Adaptive estimation of quantiles in a simulation model *

Michael Kohler[1] and Adam Krzyżak[2,†]
[1] *Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@mathematik.tu-darmstadt.de*
[2] *Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8, email: krzyzak@cs.concordia.ca*

September 25, 2016

Abstract
Let X be an $\mathbb{R}^d$ valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and set $Y = m(X)$. Given a sample of (X, Y) of size n we consider the problem of estimating the quantile of Y of a given level $\alpha \in (0, 1)$. In this setting one promising idea is to apply a surrogate quantile estimate, where in the first step the sample of (X, Y) is used to construct a surrogate estimate of m and in the second step the quantile is estimated by using the surrogate estimate of m. The construction of the surrogate estimate usually depends on a proper choice of parameters, so each parameter leads to a different quantile estimate. Given finitely many such surrogate estimates, we consider the problem of choosing the best of them in the context of quantile estimation. A data-dependent way of selecting the optimal surrogate estimate is proposed. It is shown that the pointwise error of the resulting adaptive surrogate estimate is less than twice the maximal supremum norm error of the given surrogate estimates except on a set of measure less than $c_1 \cdot \log n/n$. Furthermore it is shown that this implies that the corresponding surrogate quantile estimate achieves the rate of convergence bounded by the sum of the minimal rate of convergence of the quantile estimates corresponding to the given surrogate estimates and a term of order $\log(n)/n$. The finite sample size behaviour of this quantile estimate is illustrated by applying it

*Running title: *Adaptive estimation of quantiles*
†Corresponding author. Tel: +1-514-848-2424 ext. 3007, Fax: +1-514-848-2830

to simulated data.

AMS classification: Primary 62G05; secondary 62G20.

Key words and phrases: Adaptation, quantile estimation, rate of convergence.

1 Introduction

Let Y be a real-valued random variable with cumulative distribution function (cdf.) G_Y, i.e., $G_Y(y) = \mathbf{P}\{Y \leq y\}$. For $\alpha \in (0,1)$ denote by

$$q_{Y,\alpha} = \min\{z \in \mathbb{R} \,:\, G_Y(z) \geq \alpha\}$$

the α-quantile of Y. A simple idea to estimate $q_{Y,\alpha}$ from a sample $Y_1, \ldots, Y_n$ of Y is to use $Y_1, \ldots, Y_n$ to compute the empirical cdf.

$$\hat{G}_n(y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i \leq y\}} \tag{1}$$

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

$$\hat{q}_{Y,n,\alpha} = \min\{z \in \mathbb{R} : \hat{G}_n(z) \geq \alpha\}. \tag{2}$$

Since $\hat{q}_{Y,n,\alpha}$ 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 Y has a density g which is continuous and positive at $q_{Y,\alpha}$ we have

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

This implies

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

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

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

In this paper we consider a simulation model of a complex technical system given by $Y = m(X)$, where X is a $\mathbb{R}^d$-valued random variable which describes uncertain parameters of the system and $m : \mathbb{R}^d \to \mathbb{R}$ is a function which describes how the outcome Y of the system depends on its (random)

parameters X. We consider the problem of estimating the α–quantile $q_{m(X),\alpha}$ of $Y = m(X)$. We assume that we are allowed to evaluate m at any point and that we have available random variables X_1, X_2, ... which are independent and identically distributed as X. However, since m describes a simulation model of a complex technical system, we assume that our function m is costly to evaluate, and hence we are interested in estimation of the quantile using as few evaluations of m as possible.

One simple way of estimating the above quantile would be to compute the data

$$\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\} = \{(X_1, m(X_1)), \ldots, (X_n, m(X_n))\}, \quad (4)$$

where X, X_1, X_2, ... are independent and identically distributed random variables, and to apply the simple order statistics estimate (2) to $Y_1 = m(X_1)$, ..., $Y_n = m(X_n)$. According to (3) this results in an estimate achieving the rate of convergence $1/\sqrt{n}$, where n is the number of evaluations of m needed.

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

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

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

More precisely, assume that besides data (4) we are also given the samples

$$X_{n+1}, \ldots, X_{n+N_n}$$

for some large $N_n \in \mathbb{N}$. The surrogate quantile estimate $\hat{q}_{m_N(X),N_n}$ first computes

$$m_n(X_{n+1}), \ldots, m_n(X_{n+N_n}),$$

and then produces the corresponding empirical cdf.

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

and subsequently estimates $q_{m(X),\alpha}$ by the plug-in estimate

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

Clearly, this quantile estimate depends crucially on the chosen surrogate model. 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). 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).

In Enss et al. (2016) the rate of convergence of surrogate quantile estimates has been analyzed. It has been shown that they achieve in case of a smooth function m better rates of convergence than the rate $1/\sqrt{n}$ of the simple order statistics estimate in (2). However, all of these estimates depend on parameters (which one can restrict without loss of generality to be contained in a finite set), and choosing one of them in an application immediately leads to the following problem: Given surrogate quantile estimates

$$\hat{q}_{m_{n,p}(X),N_n,\alpha} \quad (p \in \mathcal{P}_n)$$

(where $\mathcal{P}_n$ is a finite set of parameters) and the data (4), how to select the best quantile estimate, i.e., a how to select a $\hat{p} = \hat{p}(\mathcal{D}_n) \in \mathcal{P}_n$ such that

$$|\hat{q}_{m_{n,\hat{p}}(X),N_n,\alpha} - q_{m(X),\alpha}| = O_{\mathbf{P}}\left(\min_{p\in\mathcal{P}_n} |\hat{q}_{m_{n,p}(X),N_n,\alpha} - q_{m(X),\alpha}| + \epsilon_n\right) \quad (5)$$

for some small error term $\epsilon_n > 0$ satisfying $n^\delta \cdot \epsilon_n \to 0$ $(n \to \infty)$ for some $\delta > 1/2$.

In order to solve this problem we suggest in this paper a method which enables one to choose in a data–dependent way the best surrogate estimate from a list of finitely many surrogate estimates using the data (4). We

use this approach to define an algorithm which does not need any tuning parameter to be chosen in an application, which satisfies (5) with $\epsilon_n = (\log n)/n$ and which hence achieves the optimal rate of convergence up to some additional error of order $\log(n)/n$.

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

$$\|f\|_{\infty,A} = \sup_{x \in A} |f(x)|.$$

Let $p = k + s$ for some $k \in \mathbb{N}_0$ and $0 < s \leq 1$, and let $C > 0$. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called (p, C)-smooth, if for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$ the partial derivative $\frac{\partial^k f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$ exists and satisfies

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

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

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

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

The outline of the paper is as follows: The main results are formulated in Section 2 and proven in Section 4. Section 3 illustrates our adaptive choice of an quantile by applying it to simulated data.

2 Main results

2.1 Adaptive choice of a surrogate model

Let X, X_1, X_2, ..., be independent and identically distributed random variables with values in $\mathbb{R}^d$ and let $m : \mathbb{R}^d \to \mathbb{R}$ be measurable function. In this subsection we consider the problem of choosing a suitable surrogate model for $m(X)$. Here we assume that we have given a finite parameter set $\mathcal{P}_n$, surrogate estimates

$$m_{n,p} : \mathbb{R}^d \to \mathbb{R}$$

for each $p \in \mathcal{P}_n$ together with the data

$$\mathcal{D}_n = \{(X_1, m(X_1)), \ldots, (X_n, m(X_n))\}. \tag{6}$$

Our aim is to choose a parameter

$$\hat{p} = \hat{p}(\mathcal{D}_n) \in \mathcal{P}_n$$

such that the supremum norm error between $m(x)$ and $m_{n,\hat{p}}(x)$ on some given set $K_n \subseteq \mathbb{R}^d$ is small.

To do this, we propose to minimize the maximal absolute difference between $m_{n,p}(X_i)$ and $m(X_i)$ for those X_i that are contained in K_n, i.e., we set

$$\hat{p} = \arg\min_{p \in \mathcal{P}_n} \max_{\substack{i=1,\ldots,n, \\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)|.$$

Our next result shows that the $\mathbf{P}_X$–measure of the set of all $x \in K_n$, where the error of $m_{n,\hat{p}}$ is larger than two times the best supremum norm error

$$\min_{p \in \mathcal{P}_n} \|m_{n,p} - m\|_{\infty, K_n},$$

is small.

Theorem 1. *Let X, X_1, X_2, ..., be independent and identically distributed random variables with values in $\mathbb{R}^d$ and let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Let $\mathcal{P}_n$ be some finite set and assume that for each $p \in \mathcal{P}_n$ a surrogate estimate*

$$m_{n,p} : \mathbb{R}^d \to \mathbb{R}$$

of m is given. Define

$$m_n(x) = m_{n,\hat{p}}(x) \quad (x \in \mathbb{R}^d)$$

where

$$\hat{p} = \arg\min_{p \in \mathcal{P}_n} \max_{\substack{i=1,\ldots,n, \\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)|. \tag{7}$$

Assume that

$$\frac{|\mathcal{P}_n|}{n^r} \to 0 \quad (n \to \infty)$$

for some $r > 0$.

Then we have outside of an event, whose probability tends to zero for $n \to \infty$,

$$\mathbf{P}_X \left(\left\{ x \in K_n \, : \, |m_n(x) - m(x)| > 2 \cdot \min_{p \in \mathcal{P}_n} \|m_{n,p} - m\|_{\infty, K_n} \right\} \right) \leq r \cdot \frac{\log n}{n}.$$

Remark 1. In case that the estimates $m_{n,p}$ itselves are based on data of the form (6), i.e., in case that they are based on observations of m at random points, we can split the data (6) into two halves in order to define an adaptive estimate which uses alltogether only n evaluations of m. To do this, set $n_l = \lfloor n/2 \rfloor$, let $K_n \subseteq \mathbb{R}^d$ and let $\mathcal{P}_n$ be a finite set of parameters. Assume that for each $p \in \mathcal{P}_n$ an estimate

$$m_{n_l,p}(\cdot) = m_{n_l,p}(\cdot, (X_1, m(X_1)), \ldots, (X_{n_l}, m(X_{n_l}))) : \mathbb{R}^d \to \mathbb{R}$$

of m is given. Set

$$m_n(x) = m_{n_l,\hat{p}}(x) \quad (x \in \mathbb{R}^d),$$

where

$$\hat{p} = \arg\min_{p \in \mathcal{P}_n} \max_{\substack{i=n_l+1,\ldots,n, \\ X_i \in K_n}} |m_{n_l,p}(X_i) - m(X_i)|.$$

As in the proof of Theorem 1 it is possible to show

$$\mathbf{P}_X \left(\left\{ x \in K_n \, : \, |m_n(x) - m(x)| > 2 \cdot \min_{p \in \mathcal{P}_n} \|m_{n,p} - m\|_{\infty, K_n} \right\} \Big| X_1^{n_l} \right)$$
$$\leq 2 \cdot r \cdot \frac{\log n}{n},$$

where $X_1^{n_l} = (X_1, \ldots, X_{n_l})$.

Remark 2. The estimate in Remark 1 uses splitting of the data in order to choose the optimal parameter. But at the very end this parameter is applied only with a sample of half the original size. One way to circumvent is, is to use a kind of cross-validation instead. In the sequel we define a k–fold cross-validation to choose our parameter, where $2 \leq k \leq n$ is fixed (in Section 3 below we will use $k = 5$). To do this, set

$$n_l = \left\lfloor l \cdot \frac{n}{k} \right\rfloor, \quad (l \in \{0, \ldots, k\}),$$

and set

$$\mathcal{D}_{n,l} = \left\{ (X_1, m(X_1)), \ldots, (X_{n_l}, m(X_{n_l})), (X_{n_{l+1}+1}, m(X_{n_{l+1}+1})), \ldots, (X_n, Y_n) \right\}$$

$(l \in \{1, \ldots, k\})$. Then the k–fold cross-validation choice of our parameter $p \in \mathcal{P}_n$ is

$$\hat{p} = \arg\min_{p \in \mathcal{P}_n} \max_{l=1,\ldots,k} \max_{\substack{i=n_l+1,\ldots,n_{l+1}, \\ X_i \in K_n}} \left| m_{n-(n_{l+1}-n_l),p}(X_i, \mathcal{D}_{n,l}) - m(X_i) \right|,$$

and the corresponding surrogate estimate is

$$m_n(x) = m_{n,\hat{p}}(x, \mathcal{D}_n) \quad (x \in \mathbb{R}^d).$$

The finite sample size performance of this estimate will be analyzed in Section 3 using simulated data. Whether a similar bound as the one in Theorem 1 also holds for this estimate is an open problem.

2.2 A general result for surrogate quantile estimates

Let X, X_1, X_2, ... be independent and identically distributed random variables with values in $\mathbb{R}^d$, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function, and for $\alpha \in (0,1)$ define the α–quantile of $m(X)$ by

$$q_{m(X),\alpha} = \inf\{y \in \mathbb{R} \, : \, G_{m(X)}(y) \geq \alpha\}, \tag{8}$$

where

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

is the cdf. of $m(X)$. For $n \in \mathbb{N}$ let $K_n \subseteq \mathbb{R}^d$ and let $\mathcal{P}_n$ be a finite set of parameters. Assume that for each $p \in \mathcal{P}_n$ an estimate

$$m_{n,p}(\cdot) : \mathbb{R}^d \to \mathbb{R} \tag{9}$$

of m is given. Set

$$m_n(x) = m_{n,\hat{p}}(x) \quad (x \in \mathbb{R}^d), \tag{10}$$

where

$$\hat{p} = \arg\min_{p \in \mathcal{P}_n} \max_{\substack{i=1,\ldots,n, \\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)| \tag{11}$$

is the parameter, for which the maximal error on the testing data contained in K_n is minimal. We use m_n to define a corresponding surrogate quantile estimate by

$$\hat{q}_{m_n(X),N_n,\alpha} = \inf\{y \in \mathbb{R} \, : \, \hat{G}_{m_n(X),N_n}(y) \geq \alpha\}, \tag{12}$$

where

$$\hat{G}_{m_n(X),N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{m_n(X_{n+i}) \leq y\}} \quad (y \in \mathbb{R}). \tag{13}$$

Then the following result holds.

Theorem 2. *Let X, X_1, X_2, ... be independent and indentically distributed random variables, let $m : \mathbb{R}^d \to \mathbb{R}$ be measurable, let $\alpha \in (0,1)$ and let $q_{m(X),\alpha}$ be the α–quantile of $m(X)$ defined by (8). For $n \in \mathbb{N}$ let $K_n \subseteq \mathbb{R}^d$ and let $\mathcal{P}_n$ be a finite set of parameters such that for each $p \in \mathcal{P}_n$ an estimate (9) is given, and define the corresponding quantile estimate by (10)-(13).*

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

$$\min_{p \in \mathcal{P}_n} \|m_{n,p} - m\|_{\infty,K_n} \to 0 \quad (n \to \infty), \tag{14}$$

$$\mathbf{P}_X\left(K_n^c\right) \to 0 \quad (n \to \infty), \tag{15}$$

$$N_n \to \infty \quad (n \to \infty) \tag{16}$$

and

$$\frac{|\mathcal{P}_n|}{n^r} \to 0 \quad (n \to \infty) \tag{17}$$

for some $r > 0$. Then

$$\begin{aligned} &\left|\hat{q}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha}\right| \\ &= O_{\mathbf{P}}\left(\frac{1}{\sqrt{N_n}} + \min_{p \in \mathcal{P}_n} \|m_{n,p} - m\|_{\infty,K_n} + \mathbf{P}_X\left(K_n^c\right) + \frac{\log n}{n}\right). \end{aligned}$$

2.3 Application to surrogate spline quantile estimates

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

To do this, we choose $\delta > 1$ and set $\gamma_n = (\log n)^\delta$. Next we define a spline approximand which approximates m on $[-\gamma_n, \gamma_n]^d$. In order to do this, we introduce polynomial splines, i.e., sets of piecewise polynomials satisfying a global smoothness condition, and a corresponding B-spline basis consisting of basis functions with compact support. Here our presentation is based on Kohler (2014), which in turn is an extension of the material presented in Chapters 14 and 15 of Györfi et al. (2002) to the case $d > 2$.

Choose $K \in \mathbb{N}$ and $M \in \mathbb{N}_0$, and set $u_k = k \cdot \gamma_n / K$ $(k \in \mathbb{Z})$. 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 $\mathrm{supp}(B_{k,M}) = [u_k, u_{k+M+1}]$. 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

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

(so-called hat-function). The general definition of $B_{k,M}$ can be found, e.g., in de Boor (1978), or in Section 14.1 of Györfi et al. (2002). These B-splines are basis functions of sets of univariate piecewise polynomials of degree M, where the piecewise polynomials are globally $(M-1)$–times continuously differentiable and where the M-th derivative of the functions have jump points only at the knots u_l ($l \in \mathbb{Z}$).

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

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

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 $[-\gamma_n, \gamma_n]^d$, i.e., we set

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

It can be shown by using standard arguments from spline theory, that the functions in $\mathcal{S}_{K,M}$ are in each component $(M-1)$-times continuously differentiable, that they are equal to a (multivariate) polynomial of degree less than or equal to M (in each component) on each rectangle

$$[u_{k_1}, u_{k_1+1}) \times \cdots \times [u_{k_d}, u_{k_d+1}) \quad (\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d), \tag{18}$$

and that they vanish outside of the set

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

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

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

where

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

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

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

and

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

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

$$|Q_{\mathbf{k}} f| \leq c_2 \cdot \|f\|_{\infty,[u_{k_1}, u_{k_1+M+1}] \times \cdots \times [u_{k_d}, u_{k_d+M+1}]} \tag{19}$$

for any $\mathbf{k} \in \{-M, -M+1, \ldots, K-1\}^d$, any $f : [-\gamma_n, \gamma_n]^d \to \mathbb{R}$ and some universal constant c_2, and such that Q reproduces polynomials of degree M or less (in each component) on $[-\gamma_n, \gamma_n]^d$, i.e., for any multivariate polynomial $p : \mathbb{R}^d \to \mathbb{R}$ of degree M or less (in each component) we have

$$(Qp)(x) = p(x) \quad (x \in [-\gamma_n, \gamma_n]^d) \tag{20}$$

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

Next we want to use such a quasi interpoland as surrogate estimate for m. To do this, we need to choose the degree $M \in \mathbb{N}$ and the number of knots $K \in \mathbb{N}$. For fixed values of K and M we need to evaluate m at the points

$$\left(\frac{j_1}{M \cdot K} \cdot \gamma_n, \ldots, \frac{j_d}{M \cdot K} \cdot \gamma_n \right) \quad (j_1, \ldots, j_d \in \{-M \cdot K, -M \cdot K+1, \ldots, M \cdot K\}) \tag{21}$$

in order to compute the above quasi interpoland. If we restrict the number of point evaluation to be at most $\lceil n/2 \rceil$, then our choice of M and K needs to fulfill the condition

$$(2 \cdot M \cdot K + 1)^d \leq \lceil n/2 \rceil.$$

Consequently there is a tradeoff between the choice of M and that of K, which should be both large in view of approximation power of the spline interpoland. In the sequel we use our method of Subsection 2.1 in order to choose both values in a data dependent way.

To do this, we choose $l \in \mathbb{N}$ maximal such that

$$\left(2 \cdot 2^l + 1\right)^d \leq \lceil n/2 \rceil$$

and set $K_{max} = 2^l$, hence

$$K_{max} \approx c_3 \cdot n^{1/d}.$$

Then we observe the function values of m at all the $(2K_{max}+1)^d \le \lceil n/2 \rceil$ points of the form

$$\left(\frac{j_1}{K_{max}} \cdot \gamma_n, \dots, \frac{j_d}{K_{max}} \cdot \gamma_n \right) \quad (j_1, \dots, j_d \in \{-K_{max}, -K_{max}+1, \dots, K_{max}\}). \tag{22}$$

Next we set

$$\mathcal{P}_n = \left\{ \left(2^M, \frac{K_{max}}{2^M} \right) \,:\, M \in \{1, \dots, \lceil \log \log n \rceil\} \right\}$$

and for $p \in \mathcal{P}_n$ we let $m_{n,p}$ be the quasi interpoland with parameters M and K specified by p. Since

$$2^M \cdot \frac{K_{max}}{2^M} = K_{max}$$

this quasi interpoland also does only need function values of m at the points (22).

The following result holds.

Corollary 1. *Let X be an $\mathbb{R}^d$-valued random variable, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and let $\alpha \in (0,1)$. Assume that $m(X)$ has a density which is continuous and positive at $q_{m(X),\alpha}$, that m is (p,C)-smooth for some $p > 0$ and some $C > 0$ and that*

$$\mathbf{E} \exp\left(c_4 \cdot \|X\|\right) < \infty \tag{23}$$

for some $c_4 > 0$. Let $\delta > 1$, set $\gamma_n = (\log n)^\delta$, and define the quasi interpolands $m_{n,p}$ for $p \in \mathcal{P}_n$ as above. Set

$$m_n(x) = m_{n,\hat{p}}(x) \quad (x \in \mathbb{R}^d),$$

where

$$\hat{p} = \arg\min_{p \in \mathcal{P}_n} \max_{\substack{i=1,\dots,\lfloor n/2 \rfloor, \\ X_i \in [-\gamma_n, \gamma_n]^d}} |m_n(X_i) - m(X_i)|,$$

and define $\hat{q}_{m_n(X),N_n,\alpha}$ by (12) and (13), where $N_n = n^2$. Then

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

Proof. By Theorem 2 we get

$$\left| \hat{q}_{m_n(X),N_n,\alpha} - q_{m(X),\alpha} \right|$$

$$= O_{\mathbf{P}}\left(\min_{p\in\mathcal{P}_n} \|m_{n,p} - m\|_{\infty,[-\gamma_n,\gamma_n]^d} + \mathbf{P}\left\{X \notin [-\gamma_n,\gamma_n]^d\right\} + \frac{\log n}{n}\right).$$

The definition of our spline approximand and the (p, C)-smoothness of m imply that for $2^M \geq \lfloor p \rfloor$ (where M does not depend on n) and $K = K_{max}/2^M$ we have

$$\|m_{n,(2^M,K)} - m\|_{\infty,[-\gamma_n,\gamma_n]^d} \leq c_5 \cdot (\log n)^{\delta \cdot p} \cdot n^{-p/d}$$

(cf., e.g., proof of Theorem 1 in Kohler (2014)). Furthermore, Markov inequality together with assumption (23) imply

$$\mathbf{P}\left\{X \notin [-\gamma_n,\gamma_n]^d\right\} \quad \leq \quad \mathbf{P}\left\{\|X\| > \gamma_n\right\} \leq \frac{\mathbf{E}\exp\left(c_4 \cdot \|X\|\right)}{\exp(c_4 \cdot (\log n)^{\delta})} \leq c_5 \cdot \frac{\log n}{n}.$$

The proof is complete. □

Remark 3. The spline surrogate quantile estimate is constructed such that its computation needs at most n evaluation of the function m. Compared with the simple order statistics it achieves a better rate of convergence whenever m is (p, C)–smooth for some $p > d/2$. Here the definition of the estimate does not depend on p.

Remark 4. It is an open problem how to define adaptive surrogate quantile estimates capable of achieving the rates of convergence faster than $1/n$ in case of a very smooth function m.

3 Application to simulated data

In this section we compare the finite sample size behaviour of our newly proposed adaptive surrogate quantile estimates with various other quantile estimates.

For the surrogate estimate we use a smoothing spline (as implemented in the routine *Tps()* in *R*). 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 Subsection 2.3, but is easier to implement. The routine *Tps()* in *R* allows to specify the degree of freedom *df* of the fitted surface, which is the parameter which we choose automatically in order to get an adaptive surrogate estimate. To do this we consider four different possibilities: For the first one we use the generalized L_2 cross–validation as implemented in *R*. The second one simply chooses the maximal degree of freedom (which is the sample size), which may also give reasonable results since in our data the function is observed without error. The remaining

two methods are the new ones proposed in this paper: the third one uses splitting of the sample as explained in Remark 1, and the fourth one k–fold cross-validation with $k = 5$ as described in Remark 2.

We use these four different surrogate estimates in two different ways in order to define quantile estimates. The first way is the surrogate quantile estimate $\hat{q}_{m_n(X),N_n,\alpha}$ as introduced in Section 1. The second way uses the surrogate estimate as a control variate as explained in Section 2 in Cannamela, Garnier and Ioss (2008). Here the true quantile of the control variate is replaced by a Monte Carlo estimate of the quantile. Each time the Monte Carlo estimates are based on $N_n = 50,000$ evaluations of the surrogate estimate.

We compare the resulting 8 different quantile estimates based on surrogate estimates with the simple order statistics defined by (1) and (2). So we are comparing alltogether 9 different quantile estimates. The first quantile estimate (*est. 1*) is the estimate based on the simple order statistics. The second (*est. 2*) and third (*est. 3*) quantile estimates are the surrogate quantile estimate of Section 1 and the control variate quantile estimate combined with the smoothing spline with parameter *df* chosen by generalized L_2 crossvalidation, resp. In the same way we define *est. 4* and *est. 5* by using a smoothing spline with parameter *df* chosen as the sample size, *est. 6* and *est. 7* by using a smoothing spline with parameter *df* chosen by splitting of the sample, and *est. 8* and *est. 9* by using a smoothing spline with parameter *df* chosen by 5–fold cross-validation.

We compare these nine quantile estimates in three different models, where in each model we estimate a quantile of level $\alpha = 0.95$. In each model X is chosen as a d-dimensional random variable with standard normal distribution, where in case $d > 1$ all d components of X are independent random variables with standard normal distribution. In the first model the dimension of X is $d = 1$, and we allow $n \in \{20, 200, 1000\}$ evaluations of m. In the second and third model the dimension of X is $d = 4$ and we choose $n \in \{80, 300, 1000\}$.

Since the results of our simulation depend on the randomly occurring data points, we repeat the whole procedure 100 times with independent realizations of the occurring random variables and report the medians and the interquartile ranges of the relative errors of the quantile estimates (more precisely, of the absolute values of the difference between the quantile estimates and the real quantile divided by the real quantile).

For the first model we choose

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

hence $m(X)$ has lognormal distribution. The errors of nine different esti-

mates occurring in 100 simulations for each sample size are presented in Tables 1 and 2.

n	est. 1	est. 2	est. 4	est. 6.	est. 8
20	0.264	0.015	0.015	0.057	0.015
	(0.246)	(0.024)	(0.024)	(0.135)	(0.026)
200	0.082	0.006	0.006	0.006	0.006
	(0.099)	(0.007)	(0.007)	(0.007)	(0.007)
1000	0.05	0.006	0.006	0.006	0.006
	(0.055)	(0.008)	(0.008)	(0.008)	(0.008)

Table 1: Simulation results for model 1 and three different sample sizes. Compared are the simple order statistics with the four surrogate quantile estimates based on different methods for choosing the parameter of the surrogate. Reported are the medians of the relative absolute errors of the estimates (and in brackets their interquartilerange) in 100 independent simulations.

n	est. 1	est. 2	est. 4	est. 6.	est. 8
20	0.264	0.191	0.191	0.228	0.191
	(0.246)	(0.218)	(0.218)	(0.22)	(0.218)
200	0.082	0.035	0.035	0.035	0.035
	(0.099)	(0.044)	(0.044)	(0.044)	(0.044)
1000	0.05	0.01	0.01	0.01	0.01
	(0.055)	(0.015)	(0.015)	(0.015)	(0.015)

Table 2: Simulation results for model 1 and three different sample sizes. Compared are the simple order statistics with the controlled variate quantile estimates based on different methods for choosing the parameter of the surrogate. Reported are the medians of the relative absolute errors of the estimates (and in brackets their interquartilerange) in 100 independent simulations.

Looking at the errors reported in Tables 1 and 2 we see first that all quantile estimates based on surrogate estimates clearly outperform the simple order statistics estimate, that second the surrogate quantile estimates outperform the estimates using the surrogate model as control variate, and that third the newly proposed cross-validation method for choosing the parameter of the surrogate works similar to the standard methods (generalized

cross-validation or choosing the parameter as maximal value). The splitting of the sample yields for $n = 20$ higher errors than the other surrogate quantile estimates, we guess this is due to the fact that the estimate is finally applied only to the half data size. The results for the control variate estimate are similar.

In the second model we use dimension $d = 4$ and choose

$$m(x) = \frac{1}{1 + \|x\|^2} \quad (x \in \mathbb{R}^4).$$

The errors of nine different estimates occurring in 100 simulations for each sample size are presented in Tables 3 and 4. This time our newly proposed cross-validation method is not as good as the standard methods for small sample sizes, but its performance is similar for large sample sizes. Otherwise the results are similar.

n	est. 1	est. 2	est. 4	est. 6.	est. 8
80	0.085	0.024	0.023	0.054	0.04
	(0.1)	(0.032)	(0.025)	(0.089)	(0.053)
300	0.046	0.005	0.005	0.011	0.005
	(0.054)	(0.005)	(0.005)	(0.012)	(0.005)
1000	0.022	0.004	0.004	0.004	0.003
	(0.025)	(0.004)	(0.004)	(0.004)	(0.004)

Table 3: Simulation results for model 2 and three different sample sizes. Compared are the simple order statistics with the four surrogate quantile estimates based on different methods for choosing the parameter of the surrogate. Reported are the medians of the relative absolute errors of the estimates (and in brackets their interquartilerange) in 100 independent simulations.

In both models above our newly proposed method does not outperform the generalized cross-validation or the method of choosing the degree of freedom *df* simply very large. However, at least our newly proposed cross-validation method does not perform worse than the other methods for large sample sizes. We believe that this is due to the fact that in our models above m is a very simple function of $\|x\|$, and we show next that for a slightly more complex function the situation is different and that in this case our newly proposed cross-validation method outperforms the other methods for all sample sizes.

n	est. 1	est. 2	est. 4	est. 6.	est. 8
80	0.085	0.039	0.039	0.086	0.053
	(0.1)	(0.052)	(0.053)	(0.104)	(0.067)
300	0.046	0.013	0.013	0.016	0.013
	(0.054)	(0.015)	(0.015)	(0.017)	(0.014)
1000	0.022	0.005	0.005	0.006	0.005
	(0.025)	(0.006)	(0.006)	(0.006)	(0.007)

Table 4: Simulation results for model 2 and three different sample sizes. Compared are the simple order statistics with the controlled variate quantile estimates based on different methods for choosing the parameter of the surrogate. Reported are the medians of the relative absolute errors of the estimates (and in brackets their interquartilerange) in 100 independent simulations.

In the third model we choose

$$m(x) = \begin{cases} 1 & , \|x\| \leq 3, \\ 1 + 10 \cdot \sqrt{\|x\|^2 - 9}) & , \text{elsewhere} \end{cases} \qquad (x \in \mathbb{R}^4).$$

The errors of nine different estimates occurring in 100 simulations for each sample size are presented in Tables 5 and 6.

n	est. 1	est. 2	est. 4	est. 6.	est. 8
80	0.875	0.34	0.344	0.343	0.288
	(0.548)	(0.241)	(0.242)	(0.329)	(0.229)
300	0.301	0.238	0.24	0.26	0.184
	(0.552)	(0.122)	(0.12)	(0.214)	(0.138)
1000	0.22	0.179	0.18	0.177	0.136
	(0.233)	(0.055)	(0.055)	(0.094)	(0.059)

Table 5: Simulation results for model 3 and three different sample sizes. Compared are the simple order statistics with the four surrogate quantile estimates based on different methods for choosing the parameter of the surrogate. Reported are the medians of the relative absolute errors of the estimates (and in brackets their interquartilerange) in 100 independent simulations.

Looking at the errors reported in Tables 5 and 6 we see that again all quantile estimates based on surrogate estimates clearly outperform the sim-

n	est. 1	est. 2	est. 4	est. 6.	est. 8
80	0.875	0.349	0.357	0.467	0.33
	(0.548)	(0.529)	(0.663)	(0.621)	(0.524)
300	0.301	0.195	0.196	0.191	0.171
	(0.552)	(0.2)	(0.195)	(0.355)	(0.194)
1000	0.22	0.186	0.186	0.159	0.122
	(0.233)	(0.104)	(0.088)	(0.148)	(0.12)

Table 6: Simulation results for model 3 and three different sample sizes. Compared are the simple order statistics with the controlled variate quantile estimates based on different methods for choosing the parameter of the surrogate. Reported are the medians of the relative absolute errors of the estimates (and in brackets their interquartilerange) in 100 independent simulations.

ple order statistics estimate but that this time the estimates using the surrogate model as control variate slightly outperform the surrogate quantile estimates. We see furthermore that for model 3 our newly proposed cross-validation method clearly outperforms the other methods, and that for the very large sample size also the splitting of the sample outperforms the generalized cross-validation or the method in which we choose the degree of freedom *df* simply very large.

Summarizing the above simulation results we see that our newly proposed cross-validation method never works worse than the other methods in case of large sample sizes but for model 3 clearly outperforms the other methods for all sample sizes. Here it is useful not only in combination with surrogate quantile estimates but also with quantile estimates using surrogate models as control variate.

4 Proofs

4.1 Proof of Theorem 1

We need the following in the proof of Theorem 1.

Lemma 1. *Let X, X_1, ..., X_n be independent and identically distributed random variables with values in $\mathbb{R}^d$, let $m : \mathbb{R}^d \to \mathbb{R}$ be measurable function, let $K_n \subseteq \mathbb{R}^d$ and let $\mathcal{P}_n$ be a finite set of parameters satisfying*

$$\frac{|\mathcal{P}_n|}{n^r} \to 0 \quad (n \to \infty) \tag{24}$$

for some $r > 0$. Assume that for each $p \in \mathcal{P}_n$ an estimate

$$m_{n,p}(\cdot) = m_{n,p}(\cdot) : \mathbb{R}^d \to \mathbb{R}$$

of m is given. Set

$$\hat{p} = \arg\min_{p \in \mathcal{P}_n} \max_{\substack{i=1,\dots,n, \\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)|, \tag{25}$$

$$m_n(x) = m_{n,\hat{p}}(x) \quad (x \in \mathbb{R}^d) \tag{26}$$

and

$$\hat{\beta} = \max_{\substack{i=1,\dots,n, \\ X_i \in K_n}} |m_n(X_i) - m(X_i)| = \min_{p \in \mathcal{P}_n} \max_{\substack{i=1,\dots,n, \\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)|.$$

Then we have outside of an event, whose probability tends to zero for $n \to \infty$,

$$\mathbf{P}_X \left(\left\{ x \in K_n \, : \, |m_n(x) - m(x)| > 2 \cdot \hat{\beta} \right\} \right) \leq r \cdot \frac{\log n}{n}.$$

Proof. Using (25), (26) and the definition of $\hat{\beta}$ we get

$$\begin{aligned}
&\mathbf{P}_X \left(\left\{ x \in K_n \, : \, |m_n(x) - m(x)| > 2 \cdot \hat{\beta} \right\} \right) \\
&= \mathbf{P}_X \left(\left\{ x \in K_n \, : \, |m_{n,\hat{p}}(x) - m(x)| > 2 \cdot \max_{\substack{i=1,\dots,n, \\ X_i \in K_n}} |m_{n,\hat{p}}(X_i) - m(X_i)| \right\} \right) \\
&\leq \max_{p \in \mathcal{P}_n} \mathbf{P}_X \left(\left\{ x \in K_n \, : \, |m_{n,p}(x) - m(x)| > 2 \cdot \hat{\beta}_p \right\} \right),
\end{aligned}$$

where

$$\hat{\beta}_p = \max_{\substack{i=1,\dots,n, \\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)| \quad (p \in \mathcal{P}_n).$$

Hence it suffices to show that we have outside of an event, whose probability tends to zero for $n \to \infty$,

$$\max_{p \in \mathcal{P}_n} \mathbf{P}_X \left(\left\{ x \in K_n \, : \, |m_{n,p}(x) - m(x)| > 2 \cdot \hat{\beta}_p \right\} \right) \leq r \cdot \frac{\log n}{n}. \tag{27}$$

Let $\mathcal{P}_n^*$ be the subset of $\mathcal{P}_n$ containing all those $p \in \mathcal{P}_n$ for which

$$\mathbf{P}_X \left(\{ x \in K_n \, : \, |m_{n,p}(x) - m(x)| > 0 \} \right) > r \cdot \frac{\log n}{n}$$

holds. Since

$$\begin{aligned}&\mathbf{P}_X\left(\left\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > 2 \cdot \hat{\beta}_p\right\}\right)\\ &\leq \mathbf{P}_X\left(\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > 0\}\right),\end{aligned}$$

inequality (27) is implied by

$$\max_{p \in \mathcal{P}_n^*} \mathbf{P}_X\left(\left\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > 2 \cdot \hat{\beta}_p\right\}\right) \leq r \cdot \frac{\log n}{n}. \tag{28}$$

For $p \in \mathcal{P}_n^*$ choose $\beta_p > 0$ such that

$$\begin{aligned}&\mathbf{P}_X\left(\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > 2 \cdot \beta_p\}\right)\\ &\leq r \cdot \frac{\log n}{n} \leq \mathbf{P}_X\left(\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > \beta_p\}\right).\end{aligned} \tag{29}$$

Here such a choice of β_p is possible, since

$$\beta \mapsto \mathbf{P}_X\left(\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > \beta\}\right)$$

is monotonically decreasing on $\mathbb{R}_+$, converges to zero for $\beta \to \infty$ and converges to a value greater than $r \cdot (\log n)/n$ for $\beta \to 0$. Hence if we start with some $\beta_{p,0} > 0$ such that

$$\mathbf{P}_X\left(\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > \beta_{p,0}\}\right) > r \cdot \frac{\log n}{n}$$

and set successively $\beta_{p,i+1} = 2 \cdot \beta_{p,i}$ for $i \in \mathbb{N}_0$, we will finally find some value β_p such that (29) holds.

Set

$$S_p = \{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > \beta_p\} \quad (p \in \mathcal{P}_n^*).$$

If for some $p \in \mathcal{P}_n^*$ and $j \in \{1, \ldots, n\}$ we have

$$X_j \in S_p,$$

then

$$\hat{\beta}_p = \max_{\substack{i=1,\ldots,n,\\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)| \geq |m_{n,p}(X_j) - m(X_j)| > \beta_p,$$

which implies together with the choice of β_p

$$\mathbf{P}_X\left(\left\{x \in K_n \,:\, |m_{n,p}(x) - m(x)| > 2 \cdot \hat{\beta}_p\right\}\right)$$

$$\leq \mathbf{P}_X\left(\{x \in K_n : |m_{n,p}(x) - m(x)| > 2 \cdot \beta_p\}\right)$$
$$\leq r \cdot \frac{\log n}{n}.$$

Hence on the event

$$A_n = \cap_{p \in \mathcal{P}_n^*} \{X_i \in S_p \text{ for some } 1 \leq i \leq n\}$$

the condition (28) holds, and thus it suffices to show

$$\mathbf{P}(A_n^c) \to 0 \quad (n \to 0).$$

To this end, we observe

$$\begin{aligned}
\mathbf{P}(A_n^c) &= \mathbf{P}\left(\cup_{p \in \mathcal{P}_n^*} \{X_i \notin S_p \text{ for all } 1 \leq i \leq n\}\right) \\
&\leq \sum_{p \in \mathcal{P}_n^*} \mathbf{P}\{X_i \notin S_p \text{ for all } 1 \leq i \leq n\} \\
&\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n^*} \mathbf{P}\{X_i \notin S_p \text{ for all } 1 \leq i \leq n\}.
\end{aligned}$$

For arbitrary $p \in \mathcal{P}_n^*$ the independence and identically distribution of our data imply

$$\begin{aligned}
\mathbf{P}\{X_i \notin S_p \text{ for all } 1 \leq i \leq n\} &= \prod_{i=1}^{n} (1 - \mathbf{P}_X(S_p)) \\
&\leq \left(1 - r \cdot \frac{\log n}{n}\right)^n,
\end{aligned}$$

where the last inequality follows from our definition of S_p and (29). Consequently,

$$\begin{aligned}
\mathbf{P}(A_n^c) &\leq |\mathcal{P}_n| \cdot \left(1 - r \cdot \frac{\log n}{n}\right)^n \leq |\mathcal{P}_n| \cdot \exp\left(-\frac{r \cdot n \cdot \log(n)}{n}\right) \\
&\leq \frac{|\mathcal{P}_n|}{n^r}.
\end{aligned}$$

Finally (24) implies the assertion. □

Proof of Theorem 1. Since

$$\begin{aligned}
\hat{\beta} &= \max_{\substack{i=1,\ldots,n, \\ X_i \in K_n}} |m_n(X_i) - m(X_i)| \\
&= \min_{p \in \mathcal{P}_n} \max_{\substack{i=1,\ldots,n, \\ X_i \in K_n}} |m_{n,p}(X_i) - m(X_i)|
\end{aligned}$$

$$\leq \min_{p\in\mathcal{P}_n} \|m_{n,p} - m\|_{\infty,K_n},$$

we have

$$\begin{aligned}
&\mathbf{P}_X\left(\left\{x \in K_n \,:\, |m_n(x) - m(x)| > 2 \cdot \min_{p\in\mathcal{P}_n} \|m_{n,p} - m\|_{\infty,K_n}\right\}\right) \\
&\leq \mathbf{P}_X\left(\left\{x \in K_n \,:\, |m_n(x) - m(x)| > 2 \cdot \hat{\beta}\right\}\right).
\end{aligned}$$

Application of Lemma 1 yields the assertion. □

4.2 Proof of Theorem 2

In the proof we will apply the following lemmatas.

Lemma 2. *For $y \in \mathbb{R}$ set $G_{m_n(X)}(y) = \mathbf{P}\{m_n(X) \leq y\}$, and for a set $A_n \subseteq \mathbb{R}^d$ set*

$$\beta_n = \sup_{x\in A_n} |m_n(x) - m(x)|.$$

Then the following inequality holds for all $y \in \mathbb{R}$:

$$G_{m(X)}(y - \beta_n) - \mathbf{P}_X\left(A_n^c\right) \leq G_{m_n(X)}(y) \leq G_{m(X)}(y + \beta_n) + \mathbf{P}_X\left(A_n^c\right)$$

Proof. By the definition of β_n we have for all $x \in A_n$

$$m(x) \leq y - \beta_n \quad \Longrightarrow \quad m_n(x) \leq y$$

and

$$m_n(x) \leq y \quad \Longrightarrow \quad m(x) \leq y + \beta_n.$$

This implies

$$\begin{aligned}
G_{m(X)}(y - \beta_n) &\leq \mathbf{P}_X\left(\{x \in A_n \,:\, m(x) \leq y - \beta_n\}\right) + \mathbf{P}_X\left(A_n^c\right) \\
&\leq \mathbf{P}_X\left(\left\{x \in \mathbb{R}^d \,:\, m_n(x) \leq y\right\}\right) + \mathbf{P}_X\left(A_n^c\right) \\
&= G_{m_n(X)}(y) + \mathbf{P}_X\left(A_n^c\right)
\end{aligned}$$

and

$$\begin{aligned}
G_{m_n(X)}(y) &\leq \mathbf{P}_X\left(\{x \in A_n \,:\, m_n(x) \leq y\}\right) + \mathbf{P}_X\left(A_n^c\right) \\
&\leq \mathbf{P}_X\left(\left\{x \in \mathbb{R}^d \,:\, m(x) \leq y + \beta_n\right\}\right) + \mathbf{P}_X\left(A_n^c\right) \\
&= G_{m(X)}(y + \beta_n) + \mathbf{P}_X\left(A_n^c\right).
\end{aligned}$$

□

Our next lemma is the consequence of the Dvoretzky-Kiefer-Wolfowitz inequality.

Lemma 3.

$$\sup_{y\in\mathbb{R}}\left|G_{m_n(X)}(y)-\hat{G}_{m_n(X),N_n}(y)\right| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{N_n}}\right).$$

Proof. By the Dvoretzky-Kiefer-Wolfowitz inequality (cf., Massart (1990)) we have

$$\mathbf{P}\left\{\sup_{y\in\mathbb{R}}\left|G_{m_n(X)}(y)-\hat{G}_{m_n(X),N_n}(y)\right| > c\cdot\frac{1}{\sqrt{N_n}}\right\} \le 2\cdot\exp\left(-c^2\right),$$

which implies the assertion. □

>From Lemma 2 and Lemma 3 we conclude

Lemma 4. *Let $A_n \subseteq \mathbb{R}^d$ and set*

$$\beta_n = \sup_{x\in A_n}|m_n(x)-m(x)|.$$

Let $B_n \subseteq \mathbb{R}$ be such that

$$\sup_{y\in\mathbb{R},\,|y-z|\le\beta_n \text{ for some } z\in B_n} g(y) \le c_6 < \infty. \tag{30}$$

Then

$$\sup_{y\in B_n}\left|\hat{G}_{m_n(X),N_n}(y)-G_{m(X)}(y)\right| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{N_n}}+\beta_n+\mathbf{P}_X\left(A_n^c\right)\right).$$

Proof. We have

$$\begin{aligned}
&\sup_{y\in B_n}\left|\hat{G}_{m_n(X),N_n}(y)-G_{m(X)}(y)\right| \\
&\le \sup_{y\in\mathbb{R}}\left|\hat{G}_{m_n(X),N_n}(y)-G_{m_n(X)}(y)\right| + \sup_{y\in B_n}\left|G_{m_n(X)}(y)-G_{m(X)}(y)\right| \\
&=: T_{1,n}+T_{2,n}.
\end{aligned}$$

It follows from Lemma 3 that

$$T_{1,n} = O_{\mathbf{P}}\left(\frac{1}{\sqrt{N_n}}\right).$$

Application of Lemma 2 yields

$$T_{2,n} \le \sup_{y\in B_n}\left|G_{m(X)}(y-\beta_n)-\mathbf{P}_X\left(A_n^c\right)-G_{m(X)}(y)\right|$$

$$
\begin{aligned}
&\quad + \sup_{y \in B_n} \left| G_{m(X)}(y + \beta_n) + \mathbf{P}_X\left(A_n^c\right) - G_{m(X)}(y) \right| \\
&\leq 2 \cdot \mathbf{P}_X\left(A_n^c\right) + \sup_{y \in B_n} \left| G_{m(X)}(y - \beta_n) - G_{m(X)}(y) \right| \\
&\quad + \sup_{y \in B_n} \left| G_{m(X)}(y + \beta_n) - G_{m(X)}(y) \right| \\
&\leq 2 \cdot \mathbf{P}_X\left(A_n^c\right) + 2 \cdot c_6 \cdot \beta_n,
\end{aligned}
$$

where the last inequality follows from the mean-value theorem and assumption (30). The proof is complete. □

Proof of Theorem 2. Set

$$
\delta_n = \frac{1}{\sqrt{N_n}} + 2 \cdot \min_{p \in \mathcal{P}_n} \|m_{n,p} - m\|_{\infty, K_n} + \mathbf{P}_X(K_n^c) + r \cdot \frac{\log n}{n}
$$

and observe

$$
\delta_n \to 0 \quad (n \to \infty). \tag{31}
$$

In the first step of the proof we show that it suffices to demonstrate that

$$
\lim_{c \to \infty} \limsup_{n \to \infty} \mathbf{P}\left\{ \left| \hat{q}_{m_n(X), N_n, \alpha} - q_{m(X), \alpha} \right| > c \cdot \delta_n \right\} = 0. \tag{32}
$$

This follows from

$$
\begin{aligned}
&\lim_{c \to \infty} \limsup_{n \to \infty} \mathbf{P}\Bigg\{ \left| \hat{q}_{m_n(X), N_n, \alpha} - q_{m(X), \alpha} \right| > \\
&\qquad c \cdot \left(\frac{1}{\sqrt{N_n}} + \min_{p \in \mathcal{P}_n} \|m_{n,p} - m\|_{\infty, K_n} + \mathbf{P}_X\left(K_n^c\right) + \frac{\log n}{n} \right) \Bigg\} \\
&\leq \lim_{c \to \infty} \limsup_{n \to \infty} \mathbf{P}\left\{ \left| \hat{q}_{m_n(X), N_n, \alpha} - q_{m(X), \alpha} \right| > \frac{c}{\max\{2, r\}} \cdot \delta_n \right\} \\
&= \lim_{c \to \infty} \limsup_{n \to \infty} \mathbf{P}\left\{ \left| \hat{q}_{m_n(X), N_n, \alpha} - q_{m(X), \alpha} \right| > c \cdot \delta_n \right\}.
\end{aligned}
$$

Because of the assumption that g is positive and continuous at $q_{m(X),\alpha}$, there exists $c_7 > 0$ such that for n sufficiently large we have

$$
c_7 \leq \inf_{y \in [q_{m(X),\alpha} - c \cdot \delta_n, q_{m(X),\alpha} + c \cdot \delta_n]} g(y). \tag{33}
$$

In the second step of the proof we show that (32) is implied by

$$\lim_{c\to\infty}\limsup_{n\to\infty}\mathbf{P}\left\{\sup_{y\in[q_{m(X),\alpha}-c\cdot\delta_n,q_{m(X),\alpha}+c\cdot\delta_n]}\left|\hat{G}_{m_n(X),N_n}(y)-G_{m(X)}(y)\right|\right.$$
$$\left.>\frac{c_7}{2}\cdot c\cdot\delta_n\right\}=0. \tag{34}$$

Condition (33) implies

$$\begin{aligned}
G_{m(X)}(q_{m(X),\alpha}-c\cdot\delta_n) &\leq \alpha-\left(G_{m(X)}(q_{m(X),\alpha})-G_{m(X)}(q_{m(X),\alpha}-c\cdot\delta_n)\right)\\
&\leq \alpha-c_7\cdot c\cdot\delta_n\\
&\leq \alpha+c_7\cdot c\cdot\delta_n\\
&\leq \alpha+\left(G_{m(X)}(q_{m(X),\alpha}+c\cdot\delta_n)-G_{m(X)}(q_{m(X),\alpha})\right)\\
&\leq G_{m(X)}(q_{m(X),\alpha}+c\cdot\delta_n).
\end{aligned}$$

On the event

$$\left\{\sup_{[y\in q_{m(X),\alpha}-c\cdot\delta_n,q_{m(X),\alpha}+c\cdot\delta_n]}\left|\hat{G}_{m_n(X),N_n}(y)-G_{m(X)}(y)\right|\leq\frac{c_7}{2}\cdot c\cdot\delta_n\right\}$$

this in turn implies

$$\begin{aligned}
&\hat{G}_{m_n(X),N_n}\left(q_{m(X),\alpha}-c\cdot\delta_n\right)\\
&=G_{m(X)}\left(q_{m(X),\alpha}-c\cdot\delta_n\right)\\
&\quad+\left(\hat{G}_{m_n(X),N_n}\left(q_{m(X),\alpha}-c\cdot\delta_n\right)-G_{m(X)}\left(q_{m(X),\alpha}-c\cdot\delta_n\right)\right)\\
&<\alpha\\
&<G_{m(X)}\left(q_{m(X),\alpha}+c\cdot\delta_n\right)\\
&\quad-\left(G_{m(X)}\left(q_{m(X),\alpha}+c\cdot\delta_n\right)-\hat{G}_{m_n(X),N_n}\left(q_{m(X),\alpha}+c\cdot\delta_n\right)\right)\\
&=\hat{G}_{m_n(X),N_n}\left(q_{m(X),\alpha}+c\cdot\delta_n\right),
\end{aligned}$$

from which we can conclude by the definition of $\hat{q}_{m(_nX),N_n,\alpha}$ that

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

Summarizing the above results we get

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

$$\leq \lim_{c\to\infty}\limsup_{n\to\infty}\mathbf{P}\left\{\sup_{[q_{m(X),\alpha}-c\cdot\delta_n,q_{m(X),\alpha}+c\cdot\delta_n]}\left|\hat{G}_{m_n(X),N_n}(y)-G_{m(X)}(y)\right|>\frac{c_7}{2}\cdot c\cdot\delta_n\right\},$$

which completes the second step of the proof.

In the third (and final) step of the proof we use Lemma 4 to show (34). First we observe that for n sufficiently large we have

$$\sup_{y\in[q_{m(X),\alpha}-2\cdot c\cdot\delta_n,q_{m(X),\alpha}+2\cdot c\cdot\delta_n]} g(y)\leq c_8. \tag{35}$$

Next we set

$$A_n=K_n\setminus\left\{x\in K_n\,:\,|m_n(x)-m(x)|>2\cdot\hat{\beta}\right\},$$

where

$$\begin{aligned}\hat{\beta} &= \max_{\substack{i=1,\dots,n,\\ X_i\in K_n}}|m_n(X_i)-m(X_i)|\\ &= \min_{p\in\mathcal{P}_n}\max_{\substack{i=1,\dots,n,\\ X_i\in K_n}}|m_{n,p}(X_i)-m(X_i)|\\ &\leq \min_{p\in\mathcal{P}_n}\|m_{n,p}-m\|_{\infty,K_n}.\end{aligned}$$

Clearly this choice of A_n implies

$$\beta_n=\sup_{x\in A_n}|m_n(x)-m(x)|\leq 2\cdot\hat{\beta}\leq 2\cdot\min_{p\in\mathcal{P}_n}\|m_{n,p}-m\|_{\infty,K_n}.$$

Furthermore we know that by Lemma 1 we have outside of an event, whose probability tends to zero for $n\to\infty$,

$$\begin{aligned}\mathbf{P}_X(A_n^c) &\leq \mathbf{P}_X(K_n^c)+\mathbf{P}_X\left(\left\{x\in K_n\,:\,|m_n(x)-m(x)|>2\cdot\hat{\beta}\right\}\right)\\ &\leq \mathbf{P}_X(K_n^c)+r\cdot\frac{\log n}{n}.\end{aligned}$$

Setting

$$B_n=[q_{m(X),\alpha}-c\cdot\delta_n,q_{m(X),\alpha}+c\cdot\delta_n],$$

we get the assertion by an application of Lemma 4.

The proof is complete. □

5 Acknowledgment

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

References

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

[2] 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.

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

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

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

[10] Enss, C., Kohler, M., Krzyżak, A. and Platz, R. (2016). Nonparametric quantile estimation based on surrogate models. *IEEE Transactions on Information Theory*, **62**, pp. 5727-5739.

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

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

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

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

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

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

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

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

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

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

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

[22] 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.

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