



МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М. В. ЛОМОНОСОВА
Факультет вычислительной математики и кибернетики
Кафедра системного анализа

Отчёт по практикуму

«Стохастический анализ и моделирование»

Студент 315 группы

К. Е. Летуновский

Руководитель практикума

д. ф.-м. н. С. Н. Смирнов

Москва, 2024

1 Задача 1

1. Реализовать генератор схемы Бернулли с заданной вероятностью успеха p . На основе генератора схемы Бернулли построить датчик биномиального распределения.
2. Реализовать генератор геометрического распределения; проверить для данного распределения свойство отсутствия памяти.
3. Промоделировать игру в орлянку: бесконечную последовательность независимых испытаний Бернулли с бросанием “правильной” (честной, $p = 0.5$) монеты. Величина “выигрыша” S_n определяется как сумма по n испытаниям значений 1 и -1 в зависимости от выпавшей стороны монеты. Проиллюстрировать в виде ломаной поведение нормированной суммы

$$Y(i) = \frac{S_i}{\sqrt{n}}$$

как функцию от номера испытания i для отдельно взятой траектории. Дать теоретическую оценку для значения $Y(n)$ при $n \rightarrow \infty$.

1.1 Теоретическая часть

Определение. *Схемой Бернулли с заданной вероятностью p называется эксперимент, состоящий из серии испытаний, удовлетворяющих следующим условиям:*

1. *Отсутствие взаимного влияния в испытаниях;*
2. *Воспроизводимость. Однородные испытания проводятся в сходных условиях;*
3. *Существует признак, который реализуется или не реализуется в испытании. Признак может быть отнесён к любому из испытаний.*

Определение. *Биномиальное распределение описывает число успехов в n независимых испытаниях схемы Бернулли. Вероятность получения k успехов определяется формулой:*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Замечание Среднее значение биномиального распределения равно np , а дисперсия равна $np(1-p)$.

Определение. Пусть имеется последовательность независимых испытаний Бернулли с вероятностью успеха p . Случайная величина X , которая принимает значение количества испытаний до первого успеха, имеет геометрическое распределение с параметром p . Формула вероятности имеет вид:

$$P(X = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots$$

Геометрическое распределение можно моделировать с помощью схемы Бернулли, но придется иметь дело с векторами случайной длины, что вызывает затруднения. Поэтому воспользуемся следующим фактом.

Пусть $\xi \sim \exp(k), k > 0$. Тогда $\lfloor \xi \rfloor \sim \text{Geom}(p), p = 1 - e^{-k}, k = -\ln(1 - p)$. Функция распределения ξ равна $F(x) = (1 - e^{-kx})\mathbb{I}(x \geq 0)$. Тогда $(\lfloor \xi \rfloor = n) \Leftrightarrow (n \leq \xi < n + 1)$. $F(n + 1) - F(n) = e^{-kn}(1 - e^{-k}) = (1 - p)^np$. Тогда нужно сгенерировать случайную величину, имеющую экспоненциальное распределение. $F_\xi^{-1}(y) = -\frac{1}{k} \ln(1 - y)$. Так как $F_\xi(y)$ непрерывна и строго возрастает, то $\xi \sim F_\xi^{-1}(U), U \sim U(0, 1)$. $1 - U$ имеет равномерное распределение на отрезке $[0, 1]$. Таким образом, формула для генерации геометрического распределения:

$$X = \frac{\ln(U)}{\ln(1 - p)}.$$

Утверждение. Геометрическое распределение обладает свойством отсутствия памяти.

Доказательство. Воспользуемся тем, что $\mathbb{P}(X \leq k) = 1 - (1 - p)^k, \forall k$. Тогда $\forall k, m \geq 1$ выполнено

$$\mathbb{P}(X > k + m | X > m) = \frac{\mathbb{P}(X > k + m, X > m)}{\mathbb{P}(X > m)} = \frac{(1 - p)^{k+m-1}}{(1 - p)^{m-1}} = (1 - p)^{k-1}$$

□

Игра в «орлянку». Вначале приведем формулировку центральной предельной теоремы.

Теорема 1 (Центральная предельная теорема). При $n \rightarrow \infty$, нормированная сумма независимых одинаково распределенных случайных величин $Y(n)$ стремится по распределению к нормальному распределению с математическим ожиданием 0 и дис-

персией 1:

$$Y(n) = \frac{\sum_{k=1}^n \xi_k}{\sqrt{n}} \xrightarrow{d} N(0, 1).$$

Найдем теперь теоретическую оценку. Имеем последовательность независимых случайных величин $\xi_k, k \in \mathbb{N}$, имеющих модифицированное (т.е. вместо точки 0 ненулевую вероятность имеет точка -1) распределение Бернулли с параметром $p = 0.5$. Математическое ожидание $\mathbb{E}\xi_k = 0$, дисперсия $\mathbb{D}\xi_k = 1$. $S_n = \sum_{k=1}^n \xi_k$, математическое ожидание суммы равно 0, дисперсия равна сумме дисперсий (в силу независимости случайных величин) и равна n . Тогда по центральной предельной теореме

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} < x\right) \rightarrow \mathcal{N}(0, 1), \quad n \rightarrow \infty.$$

1.2 Результаты работы программы

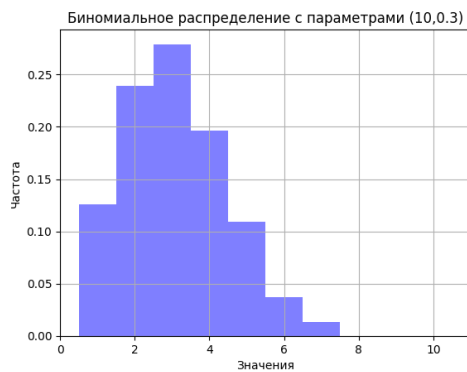


Рис. 1: Гистограмма биномиального распределения ($n=10, p=0.3$)

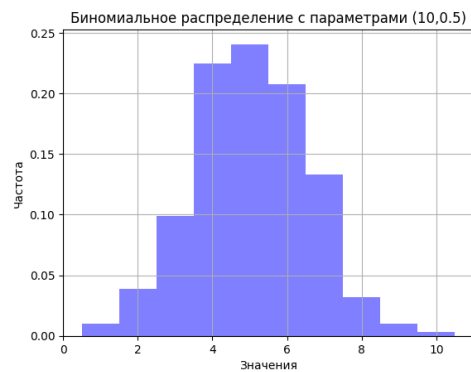


Рис. 2: Гистограмма биномиального распределения ($n=10, p=0.5$)

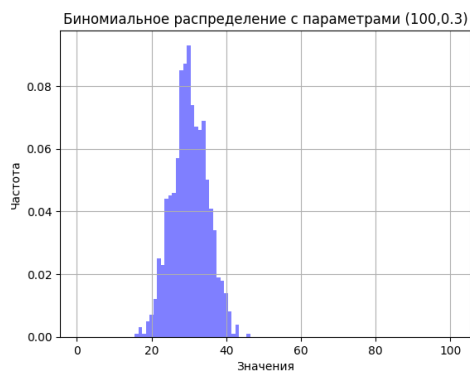


Рис. 3: Гистограмма биномиального распределения ($n=100$, $p=0.3$)

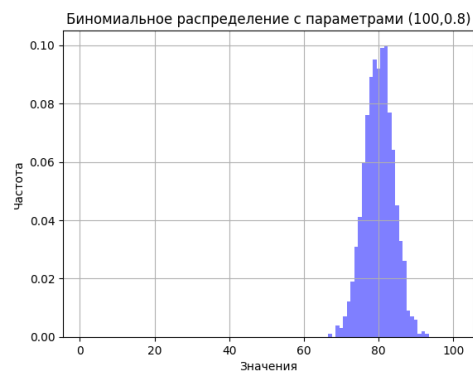


Рис. 4: Гистограмма биномиального распределения ($n=100$, $p=0.8$)

На этих картинках заметно, что в биномиальное распределение в основном сосредоточено в точке np , которая является математическим ожиданием случайной величины из этого распределения.

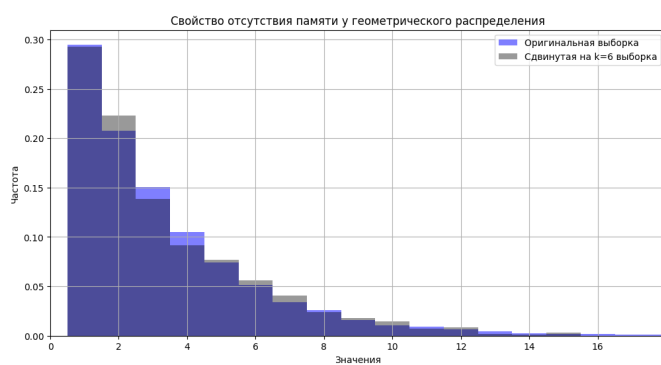


Рис. 5: Свойство отсутствия памяти у геометрического распределения с параметром $p = 0.3$

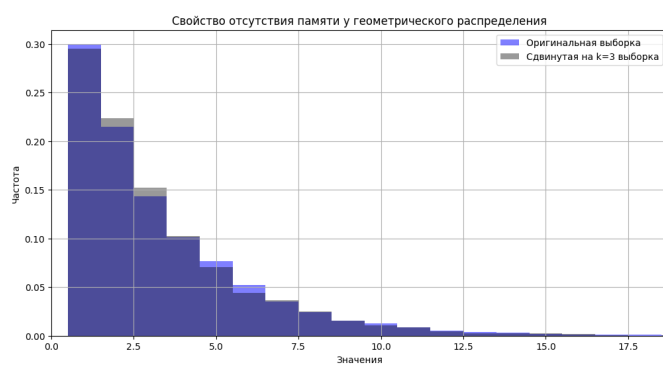


Рис. 6: Свойство отсутствия памяти у геометрического распределения с параметром $p = 0.3$

Заметно, что распределения на картинках совпадают.



Рис. 7: Нормированные суммы в игре «орлянка»

2 Задача 2

1. Построить датчик сингулярного распределения, имеющий в качестве функции распределения канторову лестницу. С помощью критерия Колмогорова убедиться в корректности работы датчика.
2. Для канторовых случайных величин с помощью критерия Смирнова проверить:
 - свойство симметричности относительно $\frac{1}{2}$, то есть случайные величины X и $1 - X$ распределены одинаково;
 - свойство самоподобия относительно деления на 3, то есть условное распределение Y при условии $Y \in [0, 1/3]$ совпадает с распределением $Y/3$.
3. Рассчитать значения математического ожидания и дисперсии для данного распределения. Сравнить теоретические значения с эмпирическими (для различных объёмов выборок) и проиллюстрировать сходимость эмпирических значений к теоретическим.

2.1 Теоретическая часть

Определение. *Сингулярное распределение Кантора определяется на множестве Кантора и имеет в качестве функции распределения канторову лестницу $C(x)$. Функция $C(x)$ обладает следующими свойствами:*

- $C(x)$ непрерывна, но не дифференцируема;
- Производная $C'(x)$ равна нулю почти всюду;
- Функция $C(x)$ является самоподобной: разделение на равные интервалы сохраняет форму функции.

Это сингулярное распределение можно моделировать через разложение числа в троичной системе счисления. В канторово множество попадают все числа, не имеющие единиц в троичной записи \implies имеем модифицированное распределение Бернулли (имеет ненулевую вероятность в точках 0 и 2) с $p = 0.5$. Генерировать бесконечную последовательность мы не можем, поэтому необходимо ввести погрешность вычислений ε . Хотим, чтобы хвост ряда $\sum_{n=1}^{\infty} \frac{2\xi_n}{3^n}$ был меньше этой погрешности, т.е.

$$\sum_{n=N+1}^{\infty} \frac{2\xi_n}{3^n} \leq \sum_{n=N+1}^{\infty} \frac{2}{3^n} = \frac{2}{3^{N+1} - \frac{1}{3}} = \frac{1}{3^N} \leq \varepsilon, \text{ откуда } N \geq \log_3\left(\frac{1}{\varepsilon}\right).$$

Тогда значение случайной величины равно $x = \sum_{n=1}^{N(\varepsilon)} \frac{2\xi_n}{3^n}$, а функция распределения в этой точке равна $F(x) = \sum_{n=1}^{N(\varepsilon)} \frac{\xi_n}{2^n}$.

Критерий Колмогорова используется для проверки гипотезы о соответствии эмпирической функции распределения теоретической. Статистика критерия $D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|$, где $F_n(x)$ — выборочная функция распределения, $F_0(x)$ — канторова лестница. Для вычисления p -value используется функция Колмогорова $F_k(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x}$. Тогда $p = 1 - F_k(\sqrt{n} D_n)$. Если $p > \alpha$, то гипотеза $H_0 : F_n(x) = F_0(x)$ принимается, иначе — отвергается (здесь α — заданный уровень значимости).

Критерий Смирнова используется для проверки гипотезы о том, что две выборки были сгенерированы из одного распределения. Рассмотрим $X = (X_1, X_2, \dots, X_n), Y = (Y_1, Y_2, \dots, Y_m)$ — выборки, элементы которых имеют распределения F и G соответственно, F_n и G_m — построенные по ним эмпирические функции распределения. Рассматривается следующая статистика: $D_{nm} = \sup_{x \in \mathbb{R}} |F_n(x) - G_m(x)|$. Далее аналогично критерию Колмогорова рассматриваем значение функции Колмогорова в точке

$\sqrt{\frac{mn}{m+n}} D_{nm}$ и соответствующее ему p -value. Если $p > \alpha$, то гипотеза $H_0 : F(x) = G(x)$ принимается, иначе — отвергается (здесь α — заданный уровень значимости).

Найдем математическое ожидание и дисперсию этой случайной величины. Воспользуемся независимостью ξ_n .

$$\mathbb{E}X = \mathbb{E} \sum_{n=1}^{+\infty} \frac{\xi_n}{3^n} = \sum_{n=1}^{+\infty} \frac{1}{3^n} \mathbb{E}\xi_n = \sum_{n=1}^{+\infty} \frac{1}{3^n} (0 \cdot 0.5 + 2 \cdot 0.5) = \frac{1}{2},$$

$$\text{var}(X) = \text{var} \left(\sum_{n=1}^{+\infty} \frac{\xi_n}{3^n} \right) = \sum_{n=1}^{+\infty} \frac{1}{3^{2n}} \cdot \text{var}(\xi_n),$$

$$\text{var}(\xi_n) = \mathbb{E}\xi_n^2 - (\mathbb{E}\xi_n)^2 = (0 \cdot 0.5 + 4 \cdot 0.5) - (0 \cdot 0.5 + 2 \cdot 0.5)^2 = 1,$$

$$\text{var}(X) = \sum_{n=1}^{+\infty} \frac{1}{3^{2n}} = \frac{1}{8}.$$

2.2 Результат работы программы

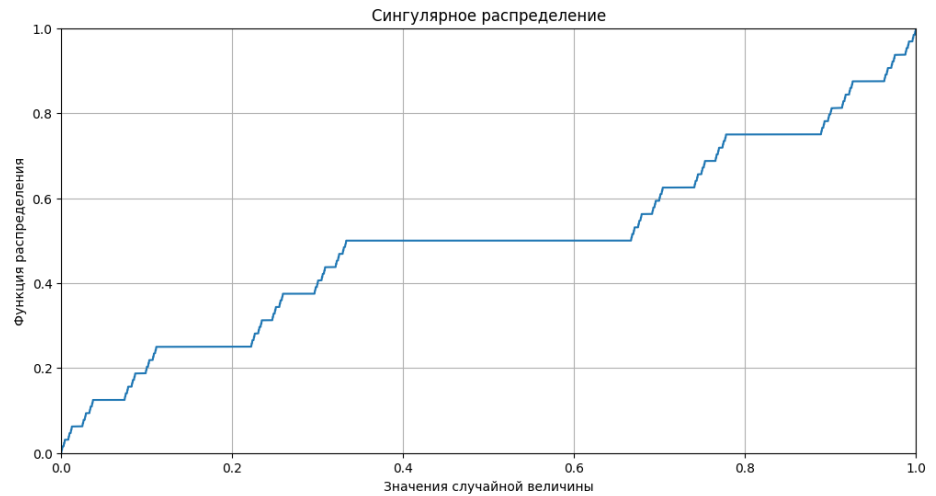


Рис. 8: Канторова лестница в качестве функции распределения случайной величины

На этой картинке показан результат работы датчика. Как видим, функция распределения очень похожа на Канторову лестницу.

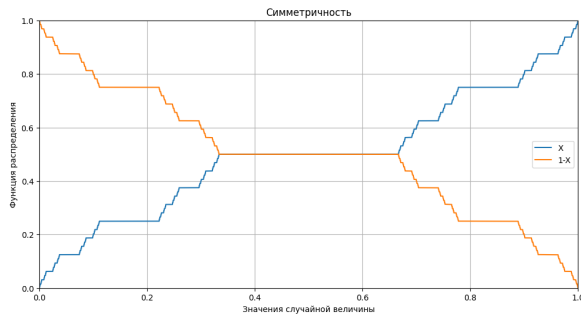


Рис. 9: Симметричность распределения

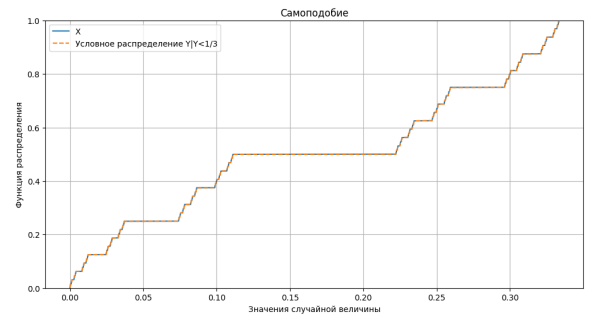


Рис. 10: Свойство самоподобия относительно деления на 3 распределения

На этих картинках свойство симметричности и самоподобия относительно деления на 3 распределения видно «невооруженным глазом».



Рис. 11: Сходимость выборочных средних

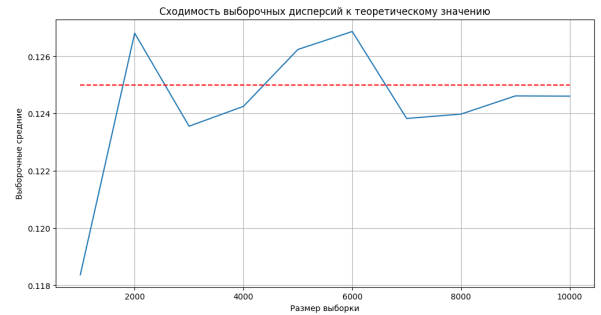


Рис. 12: Сходимость выборочных дисперсий

Здесь показана сходимость выборочных средних и выборочных дисперсий к их теоретическому значению.

3 Задача 3

1. Построить датчик экспоненциального распределения. Проверить для данного распределения свойство отсутствия памяти.
2. Пусть X_1, \dots, X_n — независимые экспоненциально распределённые случайные величины с параметрами $\lambda_1, \dots, \lambda_n$. Найти распределение случайной величины $Y = \min(X_1, \dots, X_n)$.
3. На основе датчика экспоненциального распределения построить датчик пуассоновского распределения.

4. Построить датчик пуассоновского распределения как предел биномиального распределения. Убедиться в корректности построенного датчика при помощи критерия χ^2 Пирсона.
5. Построить датчик стандартного нормального распределения методом моделирования случайных величин парами с переходом в полярные координаты (преобразование Бокса-Мюллера). Проверить при помощи t -критерия Стьюдента равенство математических ожиданий, а при помощи критерия Фишера — равенство дисперсий.

3.1 Теоретическая часть

Датчик экспоненциального распределения уже был построен методом обращения функции распределения в задании 1. Повторим формулу для генерации случайной величины, имеющей экспоненциальное распределение с параметром k :

$$\xi \sim -\frac{1}{k} \ln(U), U \sim \mathbb{U}(0, 1)$$

Найдем распределение $Y = \min(X_1, \dots, X_n)$, $X_k \sim \exp(\lambda_k)$

$$\begin{aligned} \mathbb{P}(\min\{X_1, X_2, \dots, X_n\} < x) &= 1 - \mathbb{P}(X_1 \geq x, X_2 \geq x, \dots, X_n \geq x) = 1 - \prod_{k=1}^n \mathbb{P}(X_k \geq x) = \\ &= 1 - \prod_{k=1}^n e^{-\lambda_k x} = 1 - e^{-x \sum_{k=1}^n \lambda_k} \end{aligned}$$

Таким образом, $\min\{X_1, X_2, \dots, X_n\} \sim \exp(\sum_{k=1}^n \lambda_k)$

На основе экспоненциального распределения можно построить пуассоновское. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, имеющие экспоненциальное распределение с параметром λ . Пусть $X = \max\{n : \sum_{k=1}^n \xi_k \leq 1\}$. Тогда $X - 1$ имеет пуассоновское распределение с параметром λ . Данное свойство вытекает из определения процесса Пуассона с параметром интенсивности λ : $X(1)$ имеет пуассоновское распределение с параметром λ , а время между скачками имеет экспоненциальное распределение.

Другой подход к построению датчика пуассоновского распределения дает теорема Пуассона.

Теорема 2 (Пуассон). *Рассмотрим серию испытаний Бернулли с вероятностями*

успеха p_n . Пусть $p_n \xrightarrow{n \rightarrow \infty} 0, np_n \xrightarrow{n \rightarrow \infty} \lambda > 0$. Тогда число успехов в схеме Бернулли сходится по распределению к пуассоновскому распределению с параметром λ .

Доказательство можно найти в [1].

Таким образом, можно генерировать случайную величину, имеющую пуассоновское распределение, как предел биномиального (число успехов в схеме Бернулли).

Критерий χ^2 Пирсона. Считаем частоты n_k по выборке как количество величин, равных k , теоретические частоты равны $n_k^\omega = n \frac{e^{-\lambda} \lambda^k}{k!}$. Статистика $\chi^2 = \sum_{k \in \{a_1, a_2, \dots, a_m\}} \frac{(n_k^\omega - n_k)^2}{n_k^\omega}$, где a_1, a_2, \dots, a_m — уникальные значения в выборке. p -value равно $p = 1 - F_{\chi^2}(\chi^2, \nu)$, где F_{χ^2} — функция распределения χ^2 , $\nu = m - 1$ — число степеней свободы.

Рассмотрим (X_1, X_2) — случайный вектор, компоненты которого имеют стандартное нормальное распределение. Найдём распределение полярных координат. В силу симметрии угол φ имеет равномерное распределение на отрезке $[0, 2\pi]$. Квадрат радиуса $r^2 = X_1^2 + X_2^2$ по определению имеет распределение χ^2 с двумя степенями свободы, которое совпадает с экспоненциальным с параметром $\frac{1}{2}$. Тогда $X_1 = r \cos(\varphi) \sim N(0, 1)$, $X_2 = r \sin(\varphi) \sim N(0, 1)$.

Критерий Стьюдента. Рассматривается выборка X_1, X_2, \dots, X_n из нормального распределения с неизвестными параметрами. Гипотеза $H_0 : \mathbb{E}X = \mu$. Оценка стандартного отклонения $\sigma' = \sqrt{\frac{1}{n-1} \sum_i (X_i - \bar{X})^2}$. Тогда статистика $U = \sqrt{n} \frac{\bar{X} - \mu}{\sigma'}$ имеет распределение Стьюдента с $n - 1$ степенью свободы при условии истинности H_0 . Нулевая гипотеза принимается, если $|t| < q$, где q — квантиль распределения Стьюдента с $n - 1$ степенью свободы порядка $1 - \frac{\alpha}{2}$.

Критерий Фишера. Рассматриваются две выборки X_1, X_2, \dots, X_n и Y_1, Y_2, \dots, Y_m . Гипотеза $H_0 : \sigma_1 = \sigma_2$, т.е. равенство стандартных отклонений. Статистика $V = \frac{S_X^2}{S_Y^2}$, где S_X^2, S_Y^2 выборочные дисперсии X и Y соответственно. Она имеет распределение Фишера с параметрами $n - 1$ и $m - 1$. Будем отклонять гипотезу, если $V \leq q_1$ или $V \geq q_2$, где q_1, q_2 — квантили указанного распределения Фишера порядка $\frac{\alpha}{2}$ и $1 - \frac{\alpha}{2}$ соответственно.

3.2 Результат работы программы

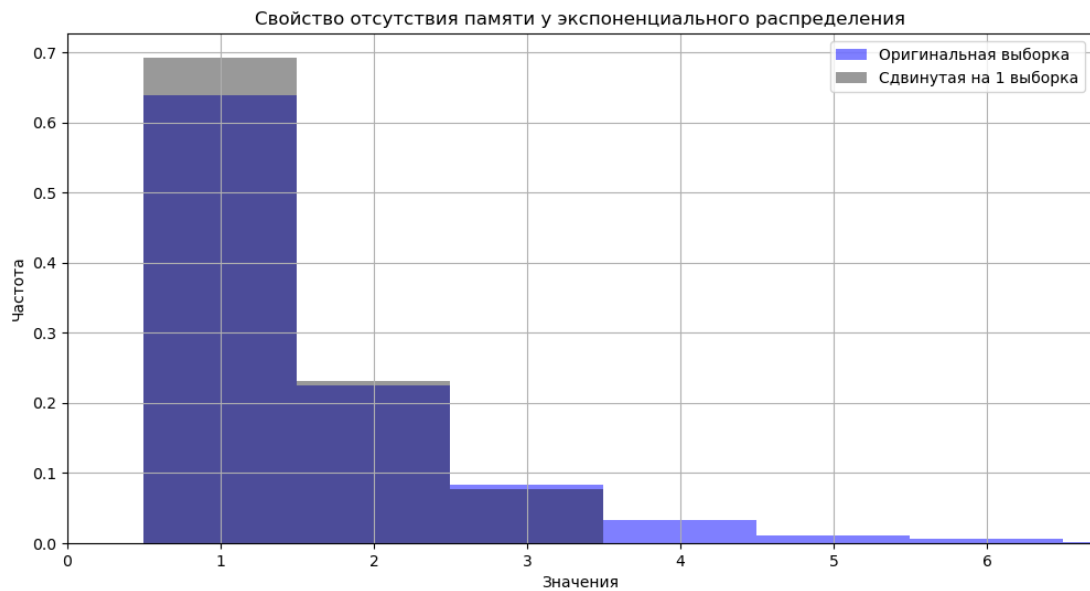


Рис. 13: Свойство отсутствия памяти для экспоненциального распределения

Данная картинка показывает свойство отсутствия памяти у экспоненциального распределения

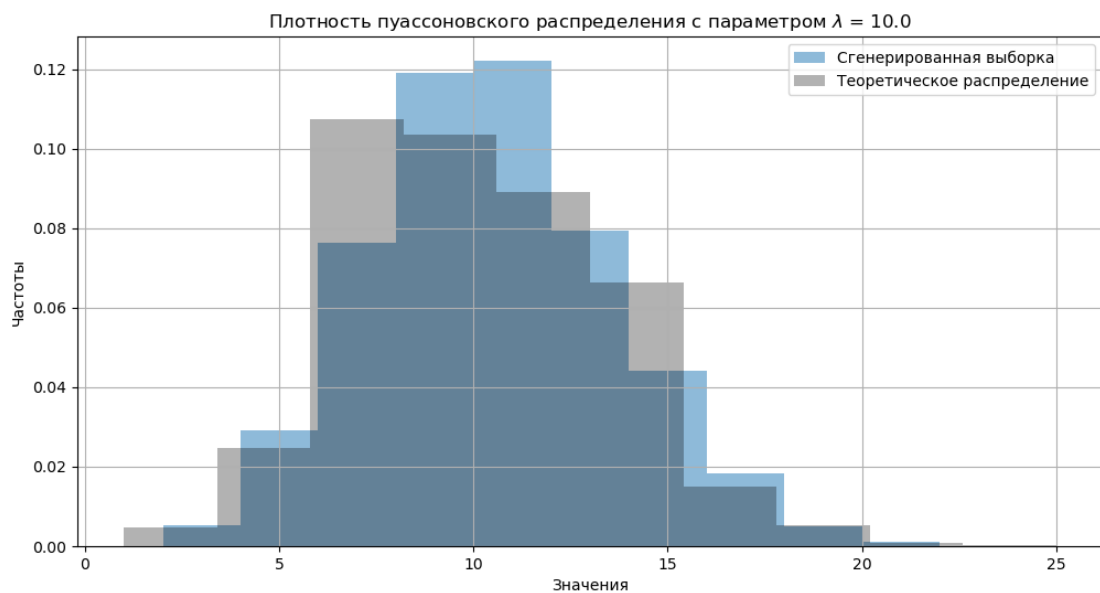


Рис. 14: Плотность распределения Пуассона, построенного при помощи датчика экспоненциального распределения

Построенное распределение совпадает с распределением Пуассона.

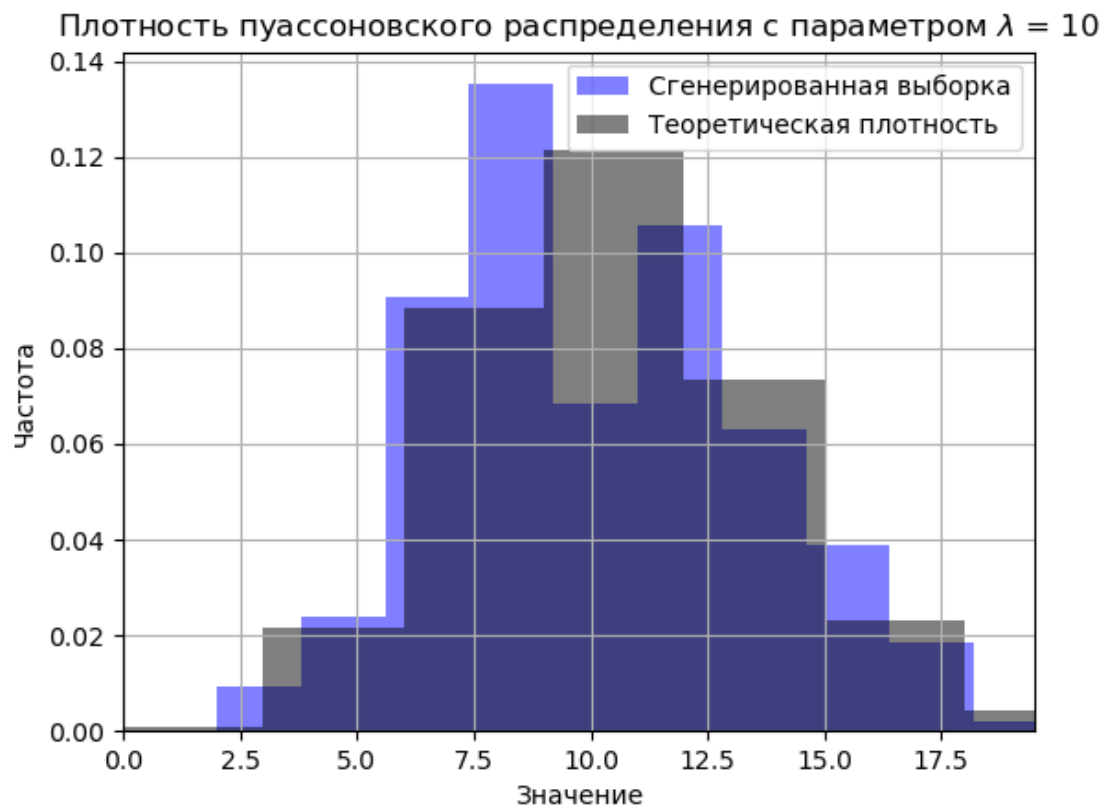


Рис. 15: Плотность распределения Пуассона, построенного как предел биномиального

Построенное распределение совпадает с распределением Пуассона.

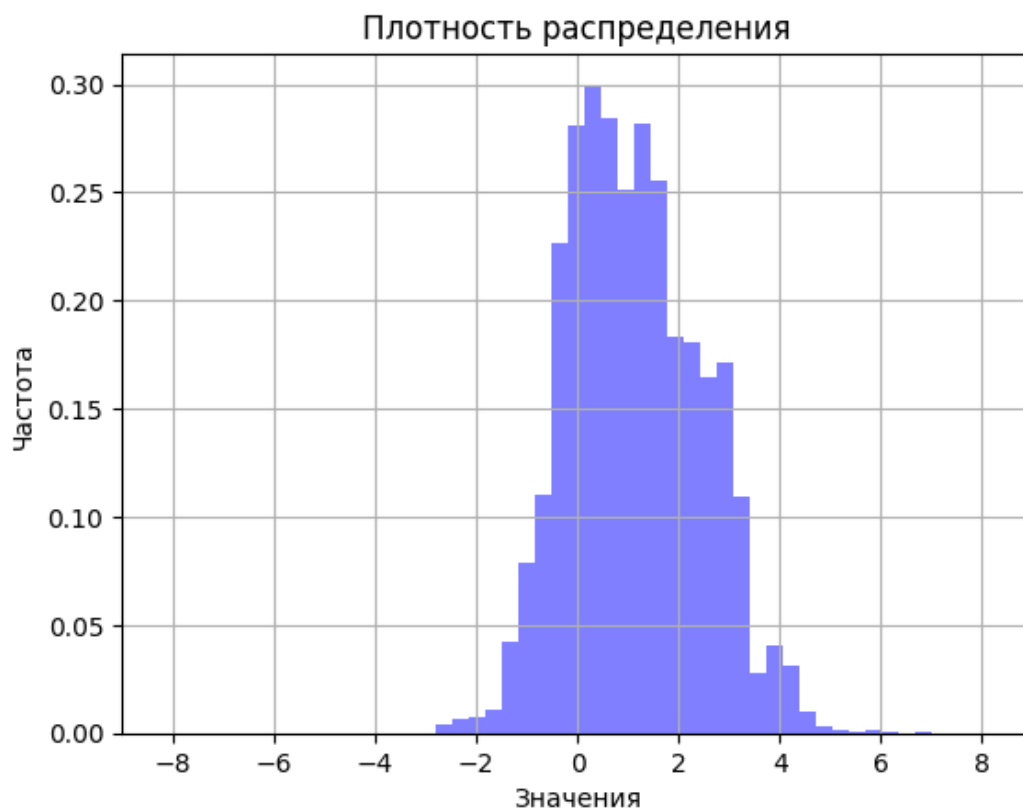


Рис. 16: Плотность нормального распределения с параметрами $(0,1)$, сгенерированного преобразованием Бокса-Мюллера

Данное распределение совпадает со стандартным нормальным.

4 Задача 4

1. Построить датчик распределения Коши.
2. На основе датчика распределения Коши с помощью метода фон Неймана построить датчик стандартного нормального распределения. При помощи графика normal probability plot убедиться в корректности построенного датчика и обосновать наблюдаемую линейную зависимость.

3. Сравнить скорость моделирования стандартного нормального распределения в заданиях 3 и 4.

4.1 Теоретическая часть

Функция распределения Коши равна $F(x) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$. Она строго монотонна и непрерывна \implies существует обратная к ней функция $F^{-1}(y) = \gamma \tan\left(\pi\left(y - \frac{1}{2}\right)\right) + x_0$. Тогда $F^{-1}(U)$ имеет распределение Коши, где U – равномерно распределенная на отрезке $[0, 1]$ случайная величина.

Пусть имеется датчик распределения Коши с плотностью $g(\cdot)$ и требуется смоделировать распределение с плотностью $f(\cdot)$ (стандартное нормальное), выполнено $\sup_{x \in \mathbb{R}} \frac{f(x)}{g(x)} \leq k$. Тогда используем следующий алгоритм:

1. Моделируем $X \sim g$ и $U \sim U(0, 1)$.
2. Если $\frac{f(x)}{g(x)} \geq kU$, то $Y = X$ и процесс завершается, иначе возврат к шагу 1.

Наилучшее приближение стандартного нормального распределения распределением Коши имеет параметры $x_0 = 0, \gamma = 1$. При этом $k = \sqrt{\frac{2\pi}{e}}$.

4.2 Результат работы программы

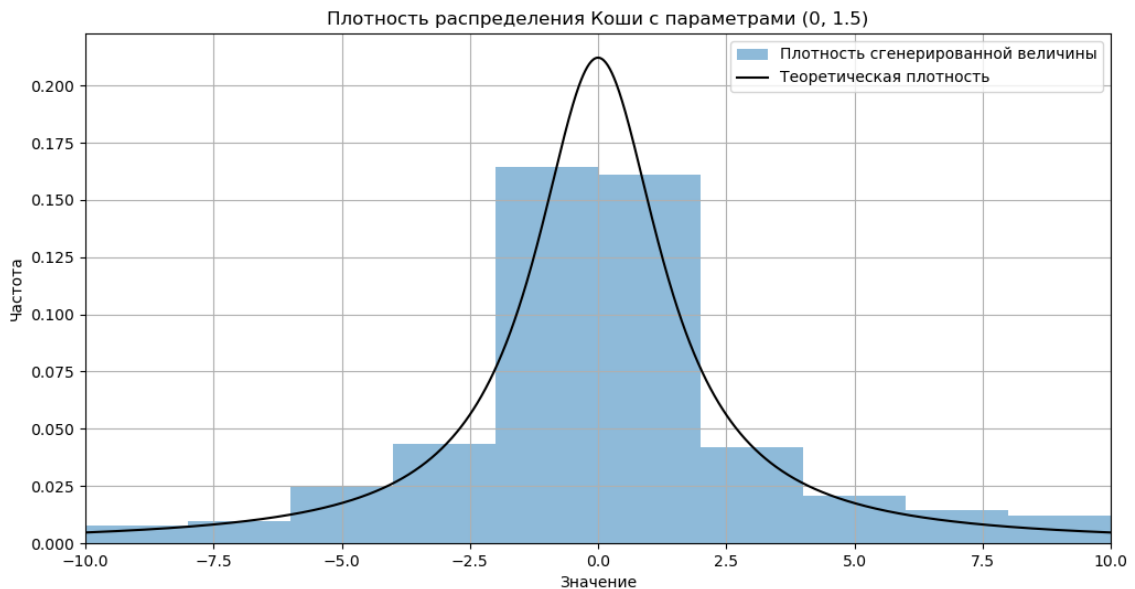


Рис. 17: Плотность распределения Коши

Normal probability plot — графический способ выяснить, принадлежит ли выборка нормальному распределению. Данные наносятся на график теоретического нормального распределения так, что если данная выборка принадлежит нормальному распределению, то точки образуют график линейной зависимости.

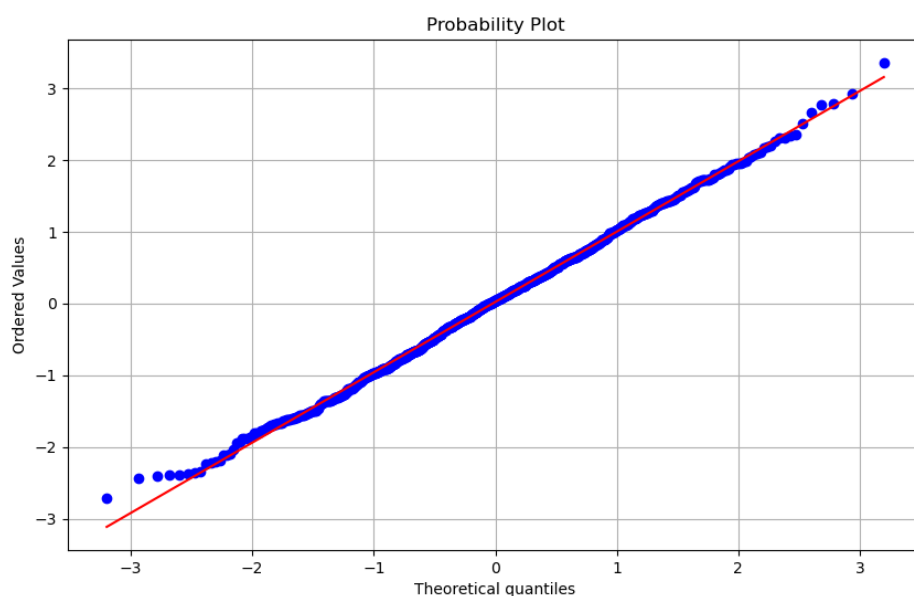


Рис. 18: График normal probability plot при $n = 1000$

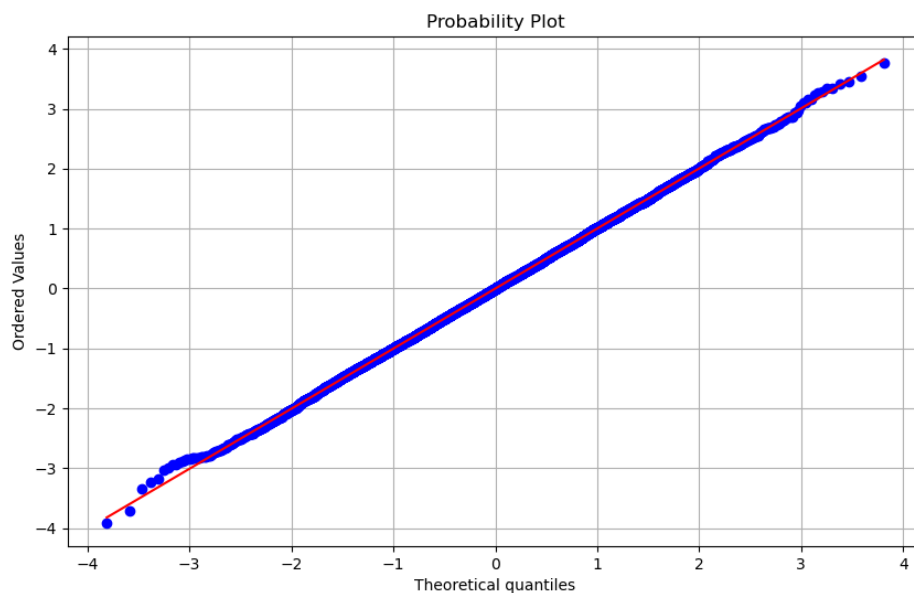


Рис. 19: График normal probability plot при $n = 10000$

Очевидно, линейная зависимость есть.

```
[162]: n = 1000
[164]: %timeit Gauss_B_M_gen(n)
       39.7 µs ± 1.22 µs per loop (mean ± std. dev. of 7 runs, 10,000 loops each)
[166]: %timeit von_Neumann_Gauss_gen(n)
       36.4 ms ± 1.18 ms per loop (mean ± std. dev. of 7 runs, 10 loops each)
[168]: n = 10000
[170]: %timeit Gauss_B_M_gen(n)
       232 µs ± 2.38 µs per loop (mean ± std. dev. of 7 runs, 1,000 loops each)
[172]: %timeit von_Neumann_Gauss_gen(n)
       357 ms ± 3.81 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
```

Рис. 20: Сравнение скорости работы при разных n

Заметно, что датчик, построенный при помощи преобразования Бокса-Мюллера, работает быстрее датчика на основе метода элиминации фон Неймана.

5 Задача 5

1. Пусть $X_i \sim N(\mu, \sigma^2)$. Убедиться эмпирически в справедливости теоремы о законе больших чисел (ЗБЧ) и центральной предельной теоремы (ЦПТ): исследовать поведение суммы $S_n = \sum_{i=1}^n X_i$ и эмпирического распределения величины:

$$\sqrt{n} \left(\frac{S_n}{n} - \mu \right).$$

2. Считая μ и σ^2 неизвестными, построить доверительные интервалы для среднего и дисперсии по имеющейся выборке.
3. Пусть $X_i \sim K(a, b)$ имеет распределение Коши с параметрами a и b . Изучить эмпирически, как ведут себя суммы S_n/n , объяснить результат и найти закон

распределения данных сумм.

5.1 Теоретическая часть

Запишем формулировки ЗБЧ и ЦПТ.

Теорема 3. ЗБЧ Пусть $X_1, X_2, \dots, X_n, \dots$ — последовательность независимых нормально распределенных случайных величин с параметрами μ и σ^2 , при этом $\mathbb{E}X_k^2 < \infty$. Рассмотрим $S_n = \sum_{k=1}^n X_k$. Справедливо следующее соотношение: $\forall \varepsilon > 0 \quad \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \rightarrow 1, n \rightarrow +\infty$.

Теорема 4. ЦПТ Пусть $X_1, X_2, \dots, X_n, \dots$ — последовательность независимых нормально распределенных случайных величин с параметрами μ и σ^2 , причем математическое ожидание конечно, а дисперсия ненулевая. Тогда $\frac{S_n - \mu n}{\sigma \sqrt{n}} \rightarrow \mathcal{N}(0, 1)$ (сходимость по распределению), $n \rightarrow +\infty$.

Доказательства этих теорем можно найти в [1] и в [2].

Доверительный интервал для математического ожидания строится согласно t-критерию Стьюдента.

Теорема 5. t-критерий Стьюдента Пусть $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Тогда статистика

$$t = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

имеет распределение Стьюдента с $(n - 1)$ степенью свободы.

Доказательство можно найти в [2].

Пусть $q(\alpha, n - 1)$ — квантили порядка α распределения Стьюдента с $n - 1$ степенями свободы. Тогда доверительный интервал для среднего $\left[\bar{X} - q\left(1 - \frac{\alpha}{2}, n - 1\right) \frac{S}{\sqrt{n}}, \bar{X} + q\left(1 - \frac{\alpha}{2}, n - 1\right) \frac{S}{\sqrt{n}}\right]$, где α — заданный уровень значимости.

Доверительный интервал для стандартного отклонения строится с помощью теоремы Фишера.

Теорема 6. Фишера Пусть $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Тогда статистика $H = \frac{(n - 1)S^2}{\sigma^2}$ имеет распределение χ^2 с $n - 1$ степенью свободы.

Тогда верно следующее равенство:

$$\mathbb{P} \left(\chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right) = 1 - \alpha,$$

где $\chi_{n-1, \frac{\alpha}{2}}^2$ и $\chi_{n-1, 1-\frac{\alpha}{2}}^2$ — квантили χ^2 -распределения порядка $\frac{\alpha}{2}$ и $1-\frac{\alpha}{2}$ соответственно.

Решая неравенство, находим $\sigma^2 \in \left[\frac{(n-1)S^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)S^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right]$.

Рассмотрим теперь вопрос о нормированной сумме случайных величин, имеющих распределение Коши с одинаковыми параметрами. Пусть $X_i \sim K(a, b)$, $S_n = \sum_{i=1}^n X_i$. Характеристическая функция $\varphi_{X_i}(t) = e^{iat-b|t|}$. Рассмотрим характеристическую функцию случайной величины $Y = \frac{S_n}{n}$:

$$\varphi_Y(t) = \mathbb{E}e^{itY} = \mathbb{E}e^{\frac{it}{n} \sum_{i=1}^n X_i} = (\mathbb{E}e^{\frac{it}{n} X_1})^n = \varphi_{X_1}(t).$$

Следовательно, $Y \sim K(a, b)$.

5.2 Результат работы программы

Для проверки на практике ЗБЧ и ЦПТ генерируем n независимых случайных величин, имеющих нормальное распределение с параметрами μ, σ^2 и рассматриваем поведение указанных в условии величин.

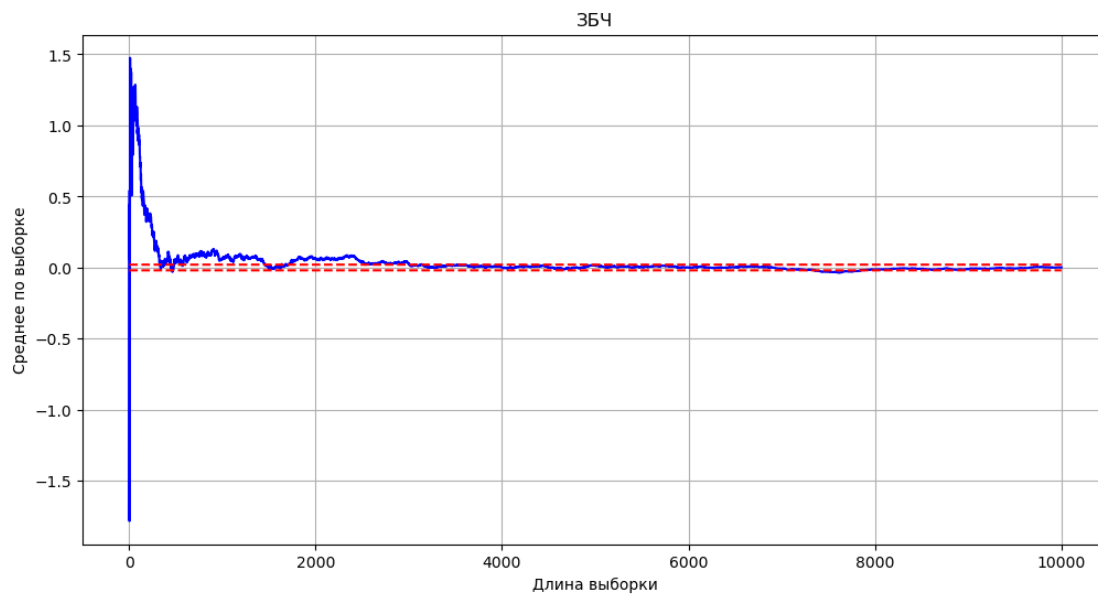


Рис. 21: График среднего стандартных нормальных случайных величин

Заметно, что среднее сгенерированных случайных величин совпадает с теоретической оценкой в ЗБЧ.

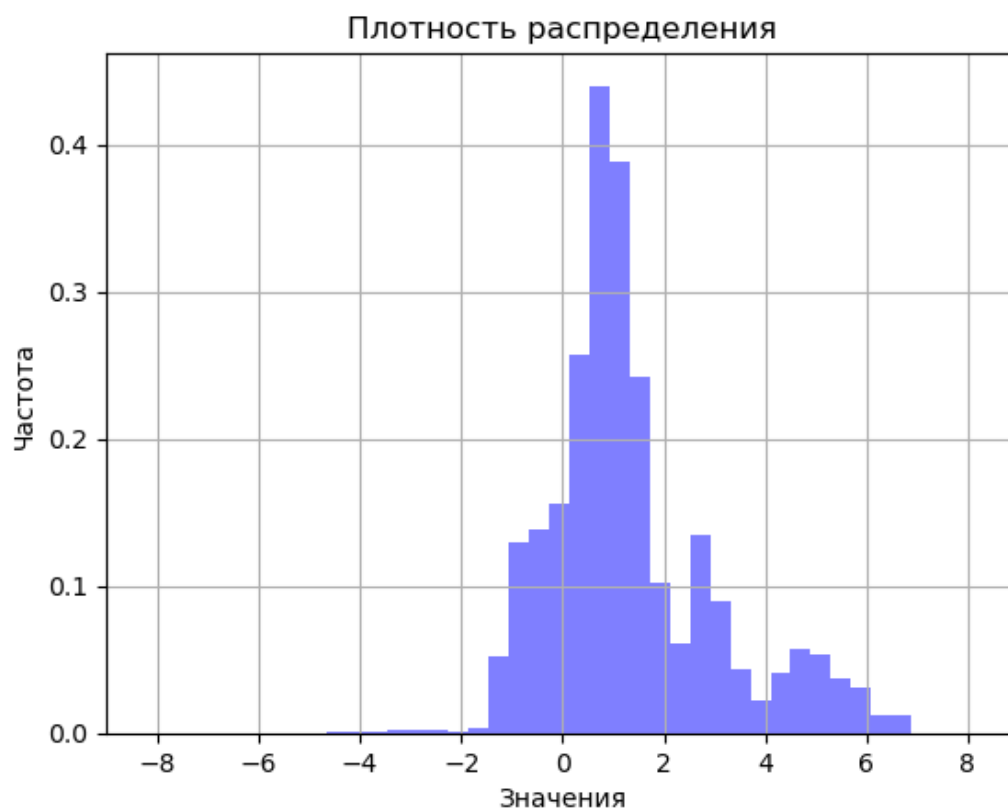


Рис. 22: Распределение центрированной нормированной суммы независимых случайных величин, имеющих стандартное нормальное распределение

Очевидно, что ЦПТ работает.

Рассмотрим теперь доверительные интервалы для оценки параметров случайной величины, имеющей нормальное распределение.

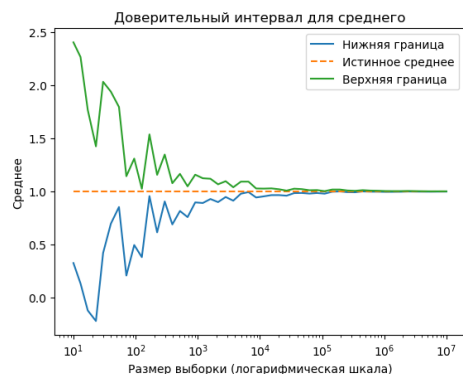


Рис. 23: Доверительный интервал для среднего, построенный с помощью t-критерия Стьюдента

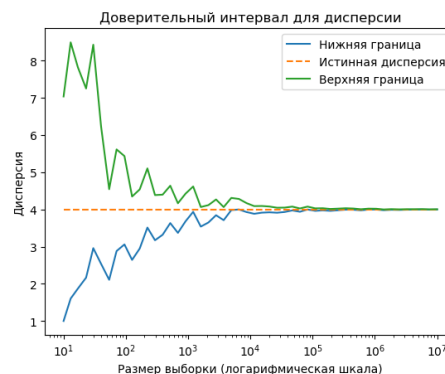


Рис. 24: Доверительный интервал для дисперсии, построенный с помощью критерия Фишера

Заметим, что доверительные интервалы дают хорошую оценку для большого размера выборки.

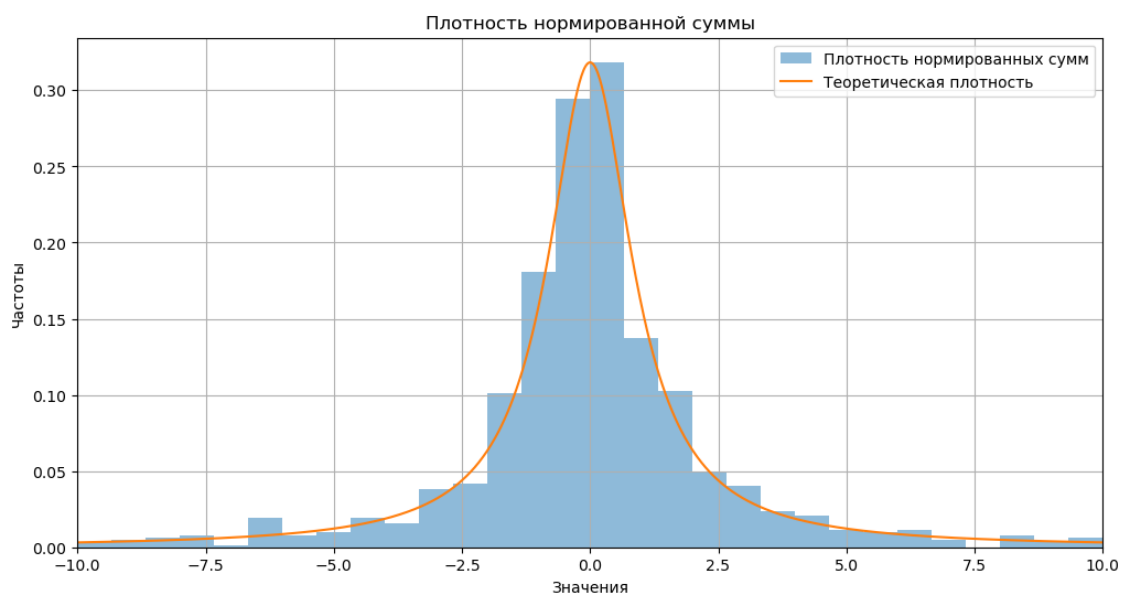


Рис. 25: Плотность нормированной суммы случайных величин, имеющих распределение Коши с параметрами $a = 0, b = 1$

Очевидно, полученная картинка достаточно хорошо согласуется с указанным в теоретической части результатом.

6 Задача 6

Вычислить следующий интеграл:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{e^{-\left(x_1^2 + \cdots + x_{10}^2 + \frac{1}{2^7 \cdot x_1^2 \cdots x_{10}^2}\right)}}{x_1^2 \cdots x_{10}^2} dx_1 dx_2 \cdots dx_{10}.$$

1. Методом Монте-Карло.
2. Методом квадратур, сводя задачу к вычислению собственного интеграла Римана.
3. Оценить точность вычислений для каждого из двух случаев.

6.1 Теоретическая часть

Найдем интеграл методом Монте-Карло. Рассмотрим $X_1, X_2, \dots, X_{10} \sim \mathcal{N}(0, \frac{1}{\sqrt{2}})$ — независимые случайные величины. Плотность каждой $f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$. Пусть $y = \left(\prod_{i=1}^{10} x_i\right)^{-2}$. Тогда подынтегральная функция равна $\pi^5 y e^{-\frac{y}{2^7}} f(x)$. Тогда весь интеграл равен математическому ожиданию $\mathbb{E}g(Y)$, где $g(y) = \pi^5 y e^{-\frac{y}{2^7}}, Y = \left(\prod_{i=1}^{10} X_i\right)^{-2}$.

Тогда по усиленному закону больших чисел $\frac{S_n}{n} = \frac{\sum_{j=1}^n g(Y_j)}{n} \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}g(Y)$. Здесь Y_j — обратные квадраты произведения 10 сгенерированных случайных величин $X_{i_j}, i_j = 1, 2, \dots, 10$.

Рассмотрим теперь вычисление интеграла методом квадратур. Заметим, что подынтегральная функция четная и можно рассматривать интегралы от 0 до $+\infty$. Сделаем замены: $x_i = \tan(\frac{\pi}{2} y_i), y_i \in (0, 1)$. После замены получим

$$I = \pi^{10} \int_0^1 \int_0^1 \cdots \int_0^1 \frac{\exp\left\{\sum_{k=1}^{10} \tan^2\left(\frac{\pi}{2} y_k\right) + \frac{1}{128} \prod_{k=1}^{10} \frac{1}{\tan^2\left(\frac{\pi}{2} y_k\right)}\right\}}{\prod_{k=1}^{10} \sin^2\left(\frac{\pi}{2} y_k\right)} dy_1 dy_2 \cdots dy_{10}.$$

Такой интеграл можно посчитать методом квадратур, т.е. приблизить его суммой площадей элементарных фигур. Формула для метода прямоугольников со значением

в середине отрезка разбиения:

$$I \approx \left(\frac{\pi}{N}\right)^{10} \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_{10} \leq N} C_{m_1, \dots, m_l}(m) f(x_1^{2i_1-1}, x_2^{2i_2-1}, \dots, x_{10}^{2i_{10}-1}),$$

где $C_{m_1, \dots, m_l}(m) = \frac{10!}{m_1! m_2! \dots m_l!}$, m_k - количество индексов в наборе $(i_1, i_2, \dots, i_{10})$, равных k .

Для метода Монте-Карло погрешность определяется через центральную предельную теорему. Вероятность того, что ошибка оценки интеграла $|\bar{I} - I|$ не превысит заданного значения ε , выражается как:

$$P(|\bar{I} - I| \leq \varepsilon) = 1 - \alpha,$$

где:

- \bar{I} — оценка интеграла по выборке,
- I — истинное значение интеграла,
- α — уровень значимости.

Погрешность оценивается по формуле:

$$\varepsilon = \frac{\sigma_n}{\sqrt{n}} K_{1-\frac{\alpha}{2}},$$

где:

- $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n \left(f(x_i) - \overline{f(x)}\right)^2$ — выборочная дисперсия функции $f(x)$,
- n — число выборок,
- $K_{1-\frac{\alpha}{2}}$ — квантиль стандартного нормального распределения.

Погрешность метода квадратур оценивается с использованием формулы для метода трапеций:

$$E_{\text{quad}} = \sum_{i=1}^k \frac{\max_{x_i \in [0,1]} |f''(x_i)|}{24} h^3,$$

где:

- k — размерность интеграла,
- $h = \frac{1}{N}$ — шаг разбиения,

- $\max_{x_i \in [0,1]} |f''(x_i)|$ — максимум второй производной функции $f(x)$ на интервале $[0, 1]$.

Погрешность метода квадратур зависит от гладкости функции $f(x)$ и точности её аппроксимации квадратичным полиномом. Чем меньше значение h , тем выше точность метода, но это увеличивает вычислительные затраты, особенно для высокомерных интегралов.

6.2 Результат работы программы

При размере выборки $n = 100000$ значение интеграла, полученное методом Монте-Карло, равно 125.60480142864812. При разбиении на $N = 15$ отрезков длины $\frac{1}{15}$ отрезка $[0, 1]$ значение интеграла, полученное методом квадратур, равно 119.30402795064616.

7 Задача 7

1. Методом случайного поиска найти минимальное значение функции f на множестве $A = \{x_1, x_2 : x_1^2 + x_2^2 \leq 1\}$, где

$$f(x_1, x_2) = x_1^3 \sin \frac{1}{x_1} + 10x_1x_2^4 \cos \frac{1}{x_2}, \quad x_1, x_2 \neq 0$$

При $x_1 = 0$ или $x_2 = 0$ функция доопределяется по непрерывности.

2. Методом имитации отжига найти минимальное значение функции Розенброка g в пространстве \mathbb{R}^2 , где

$$g(x) = (x_1 - 1)^2 + 100(x_2 - x_1^2)^2$$

3. Оценить точность и сравнить результаты со стандартными методами оптимизации.

7.1 Теоретическая часть

Метод случайного поиска. Будем генерировать 2 случайные величины $r \sim \mathcal{U}(0, 1)$ и $\varphi \sim \mathcal{U}(0, 2\pi)$. Тогда $x_1 = r \cos \varphi$, $x_2 = r \sin \varphi$. Однако при таком выборе распределения r точки будут сосредоточены в центре круга. Поэтому хотим получить, что на каждом кольце было одинаковое количество точек, не зависящее от радиуса.

Длина окружности линейно зависит от радиуса \implies плотность распределения радиуса должна быть линейной. тогда $f(r) = kr\mathbb{I}(r \leq 1)$. Найдем константу из условия нормировки плотности:

$$\int_0^1 k r dr = k/2 = 1 \implies k = 2.$$

Функция распределения тогда равна $F(r) = r^2\mathbb{I}(r \leq 1) + \mathbb{I}(r > 1)$. Такая функция строго монотонна на отрезке $[0, 1]$ и непрерывна, значит $r \sim F^{-1}(U(0, 1)) = \sqrt{U(0, 1)}$, где $U \sim \mathcal{U}(0, 1)$. Таким образом, генерируем $r \sim \sqrt{\mathcal{U}(0, 1)}$, $\varphi \sim \mathcal{U}(0, 2\pi)$. На следующем рисунке представлено указанное равномерное распределение в круге.

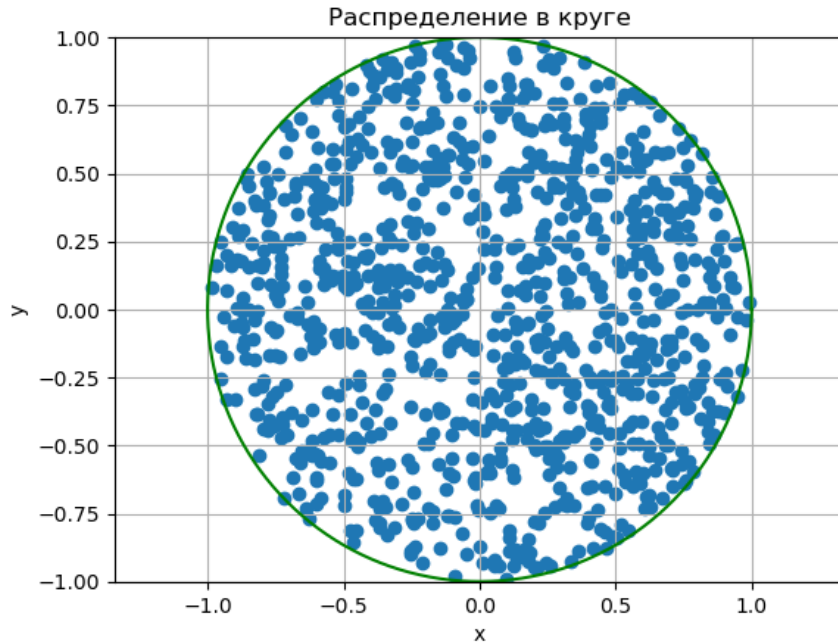


Рис. 26: Равномерное распределение точек в круге

Метод имитации отжига.

Функция выбора соседа: нормальная случайная величина со средним (x_i, y_i) и дисперсией $\sigma^2 T_i$;

Функция вероятности перехода: $\mathbb{P}(\Delta F_i) = \frac{1}{2} e^{-\Delta F_i / T_i}$;

Функция понижения температуры: $T_{i+1} = kT_i, k \in (0, 1)$.

Также нужно задать начальное состояние (x_0, y_0) и начальную температуру T .

Алгоритм: 1. Генерируется новый кандидат $(x_{i+1}, y_{i+1}) \sim (\mathcal{N}((x_i,), \sigma^2 T_i), \mathcal{N}((y_i,), \sigma^2 T_i))$ и вычисляется приращение минимизируемой функции $\Delta F_i = F(x_{i+1}, y_{i+1}) - F(x_i, y_i)$; 2. Если $\Delta F_i \leq 0$, то новое состояние равно новому кандидату, иначе переход выполняется с вероятностью $\mathbb{P}(\Delta F_i)$; 3. Новая температура вычисляется по формуле выше. Если $T_{i+1} \leq \varepsilon$, то алгоритм останавливается, иначе переход к шагу 1.

Оценим точность метода случайного поиска. Из формулы Лагранжа получаем следующую оценку:

$$|f(x_1, x_2) - f(x_1^*, x_2^*)| \leq \sup_{(x'_1, x'_2) \in A} \left\{ \sqrt{\left(\frac{\partial f(x'_1, x'_2)}{\partial x_1} \right)^2 + \left(\frac{\partial f(x'_1, x'_2)}{\partial x_2} \right)^2} \right\} \|(x_1 - x_1^*, x_2 - x_2^*)\|.$$

$$f(x_1, x_2) = x_1^3 \sin \frac{1}{x_1} + 10x_1 x_2^4 \cos \frac{1}{x_2}, \quad x_1, x_2 \neq 0$$

Посчитаем частные производные:

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 3x_1^2 \sin \frac{1}{x_1} - x_1 \cos \frac{1}{x_1} + 10x_2^4 \cos \frac{1}{x_2} \\ \frac{\partial f}{\partial x_2} &= 40x_1 x_2^3 \cos \frac{1}{x_2} + 10x_1 x_2^2 \sin \frac{1}{x_2} \end{aligned}$$

Тогда

$$|f(x_1, x_2) - f(x_1^*, x_2^*)| \leq \sqrt{14^2 + 50^2} \|(x_1 - x_1^*, x_2 - x_2^*)\|.$$

Пусть (x_1^*, x_2^*) — теоретически полученная точка минимума, $\|(x_1 - x_1^*, x_2 - x_2^*)\| = \delta$. Получаем точность $\varepsilon \leq \sqrt{2696}\delta$. Введем дополнительные параметры: пусть p — вероятность попадания хотя бы одной точки в δ -окрестность точки минимума. Тогда $p = 1 - \mathbb{P}(\forall i : \|x_i - x^*\| > \delta) = 1 - (\mathbb{P}(\|x_i - x^*\| > \delta))^n = 1 - (1 - \mathbb{P}(\|x_i - x^*\| \leq \delta))^n$, а вероятность того, что конкретная точка попала в δ -окрестность не меньше δ^2 . Таким образом, $p \geq 1 - (1 - \delta)^n$. Тогда $\varepsilon \leq \sqrt{2696} (1 - (1 - p)^{1/n})$.

7.2 Результат работы программы

При размере выборки $n = 1000$ метод случайного поиска выдал следующий результат:

- Вычисленное значение минимума равно -1.1629202559916427 ;
- Точка, в которой достигается минимум $(-0.4185848638964709, 0.8993598838245854)$.

Стандартный метод оптимизации COBYLA выдал следующий результат:

- Вычисленное значение минимума равно -1.2884805581553458 ;
- Точка, в которой достигается минимум $(-0.3573, 0.9340)$.

Значения, полученные при помощи данных методов достаточно близки.

Известно, что функция Розенброка имеет минимум в точке $(1, 1)$ и значение функции в этой точке равно 0. Проведем несколько итераций и посмотрим на график решений, найденных методом имитации отжига.

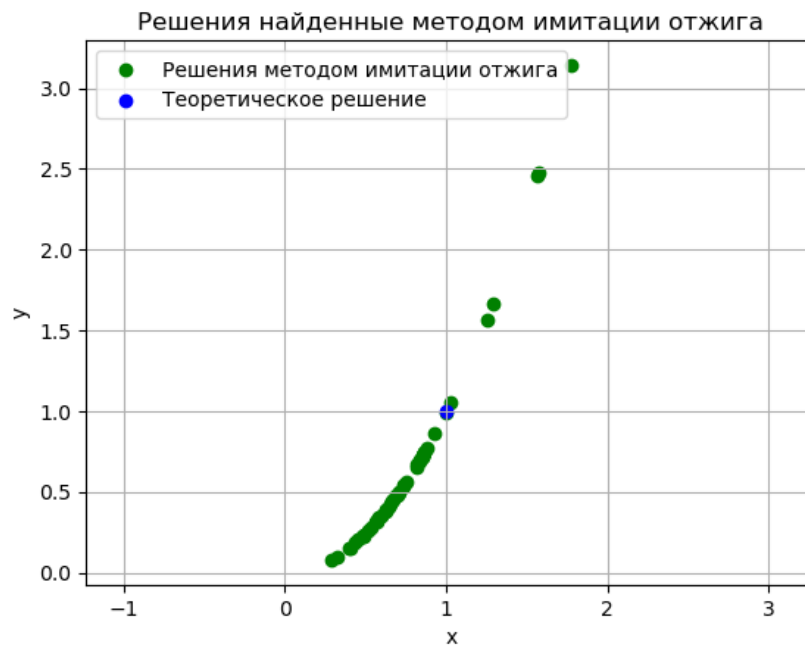


Рис. 27: Решения, найденные методом имитации отжига

Метод сопряженных градиентов выдал следующий результат:

- Вычисленное значение минимума равно $2.0056121165089653e-11$;
- Точка, в которой достигается минимум $(1.000, 1.000)$.

Как видим, решения достаточно близки к теоретическому минимуму.

8 Задача 8

Применить метод Монте-Карло к решению первой краевой задачи для двумерного уравнения Лапласа в единичном круге:

$$\begin{cases} \Delta u = 0, & (x, y) \in D = \{x, y : x^2 + y^2 \leq 1\}, \\ u|_{\partial D} = f(x, y), \\ u \in C^2(D), & f \in C(\partial D), \end{cases}$$

Для функции $f(x, y) = x^2 - y^2$ найти аналитическое решение и сравнить с полученным по методу Монте-Карло.

8.1 Теоретическая часть

Рассмотрим квадратную сетку $x_i = -1 + i * h, y_j = -1 + j * h, h = 1/n, n$ – количество точек разбиения отрезка $[-1, 1]$. Естественным образом определяются понятия соседних точек. Далее разделим все узлы сетки, принадлежащие кругу D на две группы: внутренние (все 4 соседа этой вершины лежат в круге D) и граничные (есть сосед вне круга D .) В граничных узлах u равна значению функции f в этом узле, а во внутренних в соответствии с разностной схемой:

$$u_{ij} = \frac{1}{4} (u_{i,j-1} + u_{i,j+1} + u_{i-1,j} + u_{i+1,j}).$$

Будем рассматривать задачу блуждания частицы по этой сетке. Пусть частица начинает блуждание в узле (x_i, y_j) . Вероятности переходов в соседние узлы равны $\frac{1}{4}$. Будем считать, что частица, попавшая на граничный узел останавливается в нем. Тогда возникает задача нахождения вероятности того, что частица, начав путь по сетке из вершины (x_i, y_j) , закончит его в граничной точке (x^*, y^*) . Обозначим ее через p_{ij}^* . Считаем, что частица обязательно попадет на границу. Воспользуемся формулой полной вероятности:

$$p_{ij}^* = \frac{1}{4} (p_{i,j-1}^* + p_{i,j+1}^* + p_{i-1,j}^* + p_{i+1,j}^*).$$

Дополнительно заметим, что $p_*^* = 1, p_+^* = 0$, где (x^+, y^+) – граничная точка.

Пусть частица при попадании на границу переносит температуру $f(x^*, y^*)$. Значение температуры, перенесенной из узла (x_i, y_j) является случайной величиной ξ_{ij} , вероятность переноса температуры $f(x^*, y^*)$ равна p_{ij}^* . Тогда верно следующее ра-

венство:

$$\mathbb{E}\xi_{ij} = \sum_{(x^*, y^*)} f(x^*, y^*) p_{ij}^*.$$

Обозначим эту функцию через $g(i, j)$. Для нее справедливо такое же тождество, как и для p_{ij}^* :

$$g(i, j) = \frac{1}{4}(g(i-1, j) + g(i+1, j) + g(i, j-1) + g(i, j+1)).$$

На границе для функции g верно следующее: $g(*, *) = f(x^*, y^*)$.

Тогда функция $g(i, j)$ является решением исходной задачи Дирихле.

8.2 Результат работы программы

На следующей картинке показана двумерная сетка с шагом $h = 0.05$, в узлах которой будет вычислено значение функции. Показаны только точки, лежащие внутри единичного круга.

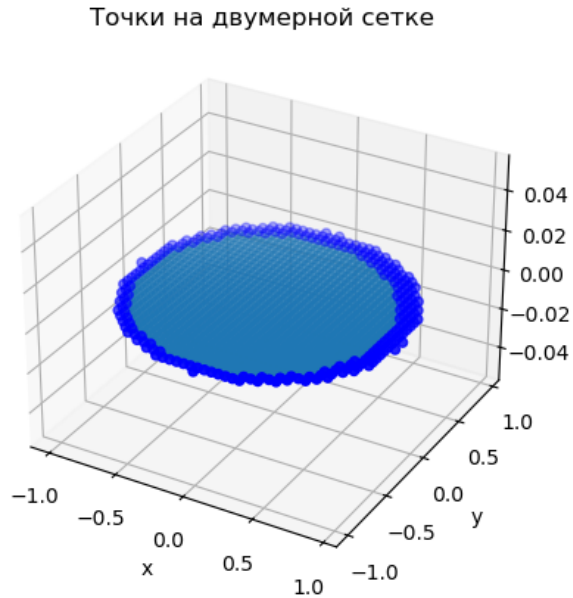


Рис. 28: Точки сетки внутри круга

Полученное при числе произведенных случайных блужданий $n = 10000$ и аналитическое решения выглядят следующим образом.

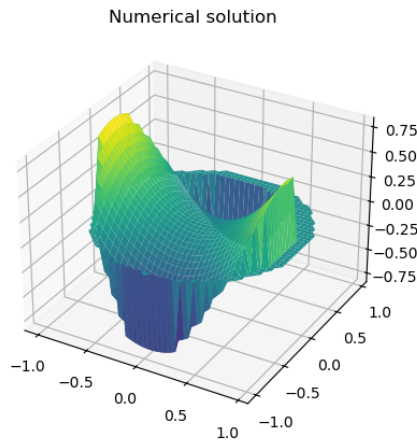


Рис. 29

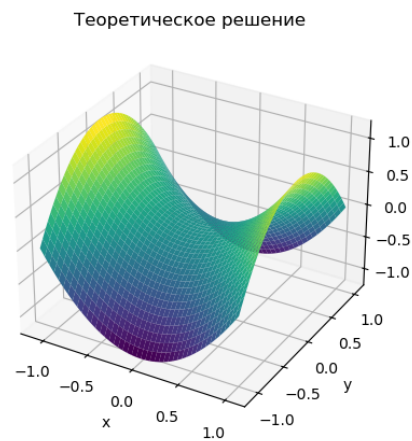


Рис. 30

Заметно, что эти решения достаточно близки.

9 Задача 9

Рассмотреть два вида гауссовских процессов:

- Винеровский процесс $W(t), t \in [0; 1], W(0) = 0$
- Процесс Орнштейна–Уленбека $X(t), t \in [0; 1], X(0) = X_0$, т.е. стационарный марковский гауссовский процесс. Начальные значения X_0 следует выбирать случайным образом так, чтобы полученный процесс был стационарным.

Для данных процессов:

1. Найти ковариационную функцию и переходные вероятности.
2. Промоделировать независимые траектории процесса с данными переходными вероятностями методом добавления разбиения отрезка.
3. Построить график траектории, не соединяя точки ломаной, с целью получения визуально непрерывной линии.

9.1 Теоретическая часть

Случайный процесс $W(t), t \in [0, 1]$ называют винеровским с параметром σ , если
 1. $W(0) = 0$ почти наверное; 2. $W(t)$ – процесс с независимыми приращениями; 3. $W(t) - W(s) \sim \mathcal{N}(0, \sigma^2(t - s))$.

Отсюда получаем $W(t) - W(0) \sim \mathcal{N}(0, \sigma^2 t)$ и $\mathbb{E}W(t) = 0, \text{Var } W(t) = \sigma^2 t$.

Если $s \leq t$, то $W(t) = W(s) + (W(t) - W(s))$. В силу независимости приращений получаем $R(s, t) = \mathbb{E}W(s)W(t) = \mathbb{E}W^2(s) + \mathbb{E}W(s)\mathbb{E}(W(t) - W(s)) = \mathbb{E}W^2(s) = \sigma^2 s$. Если $t < s$, то $W(s) = W(t) + (W(s) - W(t))$. Аналогично получаем $R(s, t) = \mathbb{E}W^2(t) = \sigma^2 t$. Объединяя полученные результаты, получаем $R(s, t) = \sigma^2 \min\{s, t\}$.

Пусть $t_1 < t_2 < t_3$. Найдем распределение $Y = (W(t_2)|W(t_1) = x_1, W(t_3) = x_3)$.
 $W(t) \sim \mathcal{N}(0, \sigma^2 t) \implies f_Y(x_2) = \frac{f_{123}(x_1, x_2, x_3)}{f_{13}(x_1, x_3)}$, где $f_{123}(x_1, x_2, x_3)$ – совместная плотность $(W(t_1), W(t_2), W(t_3))$, $f_{13}(x_1, x_3)$ – совместная плотность $(W_1(t_1), W_3(t_3))$. Известно, что плотность n -мерного нормального распределения с математическим ожиданием μ и ковариационной матрицей Σ имеет вид:

$$f(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right).$$

Используя полученную формулу для ковариационной матрицы, получим:

$$\Sigma_{13} = \sigma^2 \begin{pmatrix} t_1 & t_1 \\ t_1 & t_3 \end{pmatrix}, \quad \Sigma_{123} = \sigma^2 \begin{pmatrix} t_1 & t_1 & t_1 \\ t_1 & t_2 & t_2 \\ t_1 & t_2 & t_3 \end{pmatrix}.$$

Для простоты формулы предположим, что $t_2 = \frac{t_1 + t_3}{2}$.

Воспользовавшись пакетом `sympy`, получили дисперсию $S = \frac{\sigma^2(-t_1 + t_3)}{4}$
 и степень экспоненты $\frac{0.5x_1^2 - 2x_1x_2 + x_1x_3 + 2x_2^2 - 2x_2x_3 + 0.5x_3^2}{\sigma^2(t_1 - t_3)} =$
 $-\frac{x_1^2 - 4x_1x_2 + 2x_1x_3 + 4x_2^2 - 4x_2x_3 + x_3^2}{2\sigma^2(t_3 - t_1)} = -\frac{(2x_2 - (x_1 + x_3))^2}{2\sigma^2(t_3 - t_1)} = -\frac{(x_2 - \frac{x_1 + x_3}{2})^2}{2\frac{\sigma^2(t_3 - t_1)}{4}}.$

Таким образом $Y \sim \mathcal{N}\left(\frac{x_1 + x_3}{2}, \frac{\sigma^2(-t_1 + t_3)}{4}\right)$.

Для моделирования винеровского процесса воспользуемся методом добавления разбиения отрезка. Знаем, что $W(1) = W(1) - W(0) \sim \mathcal{N}(0, \sigma^2), W(0) = 0$. Зная значения $W(t_1)$ и $W(t_3)$, найдем $W(t_2) = W\left(\frac{t_1 + t_3}{2}\right) \sim \mathcal{N}\left(\frac{x_1 + x_3}{2}, \frac{\sigma^2(t_3 - t_1)}{4}\right)$.

Будем дробить отрезок до тех пор, пока не наберем точек не меньше требуемого количества, при этом если проходиться по всем парам соседних посчитанных значений $W(t_1)$ и $W(t_3)$, то получим $2^n + 1$ точку, где n – номер итерации прохода по всем значениям.

Процесс Орнштейна-Уленбека вводится как единственный нетривиальный гауссовский процесс, являющийся стационарным и обладающий марковским свойством. Из стационарности следует, что $\mathbb{E}X_t = a, \text{var}(X_t) = \sigma^2, K_X(s, t) = \rho(|s - t|)$. Из марковского свойства следует, что $K_X(s, t) = K_X(s, \tau)K_X(\tau, t), \forall 1 \leq s \leq \tau \leq t \leq n$. $\rho(|s - t|) = \rho(|s - \tau|)\rho(|\tau - t|) \Leftrightarrow \rho(x + y) = \rho(x)\rho(y)$. Получили уравнение Коши, которое имеет тривиальное решение и решение $K_X(t) = e^{-\lambda t}, \lambda > 0$. В силу гауссовости получаем, что все X_t независимы в совокупности, поэтому тривиальное решение не рассматриваем. Тогда $R(s, t) = \sigma^2 e^{-\lambda|s-t|}, \lambda > 0$.

Значение $X(0) = X_0$ неизвестно, поэтому будем генерировать из нормального распределения с нулевым средним и дисперсией σ^2 . Найдем распределение $Y = (X(1)|X(0) = x_0)$.

Ковариационная матрица равна $R_0 = \sigma^2 \begin{pmatrix} 1 & e^{-\lambda} \\ e^{-\lambda} & 1 \end{pmatrix}$. Тогда

$$f_Y(x_1) = \frac{f_{X_0 X_1}(x_0, x_1)}{f_{X_0}(x_0)}$$

$$f_{X_0 X_1}(x_0, x_1) = \frac{1}{\sqrt{4\pi^2 \det(R_0)}} \exp\left\{-\frac{1}{2} x R_0^{-1} x^T\right\}$$

$$f_{X_0}(x_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x_0^2}{2\sigma^2}\right\}.$$

$$R_0^{-1} = \frac{1}{\sigma^2(1 - e^{-2\lambda})} \begin{pmatrix} 1 & -e^{-\lambda} \\ -e^{-\lambda} & 1 \end{pmatrix}$$

$$x R_0^{-1} x^T = \frac{1}{\sigma^2(1 - e^{-2\lambda})} \begin{pmatrix} x_0 & x_1 \end{pmatrix} \begin{pmatrix} 1 & -e^{-\lambda} \\ -e^{-\lambda} & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = \frac{1}{\sigma^2(1 - e^{-2\lambda})} (x_0^2 - 2x_0 x_1 e^{-\lambda} + x_1^2)$$

$$\begin{aligned} -\frac{1}{2\sigma^2} \left(\frac{1}{1 - e^{-2\lambda}} (x_0^2 - 2x_0 x_1 e^{-\lambda} + x_1^2) - x_0^2 \right) &= -\frac{1}{2\sigma^2(1 - e^{-2\lambda})} (x_1^2 - 2x_0 x_1 e^{-\lambda} + x_0^2 e^{-2\lambda}) = \\ &= -\frac{(x_1 - x_0 e^{-\lambda})^2}{2\sigma^2(1 - e^{-2\lambda})} \end{aligned}$$

Получаем

$$\sigma^* = \sigma \sqrt{1 - e^{-2\lambda}}, \quad \mu^* = x_0 e^{-\lambda}.$$

Таким образом, $(X(1)|X(0) = x_0) \sim \mathcal{N}(x_0 e^{-\lambda}, \sigma^2(1 - e^{-2\lambda}))$.

Далее хотим смоделировать процесс методом разбиения отрезка. Найдем распределение $Y = (X(t_2)|X(t_1) = x_1, X(t_3) = x_3)$, $t_1 < t_3, t_2 = \frac{t_1 + t_3}{2}$.

$$f_{X_{t_2}}(x_2) = \frac{f_{123}(x_1, x_2, x_3)}{f_{13}(x_1, x_3)}.$$

$$\Sigma_{123} = \sigma^2 \begin{pmatrix} 1 & e^{-\lambda(t_2-t_1)} & e^{-\lambda(t_3-t_1)} \\ e^{-\lambda(t_2-t_1)} & 1 & e^{-\lambda(t_3-t_2)} \\ e^{-\lambda(t_3-t_1)} & e^{-\lambda(t_3-t_2)} & 1 \end{pmatrix}$$

$$\Sigma_{13} = \sigma^2 \begin{pmatrix} 1 & e^{-\lambda(t_2-t_1)} \\ e^{-\lambda(t_2-t_1)} & 1 \end{pmatrix}$$

Получили $\sigma^* = \sigma \sqrt{\tanh(d\lambda)} = \sigma \sqrt{\frac{e^{2d\lambda} - 1}{e^{2d\lambda} + 1}}$. Раскрывая скобки и приводя подобные в степени экспоненты, получим $\mu^* = \frac{e^{\lambda d}(x_1 + x_3)}{e^{2\lambda d} + 1}$.

Таким образом $(X(t_2)|X(t_1) = x_1, X(t_3) = x_3) \sim \mathcal{N}\left(\frac{e^{\lambda d}(x_1 + x_3)}{e^{2\lambda d} + 1}, \sigma^2 \frac{e^{2d\lambda} - 1}{e^{2d\lambda} + 1}\right)$

9.2 Результат работы программы

На следующем рисунке показаны траектории винеровского процесса с параметром $\sigma = 2$, разбиение отрезка $[0, 1]$ производилось на 2^{15} частей.

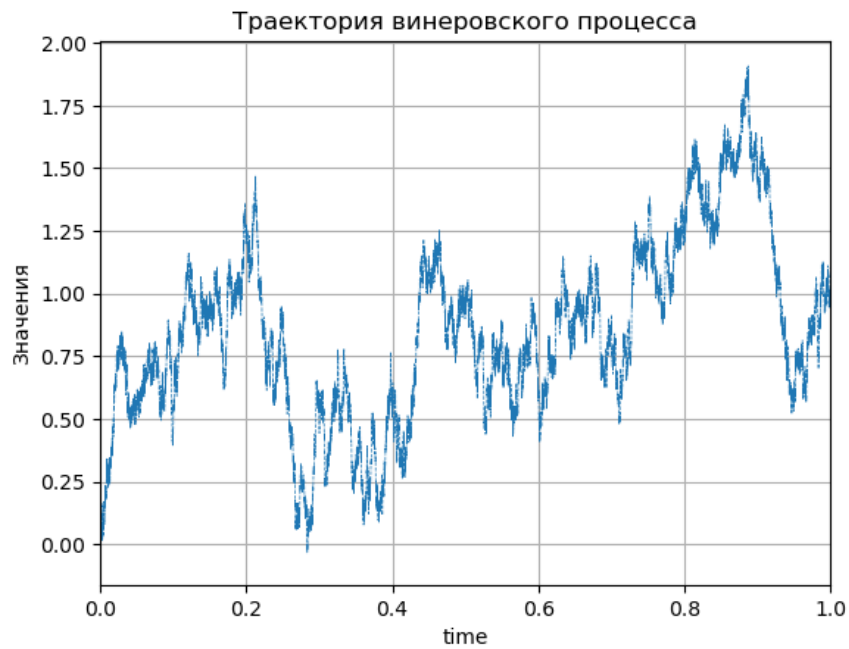


Рис. 31: Траектории винеровского процесса

На следующем рисунке показаны траектории процесса Орнштейна-Уленбека с параметрами $\sigma = 2$, $\lambda = 1$, разбиение отрезка $[0, 1]$ производилось на 2^{15} частей.

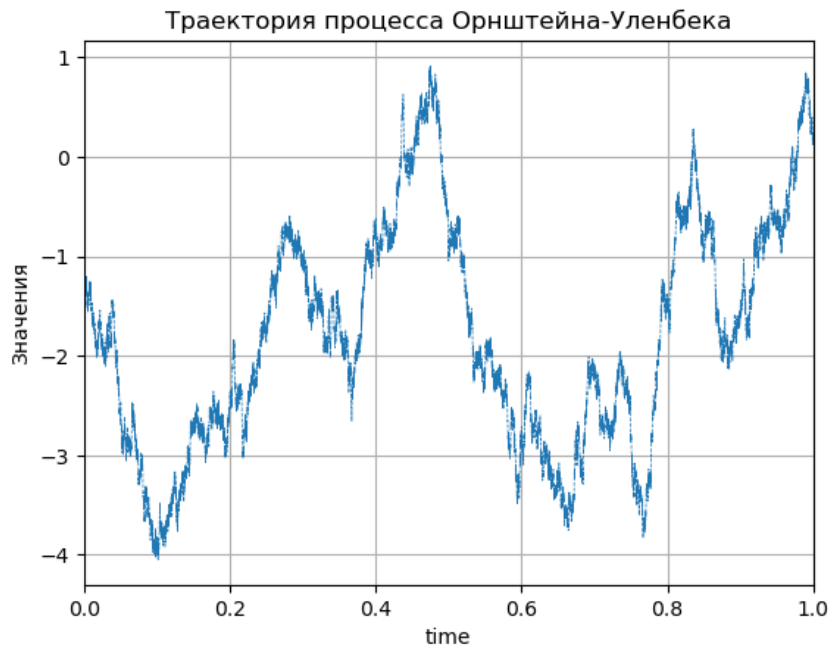


Рис. 32: Траектории процесса Орнштейна-Уленбека

10 Задача 10

Произвести фильтрацию одномерного процесса Орнштейна–Уленбека:

1. Используя генератор белого шума, добавить к реализации процесса Орнштейна–Уленбека случайную ошибку с заранее известной дисперсией. 2. При помощи одномерного фильтра Калмана оценить траекторию процесса по зашумленному сигналу, считая известными параметры шума и процесса. 3. Рассмотреть следующие виды шума: 1. Гауссов 2. Коши (шум имеет распределение Коши)

10.1 Теоретическая часть

Рассматриваем процесс Орнштейна-Уленбека на сетке, тогда его "координаты" изменяются по следующему закону:

$$x_{k+1} = ax_k + v_k$$

Динамика зашумленного сигнала:

$$y_k = x_k + \xi_k$$

По условию знаем дисперсию ошибки: $\xi_k \sim N(0, r^2)$.

Будем считать, что $v_k \sim N(0, q^2)$. Из предыдущего задания знаем, что $x_0 \sim N(0, \sigma^2)$. Для дальнейшего решения задачи необходимо найти значения параметров a, q . Зная, что x_k и v_k независимы, а также свойства процесса Орнштейна-Уленбека:

$$R(t_k, t_k) = \sigma^2,$$

$$R(t_k, t_{k+1}) = \text{cov}(x_{t_k}, x_{t_{k+1}}) = a \cdot \text{var}(x_{t_k}) = \sigma^2 e^{-\lambda(t_{k+1}-t_k)} \Rightarrow a = e^{-\lambda(t_{k+1}-t_k)},$$

$$R(t_{k+1}, t_{k+1}) = a^2 \sigma^2 + q^2 = \sigma^2 \Rightarrow q^2 = \sigma^2(1 - e^{-2\lambda(t_{k+1}-t_k)}).$$

Прогнозируемое фильтром Калмана значение x_{k+1}^* :

$$x_{k+1}^* = K_{k+1}y_{k+1} + (1 - K_{k+1})ax_k^*,$$

где K_{k+1} – коэффициент Калмана.

Далее необходимо выбрать коэффициент Калмана K_{k+1} таким, чтобы получившееся оптимальное значение прогноза x_{k+1}^* было наиболее близко к истинному значению x_{k+1} . Чтобы найти точное значение коэффициента Калмана K_{k+1} , нужно минимизировать ошибку:

$$e_{k+1} = x_{k+1} - x_{k+1}^*$$

Тогда

$$\begin{aligned} e_{k+1} &= x_{k+1} - K_{k+1}y_{k+1} - (1 - K_{k+1})ax_k^* = ax_k + v_k - K_{k+1}(ax_k + v_k + \xi_{k+1}) - (1 - K_{k+1})ax_k^* = \\ &= (1 - K_{k+1})(ae_k + v_k) - K_{k+1}\xi_{k+1}. \end{aligned}$$

Тогда необходимо минимизировать $\mathbb{E}(e_{k+1}^2) \rightarrow \min$:

$$\mathbb{E}(e_{k+1}^2) = (1 - K_{k+1})^2(a^2\mathbb{E}e_k^2 + q^2) + K_{k+1}^2r^2$$

Приравнивая производную (по K_{k+1}) выражения к 0, получим

$$K_{k+1} = \frac{a^2\mathbb{E}e_k^2 + q^2}{a^2\mathbb{E}e_k^2 + q^2 + r^2}$$

Подставляем это значение в минимизируемую функцию

$$\mathbb{E}(e_{k+1}^2) = \frac{r^2(a^2\mathbb{E}e_k^2 + q^2)}{a^2\mathbb{E}e_k^2 + q^2 + r^2}$$

Запишем алгоритм работы программы:

1) $\mathbb{E}(e_0^2) = r^2$, $x_0^* = y_0$. 2) На каждой итерации вычисляем $\mathbb{E}(e_{k+1}^2)$ по формуле, представленной выше. Далее вычисляем $K_{k+1} = \frac{\mathbb{E}(e_{k+1}^2)}{r^2}$. Зная коэффициент Калмана вычисляем x_{k+1}^* по формуле выше. Параметры программы: σ, r, λ , число точек.

Ошибка фильтрации имеет нормальное распределение и доверительный интервал имеет вид:

$$[x^* - \varepsilon, x^* + \varepsilon], \quad \varepsilon = -\sqrt{\mathbb{E}e} \cdot erf\left(\frac{\alpha}{2}\right),$$

где α - уровень значимости. Для шума с распределением Коши остаются верны те же формулы.

10.2 Результат работы программы

На следующих рисунках представлены незашумленный процесс, процесс с Гауссовым шумом с дисперсией 0.01 и процесс с применением фильтра Калмана. Также показан доверительный интервал распределения ошибки.

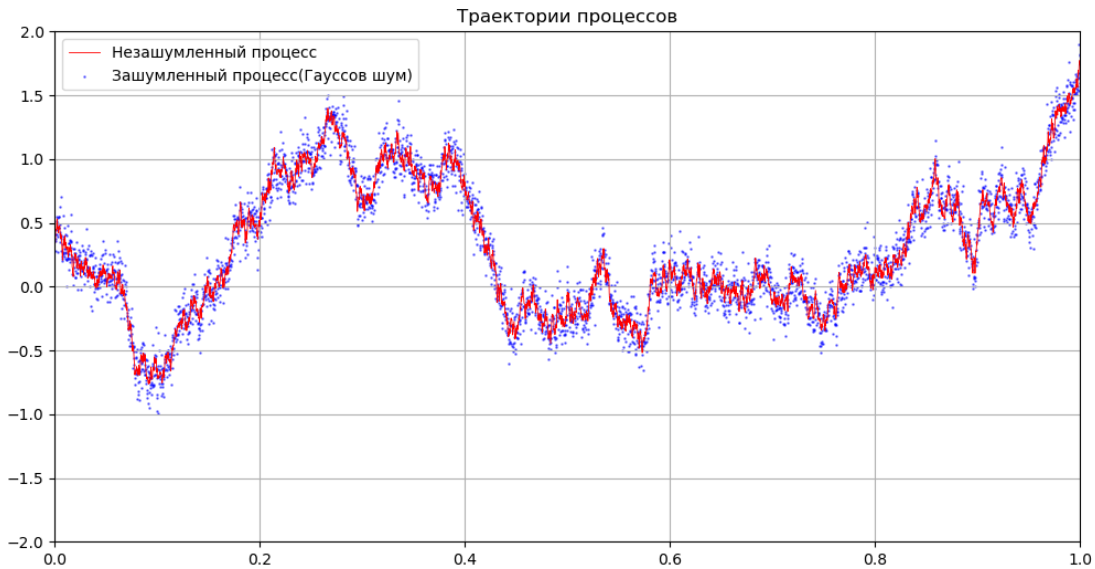


Рис. 33: Траектории процессов

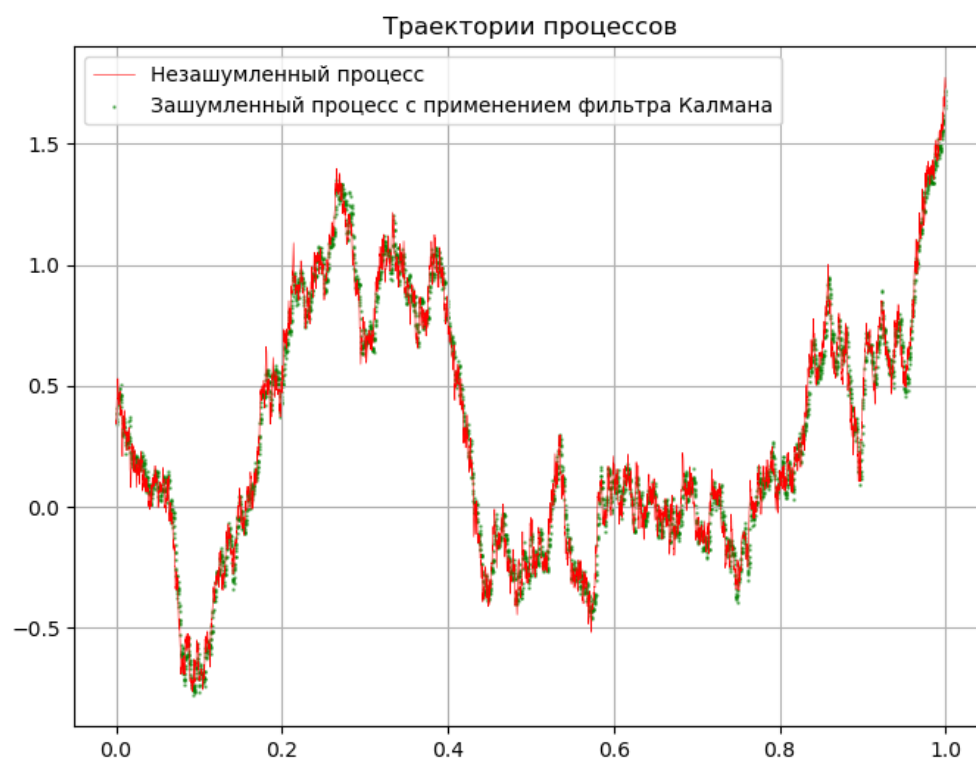


Рис. 34: Траектории процессов с применением фильтра Калмана



Рис. 35: Доверительный интервал распределения ошибки, уровень значимости $\alpha = 0.1$

Видим, что фильтр Калмана очень хорошо справляется с Гауссовым шумом. Доверительный интервал также получился

На следующих рисунках представлены незашумленный процесс, процесс с шумом Коши с параметром $\gamma = 0.2$ и процесс с применением фильтра Калмана. Также показан доверительный интервал распределения ошибки.

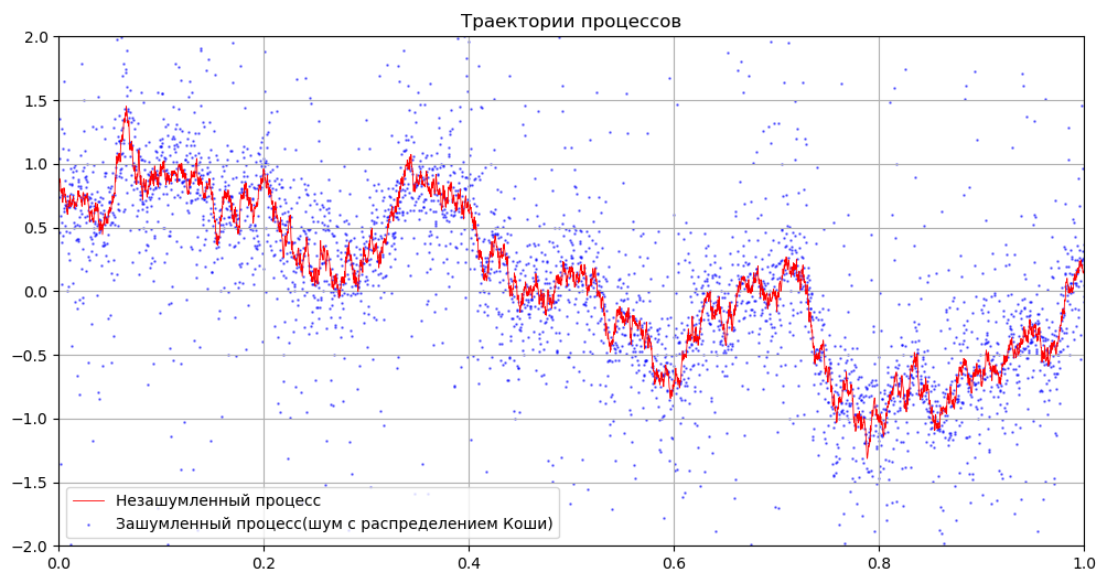


Рис. 36: Траектории процессов



Рис. 37: Траектории процессов с применением фильтра Калмана

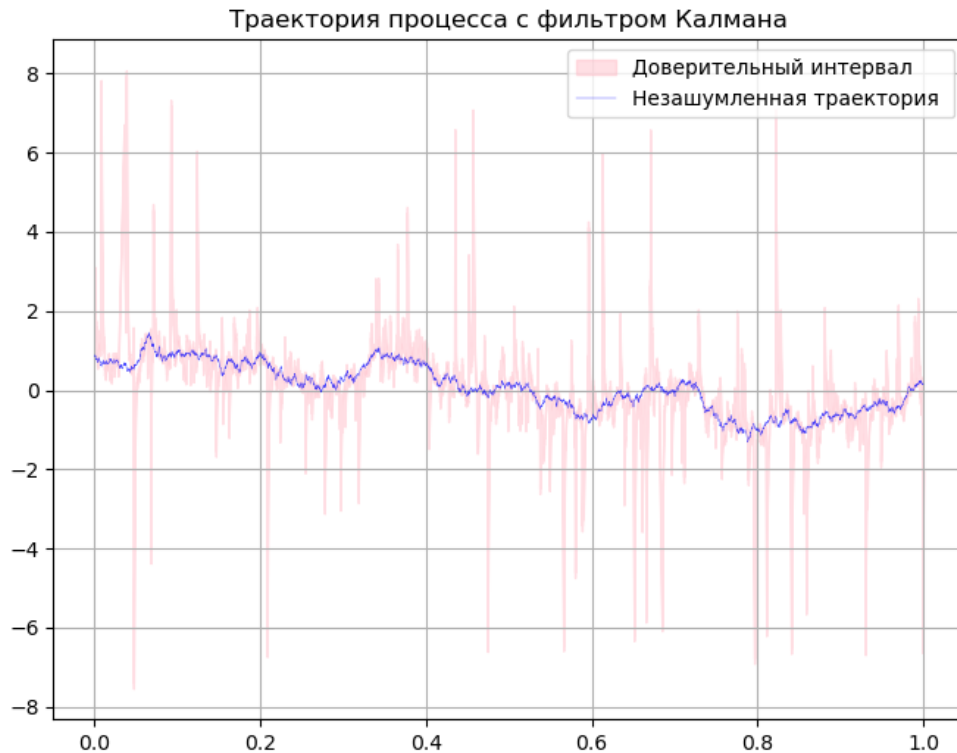


Рис. 38: Доверительный интервал распределения ошибки

Видим, что фильтр Калмана плохо справляется с шумом Коши. Доверительный интервал также отражает проблемы с данным шумом.

11 Задача 11

Построить двумерное пуассоновское поле, отвечающее сложному пуассоновскому процессу:

1. Система массового обслуживания. Первая координата поля — время поступления заявки в СМО (распределенное равномерно), а вторая — время обслуживания заявки (распределение χ^2 с десятью степенями свободы).

2. Система массового обслуживания с циклической интенсивностью $\lambda(1 + \cos(t))$ и единичными скачками. При помощи метода Льюиса и Шедлеара, свести задачу моделирования неоднородного пуассоновского процесса к моделированию двумерного пуассоновского поля, где первая координата распределена равномерно, а вторая имеет распределение Бернулли.

3. Работа страховой компании: Первая координата — момент наступления страхового случая (равномерное распределение), вторая — величина ущерба (распределение Парето). Поступление капитала считать линейным по времени со скоростью $c > 0$, начальный капитал $W > 0$.

11.1 Теоретическая часть

Рассматриваем первую координату поля на отрезке $[0, T]$. В теории пуассоновского поля λ - интенсивность поля (в нашем случае интенсивность поступления заявок = среднее число новых заявок в единицу времени). Тогда случайная считающая мера, равная числу заявок определяется следующим образом:

$$n \sim Pois(\lambda T).$$

По условию время поступления заявки распределено равномерно. Введём обозначение:

$$t_i \sim U[0, T], i = 1, 2, \dots, n.$$

После моделирования моментов поступления заявок необходимо упорядочить их по возрастанию. Пусть τ_i - время обслуживания заявки, по условию имеет распределение χ^2 с 10 степенями свободы. Вспомним важное свойство этого распределения:

$$\mathbb{E}\tau_i = 10.$$

В момент времени t_1 в обработку поступает первая заявка, следовательно, в момент времени $t_1 + \tau_1$ она будет обработана. Причём во время обработки первой заявки могут поступать другие заявки, которые образуют очередь и обрабатываются постепенно. Тогда для последующих заявок выполняется правило: если $(i - 1)$ -я заявка была обработана до момента поступления i -й заявки t_i , то она будет обработана в момент времени $t_i + \tau_i$, иначе она будет обработана через τ_i единиц после завершения обработки $(i - 1)$ заявки. Обозначим момент окончания обработки $(i - 1)$ заявки как t_i^{end} . Тогда можем посчитать количество заявок, находящихся в очереди в момент t :

$$Q(t) = \sum_{i=1}^n \mathbb{I}(t_i \leq t < t_i^{end})$$

Ввиду свойств рассматриваемого пуассоновского поля:

$$t_{i+1} - t_i \sim \text{Exp}(\lambda).$$

Тогда решение задачи можем разбить на 3 случая. Среднее время обслуживания одной заявки равно числу степеней свободы, то есть 10. Среднее время между поступлением заявок равно $\frac{1}{\lambda}$. Тогда 1) Если $\frac{1}{\lambda} > 10$, то заявки поступают медленнее, чем обрабатываются, СМО справляется с задачей. 2) Если $\frac{1}{\lambda} < 10$, то СМО не справляется с обработкой. 3) Если $\frac{1}{\lambda} = 10$, СМО находится в "равновесии".

Метод Льюиса и Шедлеара позволяет преобразовать неоднородный пуассоновский процесс в однородный. Для этого: 1) Определим верхнюю границу для интенсивности. В нашем случае максимальная интенсивность равна 2λ . 2) Генерируем пуассоновский процесс как в предыдущем пункте с постоянной интенсивностью 2λ . 3) Для каждого события, сгенерированного в однородном процессе, определяем, произошло ли оно на основе исходной функции интенсивности $\lambda(t)$. То есть удаляем t_i с вероятностью $1 - \frac{\lambda(t_i)}{2\lambda}$. Таким образом, как раз получим поле, где первая координата распределена равномерно, а вторая имеет распределение Бернулли.

Величина ущерба τ_i имеет распределение Парето с параметрами (a, b) . Знаем, что

$$\mathbb{E}\tau_i = \begin{cases} \frac{ab}{a-1}, & a > 1, \\ +\infty, & a \leq 1. \end{cases},$$

т.е. при $a \leq 1$ один страховой случай может привести к банкротству. Будем моделировать случайную величину из этого распределения методом обратных функций:

$$F_X(x) = 1 - \left(\frac{b}{x}\right)^a \implies F_X^{-1}(x) = \frac{b}{(1-x)^{1/a}}.$$

Тогда

$$X = \frac{b}{(1-Y)^{1/a}}, \quad Y \sim U[0, 1].$$

$1 - Y$ также является равномерно распределенной, значит

$$X = \frac{b}{Y^{1/a}}, \quad Y \sim U[0, 1].$$

Капитал страховой компании изменяется по правилу:

$$W(t) = W_0 + ct - \sum_{i=1}^{Q(t)} \tau_i,$$

где $Q(t)$ - число страховых случаев.

После того, как в какой то момент времени $W(t) = 0$, оставшийся вектор заполняется нулями. Алгоритм моделирования такой же как в случаях выше.

11.2 Результат работы программы

На следующих рисунках показаны различные системы, соответствующие первой части задания.



Рис. 39: Случай $\lambda = 0.5$ на отрезке времени $[0, 100]$. СМО не справляется с поступающими заявками.

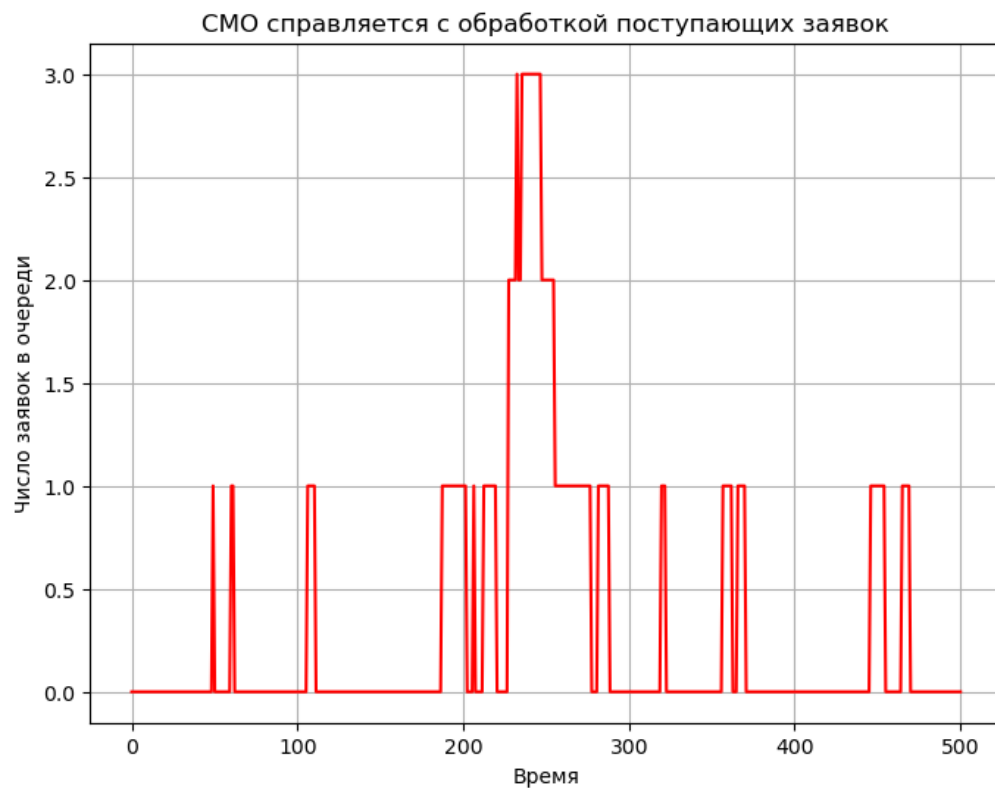


Рис. 40: Случай $\lambda = 0.07$ на отрезке времени $[0, 500]$. СМО справляется с поступающими заявками.

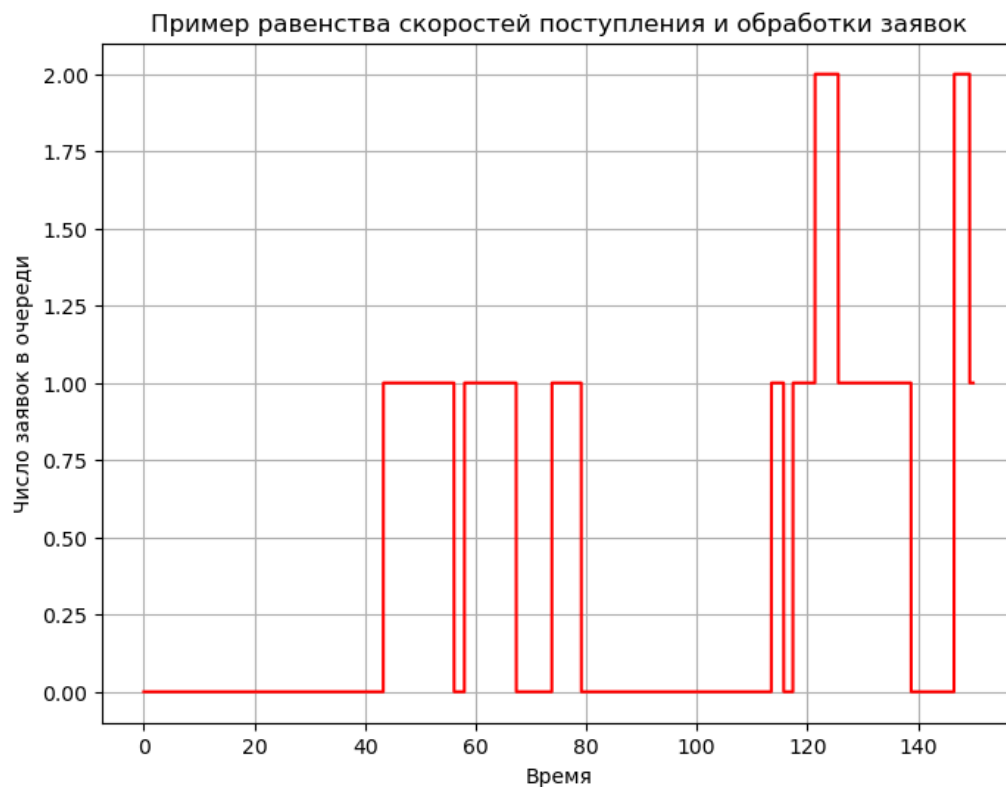


Рис. 41: Случай $\lambda = 0.1$ на отрезке времени $[0, 150]$. Пример равенства скоростей поступления и обработки заявок.

На следующей картинке показана система для второй части задания. Также изображена сама функция интенсивности поступления заявок.

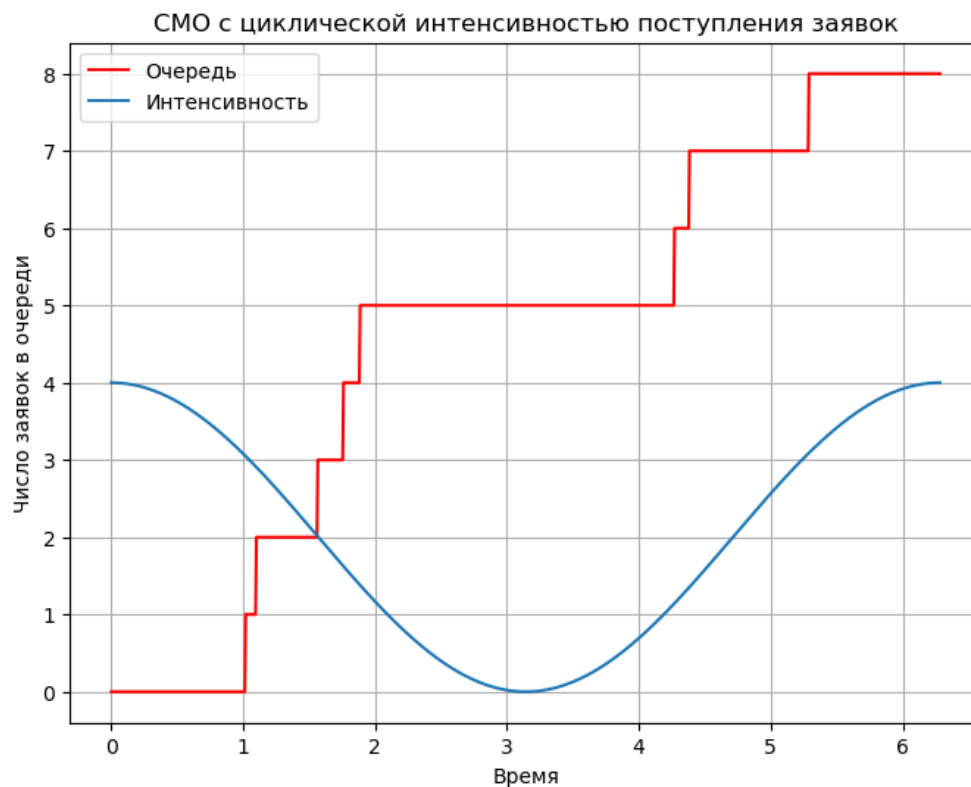


Рис. 42: СМО с циклической интенсивностью с параметром $\lambda = 2$ на отрезке времени $[0, 2\pi]$

На следующих картинках приведены системы для третьей части задания. Показаны различные случаи работы страховой компании, в том числе банкротства.



Рис. 43: Пример работы страховой компании на отрезке времени $[0, 50]$ с параметрами $\lambda = 1, a = 2, b = 2, c = 4$, начальный капитал равен 100

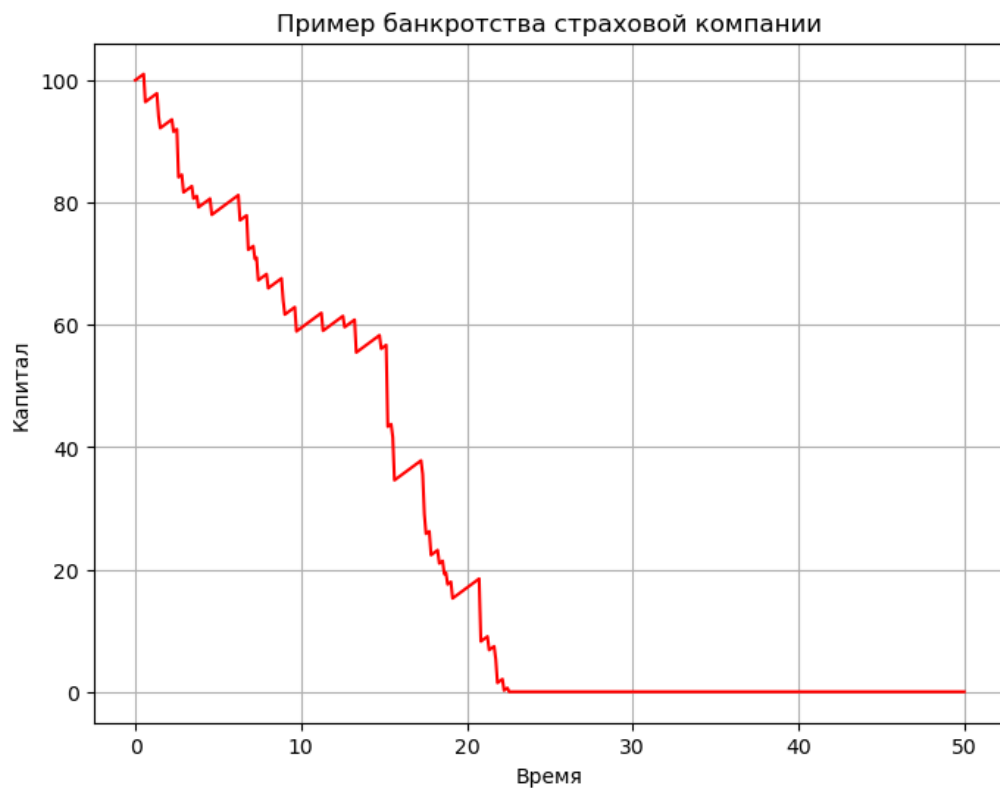


Рис. 44: Пример банкротства на отрезке времени $[0, 50]$ с параметрами $\lambda = 2, a = 2, b = 2, c = 2$, начальный капитал равен 100

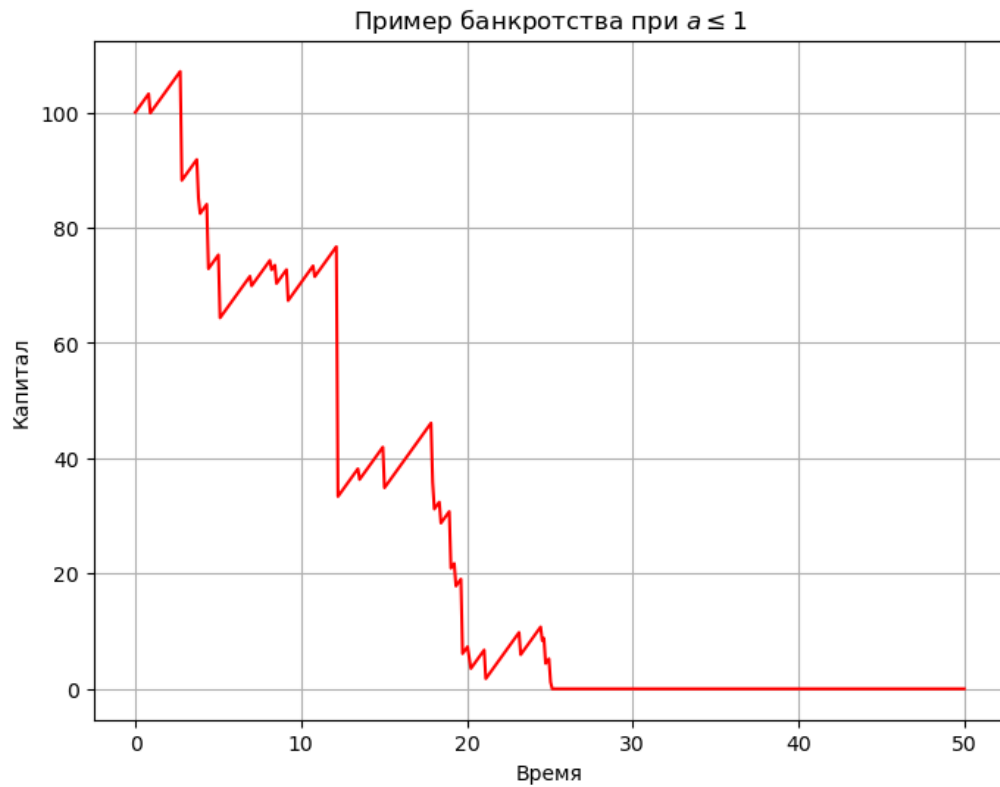


Рис. 45: Пример банкротства на отрезке времени $[0, 50]$ с параметрами $\lambda = 1, a = 0.9, b = 2, c = 4$, начальный капитал равен 100

В первом случае банкротства страховая компания совсем не справлялась с ущербом. Во втором случае банкротства страховая компания справлялась со страховыми случаями, но отдельные крупные страховые случаи обанкротили ее.

Список литературы

- [1] Ширяев А. Н. *Вероятность — 1*. МЦНМО, 2004.
- [2] Ульянов В. В. *Лекции по теории вероятностей и математической статистике*, 2022.
- [3] Смирнов С. Н. *Лекции по стохастическому анализу и моделированию*, 2024.