Kernel estimators of extreme level curves

Abdelaati Daouia(1), Laurent Gardes(2), Stéphane Girard(2,⋆) & Alexandre Lekina(2)

(1) Toulouse School of Economics (GREMAQ), University of Toulouse, Aile J.J. Laffont, 21 allée de Brienne, 31000 Toulouse, France.

(2) INRIA Rhône-Alpes, team Mistis, Inovallée, 655, av. de l'Europe, Montbonnot, 38334 Saint-Ismier cedex, France.

(⋆) Stephane.Girard@inrialpes.fr (corresponding author)

Abstract

We address the estimation of extreme level curves of heavy-tailed distributions. This problem is equivalent to estimating quantiles when covariate information is available and in the case where their order converges to one as the sample size increases. We show that, under some conditions, these so-called "extreme conditional quantiles" can still be estimated through a kernel estimator of the conditional survival function. Sufficient conditions on the rate of convergence of their order to one are provided to obtain asymptotically Gaussian distributed estimators. These results are illustrated both on simulated and real datasets.

Keywords: Conditional quantiles, heavy-tail distributions, kernel estimator, extreme-values.
AMS 2000 subject classification: 62G32, 62G30, 62E20.

1 Introduction

Let (X_i, Y_i), $i = 1, \dots, n$ be independent copies of a random pair (X, Y) in $\mathbb{R}^d \times \mathbb{R}$. We address the problem of estimating extreme level curves, defined as the graphes of the functions $x \in \mathbb{R}^d \mapsto q(\alpha_n|x) \in \mathbb{R}$ verifying

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

where $\alpha_n \to 0$ as $n \to \infty$. In such a case, $q(\alpha_n|x)$ is referred to as an extreme conditional quantile in contrast to classical conditional quantiles (sometimes called classical regression quantiles) for which $\alpha_n = \alpha$ is fixed in $(0, 1)$. While the nonparametric estimation of classical regression quantiles has been extensively studied (see for instance the seminal papers [29, 32]), less attention has been paid to extreme conditional quantiles. Parametric models are proposed in [11, 31] whereas semi-parametric methods are considered in [3, 22]. Fully non-parametric estimators have been first introduced in [10], where a local polynomial modelling of the extreme observations is used. Similarly, spline estimators are fitted in [8] through a penalized maximum likelihood method. In both cases, the authors focus on univariate covariates and on the finite sample properties of the estimators. An important literature is devoted to the particular case where the conditional distribution of Y given $X = x$ has a finite endpoint $\varphi(x)$. The function φ is referred to as the frontier and can be estimated through an estimator of the conditional quantile $q(\alpha_n|x)$ with $\alpha_n \to 0$. As an example, a kernel estimator of φ is proposed in [19] with $\alpha_n = 1/n$, the asymptotic normality being proved only in the situation where Y given $X = x$ is uniformly distributed on $[0, \varphi(x)]$. In this latter situation, regression on the extreme values of the sample has also been introduced in [16, 18, 20, 27], the case where the density of Y given $X = x$ is lower bounded being first studied by Geffroy [17]. Extensions are provided in [24, 25] to densities of Y given $X = x$ decreasing as a power of the distance from the boundary. We refer to [26] for more information on this topic.

Estimation of unconditional extreme quantiles is also widely studied since the introduction of Weissman estimator [34], dedicated to heavy tail distributions, and Dekkers and de Haan estimator [12] adapted to the general case.

In this paper, we focus on the case where the conditional distribution of Y given $X = x$ has an infinite endpoint and is heavy-tailed, an analytical characterisation of this property being given in the next section. In such a case, the frontier function does not exist and $q(\alpha_n|x) \to \infty$ as $n \to \infty$. Nevertheless, we show, under some mild conditions, that extreme regression quantiles $q(\alpha_n|x)$ can still be estimated through a kernel estimator of $\mathbb{P}(Y > .|x)$. We provide sufficient conditions on the rate of convergence of α_n to 0 under which our estimator is asymptotically Gaussian distributed.

Assumptions are introduced and discussed in Section 2. Our main results are provided in Section 3 and illustrated on simulated data in Section 4. An example of application to real data is presented in Section 5. Proofs are postponed to the Appendix.

2 Notations and assumptions

The conditional survival function (csf) of Y given $X = x$ is denoted by $\bar{F}(y|x) = \mathbb{P}(Y > y|X = x)$ and the point distribution function (pdf) of X is denoted by g. The kernel estimator of $\bar{F}(y|x)$ is defined for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ by

$$\hat{\bar{F}}_n(y|x) = \sum_{i=1}^{n} K_h(x - X_i)\mathbb{I}\{Y_i > y\} \Bigg/ \sum_{i=1}^{n} K_h(x - X_i), \tag{1}$$

where $\mathbb{I}\{.\}$ is the indicator function and $h = h_n$ is a nonrandom sequence such that $h \to 0$ as $n \to \infty$. We have also introduced $K_h(t) = K(t/h)/h^d$ where K is a pdf on $\mathbb{R}^d$. In this context, h is called the window-width. In Theorem 1, the asymptotic distribution of (1) is established when estimating small tail probabilities, *i.e* when $y = y_n$ goes to infinity with the sample size n. Similarly, the kernel estimators of conditional quantiles $q(\alpha|x)$ are defined via the generalised inverse of $\hat{\bar{F}}_n(.|x)$:

$$\hat{q}_n(\alpha|x) = \hat{\bar{F}}_n^{\leftarrow}(\alpha|x) = \inf\{t,\ \hat{\bar{F}}_n(t|x) \le \alpha\}, \tag{2}$$

for all $\alpha \in (0,1)$. Many authors are interested in this type of estimator for fixed $\alpha \in (0,1)$: weak and strong consistency are proved respectively in [32] and [14], asymptotic normality being established in [33, 30, 4]. In Theorem 2, the asymptotic distribution of (2) is investigated when estimating extreme quantiles, *i.e* when $\alpha = \alpha_n$ goes to 0 as the sample size n goes to infinity. The asymptotic behavior of such estimators depends on the nature of the conditional distribution tail. In this paper, we focus on heavy tails. More specifically, we assume that the csf satisfies

(F.1): $\bar{F}(y|x) = y^{-1/\gamma(x)}\ell(y|x)$,

where $\gamma(.)$ is a positive function of the covariate x and, for x fixed, $\ell(.|x)$ is a slowly-varying function at infinity, *i.e* for all $\lambda > 0$,

$$\lim_{y\to\infty} \frac{\ell(\lambda y|x)}{\ell(y|x)} = 1. \tag{3}$$

To summarize, **(F.1)** amounts to assuming that the conditional distribution of Y given $X = x$ is in the Fréchet maximum domain of attraction. In this context, $\gamma(x)$ is referred to as the conditional tail index since it tunes the tail heaviness of the conditional distribution of Y given $X = x$. Assumption **(F.1)** is also equivalent to stating that $\bar{F}(.|x)$ is regularly varying at infinity with index $-1/\gamma(x)$ *i.e*

$$\lim_{y\to\infty} \frac{\bar{F}(\lambda y|x)}{\bar{F}(y|x)} = \lambda^{-1/\gamma(x)}$$

for all $\lambda > 0$. As remarked in [7], p.15, one can assume that

(F.2): $\ell(.|x)$ is normalised,

without loosing generality. In such a case, the Karamata representation (see [7], Theorem 1.3.1) of the slowly-varying function can be written as

$$\ell(y|x) = c(x) \exp\left(\int_1^y \frac{\varepsilon(u|x)}{u} du\right), \tag{4}$$

where $c(.)$ is a positive function and $\varepsilon(y|x) \to 0$ as $y \to \infty$. Thus, $\ell(.|x)$ is differentiable and the auxiliary function is given by $\varepsilon(y|x) = y\ell'(y|x)/\ell(y|x)$. This function plays an important role in extreme-value theory since it drives the speed of convergence in (3) and more generally the bias of extreme-value estimators. Therefore, it may be of interest to precise how it converges to 0. For instance, in [1, 2, 21], the auxiliary function is supposed to be regularly varying and the estimation of the corresponding regular variation index is addressed. Here, we limit ourselves to assuming that

(F.3): $|\varepsilon(.|x)|$ is ultimately non-increasing.

Some Lipschitz assumptions are also required. For all $(x,x') \in \mathbb{R}^d \times \mathbb{R}^d$, the euclidean distance between x and x' is denoted by $d(x,x')$ and the following assumptions are introduced:

(L.1): There exists $c_\gamma > 0$ such that $\left|\dfrac{1}{\gamma(x)} - \dfrac{1}{\gamma(x')}\right| \leq c_\gamma d(x,x')$.

(L.2): There exist $c_\ell > 0$ and $y_0 > 1$ such that $\sup_{y \geq y_0} \left|\dfrac{\log \ell(y|x)}{\log y} - \dfrac{\log \ell(y|x')}{\log y}\right| \leq c_\ell d(x,x')$.

(L.3): There exists $c_g > 0$ such that $|g(x) - g(x')| \leq c_g d(x,x')$.

The last assumption is standard in the kernel estimation framework.

(K): K is a bounded pdf on $\mathbb{R}^d$, with support S included in B, the unit ball of $\mathbb{R}^d$.

3 Main results

Let us first focus on the estimation of small tail probabilities $\bar{F}(y_n|x)$ when $y_n \to \infty$ as $n \to \infty$. The following result provides sufficient conditions for the asymptotic normality of $\hat{\bar{F}}_n(y_n|x)$.

Theorem 1 *Suppose* **(F.1)**, **(L.1)**, **(L.2)**, **(L.3)** *and* **(K)** *hold. Let us introduce*

- $0 < a_1 < a_2 < \cdots < a_J$ *where J is a positive integer,*
- $y_n \to \infty$ *such that* $nh^d \bar{F}(y_n|x) \to \infty$ *and* $nh^{d+2}\log^2(y_n)\bar{F}(y_n|x) \to 0$ *as* $n \to \infty$,
- $y_{n,j} = a_j y_n(1+o(1))$ *for* $j = 1, \ldots, J$.

Then, for all $x \in \mathbb{R}^d$ such that $g(x) > 0$, the random vector

$$\left\{ \sqrt{nh^d \bar{F}(y_n|x)} \left(\frac{\hat{\bar{F}}_n(y_{n,j}|x)}{\bar{F}(y_{n,j}|x)} - 1 \right) \right\}_{j=1,\ldots,J}$$

is asymptotically Gaussian, centered, with covariance matrix $\frac{\|K\|_2^2}{g(x)} C(x)$ *where* $C_{j,j'}(x) = a_{j \wedge j'}^{1/\gamma(x)}$ *for* $(j,j') \in \{1,\ldots,J\}^2$.

Let us highlight that $nh^d \bar{F}(y_n|x) \to \infty$ is a necessary and sufficient condition for the almost sure presence of at least one sample point in the region $B(x,h) \times (y_n, \infty)$ of $\mathbb{R}^{d+1}$, see Lemma 3 in Appendix. Thus, this natural condition states that one cannot estimate small tail probabilities out of the sample using $\hat{\bar{F}}_n$. Considering now the estimation of extreme quantiles $q(\alpha_n|x)$ when $\alpha_n \to 0$ as $n \to \infty$, the asymptotic normality of $\hat{q}_n(\alpha_n|x)$ can be established under similar conditions.

Theorem 2 *Suppose* **(F.1)**, **(F.2)**, **(L.1)**, **(L.2)**, **(L.3)** *and* **(K)** *hold. Let us introduce*

- $\tau_1 > \tau_2 > \cdots > \tau_J > 0$ *where* J *is a positive integer,*
- $\alpha_n \to 0$ *such that* $nh^d\alpha_n \to \infty$ *and* $nh^{d+2}\alpha_n \log^2(\alpha_n) \to 0$ *as* $n \to \infty$,
- $\alpha_{n,j} = \tau_j \alpha_n(1 + o(1))$ *for* $j = 1, \ldots, J$.

Then, for all $x \in \mathbb{R}^d$ *such that* $g(x) > 0$, *the random vector*

$$\left\{ \sqrt{nh^d\alpha_n} \left(\frac{\hat{q}_n(\alpha_{n,j}|x)}{q(\alpha_{n,j}|x)} - 1 \right) \right\}_{j=1,\ldots,J}$$

is asymptotically Gaussian, centered, with covariance matrix

$$\|K\|_2^2 \frac{\gamma^2(x)}{g(x)} \Sigma \tag{5}$$

where $\Sigma_{j,j'} = 1/\tau_{j\wedge j'}$ *for* $(j, j') \in \{1, \ldots, J\}^2$.

Compared to [4], Theorem 6.4, where $\alpha_n = \alpha$ is fixed in $(0, 1)$, the asymptotic variance is larger in Theorem 2 since it involves the additional term $1/\alpha_n \to \infty$ as $n \to \infty$. Let us also highlight the important role of the tail index $\gamma(x)$. From the asymptotic variance point of view, our model is equivalent to a csf with constant tail index $\tilde{\gamma}(x) = 1$ and a pdf proportional to $\tilde{g}(x) = g(x)/\gamma^2(x)$. In such a case, the number of points in the ball $B(x, h)$ is asymptotically inversely proportional to $\gamma^2(x)$. A large value of the tail index at x thus implies a difficult estimation of $q(\alpha_n|x)$.

Remark 1 *Clearly, condition* $nh^d\alpha_n \to \infty$ *implies that, for* n *large enough,* $h > (n\alpha)^{-1/d}$. *Replacing in condition* $nh^{d+2}\alpha_n \log^2(\alpha_n) \to 0$ *entails*

$$\frac{n\alpha_n}{\log^d(1/\alpha_n)} \to \infty.$$

This condition provides a lower bound on the order of the extreme quantiles for the asymptotic normality of kernel estimators.

Remark 2 *Suppose* $n\alpha_n \log^2(\alpha_n) \to \infty$ *as* $n \to 0$. *To fulfill the assumptions of Theorem 2, one can choose* $h = \eta_n(n\alpha_n \log^2(\alpha_n))^{-1/(d+2)}$ *where* (η_n) *is a sequence tending to zero arbitrarily slowly. This choice yields an asymptotic variance proportional to*

$$\frac{1}{nh^d\alpha_n} = \eta_n^{-d} \left(\frac{n\alpha_n}{\log^d(\alpha_n)} \right)^{-\frac{2}{d+2}}.$$

Note that, for $d = 0$, *we find back the variance of estimators dedicated to unconditional extreme quantiles.*

A kernel version of Pickands estimator [28] for the conditional tail index $\gamma(x)$ can be proposed on the basis of $\hat{q}_n(.|x)$, the kernel estimator defined in (2):

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

where $k = k_n$ is an intermediate sequence *i.e* such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. The asymptotic normality of $\hat{\gamma}_n(x)$ is a consequence of Theorem 2.

Corollary 1 *Suppose* **(F.1)**, **(F.2)**, **(F.3)**, **(L.1)**, **(L.2)**, **(L.3)** *and* **(K)** *hold. Let* $\sigma_n = (kh^d)^{-1/2}$ *where k is an intermediate sequence. If $\sigma_n \to 0$, $\sigma_n^{-1} h \log(k/n) \to 0$ and $\sigma_n^{-1} \varepsilon(q(2k/n|x)|x) \to 0$ as $n \to \infty$, then, for all $x \in \mathbb{R}^d$ such that $g(x) > 0$, $\sigma_n^{-1}(\hat{\gamma}_n(x) - \gamma(x))$ converges to a centered Gaussian random variable with variance*

$$\frac{\|K\|_2^2}{g(x)} \frac{\gamma^2(x)(2^{2\gamma(x)+1}+1)^2}{4(\log 2)^2(2^{\gamma(x)}-1)^2}. \tag{6}$$

Let us remark that (6) is, up to the scale factor $\|K\|_2^2/g(x)$, the variance of the classical Pickands estimator, see for instance [23], Theorem 3.3.5. Besides, introducing

$$\hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i) \tag{7}$$

the classical kernel estimator of the pdf $g(x)$, one may obtain pointwise confidence intervals for extreme conditional quantiles by plugging $\hat{g}_n(x)$ and $\hat{\gamma}_n(x)$ in (5). Indeed, both estimates are consistent, as shown in Lemma 4 and Corollary 1.

4 Numerical experiments on simulated data

Here, we limit ourselves to unidimensional random variables X ($d = 1$) uniformly distributed on $E = [0, 1]$. Besides, Y given $X = x$ is Fréchet distributed, its csf is given by

$$\bar{F}(y|x) = \exp\left(-y^{-1/\gamma(x)}\right),$$

and the following conditional tail index has been chosen:

$$\gamma(x) = \frac{1}{2}\left(\frac{1}{10} + \sin(\pi x)\right)\left(\frac{11}{10} - \frac{1}{2}\exp\left(-64(x-1/2)^2\right)\right).$$

We focus on the estimation of conditional extreme quantiles

$$q(\alpha_n|x) = (-\log \alpha_n)^{-\gamma(x)}$$

of order $\alpha_n = 5\log(n)/n$ which is inspired from the lower bound given in Remark 1. To this end, we use the estimator introduced in (2) with a bi-quadratic kernel defined as

$$K(x) = \frac{15}{16}(1-x^2)^2 \mathbb{I}\{|x| \le 1\}.$$

The choice of the bandwidth h is an important issue. In the following, we propose a data-driven strategy to select its value in a set $\mathcal{H} = \{h_1 \le h_2 \le \cdots \le h_M\}$ where $h_1 = 1/(5 \log n)$ and $h_M = 1/2$. The minimum value h_1 is chosen to obtain approximately $2nh_1\alpha_n = 2$ observations in the area $[x - h_1, x + h_1] \times [q(\alpha_n|x), \infty)$ while the maximum value h_M corresponds to a smoothing on the whole $[0, 1]$ interval. The points $h_2, \ldots, h_{M-1}$ are regularly distributed in $[h_1, h_M]$ and $M = 50$ is used in practice. Two strategies are compared:

- The first one is derived from the cross-validation approach introduced in [35] and implemented for instance in [15]:

$$h_{cv} = \arg\min_{h \in \mathcal{H}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathbb{I}\{Y_i \ge Y_j\} - \hat{\bar{F}}_{n,-i}(Y_j|X_i) \right)^2,$$

  where $\hat{\bar{F}}_{n,-i}$ is the estimator (depending on h) given in (1) computed from the sample $\{(X_\ell, Y_\ell),\ 1 \le \ell \le n,\ \ell \ne i\}$.

- The second one is the oracle strategy which consists in minimizing a distance Δ between the estimated conditional extreme quantile and the true one:

$$h_{oracle} = \arg\min_{h \in \mathcal{H}} \Delta(q(\alpha_n|.), \hat{q}_n(\alpha_n|.)),$$

  with

$$\Delta(q(\alpha_n|.), \hat{q}_n(\alpha_n|.)) = \left\{ \frac{1}{L} \sum_{\ell=1}^{L} \left(\hat{q}_n(\alpha_n|t_\ell) - q(\alpha_n|t_\ell) \right)^2 \right\}^{1/2} \tag{8}$$

  and where $t_1, \ldots, t_L$ are regularly distributed on $[0, 1]$. Of course, this method cannot be applied in practical situations where $q(\alpha_n|.)$ is unknown. However, it provides us the lower bound on the distance Δ that can be reached with our estimator.

The finite sample performance of these strategies is assessed on $N = 100$ replications of the samples $\{(X_i, Y_i), i = 1, \ldots, n\}$ of size $n \in \{300, 1000\}$. The bandwidths $h_{cv}^{(p)}$ and $h_{oracle}^{(p)}$ are computed on each replication $p \in \{1, \ldots, N\}$ with the two strategies and the corresponding errors (8) are denoted by $\Delta_{cv}^{(p)}$ and $\Delta_{oracle}^{(p)}$. As an illustration, the empirical distributions of $\Delta_{cv}^{(p)}$ and $\Delta_{oracle}^{(p)}$, obtained on samples of size $n = 300$, are superimposed on Figure 1. It appears that the mean errors are approximately equal. Let us also remark that the cross-validation errors seem to have a heavier right-tail than the oracle errors. Similar conclusions have been drawn for $n = 1000$. Let us now focus on the estimators $\hat{q}_n(\alpha_n|.)$ computed from the bandwidths associated to the quantiles of order 10%, 50% and 90% of the cross-validation error distribution. They respectively correspond to the best 10% estimator, median estimator and worst 10% estimator obtained with the cross-validation criterion. In Figure 2 ($n = 300$) and Figure 3 ($n = 1000$) they are superimposed to the true conditional quantile as well as to the estimated one with the oracle strategy on the same replication. The vertical axis is represented in a logarithmic scale for the visualization sake. One can see that the best 10% and median estimators obtained with the cross-validation strategy are quite stable and as good as the ones obtained by the oracle strategy. In contrast, the worst 10% estimator obtained with the cross-validation strategy is less accurate than the corresponding oracle estimator, the reason being that the cross-validation strategy is more sensitive to the extreme points than the oracle one. This phenomena may explain the heavy right-tail of the cross-validation error observed on Figure 1. However, the cross-validation criterion provides satisfactory results in most of the cases. Finally, let us observe that the results obtained with $n = 300$ and $n = 1000$ are qualitatively equivalent. This phenomena is a consequence of $\alpha_{1000} < \alpha_{300}$ which means that estimating $q(\alpha_{1000}|.)$ is more difficult than estimating $q(\alpha_{300}|.)$.

5 Illustration on real data

As an illustration, we propose an application of our methodology in a hyperspectral remote sensing framework. The original data consists of $n = 3184$ pairs denoted by (S_i, P_i), $i = 1, \ldots, n$. Each S_i is a spectra (in some high-dimensional space E) representing the intensity of light energy reflected from materials on Mars as it varies across different wavelengths. The analysis of these spectral signatures allows the identification of the physical, chemical or mineralogical properties of the surface that may help to understand the geological history of planets. We refer to [5] for a detailed presentation of the physical context. Here, we focus on the physical parameter $P_i \in [0, 1]$ representing a CO_2 proportion. Reduction dimension techniques [6] have shown that a 1-dimensional predictor $X_i =< b, S_i >$ is sufficient to predict P_i, with $b \in E$ and where $< .,. >$ denotes a dot-product in E that we do not precise here. Finally, the variables of interest are $Y_i = (1 - P_i)^{-1} - (1 - P_{\min})^{-1}$ where $P_{\min}$ is chosen such that $Y_i \in [0, \infty)$. The resulting scatter-plot is depicted on Figure 4. We focus on the estimation of conditional extreme quantiles of order $\alpha_n = \zeta \log(n)/n$ with $\zeta \in \{5, 10, 20\}$. The above described procedure yields $h_{cv} \simeq 0.12$. Let us denote by Z_j, $j = 1, \ldots, m(x)$ the observations Y_i such that $X_i \in [x - h_{cv}, x + h_{cv}]$, $i = 1, \ldots, n$. Our approach relies on the property that the Z_j, $j = 1, \ldots, m(x)$ are approximately distributed from **(F.1)**. This assumption can be graphically checked on the QQ-plots obtained by drawing k log-spacings versus standard exponential quantiles:

$$\left(\log \frac{k}{j}, \log \frac{Z_{m(x)-j+1,m(x)}}{Z_{m(x)-k,m(x)}},\ j = 1, \ldots, k \right).$$

These QQ-plots are based on the property that, under **(F.1)**, the k log-spacings $\log(Z_{m(x)-j+1,m(x)}) - \log(Z_{m(x)-k+1,m(x)})$ are approximately distributed from an exponential distribution with scale parameter $\gamma(x)$. See [13], Section 6.2, for a review on exploratory data analysis methods for extremes. The obtained QQ-plots at three different locations ($x = 0.25$, $x = 0.50$ and $x = 0.75$) are presented on Figure 5 with $k = 50$. Let us note that the plots are approximately linear, confirming the adequation of the heavy-tail model **(F.1)** to the dataset. The very different slopes indicate a strong heterogeneity of the sample in terms of tail behaviour. The obtained extreme level curves are superimposed to the scatter-plot on Figure 4.

6 Appendix: Proofs

6.1 Preliminary results

The first lemma is dedicated to the control of the local variations of the csf with respect to the covariate x on a neighborhood of size h when the quantity of interest y goes to infinity.

Lemma 1 *Suppose* **(F.1)**, **(L.1)** *and* **(L.2)** *hold. If* $y_n \to \infty$ *and* $h \log y_n \to 0$ *as* $n \to \infty$*, then*

$$\sup_{d(x,x')\leq h} \left| \frac{\bar{F}(y_n|x)}{\bar{F}(y_n|x')} - 1 \right| = O(h \log y_n).$$

Proof. Assumption **(F.1)** yields

$$\begin{aligned} \left| \log\left(\frac{\bar{F}(y_n|x)}{\bar{F}(y_n|x')} \right) \right| &\leq |\log y_n| \left(\left| \frac{1}{\gamma(x)} - \frac{1}{\gamma(x')} \right| + \left| \frac{\log \ell(y_n|x)}{\log y_n} - \frac{\log \ell(y_n|x')}{\log y_n} \right| \right) \\ &\leq (c_\gamma + c_\ell) \log y_n d(x,x'), \end{aligned}$$

from **(L.1)**, **(L.2)** and since, for n large enough, $y_n > 1$. Thus,

$$\sup_{d(x,x')\leq h} \left| \log\left(\frac{\bar{F}(y_n|x)}{\bar{F}(y_n|x')} \right) \right| = O(h \log y_n) \to 0$$

as $n \to \infty$ and taking account of $\log(u+1) \sim u$ as $u \to 0$ gives the result. ■

The second lemma is also of analytical nature. It provides a second order asymptotic expansion of the quantile function.

Lemma 2 *Suppose* **(F.1)**, **(F.2)** *and* **(F.3)** *hold. Let* $0 < \lambda^- < \lambda^+ < 1$ *and* $\alpha_n \to 0$ *as* $n \to \infty$. *Then,*

$$\sup_{\lambda \in [\lambda^-, \lambda^+]} \left| \lambda^{\gamma(x)} \frac{q(\lambda\alpha_n|x)}{q(\alpha_n|x)} - 1 \right| = O(\varepsilon(q(\alpha_n|x)|x)).$$

Proof. From **(F.1)**, $\bar{F}(.|x)$ is a regularly varying function at infinity with index $-1/\gamma(x)$. As a consequence, $q(.|x)$ is a regularly varying function at 0 with index $-\gamma(x)$. We thus have

$$\Delta_n(\lambda) := \lambda^{\gamma(x)} \frac{q(\lambda\alpha_n|x)}{q(\alpha_n|x)} - 1 \to 0$$

as $n \to \infty$, entailing

$$
\begin{aligned}
\Delta_n(\lambda) &= (\log q(\lambda\alpha_n|x) - \log q(\alpha_n|x) + \gamma(x)\log\lambda)(1+o(1)) \\
&= (\varphi(\log\lambda + \log\alpha_n, x) - \varphi(\log\alpha_n, x) + \gamma(x)\log\lambda)(1+o(1)),
\end{aligned}
$$

where we have introduced $\varphi(., x) = \log q(\exp(.)|x)$. Under **(F.1)** and **(F.2)**, Karamata representation (4) holds and thus $\varphi(., x)$ is continuously differentiable. A first order expansion shows that there exists $\theta_n \in (\lambda\alpha_n, \alpha_n)$ such that

$$
\begin{aligned}
\Delta_n(\lambda) &= (\gamma(x) + \varphi'(\log\theta_n, x))\log(\lambda)(1+o(1)) \\
&= \left(\gamma(x) + \frac{\bar{F}(q(\theta_n|x)|x)}{\bar{F}'(q(\theta_n|x)|x)q(\theta_n|x)}\right)\log(\lambda)(1+o(1)) \\
&= \left(1 + \frac{1}{\dfrac{\gamma(x)\ell'(q(\theta_n|x)|x)q(\theta_n|x)}{\ell(q(\theta_n|x)|x)} - 1}\right)\gamma(x)\log(\lambda)(1+o(1)) \\
&= \left(1 + \frac{1}{\gamma(x)\varepsilon(q(\theta_n|x)|x) - 1}\right)\gamma(x)\log(\lambda)(1+o(1)) \\
&= -\gamma^2(x)\varepsilon(q(\theta_n|x)|x)\log(\lambda)(1+o(1)).
\end{aligned}
$$

Since $q(.|x)$ and $|\varepsilon(.|x)|$ are both ultimately non-increasing, it follows that $|\varepsilon(q(\theta_n|x)|x)| \le |\varepsilon(q(\alpha_n|x)|x)|$ and thus

$$
|\Delta_n(\lambda)| \le 2\gamma^2(x)\log(1/\lambda)|\varepsilon(q(\alpha_n|x)|x)| = O(\varepsilon(q(\alpha_n|x)|x)),
$$

uniformly in $\lambda \in [\lambda^-, \lambda^+]$. The result is proved. ■

The following lemma provides a geometrical interpretation of the condition $nh^d\bar{F}(y_n|x) \to \infty$ as $n \to \infty$.

Lemma 3 *Suppose* **(F.1)**, **(L.1)**, **(L.2)**, **(L.3)** *hold and let $y_n \to \infty$ such that $h\log y_n \to 0$ as $n \to \infty$. Consider the region of $\mathbb{R}^{d+1}$ defined as $R_n(x) = B(x,h) \times (y_n, \infty)$ where $x \in \mathbb{R}^d$ is such that $g(x) > 0$. Then, $\mathbb{P}(\exists i \in \{1,\dots,n\}, (X_i, Y_i) \in R_n(x)) \to 1$ as $n \to \infty$ if, and only if, $nh^d\bar{F}(y_n|x) \to \infty$.*

Proof. Standard considerations lead to

$$
\mathbb{P}(\exists i \in \{1,\dots,n\}, (X_i, Y_i) \in R_n(x)) = 1 - (1 - \mathbb{P}((X_1, Y_1) \in R_n(x))^n, \tag{9}
$$

and, in view of **(L.3)** and Lemma 1,

$$
\begin{aligned}
\mathbb{P}((X_1, Y_1) \in R_n(x))) &= \int_{B(x,h)} \bar{F}(y_n|u)g(u)du \\
&= \bar{F}(y_n|x)g(x)(1 + O(h\log y_n))\int_{B(x,h)} du \\
&= v_d h^d \bar{F}(y_n|x)g(x)(1 + O(h\log y_n)),
\end{aligned}
$$

where v_d is the volume of the unit ball in $\mathbb{R}^d$. Clearly, this probability converges to 0 as $n \to \infty$ and thus (9) can be rewritten as

$$\mathbb{P}(\exists i \in \{1,\dots,n\}, (X_i, Y_i) \in R_n(x)) = 1 - \exp\left(-v_d g(x) n h^d \bar{F}(y_n|x)(1+o(1))\right),$$

which converges to 1 if and only if $nh^d \bar{F}(y_n|x) \to \infty$. ■

Let us remark that the kernel estimator (1) can be rewritten as $\hat{\bar{F}}_n(y|x) = \hat{\psi}_n(y,x)/\hat{g}_n(x)$ where

$$\hat{\psi}_n(y,x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)\mathbb{I}\{Y_i > y\}$$

is an estimator of $\psi(y,x) = \bar{F}(y|x)g(x)$ and $\hat{g}_n(x)$ is the classical kernel estimator (7) of the pdf $g(x)$. Lemma 4 gives standard results on the kernel estimator (see [9], Proposition 2.1 and Proposition 2.2 for a proof) whereas Lemma 5 is dedicated to the asymptotic properties of $\hat{\psi}_n(y,x)$.

Lemma 4 *Suppose* **(L.3)**, **(K)** *hold. If* $nh^d \to \infty$, *then, for all* $x \in \mathbb{R}^d$,

(i) $\mathbb{E}(\hat{g}_n(x) - g(x)) = O(h)$,

(ii) $var(\hat{g}_n(x)) = \dfrac{g(x)\|K\|_2^2}{nh^d}(1+o(1))$.

Therefore, under the assumptions of the above lemma, $\hat{g}_n(x)$ converges to $g(x)$ in probability.

Lemma 5 *Suppose* **(F.1)**, **(L.1)**, **(L.2)**, **(L.3)** *and* **(K)** *hold. Let us introduce*

- $0 < a_1 < a_2 < \cdots < a_J$ *where* J *is a positive integer,*
- $y_n \to \infty$ *such that* $h \log y_n \to 0$ *and* $nh^d \bar{F}(y_n|x) \to \infty$ *as* $n \to \infty$,
- $y_{n,j} = a_j y_n(1+o(1))$ *for* $j = 1,\dots,J$.

Then, for all $x \in \mathbb{R}^d$ *such that* $g(x) > 0$,

(i) $\mathbb{E}(\hat{\psi}_n(y_{n,j},x)) = \psi(y_{n,j},x)(1+O(h\log y_n))$, *for* $j = 1,\dots,J$.

(ii) *The random vector*

$$\left\{\sqrt{nh^d\psi(y_n,x)}\left(\frac{\hat{\psi}_n(y_{n,j},x) - \mathbb{E}(\hat{\psi}_n(y_{n,j},x))}{\psi(y_{n,j},x)}\right)\right\}_{j=1,\dots,J}$$

is asymptotically Gaussian, centered, with covariance matrix $\|K\|_2^2 C(x)$ *where* $C_{j,j'}(x) = a_{j\wedge j'}^{1/\gamma(x)}$ *for* $(j,j') \in \{1,\dots,J\}^2$.

Proof. **(i)** Since the (X_i, Y_i), $i = 1,\dots,n$ are identically distributed, we have

$$\begin{aligned}\mathbb{E}(\hat{\psi}_n(y_{n,j},x)) &= \int_{\mathbb{R}^d} K_h(x-t)\bar{F}(y_{n,j}|t)g(t)dt \\ &= \int_S K(u)\bar{F}(y_{n,j}|x-hu)g(x-hu)du,\end{aligned}$$

under **(K)**. Let us now consider

$$\begin{aligned}
|\mathbb{E}(\hat{\psi}_n(y_{n,j},x)) - \psi(y_{n,j},x)| &\leq \bar{F}(y_{n,j}|x)\int_S K(u)\left|\frac{\bar{F}(y_{n,j}|x-hu)}{\bar{F}(y_{n,j}|x)}g(x-hu)-g(x)\right|du \\
&\leq \bar{F}(y_{n,j}|x)\int_S K(u)|g(x-hu)-g(x)|du \qquad (10)\\
&+ \bar{F}(y_{n,j}|x)\int_S K(u)\left|\frac{\bar{F}(y_{n,j}|x-hu)}{\bar{F}(y_{n,j}|x)}-1\right|g(x-hu)du. \qquad (11)
\end{aligned}$$

Under **(L.3)**, and since $g(x) > 0$, we have

$$(10) \leq \bar{F}(y_{n,j}|x)c_g h\int_S d(u,0)K(u)du = \psi(y_{n,j},x)O(h). \qquad (12)$$

Besides, Lemma 1 implies that

$$\sup_{u\in S}\left|\frac{\bar{F}(y_{n,j}|x-hu)}{\bar{F}(y_{n,j}|x)}-1\right| = O(h\log y_{n,j}) = O(h\log y_n)$$

and therefore, in view of (12),

$$\begin{aligned}
(11) &= \bar{F}(y_{n,j}|x)O(h\log y_n)\int_S K(u)g(x-hu)du \\
&= \bar{F}(y_{n,j}|x)g(x)O(h\log y_n)(1+o(1)) \\
&= \psi(y_{n,j},x)O(h\log y_n). \qquad (13)
\end{aligned}$$

Combining (12) and (13) concludes the first part of the proof.

(ii) Let $\beta \neq 0$ in $\mathbb{R}^J$, $\sigma_n(x) = (nh^d\psi(y_n,x))^{-1/2}$, and consider the random variable

$$\begin{aligned}
\Psi_n &= \sum_{j=1}^J \beta_j\left(\frac{\hat{\psi}_n(y_{n,j},x)-\mathbb{E}(\hat{\psi}_n(y_{n,j},x))}{\sigma_n(x)\psi(y_{n,j},x)}\right) \\
&= \sum_{i=1}^n \frac{1}{n\sigma_n(x)}\left\{\sum_{j=1}^J \frac{\beta_j K_h(x-X_i)\mathbb{I}\{Y_i\geq y_{n,j}\}}{\psi(y_{n,j},x)} - \mathbb{E}\left(\sum_{j=1}^J \frac{\beta_j K_h(x-X_i)\mathbb{I}\{Y_i\geq y_{n,j}\}}{\psi(y_{n,j},x)}\right)\right\} \\
&=: \sum_{i=1}^n Z_{i,n}.
\end{aligned}$$

Clearly, $\{Z_{i,n},\ i=1,\dots,n\}$ is a set of centered, independent and identically distributed random variables with variance

$$\mathrm{var}(Z_{i,n}) = \frac{1}{n^2h^{2d}\sigma_n^2(x)}\mathrm{var}\left(\sum_{j=1}^J \beta_j K\left(\frac{x-X_i}{h}\right)\frac{\mathbb{I}\{Y_i\geq y_{n,j}\}}{\psi(y_{n,j},x)}\right) = \frac{1}{n^2h^d\sigma_n^2(x)}\beta^t B\beta,$$

where B is the $J\times J$ covariance matrix with coefficients defined for $(j,j')\in\{1,\dots,J\}^2$ by

$$\begin{aligned}
B_{j,j'} &= \frac{A_{j,j'}}{\psi(y_{n,j},x)\psi(y_{n,j'},x)}, \\
A_{j,j'} &= \frac{1}{h^d}\mathrm{cov}\left(K\left(\frac{x-X}{h}\right)\mathbb{I}\{Y\geq y_{n,j}\},\ K\left(\frac{x-X}{h}\right)\mathbb{I}\{Y\geq y_{n,j'}\}\right) \\
&= \|K\|_2^2\mathbb{E}\left(\frac{1}{h^d}Q\left(\frac{x-X}{h}\right)\mathbb{I}\{Y\geq y_{n,j}\vee y_{n,j'}\}\right) \\
&- h^d\mathbb{E}(K_h(x-X)\mathbb{I}\{Y\geq y_{n,j}\})\mathbb{E}(K_h(x-X)\mathbb{I}\{Y\geq y_{n,j'}\}),
\end{aligned}$$

with $Q(.) =: K^2(.)/\|K\|_2^2$ also satisfying assumption **(K)**. As a consequence, the three above expectations are of the same nature. Thus, remarking that, for n large enough, $y_{n,j} \vee y_{n,j'} = y_{n,j\vee j'}$, part **(i)** of the proof implies

$$A_{j,j'} = \|K\|_2^2 \psi(y_{n,j\vee j'}, x)(1 + O(h \log y_n)) - h^d \psi(y_{n,j}, x)\psi(y_{n,j'}, x)(1 + O(h \log y_n))$$

leading to

$$B_{j,j'} = \frac{\|K\|_2^2}{\psi(y_{n,j\wedge j'}, x)}(1 + O(h \log y_n)) - h^d(1 + O(h \log y_n)) = \frac{\|K\|_2^2}{\psi(y_{n,j\wedge j'}, x)}(1 + o(1)),$$

since $\psi(y_{n,j\wedge j'}, x) \to 0$ as $n \to \infty$. Now, from the regular variation property **(F.1)**, it is easily seen that $\psi(y_{n,j\wedge j'}, x) = a_{j\wedge j'}^{-1/\gamma(x)} \psi(y_n, x)(1 + o(1))$ entailing

$$B_{j,j'} = \frac{\|K\|_2^2 C_{j,j'}(x)}{\psi(y_n, x)}(1 + o(1))$$

and therefore, $\text{var}(Z_{i,n}) \sim \|K\|_2^2 \beta^t C(x)\beta/n$, for all $i = 1, \ldots, n$. As a preliminary conclusion, the variance of Ψ_n converges to $\|K\|_2^2 \beta^t C(x)\beta$. Consequently, Lyapounov criteria for the asymptotic normality of sums of triangular arrays reduces to

$$\sum_{i=1}^{n} \mathbb{E}\,|Z_{i,n}|^3 = n\mathbb{E}\,|Z_{1,n}|^3 \to 0. \tag{14}$$

Remark that $Z_{1,n}$ is a bounded random variable:

$$|Z_{1,n}| \leq \frac{2\|K\|_\infty \sum_{j=1}^{J} |\beta_j|}{n\sigma_n(x)h^d \psi(y_{n,J}, x)} = 2\|K\|_\infty a_J^{1/\gamma(x)} \sum_{j=1}^{J} |\beta_j| \sigma_n(x)(1 + o(1))$$

and thus,

$$\begin{aligned} n\mathbb{E}\,|Z_{1,n}|^3 &\leq 2\|K\|_\infty a_J^{1/\gamma(x)} \sum_{j=1}^{J} |\beta_j| \sigma_n(x) n \text{var}(Z_{1,n})(1 + o(1)) \\ &= 2\|K\|_\infty \|K\|_2^2 a_J^{1/\gamma(x)} \sum_{j=1}^{J} |\beta_j| \beta^t C(x)\beta \sigma_n(x)(1 + o(1)) \to 0 \end{aligned}$$

as $n \to \infty$. As a conclusion, Ψ_n converges in distribution to a centered Gaussian random variable with variance $\|K\|_2^2 \beta^t C(x)\beta$ for all $\beta \neq 0$ in $\mathbb{R}^d$. The result is proved. ■

6.2 Proofs of main results

Proof of Theorem 1. Keeping in mind the notations of Lemma 5, the following expansion holds

$$\sigma_n^{-1}(x) \sum_{j=1}^{J} \beta_j \left(\frac{\hat{\bar{F}}_n(y_{n,j}|x)}{\bar{F}(y_{n,j}|x)} - 1 \right) = \frac{\Delta_{1,n} + \Delta_{2,n} - \Delta_{3,n}}{\hat{g}_n(x)}, \tag{15}$$

where

$$\begin{aligned}
\Delta_{1,n} &= g(x)\sigma_n^{-1}(x)\sum_{j=1}^J \beta_j \left(\frac{\hat\psi_n(y_{n,j},x)-\mathbb{E}(\hat\psi_n(y_{n,j},x))}{\psi(y_{n,j},x)}\right)\\
\Delta_{2,n} &= g(x)\sigma_n^{-1}(x)\sum_{j=1}^J \beta_j \left(\frac{\hat{\mathbb{E}}(\psi_n(y_{n,j},x))-\psi(y_{n,j},x)}{\psi(y_{n,j},x)}\right)\\
\Delta_{3,n} &= \left(\sum_{j=1}^J \beta_j\right)\sigma_n^{-1}(x)\left(\hat g_n(x)-g(x)\right).
\end{aligned}$$

Let us highlight that assumptions $nh^{d+2}\log^2(y_n)\bar F(y_n|x)\to 0$ and $nh^d\bar F(y_n|x)\to\infty$ imply that $h\log y_n\to 0$ as $n\to\infty$. Thus, from Lemma 5(ii), the random term $\Delta_{1,n}$ can be rewritten as

$$\Delta_{1,n} = g(x)\|K\|_2\sqrt{\beta^t C(x)\beta}\,\xi_n, \tag{16}$$

where ξ_n converges to a standard Gaussian random variable. The nonrandom term $\Delta_{2,n}$ is controlled with Lemma 5(i):

$$\Delta_{2,n} = O(\sigma_n^{-1}(x)h\log y_n) = O\left(nh^{d+2}\bar F(y_n|x)\log^2(y_n)\right)^{1/2} = o(1). \tag{17}$$

Finally, $\Delta_{3,n}$ is a classical term in kernel density estimation, which can be bounded by Lemma 4:

$$\Delta_{3,n} = O(h\sigma_n^{-1}(x)) + O_P(\sigma_n^{-1}(x)(nh^d)^{-1/2}) = O\left(nh^{d+2}\bar F(y_n|x)\right)^{1/2} + O_P(\bar F(y_n|x))^{1/2} = o_P(1). \tag{18}$$

Collecting (15)–(18), it follows that

$$\hat g_n(x)\sigma_n^{-1}(x)\sum_{j=1}^J\beta_j\left(\frac{\hat{\bar F}_n(y_{n,j}|x)}{\bar F(y_{n,j}|x)}-1\right) = g(x)\|K\|_2\sqrt{\beta^tC(x)\beta}\,\xi_n + o_P(1).$$

Finally, $\hat g_n(x)\xrightarrow{P} g(x)$ yields

$$\sqrt{nh^d\bar F(y_n|x)}\sum_{j=1}^J\beta_j\left(\frac{\hat{\bar F}_n(y_{n,j}|x)}{\bar F(y_{n,j}|x)}-1\right) = \|K\|_2\sqrt{\frac{\beta^tC(x)\beta}{g(x)}}\,\xi_n + o_P(1)$$

and the result is proved. ■

Proof of Theorem 2. Introduce for $j=1,\dots,J$,

$$\begin{aligned}
\sigma_{n,j}(x) &= q(\alpha_{n,j}|x)(nh^d\alpha_n)^{-1/2}\\
v_{n,j}(x) &= \alpha_{n,j}^{-1}\gamma(x)(nh^d\alpha_n)^{1/2}\\
W_{n,j}(x) &= v_{n,j}(x)\left(\hat{\bar F}_n(q(\alpha_{n,j}|x)+\sigma_{n,j}(x)z_j|x)-\bar F(q(\alpha_{n,j}|x)+\sigma_{n,j}(x)z_j|x)\right)\\
a_{n,j}(x) &= v_{n,j}(x)\left(\alpha_{n,j}-\bar F(q(\alpha_{n,j}|x)+\sigma_{n,j}(x)z_j|x)\right)
\end{aligned}$$

and $z_j \in \mathbb{R}$. We examine the asymptotic behavior of J-variate function defined by

$$\begin{aligned}\Phi_n(z_1,\dots,z_J) &= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\sigma_{n,j}^{-1}(x)(\hat{q}_n(\alpha_{n,j}|x) - q(\alpha_{n,j}|x) \leq z_j\right\}\right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\hat{\bar{F}}_n(q(\alpha_{n,j}|x) + \sigma_{n,j}(x)z_j|x) \leq \alpha_{n,j}\right\}\right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{W_{n,j}(x) \leq a_{n,j}(x)\right\}\right).\end{aligned}$$

Let us first focus on the nonrandom term $a_{n,j}(x)$. From assumptions **(F.1)** and **(F.2)**, Karamata representation (4) shows that $\bar{F}(.|x)$ is differentiable. Thus, for each $j \in \{1,\dots,J\}$ there exists $\theta_{n,j} \in (0,1)$ such that

$$\bar{F}(q(\alpha_{n,j}|x)|x) - \bar{F}(q(\alpha_{n,j}|x) + \sigma_{n,j}(x)z_j|x) = -\sigma_{n,j}(x)z_j\bar{F}'(q_{n,j}|x), \tag{19}$$

where $q_{n,j} = q(\alpha_{n,j}|x) + \theta_{n,j}\sigma_{n,j}(x)z_j$. It is clear that $q(\alpha_{n,j}|x) \to \infty$ and $\sigma_{n,j}(x)/q(\alpha_{n,j}|x) \to 0$ as $n \to \infty$. As a consequence, $q_{n,j} \to \infty$ and thus Karamata representation (4) entails

$$\lim_{n\to\infty} \frac{q_{n,j}\bar{F}'(q_{n,j}|x)}{\bar{F}(q_{n,j}|x)} = -1/\gamma(x). \tag{20}$$

Moreover, since $q_{n,j} \sim q(\alpha_{n,j}|x)$ as $n \to \infty$ and $\bar{F}(.|x)$ is regularly varying, it follows that $\bar{F}(q_{n,j}|x) \sim \bar{F}(q(\alpha_{n,j}|x)|x) = \alpha_{n,j}$. In view of (19) and (20), we end up with

$$a_{n,j}(x) = \frac{v_{n,j}(x)\sigma_{n,j}(x)\alpha_{n,j}z_j}{\gamma(x)q(\alpha_{n,j}|x)}(1+o(1)) = z_j(1+o(1)). \tag{21}$$

Let us now turn to the random term $W_{n,j}(x)$. Defining $a_j = \tau_j^{-\gamma(x)}$, $y_{n,j} = q(\alpha_{n,j}|x) + \sigma_{n,j}(x)z_j$ for $j = 1,\dots,J$ and $y_n = q(\alpha_n|x)$, we have $y_{n,j} \sim q(\alpha_{n,j}|x) \sim a_j y_n$ since $q(.|x)$ is regularly varying a 0 with index $-\gamma(x)$. Using the same argument, it it easily shown that $\log y_n \sim -\gamma(x)\log\alpha_n$. As a consequence, Theorem 1 applies and the random vector

$$\left\{\frac{\sqrt{nh^d\bar{F}(y_n|x)}}{v_{n,j}(x)\bar{F}(y_{n,j}|x)}W_{n,j}\right\}_{j=1,\dots,J} = (1+o(1))\left\{\frac{W_{n,j}}{\gamma(x)}\right\}_{j=1,\dots,J}$$

converges to a centered Gaussian random variable with covariance matrix $\frac{\|K\|_2^2}{g(x)}C(x)$. Taking account of (21), we obtain that $\Phi_n(z_1,\dots,z_J)$ converges to the cumulative distribution function of a centered Gaussian distribution with covariance matrix $\frac{\|K\|_2^2\gamma^2(x)}{g(x)}C(x)$ evaluated at $(z_1,\dots,z_J)$, which is the desired result. ■

Proof of Corollary 1. Introducing $\alpha_{n,j} = \tau_j\alpha_n$ for all $j \in \{1,2,3\}$ with $\tau_1 = 4$, $\tau_2 = 2$, $\tau_3 = 1$ and $\alpha_n = k/n$, Theorem 2 shows that, for $j \in \{1,2,3\}$,

$$\frac{\hat{q}_n(\alpha_{n,j}|x)}{q(\alpha_{n,j}|x)} = 1 + \sigma_n\xi_{n,j} \tag{22}$$

where $(\xi_{n,1}, \xi_{n,2}, \xi_{n,3})^t$ converges to a centered Gaussian random vector with covariance matrix

$$|K\|_2^2\gamma^2(x)/g(x)\begin{bmatrix}1/4 & 1/4 & 1/4\\ 1/4 & 1/2 & 1/2\\ 1/4 & 1/2 & 1\end{bmatrix}.$$

Replacing (22) in $\hat{\gamma}_n(x)$ yields

$$\begin{aligned}(\log 2)\hat{\gamma}_n(x) &= \log\left(\frac{q(\alpha_{n,3}|x)}{q(\alpha_{n,2}|x)}(1+\sigma_n\xi_{n,3}) - 1 - \sigma_n\xi_{n,2}\right)\\ &- \log\left(1+\sigma_n\xi_{n,2} - \frac{q(\alpha_{n,1}|x)}{q(\alpha_{n,2}|x)}(1+\sigma_n\xi_{n,1})\right)\\ &= \log\left(2^{\gamma(x)}(1+O(\varepsilon(q(\alpha_{n,2}|x)|x)))(1+\sigma_n\xi_{n,3}) - 1 - \sigma_n\xi_{n,2}\right)\\ &- \log\left(1+\sigma_n\xi_{n,2} - 2^{-\gamma(x)}(1+O(\varepsilon(q(\alpha_{n,2}|x)|x)))(1+\sigma_n\xi_{n,1})\right),\end{aligned}$$

in view of Lemma 2. As a consequence of assumption $\varepsilon(q(\alpha_{n,2}|x)|x))/\sigma_n \to 0$, we obtain

$$\begin{aligned}(\log 2)\hat{\gamma}_n(x) &= \log\left(2^{\gamma(x)} - 1 + \sigma_n(2^{\gamma(x)}\xi_{n,3} - \xi_{n,2} + o_P(1))\right)\\ &- \log\left(1 - 2^{-\gamma(x)} + \sigma_n(\xi_{n,2} - 2^{-\gamma(x)}\xi_{n,1} + o_P(1))\right).\end{aligned}$$

Standard calculations lead to

$$\begin{aligned}\sigma_n^{-1}(\log 2)(\hat{\gamma}_n(x) - \gamma(x)) &= \sigma_n^{-1}\log\left(1 + \frac{\sigma_n}{2^{\gamma(x)}-1}(2^{\gamma(x)}\xi_{n,3} - \xi_{n,2} + o_P(1))\right)\\ &- \sigma_n^{-1}\log\left(1 + \frac{\sigma_n}{1-2^{-\gamma(x)}}(\xi_{n,2} - 2^{-\gamma(x)}\xi_{n,1} + o_P(1))\right)\\ &= \frac{\xi_{n,1} - (1+2^{\gamma(x)})\xi_{n,2} + 2^{\gamma(x)}\xi_{n,3}}{2^{\gamma(x)}-1} + o_P(1),\end{aligned}$$

which converges to a centered Gaussian random variable with variance $\frac{\|K\|_2^2\gamma^2(x)(2^{2\gamma(x)+1}+1)^2}{4(2^{\gamma(x)}-1)^2 g(x)}$, and the conclusion follows. ■

Acknowledgments

The authors are grateful to Sylvain Douté for his important contribution to the hyperspectral data and for fruitful discussions.

References

[1] M.I.F. Alves, M.I. Gomes, M.I. and L. de Haan. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60:193–214, 2003.

[2] M.I.F. Alves, L. de Haan and T. Lin. Estimation of the parameter controlling the speed of convergence in extreme value theory, *Mathematical Methods of Statistics*, 12:155–176, 2003.

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

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

[5] C. Bernard-Michel, S. Dout, M. Fauvel, L. Gardes and S. Girard. Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression, *Journal of Geophysical Research - Planets*, to appear, 2009.

[6] C. Bernard-Michel, L. Gardes and S. Girard. Gaussian Regularized Sliced Inverse Regression, *Statistics and Computing*, 19:85–98, 2009.

[7] N.H. Bingham, C.M. Goldie and J.L. Teugels. *Regular Variation*, Cambridge University Press, 1987

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

[9] G. Collomb. *Estimation non paramétrique de la régression par la méthode du noyau.* PhD thesis, Université Paul Sabatier de Toulouse, 1976.

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

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

[12] A. Dekkers and L. de Haan. On the estimation of the extreme-value index and large quantile estimation, *Annals of Statistics*, 17:1795–1832, 1989.

[13] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, Springer, 1997.

[14] A. Gannoun. Estimation non paramétrique de la médiane conditionnelle, médianogramme et méthode du noyau, *Publications de l'Institut de Statistique de l'Université de Paris*, XXXXVI:11–22, 1990.

[15] A. Gannoun, S. Girard, C. Guinot and J. Saracco. Reference ranges based on nonparametric quantile regression, *Statistics in Medicine*, 21:3119-3135, 2002.

[16] L. Gardes. Estimating the support of a Poisson process via the Faber-Shauder basis and extreme values. *Publications de l'Institut de Statistique de l'Université de Paris*, XXXXVI:43–72, 2002.

[17] J. Geffroy. Sur un problème d'estimation géométrique. *Publications de l'Institut de Statistique de l'Université de Paris*, XIII:191–210, 1964.

[18] S. Girard and P. Jacob. Extreme values and Haar series estimates of point process boundaries. *Scandinavian Journal of Statistics*, 30(2):369–384, 2003.

[19] S. Girard and P. Jacob. Frontier estimation via kernel regression on high poextreme level curvewer-transformed data. *Journal of Multivariate Analysis*, 99:403–420, 2008.

[20] S. Girard and L. Menneteau. Central limit theorems for smoothed extreme value estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, 135(2):433–460, 2005.

[21] M.I. Gomes and M.J. Martins and M. Neves. Semi-parametric estimation of the second order parameter, asymptotic and finite sample behaviour. *Extremes*, 3:207–229, 2000.

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

[23] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering, Springer, 2006.

[24] P. Hall, M. Nussbaum, and S. Stern. On the estimation of a support curve of indeterminate sharpness. *Journal of Multivariate Analysis*, 62(2):204–232, 1997.

[25] W. Härdle, B. U. Park, and A. B. Tsybakov. Estimation of a non sharp support boundaries. *Journal of Multivariate Analysis*, 43:205–218, 1995.

[26] A.P. Korostelev and A.B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New-York, 1993.

[27] L. Menneteau. Multidimensional limit theorems for smoothed extreme value estimates of point processes boundaries. *ESAIM: Probability and Statistics*, 12:273–307, 2008.

[28] J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

[29] G.G. Roussas. Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.*, 40:1386–1400, 1969.

[30] T. Samanta. Non-parametric estimation of conditional quantiles. *Statistics and Probability Letters*, 7:407–412, 1989.

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

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

[33] W. Stute. Conditional empirical processes. *The Annals of Statistics*, 14:638–647, 1986.

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

[35] Q. Yao. Conditional predictive regions for stochastic processes. *Technical report*, University of Kent at Canterbury, 1999.

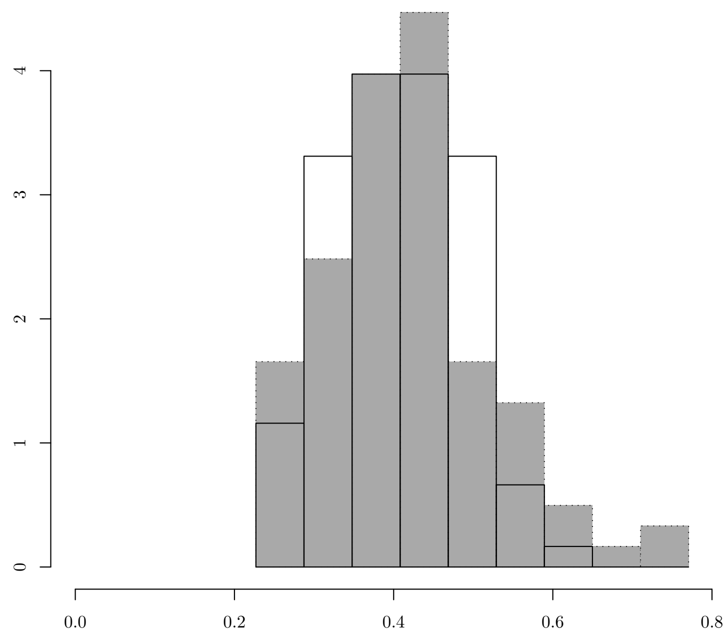

Figure 1: Comparison between the error distributions obtained with the cross-validation strategy ($\Delta_{cv}^{(p)}$, light gray) and the oracle strategy ($\Delta_{oracle}^{(p)}$, transparent) on $N = 100$ samples of size $n = 300$.

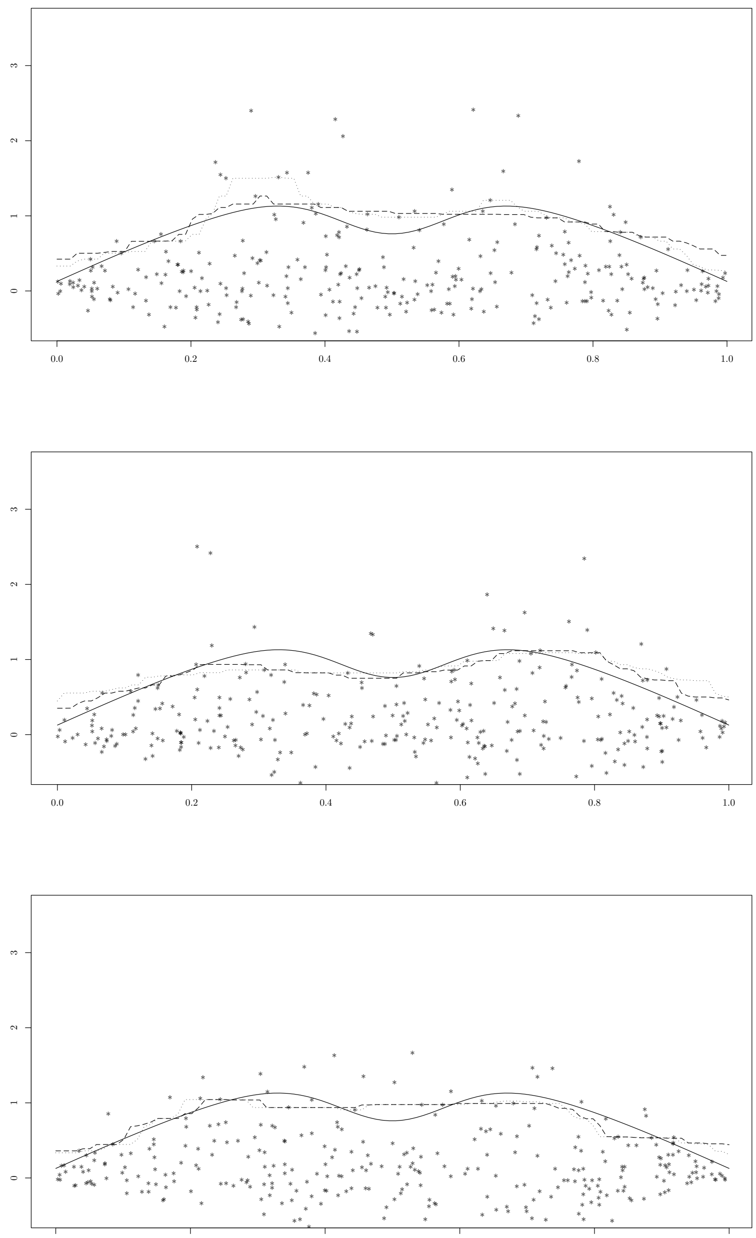

Figure 2: Comparison of the true quantile (solid line) with the estimated ones obtained by the cross-validation strategy (dotted line) and the oracle strategy (dashed line). The sample size is $n = 300$. The vertical axis is in a logarithmic scale. Top: worst 10% estimator, middle: median estimator, bottom: best 10% estimator.

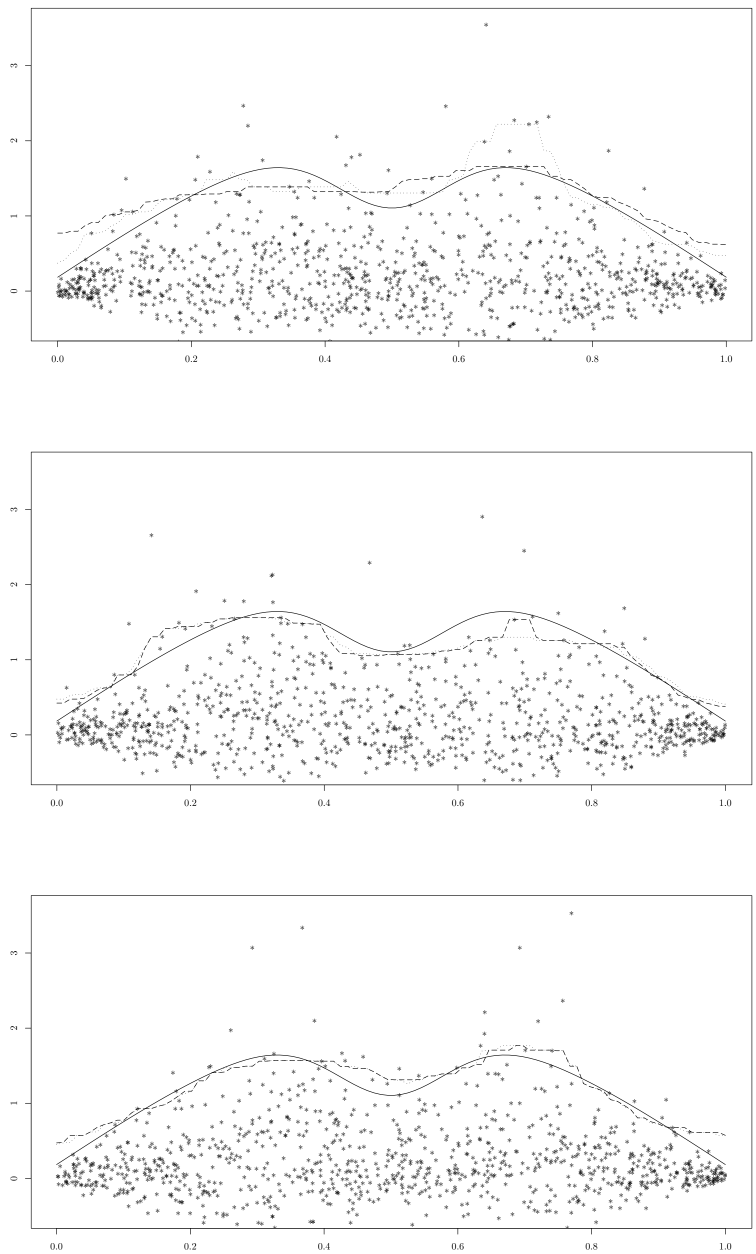

Figure 3: Comparison of the true quantile (solid line) with the ones quantiles obtained by the cross-validation strategy (dotted line) and the oracle strategy (dashed line). The sample size is $n = 1000$. The vertical axis is in a logarithmic scale. Top: worst 10% estimator, middle: median estimator, bottom: best 10% estimator.

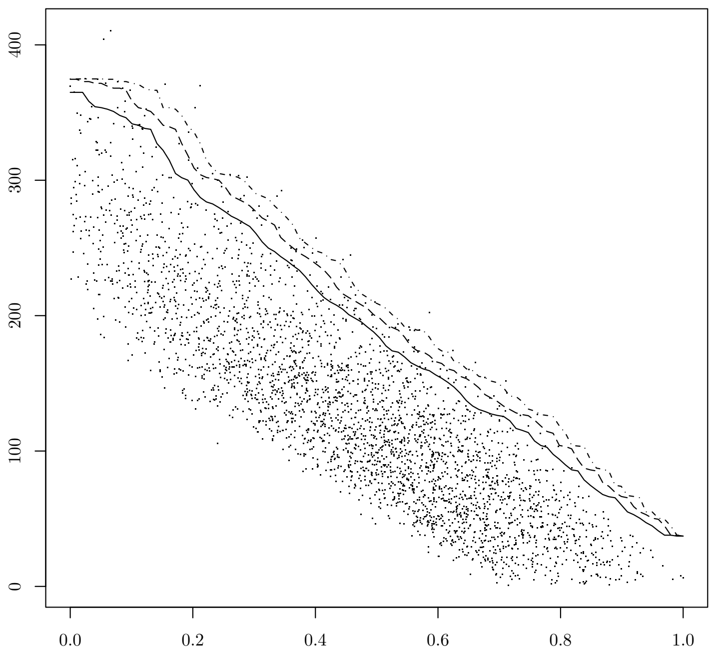

Figure 4: Real data scatter-plot (X_i, Y_i), $i = 1, \ldots, n$ and estimated extreme level curves ($\zeta = 20$: solid line, $\zeta = 10$: dashed line, $\zeta = 5$: dash-dotted line)

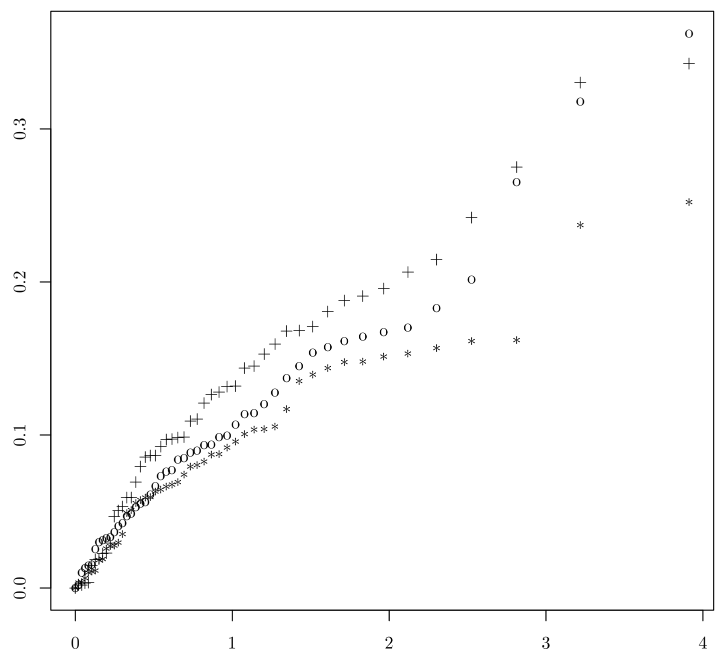

Figure 5: QQ-plots obtained at three different points: $x = 0.25$ ($\star\star\star$), $x = 0.50$ ($\circ\circ\circ$) and $x = 0.75$ ($+++$).