Optimal global rates of convergence for interpolation problems with random design *

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*

December 5, 2012

Abstract

Given a sample of a d-dimensional design variable X and observations of the corresponding values of a measurable function $m : \mathbb{R}^d \to \mathbb{R}$ without additional errors we are interested in estimating m on whole $\mathbb{R}^d$ such that the L_1 error (with integration with respect to the design measure) of the estimate is small. Under the assumption that the support of X is bounded and that m is (p, C)-smooth (i.e., roughly speaking, m is p-times continuously differentiable) we derive the minimax lower and upper bounds on the L_1 error.

AMS classification: Primary 62G08; secondary 62G05.

Key words and phrases: Interpolation, L_1–error, minimax rate of convergence, random design.

1. Introduction

Let X, X_1, X_2, ... be independent and identically distributed $\mathbb{R}^d$-valued random variables, let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and set $Y = m(X)$ and $Y_i = m(X_i)$ $(i = 1, 2, \dots)$. In the sequel we study the problem of estimating m from the data

$$\{(X_1, Y_1), \dots, (X_n, Y_n)\}. \tag{1}$$

Let

$$m_n(\cdot) = m_n(\cdot, \{(X_1, Y_1), \dots, (X_n, Y_n)\}) : \mathbb{R}^d \to \mathbb{R}$$

be an estimate of m based on the data (1). Motivated by a problem in density estimation, where m_n is used to generate additional data for the density estimate and where the error of the method crucially depends on the L_1 error of m_n (cf., Devroye, Felber and Kohler

*Running title: *Interpolation with random design*

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

(2012) and Kohler and Krzyżak (2012)), we measure in this paper the error of m_n by the L_1 error computed with respect to measure of X, i.e., by

$$\int |m_n(x) - m(x)| \, \mathbf{P}_X(dx).$$

Our main goal is to determine minimax lower bounds of the form

$$\inf_{m_n} \sup_{(X,Y) \in \mathcal{D}} \mathbf{E} \int |m_n(x) - m(x)| \, \mathbf{P}_X(dx),$$

where the infimum is taken with respect to all estimates and $\mathcal{D}$ is a suitable class of distributions of (X, Y), and to find estimates which achieve these minimax bounds up to some constant. In particular we are interested in deriving results which do not require any assumptions on the distribution of the design measure besides boundedness.

Surprisingly it turns out that the minimax rates for uniformly distributed X and for arbitrary bounded X are different. In the first case we show for univariate X and for sufficiently smooth m that the rate is better than n^{-1}, which is not possible in the latter case.

Our estimation problem can be considered as a regression estimation problem without noise in the dependent variable. The regression estimation with noise in the dependent variable has been extensively studied in the literature. The most popular estimates include kernel regression estimate (cf., e.g., Nadaraya (1964, 1970), Watson (1964), Devroye and Wagner (1980), Stone (1977, 1982), Devroye and Krzyżak (1989) or Kohler, Krzyżak and Walk (2009)), partitioning regression estimate (cf., e.g., Györfi (1981), Beirlant and Györfi (1998) or Kohler, Krzyżak and Walk (2006)), nearest neighbor regression estimate (cf., e.g., Devroye (1982) or Devroye, Györfi, Krzyżak and Lugosi (1994)), least squares estimates (cf., e.g., Lugosi and Zeger (1995) or Kohler (2000)) and smoothing spline estimates (cf., e.g., Wahba (1990) or Kohler and Krzyżak (2001)). Stone (1982) was first to derive minimax rates of convergence in this context. Our results show how minimax rates of convergence change when there is no noise in the dependent variable.

The main results are presented in Section 2 and proven in Section 3.

2. Main results

In the sequel we estimate functions which are (p, C)-smooth in the following sense:

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

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

for all $x, z \in \mathbb{R}^d$, where $\mathbb{N}_0$ is the set of non-negative integers.

One way to estimate such functions is to use a piecewise constant estimate, where the value at some point is estimated by the y value corresponding to the closest x-value in the data set (1). More precisely, for $x \in \mathbb{R}^d$ let $X_{(1,n)}(x), \ldots, X_{(n,n)}(x)$ be a permutation of $X_1, \ldots, X_n$ such that

$$\|x - X_{(1,n)}(x)\| \leq \cdots \leq \|x - X_{(n,n)}(x)\|.$$

In case of ties, i.e., in case $X_i = X_j$ for some $1 \leq i < j \leq n$, we use tie breaking by indices, i.e., we choose the data point with the smaller index first. Then we define the 1-nearest neighbor estimate by

$$m_n(x) = m_n(x, (X_1, m(X_1)), \ldots, (X_n, m(X_n))) = m(X_{(1,n)}). \tag{2}$$

For this estimate we can show the following error bound:

Theorem 1 *Assume supp*$(X) \subseteq [0,1]^d$, $0 < p \leq 1$, $C > 0$ *and let* $m : \mathbb{R}^d \to \mathbb{R}$ *be an arbitrary* (p, C)*-smooth function. Let* m_n *be defined by (2). Then*

$$\mathbf{E} \int |m_n(x) - m(x)| \, \mathbf{P}_X(dx) \leq \begin{cases} c_1 \cdot n^{-p/d} & \text{if } p < d, \\ c_2 \cdot \frac{\log n}{n} & \text{if } p = d = 1 \end{cases}$$

for some constants c_1, $c_2 \in \mathbb{R}_+$.

The rates for the 1-nearest neighbor estimate are always less than $1/n$. We show next that in case $d = 1$ we can define a nearest neighbor polynomial interpolation estimate which achieves under regularity assumption on the design distribution for smooth functions the rates that are even better than $1/n$. The estimate will depend on a parameter $k \in \mathbb{N}$. Given x and the data (1), we first choose the k nearest neighbors $X_{(1,n)}(x), \ldots, X_{(k,n)}(x)$ of x among $X_1, \ldots, X_n$, then we choose a polynomial $\hat{p}_x$ of degree $k-1$ interpolating $(X_{(1,n)}(x), m(X_{(1,n)}(x))), \ldots, (X_{(k,n)}(x), m(X_{(k,n)}(x)))$ (such a polynomial always exists and is unique when $X_{(1,n)}(x), \ldots, X_{(k,n)}(x)$ are pairwise disjoint), and define our k-nearest neighbor polynomial interpolating estimate by

$$m_{n,k}(x) = \hat{p}_x(x). \tag{3}$$

For this estimate the following error bound holds:

Theorem 2 *Let* $p \in \mathbb{N}$ *and* $C > 0$, $d = 1$ *and assume that* $m : \mathbb{R} \to \mathbb{R}$ *is* (p, C)*-smooth and that the distribution of* X *satisfies supp*$(\mathbf{P}_X) \subseteq [0,1]$,

$$\mathbf{P}\{X = x\} = 0 \quad \text{for all } x \in [0,1] \tag{4}$$

and

$$\mathbf{P}_X(S_r(x)) \geq c_3 \cdot r \tag{5}$$

for all $x \in [0,1]$ *and all* $r \in (0,1]$ *for some constant* $c_3 > 0$, *where* $S_r(x)$ *is the closed ball with radius* r *centered at* x. *Then for the* p*-nearest neighbor polynomial interpolating estimate* $m_{n,p}$ *defined by (3) the following bound holds:*

$$\mathbf{E} \int |m_{n,p}(x) - m(x)| \, \mathbf{P}_X(dx) \leq c_4 \cdot n^{-p}$$

for some $c_4 \in \mathbb{R}_+$, *where* $\mathbb{R}_+$ *is the set of positive real numbers.*

In the remaining part of the paper we investigate whether the rates of convergence presented in Theorems 1 and 2 can be improved. To do this, we present lower bounds on the expected L_1 error of any estimate.

Our first lower bound concerns estimation in case of uniformly distributed design variable. Let $\mathcal{D}_0^{(p,C)}$ be the class of all distributions of (X, Y) where X is uniformly distributed on $[0,1]^d$ and $Y = m(X)$ for some (p, C)-smooth function $m : \mathbb{R}^d \to [-1, 1]$. Then the following lower bound on the maximal expected L_1 error within the class $\mathcal{D}_0^{(p,C)}$ holds.

Theorem 3 *Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, and let $C > 0$. Then there exists a constant $c_5 > 0$ such that we have for any $n \geq 2$*

$$\inf_{m_n} \sup_{(X,Y) \in \mathcal{D}_0^{(p,C)}} \mathbf{E} \int |m_n(x) - m(x)| \, \mathbf{P}_X(dx) \geq c_5 \cdot n^{-p/d}.$$

Theorem 3 shows that the rate of convergence results presented in Theorem 1 in case $p \leq 1$ and $d > p$ and in Theorem 2 in case $d = 1$ and $p \in \mathbb{N}$ cannot be improved by more than a constant factor. Nevertheless, if we compare the results in these theorems we see that we get rates of convergence better than $1/n$ only under the strong assumptions on the distribution of the design presented in Theorem 2. So it appears that for nonparametric regression without error in the dependent variable the rates of convergence for uniformly distributed design are different than the rates of convergence for arbitrary bounded design, which is not the case for nonparametric regression with error in the dependent variable, cf., e.g., Kohler (2000) or Kohler, Krzyżak and Walk (2006, 2009). Our next result demonstrates that this is indeed the case.

Theorem 4 *Let $d \in \mathbb{N}$ be arbitrary. For any $n \geq 2$ and any estimate m_n there exists $m : \mathbb{R}^d \to [-1, 1]$ which is infinitely differentiable and vanishes outside of $[-1, 1]^d$ and a distribution of X which is concentrated on $(0, \ldots, 0)$ and $(1, \ldots, 1)$ such that*

$$\mathbf{E} \int |m_n(x) - m(x)| \, \mathbf{P}_X(dx) \geq \frac{1}{2 \cdot n}.$$

3. Proofs

3.1. Proof of Theorem 1

Since m is (p, C)-smooth and $p \leq 1$ we have

$$|m_n(X) - m(X)| = |m(X_{(1,n)}(X)) - m(X)| \leq C \cdot \|X_{(1,n)}(X) - X\|^p,$$

hence

$$\mathbf{E} \int |m_n(x) - m(x)| \, \mathbf{P}_X(dx) \quad \leq \quad C \cdot \mathbf{E} \left\{ \|X_{(1,n)}(X) - X\|^p \right\}.$$

An easy modification of the proof of Lemma 6.4 in Györfi et al. (2002) shows

$$\mathbf{E}\left\{\|X_{(1,n)}(X)-X\|^p\right\} \leq \begin{cases} c_6 \cdot n^{-p/d} & \text{if } p<d, \\ c_7 \cdot \frac{\log n}{n} & \text{if } p=d=1, \end{cases}.$$

The proof is complete. □

3.2. Proof of Theorem 2

In the proof we will need the following auxiliary result, which bounds the error of polynomial approximation in case of (p,C)-smooth functions.

Lemma 1 *Let $p \in \mathbb{N}_0$, let $C>0$ and let $m:\mathbb{R}\to\mathbb{R}$ be (p,C)-smooth. Let $x_1, \ldots, x_p$ be p distinct points in $\mathbb{R}$, and let q be the (uniquely determined) polynomial of degree $p-1$ which interpolates $(x_1, m(x_1)), \ldots, (x_p, m(x_p))$. Then for any $x \in \mathbb{R}$ we have*

$$|q(x)-m(x)| \leq \frac{C}{p!} \cdot \prod_{j=1}^{p} |x-x_j|.$$

Proof. The proof is a modification of the standard error formula for polynomial interpolation, cf., e.g., chapter 4 in Cheney and Kincaid (2008).

In case $x \in \{x_1, \ldots, x_p\}$ the assertion trivially holds, so we can assume w.l.o.g. $x \notin \{x_1, \ldots, x_p\}$. Set

$$w(t) = \prod_{j=1}^{p}(t-x_j) \quad \text{and} \quad c = \frac{m(x)-q(x)}{w(x)},$$

and define $\varphi:\mathbb{R}\to\mathbb{R}$ by

$$\varphi(t) = m(t) - q(t) - c\cdot w(t).$$

Since $m(x_i) = q(x_i)$ and $w(x_i)=0$ for $i=1,\ldots,p$ we know

$$\varphi(x_i) = 0 \quad (i=1,\ldots,p),$$

furthermore $\varphi(x)=0$. So φ has $p+1$ roots, and repeated application of the Theorem of Rolle implies that there exists

$$\xi_1, \xi_2 \in \left[\min\{x, x_1, \ldots, x_p\}, \max\{x, x_1, \ldots, x_p\}\right], \quad \xi_1 \neq \xi_2$$

such that

$$\varphi^{(p-1)}(\xi_1) = 0 = \varphi^{(p-1)}(\xi_2).$$

Since q is a polynomial of degree p its $(p-1)$-th derivative is constant, hence

$$\begin{aligned} 0 &= \varphi^{(p-1)}(\xi_1) - \varphi^{(p-1)}(\xi_2) \\ &= m^{(p-1)}(\xi_1) - m^{(p-1)}(\xi_2) - c\cdot\left(w^{(p-1)}(\xi_1) - w^{(p-1)}(\xi_2)\right) \\ &= m^{(p-1)}(\xi_1) - m^{(p-1)}(\xi_2) - c\cdot p!\cdot(\xi_1-\xi_2), \end{aligned}$$

and we get

$$\begin{aligned} m(x) - q(x) &= w(x) \cdot \frac{m^{(p-1)}(\xi_1) - m^{(p-1)}(\xi_2)}{p! \cdot (\xi_1 - \xi_2)} \\ &= \prod_{j=1}^{p} (x - x_j) \cdot \frac{m^{(p-1)}(\xi_1) - m^{(p-1)}(\xi_2)}{p! \cdot (\xi_1 - \xi_2)}. \end{aligned}$$

Using

$$|m^{(p-1)}(\xi_1) - m^{(p-1)}(\xi_2)| \leq C \cdot |\xi_1 - \xi_2|$$

this implies the assertion. □

Proof of Theorem 2. Because of (4) ties occur only with probability zero, and application of Lemma 1 yields

$$\begin{aligned} \mathbf{E} \int |m_{n,p}(x) - m(x)| \, \mathbf{P}_X(dx) &\leq c_8 \cdot \mathbf{E} \left\{ \prod_{j=1}^{p} |X - X_{(j:n)}(X)| \right\} \\ &\leq c_8 \cdot \mathbf{E} \left\{ |X - X_{(p,n)}(X)|^p \right\} \end{aligned}$$

Subdividing $X_1, \ldots X_n$ into p sets of size $\lfloor n/p \rfloor$ and computing the p first nearest neighbors $z_1, \ldots, z_p$ of X in these sets, we see that by the minimizing property in the definition of the p-th nearest neighbor we have

$$|X - X_{(p,n)}(X)|^p \leq \max_{j=1,\ldots,p} |X - z_j|^p \leq \sum_{j=1}^{p} |X - z_j|^p$$

which implies

$$\mathbf{E} \left\{ |X - X_{(p,n)}(X)|^p \right\} \leq p \cdot \mathbf{E} \left\{ |X - X_{(1, \lfloor n/p \rfloor)}(X)|^p \right\}.$$

Arguing as in the proof of Lemma 6.4 in Györfi et al. (2002) we get

$$\mathbf{E} \left\{ |X - X_{(1, \lfloor n/p \rfloor)}(X)|^p \right\} \leq \int_0^\infty \exp \left(-\lfloor n/p \rfloor \cdot \mathbf{P}_X \left(S_{\varepsilon^{1/p}}(x) \right) \right) d\varepsilon,$$

and by (5) the term on the right-hand side above is in turn bounded by

$$\int_0^\infty \exp \left(-\lfloor n/p \rfloor \cdot c_3 \cdot \varepsilon^{1/p} \right) d\varepsilon = \lfloor n/p \rfloor^{-p} \cdot \int_0^\infty \exp \left(-c_3 \cdot z^{1/p} \right) dz \leq c_9 \cdot n^{-p}.$$

The proof is complete. □

3.3. Proof of Theorem 3

The proof is a modification of the proof of Theorem 3.2 in Györfi et al. (2002), which in turn is based on a lower bound presented in Stone (1982).

Set $M_n = \lceil n^{1/d} \rceil$ and let $\{A_{n,j}\}_{j=1,\dots,M_n^d}$ be a partition of $[0,1]^d$ into cubes of side length $1/M_n$. Choose a $(p, 2^{\beta-1}C)$-smooth function $g : \mathbb{R}^d \to [-1,1]$ satisfying

$$supp(g) \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right] \quad \text{and} \quad \int |g(x)|\, dx > 0.$$

For $j \in \{1, \dots, M_n^d\}$ let $a_{n,j}$ be the center of $A_{n,j}$ and set

$$g_{n,j}(x) = M_n^{-p} \cdot g\left(M_n \cdot (x - a_{n,j})\right).$$

We index the class of functions considered by $c_n = (c_{n,1}, \dots, c_{n,M_n^d}) \in \{-1,1\}^{M_n^d}$ and define $m^{(c_n)} : \mathbb{R}^d \to [-1,1]$ by

$$m^{(c_n)}(x) = \sum_{j=1}^{M_n^d} c_{n,j} \cdot g_{n,j}(x).$$

Let m_n be an arbitrary estimate of m. As in the proof of Theorem 3.2 in Györfi et al. (2002) we can see that $m^{(c_n)}$ is (p,C)-smooth, which implies

$$\begin{aligned}
&\sup_{(X,Y) \in \mathcal{D}_0^{(p,C)}} \mathbf{E} \int |m_n(x) - m(x)|\, \mathbf{P}_X(dx) \\
&\geq \sup_{X \sim U([0,1]^d), Y = m^{(c_n)}(X) \text{ for some } c_n \in \{-1,1\}^{M_n^d}} \mathbf{E} \int_{[0,1]^d} |m_n(x) - m^{(c_n)}(x)| dx.
\end{aligned}$$

In order to bound the right-hand side of the inequality above we randomize c_n. Let $X_1, \dots, X_n$ be independent random variables uniformly distributed on $[0,1]^d$. Choose independent random variables $C_1, \dots, C_{M_n^d}$ independent from X_1,...,X_n satisfying

$$\mathbf{P}\{C_k = -1\} = \mathbf{P}\{C_k = 1\} = \frac{1}{2} \quad (k = 1, \dots, M_n^d),$$

which are also independent from $X_1, \dots, X_n$, and set

$$C_n = (C_1, \dots, C_{M_n^d})$$

and

$$Y_i = m^{(C_n)}(X_i) \quad (i = 1, \dots, n).$$

Then

$$\begin{aligned}
&\sup_{X \sim U([0,1]^d), Y = m^{(c_n)}(X) \text{ for some } c_n \in \{-1,1\}^{M_n^d}} \mathbf{E} \int_{[0,1]^d} |m_n(x) - m^{(c_n)}(x)| dx \\
&\geq \mathbf{E} \int_{[0,1]^d} |m_n(x) - m^{(C_n)}(x)| dx \\
&= \sum_{j=1}^{M_n^d} \mathbf{E} \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx
\end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{j=1}^{M_n^d} \mathbf{E}\left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \cdot I_{\{X_1,\ldots,X_n \notin A_{n,j}\}} \right\} \\
&= \sum_{j=1}^{M_n^d} \mathbf{E}\left\{ \mathbf{E}\left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \Big| \mathcal{F}_{n,j} \right\} \cdot I_{\{X_1,\ldots,X_n \notin A_{n,j}\}} \right\},
\end{aligned}
$$

where $\mathcal{F}_{n,j}$ is the σ-algebra generated by $X_1, \ldots, X_n, C_1, \ldots, C_{j-1}, C_{j+1}, \ldots, C_{M_n^d}$. If $X_1, \ldots, X_n$ are not contained in $A_{n,j}$, then $m^{(C_n)}(X_1), \ldots m^{(C_n)}(X_n)$ and hence also $m_n(x)$ are independent of C_j, which implies

$$
\begin{aligned}
&\mathbf{E}\left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \Big| \mathcal{F}_{n,j} \right\} \\
&= \frac{1}{2} \cdot \int_{A_{n,j}} |m_n(x) - g_{n,j}(x)| dx + \frac{1}{2} \cdot \int_{A_{n,j}} |m_n(x) + g_{n,j}(x)| dx \\
&= \frac{1}{2} \cdot \int_{A_{n,j}} \left(|g_{n,j}(x) - m_n(x)| + |g_{n,j}(x) + m_n(x)| \right) \, dx \\
&\geq \frac{1}{2} \cdot \int_{A_{n,j}} |(g_{n,j}(x) - m_n(x)) + (g_{n,j}(x) + m_n(x))| \, dx \\
&= \int_{A_{n,j}} |g_{n,j}(x)| \, dx,
\end{aligned}
$$

where the latter inequality follows from the triangle inequality. From this we conclude

$$
\begin{aligned}
&\sum_{j=1}^{M_n^d} \mathbf{E}\left\{ \mathbf{E}\left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \Big| \mathcal{F}_{n,j} \right\} \cdot I_{\{X_1,\ldots,X_n \notin A_{n,j}\}} \right\} \\
&\geq \sum_{j=1}^{M_n^d} \int_{A_{n,j}} |g_{n,j}(x)| \, dx \cdot \mathbf{P}\{X_1, \ldots, X_n \notin A_{n,j}\} \\
&= \sum_{j=1}^{M_n^d} M_n^{-p-d} \cdot \int |g(x)| \, dx \cdot \left(1 - \left(\frac{1}{M_n} \right)^d \right)^n \\
&= \int |g(x)| \, dx \cdot \left(\lceil n^{1/d} \rceil \right)^{-p} \cdot \left(1 - \left(\frac{1}{\lceil n^{1/d} \rceil} \right)^d \right)^n \\
&\geq \int |g(x)| \, dx \cdot \left(n^{1/d} + 1 \right)^{-p} \cdot \left(1 - \frac{1}{n} \right)^n \\
&\geq \int |g(x)| \, dx \cdot \frac{1}{2^p} \cdot n^{-p/d} \cdot \frac{1}{2}.
\end{aligned}
$$

The proof is complete. $\square$

3.4. Proof of Theorem 4

Define $f : \mathbb{R}^d \to [0,1]$ by

$$f(x) = \begin{cases} \exp\left(-\frac{\|x\|}{1-\|x\|}\right) & \text{if } \|x\| < 1, \\ 0 & \text{else,} \end{cases}$$

and for $c \in \{-1, 1\}$ set

$$m^{(c)}(x) = c \cdot f(x).$$

Clearly, $m^{(c)}$ is bounded in absolute value by one, infinitely many times differentiable and vanishes outside $[-1,1]^d$ for $c \in \{-1,1\}$. Furthermore, let $\bar{0} = (0, ..., 0)$, $\bar{1} = (1, ..., 1)$ and choose X such that

$$\mathbf{P}\{X = \bar{0}\} = \frac{1}{n} \quad \text{and} \quad \mathbf{P}\{X = \bar{1}\} = 1 - \frac{1}{n}$$

and let X, X_1, $\ldots$, X_n be independent and identically distributed. It suffices to show

$$\max_{c \in \{-1,1\}} \mathbf{E} \int |m_n(x, (X_1, m^{(c)}(X_1)), \ldots, (X_n, m^{(c)}(X_n))) - m^{(c)}(x)| \, \mathbf{P}_X(dx) \geq \frac{1}{2 \cdot n}.$$

But this follows from

$$\begin{aligned}
&\max_{c \in \{-1,1\}} \mathbf{E} \int |m_n(x, (X_1, m^{(c)}(X_1)), \ldots, (X_n, m^{(c)}(X_n))) - m^{(c)}(x)| \, \mathbf{P}_X(dx) \\
&\geq \max_{c \in \{-1,1\}} \mathbf{E} \left\{ |m_n(\bar{0}, (X_1, m^{(c)}(X_1)), \ldots, (X_n, m^{(c)}(X_n))) - m^{(c)}(\bar{0})| \cdot \frac{1}{n} \right\} \\
&\geq \frac{1}{2} \cdot \mathbf{E} \left\{ |m_n(\bar{0}, (X_1, m^{(1)}(X_1)), \ldots, (X_n, m^{(1)}(X_n))) - 1| \cdot \frac{1}{n} \right\} \\
&\qquad\qquad + \frac{1}{2} \cdot \mathbf{E} \left\{ |m_n(\bar{0}, (X_1, m^{(-1)}(X_1)), \ldots, (X_n, m^{(-1)}(X_n))) + 1| \cdot \frac{1}{n} \right\} \\
&= \frac{1}{2 \cdot n} \cdot \mathbf{E} \Bigg\{ |1 - m_n(\bar{0}, (X_1, m^{(1)}(X_1)), \ldots, (X_n, m^{(1)}(X_n)))| \\
&\qquad\qquad + |1 + m_n(\bar{0}, (X_1, m^{(-1)}(X_1)), \ldots, (X_n, m^{(-1)}(X_n)))| \Bigg\} \\
&\geq \frac{1}{2 \cdot n} \cdot \mathbf{E} \Bigg\{ \Big(|1 - m_n(\bar{0}, (X_1, m^{(1)}(X_1)), \ldots, (X_n, m^{(1)}(X_n)))| \\
&\qquad\qquad + |1 + m_n(\bar{0}, (X_1, m^{(-1)}(X_1)), \ldots, (X_n, m^{(-1)}(X_n)))| \Big) \cdot I_{\{X_1, \ldots, X_n \neq \bar{0}\}} \Bigg\} \\
&= \frac{1}{2 \cdot n} \cdot \mathbf{E} \Bigg\{ \Big(|1 - m_n(\bar{0}, (X_1, 0), \ldots, (X_n, 0))|
\end{aligned}$$

$$+|1 + m_n(\bar{0}, (X_1, 0), \ldots, (X_n, 0))|\Big) \cdot I_{\{X_1,\ldots,X_n \neq \bar{0}\}}\Bigg\}$$

$$\geq \frac{1}{2 \cdot n} \cdot \mathbf{E}\Bigg\{\Big|(1 - m_n(\bar{0}, (X_1, 0), \ldots, (X_n, 0))$$

$$+(1 + m_n(\bar{0}, (X_1, 0), \ldots, (X_n, 0))\Big| \cdot I_{\{X_1,\ldots,X_n \neq \bar{0}\}}\Bigg\}$$

$$= \frac{1}{n} \cdot \mathbf{P}\{X_1, \ldots, X_n \neq \bar{0}\} = \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right)^n \geq \frac{1}{2 \cdot n}.$$

□

References

[1] Beirlant, J. and Györfi, L. (1998). On the asymptotic L_2-error in partitioning regression estimation. *Journal of Statistical Planning and Inference*, **71**, pp. 93–107.

[2] Cheney, W. and Kincaid, D. (2008). *Numerical Mathematcs and Computing* Thomson Brooks/Cole, Belmont, CA.

[3] Devroye, L. (1982). Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **61**, pp. 467–481.

[4] Devroye, L., Felber, T., and Kohler, M. (2012). Estimation of a density using real and artificial data. To appear in *IEEE Transactions on Information Theory.*

[5] Devroye, L., Györfi, L., Krzyżak, A. and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics*, **22**, pp. 1371–1385.

[6] Devroye, L. and Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference*, **23**, pp. 71–82.

[7] Devroye, L. and Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics*, **8**, pp. 231–239.

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

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

[10] Kohler, M. (2000). Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression. *Journal of Statistical Planning and Inference*, **89**, pp. 1–23.

[11] Kohler, M. and Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Transactions on Information Theory*, **47**, pp. 3054–3058.

[12] Kohler, M. and Krzyżak, A. (2012). Adaptive density estimation based on real and artificial data. Submitted for publication.

[13] Kohler, M., Krzyżak, A. and Walk, H. (2006). Rates of convergence of for partitioning and nearest neighbor regression estimates with unbounded data. *Journal of Multivariate Analysis*, **97**, pp. 311-323.

[14] Kohler, M., Krzyżak, A. and Walk, H. (2009). Optimal global rates of convergence in nonparametric regression with unbounded data. *Journal of Statistical Planning and Inference*, **123**, pp. 1286-1296.

[15] Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Transactions on Information Theory*, **41**, pp. 677–687.

[16] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, pp. 141–142.

[17] Nadaraya, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves. *Theory of Probability and its Applications*, **15**, pp. 134–137.

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

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

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

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

Supplementary Material

A. A bound on the moment of the nearest neighbor distance

Lemma 2 *Under the assumptions of Theorem 1 we have*

$$\mathbf{E}\left\{\|X_{(1,n)}(X)-X\|^p\right\} \leq \begin{cases} c_6 \cdot n^{-p/d} & \text{if } p<d, \\ c_7 \cdot \frac{\log n}{n} & \text{if } p=d=1. \end{cases}$$

Proof.: Since $supp(X) \subseteq [0,1]^d$ implies $\|X_{(1,n)}(X)-X\| \leq \sqrt{d}$ *a.s.*, we have

$$\mathbf{E}\left\{\|X_{(1,n)}(X)-X\|^p\right\} = \int_0^{\sqrt{d}} \mathbf{P}\left\{\|X_{(1,n)}(X)-X\|^p > \varepsilon\right\} d\varepsilon.$$

Let $\varepsilon > 0$ be arbitrary. Then

$$\begin{aligned} \mathbf{P}\left\{\|X_{(1,n)}(X)-X\|^p > \varepsilon\right\} &= \int \mathbf{P}\left\{\|X_{(1,n)}(x)-x\|^p > \varepsilon\right\} \mathbf{P}_X(dx) \\ &= \int \mathbf{P}\left\{\|X_{(1,n)}(x)-x\| > \varepsilon^{1/p}\right\} \mathbf{P}_X(dx) \\ &= \int \mathbf{P}\left\{X_1,\dots,X_n \notin S_{\varepsilon^{1/p}}(x)\right\} \mathbf{P}_X(dx) \\ &= \int \left(1-\mathbf{P}_X\left(S_{\varepsilon^{1/p}}(x)\right)\right)^n \mathbf{P}_X(dx) \\ &\leq \int \exp\left(-n \cdot \mathbf{P}_X\left(S_{\varepsilon^{1/p}}(x)\right)\right) \mathbf{P}_X(dx) \\ &\leq \max_{x \in \mathbb{R}_+} x \cdot \exp(-x) \cdot \int \frac{1}{n \cdot \mathbf{P}_X\left(S_{\varepsilon^{1/p}}(x)\right)} \mathbf{P}_X(dx) \\ &\leq \frac{c_{10}}{n \cdot \varepsilon^{d/p}}, \end{aligned}$$

where the last inequality follows from Inequality (5.1) in Györfi et al. (2002). Using this we get

$$\begin{aligned} \mathbf{E}\left\{\|X_{(1,n)}(X)-X\|^p\right\} &\leq \int_0^{\sqrt{d}} \min\left\{1, \frac{c_{10}}{n \cdot \varepsilon^{d/p}}\right\} d\varepsilon \\ &\leq n^{-p/d} + \frac{c_{10}}{n} \cdot \int_{n^{-p/d}}^{\sqrt{d}} \varepsilon^{-d/p}\, d\varepsilon. \end{aligned}$$

Since

$$\frac{1}{n} \int_{n^{-p/d}}^{\sqrt{d}} \varepsilon^{-d/p}\, d\varepsilon \leq \begin{cases} c_{11} \cdot n^{-p/d} & \text{if } p<d, \\ c_{12} \cdot \frac{\log n}{n} & \text{if } p=d=1, \end{cases}$$

this implies the assertion. □