

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторной работе

Построение боксплотов Тьюки и оценка выбросов

По дисциплине «Теория вероятностей и Математическая статистика»

Выполнил

Студент гр. 3630201/80101

М.Д. Маляренко

Руководитель

к.ф.-м.н., доцент

А.Н. Баженов

« ____ » _____ 2020г.

Санкт-Петербург
2020

Содержание

1	Постановка задачи	5
2	Теория	6
2.1	Боксплот Тьюки	6
2.2	Теоретическая вероятность выбросов	6
3	Реализация	7
4	Результаты	8
4.1	Боксплоты Тьюки	8
4.2	Сравнение теоретической вероятности и экспериментальной доли выбросов	10
	Заключение	11
	Список литературы	12
	Приложение А. Репозиторий с исходным кодом	13

Список иллюстраций

1	Боксплот выборок нормального распределения	8
2	Боксплот выборок распределения Лапласа	8
3	Боксплот выборок распределения Коши	9
4	Боксплот выборок распределения Пуассона	9
5	Боксплот выборок равномерного распределения	10

Список таблиц

1	Сравнение экспериментальной доли и теоретической вероятности выбросов	10
---	---	----

1 Постановка задачи

Заданы 5 распределений случайных величин:

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

Для каждого распределения необходимо сгенерировать выборки размером в 20, 100 элементов. Для каждой выборки построить боксплот Тьюки.

Сгенерировав каждую выборку 1000 раз и определить среднюю экспериментальную долю выбросов и сравнить её с теоретической вероятностью выбросов P_B^T для рассматриваемого распределения.

2 Теория

2.1 Боксплот Тьюки

Боксплот Тьюки (а.к.а. ящик с усами) – график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей.[1]

Боксплот представляет собой прямоугольник с исходящими из боков отрезками-усами и прямой перпендикулярной усам линией внутри прямоугольника. Боксплот располагается на числовой оси. Левая и правая границы прямоугольника определяют соответственно первую выборочную квартиль Q_1 и третью выборочную квартиль Q_3 , центральная линия указывает положение медианы выборки. Отрезки-усы определяют границы статистически значимой выборки, и вычисляются по следующим формулам:

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1); \quad X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1) \quad (1)$$

, где X_1 – нижняя граница усов, X_2 – верхняя граница усов. Элементы выборки, выходящие за границы, определённой усами, называются выбросами и обозначаются небольшими кружками.

2.2 Теоретическая вероятность выбросов

Найдя по функции распределения теоретические значения первой и третьей квартили Q_1^T и Q_3^T , по формуле (1) можно вычислить теоретические границы усов X_1^T и X_2^T .

Теоретическая вероятность выбросов вычисляется по формуле:[2]

- Для непрерывных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)) \quad (2)$$

- Для дискретных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)) \quad (3)$$

, где $F(X)$ – функция распределения.

3 Реализация

Построение графиков и расчёты теоретической вероятности и экспериментальной доли выбросов производилось в среде аналитических вычислений *Maxima*. Графики построены с помощью интегрированной утилиты *gnuplot*. Полный код скрипта для среды *Maxima* представлен в репозитории *GitHub*.

4 Результаты

4.1 Боксплоты Тьюки

По полученным выборкам были построены боксплоты Тьюки. На Рис. 1-5 представлены боксплоты выборок из 20 и 100 элементов, соответствующих заданным распределениям.

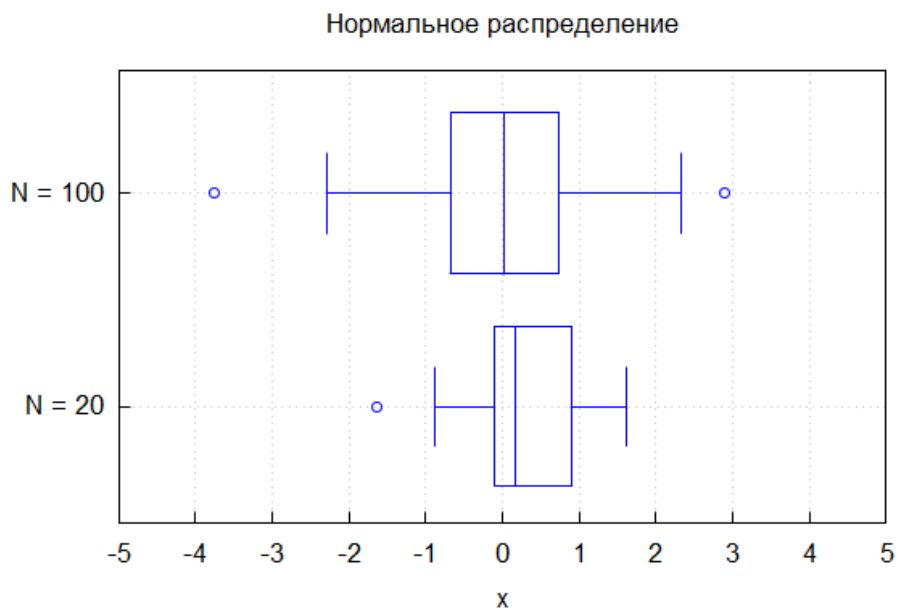


Рис. 1: Боксплот выборок нормального распределения

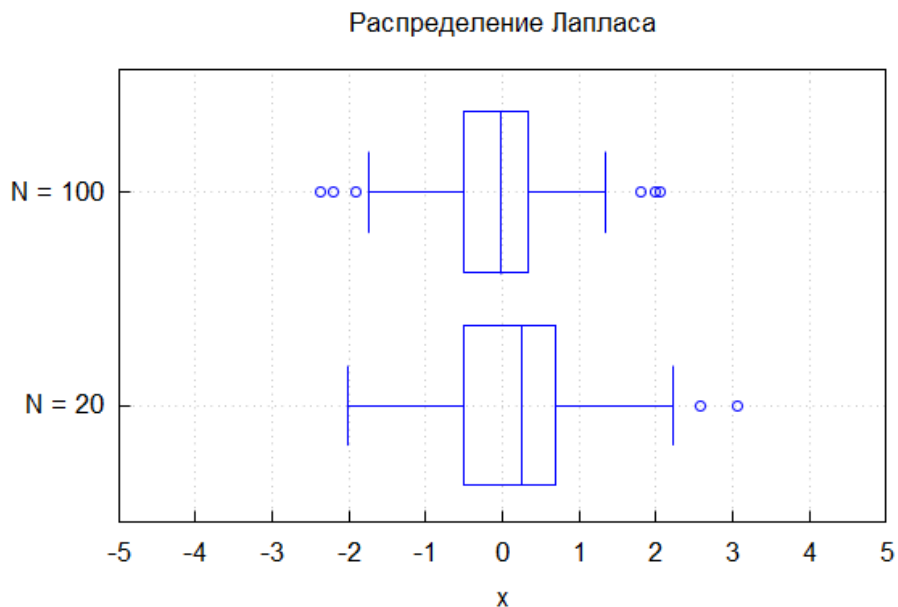


Рис. 2: Боксплот выборок распределения Лапласа

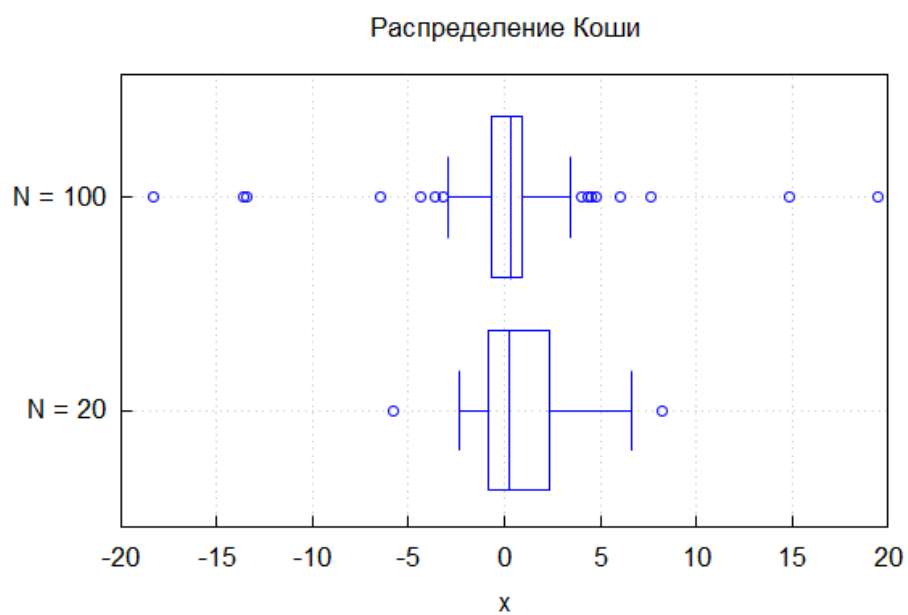


Рис. 3: Боксплот выборок распределения Коши

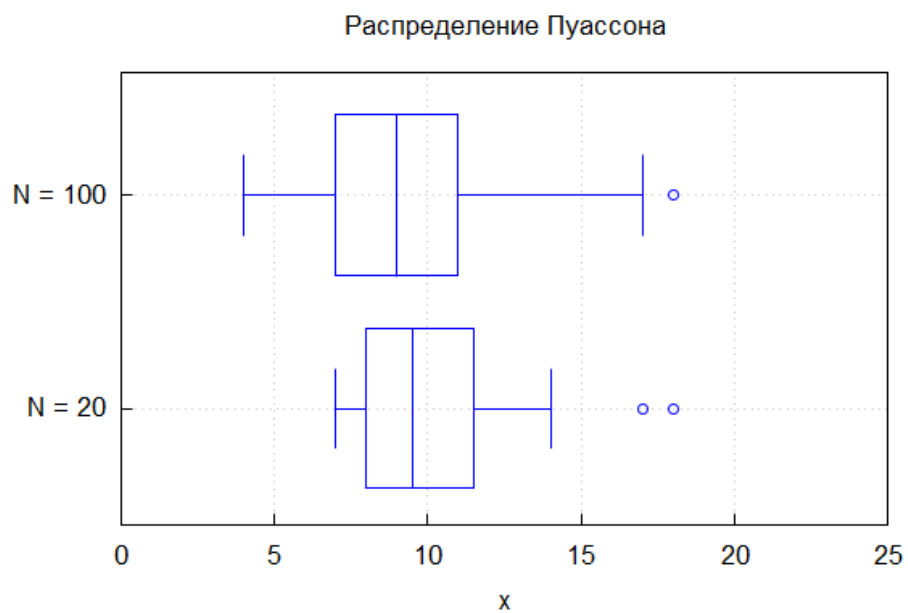


Рис. 4: Боксплот выборок распределения Пуассона

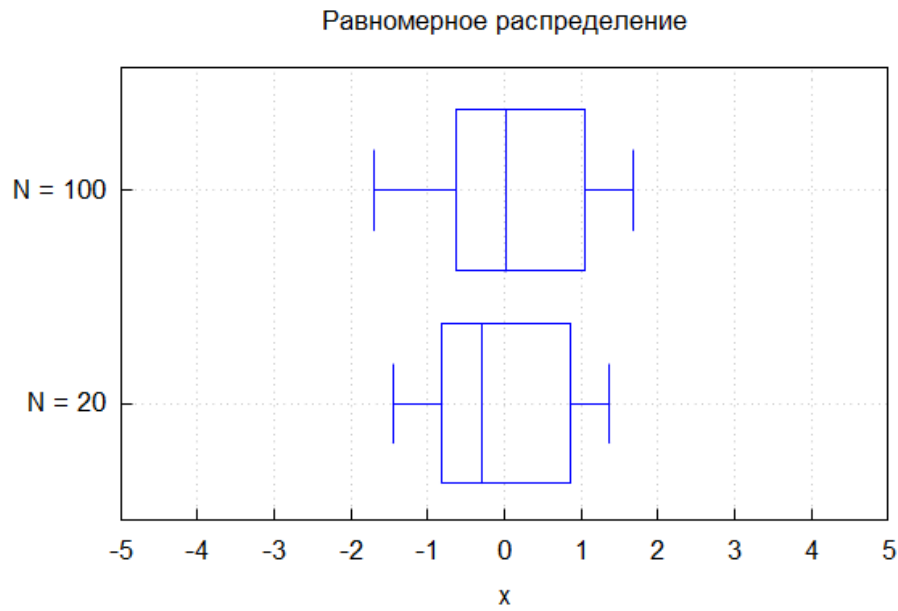


Рис. 5: Боксплот выборок равномерного распределения

4.2 Сравнение теоретической вероятности и экспериментальной доли выбросов

Для 1000 экземпляров каждой выборки была вычислена средняя доля выбросов. Теоритическая вероятность выброса P_B^T рассчитана по формуле (2).

Погрешность средней доли выброса рассчитана следующим образом

$$\Delta_z = \sqrt{\left(\overline{z^2} - \bar{z}^2\right)} \quad (4)$$

Значение экспериментальной доли выброса округлено согласно погрешности. Результаты вычислений представлены в Таблице 1.

Распределение	N	Доля выбросов	P_B^T
Нормальное	20	0.00 ± 0.02 (4)	0.0069
	100	0.007 ± 0.009	
Лапласа	20	0.06 ± 0.06	0.0625
	100	0.06 ± 0.03	
Коши	20	0.16 ± 0.08	0.156
	100	0.16 ± 0.04	
Пуассона	20	0.01 ± 0.02	0.0099
	100	0.007 ± 0.009	
Равномерное	20	0.0 ± 0.0	0.0
	100	0.0 ± 0.0	

Таблица 1: Сравнение экспериментальной доли и теоретической вероятности выбросов

Заключение

В результате лабораторной работы были построены боксплоты Тьюки по выборкам, согласно пяти заданным распределениям. По ним визуально были оценены мощности выбросов распределений.

Также была рассчитана теоретическая вероятность выбросов для каждого распределения. Это значение было сравнено с экспериментальными данными – средней долей выбросов на 1000 выборок. Сравнение показало, что экспериментальные и теоретические значения очень близки (одного порядка в худшем случае) и более сходятся, чем больше мощность выборки.

Список литературы

- [1] Теоретическое приложение к лабораторным работам №1-4 по дисциплине «Математическая статистика». – СПб.: СПбПУ, 2020. – 12 с
- [2] Box Plot – URL https://en.wikipedia.org/wiki/Box_plot. Дата обращения 23.11.2020

Приложение А. Репозиторий с исходным кодом

Исходный код скрипта для среды аналитических вычислений *Maxima* находится в репозитории GitHub – URL <https://github.com/malyarenko-md/TeorVer>