Lecture course *Statistics II*
Winter semester 2016/17
Ruprecht-Karls-Universität Heidelberg

Prof. Dr. Jan JOHANNES
Xavier LOIZEAU

Exercise sheet 7

*There are 5 exercises with a **total of 30 points** on this sheet. **14 points** are counted as **bonus**.*

Exercise 1. A common elementary density estimator is an *histogram estimator*. Let $\mathbb{p}$ be a density with support in the interval $[0,1]$ and $X_1,\dots,X_n \overset{i.i.d.}{\sim} \mathbb{p}$. For $m \in \mathbb{N}$ define bins

$$B_1 = \left[0, \frac{1}{m}\right),\ B_2 = \left[\frac{1}{m}, \frac{2}{m}\right),\ \dots,\ B_m = \left[\frac{m-1}{m}, 1\right],$$

with bin-width $h = 1/m$. Furthermore let Y_i be the number of observations in bin B_i, $\widehat{\mathbb{f}}_i := Y_i/n$ and $\mathbb{f}_i := \lambda(\mathbb{p}\,\mathbb{1}_{B_i})$, where $\mathbb{1}_{B_i}$ denotes the indicator function on the interval B_i. The histogram estimator is defined by

$$\widehat{\mathbb{p}}_h(x) = \sum_{i=1}^{m} \frac{\widehat{\mathbb{f}}_i}{h} \mathbb{1}_{B_i}(x).$$

(a) Let m and x be fixed with $x \in B_j$. Find expressions for the expectation and variance of $\widehat{\mathbb{p}}_h(x)$ in dependence of $\mathbb{f}_i$.

(b) Show that $\lim_{h\to 0} \mathbb{E}(\widehat{\mathbb{p}}_h(x)) = \mathbb{p}(x)$, if $\mathbb{p}$ is continuous.

(c) Let $\mathbb{p}$ be differentiable with absolute continuous derivative $\dot{\mathbb{p}}$ satisfying $\lambda(\dot{\mathbb{p}}^2) < \infty$. Show that

$$\text{MISE} = \frac{h^2}{12}\lambda(\dot{\mathbb{p}}^2) + \frac{1}{nh} + o(h^2) + o\left(\frac{1}{n}\right).$$

Find a value h^* for the bin-width which minimises the last expression.
Hint: Calculate first the bias and then the variance of the estimator. Considering the bias term use a Taylor development with appropriate expression for the reminder term and integrate it by decomposing the integral into sum of integrals over the bins. For the variance term a mean-value theorem for integrals might be helpful.

(d) Show that the MISE with value h^* derived in (c) for large sample sizes n equals approximately $Cn^{-2/3}$ with $C = (3/4)^{2/3}\left(\lambda(\dot{\mathbb{p}}^2)\right)^{1/3}$. Compare it with the $\text{MISE}(h_o)$ for the kernel density estimator. Which estimator would you choose, if you want to attain the fastest decay of the MISE?

(8 points)

Exercise 2. Consider the *Bart Simpson* density

$$\mathbb{p}_{Bart}(x) = \frac{1}{2}\phi(x;0,1) + \frac{1}{10}\sum_{i=0}^{4}\phi(x;(i/2)-1, 1/10),$$

where $\phi(x;\mu,\sigma^2)$ denotes the density of a $\mathfrak{N}(\mu,\sigma^2)$ normal-distribution with mean μ and variance σ^2.

(a) Use the software package `R` and the commands `dnorm` and `plot`, to define and to plot the density. The plot explains the name of the density.

(b) Use the command `rnorm` to generate a sample from a normal-distribution. Describe first theoretically an algorithm, how it can be used to generate a sample from the density $\mathbb{p}_{Bart}$ and secondly, implement the algorithm.

(c) With the command `hist` can you create histograms in `R`. Generate sufficiently many i.i.d. r.v.'s with common density $\mathbb{p}_{Bart}$ and calculate a histogram. Select appropriately the parameter `breaks` of the function `hist` and plot in addition the theoretical density. (6 points)

Exercise 3. In the following we study empirically kernel density estimation and its robustness using the software package `R`.

(a) Create a data set from a normal distribution of appropriate size using the command `rnorm`. The kernel density estimator is implemented as function `density`. Have a look at its possible parameters using the command `?density`. Plot the kernel density estimator selecting two different kernels. Compare the estimated densities with the density of the data-generating normal distribution by using the command `curve` which allows to add the true density to your plotted estimators.

(b) Visualise the behaviour of the kernel density estimator for different bandwidths. Therefore, generate three data sets from a normal distribution with sample size $n = 50, 500$, and 5000, respectively. Select the bandwidth by using the rule of thumb (Silverman). Select in addition two other interesting bandwidths and plot each estimator together with the true density.

(c) Study the behaviour of the estimator if some of the data points are outliers. Therefore, generate six data sets each of size $n = 5000$ such that, respectively, $1\%, 5\%$ and 10% of the data is not generated by a normal distribution but a Cauchy distribution with location parameter 0.5 and scale parameter 1 and a $\chi^2_5(0.5)$-distribution. How does the kernel density estimator behave? Would you say, that it is robust in the sense that it is stable w.r.t. outliers? (6 points)

In the next two exercises we develop theory for multivariate kernel density estimators.

Exercise 4. Let $X_1,\dots$ be i.i.d. $\mathbb{R}^d$-valued r.v.'s with common density $\mathbb{p}$. For $i \in [\![1,d]\!]$ let K_i be a kernel, i.e., $K_i : \mathbb{R} \to \mathbb{R}$ is integrable with $\lambda K_i = 1$, we call $K(x) = \prod_{i=1}^d K_i(x^i)$, $x = (x^1,\dots,x^d) \in \mathbb{R}^d$, a product kernel. Given a product kernel K and a diagonal matrix $H := \operatorname{diag}(h^1,\dots,h^d)$ with bandwidth-vector $h = (h^1,\dots,h^d)$a multivariate kernel density estimator of $\mathbb{p}(x)$ for $x \in \mathbb{R}^d$ is defined by

$$\widehat{\mathbb{p}}_H(x) := \frac{1}{n\det(H)} \sum_{i=1}^n K(H^{-1}(X_i - x)).$$

Let $\mathbb{p}$ be twice continuous partial-differentiable in a neighbourhood of a point $x \in \mathbb{R}^d$.

Consider a product kernel K with symmetric, bounded, compactly supported kernels K_i, $i \in [\![1, d]\!]$. Show that for $n \prod_{i=1}^{d} h^i \to \infty$ and $h^i = o(1)$ as $n \to \infty$ holds

$$\begin{aligned} \mathbb{V}\mathrm{ar}_{\mathbb{p}}(\widehat{\mathbb{p}}_H(x)) &= \frac{1}{n \prod_{i=1}^{d} h^i} \mathbb{p}(x) \prod_{i=1}^{d} \lambda(K_i^2) + o\left(\frac{1}{n \prod_{i=1}^{d} h^i}\right), \\ \mathrm{bias}_{\mathbb{p}}(x) &= \frac{1}{2} \sum_{i=1}^{d} (h^i)^2 \frac{\partial^2}{\partial (x^i)^2} \mathbb{p}(x) \lambda(\mathrm{id}^2 K_i) + o\left(\max\{(h^i)^2, i \in [\![1, d]\!]\}\right). \end{aligned}$$

(5 points)

Exercise 5. Let K be kernel as in exercise 4, $H =: \mathrm{diag}(h^1, \ldots, h^d)$ with bandwidth-vector $h = (h^1, \ldots, h^d)$, and let the density $\mathbb{p}$ be continuous in a neighbourhood of $x \in \mathbb{R}^d$. Show that for $n \prod_{i=1}^{d} h^i \to \infty$ and $h^i = o(1)$ as $n \to \infty$ holds

$$\sqrt{n \prod_{i=1}^{d} h^i} \left(\widehat{\mathbb{p}}_H(x) - \mathbb{E}_{\mathbb{p}} \widehat{\mathbb{p}}_H(x)\right) \xrightarrow{d} \mathfrak{N}\left(0, \mathbb{p}(x) \prod_{i=1}^{d} \lambda(K_i^2)\right).$$

(5 points)

A HAPPY NEW YEAR.

Handing in during the lecture on **Friday, January 13, 2017** in **fixed groups of two**.