

Санкт–Петербургский политехнический университет  
Петра Великого  
Физико-механический институт  
Кафедра «Прикладная математика»

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

Выполнил студент:  
Турченко Михаил Константинович  
группа: 5030102/90101

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

Санкт-Петербург  
2021г.

## Оглавление

1. Постановка задачи .....	3
2. Теория.....	4
2.1 Рассматриваемые 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 Эмпирическая функция распределения.....	7
2.7.3 Нахождение э. ф. р. ....	7
2.8 Оценки плотности вероятности.....	8
2.8.1 Определение .....	8
2.8.2 Ядерные оценки .....	8
3. Реализация.....	9
4. Результаты.....	9
4.1 Гистограмма и график плотности распределения.....	9
4.2 Характеристики положения и рассеяния .....	11
4.3 Боксплот Тьюки.....	15
4.4 Доля выбросов.....	17
4.5 Теоретическая вероятность выбросов.....	17
4.6 Эмпирическая функция распределения.....	17
4.7 Ядерные оценки плотности распределения.....	19
5. Обсуждение .....	25
5.1 Гистограмма и график плотности распределения.....	25
5.2 Характеристики положения и рассеяния .....	25
5.3 Доля и теоретическая вероятность выбросов .....	26
5.4 Эмпирическая функция и ядерные оценки плотности распределения .....	26

## 1. Постановка задачи

для 4 распределений:

- Нормальное распределение  $N(x, 0, 1)$ ;
- Распределение Коши  $C(x, 0, 1)$ ;
- Распределение Пуассона  $P(x, 10)$ ;
- 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:

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

Вычислить оценку дисперсии по формуле:

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

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

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

- Распределение Коши

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

- Распределение Пуассона

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

- Равномерное распределение

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

## 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.

## 2.3 Вариационный ряд

Вариационным ряд - последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются.

## 2.4 Выборочные числовые характеристики

### 2.4.1 Характеристики положения

- Выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Выборочная медиана

$$med x = \begin{cases} x_{(l+1)}, & n = 2l + 1 \\ \frac{x_{(l+1)} + x_{(l)}}{2}, & n = 2l \end{cases}$$

- Полусумма экстремальных выборочных элементов

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

- Полусумма квартилей

Выборочная квартиль  $z_p$  порядка  $p$  определяется формулой

$$z_p = \begin{cases} x_{([np]+1)}, & np \text{ дробное} \\ x_{(np)}, & np \text{ целое} \end{cases}$$

Полусумма квартилей

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

- Усеченное среднее

$$z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_i, r \sim \frac{n}{4}$$

## 2.4.2 Характеристики рассеяния

Выборочная дисперсия

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

## 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»:

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

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 звездочек.

## 2.6 Теоретическая вероятность выбросов

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

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases}$$

Теоретическая вероятность выбросов

- для непрерывных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T))$$

- для дискретных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T))$$

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

## 2.7 Эмпирическая функция распределения

### 2.7.1 Статистический ряд

Статистическим ряд – последовательность различных элементов выборки  $z_1 \dots z_k$ , расположенных в возрастающем порядке с указанием частот  $n_1 \dots n_k$ , 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:

$$F^*(X) = P^*(X < x)$$

### 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  $P^*(X < x) = \frac{1}{n} \sum_{z_i < x} n_i$ . Получаем

$$F^*(x) = \frac{1}{n} \sum_{z_i < x} n_i$$

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

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

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

$$F_n^*(x) \approx F_X(x)$$

## 2.8 Оценки плотности вероятности

### 2.8.1 Определение

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

$$\hat{f}(x) \approx f(x). \quad (22)$$

### 2.8.2 Ядерные оценки

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

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

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

$$\lim_{n \rightarrow \infty} h_n = 0; \quad \lim_{n \rightarrow \infty} nh_n = \infty$$

Такие оценки называются непрерывными ядерными.

Гауссово (нормальное) ядро

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

Правило Сильвермана

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

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



### 3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языков программирования C++ в среде разработки Microsoft Visual Studio, а также Matlab.

Ссылка на исходный код: [MathStat/Labs 1-4 at main · TurchenkoMikhail/MathStat \(github.com\)](https://github.com/TurchenkoMikhail/MathStat)

### 4. Результаты

#### 4.1 Гистограмма и график плотности распределения

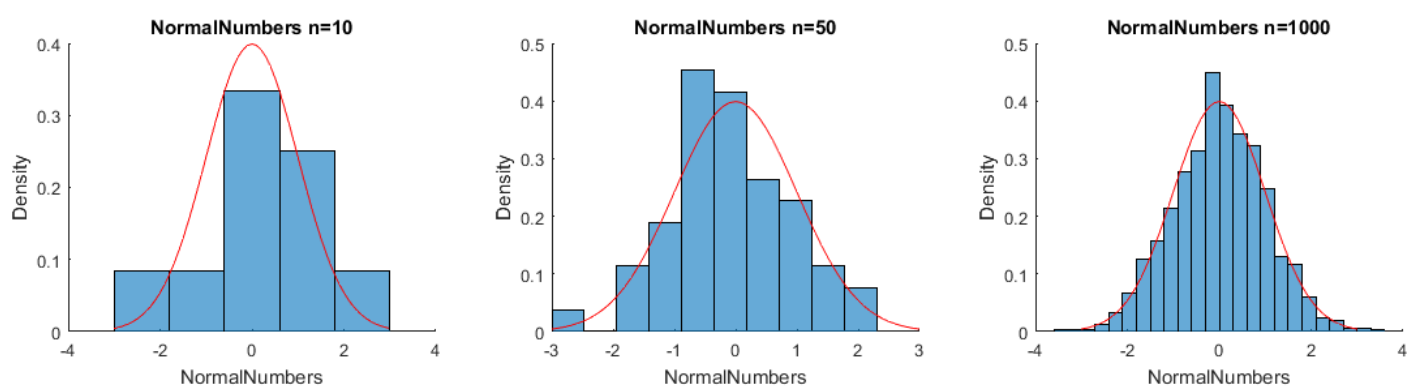


Рис. 1 Нормальное распределение

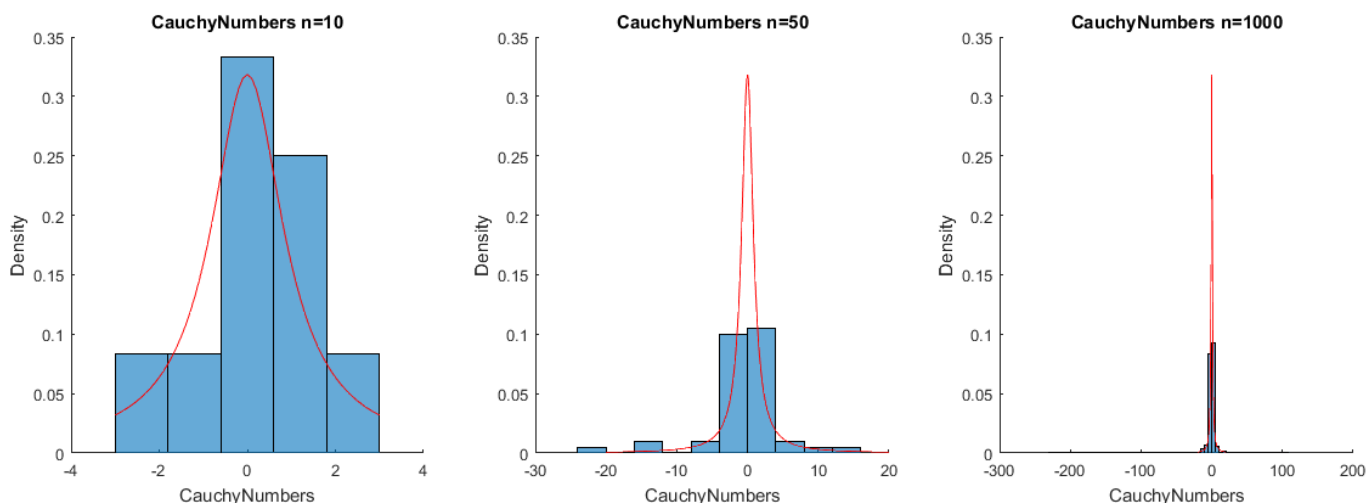


Рис. 2 Распределение Коши

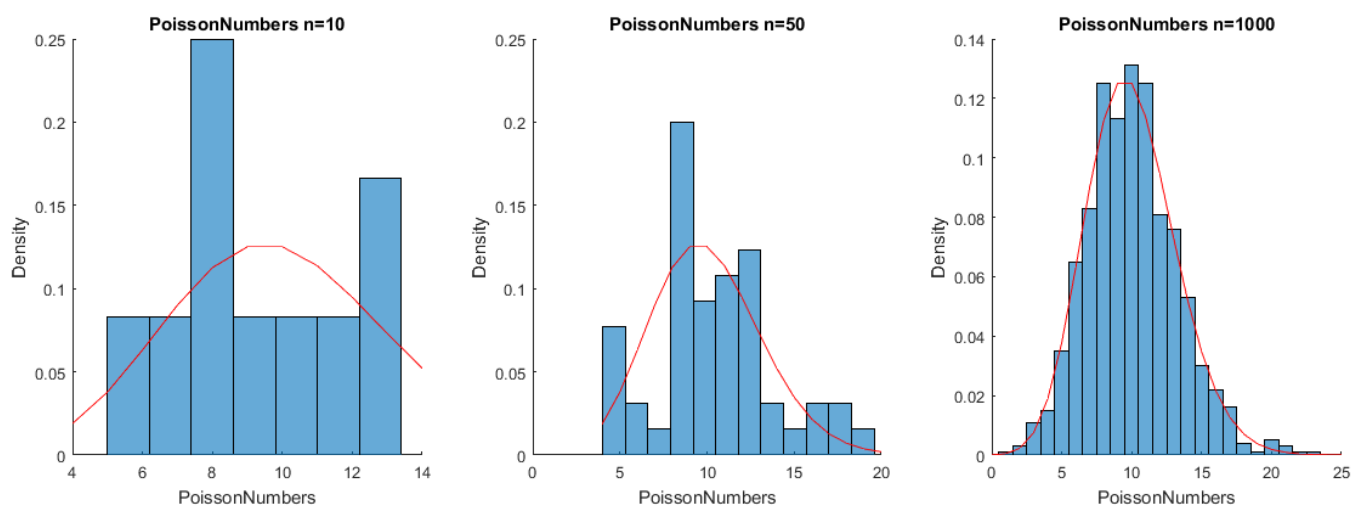


Рис. 3 Распределение Пуассона

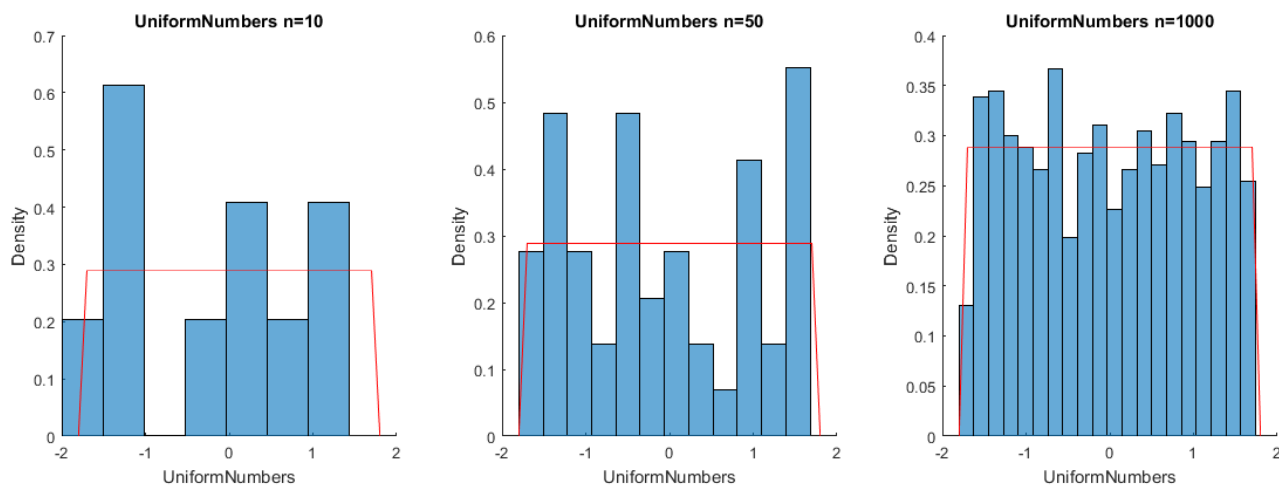


Рис. 4 Равномерное распределение

## 4.2 Характеристики положения и рассеяния

normal n = 10					
	<x>	med x	zR	zQ	ztr
E(z)	0.00278	0.00499	-0.01962	0.01198	0.00951
D(z)	0.0911	0.12557	0.17609	0.11066	0.10357
$\sqrt{D(z)}$	0,30183	0,35436	0,41964	0,33265	0,32182
$E(z) \pm \sqrt{D(z)}$	[-0,29905 , 0,30461]	[-0,34937 , 0,35935]	[-0,43926 , 0,40002]	[-0,32067 , 0,34463]	[-0,31232 , 0,33132]
	0	0	0	0	0
normal n = 100					
	<x>	med x	zR	zQ	ztr
E(z)	0.00202	0.00301	-0.0098	-0.01412	0.00317
D(z)	0.01021	0.01545	0.0951	0.01261	0.01214
$\sqrt{D(z)}$	0,101032668	0,124271879	0,308395039	0,112267983	0,110178945
$E(z) \pm \sqrt{D(z)}$	[-0,09901 , 0,10305]	[-0,12126 , 0,12728]	[-0,31818 , 0,29862]	[-0,12639 , 0,09815]	[-0,10701 , 0,11335]
	0	0	0	0	0
normal n = 1000					
	<x>	med x	zR	zQ	ztr
E(z)	0.00167	0.00077	0.00604	-0.00037	0.0014
D(z)	0.00105	0.00177	0.05829	0.00132	0.0012
$\sqrt{D(z)}$	0,032553648	0,040802451	0,241445232	0,036301653	0,035219739
$E(z) \pm \sqrt{D(z)}$	[-0,03088 , 0,03422]	[-0,04004 , 0,04156]	[-0,23541 , 0,24749]	[-0,03667 , 0,03593]	[-0,03382 , 0,03662]
	0.0	0.0	0	0.0	0.0

Таблица 1. Нормальное распределение

cauchy n = 10					
	<x>	med x	zR	zQ	Ztr
E(z)	-1.55414	0.0225447	-8.00916	0.0657576	0.0432478
D(z)	1926.02	0.328877	48004.8	2.00805	0.60502
$\sqrt{D(z)}$	43,88644438	0,573477986	219,0999772	1,417056809	0,777830316
$E(z) \pm \sqrt{D(z)}$	-	[-0,55094 , 0,59602]	-	[-1,3513 , 1,48282]	[-0,73458 , 0,82108]
	-	0	-	-	0
cauchy n = 100					
	<x>	med x	zR	zQ	Ztr
E(z)	0.0560977	0.00705747	3.71962	-0.0327477	0.00333883
D(z)	36.182	0.0239542	581316	0.0501979	0.0254582
$\sqrt{D(z)}$	6,015147546	0,154771444	762,4408174	0,224048879	0,15955626
$E(z) \pm \sqrt{D(z)}$	-	[-0,14771 , 0,16183]	-	[-0,2568 , 0,1913]	[-0,15622 , 0,1629]
	-	0	-	0	0
cauchy n = 1000					
	<x>	med x	zR	zQ	ztr
E(z)	4.07463	0.00222788	2014.77	0.00328705	0.00403019
D(z)	10209	0.00228551	2.54082e+09	0.00511592	0.00239758
$\sqrt{D(z)}$	101,0395962	0,047807008	50406,54719	0,07152566	0,04896509
$E(z) \pm \sqrt{D(z)}$	-	[-0,04558 , 0,05004]	-	[-0,06824 , 0,07482]	[-0,04494 , 0,053]
	-	0.0	-	0.0	0.0

Таблица 2. Распределение Коши

poisson n = 10					
	<x>	med x	zR	zQ	ztr
E(z)	10.045	9.9165	10.2975	9.9845	9.95033
D(z)	1.0942	1.55728	1.88924	1.32251	1.24759
$\sqrt{D(z)}$	1,04604	1,24791	1,37450	1,15000	1,11696
$E(z) \pm \sqrt{D(z)}$	[8,99896 , 11,09104]	[8,66859 , 11,16441]	[8,923 , 11,672]	[8,8345 , 11,1345]	[8,83337 , 11,06729]
	$10_{-1}^{+1}$	$10_{-1}^{+1}$	$10_{-1}^{+1}$	$10_{-1}^{+1}$	$10_{-1}^{+1}$
poisson n = 100					
	<x>	med x	zR	zQ	ztr
E(z)	9.99829	9.8495	10.9385	9.863	9.85238
D(z)	0.104427	0.2196	0.975468	0.176731	0.12761
$\sqrt{D(z)}$	0,32315	0,46861	0,98766	0,42039	0,35723
$E(z) \pm \sqrt{D(z)}$	[9,67514 , 10,32144]	[9,38089 , 10,31811]	[9,95084 , 11,92616]	[9,44261 , 10,28339]	[9,49515 , 10,20961]
	$10_{-1}^{+1}$	$10_{-1}^{+1}$	$11_{-1}^{+1}$	$10_{-1}^{+1}$	$10_{-1}^{+1}$
poisson n = 1000					
	<x>	med x	zR	zQ	ztr
E(z)	10.0013	9.998	11.6655	9.9955	9.85783
D(z)	0.0100194	0.001996	0.59686	0.00222975	0.0106565
$\sqrt{D(z)}$	0,10010	0,04468	0,77257	0,04722	0,10323
$E(z) \pm \sqrt{D(z)}$	[9,9012 , 10,1014]	[9,95332 , 10,04268]	[10,89293 , 12,43807]	[9,94828 , 10,04272]	[9,7546 , 9,96106]
	$10_{-1}^{+1}$	$10_{-1}^{+1}$	$11_{-1}^{+1}$	$10_{-1}^{+1}$	$9_{-1}^{+1}$

Таблица 3. Распределение Пуассона

uniform n = 10					
	<x>	med x	zR	zQ	ztr
E(z)	0.00541393	-0.00174422	0.00608338	0.00856121	0.00439282
D(z)	0.110836	0.242794	0.047499	0.150938	0.177693
$\sqrt{D(z)}$	0,332920411	0,492741311	0,217942653	0,3885074	0,421536475
$E(z) \pm \sqrt{D(z)}$	[-0,32751 , 0,33833]	[-0,49448 , 0,491]	[-0,21186 , 0,22402]	[-0,37995 , 0,39707]	[-0,41715 , 0,42593]
	0	0	0	0	0
uniform n = 100					
	<x>	med x	zR	zQ	ztr
E(z)	0.00126682	-0.000631133	-0.000728412	-0.0153259	-0.00140547
D(z)	0.00906393	0.0265639	0.000592985	0.0138625	0.0182497
$\sqrt{D(z)}$	0,095204674	0,162984355	0,024351283	0,117739118	0,135091451
$E(z) \pm \sqrt{D(z)}$	[-0,09393 , 0,09647]	[-0,16361 , 0,16235]	[-0,02508 , 0,02362]	[-0,13307 , 0,10241]	[-0,1365 , 0,13368]
	0.0	0	0.0	0	0
uniform n = 1000					
	<x>	med x	zR	zQ	ztr
E(z)	0.000892092	-0.000336816	3.11301e-05	-0.00047605	0.000550872
D(z)	0.00113906	0.00325909	6.22572e-06	0.00171905	0.00225572
$\sqrt{D(z)}$	0,033749963	0,05708844	0,002495139	0,041461428	0,047494421
$E(z) \pm \sqrt{D(z)}$	[-0,03286 , 0,03464]	[-0,05743 , 0,05675]	[-0,00247 , 0,00253]	[-0,04194 , 0,04098]	[-0,04694 , 0,04804]
	0.0	0.0	0.00	0.0	0.0

Таблица 4. Равномерное распределение

### 4.3 Боксплот Тьюки

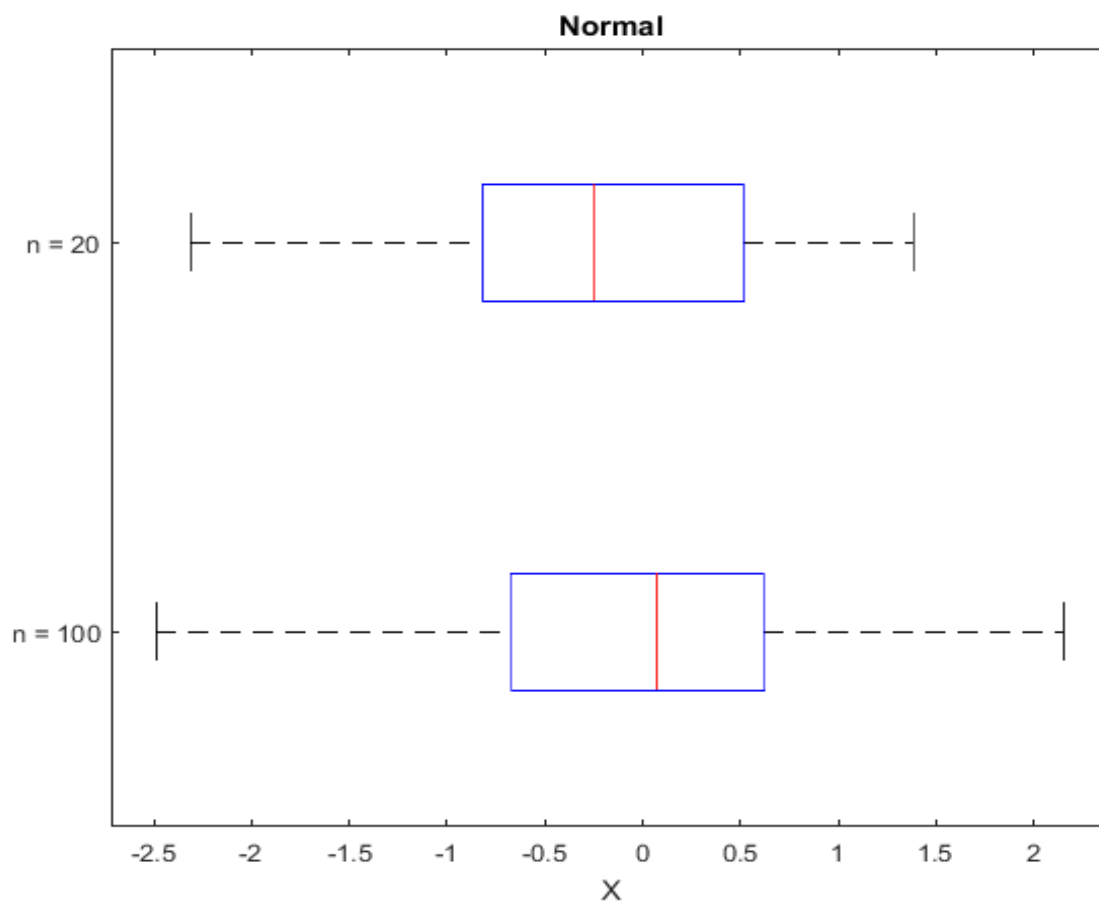


Рис. 5 Нормальное распределение

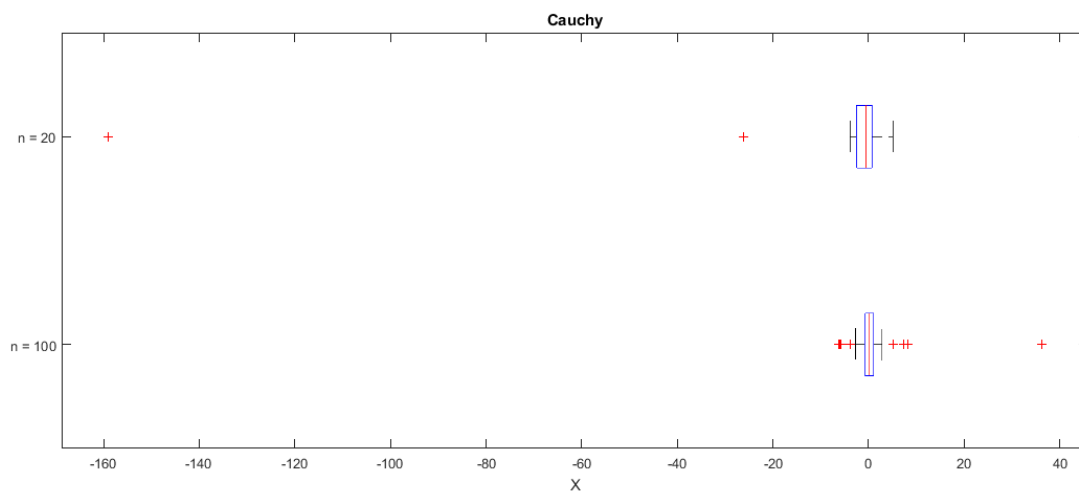


Рис. 6 Распределение Коши

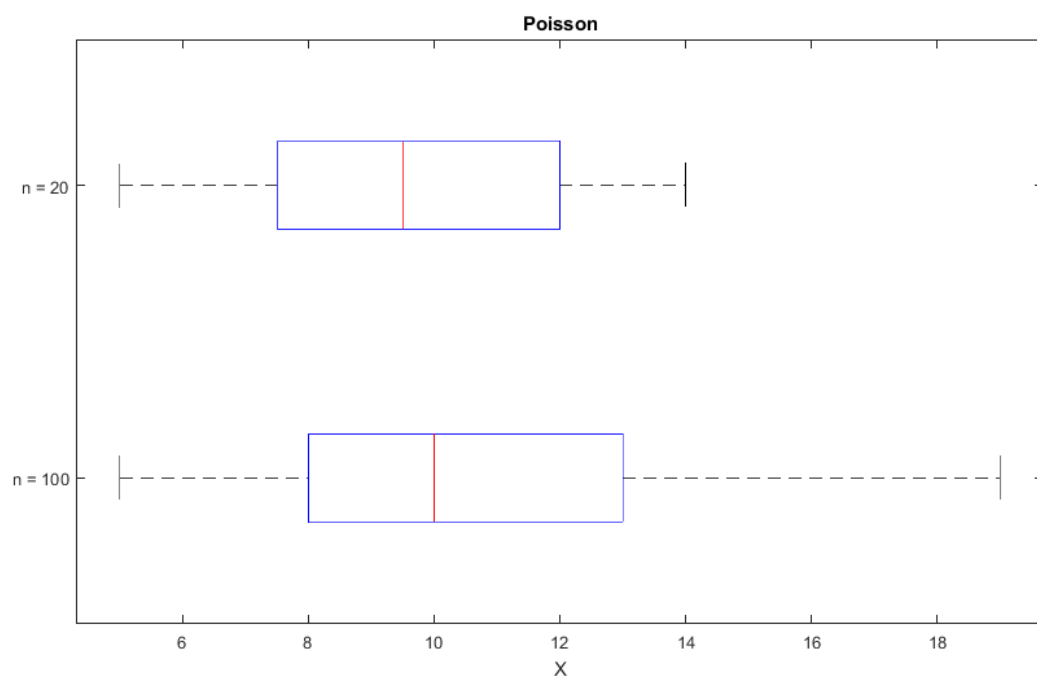


Рис. 7 Распределение Пуассона

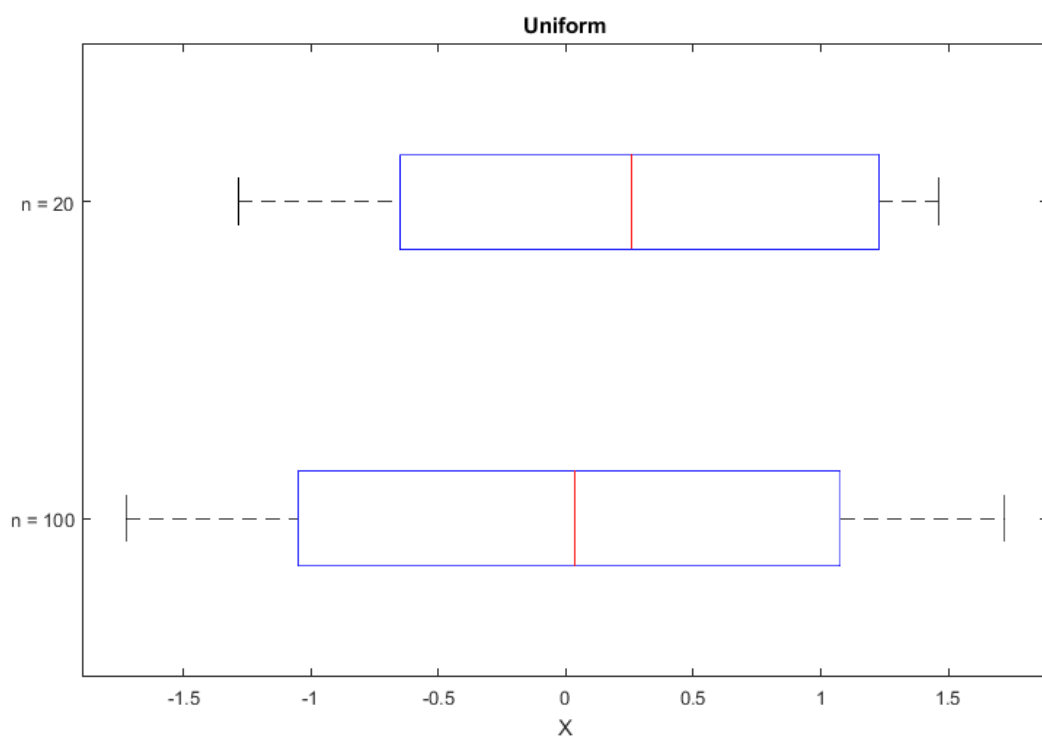


Рис. 8 Равномерное распределение



#### 4.4 Доля выбросов

Выборка	Доля выбросов
normal n = 20	0.05
normal n = 100	0.01
cauchy n = 20	0.15
cauchy n = 100	0.13
poisson n = 20	0.02
poisson n = 100	0.01
uniform n = 20	0
uniform n = 100	0

#### 4.5 Теоретическая вероятность выбросов

Распределение	$Q_1^T$	$Q_3^T$	$X_1^T$	$X_2^T$	$P_B^T$
Normal	-0.674	0.674	-2.698	2.698	0.007
Cauchy	-1	1	-4	4	0.156
Poisson	8	12	2	18	0.008
Uniform	-0.866	0.866	-3.464	3.464	0

#### 4.6 Эмпирическая функция распределения

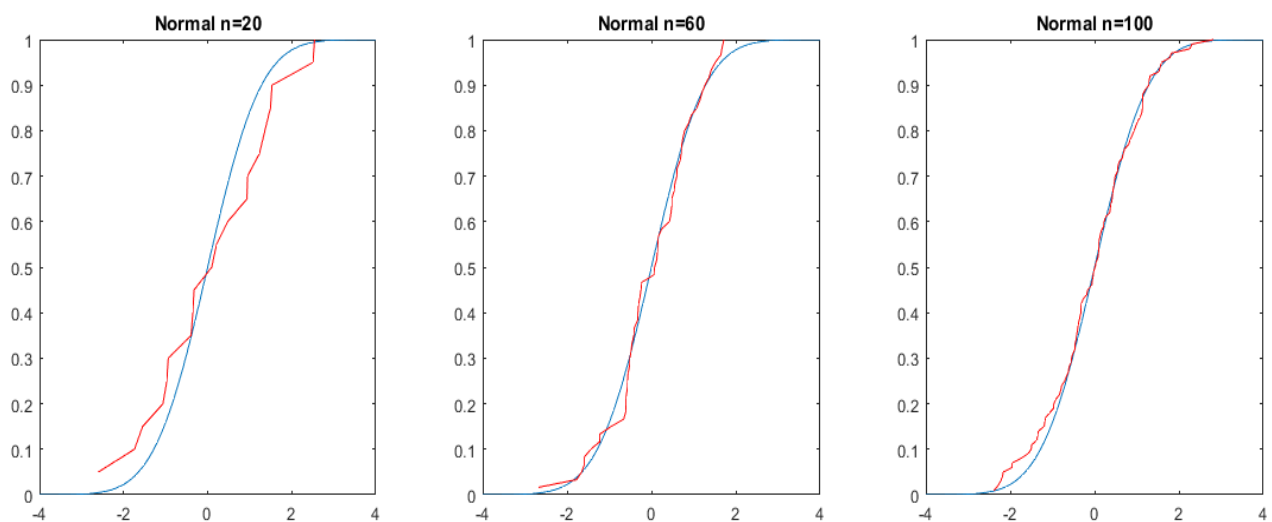


Рис. 9 Нормальное распределение

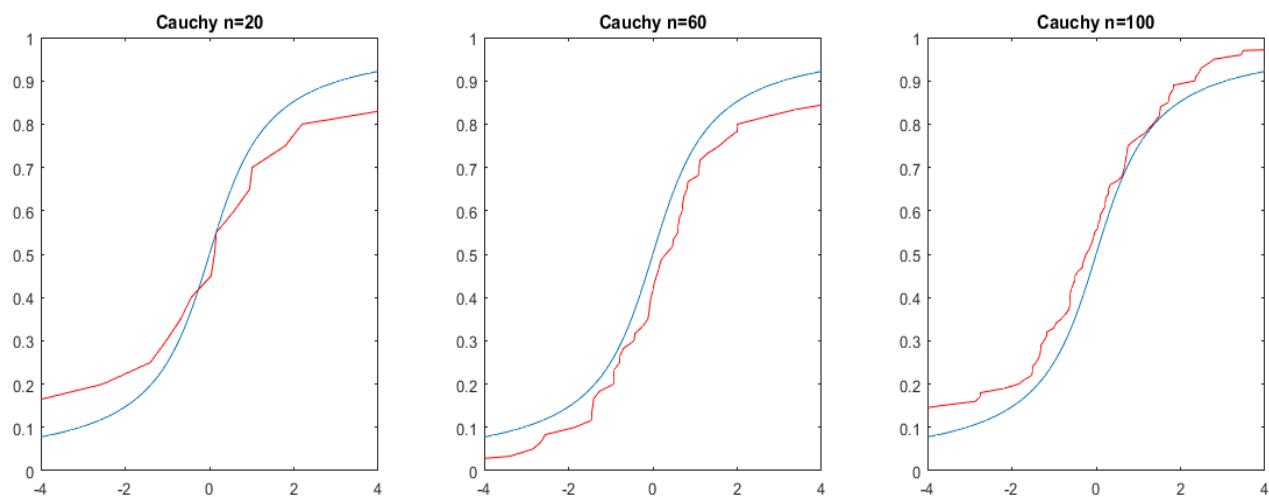


Рис. 10 Распределение Коши

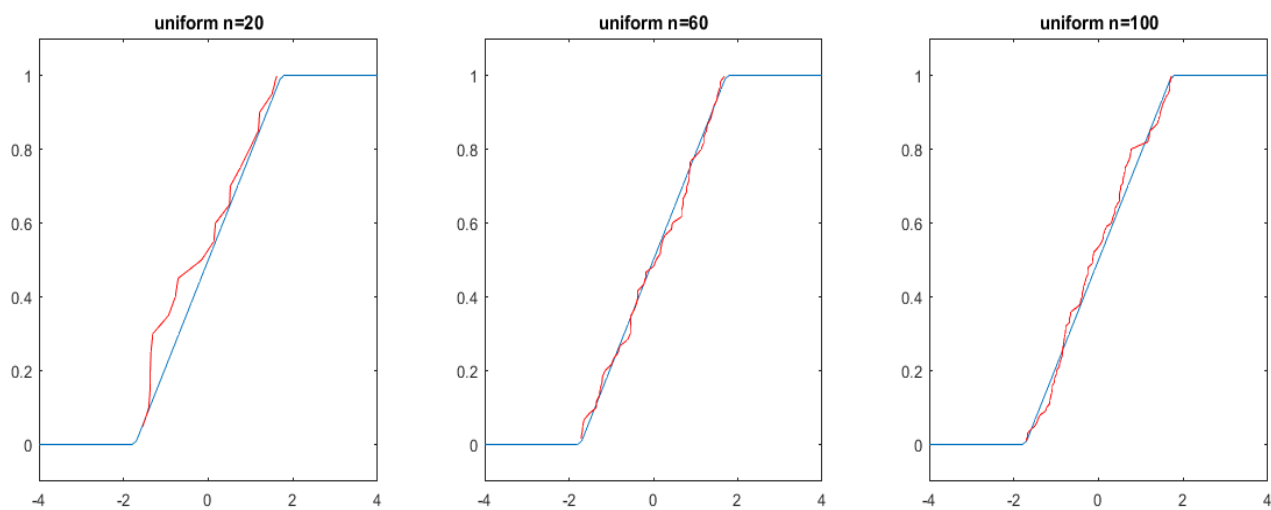


Рис. 11 Равномерное распределение

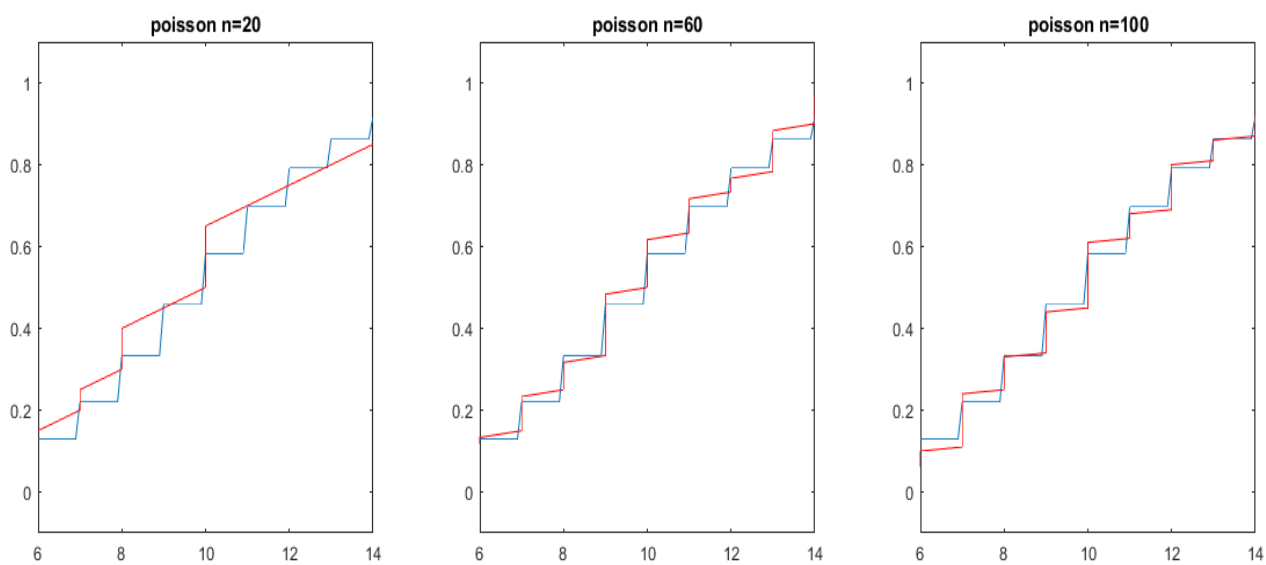


Рис. 12 Распределение Пуассона

#### 4.7 Ядерные оценки плотности распределения

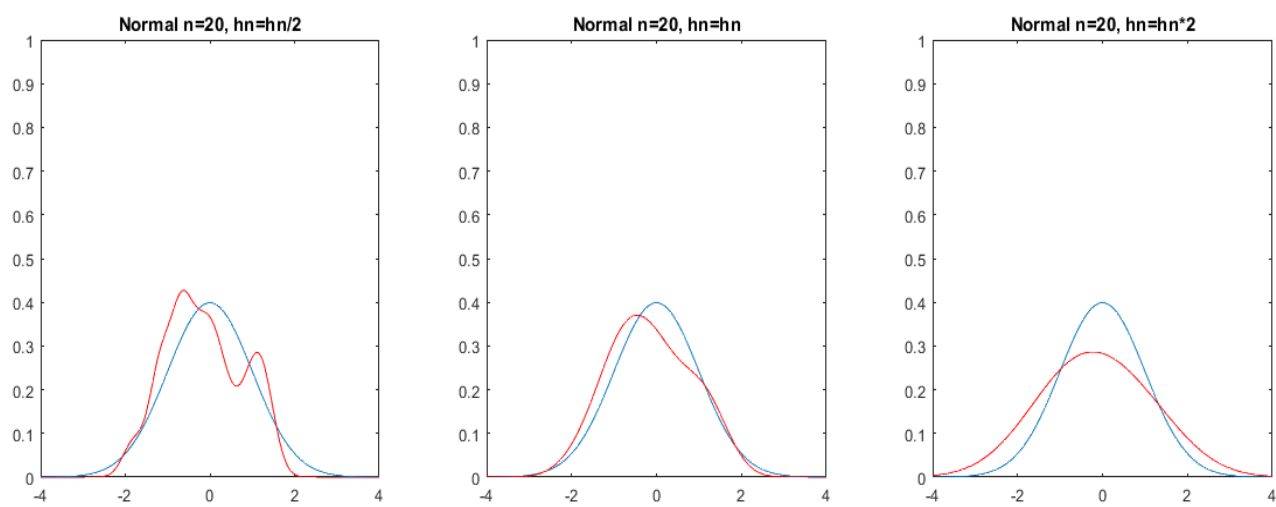


Рис. 13 Нормальное распределение,  $n=20$

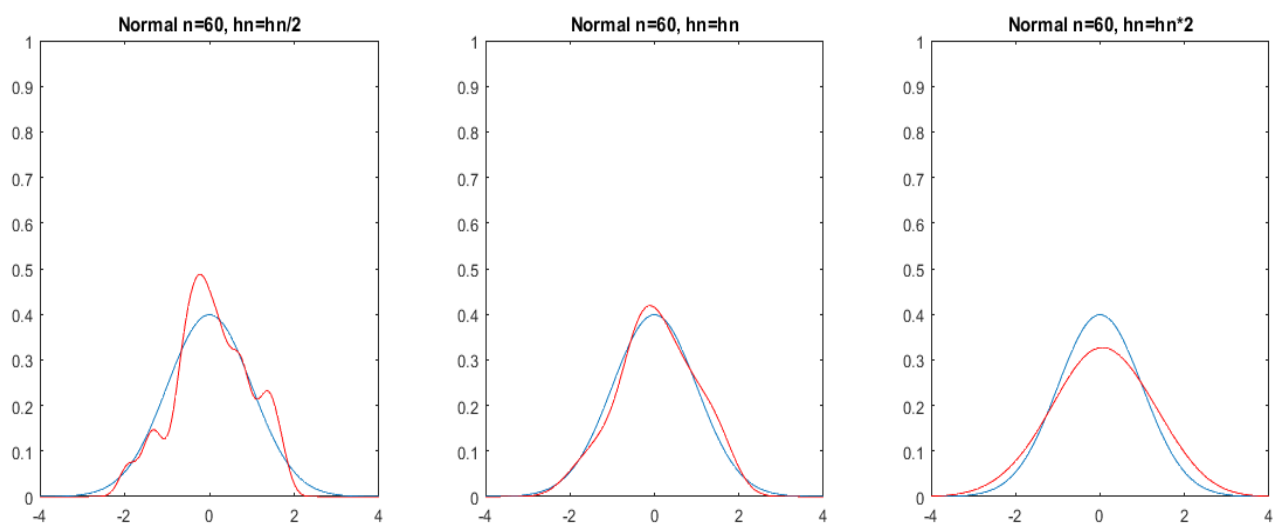


Рис. 14 Нормальное распределение,  $n=60$

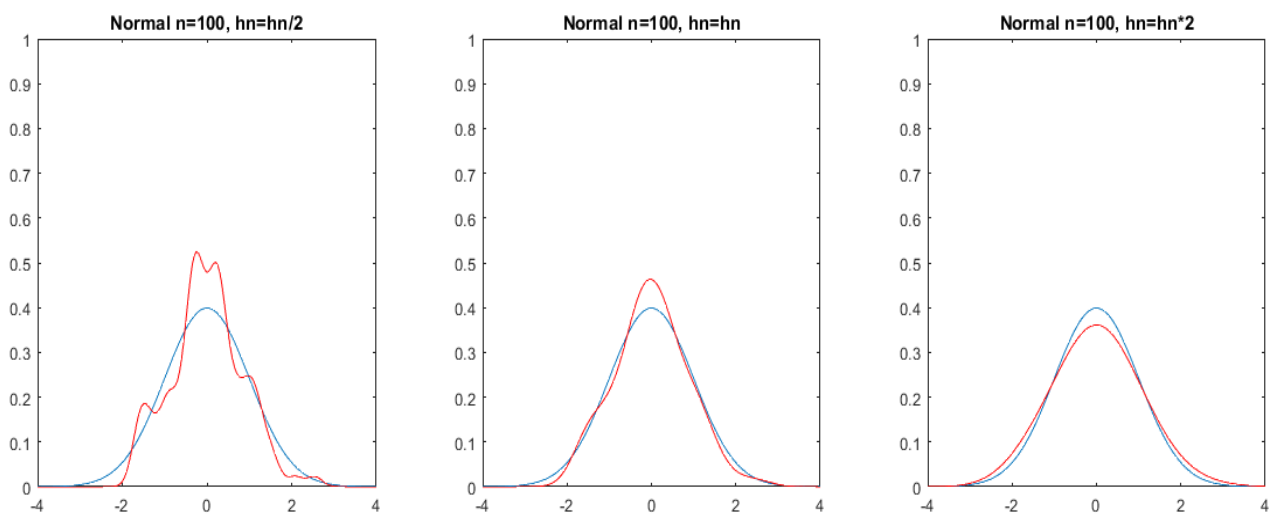


Рис. 15 Нормальное распределение,  $n=100$

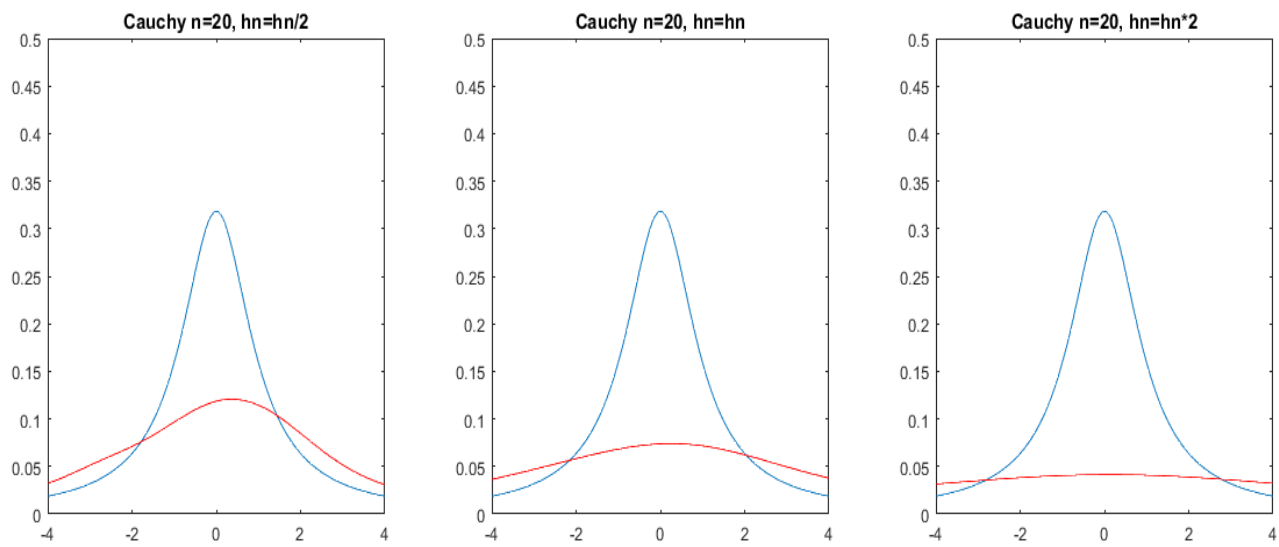


Рис. 16 Распределение Коши,  $n=20$

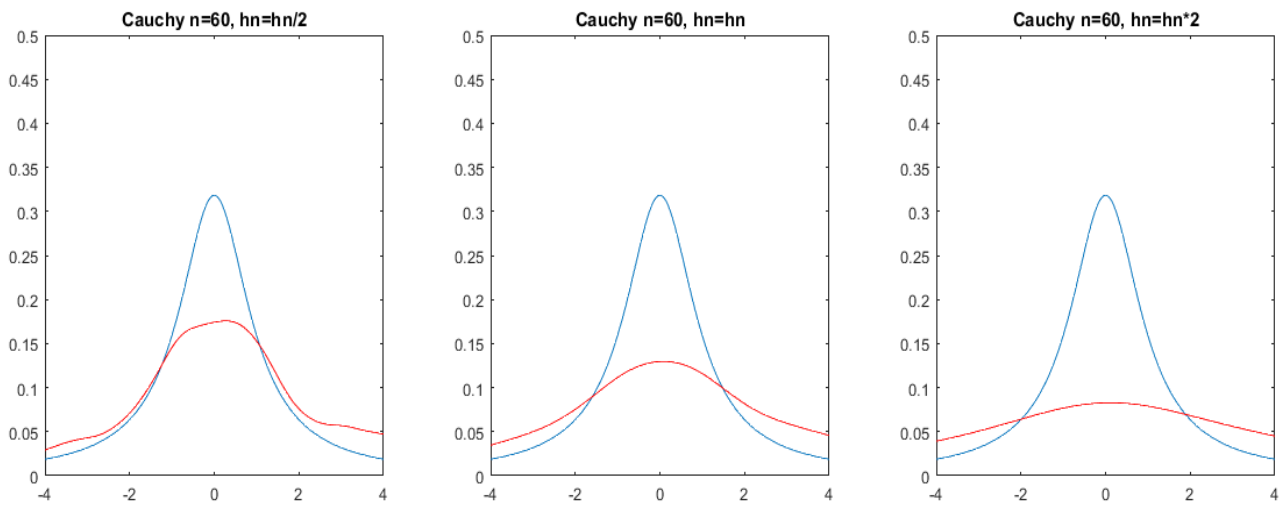


Рис. 17 Распределение Коши,  $n=60$

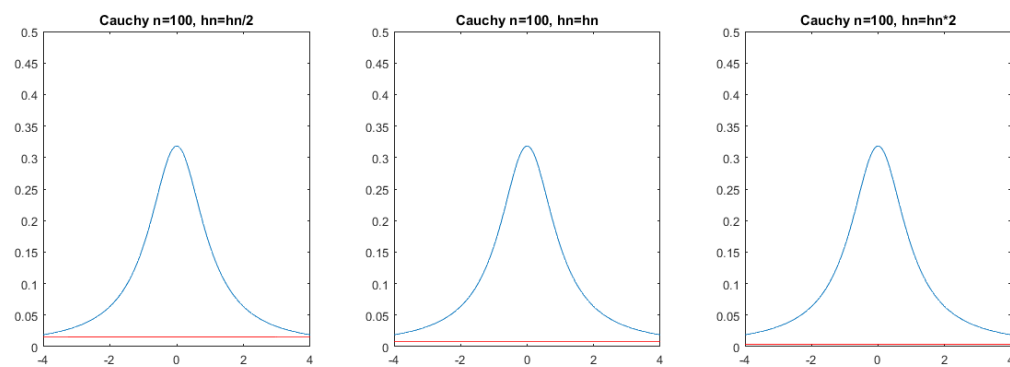


Рис. 18 Распределение Коши,  $n=100$

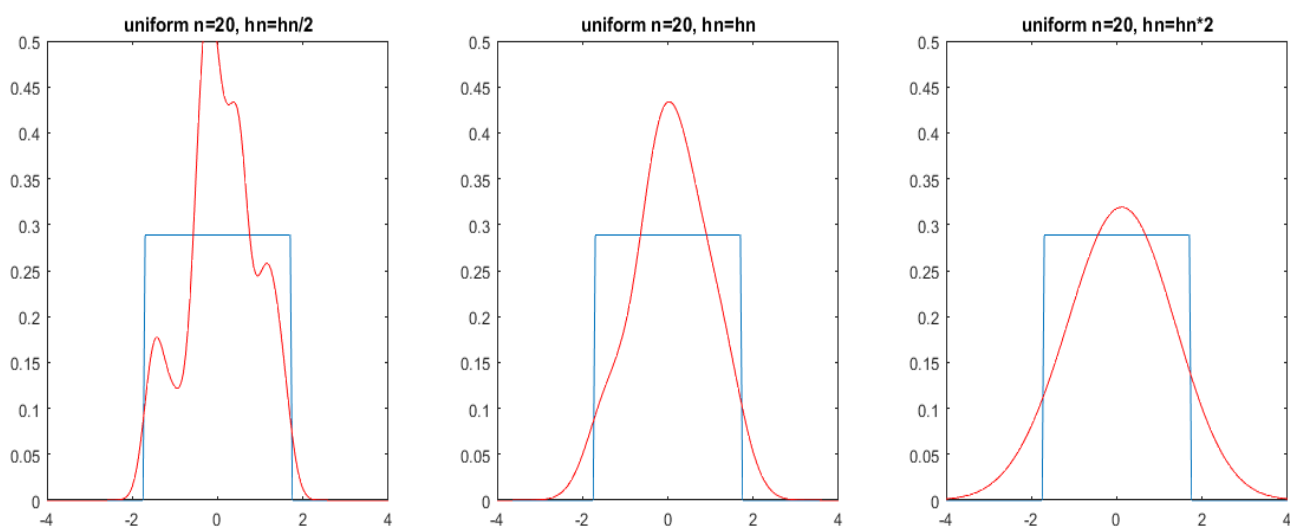


Рис. 19 Равномерное распределение,  $n=20$

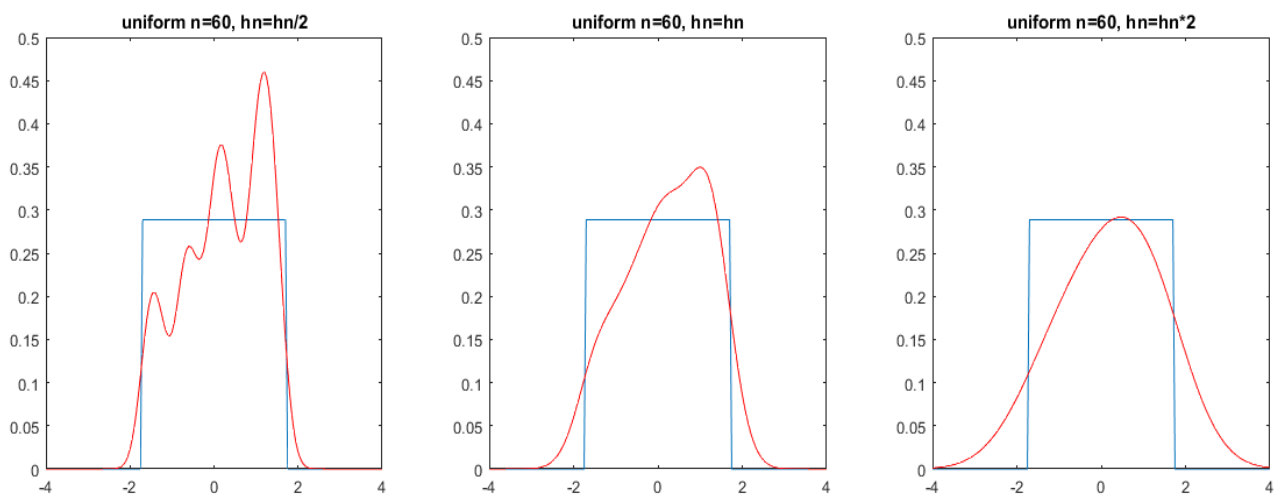


Рис. 20 Равномерное распределение,  $n=60$

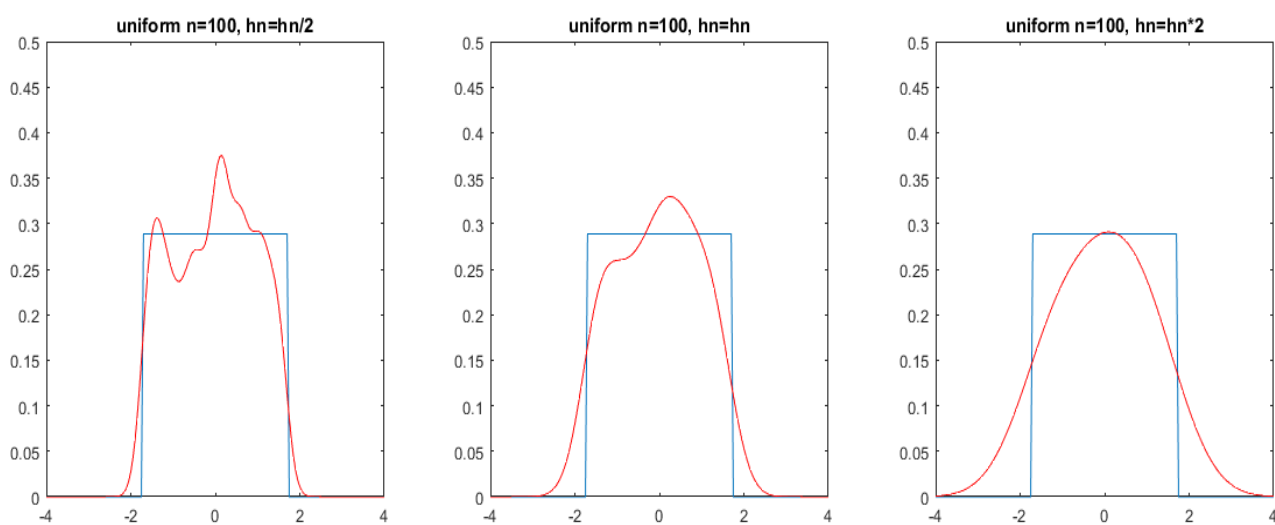


Рис. 21 Равномерное распределение,  $n=100$

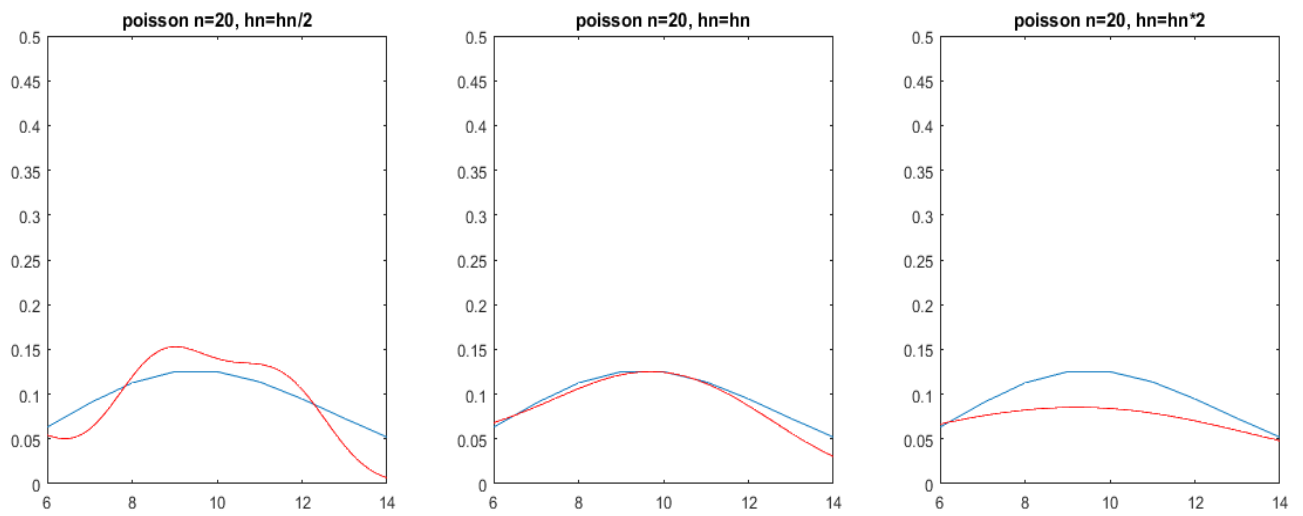


Рис. 22 Распределение Пуассона,  $n=20$

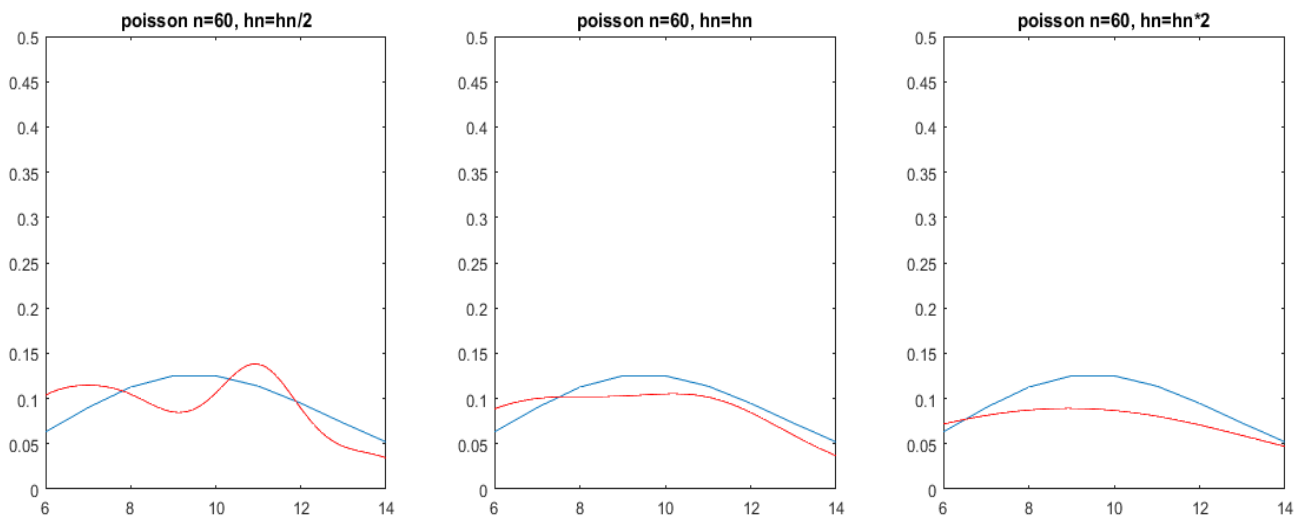


Рис. 23 Распределение Пуассона,  $n=60$



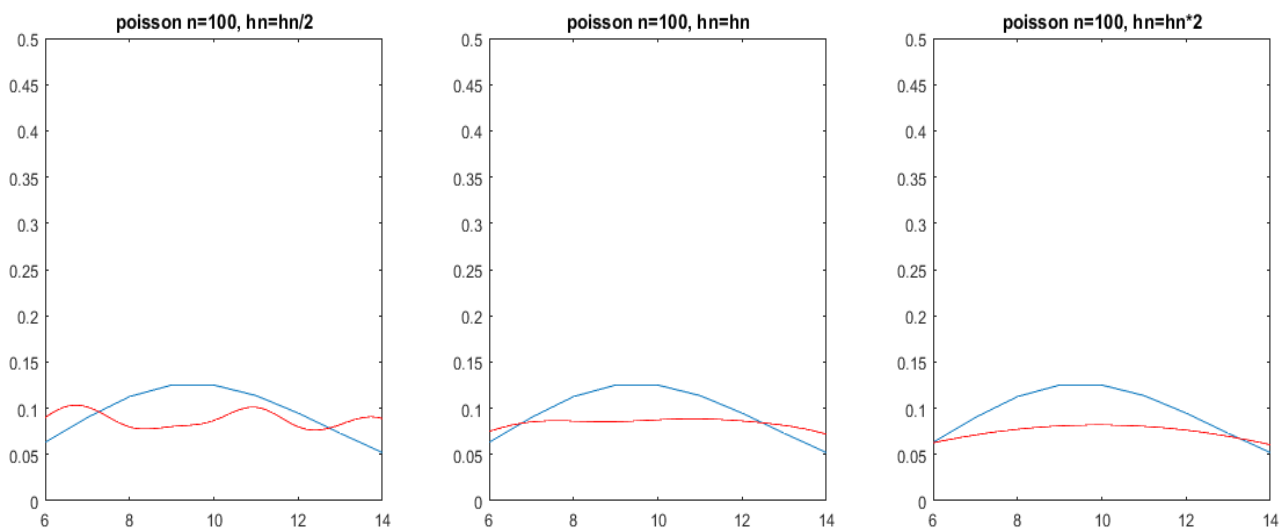


Рис. 24 Распределение Пуассона,  $n=100$

## 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 видно, что для случайных величин, сгенерированных по закону распределения Коши, более характерны всплески.

### 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. Для остальных распределений таких выбросов не наблюдается.

### 5.3 Доля и теоретическая вероятность выбросов

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.

Вновь для распределения Коши отчетливо видно большое количество выбросов.

### 5.4 Эмпирическая функция и ядерные оценки плотности распределения

Можем наблюдать на иллюстрациях с эмпирической функцией распределения, что она 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 с ростом размера выборки.

В зависимости от особенностей распределений для их описания лучше подходят разные параметры в ядерной оценке: для равномерного распределения и распределения Пуассона и нормального лучше подойдет параметр  $h_n$ , для равномерного распределения -  $2h_n$ .

Такие значения дают вид ядерной оценки наиболее близкий к плотности, характерной для данных распределений. Также можно

увидеть, что чем больше коэффициент при параметре сглаживания, тем меньше изменений знака производной у аппроксимирующей функции, вплоть до того, что при  $2h_n$  функция становится унимодальной на рассматриваемом промежутке.

Также видно, что при  $2h_n$  по полученным приближениям становится сложно сказать плотность вероятности какого распределения они должны повторять, так как они очень похожи между собой.