Estimation of extreme quantiles in a simulation model *

Michael Kohler[1] and Adam Krzyżak[2,†]

[1] *Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@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*

January 4, 2016

Abstract
A simulation model with an outcome $Y = m(X)$ is considered, where X is an $\mathbb{R}^d$-valued random variable and $m : \mathbb{R}^d \to \mathbb{R}$ is a smooth function. Estimates of the α_n–quantile $q_{m(X),\alpha_n}$ of $m(X)$ based on surrogate model of m and on importance sampling are constructed which use at most n evaluations of the function m. Results concerning the rate of convergence of the estimates are derived in case that $\alpha_n \to 1$ $(n \to \infty)$ and $n \cdot (1 - \alpha_n) \to 0$ $(n \to \infty)$. Finite sample behavior of the estimate is illustrated by simulations.

AMS classification: Primary 62G05; secondary 62G30.

Key words and phrases: Extreme quantiles, nonparametric quantile estimation, importance sampling, rates of convergence.

1 Introduction

Let X, X_1, X_2, ... be independent and identically distributed $\mathbb{R}^d$–valued random variables, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and set $Y = m(X)$, $Y_1 = m(X_1)$, ... Let $G_{m(X)}$ be the cumulative distribution function (cdf) of $m(X)$, i.e.,

$$G_{m(X)}(y) = \mathbf{P}\{m(X) \le y\} \quad (y \in \mathbb{R}).$$

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

$$\alpha_n \to 1 \quad (n \to \infty) \quad \text{and} \quad n \cdot (1 - \alpha_n) \to 0 \quad (n \to \infty), \tag{1}$$

and let

$$q_{m(X),\alpha_n} = \inf\{y \in \mathbb{R} : G_{m(X)}(y) \ge \alpha_n\}$$

*Running title: *Estimation of quantiles*

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

be the quantile of $m(X)$ of level α_n. In this paper we consider the problem of estimating $q_{m(X),\alpha_n}$ using at most n evaluations of the function m at arbitrarily chosen points.

A simple idea to estimate $q_{m(X),\alpha_n}$ is to use $m(X_1), \ldots, m(X_n)$ to compute the empirical cdf

$$\hat{G}_{m(X),n}(y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{m(X_i) \leq y\}} \tag{2}$$

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

$$\hat{q}_{m(X),n,\alpha_n} = \min\{z \in \mathbb{R} : \hat{G}_{m(X),n}(z) \geq \alpha_n\}. \tag{3}$$

Since $\hat{q}_{m(X),n,\alpha_n}$ is in fact an order statistic, results from order statistics can be used to analyze its rate of convergence. More precisely, let $Y_{1:n}, \ldots, Y_{n:n}$ be the order statistics corresponding to $Y_1 = m(X_1), \ldots, Y_n = m(X_n)$, i.e., $Y_{1:n}, \ldots, Y_{n:n}$ is a permutation of $Y_1, \ldots, Y_n$ satisfying

$$Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}.$$

Then $\hat{q}_{m(X),n,\alpha} = Y_{\lceil n \cdot \alpha_n \rceil : n}$ is the $\lceil n \cdot \alpha_n \rceil = (n - (n - \lceil n \cdot \alpha_n \rceil + 1) + 1)$–th order statistics, where $\lceil z \rceil$ is the smallest integer greater than or equal to $z \in \mathbb{R}$. In case that $\alpha_n = \alpha \in (0,1)$ for $n \in \mathbb{N}$ it is a so–called central order statistics, and, e.g., Theorem 8.5.1 in Arnold, Balakrishnan and Nagaraja (1992), implies that we have

$$\sqrt{n} \cdot g(q_{m(X),\alpha}) \cdot \frac{\hat{q}_{m(X),n,\alpha} - q_{m(X),\alpha}}{\sqrt{\alpha \cdot (1-\alpha)}} \to N(0,1) \quad \text{in distribution,}$$

whenever $m(X)$ has a density g which is continuous and positive at $q_{m(X),\alpha}$. Thus

$$|\hat{q}_{m(X),n,\alpha} - q_{m(X),\alpha}| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{4}$$

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

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

Here the rate of convergence does not depend on the distribution of $m(X)$.

In case that $\alpha_n \to 1$ $(n \to \infty)$ and $n \cdot (1 - \alpha_n) \to \infty$ $(n \to \infty)$ we have $\lceil n \cdot \alpha_n \rceil \to \infty$ $(n \to \infty)$, $n - \lceil n \cdot \alpha_n \rceil + 1 \to \infty$ $(n \to \infty)$ and $(n - \lceil n \cdot \alpha_n \rceil + 1)/n \to 0$ $(n \to \infty)$ and $\hat{q}_{m(X),n,\alpha_n}$ is so–called an intermediate order statistics. Assume that in this case $m(X)$ has a density $g : \mathbb{R} \to \mathbb{R}$ which is positive on $\mathbb{R}_+$ and which satisfies one of the so–called von Mises-type conditions (cf., e.g., Theorem 1 in Sweeting (1985)), i.e., for which (in this case) one of the following two conditions hold:

$$\lim_{y \to \infty} \frac{g(y) \cdot \int_y^\infty (1 - G_{m(X)}(t))\, dt}{(1 - G_{m(X)}(y))^2} = 1 \tag{5}$$

or

$$\lim_{y \to \infty} \frac{y \cdot g(y)}{1 - G_{m(X)}(y)} \in (0, \infty). \tag{6}$$

Then, e.g., Theorem 2.1 in Falk (1989) implies that

$$\frac{\sqrt{n} \cdot g(q_{m(X),\alpha_n})}{\sqrt{1-\alpha_n}} \cdot (\hat{q}_{m(X),n,\alpha_n} - q_{m(X),\alpha_n}) \to N(0,1) \quad \text{in distribution,}$$

from which we can conclude that

$$|\hat{q}_{m(X),n,\alpha_n} - q_{m(X),\alpha_n}| = O_{\mathbf{P}} \left(\frac{\sqrt{1-\alpha_n}}{\sqrt{n} \cdot g(q_{m(X),\alpha_n})} \right). \tag{7}$$

Here the rate of convergence depends on α_n and the rate of decay of the density g. E.g., if $m(X)$ is exponentially distributed with expectation $1/\lambda$ (which implies $q_{m(X),\alpha_n} = -\log(1-\alpha_n)/\lambda$ and $g(q_{m(X),\alpha_n}) = \lambda \cdot (1-\alpha_n)$), then the rate of convergence is $1/(\sqrt{n} \cdot \sqrt{1-\alpha_n})$.

In case that $\alpha_n \to 1$ $(n \to \infty)$ and $n \cdot (1-\alpha_n) \to 0$ $(n \to \infty)$ we have $\lceil n \cdot \alpha_n \rceil \to \infty$ $(n \to \infty)$ and $n - \lceil n \cdot \alpha_n \rceil + 1 \to 1$ $(n \to \infty)$. Consequently, $\hat{q}_{m(X),n,\alpha_n}$ is an example of an extreme order statistic, and for large n it is given by

$$\hat{q}_{m(X),n,\alpha_n} = \max\left\{m(X_1), \ldots, m(X_n)\right\}.$$

In this case the left-hand side of (7) does not in general converge to zero (since $\hat{q}_{m(X),n,\alpha_n}$ does not adapt for large n to the specific form of α_n).

In this article we study the rates of convergence of estimates of $q_{m(X),\alpha_n}$ based on an initial estimate (surrogate)

$$m_n(\cdot) = m_n(\cdot, (x_1, m(x_1)), \ldots, (x_n, m(x_n))) : \mathbb{R}^d \to \mathbb{R} \tag{8}$$

of m. Here the estimate m_n uses n evaluations of m at suitably chosen points $x_1, \ldots, x_n \in \mathbb{R}^d$. In Section 2 we will use polynomial splines to construct such an estimate in case of a smooth function $m : \mathbb{R}^d \to \mathbb{R}$ (e.g., in case when m is k–times continuously differentiable). Our estimates use this estimate of m together with either a sample $X_1, \ldots, X_{N_n}$ of X of size $N_n > n$ or a given density $f : \mathbb{R}^d \to \mathbb{R}$ of X to construct estimates $\hat{G}_{N_n}$ of the cumulative distribution function $G_{m(X)}$ of $m(X)$ and estimate the quantile by the corresponding plug-in quantile estimate

$$\hat{q}_{n,\alpha_n} = \min\{z \in \mathbb{R} : \hat{G}_{N_n}(z) \geq \alpha_n\}.$$

The estimate $\hat{G}_{N_n}$ is constructed in two different ways. For our first estimate we use the empirical cdf of the data

$$m_n(X_1), \ldots, m_n(X_{N_n}), \tag{9}$$

while for the second estimate we use importance sampling to construct from the given density f of X a new sample and define $\hat{G}_{N_n}$ as a properly renormalized empirical distribution function corresponding to the values of m_n evaluated at this sample.

In Theorems 1 and 2 below we analyze the rates of convergence of our newly proposed estimates. Of course the rates of convergence depend on the quality of the surrogate of m. If its error is small then the first estimate achieves the same rate of convergence as the order statistics but with the sample size n replaced by the size N_n of the additional sample of X. In any application there is an upper bound on the sample size N_n which can be used in computing the estimate. Consequently, it is important how the rate of convergence of the estimate depends on the number N_n of additional samples of X. Here our result on the second estimate is useful: Due to the use of importance sampling its rate of convergence decreases in N_n like $1/N_n$ in comparison to $1/\sqrt{N_n}$ for the first estimate. Furthermore we illustrate the finite sample size behaviour of our estimates by applying them to simulated data.

Extreme quantiles can be estimated by using methods from extreme value theory. The main idea there is to characterize the tail behaviour of the cdf using a so-called extreme value index, to estimate this extreme value index from data and to estimate extreme quantiles by extrapolating moderate quantiles on the basis of the estimated extreme value index (cf., e.g., Beirlant et al. (2004) and the literature cited therein). How this method can be applied to the simulation model studied in this paper, i.e., how the surrogate model can be used to estimate the extreme value index, has not been studied in the literature and is not investigated in this paper. Instead we use importance sampling (IS) combined with methods from curve estimation in our simulation model to construct completely nonparametric estimates of extreme quantiles. There is a large body of literature on importance sampling for quantile estimation. Early papers on applications of IS to Monte Carlo simulation include Glynn (1996). Efficient importance sampling for ruin problems has been investigated by Blanchet and Liu (2010) and Glasserman et al. (2002). Hong (2014) applied IS to computing value at risk using Monte Carlo techniques. Similar results for heavy tailed distributions were obtained by Hult and Svensson (2009). Chu and Nakayama (2012) and Nakayama (2014) studied confidence intervals for quantile estimates using IS and Liu and Yang (2012) provide an analysis of the bootstrap quantile variance estimator. Importance sampling has been applied to adaptive quantile estimation by Egloff and Leippold (2010). Morio (2012) applied nonparametric adaptive IS to extreme quantile estimation. Both papers do not contain any results concerning the rate of convergence of the quantile estimates. As in the current paper importance sampling has been applied in a simulation model in Kohler et al. (2014) and in Kohler, Krzyżak and Walk (2014), whereas in each paper the rates of convergence of the estimates have also been studied. The basic new idea in the current paper is that we evaluate the surrogate (and not the function of the simulation model) on the importance sampling sample, which enables us to use for this sample a much higher sample size. This in turn enables us to study the estimation of quantiles of level α_n where $n \cdot (1 - \alpha_n) \to 0$ $(n \to 0)$.

Importance sampling has also been used as standard variance reduction technique in rare event simulation, see, e.g., Siegmund (1976), Dupuis and Wang (2005) and Asmussen and Kroese (2006).

In order to construct the surrogate m_n any kind of nonparametric regression estimate suffices. For instance we can use kernel regression estimate (cf., e.g., Nadaraya (1964, 1970), Watson (1964), Devroye and Wagner (1980), Stone (1977, 1982) or Devroye and Krzyżak (1989)), partitioning regression estimate (cf., e.g., Györfi (1981) or Beirlant and Györfi (1998)), nearest neighbor regression estimate (cf., e.g., Devroye (1982) or Devroye, Györfi, Krzyżak and Lugosi (1994)), orthogonal series regression estimate (cf., e.g., Rafajłowicz (1987) or Greblicki and Pawlak (1985)), least squares estimates (cf., e.g., Lugosi and Zeger (1995) or Kohler (2000)) or smoothing spline estimates (cf., e.g., Wahba (1990) or Kohler and Krzyżak (2001)).

In the proofs of our results we apply a general result on the rate of convergence of surrogate quantile estimates derived in Enss et al. (2014).

The definitions of the estimates are provided in Section 2, the main results are presented in Section 3 and proven in Section 5. In Section 4 we illustrate the finite sample size performance of our estimates by applying them to simulated data.

2 Definition of the estimates

2.1 Definition of the spline estimate of m

In order to define spline estimates of m, 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 $l_n = (\log n)^\gamma$ for some $\gamma > 0$ and $u_k = k \cdot l_n/K$ ($k \in \mathbb{Z}$). 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

$$[u_{k_1}, u_{k_1+1}) \times \cdots \times [u_{k_d}, u_{k_d+1}) \quad (\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d), \tag{10}$$

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

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

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

$$|Q_{\mathbf{k}} f| \leq c_1 \cdot \|f\|_{\infty, [u_{k_1}, u_{k_1+M+1}] \times \cdots \times [u_{k_d}, u_{k_d+M+1}]} \tag{11}$$

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

$$(Qp)(x) = p(x) \quad (x \in K_n) \tag{12}$$

(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

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

where we assume that $n \geq (2M+1)^d$. Furthermore we choose $x_1, \dots, x_n$ such that all of the $(2M \cdot K + 1)^d$ points of the form

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

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

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

where Qm is the above defined quasi interpolant satisfying (11) and (12). The computation of Qm requires only function values of m at the points $x_1, \dots, x_n$ and hence m_n is well defined.

Let $p = k + s$ for some $k \in \mathbb{N}_0$ and some $s \in (0, 1]$. 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 $x, z \in \mathbb{R}^d$. It follows from spline theory (cf., e.g., proof of Theorem 1 in Kohler (2014)) 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 $c_2 > 0$

$$\|m_n - m\|_{\infty, K_n} := \sup_{x \in K_n} |m_n(x) - m(x)| \leq c_2 \cdot \frac{l_n^p}{n^{p/d}}, \tag{13}$$

where

$$K_n = [-l_n, l_n]^d = [-(\log n)^\gamma, (\log n)^\gamma]^d. \tag{14}$$

2.2 The first quantile estimate

For our first estimate we start by estimating the cdf $G_{m(X)}$ of $m(X)$ by the empirical cdf corresponding to the data (9), i.e., by

$$\hat{G}_{m_n(X), N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m_n(X_i) \leq y\}} \quad (y \in \mathbb{R}). \tag{15}$$

Then we construct the plug-in estimate of our quantile $q_{m(X), \alpha_n}$, i.e., we define

$$\hat{q}_{m_n(X), N_n, \alpha_n} = \inf \left\{ y \in \mathbb{R} : \hat{G}_{m_n(X), N_n}(y) \geq \alpha_n \right\}. \tag{16}$$

In this case our estimate is the $\lceil N_n \cdot \alpha_n \rceil$–th order statistic of the data (9).

2.3 The second quantile estimate

For our second estimate we assume that besides our surrogate m_n and the set K_n (cf., (14)) where it approximates m_n with supremum norm error $\delta_n > 0$ we are also given a (possible random) sequence $\eta_n \geq 3 \cdot \delta_n$ such that

$$\mathbf{P}\left\{\left|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}\right| > \eta_n/3\right\} \to 0 \quad (n \to \infty) \tag{17}$$

and the density $f : \mathbb{R}^d \to \mathbb{R}$ of X (which we assume to exist). We define a new density h_n by

$$h_n(z) = \frac{1}{c_n} \cdot \left(1_{\{z \in K_n : |m_n(z) - \hat{q}_{m_n(X),N_n,\alpha_n}| \leq \eta_n\}} + 1_{\{z \notin K_n\}}\right) \cdot f(z), \tag{18}$$

where

$$c_n = \int_{\mathbb{R}^d} \left(1_{\{z \in K_n : |m_n(z) - \hat{q}_{m_n(X),N_n,\alpha_n}| \leq \eta_n\}} + 1_{\{z \notin K_n\}}\right) \cdot f(z)\, dz. \tag{19}$$

Let $Z, Z_1, Z_2, \ldots$ be independent and identically distributed $\mathbb{R}^d$–valued random variables with density h_n, and set

$$\bar{\alpha}_n = \frac{\alpha_n - b_n}{c_n}, \tag{20}$$

where

$$b_n = \int_{\mathbb{R}^d} 1_{\{z \in K_n : m_n(z) < \hat{q}_{m_n(X),N_n,\alpha_n} - \eta_n\}} \cdot f(z)\, dz.$$

We estimate the cdf of $m(X)$ using $m_n(Z_1), \ldots, m_n(Z_{N_n})$ by a properly renormalized version of the corresponding empirical cdf, i.e., by

$$\hat{G}(y) = c_n \cdot \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m_n(Z_i) \leq y\}} + b_n \quad (y \in \mathbb{R})$$

and define the quantile estimate $\hat{q}^{(IS)}_{m_n(Z),N_n,\bar{\alpha}_n}$ by the corresponding plug-in quantile estimate, i.e., by

$$\begin{aligned} \hat{q}^{(IS)}_{m_n(Z),N_n,\bar{\alpha}_n} &= \inf\left\{y \in \mathbb{R} : \hat{G}(y) \geq \alpha_n\right\} \\ &= \inf\left\{y \in \mathbb{R} : \hat{G}_{m_n(Z),N_n}(y) \geq \bar{\alpha}_n\right\}, \end{aligned} \tag{21}$$

where

$$\hat{G}_{m_n(Z),N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m_n(Z_i) \leq y\}} \quad (y \in \mathbb{R}). \tag{22}$$

3 Main results

Our first result yields an upper bound on the rate of convergence of the surrogate quantile estimate $\hat{q}_{m_n(X),N_n,\alpha_n}$.

Theorem 1. *Let X be an $\mathbb{R}^d$–valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be measurable and assume that $m(X)$ has a density $g : \mathbb{R} \to \mathbb{R}$ which is positive on $\mathbb{R}_+$ and satisfies (5) or (6). Let $\alpha_n \in (0,1)$, let $N_n \in \mathbb{N}$ be such that*

$$\alpha_n \to 1 \quad (n \to \infty) \quad \textit{and} \quad N_n \cdot (1-\alpha_n) \to \infty \quad (n \to \infty). \tag{23}$$

Let m_n and $\hat{q}_{m_n(X),N_n,\alpha_n}$ be defined as in Section 2 and assume that

$$\sup_{x \in K_n} |m_n(x) - m(x)| \leq \delta_n \tag{24}$$

and

$$N_n \cdot \mathbf{P}\{X \notin K_n\} \to 0 \quad (n \to \infty). \tag{25}$$

Then

$$\left|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}\right| = O_{\mathbf{P}}\left(\delta_n + \frac{\sqrt{1-\alpha_n}}{\sqrt{N_n} \cdot g(q_{m(X),\alpha_n})}\right). \tag{26}$$

In particular in case that we have for some $\epsilon \in (0,1)$ and $c_3 > 0$

$$g(q_{m(X),\alpha_n}) \geq c_3 \cdot (1-\alpha_n)^{1+\epsilon} \tag{27}$$

for sufficiently large n it follows

$$\left|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}\right| = O_{\mathbf{P}}\left(\delta_n + \frac{1}{\sqrt{N_n} \cdot (1-\alpha_n)^{1/2+\epsilon}}\right). \tag{28}$$

Remark 1. In case that $m(X)$ has a Gamma distribution, condition (27) holds for any $\epsilon > 0$. Because in this case we have for any $\delta > 0$ and suitable constants $\beta, c_4, c_5 > 0$

$$c_4 \cdot e^{-(\beta+\delta)\cdot y} \leq g(y) \leq c_5 \cdot e^{-(\beta-\delta)\cdot y},$$

which implies

$$1 - \alpha_n = \int_{q_{m(X),\alpha_n}}^{\infty} g(y)\, dy \leq \frac{c_5}{\beta - \delta} \cdot e^{-(\beta-\delta)\cdot q_{m(X),\alpha_n}}.$$

From this we conclude

$$\begin{aligned} g(q_{m(X),\alpha_n}) &\geq c_4 \cdot \exp\left(-(\beta+\delta) \cdot \frac{1}{\beta-\delta} \cdot \left(\log \frac{c_5}{\beta-\delta} - \log(1-\alpha_n)\right)\right) \\ &\geq c_6 (1-\alpha_n)^{1+\epsilon}, \end{aligned}$$

where

$$\epsilon = \frac{2}{\beta-\delta} \cdot \delta.$$

Corollary 1. *Let X be an $\mathbb{R}^d$-valued random variable, $p, C > 0$ and let $m : \mathbb{R}^d \to \mathbb{R}$ be a (p,C)–smooth function. Let m_n be the spline interpolant defined in Subsection 2.1 with $M+1 \geq p$ and $l_n = (\log n)^\gamma$. Assume that $m(X)$ has a density $g : \mathbb{R} \to \mathbb{R}$ which is positive on $\mathbb{R}_+$ and satisfies (5) or (6). Let $\alpha_n \in (0,1)$ be such that*

$$\alpha_n \to 1 \quad (n \to \infty) \quad \text{and} \quad n^2 \cdot (1-\alpha_n) \to \infty \quad (n \to \infty).$$

Set $N_n = n^2$ and let $\hat{q}_{m_n(X),N_n,\alpha_n}$ be defined as in Subsection 2.2. Assume

$$n^2 \cdot \mathbf{P}\{X \notin [-(\log n)^\gamma, (\log n)^\gamma]^d\} \to 0 \quad (n \to \infty).$$

Then

$$\left|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}\right| = O_{\mathbf{P}}\left(\frac{(\log n)^{\gamma \cdot p}}{n^{p/d}} + \frac{\sqrt{1-\alpha_n}}{n \cdot g(q_{m(X),\alpha_n})}\right).$$

In particular it follows that

$$\left|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}\right| = O_{\mathbf{P}}\left(\frac{(\log n)^{\gamma \cdot p}}{n^{p/d}} + \frac{1}{n \cdot (1-\alpha_n)^{1/2+\epsilon}}\right)$$

whenever (27) holds for some $\epsilon \in (0,1)$ and $c_3 > 0$ for sufficiently large n.

Proof. Follows directly from Theorem 1 and (13). □

Remark 2. Under the assumptions of Corollary 1 it is possible to estimate quantiles of level

$$\alpha_n = 1 - n^{-r}$$

consistently for any $0 < r < 2/(1+2\cdot\epsilon)$.

Theorem 2. *Let X be an $\mathbb{R}^d$–valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be measurable and assume that $m(X)$ has a density $g : \mathbb{R}^d \to \mathbb{R}$ which is positive on $\mathbb{R}_+$ and satisfies (5) or (6). Let $\alpha_n \in (0,1)$ and $N_n \in \mathbb{N}$ be such that (23) holds. Let the importance sampling surrogate quantile estimate $\hat{q}^{(IS)}_{m_n(X),N_n,\bar{\alpha}_n}$ be defined as in Section 2 using some surrogate m_n and some (possibly random) $\eta_n \geq 3 \cdot \delta_n > 0$ satisfying (24) and the following three conditions:*

$$\mathbf{P}\{|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}| > \eta_n/3\} \to 0 \quad (n \to \infty), \tag{29}$$

$$\eta_n \geq \frac{\mathbf{P}\{X \notin K_n\}}{\sup\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq 3 \cdot \eta_n\}} \tag{30}$$

and

$$\frac{\log n}{\sqrt{N_n}} \cdot \frac{\sup\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq 3 \cdot \eta_n\}}{\inf\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq \frac{\eta_n}{3}\}} \to 0 \quad \text{in probability.} \tag{31}$$

Assume that outside of an event, whose probability tends to zero for $n \to \infty$,

$$\frac{N_n \cdot \mathbf{P}\{X \notin K_n\}}{c_n} \to 0 \quad (n \to \infty) \tag{32}$$

holds. Then

$$\left|\hat{q}^{(IS)}_{m_n(Z),N_n,\bar{\alpha}_n} - q_{m(X),\alpha_n}\right|$$
$$= O_{\mathbf{P}}\left(\log(n)\cdot\delta_n + \frac{\log(n)\cdot\eta_n}{\sqrt{N_n}}\cdot\frac{\sup\{g(y): y\in[q_{m(X),\alpha_n}-3\eta_n, q_{m(X),\alpha_n}+3\eta_n]\}}{\inf\{g(y): y\in[q_{m(X),\alpha_n}-3\eta_n, q_{m(X),\alpha_n}+3\eta_n]\}}\right).$$

Remark 3. The techniques used in the proof of Theorem 2 in order to bound c_n imply that (32) is in particular satisfied, if

$$\frac{N_n\cdot\mathbf{P}\{X\notin K_n\}}{\eta_n\cdot\inf\{g(y): y\in[q_{m(X),\alpha_n}-\eta_n/3, q_{m(X),\alpha_n}+\eta_n/3]\}}\to 0 \quad \text{in probability.}$$

Remark 4. Values of Z_1, Z_2, ... can be constructed using a rejection method. To do this one selects from values of X_1, X_2, ... successively all those values $x = X_i$ where either $x\in K_n$ and $|m_n(x)-\hat{q}_{m_n(X),N_n,\alpha_n}|\leq\eta_n$ hold or where $x\notin K_n$ holds. In this case Monte Carlo estimates of b_n and c_n can be used in order to compute $\bar{\alpha}_n$ approximately.

Remark 5. In the proof of Theorem 2 we show that the assumptions of Theorem 2 imply that outside of an event, whose probability tends to zero for $n\to\infty$, all $Z_1, \ldots, Z_{N_n}$ are contained in K_n. Hence if we construct the values of $Z_1, \ldots, Z_{N_n}$ by the rejection method described in Remark 4, then outside of an event, whose probability tends to zero for $n\to\infty$, all considered values of $x = X_i$ satisfy $x\in K_n$. This implies that outside of an event, whose probability tends to zero for $n\to\infty$, the sample of size N_n of the density h_n is identical to a sample of size N_n of the density

$$\bar{h}_n(z) = \frac{1}{\bar{c}_n}\cdot 1_{\{z\in\mathbb{R}^d:|m_n(z)-\hat{q}_{m_n(X),N_n,\alpha_n}|\leq\eta_n\}}\cdot f(z),$$

where

$$\bar{c}_n = \int_{\mathbb{R}^d} 1_{\{z\in\mathbb{R}^d:|m_n(z)-\hat{q}_{m_n(X),N_n,\alpha_n}|\leq\eta_n\}}\cdot f(z)\,dz.$$

But if the samples from two different distributions are equal, then also the importance sampling quantile estimates based on the two different densities are equal. Consequently, Theorem 2 also holds if we replace the density h_n by $\bar{h}_n$, which has the advantage that construction of the estimate does not require knowledge of the set K_n. We implement the estimate in this way in the next section.

Corollary 2. *Let the importance sampling surrogate quantile estimate* $\hat{q}^{(IS)}_{m_n(X),N_n,\bar{\alpha}_n}$ *be defined as in Section 2 using*

$$\eta_n = \log(n)\cdot\delta_n + \frac{\log n}{\sqrt{N_n}\cdot(1-\alpha_n)^{1/2+\epsilon}} \tag{33}$$

and $N_n\geq n$. *Assume that (23), (24) and*

$$\frac{N_n^{3/2}}{(\log n)\cdot(1-\alpha_n)^{1/2}}\cdot\mathbf{P}\{X\notin K_n\}\to 0 \quad (n\to\infty) \tag{34}$$

hold. Assume furthermore that the density g of $m(X)$ satisfies

$$\limsup_{n\to\infty} \frac{\sup\{g(y) : y \in [q_{m(X),\alpha_n} - 3\cdot\eta_n, q_{m(X),\alpha_n} + 3\cdot\eta_n]\}}{\inf\{g(y) : y \in [q_{m(X),\alpha_n} - 3\cdot\eta_n, q_{m(X),\alpha_n} + 3\cdot\eta_n]\}} < \infty \tag{35}$$

and, for n sufficiently large and some $\epsilon > 0$ and $c_7 > 0$,

$$g(q_{m(X),\alpha_n}) \ge c_7 \cdot (1-\alpha_n)^{1+\epsilon}. \tag{36}$$

Then

$$\left|\hat{q}^{(IS)}_{m_n(Z),N_n,\bar{\alpha}_n} - q_{m(X),\alpha_n}\right| = O_{\mathbf{P}}\left(\log n\cdot\delta_n + \frac{\log^2(n)}{N_n\cdot(1-\alpha_n)^{1/2+\epsilon}}\right). \tag{37}$$

Proof. By Theorem 1 we know that (29) is satisfied. Furthermore, (34), (35), (36) and Remark 3 imply that (30) and (32) hold. Trivially, (35) and $N_n \ge n$ imply (31). Hence the assumption of Theorem 2 are satisfied from which we conclude (via (35) and the definition of η_n) the assertion. □

Corollary 3. *Let the spline interpolant m_n be defined as in Corollary 1, set $N_n = n^2$ and let the quantile estimate $\hat{q}^{(IS)}_{m_n(X),N_n,\bar{\alpha}_n}$ be defined as in Theorem 2. Assume that m is (p,C)–smooth and that the assumption of Corollary 2 are satisfied for $K_n = [-(\log n)^\gamma, (\log n)^\gamma]^d$. Then*

$$\left|\hat{q}^{(IS)}_{m_n(Z),N_n,\bar{\alpha}_n} - q_{m(X),\alpha_n}\right| = O_{\mathbf{P}}\left(\frac{(\log n)^{\gamma\cdot p+1}}{n^{p/d}} + \frac{\log^2(n)}{n^2\cdot(1-\alpha_n)^{1/2+\epsilon}}\right).$$

Proof. Follows directly from Corollary 2 and (13). □

Remark 6. Under the assumptions of Corollary 3 it is possible to estimate quantiles of level

$$\alpha_n = 1 - n^{-r}$$

consistently for any $0 < r < 4/(1+2\cdot\epsilon)$.

Remark 7. In any application of the above estimate we have to choose η_n in some data-dependent way. To do this we propose to estimate separately the error of the surrogate and the error of the Monte Carlo approximation of the quantile corresponding to the surrogate. The error of the surrogate can be estimated by splitting the data used in the computation of the surrogate randomly into two parts of approximately equal size, by computing the surrogate with the first part of the data and by evaluating it on the second part. Here we propose to estimate the error of the surrogate by 10 times the average of the absolute pointwise errors of the surrogate on the second part of the data. In order to estimate the error due to the Monte Carlo approximation of the quantile corresponding to the surrogate, we propose to split the data used in the computation of the data into five parts of approximately equal sizes, to compute with each part of the data the quantile estimate and to use the maximal deviation between these five quantile estimates as the estimate of the error. Then we propose to choose η_n as the sum of the above two estimated errors. In the next section we will investigate the performance of this data-dependent choice of η_n by applying it to simulated data.

4 Application to simulated data

In this section we study the finite sample size behaviour of three different quantile estimates. The first one (*order stat.*) is the order statistics estimate defined by (3). The second one (*sur. quant.*) is the Monte Carlo surrogate quantile estimate of Section 2, but instead of a quasi-interpolant we use a smoothing spline (as implemented in the routine *Tps()* in *R* with smoothing parameter chosen by the generalized cross-validation as implemented in this routine). Since we apply it to data where the function is observed without additional error (i.e., in a noiseless regression estimation problem), this estimate results in an interpolating spline which gives similar result as the quasi-interpolant in Section 2, but is easier to implement. For our third estimate (*imp. quant.*) we implement our importance sampling surrogate sampling estimate as described in Remarks 4 and 5, and use the data-depend choice of η_n described in Remark 7.

We compare these three quantile estimates in three different models. In the first model the dimension of X is $d = 1$ and

$$m(x) = \exp(x) \quad (x \in \mathbb{R}).$$

In the second model the dimension of X is $d = 2$ and

$$m(x^{(1)}, x^{(2)}) = 50 \cdot \exp\left(-(x^{(1)})^2 - (x^{(2)})^2\right) \quad (x^{(1)}, x^{(2)} \in \mathbb{R}).$$

Finally in the third model the dimension of X is $d = 4$ and

$$m(x) = \sqrt{1 + \|x\|^2}.$$

Each time the d components of X are independent standard normally distributed random variables.

For all three estimates the sample sizes are chosen as $n = 100$, $n = 300$, $n = 1000$ and $n = 3000$. For the surrogate quantile estimate we use additional $N_{sur.quant.} = 10 \cdot n$ values of X, and the importance sampling surrogate quantile estimate selects its sample of Z from additional $N_{imp.samp.quant.} = 5 \cdot N_{sur.quant.} = 50 \cdot n$ values of X using the rejection method of Remark 4. Here the importance sampling surrogate quantile estimate uses a larger number of additional values of X than the surrogate quantile estimate since it needs to store and sort only the part of this data, which is selected by the rejection method and which is only a small part of this data, and can therefore be computed for much larger data sets than the surrogate quantile estimate.

The estimates are applied in all three models and with the four different sample sizes in order to estimate quantiles of level $\alpha = 0.995$ and $\alpha = 0.999$. In each case the estimates are applied to 100 different independent random samples, and the average relative absolute errors, i.e., the average of the absolute errors divided by the quantile (and in brackets its standard deviation), are listed in Table 1.

In all simulations in Table 1 the surrogate quantile estimate is better than the simple order statistics estimate, in the simulations with dimension $d \in \{1, 2\}$ its error is less than half of the error of the simple order statistics estimate and for $d = 4$ its error is again much smaller if the sample size is greater than 300. Most of the time the importance sampling quantile estimate is in turn much better than the surrogate quantile estimate, especially for $n = 3000$ its error is less than half of the error of the surrogate quantile estimate for $d \in \{1, 2\}$ and still substantially better for $d = 4$.

In Theorems 1 and 2 the error bounds on the quantile estimates consist of a sum of two terms, where compared to Theorem 1 in Theorem 2 only the second term is improved. Hence if the first term (which is related to the error of the surrogate estimate of m) dominates the sum of the two terms, we cannot expect that the importance sampling quantile estimate improves the surrogate quantile estimate. To illustrate this effect, we repeat the simulations of Table 1 for $n = 300$ with the size of the additional sample of X drastically increased (which decreases the second part of the error bound). The results are presented in Table 2. Here the importance sampling quantile estimate improves the surrogate quantile estimate in four out of six cases, but in most cases only slightly. It should be noted that the computation of the surrogate quantile estimate with such a large amount of additional data is in general not feasible since it is extremely time consuming. Also this effect will disappear as soon as the quality of the surrogate estimate is improved.

Finally we illustrate the usefulness of our newly proposed estimate by applying it to an engineering simulation model which was already presented in Enss et al. (2015), but which we present here again for the sake of completeness. Here we consider a physical model of a spring-mass-damper with active velocity feedback for the purpose of vibration isolation (cf., Figure 1). The aim is to analyze the uncertainty occuring in the maximal

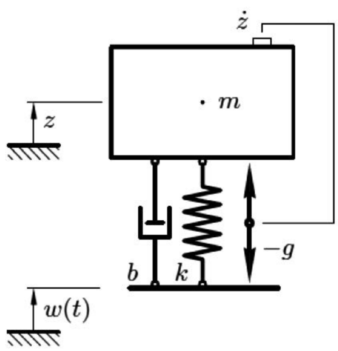

Figure 1: Spring-mass-damper with active velocity feedback (Platz and Enss (2015)).

magnification $|V_{max}|$ of the vibration amplitude in case that four parameters of the system, namely the system's mass (m), the spring's rigidity (k), the damping (b) and the active velocity feedback (g), are varied according to prespecified random processes. Based on the physical model of the spring-mass-damper, we are able to compute for given values of the above parameters the corresponding value of the maximal magnification

$$|V_{max}| = f(m, k, b, g)$$

of the vibration amplitude by a MATLAB program (cf., Platz and Enss (2015)), which needs approximately 0.2 seconds for one function evaluation. So our function $|V_{max}|$

is given by this MATLAB program, and computation of 300 function evaluations can be easily completed in approximately one minute, but computation of 100,000 values requires about 5.5 hours.

Our main interest is to predict the uncertainty of the maximal magnification of the vibration amplitude in case of uncertainty in the parameters of the spring-mass-damper. Here we model this uncertainty in the parameters by assuming that the parameters are realizations of normally distributed random variables with means and standard deviations derived from conditions typically occurring in practice. More precisely, we assume that the means of m, k, b and g are 1 kg, 1000 N/m, 0.095 Ns/m and 45 Ns/m, respectively, and their standard deviations are 0.017 kg, 33.334 N/m, 0.009 Ns/m, and 2.25 Ns/m, respectively.

We decribe the uncertainty in the corresponding (random) maximal magnification of the vibration amplitude by a confidence interval which contains the random value with probability 0.99. We estimate such a confidence interval by using estimates of the 0.995 and 0.005 quantile as upper and lower bounds of the interval, respectively. We use an order statistics with sample size $n = 100,000$ to compute a reference value for this confidence interval. This results in $|V_{max}| \in [0, 0.2522]$ dB. But if we want to estimate this interval using only $n = 300$ evaluations of our function, we get with order statistics, the surrogate quantile estimate (with $3,000$ additional values of X) and the importance sampling quantile estimate (with $15,000$ additional values of X) the intervals $[0, 0.2987]$ dB, $[0, 0.2168]$ dB and $[0, 0.2386]$ dB, resp. As we can see, the estimated interval of the importance sampling quantile estimate is much closer to our reference interval than the result of other two estimates.

5 Proofs

6 Proof of Theorem 1

Since (28) is an easy consequence of (26) and (27), we only have to prove (26). Let A_n be the event that $X_1, \ldots, X_{N_n}$ are all contained in K_n. By (25) we know that

$$\mathbf{P}(A_n^c) \leq N_n \cdot \mathbf{P}\{X \notin K_n\} \to 0 \quad (n \to \infty).$$

If A_n holds, then we have for any $i \in \{1, \ldots, N_n\}$

$$|m_n(X_i) - m(X_i)| \leq \delta_n \leq \frac{2 \cdot \delta_n}{2} + \frac{1}{2} \cdot |q_{m(X),\alpha_n} - m(X_i)|,$$

from which we conclude by Theorem 1 in Enss et al. (2014)

$$|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}| \leq 2 \cdot \delta_n + 2 \cdot |\hat{q}_{m(X),N_n,\alpha_n} - q_{m(X),\alpha_n}|.$$

This implies

$$\lim_{c \to \infty} \limsup_{n \to \infty} \mathbf{P} \left\{ \left| \hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n} \right| > c \cdot \left(\delta_n + \frac{\sqrt{1 - \alpha_n}}{\sqrt{N_n} \cdot g(q_{m(X),\alpha_n})} \right) \right\}$$

$$\leq \lim_{c\to\infty} \limsup_{n\to\infty} \left(\mathbf{P}(A_n^c) + \mathbf{P}\left\{ 2 \cdot \delta_n + 2 \cdot |\hat{q}_{m(X),N_n,\alpha_n} - q_{m(X),\alpha_n}| \right.\right.$$

$$\left.\left. > c \cdot \left(\delta_n + \frac{\sqrt{1-\alpha_n}}{\sqrt{N_n} \cdot g(q_{m(X),\alpha_n})} \right) \right\} \right)$$

$$\leq \lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\left\{ 2 \cdot |\hat{q}_{m(X),N_n,\alpha_n} - q_{m(X),\alpha_n}| > c \cdot \frac{\sqrt{1-\alpha_n}}{\sqrt{N_n} \cdot g(q_{m(X),\alpha_n})} \right\} = 0,$$

where the last equality follows from (7). □

7 Proof of Theorem 2

Set

$$\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} = \inf\left\{ y \in \mathbb{R} \,:\, \hat{G}_{m(Z),N_n}(y) \geq \bar{\alpha}_n \right\},$$

where

$$\hat{G}_{m(Z),N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m(Z_i) \leq y\}} \quad (y \in \mathbb{R}).$$

In the *first step of the proof* we observe that $Z_1, \dots, Z_{N_n} \in K_n$ implies

$$|\hat{q}^{(IS)}_{m_n(Z),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}| \leq 2 \cdot \delta_n + 2 \cdot |\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}|. \tag{38}$$

As in the proof of Theorem 1 this follows by an application of Theorem 1 in Enss et al. (2014).

In the *second step of the proof* we show that

$$N_n \cdot \mathbf{P}\{Z \notin K_n\} \to 0 \quad (n \to \infty) \tag{39}$$

holds outside of an event, whose probability tends to zero for $n \to \infty$. (Observe that the distribution of Z is random because it depends on the initial quantile estimate and therefore also the probability on the left-hand side of (39) is random).

Since

$$\begin{aligned} N_n \cdot \mathbf{P}\{Z \notin K_n\} &= N_n \cdot \int_{\mathbb{R}^d} 1_{\{z \notin K_n\}} \cdot h_n(z)\, dz \\ &= N_n \cdot \int_{\mathbb{R}^d} \frac{1}{c_n} \cdot 1_{\{z \notin K_n\}} \cdot f(z)\, dz \\ &= \frac{N_n \cdot \mathbf{P}\{X \notin K_n\}}{c_n} \end{aligned}$$

the assertion of step 2 follows directly from (32).

In the *third step of the proof* we show that outside of an event, whose probability tends to zero for $n \to \infty$, we have for all $y \in [q_{m(X),\alpha_n} - \frac{\eta_n}{3}, q_{m(X),\alpha_n} + \frac{\eta_n}{3}]$

$$G_{m(Z)}(y) = \frac{G_{m(X)}(y) - b_n}{c_n}, \tag{40}$$

where $G_{m(Z)}(y) = \mathbf{P}\{m(Z) \leq y\}$ is the cdf of $m(Z)$. Let C_n be the event that $|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}| \leq \eta_n/3$. By (29) we know that $\mathbf{P}\{C_n\} \to 1$ $(n \to \infty)$.

Let $y \in [q_{m(X),\alpha_n} - \frac{\eta_n}{3}, q_{m(X),\alpha_n} + \frac{\eta_n}{3}]$. On C_n we have for any $x \in K_n$ with $m(x) \leq y$

$$m_n(x) \leq m(x) + \delta_n \leq y + \delta_n \leq q_{m(X),\alpha_n} + \frac{\eta_n}{3} + \delta_n \leq \hat{q}_{m_n(X),N_n,\alpha_n} + \eta_n,$$

and for any $z \in K_n$ with $m_n(z) < \hat{q}_{m_n(X),N_n,\alpha_n} - \eta_n$

$$m(z) \leq m_n(z) + \delta_n \leq \hat{q}_{m_n(X),N_n,\alpha_n} - \eta_n + \delta_n \leq q_{m(X),\alpha_n} - \frac{\eta_n}{3} \leq y.$$

Hence we have

$$1_{\{m(x) \leq y\}} \cdot 1_{\{x \in K_n \,:\, m_n(x) > \hat{q}_{m_n(X),N_n,\alpha_n} + \eta_n\}} = 0$$

and

$$1_{\{m(x) \leq y\}} \cdot 1_{\{x \in K_n \,:\, m_n(x) < \hat{q}_{m_n(X),N_n,\alpha_n} - \eta_n\}} = 1_{\{x \in K_n \,:\, m_n(x) < \hat{q}_{m_n(X),N_n,\alpha_n} - \eta_n\}},$$

which implies

$$\begin{aligned}
& G_{m(Z)}(y) \\
&= \mathbf{P}\{m(Z) \leq y\} \\
&= \int_{\mathbb{R}^d} 1_{\{m(z) \leq y\}} \cdot h_n(z)\, dz \\
&= \int_{\mathbb{R}^d} 1_{\{m(z) \leq y\}} \cdot \frac{1}{c_n} \cdot \Bigg(1 - 1_{\{z \in K_n \,:\, m_n(z) < \hat{q}_{m_n(X),N_n,\alpha_n} - \eta_n\}} \\
&\qquad\qquad - 1_{\{z \in K_n \,:\, m_n(z) > \hat{q}_{m_n(X),N_n,\alpha_n} + \eta_n\}}\Bigg) \cdot f(z)\, dz \\
&= \frac{1}{c_n}\left(\int_{\mathbb{R}^d} 1_{\{m(z) \leq y\}} \cdot f(z)\, dz - \int_{\mathbb{R}^d} 1_{\{z \in K_n \,:\, m_n(z) < \hat{q}_{m_n(X),N_n,\alpha_n} - \eta_n\}} \cdot f(z)\, dz\right) \\
&= \frac{1}{c_n} \cdot \left(G_{m(X)}(y) - b_n\right).
\end{aligned}$$

In the *fourth step of the proof* we show that outside of an event, whose probability tends to zero for $n \to \infty$, we have

$$q_{m(Z),\bar{\alpha}_n} = q_{m(X),\alpha_n}. \tag{41}$$

By the definition of $q_{m(X),\alpha_n}$ and by the third step we know that outside of an event, whose probability tends to zero for $n \to \infty$, we have for $y \in [q_{m(X),\alpha_n} - \frac{\eta_n}{3}, q_{m(X),\alpha_n})$

$$G_{m(Z)}(y) = \frac{G_{m(X)}(y) - b_n}{c_n} < \frac{\alpha_n - b_n}{c_n} = \bar{\alpha}_n$$

and for $y \in [q_{m(X),\alpha_n}, q_{m(X),\alpha_n} + \frac{\eta_n}{3}]$

$$G_{m(Z)}(y) = \frac{G_{m(X)}(y) - b_n}{c_n} \geq \frac{\alpha_n - b_n}{c_n} = \bar{\alpha}_n.$$

This implies the assertion of the fourth step.

As a consequence of the fourth step of the proof we assume from now on w.l.o.g. that (41) holds.

In the *fifth step of the proof* we show that the assertion of Theorem 2 follows from

$$\begin{aligned}
&\left|\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}\right| \qquad (42)\\
&= O_{\mathbf{P}}\left(\log(n)\cdot\delta_n + \frac{\log(n)\cdot\eta_n}{\sqrt{N_n}}\cdot\frac{\sup\{g(y): y\in[q_{m(X),\alpha_n}-3\eta_n, q_{m(X),\alpha_n}+3\eta_n]\}}{\inf\{g(y): y\in[q_{m(X),\alpha_n}-3\eta_n, q_{m(X),\alpha_n}+3\eta_n]\}}\right).
\end{aligned}$$

This is a direct consequence of the steps 1 through 4, since by these steps we know that outside of an event, whose probability tends to zero for $n\to\infty$, we have

$$\begin{aligned}
\left|\hat{q}^{(IS)}_{m_n(Z),N_n,\bar{\alpha}_n} - q_{m(X),\alpha_n}\right| &= \left|\hat{q}^{(IS)}_{m_n(X),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}\right|\\
&\leq 2\cdot\delta_n + 2\cdot|\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}|.
\end{aligned}$$

In the *sixth step of the proof* we show that outside of an event, whose probability tends to zero for $n\to\infty$, we have for all $y\in[q_{m(X),\alpha_n}-\frac{\eta_n}{3}, q_{m(X),\alpha_n}+\frac{\eta_n}{3}]$

$$\left|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right| = \frac{1}{c_n}\cdot\int_{[\min\{y,q_{m(X),\alpha_n}\},\max\{y,q_{m(X),\alpha_n}\}]} g(z)\,dz.$$

This follows by applying the result of the third step, which yields

$$\begin{aligned}
\left|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right| &= \left|\frac{G_{m(X)}(y)-b_n}{c_n} - \frac{G_{m(X)}(q_{m(Z),\bar{\alpha}_n})-b_n}{c_n}\right|\\
&= \frac{1}{c_n}\cdot\left|G_{m(X)}(y) - G_{m(X)}(q_{m(Z),\bar{\alpha}_n})\right|\\
&= \frac{1}{c_n}\cdot\int_{[\min\{y,q_{m(X),\alpha_n}\},\max\{y,q_{m(X),\alpha_n}\}]} g(z)\,dz.
\end{aligned}$$

In the *seventh step of the proof* we show that outside of an event, whose probability tends to zero for $n\to\infty$, we have

$$\left|\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}\right| \leq \frac{\eta_n}{6}. \qquad (43)$$

(43) is implied by

$$\hat{G}_{m(Z),N_n}\left(q_{m(Z),\bar{\alpha}_n} - \frac{\eta_n}{6}\right) < \bar{\alpha}_n \leq \hat{G}_{m(Z),N_n}\left(q_{m(Z),\bar{\alpha}_n} + \frac{\eta_n}{6}\right).$$

Observing

$$\sup_{t\in\mathbb{R}}\left|G_{m(Z)}(t) - \hat{G}_{m(Z),N_n}(t)\right| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{N_n}}\right)$$

which follows from the Dvoretzky-Kiefer-Wolfowitz inequality (cf., e.g., Massart (1990)), we see that for (43) it suffices to show that we have outside of an event, whose probability tends to zero for $n \to \infty$,

$$G_{m(Z)}\left(q_{m(Z),\bar{\alpha}_n} - \frac{\eta_n}{6}\right) < \bar{\alpha}_n - \frac{\log n}{\sqrt{N_n}} \tag{44}$$

and

$$G_{m(Z)}\left(q_{m(Z),\bar{\alpha}_n} + \frac{\eta_n}{6}\right) > \bar{\alpha}_n + \frac{\log n}{\sqrt{N_n}}. \tag{45}$$

Let C_n again be the event that $|\hat{q}_{m_n(X),N_n,\alpha_n} - q_{m(X),\alpha_n}| \leq \eta_n/3$. On C_n we have by the result of step 6 for n sufficiently large

$$\begin{aligned}
&\bar{\alpha}_n - G_{m(Z)}\left(q_{m(Z),\bar{\alpha}_n} - \frac{\eta_n}{6}\right)\\
&= G_{m(Z)}\left(q_{m(Z),\bar{\alpha}_n}\right) - G_{m(Z)}\left(q_{m(Z),\bar{\alpha}_n} - \frac{\eta_n}{6}\right)\\
&= \frac{1}{c_n} \cdot \int_{q_{m(Z),\bar{\alpha}_n} - \frac{\eta_n}{6}}^{q_{m(Z),\bar{\alpha}_n}} g(z)\, dz\\
&= \frac{\int_{q_{m(Z),\bar{\alpha}_n} - \frac{\eta_n}{6}}^{q_{m(Z),\bar{\alpha}_n}} g(z)\, dz}{\int_{\mathbb{R}^d}\left(1_{\{z \in K_n : |m_n(z) - \hat{q}_{m_n(X),N_n,\alpha_n}| \leq \eta_n\}} + 1_{\{z \notin K_n\}}\right) \cdot f(z)\, dz}\\
&\geq \frac{\int_{q_{m(Z),\bar{\alpha}_n} - \frac{\eta_n}{6}}^{q_{m(Z),\bar{\alpha}_n}} g(z)\, dz}{\int_{\mathbb{R}^d}\left(1_{\{z \in K_n : |m(z) - q_{m(X),\alpha_n}| \leq 3\cdot\eta_n\}} + 1_{\{z \notin K_n\}}\right) \cdot f(z)\, dz}\\
&\geq \frac{\frac{\eta_n}{6} \cdot \inf\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq \frac{\eta_n}{6}\}}{6 \cdot \eta_n \cdot \sup\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq 3 \cdot \eta_n\} + \mathbf{P}\{X \notin K_n\}}\\
&\geq \frac{1}{42} \cdot \frac{\inf\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq \frac{\eta_n}{3}\}}{\sup\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq 3 \cdot \eta_n\}},
\end{aligned}$$

where the last inequality is implied by (30). Application of (31) yields (44). Analogously we obtain (45), which completes the proof of the seventh step.

In the *eighth step of the proof* we show that

$$\left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{N_n}}\right). \tag{46}$$

The definition of $\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}$ implies that we have for arbitrary $\epsilon > 0$

$$\begin{aligned}
&\left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right|\\
&\leq \left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - \hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n})\right| + \left|\hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - \bar{\alpha}_n\right|\\
&\quad + \left|\bar{\alpha}_n - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right|\\
&\leq \left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - \hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n})\right|
\end{aligned}$$

$$
\begin{aligned}
&+\hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - \hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - \epsilon) + \left|\bar{\alpha}_n - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right| \\
\leq &\left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - \hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n})\right| \\
&+ \left|\hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n})\right| \\
&+ \left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - \epsilon)\right| \\
&+ \left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - \epsilon) - \hat{G}_{m(Z),N_n}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - \epsilon)\right| + \left|\bar{\alpha}_n - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right|.
\end{aligned}
$$

The third step of the proof implies that ouside of an event, whose probability tends to zero for $n \to \infty$, $G_{m(Z)}$ is continuous in a neighborhood of $q_{m(X),\alpha_n}$ and satisfies

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

$>$From this and step 7 we can conclude

$$
\left|G_{m(Z)}(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}) - G_{m(Z)}(q_{m(Z),\bar{\alpha}_n})\right| \leq 3 \cdot \sup_{t \in \mathbb{R}} \left|G_{m(Z)}(t) - \hat{G}_{m(Z),N_n}(t)\right|.
$$

Application of the Dvoretzky-Kiefer-Wolfowitz inequality (cf., e.g., Massart (1990)) completes the eighth step of the proof.

In the *ninth step of the proof* we complete the proof of Theorem 2. Steps 6 and 7 and arguments in the proof of step 7 imply that we have outside of an event, whose probability tends to zero for $n \to \infty$,

$$
\begin{aligned}
&\left|G_{m(Z)}\left(\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}\right) - G_{m(Z)}\left(q_{m(Z),\bar{\alpha}_n}\right)\right| \\
&= \frac{1}{c_n} \cdot \int_{[\min\{\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}, q_{m(Z),\bar{\alpha}_n}\}, \max\{\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n}, q_{m(Z),\bar{\alpha}_n}\}]} g(z)\, dz \\
&\geq \frac{|\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}| \cdot \inf\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq \frac{\eta_n}{3}\}}{7 \cdot \eta_n \cdot \sup\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq 3 \cdot \eta_n\}}.
\end{aligned}
$$

By step 8 we can conclude that the probability of the event

$$
\left\{ \frac{|\hat{q}^{(IS)}_{m(Z),N_n,\bar{\alpha}_n} - q_{m(Z),\bar{\alpha}_n}| \cdot \inf\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq \frac{\eta_n}{3}\}}{7 \cdot \eta_n \cdot \sup\{g(z) \,:\, |q_{m(X),\alpha_n} - z| \leq 3 \cdot \eta_n\}} \geq \frac{\log n}{\sqrt{N_n}} \right\}
$$

tends to zero for $n \to \infty$. This implies (42). The proof is complete. □

8 Acknowledgment

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 805 and the Natural Sciences and Engineering Research Council of Canada for additional support.

References

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

[2] Asmussen, S. and Kroese, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Advances of Applied Probability* **38**, pp. 545–558.

[3] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes.* Wiley.

[4] 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.

[5] Blanchet, J. and Liu, J. (2010). Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *Journal of Applied Probability* **47**, pp. 301-322.

[6] Chu, F. and Nakayama, M. (2012). Confidence intervals for quantiles when applying variance reduction techniques. *ACM Transactions on Modeling and Computer Simulation* **22**, Art. 10, pp. 10:1–10:25.

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

[8] 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.

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

[10] 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.

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

[12] Dupuis, P. and Wang, H. (2005). Dynamic importance sampling of uniformly recurrent Markov chains. *Annals of Applied Probability* **15**, pp. 1–38.

[13] Egloff, D. and Leippold, M. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics* **38**, pp. 1244-1278.

[14] Enss, G. C., Kohler, M., Krzyżak, A. and Platz, R. (2014) Nonparametric quantile estimation based on surrogate models. Submitted for publication.

[15] Falk, M. (1989). A note on uniform asymptotic normality of intermediate order statistics. *Annals of the Institute of Statistical Mathematics*, **41**, pp. 19-29.

[16] Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2002). Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, **12**, pp. 239-269.

[17] Glynn, P. W. (1996). Importance sampling for Monte Carlo estimation of quantiles. *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation.* Publishing House of Saint Petersburg University, Saint Petersburg, pp. 180–185.

[18] Greblicki, W. and Pawlak, M. (1985). Fourier and Hermite series estimates of regression functions. *Annals of the Institute of Statistical Mathematics* **37**, pp. 443-454.

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

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

[21] Hong, J. L., Hu, Z. and Liu, G. (2014). Monte Carlo methods for value-at-risk and conditional value-at-risk: a review. *ACM Transactions on Modeling and Computer Simulation* **24**, Art. 22, pp. 22:1–22:37.

[22] Hult, H. and Svensson, J. (2009). Efficient calculation of risk measures using importance sampling - the heavy tailed case. Preprint arXiv:0909.3335

[23] Kohler, M. (2000). Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression. *Journal of Statistical Planning and Inference* **89**, pp. 1–23.

[24] Kohler, M. (2014). Optimal global rates of convergence for noiseless regression estimation problems with adaptively chosen design. *Journal of Multivariate Analysis* **132**, pp. 197-208.

[25] Kohler, M. and Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Transactions on Information Theory* **47**, pp. 3054–3058.

[26] Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Transactions on Information Theory* **41**, pp. 677–687.

[27] Kohler, M., Krzyżak, A., and Walk, H. (2014). Nonparametric recursive quantile estimation *Statistics and Probability Letters* **93**, pp. 102-107.

[28] Kohler, M., Krzyżak, A., Tent, R. and Walk, H. (2014). Nonparametric quantile estimation using importance sampling. Submitted for publication.

[29] Liu, J. and Yang, X. (2012). The convergence rate and asymptotic distribution of the bootstrap quantile variance estimator for importance sampling. *Advances in Applied Probability*, **44**, pp. 815-841.

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

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

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

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

[34] Nakayama, M. (2014). Confidence intervals for quantiles using sectioning when applying variancereduction techniques. *ACM Transactions on Modeling and Computer Simulation*, Art. 19, pp. 19:1–19:21.

[35] Platz, R. and Enss, G. C. (2015). Comparison of Uncertainty in Passive and Active Vibration Isolation. Proceedings of the $33rd$ IMAC, Feb. 2-5, 2015, Orlando, FL/USA. Accepted for publication.

[36] Rafajłowicz, E. (1987). Nonparametric orthogonal series estimators of regression: A class attaining the optimal convergence rate in L2. *Statistics and Probability Letters* **5**, pp. 219-224.

[37] Siegmund, D. (1976). Importance sampling in Monte Carlo study of sequantial tests. *Annals of Statistics* **4**. pp. 673–684.

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

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

[40] Sweeting, T. J. (1985). On domains of uniform local attraction in extreme value theory. *The Annals of Probability* **13**, pp. 196-205.

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

[42] Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia, PA.

n	model	d	α	order stat.	sur. quant.	imp. quant.
100	1	1	0.995	0.3490 (0.3018)	0.1046 (0.0773)	0.0706 (0.0629)
100	1	1	0.999	0.4478 (0.1851)	0.2727 (0.1283)	0.1812 (0.1068)
100	2	2	0.995	0.0142 (0.0156)	0.0064 (0.0055)	0.0042 (0.0041)
100	2	2	0.999	0.0192 (0.0211)	0.0060 (0.0053)	0.0041 (0.0043)
100	3	4	0.995	0.0702 (0.0510)	0.0668 (0.0450)	0.0747 (0.0337)
100	3	4	0.999	0.1174 (0.0637)	0.0700 (0.0595)	0.1138 (0.0641)
300	1	1	0.995	0.2401 (0.2546)	0.0726 (0.0515)	0.0384 (0.0296)
300	1	1	0.999	0.3398 (0.2829)	0.1387 (0.0926)	0.0837 (0.0674)
300	2	2	0.995	0.0078 (0.0076)	0.0024 (0.0019)	0.0011 (0.0008)
300	2	2	0.999	0.0044 (0.0060)	0.0013 (0.0011)	0.0006 (0.0004)
300	3	4	0.995	0.0401 (0.0356)	0.0337 (0.0238)	0.0323 (0.0127)
300	3	4	0.999	0.0713 (0.0511)	0.0515 (0.0357)	0.0537 (0.0215)
1000	1	1	0.995	0.1106 (0.0806)	0.0371 (0.0299)	0.0181 (0.0134)
1000	1	1	0.999	0.2179 (0.1759)	0.0703 (0.0485)	0.0323 (0.0196)
1000	2	2	0.995	0.0039 (0.0037)	0.0012 (0.0009)	0.0005 (0.0004)
1000	2	2	0.999	0.0027 (0.0030)	0.0006 (0.0006)	0.0002 (0.0002)
1000	3	4	0.995	0.0248 (0.0185)	0.0134 (0.0087)	0.0133 (0.0051)
1000	3	4	0.999	0.0440 (0.0296)	0.0282 (0.0213)	0.0296 (0.0086)
3000	1	1	0.995	0.0675 (0.0471)	0.0225 (0.0164)	0.0086 (0.0064)
3000	1	1	0.999	0.1263 (0.0950)	0.0362 (0.0299)	0.0165 (0.0139)
3000	2	2	0.995	0.0019 (0.0014)	0.0006 (0.0005)	0.0003 (0.0002)
3000	2	2	0.999	0.0011 (0.0011)	0.0003 (0.0002)	0.0001 (0.0001)
3000	3	4	0.995	0.0149 (0.0109)	0.0065 (0.0045)	0.0046 (0.0028)
3000	3	4	0.999	0.0273 (0.0174)	0.0162 (0.0110)	0.0144 (0.0053)

Table 1: Simulation result for the three different models with $N_{sur.quant.} = 10 \cdot n$ and $N_{imps.samp.quant.} = 50 * n$. Reported are the relative absolute errors of the estimates (and in brackets their standard deviations) in 100 independent simulations.

n	model	d	α	order stat.	sur. quant	imp. quant.
300	1	1	0.995	0.1981 (0.2158)	0.0101 (0.0150)	0.0077 (0.0158)
300	1	1	0.999	0.3281 (0.2805)	0.0647 (0.0602)	0.0634 (0.0608)
300	2	2	0.995	0.0078 (0.0077)	0.0004 (0.0004)	0.0004 (0.0004)
300	2	2	0.999	0.0052 (0.0059)	0.0004 (0.0004)	0.0004 (0.0004)
300	3	4	0.995	0.0483 (0.0343)	0.0343 (0.0060)	0.0342 (0.0059)
300	3	4	0.999	0.0591 (0.0409)	0.0597 (0.0088)	0.0595 (0.0087)

Table 2: Simulation result for the three different models with $N_{sur.quant.} = 1.000.000$ and $N_{imps.samp.quant.} = 8.000.000$.