) \rightarrow 0$ as $n \rightarrow \infty$, $\|d(X_n, X_n^*)\|_{L^p} \rightarrow 0$ and also $\mathbb{P}[d(X_n, X_n^*)] \rightarrow 0$. As a consequence, writing $\text{Lip}(f)$ for the Lipschitz constant of f ,

$$\mathbb{E}[|f(X_n) - f(X_n^*)|] \leq \text{Lip}(f)\mathbb{E}[d(X_n, X_n^*)]. \quad (6.29)$$

Since $P_n^* \xrightarrow{d} P$, we have $\mathbb{E}f(X_n^*) \rightarrow \mathbb{E}f(X)$ where $X \sim P$ and Equation (6.29) implies $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$. Since f is an arbitrary bounded Lipschitz function, we conclude by the Portmanteau Theorem that X_n converges weakly to X as $n \rightarrow \infty$, that is $P_n \xrightarrow{d} P$. \square

Analysis of proportional tail model with optimal coupling.

Ce chapitre correspond à un article en cours d'élaboration, il porte sur une généralisation du modèle "*heteroscedastic extremes*" détaillé dans le premier chapitre, afin de traiter le cadre de la regression quantile extrême. Contrairement au chapitre 4, nous développons une approche par couplage optimal moins centrée sur les processus empiriques pour traiter le problème. Dans cette approche, il est nécessaire d'apporter quelques légères modifications à la définition du modèle des queues proportionnelles, nous représenterons donc le modèle pour mettre en valeur ces changements. Nous présentons également une méthode de validation par bootstrap. Cette méthode de validation repose sur un résultat dont la preuve est en cours de finalisation. En effet nous avons décelé une faille dans la démonstration quelques jours avant l'envoi du présent manuscrit. C'est pourquoi ce resultat est, dans ce chapitre, présenté sous la forme d'une conjecture.

7.1 Analysis of a proportional tail model

7.1.1 The proportional tail model

We consider i.i.d. copies $(X_i, Y_i)_{1 \leq i \leq n}$ of (X, Y) , with real-valued response $Y \in \mathbb{R}$ and \mathbb{R}^d -valued predictor X . Following [35] the proportional tail model consists in two main assumptions. The first is about the conditional cumulative distribution functions

$$F_x(y) := \mathbb{P}(Y \leq y | X = x), \quad y \in \mathbb{R}, \quad x \in \mathbb{R}^d.$$

Assumption 1. The conditional cumulative distributions functions F_x and the unconditional distribution F of Y satisfy

$$\lim_{y \rightarrow y^*} \frac{1 - F_x(y)}{1 - F(y)} := \sigma(x) \text{ uniformly in } x \in \mathbb{R}^d, \quad (7.1)$$

where σ is the so-called skedasis function, and y^* denote the upper endpoint of F .

Moreover, we assume that there exists ε_σ such that

$$0 < \varepsilon_\sigma \leq \sigma(x) \leq \varepsilon_\sigma^{-1}, \quad x \in \mathbb{R}^d. \quad (7.2)$$

The second assumption, is made upon the tail distribution of the response Y .

Assumption 2. F belongs to the domain of attraction of some extreme value distribution G_γ with $\gamma > 0$.

Under this assumption, F is heavy tailed with index $\alpha = \gamma^{-1} > 0$. This assumption can be characterized by the extreme value condition.

$$\lim_{t \rightarrow +\infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad x > 0, \quad (7.3)$$

with U the tail quantile function defined as $U(t) = F^{\leftarrow}(1 - 1/t)$, $t > 1$, and $F^{\leftarrow}(p) = \inf\{z \in \mathbb{R} : F(z) \geq p\}$, $0 \leq p \leq 1$, the quantile function.

Remarque 7.1. *Because F belongs to the domain of attraction of G_γ with a positive γ , the upper endpoint y^* is infinite and the condition 7.1 implies that $F_x \in D(G_\gamma)$ for all $x \in \mathbb{R}^d$. The extreme value index γ does not depend on the predictor value $X = x$.*

7.1.2 Parameter estimation

We remind the estimators of the extreme value index and of the integrated skedasis function defined in chapter 5. The integrated skedasis function C defined as

$$C(x) = \int_{\{u \leq x\}} \sigma(u) du,$$

is estimated by

$$\widehat{C}_n(x) := \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{Y_i > Y_{n-k:n}, X_i \leq x\}} = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{X_{(n+1-i)} \leq x\}}, \quad x \in \mathbb{R}^d$$

with $X_{(i)}$ denoting the concomitant order statistics, that is $X_{(i)}$ is the covariate which correspond to $Y_{i:n}$, and the notation $u \leq x$ denotes the componentwise comparison of vectors. The classical Hill estimator of the extreme value index γ is defined as

$$\widehat{\gamma}_n = \frac{1}{k} \sum_{i=1}^k \log \left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}} \right).$$

In these notations, the dependency on k is implicit, since we will set $k := k(n)$ an intermediary sequence depending on n , namely : $k(n) \xrightarrow[n \rightarrow \infty]{} \infty$ and $\frac{k(n)}{n} \xrightarrow[n \rightarrow \infty]{} 0$.

The understanding of the asymptotic behaviors of \widehat{C}_n and $\widehat{\gamma}_n$ requires a more detailed description for the rates of convergence in equations (7.1) and (7.3). We define two non increasing functions A_{prop} and A_{ext} which are the rates of convergences. Write A_{prop} for

$$A_{prop}(y) = \sup_{x \in \mathbb{R}^d} \sup_{y' > y} \left| \frac{\bar{F}_x(y')}{\sigma(x) \bar{F}(y')} - 1 \right|, \quad y \geq 1 \quad (7.4)$$

with $\bar{F} := 1 - F$ and $\bar{F}_x := 1 - F_x$ the tail functions. Assertion (7.1) is equivalent to $\lim_{y \rightarrow \infty} A_{prop}(y) \rightarrow 0$. In extreme value theory, rates of convergence are usually formulated in the form of second order assumption : we assume that U satisfies

$$\lim_{y \rightarrow +\infty} \frac{\frac{U(zy)}{U(y)} - z^\gamma}{A_{ext}(y)} = z^\gamma \frac{z^\rho - 1}{\rho}, \quad (7.5)$$

with $\rho < 0$ the so called extreme value second order parameter. We moreover consider the function $A := \max(A_{ext}, A_{prop})$ which will lead rates of convergence in the model. We are interested in both consistency and asymptotic normality of \hat{C}_n and $\hat{\gamma}_n$. The C -Brownian bridge B is defined as the centered Gaussian process with covariance

$$\text{cov}(B(x_1), B(x_2)) = C(x_1 \wedge x_2) - C(x_1)C(x_2), \quad x_1, x_2 \in \mathbb{R}^d.$$

Theorem 7.1. *Under assumptions (7.1) and (7.5), for a sequence $k = k(n)$ satisfying $k \rightarrow +\infty$, $\frac{k}{n} \rightarrow 0$ and $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow 0$, we have*

$$\begin{pmatrix} \sqrt{k}(\hat{C}_n(\cdot) - C(\cdot)) \\ \sqrt{k}(\hat{\gamma}_n - \gamma) \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{d} \begin{pmatrix} B(\cdot) \\ N \end{pmatrix} \quad (7.6)$$

where $N \sim \mathcal{N}(0, \gamma^2)$, B is a C -Brownian bridge and N and B are independent.

Addendum 7.2. If we replace the condition $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow 0$ in the previous theorem by $\sqrt{k}A_{ext}\left(\frac{n}{k}\right) \rightarrow \lambda \in \mathbb{R}$ and $\sqrt{k}A_{prop}\left(\frac{n}{k}\right) \rightarrow 0$ then the convergence (7.6) hold with $N \sim \mathcal{N}(\lambda b(\rho), \gamma^2)$ with $b(\rho) = \int_0^1 \frac{z^\rho - 1}{\rho} \frac{dz}{z^2} \times \int_{\mathbb{R}^p} \sigma(x)^{-\rho} P_X(dx)$. Moreover, B and N remain independent.

Remarque 7.2. *This theorem extends Theorem 2.1 in [35], since they establish a similar result with $p = 1$, $\lambda = 0$ and $P_X = \mathcal{U}([0, 1])$.*

7.1.3 Model validation

We address here the question of validation of the proportional tail model. Since Equation (7.1) together with (7.3) imply that γ does not depend on the covariate X , it is relevant (and easier) to test whether the extreme value index γ is constant in x . In that sense, we consider the difference between several Hill estimators build on sub-samples, and the global one. We define, for every $u \in \mathbb{R}^d$,

$$\Delta_n(u) := \frac{1}{k} \sum_{i=1}^k \left(\log \left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}} \right) - \hat{\gamma}_n \right) \mathbf{1}_{\{X_{(n+1-i)} \leq u\}}.$$

Conjecture 7.1. *Under assumptions (7.1) and (7.5), for a sequence $k = k(n)$ satisfying $k \rightarrow +\infty$, $\frac{k}{n} \rightarrow 0$ and $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow 0$, we have*

$$\sqrt{k} \frac{\Delta_n(\cdot)}{\hat{\gamma}_n} \xrightarrow[k \rightarrow \infty]{d} B(\cdot), \quad \text{in } L^\infty(\mathbb{R}^d).$$

When the covariate X is real valued, the distribution of $\|B\|_{\mathbb{R}^d}$ is stochastically bounded by the Kolmogorov-Smirnov distribution. We can use the Kolmogorov-Smirnov test to check whether γ is constant over \mathbb{R}^d . On the opposite, for $d \geq 2$, the supremum of the C -Brownian bridge depends on the unknown function C and does not have a simple form. For this reason, we use a bootstrap procedure to estimate the distribution of B and provide p -values for our validation procedure.

More precisely, let $(\xi_{n,i})_{1 \leq i \leq k}$ be exchangeable non negative random variables such that $\mathbb{E}(\xi_{n,i}) = 1$ for all $i = 1, \dots, k$. We moreover assume that $\sum_{i=1}^k \xi_{n,i} = k$ with probability one.

Example 7.1. *We consider mainly the two following cases*

- *The Efron's bootstrap [34] is the case when the weights are multinomially distributed with parameters k and $(1/k, \dots, 1/k)$. This corresponds to the case of a bootstrap sample built by drawing with replacement k observations in the sample $(X_{(n+1-i)}, Y_{n+1-i:n})_{1 \leq i \leq k}$ and $(\xi_{n,i})_{1 \leq i \leq k}$ is the number of times where $(X_{(n+1-i)}, Y_{n+1-i:n})$ is redrawn.*
- *The Bayesian bootstrap : let $(\varepsilon_{n,i})_{i=1}^k$ be i.i.d exponential variables with parameter 1, independent from the data, and consider define*

$$\xi_{n,i} := \frac{\varepsilon_{n,i}}{\frac{1}{k} \sum_{j=1}^k \varepsilon_{n,j}}, \quad 1 \leq i \leq k.$$

We define the bootstrap estimators as

$$\hat{C}_{\xi,n}(x) := \frac{1}{k} \sum_{i=1}^k \xi_{n,i} \mathbb{1}_{\{X_{(n+1-i)} \leq x\}}, \quad x \in \mathbb{R}^d,$$

with $X_{(i)}$ denoting the concomitant order statistics, and

$$\hat{\gamma}_{\xi,n} := \frac{1}{k} \sum_{i=1}^k \xi_{n,i} \log \left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}} \right).$$

The asymptotic normality of those bootstrap estimators is formulated in terms of conditional weak convergence. Roughly speaking, this conditional weak convergence is the convergence in distribution of $(\hat{C}_{\xi,n}, \hat{\gamma}_{\xi,n})$ conditional on the data and with randomness induced by the bootstrap weights.

Following the notation of [82], we said that a process $\mathbb{G}_{\xi,n}$ (depending on $(\xi_{n,i})_{i=1}^k$ and on the data $(X_{(n-i+1)}, Y_{n-i+1:n}/Y_{n-k:n})_{i=1}^k$) converge weakly conditionally to ξ to a tight limit \mathbb{G} , noted $\mathbb{G}_{\xi,n} \xrightarrow[\xi]{} \mathbb{G}$, if

$$\sup_{h \in BL_1} |\mathbb{E}_{\xi}(h(\mathbb{G}_{\xi,n})) - \mathbb{E}(h(\mathbb{G}))| \xrightarrow{\mathbb{P}^*} 0 \quad \text{and}$$

$$\mathbb{E}_{\xi}(h(\mathbb{G}_{\xi,n}))^* - \mathbb{E}_{\xi}(h(\mathbb{G}_{\xi,n}))_* \xrightarrow{\mathbb{P}} 0, \quad \text{for all } h \in BL_1,$$

where BL_1 is the set of 1-Lipschitz functions that are uniformly bounded by 1, and where $\mathbb{E}_{\xi}(X)^*$ and $\mathbb{E}_{\xi}(X)_*$ are inner and outer measurable cover functions. The

notation \mathbb{E}_ξ means that the expectation is taken respectively to ξ with the data fixed. In order to obtain a weak conditional convergence result for $(\widehat{C}_{\xi,n} - \widehat{C}_n, \widehat{\gamma}_{\xi,n} - \widehat{\gamma}_n)$ we need to assume more conditions on the weights. We define the norm

$$\|\xi\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}(|\xi| > x)} dx,$$

and we assume that

$$\sup_{n \geq 1} \|\xi_{1,n} - 1\|_{2,1} < \infty, \quad (7.7)$$

$$\frac{1}{\sqrt{k}} \mathbb{E} \left(\max_{1 \leq i \leq k} |\xi_{i,n} - 1| \right) \rightarrow 0, \quad (7.8)$$

$$\frac{1}{k} \sum_{i=1}^k (\xi_{i,n} - 1)^2 \xrightarrow{\mathbb{P}} c^2 > 0. \quad (7.9)$$

Note that the Efron and Bayesian bootstrap in Example 7.1 satisfy those assumptions with $c = 1$.

Theorem 7.3. *Let $(\xi_{n,1}, \dots, \xi_{n,k})$ be non negative exchangeable random variables, independent of $(X_i, Y_i)_{1 \leq i \leq n}$, satisfying condition (7.7), (7.8) and (7.9) such that $\frac{1}{k} \sum_{i=1}^k \xi_{n,i} = 1$. Under assumptions (7.4) and (7.5), for a sequence $k = k(n)$ satisfying $k \rightarrow +\infty$, $\frac{k}{n} \rightarrow 0$ and $\sqrt{k} A\left(\frac{n}{k}\right) \rightarrow 0$, we have*

$$\sqrt{k} \begin{pmatrix} \widehat{C}_{\xi,n}(\cdot) - \widehat{C}_n(\cdot) \\ \widehat{\gamma}_{\xi,n} - \widehat{\gamma}_n \end{pmatrix} \underset{\xi}{\rightsquigarrow} \begin{pmatrix} cB(\cdot) \\ cN \end{pmatrix}, \quad (7.10)$$

where $N \sim \mathcal{N}(0, \gamma^2)$, B is a C -Brownian bridge and N and B are independent.

Note that in our framework, Bayesian bootstrap is of particular interest, since $\log(Y_{n+1-i:n}/Y_{n-k:n})$ follow the Exp(1) distribution when $Y_{n+1-i:n}/Y_{n-k:n}$ is exactly Pareto(1) distributed. Consequently, for $1 \leq i \leq k$,

$$\frac{\log\left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}}\right)}{\sum_{j=1}^k \log\left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}}\right)} = \frac{\log\left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}}\right)}{\widehat{\gamma}_k} \stackrel{d}{=} \xi_{i,n}.$$

This equality in distribution remains true with $Y_{n+1-i:n}/Y_{n-k:n}$ following a Pareto(α) distribution. The parameter α is a scale parameter of $\log(Y_{n+1-i:n}/Y_{n-k:n})$, hence it is canceled out in the ratio. Consequently, in this exact Pareto(α) model, we have

$$\left(\sqrt{k} \frac{\Delta_n(u)}{\widehat{\gamma}_n} \right)_{u \in \mathbb{R}^d} \stackrel{d}{=} \left(\sqrt{k} (\widehat{C}_{\xi,n}(u) - \widehat{C}_n(u)) \right)_{u \in \mathbb{R}^d}.$$

In order to propose a validation procedure for the proportional tail model, we build a test based on the bootstrap. Consider the statistics

$$T_n := \sup_{u \in \mathbb{R}^d} \left| \sqrt{k} \frac{\Delta_n(u)}{\widehat{\gamma}_n} \right|.$$

Under the proportional tails model assumptions, Theorem 7.1 entails that $T_n \xrightarrow{d} \sup_{u \in \mathbb{R}^d} |B(u)|$, this is what we test to validate the model. The p -value of the test is based on R bootstrap replications of

$$T_\xi^r := \sup_{u \in \mathbb{R}^d} \sqrt{k} \left| \widehat{C}_{\xi^r, n}(u) - \widehat{C}_n(u) \right|$$

with $(\xi_{n,i}^r)_{1 \leq i \leq k}$ are weights satisfying the conditions of Theorem 7.3 for $1 \leq r \leq R$. Since T_n and T_ξ^r have the same weak limit, the purpose of the test is to use R bootstrap replications of T_ξ^r to approximate a quantile of $\sup |B|$ and to compare it to the observed value of T_n . The p -value of this test is given by

$$p\text{-value} := \frac{1}{R} \sum_{r=1}^R \mathbb{1}_{\{T_\xi^r \geq T_n\}}.$$

For fixed α , say $\alpha = 5\%$ or 1% , we say that the model is valid with confidence level $1 - \alpha$, if $p\text{-value} \geq \alpha$.

7.1.4 Coupling arguments for the proof

In this section we use the same notations as in Chapters 5 and 6. Similarly to Chapter 6, we consider the empirical measure of exceedances above threshold $Y_{n-k:n}$

$$\Pi_{n,k} = \frac{1}{k} \sum_{i=1}^k \varepsilon_{(X_{(n+1-i)}, \frac{Y_{n+1-i:n}}{Y_{n-k:n}})},$$

where $X_{(i)}$ denote the concomitant order statistics.

In order to analyze the proportional tail model, the Wasserstein distance of order p considered here is defined on the space of observations $(\mathbb{R}^d \times [1, +\infty), d)$ with metric d defined as

$$d((x_1, y_1), (x_2, y_2)) = \mathbb{1}_{\{x_1 \neq x_2\}} + |\log(y_1) - \log(y_2)|, \quad x_1, x_2 \in \mathbb{R}^d, \quad y_1, y_2 \in [1, +\infty).$$

The counterpart of Proposition 6.2 in the context of regression involves the distribution of $(X, U(t)^{-1}Y) | Y > U(t)$. The construction of a coupling between this distribution and the limit distribution \mathbf{P}^* is closely related with the following lemma.

Lemme 7.1. *Let $y > 1$ and $(\tilde{X}_1, \tilde{Y}_1), \dots, (\tilde{X}_k, \tilde{Y}_k)$ be an i.i.d sample with distribution $\mathcal{L}((X, Y) | Y > y)$. Consider the empirical measure*

$$\tilde{\Pi}_{k,y} = \frac{1}{k} \sum_{i=1}^k \varepsilon_{(\tilde{X}_i, \tilde{Y}_i)}.$$

We have

$$\mathcal{L}(\tilde{\Pi}_{k,t}) = \mathcal{L}(\Pi_{n,k} | Y_{n-k:n} = y).$$

This lemma provides a simple representation of the empirical measure of exceedances conditional on the threshold : it is equal to the distribution of the empirical measure of i.i.d. observations with the suitable distribution.

Proposition 7.1. *Let $p \in [1, \infty)$, we consider the Wasserstein space $\mathcal{W}_p(\mathbb{R}^d \times [1, +\infty))$ with the underlying distance d and let $t_0 > 1$ such that $U(t_0) > 0$.*

1. *For $t \geq t_0$, under conditions (7.1) and (7.3),*

$$\lim_{t \rightarrow \infty} W_p(P_{(X, U(t)^{-1}Y)|Y > U(t)}, \mathbf{P}^*) = 0.$$

2. *If we assume moreover the second order conditions (7.5), then*

$$W_p(P_{(X, U(t)^{-1}Y)|Y > U(t)}, \mathbf{P}^*) = O(A(t)), \quad \text{as } t \rightarrow \infty.$$

7.1.5 Approximation of empirical measures

We compare the empirical measure $\Pi_{n,k}$ with

$$\Pi_k^* = \frac{1}{k} \sum_{i=1}^k \varepsilon_{(X_i^*, Y_i^*)}$$

where X_1^*, \dots, X_k^* are i.i.d with density $\sigma(x)P_X(dx)$ and, independently, Y_1^*, \dots, Y_k^* are i.i.d with Pareto distribution P_α .

Theorem 7.4. *Assume F is continuous. Let $p \in [1, \infty)$,*

1. *For $t > t_0$ such that $U(t_0) > 0$, under assumptions (7.1) and (7.3), with $k := k(n)$ an intermediary sequence,*

$$W_p^{(2)}(P_{\Pi_{n,k}|Y_{n-k:n}}, P_{\Pi_k^*}) \rightarrow 0 \quad \text{in probability.}$$

2. *Moreover, if we assume (7.5) we have*

$$W_p^{(2)}(P_{\Pi_{n,k}|Y_{n-k:n}}, P_{\Pi_k^*}) = O\left(A\left(\frac{n}{k}\right)\right) \quad \text{in probability.}$$

In order to prove Addendum 7.2, which extends Theorem 7.1 to the case where the Hill estimator is biased, We compare also the empirical measure $\Pi_{n,k}$ with the empirical measure $\Pi_{k,t}^*$ defined by

$$\Pi_{k,t}^* = \frac{1}{k} \sum_{i=1}^k \varepsilon_{\left(X_i^*, Y_i^* \left(1 + A_{ext}\left(\frac{1}{F_x(U(t))}\right)\right)^{\frac{(Y_i^*)^\rho / \gamma - 1}{\rho}}\right)}, \quad k \geq 1, t \geq 1$$

with X_1^*, \dots, X_k^* i.i.d with distribution σP_X and Y_1^*, \dots, Y_k^* i.i.d with Pareto distribution P_α .

We consider, in this case, the asymptotic normality of the Hill estimator with intermediary sequence $k = k(n)$ satisfying $\sqrt{k}A_{ext}(n/k) \rightarrow \lambda \in \mathbb{R}$ and $\sqrt{k}A_{prop}(n/k) \rightarrow 0$.

Theorem 7.5. *Let $p \in [1, \infty)$. Assume that F satisfies the second order assumptions (7.1) and (7.5) with moreover $A_{prop} = o(A_{ext})$. Then,*

$$W_p^{(2)}\left(P_{\Pi_{n,k}|Y_{n-k:n}=U(t)}, P_{\Pi_{k,t}^*}\right) = o(A_{ext}(t))$$

as $t \rightarrow \infty$.

7.1.6 Bootstrap

In this section we address a question of independent interest, namely a version of Theorem 6.1 for bootstrap empirical measures. The following results are useful in our context of model validation since it consists in an important tool of the proof of Theorem 7.3.

Let two samples (Z_1, \dots, Z_k) and (Z_1^*, \dots, Z_k^*) in a separable and complete space \mathcal{X} endowed by a distance d . We build empirical measures on these samples,

$$Q = \frac{1}{k} \sum_{i=1}^k \delta_{Z_i} \quad \text{and} \quad Q^* = \frac{1}{k} \sum_{i=1}^k \delta_{Z_i^*}.$$

We now consider (ξ_1, \dots, ξ_k) random variables taking values in $(0, +\infty)$ with expectation 1 and independent of (Z_1, \dots, Z_k) and (Z_1^*, \dots, Z_k^*) satisfying $\sum \xi_i = k$ almost surely. With these variables we can define bootstrap versions of Q and Q^* :

$$Q_\xi := \frac{1}{k} \sum_{i=1}^k \xi_i \delta_{Z_i} \quad \text{and} \quad Q_\xi^* := \frac{1}{k} \sum_{i=1}^k \xi_i \delta_{Z_i^*}.$$

Since bootstrap theorems involve conditional weak convergence, we consider the data (Z_1, \dots, Z_k) and (Z_1^*, \dots, Z_k^*) as fixed. It follows that $Q, Q^* \in \mathcal{W}_p(\mathcal{X})$ are deterministic measures, while $Q_\xi, Q_\xi^* \in \mathcal{W}_p^{(2)}(\mathcal{X})$ are random measures.

Lemma 7.2. *Let (Z_1, \dots, Z_k) and (Z_1^*, \dots, Z_k^*) be fixed. Assume moreover that $\frac{1}{k} \sum \xi_i = 1$. For $p \in [1, \infty)$, we have*

$$W_p^{(2)}(P_{Q_\xi}, P_{Q_\xi^*}) = W_p(Q, Q^*).$$

This lemma allows us to transfer bootstrap results from the limit model with distribution P^* to the model with exceedances above threshold.

7.2 Simulation study

7.2.1 Level control

The data generating process

We consider n realisations $(x_i, y_i)_{i=1}^n$ of i.i.d copies $(X, Y) \in \mathbb{R}^2 \times \mathbb{R}$ with $n = 5000$ or 10000. In each simulation X is drawn uniformly on $[0, 1]^2$, and Y is built as follows : Consider

$$\sigma : x = (x^{(1)}, x^{(2)}) \in [0, 1]^2 \mapsto e^{(x^{(1)} + x^{(2)})/2}$$

and $\gamma = 1/4$. We define $Y := \sigma(X)Y_0$ with Y_0 a random variable chosen among the following three possibilities :

- Pareto : $Y_0 = U^{-\gamma}$ with $U \sim \mathcal{U}([0, 1])$

- Fréchet : $Y_0 = -\log(U)^{-\gamma}$ with $U \sim \mathcal{U}([0, 1])$
- Burr : $Y_0 = \sqrt{(1/(1-U)^{1/(2\gamma)} - 1)}$ with $U \sim \mathcal{U}([0, 1])$

In those three first situations, γ is the extreme value index of Y_0 which belongs to the domain of attraction of Fréchet distribution. With any of those choices for Y_0 , (X, Y) follows the proportional tails model.

Our estimators are built with the ratio k/n extreme observations chosen as 10% and 20% of the total sample size. We compute the empirical quantile of $\sqrt{k}(\hat{C}_{\xi,n} - \hat{C}_n)$ by bootstrap (Bayesian and Efron) with 1000 repetitions.

The simulations

The following table presents the proportion of acceptances over 1000 replications for the test.

TABLE 7.1 – Bayesian bootstrap

α	5%				1%			
	5000		10000		5000		10000	
	k	k	k	k	k	k	k	k
Pareto	0.949	0.96	0.949	0.95	0.988	0.983	0.988	0.985
Frechet	0.939	0.965	0.962	0.949	0.988	0.99	0.989	0.99
Burr	0.953	0.967	0.952	0.952	0.981	0.991	0.988	0.989

TABLE 7.2 – Efron bootstrap

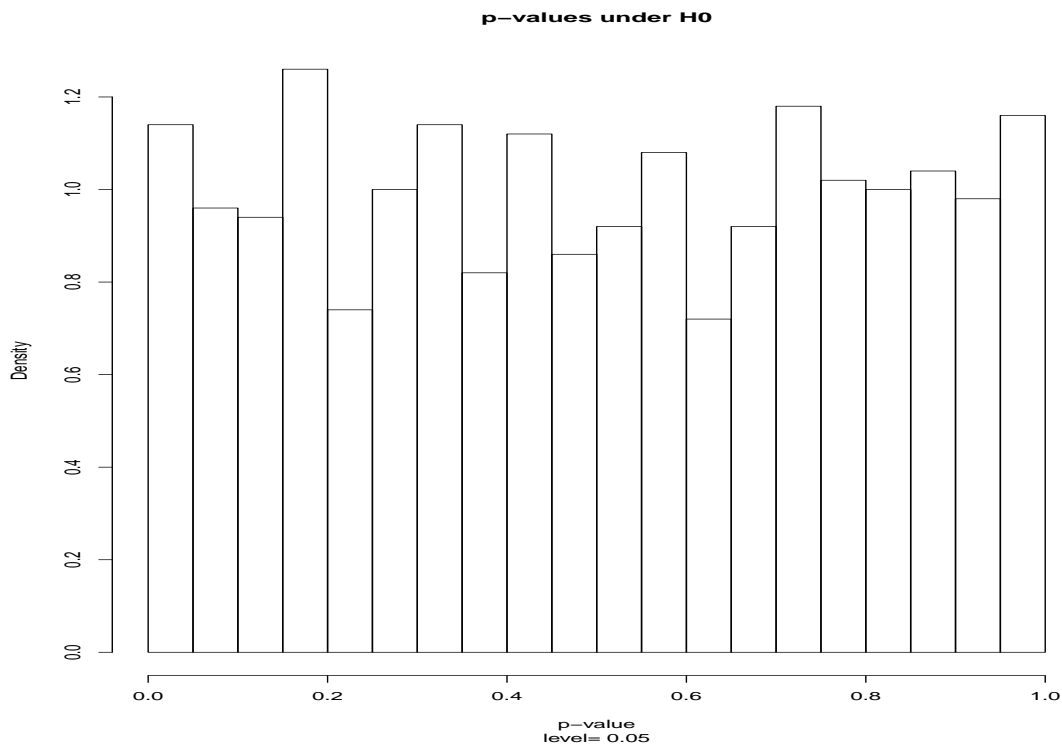
α	5%				1%			
	5000		10000		5000		10000	
	k	k	k	k	k	k	k	k
Pareto	0.94	0.949	0.95	0.949	0.994	0.989	0.99	0.991
Frechet	0.964	0.962	0.956	0.955	0.991	0.987	0.987	0.991
Burr	0.956	0.953	0.949	0.942	0.992	0.993	0.991	0.992

The observed level over 1000 replications is 95% (respectively 99%) when $\alpha = 5\%$ (respectively $\alpha = 1\%$) for the both bootstrap. Moreover, belonging to the domain of attraction of Pareto distribution rather than being exactly Pareto has no effect and does not deteriorate the level of this test.

Note that, since $\sqrt{k}\Delta_n(\cdot)/\hat{\gamma}_n$ and $\sqrt{k}(\hat{C}_{\xi,n}(\cdot) - \hat{C}_n(\cdot))$ have the same weak limit, we can compute p -values for this test. With the knowledge of p -values, we can

deduce the level since we validate the model as soon as the p -value is greater than α . The following histogram presents the distribution of 1000 p -values obtained for the sample of size $n = 5000$, $k = 1000$ with Fréchet distribution for the extremes (this corresponds to the orange tile on the table 7.1).

FIGURE 7.1 – Distribution of p -values under H_0



As expected, for sample for which proportional tails model is valid, the p -values are uniformly distributed.

7.2.2 Power study

The data generating process

The purpose of this section is to study the power of this test. To do this, we modify the model introduced in Section 7.2.1. Consider, $\gamma : x = (x^{(1)}, x^{(1)}), \varepsilon \mapsto \gamma + \varepsilon(x^{(1)} + x^{(2)})/2$ and choose the variable Y_0 among the following possibilities :

- Pareto : $Y_0 = U^{-\gamma(X,\varepsilon)}$ with $U \sim \mathcal{U}([0, 1])$
- Fréchet : $Y_0 = -\log(U)^{-\gamma(X,\varepsilon)}$ with $U \sim \mathcal{U}([0, 1])$
- Burr : $Y_0 = \sqrt{(1/(1-U)^{1/(2\gamma(X,\varepsilon))} - 1)}$ with $U \sim \mathcal{U}([0, 1])$

For $x \in [0, 1]^2$ and a fixed ε , $\gamma(x, \varepsilon)$ is the extreme value index of Y given $X = x$. In this framework, the extreme value index may depend on x . This dependency is driven by the choice of ε . If $\varepsilon = 0$, γ do not depend on x . Consequently, those samples satisfies the proportional tail model assumptions, this corresponds to the framework of Section 7.2.1. Conversely, if $\varepsilon \neq 0$, we are building samples which depart from the proportional tail model.

The power of the test is estimated with $\varepsilon = 0.1, -0.1$ and 1 . With those values of ε we are ensured that Y_0 keeps belonging to the a domain of attraction of Fréchet.

The simulations

TABLE 7.3 – Bayesian bootstrap with $\varepsilon = 0.1$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.821	0.639	0.651	0.371	0.931	0.836	0.857	0.606
Frechet	0.803	0.546	0.588	0.239	0.917	0.768	0.795	0.443
Burr	0.749	0.588	0.614	0.301	0.917	0.799	0.786	0.515

TABLE 7.4 – Bayesian bootstrap with $\varepsilon = 0.5$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.543	0.255	0.257	0.37	0.748	0.484	0.508	0.123
Frechet	0.54	0.306	0.267	0.056	0.785	0.555	0.505	0.152
Burr	0.574	0.266	0.253	0.048	0.803	0.494	0.548	0.146

TABLE 7.5 – Bayesian bootstrap with $\varepsilon = 1$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.102	0	0.005	0	0.29	0.005	0.024	0
Frechet	0.064	0	0	0	0.204	0	0.005	0
Burr	0	0	0	0	0	0	0	0

TABLE 7.6 – Bayesian bootstrap with $\varepsilon = -0.1$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.53	0.254	0.276	0.041	0.78	0.516	0.5	0.137
Frechet	0.571	0.309	0.273	0.047	0.786	0.547	0.515	0.172
Burr	0.566	0.268	0.26	0.037	0.79	0.536	0.54	0.138

TABLE 7.7 – Efron bootstrap with $\varepsilon = 0.1$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.813	0.683	0.64	0.386	0.945	0.837	0.844	0.642
Frechet	0.797	0.56	0.584	0.209	0.928	0.759	0.822	0.479
Burr	0.772	0.584	0.581	0.302	0.9	0.801	0.809	0.538

TABLE 7.8 – Efron bootstrap with $\varepsilon = 0.5$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.205	0.005	0.024	0	0.474	0.046	0.077	0
Frechet	0.144	0	0.004	0	0.377	0.011	0.051	0
Burr	0.01	0	0	0	0.054	0	0.001	0

TABLE 7.9 – Efron bootstrap with $\varepsilon = 1$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.088	0.002	0.01	0	0.317	0.003	0.012	0
Frechet	0.057	0	0.03	0	0.221	0	0.007	0
Burr	0	0	0	0	0	0	0	0

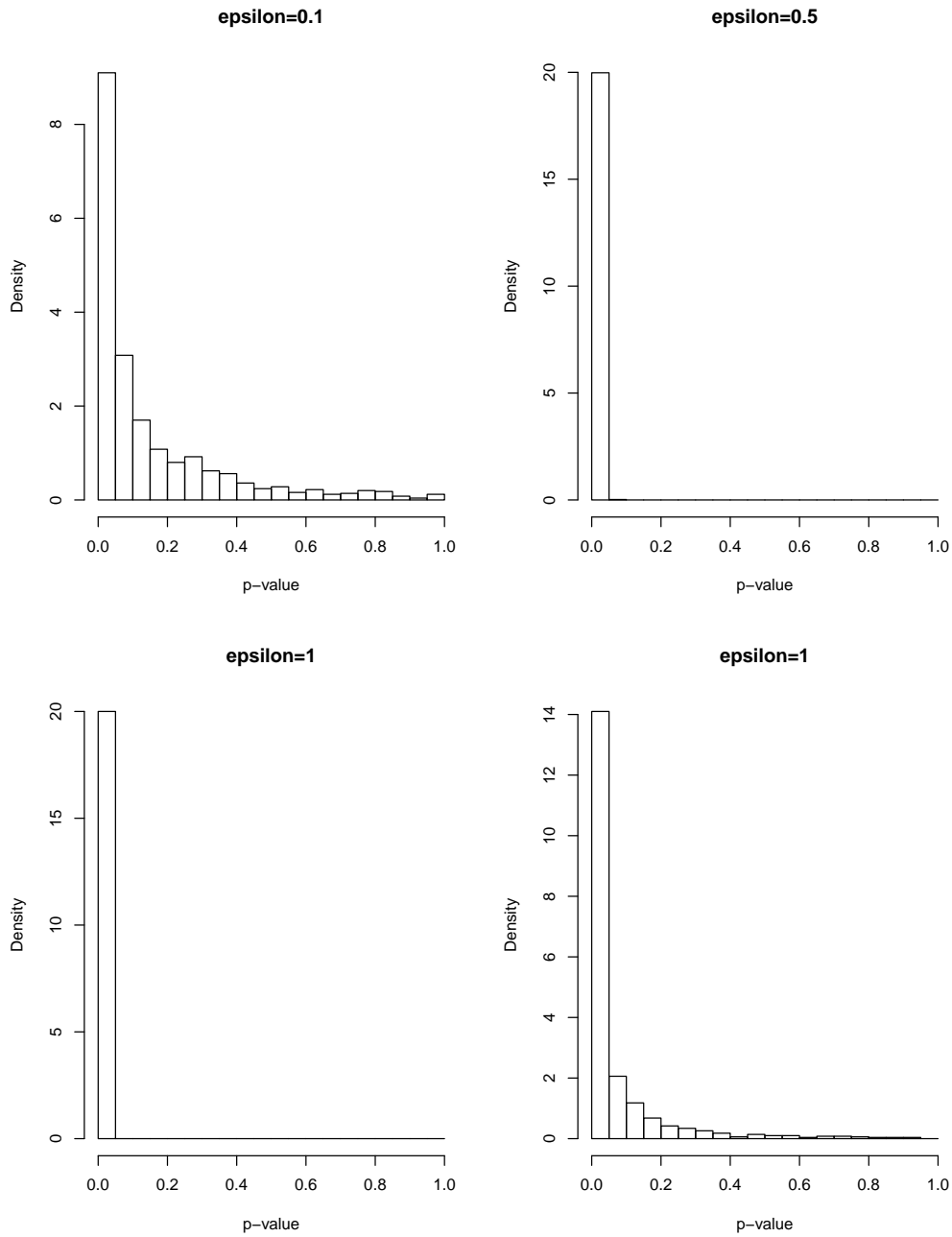
TABLE 7.10 – Efron bootstrap with $\varepsilon = -0.1$

α	5%				1%			
	5000		10000		5000		10000	
	500	1000	1000	2000	500	1000	1000	2000
Pareto	0.56	0.262	0.25	0.043	0.801	0.488	0.502	0.154
Frechet	0.561	0.28	0.274	0.046	0.762	0.532	0.52	0.16
Burr	0.585	0.288	0.284	0.046	0.796	0.497	0.497	0.153

Note that there are two factors which influence the power of the test. The power increases with k , the number of extremes. Moreover, the power also increases with $|\varepsilon|$. We can also remark again that there is no difference on the power between distribution belonging the domain of attraction of Fréchet or being exactly Pareto.

The following histograms present the distribution of p -values for the Fréchet model but when $\varepsilon \neq 0$, that is for a sample which fails the proportional tail model.

FIGURE 7.2 – Distribution of p -values under H_1



We see that the distribution of p -values concentrates around zero as $|\varepsilon|$ grows. This traduces the fact that the power of the test tends to 1 as $|\varepsilon|$ increases.

7.3 Proofs

7.3.1 Proofs related to section 7.1.4

Proof of Lemma 7.1. Consider $(Z_i, Y_i)_{1 \leq i \leq n}$ i.i.d copies of (Z, Y) $Z := (X, Y) \in \mathfrak{X} := \mathbb{R}^d \times \mathbb{R}$. Moreover consider a measurable map φ from $(\mathfrak{X} \times \mathbb{R}, \mathcal{B}(\mathfrak{X} \otimes \mathcal{B}(\mathbb{R})))$ to any measurable space. We finally define the empirical measures define

$$\Pi_k^\sharp := \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{\{Y_i > Y_{n-k:n}\}} \delta_{\varphi(X_i, Y_{n-k:n})},$$

and for each $y \in \mathbb{R}$,

$$\tilde{\Pi}_{k,y}^\sharp := \frac{1}{k} \sum_{i=1}^k \delta_{\varphi(\tilde{Z}_i, y)}$$

where \tilde{Z}_i are i.i.d with law $\mathcal{L}(Z|Y > y)$.

With those definition, we prove the Lemma 7.1 showing

$$\mathcal{L}(\Pi_k^\sharp | Y_{n-k:n} = y) = \mathcal{L}(\tilde{\Pi}_{k,y}^\sharp)$$

and specializing with the function $\varphi : (Z, Y) \mapsto Z$.

For fixed $y \in \mathbb{R}$, write \mathfrak{p}_y the law of $\tilde{\Pi}_{k,y}^\sharp$, namely

$$\mathbb{P}\left(\frac{1}{k} \sum_{i=1}^k \delta_{\varphi(\tilde{Z}_i, y)} \in A\right) := \mathfrak{p}_y(A), \quad A \in \mathcal{W}_p(\mathcal{X})$$

Let $A \in \mathcal{W}_p(\mathcal{X})$ and B be Borel, the proof consists in showing the following equality,

$$\mathbb{P}\left(\left(\frac{1}{k} \sum_{i \in I_{n,k}} \delta_{\varphi(Z_i, Y_{n-k:n})}, Y_{n-k:n}\right) \in (A, B)\right) = \int_B \mathfrak{p}_{y_{n-k}}(A) P_{Y_{n-k:n}}(dy) \quad (7.11)$$

where $I_{n,k}$ denotes the set of the indexes $1 \leq i \leq n$ such that $Y_i > Y_{n-k:n}$. We start by computing the joint distribution of Π_k^\sharp and $Y_{n-k:n}$. Given a sample (y_1, \dots, y_n) having no ties (since F is assumed continuous), there exists a unique $\sigma \in \mathfrak{S}_n$ which orders (y_1, \dots, y_n) . Hence, since (Y_1, \dots, Y_n) has almost surely no ties, we have

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{k} \sum_{i \in I_{n,k}} \delta_{\varphi(Z_i, Y_{n-k:n})} \in A \cap Y_{n-k:n} \in B \cap \exists! \sigma \text{ st } Y_{\sigma(1)} < \dots < Y_{\sigma(n)}\right) \\ &= \mathbb{P}\left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(Z_i, Y_{n-k:n})} \in A \cap Y_{n-k:n} \in B \cap \exists! \sigma \text{ st } Y_{\sigma(1)} < \dots < Y_{\sigma(n)}\right) \\ &= \sum_{\sigma \in \mathfrak{S}_n} \mathbb{P}\left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(Z_i, Y_{n-k:n})} \in A \cap Y_{n-k:n} \in B \cap \text{st } Y_{\sigma(1)} < \dots < Y_{\sigma(n)}\right) \\ &= n! \mathbb{P}\left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(Z_i, Y_{n-k:n})} \in A \cap Y_{n-k:n} \in B \cap Y_1 < \dots < Y_n\right). \end{aligned}$$

The last probability can be written as

$$\int_{\mathcal{Z}^n} \mathbb{1}_{\{\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(z_i, y_{n-k:n})} \in A\}} \mathbb{1}_{\{y_{n-k:n} \in B\}} \mathbb{1}_{\{y_1 < \dots < y_n\}} \prod_{i=1}^n P_{Z,Y}(dz_i, dy_i). \quad (7.12)$$

For a fixed y_{n-k} we have

$$\begin{aligned} \int_{\mathcal{Z}^{n-k-1}} \mathbb{1}_{\{y_1 < \dots < y_{n-k}\}} \prod_{i=1}^{n-k-1} P_{Z,Y}(z_i, y_i) &= \mathbb{P}(Y_1 < \dots < Y_{n-k-1} \leq y_{n-k}) \\ &= \frac{F(y_{n-k})^{n-k-1}}{(n-k-1)!}. \end{aligned}$$

Hence (7.12) is equal to

$$\int_B \frac{F(y_{n-k})^{n-k-1}}{(n-k-1)!} \left[\int_{\mathcal{Z}^k} \mathbb{1}_{\{\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(z_i, y_{n-k:n})} \in A\}} \mathbb{1}_{\{y_{n-k} < \dots < y_n\}} \prod_{i=n-k+1}^n P_{Z,Y}(dz_i, dy_i) \right] P_Y(dy_{n-k}). \quad (7.13)$$

Seeing that

$$\mathbb{1}_{\{y_{n-k} < \dots < y_n\}} = \mathbb{1}_{\{y_{n-k+1} < \dots < y_n\}} \prod_{i=n-k+1}^n \mathbb{1}_{\{y_{n-k} < y_i\}}$$

the integral between brackets in 7.13 is equal, for fixed y_{n-k} , to

$$\begin{aligned} &\mathbb{P} \left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(Z_i, y_{n-k:n})} \in A \cap \forall i > n-k, Y_i > y_{n-k} \cap Y_{n-k+1} < \dots < Y_n \right) \\ &= \mathbb{P} \left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(Z_i, y_{n-k:n})} \in A \cap Y_{n-k+1} < \dots < Y_n \mid \forall i > n-k, Y_i > y_{n-k} \right) \mathbb{P}(Y_i > y_{n-k})^k \\ &= \mathbb{P} \left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(\tilde{Z}_i, y_{n-k:n})} \in A \cap \tilde{Y}_{n-k+1} < \dots < \tilde{Y}_n \right) \mathbb{P}(Y_i > y_{n-k})^k \\ &= \mathbb{P} \left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(\tilde{Z}_i, y_{n-k:n})} \in A \cap \tilde{Y}_{n-k+1} < \dots < \tilde{Y}_n \right) (1 - F(y_{n-k}))^k \end{aligned}$$

The second equality holds because $\mathcal{L}(\tilde{Z}_i, \tilde{Y}_i) = \mathcal{L}((Z_i, Y_i/y) \mid Y_i > y)$. Moreover, we have

$$\begin{aligned} &\mathbb{P} \left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(\tilde{Z}_i, y_{n-k:n})} \in A \cap \tilde{Y}_{n-k+1} < \dots < \tilde{Y}_n \right) \\ &= \mathbb{P} \left(\frac{1}{k} \sum_{i=n-k+1}^n \delta_{\varphi(\tilde{Z}_i, y_{n-k:n})} \in A \right) \mathbb{P}(\tilde{Y}_{n-k+1} < \dots < \tilde{Y}_n) \\ &= \mathbf{p}_{y_{n-k}}(A) \frac{1}{k!} \end{aligned}$$

Combining those identifications we can see that the right hand side probability in (7.11) is

$$\int_B \mathbf{p}_{y_{n-k}}(A) \frac{n!}{k!(n-k-1)!} F(y_{n-k})^{n-k-1} (1 - F(y_{n-k}))^k P_Y(dy_{n-k})$$

which can be simplified into

$$\int_B \mathbf{p}_{y_{n-k}}(A) P_{Y_{n-k:n}}(dy).$$

This concludes the proof. \square

Remarque 7.3. *This proof use the fact that the sample (Y_1, \dots, Y_n) had almost surely no ties. Nevertheless, the result can be extended to the case where F is not continuous using the ties breaking algorithm given at the end of Section 3.2.*

Proof of Proposition 7.1. 1. Define, for $t \geq t_0$,

$$A_p(t) = \left(\int_{\mathbb{R}^d} \int_1^{+\infty} \left| \log \left(\frac{U_x(z/\bar{F}_x(U(t)))}{z^\gamma U(t)} \right) \right|^p \frac{dz}{z^2} P_{X|Y>U(t)}(dx) \right)^{1/p}. \quad (7.14)$$

Let a fixed $t \geq t_0$. Consider (\tilde{X}, \tilde{Y}) and (X^*, Y^*) such that $\tilde{X} \sim P_{X|Y>U(t)}$ and $X^* \sim \sigma P_X$ satisfy the maximal coupling (Lemma 5.1) and with \tilde{Y}, Y^* built as follows. Given $\tilde{X} = \tilde{x}$ and $X^* = x^*$, consider Z a standard unit Pareto random variable and set $\tilde{Y} = \frac{U_x(\bar{F}_x(U(t)))}{U(t)}$ and $Y^* = Z^\gamma$ with $\gamma = 1/\alpha$. Consequently (\tilde{X}, \tilde{Y}) and (X^*, Y^*) is a coupling between $P_{(X,U(t)^{-1}Y)|Z>t}$ and $\sigma P_X \otimes P_\alpha$. The triangular inequality for the L^p -norm entails

$$\begin{aligned} \|d((\tilde{X}, \tilde{Y}), (X^*, Y^*))\|_{L^p} &\leq \|\mathbf{1}_{\tilde{X} \neq X^*}\|_{L^p} + \|\log(\tilde{Y}) - \log(Y^*)\|_{L^p} \\ &\leq \mathbb{P}(\tilde{X} \neq X^*)^{1/p} + \|\log(\tilde{Y}) - \log(Y^*)\|_{L^p} \\ &\leq \|P_{X|Z>t} - \sigma P_X\|_{TV}^{1/p} + \|\log(\tilde{Y}) - \log(Y^*)\|_{L^p}. \end{aligned}$$

This leads to

$$\begin{aligned} W_p(P_{(X,U(t)^{-1}Y)|Z>t}, \sigma P_X \otimes P_\alpha) \\ \leq \|P_{X|Z>t} - \sigma P_X\|_{TV}^{1/p} + \|\log(\tilde{Y}) - \log(Y^*)\|_{L^p}. \end{aligned}$$

By construction, we can see that

$$\|\log(\tilde{Y}) - \log(Y^*)\|_{L^p} = \left(\int_{\mathbb{R}^d} \int_1^{+\infty} \left| \log \left(\frac{U_x(z/\bar{F}_x(U(t)))}{z^\gamma U(t)} \right) \right|^p \frac{dz}{z^2} P_{X|Y>U(t)}(dx) \right)^{1/p}.$$

Introducing this quantity in the previous inequality, we have

$$W_p(P_{(X,U(t)^{-1}Y)|Y>U(t)}, \sigma(x) P_X(dx) \otimes P_\alpha) \leq A_p(t) + \|P_{X|Y>y}(dx) - \sigma(x) P_X(dx)\|_{TV}^{1/p}.$$

The quantity $\|P_{X|Y>y}(dx) - \sigma(x) P_X(dx)\|_{TV}^{1/p}$ converge to 0 by Lemma 7.8. Moreover, the Lemma 7.6 entails that A_p has the same asymptotic behavior as

$$A'_p(t) = \left(\int_{\mathbb{R}^d} \int_1^{+\infty} \left| \log \left(\frac{U(z/\bar{F}_x(U(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} \right) \right|^p \frac{dz}{z^2} P_{X|Y>u(t)}(dx) \right)^{1/p}.$$

The extreme value condition (7.3) implies

$$\lim_{t \rightarrow +\infty} \left| \log \frac{U(z/\bar{F}_x(U(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} \right| = 0.$$

Furthermore, Potter's bounds (see *e.g.* [22] Proposition B.1.9) give a $t_1 > 0$ such that for $t > t_1$

$$\left| \frac{U(z/\bar{F}_x(U(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} \right| \leq \max((1 - \varepsilon)z^{-\varepsilon}, (1 + \varepsilon)z^\varepsilon).$$

This implies an integrable bound for $\left| \log \left(\frac{U(z/\bar{F}_x(U(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} \right) \right|^p$ and the convergence of A'_p to zero follows from the dominated convergence theorem.

2. In order to make equations easier to read, we denote by $\varphi_t := \bar{F}_x(U(t))$. The first point of this Theorem together with assumption that $A_{prop} = o(A_{ext})$ entail that we can conclude this proof by showing that $A_p(t) = O(A_{ext}(t))$. Moreover, using the decomposition (7.20) we can see that $\lim A_p(t)/A_{ext}(t) = \lim A'_p(t)/A_{ext}(t)$. The rest of the proof is focused on the study of asymptotic behavior of A'_p/A_{ext} . With the auxiliary function $f(z) = \log(U(z)/z^\gamma)$, consider the ratio

$$\frac{A'_p(t)}{A_{ext}(t)} = \left(\int_{\mathbb{R}^d} \int_1^{+\infty} \left| \frac{f(z/\varphi_t) - f(\varphi_t)}{A_{ext}(t)} \right|^p \frac{dz}{z^2} P_{X|Y>u(t)}(dx) \right)^{1/p}.$$

As a consequence of Lemma 7.7,

$$\begin{aligned} \frac{A'_p(t)}{A_{ext}(t)} &\sim \left(\int_{\mathbb{R}^d} \int_1^{+\infty} \sigma(x)^{\rho p} \left| \frac{f(z/\varphi_t) - f(\varphi_t)}{A_{ext}(1/\varphi_t)} \right|^p \frac{dz}{z^2} P_{X|Y>u(t)}(dx) \right)^{1/p} \\ &\leq \varepsilon_\sigma^\rho \left(\int_{\mathbb{R}^d} \int_1^{+\infty} \left| \frac{f(z/\varphi_t) - f(\varphi_t)}{A_{ext}(1/\varphi_t)} \right|^p \frac{dz}{z^2} P_{X|Y>u(t)}(dx) \right)^{1/p}. \end{aligned}$$

The second order assumption (7.5) entails that, in the terminology of [22], f is of extended regular variation at infinity with index ρ . Consequently, we can use Theorem B.2.18 p.383 in [22] to get, for an arbitrary $\varepsilon, \delta > 0$, there exists t_0 such that

$$\left| \frac{f(z/\varphi_t) - f(\varphi_t)}{A_{ext}(1/\varphi_t)} - \frac{z^\rho - 1}{\rho} \right| \leq \varepsilon z^{\rho+\delta}, \quad \text{for all } t \geq t_0, z > 1. \quad (7.15)$$

Then, the convergence of $A'_p(t)/A_{ext}(t)$ to 0 is a consequence of the dominated convergence theorem with dominating function provided by Equation (7.15). \square

Proof of Theorem 7.4. 1. For a random variable Z with standard unit Pareto distribution, the random variable $\frac{U_x(Z/\bar{F}_x(U(t)))}{U(t)}$ has the same distribution as Y given $Y > U(t)$ and $X = x \in \mathbb{R}^d$. We deduce that the joint conditional distribution of k top order statistics together with concomitant statistics

(($X_{(n-k+1)}, Y_{n-k+1:n}/Y_{n-k:n}$), ..., ($X_{(n)}, Y_{n:n}/Y_{n-k:n}$)) given $Y_{n-k:n} = U(t)$ is the same as (($\tilde{X}_{(1)}, \tilde{Y}_{1:k}$), ..., ($\tilde{X}_{(k)}, \tilde{Y}_{k:k}$)) with $\tilde{Y}_i = \frac{U_x(Z_i/\bar{F}_x(U(t)))}{U(t)}$ build with an i.i.d unit Pareto sample (Z_1, \dots, Z_k) and ($\tilde{X}_1, \dots, \tilde{X}_k$) are k copies of a random variable $\tilde{X} \sim P_{X|Y>U(t)}$. This way, we build a sample (\tilde{X}_1, \tilde{Y}_1), ..., (\tilde{X}_k, \tilde{Y}_k) with $\mathcal{L}(\tilde{X}, \tilde{Y}) = \mathcal{L}(X, Y|Y > U(t))$. As a consequence of Lemma 7.1, the conditional distribution of $\Pi_{n,k}$ given $Y_{n-k:n} = U(t)$ is equal to the distribution of

$$\tilde{\Pi}_{k,t} = \frac{1}{k} \sum_{i=1}^k \epsilon_{(\tilde{X}_i, \tilde{Y}_i)}.$$

Consequently, Theorem 2.3 in [11] together with Proposition 7.1 entail

$$\begin{aligned} W_p^{(2)}(P_{\Pi_{n,k}|Y_{n-k:n}=U(t)}, P_{\Pi_k^*}) &\leq W_p^{(2)}(P_{\tilde{\Pi}_{k,t}}, P_{\Pi_k^*}) \\ &\leq W_p(P_{(X,U(t)^{-1}Y)|Y>U(t)}, \sigma P_X \otimes P_\alpha) \\ &\leq A_p(t) + \|P_{X|Y>U(t)} - \sigma P_X\|_{TV}^{1/p}. \end{aligned}$$

We know moreover that $Y_{n-k:n} \xrightarrow{\mathbb{P}} \infty$ for the intermediary sequence k . Consequently, we have $1/\bar{F}(Y_{n-k:n}) \rightarrow \infty$. By Proposition 7.1 and Lemma 7.8 we get a constant M such that

$$W_p^{(2)}(P_{\Pi_{n,k}|Y_{n-k:n}}, P_{\Pi_k^*}) \leq M A_{prop}(1/\bar{F}(Y_{n-k:n})) + A_p(1/\bar{F}(Y_{n-k:n})) \rightarrow 0 \text{ in probability.}$$

2. This point is straightforward in view of the previous inequality together with Proposition 7.1. □

Proof of Theorem 7.5. Introducing, for $t > t_0$, the quantity

$$A_{p,b}(t) = \left(\int_{\mathbb{R}^d} \int_1^\infty \left| \log \frac{\frac{U_x(z/\bar{F}_x(U(t)))}{z^\gamma U(t)}}{1 + A_{ext}(1/\bar{F}_x(U(t))) \frac{z^\rho - 1}{\rho}} \right|^p \frac{dz}{z^2} P_{X|Z>t}(dx) \right)^{1/p},$$

with the same arguments as in the proof of Theorem 7.4 we have, for $t > t_0$,

$$W_p^{(2)}(P_{\Pi_{n,k}|Y_{n-k:n}=U(t)}, P_{\Pi_{k,t}^*}) \leq \|P_{X|Y>y} - \sigma P_X\|_{TV}^{1/p} + A_{p,b}(t).$$

We can remark that the asymptotic behavior of $\|P_{X|Y>y} - \sigma P_X\|_{TV}^{1/p}$ is known since we can apply Lemma 7.8 and $A_{prop} = o(A_{ext})$. Consequently the knowledge of the asymptotic behavior of $W_p^{(2)}(P_{\Pi_{n,k}|Y_{n-k:n}=U(t)}, P_{\Pi_{k,t}^*})$ only relies on $A_{p,b}$. Using the same decomposition as in Equation (7.20) in the proof of Proposition 7.1, we can see that $A_{p,b}$ has the same asymptotic behavior as

$$A'_{p,b} = \left(\int_{\mathbb{R}^d} \int_1^\infty \left| \log \frac{\frac{U(z/\bar{F}_x(U(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))}}{1 + A_{ext}(1/\bar{F}_x(U(t))) \frac{z^\rho - 1}{\rho}} \right|^p \frac{dz}{z^2} P_{X|Z>t}(dx) \right)^{1/p}.$$

Theorem B.2.18, p.383 in [22] used with the function $f : z \mapsto \log(U(z)/z^\gamma)$ implies,

$$\left| \log \frac{U(z/\bar{F}_x(u(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} - A_{ext}(1/\bar{F}_x(U(t))) \frac{z^\rho - 1}{\rho} \right| \leq \varepsilon |A_{ext}(1/\bar{F}_x(U(t)))| z^{\rho+\delta},$$

and, with $x - 2x^2 \leq \log(1+x) \leq x$, for $x \geq -1/2$, we deduce

$$\begin{aligned} & \left| \log \frac{U(z/\bar{F}_x(u(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} - \log \left(A_{ext}(1/\bar{F}_x(U(t))) \frac{z^\rho - 1}{\rho} \right) \right| \\ & \leq \varepsilon |A_{ext}(1/\bar{F}_x(U(t)))| z^{\rho+\delta} + 2A_{ext}(1/\bar{F}_x(U(t)))^2 \left| \frac{z^\rho - 1}{\rho} \right|^2. \end{aligned}$$

Now, taking the L^p -norm and ε being arbitrary, we get $A'_{p,b}(t) = o(A_{ext}(1/\bar{F}_x(U(t))))$ as $t \rightarrow \infty$. Whence the result. \square

Proof of Lemma 7.2. Let $(\tilde{Q}_\xi, \tilde{Q}_\xi^*)$ be a coupling between P_{Q_ξ} and $P_{Q_\xi^*}$. Consequently, $W_p^{(2)}(P_{Q_\xi}, P_{Q_\xi^*})^p \leq \mathbb{E}(W_p(\tilde{Q}_\xi, \tilde{Q}_\xi^*)^p)$. We deduce from Theorem 4.4 that

$$\begin{aligned} W_p(\tilde{Q}_\xi, \tilde{Q}_\xi^*)^p &= \sup_{\varphi, \psi} \int_{\mathcal{X}} \psi d\tilde{Q}_\xi^* - \int_{\mathcal{X}} \varphi d\tilde{Q}_\xi \\ &= \sup_{\varphi, \psi} \frac{1}{k} \sum_{i=1}^k \xi_i (\psi(Z_i^*) - \varphi(Z_i)) \\ &\leq \frac{1}{k} \sum_{i=1}^k \xi_i d^p(Z_i, Z_i^*), \end{aligned}$$

Where the previous supremum is taken over (φ, ψ) continuous and bounded satisfying

$$\varphi(x) - \psi(y) \leq d(x, y), \quad x, y \in \mathcal{X}. \quad (7.16)$$

Taking expectations, we have

$$\mathbb{E}(W_p(\tilde{Q}_\xi, \tilde{Q}_\xi^*)^p) \leq \frac{1}{k} \sum_{i=1}^k d^p(Z_i, Z_i^*), \quad \text{and hence}$$

$$W_p^{(2)}(P_{Q_\xi}, P_{Q_\xi^*}) \leq \left(\frac{1}{k} \sum_{i=1}^k d^p(Z_i, Z_i^*) \right)^{1/p} = W_p(Q, Q^*).$$

For the converse inequality, there exist to continuous bounded function ψ and φ fulfilling 7.16 and such that

$$W_p(Q, Q^*)^p = \int_{\mathcal{X}} \psi dQ^* - \int_{\mathcal{X}} \varphi dQ = \frac{1}{k} \sum_{i=1}^k \psi(z_i^*) - \varphi(z_i).$$

Now define

$$\begin{aligned} \varphi^\sharp : \mathcal{W}_p(\mathcal{X}) &\rightarrow \mathbb{R} & \text{and} & & \psi^\sharp : \mathcal{W}_p(\mathcal{X}) &\rightarrow \mathbb{R} \\ \Pi &\mapsto \int_{\mathcal{X}} \varphi d\Pi & & & \Pi &\mapsto \int_{\mathcal{X}} \psi d\Pi. \end{aligned}$$

Since these are Lipschitz functions, we can write

$$\begin{aligned}
W_p^{(2)}(P_{Q_\xi}, P_{Q_\xi^*}) &\geq \int_{\mathcal{W}(\mathcal{X})} \psi^\# dP_{Q_\xi^*} - \int_{\mathcal{W}(\mathcal{X})} \varphi^\# dP_{Q_\xi} \\
&\geq \mathbb{E} \left(\int_{\mathcal{X}} \psi dQ_\xi^* - \int_{\mathcal{X}} \varphi dQ_\xi \right) \\
&\geq \mathbb{E} \left(\frac{1}{k} \sum_{i=1}^k \xi_i (\psi(z_i^*) - \varphi(z_i)) \right) \\
&= \frac{1}{k} \sum_{i=1}^k \psi(z_i^*) - \varphi(z_i) = W_p(Q, Q^*)^p.
\end{aligned}$$

□

7.3.2 Proofs related to section 7.1.2

Proof of Theorem 7.1. Define the estimators on the limit model,

$$\begin{aligned}
\hat{C}_k^* &:= \frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{X_i^* \leq x\}}, \quad \text{and} \\
\hat{\gamma}_k^* &= \frac{1}{k} \sum_{i=1}^k \log(Y_i^*).
\end{aligned}$$

In virtue of Lemma 18 in [11] it is enough to show that

$$W_p(P_{\sqrt{k}(\hat{C}_k^*(\cdot) - C(\cdot), \hat{\gamma}_k^* - \gamma)} | Y_{n-k:n} = U(t), P_{\sqrt{k}(\hat{C}_k^*(\cdot) - C(\cdot), \hat{\gamma}_k^* - \gamma)}) \rightarrow 0 \quad (7.17)$$

because $\sqrt{k}(\hat{C}_k^*(\cdot) - C(\cdot), \hat{\gamma}_k^* - \gamma)$ converges in distribution to (B, N) as a straightforward consequence of the central limit Theorem and the theory of empirical processes for i.i.d samples.

Note first that \hat{C}_k^* and $\hat{\gamma}_k^*$ respectively depend only upon X^* and Y^* which are independent, so that the weak convergence of the couple is equivalent to the weak convergence of the two margins. Since Y^* have a Pareto distribution with parameter α , $\log(Y^*)$ admit a exponential distribution with parameter γ . So the convergence of $\sqrt{k}(\hat{\gamma}_k^* - \gamma)$ to a normal distribution is due to the usual central limit theorem. And, according to Theorems 2.5.2 and 2.6.7 of [82], the point is to show that the class of functions $\{f_t : x \mapsto \mathbb{1}_{(-\infty, t]}(x), t \in \mathbb{R}^d\}$ is a Vapnik-Chervonenkis class with square integrable envelope. Which is clear.

In order to prove equation (7.17), consider $L^\infty(\mathbb{R}^d) \times [1, +\infty)$ endowed with the distance D defined as

$$D((f_1, y_1), (f_2, y_2)) := \sup_{x \in \mathbb{R}^d} |f_1(x) - f_2(x)| + |\log(y_1) - \log(y_2)|.$$

We remark that $(\hat{C}_n(\cdot) - C(\cdot), \hat{\gamma}_n - \gamma)$ and $(\hat{C}_k^*(\cdot) - C(\cdot), \hat{\gamma}_k^* - \gamma)$ are obtained respectively as images of $\Pi_{n,k}$ and Π_k^* by the map

$$\begin{aligned}
\Phi : \mathcal{W}_p(\mathcal{X}) &\rightarrow L^\infty(\mathbb{R}^d) \times [1, +\infty) \\
\Pi &\mapsto (t \mapsto \int_{\mathbb{R}^d \times [1; +\infty)} \mathbb{1}_{x \leq t} \Pi(dx, dy), \int_{\mathbb{R}^d \times [1; +\infty)} \log(y) \Pi(dx, dy)).
\end{aligned}$$

We prove that Φ is Lipschitz which, together with the Point 2 of Theorem 7.4, implies equation (7.17). Consider Π_1 and $\Pi_2 \in \mathcal{W}_p(\mathcal{X})$ and $((X_1, Y_1), (X_2, Y_2))$ a coupling between Π_1 and Π_2 , we have

$$\begin{aligned} D(\varphi(\Pi_1), \varphi(\Pi_2)) &= \sup_{\mathbb{R}^d} |\mathbb{P}(X_1 \leq x) - \mathbb{P}(X_2 \leq x)| + \mathbb{E} |\log(Y_1) - \log(Y_2)| \\ &\leq \mathbb{P}(X_1 \neq X_2) + \mathbb{E} |\log(Y_1) - \log(Y_2)| \\ &\leq \mathbb{E}(d((X_1, Y_1), (X_2, Y_2))). \end{aligned}$$

Taking the infimum over all coupling we have

$$D(\varphi(\Pi_1), \varphi(\Pi_2)) \leq W_1(\Pi_1, \Pi_2) \leq W_p(\Pi_1, \Pi_2).$$

So Φ is 1-Lipschitz and this entails

$$W_p(P_{\sqrt{k}(\widehat{C}_n(\cdot) - C(\cdot), \widehat{\gamma}_k - \gamma)} | Y_{n-k:n} = U(t), P_{\sqrt{k}(\widehat{C}_k^*(\cdot) - C(\cdot), \widehat{\gamma}_k^* - \gamma)}) \leq \sqrt{k} W_p(P_{\Pi_{n,k} | Y_{n-k:n} = U(t)}, P_{\Pi_k^*})$$

and so goes to 0 when n grows to infinity. \square

Proof of Addendum 7.2. All the notations used in this proof are the same as in the proof of Theorem 7.4. Moreover, consider

$$\widehat{\gamma}_{k,t}^* = \int_1^\infty \log(y) \Pi_{k,t}^*(dy) \quad \text{and} \quad \widehat{C}_{n,t}^*(x) = \int_{\mathbb{R}^d} \mathbf{1}_{u \leq x} \Pi_{k,t}^*(du) \quad x \in \mathbb{R}^d.$$

Since the function Φ defined in the proof of Theorem 7.4 is Lipschitz, by Corollary 6 in [11] we know that

$$W_p(P_{\sqrt{k}(\widehat{C}_n - C, \widehat{\gamma}_k - \gamma)}, P_{\sqrt{k}(\widehat{C}_{n,k}^* - C, \widehat{\gamma}_{k,t}^* - \gamma)}) \leq \sqrt{k} W_p^{(2)}(P_{\Pi_{n,k} | Y_{n-k:n}}, P_{\Pi_{k,t}^*}).$$

By definition of Π_k^* and $\Pi_{k,t}^*$ we have $\widehat{C}_n = \widehat{C}_{n,t}$. Consequently, we prove the theorem if we show the convergence in distribution of

$$\sqrt{k}(\widehat{C}_{n,t}^* - C, \widehat{\gamma}_{k,t}^* - \gamma) = \sqrt{k}(\widehat{C}_n - C, \widehat{\gamma}_k^* - \gamma) + \left(0, \frac{1}{\sqrt{k}} \sum_{i=1}^k \log \left(1 + A_{ext}(1/\bar{F}_{X_i}(U(t))) \frac{Z_i^\rho - 1}{\rho}\right)\right).$$

We know that $\sqrt{k}(\widehat{C}_n^* - C, \widehat{\gamma}_k^* - \gamma)$ converge in distribution to a Gaussian process (\mathbb{B}, N') with \mathbb{B} a C -Brownian bridge and $N' \sim \mathcal{N}(0, \gamma^2)$. We compare the second term with $\sqrt{k} A_{ext}(t) \mathbb{E}(\sigma(X)^{-\rho}) b(\rho)$. Using the triangular inequality and $\log(1+x) - x \leq x^2$ we get

$$\begin{aligned} &\left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \log \left(1 + A_{ext}(1/\bar{F}_{X_i}(U(t))) \frac{Z_i^\rho - 1}{\rho}\right) - \sqrt{k} A_{ext}(t) \mathbb{E}(\sigma(X)^{-\rho}) b(\rho) \right\|_{L^p} \\ &\leq \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \left(A_{ext}(1/\bar{F}_{X_i}(U(t))) \frac{Z_i^\rho - 1}{\rho} \right)^2 \right\|_{L^p} \\ &+ \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k A_{ext}(1/\bar{F}_{X_i}(U(t))) \frac{Z_i^\rho - 1}{\rho} - \sqrt{k} A_{ext}(t) \mathbb{E}(\sigma(X)^{-\rho}) b(\rho) \right\|_{L^p}. \end{aligned}$$

Since σ is bounded, by Lemma 7.7, there exists $M > 0$ such that

$$\begin{aligned} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \left(A_{ext}(1/\bar{F}_{X_i}(U(t))) \frac{Z_i^\rho - 1}{\rho} \right)^2 \right\|_{L^p} &\leq M \sqrt{k} A_{ext}(t)^2 \left\| \frac{1}{k} \sum_{i=1}^k \left(\frac{Z_i^\rho - 1}{\rho} \right)^2 \right\|_{L^p} (1 + o_t(1))^2 \\ &= o(\sqrt{k} A_{ext}(t)). \end{aligned}$$

Where $o_t(1)$ is uniform in $x \in \mathbb{R}^d$. Moreover, by Lemma 7.7, we have

$$\begin{aligned} &\left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k A_{ext}(1/\bar{F}_{X_i}(U(t))) \frac{Z_i^\rho - 1}{\rho} - \sqrt{k} A_{ext}(t) \mathbb{E}(\sigma(X)^{-\rho}) b(\rho) \right\|_{L^p} \\ &= \sqrt{k} A_{ext}(t) \left\| \frac{1}{k} \sum_{i=1}^k \sigma(X_i)^{-\rho} (1 + o_t(1)) \frac{Z_i^\rho - 1}{\rho} - \mathbb{E}(\sigma(X)^{-\rho}) b(\rho) \right\|_{L^p} \end{aligned}$$

Introducing the quantity $\frac{1}{k} \sum_{i=1}^k \sigma(X_i)^{-\rho} b(\rho)$, the triangular inequality entails

$$\begin{aligned} &\left\| \frac{1}{k} \sum_{i=1}^k \sigma(X_i)^{-\rho} (1 + o_t(1)) \frac{Z_i^\rho - 1}{\rho} - \mathbb{E}(\sigma(X)^{-\rho}) b(\rho) \right\|_{L^p} \\ &\leq \left\| \frac{1}{k} \sum_{i=1}^k \sigma(X_i)^{-\rho} \left((1 + o_t(1)) \frac{Z_i^\rho - 1}{\rho} - b(\rho) \right) \right\|_{L^p} + \left\| \frac{1}{k} \sum_{i=1}^k (\sigma(X_i)^{-\rho} - \mathbb{E}(\sigma(X)^{-\rho})) b(\rho) \right\|_{L^p}. \end{aligned}$$

We know by assumption (7.2) that there exists a constant $M > 0$ such that

$$\begin{aligned} \left\| \frac{1}{k} \sum_{i=1}^k \sigma(X_i)^{-\rho} \left((1 + o_t(1)) \frac{Z_i^\rho - 1}{\rho} - b(\rho) \right) \right\|_{L^p} &\leq M \left\| \frac{1}{k} \sum_{i=1}^k \left(\frac{Z_i^\rho - 1}{\rho} - b(\rho) \right) \right\|_{L^p} \\ &\quad + M \left\| \frac{1}{k} \sum_{i=1}^k \frac{Z_i^\rho - 1}{\rho} \right\|_{L^p} o_t(1). \end{aligned}$$

The law of large numbers entails the convergence to zero of this term, as well as the following convergence to 0 :

$$b(\rho) \left\| \frac{1}{k} \sum_{i=1}^k \sigma(X_i)^{-\rho} - \mathbb{E}(\sigma(X)^{-\rho}) \right\|_{L^p} \xrightarrow{k \rightarrow \infty} 0.$$

Collecting the rates of convergence we get,

$$W_p \left(P_{\sqrt{k}(\hat{C}_n - C, \hat{\gamma}_k - \gamma)}, P_B \otimes \mathcal{N}(\sqrt{k} A_{ext}(t) b(\rho), \gamma^2) \right) = o_{k \rightarrow \infty}(\sqrt{k} A_{ext}(t)) + O_{k \rightarrow +\infty} \left(\frac{o_t(1)}{t} \right).$$

Consequently the result follows when $t = n/k$ with k an intermediate sequence such that $\sqrt{k} A_{ext}(k/n) \rightarrow \lambda$. \square

7.3.3 Idea of the proof of Conjecture 7.1

The aim of this section is to prove Conjecture 7.1 and to use it to derive a test procedure for the validation of the proportional tail model. We start writing Δ_k in

terms of empirical process $\mathbb{G}_{n,k} := \sqrt{k}(\Pi_{n,k} - P)$ with $P := \sigma P_X \otimes \mathcal{P}_\alpha$. Following the definition of Δ_k , it is natural to consider $\mathbb{G}_{n,k}$ as a process indexed by the class of functions

$$\mathcal{F} := \{f_u : (x, y) \in \mathbb{R}^d \times [1, \infty) \mapsto \log(y)\mathbb{1}_{(-\infty, u]}(x), u \in \mathbb{R}^d\}$$

and to study the weak convergence of this process in $L^\infty(\mathcal{F})$. The process $\mathbb{G}_{n,k}$ is not centered, this bias is voluntarily introduced in order to make the coupling with the empirical process $\mathbb{G}_k^* := \sqrt{k}(\Pi_k^* - P)$ easier. This time, this process built upon the limit Pareto sample is well centered.

Lemme 7.3. *For any $1 \leq k \leq n$ and $u \in \mathbb{R}^d$, we have*

$$\sqrt{k}\Delta_k(u) = \mathbb{G}_{n,k} \left(g\mathbb{1}_{(-\infty, u]} - \gamma\mathbb{1}_{(-\infty, u]} - C(u)g \right) - \frac{1}{\sqrt{k}}\mathbb{G}_{n,k}(g)\mathbb{G}_{n,k} \left(\mathbb{1}_{(-\infty, u]} \right) \quad (7.18)$$

where $g : (x, y) \in \mathbb{R}^d \times [1, +\infty) \mapsto \log(y)$.

Formally, the indicator in this lemma should be the application $(x, y) \in \mathbb{R}^d \times [1, +\infty) \mapsto \mathbb{1}_{(-\infty, u]}(x)$ but we write it as an indicator of a subset of \mathbb{R}^d for more clarity.

Démonstration. We compute the right hand side of Equation (7.18) for $u \in \mathbb{R}^d$. We start noticing that $\mathbb{G}_{n,k} \left(g\mathbb{1}_{(-\infty, u]} - \gamma\mathbb{1}_{(-\infty, u]} - C(u)g \right)$ worth

$$\frac{1}{\sqrt{k}} \sum_{i=1}^k \log \left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}} \right) \mathbb{1}_{\{X_{(n+1-i)} \leq u\}} - \sqrt{k} \left(\gamma C(u) + \gamma \left(\widehat{C}_n(u) - C(u) \right) + C(u)(\widehat{\gamma}_k - \gamma) \right).$$

Moreover,

$$\sqrt{k}^{-1} \mathbb{G}_{n,k}(g)\mathbb{G}_{n,k} \left(\mathbb{1}_{(-\infty, u]} \right) = \sqrt{k} \left(\widehat{C}_n(u) - C(u) \right) (\widehat{\gamma}_k - \gamma).$$

Now, noticing that

$$\gamma C(u) + \gamma \left(\widehat{C}_n(u) - C(u) \right) + C(u)(\widehat{\gamma}_k - \gamma) + \left(\widehat{C}_n(u) - C(u) \right) (\widehat{\gamma}_k - \gamma) = \widehat{C}_n(u)\widehat{\gamma}_k,$$

the right hand side in (7.18) is

$$\frac{1}{\sqrt{k}} \sum_{i=1}^k \log \left(\frac{Y_{n+1-i:n}}{Y_{n-k:n}} \right) \mathbb{1}_{\{X_{(n+1-i)} \leq u\}} - \sqrt{k}\widehat{C}_n(u)\widehat{\gamma}_k = \sqrt{k}\Delta_k(u).$$

□

Lemme 7.4. *For every sequence $k := k(n)$ such that $1 \leq k \leq n$ and $k \rightarrow \infty$,*

$$\mathbb{G}_k^* \xrightarrow[k \rightarrow \infty]{d} \mathbb{G}$$

with \mathbb{G} the centered Gaussian process in $L^\infty(\mathcal{F})$ with covariance function

$$\text{cov}(\mathbb{G}(f_1), \mathbb{G}(f_2)) = P(f_1 f_2) - P(f_1)P(f_2), \quad f_1, f_2 \in \mathcal{F}.$$

Démonstration. Following [82], this result is proved showing that \mathcal{F} is a Donsker class. Consequently, it is enough to hold that \mathcal{F} is a VC-subgraph class of functions with square integrable envelop (see e.g. [82] for more details). The class \mathcal{F} can be rewrite as $\{fg, g \in \mathcal{G}\}$ with

$$\mathcal{G} := \{g_u : (x, y) \in \mathbb{R}^d \times [1, \infty) \mapsto \mathbf{1}_{(-\infty, u]}(x), u \in \mathbb{R}^d\}$$

which is a VC-subgraph class as a collection of indicators of cells of \mathbb{R}^d (Example 2.6.1 in [82]). Consequently, \mathcal{F} is a VC-subgraph in virtue of Lemma 2.6.18 in [82]. Moreover, for every $f \in \mathcal{F}$,

$$f(x, y) \leq \log(y), \quad x \in \mathbb{R}^d, y > 1,$$

and

$$\int_{\mathbb{R}^D \times [1, \infty)} \log dP = \int_1^\infty \log^2(y) \frac{dy}{y^{1/\gamma}} = 2\gamma^2 < \infty.$$

□

The last Lemma needed to achieve the proof of Conjecture 7.1 is a coupling result between $\mathbb{G}_{n,k}$ and \mathbb{G}_k^* which allows us to say that $\mathbb{G}_{n,k}$ and \mathbb{G}_k^* have the same weak limit.

Since $\mathbb{G}_{n,k}$ and \mathbb{G}_k^* are seen as elements of $L^\infty(\mathcal{F})$ the coupling construction involves the consideration of the Wassserstein distance over $L^\infty(\mathcal{F})$ endowed by the metric

$$\|G_1 - G_2\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |G_1(f) - G_2(f)|, \quad G_1, G_2 \in L^\infty(\mathcal{F}).$$

Such a supremum is not measurable in general, but it is the case for $\|\mathbb{G}_{n,k} - \mathbb{G}_k^*\|_{\mathcal{F}}$. Since the two empirical processes are centered by the same probability measure, this is a consequence of the measurability of

$$\sup_{f \in \mathcal{F}} |(\Pi_{n,k} - \Pi_k^*)(f)|.$$

The measurability of this supremum can be obtained as a particular case of a property of the class \mathcal{F} . We say that a class \mathcal{F} is measurable if

$$((X_1, Y_1), \dots, (X_n, Y_n)) \mapsto \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n e_i f(X_i, Y_i) \right|$$

is measurable (on a completion of the underlying probability space) for all n and all $(e_1, \dots, e_n) \in \mathbb{R}^n$. Indeed, considering the countable class of functions

$$\mathcal{Q} := \{g_u : (x, y) \in \mathbb{R}^d \times [1, \infty) \mapsto \log(y) \mathbf{1}_{(-\infty, u]}(x), u \in \mathbb{Q}^d\},$$

the class \mathcal{F} is measurable since each function $f \in \mathcal{F}$ there exists a sequence $g_{u_m} \in \mathcal{Q}$ such that $g_{u_m}(x, y) \xrightarrow{m} f(x, y)$ for all $(x, y) \in \mathbb{R}^d \times [1, \infty)$ (Example 2.3.4 in [82]). A such sequence $(g_{u_m})_{m \geq 1}$ is built with a sequence $(u_m)_{m \geq 1} \subset \mathbb{Q}^d$ converging to u by density, so this sequence do not depend on (x, y) . And the point-wise convergence of $(g_{u_m})_{m \geq 1}$ holds since all function in \mathcal{F} are component-wise càd-làg (this particular type of measurable classes are called point-wise measurable classes). Consequently, taking the definition of measurable class with all $(e_i)_{i=1}^n$ equal to $1/n$ for every n , we get the measurability of $\|\mathbb{G}_{n,k} - \mathbb{G}_k^*\|_{\mathcal{F}}$.

Lemma 7.5 (Lemma to be corrected). *For $p \in [1, \infty)$ and for every $1 \leq k \leq n$ we have*

$$\mathbb{E}(\|\mathbb{G}_{n,k} - \mathbb{G}_k^*\|_{\mathcal{F}}) \leq \sqrt{k}W_p^{(2)}(P_{\Pi_{n,k}}, P_{\Pi_k^*}).$$

Since \mathcal{F} is not only constituted of 1-Lipschitz functions, the proof of this result is not as direct as we expected. Actually, this result may not be true stated in that way, but a similar is certainly true if we replace \mathcal{F} by the set

$$\mathcal{F}_M := \{f_u : (x, y) \in \mathbb{R}^d \times [1, \infty) \mapsto \min(\log(y), M)\mathbf{1}_{(-\infty, u]}(x), u \in \mathbb{R}^d, \},$$

with $M > 0$. The following sketch of proof present how to combine those lemmas to prove the Conjecture 7.1. Of course, this proof should include a corrected version of Lemma 7.5.

Sketch of proof of Conjecture 7.1. The expected coupling Lemma 7.5 gives us that $\mathbb{G}_{n,k}$ and \mathbb{G}_k^* have the same weak limit.

Consequently, writing $\mathbb{G}_{n,k} = \mathbb{G}_{n,k} - \mathbb{G}_k^* + \mathbb{G}_k^*$, Lemma 7.4 and Slutsky's lemma together imply

$$\mathbb{G}_{n,k} \xrightarrow[k \rightarrow \infty]{d} \mathbb{G}, \quad \text{in } L^\infty(\mathcal{F}).$$

Consequently we get for $u \in \mathbb{R}^d$

$$\frac{1}{\sqrt{k}}\mathbb{G}(g)\mathbb{G}(\mathbf{1}_{(-\infty, u]}) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).$$

Introducing all those convergences in the expression of Δ_k given by Lemma 7.3, we have,

$$\left(\sqrt{k}\Delta_k(u)\right)_{u \in \mathbb{R}^d} \xrightarrow[k \rightarrow \infty]{d} \left(\mathbb{G}\left(g\mathbf{1}_{(-\infty, u]} - \gamma\mathbf{1}_{(-\infty, u]} - C(u)g\right)\right)_{u \in \mathbb{R}^d}. \quad (7.19)$$

Then we can identify the limit computing the covariance function. For fixed $u_1, u_2 \in \mathbb{R}^d$, consider the covariance

$$\begin{aligned} & cov(g(Y^*)\mathbf{1}_{(-\infty, u_1]}(X^*) - \gamma\mathbf{1}_{(-\infty, u_1]}(X^*) - C(u_1)g(Y^*), \\ & g(Y^*)\mathbf{1}_{(-\infty, u_2]}(X^*) - \gamma\mathbf{1}_{(-\infty, u_2]}(X^*) - C(u_2)g(Y^*)) \end{aligned}$$

with $(X^*, Y^*) \sim \sigma P_X \otimes \mathcal{P}_{1/\gamma}$. Seeing that X^* and Y^* are independents, for $u \in \mathbb{R}^d$ $\mathbb{E}(\mathbf{1}_{(-\infty, u]}(X^*)) = C(u)$, $\mathbb{E}(\log(Y^*)) = \gamma$ and that $\mathbb{E}(\log^2(Y^*)) = 2\gamma^2$ we easily deduce the following expression of (7.19) :

$$\gamma(C(u_1 \wedge u_2) - C(u_1)C(u_2)),$$

which is the covariance function of $(\gamma B(u))_{u \in \mathbb{R}^d}$ with B the C -Brownian bridge. Moreover, exactly under the assumptions of this Theorem 7.1, $\hat{\gamma}_k$ is a consistant estimator of $\gamma > 0$. This leads, by Slutsky's lemma,

$$\left(\sqrt{k}\frac{\Delta_k(u)}{\hat{\gamma}_k}\right)_{u \in \mathbb{R}^d} \xrightarrow[k \rightarrow \infty]{d} (B(u))_{u \in \mathbb{R}^d}.$$

□

7.3.4 Bootstrap and proof of Theorem 7.3

The aim of this section is to study the bootstrap procedure with respect to the Wasserstein distance in order to have asymptotic normality of bootstrap version of estimators of γ and C . The bootstrap empirical process is defined as

$$\mathbb{G}_{\xi,n} := \sqrt{k} (\Pi_{\xi,n,k} - \Pi_{n,k})$$

where the $\Pi_{\xi,n,k}$ is the bootstrap empirical measure

$$\Pi_{\xi,n,k} = \frac{1}{k} \sum_{i=1}^k \xi_{n,i} \varepsilon_{(X_{(n+1-i)}, \frac{Y_{n+1-i:n}}{Y_{n-k:n}})}.$$

The assumption about the empirical mean of the weights entails that this bootstrap empirical measure is actually an probability measure.

$$\Pi_{\xi,k}^* = \frac{1}{k} \sum_{i=1}^k \xi_{n,i} \varepsilon_{(X_i^*, Y_i^*)}$$

with $(X_i^*, Y_i^*)_{1 \leq i \leq k}$ a sample drawn with the distribution of the limit Pareto model. That is, with X_1^*, \dots, X_k^* i.i.d with density $\sigma(x)P_X(dx)$ and, independently, Y_1^*, \dots, Y_k^* i.i.d with Pareto distribution P_α .

Proposition 7.2. *Assume F is continuous and $p \in [1, \infty)$,*

$$\sup_{h \in BL_1} |\mathbb{E}_\xi(h(\Pi_{\xi,n,k} - \Pi_{n,k})) - \mathbb{E}_\xi(h(\Pi_{\xi,k}^* - \Pi_k^*))| = O_{\mathbb{P}}(A(n/k)).$$

Proof of Proposition 7.2 . The Kantorovich duality (Theorem 4.4) leads to

$$\sup_{h \in BL_1} |\mathbb{E}_\xi(h(\Pi_{\xi,n,k} - \Pi_{n,k})) - \mathbb{E}_\xi(h(\Pi_{\xi,k}^* - \Pi_k^*))| \leq W_1^{(2)} \left(P_{\Pi_{\xi,n,k} - \Pi_{n,k} | (\mathfrak{D}_{n,k}, Y_{n-k:n})}, P_{\Pi_{\xi,k}^* - \Pi_k^* | \mathfrak{D}_{n,k}} \right)$$

where $\mathfrak{D}_{n,k}$ denote the sample $(X_{(n+1-i)}, Y_{n+1-i:n}/Y_{n-k:n}, X_i^*, Y_i^*)_{1 \leq i \leq k}$ for more readability. Since W_1 is the more stringent of all Wasserstein distances, we can replace $W_1^{(2)}$ for any distance $W_p^{(2)}$ with $p > 1$ in the previous inequality. Moreover, conditionally to $\mathfrak{D}_{n,k}$ the measures $\Pi_{n,k}$ and Π_k^* are constant elements of $\mathcal{W}_p(\mathbb{R}^d \times [1, \infty)$. Consequently we have

$$W_p^{(2)} \left(P_{\Pi_{\xi,n,k} - \Pi_{n,k} | (\mathfrak{D}_{n,k}, Y_{n-k:n})}, P_{\Pi_{\xi,k}^* - \Pi_k^* | \mathfrak{D}_{n,k}} \right) \leq W_p^{(2)} \left(P_{\Pi_{\xi,n,k} | (\mathfrak{D}_{n,k}, Y_{n-k:n})}, P_{\Pi_{\xi,k}^* | \mathfrak{D}_{n,k}} \right) + W_p(\Pi_{n,k}, \Pi_k^*).$$

We know by Lemma 7.2 that

$$W_p^{(2)} \left(P_{\Pi_{\xi,n,k} | (\mathfrak{D}_{n,k}, Y_{n-k:n})}, P_{\Pi_{\xi,k}^* | \mathfrak{D}_{n,k}} \right) = W_p(\Pi_{n,k}, \Pi_k^*).$$

Taking the expectation with respect to $\mathfrak{D}_{n,k}$ we get

$$W_p^{(2)} \left(P_{\Pi_{\xi,n,k} - \Pi_{n,k} | Y_{n-k:n}}, P_{\Pi_{\xi,k}^* - \Pi_k^*} \right) \leq W_p^{(2)} \left(P_{\Pi_{n,k} | Y_{n-k:n}}, P_{\Pi_k^*} \right).$$

Point 3) of Theorem 7.4 entails

$$W_p^{(2)} \left(P_{\Pi_{\xi,n,k} - \Pi_{n,k} | Y_{n-k:n}}, P_{\Pi_{\xi,k}^* - \Pi_k^*} \right) = O_{\mathbb{P}} \left(A_{ext} \left(\frac{1}{\bar{F}(Y_{n-k:n})} \right) \right).$$

The result follows by taking the expectation with respect to $Y_{n-k:n}$. \square

Proof of Theorem 7.3. Considering that $\sqrt{k}(\widehat{C}_{\xi,n}(\cdot) - \widehat{C}_n(\cdot), \widehat{\gamma}_{\xi,n} - \widehat{\gamma}_n)$ is the image of $\mathbb{G}_{\xi,n}$ by the 1-Lipschitz map Φ defined in the proof of Theorem 7.1 by

$$\begin{aligned} \Phi : \mathcal{W}_p(\mathcal{X}) &\rightarrow L^\infty(\mathbb{R}^d) \times [1, +\infty) \\ \Pi &\mapsto (t \mapsto \int_{\mathbb{R}^d \times [1; +\infty)} \mathbb{1}_{x \leq t} \Pi(dx, dy), \int_{\mathbb{R}^d \times [1; +\infty)} \log(y) \Pi(dx, dy)), \end{aligned}$$

this proof is done by showing $\mathbb{G}_{\xi,n} \xrightarrow[\xi]{\sim} c \mathbb{G}$ with \mathbb{G} the Borel tight limit of \mathbb{G}_n . Denoted the bootstrap empirical process upon the limit Pareto model by

$$\mathbb{G}_{\xi,n}^* := \sqrt{k} \left(\Pi_{\xi,k}^* - \Pi_k^* \right),$$

we have by Proposition 7.2

$$\begin{aligned} \sup_{h \in BL_1} |\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n})) - \mathbb{E}(c \mathbb{G})| &\leq \sup_{h \in BL_1} |\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n})) - \mathbb{E}(\mathbb{G}_{\xi,n}^*)| \\ &\quad - \sup_{h \in BL_1} |\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}^*)) - \mathbb{E}(c \mathbb{G})| \\ &\leq \sqrt{k} A_{ext} \left(\frac{1}{\overline{F}(Y_{n-k:n})} \right) - \sup_{h \in BL_1} |\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}^*)) - \mathbb{E}(c \mathbb{G})|. \end{aligned}$$

Moreover, for all $h \in BL_1$, $\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}))^* - \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}))_*$ is equal to

$$\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}))^* - \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}^*))^* + \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}^*))_* - \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}))_* + \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}^*))^* - \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}^*))_*.$$

The two first differences are both converging to 0 in probability. As a consequence of Lemma 1.2.2 in [82] we have

$$\begin{aligned} |\mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}))^* - \mathbb{E}_\xi(h(\mathbb{G}_{\xi,n}^*))^*| &\leq \mathbb{E}_\xi \left(|h(\mathbb{G}_{\xi,n}) - h(\mathbb{G}_{\xi,n}^*)| \right)^* \\ &\leq \mathbb{E}_\xi \left(\sqrt{k} W_p \left(\Pi_{\xi,n,k} - \Pi_{n,k}, \Pi_{\xi,k}^* - \Pi_k^* \right) \right)^*. \end{aligned}$$

This upper bound converges to zero to infinity by Proposition 7.2. The case of the second difference is straight forward remarking that $\mathbb{E}_\xi(T)_* = -\mathbb{E}_\xi(-T)^*$ for any random element T . After those considerations it appears that, if $\mathbb{G}_{\xi,n}^* \xrightarrow[\xi]{\sim} c \mathbb{G}$ then $\mathbb{G}_{\xi,n} \xrightarrow[\xi]{\sim} c \mathbb{G}$. Theorem 3.6.13 of [82] entails the result since the two classes of functions

$$\mathcal{F} := \{f_{x,y} : (u, v) \mapsto \mathbb{1}_{\{u \leq x\}} \mathbb{1}_{\{v > y\}}, x \in \mathbb{R}^d, y \in [1, \infty)\}$$

and

$$\mathcal{F}_\delta := \{g - f, f, g \in \mathcal{F}, \sigma P_X \otimes P_\alpha(f - g) - \sigma P_X \otimes P_\alpha(f - g)\|^2 < \delta^2\}$$

are respectively Donsker and measurable for every $\delta > 0$. We have seen that \mathcal{F} is a Donsker class in the proof of Theorem 7.1. The measurability of \mathcal{F}_δ is a consequence of a classical results which mention that the classes constituted with indicators of intervals or orthants are always measurables. □

7.4 Appendix

7.4.1 About proportional tail model

Lemme 7.6. *Assuming U continuous,*

1. *Under assumptions (7.1), (7.2) and (7.3) we have, as $t \rightarrow \infty$*

$$\frac{U_x(z/\bar{F}_x(U(t)))}{U(t)} \sim \frac{U(z/\bar{F}_x(U(t)))}{U(1/\bar{F}_x(U(t)))}, \quad x \in \mathbb{R}^d, z > 1.$$

2. *If moreover we assume (7.5) with $A_{prop} = o(A_{ext})$, we have*

$$\frac{U_x(z/\bar{F}_x(U(t)))}{U(t)} = \frac{U(z/\bar{F}_x(U(t)))}{U(1/\bar{F}_x(U(t)))} \left(1 + O_{t \rightarrow \infty} \left(A_{ext} \left(\frac{1}{\bar{F}_x(U(t))} \right) \right) \right), \quad x \in \mathbb{R}^d.$$

Démonstration. For the sake of clarity, we write $\varphi_t = \bar{F}_x(U(t))$ during this proof. This proof is based on the following decomposition

$$\frac{U_x(z/\bar{F}_x(U(t)))}{U(t)} = \frac{\sigma(x)^\gamma U(1/\bar{F}_x(U(t)))}{U_x(1/\bar{F}_x(U(t)))} \frac{U_x(z/\bar{F}_x(U(t)))}{\sigma(x)^\gamma U(z/\bar{F}_x(U(t)))} \frac{U(z/\bar{F}_x(U(t)))}{U(1/\bar{F}_x(U(t)))}. \quad (7.20)$$

This equality hold since U is continuous. Consequently, it is the same for U_x for each $x \in \mathbb{R}^d$, and we can write $U(t) = U_x(1/\bar{F}_x(U(t)))$.

1. We can bound first the two first ratios in the right hand side. Let $\varepsilon > 0$, using the proportional tail assumption (7.1) we get the existence of $y_0 > 1$ such that, for $y \geq y_0$ and every $x \in \mathbb{R}^d$,

$$U((1 - \varepsilon)\sigma(x)y) \leq U_x(y) \leq U((1 + \varepsilon)\sigma(x)y).$$

Moreover, since U satisfies the first order extreme value condition, U is regularly varying with index γ . Consequently we can apply the Potter's bounds (p.367 in [22]), which entail for all $\delta > 0$ the existence $y_1 > 1$ such that,

$$(1 - \varepsilon)\lambda^{\gamma - \delta} U(y) \leq U(\lambda y) \leq (1 + \varepsilon)\lambda^{\gamma + \delta} U(y), \quad \text{for all } y > y_1, \lambda y > y_1.$$

Consider y such that $y > \max(y_0, y_1)$ and $(1 - \varepsilon)\sigma(x)y > y_1$. Combining the previous inequalities, we get,

$$(1 - \varepsilon)^{\gamma - \delta + 1} \sigma(x)^{-\delta} \leq \frac{U_x(y)}{\sigma(x)^\gamma U(y)} \leq (1 + \varepsilon)^{\gamma + \delta + 1} \sigma(x)^\delta.$$

Since $\lim_{t \rightarrow \infty} 1/\bar{F}_x(U(t)) = +\infty$, we can set $y = 1/\bar{F}_x(U(t))$ in those inequalities, for t large enough,

$$(1 - \varepsilon)^{\gamma - \delta + 1} \sigma(x)^{-\delta} \leq \frac{U_x(1/\bar{F}_x(U(t)))}{\sigma(x)^\gamma U(1/\bar{F}_x(U(t)))} \leq (1 + \varepsilon)^{\gamma + \delta + 1} \sigma(x)^\delta.$$

Moreover, for $z \geq 1$, the same bounds hold replacing y by $z/\bar{F}_x(U(t))$ and conserving the same t_0 . Those specified bounds together with the decomposition (7.20) give

$$\begin{aligned} \frac{U(z/\bar{F}_x(U(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} \frac{(1-\varepsilon)^{\gamma-\delta+1} \sigma(x)^{-2\delta}}{(1+\varepsilon)^{\gamma+\delta+1}} &\leq \frac{U_x(z/\bar{F}_x(U(t)))}{z^\gamma U(t)} \\ &\leq \frac{U(z/\bar{F}_x(U(t)))}{z^\gamma U(1/\bar{F}_x(U(t)))} \frac{(1+\varepsilon)^{\gamma+\delta+1} \sigma(x)^{2\delta}}{(1-\varepsilon)^{\gamma-\delta+1}}. \end{aligned}$$

Together with assumption (7.2), the Point 1 follows letting $\varepsilon \rightarrow 0$.

2. Now, condition (7.1) taken with $y = z/\varphi_t$, entails the existence of a constant $M > 0$ such that,

$$\begin{aligned} &\left(\sigma(x) \bar{F}(z/\varphi_t) \left(1 + MA_{prop} \left(\frac{z}{\varphi_t} \right) \right) \right)^{-1} \\ &\leq \frac{1}{\bar{F}_x(z/\varphi_t)} \leq \left(\sigma(x) \bar{F}(z/\varphi_t) \left(1 - MA_{prop} \left(\frac{z}{\varphi_t} \right) \right) \right)^{-1}. \end{aligned}$$

Since $z \geq 1$, as A is decreasing at infinity, we have $A(z/\varphi_t) \leq A(1/\varphi_t)$. Consequently, we have

$$\begin{aligned} &\left(\sigma(x) \bar{F}(z/\varphi_t) \left(1 + MA_{prop} \left(\frac{1}{\varphi_t} \right) \right) \right)^{-1} \\ &\leq \frac{1}{\bar{F}_x(z/\varphi_t)} \leq \left(\sigma(x) \bar{F}(z/\varphi_t) \left(1 - MA_{prop} \left(\frac{1}{\varphi_t} \right) \right) \right)^{-1}. \end{aligned}$$

Taking the inverse in those inequalities we have,

$$\begin{aligned} U \left(\frac{z}{\varphi_t} \left(1 - MA_{prop} \left(\frac{1}{\varphi_t} \right) \right) \sigma(x) \right) &\leq U_x \left(\frac{z}{\varphi_t} \right) \\ &\leq U \left(\frac{z}{\varphi_t} \left(1 + MA_{prop} \left(\frac{1}{\varphi_t} \right) \right) \sigma(x) \right). \end{aligned}$$

Furthermore we can deduce from second order assumption, for t large enough, there exists a constant $M' > 0$ such that,

$$(1 - M' A_{ext}(t)) x^\gamma U(t) \leq U(tz) \leq (1 + M' A_{ext}(t)) x^\gamma U(t).$$

We can choose $M = M'$ without loss of generality. After what, we can deduce,

$$\begin{aligned} &\left(1 - MA_{ext} \left(\frac{z}{\varphi_t} \right) \right) \left(1 - MA_{prop} \left(\frac{1}{\varphi_t} \right) \right)^\gamma \sigma(x)^\gamma U(z/\varphi_t) \leq U_x(z/\varphi_t) \\ &\leq \left(1 + MA_{ext} \left(\frac{z}{\varphi_t} \right) \right) \left(1 + MA_{prop} \left(\frac{1}{\varphi_t} \right) \right)^\gamma \sigma(x)^\gamma U(z/\varphi_t) \end{aligned}$$

This leads, since A_{ext} is decreasing, to

$$\begin{aligned} \left(1 - MA_{ext} \left(\frac{1}{\varphi_t}\right)\right) \left(1 - MA_{prop} \left(\frac{1}{\varphi_t}\right)\right)^\gamma &\leq \frac{U_x(z/\varphi_t)}{\sigma(x)^\gamma U(z/\varphi_t)} \\ &\leq \left(1 + MA_{ext} \left(\frac{1}{\varphi_t}\right)\right) \left(1 + MA_{prop} \left(\frac{1}{\varphi_t}\right)\right)^\gamma \end{aligned}$$

Introducing those bounds in (7.20), we get

$$\begin{aligned} \frac{U(z/\varphi_t) \left(1 - MA_{ext} \left(\frac{1}{\varphi_t}\right)\right) \left(1 - MA_{prop} \left(\frac{1}{\varphi_t}\right)\right)^\gamma}{U(1/\varphi_t) \left(1 + MA_{ext} \left(\frac{1}{\varphi_t}\right)\right) \left(1 + MA_{prop} \left(\frac{1}{\varphi_t}\right)\right)^\gamma} &\leq \frac{U_x(z/U(z/\varphi_t))}{U(t)} \\ &\leq \frac{U(z/\varphi_t) \left(1 + MA_{ext} \left(\frac{1}{\varphi_t}\right)\right) \left(1 + MA_{prop} \left(\frac{1}{\varphi_t}\right)\right)^\gamma}{U(1/\varphi_t) \left(1 - MA_{ext} \left(\frac{1}{\varphi_t}\right)\right) \left(1 - MA_{prop} \left(\frac{1}{\varphi_t}\right)\right)^\gamma}. \end{aligned}$$

Since $A_{prop} = o(A_{ext})$, those inequalities together with a Taylor expansion entail

$$\frac{U_x(z/U(z/\varphi_t))}{U(t)} = \frac{U(z/\varphi_t)}{U(1/\varphi_t)} \left(1 + O\left(A_{ext} \left(\frac{1}{\varphi_t}\right)\right)\right).$$

□

Lemme 7.7. *Under assumptions (7.1), (7.2) and (7.5) with U continuous, uniformly in $x \in \mathbb{R}^d$, we have*

$$A_{ext} \left(\frac{1}{\bar{F}_x(U(t))}\right) \sim \sigma(x)^{-\rho} A_{ext}(t) \quad , t \rightarrow \infty.$$

Démonstration. The main assumption (7.1) of proportional tail model entails $\bar{F}_x(U(t)) \sim \sigma(x)\bar{F}(U(t)) = \sigma(x)/t$. Moreover, Theorem 2.3.3 in [22] implies that A_{ext} is regularly varying with index ρ if (7.5) hold. Consequently, Theorem B.1.4 in [22] entails, recalling that σ is bounded away from zero at infinity,

$$\lim_{t \rightarrow \infty} \frac{A_{ext}(t/\sigma(x))}{\sigma(x)^{-\rho} A_{ext}(t)} = 1, \quad \text{uniformly in } x \in \mathbb{R}^d.$$

These two considerations yield the result. □

7.4.2 About maximal coupling

The following lemma show that Equation (5.7) keep valid in the framework of optimal coupling. Moreover, we get an additional precision in the fact that the bound can be established with the function A but also with A_{prop} . This precision will be necessary in the biased case.

Lemme 7.8. *Under Assumption (7.1) we have*

$$\|P_{X|Y>y}(dx) - \sigma(x)P_X(dx)\|_{TV} = O(A_{prop}(y)), \quad \text{as } y \rightarrow \infty.$$

Proof of Lemma 7.8. For $A \in \mathcal{B}(\mathbb{R}^d)$, the conditional probability can be written as

$$P_{X|Y>y}(A) = \frac{\mathbb{P}(X \in A, Y > y)}{\mathbb{P}(Y > y)} = \frac{1}{\bar{F}(y)} \int_A \bar{F}_x(y) P_X(dx).$$

Since σ is bounded, Assumption (7.1) entails that there exist, $M > 0$ such that

$$\left| \frac{\bar{F}_x(y)}{\bar{F}(y)} - \sigma(x) \right| \leq MA_{prop}(y), \quad y > 1, x \in \mathbb{R}^d.$$

Hence, we have

$$\begin{aligned} |P(X \in A|Y > y) - \int_A \sigma(x) P_X(dx)| &= \left| \int_A \frac{\bar{F}_x(y) P_X(dx)}{\bar{F}(y)} - \int_A \sigma(x) P_X(dx) \right| \\ &\leq \int_{\mathcal{X}} \left| \frac{\bar{F}_x(y)}{\bar{F}(y)} - \sigma(x) \right| P_X(dx) \\ &= MA_{prop}(y). \end{aligned}$$

Taking the supremum over $A \in \mathcal{B}(\mathbb{R}^d)$, we have, by definition of the total variation distance

$$\|\mathcal{L}(X|Y > y) - \sigma P_X\|_{TV} \leq MA_{prop}(y)$$

which concludes the proof. □

Perspectives

Cette section présente les extensions prévues de ces travaux.

- La première extension sera traitée dans un avenir proche, car elle constitue la suite du dernier chapitre tel qu'il sera soumis sous forme d'article. Nous souhaitons comprendre le comportement du modèle des queues proportionnelles lorsqu'une des coordonnées de X n'a pas d'effet sur les extrêmes. Et ainsi fournir un test permettant de vérifier si une coordonnée a en effet un impact sur les extrêmes.
- Notre estimateur plug-in des quantiles a été étudié uniquement dans le cas où le seuil est déterministe, cependant il est plus classique de considérer $Y_{n-k:k}$. Il serait donc bienvenu d'étudier notre estimateur avec un tel seuil.
- Le troisième point est nécessaire à la réussite du second et fait écho au premier. Notre estimateur du quantile requiert l'estimation de σ , ce qui nous pousse alors naturellement vers des problématiques d'estimation de densité. Notre estimateur présente deux principales lacunes que l'on pourrait chercher à combler. La première est que, comme l'estimateur du quantile, l'estimateur à noyau proposé est construit à partir d'un seuil déterministe et il faudrait donc étudier le seuil aléatoire qui est souvent préféré. L'autre problématique est celle de la dimension. En effet, notre estimateur à noyau souffre du fléau de la dimension qui fait chuter la qualité de l'estimation lorsque la dimension de la covariable est grande. Ceci pose des problèmes importants en fonction du domaine d'application. Une solution possible serait d'utiliser des méthodes liées aux forêts aléatoires [6]. Un objectif théorique est de donner des résultats de convergences pour un estimateur construit avec des forêts particulières, comme les forêts honnêtes [84], [85] ou les forêts uniformes [16], qui sont plus simples à étudier que les forêts originales de Breiman [12]. Un estimateur de la densité basé sur ces dernières pourrait tout de même être étudié et comparé aux précédents par simulation.
- Une fois un tel estimateur mis au point, il serait alors intéressant de le comparer aux estimateurs existants et de le mettre à l'épreuve sur des jeux de données réelles.

Bibliographie

- [1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [2] P. Appell. *Le problème géométrique des déblais et remblais*. Number vol. 27 in Le problème géométrique des déblais et remblais. Gauthier-Villars et cie, 1928.
- [3] A. A. Balkema and L. de Haan. Residual life time at great age. *Ann. Probability*, 2 :792–804, 1974.
- [4] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2004. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.
- [5] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2004. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.
- [6] G. Biau and E. Scornet. A random forest guided tour. *TEST*, 25(2) :197–227, 2016.
- [7] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6) :1196–1217, 1981.
- [8] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
- [9] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

- [10] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.
- [11] B. Bobbia, C. Dombry, and D. Varron. The coupling method in extreme value theory, 2019.
- [12] L. Breiman. Random forests. *Machine Learning*, 45 :5–32, 2001.
- [13] A. Bücher and J. Segers. On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, 20(4) :839–872, 2017.
- [14] A. Bücher and J. Segers. Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2) :1427–1462, 2018.
- [15] V. Chernozhukov. Extremal quantile regression. *Ann. Statist.*, 33(2) :806–839, 2005.
- [16] S. Ciss. Random Uniform Forests. working paper or preprint, January 2015.
- [17] S. Coles. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2001.
- [18] A. Daouia, L. Gardes, and S. Girard. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B) :2557–2589, 2013.
- [19] H. A. David and H. N. Nagaraja. *Order statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.
- [20] L. de Haan. A form of regular variation and its application to the domain of attraction of the double exponential distribution. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 17 :241–258, 1971.
- [21] L. de Haan. Fighting the arch-enemy with mathematics. *Statist. Neerlandica*, 44(2) :45–68, 1990.
- [22] L. de Haan and A. Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.
- [23] L. de Haan and U. Stadtmüller. Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A*, 61(3) :381–395, 1996.
- [24] C. de Valk and J. Segers. Tails of optimal transport plans for regularly varying probability measures. arXiv :1811.12061, 2020.
- [25] A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, 17(4) :1833–1855, 1989.

- [26] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3) :458–486, 1970.
- [27] C. Dombry and A. Ferreira. Maximum likelihood estimators based on the block maxima method. *Bernoulli*, 25(3) :1690–1723, 2019.
- [28] M. D. Donsker. Justification and extension of Doob’s heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, 23 :277–281, 1952.
- [29] J. L. Doob. Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics*, 20 :393–403, 1949.
- [30] H. Drees. On smooth statistical tail functionals. *Scand. J. Statist.*, 25(1) :187–210, 1998.
- [31] H. Drees, L. de Haan, and D. Li. On large deviation for extremes. *Statist. Probab. Lett.*, 64(1) :51–62, 2003.
- [32] H. Drees, L. de Haan, and D. Li. Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statist. Plann. Inference*, 136(10) :3498–3538, 2006.
- [33] H. Drees and H. Rootzén. Limit theorems for empirical processes of cluster functionals. *Ann. Statist.*, 38(4) :2145–2186, 2010.
- [34] B. Efron. Bootstrap methods : another look at the jackknife. *Ann. Statist.*, 7(1) :1–26, 1979.
- [35] J. H. J. Einmahl, L. de Haan, and C. Zhou. Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(1) :31–51, 2016.
- [36] J. H. J. Einmahl and J. R. Magnus. Records in athletics through extreme-value theory. *J. Amer. Statist. Assoc.*, 103(484) :1382–1391, 2008.
- [37] J. J. Einmahl, Einmahl J. H. J., and L. de Haan. Limits to human life span through extreme value theory. *Journal of the American Statistical Association*, 114(527) :1075–1080, 2019.
- [38] J. El Methni, L. Gardes, S. Girard, and A. Guillou. Estimation of extreme quantiles from heavy and light tailed distributions. *J. Statist. Plann. Inference*, 142(10) :2735–2747, 2012.
- [39] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.
- [40] A. Ferreira and L. de Haan. On the block maxima method in extreme value theory : PWM estimators. *Ann. Statist.*, 43(1) :276–298, 2015.
- [41] R.A Fisher and L.H.C. Tipett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Philos. Soc.*, 24 :180–190, 1928.

- [42] M. I. Fraga Alves, M. I. Gomes, L. de Haan, and C. Neves. A note on second order conditions in extreme value theory : linking general and heavy tail conditions. *REVSTAT*, 5(3) :285–304, 2007.
- [43] L. Gardes. A general estimator for the extreme value index : applications to conditional and heteroscedastic extremes. *Extremes*, 18(3) :479–510, 2015.
- [44] L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions : An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2) :177–204, June 2010.
- [45] B. Gnedenko. Sur la distribution limite du terme maximum d’une série aléatoire. *Ann. of Math. (2)*, 44 :423–453, 1943.
- [46] J. Greenwood, J. Landwehr, NC Matalas, and J.R.M. Wallis. Probability weighted moments : Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, pages 1049–1054, 05 1979.
- [47] A. Guillou and P. Willems. Application de la théorie des valeurs extrêmes en hydrologie. *Revue de Statistique Appliquée*, pages 5–31, 2006.
- [48] E. J. Gumbel. *Statistical theory of extreme values and some practical applications. A series of lectures*. National Bureau of Standards Applied Mathematics Series No. 33. U. S. Government Printing Office, Washington, D. C., 1954.
- [49] E. J. Gumbel. *Statistics of extremes*. Columbia University Press, New York, 1958.
- [50] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5) :1163–1174, 1975.
- [51] J. Hoffmann-Jørgensen. *Stochastic processes on Polish spaces*, volume 39 of *Various Publications Series (Aarhus)*. Aarhus Universitet, Matematisk Institut, Aarhus, 1991.
- [52] J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3) :339–349, 1987.
- [53] J. R. M. Hosking, J. R. Wallis, and E. F. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3) :251–261, 1985.
- [54] J.R.M Hosking. The theory of probability weighted moments. *Research Report RC12210, IBM Thomas J. Watson Research center, Yorktown Heights, NY*, 1986.
- [55] T. Hsing. On tail index estimation using dependent data. *Ann. Statist.*, 19(3) :1547–1569, 1991.

- [56] A. F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348) :158–171, 1955.
- [57] C Jones. Simple boundary correction for kernel estimation. *Statistics and Computing*, 3(3) :135–146, 1993.
- [58] L. V. Kantorovich. On a problem of Monge. *J. Math. Sci., New York*, 133(4) :1383, 2006.
- [59] L. Kantorovitch. On the translocation of masses. *Manage. Sci.*, 5 :1–4, 1958.
- [60] L. V. Kantorovič and G. Š. Rubinštejn. On a space of completely additive functions. *Vestnik Leningrad. Univ.*, 13(7) :52–59, 1958.
- [61] M. R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York, 2008.
- [62] S. Kotz and S. Nadarajah. *Extreme value distributions*. Imperial College Press, London, 2000. Theory and applications.
- [63] T. Le Gouic and J.-M. Loubes. Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields*, 168(3-4) :901–917, 2017.
- [64] T. Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication.
- [65] S. Novak. *Extreme Value Methods with Applications to Finance*. <https://www.crcpress.com/Extreme-Value-Methods-with-Applications-to-Finance/Novak/9781439835746>. 11 2011.
- [66] V. M. Panaretos and Y. Zemel. Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.*, 6 :405–431, 2019.
- [67] R. Pether and R. Fraser. A quick reference table for extreme flood hydrology methods at uk dams. *Dams and Reservoirs*, 29(1) :41–42, 2019.
- [68] J. Pickands, III. Statistical inference using extreme order statistics. *Ann. Statist.*, 3 :119–131, 1975.
- [69] H. S. A. Potter. The mean values of certain dirichlet series, ii. *Proceedings of the London Mathematical Society*, s2-47(1) :1–19, 1942.
- [70] R.-D. Reiss. *Approximate distributions of order statistics*. Springer Series in Statistics. Springer-Verlag, New York, 1989. With applications to nonparametric statistics.
- [71] R.-D. Reiss. *A course on point processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.

- [72] R.-D. Reiss and M. Thomas. *Statistical analysis of extreme values*. Birkhäuser Verlag, Basel, second edition, 2001. From insurance, finance, hydrology and other fields, With 1 CD-ROM (Windows).
- [73] S. I. Resnick. Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin*, 27(1) :139–151, 1997.
- [74] S. I. Resnick. Heavy tail modeling and teletraffic data. *Ann. Statist.*, 25(5) :1805–1869, 1997. With discussion and a rejoinder by the author.
- [75] S. I. Resnick. *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. Probabilistic and statistical modeling.
- [76] H. Rootzén and N. Tajvidi. Can losses caused by wind storms be predicted from meteorological observations? *Scand. Actuar. J.*, (2) :162–175, 2001.
- [77] N. Ross. Fundamentals of Stein’s method. *Probab. Surv.*, 8 :210–293, 2011.
- [78] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*, volume 59 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. Reprint of the 1986 original [MR0838963].
- [79] J. A. Smith. Estimating the upper tail of flood frequency distributions. *Water Resources Research*, 23(8) :1657–1666, 1987.
- [80] H. Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.
- [81] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [82] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.
- [83] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [84] S. Wager. Asymptotic theory for random forests, 2014.
- [85] S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523) :1228–1242, 2018.
- [86] I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, 73(364) :812–815, 1978.

Résumé :

Dans le travail présenté dans ce manuscrit, nous développons deux méthodes de couplages pour la régression quantile extrême. Nous considérons des variables aléatoires $Y \in \mathbb{R}$ et $X \in \mathbb{R}^d$ et nous voulons estimer le quantile de niveau $1 - \alpha$ de Y sachant $X = x$. Ce problème est étudié dans le régime des extrêmes, c'est-à-dire que α tend vers 0.

Dans cette étude, nous introduisons le modèle des queues proportionnelles, inspiré du modèle des extrêmes hétéroscédastiques de Einmahl, de Hann et Zhou. Ce modèle revient à supposer que Y admet une distribution à queue lourde et que la queue de distribution de Y sachant $X = x$ est asymptotiquement proportionnelle à la queue de Y non conditionnelle. Nous proposons alors des estimateurs des différents paramètres de ce modèle ainsi qu'un estimateur du quantile conditionnel, ces estimateurs sont ensuite étudiés avec deux méthodes de couplage.

La première méthode repose sur le couplage de processus empiriques et se concentre sur l'estimation des paramètres du modèle ainsi que l'estimation du quantile. La seconde méthode est quant à elle reliée au transport optimal et repose sur l'utilisation des propriétés des distances de Wasserstein. Cette seconde approche nous permet d'estimer les paramètres du modèle avec la présence de biais ainsi que de développer une procédure de validation.

Abstract :

In this work, we develop two coupling approaches for extreme quantile regression. We consider i.i.d. copies of $Y \in \mathbb{R}$ and $X \in \mathbb{R}^d$ and we want an estimation of the conditional quantile of Y given $X = x$ of order $1 - \alpha$ for a very small $\alpha > 0$.

We introduce the proportional tail model which is strongly inspired by the heteroscedastic extremes developed by Einmahl, de Haan and Zhou. We assume that Y has a heavy tail with extreme value index γ and that the conditional tail of Y given $X = x$ is asymptotically proportional to the unconditional tail. We propose and study estimators of both model parameters and conditional quantile which are studied by coupling methods.

The first method is based on coupling of empirical processes while the second is related with optimal transport. Even if we establish the asymptotic normality of parameters estimators with both methods, the first is focused on the proper quantile estimation whereas the second is more focused on the estimation of γ in presence of bias and the elaboration of a validation procedure for our model.

Moreover, we also develop the optimal coupling approach in the general case of univariate extreme value theory.