

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ
ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ
ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ФИЗИКИ

Математическая статистика
Отчёт по лабораторным работам №1-4

Выполнил:

Студент: Парусов Владимир

Группа: 5030102/90201

Принял:

к. ф.-м. н., доцент

Баженов Андрей Николаевич

2022 г.

Содержание

| | |
|---|----|
| 1. Постановка задачи | 3 |
| 2. Теория | 4 |
| 2.1. Распределения | 4 |
| 2.2. Гистограмма | 4 |
| 2.2.1. Построение гистограммы | 4 |
| 2.3. Вариационный ряд | 5 |
| 2.4. Выборочные числовые характеристики | 5 |
| 2.4.1. Характеристики положения | 5 |
| 2.4.2. Характеристики рассеяния | 6 |
| 2.5. Боксплот Тьюки | 6 |
| 2.5.1. Построение | 6 |
| 2.6. Теоретическая вероятность выбросов | 6 |
| 2.7. Эмпирическая функция распределения | 7 |
| 2.7.1. Статистический ряд | 7 |
| 2.7.2. Определение | 7 |
| 2.7.3. Вычисление | 7 |
| 2.8. Оценки плотности вероятности | 7 |
| 2.8.1. Определение | 7 |
| 2.8.2. Ядерные оценки | 7 |
| 3. Реализация | 8 |
| 4. Результаты | 9 |
| 4.1. Гистограммы | 9 |
| 4.2. Характеристики положения и рассеяния | 10 |
| 4.3. Боксплот Тьюки | 12 |
| 4.4. Экспериментальная доля выбросов | 15 |
| 4.5. Теоретическая вероятность выбросов | 15 |
| 4.6. Эмпирическая функция распределения | 15 |
| 4.7. Ядерные оценки плотности распределения | 17 |
| 5. Обсуждение | 22 |
| 5.1. Гистограммы | 22 |
| 5.2. Характеристики положения и рассеяния | 22 |
| 5.3. Боксплот Тьюки | 22 |
| 5.4. Эмпирическая функция распределения. Ядерные оценки плотности | 23 |
| 6. Литература | 23 |
| 7. Ссылка на репозиторий | 23 |

Список иллюстраций

| | |
|---|----|
| 1. Гистограмма и плотность вероятности для нормального распределения | 9 |
| 2. Гистограмма и плотность вероятности для распределения Коши | 9 |
| 3. Гистограмма и плотность вероятности для распределения Лапласа | 10 |
| 4. Гистограмма и плотность вероятности для распределения Пуассона | 10 |
| 5. Гистограмма и плотность вероятности для равномерного распределения | 10 |
| 6. Нормальное распределение | 13 |
| 7. Распределение Коши | 13 |
| 8. Распределение Лапласа | 13 |
| 9. Распределение Пуассона | 14 |
| 10. Равномерное распределение | 14 |
| 11. Нормальное распределение | 15 |
| 12. Распределение Коши | 16 |
| 13. Распределение Лапласа | 16 |
| 14. Распределение Пуассона | 16 |
| 15. Равномерное распределение | 17 |
| 16. Нормальное распределение, $n = 20$ | 17 |
| 17. Нормальное распределение, $n = 60$ | 17 |
| 18. Нормальное распределение, $n = 100$ | 18 |
| 19. Распределение Коши, $n = 20$ | 18 |
| 20. Распределение Коши, $n = 60$ | 18 |
| 21. Распределение Коши, $n = 100$ | 19 |
| 22. Распределение Лапласа, $n = 20$ | 19 |
| 23. Распределение Лапласа, $n = 60$ | 19 |
| 24. Распределение Лапласа, $n = 100$ | 20 |
| 25. Распределение Пуассона, $n = 20$ | 20 |
| 26. Распределение Пуассона, $n = 60$ | 20 |
| 27. Распределение Пуассона, $n = 100$ | 21 |
| 28. Равномерное распределение, $n = 20$ | 21 |
| 29. Равномерное распределение, $n = 60$ | 21 |
| 30. Равномерное распределение, $n = 100$ | 22 |

Список таблиц

| | |
|--|----|
| 1. Нормальное распределение | 11 |
| 2. Распределение Коши | 11 |
| 3. Распределение Лапласа | 11 |
| 4. Распределение Пуассона | 12 |
| 5. Равномерное распределение | 12 |
| 6. Экспериментальная доля выбросов | 15 |
| 7. Теоретическая доля выбросов | 15 |

1. Постановка задачи

Даны пять распределений:

- Нормальное распределение: $N(x, 0, 1)$
- Распределение Коши: $C(x, 0, 1)$
- Распределение Лапласа: $L(x, 0, \frac{1}{\sqrt{2}})$
- Распределение Пуассона: $P(k, 10)$
- Равномерное распределение: $U(x, -\sqrt{3}, \sqrt{3})$

1. Сгенерировать выборки размером 10, 50 и 1000 элементов.

Построить на одном рисунке гистограмму и график плотности распределения.

2. Сгенерировать выборки размером 10, 100 и 1000 элементов. Для каждой выборки вычислить следующие статистические характеристики положения данных: \bar{x} , $medx$, z_R , z_Q , z_{tr} . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

$$E(z) = \bar{z} \quad (1)$$

Вычислить оценку дисперсии по формуле:

$$D(z) = \overline{z^2} - \bar{z}^2 \quad (2)$$

Представить полученные данные в виде таблиц.

3. Сгенерировать выборки размером 20 и 100 элементов.

Построить для них боксплот Тьюки.

Для каждого распределения определить долю выбросов экспериментально (сгенерировав выборку, соответствующую распределению 1000 раз, и вычислив среднюю долю выбросов) и сравнить с результатами, полученными теоретически.

4. Сгенерировать выборки размером 20, 60 и 100 элементов.

Построить на них эмпирические функции распределения и ядерные оценки плотности распределения на отрезке $[-4; 4]$ для непрерывных распределений и на отрезке $[6; 14]$ для распределения Пуассона.

2. Теория

2.1. Распределения

Плотности:

- Нормальное распределение

$$f(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3)$$

- Распределение Коши

$$f(x, 0, 1) = \frac{1}{\pi [1 + x^2]} \quad (4)$$

- Распределение Лапласа

$$f(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{2\sqrt{2}} e^{-\frac{|x|}{\sqrt{2}}} \quad (5)$$

- Распределение Пуассона

$$f(k, 10) = \frac{10^k}{k!} e^{-10} \quad (6)$$

- Равномерное распределение

$$f(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & x \in [-\sqrt{3}, \sqrt{3}] \\ 0, & x \notin [-\sqrt{3}, \sqrt{3}] \end{cases} \quad (7)$$

2.2. Гистограмма

2.2.1. Построение гистограммы

Множество значений, которое может принимать элемент выборки, разбивается на несколько интервалов. Чаще всего эти интервалы берут одинаковыми, но это не является строгим требованием. Эти интервалы откладываются на горизонтальной оси, затем над каждым рисуется прямоугольник. Если все интервалы были одинаковыми, то высота каждого прямоугольника пропорциональна числу элементов выборки, попадающих в соответствующий интервал. Если интервалы разные, то высота прямоугольника выбирается таким образом, чтобы его площадь была пропорциональна числу элементов выборки, которые попали в этот интервал [1].

2.3. Вариационный ряд

Вариационным ряд - последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются [2, с. 409].

2.4. Выборочные числовые характеристики

Для того, чтобы анализировать получаемые данные, вводятся характеристики положения – такие функции, которые различными способами усредняют данные и демонстрируют общие закономерности, и характеристики рассеяния – это характеристики, которые показывают насколько данные разбросаны относительно своих характеристик положения.

2.4.1. Характеристики положения

- Выборочное среднее

$$\bar{x} = \sum_{i=1}^n x_i \quad (8)$$

- Выборочная медиана

$$medx = \begin{cases} x_{(l+1)}, & n = 2l + 1 \\ \frac{x_{(l)} + x_{(l+1)}}{2}, & n = 2l \end{cases} \quad (9)$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \quad (10)$$

- Полусумма квартилей

Выборочная квартиль порядка p определяется как:

$$z_p = \begin{cases} x_{([np]+1)}, & np \in \mathbb{Q} \setminus \mathbb{Z} \\ x_{(np)}, & np \in \mathbb{Z} \end{cases}$$

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (11)$$

- Усечённое среднее

$$z_{tr} = \frac{1}{n - 2r} \sum_{i=r+1}^{n-r} x_{(i)}, r \approx \frac{n}{4} \quad (12)$$

2.4.2. Характеристики рассеяния

- Выборочная дисперсия

$$D(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (13)$$

2.5. Боксплот Тьюки

2.5.1. Построение

Границами ящика – первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выбросов). Длина «усов»:

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1) \quad (14)$$

где X_1 — нижняя граница уса, X_2 — верхняя граница уса, Q_1 — первый квартиль, Q_3 — третий квартиль. Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков [3].

2.6. Теоретическая вероятность выбросов

Выбросами считаются величины, удовлетворяющие условию:

$$\begin{cases} x < X_1 \\ x > X_2 \end{cases} \quad (15)$$

Для теоретических распределений эту величину можно получить аналитически:

$$P_{\text{в}}^T = P(x < X_1 | x > X_2) = F_X(X_1) + 1 - F_X(X_2) \quad (16)$$

где F_X — функция распределения.

Для экспериментальных — непосредственным вычислением: итерируясь по всем элементам сгенерированного распределения, увеличивать счётчик выбросов в соответствующем случае. В конце поделить полученное количество выбросов на общее число элементов в выборке.

2.7. Эмпирическая функция распределения

2.7.1. Статистический ряд

Статистический ряд – это упорядоченная по возрастанию последовательность *различных* элементов выборки $\{z_i\}_{i=\overline{1,k}}$ с указанием частот $\{n_i\}_{i=\overline{1,k}}$, с которыми элементы с соответствующим индексом встречаются в исходной выборке.

| | | | | |
|-----|-------|-------|---------|-------|
| z | z_1 | z_2 | \dots | z_k |
| n | n_1 | n_2 | \dots | n_k |

2.7.2. Определение

Эмпирической функцией распределения называется функция, которая заданному вещественному числу x сопоставляет относительную частоту события $X \leq x$, полученную по данной выборке:

$$F^*(x) = P^*(X \leq x) \quad (17)$$

2.7.3. Вычисление

Для того чтобы посчитать эмпирическую функцию вероятности в заданной точке, можно построить статистический ряд и просуммировать частоты, с которыми встречаются все элементы, меньшие, чем x – таким образом мы получим количество возможных событий $X \leq x$. Для вычисления относительной частоты остаётся разделить полученное значение на общее количество событий.

$$F^*(x) = P^*(X \leq x) = \frac{1}{n} \sum_{z_i \leq x} n_i \quad (18)$$

где n – число событий

2.8. Оценки плотности вероятности

2.8.1. Определение

Оценкой плотности вероятности называется функция $\hat{f}(x)$, построенная по выборке и приближённо равная плотности вероятности:

$$\hat{f}(x) \approx f(x) \quad (19)$$

2.8.2. Ядерные оценки

Будем оценивать плотность вероятности следующим образом:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) \quad (20)$$

Здесь $K(u)$ – ядро: функция, обладающая свойствами:

$$K \in L^1 \quad (21)$$

$$K(u) \geq 0 \quad (21')$$

$$\|K\|_{L^1} = 1 \quad (21'')$$

Эти требования также будут выполнены и для описанной суммы: (21) – поскольку L^1 – линейное пространство, (21') – очевидно, 3 – по свойствам нормы. Таким образом, \hat{f} удовлетворяет определению плотности вероятности.

Число h_n называется шириной полосы пропускания. Последовательность $\{h_n\}_{n \in \mathbb{N}}$ должна удовлетворять следующим условиям:

$$h_n \xrightarrow{n \rightarrow \infty} 0 \quad (22)$$

$$\frac{h_n}{n^{-1}} \xrightarrow{n \rightarrow \infty} \infty \quad (22')$$

При выполнении всех условий данные оценки называются ядерными.

Ядерная оценка является состоятельной в том смысле, что сходится по распределению к плотности вероятности случайной величины ([4, стр. 38, ф-ла 4]).

В качестве ядра будем использовать нормальное ядро:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (23)$$

Для вычисления ширины полосы пропускания будем использовать эмпирическое правило Сильвермана:

$$h_n = 1.06\hat{\sigma}n^{-1/5}, \quad (24)$$

где $\hat{\sigma}$ – выборочное стандартное отклонение.

3. Реализация

Данная работа реализована на языке программирования Python с использованием редактора VIM и библиотек NumPy, Matplotlib, Statsmodels, Scipy, Seaborn в ОС Ubuntu 16.04.

Отчёт подготовлен с помощью компилятора pdf_latex и среды разработки TeXStudio.

4. Результаты

4.1. Гистограммы

На нижележащих рисунках изображены гистограммы распределений и теоретические функции плотности распределения

- Нормальное распределение

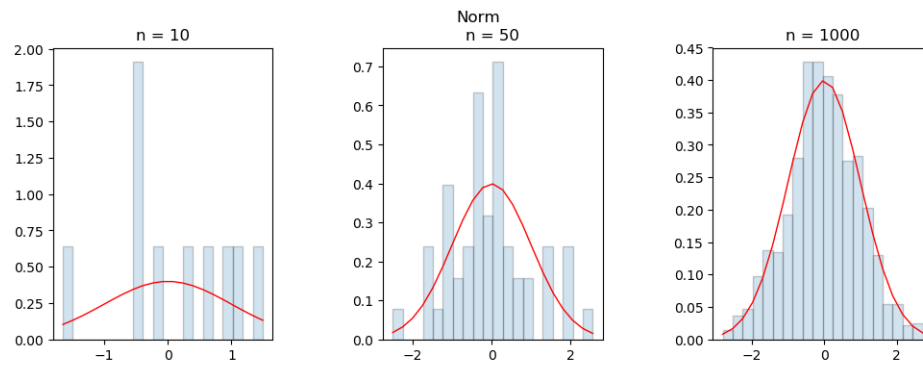


Рис. 1. Гистограмма и плотность вероятности для нормального распределения

- Распределение Коши

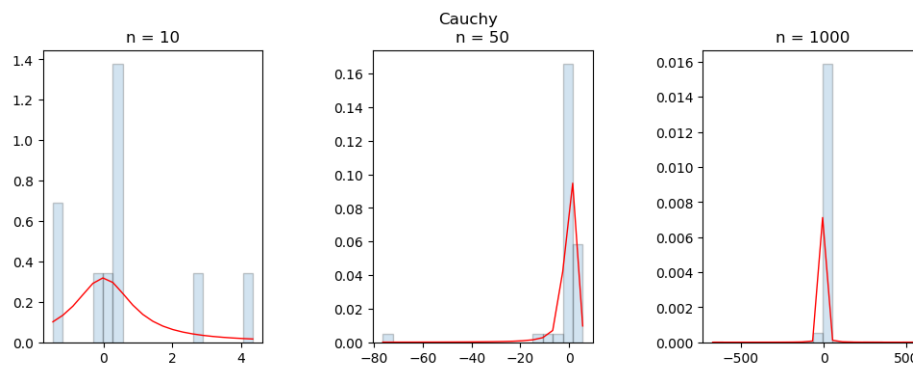


Рис. 2. Гистограмма и плотность вероятности для распределения Коши

- Распределение Лапласа

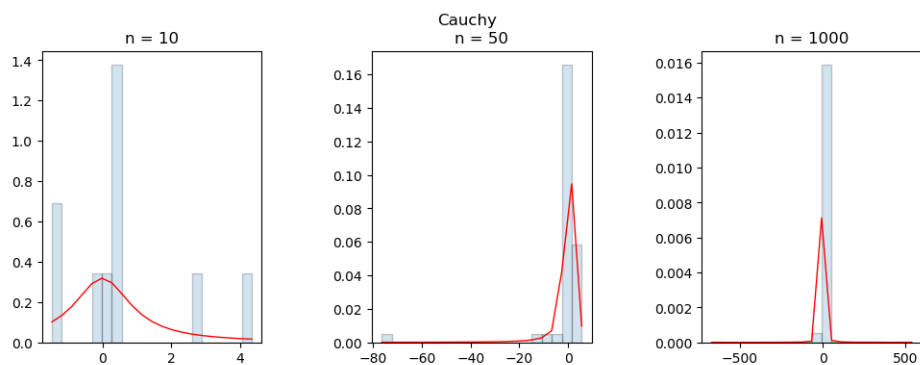


Рис. 3. Гистограмма и плотность вероятности для распределения Лапласа

- Распределение Пуассона

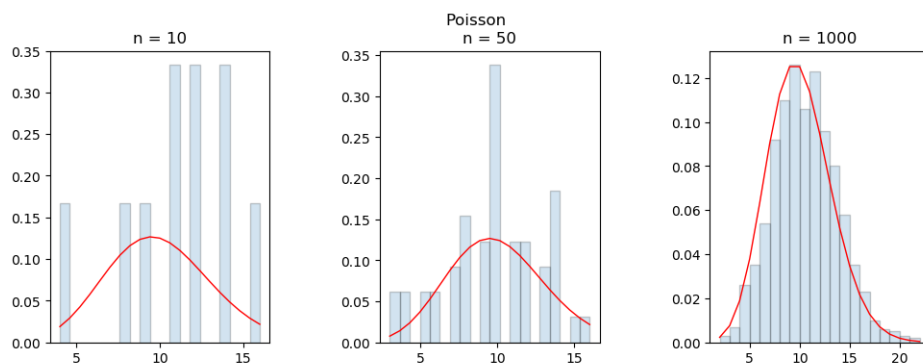


Рис. 4. Гистограмма и плотность вероятности для распределения Пуассона

- Равномерное распределение

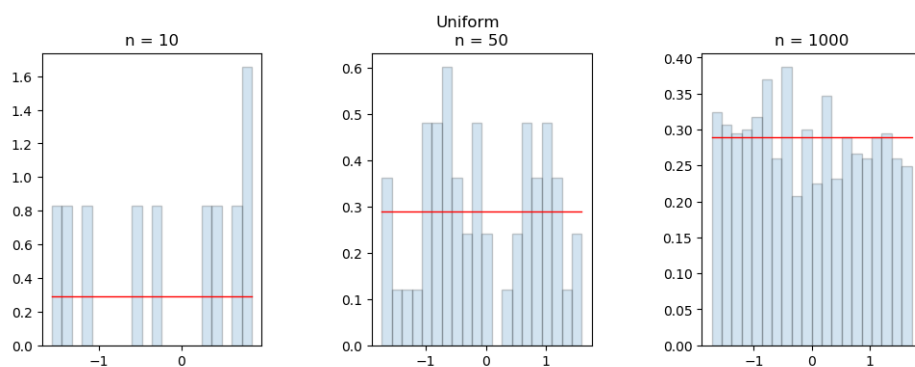


Рис. 5. Гистограмма и плотность вероятности для равномерного распределения

4.2. Характеристики положения и рассеяния

Ниже представлены таблицы полученных характеристик положения заданных распределений и характеристик рассеяния для совокупности 1000 прове-

дѐнных экспериментов

| | \bar{x} (8) | $medx$ (9) | z_R (10) | z_Q (11) | z_{tr} (12) |
|-------------|---------------|------------|------------|------------|---------------|
| Normal 10 | | | | | |
| $E(z)$ (1) | 0.0 | 0.0 | 0.0 | 0.3 | 0.0 |
| $D(z)$ (2) | 0.1 | 0.1 | 0.2 | 0.1 | 0.1 |
| Normal 100 | | | | | |
| $E(z)$ (1) | 0.00 | 0.00 | 0.00 | 0.04 | -0.01 |
| $D(z)$ (2) | 0.01 | 0.02 | 0.10 | 0.01 | 0.01 |
| Normal 1000 | | | | | |
| $E(z)$ (1) | 0.00033 | 0.001 | 0.01 | 0.005 | 0.001 |
| $D(z)$ (2) | 0.00010 | 0.002 | 0.06 | 0.001 | 0.001 |

Таблица 1. Нормальное распределение

| | \bar{x} (8) | $medx$ (9) | z_R (10) | z_Q (11) | z_{tr} (12) |
|-------------|---------------|------------|-------------|------------|---------------|
| Cauchy 10 | | | | | |
| $E(z)$ (1) | 0.0 | 0.0 | 0.2 | 1.2 | 0.0 |
| $D(z)$ (2) | 246.8 | 0.3 | 5918.0 | 5.4 | 0.6 |
| Cauchy 100 | | | | | |
| $E(z)$ (1) | 1.4 | 0.00 | 70.7 | 0.09 | 0.00 |
| $D(z)$ (2) | 11369.5 | 0.03 | 28386516.2 | 0.06 | 0.03 |
| Cauchy 1000 | | | | | |
| $E(z)$ (1) | -0.7 | 0.001 | -354.9 | 0.010 | 0.001 |
| $D(z)$ (2) | 442.9 | 0.002 | 108911495.3 | 0.005 | 0.003 |

Таблица 2. Распределение Коши

| | \bar{x} (8) | $medx$ (9) | z_R (10) | z_Q (11) | z_{tr} (12) |
|--------------|---------------|------------|------------|------------|---------------|
| Laplace 10 | | | | | |
| $E(z)$ (1) | -0.02 | 0.00 | 0.0 | 0.3 | -0.01 |
| $D(z)$ (2) | 0.09 | 0.07 | 0.4 | 0.1 | 0.07 |
| Laplace 100 | | | | | |
| $E(z)$ (1) | -0.004 | -0.001 | 0.0 | 0.038 | -0.003 |
| $D(z)$ (2) | 0.010 | 0.006 | 0.4 | 0.009 | 0.006 |
| Laplace 1000 | | | | | |
| $E(z)$ (1) | -0.0004 | -0.0002 | 0.0 | 0.000 | -0.0003 |
| $D(z)$ (2) | 0.0010 | 0.0005 | 0.4 | 0.001 | 0.0006 |

Таблица 3. Распределение Лапласа

| | \bar{x} (8) | $medx$ (9) | z_R (10) | z_Q (11) | z_{tr} (12) |
|--------------|---------------|------------|------------|------------|---------------|
| Poisson 10 | | | | | |
| $E(z)$ (1) | 10.0 | 9.9 | 10.3 | 10.9 | 9.89 |
| $D(z)$ (2) | 1.0 | 1.4 | 2.0 | 1.4 | 1.09 |
| Poisson 100 | | | | | |
| $E(z)$ (1) | 10.0 | 9.8 | 10.9 | 10.0 | 9.8 |
| $D(z)$ (2) | 0.1 | 0.2 | 1.0 | 0.2 | 0.1 |
| Poisson 1000 | | | | | |
| $E(z)$ (1) | 10.00 | 9.993 | 11.7 | 9.993 | 9.85 |
| $D(z)$ (2) | 0.01 | 0.007 | 0.7 | 0.004 | 0.01 |

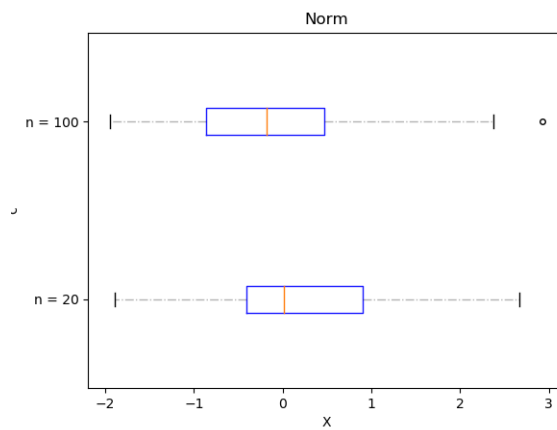
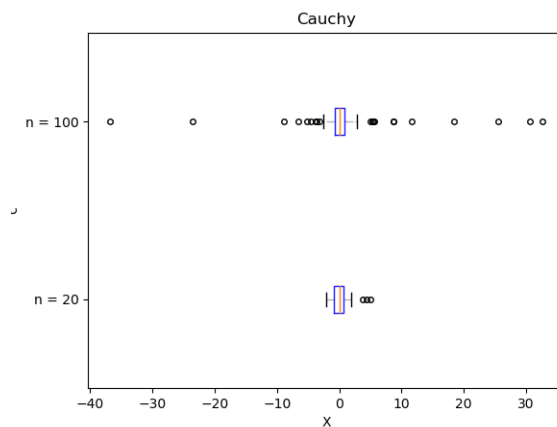
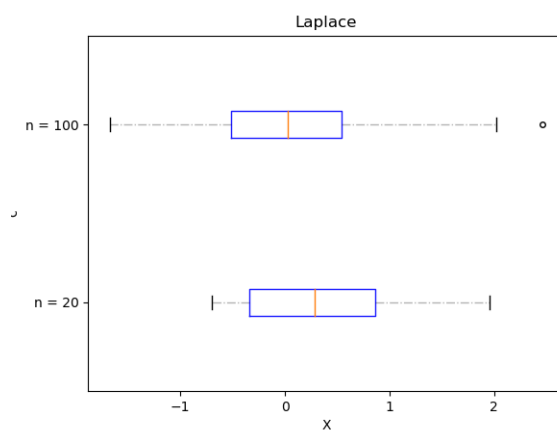
Таблица 4. Распределение Пуассона

| | \bar{x} (8) | $medx$ (9) | z_R (10) | z_Q (11) | z_{tr} (12) |
|--------------|---------------|------------|------------|------------|---------------|
| Uniform 10 | | | | | |
| $E(z)$ (1) | 0.0 | 0.0 | 0.01 | 0.3 | 0.0 |
| $D(z)$ (2) | 0.1 | 0.2 | 0.05 | 0.1 | 0.2 |
| Uniform 100 | | | | | |
| $E(z)$ (1) | 0.00 | 0.00 | 0.0007 | 0.05 | 0.00 |
| $D(z)$ (2) | 0.01 | 0.03 | 0.0005 | 0.02 | 0.02 |
| Uniform 1000 | | | | | |
| $E(z)$ (1) | 0.0003 | 0.0005 | -0.000161 | 0.005 | 0.000 |
| $D(z)$ (2) | 0.0010 | 0.0028 | 0.000006 | 0.001 | 0.002 |

Таблица 5. Равномерное распределение

4.3. Боксплот Тьюки

Ниже представлены таблицы полученных характеристик положения заданных распределений и характеристик рассеяния для совокупности 1000 проведённых экспериментов

**Рис. 6.** Нормальное распределение**Рис. 7.** Распределение Коши**Рис. 8.** Распределение Лапласа

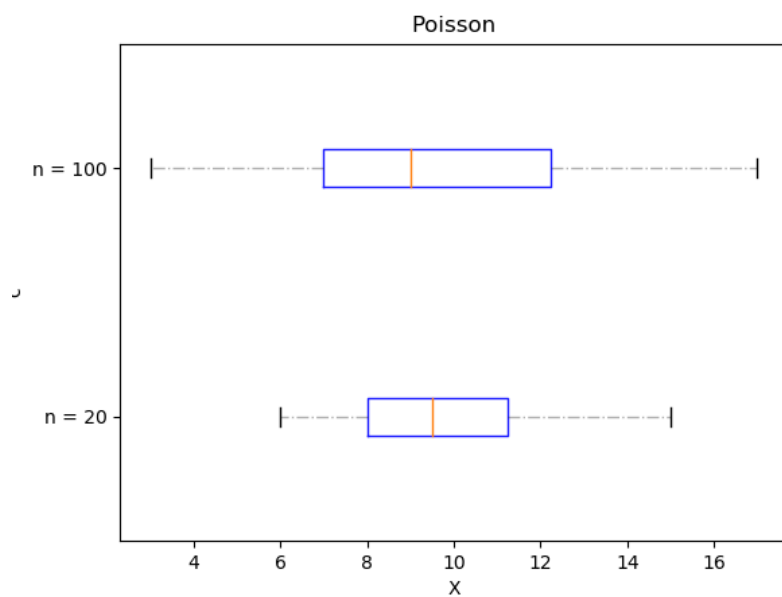


Рис. 9. Распределение Пуассона

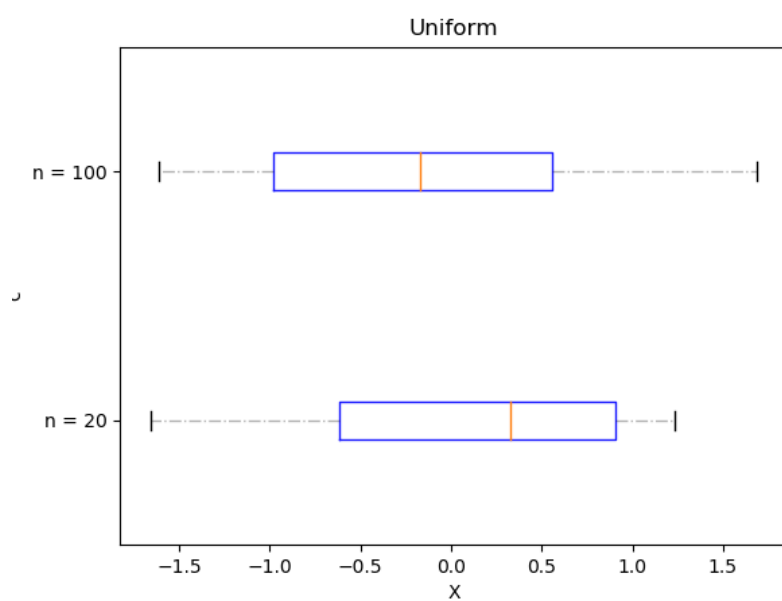


Рис. 10. Равномерное распределение

4.4. Экспериментальная доля выбросов

| Выборка | Доля выбросов |
|--------------|---------------|
| Normal, 20 | 0.02 |
| Normal, 100 | 0.01 |
| Cauchy, 20 | 0.13 |
| Cauchy, 100 | 0.16 |
| Laplace, 20 | 0.06 |
| Laplace, 100 | 0.06 |
| Poisson, 20 | 0.01 |
| Poisson, 100 | 0.01 |
| Uniform, 20 | 0.00 |
| Uniform, 100 | 0.00 |

Таблица 6. Экспериментальная доля выбросов

4.5. Теоретическая вероятность выбросов

| Распределение | $P_v^T(16)$ |
|---------------|-------------|
| Normal | 0.007 |
| Cauchy | 0.156 |
| Laplace | 0.063 |
| Poisson | 0.008 |
| Uniform | 0 |

Таблица 7. Теоретическая доля выбросов

4.6. Эмпирическая функция распределения

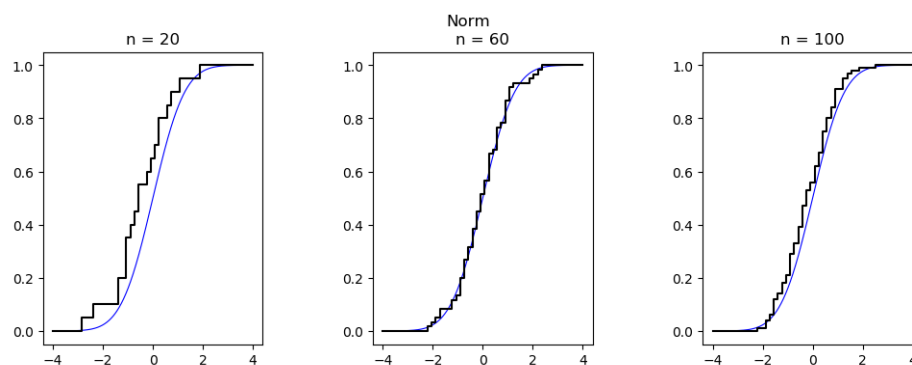


Рис. 11. Нормальное распределение

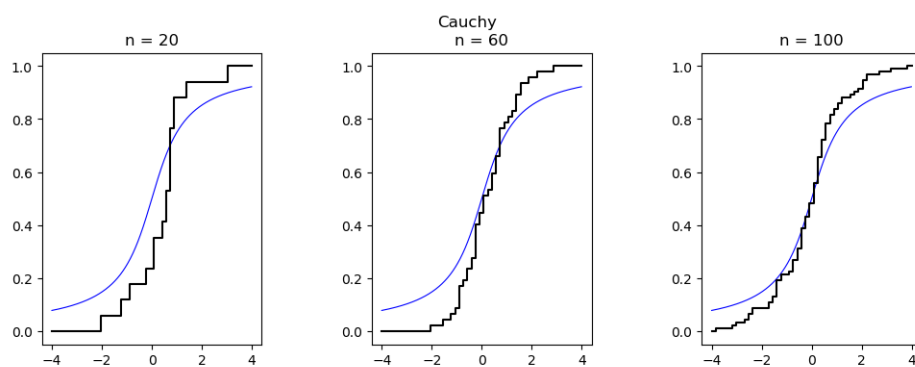


Рис. 12. Распределение Коши

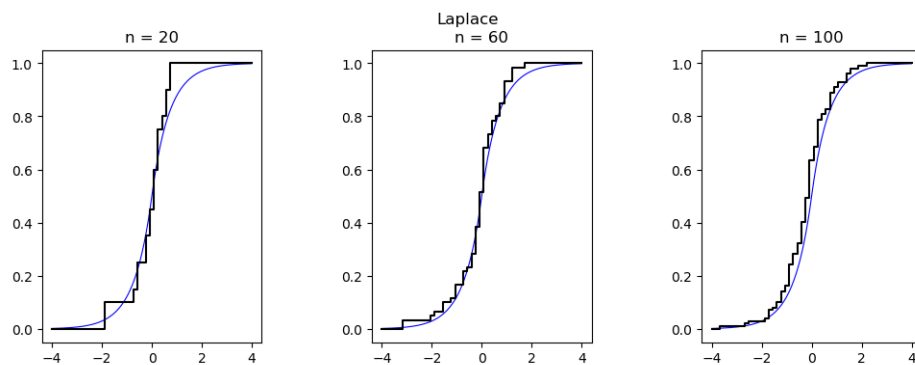


Рис. 13. Распределение Лапласа

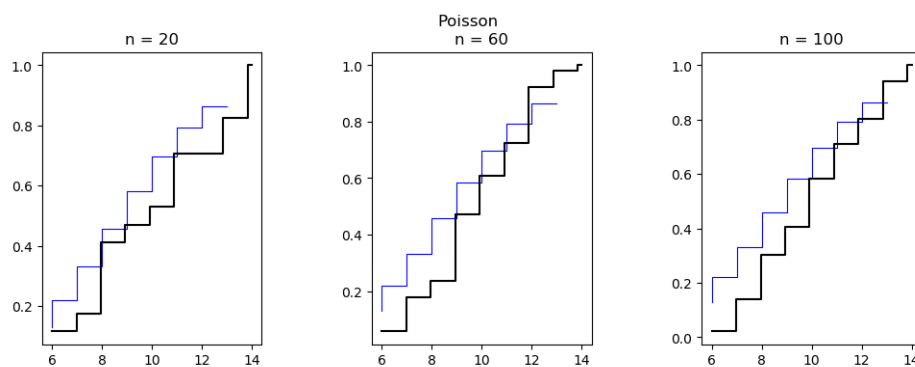


Рис. 14. Распределение Пуассона

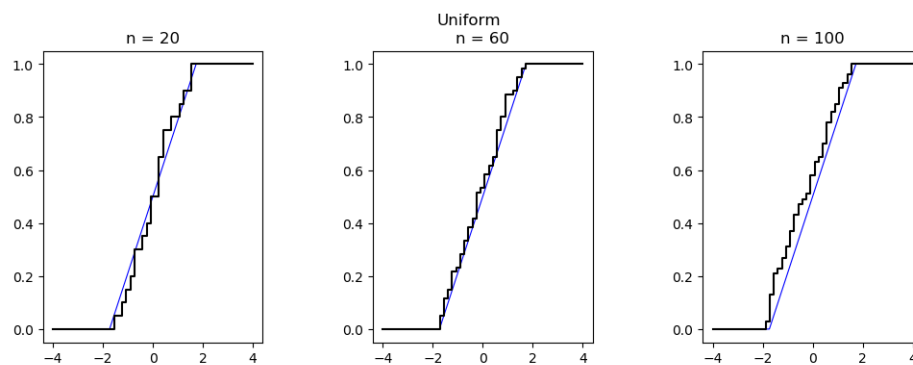
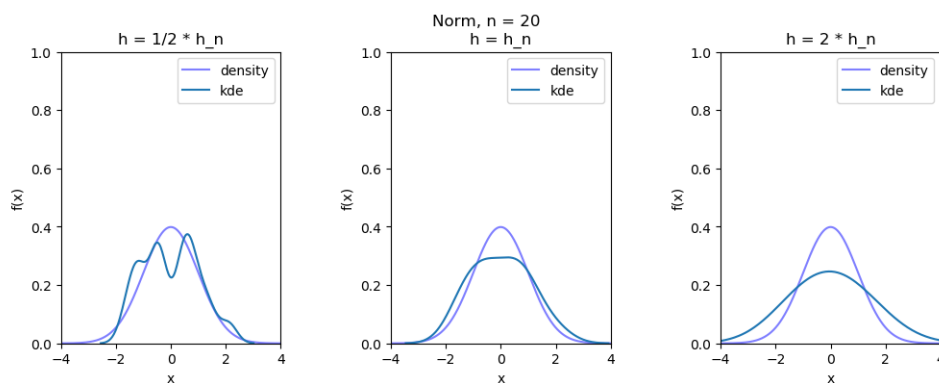
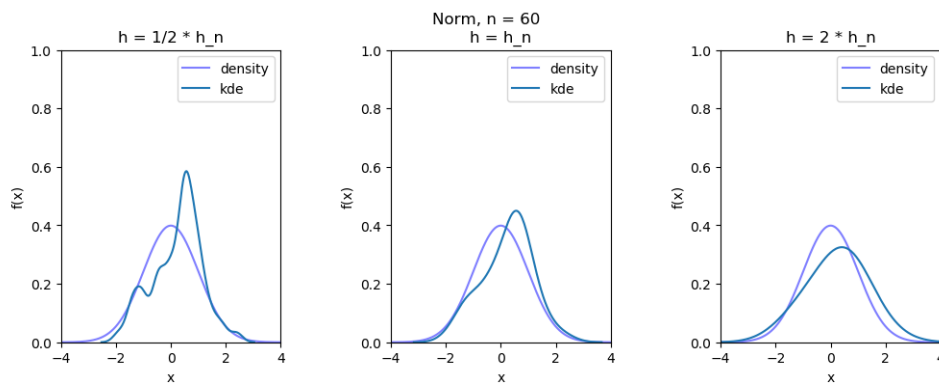


Рис. 15. Равномерное распределение

4.7. Ядерные оценки плотности распределения

Рис. 16. Нормальное распределение, $n = 20$ Рис. 17. Нормальное распределение, $n = 60$

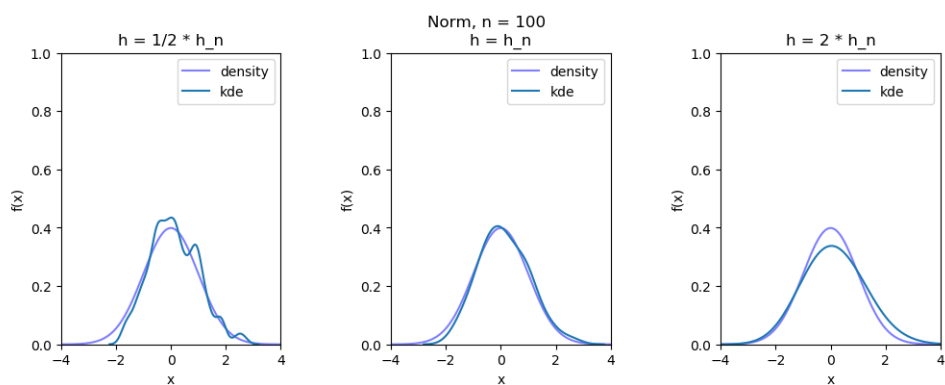


Рис. 18. Нормальное распределение, $n = 100$

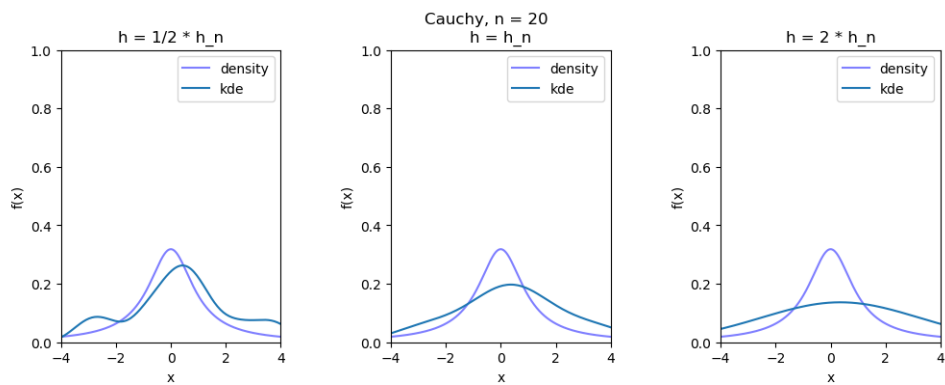


Рис. 19. Распределение Коши, $n = 20$

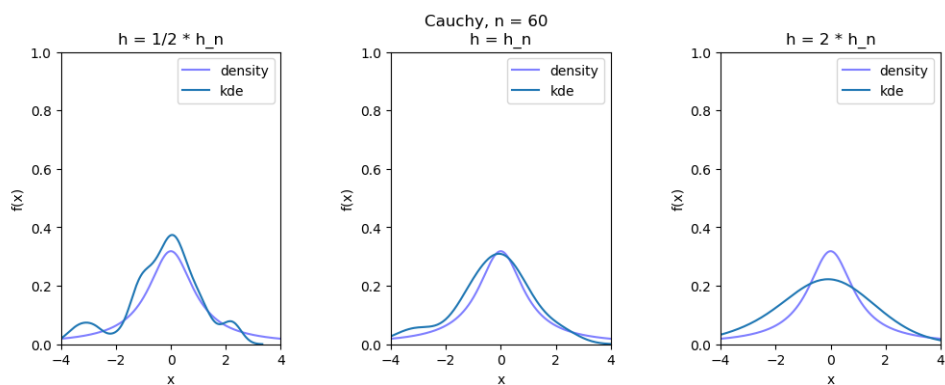
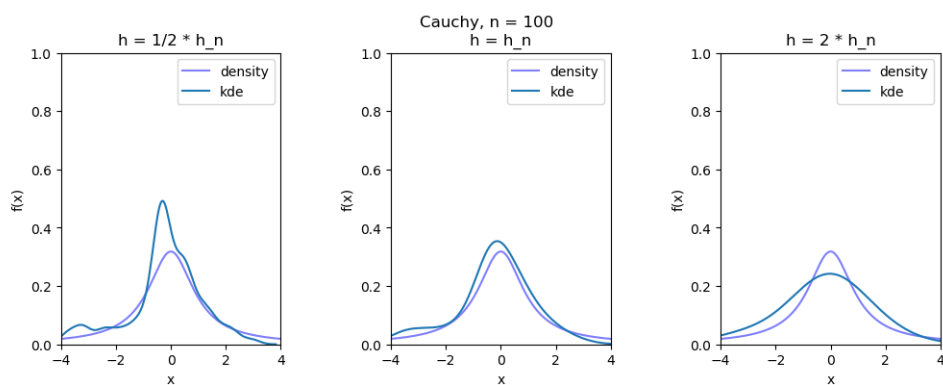
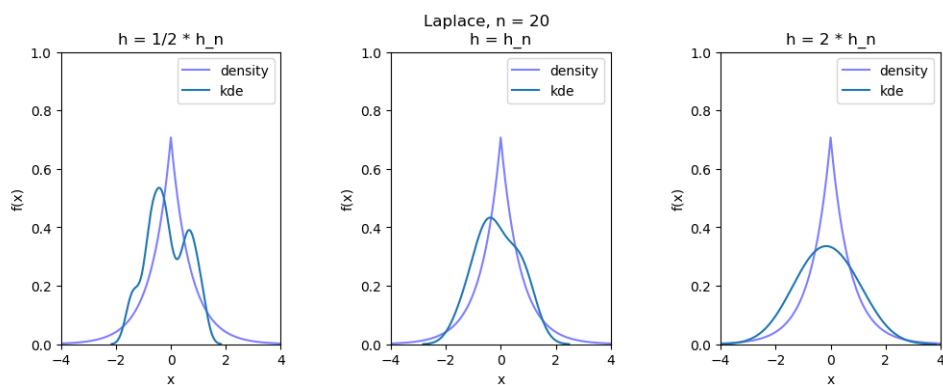
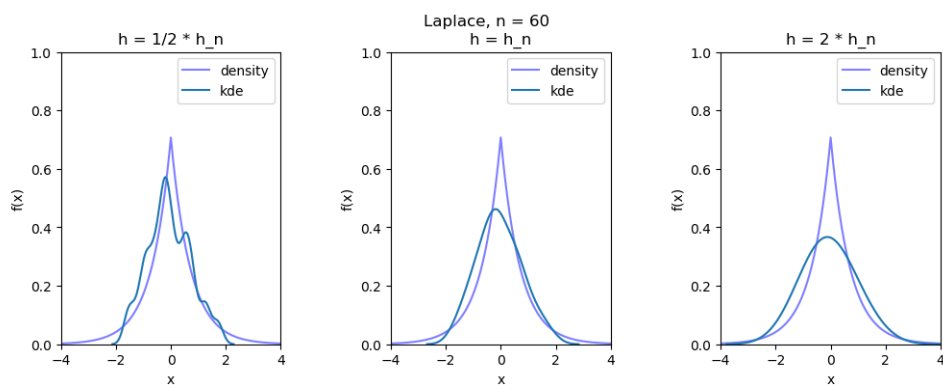


Рис. 20. Распределение Коши, $n = 60$

Рис. 21. Распределение Коши, $n = 100$ Рис. 22. Распределение Лапласа, $n = 20$ Рис. 23. Распределение Лапласа, $n = 60$

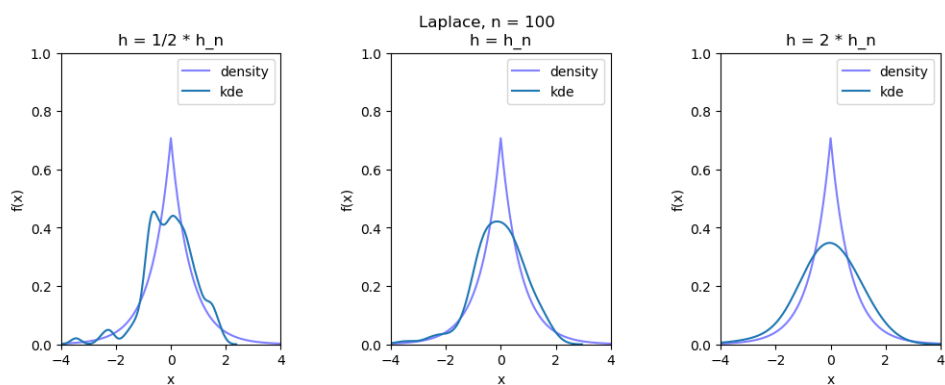


Рис. 24. Распределение Лапласа, $n = 100$

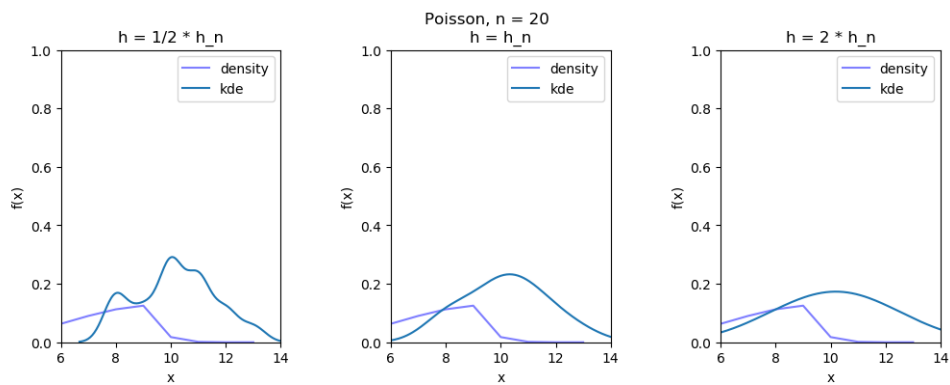


Рис. 25. Распределение Пуассона, $n = 20$

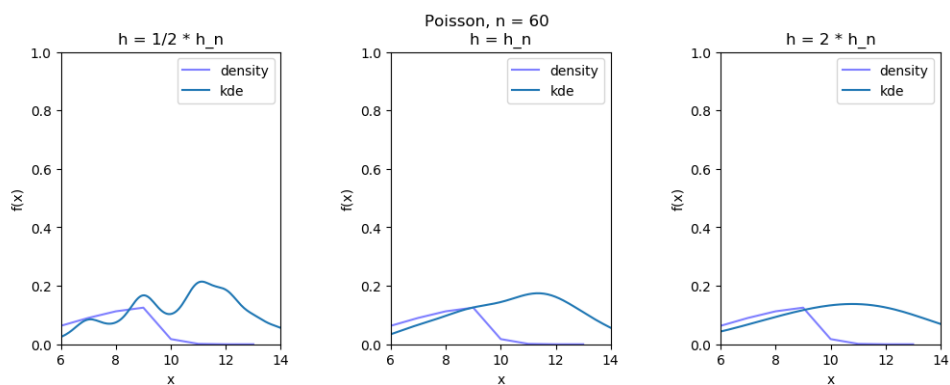
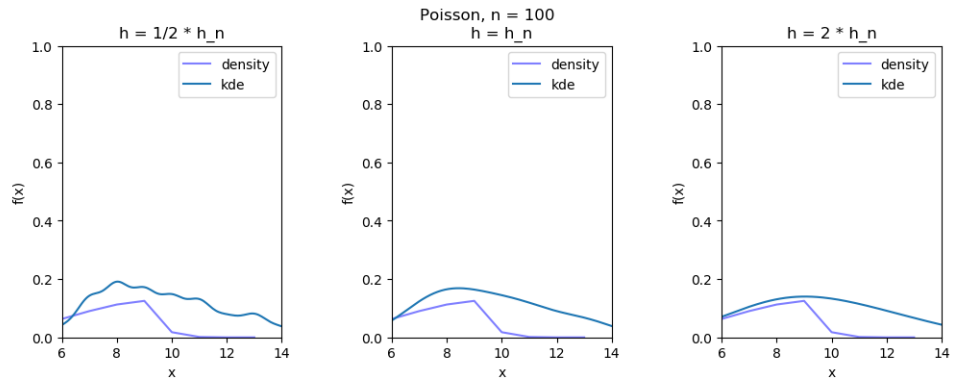
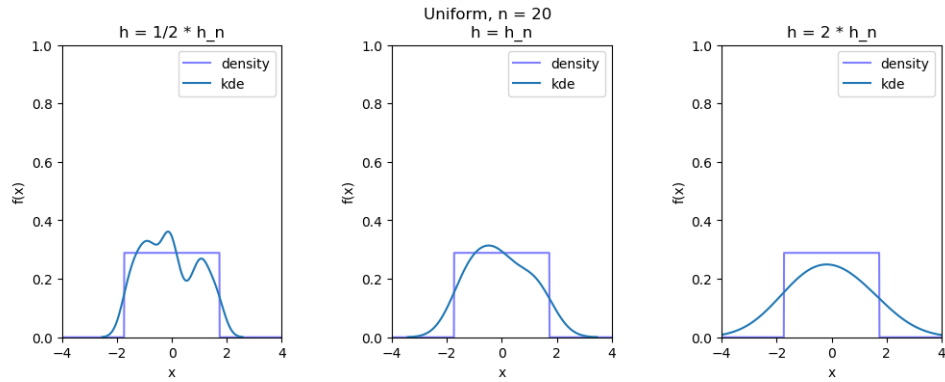
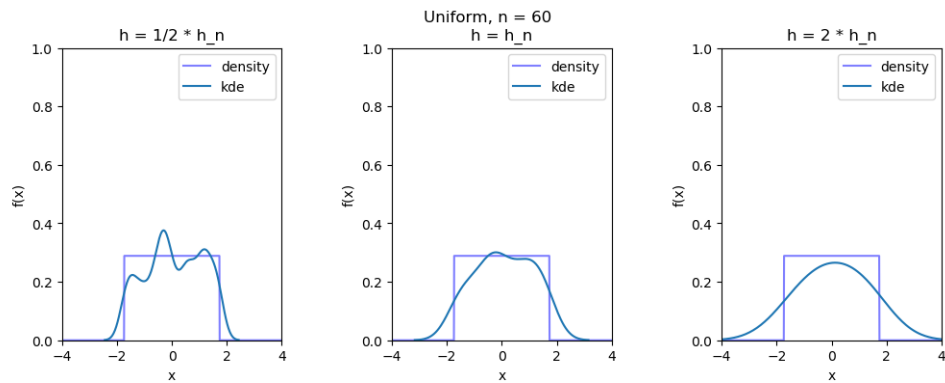


Рис. 26. Распределение Пуассона, $n = 60$

Рис. 27. Распределение Пуассона, $n = 100$ Рис. 28. Равномерное распределение, $n = 20$ Рис. 29. Равномерное распределение, $n = 60$

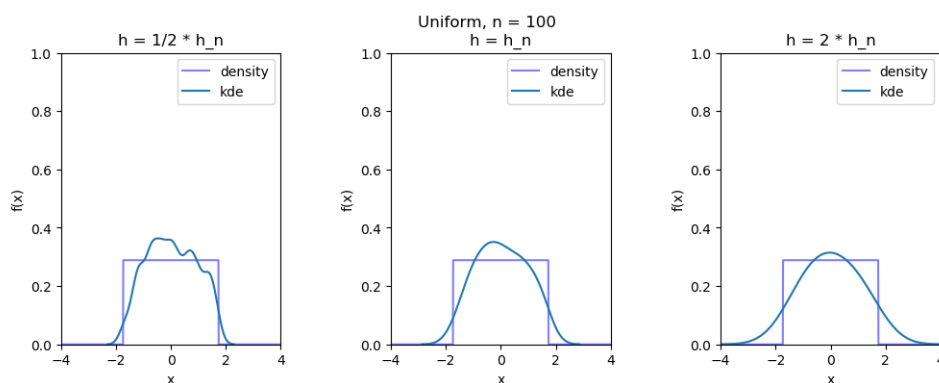


Рис. 30. Равномерное распределение, $n = 100$

5. Обсуждение

5.1. Гистограммы

По результатам проделанной работы можем сделать вывод о том, что чем больше выборка для каждого из распределений, тем ближе ее гистограмма к графику плотности вероятности того закона, по которому распределены величины сгенерированной выборки. Чем меньше выборка, тем менее она показательна - тем хуже по ней определяется характер распределения величины.

Также можно заметить, что максимумы гистограмм и плотностей распределения почти нигде не совпали. Также наблюдаются всплески гистограмм, что наиболее хорошо прослеживается на распределении Коши

5.2. Характеристики положения и рассеяния

Исходя из данных, приведенных в таблицах, можно судить о том, что дисперсия характеристик рассеяния для распределения Коши является некой аномалией: значения слишком большие даже при увеличении размера выборки - понятно, что это результат выбросов, которые мы могли наблюдать в результатах предыдущего задания.

5.3. Боксплот Тьюки

По данным, приведенным в таблице, можно сказать, что чем больше выборка, тем ближе доля выбросов будет к теоретической оценке. Снова доля выбросов для распределения Коши значительно выше, чем для остальных распределений. Равномерное распределение же в точности повторяет теоретическую оценку - выбросов мы не получали.

Боксплоты Тьюки действительно позволяют более наглядно и с меньшими усилиями оценивать важные характеристики распределений. Так, исходя из полученных рисунков, наглядно видно то, что мы довольно трудоёмко анализировали в предыдущих частях.

5.4. Эмпирическая функция распределения. Ядерные оценки плотности

Можем наблюдать на иллюстрациях с э. ф. р., что ступенчатая эмпирическая функция распределения тем лучше приближает функцию распределения реальной выборки, чем мощнее эта выборка. Заметим так же, что для распределения Пуассона и равномерного распределения отклонение функций друг от друга наибольшее.

Рисунки, посвященные ядерным оценкам, иллюстрируют сближение ядерной оценки и функции плотности вероятности для всех h с ростом размера выборки. Для распределения Пуассона наиболее ярко видно, как сглаживает отклонения увеличение параметра сглаживания h .

В зависимости от особенностей распределений для их описания лучше подходят разные параметры h в ядерной оценке: для равномерного распределения и распределения Пуассона лучше подойдет параметр $h = 2h_n$, для распределения Лапласа – $h = h_n/2$, а для нормального и Коши – $h = h_n$. Такие значения дают вид ядерной оценки наиболее близкий к плотности, характерной данным распределениям.

Также можно увидеть, что чем больше коэффициент при параметре сглаживания h_n , тем меньше изменений знака производной у аппроксимирующей функции, вплоть до того, что при $h = 2h_n$ функция становится унимодальной на рассматриваемом промежутке. Также видно, что при $h = h_n/2$ по полученным приближениям становится сложно сказать плотность вероятности какого распределения они должны повторять, так как они очень похожи между собой.

6. Литература

- [1] Histogram. URL: <https://en.wikipedia.org/wiki/Histogram>
- [2] Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.
- [3] Box plot. URL: https://en.wikipedia.org/wiki/Box_plot
- [4] [Анатолий Станислав](#) (2009) "Непараметрическая регрессия", Квантиль, №7, стр. 37-52

7. Ссылка на репозиторий

Репозиторий с исходным кодом программы: <https://github.com/sairsey/MathStats>