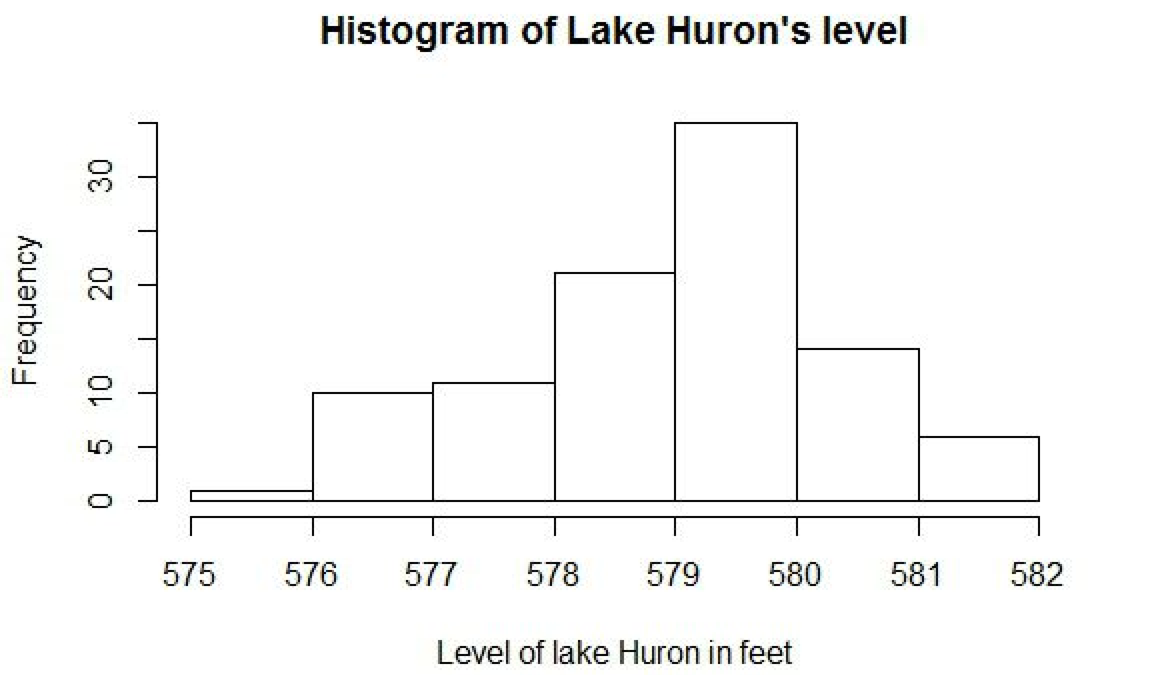
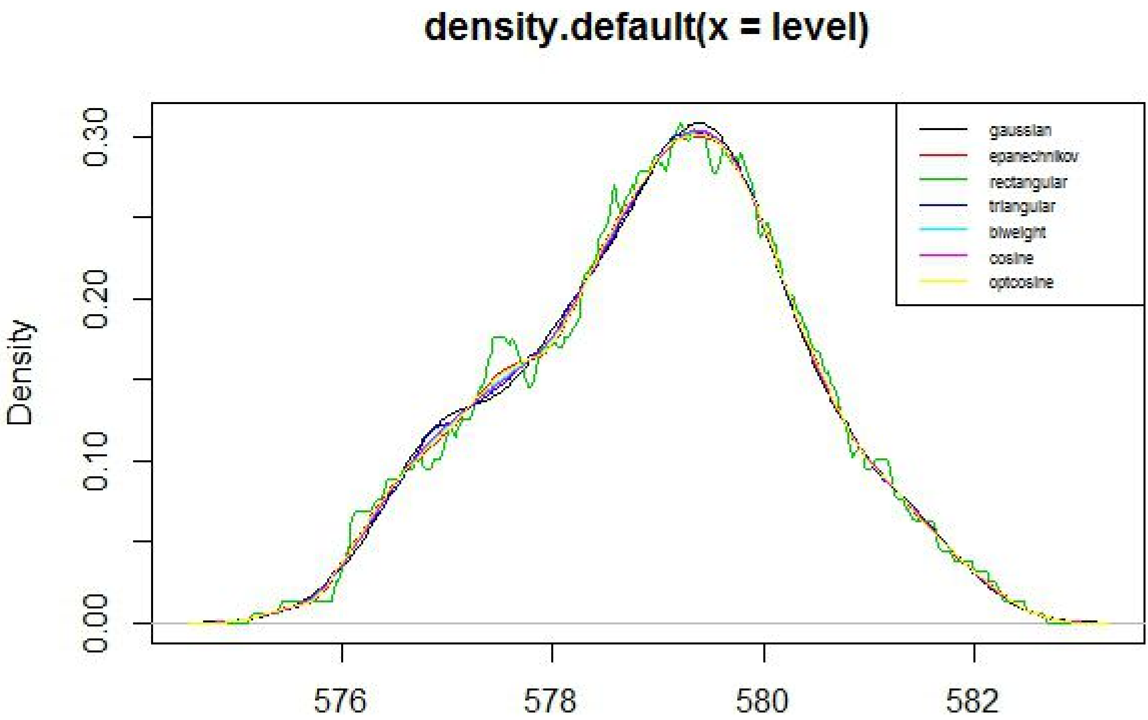
```
# 1: Constructing the histogram estimator
level <- LakeHuron
hist(level, main = paste("Histogram of" , "Lake Huron's level"), xlab = "Level of lake Huron in feet")
# As "Sturges" is a default method for hist()
```

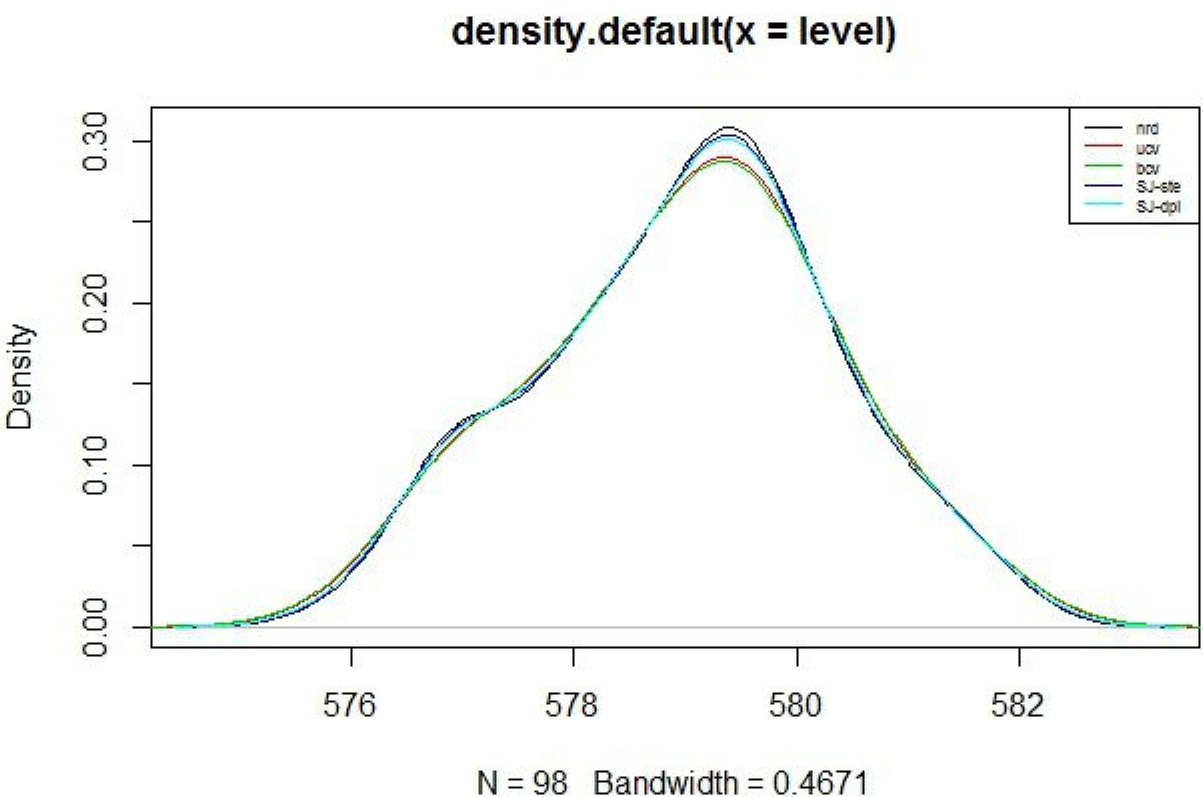


Histogram of Lake Huron's level
Frequency
0
5
10
20
30
575
576
577
578
579
580
581
582
Level of lake Huron in feet

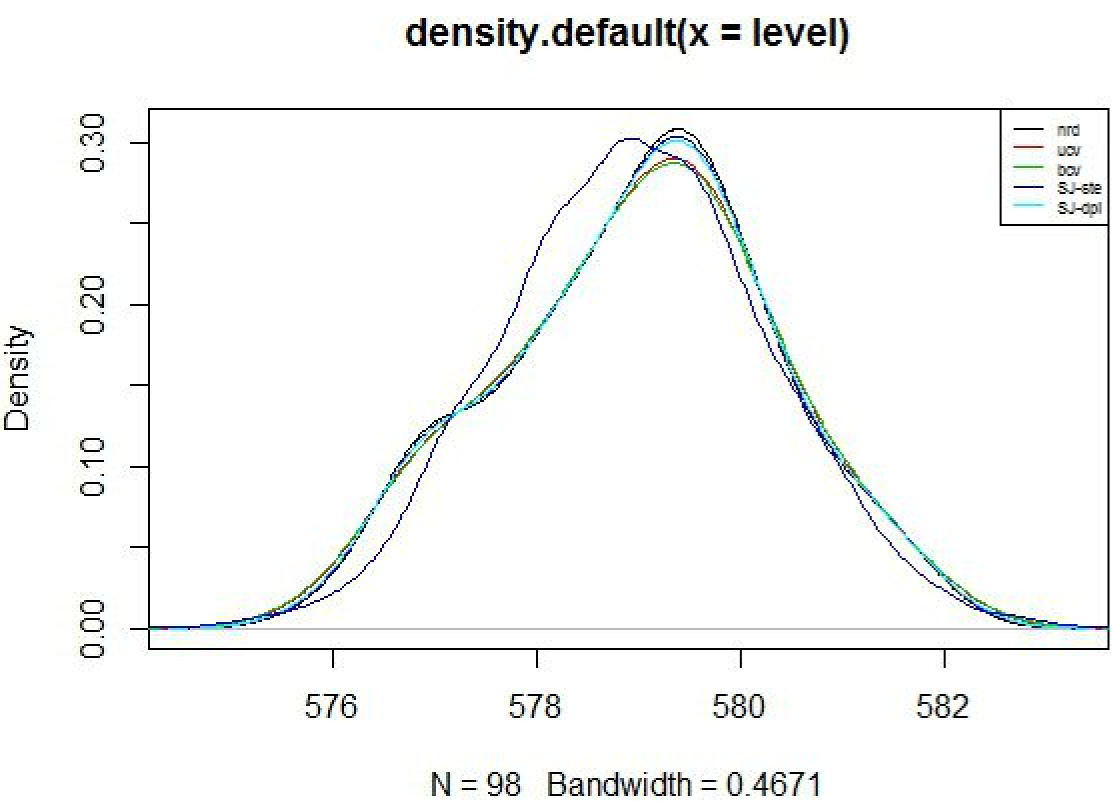
```
# 2: Constructing the kernel estimators
K <- eval(formals(density.default)$kernel)
plot(density(level), xlab="")
for (i in (2:length(K))){
  lines(density(level, kernel=K[i]), col=i) # col=i from lection
}
legend("topright", 0.25, legend=K, col=seq(K), lty=1, cex=0.5)
# Kernel's list: like here https://stat.ethz.ch/R-manual/R-devel/library/stats/html/density.html
# About legend: https://stackoverflow.com/questions/5520637/legend-in-base-r-can-fill-refrain-from
```


density.default(x = level)
Density
0.00
0.10
0.20
0.30
576
578
580
582
gaussian
epanechnikov
rectangular
triangular
biweight
cosine
optcosine

```r
# 3: Constructing the kernel estimators with different bandwidth parameters h
plot(density(level)) # Default bw='nrd0' would be selected here
h <- c('nrd', 'ucv', 'bcv', 'SJ-ste', 'SJ-dpi') # all possible h
for (i in (2:length(h))){
  lines(density(level, bw=h[i]), col=i) # same as №2
}
legend("topright", 0.25, legend=h, col=seq(h), lty=1, cex=0.5)
```

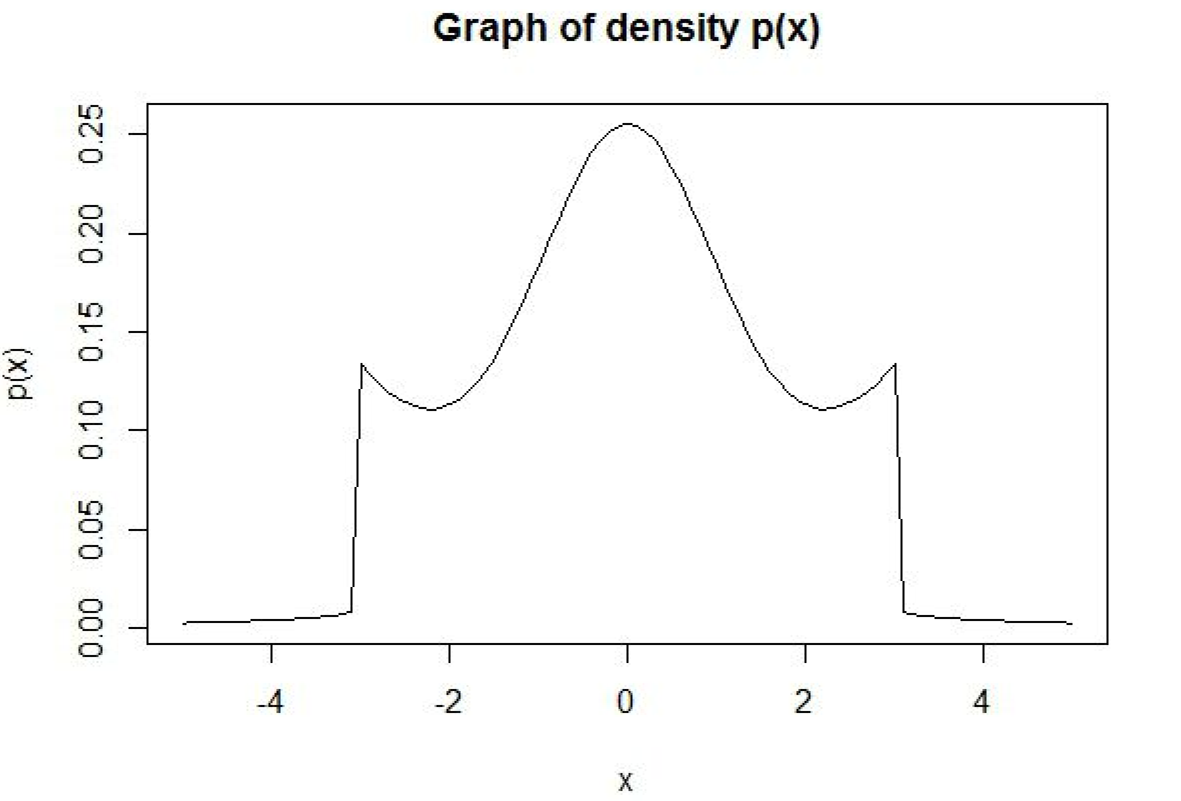

density.default(x = level)
nrd
ucv
bcv
SJ-ste
SJ-dpi
Density
0.00
0.10
0.20
0.30
576
578
580
582
N = 98   Bandwidth = 0.4671

```r
# 4: Adding pdf of N(0,1)
plot(density(level)) # Default bw='nrd0' would be selected here
h <- c('nrd', 'ucv', 'bcv', 'SJ-ste', 'SJ-dpi') # all possible h
for (i in (2:length(h))){
  lines(density(level, bw=h[i]), col=i) # same as №3
}
legend("topright", 0.25, legend=h, col=seq(h), lty=1, cex=0.5)
lines(density(rnorm(10000, mean=mean(level), sd=sd(level))), col=12)
# rnorm method as it was in classes
```



density.default(x = level)
nrd
ucv
bcv
SJ-ste
SJ-dpl
Density
0.00
0.10
0.20
0.30
576
578
580
582
N = 98 Bandwidth = 0.4671

```r
# 1: Displaying the graph of the function p(x)
p = function(x){
  p <- 0.5*dnorm(x) + 0.25*dchisq(x+3, df=2) + 0.25*dchisq(-x+3, df=2)
  return(p) # definition of function p(x) with dnorm & dchisq
}
S <- seq(-5, 5, by=0.1) # задаем ее только в точках вида a,b.
D <- c(rep(0, 100)) # вектор из нулей, в который потом записываются значения
for (i in (1:length(S))){ # плотности в соответствующих точках
  D[i] <- p(S[i]) # Это достаточно хорошо приближает реальную картину.
}
plot(S, D, type='l', main=paste("Graph of density p(x)"), xlab="x", ylab="p(x)")
# не путаем оси. Кстати, среднее действительно 0.
```



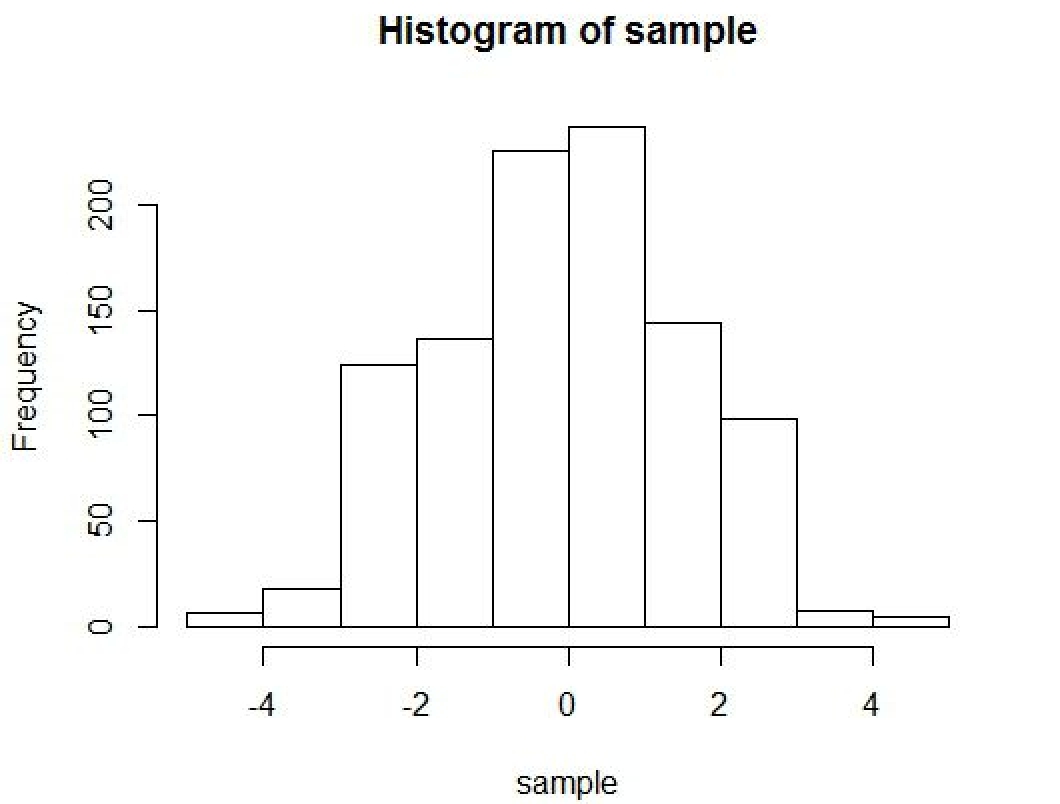
Graph of density p(x)
0.00
0.05
0.10
0.15
0.20
0.25
p(x)
-4
-2
0
2
4
x

```
# 2: Simulating a sample of length N = 1000
sample <- approx(cumsum(D)/sum(D), S, runif(1000))$y # опять же, создаем по S и D
print(head(sample, n=7)) # посмотрим 7 первых элементов
print(tail(sample, n=7)) # и посмотрим 7 последних
```

```
> print(head(sample, n=7)) # посмотрим 7 первых элементов
[1] -1.48287674  1.42126830  0.07474735 -1.36495987  1.77699368  2.41625194
[7] -1.46558127
> print(tail(sample, n=7)) # и посмотрим 7 последних
[1]  1.14133938 -1.32143474  2.75502801 -0.21843709 -0.07148242 -2.70182474
[7] -0.28691945
```

```
# 3: Constructing the histogram estimator and calculating empirical MISE
res <- hist(sample) # 'Sturges' is a default method
middles <- res$mids # из гистограммы находим центры
Q <- 10000 # по условию
s <- seq(-4, 4, by=8/Q) # длина отрезка - 8 - дробится на Q частей
MISE_emp <- 0 # значение на старте
for (i in (1:length(s))){ # для всех Q элементов смотрим, какой вклад они вносят в MISE_emp
  p_n <- res$density[which(abs(middles-s[i])==min(abs(middles-s[i])))[1]]
  p_real <- p(s[i])
  MISE_emp <- MISE_emp + (p_n-p_real)**2/Q # Все складываем и накапливаем
}
print(MISE_emp) # полученная величина эмпирического MISE
```

```
> print(MISE_emp) # полученная величина эмпирического MISE
[1] 0.0003388706
>
```


Histogram of sample
Frequency
0
50
100
150
200
-4
-2
0
2
4
sample

Новиков Лев. НА-1.

"Оценивание плотности".

Теоретическая часть:

T1. $p(x) = \frac{1}{2}\phi^{N}(x) + \frac{1}{4}\phi^{\chi^2_2}(x+3) + \frac{1}{4}\phi^{\chi^2_2}(-x+3)$

i) Во-первых, $\forall x \; p(x) \geq 0$ (т.к. $\phi^{N}(x)$ и $\phi^{\chi^2_2}(x)$ — плотности).

Во-вторых, $\int\limits_{\mathbb{R}} p(x)\,dx = \int\limits_{\mathbb{R}} \Big(\frac{1}{2}\phi^{N}(x) + \frac{1}{4}\phi^{\chi^2_2}(x+3) + \frac{1}{4}\phi^{\chi^2_2}(-x+3)\Big)dx = \frac{1}{2}\int\limits_{\mathbb{R}}\phi^{N}(x)dx + \frac{1}{4}\int\limits_{\mathbb{R}}\phi^{\chi^2_2}(x+3)\,dx + \frac{1}{4}\int\limits_{\mathbb{R}}\phi^{\chi^2_2}(-x+3)\,dx = \frac{1}{2}\cdot 1 + \frac{1}{4}\int\limits_{\mathbb{R}}\phi^{\chi^2_2}(t)\,dt + \frac{1}{4}\int\limits_{\mathbb{R}}\phi^{\chi^2_2}(u)\cdot(-du) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 1 + \frac{1}{4}\cdot 1 = 1$, нормировочное условие выполнено.

Следовательно, $p(x)$ — плотность.

ii) Для $X \sim p(x)$ найти EX и $V(X)$.

$EX = \int\limits_{\mathbb{R}} x p(x)\,dx = \int\limits_{\mathbb{R}} x\Big(\frac{1}{2}\phi^{N}(x) + \frac{1}{4}\phi^{\chi^2_2}(x+3) + \frac{1}{4}\phi^{\chi^2_2}(-x+3)\Big)dx = \frac{1}{2}\int\limits_{\mathbb{R}} x\phi^{N}(x)\,dx + \frac{1}{4}\int\limits_{\mathbb{R}} x\phi^{\chi^2_2}(x+3)\,dx + \frac{1}{4}\int\limits_{\mathbb{R}} x\phi^{\chi^2_2}(-x+3)\,dx$. $\phi^{\chi^2_2}(x) = \frac{1}{2}e^{-\frac{x}{2}}\cdot I_{\{x\geq 0\}}$.

(это, кстати, совпадает с плотностью распр-я $Exp(\frac{1}{2})$). Получаем, что: $\ldots =$

$$= \frac{1}{2}\mathbb{E}Z + \frac{1}{4}\int_{\mathbb{R}} x\cdot\frac{1}{2}e^{-\frac{x+3}{2}}\cdot I_{\{x+3\geq 0\}}dx + \frac{1}{4}\cdot\int_{\mathbb{R}} x\cdot\frac{1}{2}\cdot$$

$$\cdot e^{-\frac{(-x+3)}{2}}\cdot I_{\{-x+3\geq 0\}}dx = \frac{1}{2}\cdot 0 + \frac{1}{8}\int_{-3}^{\infty} xe^{-\frac{x+3}{2}}dx +$$

$$+ \frac{1}{8}\int_{-\infty}^{3} xe^{-\frac{(-x+3)}{2}}dx = \frac{1}{8}\cdot\int_0^{\infty}(t-3)e^{-\frac{t}{2}}d(t-3) +$$

$$+ \frac{1}{8}\int_{\infty}^{0}(3-u)e^{-\frac{u}{2}}du\cdot(-1) \quad \left(\begin{matrix}\text{Замены в интегралах} \\ t = x+3 \;\; \text{и} \;\; u = -x+3,\; du = -dx\end{matrix}\right)$$

$$= \frac{1}{8}\int_0^{\infty}(t-3)e^{-\frac{t}{2}}dt + \frac{1}{8}\int_0^{\infty}(3-u)e^{-\frac{u}{2}}du = 0$$

Т.е. $\mathbb{E}X = 0$.

$$V(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X^2 = \int_{\mathbb{R}} x^2 p(x)dx =$$

$$= \mathbb{E}Z^2\cdot\frac{1}{2} + \frac{1}{8}\cdot\int_{-3}^{\infty} x^2 e^{-\frac{x+3}{2}}dx + \frac{1}{8}\int_{-\infty}^{3} x^2 e^{-\frac{(-x+3)}{2}}dx =$$

$$= \frac{1}{2}\cdot V(Z) + \frac{1}{8}\int_0^{\infty}(t-3)^2 e^{-\frac{t}{2}}dt + \frac{1}{8}\int_0^{\infty}(3-t)^2 e^{-\frac{t}{2}}dt =$$

$$= \frac{1}{2} + \frac{1}{4}\cdot\int_0^{\infty}(t-3)^2 e^{-\frac{t}{2}}dt = \frac{1}{2} + \frac{1}{4}\int_0^{\infty}(t-3)^2 d(-2e^{-t/2}) =$$

$$= \frac{1}{2} + \frac{1}{2}(t-3)^2 e^{-t/2}\Big|_{\infty}^{0} + \frac{1}{2}\cdot\int_0^{\infty} e^{-\frac{t}{2}}d(t-3)^2 = 5 + \frac{1}{2}\cdot\int_0^{\infty}2(t-3)e^{-\frac{t}{2}}dt$$

$$= 5 + 2\cdot\int_0^{\infty} t\cdot\frac{1}{2}e^{-\frac{t}{2}}dt - 6\cdot\int_0^{\infty}\frac{1}{2}e^{-\frac{t}{2}}dt = 5 + 2\cdot 2 - 6\cdot 1 = 3$$

Т.е. $V(X) = 3$. Ответ: Для $X \sim p(x)$, $\begin{matrix}\mathbb{E}X = 0 \\ V(X) = 3\end{matrix}$.

T2. $\{X_i\}_{i=1}^{n} \sim N(0, \sigma^2)$

i) Найти оптимальное значение bandwidth (h, сглаживающего параметра)

Для гистограмм оценка $\hat{f}_n(x) = \frac{1}{nh} \cdot \sum\limits_{i=1}^{n} I_{\{x_i \in B_j\}}$,

где $B_j = [c_j - \frac{h}{2}; c_j + \frac{h}{2})$.

Нам известно, что для $f(x)$ – истинной ф-ии плотности, если $f \in C^2$, и $f' \in L^2$, то

$$MISE(\hat{f}_n(x)) = \int\limits_{\mathbb{R}} \mathbb{E}(\hat{f}_n(x) - f(x))^2 dx =$$

$$= \frac{h^2}{12} \cdot \int\limits_{\mathbb{R}} (f'(x))^2 dx + \frac{1}{nh} + o(1)$$, при $n \to \infty$, $h \to 0$.

$\Rightarrow$ Для $G(h) = \frac{1}{nh} + \frac{h^2}{12} \cdot \int\limits_{\mathbb{R}} (f'(x))^2 dx$ мы решаем задачу $G(h) \longrightarrow \min\limits_{h}$ и ищем h_{opt}.

Отсюда FOC: $G'(h) = 0 \Rightarrow$

$$\Rightarrow -\frac{1}{nh^2} + \frac{2h}{12} \cdot \int\limits_{\mathbb{R}} (f'(x))^2 dx = 0 \Rightarrow \frac{h}{6} \cdot \int\limits_{\mathbb{R}} (f'(x))^2 dx = \frac{1}{nh^2}$$

$\Rightarrow h^3 = \frac{6}{n \cdot \int\limits_{\mathbb{R}} (f'(x))^2 dx}$, откуда $h_{opt} = \left(\frac{6}{n \cdot \|f'\|_2^2}\right)^{1/3}$

В нашем случае $f(x) = (2\pi\sigma^2)^{-1/2} \cdot e^{-x^2/2\sigma^2} \Rightarrow$

$\Rightarrow f'(x) = -\frac{1}{\sqrt{2\pi}\sigma^3} \cdot e^{-x^2/2\sigma^2} \cdot x$. Следовательно:

$$h_{opt} = \left(\frac{6}{n\cdot\|f'\|_2^2}\right)^{1/3} = \left(\frac{6}{n}\right)^{1/3}\cdot\left(\int_{\mathbb{R}} \frac{x^2}{2\pi\sigma^6}\cdot e^{-\frac{x^2}{\sigma^2}}dx\right)^{-1/3} =$$

$$= \left(\frac{6}{n}\right)^{1/3}\cdot\left(\frac{1}{\sqrt{\pi}\cdot 4\sigma^3}\right)^{-1/3} = \left(\frac{24\sqrt{\pi}\,\sigma^3}{n}\right)^{1/3} = \left(\frac{3\sqrt{\pi}}{n}\right)^{1/3}\cdot 2\sigma$$

Ответ: $h_{opt} = \left(\frac{3\sqrt{\pi}}{n}\right)^{1/3}\cdot 2\sigma$.

ii) $K(x) = (2\pi)^{-1/2}\cdot e^{-x^2/2}$. $h_{opt} = ?$

Опять же, т.к. $K(x)$ – плотность (для $N(0,1)$), четна, и лежит в L^2, $f\in C^2$, и $f''\in L^2$, имеем (результат из лекций для ядерн. оценок)

$$MISE(\hat{f}_n(x)) \approx \frac{1}{nh}\cdot\int_{\mathbb{R}} K^2(u)\,du + \frac{h^4}{4}\left(\int_{\mathbb{R}} x^2\cdot K(x)\,dx\right)^2\cdot\int_{\mathbb{R}}(f''(x))^2\,dx$$

Решаем задачу $G(h) = MISE(\hat{f}_n(x)) \longrightarrow \min\limits_h$.

$$G'(h) = -\frac{1}{nh^2}\cdot\int_{\mathbb{R}} K^2(u)\,du + h^3\cdot\left(\int_{\mathbb{R}} x^2\cdot K(x)\,dx\right)^2\cdot\int_{\mathbb{R}}(f''(x))^2\,dx = 0,$$

Откуда $h_{opt} = \left(\frac{\|K\|_2^2}{n\cdot\|f''\|_2^2}\right)^{1/5}\cdot\left(\int_{\mathbb{R}} x^2 K(x)\,dx\right)^{-2/5}$

$\|K\|_2^2 = \int_{\mathbb{R}}(2\pi)^{-1}\cdot e^{-x^2}dx = \frac{1}{2\sqrt{\pi}}$ (гауссовский интеграл)

$\int_{\mathbb{R}} x^2\cdot K(x)\,dx = \mathbb{E}Z^2 = 1$ (второй момент для $X\sim N(0,1)$)

$$f''(x) = \frac{\partial}{\partial x}\left(-\frac{x}{\sqrt{2\pi}\,\sigma^3}\cdot e^{-\frac{x^2}{2\sigma^2}}\right) = \frac{x^2-\sigma^2}{\sqrt{2\pi}\,\sigma^5}\cdot e^{-\frac{x^2}{2\sigma^2}}$$

$\|f''(x)\|_2^2 = \int_{\mathbb{R}}\frac{(x^2-\sigma^2)^2}{2\pi\cdot\sigma^{10}}\cdot e^{-\frac{x^2}{\sigma^2}}dx \overset{\text{Wolfram}}{=\!=} \frac{3}{8\sqrt{\pi}\cdot\sigma^5}$. Значит,

$$h_{opt} = \left(\frac{\frac{1}{2\sqrt{\pi}}}{\frac{3n}{8\sqrt{\pi}\sigma^5}} \right)^{1/5} \cdot 1^{-2/5} = \left(\frac{4\sigma^5}{3n} \right)^{1/5} = \left(\frac{4}{3n} \right)^{1/5} \cdot \sigma$$

Ответ: $h_{opt} = \left(\frac{4}{3n} \right)^{1/5} \cdot \sigma$

iii) Проанализировать, как в AMISE зависят от σ часть, отвечающая за смещение, и часть, отвечающая за дисперсию.

а) В случае с гистограммой:

Var: $\frac{1}{nh_{opt}} = \frac{1}{2n\sigma} \cdot \left(\frac{n}{3\sqrt{\pi}} \right)^{1/3} = \frac{1}{(24n^2\sqrt{\pi})^{1/3}} \cdot \sigma^{-1} =$

$= (24n^2 \cdot \sqrt{\pi})^{-1/3} \cdot \sigma^{-1}$. Т.е чем больше σ, тем меньше часть, отвечающая за дисперсию.

Bias^2: $\frac{h_{opt}^2}{12} \cdot \int_{\mathbb{R}} (f'(x))^2 dx = \frac{1}{12} \cdot \left(\frac{3\sqrt{\pi}}{n} \right)^{2/3} (2\sigma)^2 \cdot (4\sqrt{\pi}\sigma^3)^{-1} =$

$= \frac{1}{4(3\sqrt{\pi}n^2)^{1/3}} \cdot \sigma^{-1}$ (чем больше σ, тем меньше часть, отвечающая за смещение)

б) В случае с ядерными оценками:

Var: $\frac{1}{nh_{opt}} \cdot \int_{\mathbb{R}} K^2(x) dx = \left(\frac{3}{4n^4} \right)^{1/5} \cdot \sigma^{-1} \cdot \frac{1}{2\sqrt{\pi}} = \left(\frac{3}{4n^4} \right)^{1/5} \cdot \frac{1}{2\sqrt{\pi}} \cdot \sigma^{-1}$

(Чем больше σ, тем меньше часть AMISE, отвечающая за дисперсию)

$$\text{Bias}^2: \frac{h_{opt}^4}{4}\cdot\left(\int_{\mathbb{R}} x^2\cdot K(x)\,dx\right)^2\cdot\left(\int_{\mathbb{R}}(f''(x))^2dx\right) = \frac{1}{4}\cdot\left(\frac{4}{3n}\right)^{4/5}\sigma^4\cdot$$

$$\cdot 1\cdot\frac{3}{8\sqrt{\pi}\cdot\sigma^5} = \left(\frac{3}{4}\right)^{1/5}\cdot\frac{1}{8\sqrt{\pi}\cdot n^{4/5}}\cdot\sigma^{-1}$$

Опять же, чем больше σ, тем меньше часть AMISE, отвечающая за смещение.

Ответ: Чем больше σ, тем меньше AMISE (и Var, и Bias - части).

T3. Рассчитать теоретическую эффективность

i) boxcar. $(\mathrm{eff}(K))^{-1} = \left(\left(\frac{\int_{\mathbb{R}} K^2(x)\,dx}{\int_{\mathbb{R}} K_{Ep}^2(x)\,dx}\right)^{4/5}\cdot\left(\frac{\int_{\mathbb{R}} x^2\cdot K(x)\,dx}{\int_{\mathbb{R}} x^2 K_{Ep}(x)\,dx}\right)^{2/5}\right)^{5/4}$

(здесь $K_{Ep}(x)$ — Епанечнико ядро)

Для ядра boxcar: $K(x) = \frac{1}{2}\cdot I_{\{|x|<1\}} \Rightarrow$

$$\Rightarrow (\mathrm{eff}(K))^{-1} = \frac{\int_{-1}^{1}\frac{1}{4}dx}{\int_{-1}^{1}\left(\frac{3}{4}(1-x^2)\right)^2dx}\cdot\left(\frac{\int_{-1}^{1}\frac{x^2}{2}dx}{\int_{-1}^{1}\frac{3x^2}{4}(1-x^2)dx}\right)^{1/2} =$$

$$= \frac{\frac{1}{2}\cdot\left(\frac{2}{6}\right)^{1/2}}{\frac{9}{16}\cdot\int_{-1}^{1}(1-2x^2+x^4)dx\cdot\left(\frac{3}{4}\cdot\left(\frac{x^3}{3}-\frac{x^5}{5}\right)\Big|_{-1}^{1}\right)^{1/2}} = \frac{\frac{1}{2}\cdot\left(\frac{1}{3}\right)^{1/2}}{\frac{9}{16}\left(2-\frac{4}{3}+\frac{2}{5}\right)\cdot\left(\frac{2}{3}-\frac{2}{5}\right)^{1/2}\cdot\left(\frac{3}{4}\right)^{1/2}} =$$

$$= \frac{1/2\cdot(1/3)^{1/2}}{\frac{9}{16}\cdot\frac{16}{15}\cdot\left(\frac{4}{15}\right)^{1/2}\cdot\left(\frac{3}{4}\right)^{1/2}} = \frac{1/2}{\frac{3}{15}\cdot\frac{3}{\sqrt{15}}} = \frac{5\sqrt{15}}{18}$$

$$\Rightarrow \mathrm{eff}(K) = \frac{18}{5\sqrt{15}} = \frac{6\sqrt{15}}{25} \simeq 0.929516$$

ii) Гауссовское ядро: $K(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$.

$$\mathrm{eff}(K) = \frac{\int_{\mathbb{R}} K_{Ep}^2(x)\,dx}{\int_{\mathbb{R}} K^2(x)\,dx} \cdot \left(\frac{\int_{\mathbb{R}} K_{Ep}(x) \cdot x^2\,dx}{\int_{\mathbb{R}} K(x) \cdot x^2\,dx} \right)^{1/2} =$$

$$= \frac{\int_{\mathbb{R}} K_{Ep}^2(x)\,dx \cdot \left(\int_{\mathbb{R}} K_{Ep}(x) \cdot x^2\,dx\right)^{1/2}}{\int_{\mathbb{R}} \frac{1}{2\pi} e^{-x^2}\,dx \cdot \left(\int_{\mathbb{R}} \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2}\,dx\right)^{1/2}}.$$

$$\int_{\mathbb{R}} \frac{1}{2\pi} e^{-x^2}\,dx = \frac{1}{2\sqrt{\pi}}. \quad \int_{\mathbb{R}} \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \int_{\mathbb{R}} x^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}\,dx$$

$$= \mathbb{E}\xi^2 = 1.$$

Получаем $\mathrm{eff}(K) = \dfrac{\frac{9\sqrt{3}}{15\sqrt{15}}}{\frac{1}{2\sqrt{\pi}} \cdot 1} = \dfrac{18\sqrt{\pi}}{15\sqrt{5}} = \dfrac{6\sqrt{5\pi}}{25}$

$\approx 0.951199.$

Ответ: $\mathrm{eff}(K_{Boxcar}) \simeq 0{,}929516 \simeq 92{,}95\%$.

$\mathrm{eff}(K_{Gaussian}) \simeq 0{,}951199 \simeq 95{,}12\%$.