Ann Inst Stat Math (2018) 70:439–465
https://doi.org/10.1007/s10463-016-0595-4

Nonparametric quantile estimation using importance sampling

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

Received: 10 December 2015 / Revised: 31 October 2016 / Published online: 19 January 2017

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 a 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.

Keywords Nonparametric quantile estimation · Importance sampling · Rate of convergence

✉ Adam Krzyżak
krzyzak@cs.concordia.ca

Michael Kohler
kohler@mathematik.tu-darmstadt.de

Reinhard Tent
tent@mathematik.tu-darmstadt.de

Harro Walk
walk@mathematik.uni-stuttgart.de

[1] Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany

[2] Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada

[3] Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany

1 Introduction

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

$$Y = m(X),$$

where X is an $\mathbb{R}^d$-valued random variable with density $f : \mathbb{R}^d \to \mathbb{R}$ and $m : \mathbb{R}^d \to \mathbb{R}$ is a black-box function with the possibility of expensive evaluation at arbitrarily chosen design points. Let

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

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

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

using at most n evaluations of 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

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

and to use the corresponding plug-in estimate

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

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

$$Y_{1:n} \le \cdots \le Y_{n:n}.$$

Since

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

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 et al. (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 $Z, Z_1, Z_2, \ldots$ 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) and Neddermeyer (2009) and the literature cited therein).

Quantile estimation using importance sampling has been considered by Cannamela et al. (2008), Egloff and Leippold (2010), and Morio (2012). For quantile estimation, the choice of the optimal importance sampling density is even more subtle than in the case of estimation of means. One idea which is mentioned several times in the articles cited above is to try to choose the optimal importance sampling estimate, such that the variance of the estimate

$$1 - \frac{1}{n}\sum_{i=1}^{n} I_{\{Z_i > y\}} \cdot \frac{f(Z_i)}{h(Z_i)}$$

of $G(y)$ at the point $y = q_\alpha$ is minimal. In this case, h_{opt} minimizes

$$\mathbf{Var}\left(\frac{1}{n}\sum_{i=1}^{n} I_{\{Z_i > q_\alpha\}} \cdot \frac{f(Z_i)}{h(Z_i)}\right).$$

This leads to the optimal density

$$h_{\text{opt}}(x) = \frac{I_{\{x>q_\alpha\}} \cdot f(x)}{1-\alpha},$$

in which case the above variance is zero. The corresponding order statistics estimate of q_α is the minimal value of $Z_1, \ldots, Z_n$, which usually converges rather fast towards q_α. However, of course, this nice estimate cannot be used in practice, since it relies on the knowledge of the quantity q_α.

All three papers above proposed new estimates in various models; however, only Egloff and Leippold (2010) investigated theoretical properties (consistency) of their method. None of the papers contains any results on the rates of convergence. A recursive method was proposed by Kohler et al. (2014).

As pointed out by a referee, the Approximate Bayesian Computation (ABC) is another promising approach towards quantile estimation (cf., e.g., Dunson and Taylor 2005 or Lancaster and Jun 2010). The main advantage of ABC is that it also leads to confidence intervals for the quantile. How to apply this efficiently in the simulation model considered in this paper is an open problem.

An IS approach leads to replacement of m by a surrogate function m_n which can be evaluated cheaply at arbitrary points. To construct the surrogate m_n, any kind of nonparametric regression estimate can be used. 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 et al. 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 a parametric context, the Kriging approximation method was used by Oakley (2004) for quantile estimation without IS and by Dubourg et al. (2013) for the related, but simpler problem of estimating the distribution value function of m(X) at zero employing importance sampling.

In this paper, we propose a new importance sampling quantile estimate and analyze its rates of convergence. We do this in a completely nonparametric context, using a mild smoothness assumption on m (not knowing the structure of m) in view of a good approximation by a surrogate function. The basic idea is to use an initial estimate of the quantile based on the order statistics of samples of $m(X)$ 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, so we sample only from an area, where the values are especially important for the computation of the quantile. Our final estimate of the quantile is then defined as an order statistic of $m(Z)$, where the level of the order statistic takes into account that we sample only from a part of the original density f. 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 $\frac{\log^{1.5} n}{n}$. Approximation is done by quasi-spline interpolation. In Sect. 4 dealing with simulations, also Kriging approximations are used.

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 floor and ceiling of z, i.e., 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 $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$, the partial derivative $\frac{\partial^k m}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$ exists and satisfies

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

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

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

for some finite constant $c_1 > 0$.

The estimate of the quantile is defined in Sect. 2. The main result is formulated in Sect. 3, and proofs are provided in Sect. 5. In Sect. 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 use n_1 evaluations of m to generate an initial estimate of q_α, n_2 evaluations of m to construct an approximation of m, and we 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 Sect. 1. To improve it by importance sampling, we use additional observations $(x_1, m(x_1)), \ldots, (x_{n_2}, m(x_{n_2}))$ of m at points $x_1, \ldots, x_{n_2} \in \mathbb{R}^d$ and use an estimate

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

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 $\beta_n > 0$, that is

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

Set

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

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

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

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

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

Here

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

and

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

can be computed exactly for given f and m_n. In our application below, we approximate them by Monte Carlo. 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

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

provided that, 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.

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

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

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

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

Let Z, Z_1, Z_2, …be independent and identically distributed and set:

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

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

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

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

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

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 $l_n^p / n^{p/d}$.

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 $\operatorname{supp}(B_{k,M}) = [u_k, u_{k+M+1}]$. For $M = 0$, B-spline $B_{k,0}$ is the indicator function of the interval $[u_k, u_{k+1})$, and for $M = 1$, we have

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

(so-called hat-function). The general recursive definition of $B_{k,M}$ can be found, e.g., in de Boor (1978), or in Sect. 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 Mth 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 using the 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 the 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 $t_{k,j} \in \operatorname{supp}(B_{k,M}) = [k\cdot l_n/K, (k+M+1)\cdot l_n/K]$. 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 system of linear equations), such that

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

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 $n_2 \ge (2M+1)^d$. Furthermore, we choose $x_1, \ldots, x_{n_2}$, such that all 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 are possible, since $(2M\cdot K+1)^d \le n_2$. Then, we define

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

where Qm is the above-defined quasi-interpolant satisfying Eqs. (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 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 $c_4 > 0$

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

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

3 Main results

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

Theorem 1 *Assume that X is an $\mathbb{R}^d$-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 $\bar{q}_{m(Z),\bar{\alpha},n_3}$ of q_α be defined as in Sect. 2 with $\beta_n = \frac{\log n}{\sqrt{n}}$ and assume that the regression estimate m_n satisfies (Eq. 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 of Sect. 2 is used to estimate m, then we get the following result.

Corollary 1 *Assume that X is an $\mathbb{R}^d$-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 Sect. 2 with $M \geq p - 1$ and define the estimate $\bar{q}_{m(Z),\bar{\alpha},n_3}$ of q_α as in Sect. 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) upon 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 assertion (maybe modified) 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

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

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 (cf., Algorithm 1) to simulated data and estimate the corresponding 90-, 95-, 99-, and 99.9%-quantile. For this purpose, the number n of observations is set to 200, 500, 1000, 2000, 5000, and 10,000, respectively. As suggested in Sect. 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 set to $\frac{\log(n)}{\sqrt{n}}$, and our estimate of m is the quasi-interpolant introduced in Sect. 2 with $M = 3$, $l_n = \log(n)$ and $K = K(n) = \left\lfloor (\lfloor n_2^{1/d} \rfloor - 1)/2M \right\rfloor$ or, as the case may be, a thin plate spline as implemented in the routine Tps() of the package *fields* of *R*, a free software environment for statistical computing. Since the value $\bar{\alpha}$ which is needed for our importance sampling estimate is not known in reality, we estimate it by Monte Carlo estimation. To this end, we use N additional evaluations of m_n. The value of N will be set to 20,000 and 50,000, respectively.

For comparison, we, furthermore, estimate the same quantiles using three alternative estimation methods which rely on the same number n of observed data. The first of these methods is the plug-in or order statistics estimate (OS) as defined by (2). The second quantile estimate uses the non-adaptive surrogate as a control variate (CV), as explained in Sect. 2 in Cannamela et al. (2008). This method relies on a simplified version of the function m. For this purpose, we use in our simulations the same estimate m_n which is used in our importance sampling estimation simulations. Since m_n relies on n_1 evaluations of m, for the remaining procedure of the CS algorithm $n - n_1$ evaluations of m remain. In addition, the CS method requires knowledge about the α quantile of $m_n(X)$. In the simulations, an order statistics estimate of this quantile will be used, which relies on $N = 20{,}000$, or $N = 50{,}000$ additional evaluations of m_n, respectively. The third estimation method uses the non-adaptive surrogate for controlled stratification (CS), as explained in Sect. 3.2 in Cannamela et al. (2008). Since again this method relies on simplification of function m, the estimate m_n will be used. The remaining $n - n_1$ evaluations are split approximately equally in the three different strata. Again, knowledge of the quantiles of $m_n(X)$ is needed and they will be approximated by order statistics relying on $N = 20{,}000$, or $N = 50{,}000$ additional evaluations of m_n, respectively. For the strata, we partition the interval [0, 1] at the points 0.8 and 0.9 for $\alpha = 0.9$, at the points 0.85 and 0.95 for $\alpha = 0.95$, at the points 0.9 and 0.99 for $\alpha = 0.99$, and, finally, at the points 0.95 and 0.999 for $\alpha = 0.999$.

In practice, it might occur that the value of $\bar{\alpha}$ 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 lie outside of the set the random variable Z (as defined in Sect. 2) is concentrated on. There are several ways to tackle this problem. In the following, we use a very simple strategy that uses a somewhat loose definition of the *quantile*. More precisely, if $\bar{\alpha} \leq 0$, then we just take the smallest value of our new sample, and if $\bar{\alpha} > 1$, we take the largest one.

In our first example, X is random variable having the standard normal distribution and 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

set $n1 = \lfloor n/3 \rfloor$, $n2 = \lfloor n/3 \rfloor$ and $n3 = n - n1 - n2$;
compute a first sample of $n1$ values drawn from $m(X)$, named `y1`;
set `quantile` as the $\lceil n1 \cdot \alpha \rceil$-th smallest value of `y1`;
set `lb_quantile` = `quantile` $- 4 \cdot \log(n)/\sqrt{n}$;
set `ub_quantile` = `quantile` $+ 3 \cdot \log(n)/\sqrt{n}$;
choose at most $n2$ proper design points and compute an estimate `m_n` of m;
use numerical integration to compute

1. `gamma1` $= \int_{[-\log(n),\log(n)]^d} 1_{\{x\,:\,\texttt{m\_n}(x) < \texttt{lb\_quantile}\}} \cdot f(x)\mathrm{d}x$
2. `gamma2` $= \int_{[-\log(n),\log(n)]^d} 1_{\{x\,:\,\texttt{m\_n}(x) > \texttt{ub\_quantile}\}} \cdot f(x)\mathrm{d}x$;

set `alpha_bar` $= (\alpha -$ `gamma1`$)/(1 -$ `gamma1` $-$ `gamma2`$)$;
`/* generate IS sample from new distribution */`
generate an array and set $i = 1$;
while $i \leq n3$ **do**
  generate a d-dimensional vector `z`, drawn from the distribution of X;
  if $z \notin [-\log(n), \log(n)]^d$ *or* $m\_n(z) \in [\mathit{lb\_quantile}, \mathit{ub\_quantile}]$ **then**
    store `z` as i-th element of the array and increase i by 1;
  end
end
compute function values of m at the stored vectors and set `result` as the $\lceil n3 \cdot \alpha \rceil$-th biggest of these values;
return `result`;

Algorithm 1: Proposed IS quantile estimate using at most n evaluations of the function m

estimate with a quasi interpolant of degree $M = 3$. As mentioned before, to compute γ_1 and γ_2, we use Monte Carlo estimation with $N = 20{,}000$ additional evaluations of m_n. The whole procedure is repeated 100 times for different values of n. In Table 1, the average squared error of our estimated quantile values (ASE IS) can be found with a precision of two significant digits. Here, the ASE IS is defined as $\frac{1}{100} \cdot \sum_{i=1}^{100} (q - \bar{q}_i)^2$, where q is the true quantile (which is known) and $\bar{q}_i$ denotes the ith of our estimates. As mentioned before, we estimate the quantiles using three additional estimates. The average squared errors of these can be found in Table 1 as well.

As can be seen from Table 1, the error of our newly proposed estimate is in this example much smaller than the error of the order statistics and about the size of the errors of the two additional estimates. In this example, we also computed the errors of the optimal IS density mentioned in the introduction (OPT IS in Table 1). These errors are much smaller than the errors of our IS quantile estimate; however, its definition depends on the quantile to be estimated and can, therefore, never be applied in practice.

In our second example, we set $X = (X_1, X_2)$, where random variables X_1 and X_2 are independent standard normally distributed random variables 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 which we apply in our estimate. As in our first example, we use Monte Carlo estimation relying on $N = 20{,}000$ additional evaluations of m_n to compute γ_1 and γ_2. As before, we repeat this procedure 100 times for different values of n and compare the results with those of the three additional estimates described at the beginning of this section. The results can be found in Table 2. As before, the error of our newly proposed estimate is in this

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

size of n	90%-quantile $q_{0.9} \approx 3.6022$				95%-quantile $q_{0.95} \approx 5.1803$			
	200	500	1000	2000	200	500	1000	2000
ASE OS	0.18	0.068	0.053	0.021	0.57	0.23	0.12	0.052
ASE CV	0.035	0.017	0.004	0.0027	0.32	0.078	0.017	0.011
ASE CS	0.0062	0.0021	0.0023	0.0023	0.012	0.0073	0.008	0.0063
ASE IS	0.05	0.0091	0.0048	0.0024	0.38	0.1	0.055	0.021
ASE OPT IS	0.00016	0.000026	0.0000051	0.0000017	0.0004	0.000064	0.000016	0.0000036

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

size of n	90%-quantile $q_{0.9} \approx 4.8656$				95%-quantile $q_{0.95} \approx 5.678$			
	200	500	1000	2000	200	500	1000	2000
ASE OS	0.072	0.029	0.012	0.0072	0.083	0.046	0.023	0.014
ASE CV	0.019	0.0034	0.0015	0.00082	0.054	0.013	0.0035	0.002
ASE CS	0.0018	0.001	0.00067	0.00076	0.0023	0.00093	0.0013	0.0013
ASE IS	0.021	0.0073	0.0022	0.00097	0.024	0.0056	0.0022	0.0015

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

size of n	90%-quantile $q_{0.9} \approx 4.6052$				95%-quantile $q_{0.95} \approx 5.9915$			
	200	500	1000	2000	200	500	1000	2000
ASE OS	0.2	0.091	0.033	0.02	0.36	0.12	0.076	0.032
ASE CV	0.033	0.008	0.0034	0.0024	0.19	0.033	0.011	0.0068
ASE CS	0.0063	0.0023	0.0019	0.0019	0.01	0.0041	0.0025	0.0043
ASE IS	0.081	0.0065	0.0029	0.0023	0.15	0.027	0.011	0.0096

example much smaller than the error of the order statistics and about the order of the errors of the additional estimates.

In our third example, we set $X = (X_1, X_2)$ for independent random variables X_1 and X_2 with the standard normal distribution and choose $m(x_1, x_2) = x_1^2 + x_2^2$. Consequently, $m(X)$ is Chi-square random variable with two degrees of freedom. The results for our estimate are presented in Table 3. Once again, the error of our newly proposed estimate is much smaller than the error of the order statistics and about the order of the errors of the additional estimates.

In the following two examples, we increase the dimensionality and set $X = (X_1, X_1, X_3, X_4, X_5)$ for independent standard normally distributed random variables $X_1, \ldots, X_5$. Here, in contrast to the previous approach, we cannot use quasi-

Table 4 Simulation results for $\max\{x_1, x_2, x_3, x_4, x_5\}$ and $\exp(-x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_5^2)$

m	$\max\{x_1, x_2, x_3, x_4, x_5\}$				$\exp(-x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_5^2)$			
	90%-quantile		95%-quantile		90%-quantile		95%-quantile	
	$q_{0.9} \approx 2.0365$		$q_{0.95} \approx 2.3187$		$q_{0.9} \approx 0.1998$		$q_{0.95} \approx 0.3181$	
size of n	5000	10,000	5000	10,000	5000	100,000	5000	10,000
ASE OS	0.00037	0.00017	0.00062	0.00038	0.000045	0.000028	0.00012	0.000071
ASE CV	0.000092	0.000044	0.00046	0.00014	0.000026	0.000011	0.00002	0.000013
ASE CS	0.000057	0.00004	0.0001	0.000069	0.000005	0.0000045	0.000012	0.000013
ASE IS	0.000098	0.000059	0.00013	0.000054	0.00011	0.000043	0.00032	0.00015

interpolants, based on B-splines to approximate m, since in five dimensions, this method requires too many design points. Therefore, instead, we utilize the procedure $Tps()$ of the statistics package R. One advantage over quasi-interpolants which depend on B-splines is that the design points do not have to be equidistant.

In our fourth example, function m is given by $m(x_1, x_2, x_3, x_4, x_5) = \max\{x_1, x_2, x_3, x_4, x_5\}$. Naturally, since we are in a higher dimensional case as before, the usage of a small sample size will not be sufficient. Therefore, we increase the sample size to $n = 5000$ and $n = 10{,}000$, respectively, and the number of additional evaluations of N to 50,000. Likewise before, the results for this higher dimensional example can be found in Table 4. Here, again, the error of our newly proposed estimate is much smaller than the error of the order statistics and comparable to the errors of the two additional estimates.

Our fifth example is again a function with five-dimensional input, and so, again, the thin plate splines are the interpolation method of choice. Here, m is given by $m(x_1, x_2, x_3, x_4, x_5) = \exp(-x_1^2-x_2^2-x_3^2-x_4^2-x_5^2)$ and $X = (X_1, X_1, X_3, X_4, X_5)$, whereas $X_1, \ldots, X_5$ are independent standard normal random variables. The simulation results are presented in Table 4. Even though the average squared error of our proposed method is larger than the error of the order statistic, it seems that the error of the importance sampling estimate decreases a little faster than the error of the order statistic. Yet, the results for the order statistic are still better in case of $m(x_1, x_2, x_3, x_4, x_5) = \exp(-x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_5^2)$, though all measured errors are fairly low. This is because the quantile we try to estimate is moderately high, but the findings will change when we increase the parameter α. Setting α to 99 and 99.9%, respectively, yields significant changes in the results as can be seen in Table 5. Especially in case of the 99.9% quantile, our proposed method performs significantly better than the order statistics and does even outperform the additional estimates for $n = 10{,}000$.

In our sixth example, X is a one-dimensional standard normally distributed random variable again and we choose $m(x) = \sin(x)$. Thus, $m(X)$ has several modes. To identify the true quantiles of $m(X)$, numerical approximation is used. Since in this example, the resulting ASE are very small in Table 6 and we present the summed squared error (SSE) instead which is $\sum_{i=1}^{100}(q - \bar{q}_i)^2$. As one can see in this scenario,

Table 5 Simulation results for $\max\{x_1, x_2, x_3, x_4, x_5\}$ and $\exp(-x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_5^2)$ with larger quantiles

m	$\max\{x_1, x_2, x_3, x_4, x_5\}$				$\exp(-x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_5^2)$			
	99%-quantile		99.9%-quantile		99%-quantile		99.9%-quantile	
	$q_{0.99} \approx 2,8769$		$q_{0.999} \approx 3.5400$		$q_{0.99} \approx 0.5745$		$q_{0.999} \approx 0.8104$	
size of n	5000	10,000	5000	10,000	5000	10,000	5000	10,000
ASE OS	0.0026	0.001	0.011	0.0083	0.00038	0.00016	0.00068	0.00046
ASE CV	0.0047	0.002	0.021	0.016	0.00067	0.00015	0.0027	0.00085
ASE CS	0.00024	0.00022	0.0019	0.00089	0.000044	0.000037	0.000074	0.000096
ASE IS	0.00022	0.00016	0.017	0.00082	0.0004	0.000058	0.00026	0.000045

Table 6 Simulation results for $m(x) = \sin(x)$

size of n	90%-quantile $q_{0.9} \approx 0.91568$				95%-quantile $q_{0.95} \approx 0.97744$			
	200	500	1000	2000	200	500	1000	2000
SSE OS	0.084	0.054	0.024	0.011	0.015	0.0096	0.0044	0.0015
SSE CV	0.034	0.006	0.0016	0.0013	0.015	0.0021	0.00043	0.00026
SSE CS	0.0037	0.0013	0.00096	0.0011	0.00077	0.00021	0.00022	0.00022
SSE IS	0.4	0.089	0.035	0.011	0.051	0.013	0.0063	0.0021
SSE IS*	0.47	0.0089	0.0027	0.0013	0.24	0.0055	0.00075	0.0004

the quality of our proposed importance sampling estimate is about as good as the order statistic, even if the function in the simulation model has several modes, which makes the estimation of our importance sampling estimate especially difficult. However, here, both additional estimates feature significantly smaller errors.

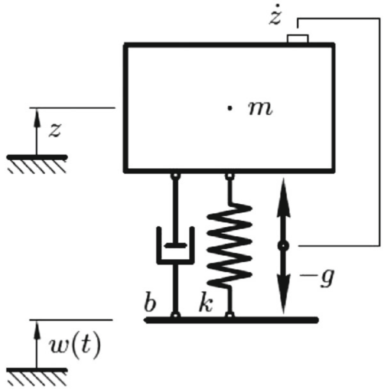
Let us recall that in the definition of the proposed quantile estimate, some parameters were chosen based on asymptotic behavior. While modifications by a constant factor do not change the asymptotic behaviour, they might change the finite sample size behaviour of the estimate. As a test, we altered the values of a_n and b_n to

$$a_n = \overline{q}_{\alpha,n_1} - \frac{\log n}{3 \cdot \sqrt{n}} \quad \text{and} \quad b_n = \overline{q}_{\alpha,n_1} + \frac{\log n}{3 \cdot \sqrt{n}},$$

and repeated the simulations of the previous example. The results can be found in Table 6 as SSE IS* and are significantly better. This shows that for practical use, a careful choice of the parameters used may improve the estimation results. How such a choice can be made is a topic for further research.

In our last example, function m is motivated by experiments of the Collaborative Research Centre 805 at the Technische Universität Darmstadt, which investigates uncertainty in load-bearing systems. Here, we use a physical model of the so-called

Fig. 1 Spring-mass-damper with active velocity feedback



spring-mass-damper with active velocity feedback, which returns the maximal magnification of the vibration amplitude, according to four chosen parameters. To put it briefly, one can think of a mass (m) that is placed on a suspension leg, resulting in some oscillations which are reduced by the system's stiffness (k) and damping (b) (cf. schematic drawing in Fig. 1). Now, in the simulation, the oscillation frequency is set to 100 times per second, and according to an accurate dynamic system model, which is described in detail in Platz and Enss (2015), the maximal occurring magnitude is simulated using the Matlab procedure *bode*. The corresponding Matlab code was kindly provided by *G*. Enss.

The previously mentioned four input parameters are named the system's mass (m), the spring's rigidity (k), the damping of the mass's oscillation (b), and the active velocity feedback (g). The active velocity feedback controls the extent of an additional counter force, depending on the velocity of the oscillations. In case g equals zero, no active damping mechanism is applied, which is called passive damping. One evaluation of function m takes approximately 0.2 s, so computation of 2000 function evaluations can be easily completed in approximately 7 min. In contrast, the computation of 100,000 values requires about 5.5 h.

In the following, we distinguish between two cases: first, the passive case, where as mentioned before the active velocity feedback g equals zero, and second, the active case, where the value of g is given by the normally distributed random variable with mean 45 and a standard deviation of 2.25. In both cases, in our simulation, the remaining variables are also normally distributed, but their means and standard deviations are different. In line with Platz and Enss (2015), the means of m, k, and b are 1, 1000, and 0.095, respectively, and their standard deviations are 0.017, 33.334, and 0.009, respectively. In the active case, we simulate the value of $x = (m, k, b, g)$ with independent random variables, as defined beforehand and use our method with $n = 2000$, to estimate the 95% quantile of the maximal magnification of the vibration amplitude.

Since the computer model was implemented in Matlab, the computation of the quantile estimate this time was done in Matlab as well. Therefore, to approximate m, a Kriging estimator provided by the Matlab toolbox DACE was used. Kriging, which is basically an interpolation method based on weighted averages of sample points, produces similar results as the thin plate spline and again has the advantage that the design points do not have to be equidistant.

As a result, we get 0.10350, as 95%-quantile. For comparison, we also estimate this quantile with order statistics and a sample of size 2000 and 100,000, respectively. In the first case, the computed value is 0.09791, whereas in the latter case, we get a value of 0.10129. Taking the order statistics with sample size 100,000 as an accurate estimate, we compare the deviations of the other two estimates from 0.10129. Here, one can see that the deviation of the order statistic with sample size 2000 is about 1.53 times larger than the deviation of our proposed method.

In the passive case, we simulate the value of $x = (m, k, b, g = 0)$ as explained before. Again, we use our method with $n = 2000$, to estimate the corresponding 95% quantile. Since now, our example is again three dimensions, we can use B-splines and so we compute our estimate both with Kriging and a quasi-interpolant of degree $M = 3$. In either case, we obtain a value of 51.929, as 95%-quantile. As before, we estimate this quantile with order statistics as well. The computed value when using a sample size of 2000 is 51.851, whereas for a sample size of 100,000, we get a value of 51.919. Again, we compare how much the estimates with sample size 2000 differ from the value 51.919 of the order statistic with sample size 100,000. Here, we observe that the deviation of the order statistic with sample size 2000 is about 6.8 times larger than the one for our proposed method. These results show that our estimate performs better than the simple estimate based on the order statistics using the same sample size.

5 Proofs

We will use the following lemma to prove Lemma 1.

Lemma 2 *Assume that X is an $\mathbb{R}^d$-valued random variable which has a density f 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 $\epsilon > 0$

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

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

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

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 $q_\alpha - \epsilon < u \leq q_\alpha$. By definition of Z, we have

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

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

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

From this, we conclude

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

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

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

The proof is complete. □

Proof of Lemma 1. 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 characterize 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 $\mathbf{P}\{C_n\}$ tends to one as $n \to \infty$. Therefore, we observe that according to (5), for all $x \in A_n$ and all $y \in B_n$, we have

$$m(x) \leq m_n(x) + \beta_n < \bar{q}_{\alpha,n_1} - 2 \cdot \frac{\log n}{\sqrt{n}} - \beta_n$$

and

$$m(y) \geq m_n(y) - \beta_n > \bar{q}_{\alpha,n_1} + 2 \cdot \frac{\log n}{\sqrt{n}}.$$

This implies

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

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

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

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

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

□

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

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

Then, $G_{m(Z)}$ is outside of an event, whose probability tends to zero for $n \to \infty$, at Lebesgue-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 $|q_\alpha - \bar{q}_{\alpha,n_1}| \leq \frac{\log n}{\sqrt{n}}$. 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

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

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

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

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

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

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

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

for Lebesgue-almost all points y of the interval I. This relation also trivially 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, 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

$$\frac{\mathrm{d}}{\mathrm{d}u} G_{m(Z)}^{-1}(u) = \frac{1}{c_2 \cdot g(G_{m(Z)}^{-1}(u))}.$$

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 neighborhood 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

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

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

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

and

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

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 are outside of an event, whose probability tends to zero for $n \to \infty$. Application of the inverse function theorem implies

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

for all $u \in \tilde{I}$. Notice that since $c_2 \ge 1$, 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. First, note 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

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

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

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

Let $G_{m(Z)}$ be the cdf of $m(Z)$, that is

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

and set

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

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

Since

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

has the same distribution as

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

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

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

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

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

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

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

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 et al. 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

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

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). 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

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

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 by 1 from above. Finally, (3) implies the assertion. □

6 Conclusion

This paper is concerned with estimation of quantiles, given a computer model of some complex technical system, which depends on some non-deterministic input. With increasing complexity, the costs to evaluate those models, measured in time or memory, increase. The usage of importance sampling techniques paves the way to improve the estimate without additional costly evaluations of the given simulation model. One of the main tasks in the application of importance sampling is to find a suitable auxiliary importance sampling density function. In this paper, a possible choice for such a density is presented and the asymptotic behaviour of the corresponding estimate is analyzed. Application on simulated data shows that our newly proposed estimate clearly outperforms the classical order statistics, and performs similarly to alternative approaches based on control variates or controlled stratification. However, in contrast to the latter two approaches, for which no convergence proof exists, we provide the proof of convergence for our newly proposed estimate.

Furthermore, simulation 6 shows that a data-dependent choice of the parameter β_n might improve the performance of our estimate. This will be studied in a forthcoming paper, together with the influence of the Monte Carlo estimation of the integrals on the rate of convergence of the estimate.

Acknowledgements The first author would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 805. The second author would like to acknowledge the support from the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2015-06412.

References

Arnold, B. C., Balakrishnan, N., Nagaraja, H. N. (1992). *A first course in order statistics*. New York: Wiley.

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

Cannamela, C., Garnier, J., Iooss, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics*, *2*(4), 1554–1580.

de Boor, C. (1978). *A practical guide to splines*. New York: Springer.

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*, 467–481.

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

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

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

Devroye, L., Györfi, L., Lugosi, G. (1996). *A probabilistic theory of pattern recognition*. New York: Springer.

Dubourg, V., Sudret, B., Deheeger, F. (2013). Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Engineering Mechanics*, *33*, 47–57.

Dunson, D. B., Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *Nonparametric Statistics*, *17*, 385–400.

Egloff, D., Leippold, M. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics*, *38*(2), 1244–1278.

Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. New York: Springer.

Greblicki, W., Pawlak, M. (1985). Fourier and Hermite series estimates of regression functions. The *Annals of the Institute of Statistical Mathematics*, *37*, 443–454.

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

Györfi, L., Kohler, M., Krzyżak, A., Walk, H. (2002). *A distribution-free theory of nonparametric regression*. New York: Springer Series in Statistics, Springer.

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

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

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

Kohler, M., Krzyżak, A., Walk, H. (2014). Nonparametric recursive quantile estimation. *Statistics and Probability Letters*, *93*, 102–107.

Lancaster, T., Jun, S. J. (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics*, *25*, 287–307.

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

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

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

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

Neddermeyer, J. C. (2009). Computationally efficient nonparametric importance sampling. *Journal of the American Statistical Association*, *104*(486), 788–802.

Oakley, J. (2004). Estimating percentiles of uncertain computer code outputs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *53*(1), 83–93.

Platz, R. Enss G.C. (2015). Comparison of uncertainty in passive and active vibration isolation. In *Proceedings of the 33rd IMAC* (pp. 15–25). Orlando, FL.

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

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, *5*, 595–645.

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

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

Wahba, G. (1990). *Spline models for observational data*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.