Nonparametric quantile estimation based on surrogate models

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

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

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

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

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

February 11, 2016

Abstract

Nonparametric estimation of a quantile $q_{m(X),\alpha}$ 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 an $\mathbb{R}^d$-valued random variable with known distribution. Monte Carlo surrogate quantile estimates are considered, where in a first step the function m is estimated by some estimate (surrogate) m_n and then the quantile $q_{m(X),\alpha}$ is estimated by a Monte Carlo estimate of the quantile $q_{m_n(X),\alpha}$. A general error bound on the error of this quantile estimate is derived which depends on the local error of the function estimate m_n, and the rates of convergence of the corresponding Monte Carlo surrogate quantile estimates are analyzed for two different function estimates. The finite sample size behavior of the estimates is investigated in simulations.

AMS classification: Primary 62G05; secondary 62G30.

Key words and phrases: Nonparametric quantile estimation, surrogate models, rate of convergence.

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

Running title: *Nonparametric quantile estimation*

1 Introduction

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

$$Y = m(X),$$

where X is an $\mathbb{R}^d$-valued random variable with known distribution $\mathbf{P}_X$ and $m : \mathbb{R}^d \to \mathbb{R}$ is a black box function which can be computed at any point $x \in \mathbb{R}^d$ but which is costly to evaluate. Let

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

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

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

using at most n evaluations of function m.

A simple idea to estimate $q_{m(X),\alpha}$ is to use observations $m(X_1), \ldots, m(X_n)$, where $X_1, \ldots, X_n$ is an i.i.d. sample of X, to compute the empirical cdf

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

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

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

Since $\hat{q}_{m(X),n,\alpha}$ is in fact an order statistic, results from order statistics, e.g., Theorem 8.5.1 in Arnold, Balakrishnan and Nagaraja (1992), imply that in case that $m(X)$ has a density g which is continuous and positive at $q_{m(X),\alpha}$ we have

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

This implies

$$|\hat{q}_{m(X),n,\alpha} - q_{m(X),\alpha}| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{3}$$

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

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

There are quite a few approaches studied already in the literature for improving the rate of convergence of the above simple quantile estimate in a simulation model of a costly-to-evaluate function. These include variance reduction techniques like control variates (cf., e.g., Hesterberg and Nelson (1998)), controlled stratification (cf., e.g., Cannamela, Garnier and Ioss (2008)) and importance sampling (cf., e.g., Glynn (1996) for a parametric and Morio (2012) for a nonparametric approach), and Bayesian methods, in particular Bayesian methods based on Gaussian process modelling (cf., e.g., Santner, Williams and Notz (2003)). For the to quantile estimation related problem of rare event simulation an extensive survey is presented in Morio et al. (2014).

In this paper we study estimates based on so–called surrogate models in non-Bayesian setting. The basic idea is to first construct an estimate m_n of m and then to estimate the quantile $q_{m(X),\alpha}$ by a Monte Carlo estimate of the quantile $q_{m_n(X),\alpha}$, where

$$q_{m_n(X),\alpha} = \inf \left\{ y \in \mathbb{R} \, : \, \mathbf{P}_X\{x \in \mathbb{R}^d \, : \, m_n(x) \leq y\} \geq \alpha \right\}.$$

Our main result concerns an analysis of the error of this Monte Carlo estimate. We show that if the local error of m_n is small in areas where $m(x)$ is close to $q_{m(X),\alpha}$, i.e., if for some small $\delta_n > 0$

$$|m_n(x) - m(x)| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |m(x) - q_{m(X),\alpha}| \quad \text{for } \mathbf{P}_X\text{-almost all } x,$$

then the error of the Monte Carlo estimate $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$ of $q_{m(X),\alpha}$ is small, i.e.,

$$\left| \hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right| = O_{\mathbf{P}} \left(\delta_n + \frac{1}{\sqrt{N_n}} \right),$$

where N_n is the sample size of the Monte Carlo estimate (cf., Theorem 1 below). We use this result to analyze the rate of convergence of two different estimates, where for the first estimate the error of m_n is globally small but where for the second estimate it is only locally small. Here we show in particular that if m is (p, C)–smooth, i.e., roughly speaking (see below for the exact definition), if m is p-times continuously differentiable, then the first estimate is able to achieve (up to some logarithmic factor) a rate of convergence of order $n^{-p/d}$ (as compared to the rate $n^{-1/2}$ of the order statistics estimate above), but the second one is able to achieve (again up to some logarithmic factor) a rate of convergence of order $n^{-(p/d)-(p/d^2)-(p^2/d^2)}$.

In order 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, 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)).

The idea of estimating the distribution of a random variable $m(X)$ by the distribution of $m_n(X)$, where m_n is a suitable surrogate (or estimate) of m, has been considered already in quite a few papers. E.g., surrogate models have been introduced and investigated with the aid of simulated and real data in connection with quadratic response surfaces in Bucher and Burgund (1990), Kim and Na (1997) and Das and Zheng (2000), in connection with support vector machines in Hurtado (2004), Deheeger and Lemaire (2010) and Bourinet, Deheeger and Lemaire (2011), in connection with neural networks in Papadrakakis and Lagaros (2002), and in connection with kriging in Kaymaz (2005) and Bichon et al. (2008). Theoretical results concerning the rate of convergence of the corresponding estimates are not derived in these papers.

As a tool to derive various versions of importance sampling algorithms surrogate models have been used in Dubourg, Sudret and Deheeger (2013) and in Kohler, Krzyżak, Tent and Walk (2014), where in the latter article theoretical results have also been provided.

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

$$\mathrm{diam}(A) = \sup\left\{\|x - z\| \,:\, x, z \in A\right\}.$$

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 $f : \mathbb{R}^d \to \mathbb{R}$ 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 $\frac{\partial^k f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$ exists and satisfies

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

for all $x, z \in \mathbb{R}^d$. In other words, a function f is (p, C)–smooth if and only if it belongs to the Hölder space $C^{k,s}$ (the space of functions that are k times differentiable and such that all their partial derivatives of order k are Hölder continuous with exponent s), with seminorm $|f|_{k,s} = \max_{|\beta|=k} |D^\beta f|_{0,s} = C$.

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

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

A general error bound on Monte Carlo surrogate quantile estimates is presented in Section 2. Results concerning the rate of convergence of estimates based on non-adaptively and adaptively chosen surrogates are presented in Section 3 and in Section 4, respectively. In Section 5 we illustrate the finite sample size performance of the estimates using simulated data. The proofs are contained in Section 6.

2 A general error bound

Let X, X_1, X_2, ... be independent and identically distributed random variables. In this section we consider a general Monte Carlo surrogate quantile estimate, which is defined as follows: In a first step data

$$(x_1, m(x_1)), \dots, (x_n, m(x_n))$$

is used to construct an estimate

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

of m. Here $x_i = X_i$ is one possible choice for the values of $x_1, \dots, x_n \in \mathbb{R}^d$, but not the only one (see Sections 3 and 4 below). Then $X_{n+1}, \dots, X_{n+N_n}$ are used to define a Monte Carlo estimate of the α-quantile of $m_n(X)$ by

$$\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} = \inf\left\{y \in \mathbb{R} : \hat{G}^{(MC)}_{m_n(X),N_n}(y) \geq \alpha\right\},$$

where

$$\hat{G}^{(MC)}_{m_n(X),N_n}(y) = \frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{m_n(X_{n+i}) \leq y\}}.$$

Intuitively it is clear that the error of m_n will influence the error of the above quantile estimate. Our main result states that for the error of the above quantile estimate it is not important that the local error of m_n is small in areas where m is far away from the quantile to be estimated.

Theorem 1 *Let X be an $\mathbb{R}^d$-valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and let $\alpha \in (0,1)$. Define the Monte Carlo surrogate quantile estimate $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$ of $q_{m(X),\alpha}$ as above. For $n \in \mathbb{N}$ let $\beta_n, \delta_n > 0$ be such that the estimate m_n satisfies*

$$|m_n(x) - m(x)| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m(X),\alpha} - m(x)| \quad \textit{for } \mathbf{P}_X\textit{–almost all } x \in [-\beta_n, \beta_n]^d, \tag{4}$$

and assume that

$$N_n \cdot \mathbf{P}\left\{X \notin [-\beta_n, \beta_n]^d\right\} \to 0 \quad (n \to \infty). \tag{5}$$

Then we have

$$\left|\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha}\right| = O_{\mathbf{P}}\left(\delta_n + \frac{1}{\sqrt{N_n}}\right).$$

Remark 1. Condition (4) is in particular satisfied if we choose

$$\delta_n = 2 \cdot \|m_n - m\|_{\infty, \text{supp}(\mathbf{P}_X) \cap [-\beta_n, \beta_n]^d},$$

so Theorem 1 implies

$$\left|\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha}\right| = O_{\mathbf{P}}\left(\|m_n - m\|_{\infty, \text{supp}(\mathbf{P}_X) \cap [-\beta_n, \beta_n]^d} + \frac{1}{\sqrt{N_n}}\right). \tag{6}$$

However, in general we can derive from Theorem 1 a much better bound on the error of the quantile estimate $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$, since it is not important for a small error bound that the error of the estimate m_n be small at points x where $m(x)$ is far away from $q_{m(X),\alpha}$ (cf., (4)).

Remark 2. In the proof of Theorem 1 we will see that (4) can be replaced by the weaker condition that with probability one the inequality

$$|m_n(X_{n+i}) - m(X_{n+i})| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m(X),\alpha} - m(X_{n+i})| \tag{7}$$

holds simultaneously for all $i \in \{1, \dots, N_n\}$ where $X_i \in [-\beta_n, \beta_n]^d$. We will use this condition in Section 4 in order to construct an adaptive surrogate.

3 A surrogate quantile estimate based on a non-adaptively chosen surrogate

In this section we choose m_n as a non-adaptively chosen spline approximand in the definition of our Monte Carlo surrogate quantile estimate.

To do this, we choose $r > 1$ and set $\beta_n = (\log n)^r$. Next we define a spline approximand which approximates m on $[-\beta_n, \beta_n]^d$. In order to do this, 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. Here our presentation is based on Kohler (2014), which in turn is an extension of the material presented in Chapters 14 and 15 of Györfi et al. (2002) to the case $d > 2$.

Choose $K \in \mathbb{N}$ and $M \in \mathbb{N}_0$, and set $u_k = k \cdot \beta_n / K$ $(k \in \mathbb{Z})$. 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 $(u_l)_{l \in \mathbb{Z}}$ and support $\text{supp}(B_{k,M}) = [u_k, u_{k+M+1}]$. In case $M = 0$ this means that $B_{k,0}$ is the indicator function of the interval $[u_k, u_{k+1})$, and for $M = 1$ we have

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

(so-called hat-function). The general 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 derivative of the functions have jump points only at the knots u_l $(l \in \mathbb{Z})$.

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

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

And we define $\mathcal{S}_{K,M}$ as the set of all linear combinations of all those of the above tensor product B-splines, where the support has nonempty intersection with $[-\beta_n, \beta_n]^d$, 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, that they are equal to a (multivariate) polynomial of degree less than or equal to M (in each component) on each rectangle

$$[u_{k_1}, u_{k_1+1}) \times \cdots \times [u_{k_d}, u_{k_d+1}) \quad (\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d), \tag{8}$$

and that they vanish outside of the set

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

Next we define spline approximands using so-called quasi interpolands: For a function $f : [-\beta_n, \beta_n]^d \to \mathbb{R}$ we define an approximating spline by

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

where

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

for some $a_{\mathbf{k},\mathbf{j}} \in \mathbb{R}$ and for suitably chosen points $t_{k,j} \in \text{supp}(B_{k,M}) \cap [-\beta_n, \beta_n]$. It can be shown that if we set

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

and

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

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

$$|Q_{\mathbf{k}} f| \leq c_1 \cdot \|f\|_{\infty,[u_{k_1}, u_{k_1+M+1}] \times \cdots \times [u_{k_d}, u_{k_d+M+1}]} \tag{9}$$

for any $\mathbf{k} \in \{-M, -M+1, \ldots, K-1\}^d$, any $f : [-\beta_n, \beta_n]^d \to \mathbb{R}$ and some universal constant c_1, and such that Q reproduces polynomials of degree M or less (in each component) on $[-\beta_n, \beta_n]^d$, i.e., for any multivariate polynomial $p : \mathbb{R}^d \to \mathbb{R}$ of degree M or less (in each component) we have

$$(Qp)(x) = p(x) \quad (x \in [-\beta_n, \beta_n]^d) \tag{10}$$

(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 interpoland. We fix the degree $M \in \mathbb{N}$ and set

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

Furthermore we choose $x_1, \ldots, x_n$ such that all of the $(2 \cdot M \cdot K + 1)^d$ points of the form

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

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

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

where Qm is the above defined quasi interpoland satisfying (9) and (10). The computation of Qm requires only function values of m at the points $x_1, \ldots, x_n$, i.e., the estimate depends on the data

$$(x_1, m(x_1)), \ldots, (x_n, m(x_n)),$$

and hence m_n is well defined.

Theorem 2 *Let X be an $\mathbb{R}^d$-valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and let $\alpha \in (0,1)$. Assume that $m(X)$ has a density which is continuous and positive at $q_{m(X),\alpha}$ and that m is (p,C)-smooth for some $p > 0$ and some $C > 0$. Let $r > 1$ and define the Monte Carlo surrogate quantile estimate $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$ of $q_{m(X),\alpha}$ as in Section 2, where m_n is the spline approximand defined above with parameter $M \geq p - 1$.*

Assume furthermore that

$$N_n \cdot \mathbf{P}\{X \notin [-(\log n)^r, (\log n)^r]^d\} \to 0 \quad (n \to \infty). \tag{11}$$

Then

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

In particular, if we set $N_n = \lceil n^{2p/d} / (\log n)^{2 \cdot r \cdot p} \rceil$ then we get

$$\left| \hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right| = O_{\mathbf{P}} \left(\frac{(\log n)^{r \cdot p}}{n^{p/d}} \right).$$

Proof. The definition of our spline approximand and the (p,C)-smoothness of m imply that

$$\|m_n - m\|_{\infty, [-\beta_n, \beta_n]^d} \leq c_2 \cdot (\log n)^{r \cdot p} \cdot n^{-p/d}$$

(cf., e.g., proof of Theorem 1 in Kohler (2014)). From this we can conclude that (4) is satisfied for

$$\delta_n = 2 \cdot c_2 \cdot (\log n)^{r \cdot p} \cdot n^{-p/d}$$

and consequently Theorem 1 yields the assertion. $\square$

Remark 3. It follows from Theorem 2 that in case that m is (p, C)-smooth for some $p > d/2$ or some $p > d$, respectively, the above Monte Carlo surrogate quantile estimate achieves a rate of convergence better than $n^{-1/2}$ or n^{-1}, respectively whenever

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

is satisfied.

Remark 4. It follows from Markov inequality that (11) is for instance satisfied if we have for some $c_3, s > 0$

$$\mathbf{E}\{\exp(c_3 \cdot \|X\|)\} < \infty \quad \text{and} \quad \frac{N_n}{n^s} \to 0 \quad (n \to \infty).$$

This holds, since by application of Markov inequality we have

$$N_n \cdot \mathbf{P}\{X \notin [-(\log n)^r, (\log n)^r]^d\} \leq N_n \cdot \mathbf{P}\{c_3 \cdot \|X\| > c_4 (\log n)^r\} \leq N_n \cdot \frac{\mathbf{E}\{\exp(c_3 \cdot \|X\|)\}}{\exp\left(c_4 (\log n)^r\right)},$$

and since $r > 1$ implies

$$\frac{n^s}{\exp\left(c_4 (\log n)^r\right)} \to 0 \quad (n \to \infty)$$

for any $s > 0$ and any $c_4 > 0$ as $s - c_4 (\log n)^{r-1} \to \infty$ as $n \to \infty$.

4 A surrogate quantile estimate based on an adaptively chosen surrogate

In the sequel we define an adaptive surrogate Monte Carlo quantile estimate. In order to simplify the presentation, we first present a simple partitioning estimate which achieves a good rate of convergence for Hölder-smooth m and then we explain how to extend the definition such that the estimate achieves a very good rate of convergence in case of higher smoothness.

Our partitioning estimate depends on a partition $\mathcal{P}_n = \{A_0, A_1, \ldots, A_{n-1}\}$ of $\mathbb{R}^d$ and on the evaluation of m at points $x_{A_0} \in A_0$, $x_{A_1} \in A_1$, $\ldots$, $x_{A_{n-1}} \in A_{n-1}$, i.e., on the data

$$(x_{A_0}, m(x_{A_0})), \ldots, (x_{A_{n-1}}, m(x_{A_{n-1}})). \tag{12}$$

For $x \in \mathbb{R}^d$ denote by $A_n(x)$ that cell $A_j \in \mathcal{P}_n$ which contains x. Then the partitioning estimate m_n is defined by

$$m_n(x) = m(x_{A_n(x)}). \tag{13}$$

The key trick in the definition of our adaptive partitioning estimate is an adaptive choice of the sets A_0, A_1, $\ldots$, A_{n-1} (the values of the points $x_{A_j} \in A_j$ are not important). Here our main goal is to define m_n such that

$$|m_n(X_{n+i}) - m(X_{n+i})| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m(X),\alpha} - m(X_{n+i})| \tag{14}$$

holds for $i \in \{1, \dots, N_n\}$ for some small $\delta_n > 0$ (cf., Remark 2).

To do this, we start by subdividing our data (12) into two parts of size

$$n_1 = \left\lceil \frac{n}{2} \right\rceil \quad \text{and} \quad n_2 = n - n_1.$$

We choose $r > 1$, set

$$C_n := [-(\log n)^r, (\log n)^r]^d$$

and partition C_n into $\lfloor n_1^{1/d} \rfloor^d$ equivolume cubes of side length $2 \cdot (\log n)^r / \lfloor n_1^{1/d} \rfloor$. We denote these cubes by A_j ($j = 1, \dots, \lfloor n_1^{1/d} \rfloor^d$), set $A_0 = \mathbb{R}^d \setminus C_n$ and let m_{n_1} be the partitioning estimate corresponding to the partition $\mathcal{P}_{n_1} = \{A_j : j = 0, 1, \dots, \lfloor n_1^{1/d} \rfloor^d\}$ of $\mathbb{R}^d$, where for $A \in \mathcal{P}_{n_1}$ the point $x_A \in A$ is arbitrarily chosen.

Assume that m is (p, C)–smooth for some $p \leq 1$. Then we have for any $x \in C_n$:

$$|m_{n_1}(x) - m(x)| \quad \leq \quad C \cdot \|x_{A_{n_1}(x)} - x\|^p \leq C \cdot \mathrm{diam}(A_{n_1}(x))^p,$$

where $\mathrm{diam}(A) = \sup\{\|x_1 - x_2\| : x_1, x_2 \in A\}$ denotes the diameter of set A. By construction of $\mathcal{P}_{n_1}$ this implies

$$|m_{n_1}(x) - m(x)| \leq \frac{1}{2} \cdot (\log n)^{r \cdot p + 1} \cdot n^{-p/d}$$

for n sufficiently large.

We use m_{n_1} to define the Monte Carlo surrogate quantile estimate

$$\hat{q}^{(MC)}_{m_{n_1}(X), N_n, \alpha} = \inf\{y \in \mathbb{R} : \hat{G}^{(MC)}_{m_{n_1}(X), N_n}(y) \geq \alpha\} \tag{15}$$

where

$$\hat{G}^{(MC)}_{m_{n_1}(X), N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m_{n_1}(X_{n+i}) \leq y\}}.$$

If

$$N_n \geq n^{2p/d} \quad \text{and} \quad N_n \cdot \mathbf{P}\{X \notin C_n\} \to 0 \quad (n \to 0),$$

then we know from the proof of Theorem 1 below that outside of an event whose probability tends to zero we have

$$\left|\hat{q}^{(MC)}_{m_{n_1}(X), N_n, \alpha} - q_{m(X), \alpha}\right| \leq \frac{(\log(n)^{r \cdot p + 1}}{n^{p/d}} \tag{16}$$

and (as already derived above) we have for any partition $\mathcal{P}_n \supseteq \mathcal{P}_{n_1}$ (which we will construct in the sequel) and the corresponding partitioning estimate m_n

$$|m_n(x) - m(x)| \leq \log(n) \cdot \mathrm{diam}(A_n(x))^p \quad \text{for all } x \in C_n. \tag{17}$$

Upon application of the triangle inequality we can conclude from (16)

$$\left|\hat{q}^{(MC)}_{m_{n_1}(X), N_n, \alpha} - m_n(X_{n+i})\right| \leq \frac{(\log n)^{r \cdot p + 1}}{n^{p/d}} + \left|q_{m(X), \alpha} - m(X_{n+i})\right| + |m(X_{n+i}) - m_n(X_{n+i})| \tag{18}$$

for $i \in \{1, \dots, N_n\}$.

If (16) (and hence also (18)) holds, then (14) holds as well for all $i \in \{1, \dots, N_n\}$ if for all $i \in \{1, \dots, N_n\}$ at least one of the following two conditions is satisfied:

$$|m_n(X_{n+i}) - m(X_{n+i})| \leq \frac{\delta_n}{2} \tag{19}$$

or

$$\begin{aligned} |m_n(X_{n+i}) - m(X_{n+i})| \quad &\leq \quad \frac{\delta_n}{2} + \frac{1}{2} \cdot |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(X_{n+i})| - \frac{1}{2} \cdot \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} \\ &\quad - \frac{1}{2} \cdot |m_n(X_{n+i}) - m(X_{n+i})|. \end{aligned} \tag{20}$$

Here (20) is equivalent to

$$3 \cdot |m_n(X_{n+i}) - m(X_{n+i})| \leq \delta_n + |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(X_{n+i})| - \frac{(\log n)^{r \cdot p+1}}{n^{p/d}}. \tag{21}$$

Using the bound (17) for the pointwise error of our estimate m_n we see that for sufficiently large n (14) holds for all $i \in \{1, \dots, N_n\}$ if for all $i \in \{1, \dots, N_n\}$ at least one of the following two conditions is satisfied:

$$2 \cdot \log(n) \cdot \mathrm{diam}(A_n(X_{n+i}))^p \leq \delta_n \tag{22}$$

or

$$3 \cdot \log(n) \cdot \mathrm{diam}(A_n(X_{n+i}))^p + \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(X_{n+i})| \leq \delta_n. \tag{23}$$

Hence we aim at choosing our partition so that the following term is small:

$$\begin{aligned} &\max_{i \in \{1,\dots,N_n\}, X_{n+i} \in C_n} \min \Bigg\{ 2 \cdot \log(n) \cdot \mathrm{diam}(A_n(X_{n+i}))^p, \\ &3 \cdot \log(n) \cdot \mathrm{diam}(A_n(X_{n+i}))^p + \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(X_{n+i})| \Bigg\}. \end{aligned} \tag{24}$$

We do this recursively as follows: Given $A_0, \dots, A_K$ for some $\lfloor n_1^{1/d} \rfloor^d \leq K \leq n - 2^d$, we choose such set A_j $(1 \leq j \leq K)$ that there exists $X_{n+i} \in A_j$ such that

$$\begin{aligned} &\min \Bigg\{ 2 \cdot \log(n) \cdot \mathrm{diam}(A_j)^p, \\ &3 \cdot \log(n) \cdot \mathrm{diam}(A_j)^p + \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(X_{n+i})| \Bigg\} \end{aligned} \tag{25}$$

is maximal (among all $X_{n+k} \in C_n, k \in \{1, \dots, N_n\}$). Then we subdivide this set into 2^d equivolume sets, and apply recursively the same procedure again until the number of evaluations of m is larger than $n - 2^d$.

As our next result (Theorem 3 below) shows that Monte Carlo surrogate quantile estimate corresponding to the partitioning estimate m_n with the adaptively chosen partition (described above) achieves the rate of convergence better than that in Theorem 2 provided the set of all x values, where $m(x)$ is "close" to the true quantile $q_{m(X),\alpha}$ is in some sense "small".

Theorem 3 *Let X be an $\mathbb{R}^d$-valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and let $\alpha \in (0,1)$. Assume that $m(X)$ has a density which is continuous and positive at $q_{m(X),\alpha}$ and that m is (p,C)–smooth for some $p \leq 1$. Define the Monte Carlo surrogate quantile estimate $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$ of $q_{m(X),\alpha}$ as in Section 2 where m_n is the adaptive partitioning estimate defined above. Assume furthermore*

$$N_n \geq n^{2p/d} \quad \text{and} \quad N_n \cdot \mathbf{P}\{X \notin [-(\log n)^r, (\log n)^r]\} \to 0 \quad (n \to \infty). \tag{26}$$

a) *Then*

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

b) *If, in addition, for some $c_5, \epsilon_0 > 0$ we have that for any $0 < h \leq \epsilon \leq \epsilon_0$*

$$\{x \in \mathbb{R}^d \,:\, m(x) \in [q_{m(X),\alpha} - \epsilon, q_{m(X),\alpha} + \epsilon]\} \tag{27}$$

can be covered by at most $c_5 \cdot \frac{\epsilon}{h^{d-1}}$ (open) balls of radius h, then

$$\left| \hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right| = O_{\mathbf{P}} \left(\frac{(\log n)^{5+2\cdot r}}{n^{(p/d)+(p/d^2)+(p^2/d^2)}} + \frac{1}{\sqrt{N_n}} \right),$$

and in particular for $N_n = n^{(2p/d)+(2\cdot p/d^2)+(2\cdot p^2/d^2)}$ we get

$$\left| \hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right| = O_{\mathbf{P}} \left(\frac{(\log n)^{5+2\cdot r}}{n^{(p/d)+(p/d^2)+(p^2/d^2)}} \right).$$

Remark 5. Let $m(x) = h(\|x\|)$ for some continuously differentiable function $h : \mathbb{R}_+ \to \mathbb{R}$, for which the derivative is less than zero on $\mathbb{R}_+$. Then the inverse function h^{-1} of h exists. Let $t = h^{-1}(q_{m(X),\alpha})$. Since we have

$$|m(x) - q_{m(X),\alpha}| \leq \epsilon \quad \Leftrightarrow \quad |h(\|x\|) - h(t)| \leq \epsilon \quad \Rightarrow \quad |\|x\| - t| \leq c_6 \cdot \epsilon$$

for some $c_6 > 0$, then set (27) is contained in

$$\left\{x \in \mathbb{R}^d \quad : \quad t - c_6 \cdot \epsilon \leq \|x\| \leq t + c_6 \cdot \epsilon \right\}.$$

Let l be the maximal number of disjoint (open) balls of radius $h/2$ that are contained in that set and let c_7 be the volume of the unit ball in $\mathbb{R}^d$. Then we have

$$l \cdot c_7 \cdot \left(\frac{h}{2}\right)^d \leq c_7 \cdot (t + c_6 \cdot \epsilon)^d - c_7 \cdot (t - c_6 \cdot \epsilon)^d,$$

from which we can conclude

$$l \leq \frac{(t + c_6 \cdot \epsilon)^d - (t - c_6 \cdot \epsilon)^d}{(h/2)^d} \leq c_8 \cdot \frac{\epsilon}{h^d}.$$

This implies that function m satisfies the assumption in Theorem 3 b) (cf., e.g., Lemma 9.2 in Györfi et al. (2002)).

Remark 6. **a)** It is possible to modify the definition of the estimate such that it achieves the rate of convergence $n^{-(p/d)-(p/d^2)-(p^2/d^2)}$ (up to some logarithmic factor) also when m is (p, C)–smooth for some $1 < p < d$. To do this, we choose $M \in \mathbb{N}$ and define for a cube

$$A = [x^{(1)}, x^{(1)} + h] \times \cdots \times [x^{(d)}, x^{(d)} + h] \subseteq \mathbb{R}^d$$

an operator Q_A as follows: In order to compute a polynomial $Q_A f$ corresponding to a function $f : \mathbb{R}^d \to \mathbb{R}$, Q_A uses function evaluations of f at $(M + 1)^d$ points

$$\left(x^{(1)} + \frac{j_1}{M} \cdot h, \ldots, x^{(d)} + \frac{j_d}{M} \cdot h,\right) \quad (j_1, \ldots, j_d \in \{0, 1, \ldots, M\}) \tag{28}$$

to construct a polynomial

$$(Q_A f)(x) = \sum_{j_1=0}^{M} \cdots \sum_{j_d=0}^{M} a_{j_1,\ldots,j_d} \cdot (x^{(1)})^{j_1} \cdots (x^{(d)})^{j_d}$$

satisfying

$$\|f - Q_A f\|_{\infty,A} \leq c_9 \cdot h^p$$

in case that f is (p, C)–smooth for some $0 < p \leq M + 1$. Such a polynomial can be constructed, e.g., similarly as the spline interpoland in Section 3, or by interpolating f at the points (28).

Then we define the partition $\{A_0, A_1, \ldots, A_{\lceil (n-1)/2^d \rceil}\}$ as above and define m_n by

$$m_n(x) = (Q_{A_n(x)} m)(x).$$

As in the proof of Theorem 3 it is possible to show that the corresponding Monte Carlo surrogate quantile estimate achieves under the assumptions of Theorem 3 b) the rate of convergence

$$\left|\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha}\right| = O_{\mathbf{P}}\left(\frac{(\log n)^{(5+2\cdot r)\cdot p}}{n^{(p/d)+(p/d^2)+(p^2/d^2)}}\right)$$

if m is (p, C)–smooth for some $0 < p \leq \min\{d, M + 1\}$.

b) In case of $(p/d) + (p/d^2) + (p^2/d^2) < 1/2$ the rate of convergence of the estimate in Theorem 3 or in Remark 6 a) is worse than the rate of convergence $n^{-1/2}$ of the Monte Carlo estimate in (3). But we can improve the rate in case that $d > 2p$ as follows: in the first step of the definition of the estimate we estimate the quantile $q_{m(X),\alpha}$ by $\hat{q}^{(MC)}_{m(X),n_1,\alpha}$, and replace the term $\frac{(\log n)^{r\cdot p+1}}{n^{p/d}}$ in (25) by the term $\frac{(\log n)}{\sqrt{n}}$. In this case it follows as in the proof of Theorem 3 that the resulting estimate satisfies under the assumptions of Theorem 3 b)

$$\left|\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha}\right| = O_{\mathbf{P}}\left(\frac{(\log n)^{p+r\cdot p+3\cdot p/d}}{n^{(p/d)+(p/(2\cdot d))+(p/d^2)}}\right).$$

However, for large d this rate of convergence may be still worse than the rate of convergence of the simple order statistic provided p is not large enough (cf., Remark 7 a) below)).

Remark 7. **a)** The rates of convergence in Theorems 2 and 3 get worse whenever d gets larger, which is a consequence of the well-known curse of dimensionality in nonparametric regression, which states that estimation of d-dimensional regression function gets harder and harder as d increases.

b) The proof of Theorem 1 utilizes order statistics and their error bound (cf., (3)) to derive the result. Since this error bound relies on the fact that α is fixed (i.e., it is an error bound on a central order statistic), it is not obvious how to generalize Theorem 1 towards the case that α depends on n and gets close to zero or one for n tending to infinity. The authors intend to study their estimate in this setting in a forthcoming paper.

c) The minimum number of evaluations of m that should be used to apply our newly proposed surrogate quantile estimates depends on function m and dimension d. In our simulations in the next section we study the finite sample size behaviour of our estimates in several examples.

5 Application to simulated data

In this section we consider the finite sample size behaviour of five different quantile estimates.

The first quantile estimate (*est. 1*) is the estimate based on order statistics and defined by (1) and (2).

All remaining estimates are based on non-adaptively or on adaptively chosen surrogate. For the non-adaptive surrogate we use a smoothing spline (as implemented in the routine *Tps()* in *R* with smoothing parameter chosen by 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-interpoland in Section 3, but is easier to implement.

The second quantile estimate (*est. 2*) uses the non-adaptive surrogate as a control variate as explained in Section 2 in Cannamela, Garnier and Ioss (2008). Here the true quantile of the control variate is replaced by a Monte Carlo estimate of the quantile using $N_n \in \{50000, 10000\}$ evaluations of the surrogate on randomly generated values of X.

The third quantile estimate (*est. 3*) uses the non-adaptive surrogate for controlled stratification with the three strata $[0, 0.85]$, $(0.85, 0.95]$ and $(0.95, 1]$ as explained in Subsection 3.2 in Cannamela, Garnier and Ioss (2008). Here $\lceil n/3 \rceil$ evaluations of m are used for the estimation of the surrogate, and the remaining $n - \lceil n/3 \rceil$ evaluations are split approximately equally in the three different strata.

The fourth quantile estimate (*est. 4*) is the non-adaptive Monte Carlo surrogate quantile estimate of Section 2 with the above smoothing spline surrogate instead of a quasi-interpoland of Section 2. The pseudocode in Algorithm 1 shows how our proposed method can be implemented.

```
Compute a sample X_1, ..., X_n of X;
Compute m(X_1), ..., m(X_n);
Use (X_1, m(X_1)), ..., (X_n, m(X_n)) to compute a thin plate spline estimate m_n;
Compute a second sample X_{n+1}, ..., X_{n+N_n} of X;
Compute m_n(X_{n+1}), ..., m_n(X_{n+N_n});
set result to the ⌈N_n · α⌉-th biggest of these values;
return result;
```

Algorithm 1: Proposed non-adaptive surrogate quantile estimate using at most n evaluations of function m.

For our fifth estimate (*est. 5*) we use $p = 1$ in our definition of the adaptive quantile estimate in Section 4 and start with an equidistant partition of $[-5, 5]^d$, where we ignore all the cells which contain none of $X_{n+1}, \ldots, X_{n+N_n}$. Furthermore we replace the factor $\frac{(\log n)^{r \cdot p+1}}{n^{p/d}}$ by $\log(n)/\sqrt{n}$, and use for $d > 2$ order statistics with sample size $\lceil n/3 \rceil$ as initial quantile estimate (and construct the partitioning estimate by using only $n - \lceil n/3 \rceil$ evaluations of m). The pseudocode in Algorithm 2 shows how our proposed method can be implemented in case $d \leq 2$ (for $d > 2$ the initial partitioning surrogate quantile estimate will be replaced by order statistics of $m(X_1), \ldots, m(X_{\lceil n/3 \rceil})$).

We compare these three quantile estimates in four different models, where in each model we estimate a quantile of level $\alpha = 0.95$. In each model X is chosen as a d-dimensional random variable with standard normal distribution, where in case $d > 1$ all d components of X are independent random variables with standard normal distribution. In the first and in the second model the dimension of X is $d = 1$, we allow $n \in \{20, 200, 2000\}$ evaluations of m, and all quantile estimates except the first one are based on $N_n = 50,000$ and (in case of $n = 2000$) $N_n = 100,000$ additionally observed values of X. In the third model the dimension of X is $d = 2$ and we choose $n \in \{30, 300, 3000\}$ and set $N_n = 50,000$ (for $n < 3000$) and $N_n = 100,000$ (for $n = 3000$) In the fourth model the dimension of X is $d = 4$ and we choose $n \in \{50, 300, 3000\}$ and set $N_n = 50,000$ (for $n < 3000$) and $N_n = 100,000$ (for $n = 3000$)

Since the result of our simulation depends on the randomly occurring data points, we repeat the whole procedure 100 times with independent realizations of the occurring random variables and report boxplots of the errors of the quantile estimates (more precisely, of the absolute values of the difference between the quantile estimates and the real quantile).

Generate sample $X_{n+1}, \dots, X_{n+N_n}$ of X;

Set $C_n := [-5,5]^d$ and $n_1 = \lceil \frac{n}{2} \rceil$;

Partition C_n into $\lfloor n_1^{1/d} \rfloor^d$ equivolume cubes;

Evaluate m at the centers of the $\lfloor n_1^{1/d} \rfloor^d$ sets A_j of this partition;

Let m_{n_1} be the corresponding partitioning estimate of m;

Compute $m_{n_1}(X_{n+1}), \dots, m_{n_1}(X_{n+N_n})$;

Let $\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha}$ be the $\lceil N_n \cdot \alpha \rceil$-th biggest of these values;

Let $\mathcal{P}_n$ be the initial partition of C_n defined above;

Set nv (number of evaluations of m) equal to $\lfloor n_1^{1/d} \rfloor^d$;

while *nv is less than $n - 2^d + 1$* **do**

> Choose set $A_j \in \mathcal{P}_n$ for which there exists some $X_{n+i} \in A_j$ such that (25) is maximal ;
>
> Subdivide A_j into 2^d equivolume sets;
>
> Replace A_j in $\mathcal{P}_n$ by the above 2^d sets;
>
> Evaluate m at the centers of the above 2^d sets;
>
> Compute the partitioning estimate m_n corresponding to $\mathcal{P}_n$;
>
> Increase nv by 2^d;

end

Compute $m_n(X_{n+1}), \dots, m_n(X_{n+N_n})$;

Let $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$ be the $\lceil N_n \cdot \alpha \rceil$-th biggest of these values;

return $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$;

Algorithm 2: Proposed adaptive surrogate quantile estimate using at most n evaluations of function m.

For the first model we choose

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

hence $m(X)$ has lognormal distribution. The boxplots of the errors of five different estimates occurring in 100 simulations for each sample size are presented in Figure 1.

In the second model we modify m in such a way that a good local approximation in an area which is important for the computation of the quantile improves the computation of the surrogate quantile estimate. To do this, we set

$$m(x) = \begin{cases} \exp(x) & , x \leq u_{0.95}, \\ \exp(u_{0.95}) & , u_{0.95} < x \leq 1.9, \\ \exp(x - 1.9 + u_{0.95}) & , \text{elsewhere}, \end{cases}$$

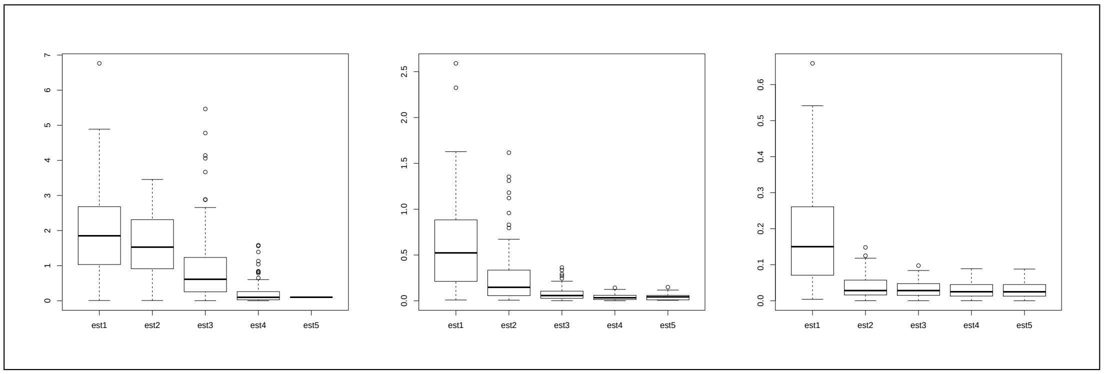

Figure 1: Boxplots of the estimation errors in model 1 for the three different sample sizes. In the left panel we have $n = 20$ and $N_n = 50,0000$, in the middle panel $n = 200$ and $N_n = 50,000$, and in the right panel $n = 2,000$ and $N_n = 100,000$.

where $u_{0.95} \approx 1.645$ is the 0.95-quantile of the standard normal distribution. The sample sizes are chosen as before, and the errors in 100 simulation for each pair of sample sizes are presented in the boxplots in Figure 2.

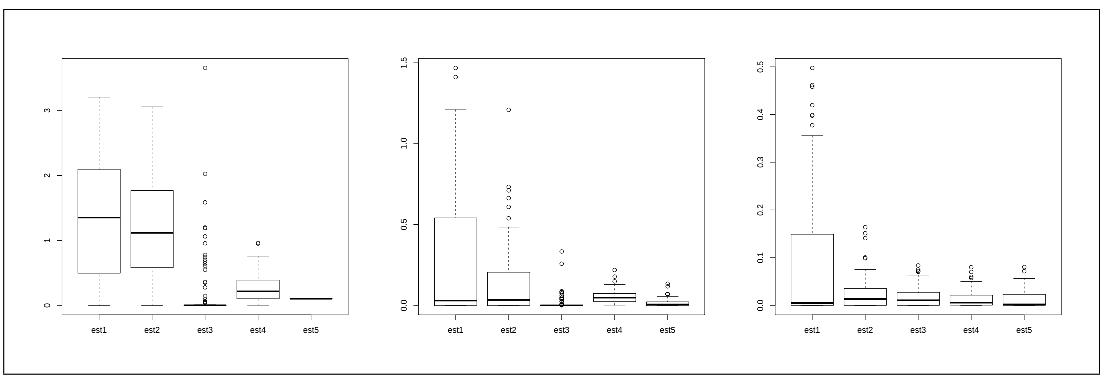

Figure 2: Boxplots of the estimation errors in model 2 for the three different sample sizes. In the left panel we have $n = 20$ and $N_n = 50,0000$, in the middle panel $n = 200$ and $N_n = 50,000$, and in the right panel $n = 2,000$ and $N_n = 100,000$.

In the third model we set $d = 2$, and we define

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

hence $m(X)$ is a monotonically increasing function of random variable which has chi-square distribution with two degrees of freedom. The sample sizes are chosen as $n = 30$ and $N_n = 50,000$, $n = 300$ and $N_n = 50,000$, and $n = 3000$ and $N_n = 100,000$, resp., and the errors in 100 simulations for each pair of sample sizes are presented in the boxplots in Figure 3.

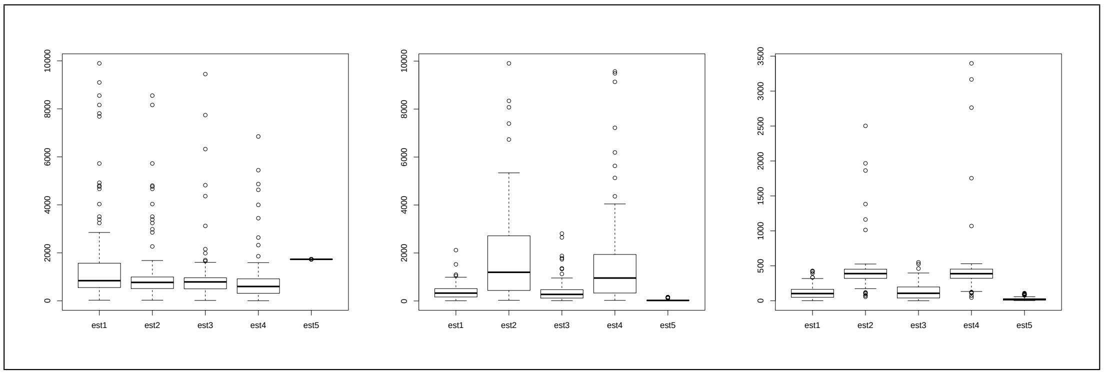

Figure 3: Boxplots of the estimation errors in model 3 for the three different sample sizes. In the left panel we have $n = 30$ and $N_n = 50,0000$, in the middle panel $n = 300$ and $N_n = 50,000$, and in the right panel $n = 3,000$ and $N_n = 100,000$.

In our fourth model we set $d = 4$ and use again a function which is constant in an area which is important for the computation of the quantile. Consequently here a good local approximation of the function is especially useful. We set

$$m(x) = \begin{cases} \sqrt{1 + \|x\|^2} & , \|x\|^2 \leq \chi^2_{0.95,4}, \\ \sqrt{1 + \chi^2_{0.95,4}} & , \chi^2_{0.95,4} \leq \|x\|^2 \leq \chi^2_{0.95,4} + 1.5, \\ \sqrt{1 + \|x\|^2 - 1.5} & , \text{elsewhere.} \end{cases}$$

The sample sizes are chosen as $n = 50$ and $N_n = 50,000$, $n = 300$ and $N_n = 50,000$, and $n = 3000$ and $N_n = 100,000$, resp., and the errors in 100 simulations for each pair of sample sizes are presented in the boxplots in Figure 4.

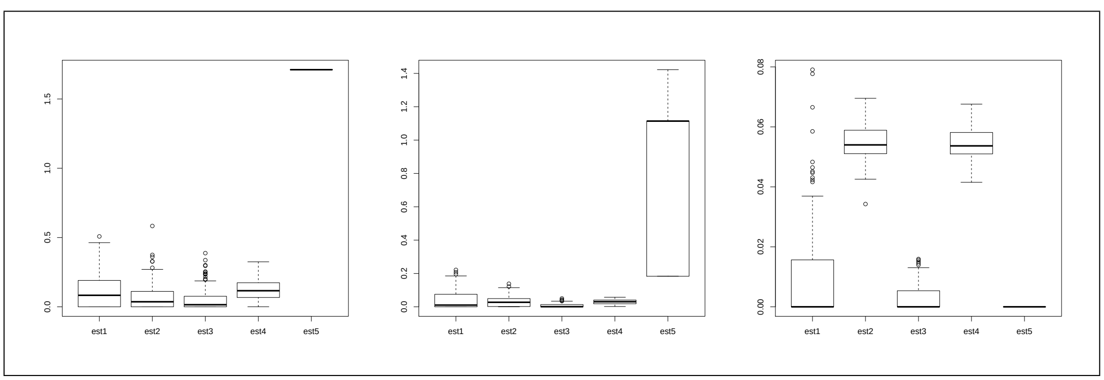

Figure 4: Boxplots of the estimation errors in model 4 for the three different sample sizes. In the left panel we have $n = 50$ and $N_n = 50,0000$, in the middle panel $n = 300$ and $N_n = 50,000$, and in the right panel $n = 3,000$ and $N_n = 100,000$.

In all four models our newly proposed adaptive partitioning Monte Carlo quantile estimate (*est. 5*) outperforms for large sample sizes all other four quantile estimates, although sometimes the control variate quantile estimate based on the non-adaptively chosen smoothing spline surrogate (*est. 3*) is better for a small sample size (e.g., for $n \in \{20, 200\}$ in model 2, for $n = 30$ in model 3 and for $n \in \{50, 300\}$ in model 4).

Finally we illustrate the usefulness of our newly proposed estimate by applying it to a simulation model in engineering. Here we consider a physical model of a spring-mass-damper with active velocity feedback for the purpose of vibration isolation (cf., Figure 5). The aim is to analyze

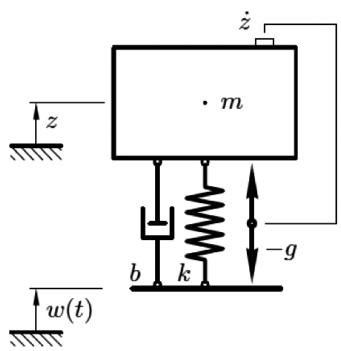

Figure 5: Spring-mass-damper with active velocity feedback (Platz and Enss (2015)).

the uncertainty occuring in the maximal 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 2,000 function evaluations can be easily completed in approximately seven minutes, but computation of 100,000 values requires about 5.5 hours.

In the following we distinguish between two cases: firstly the passive case, where the active velocity feedback g equals zero and secondly the active case, where the value of g is given by the normally distributed random variable with mean 45 Ns/m and a standard deviation of 2.25 Ns/m. In both cases in our simulation the remaining variables are also normally distributed, but their means and standard deviations are different. The means of m, k and b are 1 kg, 1000 N/m and 0.095 Ns/m, respectively and their standard deviations are 0.017 kg, 33.334 N/m and 0.009 Ns/m, respectively.

In the active case we simulate the value of $x = (m, k, b, g)$ with independent random variables, as defined above, and use order statistics with sample size $n = 100,000$ to compute a reference value of the $\alpha = 0.95$ quantile of the maximal magnification of the vibration amplitude. This yields $|V_{max}| = 0.10217\ dB$. But if we want to estimate this value using only $n = 2,000$ evaluations of our function, we get with order statistics, the surrogate quantile estimate and the adaptive partitioning quantiles estimates the values $0.096206\ dB$, $0.102315\ dB$ and $0.10119\ dB$, resp. As we can see, both the value of the surrogate quantile estimate and the adaptive partitioning estimate are much closer to our reference value than the result produced by order statistics with sample size $2,000$.

In the passive case we simulate the value of $x = (m, k, b, g = 0)$ as explained above. As before we use order statistics with sample size $n = 100,000$ and get the estimate $51.92\ dB$ as a reference value. We again compare the reference value with the value we get with sample size $n = 2,000$ in case of the order statistics ($51.916\ dB$), the surrogate quantile estimate ($51.91965\ dB$) and the adaptive partitioning quantile estimate ($51.922\ dB$). Again the last two estimates are closer to the reference value than the order statistics estimate.

6 Proofs

6.1 Proof of Theorem 1

In the proof we will apply the following two lemmata.

Lemma 1 *Let μ be an arbitrary probability measure on $\mathbb{R}^d$, let $m, \bar{m} : \mathbb{R}^d \to \mathbb{R}$ be measurable functions and let $\alpha \in (0,1)$. Set*

$$q_{m,\mu,\alpha} = \inf\left\{y \in \mathbb{R} : \mu(\{x \in \mathbb{R}^d : m(x) \leq y\}) \geq \alpha\right\})$$

and

$$q_{\bar{m},\mu,\alpha} = \inf\left\{y \in \mathbb{R} : \mu(\{x \in \mathbb{R}^d : \bar{m}(x) \leq y\}) \geq \alpha\right\}.$$

Let $\delta > 0$ and assume that m and $\bar{m}$ satisfy for μ-almost all $x \in \mathbb{R}^d$

$$|\bar{m}(x) - m(x)| \leq \frac{\delta}{2} + \frac{1}{2} \cdot |q_{m,\mu,\alpha} - m(x)|. \tag{29}$$

Then

$$|q_{\bar{m},\mu,\alpha} - q_{m,\mu,\alpha}| \leq \delta.$$

Proof. *In the first step of the proof* we show that the assertion follows from

$$\mu\left(\left\{x \in \mathbb{R}^d : \bar{m}(x) \leq q_{m,\mu,\alpha} + \delta\right\}\right) \geq \alpha \tag{30}$$

and

$$\mu\left(\left\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le q_{m,\mu,\alpha} - \delta - \epsilon\right\}\right) < \alpha \quad \text{for all } 0 < \epsilon < \delta. \tag{31}$$

To do this, assume that (30) and (31) hold. Then (30) implies

$$q_{\bar{m},\mu,\alpha} = \inf\left\{y \in \mathbb{R} \,:\, \mu\left(\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le y\} \ge \alpha\right\}\right) \le q_{m,\mu,\alpha} + \delta,$$

and since

$$\mu\left(\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le y_1\}\right) \le \mu\left(\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le y_2\}\right) \quad \text{for all } y_1 \le y_2$$

we can conclude from (31) that

$$\left\{y \in \mathbb{R} \,:\, \mu\left(\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le y\} \ge \alpha\right\}\right) \subseteq [q_{m,\mu,\alpha} - \delta, \infty),$$

which implies

$$q_{\bar{m},\mu,\alpha} = \inf\left\{y \in \mathbb{R} \,:\, \mu\left(\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le y\} \ge \alpha\right\}\right) \ge q_{m,\mu,\alpha} - \delta.$$

In the second step of the proof we show (30). Here it suffices to show

$$m(x) \le q_{m,\mu,\alpha} \quad \Longrightarrow \quad \bar{m}(x) \le q_{m,\mu,\alpha} + \delta \quad \text{for } \mu\text{–almost all } x \in \mathbb{R}^d, \tag{32}$$

because from (32) and the definition of $q_{m,\mu,\alpha}$ we get

$$\mu\left(\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le q_{m,\mu,\alpha} + \delta\}\right) \ge \mu\left(\{x \in \mathbb{R}^d \,:\, m(x) \le q_{m(X),\alpha}\}\right) \ge \alpha.$$

In order to show (32), let $x \in \mathbb{R}^d$ be such that (29) holds and assume $m(x) \le q_{m,\mu,\alpha}$. Then we get by (29)

$$\begin{aligned}
\bar{m}(x) &\le m(x) + |\bar{m}(x) - m(x)| \\
&\le m(x) + \frac{\delta}{2} + \frac{1}{2} \cdot |q_{m,\mu,\alpha} - m(x)| \\
&= m(x) + \frac{\delta}{2} + \frac{1}{2} \cdot (q_{m,\mu,\alpha} - m(x)) \\
&= \frac{1}{2} \cdot m(x) + \frac{1}{2} \cdot q_{m,\mu,\alpha} + \frac{\delta}{2} \\
&\le q_{m,\mu,\alpha} + \frac{\delta}{2},
\end{aligned}$$

where we have used $m(x) \le q_{m,\mu,\alpha}$ in the last inequality.

In the third step of the proof we show (31). To do this we will show

$$\bar{m}(x) \le q_{m,\mu,\alpha} - \delta - \epsilon \quad \Longrightarrow \quad m(x) \le q_{m,\mu,\alpha} - \epsilon \tag{33}$$

for μ-almost all $x \in \mathbb{R}^d$ and all $0 < \epsilon < \delta$, which implies (31), because if (33) holds we can conclude from the definition of $q_{m,\mu,\alpha}$ that

$$\mu\left(\{x \in \mathbb{R}^d \,:\, \bar{m}(x) \le q_{m,\mu,\alpha} - \delta - \epsilon\}\right) \le \mu\left(\{x \in \mathbb{R}^d \,:\, m(x) \le q_{m,\mu,\alpha} - \epsilon\}\right) < \alpha$$

for all $0 < \epsilon < \delta$.

Implication (33) is equivalent to

$$m(x) > q_{m,\mu,\alpha} - \epsilon \quad \Longrightarrow \quad \bar{m}(x) > q_{m,\mu,\alpha} - \delta - \epsilon \tag{34}$$

for μ-almost all $x \in \mathbb{R}^d$ and all $0 < \epsilon < \delta$.

In order to prove (34), let $x \in \mathbb{R}^d$ be such that (29) holds and let $0 < \epsilon < \delta$. Assume furthermore that $m(x) > q_{m,\mu,\alpha} - \epsilon$. Then we can conclude from (29) that

$$\begin{aligned}
\bar{m}(x) &\geq m(x) - |\bar{m}(x) - m(x)| \\
&\geq m(x) - \frac{\delta}{2} - \frac{1}{2} \cdot |q_{m,\mu,\alpha} - m(x)| \\
&= m(x) - \frac{\delta}{2} - \frac{1}{2} \cdot |(m(x) - q_{m,\mu,\alpha} + \epsilon) - \epsilon| \\
&\geq m(x) - \frac{\delta}{2} - \frac{1}{2} \cdot (m(x) - q_{m,\mu,\alpha} + \epsilon) - \frac{1}{2} \cdot \epsilon \\
&= \frac{1}{2} \cdot m(x) + \frac{1}{2} \cdot q_{m,\mu,\alpha} - \frac{\delta}{2} - \epsilon \\
&> \frac{1}{2} \cdot (q_{m,\mu,\alpha} - \epsilon) + \frac{1}{2} \cdot q_{m,\mu,\alpha} - \frac{\delta}{2} - \epsilon \\
&= q_{m,\mu,\alpha} - \frac{1}{2} \cdot \delta - \frac{3}{2} \cdot \epsilon \\
&> q_{m,\mu,\alpha} - \epsilon - \delta,
\end{aligned}$$

where the last inequality follows from $0 < \epsilon < \delta$. □

Lemma 2 *Let X be an $\mathbb{R}^d$-valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and let $\alpha \in (0,1)$. Define the Monte Carlo surrogate quantile estimate $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$ of $q_{m(X),\alpha}$ as in Section 2 and let $\hat{q}^{(MC)}_{m(X),N_n,\alpha}$ be the Monte Carlo quantile estimate of $q_{m(X),\alpha}$ based on $m(X_{n+1})$, ..., $m(X_{n+N_n})$, i.e.,*

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

where

$$\hat{G}^{(MC)}_{m(X),N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m(X_{n+i}) \leq y\}}.$$

For $n \in \mathbb{N}$ let $\delta_n > 0$ be such that the estimate m_n satisfies

$$|m_n(X_{n+i}) - m(X_{n+i})| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m(X),\alpha} - m(X_{n+i})| \quad \textit{for all } i \in \{1, \ldots N_n\}. \tag{35}$$

Then we have

$$\left| \hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right| \leq \delta_n + 2 \cdot \left| \hat{q}^{(MC)}_{m(X),N_n,\alpha} - q_{m(X),\alpha} \right|.$$

Proof. The key trick in the proof of Lemma 2 is to apply Lemma 1 with μ equal to the empirical distribution $\hat{\mathbf{P}}_X$ of $X_{n+1}, \ldots, X_{n+N_n}$. To do this, we observe first that (35) together with triangle inequality implies

$$|m_n(x) - m(x)| \leq \frac{\delta_n + |q_{m(X),\alpha} - \hat{q}^{(MC)}_{m(X),N_n,\alpha}|}{2} + \frac{1}{2} \cdot |\hat{q}^{(MC)}_{m(X),N_n,\alpha} - m(x)|$$

for $\hat{\mathbf{P}}_X$-almost all $x \in \mathbb{R}^d$. Since $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} = q_{m_n,\hat{\mathbf{P}}_X,\alpha}$ and $\hat{q}^{(MC)}_{m(X),N_n,\alpha} = q_{m,\hat{\mathbf{P}}_X,\alpha}$, an application of Lemma 1 yields

$$\left|\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - \hat{q}^{(MC)}_{m(X),N_n,\alpha}\right| \leq \delta_n + |q_{m(X),\alpha} - \hat{q}^{(MC)}_{m(X),N_n,\alpha}|.$$

By the triangle inequality we get the assertion. □

Proof of Theorem 1. Set $K_n = [-\beta_n, \beta_n]^d$ and let A_n be the event that $X_1, \ldots, X_{N_n}$ are all contained in K_n. By (5) 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 with probability one for any $i \in \{1, \ldots, N_n\}$

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

from which we conclude by Lemma 2

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

This implies

$$\begin{aligned}
&\lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\left\{\left|\hat{q}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha}\right| > c \cdot \left(\delta_n + \frac{1}{\sqrt{N_n}}\right)\right\} \\
&\leq \lim_{c\to\infty} \limsup_{n\to\infty} \left(\mathbf{P}(A_n^c) + \mathbf{P}\left\{\delta_n + 2 \cdot |\hat{q}_{m(X),N_n,\alpha} - q_{m(X),\alpha}|\right.\right. \\
&\qquad\qquad\qquad\qquad \left.\left. > c \cdot \left(\delta_n + \frac{1}{\sqrt{N_n}}\right)\right\}\right) \\
&\leq \lim_{c\to\infty} \limsup_{n\to\infty} \mathbf{P}\left\{2 \cdot |\hat{q}_{m(X),N_n,\alpha} - q_{m(X),\alpha}| > c \cdot \frac{1}{\sqrt{N_n}}\right\} = 0,
\end{aligned}$$

where the last equality follows from (3). □

6.2 Proof of Theorem 3

a) Since m is (p, C)–smooth we have for any $x \in C_n$

$$\begin{aligned}
|m_n(x) - m(x)| &= |m(x_{A_n(x)}) - m(x)| \leq C \cdot \|x_{A_n(x)} - x\|^p \leq C \cdot \mathrm{diam}(A_n(x))^p \\
&\leq C \cdot \mathrm{diam}(A_{n_1}(x))^p \leq C \cdot \left(\sqrt{d} \cdot \frac{2 \cdot (\log n)^r}{\lfloor n_1^{1/d} \rfloor}\right)^p.
\end{aligned}$$

Hence condition (4) holds with

$$\delta_n = \frac{(\log n)^{r \cdot p+1}}{n^{p/d}}$$

for n sufficiently large. Application of Theorem 1 yields the assertion.

b) Set

$$\delta_n = \max_{i \in \{1,\dots,N_n\}, X_{n+i} \in C_n} \min \Bigg\{ 2 \cdot \log(n) \cdot \mathrm{diam}(A_n(X_{n+i}))^p, \\ 3 \cdot \log(n) \cdot \mathrm{diam}(A_n(X_{n+i}))^p + \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(X_{n+i})| \Bigg\}. \tag{36}$$

In the first step of the proof we show that the assertion follows from

$$\delta_n = O_{\mathbf{P}} \left((\log n)^{5+2 \cdot r} \cdot n^{-(p/d)-(p/d^2)-(p^2/d^2)} \right). \tag{37}$$

If δ_n is defined as above, then by the construction of our estimate (cf., Section 4) and the proof of part a) of Theorem 3 (which implies that (16) holds outside of an event whose probability tends to zero) we know that m_n satisfies (4) outside of an event whose probability tends to zero. As in the proof of Theorem 1 this yields the assertion of step 1.

Set

$$C_{critical,n} := \left\{ x \in \mathbb{R}^d \, : \, m(x) \in \left[q_{m(X),\alpha} - 6 \cdot \frac{(\log n)^{r \cdot p+2}}{n^{p/d}}, q_{m(X),\alpha} + 6 \cdot \frac{(\log n)^{r \cdot p+2}}{n^{p/d}} \right] \right\},$$

and let E_n be the event that $|\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - q_{m(X),\alpha}|$ is less than or equal to $(\log n)^{r \cdot p+1}/n^{p/d}$.

In the second step of the proof we show that we have on E_n for $x \in C_n \setminus C_{critical,n}$ and for n sufficiently large

$$3 \cdot \log(n) \cdot \mathrm{diam}(A_n(x))^p + \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| \le 0. \tag{38}$$

To do this we observe that triangle inequality and (p, C)–smoothness of m imply that we have on E_n and for n sufficiently large

$$\begin{aligned} |q_{m(X),\alpha} - m(x)| &\le \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} + |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| + C \cdot \mathrm{diam}(A_n(x))^p \\ &\le \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} + |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| + \frac{(\log n)^{r \cdot p+1}}{n^{p/d}}. \end{aligned}$$

This in turn implies for $x \in C_n \setminus C_{critical,n}$ and for n sufficiently large

$$\begin{aligned} &3 \cdot \log(n) \cdot \mathrm{diam}(A_n(x))^p + \frac{(\log n)^{r \cdot p+1}}{n} - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| \\ &\le 3 \cdot \frac{(\log n)^{r \cdot p+2}}{n^{p/d}} + 3 \cdot \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} - |q_{m(X),\alpha} - m(x)| \\ &\le 6 \cdot \frac{(\log n)^{r \cdot p+2}}{n^{p/d}} - 6 \cdot \frac{(\log n)^{r \cdot p+2}}{n^{p/d}} = 0. \end{aligned}$$

In the third step of the proof we show that we have outside of an event whose probability tends to zero for all $x \in C_{critical,n} \cap C_n$ we have

$$\min \Bigg\{ 2 \cdot \log(n) \cdot \operatorname{diam}(A_n(x))^p, \\ 3 \cdot \log(n) \cdot \operatorname{diam}(A_n(x))^p + \frac{(\log n)^{r \cdot p+1}}{n^{p/d}} - \left| \hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(X_{n+i} \right| \Bigg\} \\ \leq c_6 \cdot (\log n)^{5+2 \cdot r} \cdot n^{-(p/d)-(p/d^2)-(p^2/d^2)}. \tag{39}$$

By the result of step 2 we know that on E_n and for n sufficiently large (38) holds for all $x \in C_n \setminus C_{critical,n}$. Hence as long as as any cell of the partition of the partitioning estimate, which has nonempty intersection with $C_{critical,n} \cap C_n$, does not satisfy (39), our algorithm does not choose any cell from the partition which has empty intersection $C_{critical,n} \cap C_n$. By the assumption of part b) of Theorem 3 we know that for n sufficiently large $C_{critical,n} \cap C_n$ is contained in the union of at most

$$\log(n) \cdot \frac{(\log n)^{r \cdot p+2}/n^{p/d}}{(\log n)^{r \cdot (d-1)}/n^{1-1/d}} \leq (\log n)^{3+r \cdot p - r \cdot (d-1)} \cdot n^{1-1/d-p/d}$$

many of the cubes of side length $2 \cdot (\log n)^r / \lfloor n_1^{1/d} \rfloor$ in $\mathcal{P}_{n_1}$. By construction, our algorithm does not subdivide any cube which is not contained in any of theses cubes, as long as (39) is not satisfied. But after $n_2/(2^d+1)$ many cubes are chosen, which are contained in one of the above described cubes of $\mathcal{P}_{n_1}$, we have for n sufficiently large and all $x \in C_{critical,n} \cap C_n$

$$\begin{aligned} \operatorname{diam}(A_n(x)) &\leq \log(n) \cdot \frac{(\log n)^r}{n^{1/d}} \cdot \left(\frac{n}{(\log n)^{3+r \cdot p - r \cdot (d-1)} \cdot n^{1-1/d-p/d}} \right)^{-1/d} \\ &\leq (\log n)^{4+2 \cdot r} \cdot n^{-(1/d)-(1/d^2)-(p/d^2)}, \end{aligned}$$

which implies

$$2 \log n \cdot (\operatorname{diam}(A_n(x)))^p \leq 2 (\log n)^{5+2 \cdot r} \cdot n^{-(p/d)-(p/d^2)-(p^2/d^2)}.$$

This completes the third step of the proof.

The steps 2 and 3 of the proof imply the assertion of Theorem 3 b), because by the proof of part a) of Theorem 3 we have $\mathbf{P}(E_n) \to 1$ $(n \to \infty)$ and $\mathbf{P}\{X_{n+1}, \dots, X_{n+N_n} \in C_n\} \to 1$ $(n \to \infty)$. $\square$

7 Acknowledgment

The authors would like to thank two anonymous referees for various suggestions which helped to significantly improve an early version of this manuscript. The third author would like to acknowledge the support from the Natural Sciences and Engineering Research Council of Canada. The remaining authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 805.

References

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

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

[3] Bichon, B., Eldred, M., Swiler, M., Mahadevan, S. and McFarland, J. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, **46**, pp. 2459–2468.

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

[5] Bourinte, J.-M., Deheeger, F. and Lemaire, M. (2011). Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety*, **33**, pp. 343–353.

[6] Bucher, C. and Bourgund, U. (1990). A fast and efficient response surface approach for structural reliability problems. *Structural Safety*, **7**, pp. 57-66.

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

[8] Das, P.-K. and Zheng, Y. (2000). Cumulative formation of response surface and its use in reliability analysis. *Probabilistic Engineering Mechanics*, **15**, pp. 309-315.

[9] Deheeger, F. and Lemaire, M. (2010). Support vector machines for efficient subset simulations: ^{2}SMART method. In: *Proceedings of the 10th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP10)*, Tokyo, Japan.

[10] 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.

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

[12] 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.

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

[14] Dubourg, V., Sudret, B. and Deheeger, F. (2013). Metamodel-based importance sampling for syructural reliability analysis. *Probabilistic Engineering Mechanics*, **33**, pp. 47–57.

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

[16] 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.

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

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

[19] Hesterberg, T. C. and Nelson, B. L. (1998). Control variates for probability and quantile estimation. *Management Science*, **44**, pp. 1295–1312.

[20] Hurtado, J. (2004). *Structural reliability – Statistical learning perspectives.* Vol. 17 of lecture notes in applied and computational mechanics. Springer.

[21] Kaymaz, I. (2005). Application of Kriging method to structural reliability problems. *Strutural Safety*, **27**, pp. 133–151.

[22] Kim, S.-H. and Na, S.-W. (1997). Response surface method using vector projected sampling points. *Structural Safety*, **19**, pp. 3–19.

[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 , 2014.

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

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

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

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

[29] Morio, J., Balesdent, M., Jacquemart, D. and Verge, C. (2014). A survey of rare event simulation methods for static input-output models. *Simulation Modelling Practice and Theory*, **49**, pp. 287–304.

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

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

[32] Oakley, J. (2004) Estimating percentiles of uncertain computer code outputs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 53, no. 1, pp 83-93.

[33] Papadrakakis, M. and Lagaros, N. (2002). Reliability–based structural optimization using neural networks and Monte Carlo simulation. *Computer Methods in Applied Mechanics and Engineering*, **191**, pp. 3491–3507.

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

[35] 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.

[36] Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York.

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

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

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

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