UNIVERSITÉ GASTON BERGER DE SAINT-LOUIS

ÉCOLE DOCTORALE DES SCIENCES ET DES TECHNOLOGIES

U.F.R. DE SCIENCES APPLIQUÉES ET DE TECHNOLOGIE

THÈSE

présentée par

Aboubacrène Ag AHMAD

pour obtenir le grade de

DOCTEUR EN MATHÉMATIQUES APPLIQUÉES DE L'UNIVERSITÉ GASTON BERGER

Spécialité : **Probabilités-Statistique**

Modélisation semi-paramétrique des extrêmes conditionnels

Thèse soutenue publiquement le **16 septembre 2020**, devant le jury composé de :

M. Abdou Kâ DIONGUE	Professeur Titulaire (UGB, Sénégal)	Président
M. Abdelhakim NECIR	Professeur Titulaire (Université de Biskra, Algérie)	Rapporteur
M. Laurent GARDES	Professeur Titulaire (Université de Strasbourg, France)	Rapporteur
M. Ali Souleymane DABYE	Professeur Titulaire (UGB, Sénégal)	Rapporteur
M. Papa NGOM	Professeur Titulaire (UCAD, Sénégal)	Examinateur
M. Aliou DIOP	Professeur Titulaire (UGB, Sénégal)	Co-directeur de thèse
M. Stéphane GIRARD	Directeur de recherche (Mistis, Inria, Rhône-Alpes, France)	Co-directeur de thèse
M. El Hadji DEME	Maître-Assistant (CAMES) (UGB, Sénégal)	Co-encadrant de thèse

Thèse préparée au sein du Laboratoire d'Études et de Recherches en Statistique Appliquée et Développement (LERSTAD) et de l'Équipe MISTIS de l'INRIA de Grenoble Rhône-Alpes.

Dédicaces

A ma famille.

Remerciements

Mes remerciements vont tout d'abord à mes directeurs de thèse Aliou Diop et Stéphane Girard et mon co-encadrant El Hadji Deme pour la qualité de leur encadrement et leur constante disponibilité tout au long de cette thèse. Ils ont été les principaux acteurs de la réussite de ce travail de recherche. Je leur témoigne, ici, ma très grande gratitude pour tous leurs efforts, leurs conseils mais aussi les connaissances qu'ils m'ont transmises durant ces longues années. Un grand merci à Aliou Diop pour son aide et ses encouragements depuis mon cycle de master qui a débouché sur cette thèse pour laquelle il m'a entièrement fait confiance.

Je remercie également Laurent Gardes (Université de Strasbourg, France), Abdelhakim Necir (Université Mohamed Khider de Biskra, Algérie) et Ali Souleymane Dabye (Université Gaston Berger, Sénégal) pour avoir accepté de rapporter cette thèse dans un contexte extrêmement difficile marqué par la pandémie du Covid-19. Je leur témoigne toute ma reconnaissance pour leurs efforts, suggestions et remarques.

Merci aussi à Abdou Kâ Diongue (Université Gaston Berger, Sénégal) pour m'avoir fait l'honneur d'accepter de présider mon jury et à Papa Ngom (Université Cheick Anta Diop, Sénégal) d'avoir accepté en faire partie.

J'adresse toute ma reconnaissance au CEA-MITIC (Centre d'Excellence Africain en Mathématiques, Informatique et TIC) pour son appui financier.

Merci également à l'équipe-projet SIMERGE (Statistical Inference for the Management of Extreme Risks, Genetics and Global Epidemiology) dont l'appui financier conjoint avec celui du CEA-MITIC m'a permis d'effectuer, trois fois de suite, des séjours de recherche à l'équipe MISTIS (Modelling and Inference of Complex and Structured Stochastic Systems) de l'Inria de Grenoble.

Je remercie l'équipe MISTIS pour l'accueil ainsi que la qualité des conditions de travail dont j'ai bénéficié durant mes séjours en son sein. Il me plaît d'adresser une mention très spéciale à Stéphane Girard qui, en plus de la contraignante tâche d'encadrement, s'est investi sans réserve pour que ces séjours se passent dans les meilleures conditions. En lui, j'ai trouvé un homme méthodique, simple, attentif et avec un sens de l'humour assez élevé. J'adresse aussi un grand merci à Antoine Usseglio-Carleve de la-dite équipe, avec qui j'ai eu une collaboration fructueuse dans le cadre de mes travaux de recherche, et dont l'aide m'a été très précieuse.

Merci au corps professoral et au personnel administratif de l'UFR SAT ainsi qu'à tous les camarades du Laboratoire d'Etudes et de Recherches en Statistique et Développement (LERSTAD).

J'exprime aussi mes vifs remerciements et toute ma gratitude à ma famille d'accueil à Saint-Louis. Je nomme ainsi la famille "El Hadji Deme" dont la porte m'a toujours été grande ouverte et où j'ai bénéficié d'un traitement à titre de frère. Jërë-jëf Deme !

Je saisis cette occasion pour adresser un remerciement tout particulier à Ouateni Diallo qui m'a toujours encouragé et soutenu dans mon programme de formation de troisième cycle.

Merci également à tous ceux qui ne sont pas nommément cités ici mais qui ont contribué d'une manière ou d'une autre à l'aboutissement de ce modeste travail.

Enfin, un grand merci à ma famille pour sa longue patience et son soutien ô combien important.

Résumé

L'objectif principal de cette thèse est de proposer de nouveaux estimateurs de l'indice des valeurs extrêmes (indice de queue) ainsi que des quantiles extrêmes conditionnels pour une famille de distributions à queue lourde. La famille de distributions considérée est définie à partir d'un modèle de régression avec des paramètres fonctionnels de position $a(\cdot)$ et d'échelle $b(\cdot)$ inconnus. La variable d'intérêt Y, supposée aléatoire et réelle, est simultanément mesurée avec une covariable déterministe x. Les résidus Z du modèle sont indépendants de la covariable et sont distribués suivant une loi du domaine d'attraction de Fréchet d'indice de queue γ inconnu et supposé constant.

Pour plus de souplesse que les approches purement paramétriques, nous préconisons une approche d'estimation semi-paramétrique. Aussi, la constance de l'indice de queue nous permet d'obtenir, dans le cas de petits échantillons, des estimations plus fiables que dans certaines approches purement non paramétriques existant dans la littérature.

Nous établissons les propriétés asymptotiques de nos estimateurs et présentons, sur des simulations aussi bien que sur des données réelles, des résultats permettant d'apprécier leur comportement pour des échantillons de taille finie.

Mots-clés : Valeurs extrêmes, estimation semi-paramétrique, fonctions de position et d'échelle, indice de queue, quantiles conditionnels, distributions à queue lourde.

Abstract

The main goal of this thesis is to propose new estimators of the tail-index as well as the conditional extreme quantiles in a family of heavy-tailed distributions. The considered family of distributions is defined from a regression model with a location function $a(\cdot)$ and a scale function $b(\cdot)$ which are unknown. The real random variable of interest Y is simultaneously recorded with a deterministic covariate x. The residuals Z of the model are independent of the covariate and their cumulative distribution function belongs to the Fréchet domain of attraction whose the tail-index γ is unknown and assumed to be constant. For more flexibility than purely parametric approaches, we opt for a semi-parametric estimation approach. Also, the constancy of the tail-index allows us to obtain, in the case of small samples, more reliable estimates than in certain purely non-parametric approaches existing in the literature.
We establish the asymptotic properties of our estimators and present some results allowing to appreciate their finite sample properties both on simulated and real data.

Keywords : Extreme values, semi-parametric estimation, location and scale functions, tail-index, conditional quantile, heavy-tailed distributions.

Table of contents

Introduction générale

Certains évènements (tremblements de terre, ouragans, inondations, crises financières, etc.) se produisant à travers le monde attirent particulièrement l'attention de par leurs conséquences énormes sur les plans humain, économique, financier, environnemental, etc. A titre d'exemple, nous pouvons citer :

- L'ouragan Dorian avec des pluies torrentielles et des vents frôlant les 300 km/h qui a touché une bonne partie de la côte Atlantique nord (Petites Antilles, Porto Rico, Floride, Géorgie, Les Carolines, Bahamas, Provinces de l'Atlantique) du 24 août au 10 septembre 2019, causant la mort de 84 personnes et des dégâts matériels estimés à plus de 8 milliards de dollars américains (`https://fr.wikipedia.org/wiki/Ouragan_Dorian_(2019)`).
- Le séisme de Haïti d'une magnitude de 7 à 7.3 survenu le 12 janvier 2010 et dont le bilan s'élève à plus de 230000 morts, 220000 blessés, 1.3 millions de sans abris et plusieurs bâtiments détruits (`https://fr.wikipedia.org/wiki/S%C3%A9isme_de_2010_en_Ha%C3%AFti`).

De l'impératif d'une théorie

Les évènements évoqués ci-dessus, quelle que soit leur nature, soulèvent la question très importante de la prédiction de leur apparition, leur récurrence et leur impact. Il est donc nécessaire d'élaborer une théorie qui permette de les étudier, de comprendre leur comportement et de proposer des outils fiables pour leur prédiction afin de limiter au mieux leurs conséquences. Il faut préciser que ces évènements dits **extrêmes** se caractérisent par une faible probabilité d'apparition et quand ils se produisent, ils prennent de très petites ou de très grandes valeurs et ont un grand impact. On notera donc la différence avec un évènement rare qui est un évènement dont la probabilité d'occurrence est faible. Ainsi, le fait qu'un évènement soit rare n'implique pas qu'il soit extrême, car non quantifiable (petites ou grandes valeurs). A l'inverse, tout évènement extrême est rare au sens où il a une faible probabilité de se produire.

De l'estimation des évènements extrêmes

Pour tenter de répondre à la question précédente, des modèles physiques de prédiction ont été développés en météorologie, par exemple. Cependant, ces modèles ne permettent pas de donner des prévisions à plus de quelques semaines. Une première alternative pour la prédiction de ces évènements consiste en l'utilisation d'outils mathématiques, plus spécifiquement probabilistes, en se basant sur un échantillon de données, l'idée étant de caractériser la probabilité d'occurrence d'un évènement extrême donné. Nous allons illustrer dans les deux paragraphes suivants la mise en marche de cette alternative et ses difficultés.

Un problème et son dual

A partir d'un échantillon de données, peut-on résoudre l'un des deux problèmes suivants ?

(i) Evaluer la probabilité d'occurrence d'un évènement d'amplitude supérieure à la valeur maximale de l'échantillon.

(ii) Déterminer l'amplitude de l'évènement qui est dépassée avec une certaine probabilité supposée faible.

On se rend bien compte que ces deux problèmes sont étroitement liés et révèlent des aspects importants de la prédiction des évènements extrêmes.

Une prédiction probabiliste

Considérons l'exemple du jeu de données (disponible sous le package `evir` du logiciel R) qui consiste en $n = 154$ excès (débits X) au-dessus du niveau $65m^3/s$ durant la période $1934 - 1969$ (35 ans) de la rivière Nidd située dans le Yorkshire en Angleterre. La Figure 1 représente l'histogramme de ces mesures de débits.
Déterminons dans un premier temps, la probabilité d'observer un débit de plus de $160m^3/s$. L'histogramme de la Figure 2 (a) permet d'avoir :

$$\mathbb{P}(X \geq 160) \simeq \text{nombre}(X_i \geq 160)/n = 11/154.$$

Ainsi, un tel débit est enregistré en moyenne tous les $35/11 \simeq 3.2$ ans. Cette évaluation est d'autant plus importante que si l'on veut ériger une infrastructure qui ne sera pas submergée durant au moins 3.2 ans, il faut faire de sorte qu'elle puisse contenir un flux de $160m^3/s$, une sous-évaluation exposant à un danger et une sur-évaluation entrainant des coûts de construction plus importants.
A partir respectivement des histogrammes (b) et (c) de la Figure 2, on vérifie également que la probabilité d'observer un débit de plus de $255m^3/s$, par exemple, est $3/154$ (soit un débit de ce genre tous les 11.7 ans) et que celle d'un débit dépassant $280m^3/s$ est $1/154$ (soit un débit de ce genre tous les 35 ans).

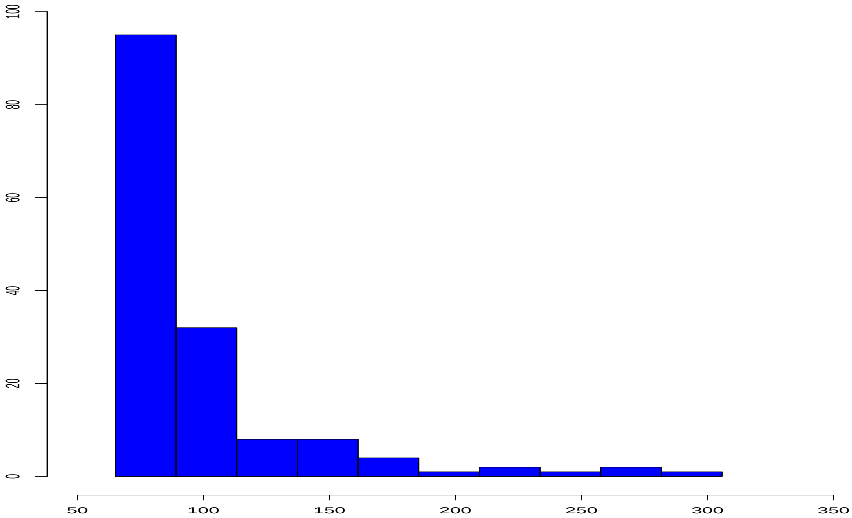

Figure 1: Histogramme des débits de la rivière Nidd.

Essayons maintenant de déterminer la probabilité d'un débit qui dépasse $500m^3/s$. L'histogramme de la Figure 2 (d) nous donne alors une probabilité nulle.

$$\mathbb{P}(X \geq 500) \simeq \text{nombre}(X_i \geq 500)/n = 0.$$

Cette probabilité est-t-elle vraiment nulle ? La réponse est bien-sûr non, juste qu'il est impossible de la déterminer à partir de l'histogramme. La même difficulté subsiste si l'on veut, par exemple, déterminer la valeur du débit centennal de la rivière, c'est-à-dire, la valeur t du débit qui est atteinte une fois par siècle : $\mathbb{P}(X \geq t) = 35/(154 \times 100) = 1/440$. Cet exemple met clairement en évidence la difficulté pratique dans la prédiction probabiliste des évènements extrêmes à partir d'un échantillon de données. Cette difficulté provient du fait que l'on dispose d'un grand nombre d'observations pour les évènements plus fréquents et peu, voire pas du tout, pour les évènements extrêmes. Une autre source de cette difficulté est que ces échantillons couvrent, en général, des périodes d'à peine une centaine d'années. Pour tenter de résoudre le problème précédent, une approche basée sur **la théorie des valeurs extrêmes** a été adoptée. Cette théorie consiste à déduire le comportement des évènements extrêmes à partir d'évènements plus fréquents en extrapolant à partir de l'échantillon de données disponibles, sous des hypothèses de régularité sur les phénomènes observés. Ses premiers développements remontent à Nicolas Bernoulli en 1709 (voir Reiss [93] et Kotz and Nadarajah [81]) alors que la première application est due à Fuller en 1914. Elle fournit une base mathématique probabiliste rigoureuse pour la construction de modèles statistiques de prédiction des évènements extrêmes.

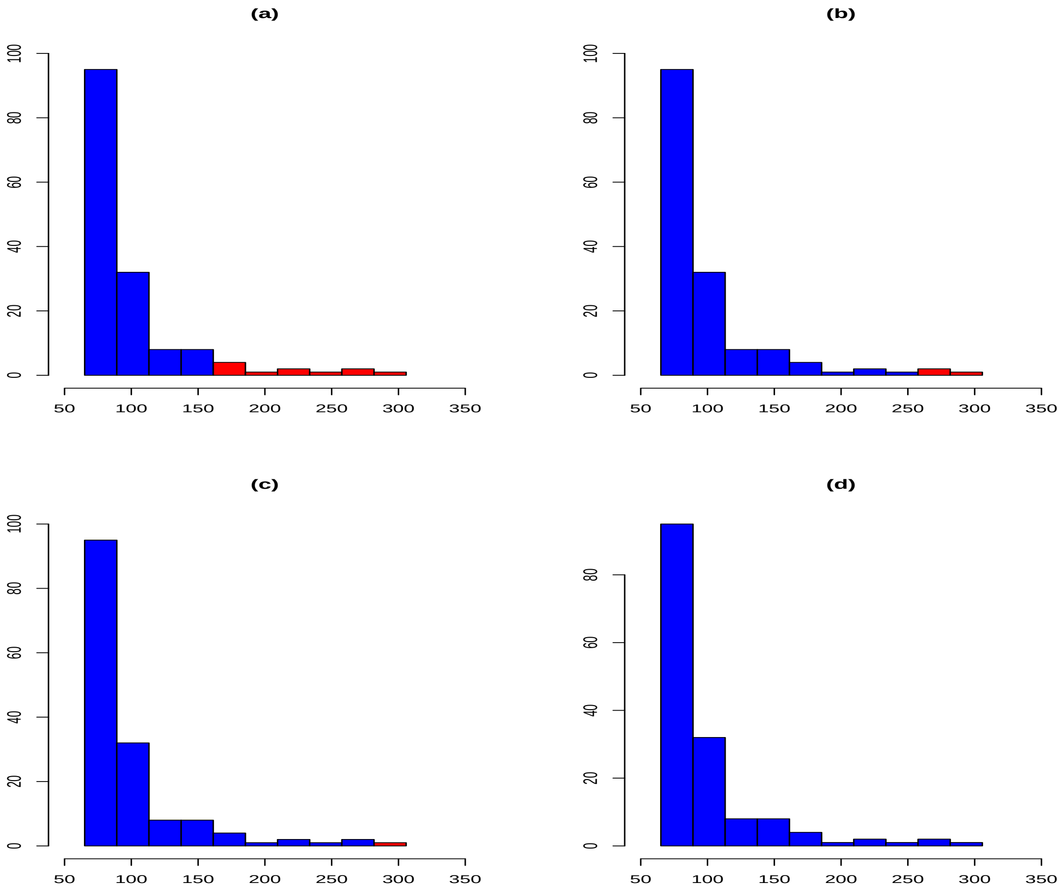

Figure 2: Histogramme des débits de la rivière Nidd. En bleu, les débits inférieurs à $160m^3/s$ (a), $255m^3/s$ (b), $280m^3/s$ (c) et $500m^3/s$ (d). En rouge, les débits supérieurs respectivement aux mêmes valeurs.

Du champ d'application

La diversité de la nature des évènements extrêmes explique la très vaste étendue du domaine d'application de la théorie des valeurs extrêmes. Historiquement, son domaine d'application reste l'hydrologie (de Haan [30], Katz et al. [79], Guillou and Willems [67], El Methni et al. [46] et Dutfoy et al. [44]), notamment suite aux travaux de Jules Emile Gumbel (Gumbel [69]) en 1954 et son ouvrage (Gumbel [70]) en 1958. Elle s'applique également dans de nombreux domaines tels que la fiabilité (Ditlevsen [41]), la climatologie (Rootzén and Tajvidi [98] et Bel et al. [12]), la météorologie (Coles and Tawn [23], Gardes and Girard [54], Ceresetti et al. [21], El Methni et al. [47] et Bechler et al. [4]), l'assurance (Beirlant and Teugels [7], Embrechts et al. [48], Resnick [96] et Rootzén and Tajvidi [97]) et la finance (Embrechts et al. [48], Embrechts et al. [49] et McNeil et al. [88]).
Pour d'autres exemples d'applications, on peut se référer au livre de Reiss and Thomas [94].

Contributions de la thèse

Cette thèse s'inscrit dans le cadre de la Statistique des valeurs extrêmes et y apporte essentiellement deux contributions :

1. La première contribution de cette thèse porte sur la construction d'un nouvel estimateur de l'indice de queue conditionnel pour une famille de distributions à queue lourde. Dans la littérature sur l'estimation de l'indice de queue conditionnel, ce dernier dépend de la covariable. Nous considérons un modèle où cet indice est supposé constant et nous y introduisons des paramètres de position et d'échelle dans un design fixe unidimensionnel. Nous adoptons une approche d'estimation semi-paramétrique et nous parvenons ainsi à construire un estimateur à performances égales que dans le cas sans covariable.
   Plus précisément, pour définir la famille de distributions à queue lourde de notre modèle, nous avons considéré une variable aléatoire réelle Y de fonction de survie conditionnelle à $x \in [0,1]$ donné, vérifiant :

   $$\overline{F}_Y(y|x) := \mathbb{P}(Y > y|x) = \overline{F}_Z\left(\frac{y - a(x)}{b(x)}\right),$$

   pour $y \geq y_0(x) > a(x)$. Les fonctions $a : [0,1] \to \mathbb{R}$ et $b : [0,1] \to \mathbb{R}_+^*$ sont respectivement les fonctions de position et d'échelle et $\overline{F}_Z$ est la fonction de survie d'une variable aléatoire Z supposée à queue lourde d'indice γ indépendant de x.
   Nous avons considéré un design fixe unidimensionnel et proposé, pour tout $x \in [0,1]$, des estimateurs des paramètres fonctionnels $a(\cdot)$ et $b(\cdot)$ sous la forme :

   $$\hat{a}_n(x) = \hat{q}_{n,Y}(\mu_2 \mid x) \text{ et } \hat{b}_n(x) = \hat{q}_{n,Y}(\mu_3 \mid x) - \hat{q}_{n,Y}(\mu_1 \mid x),$$

   où $\hat{q}_{n,Y}(\alpha \mid x),\ \alpha \in (0,1)$ est un estimateur non paramétrique du quantile conditionnel de Y et les constantes $(\mu_1, \mu_2, \mu_3) \in (0,1)^3$ sont telles que $\mu_3 < \mu_1$ et

   $$q_Z(\mu_2) = 0 \text{ et } q_Z(\mu_3) - q_Z(\mu_1) = 1,$$

   $q_Z(\cdot)$ désignant un quantile de Z. L'indice de queue est ensuite estimé à partir de statistiques de type Hill de la forme :

   $$\hat{\gamma}_n = \frac{1}{k_n}\sum_{i=0}^{k_n-1} \log \hat{Z}_{m_n-i,m_n} - \log \hat{Z}_{m_n-k_n,m_n},$$

   où k_n désigne le nombre d'extrêmes utilisées et $\hat{Z}_{m_n-k_n,m_n} \leq \cdots \leq \hat{Z}_{m_n,m_n}$ les statistiques d'ordre supérieur associées aux m_n résidus $\hat{Z}_i = (Y_i - \hat{a}_n(x_i))/\hat{b}_n(x_i)$,

$i = 1, \ldots, m_n$, retenus par la procédure d'estimation.
Nous avons établi les propriétés asymptotiques de nos estimateurs et vérifié ces propriétés pour des échantillons de taille finie aussi bien sur des données simulées que sur des données réelles d'assurance.

2. La seconde contribution est relative à la construction d'un nouvel estimateur des quantiles extrêmes conditionnels. Nous introduisons un modèle de régression avec des paramètres de position et de dispersion dépendant d'une covariable x en design fixe où la variable d'intérêt Y est liée à une variable Z de distribution supposée à queue lourde et indépendante de x. Un cadre multidimensionnel est considéré et l'estimateur que nous proposons par une approche semi-paramétrique présente une vitesse de convergence nettement meilleure que celle de certains estimateurs dans le cas avec covariable où cette vitesse est impactée par la dimensionnalité.
Plus concrètement, nous considérons la classe des modèles de régression où une variable réponse $Y \in \mathbb{R}$ est liée à une covariable déterministe multidimensionnelle $x \in \Pi \subset \mathbb{R}^d$, $d \geq 1$ par :
$$Y = a(x) + b(x)Z,$$
où la variable aléatoire $Z \in \mathbb{R}$ est supposée à queue lourde d'indice γ indépendant de x.
Ce modèle généralise celui proposé précédemment au cas multidimensionnel et les paramètres fonctionnels $a(\cdot)$ et $b\cdot)$ ainsi que l'indice de queue γ sont estimés suivant la même approche. Nous proposons alors un estimateur plugin des quantiles extrêmes conditionnels de Y, pour $x \in \Pi$, sous la forme :
$$\tilde{q}_{n,Y}(\alpha_n \mid x) = \hat{a}_n(x) + \hat{b}_n(x)\hat{q}_{n,Z}(\alpha_n), \ \alpha_n \to 0,$$
où $\hat{a}_n(x)$ et $\hat{b}_n(x)$ sont respectivement des estimateurs des fonctions de position $a(\cdot)$ et de dispersion $b(\cdot)$ et $\hat{q}_{n,Z}(\alpha_n)$ est un estimateur des quantiles extrêmes de Z.
Sous des hypothèses de régularité plus fortes que dans le cas unidimensionnel sur les paramètres fonctionnels $a(\cdot)$ et $b(\cdot)$ et sur la fonction de survie de Z, nous avons établi les propriétés asymptotiques de nos estimateurs. Nous avons non seulement mis en évidence une amélioration du terme de biais par rapport au cas unidimensionnel mais aussi l'avantage que présente notre modèle pour contourner le problème de dimension qui se pose dans les méthodes non-paramétriques. Nous avons également vérifié les propriétés de nos estimateurs pour des échantillons de taille finie aussi bien sur des données simulées que sur des données réelles de tsunami après avoir testé et retenu l'hypothèse de constance de l'indice de queue.

Organisation de la thèse

Cette thèse consiste en trois chapitres indépendants les uns des autres.
Le Chapitre 1 constitue un aperçu assez général sur la théorie des valeurs extrêmes univariées réelles. Nous y rappelons quelques résultats de cette théorie qui nous sont utiles pour la suite de cette thèse. Le Chapitre 2 propose un nouvel estimateur de l'indice de queue conditionnel pour une famille de distributions à queue lourde avec des paramètres de position et d'échelle. Le Chapitre 3 propose, quant à lui, un nouvel estimateur des quantiles extrêmes conditionnels pour des distributions à queue lourde avec des paramètres de position et de dispersion.
Nous terminons le document par une conclusion et quelques perspectives au sujet de nos travaux de recherche.

Chapter 1

Reminders on the extreme value theory

Abstract

In this chapter, we briefly recall some essential notions on the extreme value theory in the real univariate framework. Section 1.1 gives a brief presentation of this theory when Section 1.2 presents the asymptotic behavior of the largest values of a sample. We expose in Section 1.3 some tools allowing to characterize the different domains of attraction. Section 1.4 deals with the estimation of the tail-index and Section 1.5 on the different estimation methods of extreme quantiles. We present in Section 1.6 some results on the estimation of extreme quantiles in the presence of a covariate. Finally Section 1.7 constitutes a few reminders on the notion of censoring and on the estimation of the tail-index and extreme quantiles under censoring.

Summary

1.1 Presentation of the extreme value theory

The classical statistical approach is based on the study of the behavior in **average** as well as the **variability** around this average of an observed phenomenon, in particular by the use of some probabilistic tools such as the law of large numbers, the central limit theorem, etc. Despite its popularity, this approach unfortunately fails to capture rare or extreme events.

The study of the behavior of these events comes under the extreme value theory. As we mentioned in the general introduction introduction, this theory is about solving problems like the calculation of a low probability (close to zero) associated with an extreme event (see problem (i)) or the determination of the value (called quantile in statistics, Value-at-Risk (VaR) in finance or in actuarial science, return level [(1)] in hydrology, etc.) of an extreme event (see problem (ii)). To answer these two questions, it turns more towards the use of information carried by the most extreme values of the data set and not the central values as in the classic statistical approach. Thus the extreme value theory is developed, the main result of which is based on the Theorem of Fisher and Tippett [52] and Gnedenko [61] on the convergence in law of the maximum value of a sequence of independent and identically distributed random variables, and then on the result of Pickands [90] on the convergence in law of excesses above a threshold. We refer the reader to the works of Bingham et al. [16],

(1). Level exceeded on average once during a given period.

Resnick [95], Embrechts et al. [48], Beirlant et al. [10], de Haan and Ferreira [31] and Lo et al. [84] for a more complete discussion on extreme value theory.

1.2 Limit laws for extreme values

In this section, we study the asymptotic behaviour of the largest values in a sample. We introduce the notion of the law of extreme values by first looking at the law of the maximum of a sample in Paragraph 1.2.1 then by considering the excesses above a given threshold in Paragraph 1.2.2.

1.2.1 Limit law for maximum (EVD-GEV)

Let X be a random variable with distribution function F :

$$F(x) = \mathbb{P}(X \leq x) = 1 - \overline{F}(x), \tag{1.1}$$

where $\overline{F}$ is its associated survival function. Given a sample $X_1, \dots, X_n$ of independent copies of X, we focus on the behaviour of the random variable $X_{n,n} = \max(X_1, \dots, X_n)$. The distribution function of $X_{n,n}$ is given by :

$$\begin{aligned} F_{X_{n,n}}(x) := \mathbb{P}(X_{n,n} \leq x) &= \mathbb{P}(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= \mathbb{P}(X_1 \leq x) \dots \mathbb{P}(X_n \leq x) \\ &= F^n(x). \end{aligned} \tag{1.2}$$

Since F is unknown in practice, the result (1.2) is even more difficult to exploit. However, based on the properties of a distribution function, we have the following asymptotic result :

$$\lim_{n\to\infty} F_{X_{n,n}}(x) = \lim_{n\to\infty} F^n(x) = \mathbb{1}_{\{x \geq \tau_F\}}, \tag{1.3}$$

where $\mathbb{1}_{\{.\}}$ is the indicator function and $\tau_F := \sup\{x \in \mathbb{R},\ F(x) < 1\}$, with $\sup\{\emptyset\} = \infty$ by convention, denotes the right endpoint of the function F.
The result (1.3) yields that the distribution function of $X_{n,n}$ is degenerated and therefore not very informative. It is then necessary to consider the asymptotic behaviour of the suitably normalized maximum to exhibit a non-degenerated limit distribution towards which $F_{X_{n,n}}$ will converge. The following theorem gives the form of this limit distibution.

Theorem 1.1 ***(Fisher and Tippett [52] ; Gnedenko [61])*** *Let $X_1, \dots, X_n$ be independent and identically distributed random variables with distribution function F. Suppose there exists a sequence of constants $(a_n)_{n\geq 1} > 0$, and $(b_n)_{n\geq 1}$ real and a non-degenerate* *distribution $\mathscr{H}$ such that for all $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = \mathscr{H}(x),$$

then $\mathscr{H}$ belongs to the type[(2)] *of one of the following three distribution functions :*

$$\Phi_\alpha(x) = \begin{cases} 0 & \textit{si } \; x \leq 0 \\ \\ \exp\left(-x^{-\alpha}\right) & \textit{si } \; x > 0, \; \alpha > 0 \end{cases} \qquad \textit{(Fréchet)}$$

$$\Psi_\alpha(x) = \begin{cases} 1 & \textit{si } \; x \geq 0 \\ \\ \exp\left(-(-x)^{\alpha}\right) & \textit{si } \; x < 0, \; \alpha > 0 \end{cases} \qquad \textit{(Weibull)}$$

$$\Lambda(x) = \exp\left(-\exp(-x)\right) \quad \textit{pour tout } x \in \mathbb{R} \qquad \textit{(Gumbel)}$$

The normalizing constants a_n and b_n are respectively scale and location parameters and depend on the distribution of X. Theorem 1.1 has some similarity with the Central Limit Theorem (CLT) in classical statistical theory which gives the asymptotic law of the mean of a sample of independent and identically distributed random variables. Indeed, a_n plays the role of $\sigma(X)/\sqrt{n}$ ($\sigma(X)$ denoting the standard deviation of X) in the CLT and b_n the one of $\mathbb{E}(X)$ (Expectation of X). The three above distribution functions Λ, Ψ_α and Φ_α are the only possible limit laws of the normalized maximum of a sample of independent and identically distributed random variables. They are referred to as the Extreme Value Distribution (EVD). A parametrization of these three distributions into a single formula due to von Mises [104] and Jenkinson [77] called Generalized Extreme Value Distribution (GEV) is given by :

$$\mathscr{H}_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right) & \text{for all } x \text{ such that } 1+\gamma x > 0, \text{ if } \gamma \neq 0 \\ \exp\left(-\exp(-x)\right) & \text{for all } x \in \mathbb{R}, \text{ if } \gamma = 0. \end{cases} \tag{1.4}$$

The parameter γ so-called the extreme-value index or the tail-index completely characterizes the behaviour of the tail of the distribution F. Its sign also determines the notion of domain of attraction :

- if $\gamma < 0$, F is said to belong to the **Weibull** maximum domain of attraction ($F \in$ DA(Weibull)). This domain of attraction includes distributions with short tails, *i.e.* they have a finite endpoint.
- if $\gamma = 0$, F is said to belong to the **Gumbel** maximum domain of attraction ($F \in$

(2). Two non-degenerated distribution functions I and J are of same type if and only if there exist $a > 0$ and $b \in \mathbb{R}$ such that $I(ax + b) = J(x)$ for all $x \in \mathbb{R}$.

DA(Gumbel)). This domain of attraction includes distributions with light tails, *i.e.* their survival distribution function decrease as an exponential rate.

- if $\gamma > 0$, F is said to belong to the **Fréchet** maximum domain of attraction ($F \in$ DA(Fréchet)). This domain of attraction includes distribution with heavy tails, *i.e.* their survival distribution function decrease as a power function.

Figure 1.1 (left) illustrates on a standard exponential distribution, the convergence in distribution of the sequence of random variables $(a_n^{-1}(X_{n,n} - b_n))_{n\geq 1}$ to a non-degenerated limit $\mathscr{H}_0$, the normalizing constants being :

$$a_n = 1 \text{ and } b_n = \log n.$$

The rate of this convergence is illustrated in Figure 1.1 (right).

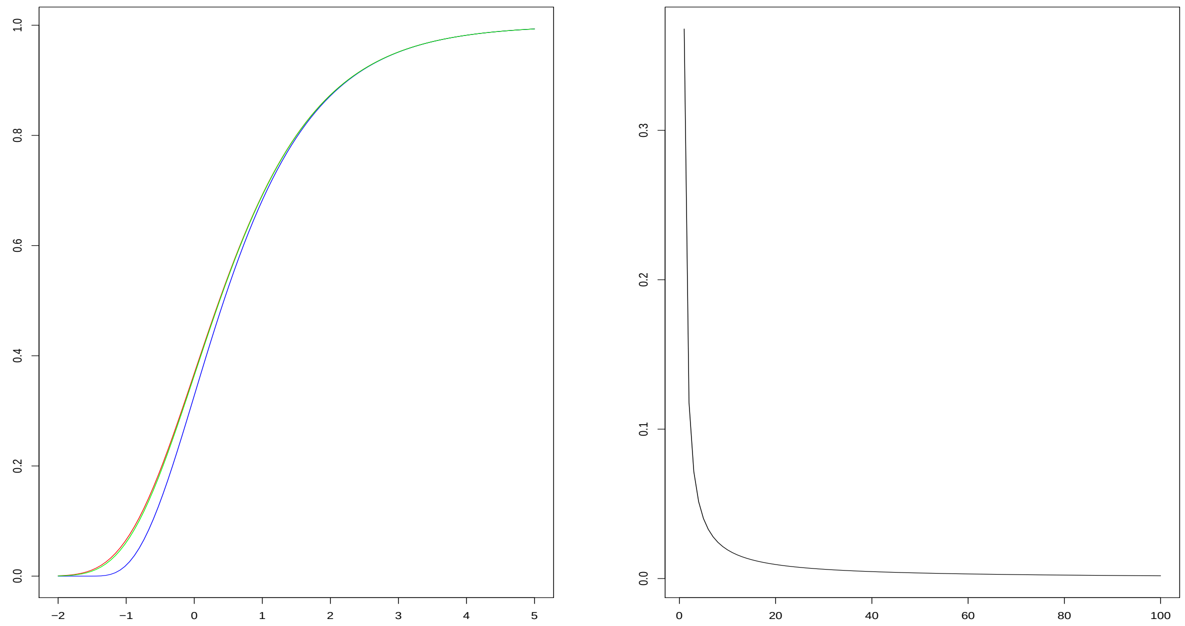

Figure 1.1: Left : Comparison on a standard exponential distribution between $\mathscr{H}_0(x)$, $\mathbb{P}\left(\frac{X_{n,n}-b_n}{a_n} \leq x\right)$ avec $n = 5$ and $\mathbb{P}\left(\frac{X_{n,n}-b_n}{a_n} \leq x\right)$ with $n = 30$. Right : Illustration of the rate of convergence in distribution of the normalized maximum of a standard exponential distribution to $\mathscr{H}_0(x)$: values of n on x-axis and $\max |\mathscr{H}_0(x) - \exp(x + \log n)^n|$ on y-axis.

Table 1.1 gives the maximum domains of attraction associated with usual distributions and Figure 1.2 illustrates for $\gamma \in \{-1, 0, 1\}$, an example of densities and distribution functions associated with the extreme-value distribution.

Domain of attraction	**Weibull** $\gamma < 0$	**Gumbel** $\gamma = 0$	**Fréchet** $\gamma > 0$
Distributions	Beta ReverseBurr Uniform	Exponential Gamma Gumbel Logistic Log-normal Gaussian Weibull	Burr Cauchy Chi-deux Fréchet Pareto Student

Table 1.1: Maximum domains of attraction associated with usual distributions.

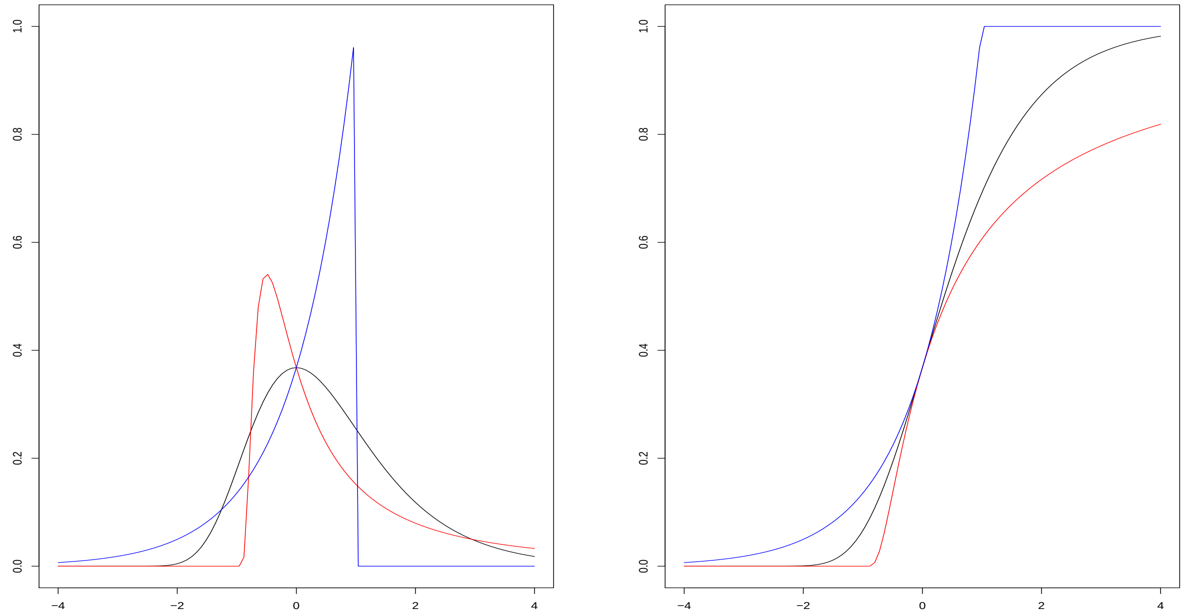

Figure 1.2: Densities (left) and distribution functions (right) associated with the extreme-value distribution ($\gamma = -1$ (blue), $\gamma = 0$ (black) and $\gamma = 1$ (red)).

It should also be noted that a more general form of GEV can be obtained from (1.4) by replacing x by $(x - \mu)/\sigma$ for $\mu \in \mathbb{R}$, $\sigma > 0$:

$$\mathscr{H}_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\left(-\left(1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}\right) & \text{for all } x \text{ such that } 1+\gamma\left(\frac{x-\mu}{\sigma}\right) > 0, \text{ if } \gamma \neq 0 \\ \exp\left(-\exp(-x)\right) & \text{for all } x \in \mathbb{R}, \text{ if } \gamma = 0. \end{cases}$$

Remark 1.1

Corresponding results for the minimum $X_{1,n} := \min(X_1, \ldots, X_n)$ can easily be obtained from the following identity :

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n).$$

For example,

$$F_{X_{1,n}}(x) = \mathbb{P}(X_{1,n} \leq x) = 1 - (1 - F(x))^n .$$

1.2.2 Limit law for excesses over threshold (POT)

In the approach by GEV, the use of the maximum leads to a loss of the information contained in the other large values of the sample. To overcome this problem, the POT (Peaks-over-Threshold) method or method of excesses over a high threshold has been introduced by Pickands [90]. The idea of this method is as follows :
Given a sample $X_1, \ldots, X_n$ of independent and identically distributed random variables a large threshold u ($u < \tau_F$) is set and the N_u observations $X_{i_1}, \ldots, X_{i_{N_u}}$ exceeding this threshold are considered. Let $Y_j = X_{i_j} - u,\ \ j = 1, \ldots, N_u$ be the excesses over the threshold u (see Figure 1.3).

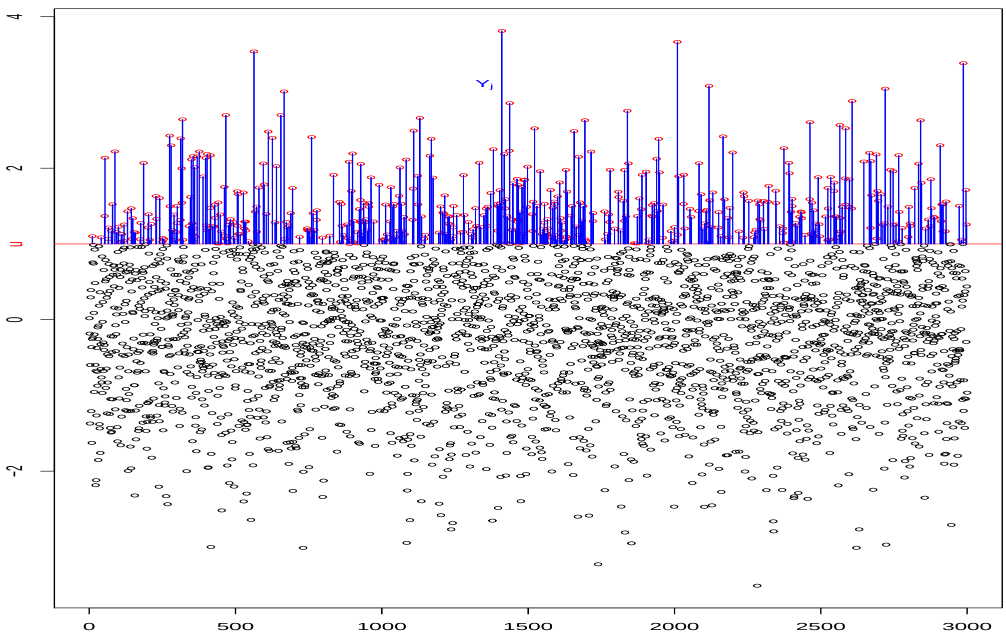

Figure 1.3: Illustration of the definition of excesses from $n = 3000$ observations of a standard normal distribution for a threshold $u = 1$. The black dots represent the observations below the threshold u, the red dots represent the observations above u and the blue segments define the excesses above u.

From F we define the distribution function F_u of the excesses Y over the threshold u for $0 \leq y \leq \tau_F - u$ by :

$$F_u(y) = \mathbb{P}(Y \leq y \mid X > u) = \mathbb{P}(X - u \leq y \mid X > u) = \frac{F(y+u) - F(u)}{1 - F(u)}, \tag{1.5}$$

or equivalently :

$$\overline{F}_u(y) = 1 - F_u(y) = \frac{\overline{F}(y+u)}{\overline{F}(u)}. \tag{1.6}$$

The objective of this method is thus to determine by which limit distribution the distribution (1.5) can be approached. Balkema and de Haan [3] and Pickands [90] have proposed the following theorem which specifies the distribution of excesses when the threshold u goes to the endpoint τ_F.

Theorem 1.2 ***(Balkema and de Haan [3] ; Pickands [90])*** *The distribution function F belongs to the maximum domain of attraction of the extreme value distribution ($F \in DA(\mathscr{H}_\gamma)$) if and only if there exists a positive function $\sigma(u)$ such that :*

$$\lim_{u \to \tau_F} \sup_{0 \leq y \leq \tau_F - u} \left| F_u(y) - G_{\gamma,\sigma(u)}(y) \right| = 0,$$

where $G_{\gamma,\sigma(u)}$ is the Generalized Pareto Distribution (GPD) defined by :

$$G_{\gamma,\sigma(u)}(y) = \begin{cases} 1 - \left(1 + \frac{\gamma y}{\sigma(u)}\right)^{-\frac{1}{\gamma}} & \text{if } \gamma \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma(u)}\right) & \text{if } \gamma = 0, \end{cases}$$

where $0 \leq y \leq \tau_F - u$ if $\gamma \geq 0$ and $0 \leq y \leq -\frac{\sigma(u)}{\gamma}$ if $\gamma < 0$.

The result of Theorem 1.2 yields that the GPD appears as the limit distribution of scaled excesses over high thresholds. It therefore plays the same role for these excesses than that of Theorem 1.1 for the suitably standardised maximum.
An illustration of this approximation is given in Figure 1.4 for excesses from a Fréchet distribution whose histogram is superimposed on the density of a GPD and Figure 1.5 where the distribution function of these excesses is superimposed on that of a GPD.

Remark 1.2

1. The tail-index γ is identical between a GEV and a GPD. This identity highlights, once again, its predominant role in the behaviour of extreme values.
2. It can be shown that the parameters of these two distributions are linked by the relation :

$$\sigma(u) = \sigma_n(u) = a_n + \gamma(u - b_n).$$

3. $G_{0,\sigma}$ is the exponential distribution with parameter $1/\sigma$ and $G_{-1,\sigma}$ is the uniform distribution on $[0, \sigma]$.

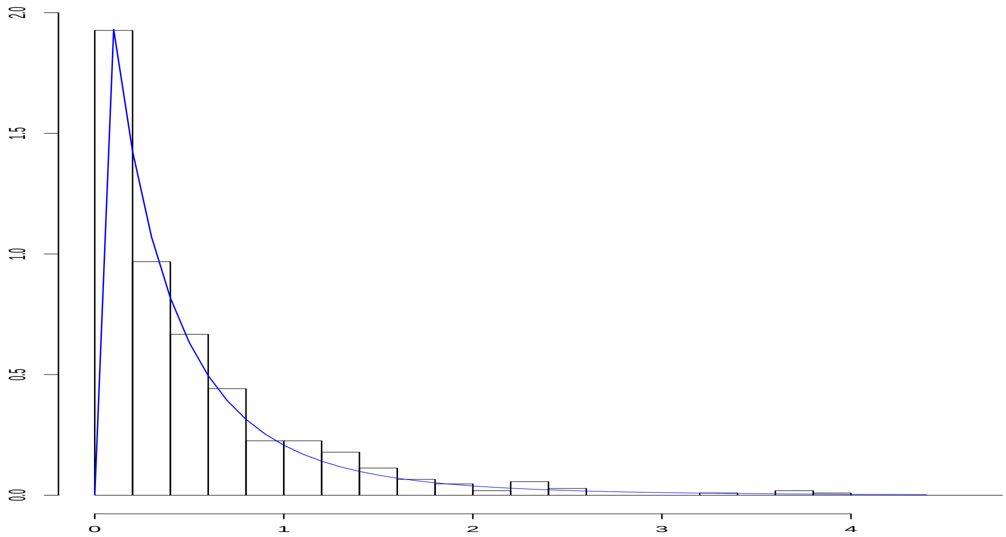

Figure 1.4: Illustration of the convergence of the excesses : Histogram of excesses from $n = 3000$ observations of a Fréchet distribution of parameter $\alpha = 4$ for a threshold $u = 1.5$ (black) and the density of a $\mathrm{GPD}_{1/\alpha,u/\alpha}$ (blue).

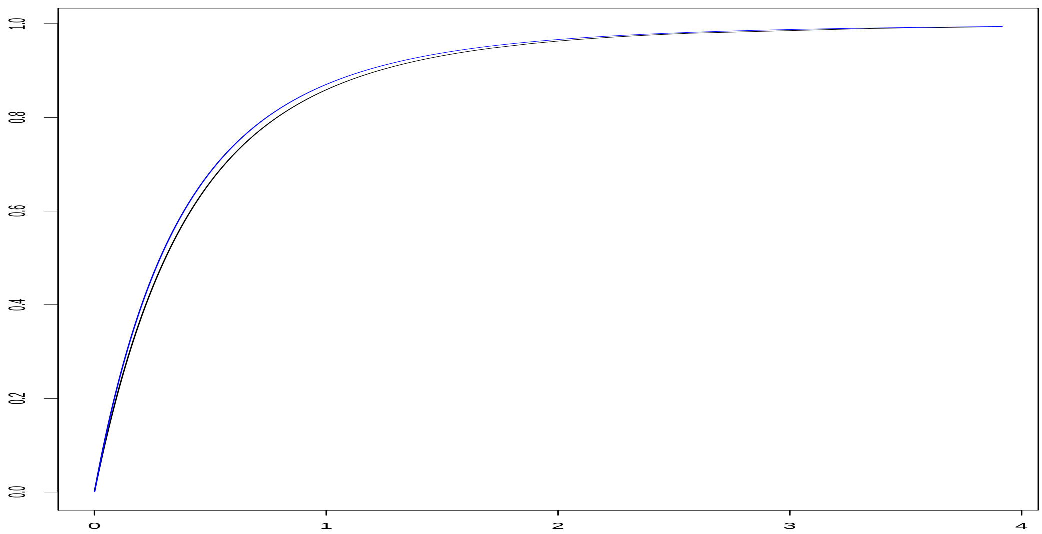

Figure 1.5: Illustration of the convergence of excesses : Distribution function of excesses from $n = 3000$ observations of a Fréchet distribution of parameter $\alpha = 4$ for a threshold $u = 1.5$ (black) and distribution function of a $\mathrm{GPD}_{1/\alpha,u/\alpha}$ (blue).

1.3 Characterization of maximum domains of attraction

The analysis of extreme values requires the characterization of belonging to a given limit distribution to one of the three previously domains of attraction. For this purpose we will limit us here to giving simple (necessary and sufficient) conditions on the distribution function F so that it belongs to a given domain of attraction. We also recall the concepts of slowly varying functions and regularly varying functions which are prerequisites for this characterization.
Next for any nondecreasing function f we introduce its generalized inverse :

$$f^{\leftarrow}(x) := \inf\{y,\ f(y) \geq x\}, \tag{1.7}$$

where by convention $\inf\{\emptyset\} = \infty$.

1.3.1 Slowly varying functions

Definition 1.1 *A positive Lebesgue measurable function ℓ is slowly varying at infinity if for all $x > 0$,*

$$\lim_{x\to\infty} \frac{\ell(\lambda x)}{\ell(x)} = 1, \ \textit{for all } \lambda > 0.$$

Typical examples of slowly varying functions at infinity are positives constants or functions converging to a positive constant, logarithms and iterated logarithms, etc.
The following theorem gives a representation of slowly varying functions called **Karamata's representation**.

Theorem 1.3 ***(Resnick [95, Corollary of Theorem 0.6])*** *The function ℓ is slowly varying if and only if it can be represented as :*

$$\forall x \geq 1,\ \ell(x) = c(x) \exp \int_1^x \frac{\epsilon(u)}{u}\, du,$$

where c and ϵ are mesurable functions such that $c(x) \to c_0 > 0$ and $\epsilon(x) \to 0$ as $x \to \infty$.

If the function c is a constant, ℓ is said to be normalized. When the function ℓ is normalized, it is differentiable :

$$\ell'(x) = \frac{\epsilon(x)\ell(x)}{x},\ \ x > 0.$$

Especially,

$$\lim_{x\to\infty} \frac{x\ell'(x)}{\ell(x)} = 0.$$

We end this part with reminders of some fundamental properties of slowly varying functions by referring the reader to Bingham et al. [16] for more details.

Proposition 1.1 ***(Bingham et al. [16, Proposition 1.3.6])*** *Let ℓ, ℓ_1 and ℓ_2 be slowly varying functions at infinity :*

1.
$$\lim_{x\to\infty} \frac{\log \ell(x)}{\log x} = 0.$$

2. *For all $\gamma > 0$,*
$$\lim_{x\to\infty} x^{\gamma}\ell(x) = \infty \text{ and } \lim_{x\to\infty} x^{-\gamma}\ell(x) = 0.$$

3. *For all $\alpha \in \mathbb{R}$, ℓ^{α} varies slowly.*

4. *$\ell_1 + \ell_2$ and $\ell_1\ell_2$ vary slowly. If moreover $\lim_{x\to\infty} \ell_2(x) = \infty$, then the composition $\ell_1 \circ \ell_2$ varies slowly.*

1.3.2 Regularly varying functions

Definition 1.2 *A positive Lebesgue measurable function f is regularly varying with index $\rho \in \mathbb{R}$ at infinity if*
$$\lim_{x\to\infty} \frac{f(\lambda x)}{f(x)} = \lambda^{\rho}, \text{ for all } \lambda > 0.$$

This property is denoted by $f \in \mathcal{RV}_{\rho}$. If $\rho = 0$, f is slowly varying. Any regularly varying function f with index ρ can be decomposed as $f(x) = x^{\rho}\ell(x)$, where ℓ is a slowly varying function.

Proposition 1.2 ***(Resnick [95, Proposition 0.5])*** *If f is a regularly varying function with index ρ then for all $0 < a < b$*
$$\lim_{x\to\infty} \sup_{\lambda\in[a,b]} \left| \frac{f(\lambda x)}{f(x)} - \lambda^{\rho} \right| = 0.$$

Proposition 1.2 establishes the local uniform convergence of a regularly varying function. Propositions 1.3, 1.4 and 1.5 list other properties of regularly varying functions that may be useful for the following.

Proposition 1.3

1. *If $f \in \mathcal{RV}_{\rho}$, then*
$$\lim_{x\to\infty} \frac{\log f(x)}{\log x} = \rho.$$
This implies,
$$\lim_{x\to\infty} f(x) = \begin{cases} 0 & \text{if } \rho < 0 \\ \infty & \text{if } \rho > 0. \end{cases}$$

2. *If* $f_1 \in \mathcal{RV}_{\rho_1}$, $f_2 \in \mathcal{RV}_{\rho_2}$, *then* $f_1 + f_2 \in \mathcal{RV}_{\max(\rho_1,\rho_2)}$. *If moreover* $\lim_{x\to\infty} f_2(x) = \infty$, *then the composition* $f_1 \circ f_2 \in \mathcal{RV}_{\rho_1\rho_2}$.
3. *If* $f \in \mathcal{RV}_\rho$ *and* $\alpha \in \mathbb{R}$, *then* $f^\alpha \in \mathcal{RV}_{\alpha\rho}$.
4. *If* $f \in \mathcal{RV}_\rho$, *then for all* $\epsilon > 0$, *there exists* t_0 *such that for all* $x \geq 1$ *and* $t \geq t_0$,

$$(1-\epsilon)x^{\rho-\epsilon} < \frac{f(tx)}{f(t)} < (1+\epsilon)x^{\rho+\epsilon}. \quad \textbf{\textit{(Potter Bounds)}}$$

The proof of Proposition 1.3 can be found in de Haan and Ferreira [31, Proposition B.1.9], Bingham et al. [16, Proposition 1.5.7] or (Resnick [95] Proposition 0.8).

Proposition 1.4 *Suppose* $f \in \mathcal{RV}_\rho$ *and* (u_n), (v_n) *satisfy,* $0 < u_n \to \infty$, $0 < v_n \to \infty$. *If* $u_n \sim v_n$, *then* $f(u_n) \sim f(v_n)$.

Proposition 1.5 ***(Inverse of a regularly varying function)***

- *If* $f \in \mathcal{RV}_\rho$, $\rho > 0$, *then* $f^{\leftarrow} \in \mathcal{RV}_{1/\rho}$.
- *If* $f \in \mathcal{RV}_\rho$, $\rho < 0$, *then* $f^{\leftarrow}(1/\cdot) \in \mathcal{RV}_{-1/\rho}$.

We refer to Bingham et al. [16, Theorem 1.5.12] for a proof of Proposition 1.5.

1.3.3 Fréchet domain of attraction

Theorem 1.4 ***(de Haan and Ferreira [31, Theorem 1.2.1])*** *F* $\in$ *DA(Fréchet) with index* $\gamma > 0$ *if and only if* $\tau_F = \infty$ *and*

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}, \ \textit{for all } x > 0. \tag{1.8}$$

Definition 1.2 and Theorem 1.4 yield that $F \in$ DA(Fréchet) if and only if $\tau_F = \infty$ and $\overline{F} \in \mathcal{RV}_{-1/\gamma}$, $\gamma > 0$. The following corollary provides an equivalent characterization to that of the Theorem 1.4 based on the tail quantile function defined by :

$$U(t) := F^{\leftarrow}(1-1/t) = \overline{F}^{\leftarrow}(1/t), \ t > 1. \tag{1.9}$$

Corollary 1.1 ***(de Haan and Ferreira [31, Corollary 1.2.10])*** *F* $\in$ *DA(Fréchet) with index* $\gamma > 0$ *if and only if* $\tau_F = \infty$ *and*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}, \ \textit{for all } x > 0. \tag{1.10}$$

In other words, $F \in$ DA(Fréchet) with index $\gamma > 0$ if and only if $\tau_F = \infty$ and U is regularly varying with index γ. In this case, Corollary 1.2.4 in de Haan and Ferreira [31, page 21] specifies a possible choice of the normalizing constants a_n and b_n :

$$a_n = U(n) \text{ and } b_n = 0.$$

Example 1.1 (Pareto distribution)
Consider a standard Pareto distribution $F(x) = 1 - x^{-1/\gamma}$, $x \geq 1$, $\gamma > 0$ with $\tau_F = \infty$. First,

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

Therefore, $1 - F = \overline{F} \in \mathcal{RV}_{-1/\gamma}$ and $F \in \text{DA(Fréchet)}$ with index $\gamma > 0$. Second, since a possible choice of the normalizing constants is

$$a_n = F^{\leftarrow}(1 - 1/n) = \left(1 - \left(1 - n^{-1}\right)\right)^{-\gamma} = n^{\gamma} \text{ and } b_n = 0,$$

one can write :

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} F^n\left(n^{\gamma} x\right) = \lim_{n\to\infty} \left(1 - n^{-1} x^{-1/\gamma}\right)^n = \exp\left(-x^{-1/\gamma}\right) = \Phi_{1/\gamma}(x),$$

which yields that the asymptotic distribution of the normalized maximum belongs to the Fréchet maximum domain of attraction.

In the literature of extreme value theory, the property (1.8) (respectively (1.10)) is often called **first-order condition**. The distributions of Fréchet domain of attraction are also called Pareto type distributions. They are applied in various fields such as meteorology (Gardes and Girard [54] and El Methni et al. [47]), hydrology (Anderson and Meerschaert [2] and El Methni et al. [46]) and finance (Bouchaud and Potters [18]), among others.

1.3.4 Weibull domain of attraction

Theorem 1.5 *$F \in DA(Weibull)$ with index $\gamma < 0$ if and only if $\tau_F < \infty$ and the function $(1 - F^*)$ is regularly varying with index $1/\gamma$, where*

$$F^*(x) = F\left(\tau_F - x^{-1}\right) \mathbb{1}_{\{x>0\}}.$$

Besides, a possible choice of the normalizing constants is :

$$a_n = \tau_F - F^{\leftarrow}(1 - 1/n) \quad et \quad b_n = \tau_F\,.$$

Theorem 1.5 shows that $F \in \text{DA(Weibull)}$ with index $\gamma < 0$ if and only if $\tau_F < \infty$ and $\overline{F}(x) = (\tau_F - x)^{-1/\gamma} \ell\Big((\tau_F - x)^{-1}\Big)$ for some slowly varying function ℓ. It is also interesting to note that the Fréchet and the Weibull domains of attraction are closely related (see Gnedenko [61] or Resnick [95]).

Example 1.2 (Uniform distribution)
Consider the uniform distribution $F(x) = x$ for $x \in [0, 1]$. First, in this case, $\tau_F = 1$ and

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

Therefore, $(1 - F^*)$ is regularly varying with index -1, then $F \in \mathrm{DA}(\text{Weibull})$ with index $\gamma = -1 < 0$. Second, taking into account this choice of the normalizing constants :

$$a_n = \tau_F - F^{\leftarrow}(1 - 1/n) = 1 - \left(1 - n^{-1}\right) = n^{-1} \text{ et } b_n = \tau_F = 1,$$

one can write :

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} F^n\left(n^{-1}x + 1\right) = \lim_{n\to\infty} \left(1 + n^{-1}x\right)^n = \exp(x) = \Psi_1(x),$$

which yields that the asymptotic distribution of the normalized maximum belongs to the Weibull maximum domain of attraction.

Applications to this domain of attraction include, among others, the work of Aarssen and de Haan [1] on the maximal life span of humans, or those of Hall [72], Falk [50] and Girard et al. [60] on the estimation of the endpoint in the Weibull domain of attraction.

1.3.5 Gumbel domain of attraction

The Gumbel domain of attraction encompasses a wide range of distributions that are difficult to obtain a simple characterization.

Theorem 1.6 *The distribution function F belongs to the Gumbel domain of attraction if and only if there exists $x_0 < \tau_F \leq \infty$ such that*

$$F(x) = 1 - c(x) \exp\left(-\int_{x_0}^{x} \frac{g(t)}{\tilde{a}(t)}\, dt\right),\ x_0 < x < \tau_F\,,$$

where a, c and g are three functions verifying $c(x) \to c > 0$, $g(x) \to 1$ and $\tilde{a}'(x) \to 0$ as $x \to \tau_F$. A possible choice of the normalizing constants is :

$$a_n = F^{\leftarrow}(1 - 1/n) \text{ and } b_n = \tilde{a}(a_n).$$

In practice, the characterization given in Theorem 1.6 is difficult to implement which denotes the complexity of the Gumbel domain of attraction who covers both distributions with finite

endpoint and ones with infinite endpoint. We refer to Resnick [95] and de Haan and Ferreira [31] for more details on the characterization of this domain of attraction.
The distributions of Gumbel domain of attraction are also called exponential-type distributions and they are used in the modeling of large claims in life insurance (Beirlant and Teugels [7]) or in hydrology (Gumbel [68; 69; 70] and de Haan [30]), etc.

1.3.6 General characterization of maximum domains of attraction

The following characterization common to the three previous domains of attraction relies on the parametrization (1.4) and the tail quantile function U (see relation (1.9)).

Theorem 1.7 ***(de Haan and Ferreira [31, Theorem 1.1.6])*** *For $\gamma \in \mathbb{R}$ the following statements are equivalent :*

1. *There exist real constants $a_n > 0$ and b_n real such that*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \mathscr{H}_\gamma(x) = \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right), \quad \textit{for all } x \textit{ with } 1+\gamma x > 0.$$

2. *There is a positive function a such that for $x > 0$,*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \textit{if } \gamma \neq 0 \\ \log x & \textit{if } \gamma = 0. \end{cases} \tag{1.11}$$

3. *There exists a positive function f such that x for which $1+\gamma x > 0$,*

$$\lim_{t\uparrow\tau_F} \frac{1 - F(t + xf(t))}{1 - F(t)} = \begin{cases} (1+\gamma x)^{-1/\gamma} & \textit{if } \gamma \neq 0 \\ e^{-x} & \textit{if } \gamma = 0. \end{cases} \tag{1.12}$$

Remark 1.3

1. If $\gamma > 0$, the condition (1.11) is equivalent to the first-order condition (1.10). Similarly, the condition (1.12) is equivalent to (1.8). In this case, a possible choice of the function a is $a(t) = U(t)$.
2. If $\gamma = 0$, the condition (1.11) implies that the function U is slowly varying. In other words, if $F \in \text{DA(Gumbel)}$ then U is a slow-varying function. In this case, a possible choice of the function a is $a(t) = \int_{U(t)}^{\tau_F} t(1-F(s))\,ds$.
3. If $\gamma < 0$, a possible choice of the function a is $a(t) = \tau_F - U(t)$.
4. A possible choice of the function f in (1.12) is $f(t) = a\,(1/(1-F(t))$, where a is the function given in (1.11).

1.4 Estimation of the extreme value index

Both the GEV approach of Paragraph 1.2.1 and the POT approach of Paragraph 1.2.2 show that the distributions of extreme values are indexed by a parameter γ (extreme value index) also called shape parameter. This parameter plays a central role in the behaviour of the shape of the distribution tail. It must then be estimated in order to better understand the nature of the extreme distribution in question.

In the literature of extreme value theory, several methods for estimating this parameter are proposed by the authors. The most widely used in practice are the Hill [74] estimator, the Pickands [90] estimator and the Dekkers et al. [36] moment estimator. One can also mention the methods based on the QQ-plot (Kratz and Resnick [82]), the graph of the mean of the excesses (Beirlant et al. [8]) and those based on the maximum likelihood (Prescott and Walden [91; 92] and Smith [100]). Estimators based on the methods of moments and weighted moments have also been introduced by Hosking et al. [76] and Hosking and Wallis [75]. As this list is not exhaustive, we refer to Embrechts et al. [48] for a more complete discussion of estimation methods for extreme value theory models.

We recall below the three most frequently used estimators of the extreme value index and their asymptotic properties.

1.4.1 The Hill estimator

The Hill [74] estimator is defined by :

$$\begin{aligned}\hat{\gamma}^{H}_{k_n} &= \frac{1}{k_n}\sum_{j=1}^{k_n} j(\log X_{n-j+1,n} - \log X_{n-j,n}) \\ &= \frac{1}{k_n}\sum_{j=1}^{k_n} \log X_{n-j+1,n} - \log X_{n-k_n,n},\end{aligned}$$

where $X_{1,n} \leq \ldots \leq X_{n,n}$ are the associated order statistics to the sample $X_1, \ldots, X_n$ and k_n is the number of the top order statistics (number of extremes) used for the estimation of γ. The construction of this estimator is based on the maximum likelihood method and the result of the Corollary 1.1 on the tail quantile function U in the case of a Pareto type distribution ($\gamma > 0$). The simplicity of the Hill estimator and the ease of its graphical interpretation, among other reasons, have made it very popular. Indeed, in the case of a Pareto distribution, the plot of the coordinates $\left(\log \frac{n+1}{j}, \log X_{n-j+1,n}, j = 1, \ldots, k_n\right)$ so-called "Pareto quantile plot" would be approximatively linear with a slope γ for small values of j, i.e. the extreme points. The Hill estimator is then an estimator of this slope. Figure 1.6 illustrates the quality of this estimate for a sample of size $n = 500$ from a Fréchet distribution with extreme value index $\gamma = 1$ and for $k_n = 100$.

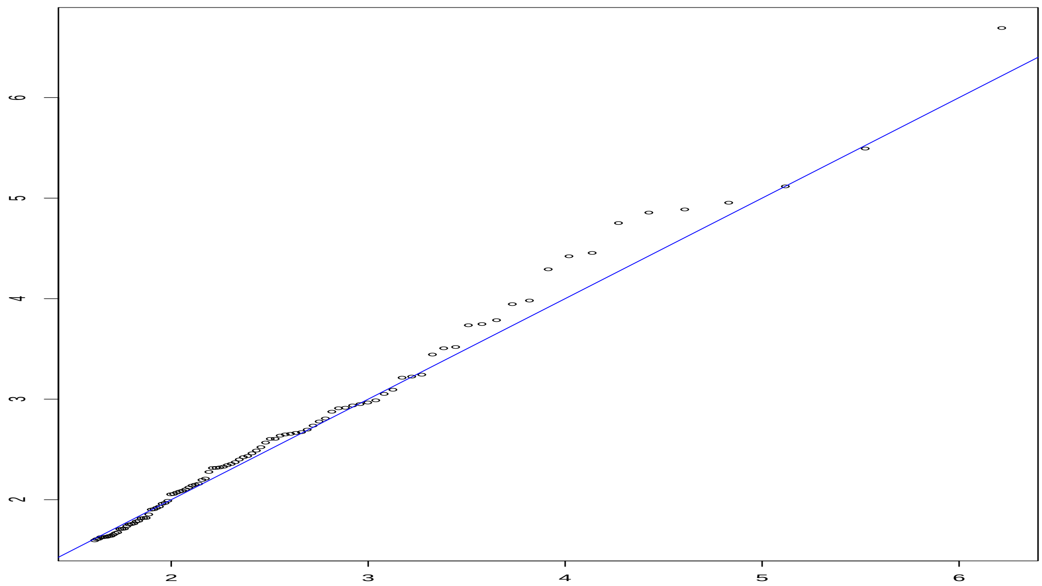

Figure 1.6: Pareto quantile plot. X-axis : $\log((n+1)/j)$. Y-axis : The line of slope γ $(y = x)$ and $\log(X_{n-j+1,n})$ for $j = 1, \dots, 100$.

Other construction methods of Hill estimator are developed in Beirlant et al. [10] and de Haan and Ferreira [31], and Diop and Lo [40] have proposed its generalized form. Significant work has also been carried out on the study of its properties. Its weak consistency has been established by Mason [85] and its strong consistency by Deheuvels et al. [33]. Asymptotic normality was established by Davis and Resnick [27], Haeusler and Teugels [71], Csörgő and Mason [24], Smith [101] and de Haan and Resnick [32], among others.

Theorem 1.8 ***(Consistency of Hill estimator)*** *Suppose $F \in DA$(Fréchet) with extreme value index $\gamma > 0$ and let $(k_n)_{n \geq 1}$ be a sequence of integers such that $1 \leq k_n < n$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.*

- *Then, $\hat{\gamma}^H_{k_n}$ converges in probability to γ.*
- *If moreover $k_n / \log\log n \to 0$ as $n \to \infty$, then $\hat{\gamma}^H_{k_n}$ converges almost surely to γ.*

More generally, the first-order condition (1.8) (or equivalently (1.10)) stated above is a sufficient condition to establish the consistency properties of extreme value estimators. To establish the asymptotic normality the so-called **second-order condition** is usually introduced :

There exist $\gamma > 0$, $\rho \leq 0$ and a positive or negative function A with $A(t) \to 0$ as $t \to \infty$ such that for all $x > 0$,

$$\lim_{t \to \infty} \frac{1}{A(t)} \left(\frac{U(tx)}{U(t)} - x^\gamma \right) = \begin{cases} x^\gamma \frac{x^\rho - 1}{\rho} & \text{if } \rho < 0 \\ x^\gamma \log x & \text{if } \rho = 0. \end{cases} \tag{1.13}$$

This condition is satisfied for most of the distributions belonging to the Fréchet domain of attraction. The function A is called a second-order auxiliary function and allows to control the rate of convergence of $U(tx)/U(t)$ to x^{γ}. Moreover, it can be shown that $|A|$ is regularly varying with index ρ (see Geluk and de Haan [59]). Thus, the parameter ρ referred to as the second-order parameter plays an important role tuning the rate of convergence of most extreme value estimators (see de Haan and Ferreira [31, Chapter 3] for examples). About the estimation of the parameter ρ and the auxiliary function A one can consult Beirlant et al. [9], Gomes et al. [65] and Deme et al. [38], for example.

Remark 1.4

The second-order condition (1.13) can also be formulated in terms of the distribution function F as follows :

There exist $\gamma > 0,\ \rho \leq 0$ and a positive or negative function A^* with $A^*(t) \to 0$ as $t \to \infty$ such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{1}{A^*(t)} \left(\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma} \right) = \begin{cases} x^{-1/\gamma} \frac{x^{\rho}-1}{\rho} & \text{if } \rho < 0 \\ x^{-1/\gamma} \log x & \text{if } \rho = 0. \end{cases} \tag{1.14}$$

In this case, $A^*(t) = A(1/(1-F(t))$ and $|A^*| \in \mathcal{RV}_{\rho/\gamma}$.

The following theorem establishes the asymptotic normality of Hill estimator.

Theorem 1.9 ***(de Haan and Ferreira [31, Theorem 3.2.5])*** *Suppose $F \in DA$(Fréchet) with extreme value index $\gamma > 0$ and let $(k_n)_{n\geq 1}$ be a sequence of integers such that $1 \leq k_n < n$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. If the condition (1.13) (or equivalently the condition (1.14)) is satisfied with $\sqrt{k_n}A(n/k_n) \to \lambda < \infty$ as $n \to \infty$, then*

$$\sqrt{k_n}\left(\hat{\gamma}^H_{k_n} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(\lambda/(1-\rho), \gamma^2\right).$$

Theorem 1.9 shows that the asymptotic variance of Hill estimator is γ^2/k_n (decreasing function of k_n) and the asymptotic bias is $A(n/k_n)/(1-\rho)$ (increasing function of k_n). Thus, for small values of k_n, the estimator $\hat{\gamma}^H_{k_n}$ is based on few observations which leads to a large variance. Conversely, for large values of k_n the estimated threshold $X_{n-k_n+1,n}$ is small so that the survival function is no longer approximatively a power function, so $\hat{\gamma}^H_{k_n}$ has a large bias (see Figure 1.7). It thus appears that the choice of the number k_n of the order statistics allowing to establish a balance between the bias and the variance remains crucial in practice. For more details on the selection of k_n, we refer to Dekkers and de Haan [35], Beirlant et al. [8], Dress and Kaufmann [43], Matthys and Beirlant [86] and Beirlant et al. [9].

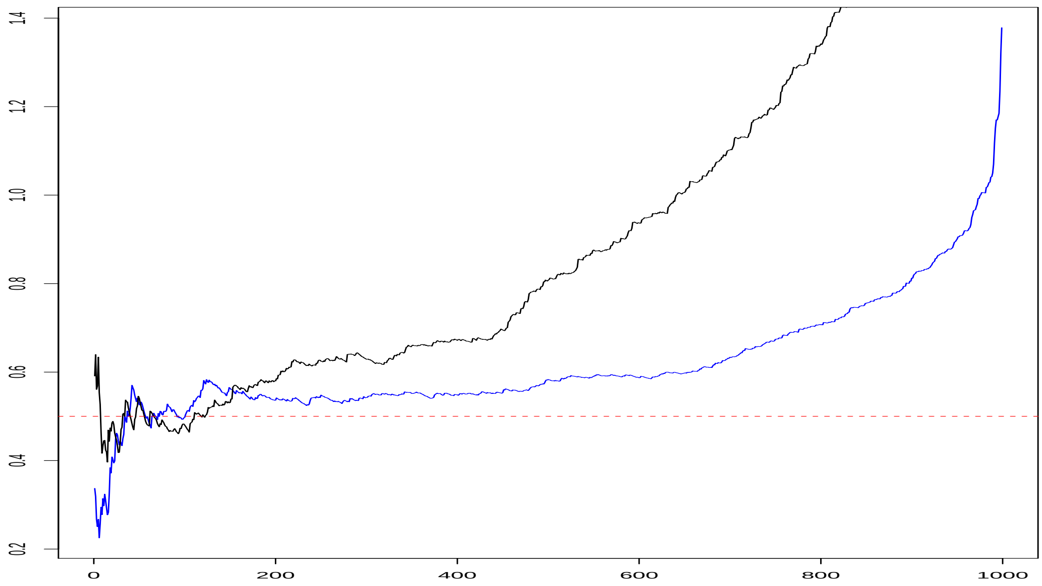

Figure 1.7: Hill estimator as a function of k_n for $n = 1000$ observations from a Fréchet dstribution (blue) and a Student disribution with 2 df (black), $\gamma = 1/2$ (red).

Since for a standard Pareto distribution the slowly varying function is constant the asymptotic bias of Hill estimator is null ($\lambda = 0$), which allows to choose k_n as large as one wants but also to easily give a confidence interval for the estimation (see Figure 1.8).

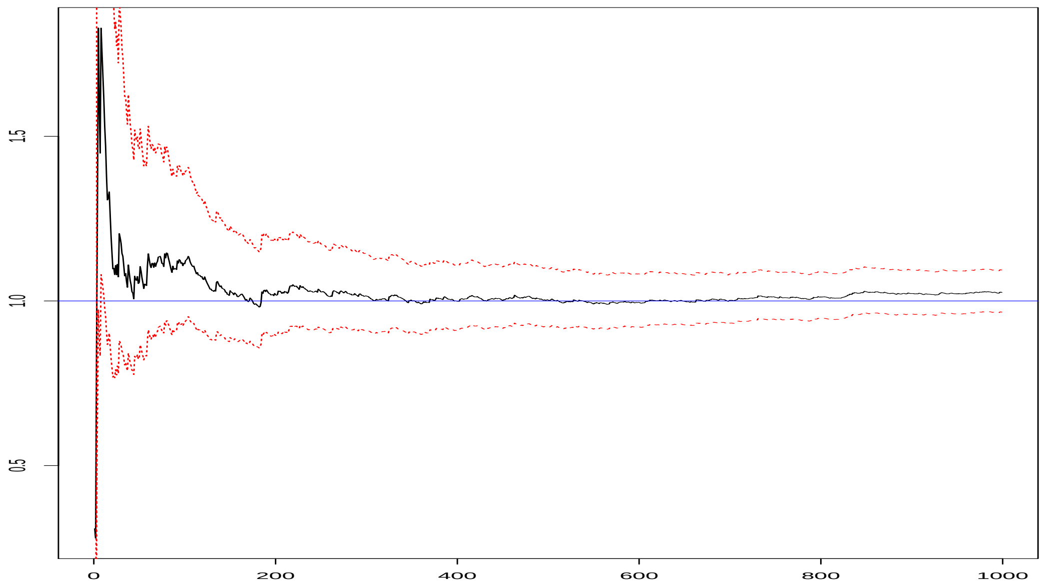

Figure 1.8: Hill estimator as a function of k_n for $n = 1000$ observations from a standard Pareto distribution (black) and its 95% confidence interval (red), $\gamma = 1$ (blue).

1.4.2 The Pickands estimator

Introduced by Pickands [90], this estimator is defined by :

$$\hat{\gamma}^P_{k_n} = \frac{1}{\log 2} \log \frac{X_{n-k_n+1,n} - X_{n-2k_n+1,n}}{X_{n-2k_n+1,n} - X_{n-4k_n+1,n}}, \; k_n = 1, \ldots, \lfloor n/4 \rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

Contrary to the Hill estimator, the Pickands estimator has the advantage of being applicable for any $\gamma \in \mathbb{R}$. On the other hand, since it uses only three order statistics (whereas Hill estimator uses k_n), it has a large asymptotic variance. Also, the maximum of the sample $X_{n,n}$ is not used, which constitutes a loss of information on the distribution tail. Theorem 1.10 establishes the asymptotic properties of Pickands estimator.

Theorem 1.10 *Suppose $F \in DA(\mathscr{H}_\gamma)$, $\gamma \in \mathbb{R}$ and let $(k_n)_{n\geq 1}$ be a sequence of integers such that $1 \leq k_n < n$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.*

- *Then, $\hat{\gamma}^P_{k_n}$ converges in probability to γ.*
- *If moreover $k_n/\log\log n \to 0$ as $n \to \infty$, then $\hat{\gamma}^P_{k_n}$ converges almost surely to γ.*
- *If the second-order condition (1.13) holds with $\sqrt{k_n}A(n/k_n) \to \lambda \in \mathbb{R}$ as $n \to \infty$, then*

$$\sqrt{k_n}\left(\hat{\gamma}^P_{k_n} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(\lambda b_{\gamma,\rho}, \sigma^2_\gamma\right),$$

with

$$b_{\gamma,\rho} := \begin{cases} 1 & \text{if } \rho = 0 \\ \frac{1-2^{1-\rho}+4^{-\rho}}{\rho^2(\log 2)^2} & \text{if } \rho < 0 = \gamma \\ \frac{4^{-\rho}\gamma\left(4^{\gamma+\rho}-1-(2^\gamma+1)\left(2^{\gamma+\rho}-1\right)\right)}{2^\gamma\rho(\gamma+\rho)(2^\gamma-1)\log 2} & \text{if } \rho < 0 \neq \gamma \end{cases}$$

and

$$\sigma^2_\gamma := \begin{cases} \frac{3}{4(\log 2)^4} & \text{if } \gamma = 0 \\ \frac{\gamma^2\left(2^{2\gamma+1}+1\right)}{4(\log 2)^2(2^\gamma-1)^2} & \text{if } \gamma \neq 0. \end{cases}$$

For more details we refer to Dekkers and de Haan [34, Theorems 2.1–2.3] or de Haan and Ferreira [31, Theorems 3.3.1 and 3.3.5].

Figure 1.9 illustrates Pickands estimator in the case of a standard exponential distribution. It shows a large variance and an asymptotically zero bias.

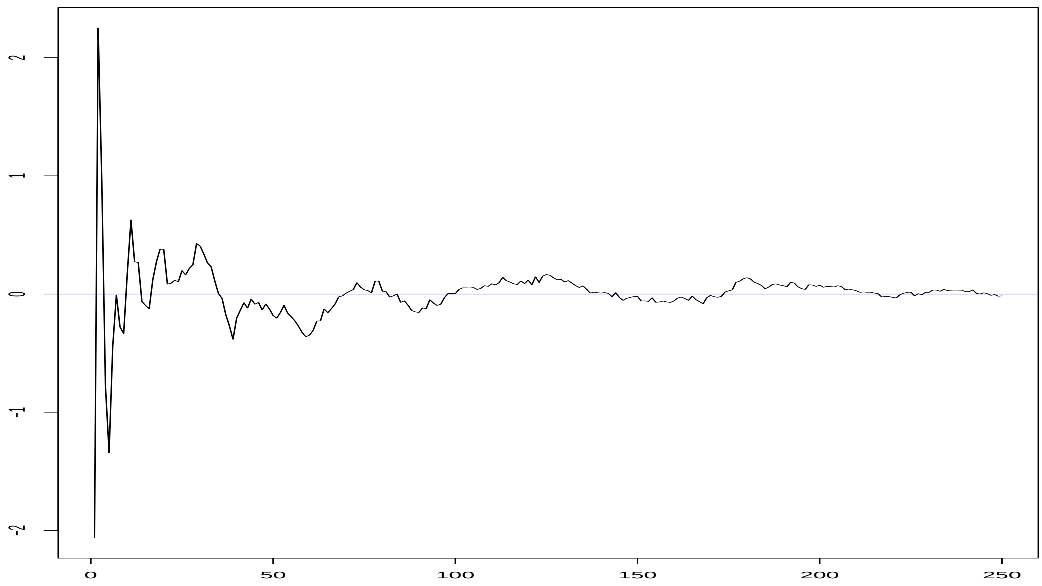

Figure 1.9: Pickands estimator as a function of k_n for $n = 1000$ observations from a standard exponential distribution (black), $\gamma = 0$ (blue).

1.4.3 The moment estimator

Proposed by Dekkers et al. [36], the moment estimator is defined by :

$$\hat{\gamma}_{k_n}^{M} = M_{k_n}^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_{k_n}^{(1)}\right)^2}{M_{k_n}^{(2)}}\right)^{-1}, \; 1 \leq k_n < n,$$

where $M_{k_n}^{(r)} = \frac{1}{k_n}\sum_{j=1}^{k_n}(\log X_{n-j+1,n} - \log X_{n-k_n,n})^r$, $r = 1, 2$. Note that $M_{k_n}^{(1)}$ corresponds to the Hill estimator. The naming of moment estimator is justified, moreover, by the fact that $M_{k_n}^{(r)}$ can be considered as empirical moments of the order r. The moment estimator is also known as the Dekkers-Einmahl-de Haan estimator. It is an extension of the Hill estimator to all domains of attraction. However, contrary to the Hill estimator, it is difficult to interpret graphically. The asymptotic properties of the estimator $\hat{\gamma}_{k_n}^{M}$ were established by Dekkers et al. [36].

Theorem 1.11 ***(Asymptotic properties of the moment estimator)*** *Suppose $F \in DA(\mathscr{H}_\gamma)$ with $\gamma \in \mathbb{R}$, $\tau_F > 0$ and let $(k_n)_{n\geq 1}$ be a sequence of integers such that $1 \leq k_n < n$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.*

- *Then, $\hat{\gamma}_{k_n}^{M}$ converges in probability to γ.*
- *If moreover $k_n/(\log n)^\delta \to \infty$ as $n \to \infty$ for some $\delta > 0$, then $\hat{\gamma}_{k_n}^{M}$ converges almost surely to γ.*

- *Under additional conditions on the distribution function F (see Dekkers et al. [36, Theorem 3.1 and Corollary 3.2]),*

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n}^{M}-\gamma\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma_\gamma^2\right),$$

where

$$\sigma_\gamma^2 := \begin{cases} 1+\gamma^2 & \text{if } \gamma \geq 0 \\ \frac{(1-\gamma)^2(1-2\gamma)\left(1-\gamma+6\gamma^2\right)}{(1-3\gamma)(1-4\gamma)} & \text{if } \gamma < 0. \end{cases}$$

For the proof of the consistency, one can see Dekkers et al. [36, Theorem 2.1] or de Haan and Ferreira [31, Theorem 3.5.2]. The proof of the asymptotic normality can be found in de Haan and Ferreira [31, Theorem 3.5.4].

Figure 1.10 illustrates the moment estimator in the cases of a standard uniform distribution.

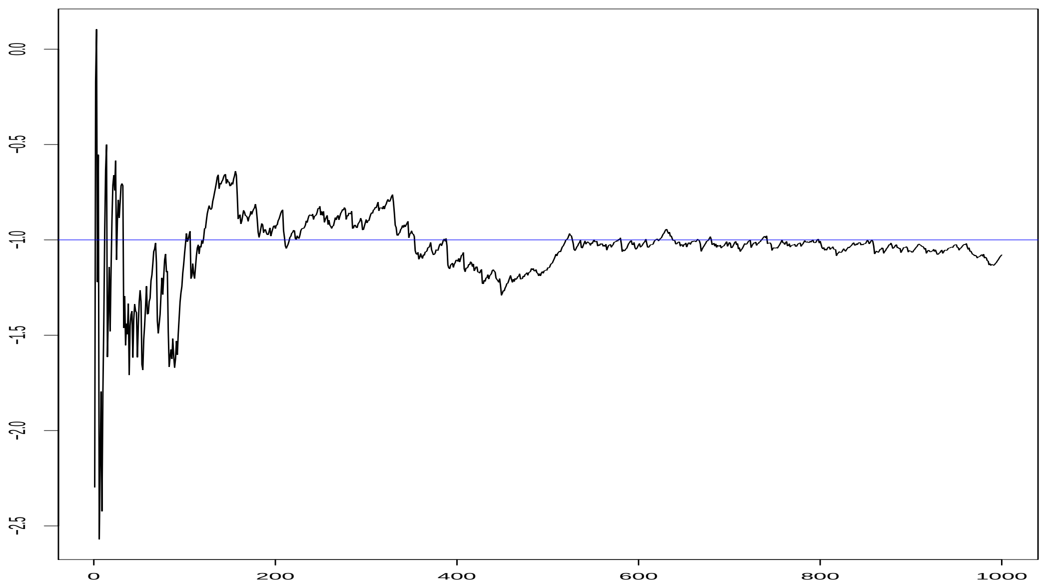

Figure 1.10: Moment estimator as function of k_n for $n = 1000$ observations from a standard uniform distribution (black), $\gamma = -1$ (blue).

A comparison of the asymptotic properties of these three estimators has been proposed in de Haan and Ferreira [31]. The Pickands estimator has a larger asymptotic variance than the others. The moment estimator has good properties when γ is close to 0. On the other hand, the Hill estimator has a lower asymptotic variance than the Pickands estimator and the moment estimator (in the case $\gamma > 0$).

Some authors including Gomes and Martins [62], Caeiro et al. [20], Gomes et al. [66] and Diebolt et al. [39], for example, have also looked at techniques to reduce the bias of extreme value index estimators, in particular, through the estimation of the second-order parameter

ρ. As mentioned above, the difficulty lies in choosing the number k_n of the order statistics to be used in order to establish a balance between bias and variance.

1.5 Estimation of the extreme quantiles

In Section 1.1 of this chapter, we have literally addressed the notion of extreme quantiles by introducing a problem that attempts to evaluate quantities whose probability of occurrence is low (close to zero). We shall now formally define the notion of quantiles and extend it to that of extreme quantiles before tackling their estimation.

Definition 1.3 *The quantile of order α associated with the survival function $\overline{F}$ is defined by :*

$$q(\alpha) := \overline{F}^{\leftarrow}(\alpha) = \inf\left\{x,\ \overline{F}(x) \leq \alpha\right\} \quad \text{with} \quad \alpha \in (0,1), \tag{1.15}$$

where $\overline{F}^{\leftarrow}$ is the generalized inverse of $\overline{F}$ (see relation (1.7)).

When the order of the quantile is a sequence α_n tending to 0 as the sample size increases this quantile is said to be extreme and the speed of convergence of α_n to 0 is then crucial for such quantiles estimation. Indeed, one may wonder with what probability the maximum of the sample is exceeded by an extreme quantile. When the random variables X_i are independent and identically distributed and α_n tends to 0, this probability is given by :

$$\begin{aligned}
\mathbb{P}\left(X_{n,n} < q(\alpha_n)\right) &= \mathbb{P}\left(\bigcap_{i=1}^{n}\left\{X_i \leq q(\alpha_n)\right\}\right) \\
&= \prod_{i=1}^{n}\mathbb{P}\left(X_i \leq q(\alpha_n)\right) \\
&= F^n\left(q(\alpha_n)\right) \\
&= \left(1-\alpha_n\right)^n \\
&= \exp\left(n\log(1-\alpha_n)\right) \\
&= \exp\left(-n\alpha_n(1+o(1))\right).
\end{aligned}$$

It is clear that the desired probability depends on the asymptotic behavior of the quantity $n\alpha_n$.

Intermediate case : If $n\alpha_n \to \infty$, then $\mathbb{P}\left(X_{n,n} < q(\alpha_n)\right) \to 0$. Therefore, the quantile to be estimated is with a high probability within the sample. This quantile is referred to as intermediate quantile, and in such a case α_n tends slowly to 0. Clearly the quantile $q(\alpha_n)$ does not tend to infinity too quickly as n goes to infinity. A natural estimator of this quantile is the $\lfloor n\alpha_n \rfloor$th largest observation of the sample :

$$\hat{q}(\alpha_n) = X_{n-\lfloor n\alpha_n \rfloor,n}. \tag{1.16}$$

The estimator (1.16) is obtained by inverting the empirical distribution function defined by :

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}},$$

i.e. :

$$\hat{q}(\alpha_n) = F_n^{\leftarrow}(1-\alpha_n) = \inf\{x,\ F_n(x) \ge 1-\alpha_n\}.$$

Under some assumptions, the estimator (1.16) is asymptotically Gaussian (see de Haan and Ferreira [31, Theorem 2.2.1]).

Extreme case : If $n\alpha_n \to 0$, then $\mathbb{P}(X_{n,n} < q(\alpha_n)) \to 1$. In this case, the quantile to be estimated is with a high probability outside the sample. Since $F_n(x) = 1$ for $x \ge X_{n,n}$, an inversion of the empirical distribution function cannnot still be used to estimate $q(\alpha_n)$. It is then necessary to extrapolate out of the sample in order to give a non-trivial estimate of $q(\alpha_n)$.
An illustration of the notions of intermediate and extreme quantiles is given in Figure 1.11 where the order of an intermediate quantile is noted β_n.

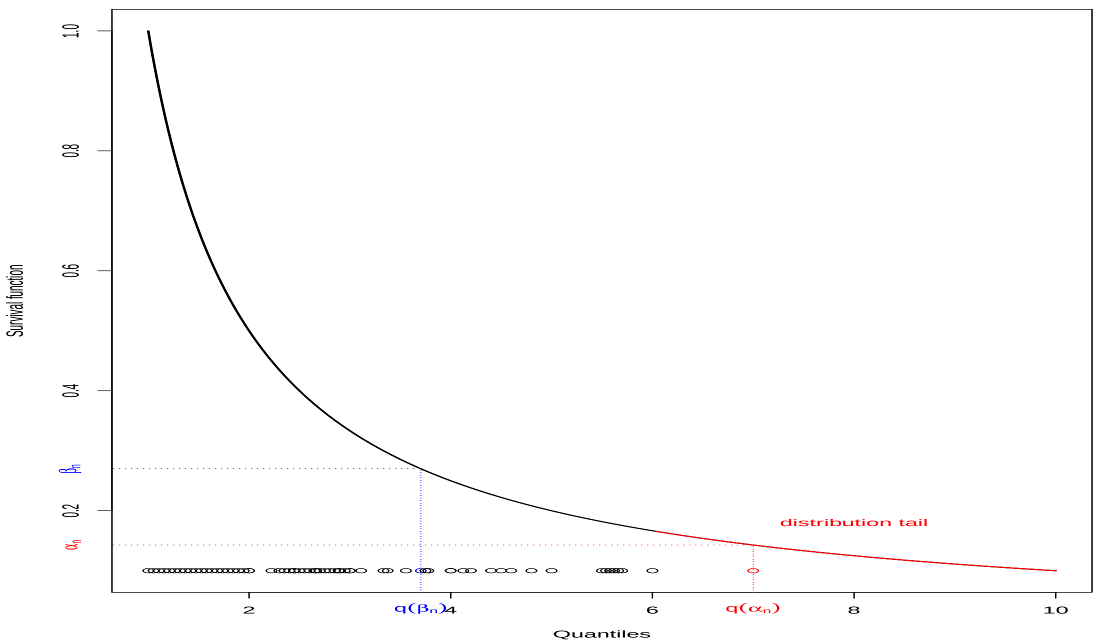

Figure 1.11: Illustration of the notions of distribution tail, intermediate quantile (blue point) and extreme quantile (red point). The black points represent the data.

Extreme value theory proposes three different approaches to the estimation of extreme quantiles. All these approaches are based on the results of the Theorems 1.1 and 1.2 as well as the characterizations of the domains of attraction introduced in the Paragraph 1.3. We present them in Paragraphs 1.5.1, 1.5.2 and 1.5.3 where it is assumed that the distribution function F belongs to one of the three domains of attraction previously introduced.

1.5.1 GEV approach

Taking into account the parametrization (1.4) the result of Theorem 1.1 yields, for n large enough, the following approximation :

$$\mathbb{P}\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \simeq \mathscr{H}_\gamma(x). \tag{1.17}$$

The relation (1.17) can be rewritten as :

$$\lim_{n\to\infty} n \log\Big(F(a_n x + b_n)\Big) = \lim_{n\to\infty} n \log\Big(1 - \bar{F}(a_n x + b_n)\Big) = \log\Big(\mathscr{H}_\gamma(x)\Big). \tag{1.18}$$

For n large enough, one can show that $a_n x + b_n \to \tau_F$, therefore, $\bar{F}(a_n x + b_n) \to 0$. A Taylor expansion of the logarithm allows to rewrite the relation (1.18) as follows :

$$\bar{F}(x) \simeq -\frac{1}{n} \log\left(\mathscr{H}_\gamma\left(\frac{x - b_n}{a_n}\right)\right). \tag{1.19}$$

Finally, by replacing $\mathscr{H}_\gamma$ with its expression (see (1.4)) in the relation (1.19)), we obtain an approximation of the distribution tail of F :

$$\bar{F}(x) \simeq \frac{1}{n}\left(1 + \gamma\frac{x - b_n}{a_n}\right)^{-1/\gamma}, \tag{1.20}$$

which is reduced, in case $\gamma = 0$, to :

$$\bar{F}(x) \simeq \frac{1}{n} \exp\left(-\frac{x - b_n}{a_n}\right). \tag{1.21}$$

By inverting the equation (1.20) we thus obtain an approximation of the extreme quantiles :

$$q(\alpha_n) \simeq b_n + \frac{a_n}{\gamma}\left(\left(\frac{1}{n\alpha_n}\right)^\gamma - 1\right). \tag{1.22}$$

In particular, for $\gamma = 0$, this approximation is reduced to :

$$q(\alpha_n) \simeq b_n - a_n \log(n\alpha_n). \tag{1.23}$$

Starting from the approximation (1.22) of the quantile function we obtain the estimate given in the following definition.

Definition 1.4 *The estimator of the extreme quantile based on the extreme value theorem is defined by :*

$$\hat{q}_n^{GEV}(\alpha_n) = \hat{b}_n + \frac{\hat{a}_n}{\hat{\gamma}_n}\left(\left(\frac{1}{n\alpha_n}\right)^{\hat{\gamma}_n} - 1\right), \tag{1.24}$$

where $\hat{a}_n$, $\hat{b}_n$ and $\hat{\gamma}_n$ are respectively estimators of the unknown parameters a_n, b_n and γ of the GEV distribution.

When $\gamma = 0$, the estimator of the extreme quantile based on the extreme value theorem is given by :

$$\hat{q}_n^{GEV}(\alpha_n) = \hat{b}_n - \hat{a}_n \log(n\alpha_n). \tag{1.25}$$

Not surprisingly, the estimation of the extreme quantiles requires the estimation of the GEV parameters. Gumbel [70] proposes within the framework of the estimation of these parameters a method called block maxima, the idea of which is based on the use of a complete sample of observations distributed according to a GEV distribution. To do this, the initial sample is subdivided into m disjoint blocks of equal size from which the maximum values are extracted. The distribution of these maxima is then approximated for m large enough by a distribution of extreme values. The parameters γ, a_n and b_n of this distribution can be estimated by the maximum likelihood method (Prescott and Walden [91; 92]) or by the methods of moments and weighted moments (Hosking et al. [76] and Hosking and Wallis [75]). For the asymptotic properties of these estimators, one can refer to the work of Smith [100], Zhou [108; 109], Dombry [42] and Ferreira and de Haan [51].
However it is preferable to use weighted moment estimators because they are often more explicit and easier to calculate but also because they give better results than maximum likelihood estimators for small or medium-sized samples.

1.5.2 GPD approach

This second approach of the estimation of extreme quantiles is based on the result of Theorem 1.2 which gives an approximation of the distribution of excesses above a given threshold. According to the relation (1.6), the change of variable $x = y + u$ yields :

$$\overline{F}(x) = \overline{F}(u)\overline{F}_u(x-u). \tag{1.26}$$

By introducing the probability p that X exceeds the threshold u, $p = \mathbb{P}(X > u) = \overline{F}(u)$ and using the result of Theorem1.2 for a sufficiently large threshold u we thus obtain the following approximation of the distribution tail :

$$\overline{F}(x) \simeq p\left(1 + \gamma\frac{x - \overline{F}^{\leftarrow}(p)}{\sigma}\right)^{-1/\gamma} \tag{1.27}$$

or, in the case $\gamma = 0$,

$$\bar{F}(x) \simeq p \exp\left(-\frac{x - \bar{F}^{\leftarrow}(p)}{\sigma}\right). \tag{1.28}$$

Recall that to have an approximation of the quantile function we need to invert the equations (1.27) and (1.28). We thus obtain :

$$q(\alpha_n) \simeq \bar{F}^{\leftarrow}(p) + \frac{\sigma}{\gamma}\left(\left(\frac{p}{\alpha_n}\right)^{\gamma} - 1\right) \tag{1.29}$$

or, for $\gamma = 0$,

$$q(\alpha_n) \simeq \bar{F}^{\leftarrow}(p) + \sigma \log\left(\frac{p}{\alpha_n}\right). \tag{1.30}$$

The threshold u given by $\bar{F}^{\leftarrow}(p)$ is an intermediate quantile that is easily estimated by inverting the empirical survival function. In practice, one chooses $p = p_n = k_n/n$, where k_n is the number of excess and one thus estimates $\bar{F}^{\leftarrow}(k_n/n)$ by $X_{n-k_n,n}$. The threshold u being chosen, it only remains to estimate the parameters γ and σ in order to obtain an estimator of the extreme quantile $q(\alpha_n)$.

Definition 1.5 *The estimator of the extreme quantile based on the GPD is defined by :*

$$\hat{q}_n^{GPD}(\alpha_n) = X_{n-k_n,n} + \frac{\hat{\sigma}_n}{\hat{\gamma}_n}\left(\left(\frac{k_n}{n\alpha_n}\right)^{\hat{\gamma}_n} - 1\right), \tag{1.31}$$

where $\hat{\gamma}_n$ and $\hat{\sigma}_n$ are respectively estimators of shape and scale parameters.

The parameters γ and σ of the GPD can be estimated by the maximum likelihood method (Smith [101] and Davison and Smith [29]) or by the methods of moments and weighted moments (Hosking and Wallis [75]).
This method has an advantage over the previous one in that it is easier to have an excess sample than a maximum sample. Two variants of this method have been presented by Breiman et al. [19] under the Exponential Tail (**ET**) and Quadratic Tail (**QT**) designations.

Remark 1.5
There is a similarity between the two expressions of the extreme quantile (1.22) and (1.29) from the GEV and GPD approaches respectively. There are three unknown parameters in each of them :

- The extreme value index γ ;
- The scale parameter σ which plays the role of a_n in the GEV approach ;
- The threshold $u = \bar{F}^{\leftarrow}(p)$ which plays the role of b_n in the GEV approach.

1.5.3 Semi-parametric approach

The semi-parametric approach is based on the characterization of the functions associated with a given domain of attraction in order to propose estimators of extreme quantiles. In this section we will restrict ourselves to Fréchet domain of attraction. This choice is based on the fact that we are mainly interested in heavy-tailed distributions and so to Weissman estimator which is, in the case of the semi-parametric approach, the main estimator of extreme quantiles based on a positive tail-index.
Recall that a distribution function F in Fréchet domain of attraction satisfies for any $\gamma > 0$:

$$\bar{F}(x) = x^{-1/\gamma}\ell(x) \ \text{ avec } \ \ell \in \mathcal{RV}_0\,.$$

By inverting the previous equation (see Definition 1.3) and using the result of the Proposition 1.5, the expression of the associated quantile function is obtained as follows :

$$q(\alpha) = \alpha^{-\gamma}L\left(\alpha^{-1}\right) \ \text{ with } \ L \in \mathcal{RV}_0\,. \tag{1.32}$$

The construction of Weissman [106] estimator is based on the equation (1.32). For all $\gamma > 0$,

$$q(\alpha_n) = \alpha_n^{-\gamma}L\left(\alpha_n^{-1}\right), \tag{1.33}$$

$$q(\beta_n) = \beta_n^{-\gamma}L\left(\beta_n^{-1}\right). \tag{1.34}$$

By dividing (1.34) by (1.33) and using Definition 1.1, we get, for β_n small enough and $\alpha_n < \beta_n$, the following approximation :

$$q(\alpha_n) \simeq q(\beta_n)\left(\frac{\beta_n}{\alpha_n}\right)^{\gamma}, \tag{1.35}$$

where $q(\beta_n)$ is an intermediate quantile. The underlying idea is to estimate the extreme quantile $q(\alpha_n)$ by extrapolating from the intermediate quantile $q(\beta_n)$, easy to estimate by inverting the empirical survival function.
Let remark that the approximation (1.35) is a particular case of the GPD approach with $\sigma = \gamma q(\beta_n)$.
Thus, replacing $q(\beta_n)$ and γ with appropriate estimators yields the Weissman estimator.

Definition 1.6 *The Weissman estimator is defined by :*

$$\hat{q}_n^W(\alpha_n) = \hat{q}_n(\beta_n)\left(\frac{\beta_n}{\alpha_n}\right)^{\hat{\gamma}_n}. \tag{1.36}$$

Weissman proposes to estimate $q(\beta_n)$ by its empirical equivalent $X_{n-\lfloor n\beta_n\rfloor,n}$ and γ by Hill's estimator (see Paragraph 1.4.1). The correction $(\beta_n/\alpha_n)^{\hat{\gamma}_n}$ is the extrapolation term. The properties of the Weissman estimator are established in Weissman [106], Embrechts et al. [48] and de Haan and Ferreira [31].

1.6 Estimation of the conditional extreme quantiles

In this section, we are interested in the estimation of extreme quantiles when the random variable of interest Y is recorded simultanuously with a covariate x. Depending on the nature of this covariate, two cases will be distinguished :

- The so-called "fixed design" setting whose data are pairs $\{(Y_i, x_i), i = 1, \ldots, n\}$ where the observations Y_i are independent real random variables and the x_i are non-random observation points.
- The so-called "random design" setting for which the data are copies $\{(Y_i, X_i), i = 1, \ldots, n\}$ of independent and identically distributed real random variables.

In both cases, the conditional distribution function of Y will be denoted by $F(\cdot|x)$ and its associated conditional survival function by $\overline{F}(\cdot|x) := 1 - F(\cdot|x)$. In this context, extreme quantiles are said to be conditional.

Definition 1.7 *The conditional extreme quantile of order α_n associated with the conditional survival function $\overline{F}(\cdot|x)$ is defined by :*

$$q(\alpha_n|x) := \overline{F}^{\leftarrow}(\alpha_n|x) = \inf\left\{t,\ \overline{F}(t|x) \leq \alpha_n\right\} \ \text{with} \ \alpha_n \to 0 \ \text{as} \ n \to \infty. \tag{1.37}$$

The same difficulties identified in the estimation of extreme quantiles in the case without covariate occur in the case with covariate beyond the maximum of the subsample of observations of Y taken in a neighbourhood of x. Moreover, in the covariate case the parameters of the GEV and GPD distributions depend on the covariate x. Therefore, to estimate the conditional extreme quantiles it will be necessary to use estimators of these parameters adapted to the conditional case. In the literature of extreme value theory there are three conditional extreme quantiles estimation approaches which we recall in Paragraphs 1.6.1 and 1.6.2 depending on the design setting. We will focus essentially on the nonparametric approach of estimation of the tail-index and extreme quantiles for heavy-tailed distributions and we will recall some estimators.

1.6.1 Estimation in a fixed design setting

Smith [102] was the first to model maxima by an extreme values distribution whose parameters are functions of the covariate. A parametric form is assumed on these functions and estimation is performed by maximum likelihood or least squares methods. Davison and Smith [29] then proposed parametric models based on excesses over a threshold. They

thus model these excesses by a GPD whose parameters are functions of the covariate, the estimation being based on the maximum likelihood method.
A semi-parametric approach has been introduced in Beirlant and Goegebeur [5] where the authors first propose a transformation of the data to obtain residuals following a Pareto-type distribution which they use in an exponential regression model where the parameters are estimated by the maximum likelihood method. In the case of a unidimensional covariate, Beirlant and Goegebeur [6] propose an adaptation of the quantile estimators proposed by Matthys and Beirlant [87] to the conditional case by replacing order statistics by quantiles estimated by the local polynomial method, see Koenker and Bassett [80]. Other semi-parametric models have also been proposed by Wang and Tsai [105].
A nonparametric estimation approach of conditional extreme quantiles has been introduced in Davison and Ramesh [28] where the authors propose polynomial fitting estimators. Beirlant et al. [10] extend these results to multidimensional covariates. Chavez-Demoulin and Davison [22] use the penalized maximum likelihood method to propose spline estimators of conditional extreme quantiles in the case of a one-dimensional covariate.
In the particular case of heavy-tailed distributions, a family of nonparametric tail-index estimators has been proposed by Gardes and Girard [53] using a moving window approach and extending to the conditional framework the estimators proposed in Beirlant et al. [9]. The authors consider a pair (Y, x) where Y is a real random variable of distribution function F assumed to be heavy-tailed ($\gamma(x) > 0$) and x is a deterministic covariate defined on a metric space E with a distance d. They propose to estimate $\gamma(t)$ for given $t \in E$ from independent observations $(Y_1, x_1), \ldots, (Y_n, x_n)$. To do this, they propose to use a selection method to retain only the variables Y_i noted $\{Z_i(t), i = 1, \ldots, m_t\}$ associated with the values x_i which are in the ball centered in t and of radius $h_{n,t}$, where $h_{n,t}$ is a positive non-random sequence tending to 0 as n tends to infinity. Denoting $Z_{1,m_t}(t) \leq \ldots \leq Z_{m_t,m_t}(t)$ the order statistics associated with the variables Z_i, the proposed family of estimators of $\gamma(t)$ is of the form :

$$\hat{\gamma}_n(t, W) = \sum_{i=1}^{k_{n,t}} i \log\left(\frac{Z_{m_t-i+1,m_t}(t)}{Z_{m_t-i,m_t}(t)}\right) W\left(\frac{i}{k_{n,t}}, t\right) \Bigg/ \sum_{i=1}^{k_{n,t}} W\left(\frac{i}{k_{n,t}}, t\right), \tag{1.38}$$

where $k_{n,t}$ is a sequence of integers such that $1 \leq k_{n,t} \leq m_t$ and $W(\cdot, t)$ is a function defined on $(0, 1)$ and such that $\int_0^1 W(s,t)\, ds \neq 0$. Under some conditions the authors establish the asymptotic normality of the estimator (1.38). Remark that when $W(s, t) = 1$ for all $s \in (0, 1)$, the estimator (1.38) is an adaptation of the Hill estimator (see Paragraph 1.4.1) to the conditional framework.
Gardes et al. [58] have also proposed a family of conditional tail-index kernel estimators of which a particular case is the estimator (1.38).
When the covariate is functional Gardes and Girard [54] have used a nearest neighbor

approach to propose estimators of conditional extreme quantiles. Gardes et al. [57] have proposed estimators of conditional extreme quantiles using the same approach as in Gardes and Girard [53]. In their paper, they propose to estimate the conditional extreme quantiles of order $1-\alpha_{m_t}$ ($q(\alpha_{m_t}, t)$, $\alpha_{m_t} \to 0$ as $n \to \infty$) in the intermediate case (see Section 1.5) by :

$$\hat{q}_1(\alpha_{m_t}, t) = Z_{m_t - \lfloor m_t \alpha_{m_t} \rfloor + 1, m_t}(t).$$

In the extreme case (see Section 1.5), they estimate the conditional quantiles by :

$$\begin{aligned} \hat{q}_2(\alpha_{m_t}, t) &= \hat{q}_1(\beta_{m_t}, t) \left(\beta_{m_t}/\alpha_{m_t}\right)^{\hat{\gamma}_n(t)} \\ &= Z_{m_t - \lfloor m_t \beta_{m_t} \rfloor + 1, m_t}(t) \left(\beta_{m_t}/\alpha_{m_t}\right)^{\hat{\gamma}_n(t)}, \end{aligned} \tag{1.39}$$

where β_{m_t} is the order of an intermediate extreme quantile and $\hat{\gamma}_n(t)$ is an estimator of the conditional tail-index. Remark that the estimator $\hat{q}_1(\alpha_{m_t}, t)$ above corresponds to the estimator (1.16) adapted to the conditional framework. Similarly, $\hat{q}_2(\alpha_{m_t}, t)$ is the Weissman estimator (see (1.36)) adapted to the conditional framework.

1.6.2 Estimation in a ramdom design setting

Semi-parametric methods have been proposed by Hall and Tajvidi [73] by combining a nonparametric trend regression model with a parametric model for extreme values. The nonparametric estimation of tail-index and conditional extreme quantiles in random design setting for heavy-tailed distributions has also been the subject of several works. In the case of a covariate of finite dimension d, Daouia et al. [25] propose a kernel estimation of the extreme quantiles. The authors consider independent copies $\{(Y_i, X_i), i = 1, \ldots, n\}$ of the random pair $(Y, X) \in \mathbb{R}^d$, where Y is a variable of interest associated with a covariate X. The conditional distribution function $\overline{F}(y|x)$ is assumed to be heavy-tailed. They first propose to estimate the conditional survival function by :

$$\hat{\overline{F}}_n(y|x) = \sum_{i=1}^{n} K_h(x - X_i) \mathbb{1}_{\{Y_i > y\}} \Big/ \sum_{i=1}^{n} K_h(x - X_i),$$

where $h = h_n$ is a non-random sequence such that $h_n \to 0$ as $n \to \infty$ and $K_h(\cdot) = K(\cdot/h)/h^d$, K (kernel) being a probability density on $\mathbb{R}^d$. They then estimate the intermediate conditional extreme quantiles (voir (1.37)) by :

$$\hat{q}_n(\alpha_n|x) = \hat{\overline{F}}_n^{\leftarrow}(\alpha_n|x) = \inf\left\{t,\ \hat{\overline{F}}_n(t|x) \le \alpha_n\right\}, \tag{1.40}$$

In the same paper they propose an adaptation of Pickands estimator (see Paragraph 1.4.2) to the conditional framework based on the estimator (1.40), of the form :

$$\hat{\gamma}_n^P(x) = \frac{1}{\log 2} \log \frac{\hat{q}_n(k/n|x) - \hat{q}_n(2k/n|x)}{\hat{q}_n(2k/n|x) - \hat{q}_n(4k/n|x)},$$

where $k = k_n$ is a sequence of integers such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. A Hill-type estimator of the tail-index adapted to the conditional framework was also proposed in the form :

$$\hat{\gamma}_n^H(x) = \sum_{j=1}^{J} (\log \hat{q}_n(\tau_j \alpha_n|x) - \log \hat{q}_n(\alpha_n|x)) \Big/ \sum_{j=1}^{J} \log(1/\tau_j),$$

where J is a positive integer and $1 = \tau_1 > \tau_2 > \ldots > \tau_J > 0$. To estimate the conditional extreme quantiles, they propose a Weissman-type estimator (see (1.36)) adapted to the conditional framework :

$$\hat{q}_n^W(\beta_n|x) = \hat{q}_n(\alpha_n|x)(\alpha_n/\beta_n)^{\hat{\gamma}_n(x)},$$

where $\hat{q}_n(\alpha_n|x)$ is the estimator of an intermediate extreme quantile and $\hat{\gamma}_n(x)$ is an estimator of $\gamma(x)$. The asymptotic normality of all the proposed estimators was established by the authors. This method of estimation of extreme quantiles was then generalized to all domains of attraction by Daouia et al. [26]. A conditional tail-index estimator was also proposed by Gardes and Stupfler [56] using a smoothed local Hill estimator.
When the covariate is functional and of infinite dimension, Gardes and Girard [55] have proposed kernel estimators of conditional extreme quantiles based on a functional Weissman estimator. They also define a large family of conditional tail-index estimators whose special cases correspond to functional versions of the Hill and Pickands estimators.

1.7 Extreme values and censoring

This section provides some reminders about the concept of censoring and the estimation of extreme values under censoring. We are introducing it in order to give a general idea of this concept in the perspective of extending our results to this.

1.7.1 Survival time

The term "survival time" refers to the time elapsed from an initial moment to the occurrence of a specific final event of interest. Examples include the time between diagnosis and recovery, the running time of a machine, the time between two disasters, etc. Survival data analysis is the study of the delay in the occurrence of this event.
Suppose that the survival time X is a positive and absolutely continuous variable of distribution function F. The variable X can also be characterized by its survival function $\overline{F}$ (see

(1.1)) :

$$\bar{F}(t) := 1 - F(t) = 1 - \mathbb{P}(X \leq t) = \mathbb{P}(X > t),\ t \geq 0, \tag{1.41}$$

which gives the probability of surviving until a fixed time t.
There are other distributions that can be used to characterize the distribution of X, such as the instantaneous hazard or risk function and the cumulative hazard function, for example.

1.7.2 Notion of censoring

One of the characteristics of survival data is the existence of incomplete observations. Indeed, they are often collected partially, notably because of the phenomenon of censoring. Censored data come from the non-access to all information. Thus, instead of observing independent and identically distributed realizations of survival time X, one observes the realization of the variable X subject to various perturbations independent or not of the event studied.

Definition 1.8 *The censoring variable C is defined by the non-observance of the event studied. If instead of observing X, we observe C, and we know that $X > C$ (respectively $X < C$, $C_1 < X < C_2$), we say that there is right-censoring (respectively left-censoring, censoring by interval).*

Here, we will limit ourselves to right-censoring, which is the most common phenomenon encountered during the collection of survival data.
For the individual i, let's consider :
— its survival time X_i ;
— its censoring variable C_i ;
— its actually observed time Z_i.
The survival time is referred to as right-censored if the individual did not experience the event at last observation. In the presence of such censoring, not all survival times are observed : for some of them, we only know that they are above a certain known value.
This censoring can be of one of the following three types :

- **Type I censoring (fixed)**

  The time is not observable beyond a fixed maximum time C. Thus, instead of observing $X_1, \ldots, X_n$, one observes $Z_i = \min(X_i, C)$, $i = 1, \ldots, n$.

- **Type II censoring (waiting)**

  One observes the survival times of n individuals until k among them have seen the event of interest occur. Let $X_{i,n}$ and $Z_{i,n}$ be the order statistics associated respectively with the variables X and Z. The date of censoring is therefore $X_{k,n}$ and the following variables are observed :

$$Z_{i,n} = \begin{cases} X_{i,n} & \text{for } \ i = 1, \ldots, k \\ X_{k,n} & \text{for } \ i = k+1, \ldots, n \, . \end{cases}$$

- **Type III censoring (random)**

  Let $C_1, \ldots, C_n$ be independent and identically distributed random variables. Instead of observing $X_1, \ldots, X_n$, one observes the variables $Z_i = \min(X_i, C_i)$ and $\delta_i = \mathbb{1}_{\{X_i \leq C_i\}}$, $i = 1, \ldots, n$, where δ_i (censoring-index) indicates the presence or absence of censoring.

1.7.3 Estimation of the survival function

The most commonly used survival function estimator when no hypothesis is made on the survival time distribution is Kaplan and Meier [78] estimator.

Definition 1.9 *Let $(Z_i, \delta_i)_{1 \leq i \leq n}$ the actually observed sample and either $(Z_{i,n}, \delta_{i,n})_{1 \leq i \leq n}$ its statistic of ascending order. The Kaplan-Meier estimator is defined by :*

$$\hat{\overline{F}}_n(t) = \prod_{i=1}^{n} \left(\frac{n-i}{n-i+1} \right)^{\delta_{i,n} \mathbb{1}\left\{ Z_{i,n} \leq t \right\}} = \prod_{i=1}^{n} \left(1 - \frac{\delta_{i,n} \mathbb{1}_{\{Z_{i,n} \leq t\}}}{n-i+1} \right). \tag{1.42}$$

Remark that if no data are censored ($\delta_{i,n} = 1$, $\forall i$) then the estimator (1.42) corresponds to the empirical survival function (proportion of individuals having survived at time t) defined by :

$$\overline{F}_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i > t\}}.$$

In a nonparametric regression context with censored data, Beran [13] proposed an adaptation of the estimator (1.42) to the conditional framework in order to estimate the conditional survival function $\overline{F}(\cdot|x) := 1 - F(\cdot|x)$, where $F(\cdot|x)$ denotes the conditional distribution function.

1.7.4 Estimation of tail-index and extreme quantiles

Several authors have been interested in the estimation of tail-index and extreme quantiles under random right-censoring.

1.7.4.1 Unconditional case

In the particular case of the heavy-tailed distributions, Gomes and Oliveira [64] were interested in estimating the tail-index. Delafosse and Guillou [37] proposed tail-index and extreme quantiles estimators for actuarial data where the censoring variable, contrary to classical models, is observed. Beirlant et al. [11] have proposed to adapt the moments estimator (see Paragraph 1.4.3) to the censoring framework and define an associated estimator of extreme quantiles. Other estimators have been proposed by Einmahl et al. [45], Gomes and Neves [63] and Worms and Worms [107], among others.

1.7.4.2 Conditional case

Ndao et al. [89] have proposed a family of tail-index estimators as well as extreme quantiles suitable for censoring. They consider independent copies $(Y_1, \ldots, Y_n)$ of non-negative random variable Y and a deterministic sample $(x_1, \ldots, x_n)$ of a variable $X \in E$, E being a bounded set in $\mathbb{R}^p$ and assume that Y is right-censored by a non-negative random variable C. From the n independent triplets actually observed (Z_i, δ_i, x_i), $i = 1, \ldots, n$ where $Z_i = \min(Y_i, C_i)$ and $\delta_i = \mathbb{1}_{Y_i \leq C_i}$, they propose when the conditional distribution function $F(\cdot \mid x)$ of Y belongs to Fréchet domain of attraction with index $\gamma(x)$, to estimate $\gamma(t)$ at any point $t \in E$. They then use the moving window method and propose, for example, to adapt to the censoring framework, a Hill-type estimator defined in Gardes and Girard [53] by :

$$\hat{\gamma}^H_{m_t}(t) := \frac{1}{k} \sum_{i=1}^{k} i \log \left(\frac{Z^t_{m_t-i+1,m_t}}{Z^t_{m_t-i,m_t}} \right),$$

where m_t designates the number of observations (Z_i, x_i) retained by the selection procedure in a neighbourhood of t, $Z^t_{1,m_t} \leq \ldots \leq Z^t_{m_t,m_t}$ are the order statistics associated with the selected values Z and k an integer such that $1 \leq k \leq m_t$. To do this, they estimate the proportion of uncensored observations among the selected k largest values Z, by :

$$\hat{p}_t = \frac{1}{k} \sum_{i=1}^{k} \delta^t_{m_t-i+1,m_t},$$

where $\delta^t_{1,m_t}, \ldots, \delta^t_{m_t,m_t}$ denote the censoring indices associated respectively with the values $Z^t_{1,m_t}, \ldots, Z^t_{m_t,m_t}$. For any $t \in E$, they thus propose to estimate $\gamma(t)$ by :

$$\hat{\gamma}^{H,C}_{m_t}(t) := \frac{\hat{\gamma}^H_{m_t}(t)}{\hat{p}_t}. \tag{1.43}$$

Other tail-index estimators are proposed in the same paper. As estimator of the conditional extreme quantiles $q(\alpha_{m_t}, t)$, $\alpha_{m_t} \to 0$ as $n \to \infty$, they propose for example :

$$\hat{q}^C(\alpha_{m_t}, t) := Z^t_{m_t-k,m_t} \left(\frac{\hat{\bar{F}}_{m_t}(Z^t_{m_t-k,m_t} \mid t)}{\alpha_{m_t}} \right)^{\hat{\gamma}^{H,C}_{m_t}(t)}, \tag{1.44}$$

wher for all $t \in E$,

$$\hat{\overline{F}}_{m_t}(y \mid t) = \prod_{i=1}^{m_t} \left(\frac{m_t - i}{m_t - i + 1} \right)^{\delta^t_{i,m_t} \mathbb{1}_{\left\{ Z^t_{i,m_t} \leq y \right\}}}$$

is the conditional Kaplan-Meier estimator of $\overline{F}(\cdot \mid t)$, based on the moving window method used. Note that (1.44) is a Weissman-type estimator that extends, in the censoring framework, the estimator of conditional extreme quantiles proposed by Gardes and Girard [54]. Other estimators of conditional extreme quantiles are also proposed in the same article. Another tail-index estimator has also been proposed by Stupfler [103].

Chapter 2

Estimation of the tail-index in a location-scale family of heavy-tailed distributions

This chapter is presented below as an article published in "Dependence Modeling" (https://doi.org/10.1515/demo-2019-0021).

Summary

Estimation of the tail-index in a location-scale family of heavy-tailed distributions

Abstract

We introduce a location-scale model for conditional heavy-tailed distributions when the covariate is deterministic. First, nonparametric estimators of the location and scale functions are introduced. Second, an estimator of the conditional extreme-value index is derived. The asymptotic properties of the estimators are established under mild assumptions and their finite sample properties are illustrated both on simulated and real data.

1 Introduction

The literature on extreme-value analysis of independent and identically distributed observations is very elaborate, see for instance [4, 13, 27]. However, the regression point of view has been less extensively studied. The goal is to describe how tail characteristics such as extreme quantiles or small exceedance probabilities of the quantity of interest Y may depend on some explanatory variable x. Furthermore, as noted in [4, Chapter 7], such covariate information allows to combine datasets from different sources which may lead to better point estimates and thus improved inference.

A parametric approach is considered in [35] where a linear trend is fitted to the expectation of the extreme-value distribution. We also refer to [12] for other examples of parametric models. Turning to semi-parametric models, [29] proposed to mix a non-parametric estimation of the trend with a parametric assumption on Y given x. Similarly, a semi-parametric estimator of γ is introduced in [3] as $\gamma(\psi(\beta' x))$ where ψ is a known link function and β is interpreted as a vector of regression coefficients. Fully non-parametric estimators have been first introduced in [7, 11] through respectively local polynomial and spline models. We also refer to [13, Theorem 3.5.2] for the approximation of the nearest neighbors distribution using the Hellinger distance and to [14] for the study of their asymptotic distribution.

Focusing on the estimation of the tail-index of the conditional distribution of Y given x, moving windows and nearest neighbors approaches are developed respectively by [16, 17] in a fixed design setting. Kernels methods are proposed in [10, 9, 20, 21, 25] to tackle the random design case. Finally, these methods have been adapted to the situation where the covariate is a random field or infinite dimensional, see respectively [1] and [18, 19].

The aim of our work is to estimate in a semi-parametric way the tail-index γ in a location-scale model for conditional heavy-tailed distributions. The so-called conditional tail-index is assumed to be constant while the location and scale parameters depend on the covariate, in a fixed design setting. The underlying idea of this model is to achieve a balance between the flexibility of non-parametric approaches (for the location and scale functions) and the stability of parametric estimators (for the conditional tail-index) compared to purely non-parametric ones. This intuition has also been implemented in [31] : An extreme-value distribution with constant extreme-value index is fitted to standardized rainfall maxima. Here, we introduce a statistical framework to assess the benefits of such approaches in terms of convergence rates of the estimators.

This paper is organized as follows. The location-scale model for heavy-tailed distribution is introduced in Section 2. The associated inference procedures are described in Section 3. Asymptotic results are provided in Section 4 while the finite sample behaviour of the estimators is illustrated in Section 5 on simulated data and in Section 6 on insurance data. Proofs are postponed to the Appendix.

2 Conditional location-scale family of heavy-tailed distributions

Let Y be a real random variable. We assume that the conditional survival function of Y given $x \in [0,1]$ can be written as

$$\bar{F}_Y(y \mid x) := \mathbb{P}(Y > y \mid x) = \bar{F}_Z\left(\frac{y - a(x)}{b(x)}\right), \tag{1}$$

for $y \geq y_0(x) > a(x)$. The functions $a : [0,1] \to \mathbb{R}$ and $b : [0,1] \to \mathbb{R}^+$ are referred to as the location and scale functions respectively while $\bar{F}_Z$ is the survival function of a real random variable Z which is assumed to be heavy-tailed :

$$\bar{F}_Z(z) = z^{-1/\gamma}\ell(z), \;\; z > 0. \tag{2}$$

Here, $\gamma > 0$ is called the conditional tail-index and ℓ is a slowly-varying function at infinity *i.e.* for all $\lambda > 0$,

$$\lim_{z\to\infty} \frac{\ell(\lambda z)}{\ell(z)} = 1.$$

$\bar{F}_Z$ is said to be regularly varying at infinity with index $-1/\gamma$. This property is denoted for short by $\bar{F}_Z \in \mathcal{RV}_{-1/\gamma}$, see [6] for a detailed account on regular variations. Combining (1) and (2) yields

$$\bar{F}_Y(y \mid x) = \left(\frac{y-a(x)}{b(x)}\right)^{-1/\gamma} \ell\left(\frac{y-a(x)}{b(x)}\right), \tag{3}$$

for $y \geq y_0(x) > a(x)$ where the functions $a(\cdot)$, $b(\cdot)$ and the conditional tail-index γ are unknown. We thus obtain a semi-parametric location-scale model for the (heavy) tail of Y given x. The main assumption is that the conditional tail-index γ is independent of the covariate. On the one hand, the proposed semi-parametric modeling offers more flexibility than purely parametric approaches. On the other hand, assuming a constant conditional tail-index γ should yield more reliable estimates in small sample contexts than purely non-parametric approaches. A similar idea is developed in [31] : An extreme-value distribution with constant extreme-value index is fitted to standardized rainfall maxima.

In the following, a fixed design setting is adopted, and thus the covariate x is supposed to be nonrandom. Model (1) can be rewritten as

$$Y = a(x) + b(x)Z, \tag{4}$$

where $x \in [0, 1]$ and Z is a random variable distributed according to (2). Starting with a n-sample $\{(Y_1, x_1), \ldots, (Y_n, x_n)\}$ from (4), it is clear that, since Z is not observed, $a(\cdot)$ and $b(\cdot)$ may only be estimated up to additive and multiplicative factors. This identifiability issue can be fixed by introducing some constraints on $\bar{F}_Z$. To this end, for all $\alpha \in (0, 1)$ consider the αth quantile of Z :

$$q_Z(\alpha) = \inf\{z \in \mathbb{R}; \bar{F}_Z(z) \leq \alpha\},$$

and assume there exist $0 < \mu_3 < \mu_2 < \mu_1 < 1$ such that

$$q_Z(\mu_2) = 0 \text{ and } q_Z(\mu_3) - q_Z(\mu_1) = 1. \tag{5}$$

From (4), it straightforwardly follows that, for all $\alpha \in (0, 1)$, the conditional quantile of Y given $x \in [0, 1]$ is

$$q_Y(\alpha \mid x) = a(x) + b(x)q_Z(\alpha), \tag{6}$$

and therefore the location and scale functions are defined in an unique way by

$$a(x) = q_Y(\mu_2 \mid x) \text{ and } b(x) = q_Y(\mu_3 \mid x) - q_Y(\mu_1 \mid x), \tag{7}$$

for all $x \in [0,1]$. This remark is the starting point of the inference procedure.

3 Inference

Let $\{(Y_1, x_1), \dots, (Y_n, x_n)\}$ be a n-sample from (4) : $Y_i = a(x_i) + b(x_i)Z_i$, $i = 1, \dots, n$ where $Z_1, \dots, Z_n$ are independent and identically distributed (iid) from (2). For the sake of simplicity, it is assumed that the design points are equidistant : $x_i = i/n$ for all $i = 1, \dots, n$ and $x_0 := 0$. This assumption could be weakened to $\max_i |x_i - x_{i-1}| = O(1/n)$ used for instance in [2, 33]. A three-stage inference procedure is adopted.

(i) First, let $\hat{q}_{n,Y}(\alpha \mid x)$ be a nonparametric estimator of the conditional quantile $q_Y(\alpha \mid x)$ where $\alpha \in (0,1)$ and $x \in [0,1]$. In view of (7), the location and scale functions are estimated for all $x \in [0,1]$ by

$$\hat{a}_n(x) = \hat{q}_{n,Y}(\mu_2 \mid x) \text{ and } \hat{b}_n(x) = \hat{q}_{n,Y}(\mu_3 \mid x) - \hat{q}_{n,Y}(\mu_1 \mid x). \tag{8}$$

(ii) Second, the non-observed $Z_1, \dots, Z_n$ can be estimated by the residuals

$$\hat{Z}_i = \frac{Y_i - \hat{a}_n(x_i)}{\hat{b}_n(x_i)}, \tag{9}$$

for all $i = 1, \dots, n$. In practice, nonparametric estimators can suffer from boundary effects [8, 32] and therefore only design points sufficiently far from 0 and 1 are considered. Let us denote by I_n the set of indices associated with such design points and set $m_n = \text{card}(I_n)$.

(iii) Finally, let (k_n) be an intermediate sequence of integers, *i.e.* such that $1 < k_n \leq n$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. The $(k_n + 1)$ top order statistics associated with the pseudo-observations $\hat{Z}_i$, $i \in I_n$ are denoted by $\hat{Z}_{m_n-k_n,m_n} \leq \dots \leq \hat{Z}_{m_n,m_n}$. The conditional tail-index is estimated using an Hill-type statistics :

$$\hat{\gamma}_n = \frac{1}{k_n} \sum_{i=0}^{k_n-1} \log \hat{Z}_{m_n-i,m_n} - \log \hat{Z}_{m_n-k_n,m_n}. \tag{10}$$

This estimator is similar to Hill estimator [30], but in our context, it is built on non iid pseudo-observations.

The proposed procedure relies on the choice of an estimator for the conditional quantiles. Here, a kernel estimator for $\bar{F}_Y(y \mid x)$ is considered (see for instance [33]). For all $(x, y) \in$

$[0,1] \times \mathbb{R}$ let

$$\hat{\bar{F}}_{n,Y}(y \mid x) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_i > y\}} \int_{x_{i-1}}^{x_i} K_h(x-t)dt, \tag{11}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function, $K_h(\cdot) := K(\cdot/h)/h$ with K a density function on $\mathbb{R}$ called a kernel and $h = h_n$ is a nonrandom sequence called the bandwidth such as $h_n \to 0$ as $n \to \infty$. The corresponding estimator of $q_Y(\alpha \mid x)$ is defined for all $(x,\alpha) \in [0,1] \times (0,1)$ by

$$\hat{q}_{n,Y}(\alpha \mid x) = \hat{\bar{F}}^{\leftarrow}_{n,Y}(\alpha \mid x) := \inf\{y;\ \hat{\bar{F}}_{n,Y}(y \mid x) \le \alpha\}. \tag{12}$$

In this context, $I_n = \{\lfloor nh \rfloor, n - \lfloor nh \rfloor\}$ and $m_n = n - 2\lfloor nh \rfloor + 1$. Remark that I_n is properly defined for all large n since $h < 1/2$ eventually. Nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution function have been extensively investigated, see, for example [5, 34, 36], among others.

4 Main results

The following general assumptions are required to establish our results. The first one gathers all the conditions to define a conditional location-scale families of heavy-tailed distributions.

(A.1) $(Y_1, x_1), \dots, (Y_n, x_n)$ are independent observations from the conditional location-scale family of heavy-tailed distributions defined by (1), (2) and (5). The functions $a(\cdot)$ and $b(\cdot)$ are continuous on $[0,1]$ and the survival function $\bar{F}_Z(\cdot)$ is continuously differentiable on $\mathbb{R}$ with associated density $f_Z(\cdot) = -\bar{F}'_Z(\cdot)$.

Under **(A.1)**, the quantile function $q_Z(\cdot)$ exists and we let $H_Z(\cdot) := 1/f_Z(q_Z(\cdot))$ the quantile density function and $U_Z(\cdot) = q_Z(1/\cdot)$ the tail quantile function of Z. The second assumption is a Lipschitz condition on the conditional survival function of Y. Lemma 1 in Appendix provides sufficient conditions on $a(\cdot)$, $b(\cdot)$ and $\bar{F}_Z(\cdot)$ such that it is verified.

(A.2) For any compact set $C \subset \mathbb{R}$, there exists $c_1 > 0$ such that for all $(s,t) \in [0,1]^2$

$$\sup_{y \in C} \left| \frac{\bar{F}_Y(y \mid s)}{\bar{F}_Y(y \mid t)} - 1 \right| \le c_1 |s-t|.$$

The next assumption is standard in the nonparametric kernel estimation framework.

(A.3) K is a bounded density with support $S \subset [-1,1]$ and verifying the Lipschitz property : There exists $c_2 > 0$ such that

$$|K(u) - K(v)| \le c_2 |u-v|$$

for all $(u,v) \in S^2$.

Under **(A.3)**, let $\|K\|_\infty = \sup_{t\in S} K(t)$ and $\|K\|_2 = \left(\int_S K^2(t)dt\right)^{1/2}$. Finally, the so-called second-order condition is introduced (see for instance [27, eq (3.2.5)] :

(A.4) For all $\lambda > 0$, as $z \to \infty$,

$$\frac{U_Z(\lambda z)}{U_Z(z)} - \lambda^\gamma \sim A(z)\lambda^\gamma \frac{\lambda^\rho - 1}{\rho},$$

where $\gamma > 0$, $\rho < 0$ and A is a positive or negative function such that $A(z) \to 0$ as $z \to \infty$.

The rationale behind **(A.4)** is the following. From [6, Theorem 1.5.12], it is clear that (2) is equivalent to $U_Z \in \mathcal{R}V_\gamma$, that is $U_Z(\lambda z)/U_Z(z) \to \lambda^\gamma$ as $z \to \infty$ for all $\lambda > 0$. The role of the second-order condition is thus to control the rate of the previous convergence thanks to the function $A(\cdot)$. Moreover, it can be shown that $|A|$ is regularly varying with index ρ, see [27, Lemma 2.2.3]. It is then clear that ρ, referred to as the second-order parameter, is a crucial quantity, tuning the rate of convergence of most extreme-value estimators, see [27, Chapter 3] for examples.

Our first result states the joint asymptotic normality of the estimators (8) of the location and scale parameters at a point $t_n \in (0, 1)$ not too close from the boundaries of the unit interval.

Theorem 1. *Assume* **(A.1)**, **(A.2)**, **(A.3)** *hold and* $f_Z(q_Z(\mu_j)) > 0$ *for* $j \in \{1, 2, 3\}$. *If* $nh \to \infty$ *and* $nh^3 \to 0$ *as* $n \to \infty$ *then, for all sequence* $(t_n) \subset [h, 1-h]$,

$$\frac{\sqrt{nh}}{b(t_n)} \begin{pmatrix} \hat{a}_n(t_n) - a(t_n) \\ \hat{b}_n(t_n) - b(t_n) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^2},\ \|K\|_2^2\, D\right),$$

where the coefficients of the matrix D *are given by*

$$\begin{aligned} D_{1,1} &= \mu_2(1-\mu_2)H_Z^2(\mu_2), \\ D_{1,2} = D_{2,1} &= \mu_2(1-\mu_1)H_Z(\mu_1)H_Z(\mu_2) - \mu_3(1-\mu_2)H_Z(\mu_2)H_Z(\mu_3), \\ D_{2,2} &= \mu_1(1-\mu_1)H_Z^2(\mu_1) - 2\mu_3(1-\mu_1)H_Z(\mu_1)H_Z(\mu_3) + \mu_3(1-\mu_3)H_Z^2(\mu_3). \end{aligned}$$

A uniform consistency result can also be established :

Theorem 2. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $I_n = \{\lfloor nh \rfloor, \dots, n - \lfloor nh \rfloor\}$ *and suppose* $nh/\log n \to \infty$ *and* $nh^3/\log n \to 0$ *as* $n \to \infty$. *Then,*

$$\sqrt{\frac{nh}{\log n}} \max_{i\in I_n} \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1) \text{ and } \sqrt{\frac{nh}{\log n}} \max_{i\in I_n} \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1).$$

Theorem 2 will reveal useful to prove that the residuals $\hat{Z}_i$ are close to the unobserved Z_i,

$i = 1, \ldots, n$. This justifies the computation of the Hill estimator (10) on the residuals. Our final main result provides the asymptotic normality of this conditional tail-index estimator.

Theorem 3. *Assume* **(A.1)**-**(A.4)** *hold. Let* (k_n) *be an intermediate sequence of integers. Suppose* $nh/(k_n \log n) \to \infty$, $nh^3/\log n \to 0$ *and* $\sqrt{k_n} A(n/k_n) \to 0$ *as* $n \to \infty$. *Then,*

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

It appears that our methodology is able to estimate the tail-index in the conditional location-scale family at the same rate $1/\sqrt{k_n}$ as in iid case, see [26] for a review. As expected, the conditional location-scale family is a more favorable situation than the purely nonparametric framework for the estimation of the conditional tail index where the rate of convergence $1/\sqrt{k_n h}$ is impacted by the covariate, see [10, Corollary 1 & 2], [9, Theorem 3] and [25, Theorem 2]. To be more specific, remark first that conditions $nh/(k_n \log n) \to \infty$ and $nh^3/\log n \to 0$ imply that $k_n = o((n/\log n)^{2/3})$. Second, following [27, Eq. (3.2.10)], if A is exactly a power function, then condition $\sqrt{k_n} A(n/k_n) \to 0$ as $n \to \infty$ yields $k_n = o(n^{-2\rho/(1-2\rho)})$. Up to logarithmic factors, the constraint is then $k_n = o(n^{(-2\rho/(1-2\rho))\wedge(2/3)})$. If $\rho \geq -1$, the rate of convergence of $\hat{\gamma}_n$ is thus $n^{\rho/(1-2\rho)}$ which is the classical rate for estimators of the tail-index, see for instance [28, Remark 3].

Let us also remark that, since $nh/(k_n \log n) \to \infty$ and since $b(\cdot)$ is lower bounded under **(A.1)**, Theorem 1 and Theorem 3 entail that

$$\sqrt{k_n} \begin{pmatrix} \hat{\gamma}_n - \gamma \\ \hat{a}_n(t_n) - a(t_n) \\ \hat{b}_n(t_n) - b(t_n) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^3}, \gamma^2 \, E\right),$$

where the coefficients of the matrix E are given by $E_{1,1} = 1$ and $E_{i,j} = 0$ if $i \in \{2, 3\}$ or $j \in \{2, 3\}$. The joint limiting distribution is degenerated since $\hat{\gamma}_n$ converges at a slower rate than $\hat{a}_n(t_n)$ and $\hat{b}_n(t_n)$.

5 Illustration on simulated data

The finite-sample performance of the estimators of the location and scale functions as well as of the conditional tail-index are illustrated on simulated data from model (4).

The location and scale functions are defined respectively by $a(x) = \cos(2\pi x)$ and $b(x) = 1 + x^2$ for $x \in [0, 1]$. Let Z_0 be a standard Student-t_ν random variable where $\nu \in \{1, 2, 4\}$ denotes the degrees of freedom (df). Let $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$ and introduce $Z = Z_0/(2q_{Z_0}(\mu_1))$ the rescaled Student random variable. By symmetry, $q_Z(\mu_2) = 0$ and $q_Z(\mu_3) = -q_Z(\mu_1)$. Besides, $q_Z(\mu_1) = q_{Z_0}(\mu_1)/(2q_{Z_0}(\mu_1)) = 1/2$ by construction and thus (5) holds. This choice also ensures that Z is heavy-tailed and that the second-order condition

(A.4) holds with conditional tail-index $\gamma = 1/\nu$ and conditional second-order parameter $\rho = -2/\nu$.

In all the experiments, $N = 100$ replications of a dataset of size $n = 1000$ are considered. The kernel function K is chosen to be the quartic (or biweight) kernel

$$K(x) = \frac{15}{16}\left(1 - x^2\right)^2 \mathbb{1}_{\{|x|\leq 1\}},$$

and the bandwidth is fixed to $h = 0.1$.

We denote respectively by $\hat{a}_{n,i}(\cdot)$, $\hat{b}_{n,i}(\cdot)$ and $\hat{\gamma}_{n,i}$ the estimates of $a(\cdot)$, $b(\cdot)$ and γ obtained on the $i-$th replication, $i \in \{1, \ldots, N\}$. The associated mean values are also computed as

$$\bar{\hat{a}}_n(\cdot) := \frac{1}{N}\sum_{i=1}^{N} \hat{a}_{n,i}(\cdot),\ \bar{\hat{b}}_n(\cdot) := \frac{1}{N}\sum_{i=1}^{N} \hat{b}_{n,i}(\cdot) \text{ and } \bar{\hat{\gamma}}_n := \frac{1}{N}\sum_{i=1}^{N} \hat{\gamma}_{n,i}.$$

The results are depicted on Figure 1 ($\nu = 1$), Figure 2 ($\nu = 2$) and Figure 3 ($\nu = 4$). On the top-left panels (a), the true conditional quantiles $q(\mu_j|\cdot)$, $j \in \{1, 2, 3\}$ are superimposed to one replication of the simulated datasets. The estimated location and scale functions $a(\cdot)$ and $b(\cdot)$ are compared with the mean estimates $\bar{\hat{a}}_n(\cdot)$ and $\bar{\hat{b}}_n(\cdot)$ on the top-right (b) and bottom-left panels (c) respectively. Finally, the estimated conditional tail-indices $\hat{\gamma}_{n,i}$, $i = 1, \ldots, N$, the mean estimated value $\bar{\hat{\gamma}}_n$ and the true conditional tail-index are displayed as functions of $k_n \in \{1, \ldots, 300\}$ on the bottom-right panels (d). As expected, it appears on Figure 1(a)–3(a) that the tail heaviness of $Y|x$ decreases as ν increases. The estimation accuracy of the location and scale function does not seem to be sensitive to ν, see Figure 1(b,c)–3(b,c). On the contrary, it appears on Figure 1(d)–3(d) that large values of ν yield a large bias in the estimation of the conditional tail-index. This trend was expected, since the conditional second-order parameter is the main driver of the bias, as explained in Section 4, and since $|\rho| = 1/(2\nu)$ for a Student distribution. Small values of $|\rho|$ in **(A.4)** entail high bias in extreme-value estimators such as Hill's statistics. A way to mitigate this bias could be to replace the conditional tail-index estimator (10) by a bias-reduced Hill-type estimators, see for instance [26].

6 Real data example

We consider here a dataset on motorcycle insurance policies and claims over the period 1994-1998 collected from the former Swedish insurance provider Wasa. The dataset is available from www.math.su.se/GLMbook and the R package `insuranceData`. We focus on two variables : the claim severity Y (defined as the ratio of claim cost by number of claims for each given policyholder) in SEK, and the age x of the policyholder in years. Removing missing data and an affine transformation of a covariate result in $n = 670$ pairs (x_i, Y_i)

with $x_i \in [0,1]$. Some graphical diagnostics have been performed in [23] to check that the heavy-tailed assumption makes sense for Y. Our goal is to estimate the conditional extreme quantile $q_Y(\alpha_n \mid x)$ where $n\alpha_n \to 0$ and $x \in (0,1)$. Two estimators are considered. The first one relies on the semi-parametric model via (6) :

$$\tilde{q}_{n,Y}(\alpha_n \mid x) = \hat{a}_n(x) + \hat{b}_n(x)\hat{q}_{n,Z}(\alpha_n),$$

and on Weissman estimator [37] applied to the pseudo-observations $\hat{Z}_i$, $i \in I_n$:

$$\hat{q}_{n,Z}(\alpha_n) = \hat{Z}_{m_n-k_n,m_n} \left(\frac{\alpha_n m_n}{k_n}\right)^{-\hat{\gamma}_n}.$$

The second one is the nonparametric conditional Weissman estimator introduced in [10] :

$$\check{q}_{n,Y}(\alpha_n \mid x) = \hat{q}_{n,Y}(k_n/m_n \mid x) \left(\frac{\alpha_n m_n}{k_n}\right)^{-\check{\gamma}_n(x)},$$

where $\hat{q}_{n,Y}(k_n/m_n \mid x)$ is defined in (12) and $\check{\gamma}_n(x)$ is an estimator of the conditional tail index. Here, we selected a recent estimator introduced in [23] and denoted by $\hat{\gamma}^{(3)}_{k_n}(x)$ in the previously mentioned paper.

As in Section 5, we set the normalizing parameters to $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$. The quartic kernel is used and the bandwidth $h = 0.065$ is chosen by the cross-validation procedure implemented in R as `h.cv`. The estimated location and scaled functions are superimposed to the dataset on Figure 4. The residuals are then computed according to (9).

To confirm that the location-scale model (3) is appropriate, Figure 5 displays a quantile-quantile plot of the weighted log-spacings within the top of the residuals against the quantiles of the standard exponential distribution. Formally, let $W_{i,m_n} = i\log(\hat{Z}_{m_n-i+1,m_n}/\hat{Z}_{m_n-i,n})$, $1 \le i \le k_n - 1$, denote the weighted log-spacings computed from the consecutive top order statistics of the residuals. It is known that, if $\hat{Z}$ is heavy-tailed with tail-index γ then, the W_{i,m_n} are approximately independent copies of an exponential random variable with mean γ, see for instance [4]. Here, the number of upper statistics is fixed to $k_n = 130$ by a visual inspection of the Hill plot (not reproduced here). The relationship appearing on Figure 5 is approximately linear, which constitutes a graphical evidence that the heavy-tail assumption (2) on Z makes sense and that the choice of k_n is appropriate.

Finally, the two conditional quantile estimators $\tilde{q}_{n,Y}(\alpha_n \mid \cdot)$ and $\check{q}_{n,Y}(\alpha_n \mid \cdot)$ are graphically compared on Figure 6 for $\alpha_n = 8/n$. Both of them yield level curves with similar shapes and located above the sample. Unsurprisingly, the estimator $\tilde{q}_{n,Y}(\alpha_n \mid \cdot)$ based on the location-scale model has a smoother behavior than $\check{q}_{n,Y}(\alpha_n \mid \cdot)$ since it relies on the assumption that the tail-index does not depend on the covariate.

7 Appendix : Proofs

Technical lemmas are collected in Paragraph 7.1 while preliminary results of general interest are provided in Paragraph 7.2. Finally, the proofs of the main results are given in Paragraph 7.3.

7.1 Auxiliary lemmas

We begin by providing some sufficient conditions such that **(A.2)** holds.

Lemma 1. *If* **(A.1)** *holds and there exist* $(c_a, c_b, c_F) \in \mathbb{R}_+^3$ *and* $m_b > 0$ *such that for all* $(y, z, t, s) \in \mathbb{R}^2 \times [0,1]^2$,

$$\begin{aligned} m_b &\leq |b(t)|, \\ |a(t) - a(s)| &\leq c_a|t - s|, \\ |b(t) - b(s)| &\leq c_b|t - s|, \\ |\log \overline{F}_Z(y) - \log \overline{F}_Z(z)| &\leq c_F|y - z|, \end{aligned}$$

then **(A.2)** *holds.*

Proof. Let us first remark that, since $|a(\cdot)|$ and $|b(\cdot)|$ are continuous functions on the compact set $[0, 1]$, they are necessarily upper bounded by some finite constants denoted by M_a and M_b. Second, consider the quantity

$$\Delta(y, t, s) := \log \overline{F}_Y(y \mid t) - \log \overline{F}_Y(y \mid s) = \log \overline{F}_Z\left(\frac{y - a(t)}{b(t)}\right) - \log \overline{F}_Z\left(\frac{y - a(s)}{b(s)}\right).$$

The Lipschitz assumption on $\log \overline{F}_Z$ yields for all $(t, s) \in [0, 1]^2$ and $y \in \mathbb{R}$:

$$\begin{aligned} |\Delta(y, t, s)| &\leq c_F \left| \frac{y - a(t)}{b(t)} - \frac{y - a(s)}{b(s)} \right| \\ &= c_F \left| \frac{y(b(s) - b(t)) + a(s)(b(t) - b(s)) + b(s)(a(s) - a(t))}{b(t)b(s)} \right| \\ &\leq \frac{c_F}{m_b^2}(|y|c_b + M_a c_b + M_b c_a)|t - s|, \end{aligned}$$

in view of the assumptions on $a(\cdot)$ and $b(\cdot)$. Let $C \subset \mathbb{R}$ be a compact set. It follows that the supremum of $|\Delta(y, t, s)|$ on $(y, t, s) \in C \times [0, 1]^2$ is bounded and thus there exists $\tilde{c} > 0$ such that

$$\sup_{y \in C} |\exp(\Delta(y, t, s)) - 1| \leq \tilde{c} \sup_{y \in C} |\Delta(y, t, s)|.$$

Letting $M_y := \sup\{|y| \in C\}$, assumption **(A.2)** holds with $c_1 = \tilde{c}\, c_F((M_y + M_a)c_b + M_b c_a)/m_b^2$. □

The next result is an adaptation of Bochner's lemma to our fixed design setting.

Lemma 2. *Let $\psi(\cdot \mid \cdot) : \mathbb{R}^p \times [0,1] \to \mathbb{R}^+$, $p \geq 1$, be a positive function and C a compact subset of $\mathbb{R}^p$. For all sequences $(t_n) \subset [h, 1-h]$ and $(y_n) \subset C$, define*

$$\psi_n(y_n \mid t_n) := \sum_{i=1}^{n} \psi(y_n \mid x_i) \int_{x_{i-1}}^{x_i} Q_h(t_n - s) ds,$$

where $x_i = i/n$ for all $i = 0, \ldots, n$ and $Q_h(\cdot) = Q(\cdot/h)/h$, with Q is a measurable positive function with support $S \subset [-1,1]$. If there exists $c > 0$ such that $\forall\, (x,s) \in [0,1]^2$,

$$\sup_{y \in C} \left| \frac{\psi(y \mid x)}{\psi(y \mid s)} - 1 \right| \leq c|x - s|,$$

then, as $n \to \infty$,

$$\left| \frac{\psi_n(y_n \mid t_n)}{\psi(y_n \mid t_n)} - \int_S Q(u) du \right| = O\left(\frac{1}{n}\right) + O(h).$$

Proof. Consider the expansion

$$\begin{aligned}
\frac{\psi_n(y_n \mid t_n)}{\psi(y_n \mid t_n)} - \int_S Q(u)du &= \frac{\sum_{i=1}^n \psi(y_n \mid x_i) \int_{x_{i-1}}^{x_i} Q_h(t_n - s)ds}{\psi(y_n \mid t_n)} - \int_S Q(u)du \\
&= \frac{\int_0^1 \psi(y_n \mid s) Q_h(t_n - s)ds - \psi(y_n \mid t_n) \int_S Q(u)du}{\psi(y_n \mid t_n)} \\
&+ \frac{\sum_{i=1}^n \psi(y_n \mid x_i) \int_{x_{i-1}}^{x_i} Q_h(t_n - s)ds - \int_0^1 \psi(y_n \mid s) Q_h(t_n - s)ds}{\psi(y_n \mid t_n)} \\
&=: \frac{T_{n,1}}{\psi(y_n \mid t_n)} + \frac{T_{n,2}}{\psi(y_n \mid t_n)},
\end{aligned}$$

and let us first focus on $T_{n,1}$. The change of variable $u = (t_n - s)/h$ yields

$$T_{n,1} = \int_{(t_n-1)/h}^{t_n/h} \psi(y_n \mid t_n - uh) Q(u)du - \psi(y_n \mid t_n) \int_S Q(u)du.$$

Since $(t_n) \subset [h, 1-h]$, it follows that $S \subset [-1,1] \subset \left[\frac{t_n - 1}{h}, \frac{t_n}{h}\right]$ and therefore

$$T_{n,1} = \int_S [\psi(y_n \mid t_n - uh) - \psi(y_n \mid t_n)] Q(u)du.$$

As a consequence, for all $y_n \in C$,

$$\left| \frac{T_{n,1}}{\psi(y_n \mid t_n)} \right| \leq \int_S \left| \frac{\psi(y_n \mid t_n - uh)}{\psi(y_n \mid t_n)} - 1 \right| Q(u)du \leq ch \int_S |u| Q(u)du = O(h). \tag{13}$$

Let us now turn to the second term

$$T_{n,2} = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} [\psi(y_n \mid x_i) - \psi(y_n \mid s)] Q_h(t_n - s) ds.$$

We have, for all $y_n \in C$,

$$\begin{aligned}
\left| \frac{T_{n,2}}{\psi(y_n \mid t_n)} \right| &\leq \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \frac{\psi(y_n \mid s)}{\psi(y_n \mid t_n)} \left| \frac{\psi(y_n \mid x_i)}{\psi(y_n \mid s)} - 1 \right| Q_h(t_n - s) ds \\
&\leq \frac{c}{\psi(y_n \mid t_n)} \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \psi(y_n \mid s) |x_i - s| Q_h(t_n - s) ds \\
&\leq \frac{c}{n\psi(y_n \mid t_n)} \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \psi(y_n \mid s) Q_h(t_n - s) ds \\
&= \frac{c}{n\psi(y_n \mid t_n)} \int_{0}^{1} \psi(y_n \mid s) Q_h(t_n - s) ds \\
&= \frac{c}{n} \left(\frac{T_{n,1}}{\psi(y_n \mid t_n)} + \int_S Q(u) du \right) \\
&= O\left(\frac{1}{n}\right), \qquad (14)
\end{aligned}$$

in view of (13). Finally, collecting (13) and (14), the conclusion follows. □

As a consequence of Lemma 2, the asymptotic bias and variance of the estimator (11) of the conditional survival function can be derived.

Lemma 3. *Suppose* **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset [h, 1-h]$ *and* $(y_n) \subset C$*, where* C *is a compact subset of* $\mathbb{R}$*, be two nonrandom sequences.*

(i) Then,

$$\left| \frac{\mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right)}{\bar{F}_Y(y_n \mid t_n)} - 1 \right| = O\left(\frac{1}{n}\right) + O(h).$$

(ii) If, moreover, $nh \to \infty$ *as* $n \to \infty$ *and* $\liminf F_Y(y_n \mid t_n) > 0$*, then*

$$var\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right) \sim \frac{\|K\|_2^2}{nh} F_Y(y_n \mid t_n) \bar{F}_Y(y_n \mid t_n).$$

Proof. (i) Remarking that

$$\mathbb{E}\left[\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\{Y_i > y_n\}} \int_{x_{i-1}}^{x_i} K_h(t_n - s) ds\right] = \sum_{i=1}^{n} \bar{F}_Y(y_n \mid x_i) \int_{x_{i-1}}^{x_i} K_h(t_n - s) ds,$$

the conclusion follows from Lemma 2.

(ii) Let us consider the expansion :

$$\begin{aligned} var\left(\hat{\overline{F}}_{n,Y}(y_n \mid t_n)\right) &= \sum_{i=1}^{n} var\left(\mathbb{1}_{\{Y_i>y_n\}} \int_{x_{i-1}}^{x_i} K_h(t_n-s)ds\right) \\ &= \sum_{i=1}^{n} \overline{F}_Y(y_n \mid x_i) S_{n,i} - \sum_{i=1}^{n} \overline{F}_Y^2(y_n \mid x_i) S_{n,i} \\ &=: T_{n,1} - T_{n,2}, \end{aligned}$$

where

$$S_{n,i} := \left(\int_{x_{i-1}}^{x_i} K_h(t_n-s)ds\right)^2 = \frac{1}{h^2}\int_{x_{i-1}}^{x_i}\int_{x_{i-1}}^{x_i} K\left(\frac{t_n-s_1}{h}\right) K\left(\frac{t_n-s_2}{h}\right) ds_1 ds_2. \tag{15}$$

Let us write

$$K\left(\frac{t_n-s_2}{h}\right) = K\left(\frac{t_n-s_1}{h}\right) + K\left(\frac{t_n-s_2}{h}\right) - K\left(\frac{t_n-s_1}{h}\right),$$

with, under **(A.3)**,

$$\left|K\left(\frac{t_n-s_2}{h}\right) - K\left(\frac{t_n-s_1}{h}\right)\right| \leq \frac{c_2|s_2-s_1|}{h} = O\left(\frac{1}{nh}\right),$$

uniformly on $(s_1, s_2) \in [x_{i-1}, x_i]^2$ and $i = 1, \dots, n$. It thus follows that

$$\begin{aligned} S_{n,i} &= \frac{1}{h^2}\int_{x_{i-1}}^{x_i}\int_{x_{i-1}}^{x_i} \left[K^2\left(\frac{t_n-s_1}{h}\right) + K\left(\frac{t_n-s_1}{h}\right) O\left(\frac{1}{nh}\right)\right] ds_1 ds_2 \\ &= \frac{1}{nh^2}\int_{x_{i-1}}^{x_i} K^2\left(\frac{t_n-s}{h}\right) ds + O\left(\frac{1}{n^2h^3}\right)\int_{x_{i-1}}^{x_i} K\left(\frac{t_n-s}{h}\right) ds. \end{aligned}$$

Defining $M(v) = K^2(v)/\|K\|_2^2$ yields

$$S_{n,i} = \frac{\|K\|_2^2}{nh}\int_{x_{i-1}}^{x_i} M_h(t_n-s)ds + O\left(\frac{1}{n^2h^2}\right)\int_{x_{i-1}}^{x_i} K_h(t_n-s)ds. \tag{16}$$

Replacing in $T_{n,1}$, we obtain :

$$\begin{aligned} T_{n,1} &= \frac{\|K\|_2^2}{nh}\left\{\sum_{i=1}^{n}\overline{F}_Y(y_n \mid x_i)\int_{x_{i-1}}^{x_i} M_h(t_n-s)ds\right. \\ &+ \left. O\left(\frac{1}{nh}\right)\sum_{i=1}^{n}\overline{F}_Y(y_n \mid x_i)\int_{x_{i-1}}^{x_i} K_h(t_n-s)ds\right\}. \end{aligned}$$

Applying Lemma 2 twice and recalling that $nh \to \infty$ as $n \to \infty$ entail

$$T_{n,1} = \frac{\|K\|_2^2}{nh}\overline{F}_Y(y_n \mid t_n)\,\left(1 + O(h) + O\left(\frac{1}{nh}\right)\right).$$

Similarly,

$$T_{n,2} = \frac{\|K\|_2^2}{nh} \overline{F}_Y^2(y_n \mid t_n) \left(1 + O(h) + O\left(\frac{1}{nh}\right)\right),$$

and the conclusion follows :

$$\begin{aligned} T_{n,1} - T_{n,2} &= \frac{\|K\|_2^2}{nh} \overline{F}_Y(y_n \mid t_n) F_Y(y_n \mid t_n) \left(1 + \frac{1}{F_Y(y_n \mid t_n)} \left(O(h) + O\left(\frac{1}{nh}\right)\right)\right) \\ &= \frac{\|K\|_2^2}{nh} \overline{F}_Y(y_n \mid t_n) F_Y(y_n \mid t_n) \,(1 + o(1)), \end{aligned}$$

under the assumption $\liminf F_Y(y_n \mid t_n) > 0$. □

The next lemma controls the error between each unobserved random variable Z_i and its estimation $\hat{Z}_i$, for all $i = 1, \dots, n$.

Lemma 4. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $I_n = \{\lfloor nh \rfloor, \dots, n - \lfloor nh \rfloor\}$ *and suppose* $nh/\log n \to \infty$ *and* $nh^3/\log n \to 0$ *as* $n \to \infty$. *Then, for all* $i \in I_n$,

$$|\hat{Z}_i - Z_i| \le R_{n,i}(1 + |Z_i|), \text{ where } \max_{i \in I_n} R_{n,i} = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh}}\right) = o_{\mathbb{P}}(1).$$

Proof. Remark that for all $i \in I_n$, one has

$$\begin{aligned} |\hat{Z}_i - Z_i| &= \left| \frac{Y_i - \hat{a}_n(x_i)}{\hat{b}_n(x_i)} - Z_i \right| = \left| \frac{a(x_i) - \hat{a}_n(x_i)}{\hat{b}_n(x_i)} + \frac{\hat{b}_n(x_i) - b(x_i)}{\hat{b}_n(x_i)} Z_i \right| \\ &\le \left| \frac{b(x_i)}{\hat{b}_n(x_i)} \right| \left(\left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| + \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| |Z_i| \right) \\ &\le \left| \frac{b(x_i)}{\hat{b}_n(x_i)} \right| \max \left\{ \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| ; \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| \right\} (1 + |Z_i|) \\ &=: \left| \frac{b(x_i)}{\hat{b}_n(x_i)} \right| \max \left\{ \left| \xi_{i,n}^{(a)} \right| ; \left| \xi_{i,n}^{(b)} \right| \right\} (1 + |Z_i|). \end{aligned}$$

Let us define, for all $i \in I_n$,

$$\xi_{i,n}^{(a)} = \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)}, \; \xi_{i,n}^{(b)} = \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \text{ and } R_{n,i} = \left| \frac{b(x_i)}{\hat{b}_n(x_i)} \right| \max \left\{ \left| \xi_{i,n}^{(a)} \right| ; \left| \xi_{i,n}^{(b)} \right| \right\}.$$

On the one hand, Theorem 2 entails

$$\begin{aligned} \max_{i \in I_n} R_{n,i} &\le \max_{i \in I_n} \left| \frac{b(x_i)}{\hat{b}_n(x_i)} \right| \max \left\{ \max_{i \in I_n} \left| \xi_{i,n}^{(a)} \right| ; \max_{i \in I_n} \left| \xi_{i,n}^{(b)} \right| \right\} \\ &= \max_{i \in I_n} \left| \frac{b(x_i)}{\hat{b}_n(x_i)} \right| O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh}}\right). \end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{P}\left(\max_{i\in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\geq 2\right) &= \mathbb{P}\left(\max_{i\in I_n}\left|\frac{1}{1+\xi_{i,n}^{(b)}}\right|\geq 2\right)\leq \mathbb{P}\left(\max_{i\in I_n}\left|\xi_{i,n}^{(b)}\right|\geq\frac{1}{2}\right)\\ &\leq \mathbb{P}\left(\sqrt{\frac{nh}{\log n}}\max_{i\in I_n}\left|\xi_{i,n}^{(b)}\right|\geq\frac{1}{2}\sqrt{\frac{nh}{\log n}}\right).\end{aligned}$$

Again, Theorem 2 shows that the following uniform consistency holds : For all $\epsilon>0$, there exists $M(\epsilon)>0$ such that

$$\mathbb{P}\left(\sqrt{\frac{nh}{\log n}}\max_{i\in I_n}\left|\xi_{i,n}^{(b)}\right|\geq M(\epsilon)\right)\leq\epsilon.$$

Now, for n large enough $(nh/\log n)^{1/2}>2M(\epsilon)$ so that

$$\mathbb{P}\left(\max_{i\in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\geq 2\right)\leq\mathbb{P}\left(\max_{i\in I_n}\sqrt{\frac{nh}{\log n}}\left|\xi_{i,n}^{(b)}\right|\geq M(\epsilon)\right)\leq\epsilon,$$

i.e. $\max_{i\in I_n}|b(x_i)/\hat{b}_n(x_i)|=O_{\mathbb{P}}(1)$. As a result,

$$\max_{i\in I_n}R_{n,i}=O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh}}\right),$$

which completes the proof of the lemma. □

Finally, Lemma 5 is an adaptation of [22, Proposition 1]. It permits to derive the error made on the estimation of the order statistics Z_{m_n-i,m_n}, $i=0,\dots,m_n-1$ from the error made on the unsorted Z_i, $i\in I_n$.

Lemma 5. *Let $I_n=\{\lfloor nh\rfloor,\dots,n-\lfloor nh\rfloor\}$ and $m_n=\operatorname{card}(I_n)$. Consider (k_n) an intermediate sequence of integers. If, for all $i\in I_n$, $|\hat{Z}_i-Z_i|\leq R_{n,i}\left(1+|Z_i|\right)$, with $\max_{i\in I_n}R_{n,i}\xrightarrow{\mathbb{P}}0$, then*

$$\max_{0\leq i\leq k_n}\left|\log\frac{\hat{Z}_{m_n-i,m_n}}{Z_{m_n-i,m_n}}\right|=O_{\mathbb{P}}\left(\max_{i\in I_n}R_{n,i}\right).$$

Proof. Remarking that $m_n=n-2\lfloor nh\rfloor+1\sim n$ as $n\to\infty$ and (2) entails that the distribution of Z has an infinite upper endpoint, the conclusion follows by applying [22, Proposition 1]. □

7.2 Preliminary results

Let $\vee$ (resp. $\wedge$) denote the maximum (resp. the minimum). The next proposition provides a joint asymptotic normality result for the estimator (11) of the conditional survival function evaluated at points depending on n.

Proposition 1. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset [h, 1-h]$ *and* $(\alpha_j)_{j=1,\dots,J}$ *a strictly decreasing sequence in* $(0,1)$. *For all* $j \in \{1,\dots,J\}$, *define* $y_{j,n} = q_Y(\alpha_j \mid t_n) + b(t_n)\epsilon_{j,n}$, *where* $\epsilon_{j,n} \to 0$ *as* $n \to \infty$. *If in addition* $nh \to \infty$ *and* $nh^3 \to 0$ *as* $n \to \infty$, *then*

$$\left\{ \sqrt{nh} \left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n) \right] \right\}_{j=1,\dots,J} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^J}, \|K\|_2^2 \, B\right),$$

where $B_{k,l} = \alpha_{k \vee l}(1 - \alpha_{k \wedge l})$ *for all* $(k,l) \in \{1,\dots,J\}^2$.

Proof. Let us first remark that, for all $j \in \{1,\dots,J\}$, in view of (6), the sequence $y_{j,n} = a(t_n) + b(t_n)(q_Z(\alpha_j) + \epsilon_{j,n})$ is bounded since $\epsilon_{j,n} \to 0$ as $n \to \infty$ and since $a(\cdot)$ and $b(\cdot)$ are continuous functions defined on compact sets. Besides, from (1), $F_Y(y_{j,n} \mid t_n) = F_Z(q_Z(\alpha_j) + \epsilon_{j,n}) \to 1 - \alpha_j > 0$ as $n \to \infty$ and thus the assumptions of Lemma 3(i,ii) are satisfied. Let $\beta \neq 0$ in $\mathbb{R}^J$, $J \geq 1$ and consider the random variable

$$\begin{aligned}
\Gamma_n &= \sum_{j=1}^{J} \beta_j \left\{ \hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n) \right\} \\
&= \sum_{j=1}^{J} \beta_j \left\{ \hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right) \right\} \\
&+ \sum_{j=1}^{J} \beta_j \left\{ \mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right) - \bar{F}_Y(y_{j,n} \mid t_n) \right\} \\
&=: \Gamma_{n,1} + \Gamma_{n,2}.
\end{aligned}$$

Let us first consider the random term :

$$\Gamma_{n,1} = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} K_h(t_n - s) ds \sum_{j=1}^{J} \beta_j \left\{ \mathbb{1}_{\{Y_i > y_{j,n}\}} - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{j,n}\}} \right) \right\} =: \sum_{i=1}^{n} T_{i,n}.$$

By definition, $\mathbb{E}(\Gamma_{n,1}) = 0$, and by independence of $Y_1, \dots, Y_n$,

$$\begin{aligned}
var(\Gamma_{n,1}) &= \sum_{i=1}^{n} var(T_{i,n}) \\
&= \sum_{i=1}^{n} \left(\int_{x_{i-1}}^{x_i} K_h(t_n - s) ds \right)^2 var\left(\sum_{j=1}^{J} \beta_j \mathbb{1}_{\{Y_i > y_{j,n}\}} \right) \\
&=: \sum_{i=1}^{n} S_{n,i} \, \beta^t \Sigma^{(i,n)} \beta, \qquad (17)
\end{aligned}$$

where $S_{n,i}$ is defined by (15) in the proof of Lemma 3, and where $\Sigma^{(i,n)}$ is the matrix whose coefficients are defined for $(k,l) \in \{1,\dots,J\}^2$ by $\Sigma^{(i,n)}_{k,l} = cov\left(\mathbb{1}_{\{Y_i > y_{k,n}\}}, \mathbb{1}_{\{Y_i > y_{l,n}\}} \right)$. In view of (16),

$$S_{n,i} = \frac{\|K\|_2^2}{nh} \int_{x_{i-1}}^{x_i} M_h(t_n - s)ds + O\left(\frac{1}{n^2h^2}\right) \int_{x_{i-1}}^{x_i} K_h(t_n - s)ds$$

and

$$\begin{aligned}
\Sigma_{k,l}^{(i,n)} &= \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{k,n} \vee y_{l,n}\}}\right) - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{k,n}\}}\right) \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{l,n}\}}\right) \\
&= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) - \bar{F}_Y(y_{k,n} \mid x_i)\bar{F}_Y(y_{l,n} \mid x_i) \\
&= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) - \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i)\bar{F}_Y(y_{k,n} \wedge y_{l,n} \mid x_i) \\
&= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) F_Y(y_{k,n} \wedge y_{l,n} \mid x_i) \\
&=: \varphi(y_{k,n}, y_{l,n} \mid x_i), \qquad (18)
\end{aligned}$$

where φ is the function $\mathbb{R}^2 \times [0,1] \to [0,1]$ defined by $\varphi(\cdot,\cdot \mid \cdot) = \bar{F}_Y(\cdot \vee \cdot \mid \cdot)F_Y(\cdot \wedge \cdot \mid \cdot)$. Replacing in (17) yields $var(\Gamma_{n,1}) = \beta^t C^{(n)} \beta$, where $C^{(n)}$ is the covariance matrix whose coefficients are defined by

$$\begin{aligned}
C_{k,l}^{(n)} &= \frac{\|K\|_2^2}{nh} \sum_{i=1}^{n} \varphi(y_{k,n}, y_{l,n} \mid x_i) \int_{x_{i-1}}^{x_i} M_h(t_n - s)ds \\
&+ O\left(\frac{1}{n^2h^2}\right) \sum_{i=1}^{n} \varphi(y_{k,n}, y_{l,n} \mid x_i) \int_{x_{i-1}}^{x_i} K_h(t_n - s)ds.
\end{aligned}$$

Applying Lemma 2 twice and recalling that $nh \to \infty$ entail

$$\begin{aligned}
C_{k,l}^{(n)} &= \frac{\|K\|_2^2}{nh} \varphi(y_{k,n}, y_{l,n} \mid t_n)(1 + O(h)) + O\left(\frac{1}{n^2h^2}\right) \varphi(y_{k,n}, y_{l,n} \mid t_n)(1 + O(h)) \\
&= \frac{\|K\|_2^2}{nh} \varphi(y_{k,n}, y_{l,n} \mid t_n)(1 + o(1)).
\end{aligned}$$

As a result,

$$var(\Gamma_{n,1}) \sim \frac{\|K\|_2^2}{nh} \beta^t B^{(n)} \beta,$$

where

$$B_{k,l}^{(n)} = \varphi(y_{k,n}, y_{l,n} \mid t_n) = \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid t_n) F_Y(y_{k,n} \wedge y_{l,n} \mid t_n).$$

Let us remark that, in view of (6),

$$\begin{aligned}
y_{k,n} - y_{l,n} &= q_Y(\alpha_k \mid t_n) - q_Y(\alpha_l \mid t_n) + b(t_n)(\epsilon_{k,n} - \epsilon_{l,n}) \\
&= b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_l) + \epsilon_{k,n} - \epsilon_{l,n}) \\
&\sim b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_l)),
\end{aligned}$$

as $n \to \infty$. Thus, assuming for instance $k < l$ implies $\alpha_k > \alpha_l$ and thus $q_Z(\alpha_k) < q_Z(\alpha_l)$ leading to $y_{k,n} < y_{l,n}$ for n large enough. More generally, $y_{k,n} \vee y_{l,n} = y_{k \vee l,n}$ and $y_{k,n} \wedge y_{l,n} =$

$y_{k\wedge l,n}$ for n large enough and thus

$$B_{k,l}^{(n)} = \bar{F}_Y(y_{k\vee l,n} \mid t_n) F_Y(y_{k\wedge l,n} \mid t_n).$$

From (1) and (6), we have

$$\bar{F}_Y(y_{k,n} \mid t_n) = \bar{F}_Z\left(\frac{y_{k,n} - a(t_n)}{b(t_n)}\right) = \bar{F}_Z\left(q_Z(\alpha_k) + \epsilon_{k,n}\right) = \alpha_k + o(1),$$

in view of the continuity of $\bar{F}_Z$. As a result, $B_{k,l}^{(n)} \to B_{k,l} = \alpha_{k\vee l}(1 - \alpha_{k\wedge l})$ as $n \to \infty$ and therefore

$$var(\Gamma_{n,1}) \sim \frac{\|K\|_2^2}{nh} \beta^t B \beta.$$

The proof of the asymptotic normality of $\Gamma_{n,1}$ is based on Lyapounov criteria for triangular arrays of independent random variables :

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \Big/ var(\Gamma_{n,1}) \to 0 \tag{19}$$

as $n \to \infty$. Let us first remark that, for all $i = 1, \ldots, n$, the random variable $T_{i,n}$ is bounded :

$$\begin{aligned} |T_{i,n}| &\leq \int_{x_{i-1}}^{x_i} K_h(t_n - s) ds \sum_{j=1}^{J} \beta_j \left| \mathbb{1}_{\{Y_i > y_{j,n}\}} - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{j,n}\}}\right)\right| \\ &\leq \int_{x_{i-1}}^{x_i} K_h(t_n - s) ds \sum_{j=1}^{J} |\beta_j| \\ &\leq \frac{\|K\|_\infty}{nh} \sum_{j=1}^{J} |\beta_j| =: \zeta_n \end{aligned}$$

in view of **(A.3)**. As a consequence,

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \leq \zeta_n \sum_{i=1}^{n} \mathbb{E}(T_{i,n}^2) = \zeta_n \sum_{i=1}^{n} var(T_{i,n}) = \zeta_n var(\Gamma_{n,1})$$

and it is thus clear that (19) holds under the assumption $nh \to \infty$. A a result,

$$\sqrt{nh}\,\Gamma_{n,1} \xrightarrow{d} \mathcal{N}\left(0, \|K\|_2^2 \, \beta^t B \beta\right). \tag{20}$$

Let us now turn to the nonrandom term

$$\Gamma_{n,2} = \sum_{j=1}^{J} \beta_j \bar{F}_Y(y_{j,n} \mid t_n) \left[\frac{\mathbb{E}\left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n)\right]}{\bar{F}_Y(y_{j,n} \mid t_n)} - 1 \right].$$

Lemma 3(i) together with the assumptions $nh^3 \to 0$ and $nh \to \infty$ as $n \to \infty$ entail

$$\sqrt{nh}|\Gamma_{n,2}| \leq \sqrt{nh} \sum_{j=1}^{J} |\beta_j| \left| \frac{\mathbb{E}\left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n)\right]}{\bar{F}_Y(y_{j,n} \mid t_n)} - 1 \right| = O(\sqrt{nh^3}) = o(1). \tag{21}$$

Finally, collecting (20) and (21), $\sqrt{nh}\Gamma_n$ converges to a centered Gaussian random variable with variance $\|K\|_2^2 \, \beta^t B \beta$, and the result follows. □

The following proposition provides the joint asymptotic normality of the estimator (12) of conditional quantiles. It can be read as an adaptation of classical results [5, 34, 36] to the location-scale setting.

Proposition 2. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset [h, 1-h]$ *and* $(\alpha_j)_{j=1,\dots,J}$ *a strictly decreasing sequence in* $(0,1)$ *such that* $f_Z(q_Z(\alpha_j)) > 0$ *for all* $j \in \{1, \dots, J\}$. *If* $nh \to \infty$ *and* $nh^3 \to 0$ *as* $n \to \infty$, *then*

$$\left\{ \frac{\sqrt{nh}}{b(t_n)} \left[\hat{q}_{n,Y}(\alpha_j \mid t_n) - q_Y(\alpha_j \mid t_n) \right] \right\}_{j=1,\dots,J} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^J}, \|K\|_2^2 \, C\right),$$

where C *is the covariance matrix defined by* $C_{k,l} = \alpha_{k \vee l}(1 - \alpha_{k \wedge l}) H_Z(\alpha_k) H_Z(\alpha_l)$ *for all* $(k,l) \in \{1, \dots, J\}^2$.

Proof. Let $(s_1, \dots, s_J) \in \mathbb{R}^J$, $\nu_{j,n} := s_j b(t_n)/\sqrt{nh}$ for all $j = 1, \dots, J$ and consider :

$$\begin{aligned} W_n(s_1, \dots, s_J) &= \mathbb{P}\left(\bigcap_{j=1}^{J} \left\{ \frac{\sqrt{nh}}{b(t_n)} \left(\hat{q}_{n,Y}(\alpha_j \mid t_n) - q_Y(\alpha_j \mid t_n) \right) \leq s_j \right\} \right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{J} \left\{ \hat{q}_{n,Y}(\alpha_j \mid t_n) \leq q_Y(\alpha_j \mid t_n) + \nu_{j,n} \right\} \right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{J} \left\{ V_{j,n} \leq v_{j,n} \right\} \right), \end{aligned}$$

where, for $j = 1, \dots, J$,

$$\begin{aligned} V_{j,n} &:= \sqrt{nh}\left[\hat{\bar{F}}_{n,Y}\left(q_Y(\alpha_j \mid t_n) + \nu_{j,n} \mid t_n \right) - \bar{F}_Y\left(q_Y(\alpha_j \mid t_n) + \nu_{j,n} \mid t_n \right) \right], \\ v_{j,n} &:= \sqrt{nh}\left[\alpha_j - \bar{F}_Y(q_Y(\alpha_j \mid t_n) + \nu_{j,n} \mid t_n) \right]. \end{aligned}$$

Let us first examine the nonrandom term $v_{j,n}$. In view of (1) and (6), it follows that

$$\begin{aligned}\overline{F}_Y(q_Y(\alpha_j \mid t_n) + \nu_{j,n} \mid t_n) &= \overline{F}_Z\left(\frac{q_Y(\alpha_j \mid t_n) + \nu_{j,n} - a(t_n)}{b(t_n)}\right) \\ &= \overline{F}_Z\left(q_Z(\alpha_j) + \frac{s_j}{\sqrt{nh}}\right).\end{aligned}$$

Since $\overline{F}_Z(\cdot)$ is differentiable, for all $j \in \{1, \ldots, J\}$, there exists $\theta_{j,n} \in (0,1)$ such that

$$\overline{F}_Z\left(q_Z(\alpha_j) + \frac{s_j}{\sqrt{nh}}\right) = \alpha_j - \frac{s_j}{\sqrt{nh}} f_Z\left(q_Z(\alpha_j) + \frac{s_j \theta_{j,n}}{\sqrt{nh}}\right).$$

In view of the continuity of $f_Z(\cdot)$ and since $s_j/\sqrt{nh} \to 0$ as $n \to \infty$, it follows that

$$f_Z\left(q_Z(\alpha_j) + \frac{s_j \theta_{j,n}}{\sqrt{nh}}\right) = \frac{1 + o(1)}{H_Z(\alpha_j)},$$

leading to

$$v_{j,n} = \frac{s_j}{H_Z(\alpha_j)}(1 + o(1)). \tag{22}$$

Let us now turn to the random variable $V_{j,n}$. For all $j = 1, \ldots, J$, let

$$y_{j,n} = q_Y(\alpha_j \mid t_n) + \nu_{j,n} = q_Y(\alpha_j \mid t_n) + b(t_n)\frac{s_j}{\sqrt{nh}} =: q_Y(\alpha_j \mid t_n) + b(t_n)\epsilon_{j,n},$$

where $\epsilon_{j,n} \to 0$ as $n \to \infty$. Then, Proposition 1 entails that

$$\left\{\sqrt{nh}\left(\hat{\overline{F}}_{n,Y}(y_{j,n} \mid t_n) - \overline{F}_Y(y_{j,n} \mid t_n)\right)\right\}_{j=1,\ldots,J} = \{V_{j,n}\}_{j=1,\ldots,J}$$

converges to a centered Gaussian random variable with covariance matrix $\|K\|_2^2\, B$. Taking account of (22) yields that W_n converges to the cumulative distribution function of a centered Gaussian distribution with covariance matrix $\|K\|_2^2\, C$, evaluated at $(s_1, \ldots, s_J)$, which is the desired result. □

The following proposition provides a uniform consistency result for the estimator (12) of conditional quantiles of Y given a sequence of design points (not too close from the boundaries 0 and 1).

Proposition 3. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $I_n = \{\lfloor nh \rfloor, \ldots, n - \lfloor nh \rfloor\}$ *and suppose* $nh/\log n \to \infty$ *and* $nh^3/\log n \to 0$ *as* $n \to \infty$. *Then, for all* $\alpha \in (0,1)$,

$$\sqrt{\frac{nh}{\log n}} \max_{i \in I_n} \left|\frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)}\right| = O_{\mathbb{P}}(1).$$

Proof. Let $\epsilon \in (0,1)$ and $\alpha \in (0,1)$. Define $v_n = (nh/\log n)^{1/2}$,

$$M(\epsilon,\alpha) = 2\|K\|_2 H_Z(\alpha)\left(\alpha(1-\alpha)\left(1-\log(\epsilon/2)\right)\right)^{1/2},$$

and for all $i \in I_n$ let $q_{i,n}^{\pm} = q_Y(\alpha \mid x_i) \pm M(\epsilon,\alpha) b(x_i)/v_n$. Let us consider the expansion :

$$\begin{aligned}
\delta_n &:= \mathbb{P}\left(v_n \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)}\right| \geq M(\epsilon,\alpha)\right) \\
&= \mathbb{P}\left(\bigcup_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)}\right| \geq M(\epsilon,\alpha)/v_n\right) \\
&= \mathbb{P}\left(\bigcup_{i\in I_n}\left\{\hat{q}_{n,Y}(\alpha \mid x_i) \geq q_{i,n}^{+}\right\} \cup \left\{\hat{q}_{n,Y}(\alpha \mid x_i) \leq q_{i,n}^{-}\right\}\right) \\
&= \mathbb{P}\left(\bigcup_{i\in I_n}\left\{\alpha \leq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right)\right\} \cup \left\{\alpha \geq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^{-} \mid x_i\right)\right\}\right) \\
&\leq \mathbb{P}\left(\bigcup_{i\in I_n}\left\{\alpha - \mathbb{E}\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right) \leq (\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y})(q_{i,n}^{+} \mid x_i)\right\}\right) \\
&+ \mathbb{P}\left(\bigcup_{i\in I_n}\left\{\alpha - \mathbb{E}\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{-} \mid x_i\right) \geq (\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y})(q_{i,n}^{-} \mid x_i)\right\}\right) \\
&=: \mathbb{P}\left(\bigcup_{i\in I_n}\alpha_{i,n}^{+} \leq \xi_{i,n}^{+}\right) + \mathbb{P}\left(\bigcup_{i\in I_n}\alpha_{i,n}^{-} \geq \xi_{i,n}^{-}\right) \\
&=: \delta_n^{+} + \delta_n^{-}.
\end{aligned}$$

Let us focus on the term δ_n^{+}. Assumption $nh/\log n \to \infty$ entails that $v_n \to \infty$ as $n \to \infty$ and thus $q_{i,n}^{+}$ is bounded. Therefore Lemma 3(i) yields

$$\begin{aligned}
\alpha_{i,n}^{+} &:= \alpha - \mathbb{E}\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right) = \alpha - \bar{F}_Y\left(q_{i,n}^{+} \mid x_i\right)(1+O(h)) \\
&= \bar{F}_Z(q_Z(\alpha)) - \bar{F}_Z\left(q_Z(\alpha) + \frac{M(\epsilon,\alpha)}{v_n}\right)(1+O(h)) \\
&= \frac{M(\epsilon,\alpha)}{v_n} f_Z\left(q_Z(\alpha) + \frac{M(\epsilon,\alpha)}{v_n}\theta\right) + O(h),
\end{aligned}$$

for some $\theta \in (0,1)$. Since $f_Z(\cdot)$ is continuous, it follows that

$$\alpha_{i,n}^{+} = \frac{M(\epsilon,\alpha)}{v_n H_Z(\alpha)}(1+o(1)) + O(h) =: \frac{\kappa_1(\epsilon,\alpha)}{v_n}(1+o(1)),$$

in view of the assumption $nh^3/\log n \to 0$ as $n \to \infty$. As a preliminary result,

$$\delta_n^+ = \mathbb{P}\left(\bigcup_{i\in I_n} \xi_{i,n}^+ \geq \frac{\kappa_1(\epsilon,\alpha)}{v_n}(1+o(1))\right) \leq \sum_{i\in I_n} \mathbb{P}\left(\xi_{i,n}^+ \geq \frac{\kappa_1(\epsilon,\alpha)}{v_n}(1+o(1))\right). \tag{23}$$

In addition,

$$\begin{aligned}\mathbb{P}\left(\xi_{i,n}^+ \geq \frac{\kappa_1(\epsilon,\alpha)}{v_n}(1+o(1))\right) &= \mathbb{P}\left((\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y})(q_{i,n}^+ \mid x_i) \geq \frac{\kappa_1(\epsilon,\alpha)}{v_n}(1+o(1))\right)\\ &:= \mathbb{P}\left(\sum_{j=1}^n \tilde{X}_j \geq \frac{\kappa_1(\epsilon,\alpha)}{v_n}(1+o(1))\right),\end{aligned} \tag{24}$$

where, for all $j = 1, \dots, n$, the random variables

$$\tilde{X}_j := \left[\mathbb{1}_{\{Y_j > q_{i,n}^+\}} - \mathbb{P}\left(Y_j > q_{i,n}^+ \mid x_i\right)\right] \int_{x_{j-1}}^{x_j} K_h(x_i - s)ds$$

are independent, centered and bounded :

$$|\tilde{X}_j| \leq \int_{x_{j-1}}^{x_j} K_h(x_i - s)ds \leq \frac{\|K\|_\infty}{nh}.$$

Lemma 3(ii) entails

$$\begin{aligned}\sum_{j=1}^n \mathbb{E}(\tilde{X}_j^2) &= var\left(\sum_{j=1}^n \tilde{X}_j\right) = var\left[\hat{\bar{F}}_{n,Y}\left(q_{i,n}^+ \mid x_i\right)\right]\\ &= \frac{\bar{F}_Y\left(q_{i,n}^+ \mid x_i\right) F_Y\left(q_{i,n}^+ \mid x_i\right)}{nh}\|K\|_2^2(1+o(1)).\end{aligned}$$

Besides, $q_{i,n}^+ \to q_Y(\alpha \mid x_i)$ as $n \to \infty$ and thus $\bar{F}_Y\left(q_{i,n}^+ \mid x_i\right) \to \alpha$ as $n \to \infty$ in view of the continuity of $\bar{F}_Y(\cdot \mid x_i)$. It follows that,

$$\sum_{j=1}^n \mathbb{E}(\tilde{X}_j^2) = \frac{\alpha(1-\alpha)}{nh}\|K\|_2^2(1+o(1)) =: \frac{\kappa_2(\alpha)}{nh}(1+o(1)).$$

Applying Bernstein's inequality for bounded random variables yields

$$
\begin{aligned}
(24) \quad &\leq \exp\left(-\frac{\frac{\kappa_1^2(\epsilon,\alpha)\log n(1+o(1))}{2nh}}{\frac{\kappa_2(\alpha)(1+o(1))}{nh}+\frac{\kappa_1(\epsilon,\alpha)(1+o(1))}{3nhv_n}}\right) \\
&= \exp\left(-\frac{\kappa_1^2(\epsilon,\alpha)\log n}{2\kappa_2(\alpha)+\frac{2\kappa_1(\epsilon,\alpha)(1+o(1))}{3v_n}}(1+o(1))\right) \\
&= \exp\left(-\frac{\kappa_1^2(\epsilon,\alpha)\log n}{2\kappa_2(\alpha)}(1+o(1))\right) \\
&= \exp\left[-2\left(1-\log(\epsilon/2)\right)\log n\left(1+o(1)\right)\right] \\
&\leq \exp\left[-\left(1-\log(\epsilon/2)\right)\log n\right], \qquad (25)
\end{aligned}
$$

for n large enough. Collecting (23)-(25) yields

$$\delta_n^+ \leq n\exp\left[-\left(1-\log(\epsilon/2)\right)\log n\right] = \exp\left(\log(\epsilon/2)\log n\right) \leq \epsilon/2$$

for n large enough. The proof that $\delta_n^- \leq \epsilon/2$ follows the same lines. As a conclusion, we have shown that, for all $\alpha \in (0,1)$ and $\epsilon \in (0,1)$ there exists $M(\epsilon,\alpha) > 0$ such that

$$\mathbb{P}\left(\sqrt{\frac{nh}{\log n}}\max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)}\right| \geq M(\epsilon,\alpha)\right) \leq \epsilon,$$

which is the desired result. □

7.3 Proofs of main results

The proof of Theorem 1 directly relies on Proposition 2 :

Proof of Theorem 1. Let us remark that

$$\frac{\sqrt{nh}}{b(t_n)}\begin{pmatrix}\hat{a}_n(t_n) - a(t_n) \\ \hat{b}_n(t_n) - b(t_n)\end{pmatrix} = \tilde{A}\xi_n,$$

where $\tilde{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$ and $\xi_n = \frac{\sqrt{nh}}{b(t_n)}\begin{pmatrix}\hat{q}_{n,Y}(\mu_3 \mid t_n) - q_Y(\mu_3 \mid t_n) \\ \hat{q}_{n,Y}(\mu_2 \mid t_n) - q_Y(\mu_2 \mid t_n) \\ \hat{q}_{n,Y}(\mu_1 \mid t_n) - q_Y(\mu_1 \mid t_n)\end{pmatrix}$.
Applying Proposition 2 with $J = 3$, $\alpha_1 = \mu_1$, $\alpha_2 = \mu_2$ and $\alpha_3 = \mu_3$ yields

$$\xi_n \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^3},\ \|K\|_2^2\, C\right),$$

where

$$C=\begin{pmatrix}\mu_1(1-\mu_1)H_Z^2(\mu_1) & \mu_2(1-\mu_1)H_Z(\mu_2)(H_Z(\mu_1) & \mu_3(1-\mu_1)H_Z(\mu_3)H_Z(\mu_1)\\ \mu_2(1-\mu_1)H_Z(\mu_2)H_Z(\mu_1) & \mu_2(1-\mu_2)H_Z^2(\mu_2) & \mu_3(1-\mu_2)H_Z(\mu_2)H_Z(\mu_3)\\ \mu_3(1-\mu_1)H_Z(\mu_3)H_Z(\mu_1) & \mu_3(1-\mu_2)H_Z(\mu_2)H_Z(\mu_3) & \mu_3(1-\mu_3)H_Z^2(\mu_3)\end{pmatrix}.$$

Therefore,

$$\tilde{A}\xi_n \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^2},\ \|K\|_2^2\,\tilde{A}C\tilde{A}^t\right),$$

and the conclusion follows from standard calculations. □

Theorem 2 is a straightforward consequence of Proposition 3 :

Proof of Theorem 2. Remarking that

$$\max_{i\in I_n}\left|\frac{\hat{a}_n(x_i)-a(x_i)}{b(x_i)}\right| = \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\mu_2\mid x_i)-q_Y(\mu_2\mid x_i)}{b(x_i)}\right|,$$

the first part of the result is a consequence of Proposition 3 applied with $\alpha=\mu_2$. Similarly,

$$\begin{aligned}\max_{i\in I_n}\left|\frac{\hat{b}_n(x_i)-b(x_i)}{b(x_i)}\right| &= \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\mu_3\mid x_i)-\hat{q}_{n,Y}(\mu_1\mid x_i)-q_Y(\mu_3\mid x_i)+q_Y(\mu_1\mid x_i)}{b(x_i)}\right|\\ &\le \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\mu_3\mid x_i)-q_Y(\mu_3\mid x_i)}{b(x_i)}\right|\\ &+ \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\mu_1\mid x_i)-q_Y(\mu_1\mid x_i)}{b(x_i)}\right|,\end{aligned}$$

and the conclusion follows from Proposition 3 successively applied with $\alpha=\mu_3$ and $\alpha=\mu_1$. □

Proof of Theorem 3. Let us consider the expansion

$$\sqrt{k_n}(\hat{\gamma}_n-\gamma)=\sqrt{k_n}(\hat{\gamma}_n-\tilde{\gamma}_n)+\sqrt{k_n}(\tilde{\gamma}_n-\gamma)=:\Upsilon_{1,n}+\Upsilon_{2,n},$$

where

$$\tilde{\gamma}_n=\frac{1}{k_n}\sum_{i=0}^{k_n-1}\log Z_{m_n-i,m_n}-\log Z_{m_n-k_n,m_n}$$

is the Hill estimator computed on the unobserved random variables $Z_1,\dots,Z_n$. The first term is controlled by remarking that

$$\begin{aligned} |\Upsilon_{1,n}| &= \frac{1}{\sqrt{k_n}} \left| \sum_{i=0}^{k_n-1} \log(\hat{Z}_{m_n-i,m_n} - \log Z_{m_n-i,m_n}) - (\log \hat{Z}_{m_n-k_n,m_n} - \log Z_{m_n-k_n,m_n}) \right| \\ &\leq \frac{1}{\sqrt{k_n}} \sum_{i=0}^{k_n-1} \left| \log \frac{\hat{Z}_{m_n-i,m_n}}{Z_{m_n-i,m_n}} \right| + \left| \log \frac{\hat{Z}_{m_n-k_n,m_n}}{Z_{m_n-k_n,m_n}} \right| \\ &\leq \sqrt{k_n} \max_{0\leq i\leq k_n} \left| \log \frac{\hat{Z}_{m_n-i,m_n}}{Z_{m_n-i,m_n}} \right|. \end{aligned}$$

Combining Lemma 4 and Lemma 5 yields

$$|\Upsilon_{1,n}| = O_{\mathbb{P}}\left(\sqrt{\frac{k_n \log n}{nh}} \right) = o_{\mathbb{P}}(1), \tag{26}$$

in view of the assumption $nh/(k_n \log n) \to \infty$ as $n \to \infty$. Let us now focus on $\Upsilon_{2,n}$. Remarking that $m_n \sim n$ as $n \to \infty$ it is clear that $m_n/k_n \to \infty$ as $n \to \infty$. Besides, since $|A| \in \mathcal{RV}_\rho$, we thus have $A(m_n/k_n) \sim A(n/k_n)$ as $n \to \infty$. Therefore, $\sqrt{k_n} A(m_n/k_n) \to 0$ as $n \to \infty$ and, since $Z_1, \dots, Z_n$ are iid from (2), classical results on Hill estimator apply, see for instance [27, Theorem 3.2.5], leading to

$$\Upsilon_{2,n} \xrightarrow{d} \mathcal{N}(0, \gamma^2). \tag{27}$$

The conclusion follows by combining (26) and (27). □

References

[1] Abdi, S., Abdi, A., Dabo-Niang, S. and Diop, A. (2010). Consistency of a nonparametric conditional quantile estimator for random fields. *Mathematical Methods of Statistics*, 19(1), 1–21.

[2] Antoch, J. and Janssen, P. (1989). Nonparametric regression M-quantiles. *Statistics and Probability Letters*, 8(4), 355–362.

[3] Beirlant, J. and Goegebeur, Y. (2003). Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis*, 42, 595–619, 2003.

[4] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004). *Statistics of extremes : theory and applications*, John Wiley and Sons.

[5] Berlinet, A., Gannoun, A. and Matzner-Løber, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles, *Statistics*, 35(2), 139–169.

[6] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation*, Encyclopedia of Mathematics and its Applications, 27, Cambridge University Press.

[7] Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society : Series C*, 54, 207–222.

[8] Cowling, A. and Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society : Series B*, 58, 551–563.

[9] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589.

[10] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20(2), 311–333.

[11] Davison, A. C. and Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society : Series B*, 62, 191–208.

[12] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society : Series B*, 52, 393–442.

[13] Falk, M., Hüsler, J. and Reiss, R. D. (2004). *Laws of small numbers : Extremes and rare events*, 2nd edition, Birkhäuser.

[14] Gangopadhyay, A. K. (1995). A note on the asymptotic behavior of conditional extremes, *Statistics and Probability Letters*, 25, 163–170.

[15] Gardes, L. (2015). A general estimator for the extreme value index : applications to conditional and heteroscedastic extremes. *Extremes*, 18(3), 479–510.

[16] Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99(10), 2368–2388.

[17] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions : an application to the estimation of extreme rainfall return levels. *Extremes*, 13(2), 177–204.

[18] Gardes, L. and Girard, S. (2012). Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics*, 6, 1715–1744.

[19] Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101(2), 419–433.

[20] Gardes, L., Guillou, A. and Schorgen, A. (2012). Estimating the conditional tail index by integrating a kernel conditional quantile estimator, *Journal of Statistical Planning and Inference*, 142(6), 1586–1598.

[21] Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1), 45–75.

[22] Girard, S. and Stupfler, G. (2019). Estimation of high-dimensional extreme conditional expectiles. *Working paper.*

[23] Girard, S., Stupfler, G. and A. Usseglio-Carleve (2019). Nonparametric extreme conditional expectile estimation. *Submitted*, `http://hal.inria.fr/hal-02114255`.

[24] Goegebeur, Y., Guillou, A. and Osmann, M. (2014). A local moment type estimator for the extreme value index in regression with random covariates. *The Canadian Journal of Statistics*, 42, 487–507.

[25] Goegebeur, Y., Guillou, A. and Stupfler, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'IHP Probabilités et Statistiques*, 51(3), 1190–1213.

[26] Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes : a review. *International Statistical Review*, 83(2), 263–292.

[27] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory : An Introduction*, New York, Springer.

[28] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1), 60–70.

[29] Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15, 153–167.

[30] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3, 1163–1174.

[31] Jalbert, J., Favre, A. C., Bélisle, C. and Angers, J. F. (2017). A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *Journal of the Royal Statistical Society : Series C*, 66(5), 941–962.

[32] Kyung-Joon, C. and Schucany, W. (1998). Nonparametric kernel regression estimation near endpoints. *Journal of Statistical Planning and Inference*, 66, 289–304.

[33] Müller, H. G. and Prewitt, K. (1991). Applications of multiparameter weak convergence for adaptive nonparametric curve estimation. In *Nonparametric Functional Estimation and Related Topics*, pp. 141–166, Springer, Dordrecht.

[34] Samanta, M. (1989). Non-parametric estimation of conditionnal quantiles. *Statistics and Probability Letters*, 7(5), 407–412.

[35] Smith, R. L. (1989). Extreme value analysis of environmental time series : an application to trend detection in ground-level ozone (with discussion). *Statistical Science*, 4, 367–393.

[36] Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5(4), 595–645.

[37] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, 73(364), 812–815.

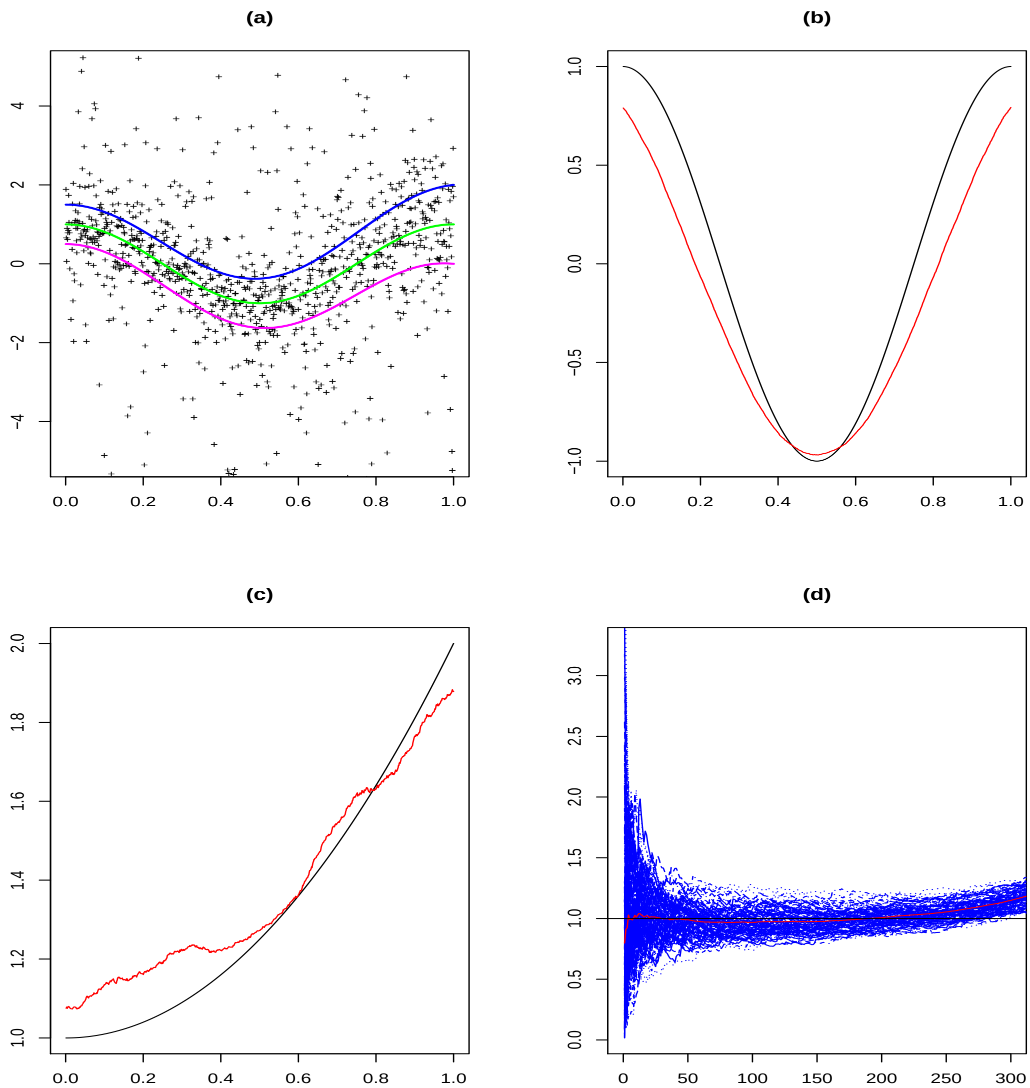

Figure 1: Student distribution with $\nu = 1$ df. (a) : Simulated data (+) and conditional quantiles $q(3/4|\cdot)$ (magenta), $q(1/2|\cdot)$ (green) and $q(1/4|\cdot)$ (blue). (b) : Location function $a(\cdot)$ (black) and mean estimate $\bar{\hat{a}}_n(\cdot)$ (red). (c) : Scale function $b(\cdot)$ (black) and mean estimate $\bar{\hat{b}}_n(\cdot)$ (red). (d) : Conditional tail-index γ (black), estimates $\hat{\gamma}_{n,i}$, $i = 1, \ldots, N$ (blue) and mean estimate $\bar{\hat{\gamma}}_n$ (red) as functions of $k_n \in \{1, \ldots, 300\}$.

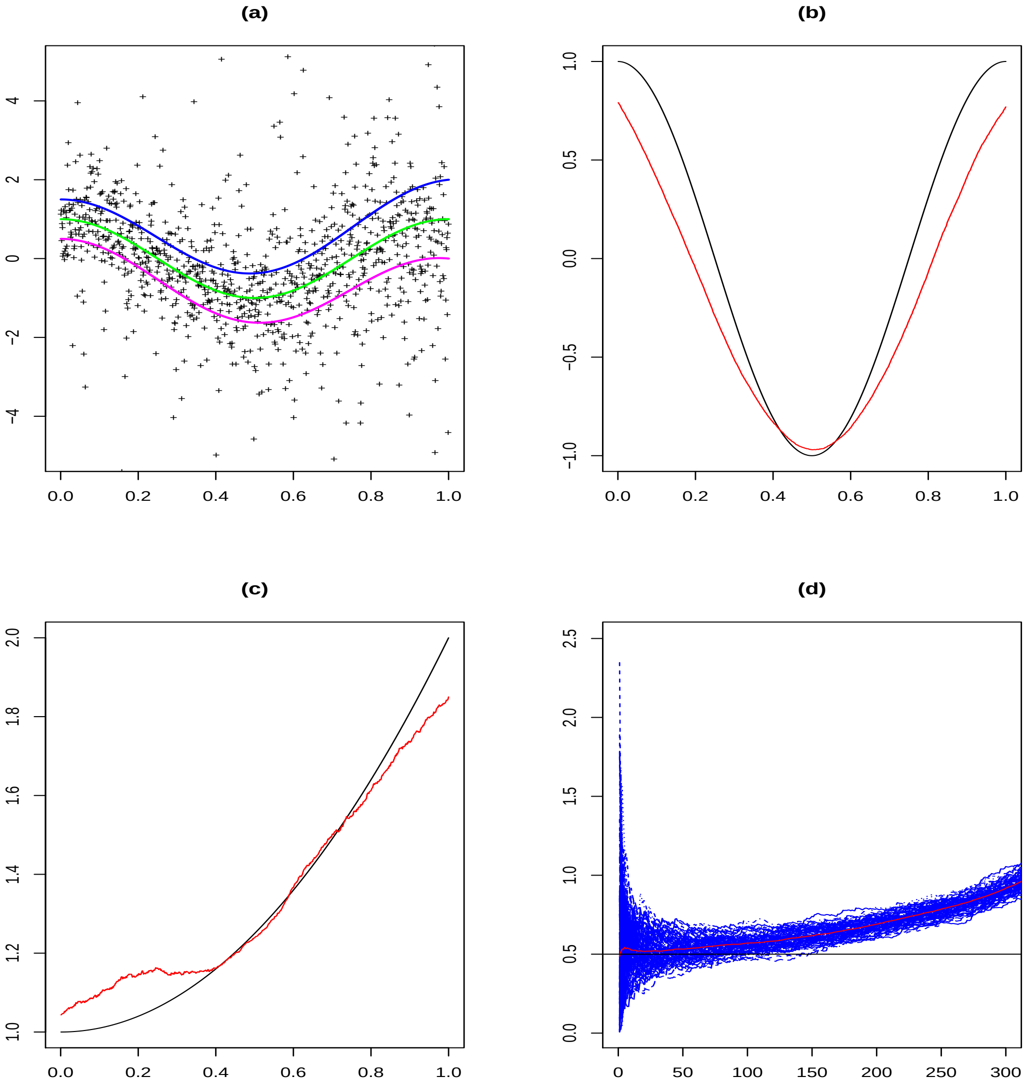

Figure 2: Student distribution with $\nu = 2$ df. (a) : Simulated data (+) and conditional quantiles $q(3/4|\cdot)$ (magenta), $q(1/2|\cdot)$ (green) and $q(1/4|\cdot)$ (blue). (b) : Location function $a(\cdot)$ (black) and mean estimate $\bar{\hat{a}}_n(\cdot)$ (red). (c) : Scale function $b(\cdot)$ (black) and mean estimate $\bar{\hat{b}}_n(\cdot)$ (red). (d) : Conditional tail-index γ (black), estimates $\hat{\gamma}_{n,i}$, $i = 1, \dots, N$ (blue) and mean estimate $\bar{\hat{\gamma}}_n$ (red) as functions of $k_n \in \{1, \dots, 300\}$.

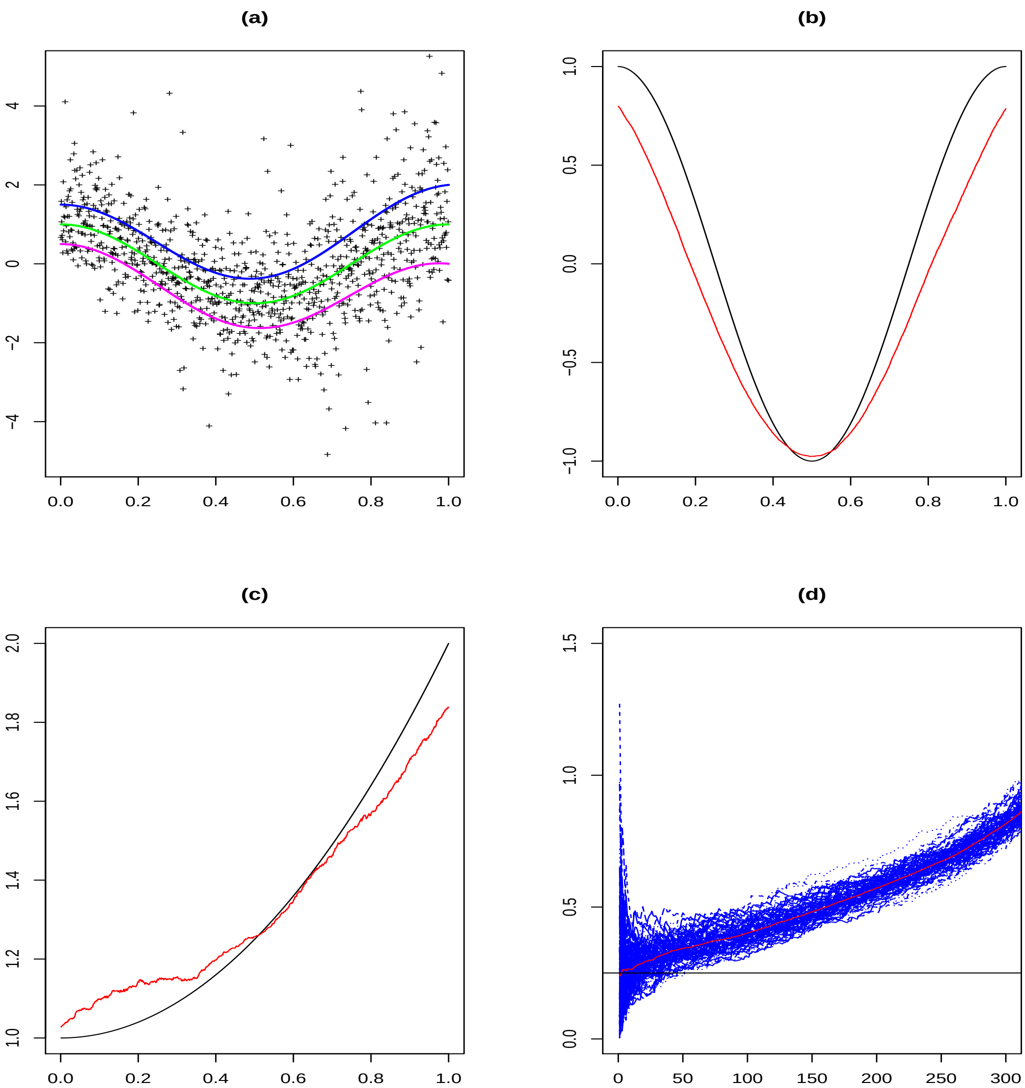

Figure 3: Student distribution with $\nu = 4$ df. (a) : Simulated data (+) and conditional quantiles $q(3/4|\cdot)$ (magenta), $q(1/2|\cdot)$ (green) and $q(1/4|\cdot)$ (blue). (b) : Location function $a(\cdot)$ (black) and mean estimate $\bar{\hat{a}}_n(\cdot)$ (red). (c) : Scale function $b(\cdot)$ (black) and mean estimate $\bar{\hat{b}}_n(\cdot)$ (red). (d) : Conditional tail-index γ (black), estimates $\hat{\gamma}_{n,i}$, $i = 1, \dots, N$ (blue) and mean estimate $\bar{\hat{\gamma}}_n$ (red) as functions of $k_n \in \{1, \dots, 300\}$.

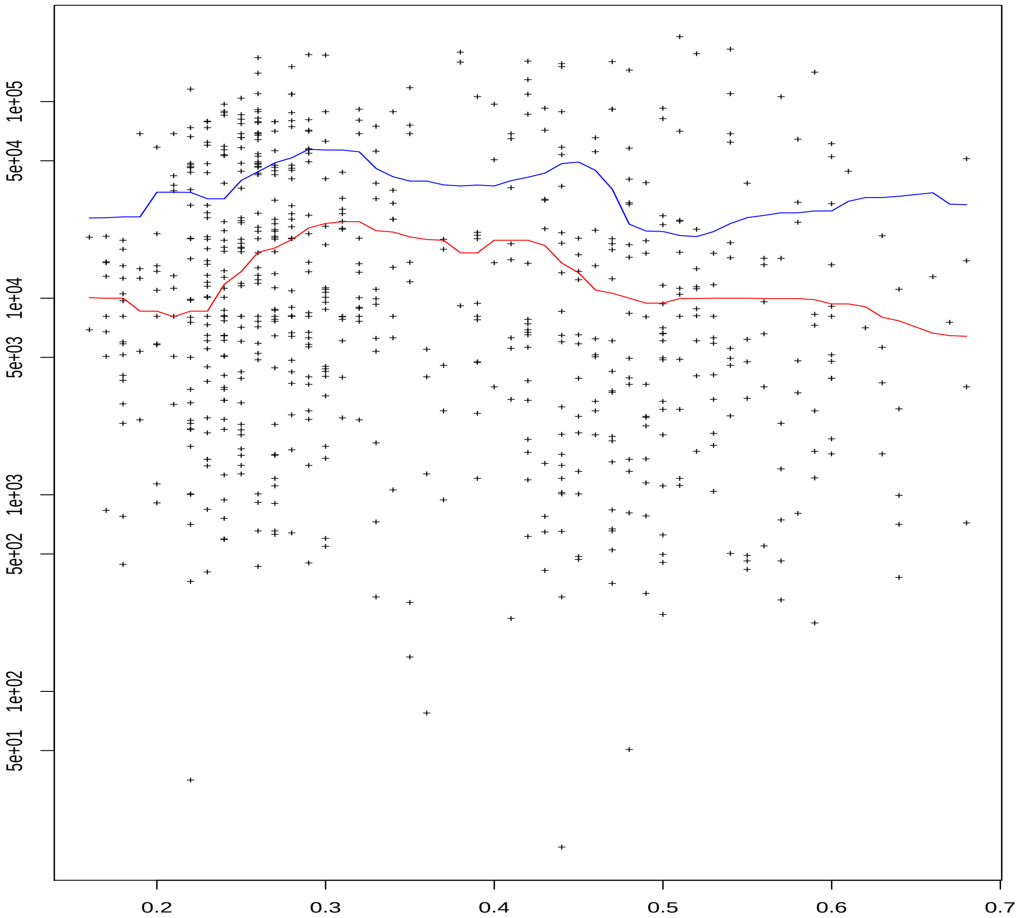

Figure 4: Illustration on motorcycle insurance data. Horizontally : Age of the policyholder (hundred of years), vertically : Claim severity (SEK, log scale). Data (+), estimated location function $\hat{a}_n(\cdot)$ (red) and scale function $\hat{b}_n(\cdot)$ (blue).

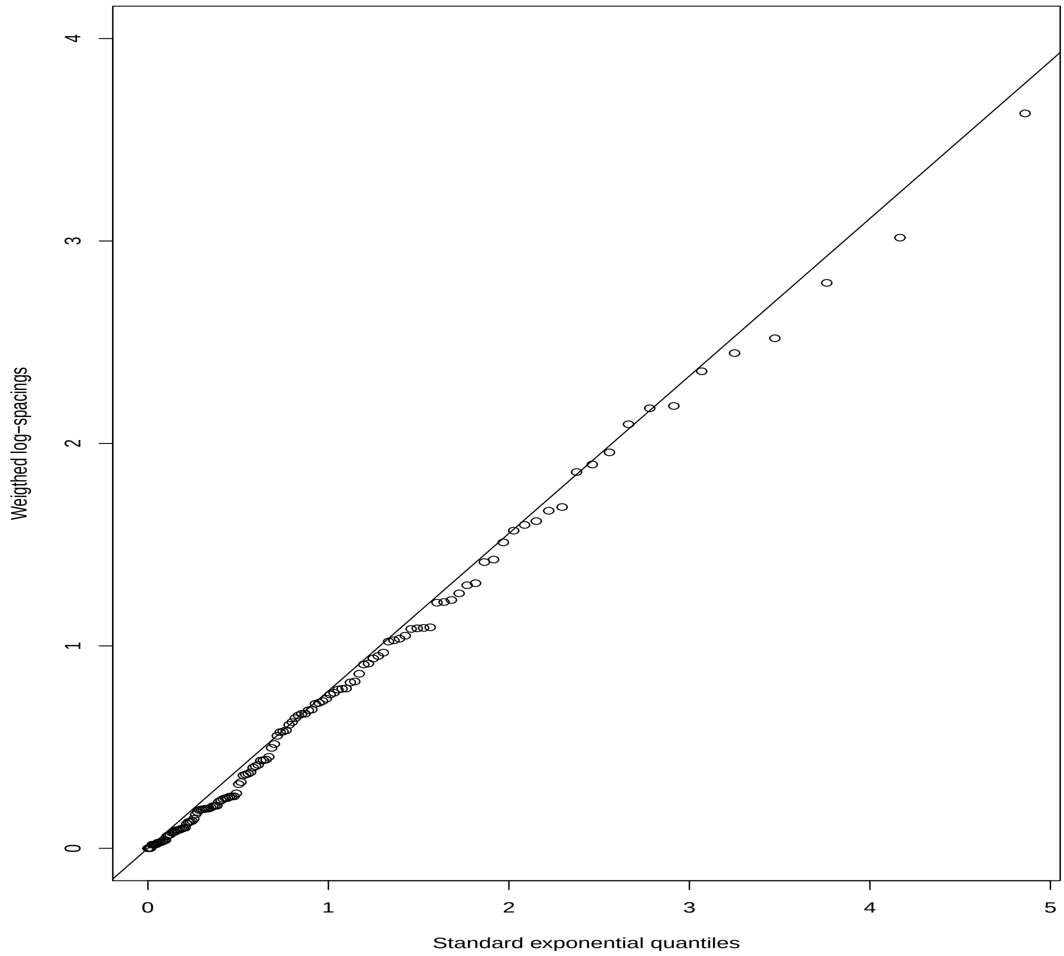

Figure 5: Illustration on motorcycle insurance data, quantile-quantile plot. Horizontally : Standard exponential quantiles, vertically : Weighted log-spacings computed on the residuals. The continuous line has slope $\hat{\gamma}_n$.

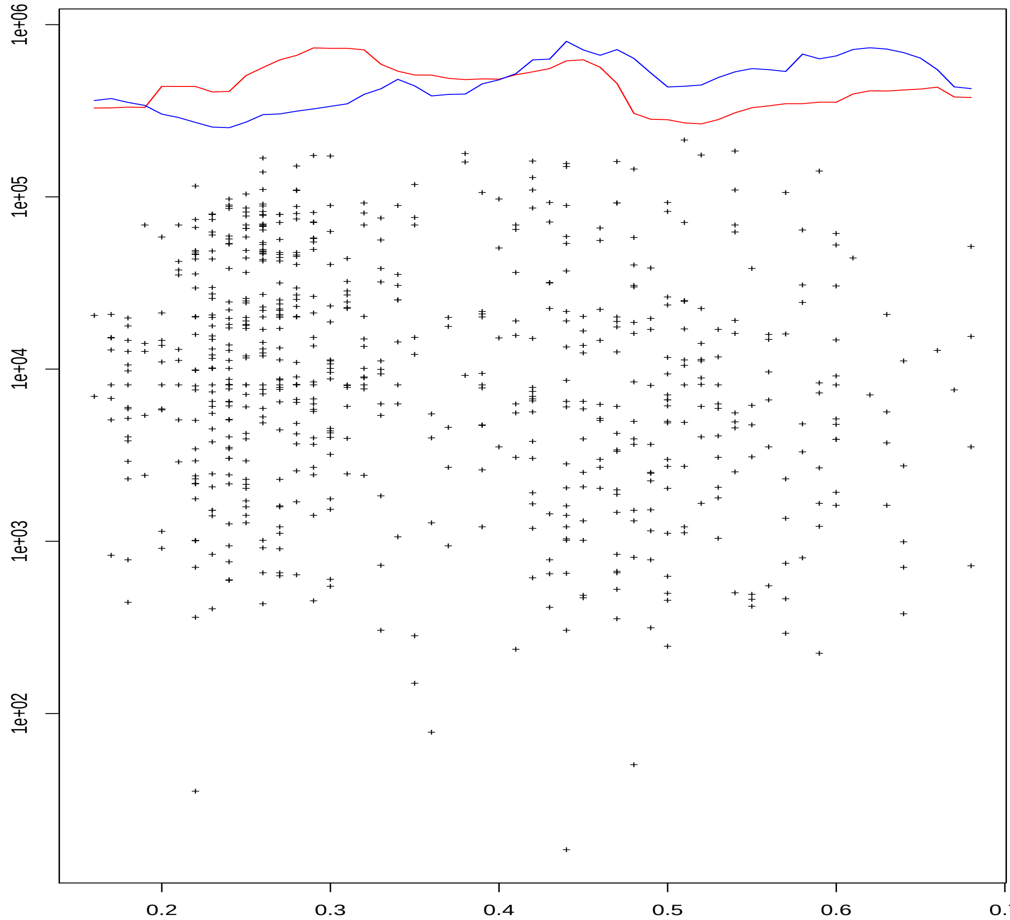

Figure 6: Illustration on motorcycle insurance data. Horizontally : Age of the policyholder (hundred of years), vertically : Claim severity (SEK, log scale). Data (+), nonparametric conditional Weissman estimator $\check{q}_{n,Y}(\alpha_n \mid \cdot)$ (blue) and semi-parametric extreme quantile estimator $\tilde{q}_{n,Y}(\alpha_n \mid \cdot)$ (red).

Chapter 3

Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model

This chapter is presented below as an article submitted for publication and online at https://hal.inria.fr/hal-02486937v2.

Summary

Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model

Abstract

We consider a location-dispersion regression model for heavy-tailed distributions when the multidimensional covariate is deterministic. In a first step, nonparametric estimators of the regression and dispersion functions are introduced. This permits, in a second step, to derive an estimator of the conditional extreme-value index computed on the residuals. Finally, a plug-in estimator of extreme conditional quantiles is built using these two preliminary steps. It is shown that the resulting semi-parametric estimator is asymptotically Gaussian and may benefit from the same rate of convergence as in the unconditional situation. Its finite sample properties are illustrated both on simulated and real tsunami data.

1 Introduction

The modeling of extreme events arises in many fields such as finance, insurance or environmental science. A recurrent statistical problem is then the estimation of extreme quantiles associated with a random variable Y, see the reference books [1, 13, 24]. In many situations, Y is recorded simultaneously with a multidimensional covariate $x \in \mathbb{R}^d$, the goal being to describe how tail characteristics such as extreme quantiles or small exceedance probabilities of the response variable Y may depend on the explanatory variable x. Motivating examples include the study of extreme rainfall as a function of the geographical location [17], the assessment of the optimal cost of the delivery activity in postal services [7], the analysis of longevity [30], the description of the upper tail of claim size distributions [1], the modeling of extremes in environmental time series [37], etc.

Here, we focus on the challenging situation where Y given x is heavy-tailed. Without additional assumptions on the pair (Y, x), the estimation of extreme conditional quantiles is addressed using nonparametric methods, see for instance the recent works of [9, 19, 21].

These methods may however suffer from the curse of dimensionality which is compounded in distribution tails by the fact that observations are rare by definition. These difficulties can be partially overcome by considering parametric models [11, 5]. Semi-parametric methods have also been considered for trend modeling in extreme events [10, 27] : A nonparametric regression model of the trend is combined with a parametric model for extreme values.

Our approach belongs to this second line of works. We assume that the response variable and the covariate are linked by a location-dispersion regression model $Y = a(x) + b(x)Z$, see [39], where Z is a heavy-tailed random variable. This model is flexible since (i) no parametric assumptions are made on $a(\cdot)$, $b(\cdot)$ and Z, (ii) it allows for heteroscedasticity via the function $b(\cdot)$. Moreover, another feature of this model is that Y inherits its tail behavior from Z and thus does not depend on the covariate x. We propose to take profit of this important property to decouple the estimation of the nonparametric and extreme structures. As a consequence, we shall show that the semi-parametric estimators of extreme conditional quantiles of Y given x are asymptotically Gaussian and may benefit from the same rate of convergence as in the unconditional situation. A similar idea is implemented in [29] : An extreme-value distribution with constant extreme-value index is fitted to standardized rainfall maxima. The theoretical study of heteroscedastic extremes has been initiated in [26] and developed in [12, 15] through the introduction of a proportional tails model. The results were applied to trend detection in rainfalls and stock market returns.

This paper is organized as follows. The location-dispersion regression model for heavy-tailed distributions is presented in more details in Section 2. The associated inference methods are described in Section 3 : Estimation of the regression and dispersion functions, estimation of the conditional tail-index and extreme conditional quantiles. Asymptotic results are provided in Section 4 while the finite sample behaviour of the estimators is illustrated in Section 5 on simulated data and in Section 6 on tsunami data. Proofs are postponed to the Appendix.

2 Location-dispersion regression model for heavy-tailed distributions

We consider the class of location-dispersion regression models, where the relation between a random response variable $Y \in \mathbb{R}$ and a deterministic covariate vector $x \in \Pi \subset \mathbb{R}^d$, $d \geq 1$ is given by

$$Y = a(x) + b(x)Z. \tag{1}$$

The real random variable Z is assumed to be heavy-tailed. Denoting by $\bar{F}_Z$ its survival function, one has

$$\bar{F}_Z(z) = z^{-1/\gamma} L(z), \; z > 0. \tag{2}$$

Here, $\gamma > 0$ is called the conditional tail-index and L is a slowly-varying function at infinity *i.e.* for all $t > 0$,

$$\lim_{z \to \infty} \frac{L(tz)}{L(z)} = 1.$$

$\bar{F}_Z$ is said to be regularly varying at infinity with index $-1/\gamma$. This property is denoted for short by $\bar{F}_Z \in \mathcal{RV}_{-1/\gamma}$, see [3] for a detailed account on regular variations. Model (1) has been introduced by [39] in the random design setting where the location function $a : \Pi \to \mathbb{R}$ and the scaling function $b : \Pi \to \mathbb{R}^+ \setminus \{0\}$ are referred to as the regression and dispersion functions respectively. Combining (1) and (2) yields

$$\bar{F}_Y(y \mid x) := \mathbb{P}(Y > y \mid x) = \bar{F}_Z\left(\frac{y - a(x)}{b(x)}\right) = \left(\frac{y - a(x)}{b(x)}\right)^{-1/\gamma} L\left(\frac{y - a(x)}{b(x)}\right), \qquad (3)$$

for $y \geq y_0(x) > a(x)$ where the functions $a(\cdot)$, $b(\cdot)$ and the conditional tail-index γ are unknown. We thus obtain a semi-parametric location-dispersion regression model for the (heavy) tail of Y given x. The main assumption is that the conditional tail-index γ is independent of the covariate. On the one hand, the proposed semi-parametric heteroscedastic modeling offers more flexibility than purely parametric approaches. On the other hand, the location-dispersion structure may circumvent the curse of dimensionality and assuming a constant conditional tail-index γ should yield more reliable estimates in small sample contexts than purely nonparametric approaches. Let us also note that, from (2) and (3), the regular variation property yields $\bar{F}_Y(y \mid x)/\bar{F}_Z(y) \to b(x)^{1/\gamma}$ as $y \to \infty$. The location-dispersion regression model can thus be interpreted as a particular case of the proportional tails model [12] with scedasis function $b(\cdot)^{1/\gamma}$. The practical consequences of this point are further discussed in Section 5.

Starting with an independent n-sample $\{(Y_1, x_1), \ldots, (Y_n, x_n)\}$ from (1), it is clear that, since Z is not observed, $a(\cdot)$ and $b(\cdot)$ may only be estimated up to additive and multiplicative factors. This identifiability issue can be fixed by introducing some constraints on the distribution of Z. To this end, for all $\alpha \in (0,1)$ consider $q_Z(\alpha) = \inf\{z \in \mathbb{R}; \bar{F}_Z(z) \leq \alpha\}$ the αth quantile of Z and let $(\mu_1, \mu_2, \mu_3) \in (0,1)^3$ such that $\mu_3 < \mu_1$ and

$$q_Z(\mu_2) = 0 \text{ and } q_Z(\mu_3) - q_Z(\mu_1) = 1. \qquad (4)$$

Let us note that the constraint (3) can always be fulfilled with *i.e.* $\mu_3 = 1/4$, $\mu_2 = 1/2$ and $\mu_1 = 3/4$ up to an affine transformation of $a(\cdot)$, $b(\cdot)$ and Z such that (1) holds. From (1), for all $\alpha \in (0,1)$, the conditional quantile of Y given $x \in \Pi$ is

$$q_Y(\alpha \mid x) = a(x) + b(x) q_Z(\alpha), \qquad (5)$$

and therefore the regression and dispersion functions are defined in an unique way by

$$a(x) = q_Y(\mu_2 \mid x) \text{ and } b(x) = q_Y(\mu_3 \mid x) - q_Y(\mu_1 \mid x), \tag{6}$$

for all $x \in \Pi$. This remark is the starting point of the inference procedure described hereafter.

3 Inference

Let us denote by λ the Lebesgue measure and $\|\cdot\|$ a norm on $\mathbb{R}^d$, $d \geq 1$. Consider $\{(Y_1, x_1), \dots, (Y_n, x_n)\}$ a n-sample from (1) : $Y_i = a(x_i) + b(x_i)Z_i$, $i = 1, \dots, n$ where $Z_1, \dots, Z_n$ are independent and identically distributed (iid) from the heavy-tailed distribution (2). We assume that the design points x_i, $i = 1, \dots, n$ are all distinct from each other and included in Π, a compact subset of $\mathbb{R}^d$ whose Lebesgue measure of the boundary is zero. Let $\{\Pi_i,\ i = 1, \dots, n\}$ be a partition of Π such that $x_i \in \Pi_i$. A three-stage inference procedure is adopted : The regression and dispersion functions are estimated nonparametrically in Paragraph 3.1, and the conditional tail-index is then computed from the residuals in Paragraph 3.2. Finally, the extreme conditional quantiles are derived by combining a plug-in method with Weissman's extrapolation device [40] in Paragraph 3.3.

3.1 Estimation of the regression and dispersion functions

The proposed procedure relies on the choice of a smoothing estimator for the conditional quantiles. Here, a kernel estimator for $\bar{F}_Y(y \mid x)$ is considered (see for instance [33, 34]). For all $(x, y) \in \Pi \times \mathbb{R}$ let

$$\hat{\bar{F}}_{n,Y}(y \mid x) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_i > y\}} \int_{\Pi_i} K_h(x - t)dt, \tag{7}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function, $K_h(\cdot) := K(\cdot/h)/h^d$ with K a density function on $\mathbb{R}^d$ called a kernel. The associated smoothing parameter $h = h_n \to 0$ as $n \to \infty$ is a nonrandom sequence called the bandwidth. The corresponding estimator of $q_Y(\alpha \mid x)$ is defined for all $(x, \alpha) \in \Pi \times (0, 1)$ by

$$\hat{q}_{n,Y}(\alpha \mid x) = \hat{\bar{F}}^{\leftarrow}_{n,Y}(\alpha \mid x) := \inf\{y;\ \hat{\bar{F}}_{n,Y}(y \mid x) \leq \alpha\}. \tag{8}$$

Nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution function have been extensively investigated, see, for example [2, 35, 38], among others. In view of (6), the regression and dispersion functions are estimated by

$$\hat{a}_n(x) = \hat{q}_{n,Y}(\mu_2 \mid x) \text{ and } \hat{b}_n(x) = \hat{q}_{n,Y}(\mu_3 \mid x) - \hat{q}_{n,Y}(\mu_1 \mid x), \tag{9}$$

for all $x \in \Pi$.

3.2 Estimation of the conditional tail-index

The non-observed $Z_1, \dots, Z_n$ are estimated by the residuals

$$\hat{Z}_i = (Y_i - \hat{a}_n(x_i))/\hat{b}_n(x_i), \tag{10}$$

for all $i = 1, \dots, n$ where $\hat{a}_n(\cdot)$ and $\hat{b}_n(\cdot)$ are given in (9). In practice, nonparametric estimators can suffer from boundary effects [6, 31] and therefore only design points sufficiently far from the boundary of Π are considered. More specifically, consider $\tilde{\Pi}^{(n)} = \{x \in \mathbb{R}^d$, such that $B(x,h) \subset \Pi\}$ the erosion of the set Π by the ball $B(0,h)$ centered at 0 and with radius h, see [36] for further details on mathematical morphology. Denote by I_n the set of indices associated with such design points $I_n = \{i \in \{1, \dots, n\}$ such that $x_i \in \tilde{\Pi}^{(n)}\}$ and let $m_n = \text{card}(I_n)$. It can be shown that $m_n = n(1 + O(h))$, see Lemma 3 in the Appendix.

Finally, let (k_n) be an intermediate sequence of integers, *i.e.* such that $1 < k_n \le n$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. The $(k_n + 1)$ top order statistics associated with the pseudo-observations $\hat{Z}_i$, $i \in I_n$ are denoted by $\hat{Z}_{m_n-k_n,m_n} \le \cdots \le \hat{Z}_{m_n,m_n}$. The conditional tail-index is estimated using a Hill-type statistic [28] :

$$\hat{\gamma}_n = \frac{1}{k_n} \sum_{i=0}^{k_n-1} \log \hat{Z}_{m_n-i,m_n} - \log \hat{Z}_{m_n-k_n,m_n}, \tag{11}$$

built on non iid pseudo-observations.

3.3 Estimation of extreme conditional quantiles

Clearly, the purely nonparametric estimator (8) cannot estimate consistently extreme quantiles of levels α_n arbitrarily small. For instance, when $n\alpha_n \to 0$, the extreme quantile is likely to be larger than the maximum observation. In such a case, an extrapolation technique is necessary to estimate the so-called extreme conditional quantile $q_Y(\alpha_n \mid x)$. To this end, we propose to take profit of the structure of the location-dispersion regression model (5) to define the plugin estimator

$$\tilde{q}_{n,Y}(\alpha_n \mid x) = \hat{a}_n(x) + \hat{b}_n(x)\hat{q}_{n,Z}(\alpha_n), \tag{12}$$

where $\hat{a}_n(x)$ and $\hat{b}_n(x)$ are given in (9) and $\hat{q}_{n,Z}(\alpha_n)$ is the Weissman type estimator [40] :

$$\hat{q}_{n,Z}(\alpha_n) = \hat{Z}_{m_n-k_n,m_n} \left(\frac{\alpha_n m_n}{k_n}\right)^{-\hat{\gamma}_n}. \tag{13}$$

Again, it should be noted that $\hat{q}_{n,Z}(\alpha_n)$ is computed from the non iid pseudo-observations $\hat{Z}_i$, $i \in I_n$. Finally, by construction, the semi-parametric estimator (12) cannot suffer from quantile crossing, a phenomenon which can occur with quantile regression techniques.

4 Main results

The following general assumptions are required to establish the asymptotic behaviour of the estimators. The first one gathers all the conditions to define a location-dispersion regression model for heavy-tailed distributions in a multidimensional fixed design setting.

(A.1) $(Y_1, x_1), \ldots, (Y_n, x_n)$ are independent observations from the location-dispersion regression model for heavy-tailed distributions defined by (1), (2) and (4) and such that

$$\max_{i=1,\ldots,n} \left|\lambda(\Pi_i) - \frac{\lambda(\Pi)}{n}\right| = o(1/n), \tag{14}$$

$$\max_{i=1,\ldots,n} \sup_{(s,t)\in\Pi_i^2} \|s-t\| = O\left(n^{-1/d}\right). \tag{15}$$

We refer to [33, 34] for this definition of the multidimensional fixed design setting.

The second assumption is a regularity condition.

(A.2) The functions $a(\cdot)$ and $b(\cdot)$ are twice continuously differentiable on Π, $b(\cdot)$ is lower bounded on Π, $b(t) \geq b_m > 0$ for all $t \in \Pi$, and the survival function $\bar{F}_Z(\cdot)$ is twice continuously differentiable on $\mathbb{R}$.

Under **(A.1)** and **(A.2)**, the quantile function $q_Z(\cdot)$ and the density $f_Z(\cdot) = -\bar{F}'_Z(\cdot)$ exist and we let $H_Z(\cdot) := 1/f_Z(q_Z(\cdot))$ the quantile density function and $U_Z(\cdot) = q_Z(1/\cdot)$ the tail quantile function of Z. Moreover, the conditional survival function of Y is twice continuously differentiable with respect to its second argument. The next assumption is standard in the nonparametric kernel estimation framework.

(A.3) K is a bounded and even density with symmetric support $S \subset B(0,1)$ the unit ball of $\mathbb{R}^d$ and verifying the Lipschitz property : There exists $c_K > 0$ such that

$$|K(u) - K(v)| \leq c_K \|u - v\|,$$

for all $(u,v) \in S^2$.

Under **(A.3)**, let $\|K\|_\infty = \sup_{t\in S} K(t)$ and $\|K\|_2 = \left(\int_S K^2(t)dt\right)^{1/2}$. Finally, the so-called second-order condition is introduced (see for instance [24, eq (3.2.5)] :

(A.4) For all $t > 0$, as $z \to \infty$,

$$\frac{U_Z(tz)}{U_Z(z)} - t^\gamma \sim A(z)t^\gamma \frac{t^\rho - 1}{\rho},$$

where $\gamma > 0$, $\rho < 0$ and A is a positive or negative function such that $A(z) \to 0$ as $z \to \infty$.

From [3, Theorem 1.5.12], property (2) is equivalent to $U_Z \in \mathcal{R}V_\gamma$, that is $U_Z(tz)/U_Z(z) \to t^\gamma$ as $z \to \infty$ for all $t > 0$. The role of the second-order condition **(A.4)** is thus to control the rate of the previous convergence thanks to the function $A(\cdot)$. Moreover, it can be shown that $|A|$ is regularly varying with index ρ, see [24, Lemma 2.2.3]. It is then clear that ρ, referred to as the (conditional) second-order parameter, is a crucial quantity, tuning the rate of convergence of most extreme-value estimators, see [24, Chapter 3] for examples. A list of distributions satisfying **(A.4)** is provided in Table 1 together with the associated values of γ and ρ. Similarly to [34], the dimension $d = 4$ plays a special role and we thus introduce

Distribution (parameters)	Density function	γ	ρ
Generalised Pareto ($\sigma, \xi > 0$)	$\sigma^{-1}\left(1+\xi t/\sigma\right)^{-1-1/\xi}$ ($t > 0$)	ξ	$-\xi$
Burr ($\alpha, \beta > 0$)	$\alpha\beta t^{\alpha-1}\left(1+t^\alpha\right)^{-\beta-1}$ ($t > 0$)	$1/(\alpha\beta)$	$-1/\beta$
Fréchet ($\alpha > 0$)	$\alpha t^{-\alpha-1}\exp\left(-t^{-\alpha}\right)$ ($t > 0$)	$1/\alpha$	-1
Fisher ($\nu_1, \nu_2 > 0$)	$\frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} t^{\nu_1/2-1}(1+\nu_1 t/\nu_2)^{-(\nu_1+\nu_2)/2}$ ($t > 0$)	$2/\nu_2$	$-2/\nu_2$
Inverse Gamma ($\alpha, \beta > 0$)	$\frac{\beta^\alpha}{\Gamma(\alpha)} t^{-\alpha-1}\exp(-\beta/t)$ ($t > 0$)	$1/\alpha$	$-1/\alpha$
Student ($\nu > 0$)	$\frac{1}{\sqrt{\nu\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$1/\nu$	$-2/\nu$

Table 1: A list of heavy-tailed distributions satisfying **(A.4)** with the associated values of γ and ρ. $\Gamma(\cdot)$ and $B(\cdot,\cdot)$ denote the Gamma and Beta functions respectively.

for all $d \ge 1$:

$$\kappa(d) = \begin{vmatrix} 4 & \text{if} \quad d \le 4 \\ 2d/(d-2) & \text{if} \quad d \ge 4. \end{vmatrix}$$

Our first result states the joint asymptotic normality of the estimators (9) of the regression and dispersion functions.

Theorem 1. *Assume* **(A.1)**, **(A.2)**, **(A.3)** *hold and* $f_Z(q_Z(\mu_j)) > 0$ *for* $j \in \{1, 2, 3\}$. *If* $nh^d \to \infty$ *and* $nh^{d+\kappa(d)} \to 0$ *as* $n \to \infty$ *then, for all sequence* $(t_n) \subset \tilde{\Pi}^{(n)}$,

$$\frac{\sqrt{nh^d}}{b(t_n)} \begin{pmatrix} \hat{a}_n(t_n) - a(t_n) \\ \hat{b}_n(t_n) - b(t_n) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^2},\ \lambda(\Pi)\|K\|_2^2\ \Sigma\right),$$

where the coefficients of the matrix Σ *are given by*

$$\begin{aligned} \Sigma_{1,1} &= \mu_2(1-\mu_2)H_Z^2(\mu_2), \\ \Sigma_{1,2} = \Sigma_{2,1} &= \mu_2(1-\mu_1)H_Z(\mu_1)H_Z(\mu_2) - \mu_3(1-\mu_2)H_Z(\mu_2)H_Z(\mu_3), \\ \Sigma_{2,2} &= \mu_1(1-\mu_1)H_Z^2(\mu_1) - 2\mu_3(1-\mu_1)H_Z(\mu_1)H_Z(\mu_3) + \mu_3(1-\mu_3)H_Z^2(\mu_3). \end{aligned}$$

A uniform consistency result can also be established :

Theorem 2. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. If, moreover,* $nh^d/\log n \to \infty$ *and* $nh^{d+\kappa(d)}/\log n \to 0$ *as* $n \to \infty$, *then,*

$$\sqrt{\frac{nh^d}{\log n}} \max_{i \in I_n} \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1) \ \textit{and} \ \sqrt{\frac{nh^d}{\log n}} \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1).$$

As a consequence of Theorem 2, one can prove that the residuals $\hat{Z}_i = (Y_i - \hat{a}_n(x_i))/\hat{b}_n(x_i)$, see (10), are close to the unobserved Z_i, $i = 1, \dots, n$.

Corollary 1. *Under the assumptions of Theorem 2, for all* $i \in I_n$,

$$|\hat{Z}_i - Z_i| \leq R_{n,i}(1 + |Z_i|), \ \textit{where} \ \max_{i \in I_n} R_{n,i} = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right) = o_{\mathbb{P}}(1).$$

Our next main result provides the asymptotic normality of the conditional tail-index estimator (11) and the Weissman estimator (13) computed on the residuals.

Theorem 3. *Assume* **(A.1)**-**(A.4)** *hold. Let* (k_n) *be an intermediate sequence of integers such that* $nh^d/(k_n \log n) \to \infty$, $nh^{d+\kappa(d)}/\log n \to 0$ *and* $\sqrt{k_n}A(n/k_n) \to \beta \in \mathbb{R}$ *as* $n \to \infty$. *Then,*

(i) $\sqrt{k_n}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2)$.

(ii) For all sequence $(\alpha_n) \subset (0, 1)$ *such that* $n\alpha_n/k_n \to 0$ *and* $\log(n\alpha_n)/\sqrt{k_n} \to 0$ *as*

$n \to \infty$,

$$\frac{\sqrt{k_n}}{\log\left(\frac{k_n}{n\alpha_n}\right)}\left(\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n)\right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2).$$

It appears that, in the location-dispersion regression model, the tail-index can be estimated at the same rate $1/\sqrt{k_n}$ as in iid case, see [22] for a review. As expected, this semi-parametric framework is a more favorable situation than the purely nonparametric one for the estimation of the conditional tail-index where the rate of convergence $1/\sqrt{k_n h^d}$ is impacted by the covariate, see for instance [9, Corollary 1 & 2], [8, Theorem 3] and [21, Theorem 2]. To be more specific, remark first that conditions $nh^d/(k_n \log n) \to \infty$ and $nh^{d+\kappa(d)}/\log n \to 0$ imply that $k_n = o\left((n/\log n)^{\kappa(d)/(d+\kappa(d))}\right)$. Second, following [24, Eq. (3.2.10)], if A is a power function, then condition $\sqrt{k_n}A(n/k_n) \to \beta$ as $n \to \infty$ yields $k_n = O\left(n^{-2\rho/(1-2\rho)}\right)$. As a conclusion, up to logarithmic factors, possible choices of sequences are then

$$h_n = n^{-1/(d+\kappa(d))} \text{ and } k_n = n^{1/(1+\max(d/\kappa(d),-1/(2\rho)))}. \tag{16}$$

If $\rho \geq -\kappa(d)/(2d)$, the rate of convergence of $\hat{\gamma}_n$ is thus $n^{\rho/(1-2\rho)}$ up to logarithmic factors which is the classical rate for estimators of the tail-index, see for instance [25, Remark 3]. For instance, in the situation where the dimension of the covariate is $d \leq 2$, then the $n^{\rho/(1-2\rho)}$ rate is reached as soon as $\rho \geq -1$. This corresponds to the challenging situation where a high bias is expected in the estimation which may occur for most usual distributions, depending on their shape parameters, see Table 1.

Theorem 4 states the asymptotic normality of the estimator (12) of extreme conditional quantiles of $Y|x$.

Theorem 4. *Assume* **(A.1)**-**(A.4)** *hold and* $f_Z(q_Z(\mu_j)) > 0$ *for* $j \in \{1,2,3\}$. *Let* (k_n) *be an intermediate sequence of integers. Suppose* $nh^d/(k_n \log n) \to \infty$, $nh^{d+\kappa(d)} \to 0$ *and* $\sqrt{k_n}A(n/k_n) \to \beta \in \mathbb{R}$ *as* $n \to \infty$. *Then, for all sequences* $(t_n) \subset \tilde{\Pi}^{(n)}$ *and* $(\alpha_n) \subset (0,1)$ *such that* $n\alpha_n/k_n \to 0$ *and* $\log(n\alpha_n)/\sqrt{k_n} \to 0$ *as* $n \to \infty$,

$$\frac{\sqrt{k_n}}{q_Z(\alpha_n)\log\left(\frac{k_n}{n\alpha_n}\right)}\left(\frac{\tilde{q}_{n,Y}(\alpha_n \mid t_n) - q_Y(\alpha_n \mid t_n)}{b(t_n)}\right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2). \tag{17}$$

Remark that $b(t_n)q_Z(\alpha_n) \sim a(t_n) + b(t_n)q_Z(\alpha_n) = q_Y(\alpha_n \mid t_n)$ and therefore (17) can be rewritten as

$$\frac{\sqrt{k_n}}{\log\left(\frac{k_n}{n\alpha_n}\right)}\left(\frac{\tilde{q}_{n,Y}(\alpha_n \mid t_n)}{q_Y(\alpha_n \mid t_n)} - 1\right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2).$$

As a comparison, the rate of convergence of purely nonparametric methods involves an extra $h^{d/2}$ factor, see for instance [18, Theorem 3] or [8, Theorem 3]. The location-dispersion regression model allows to dampen this vexing effect of the dimensionality.

Finally, a uniform consistency result is also available :

Theorem 5. *Assume* **(A.1)**-**(A.4)** *hold. Let* (k_n) *be an intermediate sequence of integers. Suppose* $nh^d/(k_n \log n) \to \infty$, $nh^{d+\kappa(d)}/\log n \to 0$ *and* $\sqrt{k_n}A(n/k_n) \to \beta \in \mathbb{R}$ *as* $n \to \infty$. *Then, for all sequence* $(\alpha_n) \subset (0,1)$ *such that* $n\alpha_n/k_n \to 0$ *and* $\log(n\alpha_n)/\sqrt{k_n} \to 0$ *as* $n \to \infty$,

$$\frac{\sqrt{k_n}}{q_Z(\alpha_n)\log\left(\frac{k_n}{n\alpha_n}\right)} \max_{i \in I_n} \left| \frac{\tilde{q}_{n,Y}(\alpha_n \mid x_i) - q_Y(\alpha_n \mid x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1).$$

5 Illustration on simulations

5.1 Experimental design

We propose to illustrate the finite-sample performance of the estimators of the conditional tail-index and the extreme conditional quantiles on simulated data from the location-dispersion regression model. For that purpose, set $d = 2$, $\Pi = [0,1]^2$ and define the regression and dispersion functions respectively by $a(x) = 1 - \cos\left(\pi\left(x^{(1)} + x^{(2)}\right)\right)$ and $b(x) = \exp\left(-\left(x^{(1)} - 0.5\right)^2 - \left(x^{(2)} - 0.5\right)^2\right)$, for $x = \left(x^{(1)}, x^{(2)}\right) \in \Pi$. Let $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$. Two distributions are considered for the heavy-tailed random variable Z :

- Let Z_0 be a standard Student-t_ν random variable where $\nu \in \{1,2,4\}$ denotes the degrees of freedom (df) and introduce $Z = Z_0/(2q_{Z_0}(\mu_3))$ the associated rescaled Student random variable. Symmetry arguments yield $q_Z(\mu_2) = 0$, $q_Z(\mu_1) = -q_Z(\mu_3)$ and $q_Z(\mu_3) = q_{Z_0}(\mu_3)/(2q_{Z_0}(\mu_3)) = 1/2$ by construction. Therefore (4) holds. This choice also ensures that Z is heavy-tailed with conditional tail-index $\gamma = 1/\nu$ and that the second-order condition **(A.4)** holds with $\rho = -2/\nu$, see Table 1.
- Let Z_0 be a Burr random variable with parameters $\alpha \in \{1,2,4\}$ and $\beta = 1$. We then introduce the translated and rescaled random variable

$$Z = \frac{Z_0 - \left(\mu_2^{-1} - 1\right)^{1/\alpha}}{\left(\mu_3^{-1} - 1\right)^{1/\alpha} - \left(\mu_1^{-1} - 1\right)^{1/\alpha}},$$

  such as (4) holds. The second-order condition **(A.4)** is also fulfilled with $\gamma = 1/\alpha$ and $\rho = -1$, see Table 1.

The design points x_i, $i = 1, \ldots, n$ are chosen on a regular grid on the unit square Π. The kernel function K is the product of two quartic (or biweight) kernels :

$$K(u,v) = \left(\frac{15}{16}\right)^2 \left(1-u^2\right)^2 \left(1-v^2\right)^2 \mathbb{1}_{\{|u|\le 1\}} \mathbb{1}_{\{|v|\le 1\}},$$

where $(u,v) \in \mathbb{R}^2$. We set $\|x\| = \max\left(|x^{(1)}|, |x^{(2)}|\right)$ so that $\tilde{\Pi}^{(n)} = [h, 1-h]^2$. The bandwidth is fixed to $h_n^* = \sigma n^{-1/6}$ following [4] and in accordance with (16), where $\sigma = 12^{-1/2}$ is the standard deviation of the coordinates of the design points. This choice is optimal for density estimation in the Gaussian case, but is also known to provide good results in other settings.

5.2 Graphical illustrations

In all the experiments, $N = 100$ replications of a dataset of size $n = 10,000$ are considered. The estimation results for the regression and dispersion functions are depicted respectively on Figure 1 and Figure 2 in the situation where Z is Student-t_ν distributed for $\nu \in \{1, 2, 4\}$. The results are visually satisfying and seem independent from the degrees of freedom. This conclusion was expected since both estimators of $a(\cdot)$ and $b(\cdot)$ are based on non-extreme quantiles, they are thus robust with respect to heavy tails.

As already noticed in Section 2, in the context of the proportional tails model, both random variables Y and Z share the same conditional tail-index γ. This parameter can thus be estimated either by (11) (computed on the residuals $\hat{Z}_i$) or by the classical Hill estimator (computed on the response variables Y_i). The associated estimation results are displayed on Figure 3 as functions of the sample fraction k_n. It first appears that working on the residuals provides much better results in terms of bias than working on the initial response variable. Second, the tail-index estimator (11) has a stronger bias for larger values of ν. These empirical results are in line with the properties of the Student distribution. Indeed, the second-order parameter $\rho = -2/\nu$ being increasing with ν, the bias of the Hill-type estimator increases as well.

In practice, the estimation of the conditional tail-index and extreme conditional quantiles require the selection of the sample fraction k_n. This parameter is selected using a mean-squared error criterion. Assuming that $A(t) = ct^\rho$, the optimal value of k_n is given by

$$k_n^* = \left(\frac{\gamma^2(1-\rho)^2}{-2\rho c^2}\right)^{\frac{1}{1-2\rho}} n^{-\frac{2\rho}{1-2\rho}},$$

see [24, Section 3.2]. Since ρ may be difficult to estimate in practice, a miss-specified value $\rho = -1$ is considered in several works dealing with bias reduction of tail-index estimators, see for instance [14] or [23]. Letting moreover $c = \sqrt{2}$ and restricting ourselves to integer values, we end up with $k_n^* = \lfloor (\check{\gamma} n)^{2/3} \rfloor$ where $\check{\gamma}$ is a prior naive estimation of γ computed with $k_n = \lfloor n^{1/2} \rfloor$ and where $\lfloor \cdot \rfloor$ denotes the floor function. Such a choice of k_n^* fulfils the assumptions of Theorem 3–5 for all three considered Burr distributions and for Student-t_ν distributions with $\nu \in \{1, 2\}$. The constraints are violated in case of the Student-t_4 distribution in order to examine the robustness of the method with respect to the choice of the pair (h, k_n) which may be challenging in practice. The estimated conditional quantiles $q_Y(1/n \,|\, \cdot)$ of extreme level $\alpha_n = 1/n$ are displayed on Figure 4. As expected, the estimated extreme conditional quantiles all share the same shape despite different variation ranges.

5.3 Quantitative assessment

In this section, we propose to highlight the performances of the extreme conditional quantile estimator (12) thanks to a comparison with a purely nonparametric one. The nonparametric estimator is based on the ideas of the moving window approach introduced in [16]. For each $x \in \tilde{\Pi}^{(n)}$, a subsample $\{(Y_i^{\circledast}, x_i^{\circledast})\}_{i=1,\dots,n^{\circledast}} = \{(Y_i, x_i),\ 1 \le i \le n,\ \text{s.t. } \|x - x_i\| < h\}$ of size $n^{\circledast} = n^{\circledast}(x, h)$ is extracted from the initial sample. Letting $k_n^{\circledast} = \lfloor \sqrt{n^{\circledast}} \rfloor$, the conditional tail-index is estimated by the (local) Hill-type statistic

$$\hat{\gamma}_n^{\circledast}(x) = \frac{1}{k_n^{\circledast}} \sum_{i=0}^{k_n^{\circledast}-1} \log Y^{\circledast}_{n^{\circledast}-i, n^{\circledast}} - \log Y^{\circledast}_{n^{\circledast}-k_n^{\circledast}, n^{\circledast}},$$

and the extreme conditional quantile $q_Y(\alpha_n \,|\, x)$ is estimated by the associated Weissman-type statistic :

$$\hat{q}^{\circledast}_{n,Y}(\alpha_n \,|\, x) = Y^{\circledast}_{n^{\circledast}-k_n^{\circledast}, n^{\circledast}} \left(\frac{\alpha_n n^{\circledast}}{k_n^{\circledast}} \right)^{-\hat{\gamma}_n^{\circledast}(x)}.$$

Another option is to re-estimate γ and $q_Y(\alpha_n \,|\, x)$ by taking $k_n^{\oplus} = \lfloor (\hat{\gamma}_n^{\circledast}(x) n^{\circledast})^{2/3} \rfloor$ in the above two estimators. The associated estimator of the extreme quantile is denoted by $\hat{q}^{\oplus}_{n,Y}(\alpha_n \,|\, x)$. The comparison between the true and estimated extreme conditional quantiles is based on a relative median-squared error (RMSE) computed on the $N = 100$ replications and the m_n design points in the square $\tilde{\Pi}^{(n)}$:

$$\text{median} \left\{ \text{median} \left\{ \left(\frac{\hat{q}^{[r]}_{n,Y}(\alpha_n \,|\, x_i)}{q_Y(\alpha_n \,|\, x_i)} - 1 \right)^2,\ x_i \in \tilde{\Pi}^{(n)} \right\},\ r \in \{1, \dots, N\} \right\},$$

where $\hat{q}^{[r]}_{n,Y}(\alpha_n \,|\, \cdot)$ denotes either $\tilde{q}_{n,Y}(\alpha_n \,|\, \cdot)$, $\hat{q}^{\circledast}_{n,Y}(\alpha_n \,|\, \cdot)$ or $\hat{q}^{\oplus}_{n,Y}(\alpha_n \,|\, \cdot)$ computed on the rth replication. Here, both Student-t_ν and Burr distributions are considered with $\nu \in \{1, 2, 4\}$, $\gamma \in \{1, 1/2, 1/4\}$, $\alpha_n = 1/n$ and $n \in \{20^2, 40^2, 60^2, 80^2, 100^2\}$. The RMSE are reported in Table 2. For both estimators, it appears that the main driver of the relative error is the tail heaviness. The nonparametric estimator even seems not to converge on the Burr distribution

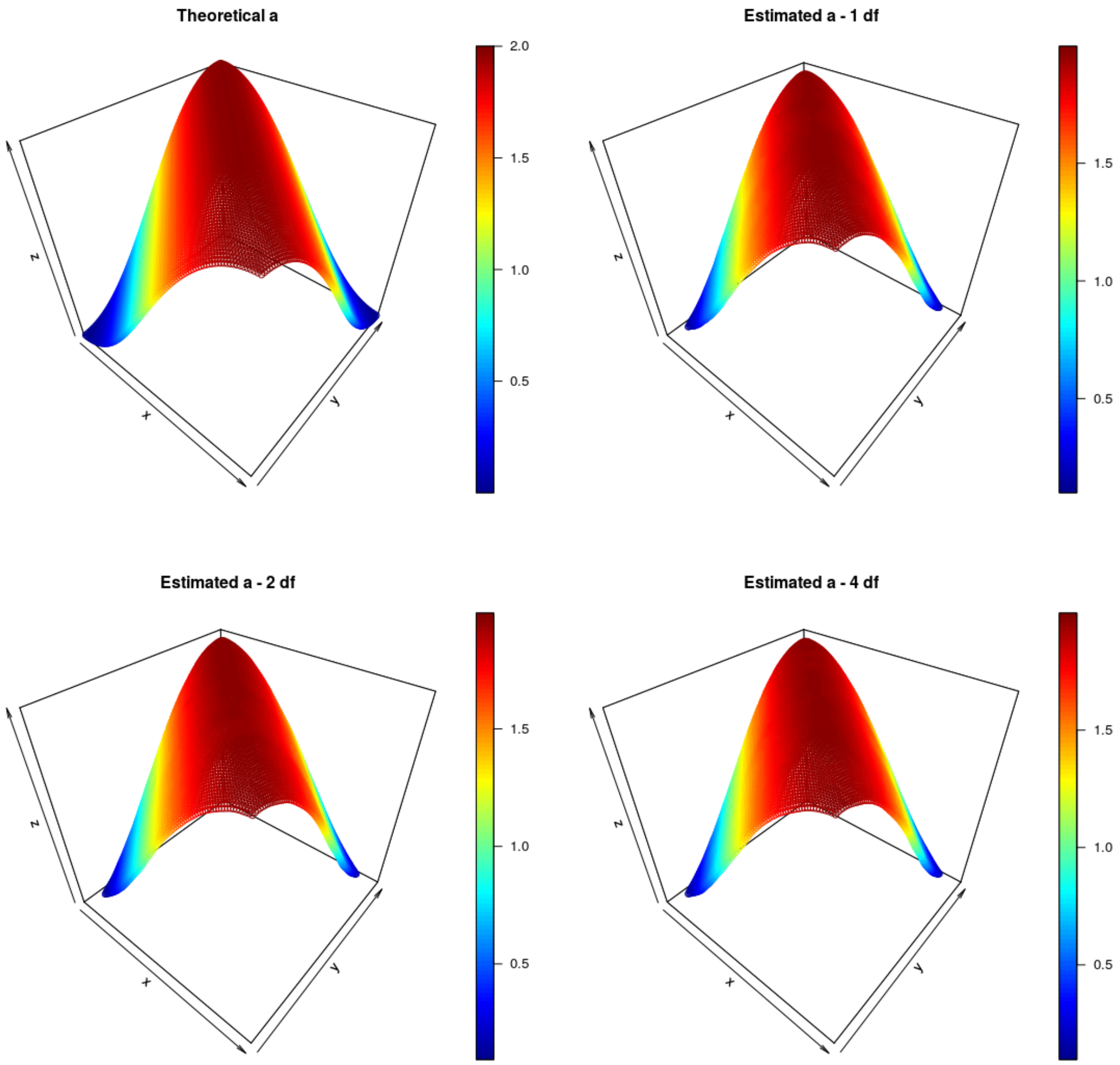

Figure 1: Simulation results obtained on a Student-t_ν distribution. From top to bottom, left to right : Theoretical function $a(\cdot)$, and means over $N = 100$ replications of estimates $\hat{a}_n(\cdot)$ computed on $n = 10,000$ observations for $\nu \in \{1, 2, 4\}$. X-axis and y-axis range between 0 and 1, the scale of the z-axis is the same for theoretical and estimated regression function.

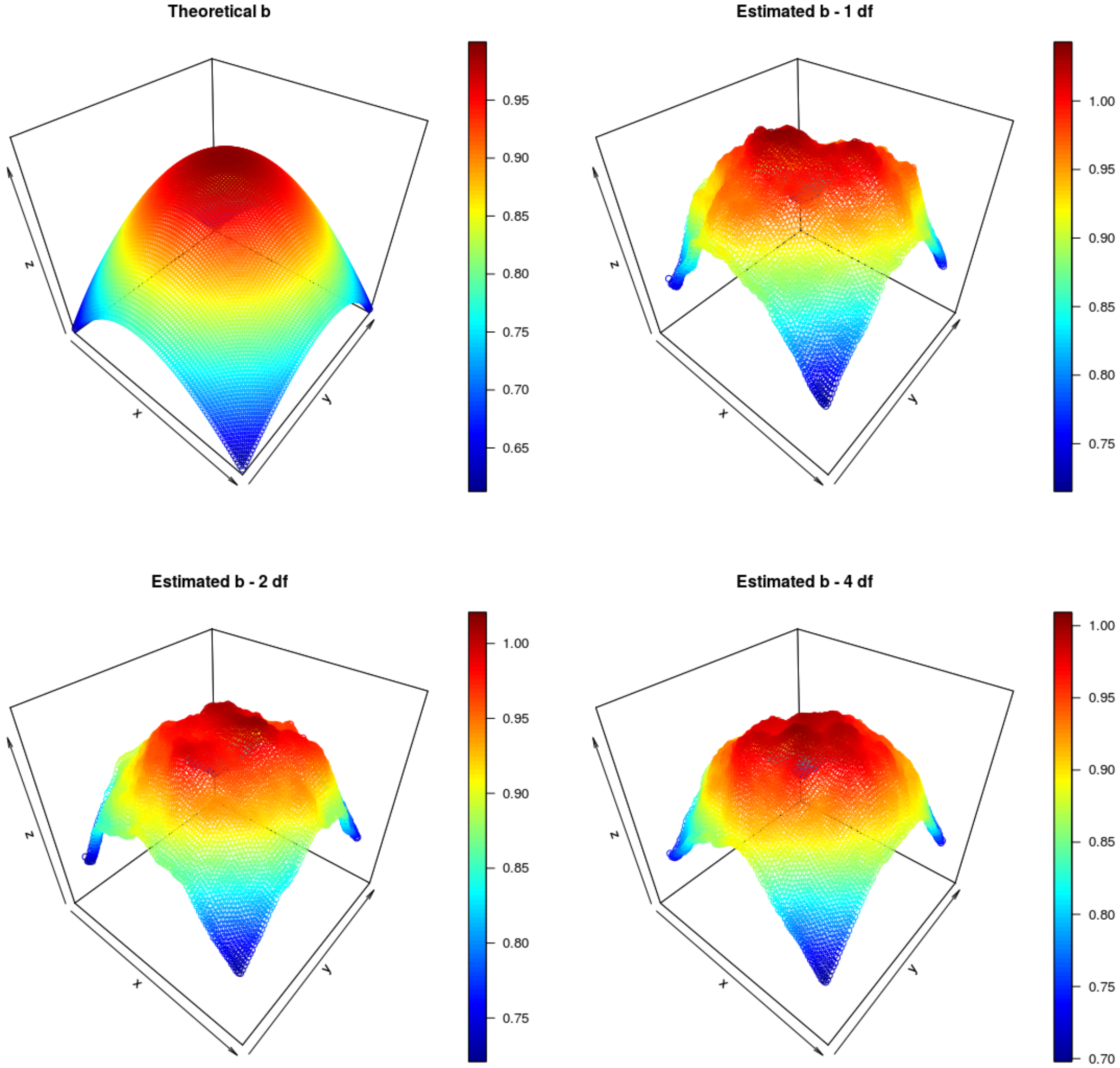

Figure 2: Simulation results obtained on a Student-t_ν distribution. From top to bottom, left to right : Theoretical function $b(\cdot)$, and means over $N = 100$ replications of estimates $\hat{b}_n(\cdot)$ computed on $n = 10,000$ observations for $\nu \in \{1, 2, 4\}$. All three coordinates range between 0 and 1.

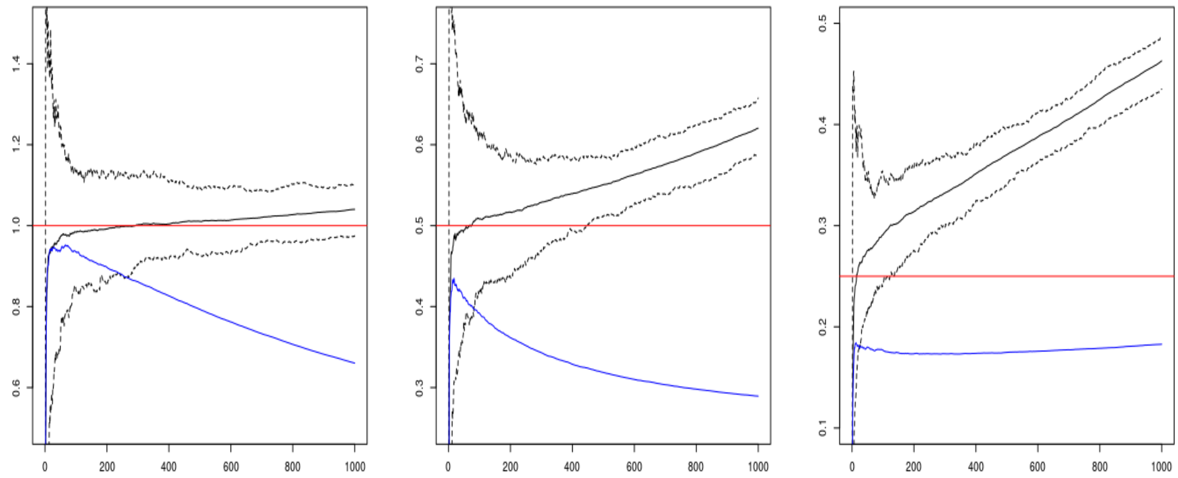

Figure 3: Simulation results obtained on a Student-t_ν distribution for $\nu = 1$ (left), $\nu = 2$ (middle) and $\nu = 4$ (right). Mean estimate of the conditional tail-index (11) (continuous black line), associated 95% empirical confidence intervals (dotted lines) and mean Hill estimate computed on the response variable (continuous blue line), as functions of the sample fraction k_n. The true value $\gamma = 1/\nu$ is depicted by a red horizontal line.

with tail-index $\gamma = 1$. Unsuprisingly, the semi-parametric estimator $\tilde{q}_{n,Y}$ provides much better results than the nonparametric ones $\hat{q}^{\circledast}_{n,Y}$ and $\hat{q}^{\oplus}_{n,Y}$: Its RMSE is smaller and converges towards 0 at a faster rate when the sample size n increases.

6 Tsunami data example

The proposed illustration is based on the "Tsunami Causes and Waves" dataset, available at https://www.kaggle.com/noaa/seismic-waves. The data include the maximum wave height recorded at several stations in the world where a tsunami occured. We focus on the 2011 Tohoku tsunami, in Japan. This earthquake was the cause of the Fukushima Daiichi nuclear disaster. Indeed, a wave height greater than 15 meters (around 50 feet) flooded the nuclear plant, protected by a seawall of only 5.7 meters (19 feet). In this context, the estimation of return levels of wave heights associated with small probability is a crucial issue. Figure 5 (top-left panel) displays the maximum wave heights $Y_1, \dots, Y_n$ (in meters) recorded the 03/11/2011 at $n = 5,364$ stations with respective latitudes $x_1^{(1)}, \dots, x_n^{(1)}$ and longitudes $x_1^{(2)}, \dots, x_n^{(2)}$. Note that the values of Y are ranging from 0 to 55.88 meters (blue to red points). We propose to estimate an extreme quantile of the wave height at each station, following the methodology introduced in Section 3. The assumption of a constant

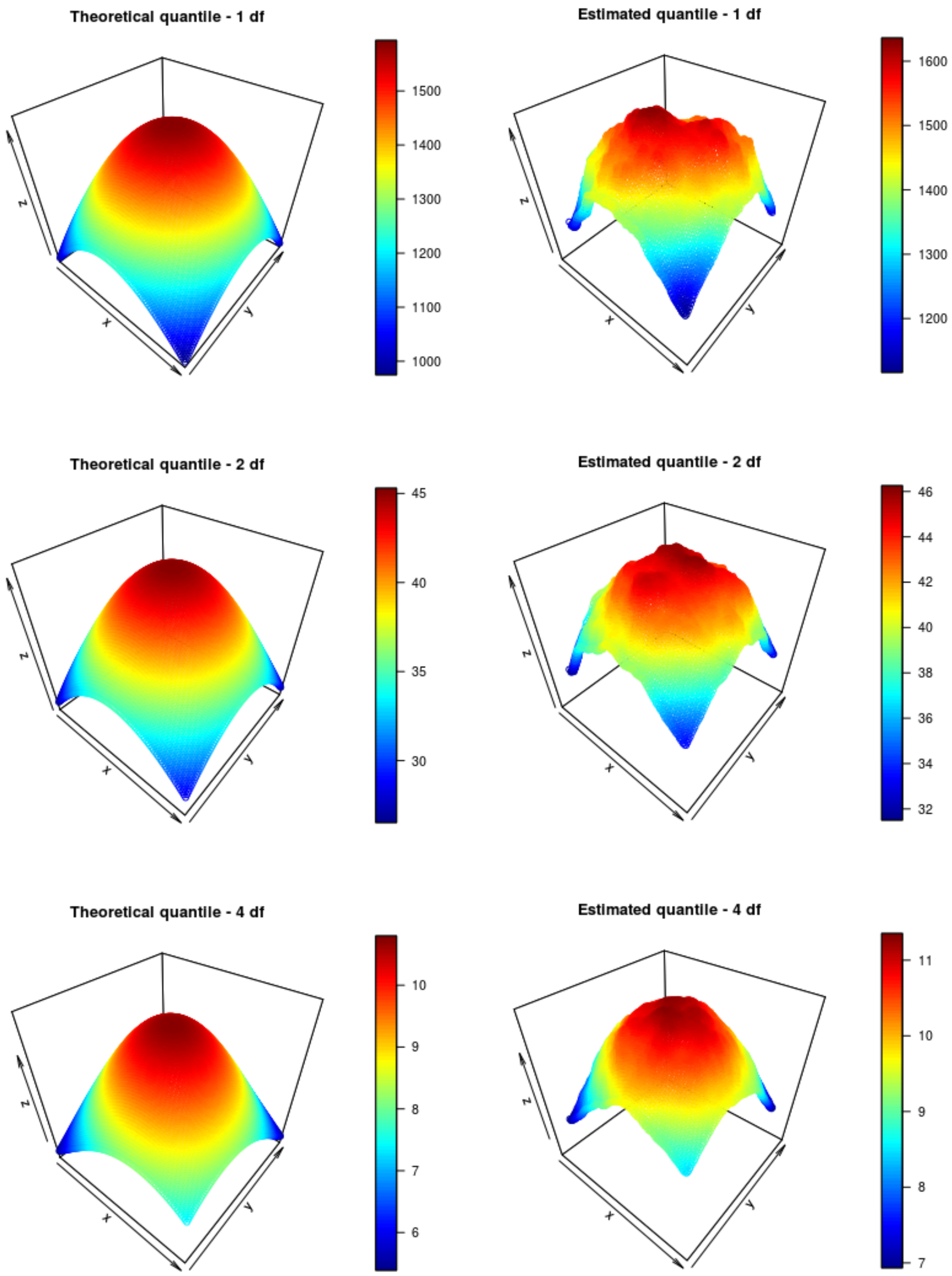

Figure 4: Simulation results obtained on a Student-t_ν distribution for $\nu = 1$ (top), $\nu = 2$ (middle) and $\nu = 4$ (bottom). Left panels : Theoretical quantiles $q_Y(1/n \,|\, \cdot)$. Right panels : Means over $N = 100$ replications of estimates $\tilde{q}_{n,Y}(1/n \,|\, .)$ computed on $n = 10,000$ observations. X-axis and y-axis range between 0 and 1, the scale of the z-axis is the same for theoretical and estimated quantiles.

n	Student, $\nu = 1$	Student, $\nu = 2$	Student, $\nu = 4$
400	0.547 (0.890, 0.976)	0.129 (0.643, 0.630)	0.062 (0.442, 0.458)
1,600	0.138 (0.867, 0.893)	0.065 (0.533, 0.458)	0.020 (0.284, 0.352)
3,600	0.145 (0.855, 0.837)	0.048 (0.477, 0.431)	0.012 (0.226, 0.306)
6,400	0.061 (0.845, 0.776)	0.032 (0.456, 0.454)	0.011 (0.206, 0.253)
10,000	0.045 (0.820, 0.723)	0.026 (0.425, 0.435)	0.013 (0.184, 0.222)
n	Burr, $\alpha = 1$, $\beta = 1$	Burr, $\alpha = 2$, $\beta = 1$	Burr, $\alpha = 4$, $\beta = 1$
400	0.525 (0.746, 0.588)	0.197 (0.329, 0.285)	0.104 (0.129, 0.176)
1,600	0.182 (0.796, 0.637)	0.068 (0.348, 0.260)	0.038 (0.124, 0.168)
3,600	0.157 (0.825, 0.625)	0.056 (0.333, 0.264)	0.023 (0.118, 0.149)
6,400	0.096 (0.827, 0.591)	0.054 (0.311, 0.271)	0.020 (0.107, 0.122)
10,000	0.070 (0.845, 0.563)	0.030 (0.301, 0.262)	0.023 (0.102, 0.107)

Table 2: Relative median squared errors associated with the estimation of the extreme conditional quantile $q_Y(1/n \,|\, \cdot)$. Results obtained with the semi-parametric estimator $\tilde{q}_{n,Y}$ and comparison with the purely nonparametric ones $(\hat{q}^{\circledast}_{n,Y}, \hat{q}^{\oplus}_{n,Y})$.

conditional tail-index can be checked thanks to the test statistic $T_{4,n}$ introduced in [12] :

$$T_{4,n} = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{\hat{\gamma}_{p_i}}{\hat{\gamma}_H} - 1 \right)^2 .$$

The idea is to compare the Hill estimate $\hat{\gamma}_H$ computed on the response variables with partial ones $\hat{\gamma}_{p_i}$ computed on non-overlapping blocks indexed by $i = 1, \dots, m$. Under the hypothesis that the conditional tail-index is constant (and additional technical assumptions), it is then shown that $k_n T_{4,n} \xrightarrow{d} \chi^2_{m-1}$, see [12] for details. Following the ideas of Paragraph 5.3, we set $k_n = k_n^{\oplus} = 72$ and we choose $m = 4$ blocks as in [12], leading to $T_{4,n} \approx 2.14$ and a $p-$value around 0.54. The hypothesis of a constant conditional tail-index cannot be rejected, and our semi-parametric approach can thus be applied on these data.

To this end, a bandwidth has to be selected. Noticing that the standard deviations of $x^{(1)}$ and $x^{(2)}$ are respectively 1.63 and 1.16, we fixed $h_n^* = 1.63 \times n^{-1/6} \simeq 0.4$. We also set $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$, these choices having no consequence in practice. The regression and dispersion functions are then estimated via (9) and depicted on the bi-dimensional map (Figure 5, top-right and bottom-left panels) and along the one-dimensional first principal axis (Figure 6, top panels). Note that the principal axis has been obtained by computing the eigenvector associated with the largest eigenvalue of the covariance matrix of the coordinates $(x_i^{(1)}, x_i^{(2)})$, $i = 1, \dots, n$. It appears that $\hat{a}_n(\cdot)$ and $\hat{b}_n(\cdot)$ have a similar shape with a peak in the neighbourhood of the epicenter, indicating a strong heteroscedasticity of the observed phenomenon.

The residuals $\hat{Z}_1, \dots, \hat{Z}_n$ are then computed from (10). In a first time, it is necessary to check whether the residuals have a heavy-tailed behavior. The common practice is to use a graphical diagnosis. Here, a quantile-quantile plot is adopted, see the bottom-right panel of Figure 6. The log-excesses $\log(\hat{Z}_{n-i+1,n}/\hat{Z}_{n-k_n^*+1,n})$ are plotted versus the quantiles $\log(k_n^*/i)$ of the standard exponential distribution, $i = 1, \ldots, k_n^*$. Note that the number of upper order statistics $k_n^* = 82$ is chosen following the approach described in Paragraph 5.2. It appears that the resulting set of points is close to the line of slope $\hat{\gamma}_n$ (computed with $k_n^* = 82$), which confirms that the heavy-tailed assumption is reasonable in this case. The proposed estimator (11) computed on the residuals as well as the Hill estimator computed on the output variables are both depicted as functions of k_n on the bottom-left panel of Figure 6. The first one features a nice stable behaviour, confirming the heavy-tail assumption, and pointing towards a tail-index close to 0.25. As a comparison, the Hill estimator computed on the original output variables is less stable and yields smaller results, in accordance with the negative bias observed on simulated data (Section 5). Finally, the extreme conditional quantile estimator (12) is evaluated at each station with the level $\alpha_n = 10/n$. The results are reported in the bottom-right panel of Figure 5. The estimated quantiles of the maximum wave height are ranging from 0 to 60.53 meters, with largest values close to the epicenter. Note that such a quantile level means that the observed values $Y_1, \ldots, Y_n$ should exceed the return levels $\tilde{q}_{n,Y}(\alpha_n \mid x_1), \ldots, \tilde{q}_{n,Y}(\alpha_n \mid x_n)$ approximately 10 times in the sample. In this particular example, there are 15 waves exceeding the return levels, this empirical result does not deviate too much from the expected number of exceedances.

7 Appendix : Proofs

Technical lemmas are collected in Paragraph 7.1 while preliminary results of general interest are provided in Paragraph 7.2. Finally, the proofs of the main results are given in Paragraph 7.3.

7.1 Auxiliary lemmas

The first result is an adaptation of Bochner's lemma (for twice differentiable functions) to the multidimensional fixed design setting.

Lemma 1. *Let $\psi(\cdot \mid \cdot) : \mathbb{R}^p \times \Pi \to \mathbb{R}^+$ be a positive, twice differentiable (with respect to its second argument) function. Let us denote by $H_2[\psi](\cdot,\cdot)$ the Hessian matrix of $\psi(\cdot \mid \cdot)$ with respect to its second argument, and assume that $H_2[\psi](\cdot,\cdot)$ is continuous on $\mathbb{R}^p \times \Pi$. Let C be a compact subset of $\mathbb{R}^p$. For all sequences $(t_n) \subset \tilde{\Pi}^{(n)}$ and $(y_n) \subset C$, define*

$$\psi_n(y_n \mid t_n) := \sum_{i=1}^{n} \psi(y_n \mid x_i) \int_{\Pi_i} Q_h(t_n - s)ds,$$

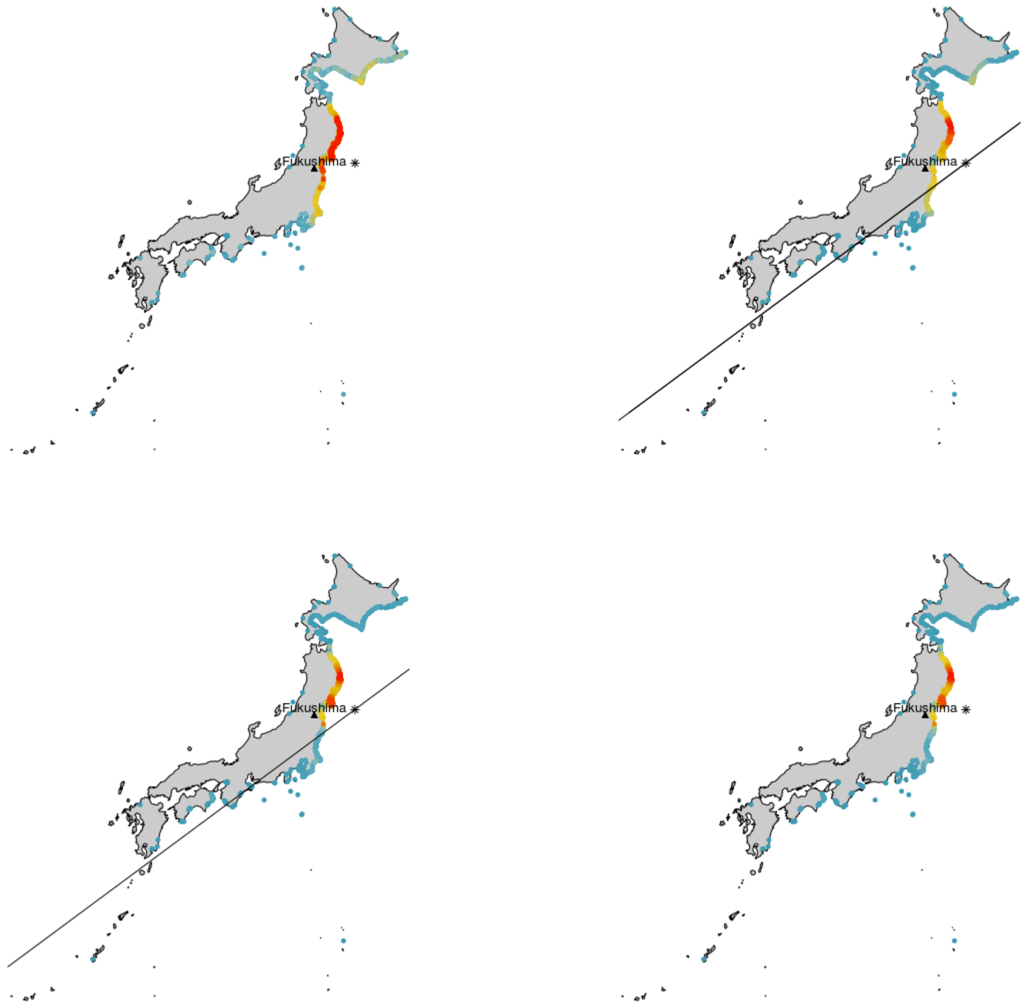

Figure 5: Results on tsunami data. Top-left : Maximum wave height recorded at each station. Top-right : Regression function estimate $\hat{a}_n(\cdot)$ at each station. Bottom-left : Dispersion function estimate $\hat{b}_n(\cdot)$ at each station. Bottom-right : Quantile estimate $\tilde{q}_{n,Y}(10/n \mid \cdot)$ at each station. On all the maps, smallest and largest values are respectively depicted in blue and red. The straight line is the principal axis $x^{(2)} = 1.64x^{(1)} + 80.35$ computed on the coordinates of the stations, and $*$ represents the epicenter of the earthquake.

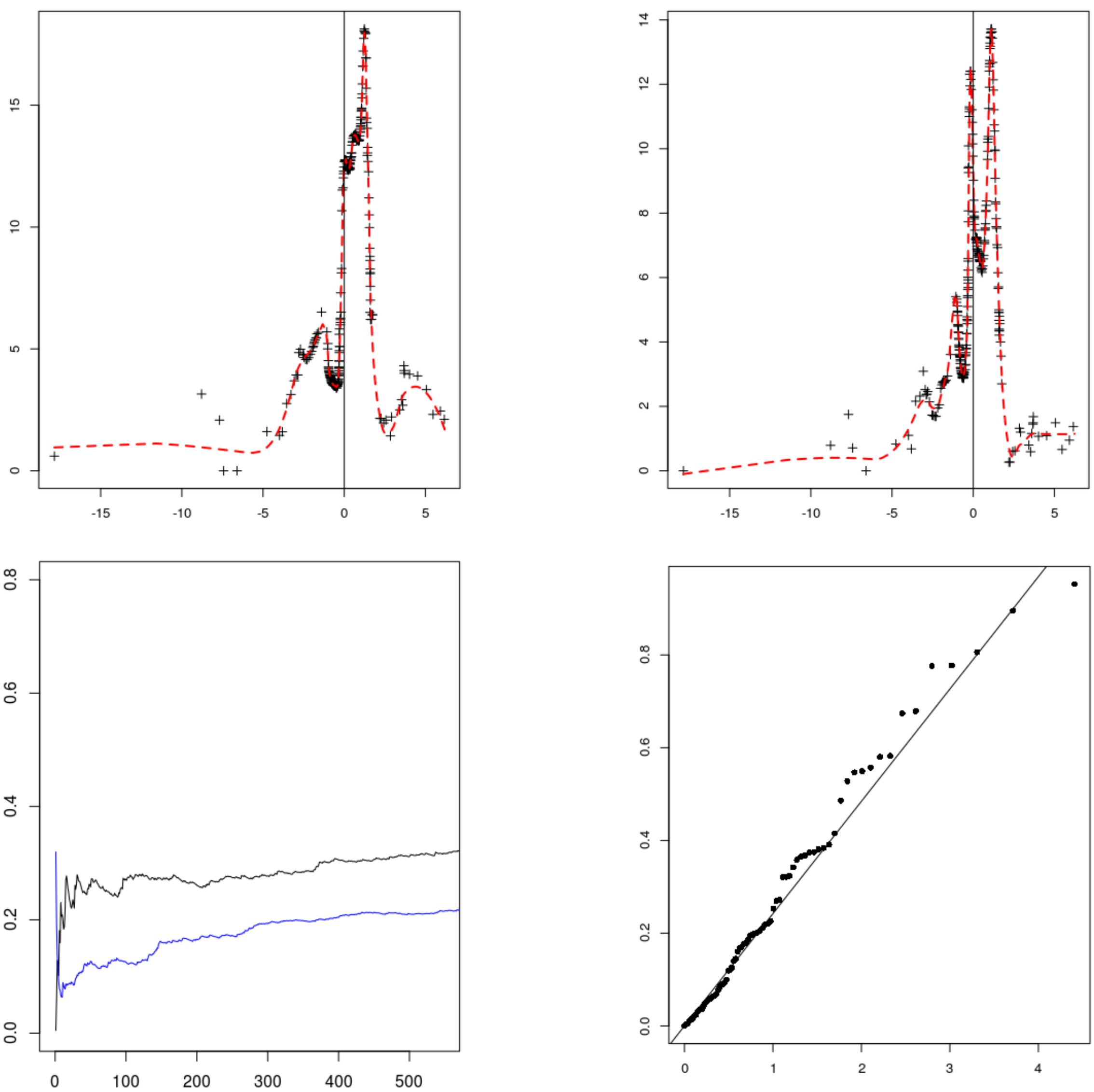

Figure 6: Results on tsunami data. Top : Regression (left) and dispersion (right) function estimates $\hat{a}_n(\cdot)$ and $\hat{b}_n(\cdot)$ along the principal axis $x^{(2)} = 1.64x^{(1)} + 80.35$. The estimates at each station (black +) are smoothed (red dashed line) for the visualization sake. The vertical black line displays the projection of the epicenter on the principal axis. Bottom left : Hill estimator (11) computed on the residuals (black line) and on the original output variables (blue line) as a function of k_n. Bottom right : Log-excesses $\log(\hat{Z}_{n-i+1,n}/\hat{Z}_{n-k_n^*+1,n})$ of the residuals versus $\log(k_n^*/i)$, $1 \le i \le k_n^* = 82$. The straight line has slope $\hat{\gamma}_n \simeq 0.25$.

where $x_i \in \Pi_i$ such that (14) and (15) hold, and $Q_h(\cdot) = Q(\cdot/h)/h^d$, where Q is an even measurable positive function with symmetric support $S \subset B(0,1)$. Then, as $n \to \infty$,

$$\psi_n(y_n \mid t_n) = \|Q\|_1 \psi(y_n \mid t_n) + O\left(n^{-1/d}\right) + O(h^2),$$

where $\|Q\|_1 = \int_S Q(u)du$.

Proof. Consider the expansion

$$\begin{aligned}
\psi_n(y_n \mid t_n) - \|Q\|_1\psi(y_n \mid t_n) &= \sum_{i=1}^{n} \psi(y_n \mid x_i) \int_{\Pi_i} Q_h(t_n - s)ds - \|Q\|_1 \psi(y_n \mid t_n) \\
&= \int_{\Pi} \psi(y_n \mid s) Q_h(t_n - s)ds - \|Q\|_1 \psi(y_n \mid t_n) \\
&+ \sum_{i=1}^{n} \psi(y_n \mid x_i) \int_{\Pi_i} Q_h(t_n - s)ds - \int_{\Pi} \psi(y_n \mid s) Q_h(t_n - s)ds \\
&=: T_{n,1} + T_{n,2}.
\end{aligned}$$

and let us first focus on $T_{n,1}$. The change of variable $u = (t_n - s)/h$ yields

$$T_{n,1} = \int_{(t_n - \Pi)/h} \psi(y_n \mid t_n - uh) Q(u) du - \|Q\|_1 \psi(y_n \mid t_n).$$

Let us remark that $x \in B(0,1)$ implies $t_n - xh \in B(t_n, h) \subset \Pi$ since $t_n \in \tilde{\Pi}^{(n)}$ and by definition of the erosion. As a consequence, $S \subset B(0,1) \subset (t_n - \Pi)/h$ and therefore

$$T_{n,1} = \int_S [\psi(y_n \mid t_n - uh) - \psi(y_n, t_n)] Q(u) du.$$

Let $\nabla_2[\psi](\cdot,\cdot)$ denote the gradient of $\psi(\cdot \mid \cdot)$ with respect to its second argument and let $\langle \cdot, \cdot \rangle$ be the usual dot product on $\mathbb{R}^d$. A second order Taylor expansion yields, for all $y_n \in C$,

$$\psi(y_n \mid t_n - uh) - \psi(y_n \mid t_n) = h \langle \nabla_2[\psi](y_n, t_n), u \rangle + O(h^2),$$

since $H_2[\psi](\cdot,\cdot)$ is bounded on compact sets. Remarking that $\int_S uQ(u)du = 0$ shows that

$$T_{n,1} = O(h^2). \tag{18}$$

Let us now turn to the second term

$$T_{n,2} = \sum_{i=1}^{n} \int_{\Pi_i} [\psi(y_n \mid x_i) - \psi(y_n \mid s)] Q_h(t_n - s) ds.$$

Since $\psi(\cdot \mid \cdot)$ is continuously differentiable with respect to its second argument, there exists

$c_\psi > 0$ such that

$$|T_{n,2}| \leq \sum_{i=1}^{n} \int_{\Pi_i} |\psi(y_n \mid x_i) - \psi(y_n \mid s)| \, Q_h(t_n - s) ds \leq c_\psi \sum_{i=1}^{n} \int_{\Pi_i} \|x_i - s\| Q_h(t_n - s) ds.$$

Moreover, under assumption (15),

$$|T_{n,2}| = \sum_{i=1}^{n} \int_{\Pi_i} Q_h(t_n - s) ds \, O\left(n^{-1/d}\right) = \int_{\Pi} Q_h(t_n - s) ds \, O\left(n^{-1/d}\right) = O\left(n^{-1/d}\right). \quad (19)$$

Finally, collecting (18) and (19), the conclusion follows. □

As a consequence of Lemma 1, the asymptotic bias and variance of the estimator (7) of the conditional survival function can be derived.

Lemma 2. *Suppose* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset \tilde{\Pi}^{(n)}$ *and* $(y_n) \subset C$ *be two nonrandom sequences with* C *a compact subset of* $\mathbb{R}$.

(i) Then,

$$\mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right) = \bar{F}_Y(y_n \mid t_n) + O\left(n^{-1/d}\right) + O(h^2).$$

(ii) If, moreover, $nh^d \to \infty$ *as* $n \to \infty$ *and* $\liminf F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n) > 0$, *then*

$$\operatorname{var}\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right) \sim \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n),$$

where F_Y *is the conditional cumulative distribution function associated with* $\bar{F}_Y$.

Proof. (i) Clearly,

$$\mathbb{E}\left[\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right] = \sum_{i=1}^{n} \bar{F}_Y(y_n \mid x_i) \int_{\Pi_i} K_h(t_n - s) ds,$$

and the conclusion follows from Lemma 1 applied with $p = 1$.
(ii) As a consequence of the independence assumption,

$$\operatorname{var}\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right) = \sum_{i=1}^{n} \bar{F}_Y(y_n \mid x_i) S_{n,i} - \sum_{i=1}^{n} \bar{F}_Y^2(y_n \mid x_i) S_{n,i} =: T_{n,1} - T_{n,2},$$

where

$$S_{n,i} := \left(\int_{\Pi_i} K_h(t_n - s) ds\right)^2 = \frac{1}{h^{2d}} \int_{\Pi_i} \int_{\Pi_i} K\left(\frac{t_n - s_1}{h}\right) K\left(\frac{t_n - s_2}{h}\right) ds_1 ds_2. \quad (20)$$

Let us write

$$K\left(\frac{t_n - s_2}{h}\right) = K\left(\frac{t_n - s_1}{h}\right) + \left[K\left(\frac{t_n - s_2}{h}\right) - K\left(\frac{t_n - s_1}{h}\right)\right],$$

with, under **(A.3)** and (15),

$$\left|K\left(\frac{t_n - s_2}{h}\right) - K\left(\frac{t_n - s_1}{h}\right)\right| \le \frac{c_K\|s_2 - s_1\|}{h} = O\left(\frac{1}{n^{1/d}h}\right),$$

uniformly on $(s_1, s_2) \in \Pi_i^2$ and $i = 1, \ldots, n$. It thus follows from (14) that

$$\begin{aligned} S_{n,i} &= \frac{1}{h^{2d}} \int_{\Pi_i}\int_{\Pi_i} \left[K^2\left(\frac{t_n - s_1}{h}\right) + K\left(\frac{t_n - s_1}{h}\right) O\left(\frac{1}{n^{1/d}h}\right)\right] ds_1 ds_2 \\ &= \frac{\lambda(\Pi)}{nh^{2d}} \int_{\Pi_i} K^2\left(\frac{t_n - s}{h}\right) ds\ (1 + o(1)) + O\left(\frac{1}{n^{1+1/d}h^{2d+1}}\right) \int_{\Pi_i} K\left(\frac{t_n - s}{h}\right) ds \\ &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \int_{\Pi_i} M_h(t_n - s) ds\ (1 + o(1)) + O\left(\frac{1}{n^{1+1/d}h^{d+1}}\right) \int_{\Pi_i} K_h(t_n - s)\, ds, \qquad (21) \end{aligned}$$

where we have defined $M(\cdot) = K^2(\cdot)/\|K^2\|_1 = K^2(\cdot)/\|K\|_2^2$. Replacing in $T_{n,1}$ yields

$$\begin{aligned} T_{n,1} &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \left\{ \sum_{i=1}^n \bar{F}_Y(y_n \mid x_i) \int_{\Pi_i} M_h(t_n - s) ds\ (1 + o(1)) \right. \\ &+ \left. O\left(\frac{1}{n^{1/d}h}\right) \sum_{i=1}^n \bar{F}_Y(y_n \mid x_i) \int_{\Pi_i} K_h(t_n - s) ds \right\}. \end{aligned}$$

Applying Lemma 1 with $p = 1$ twice and recalling that $nh^d \to \infty$ as $n \to \infty$ entail

$$\begin{aligned} T_{n,1} &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \left(\bar{F}_Y(y_n \mid t_n)\ (1 + o(1)) + O(h^2) + O\left(\frac{1}{n^{1/d}}\right)\right) \\ &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \bar{F}_Y(y_n \mid t_n)\ (1 + o(1)), \end{aligned}$$

under the assumption $\liminf F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n) > 0$. Similarly,

$$T_{n,2} = \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \bar{F}_Y^2(y_n \mid t_n)\ (1 + o(1)),$$

and the conclusion follows :

$$T_{n,1} - T_{n,2} = \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \bar{F}_Y(y_n \mid t_n) F_Y(y_n \mid t_n)\ (1 + o(1)),$$

under the assumption $\liminf F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n) > 0$. □

Finally, Lemma 3 is an adaptation of [20, Lemma 3]. It permits to derive the error made on the estimation of the order statistics Z_{m_n-i,m_n}, $i=0,\dots,m_n-1$ from the error made on the unsorted Z_i, $i\in I_n$.

Lemma 3. *Recall that* $I_n=\left\{i\in\{1,\dots,n\}\text{ such that } x_i\in\tilde\Pi^{(n)}\right\}$ *and* $m_n=\mathrm{card}(I_n)$. *Assume* $nh^d\to\infty$ *as* $n\to\infty$.

(i) Then, $m_n=n(1+O(h))$.

(ii) Consider (k_n) *an intermediate sequence of integers. If, for all* $i\in I_n$, $|\hat Z_i-Z_i|\le R_{n,i}\left(1+|Z_i|\right)$, *with* $\max_{i\in I_n}R_{n,i}\xrightarrow{\mathbb{P}}0$, *then*

$$\max_{0\le i\le k_n}\left|\log\frac{\hat Z_{m_n-i,m_n}}{Z_{m_n-i,m_n}}\right|=O_{\mathbb{P}}\left(\max_{i\in I_n}R_{n,i}\right).$$

Proof. (i) Let $C_n=\Pi\setminus\tilde\Pi^{(n)}$, $J_n=\{i\in\{1,\dots,n\}$ such that $x_i\in C_n\}$ and $N_n:=\mathrm{card}(J_n)$. For all $i\in J_n$, $x_i\in C_n$ and $nh^d\to\infty$ together with (15) entail that $\Pi_i\subset C_n$, for n large enough. Therefore, as the sets Π_i are disjoint :

$$\sum_{i\in J_n}\lambda(\Pi_i)\le\lambda(C_n)=\lambda(\Pi)-\lambda\left(\tilde\Pi^{(n)}\right)=O(h),$$

in view of the absolute continuity of the erosion with respect to Lebesgue measure, see [32]. From (14), $\lambda(\Pi_i)\sim\lambda(\Pi)/n$ uniformly on $i=1,\dots,n$ and thus $N_n=O(nh)$. Therefore, $m_n=n-N_n=n(1+O(h))$ as $n\to+\infty$.
(ii) The conclusion follows by remarking that in view of (2) the distribution of Z has an infinite upper endpoint and by applying [20, Lemma 3]. □

7.2 Preliminary results

Let $\vee$ (resp. $\wedge$) denote the maximum (resp. the minimum). The next proposition provides a joint asymptotic normality result for the estimator (7) of the conditional survival function evaluated at points depending on n.

Proposition 1. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n)\subset\tilde\Pi^{(n)}$ *and* $(\alpha_j)_{j=1,\dots,J}$ *a strictly decreasing sequence in* $(0,1)$. *For all* $j\in\{1,\dots,J\}$, *define* $y_{j,n}=q_Y(\alpha_j\mid t_n)+b(t_n)\varepsilon_{j,n}$, *where* $\varepsilon_{j,n}\to0$ *as* $n\to\infty$. *If* $nh^d\to\infty$ *and* $nh^{d+\kappa(d)}\to0$ *as* $n\to\infty$, *then*

$$\left\{\sqrt{nh^d}\left[\hat{\bar F}_{n,Y}(y_{j,n}\mid t_n)-\bar F_Y(y_{j,n}\mid t_n)\right]\right\}_{j=1,\dots,J}\xrightarrow{d}\mathcal{N}\left(0_{\mathbb{R}^J},\lambda(\Pi)\|K\|_2^2\,B\right),$$

where $B_{k,\ell}=\alpha_{k\vee\ell}(1-\alpha_{k\wedge\ell})$ *for all* $(k,\ell)\in\{1,\dots,J\}^2$.

Proof. Let us first remark that, for all $j\in\{1,\dots,J\}$, in view of (5), the sequence $y_{j,n}=a(t_n)+b(t_n)(q_Z(\alpha_j)+\varepsilon_{j,n})$ is bounded since $a(\cdot)$ and $b(\cdot)$ are continuous functions defined on compact sets and because $\varepsilon_{j,n} \to 0$ as $n \to \infty$. Besides, from (3), $F_Y(y_{j,n} \mid t_n) = F_Z(q_Z(\alpha_j) + \varepsilon_{j,n}) \to 1 - \alpha_j > 0$ as $n \to \infty$ and thus the assumptions of Lemma 2(i,ii) are satisfied. Now, let $\beta \neq 0$ in $\mathbb{R}^J$, $J \geq 1$ and consider the random variable

$$\begin{aligned}
\Gamma_n &= \sum_{j=1}^{J} \beta_j \left\{ \hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n) \right\} \\
&= \sum_{j=1}^{J} \beta_j \left\{ \hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right) \right\} \\
&+ \sum_{j=1}^{J} \beta_j \left\{ \mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right) - \bar{F}_Y(y_{j,n} \mid t_n) \right\} \\
&=: \Gamma_{n,1} + \Gamma_{n,2}.
\end{aligned}$$

The random term can be expanded as

$$\Gamma_{n,1} = \sum_{i=1}^{n} \int_{\Pi_i} K_h(t_n - s) ds \sum_{j=1}^{J} \beta_j \left\{ \mathbb{1}_{\{Y_i > y_{j,n}\}} - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{j,n}\}} \right) \right\} =: \sum_{i=1}^{n} T_{i,n}.$$

By definition, $\mathbb{E}(\Gamma_{n,1}) = 0$, and by independence of $Y_1, \dots, Y_n$,

$$\operatorname{var}(\Gamma_{n,1}) = \sum_{i=1}^{n} \left(\int_{\Pi_i} K_h(t_n - s) ds \right)^2 \operatorname{var}\left(\sum_{j=1}^{J} \beta_j \mathbb{1}_{\{Y_i > y_{j,n}\}} \right) =: \beta^t C(n) \beta,$$

where $C^{(n)}$ is the matrix whose coefficients are defined for all $(k, \ell) \in \{1, \dots, J\}^2$ by

$$C_{k,\ell}^{(n)} = \sum_{i=1}^{n} S_{n,i} \operatorname{cov}\left(\mathbb{1}_{\{Y_i > y_{k,n}\}}, \mathbb{1}_{\{Y_i > y_{\ell,n}\}} \right), \tag{22}$$

with $S_{n,i}$ being defined in (20) and expanded as (21) :

$$S_{n,i} = \frac{\lambda(\Pi) \|K\|_2^2}{nh^d} \int_{\Pi_i} M_h(t_n - s) ds \, (1 + o(1)) + O\left(\frac{1}{n^{1+1/d} h^{d+1}} \right) \int_{\Pi_i} K_h(t_n - s) ds,$$

see the proof of Lemma 2. Straightforward calculations yield

$$\begin{aligned}
\operatorname{cov}\left(\mathbb{1}_{\{Y_i > y_{k,n}\}}, \mathbb{1}_{\{Y_i > y_{\ell,n}\}} \right) &= \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i) - \bar{F}_Y(y_{k,n} \mid x_i) \bar{F}_Y(y_{\ell,n} \mid x_i) \\
&= \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i) - \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i) \bar{F}_Y(y_{k,n} \wedge y_{\ell,n} \mid x_i) \\
&= \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i) F_Y(y_{k,n} \wedge y_{\ell,n} \mid x_i) \\
&=: \varphi(y_{k,n}, y_{\ell,n} \mid x_i),
\end{aligned} \tag{23}$$

where φ is the function $\mathbb{R}^2 \times \Pi \to [0, 1]$ defined by $\varphi(\cdot, \cdot \mid .) = \bar{F}_Y(\cdot \vee \cdot \mid \cdot) F_Y(\cdot \wedge \cdot \mid \cdot)$.

Replacing in (22) yields

$$\begin{aligned}
C_{k,\ell}^{(n)} &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d}\sum_{i=1}^{n}\varphi(y_{k,n},y_{\ell,n}\mid x_i)\int_{\Pi_i}M_h(t_n-s)ds\,(1+o(1))\\
&+ O\left(\frac{1}{n^{1+1/d}h^{d+1}}\right)\sum_{i=1}^{n}\varphi(y_{k,n},y_{\ell,n}\mid x_i)\int_{\Pi_i}K_h(t_n-s)ds\\
&= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d}\left[\varphi(y_{k,n},y_{\ell,n}\mid t_n)+O\left(h^2\right)+O\left(n^{-1/d}\right)\right](1+o(1))\\
&+ O\left(\frac{1}{n^{1+1/d}h^{d+1}}\right)\left[\varphi(y_{k,n},y_{\ell,n}\mid t_n)+O\left(h^2\right)+O\left(n^{-1/d}\right)\right]\\
&= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d}\left[\varphi(y_{k,n},y_{\ell,n}\mid t_n)(1+o(1))+O\left(h^2\right)+O\left(n^{-1/d}\right)\right], \qquad (24)
\end{aligned}$$

from Lemma 1 applied twice with $p=2$ and recalling that $nh^d\to\infty$. Besides, let us remark that, in view of (5),

$$y_{k,n}-y_{\ell,n}=b(t_n)(q_Z(\alpha_k)-q_Z(\alpha_\ell)+\varepsilon_{k,n}-\varepsilon_{\ell,n})=b(t_n)(q_Z(\alpha_k)-q_Z(\alpha_\ell))(1+o(1)),$$

as $n\to\infty$. Therefore, assuming for instance $k<\ell$ implies $\alpha_k>\alpha_\ell$ and thus $q_Z(\alpha_k)<q_Z(\alpha_\ell)$ leading to $y_{k,n}<y_{\ell,n}$ for n large enough. More generally, $y_{k,n}\vee y_{\ell,n}=y_{k\vee\ell,n}$ and $y_{k,n}\wedge y_{\ell,n}=y_{k\wedge\ell,n}$ for n large enough and thus $\varphi(y_{k,n},y_{\ell,n}\mid t_n)=\bar F_Y(y_{k\vee\ell,n}\mid t_n)F_Y(y_{k\wedge\ell,n}\mid t_n)$. From (3) and (5), we have

$$\bar F_Y(y_{k,n}\mid t_n)=\bar F_Z\left(\frac{y_{k,n}-a(t_n)}{b(t_n)}\right)=\bar F_Z\left(q_Z(\alpha_k)+\varepsilon_{k,n}\right)=\alpha_k+o(1),$$

in view of the continuity of $\bar F_Z$. As a result,

$$\varphi(y_{k,n},y_{\ell,n}\mid t_n)\to B_{k,\ell}=\alpha_{k\vee\ell}(1-\alpha_{k\wedge\ell})\text{ as }n\to\infty. \qquad (25)$$

Collecting (24) and (25), one has

$$C_{k,\ell}^{(n)}=\frac{\lambda(\Pi)\|K\|_2^2}{nh^d}B_{k,\ell}(1+o(1))$$

and therefore

$$\operatorname{var}(\Gamma_{n,1})\sim\frac{\lambda(\Pi)\|K\|_2^2}{nh^d}\beta^tB\beta, \qquad (26)$$

where B is the matrix defined by the $B_{k,\ell}$ coefficients. The proof of the asymptotic normality of $\Gamma_{n,1}$ is based on Lyapounov (see Lo [83, Theorem 20, page 237]) criteria for triangular arrays of independent random variables :

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \Big/ (\mathrm{var}(\Gamma_{n,1}))^{3/2} \to 0 \tag{27}$$

as $n \to \infty$. Let us highlight that the random variables $T_{i,n}$, $i = 1, \ldots, n$, are bounded :

$$\begin{aligned} |T_{i,n}| &\leq \int_{\Pi_i} K_h(t_n - s)ds \sum_{j=1}^{J} \beta_j \left| \mathbb{1}_{\{Y_i > y_{j,n}\}} - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{j,n}\}}\right) \right| \\ &\leq \frac{\lambda(\Pi)\|K\|_\infty}{nh^d} \sum_{j=1}^{J} |\beta_j| \, (1 + o(1)) =: \zeta_n \end{aligned} \tag{28}$$

in view of **(A.3)** and (14). As a consequence, one has

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \leq \zeta_n \sum_{i=1}^{n} \mathbb{E}(T_{i,n}^2) = \zeta_n \sum_{i=1}^{n} \mathrm{var}(T_{i,n}) = \zeta_n \, \mathrm{var}(\Gamma_{n,1}),$$

leading to

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \Big/ (\mathrm{var}(\Gamma_{n,1}))^{3/2} = O\left((nh^d)^{-1/2}\right),$$

from (26) and (28). It is thus clear that (27) holds under the assumption $nh^d \to \infty$. A a result,

$$\sqrt{nh^d}\Gamma_{n,1} \xrightarrow{d} \mathcal{N}\left(0, \lambda(\Pi)\|K\|_2^2 \, \beta^t B \beta\right). \tag{29}$$

Let us now turn to the nonrandom term. Lemma 2(i) together with the assumptions $nh^d \to \infty$ and $nh^{d+\kappa(d)} \to 0$ as $n \to \infty$ entail

$$\sqrt{nh^d}|\Gamma_{n,2}| \leq \sqrt{nh^d} \sum_{j=1}^{J} |\beta_j| \left| \mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n)\right) - \bar{F}_Y(y_{j,n} \mid t_n) \right| = O\left(\sqrt{nh^{d+\kappa(d)}}\right) = o(1). \tag{30}$$

Finally, collecting (29) and (30), $\sqrt{nh^d}\Gamma_n$ converges to a centered Gaussian random variable with variance $\lambda(\Pi)\|K\|_2^2 \, \beta^t B \beta$, and the result follows. □

The following proposition provides the joint asymptotic normality of the estimator (8) of conditional quantiles. It can be read as an adaptation of classical results [2, 35, 38] to the location-dispersion regression model in the multivariate fixed design setting.

Proposition 2. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset \tilde{\Pi}^{(n)}$ *and* $(\alpha_j)_{j=1,\ldots,J}$ *a strictly decreasing sequence in* $(0, 1)$ *such that* $f_Z(q_Z(\alpha_j)) > 0$ *for all* $j \in \{1, \ldots, J\}$. *If*

$nh^d \to \infty$ and $nh^{d+\kappa(d)} \to 0$ as $n \to \infty$, then

$$\left\{ \frac{\sqrt{nh^d}}{b(t_n)} \left[\hat{q}_{n,Y}(\alpha_j \mid t_n) - q_Y(\alpha_j \mid t_n) \right] \right\}_{j=1,\ldots,J} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^J}, \lambda(\Pi)\|K\|_2^2\, C\right),$$

where C is the covariance matrix defined by $C_{k,\ell} = \alpha_{k\vee\ell}(1-\alpha_{k\wedge\ell})H_Z(\alpha_k)H_Z(\alpha_\ell)$ for all $(k,\ell) \in \{1,\ldots,J\}^2$.

Proof. Let $s = (s_1,\ldots,s_J) \in \mathbb{R}^J$, and for all $j = 1,\ldots,J$,

$$\begin{aligned}
\varepsilon_{j,n} &:= s_j/\sqrt{nh^d}, \\
\nu_{j,n} &:= b(t_n)\varepsilon_{j,n}, \\
y_{j,n} &= q_Y(\alpha_j \mid t_n) + \nu_{j,n}, \\
V_{j,n} &:= \sqrt{nh^d}\left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n)\right], \\
v_{j,n} &:= \sqrt{nh^d}\left[\alpha_j - \bar{F}_Y(y_{j,n} \mid t_n)\right].
\end{aligned}$$

These notations yield

$$W_n(s) := \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\frac{\sqrt{nh^d}}{b(t_n)}\left(\hat{q}_{n,Y}(\alpha_j \mid t_n) - q_Y(\alpha_j \mid t_n)\right) \le s_j\right\}\right) = \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{V_{j,n} \le v_{j,n}\right\}\right).$$

From (3) and (5), the nonrandom term can be rewritten as

$$v_{j,n} = \sqrt{nh^d}\left(\alpha_j - \bar{F}_Z\left(\frac{y_{j,n} - a(t_n)}{b(t_n)}\right)\right) = \sqrt{nh^d}\left(\alpha_j - \bar{F}_Z\left(q_Z(\alpha_j) + \varepsilon_{j,n}\right)\right).$$

Since $\bar{F}_Z(\cdot)$ is differentiable, for all $j \in \{1,\ldots,J\}$, there exists $\theta_{j,n} \in (0,1)$ such that

$$v_{j,n} = s_j f_Z\left(q_Z(\alpha_j) + \theta_{j,n}\varepsilon_{j,n}\right) = \frac{s_j}{H_Z(\alpha_j)}(1+o(1)), \tag{31}$$

in view of the continuity of $f_Z(\cdot)$ and since $\varepsilon_{j,n} \to 0$ as $n \to \infty$. Let us now turn to the random term. Recalling that, for all $j = 1,\ldots,J$, $y_{j,n} = q_Y(\alpha_j \mid t_n) + b(t_n)\varepsilon_{j,n}$, with $\varepsilon_{j,n} \to 0$ as $n \to \infty$, Proposition 1 entails that $\{V_{j,n}\}_{j=1,\ldots,J}$ converges to a centered Gaussian random vector with covariance matrix $\lambda(\Pi)\|K\|_2^2\, B$. Taking account of (31) yields that $W_n(s)$ converges to the cumulative distribution function of a centered Gaussian distribution with covariance matrix $\lambda(\Pi)\|K\|_2^2\, C$, evaluated at s, which is the desired result. □

The following proposition provides a uniform consistency result for the estimator (8) of conditional quantiles of Y given a sequence of multidimensional design points in $\tilde{\Pi}^{(n)}$, *i.e.* not too close from the boundary of Π.

Proposition 3. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Suppose* $nh^d/\log n \to \infty$ *and* $nh^{d+\kappa(d)}/\log n \to 0$ *as* $n \to \infty$. *Then, for all* $\alpha \in (0,1)$,

$$\sqrt{\frac{nh^d}{\log n}} \max_{i \in I_n} \left| \frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1).$$

Proof. Let $v_n = (nh^d/\log n)^{1/2}$ and for all $(\varepsilon, \alpha) \in (0,1)^2$, consider

$$\begin{aligned} \kappa_1(\varepsilon,\alpha) &= 2\|K\|_2 \left(\lambda(\Pi)\alpha(1-\alpha)\left(1 - \log(\varepsilon/2)\right)\right)^{1/2}, \\ \kappa_2(\alpha) &= \lambda(\Pi)\alpha(1-\alpha)\|K\|_2^2, \\ M(\varepsilon,\alpha) &= \kappa_1(\varepsilon,\alpha) H_Z(\alpha). \end{aligned}$$

Let us also introduce, for all $i \in I_n$,

$$\begin{aligned} q_{i,n}^{\pm} &= q_Y(\alpha \mid x_i) \pm M(\varepsilon,\alpha) b(x_i)/v_n, \\ \alpha_{i,n}^{\pm} &= \alpha - \mathbb{E}\left(\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{\pm} \mid x_i\right)\right), \\ \xi_{i,n}^{\pm} &= \left(\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y}\right)\left(q_{i,n}^{\pm} \mid x_i\right), \end{aligned}$$

so that the following expansion holds :

$$\begin{aligned} \delta_n &:= \mathbb{P}\left(v_n \max_{i \in I_n} \left| \frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)} \right| \geq M(\varepsilon,\alpha)\right) \\ &= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\hat{q}_{n,Y}(\alpha \mid x_i) \geq q_{i,n}^{+}\right\} \cup \left\{\hat{q}_{n,Y}(\alpha \mid x_i) \leq q_{i,n}^{-}\right\}\right) \\ &= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\alpha \leq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right)\right\} \cup \left\{\alpha \geq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^{-} \mid x_i\right)\right\}\right) \\ &= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\alpha_{i,n}^{+} \leq \xi_{i,n}^{+}\right\}\right) + \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\alpha_{i,n}^{-} \geq \xi_{i,n}^{-}\right\}\right) \\ &=: \delta_n^{+} + \delta_n^{-}. \end{aligned}$$

Let us focus on the first term. Assumption $nh^d/\log n \to \infty$ entails that $v_n \to \infty$ as $n \to \infty$ and thus $q_{i,n}^{+}$ is bounded. Therefore Lemma 2(i) shows that

$$\begin{aligned} \alpha_{i,n}^{+} &= \alpha - \bar{F}_Y\left(q_{i,n}^{+} \mid x_i\right) + O\left(h^2\right) + O\left(n^{-1/d}\right) \\ &= \bar{F}_Z(q_Z(\alpha)) - \bar{F}_Z\left(q_Z(\alpha) + \frac{M(\varepsilon,\alpha)}{v_n}\right) + O\left(h^2\right) + O\left(n^{-1/d}\right) \\ &= \frac{M(\varepsilon,\alpha)}{v_n} f_Z\left(q_Z(\alpha) + \frac{M(\varepsilon,\alpha)}{v_n}\theta\right) + O\left(h^2\right) + O\left(n^{-1/d}\right), \end{aligned}$$

for some $\theta \in (0,1)$, and the continuity of $f_Z(\cdot)$ then yields

$$\alpha_{i,n}^{+} = \frac{M(\varepsilon,\alpha)}{v_n H_Z(\alpha)}(1+o(1)) + O\left(h^2\right) + O\left(n^{-1/d}\right) = \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1)), \tag{32}$$

in view of the assumption $nh^{d+\kappa(d)}/\log n \to 0$ as $n \to \infty$. As a consequence,

$$\delta_n^{+} = \mathbb{P}\left(\bigcup_{i\in I_n}\left\{\xi_{i,n}^{+} \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right\}\right) \leq \sum_{i\in I_n}\mathbb{P}\left(\xi_{i,n}^{+} \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right). \tag{33}$$

Moreover,

$$\mathbb{P}\left(\xi_{i,n}^{+} \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right) \quad =: \quad \mathbb{P}\left(\sum_{j=1}^{n}\tilde{X}_j \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right), \tag{34}$$

where, for all $j = 1,\ldots,n$, the random variables

$$\tilde{X}_j := \left[\mathbb{1}_{\{Y_j > q_{i,n}^{+}\}} - \mathbb{P}\left(Y_j > q_{i,n}^{+} \mid x_i\right)\right]\int_{\Pi_j} K_h(x_i - s)ds$$

are independent, centered and bounded from (14) :

$$|\tilde{X}_j| \leq \int_{\Pi_j} K_h(x_i - s)ds \leq \frac{\lambda(\Pi)\|K\|_\infty}{nh^d}(1+o(1)).$$

Lemma 2(ii) entails

$$\begin{aligned}
\sum_{j=1}^{n}\mathbb{E}(\tilde{X}_j^2) &= \operatorname{var}\left(\sum_{j=1}^{n}\tilde{X}_j\right) = \operatorname{var}\left[\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right)\right] \\
&= \frac{\lambda(\Pi)\bar{F}_Y\left(q_{i,n}^{+} \mid x_i\right)F_Y\left(q_{i,n}^{+} \mid x_i\right)}{nh^d}\|K\|_2^2(1+o(1)), \\
&= \frac{\kappa_2(\alpha)}{nh^d}(1+o(1)),
\end{aligned}$$

since $\alpha_{i,n}^{+} \to 0$ as $n \to \infty$ from (32) and thus $\bar{F}_Y\left(q_{i,n}^{+} \mid x_i\right) \to \alpha$ as $n \to \infty$ in view of the continuity of $\overline{F}_Y(\cdot \mid x_i)$. Bernstein's inequality for bounded random variables yields

$$\begin{aligned}
(34) \quad &\leq \exp\left(-\frac{\kappa_1^2(\varepsilon,\alpha)\log n}{2\kappa_2(\alpha) + \frac{2\kappa_1(\varepsilon,\alpha)(1+o(1))}{3v_n}}(1+o(1))\right) \\
&= \exp\left(-\frac{\kappa_1^2(\varepsilon,\alpha)\log n}{2\kappa_2(\alpha)}(1+o(1))\right) \\
&= \exp\left[-2\left(1-\log(\varepsilon/2)\right)\log n\left(1+o(1)\right)\right] \\
&\leq \exp\left[-\left(1-\log(\varepsilon/2)\right)\log n\right], \qquad (35)
\end{aligned}$$

for n large enough. Collecting (33)-(35) leads to

$$\delta_n^+ \leq n\exp\left[-\left(1-\log(\varepsilon/2)\right)\log n\right] = \exp\left(\log(\varepsilon/2)\log n\right) \leq \varepsilon/2$$

for n large enough. The proof that $\delta_n^- \leq \varepsilon/2$ follows the same lines. As a conclusion, we have shown that, for all $\alpha \in (0,1)$ and $\varepsilon \in (0,1)$ there exists $M(\varepsilon,\alpha) > 0$ such that

$$\mathbb{P}\left(\sqrt{\frac{nh^d}{\log n}}\max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)}\right| \geq M(\varepsilon,\alpha)\right) \leq \varepsilon,$$

which is the desired result. □

7.3 Proofs of main results

The proof of Theorem 1 directly relies on Proposition 2 :

Proof of Theorem 1. Let us remark that

$$\frac{\sqrt{nh^d}}{b(t_n)}\begin{pmatrix}\hat{a}_n(t_n) - a(t_n) \\ \hat{b}_n(t_n) - b(t_n)\end{pmatrix} = \Omega\xi_n,$$

where $\Omega = \begin{pmatrix}0 & 1 & 0 \\ 1 & 0 & -1\end{pmatrix}$ and $\xi_n = \dfrac{\sqrt{nh^d}}{b(t_n)}\begin{pmatrix}\hat{q}_{n,Y}(\mu_3 \mid t_n) - q_Y(\mu_3 \mid t_n) \\ \hat{q}_{n,Y}(\mu_2 \mid t_n) - q_Y(\mu_2 \mid t_n) \\ \hat{q}_{n,Y}(\mu_1 \mid t_n) - q_Y(\mu_1 \mid t_n)\end{pmatrix}$.
Proposition 2 with $J = 3$ and $\alpha_j = \mu_j$, $j = 1,\dots,J$ yields that ξ_n converges in distribution to the $\mathcal{N}\left(0_{\mathbb{R}^3},\ \lambda(\Pi)\|K\|_2^2\, C\right)$ distribution where

$$C = \begin{pmatrix}
\mu_1(1-\mu_1)H_Z^2(\mu_1) & \mu_2(1-\mu_1)H_Z(\mu_2)(H_Z(\mu_1) & \mu_3(1-\mu_1)H_Z(\mu_3)H_Z(\mu_1) \\
\mu_2(1-\mu_1)H_Z(\mu_2)H_Z(\mu_1) & \mu_2(1-\mu_2)H_Z^2(\mu_2) & \mu_3(1-\mu_2)H_Z(\mu_2)H_Z(\mu_3) \\
\mu_3(1-\mu_1)H_Z(\mu_3)H_Z(\mu_1) & \mu_3(1-\mu_2)H_Z(\mu_2)H_Z(\mu_3) & \mu_3(1-\mu_3)H_Z^2(\mu_3)
\end{pmatrix}.$$

Therefore, $\Omega\xi_n \xrightarrow{d} \mathcal{N}(0_{\mathbb{R}^2}, \lambda(\Pi)\|K\|_2^2\, \Omega C\Omega^t)$ and the conclusion follows from $\Omega C \Omega^t = \Sigma$. □

Theorem 2 is a straightforward consequence of Proposition 3 :

Proof of Theorem 2. Remarking that

$$\max_{i\in I_n}\left|\frac{\hat{a}_n(x_i)-a(x_i)}{b(x_i)}\right| = \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\mu_2\mid x_i)-q_Y(\mu_2\mid x_i)}{b(x_i)}\right|,$$

the first part of the result is a consequence of Proposition 3 applied with $\alpha=\mu_2$. Similarly,

$$\begin{aligned}\max_{i\in I_n}\left|\frac{\hat{b}_n(x_i)-b(x_i)}{b(x_i)}\right| &\leq \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\mu_3\mid x_i)-q_Y(\mu_3\mid x_i)}{b(x_i)}\right| \\ &+ \max_{i\in I_n}\left|\frac{\hat{q}_{n,Y}(\mu_1\mid x_i)-q_Y(\mu_1\mid x_i)}{b(x_i)}\right|,\end{aligned}$$

and the conclusion follows from Proposition 3 with $\alpha\in\{\mu_3,\mu_1\}$. □

Proof of Corollary 1. Remark that for all $i\in I_n$, one has

$$\begin{aligned}|\hat{Z}_i-Z_i| &= \left|\frac{Y_i-\hat{a}_n(x_i)}{\hat{b}_n(x_i)}-Z_i\right| = \left|\frac{a(x_i)-\hat{a}_n(x_i)}{\hat{b}_n(x_i)}+\frac{\hat{b}_n(x_i)-b(x_i)}{\hat{b}_n(x_i)}Z_i\right| \\ &\leq \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\left(\left|\frac{\hat{a}_n(x_i)-a(x_i)}{b(x_i)}\right|+\left|\frac{\hat{b}_n(x_i)-b(x_i)}{b(x_i)}\right||Z_i|\right) \\ &\leq \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\max\left\{\left|\frac{\hat{a}_n(x_i)-a(x_i)}{b(x_i)}\right|;\left|\frac{\hat{b}_n(x_i)-b(x_i)}{b(x_i)}\right|\right\}(1+|Z_i|) \\ &=: \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\max\left\{\left|\xi_{i,n}^{(a)}\right|;\left|\xi_{i,n}^{(b)}\right|\right\}(1+|Z_i|)\,.\end{aligned}$$

Let us define, for all $i\in I_n$,

$$\xi_{i,n}^{(a)}=\frac{\hat{a}_n(x_i)-a(x_i)}{b(x_i)},\ \xi_{i,n}^{(b)}=\frac{\hat{b}_n(x_i)-b(x_i)}{b(x_i)}\ \text{and}\ R_{n,i}=\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\max\left\{\left|\xi_{i,n}^{(a)}\right|;\left|\xi_{i,n}^{(b)}\right|\right\}.$$

On the one hand, Theorem 2 entails

$$\max_{i\in I_n}R_{n,i} \leq \max_{i\in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\max\left\{\max_{i\in I_n}\left|\xi_{i,n}^{(a)}\right|;\max_{i\in I_n}\left|\xi_{i,n}^{(b)}\right|\right\} = \max_{i\in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right).$$

On the other hand,

$$\begin{aligned}\mathbb{P}\left(\max_{i\in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\geq 2\right) &= \mathbb{P}\left(\max_{i\in I_n}\left|\frac{1}{1+\xi_{i,n}^{(b)}}\right|\geq 2\right)\leq \mathbb{P}\left(\max_{i\in I_n}\left|\xi_{i,n}^{(b)}\right|\geq \frac{1}{2}\right)\\ &\leq \mathbb{P}\left(\sqrt{\frac{nh^d}{\log n}}\max_{i\in I_n}\left|\xi_{i,n}^{(b)}\right|\geq \frac{1}{2}\sqrt{\frac{nh^d}{\log n}}\right).\end{aligned}$$

Again, Theorem 2 shows that the following uniform consistency holds : For all $\varepsilon > 0$, there exists $M(\varepsilon) > 0$ such that

$$\mathbb{P}\left(\sqrt{\frac{nh^d}{\log n}}\max_{i\in I_n}\left|\xi_{i,n}^{(b)}\right|\geq M(\varepsilon)\right)\leq \varepsilon.$$

Now, for n large enough, $(nh^d/\log n)^{1/2} > 2M(\varepsilon)$ so that

$$\mathbb{P}\left(\max_{i\in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right|\geq 2\right)\leq \mathbb{P}\left(\max_{i\in I_n}\sqrt{\frac{nh^d}{\log n}}\left|\xi_{i,n}^{(b)}\right|\geq M(\varepsilon)\right)\leq \varepsilon,$$

i.e. $\max_{i\in I_n}|b(x_i)/\hat{b}_n(x_i)| = O_{\mathbb{P}}(1)$. As a result,

$$\max_{i\in I_n} R_{n,i} = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right),$$

which completes the proof of the corollary. □

Proof of Theorem 3. (i) Let us consider the expansion

$$\sqrt{k_n}(\hat{\gamma}_n-\gamma)=\sqrt{k_n}(\hat{\gamma}_n-\tilde{\gamma}_n)+\sqrt{k_n}(\tilde{\gamma}_n-\gamma)=:\Upsilon_{1,n}+\Upsilon_{2,n},$$

where

$$\tilde{\gamma}_n=\frac{1}{k_n}\sum_{i=0}^{k_n-1}\log Z_{m_n-i,m_n}-\log Z_{m_n-k_n,m_n}$$

is the Hill estimator computed on the unobserved random variables $Z_1,\dots,Z_n$. Recall that $m_n=\text{card}(I_n)$ where $I_n=\{i\in\{1,\dots,n\}$ such that $x_i\in\tilde{\Pi}^{(n)}\}$. The first term is controlled by remarking that

$$|\Upsilon_{1,n}|=\sqrt{k_n}|\hat{\gamma}_n-\tilde{\gamma}_n|\leq\sqrt{k_n}\max_{0\leq i\leq k_n}\left|\log\frac{\hat{Z}_{m_n-i,m_n}}{Z_{m_n-i,m_n}}\right|=O_{\mathbb{P}}\left(\sqrt{\frac{k_n\log n}{nh^d}}\right)=o_{\mathbb{P}}(1), \quad (36)$$

from Corollary 1 and Lemma 3(ii). Let us now focus on $\Upsilon_{2,n}$. Remarking that $m_n\sim n$ as $n\to\infty$ in view of Lemma 3(i), it is clear that $m_n/k_n\to\infty$ as $n\to\infty$. Besides, since

$|A| \in \mathcal{RV}_\rho$, we thus have $A(m_n/k_n) \sim A(n/k_n)$ as $n \to \infty$. Therefore, $\sqrt{k_n} A(m_n/k_n) \to \beta$ as $n \to \infty$ and, since $Z_1, \dots, Z_n$ are iid from (2), classical results on Hill estimator apply, see for instance [24, Theorem 3.2.5], leading to

$$\Upsilon_{2,n} \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2). \tag{37}$$

The conclusion follows from (36) and (37).
(ii) Let us introduce $v_n = \sqrt{k_n}/\log(k_n/(n\alpha_n))$ and consider the Weissman estimator computed on the unobserved random variables $Z_1, \dots, Z_n$:

$$\tilde{q}_{n,Z}(\alpha_n) = Z_{m_n-k,m_n} \left(\frac{\alpha_n m_n}{k_n}\right)^{-\tilde{\gamma}_n}.$$

The following expansion holds :

$$\begin{aligned} v_n(\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n)) &= v_n(\log \hat{q}_{n,Z}(\alpha_n) - \log \tilde{q}_{n,Z}(\alpha_n)) \\ &+ v_n(\log \tilde{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n)) \\ &=: T_{1,n} + T_{2,n}, \end{aligned}$$

with

$$|T_{1,n}| \le v_n \left| \log \frac{\hat{Z}_{m_n-k_n,m_n}}{Z_{m_n-k_n,m_n}} \right| + v_n |\hat{\gamma}_n - \tilde{\gamma}_n| \left| \log\left(\frac{\alpha_n m_n}{k_n}\right) \right| \;\; =: \;\; T_{1,1,n} + T_{1,2,n}.$$

First, $T_{1,1,n}$ is controlled by Corollary 1 and Lemma 3(ii) together with the assumptions $k_n \log n/(nh^d) \to 0$ and $k_n/(n\alpha_n) \to \infty$ as $n \to \infty$,

$$T_{1,1,n} = \frac{\sqrt{k_n}}{\log\left(\frac{k_n}{n\alpha_n}\right)} O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right) = \sqrt{\frac{k_n \log n}{nh^d}} O_{\mathbb{P}}\left(\frac{1}{\log\left(\frac{k_n}{n\alpha_n}\right)}\right) = o_{\mathbb{P}}(1). \tag{38}$$

Second, since $m_n \sim n$ as $n \to \infty$ (see Lemma 3(i)),

$$T_{1,2,n} = |\Upsilon_{1,n}|(1 + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1), \tag{39}$$

in view of (36). Collecting (38) and (39) yields

$$T_{1,n} = v_n(\log \hat{q}_{n,Z}(\alpha_n) - \log \tilde{q}_{n,Z}(\alpha_n)) = o_{\mathbb{P}}(1). \tag{40}$$

Let us now focus on $T_{2,n}$. As a consequence of [24, Theorem 4.3.8], Weissman estimator inherits its asymptotic distribution from Hill estimator :

$$v_n \left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2),$$

in view of (37). As a result,

$$T_{2,n} \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2). \tag{41}$$

The conclusion follows from (40) and (41). □

Proof of Theorem 4. Let $v_n = \sqrt{k_n}/\log(k_n/(n\alpha_n))$ and consider the following expansion :

$$\begin{aligned} &\frac{v_n}{b(t_n) q_Z(\alpha_n)} \left(\tilde{q}_{n,Y}(\alpha_n \mid t_n) - q_Y(\alpha_n \mid t_n) \right) \\ = \;& \frac{v_n}{q_Z(\alpha_n)} \left(\frac{\hat{a}_n(t_n) - a(t_n)}{b(t_n)} \right) + v_n \left(\frac{\hat{b}_n(t_n) - b(t_n)}{b(t_n)} \right) + v_n \frac{\hat{b}_n(t_n)}{b(t_n)} \left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) \\ =: \;& \frac{\sqrt{\frac{k_n}{nh^d}} \xi_n^{(a)}}{q_Z(\alpha_n) \log\left(\frac{k_n}{n\alpha_n}\right)} + \frac{\sqrt{\frac{k_n}{nh^d}} \xi_n^{(b)}}{\log\left(\frac{k_n}{n\alpha_n}\right)} + v_n \frac{\hat{b}_n(t_n)}{b(t_n)} \left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right). \end{aligned}$$

From Theorem 1, $\xi_n^{(a)} := \sqrt{nh^d} \left(\frac{\hat{a}_n(t_n) - a(t_n)}{b(t_n)} \right) = O_{\mathbb{P}}(1)$, $\xi_n^{(b)} := \sqrt{nh^d} \left(\frac{\hat{b}_n(t_n) - b(t_n)}{b(t_n)} \right) = O_{\mathbb{P}}(1)$ and thus,

$$\frac{\sqrt{\frac{k_n}{nh^d}} \xi_n^{(a)}}{q_Z(\alpha_n) \log\left(\frac{k_n}{n\alpha_n}\right)} + \frac{\sqrt{\frac{k_n}{nh^d}} \xi_n^{(b)}}{\log\left(\frac{k_n}{n\alpha_n}\right)} \xrightarrow{\mathbb{P}} 0,$$

in view of $k_n/(nh^d) \to 0$, $q_Z(\alpha_n) \to \infty$ and $n\alpha_n/k_n \to 0$ as $n \to \infty$. In addition, since $\xi_n^{(b)} = O_{\mathbb{P}}(1)$, it follows that

$$\frac{\hat{b}_n(t_n)}{b(t_n)} = 1 + \frac{\xi_n^{(b)}}{\sqrt{nh^d}} \xrightarrow{\mathbb{P}} 1. \tag{42}$$

Besides, from Theorem 3(ii),

$$v_n \left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) = v_n \left(\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n) \right) (1 + o_{\mathbb{P}}(1)) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2), \tag{43}$$

and collecting (42) and (43) yields

$$v_n \frac{\hat{b}_n(t_n)}{b(t_n)} \left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2).$$

The conclusion follows. □

Proof of Theorem 5. Recall that $v_n = \sqrt{k_n}/\log(k_n/(n\alpha_n))$. The proof follows the same lines as the one of Theorem 4 :

$$\frac{v_n}{q_Z(\alpha_n)} \max_{i \in I_n} \left| \frac{\tilde{q}_{n,Y}(\alpha_n \mid x_i) - q_Y(\alpha_n \mid x_i)}{b(x_i)} \right|$$
$$\leq \frac{v_n}{q_Z(\alpha_n)} \max_{i \in I_n} \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| + v_n \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| + v_n \left| \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right| \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} \right|.$$

From Theorem 2,

$$\frac{v_n}{q_Z(\alpha_n)} \max_{i \in I_n} \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| + v_n \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| \xrightarrow{\mathbb{P}} 0,$$

since $q_Z(\alpha_n) \to \infty$ and under the assumptions $nh^d/(k_n \log n) \to \infty$ and $n\alpha_n/k_n \to 0$ as $n \to \infty$. In addition,

$$\max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} \right| \leq \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} - 1 \right| + 1 = O_{\mathbb{P}}(1), \tag{44}$$

from Theorem 2, and

$$v_n \left| \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right| = v_n \left| (\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n))(1 + o_{\mathbb{P}}(1)) \right| = O_{\mathbb{P}}(1), \tag{45}$$

in view of Theorem 3(ii). Collecting (44) and (45) yields

$$v_n \left| \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right| \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1)$$

and the conclusion follows. □

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004). *Statistics of extremes : theory and applications*, John Wiley and Sons.

[2] Berlinet, A., Gannoun, A. and Matzner-Løber, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles, *Statistics*, 35(2), 139–169.

[3] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation*, Encyclopedia of Mathematics and its Applications, 27, Cambridge University Press.

[4] Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis : The kernel approach with S-Plus illustrations*, Oxford Statistical Science Series,

18, Oxford University Press.

[5] Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society : Series C*, 54, 207–222.

[6] Cowling, A. and Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society : Series B*, 58, 551–563.

[7] Daouia, A., Florens, J-P. and Simar, L. (2010). Frontier estimation and extreme value theory. *Bernoulli*, 16, 1039–1063.

[8] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589.

[9] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20(2), 311–333.

[10] Davison, A. C. and Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society : Series B*, 62, 191–208.

[11] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society : Series B*, 52, 393–442.

[12] Einmahl, J. H. J., de Haan, L. and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society : Series B*, 78, 31–51.

[13] Falk, M., Hüsler, J. and Reiss, R. D. (2004). *Laws of small numbers : Extremes and rare events*, 2nd edition, Birkhäuser.

[14] Feuerverger, A., and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *The Annals of Statistics*, 27(2), 760-781.

[15] Gardes, L. (2015). A general estimator for the extreme value index : applications to conditional and heteroscedastic extremes. *Extremes*, 18(3), 479–510.

[16] Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99(10), 2368–2388.

[17] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions : an application to the estimation of extreme rainfall return levels. *Extremes*, 13(2), 177–204.

[18] Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101(2), 419–433.

[19] Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1), 45–75.

[20] Girard, S., Stupfler, G. and Usseglio-Carleve, A. (2020). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Submitted*, `http://hal.inria.fr/hal-02531027`.

[21] Goegebeur, Y., Guillou, A. and Stupfler, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'IHP Probabilités et Statistiques*, 51(3), 1190–1213.

[22] Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes : a review. *International Statistical Review*, 83(2), 263–292.

[23] Gomes, M. I., and Martins, M. J. (2004). Bias reduction and explicit semi-parametric estimation of the tail index. *Journal of Statistical Planning and Inference*, 124(2), 361-378.

[24] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory : An Introduction*, New York, Springer.

[25] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1), 60–70.

[26] de Haan, L., Tank, A. K. and Neves, C. (2015). On tail trend detection : modeling relative risk. *Extremes*, 18(2), 141–178.

[27] Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15, 153–167.

[28] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3, 1163–1174.

[29] Jalbert, J., Favre, A. C., Bélisle, C. and Angers, J. F. (2017). A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *Journal of the Royal Statistical Society : Series C*, 66(5), 941–962.

[30] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity : A quantile regression survival analysis. *Journal of American Statistical Association*, 96, 458–468.

[31] Kyung-Joon, C. and Schucany, W. (1998). Nonparametric kernel regression estimation near endpoints. *Journal of Statistical Planning and Inference*, 66, 289–304.

[32] Matheron, G. (1981). Sur la négligeabilité du squelette et la continuité absolue des érosions. *Technical Report N-724*, Centre de Géostatistique et de Morphologie Mathématique, 1–26.

[33] Müller, H. G. and Prewitt, K. (1991). Applications of multiparameter weak convergence for adaptive nonparametric curve estimation. In *Nonparametric Functional Estimation and Related Topics*, pp. 141–166, Springer, Dordrecht.

[34] Müller, H. G. and Prewitt, K. (1993). Multiparameter bandwidth processes and adaptive surface smoothing. *Journal of Multivariate Analysis*, 47, 1–21.

[35] Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics and Probability Letters*, 7(5), 407–412.

[36] Serra, J. (1983). *Image analysis and mathematical morphology.* Academic Press, Inc.

[37] Smith, R. L. (1989). Extreme value analysis of environmental time series : an application to trend detection in ground-level ozone (with discussion). *Statistical Science*, 4, 367–393.

[38] Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5(4), 595–645.

[39] Van Keilegom, I. and Wang, L. (2010). Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data. *Electronic Journal of Statistics*, 4, 133–160.

[40] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, 73(364), 812–815.

Conclusion and perspectives

In this thesis, we have proposed an estimator of the conditional tail-index as well as an estimator of the conditional extreme quantiles, both constructed in a semi-parametric way. Beyond the originality of the problem dealt with, the challenge was to propose these new estimators by considering a regression model with location and dispersion functions but also to obtain better convergence rates than those obtained by some purely non-parametric approaches.

First, a one-dimensional fixed design setting was considered and we proposed estimators of the location and scale functions as well as of the conditional tail-index in the case of heavy-tailed distributions. Second, a multidimensional fixed design setting was considered and we proposed conditional extreme quantiles estimators for heavy-tailed distributions. In both cases, the asymptotic properties of the proposed estimators were established under mild assumptions.

Both the theoretical and practical obtained results confirm this research work which led to two contributions within the framework of the conditional tail-index and the conditional extreme quantiles estimation under a location-dispersion regression model.

As a perspective, it would be interesting, first of all, to generalize our work to all domains of attraction. Taking account the importance of the scope of application of the distributions in the Gumbel and Weibull domains of attraction, this extension would offer a wide range of statistical tools for estimating the tail-index as well as the conditional extreme quantiles. To this end, we could rely on the results in Daouia et al. [26] who generalize some estimators of the conditional extreme values for heavy-tailed distributions in any domain of attraction. The extension of our model to a random design setting is also a research perspective that could be very interesting, both theoretically and practically. For example, we could consider a model of the type $Y = a(X) + b(X)Z$ where Y is a real random variable recorded simultaneously with a multidimensional random covariate X and Z is a real random variable belongs to any domain of attraction and independent of X. Under the assumptions $\mathbb{E}(a(X)) = 0$ and $\mathbb{E}(b(X)) = 1$, one can prove that $\mathbb{E}(Y) = \mathbb{E}(Z)$ and we could then infer on the parameters $a(\cdot)$ and $b(\cdot)$ as well as the tail-index and the conditional extremes quantiles.

In this context, the set (denoted I_n in our case) of the indices of the residuals used for the estimation of the conditional tail-index will be random, and consequently its cardinal (m_n) too. The difficulty will therefore be taking this aspect into account.
We could also consider the case of a functional random covariate and rely on the results of the estimation of the extreme quantiles already established in Gardes and Girard [55] to propose semi-parametric estimators of the extreme values under a location-dispersion regression model.
Finally, the adaptation of our model to the case of censored random variables is also another intersting perspective. Taking account the rise of the extreme values theory in the case of censored data, it would be important to propose more tools for the estimation of extreme values in this context. As such, the estimators (1.43) and (1.44) proposed by Ndao et al. [89] in fixed design setting as well as those proposed by Stupfler [103], for example, can be used to this adaptation.

Appendix

Bochner's lemma (Bochner [17])

Let E be an ordinary Euclidian space of dimension $k \geq 1$. For any $x = (x_1, \dots, x_k) \in E$, the ordinary Lebesgue measure is denoted by dx and we also put $|x| = \left(\sum_{i=1}^{k} x_i^2\right)^{1/2}$. For a mesurable function $f(x) = f(x_1, \dots, x_k)$ in E, we introduce, if definable the approximating functions

$$f_n(x) = \frac{1}{h} \int_E f(x-u)\, K\left(\frac{u}{h}\right)\, du,$$

where $h = h_n$ is a sequence of positive constants such that $h_n \to 0$ as $n \to \infty$ and K is a bounded probability density function in E. If $f(x)$ is bounded in E

$$|f(x)| \leq M,$$

then

$$\lim_{n\to\infty} f_n(x) = f(x)$$

at every point x of continuity of $f(\cdot)$.

Bernstein's inequality (Bernstein [14])

Let $X_1, \dots, X_n$ be independent zero-mean random variables. Suppose that $|X_i| \leq M$ almost surely, for all i. Then for all $t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^{n} \mathbb{E}\,(X_i^2) + \frac{1}{3}Mt}\right).$$

Lyapounov's condition - Central limit theorem (Billingsley [15])

Let $X_1, \ldots, X_n$ be independent zero-mean random variables such that $\mathbb{E}\left(X_i^2\right) =: \sigma_i^2 < \infty$ and $s_n := \sum_{i=1}^{n} \sigma_i^2$. If for some positive ϵ the **Lyapounov's condition**

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\epsilon}} \sum_{i=1}^{n} \mathbb{E}\left(|X_i|^{2+\epsilon}\right) = 0$$

holds, then

$$\frac{1}{\sqrt{s_n}} \sum_{i=1}^{n} X_i \xrightarrow{d} \mathcal{N}(0,1).$$

Mathematical erosion (Serra [99])

The erosion of a subset X of a space E by a structuring element B (a set known a priori) is defined by :

$$\mathcal{E}_B(X) := \{x \in E, B_x \subset X\},$$

where $B_x = \{b + x, b \in B\}$ is the translate of B by $x \in E$.

Bibliography

[1] Aarssen, K. and de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies*, 4(4), 259–281. 21

[2] Anderson, P. L. and Meerschaert, M. M. (1998). Modeling river flows with heavy tails. *Water Resources Research*, 34(9), 2271–2280. 20

[3] Balkema, A. A. and de Haan, L. (1974). Residual life time at a great age. *The Annals of Probability*, 2(5), 792–804. 15

[4] Bechler, A., Bel, L. and Vrac, M. (2015). Conditional simulations of the extremal t process : application to fields of extreme precipitation. *Spatial Statistics*, 12, 109–127. 4

[5] Beirlant, J. and Goegebeur, Y. (2003). Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis*, 42, 595–619. 37

[6] Beirlant, J. and Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis*, 89, 97–118. 37

[7] Beirlant, J. and Teugels, J. L. (1992). Modelling large claims in non-life insurance. *Insurance : Mathematics and Economics*, 11(1), 17–29. 4, 22

[8] Beirlant, J., Vynckier, P. and Teugels, J. L. (1996). Excess functions and estimation of the extreme value index. *Bernoulli*, 2(4), 293–318. 23, 25

[9] Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002). On exponential representations of log-spacings of extreme order statistics. *Extremes*, 5, 157–180. 25, 37

[10] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004). *Statistics of Extremes : Theory and Applications*. John Wiley and Sons. 10, 24, 37

[11] Beirlant, J., Guillou, A., Dierckx, G. and Fils-Villetard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, 10, 151–174. 42

[12] Bel, L., Bacro, J. N. and Lantuéjoul, C. (2008). Assessing extremal dependence of environmental spatial fields. *Environmetrics*, 19, 163–182. 4

[13] Beran, R. (1981). Nonparametric regression with randomly censored data. Technical report, University of California, Berkeley. 41

[14] Bernstein, S. N. (1946). *The Theory of Probabilities.* Gastehizdat Publishing House, Moscow. 120

[15] Billingsley, P. (1986). *Probability Theory and Measure.* 2nd edition, John Wiley and Sons. 121

[16] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation.* Encyclopedia of Mathematics and its Applications, 27, Cambridge University Press. 9, 18, 19

[17] Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability.* University of California Press. 120

[18] Bouchaud, J.-P. and Potters, M. (2003). *Theory of Financial Risk and Derivative Pricing : From Statistical Physics to Risk Management.* 2nd edition, Cambridge University Press. 20

[19] Breiman, L., Stone, C. J. and Kooperberg, C. (1990). Robust confidence bounds for extreme upper quantiles. *Journal of Statistical Computation and Simulation*, 37(3–4), 127–149. 34

[20] Caeiro, F., Gomes, M. I. and Pestana, D. (2005). Direct reduction of bias of the classical hill estimator. *REVSTAT - Statistical Journal*, 3(2), 113–136. 29

[21] Ceresetti, D., Ursu, E., Carreau, J., Anquetin, S., Creutin, J. D., Gardes, L., Girard, S. and Molinié, G. (2012). Evaluation of classical spatial-analysis schemes of extreme rainfall. *Natural Hazards and Earth System Sciences*, 12, 3229–3240. 4

[22] Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Applied Statistics*, 54(1), 207–222. 37

[23] Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society : Series C*, 45(4), 463–478. 4

[24] Csörgő, S. and Mason, D. M. (1985). Central limit theorems for sums of extreme values. *Mathematical Proceedings of the Cambridge Philosophical Society*, 98(3), 547–558. 24

[25] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20(2), 311–333. 38

[26] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B), 2557–2589. 39, 118

[27] Davis, R. and Resnick, S. I. (1984). Tail estimates motivated by extreme value theory. *The Annals of Statistics*, 12(4), 1467–1487. 24

[28] Davison, A. C. and Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society : Series B*, 62(1), 191–208. 37

[29] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society : Series B*, 52(3), 393–442. 34, 36

[30] de Haan, L. (1990). Fighting the arch-enemy with mathematics. *Statistica Neerlandica*, 44(2), 45–68. 4, 22

[31] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory : An Introduction*. New York, Springer. 10, 19, 22, 24, 25, 27, 29, 31, 36

[32] de Haan, L. and Resnick, S. I. (1998). On asymptotic normality of the Hill estimator. *Stochastic Models*, 14(4), 849–866. 24

[33] Deheuvels, P., Haeusler, E. and Mason, D. M. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, 104(2), 371–381. 24

[34] Dekkers, A. L. M. and de Haan, L. (1989). On the estimation of the extreme value index and large quantile estimation. *The Annals of Statistics*, 17(4), 1795–1832. 27

[35] Dekkers, A. L. M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *Journal of Multivariate Analysis*, 47, 173–195. 25

[36] Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17(4), 1833–1855. 23, 28, 29

[37] Delafosse, E. and Guillou, A. (2004). Extreme quantiles estimation for actuarial applications. *C. R. Acad. Sci. Paris*, 1(339), 287–292. 41

[38] Deme, E. H., Gardes, L. and Girard, S. (2013). On the estimation of the second order parameter for heavy-tailed distributions. *REVSTAT - Statistical Journal*, 11 (3), 277–299. 25

[39] Diebolt, J., Gardes, L., Girard, S. and Guillou, A. (2008). Bias-reduced estimators of the weibull tail-coefficient. *Test*, 17(2), 311–331. 29

[40] Diop, A. and Lo, G. S. (2006). Generalized hill's estimator. *Journal of Theorical Statistic*, 20(2), 129–149. 24

[41] Ditlevsen, O. (1994). Distribution arbitrariness in structural reliability. *Structural Safety and Reliability,* 1241–1247. 4

[42] Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, 21(1), 420–436. 33

[43] Dress, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75, 149–172. 25

[44] Dutfoy, A., Parey, S. and Roche, N. (2014). Multivariate extreme value theory - A tutorial with applications to hydrology and meteorology. *Dependence Modeling*, 2, 30–48. 4

[45] Einmahl, J. H. J., Fils-Villetard, A. and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli*, 14(1), 207–227. 42

[46] El Methni, J., Gardes, L., Girard, S. and Guillou, A. (2012). Estimation of extreme quantiles from heavy and light tailed distributions. *Journal of Statistical Planning and Inference*, 142(10), 2735–2747. 4, 20

[47] El Methni, J., Gardes, L. and Girard, S. (2014). Nonparametric estimation of extreme risk measures from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41(4), 988–1012. 4, 20

[48] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, New York. 4, 10, 23, 36

[49] Embrechts, P., Resnick, S. I. and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2), 30–41. 4

[50] Falk, M. (1995). Some best parameter estimates for distributions with finite endpoint. *Statistics*, 27, 115–125. 21

[51] Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory : PWM estimators. *The Annals of Statistics*, 43(1), 276–298. 33

[52] Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24, Cambridge University Press, 180–190. 9, 10

[53] Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99(10), 2368-2388. 37, 38, 42

[54] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions :an application to the estimation of extreme rainfall return levels. *Extremes*, 13 (2), 177–204. 4, 20, 37, 43

[55] Gardes, L. and Girard, S. (2012). Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics*, 6, 1715–1744. 39, 119

[56] Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1), 45–75. 39

[57] Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101(2), 419-433. 38

[58] Gardes, L., Guillou, A. and Schorgen, A. (2012). Estimating the conditional tail index by integrating a kernel conditional quantile estimator. *Journal of Statistical Planning and Inference*, 142(6), 1586–1598. 37

[59] Geluk, J. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands. 25

[60] Girard, S., Guillou, A. and Stupfler, G. (2012). Estimating an endpoint with high order moments in the Weibull domain of attraction. *Statistics & Probability Letters*, 82(12), 2136–2144. 21

[61] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aléatoire. *Annals of Mathematics*, 44(3), 423–453. 9, 10, 21

[62] Gomes, M. I. and Martins, M. J. (2004). Bias reduction and explicit semi-parametric estimation of the tail index. *Journal of Statistical Planning and Inference*, 124, 361–378. 29

[63] Gomes, M. I. and Neves, M. M. (2011). Estimation of the extreme value index for randomly censored data. *Biometrical Letters*, 48(1), 1–22. 42

[64] Gomes, M. I. and Oliveira, O. (2003). Censoring estimators of a positive tail index. *Statistics & Probability Letters*, 65, 147–159. 41

[65] Gomes, M. I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 5, 387–414. 25

[66] Gomes, M. I., Martins, M. J. and Neves, M. M. (2007). Improving second order reduced bias extreme value index estimator. *REVSTAT - Statistical Journal*, 5(2), 177–207. 29

[67] Guillou, A. and Willems, P. (2006). Application de la théorie des valeurs extrêmes en hydrologie. *Revue de statistique appliquée*, 54(2), 5–31. 4

[68] Gumbel, E. J. (1941). The return period of flood flows. *The Annals of Mathematical Statistics*, 12(2), 163–190. 22

[69] Gumbel, E. J. (1954). Statistical theory of extreme values and some practical applications. *NBS Applied Mathematics Series*, 33(6). 4, 22

[70] Gumbel, E. J. (1958). *Statistics of Extremes*. New York, Columbia University Press. 4, 22, 33

[71] Haeusler, E. and Teugels, J. L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics*, 13(2), 743–756. 24

[72] Hall, P. (1982). On estimating the endpoint of a distribution. *The Annals of Statistics*, 10(2), 556–568. 21

[73] Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15(2), 153–167. 38

[74] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5), 1163–1174. 23

[75] Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3), 339–349. 23, 33, 34

[76] Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3), 251–261. 23, 33

[77] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *The Quarterly Journal of the Royal Meteorological Society*, 81(384), 158–171. 11

[78] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. 41

[79] Katz, R. W., Parlange, M. B. and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25, 1287–1304. 4

[80] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50. 37

[81] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions : Theory and Applications*. London, Imperial College Press. 3

[82] Kratz, M. and Resnick, S. I. (1996). The qq-estimator and heavy tails. *Communications in Statistics. Stochastic Models*, 12(4), 699–724. 23

[83] Lo, G. S. (2018). *Mathematical Foundations of Probability Theory.* DOI : http ://dx.doi.org/10.16929/sbs/2016.0008. 104

[84] Lo, G. S., Ngom, K. T. A. and Diallo, M. (2018). *Weak Convergence (IIA) - Functional and Random Aspects of the Univariate Extreme Value Theory.* Arxiv : 1810.01625. 10

[85] Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10(3), 754–764. 24

[86] Matthys, G. and Beirlant, J. (2000). Adaptive threshold selection in tail index estimation. In *Extremes and Integrated Risk Management*, pp. 37–57, UBS Warburg. 25

[87] Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, 13, 853–880. 37

[88] McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management : Concepts, Techniques and Tools.* Princeton University Press. 4

[89] Ndao, P., Diop, A. and Dupuy, J.-F. (2014). Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. *Computational Statistics and Data Analysis*, 79, 63–79. 42, 119

[90] Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1), 119–131. 9, 14, 15, 23, 27

[91] Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3), 723–724. 23, 33

[92] Prescott, P. and Walden, A. T. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. *Journal of Statistical Computation and Simulation*, 16(3–4), 241–250. 23, 33

[93] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics : With Applications to Nonparametric Statistics*. Springer-Verlag, New York. 3

[94] Reiss, R.-D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values : With Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser Verlag. 4

[95] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York. 10, 17, 18, 19, 21, 22

[96] Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin*, 27(1), 139–151. 4

[97] Rootzén, H. and Tajvidi, N. (1997). Extreme value statistics and wind storm losses : a case study. *Scandinavian Actuarial Journal*, 1, 70–94. 4

[98] Rootzén, H. and Tajvidi, N. (2001). Can losses caused by wind storms be predicted from meteorological observations ? *Scandinavian Actuarial Journal*, 2, 162–175. 4

[99] Serra, J. (1983). *Image Analysis and Mathematical Morphology*. Academic Press, Inc. 121

[100] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67–90. 23, 33

[101] Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, 15(3), 1174–1207. 24, 34

[102] Smith, R. L. (1989). Extreme value analysis of environmental time series : an application to trend detection in ground-level ozone (with discussion). *Statistical Science*, 4(4), 367–393. 36

[103] Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, 144, 1–24. 43, 119

[104] von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Revue de Mathématique Union Interbalcanique*, 1, 141–160. 11

[105] Wang, H. and Tsai, C. (2009). Tail index regression. *Journal of the American Statistical Association*, 104(487), 1233–1240. 37

[106] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364), 812–815. 35, 36

[107] Worms, J. and Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes*, 17, 337–358. 42

[108] Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100(4), 794–815. 33

[109] Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 101(4), 971–983. 33