Statistics and Probability Letters 81 (2011) 571–579

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Kernel adjusted density estimation

Ramidha Srihera [a], Winfried Stute [b,*]

[a] Department of Mathematics and Statistics, Thammasat University, Rangsit Center, Faculty of Science and Technology, Pathum Thani 12121, Thailand

[b] Mathematical Institute, University of Giessen, Arndtstr. 2, D-35392 Giessen, Germany

ARTICLE INFO

Article history:
Received 30 November 2010
Received in revised form 18 January 2011
Accepted 18 January 2011
Available online 26 January 2011

Keywords:
Kernel density estimator
Adaptive choice

ABSTRACT

We propose and study a kernel estimator of a density in which the kernel is adapted to the data but not fixed. The smoothing procedure is followed by a location-scale transformation to reduce bias and variance. The new method naturally leads to an adaptive choice of the smoothing parameters which avoids asymptotic expansions.

1. Introduction and main results

Since Rosenblatt (1956) and Parzen (1962) introduced the kernel estimator of an unknown density f, there have been numerous authors who studied various of its finite and large sample properties. To be more precise, let K be a given function on the real line, the "kernel", and let $h > 0$ be a given bandwidth or window size. Then, if $X_1, \ldots, X_n$ denotes an independent sample from f, the associated kernel estimator is defined as

$$f_n(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right).$$

To obtain a "bona fide" estimator, i.e., one which is itself a density, one requires

$$K \geq 0 \quad \text{and} \quad \int K(u)\mathrm{d}u = 1.$$

Silverman (1986) and Wand and Jones (1995) became standard reference books on kernel methodology. To cite only one of the many properties of $f_n(x)$, recall that for the mean square error (MSE), we have, when $\int uK(u)\mathrm{d}u = 0$ and f is twice continuously differentiable in a neighborhood of x, that

$$\begin{aligned} \mathrm{Bias} f_n(x) &:= \mathbb{E} f_n(x) - f(x) \\ &= \frac{1}{2} f''(x) h^2 \int u^2 K(u)\mathrm{d}u + O(h^3) \end{aligned}$$

and

$$\mathrm{Var}\, f_n(x) = \frac{1}{nh} f(x) \int K^2(u)\mathrm{d}u + o\left(\frac{1}{nh}\right)$$

* Corresponding author. Tel.: +49 641 99 32100; fax: +49 641 99 32029.
E-mail addresses: ramidha@mathstat.sci.tu.ac.th (R. Srihera), Winfried.Stute@math.uni-giessen.de (W. Stute).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2011.01.013

whenever $n \to \infty$ and $h \to 0$ such that $nh \to \infty$. This implies that

$$\begin{aligned} \mathrm{MSE} f_n(x) &= \mathrm{Bias}^2 f_n(x) + \mathrm{Var} f_n(x) \\ &\sim \frac{1}{4}(f''(x))^2 h^4 \left[\int u^2 K(u)\mathrm{d}u\right]^2 + \frac{1}{nh} f(x) \int K^2(u)\mathrm{d}u. \end{aligned} \tag{1.1}$$

The optimal choice of h minimizing the last expression satisfies

$$h_{\mathrm{opt}}^5 \sim \frac{1}{n} \frac{f(x)\int K^2(u)\mathrm{d}u}{\left[f''(x)\int u^2 K(u)\mathrm{d}u\right]^2}. \tag{1.2}$$

If, rather than $\mathrm{MSE} f_n(x)$ at a fixed x, one considers the integrated MSE as a measure of fit, i.e.,

$$\mathrm{MISE} = \int \mathrm{MSE} f_n(x)\mathrm{d}x,$$

then the optimal h satisfies, up to remainders,

$$h_{\mathrm{opt}}^5 = \frac{1}{n} \frac{\int K^2(u)\mathrm{d}u}{\int [f''(x)]^2 \mathrm{d}x \left[\int u^2 K(u)\mathrm{d}u\right]^2}. \tag{1.3}$$

It is known, see Silverman (1986), that the choice of K has little effect on MSE and MISE. Rather, the unknown $f(x)$ and $f''(x)$ are crucial and prevent one from a straightforward application of (1.2) or (1.3). One possibility is to choose a preliminary h^1, estimate $f(x)$ and $f''(x)$ and then compute an adapted version of h_{opt}. Another strategy is to determine h in a fully adaptive way by minimizing a cross-validated deviation between f_n and f. Finally, a third popular method consists in referring $\int [f''(x)]^2 \mathrm{d}x$ in (1.3) to a standard distribution, i.e., to compute the integral for a parametric family of centered densities with scale parameter σ, and then to apply (1.3) with an estimated σ. Silverman (1986) pointed out that this method may lead to incorrect results when the reference densities are symmetric at zero, but the true but unknown f is multimodal and thus typically has larger curvature relative to scale. Also, the first method is not fully satisfactory since it requires the subjective choice of a preliminary h^1. Finally, the cross-validated h is known to be asymptotically optimal but may show a poor behavior when sample size is small or moderate. See Feluch and Koronacki (1992).

It is the purpose of this paper to propose and study a fully adaptive approach which takes into account a modified version of the third method, in which the reference densities are associated with the true f. In other words, we shall consider the location-scale family generated by the true f. Interestingly enough, to deal with bias issues, it will not be necessary to incorporate estimators of f'' based on preliminary choices of h. Also, we shall be able to get estimates of MSE and MISE and hence adaptive choices of the smoothing parameters.

To begin with, let K_0 be a kernel from the location-scale family associated with f, i.e.,

$$K_0(u) = K_0(u, \theta, \sigma) = \sigma f(\sigma u + \theta). \tag{1.4}$$

For (1.1), with $\theta = \mathbb{E}X$ and $\sigma = 1$, we then get, e.g.,

$$\mathrm{MSE} f_n(x) \sim \frac{1}{4}(f''(x))^2 h^4 \mathrm{Var}^2 X + \frac{1}{nh} f(x) \int f^2(u)\mathrm{d}u. \tag{1.5}$$

The interesting point about (1.5) is that the bias and variance parts contain terms which reflect both the local and global behavior of f, namely $f''(x), f(x)$ and, respectively, $\mathrm{Var}\, X$ and $\int f^2$. Similarly for MISE.

For example, since typically $\mathrm{Var}\, X$ is small when $f''(x)$ is large, (1.5) demonstrates that rather than choosing a fixed K, a properly chosen kernel from (1.4) may decrease the bias. The scaling factor σ gives us more flexibility. As will be seen later this will enable us to choose K_0 so as to minimize MSE. Of course, since f in (1.4) is not available, we have to replace it by f_n from above. Hence our estimator becomes

$$\begin{aligned} \hat{f}_n(x) \equiv \hat{f}_n(x, \theta, \sigma) &= \frac{\sigma}{nh} \sum_{i=1}^n f_n\left(\sigma \frac{x - X_i}{h} + \theta\right) \\ &= \frac{\sigma}{n^2 h^2} \sum_{j=1}^n \sum_{i=1}^n K\left(\frac{\sigma x - \sigma X_i + \theta h - h X_j}{h^2}\right). \end{aligned}$$

In other words, the $\hat{f}_n$ constitute kernel estimators with the kernels taken from the location-scale family associated with a classical kernel estimator. The choice of h, σ and θ will be discussed later.

To reduce a possible bias, our final estimator will be

$$\hat{f}_n(x) = \frac{\sigma}{n(n-1)h^2} \sum_{i \neq j} K\left(\frac{\sigma x - \sigma X_i + \theta h - h X_j}{h^2}\right). \tag{1.6}$$