



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

Отчет по практикуму

«Стохастический анализ и моделирование»

Студент 415 группы

И. Р. Удовиченко

Руководитель практикума

к.ф.-м.н., доцент С. Н. Смирнов

Москва, 2020

Содержание

| | | |
|----------|---------------------------------------|-----------|
| 1 | Задание 1 | 4 |
| 1.1 | Постановка задачи | 4 |
| 1.2 | Теоретическая часть | 4 |
| 1.3 | Результаты работы программы | 5 |
| 2 | Задание 2 | 6 |
| 2.1 | Постановка задачи | 6 |
| 2.2 | Теоретическая часть | 7 |
| 2.3 | Результаты работы программы | 8 |
| 3 | Задание 3 | 9 |
| 3.1 | Постановка задачи | 9 |
| 3.2 | Теоретическая часть | 9 |
| 3.3 | Результаты работы программы | 12 |
| 4 | Задание 4 | 13 |
| 4.1 | Постановка задачи | 13 |
| 4.2 | Теоретическая часть | 13 |
| 4.3 | Результаты работы программы | 14 |
| 5 | Задание 5 | 15 |
| 5.1 | Постановка задачи | 15 |
| 5.2 | Теоретическая часть | 15 |
| 5.3 | Результаты работы программы | 17 |
| 6 | Задание 6 | 19 |
| 6.1 | Постановка задачи | 19 |
| 6.2 | Теоретическая часть | 19 |
| 6.3 | Результаты работы программы | 21 |
| 7 | Задание 7 | 21 |
| 7.1 | Постановка задачи | 21 |
| 7.2 | Теоретическая часть | 21 |
| 7.3 | Результаты работы программы | 23 |
| 8 | Задание 8 | 23 |
| 8.1 | Постановка задачи | 23 |
| 8.2 | Теоретическая часть | 24 |
| 8.3 | Результаты работы программы | 25 |

| | | |
|-----------|---------------------------------------|-----------|
| 9 | Задание 9 | 26 |
| 9.1 | Постановка задачи | 26 |
| 9.2 | Теоретическая часть | 26 |
| 9.3 | Результаты работы программы | 28 |
| 10 | Задание 10 | 28 |
| 10.1 | Постановка задачи | 28 |
| 10.2 | Теоретическая часть | 29 |
| 10.3 | Результаты работы программы | 30 |
| 11 | Задание 11 | 31 |
| 11.1 | Постановка задачи | 31 |
| 11.2 | Теоретическая часть | 32 |
| 11.3 | Результаты работы программы | 33 |

1. Задание 1

1.1. Постановка задачи

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. Рассмотреть игру в орлянку — бесконечную последовательность независимых испытаний с бросанием правильной монеты. Выигрыш S_n определяется как сумма по всем n испытаниям значений $+1$ и -1 в зависимости от выпавшей стороны. Проиллюстрировать (в виде ломаной) поведение нормированной суммы $Y(i) = S_i/\sqrt{n}$, как функцию от номера испытания $i = 1, \dots, n$ для одной отдельно взятой траектории. Дать теоретическую оценку для $Y(n)$ при $n \rightarrow \infty$.

1.2. Теоретическая часть

Пусть дана случайная величина $X \sim U[0, 1]$, рассмотрим случайную величину $Y = \mathbb{1}(X < p)$. Тогда Y принимает значение 1 с вероятностью p . Это следует из свойства: $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

Определение 1 *Биномиальным распределением случайной величины с параметрами n, p будем называть случайную величину, принимающую значения, равные числу успехов в n испытаниях Бернулли с вероятностью успеха p .*

Из определения следует способ генерирования биномиальной случайной величины:

$$X = \sum_{i=1}^n Y_i,$$

где случайные величины Y_i имеют распределение Бернулли с вероятностью успеха p .

Определение 2 *Геометрическим распределением с параметром p ($q = 1 - p$) будем называть распределение числа неудач до 1-го успеха в схеме Бернулли.*

Функция вероятности геометрического распределения имеет вид: $\mathbb{P}(X = n) = q^n p$, $n \in \mathbb{N}_0$.

Пусть случайная величина $Y \sim \exp(\lambda)$. Тогда

$$\mathbb{P}(n \leq Y < n+1) = e^{-\lambda n} (1 - e^{-\lambda}),$$

$$X = \lfloor Y \rfloor \sim \text{geom}(1 - e^{-\lambda}).$$

Тогда чтобы получить геометрическое распределение с параметром p надо взять целую часть от экспоненциально распределенной случайной величины с параметром $\lambda = -\ln(1-p)$. Экспоненциальное распределение получим обращением функции распределения: если $Z \sim U[0, 1]$, тогда $Y = -\frac{1}{\lambda} \ln(1-Z) \sim \exp(\lambda)$. Окончательно получаем $X = \left\lfloor \frac{\ln(1-Z)}{\ln(1-p)} \right\rfloor \sim \text{geom}(p)$.

Утверждение 1 Для $X \sim \text{geom}(p)$ верно свойство отсутствия памяти: $\mathbb{P}(X \geq m+n \mid X \geq m) = \mathbb{P}(X \geq n)$.

Доказательство.

$$\begin{aligned} \mathbb{P}(X \geq m+n \mid X \geq m) &= \frac{\mathbb{P}(X \geq m+n)}{\mathbb{P}(X \geq m)} = \\ &= \frac{q^{m+n} p / (1-q)}{q^m p / (1-q)} = q^n = \sum_{i=n}^{\infty} q^i p = \mathbb{P}(X \geq n). \end{aligned}$$

■

Рассмотрим игру «орлянку» со случайной величиной X :

$$X_i = \begin{cases} 1 & \text{с вероятностью } p = 0,5, \\ -1 & \text{с вероятностью } 1-p = 0,5. \end{cases}$$

Пусть $S_n = \sum_{i=1}^n X_i$ и $Y = S_n / \sqrt{n}$. Так как $\mathbb{E}S_n = 0$ и $\text{var}[S_n] = n$, то из центральной предельной теоремы следует, что $Y = S_n / \sqrt{n} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)$.

1.3. Результаты работы программы

Для проверки свойства отсутствия памяти нанесем на график эмпирическую функцию распределения, построенную по всей выборке, и эмпирическую функцию для условного распределения случайной величины, превышающей порог $\theta = 5$.

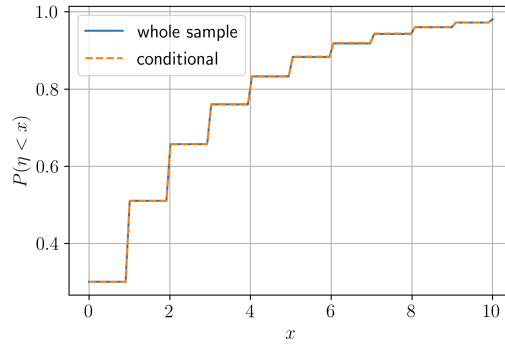


Рис. 1: Свойство отсутствия памяти для геометрического распределения.

На графике отчетливо видно наложение двух эмпирических функций распределения.

Для 3 части задания одна из траекторий для игры в орлянку:

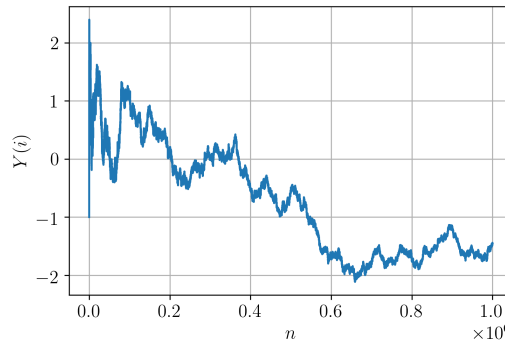


Рис. 2: Одна из возможных траекторий $Y(i)$ при игре в орлянку.

2. Задание 2

2.1. Постановка задачи

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 проверить свойство симметричности относительно 0.5 и самоподобия относительно деления деления на 3 с помощью критерия Смирнова.

3. Вычислить значение математического ожидания и дисперсии для данного распределения. Сравнить теоретические и эмпирические значения, проиллюстрировать сходимость.

2.2. Теоретическая часть

Определение 3 Канторовой лестницей называется функция f , которая строится следующим образом:

1. $f(0) = 0$, $f(1) = 1$.
2. Отрезок $[0, 1]$ разбивается на 3 равных части, затем на средней части функция f полагается равной полусумме значений на концах.
3. для первого и третьего отрезка процедура повторяется рекурсивно.

По определению канторова лестница обладает свойством фрактальности или самоподобия: $f(x) = 2f(x/3)$. Так как f монотонна и лежит в отрезке $[0, 1]$, то существует случайная величина X (сингулярная), функция распределения которой является канторовой лестницей.

Используя свойство фрактальности данной случайной величины, найдем ее математическое ожидание и дисперсию:

$$X \stackrel{d}{=} 1 - X \implies \mathbb{E}X = 0.$$

$$\mathbb{E}X^2 = \int_0^1 x^2 dF(x) = \int_0^{\frac{1}{3}} x^2 dF(x) + \int_{\frac{2}{3}}^1 x^2 dF(x).$$

Отсюда используя замену $y = 3x$ и свойство $F(x/3) = F(x)/2$ получаем

$$\text{var}[X] = \mathbb{E}X^2 = \frac{3}{8}.$$

Для проверки свойств распределения будем использовать критерии Колмогорова и Смирнова. В тесте Колмогорова проверяется гипотеза $H_0: F = F_0$ соответствия распределения некоторому наперед заданному распределению F_0 . Для проверки гипотезы строится следующая статистика:

$$D_n = \sup_x |F_n(x) - F_0(x)|,$$

где $F_n(x)$ — эмпирическая функция распределения, построенная по выборке X_1, \dots, X_n . Теорема Колмогорова утверждает, что $\sqrt{n}D_n \xrightarrow[n \rightarrow \infty]{d} K$, где K — распределение Колмогорова. Зафиксируем некоторый уровень доверия α и будем отклонять гипотезу, когда

$$\sqrt{n}D_n > K^{-1}(1 - \alpha).$$

Для проверки гипотезы о принадлежности двух выборок размеров m и n к одному распределению используется следующая статистика:

$$D_{mn} = \sup_x |F_n(x) - G_m(x)|,$$

где F_n и G_m — эмпирические функции распределения. По теореме Смирнова

$$\sqrt{\frac{mn}{m+n}} D_{mn} \xrightarrow[n \rightarrow \infty]{d} K,$$

поэтому при больших размерах выборки ($m, n > 20$) отклоняем гипотезу, если

$$\sqrt{\frac{mn}{m+n}} D_{mn} > K^{-1}(1 - \alpha).$$

2.3. Результаты работы программы

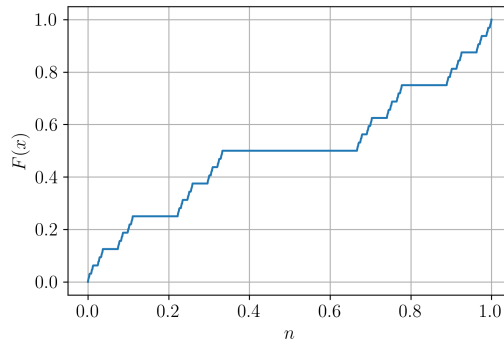


Рис. 3: Функция распределения канторовой лестницы.

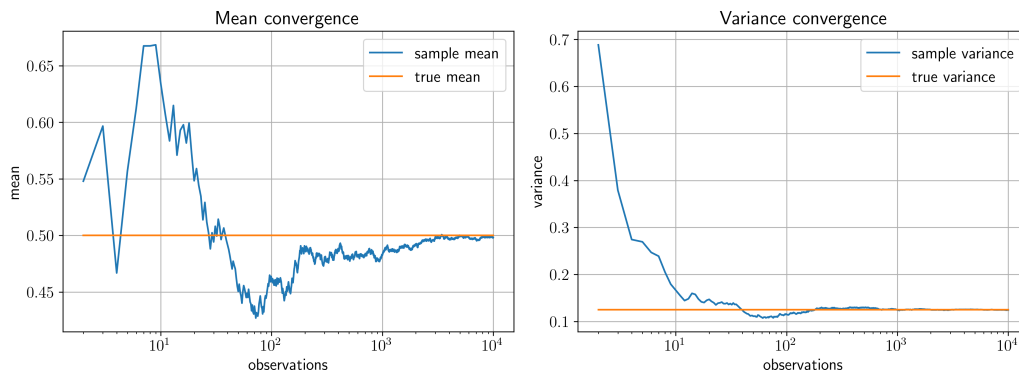


Рис. 4: Иллюстрация сходимости статистик сингулярного распределения.

3. Задание 3

3.1. Постановка задачи

1. Построить датчик экспоненциального распределения. Проверить для данного распределения свойство отсутствия памяти. Пусть X_1, \dots, X_n независимы и имеют экспоненциальное распределение с параметрами $\lambda_1, \dots, \lambda_n$. Найти распределение случайной величины $Y = \min(X_1, \dots, X_n)$.
2. На основе датчика экспоненциального распределения построить датчик пуассоновского распределения.
3. Построить датчик пуассоновского распределения как предел биномиального. С помощью критерия хи-квадрат убедиться, что получен датчик распределения Пуассона.
4. Построить датчик стандартного нормального распределения методом моделирования случайных величин с переходом в полярные координаты. Проверить при помощи t-критерия Стьюдента равенство матожиданий, а при помощи критерия Фишера равенство дисперсий.

3.2. Теоретическая часть

Определение 4 Экспоненциальным распределением с параметром λ будем называть распределение с плотностью $p(x) = \lambda e^{-\lambda x}$.

Утверждение 2 Экспоненциальное распределение обладает свойством отсутствия памяти:

$$\mathbb{P}(X \geq x + y \mid X \geq y) = \mathbb{P}(X \geq x).$$

Доказательство.

$$\mathbb{P}(X \geq x + y \mid X \geq y) = \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x} = \mathbb{P}(X \geq x).$$

■

Рассмотрим еще одно свойство экспоненциального распределения. Пусть X_1, \dots, X_n независимы и распределены экспоненциально с показателями $\lambda_1, \dots, \lambda_n$ соответственно. Пусть также $Y = \min \{X_1, \dots, X_n\}$. Тогда

$$F_Y(x) = \mathbb{P}(Y < x) = 1 - \mathbb{P}(X_1 \geq x, \dots, X_n \geq x) = 1 - \prod_{i=1}^n e^{-\lambda_i x} = 1 - e^{-x \sum_{i=1}^n \lambda_i}.$$

Таким образом, Y распределена экспоненциально с показателем $\sum_{i=1}^n \lambda_i$.

Рассмотрим два способа моделирования распределения Пуассона. Первый заключается в том, что промежутки между скачками пуассоновского процесса с параметром λ распределены экспоненциально, а случайная величина $X(1)$ имеет распределение Пуассона с параметром λ .

Второй способ дает

Теорема 1 (Пуассона) Пусть в схеме серий испытаний Бернулли с вероятностями p_n выполнено:

$$p_n \xrightarrow{n \rightarrow \infty} 0, \quad np_n \xrightarrow{n \rightarrow \infty} \lambda > 0.$$

Тогда число успехов сходится к распределению Пуассона с параметром λ . При этом верна оценка:

$$\left| \mathbb{P}(S_n \in A) - \sum_{k \in A} e^{-\lambda} \frac{\lambda^k}{k!} \right| \leq \lambda p.$$

Чтобы получить вероятность, отличающуюся от распределения Пуассона не больше чем на ε , положим $p = \frac{\varepsilon}{\lambda}$ и $n = \frac{\lambda^2}{\varepsilon}$. Тогда биномиальное распределение с такими параметрами будет хорошо приближать распределение Пуассона.

Для моделирования стандартного нормального распределения рассмотрим пару случайных величин (X_1, X_2) , которую будем рассматривать, как двумерный вектор. Пусть (r, φ) — запись вектора в полярных координатах. Из симметрии $\varphi \sim U[0, 2\pi]$. Из свойств нормального распределения $r^2 \sim \exp(\frac{1}{2})$. Тогда моделируя угол равномерным распределением и квадрат радиуса экспоненциальным, будем получать 2 стандартных нормальных случайных величины после перехода в декартовы координаты.

Критерий χ^2 Пирсона используется для проверки гипотезы равенства распределения заданному $H_0: F = F_0$. Разобьем прямую на непересекающиеся части $\Delta_1, \dots, \Delta_m$ и положим $p_i = \mathbb{P}(X \in \Delta_i \mid H_0)$. Обозначим за N_i число элементов выборки, попавших в Δ_i . Тогда статистика

$$Q = \sum_{i=1}^m \frac{(N_i - np_i)^2}{np_i} \sim \chi_{m-1}^2.$$

Тогда отклоняем гипотезу, если $\chi_{m-1}^2(Q) < \alpha$.

Пусть дана выборка X_1, \dots, X_n из нормального распределения с неизвестными параметрами. Для проверки гипотезы $H_0: \mathbb{E}X = \mu$ используется критерий Стьюдента. Положим

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Тогда статистика

$$U = \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \sim St_{n-1}.$$

Будем отклонять гипотезу, если $St_{n-1}(U) < \frac{\alpha}{2}$.

Для сравнения дисперсий двух нормально распределенных выборок используется критерий Фишера. Пусть даны две выборки X_1, \dots, X_n и Y_1, \dots, Y_m . Тогда

$$V = \frac{(m-1) \sum_{i=1}^m (X_i - \bar{X})^2}{(n-1) \sum_{i=1}^n (Y_i - \bar{Y})^2} \sim F_{m-1, n-1}.$$

Гипотеза отвергается, если $F_{m-1, n-1}(V) < \frac{\alpha}{2}$.

3.3. Результаты работы программы

Иллюстрацию свойства отсутствия памяти проведем аналогично с 1-м заданием:

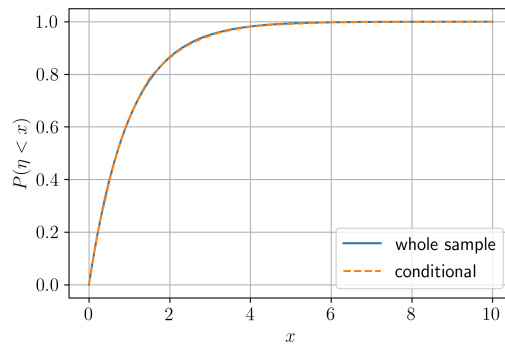


Рис. 5: Свойство отсутствия памяти для экспоненциального распределения.

Распределения минимума случайных величин проиллюстрируем, наложив на один график эмпирическую и теоретическую функции распределения (размер выборки 1000 для каждой из 100 случайных величин):

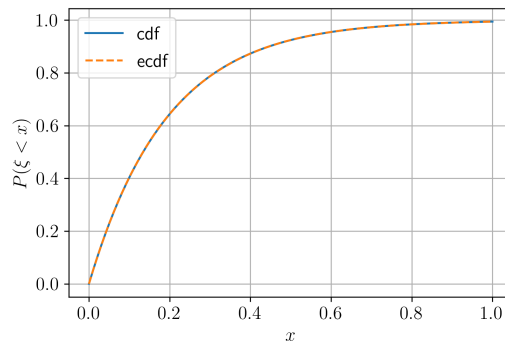


Рис. 6: Распределение минимума.

При проверке гипотез были получены следующие результаты (выборка генерировалась 1000 раз):

| Критерий | Частота отклонения | Уровень значимости |
|------------|--------------------|--------------------|
| Хи-квадрат | 3,0% | 0,95 |
| Стьюдента | 3,0% | 0,95 |
| Фишера | 6,0% | 0,95 |

4. Задание 4

4.1. Постановка задачи

1. Построить датчик распределения Коши.
2. На основе датчика распределения Коши с помощью метода фон Неймана построить датчик стандартного нормального распределения. При помощи `norm probability plot` убедиться в корректности построенного датчика и обосновать наблюдаемую линейную зависимость.
3. Сравнить скорость моделирования стандартного нормального распределения в заданиях 3 и 4.

4.2. Теоретическая часть

Определение 5 *Случайная величина ξ имеет распределение Коши с параметрами a, b ($\xi \sim C(a, b)$), если ее функция распределения имеет вид:*

$$F_{\xi}(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \left(\frac{x - a}{b} \right).$$

Для моделирования распределения Коши, воспользуемся методом обращения функции распределения:

$$F^{-1}(y) = a + b \operatorname{tg} \left(\pi \left(y - \frac{1}{2} \right) \right).$$

Метод элиминации фон Неймана подробно описан в [1]. В нашем случае:

$$p(x) = e^{-\frac{x^2}{2}}, \quad q(x) = \frac{b}{\pi [(x - a)^2 + b^2]}.$$

Наилучшие результаты в смысле частоты отклонения дают параметры $a = 0$, $b = 1$. Тогда:

$$k = \sup_x \frac{p(x)}{q(x)} = \frac{2}{\sqrt{e}}.$$

Затем будем генерировать случайные величины $\xi_n \sim Q$. Каждое значение в полученной выборке будем отбирать в конечную выборку с вероятностью $p(\xi_k)/kq(\xi_k)$.

Для тестирования полученного датчика воспользуемся norm probability plot. По определению для norm probability plot ось ординат деформируется таким образом, чтобы функция распределения для стандартной гауссовской случайной величины была линейна. Тогда если дана выборка из распределения с параметрами μ, σ , то это эквивалентно сдвигу графика стандартного распределения на μ вправо вдоль оси абсцисс и уменьшению углового коэффициента в σ раз.

4.3. Результаты работы программы

Результат работы алгоритма фон Неймана:

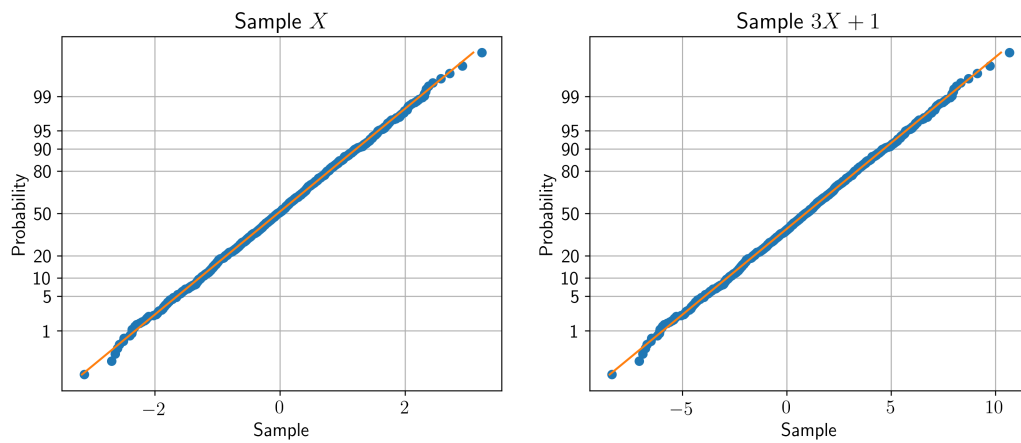


Рис. 7: Normal probability plot.

Для сравнения скорости работы алгоритмов были сгенерированы выборки разного размера. Для каждого размера выборка генерировалась 10 раз, и в качестве итогового времени бралась медиана.

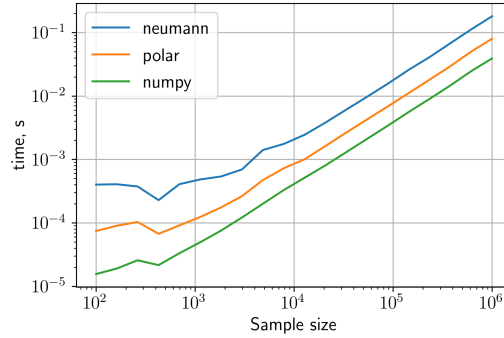


Рис. 8: Сравнение скорости работы алгоритмов генерирования выборки из нормального распределения.

5. Задание 5

5.1. Постановка задачи

1. Пусть $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Убедиться эмпирически в справедливости ЗБЧ и ЦПТ: исследовать поведение суммы S_n и эмпирического распределения величины

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

2. Считая μ и σ неизвестными, для пункта 1 построить доверительные интервалы для среднего и дисперсии.
3. Пусть $X_i \sim \mathcal{C}(a, b)$ имеет распределение Коши со сдвигом a и масштабом b . Проверить эмпирически, как ведут себя суммы $\frac{S_n}{n}$. Результат объяснить, а также найти распределения данных сумм.

5.2. Теоретическая часть

Сформулируем закон больших чисел и центральную предельную теорему.

Теорема 2 (УЗБЧ в форме Колмогорова) Пусть X_1, X_2, \dots — последовательность независимых одинаково распределенных случайных величин, имеющих конечные первые моменты. Тогда

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{n.n.} \mathbb{E}X_1.$$

Теорема 3 (ЦПТ) Пусть X_1, X_2, \dots — последовательность независимых одинаково распределенных случайных величин с конечным вторым моментом. Тогда

$$\sqrt{n} \left(\frac{S_n}{n} - \mu \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2),$$

где

$$S_n = \sum_{i=1}^n X_i, \quad \mu = \mathbb{E}X_1, \quad \sigma^2 = \text{var } X_1.$$

Доказательства представлены в [2].

Рассмотрим статистику

$$U = \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}}, \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

имеет [3] распределение Стьюдента с $n-1$ степенью свободы, можно построить доверительный интервал с уровнем доверия α следующим образом:

$$\mathbb{P}(\gamma_1 < U < \gamma_2) = \mathbb{P}\left(\bar{X} - \frac{\hat{\sigma}\gamma_2}{\sqrt{n}} < \mu < \bar{X} - \frac{\hat{\sigma}\gamma_1}{\sqrt{n}}\right) = \alpha.$$

Значения γ_1 и γ_2 выберем так, чтобы интервал был симметричным:

$$\gamma_1 = St_{n-1}^{-1}\left(\frac{1-\alpha}{2}\right), \quad \gamma_2 = St_{n-1}^{-1}\left(\frac{1+\alpha}{2}\right).$$

Итак, доверительный интервал для параметра μ имеет вид

$$\left(\bar{X} - \frac{\hat{\sigma}\gamma_2}{\sqrt{n}}, \bar{X} - \frac{\hat{\sigma}\gamma_1}{\sqrt{n}}\right).$$

Для оценки дисперсии будем использовать статистику

$$V = \frac{\sum_i (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1).$$

$$\mathbb{P}(\gamma_1 < V < \gamma_2) = \mathbb{P}\left(\frac{\sum_i (X_i - \bar{X})^2}{\gamma_2} < \sigma^2 < \frac{\sum_i (X_i - \bar{X})^2}{\gamma_1}\right) = \alpha.$$

Для построения симметричного интервала положим

$$\gamma_1 = (\chi_{n-1}^2)^{-1}\left(\frac{1-\alpha}{2}\right), \quad \gamma_2 = (\chi_{n-1}^2)^{-1}\left(\frac{1+\alpha}{2}\right).$$

Доверительный для σ^2 с уровнем доверия α имеет вид

$$\left(\frac{\sum_i (X_i - \bar{X})^2}{\gamma_2}, \frac{\sum_i (X_i - \bar{X})^2}{\gamma_1} \right).$$

Для оценки распределения суммы случайных величин, распределенных по Коши, воспользуемся тем фактом, что распределение является устойчивым с показателем 1. Отсюда следует, что если $X_i \sim \mathcal{C}(a, b)$, то $\sum_{i=1}^n X_i \sim \mathcal{C}(a, nb)$.

5.3. Результаты работы программы

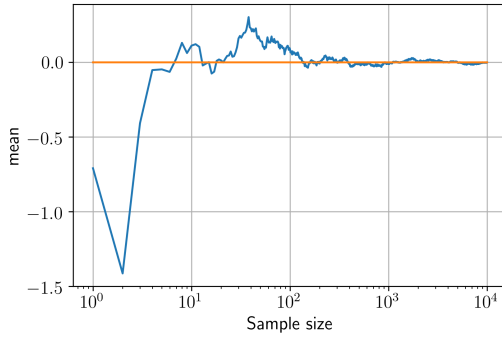


Рис. 9: Иллюстрация сходимости в ЗБЧ.

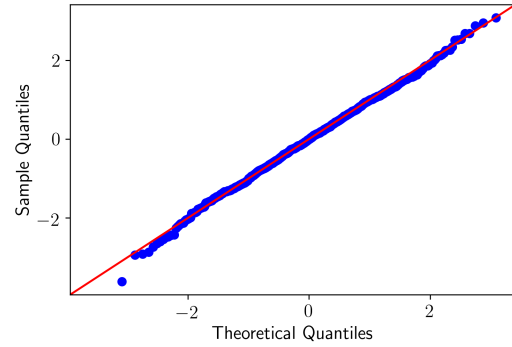


Рис. 10: Распределение нормированной суммы в ЦПТ при $n = 1000$ (Q-Q plot).

На рис. 9 проиллюстрирована сходимость частичных сумм. Предельное поведение частичных сумм видно из графика на рис. 10.

На следующих двух рисунках построены графики с доверительными интервалами для математического ожидания и дисперсии выборки с уровнем значимости $\alpha = 0,95$.

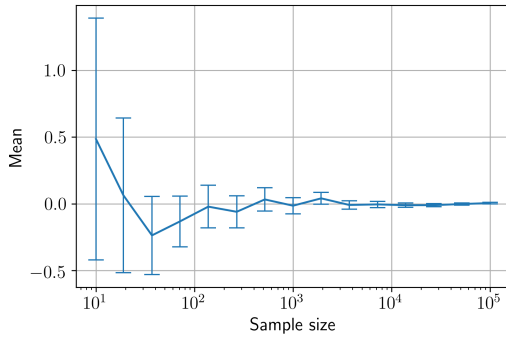


Рис. 11: Доверительные интервалы для математического ожидания.

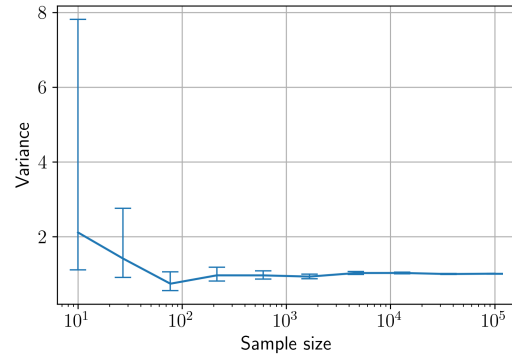


Рис. 12: Доверительные интервалы для дисперсии.

Следующие 2 графика иллюстрируют поведение выборочных средних для распределения Коши. На рис. 13 показано типичное поведение выборочного среднего в зависимости от числа элементов выборки. На рис. 14 нанесен Q-Q plot для квантилей выборки и теоретических квантилей распределения Коши.

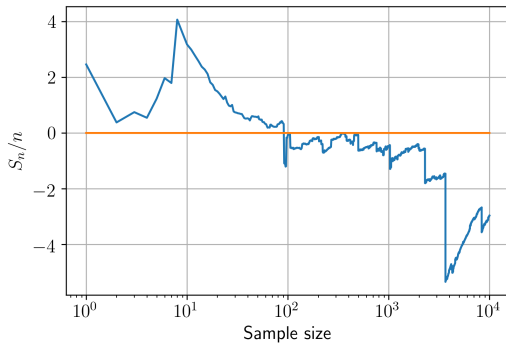


Рис. 13: Выборочное среднее S_n/n в зависимости от n .

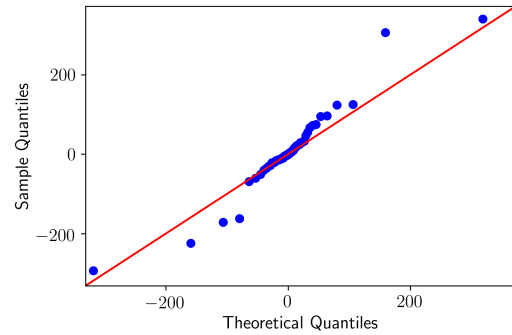


Рис. 14: Распределение выборочного среднего. 1000 выборок по 1000 испытаний. Q-Q plot для распределения Коши.

6. Задание 6

6.1. Постановка задачи

1. Посчитать интеграл

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

- методом Монте-Карло;
- методом квадратур, сводя задачу к вычислению собственного интеграла Римана.

2. Для каждого случая оценить точность вычислений.

6.2. Теоретическая часть

Рассмотрим вычисление интеграла методом Монте-Карло. Положим

$$g(\vec{x}) = \pi^5 \frac{\exp\left\{-\frac{1}{128 x_1^2 \dots x_{10}^2}\right\}}{x_1^2 \dots x_{10}^2}.$$

Тогда интеграл можно записать в виде:

$$J = \mathbb{E}g(\vec{X}),$$

где \vec{X} имеет 10-мерное нормальное распределение с нулевым средним и ковариационной матрицей $\frac{1}{2}I$. Данное математическое ожидание будем вычислять как выборочное среднее $\hat{J} = \frac{1}{n} \sum_{i=1}^n g(\vec{X}_i)$.

Так как вычислить распределение $g(\vec{X})$ не представляется возможным, для оценки точности метода Монте-Карло будем использовать неравенство Чебышева:

$$\mathbb{P}\left(\left|J - \hat{J}\right| \geq \varepsilon\right) \leq \frac{\text{var } g(\vec{X}_i)}{n\varepsilon^2} = 1 - \alpha.$$

Дисперсию будем аппроксимировать несмещенной оценкой:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(g(\vec{X}_i) - \overline{g(\vec{X}_i)}\right)^2.$$

Тогда выражение для точности:

$$\varepsilon = \frac{\hat{\sigma}}{\sqrt{n(1-\alpha)}}.$$

Для вычисления интеграла методом квадратур сделаем замену:

$$x_k = \operatorname{tg}(y_k), \quad dx_k = \frac{dy_k}{\cos^2(y_k)}, \quad k = 1, \dots, 10.$$

Тогда интеграл принимает вид:

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \dots \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \exp \left\{ - \sum_{k=1}^{10} \operatorname{tg}^2(y_k) - \frac{1}{128} \prod_{k=1}^{10} \operatorname{ctg}^2(y_k) \right\} \prod_{k=1}^{10} \frac{1}{\sin^2(y_k)} dy_1 \dots dy_{10}$$

Данный интеграл является собственным и может быть вычислен методом квадратур. Заметим, что функция является четной по каждой переменной, поэтому достаточно вычислить значение интеграла в положительном ортанте, а затем результат умножить на 2^{10} . Так же функция значение функции не меняется от перестановки переменных, поэтому для вычисления интеграла достаточно упорядочить элементы равномерной сетки в лексикографическом порядке, затем каждую точку взять с весом

$$\frac{10!}{r_1! \dots r_n!},$$

где r_i — число раз, которое повторяется i -й по порядку элемент сетки в текущем наборе из 10 точек. Для численной устойчивости рекомендуется вычислять $n!/m!$ как $\exp \{ \ln \Gamma(1+n) - \ln \Gamma(1+m) \}$.

Для оценки погрешности будем использовать разложение в ряд Тейлора с остаточным членом в форме Лагранжа функции g :

$$g(y) = g(y_0) + \langle \nabla g(\xi), \Delta y \rangle,$$

где y_0 центр гиперкуба Δ , а $\xi \in \Delta$. Интегрируя обе части по Δ , получим, что ошибка интегрирования по каждому гиперкубу равна

$$\int_{\Delta} \langle \nabla g(\xi), \Delta y \rangle dy \leq \sup_{\Delta} \|\nabla g(x)\| \cdot \|\Delta y\| \cdot \mu(\Delta) \leq \frac{\sqrt{n}}{2} h^{n+1} \sup_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \|\nabla g(x)\|,$$

а общая ошибка не превосходит $\frac{\pi^n h \sqrt{n}}{2} \sup_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \|\nabla f(x)\|$. Однако даже при

вычислении 10-мерного интеграла шаг сетки получается таким малым, что вычисление с таким шагом не представляется возможным.

6.3. Результаты работы программы

При вычислении интеграла методом Монте-Карло был получен результат $\hat{J} = 124,83$, и оценка погрешности $\varepsilon = 0,05$. Число испытаний 10^{10} . Методом квадратур был получен результат $J \approx 124,81$ на сетке размером 30 по каждому измерению.

7. Задание 7

7.1. Постановка задачи

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

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

$f(0, 0) = 0$ 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. Сравнить результаты со стандартными методами оптимизации.

7.2. Теоретическая часть

Для поиска минимума функции методом случайного поиска будем генерировать независимые равномерно распределенные в A случайные величины. Затем в полученной выборке найдем ту реализацию, на которой значение функции минимально.

Для генерирования равномерного распределения в круге найдем распределение полярных координат равномерно распределенных точек:

$$\mathbb{P}(X \in M) = \frac{1}{\pi} \iint_M dx dy = \frac{1}{\pi} \iint_M r dr d\varphi = \iint_M dr^2 d\left(\frac{\varphi}{2\pi}\right).$$

Отсюда следует, что радиус и угол независимы, угол распределен равномерно, а функция распределения радиуса имеет вид:

$$F_r(x) = x^2 [0 \leq x \leq 1].$$

Таким образом, случайная величина X с полярными координатами (r, φ) будет распределена равномерно в A , если $\varphi \sim U[0, 2\pi]$ и $\mathbb{P}(r < x) = F_r(x)$.

Оценим точность для метода случайного поиска. Потребуем, чтобы $|f(x) - f(x^*)| < \varepsilon$:

$$|f(x) - f(x^*)| \leq \|\nabla f(x^*)\| \cdot \|x - x^*\| \leq \max_{\|x\| \leq 1} \|\nabla f(x)\| \cdot \|x - x^*\| = \max_{\|x\| \leq 1} \|\nabla f(x)\| \delta$$

Оценим норму градиента:

$$\frac{\partial f}{\partial x_1} = 3x_1^2 \sin\left(\frac{1}{x_1}\right) - x_1 \cos\left(\frac{1}{x_1}\right) + 10x_2^4 \cos\left(\frac{1}{x_2}\right),$$

$$\frac{\partial f}{\partial x_2} = 40x_1x_2^3 \cos\left(\frac{1}{x_2}\right) + 10x_1x_2^2 \sin\left(\frac{1}{x_2}\right).$$

Тригонометрический функции и переменные оценим сверху по модулю числом 1. Тогда:

$$\max_{\|x\| \leq 1} \|\nabla f(x)\| \leq \sqrt{14^2 + 50^2} \leq 52$$

Таким образом

$$\delta = \frac{\varepsilon}{52}$$

Потребуем, чтобы с вероятностью не менее α хотя бы одна точка выборки попала в δ -окрестность точки минимума. Эквивалентно: с вероятностью не превосходящей $1 - \alpha$ ни одна точка не должна попасть в δ -окрестность искомой точки. Так как требуется получить верхнюю оценку, то, чтобы не учитывать границы, заменим окружность с радиусом δ на полуокружность:

$$\mathbb{P}(\forall x \in \mathbb{X}: x \notin B_\delta(x^*)) \leq \left(\frac{\pi - \pi\delta^2/2}{\pi}\right)^n = (1 - \delta^2/2)^n \leq 1 - \alpha.$$

Отсюда

$$\varepsilon = 52\delta = 52\sqrt{2 - 2^n(1 - \alpha)}$$

Описание метода имитации отжига:

1. Задается функция для минимизации $f(\cdot)$ начальная температура T_0 , начальное приближение x_0 , распределение шага по переменной \mathcal{P} и функцию понижения температуры $h(\cdot)$.
2. На каждой итерации цикла вычисляется следующее положение точки $x' = x_i + \xi$, где $\xi \sim \mathcal{P}$.
3. Если $f(x') < f(x_i)$, то $x_{i+1} = x'$, иначе $x_{i+1} = x'$ с вероятностью $p_i = \exp \left\{ \frac{f(x_i) - f(x')}{T_i} \right\}$ и $x_{i+1} = x_i$ с вероятностью $1 - p_i$.
4. Понижается температура $T_{i+1} = h(T_i)$.

7.3. Результаты работы программы

В результате работы программы для первой части задания была найдена точка минимума $(-0,367, 0,930)$. Размер выборки 10^6 , точность 0,13 с надежностью 0,95.

Для функции Розенброка методом имитации отжига был найден минимум $(1,003, 1,005)$. Было проведено 16111 итераций и 1642 раза была обновлена точка. В качестве распределения для шага было выбрано нормальное распределение с нулевым средним и стандартным отклонением 0,05. Начальное приближение — точка $(0, 0)$, температура понижалась по закону $T = T_0 k^i$, где $T_0 = 1$, $k = 0,999$, а i — номер итерации.

Стандартные методы нелинейной оптимизации в `scipy` лучше работали на функции Розенброка, но не справились с первой функцией из-за большого количества локальных экстремумов, расположенных крайне нерегулярно.

8. Задание 8

8.1. Постановка задачи

1. 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.

8.2. Теоретическая часть

Кратко опишем способ решения данной задачи, полное описание можно найти в [4]. Введем на плоскости равномерную сетку с шагом h и назовем внутренними узлы, для которых все 4 соседних узла лежат в круге. Узлы, лежащие в круге, но не удовлетворяющие этому условию, будем называть крайними.

Аппроксимируя вторые частные производные, можно свести задачу Дирихле к следующей разностной схеме:

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

для внутренних узлов, и

$$u_{ij} = f_{ij}$$

для крайних узлов.

Рассмотрим теперь следующую модель случайного блуждания:

- Частица начинает свой путь в одном из внутренних узлов x_{ij} .
- С равными вероятностями она переходит в один из четырех соседних узлов, пока частица не попадет на крайний узел x_{pq} .

Обозначим $p(x_{ij}, x_{pq})$ вероятность, начав путь в узле x_{ij} , закончить его в узле x_{pq} . Из формулы полной вероятности следует

$$p(x_{ij}, x_{pq}) = \frac{1}{4}(p(x_{i-1,j}, x_{pq}) + p(x_{i+1,j}, x_{pq}) + p(x_{i,j-1}, x_{pq}) + p(x_{i,j+1}, x_{pq})) \quad (1)$$

для внутренних узлов x_{ij} и $p(x_{ij}, x_{ij}) = 1$ для крайних узлов.

Рассмотрим теперь случайную величину $\xi_{ij} = f(x_{pq})$ — значение f в конечной точке пути. Так как $\mathbb{E}\xi_{ij} = \sum_{p,q} p(x_{ij}, x_{pq})f(x_{pq})$, используя формулу (1) связи вероятностей на соседних узлах, получим

$$\mathbb{E}\xi_{ij} = \frac{1}{4}(\mathbb{E}\xi_{i-1,j} + \mathbb{E}\xi_{i+1,j} + \mathbb{E}\xi_{i,j-1} + \mathbb{E}\xi_{i,j+1}) \text{ для внутренних узлов,}$$

$$\mathbb{E}\xi_{ij} = f_{ij} \text{ для крайних.}$$

Видно, что эти уравнения совпадают с разностной схемой, аппроксимирующей исходную задачу.

Значения $\mathbb{E}\xi_{ij}$ в каждом узле можно вычислить методом Монте–Карло, моделируя описанный выше случайных процесс.

Для оптимизации вычислений воспользуемся тем, что моделируемый процесс случайного блуждания обладает марковским свойством, поэтому при прохождении траектории через промежуточные узлы дальнейший путь частицы можно рассматривать как процесс, начавшийся в этом промежуточном узле.

Точным решением данной задачи является функция $f(x, y) = x^2 - y^2$.

8.3. Результаты работы программы

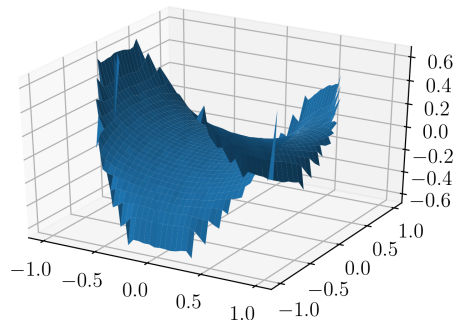


Рис. 15: Численное решение.

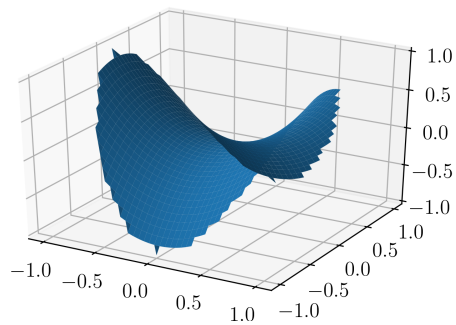


Рис. 16: Аналитическое решение.

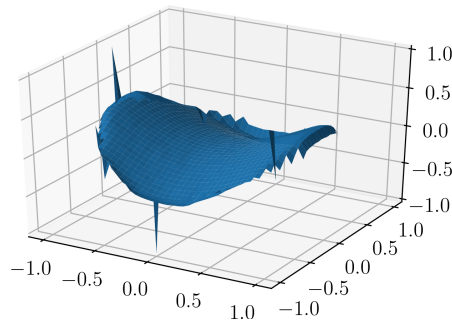


Рис. 17: Невязка

9. Задание 9

9.1. Постановка задачи

Рассмотреть два вида процессов:

- Винеровский процесс $W(t), t \in [0, 1], W(0) = 0$.
- Процесс Орнштейна–Уленбека $X(t), t \in [0, 1], X(0) = X_0$, то есть стационарный марковский гауссовский процесс. Начальное значение X_0 генерируется случайным образом так, чтобы полученный процесс бы 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 траекторий, не соединяя точки ломаной с целью получения визуально непрерывной линии.

9.2. Теоретическая часть

Опишем метод разбиения отрезка для моделирования случайного процесса.

1. Сгенерировать $X(0)$.
2. Сгенерировать $X(1) \sim \mathbb{P}(X(1) \mid X(0))$

3. Пусть случайный процесс уже сгенерирован в точках t_1 и t_3 . Тогда $X(t_2) \sim \mathbb{P}(X(t_2) \mid X(t_1), X(t_3))$.
4. Повторять, пока разбиение не получится достаточно мелким.

Найдем выражение для условных вероятностей Винеровского процесса. Пусть $t > s$, тогда

$$\begin{aligned} k(t, s) &= \mathbb{E}[W(t)W(s)] = \mathbb{E}[W(s)^2] + \mathbb{E}[W(s)(W(t) - W(s))] = \\ &= \mathbb{E}[W(s)^2] + \mathbb{E}[W(s)] \mathbb{E}[(W(t) - W(s))] = \sigma^2 \min(t, s) \end{aligned}$$

Для метода деления отрезка пополам найдем условное распределение $\mathbb{P}(W(t) = x_2 \mid W(t - \Delta t) = x_1, W(t + \Delta t) = x_3)$. По теореме Байеса и определению условной плотности имеем:

$$\begin{aligned} p(W(t) = x_2 \mid W(t - \Delta t) = x_1, W(t + \Delta t) = x_3) &= \\ &= \frac{p(W(t + \Delta t) = x_3 \mid W(t) = x_2) p(W(t) = x_2 \mid W(t - \Delta t) = x_1)}{p(W(t + \Delta t) = x_3 \mid W(t - \Delta t) = x_1)} = \\ &= \frac{\mathcal{N}(x_3 \mid x_2, \Delta t \sigma^2) \mathcal{N}(x_2 \mid x_1, \Delta t \sigma^2)}{\mathcal{N}(x_3 \mid x_1, 2\Delta t \sigma^2)} = \mathcal{N}\left(x_2 \mid \frac{x_1 + x_3}{2}, \frac{\Delta t \sigma^2}{2}\right). \end{aligned}$$

Процесс Орнштейна-Уленбека - это стационарный марковский гауссовский процесс. Без ограничения общности рассмотрим центрированный процесс. Для корреляционной функции такого процесса верны свойства:

$$\rho(t_1, t_3) = \rho(t_1, t_2) \rho(t_2, t_3), \quad t_1 < t_2 < t_3, \quad \rho(t, s) = \rho(|t - s|), \quad t < s.$$

Отсюда $\rho(a + b) = \rho(a) \rho(b)$. Учитывая то, что корреляционная функция ограничена, получаем, что

$$\rho(t, s) = \sigma^2 e^{-\lambda|t-s|}, \quad \lambda > 0.$$

Аналогично по теореме Байеса и определению условной плотности имеем:

$$\begin{aligned} p(X(t) = x_2 \mid X(t - \Delta t) = x_1, X(t + \Delta t) = x_3) &= \\ &= \mathcal{N}\left(\frac{x_1 + x_3}{2 \cosh(\lambda \Delta t)}, \sigma^2 \tanh(\lambda \Delta t)\right) \end{aligned}$$

9.3. Результаты работы программы

Следующие графики были получены при моделировании стандартного Винеровского процесса и стандартного процесса Орнштейна–Уленбека. Отрезок $[0, 1]$ был разбит на 2^{12} частей.

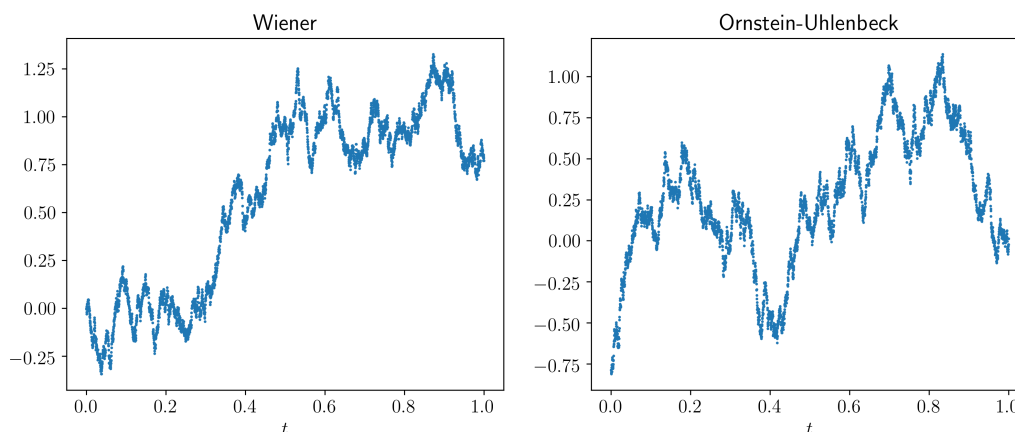


Рис. 18: Траектории процессов.

10. Задание 10

10.1. Постановка задачи

Произвести фильтрацию одномерного процесса Орнштейна–Уленбека:

1. Используя генератор белого шума, добавить случайную ошибку с известной дисперсией к реализации процесса Орнштейна–Уленбека.
2. При помощи одномерного фильтра Калмана оценить траекторию процесса по зашумленному сигналу. Параметры процесса и белого шума считать известными.
3. Рассмотреть случай, когда шум
 - является гауссовским;
 - имеет распределение Коши.

10.2. Теоретическая часть

Рассмотрим процесс с дискретным временем $t_i = hn$, где $n = \overline{0, \dots, N}$. Рассмотрим на этой сетке процесс Орнштейна–Уленбека со значениями $x_n = X(t_n)$. Для применения фильтра Калмана [1] представим процесс в виде одномерной линейной системы:

$$\begin{cases} x_{n+1} = ax_n + \nu_n, \\ \nu_n \sim \mathcal{N}(0, q), \\ x_0 \sim \mathcal{N}(0, \sigma^2). \end{cases}$$

Мы можем наблюдать только зашумленный сигнал:

$$\begin{cases} y_n = x_n + \varepsilon_n, & n = \overline{1, \dots, N}, \\ \varepsilon_n \sim \mathcal{N}(0, r), \\ y_0 = x_0. \end{cases}$$

Обозначим за $\hat{x}_{n|n}$ оценку значения x_n при известных y_1, \dots, y_n . $\hat{x}_{n|n-1}$ — экстраполяция процесса на следующий шаг в соответствии с динамической системой. Через $p_{n|n}$ будем обозначать дисперсию ошибки фильтрации на n -м шаге, через $p_{n|n-1}$ — прогнозируемую на следующем шаге дисперсию. Опишем один шаг дискретного фильтра Калмана:

1. Прогнозируем процесс и дисперсию на 1 шаг вперед:

$$\begin{cases} \hat{x}_{n|n-1} = a\hat{x}_{n-1|n-1}, \\ p_{n|n-1} = a^2 p_{n-1|n-1} + q. \end{cases}$$

2. Вычисляем отклонение прогноза от шумного сигнала и коэффициент усиления Калмана:

$$\begin{cases} \delta_n = y_n - \hat{x}_{n|n-1}, \\ k_n = \frac{p_{n|n-1}}{p_{n|n-1} + r}. \end{cases}$$

3. Фильтруем шум и оцениваем ошибку:

$$\begin{cases} \hat{x}_{n|n} = \hat{x}_{n|n-1} + k_n \delta_n, \\ p_{n|n} = (1 - k_n) p_{n|n-1}. \end{cases}$$

Так как все распределения в данной процедуре являются гауссовскими, то и ошибка будет распределена нормально. Отсюда можно выписать выражение для доверительного интервала с уровнем доверия α :

$$\Delta = \sqrt{p_{n|n}} \Phi^{-1} \left(\frac{1 + \alpha}{2} \right).$$

10.3. Результаты работы программы

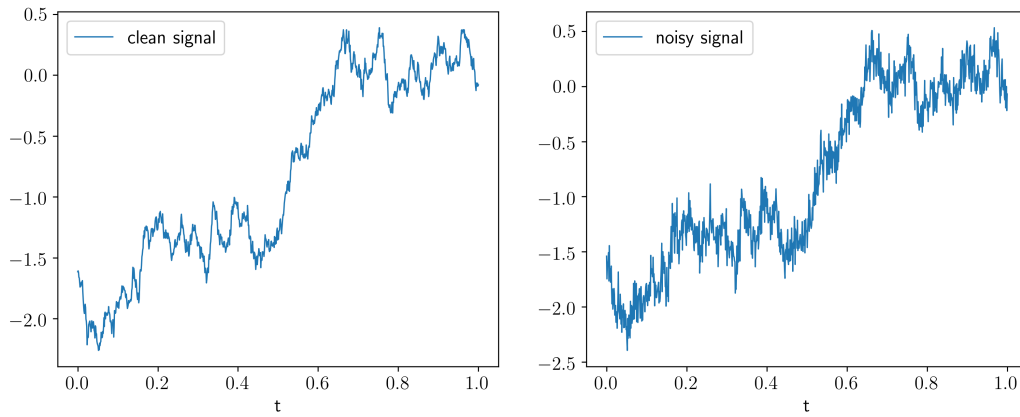


Рис. 19: Чистый и зашумленный гауссовским шумом сигналы, полученные с помощью стандартного процесса Орнштейна–Уленбека.

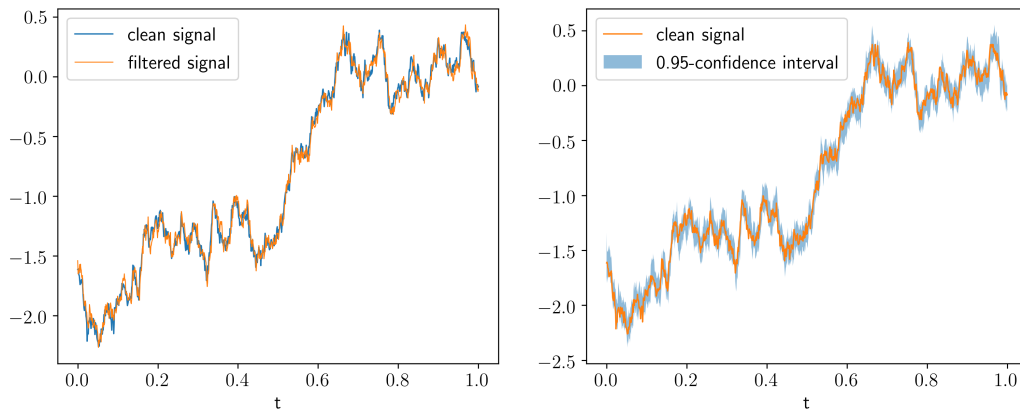


Рис. 20: Результаты фильтрации.

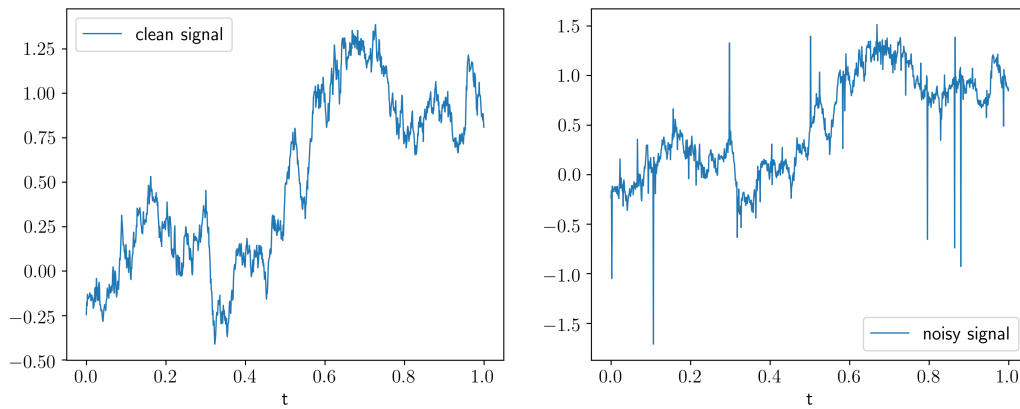


Рис. 21: Чистый и зашумленный сигналы. Шум распределен по Коши.

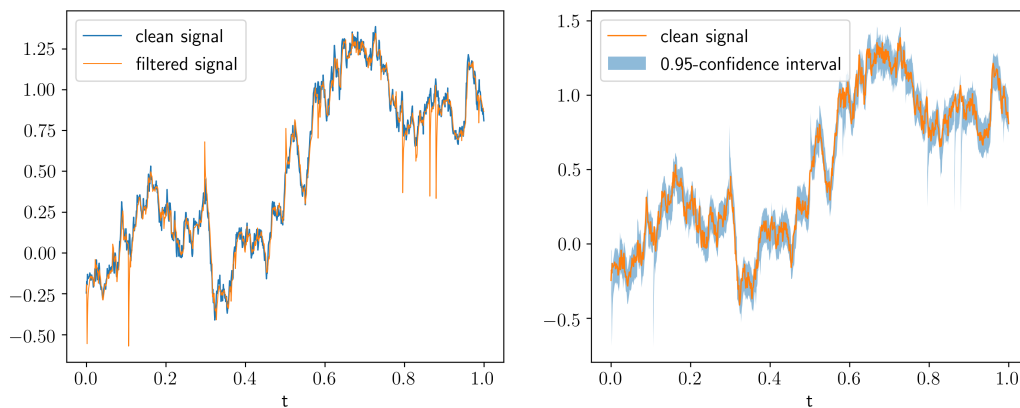


Рис. 22: Результат фильтрации.

11. Задание 11

11.1. Постановка задачи

Построить двумерное пуассоновское поле, отвечающее сложному пуассоновскому процессу:

1. Первая интерпретация: система массового обслуживания. При этом, первая координата поля — время поступления заявки в СМО (равномерное распределение), вторая — время ее обслуживания (распределение χ^2 с 10 степенями свободы).
2. Вторая интерпретация: система массового обслуживания с циклической активностью $\lambda(1 + \cos(t))$ и единичными скачками. Свести

данную задачу при помощи метода Льюиса и Шедлеара к моделированию двумерного пуассоновского поля, где первая координата имеет равномерное распределение, а вторая — распределение Бернулли.

3. Третья интерпретация: работа страховой компании. Первая координата — момент наступления страхового случая (равномерное распределение), вторая координата — величина ущерба (распределение Парето). Поступление капитала по времени линейно со скоростью $c > 0$, начальный капитал $W > 0$.
4. Для каждой системы рассмотреть все возможные случаи поведения системы в зависимости от значения параметров.

11.2. Теоретическая часть

Рассмотрим первую интерпретацию. Так как время поступления заявки имеет равномерное распределение, то мы имеем дело с пуассоновским процессом. Обозначим за λ интенсивность данного процесса. Тогда за единицу времени в СМО поступает в среднем λ заявок. Так как время обработки заявки имеет распределение $\chi^2(10)$, среднее время обработки заявки в СМО составляет $1/10$. Таким образом, возможно 3 варианта эволюции СМО:

1. $\lambda < 1/10$ — заявки в среднем обрабатываются быстрее, чем поступают новые, СМО простаивает.
2. $\lambda = 1/10$ — средняя интенсивность поступления заявок совпадает с интенсивностью их обработки. Процесс находится в равновесии.
3. $\lambda > 1/10$ — СМО не успевает обрабатывать заявки, возникает растущая очередь.

Метод Льюиса и Шедлеара для моделирования Пуассоновского процесса с интенсивностью $\lambda(t)$ очень похож на метод элиминации фон Неймана. Пусть $\lambda(t) \leq \lambda^* \forall t \geq 0$. Затем в каждый момент поступления заявки t_i с вероятностью $1 - \lambda(t_i)/\lambda^*$ отбрасываем заявку. В результате получается процесс с нужной интенсивностью. В задании $\lambda^* = 2\lambda$. Так как среднее значение интенсивности равно λ , случаи те же, что и в 1-й интерпретации.

Работу страховой компании будем моделировать с помощью следующего процесса:

$$X(t) = W + ct - \sum_{n=0}^{N(t)} \xi_n,$$

где $N(t)$ — пуассоновский процесс с интенсивностью λ , W — начальный капитал, c — скорость поступления капитала, а ξ_n имеет распределение Парето с показателем b . Распределение Парето имеет функцию распределения

$$F_{\xi_n}(x) = \begin{cases} 0, & x < 1, \\ 1 - x^{-b}, & x \geq 1. \end{cases}$$

При $b > 1$ математическое ожидание существует и $\mathbb{E}\xi_n = b/(b-1)$ и $\mathbb{E}X(t) = W + ct + \lambda tb/(b-1)$. Соответственно, критическое значение $b^* = c/(c - \lambda)$. Если хвост тяжелее, то капитал в среднем будет расти, если легче, то убывать.

11.3. Результаты работы программы

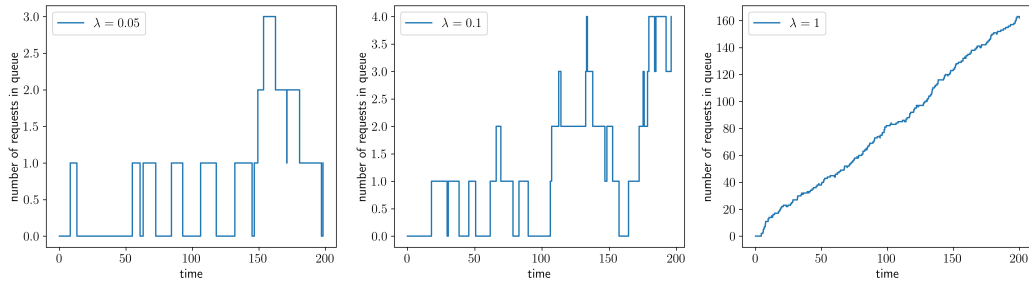


Рис. 23: Моделирование СМО с постоянной активностью при разных интенсивностях поступления заявок.

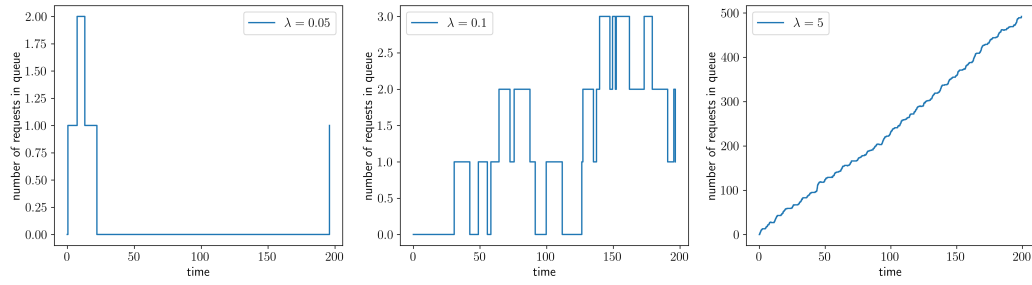


Рис. 24: Моделирование СМО с циклической активностью при разных интенсивностях поступления заявок.

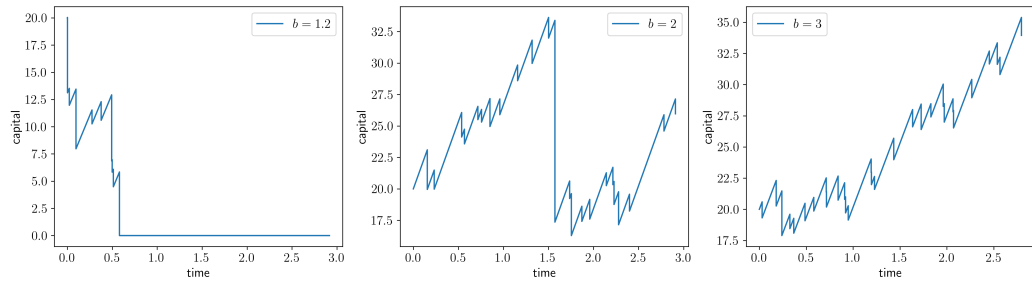


Рис. 25: Моделирование работы страховой компании. Параметры $W = 20$ и $c = 20$. Интенсивность страховых случаев $\lambda = 10$.

Список литературы

- [1] Bishop Christopher M. Pattern recognition and machine learning. Springer, 2006.
- [2] Ширяев А. Н. Вероятность. Наука, 1989.
- [3] DeGroot M. H. Probability and Statistics. Pearson, 2011.
- [4] Метод статистических испытаний / Н.П. Бусленко, Д.И. Голенко, И.М. Соболев [и др.]. Букинист, 1962.