Estimation of quantiles from data with additional measurement errors *

Matthias Hansmann† and Michael Kohler

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: hansmann@mathematik.tu-darmstadt.de, kohler@mathematik.tu-darmstadt.de

December 3, 2015

Abstract

In this paper we study the problem of estimating quantiles from data that contains additional measurement errors. The only assumption on these errors is that the average absolute measurement error converges to zero for sample size tending to infinity with probability one. In particular we do not assume that the measurement errors are independent with expectation zero. We show that the empirical measure based on the data with measurement errors leads to an estimator which approaches the quantile set asymptotically. Provided the quantile is uniquely determined, this implies that this quantile estimate is strongly consistent for the true quantile. If this assumption does not hold, we also show that we can construct estimators for the limits of the quantile set if the average absolute measurement error is bounded by a given sequence, which tends to zero for sample size tending to infinity with probability one. Furthermore, we show that there exists no estimator that is consistent for every distribution of the underlying random variable and every data containing the above mentioned measurement errors. We also derive the rate of convergence of our estimator and show that the derived rate of convergence is optimal. The results are applied in simulations and in the context of experimental fatigue tests.

AMS classification: Primary 62G05; secondary 62G20.

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

1 Introduction

Let X be a real-valued random variable with cumulative distribution function (cdf.) F, i.e., $F(x) = \mathbf{P}\{X \leq x\}$. For $\alpha \in (0,1)$ denote by

$$Q_{X,\alpha} := \{z \in \mathbb{R} : \mathbf{P}(X \leq z) \geq \alpha \quad \text{and} \quad \mathbf{P}(X \geq z) \geq 1-\alpha\}$$

*Running title: *Quantile estimation with additional measurement errors*

†Corresponding author. Tel: +49-6151-16-23371

the set of all α-quantiles of X. It has the biggest lower bound

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

and the smallest upper bound

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

More precisely, we have

$$\left[q_{X,\alpha}^{[low]}, q_{X,\alpha}^{[up]}\right) \subseteq Q_{X,\alpha} \subseteq \left[q_{X,\alpha}^{[low]}, q_{X,\alpha}^{[up]}\right]$$

and

$$Q_{X_\alpha} = \left[q_{X,\alpha}^{[low]}, q_{X,\alpha}^{[up]}\right] \quad \text{if and only if } F \text{ is continuous at } q_{X,\alpha}^{[up]}.$$

The estimation of this set or its limits $q_{X,\alpha}^{[low]}$ and $q_{X,\alpha}^{[up]}$ is well researched in the literature. For example, a simple idea to estimate $q_{X,\alpha}^{[low]}$ from a sample $X_1, \ldots, X_n$ of X is to use $X_1, \ldots, X_n$ to compute the empirical cdf.

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}} \tag{1}$$

and to estimate the quantile by the corresponding plug-in estimate

$$\hat{q}_{X,n,\alpha} = \min\{z \in \mathbb{R} : F_n(z) \geq \alpha\}, \tag{2}$$

which is in fact an order statistics (c.f., e.g., Arnold et al. (1992) for general informations).

In this paper we assume that instead of the sample $X_1, \ldots, X_n$ of X we have available only data $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ such that the average absolute error between X_i and $\bar{X}_{i,n}$ converges to zero almost surely, i.e., we assume that

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_{i,n}| \to 0 \quad a.s. \tag{3}$$

Here we do not assume anything on the measurement errors $\bar{X}_{i,n} - X_i$ $(i = 1, \ldots, n)$. In general, those errors do not need 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, 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.

The consideration of additional measurement errors 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 other similar materials to estimate quantiles of number of cycles until failure for a certain material (cf., Section 3 below).

Measurement errors of the above mentioned type have been recently considered in the context of distribution estimation (c.f. Bott et al. (2013)), nonparametric regression with random design (c.f. Kohler (2006)) and nonparametric regression with fixed design (c.f. Furer, Kohler and Krzyżak (2013), Furer and Kohler (2015)).

Since we do not assume anything on the nature of the measurement errors besides that they are asymptotically negligible in the sense that (3) 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. In this paper we investigate whether the corresponding quantile estimates are still consistent in this situation and how their rate of convergence depends on

$$\frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_{i,n}|.$$

Before we describe the results with measurement error, we will in the following summarize some results of the quantile estimation with i.i.d. data without additional measurement errors. If the quantile is uniquely determined, i.e., if the cdf. F of X fullfills

$$F(x) > \alpha \text{ if } x > q_{X,\alpha}^{[low]},$$

then $\hat{q}_{X,n,\alpha}$ is a strongly consistent estimator of $q_{X,\alpha}^{[low]}$, i.e.,

$$\hat{q}_{X,n,\alpha} \to q_{X,\alpha}^{[low]} \quad a.s.$$

(c.f., e.g., Theorem 2.2. in Puri and Ralescu (1986)). In this paper we show that this result also holds if data with the above mentioned measurement error is used instead of the i.i.d. data (see Corollary 1 below).

In case that the quantile is not uniquely determined, $\hat{q}_{X,n,\alpha}$ is no longer a strong consistent estimate of $q_{X,\alpha}^{[low]}$ (c.f., e.g., Theorem 1 in Feldman and Tucker (1966)), but it is possible to find a suitable sequence α_n, such that $\hat{q}_{X,n,\alpha_n}$ is a strong (or weak) consistent estimator for $q_{X,\alpha}^{[low]}$ for all distributions of the random variable X (c.f. Theorem 4 (or 5) in Feldman and Tucker (1966)). If we use data with measurement errors for the quantile estimation, it will turn out in the following that it is not possible to find a sequence α_n such that $\hat{q}_{\bar{X},n,\alpha_n}$ is a strong consistent estimator of $q_{X,\alpha}^{[low]}$ for all distributions of X and all corresponding data with measurement error fullfilling (3). Surprisingly, there even does not exist any general estimator that is strongly consistent for all distributions of X and all corresponding data with measurement error fullfilling (3) (see Theorem 3 below for details). But if we know an upper bound on the average measurement error, which tends to zero almost surely for sample size tending to infinity, it is possible to find sequences α_n and β_n, such that $\hat{q}_{\bar{X},n,\alpha_n}$ and $\hat{q}_{\bar{X},n,\beta_n}$ are a strongly consistent estimators of $q_{X,\alpha}^{[low]}$ and $q_{X,\alpha}^{[up]}$, respectively (see Theorem 2 below).

The rate of convergence of quantile estimates in case that we do not have additional measurement errors can be derived from the asymptotic theory of order statistics (cf., e.g., Mosteller (1946), Smirnov (1952) and Bahadur (1966)). A well known asymptotic result is (c.f., e.g., Theorem A on page 77 in Serfling (1980)), that in case that the cdf. F of X is continuous and differentiable at $q_{X,\alpha}^{[low]}$ with derivative greater than zero we have

$$\sqrt{n} \cdot F'(q_{X,\alpha}^{[low]}) \cdot \frac{\hat{q}_{X,n,\alpha} - q_{X,\alpha}^{[low]}}{\sqrt{\alpha \cdot (1-\alpha)}} \to N(0,1) \quad \text{in distribution.} \tag{4}$$

Reiss (1974) also investigated the accuracy of this normal approximation. Since (4) holds, we have

$$|\hat{q}_{X,n,\alpha} - q_{X,\alpha}^{[low]}| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{5}$$

where we write $X_n = O_{\mathbf{P}}(Y_n)$ if the nonnegative random variables X_n and Y_n satisfy

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

In this paper we investigate how additional measurement errors influence the rate of convergence of our quantile estimates. In Theorem 4 below it is shown that if the average additional measurement error is bounded from above by some $\eta_n \geq 0$, then our estimate achieves a rate of convergence of order

$$\frac{1}{\sqrt{n}} + \sqrt{\eta_n}. \tag{6}$$

At this point it is surprising that additional measurement errors of order η_n increase the error of the estimate by $\sqrt{\eta_n}$ and not only by η_n. As we show in Theorem 5 below, it is in general not possible to derive a better rate of convergence, hence (6) is the optimal rate of convergence.

Throughout this paper the following notation is used: The sets of natural numbers and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$, respectively. For a real number x, we denote by $\lfloor x \rfloor$ and $\lceil x \rceil$ the largest integer less than or equal to x and the smallest integer larger than or equal to x, respectively. We write $\to^{\mathbf{P}}$ as an abbreviation for convergence in probability and I_A for the indicator function on the set 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 illustrate the finite sample size performance of our estimates by applying them to simulated data, and we describe an application of our estimates in the context of experimental fatigue tests.

2 Main results

Let

$$\bar{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{\bar{X}_{i,n} \leq x\}}$$

be the empirical cumulative distribution function corresponding to $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$, and let

$$\hat{q}_{\bar{X},n,\alpha} = \min\{z \in \mathbb{R} : \bar{F}_n(z) \geq \alpha\}$$

be the corresponding plug-in quantile estimate.

2.1 Strong consistency

First of all we want to investigate, if the estimator $\hat{q}_{\bar{X},n,\alpha}$ approaches the quantile set $Q_{X,\alpha}$ asymptotically. The following result holds.

Theorem 1. *Let $X, X_1, X_2 \ldots$ be independent and identically distributed real valued random variables and let $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ be random variables which satisfy*

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_{i,n}| \to 0 \quad a.s. \tag{7}$$

Let $\alpha \in (0,1)$ be arbitrary. Then the above defined quantile estimate $\hat{q}_{\bar{X},n,\alpha}$ is strongly consistent in the sense that

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

where

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

for $x \in \mathbb{R}$ and a set $A \subset \mathbb{R}$.

Corollary 1. *Let $X, X_1, X_2 \ldots$ be independent and identically distributed real valued random variables and let $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ be random variables which satisfy*

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_{i,n}| \to 0 \quad a.s. \tag{8}$$

Let $\alpha \in (0,1)$ be arbitrary and assume the α-quantile is uniquely determined, i.e., assume that the cdf. F of X fullfills

$$F\,(x) > \alpha \quad \text{ for } x > q_{X,\alpha}^{[low]}. \tag{9}$$

Then the above defined quantile estimate $\hat{q}_{\bar{X},n,\alpha}$ is strongly consistent for $q_{X,\alpha}^{[low]}$, i.e.

$$\hat{q}_{\bar{X},n,\alpha} \to q_{X,\alpha}^{[low]} \quad a.s.$$

Proof. By (9) we know

$$q_{X,\alpha}^{[up]} = q_{X,\alpha}^{[low]},$$

which implies

$$Q_{X,\alpha} = \{q_{X,\alpha}^{[low]}\}.$$

So the assertion follows directly by Theorem 1. □

Remark 1. The assumption in (9) in Corollary 1 is the minimal one for obtaining

$$\hat{q}_{\bar{X},n,\alpha} \to q_{X,\alpha}^{[low]} \quad a.s.$$

If we drop it, the case $q_{X,\alpha}^{[low]} < q_{X,\alpha}^{[up]}$ with

$$F(x) = \alpha \quad \text{for} \quad x \in \left[q_{X,\alpha}^{[low]}, q_{X,\alpha}^{[up]}\right)$$

is possible. In this case we get for i.i.d. data without measurement errors

$$\mathbf{P}\left(\hat{q}_{X,n,\alpha} \le q_{X,\alpha}^{[low]} \ i.o.\right) = \mathbf{P}\left(\hat{q}_{X,n,\alpha} \ge q_{X,\alpha}^{[up]} \ i.o.\right) = 1,$$

where i.o. means infinitely often (cf., e.g., Theorem 1 in Feldman and Tucker (1966)). This implies that

$$\hat{q}_{\bar{X},n,\alpha} \to q_{X,\alpha}^{[low]} \quad a.s.$$

cannot hold in this case.

Theorem 1 tells us under which conditions $\hat{q}_{\bar{X},n,\alpha}$ converges $a.s.$ towards the set Q_{X_α}. Estimating the lower bound $q_{X,\alpha}^{[low]}$ of this set by $\hat{q}_{\bar{X},n,\alpha}$ is only possible under a suitable condition on the cdf. F of X. As our next result shows, it is possible to drop this condition, if we replace α by an appropriate sequence α_n and if we know an upper bound η_n of the average absolute measurement error, which tends to zero almost surely as n tends to infinity. This approach extends the ideas of Theorem 4 in Feldman and Tucker (1966) to data, that contains additional measurement errors.

Theorem 2. *Let $X, X_1, X_2 \ldots$ be independent and identically distributed real valued random variables with cdf. F and let $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ be random variables which satisfy*

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_{i,n}| \le \eta_n \quad a.s. \tag{10}$$

for some $\eta_n \ge 0$ satisfying $\eta_n \to 0$ $a.s.$ Let $\alpha \in (0,1)$ be arbitrary. Set

$$\alpha_n = \alpha - 2\sqrt{\frac{2 \cdot \log(\log(n/2))}{n}} - \sqrt{\eta_n} \quad \text{and} \quad \beta_n = \alpha + 2\sqrt{\frac{2 \cdot \log(\log(n/2))}{n}} + \sqrt{\eta_n}.$$

Then

$$\hat{q}_{\bar{X},n,\alpha_n} \to q_{X,\alpha}^{[low]} \quad a.s.$$

and

$$\hat{q}_{\bar{X},n,\beta_n} \to q_{X,\alpha}^{[up]} \quad a.s.$$

Remark 2. It follows from the proof of Theorem 2 that the term $2\sqrt{\frac{2 \cdot \log(\log(n/2))}{n}}$ in the definition of the sequences α_n and β_n in Theorem 2 can be replaced by any c_n satisfying $c_n \to 0$ as $n \to \infty$ and

$$c_n \ge (1+\nu) \cdot \sqrt{\frac{2 \cdot \log(\log(n/2))}{n}}$$

for some $\nu > 0$.

The definition of the sequence α_n from Theorem 2 can change if the average measurement error changes. So it is natural to ask, if there exists a sequence α_n such that $\hat{q}_{\bar{X},n,\alpha_n}$ is a strong consistent estimator of $q_{X,\alpha}^{[low]}$ for all distributions of X and all random variables $\bar{X}_{1,n}, ..., \bar{X}_{n,n}$ satisfying

$$\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}_{i,n}\right| \to 0 \quad a.s. \tag{11}$$

The following result gives a general answer to this question, by stating that it is not possible to find a sequence of quantile estimates $(\hat{q}_{n,\alpha})_{n\in\mathbb{N}}$ that is strongly consistent for all distributions of X and all random variables $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ that fullfill (11), even if the sample with measurement errors does not change each time when the sample size changes, i.e., if we haven given data $\bar{X}_1, \ldots, \bar{X}_n$ instead of $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$.

Theorem 3. *Let $\alpha \in (0,1)$ be arbitrary. There does not exist a sequence $(\hat{q}_{n,\alpha})_{n\in\mathbb{N}}$ of quantile estimates satisfying*

$$\hat{q}_{n,\alpha}\left(\bar{X}_1, \ldots, \bar{X}_n\right) \to^{\mathbf{P}} q_{X,\alpha}^{[low]}$$

for all random variables X and all random variables $\bar{X}_1, \ldots, \bar{X}_n$ satisfying

$$\frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_i| \to 0 \quad a.s. \tag{12}$$

for some independent $X_1, X_2, \ldots$ that have the same distribution as X.

Remark 3. Analogously, it is possible to show that there does not exist a sequence $(\hat{q}_{n,\alpha})_{n\in\mathbb{N}}$ of quantile estimates satisfying

$$\hat{q}_{n,\alpha}\left(\bar{X}_1, \ldots, \bar{X}_n\right) \to^{\mathbf{P}} q_{X,\alpha}^{[up]}$$

for all random variables X and all random variables $\bar{X}_1, \ldots, \bar{X}_n$ satisfying

$$\frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_i| \to 0 \quad a.s. \tag{13}$$

for some independent $X_1, X_2, \ldots$ that have the same distribution as X.

2.2 Rate of convergence

In view of the finite sample size in any application it is also useful that we investigate the rate of convergence. The following result holds.

Theorem 4. *Let $X, X_1, X_2 \dots$ be independent and identically distributed real valued random variables with cdf. F and let $\bar{X}_{1,n}, \dots, \bar{X}_{n,n}$ be random variables which satisfy*

$$\eta_n := \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}_{i,n}| \to 0 \quad a.s. \tag{14}$$

Let $\alpha \in (0,1)$ be arbitrary and assume that cdf. F of X is continuous and differentiable at $q_{X,\alpha}^{[low]}$ with derivative greater than zero. Then

$$\left| \hat{q}_{\bar{X},n,\alpha} - q_{X,\alpha}^{[low]} \right| = O_{\mathbf{P}} \left(\frac{1}{\sqrt{n}} + \sqrt{\eta_n} \right).$$

The part $\frac{1}{\sqrt{n}}$ of the rate of convergence from Theorem 4 is well known from rate of convergence of the order statistics with i.i.d. data without errors (cf. (5)). Because of (4) it is not possible to improve this part of the convergence rate by an asymptotically faster decreasing sequence. It is also known that the order statistics is asymptotically most concentrated about the distribution quantile in comparison with all other translation-equivariant and asymptotically uniformly median unbiased estimators (cf. Corollary 2 in Pfanzagl (1976)).

The part $\sqrt{\eta_n}$ of the convergence rate is due to the measurement errors of the data. We now want to investigate, whether the rate $\sqrt{\eta_n}$ is the best rate one can obtain or if it is possible to find other estimators that have a faster convergence rate. The following result holds.

Theorem 5. *Let $\alpha \in (0,1)$ be arbitrary. Under the assumptions of Theorem 3, there does not exist an estimator, that achieves a better rate of convergence than $\sqrt{\eta_n}$. More precisely, under the assumptions of Theorem 4, for every estimator $\hat{q}_{n,\alpha}$ there exists a random variable X and random variables $\bar{X}_{1,n}, ..., \bar{X}_{n,n}$ satisfying*

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

for some independent $X_1, X_2, ...$ that have the same distribution as X such that

$$\left| \hat{q}_{n,\alpha} - q_{X,\alpha}^{[low]} \right| = O_{\mathbf{P}} \left(\frac{1}{\sqrt{n}} + \sqrt{\tilde{\eta}_n} \right)$$

does not hold, whenever $\tilde{\eta}_n$ is a sequence that fullfills

$$\frac{\tilde{\eta}_n}{\eta_n} \to^{\mathbf{P}} 0.$$

3 Application to simulated and real data

In this section we want to apply the above described methods to simulated and real data and estimate 5%-, 50%-, 90%- and 95%-quantiles. First of all we will consider distributions with known quantiles in order to classify our estimation and afterwards we will apply our estimator in the context of emperimental fatigue tests.

For the first purpose we use $n = 500, 1000$ and 2000 samples. This sample size is motivated by the above mentioned application in the context of experimental fatigue tests, where we have 1222 samples. In order to reduce the randomness that is contained in the quantile estimates due to the random number generation, we repeat the quantile estimation 100 times with new random numbers and indicate the quantile estimate by an upper index i. We will compare the quantile estimates by considering the average value $\frac{1}{100}\sum_{i=1}^{100} \hat{q}^i$ and the average squared error $\frac{1}{100}\sum_{i=1}^{100}\left(\hat{q}^i - q_{X,\alpha}^{[low]}\right)^2$.

As a first example we choose $X, X_1, X_2, \cdots$ as independent and identically $\mathcal{N}(0,1)$-distributed random variables and $\bar{X}_{i,n} = X_i + \frac{1}{n}E_i$, where $E_1, \ldots, E_n$ are samples from an exponentially-distributed random variable with expectation $\lambda = 10$. Notice that we get completly new samples, when n changes. As a comparison to that, we consider, furthermore, $\bar{Y}_{i,n} = X_i + \frac{1}{i} \cdot E_i$, where the samples with bigger measurement errors are kept by. We have $\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}_{i,n}\right| \to 0 \;\; a.s.$ and $\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{Y}_{i,n}\right| \to 0 \;\; a.s.$ Since the cdf. of X is strictly increasing, we know by Corollary 1 that the estimators $\hat{q}_{\bar{X},n,\alpha}$ and $\hat{q}_{\bar{Y},n,\alpha}$ are strongly consistent for $q_{X,\alpha}^{[low]}$. This is confirmed by the average values and average squared errors shown in Table 1 for $\alpha = 0.9$ and $\alpha = 0.95$. The estimator $\hat{q}_{\bar{X},n,\alpha}$ shows even for the small samples size of $n = 500$ estimates with a small average squared error. In comparison to that the estimator $\hat{q}_{\bar{Y},n,\alpha}$ convergences slower since the samples with the bigger measurement error are kept by.

As a second example, we choose $X, X_1, X_2, \ldots$ as independent and $b(1, \frac{1}{2})$-distributed random variables, such that $\mathbf{P}(X = 0) = \mathbf{P}(X = 1) = \frac{1}{2}$. Setting $\alpha = 0.5$ leads to the lower quantile $q_{X,\alpha}^{[low]} = 0$. The average value and the average squared error of the quantile estimate $\hat{q}_{X,n,\alpha}$ are shown in Table 2. As mentioned in Remark 1, the estimator $\hat{q}_{X,n,\alpha}$ is obviously not strongly consistent for $q_{X,\alpha}^{[low]}$. However, by Theorem 2 we can modify our estimate to $\hat{q}_{X,n,\alpha_n}$ with $\alpha_n = \alpha - 2\sqrt{\frac{2\cdot\log(\log(n/2))}{n}}$. As illustrated in Table 1, this modification leads to a perfect estimation of $q_{X,\alpha}^{[low]}$. But if we use the data $\bar{X}_{i,n} = X_i + \frac{B_i}{5\cdot n^{0.1}}$, where $B_1, ..., B_n$ are i.i.d. samples from a $b(1, \frac{1}{2})$-distributed random variable, the estimator $\hat{q}_{\bar{X},n,\alpha_n}$ shows much larger errors, which are illustrated in Table 2. But since we can bound $\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}_{i,n}\right|$ by $\frac{1}{5\cdot n^{0.1}}$, Theorem 2 tells us again, that we can get a consistent esimator, if we choose the sequence $\gamma_n = \alpha - 2\sqrt{\frac{2\cdot\log(\log(n/2))}{n}} - \sqrt{\frac{1}{5\cdot n^{0.1}}}$ and consider the estimator $\hat{q}_{\bar{X},n,\gamma_n}$. The results in Table 2 show that this estimator approximates the quantile perfectly.

As a third example we choose $X, X_1, X_2, \ldots$ as independent and uniformly on $(0,1)$-

	90%-quantile $q^{[low]}_{X,\alpha} = 1.2816$			95%-quantile $q^{[low]}_{X,\alpha} = 1.6449$		
size of n	500	1000	2000	500	1000	2000
average value for $q_{\bar{X},n,\alpha}$	1.2931	1.2901	1.2894	1.6730	1.6597	1.6473
average squared error for $q_{\bar{X},n,\alpha}$	0.0066	0.0030	0.0013	0.0128	0.0047	0.0023
average value for $q_{\bar{Y},n,\alpha}$	1.4217	1.3691	1.3275	1.8289	1.7507	1.7029
average squared error for $q_{\bar{Y},n,\alpha}$	0.0248	0.0105	0.0036	0.0436	0.0164	0.0057

Table 1: Simulation results for $X, X_1, X_2, ...$ independent and identically $\mathcal{N}(0,1)$-distributed and $\bar{X}_{i,n} = X_i + \frac{1}{n} \cdot E_i$ and $\bar{Y}_{i,n} = X_i + \frac{1}{i} \cdot E_i$, where $E_1, ..., E_n$ are samples from an exponentially-distributed random variable with expectation $\lambda = 10$.

	50%-quantile $q^{[low]}_{X,\alpha} = 0$		
size of n	500	1000	2000
average value for $q_{X,n,\alpha}$	0.4000	0.4400	0.4800
average squared error for $q_{X,n,\alpha}$	0.4000	0.4400	0.4800
average value for q_{X,n,α_n}	0	0	0
average squared error for q_{X,n,α_n}	0	0	0
average value for $q_{\bar{X},n,\alpha_n}$	0.1074	0.1002	0.0935
average squared error for $q_{\bar{X},n,\alpha_n}$	0.0115	0.0100	0.0087
average value for $q_{\bar{X},n,\gamma_n}$	0	0	0
average squared error for $q_{\bar{X},n,\gamma_n}$	0	0	0

Table 2: Simulation results for $X \sim b(1, \frac{1}{2})$ and $\bar{X}_{i,n} = X_i + \frac{B_i}{5 \cdot n^{0.1}}$, where $B_1, ..., B_n$ are i.i.d. samples from a $b(1, \frac{1}{2})$-distributed random variable.

distributed random variables. As our data with additional measurement error we consider $\bar{X}_{i,n} = X_i + \frac{1}{n^{0.25}}$, which fullfills $\eta_n = \frac{1}{n} \sum_{i=1}^{n} \left|X_i - \bar{X}_{i,n}\right| \to 0 \;\; a.s.$ In order to evaluate the asymptotic behaviour of our estimates, we compute the absolute error

$$d_n = \left|\hat{q}_{\bar{X},n,\alpha} - q_{X,\alpha}\right|$$

for $\alpha = 0.9$ and sample sizes n in steps of 200. As illustrated in Figure 1, the average absolute error shows approximately the same asymptotic behaviour as $\frac{1}{\sqrt{n}} + \eta_n$ in this case. This shows that there exists data with measurement error, such that a faster convergence rate than $\frac{1}{\sqrt{n}} + \sqrt{\eta_n}$ is obtained.

Furthermore, it is also possible to construct data with measurement error, such that the absolute error of the estimator behaves asymptotically as the claimed rate $\frac{1}{\sqrt{n}} + \sqrt{\eta_n}$ from Theorem 4:

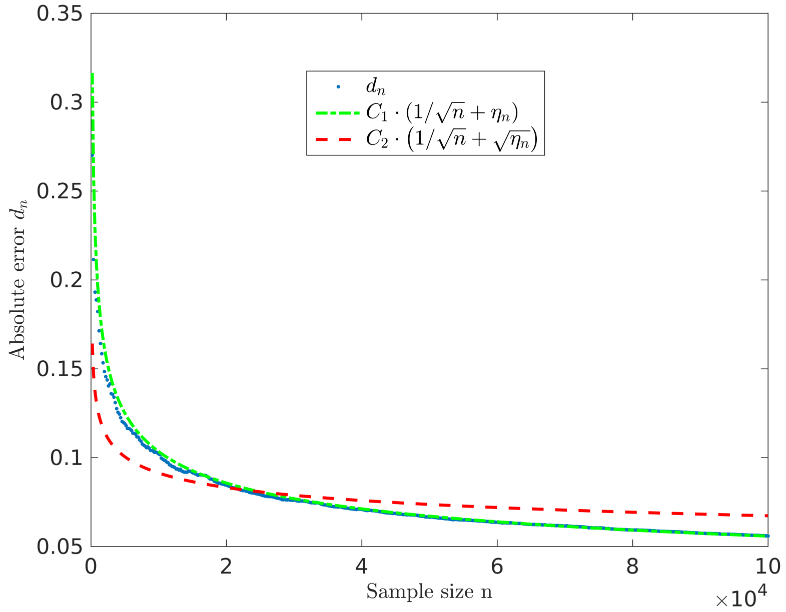

Figure 1: Typical asymptotic behaviour of $d_n = \left|\hat{q}_{\bar{X},n,\alpha} - q_{X,\alpha}\right|$ in the setting of the third example.

In the fourth example, we choose $\alpha = 0.9$ and $X, X_1, X_2, \dots$ as in example three and

$$\bar{X}_{i,n} = \begin{cases} X_i + \frac{1}{n^{0.25}} & \text{if} \quad X_i \in \left[\alpha - \frac{1}{n^{0.25}}, \alpha\right] \text{ and } X_i \text{ is one of the } \lfloor \frac{1}{n^{0.25}} \cdot n \rfloor \\ & \quad\quad \text{biggest samples of } (X_j)_{j=1,\dots,n} \text{ in } \left[\alpha - \frac{1}{n^{0.25}}, \alpha\right] \\ X_i & \text{else.} \end{cases}$$

Here

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n} \left|X_i - \bar{X}_{i,n}\right| \leq \frac{1}{n} \cdot \lfloor \frac{1}{n^{0.25}} \cdot n \rfloor \cdot \frac{1}{n^{0.25}} \to 0 \;\; a.s.$$

This leads to an absolute error d_n that has approximately the same asymptotic behaviour as $\frac{1}{\sqrt{n}} + \sqrt{\eta_n}$, as illustrated in Figure 2.

As a last example we want to apply the methods above in the context of fatigue behaviour of steel under cyclic loading. This application 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 produce structures out of one part by linear flow and bend splitting, which has several advantages concerning the material properties. In the following our main goal will be to study, whether this modified, splitted material shows better fatigue behavior under cyclic loading than the base material. Therefore for each material m data

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

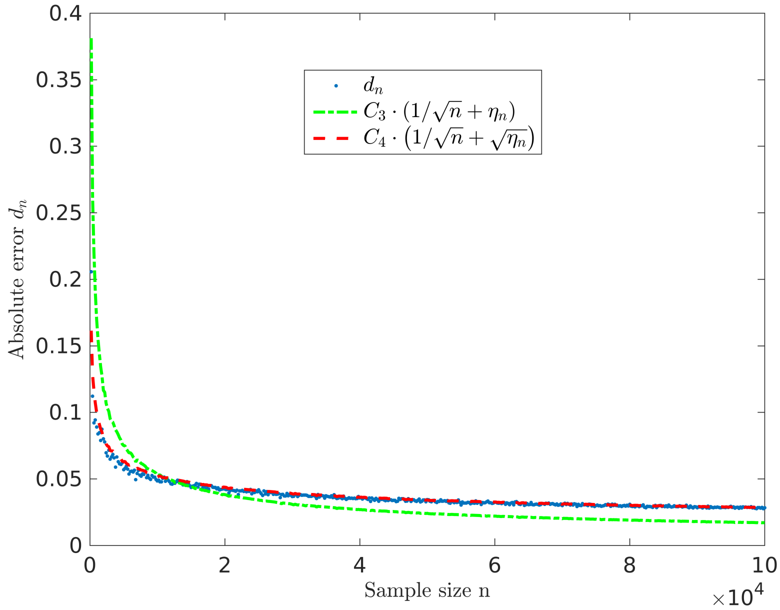

Figure 2: Typical asymptotic behaviour of $d_n = \left|\hat{q}_{\bar{X},n,\alpha} - q_{X,\alpha}\right|$ in the setting of the fourth example.

is obtained by a series of experiments, in which for a strain amplitude $\epsilon_i^{(m)}$ the number of cycles $N_i^{(m)}$ until the failure 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 amplitudes ϵ. In other words we are interested in estimating the number of cycles such that no failure occurs with a probability of 95%. 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 the model

$$N^{(m)}(\epsilon) = \mu^{(m)}(\epsilon) + \sigma^{(m)}(\epsilon) \cdot \delta \tag{15}$$

to hold, where $\mu^{(m)}(\epsilon)$ is the expected number of cycles until failure and $\sigma^{(m)}(\epsilon)$ is the standard deviation for each material m and strain amplitude ϵ. δ is an error term, which has expectation zero. In the following we will estimate the $\alpha-$quantile of δ as well as $\mu^{(m)}(\epsilon)$ and $\sigma^{(m)}(\epsilon)$, so that we can estimate 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 estimate 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 $\mu^{(m)}(\epsilon)$

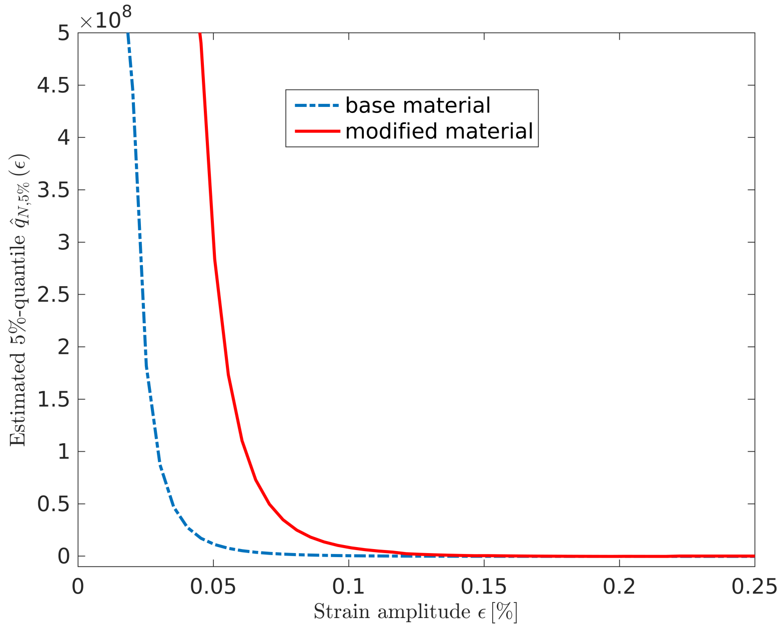

Figure 3: Comparison of the estimated 5%−quantiles of the number of cycles until the failure occurs $\hat{q}_{N,5\%}$ from the base and the modified material of ZSTE500. Here the strain amplitude is divided by the length of the material sample used in the experiments.

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, which usually needs more samples. So we augment our data points for every material m by 100 artifical ones like in Furer and Kohler (2013) and weight the Nadaraya-Watson kernel regression estimates applied to the real and the artificial data.

Thus, we can determine the data samples

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

of the random variable δ. Notice that these samples contain measurement errors because we only estimated $\mu^{(m)}(\epsilon)$ and $\sigma^{(m)}(\epsilon)$. Since we assumed in (15) that δ does not depend on the material m, we can use all above data samples to estimate the $\alpha-$quantile $\hat{q}_{\delta,\alpha}$ of δ and get an estimation of the $\alpha-$quantile of $N^{(m)}(\epsilon)$ by the transformation

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

The estimated quantiles of $N^{(m)}(\epsilon)$ for $\epsilon \in [0, 0.25]$ for the modified and the base material are illustrated in Figure 3. One can see that 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

For $\bar{\alpha} \in (0,1)$ set

$$\hat{q}_{X,n,\bar{\alpha}} = \min\{z \in \mathbb{R} : F_n(z) \geq \bar{\alpha}\},$$

where

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}}.$$

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

Lemma 1. *Let $a > 0$ be a (possibly random) finite constant and set*

$$\delta_n = \frac{1}{n}\sum_{i=1}^{n} I_{\left\{|X_i - \bar{X}_{i,n}| > a\right\}}.$$

Then it holds for $\alpha \in \mathbb{R}$ and the plug-in estimates defined above that

$$\hat{q}_{X,n,\alpha-\delta_n} - a \leq \hat{q}_{\bar{X},n,\alpha} \leq \hat{q}_{X,n,\alpha+\delta_n} + a$$

Proof. Consider

$$\bar{F}_n(x) - F_n(x+a) = \frac{1}{n}\sum_{i=1}^{n}\left(I_{\left\{\bar{X}_{i,n} \leq x\right\}} - I_{\{X_i \leq x+a\}}\right).$$

The i-th summand becomes one, if

$$\bar{X}_{i,n} \leq x \quad \text{ and } X_i > x + a.$$

In this case $\left|X_i - \bar{X}_{i,n}\right| > a$ also holds true. So we can conclude

$$\bar{F}_n(x) - F_n(x+a) \leq \frac{1}{n}\sum_{i=1}^{n} I_{\{|X_i - \bar{X}_{i,n}| > a\}} = \delta_n.$$

Analogously we can also show

$$\bar{F}_n(x) - F_n(x-a) \geq -\frac{1}{n}\sum_{i=1}^{n} I_{\{|X_i - \bar{X}_{i,n}| > a\}} = -\delta_n.$$

Hence we get

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

and

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

which yields the assertion. $\square$

4.1 Proof of Theorem 1

Let $\alpha_n \in (0,1)$ be such that

$$\alpha_n \to \alpha \quad a.s.$$

We divide the proof into three steps:
In the first step of the proof we show that

$$dist\left(\hat{q}_{X,n,\alpha_n}, Q_{X,\alpha}\right) \to 0 \quad a.s. \tag{16}$$

Therefore set

$$N := \left\{\alpha_n \to \alpha \ \left(n \to \infty\right) \quad \text{and} \quad \sup_{t \in \mathbb{R}} \left|F_n\left(t\right) - F\left(t\right)\right| \to 0 \ \left(n \to \infty\right)\right\}.$$

Notice that

$$\mathbf{P}\left(N\right) = 1$$

because of the Glivenko-Catelli theorem (cf., e.g., Theorem 12.4 in Devroye et al. (1996)) and $\alpha_n \to \alpha$ $a.s.$ Let $\epsilon > 0$ be arbitrary. We know

$$F\left(q_{X,\alpha}^{[low]} - \epsilon\right) < \alpha < F\left(q_{X,\alpha}^{[up]} + \epsilon\right). \tag{17}$$

Setting

$$\rho_1 = \min\left(\alpha - F\left(q_{X,\alpha}^{[low]} - \epsilon\right), F\left(q_{X,\alpha}^{[up]} + \epsilon\right) - \alpha\right),$$

we can conclude

$$F\left(q_{X,\alpha}^{[low]} - \epsilon\right) + \frac{\rho_1}{2} < \alpha < F\left(q_{X,\alpha}^{[up]} + \epsilon\right) - \frac{\rho_1}{2}.$$

Assume N to hold in the following. Then we can (for all $\omega \in N$) find n_0, such that for all $n \geq n_0$ we have

$$|\alpha_n - \alpha| < \frac{\rho_1}{4} \quad \text{and} \quad \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| < \frac{\rho_1}{4},$$

which implies

$$F_n\left(q_{X,\alpha}^{[low]} - \epsilon\right) < \alpha_n < F_n\left(q_{X,\alpha}^{[up]} + \epsilon\right)$$

and consequently

$$q_{X,\alpha}^{[low]} - \epsilon \leq \hat{q}_{X,n,\alpha_n} \leq q_{X,\alpha}^{[up]} + \epsilon.$$

Hence,

$$\mathbf{P}\left(\limsup_{n\to\infty} dist\left(\hat{q}_{X,n,\alpha_n}, Q_{X,\alpha}\right) \leq \epsilon\right) \geq \mathbf{P}(N) = 1.$$

Since $\epsilon >$ was arbitrary this implies the assertion.

Let $\epsilon > 0$ again be arbitrary and set

$$\delta_n = \frac{1}{n}\sum_{i=1}^{n} I_{\{|X_i - \bar{X}_{i,n}| > \epsilon\}}.$$

In the second step of the proof we show

$$\delta_n \to 0 \quad a.s. \tag{18}$$

Therefore we observe

$$\frac{1}{n}\sum_{i=1}^{n} I_{\{|X_i - \bar{X}_{i,n}| > \epsilon\}} \leq \frac{1}{\epsilon}\frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_{i,n}|,$$

which yields the assertion by (7).
Furthermore, we know by Lemma 1

$$\hat{q}_{X,n,\alpha-\delta_n} - \epsilon \leq \hat{q}_{\bar{X},n,\alpha} \leq \hat{q}_{X,n,\alpha+\delta_n} + \epsilon \tag{19}$$

In the third step of the proof we finally show the assertion. By the second step, we know $\alpha - \delta_n \to \alpha$ *a.s.* and $\alpha + \delta_n \to \alpha$ *a.s.*, so by choosing $\alpha_n = \alpha - \delta_n$ or $\alpha_n = \alpha + \delta_n$, resp., we conclude by (19) and by the first step for arbitrary $\epsilon > 0$

$$\begin{aligned} & dist\left(\hat{q}_{\bar{X},n,\alpha}, Q_{X,\alpha}\right) \\ & \leq dist\left(\hat{q}_{X,n,\alpha-\delta_n}, Q_{X,\alpha}\right) + \epsilon + dist\left(\hat{q}_{X,n,\alpha+\delta_n}, Q_{X,\alpha}\right) + \epsilon \longrightarrow 2\cdot\epsilon \quad a.s. \end{aligned} \tag{20}$$

Since $\epsilon > 0$ was arbitrary this implies the assertion. $\square$

4.2 Proof of Theorem 2

In order to proof Theorem 2, we need the following lemma, which is a straightforward extension of ideas in Theorem 4 in Feldman and Tucker (1966) to random sequences. For the sake of completeness, this lemma is proven in the appendix.

Lemma 2. *Let $\alpha \in (0,1)$ be arbitrary and $X, X_1, X_2, \dots$ be independent and identically distributed real valued random variables with cdf. F.*
(a) Let $\gamma_{n,l}$ be a (possibly random) sequence, that satisfies

$$\gamma_{n,l} + (1+\nu) \cdot \sqrt{\frac{2 \cdot \log(\log(n/2))}{n}} < \alpha \quad \textit{and} \quad \gamma_{n,l} \to \alpha \quad a.s.$$

for some $\nu > 0$. Then it holds

$$\hat{q}_{X,n,\gamma_{n,l}} \to q_{X,\alpha}^{[low]} \quad a.s. \tag{21}$$

(b) Let $\gamma_{n,r}$ be a (possibly random) sequence, that satisfies

$$\gamma_{n,r} - (1+\nu) \cdot \sqrt{\frac{2 \cdot \log(\log(n/2))}{n}} > \alpha \quad \textit{and} \quad \gamma_{n,l} \to \alpha \quad a.s.$$

for some $\nu > 0$. Then it holds

$$\hat{q}_{X,n,\gamma_{n,r}} \to q_{X,\alpha}^{[up]} \quad a.s. \tag{22}$$

Proof of Theorem 2. Set

$$\delta_n = \frac{1}{n} \sum_{i=1}^{n} I_{\{|X_i - \bar{X}_{i,n}| > \sqrt{\eta_n}\}}$$

and observe that (10) implies

$$\delta_n = \frac{1}{n} \sum_{i=1}^{n} I_{\{|X_i - \bar{X}_{i,n}| > \sqrt{\eta_n}\}} \leq \frac{1}{\sqrt{\eta_n}} \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}_{i,n}| \leq \frac{\eta_n}{\sqrt{\eta_n}} = \sqrt{\eta_n} \quad a.s. \tag{23}$$

Using Lemma 1 and (23), we can conclude that for any (random) sequence γ_n holds

$$\hat{q}_{X,n,\gamma_n - \sqrt{\eta_n}} - \sqrt{\eta_n} \leq \hat{q}_{\bar{X},n,\gamma_n} \leq \hat{q}_{X,n,\gamma_n + \sqrt{\eta_n}} + \sqrt{\eta_n} \tag{24}$$

for every $n \in \mathbb{N}$. By setting $\gamma_n = \alpha_n$ in (24) we know

$$\hat{q}_{X,n,\alpha_n - \sqrt{\eta_n}} - \sqrt{\eta_n} \leq \hat{q}_{\bar{X},n,\alpha_n} \leq \hat{q}_{X,n,\alpha_n + \sqrt{\eta_n}} + \sqrt{\eta_n} \tag{25}$$

for all $n \in \mathbb{N}$. Having regard to

$$\alpha_n + (1+\nu) \cdot \sqrt{\frac{2 \cdot \log(\log(n/2))}{n}} + \sqrt{\eta_n} < \alpha$$

for all $0 < \nu < 1$, as well as $\alpha_n \to \alpha$ *a.s.*, we also know that $\gamma_{n,l} = \alpha_n + \sqrt{\eta_n}$ and $\gamma_{n,l} = \alpha_n - \sqrt{\eta_n}$ fullfill the assumptions of Lemma 2a). So we get

$$\hat{q}_{X,n,\alpha_n-\sqrt{\eta_n}} - \sqrt{\eta_n} \to q_{X,\alpha}^{[low]} \quad a.s. \quad \text{and} \quad \hat{q}_{X,n,\alpha_n+\sqrt{\eta_n}} + \sqrt{\eta_n} \to q_{X,\alpha}^{[low]} \quad a.s.,$$

which yields

$$\hat{q}_{\bar{X},n,\alpha_n} \to q_{X,\alpha}^{[low]} \quad a.s.$$

Analogously we can show

$$\hat{q}_{\bar{X},n,\beta_n} \to q_{X,\alpha}^{[up]} \quad a.s.$$

by using Lemma 2b), which completes the proof. □

4.3 Proof of Theorem 3

Let $\alpha \in (0,1)$ be arbitrary. Assume to the contrary that there exists a sequence $(\hat{q}_{n,\alpha})_{n\in\mathbb{N}}$ of quantile estimates statisfying

$$\hat{q}_{n,\alpha}\left(\bar{X}_1, \ldots, \bar{X}_n\right) \to^{\mathbf{P}} q_{X,\alpha}^{[low]} \tag{26}$$

whenever $\bar{X}_1, \bar{X}_2, \ldots$ are such that for some independent and identically as X distributed $X_1, X_2, \ldots$ we have

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_i| \to 0 \quad a.s. \tag{27}$$

Let $X, X_1, X_2, \ldots$ be independent and indentically distributed with cdf.

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \le x < \alpha \\ \alpha & \text{if} \quad \alpha \le x < 1+\alpha \\ x-1 & \text{if} \quad 1+\alpha \le x < 2 \\ 1 & \text{if} \quad 2 \le x \end{cases}$$

and α-quantile $q_{X,\alpha}^{[low]} = \alpha$. For $k \in \mathbb{N}$ set

$$F_k(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \le x < \alpha - \frac{\alpha}{k} \\ \alpha - \frac{\alpha}{k} & \text{if} \quad \alpha - \frac{\alpha}{k} \le x < 1+\alpha - \frac{\alpha}{k} \\ x-1 & \text{if} \quad 1+\alpha - \frac{\alpha}{k} \le x < 2 \\ 1 & \text{if} \quad 2 \le x \end{cases}$$

and

$$X_i^{(k)} = \begin{cases} X_i & \text{if} \quad X_i \notin \left[\alpha - \frac{\alpha}{k}, \alpha\right] \\ X_i + 1 & \text{if} \quad X_i \in \left[\alpha - \frac{\alpha}{k}, \alpha\right] \end{cases} .$$

Then $X_1^{(k)}, X_2^{(k)}, \dots$ are independent and identically distributed random variables with cdf. F_k and α-quantile $q_{k,\alpha}^{[low]} = 1 + \alpha$. So if we set $\bar{X}_i = X_i^{(k)}$ for all $i \geq N$ with $N \in \mathbb{N}$ arbitrary, (27) is fullfilled (with X_i replaced by $X_i^{(k)}$) and we know by (26) that

$$\hat{q}_{n,\alpha}\left(\bar{X}_1, \dots, \bar{X}_n\right) \to^{\mathbf{P}} q_{k,\alpha}^{[low]} \tag{28}$$

Next we define for suitably chosen deterministic $n_0 := 0 < n_1 < n_2 < \dots$ (where $n_i \in \mathbb{N}$ for all $i \in N$) our data with measurement error by

$$\bar{X}_i = X_i^{(k)} \quad \text{if } n_{k-1} < i \leq n_k \quad (k \in \mathbb{N})\,.$$

For all $i \in \mathbb{N}$ we have

$$\mathbf{P}\left(|X_i - \bar{X}_i| = 0\right) \geq 1 - \alpha \quad \text{and} \quad \mathbf{P}\left(|X_i - \bar{X}_i| = 1\right) \leq \alpha$$

and hence

$$0 \leq \mathbf{E}\left\{|X_i - \bar{X}_i|\right\} \leq \alpha \quad \text{and} \quad \mathbf{V}\left\{|X_i - \bar{X}_i|\right\} \leq \mathbf{E}\left\{|X_i - \bar{X}_i|^2\right\} \leq \alpha.$$

So

$$\sum_{i=1}^{\infty} \frac{\mathbf{V}\{|X_i - \bar{X}_i|\}}{i^2} \leq \sum_{i=1}^{\infty} \frac{\alpha}{i^2} < \infty.$$

By a criterion which is sometimes called the Kolmogorov criterion (cf., e.g., Theorem 14.5 in Burckel and Bauer (1996)), we get

$$\frac{1}{n} \sum_{i=1}^{n} \left(|X_i - \bar{X}_i| - \mathbf{E}\{|X_i - \bar{X}_i|\}\right) \to 0 \quad a.s. \tag{29}$$

But since $|X_i - X_i^{(k)}| \geq |X_i - X_i^{(l)}|$ for all $l \geq k$ and $i \in \mathbb{N}$, we can conclude

$$\begin{aligned} 0 \leq \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}\{|X_i - \bar{X}_i|\} &= \frac{1}{n} \sum_{i=1}^{n_k} \mathbf{E}\{|X_i - \bar{X}_i|\} + \frac{1}{n} \sum_{i=n_k+1}^{n} \mathbf{E}\{|X_i - \bar{X}_i|\} \\ &\leq \frac{1}{n} \sum_{i=1}^{n_k} \alpha + \frac{1}{n} \sum_{i=n_k+1}^{n} \mathbf{E}\{|X_i - X_i^{(k)}|\} \\ &= \frac{n_k}{n} \cdot \alpha + \frac{1}{n} \sum_{i=n_k+1}^{n} \frac{\alpha}{k} \\ &\leq \frac{n_k}{n} \cdot \alpha + \frac{\alpha}{k} \quad \longrightarrow \frac{\alpha}{k} \quad (n \to \infty), \end{aligned}$$

for every $k \in \mathbb{N}$, which implies

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{E}\{|X_i - \bar{X}_i|\} \to 0$$

and finally by (29)

$$\frac{1}{n}\sum_{i=1}^{n}\left|X_i-\bar{X}_i\right| \to 0 \quad a.s.$$

So it suffies to show, that for some $\epsilon > 0$

$$\limsup_{n\to\infty} \mathbf{P}\left(\left|\hat{q}_{n,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_n\right)-q^{[low]}_{X,\alpha}\right| > \epsilon\right) > 0. \tag{30}$$

To do this we will choose n_k such that (30) holds. Let $0<\epsilon<1$ be fixed and choose n_1 such that

$$\mathbf{P}\left(\left|\hat{q}_{n_1,\alpha}\left(\bar{X}^{(1)}_1,\ldots,\bar{X}^{(1)}_{n_1}\right)-q^{[low]}_{1,\alpha}\right| > \epsilon\right) < \frac{1}{2}.$$

This is possible because of (28). Given $n_1,\ldots,n_{k-1}$, we choose $n_k > n_{k-1}$ such that

$$\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_{k-1}},\bar{X}^{(k)}_{n_{k-1}+1},\ldots,\bar{X}^{(k)}_{n_k}\right)-q^{[low]}_{k,\alpha}\right| > \epsilon\right) < \frac{1}{2},$$

which is again possible because of (28). The choice of $n_1, n_2,\ldots$ implies

$$\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_k}\right)-q^{[low]}_{k,\alpha}\right| > \epsilon\right) < \frac{1}{2}$$

and accordingly

$$\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_k}\right)-q^{[low]}_{k,\alpha}\right| \le \epsilon\right) \ge \frac{1}{2}$$

for $k\in\mathbb{N}$. Using the triangle inequality, we know

$$1=\left|q^{[low]}_{k,\alpha}-q^{[low]}_{X,\alpha}\right| \le \left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_k}\right)-q^{[low]}_{k,\alpha}\right| + \left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_k}\right)-q^{[low]}_{X,\alpha}\right|.$$

Thereby, we can conclude for any $k\in\mathbb{N}$

$$\begin{aligned}
&\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_k}\right)-q^{[low]}_{X,\alpha}\right| > 1-\epsilon\right)\\
&\ge \mathbf{P}\left(1-\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_k}\right)-q^{[low]}_{k,\alpha}\right| > 1-\epsilon\right)\\
&= \mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_1,\ldots,\bar{X}_{n_k}\right)-q^{[low]}_{k,\alpha}\right| < \epsilon\right)\\
&\ge \frac{1}{2},
\end{aligned} \tag{31}$$

which completes the proof. □

4.4 Proof of Theorem 4

For the sake of simplicity we write $q_{X,\alpha}$ for the lower α-quantile of X instead of $q^{[low]}_{X,\alpha}$. We divide the proof into two steps:
In the first step of the proof we show that if α_n is a (possibly random) sequence with

$$\alpha_n \to \alpha \quad a.s.$$

it holds

$$|\hat{q}_{X,n,\alpha_n} - q_{X,\alpha}| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}} + |\alpha_n - \alpha|\right). \tag{32}$$

Therefore it suffices to show

$$\limsup_{n\to\infty} \mathbf{P}\left(|\hat{q}_{X,n,\alpha_n} - q_{X,\alpha}| \le \frac{2\cdot c_1}{c_2}\cdot\left(\frac{1}{\sqrt{n}} + |\alpha_n - \alpha|\right)\right) \ge 1 - 2\cdot\exp\left(-2\cdot c_1^2\right)$$

for every $c_1 \ge 1$, with some finite constant $c_2 > 0$.

Since F is differentiable at $q_{X,\alpha}$ with derivative greater than zero, there exist finite constants $c_2 > 0$ and $\zeta > 0$, such that

$$c_2\,|q_{X,\alpha} - x| \le |F(q_{X,\alpha}) - F(x)| \tag{33}$$

for all x with $|q_{X,\alpha} - x| \le \zeta$.

Now set

$$B_n := \left\{\frac{2c_1}{c_2}\,|\alpha_n - \alpha| \le \frac{\zeta}{2}\right\}$$

and

$$C_n := \left\{\sup_{t\in\mathbb{R}} |F(t) - F_n(t)| \le \frac{c_1}{\sqrt{n}}\right\}.$$

We know

$$\mathbf{P}(B_n^c) \to 0 \quad (n\to\infty) \quad \text{and} \quad \mathbf{P}(C_n^c) \le 2\cdot\exp\left(-2\cdot c_1^2\right)$$

because of $\alpha_n \to \alpha$ *a.s.* and the Dvoretzky-Kiefer-Wolfowitz inequality (cf. Dvoretzky et al. (1956)) in combination with Corollary 1 in Massart (1990). Choose $n_0 \in \mathbb{N}$, such that $0 < \frac{2}{c_2}\cdot\frac{c_1}{\sqrt{n}} \le \frac{\zeta}{2}$ is fullfilled for all $n \ge n_0$. Assume in the following, that the events B_n and C_n hold and consider $n \ge n_0$. Set $\theta_n = 2c_1\cdot|\alpha_n - \alpha| + 2\cdot\frac{c_1}{\sqrt{n}}$. The assumptions imply

$$0 < \frac{1}{c_2}\cdot\theta_n = \frac{2c_1}{c_2}\cdot|\alpha_n - \alpha| + \frac{2}{c_2}\cdot\frac{c_1}{\sqrt{n}} \le \frac{\zeta}{2} + \frac{\zeta}{2} = \zeta$$

so we can conclude by (33) and $F(q_{X,\alpha}) = \alpha$

$$\theta_n = c_2\left|q_{X,\alpha} - q_{X,\alpha} - \frac{1}{c_2}\theta_n\right| \le \left|\alpha - F\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right)\right| \tag{34}$$

and

$$\theta_n = c_2\left|q_{X,\alpha} - q_{X,\alpha} + \frac{1}{c_2}\theta_n\right| \le \left|\alpha - F\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right)\right|. \tag{35}$$

Because F is differentiable at $q_{X,\alpha}$ with derivative greater zero, we know

$$F\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right) < \alpha < F\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right),$$

hence (34) and (35) imply

$$F\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right) < \alpha - \frac{\theta_n}{2} < \alpha < \alpha + \frac{\theta_n}{2} < F\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right). \tag{36}$$

Since the event C_n holds, we know

$$F_n\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right) - \frac{c_1}{\sqrt{n}} \leq F\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right)$$

and

$$F\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right) \leq F_n\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right) + \frac{c_1}{\sqrt{n}}.$$

Combining this with (36) and the definition of θ_n leads to

$$F_n\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right) < \alpha - c_1 \cdot |\alpha - \alpha_n| \leq \alpha + c_1 \cdot |\alpha - \alpha_n| < F_n\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right).$$

Since $c_1 \geq 1$ we have

$$\alpha - c_1 \cdot |\alpha - \alpha_n| \leq \alpha_n \leq \alpha + c_1 \cdot |\alpha - \alpha_n|,$$

which implies

$$F_n\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right) < \alpha_n < F_n\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right).$$

So finally we have shown

$$\mathbf{P}\left(B_n \cap C_n\right) \leq \mathbf{P}\left(F_n\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right) < \alpha_n < F_n\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right)\right),$$

which by the definition of $\hat{q}_{X,n,\alpha_n}$ and for $n \geq n_0$ leads to

$$\begin{aligned}
\mathbf{P}\left(|\hat{q}_{X,n,\alpha_n} - q_{X,\alpha}| \leq \frac{1}{c_2}\theta_n\right) &= \mathbf{P}\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n \leq \hat{q}_{X,n,\alpha_n} \leq q_{X,\alpha} + \frac{1}{c_2}\theta_n\right) \\
&\geq \mathbf{P}\left(F_n\left(q_{X,\alpha} - \frac{1}{c_2}\theta_n\right) < \alpha_n < F_n\left(q_{X,\alpha} + \frac{1}{c_2}\theta_n\right)\right) \\
&\geq \mathbf{P}\left(B_n \cap C_n\right) \\
&= 1 - \mathbf{P}\left(B_n^c \cup C_n^c\right) \\
&\geq 1 - \mathbf{P}\left(B_n^c\right) - \mathbf{P}\left(C_n^c\right) \to 1 - 2\exp\left(-2 \cdot c_1^2\right) \quad (n \to \infty).
\end{aligned}$$

This was the assertion.

Furthermore, we know (see proof of Theorem 2 in combination with (14))

$$\delta_n = \frac{1}{n}\sum_{i=1}^{n} I_{\{|X_i - \hat{X}_{i,n}| > \sqrt{\eta_n}\}} \leq \frac{\eta_n}{\sqrt{\eta_n}} = \sqrt{\eta_n} \to 0 \quad a.s. \tag{37}$$

Using (37), application of Lemma 1 yields

$$\hat{q}_{X,n,\alpha-\sqrt{\eta_n}} - \sqrt{\eta_n} \leq \hat{q}_{\bar{X},n,\alpha} \leq \hat{q}_{X,n,\alpha+\sqrt{\eta_n}} + \sqrt{\eta_n} \tag{38}$$

for all $n \in \mathbb{N}$.
In the second step of the proof we finally show the assertion. By the first step we can conclude

$$\left|\hat{q}_{X,n,\alpha-\sqrt{\eta_n}} - q_{X,\alpha}\right| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}} + \sqrt{\eta_n}\right)$$

and

$$\left|\hat{q}_{X,n,\alpha+\sqrt{\eta_n}} - q_{X,\alpha}\right| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}} + \sqrt{\eta_n}\right).$$

By (38) we know

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

which completes the proof. □

4.5 Proof of Theorem 5

Let $\alpha \in (0,1)$ be arbitrary. For the sake of simplicity we write $q_{X,\alpha}$ for the lower α-quantile of X instead of $q_{X,\alpha}^{[low]}$. Assume to the contrary that there exists an estimator $(\hat{q}_{n,\alpha})_{n\in N}$ such that for all random variables $\bar{X}_{1,n}, \bar{X}_{2,n}, \ldots$, which are such that for some independent and identically as X distributed $X_1, X_2, \ldots$ it holds

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_{i,n}| \to 0 \quad a.s., \tag{39}$$

we have

$$\lim_{c\to\infty}\limsup_{n\to\infty}\mathbf{P}\left(\left|\hat{q}_{n,\alpha}\left(\bar{X}_{1,n},\ldots,\bar{X}_{n,n}\right) - q_{X,\alpha}\right| > c\cdot\left(\frac{1}{\sqrt{n}} + \tilde{\eta}_n\right)\right) = 0, \tag{40}$$

with a sequence $\tilde{\eta}_n$ that fullfills

$$\frac{\tilde{\eta}_n}{\sqrt{\eta_n}} \to^{\mathbf{P}} 0. \tag{41}$$

Let $X, X_1, X_2, \ldots$ be independent and identically uniformly on $(0,1)$ distributed, i.e., with cdf.

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \leq x < 1 \\ 1 & \text{if} \quad x \geq 1 \end{cases}$$

and lower α-quantile $q_{X,\alpha} = \alpha$. For $k \in \mathbb{N}$ let $Y^{(k)}$ have the distribution function

$$F_k(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \le x < \alpha - \sqrt{\frac{1}{k}} \\ \alpha - \sqrt{\frac{1}{k}} & \text{if} \quad \alpha - \sqrt{\frac{1}{k}} \le x < \alpha \\ 2 \cdot (x - \alpha) + \alpha - \sqrt{\frac{1}{k}} & \text{if} \quad \alpha \le x < \alpha + \sqrt{\frac{1}{k}} \\ x & \text{if} \quad \alpha + \sqrt{\frac{1}{k}} \le x < 1 \\ 1 & \text{if} \quad 1 \le x. \end{cases}$$

In other words the distribution of the random variable $Y^{(k)}$ is obtained by shifting all mass, that is contained in the interval $\left[\alpha - \sqrt{\frac{1}{k}}, \alpha\right]$, by $\sqrt{\frac{1}{k}}$ to the right. This distribution has the lower α-quantile $q_{Y^{(k)},\alpha} = \alpha + \frac{1}{2}\sqrt{\frac{1}{k}}$. Furthermore, we set

$$X_{i,n}^{(k)} = \begin{cases} X_i + \sqrt{\frac{1}{k}} & \text{if} \quad X_i \in \left[\alpha - \sqrt{\frac{1}{k}}, \alpha\right] \text{ and } X_i \text{ is one of the } \lfloor \sqrt{\frac{1}{k}} \cdot n \rfloor \\ & \qquad \text{biggest samples of } (X_j)_{j=1,\ldots,n} \text{ in } \left[\alpha - \sqrt{\frac{1}{k}}, \alpha\right] \\ X_i & \text{else} \end{cases}$$

and notice that this is almost surely well defined, since ties occur only with probability zero because F is continuous. Now let $Y_1^{(k)}, Y_2^{(k)}, \ldots$ be independet and identically as $Y^{(k)}$ distributed. Then we know by (40) that for every $k \in \mathbb{N}$

$$\limsup_{n\to\infty} \mathbf{P}\left(\left|\hat{q}_{n,\alpha}\left(Y_1^{(k)}, \ldots, Y_n^{(k)}\right) - q_{Y^{(k)},\alpha}\right| \ge \frac{1}{4}\sqrt{\frac{1}{k}}\right) = 0. \tag{42}$$

Denote by $A_n^{(k)}$ the event, that there are not more than $\lfloor \sqrt{\frac{1}{k}} \cdot n \rfloor$ of the samples $(X_i)_{i=1,\ldots,n}$ in the interval $\left[\alpha - \sqrt{\frac{1}{k}}, \alpha\right]$. Then the de Moivre-Laplace theorem (cf., e.g., Theorem 1 and Corollary 1 on pp. 47-48 in Chow and Teicher (1978)), which is a special case of the central limit theorem for binomially-distributed random variables, implies for a $B\left(n, \sqrt{\frac{1}{k}}\right)$-distributed random variable Z, and $p = \sqrt{\frac{1}{k}}$

$$\begin{aligned} \mathbf{P}\left(A_n^{(k)}\right) &= \sum_{l=0}^{\lfloor p \cdot n \rfloor} \binom{n}{l} \cdot \mathbf{P}\left(X \in [\alpha - p, \alpha]\right)^l \cdot \mathbf{P}\left(X \notin [\alpha - p, \alpha]\right)^{n-l} \\ &= \sum_{l=0}^{\lfloor p \cdot n \rfloor} \binom{n}{l} \cdot p^l \cdot (1-p)^{n-l} \\ &= \mathbf{P}\left(Z \le \lfloor p \cdot n \rfloor\right) \\ &= \mathbf{P}\left(\frac{Z - \lfloor p \cdot n \rfloor}{\sqrt{np(1-p)}} \le 0\right) \to \frac{1}{2} \quad (n \to \infty) \end{aligned}$$

and

$$\mathbf{P}\left(\left(A_n^{(k)}\right)^c\right) \to \frac{1}{2} \quad (n \to \infty)$$

for every $k \in \mathbb{N}$. So we can conclude by (42) that for every $k \in \mathbb{N}$

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{P}\left(\left|\hat{q}_{n,\alpha}\left(X_{1,n}^{(k)}, \ldots, X_{n,n}^{(k)}\right) - q_{Y^{(k)},\alpha}\right| \geq \frac{1}{4}\sqrt{\frac{1}{k}}\right) \\
&\leq \limsup_{n\to\infty} \left[\mathbf{P}\left(\left\{\left|\hat{q}_{n,\alpha}\left(X_{1,n}^{(k)}, \ldots, X_{n,n}^{(k)}\right) - q_{Y^{(k)},\alpha}\right| \geq \frac{1}{4}\sqrt{\frac{1}{k}}\right\} \cap A_n^{(k)}\right) + \mathbf{P}\left(\left(A_n^{(k)}\right)^c\right)\right] \\
&= 0 + \frac{1}{2} = \frac{1}{2},
\end{aligned} \tag{43}$$

because if we intersect with the event $A_n^{(k)}$ the samples $X_{1,n}^{(k)}, \ldots, X_{n,n}^{(k)}$ are in fact samples drawn from the distribution of the random variable $Y^{(k)}$. So for every $k \in \mathbb{N}$ we get in particular for n large enough

$$\mathbf{P}\left(\left|\hat{q}_{n,\alpha}\left(X_{1,n}^{(k)}, \ldots, X_{n,n}^{(k)}\right) - q_{Y^{(k)},\alpha}\right| \geq \frac{1}{4}\sqrt{\frac{1}{k}}\right) \leq \frac{3}{4}. \tag{44}$$

It suffices to show, that there exists a strictly increasing sequence $(n_k)_{k\in\mathbb{N}}$ and data with measurement error $\bar{X}_{1,n_k}, \ldots, \bar{X}_{n_k,n_k}$, fullfilling (39), and $\tilde{\eta}_n$ satisfying (41), such that for every $c_3 > 0$

$$\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k}, \ldots, \bar{X}_{n_k,n_k}\right) - q_{X,\alpha}\right| > c_3 \cdot \left(\frac{1}{\sqrt{n_k}} + \tilde{\eta}_{n_k}\right)\right) \geq \frac{1}{8} \tag{45}$$

for k large enough.
We will now sequentially construct such a sequence n_k and the data $\bar{X}_{1,n_k}, \ldots, \bar{X}_{n_k,n_k}$ and show that (45) holds. Choose $n_1 \geq 1$ such that

$$\mathbf{P}\left(\left|\hat{q}_{n_1,\alpha}\left(X_{1,n_1}^{(1)}, \ldots, X_{n_1,n_1}^{(1)}\right) - q_{Y^{(1)},\alpha}\right| \geq \frac{1}{4}\sqrt{\frac{1}{1}}\right) \leq \frac{3}{4}$$

holds. This is possible because of (44). Given n_{k-1}, choose $n_k > n_{k-1}$ such that $n_k \geq k^2$ and

$$\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(X_{1,n_k}^{(k)}, \ldots, X_{n_k,n_k}^{(k)}\right) - q_{Y^{(k)},\alpha}\right| \geq \frac{1}{4}\sqrt{\frac{1}{k}}\right) \leq \frac{3}{4}.$$

hold. This is again possible because of (44). Setting

$$\begin{array}{llll}
\bar{X}_{i,n} = X_{i,n}^{(1)} & \text{for} \quad 0 < n \leq n_1 & \text{and } i = 1, ..., n & \text{and} \\
\bar{X}_{i,n} = X_{i,n}^{(k)} & \text{for} \quad n_{k-1} < n \leq n_k & \text{and } i = 1, ..., n, &
\end{array} \tag{46}$$

we can conclude for $n_{k-1} < n \leq n_k$

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}_{i,n}\right| = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - X_{i,n}^{(k)}\right| \leq \frac{1}{n}\cdot\left\lfloor\sqrt{\frac{1}{k}}\cdot n\right\rfloor\cdot\sqrt{\frac{1}{k}} \leq \frac{1}{k}$$

and in particular

$$\eta_{n_k} \leq \frac{1}{k} \quad \text{for all } k \in \mathbb{N}$$

and

$$\eta_n \to 0 \quad a.s.$$

In this way we have constructed a strictly increasing sequence $(n_k)_{k\in\mathbb{N}}$ and data with measurement error $\bar{X}_{1,n_k}, ..., \bar{X}_{n_k,n_k}$ such that

$$\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\ldots,\bar{X}_{n_k,n_k}\right) - q_{Y^{(k)},\alpha}\right| \geq \frac{1}{4}\sqrt{\frac{1}{k}}\right) \leq \frac{3}{4}. \tag{47}$$

By the triangle inequality, we know

$$\begin{aligned}\frac{1}{2}\sqrt{\frac{1}{k}} &= \left|q_{Y^{(k)},\alpha} - q_{X,\alpha}\right| \\ &\leq \left|q_{Y^{(k)},\alpha} - \hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\ldots,\bar{X}_{n_k,n_k}\right)\right| + \left|\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\ldots,\bar{X}_{n_k,n_k}\right) - q_{X,\alpha}\right|.\end{aligned} \tag{48}$$

Thereby, we can conclude for all $k \in \mathbb{N}$

$$\begin{aligned}&\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\ldots,\bar{X}_{n_k,n_k}\right) - q_{X,\alpha}\right| > c_3\cdot\left(\frac{1}{\sqrt{n_k}} + \tilde{\eta}_{n_k}\right)\right) \\ &\geq \mathbf{P}\left(\frac{1}{2}\sqrt{\frac{1}{k}} - \left|q_{Y^{(k)},\alpha} - \hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\ldots,\bar{X}_{n_k,n_k}\right)\right| > c_3\cdot\left(\frac{1}{\sqrt{n_k}} + \tilde{\eta}_{n_k}\right)\right) \\ &= \mathbf{P}\left(\frac{1}{2}\sqrt{\frac{1}{k}} - c_3\cdot\left(\frac{1}{\sqrt{n_k}} + \tilde{\eta}_{n_k}\right) > \left|q_{Y^{(k)},\alpha} - \hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\ldots,\bar{X}_{n_k,n_k}\right)\right|\right).\end{aligned}$$

Since $\eta_{n_k} \leq \frac{1}{k}$, we know by (41)

$$\frac{\tilde{\eta}_{n_k}}{\frac{1}{4}\sqrt{\frac{1}{k}}} \leq \frac{4\cdot\tilde{\eta}_{n_k}}{\sqrt{\eta_{n_k}}} \to^{\mathbf{P}} 0 \quad (k\to\infty).$$

Furthermore, since $n_k \geq k^2$ for all $k \in \mathbb{N}$ by construction, we have

$$\frac{\frac{1}{\sqrt{n_k}}}{\frac{1}{4}\sqrt{\frac{1}{k}}} \leq \frac{\frac{1}{\sqrt{k^2}}}{\frac{1}{4}\sqrt{\frac{1}{k}}} \to 0 \quad (k\to\infty),$$

which implies

$$\frac{c_3\left(\tilde{\eta}_{n_k}+\frac{1}{\sqrt{n_k}}\right)}{\frac{1}{4}\sqrt{\frac{1}{\sqrt{k}}}} \to^{\mathbf{P}} 0 \quad (k\to\infty)$$

for every $c_3>0$. So setting

$$B_k=\left\{c_3\cdot\left(\tilde{\eta}_{n_k}+\frac{1}{\sqrt{n_k}}\right)\leq \frac{1}{4}\sqrt{\frac{1}{k}}\right\}$$

yields

$$\mathbf{P}\left(B_k\right)\to 1 \;\; (k\to\infty)$$

and thus

$$\mathbf{P}\left(B_k\right)\geq \frac{7}{8}$$

for k large enough. Thereby, we finally get for every $c_3>0$ and k large enough

$$\begin{aligned}
&\mathbf{P}\left(\left|\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\dots,\bar{X}_{n_k,n_k}\right)-q_{X,\alpha}\right|>c_3\cdot\left(\tilde{\eta}_{n_k}+\frac{1}{\sqrt{n_k}}\right)\right)\\
&\geq \mathbf{P}\left(\frac{1}{2}\sqrt{\frac{1}{k}}-c_3\cdot\left(\tilde{\eta}_{n_k}+\frac{1}{\sqrt{n_k}}\right)>\left|q_{Y^{(k)},\alpha}-\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\dots,\bar{X}_{n_k,n_k}\right)\right|\right)\\
&\geq \mathbf{P}\left(\left\{\frac{1}{2}\sqrt{\frac{1}{k}}-c_3\cdot\left(\tilde{\eta}_{n_k}+\frac{1}{\sqrt{n_k}}\right)>\left|q_{Y^{(k)},\alpha}-\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\dots,\bar{X}_{n_k,n_k}\right)\right|\right\}\cap B_k\right)\\
&\geq \mathbf{P}\left(\left\{\frac{1}{4}\sqrt{\frac{1}{k}}>\left|q_{Y^{(k)},\alpha}-\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\dots,\bar{X}_{n_k,n_k}\right)\right|\right\}\cap B_k\right)\\
&\geq \mathbf{P}\left(\frac{1}{4}\sqrt{\frac{1}{k}}>\left|q_{Y^{(k)},\alpha}-\hat{q}_{n_k,\alpha}\left(\bar{X}_{1,n_k},\dots,\bar{X}_{n_k,n_k}\right)\right|\right)-\mathbf{P}\left(B_k^c\right)\\
&\geq \frac{1}{4}-\frac{1}{8}=\frac{1}{8},
\end{aligned}$$

where we have used (47) in the last inequality. This yields the assertion. □

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] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. John Wiley & Sons.

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

[3] Bott, A., Devroye, L. and Kohler, M. (2013). Estimation of a distribution from data with small measurement errors. *Electronic Journal of Statistics*, **7**, pp. 2457-2476.

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

[5] Burckel, R. and Bauer, H. (1996). *Probability Theory*. De Gruyter, Berlin.

[6] Chow, Y. S. and Teicher, H. (1978). *Probability Theory: Independence, interchangeability, martingales*. Springer, New York.

[7] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer, New York.

[8] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, **27**, pp. 642-669.

[9] Feldman, D. and Tucker, H. G. (1966). Estimation of non-unique quantiles. *Annals of Mathematical Statistics*, **37**, pp. 451-457.

[10] Furer, D. and Kohler, M. (2015). Smoothing spline regression estimation based on real and artificial data. *Metrika*, **78**, pp. 711-746.

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

[12] Kohler, M. (2006). Nonparametric regression with additional measurement errors in the dependent variable. *Journal of Statistical Planning and Inference*, **136**, pp. 3339-3361.

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

[14] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Annals of Probability*, **18**, pp. 1269-1283.

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

[16] Mosteller, F. (1946). On some useful "inefficient" statistics. *Annals of Mathematical Statistics*, **17**, pp. 377-408.

[17] Pfanzagl, J. (1976). Investigating the quantile of an unknown distribution, in *Contributions to Applied Statistics (Ziegler, W.J., ed.)*. Birkhäuser, Basel.

[18] Puri, M. L. and Ralescu, S. S. (1986). Limit theorems for random central order statistics, in *Adaptive Statistical Procedures and Related Topics*, **8**, pp. 447-475, Institute of Mathematical Statistics, Hayward, CA.

[19] Reiss, R.-D. (1974). On the accuracy of the normal approximation for quantiles. *Annals of Probability*, **2**, pp. 741-744.

[20] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley series in probability and mathematical statistics, Wiley, New York.

[21] Smirnov, N. V. (1952). Limit distributions for the terms of a variational series. *American Mathematical Society Translations (1)*, **11**, pp. 82-143.

[22] Tucker, H. G. (1967). *A Graduate Course in Probability.* Academic Press, New York.

[23] 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.

Appendix: Proof of Lemma 2

(a) It suffices to show

(i) $\mathbf{P}\left(\hat{q}_{X,n,\gamma_{n,l}} \leq q_{X,\alpha}^{[low]} - \epsilon \quad i.o.\right) = 0$ for any $\epsilon > 0$, and

(ii) $\mathbf{P}\left(\hat{q}_{X,n,\gamma_{n,l}} > q_{X,\alpha}^{[low]} \quad i.o.\right) = 0$,

where *i.o.* means infinitely often. First of all we show (i). Therefore let $\epsilon > 0$ be arbitrary. We know

$$F\left(q_{X,\alpha}^{[low]} - \epsilon\right) < \alpha.$$

Setting

$$\rho_2 = \alpha - F\left(q_{X,\alpha}^{[low]} - \epsilon\right),$$

we can conclude

$$F\left(q_{X,\alpha}^{[low]} - \epsilon\right) + \frac{\rho_2}{2} < \alpha.$$

Choose

$$N := \left\{\gamma_{n,l} \to \alpha \ (n \to \infty) \quad \text{and} \quad \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \to 0 \ (n \to \infty)\right\}.$$

As in the proof of Theorem 1 we have $\mathbf{P}(N) = 1$. We can (for every $\omega \in N$) find n_0 such that for all $n \geq n_0$ it holds

$$|\gamma_{n,l} - \alpha| \leq \frac{\rho_2}{4} \quad \text{and} \quad \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \leq \frac{\rho_2}{4}.$$

This implies (for every $\omega \in N$)

$$F_n\left(q_{X,\alpha}^{[low]} - \epsilon\right) < \gamma_{n,l}$$

and hence

$$\hat{q}_{X,n,\gamma_{n,l}} > q_{X,\alpha}^{[low]} - \epsilon$$

for n large enough. So we actually have shown

$$1 - \mathbf{P}\left(\hat{q}_{X,n,\gamma_{n,l}} \leq q_{X,\alpha}^{[low]} - \epsilon \quad i.o.\right) \geq \mathbf{P}\left(N\right) = 1,$$

which proves (i).
It remains to show (ii). Therefore set

$$U_i = 1 - 2 \cdot I_{\left\{X_i \leq q_{X,\alpha}^{[low]}\right\}} \quad \text{for } i = 1, ..., n$$

and

$$p_1 = \mathbf{P}\left(X \leq q_{X,\alpha}^{[low]}\right) \geq \alpha.$$

We know

$$\mathbf{E}\left\{U_i\right\} = 1 - 2 \cdot p_1 \leq 1 - 2 \cdot \alpha \quad \text{and} \quad s = \mathbf{V}\left\{U_i\right\} = 4p_1 \cdot (1 - p_1)$$

and

$$\sum_{i=1}^{n} U_i = n - 2n \cdot F_n\left(q_{X,\alpha}^{[low]}\right).$$

Thus,

$$\begin{aligned}
\left\{\hat{q}_{X,n,\gamma_{n,l}} > q_{X,\alpha}^{[low]}\right\} &= \left\{F_n\left(q_{X,\alpha}^{[low]}\right) < \gamma_{n,l}\right\} \\
&= \left\{-2n \cdot F_n\left(q_{X,\alpha}^{[low]}\right) > -2 \cdot \gamma_{n,l} \cdot n\right\} \\
&\subseteq \left\{\sum_{i=1}^{n} U_i \geq n - 2 \cdot \gamma_{n,l} \cdot n\right). \qquad (49)
\end{aligned}$$

Set $\psi_n = (2 \cdot n \cdot s \cdot \log(\log(n \cdot s)))^{1/2}$, which we will need in the subsequent application of Kolmogorov's law of the iterated logarithm. Observe that ψ_n is well-defined for n large enough. Since $0 \leq x \cdot (1 - x) \leq \frac{1}{4}$ for $x \in [0, 1]$, we have $0 \leq s \leq 1$ and thus $(2n \cdot \log(\log(n)))^{1/2} \geq \psi_n$. Because of

$$\alpha - \gamma_{n,l} > (1 + \nu) \cdot \sqrt{\frac{2 \cdot \log(\log(n/2))}{n}},$$

we can conclude

$$\alpha - \gamma_{n,l} \geq \frac{1 + \nu}{2} \cdot \sqrt{\frac{2 \cdot \log(\log(n))}{n}}$$

for all n large enough. Combining this with

$$1 - 2 \cdot p_1 \leq 1 - 2 \cdot \alpha,$$

we get by (49)

$$\begin{aligned}
&\mathbf{P}\left(\hat{q}_{X,n,\gamma_{n,l}} > q_{X,\alpha}^{[low]} \ \ i.o.\right) \\
&\leq \mathbf{P}\left(\sum_{i=1}^{n} U_i \geq n - 2 \cdot \gamma_{n,l} \cdot n \ \ i.o.\right) \\
&\leq \mathbf{P}\left(\sum_{i=1}^{n} U_i \geq n \cdot (1 - 2 \cdot \alpha) + 2 \cdot (\alpha \cdot n - \gamma_{n,l} \cdot n) \ \ i.o.\right) \\
&\leq \mathbf{P}\left(\sum_{i=1}^{n} U_i \geq n \cdot (1 - 2 \cdot p_1) + (1 + \nu) \cdot \psi_n \ \ i.o.\right).
\end{aligned}$$

We know by Kolmogorov's law of the iterated logarithm (cf., e.g., Theorem 1 on page 140 in Tucker (1967))

$$\mathbf{P}\left(\limsup_{n \to \infty} \frac{\sum_{i=1}^{n} U_i - n \cdot (1 - 2 \cdot p_1)}{\psi_n} = 1\right) = 1,$$

from which we can conclude

$$\mathbf{P}\left(\sum_{i=1}^{n} U_i \geq n \cdot (1 - 2 \cdot p_1) + (1 + \nu) \cdot \psi_n \ \ i.o.\right) = 0.$$

This completes the proof of (a).

(b) It suffices to show

(i) $\mathbf{P}\left(\hat{q}_{X,n,\gamma_{n,r}} > q_{X,\alpha}^{[up]} + \epsilon \quad i.o.\right) = 0$ for any $\epsilon > 0$, and

(ii) $\mathbf{P}\left(\hat{q}_{X,n,\gamma_{n,r}} < q_{X,\alpha}^{[up]} \quad i.o.\right) = 0.$

The proof of (i) is analogously to (i) in part (a). It remains to show (ii). Therefore set

$$V_i = 2 \cdot I_{\left\{X_i < q_{X,\alpha}^{[up]}\right\}} - 1 \quad \text{for } i = 1, ..., n$$

and

$$p_2 = \mathbf{P}\left(X < q_{X,\alpha}^{[up]}\right) \leq \alpha.$$

We have $\mathbf{E}\left\{V_i\right\} = 2p_2 - 1 \leq 2\alpha - 1$ and $\tilde{s} = \mathbf{V}\left\{V_i\right\} = 4p_2 \cdot (1 - p_2)$. Observe that if

$$\hat{q}_{X,n,\gamma_{n,r}} < q_{X,\alpha}^{[up]},$$

then

$$\frac{1}{n}\sum_{i=1}^{n} I_{\left\{X_i < q^{[up]}_{X,\alpha}\right\}} \geq \frac{1}{n}\sum_{i=1}^{n} I_{\left\{X_i \leq \hat{q}_{X,n,\gamma_{n,r}}\right\}} = F_n\left(\hat{q}_{X,n,\gamma_{n,r}}\right) \geq \gamma_{n,r}.$$

Thereby, we can analogously to (ii) in part (a) conclude

$$\left\{\hat{q}_{X,n,\gamma_{n,r}} < q^{[up]}_{X,\alpha}\right\} \subseteq \left\{\sum_{i=1}^{n} V_i \geq 2\cdot\gamma_{n,r}\cdot n - n\right\}.$$

Again, set $\tilde{\psi} = (2\cdot n\cdot\tilde{s}\cdot\log(\log(n\cdot\tilde{s})))^{1/2}$. Since $0 \leq x\cdot(1-x) \leq \frac{1}{4}$ for $x\in[0,1]$, we have $(2n\cdot\log(\log(n)))^{1/2} \geq \tilde{\psi}_n$. The assumption on $\gamma_{n,r}$ implies

$$\gamma_{n,r} - \alpha \geq \frac{1+\nu}{2}\cdot\sqrt{\frac{2\cdot\log(\log(n))}{n}}$$

for all n large enough. Thus, using $2\cdot\alpha - 1 \geq 2\cdot p_2 - 1$, we can conclude

$$\mathbf{P}\left(\hat{q}_{X,n,\gamma_{n,l}} < q^{[up]}_{X,\alpha}\ \ i.o.\right) \leq \mathbf{P}\left(\sum_{i=1}^{n} V_i \geq n\cdot(2\cdot p_2 - 1) + (1+\nu)\cdot\tilde{\psi}_n\ \ i.o.\right)$$

Again, by Kolmogorov's law of the iterated logarithm, we get

$$\mathbf{P}\left(\sum_{i=1}^{n} V_i \geq n\cdot(2p_2 - 1) + (1+\nu)\cdot\tilde{\psi}_n\ \ i.o.\right) = 0,$$

which completes the proof. $\square$