Nonparametric quantile estimation using importance sampling

Michael Kohler[1], Adam Krzyżak[2,*], Reinhard Tent[1] and Harro Walk[3]

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

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

[3] *Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany, email: walk@mathematik.uni-stuttgart.de.*

March 31, 2014

Abstract

Nonparametric estimation of a quantile of a random variable $m(X)$ is considered, where $m : \mathbb{R}^d \to \mathbb{R}$ is a function which is costly to compute and X is a $\mathbb{R}^d$-valued random variable with given density. An importance sampling quantile estimate of $m(X)$, which is based on a suitable estimate m_n of m, is defined, and it is shown that this estimate achieves a rate of convergence of order $\log^{1.5}(n)/n$. The finite sample size behavior of the estimate is illustrated by simulated data.

AMS classification: Primary 62G05; secondary 62G30.

Key words and phrases: Nonparametric quantile estimation, importance sampling, rate of convergence.

1 Introduction

In this paper we consider a simulation model of a complex technical system described by

$$Y = m(X),$$

where X is a $\mathbb{R}^d$-valued random variable with density $f : \mathbb{R}^d \to \mathbb{R}$ and $m : \mathbb{R}^d \to \mathbb{R}$ is a known function which is expensive to evaluate. Let

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

Running title: *Nonparametric extreme quantile estimation*

$$G(y) = \mathbf{P}\{Y \leq y\} = \mathbf{P}\{m(X) \leq y\}$$

be the cumulative distribution function (cdf) of Y. For $\alpha \in (0,1)$ we are interested in estimating quantiles of the form

$$q_\alpha = \inf\{y \in \mathbb{R} : G(y) \geq \alpha\}$$

using at most n evaluations of the function m. Here we assume that the density f of X is known.

A simple idea is to estimate q_α using an i.i.d. sample $X_1, \ldots, X_n$ of X and to compute the empirical cdf

$$G_{m(X),n}(y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{m(X_i) \leq y\}} \tag{1}$$

and to use the corresponding plug-in estimate

$$\bar{q}_{\alpha,n} = \inf\{y \in \mathbb{R} : G_{m(X),n}(y) \geq \alpha\}. \tag{2}$$

Set $Y_i = m(X_i)$ $(i = 1, \ldots, n)$ and let $Y_{1:n}, \ldots, Y_{n:n}$ be the order statistics of $Y_1, \ldots, Y_n$, i.e., $Y_{1:n}, \ldots, Y_{n:n}$ is a permutation of $Y_1, \ldots, Y_n$ such that

$$Y_{1:n} \leq \ldots \leq Y_{n:n}.$$

Since

$$\bar{q}_{\alpha,n} = Y_{\lceil n\alpha \rceil:n}$$

is in fact an order statistic, the properties of this estimate can be studied using the results from order statistics. In particular Theorem 8.5.1 in Arnold, Balakrishnan and Nagaraja (1992) implies that in case that $m(X)$ has a density g which is continuous and positive at q_α we have

$$\sqrt{n} \cdot g(q_\alpha) \cdot \frac{Y_{\lceil n\alpha \rceil:n} - q_\alpha}{\sqrt{\alpha \cdot (1-\alpha)}} \to N(0,1) \quad \text{in distribution.}$$

This implies

$$\mathbf{P}\left\{ |\bar{q}_{\alpha,n} - q_\alpha| > \frac{c_n}{\sqrt{n}} \right\} \to 0 \quad (n \to \infty) \tag{3}$$

whenever $c_n \to \infty$ $(n \to \infty)$.

In this paper we apply importance sampling (IS) to obtain a better estimate of q_α. Importance sampling is a technique to improve estimation of the expectation of a function $\phi : \mathbb{R}^d \to \mathbb{R}$ by sample averages. Instead of using an independent and identically distributed sequence $X, X_1, X_2, \ldots$ and estimating $\mathbf{E}\phi(X)$ by

$$\frac{1}{n} \sum_{i=1}^{n} \phi(X_i),$$

one can use importance sampling, where a new random variable Z with a density h satisfying for all $x \in \mathbb{R}^d$

$$\phi(x) \cdot f(x) \neq 0 \quad \Rightarrow \quad h(x) \neq 0$$

is chosen and for $Z, Z_1, Z_2, \dots$ independent and identically distributed

$$\mathbf{E}\{\phi(X)\} = \mathbf{E}\left\{\phi(Z) \cdot \frac{f(Z)}{h(Z)}\right\}$$

is estimated by

$$\frac{1}{n} \sum_{i=1}^{n} \phi(Z_i) \cdot \frac{f(Z_i)}{h(Z_i)}, \tag{4}$$

whereas we assume that $\frac{0}{0} = 0$. Here the aim is to choose h such that the variance of (4) is small (see for instance Chapter 4.6 in Glasserman (2004), Nedermayer (2009) and the literature cited therein).

Quantile estimation using importance sampling has been considered by Cannamela, Garnier and Iooss (2008), Egloff and Leippold (2010) and Morio (2012). All three papers proposed new estimates in various models, however only Egloff and Leippold (2010) investigated theoretical properties (consistency) of their method. None of the papers contain any results on the rates of convergence.

In this paper we propose a new importance sampling quantile estimate and analyze its rates of convergence. The basic idea is to use an initial estimate of the quantile based on the order statistics of samples of $m(X)$ in order to determine an interval $[a_n, b_n]$ containing the quantile. Then we construct an estimate m_n of m and restrict f to the inverse image $m_n^{-1}([a_n, b_n])$ of $[a_n, b_n]$ to construct a new random variable Z. Our final estimate of the quantile is then defined as an order statistic of $m(Z)$. Under suitable assumptions on the smoothness of m and on the tails of f we are able to show that this estimate achieves the rate of convergence of order $\frac{\log^{1.5} n}{n}$.

Throughout this paper we use the following notations: $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$ and $\mathbb{R}$ are the sets of positive integers, nonnegative integers, integers and real numbers, respectively. For a real number z we denote by $\lfloor z \rfloor$ and $\lceil z \rceil$ the largest integer less than or equal to z and the smallest integer larger than or equal to z, respectively. $\|x\|$ is the Euclidean norm of $x \in \mathbb{R}^d$. For $f : \mathbb{R}^d \to \mathbb{R}$ and $A \subseteq \mathbb{R}^d$ we set

$$\|f\|_{\infty,A} = \sup_{x \in A} |f(x)|.$$

Let $p = k + s$ for some $k \in \mathbb{N}_0$ and $0 < s \leq 1$, and let $C > 0$. A function $m : \mathbb{R}^d \to \mathbb{R}$ is called (p, C)-smooth, if for every $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$ the partial derivative $\frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$ exists and satisfies

$$\left| \frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(x) - \frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(z) \right| \leq C \cdot \|x - z\|^s$$

for all $x, z \in \mathbb{R}^d$.

For nonnegative random variables X_n and Y_n we say that $X_n = O_{\mathbf{P}}(Y_n)$ if

$$\limsup_{n \to \infty} \mathbf{P}(X_n > c_1 \cdot Y_n) = 0$$

for some finite constant $c_1 > 0$.

The estimate of the quantile is defined in Section 2. The main result is formulated in Section 3 and proofs are provided in Section 5. In Section 4 we illustrate the finite sample size performance of the estimate using simulated data.

2 Definition of the estimate

Let $n = n_1 + n_2 + n_3$ where $n_1 = n_1(n) = \lfloor n/3 \rfloor = n_2 = n_2(n)$ and $n_3 = n_3(n) = n - n_1 - n_2$. We will use n_1 evaluations of m in order to generate an initial estimate of q_α, n_2 evaluations of m to construct an approximation of m, and we will use n_3 further evaluations of m to improve our initial estimate of q_α.

Let $\bar{q}_{\alpha,n_1}$ be the quantile estimate based on order statistics introduced in Section 1. In order to improve it by importance sampling, we will use additional observations $(x_1, m(x_1))$, ..., $(x_{n_2}, m(x_{n_2}))$ of m at points $x_1, \dots, x_{n_2} \in \mathbb{R}^d$ and use an estimate

$$m_n(\cdot) = m_n(\cdot, (x_1, m(x_1)), \dots, (x_{n_2}, m(x_{n_2}))) : \mathbb{R}^d \to \mathbb{R}$$

of $m : \mathbb{R}^d \to \mathbb{R}$. Both will be specified later. Let $K_n = [-l_n, l_n]^d$ for some $l_n > 0$ such that $l_n \to \infty$ as $n \to \infty$ and assume that the supremum norm error of m_n on K_n is bounded by $\beta_n > 0$, i.e.,

$$\|m_n - m\|_{\infty, K_n} := \sup_{x \in K_n} |m_n(x) - m(x)| \leq \beta_n. \tag{5}$$

Set

$$a_n = \bar{q}_{\alpha,n_1} - 2 \cdot \frac{\log n}{\sqrt{n}} - 2 \cdot \beta_n \quad \text{and} \quad b_n = \bar{q}_{\alpha,n_1} + 2 \cdot \frac{\log n}{\sqrt{n}} + \beta_n,$$

where both quantities depend (via $\bar{q}_{\alpha,n_1}$) on the data

$$\mathcal{D}_{n_1} = \{(X_1, m(X_1)), \dots, (X_{n_1}, m(X_{n_1})\}.$$

We then replace X by a random variable Z which has the density

$$h(x) = c_2 \cdot \left(I_{\{x \in K_n \,:\, a_n \leq m_n(x) \leq b_n\}} + I_{\{x \notin K_n\}} \right) \cdot f(x)$$

where

$$c_2 = \left(\int_{\mathbb{R}^d} \left(I_{\{x \in K_n \,:\, a_n \leq m_n(x) \leq b_n\}} + I_{\{x \notin K_n\}} \right) f(x) dx \right)^{-1} = \frac{1}{1 - \gamma_1 - \gamma_2}.$$

Here

$$\gamma_1 = \mathbf{P}\{X \in K_n, m_n(X) < a_n | \mathcal{D}_{n_1}\} = \int_{\mathbb{R}^d} 1_{K_n}(x) \cdot 1_{\{x \,:\, m_n(x) < a_n\}} \cdot f(x) dx$$

and

$$\gamma_2 = \mathbf{P}\{X \in K_n, m_n(X) > b_n | \mathcal{D}_{n_1}\} = \int_{\mathbb{R}^d} 1_{K_n}(x) \cdot 1_{\{x \,:\, m_n(x) > b_n\}} \cdot f(x) dx$$

can be computed exactly for given f and m_n. In our application below we approximate them by the suitable Riemann sums. Observe that a_n and b_n depend on $\mathcal{D}_{n_1}$ and therefore the density h and the distribution of Z are random quantities. Furthermore on the event

$$\left\{ |\overline{q}_{\alpha,n_1} - q_\alpha| \leq \frac{\log n}{\sqrt{n}} \right\}$$

we have that

$$\int_{\mathbb{R}^d} \left(I_{\{x \in K_n \,:\, a_n \leq m_n(x) \leq b_n\}} + I_{\{x \notin K_n\}} \right) f(x) dx \geq \mathbf{P}\left\{ q_\alpha - \frac{\log n}{\sqrt{n}} \leq m(X) \leq q_\alpha + \frac{\log n}{\sqrt{n}} \right\} > 0, \quad (6)$$

provided, e.g., the density of $m(X)$ is positive and continuous at q_α. Hence outside of an event whose probability tends to zero for $n \to \infty$ the constant c_2 and the density h are in this case well defined. The main trick in the sequel is that we can relate the quantile q_α to a quantile of $m(Z)$ as shown in Lemma 1 below.

Lemma 1 *Assume that* (5) *holds,* $m(X)$ *has a density which is continuous and positive at* q_α *and let* Z *be a random variable defined as above. Furthermore set*

$$\bar{\alpha} = \frac{\alpha - \gamma_1}{1 - \gamma_1 - \gamma_2}$$

and

$$q_{m(Z),\bar{\alpha}} = \inf\{y \in \mathbb{R} \,:\, \mathbf{P}\{m(Z) \leq y | \mathcal{D}_{n_1}\} \geq \bar{\alpha}\}$$

where $\mathcal{D}_{n_1} = \{(X_1, m(X_1)), \dots, (X_{n_1}, m(X_{n_1})\}$. *Then we have with probability tending to one for* $n \to \infty$ *that*

$$q_\alpha = q_{m(Z),\bar{\alpha}}.$$

Let Z, Z_1, Z_2, ... be independent and identically distributed and set

$$G_{m(Z),n_3}(y) = \frac{1}{n_3} \sum_{i=1}^{n_3} I_{\{m(Z_i) \leq y\}}.$$

We estimate $q_\alpha = q_{m(Z),\bar{\alpha}}$ (which is outside of an event whose probability tends to zero for $n \to \infty$ according to Lemma 1 equal to $q_{m(Z),\bar{\alpha}}$) by

$$\begin{aligned} \bar{q}_{m(Z),\bar{\alpha},n_3} &= \inf\left\{ y \in \mathbb{R} \,:\, G_{m(Z),n_3}(y) \geq \bar{\alpha} \right\} \\ &= \inf\left\{ y \in \mathbb{R} \,:\, G_{m(Z),n_3}(y) \geq \frac{\alpha - \gamma_1}{1 - \gamma_1 - \gamma_2} \right\}. \end{aligned}$$

As before we have that $\bar{q}_{m(Z),\bar{\alpha},n_3}$ is an order statistic of $m(Z_1)$, ..., $m(Z_{n_3})$:

$$\bar{q}_{m(Z),\bar{\alpha},n_3} = m(Z)_{\lceil \bar{\alpha} \cdot n_3 \rceil : n_3}.$$

One possible choice for an estimate m_n of m is a spline approximation of m, which we introduce next. We will use well-known results from spline theory to show that if we choose the design points

$z_1, \ldots, z_n$ equidistantly in $K_n = [-l_n, l_n]^d$, then a properly defined spline approximation of a (p, C)-smooth function achieves the rate of convergence $l_n^p / n^{p/d}$.

In order to define the spline approximation, we introduce polynomial splines, i.e., sets of piecewise polynomials satisfying a global smoothness condition, and a corresponding B-spline basis consisting of basis functions with compact support as follows:

Choose $K \in \mathbb{N}$ and $M \in \mathbb{N}_0$, and set $u_k = k \cdot l_n / K$ $(k \in \mathbb{Z})$. For $k \in \mathbb{Z}$ let $B_{k,M} : \mathbb{R} \to \mathbb{R}$ be the univariate B-spline of degree M with knot sequence $(u_k)_{k \in \mathbb{Z}}$ and support $supp(B_{k,M}) = [u_k, u_{k+M+1}]$. In case $M = 0$ B-spline $B_{k,0}$ is the indicator function of the interval $[u_k, u_{k+1})$, and for $M = 1$ we have

$$B_{k,1}(x) = \begin{cases} \frac{x - u_k}{u_{k+1} - u_k} & , u_k \leq x \leq u_{k+1}, \\ \frac{u_{k+2} - x}{u_{k+2} - u_{k+1}} & , u_{k+1} < x \leq u_{k+2}, \\ 0 & , \text{elsewhere}, \end{cases}$$

(so-called hat-function). The general recursive definition of $B_{k,M}$ can be found, e.g., in de Boor (1978), or in Section 14.1 of Györfi et al. (2002). These B-splines are basis functions of sets of univariate piecewise polynomials of degree M, where the piecewise polynomials are globally $(M-1)$–times continuously differentiable and where the M-th derivatives of the functions have jump points only at the knots u_l $(l \in \mathbb{Z})$.

For $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d$ we define the tensor product B-spline $B_{\mathbf{k},M} : \mathbb{R}^d \to \mathbb{R}$ by

$$B_{\mathbf{k},M}(x^{(1)}, \ldots, x^{(d)}) = B_{k_1,M}(x^{(1)}) \cdot \ldots \cdot B_{k_d,M}(x^{(d)}) \quad (x^{(1)}, \ldots, x^{(d)} \in \mathbb{R}).$$

With these functions we define $\mathcal{S}_{K,M}$ as the set of all linear combinations of all those tensor product B-splines above, whose support has nonempty intersection with $K_n = [-l_n, l_n]^d$, i.e., we set

$$\mathcal{S}_{K,M} = \left\{ \sum_{\mathbf{k} \in \{-K-M, -K-M+1, \ldots, K-1\}^d} a_{\mathbf{k}} \cdot B_{\mathbf{k},M} \ : \ a_{\mathbf{k}} \in \mathbb{R} \right\}.$$

It can be shown by using standard arguments from spline theory, that the functions in $\mathcal{S}_{K,M}$ are in each component $(M-1)$-times continuously differentiable and that they are equal to a (multivariate) polynomial of degree less than or equal to M (in each component) on each rectangle

$$[u_{k_1}, u_{k_1+1}) \times \cdots \times [u_{k_d}, u_{k_d+1}) \quad (\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d), \tag{7}$$

and that they vanish outside the set

$$\left[-l_n - M \cdot \frac{l_n}{K}, l_n + M \cdot \frac{l_n}{K} \right]^d .$$

Next we define spline approximations using so-called quasi interpolants: For a continuous function $m : \mathbb{R}^d \to \mathbb{R}$ we define an approximating spline by

$$(Qm)(x) = \sum_{\mathbf{k} \in \{-K-M, -K-M+1, \ldots, K-1\}^d} Q_{\mathbf{k}} m \cdot B_{\mathbf{k},M}$$

where

$$Q_{\mathbf{k}}m = \sum_{\mathbf{j} \in \{0,1,\ldots,M\}^d} a_{\mathbf{k},\mathbf{j}} \cdot m(t_{k_1,j_1}, \ldots, t_{k_d,j_d})$$

for some $a_{\mathbf{k},\mathbf{j}} \in \mathbb{R}$ and some suitably chosen points $t_{k,j} \in supp(B_{k,M}) = [k \cdot l_n/K, (k+M+1) \cdot l_n/K]$. It can be shown that if we set

$$t_{k,j} = k \cdot \frac{l_n}{K} + \frac{j}{M} \cdot \frac{l_n}{K} \quad (j \in \{0, \ldots, M\}, k \in \{-K, -K-M+1, \ldots, K-1\})$$

and

$$t_{k,j} = -l_n + \frac{j}{M} \cdot \frac{l_n}{K} \quad (j \in \{0, \ldots, M\}, k \in \{-K-M, -K-M+1, \ldots, -K-1\}),$$

then there exist coefficients $a_{\mathbf{k},\mathbf{j}}$ (which can be computed by solving a linear equation system), such that

$$|Q_{\mathbf{k}}f| \leq c_3 \cdot \|f\|_{\infty,[u_{k_1}, u_{k_1+M+1}] \times \cdots \times [u_{k_d}, u_{k_d+M+1}]} \tag{8}$$

for any $\mathbf{k} \in \mathbb{Z}^d$, any continuous $f : \mathbb{R}^d \to \mathbb{R}$ and some universal constant c_1, and such that Q reproduces polynomials of degree M or less (in each component) on $K_n = [-l_n, l_n]^d$, i.e., for any multivariate polynomial $p : \mathbb{R}^d \to \mathbb{R}$ of degree M or less in each component we have

$$(Qp)(x) = p(x) \quad (x \in K_n) \tag{9}$$

(cf., e.g., Theorem 14.4 and Theorem 15.2 in Györfi et al. (2002)).

Next we define our estimate m_n as a quasi interpolant. We fix the degree $M \in \mathbb{N}$ and set

$$K = \left\lfloor \frac{\lfloor n_2^{1/d} \rfloor - 1}{2M} \right\rfloor,$$

where we assume that $n_2 \geq (2M+1)^d$. Furthermore we choose $x_1, \ldots, x_{n_2}$ such that all of the $(2M \cdot K + 1)^d$ points of the form

$$\left(\frac{j_1}{M \cdot K} \cdot l_n, \ldots, \frac{j_d}{M \cdot K} \cdot l_n \right) \quad (j_1, \ldots, j_d \in \{-M \cdot K, -M \cdot K + 1, \ldots, M \cdot K\})$$

are contained in $\{x_1, \ldots, x_{n_2}\}$, which is possible since $(2M \cdot K + 1)^d \leq n_2$. Then we define

$$m_n(x) = (Qm)(x),$$

where Qm is the above defined quasi interpolant satisfying (8) and (9). The computation of Qm requires only function values of m at the points $x_1, \ldots, x_{n_2}$ and hence m_n is well defined.

It follows from spline theory (cf., e.g., proof of Theorem 1 in Kohler (2013)) that if m is (p, C)-smooth for some $0 < p \leq M + 1$ then the above quasi interpolant m_n satisfies for some constant $c_4 > 0$

$$\|m_n - m\|_{\infty,K_n} := \sup_{x \in K_n} |m_n(x) - m(x)| \leq c_4 \cdot \frac{l_n^p}{n_2^{p/d}}, \tag{10}$$

i.e., (5) is satisfied with $\beta_n = c_4 \cdot l_n^p / n_2^{p/d}$.

3 Main results

First we show the rate of convergence result for the quantile estimate using a general estimate of m.

Theorem 1 *Assume that X is a $\mathbb{R}^d$-valued random variable which has a density with respect to the Lebesgue measure. Let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Assume that $m(X)$ has a density g with respect to the Lebesgue measure. Let $\alpha \in (0,1)$ and let q_α be the α-quantile of $m(X)$. Assume that the density g of $m(X)$ is positive at q_α and continuous on $\mathbb{R}$.*

Let the estimate $\bar{q}_{Z,\bar{\alpha},n}$ of q_α be defined as in Section 2 with $\beta_n = \frac{\log n}{\sqrt{n}}$ and assume that regression estimate m_n satisfies (5). Furthermore assume that

$$\mathbf{P}\{X \notin K_n\} = O\left(\frac{\sqrt{\log(n)}}{\sqrt{n}}\right) \tag{11}$$

Then

$$|\bar{q}_{m(Z),\bar{\alpha},n_3} - q_\alpha| = O_{\mathbf{P}}\left(\frac{\log^{1.5}(n)}{n}\right).$$

When the spline estimate from Section 2 is used to estimate m, then we get the following result.

Corollary 1 *Assume that X is a $\mathbb{R}^d$-valued random variable which has a density with respect to the Lebesgue measure. Let $m : \mathbb{R}^d \to \mathbb{R}$ be a (p,C)-smooth function for some $p > d/2$. Assume that $m(X)$ has a density g with respect to the Lebesgue measure. Let $\alpha \in (0,1)$ and let q_α be the α-quantile of $m(X)$. Assume that the density g of $m(X)$ is positive at q_α and continuous on $\mathbb{R}$.*

Let m_n be the spline estimate from Section 2 with $M \geq p-1$ and define the estimate $\bar{q}_{Z,\bar{\alpha},n}$ of q_α as in Section 2 with $\beta_n = \frac{\log n}{\sqrt{n}}$ and $l_n = \log n$. Furthermore assume that

$$\mathbf{P}\{||X|| \geq \log n\} = O\left(\frac{\sqrt{\log(n)}}{\sqrt{n}}\right). \tag{12}$$

Then

$$|\bar{q}_{m(Z),\bar{\alpha},n_3} - q_\alpha| = O_{\mathbf{P}}\left(\frac{\log^{1.5}(n)}{n}\right).$$

Proof. The assertion follows directly from Theorem 1 and inequality (10) observing that $p > d/2$ implies

$$c_4 \cdot \left(\frac{l_n^p}{n_2^{p/d}}\right) \leq \frac{\log n}{\sqrt{n}}$$

for n sufficiently large. □

Remark 1. It follows from Markov inequality that (12) is satisfied whenever

$$\mathbf{E}\left\{\exp\left(\frac{1}{2} \cdot \|X\|\right)\right\} < \infty.$$

If (12) does not hold it is possible to change the definition of l_n in Corollary 1 to get an (maybe modified) assertion under a weaker tail condition.

Remark 2. It is possible to improve the factor $\log^{1.5}(n)$ in Corollary 1, provided one changes the definition of a_n and b_n. More precisely, let $(\gamma_n)_n$ be a monotonically increasing sequence of positive real values which tends to infinity and assume

$$\mathbf{P}\{||X|| \geq \log n\} = O\left(\frac{\sqrt{\gamma_n}}{\sqrt{n}}\right).$$

Set

$$a_n = \bar{q}_{\alpha,n_1} - \frac{\sqrt{\gamma_n}}{\sqrt{n}} \quad \text{and} \quad b_n = \bar{q}_{\alpha,n_1} + \frac{\sqrt{\gamma_n}}{\sqrt{n}}.$$

By applying (3) in the proof of Theorem 1 it is possible to show that under the assumptions of Corollary 1 the estimate based on the above modified values of a_n and b_n satisfies

$$|\bar{q}_{m(Z),\bar{\alpha},n} - q_\alpha| = O_{\mathbf{P}}\left(\frac{\gamma_n}{n}\right).$$

4 Application to simulated data

In this section we apply the method described above to simulated data and estimate the corresponding 90%-quantile and 95%-quantile. For this purpose the number n of observations will be set to 200, 500, 1000 and 2000, respectively. As suggested in Section 2 we choose $n_1 = n_1(n) = \lfloor n/3 \rfloor = n_2 = n_2(n)$ and $n_3 = n_3(n) = n - n_1 - n_2$. The value of β_n will be set to $\frac{\log(n)}{\sqrt{n}}$, and our estimate of m is the quasi interpolant introduced in Section 2 with $M = 3$, $l_n = \log(n)$ and $K = K(n) = \lfloor (\lfloor n_2^{1/d} \rfloor - 1)/2M \rfloor$. We compare our estimate to the plug-in estimates corresponding to the empirical cdf of the observed data, i.e., to $\bar{q}_{0.9,n}$ and $\bar{q}_{0.95,n}$ (cf., (1) and (2)). In practice it might occur that the value of $\bar{\alpha}$, as defined in Lemma 1, is not in $(0,1)$. This is due to the fact that $\bar{\alpha}$ depends on an estimate of the quantile q_α, based on the first $\lfloor n/3 \rfloor$ samples. Now if the difference between this first estimate and the true quantile is quite large, the true quantile may lay outside of the set the random variable Z, as defined in Section 2, is concentrated on. There are several ways to tackle this problem. In the following we pursue two possible strategies. The first strategy is to alter the value of a_n or b_n, so that the true quantile will lay inside this modified set. For this notice that by definition $\bar{\alpha}$ is negative if γ_1 is larger than α and we have $\bar{\alpha} > 1$ if γ_2 is larger than $1 - \alpha$. Now in order to decrease γ_1 the value of a_n has to be decreased and if we want to decrease γ_2 the value of b_n has to be increased. So the first strategy 1 is to decrease a_n by $\log(n)/\sqrt{(n)}$ if $\bar{\alpha} \leq 0$ and to increase b_n by $\log(n)/\sqrt{(n)}$ if $\bar{\alpha} \geq 1$. This will lead to an altered version of our random variable Z and we will have to recompute $\bar{\alpha}$. We repeat this procedure until $\bar{\alpha} \in (0,1)$.

	90%-quantile $q_{0.9} \approx 3.6022$				95%-quantile $q_{0.95} \approx 5.1803$			
size of n	200	500	1000	2000	200	500	1000	2000
average value with IS start1	3.604	3.604	3.600	3.600	5.170	5.185	5.176	5.180
average value with IS strat 2	3.625	3.614	3.597	3.599	5.145	5.157	5.187	5.163
average sq. error with IS strat 1	0.026	0.005	0.002	0.0004	0.026	0.004	0.001	0.0005
average sq. error with IS strat 2	0.082	0.013	0.002	0.0005	0.221	0.062	0.024	0.014
average sq. error without IS	0.130	0.073	0.037	0.021	0.575	0.240	0.125	0.066

Table 1: Simulation results for $m(x) = \exp(x)$

Since the computation of $\bar{\alpha}$ is the most costly part of our method one might want to avoid the recomputation of $\bar{\alpha}$ as suggested by strategy one. So in the second approach we just generate a new sample of random values drawn from the distribution of Z as defined originally and use a somewhat more loose definition of the *quantile*. More precisely if we have $\bar{\alpha} \leq 0$ we will just take the smallest value of our new sample and if we have $\bar{\alpha} > 1$ we will take the largest one.

In our first example X is standard normally distributed and the function $m : \mathbb{R} \to \mathbb{R}$ is defined by $m(x) = \exp(x)$. In this case $m(X)$ is log-normally distributed. We generate a set of simulated data to which we apply our estimate with a quasi interpolant of degree $M = 3$. In order to compute γ_1 and γ_2 we use the routine *integrate()* from the basic library of the statistics package *R*. This procedure is repeated 100 times for the different values of n. The averages of our 100 estimated values of the quantiles can be found in Table 1. In addition we compute the average squared error of our estimated quantile values and the true quantile. Finally we use the above mentioned plug-in estimate to compute a reference value for our average squared error.

In our second example we set $X = (X_1, X_2)$, where random variables X_1 and X_2 independent standard normally distributed and choose $m(x_1, x_2) = 2 \cdot x_1 + x_2 + 2$. In this case $m(X)$ is normal with expectation 2 and variance $2^2 + 1^2 = 5$. As before we generate a set of simulated data on which we apply our estimate. Unlike in our first example, we now use the procedure *adaptIntegrate()* from the library *cubature* in the statistics package *R* to compute γ_1 and γ_2, since the routine *integrate()* is not applicable to multidimensional domains. As before we repeat this procedure 100 times for the different values of n and compare the results with those of the plug-in estimate described at the beginning of this section. The results can be found in Table 2.

In our third example we set $X = (X_1, X_2)$ for independent standard normally distributed random variables X_1 and X_2 and choose $m(x_1, x_2) = x_1^2 + x_2^2$. Consequently $m(X)$ is chi-square

	90%-quantile $q_{0.9} \approx 4.8656$				95%-quantile $q_{0.95} \approx 5.678$			
size of n	200	500	1000	2000	200	500	1000	2000
average value with IS strat 1	4.691	4.855	4.869	4.865	5.409	5.593	5.635	5.690
average value with IS strat 2	4.691	4.855	4.869	4.865	5.410	5.597	5.634	5.690
average sq. error with IS strat 1	0.070	0.005	0.002	0.0005	0.117	0.023	0.005	0.0008
average sq. error with IS strat 2	0.070	0.005	0.002	0.0005	0.115	0.024	0.006	0.0008
average sq. error without IS	0.093	0.029	0.012	0.005	0.124	0.038	0.021	0.008

Table 2: Simulation results for $m(x, y) = 2x + y + 2$

	90%-quantile $q_{0.9} \approx 4.6052$				95%-quantile $q_{0.95} \approx 5.9915$			
size of n	200	500	1000	2000	200	500	1000	2000
average value with IS strat 1	4.112	4.207	4.601	4.604	5.393	5.432	5.604	5.990
average value with IS strat 2	4.115	4.208	4.600	4.604	5.427	5.459	5.625	5.987
average sq. error with IS strat 1	0.329	0.206	0.001	0.0005	0.549	0.408	0.213	0.0004
average sq. error with IS strat 2	0.329	0.206	0.002	0.0005	0.570	0.437	0.196	0.002
average sq. error without IS	0.195	0.083	0.036	0.017	0.286	0.158	0.079	0.044

Table 3: Simulation results for $m(x, y) = x^2 + y^2$

random variable with two degrees of freedom. The results of our estimate are presented in Table 3.

As one can see, with our proposed procedure all quantiles are well estimated in the average (for n sufficiently large). In addition for large n the average squared error of our estimate is significantly lower compared to the plug-in estimate.

In our last example the function m is motivated by experiments of the Collaborative Research Centre 805 at the Technische Universität Darmstadt, which studies uncertainty in load-bearing systems. A simple example of such load-bearing system is a tripod. Here every leg's end is equipped with sensors to measure the axial force. Since the manufacturing process can not be assumed to be perfect, the holes where the legs are attached to the head in will differ in size. Consequently, a force applied to the tripod will not be partitioned equally between the three legs. In case that one hole is too small, the leg won't fit in and the tripod could not be used. So we'll concentrate on the case of holes with too large diameters. In this case a plugged in leg will be loose and so the center of the hole will differ from the leg's center. This difference is called excentricity, which will be measured in meters. Since the excentricity and the diameter of the hole correlate, we will use the excentricity as indicator. Let now $m : \mathbb{R}_+^3 \to \mathbb{R}$ be a function that returns the resulting load in one fixed leg, depending on the values $(x^{(1)}, x^{(2)}, x^{(3)})$ of the excentricities in the three holes. We simulate the value of $x = (x^{(1)}, x^{(2)}, x^{(3)})$ with independent and uniformly distributed random variables on $(0, 0.1)$ and use our method with $n = 2000$, to estimate the 99%-quantile of the share of the load in one observed leg. As result we get a share of 46.59% of the whole weight in the observed leg, as 99%-quantile. For comparison we also estimate this quantile with order statistics and a sample of size 2,000 and 100,0000, respectively. In the first case the computed value is 47.06% whereas in the latter case we get a value of 46.60% in the observed leg. These results show, that here our estimate performs better than the simple estimate based on order statistics using the same sample size.

5 Proofs

We will use the following lemma in order to prove Lemma 1.

Lemma 2 *Assume that X is a $\mathbb{R}^d$-valued random variable which has a density with respect to the Lebesgue measure. Let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Assume that $m(X)$ has a density g with respect to the Lebesgue measure. Let $\alpha \in (0,1)$ and let q_α be the α-quantile of $m(X)$. Assume that g is bounded away from zero in a neighborhood of q_α.*

Let A and B be subsets of $\mathbb{R}^d$ such that for some $\epsilon > 0$

$$m(x) \leq q_\alpha - \epsilon \text{ for } x \in A \quad \text{and} \quad m(x) > q_\alpha \text{ for } x \in B$$

and

$$\mathbf{P}\{X \notin A \cup B\} > 0.$$

Set

$$h(x) = c_5 \cdot I_{\{x \notin A \cup B\}} \cdot f(x)$$

where

$$c_5^{-1} = \mathbf{P}\{X \notin A \cup B\},$$

and set

$$\bar{\alpha} = \frac{\alpha - \mathbf{P}\{X \in A\}}{\mathbf{P}\{X \notin A \cup B\}}.$$

Let Z be a random variable with density h. Then

$$q_\alpha = q_{m(Z),\bar{\alpha}}.$$

Proof. Since the assumptions of the lemma imply

$$\mathbf{P}\{X \in A\} \le \mathbf{P}\{m(X) \le q_\alpha - \epsilon\} < \alpha \quad \text{and} \quad \mathbf{P}\{X \in B\} \le \mathbf{P}\{m(X) > q_\alpha\} = 1 - \alpha$$

we have

$$\bar{\alpha} = \frac{\alpha - \mathbf{P}\{X \in A\}}{1 - \mathbf{P}\{X \in A\} - \mathbf{P}\{X \in B\}} \in (0, 1].$$

Choose $\epsilon > 0$ such that g is bounded away from zero on $[q_\alpha - \epsilon, q_\alpha]$ and let $q_\alpha - \epsilon < u \le q_\alpha$. By definition of Z we have

$$\begin{aligned}
\mathbf{P}\{m(Z) \le u\} &= \int_{\mathbb{R}} I_{\{m(z) \le u\}} \mathbf{P}_Z(dz) \\
&= \int_{\mathbb{R}} I_{\{m(x) \le u\}} \cdot c_5 \cdot I_{\{x \notin A \cup B\}} \cdot f(x)\, dx.
\end{aligned}$$

The assumptions of the lemma imply that A and B are disjoint and furthermore, because of $q_\alpha - \epsilon < u \le q_\alpha$, they imply

$$I_{\{m(x) \le u\}} \cdot I_{\{x \in A\}} = I_{\{x \in A\}} \quad \text{and} \quad I_{\{m(x) \le u\}} \cdot I_{\{x \in B\}} = 0.$$

¿From this we conclude

$$\begin{aligned}
\mathbf{P}\{m(Z) \le u\} &= \int_{\mathbb{R}} I_{\{m(x) \le u\}} \cdot c_5 \cdot (1 - I_{\{x \in A\}} - I_{\{x \in B\}}) \cdot f(x)\, dx \\
&= c_5 \cdot \left(\int_{\mathbb{R}} I_{\{m(x) \le u\}} \cdot f(x)\, dx - \int_{\mathbb{R}} I_{\{x \in A\}} \cdot f(x)\, dx \right) \\
&= c_5 \cdot \left(\mathbf{P}\{m(X) \le u\} - \mathbf{P}\{X \in A\} \right).
\end{aligned}$$

Using $\mathbf{P}\{m(X) \le u\} < \alpha$ for $u < q_\alpha$, $\mathbf{P}\{m(X) \le q_\alpha\} = \alpha$ and the definition of c_5 we see that we have shown

$$\mathbf{P}\{m(Z) \le u\} < \bar{\alpha} \text{ for } q_\alpha - \epsilon < u < q_\alpha \text{ and } \mathbf{P}\{m(Z) \le q_\alpha\} = \bar{\alpha}.$$

The proof is complete. □

Proof of Lemma 1. In order to apply Lemma 2, at first we define

$$A_n := \{x \in K_n \,:\, m_n(x) < a_n\} = \left\{ x \in K_n \,:\, m_n(x) < \bar{q}_{\alpha,n_1} - 2 \cdot \frac{\log n}{\sqrt{n}} - 2 \cdot \beta_n \right\}$$

and

$$B_n := \{x \in K_n \,:\, m_n(x) > b_n\} = \left\{ x \in K_n \,:\, m_n(x) > \bar{q}_{\alpha,n_1} + 2 \cdot \frac{\log n}{\sqrt{n}} + \beta_n \right\}.$$

Here we observe that using these sets we can characterite the factor c_2 by

$$c_2^{-1} = \mathbf{P}\{X \notin A_n \cup B_n | \mathcal{D}_{n_1}\},$$

where by (6) we have $\mathbf{P}\{X \notin A_n \cup B_n | \mathcal{D}_{n_1}\} > 0$ outside of an event whose probability tends to zero for $n \to \infty$. In addition by rewriting $h(x)$ as

$$h(x) = c_2 \cdot I_{\{x \notin A_n \cup B_n\}} \cdot f(x)$$

and $\bar{\alpha}$ as

$$\bar{\alpha} = \frac{\alpha - \mathbf{P}\{X \in A_n | \mathcal{D}_{n_1}\}}{\mathbf{P}\{X \notin A_n \cup B_n | \mathcal{D}_{n_1}\}}$$

all factors are consistent with Lemma 2. Let now C_n be the event that for all $x \in A_n$ and all $y \in B_n$

$$m(x) \leq q_\alpha - \beta_n \quad \text{and} \quad m(y) > q_\alpha$$

hold. Then by Lemma 2 we get the relation

$$\mathbf{P}\{C_n\} \leq \mathbf{P}\{q_\alpha = q_{m(Z),\bar{\alpha}}\},$$

hence it suffices to show that $\mathbf{P}\{C_n\}$ is tending to one for $n \to \infty$. Therefore we observe that according to (5), for all $x \in A_n$ and all $y \in B_n$ we have

$$m(x) \leq m_n(x) + \beta_n < \bar{q}_{\alpha,n} - 2 \cdot \frac{\log n}{\sqrt{n}} - \beta_n$$

and

$$m(y) \geq m_n(y) - \beta_n > \bar{q}_{\alpha,n} + 2 \cdot \frac{\log n}{\sqrt{n}}.$$

This implies

$$\begin{aligned} \mathbf{P}\{C_n\} \quad &\geq \quad \mathbf{P}\left\{\bar{q}_{\alpha,n_1} - 2 \cdot \frac{\log n}{\sqrt{n}} - \beta_n \leq q_\alpha - \beta_n \quad \text{and} \quad \bar{q}_{\alpha,n} + 2 \cdot \frac{\log n}{\sqrt{n}} \geq q_\alpha\right\} \\ &= \quad \mathbf{P}\left\{\bar{q}_{\alpha,n_1} - 2 \cdot \frac{\log n}{\sqrt{n}} \leq q_\alpha \leq \bar{q}_{\alpha,n} + 2 \cdot \frac{\log n}{\sqrt{n}}\right\} \to 1 \qquad (n \to \infty) \end{aligned}$$

by (3), which completes the proof. □

A crucial step in the proof of Theorem 1 is to show that the inverse of the cdf of $m(Z)$ is locally differentiable at $\bar{\alpha}$ and to determine its derivative. We will do this in the next three lemmas.

Lemma 3 *Let g be the density of $m(X)$ and let A be a measurable subset of $\mathbb{R}$ with the property that for all $x \in K_n$ we have*

$$m(x) \in A \quad \Rightarrow \quad a_n \leq m_n(x) \leq b_n. \tag{13}$$

Then

$$\mathbf{P}\{m(Z) \in A | \mathcal{D}_{n_1}\} = c_2 \cdot \int_A g(y)\, dy.$$

Proof. The definition of Z, (13) and the fact that g is the density of $m(X)$ imply

$$\begin{aligned}
\mathbf{P}\{m(Z) \in A | \mathcal{D}_{n_1}\} &= \int_{\mathbb{R}} I_{\{m(z)\in A\}} \mathbf{P}_Z(dz) \\
&= \int_{\mathbb{R}} I_{\{m(x)\in A\}} \cdot c_2 \cdot \left(I_{\{x\in K_n \,:\, a_n \le m_n(x) \le b_n\}} + I_{\{x \notin K_n\}} \right) \cdot f(x)\, dx \\
&= c_2 \cdot \int_{\mathbb{R}} I_{\{m(x)\in A\}} \cdot \left(I_{\{x\in K_n\}} + I_{\{x \notin K_n\}} \right) \cdot f(x)\, dx \\
&= c_2 \cdot \int_{\mathbb{R}} I_{\{m(x)\in A\}} \cdot f(x)\, dx \\
&= c_2 \cdot \mathbf{P}\{m(X) \in A\} \\
&= c_2 \cdot \int_A g(y)\, dy.
\end{aligned}$$

□

Lemma 4 *Assume that a density g of $m(X)$ exists and let $G_{m(Z)}$ be the cdf of $m(Z)$, i.e.,*

$$G_{m(Z)}(y) = \mathbf{P}\{m(Z) \le y | \mathcal{D}_{n_1}\}.$$

Then $G_{m(Z)}$ is outside of an event, whose probability tends to zero for $n \to \infty$, at Lebegue-almost all points y of the interval

$$I := \left(q_\alpha - \frac{\log n}{\sqrt{n}}, q_\alpha + \frac{\log n}{\sqrt{n}} \right)$$

differentiable with derivative

$$G'_{m(Z)}(y) = c_2 \cdot g(y). \tag{14}$$

In particular, (14) holds for all continuity points $y \in I$ of g.

Proof. Note that the distribution of Z depends on the density h, which depends (via the estimate of the quantile) on $\mathcal{D}_{n_1}$ and hence is random itself. Now let A_n be the event that $|q_\alpha - \bar{q}_{\alpha,n_1}| \le \frac{\log n}{\sqrt{n}}$. Then (3) implies that $\mathbf{P}\{A_n\}$ tends to one for $n \to \infty$. In the following we assume that A_n holds. The next step is to show that Lemma 3 is applicable for every subset A of I when n is large. To this end notice that the inequality

$$m(x) - \beta_n \le m_n(x) \le m(x) + \beta_n$$

holds for every $x \in K_n$, due to (5). So for $x \in K_n$ with $m(x) \in I$ we have since A_n holds

$$\begin{aligned}
a_n &= \bar{q}_{\alpha,n_1} - 2 \cdot \frac{\log n}{\sqrt{n}} - 2 \cdot \beta_n \le q_\alpha - \frac{\log n}{\sqrt{n}} - 2 \cdot \beta_n \le m(x) - 2 \cdot \beta_n \le m_n(x) \\
&\le m(x) + \beta_n \le q_\alpha + \frac{\log n}{\sqrt{n}} + \beta_n \le \bar{q}_{\alpha,n_1} + 2 \cdot \frac{\log n}{\sqrt{n}} + \beta_n = b_n.
\end{aligned}$$

This and Lemma 3 (applied with $A = (\min\{y, y+h\}, \max\{y, y+h\}])$ imply that

$$\begin{aligned}
\frac{G_{m(Z)}(y+h) - G_{m(Z)}(y)}{h} &= sign(h) \cdot \frac{1}{h} \cdot \mathbf{P}\{m(Z) \in (\min\{y, y+h\}, \max\{y, y+h\}]\} \\
&= \frac{1}{h} \cdot \int_y^{y+h} c_2 \cdot g(t)\, dt,
\end{aligned}$$

for every $y \in I$ and all $h \in \mathbb{R}$ small enough to fulfill $y + h \in I$. (Here $sign(h)$ is the sign of h.) Now for h tending to zero, we get by the Lebesgue density theorem

$$G'_{m(Z)}(y) = \lim_{h \to 0} \frac{1}{h} \cdot \int_y^{y+h} c_2 \cdot g(t)\, dt = c_2 \cdot g(y)$$

for Lebesgue-almost all points y of the interval I. Trivially this relation also holds for all continuity points $y \in I$ of g. □

Observe that by definition c_2 is bounded from below by one.

Lemma 5 *Assume that the density g of $m(X)$ exists, that it is continuous on $\mathbb{R}$ and positive at q_α. Then*

$$G^{-1}_{m(Z)}(u) = \inf \left\{ y \in \mathbb{R} \, : \, G_{m(Z)}(y) \geq u \right\}$$

is outside of an event, whose probability tends to zero for $n \to \infty$, differentiable on the interval

$$\left(\bar{\alpha} - c_6 \cdot \frac{\log n}{\sqrt{n}}, \bar{\alpha} + c_6 \cdot \frac{\log n}{\sqrt{n}} \right)$$

with derivative

$$\frac{d}{du} G^{-1}_{m(Z)}(u) = \frac{1}{c_2 \cdot g(G^{-1}_{m(Z)}(u))}.$$

Proof. Observe that the premise of Lemma 4 is fulfilled, so outside of an event, whose probability tends to zero for $n \to \infty$, $G'_{m(Z)}(y) = c_2 \cdot g(y)$ holds for all points $y \in I = \left(q_\alpha - \frac{\log n}{\sqrt{n}}, q_\alpha + \frac{\log n}{\sqrt{n}} \right)$. Since we assume g to be continuous and $G'_{m(Z)}(q_\alpha) = c_2 \cdot g(q_\alpha)$ to be positive, there exists a neighbourhood U of q_α such that $g(u) > \lambda$ holds for all $u \in U$ and some constant $0 < \lambda < g(q_\alpha)$. By this we can apply the inverse function theorem on $U \cap I$. Now for n large enough the interval I will surely be a subset of U which means $U \cap I = I$ in fact. In this case we take a closer look at the range $G_{m(Z)}(I)$. Since $G_{m(Z)}$ is continuous and strictly increasing on I, we have

$$G_{m(Z)}(I) = \left(G_{m(Z)} \left(q_\alpha - \frac{\log n}{\sqrt{n}} \right) , G_{m(Z)} \left(q_\alpha + \frac{\log n}{\sqrt{n}} \right) \right).$$

Now assume $q_\alpha = q_{m(Z),\bar{\alpha}}$ to hold. Then from

$$G_{m(Z)} \left(q_\alpha - \frac{\log n}{\sqrt{n}} \right) = G_{m(Z)}(q_\alpha) - c_2 \cdot \int_{q_\alpha - \frac{\log n}{\sqrt{n}}}^{q_\alpha} g(t) dt \leq \bar{\alpha} - c_2 \cdot \lambda \cdot \frac{\log n}{\sqrt{n}}$$

and

$$G_{m(Z)} \left(q_\alpha + \frac{\log n}{\sqrt{n}} \right) = G_{m(Z)}(q_\alpha) + c_2 \cdot \int_{q_\alpha}^{q_\alpha + \frac{\log n}{\sqrt{n}}} g(t) dt \geq \bar{\alpha} + c_2 \cdot \lambda \cdot \frac{\log n}{\sqrt{n}}$$

we conclude that

$$G_{m(Z)}(I) \supseteq \left(\bar{\alpha} - c_2 \cdot \lambda \cdot \frac{\log n}{\sqrt{n}}, \bar{\alpha} + c_2 \cdot \lambda \cdot \frac{\log n}{\sqrt{n}} \right) =: \tilde{I}.$$

Notice that Lemma 1 implies that $\mathbf{P}\{q_\alpha = q_{m(Z),\bar{\alpha}}\}$ tends to one for $n \to \infty$, so we're outside of an event, whose probability tends to zero for $n \to \infty$. Application of the inverse function theorem implies

$$\frac{d}{du} G^{-1}_{m(Z)}(u) = \frac{1}{c_2 \cdot g(G^{-1}_{m(Z)}(u))}, \tag{15}$$

for all $u \in \tilde{I}$. Notice that since $c_2 \geq 1$, equality (15) holds for all $u \in \left(\bar{\alpha} - \lambda \cdot \frac{\log n}{\sqrt{n}}, \bar{\alpha} + \lambda \cdot \frac{\log n}{\sqrt{n}}\right)$ in particular. □

Proof of Theorem 1. At first notice that $q_{m(Z),\bar{\alpha}}$ implicitly depends on n. Denote by C_n the event that $q_\alpha = q_{m(Z),\bar{\alpha}}$ for $n \in \mathbb{N}$ and notice that for every $s \in \mathbb{R}$ we have

$$\mathbf{P}\left\{|\bar{q}_{m(Z),\bar{\alpha},n_3} - q_\alpha| > s\right\} \quad \leq \quad \mathbf{P}\left\{|\bar{q}_{m(Z),\bar{\alpha},n_3} - q_{m(Z),\bar{\alpha}}| > s\right\} + \mathbf{P}\left\{C_n^c\right\}.$$

Now Lemma 1 implies that $\mathbf{P}\left\{C_n^c\right\}$ is tending to zero for $n \to \infty$ and so

$$\limsup_{n\to\infty} \mathbf{P}\left\{|\bar{q}_{m(Z),\bar{\alpha},n_3} - q_\alpha| > \frac{\log^{1.5}(n)}{n}\right\} \leq \limsup_{n\to\infty} \mathbf{P}\left\{|\bar{q}_{m(Z),\bar{\alpha},n_3} - q_{m(Z),\bar{\alpha}}| > \frac{\log^{1.5}(n)}{n}\right\}. \tag{16}$$

Let $G_{m(Z)}$ be the cdf of $m(Z)$, i.e.,

$$G_{m(Z)}(y) = \mathbf{P}\{m(Z) \leq y | \mathcal{D}_{n_1}\} \quad (y \in \mathbb{R}),$$

and set

$$G^{-1}_{m(Z)}(u) = \inf\left\{y \in \mathbb{R} : G_{m(Z)}(y) \geq u\right\}.$$

Let $U, U_1, U_2, \ldots$ be independent and uniformly on $(0,1)$ distributed random variables and denote the order statistics of $U_1, \ldots, U_{n_3}$ by $U_{1:n_3}, \ldots, U_{n_3:n_3}$.

Since

$$\left(G^{-1}_{m(Z)}(U_1), \ldots, G^{-1}_{m(Z)}(U_{n_3})\right)$$

has the same distribution as

$$(m(Z_1), \ldots, m(Z_{n_3}))$$

and since $G^{-1}_{m(Z)}$ is monotonically increasing on $(0,1)$, due to (16) it suffices to show

$$\left|G^{-1}_{m(Z)}(U_{\lceil\bar{\alpha}\cdot n_3\rceil:n_3}) - G^{-1}_{m(Z)}(\bar{\alpha})\right| = O_{\mathbf{P}}\left(\frac{\log^{1.5}(n)}{n}\right).$$

It follows from Lemma 5 and the mean value theorem that outside of an event, whose probability tends to zero for $n \to \infty$, we have

$$\left|G^{-1}_{m(Z)}(U_{\lceil\bar{\alpha}\cdot n_3\rceil:n_3}) - G^{-1}_{m(Z)}(\bar{\alpha})\right| \quad = \quad \left|U_{\lceil\bar{\alpha}\cdot n_3\rceil:n_3} - \bar{\alpha}\right| \cdot \frac{1}{c_2 \cdot g(G^{-1}_{m(Z)}(D_{n_3}))}$$

where D_{n_3} is some random point between $U_{\lceil\bar{\alpha}\cdot n_3\rceil:n_3}$ and $\bar{\alpha}$, provided the distance between $U_{\lceil\bar{\alpha}\cdot n_3\rceil:n_3}$ and $\bar{\alpha}$ is less than $c_6 \cdot \log(n)/\sqrt{n}$.

Let F_U be the cdf of U and let F_{U,n_3} be the empirical cdf corresponding to $U_1, \ldots, U_{n_3}$. Then we have with probability one that

$$\begin{aligned} \left|U_{\lceil \bar{\alpha} \cdot n_3 \rceil : n_3} - \bar{\alpha}\right| &= \left|U_{\lceil \bar{\alpha} \cdot n_3 \rceil : n_3} - F_{U,n_3}(U_{\lceil \bar{\alpha} \cdot n_3 \rceil : n_3}) + \frac{\lceil \bar{\alpha} \cdot n_3 \rceil}{n_3} - \bar{\alpha}\right| \\ &\leq \sup_{t \in \mathbb{R}} |F_{U,n_3}(t) - F_U(t)| + \frac{1}{n_3}. \end{aligned}$$

This implies

$$\left|U_{\lceil \bar{\alpha} \cdot n_3 \rceil : n_3} - \bar{\alpha}\right| = O_{\mathbf{P}}\left(\frac{\sqrt{\log(n_3)}}{\sqrt{n_3}}\right)$$

(cf., e.g., Theorem 12.4 in Devroye, Györfi and Lugosi (1996)). Furthermore, by Lemma 1 we can assume that $G^{-1}_{m(Z)}(\bar{\alpha}) = q_\alpha$ holds, and we have that $G^{-1}_{m(Z)}$ is continuous at $\bar{\alpha}$ and that g is positive and continuous at q_α. Hence it suffices to show

$$\frac{1}{c_2} = \int_{\mathbb{R}^d} \left(I_{\{x \in K_n \,:\, a_n \leq m_n(x) \leq b_n\}} + I_{\{x \notin K_n\}}\right) f(x)dx = O_{\mathbf{P}}\left(\frac{\log(n)}{\sqrt{n}}\right).$$

This in turn follows from

$$\mathbf{P}\left\{X \in K_n \,:\, a_n \leq m_n(X) \leq b_n \middle| \mathcal{D}_{n_1}\right\} = O_{\mathbf{P}}\left(\frac{\log(n)}{\sqrt{n}}\right) \tag{17}$$

and

$$\mathbf{P}\left\{X \notin K_n\right\} = O_{\mathbf{P}}\left(\frac{\sqrt{\log(n)}}{\sqrt{n}}\right). \tag{18}$$

Note that (18) holds by assumption (11). In order to show (17) we assume that $|q_\alpha - \bar{q}_{\alpha,n_1}| \leq \frac{\log n}{\sqrt{n}}$. Then the definitions of a_n, b_n and β_n imply

$$\begin{aligned} &\mathbf{P}\left\{X \in K_n \,:\, a_n \leq m_n(X) \leq b_n \middle| \mathcal{D}_{n_1}\right\} \\ &\leq \mathbf{P}\left\{X \in K_n \,:\, a_n - \beta_n \leq m(X) \leq b_n + \beta_n \middle| \mathcal{D}_{n_1}\right\} \\ &= \mathbf{P}\left\{\bar{q}_{\alpha,n_1} - 2 \cdot \frac{\log n}{\sqrt{n}} - 3 \cdot \beta_n \leq m(X) \leq \bar{q}_{\alpha,n_1} + 2 \cdot \frac{\log n}{\sqrt{n}} + 2 \cdot \beta_n \middle| \mathcal{D}_{n_1}\right\} \\ &\leq \mathbf{P}\left\{q_\alpha - 3 \cdot \frac{\log n}{\sqrt{n}} - 3 \cdot \beta_n \leq m(X) \leq q_\alpha + 3 \cdot \frac{\log n}{\sqrt{n}} + 2 \cdot \beta_n \middle| \mathcal{D}_{n_1}\right\} \\ &\leq \mathbf{P}\left\{q_\alpha - 6 \cdot \frac{\log(n)}{\sqrt{n}} \leq m(X) \leq q_\alpha + 5 \cdot \frac{\log(n)}{\sqrt{n}}\right\} \\ &\leq \sup_{x \in [q_\alpha - 6\,,\, q_\alpha + 5]} g(x) \cdot 11 \cdot \frac{\log(n)}{\sqrt{n}}. \end{aligned}$$

Here we have used the fact that the continuous function g is bounded on any finite interval around q_α and that $\frac{\log(n)}{\sqrt{n}}$ is bounded from above by one. Finally (3) implies the assertion. □

6 Acknowledgment

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

References

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

[2] de Boor, C. (1978). *A Practical Guide to Splines.* Springer.

[3] Cannamela, C, Garnier, J. and Iooss, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics*, vol. 2, no. 4, pp. 1554-1580.

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

[5] Egloff, D. and Leippold, M. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics*, vol. 38, no. 2, pp. 1244-1278.

[6] Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York.

[7] 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.

[8] Kohler, M. (2013). Optimal global rates of convergence for noiseless regression estimation problems with adaptively chosen design. Submitted for publication.

[9] Morio, J. (2012). Extreme quantile estimation with nonparametric adaptive importance sampling *Simulation Modelling Practice and Theory*, vol. 27, pp. 76-89.

[10] Neddermeyer, J. C. (2009). Computationally efficient nonparametric importance sampling. *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 788-802.