



**HAL**  
open science

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

Laurent Gardes, Armelle Guillou, Claire Roman

► **To cite this version:**

Laurent Gardes, Armelle Guillou, Claire Roman. Estimation of extreme conditional quantiles under a general tail first order condition. *Annals of the Institute of Statistical Mathematics*, 2020, 72, pp.915-943. 10.1007/s10463-019-00713-7. hal-01818521

**HAL Id: hal-01818521**

**<https://hal.science/hal-01818521v1>**

Submitted on 19 Jun 2018

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

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

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

Laurent Gardes, Armelle Guillou and Claire Roman

Université de Strasbourg & CNRS, IRMA, UMR 7501,  
7 rue René Descartes, 67084 Strasbourg Cedex, France

**Abstract.** We consider the estimation of an extreme conditional quantile. In a first part, we propose a new tail condition in order to establish the asymptotic distribution of an extreme conditional quantile estimator. Next, a general class of estimators is introduced, which encompasses, among others, kernel's or nearest neighbors' types of estimators. A unified theorem of the asymptotic normality for this general class of estimators is provided under the new tail condition and illustrated on the different well-known examples. A comparison between different estimators belonging to this class is provided on a small simulation study.

**Keywords:** Extreme quantile, local estimation, asymptotic normality.

## 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  $(X_1, Y_1), \dots, (X_n, Y_n)$  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  $\hat{S}_n(\cdot|x_0)$  of  $S(\cdot|x_0)$ , the associated estimator of  $Q(\alpha|x_0)$  is given by

$$\hat{Q}_n(\alpha|x_0) := \hat{S}_n^{\leftarrow}(\alpha|x_0) = \inf\{y \in \mathbb{R}; \hat{S}_n(y|x_0) \leq \alpha\}. \quad (1)$$

Estimator (1) is called indirect since, as pointed by Racine and Li (2017), “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

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

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 Bassett (1978), Koenker et al. (1994) and He and Ng (1999). In this paper, 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  $\alpha_n$  where  $\alpha_n \rightarrow 0$  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  $\alpha$  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  $\alpha$  and of an extreme level  $\alpha_n$  is outlined in Section 2. In Section 3, this condition is specified and illustrated on many well-known examples of conditional distributions. Section 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 5 where several examples of estimators belonging to this class, among them, the kernel’s and nearest neighbors’ type-estimators, are compared. All the proofs of the results are postponed to Section 6.

## 2 Description of the methodology

The aim of this paper 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  $\alpha$ . 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  $\alpha$  is fixed, the asymptotic distribution of (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  $v_n(x_0) \rightarrow \infty$  such that for all sequence  $\varepsilon_n \rightarrow 0$

$$v_n(x_0) \left( \widehat{S}_n(y + \varepsilon_n|x_0) - S(y + \varepsilon_n|x_0) \right) \xrightarrow{d} \Lambda, \quad (2)$$

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

$$v_n(x_0) \left( \widehat{Q}_n(\alpha|x_0) - Q(\alpha|x_0) \right) \xrightarrow{d} \frac{1}{f(Q(\alpha|x_0)|x_0)} \Lambda, \quad (3)$$

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 (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

$$\mathbb{P} \left[ \sigma_n(x_0) \left( \widehat{Q}_n(\alpha|x_0) - Q(\alpha|x_0) \right) \leq z \right] = \mathbb{P}[Z_n(x_0) \leq z_n(x_0)], \quad (4)$$

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) with  $y = Q(\alpha|x_0)$ ,  $Z_n(x_0) \xrightarrow{d} \Lambda$  and since  $S(\cdot|x_0)$  is continuously differentiable,  $z_n(x_0) \rightarrow z$  as  $n \rightarrow \infty$  proving (3). Note that the asymptotic distribution of indirect estimators for a fixed level  $\alpha$  has been treated for instance by Berline et al. (2001).

**Case of an extreme level** – We consider the situation where the level of the conditional quantile is a sequence  $\alpha_n$  where  $\alpha_n \rightarrow 0$  as the sample size  $n$  increases. Replacing the level  $\alpha$  by a sequence  $\alpha_n$  does not change (at least if  $\alpha_n$  does not converge too fast to 0) the estimation procedure. We still estimate  $Q(\alpha_n|x_0)$  as in (1) just by replacing  $\alpha$  by  $\alpha_n$ . The difference lies in the assumptions required to obtain the asymptotic distribution of  $\widehat{Q}_n(\alpha_n|x_0)$ . First, instead of (2), the following kind of result for the conditional survival function estimator is required: for some well-chosen sequence  $y_n(x_0) \rightarrow y^*(x_0) := Q(0|x_0)$ , there exists a sequence  $v_n(x_0) \rightarrow \infty$  such that

$$v_n(x_0) \left( \widehat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0) \right) \xrightarrow{d} \Lambda, \quad (5)$$

for some non-degenerate distribution  $\Lambda$ . 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 (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 (4) where  $\alpha$  is replaced by  $\alpha_n$ . 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  $\alpha$  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 (5) with  $y_n(x_0) = Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)$

$$z_n(x_0) = -\alpha_n v_n(x_0) \left[ \frac{S[y_n(x_0)|x_0]}{S[Q(\alpha_n|x_0)|x_0]} - 1 \right] \rightarrow z, \quad (6)$$

as  $n \rightarrow \infty$  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  $Q(\alpha_n|x_0) \rightarrow y^*(x_0)$ , a natural general condition leading to (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  $t \in I$ ,

$$\lim_{y \uparrow y^*(x_0)} \Psi(y) \left( \frac{S[y + td(y)|x_0]}{S(y|x_0)} - 1 \right) \rightarrow \phi_{x_0}^{-1}(t), \quad (7)$$

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  $\phi(t)/t \rightarrow -1$  as  $t \rightarrow 0$ .

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 (7) and assuming that  $t_n(x_0) \rightarrow 0$ , we can show that  $z_n(x_0) \rightarrow z$  (see Section 3, Proposition 1). Next, the random sequence  $Z_n(x_0)$  is treated by (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 (7). Condition (7) is referred in what follows to as the Tail First Order condition. Under this condition and if (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 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.

### 3 A 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 1** *A survival function  $S$  satisfies a Tail First Order (TFO) condition if for some open interval  $I \subset \mathbb{R}$  containing 0, there exist positive functions  $d$  and  $\Psi$  such that for all  $t \in I$ ,*

$$\lim_{x \uparrow x^*} \Psi(x) \left( \frac{S[x + td(x)]}{S(x)} - 1 \right) = \phi^{-1}(t), \quad (8)$$

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

Note that convergence (8) entails that for all  $t \in I$  and for  $x$  large enough,  $x + td(x) < x^*$ . Consequently, the function  $\Psi$  is such that  $\Psi(x)/S(x) \rightarrow \infty$  as  $x \uparrow x^*$ . Finally, it is easy to check that  $\phi^{-1}(t)/t \rightarrow -1$  as  $t \rightarrow 0$ . As a consequence of Dini's theorem, we obtain the useful properties gathered in the next proposition.

**Proposition 1** *If  $S$  satisfies a TFO condition, the following statements are true:*

1. *Convergence in (8) holds locally uniformly on  $I$ .*
2. *For all  $t_0 \in I$ ,*

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

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

**Proposition 2** *The following statements are equivalent:*

1. *The survival function  $S$  satisfies a TFO condition.*
2. *There exist positive functions  $a$  and  $g$  such that for all  $t \in J$ ,*

$$\lim_{\alpha \rightarrow 0} \frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} = \phi(t). \quad (9)$$

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

$$\lim_{n \rightarrow \infty} [nc_n S(a_n t + b_n) - c_n] = \phi^{-1}(t). \quad (10)$$

**Remarks** – 1) The relations between the auxiliary functions involved in (8) and (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 (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  $g(\alpha) \rightarrow 0$  as  $\alpha \rightarrow 0$ .

3) An interpretation of condition (9) is based on the following remark: from the second statement of Proposition 1,

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

as  $(t, \alpha) \rightarrow 0$ . 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 3** *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  $t \in J$ ,*

$$\lim_{\alpha \rightarrow 0} \frac{Q'[\alpha + tg(\alpha)]}{Q'(\alpha)} = \Theta(t). \quad (11)$$

*If for all  $t \in J$ ,  $\int_0^t \Theta(s) ds =: \theta(t) \in \mathbb{R}$  where  $\theta$  is an increasing function on  $J$  such that  $\theta(t)/t \rightarrow 1$  as  $t \rightarrow 0$  then condition (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 a 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

$$\lim_{n \rightarrow \infty} [1 - S(a_n x + b_n)]^n = G(x), \quad (12)$$

for all point of continuity of  $G$ . Fisher and Tippett (1928) and Gnedenko (1943) 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 (12) is said to belong to the maximum domain of attraction of the extreme value distribution  $G_\gamma$ . The parameter  $\gamma \in \mathbb{R}$  is called the extreme value index. As established in de Haan and Ferreira (2006, Theorem 1.1.6), condition (12) is equivalent to assume the existence of a positive auxiliary function  $a$  and a non constant function  $\phi$  for which

$$\lim_{\alpha \rightarrow 0} \frac{Q(t\alpha) - Q(\alpha)}{a(\alpha^{-1})} = \phi(t). \quad (13)$$

From de Haan and Ferreira (2006, Theorem B.2.1), the function  $\phi$  in (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.

The aim of the next result is to show that the TFO condition introduced in this paper (see Definition 1) is weaker than (12).

**Proposition 4** *If  $S$  satisfies a TFO condition with an auxiliary function  $g$  in (9) such that  $\alpha/g(\alpha) \rightarrow c \geq 0$  as  $\alpha \rightarrow 0$  (with  $g$  continuous and strictly increasing if  $c = 0$ ) then  $S$  satisfies (12).*

As a consequence of this result, if a survival function  $S$  satisfies the TFO condition with a function  $g$  as in Proposition 4, 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 (11) is equivalent to assume that

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

for some  $\gamma \in \mathbb{R}$ . This condition coincides with condition (1.1.33) in de Haan and Ferreira (2006, 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 (9) holds with a function  $g$  such that  $\alpha/g(\alpha) \rightarrow \infty$ . 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  $t > 0$ ,

$$\lim_{x \rightarrow \infty} \frac{S(tx)}{S(x)} = 1. \quad (14)$$

It can be shown that these survival functions do not satisfy the classical first order condition (12). Note that a heavy-tailed distribution corresponds to a survival function satisfying for all  $t > 0$ ,  $S(tx)/S(x) \rightarrow t^{-1/\gamma}$  as  $x \rightarrow \infty$ , for some  $\gamma > 0$ . Hence, roughly speaking, a super heavy-tailed distribution is a heavy-tailed distribution with  $\gamma = +\infty$ .

Unfortunately, condition (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  $s > -1$ ,

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

Since super heavy-tailed distribution can be seen as a heavy-tailed distribution with  $\gamma = +\infty$ , we propose to replace in the previous limit  $\gamma$  by  $\gamma(\alpha)$  where  $\gamma(\alpha) \rightarrow \infty$  as  $\alpha \rightarrow 0$ . Since

$$\lim_{\alpha \rightarrow 0} (1+s)^{-\gamma(\alpha)} = \begin{cases} +\infty & \text{if } s \in (-1, 0), \\ 0 & \text{if } s > 0, \end{cases}$$

we must replace  $s$  by  $t/\gamma(\alpha)$  with  $t \in \mathbb{R}$  to obtain a non-degenerate limit:

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

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  $g(\alpha)/\alpha \rightarrow 0$  as  $\alpha \rightarrow 0$  such that for all  $t \in \mathbb{R}$

$$\lim_{\alpha \rightarrow 0} \frac{Q[\alpha + tg(\alpha)]}{Q(\alpha)} = e^{-t}. \quad (15)$$

It appears that convergence (15) coincides with the TFO condition with  $a(\alpha^{-1}) = Q(\alpha)$  and  $\phi(t) = e^{-t} - 1$ . As shown in Proposition 5 below, this definition is equivalent to the one introduced for instance in Fraga Alves et al. (2009) 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

$$\lim_{x \rightarrow \infty} \frac{U[x + tb(x)]}{U(x)} = e^t \quad (16)$$

with  $U(\cdot) := Q(1/\cdot)$ . Note that according to Fraga Alves et al. (2009, Lemma 4.1), condition (16) implies (14). Furthermore, the function  $b$  is such that  $b(x)/x \rightarrow 0$  as  $x \rightarrow \infty$ . Remark finally that the right endpoint of a distribution satisfying (16) is necessarily infinite. As examples of super heavy-tailed distribution satisfying (16), one can cite the standard log-Pareto distribution given by  $S(x) = [\log(x)]^{-\xi}$  with  $\xi > 0$  and the log-Weibull distribution for which  $S(x) = \exp(-\xi \log^\theta x)$ , with  $\xi > 0$  and  $\theta \in (0, 1)$ . For these two distributions, one can take  $b \sim U/U'$ .

**Proposition 5** *Conditions (15) and (16) are equivalent. The relation between the involved functions is  $b(x) = x^2 g(x^{-1})$ .*



## 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  $(X_1, Y_1), \dots, (X_n, Y_n)$  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 a TFO condition, the proposed estimators computed with an extreme level  $\alpha_n \rightarrow 0$  are asymptotically Gaussian.

### 4.1 A class of conditional quantile estimators

As mentioned in the introduction we focus in this paper 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

$$\widehat{S}_n(y|x_0) := \sum_{i=1}^n \mathcal{W}_{n,i}(x_0) \mathbb{I}_{(y, \infty)}(Y_i). \quad (17)$$

The set of weights  $\{\mathcal{W}_{n,i}(x_0), 1 \leq i \leq n\}$  is a triangular array of positive random variables depending on the data  $X_1, \dots, X_n$  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 (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 (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 (17) are called *weighted local* estimators.

The kernel based estimator introduced by Nadaraya (1964) and Watson (1964) is a classical example of weighted local estimator. This estimator is obtained by using the following random weights in (17):

$$\mathcal{W}_{n,i}^{\text{NW}}(x_0, h_n) := K\left(\frac{X_i - x_0}{h_n}\right) / \sum_{j=1}^n K\left(\frac{X_j - x_0}{h_n}\right), \quad (18)$$

where  $K$  is a density on  $\mathbb{R}^p$  and  $h_n$  is a positive non-random sequence satisfying  $h_n \rightarrow 0$  as  $n \rightarrow \infty$ . 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  $\{D_i(x_0) := \|X_i - x_0\|, i = 1, \dots, n\}$  be the distances between each observation and  $x_0$  and let  $D_{(1)}(x_0) \leq \dots \leq D_{(n)}(x_0)$  the corresponding order statistics. Denoting by  $\{r(i), i = 1, \dots, n\}$  the ranks of these distances (i.e.,  $D_{(i)}(x_0) = D_{r(i)}(x_0)$  for  $i = 1, \dots, n$ ), the  $k_n$ -NN estimator is obtained by using the following random weights in (17):

$$\mathcal{W}_{n,i}^{\text{NN}}(x_0, k_n) := [(k_n - r(i) + 1)_+]^\ell \bigg/ \sum_{j=1}^{k_n} j^\ell, \quad (19)$$

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 of (18) and (19). For instance, we can consider the random weights defined for  $\tau \in (0, 1)$  by

$$\mathcal{W}_{n,i}^{\text{LC}}(x_0, \tau, h_n, k_n) := \frac{\tau}{M_n} \mathbb{I}_{[0,1]} \left( \left\| \frac{X_i - x_0}{h_n} \right\|_\infty \right) + \frac{1 - \tau}{k_n} \mathbb{I}_{[0,1]} \left( \frac{r(i)}{k_n} \right), \quad (20)$$

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

## 4.2 Main results

Under general conditions on the random weights  $\{\mathcal{W}_{n,i}(x_0), i = 1, \dots, n\}$ , we want to establish the convergence in distribution of a normalized version of  $\hat{Q}_n(\alpha_n|x_0)$  for a level  $\alpha_n$  converging to 0 as  $n \rightarrow \infty$ . As outlined in Section 2, we first need to find a sequence  $v_n(x_0) \rightarrow \infty$  and a non-degenerate distribution  $\Lambda$  such that (under some additional assumptions)

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

for some sequence  $y_n(x_0) \uparrow y^*(x_0)$ . This is done in Theorem 1 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's 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 1** 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  $\{X = x_0\}$ . 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

$$v_n(x_0) \max_{1 \leq i \leq n} \mathcal{W}_{n,i}(x_0) \xrightarrow{a.s.} 0 \quad (21)$$

and

$$v_n(x_0) \sum_{i=1}^n \mathcal{W}_{n,i}(x_0) |S(y_n(x_0)|X_i) - S(y_n(x_0)|x_0)| \xrightarrow{\mathbb{P}} 0, \quad (22)$$

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

To understand the usefulness of conditions (21) and (22), we provide below the main ideas of the proof of Theorem 1, the complete proof being postponed to Section 6. 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

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

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:

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

see Section 6, Proposition 7. This is obtained mainly by applying the Lindeberg's theorem and only condition (21) is required. This condition is in fact equivalent to the Lindeberg's condition.

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

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

converges to 0 in probability (see Section 6, Proposition 8). 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

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

To control the bias term we need to measure the discrepancy between the two probability measures  $\mathcal{W}_{n,x_0}$  and  $\delta_{x_0}$ . 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 (22) can in fact be written in term of the Wasserstein distance as follows

$$v_n(x_0) W_1(\mathcal{W}_{n,x_0}^*, \delta_{x_0}^*) \xrightarrow{\mathbb{P}} 0, \quad (23)$$

where  $\mathcal{W}_{n,x_0}^*$  and  $\delta_{x_0}^*$  are the pushforward measures of  $\mathcal{W}_{n,x_0}$  and  $\delta_{x_0}$  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  $\alpha_n$  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 a TFO condition with positive auxiliary functions  $\Psi_{x_0} \equiv \Psi$  and  $d_{x_0} \equiv d$ .

Let  $a(1/\cdot) \equiv a_{x_0}(1/\cdot) = d[Q(\cdot|x_0)]$  and  $g(\cdot) \equiv g_{x_0}(\cdot) = \cdot/\Psi[Q(\cdot|x_0)]$ . From Proposition 2, condition **(H)** is equivalent to assume that for some open interval  $J_{x_0} = J \subset \mathbb{R}$  containing 0, one has for all  $t \in J$

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

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

**Theorem 2** *Let  $x_0 \in \mathbb{R}^p$  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  $\alpha_n m_n(x_0) \rightarrow \infty$ ,  $v_n(x_0)g(\alpha_n) \rightarrow \infty$ ,*

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

and

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

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

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

Recall that if  $g(\alpha) = \alpha$  (or equivalently  $\Psi \equiv 1$ ), condition **(H)** coincides with the classical first order condition (13) used in extreme value theory. In this case,  $\phi(t) \propto (t^{-\gamma(x_0)} - 1)/\gamma(x_0)$  where the function  $\gamma$  is referred to as the conditional extreme value index. Under (13) and if the conditions of Theorem 2 are satisfied,

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

Moreover, we know from de Haan and Ferreira (2006, Lemma 1.2.9) that under (13),  $Q(\alpha_n|x_0)/a(\alpha_n^{-1}) \rightarrow 1/\gamma_+(x_0)$ , where  $\gamma_+(x_0) = \max(\gamma(x_0), 0)$ . So, under the first order condition (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 2 for conditional super heavy-tailed distribution. Taking the definition of super heavy-tailed distributions given in Fraga Alves et al. (2009) into account, we have in this case  $a(\alpha^{-1}) = Q(\alpha|x_0)$  and  $g(\alpha)/\alpha \rightarrow 0$  as  $\alpha \rightarrow 0$ . 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 2 is proved under general conditions on the random weights used to define the conditional survival estimator (17). We close this section by applying Theorem 2 to particular weights.

**Nadaraya-Watson's weights** – Taking the weights defined in (18) leads to the well-known Nadaraya-Watson's estimator of the conditional survival function:

$$\widehat{S}_n^{\text{NW}}(y|x_0) := \sum_{i=1}^n K\left(\frac{X_i - x_0}{h_n}\right) \mathbb{I}_{(y, \infty)}(Y_i) \Big/ \sum_{i=1}^n K\left(\frac{X_i - x_0}{h_n}\right). \quad (24)$$

The corresponding conditional quantile estimator is denoted by  $\widehat{Q}_n^{\text{NW}}(\alpha_n|x_0)$ . In order to apply Theorem 2, we need to check that the Nadaraya-Watson's 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 (24). Recall that  $f$  is the pdf of  $X$ .

**Corollary 1** *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  $h_n \rightarrow 0$  and  $\alpha_n \in (0, 1)$  such that  $nh_n^p[\alpha_n \wedge (\log \log n)^{-1}] \rightarrow \infty$ ,  $\alpha_n^{-1}nh_n^pg^2(\alpha_n) \rightarrow \infty$  and*

$$\sup_{\substack{|\beta/\alpha_n - 1| \leq \xi \\ \|x - x_0\| \leq h_n}} \left| \frac{S[Q(\beta|x_0)|x]}{\beta} - 1 \right|^2 = o\left(\frac{1}{nh_n^p\alpha_n}\right) \quad (25)$$

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

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

Note that under the classical first order condition (13) (i.e., when  $g(\alpha_n) = \alpha_n$  in **(H)**), the asymptotic normality of the Nadaraya-Watson conditional quantile estimator has already been obtained in Daouia et al. (2013, Corollary 1). This last result also requires the use of condition (25) which controls the oscillations of the function  $Q(\alpha_n|\cdot)$ . Of course, the proof of Daouia et al. (2013, Corollary 1) uses arguments adapted to the Nadaraya-Watson's estimator while Theorem 2 can be used for a large range of weighted conditional survival estimators. As a consequence, conditions on  $h_n$  and  $\alpha_n$  involved in Daouia et al. (2013, Corollary 1) and in our Corollary 1 are slightly different. More precisely, the conditions in Daouia et al. (2013, Corollary 1) are  $nh_n^p\alpha_n \rightarrow \infty$  and  $nh_n^{p+2}\alpha_n \rightarrow 0$  while in our Corollary 1 it is required that  $nh_n^p\alpha_n \rightarrow \infty$  and  $nh_n^p(\log \log n)^{-1} \rightarrow \infty$ . Hence, if  $\alpha_n \log \log n \rightarrow 0$  as  $n \rightarrow \infty$  (i.e., for large quantiles), conditions on the sequences  $h_n$  and  $\alpha_n$  are weaker in Corollary 1 than in Daouia et al. (2013, Corollary 1).

**Nearest Neighbors' approach** – Now, let us consider the  $k_n$ -NN random weights defined in (19) and leading to the conditional survival function estimator

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

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  $\widehat{Q}_n^{\text{KNN}}(\alpha_n|x_0)$  is established in the following result.

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

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

The asymptotic variance  $(\ell+1)^2/(2\ell+1)$  is an increasing function of  $\ell$  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  $\widehat{Q}_n^{\text{LC}}(\alpha_n|x_0)$  of  $Q(\alpha_n|x_0)$  obtained by using the linear combination of weights introduced in (20).

**Corollary 3** *Let  $x_0 \in \mathbb{R}^p$  such that  $f$  is continuous at  $x_0$  and  $f(x_0) > 0$ . Let  $h_n \rightarrow 0$ ,  $k_n \rightarrow \infty$  and  $\alpha_n$  be sequences such that  $nh_n^p/\log \log n \rightarrow \infty$ ,  $\ell_n \alpha_n \rightarrow \infty$  with  $\ell_n := (nh_n^p \wedge k_n)$ ,  $\alpha_n^{-1} \ell_n g^2(\alpha_n) \rightarrow \infty$  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  $k_n/(nh_n^p) \rightarrow \kappa$ , we have

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

In practice, one can take  $k_n = \lfloor \kappa n h_n^p \rfloor$  with  $\kappa > 0$ . The parameter  $\kappa$  is thus a tuning parameter that has to be chosen by a data-driven procedure (see Section 5.1).

## 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 (1) for a given value of  $x_0$ . The random weights  $\{\mathcal{W}_{n,1}(x_0), \dots, \mathcal{W}_{n,n}(x_0)\}$  used in the expression of the estimator (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 Nadaraya-Watson's weights, the  $k_n$ -NN random weights or the LC-weights defined in (18), (19) and (20), where  $\lambda_n$  is equal to  $h_n$ ,  $k_n$  and  $(h_n, \kappa)$ , respectively. In the next section, we propose an adaptive procedure to select  $\lambda_n$  in practice.

### 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  $\lambda_n$  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  $\lambda_n$ , the random value  $S[\widehat{Q}_{n,-1}(\alpha_n|X_1, \lambda_n)|X_1]$  should be close to  $\alpha_n$  at least when the observed value of  $X_1$  is close to  $x_0$ . We thus propose to define our *optimal* value of  $\lambda_n$  as  $\lambda_{\text{opt}} := \arg \min\{\Lambda_n^2(\lambda), \lambda \in \mathbb{R}^d\}$ , with

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

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  $\Lambda_n$  is unknown in practice and should be estimated. We propose to use the following estimator

$$\hat{\Lambda}_n(\lambda) := \sum_{i=1}^n \mathcal{W}_{n,i}(x_0) \mathbb{I}_{\{Y_i > \hat{Q}_{n,-i}(\alpha_n | X_i, \lambda)\}} - \alpha_n. \quad (26)$$

The estimated optimal value of the hyperparameter  $\lambda_n$  is thus given by

$$\hat{\lambda}_{n,\text{opt}} := \arg \min\{\hat{\Lambda}_n^2(\lambda), \lambda \in \mathbb{R}^d\}. \quad (27)$$

The estimator (26) can be motivated by the following result.

**Proposition 6** *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 6 is satisfied for the nearest neighbors' approach with the function  $\varphi$  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  $d_{(1)}(x_0) \leq \dots \leq d_{(n-1)}(x_0)$  are the corresponding ordered values.

This is also the case for the Nadaraya-Watson's weights by using the function

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

## 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  $\alpha_n$  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 2 are satisfied for such a sequence  $\alpha_n$ . In practice, we take  $\alpha_n = 20/n$  and the quantile  $Q(\alpha_n | x_0)$  is estimated using (1). Three sets of random weights are considered:

i) Nadaraya-Watson's weights with the Epanechnikov kernel given by

$$K(u) = \frac{3}{4}(1-u^2)\mathbb{I}_{[0,1]}(|u|),$$

ii) the  $k_n$ -NN weights with  $\ell = 1$  (triangular  $k_n$ -NN weights),

iii) the linear combination of weights given in (20) with  $\tau = 1/2$  and  $k_n = \lfloor \kappa n h_n \rfloor$ .

We generate  $X$  from a standard uniform distribution and 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 (13) holds (i.e., condition **(H)** with  $g(\alpha|x_0) = \alpha$ ). The parameter  $\rho$  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 (13) with a conditional extreme value index given by  $\gamma(x) = -1/\theta_2(x) < 0$ .

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

Under this model, condition (13) is satisfied with  $\gamma(X) = 0$ .

We finally consider a model for which condition (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

$$\begin{aligned} a(\alpha^{-1}) &= Q(\alpha|x) = \exp\left\{\left[\frac{\ln(1/\alpha)}{\xi}\right]^{1/\theta(x)}\right\} \\ \text{and } g(\alpha) &= \alpha\theta(x)\xi\left[\frac{\ln(1/\alpha)}{\xi}\right]^{1-1/\theta(x)}. \end{aligned}$$

For each model,  $N = 500$  samples of size  $n = 1000$  have been generated. The hyperparameter  $\lambda_n$  is chosen according to (27) and the minimization is achieved

- over a regular grid  $\mathcal{H}$  of 20 points evenly spaced between 0.05 and 0.3 for the Nadaraya-Watson's weights,
- over a grid  $\mathcal{K}$  of 20 points evenly spaced between 100 and 600 for the Nearest-Neighbors' 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 error

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

where  $\bullet$  has to be replaced by NW (Nadaraya-Watson's weights), NN ( $k_n$ -NN approach) or LC (Linear combination of weights) and the index  $i$  refers to the  $i$ - simulation run. 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 1 to 4. 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 increase as  $|\rho|$  decreases. This is expected since the estimation is much more difficult when  $\rho$  is close to 0 where a bias in the estimation appears. Also the ERRORS increase in general when  $\gamma(\cdot)$  increases;
- Concerning model **M2**, the ERRORS increase with  $\theta_2$ , i.e., when  $\gamma(\cdot) = -1/\theta_2(\cdot)$  increases, and decreases with  $\theta_1$ . Compared to the model **M1**, the ERRORS 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**, the ERRORS are not too much sensitive on the values of  $\sigma$ , 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**, the ERRORS depend a lot on the value of  $\xi$ . Indeed, if  $\xi$  is too small, the ERRORS 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  $\xi$  ( $\xi = 1$  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 the ERROR.

## 6 Proofs

### 6.1 Proof of the results given in Section 3

**Proof of Proposition 1** –

1. Since  $S$  is decreasing and  $\phi^{-1}$  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  $t_0 \in I$ ,

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

■

**Proof of Proposition 2** – We first prove that condition (9) implies condition (8). From de Haan and Ferreira (2006, Lemma 1.1.1), one has

$$\lim_{\alpha \rightarrow 0} \frac{S[Q(\alpha) + ta(\alpha^{-1})] - \alpha}{g(\alpha)} = \phi^{-1}(t),$$

for all  $t \in I$ . Replacing  $\alpha$  by  $S(x)$  leads to

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

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 (2006, Lemma 1.1.1), we have that for all  $t \in J$ ,

$$\lim_{n \rightarrow \infty} \frac{Q[n^{-1}(1 + tc_n)] - b_n}{a_n} = \phi(t). \quad (28)$$

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

$$Q\left(\frac{1 + tc_{\lfloor \alpha^{-1} \rfloor}}{\lfloor \alpha^{-1} \rfloor}\right) \leq Q[\alpha(1 + tc_{\lfloor \alpha^{-1} \rfloor})] \leq Q\left(\frac{1 + tc_{\lfloor \alpha^{-1} \rfloor}}{\lfloor \alpha^{-1} \rfloor + 1}\right). \quad (29)$$

Using (28), we know that

$$\frac{1}{a_{\lfloor \alpha^{-1} \rfloor}} \left[ Q\left(\frac{1 + tc_{\lfloor \alpha^{-1} \rfloor}}{\lfloor \alpha^{-1} \rfloor}\right) - b_{\lfloor \alpha^{-1} \rfloor} \right] \rightarrow \phi(t). \quad (30)$$

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  $mc_m \rightarrow \infty$ , we have  $\xi_t(m) \rightarrow 1$  as  $m \rightarrow \infty$ . Dini's theorem together with (28) entail that

$$\frac{1}{a_{\lfloor \alpha^{-1} \rfloor}} \left[ Q\left(\frac{1 + tc_{\lfloor \alpha^{-1} \rfloor}}{\lfloor \alpha^{-1} \rfloor + 1}\right) - b_{\lfloor \alpha^{-1} \rfloor} \right] \rightarrow \phi(t). \quad (31)$$

Hence, by collecting (29), (30) and (31) we obtain

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

with  $g(\alpha) = \alpha c_{\lfloor \alpha^{-1} \rfloor}$ ,  $b(\alpha) = b_{\lfloor \alpha^{-1} \rfloor}$  and  $a(\alpha^{-1}) = a_{\lfloor \alpha^{-1} \rfloor}$ . We conclude the proof by remarking that

$$\begin{aligned} \frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} &= \frac{Q[\alpha + tg(\alpha)] - b(\alpha)}{a(\alpha^{-1})} + \frac{Q(\alpha) - b(\alpha)}{a(\alpha^{-1})} \\ &\rightarrow \phi(t) + \phi(0) = \phi(t). \end{aligned}$$

■

**Proof of Proposition 3** – 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 (11) concludes the proof. ■

**Proof of Proposition 4** – From Proposition 2, the TFO condition entails that  $nc_n S(a_n t + b_n) - c_n \rightarrow \phi^{-1}(t)$  as  $n \rightarrow \infty$  with  $c_n = 1/[ng(1/n)]$ ,  $a_n = a(n)$  and  $b_n = Q(1/n)$ . First assume that  $\alpha/g(\alpha) \rightarrow c$  as  $\alpha \rightarrow 0$  with  $c > 0$ . We have that  $c_n \rightarrow c > 0$  as  $n \rightarrow \infty$  and thus  $nS(a_n t + b_n) \rightarrow 1 + \phi^{-1}(t)/c$ . In particular, we have that  $S(a_n t + b_n) \rightarrow 0$  and thus that, letting  $F := 1 - S$ ,  $-nS(a_n t + b_n) \sim \ln F^n(a_n t + b_n)$  as  $n \rightarrow \infty$ . Hence,

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

showing that condition (12) is satisfied. Now, let us consider the case  $c = 0$ . From Proposition 2, we have  $nc_n S(a_n t + b_n) \rightarrow \phi^{-1}(t)$ . Let  $m_n := nc_n = 1/g(1/n) =: \tilde{g}(n)$ . Since  $g(\alpha) \rightarrow 0$  as  $\alpha \rightarrow 0$ ,  $m_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Since  $g$  is a continuous and increasing function, we have that  $\tilde{g}^{-1}(m) \rightarrow \infty$  as  $m \rightarrow \infty$ . Letting  $\tilde{a}_m := a_{\tilde{g}^{-1}(m)}$  and  $\tilde{b}_m := b_{\tilde{g}^{-1}(m)}$  we obtain the convergence

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

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

**Proof of Proposition 5** – Let us show that (16) implies (15), the converse being similar. Let  $g(\alpha) = \alpha^2 b(\alpha^{-1})$ . Since  $g(\alpha)/\alpha \rightarrow 0$  as  $\alpha \rightarrow 0$ , one has for all  $t \in \mathbb{R}$

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

as  $\alpha \rightarrow 0$ . 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 (16) is locally uniform leading to (15). ■

## 6.2 Proof of Theorem 1

As explained in Section 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.

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  $T_n := \{T_{n,i}, 1 \leq i \leq n\}$  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  $T_n := \max\{T_{n,i}, 1 \leq i \leq n\}$ , 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** – Let  $\{t_{n,i}, i = 1, \dots, n\}$  be a triangular array of real numbers satisfying

$$\min(t_{n,i}; i = 1, \dots, n) \geq 0 \text{ and } \sum_{i=1}^n t_{n,i}^2 = 1. \quad (32)$$

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  $\nu_n t_n \rightarrow 0$  as  $n \rightarrow \infty$  then, for all  $z \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sum_{i=1}^n t_{n,i} V_{n,i} \leq z \right) = \Phi(z), \quad (33)$$

where  $\Phi$  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

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

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

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

since the  $V_{n,i}$  are identically distributed and under (32). Hölder's inequality entails that  $\mathbb{E} (V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}) \leq \nu_n^{2/3} [\mathbb{P} (t_n|V_n| > \varepsilon)]^{1/3}$ . Chebyshev's inequality ensures that  $\mathbb{P} (t_n|V_n| > \varepsilon) \leq t_n^2/\varepsilon^2$  and thus  $\mathbb{E} (V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}) \leq [\nu_n t_n/\varepsilon]^{2/3} \rightarrow 0$ , as  $n \rightarrow \infty$ , by assumption. Convergence (33) is thus proved for all triangular array  $\{t_{n,i}, i = 1, \dots, n\}$  satisfying (32) with  $\nu_n t_n \rightarrow 0$ .

Now, remark that for all  $\omega \in \{\nu_n T_n \rightarrow 0\}$ , convergence (33) entails that

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

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

$$\lim_{n \rightarrow \infty} \left| \mathbb{P} \left( \sum_{i=1}^n T_{n,i} V_{n,i} \leq z \mid \{T_{n,i}; i = 1, \dots, n\} \right) - \Phi(z) \right| = 0 \text{ a.s.} \quad (34)$$

To conclude the proof, let us remark that

$$\begin{aligned} & \left| \mathbb{P} \left( \sum_{i=1}^n T_{n,i} V_{n,i} \leq z \right) - \Phi(z) \right| \\ & \leq \mathbb{E} \left[ \left| \mathbb{P} \left( \sum_{i=1}^n T_{n,i} V_{n,i} \leq z \mid \{T_{n,i}; i = 1, \dots, n\} \right) - \Phi(z) \right| \right] =: P_n. \end{aligned}$$

Since

$$\left| \mathbb{P} \left( \sum_{i=1}^n T_{n,i} V_{n,i} \leq z \mid \{T_{n,i}; i = 1, \dots, n\} \right) - \Phi(z) \right| \leq 1,$$

the dominated convergence theorem entails that

$$\lim_{n \rightarrow \infty} P_n = \mathbb{E} \left[ \lim_{n \rightarrow \infty} \left| \mathbb{P} \left( \sum_{i=1}^n T_{n,i} V_{n,i} \leq z \mid \{T_{n,i}; i = 1, \dots, n\} \right) - \Phi(z) \right| \right] = 0,$$

from (34). ■

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 7** 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 (21) holds then  $v_n(x_0) \left( \widehat{S}_n^{x_0}(y_n(x_0)) - S(y_n(x_0)|x_0) \right) \xrightarrow{d} \mathcal{N}(0, 1)$ .

**Proof** – Remark that

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

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  $n_{x_0}/\sigma_n^2(x_0) \xrightarrow{a.s.} v_n^2(x_0)$  and that

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

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

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

**Proposition 8** 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 (22) holds then  $v_n(x_0) \left( \widehat{S}_n(y_n(x_0)|x_0) - \widehat{S}_n^{x_0}(y_n(x_0)) \right) \xrightarrow{\mathbb{P}} 0$ .

**Proof** – 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 (1987, Lemma 3.4.5), one has for all  $\varepsilon > 0$ ,

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

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

$$\begin{aligned} \mathbb{E} \left[ \Delta_{n,i}(x_0) \middle| \mathbb{X} \right] &\leq \mathbb{E} \left[ \left| \mathbb{I}_{[S(y_n(x_0)|x_0) - D_{n,i}(x_0), S(y_n(x_0)|x_0) + D_{n,i}(x_0)]}(U_i) \right| \middle| \mathbb{X} \right] \\ &\leq 2D_{n,i}(x_0). \end{aligned}$$

Remarking that

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

leads to  $\mathbb{P}(|B_n(x_0)| > \varepsilon) \leq \varepsilon + \mathbb{P}[v_n(x_0)W_1(\mathcal{W}_{n,x_0}^*, \delta_{x_0}^*) > \varepsilon^2/2]$ . The result is thus proved by using assumption (23) (or equivalently (22)).  $\blacksquare$

Theorem 1 is thus proved by gathering Propositions 7 and 8.

### 6.3 Proof of Theorem 2

The proof follows the lines described in Section 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  $z \in \mathbb{R}$ ,

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

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

$$\lim_{(t,y) \rightarrow (t_0, y^*(x_0))} \frac{\Psi(y)}{t} \left( \frac{S[y + td(y)|x_0]}{S(y|x_0)} - 1 \right) = \lim_{t \rightarrow t_0} \frac{\phi^{-1}(t)}{t}. \quad (35)$$

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  $t_n(x_0) \rightarrow 0$  as  $n \rightarrow \infty$ , (35) entails that as  $n \rightarrow \infty$

$$z_n(x_0) \sim -zv_n(x_0)t_n(x_0)g(\alpha_n) = z. \quad (36)$$

Now, to prove that  $Z_n(x_0) \xrightarrow{d} \mathcal{N}(0, 1)$ , it suffices to show that conditions (21) and (22) hold for  $y_n(x_0)$ . From (36),

$$1 - \frac{S[y_n(x_0)|x_0]}{\alpha_n} \sim z\alpha_n^{-1}v_n^{-1}(x_0) = z(\alpha_n m_n(x_0))^{-1/2} \rightarrow 0, \quad (37)$$

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

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

is equivalent to condition (21) with  $y_n(x_0)$ . It remains to prove condition (22). From (37), 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,

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

and the proof is complete.  $\blacksquare$

## 6.4 Proof of Corollaries 1, 2 and 3

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 (2009, 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  $h_n \rightarrow 0$  such that  $nh_n^p / \log \log n \rightarrow \infty$ , 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, 2005) or  $K$  is an indicator function on a cell of  $\mathbb{R}^p$  (see van der Vaart and Wellner, 1996, 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, 2015, 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 1, 2 and 3 are direct consequences of Theorem 2 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  $h_n \rightarrow 0$  and  $y_n(x_0) \uparrow y^*(x_0)$  such that  $nh_n^p [S(y_n(x_0)|x_0) \wedge (\log \log n)^{-1}] \rightarrow \infty$  and*

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

one has

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

**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,

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



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  $nh_n^p S(y_n(x_0)|x_0) \rightarrow \infty$ , we have

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

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

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

from which Lemma 3 follows according to Theorem 1.  $\blacksquare$

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

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

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

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

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

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

as  $n \rightarrow \infty$ . 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  $k_n S(y_n(x_0)|x_0) \rightarrow \infty$ , we have

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

Using the bound

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

we prove Lemma 4 by applying Theorem 1.  $\blacksquare$

**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  $nh_n^p/\log \log n \rightarrow \infty$ ,  $\ell_n S(y_n(x_0)|x_0) \rightarrow \infty$  with  $\ell_n := (nh_n^p \wedge k_n)$  and*

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

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

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

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

**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  $nh_n^p / \log \log n \rightarrow \infty$  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  $n \rightarrow \infty$ . 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,

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

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

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

almost surely for all  $\kappa \in [0, \infty]$ . As a consequence, since  $\ell_n S(y_n(x_0)|x_0) \rightarrow \infty$ , condition (21) is satisfied. Finally, using the bounds obtained in the proofs of Lemmas 3 and 4, one has

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

and thus condition (22) holds. Theorem 1 concludes the proof.  $\blacksquare$

## 6.5 Proof of Proposition 6

Proposition 6 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** – Since  $(X, Y)$  and  $Z$  are independent

$$\mathbb{E}[g(X, Y, Z)] = \int g(x, y, z) \mathbb{P}_{(X,Y)}(dx, dy) \mathbb{P}_Z(dz).$$

Hence

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

The conclusion follows since  $X$  and  $Z$  are independent.  $\blacksquare$

**Proof of Proposition 6** – 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  $\mathbb{E}[W_{n,1}(x_0)] = \dots = \mathbb{E}[W_{n,n}(x_0)] = 1/n$ . It thus remains to show that

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n W_{n,i}(x_0) \mathbb{I}_{\{Y_i > \widehat{Q}_{n,-i}(\alpha_n | X_i, \lambda)\}} \right] &= \mathbb{E} \left[ W_{n,1}(x_0) \mathbb{I}_{\{Y_1 > \widehat{Q}_{n,-1}(\alpha_n | X_1, \lambda)\}} \right] \\ &= \mathbb{E} \left[ W_{n,1}(x_0) S[\widehat{Q}_{n,-1}(\alpha_n | X_1, \lambda) | X_1] \right]. \end{aligned}$$

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  $\phi$  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,  $\Psi(t, u) = \varphi(t, u) S(\phi(\alpha_n, t, u) | t)$ . ■

## References

- [1] Berline, 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.
- [2] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B), 2557–2589.
- [3] 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.
- [4] Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, 33(3), 1380–1403.
- [5] 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.
- [6] 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.
- [7] Giné, E. and Nickl, R. (2015). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, New York, USA.
- [8] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, 44(3), 423–453.
- [9] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An introduction*. Springer Series in Operations Research and Financial Engineering, Springer.

- [10] He, X. and Ng, P. (1999). Quantile splines with several covariates. *Journal of Statistical Planning and Inference*, 75, 343–352.
- [11] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.
- [12] Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81, 673–680.
- [13] Nadaraja, E.A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9, 141–142.
- [14] Owen, A.B. (1987). *Nonparametric conditional estimation*. Ph.D Dissertation, Stanford University.
- [15] Racine, J.S. and Li, K. (2017). Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach. *Journal of Econometrics*, 201, 72–94.
- [16] 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.
- [17] Watson, G.S. (1964). Smooth regression analysis. *Sankhya A*, 26, 359–372.

Table 1: ERROR based on 500 samples of size  $n = 1000$  according to the model **M1**, for three different values of  $\rho$ .

	$\rho = -2$			$\rho = -1$		
	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$ $\gamma(x_0^{(1)}) = 1/3$	0.04082	0.03314	0.03941	0.04218	0.03359	0.03975
$x_0 = x_0^{(2)}$ $\gamma(x_0^{(2)}) = 1/2$	0.07931	0.07730	0.09417	0.08199	0.08832	0.09737
$x_0 = x_0^{(3)}$ $\gamma(x_0^{(3)}) = 1/4$	0.04125	0.04121	0.04138	0.04156	0.04262	0.04320

	$\rho = -0.5$		
	NW	NN	LC
$x_0 = x_0^{(1)}$ $\gamma(x_0^{(1)}) = 1/3$	0.05060	0.03639	0.04500
$x_0 = x_0^{(2)}$ $\gamma(x_0^{(2)}) = 1/2$	0.10928	0.11655	0.12386
$x_0 = x_0^{(3)}$ $\gamma(x_0^{(3)}) = 1/4$	0.03954	0.04021	0.04841

Table 2: ERROR based on 500 samples of size  $n = 1000$  according to the model **M2**, for three different values of  $\theta_1$ .

	$\theta_1 = 1$			$\theta_1 = 2$		
	<i>NW</i>	<i>NN</i>	<i>LC</i>	<i>NW</i>	<i>NN</i>	<i>LC</i>
$x_0 = x_0^{(1)}$ $\theta_2(x_0^{(1)}) = 3$	0.00476	0.00501	0.00541	0.00190	0.00198	0.00217
$x_0 = x_0^{(2)}$ $\theta_2(x_0^{(2)}) = 2$	0.00177	0.00177	0.00197	0.00063	0.00061	0.00068
$x_0 = x_0^{(3)}$ $\theta_2(x_0^{(3)}) = 4$	0.00956	0.01077	0.01076	0.00436	0.00489	0.00509

	$\theta_1 = 3$		
	<i>NW</i>	<i>NN</i>	<i>LC</i>
$x_0 = x_0^{(1)}$ $\theta_2(x_0^{(1)}) = 3$	0.00103	0.00113	0.00116
$x_0 = x_0^{(2)}$ $\theta_2(x_0^{(2)}) = 2$	0.00032	0.00031	0.00034
$x_0 = x_0^{(3)}$ $\theta_2(x_0^{(3)}) = 4$	0.00271	0.00291	0.00303

Table 3: ERROR based on 500 samples of size  $n = 1000$  according to the model **M3**, for three different values of  $\sigma$ .

	$\sigma = 1/2$			$\sigma = 1$		
	<i>NW</i>	<i>NN</i>	<i>LC</i>	<i>NW</i>	<i>NN</i>	<i>LC</i>
$x_0 = x_0^{(1)}$ $\mu(x_0^{(1)}) = 1/3$	0.00478	0.00553	0.00556	0.00595	0.00642	0.00636
$x_0 = x_0^{(2)}$ $\mu(x_0^{(2)}) = 1/2$	0.00413	0.00428	0.00423	0.00515	0.00548	0.00516
$x_0 = x_0^{(3)}$ $\mu(x_0^{(3)}) = 1/4$	0.00672	0.00751	0.00607	0.00591	0.00728	0.00542

	$\sigma = 3/2$		
	<i>NW</i>	<i>NN</i>	<i>LC</i>
$x_0 = x_0^{(1)}$ $\mu(x_0^{(1)}) = 1/3$	0.00619	0.00725	0.00682
$x_0 = x_0^{(2)}$ $\mu(x_0^{(2)}) = 1/2$	0.00572	0.00632	0.00579
$x_0 = x_0^{(3)}$ $\mu(x_0^{(3)}) = 1/4$	0.00618	0.00749	0.00540

Table 4: ERROR based on 500 samples of size  $n = 1000$  according to the model **M4**, for three different values of  $\xi$ .

	$\xi = 1/2$			$\xi = 1$		
	<i>NW</i>	<i>NN</i>	<i>LC</i>	<i>NW</i>	<i>NN</i>	<i>LC</i>
$x_0 = x_0^{(1)}$ $\theta(x_0^{(1)}) \approx 0.871$	299.83	364.56	291.7	1.0791	1.6770	1.02256
$x_0 = x_0^{(2)}$ $\theta(x_0^{(2)}) = 0.95$	48.736	744.02	640.94	0.9140	1.5938	1.47679
$x_0 = x_0^{(3)}$ $\theta(x_0^{(3)}) = 0.83125$	153589	86805	$2 \times 10^6$	4.2959	5.4753	15.309

	$\xi = 3/2$		
	<i>NW</i>	<i>NN</i>	<i>LC</i>
$x_0 = x_0^{(1)}$ $\theta(x_0^{(1)}) \approx 0.871$	0.2345	0.2498	0.20049
$x_0 = x_0^{(2)}$ $\theta(x_0^{(2)}) = 0.95$	0.2184	0.2777	0.24919
$x_0 = x_0^{(3)}$ $\theta(x_0^{(3)}) = 0.83125$	0.3049	0.4025	0.82955