Methods for bandwidth detection in kernel conditional density estimations

Kateřina Konečná
Supervisor: prof. Ivana Horová

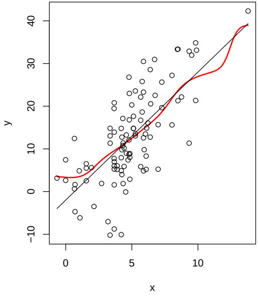
Institute of Mathematics and Statistics
Masaryk University
Brno, Czech Republic

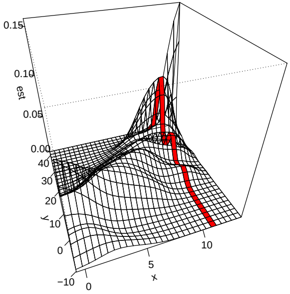
EYSM 2017
Uppsala, Sweden
$14^{th} - 18^{th}$ August 2017

Content

Motivation

- distribution of a random variable $Y|(X = x)$





Kernel smoothing, kernel function

A real valued function K satisfying

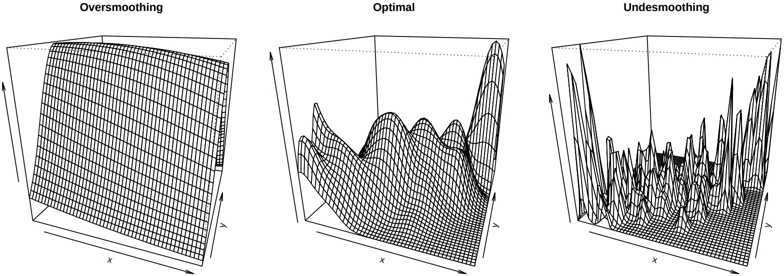
1. $K \in Lip[-1,1]$, ie. $|K(x) - K(y)| \leq L|x-y|$, $\forall x, y \in [-1,1]$, $L > 0$,
2. $supp(K) = [-1,1]$,
3. moment conditions:

$$\int_{-1}^{1} x^j K(x)\,\mathrm{d}x = \begin{cases} 1 & j = 0, \\ 0 & j = 1, \\ \beta_2(K) \neq 0 & j = 2 \end{cases}$$

is called a kernel of order 2.

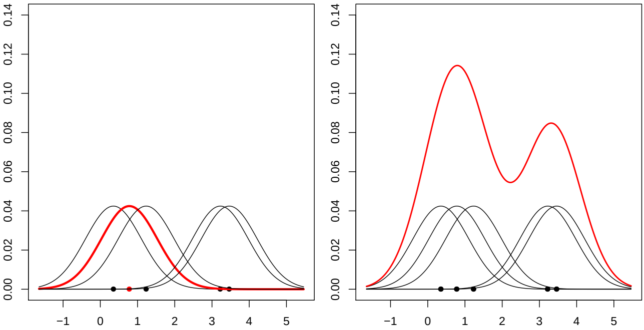
Smoothing parameters

- influence of the smoothing parameters on a final estimation



The construction of the estimator

- main idea of construction



- the conditional density of a random variable $Y|(X = x)$

$$f(y|x) = \frac{f(x,y)}{g(x)}$$

Types of kernel conditional density estimators

Kernel conditional density estimator

$$\hat{f}(y|x) = \sum_{i=1}^{n} w_i(x) K_{h_y}(y - Y_i),$$

- Nadaraya-Watson weights ([1])

$$w_i(x) = \frac{K_{h_x}(x - X_i)}{\sum_{j=1}^{n} K_{h_x}(x - X_j)},$$

- local linear weights ([2])

$$w_i(x) = \frac{\left(\hat{s}_2(x) - (x - X_i)\hat{s}_1(x)\right) K_{h_x}(x - X_i)}{\hat{s}_0(x)\hat{s}_2(x) - \hat{s}_1^2(x)},$$

where $\hat{s}_r(x) = \frac{1}{n}\sum_i (x - X_i)^r K_{h_x}(x - X_i)$

Asymptotic Bias (AB) and Asymptotic Variance (AV)

Theorem. ([5]) Let $h_x \to 0$, $h_y \to 0$, $nh_x h_y \to \infty$ as $n \to \infty$. The asymptotic bias (AB) and asymptotic variance (AV) are given by the expressions

$$\mathrm{AB}\left\{\hat{f}_{NW}(y|x)\right\} = \frac{1}{2}h_x^2\beta_2(K)\left(2\frac{g'(x)}{g(x)} + \frac{\partial^2 f(y|x)}{\partial x^2}\right) + \frac{1}{2}h_y^2\beta_2(K)\frac{\partial^2 f(y|x)}{\partial y^2},$$

$$\mathrm{AV}\left\{\hat{f}_{NW}(y|x)\right\} = \frac{1}{nh_x h_y} \cdot \frac{R^2(K)f(y|x)}{g(x)},$$

where $R(K) = \int K^2(t)\,\mathrm{d}t$.

- optimal values of bandwidths

$$\mathrm{MSE}\left\{\hat{f}_{NW}(y|x)\right\} \to \mathrm{MISE}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} \to \mathrm{AMISE}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} \to h_x^*,\ h_y^*$$

Optimal values of the smoothing parameters I

- local measurement of the quality: Asymptotic Mean Squared Error

$$\mathrm{AMSE}\left\{\hat{f}_{NW}(y|x)\right\} = \left(\mathrm{AB}\left\{\hat{f}_{NW}(y|x)\right\}\right)^2 + \mathrm{AV}\left\{\hat{f}_{NW}(y|x)\right\}$$

$$= \left(\frac{1}{2}h_x^2\beta_2(K)\left(2\frac{g'(x)}{g(x)} + \frac{\partial^2 f(y|x)}{\partial x^2}\right) + \frac{1}{2}h_y^2\beta_2(K)\frac{\partial^2 f(y|x)}{\partial y^2}\right)^2 + \frac{R^2(K)}{nh_xh_y}\cdot\frac{f(y|x)}{g(x)}$$

- global measurement of the quality of the estimate: Mean Integrated Squared Error

$$\mathrm{MISE}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} = \iint \mathrm{E}\left\{\left(\hat{f}_{NW}(y|x) - f(y|x)\right)^2\right\} g(x)\,\mathrm{d}x\,\mathrm{d}y$$

- Asymptotic Mean Integrated Squared Error

$$\mathrm{AMISE}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} = \frac{c_1}{nh_xh_y} + c_2h_x^4 + c_3h_y^4 + c_4h_x^2h_y^2$$

Optimal values of the smoothing parameters II

Optimal values of the smoothing parameters: $\mathrm{AMISE}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} \to \min$

$$h_x^* = n^{-1/6} c_1^{1/6} \left(4 \left(\frac{c_3^5}{c_4} \right)^{1/4} + 2c^5 \left(\frac{c_3}{c_4} \right)^{3/4} \right)^{-1/6},$$

$$h_y^* = h_x \left(\frac{c_3}{c_4} \right)^{1/4} = n^{-1/6} c_1^{1/6} \left(4 \left(\frac{c_4^5}{c_3} \right)^{1/4} + 2c_5 \left(\frac{c_4}{c_3} \right)^{3/4} \right)^{-1/6}$$

and

$$c_1 = \int R^2(K)\,\mathrm{d}x,$$

$$c_2 = \iint \frac{1}{4}\beta_2^2(K) \left(2\frac{g'(x)}{g(x)} + \frac{\partial^2 f(y|x)}{\partial x^2} \right)^2 g(x)\,\mathrm{d}y\,\mathrm{d}x,$$

$$c_3 = \iint \frac{1}{4}\beta_2^2(K) \left(\frac{\partial^2 f(y|x)}{\partial y^2} \right)^2 g(x)\,\mathrm{d}y\,\mathrm{d}x,$$

$$c_4 = \iint \frac{1}{2}\beta_2^2(K) \left(2\frac{g'(x)}{g(x)} + \frac{\partial^2 f(y|x)}{\partial x^2} \right) \frac{\partial^2 f(y|x)}{\partial y^2} g(x)\,\mathrm{d}y\,\mathrm{d}x.$$

The cross-validation method ([3], [4])

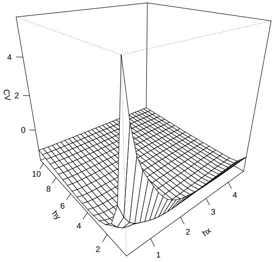
- proper estimation of the Integrated Squared Error

$$CV\left(h_x, h_y\right) = \frac{1}{n}\sum_{i=1}^{n}\int \hat{f}_{-i,NW}\left(y|X_i\right)^2 \,\mathrm{d}y - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i,NW}\left(Y_i|X_i\right),$$

where $\hat{f}_{-i,NW}\left(Y_i|X_i\right)$ is the estimate in the point $\left(X_i, Y_i\right)$ using points $\{\left(X_j, Y_j\right), j \neq i\}$

- bandwidth estimation:

$$\left(\hat{h}_x^{CV}, \hat{h}_y^{CV}\right) = \underset{(h_x, h_y)}{\arg\min}\, CV(h_x, h_y)$$



The iterative method ([6])

- proper relation between the terms setting AMISE:

$$\mathrm{AIV}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} - 2\mathrm{AISB}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} = 0$$

- an approximation of the AISB term:

$$\widehat{\mathrm{ISB}}\left\{\hat{f}_{NW}(\cdot|\cdot)\right\} = \iint \left(\widehat{\mathrm{bias}}\left\{\hat{f}_{NW}(y|x)\right\}\right)^2 \hat{g}(x)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \iint \left(\frac{\sum_i K_{h_x\sqrt{2}}(x-X_i)\,K_{h_y\sqrt{2}}(y-Y_i)}{\sum_i K_{h_x\sqrt{2}}(x-X_i)} - \hat{f}_{NW}(y|x)\right)^2 \hat{g}(x)\,\mathrm{d}x\,\mathrm{d}y$$

- additional equation:

$$\hat{h}_y = \hat{c}\hat{h}_x,$$

$\hat{c}$ given by the reference rule ([1])

The leave-one-out maximum likelihood method

- modified likelihood function

$$\mathcal{L}\left(h_x, h_y \mid \mathbf{X}, \mathbf{Y}\right) = \prod_{j=1}^{n} \hat{f}_{-j,NW}\left(Y_j \mid X_j\right)$$

- the estimations of the smoothing parameters:

$$(\hat{h}_x^{\mathcal{L}}, \hat{h}_y^{\mathcal{L}}) = \underset{(h_x, h_y)}{\arg\max}\, \mathcal{L}\left(h_x, h_y \mid \mathbf{X}, \mathbf{Y}\right).$$

- the likelihood function (left) and modified likelihood function (right)

The leave-one-out maximum likelihood method

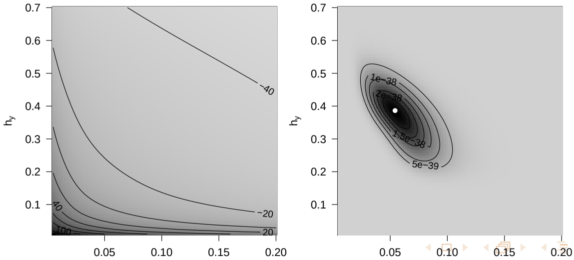
- modified likelihood function

$$\mathcal{L}\left(h_x, h_y \mid \mathbf{X}, \mathbf{Y}\right) = \prod_{j=1}^{n} \hat{f}_{-j,NW}\left(Y_j \mid X_j\right)$$

- the estimations of the smoothing parameters:

$$(\hat{h}_x^{\mathcal{L}}, \hat{h}_y^{\mathcal{L}}) = \underset{(h_x, h_y)}{\arg\max}\, \mathcal{L}\left(h_x, h_y \mid \mathbf{X}, \mathbf{Y}\right).$$

- the likelihood function (left) and modified likelihood function (right)



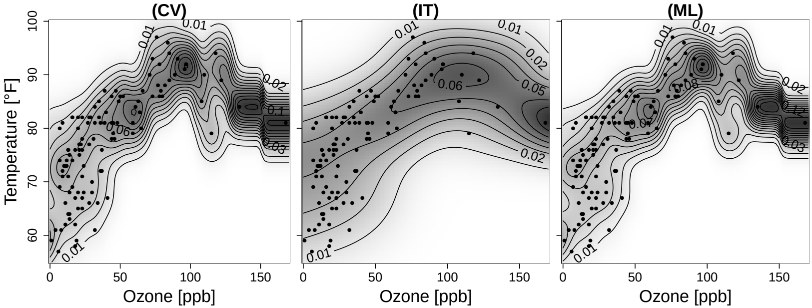
Data introduction

- the `airquality` data from the `datasets` package in R
- daily air quality in New York, May 1 to September 30, 1973
  - Ozone: Mean ozone in parts per billion
  - Temp: Maximum daily temperature in degrees Fahrenheit
- 153 observations in total, 116 observations included due to missing values
- comparison of the proposed methods: cross-validation method (CV), the iterative method (IT), the leave-one-out maximum likelihood method (ML)

Results

- Temperature | Ozone

method	$\hat{h}_x$	$\hat{h}_y$	computational time [s]
CV	5.61	3.29	182
IT	21.87	4.39	67.6
ML	6.28	2.89	31.3



References

[1] D. M. Bashtannyk and R. J. Hyndman. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3): 279 – 298, 2001.

[2] De Gooijer, J. G., Zerom, D. On conditional density estimation. *Statistica Neerlandica*, vol. 57, no. 2, 2003, p. 159–176.

[3] J. Fan and T. H. Yim. A crossvalidation method for estimating conditional densities. *Biometrika*, 91(4): 819–834, 2004.

[4] B. E. Hansen. Nonparametric conditional density estimation. *Unpublished manuscript*, 2004.

[5] R. J. Hyndman, D. M. Bashtannyk, and G. K. Grunwald. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5(4): 315–336, 1996.

[6] K. Konečná and I. Horová. *Conditional Density Estimations*, pages 15–31. International Society for the Advancement of Science and Technology (ISAST), Athens, 2014.