Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторным работам № 1, 2

По дисциплине «Теория вероятностей и Математическая статистика»

Выполнил
Студент гр. 3630201/80101 ________ Печеный Н. А.

Руководитель
к.ф.-м.н., доцент ________ Баженов А. Н.

«___»____________ 2020г.

Санкт-Петербург
2020

Содержание

Список иллюстраций

Список таблиц

1 Постановка задачи

Для 5 распределений:

- Нормальное распределение $N(x, 0, 1)$
- Распределение Коши $C(x, 0, 1)$
- Распределение Лапласа $L(x, 0, 1/\sqrt{2})$
- 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: $\overline{x}, med\ x, z_R, z_Q, z_{tr}$. Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

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

Вычислить оценку дисперсии по формуле:

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

Представить полученные данные в виде таблиц.

2 Теория

2.1 Рассматриваемые распределения

Ниже приведены плотности вероятности для рассматриваемых распределений [2]:

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

- Распределение Коши
$$C(x, 0, 1/\sqrt{2}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{4}$$

- Распределение Лапласа
$$L(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \tag{5}$$

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

- Равномерное распределение
$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| <= \sqrt{3} \\ 0, & |x| >= \sqrt{3} \end{cases} \tag{7}$$

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].

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

$$med\,x = \begin{cases} x_{l+1}, & n = 2l+1 \\ \dfrac{x_l + x_{l+1}}{2}, & n = 2l \end{cases} \tag{9}$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_1 + x_n}{2} \tag{10}$$

- Полусумма квартилей

  Выборочная квартиль z_p порядка p определяется формулой

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

  Полусумма квартилей:

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \tag{11}$$

- Усечённое среднее

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

2.4.2 Характеристики рассеяния

Выборочная дисперсия

$$D = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \tag{13}$$

3 Реализация

Расчёты проводились в среде аналитических вычислений Maxima. Для генерации выборок и создания и отрисовки графиков были использованы библиотечные функции среды разработки. Код скрипта представлен в репозитории на GitHub, ссылка на репозиторий находится в **Приложении А**.

4 Результаты

4.1 Гистограмма и график плотности распределения

Рис. 1: Нормальное распределение $N(x, 0, 1)$

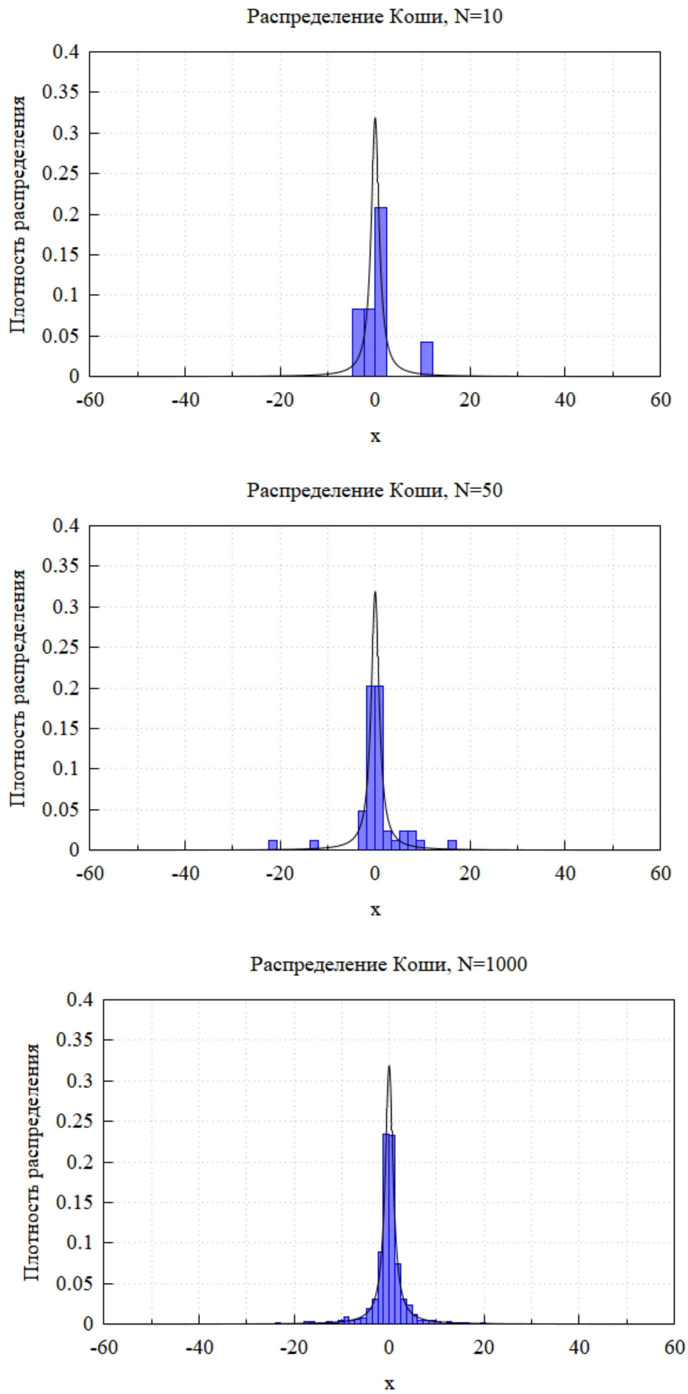

Рис. 2: Распределение Коши $C(x, 0, 1)$

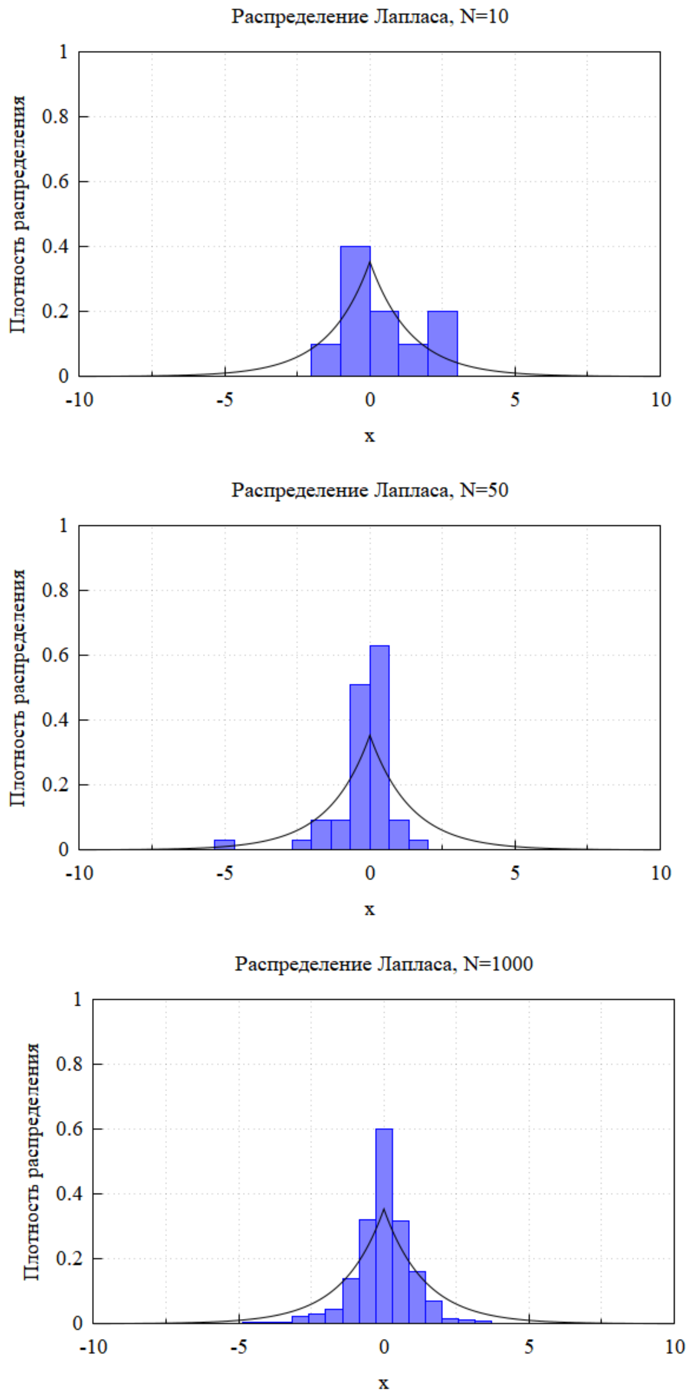

Рис. 3: Распределение Лапласа $L(x, 0, 1/\sqrt{2})$

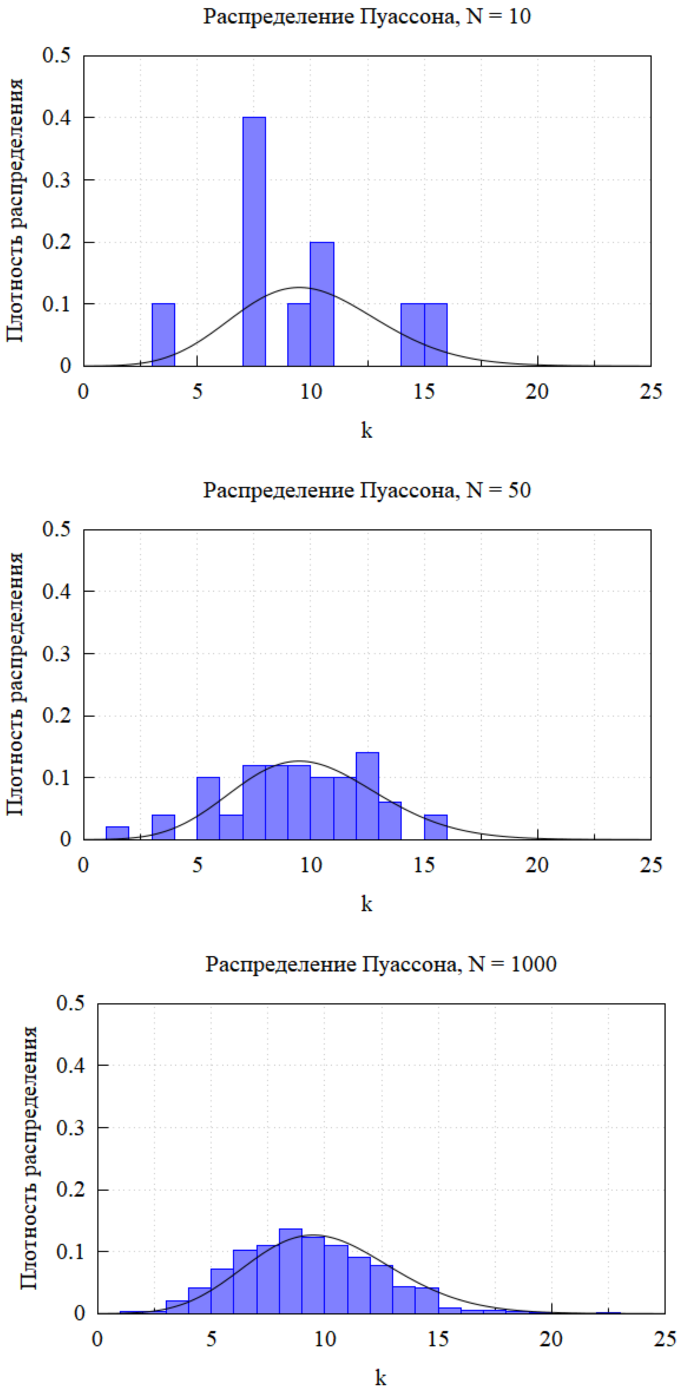

Рис. 4: Распределение Пуассона $P(k, 10)$

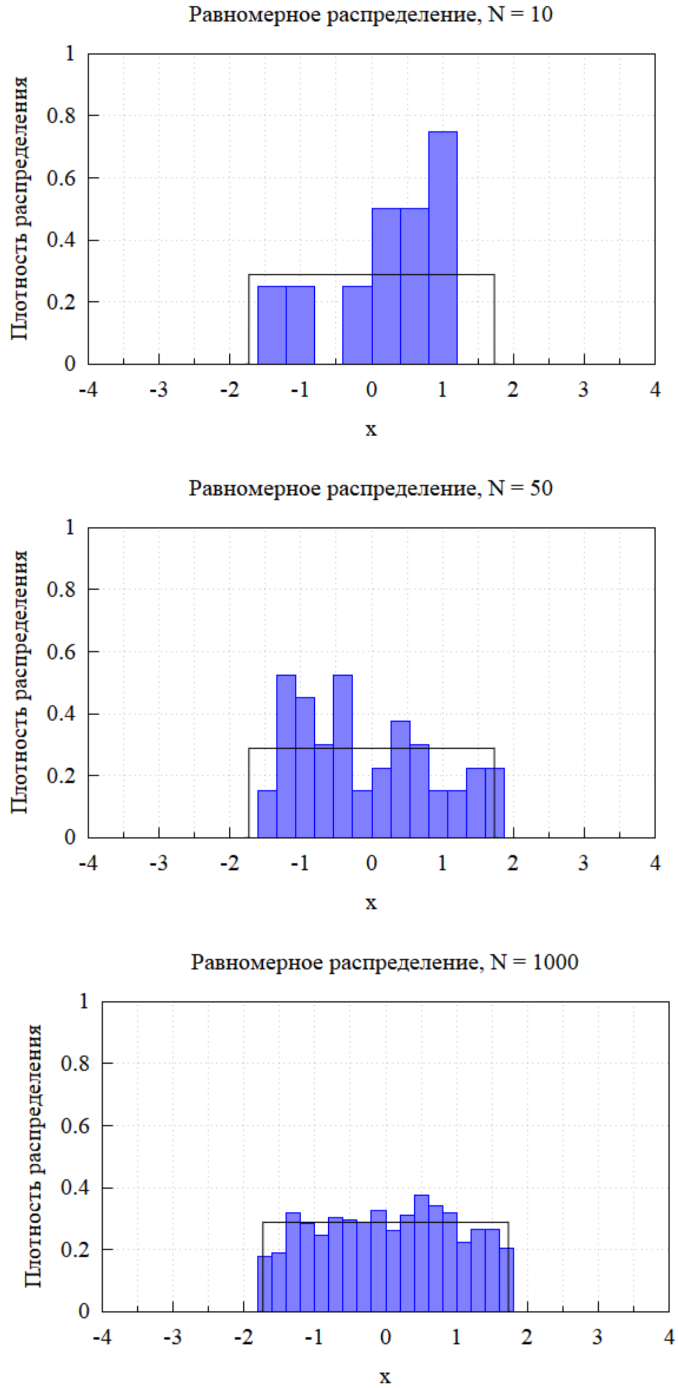

Рис. 5: Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

4.2 Характеристики положения и рассеяния

Погрешность среднего значения характеристики выборки рассчитывалась как $\Delta_z = \sqrt{D(z)}$.

Нормальное распределение						
		$\overline{x}$ (8)	$med\ x$ (9)	z_R (10)	z_Q (11)	z_{tr} (12)
$N = 10$	$E(z)$ (1) $\pm \Delta_z$	0.0 ± 0.3	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.4
	$D(z)$ (2)	0.0978	0.4922	0.4740	0.5120	0.1674
$N = 100$	$E(z) \pm \Delta_z$	0.0 ± 0.1	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.1
	$D(z)$	0.0102	0.4970	0.5166	0.5312	0.0210
$N = 1000$	$E(z) \pm \Delta_z$	0.00 ± 0.03	0.0 ± 0.7	0.0 ± 0.6	0.0 ± 0.7	0.00 ± 0.04
	$D(z)$	0.0010	0.5211	0.4312	0.5052	0.0019

Таблица 1: Характеристики выборок нормального распределения

Распределение Коши						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	1 ± 15	0 ± 23	-1 ± 20	2 ± 58	0 ± 19
	$D(z)$	240	548	409	3467	371
$N = 100$	$E(z) \pm \Delta_z$	2 ± 76	2 ± 56	1 ± 19	1 ± 56	3 ± 117
	$D(z)$	5797	3084	371	3135	13831
$N = 1000$	$E(z) \pm \Delta_z$	-1 ± 23	-9 ± 337	1 ± 38	0 ± 34	-2 ± 45
	$D(z)$	517	11385	1435	1171	1996

Таблица 2: Характеристики выборок распределения Коши

Распределение Лапласа						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	0.0 ± 0.3	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.4
	$D(z)$	0.106	0.485	0.559	0.501	0.173
$N = 100$	$E(z) \pm \Delta_z$	0.0 ± 0.1	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.1
	$D(z)$	0.010	0.498	0.486	0.484	0.019
$N = 1000$	$E(z) \pm \Delta_z$	0.00 ± 0.03	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.00 ± 0.04
	$D(z)$	0.001	0.528	0.467	0.540	0.002

Таблица 3: Характеристики выборок распределения Лапласа

Распределение Пуассона						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	10 ± 1	10 ± 2	10 ± 2	10 ± 2	10 ± 1
	$D(z)$	1.05	5.01	4.57	5.28	1.77
$N = 100$	$E(z) \pm \Delta_z$	10.0 ± 0.3	10.0 ± 2.3	9.9 ± 2.3	10.0 ± 2.2	10.0 ± 0.4
	$D(z)$	0.10	5.13	5.17	4.97	0.19
$N = 1000$	$E(z) \pm \Delta_z$	9.99 ± 0.09	9.95 ± 2.23	9.98 ± 2.24	10.02 ± 2.22	9.99 ± 0.14
	$D(z)$	0.010	4.987	5.004	4.937	0.019

Таблица 4: Характеристики выборок распределения Пуассона

Равномерное распределение						
		$\overline{x}$	$med\ x$	z_R	z_Q	z_{tr}
$N = 10$	$E(z) \pm \Delta_z$	0.0 ± 0.3	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.4
	$D(z)$	0.096	0.499	0.496	0.493	0.166
$N = 100$	$E(z) \pm \Delta_z$	0.0 ± 0.1	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.1
	$D(z)$	0.011	0.496	0.536	0.478	0.020
$N = 1000$	$E(z) \pm \Delta_z$	0.00 ± 0.03	0.0 ± 0.7	0.0 ± 0.7	0.0 ± 0.7	0.00 ± 0.04
	$D(z)$	0.001	0.526	0.517	0.479	0.002

Таблица 5: Характеристики выборок равномерного распределения

5 Заключение

5.1 Гистограмма и график плотности распределения

В рамках данной работы были построены гистограммы с наложенными на них графиками плотности распределения случайных величин. По результатам проделанной работы можно 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.
Следует обратить внимание на тот факт, что максимум графика плотности распределения зачастую совпадает с максимумом (хотя бы в локальном смысле) гистограммы распределения, что особенно хорошо видно при больших выборках. Также наблюдаются всплески гистограмм, что лучше всего прослеживается на распределении Коши.

5.2 Характеристики положения и рассеяния

В рамках работы были вычислены различные характеристики положения и рассеяния для рассматриваемых видов распределения. Следует отметить высокие значения дисперсии почти для всех характеристик в распределении Коши, которые имеют тенденцию расти с ростом выборки, что объясняется склонностью данного распределения к выбросам. Такого не наблюдается в других распределениях.

Список литературы

[1] Histogram `https://en.wikipedia.org/wiki/Histogram` Дата обращения 5.12.2020

[2] Теоретическое приложение к лабораторным работам №1-4 по дисциплине «Математическая статистика». – СПб.: СПбПУ, 2020. – 12 с

Приложение А. Репозиторий с исходным кодом

Ссылка на репозиторий GitHub с исходным кодом: `https://github.com/pchn/TeorVer`