Лабораторная работа № 3.1. Датчик непрерывной случайной величины

1. Постановка задачи:

- Требуется реализовать датчик непрерывной случайной величины в виде функции для метода обратной функции, со следующей плотностью вероятности:

$$p(x) = \frac{7(x+1)^{3/4}}{4(2^{7/4}-1)},\ 0 \le x \le 1.$$

- Требуется аналитически получить из заданной плотности распределения функцию распределения и обратную ей функцию распределения.
- Вычислить математическое ожидание и дисперсию моделируемой случайной величины. Для конечного матожидания и дисперсии моделируемой случайной величины требуется аналитически получить формулу для вычисления матожидания и дисперсии.
- С помощью реализованного датчика получить выборку, вычислить выборочное среднее и несмещённую выборочную дисперсию, сравнить с теоретическим.

Примечание 3.1.1: метод обратной функции - это метод генерации случайных величин, который используется для получения случайных чисел с заданным распределением.

2. Математическая составляющая задачи:

- **Аналитическое получение функции распределения F(x):**

Дана плотность вероятности:

$$p(x) = \frac{7 \cdot (x+1)^{3/4}}{4 \cdot (2^{7/4}-1)}, \quad 0 \le x \le 1$$

Функция распределения F(x) получается интегрированием плотности вероятности:

$$F(x) = \int_0^x p(t)\,dt$$

Подставим p(t):

$$F(x) = \int_0^x \frac{7 \cdot (t+1)^{3/4}}{4 \cdot (2^{7/4}-1)}\,dt$$

Вынесем константу за знак интеграла:

$$F(x) = \frac{7}{4 \cdot (2^{7/4}-1)} \int_0^x (t+1)^{3/4}\,dt$$

Интеграл от $(t+1)^{3/4}$:

$$\int (t+1)^{3/4}\,dt = \frac{4}{7}(t+1)^{7/4} + C$$

Подставим пределы интегрирования:

$$F(x) = \frac{7}{4\cdot(2^{7/4}-1)} \cdot \left[\frac{4}{7}(x+1)^{7/4} - \frac{4}{7}(0+1)^{7/4}\right]$$

$$F(x) = \frac{7}{4\cdot(2^{7/4}-1)} \cdot \frac{4}{7}\left[(x+1)^{7/4} - 1\right]$$

$$F(x) = \frac{(x+1)^{7/4}-1}{2^{7/4}-1}$$

- **Аналитически получим обратную функцию распределения $F^{-1}(y)$:**

$$y = \frac{(x+1)^{7/4}-1}{2^{7/4}-1}$$

$$y\cdot(2^{7/4}-1) = (x+1)^{7/4} - 1$$

$$(x+1)^{7/4} = y\cdot(2^{7/4}-1) + 1$$

$$x+1 = \left(y\cdot(2^{7/4}-1)+1\right)^{4/7}$$

$$x = \left(y\cdot(2^{7/4}-1)+1\right)^{4/7} - 1$$

- **Математическое ожидание E[X]:**

$$E[X] = \int_0^1 x \cdot p(x)\, dx$$

$$E[X] = \int_0^1 x \cdot \frac{7 \cdot (x+1)^{3/4}}{4 \cdot (2^{7/4} - 1)}\, dx$$

$$E[X] = \frac{7}{4 \cdot (2^{7/4} - 1)} \int_0^1 x \cdot (x+1)^{3/4}\, dx$$

Для вычисления интеграла сделаем замену u = x + 1, тогда du = dx, и x = u - 1:

$$\int (u-1) \cdot u^{3/4}\, du = \int u^{7/4}\, du - \int u^{3/4}\, du = \frac{4}{11} u^{11/4} - \frac{4}{7} u^{7/4}$$

Подставим пределы u = 1 и u = 2:

$$\left[\frac{4}{11}(2)^{11/4} - \frac{4}{7}(2)^{7/4}\right] - \left[\frac{4}{11}(1)^{11/4} - \frac{4}{7}(1)^{7/4}\right]$$

Упростим:

$$E[X] = \frac{7}{4 \cdot (2^{7/4} - 1)} \cdot \left(\frac{4}{11} \cdot 2^{11/4} - \frac{4}{7} \cdot 2^{7/4} - \frac{4}{11} + \frac{4}{7}\right)$$

Вычислим значение:

$$E[X] \approx 0.6$$

- **Дисперсия Var(X):**

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

Вычислим второй момент $E[X^2]$:

$$E[X^2] = \int_0^1 x^2 \cdot p(x)\, dx = \int_0^1 x^2 \cdot \frac{7 \cdot (x+1)^{3/4}}{4 \cdot (2^{7/4} - 1)}\, dx$$

Упростим:

$$E[X^2] = \frac{7}{4 \cdot (2^{7/4} - 1)} \int_0^1 x^2 \cdot (x+1)^{3/4}\, dx$$

Вычислим интеграл:

$$\int x^2 \cdot (x+1)^{3/4}\, dx$$

Для вычисления интеграла сделаем замену u = x + 1, тогда du = dx, и x = u - 1:

$$\int (u-1)^2 \cdot u^{3/4}\, du = \int (u^2 - 2u + 1) \cdot u^{3/4}\, du = \int u^{11/4}\, du - 2\int u^{7/4}\, du + \int u^{3/4}\, du$$

Вычислим каждый интеграл:

$$\int u^{11/4}\, du = \frac{4}{15}u^{15/4}, \quad \int u^{7/4}\, du = \frac{4}{11}u^{11/4}, \quad \int u^{3/4}\, du = \frac{4}{7}u^{7/4}$$

Подставим пределы u = 1 и u = 2:

$$E[X^2] = \frac{7}{4 \cdot (2^{7/4} - 1)} \cdot \left(\frac{4}{15} \cdot 2^{15/4} - \frac{8}{11} \cdot 2^{11/4} + \frac{4}{7} \cdot 2^{7/4} - \frac{4}{15} + \frac{8}{11} - \frac{4}{7} \right)$$

Вычислим значение:

$$E[X^2] \approx 0.45$$

Теперь дисперсия Var(X):

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

$$\mathrm{Var}(X) \approx 0.09$$

<u>3. Результаты выполнения задачи:</u>

В ходе выполнения лабораторной работы был реализован датчик непрерывной случайной величины с заданной плотностью вероятности, используя метод обратной функции. Программа высчитывала математическое ожидание и дисперсию, варьируя размерность выборки от 10 до 1,000,000. Для различных размеров выборки были получены следующие результаты:

```
Размер выборки: 10
Теоретическое матожидание: 0.541964
Выборочное среднее: 0.642854
Теоретическая дисперсия: 0.081335
Выборочная дисперсия: 0.081514

Размер выборки: 100
Теоретическое матожидание: 0.541964
Выборочное среднее: 0.529408
Теоретическая дисперсия: 0.081335
Выборочная дисперсия: 0.089772

Размер выборки: 1000
Теоретическое матожидание: 0.541964
Выборочное среднее: 0.546873
Теоретическая дисперсия: 0.081335
Выборочная дисперсия: 0.082828

Размер выборки: 10000
Теоретическое матожидание: 0.541964
Выборочное среднее: 0.536414
Теоретическая дисперсия: 0.081335
Выборочная дисперсия: 0.080987

Размер выборки: 100000
Теоретическое матожидание: 0.541964
Выборочное среднее: 0.541720
Теоретическая дисперсия: 0.081335
Выборочная дисперсия: 0.081143

Размер выборки: 1000000
Теоретическое матожидание: 0.541964
Выборочное среднее: 0.542519
Теоретическая дисперсия: 0.081335
Выборочная дисперсия: 0.081310
```

Удобное представление в виде таблицы:

Размер выборки	Теоретическое матожидание	Выборочное среднее	Теоретическая дисперсия	Выборочная дисперсия
10	0.541964	0.642854	0.081335	0.081514
100	0.541964	0.529408	0.081335	0.089772
1000	0.541964	0.546873	0.081335	0.082828
10000	0.541964	0.536414	0.081335	0.080987
100000	0.541964	0.541720	0.081335	0.081143
1000000	0.541964	0.542519	0.081335	0.081310

Анализ результатов:

- Выборочное среднее: с увеличением размера выборки выборочное среднее стремится к теоретическому значению математического ожидания. Для выборки из 1,000,000 элементов выборочное среднее составило 0.542519, что очень близко к теоретическому значению 0.541964
- Выборочная дисперсия: аналогично выборочному среднему, выборочная дисперсия с увеличением размера выборки приближается к теоретическому значению дисперсии. Для выборки 1,000,000 элементов выборочная дисперсия

составила 0.081310, что практически совпадает с теоретическим значением 0.081335

4. Вывод:

Результаты лабораторной работы подтверждают корректность реализации датчика непрерывной случайной величины и аналитических значений. С увеличением размера выборки выборочные значения математического ожидания и дисперсии всё больше приближаются к теоретическим значениям, что свидетельствует о правильности проведённых вычислений и реализации метода обратной функции.

Лабораторная работа № 3.2. Датчик нормальной случайной величины

1. Постановка задачи:

- Требуется реализовать датчик нормальной случайной величины для метода суммирования в виде функции со следующими параметрами: α - математическое ожидание, σ - среднеквадратическое отклонение.
- Получить с помощью датчика выборку, построить гистограмму с задаваемым шагом h гистограммы.
- По выборке проверить экспериментальное выполнение правила “трёх сигм”.

Примечание 3.2.1: метод суммирования - это способ генерации нормальной случайной величины, основанный на центральной предельной теореме.

Примечание 3.2.2: центральная предельная теорема (ЦПТ) - это утверждение, что сумма большого числа независимых случайных величин, распределённых по любому распределению, приближается к нормальному распределению.

Примечание 3.2.3: правило “трёх сигм” - это эмпирическое правило, которое описывает, как данные распределяются в нормальном распределении:

- 1 сигма (σ): примерно 68.27% значений нормального распределения находятся в пределах одного стандартного отклонения от среднего значения.
- 2 сигмы (2σ): примерно 95.45% значений находятся в пределах двух стандартных отклонений от среднего.
- 3 сигмы (3σ): примерно 99.73% значений находятся в пределах трёх стандартных отклонений от среднего.

2. Математическая составляющая задачи:

★ **Функция для генерации нормальной случайной величины методом суммирования:**

- Суммируем n равномерно распределённых случайных величин U_i, где:
  $$U_i \sim \text{Uniform}(0,1).$$
- Преобразование суммы равномерно распределённых случайных величин в стандартную нормальную случайную величину Z с использованием центральной предельной теоремы:
  $$Z = \frac{\sum_{i=1}^{n} U_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$
- Масштабирование и сдвиг стандартной нормальной величины Z для получения нормальной случайной величины с математическим ожиданием α и среднеквадратичным отклонением σ:
  $X = a + \sigma \cdot Z$, где $Z \sim N(0,1)$, а $X \sim N(a, \sigma^2)$

★ **Функция для генерации выборки нормальных случайных величин:**

- Генерация выборки нормальных случайных величин:

$\text{sample} = [X_1, X_2, \dots, X_{\text{size}}]$, где
$X_i \sim N(a, \sigma^2)$

★ **Функция для построения гистограммы:**
- Определение границ:

$$\text{bins} = [x_{\min}, x_{\min} + h, x_{\min} + \ 2h, \dots, x_{\max} + h]$$

- Определяем массив значения для построения теоретической кривой нормального распределения:

$$x = [a - 4\sigma, a - 4\sigma + \Delta, \dots, a + \ 4\sigma]$$

- Вычисление теоретической плотности нормального распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

★ **Функция для проверки правила “трёх сигм”:**
- Подсчёт количества элементов, попадающих в интервалы:

$[a - \sigma, a + \ \sigma]$, $[a - 2\sigma, a + 2\sigma]$, и $[a - 3\sigma, a + 3\sigma]$.

$$\text{withinSigma1} = \textstyle\sum_{i=1}^{\text{size}} \mathbb{I}(a - \sigma \leq \ X_i \leq a + \sigma)$$
$$\text{withinSigma2} = \textstyle\sum_{i=1}^{\text{size}} \mathbb{I}(a - 2\sigma \leq \ X_i \leq a + 2\sigma)$$
$$\text{withinSigma3} = \textstyle\sum_{i=1}^{\text{size}} \mathbb{I}(a - 3\sigma \leq \ X_i \leq a + 3\sigma)$$

- Определение долей (пропорций) элементов в выборке, которые попадают в соответствующие интервалы, заданные правилом “трёх сигм”:

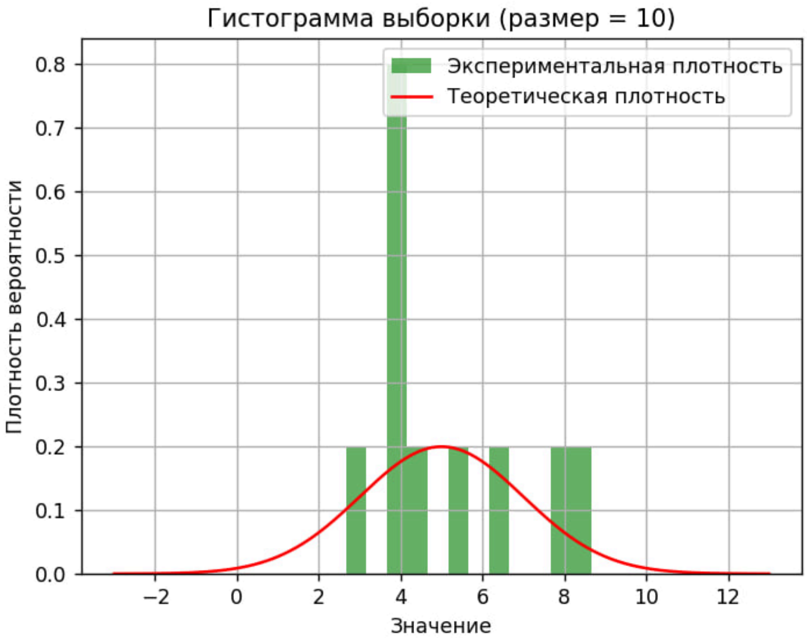
$$\text{proportionSigma1} = \frac{\text{withinSigma1}}{\text{size}}$$
$$\text{proportionSigma2} = \frac{\text{withinSigma2}}{\text{size}}$$
$$\text{proportionSigma3} = \frac{\text{withinSigma3}}{\text{size}}$$

3. Результаты выполнения задачи:

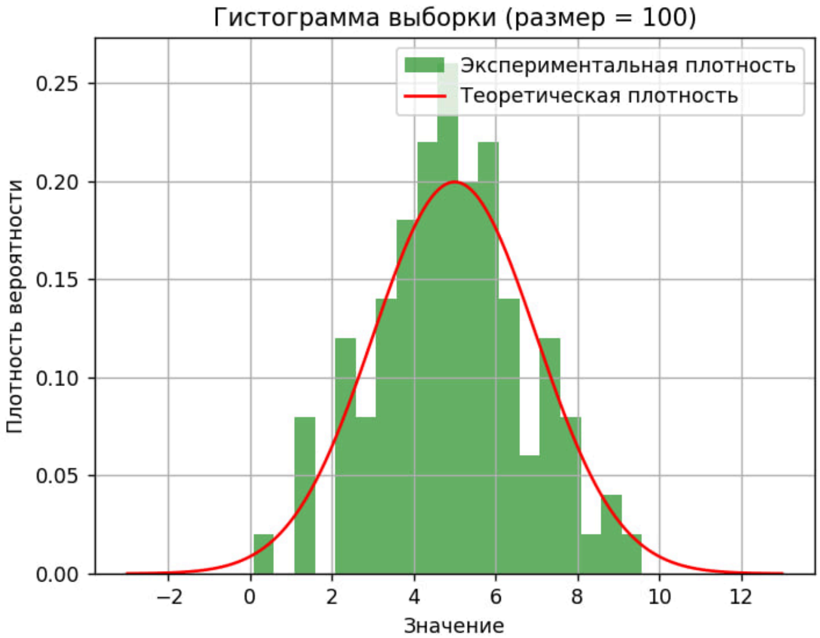
В результате выполнения лабораторной работы было проверено выполнение закона “трёх сигм” для нормального распределения, используя метод суммирования для генерации случайных величин. Программа использовала выборку, размер которой варьировался от 10 1,000,000. Для различных размеров выборки были получены следующие результаты:

- Малые размеры выборки (10, 100):
  небольшие выборки показывают, что доли значительно отклоняются от теоретических значений, особенно в пределах одной сигмы:

```
Размер выборки: 10
Доля элементов в пределах 1 сигмы: 0.7000 (ожидается ~0.6827)
Доля элементов в пределах 2 сигм: 1.0000 (ожидается ~0.9545)
Доля элементов в пределах 3 сигм: 1.0000 (ожидается ~0.9973)
```

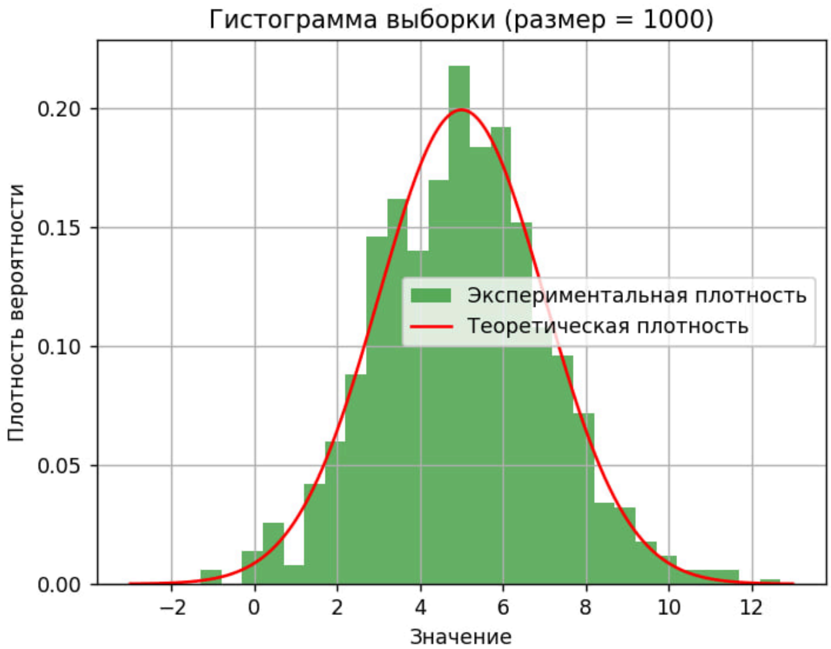



```
Размер выборки: 100
Доля элементов в пределах 1 сигмы: 0.7100 (ожидается ~0.6827)
Доля элементов в пределах 2 сигм: 0.9800 (ожидается ~0.9545)
Доля элементов в пределах 3 сигм: 1.0000 (ожидается ~0.9973)
```



- Средние размеры выборки (1000, 10000):
  средние выборки показывают, что доли становятся ближе к теоретическим значениям, особенно в пределах двух сигм:

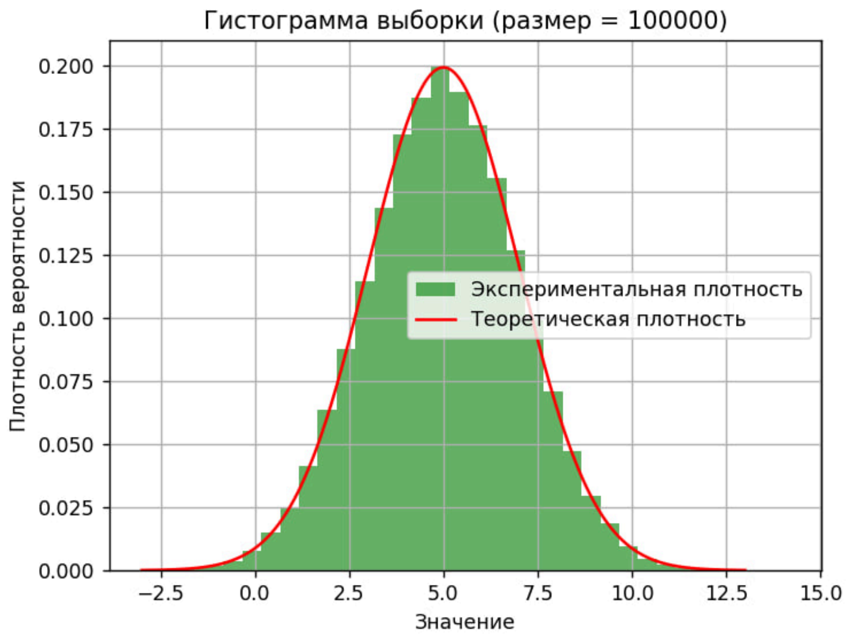
```
Размер выборки: 1000
Доля элементов в пределах 1 сигмы: 0.6720 (ожидается ~0.6827)
Доля элементов в пределах 2 сигм: 0.9460 (ожидается ~0.9545)
Доля элементов в пределах 3 сигм: 0.9910 (ожидается ~0.9973)
```



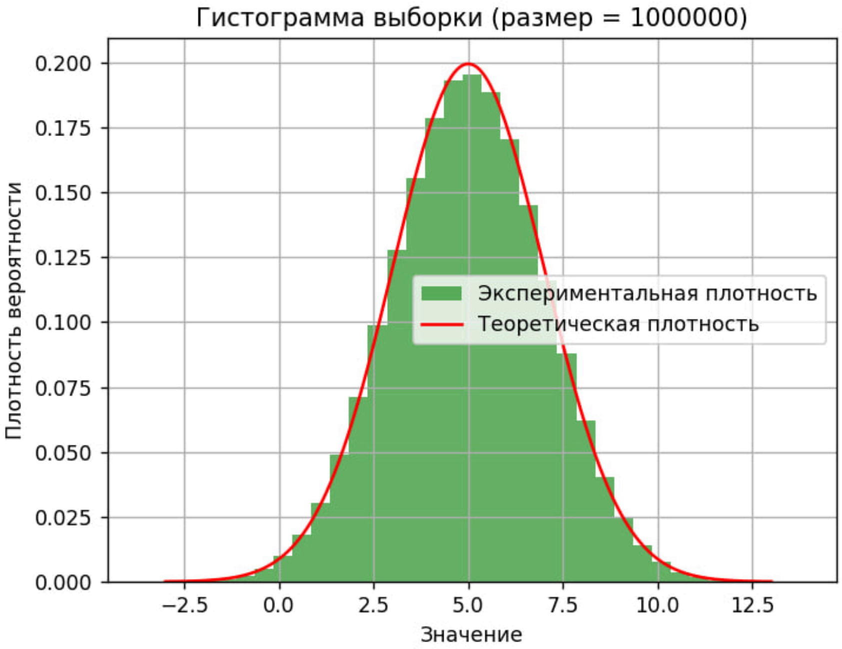

```
Размер выборки: 10000
Доля элементов в пределах 1 сигмы: 0.6641 (ожидается ~0.6827)
Доля элементов в пределах 2 сигм: 0.9528 (ожидается ~0.9545)
Доля элементов в пределах 3 сигм: 0.9981 (ожидается ~0.9973)
```

- Большие размеры выборки (100,000 и более):
  большие выборки показывают, что доли практически идеально совпадают с теоретическими значениями:

```
Размер выборки: 100000
Доля элементов в пределах 1 сигмы: 0.6784 (ожидается ~0.6827)
Доля элементов в пределах 2 сигм: 0.9549 (ожидается ~0.9545)
Доля элементов в пределах 3 сигм: 0.9978 (ожидается ~0.9973)
```




```
Размер выборки: 1000000
Доля элементов в пределах 1 сигмы: 0.6787 (ожидается ~0.6827)
Доля элементов в пределах 2 сигм: 0.9556 (ожидается ~0.9545)
Доля элементов в пределах 3 сигм: 0.9980 (ожидается ~0.9973)
```



<u>4. Вывод:</u>
На основании проведённых тестов можно заключить, что с увеличением размера выборки экспериментальные доли элементов, попадающих в заданные интервалы, всё ближе подходят к теоретическим значениям, что подтверждает выполнение правила "трёх сигм" для нормального распределения. Отклонения в малых выборках объясняются статистической случайностью.