

TP1 : Estimation de la Densité

Evgenii Chzhen & VONG Henry

13 october 2015

Contents

Introduction	1
1. Influence of choice of m parametr in histogramm	1
2. Influence of the choice of the kernel	2
3. Chosing optimal h by minimazing $J(h)$	3
3.1. Transformed data	4
4. Working with real data	5
References	5

Introduction

We consider a random variables X with the following density function :

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} + \frac{1}{2} \mathbb{1}_{[1,2]}(x).$$

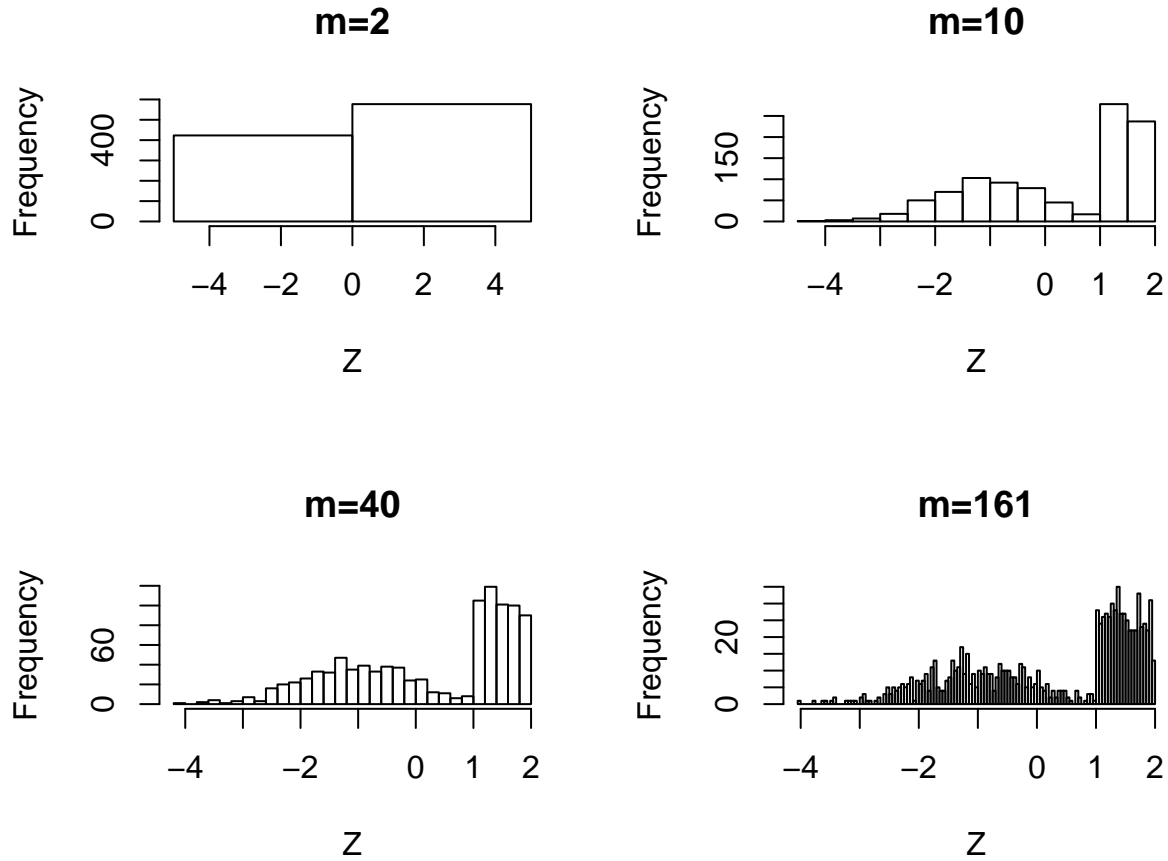
We simulate $n = 1000$ obsevation of variable X from given density.

1. Influence of choice of m parametr in histogramm

I this section we study estimation of a density function by histogramm. This methods can be represented as it follows: take a support of the density $p(x) \in [-M, M]$ and a parametr $m > 0$, consider following intervals $C_j = [\frac{(j-1)2M}{m}, \frac{j2M}{m}]$ for $j \in \{1, \dots, m\}$ and compute the following estimation

$$\tilde{p}_{n,m}(x) = \frac{m}{2M} \sum_{j=1}^m \left[\mathbb{1}_{C_j}(x) \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{C_j}(X_k) \right].$$

We use different values of m to understand how it influences the estimation.



Corollary: It can be seen that small values are useless as well as large ones. With small values we have lack of information and with large ones we overdefine our density. One can see that $10 \leq m \leq 40$ are visually quite good.

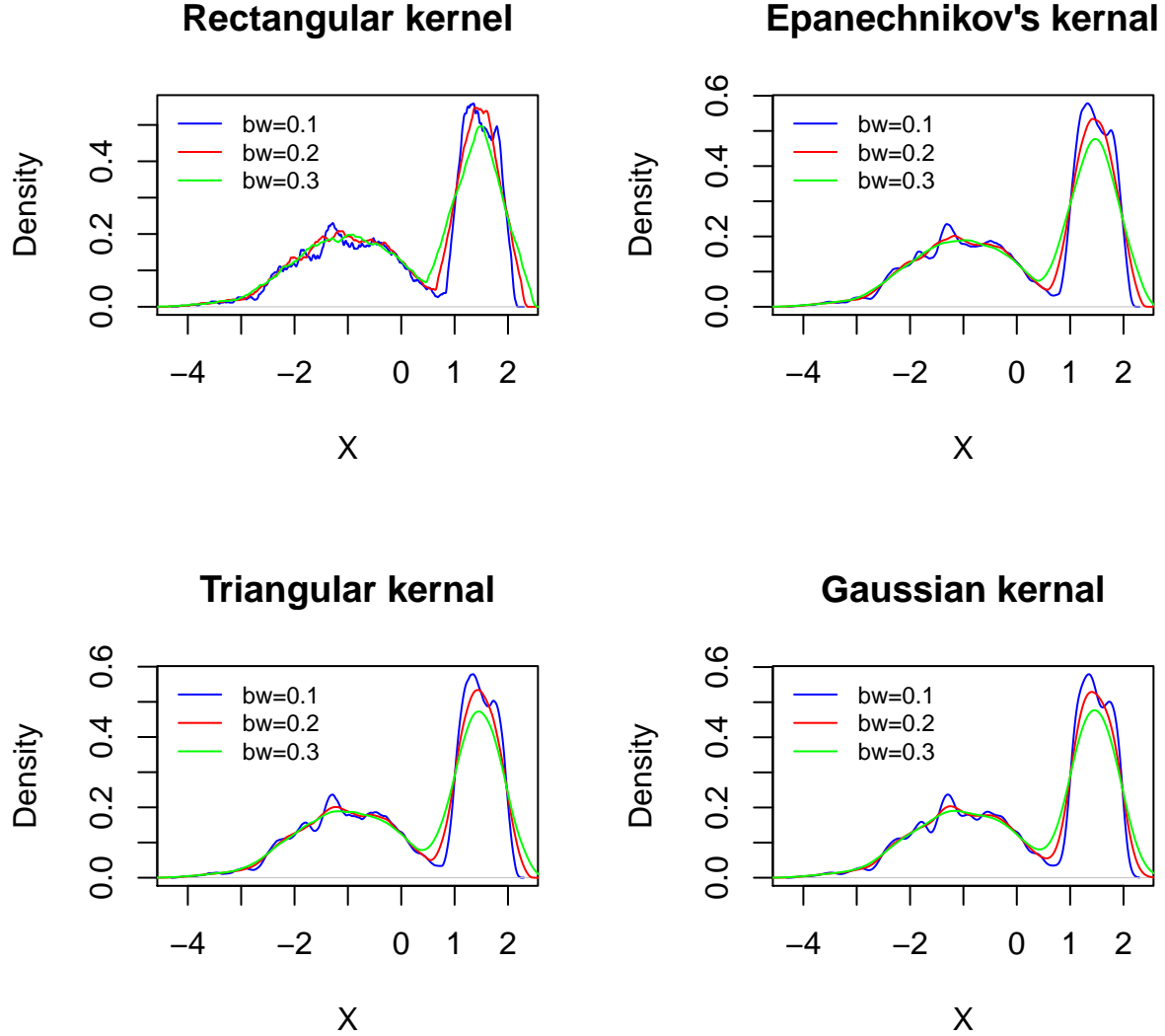
2. Influence of the choice of the kernel

Now we consider a kernel density estimator which can be represented as it follows :

$$\hat{p}_n(x) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{x - X_k}{h}\right), \quad x \in \mathbb{R}.$$

In this section we test 4 different kernels:

Type of Kernel	Kernel function
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{\{ u \leq 1\}}$
Gaussian	$K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$
Triangle	$K(u) = (1 - u)\mathbb{1}_{\{ u \leq 1\}}$
Rectangle	$K(u) = \frac{1}{2}\mathbb{1}_{\{ u \leq 1\}}$



Corollary: One can see that the rectangular KDE consists of noise for all values of h , whenever the other three kernels are giving smoother results. As a common tendency one can see that $h = 0.3$ is visually more appropriate for all the kernels.

3. Chosing optimal h by minimazing $J(h)$.

We consider the following functional

$$J(h) = \mathbb{E} \int (\hat{p}_{n,h} - p(x))^2 dx - \int p^2(x) dx.$$

One can notice that last term of the functional $J(h)$ is a constant therefore one can rewrite

$$\operatorname{argmin}_{h>0} J(h) = \operatorname{argmin}_{h>0} \mathbb{E} \left[\int \hat{p}_{n,h}^2 dx - \int 2\hat{p}_{n,h}p(x) dx \right]$$

Let $\tilde{J}(h) = \mathbb{E} \left[\int \hat{p}_{n,h}^2 dx - \int 2\hat{p}_{n,h}p(x) dx \right]$ we propose an unbiased estimator of $\tilde{J}(h)$ as it follows

$$CV = \int \hat{p}_{n,h}^2 dx - \frac{2}{n} \sum_{j=1}^n \hat{p}_{n,h}^{-j}(X_j)$$

where $\hat{p}_{n,h}^{-j}$ is a kernel density estimation of $p(x)$ without observation X_j clearly,

$$\hat{p}_{n,h}^{-j}(x) = \frac{1}{(n-1)h} \sum_{i \neq j} K\left(\frac{x - X_i}{h}\right)$$

this technique is known as cross-validation, see for instance (Tsybakov 2008, 27–31). Finally one can obtain an optimal h as

$$\hat{h} = \operatorname{argmin}_{h>0} \int \hat{p}_{n,h}^2 dx - \frac{2}{n} \sum_{j=1}^n \hat{p}_{n,h}^{-j}(X_j)$$

```
## $minimum
## [1] 0.05884854
##
## $objective
## [1] -0.3292468
```

We obtained that $\hat{h}_1 = 0.0588$

3.1. Transformed data

In this section we study an effect of two types of transformation of data. Firstly we consider a translated data, for this task we move our observations by 800. And since the support of $p(x)$ was not changed we expect to receive the same result as in previous section.

```
## $minimum
## [1] 0.05857286
##
## $objective
## [1] -0.3292469
```

We obtained that $\hat{h}_2 = 0.0586$. One can notice that $\hat{h}_1 \approx \hat{h}_2$.

Secondly we consider scaled data, for this purpose we divide our observations by 5. Since the support of $p(x)$ was changed we expect to receive h five times smaller than in previous section.

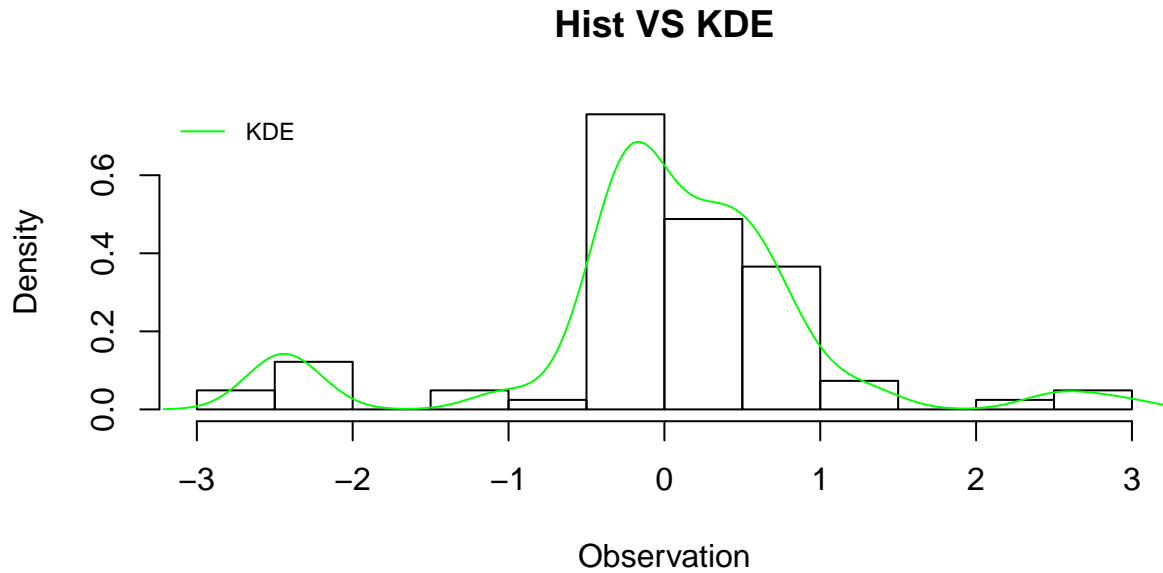
```
## $minimum
## [1] 0.01172326
##
## $objective
## [1] -1.646238
```

We obtained that $\hat{h}_3 = 0.0117$. One can notice that $\hat{h}_1 \approx \hat{h}_2 \approx 5\hat{h}_3$, which is in total agreement with our expectations.

Corollary: A translation is not changing the optimal value of h , however a rescaling by m (i.e. $X' = \frac{X}{m}$) changing the optimal value by m times.

4. Working with real data

For this part we consider a dataset Galaxies which consists of a numeric vector of velocities in km/sec of 82 galaxies from 6 well-separated conic sections of an unfilled survey of the Corona Borealis region. Multimodality in such surveys is evidence for voids and superclusters in the far universe.



Comparing the estimators: We use the histogram if we want to describe our sample, and the kernel density estimator (KDE) if we want to describe the hypothesized underlying distribution. Since we don't have any knowledge and assumptions on distribution of the Galaxies dataset we would rather use histogram than KDE.

References

Tsybakov, Alexandre B. 2008. *Introduction to Nonparametric Estimation*. Springer.