Statistics and Probability Letters 123 (2017) 17–22

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view[☆]

Kairat Mynbaev [a], Carlos Martins-Filho [b,c,*]

[a] International School of Economics, Kazakh-British Technical University, Tolebi 59, Almaty 050000, Kazakhstan
[b] Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA
[c] IFPRI, 2033 K Street NW, Washington, DC 20006-1002, USA

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form 19 November 2016
Accepted 20 November 2016
Available online 5 December 2016

MSC:
62G07
62G10
62G20

Keywords:
Kernel density estimation
Higher order kernels
Bias reduction

ABSTRACT

We define a new bandwidth-dependent kernel density estimator that improves existing convergence rates for the bias, and preserves that of the variation, when the error is measured in L_1. No additional assumptions are imposed to the extant literature.

1. Introduction

Given a sequence of $n \in \mathbb{N}$ independent realizations $\{X_j\}_{j=1}^n$ of the random variable X, having density f on $\mathbb{R}$, the Rosenblatt–Parzen kernel estimator (Rosenblatt, 1956; Parzen, 1962) of f is given by

$$f_n(x) = \frac{1}{n}\sum_{j=1}^{n}(S_{h_n}K)(x - X_j), \quad (1.1)$$

where S_{h_n} is an operator defined by

$$(S_{h_n}K)(x) = \frac{1}{h_n}K\left(\frac{x}{h_n}\right), \quad (1.2)$$

K is a kernel, i.e., a function on $\mathbb{R}$ such that $\int K(x)dx = 1$ and $h_n > 0$ is a non-stochastic bandwidth such that $h_n \to 0$ as $n \to \infty$.[1]

☆ We thank an anonymous referee and an Associate Editor for excellent comments that improved this note significantly.
* Corresponding author at: Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA.
E-mail addresses: kairat_mynbayev@yahoo.com (K. Mynbaev), carlos.martins@colorado.edu, c.martins-filho@cgiar.org (C. Martins-Filho).
1 Throughout this note, integrals are over $\mathbb{R}$, unless otherwise specified.

http://dx.doi.org/10.1016/j.spl.2016.11.019
0167-7152/

One of the most natural and mathematically sound (Devroye and Györfi, 1985; Devroye, 1987) criteria to measure the performance of f_n as an estimator of f is the L_1 distance $\int |f_n - f|$. In particular, given that this distance is a random variable (measurable function of $\{X_j\}_{j=1}^n$) it is convenient to focus on $E\left(\int |f_n - f|\right)$, where E denotes the expectation taken using f. For this criterion, there is a simple bound (Devroye, 1987, p. 31)

$$E\left(\int |f_n - f|\right) \le \int |(f * S_{h_n}K) - f| + E\left(\int |f_n - f * S_{h_n}K|\right),$$

where for arbitrary $f, g \in L_1$, $(f*g)(x) = \int g(y)f(x-y)dy$ is the convolution of f and g. The term $\int |f * S_{h_n}K - f|$ is called bias over $\mathbb{R}$ and $E\left(\int |f_n - f * S_{h_n}K|\right)$ is called the variation over $\mathbb{R}$. There exists a large literature devoted to establishing conditions on f and K that assure suitable rates of convergence of the bias to zero as $n \to \infty$ (see, *inter alia*, Silverman, 1986; Devroye, 1987; Tsybakov, 2009). In particular, if K is of order s, i.e., $\alpha_j(K) = 0$ for $j = 1, \ldots, s-1$ and $\alpha_s(K) \neq 0$, where $\alpha_j(K) = \int t^j K(t)dt$ is the jth moment of K, and f has an integrable derivative $f^{(s)}$, then $\int |f * S_{h_n}K - f|$ is of order $O(h_n^s)$ and this order cannot be improved, see, e.g., Devroye (1987, Theorem 7.2). In this note, we show that if in (1.2) the kernel is allowed to depend on n, then the order $O(h_n^s)$ can be replaced by the order $o(h_n^s)$, without increasing the order of the kernel or the smoothness of the density. In addition, another result from Devroye (1987) states that if K is a kernel of order greater than s and the derivative $f^{(s)}$ is a-Lipschitz then the bias is of order $O(h_n^{s+a})$. We achieve the same rate of convergence with kernels of order s.

2. Main results

Let L_1 and C denote the spaces of integrable and (bounded) continuous functions on $\mathbb{R}$ with norms $\|f\|_1 = \int |f|$ and $\|f\|_C = \sup |f|$, and $\beta_s(K) = \int |t|^s |K(t)|\, dt$. Let $\{K_n\}$ be a sequence of kernels and define

$$\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n}(S_{h_n}K_n)(x - X_j).$$

In the following Theorem 1, the density f has the same degree of smoothness and the kernels K_n are of the same order as in Devroye (1987, Theorem 7.2), but the bias is of order $o(h_n^s)$ instead of $O(h_n^s)$. This results because the kernels depend on n and have "disappearing" moments of order s.

Theorem 1. *Let $\{K_n\}$ be a sequence of kernels of order s such that: 1. $\alpha_s(K_n) \to 0$; 2. $\{u^s K_n(u)\}$ is uniformly integrable. For all f with absolutely continuous $f^{(s-1)}$ and $f^{(s)} \in L_1$, we have $\|f * S_{h_n}K_n - f\|_1 = o(h_n^s)$.*

Proof. Note that since K_n is a kernel

$$f * S_{h_n}K_n(x) - f(x) = \int K_n(t)[f(x - h_n t) - f(x)]dt. \tag{2.1}$$

Since f is s-times differentiable, by Taylor's Theorem,

$$f(x - h_n t) - f(x) = \sum_{j=1}^{s-1}\frac{f^{(j)}(x)}{j!}(-h_n t)^j + \int_x^{x-h_n t}\frac{(x - h_n t - u)^{s-1}}{(s-1)!}f^{(s)}(u)du.$$

Furthermore, given that K_n is of order s,

$$f * S_{h_n}K_n(x) - f(x) = \frac{1}{(s-1)!}\int\int_x^{x-h_n t}(x - h_n t - u)^{s-1}f^{(s)}(u)duK_n(t)dt. \tag{2.2}$$

Letting $\lambda = -\frac{u-x}{h_n t}$ we have

$$\int_x^{x-h_n t}(x - h_n t - u)^{s-1}f^{(s)}(u)du = (-h_n t)^s\int_0^1 f^{(s)}(x - h_n\lambda t)(1-\lambda)^{s-1}d\lambda. \tag{2.3}$$

Substituting (2.3) into (2.2) we obtain

$$f * S_{h_n}K_n(x) - f(x) = \frac{(-h_n)^s}{s!}\int\int_0^1 f^{(s)}(x - h_n\lambda t)s(1-\lambda)^{s-1}d\lambda t^s K_n(t)dt. \tag{2.4}$$

Since $\int_0^1(1-\lambda)^{s-1}d\lambda = \frac{1}{s}$, we have that

$$\frac{(-h_n)^s}{(s-1)!}\int\int_0^1 f^{(s)}(x)(1-\lambda)^{s-1}d\lambda t^s K_n(t)dt = \frac{(-h_n)^s}{s!}f^{(s)}(x)\int t^s K_n(t)dt. \tag{2.5}$$