Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики

Кафедра «Прикладная математика»

Отчёт
по лабораторным работам №1-4
по дисциплине
«Математическая статистика»

Выполнил студент:

...
группа: ...

Проверил:

к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

1 Постановка задачи **4**

2 Теория **5**
2.1 Рассматриваемые распределения 5
2.2 Гистограмма 5
2.2.1 Построение гистограммы 5
2.3 Вариационный ряд 5
2.4 Выборочные числовые характеристики 6
2.4.1 Характеристики положения 6
2.4.2 Характеристики рассеяния 6
2.5 Боксплот Тьюки 7
2.5.1 Построение 7
2.6 Теоретическая вероятность выбросов 7
2.7 Эмпирическая функция распределения 7
2.7.1 Статистический ряд 7
2.7.2 Эмпирическая функция распределения 8
2.7.3 Нахождение э. ф. р. 8
2.8 Оценки плотности вероятности 8
2.8.1 Определение 8
2.8.2 Ядерные оценки 8

3 Реализация **9**

4 Результаты **10**
4.1 Гистограмма и график плотности распределения 10
4.2 Характеристики положения и рассеяния 11
4.3 Боксплот Тьюки 14
4.4 Доля выбросов 17
4.5 Теоретическая вероятность выбросов 18
4.6 Эмпирическая функция распределения 18
4.7 Ядерные оценки плотности распределения 20

5 Обсуждение **27**
5.1 Гистограмма и график плотности распределения 27
5.2 Характеристики положения и рассеяния 27
5.3 Доля и теоретическая вероятность выбросов 27
5.4 Эмпирическая функция и ядерные оценки плотности распределения 28

Литература **28**

Список иллюстраций

1 Нормальное распределение 10
2 Распределение Коши 10
3 Распределение Лапласа 10
4 Распределение Пуассона 11
5 Равномерное распределение 11
6 Нормальное распределение 14
7 Распределение Коши 15
8 Распределение Лапласа 15
9 Распределение Пуассона 16
10 Равномерное распределение 16
11 Нормальное распределение 18
12 Распределение Коши 18
13 Распределение Лапласа 19
14 Распределение Пуассона 19
15 Равномерное распределение 19
16 Нормальное распределение, $n = 20$ 20
17 Нормальное распределение, $n = 60$ 20
18 Нормальное распределение, $n = 100$ 21
19 Распределение Коши, $n = 20$ 21
20 Распределение Коши, $n = 60$ 22
21 Распределение Коши, $n = 100$ 22
22 Распределение Лапласа, $n = 20$ 23
23 Распределение Лапласа, $n = 60$ 23
24 Распределение Лапласа, $n = 100$ 24
25 Распределение Пуассона, $n = 20$ 24
26 Распределение Пуассона, $n = 60$ 25
27 Распределение Пуассона, $n = 100$ 25
28 Равномерное распределение, $n = 20$ 26
29 Равномерное распределение, $n = 60$ 26
30 Равномерное распределение, $n = 100$ 27

Список таблиц

1 Таблица распределения 8
2 Нормальное распределение 12
3 Распределение Коши 12
4 Распределение Лапласа 13
5 Распределение Пуассона 13
6 Равномерное распределение 14
7 Доля выбросов 17
8 Теоретическая вероятность выбросов 18

1 Постановка задачи

Для 5 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: $\bar{x}, med\ x, z_R, z_Q, z_{tr}$. Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

   $$E(z) = \overline{z} \tag{1}$$

   Вычислить оценку дисперсии по формуле:

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

   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 Рассматриваемые распределения

Плотности:

- Нормальное распределение

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

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \tag{4}$$

- Распределение Лапласа

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

- Распределение Пуассона

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

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при} \quad |x| \leq \sqrt{3} \\ 0 & \text{при} \quad |x| > \sqrt{3} \end{cases} \tag{7}$$

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

2.4.1 Характеристики положения

- Выборочное среднее

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

- Выборочная медиана

$$med\, x = \begin{cases} x_{(l+1)} & \text{при} \quad n = 2l+1 \\ \dfrac{x_{(l)} + x_{(l+1)}}{2} & \text{при} \quad n = 2l \end{cases} \tag{9}$$

- Полусумма экстремальных выборочных элементов

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

- Полусумма квартилей

  Выборочная квартиль z_p порядка p определяется формулой

$$z_p = \begin{cases} x_{([np]+1)} & \text{при } \ np \ \text{дробном,} \\ x_{(np)} & \text{при } \ np \ \text{целом.} \end{cases} \tag{11}$$

  Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \tag{12}$$

- Усечённое среднее

$$z_{tr} = \frac{1}{n-2r}\sum_{i=r+1}^{n-r} x_{(i)}, \quad r \approx \frac{n}{4} \tag{13}$$

2.4.2 Характеристики рассеяния

Выборочная дисперсия

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{14}$$

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»:

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), \;\; X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \tag{15}$$

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

Можно вычислить теоретические первый и третий квартили распределений – $Q_1^{\text{т}}$ и $Q_3^{\text{т}}$. По ф-ле (15) – теоретические нижнюю и верхнюю границы уса – $X_1^{\text{т}}$ и $X_2^{\text{т}}$. Выбросы – величины x:

$$\begin{bmatrix} x < X_1^{\text{т}} \\ x > X_2^{\text{т}} \end{bmatrix} \tag{16}$$

Теоретическая вероятность выбросов:

- для непрерывных распределений

$$P_{\text{в}}^{\text{т}} = P(x < X_1^{\text{т}}) + P(x > X_2^{\text{т}}) = F(X_1^{\text{т}}) + \Big(1 - F(X_2^{\text{т}})\Big). \tag{17}$$

- для дискретных распределений

$$P_{\text{в}}^{\text{т}} = P(x < X_1^{\text{т}}) + P(x > X_2^{\text{т}}) = \Big(F(X_1^{\text{т}}) - P(x = X_1^{\text{т}})\Big) + \Big(1 - F(X_2^{\text{т}})\Big). \tag{18}$$

Выше $F(X) = P(x \leq X)$ - функция распределения.

2.7 Эмпирическая функция распределения

2.7.1 Статистический ряд

Статистическим ряд – последовательность различных элементов выборки $z_1, z_2, \ldots, z_k$, расположенных в возрастающем порядке с указанием частот $n_1, n_2, \ldots, n_k$, с которыми эти элементы содержатся в выборке. Обычно записывается в виде таблицы.

2.7.2 Эмпирическая функция распределения

Эмпирическая (выборочная) функция распределения (э. ф. р.) – относительная частота события $X < x$, полученная по данной выборке:

$$F_n^*(x) = P^*(X < x). \tag{19}$$

2.7.3 Нахождение э. ф. р.

Для получения относительной частоты $P^*(X < x)$ просуммируем в статистическом ряде, построенном по данной выборке, все частоты n_i, для которых элементы z_i статистического ряда меньше x. Тогда $P^*(X < x) = \frac{1}{n} \sum\limits_{z_i<x} n_i$. Получаем

$$F^*(x) = \tfrac{1}{n} \sum_{z_i<x} n_i. \tag{20}$$

$F^*(x)$ — функция распределения дискретной случайной величины X^*, заданной таблицей распределения

X^*	z_1	z_2	...	z_k
P	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_k}{n}$

Таблица 1: Таблица распределения

Эмпирическая функция распределения является оценкой, т. е. приближённым значением, генеральной функции распределения

$$F_n^*(x) \approx F_X(x). \tag{21}$$

2.8 Оценки плотности вероятности

2.8.1 Определение

Оценкой плотности вероятности $f(x)$ называется функция $\widehat{f}(x)$, построенная на основе выборки, приближённо равная $f(x)$

$$\widehat{f}(x) \approx f(x). \tag{22}$$

2.8.2 Ядерные оценки

Представим оценку в виде суммы с числом слагаемых, равным объёму выборки:

$$\widehat{f}_n(x) = \tfrac{1}{nh_n} \sum_{i=1}^{n} K\left(\tfrac{x-x_i}{h_n}\right). \tag{23}$$

Здесь функция $K(u)$, называемая ядерной (ядром), непрерывна и является плотностью вероятности, x_1, ... , x_n — элементы выборки, $\{h_n\}$ — любая последовательность положительных чисел, обладающая свойствами

$$h_n \underset{n\to\infty}{\longrightarrow} 0; \qquad \frac{h_n}{n^{-1}} \underset{n\to\infty}{\longrightarrow} \infty. \tag{24}$$

Такие оценки называются непрерывными ядерными [2, с. 421-423].

Гауссово (нормальное) ядро [4, с. 38]

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}. \tag{25}$$

Правило Сильвермана [4, с. 44]

$$h_n = 1.06\hat{\sigma}n^{-1/5}, \tag{26}$$

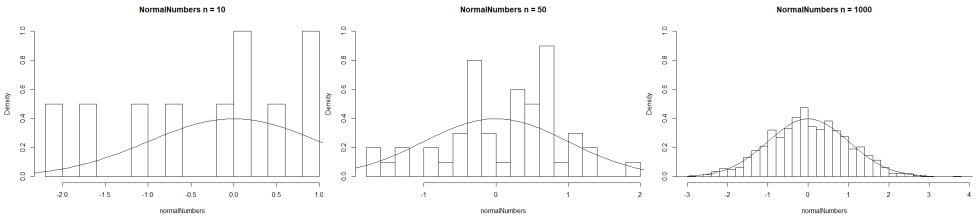
где $\hat{\sigma}$ - выборочное стандартное отклонение.

3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования R в среде разработки RStudio. Исходный код лабораторной работы приведён в приложении.

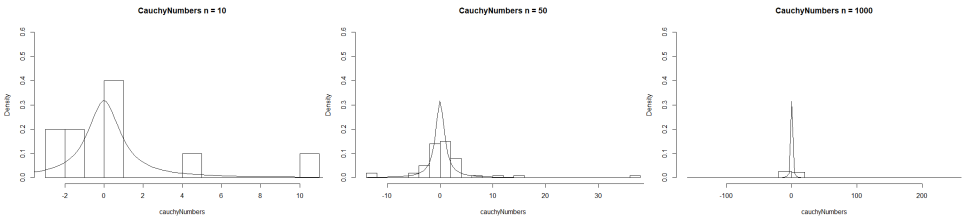
4 Результаты

4.1 Гистограмма и график плотности распределения



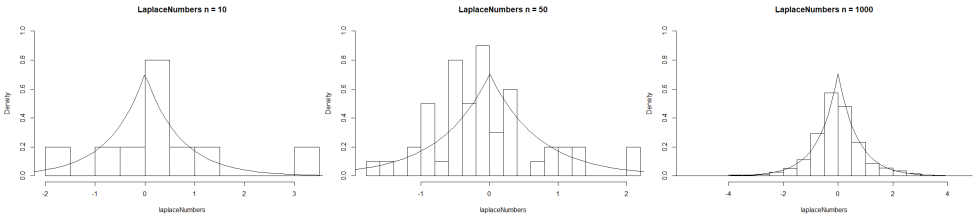
$n = 10$ $n = 50$ $n = 1000$

Рис. 1: Нормальное распределение



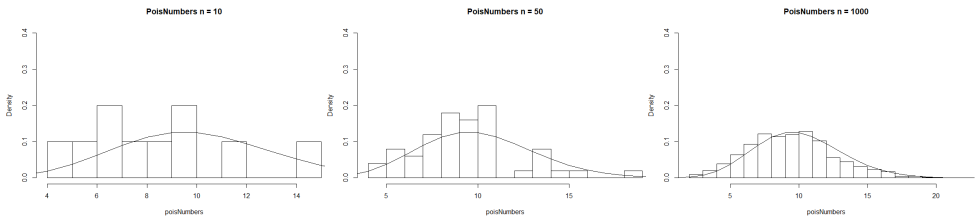
$n = 10$ $n = 50$ $n = 1000$

Рис. 2: Распределение Коши



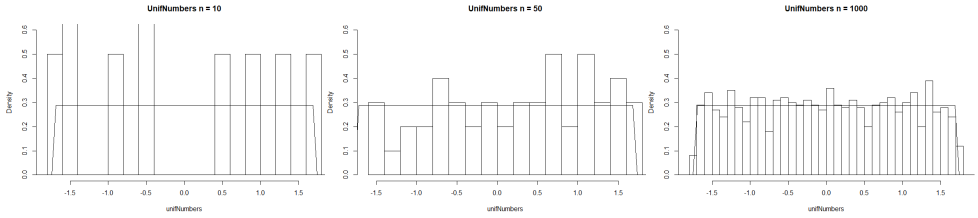
$n = 10$ $n = 50$ $n = 1000$

Рис. 3: Распределение Лапласа



$n = 10$ $n = 50$ $n = 1000$

Рис. 4: Распределение Пуассона



$n = 10$ $n = 50$ $n = 1000$

Рис. 5: Равномерное распределение

4.2 Характеристики положения и рассеяния

Как было проведено округление:
В оценке $x = E \pm D$ вариации подлежит первая цифра после точки.
В данном случае $x = 0.0 \pm 0.1k$,
k − зависит от доверительной вероятности и вида распределения (рассматривается в дальнейшем цикле лабораторных работ)
Округление сделано для $k = 1$

normal n = 10					
	$\bar{x}$ (8)	$med\, x$ (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$ (1)	0.012	0.017	0.014	0.007	0.013
$D(z)$ (2)	0.097	0.136	0.201	0.112	0.121
normal n = 100					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	0.002	-0.002	0.016	-0.014	0.000
$D(z)$	0.010	0.016	0.085	0.013	0.012
normal n = 1000					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	0.000894	0.000890	-0.004452	-0.000313	0.001087
$D(z)$	0.00095	0.00150	0.06169	0.00123	0.00116

Таблица 2: Нормальное распределение

cauchy n = 10					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	0.961	0.015	4.624	0.070	0.022
$D(z)$	857.404	0.301	21094.729	1.202	0.338
cauchy n = 100					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	1.1034	-0.0071	52.9826	-0.0393	-0.0067
$D(z)$	370.392	0.024	899464.699	0.054	0.026
cauchy n = 1000					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	-0.567	-0.002	-303.307	-0.004	-0.001
$D(z)$	1192.6058	0.0026	297518700	0.0054	0.0028

Таблица 3: Распределение Коши

laplace n = 10					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	-0.008	-0.002	-0.021	-0.000	-0.001
$D(z)$	0.098	0.070	0.402	0.099	0.067
laplace n = 100					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	-0.004	-0.001	-0.046	-0.015	-0.002
$D(z)$	0.011	0.006	0.416	0.010	0.007
laplace n = 1000					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0012	-0.0004	-0.0038	-0.0025	-0.0010
$D(z)$	0.00102	0.00053	0.40565	0.00104	0.00064

Таблица 4: Распределение Лапласа

pois n = 10					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	10.017	9.858	10.329	9.952	9.872
$D(z)$	1.087	1.607	1.923	1.351	1.387
pois n = 100					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	9.986	9.828	10.924	9.854	9.843
$D(z)$	0.095	0.223	0.987	0.142	0.115
pois n = 1000					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	9.999	9.996	11.636	9.995	9.857
$D(z)$	0.010	0.004	0.596	0.003	0.011

Таблица 5: Распределение Пуассона

unif n = 10					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	0.0076	0.0084	0.0079	0.0063	0.0113
$D(z)$	0.099	0.221	0.046	0.134	0.188
unif n = 100					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	0.004	0.008	0.000	-0.014	0.007
$D(z)$	0.009	0.028	0.001	0.014	0.018
unif n = 1000					
	$\bar{x}$	$med\, x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0012	-0.0019	0.0001	-0.0033	-0.0020
$D(z)$	0.0009	0.0027	0.0000	0.0015	0.0018

Таблица 6: Равномерное распределение

4.3 Боксплот Тьюки

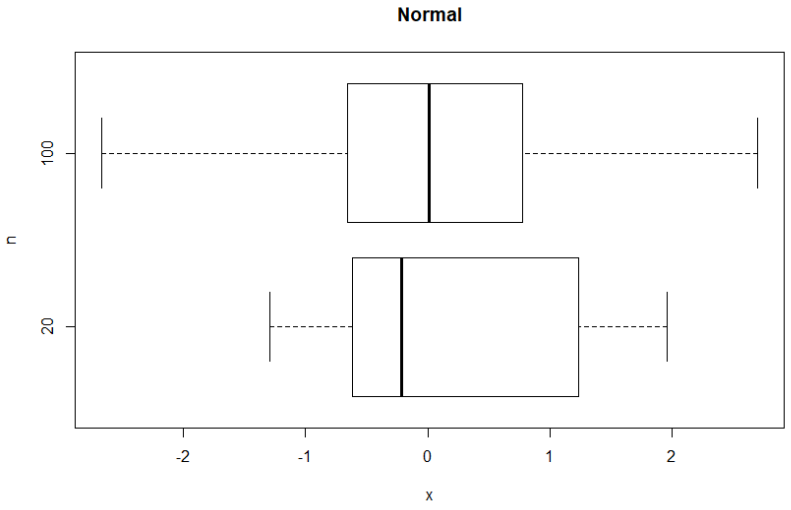


Рис. 6: Нормальное распределение

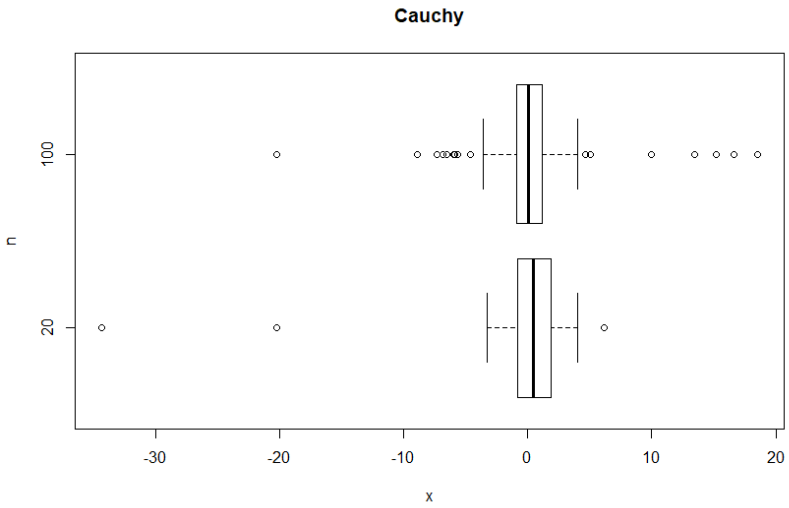


Рис. 7: Распределение Коши

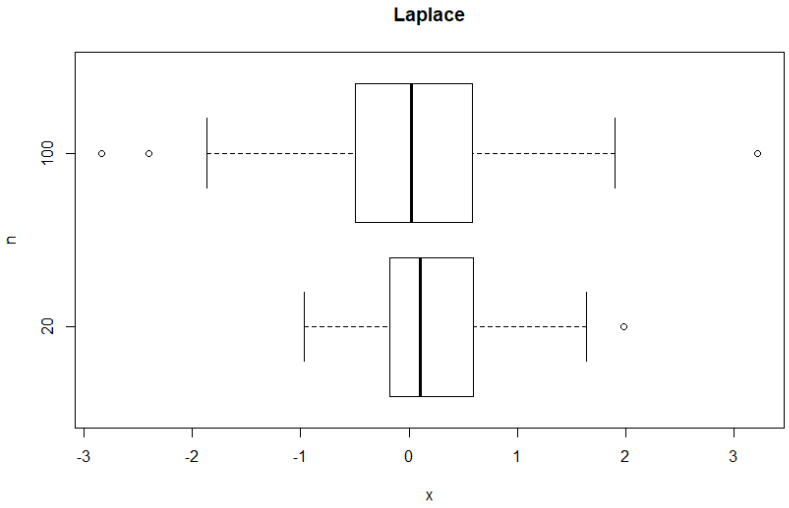


Рис. 8: Распределение Лапласа

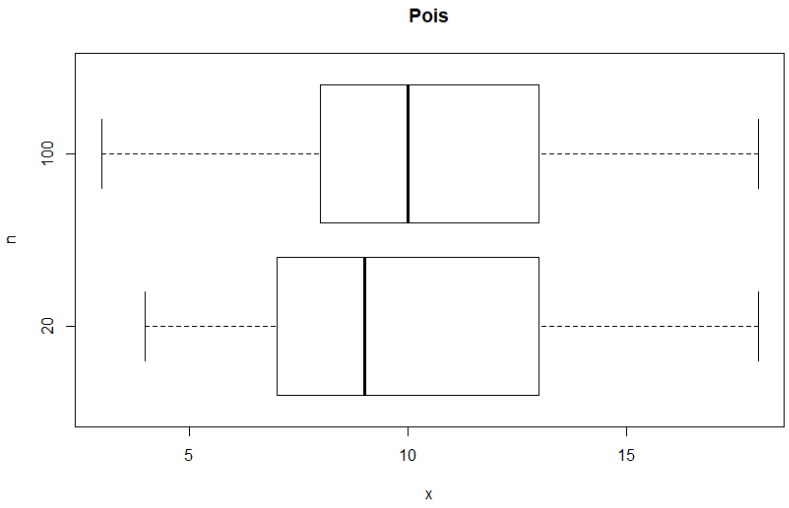


Рис. 9: Распределение Пуассона

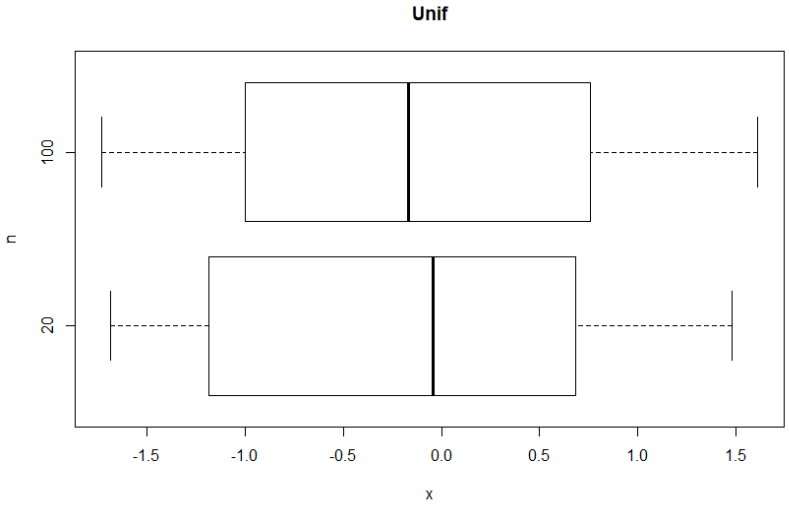


Рис. 10: Равномерное распределение

4.4 Доля выбросов

Округление доли выбросов:

Выборка случайна, поэтому в качестве оценки рассеяния можно взять дисперсию пуассоновского потока: $D_n \approx \sqrt{n}$
Доля $p_n = D_n/n = 1/\sqrt{n}$
Для $n = 20$: $p_n = 1/\sqrt{20}$ – примерно 0.2 или 20%
Для $n = 100$: $p_n = 0.1$ или 10%
Исходя из этого можно решить, сколько знаков оставлять в доле выбросов.

Выборка	Доля выбросов
normal n = 20	0.02
normal n = 100	0.01
cauchy n = 20	0.15
cauchy n = 100	0.16
laplace n = 20	0.07
laplace n = 100	0.06
pois n = 20	0.02
pois n = 100	0.01
unif n = 20	0
unif n = 100	0

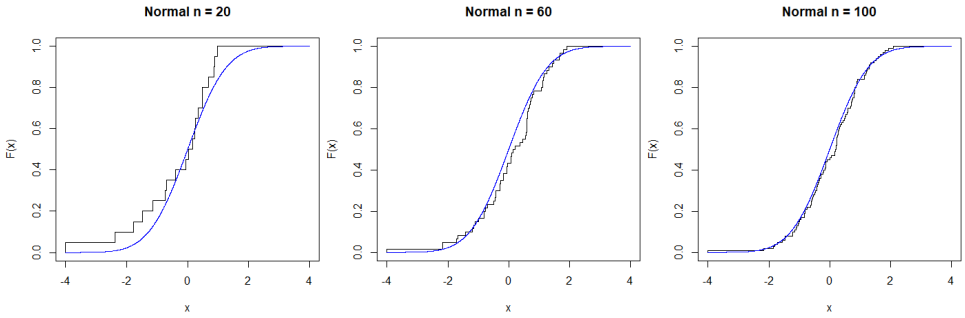
Таблица 7: Доля выбросов

4.5 Теоретическая вероятность выбросов

Распределение	$Q_1^{\text{т}}$	$Q_3^{\text{т}}$	$X_1^{\text{т}}$ (15)	$X_2^{\text{т}}$ (15)	$P_{\text{в}}^{\text{т}}$ (17), (18)
Нормальное распределение	-0.674	0.674	-2.698	2.698	0.007
Распределение Коши	-1	1	-4	4	0.156
Распределение Лапласа	-0.490	0.490	-1.961	1.961	0.063
Распределение Пуассона	8	12	2	18	0.008
Равномерное распределение	-0.866	0.866	-3.464	3.464	0

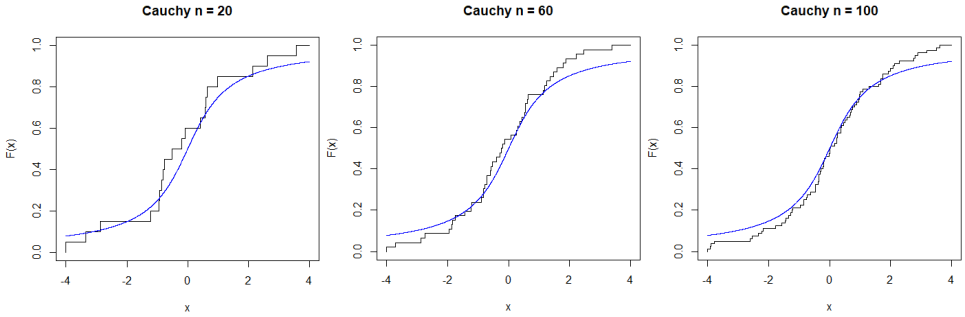
Таблица 8: Теоретическая вероятность выбросов

4.6 Эмпирическая функция распределения



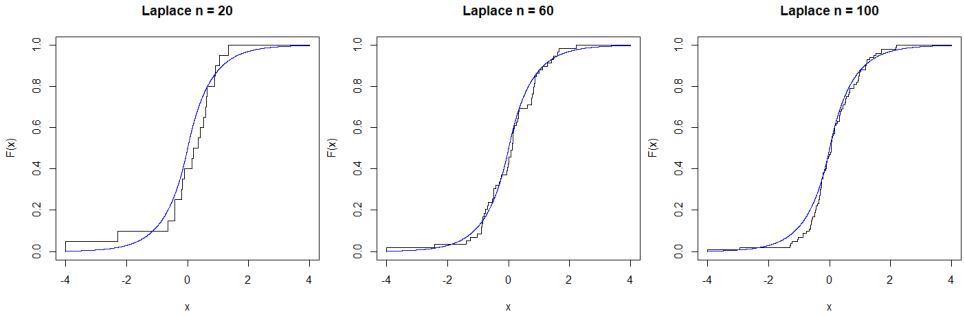
$n = 20$ $n = 60$ $n = 100$

Рис. 11: Нормальное распределение



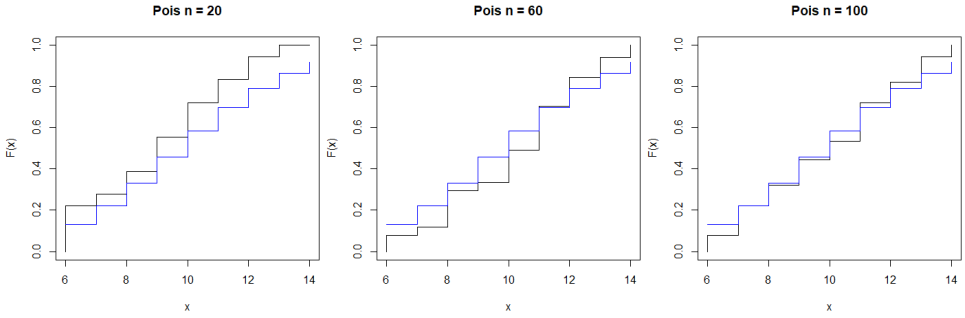
$n = 20$ $n = 60$ $n = 100$

Рис. 12: Распределение Коши



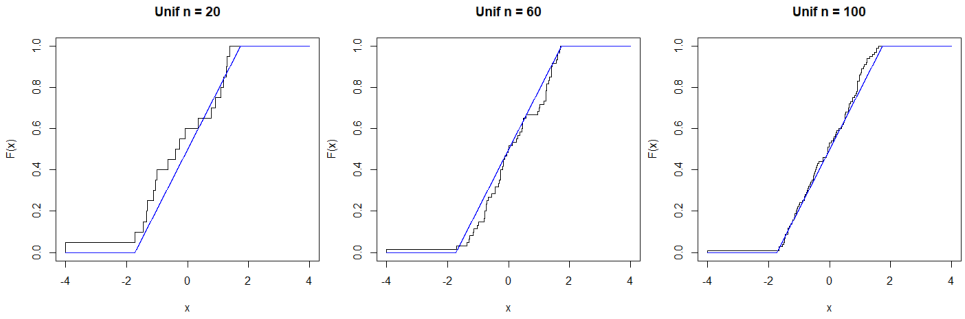
$n = 20$ $n = 60$ $n = 100$

Рис. 13: Распределение Лапласа



$n = 20$ $n = 60$ $n = 100$

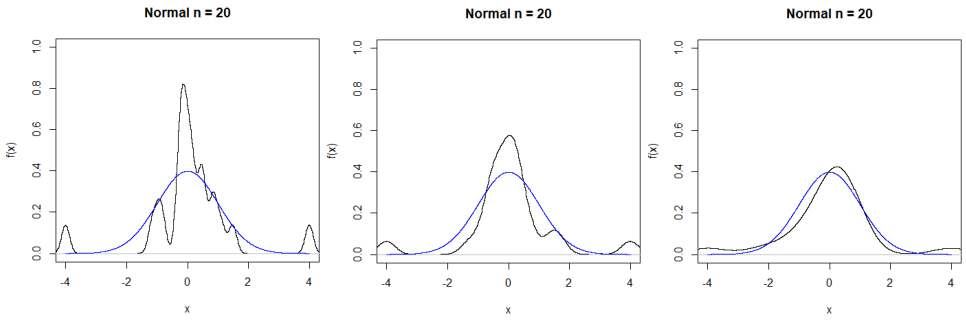
Рис. 14: Распределение Пуассона



$n = 20$ $n = 60$ $n = 100$

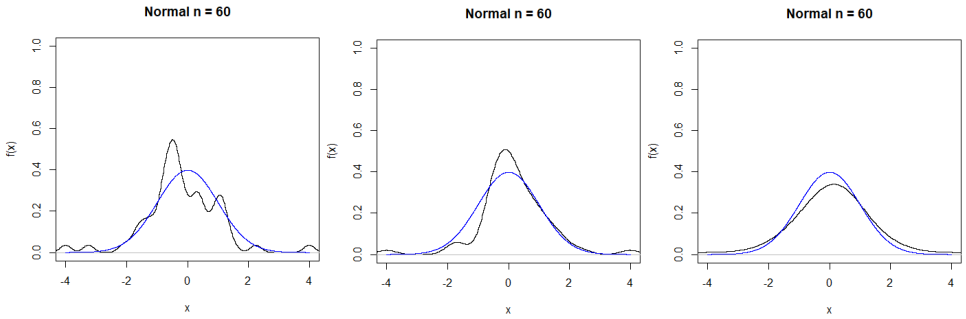
Рис. 15: Равномерное распределение

4.7 Ядерные оценки плотности распределения



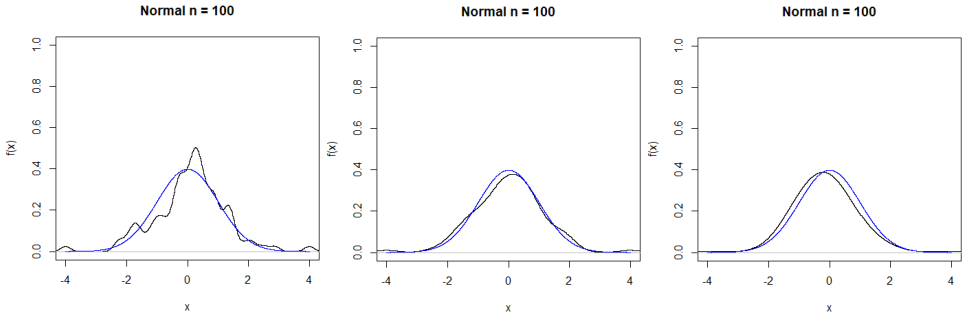
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 16: Нормальное распределение, $n = 20$



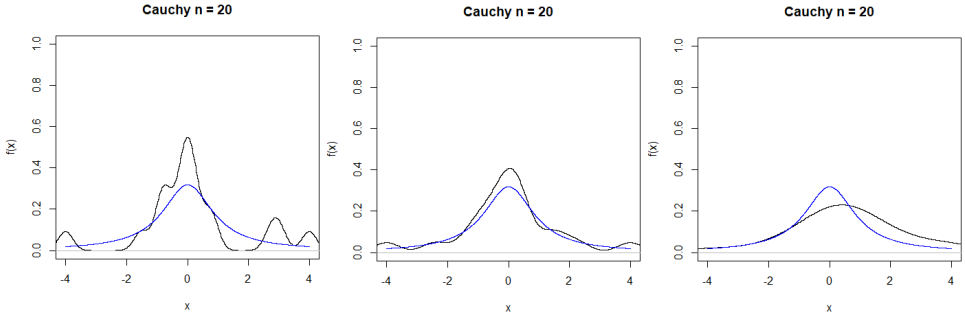
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 17: Нормальное распределение, $n = 60$



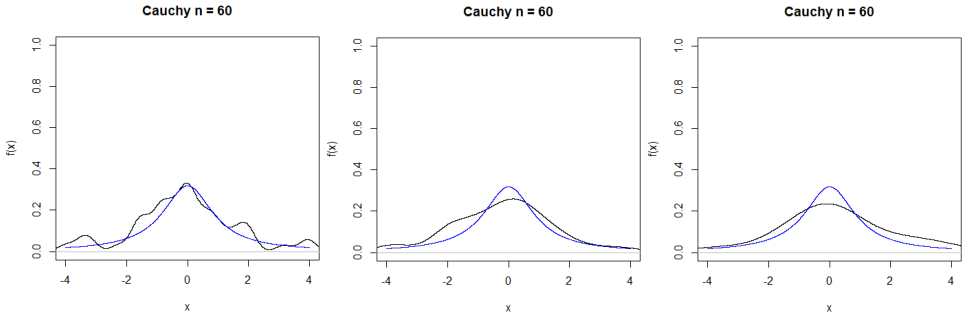
$h = h_n/2$ $h = h_n$ $h = 2h_n$

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



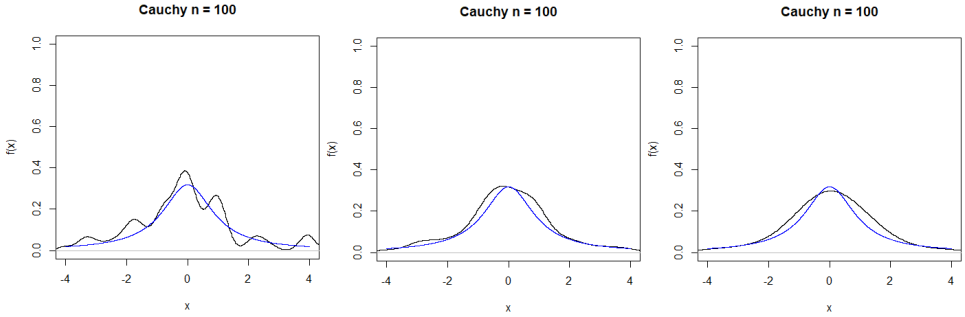
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 19: Распределение Коши, $n = 20$



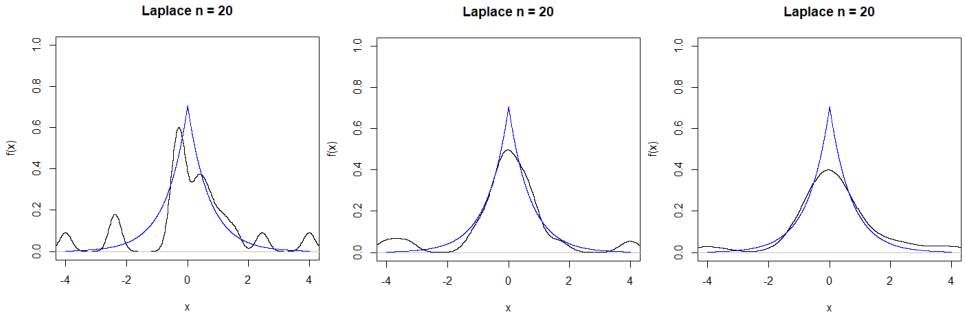
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 20: Распределение Коши, $n = 60$



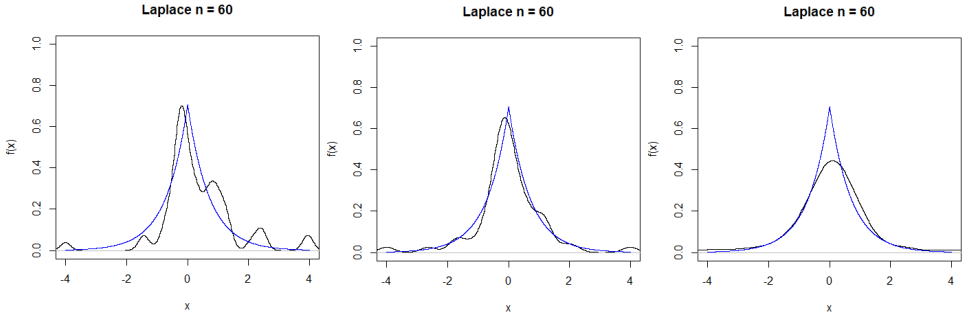
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 21: Распределение Коши, $n = 100$



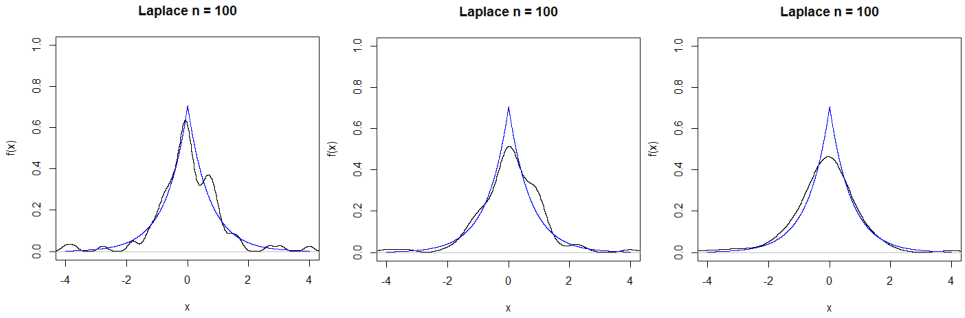
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 22: Распределение Лапласа, $n = 20$



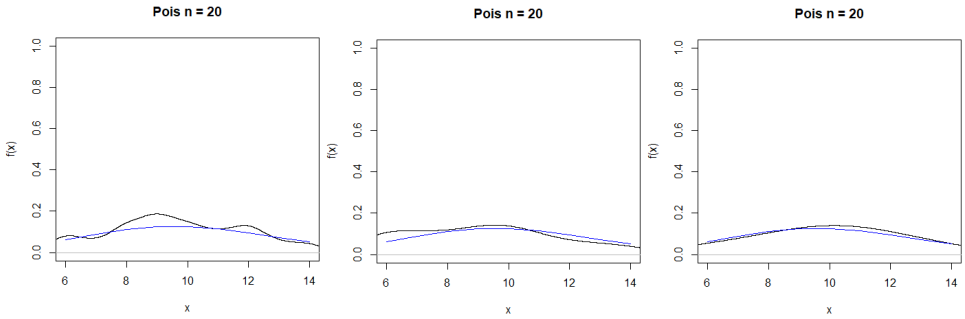
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 23: Распределение Лапласа, $n = 60$



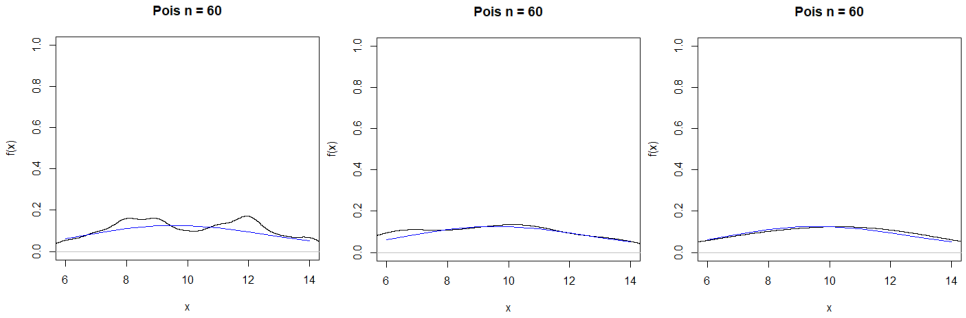
$h = h_n/2$ $h = h_n$ $h = 2h_n$

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



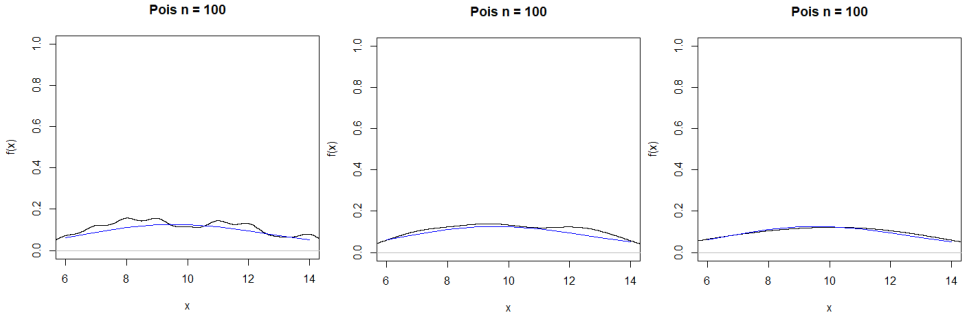
$h = h_n/2$ $h = h_n$ $h = 2h_n$

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



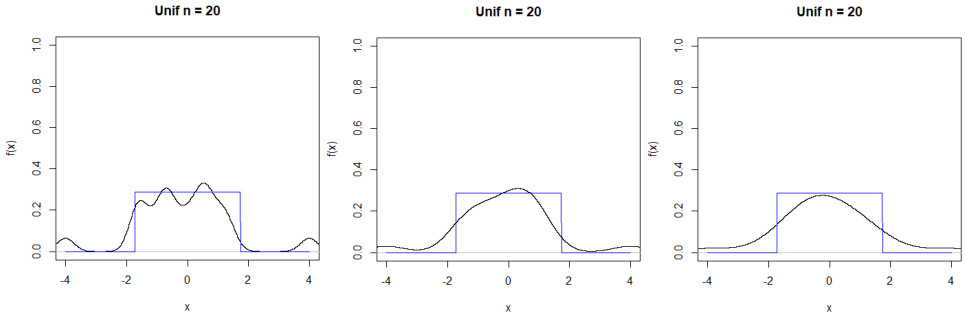
$h = h_n/2$ $h = h_n$ $h = 2h_n$

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



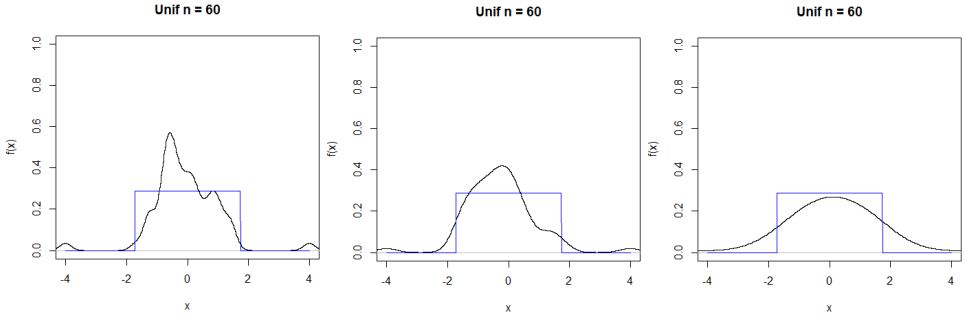
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 27: Распределение Пуассона, $n = 100$



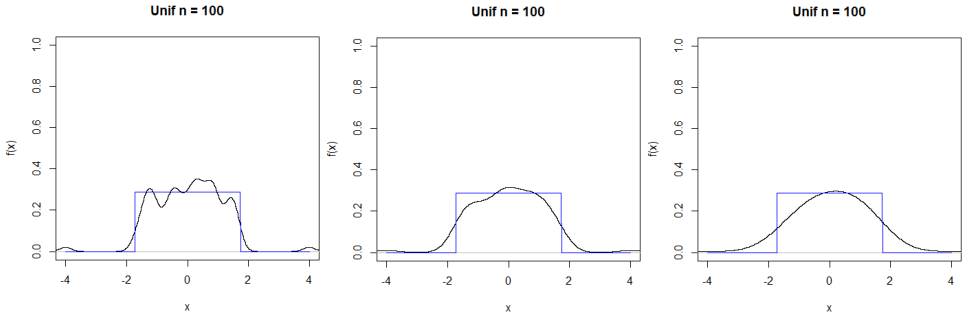
$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 28: Равномерное распределение, $n = 20$



$h = h_n/2$ $h = h_n$ $h = 2h_n$

Рис. 29: Равномерное распределение, $n = 60$



$h = h_n/2$ $h = h_n$ $h = 2h_n$

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

5 Обсуждение

5.1 Гистограмма и график плотности распределения

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 Доля и теоретическая вероятность выбросов

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 тео-

ретическую 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 и ядерные оценки плотности распределения

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 $\hat{h_n}$, 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 $h = 2h_n$ по полученным приближениям становится сложно сказать плотность вероятности какого распределения они должны повторять, так как они очень похожи между собой.

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] Анатольев, Станислав (2009) «Непараметрическая регрессия», Квантиль, №7, стр. 37-52.