

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторным работам 1-4

По дисциплине «Теория вероятностей и математическая статистика»

Выполнил

Студент гр.3630201/80101

В.Н. Сеннов

Руководитель

доцент к.ф.-м.н.

А.Н. Баженов

«___» _____ 202__г.

Санкт-Петербург
2020

Содержание

1	Постановка задачи	4
2	Математическое описание	5
2.1	Рассматриваемые распределения	5
2.1.1	Нормальное распределение	5
2.1.2	Распределение Коши	5
2.1.3	Распределение Лапласа	5
2.1.4	Распределение Пуассона	6
2.1.5	Равномерное распределение	6
2.2	Гистограмма	6
2.3	Вариационный ряд	7
2.4	Выборочные числовые характеристики	7
2.5	Характеристики рассеяния	7
2.6	Построение боксплотов	8
2.7	Теоретическая вероятность выбросов	8
2.8	Ядерные оценки	8
2.9	Гауссово ядро	8
2.10	Выбор ширины полосы	9
3	Особенности реализации	10
4	Результаты работы программы	12
	Заключение	22
	Список литературы	23
A	Репозиторий с исходным кодом	24

Список таблиц

1	Числовые характеристики выборки, соответствующей нормальному распределению	14
2	Числовые характеристики выборки, соответствующей распределению Коши . .	15
3	Числовые характеристики выборки, соответствующей распределению Лапласа .	15
4	Числовые характеристики выборки, соответствующей распределению Пуассона	15
5	Числовые характеристики выборки, соответствующей равномерному распределению	16
6	Доли выбросов и теоретические вероятности выбросов P_{sp}	19

Список иллюстраций

1	Графики для нормального распределения	12
2	Графики для распределения Коши	12
3	Графики для распределения Лапласа	13
4	Графики для распределения Пуассона	13
5	Графики для равномерного распределения	14
6	Боксплот для выборки, соответствующей нормальному распределению	16
7	Боксплот для выборки, соответствующей распределению Коши	17
8	Боксплот для выборки, соответствующей распределению Лапласа	17
9	Боксплот для выборки, соответствующей распределению Пуассона	18
10	Боксплот для выборки, соответствующей равномерному распределению	18
11	Ядерные оценки плотности для нормального распределения, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка.	19
12	Ядерные оценки плотности для распределения Коши, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка. . .	20
13	Ядерные оценки плотности для распределения Лапласа, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка. . .	20
14	Ядерные оценки плотности для распределения Пуассона, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка. .	21
15	Ядерные оценки плотности для равномерного распределения, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка.	21

1 Постановка задачи

1. Для заданных распределений нужно сгенерировать выборки размером 10, 100, 1000 элементов. Для каждой выборки нужно построить гистограмму и теоретическую функцию плотности вероятности на одном рисунке.
2. Для заданных распределений нужно сгенерировать выборки размером 10, 100, 1000 элементов. Для каждой выборки нужно найти следующие характеристики:
 - (a) среднее значение;
 - (b) медиану;
 - (c) полусумму экстремальных значений;
 - (d) полусумму квартилей;
 - (e) усеченное среднее.

Это нужно повторить 1000 раз и найти среднее значение и дисперсию этих характеристик для каждого размера выборки

3. Для заданных распределений нужно сгенерировать выборки размером 20, 100 элементов. Для каждой выборки нужно построить боксплот и рассчитать долю выбросов. Долю выбросов нужно посчитать 1000 раз и взять среднее значение. Также необходимо рассчитать теоретическую вероятность выбросов.
4. Для заданных распределений нужно сгенерировать выборки размером 20, 60 и 100 элементов. Для каждой выборки нужно построить ядерные оценки плотности со следующими значением ширины полосы:
 - (a) ширина полосы по правилу Сильвермана;
 - (b) половина ширины полосы по правилу Сильвермана;
 - (c) ширина в два раза больше ширины полосы по правилу Сильвермана.

Для всех распределений необходимо построить на одном графике теоретическую плотность и ядерную оценку. Для непрерывных распределений график необходимо построить на промежутке $[-4; 4]$, для распределения Пуассона на $[6; 14]$.

Заданные распределения:

1. Нормальное (гауссово) распределение с параметрами $\mu = 0$, $\sigma = 1$;
2. Распределение Коши с параметрами $\mu = 0$, $\lambda = 1$;
3. Распределение Лапласа с параметрами $\mu = 0$, $\lambda = \frac{1}{\sqrt{2}}$;
4. Распределение Пуассона с параметром $\mu = 10$;
5. Равномерное распределение с параметрами $a = -\sqrt{3}$, $b = \sqrt{3}$.

2 Математическое описание

2.1 Рассматриваемые распределения

2.1.1 Нормальное распределение

Рассматриваемое распределение имеет следующую функцию плотности вероятности:

$$f_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (1)$$

Функция распределения имеет следующий вид:

$$F_n(x) = 0.5 + \Phi(x), \quad (2)$$

где $\Phi(x)$ — нормированная функция Лапласа [2].

Сгенерировать случайную величину, соответствующую данному нормальному распределению можно по следующей формуле:

$$x = \sum_{i=1}^{12} r_i - 6, \quad (3)$$

где r_i — случайная величина из интервала $(0, 1)$ [1].

2.1.2 Распределение Коши

Рассматриваемое распределение имеет следующую функцию плотности вероятности:

$$f_c(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2} \quad (4)$$

Функция распределения имеет следующий вид:

$$F_c(x) = 0.5 + \frac{\operatorname{arctg} x}{\pi} \quad (5)$$

Сгенерировать случайную величину, соответствующую данному распределению можно по следующей формуле:

$$x = \operatorname{tg}(2\pi r_i), \quad (6)$$

где r_i — случайная величина из интервала $(0, 1)$ [1].

2.1.3 Распределение Лапласа

Рассматриваемое распределение имеет следующую функцию плотности вероятности:

$$f_L(x) = \frac{1}{2\sqrt{2}} e^{-\frac{|x|}{\sqrt{2}}} \quad (7)$$

Функция распределения задается следующим образом:

$$F_L(x) = \begin{cases} \frac{1}{2} e^{\frac{x}{\sqrt{2}}}, & x \geq 0 \\ 1 - \frac{1}{2} e^{-\frac{x}{\sqrt{2}}}, & x < 0 \end{cases} \quad (8)$$

Сгенерировать случайную величину, соответствующую данному распределению можно по следующей формуле:

$$x = \sqrt{2} \ln(r_{i+1}/r_i) \quad (9)$$

где r_i — случайная величина из интервала $(0, 1)$ [1].

2.1.4 Распределение Пуассона

Рассматриваемое распределение задается формулой:

$$p_k = \frac{10^k}{k!} e^{-10} \quad (10)$$

Случайная величина, соответствующая данному распределению может быть сгенерирована по алгоритму 1 [1].

Алгоритм 1 Генерирование случайной величины согласно распределению Пуассона

Вход: μ — параметр распределения, r_i — случайная величина из интервала $(0; 1)$

Выход: x — случайная величина

```
 $p \leftarrow e^{-\mu}$   
 $x \leftarrow 0$   
 $r \leftarrow r_i - p$   
while  $r \geq 0$  do  
   $x \leftarrow x + 1$   
   $p \leftarrow p\mu/x$   
   $r \leftarrow r - p$   
end while  
return  $x$ 
```

2.1.5 Равномерное распределение

Рассматриваемое равномерное распределение имеет следующую функцию плотности вероятности:

$$f_U(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & -\sqrt{3} \leq x \leq \sqrt{3} \\ 0, & \text{иначе} \end{cases} \quad (11)$$

2.2 Гистограмма

Гистограмма в математической статистике — это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него [5].

В процессе построения гистограммы числовой промежутков, покрывающий значения элементов выборки, разбивают на несколько интервалов. Эти интервалы называются бинами. Для каждого бина подсчитывается количество элементов выборки, попавших в него. Далее на графике строятся прямоугольники, основание которых совпадает с бинами, а высота пропорциональна количеству элементов, попавших в этот бин.

В рамках данной работы брался отрезок $[x_{min}; x_{max}]$, где x_{min} и x_{max} — минимальный и максимальный элементы выборки. Отрезок делился на бины равной ширины. Количество бинов n определялось по эвристике:

$$n = \left\lceil \sqrt{N} \right\rceil, \quad (12)$$

где N — количество элементов в выборке.

Гистограмма задавалась следующей функцией:

$$h(x) = \begin{cases} \frac{n_k}{N \cdot b}, & x \in [x_{min}; x_{max}] \\ 0, & \text{иначе} \end{cases} \quad (13)$$

где b — ширина одного бина, N — количество элементов в выборке, x_{min} и x_{max} — минимальный и максимальный элементы выборки, $k = \left\lceil \frac{x - x_{min}}{b} \right\rceil$, n_k — количество элементов выборки в k -том бине.

2.3 Вариационный ряд

Вариационным рядом называется последовательность элементов выборки, расположенных в неубывающем порядке, причем одинаковые элементы повторяются [5].

2.4 Выборочные числовые характеристики

Для вариационного ряда x_1, x_2, \dots, x_n определяют следующие числовые характеристики:

1. Выборочное среднее:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (14)$$

2. Выборочная медиана:

$$\text{med } x = \begin{cases} x_{l+1} & \text{если } n = 2l + 1 \\ \frac{x_l + x_{l+1}}{2} & \text{если } n = 2l \end{cases} \quad (15)$$

3. Полусумма экстремалей:

$$z_R = \frac{x_1 + x_n}{2} \quad (16)$$

4. Выборочный квантиль порядка p :

$$z_p = x_{[np]}$$

5. Полусумма квартилей:

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (17)$$

6. Усеченное среднее:

$$z_{tr} = \frac{1}{n - 2r} \sum_{i=r+1}^{n-r} x_i, \quad r \approx \frac{n}{4} \quad (18)$$

[2]

2.5 Характеристики рассеяния

Для выборки можно рассчитать выборочную дисперсию по формуле:

$$D = \overline{x^2} - (\bar{x})^2, \quad (19)$$

где \bar{x} — выборочное среднее, а $\overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$.

2.6 Построение боксплотов

Боксплотом называется диаграмма, компактно изображающая распределение одномерной величины. [3]

Диаграмма состоит из следующих частей:

1. Ящик. Границами ящика являются верхний и нижний квартили. Внутри ящика проводится линия — медиана.
2. Усы. Вне ящика изображают линии, напоминающие усы. Длина усов равна либо $3/2$ от межквартильного расстояния, либо разнице экстремального значения с квартилем. Тогда для выборки x_1, \dots, x_n концы усов будут иметь координаты X_1 и X_2 :

$$X_1 = \max \{x_1, z_{1/4} - 3/2 (z_{3/4} - z_{1/4})\} \quad (20)$$

$$X_2 = \min \{x_n, z_{3/4} + 3/2 (z_{3/4} - z_{1/4})\} \quad (21)$$

3. Выбросы. Значения выборки, меньшие X_1 или большие X_2 отображаются на диаграмме кружочками.

[3]

2.7 Теоретическая вероятность выбросов

По формулам (20) и (21) можно вычислить теоретические значения X_1^T и X_2^T . Выбросом считается значение случайной величины, меньшее X_1^T или большее X_2^T . Тогда теоретическая вероятность выброса может быть рассчитана по формуле:

$$P_{sp} = P(x \in (-\infty; X_1^T) \cup (X_2^T; +\infty)) = F(X_1^T) + (1 - F(X_2^T)), \quad (22)$$

где $F(x)$ — функция распределения [5].

2.8 Ядерные оценки

Оценкой плотности вероятности $f(x)$ называется функция $\hat{f}(x)$, построенная на основе выборке и приближенно равная $f(x)$ [5].

Ядерной оценкой плотности $\hat{f}_n(x)$, построенной на основе выборки x_1, \dots, x_n называется функция:

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - x_i}{h}\right), \quad (23)$$

где $h > 0$ — некий параметр, называемый шириной полосы, $K(x)$ — ядерная функция, для нее выполняется $\int_{-\infty}^{+\infty} K(t)dt = 1$ [4].

2.9 Гауссово ядро

В данной лабораторной в качестве ядерной функции было использовано Гауссово ядро — плотность вероятности стандартного нормального распределения:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (24)$$

2.10 Выбор ширины полосы

В рамках данной работы ширина полосы определялась по правилу Сильвермана:

$$h_n = 1.06\hat{\sigma}n^{-1/5}, \quad (25)$$

где n — размер выборки, $\hat{\sigma}$ — выборочное стандартное отклонение.

3 Особенности реализации

Программа для выполнения лабораторной была написана на языке Python 3.8.2. Для построения графиков использовалась библиотека Matplotlib.

Программа состоит из 7 модулей:

1. Модуль **distributions**. Содержит классы, реализующие необходимые распределения. Все классы унаследованы от класса **AbstractDistribution**. Он имеет 2 поля:
 - **name: str** — название распределения, используется для подписи на на графиках;
 - **parameters: dict** — словарь с параметрами распределения.

Класс имеет следующие методы:

- **Конструктор** — инициализирует поля класса.
- **x()->float** — возвращает случайную величину, соответствующую распределению.
- **f(x:float)->float** — возвращает значение функции плотности вероятности в точке **x**.
- **F(x:float)->float** — возвращает значение функции распределения в точке **x**.
- **discrete()** — возвращает True, если распределение дискретное, иначе False. По умолчанию возвращает False.

От этого класса унаследованы классы **Normal**, **Cauchy**, **Laplace**, **Poisson**, **Uniform**. Они реализуют формулы для вычислений, приведенные в разделе 2.1. Интерес представляет только метод вычисления функции распределения для нормального распределения (2). Значение функции вычисляется по таблице, сохраненной в отдельном файле. Генерация случайной величины производилась согласно формулам (3), (6), (9) и алгоритму 1. Также стоит указать, что диапазон возможных значений случайной величины, соответствующей распределению Коши был насильно ограничен до интервала $(-1591; 1591)$, чтобы не возникало ошибок области определения математических функций.

2. Модуль **histogram**. Содержит класс **Histogram**, реализующий преобразование выборки в гистограмму. Количество бинов вычисляется при помощи эвристики (12). Высота столбца в гистограмме вычисляется по формуле (13).
3. Модуль **plot**. Модуль верхнего уровня для лабораторной №1, в нем инициализируются заданные распределения, генерируются выборки и отображаются графики.
4. Модуль **characteristics**. В нем на основе формул (14-18) рассчитываются числовые характеристики. При помощи формулы (14) рассчитывается среднее значение для каждой характеристики, при помощи формулы (19) рассчитывается дисперсия.
5. Модуль **boxplots**. В нем строятся необходимые боксплоты с помощью функции библиотеки Matplotlib.
6. Модуль **spikes**. В нем на основе формул (20) и (21) рассчитывается средняя доля выбросов.

7. Модуль **kernel**. В нем отдельной функцией реализовано Гауссово ядро по формуле (24). Также отдельной функцией было реализовано вычисление значения ядерной оценки по формуле (23) с шириной полосы согласно правилу Сильвермана (25).

В приложении А приведена ссылка на репозиторий с исходным кодом.

4 Результаты работы программы

В этом разделе представлены результаты работы программы.

На рис. 1-5 представлены графики функции плотности вероятности $f(x)$ и функции распределения $F(x)$ для выборок разного размера.

На рис. 1 представлены графики для нормального распределения. Теоретические значения вычислены по формулам 1 и 2.

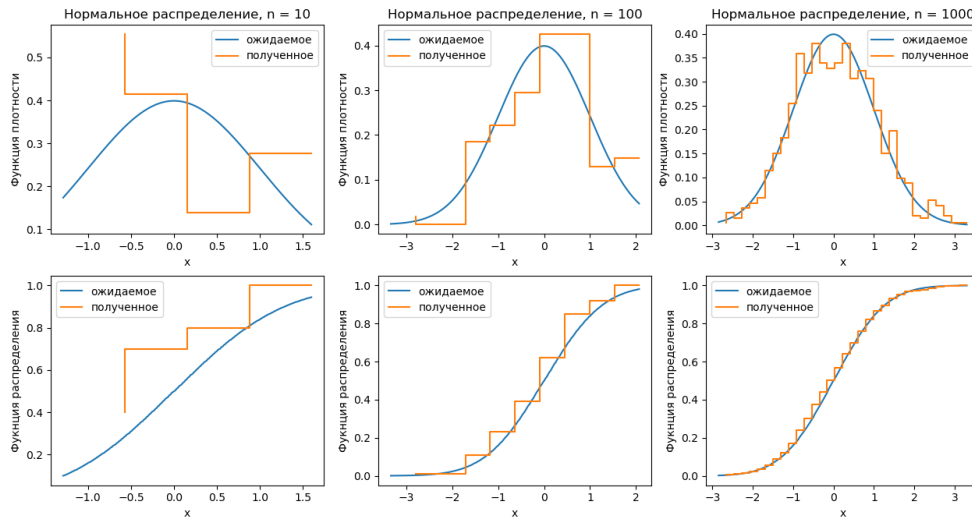


Рис. 1: Графики для нормального распределения

На рис. 2 представлены графики для распределения Коши. Теоретические значения вычислены по формулам (4) и (5).

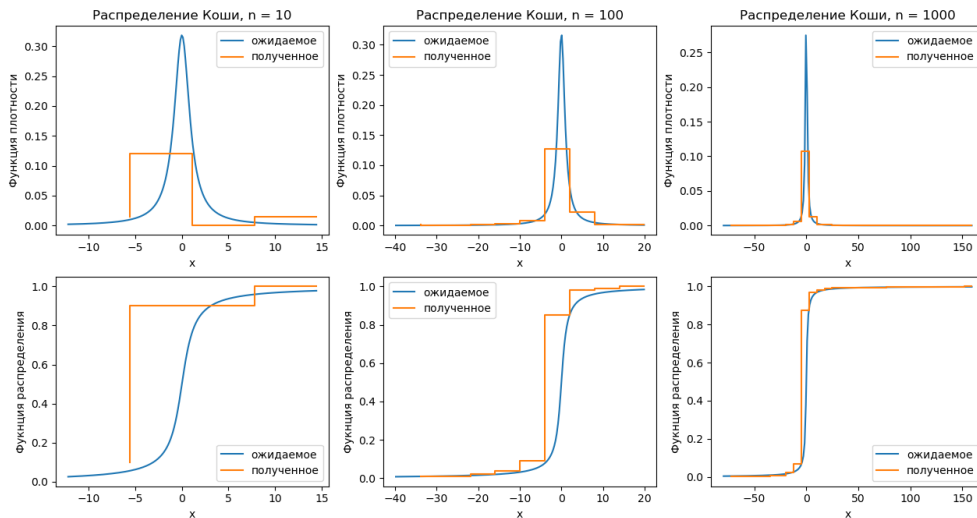


Рис. 2: Графики для распределения Коши

На рис. 3 представлены графики для распределения Лапласа. Теоретические значения вычислены по формулам (7) и (8).

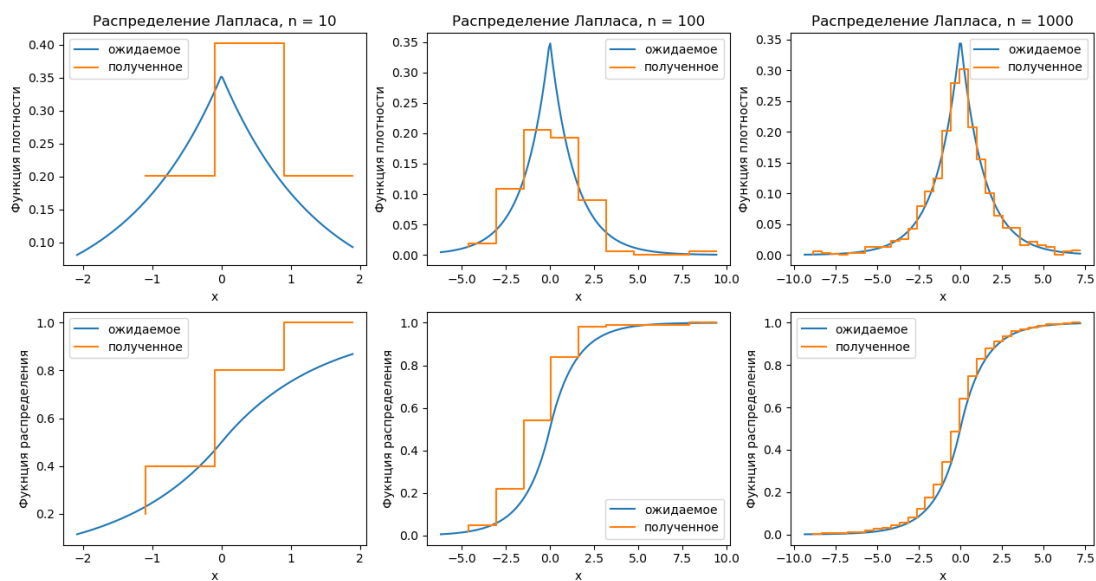


Рис. 3: Графики для распределения Лапласа

На рис. 4 представлены графики для распределения Пуассона. Теоретические значения вычислены по формуле (10).

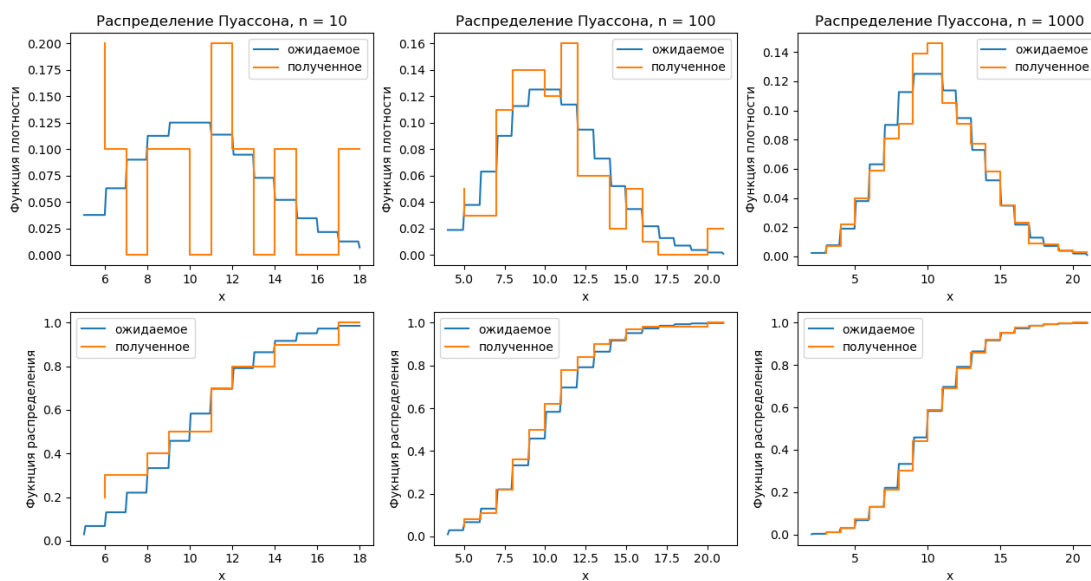


Рис. 4: Графики для распределения Пуассона

На рис. 5 представлены графики для равномерного распределения. Теоретические значения вычислены по формуле (11).

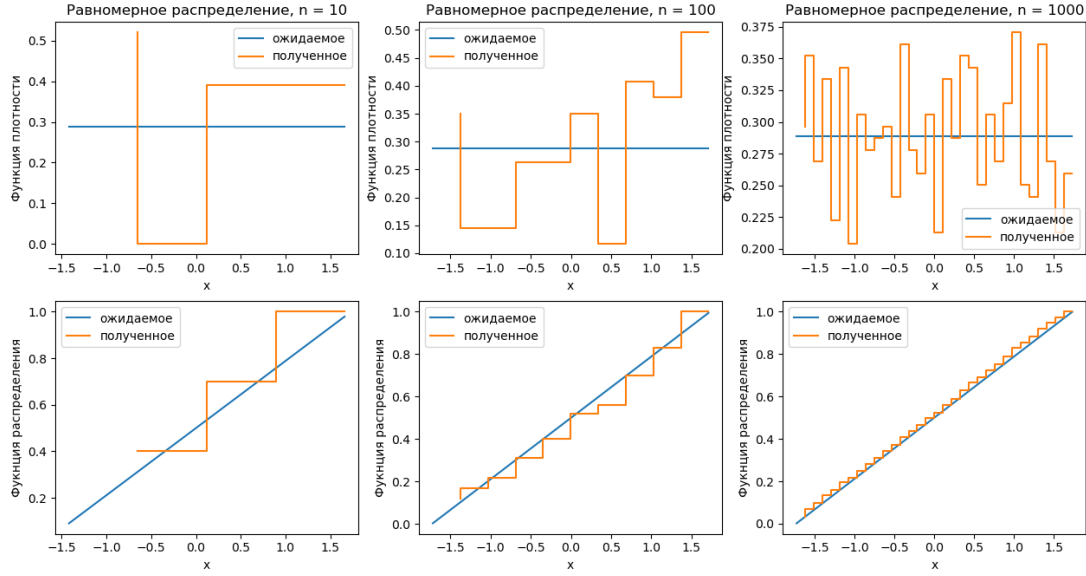


Рис. 5: Графики для равномерного распределения

В таблицах 1-5 представлены результаты расчета числовых характеристик выборки. Выборочные характеристики \bar{x} - z_{tr} были рассчитаны по формулам (14-18). Расчет погрешности для средней величины был произведен по формуле:

$$\Delta_z = \sqrt{D(z)}$$

Нормальное распределение						
		\bar{x}	$\text{med } x$	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$	0.0 ± 0.4	0.0 ± 0.4	0.0 ± 0.5	0.0 ± 0.4	0.0 ± 0.4
	D_z	0.103	0.146	0.183	0.124	0.124
$n = 100$	$\bar{z} \pm \Delta_z$	0.0 ± 0.1	0.0 ± 0.2	0.0 ± 0.3	0.0 ± 0.2	0.0 ± 0.2
	D_z	0.009	0.015	0.074	0.012	0.011
$n = 1000$	$\bar{z} \pm \Delta_z$	0.00 ± 0.04	0.00 ± 0.04	0.0 ± 0.2	0.00 ± 0.04	0.00 ± 0.04
	D_z	0.001	0.002	0.042	0.001	0.001

Таблица 1: Числовые характеристики выборки, соответствующей нормальному распределению

Распределение Коши						
		\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$ D_z	0 ± 4 15.4	0.0 ± 0.6 0.35	0 ± 20 296	0.0 ± 1.1 1.16	0.0 ± 0.8 0.519
$n = 100$	$\bar{z} \pm \Delta_z$ D_z	0.0 ± 1.2 1.521	0.0 ± 0.2 0.024	0.0 ± 40 1190	0.0 ± 0.3 0.050	0.0 ± 0.2 0.024
$n = 1000$	$\bar{z} \pm \Delta_z$ D_z	0.00 ± 0.5 0.172	0.00 ± 0.06 0.002	0 ± 20 400	0.00 ± 0.08 0.005	0.00 ± 0.06 0.003

Таблица 2: Числовые характеристики выборки, соответствующей распределению Коши

Распределение Лапласа						
		\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$ D_z	0.0 ± 0.6 0.361	0.0 ± 0.6 0.284	0.0 ± 1.2 1.379	0.0 ± 0.6 0.369	0.0 ± 0.6 0.278
$n = 100$	$\bar{z} \pm \Delta_z$ D_z	0.0 ± 0.2 0.041	0.0 ± 0.2 0.027	0.0 ± 1.0 1.025	0.0 ± 0.2 0.040	0.0 ± 0.2 0.028
$n = 1000$	$\bar{z} \pm \Delta_z$ D_z	0.00 ± 0.07 0.0036	0.00 ± 0.05 0.0019	0.0 ± 0.6 0.3298	0.00 ± 0.06 0.0034	0.00 ± 0.05 0.0021

Таблица 3: Числовые характеристики выборки, соответствующей распределению Лапласа

Распределение Пуассона						
		\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$ D_z	10.0 ± 1.0 1.019	9.8 ± 1.2 1.442	10 ± 2 1.762	9.9 ± 1.2 1.245	9.9 ± 1.1 1.156
$n = 100$	$\bar{z} \pm \Delta_z$ D_z	10.0 ± 0.3 0.097	9.8 ± 0.5 0.211	10.9 ± 0.8 0.689	9.9 ± 0.4 0.152	9.8 ± 0.4 0.121
$n = 1000$	$\bar{z} \pm \Delta_z$ D_z	10.0 ± 0.1 0.0096	10.00 ± 0.04 0.0019	11.3 ± 0.4 0.1105	10.00 ± 0.06 0.0027	9.86 ± 0.10 0.0109

Таблица 4: Числовые характеристики выборки, соответствующей распределению Пуассона

Равномерное распределение						
		\bar{x}	med x	z_R	z_Q	z_{tr}
$n = 10$	$\bar{z} \pm \Delta_z$ D_z	0.0 ± 0.4 0.106	0.0 ± 0.5 0.232	0.0 ± 0.3 0.046	0.0 ± 0.4 0.141	0.0 ± 0.5 0.166
$n = 100$	$\bar{z} \pm \Delta_z$ D_z	0.00 ± 0.11 0.0103	0.0 ± 0.2 0.232	0.00 ± 0.03 0.0462	-0.02 ± 0.12 0.141	0.0 ± 0.2 0.166
$n = 1000$	$\bar{z} \pm \Delta_z$ D_z	0.00 ± 0.04 0.0009	0.00 ± 0.06 0.0028	0.000 ± 0.003 0.00005	0.00 ± 0.04 0.0014	0.00 ± 0.05 0.0019

Таблица 5: Числовые характеристики выборки, соответствующей равномерному распределению

На рис. 6 изображен боксплот для выборки, соответствующей нормальному распределению.

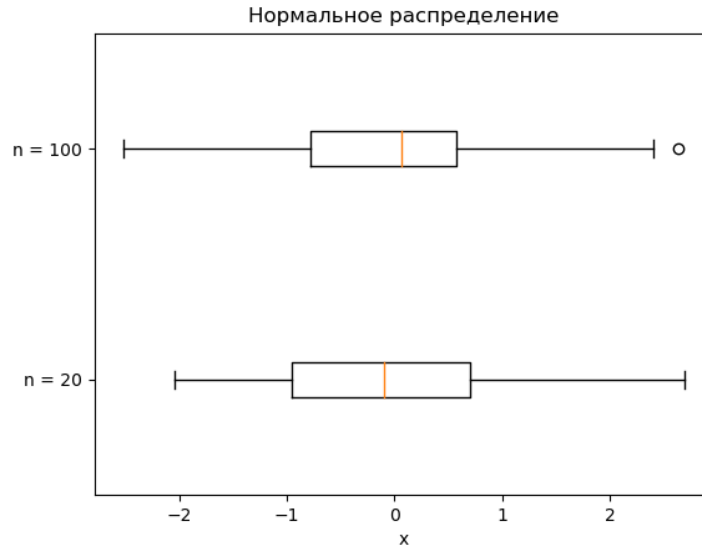


Рис. 6: Боксплот для выборки, соответствующей нормальному распределению

На рис. 7 изображен боксплот для выборки, соответствующей распределению Коши.

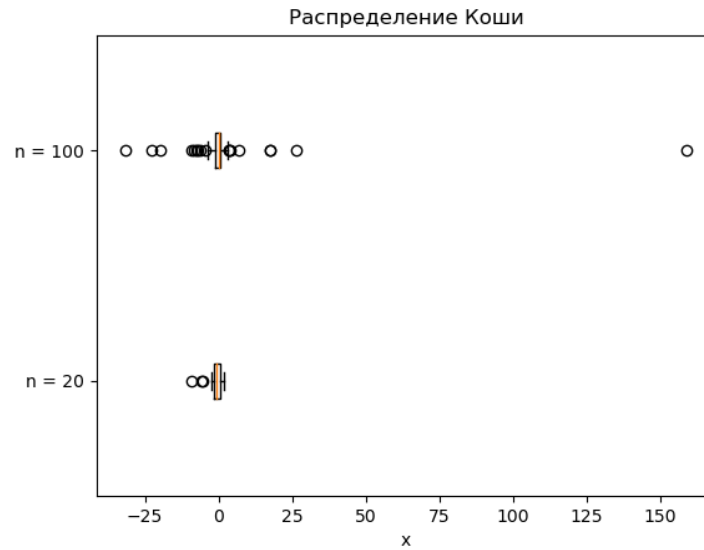


Рис. 7: Боксплот для выборки, соответствующей распределению Коши

На рис. 8 изображен боксплот для выборки, соответствующей распределению Лапласа.

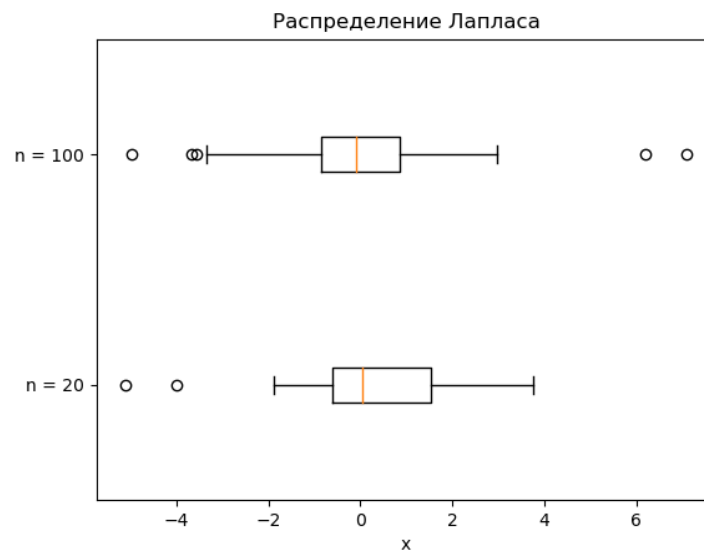


Рис. 8: Боксплот для выборки, соответствующей распределению Лапласа

На рис. 9 изображен боксплот для выборки, соответствующей распределению Пуассона.

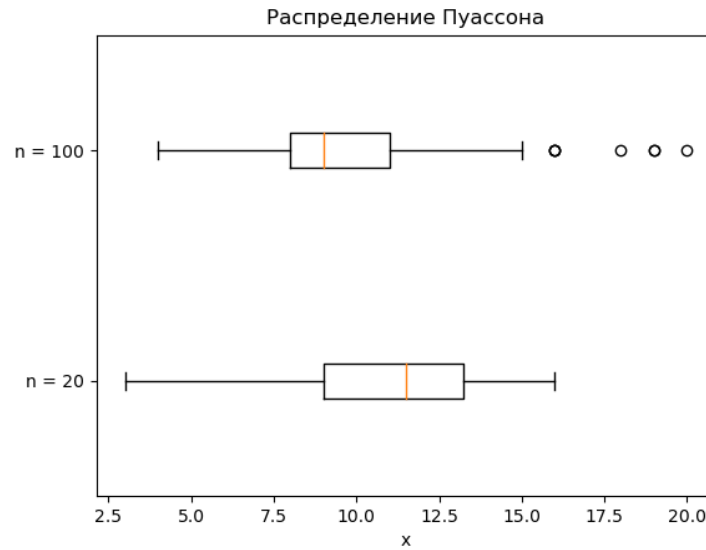


Рис. 9: Боксплот для выборки, соответствующей распределению Пуассона

На рис. 10 изображен боксплот для выборки, соответствующей равномерному распределению.

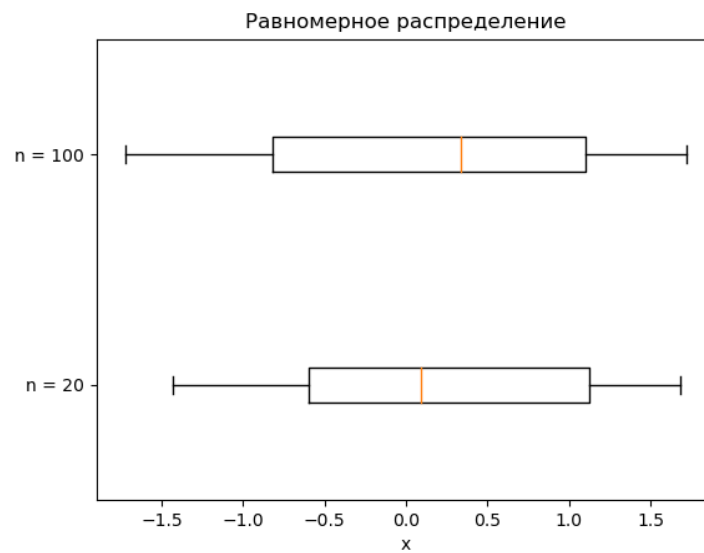


Рис. 10: Боксплот для выборки, соответствующей равномерному распределению

В таблице 6 представлены полученные экспериментально доли выбросов и теоретические значения. Экспериментальные значения приведены с округлением, погрешность рассчитана по следующей формуле:

$$\Delta_z = \sqrt{\frac{1}{n} \sum z_i^2 - \frac{1}{n} \left(\sum z_i \right)^2}$$

Распределение	n	Доля выбросов	P_{sp}
Нормальное	20	0.02	0.0069
	100	0.007	
Коши	20	0.14	0.156
	100	0.15	
Лапласа	20	0.06	0.062
	100	0.06	
Пуассона	20	0.02	0.010
	100	0.01	
Равномерное	20	0.001	0
	100	0.0	

Таблица 6: Доли выбросов и теоретические вероятности выбросов P_{sp}

Ядерные оценки плотности для графиков на рис. 11-15 были вычислены по формулам (23) и (24).

На рис. 11 представлены ядерные оценки для нормального распределения.

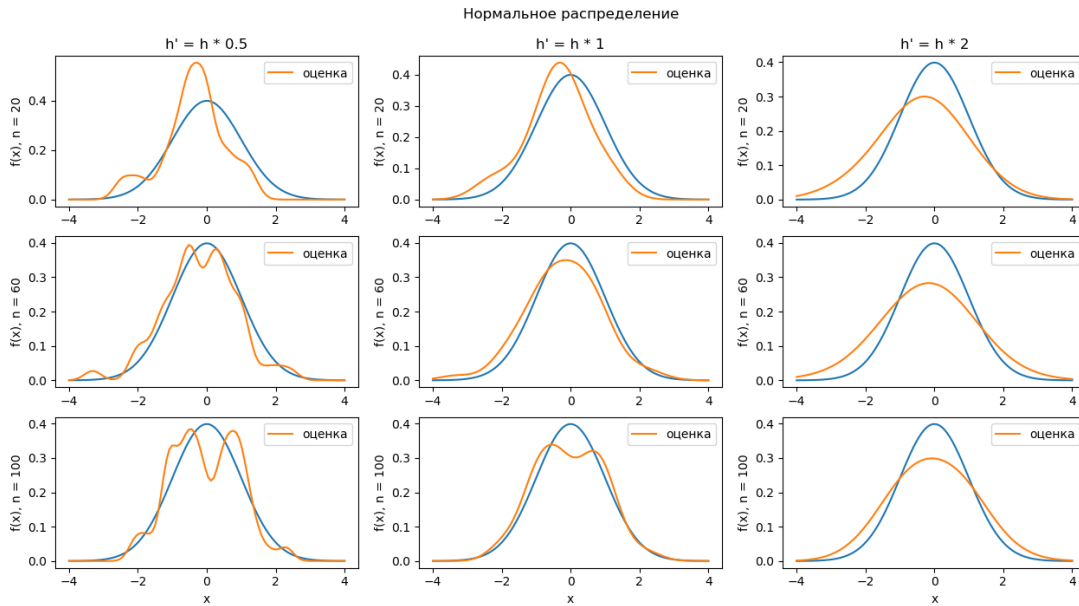


Рис. 11: Ядерные оценки плотности для нормального распределения, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 12 представлены ядерные оценки для распределения Коши.

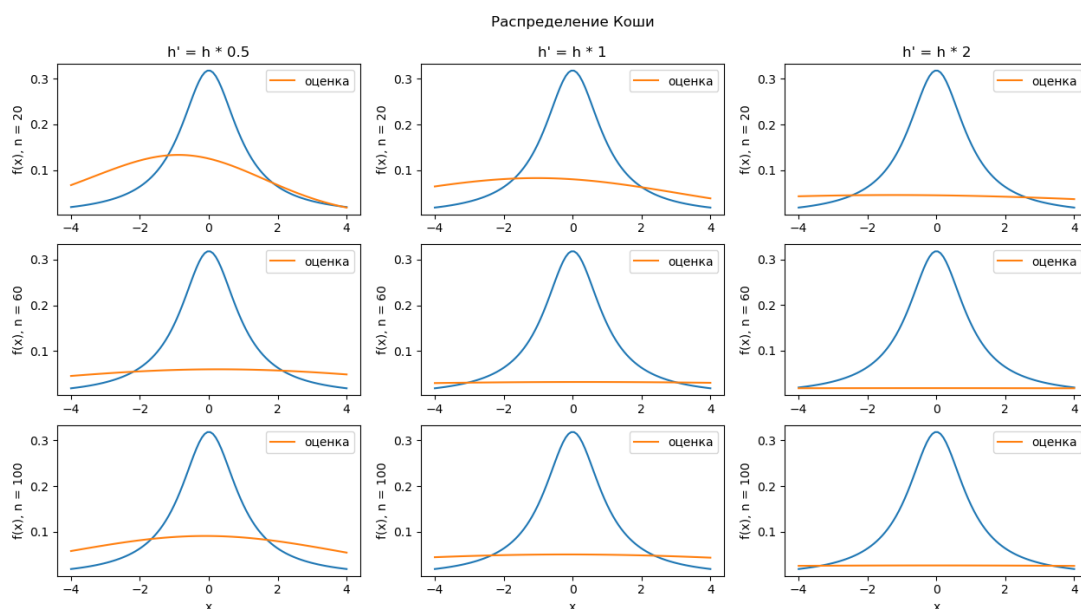


Рис. 12: Ядерные оценки плотности для распределения Коши, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 13 представлены ядерные оценки для распределения Лапласа.

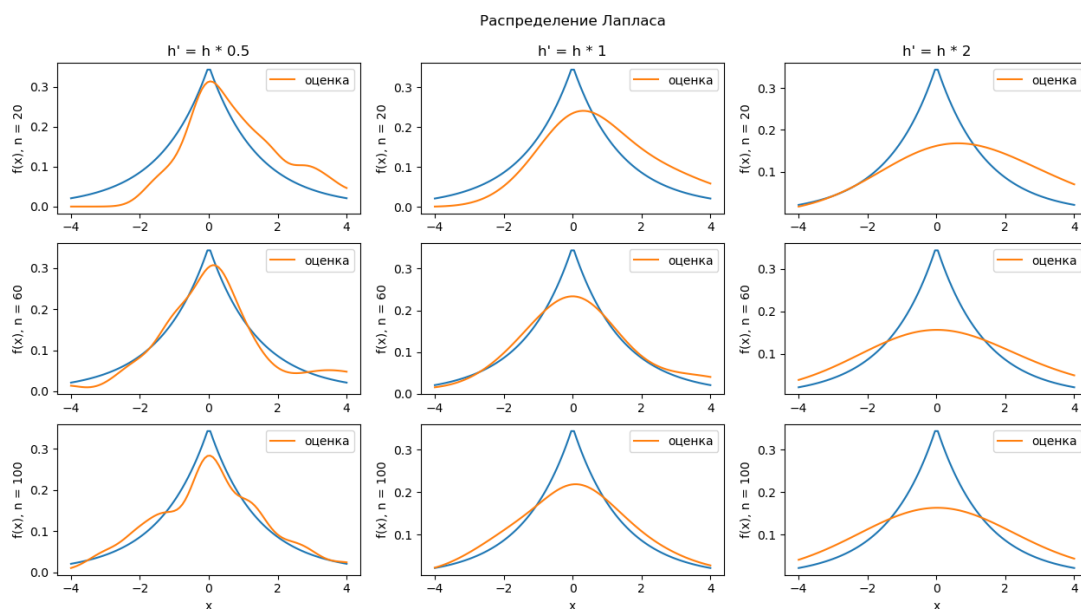


Рис. 13: Ядерные оценки плотности для распределения Лапласа, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 14 представлены ядерные оценки для распределения Пуассона.

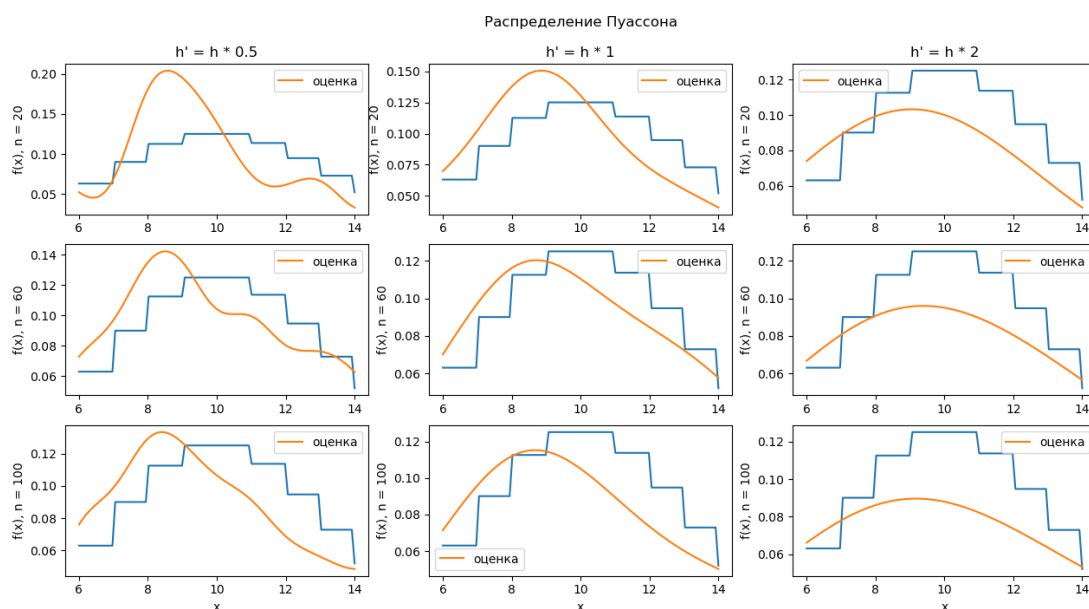


Рис. 14: Ядерные оценки плотности для распределения Пуассона, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 15 представлены ядерные оценки для равномерного распределения.

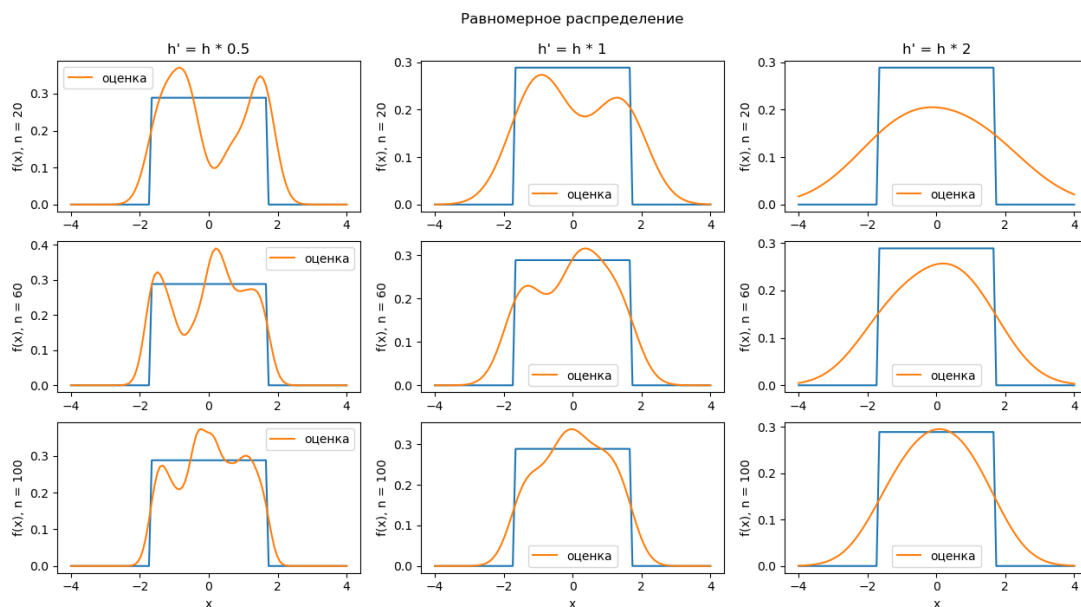


Рис. 15: Ядерные оценки плотности для равномерного распределения, где h — ширина полосы по формуле (25), h' — ширина полосы, с которой строилась ядерная оценка.

Заключение

1. В рамках лабораторной работы были сгенерированы выборки трех разных размеров для пяти различных распределений. Были построены графики гистограмм полученных выборок и теоретических функций.
Заметно, что при малой выборке гистограммы лишь отдаленно напоминают теоретические функции, но при большей выборке функции распределения становятся очень похожи.
2. В рамках лабораторной работы были рассчитаны средние значения числовых характеристик выборок разного размера для заданных распределений.
Заметно, что при большем количестве элементов выборки значения среднего и медианы точнее соответствует математическому ожиданию и теоретическому значению медианы. Стоит отметить, что при реализации генерации выборки, соответствующей распределению Коши, был намеренно ограничен диапазон значений до интервала $(-1591; 1591)$, чтобы избежать чересчур больших значений случайной величины. Несмотря на это, заметно, что дисперсия при подсчете среднего больше, чем для других распределений. Заметно также, что для медианы и усеченного среднего этого эффекта не наблюдается.
3. В рамках лабораторной работы были построены боксплоты для заданных распределений, были вычислены доли выбросов и теоретические вероятности выбросов.
По построенным боксплотам можно легко отличить на вид распределение Лапласа и Коши, а вот нормальное распределение и равномерное распределение оказываются очень похожи.
Заметно, что если выборка больше, то доля выбросов ближе к теоретической. Также заметно, что разные распределения весьма заметно отличаются долями выбросов.
4. В рамках лабораторной работы были построены ядерные оценки плотности для заданных распределений с различными значениями ширины полосы. Для полученных оценок были построены графики.
По построенным графикам видно, что чем меньше значение ширины полосы, тем чувствительнее ядерная оценка к отдельным значениям выборки, тем менее она гладкая. Судя по построенным графикам, можно сказать, что для нормального распределения, равномерного распределения и для распределения Пуассона лучше выбирать ширину полосы согласно правилу Сильвермана. Для распределения Лапласа лучше взять значение ширины полосы вдвое меньше. Для распределения Коши стоит брать еще меньшее значение ширины полосы.
Также стоит заметить, что для распределения Коши оценка получилась слишком сглаженная из-за того, что выборочное стандартное отклонение чересчур велико. Такая оценка не может быть названа состоятельной.
Также можно заметить, что при увеличении размера выборки точность ядерной оценки не сильно возрастает.

Программа для лабораторной была написана языке Python 3.8.2, для построения графиков использовалась библиотека Matplotlib.

Список литературы

- [1] Вадзинский Р. Н. Справочник по вероятностным распределениям. — СПб.: Наука, 2001. — 295 с., ил. 116.
- [2] Максимов Ю. Д. Математика. Теория вероятностей и случайных процессов; Учебное пособие / Под ред. В. И. Антонова. — СПб.: Изд-во Политехн. ун-та, 2008. — 384 с.
- [3] Box plot. // Wikipedia, the free encyclopedia. — URL: https://en.wikipedia.org/wiki/Box_plot. — (дата обращения: 09.11.2020)
- [4] Kernel density estimation. // Wikipedia, the free encyclopedia. — URL: https://en.wikipedia.org/wiki/Kernel_density_estimation. — (дата обращения: 18.11.2020)
- [5] Теоретическое приложение к лабораторным работам №1-4 по дисциплине «Математическая статистика». — Спб.: Санкт-Петербургский политехнический университет, 2020. — 12 с.

А Репозиторий с исходным кодом

Исходный код программы для данной лабораторной размещен на сервисе GitHub.

Ссылка на репозиторий: <https://github.com/Vovan-S/TV-Lab1>.