STAPRO: 8189 Model 3G pp. 1–7 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On the utility of asymptotic bandwidth selectors for spatially adaptive kernel density estimation

Tilman M. Davies [a,*], Claire R. Flynn [b,a], Martin L. Hazelton [c]

[a] Department of Mathematics & Statistics, University of Otago, Dunedin, New Zealand
[b] Metservice, Wellington, New Zealand
[c] Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

ARTICLE INFO

Article history:
Received 29 August 2017
Received in revised form 13 January 2018
Accepted 26 February 2018
Available online xxxx

Keywords:
Global bandwidth
AMISE
Spatial point process
Kernel smoothing
Normal scale rule

ABSTRACT

Implementation of the spatially adaptive kernel estimator relies on choice of a 'global bandwidth'. We derive the closed-form asymptotic bias for this estimator with the aim of developing relevant selectors, and note non-uniform convergence hinders its usability for this purpose.

1. Introduction

Estimation of the density or intensity function from planar point process data is arguably the most fundamental inferential problem in spatial statistics. By far the most common approach in practice is to apply kernel smoothing. The majority of the literature has focused on kernel density (and intensity) estimation using a spatially uniform amount of smoothing controlled by a fixed bandwidth. However, spatial datasets frequently exhibit pronounced inhomogeneity. For example, data in geographical epidemiology typically combine tight clusters of points identifying disease cases in towns and cities with large sparsely populated regions. This leads to huge spatial variation in smoothing requirements, with less wanted in urban areas to retain detail, and much more required elsewhere to avoid stochastic artefacts (e.g. Davies and Hazelton, 2010).

In response, spatially adaptive kernel density estimation has become increasingly common for planar data. A popular version of this type of methodology is the sample point adaptive density estimator, defined by

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{h_i}(\boldsymbol{x} - \boldsymbol{x}_i) \quad (1)$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are the bivariate coordinates of n independent, identically distributed observations, and $K_h(\boldsymbol{x}) = h^{-2}K(\boldsymbol{x}/h)$ is a scaled kernel function defined in terms of an unscaled kernel K which is a spherically symmetric probability density function. The scaling factor h is referred to as the *bandwidth*, and determines the degree of smoothing. In (1) this is allowed

* Corresponding author.
E-mail address: tdavies@maths.otago.ac.nz (T.M. Davies).

https://doi.org/10.1016/j.spl.2018.02.067

to vary from data point to data point, in principle permitting the degree of smoothing to adapt to local requirements. In contrast, $h_1 = h_2 = \cdots = h_n$ for fixed bandwidth estimation.

The preeminent implementation of the sample point adaptive density estimator is due to Abramson (1982). He proposed setting the local bandwidths according to

$$h_i = h(\boldsymbol{x}_i) = h_0 f(\boldsymbol{x}_i)^{-1/2} \tag{2}$$

where h_0 is the *global bandwidth* and $f(\boldsymbol{x}_i)^{-1/2}$ a local adjustment factor. The form of the latter is very intuitive, indicating the need for progressively more smoothing as the density becomes smaller. Eq. (2) is often rewritten as $h_i = \tilde{h}_0 f(\boldsymbol{x}_i)^{-1/2}\gamma^{-1}$ where γ is the geometric mean of the local adjustments, so that $\tilde{h}_0$ operates on a comparable scale to a fixed bandwidth. For simplicity we tend to work with h_0 for mathematical developments, and $\tilde{h}_0$ when comparing empirical results between the fixed and adaptive cases.

The bandwidths defined by (2) can become arbitrarily large in the tails of f. This can impact negatively both on the practical performance of the adaptive estimator and its theoretical properties. A solution is to clip the local bandwidths at some pre-specified maximum size (Abramson, 1982; Hall and Marron, 1988). One way of achieving this is to truncate above all bandwidths at some (large) multiple of the smallest local bandwidth. This methodology is employed in the numerical experiments in Section 3.

While the local bandwidth factors describe the relative smoothing requirements from location to location, the overall performance of the adaptive estimator $\hat{f}$ still depends on the choice of global bandwidth h_0. In deriving a methodology to do this, it seems sensible to investigate what lessons can be learned from the extensive literature on bandwidth selection for fixed bandwidth estimators. There are a range of types of bandwidth selectors that have proven effective for density estimation, including normal reference rules (Silverman, 1986), plug-in methods (e.g. Sheather and Jones, 1991) and cross-validation techniques (e.g. Hall et al., 1992). A unifying theme with all these methods is the use of large-sample asymptotic analysis for a deterministic n and a diminishing bandwidth i.e. as $n \to \infty$, $h \to 0$ and $nh \to \infty$. Their success reflects the fact that the asymptotics tend to provide rather good approximations to finite sample behaviour in fixed kernel estimation. In what follows, we also proceed under these conditions.

For the adaptive estimator $\hat{f}$, the only practical method for selecting h_0 to have been published thus far is an unbiased cross-validation technique suggested by Silverman (1986). The development of more sophisticated approaches has been hampered by the complexity of the asymptotic analysis. In particular, while explicit asymptotic forms of the mean and variance of $\hat{f}$ have been published for the univariate adaptive density estimator (Silverman, 1986; Hall and Marron, 1988), there are no corresponding results for the bivariate case in the literature. The first purpose of this paper is to fill this gap by providing explicit results in the bivariate case; see Section 2.

A critical feature of the existing asymptotics for the adaptive estimator is that the expansions for bias and variance do not converge uniformly over $\mathbb{R}^2$ (Hall and Marron, 1988). This immediately raises doubts about the utility of these results as a basis for developing practical data-driven selectors for h_0. The second goal of this article is to explore this matter in some detail. We pay particular attention to the behaviour of the bias, contrasting asymptotic approximations with true finite sample behaviour (as described by very large numbers of simulations). The results are presented in Section 3 and a discussion follows in Section 4. In the supplementary materials we provide fuller plots of some simulation results and further illustrate practical ramifications through a real-world example.

2. Asymptotic results

In what follows, we assume standard regularity conditions. Specifically, we assume that all fourth order partial derivatives of f are continuous over $\mathbb{R}^2$, and that $\int K(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$; $\int \boldsymbol{x}K(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0$; and $\int x_v^2 K(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$, with $v \in \{1, 2\}$ indexing the bivariate coordinate $\boldsymbol{x} = [x_1, x_2]^\top$. As mentioned earlier, we assume that any sample used to estimate the density f is comprised of independent and identically distributed observations.

2.1. Fixed bandwidth

Explicitly denote the fixed bandwidth estimator by $\bar{f}(\boldsymbol{x})$; we obtain $\bar{f}$ from (1) when setting $h_i = h$ for all i. The asymptotic properties of $\bar{f}$ are well established (see e.g. Silverman, 1986; Wand and Jones, 1995). Both asymptotic mean and variance expressions are uniformly convergent over $\mathbb{R}^2$, and hence can safely be integrated. Squaring the bias, summing with the variance and integrating over $\mathbb{R}^2$ yields the familiar asymptotic mean integrated squared error (AMISE);

$$\mathrm{AMISE}\left[\bar{f}\right] = \frac{1}{nh^2}R(K) + \frac{1}{4}h^4 \int \mathrm{trace}\left[\mathcal{H}_f(\boldsymbol{z})\right]\,\mathrm{d}\boldsymbol{z}, \tag{3}$$

where $R(K) = \int K(\boldsymbol{z})^2\,\mathrm{d}\boldsymbol{z}$ and $\mathcal{H}_f(\boldsymbol{z})$ is the 2×2 Hessian matrix of f. The AMISE typically provides an excellent approximation to the exact mean integrated squared error of $\bar{f}(\boldsymbol{x})$, and has proven highly successful as the basis for development of practical bandwidth selectors.