Nonparametric quantile estimation based on surrogate models

Georg C. Enss¹, Michael Kohler², Adam Krzyżak³,∗ and Roland Platz⁴

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

December 5, 2014

Abstract
Nonparametric estimation of a quantile $q_{m(X),\alpha}$ of a random variable $m(X)$ is considered, where $m : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 \( P_X \) 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

\[ G(y) = P\{Y \leq y\} = P\{m(X) \leq y\} \]

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 the 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

\[ \hat{G}_{m(X), n}(y) = \frac{1}{n} \sum_{i=1}^{n} I\{m(X_i) \leq y\} \]

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

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

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

\[ |\hat{q}_{m(X), n, \alpha} - q_{m(X), \alpha}| = O_P \left( \frac{1}{\sqrt{n}} \right), \]  

where we write \( X_n = O_P(Y_n) \) if the nonnegative random variables \( X_n \) and \( Y_n \) satisfy

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

In this paper we construct estimates which achieve (under suitable assumptions) better rates of convergence. 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

\[ q_{m_n(X), \alpha} = \inf\{y \in \mathbb{R} : P_X\{x \in \mathbb{R}^d : m_n(x) \leq y\} \geq \alpha\}. \]
Our main result analyzes 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 \( \delta_n > 0 \)

\[
|m_n(x) - m(x)| \leq \frac{\delta_n}{2} + \frac{1}{2} |m(x) - q_{m(X),\alpha}| \quad \text{for } P_X\text{-almost all } x,
\]

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

\[
\left| q_{m_n(X),N_n,\alpha}^{(MC)} - q_{m(X),\alpha} \right| = O_P \left( \delta_n + \frac{1}{\sqrt{N_n}} \right),
\]

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 \( n^{-2p/d} \).

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., Győrfi (1981) or Beirlant and Győrfi (1998)), nearest neighbor regression estimate (cf., e.g., Devroye (1982) or Devroye, Győrfi, 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: \( \mathbb{N}, \mathbb{N}_0, \mathbb{Z} \) and \( \mathbb{R} \) are the sets of positive integers, nonnegative integers, integers and real numbers, respectively. For a real number \( z \) we denote by \( \lfloor z \rfloor \) and \( \lceil z \rceil \) the largest integer less than or equal to \( z \) and the smallest integer larger than or equal to \( z \), respectively. \( \| x \| \) is the Euclidean norm of \( x \in \mathbb{R}^d \), and the diameter of a set \( A \subseteq \mathbb{R}^d \) is denoted by

\[
\text{diam}(A) = \sup \{ \| x - z \| : x, z \in A \}.
\]

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 \( \frac{\partial^k m}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \) exists and satisfies

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

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

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

\[
\lim_{c \to \infty} \limsup_{n \to \infty} \mathbb{P}(X_n > c \cdot Y_n) = 0.
\]

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 \( X, X_1, X_2, \ldots \) 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

\[(x_1, m(x_1)), \ldots, (x_n, m(x_n))\]

is used to construct an estimate

\[m_n(\cdot) = m_n(\cdot, (x_1, m(x_1)), \ldots, (x_n, m(x_n))) : \mathbb{R}^d \to \mathbb{R}\]
of \( m \). Here \( x_i = X_i \) is one possible choice for the values of \( x_1, \ldots, x_n \in \mathbb{R}^d \), but not the only one (see Sections 3 and 4 below). Then \( X_{n+1}, \ldots, X_{n+N_n} \) are used to define a Monte Carlo estimate of the \( \alpha \)-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 \( m : \mathbb{R}^d \rightarrow \mathbb{R} \) 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 and let \( q^{(MC)}_{m(X), N_n, \alpha} \) be the Monte Carlo quantile estimate of \( q_m(X, \alpha) \) based on \( m(X_{n+1}), \ldots, m(X_{n+N_n}) \), 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

\[
|m_n(X_{n+i}) - m(X_{n+i})| \leq \frac{\delta_n}{2} + \frac{1}{2} |q_m(X, \alpha) - m(X_{n+i})| \quad \text{for all } i \in \{1, \ldots, N_n\}.
\]

(4)

Then we have

\[
|\hat{q}^{(MC)}_{m_n(X), N_n, \alpha} - q_m(X, \alpha)| \leq \delta_n + 2 \cdot |\hat{q}^{(MC)}_{m(X), N_n, \alpha} - q_m(X, \alpha)|.
\]

We immediately conclude from (3)

**Corollary 1** Let \( X \) be an \( \mathbb{R}^d \)-valued random variable, let \( m : \mathbb{R}^d \rightarrow \mathbb{R} \) be a measurable function and let \( \alpha \in (0, 1) \). Assume that \( m(X) \) has a density which is continuous and positive at \( q_m(X, \alpha) \). 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 \( \delta_n > 0 \) be such that the estimate \( m_n \) satisfies (4) with probability one. Then

\[
|\hat{q}^{(MC)}_{m_n(X), N_n, \alpha} - q_m(X, \alpha)| = O_P \left( \delta_n + \frac{1}{\sqrt{N_n}} \right)
\]

**Proof.** By (3) we know

\[
|\hat{q}^{(MC)}_{m(X), N_n, \alpha} - q_m(X, \alpha)| = O_P \left( \frac{1}{\sqrt{N_n}} \right).
\]

This and Theorem 1 yield the assertion. \( \square \)
Remark 1. If
\[ |m_n(x) - m(x)| \leq \frac{\delta_n}{2} + \frac{1}{2} |m(x) - q_{m(X),\alpha}| \quad \text{for } P_X\text{-almost all } x, \tag{5} \]
then (4) holds with probability one.

Remark 2. Condition (4) is in particular satisfied if we choose
\[ \delta_n = 2 \cdot \|m_n - m\|_{\infty, \text{supp}(P_X)}, \]
so Corollary 1 implies
\[ \left| \hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right| = O_P \left( \|m_n - m\|_{\infty, \text{supp}(P_X)} + \frac{1}{\sqrt{N_n}} \right). \tag{6} \]
However, in general we can derive from Corollary 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., (5)).

Remark 3. If the support of \( X \) is unbounded, it might be difficult to construct estimates for which the error is uniformly small on the whole support as requested in Remark 2. But under suitable assumptions on the tails of \( X \) it suffices to approximate \( m \) on a compact set. Indeed, let \( \beta_n > 0 \) be such that
\[ N_n \cdot P\{X \notin [-\beta_n, \beta_n]^d\} \to 0 \quad (n \to \infty). \]
As in the proof of Theorem 2 below it is possible to conclude from Theorem 1 that in this case
\[ \left| \hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right| = O_P \left( \|m_n - m\|_{\infty, \text{supp}(P_X)\cap [-A_n,A_n]^d} + \frac{1}{\sqrt{N_n}} \right). \]

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 \( \alpha > 0 \) and set \( \beta_n = \log(n)^\alpha \). 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 as follows:

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

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

And we define $S_{K,M}$ 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

$$S_{K,M} = \left\{ \sum_{k \in \{-K - M, -K - M + 1, \ldots, K - 1\}^d} a_k \cdot B_{k,M} : a_k \in \mathbb{R} \right\}.$$

It can be shown by using standard arguments from spline theory, that the functions in $S_{K,M}$ 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

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

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 $m : [-\beta_n, \beta_n]^d \to \mathbb{R}$ we define an approximating spline by

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

where

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

for some $a_{k,j} \in \mathbb{R}$ and for suitably chosen points $t_{k,j} \in supp(B_{k,M}) \cap [-\beta_n, \beta_n]$. 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 \( a_{k,j} \) (which can be computed by solving a linear equation system), such that
\[
|Q_k f| \leq c_1 \cdot \|f\|_{\infty, [u_{k_1}, u_{k_1+M+1}] \times \cdots \times [u_{k_d}, u_{k_d+M+1}]} \quad (8)
\]
for any \( k \in \{-M, -M + 1, \ldots, K - 1\}^d \), 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
\[
(Qp)(x) = p(x) \quad (x \in [-\beta_n, \beta_n]^d) \quad (9)
\]
(cf., e.g., Theorem 14.4 and Theorem 15.2 in Győrfi et al. (2002)).

Next we define our estimate \( m_n \) as a quasi interpoland. We fix the degree \( M \in \mathbb{N} \) and set
\[
K = K_n = \left\lfloor \frac{n^{1/d} - 1}{2 \cdot M} \right\rfloor.
\]
Furthermore we choose \( x_1, \ldots, x_n \) such that all of the \((2 \cdot M \cdot K + 1)^d \) points of the form
\[
\left( \frac{j_1}{M \cdot K} \cdot A_n, \ldots, \frac{j_d}{M \cdot K} \cdot A_n \right) \quad (j_1, \ldots, j_d \in \{-M \cdot K, -M \cdot K + 1, \ldots, M \cdot K\})
\]
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 (8) and (9). 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 \( q_\alpha \) and that \( m \) is \((p, C)\)-smooth for some \( p > 0 \) and some \( C > 0 \). Define the Monte Carlo surrogate quantile estimate \( q_{m_n(X), N_n, \alpha}^{MC} \) of \( q_{m(X), \alpha} \) as in Section 2, where \( m_n \) is the spline approximant defined above with parameter \( M \geq p - 1 \).

Assume furthermore that
\[
N_n \cdot \mathbb{P}\{X \notin [-\log(n)^\alpha, \log(n)^\alpha]^d\} \to 0 \quad (n \to \infty). \quad (10)
\]
Then
\[ \left| \hat{q}_{m_n(X),N_n,\alpha}^{(MC)} - q_m(X,\alpha) \right| = O_P \left( \frac{\log(n)^\alpha p}{n^{p/d}} + \frac{1}{\sqrt{N_n}} \right). \]

In particular, if we set \( N_n = \left\lceil \frac{n^{2p/d}}{\log(n)^2 \alpha} \right\rceil \) then we get
\[ \left| \hat{q}_{m_n(X),N_n,\alpha}^{(MC)} - q_m(X,\alpha) \right| = O_P \left( \frac{\log(n)^\alpha p}{n^{p/d}} \right). \]

**Remark 4.** It follows from Theorem 2 that in case that \( m \) be \((p,C)\)-smooth for some \( p > d/2 \) or some \( p > d \), respectively, and that
\[ n^{2p/d} \cdot P\{X \notin [-\log(n)^\alpha, \log(n)^\alpha]\} \to 0 \quad (n \to \infty), \]
then the above Monte Carlo surrogate quantile estimate achieves a rate of convergence better than \( n^{-1/2} \) or \( n^{-1} \), respectively.

**Remark 5.** It follows from Markov inequality that (10) is for instance satisfied if
\[ E\{\exp(X)\} < \infty \quad \text{and} \quad \frac{N_n}{n^\alpha} \to 0 \quad (n \to \infty). \]

### 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 in case that the function \( m \) is Hölder-smooth, and then we will 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 \( P_n = \{A_0, A_1, \ldots, A_{n-1}\} \) 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
\[ (x_{A_0}, m(x_{A_0})), \ldots, (x_{A_{n-1}}, m(x_{A_{n-1}})). \] (11)

For \( x \in \mathbb{R}^d \) denote by \( A_n(x) \) that cell \( A_j \in P_n \) which contains \( x \). Then the partitioning estimate \( m_n \) is defined by
\[ m_n(x) = m(x_{A_n(x)}). \] (12)

The key trick in the definition of our adaptive partitioning estimate is the 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
\[ |m_n(X_{n+i}) - m(X_{n+i})| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_m(X) - m(X_{n+i})| \] (13)
holds for \( i \in \{1, \ldots, N_n\} \) for some small \( \delta_n > 0 \).

To do this, we start by subdividing our data (11) into two parts of size

\[
n_1 = \lceil \frac{n}{2} \rceil \quad \text{and} \quad n_2 = n - n_1.
\]

We choose \( \alpha > 0 \), set

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

and partition \( C_n \) into \( [n_1^{1/d}]^d \) equivolume cubes of side length \( 2 \cdot \log(n)^\alpha/[n_1^{1/d}] \). We denote these cubes by \( A_j \ (j = 1, \ldots, [n_1^{1/d}]^d) \), set \( A_0 = \mathbb{R}^d \setminus C_n \) and let \( m_{n_1} \) be the partitioning estimate corresponding to the partition \( P_n = \{A_j : j = 0, 1, \ldots, [n_1^{1/d}]^d\} \) of \( \mathbb{R}^d \), where for \( A \in P_n \) 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 \( x \in C_n \):

\[
|m_{n_1}(x) - m(x)| \leq C \cdot \|x_{A_{n_1}}(x) - x\|^p \leq C \cdot \text{diam}(A_{n_1}(x))^p,
\]

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

\[
|m_{n_1}(x) - m(x)| \leq \log(n)^{\alpha p + 1} \cdot n^{-p/d}
\]

for \( n \) sufficiently large.

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

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

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

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

then we know from the proof of Theorem 2 below that we have outside of an event whose probability tends to zero

\[
\left|\hat{q}^{(MC)}_{m_{n_1}(X), N_n, \alpha} - q_{m(X), \alpha}\right| \leq \frac{\log(n)^{\alpha p + 1}}{np/d}
\]

and (as already derived above)

\[
|m_{n}(x) - m(x)| \leq \log(n) \cdot \text{diam}(A_{n}(x))^p \quad \text{for all} \ x \in C_n.
\]

By the triangle inequality we can conclude from (15)

\[
\left|\hat{q}^{(MC)}_{m_{n_1}(X), N_n, \alpha} - m_n(X_{n+i})\right| \leq \frac{\log(n)^{\alpha p + 1}}{np/d} + |q_{m(X), \alpha} - m(X_{n+i})| + |m(X_{n+i}) - m_n(X_{n+i})|
\]
for $i \in \{1, \ldots, N_n\}$.

If (15) (and hence also (17)) holds, then (13) holds as well for all $i \in \{1, \ldots, N_n\}$ if for all $i \in \{1, \ldots, N_n\}$ at least one of the following two conditions is satisfied:

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

(18)

or

$$|m_n(X_{n+i}) - m(X_{n+i})| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m_n(X),N_n,\alpha}^{(MC)} - m_n(X_{n+i})| - \frac{1}{2} \cdot |m_n(X_{n+i}) - m(X_{n+i})|.\quad (19)$$

Here (19) is equivalent to

$$3 \cdot |m_n(X_{n+i}) - m(X_{n+i})| \leq \delta_n + |q_{m_n(X),N_n,\alpha}^{(MC)} - m_n(X_{n+i})| - \frac{\log(n)^{\alpha-p+1}}{n^{p/d}}.\quad (20)$$

Using the bound (16) on the pointwise error of our estimate $m_n$ we see that for sufficiently large $n$ (13) holds for all $i \in \{1, \ldots, N_n\}$ if for all $i \in \{1, \ldots, N_n\}$ at least one of the following two conditions is satisfied:

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

(21)

or

$$3 \cdot \log(n) \cdot \text{diam}(A_n(X_{n+i}))^p + \frac{\log(n)^{\alpha-p+1}}{n^{p/d}} - |q_{m_n(X),N_n,\alpha}^{(MC)} - m_n(X_{n+i})| \leq \delta_n.\quad (22)$$

Hence the aim in choosing our partition is to choose it such that the following term is small:

$$\max_{i \in \{1, \ldots, N_n\}} \min_{X_{n+i} \in C_n} \left\{ 2 \cdot \log(n) \cdot \text{diam}(A_n(X_{n+i}))^p, 3 \cdot \log(n) \cdot \text{diam}(A_n(X_{n+i}))^p + \frac{\log(n)^{\alpha-p+1}}{n^{p/d}} - |q_{m_n(X),N_n,\alpha}^{(MC)} - m_n(X_{n+i})| \right\}.\quad (23)$$

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

$$\min \left\{ 2 \cdot \log(n) \cdot \text{diam}(A_j)^p, 3 \cdot \log(n) \cdot \text{diam}(A_j)^p + \frac{\log(n)^{\alpha-p+1}}{n^{p/d}} - |q_{m_n(X),N_n,\alpha}^{(MC)} - m_n(X_{n+i})| \right\}.\quad (24)$$

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

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

**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 \(q_\alpha\) and that \(m\) is \((p, C)\)-smooth for some \(p \leq 1\). Define the Monte Carlo surrogate quantile estimate \(\hat{q}_{m_n(X), N, \alpha}^{(MC)}\) of \(q_{m(X), \alpha}\) as in Section 2 where \(m_n\) is the adaptive partitioning estimate defined above. Assume furthermore

\[
N_n \geq n^{2p/d} \quad \text{and} \quad N_n \cdot P\{X / \in [\log(n)\alpha, \log(n)\alpha]\} \to 0 \quad (n \to \infty).
\] (25)

**a)** Then

\[
\left| \hat{q}_{m_n(X), N, \alpha}^{(MC)} - q_{m(X), \alpha} \right| = O_P\left( \frac{\log(n)^{\alpha p+1}}{n^{p/d}} \right).
\]

**b)** If, in addition, for some \(\epsilon_0 > 0\)

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

is contained in a cube of length less than \(c_2 \cdot \epsilon^{1/p}\) for all \(0 < \epsilon < \epsilon_0\), then

\[
\left| \hat{q}_{m_n(X), N, \alpha}^{(MC)} - q_{m(X), \alpha} \right| = O_P\left( \frac{\log(n)^{\alpha p+3}}{n^{2p/d}} + \frac{1}{\sqrt{N_n}} \right),
\]

and in particular for \(N_n = n^{4p/d}\) we get

\[
\left| \hat{q}_{m_n(X), N, \alpha}^{(MC)} - q_{m(X), \alpha} \right| = O_P\left( \frac{\log(n)^{\alpha p+3}}{n^{2p/d}} \right).
\]

**Remark 6. a)** It is possible to modify the definition of the estimate such that it achieves the rate of convergence \(n^{-2p/d}\) (up to some logarithmic factor) also in case that \(m\) be \((p, C)\)-smooth with \(p > 1\). 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 \(Q_A m\) corresponding to a function \(m : \mathbb{R}^d \to \mathbb{R}\), \(Q_A\) uses function evaluations of \(m\) at the \((M + 1)^d\) points

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

(26)

to construct a polynomial

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

satisfying

\[
\|m - Q_A m\|_{\infty, A} \leq c_3 \cdot h^p
\]

12
in case that \( m \) be \((p,C)\)-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 \( m \) at the points (26).

Then we define the partition \( \{ A_0, A_1, \ldots, A_{(n-1)/2} \} \) 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
\[
\left| \hat{q}^{(MC)}_{m_n(X), N_n, \alpha} - q_{m(X), \alpha} \right| = O_p \left( \frac{\log(n)^{\alpha p + 3}}{n^{2p/d}} \right)
\]
if \( m \) be \((p,C)\)-smooth for some \( 0 < p \leq M + 1 \).

b) In case of \( d > 4p \) 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: We estimate in the first step of the definition of the estimate the quantile \( q_{m(X), \alpha} \) by \( \hat{q}^{(MC)}_{m(X), n_1, \alpha} \), and replace in (24) the term
\[
\frac{\log(n)^{\alpha p + 1}}{n^{p/d}}
\]
by
\[
\frac{\log(n)}{\sqrt{n}}
\]
In this case it follows as in the proof of Theorem 3 that the resulting estimate satisfies under the assumptions of Theorem 3 b)
\[
\left| \hat{q}^{(MC)}_{m_n(X), N_n, \alpha} - q_{m(X), \alpha} \right| = O_p \left( \frac{\log(n)}{n^{\frac{d}{2d+2}}} \right).
\]

5 Application to simulated data

In this section we consider the finite sample size behaviour of three different quantile estimates. The first one is based on the simple order statistics, the second one is a Monte Carlo surrogate quantile estimate based on a non-adaptively chosen surrogate and the third one is the adaptive surrogate quantile estimate of Section 4. Here the first one (order. stat.) is the order statistics estimate defined by (1) and (2). The second one (sur. quant.) is the non-adaptive Monte Carlo surrogate quantile estimate of Section 2, but instead of a quasi-interpoland we use a smoothing spline (as implemented in the routine \( \text{Tps}() \) 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. For our third estimate (part. quant.) 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 cells which contain none of the \( X_{n+1}, \ldots, X_{n+N_n} \). Furthermore we replace the factor \( \frac{\log(n)^{p+1}}{n^{p/d}} \) by \( \log(n)/\sqrt{n} \).

We compare these three quantile estimates in four different models. In the first and in the second model the dimension is \( d = 1 \), we allow \( n \in \{20, 200, 2000\} \) evaluations of \( m \), and the second and the third quantile estimates are based on \( N_n = 50,000 \) and (in case of \( n = 2000 \)) \( N = 100,000 \) additionally observed values of \( X \). 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).

For the first model we choose

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

hence \( m(X) \) is lognormally distributed. The boxplots of the errors of the three different estimates occurring in the 100 simulations for each sample sizes are presented in Figure 1.

---

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

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

\[
m(x) = \begin{cases} 
\exp(x), & x \leq u_{0.95}, \\
\exp(u_{0.95}), & u_{0.95} < x \leq 1.9, \\
\exp(x - 1.9 + u_{0.95}), & \text{else},
\end{cases}
\]
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 the 100 simulation for each pair of sample sizes are presented in the boxplots in Figure 2.

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 the 100 simulation for each pair of sample sizes are presented in the boxplots in Figure 3.

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

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

In the third model we set \( d = 2 \), and we define

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

Figure 3: Boxplots of the occurring estimation errors in model 3 for the three different sample sizes. In the left panel we have \( n = 30 \) and \( N = 50,000 \), in the middle panel \( n = 300 \) and \( N = 50,000 \), and in the right panel \( n = 3,000 \) and \( N = 100,000 \).
of the function is especially useful. We set
\[
m(x) = \begin{cases} 
sqrt{1 + \|x\|^2}, & \|x\|^2 \leq \chi^2_{0.95,4}; \\
\sqrt{1 + \chi^2_{0.95,4}}, & \chi^2_{0.95,4} \leq \|x\|^2 \leq \chi^2_{0.95,4} + 1.9, \\
\sqrt{1 + \|x\|^2 - 1.9 + \chi^2_{0.95,4}}, & \text{else.}
\end{cases}
\]

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 the 100 simulation for each pair of sample sizes are presented in the boxplots in Figure 3.

![Boxplots](image)

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

In the first model both the surrogate Monte Carlo quantile estimate and the adaptively chosen partitioning Monte Carlo quantile estimate outperform the order statistics, where the surrogate Monte Carlo quantile estimate is slightly better than the partitioning estimate. But in the other three models the adaptively chosen partitioning Monte Carlo quantile estimate outperforms for large sample sizes simultaneously the surrogate Monte Carlo quantile estimate and the order statistics.

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 the uncertainty occurring 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 $|V_{\text{max}}|$ 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 $N\text{s/m}$ and a standard deviation of 2.25 $N\text{s/m}$. 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 $N\text{s/m}$, respectively and their standard deviations are 0.017 $kg$, 33.334 $N/m$ and 0.009 $N\text{s/m}$, respectively.

In the active case we simulate the value of $x = (m, k, b, g)$ with independent random variables, as defined above, and use an 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 as result $|V_{\text{max}}| = 0.10217 \text{ dB}$. 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 of the 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 an 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 lemma.

Lemma 1 Let $X$ be an $\mathbb{R}^d$-valued random variable, let $m, \bar{m} : \mathbb{R}^d \to \mathbb{R}$ be measurable functions and let $\alpha \in (0, 1)$. Set

$$q_{m(X), \alpha} = \inf \{y \in \mathbb{R} : \mathbb{P}\{m(X) \leq y\} \geq \alpha\} \quad \text{and} \quad q_{\bar{m}(X), \alpha} = \inf \{y \in \mathbb{R} : \mathbb{P}\{\bar{m}(X) \leq y\} \geq \alpha\}.$$ 

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

$$|\bar{m}(x) - m(x)| \leq \frac{\delta}{2} + \frac{1}{2} \cdot |q_{m(X), \alpha} - m(x)|. \quad (27)$$

Then

$$|q_{\bar{m}(X), \alpha} - q_{m(X), \alpha}| \leq \delta.$$ 

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

$$\mathbb{P}\{\bar{m}(X) \leq q_{m(X), \alpha} + \delta\} \geq \alpha \quad (28)$$

and

$$\mathbb{P}\{\bar{m}(X) \leq q_{m(X), \alpha} - \delta - \epsilon\} < \alpha \quad \text{for all } 0 < \epsilon < \delta. \quad (29)$$

To do this, assume that (28) and (29) hold. Then (28) implies

$$q_{\bar{m}(X), \alpha} = \inf \{y \in \mathbb{R} : \mathbb{P}\{\bar{m}(X) \leq y\} \geq \alpha\} \leq q_{m(X), \alpha} + \delta,$$

and since

$$\mathbb{P}\{\bar{m}(X) \leq y_1\} \leq \mathbb{P}\{\bar{m}(X) \leq y_2\} \quad \text{for all } y_1 \leq y_2$$

we can conclude from (29) that

$$\{y \in \mathbb{R} : \mathbb{P}\{\bar{m}(X) \leq y\} \geq \alpha\} \subseteq [q_{m(X), \alpha} - \delta, \infty),$$

which implies

$$q_{\bar{m}(X), \alpha} = \inf \{y \in \mathbb{R} : \mathbb{P}\{\bar{m}(X) \leq y\} \geq \alpha\} \geq q_{m(X), \alpha} - \delta.$$

In the second step of the proof we show (28). Here it suffices to show

$$[\bar{m}(X) \leq q_{m(X), \alpha} + \delta] \supseteq [m(X) \leq q_{m(X), \alpha}] \quad a.s., \quad (30)$$
because from (30) and the definition of \( q_{m(X),\alpha} \) we get

\[
P[\bar{m}(X) \leq q_{m(X),\alpha} + \delta] \geq P[m(X) \leq q_{m(X),\alpha}] \geq \alpha.
\]

(30) follows from

\[
m(x) \leq q_{m(X),\alpha} \implies \bar{m}(x) \leq q_{m(X),\alpha} + \delta
\]

for \( P_X \)-almost all \( x \in \mathbb{R}^d \), which we show next.

Let \( x \in \mathbb{R}^d \) be such that (27) holds and assume \( m(x) \leq q_{m(X),\alpha} \). Then we get by (27)

\[
\bar{m}(x) \leq m(x) + |\bar{m}(x) - m(x)| \\
\leq m(x) + \frac{\delta}{2} + \frac{1}{2} |q_{m(X),\alpha} - m(x)| \\
= m(x) + \frac{\delta}{2} + \frac{1}{2} (q_{m(X),\alpha} - m(x)).
\]

In case that \( q_{m(X),\alpha} - m(x) \leq \delta \) holds this together with \( m(x) \leq q_{m(X),\alpha} \) implies

\[
\bar{m}(x) \leq q_{m(X),\alpha} + \frac{\delta}{2} + \frac{\delta}{2} = q_{m(X),\alpha} + \delta.
\]

And in case \( q_{m(X),\alpha} - m(x) > \delta \) we can conclude

\[
\bar{m}(x) \leq m(x) + \frac{1}{2} (q_{m(X),\alpha} - m(x)) + \frac{1}{2} (q_{m(X),\alpha} - m(x)) \leq q_{m(X),\alpha} + \delta.
\]

Hence in both cases we have \( \bar{m}(x) \leq q_{m(X),\alpha} + \delta \), which completes the second step of the proof.

In the third step of the proof we show (29). To do this we will show

\[
[\bar{m}(X) \leq q_{m(X),\alpha} - \delta - \epsilon] \subseteq [m(X) \leq q_{m(X),\alpha} - \epsilon] \ a.s. \tag{32}
\]

for all \( 0 < \epsilon < \delta \), which implies (29), because if (32) holds we can conclude from the definition of \( q_{m(X),\alpha} \) that

\[
P[\bar{m}(X) \leq q_{m(X),\alpha} - \delta - \epsilon] \leq P[m(X) \leq q_{m(X),\alpha} - \epsilon] < \alpha
\]

for all \( 0 < \epsilon < \delta \).

(32) is equivalent to

\[
[m(X) > q_{m(X),\alpha} - \epsilon] \subseteq [\bar{m}(X) > q_{m(X),\alpha} - \delta - \epsilon] \ a.s. \tag{33}
\]

for all \( 0 < \epsilon < \delta \), which is implied by

\[
m(x) > q_{m(X),\alpha} - \epsilon \implies \bar{m}(x) > q_{m(X),\alpha} - \delta - \epsilon \tag{34}
\]

for \( P_X \)-almost all \( x \in \mathbb{R}^d \) and all \( 0 < \epsilon < \delta \).
To prove (34) let $x \in \mathbb{R}^d$ be such that (27) holds and let $0 < \epsilon < \delta$. Assume furthermore that $m(x) > q_{m(X),\alpha} - \epsilon$. Then we can conclude from (27) that

$$m(x) = m(x) - |\tilde{m}(x) - m(x)| \geq m(x) - \frac{\delta}{2} - \frac{1}{2} |q_{m(X),\alpha} - m(x)|.$$  

In case that $|q_{m(X),\alpha} - m(x)| < \delta$ holds this together with our assumption $m(x) > q_{m(X),\alpha} - \epsilon$ implies

$$\tilde{m}(x) \geq m(x) - \delta > q_{m(X),\alpha} - \epsilon - \delta.$$  

And in case that $|q_{m(X),\alpha} - m(x)| \geq \delta$ holds we can conclude

$$\tilde{m}(x) \geq m(x) - |q_{m(X),\alpha} - m(x)| \geq m(x) - |q_{m(X),\alpha} - m(x) - \epsilon| - \epsilon = m(x) - (m(x) + \epsilon - q_{m(X),\alpha}) - \epsilon = q_{m(X),\alpha} - 2\epsilon \geq q_{m(X),\alpha} - \delta - \epsilon,$$  

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

**Proof of Theorem 1.** Lemma 1 depends only on the distribution $P_X$ of $X$ (and not on the concrete random variable). The key trick in the proof of Theorem 1 is to apply it with the empirical distribution $\hat{P}_X$ of $X_{n+1}, \ldots, X_{n+N_n}$ as distribution of $X$. To do this, we observe first that (4) together with triangle inequality implies that we have for $\hat{P}_X$-almost all $x \in \mathbb{R}^d$

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

Here $\hat{q}^{(MC)}_{m(X),N_n,\alpha}$ is by definition the $\alpha$-quantile corresponding to $m(X)$ if we choose $\hat{P}_X$ as the distribution of $X$. Application of Lemma 1 yields

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

An application of the triangle inequality yields the assertion. \hfill \Box

**6.2 Proof of Theorem 2**

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

$$\|m_n - m\|_{\infty,[-\beta_n,\beta_n]^d} \leq c_4 \cdot \log(n)^\alpha \cdot n^{-p/d}.$$  

(cf., e.g., proof of Theorem 1 in Kohler (2014)). Hence on the event that $X_{n+1}, \ldots, X_{n+N_n} \in [-\beta_n,\beta_n]^d$ (4) holds if we choose $\delta_n = 2 \cdot c_4 \cdot \log(n)^\alpha \cdot n^{-p/d}$, and we get from Theorem 1

$$\left|\hat{q}^{(MC)}_{m_n(X),N_n,\alpha} - \hat{q}_{m(X),\alpha}\right| \leq 2 \cdot c \cdot \log(n)^\alpha \cdot n^{-p/d} + 2 \cdot |\hat{q}^{(MC)}_{m(X),N_n,\alpha} - q_{m(X),\alpha}|.$$  

20
Since (10) implies
\[ P \{ X_{n+1}, \ldots, X_{n+N_n} \in [-\beta_n, \beta_n]^d \} \to 1 \quad (n \to \infty), \]
this together with (3) yields the assertion. \( \square \)

### 6.3 Proof of Theorem 3

a) Since \( m \) is \((p,C)\)-smooth we have for any \( x \in C_n \)
\[
|m_n(x) - m(x)| = |m(x_{A_n}(x)) - m(x)| \leq C \cdot \| x_{A_n}(x) - x \|^p \leq C \cdot \text{diam}(A_n(x))^p \\
\leq C \cdot \text{diam}(A_n(x))^p \leq C \cdot \left( \sqrt{d \cdot \frac{2 \cdot \log(n)^\alpha}{[n_1^{1/d}]} \right)^p.
\]

Hence on the event \( X_1, \ldots, X_n \in C_n \) condition (4) holds with
\[
\delta_n = \frac{\log(n)^{\alpha p+1}}{n^{p/d}}
\]
for \( n \) sufficiently large. From this we get the assertion as in the proof of Theorem 2.

b) Set
\[
\delta_n = \max_{i \in \{1, \ldots, N_n\}, X_{n+i} \in C_n} \min \left\{ 2 \cdot \log(n) \cdot \text{diam}(A_n(X_{n+i}))^p, \right. \\
\left. 3 \cdot \log(n) \cdot \text{diam}(A_n(X_{n+i}))^p + \frac{\log(n)^{\alpha+1}}{n^{p/d}} - |q^{(MC)}_{m_n, \alpha} - m_n(X_{n+i})| \right\}.
\] (35)

*In the first step* of the proof we show that the assertion follows from
\[
\delta_n = O_P \left( \frac{\log(n)^{\alpha p+3}}{n^{2p/d}} \right). \quad (36)
\]

If \( \delta_n \) 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 (15) 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. Application of Theorem 1 and (3) yields the assertion of step 1.

By the assumption of part b) of Theorem 3 we know that
\[
C_{\text{critical},n} := \left\{ x \in \mathbb{R}^d : m(x) \in \left[ q_m(X, \alpha) - 5 \cdot \frac{\log(n)^{\alpha p+2}}{n^{p/d}}, q_m(X, \alpha) + 5 \cdot \frac{\log(n)^{\alpha p+2}}{n^{p/d}} \right] \right\}
\]
is contained in a cube of sidelength less than \( c_5 \cdot \frac{\log(n)^{\alpha+2/p}}{n^{1/d}} \). Let \( E_n \) be the event that be the event that \( |q^{(MC)}_{m_n, \alpha} - q_m(X, \alpha)| \) is less than or equal to \( \log(n)^{\alpha p+1}/n^{p/d} \).

*In the second step* of the proof we show that we have on \( E_n \) for \( x \in C_n \setminus C_{\text{critical},n} \) and for \( n \) sufficiently large
\[
3 \cdot \log(n) \cdot \text{diam}(A_n(x))^p + \frac{\log(n)^{\alpha p+1}}{n^{p/d}} - |q^{(MC)}_{m_n, \alpha} - m_n(x)| \leq 0. \] (37)
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

\[
|q^\ast(x,\alpha) - m(x)| \leq \frac{\log(n)^{\alpha + 1}}{np/d} + |q^{(MC)}_{m,n}(x,\alpha) - m_n(x)| + C \cdot \text{diam}(A_n(x))^p
\]

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

\[
3 \cdot \log(n) \cdot \text{diam}(A_n(x))^p + \frac{\log(n)^{\alpha + 1}}{n} - |q^{(MC)}_{m,n}(x,\alpha) - m_n(x)|
\]

\[
\leq 3 \cdot \frac{\log(n)^{\alpha + 2}}{np/d} + 2 \cdot \frac{\log(n)^{\alpha + 1}}{np/d} - |q^\ast(x,\alpha) - m(x)|
\]

\[
\leq 5 \cdot \frac{\log(n)^{\alpha + 2}}{np/d} - 5 \cdot \frac{\log(n)^{\alpha + 2}}{np/d} = 0.
\]

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

\[
\min \left\{ 2 \cdot \log(n) \cdot \text{diam}(A_n(x))^p, \right. \]

\[
3 \cdot \log(n) \cdot \text{diam}(A_n(x))^p + \frac{\log(n)^{\alpha + 1}}{np/d} - \left| q^{(MC)}_{m,n}(X_n,\alpha) - m_n(X_{n+1}) \right| \right\}
\]

\[
\leq c_6 \cdot \frac{\log(n)^{\alpha + 3}}{n^{2p/d}}.
\] (38)

By the result of step 2 we know that on \(E_n\) and for \(n\) sufficiently large (37) holds for all \(x \in C_n \setminus C_{\text{critical},n}\). Hence as long as as any cell of the partition of the partitioning estimate, which has nonempty intersection with \(C_{\text{critical},n} \cap C_n\), does not satisfy (38), our algorithm does not choose any cell from the partition which has empty intersection \(C_{\text{critical},n} \cap C_n\). By the assumption of part b) of Theorem 3 we know that \(C_{\text{critical},n} \cap C_n\) is contained in a cube of sidelength less than \(\text{const} \cdot \log(n)^{\alpha + 2}/n^{1/d}\). But after \(n_2/2^d\) of the elements of the cells of the partition which have nonempty intersection with \(C_{\text{critical},n} \cap C_n\) are chosen we have for all \(x \in C_{\text{critical},n} \cap C_n\)

\[
\text{diam}(A_n(x)) \leq c_7 \cdot \frac{\log(n)^{\alpha + 2}/n^{1/d}}{n_2^{1/d}} \cdot \frac{1}{n_2^{1/d}} \leq c_8 \cdot \frac{\log(n)^{\alpha + 2}/n^{2/d}}{n_2^{2/d}},
\]

which implies the assertion of 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 \(P(E_n) \to 1 (n \to \infty)\) and \(P\{X_{n+1}, \ldots, X_{n+N_n} \in C_n\} \to 1 (n \to \infty)\). 

\[\square\]

7 Acknowledgment

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


