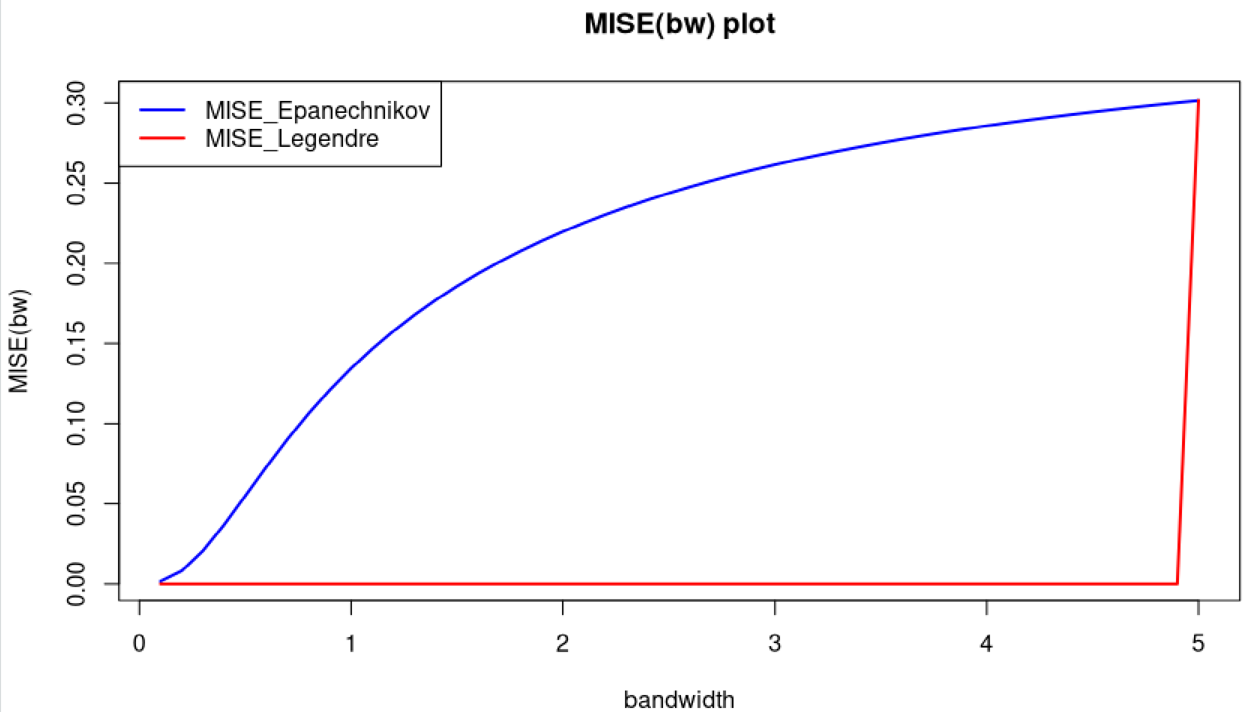
N1

```
# task N1

n = 1000
shape = 3
rate = 4

# (i)
# моделируем выборку
X = rgamma(n, shape = shape, rate = rate)

# (ii)
# значения bandwidth
bws = seq(from = 0.1, to = 5, by = 0.1)
MISE_E = rep(0, length(bws))

# x из отрезка [0, 2]
xms = seq(from = 0, to = 2, by = 0.01)
ps = rep(0, length(xms)) # истинные значения плотности
pns = rep(0, length(xms)) # значения плотности с ядром Епанечникова
dps = rep(0, length(xms)) # разность значений плотности

for(j in 1:length(bws)){
  pns = density(X, n = 201, from=0, to=2, bw=bws[j])$y
  for(i in 1:length(ps)){
    ps[i]=dgamma(xms[i], shape=shape, rate=rate)
    dps[i] = (ps[i] - pns[i])^2
  }
  MISE_E[j] = mean(dps)
}

bws[which.min(MISE_E)] # минимум MISE при bandwidth = 0.1
```

```
# (iii)
# ядро с  первыми 3 многочленами Лежандра на отрезке [-1, 1]
K = function(x){
  9/8 - 15/8 * x^2
}

pns2 = rep(0, length(xms)) # значения плотности с ядром из задания
ks = rep(0, length(X)) # значения плотности с ядром из задания
MISE_L = rep(0, length(bws))

for (a in length(bws)){
  for (i in length(pns2)){
    xi = xms[i]
    n = 0
    for (j in length(ks)){
      dlt = (X[j] - xi) / bws[a]
      if (dlt > 1 | dlt < -1){
        ks[j] = 0
      } else {
        ks[j] = K(dlt)
        n = n + 1
      }
    }
    pns2[i] = sum(ks) / (bws[a] * n)
    print(sum(ks))
    dps[i] = (ps[i] - pns2[i])^2
  }
  MISE_L[a] = mean(dps)
}

bws[which.min(MISE_L)] # минимум MISE при bandwidth = 0.1

# Графически сравниваем оценки MISE с ядром Епанечникова и с ядром (2),
# в зависимости от значения параметра bandwidth.

plot(bws, MISE_E, type='l', col = 'blue', lwd = 2,
     main="MISE(bw) plot", xlab="bandwidth", ylab="MISE(bw)")
lines(bws, MISE_L, col = 'red', lwd = 2)
legend(x = 'topleft', col = c('blue', 'red'), lwd = 2,
       legend = c('MISE_Epanechnikov', 'MISE_Legendre'))
```


MISE(bw) plot
MISE_Epanechnikov
MISE_Legendre
MISE(bw)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
4
5
bandwidth

T1

T1 Для оптимального bandwidth и ядра Епанечникова воспользуемся формулой $MISE(\hat{p}_n) = \frac{3^{4/5}}{5^{1/5}\cdot 4}\cdot\left(\int (p''(x))^2 dx\right)^{1/5}$ [1.27, Цыбаков].

В задании используем выборку с гамма-распределением $\Gamma(k,\theta)=\Gamma(3,\frac{1}{4})$.

Тогда $p(x)=\frac{1}{\Gamma(k)\cdot\theta^k}x^{k-1}\cdot e^{-\frac{x}{\theta}} = \frac{4^3}{\Gamma(3)}\cdot x^2\cdot e^{-4x} = \frac{2^6}{2}x^2e^{-4x} = 2^5\cdot x^2\cdot e^{-4x}$

Найдём 2ю производную $p''(x) = 2^5\left(2x\cdot e^{-4x}-4x^2\cdot e^{-4x}\right)' = 2^6\left(e^{-4x}-4x\cdot e^{-4x}-\right.$

$\left.-4x\cdot e^{-4x}+8x^2e^{-4x}\right) = 2^6\cdot e^{-4x}\cdot(1-8x+8x^2)$

Вычислим отдельно интеграл квадрата второй производной плотности. Пределами интегрирования будут, согласно области определения x для гамма-распределения, $x\in[0;+\infty)$.

$$\int_0^{+\infty}(p''(x))^2dx = \int_0^{+\infty}\left(2^6\cdot e^{-4x}\cdot(1-8x+8x^2)\right)^2dx = 2^{12}\int_0^{+\infty}e^{-8x}\cdot(1-8x+8x^2-8x+8^2x^2-8^2x^3+$$

$$+8x^2-8^2x^3+8^2x^4)\,dx = 2^{12}\int_0^{+\infty}e^{-8x}\cdot(1-2^4x+2^4\cdot 5\cdot x^2-2^7\cdot x^3+2^6\cdot x^4)dx = \left[\begin{array}{l} y=-8x\ ;\ dx=-\frac{1}{8}dy \\ x=-\frac{1}{8}y\ ;\ dy=-8dx\end{array}\right] =$$

$$=-2^9\int_0^{-\infty}e^y\cdot\left(1+2y+\frac{5}{4}y^2+\frac{1}{4}y^3+\frac{1}{2^6}y^4\right)dy = \left[\begin{array}{l} u=(1+2y+\frac{5}{4}y^2+\frac{1}{4}y^3+\frac{1}{2^6}y^4) \Rightarrow du=(2+\frac{5}{2}y+\frac{3}{4}y^2+\frac{1}{16}y^3)dy \\ dv=e^ydy \Rightarrow v=\int e^ydy=e^y\end{array}\right]=$$

$$=-2^9\left(e^y\left(1+2y+\frac{5}{4}y^2+\frac{1}{4}y^3+\frac{1}{64}y^4\right)-\int_0^{-\infty}e^y\left(2+\frac{5}{2}y+\frac{3}{4}y^2+\frac{1}{16}y^3\right)dy\right) = \left[\begin{array}{l} u=(2+\frac{5}{2}y+\frac{3}{4}y^2+\frac{1}{16}y^3)\Rightarrow \\ \Rightarrow du=(5/2+3/2y+3/16y^2)dy \\ dv=e^ydy\Rightarrow v=e^y\end{array}\right]=$$

$$=-2^9\left(e^y\left(1+2y+\frac{5}{4}y^2+\frac{1}{4}y^3+\frac{1}{64}y^4\right)-e^y\left(2+\frac{5}{2}y+\frac{3}{4}y^2+\frac{1}{16}y^3\right)+\int_0^{-\infty}e^y\left(\frac{5}{2}+\frac{3}{2}y+\frac{3}{16}y^2\right)dy\right) = \left[\begin{array}{l} u=(\frac{5}{2}+\frac{3}{2}y+\frac{3}{16}y^2)\Rightarrow \\ \Rightarrow du=(3/2+3/8y)dy \\ dv=e^ydy\Rightarrow v=e^y\end{array}\right]=$$

$$=-2^9\left(e^y\left(1+2y+\frac{5}{4}y^2+\frac{1}{4}y^3+\frac{1}{64}y^4-2-\frac{5}{2}y-\frac{3}{4}y^2-\frac{1}{16}y^3+\frac{5}{2}+\frac{3}{2}y+\frac{3}{16}y^2\right)-\int_0^{-\infty}e^y\left(\frac{3}{2}+\frac{3}{8}y\right)dy\right) = \left[\begin{array}{l} u=(\frac{3}{2}+\frac{3}{8}y)\Rightarrow \\ \Rightarrow du=3/8\,dy \\ dv=e^ydy\Rightarrow v=e^y\end{array}\right]=$$

$$=-2^9\left(e^y\left(\frac{1}{64}y^4+\frac{3}{16}y^3+\frac{11}{16}y^2+y+\frac{3}{2}\right)-e^y\left(\frac{3}{2}+\frac{3}{8}y\right)+\int_0^{-\infty}e^y\cdot\frac{3}{8}dy\right) = -2^9\left(e^y\left(\frac{1}{64}y^4+\frac{3}{16}y^3+\frac{11}{16}y^2+y+\frac{3}{2}-\right.\right.$$

$$\left.\left.-\frac{3}{2}-\frac{3}{8}y+\frac{3}{8}\right)\right)\Big|_0^{-\infty} = -2^9\cdot e^y\cdot\left(\frac{1}{64}y^4+\frac{3}{16}y^3+\frac{11}{16}y^2+\frac{5}{8}y+\frac{3}{8}\right)\Big|_0^{-\infty} = \left[\begin{array}{l}\text{обратная замена} \\ y=-8x\end{array}\right] =$$

$$=-2^9\cdot e^{-8x}\left(64x^4-3\cdot 2^5x^3+11\cdot 4\cdot x^2-5x+\frac{3}{8}\right)\Big|_0^{+\infty} = \left[\begin{array}{l}\text{заменим} \\ \text{верхний предел} \\ \text{интегрирования} \\ \text{на } b\end{array}\right] = \lim_{b\to+\infty}\left(-2^9\cdot\left(e^{-8b}\cdot\left(64b^4-\right.\right.\right.$$

$$\left.\left.\left.-3\cdot 2^5\cdot b^3+11\cdot 4\cdot b^2-5b+\frac{3}{8}\right)\right)\right)+2^9\cdot\left(e^{-8\cdot 0}\cdot\frac{3}{8}\right) = 2^6\cdot 3-2^9\cdot\lim_{b\to+\infty}\frac{64b^4-3\cdot 2^5b^3+44b^2-5b+\frac{3}{8}}{e^{8b}} =$$

$$=\left[\begin{array}{l}\text{многочлен растёт медленнее,} \\ \text{чем экспонента (можно} \\ \text{посчитать по правилу Лопиталя)} \\ \text{и весь предел тогда стремится} \\ \text{к 0 при } b\to+\infty\end{array}\right] = 2^6\cdot 3$$

Тогда $MISE(\hat{p}_n) = \frac{3^{4/5}}{5^{4/5} \cdot 4} \cdot (2^6 \cdot 3)^{1/5} = \frac{3 \cdot 2 \cdot 2^{1/5}}{5^{4/5} \cdot 2^2} = \frac{3}{5^{4/5} \cdot 2^{4/5}} \approx 1{,}249$

Для нахождения h_{min} воспользуемся утверждением 1.7 из книги Цыбакова „Introduction to nonparametric estimation"

$h = n^{-1/5} \cdot \varepsilon^{-1} \cdot \int K^2(u)du$, где $\varepsilon = MISE(\hat{p}_n)$ для ядра Епанечникова с оптимальным значением $h_{opt\ Epanechnikov}$, $K(u) = \sum_{k=0}^{2} \varphi_k(0)\varphi_k(u) \mathbb{I}\{|u| \le 1\}$ из условий задания. Распишем $K(u) = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + 0 - \frac{\sqrt{5}}{2\sqrt{2}} \cdot \frac{\sqrt{5}}{2\sqrt{2}}(3u^2-1) = \frac{1}{2} - \frac{5}{8}(3u^2-1) =$

$= \frac{9}{8} - \frac{15}{8}u^2$, $u \in [-1;1]$

Посчитаем интеграл отдельно с пределами интегрирования, совпадающими с областью определения $u \in [-1;1]$.

$\int_{-1}^{1} K^2(u)du = \int_{-1}^{1} \left(\frac{9}{8} - \frac{15}{8}u^2\right)^2 du = \int_{-1}^{1} \frac{1}{8^2}\left(9^2 - 2 \cdot 9 \cdot 15u^2 + 15^2u^4\right)du = \frac{1}{64}\int_{-1}^{1}(81 - 18 \cdot 15u^2 + 15^2u^4)du =$

$= \frac{1}{64}\left(81u - \frac{18 \cdot 15 \cdot u^3}{3} + \frac{15^2 \cdot u^5}{5}\right)\Big|_{-1}^{1} = \left[\begin{array}{l}\text{т.к. функция чётная} \\ \text{подставляем один предел} \\ \text{и умножаем всё на 2}\end{array}\right] = \frac{1}{32}(81 - 6 \cdot 15 + 3 \cdot 15) = \frac{81-45}{32} = \frac{36}{32}$

Находим h при условии $n = 1000$ (из задания)

$h = 1000^{-1/5} \cdot \frac{1}{1{,}249} \cdot \frac{36}{32} = \frac{9}{\sqrt[5]{1000} \cdot 1{,}249 \cdot 8} \approx 0{,}226$

Ответ: наименьшее значение h, при котором ядерная оценка с ядром из первых трёх многочленов Лежандра на отрезке $[-1;1]$ лучше любой ядерной оценки с ядром Епанечникова в смысле асимптотического поведения MISE, приближённо равно 0,226.

T2

T2 Необходимо доказать, что оценка $\hat{G}(h)$ является несмещённой оценкой $G(h)$, то есть при $n \to \infty$

$$\mathbb{E}(\hat{G}(h)) = G(h)$$

Выпишем оценку $\hat{G}(h)$ и функцию $G(h)$ из условия задачи.

$$\hat{G}(h) = \int \hat{p}_n^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n} \hat{p}_{n,-i}(x_i)$$

$$G(h) = \mathbb{E}\left(\int \hat{p}_n^2(x)\,dx - 2\int \hat{p}_n\, p(x)\,dx\right)$$

Перепишем равенство, которое необходимо доказать, подставив в него сами выражения для $G(h)$ и $\hat{G}(h)$.

$$\mathbb{E}\left(\int \hat{p}_n^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n} \hat{p}_{n,-i}(x_i)\right) = \mathbb{E}\left(\int \hat{p}_n^2(x)\,dx - 2\int \hat{p}_n\, p(x)\,dx\right)$$

Несложно заметить, что чтобы левая и правая часть были равны, достаточно доказать, что $\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{p}_{n,-i}(x_i)\right) = \mathbb{E}\left(\int \hat{p}_n\, p(x)\,dx\right)$ ⊛

Ядерная оценка плотности для leave-one-out подхода для i-того x:

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

Так как X_i - независимые одинаково распределённые случайные величины, перейдём от матожидания среднего к выражению для плотности одного X, например, X_1. Тогда левая часть ⊛

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{p}_{n,-i}(x_i)\right) = \mathbb{E}\left(\hat{p}_{n,-1}(X_1)\right) = \left[\begin{array}{c}\text{подставляем}\\ \text{ядерную оценку}\end{array}\right] = \mathbb{E}\left(\frac{1}{(n-1)h}\cdot\sum_{j\neq 1} K\left(\frac{X_j - X_1}{h}\right)\right)$$

Нам нужно найти матожидание выражения, в котором присутствуют две случайные величины, при этом независимые.

Тогда при переходе от матожидания от совместной плотности получаем двойной интеграл по первой (обозначим за x) и по второй (z) случайным величинам.

Также отметим, что в последнем выражении суммирование производится $(n-1)$ раз, и под матожиданием сумма заменится на $(n-1)$ одинаковое слагаемое.

Получаем:

$$\mathbb{E}\left(\frac{1}{(n-1)h}\sum_{j\neq 1} K\left(\frac{X_j - X_1}{h}\right)\right) = \frac{\cancel{n-1}}{\cancel{(n-1)}h}\int K\left(\frac{x-z}{h}\right)p(z)\int p(x)\,dz\,dx = \frac{1}{h}\int K\left(\frac{x-z}{h}\right)p(z)\int p(x)\,dx\,dz$$

Рассмотрим правую часть ⊛.

Здесь уместны те же рассуждения про совместную плотность.

Однако теперь мы вычисляем $\hat{p}_n$ для всех n иксов.

$$\hat{p}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$

Тогда под матожиданием получим n одинаковых слагаемых и выражение примет вид (с переобозначенными переменными, как в левой части):

$$\mathbb{E}\left(\int \hat{p}_n\, p(x)\,dx\right) = \mathbb{E}\left(\int p(x)\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)dx\right) = \frac{\cancel{n}}{\cancel{n}h}\int p(x)\int K\left(\frac{x-z}{h}\right)p(z)\,dz\,dx$$

Получили одинаковые выражения для левой и правой части ⊛, а это значит, что $\hat{G}(h)$ является несмещённой оценкой $G(h)$.