МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

**«САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО»**

Институт Прикладной математики и механики

Кафедра: «Телематика (при ЦНИИ РТК)»

Направление 02.03.01 Математика и компьютерные науки

Отчёт по дисциплине:
«Теория вероятности и математическая статистика »

Лабораторная работа № 1

«Типы распределений и построение соответствующих им гистограмм»

Обучающийся: ______________ Фомина Дарья Дмитриевна

Руководитель: ______________ Баженов Александр Николаевич

«____» ____________ 20__ г.

Санкт-Петербург 2020

Содержание

1 Постановка задачи

Для пяти распределений:

1. Нормальное распределение $\mu = 0, \sigma = 1$;
2. Распределние Коши с параметрами $\mu = 0, \lambda = 1$;
3. Распределение Лапласа с параметрами $\mu = 0, \lambda = \frac{1}{\sqrt{2}}$;
4. Распределение Пуассона с параметром $\mu = 10$;
5. Равномерное распределение с параметрами $a = -\sqrt{3}, b = \sqrt{3}$;

Сгенерировать выборки размером 10, 100, 1000 элементов. Для каждой выборки нужно построить гистрограмму и теоретическую функцию плотности вероятности на одном рисунке.

2 Математическое описание

2.1 Рассматриваемые распределения

2.1.1 Нормальное распределение

Нормальное распределение (также называемое распределением Гаусса или Гаусса - Лапласа) — распределение вероятностей, которое в одномерном случае задаётся функцией плотности вероятности, совпадающей с функцией Гаусса:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Нормальное распределение зависит от двух параметров μ - математического ожидания и σ - среднеквадратического отклонения.

Стандартным нормальным распределением называется нормальное распределение с математическим ожиданием $\mu = 0$ и стандартным отклонением $\sigma = 1$. [2]

Функция плотности вероятности для стандартного нормального распределения
при $\sigma = 1$ и $\mu = 0$:

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

Функция распределения имеет следующий вид:

$$F(x) = 0.5 + \Phi_0\left(\frac{x-\mu}{\sigma}\right) \tag{3}$$

При $\sigma = 1$ и $\mu = 0$:

$$F(x) = 0.5 + \Phi_0(x), \tag{4}$$

где $\Phi_0(x)$ - функция Лапласа.

Сгенерировать случайную величину, соответствующую данному нормальному распределению можно по следующей формуле:

$$x_i = \sum_{i=1}^{12} r_i - 6, \tag{5}$$

где r_i - стандартная равномерная случайная величина из интервала (0, 1),
x_i - стандартные нормальные случайные числа. [1]

2.1.2 Распределение Коши

Распределение Коши описывает отношение двух нормально распределенных случайных величин. В отличие от других распределений, для распределения Коши не определены матожидание и дисперсия. Для описания распределения используются коэффициенты сдвига x_0 и масштаба γ. [2]

Распределение Коши является бесконечно делимым: сумма независимых случайных величин, распределённых по Коши, также распределена по Коши.

Плотность вероятности для распределения Коши:

$$f(x, x_0, \gamma) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x-x_0}{\gamma}\right)^2\right)} \tag{6}$$

При $\gamma = 1$ и $x_0 = 0$:

$$f(x, 0, 1) = \frac{1}{\pi\left(1 + x^2\right)} \tag{7}$$

Функция распределения имеет следующий вид:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} arctg \frac{(x - x_0)}{\gamma} \tag{8}$$

При $\gamma = 1$ и $x_0 = 0$:

$$F(x) = \frac{1}{2} + \frac{arctgx}{\pi} \tag{9}$$

Сгенерировать случайную величину, соответствующую данному распределению Коши можно по следующей формуле:

$$x_i = tg(2\pi r_i), \tag{10}$$

где r_i - стандартная равномерная случайная величина из интервала (0, 1). [1]

2.1.3 Распределение Лапласа

Распределение Лапласа (двойное эспоненциальное распределение) — в теории вероятностей это непрерывное распределение случайной величины. Распределение применяется для моделирования обработки сигналов, в моделировании биологических процессов, экономике и финансах. [2]

Плотность вероятности для распределения Лапласа:

$$f(x, \beta, \alpha) = \frac{\alpha}{2} e^{-\alpha|x-\beta|}, \tag{11}$$

где $\alpha > 0$ — параметр масштаба,$-\infty < \beta < +\infty$ — параметр сдвига.

При $\beta = 0$ и $\alpha = \frac{1}{\sqrt{2}}$:

$$f(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{2\sqrt{2}} e^{-\frac{|x|}{\sqrt{2}}} \tag{12}$$

Функция распределения задается следующим образом:

$$F_L(x) = \begin{cases} \frac{1}{2}\, e^{\frac{x}{\sqrt{2}}}, & x \geq 0; \\ 1 - \frac{1}{2}\, e^{-\frac{x}{\sqrt{2}}}, & x < 0 \end{cases} \tag{13}$$

Сгенерировать случайную величину, соответствующую анному распределению Лапласа можно по следующей формуле:

$$x_i = \sqrt{2}\, ln\,(r_{i+1}/r_i), \tag{14}$$

где r_i - стандартная равномерная случайная величина из интервала (0, 1). [1]

2.1.4 Распределение Пуассона

Распределение Пуассона — распределение дискретного типа случайной величины, представляющей собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга. [2]

Плотность вероятности для распределения Пуассона:

$$f(k, \lambda) = \frac{\mu^x}{x!}\, e^{-\mu}, \tag{15}$$

где $\mu > 0$ — математическое ожидание случайной величины (среднее количество событий за фиксированный промежуток времени),

При $\mu = 10$:

$$f(k, 10) = \frac{10^x}{x!}\, e^{-10} \tag{16}$$

Функция распределения задается следующим образом:

$$F(k) = \begin{cases} 0, & x \leq 0; \\ e^{-\mu} \sum_{i=0}^{k} \frac{(\mu)^i}{i!}, & k < x \leq k+1 \end{cases} \tag{17}$$

Случайная величина, соответствующая данному распределению Пуассона может быть сгенерирована по следующему алгоритму:

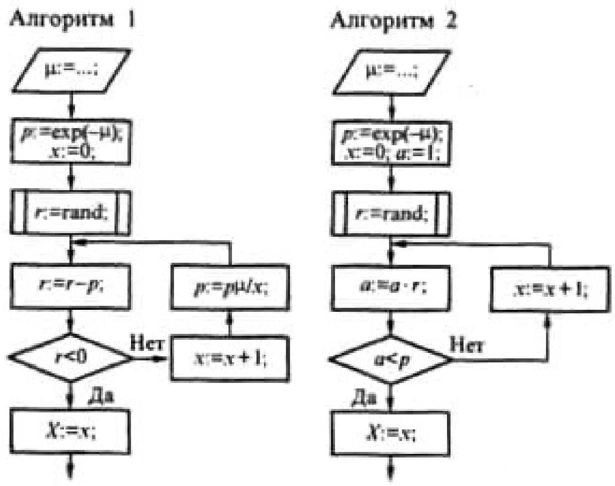

Рис. 1: Блок-схемы алгоритмов генерирования пуассоновских случайных чисел

Алгоритм 1 реализует стандартный способ имитационного моделирования дискретных величин. Алгоритм 2 основан на связи пуассоновского распределения с показательным и эрланговским распределениями. (рис. 1)

2.1.5 Равномерное распределение

Непрерывное равномерное распределение — распределение случайной вещественной величины, принимающей значения, принадлежащие некоторому промежутку конечной длины, характеризующееся тем, что плотность вероятности на этом промежутке почти всюду постоянна.[2]

Плотность вероятности для равномерного распределения:

$$f(x,\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha}, & x \in [\alpha,\beta] \\ 0, & x \notin [\alpha,\beta] \end{cases}, \tag{18}$$

где α, β- границы области возможных значений случайной величины. Левая граница α области возможных значений случайной величины является параметром положения, а длина $\beta - \alpha$ области возможных значений - параметром масштаба.

При $a = -\sqrt{3}$ и $b = \sqrt{3}$:

$$f(x,-\sqrt{3},\sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & x \in [-\sqrt{3},\sqrt{3}] \\ 0, & x \notin [-\sqrt{3},\sqrt{3}] \end{cases}. \tag{19}$$

2.2 Гистограмма

Гистограмма в математической статистике — это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него.

Для построения гистограммы промежуток, которому принадлежат элементы выборки, разбивается на несколько интервалов. Эти интервалы называются бинами. Далее для каждого бина подсчитывается количество элементов выборки, попавших в него. Затем на графике строятся прямоугольники, основание которых совпадает с бинами, а высота пропорциональна количеству, попавших в этот бин.

Количество бинов определяется по формуле:

$$n = [\sqrt{N}], \tag{20}$$

где N - количество элементов в выборке.

Гистограмма задается следующей функцией:

$$h(x) \begin{cases} \frac{n_k}{Nb}, & x \in [x_{min}, x_{max}] \\ 0, & \text{иначе} \end{cases}, \tag{21}$$

где b - ширина одного бина, N - количество элементов в выборке, x_{min}, x_{max} - минимальный и максимальный элементы выборки, n_k- количество элементов в k-том бине.

3 Особенности реализации

Данная лабораторная работа реализована на языке программирования Python с использованием IDLE Python 3.9 и библиотек NumPy и Matplotlib, предназначенной для графического представления графиков и гистограмм.

Для каждого из рассматриваемых распределений были реализованы классы. Каждый класс содержит такие методы как:

- `def_init_(distr)`- конструктор, который инициализирует поля класса: принимает название распределения и параметры для данного распределения;
- x()->float - возвращает случайную величину;
- def f(x)->float - возвращает функцию плотности вероятности от заданной точки;
- def F(x)->float - возвращает значение функции распределения в точке x.
- def x(distr)->float - принимает название распределения и генерирует случайную величину в соответствии с заданным распределением.
- def interval (distr) - возращает интервал для построения графиков.

Методы классов распределений реализуют указанные в разделе 2.1 формулы для вычислений соответствующих величин.

Так же был реализован класс Histogram, реализующий представление выборки в гистограмму. Количество бинов вычисляется по формуле 20. Высота столбца в гистограмме вычисляется по формуле 21.

В приложении А размещена ссылка на исходный код.

4 Результаты работы программы

На рис. 2 - 6 представлены результаты работы программы. На каждом из рисунков представлены график функции плотности вероятности f(x) и график функции распределения F(x) для выборок размерами 10, 100 и 1000 элементов.

На рис. 2 представлены графики для нормального распределения. Теоретические значения вычислены по формулам 2 и 4.

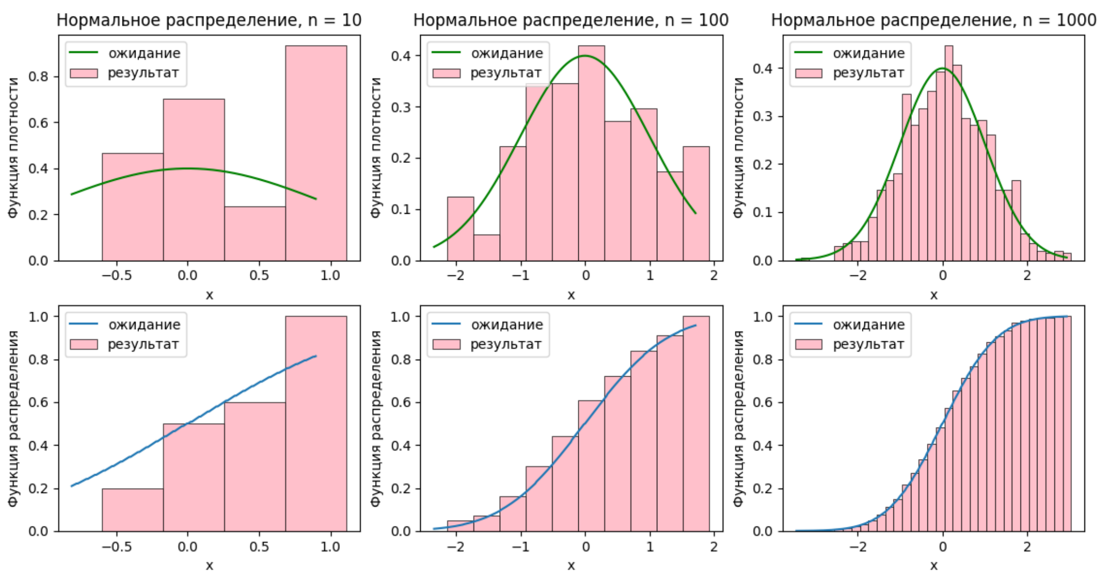

Рис. 2: Графики функций нормального распределения.

На рис. 3 представлены графики для распределения Коши. Теоретические значения вычислены по формулам 7 и 9.

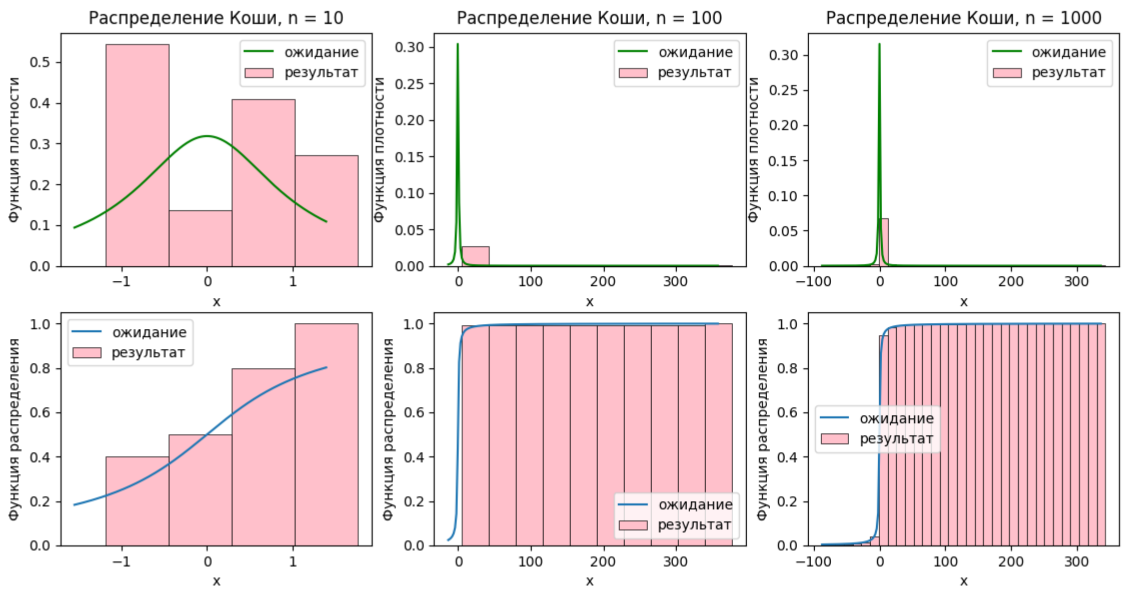

Рис. 3: Графики функций распределения Коши.

На рис. 4 представлены графики для распределения Лапласа. Теоретические значения вычислены по формулам 12 и 13.

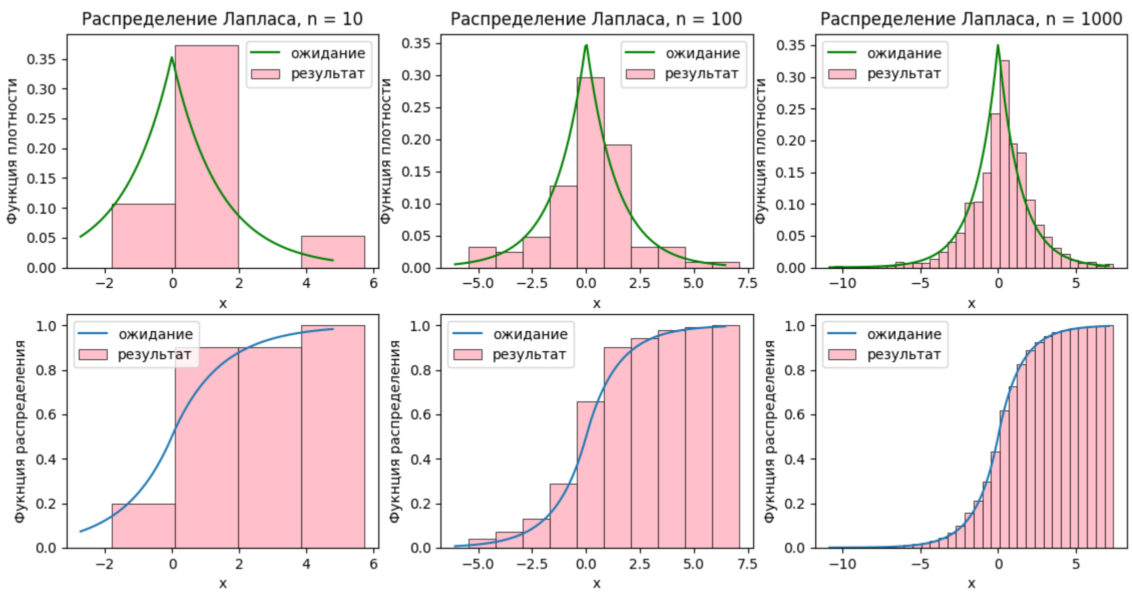

Рис. 4: Графики функций распределения Лапласа.

На рис. 5 представлены графики для распределения Пуассона. Теоретические значения вычислены по формулам 16 и 17.

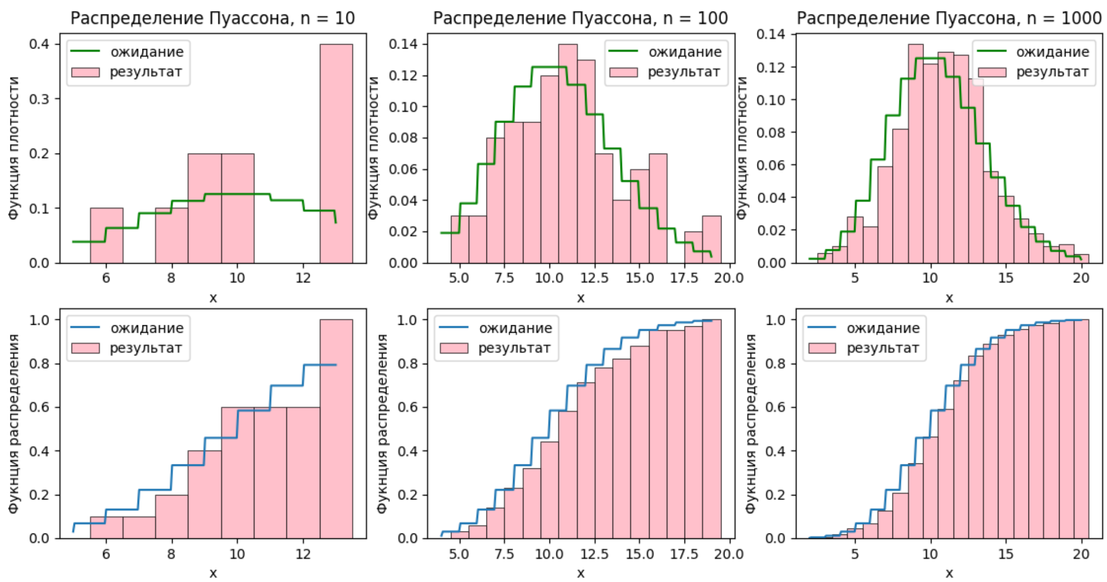

Рис. 5: Графики функций распределения Пуассона.

На рис. 6 представлены графики для равномерного распределения. Теоретические значения вычислены по формуле 19.

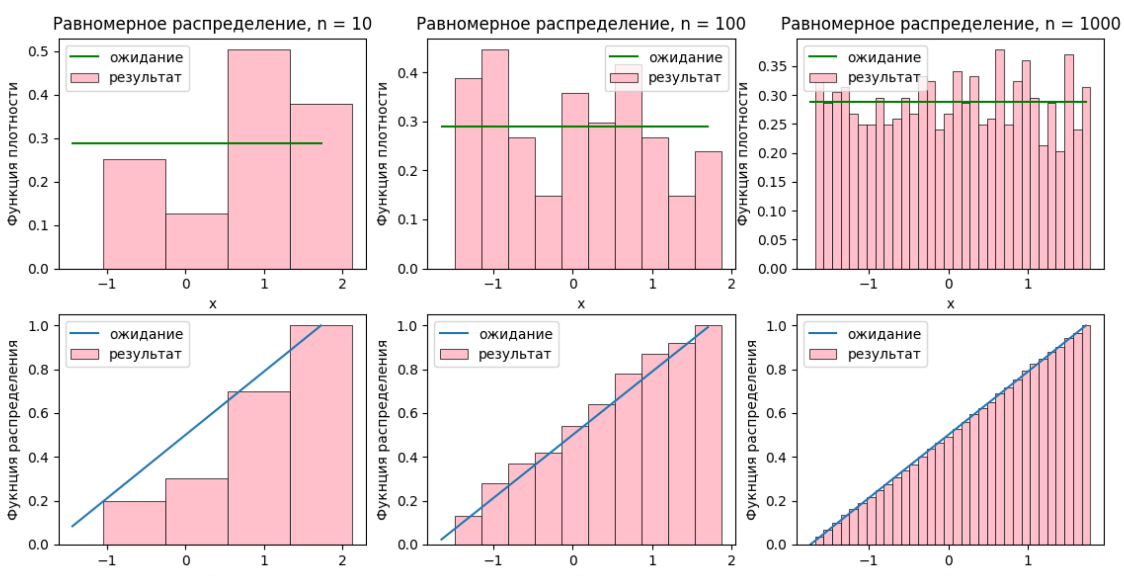

Рис. 6: Графики функций равномерного распределения.

Заключение

В ходе данной лабораторной работы были сгенерированы выборки размером 10, 100 и 1000 для пяти распределений: нормального распределения, распределения Коши, Лапласа, Пуассона и равномерного распределения. Так же были построены гистограммы полученных выборок и теоретических функций.

При изучении полученных гистограмм было установлено, что при малой выборке гистограммы лишь отдаленно напоминают теоретические функции, но при большой выборке гистограммы демонстрируют близкий к теоретическому результат.

Отчёт подготовлен с помощью компилятора pdflatex и среды разработки TeXworks.

Список литературы

[1] Р.Н. Вадзинский. Справочник по вероятностным распределениям. СПб: Наука, 2001г. 294с.

[2] Ю.Д. Максимов. Математика. Теория вероятностей и случайных процессов: Учебн. пособие Под ред. В.И. Антонова. - СПб.: Изд-во Политехн. ун-та, 2008. - 384 с.

A Приложение

Исходный код программы размещен на сервисе GitHub.
Ссылка на репозиторий: https://github.com/foria1405/TV/tree/main/lab1