Statistics and Probability Letters 94 (2014) 106–112

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Improving bias in kernel density estimation

Kairat T. Mynbaev[a,*], Saralees Nadarajah[b], Christopher S. Withers[c], Aziza S. Aipenova[d]

[a] *Kazakh-British Technical University, Almaty 050000, Kazakhstan*
[b] *School of Mathematics, University of Manchester, Manchester M13 9PL, UK*
[c] *Applied Mathematics Group, Industrial Research Limited, Lower Hutt, New Zealand*
[d] *Kazakh National University, Almaty, Kazakhstan*

ARTICLE INFO

Article history:
Received 26 January 2014
Received in revised form 11 July 2014
Accepted 11 July 2014
Available online 16 July 2014

MSC:
62G07
62G10
62G20

Keywords:
Density estimation
Bias
Higher order kernel

ABSTRACT

For order q kernel density estimators we show that the constant b_q in *bias* $= b_q h^q + o(h^q)$ can be made arbitrarily small, while keeping the variance bounded. A data-based selection of b_q is presented and Monte Carlo simulations illustrate the advantages of the method.

1. Introduction

Let f denote a density, K an integrable function on $\mathbb{R}$ such that $\int K dt = 1$ and let $X_1, \ldots, X_n$ be i.i.d. random variables with density f. Consider the kernel estimator of $f(x)$

$$f_h(x, K) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h}K\left(\frac{x - X_j}{h}\right), \quad h > 0. \tag{1}$$

Denote $\alpha_i(K) = \int x^i K(x)dx$ the ith moment of K and let K be a kernel of order q, that is $\alpha_j(K) = 0,\ j = 1, \ldots, q-1,\ \alpha_q(K) \neq 0$. It is well-known that the bias is proportional to $\alpha_q(K)h^q$ if f is q-smooth in some sense (Devroye, 1987; Scott, 1992; Silverman, 1986; Wand and Jones, 1995).

The usual approach is to stick to some K and be content with the resulting $\alpha_q(K)$. The purpose of this paper is to show that it pays to reduce $\alpha_q(K)$ by choosing a suitable K. Despite the bias being proportional to $\alpha_q(K)h^q$, the benefits of the suggested approach are not obvious because as the qth moment is made smaller, the variance of the estimator may go up. Our construction of K allows us to control the variance. Our results imply that among all kernels of order q with uniformly

* Corresponding author. Tel.: +7 701 716 5134.
E-mail addresses: kairat_mynbayev@yahoo.com (K.T. Mynbaev), Saralees.nadarajah@manchester.ac.uk (S. Nadarajah), kit.withers@gmail.com (C.S. Withers), a.aipenova@mail.ru (A.S. Aipenova).

http://dx.doi.org/10.1016/j.spl.2014.07.014
0167-7152/

bounded variances there is no kernel with the least nonzero $|\alpha_q(K)|$. The issue of selecting the kernel order does not arise in the approach suggested in Mynbaev and Martins Filho (2010).

In case of L_1 convergence the main idea can be illustrated using the corresponding bias notion from Devroye (1987). Let bias be defined as $\int|f\star K_h-f|dt$ where $K_h(x)=K(x/h)/h$. If K is of order q, f has $q-1$ absolutely continuous derivatives and an integrable derivative $f^{(q)}$, then by Devroye (1987, Theorem 7.2)

$$q!\int|f\star K_h-f|dt\Big/\left(h^q\int|f^{(q)}|dt\right)\to\alpha_q(K),\quad h\downarrow 0.$$

Here $\alpha_q(K)$ can be made as small as desired using our Theorem 2.

We call a free-lunch effect the fact that $\alpha_q(K)$ can be made as small as desired, without increasing the density smoothness or the kernel order. Of course, in finite samples bias cannot be eliminated completely. Put it differently, for very small $\alpha_q(K)$ sample variance starts to dominate the effect of small bias.

For simplicity, in our main results in Section 2 we consider only classical smoothness characteristics. The simulation results in Section 3 compare our kernel performance with that of three well-known kernel families. The overall conclusion is that a better estimation performance is not necessarily a consequence of some optimization criterion and can be achieved by directly targeting the bias of the estimator. All proofs are provided in Appendix A, while Appendix B contains the code used for simulations (both are supplied as supplemental material, see Appendix A).

2. Main results

Multiplication by polynomials (Deheuvels, 1977; Wand and Schucany, 1990) is one of several ways to construct higher-order kernels. Withers and Nadarajah (2013) have explored the procedure of transforming a kernel K into a higher-order kernel $T_{\boldsymbol{a}}K$ via multiplication of K by a polynomial of order q, $(T_{\boldsymbol{a}}K)(t)=\left(\sum_{i=0}^q a_it^i\right)K(t)$, with a suitably chosen vector of coefficients $\boldsymbol{a}=(a_0,\ldots,a_q)'\in\mathbb{R}^{q+1}$. Unlike several authors who chose the polynomial subject to some optimization criterion (see Berlinet, 1993, Fan and Hu, 1992, Gasser and Muller, 1979, Lejeune and Sarda, 1992 and Wand and Schucany, 1990) Withers and Nadarajah with their definition of the polynomial directly targeted moments of the resulting kernel. In their Theorem 2.1, they defined a polynomial transformation in such a way that the moments of the new kernel numbered 1 through $q-1$ are zero. They did not notice that the qth moment can be targeted in the same way and can be made as small as desired and that the variance of the resulting estimator retains the order $1/(nh)$ as the qth moment is manipulated. This is what we do here. Besides, we show that not only variance but all the higher-order terms in h in the Taylor decomposition of the bias and variance can be controlled not to increase.

We do this under two sets of assumptions. The first set is that the density is infinitely differentiable and all moments of K exist and the second is that the density has a finite number of derivatives and the kernel and its square possess a finite number of moments. We give complete proofs for the first set, because part of the argument is new and it can be extended to justify some formal infinite decompositions from Withers and Nadarajah (2013). The proof for the second set goes more along traditional lines (except for controlling higher-order terms) and is therefore omitted.

Let $\beta_j(K)=\int_{\mathbb{R}}|K(t)t^j|dt$ denote the jth absolute moment of K. The estimator of $f^{(l)}(x)$ is obtained by differentiating both sides of (1) l times.

Theorem 1. *Suppose that f is infinitely differentiable and K has a continuous derivative of order l. Further assume that K and $K^{(l)}$ have absolute moments of all orders,*

$$\limsup_{j\to\infty}\left|\frac{f^{(j)}(x)}{j!}\max\left\{\beta_{j+1}(K),\beta_{j+1}(K^{(l)})\right\}\right|^{1/j}=0,\tag{2}$$

$$\left\|K^{(l)}\right\|_{C(\mathbb{R})}=\sup_{t\in\mathbb{R}}\left|K^{(l)}(t)\right|<\infty.\tag{3}$$

Then

$$Ef_h^{(l)}(x,K)=\sum_{i=0}^{\infty}\frac{f^{(i+l)}(x)}{i!}(-h)^i\alpha_i(K),\tag{4}$$

$$\operatorname{var}\left(f_h^{(l)}(x,K)\right)=\frac{1}{nh^{2l+1}}\left\{\sum_{i=0}^{\infty}\frac{f^{(i)}(x)\alpha_i(M)}{i!}(-h)^i-h\left[h^lEf_h^{(l)}(x,K)\right]^2\right\}\tag{5}$$

where $M=\left[K^{(l)}\right]^2$ and the series converge for all $h\in\mathbb{R}$. Consequently, if K is a kernel of order q, then

$$Ef_h^{(l)}(x,K)-f^{(l)}(x)=\frac{f^{(q+l)}(x)}{q!}(-h)^q\alpha_q(K)+O(h^{q+1}),\tag{6}$$

$$\operatorname{var}\left(f_h^{(l)}(x,K)\right)=\frac{1}{nh^{2l+1}}\left\{f(x)\int_{\mathbb{R}}M(t)dt+O(h)\right\}.\tag{7}$$