european
social fund in the
czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

OP Education
for Competitiveness

INVESTMENTS
IN EDUCATION
DEVELOPMENT

Streamlining the Applied Mathematics Studies
at Faculty of Science of Palacký University in Olomouc
CZ.1.07/2.2.00/15.0243

International Conference
Olomoucian Days of Applied Mathematics

ODAM 2011

Department of Mathematical analysis
and Applications of Mathematics
Faculty of Science
Palacký University Olomouc

Gregor Reisch: Margarita Philosophica (1508)

Multivariate kernel density estimate

Ivana Horová

joint work with

Kamila Vopatová – Jan Koláček – Jiří Zelinka – Martin Řezáč

Department of Mathematics and Statistics
Masaryk University, Brno

ODAM 2011

devoted to Professor Lubomír Kubáček in occasion of his 80^{th} birthday

research supported by MŠMT project LC06024 – V08

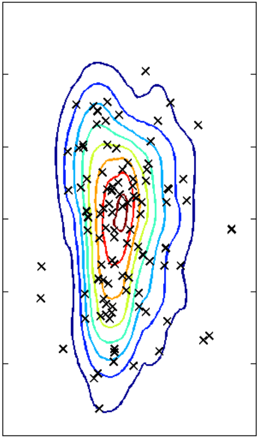
Outline

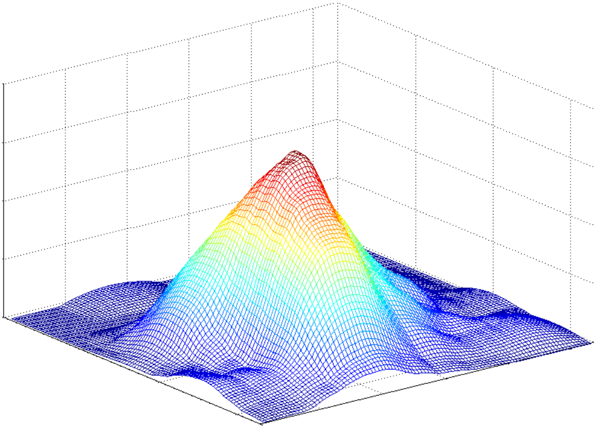
1. Kernel density estimate
2. Kernel density gradient estimate
3. Simulations
4. Real data

Outline

1. Kernel density estimate
2. Kernel density gradient estimate
3. Simulations
4. Real data

Motivation





Kernel density estimate: KDE

A kernel density estimate for a d-variate random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ drawn from a density f is defined as

$$\hat{f}(\mathbf{x}, H) = \frac{1}{n}\sum_{i=1}^{n} K_H(\mathbf{x} - \mathbf{X}_i) = \frac{1}{n}|H|^{-1/2}\sum_{i=1}^{n} K\left(H^{-1/2}(\mathbf{x} - \mathbf{X}_i)\right),$$

where

- K ⇾ a d-variate kernel function satisfying $\int_{\mathbb{R}^d} K(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1$
- H ⇾ a symmetric positive definite matrix called the bandwidth matrix
- $|H|$ denotes a determinant of H
- $\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$

Notation and assumptions

(A1) $K \multimap$ a symmetric probability density function: $\int_{\mathbb{R}^d} K(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1$, $\int_{\mathbb{R}^d} \mathbf{x} K(\mathbf{x}) dx = 0$, $\int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^T K(\mathbf{x}) dx = \beta_2(K) I_d$, $I_d \multimap d \times d$ identity matrix,

(A2) $R(K) = \int_{\mathbb{R}^d} K^2(\mathbf{x})\, \mathrm{d}\mathbf{x}$,

(A3) $R(g) = \int_{\mathbb{R}^d} g(\mathbf{x}) g(\mathbf{x})^T\, \mathrm{d}\mathbf{x}$ for any square integrable vector valued function g.

- $A^{\otimes r} \multimap$ the r^{th} Kronecker power of a matrix A, $A^{\otimes 1} = A$, $A^{\otimes 0} = 1$
- $D^{\otimes r} f(\mathbf{x}) \multimap$ the vector containing all partial derivatives of the order r of f at $\mathbf{x}$, i.e. if $f : \mathbb{R}^d \to \mathbb{R} \Rightarrow D^{\otimes r} f(\mathbf{x}) \in \mathbb{R}^{d^r}$, $D^{\otimes 1} f = Df$ is a gradient of f
- $\operatorname{vec} H \multimap d^2 \times 1$ vector obtained by stacking columns of H

Bandwidth matrix H

- The most important factor
- It induces orientation of kernel and controls a spread of a kernel

(B1) $\mathcal{H}_{\mathcal{F}}$: a class of symmetric positive definite $d \times d$ matrices

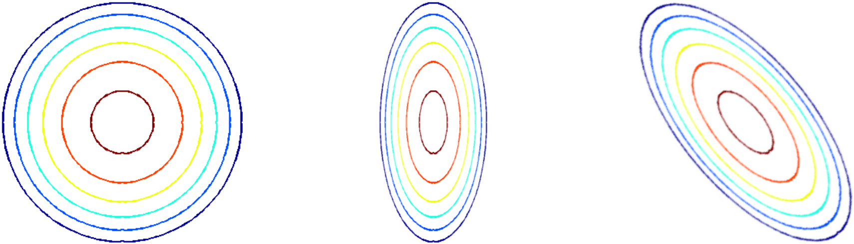
(B2) $\mathcal{H}_{\mathcal{D}} \subset \mathcal{H}_{\mathcal{F}}$: a subclass of diagonal positive definite matrices

(B3) $\mathcal{H}_{\mathcal{S}} \subset \mathcal{H}_{\mathcal{D}}$: a subclass of matrices $\mathcal{H}_{\mathcal{S}} = \{h^2 \cdot I_d, h > 0\}$

Bandwidth matrix H

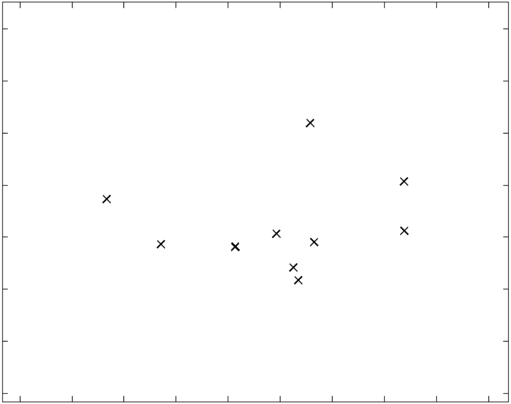
How does matrix H affect the shape of the kernel (bivariate case)

$$\mathcal{H}_{\mathcal{S}} : h^2 \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathcal{H}_{\mathcal{D}} : \begin{pmatrix} h_1^2 & 0 \\ 0 & h_2^2 \end{pmatrix} \qquad \mathcal{H}_{\mathcal{F}} : \begin{pmatrix} h_1^2 & h_{12} \\ h_{12} & h_2^2 \end{pmatrix}$$



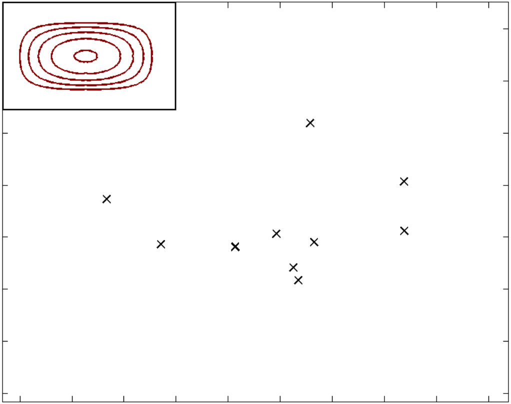
Sample – Kernel – Density

For given data



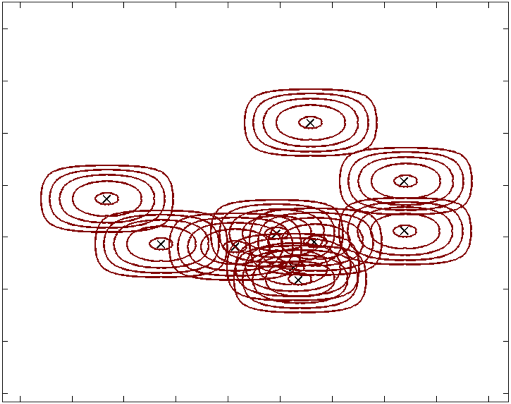
Sample – Kernel – Density

we choose a kernel, e.g. Epanechnikov product kernel



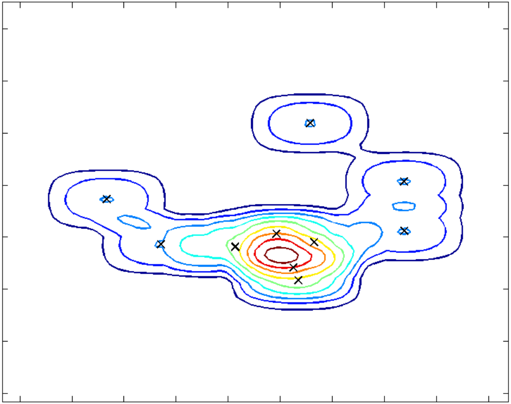
Sample – Kernel – Density

evaluate kernel function in each point



Sample – Kernel – Density

and get contours of reconstructed density



MISE

The quality of the estimate $\hat{f}$ is measured by the **M**ean **I**ntegrated **S**quare **E**rror

$$\begin{aligned} MISE(H) &= E\int\left(\hat{f}(\mathbf{x},H)-f(\mathbf{x})\right)^2 \, \mathrm{d}\mathbf{x} \\ &= \frac{1}{n}\int\left[(K_H^2 * f)(\mathbf{x})-(K_H * f)^2(\mathbf{x})\right] \, \mathrm{d}\mathbf{x} \\ &\quad + \int\left[(K_H * f)(\mathbf{x})-f(\mathbf{x})\right]^2 \, \mathrm{d}\mathbf{x} \\ &= \int \mathrm{Var}\,\hat{f}(\mathbf{x},H)\,\mathrm{d}\mathbf{x}+\int \mathrm{Bias}^2\,\hat{f}(\mathbf{x},H)\,\mathrm{d}\mathbf{x} \end{aligned}$$

where $*$ denotes a convolution.

AMISE

The MISE can be approximated by AMISE – **A**symptotic **M**ean **I**ntegrated **S**quare **E**rror

Assumptions:

- All the second derivatives of f are piecewise continuous and square integrable
- $H = H_n$ is a sequence of bandwidth matrices such that $n^{-1}|H|^{-1/2}$ and all entries of H approach zero as $n \to \infty$
- K satisfies assumptions (A1)

$$\mathrm{AMISE}(H) = \underbrace{n^{-1}|H|^{-1/2}R(K)}_{\mathrm{AIVar}(H)} + \underbrace{\frac{\beta_2(K)^2}{4}(\operatorname{vec} H)^T R(D^{\otimes 2} f)(\operatorname{vec} H)}_{\mathrm{AIBias}^2(H)},$$

Optimal bandwidth matrix H

- Optimal H with respect to MISE

$$H_{\text{MISE}} = \arg\min_{\mathcal{H}} \text{MISE}(H)$$

- Optimal H with respect to AMISE

$$H_{\text{AMISE}} = \arg\min_{\mathcal{H}} \text{AMISE}(H)$$

Relative rate of convergence

$$\text{vec}(H_{\text{AMISE}} - H_{\text{MISE}}) = O(J_d n^{-2/(d+4)}) \,\text{vec}\, H_{\text{MISE}},$$

where J_d is a $d \times d$ matrix of ones.

Choice of the optimal bandwidth matrix

$$\frac{\partial \operatorname{AMISE}(H)}{\partial \operatorname{vec} H} = D_H \operatorname{AMISE}(H) = -\frac{1}{2} n^{-1} |H|^{-1/2} R(K) \operatorname{vec} H^{-1} + \frac{\beta_2(K)^2}{2} R(D^{\otimes 2} f) \operatorname{vec} H$$

H_{AMISE} is the solution of the equation $D_H \operatorname{AMISE}(H) = \mathbf{0}$. For $d > 2$ there is not close form expression for the solution of this equation.

Lemma

Let H_{AMISE} be a minimum of $\operatorname{AMISE}(H)$. Then

$$\operatorname{AIVar}(H_{\mathrm{AMISE}}) = \frac{4}{d} \operatorname{AIBias}^2(H_{\mathrm{AMISE}}).$$

Remark. $H_{\mathrm{AMISE}} = O(J_d n^{-2/(d+4)})$ and $\operatorname{AMISE}(H_{\mathrm{AMISE}}) = O(n^{-4/(d+4)})$.

Data-driven bandwidth matrix selectors

- The least square cross-validation (LSCV) targets MISE and employs the objective function

$$\mathrm{LSCV}(H) = \int_{\mathbb{R}^d} \hat{f}^2(\mathbf{x}, H)\, \mathrm{d}\mathbf{x} - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(\mathbf{X}_i, H)$$

where $$\hat{f}_{-i}(\mathbf{X}_i, H) = \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{n} K_H(\mathbf{X}_i - \mathbf{X}_j), \quad i = 1, \ldots, n.$$

$$\widehat{H}_{\mathrm{LSCV}} = \arg\min_{\mathcal{H}} \mathrm{LSCV}(H)$$

$E\left[\mathrm{LSCV}(H)\right] = \mathrm{MISE}(H) - R(f).$

- Biased cross-validation (BCV) involves estimation of AMISE
- Smooth cross-validation (SCV) is a hybrid of LSCV and BCV
- Plug-in method (PI) estimates the functional $R(D^{\otimes 2} f)$ in the AMISE

Estimate of AMISE

$$\widehat{\text{AMISE}}(H) = \int_{\mathbb{R}^d} \widehat{\text{Var}}(\hat{f}(\mathbf{x}, H))\ \text{d}\mathbf{x} + \int_{\mathbb{R}^d} \widehat{\text{Bias}}^2(\hat{f}(\mathbf{x}, H))\ \text{d}\mathbf{x}$$
$$= \text{AI}\,\widehat{\text{Var}}(H) + \text{AI}\,\widehat{\text{Bias}}^2(H),$$

where

$$\text{AI}\,\widehat{\text{Var}}(H) = \frac{1}{n}|H|^{-1/2}R(K),$$
$$\text{AI}\,\widehat{\text{Bias}}(H) = \frac{1}{n^2}\sum_{i,j=1}^{n}\big(K_H * K_H * K_H * K_H - 2K_H * K_H * K_H$$
$$+ K_H * K_H\big)(\mathbf{X}_i - \mathbf{X}_j)$$

Let

$$\widehat{H}_{\text{AMISE}} = \arg\min_{\mathcal{H}} \widehat{\text{AMISE}}(H).$$

API – Method for data-driven bandwidth matrix selector

API method is based on the Lemma, i.e. to select such a matrix $\widehat{H}_{\text{AMISE}}$ for which the equation

$$\text{AIVar}(H) = \frac{4}{d}\,\text{AIBias}^2(H)$$

is satisfied. This equation can be rewritten as

$$|H|^{1/2} = \frac{dR(K)}{4ng(H)},$$

where

$$g(H) = \sum_{i,j=1}^{n} \big(K_H * K_H * K_H * K_H - 2K_H * K_H * K_H + K_H * K_H\big)(\mathbf{X}_i - \mathbf{X}_j).$$

API

Previous equation is nonlinear equation for $d(d+1)/2$ unknowns – entries of $\widehat{H}_{\text{AMISE}}$. Additional equations:

- $\mathcal{H} = \mathcal{H}_{\mathcal{S}} \multimap$ the only equation for h
- $\mathcal{H} = \mathcal{H}_{\mathcal{D}} \multimap \widehat{H}_{\text{AMISE}} = \text{diag}(h_1^2, \ldots, h_d^2)$
  Scott's rule in $\mathbb{R}^d$: $\hat{h}_i = \hat{\sigma}_i n^{-1/(d+4)}$ for $i = 1, \ldots, d$, $\Rightarrow \hat{h}_i = \frac{\hat{\sigma}_i}{\hat{\sigma}_1}\hat{h}_1$,
  $\hat{\sigma}$ is a sample standard deviation estimate
- $\mathcal{H} = \mathcal{H}_{\mathcal{F}} \multimap$ we can adopt a similar idea as in the case of the diagonal matrix: $\widehat{\Sigma} = \left(\hat{\sigma}_{ij}\right)_{i,j=1}^{d}$ is an estimate of a sample covariance matrix:
  $h_1^2 = h_{11} = \hat{\sigma}_{11} n^{-2/(d+4)}$, $\qquad h_i^2 = h_{ii} = \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{11}} h_{11}$ for $i = 2, \ldots, d$,
  $h_{ij} = \frac{\text{sign}\,\hat{\sigma}_{ij} |\hat{\sigma}_{ij}|}{\hat{\sigma}_{11}} h_{11}$ for $i, j = 2, \ldots, d$, $i \neq j$

API – special case

$d = 2,\ H \in \mathcal{H}_{\mathcal{D}}$

$$\mathrm{AMISE}(h_1, h_2) = \underbrace{\frac{1}{nh_1 h_2} R(K)}_{\mathrm{AIVar}} + \underbrace{\frac{1}{4}\beta_2(K)^2 (h_1^4 \psi_{40} + 2h_1^2 h_2^2 \psi_{22} + h_2^4 \psi_{04})}_{\mathrm{AIBias}^2},$$

where

$$\psi_{k\ell} = \int \left(\frac{\partial^2 f}{\partial x_1^2}\right)^{k/2} \left(\frac{\partial^2 f}{\partial x_2^2}\right)^{\ell/2} \mathrm{d}\mathbf{x} \qquad k, \ell = 0, 2, 4, \quad k + \ell = 4$$

API1 and API2 methods

API 1 method:

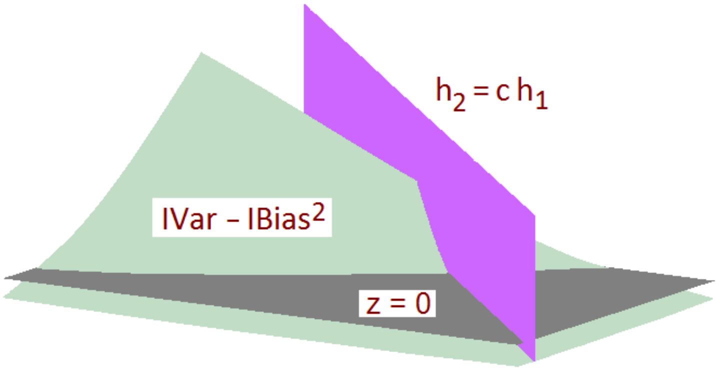
$$g(\hat{h}_1, \hat{h}_2) = \frac{n}{2} R(K) \qquad \hat{h}_2 = \frac{\hat{\sigma}_2}{\hat{\sigma}_1} \hat{h}_1$$

API 2 method:

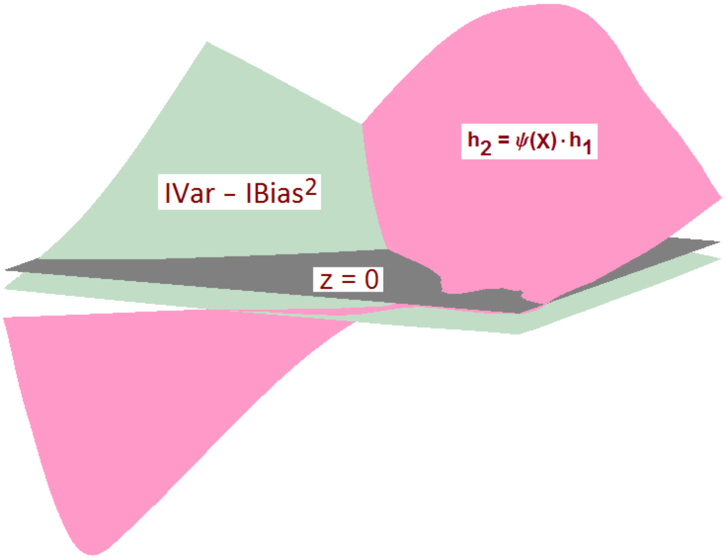
$$g(\hat{h}_1, \hat{h}_2) = \frac{n}{2} R(K) \qquad \hat{h}_2 = \left(\frac{\hat{\psi}_{40}}{\hat{\psi}_{04}} \right)^{1/4} \hat{h}_1$$

- $\hat{\psi}_{04} = \frac{1}{n^2 h_1 h_2^5} \sum_{i,j=1}^{n} C_K \left(\frac{X_{1j} - X_{1i}}{h_1} \right) C_{K''} \left(\frac{X_{2j} - X_{2i}}{h_2} \right)$
- $\hat{\psi}_{40} = \frac{1}{n^2 h_1^5 h_2} \sum_{i,j=1}^{n} C_{K''} \left(\frac{X_{1j} - X_{1i}}{h_1} \right) C_K \left(\frac{X_{2j} - X_{2i}}{h_2} \right)$
- $C_K(x) = \int K(t) K(x-t) \, \mathrm{d}t, \qquad C_{K''}(x) = \int K''(t) K(x-t) \, \mathrm{d}t$

API1 method – basis



API2 method – basis



Outline

1. Kernel density estimate
2. Kernel density gradient estimate
3. Simulations
4. Real data

Kernel density gradient estimate

A kernel estimate of a gradient Df is defined as

$$\widehat{Df}(\mathbf{x}, H) = \frac{1}{n} \sum_{i=1}^{n} DK_H(\mathbf{x} - \mathbf{X}_i),$$

where $DK_H(\mathbf{x}) = |H|^{-1/2} H^{-1/2} DK(H^{-1/2}\mathbf{x})$,
MISE is the measure of the quality of the estimate

$$\begin{aligned} \mathrm{MISE}(\widehat{Df}, H) &= \int E\Big(\widehat{Df}(\mathbf{x}, H) - Df(\mathbf{x})\Big)\Big(\widehat{Df}(\mathbf{x}, H) - Df(\mathbf{x})\Big)^T \, \mathrm{d}\mathbf{x} \\ &= \int \mathrm{Var}(\widehat{Df}(\mathbf{x}, H)) \, \mathrm{d}\mathbf{x} + \int \|E\widehat{Df}(\mathbf{x}, H) - Df(\mathbf{x})\|_2^2 \, \mathrm{d}\mathbf{x} \end{aligned}$$

$\|\cdot\|_2$ denotes Euclidean norm.

TAMISE

MISE is a matrix for a gradient estimate. Duong et al. (2008) proposed to use the **T**race of the **A**symptotic **M**ean **I**ntegrated **S**quare **E**rror: TAMISE

$$\begin{aligned}\mathrm{TAMISE}(H) = {} & n^{-1}|H|^{-1/2}\,\mathrm{tr}\big[H^{-1}R(DK)\big] \\ & + \frac{\beta_2(K)^2}{4}\,\mathrm{tr}(I_d \otimes \mathrm{vec}^T H)R(D^{\otimes 3}f)(I_d \otimes \mathrm{vec}\, H).\end{aligned}$$

TAMISE – special case

In the bivariate case with a diagonal bandwidth matrix TAMISE can be rewritten in the form

$$\mathrm{TAMISE}(H) = \underbrace{= \frac{1}{nh_1^3 h_2^3}\left(h_2^2 R(\partial_1 K) + h_1^2 R(\partial_2 K)\right)}_{\mathrm{TIVar}} + \underbrace{\frac{1}{4}\beta_2(K)^2\left(h_1^4(\psi_{60} + \psi_{42}) + 2h_1^2 h_2^2(\psi_{42} + \psi_{24}) + h_2^4(\psi_{24} + \psi_{06})\right)}_{\mathrm{TIBias}^2},$$

where $R(\partial_i K) = \int \left(\frac{\partial K(\mathbf{x})}{\partial x_i}\right)^2 \mathrm{d}\mathbf{x},\ i = 1, 2.$

Bandwidth matrix choice

Let H_T be a bandwidth matrix minimizing TAMISE:

$$H_T = \arg\min_{\mathcal{H}} \mathrm{TAMISE}(H).$$

Then

$$H_T = O(J_d n^{-2/(d+6)}), \qquad \textit{and} \qquad \mathrm{TAMISE}(H_T) = O(n^{-4/(d+6)}),$$

H_T is the solution of

$$D_H \,\mathrm{TAMISE} = \frac{\partial\, \mathrm{TAMISE}(H)}{\partial \operatorname{vec} H} = \mathbf{0}$$

$\rightharpoonup$ there is not any explicit solution.

Data - driven bandwidth matrix choice

Practical bandwidth matrix choice

In the case of diagonal bandwidth matrix ($H_T \in \mathcal{H}_{\mathcal{D}}$):

$$\widehat{H}_T = \operatorname{diag}(\hat{h}_{T1}^2, \ldots, \hat{h}_{Td}^2),$$
$$\hat{h}_{Ti}^2 = \hat{h}_i^2 n^{\frac{4}{(d+4)(d+6)}} \left(\hat{\sigma}_i^2\right)^{\frac{4}{(d+4)(d+6)}},$$

where $\hat{h}_i$ $(i = 1, \ldots, d)$ are optimal bandwidths for density estimate, $\hat{\sigma}_i$ $(i = 1, \ldots, d)$ are estimates of sample standard deviations.

TAMISE estimate and TAPI method

Lemma

Let H_T *be a minimizer of* TAMISE*. Then*

$$\frac{d+2}{4}\,\mathrm{TIVar}(H_T) = \mathrm{TIBias}^2(H_T).$$

This equation can be rewritten as

$$|H|^{1/2} = \frac{d+2}{4n}\,\frac{\mathrm{tr}\left[H^{-1}R(DK)\right]}{\mathrm{TIBias}^2(H)}$$

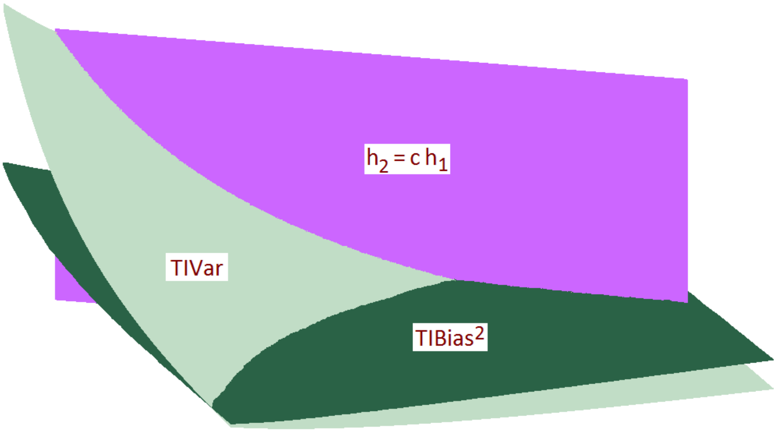
TAPI method

The idea of TAPI method is the same as in the case of density estimate, we use a suitable estimate of $\mathrm{TIBias}^2(H)$

$$\widehat{\mathrm{TIBias}}^2(H) = \operatorname{tr} \frac{1}{n^2} \sum_{i,j=1}^{n} \int \left[(K_H * DK_H - DK_H)(\mathbf{x} - \mathbf{X}_i)\right] \times \\ \times \left[(K_H * DK_H - DK_H)(\mathbf{x} - \mathbf{X}_j)\right]^T \,\mathrm{d}\mathbf{x}.$$

The additional equations can be obtained by means of practical bandwidth choice and Scott's rule.

TAPI method – basis



Outline

1 Kernel density estimate

2 Kernel density gradient estimate

3 Simulations

4 Real data

Simulations

- density $f \sim N_2(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2, \varrho)$
- normal kernel
$$K(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2}{2} - \frac{x_2^2}{2}\right)$$
- number of observation $n = 100$
- number of repetition $R = 100$

Simulations

The integrated square error

- for density f:

$$\mathrm{ISE}(\hat{f}, H) = \int \left[\hat{f}(\mathbf{x}, H) - f(\mathbf{x})\right]^2 \, \mathrm{d}\mathbf{x}$$

$\overline{\mathrm{ISE}} = \operatorname{avg} \mathrm{ISE}(\hat{f}, H),$

- for density gradient Df:

$$\mathrm{TISE}(\widehat{Df}, H) = \operatorname{tr} \int \left[\widehat{Df}(\mathbf{x}, H) - Df(\mathbf{x})\right] \left[\widehat{Df}(\mathbf{x}, H) - Df(\mathbf{x})\right]^T \, \mathrm{d}\mathbf{x}$$

$\overline{\mathrm{TISE}} = \operatorname{avg} \mathrm{TISE}(\widehat{Df}, H),$

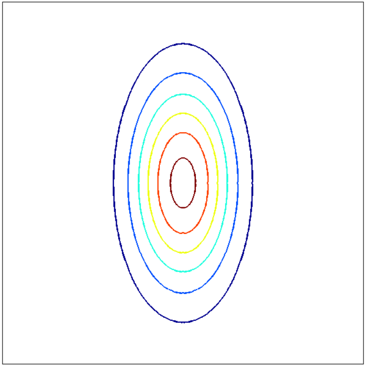
where the average is taken over simulated samples.

Normal A

$$\mathbf{X} \sim N_2(0, 0; 1/4, 1, 0)$$

ISE	avg	std
$H_{\text{API}\,1}$	0.0097	0.0043
H_{AMISE}	0.0234	0.0030

TISE	avg	std
H_{TAPI}	0.0801	0.0306
H_{TAMISE}	0.0866	0.0263

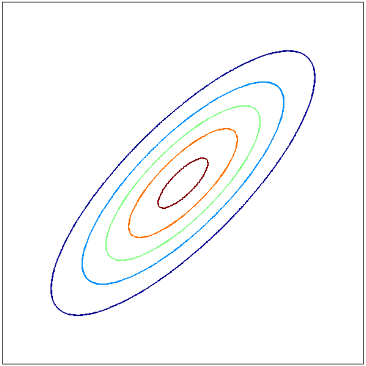


Normal L

$$\mathbf{X} \sim N_2(0, 0; 1, 1, 4/5)$$

ISE	avg	std
$H_{\mathrm{API}\,1}$	0.0077	0.0035
H_{AMISE}	0.0224	0.0031

TISE	avg	std
H_{TAPI}	0.0693	0.0295
H_{TAMISE}	0.0755	0.0250

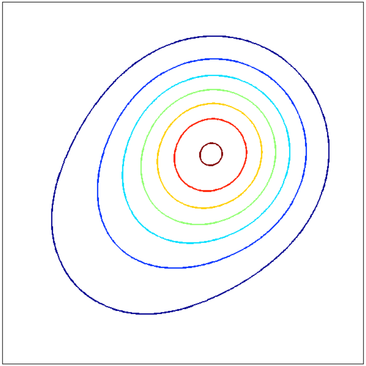


Normal D

$$\mathbf{X} \sim 1/5N_2(0,0;1,1,0) + 1/5N_2(1/2,1/2;4/9,4/9,0) \\ +3/5N_2(13/12,13/12;25/81,25/81,0)$$

ISE	avg	std
$H_{\mathrm{API\,1}}$	0.0109	0.0046
H_{AMISE}	0.0333	0.0045

TISE	avg	std
H_{TAPI}	0.0889	0.0294
H_{TAMISE}	0.0907	0.0265

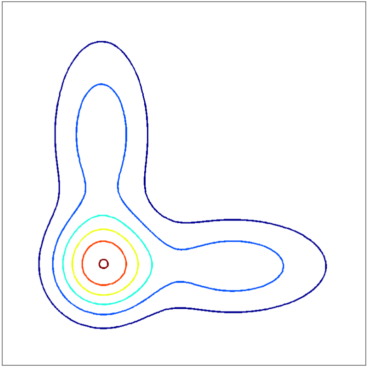


Normal P

$$\mathbf{X} \sim 1/3N_2(0,0;1,1,0) + 1/3N_2(0,4;1,4,0) + 1/3N_2(4,0;4,1,0)$$

ISE	avg	std
$H_{\text{API}\,1}$	0.0027	0.0008
H_{AMISE}	0.0027	0.0007

TISE	avg	std
H_{TAPI}	0.0055	0.0013
H_{TAMISE}	0.0049	0.0010

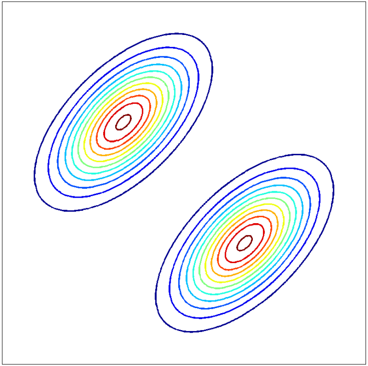


Normal R

$$\mathbf{X} \sim 1/2N_2(1,-1;4/9,4/9,3/5) + 1/2N_2(-1,1;4/9,4/9,3/5)$$

ISE	avg	std
$H_{\mathrm{API}\,1}$	0.0189	0.0040
H_{AMISE}	0.0146	0.0029

TISE	avg	std
H_{TAPI}	0.1709	0.0242
H_{TAMISE}	0.1049	0.0269



ISE results

The average of ISE with a standard deviation:

density	ISE($H_{\mathrm{API}\,1}$)		ISE(H_{AMISE})	
	avg	std	avg	std
A	0.0097	0.0043	0.0234	0.0030
L	0.0077	0.0035	0.0224	0.0031
D	0.0109	0.0046	0.0333	0.0045
P	0.0027	0.0008	0.0027	0.0007
R	0.0189	0.0040	0.0146	0.0029

TISE results

The average of TISE with a standard deviation:

density	TISE(H_{TAPI})		TISE(H_{TAMISE})	
	avg	std	avg	std
A	0.0801	0.0306	0.0866	0.0263
L	0.0693	0.0295	0.0755	0.0250
D	0.0889	0.0294	0.0907	0.0265
P	0.0055	0.0013	0.0049	0.0010
R	0.1709	0.0242	0.1049	0.0269

Outline

1. Kernel density estimate
2. Kernel density gradient estimate
3. Simulations
4. Real data

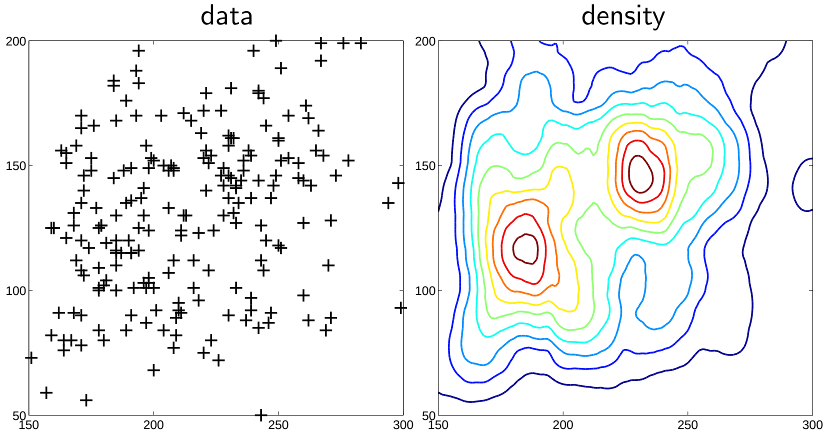
Lipids

Concentration of plasma cholesterol and plasma triglyceride taken on 320 patients with chest pain in a heart disease study.

- X_1 - cholesterol [mg/100 ml]
- X_2 - triglyceride [mg/100 ml]

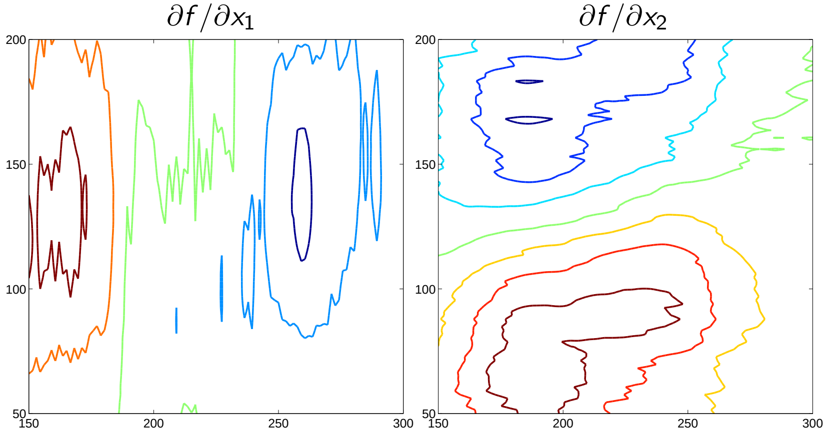
Lipids: density

The reconstructed density with a diagonal bandwidth matrix $H = \text{diag}(14.99^2, 25.58^2)$.



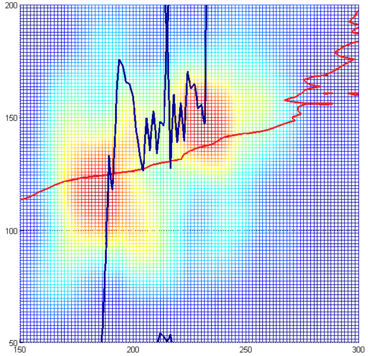
Lipids: density gradient

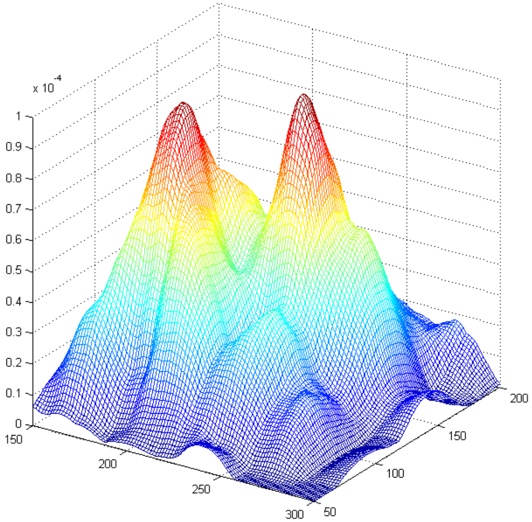
The reconstructed density gradient with a diagonal bandwidth matrix $H_T = \text{diag}(31.57^2, 68.34^2)$.



Lipids: $\hat{f}$ and $\widehat{Df} = 0$

$\partial f/\partial x_1 = 0$ $\partial f/\partial x_2 = 0$





References

- J.E. Chacón, T. Duong, M.P. Wand: Asymptotics for general multivariate kernel density derivative estimators, *(preprint)*, 2008.
- T. Duong, A. Cowling, I. Koch, M.P. Wand: Feature significance for multivariate kernel density estimation, *Computational Statistics and Data Analysis*, vol. 52, pp. 4225–4242, 2008.
- I. Horová, J. Koláček, K. Vopatová: Visualization and Bandwidth Matrix Choice, *Communications in Statistics – Theory and Methods*, to appear.
- D.W. Scott: *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York: John Wiley and Sons, 1992.
- K. Vopatová, I. Horová, J. Koláček: Bandwidth Matrix Choice for Bivariate Kernel Density Derivative, *Proceedings of IWSM*, pp. 561-564, 2010.
- M.P. Wand, M.C. Jones: *Kernel Smoothing*, London: Chapman and Hall, 1995.