Thèse

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE

UMR 7501

Strasbourg

présentée pour obtenir le grade de docteur de
l'Université de Strasbourg
Spécialité MATHÉMATIQUES APPLIQUÉES

Claire Roman

Étude des valeurs extrêmes en présence d'une covariable de grande dimension

Soutenue le 12 septembre 2019
devant la commission d'examen

Armelle Guillou, directeur de thèse
Laurent Gardes, directeur de thèse
Valérie Chavez-Demoulin, rapporteur
Clément Dombry, rapporteur

irma.math.unistra.fr

Institut de Recherche Mathématique Avancée

Université de Strasbourg

Extreme values study with high-dimensional covariate

Claire Roman

Contents

Publication

Article accepté :

Estimation of extreme conditional quantiles under a general tail first order condition, *Annals of the Institute of Statistical Mathematics* (avec A. Guillou et L. Gardes, Chapitre 2).

Chapter 1

Introduction

La théorie des valeurs extrêmes (TVE) constitue une branche importante des statistiques qui n'a cessé de gagner en importance dans plusieurs domaines ces dernières années. Elle permet d'étudier et de modéliser des évènements que l'on nomme rares, c'est-à-dire dont la probabilité d'apparition est faible. Contrairement à la statistique classique, qui met l'accent sur l'étude des comportements moyens d'un phénomène (calcul de l'espérance, de la variance, loi forte des grands nombres, théorème central limite, etc.), on se concentre sur les comportements exceptionnels, c'est-à-dire sur la *queue de distribution* de la loi modélisant au mieux le phénomène étudié. L'étude d'évènements rares ne peut se faire avec des outils de la statistique classique qui présentent rapidement des limites étant donné le nombre très restreint de données dont on peut disposer. Le rôle de la théorie des valeurs extrêmes est d'*extrapoler* au delà des données utilisables par le biais d'outils différents et efficaces. Les premières motivations du développement de cette théorie ont été les problèmes hydrologiques et environnementaux (inondations, crues, chocs pétroliers, etc.) dont la gestion est une préoccupation importante de par leurs impacts économiques et sociaux. Elle prend également une part active en finance (Embrechts et al. [12]) ou encore en sciences humaines, notamment sur la question de l'âge maximal que peut atteindre un être humain (Thatcher [37]).

Un des problèmes fondamentaux en théorie des valeurs extrêmes est l'estimation de quantités appelées *quantiles extrêmes*. Soit $S = 1 - F$ la fonction de survie associée à une fonction de répartition F. Le quantile $Q(\alpha)$ d'ordre $\alpha \in \;]0,1[$ associé à S est défini par l'inverse généralisé de S, c'est-à-dire

$$Q(\alpha) = S^{\leftarrow}(\alpha) = \inf \{y \in \mathbb{R} \mid S(y) \leq \alpha\}.$$

Un des exemples les plus connus qui explique au mieux la notion de quantile est le problème de hauteur de digue aux Pays-Bas (de Haan [24]). Une violente tempête est survenue aux Pays-Bas en 1953 pendant laquelle la mer a débordé par dessus les digues, ce qui a engendré d'importantes inondations et causé de nombreux dégâts tant sur le plan matériel que sur le plan humain. Une des questions posées aux mathématiciens a été de calculer la hauteur que devait avoir une digue afin que la mer passe par dessus avec une probabilité très faible α (de l'ordre de 10^{-4} à 10^{-3}). Cela revient en fait à résoudre un problème d'estimation de quantiles dont l'ordre est proche de 0, c'est-à-dire de quantiles extrêmes.

Dans cette thèse, on s'intéresse à l'estimation de quantiles extrêmes dans un certain contexte. On associe une variable aléatoire réelle Y à une covariable X de grande dimension, c'est-à-dire $X \in \mathbb{R}^p$ avec $p \in \mathbb{N} \setminus \{0\}$. Si on pose $x_0 \in \text{supp}(X)$ où $\text{supp}(X)$ est le support de X, alors on cherche à estimer les quantiles extrêmes conditionnels pour α proche de 0 définis par

$$Q(\alpha|x_0) = \inf \{y \in \mathbb{R} \mid S(y|x_0) \leq \alpha\}$$

où $S(\cdot|x_0)$ est la fonction de survie de la distribution de Y sachant $\{X = x_0\}$. Étant donné un estimateur $\tilde{S}_n(\cdot|x_0)$ de $S(\cdot|x_0)$, il est naturel de définir un estimateur de quantiles par la simple relation d'inverse généralisé et ainsi obtenir l'estimateur de quantiles

$$\tilde{Q}_n(\alpha|x_0) := \tilde{S}_n^{\leftarrow}(\alpha|x_0).$$

Il s'agit de la méthode *indirecte* d'estimation de quantiles. C'est par cette première remarque que débute le travail de cette thèse. Avant de formaliser correctement le problème, rappelons d'abord quelques généralités sur la théorie des valeurs extrêmes dans le cadre univarié.

1.1 Généralités sur la TVE dans le cadre univarié

Soient $Y_1, \ldots, Y_n$ n copies d'une variable aléatoire Y ayant comme fonction de répartition F.

1.1.1 Comportement asymptotique de max$(Y_1, \ldots, Y_n)$

En théorie des valeurs extrêmes, on se concentre sur la queue de distribution et donc sur le comportement asymptotique de la suite $M_n = \max(Y_1, \ldots, Y_n)$. On peut également s'intéresser à la suite $\min(Y_1, \ldots, Y_n)$ mais tout résultat sur M_n est suffisant puisqu'on a la relation $\min(Y_1, \ldots, Y_n) = -\max(-Y_1, \ldots, -Y_n)$. Par un calcul simple, la distribution du maximum M_n est donnée pour $x \in \mathbb{R}$ par

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

et ainsi l'étude du comportement asymptotique de M_n revient à étudier celui de F^n. Une première remarque que l'on peut faire est que le maximum M_n converge vers une variable *dégénérée*, c'est-à-dire une variable qui est égale à une constante presque sûrement. En effet, par le lemme de Borel-Cantelli, on montre que

$$M_n \xrightarrow{\text{p.s}} x_F$$

quand n tend vers $+\infty$ où x_F est le *point terminal* de F c'est-à-dire $x_F := \sup \{x \in \mathbb{R} \mid F(x) < 1\}$ qui peut être fini ou infini. Ce résultat ne nous donne pas d'informations essentielles comme le comportement de la queue de distribution ou encore une vitesse de convergence. Un moyen de contourner le problème est de renormaliser la variable M_n et de montrer que F appartient à un *domaine d'attraction.*

Definition 1 *Soit G une fonction de répartition non-dégénérée. On dit que F appartient au domaine d'attraction de G (et on note $F \in \mathcal{DA}(G)$) s'il existe des suites réelles $a_n > 0$ et b_n telles que en tout point de continuité x de G,*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \Leftrightarrow \frac{M_n - b_n}{a_n} \xrightarrow{\text{d}} Z$$

où Z admet pour fonction de répartition la fonction G.

Une très grande majorité de lois appartient à un domaine d'attraction, d'où l'intérêt d'étudier davantage cette notion. Les mathématiciens Fisher et Tippett [14] ainsi que Gnedenko [21] ont montré que si la fonction de répartition F appartient à un domaine d'attraction d'une fonction G, alors elle appartient également au domaine d'attraction d'une fonction explicite G_γ qui dépend d'un unique paramètre $\gamma \in \mathbb{R}$. Plus précisément, nous avons le théorème suivant.

Theorem 1 *Supposons que $F \in \mathcal{DA}(G)$ où G est non-dégénérée. Alors il existe des constantes $a, b, \gamma \in \mathbb{R}$ telles que $G(x) = G_\gamma(ax + b)$ pour tout $x \in \mathbb{R}$, où*

$$G_\gamma(x) := \begin{cases} \exp\left[-(1+\gamma x)^{-1/\gamma}\right] & si\ \gamma \neq 0\ et\ 1+\gamma x > 0, \\ \exp\left[-\exp(-x)\right] & si\ \gamma = 0. \end{cases}$$

Ce théorème est considéré comme étant le théorème fondamental de la théorie des valeurs extrêmes et l'équivalent du théorème central limite dans le domaine des statistiques classiques. La fonction G_γ est connue sous le nom de représentation de Jenkinson - Von Mises de la loi des valeurs extrêmes ou encore de distribution GEV (pour Generalised Extreme-Value). Le paramètre γ est appelé *indice des valeurs extrêmes* et renseigne sur le comportement de la queue de distribution de F. Selon son signe, on peut distinguer trois sortes de distributions GEV :

- Si $\gamma > 0$, on dit que F appartient au domaine d'attraction de *Fréchet* (noté $F \in \mathcal{DA}(\text{Fréchet})$) et la distribution GEV est donnée par :

  $$G_\gamma^{\mathrm{F}}(x) = \begin{cases} 0 & \text{si } x < 0, \\ \exp(-x^{-1/\gamma}) & \text{si } x \geq 0. \end{cases}$$

  Le point terminal x_F est dans ce cas $+\infty$ et la fonction de survie S est à décroissance polynomiale. Les fonctions de répartition $F \in \mathcal{DA}(\text{Fréchet})$ sont dites à queues lourdes (ou heavy tailed distributions en anglais).

- Si $\gamma < 0$, on dit que F appartient au domaine d'attraction de *Weibull* (noté $F \in \mathcal{DA}(\text{Weibull})$) et la distribution GEV est donnée par :

  $$G_\gamma^{\mathrm{W}}(x) = \begin{cases} \exp[-(-x)^{-1/\gamma}] & \text{si } x \leq 0, \\ 1 & \text{si } x > 0. \end{cases}$$

  Le point terminal x_F est fini et S est à décroissance polynomiale.

- Si $\gamma = 0$, on dit que F appartient au domaine d'attraction de *Gumbel* (noté $F \in \mathcal{DA}(\text{Gumbel})$) et la distribution GEV est donnée par :

$$G_\gamma^{\text{G}}(x) = \exp\left[-\exp(-x)\right] \quad \text{pour tout } x \in \mathbb{R}.$$

  Le point terminal x_F peut être fini ou infini et S est à décroissance exponentielle.

La Figure 1.1 présente trois exemples de densités associées à la distribution GEV, obtenues par simple dérivation de la fonction G_γ.

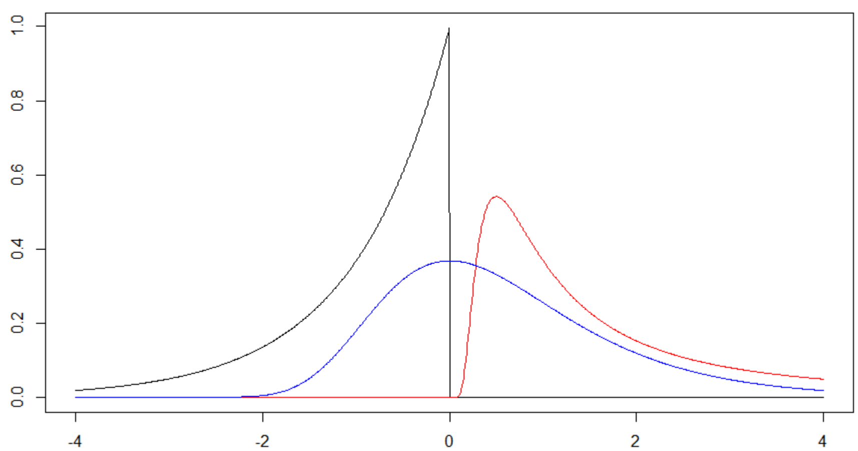


Figure 1.1: Densité de la distribution GEV pour $\gamma = 1$ (courbe rouge), $\gamma = 0$ (courbe bleue) et $\gamma = -1$ (courbe noire).

1.1.2 Caractérisation des domaines d'attraction

Une question naturelle est de se demander quelles conditions doivent vérifier les distributions pour qu'elles appartiennent à un domaine d'attraction et si c'est le cas, à quel des trois domaines. L'étude des fonctions à *variations régulières à l'infini* joue un grand rôle dans cette caractérisation. Une étude approfondie de cette notion se trouve dans le livre de Bingham et al. [4].

Definition 2 *Une fonction mesurable* $f : \mathbb{R} \longrightarrow [0, +\infty)$ *est dite à variations régulières à l'infini d'indice* $\alpha \in \mathbb{R}$ *(et on note* $f \in \mathcal{RV}(\alpha)$*), si* $\forall t > 0$

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

Dans le cas particulier où $\alpha = 0$*, on dit que* f *est à variations lentes à l'infini.*

L'étude des fonctions à variations régulières se ramène à étudier uniquement les fonctions à variations lentes. En effet, il est facile de montrer la proposition suivante.

Proposition 1 *Pour toute fonction* $f : \mathbb{R} \longrightarrow [0, +\infty)$ *appartenant à* $\mathcal{RV}(\alpha)$ *avec* $\alpha \in \mathbb{R}$*, alors pour tout* $x \in \mathbb{R}$*,*

$$f(x) = x^{\alpha} L(x)$$

où L *est une fonction à variations lentes à l'infini.*

Comme exemples simples de fonctions à variations lentes à l'infini, on peut citer les fonctions constantes, les fonctions qui convergent vers une constante ou encore la fonction logarithme. D'autres exemples sont les fonctions appartenant à la *classe de Hall* introduite en 1982. Les fonctions L de cette classe sont telles que

$$\exists M > 0,\ \forall x \geq M,\ L(x) = C + Dx^{\beta}(1 + o(1))$$

où $o(1)$ converge vers 0 quand x tend vers l'infini avec $C > 0$, $\beta \leq 0$ et $D \in \mathbb{R}^*$. On peut également définir des fonctions à variations rapides où l'on considère les cas $\alpha = +\infty$ ou $\alpha = -\infty$.

Definition 3 *Une fonction mesurable* $f : \mathbb{R} \longrightarrow [0, +\infty)$ *est dite à variations rapides d'indice* $-\infty$ *(resp.* $+\infty$*) si elle est positive et si*

$$\lim_{x \to \infty} \frac{f(tx)}{f(x)} = \begin{cases} +\infty \quad (resp.\ 0) & si\ 0 \leq t < 1, \\ 0 \quad (resp.\ +\infty) & si\ t > 1. \end{cases}$$

Comme exemples de fonctions à variations rapides, on la fonction $x \mapsto \exp(x)$ avec $\alpha = +\infty$ et la fonction $x \mapsto \exp(-x)$ avec $\alpha = -\infty$.
On peut aussi s'intéresser à l'inverse généralisé d'une fonction à variations régulières à l'infini qui est croissante ou décroissante.

Proposition 2 *Si $f \in \mathcal{RV}(\alpha)$ avec $\alpha \geq 0$ et est une fonction croissante telle que $f(x) \to +\infty$ lorsque $x \to \infty$, alors $f^{\leftarrow} \in \mathcal{RV}(1/\alpha)$.*

Corollary 1 *Si $f \in \mathcal{RV}(\alpha)$ avec $\alpha < 0$ est une fonction décroissante, alors la fonction $f^{\leftarrow}(1/\cdot)$ est à variations régulières à l'infini d'indice $-1/\alpha$.*

Enfin, on peut différencier les trois domaines d'attraction de la manière suivante.

Domaine de Fréchet et de Weibull

En utilisant la théorie des fonctions à variations régulières, on peut obtenir un critère utile pour déterminer si une fonction de répartition F appartient à un domaine d'attraction.

Theorem 2 *Soit $\gamma \neq 0$, F une fonction de répartition, S la fonction de survie associée et x_F le point terminal de F.*

- $F \in \mathcal{DA}($Fréchet) si et seulement si $x_F = +\infty$ et

  $$S(x) = x^{-1/\gamma} L(x)$$

  ou, de manière équivalente,

  $$Q(\alpha) = \alpha^{-\gamma} \ell(\alpha^{-1}), \quad \alpha \in [0, 1]$$

  où L et ℓ sont des fonctions à variations lentes à l'infini et $Q(\alpha) = S^{\leftarrow}(\alpha)$. Des suites possibles de normalisation sont $a_n = Q(1/n)$ et $b_n = 0$.

- $F \in \mathcal{DA}$(Weibull) si et seulement si $x_F < \infty$ et si

  $$S^*(x) = x^{1/\gamma} L(x)$$

  ou, de manière équivalente,

  $$Q(\alpha) = x_F - \alpha^{-\gamma} \ell(\alpha^{-1})$$

  où L et ℓ sont des fonctions à variations lentes à l'infini et où S^* est une fonction de survie telle que la fonction de répartition associée F^* est définie par

  $$F^*(x) := \begin{cases} 1 & \text{si } x < 0 \\ 1 - F(x_F - 1/x) & \text{si } x \geq 0. \end{cases}$$

  Des suites possibles de normalisation sont $a_n = x_F - Q(1/n)$ et $b_n = x_F$.

On peut remarquer que les domaines de Fréchet et de Weibull sont étroitement liés. En effet, on peut passer de l'un à l'autre par un simple changement de variable. Supposons qu'une variable aléatoire Y ait pour fonction de répartition $F \in \mathcal{DA}$(Fréchet) et soit x_F un réel fixé. Alors, la fonction de répartition de la variable $x_F - 1/Y$ appartient à $\mathcal{DA}$(Weibull) et son point terminal est x_F. Inversement, si $F \in \mathcal{DA}$(Weibull) avec pour point terminal x_F, alors la fonction de répartition de $(x_F - Y)^{-1}$ appartient à $\mathcal{DA}$(Fréchet). Voici quelques exemples de loi :

- Domaine d'attraction de Fréchet : Fréchet, Cauchy, Burr, Pareto, Log-Gamma, Student

- Domaine d'attraction de Weibull : Uniforme, Beta.

Domaine de Gumbel

La caractérisation du domaine de Gumbel est plus compliquée. Il n'existe pas de représentations simples. Un sous-ensemble intéressant de distributions sont les lois de *type Weibull*, c'est-à-dire les distributions définies par :

$$S(x) = \exp(-x^\theta L(x))$$

où L est une fonction à variations lentes à l'infini et $\theta > 0$ est l'indice de queue de Weibull. Un exemple de suite de normalisation est $b_n = Q(1/n)$. La suite a_n est complexe à obtenir. Un grand nombre de distributions sont de ce type là. On peut citer la loi Normale, Log-Normale, Weibull, Gamma et Exponentielle.
Cependant, il est possible de caractériser les fonctions de répartition en terme de fonctions à variations rapides d'indice $-\infty$.

Proposition 3 *Soit F une fonction de répartition appartenant au domaine d'attraction de Gumbel de point terminal x_F et soit S la fonction de survie associée.*

- *si $x_F = +\infty$ alors S est une fonction à variations rapides d'indice $-\infty$.*
- *si $x_F < +\infty$ alors $S(x_F - 1/\cdot)$ est une fonction à variations rapides d'indice $-\infty$.*

De nombreux exemples supplémentaires de lois appartenant à des domaines d'attraction se trouvent dans Embrechts et al. [12] tableaux 3.4.2, 3.4.3 et 3.4.4.

Remarque 1 *Trouver un domaine d'attraction pour une fonction de répartition F revient en fait à définir une suite $u_n = u_n(x) = a_n x + b_n$ telle que $\mathbb{P}(M_n \leq u_n)$ converge vers une limite qui n'est pas toujours égale à 0 ou 1. Comme il est expliqué en détail dans Leadbetter et al. [30] et plus succinctement dans Embrechts et al. [12], cela exige certaines conditions de régularité sur F. Supposons que x_F est fini. Il est alors impossible de trouver une telle suite u_n pour une fonction de répartition F qui présente un "saut" en x_F, c'est-à-dire si*

$$\lim_{x \to x_F -} F(x) \neq 1,$$

ce qui est le cas pour la loi binomiale par exemple. Il existe d'autre critères qui permettent de décider si une fonction de répartition appartient à un domaine d'attraction, peu importe que x_F soit fini ou non (voir Leadbetter et al. [30] Corollaire 1.5.2 et Théorème 1.7.13).

1.1.3 Distributions super heavy-tailed

Une autre classe de fonctions, que l'on appelle *classe des distributions super heavy-tailed* ne présente pas les caractéristiques nécessaires pour appartenir à un domaine d'attraction. Une distribution super heavy-tailed est une distribution dont la fonction de survie S est une fonction à variations lentes à l'infini, c'est-à-dire pour tout $t > 0$, on a

$$\lim_{x \to +\infty} \frac{S(tx)}{S(x)} = 1.$$

On rappelle que les fonctions de survie des distributions appartenant au domaine de Fréchet ($\gamma > 0$) sont des fonctions à variations régulières à l'infini d'indice $-1/\gamma$. Par abus de langage, une distribution super heavy-tailed peut être considérée comme étant une distribution à queue lourde ayant comme indice des valeurs extrêmes $\gamma = +\infty$. Comme exemples, on peut citer les lois :

- Log-Pareto : $S(x) = [\ln(x)]^{-\xi}$, $x > 0$, $\xi > 0$;

- Log-Weibull : $S(x) = \exp[-\xi \ln^{\theta}(x)]$, $x > 0$, $\xi > 0$;

- Log-Cauchy : $S(x) = \frac{1}{2} - \frac{1}{\pi}\text{arctan}\left(\frac{\ln(x)-\mu}{\xi}\right)$, $x > 0$, $\mu \in \mathbb{R}$, $\xi > 0$.

Cela concerne toutes les lois où S est à décroissance logarithmique. La Figure 1.2 compare l'allure d'une loi Log-Pareto avec trois lois appartenant à un domaine d'attraction.

La majorité des résultats en théorie des valeurs extrêmes concerne les distributions vérifiant le théorème fondamental érigé par Fisher, Tippett et Gnedenko. Il peut alors être intéressant de tenter une nouvelle approche afin que les résultats puissent se généraliser à cette classe de distributions.

1.2 Estimation de quantiles extrêmes

1.2.1 Définition

On rappelle qu'un quantile d'ordre $\alpha \in \,]0,1[$ est défini par l'inverse généralisé d'une fonction de survie S. Autrement dit, si on désigne par $Q(\alpha)$ le quantile,

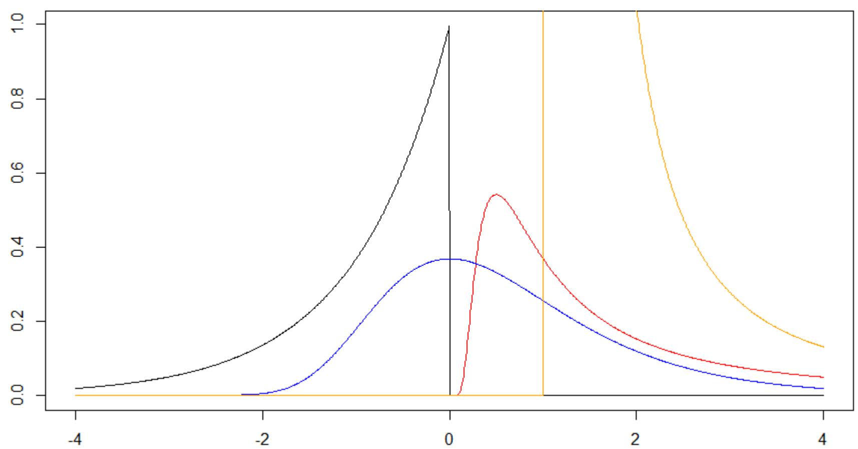


Figure 1.2: Densité de la loi GEV pour $\gamma = 1$ (courbe rouge), $\gamma = 0$ (courbe bleue), $\gamma = -1$ (courbe noire) et d'une loi Log-Pareto (courbe orange).

alors

$$Q(\alpha) = S^{\leftarrow}(\alpha) = \inf \{y \in \mathbb{R} \mid S(y) \leq \alpha\}$$

avec la convention $\inf \{\varnothing\} = +\infty$. On remarque que le quantile d'ordre 1 est égal à x_F, le point terminal de F. L'objectif, en théorie des valeurs extrêmes, est d'estimer les quantiles $Q(\alpha)$ où α est proche de 0. Une des méthodes classiques est la méthode dite indirecte qui consiste à définir tout d'abord un estimateur de la fonction de survie S et en déduire, par la relation d'inverse généralisé, un estimateur de Q. Comme estimateur de S, on pense naturellement à l'estimateur empirique : pour tout $y \in \mathbb{R}$

$$\check{S}_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{Y_i > y\}}$$

et un estimateur possible de $Q(\alpha)$ est donc

$$\check{Q}_n(\alpha) = \sum_{i=1}^{n} Y_{n-i+1,n} \mathbb{I}_{\left\{\alpha \,\in\, \left[\frac{i-1}{n}, \frac{i}{n}\right[\right\}} = Y_{n-\lfloor n\alpha \rfloor, n}$$

où $\lfloor x \rfloor$ désigne la partie entière de x et $Y_{1,n} \leq \ldots \leq Y_{n,n}$ désigne l'échantillon rangé par ordre croissant. On se rend compte rapidement de la limite de cette méthode. En effet, dès que $\alpha < 1/n$, notre estimateur est "bloqué" à la statistique $Y_{n,n}$. Ce qui n'apporte pas de réponse satisfaisante. La zone $[X_{n,n}, +\infty)$ est appelée la queue de distribution de S et un quantile appartenant à la queue de distribution est appelé un quantile extrême. La notion de quantile dépend donc de l'échantillon et de sa taille n. L'intérêt se porte alors sur les quantités $Q(\alpha_n)$ où α_n est une suite qui converge vers 0. On classifie les quantiles en deux catégories.

Definition 4 *Soit α_n une suite qui converge vers 0.*

- *On dit d'un quantile qu'il est intermédiaire si $n\alpha_n \to +\infty$.*
- *On dit d'un quantile qu'il est extrême si $n\alpha_n \to c \geq 0$.*

1.2.2 Estimation d'un quantile extrême par extrapolation

La question qui se pose est comment peut-on estimer des tels quantiles si les outils de la statistique classique ne fonctionnent pas ? Une des réponses possibles se trouve dans l'extrapolation de données. Considérons deux suites α_n et β_n convergeant vers 0 telles que $\beta_n/\alpha_n \to 0$ quand n tend vers l'infini. L'extrapolation consiste à définir un estimateur de quantiles intermédiaire $\tilde{Q}_n(\alpha_n)$ et d'en déduire un estimateur de quantiles extrêmes $\tilde{Q}_n^{(E)}(\beta_n)$ par une relation qui les lie.

Une approche non paramétrique

Dans de Haan et Ferreira [25] Théorème 1.1.6, il est stipulé que F appartient à un domaine d'attraction si et seulement s'il existe une fonction positive a telle que pour tout $u > 0$,

$$\lim_{\alpha \to 0} \frac{Q(u\alpha) - Q(\alpha)}{a(1/\alpha)} = \int_1^{1/u} s^{\gamma-1} \mathrm{d}s =: L_\gamma(1/u). \tag{1.1}$$

Cette relation s'appelle *relation du premier ordre.* En remplaçant u par β_n/α_n, on obtient l'approximation

$$Q(\beta_n) \approx Q(\alpha_n) + a(1/\alpha_n)L_\gamma(\alpha_n/\beta_n).$$

Ainsi, notre estimateur de quantiles extrêmes est défini par :

$$\tilde{Q}_n^{(E)}(\beta_n) := \tilde{Q}_n(\alpha_n) + \tilde{a}_n(1/\alpha_n)L_{\tilde{\gamma}_n}(\alpha_n/\beta_n) \tag{1.2}$$

où $\tilde{Q}_n(\alpha_n)$, $\tilde{a}_n(1/\alpha_n)$ et $\tilde{\gamma}_n$ sont des estimateurs de $Q(\alpha_n)$, $a(1/\alpha_n)$ et γ. Plusieurs conditions sont nécessaires sur ces estimateurs afin d'obtenir une normalité asymptotique pour $\tilde{Q}_n^{(E)}(\beta_n)$. En premier lieu, une condition du *second ordre* sur la relation (1.1) : il existe une fonction A qui ne change pas de signe et qui vérifie $A(t) \to 0$ quand $t \to \infty$ telle que

$$\lim_{\alpha \to 0} \frac{\frac{Q(u\alpha)-Q(\alpha)}{a(1/\alpha)} - L_\gamma(1/u)}{A(1/\alpha)} = \int_1^{1/u} r^{\gamma-1} \int_1^r t^{\rho-1} \mathrm{d}t \, \mathrm{d}r \tag{1.3}$$

où $\rho \leq 0$ est appelé le paramètre de second-ordre. On a également besoin d'une condition de convergence en loi du triplet suivant :

$$\sqrt{n\alpha_n}\left(\tilde{\gamma}_n - \gamma, \frac{\tilde{a}_n(1/\alpha_n)}{a(1/\alpha_n)} - 1, \frac{\tilde{Q}_n(\alpha_n) - Q(\alpha_n)}{a(1/\alpha_n)}\right) \xrightarrow{\mathrm{d}} (\Gamma, \Lambda, \Theta) \tag{1.4}$$

où le triplet $(\Gamma, \Lambda, \Theta)$ suit une loi normale multidimensionnelle dont le vecteur espérance peut dépendre de γ et ρ et dont la matrice de covariance ne peut dépendre que de γ. Ainsi, on peut énoncer le théroème suivant.

Theorem 3 (de Haan et Ferreira [25] Théroème 4.3.1) *Soient deux suites α_n et β_n convergeant vers 0 telles que $\beta_n/\alpha_n \to 0$. Supposons qu'il existe une fonction A ne changeant pas de signe telle que $A(t) \to 0$ quand $t \to \infty$ et que (1.3) soit satisfaite. Si on a les conditions suivantes :*

- *le paramètre ρ est négatif ou égal à 0 avec γ négatif;*
- *$n\alpha_n \to \infty$, $\ln(\alpha_n/\beta_n)/\sqrt{n\alpha_n} \to 0$ et $\sqrt{n\alpha_n}A(1/\alpha_n) \to \lambda \in \mathbb{R}$ quand n tend vers l'infini;*
- *la condition (1.4) est vérifiée pour des estimateurs pertinents de $Q(\alpha_n)$, $a(1/\alpha_n)$ et γ;*

alors,

$$\sqrt{n\alpha_n}\,\frac{\tilde{Q}_n^{(E)}(\beta_n)-Q(\beta_n)}{a(1/\alpha_n)q_\gamma(\alpha_n/\beta_n)} \xrightarrow{\mathrm{d}} \Gamma+(\gamma_-)^2\Theta-\gamma_-\Lambda-\lambda\frac{\gamma_-}{\gamma_-+\rho}$$

où $\gamma_- := \min(0,\gamma)$ *et où pour tout* $t>1$*,*

$$q_\gamma(t) := \int_1^t s^{\gamma-1}\ln(s)\mathrm{d}s.$$

On peut remarquer que cette méthode d'extrapolation présente des limites. En effet, La condition $\ln(\alpha_n/\beta_n)/\sqrt{n\alpha_n} \to 0$ quand n tend vers l'infini impose que pour tout $\epsilon>0$ on a $\beta_n > n^{-1}\exp(-\epsilon\sqrt{n\alpha_n})$ pour n suffisamment grand. L'ordre du quantile extrême ne peut donc être arbitrairement petit. En ce qui concerne le choix des estimateurs, un des estimateurs classiques de quantiles intermédiaires $Q(\alpha_n)$ est tout simplement l'estimateur empirique $Y_{n-\lfloor n\alpha_n\rfloor,n}$. Dans de Haan et Ferreira [25] Théorème 2.2.1, il est montré que cet estimateur est asymptotiquement gaussien.

Theorem 4 (de Haan et Ferreira [25] Théorème 2.2.1) *Supposons que la condition du second ordre* (1.3) *est satisfaite pour* $\gamma\in\mathbb{R}$ *et* $\rho\le 0$*. Soit* α_n *une suite convergeant vers* 0 *telle que* $n\alpha_n\to+\infty$ *quand* $n\to+\infty$*. Si*

$$\lim_{n\to+\infty}\sqrt{n\alpha_n}A(1/\alpha_n) < +\infty$$

alors

$$\sqrt{n\alpha_n}\frac{Y_{n-\lfloor n\alpha_n\rfloor,n}-Q(\alpha_n)}{a(1/\alpha_n)} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Pour le choix de $\tilde{\gamma}_n$ et $\tilde{a}_n(1/\alpha_n)$, un exemple est donné dans de Haan et Ferreira [25] Section 4.3.2 avec l'estimateur des moments, qui est un estimateur général de $\gamma\in\mathbb{R}$. Il existe cependant des cas plus simples.

Une approche semi-paramétrique simplifiée pour le cas $\gamma>0$

Si l'on suppose au préalable que $\gamma>0$, c'est-à-dire que $F\in\mathcal{DA}$(Fréchet), alors on peut obtenir un estimateur de quantiles extrêmes plus simple à étudier. On a pu voir que la fonction de survie associée S s'écrivait :

$$S(x) = x^{-1/\gamma}L(x)$$

où L est une fonction à variations lentes à l'infini. De manière équivalente, on peut aussi écrire cette relation en termes de quantiles, c'est-à-dire, pour tout $\alpha \in [0,1]$,

$$Q(\alpha) = \alpha^{-\gamma}\ell(\alpha^{-1})$$

où ℓ est une fonction à variations lentes à l'infini. Si l'on considère deux suites α_n et β_n qui convergent vers 0 telles que $\beta_n/\alpha_n \to 0$, alors par la définition de fonctions à variations lentes à l'infini on obtient

$$Q(\beta_n) \approx Q(\alpha_n)\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma}.$$

Un estimateur possible de quantiles extrêmes est celui de Weissman (1978) défini par

$$\tilde{Q}^{(E)}_{n,\text{pos}}(\beta_n) = Y_{n-\lfloor n\alpha_n\rfloor,n}\left(\frac{\alpha_n}{\beta_n}\right)^{\hat{\gamma}_n}$$

et étudié dans de Haan et Ferreira [25] Théorème 4.3.8.

Theorem 5 (de Haan et Ferreira [25] Théorème 4.3.8) *Soient deux suites α_n et β_n convergeant vers 0 telle que $n\alpha_n \to +\infty$ et $\beta_n/\alpha_n \to 0$. Supposons qu'il existe une fonction A telle que $A(t) \to 0$ quand $t \to +\infty$ et*

$$\lim_{\alpha\to 0}\frac{\frac{Q(u\alpha)}{Q(\alpha)} - u^{-\gamma}}{A(1/\alpha)} = u^{-\gamma}\frac{u^{-\rho}-1}{\rho}.$$

Si on a les conditions :

- *le paramètre ρ est strictement négatif;*
- $\sqrt{n\alpha_n}A(1/\alpha_n) \to \lambda \in \mathbb{R}$ *et* $\ln(\alpha_n/\beta_n)/\sqrt{n\alpha_n}$ *quand* $n \to +\infty$*;*
- $\sqrt{n\alpha_n}(\tilde{\gamma}_n - \gamma) \xrightarrow{\text{d}} \Gamma$ *où* Γ *suit une loi normale;*

alors,

$$\frac{\sqrt{n\alpha_n}}{\ln(\alpha_n/\beta_n)}\left(\frac{\hat{Q}^{(E)}_{n,pos}(\beta_n)}{Q(\beta_n)} - 1\right) \xrightarrow{\text{d}} \Gamma.$$

Comme exemple d'estimateur de $\gamma > 0$, on peut citer l'estimateur de Hill introduit en 1975. Une étude détaillée est faite dans de Haan et Ferreira [25].

1.3 Contexte de la thèse et résumés des chapitres

1.3.1 Contexte de la thèse et problématiques

On considère un couple de variables aléatoires $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$ avec $p \in \mathbb{N} \setminus \{0\}$ et soit $x_0 \in \text{supp}(X)$. On pose $S(\cdot|x_0)$ et $Q(\cdot|x_0)$ la fonction de survie et la fonction quantile associée de la distribution de Y sachant l'évènement $\{X = x_0\}$. On suppose que (X, Y) admet une densité de probabilité et note par f la densité marginale de X. L'objectif de la thèse est d'étudier des estimateurs de la fonction de survie conditionnelle qui dépendent de *fonctions poids* pour en déduire des estimateurs de quantiles extrêmes conditionnels par la méthode indirecte. Considérons n copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ de (X, Y). Les estimateurs de $S(\cdot|x_0)$ que nous étudions sont de la forme :

$$\hat{S}_n(y|x_0) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \mathbb{I}_{\{Y_i > y\}} \tag{1.5}$$

pour $y \in \mathbb{R}$ et où l'ensemble des poids $\{\mathcal{W}_{n,i}(x_0), 1 \leq i \leq n\}$ est un tableau triangulaire de variables aléatoires positives dépendant des données $X_1, \ldots, X_n$ et de x_0 et vérifiant

$$\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) = 1.$$

Le rôle des poids $\mathcal{W}_{n,i}(x_0)$ est de sélectionner les covariables qui sont le plus "proches" de x_0 afin que l'information utilisée pour une telle estimation soit la plus pertinente possible. Ce genre d'estimateurs est plus communément appelé *estimateur local*. On obtient naturellement l'estimateur de quantiles :

$$\hat{Q}_n(\alpha|x_0) := \hat{S}_n^{\leftarrow}(\alpha|x_0) \tag{1.6}$$

pour $\alpha \in]0, 1[$. Nous pouvons déjà citer deux exemples fondamentaux qui font partie des estimateurs locaux.

- L'estimateur de Nadaraya-Watson introduit par Nadaraya (1964) et Watson (1964) où les poids sont définis par

$$\mathcal{W}_{n,i}^{\text{NW}}(x_0) := K\left(\frac{X_i - x_0}{h_n}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{X_j - x_0}{h_n}\right)$$

avec K une fonction noyau et h_n une suite positive de réels qui converge vers 0 quand n tend vers l'infini et que l'on appelle *fenêtre*. Il s'agit d'un paramètre de *lissage*. Cette suite contrôle la convergence de l'estimateur ainsi que l'équilibre biais-variance. Une grande fenêtre s'accompagne d'une faible variance (la moyenne étant effectuée sur plus d'observations) mais d'un biais plus important (la fonction étant presque constante). Une petite fenêtre aura l'effet inverse. Il n'est pas possible de déterminer exactement quelle suite h_n permet d'avoir des résultats optimaux. Cependant, il est usuel d'utiliser une méthode de validation croisée pour obtenir un choix optimal de h_n. Le choix de la fonction noyau K n'est pas crucial (voir Davison et al. [7]). Les fonctions noyau classiques sont le noyau Uniforme, Triangle, Epanechnikov, Quadratique ou encore Gaussien.

- L'estimateur des k_n plus proches voisins : Soit k_n une suite d'entiers et soit $||\cdot||$ une norme sur $\mathbb{R}^p$. La méthode des k_n plus proches voisins consiste à sélectionner les k_n observations X_i qui sont les plus proches en termes de distances de x_0. Si on désigne par $r(i)$, $1 \leq i \leq k_n$ le rang de $||X_i - x_0||$, alors l'estimateur des k_n plus proches voisins est défini pour $\ell \in \mathbb{N}$,

$$\mathcal{W}_{n,i}^{\mathrm{KNN}}(x_0) := [(k_n - r(i) + 1)_+]^\ell \Bigg/ \sum_{j=1}^{k_n} j^\ell .$$

  où $(x)_+ = \max(0, x)$. Lorsque $\ell = 0$, il s'agit de l'estimateur des k_n plus proches voisins uniforme. Le même poids est assigné aux k_n observations sélectionnées. Pour $\ell \neq 0$, plus l'observation X_i est proche de x_0, plus on lui assignera un poids important. Pour $\ell = 1$ (resp. $\ell = 2$), on dit qu'il s'agit de l'estimateur des k_n plus proches voisins triangulaire (resp. quadratique).

L'objectif reste l'estimation de quantiles extrêmes. Soient donc α_n et β_n deux suites qui convergent vers 0 telles que $\beta_n/\alpha_n \to 0$. Deux cas de figure se présentent.

- Supposons que $S(\cdot|x_0)$ appartiennent à un domaine d'attraction, c'est-à-dire qu'il existe des suites de réels $a_n > 0$ et b_n et une fonction de répartition non-dégénérée G tels que

$$\lim_{n\to+\infty}[1 - S(a_n z + b_n|x_0)]^n = G(z|x_0)$$

  pour tout point de continuité z de $G(\cdot|x_0)$. Alors, d'après de Haan et Ferreira [25] Théorème 1.1.6, il existe une fonction positive $a(\cdot|x_0)$ telle que pour tout $u > 0$,

$$\lim_{\alpha\to 0}\frac{Q(u\alpha|x_0) - Q(\alpha|x_0)}{a(1/\alpha|x_0)} = L_{\gamma(x_0)}(1/u)$$

  où $\gamma(x_0)$ est l'indice des valeurs extrêmes conditionnel associé. On a alors l'estimateur de quantiles extrêmes :

$$\hat{Q}_n^{(\mathrm{E})}(\beta_n|x_0) := \hat{Q}_n(\alpha_n|x_0) + \tilde{a}_n(1/\alpha_n|x_0)L_{\tilde{\gamma}_n(x_0)}(\alpha_n/\beta_n)$$

  où $\tilde{a}_n(1/\alpha_n|x_0)$ et $\tilde{\gamma}_n(x_0)$ sont des estimateurs de $a(1/\alpha_n|x_0)$ et $\gamma(x_0)$ encore à déterminer.

- Supposons que $S(\cdot|x_0)$ n'appartienne pas forcément à un domaine d'attraction. Est-il alors possible de trouver une relation asymptotique permettant de définir des quantiles extrêmes dans un cadre plus général ? Par exemple pour des distributions super heavy-tailed ?

Enfin, nous pouvons aussi nous interroger sur le problème de la dimension de la covariable X. Il est bien connu que plus la dimension des variables aléatoires augmente, plus les méthodes d'estimations sont difficiles. Il s'agit du *fléau de la dimension* ou *curse of dimensionality* en anglais. Lorsque la dimension est grande, les observations sont éparpillées dans l'espace et leurs positions sont très éloignées de celle de x_0. C'est ce que nous pouvons observer sur la Figure 1.3. Nous simulons 400 variables aléatoires suivant une loi uniforme standard et regardons quelles sont les 100 observations les plus proches de $x_0 = 1/2$ (dimension 1), $x_0 = (1/2, 1/2)$ (dimension 2) et $x_0 = (1/2, 1/2, 1/2)$ (dimension 3).
Nous pouvons constater que plus la dimension augmente, plus il est difficile de contrôler la distance pour qu'il y ait un nombre suffisant d'observations prises en compte. Cela peut poser problème par exemple pour l'estimateur

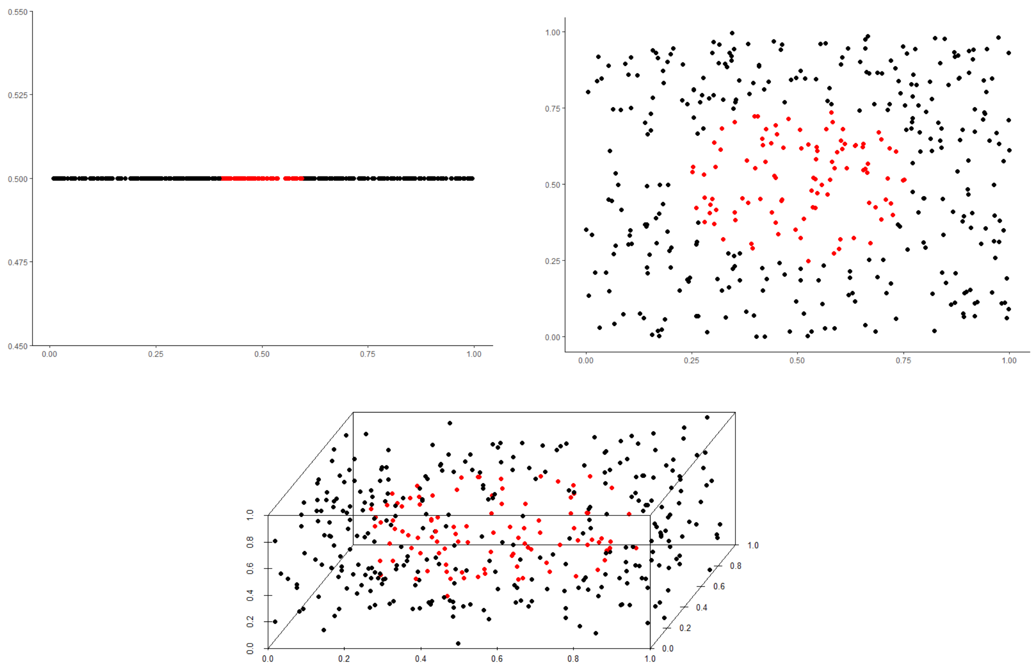


Figure 1.3: Simulations de 400 variables aléatoires suivant une loi uniforme standard en dimension 1, 2 et 3. En rouge : les 100 plus proches observations des points $x_0 = 1/2$ (dimension 1), $x_0 = (1/2, 1/2)$ (dimension 2) et $x_0 = (1/2, 1/2, 1/2)$ (dimension 3).

de Nadaraya-Watson où la suite h_n ne peut plus être correctement choisie. Cela engendre des problèmes de biais importants et il est donc important de réduire la dimension sans perdre de l'information. La méthode d'Analyse en Composantes Principales (ACP) est l'une des méthodes de réduction de dimension les plus connues ou encore la méthode de Régression Inverse par Tranchage (Sliced Reverse Regression (SIR) en anglais) introduite dans Li [31]. L'idée en général est de considérer une matrice B de taille $p \times q$ avec $q < p$ et de rang maximal telle que la distribution de Y sachant $B^\top X$ soit la même que celle de Y sachant X. On dit alors que X et Y sont indépendantes conditionnellement à $B^\top X$. Ainsi, on pourrait remplacer les observations X_i par $B^\top X_i$ sans perdre de l'information et obtenir des estimateurs plus performants puisque nos nouvelles covariables ont une dimension moindre $(q < p)$. Il serait donc intéressant de voir s'il est possible de définir un estimateur de quantiles extrêmes définit en fonction de nouvelles observations $B^\top X_i$ qui soit plus efficace que l'estimateur standard, c'est-à-dire où aucune méthode de réduction de dimension n'est appliquée.

1.3.2 Résumés des chapitres

Chapitre 2

Dans le chapitre 2, on introduit une nouvelle condition semblable à la condition du premier ordre (1.1) mais qui est plus générale. Il s'agit de la *Tail First Order Condition* (TFO). Soit $I \in \mathbb{R}$ et $J \in \mathbb{R}$ deux intervalles ouverts contenant 0 et y^* le point terminal de S. Une distribution S vérifie la condition TFO s'il existe des fonctions positives d et Ψ telles que pour tout $t \in I$,

$$\lim_{y \to y^*} \Psi(y) \left(\frac{S[y + td(y)]}{S(y)} - 1 \right) = \phi^{-1}(t),$$

où ϕ^{-1} est l'inverse d'une fonction $\phi : J \to I$ continue et strictement décroissante telle que $\phi(t)/t \to -1$ quand $t \to 0$. Une version équivalente en terme de quantiles est la suivante. Une distribution S vérifie la condition TFO s'il existe des fonctions positives a et g telles que

$$\lim_{\alpha \to 0} \frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(1/\alpha)} = \phi(t).$$

où $Q(\alpha) = S^{\leftarrow}(\alpha)$. On remarque que si g est l'identité et que $\phi(t) = L_\gamma(1/(t+1))$, alors on retrouve la condition du premier ordre (1.1). Les distributions appartenant à un domaine d'attraction vérifient alors la condition TFO. On exposera dans ce chapitre les motivations et les particularités de cette condition. On montrera également que les distributions super heavy-tailed vérifient la condition TFO. En fixant $x_0 \in \mathbb{R}^p$, $p \in \mathbb{N} \setminus \{0\}$, on peut évidemment donner une version conditionnelle de la condition TFO avec $J = J_{x_0}$, $\phi(\cdot) = \phi(\cdot|x_0)$, $\Psi = \Psi_{x_0}$ et $d = d_{x_0}$, ou $g(\cdot) = g(\cdot|x_0)$ et $a(\cdot) = a(\cdot|x_0)$.
On s'intéresse ensuite à l'estimateur $\hat{Q}(\alpha_n|x_0)$ défini en (1.6) avec α_n une suite qui converge vers 0. Le but est de montrer sa normalité asymptotique sous certaines conditions. Une première étape consiste à montrer la normalité asymptotique de $\hat{S}_n(y_n(x_0)|x_0)$ défini en (1.5) où $y_n(x_0)$ est une suite qui converge vers le point terminal de $S(\cdot|x_0)$. Il est nécessaire d'introduire

une telle suite $y_n(x_0)$ afin d'obtenir des renseignements sur la queue de distribution. Soient deux suites $\sigma_n(x_0)$ et $v_n(x_0)$ convergeant vers $+\infty$ quand $n \to +\infty$ et qui dépendent de x_0. Alors, pour tout $z \in \mathbb{R}$,

$$\mathbb{P}\left[\sigma_n(x_0)\left(\hat{Q}(\alpha_n|x_0) - Q(\alpha_n|x_0)\right) \leq z\right] = \mathbb{P}[Z_n(x_0) \leq z_n(x_0)],$$

où

$$Z_n(x_0) := v_n(x_0)\left[\hat{S}_n(Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)|x_0) - S(Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)|x_0)\right]$$

et

$$z_n(x_0) := -\alpha_n v_n(x_0)\left[\frac{S(Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)|x_0)}{\alpha_n} - 1\right].$$

En supposant que $S(Q(\alpha|x_0)) = \alpha$ pour α suffisamment petit, que S vérifie la condition TFO et en posant $y_n(x_0) = Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)$, $\sigma_n^{-1}(x_0) = \alpha_n v_n(x_0)/[\Psi(Q(\alpha_n|x_0)|x_0)d(Q(\alpha_n|x_0)|x_0)]$ et $t_n^{-1}(x_0) := \sigma_n(x_0)d[Q(\alpha_n|x_0)]$, alors on peut montrer que $z_n(x_0) \to z$ quand $n \to \infty$. Ainsi, si on a la convergence

$$v_n(x_0)[\hat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0)] \xrightarrow{\mathrm{d}} \mathcal{N}(0,1), \tag{1.7}$$

on peut en déduire la normalité asymptotique

$$\sigma_n(x_0)[\hat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)] \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Pour montrer une telle convergence en loi (1.7), nous nous sommes inspirés des travaux de Owen dans sa thèse [34] qui étudie les mêmes estimateurs mais dans le cas classique, c'est-à-dire les estimateurs $\hat{S}_n(y|x_0)$ pour $y \in \mathbb{R}$. Nous avons adapté ses méthodes de démonstration au cas extrême. L'idée est d'introduire la quantité

$$n_{x_0} := \left(\sum_{i=1}^{n} \mathcal{W}_{n,i}^2(x_0)\right)^{-1}$$

qui correspond approximativement au nombre de covariables prises en compte dans l'estimation. On a alors le théorème suivant.

Theorem 6 *Soit $x_0 \in \mathbb{R}^p$ tel que $f(x_0) > 0$ et soit $y_n(x_0)$ une suite convergeant vers le point terminal $y^*(x_0)$ de la distribution de Y sachant $\{X =$*

$x_0\}$. On suppose qu'il existe une suite $m_n(x_0)$ telle que $n_{x_0}/m_n(x_0) \xrightarrow{\text{p.s}} 1$ et on pose $v_n^2(x_0) := m_n(x_0)/S(y_n(x_0)|x_0)$. Si on a les conditions

$$v_n(x_0) \max_{1\le i\le n} \mathcal{W}_{n,i}(x_0) \xrightarrow{\text{a.s}} 0 \tag{1.8}$$

et

$$v_n(x_0) \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)|S(y_n(x_0)|X_i) - S(y_n(x_0)|x_0)| \xrightarrow{\text{d}} \mathcal{N}(0,1) \tag{1.9}$$

alors

$$v_n(x_0)\left(\hat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0)\right) \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

L'idée de la preuve est de définir pour $i = 1, \ldots, n$ les variables $Y_i^{x_0} := Q(U_i|x_0)$ où $U_1, \ldots, U_n$ est une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi uniforme standard et indépendantes des X_i. Ainsi,

$$\begin{aligned}\hat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0) &= \left[\hat{S}_n^{x_0}(y_n(x_0)) - S(y_n(x_0)|x_0)\right] \\ &\quad + \left[\hat{S}_n(y_n(x_0)|x_0) - \hat{S}_n^{x_0}(y_n(x_0))\right]\end{aligned}$$

où pour tout $y \in \mathbb{R}$,

$$\hat{S}_n^{x_0}(y) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{\{Y_i^{x_0}>y\}}.$$

Le premier terme correspond au *terme de variance* et le deuxième au *terme de biais*. La première étape consiste à montrer que le terme de variance après normalisation

$$v_n(x_0)[\hat{S}_n^{x_0}(y_n(x_0)) - S(y_n(x_0)|x_0)]$$

converge vers une loi normale centrée réduite. Cela nécessite la condition (1.8) et le théorème de Lindeberg. Il reste ensuite à montrer que le terme

$$v_n(x_0)[\hat{S}_n(y_n(x_0)|x_0) - \hat{S}_n^{x_0}(y_n(x_0))]$$

converge vers 0 en probabilité en utilisant la condition (1.9). Ainsi, par la méthode expliquée précédemment, on obtient la normalité asymptotique de $\hat{Q}_n(\alpha_n|x_0)$.

Theorem 7 *Soient $x_0 \in \mathbb{R}^p$ tel que $f(x_0) > 0$, $\alpha_n \to 0$ quand $n \to +\infty$ et on suppose que $S(\cdot|x_0)$ vérifie la condition TFO. Supposons qu'il existe une suite $m_n(x_0)$ telle que $n_{x_0}/m_n(x_0) \xrightarrow{\text{p.s}} 1$ et soit $v_n^2(x_0) = m_n(x_0)/\alpha_n$. Si $\alpha_n m_n(x_0) \to +\infty$, $v_n(x_0) g(\alpha_n|x_0) \to +\infty$,*

$$v_n(x_0) \max_{1 \leq i \leq n} \mathcal{W}_{n,i}(x_0) \xrightarrow{\text{p.s}} 0$$

et

$$[\alpha_n m_n(x_0)]^{1/2} \sup_{|\beta/\alpha_n - 1| \leq \xi} \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \left| \frac{S[Q(\beta|x_0)|X_i]}{\beta} - 1 \right| \xrightarrow{\mathbb{P}} 0,$$

pour un certain $\xi \in]0,1[$, alors

$$v_n(x_0) \frac{g(\alpha_n|x_0) Q(\alpha_n|x_0)}{a(1/\alpha_n|x_0)} \left(\frac{\hat{Q}_n(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

Les conditions nécessaires pour cette normalité sont sans surprise sur la nature de la suite α_n. Il ne peut s'agir d'une suite qui est l'ordre d'un quantile extrême. En effet, on retrouve une condition sur la vitesse de convergence de α_n vers 0 qui impose qu'elle ne converge pas vers 0 de manière suffisamment rapide. Il y a donc tout intérêt à définir un estimateur de quantiles extrêmes par extrapolation, par exemple en utilisant la condition TFO. C'est l'objet du chapitre 3. Enfin, on étudie les deux cas particuliers : estimateur de Nadaraya-Watson et estimateur des plus proches voisins et on valide les résultats avec des simulations. Une méthode de validation croisée est présentée afin de choisir de manière optimale les paramètres de sélection h_n et k_n. On teste également la performance de nos estimateurs sur un jeu de données réelles.

Chapitre 3

Dans le chapitre 3, on s'intéresse à un estimateur de quantiles extrêmes conditionnels vérifiant la condition TFO. Cependant, la qualité de cet estimateur dépend étroitement de la pertinence de l'approximation

$$Q(\beta|x_0) \approx Q(\alpha|x_0) + a(1/\alpha|x_0)\phi(t|x_0)$$

avec $t = (\beta - \alpha)/g(\alpha|x_0)$ et $\beta < \alpha$ qui elle même dépend de la distribution sous-jacente. Malheureusement, lorsque l'on compare cette approximation à

la vraie valeur $Q(\beta|x_0)$ les erreurs sont très importantes dans le cas des distributions super heavy-tailed. On se contente donc de la condition habituelle du premier ordre pour définir notre estimateur de quantiles extrêmes. C'est-à-dire, on étudie l'estimateur défini en (1.2).
Il est montré dans ce chapitre que cet estimateur est consistant, la preuve étant similaire à de Haan et Ferreira [25] Théorème 4.3.1. Cela nécessite une condition du second-ordre mais aussi de résultats de consistance sur les estimateurs $\tilde{Q}(\alpha_n|x_0)$, $\tilde{a}_n(1/\alpha_n|x_0)$ et $\tilde{\gamma}_n(x_0)$. Posons

$$\mathrm{ERV}(\alpha, t|x_0) := \frac{Q(t\alpha|x_0) - Q(\alpha|x_0)}{a(1/\alpha|x_0)} - L_{\gamma(x_0)}(1/t)$$

et **(H2)** la condition du second-ordre :
(H2) Il existe une fonction $A(\cdot|x_0)$ ne changeant pas de signe et une constante $\rho(x_0) < 0$ telles que $A(y|x_0) \to 0$ quand $y \to +\infty$ et pour tout $t > 0$,

$$\lim_{\alpha \to 0} \frac{\mathrm{ERV}(\alpha, t|x_0)}{A(1/\alpha|x_0)} = \int_1^{1/t} r^{\gamma(x_0)-1} L_{\rho(x_0)}(r) \mathrm{d}s$$

On a alors le théorème suivant :

Theorem 8 *Soient α_n et β_n deux suites convergeant vers 0 telles que $\beta_n/\alpha_n \to 0$ quand $n \to +\infty$. Supposons que la condition **(H2)** soit satisfaite avec $\rho(x_0) < 0$ ou ($\rho(x_0) \leq 0$ avec $\gamma(x_0) < 0$) et qu'il existe une suite $\tau_n(x_0)$ convergeant vers 0 quand $n \to +\infty$ telle que*

$$\frac{\tilde{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)}{a(1/\alpha_n|x_0)} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)), \quad \frac{\tilde{a}_n(1/\alpha_n|x_0)}{a(1/\alpha_n|x_0)} = 1 + \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$$

et $\tilde{\gamma}_n(x_0) - \gamma(x_0) = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$. Si $\tau_n(x_0)\ln(\alpha_n/\beta_n) \to 0$ et $A(1/\alpha_n|x_0) = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$, alors

$$\frac{\tilde{Q}_n^{(E)}(\beta_n|x_0) - Q(\beta_n|x_0)}{a(1/\alpha_n|x_0) q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)).$$

Il reste encore à choisir des estimateurs vérifiant ces conditions. Si on consière $\tilde{Q}_n(\alpha_n|x_0) = \hat{Q}_n(\alpha_n|x_0)$ défini en (1.6), il est clair que sa consistance est vérifiée par le chapitre 2 pour $\tau_n(x_0) = \{\ln[\alpha_n m_n(x_0)]/[\alpha_n m_n(x_0)]\}^{1/2}$. Le choix pour $\tilde{\gamma}(x_0)$ s'oriente vers celui défini dans l'article de Gardes [17]. Il s'agit d'un estimateur qui est fonctionnelle de $\hat{Q}_n(\alpha_n|x_0)$ et qui peut être utilisé dans plusieurs modèles, y compris celui où l'on considère une variable à grande dimension. Pour $\kappa \in (0,1)$ et φ une fonction positive et bornée sur $[\kappa, 1]$, soit $\tilde{\Psi}$ une fonction décroissante définie pour $s \geq 0$ par $\tilde{\Psi}(s) = 0$ et pour $s < 0$ par

$$\tilde{\Psi}(s) := \left(\int_\kappa^1 \varphi(u) L_s(1/u) \mathrm{d}u\right)^2 \Big/ \int_\kappa^1 \varphi(u) L_s^2(1/u) \mathrm{d}u \, .$$

Pour tout $\delta \in \mathbb{N}$ et toute fonction décroissante continue à droite U et tout $\alpha \in]0,1[$, soit

$$\mathcal{T}_\alpha^{(\delta)}(U) := \int_\kappa^1 \varphi(u) \left(\ln \frac{U(u\alpha)}{U(\alpha)}\right)^\delta \mathrm{d}u \Big/ \left(\int_\kappa^1 \varphi(u) L_0(1/u) \mathrm{d}u\right)^\delta .$$

On considère l'estimateur

$$\begin{aligned} \hat{\gamma}_n(x_0) &:= \hat{\gamma}_{n,+}(x_0) + \hat{\gamma}_{n,-}(x_0) \\ &= \mathcal{T}_{\alpha_n}^{(1)}(\hat{Q}_n(\cdot|x_0)) + \tilde{\Psi}^{\leftarrow}\left(\frac{[\mathcal{T}_{\alpha_n}^{(1)}(\hat{Q}_n(\cdot|x_0))]^2}{\mathcal{T}_{\alpha_n}^{(2)}(\hat{Q}_n(\cdot|x_0))}\right). \end{aligned}$$

Dans le même contexte, on peut définir un estimateur de $a(1/\alpha_n|x_0)$:

$$\hat{a}_n(1/\alpha_n|x_0) := \tilde{\mathcal{T}}_{\alpha_n}\left(\hat{Q}_n(\cdot|x_0); \hat{\gamma}_{n,-}(x_0)\right),$$

où $\tilde{\mathcal{T}}_\alpha(U,s)$ est donné pour toute fonction décroissante et continue à droite par

$$\tilde{\mathcal{T}}_\alpha(U,s) := U(\alpha) \int_\kappa^1 \varphi(u) \ln \frac{U(u\alpha)}{U(\alpha)} \mathrm{d}u \Big/ \int_\kappa^1 \varphi(u) L_s(1/u) \mathrm{d}u \, .$$

Pour la motivation de ces estimateurs on peut se référer à Gardes [17].

Proposition 4 *Soit α_n une suite convergeant vers 0 quand n tend vers l'infini. Supposons qu'il existe une suite $m_n(x_0)$ telle que $n_{x_0}/m_n(x_0) \xrightarrow{\text{p.s}} 1$ et $\alpha_n m_n(x_0) \to +\infty$ quand $n \to \infty$ et soit $\tau_n(x_0)^2 := \{\ln[\alpha_n m_n(x_0)]/[\alpha_n m_n(x_0)]\}$. Si on a les conditions suivantes :*

- *la condition **(H2)** est satisfaite;*
- *$A(1/\alpha_n|x_0) = o(\tau_n(x_0))$ et*

$$\frac{a(1/\alpha_n|x_0)}{Q(\alpha_n|x_0)} - \gamma_+(x_0) = o(\tau_n(x_0)); \tag{1.10}$$

- *il existe une constante C_X telle que*

$$m_n(x_0)\max_{1\le i\le n}\mathcal{W}_{n,i}(x_0) < C_X \ p.s; \tag{1.11}$$

- *il existe $\xi > 0$ tel que*

$$[m_n(x_0)\alpha_n]^{1/2}\sup_{\beta/\alpha_n\in\mathcal{J}_\xi}\sum_{i=1}^{n}\mathbb{E}\left[\mathcal{W}_{n,i}(x_0)\left|\frac{S[Q(\beta|x_0)|X_i]}{\beta}-1\right|\right] = o(\tau_n(x_0)) \tag{1.12}$$

avec $J_\xi := [(1-\xi)\kappa, 1+\xi]$, alors

$$\hat{\gamma}_n(x_0)-\gamma(x_0) = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)) \quad et \quad \frac{\hat{a}_n(1/\alpha_n|x_0)}{a(1/\alpha_n|x_0)} = 1+\mathcal{O}_{\mathbb{P}}(\tau_n(x_0)).$$

La condition (1.12) contrôle les oscillations de la fonction $x \mapsto S(y|x)$ pour des grandes valeurs de y. Ainsi, on établit la consistance suivante.

Corollary 2 *Soient deux suites α_n et β_n convergeant vers 0 quand n tend vers l'infini telles que $\beta_n/\alpha_n \to 0$. Supposons qu'il existe une suite $m_n(x_0)$ telle que $n_{x_0}/m_n(x_0) \xrightarrow{\text{p.s}} 1$ et $\alpha_n m_n(x_0) \to +\infty$. Soit $\tau_n(x_0) := \{\ln[\alpha_n m_n(x_0)]/[\alpha_n m_n(x_0)]\}^{1/2}$. Supposons également que la condition* ***(H2)*** *est satisfaite pour $\rho(x_0)$ ou ($\rho \le 0$ et $\gamma(x_0) < 0$). Si $\tau_n(x_0)\ln(\alpha_n/\beta_n) \to 0$, $A(1/\alpha_n|x_0) = o(\tau_n(x_0))$ et si les conditions (1.10), (1.11) et (1.12) sont vérifiées, alors*

$$\frac{\hat{Q}_n^{(E)}(\beta_n|x_0)-Q(\beta_n|x_0)}{a(1/\alpha_n|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$$

où

$$\hat{Q}_n^{(E)}(\beta_n|x_0) := \hat{Q}_n(\alpha_n|x_0)+\hat{a}_n(1/\alpha_n|x_0)L_{\hat{\gamma}_n(x_0)}(\alpha_n/\beta_n).$$

Enfin, le problème du fléau de la dimension est traité dans une dernière partie. On applique la méthode de réduction de dimension définie dans Gardes [18] qui est spécialement dédiée au cas des extrêmes. En effet, il n'est pas utile de supposer l'indépendance conditionnelle sur tout le support des distributions. Comme on ne s'intéresse qu'à la queue de distribution, on suppose la condition TCI (Tail Conditional Independence):

(TCI) Soit B_0 une matrice de taille $p \times q$ avec $q < p$ de rang maximal. La variable aléatoire Y est TCI de X sachant $B_0^\top X$ si pour tout $\epsilon > 0$ il existe $\tilde{\kappa}$ tel que pour tout $\delta \in (0, \tilde{\kappa}]$,

$$\mathbb{P}\left[\left|\frac{\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|X)}{\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|B_0^\top X)} - 1\right| \leq \epsilon\right] = 1,$$

où pour tout $\delta > 0$, $\mathcal{Y}_\delta$ est une fonction mesurable définie pour tout $z \in \text{supp}(B_0^\top X)$ par $\mathcal{Y}_\delta(z) := Q_{B_0}(0|z) - \delta$ si $Q_{B_0}(0|z) < +\infty$ et $\mathcal{Y}_\delta(z) := \delta^{-1}$ si $Q_{B_0}(0|z) = +\infty$ et où $Q_{B_0}(\cdot|z)$ est le quantile associé à la distribution conditionnelle de Y sachant $\{B_0^\top X = z\}$.
Cela signifie que presque sûrement les distributions de Y sachant X et Y sachant $B_0^\top X$ partagent le même point terminal. Dans la définition de l'estimateur $\hat{S}_n(y|x_0)$, on remplace les poids $\mathcal{W}_{n,i}(x_0)$ par les poids $\mathcal{W}_{n,i}(B_0, x_0)$ qui dépendent de la matrice B_0. On obtient alors l'estimateur de quantile conditionnel pour $\alpha \in]0, 1[$,

$$\hat{Q}_n(\alpha_n|B_0, x_0) := \inf \{y \in \mathbb{R} \mid \hat{S}_n(y|B_0, x_0) \leq \alpha\},$$

où

$$\hat{S}_n(y|B_0, x_0) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(B_0, x_0)\mathbb{I}_{\{Y_i > y\}}.$$

L'estimateur de quantiles extrêmes définis par extrapolation est donc

$$\hat{Q}_n^{(E)}(\beta_n|B_0, x_0) := \hat{Q}_n(\alpha_n|B_0, x_0) + \hat{a}_n(1/\alpha_n|B_0, x_0)L_{\hat{\gamma}_n(B_0,x_0)}(\alpha_n/\beta_n)$$

où $\hat{a}_n(1/\alpha_n|B_0, x_0)$ et $\hat{\gamma}_n(B_0, x_0)$ sont les estimateurs $\hat{a}_n(1/\alpha_n|x_0)$ et $\hat{\gamma}_n(x_0)$ où on a remplacé $\mathcal{W}_{n,i}(x_0)$ par $\mathcal{W}_{n,i}(B_0, x_0)$. La consistance de cet estimateur de quantiles extrêmes se montre de la même façon que pour l'estimateur de quantiles extrêmes sans méthode de réduction de dimension. Enfin, les résultats sont validés par des études de simulation pour les deux cas particuliers : estimateur de Nadaraya-Watson et estimateur des plus proches voisins.

Chapter 2

Estimation of extreme conditional quantiles under a general tail first order condition

2.1 Introduction

To describe the dependence between a real-valued random variable Y and an explanatory random vector X of dimension $p \in \mathbb{N} \setminus \{0\}$, different approaches can be used. The most common one is perhaps provided by the conditional mean $m(X) := \mathbb{E}(Y|X)$, which gives information on the central part of the conditional distribution. However, depending on the applications in mind, it can be also of interest to consider a conditional quantile instead of $m(X)$ (e.g., median or quartile). To be more specific, denoting by $S(\cdot|x_0) := \mathbb{P}(Y > \cdot|X = x_0)$ the conditional survival function of Y given $\{X = x_0\}$ for some $x_0 \in \mathbb{R}^p$ in the support of X, the conditional quantile of level $\alpha \in [0, 1]$ of Y given $\{X = x_0\}$ is $Q(\alpha|x_0) := S^{\leftarrow}(\alpha|x_0) = \inf\{y \in \mathbb{R};\ S(y|x_0) \leq \alpha\}$ with the convention $\inf\{\emptyset\} = +\infty$. This conditional quantile presents the advantage to be more robust than the classical conditional mean.

Given n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), one question of interest is of course the estimation of the conditional quantile $Q(\alpha|x_0)$ in a nonparametric way. There exist numerous estimation methods in the literature. The most common one is the *indirect* method: starting from a suitable estimator $\widehat{S}_n(\cdot|x_0)$ of $S(\cdot|x_0)$, the associated estimator of $Q(\alpha|x_0)$ is given by

$$\widehat{Q}_n(\alpha|x_0) := \widehat{S}_n^{\leftarrow}(\alpha|x_0) = \inf\{y \in \mathbb{R};\ \widehat{S}_n(y|x_0) \leq \alpha\}. \tag{2.1}$$

Estimator (2.1) is called indirect since, as pointed by Racine and Li [31], "one estimates a conditional survival function, and then, one 'backs out' the inferred quantile via inversion".

An alternative way to estimate a conditional quantile is by using the so-called check function defined for $\alpha \in [0,1]$ by $\rho_\alpha(v) := v[\alpha - \mathbb{I}_{(-\infty,0]}(v)]$ where for any $A \subset \mathbb{R}$, $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise. Indeed, since the conditional quantile is also defined by

$$Q(\alpha|x_0) = \arg\min_{\tau\in[0,1]} \mathbb{E}\left[\rho_\alpha(Y-\tau)\middle|\, X = x_0\right],$$

the estimation of $Q(\alpha|x_0)$ can be achieved by replacing the conditional expectation by a suitable estimator and then by solving the minimization problem. This method of estimation was investigated among others by Koenker and Basset [28], Koenker et al. [29] and He and Ng [26]. In this chapter, we focus on the so-called indirect method.

In some applications, we are interested in the tail of the conditional distribution rather than on its central part. In this case, instead of looking at the conditional quantile of level $\alpha \in [0,1]$, we consider an extreme conditional quantile, i.e., a conditional quantile of level α_n where $\alpha_n \to 0$ as the sample size n increases.

To obtain the asymptotic distribution of an indirect conditional quantile estimator, the following two-step procedure can be used. First, we establish the asymptotic distribution of the associated conditional survival function estimator. Next, a delta-type method is used to deduce the result on the conditional quantile estimator from this first step. This requires an additional condition on the conditional survival function. When the level α is fixed, this condition is simply that $S(\cdot|x_0)$ is continuously differentiable. However, in case of an extreme level, this condition is much more complicated. In this work, we introduce a new general condition, called *Tail First Order Condition*, which is the cornerstone to obtain the asymptotic distribution of any indirect conditional quantile estimator. As we will see, this condition is more flexible than the one classically used in extreme value theory.

To understand where the Tail First Order condition comes from, the main ingredients of the proof of the asymptotic normality in case of a fixed level α and of an extreme level α_n is outlined in Section 2.2. In Section 2.3, this condition is specified and illustrated on many well-known examples of conditional distributions. Section 2.4 is devoted to the study of a general class of extreme conditional quantile estimators. In particular, a unified theorem for the asymptotic normality is established. A simulation study is provided in Section 2.5 where several examples of estimators belonging to this class, among them, the kernel and nearest neighbors type estimators, are compared. Their performance is finally illustrated in Section 2.6 on a real dataset on earthquake magnitudes. The proofs of the main results are postponed to Section 2.7.

2.2 Description of the methodology

The aim of this chapter is to show the asymptotic normality of a general class of indirect type of conditional quantile estimators when the level is extreme. This requires a condition, which is not usual in the case of a fixed level α. To understand where this condition comes from we briefly start to present the simple case where the level is fixed, and then, we outline the main differences when it is assumed to be extreme, and we introduce the required condition in that context.

Case where the level is fixed – When the level α is fixed, the asymptotic distribution of (2.1) can be deduced from the one of the conditional survival function estimator $\widehat{S}_n(\cdot|x_0)$. More precisely, if we assume that for some $y \in \mathbb{R}$, there exists a sequence $v_n(x_0) \to \infty$ such that for all sequence $\varepsilon_n \to 0$

$$v_n(x_0)\left(\widehat{S}_n(y+\varepsilon_n|x_0) - S(y+\varepsilon_n|x_0)\right) \xrightarrow{d} \Lambda, \tag{2.2}$$

where Λ is some non-degenerate distribution, then if $S(\cdot|x_0)$ is a continuously differentiable function with $S[Q(\alpha|x_0)|x_0] = \alpha$

$$v_n(x_0)\left(\widehat{Q}_n(\alpha|x_0) - Q(\alpha|x_0)\right) \xrightarrow{d} \frac{1}{f(Q(\alpha|x_0)|x_0)}\Lambda, \tag{2.3}$$

where $f(\cdot|x_0)$ is the probability density function of Y given $\{X = x_0\}$ with $f(Q(\alpha|x_0)|x_0) \neq 0$. The proof of (2.3) is based on the following remark: for all $z \in \mathbb{R}$, letting $\sigma_n(x_0) := v_n(x_0) f(Q(\alpha|x_0)|x_0)$, one has

$$\mathbb{P}\left[\sigma_n(x_0)\left(\widehat{Q}_n(\alpha|x_0) - Q(\alpha|x_0)\right) \leq z\right] = \mathbb{P}[Z_n(x_0) \leq z_n(x_0)], \qquad (2.4)$$

where,

$$Z_n(x_0) := v_n(x_0)\left(\widehat{S}_n(Q(\alpha|x_0) + z\sigma_n^{-1}(x_0)|x_0) - S(Q(\alpha|x_0) + z\sigma_n^{-1}(x_0)|x_0)\right)$$

and $z_n(x_0) := v_n(x_0)[\alpha - S(Q(\alpha|x_0) + z\sigma_n^{-1}(x_0)|x_0)]$. From (2.2) with $y = Q(\alpha|x_0)$, $Z_n(x_0) \xrightarrow{d} \Lambda$ and since $S(\cdot|x_0)$ is continuously differentiable, $z_n(x_0) \to z$ as $n \to \infty$ proving (2.3). Note that the asymptotic distribution of indirect estimators for a fixed level α has been treated for instance by Berlinet et al. [3].

Case of an extreme level – We consider the situation where the level of the conditional quantile is a sequence α_n where $\alpha_n \to 0$ as the sample size n increases. Replacing the level α by a sequence α_n does not change (at least if α_n does not converge too fast to 0) the estimation procedure. We still estimate $Q(\alpha_n|x_0)$ as in (2.1) just by replacing α by α_n. The difference lies in the assumptions required to obtain the asymptotic distribution of $\widehat{Q}_n(\alpha_n|x_0)$. First, instead of (2.2), the following kind of result for the conditional survival function estimator is required: for some well-chosen sequence $y_n(x_0) \to y^*(x_0) := Q(0|x_0)$, there exists a sequence $v_n(x_0) \to \infty$ such that

$$v_n(x_0)\left(\widehat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0)\right) \xrightarrow{d} \Lambda, \qquad (2.5)$$

for some non-degenerate distribution Λ. Of course, the sequence $v_n(x_0)$ depends on the sequence $y_n(x_0)$. Since $y^*(x_0)$ is the right endpoint, convergence (2.5) focus on the asymptotic behavior of $\widehat{S}_n(\cdot|x_0)$ in the right tail of the conditional distribution. To obtain the asymptotic distribution of $\widehat{Q}_n(\alpha_n|x_0)$, we start again with (2.4) where α is replaced by α_n. In the extreme level case, the main difficulty is to deal with the non-random sequence $z_n(x_0)$. More specifically, assuming that $S[Q(\alpha|x_0)|x_0] = \alpha$ at least for α small enough, we need to find a general condition on the conditional distribution ensuring that for a well-chosen sequence $\sigma_n(x_0)$ and for a sequence $v_n(x_0)$ satisfying (2.5) with $y_n(x_0) = Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)$

$$z_n(x_0) = -\alpha_n v_n(x_0) \left[\frac{S[y_n(x_0)|x_0]}{S[Q(\alpha_n|x_0)|x_0]} - 1 \right] \to z, \tag{2.6}$$

as $n \to \infty$ for all $z \in \mathbb{R}$. Obviously, assuming that $S(\cdot|x_0)$ is a continuously differentiable function is not relevant here and the sequence $\sigma_n(x_0)$ is not necessarily equal to $v_n(x_0)f(Q(\alpha_n|x_0)|x_0)$. Since $Q(\alpha_n|x_0) \to y^*(x_0)$, a natural general condition leading to (2.6) is to assume that for some open interval $I_{x_0} = I \subset \mathbb{R}$ containing 0, there exist positive functions $d_{x_0} \equiv d$ and $\Psi_{x_0} \equiv \Psi$ such that for all $t \in I$,

$$\lim_{y \uparrow y^*(x_0)} \Psi(y) \left(\frac{S[y + td(y)|x_0]}{S(y|x_0)} - 1 \right) \to \phi_{x_0}^{-1}(t), \tag{2.7}$$

where $\phi_{x_0}^{-1} \equiv \phi^{-1}$ is the inverse of a continuous and strictly decreasing function $\phi_{x_0} \equiv \phi$ such that $\phi(t)/t \to -1$ as $t \to 0$.
Indeed, taking $\sigma_n(x_0) = \alpha_n v_n(x_0)/[\Psi(Q(\alpha_n|x_0))d(Q(\alpha_n|x_0))]$ and $t_n^{-1}(x_0) := \sigma_n(x_0)d[Q(\alpha_n|x_0)]$, we obtain

$$z_n(x_0) = -\frac{\Psi[Q(\alpha_n|x_0)]}{t_n(x_0)} \left(\frac{S[Q(\alpha_n|x_0) + zt_n(x_0)d[Q(\alpha_n|x_0)]|x_0]}{S[Q(\alpha_n|x_0)|x_0]} - 1 \right).$$

Under (2.7) and assuming that $t_n(x_0) \to 0$, we can show that $z_n(x_0) \to z$ (see Section 2.3, Proposition 5). Next, the random sequence $Z_n(x_0)$ is treated by (2.5). To sum up, in the extreme level case, a natural condition on $S(\cdot|x_0)$ to establish the asymptotic distribution of the conditional quantile estimator is (2.7). Condition (2.7) is referred in what follows to as the Tail First Order condition. Under this condition and if (2.5) holds with $y_n(x_0) := Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)$, we have $\sigma_n(x_0)(\widehat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)) \xrightarrow{d} \Lambda$. We show in Section 2.3 that this Tail First Order condition is satisfied by a larger class of conditional distributions than the one satisfying the condition classically used in extreme value theory. Note that while on the fixed level case, the rate of convergence of $\widehat{Q}_n(\alpha|x_0)$ is proportional to $v_n(x_0)$ this is no longer the case when estimating an extreme conditional quantile.

2.3 The Tail First Order condition

The Tail First Order condition is related to the conditional distribution of Y given $\{X = x_0\}$ for some $x_0 \in \mathbb{R}^p$ in the support of X. Since x_0 is fixed, the dependence on x_0 can be omitted. This is what we do in all this section. For a given (conditional) survival function S, we denote by $Q = S^{\leftarrow}$ the associated quantile and by $x^* = S^{\leftarrow}(0)$ the right endpoint.

Definition 5 A survival function S satisfies the Tail First Order (TFO) condition if for some open interval $I \subset \mathbb{R}$ containing 0, there exist positive functions d and Ψ such that for all $t \in I$,

$$\lim_{x \uparrow x^*} \Psi(x) \left(\frac{S[x + td(x)]}{S(x)} - 1 \right) = \phi^{-1}(t), \tag{2.8}$$

where ϕ^{-1} is the inverse of a continuous and strictly decreasing function $\phi : J \to I$ such that $\phi(t)/t \to -1$ as $t \to 0$.

Note that convergence (2.8) entails that for all $t \in I$ and for x large enough, $x + td(x) < x^*$. Consequently, the function Ψ is such that $\Psi(x)/S(x) \to \infty$ as $x \uparrow x^*$. Finally, it is easy to check that $\phi^{-1}(t)/t \to -1$ as $t \to 0$. As a consequence of Dini's theorem, we obtain the useful properties gathered in the next proposition.

Proposition 5 *If S satisfies the TFO condition, the following statements are true:*
1. Convergence in (2.8) holds locally uniformly on I.
2. For all $t_0 \in I$,

$$\lim_{(t,x) \to (t_0, x^*)} \frac{\Psi(x)}{t} \left(\frac{S[x + td(x)]}{S(x)} - 1 \right) = \lim_{t \to t_0} \frac{\phi^{-1}(t)}{t}.$$

We give in the next result some equivalent reformulations of the TFO condition.

Proposition 6 *The following statements are equivalent:*
1. The survival function S satisfies the TFO condition.
2. There exist positive functions a and g such that for all $t \in J$,

$$\lim_{\alpha \to 0} \frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} = \phi(t). \tag{2.9}$$

3. There exist sequences $a_n > 0$, $b_n \in \mathbb{R}$ and $c_n > 0$ with $nc_n \to \infty$ such that for all $t \in I$,

$$\lim_{n \to \infty} [nc_n S(a_n t + b_n) - c_n] = \phi^{-1}(t). \tag{2.10}$$

Remarks – 1) The relations between the auxiliary functions involved in (2.8) and (2.9) are: $d(\cdot) = a(1/S(\cdot))$ and $\Psi(\cdot) = S(\cdot)/g(S(\cdot))$.
2) A possible choice for the sequences a_n, b_n and c_n in (2.10) is $a_n = a(n)$, $b_n = Q(1/n)$ and $c_n = 1/[ng(1/n)]$. It is also easy to check that necessarily $g(\alpha) \to 0$ as $\alpha \to 0$.
3) An interpretation of condition (2.9) is based on the following remark: from the second statement of Proposition 5,

$$\frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{tg(\alpha)} \sim -\frac{a(\alpha^{-1})}{g(\alpha)},$$

as $(t, \alpha) \to (0, 0)$. Hence, one can see the function $-a(\alpha^{-1})/g(\alpha)$ as the derivative of Q near 0 and in the direction of $g(\alpha)$. This heuristic is confirmed by the next result which provides a sufficient condition for the TFO condition.

Proposition 7 *Assume that Q is a differentiable function and that for some open interval $J \subset \mathbb{R}$ containing 0, there exists a positive function g such that for all $t \in J$,*

$$\lim_{\alpha \to 0} \frac{Q'[\alpha + tg(\alpha)]}{Q'(\alpha)} = \Theta(t). \tag{2.11}$$

If for all $t \in J$, $\int_0^\top \Theta(s)ds =: \theta(t) \in \mathbb{R}$ where θ is an increasing function on J such that $\theta(t)/t \to 1$ as $t \to 0$ then condition (2.9) holds with $\phi(t) = -\theta(t)$ and $a(\alpha^{-1}) = -Q'(\alpha)g(\alpha)$.

We conclude this section by giving examples of distributions satisfying the TFO condition.
Maximum domain of attraction – In extreme value theory, in order to make inference on the tail of a distribution S, we classically assume that there exist sequences $a_n > 0$ and b_n and a non-degenerate distribution function G for which

$$\lim_{n \to \infty} [1 - S(a_n x + b_n)]^n = G(x), \tag{2.12}$$

for all point of continuity of G. Fisher and Tippett [14] and Gnedenko [21] show that $G(x) = G_\gamma(ax + b)$ for some $a > 0$ and $b \in \mathbb{R}$ where

$$G_\gamma(x) = \exp\left[-(1 + \gamma x)^{-1/\gamma}\right],$$

for all x such that $1 + \gamma x > 0$. A survival function S satisfying (2.12) is said to belong to the maximum domain of attraction of the extreme value distribution G_γ. The parameter $\gamma \in \mathbb{R}$ is called the extreme value index. As established in de Haan and Ferreira [25] Theorem 1.1.6, condition (2.12) is equivalent to assume the existence of a positive auxiliary function a and a non constant function ϕ for which

$$\lim_{\alpha \to 0} \frac{Q(t\alpha) - Q(\alpha)}{a(\alpha^{-1})} = \phi(t). \tag{2.13}$$

From de Haan and Ferreira [25] Theorem B.2.1, the function ϕ in (2.13) is necessarily of the form $\phi(t) = c(t^{-\gamma} - 1)/\gamma$ for some $c \neq 0$ and where $\gamma \in \mathbb{R}$ is always the extreme value index. We say that the distribution is of extended regular variation (ERV).

The aim of the next result is to show that the TFO condition introduced in this chapter (see Definition 5) is weaker than (2.12).

Proposition 8 *If S satisfies the TFO condition with an auxiliary function g in (2.9) such that $\alpha/g(\alpha) \to c \geq 0$ as $\alpha \to 0$ (with g continuous and strictly increasing if $c = 0$) then S satisfies (2.12).*

As a consequence of this result, if a survival function S satisfies the TFO condition with a function g as in Proposition 8, then S also satisfies the TFO condition with $g(\alpha) = \alpha$ and in this case the TFO condition coincides with the classical extreme value condition. Remark also that in this situation (i.e., $g(\alpha) = \alpha$), condition (2.11) is equivalent to assume that

$$\lim_{\alpha \to 0} \frac{Q'(t\alpha)}{Q'(\alpha)} = t^{-\gamma-1},$$

for some $\gamma \in \mathbb{R}$. This condition coincides with condition (1.1.33) in de Haan and Ferreira [25] Corollary 1.1.10.

At this step, a natural question is: "Can we find survival functions that satisfy the TFO condition but not the classical extreme value one ?" Roughly speaking, this is equivalent to find survival functions S such that (2.9) holds with a function g such that $\alpha/g(\alpha) \to \infty$. An example of such survival functions is given by super heavy-tailed distributions.

Super heavy-tailed distributions − The term *super heavy-tailed* is often attached in the literature to a distribution with a slowly varying survival function S, i.e., such that for all $t > 0$,

$$\lim_{x \to \infty} \frac{S(tx)}{S(x)} = 1. \tag{2.14}$$

It can be shown that these survival functions do not satisfy the classical first order condition (2.12). Note that a heavy-tailed distribution corresponds to a survival function satisfying for all $t > 0$, $S(tx)/S(x) \to t^{-1/\gamma}$ as $x \to \infty$, for some $\gamma > 0$. Hence, roughly speaking, a super heavy-tailed distribution is a heavy-tailed distribution with $\gamma = +\infty$.

Unfortunately, condition (2.14) is not precise enough for the study of super heavy-tailed distribution. To define more precisely the class of super heavy-tailed distribution, we start by remarking that for heavy-tailed distributions, there exists $\gamma > 0$ such that for all $s > -1$,

$$\lim_{\alpha \to 0} \frac{Q[(1+s)\alpha]}{Q(\alpha)} = (1+s)^{-\gamma}.$$

Since super heavy-tailed distribution can be seen as a heavy-tailed distribution with $\gamma = +\infty$, we propose to replace in the previous limit γ by $\gamma(\alpha)$ where $\gamma(\alpha) \to \infty$ as $\alpha \to 0$, and s by $t/\gamma(\alpha)$ with $t \in \mathbb{R}$ to obtain a non-degenerate limit:

$$\lim_{\alpha \to 0} \frac{Q[(1+t/\gamma(\alpha))\alpha]}{Q(\alpha)} = \mathrm{e}^{-t}.$$

The class of super heavy-tailed distributions can thus be defined by the set of distributions for which there exists a positive function g with $g(\alpha)/\alpha \to 0$ as $\alpha \to 0$ such that for all $t \in \mathbb{R}$

$$\lim_{\alpha \to 0} \frac{Q[\alpha + tg(\alpha)]}{Q(\alpha)} = \mathrm{e}^{-t}. \tag{2.15}$$

It appears that convergence (2.15) coincides with the TFO condition with $a(\alpha^{-1}) = Q(\alpha)$ and $\phi(t) = \mathrm{e}^{-t} - 1$. As shown in Proposition 9 below, this definition is equivalent to the one introduced for instance in Fraga Alves et al. [15] where the class of super heavy-tailed distributions is defined as the set of distributions for which there exists a positive function b such that

$$\lim_{x \to \infty} \frac{U[x + tb(x)]}{U(x)} = \mathrm{e}^{t} \tag{2.16}$$

with $U(\cdot) := Q(1/\cdot)$. Note that according to Fraga Alves et al. [15] Lemma 4.1, condition (2.16) implies (2.14). Furthermore, the function b is such that $b(x)/x \to 0$ as $x \to \infty$. Remark finally that the right end-point of a distribution satisfying (2.16) is necessarily infinite. As examples of super heavy-tailed distribution satisfying (2.16), one can cite the standard log-Pareto distribution given by $S(x) = [\ln(x)]^{-\xi}$ with $\xi > 0$ and the log-Weibull distribution for which $S(x) = \exp(-\xi \ln^{\theta} x)$, with $\xi > 0$ and $\theta \in (0, 1)$. For these two distributions, one can take $b \sim U/U'$.

Proposition 9 *Conditions (2.15) and (2.16) are equivalent. The relation between the involved functions is* $b(x) = x^2 g(x^{-1})$.

2.4 Extreme conditional quantile estimation

Let (X, Y) be a random vector taking its values in $\mathbb{R}^p \times \mathbb{R}$. In all what follows, we assume that (X, Y) admits a probability density function (pdf). The marginal pdf of X is denoted by f. As in the introduction, for all $x_0 \in \mathbb{R}^p$, let $S(\cdot|x_0)$ and $Q(\cdot|x_0)$ be the survival function and the quantile function of the conditional distribution of Y given $\{X = x_0\}$, respectively. Given n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), the first part of this section is dedicated to the presentation of a large class of estimators of $Q(\cdot|x_0)$. In the second part, we show that under the TFO condition, the proposed estimators computed with an extreme level $\alpha_n \to 0$ are asymptotically Gaussian.

2.4.1 A class of conditional quantile estimators

As mentioned in the introduction we focus in this chapter on indirect estimators of $Q(\cdot|x_0)$. The first step is thus the estimation of the conditional survival function $S(\cdot|x_0)$. We consider estimators of the form

$$\widehat{S}_n(y|x_0) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \mathbb{I}_{(y,\infty)}(Y_i). \tag{2.17}$$

The set of weights $\{\mathcal{W}_{n,i}(x_0),\ 1 \le i \le n\}$ is a triangular array of positive random variables depending on the data $X_1, \ldots X_n$ as well as on x_0 such that

$$\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) = 1.$$

These properties on the random weights ensure that $\widehat{S}_n(\cdot|x_0)$ is a well-defined distribution function. This is crucial to estimate the conditional quantile by inverting estimator (2.17). This class of estimators encompasses various classical estimators of the conditional distribution function, see below for some examples. The indirect estimator of the conditional quantile of level $\alpha \in (0,1)$ is thus defined as in (2.1) by

$$\widehat{Q}_n(\alpha|x_0) := \widehat{S}_n^{\leftarrow}(\alpha|x_0) = \inf\{y \in \mathbb{R};\ \widehat{S}_n(y|x_0) \leq \alpha\}.$$

Of course, the main feature of the weights $\{\mathcal{W}_{n,i}(x_0),\ 1 \leq i \leq n\}$ is to select a set of data around x_0. For this reason, estimator of the form (2.17) are called *weighted local* estimators.

The kernel based estimator introduced by Nadaraya [33] and Watson [39] is a classical example of weighted local estimator. This estimator is obtained by using the following random weights in (2.17):

$$\mathcal{W}_{n,i}^{\mathrm{NW}}(x_0, h_n) := K\left(\frac{X_i - x_0}{h_n}\right) \Bigg/ \sum_{j=1}^{n} K\left(\frac{X_j - x_0}{h_n}\right), \tag{2.18}$$

where K is a density on $\mathbb{R}^p$ and h_n is a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$. Typically, the probability density function K has a unique mode at 0 in order to give the largest values of the weights for the observations close to x_0.

Another possibility to select the observations is to take the k_n observations which are closest to the reference position x_0. This approach is called the k_n-Nearest Neighbors (k_n-NN) method. More specifically, for some norm $\|\cdot\|$ on $\mathbb{R}^p$, let $\{D_i(x_0) := \|X_i - x_0\|,\ i = 1, \ldots, n\}$ be the distances between each observation and x_0 and let $D_{(1)}(x_0) \leq \ldots \leq D_{(n)}(x_0)$ the corresponding order statistics. Denoting by $\{r(i),\ i = 1, \ldots, n\}$ the ranks of these distances (i.e., $D_{(i)}(x_0) = D_{r(i)}(x_0)$ for $i = 1, \ldots, n$), the k_n-NN estimator is obtained by using the following random weights in (2.17):

$$\mathcal{W}_{n,i}^{\mathrm{NN}}(x_0, k_n) := [(k_n - r(i) + 1)_+]^{\ell} \Bigg/ \sum_{j=1}^{k_n} j^{\ell}, \tag{2.19}$$

where $(\cdot)_+$ stands for the positive part function and $\ell \in \mathbb{N}$. For instance, by taking $\ell = 0$ (with the convention $0^0 = 0$), we affect the same weight to the k_n closest observations. The corresponding weights are referred to as uniform k_n-NN weights. The choice $\ell = 1$ (resp., $\ell = 2$) leads to triangular k_n-NN weights (resp., quadratic k_n-NN weights).

Roughly speaking, the main difference between these two sets of weights is that the kernel based estimator averages over all observations which are within a fixed distance, whereas the k_n-NN approach averages over a fixed number of observations which might be arbitrarily far away. Of course, one can also think about a linear combination (LC) of (2.18) and (2.19). For instance, we can consider the random weights defined for $\tau \in (0,1)$ by

$$\mathcal{W}_{n,i}^{\mathrm{LC}}(x_0, \tau, h_n, k_n) := \frac{\tau}{M_n} \mathbb{I}_{[0,1]}\left(\left\| \frac{X_i - x_0}{h_n} \right\|_\infty \right) + \frac{1-\tau}{k_n} \mathbb{I}_{[0,1]}\left(\frac{r(i)}{k_n} \right), \quad (2.20)$$

where M_n is the random number of random variables among $\{X_1, \ldots, X_n\}$ that belong to $\mathcal{B}_{x_0}(h_n)$, the closed ball with respect to $\|\cdot\|_\infty$ centered at x_0 and with radius h_n.

2.4.2 Main results

Under general conditions on the random weights $\{\mathcal{W}_{n,i}(x_0),\ i = 1, \ldots, n\}$, we want to establish the convergence in distribution of a normalized version of $\widehat{Q}_n(\alpha_n|x_0)$ for a level α_n converging to 0 as $n \to \infty$. As outlined in Section 2.2, we first need to find a sequence $v_n(x_0) \to \infty$ and a non-degenerate distribution Λ such that (under some additional assumptions)

$$v_n(x_0) \left(\widehat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0) \right) \xrightarrow{d} \Lambda,$$

for some sequence $y_n(x_0) \uparrow y^*(x_0)$. This is done in Theorem 9 where the following notation is used

$$n_{x_0} := \left(\sum_{i=1}^{n} \mathcal{W}_{n,i}^2(x_0) \right)^{-1}.$$

Note that the random variable n_{x_0} corresponds, roughly speaking, to the number of observations used in the estimation procedure. For instance, for the Nadaraya-Watson (NW) weights with the uniform kernel $K(\cdot) \propto \mathbb{I}_{[0,1]}(\|\cdot\|_\infty)$, it is easy to check that n_{x_0} is exactly the number of points in $\mathcal{B}_{x_0}(h_n)$. For the uniform k_n-NN weights, one has $n_{x_0} = k_n$, the number of nearest neighbors.

Theorem 9 *Let $x_0 \in \mathbb{R}^p$ such that $f(x_0) > 0$ and let $y_n(x_0)$ be a sequence converging to the right endpoint $y^*(x_0)$ of the conditional distribution of Y*

given that $\{X = x_0\}$. *Assume that there exists a sequence* $m_n(x_0)$ *such that* $n_{x_0}/m_n(x_0) \xrightarrow{a.s.} 1$ *and let* $v_n^2(x_0) := m_n(x_0)/S(y_n(x_0)|x_0)$. *Under the conditions*

$$v_n(x_0) \max_{1 \le i \le n} \mathcal{W}_{n,i}(x_0) \xrightarrow{a.s.} 0 \tag{2.21}$$

and

$$v_n(x_0) \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \left| S(y_n(x_0)|X_i) - S(y_n(x_0)|x_0) \right| \xrightarrow{\mathbb{P}} 0, \tag{2.22}$$

we have that $v_n(x_0)\left(\widehat{S}_n\left(y_n(x_0)|x_0\right) - S\left(y_n(x_0)|x_0\right)\right) \xrightarrow{d} \mathcal{N}(0,1)$.

To understand the usefulness of conditions (2.21) and (2.22), we provide below the main ideas of the proof of Theorem 9, the complete proof being postponed to Section 2.7. Let $Y_i^{x_0} := Q(U_i|x_0)$ where $U_1, U_2, \dots$ are independent standard uniform random variables, independent of the X_i. The random vectors $\{(X_i, Q(U_i|X_i)),\ i = 1, \dots, n\}$ are thus independent and distributed as (X, Y), which implies that

$$\widehat{S}_n(y_n(x_0)|x_0) \stackrel{d}{=} \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \mathbb{I}_{(y_n(x_0),\infty)}(Q(U_i|X_i)).$$

In other words, one can work as if $Y_i = Q(U_i|X_i)$. The starting point of the proof is the decomposition

$$\begin{aligned} \widehat{S}_n\left(y_n(x_0)|x_0\right) - S\left(y_n(x_0)|x_0\right) &= \left[\widehat{S}_n^{x_0}\left(y_n(x_0)\right) - S\left(y_n(x_0)|x_0\right)\right] \\ &+ \left[\widehat{S}_n\left(y_n(x_0)|x_0\right) - \widehat{S}_n^{x_0}\left(y_n(x_0)\right)\right], \end{aligned}$$

where for all $y \in \mathbb{R}$,

$$\widehat{S}_n^{x_0}(y) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \mathbb{I}_{(y,\infty)}(Y_i^{x_0}).$$

Since $\mathbb{E}[\widehat{S}_n^{x_0}(y_n(x_0))] = S(y_n(x_0)|x_0)$, the first term corresponds to the *variance term* and the second one to the *bias term*.

The first part of the proof consists in establishing the asymptotic normality of the normalized variance term given by:

$$v_n(x_0)\left[\widehat{S}_n^{x_0}\left(y_n(x_0)\right) - S\left(y_n(x_0)|x_0\right)\right],$$

see Section 2.7, Proposition 11. This is obtained mainly by applying the Lindeberg theorem and only condition (2.21) is required. This condition is in fact equivalent to the Lindeberg condition.

In the second part of the proof, we show that the bias term given by

$$B_n(x_0) := v_n(x_0)\left[\widehat{S}_n\left(y_n(x_0)|x_0\right) - \widehat{S}_n^{x_0}\left(y_n(x_0)\right)\right]$$

converges to 0 in probability (see Section 2.7, Proposition 12). The proof is based on the following remark. Let $\mathcal{W}_{n,x_0}$ be the discrete random measure define for all $A \in \mathcal{B}(\mathbb{R}^p)$ by

$$\mathcal{W}_{n,x_0}(A) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\delta_{X_i}(A).$$

Straightforward calculation leads to

$$\begin{aligned} & \widehat{S}_n(y_n(x_0)|x_0) - \widehat{S}_n^{x_0}(y_n(x_0)) \\ = \; & \int \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{(y_n(x_0),\infty)}(Q(U_i|\cdot))\left(d\mathcal{W}_{n,x_0} - d\delta_{x_0}\right). \end{aligned}$$

To control the bias term we need to measure the discrepancy between the two probability measures $\mathcal{W}_{n,x_0}$ and δ_{x_0}. A useful distance between probability measures is the Wasserstein distance defined for all probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ by $W_1(\mathbb{P}_1, \mathbb{P}_2) = \inf\{[\mathbb{E}(|X_1 - X_2|)],\ X_1 \sim \mathbb{P}_1,\ X_2 \sim \mathbb{P}_2\}$. Condition (2.22) can in fact be written in term of the Wasserstein distance as follows

$$v_n(x_0)W_1(\mathcal{W}_{n,x_0}^*, \delta_{x_0}^*) \xrightarrow{\mathbb{P}} 0, \qquad (2.23)$$

where $\mathcal{W}_{n,x_0}^*$ and $\delta_{x_0}^*$ are the pushforward measures of $\mathcal{W}_{n,x_0}$ and δ_{x_0} by the measurable function $S(y_n(x_0)|\cdot)$.

We have now all the ingredients to establish the asymptotic distribution of the conditional quantile estimator of level α_n obtained by inverting the estimator $\widehat{S}_n(\cdot|x_0)$. This requires the following first order condition on the conditional distribution of Y given $\{X = x_0\}$.

(H) The conditional survival function $S(\cdot|x_0)$ satisfies the TFO condition with positive auxiliary functions $\Psi_{x_0} \equiv \Psi$ and $d_{x_0} \equiv d$.

Let $a(1/\cdot) \equiv a_{x_0}(1/\cdot\,|x_0) = d[Q(\cdot|x_0)]$ and $g(\cdot) \equiv g(\cdot|x_0) = \cdot/\Psi[Q(\cdot|x_0)]$. From Proposition 6, condition **(H)** is equivalent to assume that for some open interval $J_{x_0} = J \subset \mathbb{R}$ containing 0, one has for all $t \in J$

$$\lim_{\alpha\to 0} \frac{Q(\alpha + tg(\alpha)|x_0) - Q(\alpha|x_0)}{a(\alpha^{-1})} = \phi_{x_0}(t),$$

where $\phi_{x_0} \equiv \phi$ is a continuous and strictly decreasing function such that $\phi(t)/t \to -1$ as $t \to 0$.

Theorem 10 *Let $x_0 \in \mathbb{R}^p$ such that $f(x_0) > 0$ and assume that condition* **(H)** *holds. Assume that there exists a sequence $m_n(x_0)$ such that $n_{x_0}/m_n(x_0) \xrightarrow{a.s.} 1$ and let $v_n^2(x_0) := m_n(x_0)/\alpha_n$. If $\alpha_n m_n(x_0) \to \infty$, $v_n(x_0)g(\alpha_n) \to \infty$,*

$$v_n(x_0) \max_{1\le i\le n} \mathcal{W}_{n,i}(x_0) \xrightarrow{a.s.} 0$$

and

$$[\alpha_n m_n(x_0)]^{1/2} \sup_{|\beta/\alpha_n - 1|\le\xi} \sum_{i=1}^n \mathcal{W}_{n,i}(x_0) \left| \frac{S[Q(\beta|x_0)|X_i]}{\beta} - 1 \right| \xrightarrow{\mathbb{P}} 0, \qquad (2.24)$$

for some $\xi \in (0,1)$ then

$$v_n(x_0) \frac{g(\alpha_n)Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)} \left(\frac{\widehat{Q}_n(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}(0,1).$$

Recall that if $g(\alpha) = \alpha$ (or equivalently $\Psi \equiv 1$), condition **(H)** coincides with the classical first order condition (2.13) used in extreme value theory. In this case, $\phi(t) \propto (t^{-\gamma(x_0)} - 1)/\gamma(x_0)$ where the function γ is referred to as the conditional extreme value index. Under (2.13) and if the conditions of Theorem 10 are satisfied,

$$[\alpha_n m_n(x_0)]^{1/2} \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1})} \left(\frac{\widehat{Q}_n(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}(0,1).$$

Moreover, we know from de Haan and Ferreira [25] Lemma 1.2.9 that under (2.13), $Q(\alpha_n|x_0)/a(\alpha_n^{-1}) \to 1/\gamma_+(x_0)$, where $\gamma_+(x_0) = \max(\gamma(x_0), 0)$.

So, under the first order condition (2.13), the worst rate of convergence is achieved when $\gamma(x_0) > 0$. This was expected since the case $\gamma(x_0) > 0$ corresponds to heavy-tailed distributions.

Let us now focus on the rate of convergence in Theorem 10 for conditional super heavy-tailed distribution. Taking the definition of super heavy-tailed distributions given in Fraga Alves et al. [15] into account, we have in this case $a(\alpha^{-1}) = Q(\alpha|x_0)$ and $g(\alpha)/\alpha \to 0$ as $\alpha \to 0$. Hence, for these distributions,

$$[\alpha_n m_n(x_0)]^{1/2} \frac{g(\alpha_n)}{\alpha_n} \left(\frac{\widehat{Q}_n(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}(0,1).$$

Not surprisingly, this rate is worse than the one for heavy-tailed distributions.

Theorem 10 is proved under general conditions on the random weights used to define the conditional survival estimator (2.17). We close this section by applying Theorem 10 to particular weights.

Nadaraya-Watson weights – Taking the weights defined in (2.18) leads to the well-known NW estimator of the conditional survival function:

$$\widehat{S}_n^{\mathrm{NW}}(y|x_0) := \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h_n}\right) \mathbb{I}_{(y,\infty)}(Y_i) \Bigg/ \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h_n}\right). \quad (2.25)$$

The corresponding conditional quantile estimator is denoted by $\widehat{Q}_n^{\mathrm{NW}}(\alpha_n|x_0)$. In order to apply Theorem 10, we need to check that the NW weights satisfy the required conditions. To this aim, we assume the following on the kernel function K:

(K) the kernel K is either an indicator function on a cell of $\mathbb{R}^p$ or such that $K(x) = L(\|x\|)$ where L is of bounded variation, continuous on $(0,\infty)$ and with support $[0,1]$.

It is very easy to check that **(K)** is satisfied for a large range of usual kernels such as the uniform kernel ($K(t) \propto \mathbb{I}_{[0,1]}(\|t\|_\infty)$), triangular (with $L(t) \propto 1-t$), Epanechnikov kernel ($L(t) \propto 1-t^2$), biweight kernel ($L(t) \propto (1-t^2)^2$), etc. We can now state the convergence in distribution of the conditional survival estimator (2.25). Recall that f is the pdf of X.

Corollary 3 *Let $x_0 \in \mathbb{R}^p$ such that f is continuous at x_0 and $f(x_0) > 0$ and let K be a kernel satisfying* **(K)**. *Under* **(H)**, *for sequences $h_n \to 0$ and $\alpha_n \in (0,1)$ such that $nh_n^p[\alpha_n \wedge (\ln\ln n)^{-1}] \to \infty$, $\alpha_n^{-1} nh_n^p g^2(\alpha_n) \to \infty$ and*

$$\sup_{\substack{|\beta/\alpha_n - 1| \le \xi \\ \|x - x_0\| \le h_n}} \left| \frac{S[Q(\beta|x_0)|x]}{\beta} - 1 \right|^2 = o\left(\frac{1}{nh_n^p \alpha_n} \right), \tag{2.26}$$

for some $\xi \in (0,1)$ we have

$$\frac{g(\alpha_n)}{\alpha_n} \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1})} (nh_n^p \alpha_n)^{1/2} \left(\frac{\widehat{Q}_n^{\mathrm{NW}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\|K\|_2^2}{f(x_0)}\right).$$

Note that under the classical first order condition (2.13) (i.e., when $g(\alpha_n) = \alpha_n$ in **(H)**, see (2.9) and the remarks below Proposition 6), the asymptotic normality of the NW conditional quantile estimator has already been obtained in Daouia et al. [6] Corollary 1. This last result also requires the use of condition (2.26) which controls the oscillations of the function $Q(\alpha_n|\cdot)$. Of course, the proof of Daouia et al. [6] Corollary 1 uses arguments adapted to the NW estimator while Theorem 10 can be used for a large range of weighted conditional survival estimators. As a consequence, conditions on h_n and α_n involved in Daouia et al. [6] Corollary 1 and in our Corollary 3 are slightly different. More precisely, the conditions in Daouia et al. [6] Corollary 1 are $nh_n^p \alpha_n \to \infty$ and $nh_n^{p+2}\alpha_n \to 0$ while in our Corollary 3 it is required that $nh_n^p\alpha_n \to \infty$ and $nh_n^p(\ln\ln n)^{-1} \to \infty$. Hence, if $\alpha_n \ln\ln n \to 0$ as $n \to \infty$ (i.e., for large quantiles), conditions on the sequences h_n and α_n are weaker in Corollary 3 than in Daouia et al. [6] Corollary 1.

Nearest Neighbors approach – Now, let us consider the k_n-NN random weights defined in (2.19) and leading to the conditional survival function estimator

$$\widehat{S}_n^{KNN}(y|x_0) := \sum_{i=1}^n [(k_n - r(i) + 1)_+]^\ell \mathbb{I}_{(y,\infty)}(Y_i) \Bigg/ \sum_{j=1}^{k_n} j^\ell ,$$

with $k_n \in \{1, \dots, n\}$, $\ell \in \mathbb{N}$ and $r(i)$ is the rank of $\|X_i - x_0\|$ among the random variables $X_1, \dots, X_n$. The asymptotic normality of the k_n-NN conditional quantile estimator $\widehat{Q}_n^{\mathrm{KNN}}(\alpha_n|x_0)$ is established in the following result.

Corollary 4 *Let $x_0 \in \mathbb{R}^p$ such that $f(x_0) > 0$. Under* **(H)**, *for sequences $k_n \to \infty$ and $\alpha_n \in (0,1)$ such that $k_n\alpha_n \to \infty$, $\alpha_n^{-1} k_n g^2(\alpha_n) \to \infty$ and*

$$(k_n\alpha_n) \sup_{\substack{|\beta/\alpha_n - 1| \leq \xi \\ \|x - x_0\| \leq D_{(k_n)}(x_0)}} \left| \frac{S[Q(\beta|x_0)|x]}{\beta} - 1 \right|^2 \xrightarrow{\mathbb{P}} 0,$$

for some $\xi \in (0,1)$, we have

$$\frac{g(\alpha_n)}{\alpha_n} \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1})} (k_n\alpha_n)^{1/2} \left(\frac{\widehat{Q}_n^{\mathrm{KNN}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{(\ell+1)^2}{2\ell+1}\right).$$

The asymptotic variance $(\ell+1)^2/(2\ell+1)$ is an increasing function of ℓ and thus the best choice (at least in term of variance) seems to be $\ell = 0$, i.e., when the same weight $1/k_n$ is affected to the k_n observations closest to x_0.

Linear combination of weights – We finally focus on the estimator $\widehat{Q}_n^{\mathrm{LC}}(\alpha_n|x_0)$ of $Q(\alpha_n|x_0)$ obtained by using the LC weights introduced in (2.20).

Corollary 5 *Let $x_0 \in \mathbb{R}^p$ such that f is continuous at x_0 and $f(x_0) > 0$. Let $h_n \to 0$, $k_n \to \infty$ and α_n be sequences such that $nh_n^p/\ln\ln n \to \infty$, $\ell_n\alpha_n \to \infty$ with $\ell_n := (nh_n^p \wedge k_n)$, $\alpha_n^{-1}\ell_n g^2(\alpha_n) \to \infty$ and*

$$(\ell_n\alpha_n) \sup_{\substack{|\beta/\alpha_n - 1| \leq \xi \\ \|x - x_0\| \leq (h_n \vee D_{(k_n)}(x_0))}} \left| \frac{S[Q(\beta|x_0)|x]}{\beta} - 1 \right|^2 \xrightarrow{\mathbb{P}} 0,$$

for some $\xi \in (0,1)$. Under **(H)** *and if there exists $\kappa \in [0,\infty]$ such that $k_n/(nh_n^p) \to \kappa$, we have*

$$\frac{g(\alpha_n)}{\alpha_n} \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1})} (\ell_n\alpha_n)^{1/2} \left(\frac{\widehat{Q}_n^{\mathrm{LC}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{C^2(\kappa)}{2^p f(x_0)}\right).$$

In practice, one can take $k_n = \lfloor \kappa n h_n^p \rfloor$ with $\kappa > 0$. The parameter κ is thus a tuning parameter that has to be chosen by a data-driven procedure (see Section 2.5.1).

2.5 Simulation study

In this section, we are interested in the finite sample behavior of the estimator $\widehat{Q}_n(\alpha_n|x_0)$ defined in (2.1) for a given value of x_0. The random weights $\{\mathcal{W}_{n,1}(x_0), \ldots, \mathcal{W}_{n,n}(x_0)\}$ used in the expression of the estimator (2.17) of the conditional survival function $S(\cdot|x_0)$ often depend on an hyperparameter $\lambda_n \in \mathbb{R}^d$, $d \in \mathbb{N} \setminus \{0\}$, useful in order to control the smoothness of the estimator. This is the case for instance for the NW weights, the k_n-NN random weights or the LC weights defined in (2.18), (2.19) and (2.20), where λ_n is equal to h_n, k_n and (h_n, κ), respectively. In the next section, we propose an adaptive procedure to select λ_n in practice.

2.5.1 Choice of the hyperparameter

For $t \in \mathbb{R}^p$, let us denote by $\widehat{Q}_n(\alpha_n|t, \lambda_n)$ an estimator of $Q(\alpha_n|t)$ depending on an hyperparameter λ_n and by $\widehat{Q}_{n,-i}(\alpha_n|t, \lambda_n)$ the estimator computed without the random pair (X_i, Y_i).

Our procedure of selection is based on the following simple remark: for a good choice of λ_n, the random value $S[\widehat{Q}_{n,-1}(\alpha_n|X_1, \lambda_n)|X_1]$ should be close to α_n at least when the observed value of X_1 is close to x_0. We thus propose to define our *optimal* value of λ_n as $\lambda_{\text{opt}} := \arg\min\{\Lambda_n^2(\lambda),\ \lambda \in \mathbb{R}^d\}$, with

$$\Lambda_n(\lambda) := \mathbb{E}\left[\frac{\mathcal{W}_{n,1}(x_0)}{\mathbb{E}\left[\mathcal{W}_{n,1}(x_0)\right]} S[\widehat{Q}_{n,-1}(\alpha_n|X_1, \lambda)|X_1]\right] - \alpha_n.$$

Note that the proximity of X_1 and x_0 is controlled by the random weight $\mathcal{W}_{n,1}(x_0)$. Of course, the function Λ_n is unknown in practice and should be estimated. We propose to use the following estimator

$$\hat{\Lambda}_n(\lambda) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \mathbb{I}_{\{Y_i > \widehat{Q}_{n,-i}(\alpha_n|X_i, \lambda)\}} - \alpha_n. \tag{2.27}$$

The estimated optimal value of the hyperparameter λ_n is thus given by

$$\widehat{\lambda}_{n,\text{opt}} := \arg\min\{\hat{\Lambda}_n^2(\lambda),\ \lambda \in \mathbb{R}^d\}. \tag{2.28}$$

The estimator (2.27) can be motivated by the following result.

Proposition 10 *If there exists a function $\varphi : \mathbb{R}^p \times \mathbb{R}^{p\times(n-1)} \mapsto [0,\infty)$ such that for all $i = 1,\dots,n$, $\mathcal{W}_{n,i}(x_0) = \varphi(X_i, \mathbb{X}_{-i})$ where the matrix $\mathbb{X}_{-i}$ is given by $[X_1,\dots,X_{i-1},X_{i+1},\dots,X_n]$ then $\mathbb{E}[\hat\Lambda_n(\lambda)] = \Lambda_n(\lambda)$ for all $\lambda \in \mathbb{R}^d$.*

Note that the assumption of Proposition 10 is satisfied for the NN approach with the function φ defined for $t \in \mathbb{R}^p$ and $u = [u_1,\dots,u_{n-1}] \in \mathbb{R}^{p\times(n-1)}$ by $\varphi(t,u) = \lambda^{-1}\mathbb{I}_{\{\|t-x_0\|<d_{(\lambda)}(x_0)\}}$, where $d_i(x_0) = \|u_i - x_0\|$, $i = 1,\dots,n-1$ and $d_{(1)}(x_0) \le \ldots \le d_{(n-1)}(x_0)$ are the corresponding ordered values.
This is also the case for the NW weights by using the function

$$\varphi(t,u) = K[(t-x_0)/\lambda] \Bigg/ \left(\sum_{i=1}^{n-1} K[(u_i - x_0)/\lambda] + K[(t-x_0)/\lambda]\right).$$

2.5.2 Finite sample behavior

Using a sample of size n from the random vector (X,Y), we are interested in estimating an extreme conditional quantile in the situation where the quantile level α_n is not too small. We consider the situation where X is a real-valued random variable ($p = 1$). In a theoretical point of view, we assume that the conditions of Theorem 10 are satisfied for such a sequence α_n. In practice, we take $\alpha_n = 20/n$ and the quantile $Q(\alpha_n|x_0)$ is estimated using (2.1). Three sets of random weights are considered:
i) the NW weights with the Epanechnikov kernel: $K(u) = \frac{3}{4}(1-u^2)\mathbb{I}_{[0,1]}(|u|)$,
ii) the k_n-NN weights with $\ell = 1$ (triangular k_n-NN weights),
iii) the LC weights given in (2.20) with $\tau = 1/2$ and $k_n = \lfloor \kappa n h_n \rfloor$.

Although the theory on our estimators is valid without any assumption on the distribution of X, from a practical point of view, the estimation is very difficult in case of unbounded distribution, especially at the border. For this reason, we illustrate our methodology in the case of a bounded distribution, namely the standard uniform distribution. The four following models have been considered for the conditional survival distribution function of Y given X:

M1 – Conditional Burr distribution:

$$S(y|X) = \left(1 + y^{-\rho/\gamma(X)}\right)^{1/\rho}, y > 0,$$

where $\rho < 0$ and for all $x \in (0,1)$, $\gamma(x) = 2x(1-x)$.

It is well-known that for this model, condition (2.13) holds (i.e., condition **(H)** with $g(\alpha|x_0) = \alpha$, see, e.g., Embrechts et al. [12] Table 3.4.2). The parameter ρ is referred in the literature to as the second order parameter and it affects the bias of the estimator.

M2 – Conditional Beta distribution with parameters $\theta_1 > 0$ and $\theta_2(X)$ where for all $x \in (0,1)$, $\theta_2(x) = 1/[2x(1-x)]$.

This conditional distribution satisfies condition (2.13) with a conditional extreme value index given by $\gamma(x) = -1/\theta_2(x) < 0$ (see, e.g., Embrechts et al. [12] Table 3.4.3).

M3 – Conditional Gaussian distribution with mean $\mu(X) = 2X(1-X)$ and variance σ^2.

Under this model, condition (2.13) is satisfied with $\gamma(X) = 0$ (see, e.g., Embrechts et al. [12] Table 3.4.4).
We finally consider a model for which condition (2.13) does not hold.

M4 – Conditional super heavy-tailed distribution:

$$S(y|X) = \exp\left\{-\xi[\ln(y)]^{\theta(X)}\right\}, y > 1,$$

with $\xi > 0$ and $\theta(x) = 19(x+1/2)(3/2-x)/20 \in [0, 0.95]$.

One can check that this conditional distribution satisfies condition **(H)** with

$$a(\alpha^{-1}) = Q(\alpha|x) = \exp\left\{\left[\frac{\ln(1/\alpha)}{\xi}\right]^{1/\theta(x)}\right\}$$

$$\text{and} \quad g(\alpha) = \alpha\theta(x)\xi\left[\frac{\ln(1/\alpha)}{\xi}\right]^{1-1/\theta(x)}.$$

For each model, $N = 500$ samples of size $n = 1000$ have been generated. The hyperparameter λ_n is chosen according to (2.28) and the minimization is achieved

- over a regular grid $\mathcal{H}$ of 20 points evenly spaced between 0.05 and 0.3 for the NW weights,

- over a grid $\mathcal{K}$ of 20 points evenly spaced between 100 and 600 for the NN weights,

- over the grid $\mathcal{H} \times \mathcal{F}$ where $\mathcal{F}$ is a grid of 5 evenly spaced points between 0.9 and 1.1.

The accuracy of the estimators is measured by the errors

$$RMSE := \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\widehat{Q}_n^{\bullet,\mathrm{i}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1\right]^2} \text{ and } ARE := \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\widehat{Q}_n^{\bullet,\mathrm{i}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1\right|,$$

where $\bullet$ has to be replaced by NW, NN or LC and the index i refers to the $i-$simulation run. The error RMSE corresponds to the root mean squared error of the ratio between the estimates and the true quantile value. The error ARE is the average over all replications of the absolute value of the relative error. The estimation of $Q(\alpha_n|x_0)$ is done at three different positions: $x_0 := x_0^{(1)} = (1-\sqrt{1/3})/2 \approx 0.211$, $x_0 = x_0^{(2)} = 1/2$ and $x_0 = x_0^{(3)} = (1+\sqrt{1/2})/2 \approx 0.854$. The results are gathered in Tables 2.1 to 2.4 (see Section 2.8). Based on these simulations, we can draw the following conclusions:

- The three methods, NW, NN and LC, perform similarly for the models **M1**-**M3**;

- Concerning model **M1**, the errors (RMSE and ARE) increase as $|\rho|$ decreases. This is expected since the estimation is much more difficult when ρ is close to 0 where a bias in the estimation appears. Also the errors increase in general when $\gamma(\cdot)$ increases;

- Concerning model **M2**, both RMSE and ARE increase with θ_2, i.e., when $\gamma(\cdot) = -1/\theta_2(\cdot)$ increases, and decreases with θ_1. Compared to the model **M1**, the RMSE and ARE are considerably smaller, but this is not surprising since the conditional extreme value index is negative in model **M2**, which means that the observations are bounded;

- Concerning model **M3**, RMSE and ARE are not too much sensitive on the values of σ, nor on x_0. In general the orders of the errors are intermediate between those obtained in the case $\gamma(\cdot) > 0$ (model **M1**) and $\gamma(\cdot) < 0$ (model **M2**);

- Concerning model **M4**, RMSE and ARE depend a lot on the value of ξ. Indeed, if ξ is too small, both RMSE and ARE increase drastically and in that case the variability of the results is probably too large to allow a more precise interpretation of the results. For larger values of ξ ($\xi = 1$ or $3/2$), the errors are more reasonable, although larger than for the others models. In that case, a slight increase in $\theta(\cdot)$ implies in general a decrease in RMSE and ARE.

To complete the simulation study, we compare in Figure 2.1 the boxplots of the estimates of $Q(\alpha_n|x_0)$ with the three weights (NW, NN and LC) for model **M1** when $\rho = -1/2$, which corresponds to a difficult case, and $x_0 = x_0^{(3)}$. The horizontal line indicates the true value of the conditional quantile. As is clear from this figure, the three methods perform similarly and well, with almost no bias and a sampling distribution of the estimates symmetric. Since the boxplots for the other considered cases (model and values of x_0 and parameters) are similar, they are omitted.

2.6 Real data analysis

As an illustration, we consider in this section the world catalogue of earthquake from 2002 until 2017 which contains information such as the longitude, latitude and seismic moment of earthquakes. The seismic moment denoted by M_S is a physical quantity which illustrates the severity of an earthquake. It is a measure of the energy released by a seism and whose unit is the dyne-centimeters. The dataset considered in this section, of size 15000, is part of the Global Centroid Moment Tensor database, which can be uploaded freely on http://www.globalcmt.org/CMTsearch.html (Dziewonski et al. [10]; Ekström et al. [11]). Note also that this database has already been used in the extreme value framework, but on different periods, by Goegebeur et al. [22, 23]. Being able to model accurately the tail of the earthquake energy distribution is clearly of interest since severe earthquakes may cause important damage and serious losses.
Although we want to study the tail behavior at a specific, fixed, location, the extreme conditional quantiles estimates have to take into account that earthquakes happen at a random location. Thus, this dataset is particularly suited for illustration of our local estimation method. Note that the scientists prefer to convert the seismic moment M_S into the magnitude moment M_W, defined as

$$M_W = \frac{2}{3}\ln_{10}(M_S) - \frac{32}{3}$$

which is a dimensionless value. A value $M_W > 9$ indicates an extreme earthquake which may cause severe damages and losses whereas a value $M_W < 6$ corresponds to a moderated one. Our interest is thus on the distribution of M_W given the location (in latitude and longitude) of the earthquake. The five-number summary of M_W is given below:

Min.	1st Qu.	Median	3rd Qu.	Max.
5.224	5.617	5.778	6.052	9.75

It appears that between 2002 and 2017, approximatively 75% of the earthquakes can be classified as *moderate*. Concerning the points in the covariate space where we want to do our estimation, we use locations where an earthquake has already happened. In order to determine the neighborhood of these locations, we compute the distance in kilometers to every other earthquake position using the formula

$$R\,\mathrm{Arcos}\left(\cos(\phi_1)\cos(\phi_2)\cos(\phi_1-\phi_2)+\sin(\psi_1)\sin(\psi_2)\right),$$

which gives the spherical distance between two points with longitude and latitude (ϕ_1, ψ_1) and (ϕ_2, ψ_2), respectively, expressed in radian (see, e.g., Weisstein [40]). Here, it is assumed that the earth is a perfect sphere, with radius $R = 6371$km.

We estimate the extreme quantile of level $\alpha_n = 20/15000$, and the hyperparameters are selected as described in Section 2.5. The same grid as the one used in Goegebeur et al. [22], i.e., $\mathcal{H} = \{200, 300, ..., 2000\}$, has been used for the NW weights, and for the NN weights, we use a grid $\mathcal{K}$ of 19 evenly spaced points between 1 and 50. Note that the LC method is not considered here since it does not outperform the others two methods as seen in Section 2.5. The level plot of our quantile estimates is given in Figure 2.2 for the NW (left panel) and NN (right panel) weights, respectively. Note that this figure focuses on the Asia-Pacific region, since it is part of the well-known Ring of Fire, an area where many earthquakes and volcanic eruptions occur. The two panels of the figure are slightly different but, as expected, we can observe in both level plots that the seismic activity is intense, especially in Japan and Thailand where we can observe earthquakes with magnitude moment beyond 9. Finally, among all extreme quantile estimates of level 20/15000 calculated with NW weights (resp. NN weights), we have a proportion of 1.5% (resp. 1.25%) for which $M_W > 9$ and 60.75% (resp. 61.25%) for which $M_W < 7$.

2.7 Proofs

2.7.1 Proof of the results given in Section 2.3

Proof of Proposition 5 –
1. Since S is decreasing and ϕ^{-1} is a continuous function, statement 1. is a direct consequence of Dini's theorem.
2. It suffices to remark that from the first statement, one has for all $t_0 \in I$,

$$\lim_{(t,x)\to(t_0,x^*)} \Psi(x)\left(\frac{S[x+td(x)]}{S(x)} - 1\right) = \lim_{t\to t_0} \phi^{-1}(t). \qquad \blacksquare$$

Proof of Proposition 6 – We first prove that condition (2.9) implies condition (2.8). From de Haan and Ferreira [25] Lemma 1.1.1, one has, for all $t \in I$,

$$\lim_{x\to x^*} \frac{S[x+ta(1/S(x))] - S(x)}{g[S(x)]} = \phi^{-1}(t).$$

Taking $a_n = d(Q(1/n))$, $b_n = Q(1/n)$ and $c_n = \Psi(Q(1/n))$, we easily show that 1. $\Rightarrow$ 3.
Finally, let us prove that 3. $\Rightarrow$ 2. From de Haan and Ferreira [25] Lemma 1.1.1, we have that for all $t \in J$,

$$\lim_{n\to\infty} \frac{Q[n^{-1}(1+tc_n)] - b_n}{a_n} = \phi(t). \tag{2.29}$$

Hence, since Q is decreasing and $\lfloor\alpha^{-1}\rfloor \le \alpha^{-1} < \lfloor\alpha^{-1}\rfloor + 1$,

$$Q\left(\frac{1+tc_{\lfloor\alpha^{-1}\rfloor}}{\lfloor\alpha^{-1}\rfloor}\right) \le Q[\alpha(1+tc_{\lfloor\alpha^{-1}\rfloor})] \le Q\left(\frac{1+tc_{\lfloor\alpha^{-1}\rfloor}}{\lfloor\alpha^{-1}\rfloor+1}\right). \tag{2.30}$$

Using (2.29), we know that

$$\frac{1}{a_{\lfloor\alpha^{-1}\rfloor}}\left[Q\left(\frac{1+tc_{\lfloor\alpha^{-1}\rfloor}}{\lfloor\alpha^{-1}\rfloor}\right) - b_{\lfloor\alpha^{-1}\rfloor}\right] \to \phi(t). \tag{2.31}$$

Moreover,

$$Q[(\lfloor\alpha^{-1}\rfloor+1)^{-1}(1+tc_{\lfloor\alpha^{-1}\rfloor})] = Q\{\lfloor\alpha^{-1}\rfloor^{-1}[1+tc_{\lfloor\alpha^{-1}\rfloor}\xi_t(\lfloor\alpha^{-1}\rfloor)]\},$$

where for all $m \in \mathbb{N}$,

$$\xi_t(m) := \frac{m}{1+m}\left(1 - \frac{1}{tmc_m}\right).$$

Since $mc_m \to \infty$, we have $\xi_t(m) \to 1$ as $m \to \infty$. Dini's theorem together with (2.29) entail that

$$\frac{1}{a_{\lfloor \alpha^{-1} \rfloor}} \left[Q\left(\frac{1 + tc_{\lfloor \alpha^{-1} \rfloor}}{\lfloor \alpha^{-1} \rfloor + 1} \right) - b_{\lfloor \alpha^{-1} \rfloor} \right] \to \phi(t). \tag{2.32}$$

Hence, by collecting (2.30), (2.31) and (2.32) we obtain

$$\frac{Q[\alpha + tg(\alpha)] - b(\alpha)}{a(\alpha^{-1})} \to \phi(t), \tag{2.33}$$

with $g(\alpha) = \alpha c_{\lfloor \alpha^{-1} \rfloor}$, $b(\alpha) = b_{\lfloor \alpha^{-1} \rfloor}$ and $a(\alpha^{-1}) = a_{\lfloor \alpha^{-1} \rfloor}$. Using twice the convergence (2.33) yields

$$\frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} \to \phi(t) - \phi(0) = \phi(t). \qquad \blacksquare$$

Proof of Proposition 7 – It suffices to remark that

$$\frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} = \frac{Q'(\alpha) g(\alpha)}{a(\alpha^{-1})} \int_0^t \frac{Q'[\alpha + sg(\alpha)]}{Q'(\alpha)} ds.$$

The local uniform convergence (2.11) concludes the proof. $\blacksquare$

Proof of Proposition 8 – From Proposition 6, the TFO condition entails that $nc_n S(a_n t + b_n) - c_n \to \phi^{-1}(t)$ as $n \to \infty$ with $c_n = 1/[ng(1/n)]$, $a_n = a(n)$ and $b_n = Q(1/n)$. First assume that $\alpha/g(\alpha) \to c$ as $\alpha \to 0$ with $c > 0$. We have that $c_n \to c > 0$ as $n \to \infty$ and thus $nS(a_n t + b_n) \to 1 + \phi^{-1}(t)/c$. In particular, we have that $S(a_n t + b_n) \to 0$ and thus that, letting $F := 1 - S$, $-nS(a_n t + b_n) \sim \ln F^n(a_n t + b_n)$ as $n \to \infty$. Hence,

$$\lim_{n \to \infty} F^n(a_n t + b_n) = G(t) = \exp\left[-\left(1 + \frac{\phi^{-1}(t)}{c} \right) \right],$$

showing that condition (2.12) is satisfied. Now, let us consider the case $c = 0$. From Proposition 6, we have $nc_n S(a_n t + b_n) \to \phi^{-1}(t)$. Let $m_n := nc_n = 1/g(1/n) =: \tilde{g}(n)$. Since $g(\alpha) \to 0$ as $\alpha \to 0$, $m_n \to \infty$ as $n \to \infty$. Since g is a continuous and increasing function, we have that $\tilde{g}^{-1}(m) \to \infty$ as $m \to \infty$. Letting $\tilde{a}_m := a_{\tilde{g}^{-1}(m)}$ and $\tilde{b}_m := b_{\tilde{g}^{-1}(m)}$ we obtain the convergence

$$\lim_{m \to \infty} mS(\tilde{a}_m t + \tilde{b}_m) = \phi^{-1}(t).$$

The end of the proof is similar to the one in the case $c > 0$. $\blacksquare$

Proof of Proposition 9 – Let us show that (2.16) implies (2.15), the converse being similar. Let $g(\alpha) = \alpha^2 b(\alpha^{-1})$. Since $g(\alpha)/\alpha \to 0$ as $\alpha \to 0$, one has for all $t \in \mathbb{R}$

$$\Delta(\alpha, t) := \frac{\alpha}{g(\alpha)} \left[\left(1 + t\frac{g(\alpha)}{\alpha}\right)^{-1} - 1 \right] \to -t,$$

as $\alpha \to 0$. Hence,

$$\frac{Q[\alpha + tg(\alpha)]}{Q(\alpha)} = \frac{U[\alpha^{-1} + b(\alpha^{-1})\Delta(\alpha, t)]}{U(\alpha^{-1})}.$$

From Dini's theorem, the convergence (2.16) is locally uniform leading to (2.15). ■

2.7.2 Proof of Theorem 9

As explained in Section 2.4.2, the asymptotic normality of the conditional survival estimator is established in two steps: a) prove the asymptotic normality of the variance term and b) show that the bias term is negligible. These two steps are based on technical results given below.

The first step is a direct consequence of the following lemma.

Lemma 1 *Let* $\{V_{n,1}, V_{n,2}, \dots, V_{n,n}\}$ *be a triangular array of independent copies of a centered random variable* V_n. *Assume that* $\mathbb{E}(V_n^2) = 1$ *and* $\mathbb{E}(|V_n|^3) < \infty$. *Let* $\mathcal{T}_n := \{T_{n,i},\ 1 \le i \le n\}$ *be a triangular array of positive random variables independent of the* $V_{n,i}$ *and such that* $T_{n,1}^2 + \dots + T_{n,n}^2 = 1$. *For* $T_n := \max\{T_{n,i},\ 1 \le i \le n\}$, *if* $\mathbb{E}(|V_n|^3)T_n \xrightarrow{a.s.} 0$ *then*

$$\sum_{i=1}^{n} T_{n,i} V_{n,i} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof of Lemma 1 – Let $\{t_{n,i},\ i = 1, \dots, n\}$ be a triangular array of real numbers satisfying

$$\min\left(t_{n,i};\ i = 1, \dots, n\right) \ge 0 \text{ and } \sum_{i=1}^{n} t_{n,i}^2 = 1. \tag{2.34}$$

Let $t_n := \max(t_{n,i};\ i=1,\dots,n)$ and $\nu_n := \mathbb{E}(|V_n|^3)$. In a first step, let us show that if $\nu_n t_n \to 0$ as $n\to\infty$ then, for all $z\in\mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^n t_{n,i}V_{n,i} \le z\right) = \Phi(z), \tag{2.35}$$

where Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$ distribution. Since the $V_{n,i}$ are independent and centered random variables, it suffices to prove that the Lindeberg condition is satisfied, i.e., that

$$\lim_{n\to\infty} \sum_{i=1}^n t_{n,i}^2 \mathbb{E}\left(V_{n,i}^2 \mathbb{I}_{\{t_{n,i}|V_{n,i}|>\varepsilon\}}\right) = 0,$$

for all $\varepsilon > 0$. Since $t_{n,i} \le t_n$ for all $i \in \{1,\dots,n\}$,

$$\sum_{i=1}^n t_{n,i}^2 \mathbb{E}\left(V_{n,i}^2 \mathbb{I}_{\{t_{n,i}|V_{n,i}|>\varepsilon\}}\right) \le \sum_{i=1}^n t_{n,i}^2 \mathbb{E}\left(V_{n,i}^2 \mathbb{I}_{\{t_n|V_{n,i}|>\varepsilon\}}\right) = \mathbb{E}\left(V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}\right),$$

since the $V_{n,i}$ are identically distributed and under (2.34).
Hölder's inequality entails that $\mathbb{E}\left(V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}\right) \le \nu_n^{2/3}\left[\mathbb{P}\left(t_n|V_n|>\varepsilon\right)\right]^{1/3}$. Chebyshev's inequality ensures that $\mathbb{P}\left(t_n|V_n|>\varepsilon\right) \le t_n^2/\varepsilon^2$ and thus $\mathbb{E}\left(V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}\right) \le [\nu_n t_n/\varepsilon]^{2/3} \to 0$, as $n\to\infty$, by assumption. Convergence (2.35) is thus proved for all triangular array $\{t_{n,i},\ i=1,\dots,n\}$ satisfying (2.34) with $\nu_n t_n \to 0$.
Now, remark that for all $\omega \in \{\nu_n T_n \to 0\}$, convergence (2.35) entails that

$$\begin{aligned} & \left|\mathbb{P}\left(\sum_{i=1}^n T_{n,i}V_{n,i} \le z \,\middle|\, \{T_{n,i} = T_{n,i}(\omega);\ i=1,\dots,n\}\right) - \Phi(z)\right| \\ = & \left|\mathbb{P}\left(\sum_{i=1}^n T_{n,i}(\omega)V_{n,i} \le z\right) - \Phi(z)\right| \to 0, \end{aligned}$$

as $n\to\infty$. Note that the last equality is true since the $T_{n,i}$ are independent of the $V_{n,i}$. Hence, since $\mathbb{P}[\nu_n T_n \to 0] = 1$,

$$\lim_{n\to\infty} \left|\mathbb{P}\left(\sum_{i=1}^n T_{n,i}V_{n,i} \le z \,\middle|\, \{T_{n,i};\ i=1,\dots,n\}\right) - \Phi(z)\right| = 0 \text{ a.s.} \tag{2.36}$$

To conclude the proof, let us remark that

$$\begin{aligned}
&\lim_{n\to\infty}\left|\mathbb{P}\left(\sum_{i=1}^{n} T_{n,i}V_{n,i}\le z\right)-\Phi(z)\right| \\
\le\;& \lim_{n\to\infty}\mathbb{E}\left[\left|\mathbb{P}\left(\sum_{i=1}^{n} T_{n,i}V_{n,i}\le z\,\middle|\,\{T_{n,i};\ i=1,\dots,n\}\right)-\Phi(z)\right|\right] \\
\le\;& \mathbb{E}\left[\lim_{n\to\infty}\left|\mathbb{P}\left(\sum_{i=1}^{n} T_{n,i}V_{n,i}\le z\,\middle|\,\{T_{n,i};\ i=1,\dots,n\}\right)-\Phi(z)\right|\right]=0,
\end{aligned}$$

by the dominated convergence theorem and (2.36). ■

We can now establish the asymptotic normality of the variance term. Let $\sigma_n^2(x_0) := S(y_n(x_0)|x_0)[1-S(y_n(x_0)|x_0)]$ and recall that $m_n(x_0)$ is a sequence such that $n_{x_0}/m_n(x_0) \xrightarrow{a.s.} 1$ and that $v_n^2(x_0) = m_n(x_0)/S(y_n(x_0)|x_0)$.

Proposition 11 *For $x_0 \in \mathbb{R}^p$, let $y_n(x_0)$ be a sequence converging to the right endpoint $y^*(x_0)$ of the conditional distribution of Y given $\{X = x_0\}$. If condition (2.21) holds then* $v_n(x_0)\left(\widehat{S}_n^{x_0}\left(y_n(x_0)\right) - S\left(y_n(x_0)|x_0\right)\right) \xrightarrow{d} \mathcal{N}(0,1)$.

Proof of Proposition 11 – Remark that

$$\left(\frac{n_{x_0}}{\sigma_n^2(x_0)}\right)^{1/2}\left(\widehat{S}_n^{x_0}\left(y_n(x_0)\right)-S\left(y_n(x_0)|x_0\right)\right)=\sum_{i=1}^{n}T_{n,i}(x_0)V_{n,i}(x_0),$$

with $T_{n,i}(x_0) := (n_{x_0})^{1/2}\mathcal{W}_{n,i}(x_0)$ and

$$V_{n,i}(x_0) := [\sigma_n(x_0)]^{-1}\left(\mathbb{I}_{\{Y_i^{x_0}>y_n(x_0)\}}-S(y_n(x_0)|x_0)\right).$$

It thus suffices to apply Lemma 1 after remarking that $n_{x_0}/\sigma_n^2(x_0) \overset{a.s.}{\sim} v_n^2(x_0)$ and that

$$\mathbb{E}(|V_{n,1}(x_0)|^3)=\sigma_n^{-1}(x_0)\left\{[S(y_n(x_0)|x_0)]^2+[1-S(y_n(x_0)|x_0)]^2\right\}\sim\sigma_n^{-1}(x_0),$$

as $n\to\infty$, since $S(y_n(x_0)|x_0)\to 0$. ■

The second step of the proof is treated in the following result.

Proposition 12 *Let $x_0 \in \mathbb{R}^p$ and $y_n(x_0)$ be a sequence converging to the right endpoint $y^*(x_0)$ of the conditional distribution of Y given $\{X = x_0\}$. If condition (2.22) holds then* $v_n(x_0)\left(\widehat{S}_n\left(y_n(x_0)|x_0\right) - \widehat{S}_n^{x_0}\left(y_n(x_0)\right)\right) \xrightarrow{\mathbb{P}} 0.$

Proof of Proposition 12 – Let $U_1, \dots, U_n$ be independent uniform random variables independent of the X_i. Since $Y_i^{x_0} = Q(U_i|x_0)$ and $Y_i \stackrel{d}{=} Q(U_i|X_i)$ for all $i \in \{1, \dots, n\}$,

$$B_n(x_0) \stackrel{d}{=} v_n(x_0) \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\left[\mathbb{I}_{(-\infty, S(y_n(x_0)|X_i))} - \mathbb{I}_{(-\infty, S(y_n(x_0)|x_0))}\right](U_i).$$

From Owen [34, Lemma 3.4.5], one has for all $\varepsilon > 0$,

$$\mathbb{P}(|B_n(x_0)| > \varepsilon) \quad \leq \quad \varepsilon + \mathbb{P}\left\{ v_n(x_0) \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \mathbb{E}\left[\Delta_{n,i}(x_0) \middle| \mathbb{X}\right] > \varepsilon^2 \right\},$$

where $\mathbb{X} := (X_1, \dots, X_n)$ and

$$\Delta_{n,i}(x_0) := \left|\mathbb{I}_{(-\infty, S(y_n(x_0)|X_i))} - \mathbb{I}_{(-\infty, S(y_n(x_0)|x_0))}\right|(U_i).$$

Introducing the quantity $D_{n,i}(x_0) := |S(y_n(x_0)|X_i) - S(y_n(x_0)|x_0)|$, it is easy to check that $\mathbb{E}[\Delta_{n,i}(x_0)|\mathbb{X}] \leq 2D_{n,i}(x_0)$. Remarking that

$$\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) D_{n,i}(x_0) = W_1\left(\mathcal{W}^*_{n,x_0}, \delta^*_{x_0}\right)$$

leads to $\mathbb{P}(|B_n(x_0)| > \varepsilon) \leq \varepsilon + \mathbb{P}\left[v_n(x_0) W_1\left(\mathcal{W}^*_{n,x_0}, \delta^*_{x_0}\right) > \varepsilon^2/2\right]$. The result is thus proved by using assumption (2.23) (or equivalently (2.22)). ■

Theorem 9 is thus proved by gathering Propositions 11 and 12. ■

2.7.3 Proof of Theorem 10

The proof follows the lines described in Section 2.2. Let us introduce the sequences $t_n^{-1}(x_0) := -v_n(x_0) g(\alpha_n)$ and $\sigma_n^{-1}(x_0) = a(\alpha_n^{-1}) t_n(x_0)$. It is easy to check that for all $z \in \mathbb{R}$,

$$\mathbb{P}\left\{\sigma_n(x_0)[\widehat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)] \leq z\right\} = \mathbb{P}\left\{Z_n(x_0) \leq z_n(x_0)\right\},$$

where $y_n(x_0) := Q(\alpha_n|x_0) + \sigma_n^{-1}(x_0)z$, $z_n(x_0) = v_n(x_0)[\alpha_n - S(y_n(x_0)\big|x_0)]$ and $Z_n(x_0) := v_n(x_0)[\widehat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0)]$. From Proposition 5, condition **(H)** entails that for all $t_0 \in I$,

$$\lim_{(t,y)\to(t_0,y^*(x_0))} \frac{\Psi(y)}{t}\left(\frac{S[y+td(y)|x_0]}{S(y|x_0)} - 1\right) = \lim_{t\to t_0} \frac{\phi^{-1}(t)}{t}. \tag{2.37}$$

Since $y_n(x_0) = Q(\alpha_n|x_0) + a(\alpha_n^{-1})t_n(x_0)z = Q(\alpha_n|x_0) + d(Q(\alpha_n|x_0))t_n(x_0)z$ with $t_n(x_0) \to 0$ as $n \to \infty$, (2.37) entails that as $n \to \infty$

$$z_n(x_0) \sim -zv_n(x_0)t_n(x_0)g(\alpha_n) = z. \tag{2.38}$$

Now, to prove that $Z_n(x_0) \xrightarrow{d} \mathcal{N}(0,1)$, it suffices to show that conditions (2.21) and (2.22) hold for $y_n(x_0)$. From (2.38),

$$1 - \frac{S[y_n(x_0)|x_0]}{\alpha_n} \sim z\alpha_n^{-1}v_n^{-1}(x_0) = z(\alpha_n m_n(x_0))^{-1/2} \to 0, \tag{2.39}$$

as $n \to \infty$ and thus $S[y_n(x_0)|x_0] \sim \alpha_n$. This entails that condition

$$v_n(x_0) \max_{1\le i\le n} \mathcal{W}_{n,i}(x_0) \xrightarrow{a.s.} 0$$

is equivalent to condition (2.21) with $y_n(x_0)$. It remains to prove condition (2.22). From (2.39), there exists $\xi > 0$ such that for n large enough, $S(y_n(x_0)|x_0) \in [(1-\xi)\alpha_n, (1+\xi)\alpha_n]$. Hence, for n large enough,

$$\sum_{i=1}^n \mathcal{W}_{n,i}(x_0)\left|\frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)} - 1\right| \le \sup_{|\beta/\alpha_n - 1|\le\xi} \sum_{i=1}^n \mathcal{W}_{n,i}(x_0)\left|\frac{S[Q(\beta|x_0)|X_i]}{\beta} - 1\right|,$$

and the proof is complete. ■

2.7.4 Proof of Corollaries 3, 4 and 5

We first recall a useful result dealing with the almost sure convergence of the statistic

$$\hat{f}_n(x) := \frac{1}{nh_n^p}\sum_{i=1}^n K\left(\frac{X_i - x_0}{h_n}\right),$$

which is the kernel estimator of the density f of the random value X. The following result can be found for instance in Dony and Einmahl [8] Corollary 2.1.

Lemma 2 *Let $x \in \mathbb{R}^p$ such that f is continuous at x and $f(x) > 0$. If the kernel K is a bounded density with support included in the unit ball $\mathcal{U}_p$ of $\mathbb{R}^p$ and if $\mathcal{K} := \{K(\gamma(t - \cdot)),\ \gamma > 0,\ t \in \mathbb{R}^p\}$, is a pointwise measurable Vapnik-Chervonenkis (VC) type class of functions from $\mathbb{R}^p$ to $\mathbb{R}$ then for a sequence $h_n \to 0$ such that $nh_n^p / \ln\ln n \to \infty$, we have that $\hat{f}_n(x) \xrightarrow{a.s.} f(x)$.*

Conditions on the family $\mathcal{K}$ of functions are not easy to check in practice. Nevertheless, the measurability condition on $\mathcal{K}$ is satisfied whenever K is right-continuous (see Einmahl and Mason [13]) or K is an indicator function on a cell of $\mathbb{R}^p$ (see van der Vaart and Wellner [38] Example 2.3.4). Concerning the VC condition, it is satisfied for kernel function K such that $K(x) = L(\|x\|)$ where L is of bounded variation (see Giné and Nickl [20] Exercice 3.6.13). For the sake of simplicity, we have preferred to replace in Lemma 2 all the conditions involving the kernel function by the stronger (but simpler to check) condition **(K)**.

Corollaries 3, 4 and 5 are direct consequences of Theorem 10 and of the three following lemmas establishing the asymptotic distribution of the corresponding conditional survival function estimators.

Lemma 3 *Let $x_0 \in \mathbb{R}^p$ such that f is continuous at x_0 and $f(x_0) > 0$ and let K be a kernel satisfying* **(K)**. *For sequences $h_n \to 0$ and $y_n(x_0) \uparrow y^*(x_0)$ such that $nh_n^p[S(y_n(x_0)|x_0) \wedge (\ln\ln n)^{-1}] \to \infty$ and*

$$\sup_{\|x-x_0\| \le h_n} \left| \frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1 \right|^2 = o\left(\frac{1}{nh_n^p S(y_n(x_0)|x_0)} \right),$$

one has

$$(nh_n^p S(y_n(x_0)|x_0))^{1/2} \left(\frac{\widehat{S}_n^{\mathrm{NW}}(y_n(x_0)|x_0)}{S(y_n(x_0)|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\|K\|_2^2}{f(x_0)}\right).$$

Proof of Lemma 3 – Let $\tilde{K} := K^2 / \|K\|_2^2$ where $\|K\|_2^2 := \int_{\mathcal{U}_p} K^2(y)dy$. It is easy to check that $\tilde{K}$ also satisfy condition **(K)**. Hence, Lemma 2 entails that almost surely,

$$\lim_{n\to\infty} \frac{\|K\|_2^2}{nh_n^p} n_{x_0} = \lim_{n\to\infty} \hat{f}_n^2(x_0) \Bigg/ \left[\frac{1}{nh_n^p} \sum_{i=1}^n \tilde{K}\left(\frac{x_0 - X_i}{h_n} \right) \right] = f(x_0).$$

Hence, almost surely, $n_{x_0} \sim f(x_0)nh_n^p/\|K\|_2^2 =: m_n(x_0)$. It is easy to infer that, as soon as $nh_n^p S(y_n(x_0)|x_0) \to \infty$, we have

$$\frac{m_n(x_0)}{S(y_n(x_0)|x_0)} \left(\max_{1\le i\le n} \mathcal{W}_{n,i}^{\mathrm{NW}}(x_0,h_n) \right)^2 \le \frac{f(x_0)}{\|K\|_2^2} \frac{1}{nh_n^p S(y_n(x_0)|x_0)} \frac{\|K\|_\infty^2}{\hat{f}_n^2(x_0)} \xrightarrow{a.s.} 0.$$

Similarly, using Assumption (**K**), we have

$$\sum_{i=1}^n \mathcal{W}_{n,i}^{\mathrm{NW}}(x_0,h_n) \left| \frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)} - 1 \right| \le \sup_{\|x-x_0\|\le h_n} \left| \frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1 \right|$$

from which Lemma 3 follows according to Theorem 1. ∎

Lemma 4 *Let $x_0 \in \mathbb{R}^p$. For sequences k_n and $y_n(x_0)$ such that, as $n \to \infty$, $y_n(x_0) \uparrow y^*(x_0)$, $k_n S(y_n(x_0)|x_0) \to \infty$ and*

$$\sup_{\|x-x_0\|\le D_{(k_n)}(x_0)} \left| \frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1 \right|^2 = o\left(\frac{1}{k_n S(y_n(x_0)|x_0)} \right),$$

with $D_{(k_n)}(x_0) = \|X_{r(k_n)} - x_0\|$, one has

$$(k_n S(y_n(x_0)|x_0))^{1/2} \left(\frac{\widehat{S}_n^{\mathrm{KNN}}(y_n(x_0)|x_0)}{S(y_n(x_0)|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{(\ell+1)^2}{2\ell+1}\right).$$

Proof of Lemma 4 – First, remark that since $k_n \to \infty$ as $n \to \infty$,

$$\frac{(\ell+1)^2}{2\ell+1} \frac{n_{x_0}}{k_n} = \frac{(\ell+1)^2}{k_n(2\ell+1)} \left(\sum_{i=1}^{k_n} i^\ell \right)^2 \Big/ \sum_{i=1}^{k_n} i^{2\ell} \to 1,$$

as $n \to \infty$. Thus, $n_{x_0} \sim m_n(x_0)$ with $m_n(x_0) = (2\ell+1)/(\ell+1)^2 k_n$. As soon as $k_n S(y_n(x_0)|x_0) \to \infty$, we have

$$\frac{m_n(x_0)}{S(y_n(x_0)|x_0)} \left(\max_{1\le i\le n} \mathcal{W}_{n,i}^{\mathrm{NN}}(x_0,k_n) \right)^2 = \frac{2\ell+1}{k_n S(y_n(x_0)|x_0)} \to 0.$$

Using the bound

$$\sum_{i=1}^n \mathcal{W}_{n,i}^{\mathrm{NN}}(x_0,k_n) \left| \frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)} - 1 \right| \le \sup_{\|x-x_0\|\le D_{(k_n)}(x_0)} \left| \frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1 \right|,$$

we prove Lemma 4 by applying Theorem 1. ∎

Lemma 5 *Let* $x_0 \in \mathbb{R}^p$ *such that* f *is continuous at* x_0 *and* $f(x_0) > 0$. *Let* h_n, k_n *and* $y_n(x_0) \uparrow y^*(x_0)$ *be sequences such that* $nh_n^p/\ln\ln n \to \infty$, $\ell_n S(y_n(x_0)|x_0) \to \infty$ *with* $\ell_n := (nh_n^p \wedge k_n)$ *and*

$$\sup_{\|x-x_0\|\le(h_n\vee D_{(k_n)}(x_0))}\left|\frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)}-1\right|^2 = o\left(\frac{1}{\ell_n S(y_n(x_0)|x_0)}\right).$$

If there exists $\kappa \in [0,\infty]$ *such that* $k_n/(nh_n^p) \to \kappa$ *then*

$$(\ell_n S(y_n(x_0)|x_0))^{1/2}\left(\frac{\widehat{S}_n^{\mathrm{LC}}(y_n(x_0)|x_0)}{S(y_n(x_0)|x_0)}-1\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{C^2(\kappa)}{2^p f(x_0)}\right),$$

where $C^2(\kappa) := (1\wedge\kappa^{-1})\left[\kappa\tau^2 + 2^p f(x_0)(1-\tau)^2 + 2\tau(1-\tau)\left(\kappa\wedge 2^p f(x_0)\right)\right].$

■

Proof of Lemma 5 – We start by remarking that

$$\sum_{i=1}^n \mathbb{I}_{[0,1]}\left(\left\|\frac{X_i-x_0}{h_n}\right\|_\infty\right)\mathbb{I}_{[0,1]}\left(\frac{r(i)}{k_n}\right) = k_n\wedge M_n.$$

Then, straightforward calculation shows that

$$n_{x_0}^{-1} = \frac{\tau^2}{M_n} + \frac{2\tau(1-\tau)}{k_n\vee M_n} + \frac{(1-\tau)^2}{k_n}.$$

Next, since by assumption $nh_n^p/\ln\ln n \to \infty$ and since the uniform kernel satisfies condition **(K)**, Lemma 2 ensures that $(2h_n)^{-p}n^{-1}M_n \xrightarrow{a.s.} f(x_0)$ as $n\to\infty$. Hence, as a first conclusion, $n_{x_0} \sim \ell_n 2^p f(x_0) C^{-2}(\kappa) =: m_n(x_0)$ almost surely. Furthermore,

$$\max_{1\le i\le n}\mathcal{W}_{n,i}^{\mathrm{LC}}(x_0,\tau,h_n,k_n) \le \frac{\tau}{M_n} + \frac{1-\tau}{k_n}.$$

Hence, using again the almost sure convergence $(2h_n)^{-p}n^{-1}M_n \to f(x_0)$,

$$\lim_{n\to\infty}\ell_n\max_{1\le i\le n}\mathcal{W}_{n,i}^{\mathrm{LC}}(x_0,\tau,h_n,k_n) = \frac{\tau(\kappa\wedge 1)}{2^p f(x_0)} + (1-\tau)(\kappa^{-1}\wedge 1),$$

almost surely for all $\kappa\in[0,\infty]$. As a consequence, since $\ell_n S(y_n(x_0)|x_0)\to\infty$, condition (2.21) is satisfied. Finally, using the bounds obtained in the proofs of Lemmas 3 and 4, one has

$$\sum_{i=1}^n \mathcal{W}_{n,i}^{\mathrm{LC}}(x_0,\tau,h_n,k_n)\left|\frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)}-1\right| \le \sup_{\|x-x_0\|\le h_n\vee D_{(k_n)}(x_0)}\left|\frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)}-1\right|,$$

and thus condition (2.22) holds. Theorem 1 concludes the proof. ■

2.7.5 Proof of Proposition 10

Proposition 10 is a consequence of the following lemma.

Lemma 6 *Let $(X, Y, Z)^\top$ be a random vector for which (X, Y) and Z are independent. Let g be a measurable function such that $g(X, Y, Z)$ is integrable. One has $\mathbb{E}[g(X, Y, Z)] = \mathbb{E}[\Psi(X, Z)]$, where $\Psi(x, z) := \mathbb{E}[g(x, Y, z)|X = x]$.*

Proof of Lemma 6 – Since (X, Y) and Z are independent

$$\begin{aligned}
\mathbb{E}[g(X, Y, Z)] &= \int\int \left(\int g(x, y, z) \mathbb{P}_Y(dy|X = x) \right) \mathbb{P}_X(dx)\mathbb{P}_Z(dz) \\
&= \int\int \Psi(x, z)\mathbb{P}_X(dx)\mathbb{P}_Z(dz).
\end{aligned}$$

The conclusion follows since X and Z are independent. ■

Proof of Proposition 10 – First remark that the assumption on the weights entails that the $W_{n,i}(x_0)$ are identically distributed. Furthermore, since the $W_{n,i}(x_0)$ sum to 1, it is clear that $\mathbb{E}[W_{n,1}(x_0)] = \ldots = \mathbb{E}[W_{n,n}(x_0)] = 1/n$. It thus remains to show that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{\{Y_i > \widehat{Q}_{n,-i}(\alpha_n|X_i,\lambda)\}}\right] = \mathbb{E}\left[\mathcal{W}_{n,1}(x_0)S[\widehat{Q}_{n,-1}(\alpha_n|X_1, \lambda)|X_1]\right].$$

We apply Lemma 6 with $X = X_1$, $Y = Y_1$, $Z = \mathbb{X}_{-1}$ and $g(t, y, u) = \varphi(t, u)\mathbb{I}_{\{y>\phi(\alpha_n,t,u)\}}$ where the function ϕ is such that

$$\widehat{Q}_{n,-1}(\alpha_n|X_1, \lambda) = \phi(\alpha_n, X_1, \mathbb{X}_{-1}).$$

The conclusion is straightforward since, with the notation of Lemma 6, $\Psi(t, u) = \varphi(t, u)S(\phi(\alpha_n, t, u)|t)$. ■

2.8 Tables and figures

Table 2.1: RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M1**, for three different values of ρ and x_0 and three different weights : Nadaraya-Watson (NW), Nearest Neighbors (NN) and linear combination of both (LC).

	$\rho = -2$			$\rho = -1$			$\rho = -0.5$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	0.20	0.18	0.20	0.21	0.18	0.20	0.22	0.19	0.21
$\gamma(x_0^{(1)}) = 1/3$	0.15	0.14	0.15	0.15	0.14	0.15	0.16	0.15	0.16
$x_0 = x_0^{(2)}$	0.28	0.28	0.31	0.29	0.30	0.31	0.33	0.34	0.35
$\gamma(x_0^{(2)}) = 1/2$	0.20	0.20	0.21	0.21	0.21	0.21	0.23	0.24	0.23
$x_0 = x_0^{(3)}$	0.20	0.20	0.20	0.20	0.21	0.21	0.20	0.20	0.22
$\gamma(x_0^{(3)}) = 1/4$	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16

Table 2.2: RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M2**, for three different values of θ_1 and x_0 and three different weights : Nadaraya-Watson (NW), Nearest Neighbors (NN) and linear combination of both (LC).

	$\theta_1 = 1$			$\theta_1 = 2$			$\theta_1 = 3$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	0.07	0.07	0.07	0.04	0.04	0.05	0.03	0.03	0.03
$\theta_2(x_0^{(1)}) = 3$	0.06	0.06	0.06	0.04	0.04	0.04	0.03	0.03	0.03
$x_0 = x_0^{(2)}$	0.04	0.04	0.04	0.03	0.02	0.03	0.02	0.02	0.02
$\theta_2(x_0^{(2)}) = 2$	0.03	0.03	0.04	0.02	0.02	0.02	0.01	0.01	0.01
$x_0 = x_0^{(3)}$	0.10	0.10	0.10	0.07	0.07	0.07	0.05	0.05	0.06
$\theta_2(x_0^{(3)}) = 4$	0.08	0.08	0.08	0.05	0.06	0.06	0.04	0.04	0.04

Table 2.3: RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M3**, for three different values of σ and x_0 and three different weights : Nadaraya-Watson (NW), Nearest Neighbors (NN) and linear combination of both (LC).

	$\sigma = 1/2$			$\sigma = 1$			$\sigma = 3/2$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.09	0.08
$\mu(x_0^{(1)}) = 1/3$	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06
$x_0 = x_0^{(2)}$	0.06	0.07	0.07	0.07	0.07	0.07	0.08	0.08	0.08
$\mu(x_0^{(2)}) = 1/2$	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06
$x_0 = x_0^{(3)}$	0.08	0.09	0.08	0.08	0.09	0.07	0.08	0.09	0.07
$\mu(x_0^{(3)}) = 1/4$	0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06

Table 2.4: RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M4**, for three different values of ξ and x_0 and three different weights : Nadaraya-Watson (NW), Nearest Neighbors (NN) and linear combination of both (LC).

	$\xi = 1/2$			$\xi = 1$			$\xi = 3/2$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	17.3	19.1	17.1	1.04	1.29	1.01	0.48	0.50	0.45
$\theta(x_0^{(1)}) \approx 0.871$	4.02	4.75	3.93	0.58	0.65	0.58	0.32	0.32	0.30
$x_0 = x_0^{(2)}$	6.98	27.3	25.3	0.96	1.26	1.22	0.47	0.53	0.50
$\theta(x_0^{(2)}) = 0.95$	1.91	3.07	2.81	0.49	0.54	0.50	0.28	0.30	0.27
$x_0 = x_0^{(3)}$	392	295	1414	2.07	2.34	3.91	0.55	0.63	0.91
$\theta(x_0^{(3)}) = 0.83125$	35.0	30.2	102	0.81	0.88	0.97	0.37	0.39	0.41

Figure 2.1: Boxplots of $\widehat{Q}_n^{\mathrm{NW}}(\alpha_n|x_0)$ (A), $\widehat{Q}_n^{\mathrm{NN}}(\alpha_n|x_0)$ (B) and $\widehat{Q}_n^{\mathrm{LC}}(\alpha_n|x_0)$ (C) for the model **M1** when $\rho = -1/2$, $x_0 = (1+\sqrt{1/2})/2$, $\alpha_n = 20/n$ and $n = 1000$. The horizontal line indicates the true value of the conditional quantile.

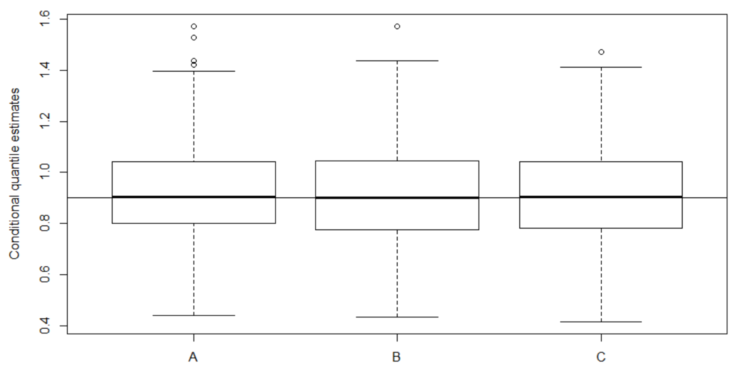
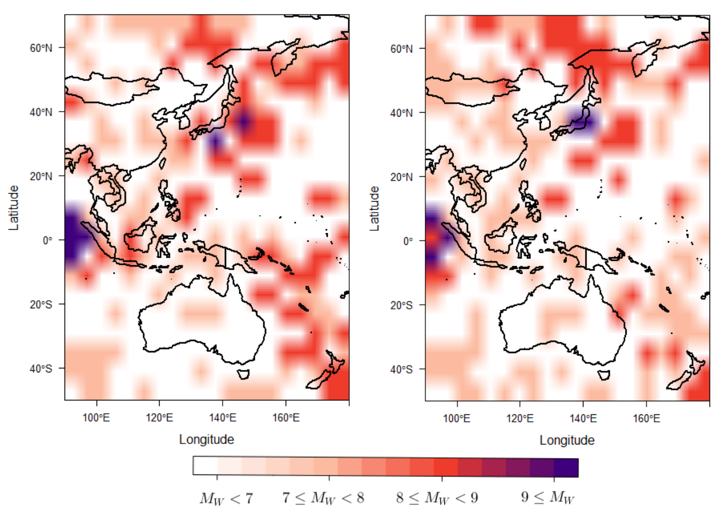


Figure 2.2: Level plot of the conditional extreme quantiles of order 20/15000 in the Asia-Pacific region with NW weights (left panel) and NN weights (right panel).



Chapter 3

Extreme conditional quantile estimation with large-dimensional covariates

3.1 Introduction

One of the main objective of extreme value theory is to predict the statistical properties of events that have never been observed, or at least rarely. This can happen for instance in finance (e.g., to predict crises), in insurance (e.g., to predict large claims due to catastrophic events), or in environmental science (e.g., for particulate matter concentrations). Again, as in Chapter 2, for these types of applications, we are often in a regression context where some covariate information (e.g., geographical location) are recorded simultaneously with the quantity of interest. Thus our aim is still to estimate a α_n-quantile at some target value x_0 on the basis of $m_n(x_0)$ observations. However, the quantile proposed in Chapter 2 is no longer valid since it requires $m_n(x_0)\alpha_n \to \infty$, thus a level which is not enough extreme. To overcome this issue, we need to extrapolate outside the dataset, and thus to estimate an extreme quantile such that $m_n(x_0)\alpha_n \to c \geq 0$. From a theoretical point of view, extreme value theory provides us the solid theoretical framework for this extrapolation and it allows us to define estimators of such quantities and to study their asymptotic properties. However, from the practical point of view, since more and more data are stored nowadays, we are often faced with high dimensional covariates problems. In that situa-

tion, inference on the conditional distribution of Y given X becomes difficult since the space is sparsely populated by data points. This is the well-known curse of dimensionality problem. In this chapter, we propose to combine extreme value theory with dimension reduction methods in order to estimate a conditional extreme quantile based on extrapolation in case of a high dimensional vector of covariates. The combination of these two theories is, as far as we know, still in its infancy. We are only aware of a first attempt by Gardes [18] where a dimension reduction method has been adapted to the study of conditional tail distributions. To deal with high dimensional covariates, a classical method is to assume the existence of a $q \times p$ full rank matrix B_0 (with $q < p$) such that the conditional distributions of Y given $B_0^\top X$ and Y given X are the same. In other words, X and Y are assumed to be independent conditionally on $B_0^\top X$. For a comprehensive discussion on conditional independence, we refer to Basu and Pereira [1]. In the literature, this model is referred to the multiple-index model (single-index model if $q = 1$) and the subspace spanned by the columns of B_0 is called the Dimension Reduction (DR) subspace. Among the contributions on the estimation of the DR subspace, one can cite the Slice Inverse Regression method by Li [31], the Slice Average Variance Estimation method by Cook and Weisberg [5] and the Principal Hessian Directions method, see Li [32]. The existence of a DR subspace is assumed by many authors in order to study the link between Y and the explanatory variable X.
The remainder of the chapter is organized as follows. In Section 3.2, we define our extrapolated conditional quantile estimator, whose main asymptotic properties are established in Section 3.3. The strategy to overcome the problem of the dimension of the vector of covariates is described in Section 3.4, whereas simulations are provided in Section 3.5 to illustrate the efficiency of our estimator. All the proofs are postponed to Section 3.6.

3.2 Extreme conditional quantile estimation

Let (X, Y) be a random vector taking its values in $\mathbb{R}^p \times \mathbb{R}$. For all $x_0 \in \mathbb{R}^p$ let us denote by $S(\cdot|x_0) := \mathbb{P}(Y > \cdot|X = x_0)$ its conditional survival function. Starting from n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), our goal is the estimation of the conditional quantile of order $\beta_n \to 0$ defined for $x_0 \in \mathbb{R}^p$ by

$$Q(\beta_n|x_0) := S^{\leftarrow}(\beta_n|x_0) = \inf\{y \in \mathbb{R};\ S(y|x_0) \leq \beta_n\},$$

with the convention $\inf\{\emptyset\} = +\infty$. In Chapter 2, an estimator of $Q(\alpha_n|x_0)$ has been proposed under the first tail condition (TFO) (see Definition 5). According to Theorem 10, the estimator $\widehat{Q}_n(\alpha_n|x_0)$ defined in (2.1) is consistent if $\alpha_n m_n(x_0) \to \infty$ where $m_n(x_0)$ corresponds, roughly speaking, to the number of observations used in the estimation procedure. This condition is obviously a constraint on the rate of convergence of the level α_n to 0. In what follows, a level α_n satisfying $\alpha_n m_n(x_0) \to \infty$ as $n \to \infty$ is called an intermediate level.

In this chapter, we are interested in the estimation of a conditional quantile $Q(\beta_n|x_0)$ where the level β_n is a proper extreme level, i.e., such that $\beta_n m_n(x_0) \to c \geq 0$.

Estimation with condition (TFO) – Let us assume that condition **(H)** holds. For all sequence $\beta_n < \alpha_n$ where α_n is an intermediate level, one has for n large enough that

$$Q(\beta_n|x_0) \approx Q(\alpha_n|x_0) + a(\alpha_n^{-1}|x_0)\phi_{x_0}(t_n), \tag{3.1}$$

with $t_n = (\beta_n - \alpha_n)/g(\alpha_n|x_0)$. To estimate $Q(\beta_n|x_0)$, a natural way is to replace $Q(\alpha_n|x_0)$, $a(\cdot|x_0)$, $\phi_{x_0}(\cdot)$ and $g(\cdot|x_0)$ by suitable estimators. Of course, the quality of the estimation strongly depends on the accuracy of the approximation (3.1) which depends on the underlying conditional distribution. To illustrate this last point, let us consider the following example. Let (X, Y) be a random vector such that $\mathbb{P}(Y > y) = (1 + y^{1/3})^{-3}$. To simplify, we assume that X and Y are independent (corresponding to the non-conditional situation). Our goal is to evaluate the accuracy of the approximation (3.1) for the random vector $(X, \zeta(Y))$ where $\zeta : (0, \infty) \to \mathbb{R}$ is a continuous and increasing function. Since X and Y are independent, the conditional quantile of $\zeta(Y)$ given X is given by

$$Q_\zeta(\alpha|X) = Q_\zeta(\alpha) = \zeta\left(\left(\alpha^{-1/3} - 1\right)^3\right).$$

Different functions ζ are investigated: for $y > 0$, $\zeta(y) = \zeta_{\mathrm{WT}}(y) := \ln(y)$, $\zeta(y) = \zeta_{\mathrm{HT}}(y|\gamma) := y^\gamma$ for some $\gamma > 0$ and $\zeta(y) = \zeta_{\mathrm{SHT}}(y) := \exp(y)$. For the function ζ_{WT}, condition **(H)** holds with $a(\cdot|x) = 1$, $g(\alpha|x) = \alpha$ and $\phi_x(t) = -\ln(1 + t)$. In other words, for this choice of function ζ, condition (2.13) (see Chapter 2) is satisfied and thus Y belongs to the maximum domain of attraction of G_γ with $\gamma = 0$. In the literature, the distribution of $\ln(Y)$

is called a Weibull-tail distribution (see for instance Beirlant *et al.* [2] and Gardes and Girard [19]). Approximation (3.1) is then given for all $\beta < \alpha$ by

$$Q_{\zeta_{\mathrm{WT}}}(\beta) \approx \tilde{Q}_{\zeta_{\mathrm{WT}}}(\beta|\alpha) := Q_{\zeta_{\mathrm{WT}}}(\alpha) + \ln\left(\frac{\alpha}{\beta}\right).$$

For the function $\zeta_{\mathrm{HT}}(\cdot|\gamma)$ condition (2.13) is also satisfied with $a(\alpha^{-1}|x) = \gamma Q_{\zeta_{\mathrm{HT}}(\cdot|\gamma)}(\alpha)$ and with an extreme-value index given by γ. The distribution of Y is in this case heavy-tailed. For all $\beta < \alpha$, we obtain the approximation

$$Q_{\zeta_{\mathrm{HT}}(\cdot|\gamma)}(\beta) \approx \tilde{Q}_{\zeta_{\mathrm{HT}}(\cdot|\gamma)}(\beta|\alpha) := Q_{\zeta_{\mathrm{HT}}(\cdot|\gamma)}(\alpha)\left(\frac{\alpha}{\beta}\right)^{\gamma}.$$

Finally, choosing the function ζ_{SHT} leads to a super heavy-tailed distribution (see Section 2.3 for more details). Indeed, one can show that in this situation, condition **(H)** holds with $a(\alpha^{-1}|x) = Q_{\zeta_{\mathrm{SHT}}}(\alpha)$ and $g(\alpha|x) = \alpha^2$. The approximation is given for all $\beta < \alpha$ by

$$Q_{\zeta_{\mathrm{SHT}}}(\beta) \approx \tilde{Q}_{\zeta_{\mathrm{SHT}}}(\beta|\alpha) := Q_{\zeta_{\mathrm{SHT}}}(\alpha)\exp\left(\frac{\alpha-\beta}{\alpha^2}\right).$$

The accuracy of the approximation is measured by the relative error function. More specifically, the error between a value v and its approximation v_{app} is defined by

$$\mathrm{erf}(v, v_{\mathrm{app}}) := \max\left(\frac{v}{v_{\mathrm{app}}}, \frac{v_{\mathrm{app}}}{v}\right) - 1.$$

Fixing α to 0.05, this error is represented on Figure 3.1 as a function of β/α for the four functions ζ_{WT}, $\zeta_{\mathrm{HT}}(\cdot|1)$, $\zeta_{\mathrm{HT}}(\cdot|2)$ and ζ_{SHT}. As expected, the approximation deteriorates when β moves away from α. One can note also that the heavier the tail is, worse is the approximation. For super heavy-tailed distribution, the estimation of conditional quantile for very small levels β seems unfortunately hopeless (at least by using approximation (3.1)).
We thus propose to focus on the situation where condition (2.13) holds (i.e., if the conditional distribution belongs to a maximum domain of attraction).

Estimation in a maximum domain of attraction – Condition (2.13) leads to the approximation given by

$$Q(\beta_n|x_0) \approx Q(\alpha_n|x_0) + a(\alpha_n^{-1}|x_0)L_{\gamma(x_0)}\left(\frac{\alpha_n}{\beta_n}\right),$$

where for all $\xi \in \mathbb{R}$ and $x > 0$, $L_\xi(x) = \int_1^x s^{\xi-1} ds$. Replacing the unknown quantity $Q(\alpha_n|x_0)$, $a(\alpha_n^{-1}|x_0)$ and $\gamma(x_0)$ by some estimators denoted by $\check{Q}_n(\alpha_n|x_0)$, $\check{a}_n(\alpha_n^{-1}|x_0)$ and $\check{\gamma}_n(x_0)$, we obtain an estimator of $Q(\beta_n|x_0)$ given by

$$\check{Q}_n^{(E)}(\beta_n|x_0) := \check{Q}_n(\alpha_n|x_0) + \check{a}_n(\alpha_n^{-1}|x_0) L_{\check{\gamma}_n(x_0)}\left(\frac{\alpha_n}{\beta_n}\right). \tag{3.2}$$

The aim of the next section is to establish the consistency of (3.2).

3.3 Main results

Let us assume that the conditional distribution of Y given $X = x_0$ belongs to a maximum domain of attraction with conditional extreme value index $\gamma(x_0)$. Letting $\text{ERV}(\alpha, t|x_0)$ (ERV for extended regular variation) defined by

$$\text{ERV}(\alpha, t|x_0) := \frac{Q(t\alpha|x_0) - Q(\alpha|x_0)}{a(\alpha^{-1}|x_0)} - L_{\gamma(x_0)}\left(t^{-1}\right)$$

we thus have that for all $t > 0$, $\text{ERV}(\alpha, t|x_0) \to 0$ as $\alpha \to 0$. To establish the consistency of (3.2), this condition must be strengthen by the following second order condition.

(H2) There exists a function $A(\cdot|x_0)$ not changing sign and a constant $\rho(x_0) \leq 0$ such that $A(y|x_0) \to 0$ as $y \to \infty$ and for all $t > 0$

$$\lim_{\alpha \to 0} \frac{\text{ERV}(\alpha, t|x_0)}{A(\alpha^{-1}|x_0)} = \int_1^{t^{-1}} r^{\gamma(x_0)-1} L_{\rho(x_0)}(r) dr.$$

We are now in position to state our first main result.

Theorem 11 *Assume that condition* **(H2)** *holds with $\rho(x_0) < 0$ or ($\rho(x_0) \leq 0$ and $\gamma(x_0) < 0$). Suppose also that there exist a sequence $\tau_n(x_0)$ converging to 0 as $n \to \infty$ and such that*

$$\frac{\check{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)), \ \frac{\check{a}_n(\alpha_n^{-1}|x_0)}{a(\alpha_n^{-1}|x_0)} = 1 + \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)),$$

and $\check{\gamma}_n(x_0) - \gamma(x_0) = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$. *If* $\alpha_n/\beta_n \to \infty$, $\tau_n(x_0)\ln(\alpha_n/\beta_n) \to 0$ *and* $A(\alpha_n^{-1}|x_0) = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$ *then*

$$\frac{\check{Q}_n^{(E)}(\beta_n|x_0) - Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$$

where for $t > 1$,

$$q_{\gamma(x_0)}(t) := \int_1^t s^{\gamma(x_0)-1}\ln(s)ds.$$

Application – Let us now apply Theorem 11 to specific estimators of $Q(\alpha_n|x_0)$, $a(\alpha_n^{-1}|x_0)$ and $\gamma(x_0)$. For the conditional quantile $Q(\alpha_n|x_0)$, we consider the estimator $\widehat{Q}_n(\alpha_n|x_0)$ introduced in Chapter 2 which is obtained by inverting the conditional survival function estimator

$$\widehat{S}_n(y|x_0) := \sum_{i=1}^n \mathcal{W}_{n,i}(x_0)\mathbb{I}_{(y,\infty)}(Y_i).$$

According to Theorem 10 in Chapter 2, we have (under some technical assumptions) that

$$[\alpha_n m_n(x_0)]^{1/2}\frac{\widehat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)} \xrightarrow{d} \mathcal{N}(0,1),$$

where we recall that $m_n(x_0) \overset{a.s.}{\sim} n_{x_0}$ with

$$n_{x_0} := \left(\sum_{i=1}^n \mathcal{W}_{n,i}(x_0)\right)^{-1}.$$

Hence, introducing the sequence $\tau_n(x_0) := \{\ln[\alpha_n m_n(x_0)]/[\alpha_n m_n(x_0)]\}^{1/2}$, we have that

$$\frac{\widehat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)). \tag{3.3}$$

For the estimation of $\gamma(x_0)$ and $a(\alpha_n^{-1}|x_0)$, we propose to use the class of statistics introduced in Gardes [17, 18]. Let us first introduce some notations. For $\kappa \in (0,1)$ and $\varphi(\cdot)$ a positive and bounded function on $[\kappa, 1]$, let $\Psi(\cdot)$ be the decreasing function defined for $s \geq 0$ by $\Psi(s) = 0$ and for $s \leq 0$ by

$$\Psi(s) := \left(\int_\kappa^1 \varphi(u)L_s(1/u)du\right)^2 \Big/ \int_\kappa^1 \varphi(u)L_s^2(1/u)du\ .$$

In addition, for all $\delta \in \mathbb{N}$, for all non-increasing right-continuous function $U(\cdot)$ and all $\alpha \in (0,1)$, let

$$\mathcal{T}_\alpha^{(\delta)}(U) := \int_\kappa^1 \varphi(u) \left(\ln \frac{U(u\alpha)}{U(\alpha)} \right)^\delta du \Bigg/ \left(\int_\kappa^1 \varphi(u) L_0(1/u) du \right)^\delta .$$

The estimator $\widehat{\gamma}_n(x_0)$ of $\gamma(x_0)$ is given by

$$\begin{aligned} \widehat{\gamma}_n(x_0) &:= \widehat{\gamma}_{n,+}(x_0) + \widehat{\gamma}_{n,-}(x_0) \\ &= \mathcal{T}_{\alpha_n}^{(1)}(\widehat{Q}_n(\cdot|x_0)) + \Psi^{\leftarrow} \left(\frac{[\mathcal{T}_{\alpha_n}^{(1)}(\widehat{Q}_n(\cdot|x_0))]^2}{\mathcal{T}_{\alpha_n}^{(2)}(\widehat{Q}_n(\cdot|x_0))} \right). \end{aligned} \quad (3.4)$$

Concerning the estimation of $a(\alpha_n^{-1}|x_0)$, we consider the statistic

$$\widehat{a}_n(\alpha_n^{-1}|x_0) = \widetilde{\mathcal{T}}_{\alpha_n} \left(\widehat{Q}_n(\cdot|x_0); \widehat{\gamma}_{n,-}(x_0) \right), \quad (3.5)$$

where $\widetilde{\mathcal{T}}_\alpha(U, \gamma_-)$ is given for all non-increasing and right-continuous function $U(\cdot)$, for all $s \leq 0$ and for all $\alpha \in (0,1)$ by

$$\widetilde{\mathcal{T}}_\alpha(U, s) := U(\alpha) \int_\kappa^1 \varphi(u) \ln \frac{U(u\alpha)}{U(\alpha)} du \Bigg/ \int_\kappa^1 \varphi(u) L_s(1/u) du .$$

Note that the previous defined estimators depend on the choice of a parameter $\kappa \in (0,1)$ and a positive and bounded function $\varphi(\cdot)$ on $[\kappa, 1]$. In order to not overload the notations, this dependence has been omitted. For a motivation of the expressions of $\widehat{\gamma}_n(x_0)$ and $\widehat{a}_n(\alpha_n^{-1}|x_0)$, see Gardes [18] Proposition 2. The consistency of these estimators is established in the next proposition. The following notations are required. For $\alpha \in (0,1)$, let

$$\Delta_a(\alpha|x_0) := \frac{a(\alpha^{-1}|x_0)}{Q(\alpha|x_0)} - \gamma_+(x_0).$$

According to Fraga Alves et al. [16] Lemma 3.1, condition (2.13) entails that $\Delta_a(\alpha|x_0) \to 0$ as $\alpha \to 0$. The next quantity is introduced in order to control the oscillations of the function $x \mapsto S(y|x)$ for large values of y: for $\xi \in (0,1)$ and $\alpha \in (0,1)$, let

$$\omega(\alpha, \xi|x_0) := \sup_{\beta/\alpha \in \mathcal{J}_\xi} \sum_{i=1}^n \mathbb{E} \left[\mathcal{W}_{n,i}(x_0) \left| \frac{S[Q(\beta|x_0)|X_i]}{\beta} - 1 \right| \right],$$

with $\mathcal{J}_\xi := [(1-\xi)\kappa, 1+\xi]$.

Proposition 13 *Assume that there exists a sequence* $m_n(x_0)$ *such that* $m_n(x_0) \overset{a.s.}{\sim} n_{x_0}$ *and let* $\tau_n(x_0) := \{\ln[\alpha_n m_n(x_0)]/[\alpha_n m_n(x_0)]\}^{1/2} \to 0$. *Under condition* **(H2)**, *if* $A(\alpha_n^{-1}|x_0) = o(\tau_n(x_0))$, $\Delta_a(\alpha|x_0) = o(\tau_n(x_0))$, *if there exists a positive constant* C_X *such that*

$$m_n(x_0) \max_{1\leq i\leq n} \mathcal{W}_{n,i}(x_0) < C_X \text{ a.s.} \tag{3.6}$$

and if there exists $\xi > 0$ *such that*

$$[m_n(x_0)\alpha_n]^{1/2}\omega(\alpha_n, \xi|x_0) = o(\tau_n(x_0)), \tag{3.7}$$

then $\widehat{\gamma}_n(x_0) - \gamma(x_0) = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$ *and*

$$\frac{\widehat{a}_n(\alpha_n^{-1}|x_0)}{a(\alpha_n^{-1}|x_0)} = 1 + \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)).$$

Note that, using Markov's inequality, condition (3.7) is a stronger version of condition (2.24) in Theorem 10, Chapter 2. As a consequence, under the assumptions of Proposition 13, (3.3) holds. The consistency of the estimator

$$\widehat{Q}_n^{(E)}(\beta_n|x_0) := \widehat{Q}_n(\alpha_n|x_0) + \widehat{a}_n(\alpha_n^{-1}|x_0)L_{\widehat{\gamma}_n(x_0)}\left(\frac{\alpha_n}{\beta_n}\right),$$

where $\widehat{\gamma}_n(x_0)$ and $\widehat{a}_n(\alpha_n^{-1}|x_0)$ are given in (3.4) and (3.5) is then a direct consequence of Theorem 11 and Proposition 13.

Corollary 6 *Assume that there exists a sequence* $m_n(x_0)$ *such that* $m_n(x_0) \overset{a.s.}{\sim} n_{x_0}$ *and let* $\tau_n(x_0) := \{\ln[\alpha_n m_n(x_0)]/[\alpha_n m_n(x_0)]\}^{1/2} \to 0$. *Assume also that condition* **(H2)** *holds with* $\rho(x_0) < 0$ *or (*$\rho(x_0) \leq 0$ *and* $\gamma(x_0) < 0$*). If* $\alpha_n/\beta_n \to \infty$, $\tau_n(x_0)\ln(\alpha_n/\beta_n) \to 0$, $A(\alpha_n^{-1}|x_0) = o_{\mathbb{P}}(\tau_n(x_0))$ *and* $\Delta_a(\alpha|x_0) = o(\tau_n(x_0))$ *then, under conditions (3.6) and (3.7),*

$$\frac{\widehat{Q}_n^{(E)}(\beta_n|x_0) - Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)).$$

3.4 Estimation for a large-dimensional covariate

When the dimension p of the explicative variable X is large, it is well-known that the data become sparse. As a consequence, the observations selected by the weights $\mathcal{W}_{n,i}(x_0)$, $i = 1, \ldots, n$ (for instance (2.18) or (2.19)) can be located far away from the target x_0. This problem is often referred in the literature to as the *curse of dimensionality*. In a large dimension setting, it is thus necessary to find more appropriate weights. A classical approach is to assume that it exists a $p \times q$ full rank matrix B_0 (with $q < p$) such that the conditional distributions of Y given $B_0^\top X$ and Y given X are the same (Li [31]). In this case, we say that X and Y are independent conditionally on $B_0^\top X$ (in symbols $X \perp\!\!\!\perp Y | B_0^\top X$). As mentioned by Gardes [18], this assumption can be too strong when we are interested in the tail of the conditional distribution. The Conditional Independence assumption is replaced by the Tail Conditional Independence (TCI) assumption. For $\alpha \in (0,1)$ and $z \in \text{supp}(B_0^\top X)$, let $Q_{B_0}(\alpha|z)$ be the quantile of level α associated to the conditional distribution of Y given $B_0^\top X = z$.

(TCI) The random variable Y is tail conditionally independent of X given $B_0^\top X$ if for all $\epsilon > 0$ there exists $\kappa > 0$ such that for all $\delta \in (0, \kappa]$,

$$\mathbb{P}\left[\left|\frac{\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|X)}{\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|B_0^\top X)} - 1\right| \leq \epsilon\right] = 1,$$

where for $\delta > 0$, $\mathcal{Y}_\delta(\cdot)$ is a measurable function defined for all $z \in \text{supp}(B_0^\top X)$ by $\mathcal{Y}_\delta(z) := Q_{B_0}(0|z) - \delta$ if $Q_{B_0}(0|z) < +\infty$ and $\mathcal{Y}_\delta(z) := \delta^{-1}$ if $Q_{B_0}(0|z) = +\infty$.

Note that under (**TCI**), the distributions of Y given X and Y given $B_0^\top X$ share the same right endpoint, i.e., for all $x \in \mathcal{A}$ with $\mathbb{P}(X \in \mathcal{A}) = 1$, $Q_{B_0}(0|B_0^\top x) = Q(0|x)$. In particular, condition (**TCI**) entails that for all $x \in \mathcal{A}$,

$$\lim_{y \uparrow Q(0|x)} \frac{S_{B_0}(y|B_0^\top x)}{S(y|x)} = 1,$$

where $S_{B_0}(y|z) := \mathbb{P}(Y > y|B_0^\top X = z)$. Obviously, for any regular $q \times q$ matrix D, if Y is tail conditionally independent of X given $B_0^\top X$ then Y is also tail conditionally independent of X given $DB_0^\top X$. To ensure the unicity of the matrix B_0, it is assumed in what follows that $B_0 \in \mathcal{B}$ where $B \in \mathcal{B}$ if the q columns of B are the first normalized q linearly independent columns of the orthogonal projection matrix on the subspace spanned by B.

3.4.1 Estimation in the case where B_0 is known

Recall that our goal is to find more appropriate weights $\mathcal{W}_{n,i}(x_0)$ to select the observations. For any norm $\|\cdot\|_p$ on $\mathbb{R}^p$, the weights proposed in Chapter 2 (Nadaraya-Watson (2.18) or Nearest-Neighbors (2.19)) select the observations X_i such that $\|X_i - x_0\|_p$ is closed enough to 0. When p is large, these observations can be located far away from x_0 due to the sparcity of $\mathbb{R}^p$. Under **(TCI)**, it seems more appropriate to select the observations X_i such that $\|B_0^\top(X_i - x_0)\|_q$ is closed enough to 0. Since $q < p$, one can expect that the selected observations will be more relevant for the study of the conditional tail distribution. We thus propose to replace the weights $\mathcal{W}_{n,i}(x_0)$ by some weights $\mathcal{W}_{n,i}(B_0, x_0)$ summing to 1 and depending on the matrix B_0 satisfying **(TCI)**. The conditional quantile of level α is then estimated by

$$\widehat{Q}_n(\alpha|B_0, x_0) := \inf\{y;\ \widehat{S}_n(y|B_0, x_0) \leq \alpha\}, \tag{3.8}$$

where

$$\widehat{S}_n(y|B_0, x_0) := \sum_{i=1}^n \mathcal{W}_{n,i}(B_0, x_0)\mathbb{I}_{(y,\infty)}(Y_i).$$

For a given intermediate sequence α_n, an estimator of $Q(\beta_n|x_0)$ where the level β_n is such that $\beta_n/\alpha_n \to 0$ is given by

$$\widehat{Q}_n^{(E)}(\beta_n|B_0, x_0) := \widehat{Q}_n(\alpha_n|B_0, x_0) + \widehat{a}_n(\alpha_n^{-1}|B_0, x_0)L_{\widehat{\gamma}_n(B_0,x_0)}\left(\frac{\alpha_n}{\beta_n}\right),$$

where $\widehat{a}_n(\alpha_n^{-1}|B_0, x_0)$ and $\widehat{\gamma}_n(B_0, x_0)$ are obtained by replacing in (3.5) and (3.4) the weights $\mathcal{W}_{n,i}(x_0)$ by $\mathcal{W}_{n,i}(B_0, x_0)$. The consistency of $\widehat{Q}_n^{(E)}(\beta_n|B_0, x_0)$ is established in the next theorem. The oscillations of the function $z \mapsto S_{B_0}(y|z)$ for large values of y are controlled by the following quantity: for $\xi \in (0, 1)$ and $\alpha \in (0, 1)$, let

$$\omega(\alpha, \xi|B_0, x_0) := \sup_{\beta/\alpha \in \mathcal{J}_\xi} \sum_{i=1}^n \mathbb{E}\left[\mathcal{W}_{n,i}(B_0, x_0)\left|\frac{S_{B_0}[Q_{B_0}(\beta|B_0^\top x_0)|B_0^\top X_i]}{\beta} - 1\right|\right],$$

with $\mathcal{J}_\xi := [(1-\xi)\kappa, 1+\xi]$. Let also

$$n_{B_0,x_0} := \left(\sum_{i=1}^n \mathcal{W}_{n,i}(B_0, x_0)\right)^{-1}.$$

Theorem 12 *Assume that condition* **(H2)** *holds with* $\rho(x_0) < 0$ *or* ($\rho(x_0) \leq 0$ *and* $\gamma(x_0) < 0$) *and that there exits a matrix* $B_0 \in \mathcal{B}$ *such that*

$$\lim_{\alpha \to 0} \frac{1}{A(\alpha^{-1}|x_0)} \left(\frac{S[Q_{B_0}(\alpha|B_0^\top x_0)|x_0]}{\alpha} - 1 \right) = 0, \tag{3.9}$$

Assume also that there exists a sequence $m_n(B_0, x_0)$ *such that* $m_n(B_0, x_0) \overset{a.s.}{\sim} n_{B_0,x_0}$ *and let*

$$\tau_n(B_0, x_0) := \{\ln[\alpha_n m_n(B_0, x_0)]/[\alpha_n m_n(B_0, x_0)]\}^{1/2}.$$

If $\alpha_n/\beta_n \to \infty$, $\tau_n(B_0, x_0)\ln(\alpha_n/\beta_n) \to 0$, $A(\alpha_n^{-1}|x_0) = o(\tau_n(B_0, x_0))$, $\Delta_a(\alpha|x_0) = o(\tau_n(B_0, x_0))$, *if there exists a positive constant* C_X *such that*

$$m_n(B_0, x_0) \max_{1 \leq i \leq n} \mathcal{W}_{n,i}(B_0, x_0) < C_X \ \ a.s. \tag{3.10}$$

and if there exists $\xi > 0$ *such that*

$$[m_n(B_0, x_0)\alpha_n]^{1/2} \omega(\alpha_n, \xi|B_0, x_0) = o(\tau_n(B_0, x_0)), \tag{3.11}$$

then,

$$\frac{\widehat{Q}_n^{(E)}(\beta_n|B_0, x_0) - Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0) q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n(B_0, x_0)).$$

Under condition **(TCI)**, we know that $\alpha^{-1} S[Q_{B_0}(\alpha|B_0^\top x_0)|x_0]$ converges to 1 as $\alpha \to 0$. The rate of convergence of this limit is provided by condition (3.9). Note also that if condition **(H2)** holds then condition (3.9) ensures that for all $t > 0$

$$\lim_{\alpha \to 0} \frac{\mathrm{ERV}(\alpha, t|B_0, x_0)}{A(\alpha^{-1}|x_0)} = \int_1^{t^{-1}} r^{\gamma(x_0)-1} L_{\rho(x_0)}(r) dr,$$

with

$$\mathrm{ERV}(\alpha, t|B_0, x_0) := \frac{Q_{B_0}(t\alpha|B_0^\top x_0) - Q_{B_0}(\alpha|B_0^\top x_0)}{a(\alpha^{-1}|x_0)} - L_{\gamma(x_0)}(t^{-1}),$$

see Lemma 9. In other words, the conditional distribution of Y given $B_0^\top X$ belongs to a maximum domain of attraction and satisfies a second-order condition.

Examples of weights – First, one can adapt the Nadaraya-Watson weights defined in (2.18). More specifically, a possible set of weights is given by

$$\mathcal{W}_{n,i}^{\mathrm{NW}}(B_0, x_0, h_n) := K\left(\frac{B_0^\top(X_i - x_0)}{h_n}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{B_0^\top(X_j - x_0)}{h_n}\right),$$

where $h_n > 0$ and $K(\cdot)$ is a density on $\mathbb{R}^q$. As in Chapter 2, we assume the following on the kernel function K:

(K) the kernel K is either an indicator function on a cell of $\mathbb{R}^q$ or such that $K(x) = L(\|x\|_q)$ where L is of bounded variation, continuous on $(0, \infty)$ and with support $[0, 1]$.

Note that when h_n converges to 0, these weights select the observations X_i such that $B_0^\top X_i \approx B_0^\top x_0$. For this set of weights, the obtained estimator of $Q(\beta_n|x_0)$ is denoted by $\widehat{Q}_n^{(NW,E)}(\beta_n|B_0, x_0)$. Its consistency is established in the following result.

Corollary 7 *Assume that conditions* **(K)** *and* **(H2)** *hold with $\rho(x_0) < 0$ or ($\rho(x_0) \leq 0$ and $\gamma(x_0) < 0$) and that there exits a matrix $B_0 \in \mathcal{B}$ such that (3.9) holds. Assume also that $B_0^\top X$ admits a density f_{B_0} continuous at x_0 and such that $f_{B_0}(x_0) > 0$ and let*

$$\tau_n^{\mathrm{NW}}(B_0, x_0) := \left(\frac{\ln(nh_n^q\alpha_n)}{nh_n^q\alpha_n}\right)^{1/2}.$$

If $\alpha_n/\beta_n \to \infty$, $nh_n^q[\alpha_n \wedge (\ln\ln n)^{-1}] \to \infty$, $\tau_n^{\mathrm{NW}}(B_0, x_0)\ln(\alpha_n/\beta_n) \to 0$, $A(\alpha_n^{-1}|x_0) = o(\tau_n^{\mathrm{NW}}(B_0, x_0))$, $\Delta_a(\alpha|x_0) = o(\tau_n^{\mathrm{NW}}(B_0, x_0))$, and if there exists $\xi > 0$ such that

$$\sup_{\substack{\beta/\alpha_n \in \mathcal{J}_\xi \\ \|B_0^\top(x-x_0)\|_q \leq h_n}} \left|\frac{S_{B_0}[Q_{B_0}(\beta|B_0^\top x_0)|B_0^\top x]}{\beta} - 1\right| = o\left(\frac{\tau_n^{\mathrm{NW}}(B_0, x_0)}{[nh_n^q\alpha_n]^{1/2}}\right),$$

with $\mathcal{J}_\xi = [(1-\xi)\kappa, 1+\xi]$ then,

$$\frac{\widehat{Q}_n^{(NW,E)}(\beta_n|B_0, x_0) - Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n^{\mathrm{NW}}(B_0, x_0)).$$

The proof of Corollary 7 follows the same lines as the proof of Corollary 3 in Chapter 2 and is thus omitted. Taking $B_0 = I_p$ is this result (in this case condition (3.9) is always satisfied but no reduction of dimension is done), the rate of convergence $\tau_n^{\mathrm{NW}}(I_p, x_0)$ of $\widehat{Q}_n^{(NW,E)}(\beta_n|I_p, x_0)$ is such that

$$\frac{\tau_n^{\mathrm{NW}}(B_0, x_0)}{\tau_n^{\mathrm{NW}}(I_p, x_0)} = \mathcal{O}(h_n^{p-q}) = o(1),$$

if $q < p$. As a consequence and as expected, if condition (3.9) holds with a $q \times q$ matrix B_0 ($q < p$) then the estimator $\widehat{Q}_n^{(NW,E)}(\beta_n|B_0, x_0)$ is theoretically better than the estimator $\widehat{Q}_n^{(NW,E)}(\beta_n|I_p, x_0)$ obtained without reducing the dimension of the covariate.

Another possibility is to adapt the Nearest-Neighbors weights given in (2.19). Let $\{D_i(B_0, x_0) := \|B_0^\top(X_i - x_0)\|_q,\ i = 1, \ldots, n\}$ be the set of distances between $B_0^\top X_i$ and $B_0^\top x_0$ and let $D_{(1)}(B_0, x_0) \leq \ldots \leq D_{(n)}(B_0, x_0)$ be the ordered distances. Denoting by $\{r(B_0, i),\ i = 1, \ldots, n\}$ the ranks of these distances, the nearest-neighbors weights are given by

$$\mathcal{W}_{n,i}^{\mathrm{NN}}(B_0, x_0, k_n) := [(k_n - r(B_0, i) + 1)_+]^\ell \Bigg/ \sum_{j=1}^{k_n} j^\ell ,$$

where $\ell \in \mathbb{N}$ and $k_n \in \{1, \ldots, n\}$. The obtained estimator is denoted $\widehat{Q}_n^{(NN,E)}(\beta_n|B_0, x_0)$.

Corollary 8 *Assume that condition* **(H2)** *holds with $\rho(x_0) < 0$ or ($\rho(x_0) \leq 0$ and $\gamma(x_0) < 0$) and that there exits a matrix $B_0 \in \mathcal{B}$ such that (3.9) holds. Assume also that $B_0^\top X$ admits a density f_{B_0} such that $f_{B_0}(x_0) > 0$ and let*

$$\tau_n^{\mathrm{NN}}(B_0, x_0) := \left(\frac{\ln(k_n \alpha_n)}{k_n \alpha_n}\right)^{1/2}.$$

If $\alpha_n/\beta_n \to \infty$, $k_n \alpha_n \to \infty$, $\tau_n^{\mathrm{NN}}(B_0, x_0) \ln(\alpha_n/\beta_n) \to 0$, $A(\alpha_n^{-1}|x_0) = o(\tau_n^{\mathrm{NN}}(B_0, x_0))$, $\Delta_a(\alpha|x_0) = o(\tau_n^{\mathrm{NN}}(B_0, x_0))$, and if there exists $\xi > 0$ such that

$$\sup_{\substack{\beta/\alpha_n \in \mathcal{J}_\xi \\ \|B_0^\top(x - x_0)\|_q \leq r_n}} \left| \frac{S_{B_0}[Q_{B_0}(\beta|B_0^\top x_0)|B_0^\top x]}{\beta} - 1 \right| = o\left(\frac{\tau_n^{\mathrm{NN}}(B_0, x_0)}{[k_n \alpha_n]^{1/2}}\right),$$

with $\mathcal{J}_\xi = [(1-\xi)\kappa, 1+\xi]$ *and* $r_n^q := 2k_n/(nf_{B_0}(x_0))$ *then,*

$$\frac{\widehat{Q}_n^{(NN,E)}(\beta_n|B_0,x_0) - Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n^{\mathrm{NN}}(B_0,x_0)).$$

3.4.2 Estimation of B_0

Of course in practice the matrix B_0 is unknown. In order to obtain a proper estimator of $Q(\beta_n|x_0)$, we propose to replace in the expression of $\widehat{Q}_n^{(E)}(\beta_n|B_0,x_0)$ the matrix B_0 by the estimator introduced in Gardes [18]. Let us recall the expression of this estimator. Let us first introduce some notations. For $J \in \mathbb{N}^*$ and for any matrix $B \in \mathcal{B}$, let $\{\Pi_j(B^\top X),\ j = 1,\dots,J\}$ be a random partition of $\mathrm{supp}(X)$. In the sequel, the following partition is considered: Let $D = [d_1,\dots,d_p]$ be a $p \times p$ orthogonal matrix such that $\mathrm{span}(B) = \mathrm{span}(\{d_1,\dots,d_q\})$. Let $m(B^\top x)$ be the conditional marginal median of X given $B^\top X = B^\top x$ and for $\ell \in \{1,\dots,p-q\}$, let us introduce the half spaces

$$\begin{aligned} E_\ell(B^\top x) &:= \{s \in \mathbb{R}^p;\ d_{\ell+q}^\top s > d_{\ell+q}^\top m(B^\top x)\} \\ \text{and} \quad \bar{E}_\ell(B^\top x) &:= \{s \in \mathbb{R}^p;\ d_{\ell+q}^\top s \le d_{\ell+q}^\top m(B^\top x)\}. \end{aligned}$$

An element of the partition $\{\Pi_j(B^\top X = B^\top x),\ j = 1,\dots,J\}$ is the intersection of $p-q$ half spaces. More specifically, an element of the partition is a set $E_1^* \cap \ldots \cap E_{p-q}^*$ where for $\ell \in \{1,\dots,p-q\}$, $E_\ell^* \in \{E_\ell(B^\top x), \bar{E}_\ell(B^\top x)\}$. There is thus $J = 2^{p-q}$ elements in the partition. Obviously, if $\mathrm{supp}(X) = \mathbb{R}^p$, then, for all $x \in \mathbb{R}^p$ and, for all $j \in \{1,\dots,J\}$, $\mathbb{P}(X \in \Pi_j(B^\top X)|B^\top X = B^\top x) > 0$.
In practice, since the conditional marginal median $m(B^\top x)$ is unknown, it is replaced by its empirical estimator $\widehat{m}_n(B^\top x) := (\widehat{m}_{n,j}(B^\top x),\ j = 1,\dots,J)^\top$ where for $j = 1,\dots,J$, $\widehat{m}_{n,j}(B^\top x)$ is the empirical median of the j-th component of the observations falling in $D(B^\top x, H_n)$. This choice of random partition ensures that, for all $x \in \mathrm{supp}(X)$, the number of available observations in each element of $D(B^\top x, H_n) \cap \Pi_j(B^\top X = B^\top x)$ is approximatively the same.

Let us now introduce the function $T : (0,1) \times \mathcal{B} \mapsto [0,\infty]$ defined by

$$T(\alpha, B) := \sum_{j=1}^{J} \left\{ \mathbb{E}\left[\frac{\mathbb{P}(\{Y > Q(\alpha|B^\top X)\} \cap \{X \in \Pi_j(B^\top X)\}|B^\top X)}{\alpha \mathbb{P}(X \in \Pi_j(B^\top X)|B^\top X)} \right] - 1 \right\}^2.$$

According to the second statement of Gardes [18] Theorem 1, the quantity $T(\alpha, B_0)$ is close to 0 for small values of α. This argument suggests that an approximation of B_0 can be obtained by minimizing the function $T(\alpha, B)$ with α small. This naturally motivates us to estimate B_0 by minimizing over $\mathcal{B}$ a reasonable estimator of $T(\alpha, B)$ with α sufficiently small. To construct this estimator, let us introduce a sequence (α_n) converging to 0 with the sample size. The sample analog of $T(\alpha_n, B)$ is given by:

$$\frac{1}{n^2}\sum_{j=1}^{J}\left\{\sum_{i=1}^{n}\left(\frac{\Phi_{n,j}(B^\top X_i)}{\alpha_n p_j(B^\top X_i)}-1\right)\right\}^2, \tag{3.12}$$

with for $B \in \mathcal{B}$, $z \in \text{supp}(B^\top X)$ and $j \in \{1, \dots, J\}$, $p_j(z) := \mathbb{P}(X \in \Pi_j(B^\top X)|B^\top X = z)f_B(z)$ (where $f_B(\cdot)$ is the probability density function of $B^\top X$) and

$$\frac{\Phi_{n,j}(z)}{f_B(z)} := \mathbb{P}\left(\{Y > Q(\alpha|B^\top X = z)\} \cap \{X \in \Pi_j(B^\top X)\}|B^\top X = z\right).$$

Obviously, in practice, random variables $\Phi_{n,j}(B^\top X_i)$ and $p_j(B^\top X_i)$ are not observed and must be replaced by their respective estimators:

$$\widehat{\Phi}_{n,j}(B^\top X_i) := \sum_{\ell\neq i}\mathcal{W}_{n,\ell}(B, X_i)\mathbb{I}_{\{Y_\ell > \widehat{Q}_{n,-i}(\alpha_n|B,X_i)\}}\mathbb{I}_{\{X_\ell\in\Pi_j(B^\top X_i)\}}, \tag{3.13}$$

where $\widehat{Q}_{n,-i}(\alpha_n|B, X_i)$ is the conditional quantile estimator defined in (3.8) computed without the pair (X_i, Y_i) and

$$\widehat{p}_j(B^\top X_i) := \sum_{\ell\neq i}\mathcal{W}_{n,\ell}(B, X_i)\mathbb{I}_{\{X_\ell\in\Pi_j(B^\top X_i)\}}. \tag{3.14}$$

We can now introduce our estimator of B_0:

$$\widehat{B}_{0,n} := \arg\min_{B\in\mathcal{B}} \widehat{T}_n(B), \tag{3.15}$$

where $\widehat{T}_n(B)$ is obtained by replacing in (3.12) the unobserved random variables $\Phi_{n,j}(B^\top X_i)$ and $p_j(B^\top X_i)$ by $\widehat{\Phi}_{n,j}(B^\top X_i)$ and $\widehat{p}_j(B^\top X_i)$.
In practice, since the construction of the partition $\{\Pi_j(B^\top X_i),\ j = 1, \dots, J\}$ is time consuming, the sample analog of $T(\alpha_n, B)$ given in (3.12) is computed on a subsample of size $n_0 = \lfloor n/40 \rfloor$. The minimization problem (3.15) is solved by a coordinate search method (see Gardes [18] for more details).

3.5 Simulation study

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n independent copies of a random vector (X, Y) where X is a $\mathbb{R}^p$-valued random vector and Y a $\mathbb{R}$-valued random variable. We are interested in the finite sample performance of the estimators of $Q(\beta_n|x_0)$ introduced in this chapter where $\beta_n \to 0$ and $x_0 \in \mathbb{R}^p$.

Model settings – In what follows, the p components of X are independent and distributed as a Gaussian distribution with mean 1/2 and standard deviation $\sigma = 1/3$. Let B_0 and B_1 be two $p \times 1$ matrices, the conditional quantile function of Y given $\{X = x_0\}$ is given by

(M1) Conditional Burr distribution: for $\alpha \in (0, 1)$,

$$Q_1(\alpha|x_0) := \left(\alpha^{-g_1(B_1^\top x_0)} - 1\right)^{g_0(B_0^\top x_0)/g_1(B_1^\top x_0)},$$

(M2) Conditional reverse Burr distribution: for $\alpha \in (0, 1)$,

$$Q_2(\alpha|x_0) := g_2(B_0^\top x_0) - \frac{1}{Q_1(\alpha|x_0)},$$

(M3) Conditional Weibull-tail distribution: for $\alpha \in (0, 1)$,

$$Q_3(\alpha|x_0) := [\ln(1/\alpha)]^{g_0(B_0^\top x_0)} (1-\alpha)^{g_0(B_0^\top x_0)/g_1(B_1^\top x_0)}.$$

The positive functions g_0 and g_1 are defined for all $z \in \mathbb{R}$ by

$$g_0(z) = \frac{1}{3}\mathbb{I}_{\{z \le 1/4\}} + \frac{1}{9}(1 - 8z)\mathbb{I}_{\{1/4 < z \le 1\}} + \mathbb{I}_{\{z > 1\}},$$

and

$$\begin{aligned} g_1(z) &= 2\mathbb{I}_{\{z \le 1/4\}} + (4z - 1)\mathbb{I}_{\{1/4 < z \le 3/8\}} + 4(1 - z)\mathbb{I}_{\{3/8 < z \le 1/2\}} \\ &+ (6 - 8z)\mathbb{I}_{\{1/2 < z \le 5/8\}} + (8z - 4)\mathbb{I}_{\{5/8 < z \le 3/4\}} + 2\mathbb{I}_{\{z > 3/4\}}. \end{aligned}$$

The real-valued function g_2 is given by $g_2(z) = 2z$.

These three conditional distributions belong to a maximum domain of attraction (MDA). Model **(M1)** is in the Fréchet MDA with tail index

$\gamma(x_0) = g_0(B_0^\top x_0) > 0$. The second order parameter is $\rho(x_0) = -g_1(B_1^\top x_0)$. Model **(M2)** is in the Weibull MDA with extreme value index $\gamma(x_0) = -g_0(B_0^\top x_0) < 0$ and finite right endpoint $g_2(B_0^\top x_0) > 0$. Finally, model **(M3)** is in the Weibull MDA with extreme value index $\gamma(x_0) = 0$. One can show that all these models satisfy condition **(TCI)** with the matrix B_0 (we thus have $q = 1$).

Definition of the estimators – Let us now defined the estimators considered in this simulation study.
Nadaraya-Watson's estimators: For any full rank $p \times r$ matrix with $0 < r \leq p$ and for $i \in \{1, \ldots, n\}$, let

$$\mathcal{W}_{n,i}^{\mathrm{NW}}(B, x_0, h_n) := K\left(\frac{B^\top (X_i - x_0)}{h_n}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{B^\top (X_j - x_0)}{h_n}\right),$$

where $h_n > 0$ and for $u = (u_1, \ldots, u_r)^\top \in \mathbb{R}^r$,

$$K(u) = \left(\frac{3}{4}\right)^r \prod_{j=1}^{r} (1 - u_j^2) \mathbb{I}_{\{\|u\|_\infty \leq 1\}},$$

where $\|u\|_\infty = \max\{|u_1|, \ldots, |u_r|\}$. Taking these set of weights in the expression of the estimators $\widehat{Q}_n(\alpha_n|B, x_0)$, $\widehat{a}_n(\alpha_n^{-1}|B, x_0)$ and $\widehat{\gamma}_n(B, x_0)$ leads to the estimator

$$\widehat{Q}_n^{(NW,E)}(\beta_n|B, x_0) = \widehat{Q}_n(\alpha_n|B, x_0) + \widehat{a}_n(\alpha_n^{-1}|B, x_0) L_{\widehat{\gamma}_n(B, x_0)}\left(\frac{\alpha_n}{\beta_n}\right). \quad (3.16)$$

Note that this estimator depends on two hyper-parameters: the bandwidth h_n and the intermediate sequence α_n. Let us also mention that $\widehat{a}_n(\alpha_n^{-1}|B, x_0)$ and $\widehat{\gamma}_n(B, x_0)$ have been computed with $\kappa = 0.02$ and $\varphi(\cdot) = \ln(1/\cdot)$.
Nearest Neighbor's estimators: For any full rank $p \times r$ matrix with $0 < r \leq p$, let $\{D_i(B, x_0) := \|B^\top (X_i - x_0)\|_r,\ i = 1, \ldots, n\}$ be the set of distances between $B^\top X_i$ and $B^\top x_0$ and let $D_{(1)}(B, x_0) \leq \ldots \leq D_{(n)}(B, x_0)$ be the ordered distances. Denoting by $\{r(B, i),\ i = 1, \ldots, n\}$ the ranks of these distances, let

$$\mathcal{W}_{n,i}^{\mathrm{NN}}(B, x_0, k_n) := \frac{2(k_n - r(B, i) + 1)_+}{k_n(k_n + 1)}.$$

Taking these set of weights yields to the following estimator of $Q(\beta_n|x_0)$:

$$\widehat{Q}_n^{(NN,E)}(\beta_n|B,x_0) = \widehat{Q}_n(\alpha_n|B,x_0) + \widehat{a}_n(\alpha_n^{-1}|B,x_0) L_{\widehat{\gamma}_n(B,x_0)}\left(\frac{\alpha_n}{\beta_n}\right). \quad (3.17)$$

This estimator depends on the choices of the number of nearest neighbors k_n and the intermediate sequence α_n.
For the matrix B in (3.16) or (3.17), a first possibility is to take $B = B_0$ since condition **(TCI)** is satisfied with the matrix B_0. Of course this choice is theoretically the best one but in practice the matrix B_0 is unknown. It can be estimated by $\widehat{B}_{0,n}$ defined in (3.15) by taking in (3.13) and (3.14) the weights $\mathcal{W}_{n,i}^{\mathrm{NW}}(B,x_0,h_n)$ or $\mathcal{W}_{n,i}^{\mathrm{NN}}(B,x_0,k_n)$. The statistic $\widehat{B}_{0,n}$ depends on h_n (or k_n) and α_n. In what follows we take $h_n = 0.2$ (or $k_n = 1500$) and $\alpha_n = 0.1$, these "optimal" values coming from an extensive simulation study. Another way to estimate $Q(\beta_n|x_0)$ is thus to take $B = \widehat{B}_{0,n}$ in (3.16) or (3.17). Finally, if one does not assume the validity of condition **(TCI)**, natural estimators are obtained by taking $B = I_p$ in (3.16) or (3.17). As mentionned before these last estimators are expected to perform very badly when the dimension p increases.

Choice of the hyper-parameters – Let λ_n be the sequence equal to h_n or k_n depending on the considered estimator (Nadaraya-Watson or Nearest-Neighbor). The parameters λ_n and α_n required to compute the previous estimators are selected according to the following data driven procedure. For $B \in \{B_0, \widehat{B}_{0,n}, I_p\}$, let

$$\lambda_n^{\mathrm{opt},\bullet}(B,\alpha) := \arg\min_{\lambda\in\mathcal{G}} \left\{ \left(\sum_{i=1}^n \mathcal{W}_{n,i}^{\bullet}(B,x_0,\lambda) \mathbb{I}_{\{Y_i > \widehat{Q}_{n,-i}(\alpha|B,x_0)\}} - \alpha \right)^2 \right\},$$

where $\bullet$ has to be replaced by NW or NN and $\mathcal{G}$ is a grid of 20 evenly spaced points between 0.05 and 0.5 if $\bullet = NW$ or a grid of 20 evenly spaced points between 300 and 3000 if $\bullet = NN$. The selected value of α_n is given by

$$\alpha_n^{\mathrm{opt},\bullet}(B) = \arg\min_{\alpha\in\mathcal{A}} \int_{0.03}^{0.1} \left(\frac{\widehat{Q}_n^{(\bullet,E)}(\beta|B,x_0,\lambda_n^{\mathrm{opt},\bullet}(B,\alpha))}{\widehat{Q}_n(\beta|B,x_0)} - 1 \right)^2 d\beta$$

where $\widehat{Q}_n(\beta|B,x_0)$ is the estimator defined in (3.8) and computed with the weights $\mathcal{W}_{n,i}^{\bullet}(B,x_0,\lambda_n^{\mathrm{opt},\bullet}(B,\alpha))$. Note that the interval on which we integrate has been found after an extensive simulation study. The set $\mathcal{A}$ is a grid of 10 evenly spaced points between 0.05 and 0.25. The selected value of λ_n is then given for $B \in \{B_0, \widehat{B}_{0,n}, I_p\}$ by

$$\lambda_n^{\text{opt},\bullet}(B) := \lambda_n^{\text{opt},\bullet}(B, \alpha_n^{\text{opt},\bullet}(B)).$$

except when $B = I_p$ and $\bullet = NW$. In order to keep the same rate of convergence for the Nadaraya-Watson's estimators, we take when $B = I_p$,

$$\lambda_n^{\text{opt,NW}}(I_p) := \left[\lambda_n^{\text{opt,NW}}(B, \alpha_n^{\text{opt,NW}}(B))\right]^{1/p}$$

Simulations results – For each model, we generate $N = 100$ samples of size $n = 4000$. Three values of the covariate's dimension are considered: $p = 2$ with $B_0 = (1,2)^\top/\sqrt{5}$ and $B_1 = (0,1)^\top$; $p = 3$ with $B_0 = (1,2,0)^\top/\sqrt{5}$ and $B_1 = (0,1,1)^\top$ and $p = 4$ with $B_0 = (1,2,0,0)^\top/\sqrt{5}$ and $B_1 = (0,1,1,1)^\top$. In order to compare the previously defined estimators, we compute the error given for $B \in \{B_0, \widehat{B}_{0,n}, I_p\}$ by

$$E_Q(\bullet, B) := \sum_{x_0 \in \mathcal{X}} \sum_{j=1}^{N} \left(\frac{\widehat{Q}_{n,j}^{(\bullet,E)}(\beta_n | B, x_0)}{Q(\beta_n, x_0)} - 1 \right)^2,$$

where $\widehat{Q}_{n,j}^{(\bullet,E)}(\beta_n|B, x_0)$ is the estimator $\widehat{Q}_n^{(\bullet,E)}(\beta_n|B, x_0)$ computed with the j-th replicated sample and $\mathcal{X} := \{cu^\top,\ c \in \{1/4, 2/5, 1/2, 3/5, 3/4\}\}$ with $u = (1, \cdots, 1)^\top \in \mathbb{R}^p$. The results are summarized in Tables 3.1 and 3.2 for two different values of the quantile level: $\beta_n = 1/n$ and $\beta_n = 1/(2n)$. From the simulations, we can infer that, as expected, the smallest error is obtained in case $B = B_0$, but estimating the matrix B_0 by $\widehat{B}_{0,n}$ does not too much deteriorate the performance of the extreme conditional quantile. On the contrary, the estimator based on $B = I_p$ performs very poorly, especially for model **(M1)** corresponding to heavy-tailed distributions. In addition, the error $E_Q(\bullet, B)$ does not seem to be as much affected by the dimension p of the covariate in the case $B \in \{B_0, \widehat{B}_{0,n}\}$, although when I_p is used, it clearly explodes with the dimension especially for heavy-tailed distributions, the deterioration being less sensitive for the two other models. Finally, as one might expect, the smallest the level of the quantile is, the large is the error and thus worse is the estimation. On the figure 3.2, we can observe that if $h_n \in [0.1, 0.4]$, the errors $E_Q(NW, B_0)$ and $E_Q(NW, \hat{B}_{0,n})$ are minimal while $E_Q(NW, I_p)$ displays high values whatever the value of h_n. It shows that the selection of the parameter h_n is relevant when we use the dimension reduction method and not when we consider $B = I_p$.

3.6 Proofs

3.6.1 Proof of Theorem 11

We start with the decomposition

$$\frac{\check{Q}_n^{(E)}(\beta_n|x_0) - Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = T_{1,n} + T_{2,n} + T_{3,n} + T_{4,n},$$

where

$$\begin{aligned}
T_{1,n} &:= \frac{\check{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)}, \\
T_{2,n} &:= \frac{\check{a}_n(\alpha_n^{-1}|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)}\left[L_{\check{\gamma}_n(x_0)}(\alpha_n/\beta_n) - L_{\gamma(x_0)}(\alpha_n/\beta_n)\right], \\
T_{3,n} &:= \frac{1}{q_{\gamma(x_0)}(\alpha_n/\beta_n)}\left(\frac{\check{a}_n(\alpha_n^{-1}|x_0)}{a(\alpha_n^{-1}|x_0)} - 1\right)L_{\gamma(x_0)}(\alpha_n/\beta_n) \\
\text{and } T_{4,n} &:= \frac{1}{q_{\gamma(x_0)}(\alpha_n/\beta_n)}L_{\gamma(x_0)}(\alpha_n/\beta_n) + \frac{Q(\alpha_n|x_0) - Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)}.
\end{aligned}$$

Let us first focus on the term $T_{1,n}$. Since

$$q_{\gamma(x_0)}(t) = \begin{cases} t^{\gamma(x_0)}\ln(t)/\gamma(x_0) & \text{if } \gamma(x_0) > 0, \\ \ln^2(t)/2 & \text{if } \gamma(x_0) = 0, \\ 1/\gamma^2(x_0) & \text{if } \gamma(x_0) < 0, \end{cases} \tag{3.18}$$

one has that $q_{\gamma(x_0)}(\alpha_n/\beta_n) \to \infty$. The assumption on $\check{Q}_n(\alpha_n|x_0)$ leads to

$$T_{1,n} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)). \tag{3.19}$$

Now, using the assumption on the estimator $\check{a}_n(\alpha_n^{-1}|x_0)$,

$$T_{2,n} = [\check{\gamma}_n(x_0) - \gamma(x_0)][1 + \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))]\left\{\frac{I_n}{q_{\gamma(x_0)}(\alpha_n/\beta_n)} + 1\right\},$$

where

$$I_n := \int_1^{\alpha_n/\beta_n} s^{\gamma(x_0)-1}\left(\frac{\exp[(\check{\gamma}_n(x_0)-\gamma(x_0))\ln(s)]-1}{(\check{\gamma}_n(x_0)-\gamma(x_0))\ln(s)}-1\right)\ln(s)ds.$$

Using the assumption on the estimator $\check{\gamma}_n(x_0)$ and since $\tau_n(x_0)\to 0$, it appears that $[\check{\gamma}_n(x_0)-\gamma(x_0)][1+\mathcal{O}_{\mathbb{P}}(\tau_n(x_0))]=\mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$. Furthermore, using the inequality

$$\left|\frac{\exp(x)-1}{x}-1\right|\leq \exp(|x|)-1,$$

yields to

$$\begin{aligned}
|I_n| &\leq \int_1^{\alpha_n/\beta_n} s^{\gamma(x_0)-1}\left(\exp[|\check{\gamma}_n(x_0)-\gamma(x_0)|\ln(s)]-1\right)\ln(s)ds\\
&\leq q_{\gamma(x_0)}(\alpha_n/\beta_n)\left(\exp[|\check{\gamma}_n(x_0)-\gamma(x_0)|\ln(\alpha_n/\beta_n)]-1\right).
\end{aligned}$$

Finally, using the Taylor expansion $\exp(u)-1=o(u)$ as $u\to 0$ and since by assumption $|\check{\gamma}_n(x_0)-\gamma(x_0)|\ln(\alpha_n/\beta_n)\xrightarrow{\mathbb{P}}0$, one has

$$T_{2,n}=\mathcal{O}_{\mathbb{P}}(\tau_n(x_0)). \tag{3.20}$$

Let us now deal with the term $T_{3,n}$. As $t\to\infty$,

$$L_{\gamma(x_0)}(t)\sim\begin{cases} t^{\gamma(x_0)}/\gamma(x_0) & \text{if } \gamma(x_0)>0,\\ \ln(t) & \text{if } \gamma(x_0)=0,\\ -1/\gamma(x_0) & \text{if } \gamma(x_0)<0.\end{cases}$$

As a consequence $L_{\gamma(x_0)}(\alpha_n/\beta_n)/q_{\gamma(x_0)}(\alpha_n/\beta_n)=\mathcal{O}(1)$ and thus, under the condition on the estimator $\check{a}_n(\alpha_n^{-1}|x_0)$,

$$T_{3,n}=\mathcal{O}_{\mathbb{P}}(\tau_n(x_0)). \tag{3.21}$$

Finally, we focus on the non-random term $T_{4,n}$. Remark first that

$$\begin{aligned}
T_{4,n} &= -A(\alpha_n^{-1}|x_0)\frac{L_{\gamma(x_0)}(\alpha_n/\beta_n)}{q_{\gamma(x_0)}(\alpha_n/\beta_n)}\\
&\times \frac{1}{A(\alpha_n^{-1}|x_0)}\left(\frac{Q(\beta_n|x_0)-Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)}L^{-1}_{\gamma(x_0)}(\alpha_n/\beta_n)-1\right).
\end{aligned}$$

The second order condition entails (see de Haan and Ferreira [25] Lemma 4.3.5) that

$$\lim_{n\to\infty} \frac{1}{A(\alpha_n^{-1}|x_0)} \left(\frac{Q(\beta_n|x_0) - Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)} L^{-1}_{\gamma(x_0)}(\alpha_n/\beta_n) - 1 \right) = -\frac{1}{(\gamma(x_0))_- + \rho(x_0)},$$

where $(\cdot)_-$ is the negative part function. Since by assumption $A(\alpha_n^{-1}|x_0) = \mathcal{O}(\tau_n(x_0))$ and since $L_{\gamma(x_0)}(\alpha_n/\beta_n)/q_{\gamma(x_0)}(\alpha_n/\beta_n) = \mathcal{O}(1)$, we obtain that

$$T_{4,n} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)). \tag{3.22}$$

Collecting (3.19) to (3.22) conclude the proof. ■

3.6.2 Proof of Proposition 13

Some preliminaries results are required for the proof of Proposition 13.

Lemma 7 *Let $\{X_{n,i}(\cdot),\ i = 1,\ldots,n\}$ be independent, non-decreasing and positive processes defined on the interval $[a,b]$ with $-\infty < a < b < +\infty$. For all $v \in [a,b]$, let*

$$\nu_n(v) = \sum_{i=1}^n \mathbb{E}(X_{n,i}(v)),$$

and assume that for all $n \in \mathbb{N}$, $\nu_n(a) > 0$ and $\nu_n(b) < \infty$. Let us also introduce the sequences $\tau_n := [\ln(\nu_n(b))/\nu_n(b)]^{1/2}$ and

$$\bar{\nu}_n := \sup\left\{ \left| \frac{\nu_n(v)}{\nu_n(v')} - 1 \right|,\ v \in [a,b] \text{ with } \frac{|v - v'|}{b-a} \le [\nu_n(b)]^{-1/2} \right\}.$$

If there exist a positive constant C_X such that

$$\max_{i=1,\ldots,n} X_{n,i}(b) \le C_X \text{ a.s.}$$

and if there exists $N_1 \in \mathbb{N}$ and positive constants C_ν and $\bar{C}_\nu$ such that for all $n \ge N_1$, $\tau_n^{-1}\bar{\nu}_n \le \bar{C}_\nu$ and $\nu_n(a)/\nu_n(b) \ge C_\nu$, then for all $n \ge N_1$, there exists a positive constant $C(C_\nu, C_X, \bar{C}_\nu)$ such that

$$\mathbb{P}\left[\tau_n^{-1} \sup_{v\in[a,b]} \left| \sum_{i=1}^n \frac{X_{n,i}(v)}{\nu_n(v)} - 1 \right| > C(C_\nu, C_X, \bar{C}_\nu) \right] \le 2\frac{\lceil \nu_n(b) \rceil^{1/2}}{\nu_n(b)}. \tag{3.23}$$

Note that this result is a slightly more precise version of Gardes [17, Lemma 6]. Note also that if we assume in addition that $\nu_n(a) \to \infty$ as $n \to \infty$, then a direct consequence of (3.23) is that

$$\sup_{v\in[a,b]} \left| \sum_{i=1}^{n} \frac{X_{n,i}(v)}{\nu_n(v)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(\tau_n).$$

Proof – Let us introduce the sequence $r_n := \lceil \nu_n(b) \rceil^{1/2} + 1$ where $\lceil\cdot\rceil$ is the ceiling function. For $j = 1, \ldots, r_n$, let

$$\theta_n(j) := a + (j-1)\frac{b-a}{r_n - 1} \in [a,b].$$

Clearly, for all $v \in [a,b]$, there exists $j_v \in \{1, \ldots, r_n - 1\}$ such that $\theta_n(j_v) \le v \le \theta_n(j_v+1)$. Let

$$\widehat{\varphi}_n(\cdot) := \sum_{i=1}^{n} X_{n,i}(\cdot).$$

Since $\widehat{\varphi}_n$ is a non-decreasing function,

$$\begin{aligned} \left| \frac{\widehat{\varphi}_n(v)}{\nu_n(v)} - 1 \right| \;\; &\le \;\; \frac{1}{\nu_n(a)} \Big\{ \left|\widehat{\varphi}_n[\theta_n(j_v+1)] - \nu_n[\theta_n(j_v+1)]\right| \\ &+ \;\; 2\left|\widehat{\varphi}_n[\theta_n(j_v)] - \nu_n[\theta_n(j_v)]\right| + 2\left(\nu_n[\theta_n(j_v+1)] - \nu_n[\theta_n(j_v)]\right) \Big\} \end{aligned}$$

As a consequence,

$$\sup_{v\in[a,b]} \left| \frac{\widehat{\varphi}_n(v)}{\nu_n(v)} - 1 \right| \le T_{n,1} + T_{n,2}, \tag{3.24}$$

where

$$T_{n,1} := \frac{3}{\nu_n(a)} \max_{1\le j\le r_n-1} \left(\nu_n[\theta_n(j+1)] - \nu_n[\theta_n(j)] \right),$$

and

$$T_{n,2} := \frac{3}{\nu_n(a)} \max_{1\le j\le r_n-1} \left| \widehat{\varphi}_n[\theta_n(j)] - \nu_n[\theta_n(j)] \right|.$$

Let us first focus on $T_{n,1}$. Since for all $j \in \{1, \ldots, r_n - 1\}$, $\theta_n(j+1) - \theta_n(j) = (b-a)/(r_n-1) \le (b-a)[\nu_n(b)]^{-1/2}$, one has

$$T_{n,1} \le \frac{3\nu_n(b)}{\nu_n(a)} \left| \frac{\nu_n[\theta_n(j+1)]}{\nu_n[\theta_n(j)]} - 1 \right| \le \frac{3\bar{\nu}_n}{C_\nu}.$$

Since for $n \geq N_1$, $\tau_n^{-1}\bar{\nu}_n \leq \bar{C}_\nu$, we obtain that for all $n \geq N_1$,

$$\tau_n^{-1} T_{n,1} \leq \frac{3\bar{C}_\nu}{C_\nu}. \tag{3.25}$$

Let us now treat the term $T_{n,2}$. For all $C > 0$,

$$\begin{aligned}
\mathbb{P}[\tau_n^{-1} T_{n,2} > C] &\leq \mathbb{P}\left[\tau_n^{-1} \frac{3}{C_\nu} \max_{1 \leq j \leq r_n - 1} \left| \frac{\widehat{\varphi}_n[\theta_n(j)]}{\nu_n[\theta_n(j)]} - 1 \right| > C \right] \\
&\leq \sum_{j=1}^{r_n - 1} \mathbb{P}\left[\tau_n^{-1} \left| \frac{\widehat{\varphi}_n[\theta_n(j)]}{\nu_n[\theta_n(j)]} - 1 \right| > \frac{C C_\nu}{3} \right].
\end{aligned}$$

Using a multiplicative form of the Chernoff's inequality for bounded variables (see for instance Dubhashi and Panconesi [9] Theorem 1.1), one has

$$\begin{aligned}
\mathbb{P}[\tau_n^{-1} T_{n,2} > C] &\leq 2 \sum_{j=1}^{r_n - 1} \exp\left(-\frac{C^2 C_\nu^2}{27 C_X} \frac{\nu_n[\theta_n(j)]}{\nu_n(b)} \ln \nu_n(b) \right) \\
&\leq 2 \lceil \nu_n(b) \rceil^{1/2} \exp\left(-\frac{C^2 C_\nu^3}{27 C_X} \ln \nu_n(b) \right).
\end{aligned}$$

Replacing C by $3C_\nu^{-3/2}(3C_X)^{1/2}$ leads to

$$\mathbb{P}[\tau_n^{-1} T_{n,2} > C] \leq 2 \frac{\lceil \nu_n(b) \rceil^{1/2}}{\nu_n(b)}. \tag{3.26}$$

Now, using (3.24), one has for all $C > 0$ that,

$$\mathbb{P}\left[\tau_n^{-1} \sup_{u \in [a,b]} \left| \frac{\widehat{\varphi}_n(u)}{\nu_n(u)} - 1 \right| \geq C \right] \leq \mathbb{I}_{\{\tau_n^{-1} T_{n,1} \geq C/2\}} + \mathbb{P}\left[\tau_n^{-1} T_{n,2} \geq \frac{C}{2}\right].$$

Replacing C by $C(C_\nu, C_X, \bar{C}_\nu)$ and using (3.25) and (3.26) conclude the proof. ∎

Lemma 7 is used to establish the following concentration inequality.

Lemma 8 *Let $V_n(\cdot)$ be a non-decreasing and positive stochastic process defined on $[a,b]$ with $-\infty < a < b < +\infty$. Denote by $\{V_{n,1}(\cdot), \ldots, V_{n,n}(\cdot)\}$ a sequence of n independent copies of $V_n(\cdot)$ and by $\{T_{n,1}, \ldots, T_{n,n}\}$ a triangular array of positive random variables independent of the $V_{n,i}(\cdot)$ such* *that* $T_{n,1} + \ldots + T_{n,n} = 1$. *For* $v \in [a,b]$, *let* $\nu_n(v) = \mathbb{E}(V_n(v))$, $\tau_n = [\ln(\nu_n(b))/\nu_n(b)]^{1/2}$ *and*

$$\bar{\nu}_n := \sup\left\{\left|\frac{\nu_n(v)}{\nu_n(v')} - 1\right|,\ v \in [a,b] \text{ with } \frac{|v-v'|}{b-a} \leq [\nu_n(b)]^{-1/2}\right\}.$$

If there exists a positive sequence $\bar{b}_n$ *such that* $V_n(b) < \bar{b}_n$ *a.s., a positive constant* C_X *such that*

$$\max_{i=1,\ldots,n} \bar{b}_n T_{n,i} \leq C_X \text{ a.s.}$$

and if there exist $N_1 \in \mathbb{N}$ *and positive constants* C_ν *and* $\bar{C}_\nu$ *such that for all* $n \geq N_1$, $\tau_n^{-1}\bar{\nu}_n \leq \bar{C}_\nu$ *and* $\nu_n(a)/\nu_n(b) \geq C_\nu$, *then there exists a positive constant* $C(C_\nu, C_X, \bar{C}_\nu)$ *such that*

$$\mathbb{P}\left[\tau_n^{-1} \sup_{v\in[a,b]} \left|\sum_{i=1}^n \frac{T_{n,i}V_{n,i}(v)}{\nu_n(v)} - 1\right| > C(C_\nu, C_X, \bar{C}_\nu)\right] \leq 2\frac{\lceil \nu_n(b)\rceil^{1/2}}{\nu_n(b)}. \quad (3.27)$$

Proof – Let $\{t_{n,i},\ i = 1,\ldots,n\}$ be a triangular array of positive numbers, summing to 1 and such that

$$\bar{b}_n \max_{1\leq i\leq n} t_{n,i} \leq C_X.$$

Let us also introduce the stochastic process

$$\tilde{\varphi}_n(v; t_{n,1},\ldots,t_{n,n}) := \sum_{i=1}^n t_{n,i}V_{n,i}(v),\ v \in [a,b].$$

Since $\mathbb{E}[\tilde{\varphi}_n(v; t_{n,1},\ldots,t_{n,n})] = \nu_n(v)$ and

$$\max_{1\leq i\leq n} t_{n,i}V_{n,i}(b) \leq C_X \text{ a.s.}$$

one can use Lemma 7. Thus, there exists $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$,

$$\mathbb{P}\left[\tau_n^{-1} \sup_{v\in[a,b]} \left|\frac{\tilde{\varphi}_n(v; t_{n,1},\ldots,t_{n,n})}{\nu_n(v)} - 1\right| > C\right] \leq 2\frac{\lceil \nu_n(b)\rceil^{1/2}}{\nu_n(b)},$$

where $C := C(C_\nu, C_X, \bar{C}_\nu)$. As a consequence, since the $V_{n,i}$'s and the $T_{n,i}$'s are independent, one has for all ω satisfying the condition

$$\bar{b}_n \max_{1\leq i\leq n} T_{i,n}(\omega) \leq C_X,$$

that for all $n \geq N_1$

$$\begin{aligned}
&\mathbb{P}\left[\tau_n^{-1} \sup_{v\in[a,b]}\left|\frac{\tilde{\varphi}_n(v;T_{n,1}(\omega),\ldots,T_{n,n}(\omega))}{\nu_n(v)}-1\right|>C\middle|T_{n,i}=T_{n,i}(\omega),\ i=1,\ldots,n\right]\\
=\ &\mathbb{P}\left[\tau_n^{-1} \sup_{v\in[a,b]}\left|\frac{\tilde{\varphi}_n(v;T_{n,1}(\omega),\ldots,T_{n,n}(\omega))}{\nu_n(v)}-1\right|>C\right]\leq 2\frac{\lceil\nu_n(b)\rceil^{1/2}}{\nu_n(b)}.
\end{aligned}$$

Hence, for all $n \geq N_1$,

$$\begin{aligned}
&\mathbb{P}\left[\tau_n^{-1} \sup_{v\in[a,b]}\left|\frac{\widehat{\varphi}_n(v)}{\nu_n(v)}-1\right|>C\right]\\
=\ &\mathbb{E}\left\{\mathbb{P}\left[\tau_n^{-1} \sup_{v\in[a,b]}\left|\frac{\tilde{\varphi}_n(v;T_{n,1},\ldots,T_{n,n})}{\nu_n(v)}-1\right|>C\middle|T_{n,i},\ i=1,\ldots,n\right]\right\}\\
\leq\ &2\frac{\lceil\nu_n(b)\rceil^{1/2}}{\nu_n(b)},
\end{aligned}$$

and the proof is complete. ∎

The last preliminary result required to prove Proposition 13 is a uniform consistency property on the conditional quantile estimator.

Proposition 14 *Assume that there exists a sequence $m_n(x_0)$ such that $m_n(x_0) \overset{a.s.}{\sim} n_{x_0}$ and let $\tau_n(x_0) := \{\ln[\alpha_n m_n(x_0)]/[\alpha_n m_n(x_0)]\}^{1/2} \to 0$. Under condition* **(H2)**, *suppose that $A(\alpha_n^{-1}|x_0) = o_{\mathbb{P}}(\tau_n(x_0))$, that there exists a positive constant C_X such that*

$$m_n(x_0) \max_{1\leq i\leq n} \mathcal{W}_{n,i}(x_0) < C_X \ a.s.$$

and that there exists $\xi > 0$ such that

$$[m_n(x_0)\alpha_n]^{1/2}\omega(\alpha_n,\xi|x_0) = o(\tau_n(x_0)).$$

For all sequence w_n converging to 0, letting $y_n(t|x_0) := Q(t\alpha_n|x_0) + w_n a(\alpha_n^{-1}|x_0)$, one has that

$$\sup_{t\in[\kappa,1]}\left|\frac{\widehat{S}_n(y_n(t|x_0)|x_0)}{S(y_n(t|x_0)|x_0)}-1\right| = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)).$$

Proof of Proposition 14 – Let us first introduce the following notations:

$$\nu_n(t|x_0) := m_n(x_0)S[y_n(t|x_0)|x_0] \text{ and } \widehat{\varphi}_n(t|x_0) := m_n(x_0)\widehat{S}_n\left[y_n(t|x_0)|\, x_0\right].$$

Note that the functions $\nu_n(\cdot|x_0)$ and $\widehat{\varphi}_n(\cdot|x_0)$ are non-decreasing. Let us introduce the sequence $r_n := \lceil \nu_n(1|x_0)\rceil^{1/2} + 1$. For $j = 1, \ldots, r_n$, let

$$\theta_n(j) := \kappa + (j-1)\frac{1-\kappa}{r_n - 1} \in [\kappa, 1].$$

Mimicking the proof of Lemma 7, we show that

$$\sup_{t\in[\kappa,1]}\left|\frac{\widehat{\varphi}_n(t|x_0)}{\nu_n(t|x_0)} - 1\right| \leq D_{n,1} + D_{n,2}, \tag{3.28}$$

where,

$$D_{n,1} := \frac{3}{\nu_n(\kappa|x_0)}\max_{1\leq j\leq r_n-1}\left(\nu_n[\theta_n(j+1)|x_0] - \nu_n[\theta_n(j)|x_0]\right),$$

and

$$D_{n,2} := \frac{3}{\nu_n(\kappa|x_0)}\max_{1\leq j\leq r_n-1}\left|\widehat{\varphi}_n[\theta_n(j)|x_0] - \nu_n[\theta_n(j)|x_0]\right|.$$

Let us first focus on the term $D_{n,1}$. By assumption, $y_n(t|x_0) = Q(\alpha_n|x_0) + a(\alpha_n^{-1}|x_0)[L_{\gamma(x_0)}(1/t)+o(\tau_n(x_0))+w_n]$ locally uniformly on $(0,\infty)$. From Vervaat's lemma (see de Haan and Ferreira [25] Lemma A.0.2) we have (locally uniformly on $(0,\infty)$)

$$\frac{\nu_n(t|x_0)}{m_n(x_0)} = \alpha_n\left\{L^{\leftarrow}_{\gamma(x_0)}[L_{\gamma(x_0)}(t|x_0) + o(\tau_n(x_0)) + w_n] + o(\tau_n(x_0))\right\}. \tag{3.29}$$

Since $L^{\leftarrow}_{\gamma(x_0)}$ is locally lipschitz-continuous, we have for n large enough that $\nu_n(t|x_0) = m_n(x_0)t\alpha_n(1 + o(1))$, locally uniformly. Furthermore, for all $\xi \in (0,1)$, we have for n large enough that

$$(1-\xi)\kappa \leq \frac{\nu_n(t|x_0)}{m_n(x_0)\alpha_n} \leq 1 + \xi, \tag{3.30}$$

and

$$\frac{1-\xi}{1+\xi}\kappa \leq \frac{\nu_n(1|x_0)}{\nu_n(\kappa|x_0)} \leq \left[\frac{1-\xi}{1+\xi}\kappa\right]^{-1}. \tag{3.31}$$

Now, for all $(t,t') \in [\kappa,1]^2$, we have from (3.29) that

$$\begin{aligned} |\nu_n(u|x_0) - \nu_n(u'|x_0)| &= m_n(x_0)\alpha_n\Big\{L^{\leftarrow}_{\gamma(x_0)}[L_{\gamma(x_0)}(t|x_0) + o(\tau_n(x_0)) + w_n] \\ &- L^{\leftarrow}_{\gamma(x_0)}[L_{\gamma(x_0)}(t'|x_0) + o(\tau_n(x_0)) + w_n] + o(\tau_n(x_0))\Big\}. \end{aligned}$$

Since $L_{\gamma(x_0)}$ and $L^{\leftarrow}_{\gamma(x_0)}$ are locally Lipschitz-continuous, there exists a positive constant L_1 such that for n large enough

$$|\nu_n(t|x_0) - \nu_n(t'|x_0)| \le m_n(x_0)\alpha_n\left[L_1|t-t'| + o(\tau_n(x_0))\right].$$

Using (3.30) we obtain that

$$\sup_{t\in[\kappa,1]}\left\{\left|\frac{\nu_n(t|x_0)}{\nu_n(t'|x_0)} - 1\right|, \frac{|t-t'|}{1-\kappa} \le [\nu_n(1|x_0)]^{-1/2}\right\} = o(\tau_n(x_0)). \qquad (3.32)$$

Since $|\theta_n(j) - \theta_n(j+1)| = (1-\kappa)/(r_n-1) \le (1-\kappa)[\nu_n(1|x_0)]^{-1/2}$, collecting (3.31) and (3.32) lead to

$$D_{n,1} = \mathcal{O}[\tau_n(x_0)]. \qquad (3.33)$$

Let us now treat the term $D_{n,2}$. We start with the following expansion

$$\widehat{S}_n[y_n(t|x_0)|x_0] = \widehat{S}^{x_0}_n[y_n(t|x_0)] + \left(\widehat{S}_n[y_n(t|x_0)|x_0] - \widehat{S}^{x_0}_n[y_n(t|x_0)]\right),$$

where we have used the notations introduced in Chapter 2, Section 2.4.2. Hence,

$$\begin{aligned} D_{n,2} &\le \frac{3}{\nu_n(\kappa|x_0)}\max_{1\le j\le r_n-1}\left|m_n(x_0)\widehat{S}^{x_0}_n[y_n(\theta_n(j)|x_0)] - \nu_n[\theta_n(j)|x_0]\right| \\ &+ \frac{3}{\nu_n(\kappa|x_0)}\max_{1\le j\le r_n-1}\left|\widehat{\varphi}_n[\theta_n(j)|x_0] - m_n(x_0)\widehat{S}^{x_0}_n[y_n(\theta_n(j)|x_0)]\right| \\ &=: D^{(1)}_{n,2} + D^{(2)}_{n,2}. \end{aligned}$$

First, to deal with $D^{(1)}_{n,2}$, let us introduce the stochastic process

$$\widehat{\Phi}_n(t|x_0) := m_n(x_0)\widehat{S}^{x_0}_n[y_n(t|x_0)] = \sum_{i=1}^n T_{n,i}V_{n,i}(t),$$

where $T_{n,i} := \mathcal{W}_{n,i}(x_0)$ and $V_{n,i}(t) = m_n(x_0)\mathbb{I}_{(y_n(t|x_0),\infty)}(Y_i^{(x_0)})$. Let us check that this process satisfies the conditions of Lemma 8. First, it is clear that $V_{n,1}(\cdot), \ldots, V_{n,n}(\cdot)$ are independent stochastic processes distributed as the stochastic process

$$V_n(t) := m_n(x_0)\mathbb{I}_{(y_n(t|x_0),\infty)}(Q(U|x_0)),$$

where U is a standard uniform random variable. The process $V_n(\cdot)$ is clearly non-decreasing and positive. Furthermore, the $T_{n,i}$ sum to 1 and are independent of the $V_{n,i}(\cdot)$. It is also clear that $V_n(1) \leq m_n(x_0)$ everywhere and that by assumption, there exists a positive constant C_X such that

$$m_n(x_0) \max_{1\leq i\leq n} T_{n,i} \leq C_X.$$

Since $\mathbb{E}[\widehat{\Phi}_n(t|x_0)] = \nu_n(t|x_0) \to \infty$ and since (3.31) and (3.32) hold, we obtain as a direct consequence of Lemma 8 that

$$D_{n,2}^{(1)} = \mathcal{O}_{\mathbb{P}}[\tau_n(x_0)]. \tag{3.34}$$

Let us now focus on the last term $D_{n,2}^{(2)}$. Denoting by

$$B_{n,j}(x_0) := \widehat{\varphi}_n[\theta_n(j)|x_0] - m_n(x_0)\widehat{S}_n^{x_0}[y_n(\theta_n(j)|x_0)],$$

we have that

$$D_{n,2}^{(2)} = \frac{3}{\nu_n(\kappa|x_0)} \max_{1\leq j\leq r_n-1} |B_{n,j}(x_0)|.$$

Moreover, for all $\varepsilon > 0$,

$$\mathbb{P}\left[\frac{\tau_n^{-1}(x_0)}{\nu_n(\kappa|x_0)} \max_{1\leq j\leq r_n-1} |B_{n,j}(x_0)| > \varepsilon\right] \leq \sum_{j=1}^{r_n-1} \mathbb{P}\left[\frac{\tau_n^{-1}(x_0)}{\nu_n(\kappa|x_0)} |B_{n,j}(x_0)| > \varepsilon\right].$$

Since for all $j = 1, \ldots, r_n - 1$, $\mathbb{E}[|B_{n,j}(x_0)|] < \infty$, Markov's inequality entails that

$$\mathbb{P}\left[\frac{\tau_n^{-1}(x_0)}{\nu_n(\kappa|x_0)} \max_{1\leq j\leq r_n-1} |B_{n,j}(x_0)| > \varepsilon\right] \leq \frac{\varepsilon^{-1}\tau_n^{-1}(x_0)}{\nu_n(\kappa|x_0)} \sum_{j=1}^{r_n-1} \mathbb{E}[|B_{n,j}(x_0)|].$$

Let $\mathbb{X} := (X_1, \ldots, X_n)^\top$ and for all $i = 1, \ldots, n$ and $t \in [\kappa, 1]$, let

$$\Delta_{n,i}(t|x_0) := \left|\mathbb{I}_{(-\infty,S[y_n(t|x_0)|X_i])} - \mathbb{I}_{(-\infty,S[y_n(t|x_0)|x_0])}\right|(U_i),$$

where the random variables $U_1, \ldots, U_n$ are introduced in Chapter 2, Section 2.4.2. One has for all $j = 1, \ldots, r_n - 1$

$$\mathbb{E}[|B_{n,j}(x_0)|] \leq m_n(x_0)\mathbb{E}\left\{\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{E}\left[\Delta_{n,i}[\theta_n(j)|x_0]|\,\mathbb{X}\right]\right\}.$$

For all $t \in [\kappa, 1]$ and $i = 1, \ldots, n$, introducting the quantity $d_{n,i}(t|x_0) := m_n(x_0)|S[y_n(t|x_0)|X_i] - S[y_n(t|x_0)|x_0]|$ we obtain that

$$\mathbb{E}[|B_{n,j}(x_0)|] \leq 2\mathbb{E}\left\{\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)d_{n,i}[\theta_n(j)|x_0]\right\}.$$

Hence, for all $\varepsilon > 0$,

$$\mathbb{P}\left[\tau_n^{-1}(x_0)D_{n,2}^{(2)} > \varepsilon\right] \leq \frac{2\varepsilon^{-1}\tau_n^{-1}(x_0)}{\nu_n(\kappa|x_0)}(r_n - 1) \sup_{t\in[\kappa,1]} \mathbb{E}\left\{\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)d_{n,i}(t|x_0)\right\}.$$

Using (3.30) and (3.31), for n large enough, there exist ξ and a constant $C_\xi > 0$ such that

$$\mathbb{P}\left[\tau_n^{-1}(x_0)D_{n,2}^{(2)} > \varepsilon\right] \leq C_\xi\tau_n^{-1}(x_0)[m_n(x_0)\alpha_n]^{1/2}\omega(\alpha_n, \xi|x_0).$$

Since $[m_n(x_0)\alpha_n]^{1/2}\omega(\alpha_n, \xi|x_0) = o(\tau_n(x_0))$, one has

$$D_{n,2}^{(2)} = \mathcal{O}_{\mathbb{P}}[\tau_n(x_0)]. \tag{3.35}$$

Collecting (3.28), (3.33), (3.34) and (3.35) conclude the proof. ■

We are now in position to prove Proposition 13.

Proof of Proposition 13 First, the consistency of $\widehat{\gamma}_n(x_0)$ is a direct consequence of Gardes [17] Theorem 1. Indeed, Proposition 14 ensures that the assumptions of [17] Theorem 1 are satisfied. We have in turn that

$$\widehat{\gamma}_{n,-}(x_0) - \gamma_-(x_0) = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)) \tag{3.36}$$

and,

$$\frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)}\mathcal{T}_{\alpha_n}\left(\widehat{Q}_n(\cdot|x_0)\right) - \frac{\int_\kappa^1 \varphi(u)L_{\gamma_-(x_0)}(1/u)du}{\int_\kappa^1 \varphi(u)L_0(1/u)du} = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)), \tag{3.37}$$

see Gardes [17] eq. (30). We are now interested in showing the consistency of the estimator $\widehat{a}_n(\alpha_n^{-1}|x_0)$. We start with

$$\begin{aligned}\frac{\widehat{a}_n(\alpha_n^{-1}|x_0)}{a(\alpha_n^{-1}|x_0)} &= \frac{\widehat{Q}_n(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)}\mathcal{T}_{\alpha_n}^{(1)}\left(\widehat{Q}_n(\cdot|x_0)\right)\int_\kappa^1 \varphi(u)L_0(1/u)du \\ & \Big/ \int_\kappa^1 \varphi(u)L_{\widehat{\gamma}_{n,-}(x_0)}(1/u)du.\end{aligned}$$

Using the inequality $|1-\exp(x)| \leq |x| + x^2$ that holds for all $x < \ln(2)$, we have for all $u \in (\nu, 1)$,

$$\begin{aligned}&\left|L_{\widehat{\gamma}_{n,-}(x_0)}(1/u) - L_{\gamma_-(x_0)}(1/u)\right| \\ \leq\ & \int_1^{1/\kappa} v^{\gamma_-(x_0)-1}\left|\exp[(\widehat{\gamma}_{n,-}(x_0) - \gamma_-(x_0)))\ln(v)] - 1\right| dv \\ \leq\ & |\widehat{\gamma}_{n,-}(x_0) - \gamma_-(x_0)|\, q_{\gamma_-(x_0)}(1/\kappa)\left(1 + \ln^2(\kappa)\,|\widehat{\gamma}_{n,-}(x_0) - \gamma_-(x_0)|\right).\end{aligned}$$

Hence, from (3.36),

$$\left|L_{\widehat{\gamma}_{n,-}(x_0)}(1/u) - L_{\gamma_-(x_0)}(1/u)\right| = \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)). \tag{3.38}$$

It is then straightforward to check that

$$\int_\kappa^1 \varphi(u)L_{\widehat{\gamma}_{n,-}(x_0)}(1/u)du \Big/ \int_\kappa^1 \varphi(u)L_{\gamma_-(x_0)}(1/u)du = 1 + \mathcal{O}_{\mathbb{P}}(\tau_n(x_0)).$$

As a consequence, since under the assumptions of Proposition 13, $\widehat{Q}_n(\alpha_n|x_0)/Q(\alpha_n|x_0) = 1 + \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))$,

$$\begin{aligned}\frac{\widehat{a}_n(\alpha_n^{-1}|x_0)}{a(\alpha_n^{-1}|x_0)} &= \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)}\mathcal{T}_{\alpha_n}^{(1)}\left(\widehat{Q}_n(\cdot|x_0)\right)\frac{\int_\kappa^1 \varphi(u)L_0(1/u)du}{\int_\kappa^1 \varphi(u)L_{\gamma_-(x_0)}(1/u)du} \\ &\times\ (1 + \mathcal{O}_{\mathbb{P}}(\tau_n(x_0))).\end{aligned}$$

Equation (3.37) concludes the proof. ■

3.6.3 Proof of Theorem 12

This proof is based on the following lemma.

Lemma 9 *Under* **(H2)** *and (3.9),* *i)*

$$\lim_{\alpha\to 0}\frac{Q_{B_0}(\alpha|B_0^\top x_0)-Q(\alpha|x_0)}{A(\alpha^{-1}|x_0)a(\alpha^{-1}|x_0)}=0.$$

ii) For all $t>0$

$$\lim_{\alpha\to 0}\frac{\mathrm{ERV}(\alpha,t|B_0,x_0)}{A(\alpha^{-1}|x_0)}=\int_1^{t^{-1}} r^{\gamma(x_0)-1}L_{\rho(x_0)}(r)dr,$$

with

$$\mathrm{ERV}(\alpha,t|B_0,x_0):=\frac{Q_{B_0}(t\alpha|B_0^\top x_0)-Q_{B_0}(\alpha|B_0^\top x_0)}{a(\alpha^{-1}|x_0)}-L_{\gamma(x_0)}(t^{-1}),$$

iii) If there exists a sequence τ_n such that $\Delta_a(\alpha|x_0)=o(\tau_n)$ and if $A(\alpha^{-1}|x_0)=o(\tau_n)$ then

$$\frac{a(\alpha^{-1}|x_0)}{Q_{B_0}(\alpha|B_0^\top x_0)}-\gamma_+(x_0)=o(\tau_n).$$

Proof of Lemma 9 Let $\varepsilon>0$. Condition (3.9) entails that for sufficiently small α,

$$\alpha(1-\varepsilon A(\alpha^{-1}|x_0))\leq S[Q_{B_0}(\alpha|B_0^\top x_0)|x_0]\leq \alpha(1+\varepsilon A(\alpha^{-1}|x_0)).$$

Hence, applying the function $Q(\cdot|x_0)$ in each members of this inequation yields

$$Q[\alpha(1+\varepsilon A(\alpha^{-1}|x_0))|x_0]\leq Q_{B_0}(\alpha|B_0^\top x_0)\leq Q[\alpha(1-\varepsilon A(\alpha^{-1}|x_0))|x_0].$$

As a consequence

$$\frac{Q_{B_0}(\alpha|B_0^\top x_0)-Q(\alpha|x_0)}{a(\alpha^{-1}|x_0)}\leq\frac{Q[\alpha(1-\varepsilon A(\alpha^{-1}|x_0))|x_0]-Q(\alpha|x_0)}{a(\alpha^{-1}|x_0)}.$$

Since $\mathrm{ERV}(\alpha,t|B_0,x_0)\to 0$ locally uniformly on $t\in(0,\infty)$, for all $\tilde{\varepsilon}>0$, one has for α small enough

$$\frac{Q_{B_0}(\alpha|B_0^\top x_0)-Q(\alpha|x_0)}{a(\alpha^{-1}|x_0)}\leq(1+\tilde{\varepsilon})L_{\gamma(x_0)}\left(\frac{1}{1-\varepsilon A(\alpha^{-1}|x_0)}\right)$$

Since the function $L_{\gamma(x_0)}(1/\cdot)$ is locally a Lipschitz function, there exists a positive constant C_1 such that

$$\frac{Q_{B_0}(\alpha|B_0^\top x_0) - Q(\alpha|x_0)}{a(\alpha^{-1}|x_0)} \leq C_1(1+\tilde{\varepsilon})\varepsilon A(\alpha^{-1}|x_0).$$

In the same way, one can prove that there exists a positive constant C_2 such that

$$\frac{Q_{B_0}(\alpha|B_0^\top x_0) - Q(\alpha|x_0)}{a(\alpha^{-1}|x_0)} \geq -C_2(1-\tilde{\varepsilon})\varepsilon A(\alpha^{-1}|x_0),$$

proving the first part of the lemma. To prove ii), remark that

$$\begin{aligned}\mathrm{ERV}(\alpha, t|B_0, x_0) &= \mathrm{ERV}(\alpha, t|x_0) + \frac{Q_{B_0}(t\alpha|B_0^\top x_0) - Q(t\alpha|x_0)}{a(\alpha^{-1}|x_0)} \\ &- \frac{Q_{B_0}(\alpha|B_0^\top x_0) - Q(\alpha|x_0)}{a(\alpha^{-1}|x_0)}.\end{aligned}$$

Part i) of this lemma together with the fact that $A(\cdot|x_0)$ and $a(\cdot|x_0)$ are regularly varying functions conclude the proof.
Let us now focus on the point iii). We start with the decomposition

$$\frac{a(\alpha^{-1}|x_0)}{Q_{B_0}(\alpha|B_0^\top x_0)} - \gamma_+(x_0) = \Delta_a(\alpha|x_0) + \frac{a(\alpha^{-1}|x_0)}{Q(\alpha|x_0)} \frac{Q(\alpha|x_0) - Q_{B_0}(\alpha|B_0^\top x_0)}{Q_{B_0}(\alpha|B_0^\top x_0)}.$$

By assumption, $\Delta_a(\alpha^{-1}|x_0) = o(\tau_n)$. Moreover, using the point i) and since $a(\alpha^{-1}|x_0)/Q(\alpha|x_0) \to \gamma_+(x_0)$,

$$\frac{a(\alpha^{-1}|x_0)}{Q(\alpha|x_0)} \frac{Q(\alpha|x_0) - Q_{B_0}(\alpha|B_0^\top x_0)}{Q_{B_0}(\alpha|B_0^\top x_0)} = \mathcal{O}\left(\frac{a(\alpha^{-1}|x_0)}{Q_{B_0}(\alpha|B_0^\top x_0)} A(\alpha^{-1}|x_0)\right).$$

The second point of this lemma and Fraga Alves et al. [16] Lemma 3.1 entail that $a(\alpha^{-1}|x_0)/Q_{B_0}(\alpha|B_0^\top x_0) \to \gamma_+(x_0)$ as $\alpha \to 0$ and thus

$$\frac{a(\alpha^{-1}|x_0)}{Q(\alpha|x_0)} \frac{Q(\alpha|x_0) - Q_{B_0}(\alpha|B_0^\top x_0)}{Q_{B_0}(\alpha|B_0^\top x_0)} = \mathcal{O}(A(\alpha^{-1}|x_0)) = o(\tau_n),$$

by assumption. The proof is then complete. ■

Proof of Theorem 12 – We start with the decomposition

$$\frac{\widehat{Q}_n^{(E)}(\beta_n|B_0,x_0)-Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \frac{\widehat{Q}_n^{(E)}(\beta_n|B_0,x_0)-Q_{B_0}(\beta_n|B_0^\top x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} + \frac{Q_{B_0}(\beta_n|B_0^\top x_0)-Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)}$$

According to Lemma 9, we have all the required conditions ensuring that

$$\frac{\widehat{Q}_n^{(E)}(\beta_n|B_0,x_0)-Q_{B_0}(\beta_n|B_0^\top x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n(B_0,x_0)).$$

To show that, it suffices to use Corollary 6 where the conditional quantile of Y given X is replaced by the conditional quantiles of Y given $B_0^\top X$. Furthermore, part i) of Lemma 9 combining with (3.18) and the regularly varying properties of $A(\cdot|x_0)$ and $a(\cdot|x_0)$ with parameters $\rho(x_0)$ and $\gamma(x_0)$ respectively, yields to

$$\frac{Q_{B_0}(\beta_n|B_0^\top x_0)-Q(\beta_n|x_0)}{a(\alpha_n^{-1}|x_0)q_{\gamma(x_0)}(\alpha_n/\beta_n)} = \mathcal{O}\left(\frac{A(\alpha_n^{-1}|x_0)}{q_{\gamma(x_0)}(\alpha_n/\beta_n)}\right) = o(\tau_n(B_0,x_0)),$$

since by assumption $A(\alpha_n^{-1}|x_0) = o(\tau_n(B_0,x_0))$. The proof is then complete. ■

3.6.4 Proof of Corollary 8

Let us first establish the following result.

Lemma 10 *Assume that the distribution of $B_0^\top X$ admits a probability density function f_{B_0} such that $f_{B_0}(x_0) > 0$. If $k_n/(\ln\ln n) \to \infty$ and $n/k_n \to \infty$ then, for*

$$r_n = \left(\frac{2}{f_{B_0}(x_0)}\frac{k_n}{n}\right)^{1/q},$$

one has $\mathbb{P}(D_{(k_n)}(B_0,x_0) \le r_n) = 1$ for n large enough.

Proof of Lemma 10 – Let

$$N_n := \sum_{i=1}^{n} \mathbb{I}_{(-\infty,r_n)}(D_{(k_n)}(B_0,x_0))$$

be the number of covariates in the ball of center x_0 and radius $r_n = (2k_n/[nf_{B_0}(x_0)])^{1/q}$. To prove Lemma 10, it suffices to show that for n large enough, $\mathbb{P}[N_n \geq k_n] = 1$. From Lemma 2, since $nr_n^q/[\ln \ln n] \to \infty$, one as $N_n/(nr_n^q) \xrightarrow{a.s.} f_{B_0}(x_0)$. Hence, for n large enough,

$$\mathbb{P}\left[\frac{N_n}{nr_n^q} > \frac{f_{B_0}(x_0)}{2}\right] = 1.$$

The end of the proof is straightforward. ■

Lemma 10 entails that

$$\sum_{i=1}^{n} \mathbb{E}\left[\mathcal{W}_{n,i}(B_0, x_0)\left|\frac{S_{B_0}[Q_{B_0}(\beta|B_0^\top x_0)|B_0^\top X_i]}{\beta} - 1\right|\right]$$
$$\leq \sup_{\|B_0^\top(x-x_0)\|_q \leq r_n}\left|\frac{S_{B_0}[Q_{B_0}(\beta|B_0^\top x_0)|B_0^\top x]}{\beta} - 1\right|.$$

From this point, the proof of Corollary 8 follows the same lines as the proof of Corollary 4 in Chapter 2. ■

3.7 Tables and figure

Figure 3.1: Relative error as a function of β/α with $\alpha = 0.05$. Full line: ζ_{WT}, dashed line: $\zeta_{\mathrm{HT}}(\cdot|1)$, dotted line: $\zeta_{\mathrm{HT}}(\cdot|2)$ and dashed-dotted line: ζ_{SHT}.

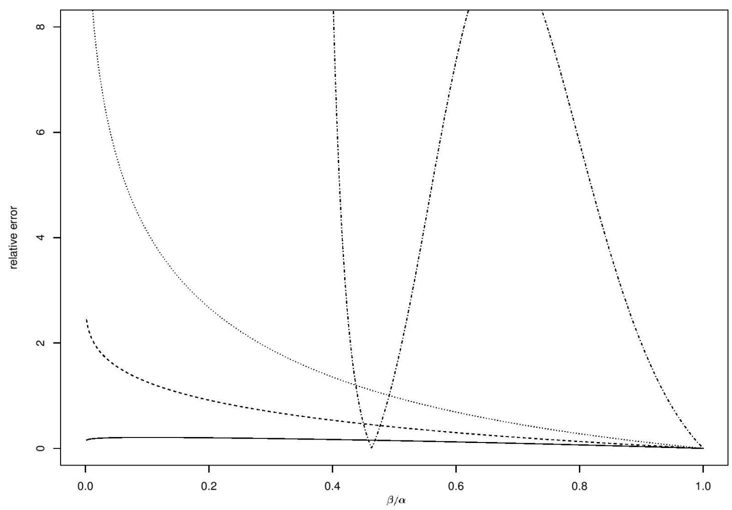


Figure 3.2: Representation of the error $E_Q(NW, B)$ with $p = 3$ for $\beta_n = 1/n$ as a function of $\lambda_n = h_n$ with $B = B_0$ (full line), $B = \hat{B}_{0,n}$ (dashed line) and $B = I_p$ (dotted line). The sequence α_n is fixed to $\alpha_n = 0.10$ and the observations are generated from model (**M1**).

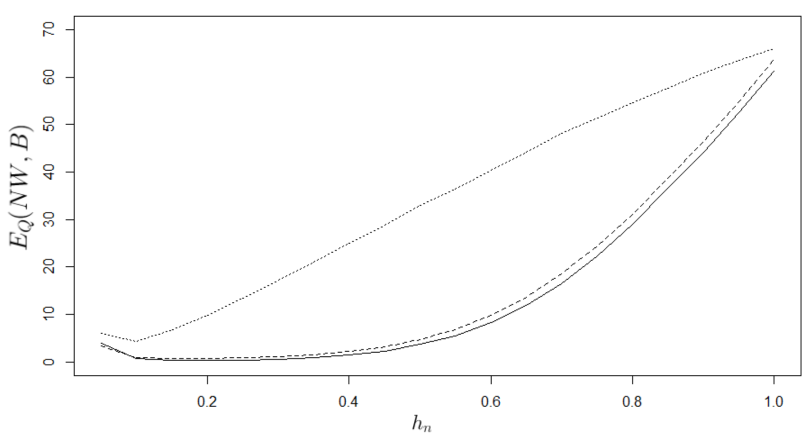


Table 3.1: Values of $E_Q(\bullet, B)$ for $B \in \{B_0, \widehat{B}_{0,n}, I_p\}$ for models **(M1)**, **(M2)** and **(M3)** with $\beta_n = 1/n$, for (1) $p = 2$, (2) $p = 3$ and (3) $p = 4$ and for two different weights: $\bullet = NW$ and $\bullet = NN$.

		$B = B_0$		$B = \hat{B}_{0,n}$		$B = I_p$	
		NW	NN	NW	NN	NW	NN
	(1)	0.549	0.660	0.554	0.744	3.169	1.208
(M1)	(2)	0.630	0.409	0.690	0.429	7.603	1.948
	(3)	0.642	0.584	0.988	0.808	15.404	5.391
	(1)	0.057	0.019	0.057	0.023	0.529	0.110
(M2)	(2)	0.092	0.023	0.096	0.024	1.003	0.282
	(3)	0.091	0.018	0.088	0.061	1.295	0.441
	(1)	0.057	0.065	0.063	0.069	0.244	0.139
(M3)	(2)	0.091	0.073	0.187	0.088	0.531	0.279
	(3)	0.100	0.065	0.361	0.219	0.661	0.423

Table 3.2: Values of $E_Q(\bullet, B)$ for $B \in \{B_0, \widehat{B}_{0,n}, I_p\}$ for models **(M1)**, **(M2)** and **(M3)** with $\beta_n = 1/(2n)$, for (1) $p = 2$, (2) $p = 3$ and (3) $p = 4$ and for two different weights: $\bullet = NW$ and $\bullet = NN$.

		$B = B_0$		$B = \hat{B}_{0,n}$		$B = I_p$	
		NW	NN	NW	NN	NW	NN
	(1)	0.752	0.956	0.761	1.109	4.645	1.766
(M1)	(2)	0.879	0.555	1.007	0.584	11.639	2.945
	(3)	0.928	0.845	1.530	1.199	25.278	8.840
	(1)	0.063	0.025	0.063	0.032	0.566	0.126
(M2)	(2)	0.154	0.030	0.158	0.032	1.682	0.316
	(3)	0.099	0.022	0.095	0.077	1.343	0.485
	(1)	0.086	0.096	0.096	0.101	0.357	0.196
(M3)	(2)	0.126	0.106	0.262	0.132	0.730	0.396
	(3)	0.140	0.096	0.504	0.317	0.918	0.612

Chapter 4

Conclusion et Perspectives

On considère un couple de variable $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$ où $p \in \mathbb{N} \setminus \{0\}$. L'objectif de cette thèse a été de développer une méthode d'estimation de quantiles extrêmes conditionnels par extrapolation de la distribution de Y sachant $\{X = x_0\}$ où x_0 appartient au support de X.

Dans le chapitre 2, nous avons introduit la condition TFO (Tail First Order Condition) qui est vérifiée non seulement par les distributions appartenant à un domaine d'attraction mais aussi par les distributions super heavy-tailed dont les fonctions de survie sont des fonctions à variations lentes. En supposant qu'une fonction de survie vérifie la condition TFO et sous d'autres conditions, on a montré la normalité asymptotique d'estimateurs de quantiles intermédiaires $\hat{Q}_n(\alpha_n|x_0) := \hat{S}_n^{\leftarrow}(\alpha_n|x_0)$ où pour tout $y \in \mathbb{R}$,

$$\hat{S}_n(y|x_0) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{\{Y_i > y\}}$$

et avec $\alpha_n \to 0$. Une première étape a consisté à montrer la normalité asymptotique de l'estimateur $\hat{S}_n(y_n(x_0)|x_0)$ où $y_n(x_0)$ est une suite convergeant vers le point terminal de $S(\cdot|x_0)$ et d'en déduire par une méthode delta celle de $\hat{Q}(\alpha_n|x_0)$. Nous avons également étudié deux cas particuliers de ces estimateurs : le cas de l'estimateur de Nadaraya-Watson et des plus proches voisins.

Dans le chapitre 3, nous avons vu que la condition TFO ne permet pas d'obtenir d'estimation pertinente de quantiles extrêmes dans le cas où $S(\cdot|x_0)$ est une distribution super heavy-tailed. Nous avons donc étudié la classe d'estimateurs de quantiles extrêmes obtenus par la condition du premier ordre classique :

$$\hat{Q}_n^{(E)}(\beta_n|x_0) := \hat{Q}_n(\alpha_n|x_0) + \hat{a}_n(1/\alpha_n|x_0)L_{\hat{\gamma}_n(x_0)}(\alpha_n/\beta_n)$$

où β_n est une suite convergeant vers 0 telle que $\beta_n/\alpha_n \to 0$ et où $\hat{a}_n(1/\alpha_n|x_0)$ et $\hat{\gamma}_n(x_0)$ sont des estimateurs de $a(1/\alpha_n|x_0)$ et $\gamma(x_0)$ proposés par Gardes [17]. Nous avons établi la consistance de cette classe d'estimateurs de quantiles extrêmes. Enfin, nous avons appliqué une méthode de réduction de dimension pour cette classe d'estimateurs qui nous permet d'avoir de meilleures performances lorsque la dimension de la covariable est grande. Comme dans le chapitre 2, les cas particuliers de l'estimateur de Nadaraya-Watson et des plus proches voisins ont été étudiés.

Plusieurs perspectives apparaissent à l'issue de cette thèse. On peut citer notamment les points suivants :

- Il reste à établir la normalité asymptotique de l'estimateur $\hat{Q}_n^{(E)}(\beta_n|x_0)$. Cela suggère de montrer notamment la normalité asymptotique du triplet :

$$\left(\hat{\gamma}_n(x_0) - \gamma(x_0), \frac{\hat{a}_n(1/\alpha_n|x_0)}{a(1/\alpha_n|x_0)} - 1, \frac{\hat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)}{a(1/\alpha_n|x_0)}\right).$$

- On pourrait utiliser d'autres estimateurs de $a(1/\alpha_n|x_0)$ et $\gamma(x_0)$. Dans Stupfler [36], l'auteur a généralisé l'estimateur des moments au cas conditionnel, où X peut être de grande dimension. Cependant, la méthode de sélection des covariables se fait selon les poids uniformes :

$$\mathcal{W}_{n,i}(x_0) = \frac{\mathbb{I}_{\{X_i \in B(x_0,h_n)\}}}{N_n(x_0,h_n)}$$

  où $B(x_0, h_n) := \{x \in \mathbb{R}^p, ||x - x_0|| \leq h_n\}$ et $N_n(x_0, h_n) := \sum_{i=1}^{n} \mathbb{I}_{\{X_i \in B(x_0,h_n)\}}$. Il serait intéressant d'adapter la méthode de Stupfler [36] au cas où les poids $\{\mathcal{W}_{n,i}(x_0), 1 \leq i \leq n\}$ forment un tableau triangulaire de variables aléatoires.

- On pourrait s'intéresser aux propriétés de l'estimateur $\hat{B}_{0,n}$ de la matrice B_0 utilisée dans la méthode de réduction de dimension, notamment à sa consistance et pour cela rependre les idées de Ichimura [27] Théorème 5.1.

Bibliography

[1] Basu, D. and Pereira, C.A.B (1983). Conditional independence in statistics. *Sankhya*, **45**, 324–337.

[2] Beirlant, J., Broniatowski, M., Teugels, J. and Vynckier, P. (1995). The mean residual life function at great age: Applications to tail estimation. *Journal of Statistical Planning and Inference*, **45(1-2)**, 21–48.

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

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

[5] Cook, R.D. and Weisberg, S. (1991). Discussion of "Sliced Inverse Regression for dimension reduction". *Journal of the American Statistical Association*, **86**, 328–332.

[6] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B), 2557–2589.

[7] Davison, A.C. and Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 62(1), 191–208.

[8] Dony, J. and Einmahl, U. (2009). Uniform in bandwidth consistency of kernel regression estimators at fixed point. *In High Dimensional Probability V: The Luminy Volume*, (pp. 308–325), Institute of Mathematical Statistics, Beachwood, Ohio, USA.

[9] Dubhashi, D. and Panconesi, A. (2009). *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press, New York.

[10] Dziewonski, A.M., Chou, T.A. and Woodhouse, J.H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, 86, 2825–2852.

[11] Ekström, G., Nettles, M. and Dziewonski, A.M. (2012). The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200–201, 1–9.

[12] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling extremal events. For insurance and finance.* Springer-Verlag, Berlin.

[13] Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, 33(3), 1380–1403.

[14] 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(2), 180–190.

[15] Fraga Alves, I., de Haan, L. and Neves, C. (2009). A test procedure for detecting super-heavy tails. *Journal of Statistical Planning and Inference*, 139, 213–227.

[16] Fraga Alves, M. I., Gomes, M. I., de Haan, L. and Neves, C. (2007). A note on second order conditions in extreme value theory: linking general and heavy tail conditions, *REVSTAT- Statistical Journal*, **5(3)**, 285–304.

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

[18] Gardes, L. (2018). Tail dimension reduction for extreme quantile estimation, *Extremes*, **21(1)**, 57–95.

[19] Gardes, L. and Girard, S. (2008). Estimation of the Weibull tail-coefficient with linear combination of upper order statistics, *Journal of Statistical Planning and Inference*, **138(5)**, 1416–1427.

[20] Giné, E. and Nickl, R. (2015). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, New York, USA.

[21] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3), 423–453.

[22] 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.

[23] Goegebeur, Y., Guillou, A. and Osmann, M. (2017). A local moment type estimator for an extreme quantile in regression with random covariates. *Communication in Statistics - Theory and Methods*, 46, 319–343.

[24] de Haan, L. (2008). Fighting the archenemy with mathematics. *Statistica Neerlandica*, 44, 45–68.

[25] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An introduction.* Springer Series in Operations Research and Financial Engineering, Springer, New York.

[26] He, X. and Ng, P. (1999). Quantile splines with several covariates. *Journal of Statistical Planning and Inference*, 75, 343–352.

[27] Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single index models. *Journal of Econometrics*, 58(1), 71–120.

[28] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.

[29] Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81, 673–680.

[30] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New-York-Berlin.

[31] Li, K.C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**, 316–327.

[32] Li, K.C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *Journal of the American Statistical Association*, **87**,1025–1039.

[33] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9, 141–142.

[34] Owen, A.B. (1987). *Nonparametric conditional estimation.* Ph.D Dissertation, Stanford University.

[35] Racine, J.S. and Li, K. (2017). Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach. *Journal of Econometrics*, 201, 72–94.

[36] Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics*, 7, 2298–2343.

[37] Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 162, 5–43.

[38] van der Vaart, A.W. and Wellner, J.A. (1996), *Weak convergence and empirical processes with applications to statistics.* Springer Series in Statistics, Springer-Verlag, Berlin.

[39] Watson, G.S. (1964). Smooth regression analysis. *Sankhya A*, 26, 359–372.

[40] Weisstein, E. W. (2003). *CRC Concise Encyclopedia of Mathematics (Second edition)*, Chapman and Hall, Boca Raton.

Dans cette thèse, on cherche à estimer des quantiles extrêmes conditionnels par la méthode d'inversion d'estimateurs locaux de la fonction de survie associée. Ces estimateurs dépendent de fonctions poids qui permettent à partir d'un échantillon de sélectionner les covariables les plus pertinentes.

Dans un premier chapitre, on s'intéresse à la normalité asymptotique de ces estimateurs. Celle-ci nécessite l'introduction d'une nouvelle condition sur la distribution d'intérêt appelée Tail First Order condition. Il est montré que cette condition est vérifiée non seulement par les distributions satisfaisant le théorème de Gnedenko-Fisher-Tippet mais également par les distributions super heavy-tailed. D'autres conditions, plus classiques, sont imposées notamment sur la nature du quantile qui doit être intermédiaire.

Dans un deuxième chapitre, on définit un nouvel estimateur de quantiles extrêmes par extrapolation et on montre sa consistance. Le problème de la dimension de la covariable est également traité.

Dans les deux chapitres, des cas particuliers sont étudiés dont le célèbre estimateur de Nadaraya-Watson ou encore l'estimateur des plus proches voisins. Les performances des différents estimateurs sont testés avec des études de simulation à distance finie. Une application à un jeu de données réelles a également été faite.

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE
UMR 7501
Université de Strasbourg
CNRS
IRMA, UMR 7501
7 rue René Descartes
F-67000 STRASBOURG

Tél. 03 68 85 01 29

irma.math.unistra.fr
irma@math.unistra.fr

IRMA 2019/007