L_1-consistent estimation of the density of residuals in random design regression models *

Luc Devroye,[1,], Tina Felber[2†], Michael Kohler[2] and Adam Krzyżak[3]

[1] *School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6, email: lucdevroye@gmail.com*

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

[3] *Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8, email: krzyzak@cs.concordia.ca*

May 30, 2011

Abstract

In this paper we study the problem of estimating the density of the error distribution in a random design regression model, where the error is assumed to be independent of the design variable. Our main result is that the L_1 error of the kernel density estimate applied to residuals of a consistent regression estimate converges with probability one to zero regardedless of the form of the true density. We demonstrate that this result is in general no longer true if the error distribution and the design variable are dependent.

AMS classification: Primary 62G07; secondary 62G20.

Key words and phrases: Density estimation, L_1 error, residuals, nonparametric regression, universal consistency.

1 Introduction

In this paper we study the problem of estimating the density of the error distribution in a random design regression model. More precisely, we assume that we have given data

$$\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\},$$

where

$$Y_i = m(X_i) + \epsilon_i$$

for some $m : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{R}^d$-valued random (design) variables $X_1, \dots, X_n$ and real-valued random variables $\epsilon_1, \dots, \epsilon_n$ with zero expectation. We assume that $(X_1, \epsilon_1), (X_2, \epsilon_2), \dots$

*Running title: *Estimation of the density of residuals*

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

are independent and identically distributed and that X_1, ϵ_1 are independent. We let X be a random variable distributed as X_1 and ϵ be a random variable with density f with respect to Lebesgue measure. It is assumed that all ϵ_i's are distributed as ϵ. We use $\mathbf{P}_X$ for the probability law of a random variable X. We consider the problem of estimating f from the data $\mathcal{D}_n$.

Estimating the density of the error distribution in nonparametric regression models has been dealt with by several researchers. Ahmad (1992) showed that under a Lipschitz-condition of the kernel function, the Parzen-Rosenblatt density estimator (Parzen (1962), Rosenblatt (1956)) converges in probability at every continuity point to the real density of the residuals. In case of a continuous error density, the same estimator is pointwise and uniformly consistent (Cheng (2004)), and, in addition, the histogram error density estimator is uniformly and in L_1 consistent (Cheng (2002)). Efromovich (2005) investigated in a homeoscedastic regression model estimates which are as good as estimates using an oracle that knows the underlying regression errors. In the heteroscedastic nonparametric regression model, where the Y_i's have different variances, Efromovich (2006) generalized his optimal estimation for a twice differentiable error density with finite support. Estimators of the residual distribution function include that of Akritas and Van Keilegom (2001), who extended the results of Durbin (1973) and Loynes (1980) to a weak convergence result for a distribution function estimator in a nonparametric heteroscedastic regression model. The empirical distribution function of residuals was used as an estimator in an heteroscedastic model with multivariate covariates by Neumeyer and Van Keilgom (2010). For general results in density estimation we refer to the books of Devroye and Györfi (1985), Devroye (1987) and Devroye and Lugosi (2000).

In this paper the main aim is to derive L_1-consistent estimates of f. This is important, because Scheffé's Lemma implies that the L_1 error of a density estimate equals twice the total variation distance (see, e.g., Devroye and Györfi (1985)) and hence an L_1-consistent density estimate allows simultaneous estimation of all probabilities. In order to estimate the density of the error distribution, we split the data in two parts, use the first part to compute a regression estimate, compute its residuals on the second part and use them as data for a standard kernel density estimate. We show that the resulting density estimate is strongly consistent in L_1 for all densities f provided we use in the first step a consistent regression estimate. Furthermore, we show that this result does no longer hold in case that the design variable and the errors in the regression model are dependent. The estimates are defined in Section 2, the main results are presented in Section 3, and Section 4 contains the proofs.

2 Definition of the estimates

We start by splitting the sample in two parts: a first part of size n' and the second part containing the $n'' = n - n'$ remaining data points. We assume that we have given a regression estimate

$$m_{n'}(\cdot) = m_{n'}(\cdot, \mathcal{D}_{n'}) : \mathbb{R}^d \to \mathbb{R}$$

satisfying

$$\int |m_{n'}(x) - m(x)|\mathbf{P}_X(dx) \to 0 \quad a.s.$$

(for $n \to \infty$), which we apply to the first part of the data. We use $m_{n'}$ to compute the residuals on the second part of the data, i.e., we set

$$\hat{\epsilon}_j = Y_{n'+j} - m_{n'}\left(X_{n'+j}\right) \qquad (j = 1, \dots, n'')$$

and we use them to compute the Parzen-Rosenblatt kernel density estimate

$$f_n(x) = \frac{1}{n'' h_{n''}} \sum_{j=1}^{n''} K\left(\frac{x - \hat{\epsilon}_j}{h_{n''}}\right).$$

Here $h_{n''} > 0$ is the bandwidth, and $K : \mathbb{R} \to \mathbb{R}_+$ is the kernel function, which we assume to be a density [in general, only integrability to one is needed, but we will be happy with the subclass of densities].

3 Main results

Our first result deals with the consistency of f_n.

Theorem 1. *Let $X_1, X_2, \dots$ be i.i.d. $\mathbb{R}^d$-valued random variables distributed as X, and let $\epsilon_1, \epsilon_2, \dots$ be i.i.d. real-valued random variables with $\mathbf{E}\{\epsilon_1\} = 0$, and having common density f, and assume that both sequences are also independent of each other. Define*

$$Y_i = m(X_i) + \epsilon_i \qquad (i \in \mathbb{N})$$

for some function $m : \mathbb{R}^d \to \mathbb{R}$, and let the estimate f_n be defined as in Section 2. Assume that K is a density with compact support satisfying

$$\int_{\mathbb{R}} K^2(u)\, du < \infty, \tag{1}$$

such that

$$h_n \to 0 \quad (n \to \infty) \qquad \text{and} \qquad n \cdot h_n \to \infty \quad (n \to \infty) \tag{2}$$

and that the regression estimate m_n satisfies

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)|\, \mathbf{P}_{X_1}(dx) \to 0 \quad a.s. \tag{3}$$

Finally, assume that both $n' \to \infty$ and $n'' \to \infty$ as $n \to \infty$. Then

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

Remark 1. The L_1 error of the above density estimate tends to zero regardedless of the form of the density of ϵ_1, provided the regression estimate is $L_1(\mathbf{P}_{X_1})$-consistent. By Cauchy-Schwarz, (3) is implied by

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \, \mathbf{P}_{X_1}(dx) \to 0 \quad \text{a.s.}, \tag{4}$$

and if we assume $\mathbf{E}\left\{Y_1^2\right\} < \infty$ there are many different estimates which are universally consistent in the sense that (4) holds for all distributions, cf., e.g., Devroye et al. (1994), Györfi and Walk (1996, 1997), Kohler and Krzyżak (2001), Lugosi and Zeger (1995), Nobel (1996), Walk (2002), or Györfi et al. (2002) and the literature cited therein. So under this additional assumption our estimate of the density of the error distribution in our regression model is strongly consistent in L_1 for **all** densities.

In Theorem 1, the generic pair (X, ϵ) is independent. Without this independence condition, one has to be much more careful. Noting that Theorem 1 was formulated in terms of a very general L_1-consistent regression estimate (3), the lack of independence will force one to at least make a specific choice of regression function estimate. This is captured in Theorem 2.

Theorem 2. *Let $d = 1$. Assume that $|n' - n''| \leq 1$, $n = n' + n''$. There exists a regression function m and a distribution of (X, ϵ) with $\mathbf{E}\{\epsilon|X\} = 0$, such that ϵ has a density f (with respect to the Lebesgue-measure) , and such that for any sequence of bandwiths h_n satisfying (2) there exists regression estimates m_n satisfying (3) with the property that the corresponding density estimate f_n from Section 2 in case of the naive kernel (i.e., the kernel $K(x) = (1/2)I_{[-1,1]}(x)$):*

$$\limsup_{n\to\infty} \int_{\mathbb{R}} |f_n(x) - f(x)| \, dx \geq 1.$$

Remark 2. Theorem 1 showed that when the pair (X, ϵ) is independent, the density estimate of Section 2 is strongly L_1-consistent whenever the regression estimate is strongly $L_1(\mathbf{P}_{X_1})$-consistent. By Theorem 2 this is no longer true in all generality if we omit the independence assumption. The counterexample in Theorem 2 has two special properties—first of all, conditional on X, ϵ is a shifted Bernoulli random variable without a density, but its marginal distribution has a density. Secondly, and more importantly, the example uses a special regression estimate that one would not encounter in statistical practice. It is still an open problem whether there exists any regression estimate such that the corresponding density estimate from Section 2 is strongly L_1-consistent for all distributions of (X, Y) where $\epsilon = Y - \mathbf{E}\{Y|X\}$ has a density f and where $\mathbf{E}|Y| < \infty$.

4 Proofs

4.1 Proof of Theorem 1

First we show

$$\int_{\mathbb{R}} |f_n(x) - f(x)|\, dx - \mathbf{E}\left\{ \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx \;\middle|\; \mathcal{D}_{n'} \right\} \to 0 \tag{5}$$

almost surely (applied conditioned on $\mathcal{D}_{n'}$).
Define

$$g(\hat{\epsilon}_1, \ldots, \hat{\epsilon}_{n''}) := \int_{\mathbb{R}} \left| \frac{1}{n'' \cdot h_{n''}} \sum_{j=1}^{n''} K\left(\frac{x - \hat{\epsilon}_j}{h_{n''}}\right) - f(x) \right| dx.$$

Scheffé's Lemma implies

$$\begin{aligned}
&\left| g(\hat{\epsilon}_1, \ldots, \hat{\epsilon}_{n''}) - g(\hat{\epsilon}_1, \ldots, \hat{\epsilon}_{i-1}, \hat{\epsilon}_i', \hat{\epsilon}_{i+1}, \hat{\epsilon}_{n''}) \right| \\
&\leq \int_{\mathbb{R}} \left| \frac{1}{h_{n''}} K\left(\frac{x - \hat{\epsilon}_i}{h_{n''}}\right) - \frac{1}{h_{n''}} K\left(\frac{x - \hat{\epsilon}_i'}{h_{n''}}\right) \right| dx \\
&= 2 \cdot \int_{\mathbb{R}} \left(\frac{1}{h_{n''}} K\left(\frac{x - \hat{\epsilon}_i}{h_{n''}}\right) - \frac{1}{h_{n''}} K\left(\frac{x - \hat{\epsilon}_i'}{h_{n''}}\right) \right)_+ dx \\
&\leq 2.
\end{aligned}$$

From McDiarmid's inequality (McDiarmid (1989); see also Devroye (1991) and Theorem A.2. in Györfi et. al. (2002)), we see that

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbf{P}\left[\left| \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx - \mathbf{E}\left\{ \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx \;\middle|\; \mathcal{D}_{n'} \right\} \right| \geq \epsilon \middle| \mathcal{D}_{n'} \right] \\
&\leq \sum_{n=1}^{\infty} \exp\left(\frac{-n'' \epsilon^2}{2} \right) < \infty.
\end{aligned}$$

We obtain (5) by an application of the lemma of Borel and Cantelli.

Next we show

$$\mathbf{E}\left\{ \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx \;\middle|\; \mathcal{D}_{n'} \right\} \to 0.$$

Scheffé's Lemma implies that

$$\int_{\mathbb{R}} |f_n(x) - f(x)|\, dx = 2 \cdot \int_{\mathbb{R}} (f(x) - f_n(x))_+ \, dx \leq 2 \cdot \int_B |f(x) - f_n(x)|\, dx + 2 \cdot \int_{B^c} f(x)\, dx.$$

Hence it suffices to show

$$\mathbf{E}\left\{ \int_B |f_n(x) - f(x)|\, dx \;\middle|\; \mathcal{D}_{n'} \right\} \to 0 \qquad (n \to \infty)$$

for any compact set $B \subseteq \mathbb{R}$. Let B be an arbitrary compact set in $\mathbb{R}$. Set

$$f_n^*(x) = \frac{1}{n'' \cdot h_{n''}} \sum_{j=1}^{n''} \mathbf{E}\left\{ K\left(\frac{x-\hat{\epsilon}_j}{h_{n''}}\right) \,\Big|\, \mathcal{D}_{n'} \right\}.$$

Then

$$\begin{aligned}
&\mathbf{E}\left\{ \int_B |f_n(x) - f(x)|\, dx \,\Big|\, \mathcal{D}_{n'} \right\} \\
&\leq \mathbf{E}\left\{ \int_B |f_n(x) - f_n^*(x)|\, dx \,\Big|\, \mathcal{D}_{n'} \right\} + \mathbf{E}\left\{ \int_B |f_n^*(x) - f(x)|\, dx \,\Big|\, \mathcal{D}_{n'} \right\}.
\end{aligned}$$

In the *first step of the proof* we show

$$\mathbf{E}\left\{ \int_B |f_n(x) - f_n^*(x)|\, dx \,\Big|\, \mathcal{D}_{n'} \right\} \to 0 \qquad (n \to \infty).$$

By Cauchy-Schwarz and the inequality of Jensen we have

$$\begin{aligned}
&\mathbf{E}\left\{ \int_B |f_n(x) - f_n^*(x)|\, dx \,\Big|\, \mathcal{D}_{n'} \right\} \\
&\leq \mathbf{E}\left\{ \sqrt{\int_B 1\ dx} \cdot \sqrt{\int_B |f_n(x) - f_n^*(x)|^2\, dx} \,\Bigg|\, \mathcal{D}_{n'} \right\} \\
&\leq \sqrt{\int_B 1\ dx} \cdot \sqrt{\mathbf{E}\left\{ \int_B |f_n(x) - f_n^*(x)|^2\, dx \,\Big|\, \mathcal{D}_{n'} \right\}}.
\end{aligned}$$

Now

$$\begin{aligned}
&\mathbf{E}\left\{ \int_B |f_n(x) - f_n^*(x)|^2\, dx \,\Big|\, \mathcal{D}_{n'} \right\} \\
&= \int_B \mathbf{E}\left\{ \left| \frac{1}{n'' \cdot h_{n''}} \sum_{j=1}^{n''} \left(K\left(\frac{x-\hat{\epsilon}_j}{h_{n''}}\right) - \mathbf{E}\left\{ K\left(\frac{x-\hat{\epsilon}_j}{h_{n''}}\right) \,\Big|\, \mathcal{D}_{n'} \right\} \right) \right|^2 \,\Bigg|\, \mathcal{D}_{n'} \right\} dx \\
&\leq \frac{1}{n''^2 \cdot h_{n''}^2} \sum_{j=1}^{n''} \int_B \mathbf{E}\left\{ K^2\left(\frac{x-\hat{\epsilon}_j}{h_{n''}}\right) \,\Big|\, \mathcal{D}_{n'} \right\} dx \\
&= \frac{1}{n''^2 \cdot h_{n''}^2} \sum_{j=1}^{n''} \int_B \int_{\mathbb{R}} K^2\left(\frac{x-u-m(X_{n'+j}) + m_{n'}(X_{n'+j})}{h_{n''}}\right) f(u)\ du\ dx \\
&= \frac{1}{n''^2 \cdot h_{n''}} \sum_{j=1}^{n''} \int_{\mathbb{R}} f(u) \int_B \frac{1}{h_{n''}} K^2\left(\frac{x-u-m(X_{n'+j}) + m_{n'}(X_{n'+j})}{h_{n''}}\right) dx\ du \\
&= \frac{1}{n'' \cdot h_{n''}} \int_{\mathbb{R}} K^2(z)\ dz \to 0 \qquad (n \to \infty).
\end{aligned}$$

In the *second step of the proof* we show

$$\mathbf{E}\int_{\mathbb{R}}\left|f_n^*(x)-\mathbf{E}\left\{\frac{1}{h_{n''}}K\left(\frac{x-\epsilon_1}{h_{n''}}\right)\right\}\right|dx\to 0\qquad (n\to\infty).$$

To do this we observe

$$\begin{aligned}
&\int_{\mathbb{R}}\left|f_n^*(x)-\mathbf{E}\left\{\frac{1}{h_{n''}}K\left(\frac{x-\epsilon_1}{h_{n''}}\right)\right\}\right|dx\\
&=\int_{\mathbb{R}}\Big|\frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-u+m_{n'}(X_{n'+j})-m(X_{n'+j})}{h_{n''}}\right)f(u)\,du\\
&\qquad\qquad -\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-u}{h_{n''}}\right)f(u)\,du\Big|\,dx\\
&\le\int_{\mathbb{R}}\frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-z}{h_{n''}}\right)\left|f\left(z-m(X_{n'+j})+m_{n'}(X_{n'+j})\right)-f(z)\right|dz\,dx\\
&\le\frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-z}{h_{n''}}\right)\cdot\big|f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))\\
&\qquad\qquad\qquad -g_f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))\big|dz\,dx\\
&\quad+\frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-z}{h_{n''}}\right)\cdot\left|g_f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))-g_f(z)\right|dz\,dx\\
&\quad+\frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-z}{h_{n''}}\right)\cdot|g_f(z)-f(z)|\,dz\,dx\\
&=:T_{1,n}+T_{2,n}+T_{3,n}
\end{aligned}$$

with an arbitrary density $g_f:\mathbb{R}^d\to\mathbb{R}_+$.

Application of Fubini's theorem yields

$$\begin{aligned}
T_{1,n} &= \frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}}\left|f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))-g_f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))\right|\\
&\qquad\cdot\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-z}{h_{n''}}\right)\,dx\,dz\\
&=\int_{\mathbb{R}}|f(z)-g_f(z)|\,\,dz
\end{aligned}$$

and

$$
\begin{aligned}
T_{3,n} &= \frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}} |g_f(z)-f(z)|\cdot\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-z}{h_{n''}}\right)\,dx\,dz\\
&= \int_{\mathbb{R}} |f(z)-g_f(z)|\,\,dz.
\end{aligned}
$$

Choosing g_f as a density which approximates f in L_1, both terms can be made arbitrary small.

Assume that g_f is Lipschitz-continous with Lipschitz-constant $L>0$ and let C be an arbitrary compact set in $\mathbb{R}$. By Scheffé's Lemma,

$$
\begin{aligned}
T_{2,n} &= \frac{1}{n''}\sum_{j=1}^{n''}\int_{\mathbb{R}} \left|g_f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))-g_f(z)\right|\cdot\int_{\mathbb{R}}\frac{1}{h_{n''}}K\left(\frac{x-z}{h_{n''}}\right)\,dx\,dz\\
&= \frac{1}{n''}\sum_{j=1}^{n''}2\cdot\int_{\mathbb{R}}\left(g_f(z)-g_f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))\right)_+\,dz\\
&\le 2\cdot\frac{1}{n''}\sum_{j=1}^{n''}\int_C \left|g_f(z)-g_f(z+m_{n'}(X_{n'+j})-m(X_{n'+j}))\right|\,dz+2\cdot\int_{C^c} g_f(z)\,dz\\
&\le 2\cdot L\cdot\frac{1}{n''}\sum_{j=1}^{n''}\left|m_{n'}(X_{n'+j})-m(X_{n'+j})\right|\int_C 1\,dz\,\,+2\cdot\int_{C^c} g_f(z)\,dz.
\end{aligned}
$$

So

$$
\limsup_{n\to\infty}\mathbf{E}\left\{T_{2,n}\,\middle|\,\mathcal{D}_{n'}\right\}\;\le\;2L\cdot\int_C 1\,dz\cdot\lim_{n\to\infty}\int_{\mathbb{R}^d}|m_{n'}(x)-m(x)|\,\mathbf{P}_{X_1}(dx)\,\,+2\int_{C^c} g_f(z)\,dz
$$

and with $C\uparrow\mathbb{R}$ we get

$$
\mathbf{E}\left\{T_{2,n}\,\middle|\,\mathcal{D}_{n'}\right\}\to 0\qquad a.s.
$$

From Devroye and Györfi (1985), Chapter 3, Theorem 1 we know

$$
\int_{\mathbb{R}}\left|\frac{1}{h_{n''}}\mathbf{E}\left\{K\left(\frac{x-\epsilon_1}{h_{n''}}\right)\right\}-f(x)\right|\,dx\to 0\qquad a.s.
$$

and the assertion is proved. □

4.2 Proof of Theorem 2

Without loss of generality, we assume that n is even, so that $n'=n''=n/2$. Let $i\in\{1,\ldots,n\}$ and $j\in\{1,\ldots,n''\}$. Let X_i be i.i.d. and uniformly distributed on $[0,1]$, and define

$$Y_i = \begin{cases} X_i & \text{with probability } \frac{1}{2} \\ -X_i & \text{with probabilty } \frac{1}{2}. \end{cases}$$

Then Y_i is uniformly distributed on $[-1,1]$, $\mathbf{E}\left\{Y_i \mid X_i\right\} = 0$ and $\epsilon_i = Y_i - \mathbf{E}\left\{Y_i \mid X_i\right\} = Y_i$. Define

$$m_{n'}(x) = x - x'$$

with $x' = \lfloor \frac{x}{\sqrt{h_{n'}}} \rfloor \cdot \sqrt{h_{n'}} = \lfloor \frac{x}{\sqrt{h_{n''}}} \rfloor \cdot \sqrt{h_{n''}}$. Since

$$|m_{n'}(x)| \leq \sqrt{h_{n'}} \to 0 \qquad (n \to \infty)$$

assumption (3) is valid. Estimating the residuals as described in Section 2 we get

$$\hat{\epsilon}_j = Y_{n'+j} - m_{n'}(X_{n'+j}) = \begin{cases} X'_{n'+j} & \text{with probability } \frac{1}{2} \\ -2X_{n'+j} + X'_{n'+j} & \text{with probability } \frac{1}{2}, \end{cases}$$

where $X'_{n'+j} = \lfloor \frac{X_{n'+j}}{\sqrt{h_{n''}}} \rfloor \cdot \sqrt{h_{n''}}$. Observe that $-2X_{n'+j} + X'_{n'+j} \leq 0$. Define

$$A_n := \bigcup_{k=0}^{\infty} \left[k\sqrt{h_n} - h_n, k\sqrt{h_n} + h_n \right].$$

Note that these intervals are disjoint if $h_n < \frac{1}{4}$. So there exists a natural number $N \in \mathbb{N}$ such that for every $n'' > N$, we have

$$\begin{aligned} \int_{[0,1]\cap A^c_{n''}} f(x)dx &= \sum_{k=0}^{\ell_{n''}-1} \int_{k\sqrt{h_{n''}}+h_{n''}}^{(k+1)\sqrt{h_{n''}}-h_{n''}} \frac{1}{2}dx \\ &= \frac{1}{2}\ell_{n''}\left(\sqrt{h_{n''}} - 2h_{n''}\right) \end{aligned}$$

with $\ell_{n''} := \max\left\{k \in \mathbb{N} : k\sqrt{h_{n''}} \leq 1\right\} = \lfloor \frac{1}{\sqrt{h_{n''}}} \rfloor$.

If $x \in A^c_{n''} \cap [0,1]^d$ then $\left|x - k\sqrt{h_{n''}}\right| > h_{n''}$ for every $k \in \mathbb{N}$ and consequently $|x - \hat{\epsilon}_j| > h_{n''}$ for all $j \in \mathbb{N}$. Using the naive kernel in the definition of f_n, we have for every $n'' > N$,

$$\int_{[0,1]\cap A^c_{n''}} f_n(x)\ dx \;=\; \int_{[0,1]\cap A^c_{n''}} \frac{1}{2n''h_{n''}} \sum_{j=1}^{n''} \mathbf{1}_{\{|x-\hat{\epsilon}_j| \leq h_{n''}\}}(x)dx = 0.$$

By Scheffé's Lemma, we see that for every $n'' > N$,

$$\begin{aligned} \int |f_n(x) - f(x)|\, dx &= 2 \cdot \sup_{B\in\mathcal{B}} \left| \int_B f(x)dx - \int_B f_n(x)dx \right| \\ &\geq 2 \cdot \int_{[0,1]\cap A^c_{n''}} f(x)dx - 2 \cdot \int_{[0,1]\cap A^c_{n''}} f_n(x)dx \\ &= \ell_{n''}\sqrt{h_{n''}} - \ell_{n''}\sqrt{h_{n''}}\sqrt{h_{n''}} \\ &\to 1 \qquad (n \to \infty) \end{aligned}$$

and the proof is complete.

References

[1] Ahmad, I. A. (1992). Residuals density estimation in nonparametric regression. *Statistics and Probability Letters*, **14**, pp. 133–139.

[2] Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric Estimation of the Residual Distribution. *Board of the Foundation of the Scandinavian Journal of Statistics, Blackwell Publishers Ltd*, **28**, pp. 549–567.

[3] Cheng, F. (2002). Consistency of error density and distribution function estimators in nonparametric regression. *Statistics and Probability Letters*, **59**, pp. 257–270.

[4] Cheng, F. (2004). Weak and strong uniform consistency of a kernel error density estimator in nonparametric regression. *Journal of Statistical Planning and Inference*, **119**, pp. 95–107.

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

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

[7] Devroye, L. (1991). Exponential inequalities in nonparametric estimation. *Nonparametric Functional Estimation and Related Topics*, edited by G. Roussas, pp. 31–44, Kluwer Academic Publishers, Dordrecht.

[8] 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.

[9] Devroye, L. and Lugosi, G. (2000). Combinatorial Methods in Density Estimation. *Springer-Verlag*, New York.

[10] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics*, **1**, pp. 279–290.

[11] Efromovich, S. (2005). Estimation of the density of regression errors. *Annals of Statistics*, **33**, pp. 2194–2227.

[12] Efromovich, S. (2006). Optimal nonparametric estimation of the density of regression errors with finite support. *AISM*, **59**, pp. 617–654.

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

[14] Györfi, L. and Walk, H. (1996). On the strong universal consistency of a series type regression estimate. *Mathematical Methods of Statistics*, **5**, pp. 332–342.

[15] Györfi, L. and Walk, H. (1997). On the strong universal consistency of a recursive regression estimate by Pál Révész. *Statistics and Probability Letters*, **31**, pp. 177–183.

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

[17] Loynes, R. M. (1980). The empirical sample distribution function of residuals from generalized regression. *Annals of Statistics*, **8**, pp. 285–298.

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

[19] McDiarmid, C. (1989). On the method of bounded differences. *Surveys in Combinatorics 1989*, vol. 141, pp. 148–188, London Mathematical Society Lecture Notes Series, Cambridge University Press, Cambridge.

[20] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, **101**, pp. 1067–1078.

[21] Nobel, A. (1996). Histogram regression estimation using data-dependent partitions. *Annals of Statistics*, **24**, pp. 1084–1105.

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

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

[24] Walk, H. (2002). Almost sure convergence properties of Nadaraya–Watson regression estimates. In: M. Dror, P. L'Ecuyer and F. Szidarovszky (editors), *Modeling Uncertainty: An Examination of its Theory, Methods and Applications*, pp. 201-223, Kluwer Academic Publishers, Dordrecht.