On Bandwidth Choice for Spatial Data Density Estimation by Zhenyu Jiang, Nengxiang Ling, Zudi Lu, Dag Tjostheim, Qiang Zhang

Di SU

Department of Statistics, The Chinese University of Hong Kong

Summer Lab Meeting

Aug. 4th, 2021

Overview

1. Introduction
2. Spatial Adaptive Density Estimation
3. Spatial Cross-Validation Choice of Adaptive Bandwidth
   - Scenario 1: Pilot Density Given
   - Scenario 2: Pilot Density Estimated
4. Extension
5. Numerical Finite Sample Performances

Table of Contents

1. Introduction
2. Spatial Adaptive Density Estimation
3. Spatial Cross-Validation Choice of Adaptive Bandwidth
   - Scenario 1: Pilot Density Given
   - Scenario 2: Pilot Density Estimated
4. Extension
5. Numerical Finite Sample Performances

Introduction

Spatial kernel density estimation is a non-parametric smoothing method that has provided a powerful methodology to analyze spatially dependent data collected from real life.

For generality, consider the data to be the observations from

$$\{X_{\boldsymbol{i}} = (X_{\boldsymbol{i}}^{(1)}, X_{\boldsymbol{i}}^{(2)}, \ldots, X_{\boldsymbol{i}}^{(d)})\},$$

which is a d-dimensional stationary random field defined on a common probability space $(\Omega, \mathcal{F}, \mathcal{P})$ observed over a rectangular region defined by

$$\mathcal{I}_{\boldsymbol{n}} = \left\{ \boldsymbol{i} = (i_1, i_2, \ldots, i_N) \in \mathbb{Z}^N \mid 1 \leqslant i_k \leqslant n_k, k = 1, 2, \ldots, N \right\}.$$

It is popular to estimate f, the common density function of $X_{\boldsymbol{i}}$, by kernel density estimator.

Introduction

Kernel density estimation requires the choice of a smoothing parameter (bandwidth). In particular, adaptive bandwidth selection is being considered. The idea of adaptive density estimation is popular in application (see Davies and Hazelton (2010) and Lemke et al. (2015))

The advantage of an adaptive bandwidth is that it is attempting to enhance local, or observation-wise, smoothing, rather than a one bandwidth fits all type of smoothing.

Introduction

In the special case of density estimation for independent or time series data, bandwidth selection methods can be categorized into two generations according to Jones et al. (1996).

- 'First-generation': Cross-Validation (CV) methods. (Silverman (1986), Fan and Gijbels (1996), Fan and Yao (2003) and Gao (2007))
    - Non-adaptive CV bandwidth selection for independent and time series data. (Hall (1983), Stone (1984), Marron and Hardle (1986), Marron (1985, 1987), Hart and Vieu (1990), Scott (1992) and Kim and Cox (1997))
    - This paper generalizes the CV idea to the choice of adaptive bandwidth for spatial data.
    - The CV bandwidth looks a more natural and implementable option for the AKDE.

Introduction

- 'Second-generation': Basically plug-in or bootstrap based, which rely on selection of pilot bandwidths (often by rule of thumb). (e.g. Sheather and Jones (1991) and Marron (1992))
  - Perform better than the first-generation methods for independent data (Sheather and Jones (1991), Cao et al. (1994), Chiu (1996) and Jones et al. (1996)).
  - For adaptive kernel density estimation, no plug-in method has been seen even for independent data in the literature.
  - The proposal in this paper may also be seen as 'second-generation' in view of the use of pilot density.

Table of Contents

1. Introduction
2. Spatial Adaptive Density Estimation
3. Spatial Cross-Validation Choice of Adaptive Bandwidth
   - Scenario 1: Pilot Density Given
   - Scenario 2: Pilot Density Estimated
4. Extension
5. Numerical Finite Sample Performances

Adaptive KDE

- The total sample size in $\mathcal{I}_{\boldsymbol{n}}$ is denoted as $\tilde{\boldsymbol{n}} = \Pi_{k=1}^{N} n_k$ for $\boldsymbol{n} = (n_1, n_2, \ldots, n_N) \in \mathbb{Z}^N$ with strictly positive integer co-ordinates $n_1, n_2, \ldots, n_N$.
- As in Hallin et al. (2004 b), we write that $\boldsymbol{n} \to \infty$ if $\min_{1 \leqslant k \leqslant N} \{n_k\} \to \infty$, without requiring $\max_{1 \leqslant j,k \leqslant N} \{n_j / n_k\} \leqslant C$ for some $0 < C < \infty$ given in Tran (1990), allowing for multidirectional convergence in the sample size.

Spatial Kernel Density Estimator

$$\check{f}_{\boldsymbol{n}}(x) = \frac{1}{\tilde{\boldsymbol{n}}} \sum_{\substack{i_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} \frac{1}{h^d_{(i_1,i_2,\ldots,i_N)}} K\left(\frac{x - X_{(i_1,i_2,\ldots,i_N)}}{h_{(i_1,i_2,\ldots,i_N)}}\right), \quad x \in \mathbb{R}^d, \tag{1}$$

where the summation symbol stands for the N-fold summations $\sum_{i_1=1}^{n_1} \Sigma_{i_2=1}^{n_2} \ldots \sum_{i_N=1}^{n_N}$ and $K : \mathbb{R}^d \to \mathbb{R}^+$ is a kernel function.

Adaptive Bandwidth

Adaptive Bandwidth

$$h_{\boldsymbol{i}} \equiv h_{(i_1,i_2,\ldots,i_N)} \equiv h_{\boldsymbol{i}}\left(h_0; f, \delta\right) = \frac{h_0}{f\left(X_{(i_1,i_2,\ldots,i_N)}\right)^{\delta}\gamma_f} \tag{2}$$

where $\gamma_f = \left\{ \prod_{\substack{i_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} f\left(X_{(i_1,i_2,\ldots,i_N)}\right)^{-\delta} \right\}^{1/\tilde{\boldsymbol{n}}}$.

To determine the adaptive bandwidth is to determine the following three components:

- h_0. Inclusion of the geometric mean term γ_f in equation (2), as noted in Silverman (1986), is to free the bandwidth factors from dependence on the scale of the data, allowing the global bandwidth h_0 to be considered on the same scale as in the corresponding fixed non-adaptive bandwidth estimate of f.

Adaptive Bandwidth

- δ. In adaptive estimation for independent data, it has been suggested by Abramson (1982a, b) that $\delta = 1/2$ is optimal when f is second-order differentiable.
- f. The pilot density estimator of f which is chosen to be the traditional KDE $\hat{f}_{\boldsymbol{n}}(x)$ with a fixed bandwidth. It can be seen as a special case of (1) with $\delta = 0$ defined as

$$\hat{f}_{\boldsymbol{n}}(x) = \frac{1}{\tilde{\boldsymbol{n}} h^d} \sum_{i_k=1}^{n_k} \quad K\left(\frac{x - X_{(i_1, i_2, \ldots, i_N)}}{h}\right), \quad x \in \mathbb{R}^d, \qquad (3)$$
$$\forall k = 1, 2, \ldots, N$$

where $h = h_{\boldsymbol{n}} \to 0$ as $\boldsymbol{n} \to \infty$.

Adaptive Bandwidth

Remarks

- With the help of $\hat{f}_{\boldsymbol{n}}(x)$, the design of the estimator is then reduced to the selection of the pilot bandwidth $h_{\boldsymbol{n}}$ and the global bandwidth h_0.
- The selection of $h_{\boldsymbol{n}}$ is equivalent to the selection of h_0 with $\delta = 0$.

Table of Contents

1. Introduction
2. Spatial Adaptive Density Estimation
3. Spatial Cross-Validation Choice of Adaptive Bandwidth
    - Scenario 1: Pilot Density Given
    - Scenario 2: Pilot Density Estimated
4. Extension
5. Numerical Finite Sample Performances

Spatial Cross-Validation (SCV)

With the non-adaptive case, performance of leave-five-out CV seems very similar to that of leave-one-out CV (Lu et al. (2014)). Further, leave-one-out CV is more popular with spatial data (LeRest et al. (2014)).

In the following, the authors adopted leave-one-out CV for its simplicity in applications.

Scenario 1: Pilot Density Given

First consider the choice of the bandwidth (2) when the pilot density function f is given, then the choice reduces to the selection of the global bandwidth h_0.

Spatial Cross-Validation With a Given Pilot Density

$$\mathrm{SCV}_{\delta}\left(h_0\right) \equiv \mathrm{CV}_{\delta}\left(h_0\right) = \int_{\mathbb{R}^d} \check{f}_{\boldsymbol{n}}^2(x) w(x) \mathrm{d}x - \frac{2}{\tilde{\boldsymbol{n}}} \sum_{\substack{i_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} \check{f}_{\boldsymbol{n}}^{(\boldsymbol{i})}\left(X_{\boldsymbol{i}}\right) w\left(X_{\boldsymbol{i}}\right), \tag{4}$$

where SCV_{δ} stands for SCV for the global bandwidth h_0 with a given δ and some non-negative weight function $w(\cdot)$.

Scenario 1: Pilot Density Given

Spatial Cross-Validation With a Given Pilot Density (cont.)

And $\check{f}_{\boldsymbol{n}}^{(\boldsymbol{i})}(x)$ is the adaptive kernel estimator of f based on data points except $X_{\boldsymbol{i}}$ which is defined as

$$\check{f}_{\boldsymbol{n}}^{(\boldsymbol{i})}(x) = \frac{1}{\tilde{\boldsymbol{n}} - 1} \sum_{\substack{j_k=1 \\ \forall k=1,2,\dots,N \\ \exists k: j_k \neq i_k}}^{n_k} \frac{1}{h_{(j_1,j_2,\dots,j_N)}^{d}} K\left(\frac{x - X_{(j_1,j_2,\dots,j_N)}}{h_{(j_1,j_2,\dots,j_N)}}\right), \tag{5}$$

Then, the extended CV optimal smoothing parameter is defined by

$$\check{h}_0(f,\delta) \equiv \check{h}_0 = \arg \min_{h_0 \in \mathcal{H}_{\boldsymbol{n}}} \mathrm{SCV}_\delta\left(h_0\right). \tag{6}$$

Scenario 1: Pilot Density Given

The adaptive bandwidth includes the fixed bandwidth as a special case. When $\delta = 0$, the adaptive KDE (AKDE) (1) reduces to the non-adaptive (fixed bandwidth) KDE (3) which is independent of the pilot f. Therefore, the bandwidth h of estimator (3) can be selected following the above procedure with $\delta = 0$, i.e. to minimize an estimated ISE defined by

$$\mathrm{SCV}_0(h) \equiv \mathrm{CV}_0(h) = \int_{\mathbb{R}^d} \hat{f}_{\boldsymbol{n}}^2(x) w(x) \mathrm{d}x - \frac{2}{\tilde{\boldsymbol{n}}} \sum_{\substack{i_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} \hat{f}_{\boldsymbol{n}}^{(i_1,\ldots,i_N)}\left(X_{(i_1,\ldots,i_N)}\right) w\left(X_{(i_1,\ldots,i_N)}\right),$$

where $\hat{f}_{\boldsymbol{n}}^{(i_1,\ldots,i_N)}(x)$ is the kernel estimator of f based on $X_{\boldsymbol{j}}'\mathrm{s}, \boldsymbol{j} = (j_1, \ldots, j_N) \neq \boldsymbol{i} = (i_1, \ldots, i_N)$. Then, the SCV optimal smoothing parameter for equation (3) is defined by

$$\hat{h} = \arg\min_{h \in \mathcal{H}_{\boldsymbol{n}}} \mathrm{SCV}_0(h).$$

Measure of Optimality

There is a growing body of opinion for using performance criteria, which are not just the MISE (see Mammen (1990), Jones (1991), Hardle and Vieu (1992) and Loader (1999)), targeting at estimating the unknown probability density function (in the context of this paper).

The authors concretely measured the optimality of the selected bandwidth by considering the widely used ISE, MISE and ASE, defined respectively as follows.

Integrated Squared Error (ISE)

$$d_{\mathrm{I}}\left(\check{f}_{\boldsymbol{n}}, f\right)(h) = \mathrm{ISE}(h) = \int_{\mathbb{R}^d} \left\{\check{f}_{\boldsymbol{n}}(x) - f(x)\right\}^2 w(x)\mathrm{d}x. \tag{7}$$

Mean Integrated Squared Error (MISE)

$$d_{\mathrm{M}}\left(\check{f}_{\boldsymbol{n}}, f\right)(h) = \mathrm{MISE}(h) = E\left[\int_{\mathbb{R}^d} \left\{\check{f}_{\boldsymbol{n}}(x) - f(x)\right\}^2 w(x)\mathrm{d}x\right]. \tag{8}$$

Measure of Optimality

Approximated Squared Error (ASE)

$$d_{\mathrm{A}}\left(\check{f}_{\boldsymbol{n}}, f\right)(h) = \mathrm{ASE}(h) = \frac{1}{\tilde{n}} \sum_{\substack{j_k=1\\ \forall k=1,2,\dots,N}}^{n_k} \left\{\check{f}_{\boldsymbol{n}}\left(X_{(j_1,\dots,j_N)}\right) - f\left(X_{(j_1,\dots,j_N)}\right)\right\}^2 w\left(X_{(j_1,\dots,j_N)}\right). \tag{9}$$

Denote

$$R = \frac{1}{\tilde{\boldsymbol{n}}} \sum_{\substack{j_k=1\\ \forall k=1,2,\dots,N}}^{n_k} f\left(X_{(j_1,\dots,j_N)}\right) w\left(X_{(j_1,\dots,j_N)}\right) - E\left\{f\left(X_{(j_1,\dots,j_N)}\right) w\left(X_{(j_1,\dots,j_N)}\right)\right\} \tag{10}$$

and

$$T = -\int_{\mathbb{R}^d} f(x)^2 w(x)\mathrm{d}x - 2R$$

Assumptions

Some assumptions are needed to establish the optimality of the SCV-selected adaptive bandwidth.

Assumptions

- (K1) The kernel function K is a bounded function symmetric with respect to zero, as well as Holder continuous and compactly supported, satisfying $\int_{\mathbb{R}^d} K(t)dt = 1$.
- (K2) For any non-negative components of $(i_1, i_2, \ldots, i_d)$ with $i_1 + i_2 + \ldots + i_d \leq r$, denote

  $$S\left(K, i_1, i_2, \ldots, i_d\right) = \int_{\mathbb{R}^d} t_1^{i_1} \ldots t_d^{i_d} K(t)dt$$

  which satisfies the properties of r th-order kernels that

  $$S\left(K, i_1, \cdots, i_d\right) = 0, \text{ when for any } j, i_j < r, \text{ with } i_1 + i_2 + \ldots + i_d > 0$$

  and $0 < \left|S\left(K, i_1, \cdots, i_d\right)\right| < \infty$, if there is some j such that $i_j = r$ where r is a positive integer given in (D1) below.

Assumptions

Assumptions

- (K3) The convolution of kernel function K with itself, $\tilde{K}$, is absolutely integrable.
- (K4) The kernel function K is differentiable, and its characteristic function, $\psi_K(t) = \int_{\mathbb{R}^d} e^{\iota t' u} K(u) du$, with $\iota^2 = -1$, satisfies $|\psi_K(t)| \leq c_K \, |\psi_{\mathcal{N}}(t)|$, where $\psi_{\mathcal{N}}(t) = e^{-t't/2}$ is the characteristic function of d-dimensional standard normal distribution, $c_K > 0$ is a constant and t' is the transpose of $t \in \mathbb{R}^d$
- (D1) The bounded density function f is Hölder continuous with rth order continuous differentiations.
- (D2)(i) The joint probability density function $f_{\mathrm{i},\mathrm{j}}(x, y)$ of X_{i} and X_{j} exists and satisfies $|f_{\mathrm{i},\mathrm{j}}(x, y) - f(x)f(y)| \leq C$ for all x, y and all $\mathrm{i} \neq \mathrm{j}$. (ii) The conditional probability density function $\hat{f}_{\mathrm{i}_1,\cdots,\mathrm{i}_s|\mathrm{j}_1,\cdots,\mathrm{j}_s}(x_1, \cdots, x_s \mid y_1, \cdots, y_s)$ of $(X_{\mathrm{i}_1}, \cdots, X_{\mathrm{i}_s})$ given $(X_{\mathrm{j}_1} = y_1, \cdots, X_{\mathrm{j}_s} = y_s)$ is bounded for all $x_k, y_k \in \mathbb{R}^d$ and all $\mathrm{i}_k \neq \mathrm{j}_k, k = 1, \cdots, s$, with $1 \leq s \leq 2r$

Assumptions

Assumptions

- (W) $w(.)$ is bounded and integrable with a compact support $S_w \subset \mathbb{R}^d$.
- (D3) The density function $f(\cdot)$ is bounded away from zero on S_w, that is $\inf_{x \in S_w} f(x) \geq c_w > 0$
- (H) $h \in \mathcal{H}_{\mathrm{n}} = \left[a\tilde{\mathrm{n}}^{-\frac{1}{2r+d}}, b\tilde{\mathrm{n}}^{-\frac{1}{2r+d}}\right]$ for some constants a and b with $0 < a < b < \infty$.
- (M) The mixing coefficient $\varphi(t)$ satisfies

  $$\varphi(t) = \mathcal{O}\left(t^{-\mu}\right)$$

  with $\mu > 2Nr(2 - 2/q)/(1 - 2/q)$ for some $q > 2$.

Criteria Equivalence

The SCV criterion proposed is asymptotically equivalent to the ISE criterion in the selection of bandwidth with spatial kernel density estimation, when compared with the MISE.

Theorem 1.

- (a) When $\delta = 0$, under assumptions (K1), (K2), (D1), (D2), (M), (H) and (W), we have

$$\frac{|\mathrm{SCV}_\delta(h) - \mathrm{ISE}(h) - T|}{\mathrm{MISE}(h)} = \mathcal{O}_P\left(\tilde{\boldsymbol{n}}^{-d/\{2(2r+d)\}}\right). \tag{11}$$

  Further, if assumption (K3) is satisfied, we have

$$\sup_{h \in \mathcal{H}_{\boldsymbol{n}}} \frac{|\mathrm{SCV}_\delta(h) - \mathrm{ISE}(h) - T|}{\mathrm{MISE}(h)} = o_P(1), \text{ as } \boldsymbol{n} \to \infty. \tag{12}$$

- (b) When $\delta > 0$, in addition to the conditions in (a), if assumptions (K4) and (D3) are satisfied, then conclusions (11) and (12) hold true.

Criteria Equivalence

Both the ISE and the ASE are asymptotically equivalent to the MISE criterion in bandwidth selection with spatial kernel density estimation.

Theorem 2.

- (a) When $\delta = 0$, under assumptions $(\mathrm{K1}) - (\mathrm{K3}), (\mathrm{D1})$ and $(\mathrm{D2}), (\mathrm{M}), (\mathrm{H})$ and (W), we have

$$\sup_{h \in \mathcal{H}_{\boldsymbol{n}}} \left| \frac{\mathrm{ISE}(h) - \mathrm{MISE}(h)}{\mathrm{MISE}(h)} \right| = o_p(1), \qquad \text{and} \qquad \sup_{h \in \mathcal{H}_{\boldsymbol{n}}} \left| \frac{\mathrm{ASE}(h) - \mathrm{MISE}(h)}{\mathrm{MISE}(h)} \right| = o_p(1)$$

  as $\boldsymbol{n} \to \infty$

- (b) When $\delta > 0$, in addition to the conditions in (a), if assumptions (K4) and (D3) are satisfied, then conclusions above hold true.

Optimality

The bandwidth $\hat{h}$ that is selected by the suggested CV is asymptotically optimal in terms of the criteria involving the ISE and the ASE as well as the MISE for spatial kernel density estimation.

Theorem 3.

- (a) When $\delta = 0$, under the conditions for part (a) of theorem 2, we have

$$\frac{d\left(\check{f}_{\boldsymbol{n}}, f\right)\left(\check{h}_0\right)}{\inf_{h \in \mathcal{H}_{\mathrm{n}}} d\left(\check{f}_{\boldsymbol{n}}, f\right)(h)} \rightarrow 1$$

  in probability, as $\boldsymbol{n} \rightarrow \infty$, where d is any of $d_{\mathrm{I}}, d_{\mathrm{A}}$ and d_{M}, and $\check{f}_{\boldsymbol{n}}(x) = \hat{f}_{\boldsymbol{n}}(x)$ and $\check{h}_0 = \hat{h}$ respectively.
- (b) When $\delta > 0$, in addition to the conditions in (a), if assumptions (K4) and (D3) are satisfied, then conclusion above holds true.

Scenario 2: Pilot Density Estimated

- In practice, we may not know the pilot density and thus need a pilot estimator $\hat{f}_{\boldsymbol{n}}$, hence the practical AKDE can be defined as

$$\check{\check{f}}_{\boldsymbol{n}}(x) = \frac{1}{\tilde{\boldsymbol{n}}} \sum_{\substack{i_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} \frac{1}{\check{h}^d_{(i_1,i_2,\ldots,i_N)}} K\left(\frac{x - X_{(i_1,i_2,\ldots,i_N)}}{\check{h}_{(i_1,i_2,\ldots,i_N)}}\right), \quad x \in \mathbb{R}^d, \tag{13}$$

  where $\check{h}_{(i_1,i_2,\ldots,i_N)} = h_{(i_1,i_2,\ldots,i_N)}\left(h_0; \hat{f}_{\boldsymbol{n}}, \delta\right)$ with $\hat{f}_{\boldsymbol{n}}$ replacing f in equation (2).

- Then the SCV optimal smoothing parameter is defined by

$$\check{\check{h}}_0\left(\hat{f}_{\boldsymbol{n}}, \delta\right) \equiv \check{\check{h}}_0 = \arg\min_{h_0 \in \mathcal{H}_{\boldsymbol{n}}} S\check{\mathrm{C}}\mathrm{V}_{\delta}\left(h_0\right), \tag{14}$$

  where $\mathrm{S\check{C}V}_{\delta}\left(h_0\right)$ is as defined in equation (4) with $h_{(i_1,i_2,\ldots,i_N)}$ replaced by $\check{h}_{(i_1,i_2,\ldots,i_N)}$.

gsj

Optimality

The optimality of the SCV selected bandwidth can be extended to the practical AKDE.

Theorem 4.

Under the conditions for part (b) of theorem 3 with $\sup_{x\in S_w}\left|\hat{f}_{\boldsymbol{n}}(x)-f(x)\right|\to 0$ in probability, we have

$$\frac{d\left(\check{\check{f}}_{\boldsymbol{n}}, f\right)\left(\check{\check{h}}_0\right)}{\inf_{h_0\in\mathcal{H}_{\boldsymbol{n}}} d\left(\check{\check{f}}_{\boldsymbol{n}}, f\right)(h_0)} \to 1$$

in probability as $\boldsymbol{n}\to\infty$, where d is any of $d_{\mathrm{I}}, d_{\mathrm{A}}$ and d_{M}.

We call this property an oracle property in the sense that the asymptotic optimality is achieved as if f were known.
CV-selected $\check{h}_0$ with the true f: the oracle CV bandwidth.
CV-selected h_0 with f replaced by $\hat{f}_{\boldsymbol{n}}$: the estimated adaptive CV bandwidth for h_0.

Table of Contents

1. Introduction
2. Spatial Adaptive Density Estimation
3. Spatial Cross-Validation Choice of Adaptive Bandwidth
    - Scenario 1: Pilot Density Given
    - Scenario 2: Pilot Density Estimated
4. Extension
5. Numerical Finite Sample Performances

Extension

Spatial Trends

As done in Hallin et al. (2009) and Lu et al. (2014), the R package `sm` can be used to remove the spatial trends, and it can be proved that the theoretical results for the (estimated) detrended data, say $\hat{X}_{\boldsymbol{i}}$, replacing the unobservable stationary $X_{\boldsymbol{i}}$ in the above sections can still hold under some appropriate conditions (see section 3 of Hallin et al. (2009)).

Regression

Similarly, with a more involved regression setting, ideas and techniques developed from this paper will be useful in adaptive bandwidth selection for conditional regression of lattice data (Hallin et al. (2004b)). CV-based adaptive bandwidth selection can be extended more naturally than other procedures for spatial regression, which is beyond the scope of this paper and will be left for future research.

Table of Contents

1. Introduction

2. Spatial Adaptive Density Estimation

3. Spatial Cross-Validation Choice of Adaptive Bandwidth
   - Scenario 1: Pilot Density Given
   - Scenario 2: Pilot Density Estimated

4. Extension

5. Numerical Finite Sample Performances

Simulations

- The proposed spatial adaptive CV bandwidth choice outperforms the existing non-adaptive R routines such as the 'rule of thumb' and the so-called 'second-generation' Sheather and Jones (1991) bandwidths both for moderate sizes of spatial samples and in particular for big spatial data sets.
- $f(x) = 0.4\phi_{(\mu_1=-1,\sigma_1^2=0.1656)}(x) + 0.3\phi_{(\mu_2=0.4,\sigma_2^2=0.088)}(x) + 0.3\phi_{(\mu_3=1.5,\sigma_3^2=0.352)}(x)$.
- Our empirical application to a set of spatial soil data will further illustrate that non-Gaussian features of the data are more significantly identified by spatial adaptive density estimation.

Conclusion

Conclusion

- Extended CV selection to spatial adaptive KDE.
- Derived optimality results.
- Finite sample performance better than rule of thumb and robust to non-Gaussian data.

Thank You!