Estimating quantiles in imperfect simulation models using conditional density estimation [*]

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*

November 16, 2017

Abstract

In this article we consider the problem of estimating quantiles related to the outcome of experiments with a technical system given the distribution of the input together with an (imperfect) simulation model of the technical system, and (few) data points from the technical system. The distribution of the outcome of the technical system is estimated in a regression model, where the distribution of the residuals is estimated on the basis of a conditional density estimate. It is shown how Monte Carlo can be used to estimate quantiles of the outcome of the technical system on the basis of the above estimates, and the rate of convergence of the quantile estimate is analyzed. Under suitable assumptions it is shown that this rate of convergence is faster than the rate of convergence of standard estimates which ignore either the (imperfect) simulation model or the data from the technical system, hence it is crucial to combine both kinds of information. The results are illustrated by applying the estimates to simulated and real data.

AMS classification: Primary 62G05; secondary 62P30.

Key words and phrases: Conditional density estimation, quantile estimation, imperfect models, L_1 error, surrogate models, uncertainty quantification.

1 Introduction

The design of complex technical systems by engineers always has to take into account some kind of uncertainty. This uncertainty might occur because of lack of knowledge about future use or about properties of the materials used to build the technical system (e.g., the exact value of the damping coefficient of a spring–mass damper). In order to take this uncertainty into account, we model in the sequel the outcome Y of the technical system by a random variable. For simplicity we restrict ourselves to the case that Y is a

[*]Running title: *Estimating quantiles in imperfect simulation models*

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

real–valued random variable. Thus we are interested in properties of the distribution of Y, e.g., we are interested in quantiles

$$q_{Y,\alpha} = \min\{y \in \mathbb{R} : \mathbf{P}\{Y \leq y\} \geq \alpha\} \tag{1}$$

for $\alpha \in (0,1)$ (which describe for α close to one values which we expect to be upper bounds on the values occurring in an application), or in the density $g_Y : \mathbb{R} \to \mathbb{R}$ of Y with respect to the Lebesgue-Borel measure, which we assume later to exist.

In the sequel we model the lack of knowledge about the future use of the system or about properties of materials used in it by introducing an additional $\mathbb{R}^d$–valued random variable X, which contains values for uncertain parameters describing the system or its future use, and from which we assume either to know the distribution or are able to generate an arbitrary number of independent realizations. Furthermore we assume that we have available a model describing the relation between X and Y by a function $\bar{m} : \mathbb{R}^d \to \mathbb{R}$. This function $\bar{m}$ might be constructed by using a physical model of our technical system, and in some sense $\bar{m}(X)$ is an approximation of Y. However, as all models our model is imperfect in the sense that $Y = \bar{m}(X)$ does not hold. This might be due to the fact that Y cannot be exactly characterized by a function of X (since X might not describe the randomness of Y completely), or since our relation between Y and X is not correctly specified by $\bar{m}$, or because of both. So although we know $\bar{m}$ and can generate an arbitrary number of independent copies X_1, X_2, ... of X, we cannot use $\bar{m}(X_1)$, $\bar{m}(X_2)$, ... as observations of Y, since there is an error between these values and a sample of Y.

In order to control this error, we assume that we have available $n \in \mathbb{N}$ observations of the Y–values corresponding to the first n values of X. To formulate our prediction problem precisely, let (X,Y), (X_1,Y_1), (X_2,Y_2), ... be independent and identically distributed and let $L_n, N_n \in \mathbb{N}$. We assume that we are given the data

$$\begin{aligned}&(X_1,Y_1),\dots,(X_n,Y_n),(X_{n+1},\bar{m}(X_{n+1})),\dots,(X_{n+L_n},\bar{m}(X_{n+L_n})),\\&X_{n+L_n+1},\dots,X_{n+L_n+N_n},\end{aligned} \tag{2}$$

and we want to use this data in order to estimate the quantiles $q_{Y,\alpha}$ or the density g_Y of Y (which we later assume to exist). The main difficulty in solving this problem is that the sample size n of the observations of Y (which corresponds to the number of experiments we are making with the technical system) is rather small (since these experiments are time consuming or costly).

Before we describe various existing approaches to solve this problem in the literature, we will illustrate the problem by an example. Here we consider a demonstrator for a suspension strut, which was built at Technische Universität Darmstadt and which serves as an academic demonstrator to study uncertainty in load distributions and the ability to control vibrations, stability and load paths in suspension struts such as aircraft landing gears. The photo of this suspension strut and its experimental test setup is shown in Figure 1 (left), a CAD illustration of this suspension strut can be found in Figure 1 (middle).

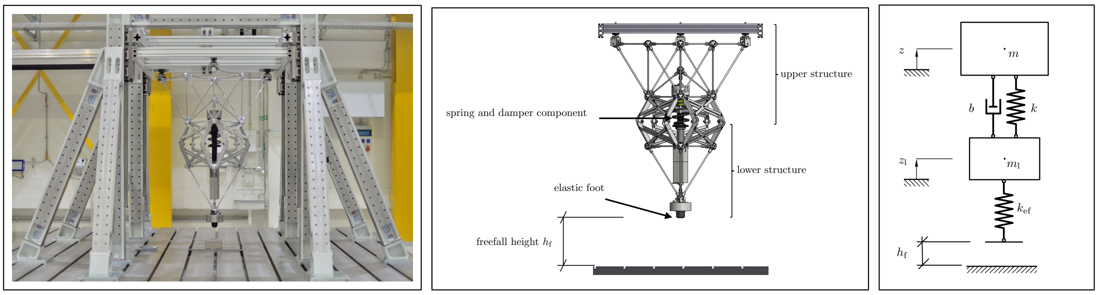

Figure 1: A photo of the demonstrator of a suspension strut and its experimental test setup (left), a CAD illustration of the suspension strut (middle) and illustration of a simplified model of the suspension strut (right).

This suspension strut consists of an upper and lower structure, where the lower structure contains a spring–damper component and an elastic foot. The spring–damper component transmits the axial forces between the upper and lower structures of the suspension strut. The aim of our analysis is the analysis of the behaviour of the maximum relative compression of the spring damper component in case that the free fall height is chosen randomly. Here we assume that the free fall heights are independent normally distributed with mean 0.05 meter and standard deviation 0.0057 meter.

We analyze the uncertainty in the maximum relative compression in our suspension strut using a simplified mathematical model of the suspension strut (cf., Figure 1 (right)), where the upper and the lower structures of the suspension strut are two lump masses m and m_1, the spring damper component is represented by a stiffness parameter k and a suitable damping coefficient b, and the foot is represented by another stiffness parameter k_{ef}. Using a linear stiffness and an axiomatic damping it is possible to compute the maximum relative compression by solving a differential equation using Runge-Kutta algorithm (cf., model a) in Mallapur and Platz (2017)). Figure 2 shows $L_n = 500$ data points from the computer experiment and also $n = 20$ experimental data points. Since they do not look like they come from the same source, our computer experiment is obviously imperfect. Our aim in the sequel is to us the $n = 20$ data points from our experiments with the suspension strut together with the $L_n = 500$ data points from the computer experiments in order to analyze the uncertainty in the above described experiments with the suspension strut. This can be done, e.g., by making some statistical inference about quantiles or the density of the maximal occurring compression in experiments with the suspension strut.

There are various possible approaches to solve the above estimation problem. The simplest idea is to ignore the model $\bar{m}(X)$ completely and to make inference about $q_{Y,\alpha}$ and g_Y using only the observations

$$Y_1, \ldots, Y_n \tag{3}$$

of Y. E.g., we can estimate the quantile $q_{Y,\alpha}$ by the plug–in estimate

$$\hat{q}_{Y,n,\alpha} = \min\left\{ y \in \mathbb{R} \, : \, \hat{G}_{Y,n}(y) \geq \alpha \right\} \tag{4}$$

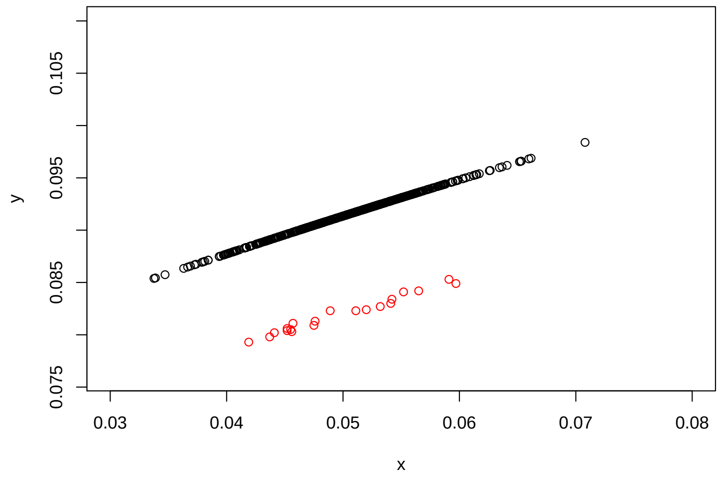

Figure 2: Data from $L_n = 500$ computer experiments (in black) together with data (in red) from $n = 20$ experiments with the suspension strut in Figure 1 (left panel).

corresponding to the estimate

$$\hat{G}_{Y,n}(y) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,y]}(Y_i)$$

of the cumulative distribution function (cdf.) $G(y) = \mathbf{P}\{Y \leq y\}$ of Y, which result in an order statistics as an estimate of the quantile. Or we can estimate the density g_Y of Y by the well–known kernel density estimate of Rosenblatt (1956) and Parzen (1962), where we first choose a density $K : \mathbb{R} \to \mathbb{R}$ (so–called kernel) and a so–called bandwidth $h_n > 0$ and define our estimate by

$$\hat{g}_{Y,n}(y) = \frac{1}{n \cdot h_n} \cdot \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h_n}\right).$$

However, since the sample size n of our data (3) is rather small, this will in general not lead to satisfying results.

Another simple idea is to ignore the real data (3), and to use the model data

$$\bar{m}(X_{n+1}), \ldots, \bar{m}(X_{n+L_n}) \tag{5}$$

as a sample of Y with additional measurement errors, and to use this sample to define quantile and density estimates as above. In this way we estimate $q_{Y,\alpha}$ by

$$\hat{q}_{\bar{m}(X),L_n,\alpha} = \min\left\{ y \in \mathbb{R} \,:\, \hat{G}_{\bar{m}(X),L_n}(y) \geq \alpha \right\} \tag{6}$$

where

$$\hat{G}_{\bar{m}(X),N_n}(y) = \frac{1}{L_n} \sum_{i=1}^{L_n} I_{(-\infty,y]}(\bar{m}(X_{n+i})),$$

and we can estimate the density g of Y by

$$\hat{g}_{\bar{m}(X),L_n}(y) = \frac{1}{L_n \cdot h_{L_n}} \cdot \sum_{i=1}^{L_n} K\left(\frac{y - \bar{m}(X_{n+i})}{h_{L_n}}\right).$$

Since the function $\bar{m}$ of our model $\bar{m}(X)$ of Y might be costly to evaluate (e.g., in case that its values are defined as solutions of a complicated partially differential equation) and consequently L_n might not be really large, it makes sense to use in a first step the data

$$(X_{n+1}, \bar{m}(X_{n+1})), \dots, (X_{n+L_n}, \bar{m}(X_{n+L_n}))$$

to compute a surrogate model

$$\hat{m}_{L_n}(\cdot) = \hat{m}_{L_n}(\cdot, (X_{n+1}, \bar{m}(X_{n+1})), \dots, (X_{n+L_n}, \bar{m}(X_{n+L_n})) : \mathbb{R}^d \to \mathbb{R}$$

of $\bar{m}$, and to compute in the second step the quantile and density estimates $\hat{q}_{\hat{m}_{L_n}(X),N_n,\alpha}$ and $\hat{g}_{\bar{m}_{L_n}(X),N_n}$ using the data

$$\hat{m}_{L_n}(X_{n+L_n+1}), \dots, \hat{m}_{L_n}(X_{n+L_n+N_n}).$$

Surrogate models have been introduced and investigated with the aid of the simulated and real data in connection with the quadratic response surfaces in Bucher and Burgund (1990), Kim and Na (1997) and Das and Zheng (2000), in context of 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 context of kriging in Kaymaz (2005) and Bichon et al. (2008).

Under the assumption that we have $\bar{m}(X) = Y$, the above estimates have been theoretically analyzed in Devroye, Felber and Kohler (2013), Bott, Felber and Kohler (2015), Felber, Kohler and Krzyżak (2015a, 2015b), Enss et al. (2016) and Kohler and Krzyżak (2017a).

However, in practice there usually will be an error in the approximation of Y by $\bar{m}(X)$, and it is unclear how this error influences the error of the quantile and density estimates.

Kohler et al. (2016) and Kohler and Krzyżak (2016) used the data

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

obtained by experiments with the technical system in order to control this error. In particular, confidence intervals for quantiles and confidence bands for densities are derived there. Wong, Storlie and Lee (2017) used the above data of the technical system in order to calibrate a computer model and estimated the error of the resulting model by using bootstrap. Kohler and Krzyżak (2017b) used this data in order to improve the surrogate model and analyzed the density estimate based on the improved surrogate model.

Kohler et al. (2016) and Kohler and Krzyżak (2016, 2017b) try to approximate Y by some function of X and make statistical inference on the basis of this approximation. Wong, Storlie and Lee (2017) do this similarly, but take into account additional measurement errors of the y-values. The basic new idea in this article is to estimate instead a regression model

$$Y = \bar{m}(X) + \bar{\epsilon}, \tag{7}$$

where

$$\bar{\epsilon} = Y - \bar{m}(X)$$

is the residual error of our model $\bar{m}(X)$, which is not related to measurement errors but instead is due to the fact, that an approximation of Y by a function of X cannot be perfect. In this model we estimate simultaneously $\bar{m}$ and the conditional distribution $\mathbf{P}_{\bar{\epsilon}|X=x}$ of $\bar{\epsilon}$ given $X = x$. As soon as we have available estimates $\hat{m}_{L_n}$ and $\mathbf{P}_{\bar{\epsilon}|X=x}$ for both, we generate data

$$\hat{m}_{L_n}(X_{n+L_n+1}) + \hat{\epsilon}(X_{n+L_n+1}), \dots, \hat{m}_{L_n}(X_{n+L_n+N_n}) + \hat{\epsilon}(X_{n+L_n+N_n})$$

(where $\hat{\epsilon}(x)$ has the distribution $\hat{\mathbf{P}}_{\bar{\epsilon}|X=x}$ conditioned on $X = x$) and use this data to define corresponding quantile estimates.

We assume in the sequel that the conditional distribution of $\bar{\epsilon}$ given X has a density with respect to the Lebesgue-Borel measure. In order to estimate this conditional density, we use the well-known conditional kernel density estimate introduced already in Rosenblatt (1969). Concerning existing results on conditional density estimates we refer to Fan, Yao and Tong (1996), Fan and Yim (2004), Gooijer and Zerom (2003), Efromovich (2007), Bott and Kohler (2016, 2017) and the literature cited therein.

Our main result, which is formulated in Section 3, shows that our newly proposed quantile estimates achieve under suitable regularity condition rates of convergence, which are faster than the rates of convergence of the estimates (4), (6) and the modifications of (6) using $\hat{m}_{L_n}$ instead of $\bar{m}$. Furthermore we show with simulated data that in the situations which we consider in our simulations this effect also occurs for finite sample sizes, and illustrate the usefulness of our newly proposed method by applying it to a spring-damper system introduced earlier.

Throughout this paper we use the following notation: $\mathbb{N}$, $\mathbb{N}_0$ and $\mathbb{R}$ are the sets of positive integers, nonnegative integers and real numbers, respectively. Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, 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

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

for all $x, z \in \mathbb{R}^d$. If X is a random variable, then $\mathbf{P}_X$ is the corresponding distribution, i.e., the measure associated with the random variable. If (X, Y) is a $\mathbb{R}^d \times \mathbb{R}$–valued random variable and $x \in \mathbb{R}^d$, then $\mathbf{P}_{Y|X=x}$ denotes the conditional distribution of Y

given $X = x$. Let $D \subseteq \mathbb{R}^d$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be a real-valued function defined on $\mathbb{R}^d$. We write $x = \arg\min_{z \in D} f(z)$ if $\min_{z \in \mathcal{D}} f(z)$ exists and if x satisfies

$$x \in D \quad \text{and} \quad f(x) = \min_{z \in \mathcal{D}} f(z).$$

For $x \in \mathbb{R}^d$ and $r > 0$ we denote the (closed) ball with center x and radius r by $S_r(x)$. If A is a set, then I_A is the indicator function corresponding to A, i.e., the function which takes on the value 1 on A and is zero elsewhere. For $A \subseteq \mathbb{R}$ we denote the infimum of A by $\inf A$, where we use the convention $\inf \emptyset = \infty$. If $x \in \mathbb{R}$, then we denote the smallest integer greater than or equal to x by $\lceil x \rceil$.

The outline of this paper is as follows: In Section 2 the construction of the newly proposed quantile estimate is explained. The main results are presented in Section 3 and proven in Section 5. The finite sample size performance of our estimates is illustrated in Section 4 by applying it to simulated and real data.

2 Definition of the estimate

In the sequel we assume that we are given data (2), where $n, L_n, N_n \in \mathbb{N}$, the $\mathbb{R}^d \times \mathbb{R}$ valued random variables (X, Y), (X_1, Y_1), (X_2, Y_2), ... are independent and identically distributed, and where $\bar{m} : \mathbb{R}^d \to \mathbb{R}$ is measurable. Our aim is to estimate the quantile $q_{Y,\alpha}$ defined in (1) for some $\alpha \in (0, 1)$.

To do this, we start by constructing an estimate of $\bar{m}$. For this we use the data

$$(X_{n+1}, \bar{m}(X_{n+1})), \dots, (X_{n+L_n}, \bar{m}(X_{n+L_n}))$$

and define the penalized least squares estimates of $\bar{m}$ by

$$\tilde{m}_{L_n}(\cdot) = \arg \min_{f \in W^k(\mathbb{R}^d)} \left(\frac{1}{L_n} \sum_{i=1}^{L_n} (\bar{m}(X_{n+i}) - f(X_{n+i}))^2 + \lambda_{L_n} \cdot J_k^2(f) \right)$$

and

$$\hat{m}_{L_n}(x) = T_{\beta_{L_n}}(\tilde{m}_{L_n}(x)) \quad (x \in \mathbb{R}^d)$$

for some $\beta_{L_n} > 0$, where $k \in \mathbb{N}$ with $2k > d$,

$$J_k^2(f) = \sum_{\alpha_1, \dots, \alpha_d \in \mathbb{N},\, \alpha_1 + \dots + \alpha_d = k} \frac{k!}{\alpha_1! \cdot \dots \cdot \alpha_d!} \int_{\mathbb{R}^d} \left| \frac{\partial^k f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(x) \right|^2 dx$$

is a penalty term penalizing the roughness of the estimate, $W^k(\mathbb{R}^d)$ denotes the Sobolev space

$$\left\{ f \,:\, \frac{\partial^k f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \in L_2(\mathbb{R}^d) \text{ for all } \alpha_1, \dots, \alpha_d \in \mathbb{N} \text{ with } \alpha_1 + \dots + \alpha_d = k \right\},$$

and where $\lambda_{L_n} > 0$, $T_L(x) = \max\{-L, \min\{L, x\}\}, L > 0$ is the truncation operator and $L_2(\mathbb{R}^d)$ denotes square integrable functions on $\mathbb{R}^d$. The condition $2k > d$ implies that the functions in $W^k(\mathbb{R}^d)$ are continuous and hence the value of a function at a point is well defined.

Then we compute the residuals of this estimate on the data (X_1, Y_1), ..., (X_n, Y_n), i.e., we set

$$\hat{\epsilon}_i = Y_i - \hat{m}_{L_n}(X_i) \quad (i = 1, \dots, n). \tag{8}$$

We use these residuals in order to estimate the conditional distribution of $\bar{\epsilon} = Y - \bar{m}(X)$ given $X = x$. Here we assume that this distribution has a density and estimate this density by applying a conditional density estimator to the data

$$(X_1, Y_1 - \hat{m}_{L_n}(X_1)), \dots, (X_n, Y_n - \hat{m}_{L_n}(X_n)).$$

To do this, we set $G = I_{[-1,1]}$ and let $K : \mathbb{R} \to \mathbb{R}$ be a density, let $h_n, H_n > 0$ and set

$$\hat{g}_{\hat{\epsilon}|X}(y, x) = \frac{\sum_{i=1}^n G\left(\frac{\|x - X_i\|}{H_n}\right) \cdot K\left(\frac{y - (Y_i - \hat{m}_{L_n}(X_i))}{h_n}\right)}{h_n \cdot \sum_{j=1}^n G\left(\frac{\|x - X_j\|}{H_n}\right)} \tag{9}$$

Once we have constructed the estimates $\hat{m}_n$ and $\hat{g}_{\hat{\epsilon}|X}$ we construct a sample of size N_n of the distribution of

$$\hat{m}_{L_n}(X) + \hat{\epsilon}(X),$$

where the random variable $\hat{\epsilon}(X)$ has the conditional density $\hat{g}_{\hat{\epsilon}|X}(\cdot, X)$ given X, and estimate the quantile by the empirical quantile corresponding to this sample. To do this we use an inversion method: We define for $u \in (0, 1)$ and $x \in \mathbb{R}^d$

$$F_n^{-1}(u, x) = \inf\left\{y \in \mathbb{R} \,:\, \int_{-\infty}^{y} \hat{g}_{\hat{\epsilon}|X}(z, x)\, dz \geq u\right\},$$

choose independent and identically uniformly on $(0, 1)$ distributed random variables U_1, U_2, ..., such that they are independent of all other previously introduced random variables, and set

$$\hat{Y}_{n+L_n+i} = F_n^{-1}(U_i, X_{n+L_n+i}) \quad (i = 1, \dots, N_n).$$

This implies in case

$$\int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z, X_{n+L_n+i})\, dz = 1$$

that $\hat{Y}_{n+L_n+i}$ conditioned on X_{n+L_n+i} has the density $\hat{g}_{\hat{\epsilon}|X}(\cdot, X_{n+L_n+i}))$.

With these random variables we estimate the cdf. of Y by

$$\hat{G}_{\hat{Y},N_n}(y) = \frac{1}{N_n} \sum_{i=1}^{N_n} I_{\{\hat{Y}_{n+L_n+i} \leq y\}},$$

and use the corresponding plug–in estimate

$$\hat{q}_{\hat{Y},N_n,\alpha} = \min\left\{y \in \mathbb{R} \,:\, \hat{G}_{\hat{Y},N_n}(y) \geq \alpha\right\}$$

as an estimate of $q_{Y,\alpha}$.

3 Main result

Our main result is the following theorem, which gives a nonasymptotic bound on the error of our quantile estimate.

Theorem 1 *Let (X,Y), (X_1,Y_1), (X_2,Y_2), ... be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$–valued random variables, and let $\bar{m} : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Let $g_{\bar{\epsilon}|X} : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ be a measurable function with the property that $g_{\bar{\epsilon}|X}(\cdot, X)$ is a density of the conditional distribution of $\bar{\epsilon} = Y - \bar{m}(X)$ given X. Assume that the following regularity conditions hold for some $C_1, C_2 > 0$, $r, s \in (0,1]$:*

(A1) $|g_{\bar{\epsilon}|X}(y,x_1) - g_{\bar{\epsilon}|X}(y,x_2)| \leq C_1 \cdot \|x_1 - x_2\|^r$ *for all* $x_1, x_2 \in \mathbb{R}^d, y \in \mathbb{R}$,

(A2) $|g_{\bar{\epsilon}|X}(u,x) - g_{\bar{\epsilon}|X}(v,x)| \leq C_2 \cdot |u - v|^s$ *for all* $u, v \in \mathbb{R}, x \in \mathbb{R}^d$

Let $n, L_n, N_n \in \mathbb{N}$ and assume $N_n^2 \geq 8 \cdot \log n$. For $\alpha \in (0,1)$ define the estimate $\hat{q}_{\hat{Y},N_n,\alpha}$ of the quantile $q_{Y,\alpha}$ (given by (1)) as in Section 2, where $h_n, H_n > 0$, G is the naive kernel and where $K : \mathbb{R}^d \to \mathbb{R}$ is bounded and symmetric density, which decreases monotonically on $\mathbb{R}_+$ and which satisfies

$$\int K^2(z)\, dz < \infty \quad \textit{and} \quad \int K(z) \cdot |z|^s dz < \infty.$$

Let $\gamma_n > 0$, assume $2 \cdot \sqrt{d} \cdot \gamma_n \geq H_n$, and for $x \in \mathbb{R}^d$ let $-\infty < a_n(x) \leq b_n(x) < \infty$. Set

$$\epsilon_n = 4 \cdot \mathbf{E} \int_{\mathbb{R}^d} |\hat{m}_{L_n}(x) - \bar{m}(x)|^2 \mathbf{P}_X(dx),$$

$$\begin{aligned}
\delta_n \;=\; & \frac{8 \cdot K(0) \cdot (4 \cdot \sqrt{d})^d \gamma_n^d}{h_n \cdot H_n^d} \cdot \mathbf{E} \int_{\mathbb{R}^d} |\hat{m}_{L_n}(x) - \bar{m}(x)|\, \mathbf{P}_X(dx) \\
& + 8 \cdot c_1 \cdot \Bigg(\sqrt{\frac{\int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)|\, \mathbf{P}_X(dx) \cdot \gamma_n^d}{n \cdot H_n^d \cdot h_n}} \\
& + \frac{4 \cdot \gamma_n^d}{n \cdot H_n^d} + 4 \cdot \int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)|\, \mathbf{P}_X(dx) \cdot (C_1 \cdot H_n^r + C_2 \cdot h_n^s) \Bigg) \\
& + 8 \cdot \mathbf{P}_X(\mathbb{R}^d \setminus [-\gamma_n,\gamma_n]^d) + 8 \cdot \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]^c} g_{\bar{\epsilon}|X}(y,x)\, dy\, \mathbf{P}_X(dx)
\end{aligned}$$

where

$$c_1 = \max \left\{ 1, \sqrt{2 \cdot (4 \cdot \sqrt{d})^d \cdot \int K^2(z)\, dz}, (4 \cdot \sqrt{d})^d, \int K(z) \cdot |z|^s dz \right\}$$

and

$$\eta_n = 4 \cdot \mathbf{P}\left\{ X \in \mathbb{R}^d \setminus [-\gamma_n,\gamma_n]^d \right\} + 4 \cdot \frac{(4 \cdot \sqrt{d})^d \cdot \gamma_n^d}{n \cdot H_n^d}.$$

Let $e_n > 0$ and assume that the cdf. of Y satisfies

$$G_Y(q_{Y,\alpha} + e_n - \epsilon_n^{1/3}) - G_Y(q_{Y,\alpha}) > \epsilon_n^{1/3} + \sqrt{\frac{\log N_n}{N_n}} + \delta_n + \eta_n \tag{10}$$

and

$$G_Y(q_{Y,\alpha}) - G_Y(q_{Y,\alpha} - e_n + \epsilon_n^{1/3}) > \epsilon_n^{1/3} + \sqrt{\frac{\log N_n}{N_n}} + \delta_n + \eta_n. \tag{11}$$

Then

$$\mathbf{P}\left\{ \left| \hat{q}_{\hat{Y},N_n,\alpha} - q_{Y,\alpha} \right| > (\log n) \cdot e_n \right\} \leq \frac{1}{\log n}.$$

Remark 1. Assume that Y has a density $g_Y : \mathbb{R} \to \mathbb{R}$ with respect to the Lebesgue measure which satisfies for some $c_2, c_3 > 0$

$$g_Y(y) > c_2 \quad \text{for all} \quad y \in [q_{Y,\alpha} - c_3, q_{Y,\alpha} + c_3]. \tag{12}$$

Assume that positive $\epsilon_n, \delta_n, \eta_n$ defined in Theorem 1 satisfy

$$\left(1 + \frac{1}{c_2}\right) \cdot \left(\epsilon_n^{1/3} + \delta_n + \eta_n + \sqrt{\frac{\log N_n}{N_n}} \right) \leq c_3, \tag{13}$$

and set

$$e_n = \left(1 + \frac{1}{c_2}\right) \cdot \left(\epsilon_n^{1/3} + \delta_n + \eta_n + \sqrt{\frac{\log N_n}{N_n}} \right).$$

Then (10) and (11) hold, and consequently we can conclude from Theorem 1

$$\mathbf{P}\left\{ \left| \hat{q}_{\hat{Y},N_n,\alpha} - q_{Y,\alpha} \right| > \left(1 + \frac{1}{c_2}\right) \cdot (\log n) \cdot \left(\epsilon_n^{1/3} + \delta_n + \eta_n + \sqrt{\frac{\log N_n}{N_n}} \right) \right\} \leq \frac{1}{\log n}.$$

Indeed, the assumptions above imply

$$0 \leq e_n - \epsilon_n^{1/3} \leq c_3.$$

Consequently because of the assumption on the density of Y we have

$$G_Y(q_{Y,\alpha} + e_n - \epsilon_n^{1/3}) - G_Y(q_{Y,\alpha}) \geq c_2(e_n - \epsilon_n^{1/3})).$$

By the definition of e_n we have

$$c_2(e_n - \epsilon_n^{1/3}) - \epsilon_n^{1/3} - \sqrt{\frac{\log N_n}{N_n}} - \delta_n - \eta_n > 0,$$

which implies (10). In the same way one can show (11).

Remark 2. Set $\gamma_n = \log(n)$. Under suitable smoothness assumptions on $\bar{m} : \mathbb{R}^d \to \mathbb{R}$, suitable assumptions on the tails of $\|X\|$ and in case that λ_{L_n} and β_{L_n} are suitably chosen it is well-known that the expected L_2 error of the smoothing spline estimate satisfies

$$\mathbf{E} \int_{\mathbb{R}^d} |\hat{m}_{L_n}(x) - \bar{m}(x)|^2 \mathbf{P}_X(dx) \leq c_4 \cdot \left(\frac{\log L_n}{L_n} \right)^{2k/(2k+d)}$$

(cf., e.g. Theorem 2 in Kohler and Krzyżak (2017b)). Thus for L_n large compared to n and under suitable assumptions on the tails of $\|X\|$ and on the tails of the conditional distribution of $\bar{\epsilon}$ given X it follows from Remark 1 that the error of our quantile estimate in Theorem 1 is up to some constant given by

$$(\log n) \cdot \Bigg(\sqrt{\frac{\int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx) \cdot (\log n)^d}{n \cdot H_n^d \cdot h_n}} + \frac{(\log n)^d}{n \cdot H_n^d}$$
$$+ \int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx) \cdot (C_1 \cdot H_n^r + C_2 \cdot h_n^s) \Bigg). \tag{14}$$

Minimizing the expression above with respect to h_n and H_n as in the proof of Corollary 2 in Bott and Kohler (2017), shows that in case of a suitable choice of the bandwidths $h_n, H_n > 0$ the error of our quantile estimate in Theorem 1 is up to some logarithmic factor given by the minimum of

$$C_1^{\frac{d}{r+d}} \cdot \Big(\int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx)\Big)^{\frac{d}{r+d}} \cdot n^{-\frac{r}{r+d}}$$
$$+ C_1^{\frac{ds}{(r+d)(2s+1)}} \cdot C_2^{\frac{1}{2s+1}} \Big(\int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx)\Big)^{\frac{(r+d)(s+1)+ds}{(r+d)(2s+1)}} \cdot n^{-\frac{rs}{(r+d)(2s+1)}}$$

and

$$C_1^{\frac{(2s+1)d}{r(2s+1)+ds}} \cdot C_2^{-\frac{d}{r(2s+1)+ds}} \cdot \Big(\int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx)\Big)^{\frac{ds}{r(2s+1)+ds}} \cdot n^{-\frac{r(2s+1)}{r(2s+1)+ds}}$$
$$+ C_1^{\frac{ds}{r(2s+1)+ds}} \cdot C_2^{\frac{r}{r(2s+1)+ds}} \cdot \Big(\int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx)\Big)^{\frac{r(s+1)+ds}{r(2s+1)+ds}} \cdot n^{-\frac{rs}{r(2s+1)+ds}}.$$

In case of $\int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx)$ small, these terms might get much smaller than the well-known rate of convergence $1/\sqrt{n}$ of the simple quantile estimate (4), and in case of imperfect models they will also be smaller than the rate of convergence of the surrogate quantile estimate.

Remark 3. The results in Remark 2 require that the parameters of the estimates (e.g., h_n and H_n) are suitably chosen. A data-dependent way of chosing these parameters in an application will be proposed in the next section, and by using simulated data it will be shown that in this case our newly proposed estimates outperform the other estimates for finite sample size in the situations which we consider there.

4 Application to simulated and real data

In this section we illustrate the finite sample size performance of our estimates by applying them to simulated and real data. We start with an application to simulated data, where we compare the simple order statistics estimate (*est. 1*) defined by (4) and a surrogate quantile estimate (*est. 2*) defined by (6) (where we replace $\bar{m}$ by $\hat{m}_{L_n}$ and evaluate this function on N_n x–values) with our newly proposed estimate based on estimation of the conditional density (*est. 3*) as defined in Section 2.

In the implementation of *est. 2* and *est. 3* we use thin plate splines (with smoothing parameter chosen by generalized cross validation) as implemented by the routine *Tps()* of *R* in order to estimate a surrogate model for our computer experiment. Here the implementation of the surrogate quantile estimate *est. 2* computes a sample of size $N_n = 100,000$ of $\hat{m}_{L_n}(X)$ and estimates the quantile by the corresponding order statistics.

In the implementation of our newly proposed *est. 3* we use the naive kernel $G(x) = I_{[-1,1]}(x)$ and the Epanechnikov kernel $K(y) = (3/4)\cdot(1-y^2)_+$ for the conditional density estimate

$$\hat{g}_{\hat{\epsilon}|X}(y,x) = \frac{\sum_{i=1}^{n} G\left(\frac{\|x-X_i\|}{H}\right) \cdot K\left(\frac{y-(Y_i-\hat{m}_{L_n}(X_i))}{h}\right)}{h \cdot \sum_{j=1}^{n} G\left(\frac{\|x-X_j\|}{H}\right)}.$$

Here the bandwidths h and H are chosen in a data dependent way from the sets

$$\mathcal{P}_h = \left\{2 \cdot 2^{-l} \cdot IQR(Y_1 - \hat{m}_{L_n}(X_1), \dots, Y_n - \hat{m}_{L_n}(X_n)) \,:\, l \in \{0,1,\dots,4\}\right\}$$

and

$$\mathcal{P}_H = \left\{2 \cdot 2^{-l} \cdot IQR(X_1,\dots,X_n) \,:\, l \in \{0,1,\dots,4\}\right\}$$

by the combinatorial method proposed by Bott and Kohler (2016), where IQR denotes an interquartile range, i.e., the distance between 25th and 75th percentiles. To do this we choose them by minimizing

$$\max_{\substack{h_1,h_2 \in \mathcal{P}_h,\\ H_1,H_2 \in \mathcal{P}_H}} \left| \frac{1}{n_t} \sum_{i=n_l+1}^{n} \int_{A_i(h_1,H_1,h_2,H_2)} \hat{g}_{\hat{\epsilon}|X}^{(n_l,(h,H)}(y,X_i)\,dy - \frac{1}{n_t} \cdot \sum_{i=n_l+1}^{n} I_{A_i(h_1,H_1,h_2,H_2)}(Y_i) \right|$$

with respect to $h \in \mathcal{P}_h$ and $H \in \mathcal{P}_H$, where $n_l = \lfloor n/2 \rfloor$, $n_t = n - n_l$,

$$\hat{g}_{\hat{\epsilon}|X}^{(n_l,(h,H))}(y,x) = \frac{\sum_{i=1}^{n_l} G\left(\frac{\|x-X_i\|}{H}\right) \cdot K\left(\frac{y-(Y_i-\hat{m}_{L_n}(X_i))}{h}\right)}{h \cdot \sum_{j=1}^{n_l} G\left(\frac{\|x-X_j\|}{H}\right)}$$

and

$$A_i(h_1,H_1,h_2,H_2) = \left\{y \in \mathbb{R} \,:\, \hat{g}_{\hat{\epsilon}|X}^{(n_l,(h_1,H_1))}(y,X_i) > \hat{g}_{\hat{\epsilon}|X}^{(n_l,(h_2,H_2))}(y,X_i)\right\}.$$

In the implementation of this method we approximate the integral

$$\int_{A_i(h_1,H_1,h_2,H_2)} \hat{g}_{\hat{\epsilon}|X}^{(n_l,(h,H)}(y,X_i)\,dy$$

by a Rieman sum based on an equidistant grid of

$$\Big[\min\{Y_1 - \hat{m}_{L_n}(X_1), \dots, Y_n - \hat{m}_{L_n}(X_n)\} - \max_{h \in \mathcal{P}_h} h,$$
$$\max\{Y_1 - \hat{m}_{L_n}(X_1), \dots, Y_n - \hat{m}_{L_n}(X_n)\} + \max_{h \in \mathcal{P}_h} h\Big]$$

consisting of 200 grid points (which enables an "efficient" implementation of the above minimization problem by first computing of $\hat{g}_{\hat{\epsilon}|X}^{(n_l,(h,H)}(y, X_i)$ for all grid points y, all $h \in \mathcal{P}_h$, all $H \in \mathcal{P}_H$ and all $i = n_l + 1, \dots, n$). After the computation of $\hat{g}_{\epsilon|X}$ we use the inversion method to generate random variables with the conditional density $\hat{g}_{\epsilon|X}(\cdot, X_i)$. Here we do not have to consider values outside of the above interval, since our density estimate is zero outside of this interval. In order to implement the inversion method we discretize the corresponding conditional cumulative distribution function

$$\begin{aligned}
\hat{G}_{\hat{\epsilon}|X}(y, X_i) &= \int_{-\infty}^{y} \hat{g}_{\hat{\epsilon}|X}(z, X_i)\, dz \\
&= \frac{\sum_{i=1}^{n_l} G\left(\frac{\|x-X_i\|}{H}\right) \cdot \int_{-\infty}^{y} K\left(\frac{z-(Y_i-\hat{m}_{L_n}(X_i))}{h}\right) dz}{h \cdot \sum_{j=1}^{n_l} G\left(\frac{\|x-X_j\|}{H}\right)}
\end{aligned}$$

by considering only its values on an equidistant grid of

$$\Big[\min\{Y_1 - \hat{m}_{L_n}(X_1), \dots, Y_n - \hat{m}_{L_n}(X_n)\} - h,$$
$$\max\{Y_1 - \hat{m}_{L_n}(X_1), \dots, Y_n - \hat{m}_{L_n}(X_n)\} + h\Big]$$

consisting of 1000 points, and by approximating the above integral by a Rieman sum corresponding to this grid. This enables again an "efficient" computation of the values of the conditional empirical cumulative distribution function by computing in advance

$$K\left(\frac{z - (Y_i - \hat{m}_{L_n}(X_i))}{h}\right)$$

for all grid points z and all $i = 1, \dots, n$. Using so computed values of the random variables we compute a sample of size $N_n = 100,000$ of Y and estimate the quantile by the corresponding order statistics.

We compare the above three estimates in the regression model

$$Y = m(X) + \epsilon,$$

where X is a standard normally distributed random variable,

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

and the conditional distribution of ϵ given X is normally distributed with mean zero and standard deviation

$$\sigma(X) = \sigma \cdot (0.25 + X \cdot (1 - X)).$$

Here $\sigma > 0$ is a parameter of our distribution for which we allow the values 0.5, 1 and 2. Furthermore we assume that our simulation model is based on the function

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

where $\delta \in \mathbb{R}$ is the constant model error of our model for which we consider the values 0 (i.e., no error) and 1 (i.e., negative error). Here we consider a negative value for the model error, since the surrogate quantile estimate tends to underestimate the quantile in the above example, so that a positive error might accidentally improve the surrogate quantile estimate.

We apply our estimates to samples of size $n \in \{20, 50, 100\}$ of (X, Y) and $L_n = 500$ of $(X, \bar{m}(X))$, and use them to estimate quantiles of order $\alpha = 0.95$ and $\alpha = 0.99$.

In order to judge the errors of our quantile estimate, we use a simple order statistics with sample size $1,000,000$ applied to a sample of Y as a reference value for the (unkonwn) quantile $q_{Y,\alpha}$ and compute the relative errors

$$\frac{|\hat{q}_{Y,\alpha} - q_{Y,\alpha}|}{q_{Y,\alpha}}.$$

Of course, our estimates $\hat{q}_{Y,\alpha}$ and hence also the above relative errors depend on the random samples selected above, and hence are random. Therefore we repeat the computation of the above error 100 times with newly generated independent samples and report the median and the interquantile ranges of the 100 errors in each of the considered cases for α, σ, δ and n, which results in errors for $2 \cdot 3 \cdot 2 \cdot 3 = 36$ different situations. The values we obtained in case $\alpha = 0.95$ and in case $\alpha = 0.99$ are reported in Tables 1 and 2, resp.

Looking at the results in Tables 1 and 2 wee see that our newly proposed estimate outperforms the order statistics estimate in all 36 settings of the simulations. Furthermore it outperfortms the surrogate quantile estimates whenever the model error is not zero, and also in case of the model error being zero whenever σ is large. There are a few cases with small σ value and zero model error where the surrogate quantile estimate is better than our newly proposed estimate, but in this case the difference between the errors is not large in contrast to the improvement of the error of the surrogate quantile estimate by our newly proposed estimate in most of the other cases.

Finally we illustrate the usefulness of our newly proposed method for uncertainty quantification by using it in analysis of the uncertainty occurring in experiments with the suspension strut in Figure 1 (left) described in the Introduction. We use the results of $L_n = 500$ computer experiments to construct a surrogate estimate $\hat{m}_{L_n}$ as described above, and we apply the method proposed in Section 2 to compute the conditional density of the residuals. To do this, we choose as described above the bandwidths h and H from the sets

$$\mathcal{P}_h = \{0.000766, 0.000383, 0.000191, 0.000096, 0.000048\}$$

and

$$\mathcal{P}_H = \{0.0174, 0.0087, 0.0043, 0.0022, 0.0011\}$$

n	σ	δ	est. 1	est. 2	est. 3
20	0.5	0	0.2876 (0.2315)	**0.0181** (0.0071)	0.0242 (0.0285)
20	0.5	1	0.2971 (0.3107)	0.2066 (0.0084)	**0.0254** (0.0458)
20	1	0	0.2911 (0.2617)	0.0950 (0.0078)	**0.0902** (0.0690)
20	1	1	0.2990 (0.2844)	0.2679 (0.0081)	**0.0844** (0.0869)
20	2	0	0.3511 (0.2645)	0.2771 (0.0070)	**0.1804** (0.1572)
20	2	1	0.3082 (0.3147)	0.4158 (0.0070)	**0.1816** (0.1514)
50	0.5	0	0.1595 (0.1638)	**0.0182** (0.0085)	0.0275 (0.0360)
50	0.5	1	0.2058 (0.2209)	0.2069 (0.0089)	**0.0224** (0.0360)
50	1	0	0.1579 (0.1584)	0.0941 (0.0079)	**0.0882** (0.0815)
50	1	1	0.2095 (0.2378)	0.2684 (0.0074)	**0.0768** (0.0830)
50	2	0	0.2361 (0.3509)	0.2757 (0.0061)	**0.1316** (0.1902)
50	2	1	0.2808 (0.2220)	0.4155 (0.0068)	**0.1428** (0.1550)
100	0.5	0	0.1210 (0.1312)	**0.0162** (0.0079)	0.0219 (0.0303)
100	0.5	1	0.1260 (0.1480)	0.2063 (0.0093)	**0.0211** (0.0371)
100	1	0	0.1269 (0.1574)	0.0930 (0.0084)	**0.0647** (0.0796)
100	1	1	0.1590 (0.1721)	0.2679 (0.0078)	**0.0732** (0.0735)
100	2	0	0.1269 (0.1835)	0.2760 (0.0060)	**0.0922** (0.0902)
100	2	1	0.1799 (0.1870)	0.4167 (0.0061)	**0.1143** (0.1238)

Table 1: Simulation results in case $\alpha = 0.95$. Reported are the median (and in brackets the interquartile range) of the 100 relative errors for each of our three estimates.

by using the combinatorial method of Bott and Kohler (2016). This results in $h = 0.000191$ and $H = 0.0043$. As described above we use the corresponding density estimate together with the surrogate model to generate an approximate sample of size $100,000$ of Y and estimate the $\alpha = 0.95$ quantile of Y by the corresponding order statistics, which results in the estimate 0.0855. In contrast the simple order statistics estimate of the quantile based only on the $n = 20$ experimental data points yields the smaller value 0.0849.

5 Proofs

5.1 Estimation of quantiles on the basis of conditional density estimates

Let (X,Y), (X_1,Y_1), (X_2,Y_2), ... be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$–valued random vectors and let $\bar{m} : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Assume that the conditional distribution of $\bar{\epsilon} = Y - \bar{m}(X)$ given X has the density $g_{\bar{\epsilon}|X}(\cdot, X) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with respect to the Lebesgue-Borel-measure, where $g_{\bar{\epsilon}|X} : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is measurable. Let $n, L_n, N_n \in \mathbb{N}$ and set

$$\mathcal{D}_n = \{(X_1,Y_1), \dots, (X_n,Y_n), (X_{n+1}, \bar{m}(X_{n+1})), \dots, (X_{n+L_n}, \bar{m}(X_{n+L_n}))\}.$$

n	σ	δ	est. 1	est. 2	est. 3
20	0.5	0	0.4711 (0.3179)	**0.0149** (0.0127)	0.0210 (0.0224)
20	0.5	1	0.4030 (0.3437)	0.1122 (0.0146)	**0.0523** (0.0918)
20	1	0	0.4430 (0.4022)	0.1125 (0.0124)	**0.1102** (0.0286)
20	1	1	0.4142 (0.3468)	0.1995 (0.0128)	**0.1284** (0.1059)
20	2	0	0.5208 (0.4348)	0.3601 (0.0125)	**0.3388** (0.1051)
20	2	1	0.5321 (0.3556)	0.4223 (0.0100)	**0.3569** (0.1328)
50	0.5	0	0.3172 (0.3565)	**0.0160** (0.0159)	0.0224 (0.0425)
50	0.5	1	0.2518 (0.2873)	0.1122 (0.0150)	**0.0291** (0.0643)
50	1	0	0.3137 (0.2480)	**0.1140** (0.0177)	0.1197 (0.0787)
50	1	1	0.3490 (0.3389)	0.2009 (0.0138)	**0.1211** (0.1055)
50	2	0	0.3059 (0.3802)	0.3578 (0.0107)	**0.2475** (0.2130)
50	2	1	0.2993 (0.4137)	0.4215 (0.0094)	**0.2556** (0.2473)
100	0.5	0	0.2439 (0.2368)	**0.0135** (0.0136)	0.0275 (0.0422)
100	0.5	1	0.2120 (0.3256)	0.1130 (0.0191)	**0.0390** (0.0476)
100	1	0	0.2125 (0.3053)	0.1114 (0.0152)	**0.1085** (0.0925)
100	1	1	0.2457 (0.2612)	0.1986 (0.0164)	**0.0978** (0.0886)
100	2	0	0.2644 (0.2248)	0.3608 (0.0107)	**0.1785** (0.1993)
100	2	1	0.2544 (0.3104)	0.4214 (0.0110)	**0.1686** (0.2273)

Table 2: Simulation results in case $\alpha = 0.99$. Reported are the median (and in brackets the interquartile range) of the 100 relative errors for each of our three estimates.

Let $\hat{m}_{L_n}(\cdot) = \hat{m}_{L_n}(\cdot, \mathcal{D}_n) : \mathbb{R}^d \to \mathbb{R}$ and let

$$\hat{g}_{\hat{\epsilon}|X}(\cdot,\cdot) = \hat{g}_{\hat{\epsilon}|X}(\cdot,\cdot,\mathcal{D}_n) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$$

be a measurable function satisfying

$$\hat{g}_{\hat{\epsilon}|X}(y,x) \geq 0 \quad \text{for all } y \in \mathbb{R}, x \in \mathbb{R}^d.$$

Let U, U_1, U_2, ... be independent random variables which are uniformly distributed on $(0,1)$ and which are idependent of (X,Y), (X_1,Y_1), ... and set

$$\hat{\epsilon} = \inf \left\{ y \in \mathbb{R} \, : \, \int_{-\infty}^{y} \hat{g}_{\hat{\epsilon}|X}(z,X) \, dz \geq U \right\}$$

and

$$\hat{\epsilon}_i = \inf \left\{ y \in \mathbb{R} \, : \, \int_{-\infty}^{y} \hat{g}_{\epsilon|X}(z,X_i) \, dz \geq U_i \right\} \quad (i \in \mathbb{N}).$$

Set

$$\hat{Y} = \hat{m}_{L_n}(X) + \hat{\epsilon} \quad \text{and} \quad \hat{Y}_i = \hat{m}_{L_n}(X_i) + \hat{\epsilon}_i \quad (i \in \{n+L_n+1, n+L_n+2, \ldots, n+L_n+N_n\}).$$

For $\alpha \in (0,1)$ set

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

where

$$G_Y(y) = \mathbf{P}\{Y \le y\},$$

and

$$\hat{q}_{\hat{Y},N_n,\alpha} = \min\left\{y \in \mathbb{R} \,:\, \hat{G}_{\hat{Y},N_n}(y) \ge \alpha\right\},$$

where

$$\hat{G}_{\hat{Y},N_n}(y) = \frac{1}{N_n} \sum_{i=n+L_n+1}^{N_n} I_{\{\hat{Y}_i \le y\}}.$$

Lemma 1 *Let $\alpha \in (0,1)$, $n \in \mathbb{N}$ and $L_n, N_n \in \mathbb{N}$ and define the estimate $\hat{q}_{Y,N_n,\alpha}$ of $q_{Y,\alpha}$ as above. Assume that $\hat{g}_{\hat{\epsilon}|X}$ satisfies*

$$\hat{g}_{\hat{\epsilon}|X}(y,x) \ge 0 \quad (y \in \mathbb{R}, x \in \mathbb{R}^d) \quad \text{and} \quad \int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(y,x)\,dy \le 1 \quad (x \in \mathbb{R}^d). \tag{15}$$

Let $\epsilon_n, \delta_n, \eta_n, e_n > 0$ be such that

$$G_Y(q_{Y,\alpha} + e_n - \epsilon_n^{1/3}) - G_Y(q_{Y,\alpha}) > \epsilon_n^{1/3} + \sqrt{\frac{\log N_n}{N_n}} + \delta_n + \eta_n \tag{16}$$

and

$$G_Y(q_{Y,\alpha}) - G_Y(q_{Y,\alpha} - e_n + \epsilon_n^{1/3}) > \epsilon_n^{1/3} + \sqrt{\frac{\log N_n}{N_n}} + \delta_n + \eta_n. \tag{17}$$

Then

$$\begin{aligned}
\mathbf{P}\left\{|\hat{q}_{Y,N_n,\alpha} - q_{Y,\alpha}| > e_n\right\} \quad \le \quad & \mathbf{P}\left\{\frac{1}{N_n}\sum_{i=1}^{N_n} |\hat{m}_n(X_{n+L_n+i}) - \bar{m}(X_{n+L_n+i})|^2 > \epsilon_n\right\} \\
& + \mathbf{P}\left\{\int_{\mathbb{R}^d}\int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)|\,dy\,\mathbf{P}_X(dx) > \delta_n\right\} \\
& + \mathbf{P}\left\{\mathbf{P}\left\{\int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z,X)\,dz \neq 1 \big| \mathcal{D}_n\right\} > \eta_n\right\} + \frac{2}{N_n^2}.
\end{aligned}$$

Proof. Set

$$\bar{Y} = \bar{m}(X) + \hat{\epsilon}, \quad \bar{Y}_i = \bar{m}(X_i) + \hat{\epsilon}_i \quad (i \in \mathbb{N}),$$

$$G_{\bar{Y}}(y) = \mathbf{P}\{\bar{Y} \le y | \mathcal{D}_n\} \quad \text{and} \quad \hat{G}_{\bar{Y},N_n}(y) = \frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{\bar{Y}_{n+L_n+i} \le y\}}.$$

By the Dvoretzky-Kiefer-Wolfowitz inequality (cf., Massart (1990)) applied conditionally on $\mathcal{D}_n$ we get

$$\mathbf{P}\left\{\sup_{y \in \mathbb{R}} \left|G_{\bar{Y}}(y) - \hat{G}_{\bar{Y},N_n}(y)\right| > \sqrt{\frac{\log N_n}{N_n}}\right\} \le 2 \cdot \exp\left(-2 \cdot N_n \cdot \frac{\log N_n}{N_n}\right) = \frac{2}{N_n^2}.$$

Since

$$\mathbf{P}\left\{|\hat{q}_{Y,N_n,\alpha} - q_{Y,\alpha}| > e_n\right\}$$

$$
\begin{aligned}
&\leq \mathbf{P}\Bigg\{ |\hat{q}_{Y,N_n,\alpha} - q_{Y,\alpha}| > e_n, \frac{1}{N_n} \sum_{i=1}^{N_n} |\hat{m}_{L_n}(X_{n+L_n+i}) - \bar{m}(X_{n+L_n+i})|^2 \leq \epsilon_n, \\
&\qquad \int_{\mathbb{R}^d} \int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx) \leq \delta_n, \\
&\qquad \mathbf{P}\left\{ \int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z,X) \, dz \neq 1 \big| \mathcal{D}_n \right\} \leq \eta_n, \sup_{y \in \mathbb{R}} \left| G_{\bar{Y}}(y) - \hat{G}_{\bar{Y},N_n}(y) \right| \leq \sqrt{\frac{\log N_n}{N_n}} \Bigg\} \\
&\quad + \mathbf{P}\left\{ \frac{1}{N_n} \sum_{i=1}^{N_n} |\hat{m}_{L_n}(X_{n+L_n+i}) - \bar{m}(X_{n+L_n+i})|^2 > \epsilon_n \right\} \\
&\quad + \mathbf{P}\left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx) > \delta_n \right\} \\
&\quad + \mathbf{P}\left\{ \mathbf{P}\left\{ \int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z,X) \, dz \neq 1 \big| \mathcal{D}_n \right\} > \eta_n \right\} \\
&\quad + \mathbf{P}\left\{ \sup_{y \in \mathbb{R}} \left| G_{\bar{Y}}(y) - \hat{G}_{\bar{Y},N_n}(y) \right| > \sqrt{\frac{\log N_n}{N_n}} \right\},
\end{aligned}
$$

it suffices to show that

$$
\frac{1}{N_n} \sum_{i=1}^{N_n} |\hat{m}_{L_n}(X_{n+L_n+i}) - \bar{m}(X_{n+L_n+i})|^2 \leq \epsilon_n, \tag{18}
$$

$$
\int_{\mathbb{R}^d} \int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx) \leq \delta_n, \tag{19}
$$

$$
\mathbf{P}\left\{ \int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z,X) \, dz \neq 1 \big| \mathcal{D}_n \right\} \leq \eta_n, \tag{20}
$$

and

$$
\sup_{y \in \mathbb{R}} \left| G_{\bar{Y}}(y) - \hat{G}_{\bar{Y},N_n}(y) \right| \leq \sqrt{\frac{\log N_n}{N_n}} \tag{21}
$$

imply

$$
|\hat{q}_{Y,N_n,\alpha} - q_{Y,\alpha}| \leq e_n. \tag{22}
$$

By the definition of $\hat{q}_{Y,N_n,\alpha}$ we know that (22) is implied by

$$
\hat{G}_{\hat{Y},N_n}(q_{Y,\alpha} + e_n) \geq \alpha \tag{23}
$$

and

$$
\hat{G}_{\hat{Y},N_n}(q_{Y,\alpha} - e_n) < \alpha, \tag{24}
$$

so it suffices to show that (18)-(21) imply (23) and (24), what we do next.

So assume from now on that (18)-(21) hold. Before we start with the proof of (23) we show

$$
\sup_{y \in \mathbb{R}} |G_Y(y) - G_{\bar{Y}}(y)| \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx) + \eta_n. \tag{25}
$$

Indeed, we observe first

$$\begin{aligned}
G_Y(y) &= \mathbf{P}\{Y \le y\} \\
&= \mathbf{E}\left\{\mathbf{P}\left\{\bar{m}(X) + \bar{\epsilon} \le y \big| X\right\}\right\} \\
&= \mathbf{E}\left\{\mathbf{P}\left\{\bar{\epsilon} \le y - \bar{m}(X) \big| X\right\}\right\} \\
&= \mathbf{E}\left\{\int_{-\infty}^{y-\bar{m}(X)} g_{\bar{\epsilon}|X}(z, X)\, dz\right\} \\
&= \int_{\mathbb{R}^d} \int_{-\infty}^{y-\bar{m}(x)} g_{\bar{\epsilon}|X}(z, x)\, dz\, \mathbf{P}_X(dx) \\
&= \int_{\mathbb{R}^d} \int_{-\infty}^{y} g_{\bar{\epsilon}|X}(z - \bar{m}(x), x)\, dz\, \mathbf{P}_X(dx).
\end{aligned}$$

Furthermore we have

$$\begin{aligned}
&G_{\bar{Y}}(y) \\
&= \mathbf{E}\left\{\mathbf{P}\left\{\hat{\epsilon} \le y - \bar{m}(X) \big| X, \mathcal{D}_n\right\} \big| \mathcal{D}_n\right\} \\
&= \mathbf{E}\Bigg\{ I_{\{\int_{\mathbb{R}} \hat{g}_{\bar{\epsilon}|X}(u,X)\, du = 1\}} \cdot \mathbf{P}\left\{\hat{\epsilon} \le y - \bar{m}(X) \big| X, \mathcal{D}_n\right\} \\
&\qquad\qquad + I_{\{\int_{\mathbb{R}} \hat{g}_{\bar{\epsilon}|X}(u,X)\, du \ne 1\}} \cdot \mathbf{P}\left\{\hat{\epsilon} \le y - \bar{m}(X) \big| X, \mathcal{D}_n\right\} \big| \mathcal{D}_n \Bigg\} \\
&= \mathbf{E}\Bigg\{ I_{\{\int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(u,X)\, du = 1\}} \cdot \int_{-\infty}^{y-\bar{m}(X)} \hat{g}_{\hat{\epsilon}|X}(z, X)\, dz \\
&\qquad\qquad + I_{\{\int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(u,X)\, du \ne 1\}} \cdot \mathbf{P}\left\{\hat{\epsilon} \le y - \bar{m}(X) \big| X, \mathcal{D}_n\right\} \big| \mathcal{D}_n \Bigg\} \\
&= \int_{\mathbb{R}^d} \int_{-\infty}^{y} \hat{g}_{\hat{\epsilon}|X}(z - \bar{m}(x), X)\, dz\, \mathbf{P}_X(dx) \\
&\quad + \mathbf{E}\left\{ I_{\{\int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(u,X)\, du \ne 1\}} \cdot \left(\mathbf{P}\left\{\hat{\epsilon} \le y - \bar{m}(X) \big| X, \mathcal{D}_n\right\} - \int_{-\infty}^{y-\bar{m}(X)} \hat{g}_{\hat{\epsilon}|X}(z, X)\, dz \right) \big| \mathcal{D}_n \right\}.
\end{aligned}$$

Since we have

$$\left| \mathbf{P}\left\{\hat{\epsilon} \le y - \bar{m}(X) \big| X\right\} - \int_{-\infty}^{y-\bar{m}(X)} \hat{g}_{\hat{\epsilon}|X}(z, X)\, dz \right| \le 1 \quad a.s.,$$

which follows from assumption (15)) and which implies

$$\begin{aligned}
&\left| \mathbf{E}\left\{ I_{\{\int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(u,X)\, du \ne 1\}} \cdot \left(\mathbf{P}\left\{\hat{\epsilon} \le y - \bar{m}(X) \big| X, \mathcal{D}_n\right\} - \int_{-\infty}^{y-\bar{m}(X)} \hat{g}_{\hat{\epsilon}|X}(z, X)\, dz \right) \big| \mathcal{D}_n \right\} \right| \\
&\le \mathbf{P}\left\{ \int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z, X)\, dz \ne 1 \big| \mathcal{D}_n \right\} \le \eta_n,
\end{aligned}$$

and

$$
\begin{aligned}
&\sup_{y\in\mathbb{R}}\left|\int_{\mathbb{R}^d}\int_{-\infty}^{y} g_{\bar{\epsilon}|X}(z-\bar{m}(x),x)\,dz\,\mathbf{P}_X(dx)-\int_{\mathbb{R}^d}\int_{-\infty}^{y} \hat{g}_{\hat{\epsilon}|X}(z-\bar{m}(x),x)\,dz\,\mathbf{P}_X(dx)\right|\\
&\leq \sup_{y\in\mathbb{R}}\int_{\mathbb{R}^d}\int_{-\infty}^{y} |g_{\bar{\epsilon}|X}(z-\bar{m}(x),x)-\hat{g}_{\hat{\epsilon}|X}(z-\bar{m}(x),x)|\,dz\,\mathbf{P}_X(dx)\\
&=\int_{\mathbb{R}^d}\int_{\mathbb{R}} |g_{\bar{\epsilon}|X}(z-\bar{m}(x),x)-\hat{g}_{\hat{\epsilon}|X}(z-\bar{m}(x),x)|\,dz\,\mathbf{P}_X(dx)\\
&=\int_{\mathbb{R}^d}\int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x)-g_{\bar{\epsilon}|X}(y,x)|\,dy\,\mathbf{P}_X(dx)
\end{aligned}
$$

this implies (25).

Next we prove (23). Using (18), (21), (25) and (19) we get

$$
\begin{aligned}
&\hat{G}_{\hat{Y},N_n}(q_{Y,\alpha}+e_n)\\
&\geq \frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{\hat{Y}_{n+L_n+i}\leq q_{Y,\alpha}+e_n,|\hat{Y}_{n+L_n+i}-\bar{Y}_{n+L_n+i}|\leq \epsilon_n^{1/3}\}}\\
&\geq \frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{\bar{Y}_{n+L_n+i}\leq q_{Y,\alpha}+e_n-\epsilon_n^{1/3},|\hat{Y}_{n+L_n+i}-\bar{Y}_{n+L_n+i}|\leq \epsilon_n^{1/3}\}}\\
&\geq \frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{\bar{Y}_{n+L_n+i}\leq q_{Y,\alpha}+e_n-\epsilon_n^{1/3}\}}-\frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{|\hat{Y}_{n+L_n+i}-\bar{Y}_{n+L_n+i}|>\epsilon_n^{1/3}\}}\\
&\geq \frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{\bar{Y}_{n+L_n+i}\leq q_{Y,\alpha}+c\cdot e_n-\epsilon_n^{1/3}\}}-\frac{1}{N_n}\sum_{i=1}^{N_n} \frac{|\hat{Y}_{n+L_n+i}-\bar{Y}_{n+L_n+i}|^2}{\epsilon_n^{2/3}}\\
&= \frac{1}{N_n}\sum_{i=1}^{N_n} I_{\{\bar{Y}_{n+L_n+i}\leq q_{Y,\alpha}+c\cdot e_n-\epsilon_n^{1/3}\}}-\frac{1}{\epsilon_n^{2/3}}\cdot\frac{1}{N_n}\sum_{i=1}^{N_n} |\hat{m}_{L_n}(X_{n+L_n+i})-\bar{m}(X_{n+L_n+i})|^2\\
&\geq \hat{G}_{\bar{Y},N_n}(q_{Y,\alpha}+e_n-\epsilon_n^{1/3})-\epsilon_n^{1/3}\\
&\geq G_{\bar{Y}}(q_{Y,\alpha}+e_n-\epsilon_n^{1/3})-\epsilon_n^{1/3}-\sup_{y\in\mathbb{R}}\left|G_{\bar{Y}}(y)-\hat{G}_{\bar{Y},N_n}(y)\right|\\
&\geq G_{\bar{Y}}(q_{Y,\alpha}+e_n-\epsilon_n^{1/3})-\epsilon_n^{1/3}-\sqrt{\frac{\log N_n}{N_n}}\\
&\geq G_{Y}(q_{Y,\alpha}+e_n-\epsilon_n^{1/3})-\epsilon_n^{1/3}-\sqrt{\frac{\log N_n}{N_n}}-\sup_{y\in\mathbb{R}}|G_Y(y)-G_{\bar{Y}}(y)|\\
&\geq G_{Y}(q_{Y,\alpha}+e_n-\epsilon_n^{1/3})-\epsilon_n^{1/3}-\sqrt{\frac{\log N_n}{N_n}}-\delta_n-\eta_n\\
&> G_Y(q_{Y,\alpha})=\alpha,
\end{aligned}
$$

where the last inequality follows from (16).

In the same way we argue that

$$\hat{G}_{\hat{Y},N_n}(q_{Y,\alpha}-e_n)$$

$$
\begin{aligned}
&\le G_Y(q_{Y,\alpha} - e_n + \hat{\epsilon}_n^{1/3}) + \epsilon_n^{1/3} + \sqrt{\frac{\log N_n}{N_n}} + \delta_n + \eta_n \\
&< \alpha,
\end{aligned}
$$

which finishes the proof. $\square$

5.2 A bound on the L_1 error of a conditional density estimate

Lemma 2 *Let (X,Y), (X_1,Y_1), ..., (X_n,Y_n) be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$–valued random vectors. Assume that the conditional distribution $\mathbf{P}_{Y|X}$ of Y given X has the density $g_{Y|X}(\cdot, X) : \mathbb{R} \to \mathbb{R}$ with respect to the Lebegue-Borel measure, where*

$$
g_{Y|X} : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}
$$

is a measurable function which satisfies

$$
|g_{Y|X}(y,x_1) - g_{Y|X}(y,x_2)| \le C_1 \cdot \|x_1 - x_2\|^r \quad \textit{for all } x_1, x_2 \in \mathbb{R}^d, y \in \mathbb{R}, \tag{26}
$$

and

$$
|g_{Y|X}(u,x) - g_{Y|X}(v,x)| \le C_2 \cdot |u-v|^s \quad \textit{for all } u, v \in \mathbb{R}, x \in \mathbb{R}^d \tag{27}
$$

for some $r, s \in (0,1]$ and some $C_1, C_2 > 0$. Let $\gamma_n > 0$. For $x \in \mathbb{R}^d$ let $-\infty < a_n(x) \le b_n(x) < \infty$ be such that

$$
\int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx) < \infty. \tag{28}
$$

Set $G = I_{[-1,1]}$ and let $K : \mathbb{R} \to \mathbb{R}$ be a density satisfying

$$
\int_{\mathbb{R}} K^2(z)\, dz < \infty \quad \textit{and} \quad \int_{\mathbb{R}} K(z) \cdot |z|^s dz < \infty.
$$

Let $h_n, H_n > 0$ be such that $2 \cdot \sqrt{d} \cdot \gamma_n \ge H_n$, and set

$$
\hat{g}_{Y|X}(y,x) = \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right) \cdot K\left(\frac{y-Y_i}{h_n}\right)}{h_n \cdot \sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)}, \tag{29}
$$

where $\frac{0}{0} := 0$. Then

$$
\begin{aligned}
&\mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} \left|\hat{g}_{Y|X}(y,x) - g_{Y|X}(y,x)\right| \, dy \, \mathbf{P}_X(dx) \\
&\le c_1 \cdot \left(\sqrt{\frac{\int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx) \cdot \gamma_n^d}{n \cdot H_n^d \cdot h_n}} \right. \\
&\qquad \left. + \frac{\gamma_n^d}{n \cdot H_n^d} + \int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx) \cdot (C_1 \cdot H_n^r + C_2 \cdot h_n^s) \right),
\end{aligned}
$$

where the constant

$$c_1 = \max\left\{1, \sqrt{2\cdot(4\cdot\sqrt{d})^d\cdot\int K^2(z)\,dz}, (4\cdot\sqrt{d})^d, \int K(z)\cdot|z|^s dz\right\}$$

does not depend on $\mathbf{P}_{(X,Y)}$, C_1 *or* C_2.

In the proof we will need the following well-known auxiliary result:

Lemma 3 *Let* $n\in\mathbb{N}$, *let* $H_n,\gamma_n >$ *be such that* $2\cdot\sqrt{d}\cdot\gamma_n \geq H_n$, *and let* X *be an* $\mathbb{R}^d$*–valued random variable. Then it holds:*

$$\int_{[-\gamma_n,\gamma_n]^d} \frac{1}{n\cdot\mathbf{P}_X(S_{H_n}(x))}\,\mathbf{P}_X(dx) \leq (4\cdot\sqrt{d})^d\cdot\frac{\gamma_n^d}{n\cdot H_n^d}.$$

Proof. The assertion follows from the proof of equation (5.1) in Györif et al. (2002), a complete proof is available from the authors on request. □

Proof of Lemma 2. By triangle inequality we have

$$\begin{aligned}
&\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\hat{g}_{Y|X}(y,x)-g_{Y|X}(y,x)\right|\,dy\,\mathbf{P}_X(dx) \qquad (30)\\
&\leq \mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\hat{g}_{Y|X}(y,x)-\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|\,dy\,\mathbf{P}_X(dx)\\
&\quad +\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}-g_{Y|X}(y,x)\right|\,dy\,\mathbf{P}_X(dx).
\end{aligned}$$

In the first step of the proof we show

$$\begin{aligned}
&\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\hat{g}_{Y|X}(y,x)-\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|\,dy\,\mathbf{P}_X(dx)\\
&\leq \sqrt{2\cdot(4\cdot\sqrt{d})^d\cdot\int K^2(z)\,dz}\cdot\sqrt{\frac{\int_{[-\gamma_n,\gamma_n]^d}|b_n(x)-a_n(x)|\,\mathbf{P}_X(dx)\cdot\gamma_n^d}{n\cdot H_n^d\cdot h_n}}. \qquad (31)
\end{aligned}$$

The inequality of Cauchy-Schwarz implies

$$\begin{aligned}
&\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\hat{g}_{Y|X}(y,x)-\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|\,dy\,\mathbf{P}_X(dx)\\
&= \mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\mathbf{E}\left\{1\cdot\left|\hat{g}_{Y|X}(y,x)-\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|\,\middle|X_1^n\right\}\,dy\,\mathbf{P}_X(dx)\\
&\leq \mathbf{E}\sqrt{\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\mathbf{E}\left\{1^2\middle|X_1^n\right\}\,dy\,\mathbf{P}_X(dx)}\\
&\quad \cdot\mathbf{E}\sqrt{\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\mathbf{E}\left\{\left|\hat{g}_{Y|X}(y,x)-\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|^2\middle|X_1^n\right\}\,dy\,\mathbf{P}_X(dx)}
\end{aligned}$$

$$\leq \sqrt{\int_{[-\gamma_n,\gamma_n]^d} |b_n(x)-a_n(x)|\, \mathbf{P}_X(dx)}$$

$$\cdot \sqrt{\mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} \mathbf{E}\left\{ \left| \hat{g}_{Y|X}(y,x) - \mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|^2 \middle| X_1^n \right\} \, dy\, \mathbf{P}_X(dx)},$$

hence it suffices to show

$$\mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} \mathbf{E}\left\{ \left| \hat{g}_{Y|X}(y,x) - \mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|^2 \middle| X_1^n \right\} \, dy\, \mathbf{P}_X(dx)$$
$$\leq 2 \cdot (4 \cdot \sqrt{d})^d \cdot \int K^2(z)\, dz \cdot \frac{\gamma_n^d}{n \cdot H_n^d \cdot h_n}. \tag{32}$$

To show this, we observe first that the independence of the data implies

$$\mathbf{E}\left\{ \left| \hat{g}_{Y|X}(y,x) - \mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|^2 \middle| X_1^n \right\}$$
$$= \mathbf{E}\left\{ \left| \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right) \cdot \left(K\left(\frac{y-Y_i}{h_n}\right) - \mathbf{E}\left\{K\left(\frac{y-Y_i}{h_n}\right) \middle| X_i\right\}\right)}{h_n \cdot \sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)} \right|^2 \middle| X_1^n \right\}$$
$$= \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right)^2 \cdot \mathbf{E}\left\{ \left| K\left(\frac{y-Y_i}{h_n}\right) - \mathbf{E}\left\{K\left(\frac{y-Y_i}{h_n}\right) \middle| X_i\right\}\right|^2 \middle| X_i \right\}}{h_n^2 \cdot \left(\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)\right)^2}$$
$$\leq \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right)^2 \cdot \mathbf{E}\left\{ \left| K\left(\frac{y-Y_i}{h_n}\right)\right|^2 \middle| X_i \right\}}{h_n^2 \cdot \left(\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)\right)^2}$$
$$= \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right) \cdot \int_{\mathbb{R}} K^2\left(\frac{y-u}{h_n}\right) \cdot g_{Y|X}(u,X_i)\, du}{h_n^2 \cdot \left(\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)\right)^2}.$$

Hence,

$$\mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} \mathbf{E}\left\{ \left| \hat{g}_{Y|X}(y,x) - \mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}\right|^2 \middle| X_1^n \right\} \, dy\, \mathbf{P}_X(dx)$$
$$\leq \mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right) \cdot \int_{\mathbb{R}} \int_{\mathbb{R}} K^2\left(\frac{y-u}{h_n}\right)\, dy \cdot g_{Y|X}(u,X_i)\, du}{h_n^2 \cdot \left(\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)\right)^2} \mathbf{P}_X(dx)$$
$$= \mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right) \cdot \int_{\mathbb{R}} K^2(z)\, dz \cdot h_n \cdot \int_{\mathbb{R}} g_{Y|X}(u,X_i)\, du}{h_n^2 \cdot \left(\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)\right)^2} \mathbf{P}_X(dx)$$
$$= \frac{\int K^2(z)\, dz}{h_n} \cdot \mathbf{E}\left\{ \frac{I_{[-\gamma_n,\gamma_n]^d}(X)}{\sum_{j=1}^n G\left(\frac{\|X-X_j\|}{H_n}\right)} \cdot I_{\left\{\sum_{j=1}^n G\left(\frac{\|X-X_j\|}{H_n}\right)>0\right\}} \right\}.$$

Application of Lemma 4.1 in Györif et al. (2002) and Lemma 3 yields

$$\mathbf{E}\left\{\frac{I_{[-\gamma_n,\gamma_n]^d}(X)}{\sum_{j=1}^n G\left(\frac{\|X-X_j\|}{H_n}\right)}\cdot I_{\left\{\sum_{j=1}^n G\left(\frac{\|X-X_j\|}{H_n}\right)>0\right\}}\right\}$$
$$\leq \int_{[-\gamma_n,\gamma_n]^d}\frac{2}{(n+1)\cdot\mathbf{P}_X(S_{H_n}(x))}\,\mathbf{P}_X(dx)\leq 2\cdot(4\cdot\sqrt{d})^d\cdot\frac{\gamma_n^d}{n\cdot H_n^d},$$

which completes the proof of (32).

In the second step of the proof we show

$$\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}-g_{Y|X}(y,x)\right|\,dy\,\mathbf{P}_X(dx).$$
$$\leq (4\cdot\sqrt{d})^d\cdot\frac{\gamma_n^d}{n\cdot H_n^d}+\int_{[-\gamma_n,\gamma_n]^d}|b_n(x)-a_n(x)|\,\mathbf{P}_X(dx)\cdot C_1\cdot H_n^r$$
$$+\int K(z)\cdot|z|^s dz\cdot\int_{[-\gamma_n,\gamma_n]^d}|b_n(x)-a_n(x)|\,\mathbf{P}_X(dx)\cdot C_2\cdot h_n^s. \quad (33)$$

Using the independence of the data and arguing similar as in the proof of inequality (32) we get

$$\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\mathbf{E}\{\hat{g}_{Y|X}(y,x)|X_1^n\}-g_{Y|X}(y,x)\right|\,dy\,\mathbf{P}_X(dx)$$
$$=\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right)\cdot\int_{\mathbb{R}}K\left(\frac{y-u}{h_n}\right)\cdot g_{Y|X}(u,X_i)\,du}{h_n\cdot\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)}\right.$$
$$\left.-g_{Y|X}(y,x)\right|\,dy\,\mathbf{P}_X(dx)$$
$$=\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}I_{\left\{\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)=0\right\}}g_{Y|X}(y,x)\,dy\,\mathbf{P}_X(dx)$$
$$+\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\sum_{i=1}^n\frac{G\left(\frac{\|x-X_i\|}{H_n}\right)}{\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)}\right.$$
$$\left.\cdot\int_{\mathbb{R}}\frac{1}{h_n}\cdot K\left(\frac{y-u}{h_n}\right)\cdot(g_{Y|X}(u,X_i)-g_{Y|X}(y,x))\,du\right|\,dy\,\mathbf{P}_X(dx)$$
$$\leq\int_{[-\gamma_n,\gamma_n]^d}\mathbf{P}\left\{\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)=0\right\}\mathbf{P}_X(dx)$$
$$+\mathbf{E}\int_{[-\gamma_n,\gamma_n]^d}\int_{[a_n(x),b_n(x)]}\left|\sum_{i=1}^n\frac{G\left(\frac{\|x-X_i\|}{H_n}\right)}{\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)}\right.$$

$$
\cdot \int_{\mathbb{R}} \frac{1}{h_n} \cdot K\left(\frac{y-u}{h_n}\right) \cdot (g_{Y|X}(u, X_i) - g_{Y|X}(y,x))\, du \Bigg|\, dy\, \mathbf{P}_X(dx)
$$

$$
\begin{aligned}
&\leq \int_{[-\gamma_n,\gamma_n]^d} (1 - \mathbf{P}_X(S_{H_n}(x)))^n\, \mathbf{P}_X(dx) \\
&\quad + \mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} \sum_{i=1}^n \frac{G\left(\frac{\|x-X_i\|}{H_n}\right)}{\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)} \\
&\qquad \cdot \int_{\mathbb{R}} \frac{1}{h_n} \cdot K\left(\frac{y-u}{h_n}\right) \cdot |g_{Y|X}(u, X_i) - g_{Y|X}(y,x))| du\, dy\, \mathbf{P}_X(dx).
\end{aligned}
$$

By Lemma 3 we get

$$
\begin{aligned}
\int_{[-\gamma_n,\gamma_n]^d} (1 - \mathbf{P}_X(S_{H_n}(x)))^n\, \mathbf{P}_X(dx) &\leq \max_{z \in \mathbb{R}_+} z \cdot e^{-z} \cdot \int_{[-\gamma_n,\gamma_n]^d} \frac{1}{n \cdot \mathbf{P}_X(S_{H_n}(x))}\, \mathbf{P}_X(dx) \\
&\leq (4 \cdot \sqrt{d})^d \cdot \frac{\gamma_n^d}{n \cdot H_n^d}.
\end{aligned}
$$

Furthermore, by triangle inequality and assumptions (26) and (27), which imply

$$
\begin{aligned}
|g_{Y|X}(u, X_i) - g_{Y|X}(y,x))| &\leq |g_{Y|X}(u, X_i) - g_{Y|X}(u,x)| + |g_{Y|X}(u,x) - g_{Y|X}(y,x)| \\
&\leq C_1 \cdot \|X_i - x\|^r + C_2 \cdot |y-u|^s
\end{aligned}
$$

we get

$$
\begin{aligned}
&\mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} \sum_{i=1}^n \frac{G\left(\frac{\|x-X_i\|}{H_n}\right)}{\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)} \\
&\qquad \cdot \int_{\mathbb{R}} \frac{1}{h_n} \cdot K\left(\frac{y-u}{h_n}\right) \cdot |g_{Y|X}(u, X_i) - g_{Y|X}(y,x))| du\, dy\, \mathbf{P}_X(dx) \\
&\leq \mathbf{E} \int_{[-\gamma_n,\gamma_n]^d} \sum_{i=1}^n \frac{G\left(\frac{\|x-X_i\|}{H_n}\right)}{\sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)} \cdot C_1 \cdot \|X_i - x\|^r \cdot |b_n(x) - a_n(x)|\, \mathbf{P}_X(dx) \\
&\quad + \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} \int_{\mathbb{R}} \frac{1}{h_n} \cdot K\left(\frac{y-u}{h_n}\right) \cdot C_2 \cdot |u-y|^s\, du\, dy\, \mathbf{P}_X(dx) \\
&\leq C_1 \cdot H_n^r \cdot \int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)|\, \mathbf{P}_X(dx) \\
&\quad + \int K(z) \cdot |z|^s dz \cdot C_2 \cdot h_n^s \cdot \int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)|\, \mathbf{P}_X(dx).
\end{aligned}
$$

Summarizing the above results we get the assertion. □

5.3 Proof of Theorem 1

In the proof of Theorem 1 we will use Lemma 1, Lemma 2 and the following auxiliary result from Bott, Felber and Kohler (2015).

Lemma 4 *Let $K : \mathbb{R} \to \mathbb{R}$ be a symmetric and bounded density which is monotonically decreasing on $\mathbb{R}_+$. Then it holds*

$$\int \left| K\left(\frac{y - z_1}{h_n}\right) - K\left(\frac{y - z_2}{h_n}\right) \right| dy \leq 2 \cdot K(0) \cdot |z_1 - z_2|$$

for arbitrary $z_1, z_2 \in \mathbb{R}$.

Proof. See Lemma 1 in Bott, Felber and Kohler (2015). □

Proof of Theorem 1. By Lemma 1 and Markov inequality it suffices to show

$$\mathbf{E}\left\{ \frac{1}{N_n} \sum_{i=1}^{N_n} |\hat{m}_{L_n}(X_{n+L_n+i}) - \bar{m}(X_{n+L_n+i})|^2 \right\} \leq \frac{\epsilon_n}{4}, \tag{34}$$

$$\mathbf{E}\left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx) \right\} \leq \frac{\delta_n}{4} \tag{35}$$

and

$$\mathbf{P}\left\{ \int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z,X) \, dz \neq 1 \right\} \leq \frac{\eta_n}{4}. \tag{36}$$

In the first step of the proof we observe that (34) is a trivial consequence of the independence of the data and the definition of ϵ_n.

In the second step of the proof we show (35). In case $\sum_{j=1}^n G\left(\frac{\|x - X_j\|}{H_n}\right) \neq 0$ we have that $\hat{g}_{\hat{\epsilon}|X}(\cdot, x)$ is a density, and we can conclude by the Lemma of Scheffé and triangle inequality

$$\begin{aligned}
&\int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \\
&\leq 2 \cdot \int_{\mathbb{R}} \left(g_{\bar{\epsilon}|X}(y,x) - \hat{g}_{\hat{\epsilon}|X}(y,x)\right)_+ \, dy \\
&\leq 2 \cdot \int_{[a_n(x), b_n(x)]} \left(g_{\bar{\epsilon}|X}(y,x) - \hat{g}_{\hat{\epsilon}|X}(y,x)\right)_+ \, dy + 2 \cdot \int_{[a_n(x), b_n(x)]^c} g_{\bar{\epsilon}|X}(y,x) \, dy \\
&\leq 2 \cdot \int_{[a_n(x), b_n(x)]} \left|g_{\bar{\epsilon}|X}(y,x) - \hat{g}_{\hat{\epsilon}|X}(y,x)\right| \, dy + 2 \cdot \int_{[a_n(x), b_n(x)]^c} g_{\bar{\epsilon}|X}(y,x) \, dy.
\end{aligned}$$

In case $\sum_{j=1}^n G\left(\frac{\|x - X_j\|}{H_n}\right) = 0$ we have

$$\hat{g}_{\hat{\epsilon}|X}(y,x) = 0 \quad \text{for all } y \in \mathbb{R},$$

and the above sequence of inequalities does trivially hold.

Using this we get

$$\mathbf{E}\left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx) \right\}$$

$$
\begin{aligned}
&\leq 2 \cdot \mathbf{E}\left\{\int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx)\right\} \\
&\quad +2 \cdot \mathbf{P}_X(\mathbb{R}^d \setminus [-\gamma_n,\gamma_n]^d) + 2 \cdot \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]^c} g_{\bar{\epsilon}|X}(y,x) \, dy \, \mathbf{P}_X(dx) \\
&\leq 2 \cdot \mathbf{E}\left\{\int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} |\hat{g}_{\hat{\epsilon}|X}(y,x) - \hat{g}_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx)\right\} \\
&\quad +2 \cdot \mathbf{E}\left\{\int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} |\hat{g}_{\bar{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx)\right\} \\
&\quad +2 \cdot \mathbf{P}_X(\mathbb{R}^d \setminus [-\gamma_n,\gamma_n]^d) + 2 \cdot \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]^c} g_{\bar{\epsilon}|X}(y,x) \, dy \, \mathbf{P}_X(dx),
\end{aligned}
$$

where

$$
\hat{g}_{\bar{\epsilon}|X}(y,x) = \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right) \cdot K\left(\frac{y-(Y_i-\bar{m}(X_i))}{h_n}\right)}{h_n \cdot \sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)}.\}
$$

Application of Lemma 4 yields

$$
\begin{aligned}
&\int_{[a_n(x),b_n(x)]} |\hat{g}_{\hat{\epsilon}|X}(y,x) - \hat{g}_{\bar{\epsilon}|X}(y,x)| \, dy \\
&\leq \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right)}{h_n \cdot \sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)} \cdot \int_{\mathbb{R}} \left| K\left(\frac{y-(Y_i-\hat{m}_{L_n}(X_i))}{h_n}\right) - K\left(\frac{y-(Y_i-\bar{m}(X_i))}{h_n}\right)\right| dy \\
&\leq 2 \cdot K(0) \cdot \frac{\sum_{i=1}^n G\left(\frac{\|x-X_i\|}{H_n}\right) \cdot |\hat{m}_{L_n}(X_i) - \bar{m}(X_i)|}{h_n \cdot \sum_{j=1}^n G\left(\frac{\|x-X_j\|}{H_n}\right)} \\
&\leq \frac{2 \cdot K(0)}{h_n} \cdot \sum_{i=1}^n \frac{|\hat{m}_{L_n}(X_i) - \bar{m}(X_i)|}{\left(1 + \sum_{j \in \{1,\dots,n\} \setminus \{i\}} G\left(\frac{\|x-X_j\|}{H_n}\right)\right)} |,
\end{aligned}
$$

where the last inequality followed from the fact that G is the naive kernel. Using this together with the independence of the data, Lemma 4.1 in Györfi et al. (2002) and Lemma 3 we get

$$
\begin{aligned}
&\mathbf{E}\left\{\int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} |\hat{g}_{\hat{\epsilon}|X}(y,x) - \hat{g}_{\epsilon|X}(y,x)| \, dy \, \mathbf{P}_X(dx)\right\} \\
&\leq \frac{2 \cdot K(0)}{h_n} \cdot \int_{[-\gamma_n,\gamma_n]^d} \sum_{i=1}^n \mathbf{E}\left\{\frac{1}{\left(1 + \sum_{j \in \{1,\dots,n\} \setminus \{i\}} G\left(\frac{\|x-X_j\|}{H_n}\right)\right)}\right\} \\
&\quad \cdot \mathbf{E} \int_{\mathbb{R}^d} |\hat{m}_{L_n}(x) - \bar{m}(x)| \, \mathbf{P}_X(dx)
\end{aligned}
$$

$$\leq \frac{2 \cdot K(0) \cdot (4 \cdot \sqrt{d})^d \cdot \gamma_n^d}{h_n \cdot H_n^d} \cdot \mathbf{E} \int_{\mathbb{R}^d} |\hat{m}_{L_n}(x) - \bar{m}(x)| \, \mathbf{P}_X(dx).$$

Application of Lemma 2 yields

$$\mathbf{E}\left\{ \int_{[-\gamma_n,\gamma_n]^d} \int_{[a_n(x),b_n(x)]} |\hat{g}_{\hat{\epsilon}|X}(y,x) - g_{\bar{\epsilon}|X}(y,x)| \, dy \, \mathbf{P}_X(dx) \right\}$$
$$\leq c_1 \cdot \left(\sqrt{\frac{\int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx) \cdot \gamma_n^d}{n \cdot H_n^d \cdot h_n}} \right.$$
$$\left. + \frac{\gamma_n^d}{n \cdot H_n^d} + \int_{[-\gamma_n,\gamma_n]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx) \cdot (C_1 \cdot H_n^\alpha + C_2 \cdot h_n^r) \right).$$

Summarizing the above results, the proof of (35) is complete.

In the third step of the proof we show (36). As in the proof of Lemma 2 we get

$$\mathbf{P}\left\{ \int_{\mathbb{R}} \hat{g}_{\hat{\epsilon}|X}(z,X) \, dz \neq 1 \right\}$$
$$= \mathbf{P}\left\{ \sum_{j \in \{1,\ldots,n\}} G\left(\frac{\|X - X_j\|}{H_n} \right) = 0 \right\}$$
$$\leq \mathbf{P}\left\{ X \in \mathbb{R}^d \setminus [-\gamma_n,\gamma_n]^d \right\} + \mathbf{P}\left\{ X \in [-\gamma_n,\gamma_n]^d, \sum_{j \in \{1,\ldots,n\}} G\left(\frac{\|X - X_j\|}{H_n} \right) = 0 \right\}$$
$$\leq \mathbf{P}\left\{ X \in \mathbb{R}^d \setminus [-\gamma_n,\gamma_n]^d \right\} + \frac{2 \cdot (4 \cdot \sqrt{d})^d \cdot \gamma_n^d}{n \cdot H_n^d}.$$

Summarizing the above results, the proof is complete. □

6 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 under Grant RGPIN 2015-06412.

References

[1] 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.

[2] Bott, A. K., Felber, T., and Kohler, M. (2015). Estimation of a density in a simulation model. *Journal of Nonparametric Statistics*, **27**, pp. 271-285.

[3] Bott, A. and Kohler, M. (2016). Adaptive estimation of a conditional density. *International Statistical Review* **84**, pp. 291-316.

[4] Bott, A. and Kohler, M. (2017). Nonparametric estimation of a conditional density. *Annals of the Institute of Statistical Mathematics* **69**, pp. 189-214.

[5] Bourinet, 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] 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] Devroye, L., Felber, T., and Kohler, M. (2013). Estimation of a density using real and artificial data. *IEEE Transactions on Information Theory*, **59**, No. 3, pp. 1917-1928.

[10] Efromovich, S. (2007). Conditional density estimation in a regression setting. *Annals of Statistics* **35**, pp. 2504–2535.

[11] 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.

[12] Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, pp. 189–206.

[13] Fan, J. and Yim, T. H. (2004). A crossvalidation method for estimating conditional densities. *Biometrika*, **91**, pp. 819–834.

[14] Felber, T., Kohler, M., and Krzyżak, A. (2015a). Adaptive density estimation based on real and artificial data. *Journal of Nonparametric Statistics*, **27**, pp. 1-18.

[15] Felber, T., Kohler, M., and Krzyżak, A. (2015b). Density estimation with small measurement errors. *IEEE Transactions on Information Theory*, **61**, pp. 3446-3456.

[16] Gooijer, J. G. D. and Zerom, D. (2003). On conditional density estimation. *Statistica Neerlandica*, **57**, pp. 159–176.

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

[18] Hurtado, J. (2004). *Structural Reliability – Statistical Learning Perspectives.* Vol. 17 of lecture notes in applied and computational mechanics. Springer.

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

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

[21] Kohler, M. and Krzyżak, A. (2016). Estimation of a density from an imperfect simulation model. Submitted for publication.

[22] Kohler, M. and Krzyżak, A. (2017a). Adaptive estimation of quantiles in a simulation model. To appear in *IEEE Transactions on Information Theory.*

[23] Kohler, M. and Krzyżak, A. (2017b). Improving a surrogate model in uncertainty quantification by real data. Submitted for publication.

[24] Kohler, M., Krzyżak, A., Mallapur, S., and Platz, R. (2016). Uncertainty Quantification in Case of Imperfect Models: A Non-Bayesian Approach. Submitted for publication.

[25] Mallapur, S., and Platz, R. (2017). Quantification and Evaluation of Uncertainty in the Mathematical Modelling of a Suspension Strut using Bayesian Model Validation Approach. Proceedings of the International Modal Analysis Conference IMAC-XXXV, Garden Grove, California, USA, Paper 117, 30. Jan - 2. Feb., 2017.

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

[27] 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.

[28] Parzen, E. (1962). On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, **33**, pp. 1065–1076.

[29] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, pp. 832–837.

[30] Rosenblatt, M. (1969). Conditional probability density and regression estimates. *Multivariate Analysis II* (Ed. P.R. Krishnaiah), pp. 25-31. Academic Press, New York.

[31] Wong, R. K. W., Storlie, C. B., and Lee, T. C. M. (2017). A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society, Series B*, **79**, pp. 635–648.

Supplementary material for the referees

Proof of Lemma 3. Partition $[-\gamma_n, \gamma_n]^d$ into

$$N = \left(\left\lceil \frac{2 \cdot \gamma_n}{H_n/\sqrt{d}} \right\rceil \right)^d \leq \left(\frac{4 \cdot \sqrt{d} \cdot \gamma_n}{H_n} \right)^d$$

many cubes $A_1, \ldots, A_N$ of side length at most $H_n/\sqrt{d}$. Let x_i be the center of A_i. Then $A_i \subseteq S_{H_n/2}(x_i)$ and $x \in S_{H_n/2}(x_i)$ implies $S_{H_n}(x) \supseteq S_{H_n/2}(x_i)$. Consequently we have

$$\int_{[-\gamma_n,\gamma_n]^d} \frac{1}{n \cdot \mathbf{P}_X(S_{H_n}(x))} \mathbf{P}_X(dx) \leq \sum_{i=1}^N \int_{S_{H_n/2}(x_i)} \frac{1}{n \cdot \mathbf{P}_X(S_{H_n}(x))} \mathbf{P}_X(dx)$$
$$\leq \sum_{i=1}^N \int_{S_{H_n/2}(x_i)} \frac{1}{n \cdot \mathbf{P}_X(S_{H_n/2}(x_i))} \mathbf{P}_X(dx) \leq \frac{N}{n},$$

which implies the assertion. □

Proof of Remark 2.

In what follows we show how to choose bandwidths h_n and H_n in order to minimize the expression (14). Here we ignore constants and logarithmic factors, so the aim is to minimize

$$\sqrt{\frac{A_n}{n \cdot H_n^d \cdot h_n}} + \frac{1}{n \cdot H_n^d} + A_n \cdot (C_1 \cdot H_n^r + C_2 \cdot h_n^s) \tag{37}$$

with respect to $h_n > 0$ and $H_n > 0$, where we have used the abbreviation

$$A_n = \int_{[-\log(n),\log(n)]^d} |b_n(x) - a_n(x)| \, \mathbf{P}_X(dx).$$

If (h_n, H_n) minimizes (37), then h_n satisfies

$$\sqrt{\frac{A_n}{n \cdot H_n^d \cdot h_n}} = A_n \cdot C_2 \cdot h_n^s.$$

The last equation is equivalent to

$$\frac{1}{C_2^2 \cdot A_n \cdot n \cdot H_n^d} = h_n^{2s+1},$$

from which we conclude

$$h_n = C_2^{-2/(2s+1)} \cdot A_n^{-1/(2s+1)} \cdot n^{-1/(2s+1)} \cdot H_n^{-d/(2s+1)}.$$

Plugging that bandwidth back into (37) yields

$$\frac{1}{n \cdot H_n^d} + A_n \cdot C_1 \cdot H_n^r + 2 \cdot A_n \cdot C_2 \cdot h_n^s$$

$$= \frac{1}{n \cdot H_n^d} + A_n \cdot C_1 \cdot H_n^r + 2 \cdot C_2^{1/(2s+1)} \cdot A_n^{(s+1)/(2s+1)} \cdot n^{-s/(2s+1)} \cdot H_n^{-d \cdot s/(2s+1)}. \tag{38}$$

So the optimal $H_n > 0$ minimizes (38), and it is easy to see, that a value $H_n > 0$ which minimizes (38) does indeed exist.

The optimal $H_n > 0$ minimizing (38) must satisfy

$$\frac{-d}{n} \cdot H_n^{-d-1} + r \cdot A_n \cdot C_1 \cdot H_n^{r-1} - \frac{d \cdot s}{2s+1} \cdot 2 \cdot C_2^{1/(2s+1)} \cdot A_n^{(s+1)/(2s+1)} \cdot n^{-s/(2s+1)} \cdot H_n^{-1-d \cdot s/(2s+1)} = 0,$$

which we can rewrite as

$$\frac{d}{n} \cdot H_n^{-d} + \frac{d \cdot s}{2s+1} \cdot 2 \cdot C_2^{1/(2s+1)} \cdot A_n^{(s+1)/(2s+1)} \cdot n^{-s/(2s+1)} \cdot H_n^{-d \cdot s/(2s+1)} - r \cdot A_n \cdot C_1 \cdot H_n^r = 0 \tag{39}$$

For the optimal H_n, either the first term on the left–hand side of (39) will be larger than the second one, or not. In the first case the optimal H_n lies between the solutions of

$$\frac{d}{n} \cdot H_n^{-d} = r \cdot A_n \cdot C_1 \cdot H_n^r$$

and of

$$2 \cdot \frac{d}{n} \cdot H_n^{-d} = r \cdot A_n \cdot C_1 \cdot H_n^r,$$

and in the second case it lies between the solutions of

$$\frac{d \cdot s}{2s+1} \cdot 2 \cdot C_2^{1/(2s+1)} \cdot A_n^{(s+1)/(2s+1)} \cdot n^{-s/(2s+1)} \cdot H_n^{-d \cdot s/(2s+1)} = r \cdot A_n \cdot C_1 \cdot H_n^r$$

and

$$2 \cdot \frac{d \cdot s}{2s+1} \cdot 2 \cdot C_2^{1/(2s+1)} \cdot A_n^{(s+1)/(2s+1)} \cdot n^{-s/(2s+1)} \cdot H_n^{-d \cdot s/(2s+1)} = r \cdot A_n \cdot C_1 \cdot H_n^r.$$

If we ignore again all constants we get that the optimal $H_n > 0$ either satisfies

$$\frac{1}{n \cdot H_n^d} = A_n \cdot C_1 \cdot H_n^r$$

or

$$C_2^{1/(2s+1)} \cdot A_n^{(s+1)/(2s+1)} \cdot n^{-s/(2s+1)} \cdot H_n^{-d \cdot s/(2s+1)} = A_n \cdot C_1 \cdot H_n^r.$$

In the first case we get

$$H_n^{r+d} = A_n^{-1} \cdot C_1^{-1} \cdot n^{-1},$$

which implies

$$H_n = A_n^{-1/(r+d)} \cdot C_1^{-1/(r+d)} \cdot n^{-1/(r+d)}, \tag{40}$$

and in the second case we get

$$H_n^{(r \cdot (2s+1)+d \cdot s)/(2s+1)} = C_1^{-1} \cdot A_n^{-s/(2s+1)} \cdot C_2^{1/(2s+1)} \cdot n^{-s/(2s+1)},$$

from which we get

$$H_n = C_1^{-\frac{2s+1}{r\cdot(2s+1)+d\cdot s}} \cdot C_2^{\frac{1}{r\cdot(2s+1)+d\cdot s}} \cdot A_n^{-\frac{s}{r\cdot(2s+1)+d\cdot s}} \cdot n^{-\frac{s}{r\cdot(2s+1)+d\cdot s}}. \tag{41}$$

Plugging (40) into (38) and ignoring again all constants yields as an upper bound on the error

$$\begin{aligned} & A_n \cdot C_1 \cdot H_n^r + C_2^{1/(2s+1)} \cdot A_n^{(s+1)/(2s+1)} \cdot n^{-s/(2s+1)} \cdot H_n^{-d\cdot s/(2s+1)} \\ &= C_1^{\frac{d}{r+d}} \cdot A_n^{\frac{d}{r+d}} \cdot n^{-\frac{r}{r+d}} + C_1^{\frac{ds}{(r+d)(2s+1)}} \cdot C_2^{\frac{1}{2s+1}} A_n^{\frac{(r+d)(s+1)+ds}{(r+d)(2s+1)}} \cdot n^{-\frac{rs}{(r+d)(2s+1)}}. \end{aligned} \tag{42}$$

And plugging (41) into (38) and ignoring again all constants yields as an upper bound on the error

$$\begin{aligned} & \frac{1}{n \cdot H_n^d} + A_n \cdot C_1 \cdot H_n^r \\ &= C_1^{\frac{(2s+1)d}{r(2s+1)+ds}} \cdot C_2^{-\frac{d}{r(2s+1)+ds}} \cdot A_n^{\frac{ds}{r(2s+1)+ds}} \cdot n^{-\frac{r(2s+1)}{r(2s+1)+ds}} \\ &\quad + C_1^{\frac{ds}{r(2s+1)+ds}} \cdot C_2^{\frac{r}{r(2s+1)+ds}} \cdot A_n^{\frac{r(s+1)+ds}{r(2s+1)+ds}} \cdot n^{-\frac{rs}{r(2s+1)+ds}}. \end{aligned} \tag{43}$$

From this we can conclude that (up to a logarithmic factor) the minimal value of (14) is given by the minimum of (42) and (43).