Estimation of conditional quantiles from data with additional measurement errors *

Matthias Hansmann† and Michael Kohler[1]
Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: hansmann@mathematik.tu-darmstadt.de, kohler@mathematik.tu-darmstadt.de

June 1, 2016

Abstract
In this paper we study the problem of estimating conditional quantiles from data that contains additional measurement errors. The only assumption on these errors is that a weighted sum of the absolute errors tends to zero with probability one for sample size tending to infinity. We show that the plug-in quantile estimate corresponding to a local averaging estimate of the conditional distribution function (codf.) approaches the quantile set asymptotically, presumed that the local averaging estimate of the codf. is pointwise strongly consistent. Furthermore, we show that the above mentioned local assumption on the measurement errors can not be replaced by a global one. We also investigate the rate of convergence and show that our plug-in estimate archieves at least the same pointwise rate of convergence as the local averaging estimate of the codf. Finally, the results are applied in simulations and in the context of experimental fatigue tests.

AMS classification: Primary 62G05; secondary 62G20.

Key words and phrases: Conditional quantile estimation, consistency, experimental fatigue tests, rate of convergence.

1. Introduction

Let (X, Y) be a random vector, such that X is $\mathbb{R}^d$- valued and Y is real-valued, with conditional distribution function (codf.) F, i.e., $F(y,x) = \mathbf{P}\{Y \le y | X = x\}$. Conveniently, we write $\mathbf{P}(\cdot| x)$ instead of $\mathbf{P}(\cdot| X = x)$. For $\alpha \in (0,1)$ denote by

$$Q_{Y,\alpha}(x) := \{z \in \mathbb{R} : \mathbf{P}(Y \le z|x) \ge \alpha \quad \text{and} \quad \mathbf{P}(Y \ge z|x) \ge 1-\alpha\}$$

the set of all conditional α-quantiles of Y given $X = x$. More precisely for $x \in \mathbb{R}^d$ fixed, we have

$$Q_{Y,\alpha}(x) = \left[q^{[low]}_{Y,\alpha}(x), q^{[up]}_{Y,\alpha}(x)\right],$$

*Running title: *Estimation of conditional quantiles*
†Corresponding author. Tel: +49-6151-16-23371

where

$$q_{Y,\alpha}^{[low]}\left(x\right) := \min\left\{z \in \mathbb{R} : F\left(z, x\right) \geq \alpha\right\}$$

is the lower α-quantile and

$$q_{Y,\alpha}^{[up]}\left(x\right) := \sup\{z \in \mathbb{R} : F\left(z, x\right) \leq \alpha\}$$

is the upper α-quantile.

A brief overview on estimates of conditional quantiles is given in Yu et al. (2003). One idea to construct estimates, approaching the quantile set $Q_{Y,\alpha}\left(x\right)$ asymptotically for some fixed $x \in \mathbb{R}^d$, is to use an independent and identically distributed (i.i.d.) sample (X_1, Y_1), ..., (X_n, Y_n) of (X, Y) to compute a local averaging estimate

$$F_n(y, x) = \sum_{i=1}^{n} I_{\{Y_i \leq y\}} W_{n,i}\left(x\right) \tag{1}$$

of the codf. and to estimate the conditional quantile by the corresponding plug-in estimate

$$\hat{q}_{Y,n,\alpha}\left(x\right) = \min\{z \in \mathbb{R} : F_n(z, x) \geq \alpha\}. \tag{2}$$

Here $W_{n,i}\left(x\right)$ for $i = 1, ..., n$ are so called subprobability weights, i.e.,

$$W_{n,i}\left(x\right) \geq 0 \quad \text{for } i = 1, ..., n \quad \text{and} \sum_{i=1}^{n} W_{n,i}\left(x\right) \leq 1 \quad \text{for all } x \in \mathbb{R}^d,$$

which can depend on the samples $X_1, ..., X_n$.

The estimator introduced in (2) was considered by Stone (1977), in particular, he has shown in Theorem 3, that the conditional quantile estimate converges towards the quantile set in probability and in L_r for every $r \geq 1$ under some conditions on the weights $W_{n,i}$. These conditions are for example fullfilled by the weights of the nearest-neighbor estimate, presumed that ties occur only with probability zero (see proof of Theorem 6.1. in Györfi et al. (2002) for details and Chapter 6 in Györfi et al. (2002) for a definition of the estimate). This consistency result can also be extended to the weights of the partioning (see Chapter 4 in Györfi et al. (2002) for a definition) and kernel estimate (cf., e.g., Samanta (1989) for a pointwise consistency result). The kernel estimate of the codf. is defined by the weights

$$W_{n,i}\left(x\right) = \frac{K\left(\frac{x - X_i}{h_n}\right)}{\sum\limits_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)}, \tag{3}$$

where $0/0 = 0$ by definition (cf., e.g., Nadaraya (1964) and Watson (1964)). Here $h_n > 0$ is the so-called bandwidth and $K : \mathbb{R}^d \to \mathbb{R}$ is a so-called kernel function, e.g., the so-called naive kernel defined by

$$K\left(x\right) = I_{\{||x|| \leq 1\}} \quad \text{for all } x \in \mathbb{R}^d.$$

Almost sure rates of convergence for the plug-in conditional quantile estimate with kernel and nearest-neighbor weigths have been shown in Bhattacharya and Gangopadhyay (1990) by deriving a Bahadur-type representation (cf., e.g., Bahadur (1966)).

Other plug-in conditional quantile estimates have for example been considered by Stute (1986a), who showed asymptotic normality of the plug-in quantile estimates corresponding to a nearest neighbor-type estimator of the codf. The asymptotic behaviour of the mean squared error of a plug-in quantile estimate corresponding to a double-kernel local linear estimate of the codf. has been investigated by Yu and Jones (1998). Furthermore, Troung (1989) introduced a kernel estimator for the conditional median and showed that it achieves the optimal global rate of convergence in the sense of Stone (1982) both pointwise and in the L_r-norm restricted to a compact for all $1 \leq r \leq \infty$.

Since the estimate introduced in (2) depends on an estimate of the codf., it is also of great interest to study results concerning nonparametric estimates of the codf. But this is actually a special type of nonparametric regression. Rates of convergence in probability for the kernel regression estimate have been obtained in Krzyżak and Pawlak (1987) and in Györfi (1981) for the nearest neighbor regression estimate. Uniform almost sure rates of convergence for regression estimates have been shown in Härdle et al. (1988) by considering a more gernal setting of kernel-type estimators of conditional functionals. Optimal global rates of convergence for nonparametric regression estimates have been shown by Stone (1982). Other estimates of the codf. have been proposed by Hall et al. (1999), who studied the rate of convergence of a weighted kernel estimator. Cai (2002) showed asymptotic normality of this estimate of the codf. and of the corresponding plug-in conditional quantile estimate in the context of α-mixing time series. Furthermore, Hall and Yao (2005) used a dimension reduction technique to approximate the codf. and study the asymptotic properties. Preadjusted local averaging estimates of the codf. were proposed by Veraverbeke et al. (2014), who proved results concerning the uniform rate of convergence.

Another approach to obtain conditional quantile estimates, without using an estimate of the codf., is based on the fact that the quantile set $Q_{Y,\alpha}(x)$ consists of exactly those points $q \in \mathbb{R}$, that minimize the conditional risk

$$\mathbf{E}\left\{\rho_\alpha(Y-q) \,|\, X=x\right\}, \tag{4}$$

where

$$\rho_\alpha(t) = t \cdot \left(\alpha - I_{\{t<0\}}\right)$$

is the so called pinball loss function. Under the assumption that the conditional quantile is unique, i.e., that $Q_{Y,\alpha}(x)$ consists only of one point, several authors approximated the conditional risk in (4) and proposed a parametric conditional quantile estimate that minimizes the approximated conditional risk. See Koenker (2005) for a detailed overview on this approach, which is usually referred to as quantile regression in the literature. The approximation of the conditional risk function using an i.i.d. sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ of (X,Y) can for example be done by an empirical risk function (cf., e.g., Koenker and Bassett (1978,1982), Lejeune and Sarda (1988)) or by empirical kernel weighted risk function (cf., e.g., Chaudhuri (1991), Yu and Jones (1997,1998)). Furthermore, Powell (1986) used the idea of minimizing an empirical risk function to estimate conditional quantiles in a censored regression model.

In order to generalize the above technique to nonpolynomial approaches for the quantile function, Takeuchi et al. (2006) proposed to consider the minimization problem in (4) over a reproducing kernel Hilbert space of possible quantile functions. As an estimate for the quantile function, they chose a function from this Hilbert space that minimizes the empirical risk functional plus a regularity term. For this estimate Christmann and Steinwart (2008) showed consistency results.

Usefulness and possible applications of estimates of conditional quantiles have for example been illustrated in the context of income evaluation (cf., e.g., Hogg (1975) and Fan and Gijbels (1996)), human medicine (cf., e.g., Cole (1988) and Cole and Green (1992)) and finance (cf., e.g., Cai and Wang (2008)). See also Yu et al. (2003) for an overview of further applications.

In this paper we assume that instead of the i.i.d. sample (X_1, Y_1), ..., (X_n, Y_n) of (X, Y) we have available only data $\left(X_1, \bar{Y}_{1,n}\right)$, ..., $\left(X_n, \bar{Y}_{n,n}\right)$ such that a weigthed sum of absolute errors between Y_i and $\bar{Y}_{i,n}$ converges to zero almost surely, i.e., we assume that

$$\sum_{i=1}^{n} |Y_i - \bar{Y}_{i,n}| \cdot W_{n,i}(x) \to 0 \quad a.s. \text{ for } \mathbf{P}_X\text{–almost every } x, \tag{A1}$$

where $\mathbf{P}_X$ is the of X induced measure on $\left(\mathbb{R}^d, \mathcal{B}_d\right)$, i.e., $\mathbf{P}_X(B) = \mathbf{P}(X \in B)$ for every $B \in \mathcal{B}_d$.

Here we do not assume anything on the measurement errors $\bar{Y}_{i,n} - Y_i$ $(i = 1, \ldots, n)$. In particular, we do not assume that those errors have to be random and in case that they are random they do not need to be independent or identically distributed and they do not need to have expectation zero or a density w.r.t. the Lebesgue measure, so estimates for convolution problems (see, e.g., Meister (2009) and the literature cited therein) are not applicable in the context of this paper. Note also that our set-up is triangular.

Since we do not assume anything on the nature of the measurement errors besides that they are pointwise asymptotically negligible in the sense that (A1) holds, it seems to be a natural idea to ignore them completely and to try to use the same estimates as in the case that an independent and identically distributed sample is given.

The investigation of additional measurement errors in the dependent variable is motivated by experimental fatigue tests from the Collaborative Research Center 666 at the Technische Universität Darmstadt, where we have to use measured data from similar materials to obtain a sufficient large number of samples to estimate the quantiles of the number of cycles until failure (cf., Section 3 below).

Additional measurement errors in the covariate have for example been considered in the context of quantile regression by He and Liang (2000), who assumend that the errors in X and Y are independent and have a common symmetric distribution, and by Wei and Carroll (2009), who assumend that the measurement error in the sample of the covariate $\bar{Y}_i$ has an α-quantile of zero conditional on X_i. However, in both references the authors assumed that X and Y fullfill a linear model. Although Schennach (2008) does not make this linearity assumption, she also proposed only a consistent quantile regression estimate for measurement errors in the covariate and not in the dependent variable. Furthermore, other non-i.i.d. data in the context of conditional quantile estimation typically occurs in connection with autoregressive time series models and has for example been considered by Portnoy (1991), Koenker and Zhao (1996) and Xiao and Koenker (2009).

In the context of quantile estimation additional measurement errors have been considered in Hansmann and Kohler (2016). In this paper we extend the results to conditional quantile estimation. In particular, we investigate whether the above defined plug-in estimate is strongly universally pointwise consistent for the quantile set, if the data contains additional measurement errors fullfilling (A1). We show in Theorem 1 that this holds, presumed that the estimate of the codf. fullfills for every $y \in \mathbb{R}$

$$F_n(y,x) \to F(y,x) \quad a.s. \text{ for } \mathbf{P}_X\text{–almost every } x.$$

In Corollary 1 we proof that this assumption is for example fullfilled by the kernel estimate of the codf. and obtain a more general consistency result than Theorem 1 in Samanta (1989) with weaker assumptions. Moreover, we show in Theorem 2 that the local assumption in (A1) on the measurement error can not be replaced by

$$\frac{1}{n}\sum_{i=1}^{n} \left|Y_i - \bar{Y}_{i,n}\right| \to 0 \quad a.s.,$$

i.e., by the assumption that the average global sum of the absolute error tends to zero with probability one.

Furthermore, we investigate how the additional measurement error influence the pointwise rate of convergence in probability. As we show in Theorem 3, if our estimate of the codf. archieves pointwise in x and locally uniform in y a rate of convergence of r_n and if we know a pointwise upper bound $\eta_n(x)$ on the measurement error in (A1), our plug-in estimate obtains a pointwise rate of convergence in probability of

$$r_n + \sqrt{\eta_n(x)}.$$

In particular, in Corollary 2 we show that it is possible with the kernel estimate of the codf. to obtain the rate $r_n = (\log(n)/n)^{p/(2p+d)}$.

Throughout this paper the following notation is used: We write $V_n = O_{\mathbf{P}}(W_n)$ if the nonnegative random variables V_n and W_n satisfy

$$\lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\{V_n > c \cdot W_n\} = 0.$$

The sets of natural positive, natural nonnegative and real numbers are denoted by $\mathbb{N}$, $\mathbb{N}_0$ and $\mathbb{R}$, respectively. We write $\to^{\mathbf{P}}$ as an abbreviation for convergence in probability and I_A for the indicator function of the set A. We denote the Euclidian Norm on $\mathbb{R}^d$ by $||\cdot||$. For $z \in \mathbb{R}$ and a set $A \subseteq \mathbb{R}$, we define the distance from z to A as

$$dist(z,A) := \inf_{a\in A} |z-a|.$$

The outline of the paper is as follows: The main results are formulated in Section 2 and proven in Section 4. In Section 3 we apply our estimate to simulated data to illustrate the finite sample size performance, and we present an application of our estimates in the context of experimental fatigue tests.

2. Main results

Let

$$\bar{F}_n(y,x) = \sum_{i=1}^{n} I_{\{\bar{Y}_{i,n} \le y\}} W_{n,i}(x) \tag{6}$$

be a local averaging estimate of the codf. $F(y,x)$ corresponding to the data $\left(X_1, \bar{Y}_{1,n}\right)$, $\ldots$, $\left(X_n, \bar{Y}_{n,n}\right)$ and let

$$\hat{q}_{\bar{Y},n,\alpha}(x) = \min\{z \in \mathbb{R} : \bar{F}_n(z,x) \ge \alpha\} \tag{7}$$

be the corresponding plug-in estimate.

2.1. Consistency

First of all we want to investigate, under which conditions the estimator $\hat{q}_{\bar{Y},n,\alpha}(x)$ is pointwise strongly consistent for the quantile set $Q_{X,\alpha}(x)$. The following result holds.

Theorem 1. *Let $(X,Y), (X_1,Y_1), (X_2,Y_2)\ldots$ be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$-valued random vectors and let $\alpha \in (0,1)$ be arbitrary. Let the subprobability weights $W_{n,i}$ be such that the local averaging estimate F_n of the codf. F as defined in (1) fullfills for every $y \in \mathbb{R}$*

$$F_n(y,x) \to F(y,x) \quad a.s. \;\; for\ \mathbf{P}_X\text{–almost every } x. \tag{8}$$

Furthermore let $\bar{Y}_{1,n}, \ldots, \bar{Y}_{n,n}$ be random variables, which fullfill

$$\sum_{i=1}^{n} |Y_i - \bar{Y}_{i,n}| \cdot W_{n,i}(x) \to 0 \quad a.s. \;\; for\ \mathbf{P}_X\text{–almost every } x \tag{A1}$$

and let $\bar{F}_n$ be the local averaging estimate defined in (6) with weights $W_{n,i}$. Then the quantile estimate $\hat{q}_{\bar{Y},n,\alpha}(x)$ defined in (7) is strongly consistent in the sense that

$$dist\left(\hat{q}_{\bar{Y},n,\alpha}(x), Q_{Y,\alpha}(x)\right) \to 0 \quad a.s. \;\; for\ \mathbf{P}_X\text{–almost every } x.$$

In the following corollary we formulate sufficient conditions for the strong pointwise consistency of the plug-in quantile estimate corresponding to a kernel estimate of the conditional distribution function, defined by the weights in (3).

Corollary 1. *Let $(X,Y), (X_1,Y_1), (X_2,Y_2)\ldots$ be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$-valued random vectors and let $\alpha \in (0,1)$ be arbitrary. Assume that K is the naive kernel and that the bandwidth $h_n > 0$ fullfills*

$$h_n \to 0 \quad and \quad n \cdot h_n^d / \log(n) \to \infty \quad (n \to \infty). \tag{K1}$$

Let $\bar{Y}_{1,n}, \ldots, \bar{Y}_{n,n}$ be random variables, which fullfill

$$\sum_{i=1}^{n} \frac{|Y_i - \bar{Y}_{i,n}| \cdot K\left(\frac{x-X_i}{h_n}\right)}{\sum\limits_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \to 0 \quad a.s. \ \textit{for } \mathbf{P}_X\textit{–almost every } x. \tag{A2}$$

Furthermore let $\bar{F}_n$ be the corresponding kernel estimate of the codf. with kernel K and bandwidth h_n. Then the quantile estimate $\hat{q}_{\bar{Y},n,\alpha}(x)$ defined in (7) is strongly consistent in the sense that

$$dist\left(\hat{q}_{\bar{Y},n,\alpha}(x), Q_{Y,\alpha}(x)\right) \to 0 \quad a.s. \ \textit{for } \mathbf{P}_X\textit{–almost every } x.$$

Proof. Theorem 1 in Stute (1986b) implies (8) for the kernel estimate and therefore Theorem 1 yields the assertion (see also Theorem 25.11 in Györfi et al. (2002) for an alternative proof of (8) for the kernel estimate). □

Remark 1. Analogous results can be shown for plug-in quantile estimate corresponding to the partitioning and nearest neighbor estimate of the codf., using the results of Theorems 25.6. and 25.17. in Györfi et al. (2002), respectively.

In Theorem 1 it was assumed that (A1) holds, which says that the (locally) weighted sum of the absolute errors tends to zero almost surely. We now want to investigate, whether this assumption can be replaced by

$$\frac{1}{n} \sum_{i=1}^{n} \left|Y_i - \bar{Y}_{i,n}\right| \to 0 \quad a.s. \tag{12}$$

i.e., by the assumption that the average (global) sum of the absolute error tends to zero almost surely. Our next result shows that this is not possible in general.

Theorem 2. *The assumption (12) is not strong enough to obtain the results of Theorem 1. More precisely, let $W_{n,i}$ be the weights of the kernel estimate, with naive kernel K and a positive bandwidth h_n that fullfills the assumptions of Corollary 1. Then there exist independent and indentically distributed random vectors $(X, Y), (X_1, Y_1), (X_2, Y_2) \ldots$ and random variables $\bar{Y}_{1,n}, \ldots, \bar{Y}_{n,n}$ which fullfill (12) such that*

$$dist\left(\hat{q}_{\bar{Y},n,\alpha}(x), Q_{Y,\alpha}(x)\right) \to 0 \quad a.s.$$

does not hold for $\mathbf{P}_X$–almost every x.

2.2. Rate of convergence

Next we investigate the rates of convergence of our quantile estimates. The following result holds, which relates the locally uniform rate of convergence of the estimate of the codf. to the rate of the plug-in quantile estimate.

Theorem 3. *Let $(X,Y),(X_1,Y_1),(X_2,Y_2)\dots$ be independent and identically distributed $\mathbb{R}^d\times\mathbb{R}$-valued random vectors and let $\alpha\in(0,1)$ be arbitrary. Assume that the codf. $F(\cdot,x)$ is continuous and differentiable at $q_{Y,\alpha}^{[low]}(x)$ with derivative greater than zero for $\mathbf{P}_X$–almost every $x\in\mathbb{R}^d$. Let $W_{n,i}$ be subprobability weights, which are such that the local averaging estimate $F_n(y,x)$ converges to the codf. $F(y,x)$ for $\mathbf{P}_X$–almost every x locally uniform in y in probability with deterministic rate $r_n>0$, satisfying $r_n\to 0$ as $n\to\infty$, in the sense that for $\mathbf{P}_X$–almost every x there exists $\gamma(x)>0$ such that*

$$\sup_{\left|y-q_{Y,\alpha}^{[low]}(x)\right|\le\gamma(x)}|F_n(y,x)-F(y,x)|=O_{\mathbf{P}}(r_n)\,. \tag{13}$$

Furthermore let $\bar{Y}_{1,n},\dots,\bar{Y}_{n,n}$ be random variables, which fullfill

$$\eta_n(x):=\sum_{i=1}^{n}|Y_i-\bar{Y}_{i,n}|\cdot W_{n,i}(x)\to^{\mathbf{P}}0\qquad\textit{for } \mathbf{P}_X\textit{–almost every } x, \tag{A3}$$

let $\bar{F}_n$ be the corresponding local averaging estimate of the codf. with weights $W_{n,i}$ and let $\hat{q}_{\bar{Y},n,\alpha}(x)$ by the quantile estimate defined in (7). Then

$$\left|\hat{q}_{\bar{Y},n,\alpha}(x)-q_{Y,\alpha}^{[low]}(x)\right|=O_{\mathbf{P}}\left(r_n+\sqrt{\eta_n(x)}\right)\quad\textit{for } \mathbf{P}_X\textit{–almost every } x.$$

As in Section 2.1, we will apply the above result to the quantile estimate corresponding to the kernel estimate of the codf. Therefore we have to assume that for $\mathbf{P}_X$–almost all x the codf. $F(y,x)$ is locally Hölder continuous in x with exponent $0<p\le 1$, locally uniform in y. More precisely, we assume that for $\mathbf{P}_X$–almost every x there exist finite constants $C(x),\kappa_1(x),\kappa_2(x)>0$ such that

$$\sup_{\left|y-q_{Y,\alpha}^{[low]}(x)\right|\le\kappa_1(x)}|F(y,x)-F(y,z)|\le C(x)\cdot||x-z||^p \tag{15}$$

for all $z\in\mathbb{R}^d$ with $||z-x||\le\kappa_2(x)$. The following result will we be proven in Section 4.4.

Corollary 2. *Let $(X,Y),(X_1,Y_1),(X_2,Y_2)\dots$ be independent and identically distributed $\mathbb{R}^d\times\mathbb{R}$-valued random vectors and let $\alpha\in(0,1)$ be arbitrary. Assume that the codf. $F(\cdot,x)$ is continuous and differentiable at $q_{Y,\alpha}^{[low]}(x)$ with derivative greater than zero for $\mathbf{P}_X$–almost every $x\in\mathbb{R}^d$ and that F fullfills the smoothness assumption in (15) for some $0<p\le 1$. Let K be the naive kernel and let the bandwidth $h_n>0$ fullfill (K1). Furthermore let $\bar{Y}_{1,n},\dots,\bar{Y}_{n,n}$ be random variables, which satisfy*

$$\eta_n(x):=\sum_{i=1}^{n}\frac{|Y_i-\bar{Y}_{i,n}|\cdot K\left(\frac{x-X_i}{h_n}\right)}{\sum\limits_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)}\to^{\mathbf{P}}0\ \ \textit{for } \mathbf{P}_X\textit{–almost every } x. \tag{A4}$$

Let $\bar{F}_n$ be the corresponding kernel estimate of the codf. with naive kernel K and bandwidth h_n and let $\hat{q}_{\bar{Y},n,\alpha}(x)$ be the quantile estimate defined in (7). Then

$$\left|\hat{q}_{\bar{Y},n,\alpha}(x) - q_{Y,\alpha}^{[low]}(x)\right| = O_{\mathbf{P}}\left(\sqrt{\frac{\log(n)}{n \cdot h_n^d}} + h_n^p + \sqrt{\eta_n(x)}\right) \quad \textit{for } \mathbf{P}_X\textit{-almost every } x.$$

In particular, the choice of $h_n = \tilde{c} \cdot \left(\frac{\log(n)}{n}\right)^{\frac{1}{2p+d}}$ leads to

$$\left|\hat{q}_{\bar{Y},n,\alpha}(x) - q_{Y,\alpha}^{[low]}(x)\right| = O_{\mathbf{P}}\left(\left(\frac{\log(n)}{n}\right)^{\frac{p}{2p+d}} + \sqrt{\eta_n(x)}\right) \quad \textit{for } \mathbf{P}_X\textit{-almost every } x.$$

Remark 2. Similar results can analogously be shown for the plug-in quantile estimates corresponding to the nearest neighbor and partitioning estimate of the codf. In both cases it is possible to archieve a rate of convergence in probability of $\left(\frac{\log(n)}{n}\right)^{\frac{p}{2p+d}} + \sqrt{\eta_n(x)}$ by choosing a sufficient number of nearest neighbors that have to be considered or special cubic partitions, respectively.

3. Application to simulated and real data

In this section we apply the above described methods to simulated and real data and estimate $5\%-$, $50\%-$, $90\%-$ and $95\%-$quantiles. Therefore we use the kernel weights with naive kernel in our local averaging estimate of the codf., where we choose the bandwidth h_n data-dependent from the set $\{0.05, 0.1, 0.2, 0.3\}$ by cross-validation w.r.t. the estimate of the codf. (cf. Section 8 in Györfi et al. (2002)). In order to classify our estimates, we firstly consider distributions with known quantiles, afterwards we will apply our estimator in the context of experimental fatigue tests.

For the first purpose we use samples of sample sizes $n = 500, 1000$ and 2000. The consideration of these sample sizes is motivated by the application in the context of experimental fatigue tests, where we have 1222 data points. We consider the maximum absolute error

$$err_{max} := \max_{i=1,\dots,M}\left|\hat{q}_{\bar{Y},n,\alpha}(x_i) - q_{Y,\alpha}^{[low]}(x_i)\right|$$

on an equidistant grid $x_1, \dots, x_M$ for some fixed number $M \in \mathbb{N}$. Due to the random number generation in our simulated data, our quantile estimates contain randomness, therefore we repeat the quantile estimation 100 times with new random numbers and subscript our maximum absolute errors by an upper index i. We will compare our estimates by considering the average value $\frac{1}{100}\sum_{i=1}^{100} err_{max}^i$ of the maximum absolute error.

As a first example we choose $(X,Y), (X_1,Y_1), (X_2,Y_2), \dots$ as independent and identically distributed random vectors such that X is uniform distributed on $(0,2)$ and Y is normal-distributed with mean $X \cdot (1-X)$ and variance 1. As data with measurement error we set $\bar{Y}_{i,n} = Y_i + \frac{20}{n}$. Observe that we get completely new samples, when n changes.

	90%-quantile			95%-quantile		
size of n	500	1000	2000	500	1000	2000
average value of err_{max} for $q_{Y,n,\alpha}$	0.5076	0.3836	0.3081	0.5632	0.4290	0.3548
average value of err_{max} for $q_{\bar{Y},n,\alpha}$	0.5767	0.4174	0.3262	0.6247	0.4644	0.3726
average value of err_{max} for $q_{\bar{Z},n,\alpha}$	1.4244	0.8121	0.5508	4.1559	1.2767	0.6661

Table 1: Average value of err_{max} for $q_{Y,n,\alpha}$, $q_{\bar{Y},n,\alpha}$ and $q_{\bar{Z},n,\alpha}$ in the first example.

	90%-quantile			95%-quantile		
size of n	500	1000	2000	500	1000	2000
average value of err_{max} for $q_{Y,n,\alpha}$	0.5403	0.3933	0.3513	0.7363	0.5804	0.4890
average value of err_{max} for $q_{\bar{Y},n,\alpha}$	0.7554	0.6377	0.6020	1.0396	0.8688	0.8386

Table 2: Average value of err_{max} for $q_{Y,n,\alpha}$ and $q_{\bar{Y},n,\alpha}$ in the second example.

As a comparison to that we also consider $\bar{Z}_{i,n} = Y_i + \frac{20}{i}$, where the samples with bigger measurement errors are kept by. The grid $x_1, \ldots, x_{20}$ is chosen equidistantly on $[0, 2]$. Corollary 1 implies that $q_{Y,n,\alpha}(x)$, $q_{\bar{Y},n,\alpha}(x)$ and $q_{\bar{Z},n,\alpha}(x)$ are strongly consistent estimates for the quantile set $Q_{Y,\alpha}(x)$ for $\mathbf{P}_X$–almost all $x \in \mathbb{R}$, which in fact only consists of one point, that is equal to $q_{Y,\alpha}^{[low]}(x) = q_{Y,\alpha}^{[up]}(x)$. This result is confirmed by the average values of the maximum absolute error in Table 1. Especially for small sample sizes the estimator $q_{\bar{Y},n,\alpha}(x)$ yields smaller average squared errors than the estimator $q_{\bar{Z},n,\alpha}(x)$. This is due to the fact that the samples with bigger measurement errors are kept by. Furthermore, we can observe that the main part of the maximum error of $q_{\bar{Y},n,\alpha}(x)$ is not due to measurement errors, because the maximum error of the estimator $q_{Y,n,\alpha}(x)$ is not much smaller.

As a second example we choose $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots$ as independent and identically distributed random vectors such that X is normal distributed with mean 0 and variance 1 and Y is exponentially-distributed with mean $\left|\sqrt{X}\right|$. As data with measurement error we choose $\bar{Y}_{i,n} = Y_i + U_{i,n}$ where $U_{1,n}, \ldots, U_{n,n}$ are independent and uniformly on $(0, 1/n)$-distributed random variables. The grid $x_1, \ldots, x_{20}$ for the evaluation of the maximum error is chosen equidistanly on $[0, 1]$. As in the first example, we can conclude the strong consistency of the estimator $q_{\bar{Y},n,\alpha}(x)$ for $\mathbf{P}_X$–almost all $x \in \mathbb{R}$ by Corollary 1, which is confirmed by the average maximum errors in Table 2.

As a third example we choose $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots$ as independent and identically distributed random vectors with a discrete covariate, namely X as uniformly on $\{1, 2, 3, 4, 5\}$ distributed random variable. Furthermore we choose Y as χ^2-distributed with X degrees of freedom. As data with measurement error we set $\bar{Y}_{i,n} = Y_i + \epsilon_{i,n}$, where $\epsilon_{1,n}, \ldots, \epsilon_{n,n}$ are normal distributed random variables with mean and variance $1/n$, such that $\epsilon_{1,n}, \ldots, \epsilon_{n,n}, (X_1, Y_1), \ldots, (X_n, Y_n)$ are independent. Since the covariate is discrete, we change our set for the data-dependent choice of the bandwidth to $\{0.5, 1, 2\}$. The grid points for the evaluation of the maximum error are chosen as $x_i = i$ for $i = 1, 2, 3, 4, 5$. Again, Corollary 1 implies the strong consistency of the estimators

	90%-quantile			95%-quantile		
size of n	500	1000	2000	500	1000	2000
average value of err_{max} for $q_{Y,n,\alpha}$	1.1155	0.7618	0.5633	1.4971	1.1288	0.7463
average value of err_{max} for $q_{\bar{Y},n,\alpha}$	1.1165	0.7743	0.5854	1.5208	1.1752	0.7731

Table 3: Average value of err_{max} for $q_{Y,n,\alpha}$ and $q_{\bar{Y},n,\alpha}$ in the third example.

$q_{Y,n,\alpha}(x)$ and $q_{\bar{Y},n,\alpha}(x)$ for $x = 1, 2, 3, 4, 5$, which is confirmed by the average maximum errors in Table 3.

As a fourth example we want to consider a setting, in which the absolute error

$$\left|\hat{q}_{\bar{Y},n,\alpha}(x) - q_{Y,\alpha}(x)\right|$$

for some fixed x archieves asymptotically the claimed rate of convergence of Corollary 1, which actually shows in an empirical way that this rate can not be improved. Therefore we choose $(X,Y), (X_1,Y_1), (X_2,Y_2), \ldots$ as independent and identically distributed random vectors such that X is uniformly distributed on $(0,1)$ and Y is uniformly distributed on $(0,X)$. In this setting assumption (15) is fullfilled with a maximum value of $p = 1$. We choose $\alpha = 0.5$, the bandwidth h_n to be the asymptotically optimal one from Corollary 2, i.e.,

$$h_n = \frac{1}{5} \cdot \left(\frac{\log(n)}{n}\right)^{\frac{1}{3}}$$

and consider

$$\bar{Y}_{i,n} = \begin{cases} Y_i + \frac{10}{n^{0.1}} & \text{if} \quad K\left(\frac{X_i - 0.5}{h_n}\right) = 1, Y_i \in \left[q_{Y,\alpha}(0.5) - \frac{1}{n^{0.1}}, q_{Y,\alpha}(0.5)\right] \text{ and } Y_i \text{ is} \\ & \quad \text{one of the } \lfloor \frac{1}{n^{0.1}} \cdot \sum_{i=1}^{n} K\left(\frac{X_i - 0.5}{h_n}\right) \rfloor \text{ biggest samples of} \\ & \quad (Y_j)_{j=1,\ldots,n} \text{ in } \left[q_{Y,\alpha}(0.5) - \frac{1}{n^{0.1}}, q_{Y,\alpha}(0.5)\right] \\ Y_i & \text{else,} \end{cases}$$

where $q_{Y,\alpha}(0.5)$ is an abbreviation for the lower and upper α–quantile of Y conditional on $X = 0.5$, which are equal in this case. Clearly, for all $x \in \mathbb{R}$ with $x \neq 0.5$

$$\eta_n(x) = 0$$

is fullfilled for all n large enough, since the measurement error in Y_i does only occur, if $||X_i - 0.5|| \leq h_n$ and $h_n \to 0$ as $n \to \infty$. For $x = 0.5$ we have

$$\begin{aligned} \eta_n(0.5) &= \sum_{i=1}^{n} \frac{\left|Y_i - \bar{Y}_{i,n}\right| K\left(\frac{X_i - 0.5}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - 0.5}{h_n}\right)} \\ &\leq \frac{1}{\sum_{i=1}^{n} K\left(\frac{X_i - 0.5}{h_n}\right)} \cdot \frac{10}{n^{0.1}} \cdot \lfloor \frac{1}{n^{0.1}} \cdot \sum_{i=1}^{n} K\left(\frac{X_i - 0.5}{h_n}\right) \rfloor \leq \frac{10}{n^{0.2}} \to 0 \quad a.s. \end{aligned}$$

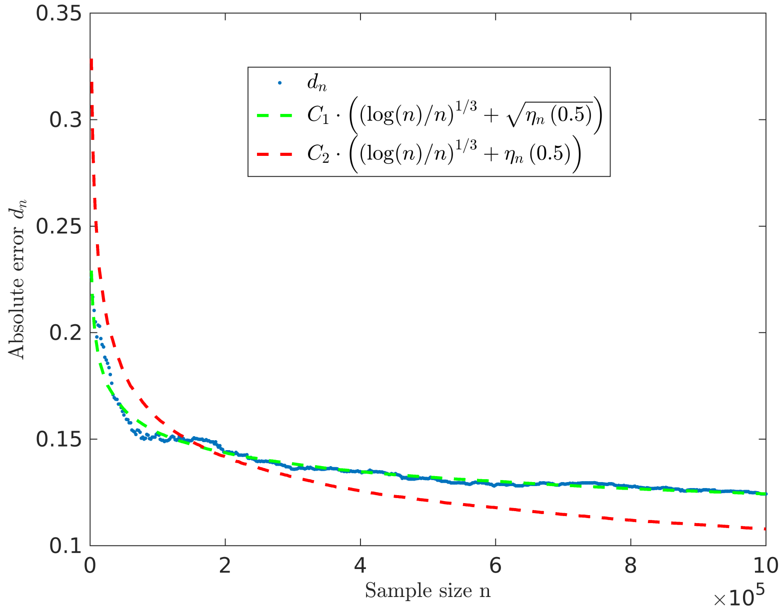

Figure 1: Typical asymptotic behaviour of $d_n = \left|\hat{q}_{\bar{Y},n,\alpha}(0.5) - q_{Y,\alpha}(0.5)\right|$ in the setting of the fourth example.

In this way we defined data with measurement errors such that the corresponding quantile is shifted to the right by some error proportional to $n^{-0.1}$ (if there are not too many samples in the relevant interval, which is asymptotically fullfilled), but the weighted sum of absolute errors η_n tends to zero faster than $10 \cdot n^{-0.2}$. In order to investigate the asymptotic behaviour of our estimate $\hat{q}_{\bar{Y},n,\alpha}(x)$, we consider the absolute error

$$d_n = \left|\hat{q}_{\bar{Y},n,\alpha}(0.5) - q_{Y,\alpha}(0.5)\right|$$

of the estimated α-quantile of Y conditional on $X = 0.5$ for sample sizes n in steps of 2000. As illustrated in Figure 1, the absolute error d_n has approximately the same asymptotic behaviour as the claimed rate $\left(\frac{\log(n)}{n}\right)^{1/3} + \sqrt{\eta_n(0.5)}$ of Corollary 2. In particular, it shows up that a rate of convergence which contains η_n instead of $\sqrt{\eta_n}$ is too fast in this setting. The occuring constants were chosen empirically and take the values $C_1 = C_2 = 0.14$.

As a last example we give an application of the methods above in the context of fatigue behaviour of steel under cyclic loading, that is motivated by experiments of the Collaborative Research Center 666 at the Technische Universität Darmstadt, which studies integral sheet metal design with higher order bifurcations. Here the main idea is to obtain several advantages concerning the material properties by producing structures out of one part by linear flow and bend splitting. Our main goal will be in the following to study, whether this modified, splitted material shows better fatigue behavior under cyclic loading than the base material. Therefore for each material m data

$$\left\{\left(\epsilon_1^{(m)},\left(N_1^{(m)},\sigma_1^{(m)}\right)\right),...,\left(\epsilon_{l_m}^{(m)},\left(N_{l_m}^{(m)},\sigma_{l_m}^{(m)}\right)\right)\right\}$$

is obtained by a series of experiments, in which for a strain amplitude $\epsilon_i^{(m)}$ the number of cycles $N_i^{(m)}$ until failure and the corresponding stress amplitude $\sigma_i^{(m)}$ is determined. We have available a database of 132 materials with 1222 of the above data points in total. This data will be used to compare the estimated 5%–quantiles of the number of cycles until failure from the modified and the base material of ZStE500 for different strain amplitudes ϵ. Since these 5%–quantiles are equal to the lower bounds of the one-sided 95%–level confidence intervals, we actually estimate the number of cycles such that no failure occurs with a probability of approximately 95%. However, since the above mentioned experiments are very time consuming, we only have available 4 to 35 data points per material, which is not enough for a nonparametric estimation. In order to nevertheless estimate the quantile of the number of cycles until failure, we assume that the model

$$N^{(m)}(\epsilon)=\mu^{(m)}(\epsilon)+\sigma^{(m)}(\epsilon)\cdot\delta^{(m)} \tag{17}$$

holds, where $\mu^{(m)}(\epsilon)$ is the expected number of cycles until failure, $\sigma^{(m)}(\epsilon)$ is the standard deviation for each material m and strain amplitude ϵ and where the error term $\delta^{(m)}$ has expectation zero for each material m. In the following we will estimate the $\alpha-$quantile of $\delta^{(m)}$ as well as $\mu^{(m)}(\epsilon)$ and $\sigma^{(m)}(\epsilon)$, so that we get an estimate of the $\alpha-$quantile of $N^{(m)}(\epsilon)$ by a simple linear transformation. For this purpose we use a similar approach as in Bott and Kohler (2015):

In order to obtain an estimate $\hat{\mu}^{(m)}(\epsilon)$ of the expected number of cycles $\mu^{(m)}(\epsilon)$, we apply a standard-method from the literatur (cf. Williams, Lee and Rilly (2002)), which uses the measured data to estimate the coefficients $p=\left(\sigma_f^{'},\epsilon_f^{'},b,c\right)$ of the strain life curve according to Coffin-Morrow-Manson (cf. Manson (1965)) by linear regression and estimate $\mu^{(m)}(\epsilon)$ from the corresponding strain life curve.

The estimation of the standard deviation $\sigma^{(m)}(\epsilon)$ is more complicated, since we need to apply a nonparametric estimator to the squared deviations $Y_i^{(m)}=\left(N_i^{(m)}-\hat{\mu}_i^{(m)}\right)^2$ $(i=1,\dots,l_m)$ for each material m, which usually needs more samples. So we augment our data points per material m by 100 artifical ones as in Furer and Kohler (2013):

At first, we interpolate the squared deviations $Y_i^{(k)}$ for each material $k\neq m$ on a grid of 100 equidistant strain amplitudes ϵ. In order to genereate an artifical data point at a fixed grid point, we only want to use interpolated values from materials, that are similar to the material m, because we can assume that similar materials yield similar fatigue behaviour. This simularity is measured using 5 static material properties, namely Young's modulus, the yield limit for 0.2% residual elongation, the tensile strength, the static strength coefficient and the static strain hardening exponent. We apply the Nadaraya-Watson kernel regression estimates to the static material properties as covariate and the interpolated data as dependent variable to obtain the 100 artifical data points (one at each grid point) per material m. Finally, the estimation of the standard deviation

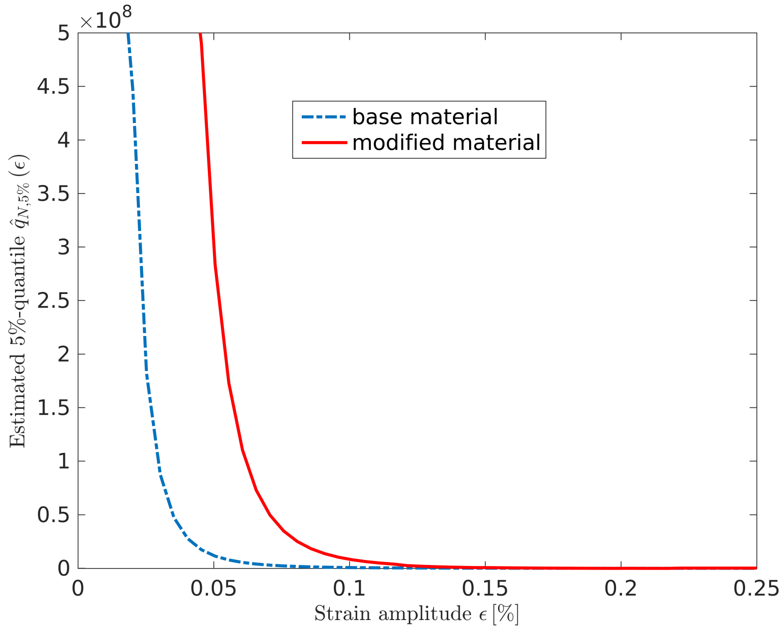

Figure 2: Comparison of the estimated 5%−quantiles of the number of cycles until the failure occurs $\hat{q}_{N,5\%}$ from the base and the modified material of ZSTE500.

$\sigma^{(m)}(\epsilon)$ is done by weighting the Nadaraya-Watson kernel regression estimates applied to the real and the artificial data of the squared deviations as dependent variable and the corresponding ϵ-values as covariate.

Thus, we can finally determine the data samples

$$\hat{\delta}_i^{(m)} = \frac{N_i^{(m)} - \hat{\mu}_i^{(m)}}{\hat{\sigma}_i^{(m)}} \quad \text{for } i = 1, ..., l_m$$

of the random variables $\delta^{(m)}$ for each material m. Notice that these samples contain measurement errors because we only estimated $\mu^{(m)}(\epsilon)$ and $\sigma^{(m)}(\epsilon)$. Since we only have available 4 to 35 of the above data samples per material, we will use data samples from other materials, that have similar static material properties (with the same justification as above).

Analogously to the estimation of the standard deviation, the consideration of similar materials in the quantile estimation is done by using the plug-in quantile estimate corresponding to the kernel estimate of the codf. with the static material properties as covariate X_i and the data samples of $\delta^{(m)}$ as the dependent variable. Evaluating this quantile estimate at the static material properties $x = X^{(m)}$ of some material m leads to an estimate $\hat{q}_{\delta^{(m)},\alpha}$, which can be transformed to an estimate of the α–quantile of $N^{(m)}(\epsilon)$ by

$$\hat{q}_{N^{(m)},\alpha}(\epsilon) = \hat{\sigma}^{(m)}(\epsilon) \cdot \hat{q}_{\delta^{(m)},\alpha} + \hat{\mu}^{(m)}(\epsilon) \,.$$

The estimated quantiles of $N^{(m)}(\epsilon)$ for $\epsilon \in [0, 0.25]$ for the modified and the base material are illustrated in Figure 2, where the strain amplitude ϵ is divided by the length of the material sample in the experiments. Here the material shows much better fatigue behaviour after the flow splitting, which confirms the conjecture that the strain hardening occuring during the flow splitting improves the fatigue behaviour of materials.

4. Proofs

In two of the proofs in this section we use the following lemma, which relates the plug-in estimate with data containing additional measurement error to plug-in estimates with i.i.d. data without additional measurement error.

Lemma 1. *Let $A > 0$ be a random variable, $x \in \mathbb{R}$ be arbitrary and $W_{n,i}(x) \geq 0$. Set*

$$\delta_n(x) = \sum_{i=1}^{n} I_{\left\{\left|Y_i - \bar{Y}_{i,n}\right| > A\right\}} \cdot W_{n,i}(x).$$

Then it holds for $\alpha \in (0,1)$ and the plug-in estimates defined in (2) and (7) that

$$\hat{q}_{Y,n,\alpha-\delta_n(x)}(x) - A \leq \hat{q}_{\bar{Y},n,\alpha}(x) \leq \hat{q}_{Y,n,\alpha+\delta_n(x)}(x) + A.$$

Proof. Consider

$$\bar{F}_n(y,x) - F_n(y+A,x) = \sum_{i=1}^{n} \left(I_{\left\{\bar{Y}_{i,n} \leq y\right\}} - I_{\{Y_i \leq y+A\}} \right) \cdot W_{n,i}(x).$$

The term $I_{\left\{\bar{Y}_{i,n} \leq y\right\}} - I_{\{Y_i \leq y+A\}}$ becomes one, if

$$\bar{Y}_{i,n} \leq y \quad \text{and } Y_i > y + A.$$

In this case $\left|Y_i - \bar{Y}_{i,n}\right| > A$ also holds true. Since $W_{n,i}(x)$ is nonnegative, we can conclude

$$\bar{F}_n(y,x) - F_n(y+A,x) \leq \sum_{i=1}^{n} I_{\left\{\left|Y_i - \bar{Y}_{i,n}\right| > A\right\}} \cdot W_{n,i}(x) = \delta_n(x).$$

Analogously we can show

$$\bar{F}_n(y,x) - F_n(y-A,x) \geq -\sum_{i=1}^{n} I_{\left\{\left|Y_i - \bar{Y}_{i,n}\right| > A\right\}} \cdot W_{n,i}(x) = -\delta_n(x).$$

Hence we get

$$\begin{aligned}
\hat{q}_{\bar{Y},n,\alpha}(x) &= \min\left\{z \in \mathbb{R} : \bar{F}_n(z,x) \geq \alpha\right\} \\
&= \min\left\{z \in \mathbb{R} : \bar{F}_n(z,x) - F_n(z+A,x) + F_n(z+A,x) \geq \alpha\right\} \\
&\geq \min\left\{z \in \mathbb{R} : \delta_n(x) + F_n(z+A,x) \geq \alpha\right\} \\
&= \min\left\{z \in \mathbb{R} : F_n(z,x) \geq \alpha - \delta_n(x)\right\} - A \\
&= \hat{q}_{Y,n,\alpha-\delta_n(x)}(x) - A
\end{aligned}$$

and

$$\begin{aligned}\hat{q}_{\bar{Y},n,\alpha}(x) &= \min\left\{z \in \mathbb{R} : \bar{F}_n(z,x) \geq \alpha\right\} \\ &= \min\left\{z \in \mathbb{R} : \bar{F}_n(z,x) - F_n(z-A,x) + F_n(z-A,x) \geq \alpha\right\} \\ &\leq \min\left\{z \in \mathbb{R} : -\delta_n(x) + F_n(z-A,x) \geq \alpha\right\} \\ &= \min\left\{z \in \mathbb{R} : F_n(z,x) \geq \alpha + \delta_n(x)\right\} + A \\ &= \hat{q}_{Y,n,\alpha+\delta_n(x)}(x) + A,\end{aligned}$$

which yields the assertion. □

4.1. Proof of Theorem 1

Let $\alpha \in (0,1)$ be fixed. We divide the proof into three steps:
In the first step of the proof we show that if $\alpha_n(x)$ is a (possibly random) sequence with

$$\alpha_n(x) \to \alpha \quad a.s. \quad \text{for } \mathbf{P}_X\text{–almost every } x,$$

then

$$dist\left(\hat{q}_{Y,n,\alpha_n(x)}(x), Q_{Y,\alpha}(x)\right) \to 0 \quad a.s. \text{ for } \mathbf{P}_X\text{–almost every } x.$$

Therefore it suffices to show for $\mathbf{P}_X$–almost all $x \in \mathbb{R}^d$ and for all $\epsilon > 0$

(i) $\mathbf{P}\left(\hat{q}_{Y,n,\alpha_n(x)}(x) \leq q^{[low]}_{Y,\alpha}(x) - \epsilon \quad i.o.\right) = 0$, and

(ii) $\mathbf{P}\left(\hat{q}_{Y,n,\alpha_n(x)}(x) > q^{[up]}_{Y,\alpha}(x) + \epsilon \quad i.o.\right) = 0$,

where $i.o.$ means infinitely often. Next we will show (i). Therefore let $\epsilon > 0$ be arbitrary. By the definition of $q^{[low]}_{Y,\alpha}$, we have

$$F\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right) < \alpha.$$

Set

$$\rho_1(x) = \alpha - F\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right) > 0.$$

Since (8) holds, we have

$$F_n\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right) \to F\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right) \quad a.s. \quad \text{for } \mathbf{P}_X\text{–almost all } x.$$

Set

$$\begin{aligned}N(x) := \{&\alpha_n(x) \to \alpha \;\; (n \to \infty) \qquad \text{and} \\ &F_n\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right) \to F\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right) \;\; (n \to \infty)\Big\},\end{aligned}$$

such that $\mathbf{P}(N(x)) = 1$ for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$. So for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$ it holds (ω-wise, for all $\omega \in N(x)$)

$$\left|F_n\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right) - F\left(q^{[low]}_{Y,\alpha}(x) - \epsilon, x\right)\right| < \frac{\rho_1(x)}{2} \quad \text{and} \quad |\alpha_n(x) - \alpha| < \frac{\rho_1(x)}{2}$$

for all n large enough, which implies for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$

$$F_n\left(q_{Y,\alpha}^{[low]}(x)-\epsilon, x\right) < F\left(q_{Y,\alpha}^{[low]}(x)-\epsilon, x\right) + \frac{\rho_1(x)}{2} = \alpha - \frac{\rho_1(x)}{2} < \alpha_n(x)$$

for all n large enough by the definition of ρ_1. But then for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$

$$\hat{q}_{Y,n,\alpha_n(x)}(x) > q_{Y,\alpha}^{[low]}(x) - \epsilon$$

holds for all n large enough (ω-wise for all $\omega \in N(x)$) by the definition of $\hat{q}_{Y,n,\alpha_n(x)}$. We finally have shown

$$\mathbf{P}\left(\hat{q}_{Y,n,\alpha_n(x)}(x) \le q_{Y,\alpha}^{[low]}(x) - \epsilon \;\; i.o.\right) \le \mathbf{P}\left(N(x)^c\right) = 1 - \mathbf{P}\left(N(x)\right) = 0,$$

for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$, which was the assertion of part (i). Similarly, (ii) can be shown, which yields the assertion of the first step of the proof.
Let $\epsilon > 0$ be arbitrary and set

$$\delta_n(x) = \sum_{i=1}^{n} I_{\left\{\left|Y_i - \bar{Y}_{i,n}\right| > \epsilon\right\}} \cdot W_{n,i}(x).$$

In the second step of the proof we show

$$\delta_n(x) \to 0 \quad a.s. \text{ for } \mathbf{P}_X\text{–almost all } x.$$

Therefore observe

$$\delta_n(x) \le \sum_{i=1}^{n} \frac{\left|Y_i - \bar{Y}_{i,n}\right|}{\epsilon} \cdot W_{n,i}(x),$$

because the $W_{n,i}$ are nonnegative. So (A1) implies the assertion.
By Lemma 1 we know

$$\hat{q}_{Y,n,\alpha-\delta_n(x)}(x) - \epsilon \le \hat{q}_{\bar{Y},n,\alpha}(x) \le \hat{q}_{Y,n,\alpha+\delta_n(x)}(x) + \epsilon. \tag{18}$$

In the third step of the proof we finally show the assertion. By the second step, we know $\alpha - \delta_n(x) \to \alpha \;\; a.s.$ and $\alpha + \delta_n(x) \to \alpha \;\; a.s.$ for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$, so by choosing $\alpha_n(x) = \alpha - \delta_n(x)$ or $\alpha_n(x) = \alpha + \delta_n(x)$, resp., we can conclude by (18) and by the first step for arbitrary $\epsilon > 0$

$$\begin{aligned} &dist\left(\hat{q}_{\bar{Y},n,\alpha}(x), Q_{Y,\alpha}(x)\right) \\ &\le dist\left(\hat{q}_{Y,n,\alpha-\delta_n(x)}(x), Q_{Y,\alpha}(x)\right) + \epsilon + dist\left(\hat{q}_{Y,n,\alpha+\delta_n(x)}(x), Q_{Y,\alpha}(x)\right) + \epsilon \longrightarrow 2\cdot\epsilon \quad a.s. \end{aligned}$$

for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$. Since $\epsilon > 0$ was arbitrary this implies the assertion. □

4.2. Proof of Theorem 2

Let $X, Y, X_1, Y_1, X_2, Y_2, \ldots$ be independent and identically uniformly on $(0,1)$ distributed random variables and let $\alpha \in (0,1)$ be fixed. Then

$$q_{Y,\alpha}^{[up]}(x) = q_{Y,\alpha}^{[low]}(x) \quad \text{for } \mathbf{P}_X\text{–almost every } x. \tag{19}$$

Let (h_n) be a positive sequence that fullfills the assumptions of Corollary 1, i.e.

$$h_n \to 0 \quad \text{and} \quad n \cdot h_n / \log(n) \to \infty \quad (n \to \infty)$$

For $n \in \mathbb{N}$ there exist (uniquely determined) $m \in \mathbb{N}_0$ and $k \in \{0, 1, 2, 3, ..., 2^m - 1\}$, such that $n = 2^m + k$. Set

$$A_n = \left[\frac{2k+1}{2^{m+1}} - \frac{3}{2} \cdot h_n, \frac{2k+1}{2^{m+1}} + \frac{3}{2} \cdot h_n\right]$$

and

$$\bar{Y}_{i,n} = Y_i + I_{\{X_i \in A_n\}} \quad \text{for } i = 1, .., n \quad \text{and all } n \in \mathbb{N}.$$

Since

$$\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left|\bar{Y}_{i,n} - Y_i\right|\right\} = \mathbf{P}(X \in A_n),$$

we get for arbitrary $\epsilon > 0$, using Hoeffding's inequality (cf., e.g., Hoeffding (1963)),

$$\sum_{n=1}^{\infty}\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(\left|\bar{Y}_{i,n} - Y_i\right| - \mathbf{P}(X_i \in A_n)\right)\right| > \epsilon\right) \le \sum_{n=1}^{\infty} 2 \cdot \exp\left(-2n\epsilon^2\right) < \infty.$$

Together with the Lemma of Borel-Cantelli this implies

$$\frac{1}{n}\sum_{i=1}^{n}\left|\bar{Y}_{i,n} - Y_i\right| - \frac{1}{n}\sum_{i=1}^{n}\mathbf{P}(X_i \in A_n) \to 0 \quad a.s.,$$

and thus

$$\frac{1}{n}\sum_{i=1}^{n}\left|\bar{Y}_{i,n} - Y_i\right| \to 0 \quad a.s.,$$

because $\frac{1}{n}\sum_{i=1}^{n}\mathbf{P}(X_i \in A_n) = \mathbf{P}(X \in A_n) \le 3h_n \to 0$ as $n \to \infty$. For $n \in \mathbb{N}$ choose $m, k \in \mathbb{N}$ as above and set

$$B_n = \left[\frac{2k+1}{2^{m+1}} - \frac{h_n}{2}, \frac{2k+1}{2^{m+1}} + \frac{h_n}{2}\right]$$

which is of one third of the length of A_n. Now consider $x \in B_n$. Then

$$\left|x - \frac{2k+1}{2^{m+1}}\right| \le \frac{h_n}{2}.$$

If

$$|X_i - x| \leq h_n,$$

this yields

$$\left| X_i - \frac{2k+1}{2^{m+1}} \right| \leq |X_i - x| + \left| x - \frac{2k+1}{2^{m+1}} \right| \leq \frac{3}{2} \cdot h_n$$

and therefore $X_i \in A_n$. Thus, for any $x \in B_n$ we have

$$\bar{F}_n(y, x) = \frac{\sum_{i=1}^{n} I_{\left\{Y_i + I_{\{X_i \in A_n\}} \leq y\right\}} K\left(\frac{X_i - x}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)} = \frac{\sum_{i=1}^{n} I_{\{Y_i + 1 \leq y\}} K\left(\frac{X_i - x}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)} = F_n(y - 1, x)$$

and consequently,

$$\hat{q}_{\bar{Y},n,\alpha}(x) = \hat{q}_{Y,n,\alpha}(x) + 1.$$

Hence, it suffices to show, that

$$C = \{x \in \mathbb{R} : x \in B_n \;\; i.o.\}$$

(where *i.o.* means infinitely often) is not a set of $\mathbf{P}_X$-measure zero. Because then for every $x \in C$, we can find a subsequence n_k, such that

$$\hat{q}_{\bar{Y},n_k,\alpha}(x) = \hat{q}_{Y,n_k,\alpha}(x) + 1 \longrightarrow q_{Y,\alpha}^{[low]}(x) + 1 \neq q_{Y,\alpha}^{[low]}(x) \quad a.s.$$

holds by Corollary 1. In the following we will proof the stronger assertion

$$\mathbf{P}_X(C) = 1,$$

which is implied by the existence of an index M, such that

$$[0, 1] \subseteq \bigcup_{n=2^m}^{2^{m+1}-1} B_n \tag{20}$$

for all $m \geq M$. To proof this, we will show, that there exists an index M such that

(i) $\frac{2k+1}{2^{m+1}} - \frac{h_{2^m+k}}{2} \leq \frac{2k}{2^{m+1}}$ and

(ii) $\frac{2k+2}{2^{m+1}} \leq \frac{2k+1}{2^{m+1}} + \frac{h_{2^m+k}}{2}$

for all $m \geq M$ and all $k = 0, 1, \ldots, 2^m - 1$, because in this case B_{2^m+k} covers

$$\left[\frac{2k}{2^{m+1}}, \frac{2k+2}{2^{m+1}}\right]$$

for $k = 0, 1, \ldots, 2^m - 1$ and therefore (20) is fullfilled. Clearly (i) and (ii) are equivalent. So it suffices to prove (ii), which is equivalent to

$$1 + \frac{k}{2^m} \leq (2^m + k) \cdot h_{2^m+k}.$$

But since $1 + \frac{k}{2^m} \leq 2$ and $l \cdot h_l \to \infty$ as $l \to \infty$, this clearly has to be fullfilled for all m large enough. The proof is complete. $\square$

4.3. Proof of Theorem 3

Since the codf. $F(\cdot, x)$ is differentiable at $q_{Y,\alpha}^{[low]}(x)$ with derivative greater than zero for $\mathbf{P}_X$–almost every x, the upper and lower quantile are equal in this case. So for the sake of simplicity we just write $q_{Y,\alpha}(x)$.
In the first step of the proof we show that if $\alpha_n(x)$ is a (possibly random) sequence with

$$\alpha_n(x) \to^{\mathbf{P}} \alpha \quad \text{for } \mathbf{P}_X\text{–almost all } x,$$

it holds

$$\left|\hat{q}_{Y,n,\alpha_n(x)}(x) - q_{Y,\alpha}(x)\right| = O_{\mathbf{P}}\left(r_n + |\alpha_n(x) - \alpha|\right) \quad \text{for } \mathbf{P}_X\text{–almost all } x. \tag{21}$$

Therefore, denote by D the set of all $x \in \mathbb{R}^d$ such that (13) holds, $F(\cdot, x)$ is differentiable at $q_{Y,\alpha}(x)$ with derivative greater than zero and $\alpha_n(x) \to^{\mathbf{P}} \alpha$. Clearly $\mathbf{P}_X(D) = 1$. Now consider a fixed $x \in D$. Since $F(\cdot, x)$ is differentiable at $q_{Y,\alpha}(x)$ with derivative greater than zero, there exist finite constants $c_1 = c_1(x) > 0$ and $\zeta = \zeta(x) > 0$ with $\gamma(x) \geq \zeta$, such that

$$c_1 |q_{Y,\alpha}(x) - y| \leq |F(q_{Y,\alpha}(x), x) - F(y, x)| \tag{22}$$

for all y with $|q_{Y,\alpha}(x) - y| \leq \zeta$. Let $c \geq 1$ be arbitrary and set

$$E_n := \left\{ \frac{2c}{c_1} |\alpha_n(x) - \alpha| \leq \frac{\zeta}{2} \right\}.$$

and

$$G_n(c) := \left\{ \sup_{\left|y - q_{Y,\alpha}(x)\right| \leq \gamma(x)} |F(y, x) - F_n(y, x)| \leq c \cdot r_n \right\}.$$

The assumptions $\alpha_n(x) \to^{\mathbf{P}} \alpha$ and (13) imply

$$\lim_{n\to\infty} \mathbf{P}(E_n^c) = 0 \quad \text{and} \quad \lim_{c\to 0} \limsup_{n\to\infty} \mathbf{P}(G_n(c)^c) = 0.$$

Now choose $n_0(x) \in \mathbb{N}$, such that $0 < \frac{2c \cdot r_n}{c_1} \leq \frac{\zeta}{2}$ is fullfilled for all $n \geq n_0(x)$. Assume in the following that the events E_n and $G_n(c)$ hold and consider $n \geq n_0(x)$. Set

$$\theta_n = \theta_n(x) = 2c \cdot |\alpha_n(x) - \alpha| + 2c \cdot r_n.$$

The assumptions imply

$$0 < \frac{1}{c_1} \cdot \theta_n = \frac{2c}{c_1} \cdot |\alpha_n(x) - \alpha| + \frac{2c \cdot r_n}{c_1} \leq \frac{\zeta}{2} + \frac{\zeta}{2} = \zeta \leq \gamma(x),$$

so we can conclude by (22) and $F(q_{Y,\alpha}(x), x) = \alpha$

$$\theta_n = c_1 \left| q_{Y,\alpha}(x) - q_{Y,\alpha}(x) - \frac{1}{c_1}\theta_n \right| \leq \left| \alpha - F\left(q_{Y,\alpha}(x) + \frac{1}{c_1}\theta_n, x \right) \right| \tag{23}$$

and

$$\theta_n = c_1 \left| q_{Y,\alpha}(x) - q_{Y,\alpha}(x) + \frac{1}{c_1}\theta_n \right| \leq \left| \alpha - F\left(q_{Y,\alpha}(x) - \frac{1}{c_1}\theta_n, x \right) \right|. \tag{24}$$

Because $F(\cdot, x)$ is differentiable at $q_{Y,\alpha}(x)$ with derivative greater than zero, we know

$$F\left(q_{Y,\alpha}(x) - \frac{1}{c_1}\theta_n, x \right) < \alpha < F\left(q_{Y,\alpha}(x) + \frac{1}{c_1}\theta_n, x \right).$$

Thus (23) and (24) imply

$$F\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1}, x \right) \leq \alpha - \theta_n < \alpha - \frac{\theta_n}{2} < \alpha < \alpha + \frac{\theta_n}{2} < \alpha + \theta_n \leq F\left(q_{Y,\alpha}(x) + \frac{\theta_n}{c_1}, x \right). \tag{25}$$

Since the event $G_n(c)$ holds and $\frac{1}{c_1}\theta_n \leq \gamma(x)$, we know

$$F_n\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1}, x \right) - c \cdot r_n \leq F\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1}, x \right)$$

and

$$F\left(q_{Y,\alpha}(x) + \frac{\theta_n}{c_1}, x \right) \leq F_n\left(q_{Y,\alpha}(x) + \frac{\theta_n}{c_1}, x \right) + c \cdot r_n.$$

Combining this with (25) and the definition of θ_n leads to

$$F_n\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1}, x \right) < \alpha - \frac{\theta_n}{2} + c \cdot r_n = \alpha - c \cdot |\alpha_n(x) - \alpha|$$

and

$$\alpha + c \cdot |\alpha_n(x) - \alpha| = \alpha + \frac{\theta_n}{2} - c \cdot r_n < F_n\left(q_{Y,\alpha}(x) + \frac{\theta_n}{c_1}, x \right).$$

The assumption $c \geq 1$ implies

$$\alpha - c \cdot |\alpha_n(x) - \alpha| \leq \alpha_n(x) \leq \alpha + c \cdot |\alpha_n(x) - \alpha|$$

and thus

$$F_n\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1}, x \right) < \alpha_n(x) < F_n\left(q_{Y,\alpha}(x) + \frac{\theta_n}{c_1}, x \right).$$

So finally we have shown

$$\mathbf{P}\left(E_n \cap G_n(c)\right) \leq \mathbf{P}\left(F_n\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1}, x \right) < \alpha_n < F_n\left(q_{Y,\alpha}(x) + \frac{\theta_n}{c_1}, x \right) \right)$$

for all $c \geq 1$, which by the definition of $\hat{q}_{Y,n,\alpha_n(x)}(x)$ leads to

$$\begin{aligned}
&\lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(\left| \hat{q}_{Y,n,\alpha_n(x)} - q_{Y,\alpha}(x) \right| \leq \frac{1}{c_1}\theta_n \right) \\
&= \lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1} \leq \hat{q}_{Y,n,\alpha_n(x)} \leq q_{Y,\alpha}(x) + \frac{\theta_n}{c_1} \right)
\end{aligned}$$

$$
\begin{aligned}
&\geq \lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(F_n\left(q_{Y,\alpha}(x) - \frac{\theta_n}{c_1}, x\right) < \alpha_n(x) < F_n\left(q_{Y,\alpha}(x) + \frac{\theta_n}{c_1}, x\right)\right) \\
&\geq \lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(E_n \cap G_n(c)\right) \\
&\geq \lim_{c\to\infty} \limsup_{n\to\infty} \left(1 - \mathbf{P}\left(E_n^c\right) - \mathbf{P}\left(G_n(c)^c\right)\right) = 1.
\end{aligned}
$$

Since $x \in D$ was arbitrary, this yields the assertion.
Set

$$
\delta_n(x) = \sum_{i=1}^n I_{\left\{\left|Y_i - \bar{Y}_{i,n}\right| > \sqrt{\eta_n(x)}\right\}} \cdot W_{n,i}(x).
$$

In the second step of the proof we show

$$
\delta_n(x) \to^{\mathbf{P}} 0 \quad \text{for } \mathbf{P}_X\text{–almost all } x.
$$

Therefore observe

$$
\delta_n(x) \leq \sum_{i=1}^n \frac{\left|Y_i - \bar{Y}_{i,n}\right|}{\sqrt{\eta_n(x)}} \cdot W_{n,i}(x) = \frac{\eta_n(x)}{\sqrt{\eta_n(x)}} = \sqrt{\eta_n(x)}, \tag{26}
$$

which yields the assertion because of (A3). Combining Lemma 1 with (26) leads to

$$
\hat{q}_{Y,n,\alpha-\sqrt{\eta_n(x)}}(x) - \sqrt{\eta_n(x)} \leq \hat{q}_{\bar{Y},n,\alpha} \leq \hat{q}_{Y,n,\alpha+\sqrt{\eta_n(x)}}(x) + \sqrt{\eta_n(x)} \tag{27}
$$

for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$.
In the third step of the proof we finally show the assertion. By the first step we know

$$
\left|\hat{q}_{Y,n,\alpha-\sqrt{\eta_n(x)}}(x) - q_{Y,\alpha}(x)\right| = O_{\mathbf{P}}\left(r_n + \sqrt{\eta_n(x)}\right)
$$

and

$$
\left|\hat{q}_{Y,n,\alpha+\sqrt{\eta_n(x)}}(x) - q_{Y,\alpha}(x)\right| = O_{\mathbf{P}}\left(r_n + \sqrt{\eta_n(x)}\right).
$$

for $\mathbf{P}_X$–almost every x. By (27) we can conclude

$$
\begin{aligned}
&\left|\hat{q}_{\bar{Y},n,\alpha}(x) - q_{Y,\alpha}(x)\right| \\
&\leq \left|\hat{q}_{Y,n,\alpha-\sqrt{\eta_n(x)}}(x) - \sqrt{\eta_n(x)} - q_{Y,\alpha}(x)\right| + \left|\hat{q}_{Y,n,\alpha+\sqrt{\eta_n(x)}}(x) + \sqrt{\eta_n(x)} - q_{Y,\alpha}(x)\right| \\
&\leq \left|\hat{q}_{Y,n,\alpha-\sqrt{\eta_n(x)}}(x) - q_{Y,\alpha}(x)\right| + \left|\hat{q}_{Y,n,\alpha+\sqrt{\eta_n(x)}}(x) - q_{Y,\alpha}(x)\right| + 2\sqrt{\eta_n(x)},
\end{aligned}
$$

which completes the proof. □

4.4. Proof of Corollary 2

In order to prove Corollary 2, we need the following lemma, which extends Theorem 9.1. in Györfi et al. (2002) to kernel weighted sums. Therefore denote by S_{x,h_n} the closed ball around x with radius h_n and by $\mathcal{N}_1\left(\epsilon, \mathcal{G}, y_1^n\right)$ the minimal size of an ϵ-cover of $\mathcal{G}$ on

$y_1^n = (y_1, \ldots, y_n) \in \mathbb{R}^n$, which is defined as a finite collection of functions $g_1, \ldots, g_N : \mathbb{R}^d \to \mathbb{R}$ with the property that for every $g \in \mathcal{G}$ there is a $j = j(g) \in \{1, \ldots, N\}$, such that

$$\frac{1}{n} \sum_{i=1}^{n} |g(y_i) - g_j(y_i)| < \epsilon.$$

Lemma 2. *Let $x \in \mathbb{R}^d$ be arbitrary, $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed $\mathbb{R}^{d+1}$-valued random vectors, set $Y_1^n = (Y_1, \ldots, Y_n)$ and let $\mathcal{G}$ be a set of functions $g : \mathbb{R} \to [0,1]$. Furthermore let K be the naive kernel and let $\epsilon > 0$ be arbitrary. Then $n \cdot \mathbf{P}_X(S_{x,h_n}) \geq 8/\epsilon^2$ implies*

$$\mathbf{P}\left(\exists g \in \mathcal{G} : \left|\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right)\right\}\right| > \epsilon\right)$$
$$\leq \mathbf{E}\left\{8 \cdot \min\left[\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}, Y_1^n\right) \cdot \exp\left(-\frac{n \cdot \epsilon^2}{128 \cdot \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}\right), 1\right]\right\}.$$

Proof of Lemma 2. Since the proof is similar to the proof of Theorem 9.1. in Györfi et al. (2002), we give only an outline of the proof. A complete proof is available from the authors on request. Choose random vectors $(X_1', Y_1'), \ldots, (X_n', Y_n')$, such that $(X_1, Y_1), \ldots, (X_n, Y_n), (X_1', Y_1'), \ldots, (X_n', Y_n')$ are independent and identically distributed, and choose independent and uniformly over $\{-1, 1\}$ distributed random variables $U_1, \ldots, U_n$, which are independent of all previously introduced random variables. As in the proof of Theorem 9.1. in Györfi et al. (2002) we get

$$\mathbf{P}\left(\exists g \in \mathcal{G} : \left|\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right)\right\}\right| > \epsilon\right)$$
$$\leq 2 \cdot \mathbf{P}\left(\exists g \in \mathcal{G} : \left|\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right) - \frac{1}{n}\sum_{i=1}^{n} g(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right)\right| > \frac{\epsilon}{2}\right)$$
$$\leq 4 \cdot \mathbf{P}\left(\exists g \in \mathcal{G} : \left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right)\right| > \frac{\epsilon}{4}\right)$$
$$\leq 4 \cdot \mathbf{E}\left\{\min\left[\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}, Y_1^n\right)\right.\right.$$
$$\left.\left.\cdot \max_{g \in \mathcal{G}_{\frac{\epsilon}{8}}, (x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}} \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(x_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| > \frac{\epsilon}{8}\right), 1\right]\right\}.$$

Application of Hoeffding's inequality yields the assertion. □

Proof of Corollary 2. Choose an arbitrary $x \in \mathbb{R}^d$ such that (15) holds for some constants $C(x), \kappa_1(x), \kappa_2(x) > 0$. Set

$$F_n^E(y,x) = \frac{\mathbf{E}\left\{\sum\limits_{i=1}^{n} I_{\{Y_i \le y\}} \cdot K\left(\frac{X_i - x}{h_n}\right)\right\}}{\mathbf{E}\left\{\sum\limits_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\right\}} = \frac{\int\limits_{S_{x,h_n}} F(y,t)\,\mathbf{P}_X(\mathrm{d}t)}{\mathbf{P}_X(S_{x,h_n})}$$

if $\mathbf{P}_X(S_{x,h_n}) > 0$ and $F_n^E(y,x) = 0$ otherwise. In order to apply Theorem 3, it remains to show (13) for

$$r_n = \sqrt{\frac{\log(n)}{n \cdot h_n^d}} + h_n^p.$$

Therefore we choose $\gamma(x) = \kappa_1(x)$ and decompose the error into

$$\begin{aligned}
&\sup_{|y-q_{Y,\alpha}(x)| \le \gamma(x)} |F(y,x) - F_n(y,x)| \\
\le\ &\sup_{|y-q_{Y,\alpha}(x)| \le \gamma(x)} \left|F(y,x) - F_n^E(y,x)\right| + \sup_{|y-q_{Y,\alpha}(x)| \le \gamma(x)} \left| F_n^E(y,x) - \frac{\frac{1}{n}\sum\limits_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}{\mathbf{P}_X(S_{x,h_n})} \cdot F_n(y,x) \right| \\
&+ \sup_{|y-q_{Y,\alpha}(x)| \le \gamma(x)} |F_n(y,x)| \cdot \left| \frac{\frac{1}{n}\sum\limits_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}{\mathbf{P}_X(S_{x,h_n})} - 1 \right| \\
=:\ & J_{n,1}(x) + J_{n,2}(x) + J_{n,3}(x).
\end{aligned}$$

Hence it suffices to show, that we obtain the claimed rate of convergence for all three summands for $\mathbf{P}_X$–almost all x, which we will do in the following three steps.
In the first step of the proof we show

$$J_{n,1}(x) = O_{\mathbf{P}}(h_n^p) \quad \text{for } \mathbf{P}_X\text{–almost every } x.$$

Therefore we set

$$L_n := \left\{x \in \mathbb{R}^d : \mathbf{P}_X(S_{x,h_n}) > 0\right\},$$

rewrite $J_{n,1}(x)$ and use assumption (15), which is applicable since $h_n \le \kappa_2(x)$ for all n sufficient large, to obtain for $\mathbf{P}_X$–almost every x

$$\begin{aligned}
J_{n,1}(x) &= \sup_{|y-q_{Y,\alpha}(x)| \le \gamma(x)} \left\{ \left| \frac{\int\limits_{S_{x,h_n}} F(y,x) - F(y,t)\,\mathbf{P}_X(\mathrm{d}t)}{\mathbf{P}_X(S_{x,h_n})} \right| I_{\{x \in L_n\}} + F(y,x) \cdot I_{\{x \in L_n^c\}} \right\} \\
&\le \left| \frac{\int\limits_{S_{x,h_n}} C(x) \cdot ||x-t||^p \mathbf{P}_X(\mathrm{d}t)}{\mathbf{P}_X(S_{x,h_n})} \right| I_{\{x \in L_n\}} + I_{\{x \in L_n^c\}} \le C(x) \cdot h_n^p + I_{\{x \in L_n^c\}}
\end{aligned}$$

for all n large enough. Since $L_n^c \subseteq supp\left(\mathbf{P}_X\right)^c$, and $\mathbf{P}_X\left(supp\left(\mathbf{P}_X\right)\right) = 1$, we have

$$I_{\{x \in L_n^c\}} = 0 \quad \text{for } \mathbf{P}_X\text{–almost every } x. \tag{28}$$

which implies the assertion.
In the second step of the proof we show

$$J_{n,2}\left(x\right) = O_{\mathbf{P}}\left(\sqrt{\frac{\log\left(n\right)}{n \cdot h_n^d}}\right) \quad \text{for } \mathbf{P}_X\text{–almost every } x.$$

Therefore, we only need to consider $x \in L_n$, since L_n^c is a set of $\mathbf{P}_X$-measure zero by (28). Set

$$\mathcal{G} = \left\{g : \mathbb{R} \to [0,1] : \quad g\left(z\right) = I_{\{z \leq y\}} \; y \in \mathbb{R}\right\}$$

and observe

$$\begin{aligned}
&\mathbf{P}\left(J_{n,2}\left(x\right) > c \cdot \sqrt{\frac{\log\left(n\right)}{n \cdot h_n^d}}\right) \\
&\leq \mathbf{P}\left(\sup_{y \in \mathbb{R}}\left|\frac{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}{\mathbf{P}_X\left(S_{x,h_n}\right)} \cdot F_n\left(y,x\right) - F_n^E\left(y,x\right)\right| > c \cdot \sqrt{\frac{\log\left(n\right)}{n \cdot h_n^d}}\right) \\
&= \mathbf{P}\left(\sup_{g \in \mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n} g\left(Y_i\right) K\left(\frac{X_i - x}{h_n}\right) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g\left(Y_i\right) K\left(\frac{X_i - x}{h_n}\right)\right\}\right| > \epsilon_n\right),
\end{aligned} \tag{29}$$

where

$$\epsilon_n = c \cdot \mathbf{P}_X\left(S_{x,h_n}\right) \cdot \sqrt{\frac{\log n}{n \cdot h_n^d}}.$$

Define the event

$$Q_n := Q_n\left(x\right) = \left\{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \leq \frac{3}{2} \cdot \mathbf{P}_X\left(S_{x,h_n}\right)\right\}.$$

Lemma 24.6. in Györfi et al. (2002), which states

$$\limsup_{n \to \infty} \frac{h_n^d}{\mathbf{P}_X\left(S_{x,h_n}\right)} < \infty \quad \text{for } \mathbf{P}_X\text{–almost every } x, \tag{30}$$

implies that for every $c > 0$ and $\mathbf{P}_X$–almost every x

$$n \geq \frac{8}{\epsilon_n^2} \cdot \mathbf{P}_X\left(S_{x,h_n}\right) = \frac{8n \cdot h_n^d}{c^2 \cdot \log\left(n\right) \cdot \mathbf{P}_X\left(S_{x,h_n}\right)}$$

is fullfilled for all n large enough. Thus, we can apply Lemma 2 to bound the last probability in (29). Since $\mathcal{G}$ is a set of indicator functions I_A with $A \in \mathcal{A} = \{(-\infty, y] : y \in \mathbb{R}\}$, we have

$$\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}, y_1^n\right) \leq s(\mathcal{A}, n),$$

for every $y_1^n = (y_1, \dots, y_n)$, where $s(\mathcal{A}, n)$ denotes the n-th shatter coefficient of the set $\mathcal{A}$ (see Definition 9.5. in Györfi et al. (2002)). By Theorem 9.3. and Example 9.1. in Györfi et al. (2002), it follows

$$s(\mathcal{A}, n) \leq n + 1 \leq 2n.$$

Using the bound for $\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}, Y_1^n\right)$, we finally obtain for $\mathbf{P}_X$–almost every $x \in L_n$

$$\begin{aligned}
&\mathbf{P}\left(J_{n,2}(x) > c \cdot \sqrt{\frac{\log(n)}{n \cdot h_n^d}}\right) \\
&\leq \mathbf{E}\left\{8 \cdot \min\left[\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}, Y_1^n\right) \cdot \exp\left(-\frac{n \cdot c^2 \cdot \log(n) \cdot \mathbf{P}_X(S_{x,h_n})^2}{128 \cdot n \cdot h_n^d \cdot \frac{1}{n}\sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)}\right), 1\right]\right\} \\
&\leq 16n \cdot \mathbf{E}\left\{\exp\left(-\frac{c^2 \cdot \log(n) \cdot \mathbf{P}_X(S_{x,h_n})^2}{128 \cdot h_n^d \cdot \frac{3}{2} \cdot \mathbf{P}_X(S_{x,h_n})}\right) I_{Q_n}\right\} + 8 \cdot \mathbf{P}(Q_n^c) \\
&= 16n \cdot \exp\left(-\frac{c^2 \cdot \log(n) \cdot \mathbf{P}_X(S_{x,h_n})}{192 \cdot h_n^d}\right) + 8 \cdot \mathbf{P}(Q_n^c)
\end{aligned}$$

for all n large enough. Since (30) implies

$$\liminf_{n\to\infty} \frac{\mathbf{P}_X(S_{x,h_n})}{192 \cdot h_n^d} > 0,$$

the left term on the right hand side tends to zero as $n \to \infty$ for $c > 0$ large enough. Hence, it suffices to show for $\mathbf{P}_X$–almost every $x \in L_n$

$$\mathbf{P}(Q_n^c) \to 0 \quad (n \to \infty).$$

Therefore observe that Chebyshev's inequality and the independence of $X_1, \dots X_n$ implies

$$\begin{aligned}
\mathbf{P}(Q_n^c) &\leq \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) - \mathbf{P}_X(S_{x,h_n})\right| > \frac{1}{2} \cdot \mathbf{P}_X(S_{x,h_n})\right) \qquad (31) \\
&\leq \frac{\mathbf{V}\left\{\frac{1}{n}\sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)\right\}}{\frac{1}{4} \cdot \mathbf{P}_X(S_{x,h_n})^2} \leq \frac{4 \cdot \mathbf{E}\left\{K\left(\frac{X_i - x}{h_n}\right)^2\right\}}{n \cdot \mathbf{P}_X(S_{x,h_n})^2} = \frac{4}{n \cdot h_n^d} \cdot \frac{h_n^d}{\mathbf{P}_X(S_{x,h_n})}
\end{aligned}$$

Thus, the assertion follows by (30) and (K1).
In the third step of the proof we show

$$J_{n,3}(x) = O_{\mathbf{P}}\left(\sqrt{\frac{\log(n)}{n \cdot h_n^d}}\right) \quad \text{for } \mathbf{P}_X\text{–almost every } x.$$

Therefore observe

$$
\begin{aligned}
J_{n,3}(x) &= \sup_{\left|y-q_{Y,\alpha}(x)\right|\leq\gamma(x)} \left| \frac{\sum_{i=1}^{n} I_{\{Y_i\leq y\}}\cdot K\left(\frac{X_i-x}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)} \right| \cdot \left| \frac{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)}{\mathbf{P}_X\left(S_{x,h_n}\right)} - 1 \right| \\
&\leq \left| \frac{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)}{\mathbf{P}_X\left(S_{x,h_n}\right)} - 1 \right|,
\end{aligned}
$$

since K is nonnegative. Using Chebyshev's inequality, we finally get for $x \in L_n$

$$
\begin{aligned}
\mathbf{P}\left(J_{n,3}(x) > c\cdot\sqrt{\frac{\log(n)}{n\cdot h_n^d}}\right) &\leq \mathbf{P}\left(\left| \frac{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)}{\mathbf{P}_X\left(S_{x,h_n}\right)} - 1 \right| > c\cdot\sqrt{\frac{\log(n)}{n\cdot h_n^d}}\right) \\
&\leq \mathbf{V}\left\{\frac{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)}{\mathbf{P}_X\left(S_{x,h_n}\right)}\right\}\cdot\frac{n\cdot h_n^d}{c^2\cdot\log(n)} \\
&\leq \frac{\mathbf{E}\left\{K\left(\frac{X_i-x}{h_n}\right)^2\right\}}{n\cdot\mathbf{P}_X\left(S_{x,h_n}\right)^2}\cdot\frac{n\cdot h_n^d}{c^2\cdot\log(n)} = \frac{h_n^d}{\mathbf{P}_X\left(S_{x,h_n}\right)}\cdot\frac{1}{c^2\cdot\log(n)},
\end{aligned}
$$

where we have used that K is the naive kernel. The assertion follows directly by the definition of $O_{\mathbf{P}}$ using (28) and (30). □

5. Acknowledgement

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 666.

References

[1] Bahadur, R. R. (1966). A note on quantiles in large samples. *Annals of Mathematical Statistics*, **37**, pp. 577-580.

[2] Bhattacharya, P. K. and Gangopadhyay, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics*, **18**, pp. 1400-1415.

[3] Bott, A. and Kohler, M. (2015). Nonparametric estimation of a conditional density. To appear in *Annals of the Institute of Statistical Mathematics*.

[4] Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory*, **18**, pp. 169-192.

[5] Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics*, **147**, pp. 120-130.

[6] Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics*, **19**, pp. 760-777.

[7] Christmann, A. and Steinwart, I. (2008). Consistency of kernel-based quantile regression. *Applied Stochastic Models in Business and Industry*, **24**, pp. 171-183.

[8] Cole, T. J. (1988). Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society: Series A*, **151**, pp. 385-418.

[9] Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: The lms method and penalized likelihood. *Statistics in Medicine*, **11**, pp. 1305-1319.

[10] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

[11] Furer, D., Kohler, M. und Krzyżak, A. (2013). Fixed design regression estimation based on real and artificial data. *Journal of Nonparametric Statistics*, **25**, pp. 223-241.

[12] Györfi, L. (1981). The rate of convergence of k_n-NN regression estimates and classification rules. *IEEE Transactions on Information Theory*, **27**, pp. 362-364.

[13] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression.* Springer, New York.

[14] Härdle, W., Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Annals of Statistics*, **16**, pp. 1428-1449.

[15] Hall, P., Wolff, R. C. L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, **94**, pp. 154-163.

[16] Hall, P. and Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Annals of Statistics*, **33**, pp. 1404-1421.

[17] Hansmann, M. and Kohler, M. (2016). Estimation of quantiles from data with additional measurement errors. *Submitted for publication.*

[18] He, X. and Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica*, **10**, pp. 129-140.

[19] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**, pp. 13-30.

[20] Hogg, R. V. (1975). Estimates of percentile regression lines using salary data. *Journal of the American Statistical Association*, **70**, pp. 56-59.

[21] Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, Cambridge.

[22] Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, **46**, pp. 33-50.

[23] Koenker, R. and Bassett, G. (1982). An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association*, **77**, pp. 407-415.

[24] Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for Arch models. *Econometric Theory*, **12**, pp. 793-813.

[25] Krzyżak, A. and Pawlak, M. (1987). The pointwise rate of convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference*, **16**, pp. 159-166.

[26] Lejeune, M. G. and Sarda, P. (1988). Quantile regression: a nonparametric approach. *Computational Statistics and Data Analysis*, **6**, pp. 229-239.

[27] Manson, S. S. (1965). Fatigue: A complex subject - some simple approximation. *Experimental Mechanics*, **5**, pp. 193-226.

[28] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics.* Lecture Notes in Statistics, **193**. Springer, Berlin.

[29] Nadaraya, E. A. (1964). On estimating regression. *Theory of probability and its Applications*, **9**, pp. 141-142.

[30] Portnoy, S. (1991). Asymptotic behavior of regression quantiles in non-stationary, dependent cases. *Journal of Multivariate Analysis*, **38**, pp. 100-113.

[31] Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, **32**, pp. 143-155.

[32] Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics and Probability Letters*, **7**, pp. 407-412.

[33] Schennach, S. M. (2008). Quantile regression with mismeasured covariates. *Econometric Theory*, **24**, pp. 1010-1043.

[34] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics*, **5**, pp. 595-620.

[35] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, **10**, pp. 1040-1053.

[36] Stute, W. (1986a). Conditional empirical processes. *Annals of Statistics*, **14**, pp. 638-647.

[37] Stute, W. (1986b). On almost sure convergence of conditional empirical distribution functions. *Annals of Probability*, **14**, pp. 891-901.

[38] Takeuchi, I., Le, Q. V., Sears, T. D. and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, **7**, pp. 1231-1264.

[39] Truong, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Annals of Statistics*, **17**, pp. 606-617.

[40] Veraverbeke, N., Gijbels, I. and Omelka, M. (2014). Preadjusted non-parametric estimation of a conditional distribution function. *Journal of the Royal Statistical Society: Series B*, **76**, pp. 399-438.

[41] Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A*, **26**, pp. 359-372.

[42] Wei, Y. and Carroll, R. J. (2009). Quantile regression with measurement error. *Journal of the American Statistical Association*, **104**, pp. 1129-1143.

[43] Williams, C. R., Lee, Y.-L. and Rilly, J. T. (2002). A practical method for statistical analysis of strain-life fatigue data. *International Journal of Fatigue*, **25**, pp. 427-436.

[44] Xiao, Z. and Koenker, R. (2009). Conditional quantil e estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, **104**, pp. 1696-1712.

[45] Yu, K. and Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics and Data Analysis*, **25**, pp. 159-166.

[46] Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, pp. 228-237.

[47] Yu, K., Lu, Z., and Stander, J. (2003). Quantile Regression: Applications and Current Research Areas. *Journal of the Royal Statistical Society: Series D*, **52**, pp. 331-350.

A. Supplementary material for the referees

A.1. Proof of Theorem 2: proof of (ii)

By the definition of $q_{Y,\alpha}^{[up]}$, we have

$$\alpha < F\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right).$$

Set

$$\rho_2(x) = F\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) - \alpha > 0.$$

Again, (8) implies

$$F_n\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) \to F\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) \quad a.s. \quad \text{for } \mathbf{P}_X\text{–almost all } x.$$

Set

$$\begin{aligned}\tilde{N}(x) := \{&\alpha_n(x) \to \alpha \ \ (n \to \infty) \quad \text{and} \\ &F_n\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) \to F\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) \ (n \to \infty)\Big\},\end{aligned}$$

such that $\mathbf{P}\left(\tilde{N}(x)\right) = 1$ for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$. So for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$ it holds (ω-wise for all $\omega \in \tilde{N}(x)$)

$$\left|F_n\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) - F\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right)\right| < \frac{\rho_2(x)}{2} \quad \text{and} \quad |\alpha_n(x) - \alpha| < \frac{\rho_2(x)}{2}$$

for all n large enough, which implies for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$

$$F_n\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) > F\left(q_{Y,\alpha}^{[up]}(x) + \epsilon, x\right) - \frac{\rho_2(x)}{2} = \alpha + \frac{\rho_2(x)}{2} > \alpha_n(x)$$

for all n large enough. But then for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$

$$\hat{q}_{Y,n,\alpha_n(x)}(x) \leq q_{Y,\alpha}^{[up]}(x) + \epsilon$$

holds (ω-wise for all $\omega \in \tilde{N}(x)$) for all n large enough by the definition of $\hat{q}_{Y,n,\alpha_n(x)}$. Finally we have shown

$$\mathbf{P}\left(\hat{q}_{Y,n,\alpha_n(x)}(x) > q_{Y,\alpha}^{[up]}(x) + \epsilon \ \ i.o.\right) \leq \mathbf{P}\left(\tilde{N}(x)^c\right) = 1 - \mathbf{P}\left(\tilde{N}(x)\right) = 0,$$

for $\mathbf{P}_X$–almost every $x \in \mathbb{R}^d$, which was the assertion of part (ii). $\square$

A.2. Proof of Lemma 2

In the following, we extend the arguments of the proof of Theorem 9.1. in Györfi et al. (2002).

Step 1: Symmetrization by a ghost sample.
Choose random vectors $(X_1', Y_1'), \ldots, (X_n', Y_n')$, such that $(X_1, Y_1), \ldots, (X_n, Y_n)$, $(X_1', Y_1'), \ldots, (X_n', Y_n')$ are independent and identically distributed. Let g^* be a function $g \in \mathcal{G}$, such that

$$\left| \frac{1}{n} \sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right) - \mathbf{E}\left\{\frac{1}{n} \sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right)\right\}\right| \ > \ \epsilon,$$

if there exists any such function, otherwise let g^* be an arbitrary function in $\mathcal{G}$. Set

$$\mathcal{D}_n = ((X_1, Y_1), \ldots, (X_n, Y_n))$$

and

$$\mathcal{D}_n' = \left((X_1', Y_1'), \ldots, (X_n', Y_n')\right).$$

Chebyshev's inequality and the independence of $(X_1', Y_1'), \ldots, (X_n', Y_n')$ yield

$$\mathbf{P}\left\{\left|\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right) \middle| \mathcal{D}_n\right\} - \frac{1}{n}\sum_{i=1}^{n} g^*(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right)\right| > \frac{\epsilon}{2} \,\middle|\, \mathcal{D}_n\right\}$$
$$\leq \frac{4}{\epsilon^2 \cdot n^2} \cdot \sum_{i=1}^{n} \mathbf{V}\left\{g^*(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right)\middle| \mathcal{D}_n\right\}$$
$$\leq \frac{4}{\epsilon^2 \cdot n^2} \cdot \sum_{i=1}^{n} \mathbf{E}\left\{\left(g^*\left(Y_i'\right)\right)^2 \cdot K\left(\frac{X_i' - x}{h_n}\right)^2 \middle| \mathcal{D}_n\right\} \leq \frac{4 \cdot \mathbf{P}_X\left(S_{x,h_n}\right)}{\epsilon^2 \cdot n},$$

where we have used the upper bound 1 of the functions $g \in \mathcal{G}$ and that K is the naive kernel. Thus, for $n \geq 8 \cdot \mathbf{P}_X\left(S_{x,h_n}\right)/\epsilon^2$, it follows

$$\mathbf{P}\left\{\left|\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right) \middle| \mathcal{D}_n\right\} - \frac{1}{n}\sum_{i=1}^{n} g^*(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right)\right| \leq \frac{\epsilon}{2} \,\middle|\, \mathcal{D}_n\right\}$$
$$\geq \frac{1}{2}.$$

Therefore, we can conclude for $n \geq 8 \cdot \mathbf{P}_X\left(S_{x,h_n}\right)/\epsilon^2$

$$\mathbf{P}\left(\exists g \in \mathcal{G} : \left|\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right) - \frac{1}{n}\sum_{i=1}^{n} g(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right)\right| > \frac{\epsilon}{2}\right)$$
$$\geq \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right) - \frac{1}{n}\sum_{i=1}^{n} g^*(Y_i') \cdot K\left(\frac{X_i' - x}{h_n}\right)\right| > \frac{\epsilon}{2}\right)$$

$$
\begin{aligned}
\geq & \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\middle|\mathcal{D}_n\right\}\right|>\epsilon,\right.\\
&\left.\left|\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)-\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\middle|\mathcal{D}_n\right\}\right|\leq\frac{\epsilon}{2}\right)\\
\geq & \mathbf{E}\left\{I_{\left\{\left|\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\middle|\mathcal{D}_n\right\}\right|>\epsilon\right\}}\right.\\
&\left.\cdot\,\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)-\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\middle|\mathcal{D}_n\right\}\right|\leq\frac{\epsilon}{2}\middle|\mathcal{D}_n\right)\right\}\\
\geq & \frac{1}{2}\cdot\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\middle|\mathcal{D}_n\right\}\right|>\epsilon\right)\\
= & \frac{1}{2}\cdot\mathbf{P}\left(\exists g\in\mathcal{G}:\ \left|\frac{1}{n}\sum_{i=1}^{n} g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)\right\}\right|>\epsilon\right)
\end{aligned}
$$

Finally, we obtain for $n \geq 8\cdot\mathbf{P}_X\left(S_{x,h_n}\right)/\epsilon^2$

$$
\begin{aligned}
&\mathbf{P}\left(\exists g\in\mathcal{G}:\ \left|\frac{1}{n}\sum_{i=1}^{n} g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-\mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)\right\}\right|>\epsilon\right)\\
&\leq 2\cdot\mathbf{P}\left(\exists g\in\mathcal{G}:\ \left|\frac{1}{n}\sum_{i=1}^{n} g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-\frac{1}{n}\sum_{i=1}^{n} g(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\right|>\frac{\epsilon}{2}\right).
\end{aligned}
$$

Step 2: Introduction of additional randomness by random signs
Let $U_1,\ldots,U_n$ be independent and uniformly over $\{-1,1\}$ distributed random variables, which are independent of $(X_1,Y_1),\ldots,(X_n,Y_n),(X_1',Y_1'),\ldots,(X_n',Y_n')$. Since $(X_1,Y_1),\ldots,(X_n,Y_n),(X_1',Y_1'),\ldots,(X_n',Y_n')$ are i.i.d., the joint distribution of $\mathcal{D}_n,\mathcal{D}_n'$ is not affected if one ranomly interchanges the corresponding components of $\mathcal{D}_n$ and $\mathcal{D}_n'$. Thus,

$$
\begin{aligned}
&\mathbf{P}\left(\exists g\in\mathcal{G}:\ \left|\frac{1}{n}\sum_{i=1}^{n}\left[g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-g(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\right]\right|>\frac{\epsilon}{2}\right)\\
&=\mathbf{P}\left(\exists g\in\mathcal{G}:\ \left|\frac{1}{n}\sum_{i=1}^{n}U_i\cdot\left[g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)-g(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\right]\right|>\frac{\epsilon}{2}\right)\\
&\leq\mathbf{P}\left(\exists g\in\mathcal{G}:\ \left|\frac{1}{n}\sum_{i=1}^{n}U_i\cdot g(Y_i)\cdot K\left(\frac{X_i-x}{h_n}\right)\right|>\frac{\epsilon}{4}\right)\\
&\quad+\mathbf{P}\left(\exists g\in\mathcal{G}:\ \left|\frac{1}{n}\sum_{i=1}^{n}U_i\cdot g(Y_i')\cdot K\left(\frac{X_i'-x}{h_n}\right)\right|>\frac{\epsilon}{4}\right) \qquad (32)
\end{aligned}
$$

$$=2 \cdot \mathbf{P}\left(\exists g \in \mathcal{G}: \left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(Y_i) \cdot K\left(\frac{X_i - x}{h_n}\right)\right| > \frac{\epsilon}{4}\right) \tag{33}$$

Step 3: Conditioning and introduction of a covering.
We condition the right probability in (33) on $\mathcal{D}_n$, which is equivalent to fixing $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^{d+1}$ and to considering

$$\mathbf{P}\left(\exists g \in \mathcal{G}: \left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| > \frac{\epsilon}{4}\right). \tag{34}$$

Fix $g \in \mathcal{G}$ and let $\mathcal{G}_{\frac{\epsilon}{8}}$ be an L_1 $\frac{\epsilon}{8}$-cover on $y_1^n = (y_1, \ldots, y_n)$ of minimal size. Then there exists $\bar{g} \in \mathcal{G}_{\frac{\epsilon}{8}}$ such that

$$\frac{1}{n}\sum_{i=1}^{n} |(g(y_i) - \bar{g}(y_i)| < \frac{\epsilon}{8}. \tag{35}$$

W.l.o.g. we assume that $0 \le \bar{g}(z) \le 1$, otherwise we truncate $\bar{g}$ at 0 and 1 and observe that (35) is still fullfilled in this case. Then

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} U_i\, g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| &\le \left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot \bar{g}(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} |U_i| \cdot |g(y_i) - \bar{g}(y_i)| \cdot \left|K\left(\frac{x_i - x}{h_n}\right)\right| \\ &\le \left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot \bar{g}(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| + \frac{\epsilon}{8}.\end{aligned}$$

Thus, taking into account that the size of $\mathcal{G}_{\frac{\epsilon}{8}}$ is $\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}, y_1^n\right)$, we finally obtain

$$\begin{aligned}&\mathbf{P}\left(\exists g \in \mathcal{G}: \left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| > \frac{\epsilon}{4}\right) \\ &\le \mathbf{P}\left(\exists g \in \mathcal{G}_{\frac{\epsilon}{8}}: \left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| + \frac{\epsilon}{8} > \frac{\epsilon}{4}\right) \\ &\le \min\left[\left|\mathcal{G}_{\frac{\epsilon}{8}}\right| \cdot \max_{g \in \mathcal{G}_{\frac{\epsilon}{8}}} \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| > \frac{\epsilon}{8}\right), 1\right] \\ &= \min\left[\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}, y_1^n\right) \cdot \max_{g \in \mathcal{G}_{\frac{\epsilon}{8}}} \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_i \cdot g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right)\right| > \frac{\epsilon}{8}\right), 1\right]\end{aligned}$$

Step 4: Application of Hoeffding's inequality.
Observing that $U_1 \cdot g(y_1) \cdot K\left(\frac{x_1 - x}{h_n}\right), \ldots, U_n \cdot g(y_n) \cdot K\left(\frac{x_n - x}{h_n}\right)$ are independent random variables with

$$-g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right) \le U_i \cdot g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right) \le g(y_i) \cdot K\left(\frac{x_i - x}{h_n}\right) \quad \text{for } i = 1, \ldots, n,$$

Hoeffding's inequality yields

$$\begin{aligned}
\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_i g(y_i)\cdot K\left(\frac{x_i-x}{h_n}\right)\right| > \frac{\epsilon}{8}\right) &\leq 2\cdot\exp\left(-\frac{2\cdot n\cdot\left(\frac{\epsilon}{8}\right)^2}{4\cdot\frac{1}{n}\sum\limits_{i=1}^{n} g(y_i)^2\cdot K\left(\frac{x_i-x}{h_n}\right)^2}\right)\\
&\leq 2\cdot\exp\left(-\frac{n\cdot\epsilon^2}{128\cdot\frac{1}{n}\sum\limits_{i=1}^{n} K\left(\frac{x_i-x}{h_n}\right)}\right),
\end{aligned}$$

where we have used the upper bound 1 of $g\in\mathcal{G}$ and that K is the naive kernel. The assertion follows by combining this with the results of the previous steps.