Estimation of a time–dependent density *

Ann-Kathrin Bott, Tina Felber and Michael Kohler†

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: abott@mathematik.tu-darmstadt.de, tfelber@mathematik.tu-darmstadt.de, kohler@mathematik.tu-darmstadt.de

August 22, 2014

Abstract

The problem of estimating a time–dependent density at each time point $t \in [0, 1]$ given independent samples of the density at discrete time points in $[0, 1]$ is considered. It is assumed that the distribution corresponding to the density of time t depends smoothly on t. The error of the estimate is measured pointwise by the L_1–error. Results concerning consistency and rate of convergence of a local average of kernel density estimates of the density at the discrete time points are presented. The finite sample size performance of the estimate is illustrated by applying it to simulated and real data.

AMS classification: Primary 62G07; secondary 62G20.

Key words and phrases: Density estimation, fixed design functional regression, functional response variable, L_1–error, time–dependent density, rate of convergence.

1 Introduction

Let $(X_t)_{t \in [0,1]}$ be an $\mathbb{R}$–valued stochastic process. For discrete time points $t_1, \ldots, t_N \in [0, 1]$ we assume that we have given independent samples

$$\mathcal{D}_N^{(t_k)} = \left\{ X_1^{(t_k)}, \ldots, X_{n_{k,N}}^{(t_k)} \right\} \tag{1}$$

of X_{t_k} $(k = 1, \ldots, N)$, and we are interested in estimating all marginal distributions μ_t of X_t $(t \in [0, 1])$.

If we try to reconstruct a distribution μ of some real–valued random variable X given an independent sample of X of size n, it is well–known that there does not exist any estimate $\hat{\mu}_n$ which is consistent with respect to the total variation error

$$\sup_{B \in \mathcal{B}} |\hat{\mu}_n(B) - \mu(B)|$$

(where $\mathcal{B}$ denotes the Borel sigma-field) for all distributions (cf. Devroye and Györfi (1990)). However, in case that a density f of μ with respect to the Lebesgue–Borel

*Running title: *Estimation of a time–dependent density*

†Corresponding author: Tel. +49-6151-16-6846, Fax. +49-6151-16-6822

measure exists, it is possible to construct universally L_1–consistent estimates of this density, i.e., there exist estimates f_n based on independent samples of f of size n satisfying

$$\int_{\mathbb{R}} |f_n(x) - f(x)|\, dx \to 0 \quad a.s.$$

no matter which form the underlying density f has (cf., e.g., Devroye (1983)). If we use such a sequence of density estimates to define corresponding distribution estimates via

$$\hat{\mu}_n(B) = \int_B f_n(x)\, dx \quad (B \in \mathcal{B}),$$

then the Lemma of Scheffé (cf., e.g., Theorem 1 in Chapter 1 of Devroye and Györfi (1985)) implies

$$\sup_{B \in \mathcal{B}} |\hat{\mu}_n(B) - \mu(B)| = \frac{1}{2} \cdot \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx$$

and hence $\hat{\mu}_n$ is consistent in total variation for all distributions which have a density with respect to the Lebesgue-Borel measure.

In the sequel we assume that X_t has a density $f(\cdot, t) : \mathbb{R} \to \mathbb{R}$ with respect to the Lebesgue–Borel measure for all $t \in [0, 1]$. The task is to construct distribution estimates $\hat{\mu}_{N,t}$ of μ_t which depend only on the data sets (1) for $k = 1, \dots, N$ such that

$$\sup_{B \in \mathcal{B}} |\hat{\mu}_{N,t}(B) - \mu_t(B)|$$

is "small" for every $t \in [0, 1]$. Again, this can be achieved by the Lemma of Scheffé if we construct density estimates $f_N(\cdot, t)$ such that

$$\int_{\mathbb{R}} |f_N(x, t) - f(x, t)|\, dx$$

is "small" for every $t \in [0, 1]$.

We proceed in two steps in order to construct such estimates: First we define standard kernel density estimates (cf., e.g., Rosenblatt (1956), Parzen (1962)) of $f(\cdot, t_k)$ using the data set (1), i.e., we define

$$f_N^{(t_k)}(x) = \frac{1}{n_{k,N} \cdot h_{k,N}} \cdot \sum_{i=1}^{n_{k,N}} K\left(\frac{x - X_i^{(t_k)}}{h_{k,N}}\right) \quad (x \in \mathbb{R}) \tag{2}$$

with some kernel function $K : \mathbb{R} \to \mathbb{R}$, which is a density (e.g., the naive kernel $K(u) = 1/2 \cdot \mathbb{1}_{[-1,1]}(u)$), and some bandwidth $h_{k,N} > 0$, which is the smoothing parameter of the estimate $f_N^{(t_k)}$. Then we use local averaging of these density estimates in order to define an estimate $f_N(\cdot, t)$ of $f(\cdot, t)$ for arbitrary $t \in [0, 1]$. Here we choose another kernel $H : \mathbb{R} \to \mathbb{R}$ (e.g., again the naive kernel) and a bandwidth $h_N > 0$ and define

$$f_N(x, t) = \frac{\sum_{k=1}^N H\left(\frac{t - t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot f_N^{(t_k)}(x)}{\sum_{j=1}^N H\left(\frac{t - t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \quad (x \in \mathbb{R},\ t \in [0, 1]). \tag{3}$$

In the sequel we show that under suitable conditions on the bandwidth the estimate f_N is pointwise L_1–consistent for all densities which depend smoothly on t. Furthermore, we analyze the rate of convergence of this estimate and present a data–driven method to choose the bandwidth h_N.

Our estimate can be regarded as a kernel estimate applied to the functional dataset

$$\left(t_1, \frac{1}{n_{1,N} h_{1,N}} \sum_{i=1}^{n_{1,N}} K\left(\frac{x - X_i^{(t_1)}}{h_{1,N}}\right)\right), \dots, \left(t_N, \frac{1}{n_{N,N} h_{N,N}} \sum_{i=1}^{n_{N,N}} K\left(\frac{x - X_i^{(t_N)}}{h_{N,N}}\right)\right).$$

For an introduction to functional data analysis we refer to the monographs Ramsay and Silverman (1997, 2002, 2005), Ferraty and View (2006) and Ferraty and Romain (2011). There exists a vast literature on functional nonparametric regression when the response variable is scalar (see, e.g., Burba, Ferraty and Vieu (2009), Ferraty et al. (2010) and Masry (2005) and the literature cited therein). But there are very few results for the current case when the response variable is functional, cf., e.g., Bosq and Delecroix (1985), Lecoutre (1990) and Ferraty et al. (2011). The case of a functional response variable, which is a density, is not considered in these articles.

A comprehensive introduction in L_1–consistent density estimation can be found in Devroye and Györfi (1985), Devroye (1987) and Devroye and Lugosi (2001).

Throughout the paper the following notation is used: The sets of natural numbers and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$, resp. $\mathcal{B}$ denotes the set of all Borel sets in $\mathbb{R}$ and $\mathbb{1}_B$ denotes the indicator function of the set B.

The outline of the paper is as follows: The main results are presented in Section 2, Section 3 illustrates the estimate by applying it to simulated and real data, and Section 4 contains the proofs.

2 Main results

2.1 Consistency

In our first theorem we present sufficient conditions for the consistency of our estimate.

Theorem 1 *Let $(X_t)_{t\in[0,1]}$ be an $\mathbb{R}$–valued stochastic process such that X_t has a density $f(\cdot,t) : \mathbb{R} \to \mathbb{R}$ with respect to the Lebesgue–Borel measure. For $N \in \mathbb{N}$ let $t_1, \dots, t_N \in [0,1]$, $h_N > 0$ and let K be a symmetric density satisfying*

$$\int_{\mathbb{R}} K^2(u)\, du < \infty \quad \text{and} \quad \int_{\mathbb{R}} |u| \cdot K(u)\, du < \infty.$$

For each $k \in \{1, \dots, N\}$ let $f_N^{(t_k)} = f_N^{(t_k)}(\cdot, \mathcal{D}_N^{(t_k)})$ be the estimate of $f(\cdot, t_k)$ defined by (2). Let $H : \mathbb{R} \to \mathbb{R}$ be a nonnegative bounded function with compact support satisfying $H \geq c_1 \cdot \mathbb{1}_{(-\delta,\delta)}$ for some $\delta, c_1 > 0$ and define the estimate $f_N(\cdot,t)$ of $f(\cdot,t)$ by (3). Assume that for all $k \in \{1, \dots, N\}$

$$X_{t_k}, X_1^{(t_k)}, \dots, X_{n_{k,N}}^{(t_k)} \text{ are independent and identically distributed} \tag{4}$$

and that

$$X_1^{(t_1)}, \ldots, X_{n_{1,N}}^{(t_1)}, \ldots, X_1^{(t_N)}, \ldots, X_{n_{N,N}}^{(t_N)} \text{ are independent.} \tag{5}$$

Assume furthermore that

$$h_N \to 0 \quad (N \to \infty), \tag{6}$$

$$\inf_{k \in \{1,\ldots,N\}} \frac{|t - t_k|}{h_N} \to 0 \quad (N \to \infty) \quad \text{for every } t \in [0,1], \tag{7}$$

$$\sum_{j=1}^{N} H\left(\frac{t - t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N} \to \infty \quad (N \to \infty) \quad \text{for every } t \in [0,1] \tag{8}$$

and

$$\max_{j \in \{1,\ldots,N\}} h_{j,N} \to 0 \quad (N \to \infty). \tag{9}$$

Then the estimate f_N satisfies

$$\mathbf{E} \int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\, dx \to 0 \quad (N \to \infty) \tag{10}$$

for every $t \in [0,1]$ and for all functions f such that $f(\cdot, t)$ is a density for all $t \in [0,1]$ and such that

$$\sup_{s,t \in [0,1], |s-t| < \delta} \int |f(x,s) - f(x,t)| dx \to 0 \quad (\delta \to 0). \tag{11}$$

If, in addition,

$$\sum_{N=1}^{\infty} \exp\left(-\frac{\delta^2}{2} \frac{\left(\sum_{k=1}^{N} H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N}\right)^2}{\sum_{l=1}^{N} H^2\left(\frac{t-t_l}{h_N}\right) \cdot n_{l,N} \cdot h_{l,N}^2} \right) < \infty \tag{12}$$

for every $\delta > 0$, then the convergence in (10) holds even almost surely.

Remark 1. In case $t_k = k/N$ and $n_{k,N} = n_N$ $(k = 1, \ldots, N)$ conditions (6)–(9) are satisfied provided we choose h_N and $h_{k,N} = \tilde{h}_N$ $(k = 1, \ldots, N)$ such that

$$h_N \to 0 \quad (N \to \infty), \quad \tilde{h}_N \to 0 \quad (N \to \infty), \quad N \cdot h_N \to \infty \quad (N \to \infty)$$

and

$$N \cdot h_N \cdot n_N \cdot \tilde{h}_N \to \infty \quad (N \to \infty).$$

And in this case

$$\frac{\left(\sum_{k=1}^{N} H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N}\right)^2}{\sum_{l=1}^{N} H^2\left(\frac{t-t_l}{h_N}\right) \cdot n_{l,N} \cdot h_{l,N}^2} \geq n_N$$

implies that (12) holds whenever

$$\sum_{N=1}^{\infty} \exp\left(-\frac{\delta^2}{2} \cdot n_N\right) < \infty.$$

2.2 Rate of convergence

Next we study the rate of convergence of our estimate. For simplicity we consider here only equidistant time points, where for each time point we have the same sample size. Furthermore, we choose the bandwidths independent of the time point and H as the naive kernel.

Theorem 2 *Let $(X_t)_{t\in[0,1]}$ be an $\mathbb{R}$-valued stochastic process such that X_t has a density $f(\cdot,t):\mathbb{R}\to\mathbb{R}$ with respect to the Lebesgue-Borel measure. Let $N\in\mathbb{N}$, set $t_k=k/N$ $(k=1,\ldots,N)$, let $h_N>0$ and let K be a symmetric density satisfying*

$$\int_{\mathbb{R}} K^2(u)\,du<\infty \quad \text{and} \quad \int_{\mathbb{R}} |u|\cdot K(u)\,du<\infty.$$

Assume that the sample sizes of the data sets (1) are equal, i.e., assume that

$$n_{1,N}=\cdots=n_{N,N}=n_N.$$

Let $\tilde{h}_N>0$. For each $k\in\{1,\ldots,N\}$ let $f_N^{(t_k)}=f_N^{(t_k)}(\cdot,\mathcal{D}_N^{(t_k)})$ be defined by

$$f_N^{(t_k)}(x)=\frac{1}{n_N\cdot\tilde{h}_N}\cdot\sum_{i=1}^{n_N}K\left(\frac{x-X_i^{(t_k)}}{\tilde{h}_N}\right)\quad (x\in\mathbb{R}).$$

Let $H=1/2\cdot\mathbb{1}_{[-1,1]}$ be the naive kernel and define the estimate $f_N(\cdot,t)$ of $f(\cdot,t)$ by (3) for some $h_N\geq 1/N$.
Assume that there exists a compact set $B\in\mathcal{B}$ such that

$$f(x,t)=0 \quad \text{for all } x\notin B \text{ and all } t\in[0,1], \tag{13}$$

that $f(\cdot,t)$ is Hölder continuous with exponent $r\in(0,1]$ and with Hölder constant $C>0$ for all $t\in[0,1]$, i.e.,

$$|f(x,t)-f(y,t)|\leq C\cdot|x-y|^r \quad \text{for all } x,y\in\mathbb{R} \text{ and all } t\in[0,1], \tag{14}$$

and that $f(x,\cdot)$ is Hölder continuous with exponent $p\in(0,1]$ and with Hölder constant $C>0$ for all $x\in\mathbb{R}$, i.e.,

$$|f(x,s)-f(x,t)|\leq C\cdot|s-t|^p \quad \text{for all } s,t\in[0,1] \text{ and all } x\in\mathbb{R}. \tag{15}$$

Assume furthermore that for all $k\in\{1,\ldots,N\}$

$$X_{t_k},X_1^{(t_k)},\ldots,X_{n_N}^{(t_k)} \text{ are independent and identically distributed} \tag{16}$$

and that

$$X_1^{(t_1)},\ldots,X_{n_N}^{(t_1)},\ldots,X_1^{(t_N)},\ldots,X_{n_N}^{(t_N)} \text{ are independent.} \tag{17}$$

Then we have

$$\mathbf{E}\int |f_N(x,t) - f(x,t)|\,dx = O\left(\sqrt{\frac{1}{N\cdot h_N\cdot n_N\cdot \tilde{h}_N}} + \tilde{h}_N^r + h_N^p\right) \tag{18}$$

for every $t \in [0,1]$*. In particular, if we set* $h_N = c_2\cdot(N\cdot n_N)^{-\frac{r}{(2r+1)\cdot p+r}}$ *and* $\tilde{h}_N = c_3\cdot(N\cdot n_N)^{-\frac{p}{(2r+1)\cdot p+r}}$ *we get*

$$\mathbf{E}\int |f_N(x,t) - f(x,t)|\,dx = O\left((N\cdot n_N)^{-\frac{p\cdot r}{(2r+1)\cdot p+r}}\right)$$

for every $t \in [0,1]$.

2.3 Data–dependent choice of the bandwidth

In this section we present a data–driven method to choose the bandwidth h_N of our estimate f_N. Our aim is to to choose the bandwidth such that the L_1–error of the estimate $f_N(\cdot,t)$ is small for any $t \in [0,1]$, and to achieve this we try to choose the bandwidth such that the maximal L_1–error of $f_N(\cdot,t)$ is small for $t \in \{t_1,\ldots,t_N\}$. To do this, we adapt the combinatorial method of Devroye and Lugosi (2001) to our setting. First, we fix a vector of sample sizes

$$m_N = (m_{1,N},\ldots,m_{N,N}) \in \mathbb{Z}_+^N \quad \text{with} \quad m_{k,N} \le \frac{n_{k,N}}{2}, \qquad k = 1,\ldots,N,$$

and define a class of density estimates $\mathcal{F} = \left\{\tilde{f}_{m_N,h_N} : h_N > 0\right\}$, where

$$\tilde{f}_{m_N,h_N}(x,t) = \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot m_{k,N}\cdot h_{k,N}\cdot f_{m_{k,N}}^{(t_k)}(x)}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot m_{j,N}\cdot h_{j,N}}$$

is our estimate (3) based on standard kernel density estimates

$$f_{m_{k,N}}^{(t_k)}(x) = \frac{1}{m_{k,N}\cdot h_{k,N}}\cdot\sum_{i=1}^{m_{k,N}} K\left(\frac{x - X_i^{(t_k)}}{h_{k,N}}\right) \qquad (x \in \mathbb{R})$$

evaluated on the first parts of the data sets (1). Here, we choose H and K as naive kernels and $h_{k,N}$ are the smoothing parameters, which can be determined, e.g., by cross-validation or by the combinatorial method of Devroye and Lugosi (2001). Then we project the empirical measure $\mu_{m_{k,N}}$ defined on the held–out data, i.e.,

$$\mu_{m_{k,N}}(A) = \frac{1}{n_{k,N} - m_{k,N}}\cdot\sum_{i=m_{k,N}+1}^{n_{k,N}} \mathbb{1}_A\left(X_i^{(t_k)}\right),$$

on $\mathcal{F}$. For this purpose, we define the sets

$$\mathcal{A}_{t_k} = \left\{ \left\{ x \in \mathbb{R} : \tilde{f}_{m_N,h_1}(x,t_k) > \tilde{f}_{m_N,h_2}(x,t_k) \right\} : h_1, h_2 > 0 \right\},$$

$k = 1, \ldots, N$, for fixed N, and choose the bandwidth $\hat{h}_N$ by minimizing the distance

$$\Delta_{h_N} := \max_{k \in \{1,\ldots,N\}} \sup_{A \in \mathcal{A}_{t_k}} \left| \int_A \tilde{f}_{m_N,h_N}(x,t_k)\ dx - \mu_{m_{k,N}}(A) \right|$$

over all $h_N > 0$. Since it is possible that the minimum of Δ_{h_N} does not exist, we select $\hat{h}_N > 0$ such that

$$\Delta_{\hat{h}_N} < \inf_{h_N > 0} \Delta_{h_N} + \frac{1}{N}$$

and define

$$\hat{f}_N = \tilde{f}_{m_N,\hat{h}_N}.$$

The next theorem states that the expected maximal L_1–error of this estimate is almost as small as that of the best estimate in $\mathcal{F}$.

Theorem 3 *Let $\hat{f}_N$ be defined as above. Then*

$$\begin{aligned} &\mathbf{E}\left\{ \max_{k \in \{1,\ldots,N\}} \int \left| \hat{f}_N(x,t_k) - f(x,t_k) \right|\ dx \right\} \\ &\leq 3\,\mathbf{E}\left\{ \inf_{h_N > 0} \max_{k \in \{1,\ldots,N\}} \int \left| \tilde{f}_{m_N,h_N}(x,t_k) - f(x,t_k) \right|\ dx \right\} \\ &\quad + 32 \cdot \sqrt{\frac{\log\left(2^{20} \cdot N^{17} \cdot \max_{k \in \{1,\ldots,N\}} m_{k,N}^{16}\right)}{\min_{k \in \{1,\ldots,N\}} m_{k,N}}} + \frac{7}{N}. \end{aligned}$$

3 Application to simulated and real data

In this section we apply our estimate to real and simulated data.

In the next three examples we assume that we have given $N = 101$ equidistant time points and $n = 20$ data points $X_1^{(t_k)}, \ldots, X_{20}^{(t_k)}$ for each time point $t_k \in [0,1]$, $k = 1, \ldots, 101$. We use the Gaussian kernel for the kernel functions H and K. We set $m_{k,n} = \lfloor \frac{n}{2} \rfloor = 10$ for every $k = 1, \ldots, 101$. We use the same bandwidth $\tilde{h}_N = h_{1,N} = \cdots = h_{N,N}$ for each time point and choose the bandwidths (H_N, h_N) of our estimate f_N from the set

$$\{0.05, 0.1, 0.2, 0.4, 0.8\} \times \{0.01, 0.05, 0.1, 0.25, 0.5, 1\}$$

via the data-dependent method described in Section 2.3. Then we compare the proposed estimate with the estimate of Rosenblatt and Parzen which can only be calculated for the observed time points t_k, $k = 1, \ldots, 101$, using for each time point all $n = 20$ data points corresponding to the time point. For this estimate the bandwidth is chosen by cross-validation. In our newly proposed algorithm all integrals are approximated by Riemann sums.

In our first example we choose $(X_t)_{t\in[0,1]}$ as exponentially distributed with variable rate $1.5-t$. Since the result of our simulation depends on the randomly occuring data points, we repeat the whole procedure 100 times with independent realizations of the occuring random variables for each observed time point. Figure 1 shows boxplots of the average and maximal occured L_1–error for each time point. The mean of the average (with respect to the 101 time points) L_1–errors of the proposed estimate (0.3018) is less than the mean average L_1–error of the Rosenblatt-Parzen density estimate (0.5568). In addition, the mean of the maximal (with respect to the 101 time points) L_1–errors of our estimate (0.4736) is lower than the mean of the maximal L_1–errors of the Rosenblatt and Parzen estimate (1.0909).

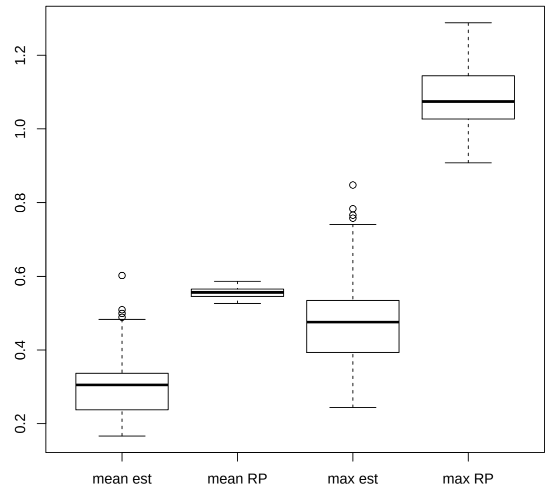

Figure 1: Boxplots of the average and maximal occuring L_1–errors in the first model

In our second example we choose $(X_t)_{t\in[0,1]}$ as weibull distributed with shape parameter 1 and variable scale parameter $1+t$. In Figure 2 we compare again boxplots of the average and maximal occuring L_1–errors of the two estimates. The mean of the average (0.3252) and the maximal (0.4652) L_1–error of our estimate are much lower than the ones of Rosenblatt and Parzen (0.4725 and 0.8627).
In Figure 3 we repeat the same procedure with gamma distributed process $(X_t)_{t\in[0,1]}$ with shape parameter $0.5+t$ and scale parameter 2. The mean of the average (0.2499) and the maximal L_1–errors (0.3686) are again lower than the mean average (0.3099) and mean maximal L_1–error (0.7123) of the Rosenblatt–Parzen estimate.

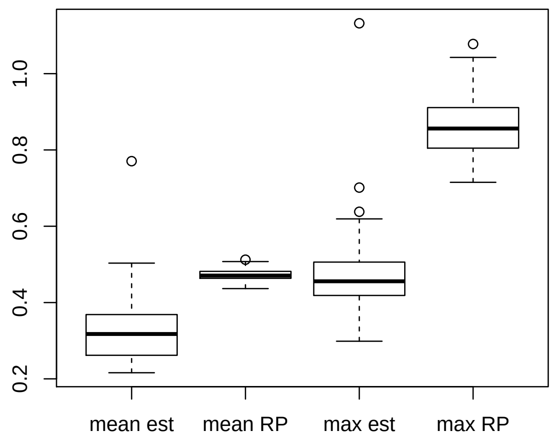

Figure 2: Boxplots of the average and maximal occuring L_1–errors in the second model

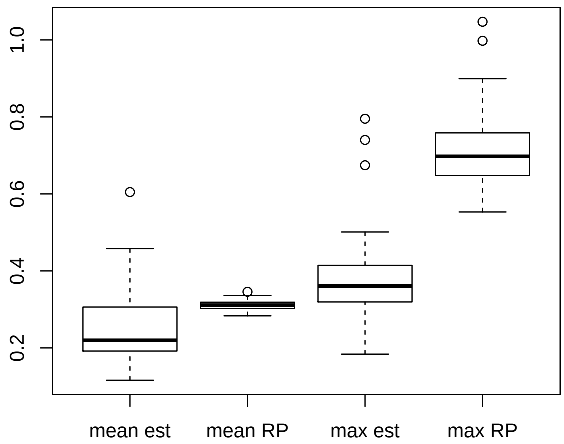

Figure 3: Boxplots of the average and maximal occuring L_1–errors in the third model

Finally, we illustrate the usefulness of our estimate by applying it to real data gained by the Collaborative Research Foundation 805 which is interested in the measurement of uncertainty in load-bearing systems like bearing structures of aeroplanes. We consider the load distribution in the three legs of a tripod (Figure 4), which is an example of a load-bearing system in mechanical engineering. In the experiments a static force is applied on the tripod. On the bottom side of the legs force sensors are mounted to measure the legs' axial force. If the holes where the legs are plugged in have exactly the same diameter, a third of the general load should be measured in each leg. Unfortunately, such

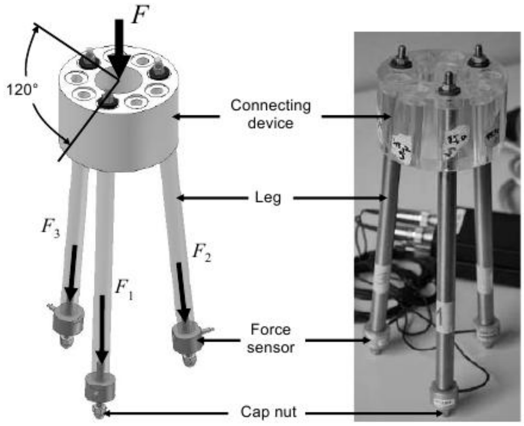

Figure 4: Tripod

an accurate drilling is not possible in the manufacturing process. Since there is always a small deviation, the force is distributed nonuniformly in the three legs. Due to wear engineers expect that the diameters expand over time. To examine the influence of a larger diameter to the force distribution in the three legs, we assume that the diameter in the first leg behaves like a standard normally distributed process expand over time and behave like independent normally distributed processes $(X_t^{(1)})_{t\in[0,1]}$ with expectation $15 + 2 \cdot t$ (in mm) and standard deviation 0.5. The diameters of the other two legs are independet and normally distributed with expectation 15 and standard deviaton 0.5. Based on the physical model of the tripod we are able to calculate the resulting load in each leg in dependence of the values of the three diameters at time $t \in [0, 1]$. For simplicity, we consider only the first leg of the tripod. To estimate the density of the resulting load in the first leg of the tripod we calculate our density estimate as described before using $n_k = 20$ measurements at each of the 101 equidistant time points. The bandwidths (H_N, h_N) of the estimate is selected by our data-driven method from the set

$$\{0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15, 0.2\} \times \\ \{0.00001, 0.00015, 0.0002, 0.00025, 0.00003, 0.00005, 0.0001\}$$

and is selected as $H_N = 0.15$ and $\tilde{h}_N = 0.00015$. Figure 5 shows how the estimated density changes over time. From the figure it can be seen that less force is distributed into the first leg as the diameter increases.

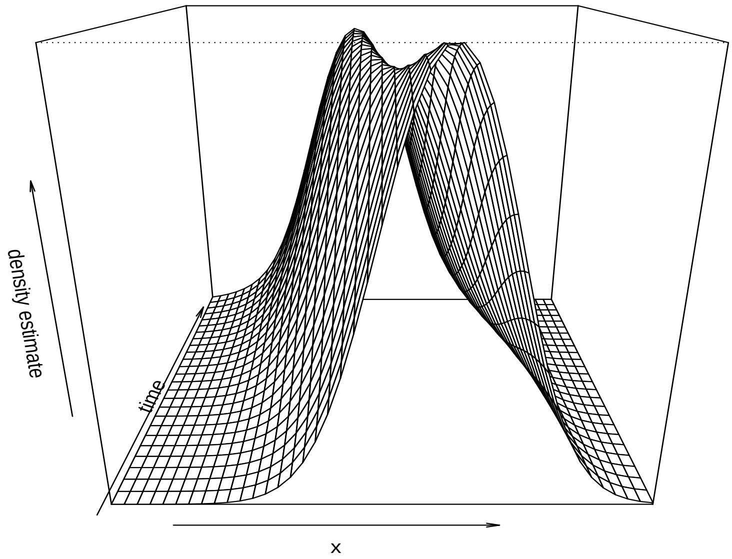

Figure 5: Density estimate for the tripod

4 Proofs

4.1 Proof of Theorem 1

Let $\epsilon > 0$ be arbitary. Because of (11) we can choose $\tilde{K} \in \mathbb{N}$ and $s_1, \ldots, s_{\tilde{K}} \in [0,1]$ such that

$$\min_{k \in \{1,\ldots,\tilde{K}\}} \int_{\mathbb{R}} |f(x,t) - f(x,s_k)|\, dx < \epsilon \tag{19}$$

for all $t \in [0,1]$. For $k \in \{1,\ldots,\tilde{K}\}$ let g_k be a continuously differentiable density with compact support satisfying

$$\int_{\mathbb{R}} |g_k(x) - f(x,s_k)|\, dx < \epsilon$$

and let B be the (compact) union of the supports of $g_1, \ldots, g_{\tilde{K}}$. Set $z_+ = \max\{z, 0\}$ for all $z \in \mathbb{R}$. Let $t \in [0,1]$ be arbitrary. Choose $\tilde{k}_t \in \{1,\ldots,\tilde{K}\}$ such that

$$\int_{\mathbb{R}} |f(x,t) - f(x,s_{\tilde{k}_t})|\, dx < \epsilon.$$

By the triangle inequality and the Lemma of Scheffé (cf., e.g., Theorem 1 in Chapter 1 of Devroye and Györfi (1985)) we have

$$
\begin{aligned}
&\mathbf{E}\int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\,dx \\
&< \mathbf{E}\int_{\mathbb{R}} |f_N(x,t) - f(x,s_{\tilde{k}_t})|\,dx + \epsilon \\
&< \mathbf{E}\int_{\mathbb{R}} |f_N(x,t) - g_{\tilde{k}_t}(x)|\,dx + 2\,\epsilon \\
&= 2\,\mathbf{E}\int_{\mathbb{R}} (g_{\tilde{k}_t}(x) - f_N(x,t))_+\,dx + 2\,\epsilon \\
&= 2\,\mathbf{E}\int_{B} (g_{\tilde{k}_t}(x) - f_N(x,t))_+\,dx + 2\,\epsilon \\
&\leq 2\,\mathbf{E}\int_{B} |f_N(x,t) - \mathbf{E}\left\{f_N(x,t)\right\}|\,dx + 2\int_{B} |\mathbf{E}\left\{f_N(x,t)\right\} - f(x,t)|\,dx \\
&\quad + 2\int_{B} \left|f(x,t) - f(x,s_{\tilde{k}_t})\right|\,dx + 2\int_{B} |f(x,s_{\tilde{k}_t}) - g_{\tilde{k}_t}(x)|\,dx + 2\,\epsilon \\
&< \mathbf{E}\{T_{1,N}\} + T_{2,N} + 6\,\epsilon,
\end{aligned}
$$

where

$$
T_{1,N} = 2\int_{B}\left|\frac{\sum_{k=1}^{N} H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot (f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\})}{\sum_{j=1}^{N} H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}}\right|\,dx
$$

and

$$
T_{2,N} = 2\int_{B}\left|f(x,t) - \frac{\sum_{k=1}^{N} H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot \mathbf{E}\{f_N^{(t_k)}(x)\}}{\sum_{j=1}^{N} H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}}\right|\,dx.
$$

First, we bound $\mathbf{E}\{T_{1,N}\}$. Let $|B|$ be the Lebesgue measure of the set B. By Cauchy-Schwarz inequality, Jensen inequality, theorem of Fubini and assumption (5), which implies that $f_N^{(t_j)}(x)$ and $f_N^{(t_k)}(x)$ are independent for $j \neq k$ and hence

$$
\mathbf{E}\left\{\left(f_N^{(t_j)}(x) - \mathbf{E}\{f_N^{(t_j)}(x)\}\right)\cdot\left(f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right)\right\} = 0
$$

for all $x \in \mathbb{R}$, $j,k \in \{1,\dots,N\}$ with $j \neq k$, we get

$$
\begin{aligned}
&\mathbf{E}\{T_{1,N}\} \\
&\leq 2\cdot\sqrt{|B|}\cdot\mathbf{E}\left\{\left(\int_{B}\left|\frac{\sum_{k=1}^{N} H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot\left(f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right)}{\sum_{j=1}^{N} H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}}\right|^2 dx\right)^{1/2}\right\}
\end{aligned}
$$

$$\leq 2 \cdot \sqrt{|B|} \cdot \left(\mathbf{E} \left\{ \int_B \left| \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot \left(f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right)}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \right|^2 dx \right\} \right)^{1/2}$$

$$= 2 \cdot \sqrt{|B|} \cdot \left(\frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)^2 \cdot n_{k,N}^2 \cdot h_{k,N}^2 \cdot \int_B \mathbf{E}\left\{ \left| f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right|^2 \right\} dx}{\left(\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}\right)^2} \right)^{1/2}.$$

Since H is nonnegative and bounded we have $H^2(z) \leq c_4 \cdot H(z)$ for all $z \in \mathbb{R}$. Furthermore, using (4) and the theorem of Fubini we get

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbf{E} \left| f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right|^2 dx \\
&\leq \int_{\mathbb{R}} \frac{1}{n_{k,N}} \cdot \frac{1}{h_{k,N}^2} \cdot \int_{\mathbb{R}} K^2\left(\frac{x-u}{h_{k,N}}\right) \cdot f(u,t_k)\, du\, dx \\
&= \frac{1}{n_{k,N} \cdot h_{k,N}} \cdot \int_{\mathbb{R}} f(u,t_k) \cdot \int_{\mathbb{R}} \frac{1}{h_{k,N}} \cdot K^2\left(\frac{x-u}{h_{k,N}}\right) dx\, du \\
&= \frac{\int_{\mathbb{R}} K^2(z)\, dz}{n_{k,N} \cdot h_{k,N}}. \qquad (20)
\end{aligned}$$

This implies

$$\begin{aligned}
\mathbf{E}\{T_{1,n}\} &\leq 2 \cdot \sqrt{|B|} \cdot \sqrt{c_4} \cdot \left(\frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot c_5}{\left(\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}\right)^2} \right)^{1/2} \\
&\leq \frac{c_6}{\sqrt{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}}}
\end{aligned}$$

for n sufficiently large.

Next we bound $T_{2,N}$. Because of (6) and the compact support of H we can conclude from (11) that $H((t-t_k)/h_N) > 0$ implies

$$\int |f(x,t) - f(x,t_k)|\, dx < \epsilon \qquad (21)$$

for all $t \in [0,1]$ for N sufficiently large. For $t \in [0,1]$ choose again $\tilde{k}_t \in \{1, \ldots, \tilde{K}\}$ such that $g^* = g_{\tilde{k}_t}$ satisfies

$$\int |f(x,t) - g^*(x)|\, dx \leq \int |f(x,t) - f(x, s_{\tilde{k}_t})|\, dx + \int |f(x, s_{\tilde{k}_t}) - g^*(x)|\, dx < 2\,\epsilon.$$

We have

$T_{2,N}$

$$
\begin{aligned}
&= 2\int_B \left| \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot \int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot f(u,t_k)\,du}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}} - f(x,t)\right| dx \\
&\leq 2\cdot \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot \int_B\int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot |f(u,t_k)-f(u,t)|\,du\,dx}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}} \\
&\quad + 2\int_B \left| \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot \int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot f(u,t)\,du}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}} - f(x,t)\right| dx \\
&\leq 2\cdot \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot \int_B\int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot |f(u,t_k)-f(u,t)|\,du\,dx}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}} \\
&\quad + 2\cdot \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot \int_B\int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot |f(u,t)-g^*(u)|\,du\,dx}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}} \\
&\quad + 2\int_B \left| \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N}\cdot \int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot g^*(u)\,du}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}} - g^*(x)\right| dx \\
&\quad + 2\int_B |g^*(x)-f(x,t)|\,dx \\
&= T_{3,N}+T_{4,N}+T_{5,N}+T_{6,N}.
\end{aligned}
$$

Using the theorem of Fubini and the fact that K is a density we see that

$$
\begin{aligned}
&\int_B\int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot |f(u,t_k)-f(u,t)|\,du\,dx \\
&= \int_{\mathbb{R}}\int_B \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\,dx\cdot |f(u,t_k)-f(u,t)|\,du \\
&\leq \int_{\mathbb{R}} |f(u,t_k)-f(u,t)|\,du
\end{aligned}
$$

and

$$
\int_B\int_{\mathbb{R}} \frac{1}{h_{k,N}}\cdot K\left(\frac{x-u}{h_{k,N}}\right)\cdot |f(u,t)-g^*(u)|\,du\,dx \leq \int_{\mathbb{R}} |f(u,t)-g^*(u)|\,du
$$

from which we conclude via (21) and the choice of g^* that we have $T_{3,N} < 2\,\epsilon$ and $T_{4,N} < 4\,\epsilon$. Furthermore, again by choice of g^* we have $T_{6,N} < 4\,\epsilon$. It remains to bound $T_{5,N}$. Since $g_1, \ldots, g_{\tilde{K}}$ are continuously differentiable with bounded support, they are also Lipschitz continuous. Denote the maximum of the Lipschitz constants of $g_1, \ldots, g_{\tilde{K}}$ by C. Because of (7) and $H \geq c_1 \cdot \mathbb{1}_{(-\delta,\delta)}$ we can assume w.l.o.g. that $\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)\cdot n_{k,N}\cdot h_{k,N} > 0$ and hence for N sufficiently large

$T_{5,N}$

$$\leq 2 \cdot \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot \int_B \int_{\mathbb{R}} \frac{1}{h_{k,N}} \cdot K\left(\frac{x-u}{h_{k,N}}\right) \cdot |g^*(u) - g^*(x)|\, du\, dx}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}}$$

$$\leq 2 \cdot \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot \int_B \int_{\mathbb{R}} \frac{1}{h_{k,N}} \cdot K\left(\frac{x-u}{h_{k,N}}\right) \cdot C \cdot |x-u|\, du\, dx}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}}$$

$$\leq 2 \cdot \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot C \cdot |B| \cdot h_{k,N} \cdot \int_{\mathbb{R}} |z|\, K(z)\, dz}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}}$$

$$\leq 2 \cdot C \cdot |B| \cdot \int_{\mathbb{R}} |z|\, K(z)\, dz \cdot \max_{k \in \{1,\ldots,N\}} h_{k,N}\,.$$

Summarizing the above results we see that for N sufficiently large

$$\mathbf{E} \int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\, dx$$
$$\leq \frac{c_6}{\sqrt{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}}} + c_7 \cdot \max_{k \in \{1,\ldots,N\}} h_{k,N} + 10\,\epsilon.$$

Because of (8) and (9) we conclude

$$\limsup_{N \to \infty} \mathbf{E} \int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\, dx \leq 10\,\epsilon,$$

and since $\epsilon > 0$ was arbitrary this implies (10).

In order to show the second assertion, we observe

$$\begin{aligned} \int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\ dx \quad &= \quad \int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\ dx - \mathbf{E}\left\{\int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\ dx\right\} \\ &\quad + \mathbf{E}\left\{\int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\ dx\right\} \end{aligned}$$

for every $t \in [0,1]$. The second term can be bounded like before. For the first term, let $z_1^{(t_k)}, \ldots, z_{n_{k,N}}^{(t_k)}, \bar{z}_{j_l}^{(t_l)} \in \mathbb{R}$ for $k = 1, \ldots, N$, $l = 1, \ldots, N$, $j_l = 1, \ldots, n_{l,N}$. Define

$$\begin{aligned} &g(z_1^{(t_1)}, \ldots, z_{n_{1,N}}^{(t_1)}, \ldots, z_1^{(t_N)}, \ldots, z_{n_{N,N}}^{(t_N)}) \\ &= \int_{\mathbb{R}} \left| \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \sum_{i=1}^{n_{k,N}} K\left(\frac{x - z_i^{(t_k)}}{h_{k,N}}\right)}{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N}} - f(x,t) \right| dx \end{aligned}$$

for every $t \in [0,1]$. With $||a| - |b|| \leq |a-b|$ for $a, b \in \mathbb{R}$ and the Lemma of Scheffé it holds

$$\left| g(z_1^{(t_1)}, \ldots, z_{j_l}^{(t_l)}, \ldots, z_{n_{N,N}}^{(t_N)}) - g(z_1^{(t_1)}, \ldots, z_{j_l-1}^{(t_l)}, \bar{z}_{j_l}^{(t_l)}, z_{j_l+1}^{(t_l)}, \ldots, z_{n_{N,N}}^{(t_N)}) \right|$$

$$\leq \int_{\mathbb{R}} \left| \frac{H\left(\frac{t-t_l}{h_N}\right) \cdot \left(K\left(\frac{x-z_{j_l}^{(t_l)}}{h_{l,N}}\right) - K\left(\frac{x-\bar{z}_{j_l}^{(t_l)}}{h_{l,N}}\right)\right)}{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N}} \right| dx$$

$$= \frac{2 \cdot H\left(\frac{t-t_l}{h_N}\right) \cdot h_{l,N}}{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N}} \int_{\mathbb{R}} \left(\frac{1}{h_{l,N}} K\left(\frac{x - z_{j_l}^{(t_l)}}{h_{l,N}}\right) - \frac{1}{h_{l,N}} K\left(\frac{x - \bar{z}_{j_l}^{(t_l)}}{h_{l,N}}\right)\right)_+ dx$$

$$\leq \frac{2 \cdot H\left(\frac{t-t_l}{h_N}\right) \cdot h_{l,N}}{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N}}$$

for all $l \in \{1, \ldots, N\}$, $j_l \in \{1, \ldots, n_{l,N}\}$. Let $\delta > 0$ be arbitrary. Using the inequality of McDiarmid (cf., e.g., Theorem A.2. in Györfi et al. (2002)) and assumption (12) we have

$$\sum_{N=1}^{\infty} \mathbf{P}\left(\left|\int_{\mathbb{R}} |f_N(x,t) - f(x,t)| \, dx - \mathbf{E}\left\{\int_{\mathbb{R}} |f_N(x,t) - f(x,t)| \, dx\right\}\right| \geq \delta\right)$$

$$= \sum_{N=1}^{\infty} \mathbf{P}\left(\left|g\left(X_1^{(t_1)}, \ldots, X_{n_{N,N}}^{(t_N)}\right) - \mathbf{E}\left\{g\left(X_1^{(t_1)}, \ldots, X_{n_{N,N}}^{(t_N)}\right)\right\}\right| \geq \delta\right)$$

$$\leq \sum_{N=1}^{\infty} 2 \cdot \exp\left(-\frac{\delta^2}{2} \frac{\left(\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N}\right)^2}{\sum_{l=1}^N H^2\left(\frac{t-t_l}{h_N}\right) \cdot n_{l,N} \cdot h_{l,N}^2}\right) < \infty.$$

An application of the lemma of Borel Cantelli implies the assertion. □

4.2 Preliminaries to the proof of Theorem 2

In this subsection we formulate and prove two general results, which we will use to prove Theorem 2.

Theorem 4 *Let $(X_t)_{t\in[0,1]}$ be an $\mathbb{R}$-valued stochastic process such that X_t has a density $f(\cdot,t) : \mathbb{R} \to \mathbb{R}$ with respect to the Lebesgue-Borel measure. Let $N \in \mathbb{N}$, $t_1, \ldots, t_N \in [0,1]$ and for each $k \in \{1, \ldots, N\}$ let $f_N^{(t_k)} = f_N^{(t_k)}(\cdot, \mathcal{D}_N^{(t_k)})$ be an estimate of $f(\cdot, t_k)$. Let $H : \mathbb{R} \to \mathbb{R}$ be a nonnegative bounded function with compact support and define the estimate $f_N(\cdot, t)$ of $f(\cdot, t)$ by (3).*
Assume that there exists a compact set $B \in \mathcal{B}$ such that

$$f(x,t) = 0 \quad \text{for all } x \notin B \text{ and all } t \in [0,1]. \tag{22}$$

Assume furthermore that the estimates $f_N^{(t_1)}, \ldots, f_N^{(t_N)}$ are densities satisfying for some $r \in (0,1]$ and some constants $c_8, c_9 > 0$

$$\int_{\mathbb{R}} \mathbf{E}\left|f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right|^2 dx \leq \frac{c_8}{n_{k,N} \cdot h_{k,n}} \quad (k \in \{1, \ldots, N\}), \tag{23}$$

$$\int_{\mathbb{R}} \left| \mathbf{E}\{f_N^{(t_k)}(x)\} - f(x,t_k) \right| dx \leq c_9 \cdot h_{k,N}^r \quad (k \in \{1,\dots,N\}) \tag{24}$$

and

$$f_N^{(t_j)}(x),\ f_N^{(t_k)}(x) \text{ independent } \quad (x \in \mathbb{R},\ j,k \in \{1,\dots,N\},\ j \neq k). \tag{25}$$

Then we have for any $t \in [0,1]$ *such that* $\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N} > 0$ *:*

$$\begin{aligned}
&\mathbf{E} \int |f_N(x,t) - f(x,t)|\, dx \\
&\leq \frac{c_{10}}{\sqrt{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}}} + 2 \cdot c_9 \cdot \max_{k:\, H(\frac{t-t_k}{h_N})>0} h_{k,N}^r \\
&\quad +2 \cdot \sup_{s \in [0,1]\,:\, H(\frac{t-s}{h_N})>0} \int_{\mathbb{R}} |f(x,t) - f(x,s)|\, dx.
\end{aligned}$$

Proof of Theorem 4. For $z \in \mathbb{R}$ set $z_+ = \max\{z, 0\}$. By the Lemma of Scheffé (cf., e.g., Theorem 1 in Chapter 1 of Devroye and Györfi (1985)) and assumption (22) we have

$$\int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\, dx = 2\int_{\mathbb{R}} (f(x,t) - f_N(x,t))_+\, dx = 2\int_B (f(x,t) - f_N(x,t))_+\, dx.$$

Using

$$(a+b)_+ \leq |a| + |b| \quad (a, b \in \mathbb{R})$$

we get

$$\int_{\mathbb{R}} |f_N(x,t) - f(x,t)|\, dx \leq T_{1,N} + T_{2,N}$$

where

$$T_{1,N} = 2\int_B \left| \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot (f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\})}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \right| dx$$

and

$$T_{2,N} = 2\int_B \left| f(x,t) - \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot \mathbf{E}\{f_N^{(t_k)}(x)\}}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \right| dx.$$

First, we bound $\mathbf{E}\{T_{1,N}\}$ with similar arguments as in the proof of Theorem 1. Let $|B|$ be the Lebesgue measure of the set B. By Cauchy-Schwarz inequality, Jensen inequality, the theorem of Fubini and assumption (25), which implies

$$\mathbf{E}\left\{ \left(f_N^{(t_j)}(x) - \mathbf{E}\{f_N^{(t_j)}(x)\}\right) \cdot \left(f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right) \right\} = 0$$

for all $x \in \mathbb{R}$, $j, k \in \{1, \dots, N\}$ with $j \neq k$, we get

$$\mathbf{E}\{T_{1,N}\}$$

$$\leq 2 \cdot \sqrt{|B|} \cdot \left(\mathbf{E}\left\{ \int_B \left(\frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot \left(f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right)}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \right)^2 dx \right\}^{1/2} \right)$$

$$\leq 2 \cdot \sqrt{|B|} \cdot \left(\frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right)^2 \cdot n_{k,N}^2 \cdot h_{k,N}^2 \cdot \int_B \mathbf{E}\left\{ \left| f_N^{(t_k)}(x) - \mathbf{E}\{f_N^{(t_k)}(x)\} \right|^2 \right\} dx}{\left(\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}\right)^2} \right)^{1/2}.$$

Since H is nonnegative and bounded we have $H^2(z) \leq c_{11} \cdot H(z)$ for all $z \in \mathbb{R}$. Using this and assumption (23) we get

$$\begin{aligned}
\mathbf{E}\{T_{1,n}\} &\leq 2 \cdot \sqrt{|B|} \cdot \sqrt{c_{11}} \cdot \left(\frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot c_8}{\left(\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}\right)^2} \right)^{1/2} \\
&= \frac{c_{10}}{\sqrt{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}}}.
\end{aligned}$$

Next, we bound $T_{2,N}$. Using triangle inequality, condition (24) and the fact that H is nonnegative we get

$$\begin{aligned}
T_{2,n} &\leq 2 \int_B \left| f(x,t) - \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot f(x,t_k)}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \right| dx \\
&\quad + 2 \int_B \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot \left| f(x,t_k) - \mathbf{E}\{f_N^{(t_k)}(x)\} \right|}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} dx \\
&\leq 2 \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot \int_B |f(x,t) - f(x,t_k)| \, dx}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \\
&\quad + 2 \frac{\sum_{k=1}^N H\left(\frac{t-t_k}{h_N}\right) \cdot n_{k,N} \cdot h_{k,N} \cdot c_9 \cdot h_{k,N}^r}{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \cdot n_{j,N} \cdot h_{j,N}} \\
&\leq 2 \cdot \sup_{s \in [0,1] \,:\, H(\frac{t-s}{h_N}) > 0} \int_{\mathbb{R}} |f(x,t) - f(x,s)| \, dx + 2 \cdot c_9 \cdot \max_{k:\, H(\frac{t-t_k}{h_N}) > 0} h_{k,N}^r.
\end{aligned}$$

The proof is complete. □

Theorem 5 *Let* $(X_t)_{t \in [0,1]}$ *be an* $\mathbb{R}$*-valued stochastic process such that* X_t *has a density* $f(\cdot,t) : \mathbb{R} \to \mathbb{R}$ *with respect to the Lebesgue-Borel measure. Let* $N \in \mathbb{N}$, $t_1, \dots, t_N \in [0,1]$, $h_N > 0$ *and let* K *be a symmetric density satisfying*

$$\int_{\mathbb{R}} K^2(u) \, du < \infty \quad \text{and} \quad \int_{\mathbb{R}} |u| \cdot K(u) \, du < \infty.$$

For each $k \in \{1,\ldots,N\}$ let $f_N^{(t_k)} = f_N^{(t_k)}(\cdot, \mathcal{D}_N^{(t_k)})$ be the estimate of $f(\cdot,t_k)$ defined by (2). Let $H:\mathbb{R}\to\mathbb{R}$ be a nonnegative bounded function with compact support and define the estimate $f_N(\cdot,t)$ of $f(\cdot,t)$ by (3).
Assume that there exists a compact set $B \in \mathcal{B}$ such that

$$f(x,t) = 0 \quad \text{for all } x \notin B \text{ and all } t \in [0,1], \tag{26}$$

and that $f(\cdot,t_k)$ is Hölder continuous with exponent $r \in (0,1]$ and with Hölder constant $C>0$ for all $k \in \{1,\ldots,N\}$, i.e.,

$$|f(x,t_k) - f(y,t_k)| \le C\cdot|x-y|^r \quad \text{for all } x,y\in\mathbb{R} \text{ and all } k \in \{1,\ldots,N\}. \tag{27}$$

Assume furthermore that for all $k \in \{1,\ldots,N\}$

$$X_{t_k}, X_1^{(t_k)}, \ldots, X_{n_{k,N}}^{(t_k)} \text{ are independent and identically distributed} \tag{28}$$

and that

$$X_1^{(t_1)}, \ldots, X_{n_{1,N}}^{(t_1)}, \ldots, X_1^{(t_N)}, \ldots, X_{n_{N,N}}^{(t_N)} \text{ are independent.} \tag{29}$$

Then we have for any $t\in[0,1]$ such that $\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N} > 0:$

$$\begin{aligned}
&\mathbf{E}\int |f_N(x,t) - f(x,t)|\,dx \\
&\le \frac{c_{10}}{\sqrt{\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)\cdot n_{j,N}\cdot h_{j,N}}} + 2\cdot c_9 \cdot \max_{k:\,H(\frac{t-t_k}{h_N})>0} h_{k,N}^r \\
&\quad +2\cdot \sup_{s\in[0,1]\,:\,H(\frac{t-t_k}{h_N})>0} \int_{\mathbb{R}} |f(x,t) - f(x,s)|\,dx.
\end{aligned}$$

Proof of Theorem 5. The result follows directly from Theorem 4 provided we can show that the conditions (23), (24) and (25) are satisfied. Clearly, (25) follows from (29). The proof of Theorem 1 (cf., (20)) implies (23). Finally, by using the Lemma of Scheffé, (26), (28) and (27) we observe

$$\begin{aligned}
&\int_{\mathbb{R}} \left|\mathbf{E}\{f_N^{(t_k)}(x)\} - f(x,t_k)\right| dx \\
&= 2\int_{\mathbb{R}} \left(f(x,t_k) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right)_+ dx \\
&= 2\int_{B} \left(f(x,t_k) - \mathbf{E}\{f_N^{(t_k)}(x)\}\right)_+ dx \\
&\le 2\cdot|B|\cdot\sup_{x\in B}\left|f(x,t_k) - \frac{1}{h_{k,N}}\int_{\mathbb{R}} K\left(\frac{x-u}{h_{k,N}}\right)\cdot f(u,t_k)\,du\right| \\
&\le 2\cdot|B|\cdot\sup_{x\in B}\frac{1}{h_{k,N}}\int_{\mathbb{R}} K\left(\frac{x-u}{h_{k,N}}\right)\cdot |f(x,t_k) - f(u,t_k)|\,du
\end{aligned}$$

$$\leq 2 \cdot |B| \cdot \sup_{x \in B} \frac{1}{h_{k,N}} \int_{\mathbb{R}} K\left(\frac{x-u}{h_{k,N}}\right) \cdot C \cdot |x-u|^r \, du$$
$$\leq 2 \cdot |B| \cdot C \cdot \int_{\mathbb{R}} |z|^r K(z) \, dz \cdot h_{k,N}^r.$$

(Here $|B|$ denotes again the Lebesgue measure of B). Since

$$\int_{\mathbb{R}} |z|^r K(z) \, dz \leq \int_{\mathbb{R}} (1+|z|) \cdot K(z) \, dz = 1 + \int_{\mathbb{R}} |z| \, K(z) \, dz < \infty$$

this implies (24), which completes the proof. □

4.3 Proof of Theorem 2

Because of $h_N \geq 1/N$ and the choice of t_k and H we have $\sum_{j=1}^N H((t-t_j)/h_N) > 0$ for all $t \in [0,1]$. Application of Theorem 5 yields

$$\mathbf{E} \int |f_N(x,t) - f(x,t)| \, dx \quad \leq \quad \frac{c_{10}}{\sqrt{n_N \cdot \tilde{h}_N \cdot \sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right)}} + c_9 \cdot h_N^r$$
$$+2 \cdot \sup_{s,t \in [0,1] \, : \, |s-t| \leq h_N} \int_{\mathbb{R}} |f(x,t) - f(x,s)| \, dx.$$

Again by the choice of t_k and H we get $\sum_{j=1}^N H\left(\frac{t-t_j}{h_N}\right) \geq \frac{1}{2} \cdot (2 \cdot N \cdot h_N - 1) = N \cdot h_N - \frac{1}{2}$. Furthermore, by the Lemma of Scheffé and assumptions (13) and (15) we get

$$\sup_{s,t \in [0,1] \, : \, |s-t| \leq h_N} \int_{\mathbb{R}} |f(x,t) - f(x,s)| \, dx$$
$$= \sup_{s,t \in [0,1] \, : \, |s-t| \leq h_N} 2 \cdot \int_B (f(x,t) - f(x,s))_+ \, dx \leq c_{11} \cdot h_N^p,$$

which proves (18). Plugging in the values of h_N and $\tilde{h}_N$ we get the second assertion. □

4.4 Proof of Theorem 3

Define

$$\Delta = \max_{k \in \{1,\ldots,N\}} \sup_{A \in \mathcal{A}_{t_k}} \left| \int_A f(x,t_k) \, dx - \mu_{m_{k,N}}(A) \right|$$

and let $\tilde{f}_{m_N, h_N^*} \in \mathcal{F}$ be any estimate which satisfies

$$\max_{k \in \{1,\ldots,N\}} \int \left| \tilde{f}_{m_N, h_N^*}(x,t_k) - f(x,t_k) \right| \, dx$$
$$\leq \inf_{h_N > 0} \max_{k \in \{1,\ldots,N\}} \int \left| \tilde{f}_{m_N, h_N}(x,t_k) - f(x,t_k) \right| \, dx + \frac{1}{4 \cdot N}.$$

W.l.o.g. we assume $h_N^* \neq \hat{h}_N$. In the *first step* of the proof we show

$$\begin{aligned}
&\max_{k\in\{1,\dots,N\}} \int \left| \hat{f}_N(x,t_k) - f(x,t_k) \right| \, dx \\
&\leq 3 \cdot \max_{k\in\{1,\dots,N\}} \int \left| \tilde{f}_{m_N,h_N^*}(x,t_k) - f(x,t_k) \right| \, dx + 4\,\Delta + \frac{2}{N}. \qquad (30)
\end{aligned}$$

It is very similar to the one of Theorem 6.3 in Devroye and Lugosi (2001) but we present it here for the sake of completeness. By triangle inequality it holds

$$\begin{aligned}
&\max_{k\in\{1,\dots,N\}} \int \left| \hat{f}_N(x,t_k) - f(x,t_k) \right| \, dx \qquad (31) \\
&\leq \max_{k\in\{1,\dots,N\}} \int \left| \hat{f}_N(x,t_k) - \tilde{f}_{m_N,h_N^*}(x,t_k) \right| \, dx + \max_{k\in\{1,\dots,N\}} \int \left| \tilde{f}_{m_N,h_N^*}(x,t_k) - f(x,t_k) \right| \, dx.
\end{aligned}$$

Define

$$G_{t_k} = \left\{ x \in \mathbb{R} : \, \hat{f}_N(x,t_k) > \tilde{f}_{m_N,h_N^*}(x,t_k) \right\}.$$

With Scheffé's identity, triangle inequality and the definition of $\hat{h}_N$ and h_N^* we get for the first term

$$\begin{aligned}
&\max_{k\in\{1,\dots,N\}} \int \left| \hat{f}_N(x,t_k) - \tilde{f}_{m_N,h_N^*}(x,t_k) \right| \, dx \\
&= 2 \max_{k\in\{1,\dots,N\}} \left(\int_{G_{t_k}} \hat{f}_N(x,t_k)\,dx - \int_{G_{t_k}} \tilde{f}_{m_N,h_N^*}(x,t_k)\,dx \right) \\
&\leq 2 \max_{k\in\{1,\dots,N\}} \sup_{A\in\mathcal{A}_{t_k}} \left| \int_A \hat{f}_N(x,t_k)\,dx - \int_A \tilde{f}_{m_N,h_N^*}(x,t_k)\,dx \right| \\
&\leq 2 \max_{k\in\{1,\dots,N\}} \sup_{A\in\mathcal{A}_{t_k}} \left| \int_A \hat{f}_N(x,t_k)\,dx - \mu_{m_{k,N}}(A) \right| \\
&\quad + 2 \max_{k\in\{1,\dots,N\}} \sup_{A\in\mathcal{A}_{t_k}} \left| \mu_{m_{k,N}}(A) - \int_A \tilde{f}_{m_N,h_N^*}(x,t_k)\,dx \right| \\
&\leq 2 \cdot \left(\inf_{h_N>0} \Delta_{h_N} + \frac{1}{N} \right) + 2\,\Delta_{h_N^*} \\
&\leq 4\,\Delta_{h_N^*} + \frac{2}{N} \\
&\leq 4 \max_{k\in\{1,\dots,N\}} \sup_{A\in\mathcal{A}_{t_k}} \left| \int_A \tilde{f}_{m_N,h_N^*}(x,t_k)\,dx - \int_A f(x,t_k)\,dx \right| \\
&\quad + 4 \max_{k\in\{1,\dots,N\}} \sup_{A\in\mathcal{A}_{t_k}} \left| \int_A f(x,t_k)\,dx - \mu_{m_{k,N}}(A)\,dx \right| + \frac{2}{N} \\
&\leq 4 \max_{k\in\{1,\dots,N\}} \sup_{B\in\mathcal{B}} \left| \int_B \tilde{f}_{m_N,h_N^*}(x,t_k)\,dx - \int_B f(x,t_k)\,dx \right| + 4\,\Delta + \frac{2}{N} \\
&= 2 \max_{k\in\{1,\dots,N\}} \int \left| \tilde{f}_{m_N,h_N^*}(x,t_k) - f(x,t_k) \right| \, dx + 4\,\Delta + \frac{2}{N}.
\end{aligned}$$

This together with (31) leads to (30).

In the *second step* of the proof we bound $\mathbf{E}(\Delta)$. Let $\epsilon > 0$ be arbitrary. Since Δ is nonnegative, we have

$$\mathbf{E}(\Delta) \;=\; \int_0^\infty \mathbf{P}\left(\Delta > u\right) du \;\leq\; \epsilon + \int_\epsilon^\infty \mathbf{P}\left(\Delta > u\right) du. \tag{32}$$

Let $u > 0$ be arbitrary. With a well-known result of Vapnik and Chervonenkis (1971) (cf., e.g., Theorem 12.5 in Devroye, Györfi and Lugosi (1996)) it holds

$$\begin{aligned} \mathbf{P}\left(\Delta > u\right) &\leq \sum_{k=1}^N \mathbf{P}\left(\sup_{A\in\mathcal{A}_{t_k}} \left|\int_A f(x,t_k)\,dx - \mu_{m_{k,N}}(A)\right| > u\right) \\ &\leq \sum_{k=1}^N 8\cdot s\left(\mathcal{A}_{t_k}, m_{k,N}\right)\cdot \exp\left(-\frac{m_{k,N}\cdot u^2}{32}\right), \end{aligned} \tag{33}$$

where

$$s\left(\mathcal{A}_{t_k}, m_{k,N}\right) = \max_{x_1,\ldots,x_{m_{k,N}}\in\mathbb{R}} \#\left\{A\cap\left\{x_1,\ldots,x_{m_{k,N}}\right\} : A\in\mathcal{A}_{t_k}\right\}$$

is the $m_{k,N}$–th shatter coefficient of $\mathcal{A}_{t_k}$, which we bound in the sequel. For this purpose, we have to count the number of subsets of $\left\{x_1,\ldots,x_{m_{k,N}}\right\}$ which can be picked out from sets of the form

$$\left\{x\in\mathbb{R} : \tilde{f}_{m_N,h_1}(x,t_k) > \tilde{f}_{m_N,h_2}(x,t_k)\right\},$$

h_1, $h_2 > 0$. Since H and K are naive kernels, this number is upper bounded by the number of subsets of $\left\{x_1,\ldots,x_{m_{k,N}}\right\}$ which can be picked out from sets of the form

$$\begin{aligned} \Bigg\{x\in\mathbb{R} : a_1\cdot\sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{h_1}\right)\cdot\sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x-X_i^{(t_k)}}{h_{k,N}}\right) \\ > a_2\cdot\sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{h_2}\right)\cdot\sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x-X_i^{(t_k)}}{h_{k,N}}\right)\Bigg\} \end{aligned} \tag{34}$$

for arbitrary $a_1, a_2 > 0$. We start by counting the number of vectors of the form

$$\begin{aligned} \Bigg(\sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{h_1}\right)\cdot\sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x_v-X_i^{(t_k)}}{h_{k,N}}\right), \\ \sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{h_2}\right)\cdot\sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x_v-X_i^{(t_k)}}{h_{k,N}}\right)\Bigg), \end{aligned}$$

$v\in\{1,\ldots,m_{k,N}\}$. The components of these vectors are natural numbers bounded by $N\cdot m_{k,N}$. Consequently, the number of different vectors does not exceed

$$L_{k,N} := \left(1+N\cdot m_{k,N}\right)^2.$$

In the following we proceed as in the proof of Lemma 11.1 in Devroye and Lugosi (2001). Let

$$\left(z_1^{(1)}, z_2^{(1)}\right), \dots, \left(z_1^{(L_{k,N})}, z_2^{(L_{k,N})}\right)$$

be all these vectors. If

$$\mathbb{1}_{\left\{a_1 \cdot \sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{h_1}\right) \cdot \sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x_v - X_i^{(t_k)}}{h_{k,N}}\right) \;>\; a_2 \cdot \sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{h_2}\right) \cdot \sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x_v - X_i^{(t_k)}}{h_{k,N}}\right)\right\}}$$
$$\neq \mathbb{1}_{\left\{a_3 \cdot \sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{\bar{h}_1}\right) \cdot \sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x_v - X_i^{(t_k)}}{h_{k,N}}\right) \;>\; a_4 \cdot \sum_{j=1}^N \mathbb{1}_{[-1,1]}\left(\frac{t-t_j}{\bar{h}_2}\right) \cdot \sum_{i=1}^{m_{k,N}} \mathbb{1}_{[-1,1]}\left(\frac{x_v - X_i^{(t_k)}}{h_{k,N}}\right)\right\}}$$

for some $v \in \{1, \dots, m_{k,N}\}$, $a_1, a_2, a_3, a_4, h_1, h_2, \bar{h}_1, \bar{h}_2 > 0$, then

$$\mathbb{1}_{\{a_1 \cdot z_1^{(w)} \;>\; a_2 \cdot z_2^{(w)}\}} \neq \mathbb{1}_{\{a_3 \cdot z_1^{(\bar{w})} \;>\; a_4 \cdot z_2^{(\bar{w})}\}}$$

for some $w, \bar{w} \in \{1, \dots, L_{k,N}\}$. Consequently, the number of subsets which can be picked out by the sets of the form (34) is upper bounded by the $L_{k,N}^2$–th shatter coefficient of the set

$$\mathcal{C} := \left\{ \left\{ (z_1, z_2, z_3, z_4) \in \mathbb{R}^4 : \sum_{i=1}^4 a_i \cdot z_i > 0 \right\} : a_1, a_2, a_3, a_4 > 0 \right\}.$$

Theorem 9.3 and Theorem 9.5 of Györfi et al. (2002) lead to

$$s\left(\mathcal{C}, L_{k,N}\right) \leq \left(1 + L_{k,N}^2\right)^4.$$

Using this and (33) we have for $\epsilon \geq 1/N$

$$\begin{aligned}
\mathbf{E}(\Delta) &\leq \epsilon + \int_\epsilon^\infty \sum_{k=1}^N 8 \cdot \left(1 + L_{k,N}^2\right)^4 \cdot \exp\left(-\frac{m_{k,N} \cdot u^2}{32}\right) du \\
&\leq \epsilon + 8 \cdot \sum_{k=1}^N \left(1 + L_{k,N}^2\right)^4 \int_\epsilon^\infty \exp\left(-\frac{m_{k,N} \cdot \epsilon \cdot u}{32}\right) du \\
&= \epsilon + 8 \cdot \sum_{k=1}^N \left(1 + L_{k,N}^2\right)^4 \cdot \frac{32}{m_{k,N} \cdot \epsilon} \cdot \exp\left(-\frac{m_{k,N} \cdot \epsilon^2}{32}\right) \\
&\leq \epsilon + 256 \cdot N \cdot \frac{2^{20} \cdot N^{16} \cdot \max_{k \in \{1,\dots,N\}} m_{k,N}^{16}}{\min_{k \in \{1,\dots,N\}} m_{k,N}} \cdot \exp\left(-\frac{\min_{k \in \{1,\dots,N\}} m_{k,N} \cdot \epsilon^2}{32}\right) \\
&= \epsilon + 256 \cdot \exp\left(\log\left(\frac{2^{20} \cdot N^{17} \cdot \max_{k \in \{1,\dots,N\}} m_{k,N}^{16}}{\min_{k \in \{1,\dots,N\}} m_{k,N}}\right) - \frac{\min_{k \in \{1,\dots,N\}} m_{k,N} \cdot \epsilon^2}{32}\right) \\
&\leq \epsilon + 256 \cdot \exp\left(-\frac{\min_{k \in \{1,\dots,N\}} m_{k,N} \cdot \epsilon^2}{64}\right),
\end{aligned}$$

provided we choose ϵ such that

$$\frac{\min_{k\in\{1,\dots,N\}} m_{k,N} \cdot \epsilon^2}{64} \geq \log\left(\frac{2^{20} \cdot N^{17} \cdot \max_{k\in\{1,\dots,N\}} m_{k,N}^{16}}{\min_{k\in\{1,\dots,N\}} m_{k,N}}\right).$$

With

$$\epsilon = \max\left\{\frac{1}{N}, \sqrt{\frac{64}{\min_{k\in\{1,\dots,N\}} m_{k,N}} \cdot \log\left(\frac{2^{20} \cdot N^{17} \cdot \max_{k\in\{1,\dots,N\}} m_{k,N}^{16}}{\min_{k\in\{1,\dots,N\}} m_{k,N}}\right)}\right\}$$

we get the assertion. □

5 Acknowledgment

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Center 805.

References

[1] Bosq, D. and Delecroix, M. (1985). Nonparametric prediction of a Hilbert-space valued random variable. *Stochastic Processes and their Applications*, **19**, pp. 271–280.

[2] Burba, F., Ferraty, F. and Vieu, P. (2009). K-Nearest Neighbour method in functional nonparametric regression. *Journal of Nonparametric Statistics*, **21**, pp. 453–469.

[3] Devroye, L. (1983). The equivalence in L1 of weak, strong and complete convergence of kernel density estimates. *The Annals of Statistics*, **11**, pp. 896–904.

[4] Devroye, L. (1987). A Course in Density Estimation. *Birkhäuser*, Basel.

[5] Devroye, L. and Györfi, L. (1985). Nonparametric Density Estimation. The L1 view. *Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley and Sons*, New York.

[6] Devroye, L. and Györfi, L. (1990). No empirical probability measure can converge in the total variation sense for all distributions. *The Annals of Statistics*, **18**, pp. 1496–1499.

[7] Devroye, L., Györfi, L. and Lugosi, G. (1996). A Probabilistic Theory of Pattern Recognition. *Springer*, New York.

[8] Devroye, L. and Lugosi, G. (2001). Combinatorial Methods in Density Estimation. *Springer-Verlag*, New York.

[9] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical Planning and Inference*, **140**, pp. 335–352.

[10] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. (2011). Kernel regression with functional response. *Electronic Journal of Statistics*, **5**, pp. 159–171.

[11] Ferraty, F. and Romain, Y. (2011). The Oxford Handbook of Functional Data Analysis. *Oxford University Press*, New York.

[12] Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Prctice. *Springer Series in Statistics*, Springer, New York.

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

[14] Lecoutre, J. P. (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statistics and Probability Letters*, **10**, pp. 145–149.

[15] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications*, **115**, pp. 155–177.

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

[17] Ramsay, J. O., and Silverman, B. W. (1997). Functional Data Analysis. *Springer*, New York.

[18] Ramsay, J. O., and Silverman, B. W. (2002). Applied Functional Data Analysis: Methods and case studies. *Springer*, New York.

[19] Ramsay, J. O., and Silverman, B. W. (2005). Functional Data Analysis. *2nd edition Springer*, New York.

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

[21] Vapnik, V. and Chervonenkis, A. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, **16**, pp. 264–280.