Nonparametric estimation of a maximum of quantiles

Georg C. Enss[1], Benedict Götz[1], Michael Kohler[2], Adam Krzyżak[3,*] and Roland Platz[4].

[1] *Fachgebiet Systemzuverlässigkeit und Maschinenakustik SzM, Technische Universität Darmstadt, Magdalenenstr. 4, 64289 Darmstadt, Germany, email: {enss, goetz}@szm.tu-darmstadt.de*

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

[3] *Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8, email: krzyzak@cs.concordia.ca*

[4] *Fraunhofer Institute for Structural Durability and System Reliability LBF, Bartningstr. 47, 64289 Darmstadt, Germany, email: roland.platz@lbf.fraunhofer.de*

May 16, 2014

Abstract

A simulation model of a complex system is considered, for which the outcome is described by $m(p, X)$, where p is a parameter of the system, X is a random input of the system and m is a real-valued function. The maximum (with respect to p) of the quantiles of $m(p, X)$ is estimated. The quantiles of $m(p, X)$ of a given level are estimated for various values of p by an order statistic of values $m(p_i, X_i)$ where $X, X_1, X_2, \ldots$ are independent and identically distributed and where p_i is close to p, and the maximal quantile is estimated by the maximum of these quantile estimates. Under assumptions on the smoothness of the function describing the dependency of the values of the quantiles on the parameter p the rate of convergence of this estimate is analyzed. The finite sample size behavior of the estimate is illustrated by simulated data and by applying it in a simulation model of a real mechanical system.

AMS classification: Primary 62G05; secondary 62G30.

Key words and phrases: Conditional quantile estimation, maximal quantile, rate of convergence, supremum norm error.

*Corresponding author. Tel: +1-514-848-2424 ext. 3007, Fax: +1-514-848-2830

Running title: *Estimation of the maximal quantile*

1 Introduction

We consider a simulation model of a complex system described by

$$Y = m(p, X),$$

where p is a parameter of the system from some compact subset $\mathcal{P}$ of $\mathbb{R}^l$, X is a $\mathbb{R}^d$-valued random variable with known distribution and $m : \mathcal{P} \times \mathbb{R}^d \to \mathbb{R}$ is a given (measurable) function. Let

$$G_p(y) = \mathbf{P}\{m(p, X) \leq y\}$$

be the cumulative distribution function (cdf) of Y for a given value of p. For $\alpha \in (0, 1)$ let

$$q_{p,\alpha} = \min\{y \in \mathbb{R} : G_p(y) \geq \alpha\}$$

be the α-quantile of $m(p, X)$. Using at most n function evaluations $m(p_i, x_i)$ of m for arbitrarily chosen values

$$(p_1, x_1), \ldots, (p_n, x_n) \in \mathcal{P} \times \mathbb{R}^d$$

we are interested in estimating the maximal α-quantile

$$\sup_{p \in \mathcal{P}} q_{p,\alpha}.$$

This problem is motivated by research carried out at the Collaborative Research Centre SFB 805 at the Technische Universität Darmstadt which focuses on uncertainty in load-carrying systems such as truss structures. The truss structure under consideration is a typical system for load distribution in mechanical engineering (cf., Figure 1). It is made of ten nodes and 24 equal beams

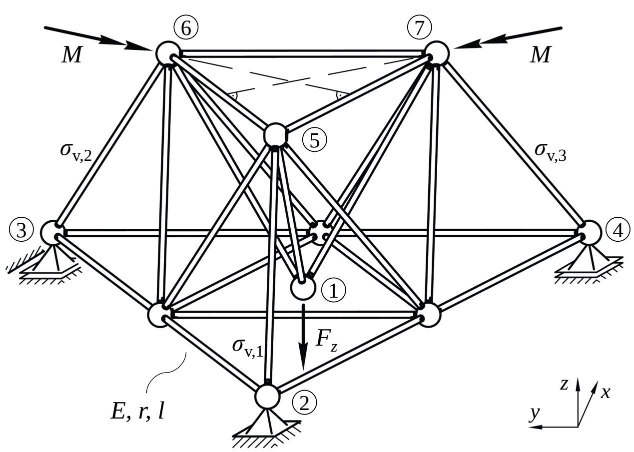

Figure 1: Truss structure

with circular cross sections and fixed boundary conditions. Four equal tetrahedron structures may be assembled into modules. The truss is supported statically determined at nodes 2, 3 and 4. A single force F_z in z-direction acts at node 1 and two symmetric moments M act at nodes 6 and 7 as shown in Figure 1. Due to the non-symmetric loading, the maximum equivalent stress σ_{v} (cf., e.g., Gross et al. (2007)) is not equal in every beam. It is influenced not only by the non-symmetric loading via F_z and M but also by the beam's material parameter Young's modulus E as well as geometric parameters radius r and length l. We assume that from the manufacturing process of the truss structure we know for each beam parameter E, r and l the corresponding ranges $[a, b]$. The parameter vector $p = (E, r, l)$ describes the parameters for all 24 equal beams within the truss structure.

The occurring random force F_z is normally distributed and the moments M have truncated normal distribution restricted to positive half-axis. A numerical finite element model is built that calculates the maximum equivalent stress $Y = \sigma_{\mathrm{v},1}$, $Y = \sigma_{\mathrm{v},2}$ and $Y = \sigma_{\mathrm{v},3}$ in the beams 1, 2 and 3, resp., that are the members with the highest equivalent stresses in this specific truss structure. So our parameter set is $\mathcal{P} = [a_E, b_E] \times [a_r, b_r] \times [a_l, b_l]$. The function $m : \mathcal{P} \times \mathbb{R}_+ \to \mathbb{R}_+$ describes the physical model of the truss structure and $Y = m(p, X)$ is the occurring maximum equivalent stress in beam 1, 2 and 3, resp., of the truss structure with load vector $X = (F_z, M)$.

We are interested in quantifying the uncertainty in the random quantity $Y = m(p, X)$ which we characterize by determining an interval in which the value of Y is contained with high probability. Such an interval can be determined by computing quantiles of Y for α close to one (leading to the right end point of the interval) and α close to zero (leading to the left end point of the interval). However, we do not know the exact value of $p \in \mathcal{P}$. Hence instead of computing the right end point of the interval by a quantile of Y we compute for each $p \in \mathcal{P}$ the α-quantile $q_{p,\alpha}$ of $m(p, X)$ and use $\sup_{p\in\mathcal{P}} q_{p,\alpha}$ for α close to one as the right end point of the interval. Similarly we can construct the left end point of the interval by computing $\inf_{p\in\mathcal{P}} q_{p,\alpha}$ for α close to zero.

In the sequel we will describe how we can define estimates of these quantities. In order to simplify the notation we consider only $\sup_{p\in\mathcal{P}} q_{p,\alpha}$, the other value can be estimated similarly.

For $p \in \mathcal{P}$ we estimate the α-quantile of G_p by the α-quantile of an empirical cumulative distribution function corresponding to data points $m(p_i, X_i)$, where p_i is close to p and X, X_1, X_2, … are independent and identically distributed random variables. Then we estimate the maximum quantile by the maximum of these quantile estimates. Under suitable assumptions on the smoothness of the function describing the dependency of the quantiles on the parameter we analyze the rate of convergence of this estimate. The finite sample size behavior of the estimate is illustrated using simulated data.

In our proofs the main trick is to relate the error of our quantile estimates to the behavior of the order statistics of a fixed sequence of independent random variables uniformly distributed on $(0, 1)$.

The estimation problem in this paper is estimation of a quantile function in a fixed design regression problem. Usually this kind of problem is studied in the literature in the context of random design regression, see, e.g., Yu, Lu and Stander (2003), Koenker (2005), and the literature cited therein. In the context of random design regression the rate of convergence results for local averaging estimates of conditional quantiles have been derived in Bhattacharya and Gangopadhya (1990). Results concerning estimates based on support vector machines can be found in Steinwart and Christmann (2011). The problem of estimating the maximal quantile was not considered in the papers mentioned above.

The concept of local averaging considered in this paper is also very popular in the context of nonparametric regression, cf., e.g., Nadaraya (1964, 1970), Watson (1964), Stone (1977), Devroye and Wagner (1980), Györfi (1981), Devroye (1982), Devroye and Krzyżak (1989), Devroye, Györfi, Krzyżak and Lugosi (1994) or Beirlant and Györfi (1998).

Throughout the paper we use the following notation: $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{R}_+$ are the set of positive integers, real numbers, and nonnegative real numbers, resp. For $z \in \mathbb{R}$ we denote the smallest integer greater than or equal to z by $\lceil z \rceil$. For a set A its indicator function (which takes on value one on A and zero outside of A) is denoted by I_A. $\|x\|_\infty = \max\{x^{(1)}, \ldots, x^{(d)}\}$ is the supremum norm of a vector $x = (x^{(1)}, \ldots, x^{(d)})^T \in \mathbb{R}^d$. For $z_1, \ldots, z_n \in \mathbb{R}$ the corresponding order statistics are denoted by $z_{1:n}, \ldots, z_{n:n}$, i.e., $z_{1:n}, \ldots, z_{n:n}$ is a permuation of $z_1, \ldots, z_n$ satisfying $z_{1:n} \leq \cdots \leq z_{n:n}$. Similarly, the order statistics of real valued random variables are defined. If Z_n are real valued random variables and $\delta_n > 0$ are positive real numbers, we write $Z_n = O_{\mathbf{P}}(\delta_n)$ if

$$\lim_{c \to \infty} \limsup_{n \to \infty} \mathbf{P}\{|Z_n| > c \cdot \delta_n\} = 0.$$

The definition of our estimate of the maximum quantile is given in Section 2, the main result is formulated in Section 3 and proven in Section 5. The finite sample size behaviour of the estimate is illustrated in Section 4 using simulated data.

2 Definition of the estimate

In order to simplify the notation we assume in the sequel for all of our theoretical considerations that the set of parameters is given by $\mathcal{P} = [0, 1]^l$.

Let $X, X_1, X_2, \ldots$ be independent and identically distributed and let $p_1, \ldots, p_n$ be equidistantly chosen from $\mathcal{P}$. In the sequel we estimate the maximum quantile $\sup_{p \in \mathcal{P}} q_{p,\alpha}$ from the data

$$\mathcal{D}_n = \{((p_1, X_1), m(p_1, X_1)), \ldots, ((p_n, X_n), m(p_n, X_n))\}.$$

For $p \in \mathcal{P}$ we estimate the cumulative distribution function

$$G_p(y) = \mathbf{P}\{m(p, X) \le y\}$$

by the empirical cumulative distribution function based on all those $m(p_i, X_i)$ where the distance between p_i and p in supremum norm is at most h_n. Here $h_n > 0$ is a parameter of the estimate. In order to define this estimate, let

$$K = I_{[-1,1]^l}$$

be the naive kernel. Then

$$\hat{G}_p(y) = \frac{\sum_{i=1}^n I_{\{m(p_i,X_i) \le y\}} \cdot K\left(\frac{p-p_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{p-p_j}{h_n}\right)}$$

is our estimate of $G_p(y)$.

Next, the quantile

$$q_{p,\alpha} = \min\{y \in \mathbb{R} : G_p(y) \ge \alpha\}$$

is estimated by the plug-in estimate

$$\hat{q}_{p,\alpha} = \min\left\{y \in \mathbb{R} : \hat{G}_p(y) \ge \alpha\right\}.$$

Finally, we estimate the maximal quantile

$$\sup_{p \in \mathcal{P}} q_{p,\alpha}$$

by the maximum of the estimated quantile $\hat{q}_{p,\alpha}$ for $p \in \{p_1, \dots, p_n\}$, i.e., by

$$\hat{M}_n = \max_{i=1,\dots,n} \hat{q}_{p_i,\alpha}. \tag{1}$$

3 Main results

Our main result is the following bound on the error of our maximal quantile estimate.

Theorem 1 *Let $\alpha \in (0,1)$, let $\mathcal{P} = [0,1]^l$ and let $m : \mathcal{P} \times \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Let X be a $\mathbb{R}^d$-valued random variable and for $p \in \mathcal{P}$ let G_p be the cumulative distribution function of $m(p, X)$. Let*

$$q_{p,\alpha} = \min\{y \in \mathbb{R} : G_p(y) \ge \alpha\}$$

be the α-quantile of $m(p, X)$.

Assume that there exists $\delta > 0$ such that for some $c_1 > 0$ for all $p \in \mathcal{P}$ the derivative of G_p exists on $[q_{p,\alpha} - \delta, q_{p,\alpha} + \delta]$ and is continuous on this interval and greater than c_1. Let the estimate $\hat{M}_n$ be defined as in (1) for some $h_n > 0$ satisfying

$$h_n \to 0 \quad (n \to \infty) \quad \text{and} \quad n \cdot h_n^l \to \infty \quad (n \to \infty).$$

Then

$$\left|\hat{M}_n - \sup_{p\in\mathcal{P}} q_{p,\alpha}\right| = O_{\mathbf{P}}\left(\frac{\log(n)}{\sqrt{n\cdot h_n^l}} + \sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\le h_n} |q_{p_1,\alpha} - q_{p_2,\alpha}|\right).$$

If we impose some smoothness condition on the function describing the dependency of the values of the quantiles on the parameter p, we can derive from the above theorem the following rate of convergence result.

Corollary 1 *Assume that the assumptions of Theorem 1 hold and that, in addition, $q_{p,\alpha}$ is (r,C)-smooth as a function of $p\in\mathcal{P}$ for some $r\in(0,1]$ and $C>0$, i.e.,*

$$|q_{p_1,\alpha} - q_{p_2,\alpha}| \le C\cdot\|p_1-p_2\|_\infty^r \quad (p_1,p_2\in\mathcal{P}).$$

Set

$$h_n = c\cdot\left(\frac{\log^2(n)}{n}\right)^{1/(2r+l)}$$

and define the estimate $\hat{M}_n$ as in Section 2. Then

$$\left|\hat{M}_n - \sup_{p\in\mathcal{P}} q_{p,\alpha}\right| = O_{\mathbf{P}}\left(\left(\frac{\log^2(n)}{n}\right)^{r/(2r+l)}\right).$$

Proof. Since $q_{p,\alpha}$ is (r,C)-smooth as a function of $p\in\mathcal{P}$ we can conclude from Theorem 1

$$\left|\hat{M}_n - \sup_{p\in\mathcal{P}} q_{p,\alpha}\right| = O_{\mathbf{P}}\left(\frac{\log(n)}{\sqrt{n\cdot h_n^l}} + C\cdot h_n^r\right).$$

The definition of h_n implies the result. □

Remark 1. The rate of convergence in Corollary 1 is (up to some logarithmic factor) the same as the optimal minimax rate of convergence for estimation of a (r,C)–smooth function on a compact subset of $\mathbb{R}^l$ in supremum norm derived in Stone (1982).

Remark 2. In any application of the above estimate we have to select the bandwidth h_n of the estimate in a data-driven way. We suggest to choose the bandwidth in a optimal way in view of estimation of $G_p(y)$ by $\hat{G}_p(y)$ ($p\in\mathcal{P}$) for suitable chosen $y\in\mathbb{R}$. To do this, we propose to use a version of the well-known splitting of the sample technique in nonparametric regression (cf., e.g., Chapter 7 in Györfi et al. (2002)). More precisely, let us assume that we have available n additional random variables $\bar{X}_1, \dots, \bar{X}_n$ such that $X, X_1, \dots, X_n, \bar{X}_1, \dots, \bar{X}_n$ are independent and identically distributed, and that we observe $m(p_i,\bar{X}_i)$ for these random variables. We choose y as the α-quantile of the empirical cumulative distribution function corresponding to $m(p_1,X_1)$, $\dots$, $m(p_n,X_n)$, and choose h_n by minimizing

$$\frac{1}{n}\sum_{i=1}^n \left|I_{\{m(p_i,\bar{X}_i)\le y\}} - \hat{G}_{p_i}(y)\right|^2.$$

In the next section we will investigate the performance of this data-driven way of choosing the bandwidth using simulated data.

4 Application to simulated data

In this section we illustrate the finite sample size behaviour of our estimate (in particular in view of the data-driven choice of the bandwidth in Remark 2) with the aid of simulated data.

In all simulations we choose the sample size for the estimate of the maximal quantile as $n = 20,000$, where we use half of the data to choose the bandwidth of the estimate from a finite set of bandwidths as described in Remark 2. In each simulation we compute the absolute difference between the true value of the maximal quantile and firstly the estimate with the data-driven choice of the bandwidth and secondly with the estimate with sample size $n/2$ applied with that value of the bandwidth which leads to the minimal error for the given set of data. The level of the quantile is chosen as $\alpha = 0.95$. We repeat each simulation 100 times and report the mean values and the standard deviations of the resulting 100 error values for the two estimates.

In our first example we choose X as a standard normally distributed random variable, $\mathcal{P} = [0,1]$ and define

$$m(p, X) = 0.5 \cdot \sin(2 \cdot \pi \cdot p) + X.$$

Hence $m(p,X)$ is normally distributed with mean $0.5 \cdot \sin(2 \cdot \pi \cdot p)$ and variance 1. In this case the $\alpha = 0.95$ quantile of $m(p,X)$ is given by $q_{p,\alpha} \approx 0.5 \cdot \sin(2 \cdot \pi \cdot p) + 1.64$. We choose in this example the bandwidth from the set $\{0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4\}$. In Figure 2 we show a typical simulation for the two quantile estimates using the data-driven bandwidth and the optimal bandwidth, respectively. The mean values (standard deviations) of our errors are in case of our data driven choice of the bandwidth approximately given by 0.044 (0.034) as compared to 0.017 (0.015) in case of the optimal bandwidth choice, which is not applicable in practice since it depends on the unknown maximal value. Hence the mean absolute error of our estimate using the data-driven choice of the bandwidth is approximately only 2.5-times larger than the mean absolute error of the estimate using the optimal choice of the bandwidth, which is not known in practice.

In our second example we again choose random variable X with a standard normal distribution and $\mathcal{P} = [0,1]$, but this time we define m by

$$m(p, x) = \exp(p + x),$$

which implies that $m(p,X)$ is lognormally distributed and its logarithm has mean p and variance 1. Again the α-quantiles are known and can be computed with a statistics package. The bandwidth is chosen from the same set as in the first example. In Figure 3 we show typical simulation for the

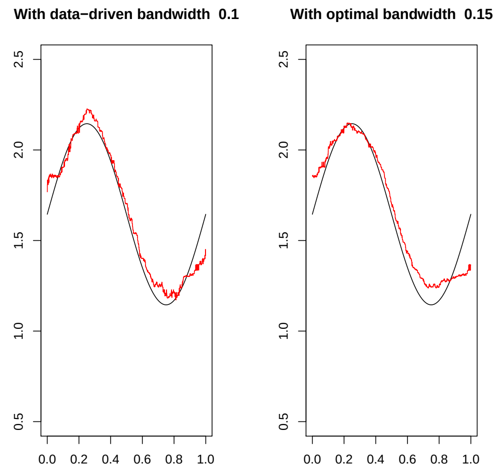

Figure 2: Typical simulation in the first model

two quantile estimates using the data-driven bandwidth and the optimal bandwidth, respectively. In this example the mean values (standard deviations) of our errors are in case of our data driven

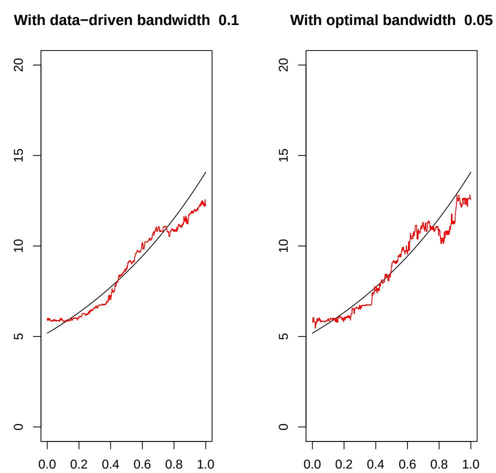

Figure 3: Typical simulation in the second model

choice of the bandwidth approximately given by 0.827 (0.517) as compared to 0.344 (0.285) in case of the optimal bandwidth choice, hence the mean error of the data-driven bandwidth choice is within a factor of 2.5 of the mean value of the optimal (and in practice not available) bandwidth choice.

In our third example we choose $X = (X^{(1)}, X^{(2)}, X^{(3)})^T$ where $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$ are independent standard normally distributed random variables, $\mathcal{P} = [-0.5, 0.5]^3$ and

$$m(p, x) = (p^{(1)} + x^{(1)})^2 + (p^{(2)} + x^{(2)})^2 + (p^{(3)} + x^{(1)})^2.$$

In this case $m(p, X)$ is non-central chi-square random variable with 3 degrees of freedom and non-centrality factor $(p^{(1)})^2 + (p^{(2)})^2 + (p^{(3)})^2$. Again the α-quantiles are known and can be computed with a statistics package. The bandwidth is chosen from the same set as above. In this example the mean values (standard deviations) of our errors are in case of our data driven choice of the bandwidth approximately given by 0.399 (0.376) as compared to 0.207 (0.144) in case of the optimal bandwidth choice, hence the mean error of the data-driven bandwidth choice is within a factor of approximately 2 of the mean value of the optimal (not available in practice) bandwidth choice.

Finally we apply our newly proposed estimation method in the application described in Section 1. The numerical simulation model of the truss structure in Figure 1 is obtained by finite elements with the commercial software ANSYS. Each of the 24 beams is modeled with the same parameters E, r and l. Timoshenko beam theory is applied to all beams by using BEAM189 elements in ANSYS. The 10 connecting nodes are modeled as rigid links using MPC184 elements. Deflection is constrained at node 2 in x-, y- and z-direction, at node 3 in y- and z-direction and at node 4 in z-direction such that the truss is supported statically determined.

The beam parameters E, r and l are assumed to be in the intervals given in Table 1. Random

Table 1: Uncertain parameters

parameter p	value	unit
Young's modulus E	$[69.7, 73.7] \cdot 10^9$	$\mathrm{N/m^2}$
radius r	$[0.0049, 0.0051]$	m
length l	$[0.3995, 0.4005]$	m

loading is applied to node 1 as a single force F_z in z-direction and two symmetric moments with magnitude M that act around axes in the x-y-plane in an angle of $+60°$ and $-60°$ from x-direction at nodes 6 and 7, respectively. The load F_z is normally distributed with mean value μ_{Fz} and standard deviation σ_{Fz}. The moments M have truncated normal distribution restricted to positive half axis, where the underlying normal distribution has mean value $\mu_M = 0$ and standard deviation σ_M, see Table 2. To obtain the maximum equivalent stresses $\sigma_{\mathrm{v},1}$, $\sigma_{\mathrm{v},2}$ and $\sigma_{\mathrm{v},3}$, a static analysis is carried out for each parameter set and load set. Two independent random load sets were generated for each combination of 41 values of each parameter set distributed equidistantly in the parameter intervals. In total, $2 \cdot 41^3$ deterministic calculations were carried out.

Application of the newly proposed estimate of the maximum 0.95-quantile to this data results in predicted maximum stresses of $7.05 \cdot 10^6$ $\mathrm{N/m^2}$, $17.67 \cdot 10^6$ $\mathrm{N/m^2}$ and $17.67 \cdot 10^6$ $\mathrm{N/m^2}$ in the

Table 2: Uncertain loads

load X	value	unit
load F_z	$\mu_{Fz} = -1000$, $\sigma_{Fz} = 33.33$	N
moments M	$\mu_M = 0$, $\sigma_M = 3.33$	Nm

three considered beams, respectively. In contrast, if we ignore the uncertainty in the parameters and just compute 0.95-quantiles using the mean values of these parameters and a data set of 41^3 corresponding values, we get instead $6.78 \cdot 10^6$ N/m^2, $16.80 \cdot 10^6$ N/m^2 and $16.80 \cdot 10^6$ N/m^2, hence we get values which are 3.8%, 4.9% and 4.9% lower, respectively.

5 Proofs

5.1 Auxiliary results

In this subsection we formulate and prove two lemmas which we will need in the proof of Theorem 1. Our first lemma shows that if we mix coupled samples from several different distributions, the empirical quantile of the mixed samples lies with probability one between the minimal and the maximal quantiles of the individual samples. Here we use the fact that we do not change the joint distribution of our samples if we assume that the random variables are defined by applying the inverse of the cumulative distribution function to uniformly distributed random variables.

More precisely, our setting is the following: Let $M \in \mathbb{N}$ be the number of samples which we mix, and for $k \in \{1, \ldots, M\}$ let $H^{(k)} : (0,1) \to \mathbb{R}$ be a monotonically increasing function (which we will choose later as the inverse function of a cumulative distribution function). Let $N \in \mathbb{N}$, let $u_1, \ldots, u_N \in (0,1)$ and denote by $u_{1:N}, \ldots, u_{N:N}$ the corresponding order statistics. We define M different coupled samples by

$$z_i^{(k)} = H^{(k)}(u_i) \quad (i = 1, \ldots, N)$$

for $k \in \{1, \ldots, M\}$, and we mix them by choosing

$$z_i \in \{z_i^{(1)}, \ldots, z_i^{(M)}\}$$

for $i = 1, \ldots, N$.

Let

$$\hat{G}^{(k)}(y) = \frac{1}{N} \cdot \sum_{i=1}^{N} I_{\{z_i^{(k)} \leq y\}} \quad \text{and} \quad \hat{G}(y) = \frac{1}{N} \cdot \sum_{i=1}^{N} I_{\{z_i \leq y\}}$$

be the empirical cumulative distribution functions corresponding to $z_1^{(k)}, \ldots, z_N^{(k)}$ and $z_1, \ldots, z_N$, resp., and let

$$\hat{q}_\alpha^{(k)} = \min\left\{y \in \mathbb{R} \, : \, \hat{G}^{(k)}(y) \geq \alpha\right\} \quad \text{and} \quad \hat{q}_\alpha = \min\left\{y \in \mathbb{R} \, : \, \hat{G}(y) \geq \alpha\right\}$$

be the corresponding quantile estimates. Then the following result holds.

Lemma 1 *Define* $\hat{q}_\alpha^{(k)}$ *and* $\hat{q}_\alpha$ *as above. Then*

$$\min_{k=1,\ldots,M} \hat{q}_\alpha^{(k)} \leq \hat{q}_\alpha \leq \max_{k=1,\ldots,M} \hat{q}_\alpha^{(k)}.$$

Proof. By construction we have

$$\hat{q}_\alpha^{(k)} = H^{(k)}(u_{\lceil N\cdot\alpha\rceil:N}) \tag{2}$$

for $k \in \{1, \ldots, M\}$. Furthermore, for $k_1, \ldots, k_N \in \{1, \ldots, M\}$ suitably chosen we have that

$$H^{(k_1)}(u_{1:N}), \ldots, H^{(k_N)}(u_{N:N})$$

is a permuation of $z_1, \ldots, z_N$.

Choose $k_{min}, k_{max} \in \{1, \ldots, M\}$ such that

$$\min_{k=1,\ldots,M} \hat{q}_\alpha^{(k)} = \hat{q}_\alpha^{(k_{min})} \quad \text{and} \quad \max_{k=1,\ldots,M} \hat{q}_\alpha^{(k)} = \hat{q}_\alpha^{(k_{max})}.$$

Then by (2) we have for any $k \in \{1, \ldots, M\}$

$$H^{(k_{min})}(u_{\lceil N\cdot\alpha\rceil:N}) \leq H^{(k)}(u_{\lceil N\cdot\alpha\rceil:N}) \leq H^{(k_{max})}(u_{\lceil N\cdot\alpha\rceil:N}).$$

Using this relation and the monotonicity of $H^{(k)}$ we get for any $i \in \{1, \ldots, \lceil N \cdot \alpha\rceil\}$

$$\begin{aligned} H^{(k_i)}(u_{i:N}) &\leq H^{(k_i)}(u_{\lceil N\cdot\alpha\rceil:N}) \\ &\leq H^{(k_{max})}(u_{\lceil N\cdot\alpha\rceil:N}) \\ &= \max_{k=1,\ldots,M} \hat{q}_\alpha^{(k)} \end{aligned}$$

and for any $i \in \{\lceil N \cdot \alpha\rceil, \ldots, N\}$

$$\begin{aligned} H^{(k_i)}(u_{i:N}) &\geq H^{(k_i)}(u_{\lceil N\cdot\alpha\rceil:N}) \\ &\geq H^{(k_{min})}(u_{\lceil N\cdot\alpha\rceil:N}) \\ &= \min_{k=1,\ldots,M} \hat{q}_\alpha^{(k)}. \end{aligned}$$

Now, let $z_{1:N}, \ldots, z_{N:N}$ be the order statistics of $z_1, \ldots, z_n$. Then the above inequalities imply

$$\min_{k=1,\ldots,M} \hat{q}_\alpha^{(k)} \leq z_{\lceil N\cdot\alpha\rceil:N} \leq \max_{k=1,\ldots,M} \hat{q}_\alpha^{(k)}.$$

Since $z_{\lceil N\cdot\alpha\rceil:N} = \hat{q}_\alpha$ this completes the proof. □

Our second lemma relates the error of the quantile estimate to the behavior of order statistics corresponding to some fixed sequence of independent random variables distributed uniformly on $(0, 1)$.

Lemma 2 *Let $p \in \mathcal{P}$ be fixed and let the estimate $\hat{q}_{p,\alpha}$ be defined as in Section 2. Assume that there exists a constant $c_1 > 0$ such that for some $\delta > 0$ we have that $G_{\bar{p}}$ is continuously differentiable on $[q_{\bar{p},\alpha} - \delta, q_{\bar{p},\alpha} + \delta]$ with derivative bounded from below by c_1, for all $\bar{p} \in \mathcal{P}$ satisfying $\|\bar{p} - p\|_\infty \leq \delta$.*

Let

$$N = |\{1 \leq i \leq n \, : \, \|p - p_i\|_\infty \leq h_n\}|$$

be the number of parameters p_i which have at most supnorm distance h_n to p. Let $U_1, \ldots, U_N$ be independent random variables distributed uniformly on $(0,1)$ and denote their order statistics by $U_{1:N}, \ldots, U_{N:N}$. Then we have for any $0 < \epsilon < \delta$ satisfying $\delta \geq h_n$:

$$\mathbf{P}\left\{ |\hat{q}_{p,\alpha} - q_{p,\alpha}| > \epsilon + \sup_{\bar{p} \in \mathcal{P}: \|\bar{p} - p\|_\infty \leq h_n} |q_{\bar{p},\alpha} - q_{p,\alpha}| \right\} \leq 2 \cdot \mathbf{P}\left\{ \left| U_{\lceil N \cdot \alpha \rceil : N} - \alpha \right| > c_1 \cdot \epsilon \right\}.$$

Proof. For $\bar{p} \in \mathcal{P}$ let

$$G_{\bar{p}}^{-1}(u) = \min\{y \in \mathbb{R} \, : \, G_{\bar{p}}(y) \geq u\} \quad (u \in (0,1))$$

be the generalized inverse of $G_{\bar{p}}$. Then it is well-known that $G_{\bar{p}}^{-1}(U_1)$ has the same distribution as $m(\bar{p}, X_1)$. Since $\hat{G}_p$ is by definition (observe that K is the naive kernel) the empirical cumulative distribution function of $m(p_{i_1}, X_{i_1}), \ldots, m(p_{i_N}, X_{i_N})$ for suitable chosen pairwise distinct $i_1, \ldots, i_N \in \{1, \ldots, n\}$, we see that it has the same distribution as the empirical cumulative distribution function $\hat{G}_N$ of $G_{p_{i_1}}^{-1}(U_1), \ldots, G_{p_{i_N}}^{-1}(U_N)$. Consequently, $\hat{q}_{p,\alpha}$ has the same distribution as the quantile estimate $\hat{q}_\alpha^U$ corresponding to $\hat{G}_N$ which implies for any $\eta > 0$

$$\mathbf{P}\{|\hat{q}_{p,\alpha} - q_{p,\alpha}| > \eta\} = \mathbf{P}\left\{ \left|\hat{q}_\alpha^U - q_{p,\alpha}\right| > \eta \right\}.$$

Let

$$\hat{G}^{(k)}(y) = \frac{1}{N} \sum_{i=1}^N I_{\{G_{p_{i_k}}^{-1}(U_i) \leq y\}}$$

be the cumulative emprirical distribution function of $G_{p_{i_k}}^{-1}(U_1), \ldots, G_{p_{i_k}}^{-1}(U_N)$ and denote by $\hat{q}_\alpha^{(k)}$ the corresponding quantile estimate ($k = 1, \ldots, N$). Application of Lemma 1 yields for any $\eta > 0$

$$\begin{aligned} \mathbf{P}\left\{ \left|\hat{q}_\alpha^U - q_{p,\alpha}\right| > \eta \right\} &\leq \mathbf{P}\left\{ \left| \min_{k=1,\ldots,N} \hat{q}_\alpha^{(k)} - q_{p,\alpha} \right| > \eta \right\} + \mathbf{P}\left\{ \left| \max_{k=1,\ldots,N} \hat{q}_\alpha^{(k)} - q_{p,\alpha} \right| > \eta \right\} \\ &\leq 2 \cdot \mathbf{P}\left\{ \max_{k=1,\ldots,N} \left| \hat{q}_\alpha^{(k)} - q_{p_{i_k},\alpha} \right| + \max_{k=1,\ldots,N} \left| q_{p_{i_k},\alpha} - q_{p,\alpha} \right| > \eta \right\}. \end{aligned}$$

Summarizing the above results we see that we have shown

$$\begin{aligned} &\mathbf{P}\left\{ |\hat{q}_{p,\alpha} - q_{p,\alpha}| > \epsilon + \sup_{\{\bar{p} \in \mathcal{P}: \|\bar{p} - p\|_\infty \leq h_n\}} |q_{\bar{p},\alpha} - q_{p,\alpha}| \right\} \\ &\leq 2 \cdot \mathbf{P}\left\{ \max_{k=1,\ldots,N} \left| \hat{q}_\alpha^{(k)} - q_{p_{i_k},\alpha} \right| > \epsilon \right\} \\ &= 2 \cdot \mathbf{P}\left\{ \max_{k=1,\ldots,N} \left| G_{p_{i_k}}^{-1}(U_{\lceil N \cdot \alpha \rceil : N}) - G_{p_{i_k}}^{-1}(\alpha) \right| > \epsilon \right\}. \end{aligned}$$

By the assumptions of Lemma 2, on $[q_{p_{i_k},\alpha} - \delta, q_{p_{i_k},\alpha} + \delta]$ the derivative of $G_{p_{i_k}}$ exists and is bounded from below by $c_1 > 0$. Consequently, on $[\alpha - c_1 \cdot \delta, \alpha + c_1 \cdot \delta]$, the inverse function of $G_{p_{i_k}}$ is continuously differentiable with derivative bounded from above by $1/c_1$. So in case that $\left|U_{\lceil N\cdot\alpha\rceil:N} - \alpha\right| < c_1 \cdot \delta$ we can conclude from the mean value theorem that we have

$$\left|G_{p_{i_k}}^{-1}(U_{\lceil N\cdot\alpha\rceil:N}) - G_{p_{i_k}}^{-1}(\alpha)\right| \leq \frac{1}{c_1} \cdot \left|U_{\lceil N\cdot\alpha\rceil:N} - \alpha\right|.$$

This implies

$$\mathbf{P}\left\{\max_{k=1,\dots,N}\left|G_{p_{i_k}}^{-1}(U_{\lceil N\cdot\alpha\rceil:N}) - G_{p_{i_k}}^{-1}(\alpha)\right| > \epsilon\right\} \leq \mathbf{P}\left\{\left|U_{\lceil N\cdot\alpha\rceil:N} - \alpha\right| > c_1 \cdot \epsilon\right\},$$

which completes the proof. □

5.2 Proof of Theorem 1

Since

$$\left|\hat{M}_n - \sup_{p\in\mathcal{P}} q_{p,\alpha}\right| \quad \leq \quad \left|\hat{M}_n - \max_{i=1,\dots,n} q_{p_i,\alpha}\right| + \left|\max_{i=1,\dots,n} q_{p_i,\alpha} - \sup_{p\in\mathcal{P}} q_{p,\alpha}\right|$$

it suffices to show

$$\left|\hat{M}_n - \max_{i=1,\dots,n} q_{p_i,\alpha}\right| = O_{\mathbf{P}}\left(\frac{\log(n)}{\sqrt{n^{1/l}\cdot h_n}} + \sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq h_n} |q_{p_1,\alpha} - q_{p_2,\alpha}|\right) \tag{3}$$

and

$$\left|\max_{i=1,\dots,n} q_{p_i,\alpha} - \sup_{p\in\mathcal{P}} q_{p,\alpha}\right| = O\left(\frac{\log(n)}{\sqrt{n^{1/l}\cdot h_n}} + \sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq h_n} |q_{p_1,\alpha} - q_{p_2,\alpha}|\right). \tag{4}$$

Because of

$$\begin{aligned}\left|\max_{i=1,\dots,n} q_{p_i,\alpha} - \sup_{p\in\mathcal{P}} q_{p,\alpha}\right| &= \sup_{p\in\mathcal{P}} q_{p,\alpha} - \max_{i=1,\dots,n} q_{p_i,\alpha}\\ &= \sup_{p\in\mathcal{P}} \min_{i=1,\dots,n} (q_{p,\alpha} - q_{p_i,\alpha})\\ &\leq \sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq 1/n^{1/l}} |q_{p_1,\alpha} - q_{p_2,\alpha}|,\end{aligned}$$

(4) follows from $n^{-1/l} \leq h_n$ for n sufficiently large, which is implied by $n \cdot h_n^l \to \infty$ $(n \to \infty)$.

In order to prove (3) we observe that for $a_i, b_i \in \mathbb{R}$ $(i = 1, \dots, n)$ we have

$$\left|\max_{i=1,\dots,n} a_i - \max_{i=1,\dots,n} b_i\right| \leq \max_{i=1,\dots,n} |a_i - b_i|, \tag{5}$$

which together with the union bound implies for arbitrary $\epsilon > 0$

$$\begin{aligned}&\mathbf{P}\left\{\left|\hat{M}_n - \max_{i=1,\dots,n} q_{p_i,\alpha}\right| > \epsilon + \sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq h_n} |q_{p_1,\alpha} - q_{p_2,\alpha}|\right\}\\ &\leq \mathbf{P}\left\{\max_{i=1,\dots,n} |\hat{q}_{p_i,\alpha} - q_{p_i,\alpha}| > \epsilon + \sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq h_n} |q_{p_1,\alpha} - q_{p_2,\alpha}|\right\}\\ &\leq n \cdot \max_{i=1,\dots,n} \mathbf{P}\left\{|\hat{q}_{p_i,\alpha} - q_{p_i,\alpha}| > \epsilon + \sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq h_n} |q_{p_1,\alpha} - q_{p_2,\alpha}|\right\}.\end{aligned} \tag{6}$$

Let U_i be defined as in Lemma 2 and let F_N be the empirical cumulative distribution function corresponding to $U_1, \dots, U_N$ and let F be the cumulative distribution function of U_1. Then

$$\begin{aligned} \mathbf{P}\left\{\left|U_{\lceil N\cdot\alpha\rceil:N}-\alpha\right|>c_1\cdot\epsilon\right\} &= \mathbf{P}\left\{\left|F(U_{\lceil N\cdot\alpha\rceil:N})-F_N(U_{\lceil N\cdot\alpha\rceil:N})+\frac{\lceil N\cdot\alpha\rceil}{N}-\alpha\right|>c_1\cdot\epsilon\right\} \\ &\leq \mathbf{P}\left\{\sup_{t\in\mathbb{R}}|F_N(t)-F(t)|>c_1\cdot\epsilon-\frac{1}{N}\right\}. \end{aligned}$$

In case that $c_1 \cdot \epsilon > \frac{1}{N}$ we can bound the last probability using well-known results from Vapnik-Chervonenkis theory (cf., e.g., Theorem 12.4 in Devroye, Györfi and Lugosi (1996)) and get

$$\mathbf{P}\left\{\left|U_{\lceil N\cdot\alpha\rceil:N}-\alpha\right|>c_1\cdot\epsilon\right\}\leq 8\cdot(N+1)\cdot\exp\left(-N\cdot\left(c_1\cdot\epsilon-\frac{1}{N}\right)^2/32\right).$$

Using this, $N \approx n \cdot h_n^l$, Lemma 2 and (6) we conclude

$$\begin{aligned} &\mathbf{P}\left\{\left|\hat{M}_n-\max_{i=1,\dots,n}q_{p_i,\alpha}\right|>\epsilon+\sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq h_n}|q_{p_1,\alpha}-q_{p_2,\alpha}|\right\} \\ &\leq c_3\cdot n\cdot(n\cdot h_n^l+1)\cdot\exp\left(-c_3\cdot n\cdot h_n^l\cdot\left(c_1\cdot\epsilon-\frac{1}{n\cdot h_n^l}\right)^2\right), \end{aligned}$$

which implies

$$\left|\hat{M}_n-\max_{i=1,\dots,n}q_{p_i,\alpha}\right|=O_{\mathbf{P}}\left(\frac{\log(n)}{n\cdot h_n^l}+\sup_{p_1,p_2\in\mathcal{P}\,:\,\|p_1-p_2\|_\infty\leq h_n}|q_{p_1,\alpha}-q_{p_2,\alpha}|\right).$$

This completes the proof. □

6 Acknowledgements

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre SFB 805.

References

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

[2] Beirlant, J. and Györfi, L. (1998). On the asymptotic L_2-error in partitioning regression estimation. *Journal of Statistical Planning and Inference*, **71**, pp. 93–107.

[3] Devroye, L. (1982). Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **61**, pp. 467–481.

[4] Devroye, L., Györfi, L., Krzyżak, A., and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics*, **22**, pp. 1371–1385.

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

[6] Devroye, L. and Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference*, **23**, pp. 71–82.

[7] Devroye, L. and Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics*, **8**, pp. 231–239.

[8] Gross, D., Hauger, W., Schröder, J. and Wall, W. A. (2007). *Technische Mechanik Band 2: Elastostatik.* Springer-Verlag, New York.

[9] Györfi, L. (1981). Recent results on nonparametric regression estimate and multiple classification. *Problems of Control and Information Theory*, **10**, pp. 43–52.

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

[11] Koenker, R. (2005). *Quantile regression.* Cambridge University Press.

[12] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, pp. 141–142.

[13] Nadaraya, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves. *Theory of Probability and its Applications*, **15**, pp. 134–137.

[14] Steinwart, I. and Christmann, A. (2011). Estimation conditional quantiles with the help of the pinball loss. *Bernoulli* **17**, pp. 211-225.

[15] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statististics*, **5**, pp. 595–645.

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

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

[18] Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: application and current research areas. *Journal of the Royal Statistical Society, Series D* **52**, pp. 331–350.