

On uniform error of kernel estimate of discontinuous regression function

Alfredas RAČKAUSKAS (VU, MII), Aurelijus TAMULIS (MII)
e-mail: tamulis@maf.vu.lt

1. Introduction and results

Suppose the measurements Y_i at fixed design points $t_i = i/n$, $i = 1, \dots, n$, satisfy the following regression model:

$$Y_i = g(t_i) + \epsilon_i, \quad 1 \leq i \leq n, \quad (1)$$

where g is an unknown regression function and ϵ_i are independent identically distributed errors with zero mean $E(\epsilon_i) = 0$ and finite variance $\text{var}(\epsilon_i) = \sigma^2 < \infty$. We will assume that the function g has a jump of magnitude Δ at a change-point τ , $0 < \tau < 1$ and is smooth elsewhere. More precisely we assume that

$$g(t) = f(t) + \Delta 1_{[\tau, 1]}(t), \quad t \in [0, 1],$$

where f is a continuous function. We will consider the case where neither τ nor Δ are known. For this model Kee–Hoon Kang *et al.* [1] suggested the following procedure of kernel estimation of the function g . Using preliminary estimators of τ and Δ , say $\hat{\tau}$ and $\hat{\Delta}$ respectively, one computes the adjusted data

$$Y_i^* = Y_i - \hat{\Delta} 1_{[\hat{\tau}, 1]}(t_i), \quad 1 \leq i \leq n.$$

The regression function g is estimated then by

$$\hat{g}^*(t) = \frac{1}{h} \sum_{i=1}^n Y_i^* \int_{s_{i-1}}^{s_i} K^*\left(\frac{t-u}{h}\right) du + \hat{\Delta} 1_{[\hat{\tau}, 1]}(t), \quad (2)$$

where $s_0 = 0$, $s_n = 1$, $s_k = (t_k + t_{k+1})/2$, $k = 1, \dots, n-1$. The kernel K^* is defined by (denoting $v = (t-u)/h$):

$$K^*(v) = \begin{cases} K_+(v, q), & \text{for } 0 \leq t \leq h \text{ with } q = \frac{t}{h}, \\ K(v), & \text{for } h \leq t \leq 1-h, \\ K_-(v, q), & \text{for } 1-h \leq t \leq 1 \text{ with } q = \frac{1-t}{h}, \end{cases}$$

where, $K_+(\cdot, q)$ and $K_-(\cdot, q)$ are kernel functions of order d with supports $[-1, q]$ and $[-q, 1]$ respectively, K is a kernel function of order d with support $[-1, 1]$ and $h = h(n)$ is a sequence of bandwidths which is required to satisfy $h \rightarrow 0, nh \rightarrow \infty$ as $n \rightarrow \infty$.

Particularly, in [1] it is proved that under certain smoothness conditions on the kernels $K, K_{\pm}(\cdot, q)$ the integrated square error of the estimator \hat{g}^* computed with Müller's [2] estimators $\hat{\tau}$ and $\hat{\Delta}$ of τ and Δ respectively is

$$\int_0^1 |\hat{g}^*(t) - g(t)|^2 dt = O_P(n^{-2d/(2d+1)}). \tag{3}$$

The main aim of this paper is to estimate the Skorochod's distance $d_S(\hat{g}^*, g)$. Let us recall that

$$d_S(x, y) = \inf_{\lambda \in \Lambda} \left\{ \sup_{0 \leq t \leq 1} |x(t) - y(\lambda(t))| + \sup_{0 \leq t \leq 1} |t - \lambda(t)| \right\},$$

for $x, y \in D[0, 1]$, where Λ is a class of continuous monotonically increasing functions such that $\lambda(0) = 0$ and $\lambda(1) = 1$ (see [3] for more information). We do not assume that the jump point τ or its size Δ are estimated in some particular way.

Theorem 1. *Assume that the kernels $K, K_+(\cdot, q)$ and $K_-(\cdot, q)$ are continuously differentiable. Assume that the function $f(t), t \in [0, 1]$ satisfies Hölder condition with exponent one and assume $E|\xi_1|^p < \infty$ for some $p > 2$. Then for the estimator \hat{g}^* given by (2) it holds*

$$Ed_S(\hat{g}^*, g) \leq E|\hat{\tau} - \tau| + c_1 E|\hat{\Delta} - \Delta| + c_2 \left(h + \frac{1}{h\sqrt{n}} \right) + c_3 h^{-1} E(\hat{\Delta}|\tau - \hat{\tau}|),$$

where $c_1, c_2, c_3 > 0$ are constants depending on kernels k and Hölder constant of the function f .

2. Proof of Theorem 1

For a $x \in D[0, 1]$ we denote

$$\|x\|_{\infty} = \sup_{0 \leq t \leq 1} |x(t)|.$$

By the definition of the metric d_S we have

$$Ed_S(\hat{g}^*, g) \leq E\|\hat{g}^* - g \circ \lambda\|_{\infty} + E \sup_{0 \leq t \leq 1} |t - \lambda(t)|$$

for any random function $\lambda \in \Lambda$. Fix the function $\lambda \in \Lambda$ which is obtained by linearly interpolating the points $(0, 0), (\hat{\tau}, \tau), (1, 1)$. Hence

$$\max_{t \in [0, 1]} |t - \lambda(t)| = |\hat{\tau} - \tau|. \tag{4}$$

Therefore we need to estimate

$$I = E\|\widehat{g^*} - g \circ \lambda\|_\infty \leq \|I_1\|_\infty + \|I_2\|_\infty + \|I_3\|_\infty + \|I_4\|_\infty,$$

where

$$\begin{aligned} I_1(t) &= \sum_{i=1}^n f(t_i) \int_{s_{i-1}}^{s_i} K_h^*(t-u)du - f(\lambda(t)), \\ I_2(t) &= \widehat{\Delta}1_{[\widehat{\tau},1]}(t) - \Delta 1_{[\tau,1]}(\lambda(t)), \\ I_3(t) &= \sum_{i=1}^n (\Delta 1_{[\tau,1]}(t_i) - \widehat{\Delta}1_{[\widehat{\tau},1]}(t_i)) \int_{s_{i-1}}^{s_i} K_h^*(t-u)du, \\ I_4(t) &= \sum_{j=1}^n \varepsilon_j \int_{s_{j-1}}^{s_j} K_h^*(t-u)du, \end{aligned}$$

and where $K_h^*(\cdot) = h^{-1}K^*(\cdot/h)$. By the definition of λ it follows that $t \geq \widehat{\tau}$ if and only if $\lambda(t) \geq \tau$. Hence

$$\|I_2\|_\infty = |\widehat{\Delta} - \Delta|. \tag{5}$$

To estimate $\|I_1\|_\infty$, note that

$$\sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_h^*(t-u)du = \int_0^1 K_h^*(t-u)du = 1.$$

So

$$I_1(t) = \sum_{i=1}^n [f(t_i) - f(t)] \int_{s_{i-1}}^{s_i} K_h^*(t-u)du + f(t) - f(\lambda(t)).$$

Evidently

$$|f(t) - f(\lambda(t))| \leq \|f\|_1 \sup_{0 \leq t \leq 1} |t - \lambda(t)| = \|f\|_1 |\widehat{\tau} - \tau|,$$

where $\|f\|_1 = \sup_{s \neq t} |f(s) - f(t)|/|t - s|$. Noting that $\int_{s_{i-1}}^{s_i} K_h^*(t-u)du \neq 0$ only if $[s_{i-1}, s_i] \cap A(t, h) \neq \emptyset$, where

$$A(t, h) = \begin{cases} [t-h, t+h], & \text{if } h \leq t \leq 1-h, \\ [t, t+h], & \text{if } 0 \leq t < h, \\ [t-h, t], & \text{if } 1-h < t \leq 1. \end{cases}$$

We have

$$\begin{aligned} & \sum_{i=1}^n [f(t_i) - f(t)] \int_{s_{i-1}}^{s_i} K_h^*(t-u) du \\ & \leq \|f\|_1 \sum_{i=1}^n |t_i - t| \int_{s_{i-1}}^{s_i} K_h^*(t-u) du \leq \|f\|_1 (h + 1/n). \end{aligned}$$

Hence

$$\|I_1\|_\infty \leq \|f\|_1 \left(|\hat{\tau} - \tau| + h + \frac{1}{n} \right). \tag{6}$$

Now we estimate I_3

$$\begin{aligned} I_3(t) &= \sum_{i=1}^n (\Delta 1_{[\tau,1]}(t_i) - \widehat{\Delta} 1_{[\hat{\tau},1]}(t_i)) \int_{s_{i-1}}^{s_i} K_h^*(t-u) du = \\ &= \sum_{i=1}^n (\Delta 1_{[\tau,1]}(t_i) - \widehat{\Delta} 1_{[\tau,1]}(t_i) + \widehat{\Delta} 1_{[\tau,1]}(t_i) - \widehat{\Delta} 1_{[\hat{\tau},1]}(t_i)) \int_{s_{i-1}}^{s_i} K_h^*(t-u) du \\ &= \sum_{i=1}^n (\Delta - \widehat{\Delta}) 1_{[\tau,1]}(t_i) + \widehat{\Delta} 1_{[\hat{\tau},\tau]}(t_i) \int_{s_{i-1}}^{s_i} K_h^*(t-u) du. \end{aligned}$$

We have

$$\begin{aligned} & \sum_{i=1}^n (\Delta - \widehat{\Delta}) 1_{[\tau,1]}(t_i) \int_{s_{i-1}}^{s_i} K_h^*(t-u) du \\ & \leq |\Delta - \widehat{\Delta}| \widehat{\Delta} \sum_{i=1}^n 1_{[\hat{\tau},\tau]}(t_i) \int_{s_{i-1}}^{s_i} K_h^*(t-u) du \leq \widehat{\Delta} \frac{|\tau - \hat{\tau}|}{h}. \end{aligned}$$

Hence

$$\|I_3\|_\infty \leq |\Delta - \widehat{\Delta}| + \widehat{\Delta} |\tau - \hat{\tau}| h^{-1}.$$

In order to estimate $E\|I_4\|_\infty$, we need some preparation.

Let $T = [0, 1)$ and let d be a pseudo metric on the set T . For $\tau > 0$, denote by $N(T, d, \tau)$ the minimal number of open d -balls of radius τ which form a covering of T . Denote further by $D = D(T)$ the d -diameter of T . The following Lemma is a corollary from Theorem 1.11 in Ledoux and Talagrand [4].

Lemma 2. Let $X = (X_t)_{t \in [0,1]}$ be a random process in L_p such that for all $s, t \in [0, 1]$

$$E|X_s - X_t|^p \leq d^p(s, t),$$

where d is a pseudo metric on the set $[0, 1]$. Then, if

$$\int_0^D N^{1/p}(T, d, \tau) d\tau < \infty,$$

we have

$$E \sup_{s, t \in T} |X_s - X_t| \leq 8 \int_0^D N^{1/p}(T, d, \tau) d\tau.$$

Set

$$k_i(t) = \int_{s_{i-1}}^{s_i} K^* \left(\frac{t-u}{h} \right) du, \quad i = 1, \dots, n; \quad t \in [0, 1].$$

Let $p > 2$. By Rosenthal's inequality we have for all $s, t \in [0, 1]$

$$\begin{aligned} E|I_4(t) - I_4(s)|^p &\leq h^{-p} E \left| \sum_{i=1}^n \epsilon_i (k_i(t) - k_i(s)) \right|^p \\ &\leq ch^{-p} \left(\sum_{i=1}^n (k_i(t) - k_i(s))^2 \right)^{p/2} + ch^{-p} E|\xi_1|^p \sum_{k=1}^n |k_i(t) - k_i(s)|^p \\ &\leq ch^{-p} (1 + E|\xi_1|^p) \left(\sum_{i=1}^n (k_i(t) - k_i(s))^2 \right)^{p/2}, \end{aligned}$$

where the constant $c > 0$ depends on p only. Observing that

$$|k_i(t) - k_i(s)| \leq \|K^*\|_1 \frac{|t-s|}{hn},$$

and

$$\sum_{i=1}^n k_i(t) = h,$$

we obtain

$$E|I_4(t) - I_4(s)|^p \leq c(1 + E|\xi_1|^p) |t-s|^{p/2} h^{-p} n^{-p/2}.$$

By Lemma 2

$$E\|I_4\|_\infty \leq Ch^{-1}n^{-1/2}.$$

Collecting estimates we complete the proof.

References

- [1] Kee-Hoon Kang, Ja-Yong Koo, Cheol-Woo Park, Kernel estimation of discontinuous regression functions, *Stat. Probab. Letters*, **47**, 277–285 (2001).
- [2] H.G. Müller, Change-point in nonparametric regression analysis, *Ann. Statist.*, **20**, 737–761 (1992).
- [3] I.I. Gichman, A.V. Skorochod, *Theory of Stochastic Processes*, I, Nauka, Moscow (1971).
- [4] M. Ledoux, M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, Heidelberg (1991).

Apie trūkios regresijos funkcijos branduolinio įvertinimo tolygią paklaidą

A. Račkauskas, A. Tamulis

Gautas Skorochoodo erdvei atstumo tarp regresijos funkcijos, turinčios trūkio tašką, ir jos branduolinio įverčio įvertinimas.