Nonparametric Quantile Estimation Using Surrogate Models and Importance Sampling

Michael Kohler[1] and Reinhard Tent[1,*]

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

January 05, 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 quantile estimates are constructed by estimating m by some estimate (surrogate) m_n and then by using an initial quantile estimate together with importance sampling to construct an importance sampling surrogate quantile estimate. 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 importance sampling surrogate quantile estimates are analyzed for two different function estimates. The finite sample size behavior of the estimates is investigated by applying them to simulated data.

AMS classification: Primary 62G05; secondary 62G30.

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

*Corresponding author. Tel: +49-6151-16-5288

Running title: *Nonparametric quantile estimation*

1 Introduction

Many physical phenomena are nowadays described and modeled with mathematical tools, and instead of real experiments so–called computer experiments are used to analyze physical phenomena. In this paper we assume that we have given a simulation model of a such a complex physical phenomenon described by

$$Y = m(X).$$

Here 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. The aim of the analysis of this simulation model is the prediction of the uncertainty of the random variable Y. Therefore we determine values which are with large probabilities upper or lower bounds on the (random) value of Y. More precisely, let

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

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

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

using at most n evaluations of the function m.

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

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

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

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

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

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

One well-known idea in the literature for the construction of estimates which achieve better results than the above simple order statistics is to use in a first step suitably chosen evaluations of m to construct an estimate (surrogate) m_n of m, and to use this estimate (maybe together with a few additional evaluations of m) to construct a quantile estimate. Here most of the approaches in the literature use a Bayesian approach to construct the surrogate m_n, e.g., 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.

In this paper we use instead results from non-Bayesian curve estimation in order to construct the surrogate m_n. Here in principle 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)).

Enss et al. (2014) considered so–called surrogate quantile estimates, where $q_{m(X),\alpha}$ is estimated by a Monte Carlo estimate of the quantile $q_{m_n(X),\alpha}$, i.e., by a Monte Carlo estimate of

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

It was shown there 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 $\delta_n > 0$

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

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

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

Here N_n is the sample size of the Monte Carlo estimate.

In Kohler et al. (2014) the surrogate m_n was used to construct so-called importance sampling quantile estimates. Importance sampling is a technique to improve estimation of the expectation of a function $\phi : \mathbb{R}^d \to \mathbb{R}$ by sample averages. Instead of using an independent and identically distributed sequence $X, X_1, X_2, \ldots$ and estimating $\mathbf{E}\phi(X)$ by

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

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

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

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

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

is estimated by

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

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

Quantile estimation using importance sampling has been considered by Cannamela, Garnier and Iooss (2008), Egloff and Leippold (2010) and Morio (2012). All three papers proposed new estimates in various models, however only Egloff and Leippold (2010) investigated theoretical properties (consistency) of their method. In Kohler et al. (2014) the rates of convergence of a newly proposed importance sampling quantile estimate have been analyzed. The basic idea there was to use an initial estimate of the quantile based on the order statistics of samples of $m(X)$ in order to determine an interval $[a_n, b_n]$ containing the quantile. Then an estimate m_n of m was constructed and f was restricted to the inverse image $m_n^{-1}([a_n, b_n])$ of $[a_n, b_n]$ to construct a new random variable Z, so only from an area, where the values are especially important for the computation of the quantile, values have been sampled. The final estimate of the quantile is then defined as an order statistic of $m(Z)$, where the level of the order statistics takes into account that it was sampled only from a part of the original density f. Under suitable assumptions on the smoothness of m and on the tails of f it was shown that this estimate achieves the rate of convergence of order $\frac{\log^{1.5} n}{n}$. This result requires that the supremum norm error of the estimate m_n of m is small. Furthermore the computation of the level of the quantile of $m(Z)$ requires the numerically exact computation of several integrals, which is in general not possible, and very time consuming if one wants to compute these integrals with high accuracy.

In this paper we extend the results from Kohler et al. (2014) in such a way that we firstly require only that a local error of m_n is small, that secondly the integrals are replaced by corresponding Monte Carlo estimates, and that thirdly for suitable smooth functions we achieve rates of convergence which are simultaneously better than $\frac{\log^{1.5} n}{n}$ and better than the rates of convergence proven in Enss et al. (2014).

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

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

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 $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$ the partial derivative $\frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$ exists and satisfies

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

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

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

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

A general error bound on our quantile estimate is presented in Section 2 and applied to global and local surrogate estimates in Sections 3 and 4, resp. Section 5 illustrates the finite sample size behaviour of our estimates by applying them to 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 with values in $\mathbb{R}^d$, which have a density f with respect to the Lebesgue measure. In this section we consider a general importance sampling surrogate quantile estimate, which depends on an estimate m_n of m and some initial quantile estimate $\tilde{q}_{m(X),n,\alpha}$. Since m is costly to evaluate, we want to restrict the number of evaluations by some natural number n. This number we split into three parts $n_1, n_2, n_3 \geq 0$ where $n_1 = n_2 = \lceil n/3 \rceil$ and $n_3 = n - n_1 - n_2$. We use $n_1 + n_2$ evaluations of m to generate the estimate m_n of m and to generate $\tilde{q}_{m(X),n,\alpha}$. Eventually we use the remaining n_3 evaluations to generate our final quantile estimate. Furthermore we use additional values of X, i.e., X_{n+1}, X_{n+2}, $\ldots$, in particular in order to construct Monte Carlo estimates of some integrals involving m_n (but not m).

The basic assumption on the estimate m_n is that it satisfies for some compact set $K_n \subseteq \mathbb{R}^d$ and some $\delta_n > 0$ with $\delta_n \to 0$ for $n \to \infty$ the relation

$$|m_n(x) - m(x)| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m(X),\alpha} - m(x)| \quad \text{for all } x \in K_n, \tag{5}$$

where we have in addition

$$0 < \frac{1}{\delta_n + \log(n) \cdot \eta_n} \cdot \mathbf{P}\{X \notin K_n\} \to 0 \quad (n \to \infty). \tag{6}$$

Furthermore, we assume that the initial quantile estimate $\tilde{q}_{m(X),n,\alpha}$ satisfies

$$\left|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}\right| = O_{\mathbf{P}}(\eta_n) \tag{7}$$

for some $\eta_n > 0$, where $\lim_{n\to\infty} \eta_n = 0$. In the sequel we use this initial quantile estimate together with m_n to construct an importance sampling surrogate quantile estimate.

Depending on m_n and $\tilde{q}_{m(X),n,\alpha}$ define $h_n : \mathbb{R}^d \to \mathbb{R}$ by

$$h_n(z) = \frac{1}{c_n} \cdot \left(I_{\{z \in K_n \,:\, |m_n(z) - \tilde{q}_{m(X),n,\alpha}| \leq \log(n) \cdot \eta_n + 2 \cdot \delta_n\}} + I_{\{z \notin K_n\}} \right) \cdot f(z),$$

where

$$c_n = \int_{\mathbb{R}^d} \left(I_{\{z \in K_n \,:\, |m_n(z) - \tilde{q}_{m(X),n,\alpha}| \leq \log(n) \cdot \eta_n + 2 \cdot \delta_n\}} + I_{\{z \notin K_n\}} \right) \cdot f(z) \, dz.$$

Let $Z_1, \dots, Z_{n_3}$ be independent and identically distributed with density h_n, define

$$\hat{G}_n^{(IS)}(y) = \frac{1}{n_3} \sum_{i=1}^{n_3} I_{\{m(Z_i) \leq y\}},$$

and set

$$\bar{\alpha} = \frac{\alpha - \int_{\mathbb{R}^d} \left(I_{\{z \in K_n \, : \, m_n(z) < \tilde{q}_{m(X),n,\alpha} - \log(n) \cdot \eta_n - 2 \cdot \delta_n\}} \right) \cdot f(z) \, dz}{c_n}.$$

As shown in Kohler, Krzyżak, Tent and Walk (2014) a natural choice for our importance sampling surrogate quantile estimate is $\inf\{y \in \mathbb{R} \, : \, \hat{G}_n^{(IS)}(y) \geq \bar{\alpha}\}$. Since it is hard to compute the integrals in the definition of $\bar{\alpha}$ exactly, we estimate $\bar{\alpha}$ by

$$\hat{\alpha} = \frac{\alpha - \frac{1}{n_{\hat{\alpha}}} \sum_{i=1}^{n_{\hat{\alpha}}} I_{\{z \in K_n \, : \, m_n(z) < \tilde{q}_{m(X),n,\alpha} - \log(n) \cdot \eta_n - 2 \cdot \delta_n\}}(X_{n+i})}{\frac{1}{n_{\hat{\alpha}}} \sum_{i=1}^{n_{\hat{\alpha}}} \left(I_{\{z \in K_n \, : |m_n(z) - \tilde{q}_{m(X),n,\alpha}| \leq \log(n) \cdot \eta_n + 2 \cdot \delta_n\}}(X_{n+i}) + I_{\{z \notin K_n\}}(X_{n+i}) \right)},$$

for some $n_{\hat{\alpha}} \in \mathbb{N}$. Then we define our importance sampling surrogate quantile estimate by

$$\hat{q}_{n,\hat{\alpha}}^{(IS)} = \inf\{y \in \mathbb{R} \, : \, \hat{G}_n^{(IS)}(y) \geq \hat{\alpha}\}.$$

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

Let m_n be an estimate of m and assume that (5), (6) and (7) hold for some $\delta_n, \eta_n > 0$ satisfying

$$\delta_n \to 0 \quad \textit{and} \quad \log(n) \cdot \eta_n \to 0 \qquad (n \to \infty). \tag{8}$$

Define the importance sampling surrogate quantile estimate $\hat{q}_{n,\alpha}^{(IS)}$ as above, where

$$n_{\hat{\alpha}} \geq \frac{n}{(\delta_n + \eta_n)^2}. \tag{9}$$

Then

$$\left| \hat{q}_{n,\alpha}^{(IS)} - q_{m(X),\alpha} \right| = O_{\mathbf{P}} \left(\frac{\delta_n + \eta_n \cdot \log(n)}{\sqrt{n}} \right).$$

Remark 1. If we choose the initial quantile estimate as surrogate quantile estimate corresponding to m_n and N_n is sufficiently large, then Theorem 2 in Enss et al. (2014) implies that under suitable assumptions on $\mathbf{P}\{X \notin K_n\}$ condition (7) holds with $\eta_n = \delta_n$. Hence in this case Theorem 1 implies that our importance sampling surrogate quantile estimate improves the rate δ_n of the surrogate quantile estimate to $\delta_n \cdot \log(n)/\sqrt{n}$.

Remark 2. Values of $Z_1, Z_2, \dots$ can be constructed using a rejection method. To do this one selects from values of $X_1, X_2, \dots$ successively all those values $x = X_i$ where either $x \in K_n$ and $|m_n(x) - \hat{q}_{m_n(X),N_n,\alpha_n}| \leq \log(n) \cdot \eta_n + 2\delta_n$ hold or where $x \notin K_n$ holds.

3 An importance sampling 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 $\gamma > 0$ and set $\beta_n = \log(n)^\gamma$. 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 $supp(B_{k,M}) = [u_k, u_{k+M+1}]$, see, e.g., de Boor (1978), or 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

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

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

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

and that they vanish outside of the set

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

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

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

where

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

for some $a_{\mathbf{k},\mathbf{j}} \in \mathbb{R}$ and 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 $a_{\mathbf{k},\mathbf{j}}$ (which can be computed by solving a linear equation system), such that

$$|Q_{\mathbf{k}} f| \leq c_1 \cdot \|f\|_{\infty, [u_{k_1}, u_{k_1+M+1}] \times \cdots \times [u_{k_d}, u_{k_d+M+1}]} \tag{10}$$

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

$$(Qp)(x) = p(x) \quad (x \in [-\beta_n, \beta_n]^d) \tag{11}$$

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

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

$$K = K_n = \left\lfloor \frac{\lfloor n_1^{1/d} \rfloor - 1}{2 \cdot M} \right\rfloor.$$

Furthermore we choose $x_1, \dots, x_{n_1}$ such that all of the $(2 \cdot M \cdot K + 1)^d$ points of the form

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

are contained in $\{x_1, \dots, x_{n_1}\}$, which is possible since $(2 \cdot M \cdot K + 1)^d \leq n_1$. Then we define

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

where Qm is the above defined quasi interpolant satisfying (10) and (11). The computation of Qm requires only function values of m at the points $x_1, \dots, x_{n_1}$, i.e., the estimate depends on the data

$$(x_1, m(x_1)), \dots, (x_{n_1}, m(x_{n_1})),$$

and hence m_n is well defined.

For our initial quantile estimate $\tilde{q}_{m(X),n,\alpha}$ we use $X_{n+n_{\hat{\alpha}}+1}, \dots, X_{n+n_{\hat{\alpha}}+N_n}$ 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

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

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 on $\mathbb{R}$ and positive at $q_{m(X),\alpha}$ and that m is (p,C)-smooth for some $p > 0$ and some $C > 0$. Define the Monte Carlo surrogate importance sampling quantile estimate $\hat{q}^{(IS)}_{n,\hat{\alpha}}$ of $q_{m(X),\alpha}$ as in Section 2, where m_n is the spline approximand defined above with parameter $M \geq p-1$, where $\hat{q}_{m(X),n,\alpha}$ is chosen as the corresponding surrogate quantile estimate $\hat{q}^{(MC)}_{m_n(X),N_n,\alpha}$ defined above, and where the construction of the estimates uses the parameters*

$$\delta_n = \eta_n = \frac{\log(n)^{\gamma \cdot p}}{n^{p/d}}, \quad n_{\hat{\alpha}} = n^{1+2p/d} \quad \text{and} \quad N_n = \lceil n^{2p/d} / \log(n)^{2 \cdot \gamma \cdot p} \rceil. \tag{12}$$

Assume

$$N_n \cdot \mathbf{P}\{X \notin [-\log(n)^\gamma, \log(n)^\gamma]^d\} \to 0 \quad (n \to \infty). \tag{13}$$

Then

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

Proof. By the proof of Theorem 2 in Enss et al. (2014) we know that (5) and (7) hold with δ_n and η_n defined as in (12) and $K_n = [-\log(n))^\gamma, \log(n))^\gamma]^d$. Application of Theorem 1 yields the desired result. □

Remark 3. In any application the smoothness of m (measured by p and C) will be unknown hence the above choice of δ_n and η_n is not possible. In Section 5 we propose a data-driven choice for the values of δ_n and η_n and investigate its finite sample performance using simulated data.

4 An importance sampling surrogate quantile estimate based on an adaptively chosen surrogate

In this section we define an importance sampling surrogate quantile estimate based on an adaptive partitioning estimate m_n of m. Here m_n depends on $n_1 + n_2$ evaluations of m whereas our inital quantile estimate $\tilde{q}_{m(X),n,\alpha}$ requires $N_n > 0$ additional evaluations of m_n but none of m.

Our partitioning estimate depends on a partition $\mathcal{P}_{n_1+n_2} = \{A_0, A_1, \ldots, A_{n_1+n_2-1}\}$ of $\mathbb{R}^d$ and on the evaluation of m at points $x_{A_0} \in A_0$, $x_{A_1} \in A_1$, $\ldots$, $x_{A_{n_1+n_2-1}} \in A_{n_1+n_2-1}$, i.e., on the data

$$(x_{A_0}, m(x_{A_0})), \ldots, (x_{A_{n_1+n_2-1}}, m(x_{A_{n_1+n_2-1}})). \tag{14}$$

For $x \in \mathbb{R}^d$ denote by $A_{n_1+n_2}(x)$ that cell $A_j \in \mathcal{P}_{n_1+n_2}$ which contains x. Then the partitioning estimate m_n is defined by

$$m_n(x) := m_{n_1+n_2}(x) = m(x_{A_{n_1+n_2}(x)}). \tag{15}$$

The key trick in the definition of our adaptive partitioning estimate is the adaptive choice of the sets A_0, A_1, $\ldots$, $A_{n_1+n_2-1}$ (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

$$|m_n(x) - m(x)| \leq \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m(X),\alpha} - m(x)| \tag{16}$$

holds for all $x \in K_n := [-\log(n), \log(n)]^d$ and for some small $\delta_n > 0$.

To do this, we start by partitioning $K_n = [-\log(n), \log(n)]^d$ into $\lfloor n_1^{1/d} \rfloor^d$ equivolume cubes of side length $2 \cdot \log(n)/\lfloor n_1^{1/d} \rfloor$. We denote these cubes by A_j $(j = 1, \ldots, \lfloor n_1^{1/d} \rfloor^d)$, set $A_0 = \mathbb{R}^d \setminus K_n$ and let m_{n_1} be the partitioning estimate corresponding to the partition $\mathcal{P}_{n_1} = \{A_j : j = 0, 1, \ldots, \lfloor n_1^{1/d} \rfloor^d\}$ 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 $x \in K_n$:

$$|m_{n_1}(x) - m(x)| \quad \leq \quad C \cdot \|x_{A_{n_1}(x)} - x\|^p \leq C \cdot diam(A_{n_1}(x))^p \leq \log(n)^{p+1} \cdot n^{-p/d}$$

for n sufficiently large. We use m_{n_1} to define the Monte Carlo surrogate quantile estimate

$$\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} = \inf\{y \in \mathbb{R} : \hat{G}^{(MC)}_{m_{n_1}(X),N_n}(y) \geq \alpha\}, \tag{17}$$

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

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

then the proof of Theorem 2 in Enss et al. (2014) implies that we have outside of an event whose probability tends to zero

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

and (as already derived above)

$$|m_{n_1}(x) - m(x)| \leq \log(n) \cdot diam(A_{n_1}(x))^p \quad \text{for all } x \in K_n. \tag{19}$$

Remark that until now we just used n_1 data points and have n_2 additional data left and that if we had used the whole available data in our previous approach the right-hand side of (18) would only be improved by some constant factor. Our goal is to use the remaining n_2 data points more efficiently. To this end we derive a better estimate of how big our approximation error is on each set of our partition and use the remaining data points to refine those sets which bring the worst results.

For this purpose we show next that for a given δ_n condition (5) holds, when for all $x \in K_n$ we have

$$|m_n(x) - m(x)| \leq \frac{\delta_n}{2} \tag{20}$$

or

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

Whereas condition (20) is quite obvious the usefulness of condition (21) follows for n large enough from triangle inequality and (18) by

$$
\begin{aligned}
3 \cdot |m_n(x) - m(x)| &\le \delta_n + |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| - \frac{\log(n)^{p+1}}{n_1^{p/d}} \\
&\le \delta_n - \frac{\log(n)^{p+1}}{n_1^{p/d}} + |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - q_{m(X),\alpha}| \\
&\quad + |q_{m(X),\alpha} - m(x)| + |m(x) - m_n(x)| \\
&\le \delta_n + |q_{m(X),\alpha} - m(x)| + |m(x) - m_n(x)|,
\end{aligned}
\tag{22}
$$

which by rearrangement implies

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

From this we conclude that δ_n can be chosen in such a way that for every $x \in K_n$ at least one of the following conditions holds:

$$
\begin{aligned}
\delta_n &\ge 2 \cdot \log(n) \cdot \mathrm{diam}(A_{n_1+n_2}(x))^p, \\
\delta_n &\ge 3 \cdot \log(n)\, \mathrm{diam}(A_{n_1+n_2}(x))^p - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| + \frac{\log(n)^{p+1}}{n_1^{p/d}}.
\end{aligned}
$$

Hence in order to minimize δ_n the goal is to refine our partition in a way such that

$$
\sup_{x \in K_n} \min \Bigg\{ 2 \cdot \log(n) \cdot \mathrm{diam}(A_{n_1+n_2}(x))^p, \\
3 \cdot \log(n) \cdot \mathrm{diam}(A_{n_1+n_2}(x))^p - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| + \frac{\log(n)^{p+1}}{n_1^{p/d}} \Bigg\}
$$

is small. This can be accomplished recursively, by subdividing that cube A^* of our partition of K_n into 2^d smaller equivolume, cubes for which

$$
\min \Bigg\{ 2 \cdot \log(n) \cdot \mathrm{diam}(A^*)^p, \\
3 \cdot \log(n) \cdot \mathrm{diam}(A^*)^p - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x_{A^*})| + \frac{\log(n)^{p+1}}{n_1^{p/d}} \Bigg\}
$$

is maximal. Now for each of our new formed cubes we choose one of our remaining data points to lie in there. This procedure is repeated until more than $n_1 + n_2 - 1 - 2^d$ data points are used.

After that we define a new surrogate quantile estimate by

$$
\hat{q}^{(MC)}_{m_{n_1+n_2}(X),N_n,\alpha} = \inf\{y \in \mathbb{R} : \hat{G}^{(MC)}_{m_{n_1+n_2}(X),N_n}(y) \ge \alpha\}
$$

where

$$\hat{G}^{(MC)}_{m_{n_1+n_2}(X),N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m_{n_1+n_2}(X_{n+i}) \le y\}},$$

and use this quantile estimate as quantile estimate $\tilde{q}_{m(X),n,\alpha}$ in the definition of the importance sampling surrogate quantile estimate $\hat{q}^{(IS)}_{n,\hat{\alpha}}$ in Section 2.

Theorem 3 *Let X be an $\mathbb{R}^d$-valued random variable which has a density f with respect to the Lebesgue measure, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and let $\alpha \in (0,1)$. Let $q_{m(X),\alpha}$ be the α-quantile of $m(X)$ and assume that $m(X)$ has a density g with respect to the Lebesgue-Borel measure which is continuous on $\mathbb{R}$ and positive at $q_{m(X),\alpha}$. Assume furthermore that m is (p,C)–smooth for some $p \in (0,1]$ and $C > 0$.*

Set $n_1 = n_2 = \lceil n/3 \rceil$ and $n_3 = n - n_1 - n_2$. Define the importance sampling surrogate quantile estimate $\hat{q}^{(IS)}_{n,\alpha}$ as above, where

$$\eta_n = \frac{\log(n)^p}{n^{p/d}} \quad \textit{and} \quad \delta_n = \frac{\log(n)^{p+1}}{n^{p/d}}.$$

Assume

$$N_n \ge n^{2p/d} \quad \textit{and} \quad N_n \cdot \mathbf{P}\{X \notin K_n\} \to 0 \qquad (n \to \infty). \tag{23}$$

a) When the integrals are estimated by a Monte Carlo estimate based on $n_{\hat{\alpha}} \ge n^{1+2p/d} \cdot \log(n)^{-2p}$ samples, we have

$$\left| \hat{q}^{(IS)}_{n,\alpha} - q_{m(X),\alpha} \right| = O_{\mathbf{P}} \left(\frac{\log(n)^{p+1}}{n^{1/2+p/d}} \right).$$

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 side length less than $\kappa_1 \cdot \epsilon^{1/p}$ for all $0 < \epsilon < \epsilon_0$, and if we change the above definitions of η_n, δ_n and $n_{\hat{\alpha}}$ to

$$\eta_n = \frac{\log(n)^{p+3}}{n^{2p/d}}, \quad \delta_n = \frac{\log(n)^{p+4}}{n^{2p/d}} \quad \textit{and} \quad n_{\hat{\alpha}} = \left\lceil n^{1+4p/d} \cdot \log(n)^{-2p} \right\rceil$$

then

$$\left| \hat{q}^{(IS)}_{n,\alpha} - q_{m(X),\alpha} \right| = O_{\mathbf{P}} \left(\frac{\log(n)^{p+4}}{n^{1/2+2p/d}} \right).$$

5 Application to simulated data

In this section we study the finite sample size behaviour of three different quantile estimates.

The first one (*order stat.*) is the order statistics estimate defined by (2). The second one (*sur. quant.*) is the Monte Carlo surrogate quantile estimate, which was used as the initial quantile estimate in Theorem 2 (applied to the whole sample $\mathcal{D}$ consisting of n data points), and evaluated on $N_n = 100,000$ additional values of X. Instead of a quasi-interpolant we use a smoothing spline (as implemented in the routine *Tps()* in *R* with smoothing parameter chosen by the generalized cross-validation as implemented in this routine). Since we apply it to data where the function is observed without additional error (i.e., in a noiseless regression estimation problem), this estimate results in an interpolating spline which gives similar result as the quasi-interpolant in Section 2, but is easier to implement.

For our third estimate (*imp. quant.*) we implement our importance sampling surrogate sampling estimate. Here we generate the values of Z by the rejection method described in Remark 2. Since the approximation error of m_n is important in the algorithm of our method but unknown in reality this error is to be approximated. To this end $\mathcal{D}_1$ denotes the first $\lfloor n/2 \rfloor$ data of $\mathcal{D}$. Now the data of $\mathcal{D}_1$ are divided into learning and testing data of equal size. With our learning data another surrogate of m is generated by the same method as before. This new surrogate is then evaluated on the testing data and the absolute differences between the true values and the approximated values are computed. The overall approximation error is then estimated by the median of these values.

In the next step the α-quantile of $m_n(X)$ is approximated by order statistics using a data set $\mathcal{D}_2$ which consists of N_n realizations of X and the evaluated data $m_n(\mathcal{D}_2)$. Here again the error of this estimate is important in our method but unknown. So anew we estimate this error by splitting the sample. More exactly we split $\mathcal{D}_1$ as well as $\mathcal{D}_2$ in five parts. Now for $i \in \{1, \dots 5\}$ the estimate $m_n^{(i)}$ of m is computed as m_n but using only the data $\mathcal{D}_1$ without the i-th part. Then only the i-th part of $\mathcal{D}_2$ is given to $m_n^{(i)}$ to compute an estimate of the α-quantile of $m_n^{(i)}(X)$ as before. We so achieve five new estimates of the quantile from which some are likely to be smaller than the true quantile and some are likely to be greater. We approximate the error of our estimation by taking the interquartile range of these five data.

In practical tests we had to realize that sometimes the simple order statistics led to a more reliable estimation of the quantile than the *sur. quant* method. Since we already generated data $m(\mathcal{D}_1)$ which are realizations of $m(X)$, we decided to compute the order statistics of $m(\mathcal{D}_1)$ as well to estimate the quantile. As before we approximate the error of this estimate by splitting $m(\mathcal{D}_1)$ into five parts, and computing five quantile estimates using order statistics with only one of those partial data at a time. If the interquartile range of these estimates is lower than the corresponding interquartile range of the above surrogate quantile estimates, the order statistics is chosen as initial quantile estimate, otherwise the surrogate quantile estimate is used.

The approximation of the new level $\bar{\alpha}$ of our quantile uses integration by Monte Carlo method with $n_{\hat{\alpha}} = 200,000$ data. In order to simplify the computation, the first $\lceil n/2 \rceil$ feasible data points which are generated here, are used as realizations of our random variable Z in the last step of our method.

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

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

In the second model the dimension of X is $d = 4$ and

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

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

$$m(x) = \sqrt{1 + \|x\|^2} \quad (x \in \mathbb{R}^d).$$

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

For all three estimates the sample sizes are chosen as $n = 100$, $n = 300$ and $n = 1000$. The estimates are applied in all three models and with the three different sample sizes in order to estimate quantiles of level $\alpha = 0.95$ and $\alpha = 0.99$. In each case the estimates are applied to 100 different independent random samples, and the median of the relative absolute errors, i.e., the median of the absolute errors divided by the quantile (and in brackets its interquartile range), are listed in Table 1.

From Table 1 we see that for $d = 1$ the surrogate quantile estimate and the importance sampling quantile estimate perform similarly and both improve the order statistcs for $n \leq 300$ substantially. For $d > 1$ the surrogate quantile estimate substantially improves the order statistics in 7 out of 12 settings of the simulations, but in contrast the importance sampling quantile estimate improves the order statistics in all 12 settings substantially.

Finally we illustrate the usefulness of our newly proposed estimate by applying it to an engineering simulation model. Here we consider a physical model of a spring-mass-damper with active velocity feedback for the purpose of vibration isolation (cf., Figure 1). The aim is to analyze the uncertainty occuring in the maximal 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

model	d	α	n	order stat.	sur. quant.	imp. quant.
1	1	0.95	100	0.1394 (0.1478)	0.0045 (0.0053)	0.0036 (0.0047)
1	1	0.95	300	0.0810 (0.1103)	0.0050 (0.0059)	0.0033 (0.0051)
1	1	0.95	1000	0.0480 (0.0649)	0.0043 (0.0043)	0.0045 (0.0048)
1	1	0.99	100	0.2608 (0.2364)	0.0122 (0.0210)	0.0074 (0.0123)
1	1	0.99	300	0.1341 (0.1604)	0.0061 (0.0104)	0.0050 (0.0089)
1	1	0.99	1000	0.0766 (0.0965)	0.0079 (0.0107)	0.0080 (0.0096)
2	4	0.95	100	0.1260 (0.1295)	0.0537 (0.0579)	0.0524 (0.0783)
2	4	0.95	300	0.0614 (0.0871)	0.0160 (0.0122)	0.0053 (0.0078)
2	4	0.95	1000	0.0419 (0.0460)	0.0043 (0.0045)	0.0036 (0.0038)
2	4	0.99	100	0.1190 (0.1209)	0.1302 (0.0799)	0.0345 (0.0391)
2	4	0.99	300	0.0565 (0.0630)	0.0426 (0.0210)	0.0069 (0.0085)
2	4	0.99	1000	0.0375 (0.0373)	0.0080 (0.0086)	0.0032 (0.0042)
3	4	0.95	100	0.0282 (0.0418)	0.0211 (0.0085)	0.0044 (0.0045)
3	4	0.95	300	0.0152 (0.0258)	0.0041 (0.0035)	0.0017 (0.0023)
3	4	0.95	1000	0.0110 (0.0147)	0.0012 (0.0014)	0.0014 (0.0020)
3	4	0.99	100	0.0579 (0.0612)	0.0578 (0.0173)	0.0058 (0.0082)
3	4	0.99	300	0.0299 (0.0415)	0.0233 (0.0065)	0.0043 (0.0052)
3	4	0.99	1000	0.0146 (0.0181)	0.0075 (0.0038)	0.0025 (0.0048)

Table 1: Simulation result for the three different models. Reported are the median of the relative absolute errors of the estimates (and in brackets their interquartile range) in 100 independent simulations.

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 100 function evaluations can be easily completed in approximately 20 seconds, but computation of 100,000 values requires about 5.5 hours.

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

We decribe the uncertainty in the corresponding (random) maximal magnification of the vibration amplitude by a confidence interval which contains the random value with probability 0.95. We estimate such a confidence interval by using 0 as a lower bound and an estimate of the 0.95

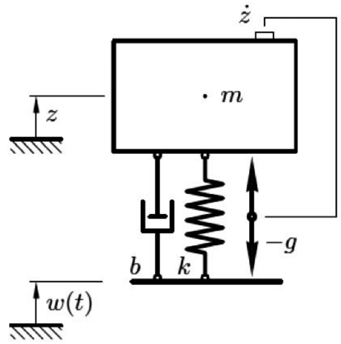

Figure 1: Spring-mass-damper with active velocity feedback (Platz and Enss (2015)).

quantile as an upper bound of the interval. We use an order statistics with sample size $n = 100,000$ to compute a reference value for this confidence interval. This results in $|V_{max}| \in [0, 0.1022]$ dB. But if we want to estimate this interval using only $n = 100$ evaluations of our function, we get with order statistics, the surrogate quantile estimate (with $50,000$ additional values of X) and the importance sampling quantile estimate (with $100,000$ additional values of X) the intervals $[0, 0.0827]$ dB, $[0, 0.1044]$ dB and $[0, 0.1038]$ dB, resp. As we can see, the estimated interval of the importance sampling quantile estimate is closer to our reference interval than the results of other two estimates.

6 Proofs

6.1 Proof of Theorem 1

We first proof a general result about importance sampling quantile estimation. Here we consider a general importance sampling estimate, where the importance sampling density is defined by restricting the density f of X to a general inverse image of a function estimate m_n. More precisely, we let $m_n, \hat{a}_n, \hat{b}_n : \mathbb{R}^d \to \mathbb{R}$ and $K_n \subseteq \mathbb{R}^d$ and set

$$h_n(z) = \frac{1}{c_n} \left(I_{\{z \in K_n \,:\, \hat{a}_n(z) \le m_n(z) \le \hat{b}_n(z)\}} + I_{\{z \notin K_n\}} \right) \cdot f(z) \quad (z \in \mathbb{R}^d),$$

where

$$c_n = \int_{\mathbb{R}^d} \left(I_{\{z \in K_n \,:\, \hat{a}_n(z) \le m_n(z) \le \hat{b}_n(z)\}} + I_{\{z \notin K_n\}} \right) \cdot f(z) \, dz.$$

Let Z, Z_1, $\dots$, Z_{n_3} be independent and identically distributed random variables with density h_n. Define

$$\hat{G}_n^{(IS)}(y) = \frac{1}{n_3} \sum_{i=1}^{n_3} I_{\{m(Z_i) \le y\}},$$

and

$$\hat{q}^{(IS)}_{n,\hat{\alpha}} = \inf\{y \in \mathbb{R} \, : \, \hat{G}^{(IS)}_n(y) \geq \hat{\alpha}\},$$

where $\hat{\alpha}$ is an estimate of

$$\bar{\alpha} = \frac{\alpha - \int_{\mathbb{R}^d} I_{\{z \in K_n \, : \, m_n(z) < \hat{a}_n(z)\}} \cdot f(z)\, dz}{c_n}.$$

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

Let $m_n, \hat{a}_n, \hat{b}_n : \mathbb{R}^d \to \mathbb{R}$ be functions which depend on the first $n_1 + n_2$ evaluations of m, let $K_n \subseteq \mathbb{R}^d$, and let $\hat{\alpha}$ be some estimate of $\bar{\alpha}$. Assume that for some $\delta_n, \epsilon_n > 0$ the following four conditions hold outside of an event, whose probability tends to zero for n tending to infinity:

$$0 < c_n \leq \kappa_2 \cdot (\log(n) \cdot \eta_n + \delta_n)\,, \tag{24}$$

for some constant $\kappa_2 > 0$,

$$m_n(x) < \hat{a}_n(x) \implies m(x) < q_{m(X),\alpha} - \epsilon_n \quad \text{and} \quad m_n(x) > \hat{b}_n(x) \implies m(x) > q_{m(X),\alpha} \tag{25}$$

for all $x \in K_n$,

$$q_{m(X),\alpha} - \delta_n - \eta_n \leq m(x) \leq q_{m(X),\alpha} + \delta_n + \eta_n \implies \hat{a}_n(x) \leq m_n(x) \leq \hat{b}_n(x) \tag{26}$$

for all $x \in K_n$, and

$$|\bar{\alpha} - \hat{\alpha}| \leq \frac{\kappa_3}{\log^2(n)}, \tag{27}$$

for some constant $\kappa_3 > 0$. Assume furthermore that (8) holds. Define the importance sampling surrogate quantile estimate $\hat{q}^{(IS)}_{n,\hat{\alpha}}$ as above. Then

$$\left| \hat{q}^{(IS)}_{n,\hat{\alpha}} - q_{m(X),\alpha} \right| = O_{\mathbf{P}}\left((\delta_n + \eta_n \cdot \log(n)) \cdot \left(n^{-1/2} + |\bar{\alpha} - \hat{\alpha}| \right) \right).$$

In the proof we will need the following two lemmas.

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

Assume that g is bounded away from zero in a neighborhood of $q_{m(X),\alpha}$.

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

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

and

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

Set

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

where

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

and set

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

Let Z be a random variable with density h and let $q_{m(Z),\bar{\alpha}}$ *be the* $\bar{\alpha}$*-quantile of* $m(Z)$*. Then*

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

Proof. See Lemma 2 in Kohler et al. (2014). □

Lemma 3 *Assume that X is a* $\mathbb{R}^d$*-valued random variable which has a density f with respect to the Lebesgue measure. Let* $m : \mathbb{R}^d \to \mathbb{R}$ *be a measurable function. Assume that* $m(X)$ *has a density g with respect to the Lebesgue measure. Let A be a measurable subset of* $\mathbb{R}$ *with the property that for all* $x \in K_n$ *we have*

$$m(x) \in A \quad \Rightarrow \quad \hat{a}_n(x) \leq m_n(x) \leq \hat{b}_n(x). \tag{28}$$

Let Z be defined as in Lemma 1. Then

$$\mathbf{P}_{m(Z)}\{A\} = \frac{1}{c_n} \cdot \int_A g(y)\, dy.$$

Proof. The proof is a modification of the proof of Lemma 3 in Kohler et al. (2014). By definition of Z and (28) we have

$$\begin{aligned}
\mathbf{P}_{m(Z)}\{A\} &= \int_{\mathbb{R}} I_{\{m(z)\in A\}} \cdot \frac{1}{c_n} \left(I_{\{z\in K_n\,:\,\hat{a}_n(z)\le m_n(z)\le \hat{b}_n(z)\}} + I_{\{z\notin K_n\}} \right) \cdot f(z)\,dz \\
&= \int_{\mathbb{R}} \frac{1}{c_n} \left(I_{\{m(z)\in A\}} \cdot I_{\{z\in K_n\}} + I_{\{m(z)\in A\}} \cdot I_{\{z\notin K_n\}} \right) \cdot f(z)\,dz \\
&= \frac{1}{c_n} \int_{\mathbb{R}} I_{\{m(z)\in A\}} \cdot f(z)\,dz \\
&= \frac{1}{c_n} \cdot \int_A g(y)\,dy,
\end{aligned}$$

where the last equality follows from the facts that f is the density of X and g is the density of $m(X)$. □

Proof of Lemma 1. Set $G_{m(Z)}(y) = \mathbf{P}\{m(Z) \le y\}$,

$$q_{m(Z),\bar{\alpha}} = \inf\left\{ y \in \mathbb{R} \,:\, G_{m(Z)}(y) \ge \bar{\alpha} \right\}$$

and

$$q_{m(Z),\hat{\alpha}} = \inf\left\{ y \in \mathbb{R} \,:\, G_{m(Z)}(y) \ge \hat{\alpha} \right\}.$$

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

$$q_{m(X),\alpha} = q_{m(Z),\bar{\alpha}}. \tag{29}$$

Set

$$A_n := \{x \in K_n \,:\, m_n(x) < \hat{a}_n(x)\} \quad \text{and} \quad B_n := \{x \in K_n \,:\, m_n(x) > \hat{b}_n(x)\}.$$

Using these sets c_n can be rewritten as

$$c_n = \mathbf{P}\{X \notin A_n \cup B_n \mid \mathcal{D}\}.$$

Here $\mathcal{D}$ stands for the first $n_1 + n_2$ sample points at which m is evaluated in order to generate m_n, $\hat{a}_n$ and $\hat{b}_n$. Since the choice of these points is left open they might be chosen randomly and so the sets A_n and B_n would be random themselves. In that case it is necessary to take the sample in condition. If on the other hand these points are chosen deterministically, the above condition could be left out but doesn't interfere if retained, as well. Now by (25) for $x \in A_n$ and $y \in B_n$ outside of an event whose probability goes to zero for n tending to infinity we have $m(x) \le q_{m(X),\alpha} - \epsilon_n$ and $m(y) > q_{m(X),\alpha}$. Application of Lemma 2 yields the assertion.

Remark as well that the assumptions above imply

$$\mathbf{P}\{X \in A_n \mid \mathcal{D}\} \leq \mathbf{P}\{m(X) \leq q_{m(X),\alpha} - \epsilon_n\} < \alpha$$

and

$$\mathbf{P}\{X \in B_n \mid \mathcal{D}\} \leq \mathbf{P}\{m(X) > q_{m(X),\alpha}\} = 1 - \alpha,$$

from which we conclude

$$\bar{\alpha} = \frac{\alpha - \mathbf{P}\{X \in A_n \mid \mathcal{D}\}}{1 - \mathbf{P}\{X \in A_n \mid \mathcal{D}\} - \mathbf{P}\{X \in B_n \mid \mathcal{D}\}} \in (0, 1]. \tag{30}$$

In the second step of the proof we show that outside of an event, whose probability tends to zero for $n \to \infty$, we have for all $y \in (q_{m(Z),\bar{\alpha}} - \delta_n - \eta_n, q_{m(Z),\bar{\alpha}} + \delta_n + \eta_n)$

$$|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\bar{\alpha}})| \geq c_1 \cdot \frac{|y - q_{m(Z),\bar{\alpha}}|}{\log(n) \cdot \eta_n + \delta_n}. \tag{31}$$

W.l.o.g. we can assume that (26) and $q_{m(X),\alpha} = q_{m(Z),\bar{\alpha}}$ hold. Set

$$I = (q_{m(X),\alpha} - \delta_n - \eta_n, q_{m(X),\alpha} + \delta_n + \eta_n).$$

Then, due to (26), by applying Lemma 3 we get for any subset $A \subset I$, outside of an event, whose probability tends to zero for $n \to \infty$,

$$\mathbf{P}_{m(Z)}(A) = c_n^{-1} \cdot \int_A g(y)dy \geq c_n^{-1} \cdot \inf_{y \in A} g(y) \cdot |A|,$$

where $|A|$ denotes the Lebesgue measure of A. Let $y \in (q_{m(Z),\bar{\alpha}} - \delta_n - \eta_n, q_{m(Z),\bar{\alpha}} + \delta_n + \eta_n)$. With

$$A = [\min\{y, q_{m(Z),\bar{\alpha}}\}, \max\{y, q_{m(Z),\bar{\alpha}}\}]$$

we get

$$|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\bar{\alpha}})| = \mathbf{P}_{m(Z)}(A) \geq c_n^{-1} \cdot \inf_{y \in A} g(y) \cdot |y - q_{m(Z),\bar{\alpha}}|.$$

Since $\eta_n \to 0$ $(n \to \infty)$ and $\delta_n \to 0$ $(n \to \infty)$ by (8) and since g is continuous on $\mathbb{R}$ and positive at $q_{m(X),\alpha}$ we know from step one that

$$\inf_{y \in A} g(y) \geq c_0 > 0$$

for some constant $c_0 > 0$, on an event whose probability tends to 1 for $n \to \infty$. Together with (24) this implies the assertion of the second step.

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

$$|q_{m(Z),\bar{\alpha}} - q_{m(Z),\hat{\alpha}}| \leq \frac{2}{c_1} \cdot (\log(n) \cdot \eta_n + \delta_n) \cdot |\hat{\alpha} - \bar{\alpha}|. \tag{32}$$

W.l.o.g. we can assume that (27), (31) and $\bar{\alpha} \neq \hat{\alpha}$ hold. Because of (27) we can assume furthermore that w.l.o.g. we have

$$\frac{2}{c_1} \cdot (\log(n) \cdot \eta_n + \delta_n) \cdot |\hat{\alpha} - \bar{\alpha}| < \eta_n + \delta_n.$$

By (31) we can conclude that $y_1 = q_{m(Z),\bar{\alpha}} + \frac{1}{c_1} \cdot (\log(n) \cdot \eta_n + \delta_n) \cdot |\hat{\alpha} - \bar{\alpha}|$ satisfies

$$\begin{aligned} G_{m(Z)}(y_1) &= G_{m(Z)}(q_{m(Z),\bar{\alpha}}) + |G_{m(Z)}(y_1) - G_{m(Z)}(q_{m(Z),\bar{\alpha}})| \\ &\geq \bar{\alpha} + c_1 \cdot \frac{|y_1 - q_{m(Z),\bar{\alpha}}|}{\log(n) \cdot \eta_n + \delta_n} = \bar{\alpha} + |\hat{\alpha} - \bar{\alpha}| \geq \hat{\alpha}, \end{aligned}$$

and that $y_2 = q_{m(Z),\bar{\alpha}} - \frac{2}{c_1} \cdot (\log(n) \cdot \eta_n + \delta_n) \cdot |\hat{\alpha} - \bar{\alpha}|$ satisfies

$$\begin{aligned} G_{m(Z)}(y_2) &= G_{m(Z)}(q_{m(Z),\bar{\alpha}}) - |G_{m(Z)}(y_2) - G_{m(Z)}(q_{m(Z),\bar{\alpha}})| \\ &\leq \bar{\alpha} - c_1 \cdot \frac{|y_2 - q_{m(Z),\bar{\alpha}}|}{\log(n) \cdot \eta_n + \delta_n} = \bar{\alpha} - 2 \cdot |\hat{\alpha} - \bar{\alpha}| < \hat{\alpha}, \end{aligned}$$

which implies the assertion.

In the fourth step of the proof we show that outside of an event, whose probability tends to zero for $n \to \infty$, we have for all $y \in (q_{m(Z),\hat{\alpha}} - \frac{1}{2} \cdot (\delta_n + \eta_n), q_{m(Z),\hat{\alpha}} + \frac{1}{2} \cdot (\delta_n + \eta_n))$

$$|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})| \geq c_1 \cdot \frac{|y - q_{m(Z),\hat{\alpha}}|}{\log(n) \cdot \eta_n + \delta_n}. \tag{33}$$

W.l.o.g. we can assume that (27), (29), (32) and

$$\frac{2}{c_1} \cdot (\log(n) \cdot \eta_n + \delta_n) \cdot |\hat{\alpha} - \bar{\alpha}| < \frac{1}{2} \cdot (\eta_n + \delta_n)$$

hold. Set

$$A = [\min\{y, q_{m(Z),\hat{\alpha}}\}, \max\{y, q_{m(Z),\hat{\alpha}}\}].$$

Then

$$A \subseteq (q_{m(X),\alpha} - \delta_n - \eta_n, q_{m(X),\alpha} + \delta_n + \eta_n),$$

from which we can conclude as in the second step of the proof that

$$|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})| = \mathbf{P}_{m(Z)}(A) \geq c_n^{-1} \cdot \inf_{y \in A} g(y) \cdot |y - q_{m(Z),\hat{\alpha}}| \geq c_1 \cdot \frac{|y - q_{m(Z),\hat{\alpha}}|}{\log(n) \cdot \eta_n + \delta_n}.$$

Here the last inequality follows from condition (24).

In the fifth step of the proof we show that outside of an event, whose probability tends to zero for $n \to \infty$, we have for all $y \in \mathbb{R}$

$$|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})| \geq c_1 \cdot \frac{\min\{|y - q_{m(Z),\hat{\alpha}}|, \frac{\delta_n + \eta_n}{4}\}}{\log(n) \cdot \eta_n + \delta_n}. \tag{34}$$

This follows directly from the fourth step, since by monotonicity of $G_{m(Z)}$ we have

$$|G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})| \geq |G_{m(Z)}(\bar{y}) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})|$$

where

$$\bar{y} = \min\left\{\max\left\{y, q_{m(Z),\hat{\alpha}} - \frac{\delta_n + \eta_n}{4}\right\}, q_{m(Z),\hat{\alpha}} + \frac{\delta_n + \eta_n}{4}\right\}.$$

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

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

By the proof of the fourth step we know that for $y \in (q_{m(Z),\hat{\alpha}} - \frac{1}{2} \cdot (\delta_n + \eta_n), q_{m(Z),\hat{\alpha}} + \frac{1}{2} \cdot (\delta_n + \eta_n))$ we have outside of an event, whose probability tends to zero for $n \to \infty$,

$$G_{m(Z)}(y) - G_{m(Z)}(q_{m(Z),\hat{\alpha}}) = \frac{1}{c_n} \cdot \int_{\min\{y, q_{m(Z),\hat{\alpha}}\}}^{\max\{y, q_{m(Z),\hat{\alpha}}\}} g(y) \, dy.$$

From this we can conlude, that $G_{m(Z)}$ is in a surrounding of $q_{m(Z),\hat{\alpha}}$ continuous and strictly increasing, which implies the assertion.

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

$$\left|G_{m(Z)}(\hat{q}_{n,\hat{\alpha}}^{(IS)}) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})\right| \leq 3 \cdot \sup_{t \in \mathbb{R}} \left|G_{m(Z)}(t) - \hat{G}_n^{(IS)}(t)\right|. \tag{35}$$

From the sixth step of the proof we deduce that we have for arbitrary $\varepsilon > 0$

$$\begin{aligned}
&\left|G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})\right| \\
&\leq \left|G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - \hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}})\right| + \left|\hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - \hat{\alpha}\right| + \left|\hat{\alpha} - G_{m(Z)}(q_{m(Z),\hat{\alpha}})\right| \\
&\leq \left|G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - \hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}})\right| + \hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - \hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}} - \varepsilon) + \left|\hat{\alpha} - G_{m(Z)}(q_{m(Z),\hat{\alpha}})\right| \\
&\leq \left|G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - \hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}})\right| + \left|\hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}})\right| \\
&\quad + \left|G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}} - \varepsilon)\right| + \left|G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}} - \varepsilon) - \hat{G}^{(IS)}_n(\hat{q}^{(IS)}_{n,\hat{\alpha}} - \varepsilon)\right| \\
&\quad + \left|\hat{\alpha} - G_{m(Z)}(q_{m(Z),\hat{\alpha}})\right| \\
&\leq 3 \cdot \sup_{t \in \mathbb{R}} \left|G_{m(Z)}(t) - \hat{G}^{(IS)}_n(t)\right| + \left|G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}} - \varepsilon)\right|.
\end{aligned}$$

Since ε was chosen arbitrarily it follows that

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

if $G_{m(Z)}$ is continuous in $\hat{q}^{(IS)}_{n,\hat{\alpha}}$. Since we already showed in step number six that $G_{m(Z)}$ is continuous on $(q_{m(Z),\hat{\alpha}} - \frac{1}{2} \cdot (\delta_n + \eta_n), q_{m(Z),\hat{\alpha}} + \frac{1}{2} \cdot (\delta_n + \eta_n))$ it suffices to show that outside of an event whose probability tends to zero for $n \to \infty$ we have

$$\left|\hat{q}^{(IS)}_{n,\hat{\alpha}} - q_{m(Z),\hat{\alpha}}\right| \leq \frac{1}{3} \cdot (\delta_n + \eta_n).$$

The above inequality however is an implication of

$$\hat{G}^{(IS)}_n \left(q_{m(Z),\hat{\alpha}} - \frac{1}{3} \cdot (\delta_n + \eta_n)\right) < \hat{\alpha} \leq \hat{G}^{(IS)}_n \left(q_{m(Z),\hat{\alpha}} + \frac{1}{3} \cdot (\delta_n + \eta_n)\right),$$

which in turn follows from

$$G_{m(Z)} \left(q_{m(Z),\hat{\alpha}} - \frac{1}{3} \cdot (\delta_n + \eta_n)\right) \quad < \quad \hat{\alpha} - \frac{\log(n)}{\sqrt{n_3}}, \tag{36}$$

$$G_{m(Z)} \left(q_{m(Z),\hat{\alpha}} + \frac{1}{3} \cdot (\delta_n + \eta_n)\right) \quad > \quad \hat{\alpha} + \frac{\log(n)}{\sqrt{n_3}} \tag{37}$$

and the Dvoretzky-Kiefer-Wolfowitz inequality (cf., e.g., Dvoretzky, Kiefer and Wolfowitz (1956) and Massart (1990)), which states that

$$\sup_{t \in \mathbb{R}} \left|G_{m(Z)}(t) - \hat{G}^{(IS)}_n(t)\right| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n_3}}\right).$$

By steps 6 and 4 we have

$$\begin{aligned}
\hat{\alpha} - G_{m(Z)}\left(q_{m(Z),\hat{\alpha}} - \frac{1}{3}\cdot(\delta_n+\eta_n)\right) &= G_{m(Z)}\left(q_{m(Z),\hat{\alpha}}\right) - G_{m(Z)}\left(q_{m(Z),\hat{\alpha}} - \frac{1}{3}\cdot(\delta_n+\eta_n)\right)\\
&\geq c_1\cdot\frac{\frac{1}{3}\cdot(\delta_n+\eta_n)}{\log(n)\cdot\eta_n+\delta_n}\\
&\geq \frac{c_1}{3\cdot\log(n)}\\
&> \frac{\log(n)}{\sqrt{n_3}},
\end{aligned}$$

which proves (36). Inequality (37) can be proven analogously.

In the eighth step of the proof we prove the assertion of Lemma 1.

To do this, set

$$s_n = \frac{2\cdot c}{c_1}\cdot(\delta_n+\eta_n\cdot\log(n))\cdot(n^{-1/2}+|\hat{\alpha}-\bar{\alpha}|)$$

for some $c \geq 1$, and let A_n be the event that (29), (32) and (35) hold.

Then the results of Step 1, Step 3, Step 5 and Step 7 together with the Dvoretzky-Kiefer-Wolfowitz inequality imply for n large enough

$$\begin{aligned}
&\mathbf{P}\left\{|\hat{q}^{(IS)}_{n,\hat{\alpha}} - q_{m(X),\alpha}| > s_n\right\}\\
&\leq \mathbf{P}\left\{|\hat{q}^{(IS)}_{n,\hat{\alpha}} - q_{m(Z),\hat{\alpha}}| + |q_{m(Z),\hat{\alpha}} - q_{m(Z),\bar{\alpha}}| + |q_{m(Z),\bar{\alpha}} - q_{m(X),\alpha}| > s_n\right\}\\
&\leq \mathbf{P}\{A_n^c\} + \mathbf{P}\left\{A_n \text{ holds and } |\hat{q}^{(IS)}_{n,\hat{\alpha}} - q_{m(Z),\hat{\alpha}}| > \frac{2\cdot c}{c_1}\cdot(\delta_n+\eta_n\cdot\log(n))\cdot n^{-1/2}\right\}\\
&\leq \mathbf{P}\{A_n^c\} + \mathbf{P}\left\{A_n \text{ holds and } \min\left\{|\hat{q}^{(IS)}_{n,\hat{\alpha}} - q_{m(Z),\hat{\alpha}}|, \frac{\delta_n+\eta_n}{4}\right\} > \frac{2\cdot c}{c_1}\cdot(\delta_n+\eta_n\cdot\log(n))\cdot n^{-1/2}\right\}\\
&\leq \mathbf{P}\{A_n^c\} + \mathbf{P}\left\{A_n \text{ holds and } |G_{m(Z)}(\hat{q}^{(IS)}_{n,\hat{\alpha}}) - G_{m(Z)}(q_{m(Z),\hat{\alpha}})| > 2c\cdot n^{-1/2}\right\}\\
&\leq \mathbf{P}\{A_n^c\} + \mathbf{P}\left\{\sup_{t\in\mathbb{R}}\left|G_{m(Z)}(t) - \hat{G}^{(IS)}_n(t)\right| > \frac{2c}{3}\cdot n^{-1/2}\right\}\\
&\leq \mathbf{P}\{A_n^c\} + 2\cdot\exp\left(-2\cdot n_3\cdot\left(\frac{2c}{3}\cdot n^{-1/2}\right)^2\right).
\end{aligned}$$

Now for $n \to \infty$ this term converges to $2\cdot\exp\left(-8\cdot c^2/27\right)$. With $c \to \infty$ we get the assertion. □

Proof of Theorem 1. Set

$$\hat{a}_n(x) = \tilde{q}_{m(X),n,\alpha} - \log(n)\cdot\eta_n - 2\cdot\delta_n \quad \text{and} \quad \hat{b}_n(x) = \tilde{q}_{m(X),n,\alpha} + \log(n)\cdot\eta_n + 2\cdot\delta_n.$$

The assertion follows from Lemma 1, provided we can show that with these definitions of $\hat{a}_n$ and $\hat{b}_n$ the assumptions (24), (25), (26) and

$$|\hat{\alpha} - \bar{\alpha}| = O_{\mathbf{P}}\left(n^{-1/2}\right) \tag{38}$$

are satisfied.

Proof of (24): At first we notice that $0 < c_n$ is fulfilled by definition of c_n and condition (6). So it suffices to show the second inequality in (24). To this end we show first that outside of an event, whose probability tends to zero for $n \to \infty$, for all $x \in K_n$

$$\tilde{q}_{m(X),n,\alpha} - \log(n) \cdot \eta_n - 2 \cdot \delta_n \leq m_n(x) \leq \tilde{q}_{m(X),n,\alpha} + \log(n) \cdot \eta_n + 2 \cdot \delta_n \tag{39}$$

implies

$$\left|m(x) - q_{m(X),\alpha}\right| \leq 5 \cdot \delta_n + 4 \cdot \log(n) \cdot \eta_n.$$

To do this, observe that on the event

$$\left\{\left|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}\right| \leq \log(n) \cdot \eta_n\right\}$$

(39) implies

$$\left|m_n(x) - q_{m(X),\alpha}\right| \leq 2 \cdot \log(n) \cdot \eta_n + 2 \cdot \delta_n. \tag{40}$$

From (5) and the triangle inequality we get furthermore

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

from which we conclude

$$\begin{aligned} |m_n(x) - m(x)| \quad &\leq \quad \delta_n + |q_{m(X),\alpha} - m_n(x)| \\ &\leq \quad 3 \cdot \delta_n + 2 \cdot \log(n) \cdot \eta_n. \end{aligned}$$

But this together with (40) in turn implies

$$\left|m(x) - q_{m(X),\alpha}\right| \leq |m(x) - m_n(x)| + \left|m_n(x) - q_{m(X),\alpha}\right| \leq 5 \cdot \delta_n + 4 \cdot \log(n) \cdot \eta_n.$$

Now, by (7) $\{|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}| \leq \log(n) \cdot \eta_n\}$ is an event whose probability tends to one as n tends to infinity and so we see that outside of an event, whose probability tends to zero for $n \to \infty$, we have

$$
\begin{aligned}
c_n &= \int_{\mathbb{R}^d} \left(I_{\{z\in K_n\,:\,|m_n(x)-\tilde{q}_{m(X),n,\alpha}|\le \log(n)\cdot\eta_n+2\cdot\delta_n\}} + I_{\{z\notin K_n\}} \right)\cdot f(z)\,dz \\
&\le \int_{\mathbb{R}^d} \left(I_{\{z\in K_n\,:\,|m(x)-q_{m(X),\alpha}|\le 5\cdot\delta_n+4\cdot\log(n)\cdot\eta_n\}} + I_{\{z\notin K_n\}} \right)\cdot f(z)\,dz.
\end{aligned}
$$

According to (6) for n large enough

$$
\int_{\mathbb{R}^d} I_{\{z\notin K_n\}}\cdot f(z)dz = \mathbf{P}\{X\notin K_n\} < \log(n)\cdot\eta_n+\delta_n,
$$

while for the first summand in the above integral we get

$$
\begin{aligned}
& \int_{\mathbb{R}^d} I_{\{z\in K_n\,:\,|m(z)-q_{m(X),\alpha}|\le 5\cdot\delta_n+4\cdot\log(n)\cdot\eta_n\}}\cdot f(z)\,dz \\
\le\;& \int_{\mathbb{R}^d} I_{\{|y-q_{m(X),\alpha}|\le 5\cdot\delta_n+4\cdot\log(n)\cdot\eta_n\}}\cdot g(y)\,dy \\
\le\;& (8\cdot\log(n)\cdot\eta_n+10\cdot\delta_n)\cdot\max_{y\in\mathcal{I}_n} g(y),
\end{aligned}
$$

where $\mathcal{I}_n = [q_{m(X),\alpha}-4\cdot\log(n)\cdot\eta_n-5\cdot\delta_n\ ,\ q_{m(X),\alpha}+4\cdot\log(n)\cdot\eta_n+5\cdot\delta_n]$. Now since g is assumed to be continuous and $4\cdot\log(n)\cdot\eta_n+5\cdot\delta_n$ tends to zero for $n\to\infty$ even the value

$$
\max_{n\in\mathbb{N}}\max_{y\in\mathcal{I}_n} g(y)
$$

exists, implying we can bound c_n by $\kappa_2\cdot(\log(n)\cdot\eta_n+\delta_n)$. So the second inequality of (24) is fulfilled as well.

Proof of (25): In the following assume that the event

$$
\left\{\left|\tilde{q}_{m(X),n,\alpha}-q_{m(X),\alpha}\right|\le\log(n)\cdot\eta_n\right\}
$$

holds. Furthermore let $x\in K_n$ be such that

$$
m_n(x) < \hat{a}_n(x) = \tilde{q}_{m(X),n,\alpha}-\log(n)\cdot\eta_n-2\cdot\delta_n. \tag{41}
$$

Now from (5) we conclude that

$$
m(x)\le m_n(x)+\frac{\delta_n}{2}+\frac{1}{2}\cdot|q_{m(x),\alpha}-m(x)|,
$$

which together with (41) implies

$$\begin{aligned} m(x) &< \tilde{q}_{m(X),n,\alpha} - \log(n) \cdot \eta_n - 2 \cdot \delta_n + \frac{\delta_n}{2} + \frac{1}{2} \cdot |q_{m(x),\alpha} - m(x)| \\ &\leq q_{m(X),\alpha} - \frac{3}{2} \cdot \delta_n + \frac{1}{2} \cdot |q_{m(x),\alpha} - m(x)|. \end{aligned}$$

Here for the last inequality we used that $|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}| \leq \log(n) \cdot \eta_n$ by assumption. Now in case $m(x) \leq q_{m(X),\alpha}$ the above inequality is equivalent to

$$m(x) < q_{m(X),\alpha} - \delta_n,$$

whereas in case $q_{m(X),\alpha} < m(x)$ it is equivalent to

$$m(x) < q_{m(X),\alpha} - 3 \cdot \delta_n.$$

So the first implication in (25) holds for $\epsilon_n = \delta_n$. Similarly we proof the second assertion by showing that for $x \in K_n$ satisfying

$$m_n(x) > \hat{b}_n(x) = \tilde{q}_{m(X),n,\alpha} + \log(n) \cdot \eta_n + 2 \cdot \delta_n,$$

we get by (5)

$$\begin{aligned} m(x) &\geq m_n(x) - \frac{\delta_n}{2} - \frac{1}{2} \cdot |q_{m(x),\alpha} - m(x)| \\ &> \tilde{q}_{m(X),n,\alpha} + \log(n) \cdot \eta_n + 2 \cdot \delta_n - \frac{\delta_n}{2} - \frac{1}{2} \cdot |q_{m(x),\alpha} - m(x)| \\ &\geq q_{m(X),\alpha} + \frac{3}{2} \cdot \delta_n - \frac{1}{2} \cdot |q_{m(x),\alpha} - m(x)|. \end{aligned}$$

Here again the last inequality follows from $|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}| \leq \log(n) \cdot \eta_n$. This time in case $m(x) < q_{m(X),\alpha}$ the above inequality is equivalent to

$$m(x) > q_{m(X),\alpha} + 3 \cdot \delta_n,$$

whereas in case $q_{m(X),\alpha} \leq m(x)$ we get

$$m(x) > q_{m(X),\alpha} + \delta_n.$$

Proof of (26): Assume that the event

$$\left\{ \left| \tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha} \right| \leq \frac{1}{2} \cdot \log(n) \cdot \eta_n \right\}$$

holds, and let $x \in K_n$ be such that

$$q_{m(X),\alpha} - \delta_n - \eta_n \leq m(x) \leq q_{m(X),\alpha} + \delta_n + \eta_n.$$

Then (5) implies

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

from which we conclude

$$\begin{aligned} \hat{a}_n(x) = \tilde{q}_{m(X),n,\alpha} - \log(n) \cdot \eta_n - 2 \cdot \delta_n \quad &\leq \quad q_{m(x),\alpha} - \frac{1}{2} \cdot \log(n) \cdot \eta_n - 2 \cdot \delta_n \\ &\leq \quad m(x) - \delta_n + \eta_n \cdot \left(1 - \frac{1}{2} \cdot \log(n)\right) \\ &\leq \quad m_n(x) + \eta_n \cdot \left(\frac{3}{2} - \frac{1}{2} \cdot \log(n)\right), \end{aligned}$$

and in the same way

$$\hat{b}_n(x) \ \geq \ q_{m(x),\alpha} + \frac{1}{2} \cdot \log(n) \cdot \eta_n + 2 \cdot \delta_n \ \geq \ m_n(x) - \eta_n \cdot \left(\frac{3}{2} - \frac{1}{2} \cdot \log(n)\right).$$

Now for $3 < \log(n)$ this implies (26).

Proof of (38): Set

$$J = \int_{\mathbb{R}^d} \left(I_{\{z \in K_n \ : \ m_n(z) < \tilde{q}_{m(X),n,\alpha} - \log(n) \cdot \eta_n - 2 \cdot \delta_n\}} \right) \cdot f(z) \, dz,$$

$$\hat{J} = \frac{1}{n_{\hat{\alpha}}} \sum_{i=1}^{n_{\hat{\alpha}}} I_{\{z \in K_n \ : \ m_n(z) < \tilde{q}_{m(X),n,\alpha} - \log(n) \cdot \eta_n - 2 \cdot \delta_n\}}(X_{n+i}),$$

and

$$\hat{c}_n = \frac{1}{n_{\hat{\alpha}}} \sum_{i=1}^{n_{\hat{\alpha}}} \left(I_{\{z \in K_n : |m_n(z) - \tilde{q}_{m(X),n,\alpha}| \leq \log(n) \cdot \eta_n + 2 \cdot \delta_n\}}(X_{n+i}) + I_{\{z \notin K_n\}}(X_{n+i}) \right).$$

Then

$$\bar{\alpha} = \frac{\alpha - J}{c_n} \quad \text{and} \quad \hat{\alpha} = \frac{\alpha - \hat{J}}{\hat{c}_n}.$$

The inequality of Markov implies

$$|\hat{J} - J| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n_{\hat{\alpha}}}}\right) \quad \text{and} \quad |\hat{c}_n - c_n| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n_{\hat{\alpha}}}}\right).$$

Furthermore we know by (26) and by our choice of $\hat{a}_n(x)$ and $\hat{b}_n(x)$

$$\begin{aligned} c_n &= \int_{\mathbb{R}^d} \left(I_{\{z \in K_n : \hat{a}_n(z) \le m_n(z) \le \hat{b}_n(z)\}} + I_{\{z \notin K_n\}} \right) \cdot f(z)\, dz \\ &\ge \int_{\mathbb{R}^d} \left(I_{\{z \in K_n : q_{m(X),\alpha} - \delta_n - \eta_n \le m(z) \le q_{m(X),\alpha} + \delta_n + \eta_n\}} + I_{\{z \notin K_n\}} \right) \cdot f(z)\, dz \\ &\ge \int_{q_{m(X),\alpha} - \delta_n - \eta_n}^{q_{m(X),\alpha} + \delta_n + \eta_n} g(y)\, dy \\ &\ge c \cdot (\delta_n + \eta_n), \end{aligned}$$

where the last inequality follows for large n from the fact that g is continuous on $\mathbb{R}$ and positive at $q_{m(X),\alpha}$. This together with (9) implies that outside of an event whose probability tends to zero for $n \to \infty$ we have

$$|c_n - \hat{c}_n| \le \frac{1}{2} \cdot c_n.$$

Additionally we know from the proof of Lemma 1 (cf., (30)) that $0 < \alpha - J \le c_n$ and so we conclude

$$\frac{|\alpha - \hat{J}|}{|\hat{c}_n|} \le \frac{|\alpha - J| + |J - \hat{J}|}{|\hat{c}_n - c_n + c_n|} \le \frac{c_n + |J - \hat{J}|}{c_n/2}.$$

Using (9) and

$$|\hat{\alpha} - \bar{\alpha}| = \left| \frac{\alpha - \hat{J}}{\hat{c}_n} - \frac{\alpha - J}{c_n} \right| \le \frac{|\alpha - \hat{J}|}{\hat{c}_n} \cdot \frac{|c_n - \hat{c}_n|}{c_n} + \frac{|\hat{J} - J|}{c_n}$$

we can conclude

$$|\hat{\alpha} - \bar{\alpha}| = O_{\mathbf{P}}\left(\frac{1}{(\delta_n + \eta_n) \cdot \sqrt{n_{\hat{\alpha}}}}\right) = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}}\right),$$

which implies the assertion.

Since most of the steps required an event of the form $\{|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}| \le \lambda \cdot \log(n) \cdot \delta_n\}$ to hold it is important to notice that by (7) for any such event the probability to hold tends to one for $n \to \infty$. This completes the proof. □

6.2 Proof of Theorem 3

Proof of Theorem 3.

a) At first we show that in the given setting conditions (5), (6) and (7) are fulfilled with

$$\eta_n = \frac{\log(n)^p}{n^{p/d}} \quad \text{and} \quad \delta_n = \tilde{C} \cdot \eta_n,$$

where $\tilde{C}$ only depends on the dimension d and the smoothness of m. Since m is Hölder-continuous we have for $n \geq 6^d$ and for any $x \in K_n$

$$\begin{aligned} |m_{n_1+n_2}(x) - m(x)| &\leq C \cdot \text{diam}(A_{n_1+n_2}(x))^p \leq C \cdot \text{diam}(A_{n_1}(x))^p \\ &\leq C \cdot \left(d \cdot 2 \cdot \frac{\log(n)}{\lfloor n_1^{1/d} \rfloor} \right)^p \leq C \cdot (12d)^p \cdot \frac{\log(n)^p}{n^{p/d}}. \end{aligned}$$

Hence condition (5) holds with $\tilde{C} = 2 \cdot C \cdot (12d)^p$ and $\delta_n = \tilde{C} \cdot \log(n)^p \cdot n^{-p/d}$. Furthermore the assumption that $N_n \cdot \mathbf{P}\{X \notin K_n\}$ tends to zero for $n \to \infty$ implies that $\mathbf{P}\{X_{n+1}, \dots, X_{n+N_n} \notin K_n\}$ tends to zero as n tends to infinity and so from Theorem 1 in Enss, Kohler, Krzyżak and Platz (2014) we conclude that

$$|\hat{q}^{(MC)}_{m_{n+n_2}(X),N_n,\alpha} - q_{m(X),\alpha}| \leq \tilde{C} \cdot \frac{\log(n)^p}{n^{p/d}} + 2 \cdot |\hat{q}^{(MC)}_{m(X),N_n,\alpha} - q_{m(X),\alpha}|. \tag{42}$$

By (3) this in turn implies

$$|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}| = |\hat{q}^{(MC)}_{m_{n_1+n_2}(X),N_n,\alpha} - q_{m(X),\alpha}| = O_{\mathbf{P}}\left(\frac{\log(n)^p}{n^{p/d}}\right), \tag{43}$$

whenever $N_n \geq n_1^{2p/d} \cdot \log(n)^{-2p}$, which implies assertion (7) with $\eta_n = \log(n)^p \cdot n^{-p/d}$ as well as condition (6). Now Theorem 1 yields the assertion.

b) Notice that the assertion follows from Theorem 1, if we can show that outside of an event, whose probability tends to zero for $n \to \infty$, (5) and (7) hold with

$$\delta_n = \frac{\log(n)^{p+3}}{n^{2p/d}} \quad \text{and} \quad \eta_n = const \cdot \frac{\log(n)^{p+3}}{n^{2p/d}}. \tag{44}$$

In the following let E_n be the event that $|\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - q_{m(X),\alpha}| \leq \log(n)^{p+1}/n^{p/d}$ holds and define

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

Now the verification that outside of an event, whose probability tends to zero for $n \to \infty$, (5) and

(7) hold with δ_n and η_n as in (44), follows in three steps.

In the first step of our proof we show that we have on E_n for n large enough and for all $x \in K_n \setminus C_{\text{critical},n}$

$$3 \cdot \log(n) \cdot \text{diam}(A(x))^p + \frac{\log(n)^{p+1}}{n^{p/d}} - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| \leq 0. \tag{45}$$

To do this we observe that triangle inequality and the Hölder-continuity of m imply that we have on E_n and for n large

$$\begin{aligned} |q_{m(X),\alpha} - m(x)| &\leq \frac{\log(n)^{p+1}}{n^{p/d}} + |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| + C \cdot \text{diam}(A(x))^p \\ &\leq 2 \cdot \frac{\log(n)^{p+1}}{n^{p/d}} + |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)|. \end{aligned}$$

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

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

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

$$\begin{aligned} \min \Bigg\{ & 2 \cdot \log(n) \cdot \text{diam}(A(x))^p, \\ & 3 \cdot \log(n) \cdot \text{diam}(A(x))^p - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| + \frac{\log(n)^{p+1}}{n_1^{p/d}} \Bigg\} \\ \leq & \kappa_5 \cdot \frac{\log(n)^{p+3}}{n^{2p/d}}, \end{aligned} \tag{46}$$

for some constant $\kappa_5 > 0$. By the results of step 1 we know that on E_n and for n sufficiently large (46) holds for all $x \in K_n \setminus C_{\text{critical},n}$. Hence as long as there exists some cube A in our partition such that $A \cap K_n \cap C_{\text{critical},n}$ is nonempty and which does not fulfill (46), our algorithm does not choose any of those cubes which are subsets of $K_n \setminus C_{\text{critical},n}$. By the assumption of part b) of Theorem 3 we know that $C_{\text{critical},n}$ is contained in a cube of side length less than $\kappa_1 \cdot \log(n)^{1+2/p}/n^{1/d}$. But after $n_2/2^d$ of the elements of the cubes of the partition, which have nonempty intersection with $K_n \cap C_{\text{critical},n}$ and which do not satisfy (46), are chosen we have for all $x \in K_n \cap C_{\text{critical},n}$

$$\text{diam}(A(x)) \leq c \cdot \frac{\log(n)^{1+2/p}}{n^{1/d}} \cdot \frac{1}{n_2^{1/d}} \leq c_2 \cdot \frac{\log(n)^{1+2/p}}{n^{2/d}},$$

which implies the assertion of the second step.

In the third step of the proof we actually show that (44) holds. By setting

$$\delta_n \quad := \quad \max_{x \in K_n} \min \Bigg\{ 2 \cdot \log(n) \cdot \operatorname{diam}(A(x))^p, \\ 3 \cdot \log(n) \cdot \operatorname{diam}(A(x))^p - |\hat{q}^{(MC)}_{m_{n_1}(X),N_n,\alpha} - m_n(x)| + \frac{\log(n)^{p+1}}{n_1^{p/d}} \Bigg\}$$

we conclude (as described in section 4) that for any $x \in K_n$

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

holds. So Theorem 1 in Enss, Kohler, Krzyżak and Platz (2014) implies that

$$\left|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}\right| = \left|\hat{q}^{(MC)}_{m_{n_1+n_2}(X),N_n,\alpha} - q_{m(X),\alpha}\right| = O_{\mathbf{P}}\left(\delta_n + \left|\hat{q}^{(MC)}_{m(X),N_n,\alpha} - q_{m(X),\alpha}\right|\right),$$

provided that

$$\mathbf{P}\left\{X_{n+1}, \ldots, X_{n+N_n} \in K_n\right\} \to 1 \quad (n \to \infty) \tag{47}$$

holds. (47) is a direct implication of (23), and by the same arguments as in the proof of Theorem 3 a) we see

$$\left|\tilde{q}_{m(X),n,\alpha} - q_{m(X),\alpha}\right| = O_{\mathbf{P}}\left(\delta_n + \frac{1}{\sqrt{N_n}}\right). \tag{48}$$

Now by step two we know that $\delta_n \leq c_6 \cdot \frac{\log(n)^{p+3}}{n^{2p/d}}$, whereas $1/\sqrt{N_n} \leq n^{-2p/d}$ by assumption, which completes the proof. □

7 Acknowledgment

The 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] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, **27**, pp. 642–669.

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

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

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

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

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

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

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

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

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

[24] 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.

[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] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability* **18 (3)**, pp. 1269–1283,

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

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

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

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

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

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