Estimation of a jump point in random design regression

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*

March 20, 2013

Abstract

Given an independent and identically distributed sample of the distribution of an $\mathbb{R} \times \mathbb{R}$-valued random vector (X, Y), where the corresponding regression function $m(x) = \mathbf{E}\{Y|X = x\}$ is piecewise continuous, the problem of estimation of the maximal jump point of the regression function is considered. Estimates are constructed which converge almost surely to the maximal jump point whenever the support of the independent variable X is a compact interval and the dependent variable Y satisfies some weak integrability condition.

AMS classification: Primary 62G08; secondary 62G20.

Key words and phrases: Jump point, nonparametric regression, strong consistency.

1 Introduction

Let (X, Y), (X_1, Y_1), (X_2, Y_2) ... be independent and identically distributed random variables with values in $\mathbb{R} \times \mathbb{R}$. Assume $\mathbf{E}|Y| < \infty$, let $m(x) = \mathbf{E}\{Y|X = x\}$ be the so-called

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

Running title: *Jump point estimation*

regression function, and let $\mu = \mathbf{P}_X$ be the distribution of the design variable X. Assume that m is uniformly continuous except for finitely many jump points, i.e., assume that there exist $N \in \mathbb{N}$, $z_1, \dots, z_N \in \mathbb{R}$ and $L : \mathbb{R}_+ \to \mathbb{R}_+$ such that $L(h) \to 0$ $(h \to 0)$ and for all $x, y \in \mathbb{R}, x < y$ with the property that $[x, y]$ does not contain any of the $z_1, \dots, z_N$ we have

$$|m(x) - m(y)| \leq L(|x - y|).$$

In this paper we consider the problem of estimating the location and the size of the maximal jump of m. More precisely, let

$$m^+(x) = \lim_{h \to 0, h > 0} m(x + h) \quad \text{and} \quad m^-(x) = \lim_{h \to 0, h > 0} m(x - h)$$

be the right-hand and left-hand limits of m. Then

$$\Delta(z) = |m^+(z) - m^-(z)|$$

is the size of the jump of m at z. Let $[a, b]$ be the support of X, which we assume to be a compact interval, and denote by z^* the location of the jump with the maximal size within (a, b), i.e.,

$$\Delta := \Delta(z^*) = \sup_{z \in (a,b)} \Delta(z). \tag{1}$$

Given the data

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

we want to construct estimates

$$\hat{\Delta}_n = \hat{\Delta}_n(\mathcal{D}_n) \quad \text{and} \quad \hat{z}_n = \hat{z}_n(\mathcal{D}_n)$$

such that

$$\hat{\Delta}_n \to \Delta \quad a.s.$$

and

$$\hat{z}_n \to z^* \quad a.s.$$

as $n \to \infty$.

Of course, the last convergence will be only possible in case that z^* is unique.

The most popular estimates for nonparametric regression include kernel regression estimate (cf., e.g., Watson (1964), Stone (1977) or Devroye and Krzyżak (1989)), partitioning regression estimate (cf., e.g., Beirlant and Györfi (1998)), nearest neighbor regression estimate (cf., e.g., Devroye, Györfi, Krzyżak and Lugosi (1994), or Zhao (1987)), local polynomial kernel estimates (cf., e.g., Stone (1982)), least squares estimates (cf., e.g., Lugosi and Zeger (1995)) or smoothing spline estimates (cf., e.g., Kohler and Krzyżak (2001)). The main theoretical results are summarized in the monograph by Györfi et al. (2002). Modifications of several of these estimates have already been applied to jump point regression in a random design setting in various papers. E.g., rate of convergence results have been derived in Gijbels, Hall and Kneip (1999) and Ma and Yang (2011), a data-driven choice of the bandwidth of kernel based jump point estimators has been investigated in Gijbels and Goderniaux (2004a), and jump points of the derivative of a regression function have been estimated in Gijbels and Goderniaux (2004b). But most papers for jump point regression derive results in the fixed design regression setting, see, e.g., Desmet and Gijbels (2011), Gijbels, Lambert and Qiu (2007), Jose and Ismail (2001) or Wu and Chu (1993) and the literature cited therein. Related techniques are also applied in change point estimation in connection with time series, see, e.g., Carlstein (1988), Hariz, Wylie and Zhang (2007), Lee (2011) or Rafajłowicz, Pawlak and Steland (2010).

In this paper we consider a standard kernel estimate of $\Delta(z)$, and the aim is to derive consistency results for this estimate under much more general conditions than usually considered in the literature, in particular we avoid any assumptions stipulating that the distribution of the design has a density with respect to the Lebesgue-Borel measure. If we want to avoid this assumption, we could use techniques from Devroye (1978a, 1978b) and try to construct estimates of m^+ and m^- which are consistent in the supremum norm whenever the distribution of X has the property that the probability of an interval is always greater than or equal to a constant times the length of the interval. However, in this paper we want to avoid even such an assumption. The key trick which allows us to derive consistency results for the estimates under even weaker conditions is that we use a data-dependent modification of the bandwidth of the kernel estimates: we start with some fixed value depending on the sample size and increase it until the intervals to the left and to the right of the point considered contain enough data points. We show consistency of our method under rather weak conditions: we assume that the regression function is uniformly continuous except for finitely many jump points, that the support of X is a compact interval and that Y satisfies some rather weak integrability condition. We prove that our estimates are strongly consistent in a sense that the estimates of the maximal jump size and of the jump point converge almost surely to the true values provided the sample size goes to infinity.

1.1 Notation

Throughout this paper we use the following notations: μ denotes the distribution of X and $m(x) = \mathbf{E}\{Y|X = x\}$ is the regression function of (X, Y).

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

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

For $A \subseteq \mathbb{R}$ let I_A be the indicator function of A, i.e.,

$$I_A(x) := \begin{cases} 1 & \text{if} \quad x \in A, \\ 0 & \text{if} \quad x \notin A, \end{cases}$$

for $x \in \mathbb{R}$. Furthermore we define

$$\log_+ z := \begin{cases} \log(z) & \text{if} \quad z \geq 1, \\ 0 & \text{if} \quad z < 1, \end{cases}$$

for $z \in \mathbb{R}_+$.

1.2 Outline

The definition of the estimates are given in Section 2, the main result is formulated in Section 3 and proven in Section 4.

2 Definition of the estimate

Set $K_+ = I_{(0,1]}$ and $K_- = I_{[-1,0)}$ and define kernel type estimates of m^+ and m^- as follows: First choose $h_n > 0$, which is a parameter of the estimate. In our estimate we increase h_n in a data-dependent way such that the intervals to the left and to the right contain enough data points. To do this, we let μ_n be the empirical distribution of $X_1, \ldots,$ X_n, i.e.,

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} I_A(X_i) \quad (A \subseteq \mathbb{R}),$$

and we set

$$h_n(x) = \inf \left\{ h \geq h_n \quad : \quad \mu_n([x-h, x)) \geq 1/\log(n) \quad \text{and} \quad \mu_n((x, x+h]) \geq 1/\log(n) \right\},$$

where $\inf \emptyset = \infty$. In case $h_n(x) < \infty$ we define

$$m_n^+(x) = \frac{\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right) \cdot Y_i}{\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right)}$$

and

$$m_n^-(x) = \frac{\sum_{i=1}^n K_-\left(\frac{x-X_i}{h_n(x)}\right) \cdot Y_i}{\sum_{i=1}^n K_-\left(\frac{x-X_i}{h_n(x)}\right)}.$$

Using m_n^+ and m_n^-, we define an estimate of $\Delta(x)$ by

$$\Delta_n(x) = |m_n^+(x) - m_n^-(x)|$$

and estimate Δ by

$$\hat{\Delta}_n = \max_{x \in \mathbb{R} \,:\, h_n(x) < \infty} \Delta_n(x)$$

and z^* by

$$\hat{z}_n = \arg \max_{x \in \mathbb{R} \,:\, h_n(x) < \infty} \Delta_n(x).$$

3 Main result

Let the estimates $\hat{\Delta}_n$ and $\hat{z}_n$ be defined as in the previous section. Then the following result is valid:

Theorem 1 *Let (X, Y), (X_1, Y_1), ... be independent and identically distributed random variables with values in $\mathbb{R} \times \mathbb{R}$. Assume that the support of X is a compact interval,*

$$\mathbf{E}\left(|Y| \log_+ |Y|\right) < \infty, \tag{2}$$

and that the regression function is uniformly continuous except for finitely many jump points, i.e., assume that there exist $N \in \mathbb{N}$, $z_1, \ldots, z_N \in \mathbb{R}$ and $L : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$L(h) \to 0$ $(h \to 0)$ and for all $x, y \in \mathbb{R}, x < y$ with the property that $[x, y]$ does not contain any of the $z_1, \dots, z_N$ we have

$$|m(x) - m(y)| \leq L(|x - y|).$$

Let Δ and z^* be defined as in Section 1 (cf., (1), where $[a, b] = supp(X)$) and define the estimates $\hat{\Delta}_n$ and $\hat{z}_n$ as in the previous section. Assume that $h_n > 0$ satisfies

$$h_n \to 0 \quad (n \to \infty). \tag{3}$$

Then the following assertions hold:

a)

$$\hat{\Delta}_n \to \Delta \quad a.s. \tag{4}$$

as $n \to \infty$.

b) *If, in addition, z^* is uniquely defined, then*

$$\hat{z}_n \to z^* \quad a.s. \tag{5}$$

as $n \to \infty$.

Remark. It follows from the proof of Theorem 1 that in case that z^* is not uniquely defined but the other assumptions in Theorem 1 hold we have

$$\min_{i=1,\dots,N \,:\, \Delta(z_i)=\Delta} |\hat{z}_n - z_i| \to 0 \quad a.s.$$

4 Proofs

Lemma 1 *Assume that (3) holds and define*

$$m^+_{h_n}(x) = \frac{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right) \cdot Y|\mathcal{D}_n\right\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\}} \quad \text{and} \quad m^-_{h_n}(x) = \frac{\mathbf{E}\left\{K_-\left(\frac{x-X}{h_n(x)}\right) \cdot Y|\mathcal{D}_n\right\}}{\mathbf{E}\left\{K_-\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\}}.$$

Then

$$\sup_{x\in\mathbb{R} \,:\, h_n(x)<\infty} \left|m^+_n(x) - m^+_{h_n}(x)\right| \to 0 \quad a.s. \tag{6}$$

and

$$\sup_{x\in\mathbb{R} \,:\, h_n(x)<\infty} \left|m^-_n(x) - m^-_{h_n}(x)\right| \to 0 \quad a.s. \tag{7}$$

as $n \to \infty$.

Proof. The proof is a more or less straightforward modification of results from Kohler, Krzyżak and Walk (2011) and Hertel and Kohler (2013). For the sake of completeness we nevertheless present in the sequel a complete version of the proof of (6). (7) can be proven in the same way.

Let $s \in (0, \frac{1}{2})$ be arbitrary. Using well-known results from VC-theory (cf., e.g., Theorem 12.5 and Theorem 13.7 in Devroye, Györfi and Lugosi (1996)) we get

$$\mathbf{P}\left\{ \sup_{x\in\mathbb{R},h>0} |\mu_n((x,x+h]) - \mathbf{P}_X((x,x+h])| > \epsilon \right\} \leq 8 \cdot n^2 \cdot e^{-\frac{n\cdot\epsilon^2}{32}}.$$

Furthermore we can conclude from Theorem 9.1 in Györfi et al. (2002) and Theorem 13.7 in Devroye, Györfi and Lugosi (1996)

$$\mathbf{P}\left\{ \sup_{x\in\mathbb{R},h>0} \left| \frac{1}{n}\sum_{i=1}^{n} Y_i \cdot 1_{\{|Y_i|\leq n^s\}} \cdot K_+\left(\frac{x-X_i}{h}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\leq n^s\}} \cdot K_+\left(\frac{x-X}{h}\right)\} \right| > \epsilon \right\}$$
$$\leq 8 \cdot n^2 \cdot e^{-\frac{n\cdot\epsilon^2}{128\cdot n^{2s}}}.$$

Application of the Borel-Cantelli lemma yields

$$\frac{\sup_{x\in\mathbb{R},h>0} |\mu_n((x,x+h]) - \mathbf{P}_X((x,x+h])|}{\log n/\sqrt{n}} \to 0 \quad a.s., \tag{8}$$

and

$$\frac{\sup_{x\in\mathbb{R},h>0} \left| \frac{1}{n}\sum_{i=1}^{n} Y_i \cdot 1_{\{|Y_i|\leq n^s\}} \cdot K_+\left(\frac{x-X_i}{h}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\leq n^s\}} \cdot K_+\left(\frac{x-X}{h}\right)\} \right|}{\log n/n^{\frac{1}{2}-s}} \to 0 \quad a.s. \tag{9}$$

as $n \to \infty$.

By definition of $h_n(x)$ we have in case $h_n(x) < \infty$

$$\frac{1}{n}\sum_{i=1}^{n} K_+\left(\frac{x-X_i}{h_n(x)}\right) = \mu_n((x, x+h_n(x)]) \geq \frac{1}{\log(n)}, \tag{10}$$

which implies

$$\left| \frac{\sum_{i=1}^{n} Y_i \cdot K_+\left(\frac{x-X_i}{h_n(x)}\right)}{\sum_{i=1}^{n} K_+\left(\frac{x-X_i}{h_n(x)}\right)} - \frac{\mathbf{E}\left\{Y \cdot K_+\left(\frac{x-X}{h_n(x)}\right)\right\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\}} \right|$$
$$= \left| \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i \cdot 1_{\{|Y_i|>n^s\}} \cdot K_+\left(\frac{x-X_i}{h_n(x)}\right)}{\frac{1}{n}\sum_{i=1}^{n} K_+\left(\frac{x-X_i}{h_n(x)}\right)} + \frac{\sum_{i=1}^{n} Y_i \cdot 1_{\{|Y_i|\leq n^s\}} \cdot K_+\left(\frac{x-X_i}{h_n(x)}\right)}{\sum_{i=1}^{n} K_+\left(\frac{x-X_i}{h_n(x)}\right)} \right.$$

$$
-\frac{\mathbf{E}\left\{Y\cdot 1_{\{|Y|\le n^s\}}\cdot K_+\left(\frac{x-X}{h_n(x)}\right)\right\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\}}-\frac{\mathbf{E}\left\{Y\cdot 1_{\{|Y|>n^s\}}\cdot K_+\left(\frac{x-X}{h_n(x)}\right)\right\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\}}\Bigg|
$$

$$
\begin{aligned}
&\le \frac{\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot 1_{\{|Y_i|>n^s\}}}{1/\log(n)}+\frac{\mathbf{E}\{|Y|\cdot 1_{\{|Y|>n^s\}}\}}{\mathbf{P}_X((x,x+h_n(x)])}\\
&\quad+\left|\frac{\sum_{i=1}^n Y_i\cdot 1_{\{|Y_i|\le n^s\}}\cdot K_+\left(\frac{x-X_i}{h_n(x)}\right)}{\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right)}-\frac{\mathbf{E}\{Y\cdot 1_{\{|Y|\le n^s\}}K_+\left(\frac{x-X}{h_n(x)}\right)\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\}}\right|\\
&=: T_{1,n}+T_{2,n}+T_{3,n}.
\end{aligned}
$$

Hence it suffices to show

$$
\sup_{x\in\mathbb{R}:h_n(x)<\infty} T_{i,n}\to 0 \quad a.s. \tag{11}
$$

as $n\to\infty$ for $i\in\{1,2,3\}$.

For $i=1$ we have for any $L>1$ and n sufficiently large

$$
\begin{aligned}
\frac{\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot 1_{\{|Y_i|>n^s\}}}{1/\log(n)} &\le \frac{\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot \frac{\log|Y_i|}{\log(n^s)}\cdot 1_{\{|Y_i|>n^s\}}}{1/\log(n)}\\
&= \frac{1}{s}\cdot\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot\log|Y_i|\cdot 1_{\{|Y_i|>n^s\}}\\
&\le \frac{1}{s}\cdot\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot\log|Y_i|\cdot 1_{\{|Y_i|>L\}}\\
&\to \frac{1}{s}\cdot\mathbf{E}\left\{|Y|\cdot\log|Y|\cdot 1_{\{|Y|>L\}}\right\} \quad a.s.
\end{aligned}
$$

as $n\to\infty$ by (2) and by the strong law of large numbers. And because of (2) we get

$$
\mathbf{E}\left\{|Y|\cdot\log|Y|\cdot 1_{\{|Y|>L\}}\right\}\to 0
$$

for $L\to\infty$, from which (11) follows for $i=1$.

For $i=2$ we observe

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}:h_n(x)<\infty}\frac{\mathbf{E}\{|Y|\cdot 1_{\{|Y|>n^s\}}\}}{\mathbf{P}_X((x,x+h_n(x)])}\\
&\le \sup_{x\in\mathbb{R}:h_n(x)<\infty}\frac{\mathbf{E}\{|Y|\cdot\frac{\log(|Y|)}{\log(n^s)}\cdot 1_{\{|Y|>n^s\}}\}}{\mu_n((x,x+h_n(x)])-(\mu_n((x,x+h_n(x)])-\mathbf{P}_X((x,x+h_n(x)])}\\
&\le \frac{1}{s}\cdot\frac{\mathbf{E}\{|Y|\cdot\log(|Y|)\cdot 1_{\{|Y|>n^s\}}\}}{\log(n)\cdot\left(1/\log(n)-\sup_{x\in\mathbb{R}:h_n(x)<\infty}|\mu_n((x,x+h_n(x)])-\mathbf{P}_X((x,x+h_n(x)])|\right)}\\
&= \frac{1}{s}\cdot\frac{\mathbf{E}\{|Y|\cdot\log(|Y|)\cdot 1_{\{|Y|>n^s\}}\}}{1-\log(n)\cdot\sup_{x\in\mathbb{R}:h_n(x)<\infty}|\mu_n((x,x+h_n(x)])-\mathbf{P}_X((x,x+h_n(x)])|}.
\end{aligned}
$$

Because of (2) we have

$$\mathbf{E}\{|Y| \cdot \log(|Y|) \cdot 1_{\{|Y|>n^s\}}\} \to 0 \quad (n \to \infty),$$

and together with (8) this implies (11) for $i = 2$.

In order to show (11) for $i = 3$ we observe that because of (10) we have for any $x \in \mathbb{R}$ satisfying $h_n(x) < \infty$

$$\begin{aligned}
&\left| \frac{\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i| \le n^s\}} \cdot K_+\left(\frac{x-X_i}{h_n(x)}\right)}{\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right)} - \frac{\mathbf{E}\{Y \cdot 1_{\{|Y|\le n^s\}} K_+\left(\frac{x-X}{h_n(x)}\right)\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\}} \right| \\
&= \Bigg| \frac{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\} \cdot \left(\frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^s\}} \cdot K_+\left(\frac{x-X_i}{h_n(x)}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\le n^s\}} \cdot K_+\left(\frac{x-X}{h_n(x)}\right)\}\right)}{\frac{1}{n}\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right) \cdot \mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\}} \\
&\qquad + \frac{\mathbf{E}\{Y \cdot 1_{\{|Y|\le n^s\}} \cdot K_+\left(\frac{x-X}{h_n(x)}\right)\} \cdot \left(\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\} - \frac{1}{n}\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right)\right)}{\frac{1}{n}\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right) \cdot \mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\}} \Bigg| \\
&\le \frac{\left|\frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^s\}} \cdot K_+\left(\frac{x-X_i}{h_n(x)}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\le n^s\}} \cdot K_+\left(\frac{x-X}{h_n(x)}\right)\}\right|}{\mu_n((x, x+h_n(x)])} \\
&\qquad + \frac{\mathbf{E}\{|Y| \cdot 1_{\{|Y|\le n^s\}} \cdot K_+\left(\frac{x-X}{h_n(x)}\right)\}}{\mathbf{P}_X((x, x+h_n(x)])} \cdot \frac{\left|\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)\right\} - \frac{1}{n}\sum_{i=1}^n K_+\left(\frac{x-X_i}{h_n(x)}\right)\right|}{1/\log(n)} \\
&\le \frac{\left|\frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^s\}} \cdot K_+\left(\frac{x-X_i}{h_n(x)}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\le n^s\}} \cdot K_+\left(\frac{x-X}{h_n(x)}\right)\}\right|}{1/\log(n)} \\
&\quad + \frac{\log(n)^3/\sqrt{n} \cdot \mathbf{E}\{|Y| \cdot 1_{\{|Y|\le n^s\}}\}}{\log(n) \cdot (1/\log(n) - (\mu_n((x, x+h_n(x)]) - \mathbf{P}_X((x, x+h_n(x)])}\\
&\qquad\qquad \cdot \frac{|\mathbf{P}_X((x, x+h_n(x)]) - \mu_n((x, x+h_n(x)])|}{\log(n)/\sqrt{n}}.
\end{aligned}$$

Because of

$$\log(n)^3/\sqrt{n} \cdot \mathbf{E}\{|Y| \cdot 1_{\{|Y|\le n^s\}}\} \le \log(n)^3/\sqrt{n} \cdot \mathbf{E}\{|Y|\} \to 0 \quad (n \to \infty)$$

and

$$\liminf_{n\to\infty} \log(n) \cdot (1/\log(n) - \sup_{x\in\mathbb{R}} |\mu_n((x, x+h_n(x)]) - \mathbf{P}_X((x, x+h_n(x)])|) > 0$$

(which follows from (8)) we conclude from (8) and (9) that (11) also holds for $i = 3$. The proof is complete. □

Lemma 2 *Define*

$$\Delta_{h_n} = \sup_{x \in \mathbb{R}\,:\,h_n(x)<\infty} \left| m^+_{h_n}(x) - m^-_{h_n}(x) \right| .$$

Then

$$\hat{\Delta}_n - \Delta_{h_n} \to 0 \quad a.s.$$

as $n \to \infty$.

Proof. Using

$$\left| \sup_{x \in A} f(x) - \sup_{x \in A} g(x) \right| \le \sup_{x \in A} |f(x) - g(x)|$$

and the second triangle inequality we get

$$\begin{aligned}
&\left| \hat{\Delta}_n - \Delta_{h_n} \right| \\
&\le \sup_{x \in \mathbb{R}\,:\,h_n(x)<\infty} \left| \left| m^+_n(x) - m^-_n(x) \right| - \left| m^+_{h_n}(x) - m^-_{h_n}(x) \right| \right| \\
&\le \sup_{x \in \mathbb{R}\,:\,h_n(x)<\infty} \left| m^+_n(x) - m^+_{h_n}(x) \right| + \sup_{x \in \mathbb{R}\,:\,h_n(x)<\infty} \left| m^-_n(x) - m^-_{h_n}(x) \right| .
\end{aligned}$$

Lemma 1 implies the assertion. □

Lemma 3 *Assume that the support of* X *is a compact interval and that (3) holds. Then*

$$\sup_{x \in \mathbb{R}\,:\,h_n(x)<\infty} h_n(x) \to 0 \quad a.s.$$

as $n \to \infty$.

Proof. We have

$$h_n(x) = \max\{h^+_n(x), h^-_n(x)\}$$

where

$$h^+_n(x) = \inf \left\{ h \ge h_n \quad : \quad \mu_n((x, x+h]) \ge 1/\log(n) \right\}$$

and

$$h^-_n(x) = \inf \left\{ h \ge h_n \quad : \quad \mu_n([x-h, x)) \ge 1/\log(n) \right\},$$

hence it suffices to show

$$\sup_{x \in \mathbb{R}\,:\,h_n(x)<\infty} h^+_n(x) \to 0 \quad a.s. \tag{12}$$

and

$$\sup_{x \in \mathbb{R}\,:\,h_n(x)<\infty} h^-_n(x) \to 0 \quad a.s. \tag{13}$$

as $n \to \infty$.

(12) follows from

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}\,:\,h_n(x)<\infty} h_n^+(x) \le \epsilon \quad a.s. \tag{14}$$

for all $\epsilon > 0$, which we show next. Assume to the contrary that (14) does not hold. Then we can find $\epsilon > 0$ and a sequence of random points $x_{n_k} \in \mathbb{R}$ such that $h_{n_k}(x_{n_k}) < \infty$ and with probability greater than zero for all $k \in \mathbb{N}$ we have $h_{n_k}^+(x_{n_k}) \ge \epsilon$. The first property implies that x_{n_k} is contained in the support $supp(X) = [a, b]$ of X. Since $[a, b]$ is a compact interval, we know by the theorem of Bolzano-Weierstrass that a (random) subsequence of $(x_{n_k})_k$ converges almost surely to some random point $x \in [a, b]$. W.l.o.g. we denote this subsequence again by $(x_{n_k})_k$. Because of $h_{n_k}^+(x_{n_k}) \ge \epsilon$ for all $k \in \mathbb{N}$ with probability greater than zero we have $x_{n_k} \le b - \epsilon$, and consequently we have $x \in [a, b - \epsilon]$ with probability greater than zero. But then the (random) interval $(x + \epsilon/4, x + \epsilon/2]$ contains a (random) positive mass, which is bounded away from zero (since we can cover $[a, b]$ by finitely many disjoint intervals of length $\epsilon/16$, where each interval has positive mass, and each interval of length $\epsilon/4$ will contain at least one of these intervals of length $\epsilon/16$), which implies (via the Glivenko-Cantelli theorem) that it will with probability one eventually contain more than $n/\log(n)$ data points and consequently we have that with probability one $h_{n_k}^+(x_{n_k}) \le \epsilon/2$ for k large enough, which is a contradiction to $h_{n_k}^+(x_{n_k}) \ge \epsilon$ with probability greater than zero. This proves (12), and a straightforward modification of the above arguments leads to (13). The proof is complete. □

Lemma 4 *Assume that (3) holds, that $supp(X) = [a, b]$ is a compact interval and that m is uniformly continuous except for finitely many jump points. Then*

$$\Delta_{h_n} \to \Delta \quad a.s.$$

as $n \to \infty$.

Proof. Because of Lemma 3 we can assume without loss of generality that in case $h_n(x) < \infty$ we have that $h_n(x)$ is so small that the interval $[x - h_n(x), x + h_n(x)]$ contains at most one of the finitely many jump points of m. Let $x \in \mathbb{R}$ be such that $h_n(x) < \infty$, which implies that with probability one x has the property

$$\mathbf{E}\left\{K_-\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\} > 0 \quad \text{and} \quad \mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\} > 0.$$

Consequently,

$$
\begin{aligned}
&\left|m^+_{h_n}(x) - m^-_{h_n}(x)\right| \\
&= \left| \frac{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right) \cdot m(X)|\mathcal{D}_n\right\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\}} - \frac{\mathbf{E}\left\{K_-\left(\frac{x-X}{h_n(x)}\right) \cdot m(X)|\mathcal{D}_n\right\}}{\mathbf{E}\left\{K_-\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\}} \right| \\
&= \left| \frac{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right) \cdot (m(X) - m^+(x))|\mathcal{D}_n\right\}}{\mathbf{E}\left\{K_+\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\}} + \left(m^+(x) - m^-(x)\right) \right. \\
&\qquad \left. - \frac{\mathbf{E}\left\{K_-\left(\frac{x-X}{h_n(x)}\right) \cdot (m(X) - m^-(x))|\mathcal{D}_n\right\}}{\mathbf{E}\left\{K_-\left(\frac{x-X}{h_n(x)}\right)|\mathcal{D}_n\right\}} \right| \\
&\leq 3 \cdot L(h_n(x)) + \Delta.
\end{aligned}
$$

By Lemma 3 and $L(h) \to 0$ $(h \to 0)$ we conclude

$$
\limsup_{n\to\infty} \Delta_{h_n} \leq \Delta \quad a.s.
$$

Furthermore, for $z^* \in (a, b)$ such that

$$
|m^+(z^*) - m^-(z^*)| = \Delta
$$

we have that both of the intervals $[a, z^*)$ and $(z^*, b]$ contain positive mass, so by the Glivenko-Cantelli theorem we know that with probability one $h_n(z^*) < \infty$ for n large enough. By Lemma 2 we can assume w.l.o.g. that $h_n(z^*)$ is so small that the only jump point of m within $[z^* - h_n(z^*), z^* + h_n z^*]$ is z^* itself. Consequently,

$$
\begin{aligned}
&|m^+_{h_n}(z^*) - m^-_{h_n}(z^*)| \\
&= \left|(m^+(z^*) - m^-(z^*)) - (m^+(z^*) - m^+_{h_n}(z^*)) - (m^-_{h_n}(z^*) - m^-(z^*))\right| \\
&\geq |m^+(z^*) - m^-(z^*)| - |m^+(z^*) - m^+_{h_n}(z^*)| - |m^-_{h_n}(z^*) - m^-(z^*)| \\
&\geq \Delta - 2 \cdot L(h_n(z^*)).
\end{aligned}
$$

By using again Lemma 3 and $L(h) \to 0$ $(h \to 0)$ we can conclude

$$
\liminf_{n\to\infty} \Delta_{h_n} \geq \Delta \quad a.s.
$$

The proof is complete. □

Lemma 5 *Assume that the assumptions of Theorem 1 b) hold. Then*

$$\hat{z}_n \to z^* \quad a.s.$$

Proof. It suffices to show

$$\limsup_{n\to\infty} |\hat{z}_n - z^*| \le \epsilon \quad a.s. \tag{15}$$

for all $\epsilon > 0$. Assume to the contrary that (15) does not hold. Then we can find $\epsilon > 0$ and a subsequence $(\hat{z}_{n_k})_k$ of $(\hat{z}_n)_n$ such that

$$|\hat{z}_{n_k} - z^*| \ge \epsilon$$

for all k with probability greater than zero. Since $h_{n_k}(\hat{z}_{n_k}) < \infty$ by definition of $\hat{z}_{n_k}$ we can conclude from Lemma 3 that we have with probability one

$$h_{n_k}(\hat{z}_{n_k}) \le \epsilon$$

for k large enough and that for k large enough $[\hat{z}_{n_k} - h_{n_k}(\hat{z}_{n_k}), \hat{z}_{n_k} + h_{n_k}(\hat{z}_{n_k})]$ contains at most one of the jump points of m. Hence neither $[\hat{z}_{n_k} - h_{n_k}(\hat{z}_{n_k}), \hat{z}_{n_k})$ nor $(\hat{z}_{n_k}, \hat{z}_{n_k} + h_{n_k}(\hat{z}_{n_k})]$ can contain z^*. Using this we conclude that with probability one

$$\begin{aligned}
&\left| m^+_{h_{n_k}}(\hat{z}_{n_k}) - m^-_{h_{n_k}}(\hat{z}_{n_k}) \right| \\
&= \left| \frac{\mathbf{E}\left\{ K_+\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) \cdot m(X) | \mathcal{D}_n \right\}}{\mathbf{E}\left\{ K_+\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) | \mathcal{D}_n \right\}} - \frac{\mathbf{E}\left\{ K_-\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) \cdot m(X) | \mathcal{D}_n \right\}}{\mathbf{E}\left\{ K_-\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) | \mathcal{D}_n \right\}} \right| \\
&= \left| \frac{\mathbf{E}\left\{ K_+\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) \cdot (m(X) - m^+(\hat{z}_{n_k})) | \mathcal{D}_n \right\}}{\mathbf{E}\left\{ K_+\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) | \mathcal{D}_n \right\}} + \left(m^+(\hat{z}_{n_k}) - m^-(\hat{z}_{n_k}) \right) \right. \\
&\qquad \left. - \frac{\mathbf{E}\left\{ K_-\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) \cdot (m(X) - m^-(\hat{z}_{n_k})) | \mathcal{D}_n \right\}}{\mathbf{E}\left\{ K_-\left(\frac{\hat{z}_{n_k}-X}{h_n(\hat{z}_{n_k})}\right) | \mathcal{D}_n \right\}} \right| \\
&\le 3 \cdot L(h_{n_k}(\hat{z}_{n_k})) + \max_{i \in \{j \in \{1,\dots,N\} : z_j \ne z^*\}} \Delta(z_i)
\end{aligned}$$

for k large enough. By Lemma 3 and $L(h) \to 0$ $(h \to 0)$ we conclude

$$\limsup_{k\to\infty} \Delta_{h_{n_k}}(\hat{z}_{n_k}) = \limsup_{k\to\infty} \left| m^+_{h_n}(\hat{z}_{n_k}) - m^-_{h_n}(\hat{z}_{n_k}) \right| \le \max_{i \in \{j \in \{1,\dots,N\} : z_j \ne z^*\}} \Delta(z_i)$$

Furthermore, the proof of Lemma 2 implies

$$\Delta_{h_n}(\hat{z}_n) - \hat{\Delta}_n(\hat{z}_n) \to 0 \quad a.s. \quad (n \to \infty)$$

and by Lemma 2 and Lemma 4 we know

$$\hat{\Delta}_n(\hat{z}_n) = \hat{\Delta}_n \to \Delta \quad a.s.$$

as $n \to \infty$, hence

$$\begin{aligned}
\Delta &= \lim_{k\to\infty} \hat{\Delta}_{n_k}(\hat{z}_{n_k}) \\
&= \lim_{k\to\infty} \left(\hat{\Delta}_{n_k}(\hat{z}_{n_k}) - \Delta_{h_{n_k}}(\hat{z}_{n_k}) + \Delta_{h_{n_k}}(\hat{z}_{n_k}) \right) \\
&= \limsup_{k\to\infty} \Delta_{h_{n_k}}(\hat{z}_{n_k}) \\
&\leq \max_{i\in\{j\in\{1,\dots,N\}\,:\, z_j\neq z^*\}} \Delta(z_i) < \Delta,
\end{aligned}$$

which is the desired contradiction. The proof is complete. □

Proof of Theorem 1. Part a) follows from Lemma 2 and Lemma 4, part b) is proven in Lemma 5. □

References

[1] Beirlant, J. and Györfi, L. (1998). On the asymptotic normality of the L2-error in partitioning regression estimation *Journal of Statistical Planning and Inference* **71**, pp. 93–107.

[2] Carlstein, E. (1988). Nonparametric change-point estimation. *Annals of Statistics* **16**, pp. 188-197.

[3] Desmet, L. and Gijbels, I. (2011). Curve fitting under jump and peak irregularities using local linear regression. *Communications in Statistics - Theory and Methods* **40**, pp. 4001-4020.

[4] Devroye, L. (1978a). The uniform convergence of the Nadarya-Watson regression function estimate. *Canadian Journal of Statistics* **6**, pp. 179-191.

[5] Devroye, L. (1978b). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transactions on Information Theory* **24**, pp. 142-151.

[6] 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.

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

[8] 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.

[9] Gijbels, I. and Goderniaux, A.-C. (2004a). Bandwidth selection for changepoint estimation in nonparametric regression. *Technometrics* **46**, pp. 76-86.

[10] Gijbels, I. and Goderniaux, A.-C. (2004b). Data-driven discontinuity detection in derivatives of a regression function. *Communications in Statistics - Theory and Methods* **33**, pp. 851-871.

[11] Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.***51**, pp. 231-251.

[12] Gijbels, I., Lambert, A. and Qiu, P. (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Ann. Inst. Statist. Math.* **59**, pp. 235-272.

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

[14] Hariz, S. B., Wylie, J. J. and Zhang, Q. (2007). Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *Annals of Statistics* **35**, pp. 1802-1826.

[15] Hertel, I. and Kohler, M. (2013). Estimation of the optimal design of a nonlinear parametric regression problem via Monte Carlo experiments. *Computational Statistics and Data Analysis* (to appear).

[16] Jose, C. T. and Ismail, B. (2001). Nonparametric inference on jump regression surface. *Nonparametric Statistics* **13**, pp. 791-813.

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

[18] Kohler, M., Krzyżak, A. and Walk, H. (2011). Estimation of the essential supremum of a regression function. *Statistics and Probability Letters* **81**, pp. 685-693.

[19] Lee, T.-S. (2011). Change-point problems: bibliography and review. *Journal of Statistical Theory and Practice* **4**, pp. 643-662.

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

[21] Ma, S. and Yang, L. (2011). A jump-detecting procedure based on spline estimation. *Journal of Nonparametric Statistics* **23**, pp. 67-81.

[22] Rafajłowicz, E., Pawlak, M. and Steland, A. (2010). Nonparametric sequential change-point detection by a vertically trimmed box method. *IEEE Transactions on Information Theory* **56**, pp. 3621-3634.

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

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

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

[26] Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. *Annals of Statistics* **21**, pp. 1545-1566.

[27] Zhao, L. C. (1987). Exponential bounds of mean error for the nearest neighbor estimates of regression functions. *Journal of Multivariate Analysis* **21**, pp. 168–178.