Improving bias in nonparametric density estimation: L_1 view

Kairat Mynbaev*
Department of Economics, University of Colorado, Boulder, Colorado, USA - kairat_mynbayev@yahoo.com

Carlos Martins Filho
Department of Economics, University of Colorado, Boulder, Colorado, USA & IFPRI, Washington, DC, USA - Carlos.Martins@colorado.edu

Abstract

Density estimation is important for many applications, including densities of returns on financial assets. Researchers are concerned with the precision of estimation (bias) while keeping variability of estimators low. There are well-established results regarding the order of bias, when the density is sufficiently smooth and the kernel used for estimation is of a corresponding order. In this paper we show that two of these results can be improved further. We define a new kernel density estimator and show that it improves the existing convergence estimates when the error is measured in L_1. In one result, the bias is shown to be $o(h^s)$ instead of the conventional $O(h^s)$, assuming that the density is s-smooth and the kernel is of order s. In the other result, for a density whose sth order derivative is α-Lipschitz, we prove that the bias is of order $h^{s+\alpha}$, provided that the kernel order is s, while the extant papers require "larger than s".

Keywords: Rosenblatt-Parzen estimator; convergence; higher order kernel.

1. Introduction

In this paper we explore the benefits of using kernels with a disappearing higher-order moment.
Let f be a density on R. A kernel estimate of $f(x)$ is defined by $f_n(x) = \frac{1}{n}\sum_{j=1}^{n}(S_hK)(x - X_j)$. Here K is a kernel, that is, a function on R that integrates to 1, $X_1, ..., X_n$ are independent random variables with density f, and the operator S_h is defined by

$$(S_hK)(x) = \frac{1}{h}K\left(\frac{1}{h}\right), \tag{1}$$

where the parameter $h > 0$ is a bandwidth.
One of the most natural ways to measure the performance of f_n in estimating f is by using the L_1 distance $E\int|f_n - f|$, see Devroye, L., & Györfi, L. (1985). For this distance there is a simple bound (see Devroye, 1987)

$$E\int|f_n - f| \le \int|f * S_hK - f| + E\int|f_n - f * S_hK|.$$

The term $\int|f * S_hK - f|$ is called bias and $\int|f_n - f * S_hK|$ is called variation. Denote

$$\alpha_s(K) = \int_R t^sK(t)dt,\ \beta_s(K) = \int_R|t^sK(t)|\,dt$$

the sth moment and absolute moment of K, resp. K is called a kernel of order s if $\alpha_0(K) = 1$, $\alpha_j(K) = 0$ for $j = 1, ..., s-1$, and $\alpha_s(K) \ne 0$. It is an established truth that if K is a kernel of order s and f has an integrable derivative $f^{(s)}$, then $\int|f * S_hK - f|$ is of order $O(h^s)$, and this order cannot be improved, see, e.g., Devroye (1987), Theorem 7.2. We show that if in (1) the kernel is allowed to depend on the bandwidth, then the $O(h^s)$ can be replaced by $o(h^s)$, without increasing the kernel order or density smoothness requirement. Another result from Devroye (1987) states that if K is a kernel of order $> s$ and the derivative $f^{(s)}$ is Lipschitz α, then the bias is of order $O(h^{s+\alpha})$. We achieve the same rate of convergence with kernels of order s.

2. Main results

L_1 and C denote the spaces of integrable and continuous functions on R provided with the norms $\|f\|_1 = \int |f|$ and $\|f\|_C = \sup |f|$, respectively. Let $\mathcal{B}_s$ denote the space of functions with the norm $\|K\|_{\mathcal{B}_s} = \beta_0(K) + \beta_s(K)$. Take functions $K^{(0)}, K^{(s)} \in \mathcal{B}_s$ such that

$$\begin{aligned} \alpha_0(K^{(0)}) &= 1,\ \alpha_j(K^{(0)}) = 0 \text{ for } j = 1, ..., s, \\ \alpha_j(K^{(s)}) &= 0 \text{ for } 0 = 1, ..., s-1,\ \alpha_s(K^{(s)}) = 1. \end{aligned} \tag{2}$$

The kernel $K_h = K^{(0)} + hK^{(s)}$, $0 \le h \le 1$, has a disappearing moment $\alpha_s(K_h) = h$ of order s. Any kernel of order s can be decomposed as $K = K^{(0)} + \alpha_s(K^{(s)})K^{(s)}$, so the conventional kernels obtain from ours with $h = \alpha_s(K^{(s)})$. In the next theorem the density has the same degree of smoothness and the kernel is of the same order as in Devroye(1987), Theorem 7.2 but the bias is of order $o(h^s)$ instead of $O(h^s)$ because we put to good use the "residual smoothness" of $f^{(s)}$ entailed by the membership $f^{(s)} \in L_1$.
Theorem 1. Let K_h be the kernel defined above. For all f with absolutely continuous $f^{(s-1)}$ and summable $f^{(s)}$ one has $\int |f * S_h K_h - f| = o(h^s)$.
Remark. By Young's inequality

$$\begin{aligned} E\int |f_n - f * S_h K_h| &\le E\int |f_n - f * S_h K^{(0)}| + h\int |f * S_h K^{(s)}| \\ &\le E\int |f_n - f * S_h K^{(0)}| + h\int |f| \int |K^{(s)}|. \end{aligned}$$

Therefore the expected value of variation of the proposed estimator asymptotically is not larger than $E\int |f_n - f * S_h K^{(0)}|$.
Next we provide an analog of the result from Devroye (1987) for densities with derivatives satisfying the Lipschitz condition. While it is possible to consider densities with compact support, as in Devroye (1987), we prefer to avoid this assumption because there is an issue of properly generalizing the Lipschitz condition for unbounded domains. We say that a function g defined on R satisfies a *global Lipschitz condition* of order $\alpha \in (0,1)$ if there exist positive functions $l(x), r(x)$ such that

$$|g(x-h) - g(x)| \le l(x)|h|^\alpha \text{ for } |h| \le r(x),\ x \in R. \tag{3}$$

The function l is called a Lipschitz constant and the function r is called a Lipschitz radius. Typically, for good densities $l(x) \to 0$, $r(x) \to \infty$ as $|x| \to \infty$, see Mynbaev, K., & Martins-Filho, C. (2010). The class $W^{s+\alpha}(l, r)$ is defined as the set of functions with an absolutely continuous derivative $f^{(s-1)}$ such that $g = f^{(s)}$ satisfies the global Lipschitz condition (3). In Devroye (1987), Theorem 7.1 the bias order $O(h^{s+\alpha})$ is achieved for kernel orders $> s$, while we in the following theorem obtain the same order of bias for kernel order s.
Theorem 2. Suppose the density f belongs to $W^{s+\alpha}(l, r)$ where

$$\int_R [l(x) + r^{-\alpha}(x)]dx < \infty$$

and the derivative $f^{(s)}$ belongs to $L_1 \cap C$. Let $K^{(0)}, K^{(s)}$ satisfy (2), $K^{(0)} \in \mathcal{B}_{s+\alpha}$, $K^{(s)} \in \mathcal{B}_s$. Put $K_h = K^{(0)} + h^\alpha K^{(s)}$, $0 \le h \le 1$. Then K_h is a kernel of order s for $h > 0$ and $\int |f * S_h K_h - f| = O(h^{s+\alpha})$.
Remark. The condition $K^{(0)} \in \mathcal{B}_{s+\alpha}$ and the definition $K_h = K^{(0)} + h^\alpha K^{(s)}$ can be replaced by $K^{(0)} \in \mathcal{B}_{s+1}$ and $K_h = K^{(0)} + hK^{(s)}$, resp., without affecting the conclusion.

3. Conclusions

To have a certain order of bias, the existing results are valid under two essential conditions: the kernel order and the density smoothness should be the same. Under these conditions, the bias order is known to be precise (cannot be improved). However, so far the kernel has been fixed. We have shown that if one considers a family of kernels, then the order of bias can be reduced, even though all kernels in the family are of the same order. The estimation procedure is as simple as the classical Rosenblatt-Parzen estimator.

References

Devroye, L., & Györfi, L. (1985). Nonparametric density estimation: The L_1 view. Wiley & Sons, New York.

Devroye, L. (1987). A course in density estimation. Birkhäuser, Boston, Boston, MA.

Mynbaev, K., & Martins-Filho, C. (2010). Bias reduction in kernel density estimation via Lipschitz conditions. J. Nonparametr. Stat. 22, 219–235.

Bretagnolle, J., & Huber, C. (1979). Estimation des densites: risque minimax, Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 47, 119–137.