



Московский государственный университет имени М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра системного анализа

Отчет по компьютерному практикуму к курсу

«Стохастический анализ и моделирование»

Студент 415 группы
А. А. Владимиров

Руководитель практикума
к.ф.-м.н., доцент С. Н. Смирнов

Москва, 2022

Содержание

1	Задание 1	4
1.1	Реализация схемы Бернулли и биномиального распределения	4
1.2	Геометрическое распределение	4
1.3	Игра в орлянку	5
2	Задание 2	6
2.1	Датчик для канторова распределения	6
2.2	Свойства симметричности и самоподобия	9
2.3	Математическое ожидание и дисперсия	9
3	Задание 3	10
3.1	Экспоненциальное распределение	11
3.2	Датчик пуассоновского распределения 1	13
3.3	Датчик пуассоновского распределения 2	13
3.4	Датчик стандартного нормального распределения 1	14
4	Задание 4	16
4.1	Датчик распределения Коши	16
4.2	Датчик стандартного нормального распределения 2	16
4.3	Сравнение реализаций датчика стандартного нормального распределения	17
5	Задание 5	18
5.1	Проверка ЗБЧ и ЦПТ	18
5.2	Доверительные интервалы	18
5.3	Поведение сумм S_n/n для распределения Коши	20
6	Задание 6	21
6.1	Численное интегрирование	21
6.2	Точность вычислений	22
7	Задание 7	23
7.1	Случайный поиск	23
7.2	Отжиг	23
7.3	Оценка точности	24
8	Задание 8	24
9	Задание 9	25
9.1	Винеровский процесс	26
9.2	Процесс Орнштейна-Уленбека	26
10	Задание 10	27

11 Задание 11	29
11.1 СМО	29
11.2 СМО с циклической интенсивностью	29
11.3 Работа страховой компании	30

1 Задание 1

Считается доступным лишь генератор равномерно распределенной на отрезке $[0, 1]$ случайной величины $\eta \sim \text{Uni}(0, 1)$. Требуется:

1. Реализовать генератор схемы Бернулли с заданной вероятностью успеха p . На основе генератора схемы Бернулли построить датчик для 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 1 и -1 в зависимости от выпавшей стороны. Проиллюстрировать (в виде ломанной) поведение нормированной суммы $Y(i) = S_i/\sqrt{n}$, как функцию от номера испытания $i = 1, \dots, n$ для одной отдельно взятой траектории. Дать теоритическую оценку для $Y(n)$ при $n \rightarrow \infty$.

1.1 Реализация схемы Бернулли и биномиального распределения

Под схемой Бернулли понимается серия однородных независимых испытаний, каждое из которых с вероятностью p оканчивается успехом или, с вероятностью $q = 1 - p$, неудачей.

Чтобы практически реализовать схему Бернулли нужно получить н.о.р.с.в. $\xi_i \sim \text{Bern}(p), i = 1, \dots, n$. Для этого достаточно выразить ξ_i через η следующим образом: $\xi_i = \mathbb{I}(\eta < p) + \mathbb{I}(\eta \geq p)$, т.е.

$$\xi_i = \begin{cases} 1, & \eta < p, \\ 0, & \eta \geq p. \end{cases}$$

В свою очередь $\beta \sim \text{Bin}(n, p)$ можно представить как $\beta = \sum_{i=1}^n \xi_i$.

Программа, по описанной выше схеме моделирующая $\text{Bin}(16, 0.5)$, дает следующий результат (рис. 1а).

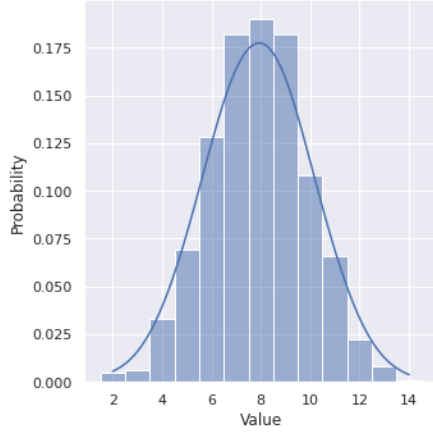
1.2 Геометрическое распределение

Под геометрическим распределением подразумевают распределение случайной величины γ равной числу неудач до первого успеха в серии испытаний Бернулли. $\mathbb{P}(\gamma = n) = q^n p$.

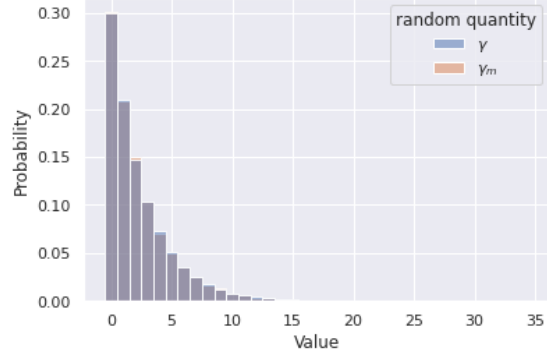
Случайная величина $\gamma \sim \text{Geom}(p)$ представима как

$$\gamma = \max\{n \in \mathbb{N} \cup \{0\} : \xi_i = 0, i = 1, \dots, n\}.$$

Геометрическое распределение обладает свойством отсутствия памяти, т.е.



(a) Биномиальное распределение



(b) Отсутствие памяти ($m = 5$)

Рис. 1

$$\mathbb{P}(\gamma > m + n \mid \gamma \geq m) = \mathbb{P}(\gamma > n), \quad \forall m, n \in \mathbb{N} \cup \{0\}.$$

Это свойство можно переформулировать. Пусть $\gamma \sim \text{Geom}(p)$ — случайная величина, определенная на вероятностном пространстве $(\Omega, \mathcal{A}, \mathbb{P})$. Свойство отсутствия памяти сл. в. γ означает, что

$$\gamma_m \sim \gamma, \quad \forall m \in \mathbb{N} \cup \{0\},$$

где $\gamma_m := (\gamma|_{\Omega_m} - m)$, $\Omega_m = \gamma^{-1}(\gamma \geq m) \in \mathcal{A}$.

То есть для каждого m распределение случайной величины γ отличается ровно на константу m от распределения сл.в. γ , индуцированной на вероятностное подпространство Ω_m .

Проверка свойства отсутствия памяти проведена численным моделированием распределений сл.в. γ и γ_m (рис. 1b).

1.3 Игра в орлянку

Даны n н.о.р.с.в.

$$\theta_j : \mathbb{P}(\theta_j = 1) = \mathbb{P}(\theta_j = -1) = \frac{1}{2}, \quad j = 1, \dots, n.$$

Рассматривается нормированная сумма

$$Y(i) = \frac{S_i}{\sqrt{n}}, \quad \text{где } S_i = \sum_{j=1}^n \theta_j,$$

пример поведения которой проиллюстрирован на рис. 2.

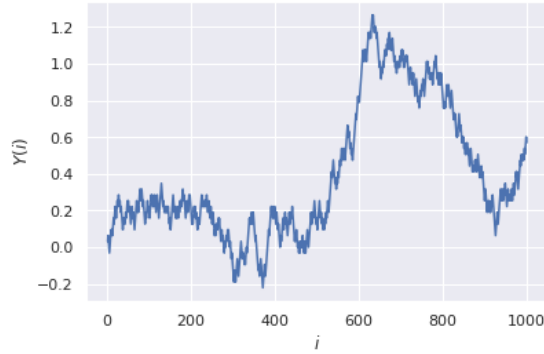


Рис. 2: Игра в орлянку ($n = 1000$)

Для оценки $Y(n)$ при $n \rightarrow \infty$, нам потребуется

Теорема (Центральная предельная теорема). Пусть ξ_1, ξ_2, \dots — последовательность независимых одинаково распределенных (невыврожденных) случайных величин с $\mathbb{E} \xi_1^2 \leq \infty$ и $S_n = \xi_1 + \dots + \xi_n$. Тогда

$$\frac{S_n - \mathbb{E} S_n}{\sqrt{\mathbb{D} S_n}} \xrightarrow{d} \text{Norm}(0, 1).$$

Действительно

$$\frac{S_n - \mathbb{E} S_n}{\sqrt{\mathbb{D} S_n}} = \frac{S_n - 0}{\sqrt{n \mathbb{D} \theta_1}} = \frac{S_n}{\sqrt{n}} = Y(n) \xrightarrow{d} \text{Norm}(0, 1).$$

2 Задание 2

1. Построить датчик сингулярного распределения, имеющий в качестве функции распределения канторову лесницу. С помощью критерия Колмогорова убедиться в корректности работы датчика.
2. Для канторовых случайных величин проверить свойство симметричности относительно $\frac{1}{2}$ (X и $1-X$ распределены одинаково) и самоподобия относительно деления на 3 (условное распределение Y при условии $Y \in [0, \frac{1}{3}]$ совпадает с распределением $Y/3$) с помощью критерия Смирнова.
3. Вычислить значение математического ожидания и дисперсии для данного распределения. Сравнить теоретические значения с эмпирическими для разного объема выборок. Проиллюстрировать сходимость.

2.1 Датчик для канторова распределения

Случайная величина имеет канторово распределение, если ее функция распределения — канторова лесница.

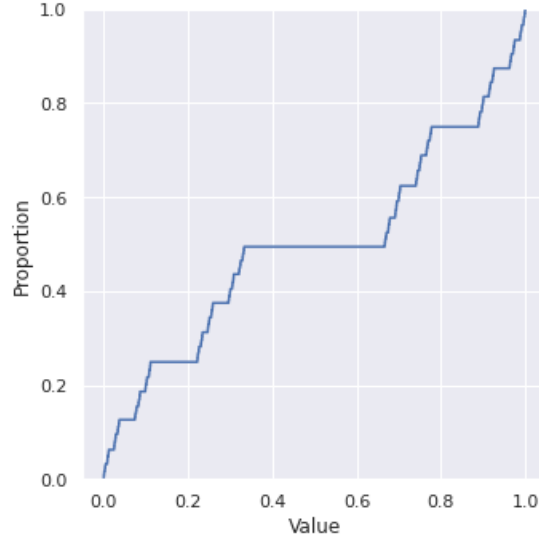


Рис. 3: Э.ф.р. канторовой сл.в.

Носителем канторова распределения является канторово множество, представимое как счетное пересечение множеств

$$\begin{aligned}
C_0 &= [0, 1] \\
C_1 &= [0, 1/3] \cup [2/3, 1] \\
C_2 &= [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1] \\
&\dots, \\
C_n &= \bigcup_{i=1, \dots, 2^n} [a_i, b_i]
\end{aligned}$$

Причем $\mathbb{P}([a_i, b_i]) = \frac{1}{2^n}$, $\forall i$, что дает естественный способ разложения сл.в. $\delta \sim \text{Cant}$ в ряд по $\xi \sim \text{Bern}(0.5)$

$$\delta = \sum_{i=1}^{\infty} \frac{2}{3^i} \xi_i. \quad (1)$$

Таким образом, для моделирования канторовской случайной величины достаточно провести достаточно большое число испытаний бернулли и посчитать сумму приближающую ряд (1).

Для проверки корректности полученного датчика напомним следующее.

Статистикой критерия Колмогорова является величина

$$\sqrt{n}D_n = \sup_{-\infty < x < \infty} \left| \hat{F}_n(x) - F(x) \right|,$$

где $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x)$ — эмпирическая функция распределения.

Теорема (Колмогоров). *Если функция распределения элементов выборки $F(x)$ непрерывна, то для $x > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{n}D_n \leq x) = K(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}.$$

Проверяемая гипотеза с заданным *уровнем значимости* α отвергается, если на полученной выборке (X_1, \dots, X_n) значение статистики неправдоподобно велико, т.е.

$$\sqrt{n}D_n(X_1, \dots, X_n) \geq x_{1-\alpha}, \quad (2)$$

где $x_{1-\alpha}$ — наименьшее значение, удовлетворяющее условию

$$\mathbb{P}(\sqrt{n}D_n \geq x_{1-\alpha}) \leq \alpha.$$

В силу теоремы Колмогорова “в пределе”

$$\mathbb{P}(\sqrt{n}D_n \geq x_{1-\alpha}) = 1 - K(x_{1-\alpha}) = 1 - (1 - \alpha) = \alpha,$$

и $x_{1-\alpha}$ есть не что иное как $(1 - \alpha)$ -квантиль функции $K(x)$.

Таким образом для значения статистики $\sqrt{n}D_n(X)$ на некоторой определенной выборке X справедливо

$$\mathbb{P}(\sqrt{n}D_n \geq \sqrt{n}D_n(X)) = 1 - K(\sqrt{n}D_n(X)) = 1 - (1 - \alpha(X)) = \alpha(X)$$

тем самым $\sqrt{n}D_n - (1 - \alpha(X))$ -квантиль, и условие (2) выполнено для тех и только тех $x_{1-\alpha}$, что

$$\begin{aligned} x_{1-\alpha(X)} &\geq x_{1-\alpha} \Leftrightarrow \\ 1 - \alpha(X) &\geq 1 - \alpha \Leftrightarrow \\ \alpha &\geq \alpha(X) \end{aligned}$$

Выходит что для полученной в результате серии испытаний статистики $\sqrt{n}D_n(X)$ значение $1 - K(\sqrt{n}D_n(X)) = \alpha(X)$ означает, что гипотезу *следует отвергнуть* тогда и только тогда, когда был принят уровень значимости α больший чем величина $\alpha(X)$. И обратно, гипотеза *может быть принята* для любого уровня значимости α меньшего чем $\alpha(X)$.

Поскольку функция $F(x)$ непрерывна и не убывает, а $\hat{F}_n(x)$ — кусочно-постоянна, то D_n можно вычислить по формуле

$$D_n(x_1, \dots, x_n) = \max_{1 \leq i \leq n} \left\{ \frac{1}{n} - F(x_{(i)}), F(x_{(i)}) - \frac{i-1}{n} \right\}.$$

Посчитанная таким образом на некоторой выборке (X_1, \dots, X_n) значений датчика δ статистика D_n составляет приблизительно 0.0087. Величина $1 - K(\sqrt{n}D_n) = \alpha(X)$ равна приблизительно 0.9999, что даёт нам основания принять гипотезу о корректности построенного датчика δ , с уровнем значимости, например, 20%.

2.2 Свойства симметричности и самоподобия

Для проверки требуемых свойств необходимо к двум сгенерированным выборкам (X_1, \dots, X_n) , (Y_1, \dots, Y_m) применить критерий Смирнова, статистикой которого служит величина

$$D_{n,m} = \sup_x \left| \hat{F}_n(x) - \hat{G}_n(x) \right|,$$

где $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x)$, $\hat{G}_n(x) = \frac{1}{m} \sum_{j=1}^m \mathbb{I}(Y_j \leq x)$.

Известен следующий результат

Теорема (Смирнов). *Если гипотеза однородности верна (подробнее см. [3]), то имеет место сходимость*

$$\mathbb{P} \left(\sqrt{nm/(n+m)} D_{n,m} \leq x \right) \rightarrow K(x) \text{ при } n, m \rightarrow \infty,$$

где $K(x)$ — функция распределения Колмогорова.

Для нахождения статистики достаточно произвести вычисления по формулам

$$D_{n,m} = \max\{D_{n,m}^+, D_{n,m}^-\},$$

где

$$D_{n,m}^+ = \sup_x \left(\hat{F}_n(x) - \hat{G}_n(x) \right) = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - \hat{G}_n(X_{(i)}) \right\}$$

$$D_{n,m}^- = \sup_x \left(\hat{G}_n(x) - \hat{F}_n(x) \right) = \max_{1 \leq j \leq m} \left\{ \frac{j}{m} - \hat{F}_n(Y_{(j)}) \right\}.$$

В результате компьютерных вычислений для случайных величин δ и $1 - \delta$ получены значения $D_{n,m} \approx 0.489$ и $1 - K(\sqrt{nm/(n+m)} D_{n,m}) \approx 0.1811$, что позволяет принять гипотезу об одиноковой распределенности на уровне значимости 10%.

Соответствующие результаты для $\delta/3$ и $\delta|_{[0,3]}$ составляют приблизительно 0.3482 и 0.1427 соответственно, что подтверждает свойство самоподобия на уровне значимости 10%.

2.3 Математическое ожидание и дисперсия

Случайная величина δ обладает свойством самоподобия, т.е. для ее функции распределения $F(\cdot)$ выполнено:

- $F(x) = \frac{F(3x)}{2}$, при $\frac{2}{3} < x < 1$,
- $F(x) = \frac{1}{2} + \frac{F(3x-2)}{2}$, при $\frac{2}{3} < x < 1$.

Используя это свойство вычислим математическое ожидание

$$\begin{aligned}\mathbb{E} \delta &= \int_{-\infty}^{+\infty} x dF(x) = \int_0^{1/3} x dF(x) + \int_{2/3}^1 x dF(x) = \frac{1}{2} \int_0^{1/3} x dF(3x) + \\ &+ \frac{1}{2} \int_{2/3}^1 x d(1/2 + F(3x - 2)) = \frac{1}{2} \int_0^{1/3} \frac{y}{3} dF(y) + \frac{1}{2} \int_0^1 \frac{y+2}{3} dF(y) = \\ &\frac{1}{6} \int_0^1 y dF(y) + \frac{1}{6} \int_0^1 y dF(y) + \frac{1}{3} \int_0^1 dF(y) = \frac{1}{3} \mathbb{E} \delta + \frac{1}{3}.\end{aligned}$$

Следовательно

$$\mathbb{E} \delta = \frac{1}{2}.$$

Аналогично дисперсия:

$$\begin{aligned}\mathbb{E} \delta^2 &= \int_0^{1/3} x^2 dF(x) + \int_{2/3}^1 x^2 dF(x) = \frac{1}{2} \int_0^{1/3} \left(\frac{y}{3}\right)^2 dF(y) + \\ &+ \frac{1}{2} \int_0^1 \left(\frac{y+2}{3}\right)^2 dF(y) = \frac{1}{9} \mathbb{E} \delta^2 + \frac{2}{9} \mathbb{E} \delta + \frac{2}{9} = \frac{1}{9} \mathbb{E} \delta^2 + \frac{1}{3}.\end{aligned}$$

Т.е.

$$\mathbb{E} \delta^2 = \frac{3}{8},$$

и

$$\mathbb{D} \delta = \mathbb{E} (\delta - \mathbb{E} \delta)^2 = \mathbb{E} \delta^2 - (\mathbb{E} \delta)^2 = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$$

Сходимость эмпирических показателей к теоретическим показана на рис. 4.

3 Задание 3

1. Построить датчик экспоненциального распределения. Проверить для данного распределения свойство отсутствия памяти. Пусть X_1, X_2, \dots, X_n — независимые экспоненциально распределенные с.в. с параметрами $\lambda_1, \lambda_2, \dots, \lambda_n$ соответственно. Найти распределение случайной величины $Y = \min(X_1, X_2, \dots, X_n)$.
2. На основе датчика экспоненциального распределения построить датчик пуассоновского распределения.

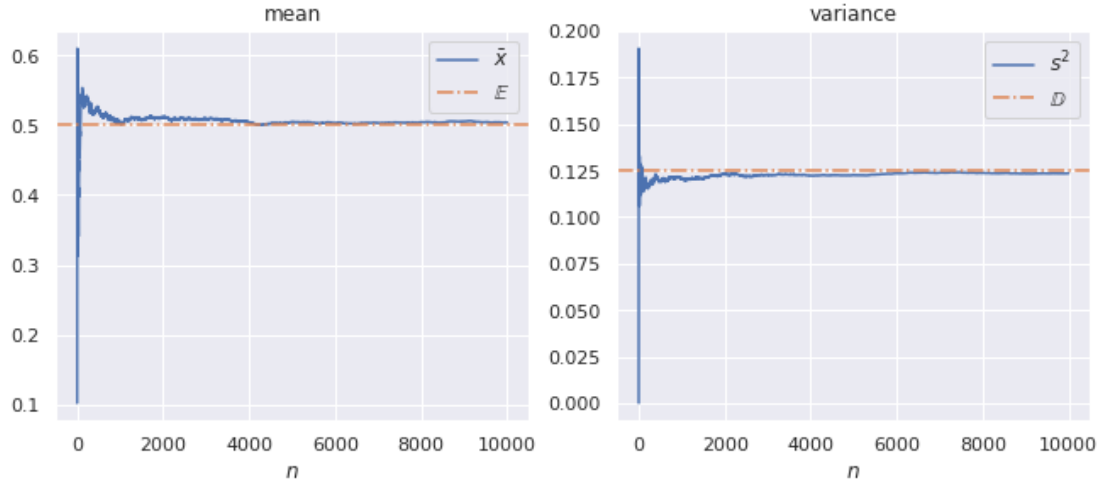


Рис. 4

3. Построить датчик пуассоновского распределения как предел биномиального распределения. С помощью критерия хи-квадрат Пирсона убедиться, что получен датчик распределения Пуассона.
4. Построить датчик стандартного нормального распределения методом моделирования случайных величин парами с переходом в полярные координаты. Проверить при помощи t -критерия Стьюдента равенство математических ожиданий, а при помощи критерия Фишера равенство дисперсий.

3.1 Экспоненциальное распределение

Случайная величина ε имеет экспоненциальное распределение с параметром λ , если ее функция распределения имеет вид

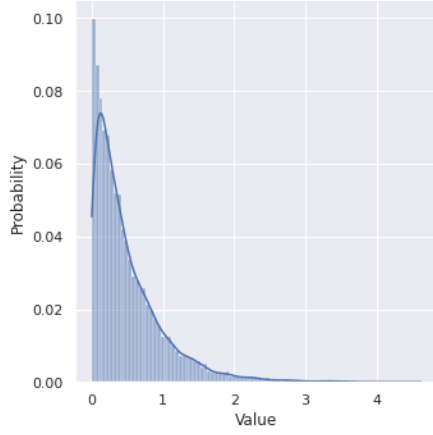
$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (3)$$

Мы будем моделировать $\varepsilon \sim \text{Exp}(\lambda)$ с помощью метода обратной функции¹, согласно которому

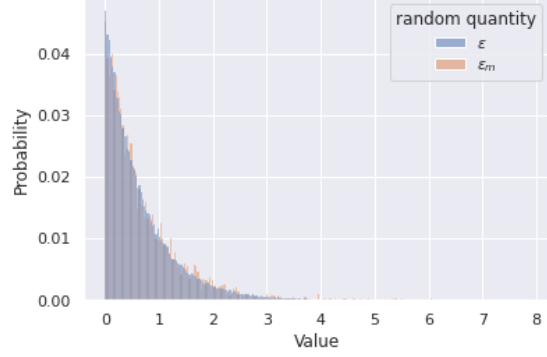
$$\varepsilon \sim -\frac{1}{\lambda} \ln \eta,$$

где $\eta \sim \text{Uni}(0, 1)$. Результат моделирования см. рис. 5а

¹см. [3] гл. 4 §1 “Метод обратной функции”.



(a) Э.ф.п. экспоненциального распределения



(b) Отсутствие памяти ($m = 2$)

Рис. 5

Экспоненциальное распределение обладает свойством отсутствия памяти (рис. 5b), т.е. аналогично § 1.2:

$$\varepsilon_m \sim \varepsilon, \quad \forall m \in \mathbb{N} \cup \{0\},$$

где $\varepsilon_m := (\gamma|_{\Omega_m} - m)$, $\Omega_m = \varepsilon^{-1}(\varepsilon \geq m) \in \mathcal{A}$.

Распределение сл.в. $Y = \min(X_1, X_2, \dots, X_n)$ имеет функцию распределения

$$\begin{aligned} F(y) &= \mathbb{P}(Y < y) = 1 - \mathbb{P}(Y \geq 1) = 1 - \mathbb{P}(\min(X_1, X_2, \dots, X_n) \geq 1) = \\ &= 1 - \mathbb{P}(X_1 \geq y, \dots, X_n \geq y) = \{\text{в силу независимости сл.в. } X_i\} = 1 - \prod_{i=1}^n \mathbb{P}(X_i \geq y) = \\ &= 1 - \prod_{i=1}^n (1 - F_{Exp(\lambda_i)}) = 1 - \prod_{i=1}^n (1 - (1 - e^{-\lambda_i y})) = 1 - \prod_{i=1}^n e^{-\lambda_i y} = 1 - e^{-(\sum_{i=1}^n \lambda_i)y}. \end{aligned}$$

Таким образом

$$Y \sim \text{Exp}\left(\sum_{i=1}^n \lambda_i\right)$$

Сравните результаты, посчитанные для обоих представлений Y на рис. 6.

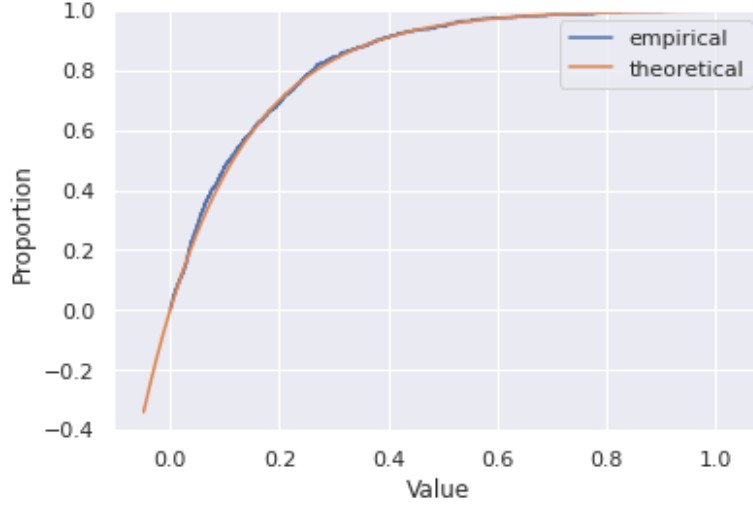


Рис. 6: Э.ф.р. $\text{Exp}(\lambda)$

3.2 Датчик пуассоновского распределения 1

Случайная величина π имеет распределение Пуассона с параметром $\lambda > 0$, если $\mathbb{P}(\pi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Для $\pi \sim \text{Pois}(\lambda)$ верно следующее представление ²

$$\pi = \max_{\mathbb{N} \cup \{0\}} (n \mid S_n = \sum_{i=1}^n \varepsilon_i \leq 1),$$

где $\varepsilon_1, \varepsilon_2, \dots$ — экспоненциальные с параметром λ н.о.р.с.в.

Результат моделирования см. рис. 7а

3.3 Датчик пуассоновского распределения 2

Известно, что распределение $\text{Pois}(\lambda)$ получается из биномиального $\text{Bin}(n, p)$ предельным переходом при $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \lambda$. Результат моделирования распределения пуассона на основе этого принципа предоставлен на рис. 7б.

Проверим корректность нашего датчика при помощи критерия хи-квадрат Пирсона.

Пусть X_1, \dots, X_n — выборка. Разобьем множество значений ξ_1 на N промежутков (возможно, бесконечных) $\Delta_j = (a_j, b_j]$, $j = 1, \dots, N$. Положим $p_j = \mathbb{P}(X_1 \in \Delta_j)$, а случайные величины ν_j — равными количеству элементов выборки в Δ_j .

²см. [3] гл. 5 §1 “Моделирование дискретных величин”.

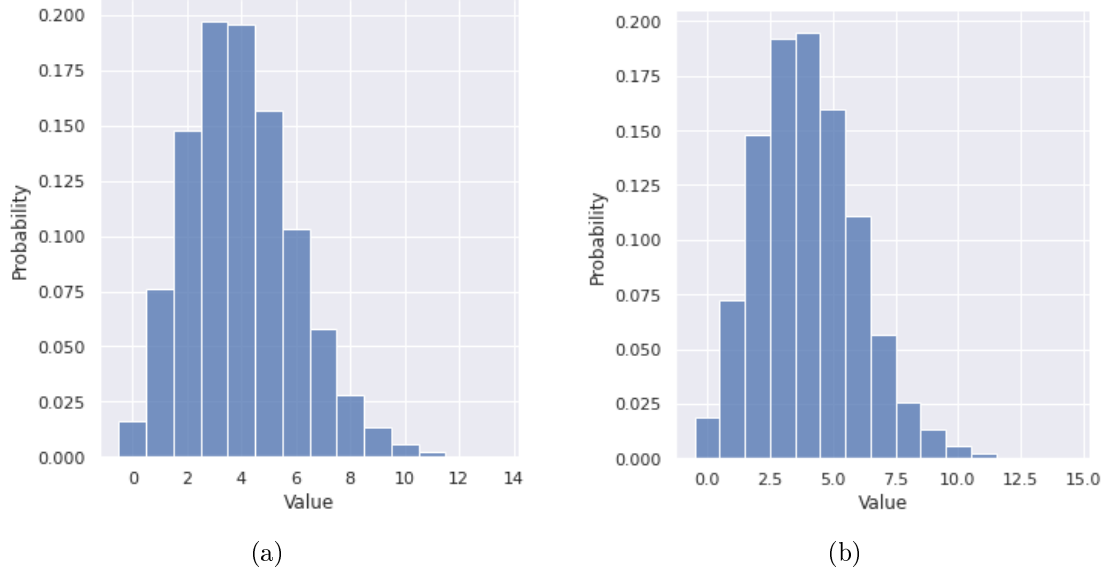


Рис. 7: Гистограммы датчиков распределения Пуассона

Так, статистикой критерия Пирсона служит величина

$$X_n^2 = \sum_{j=1}^N \frac{(\nu_j - np_j)^2}{p_j}, \quad (4)$$

предел которой, при $n \rightarrow \infty$ имеет распределение χ_{N-1}^2 .

Поскольку распределение Пуассона дискретно, промежутки Δ_j зададим как $\Delta_j = \{j - 1\}, j = 1, \dots, N - 1$; $\Delta_N = \{j \in \mathbb{N} : j > N - 1\}$. N положим равным приблизительно $\log_2 n$. Теперь вычислим соответствующие частоты, вероятности и статистику. По результатам вычислений гипотеза о корректности принимается с уровнем значимости 5%.

3.4 Датчик стандартного нормального распределения 1

Нормальное распределение — абсолютно непрерывное распределение с плотностью $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, где параметры μ, σ — математическое ожидание и среднеквадратическое отклонение соответственно.

С помощью нелинейного преобразования пары н.о.р.с.в. $\eta_1, \eta_2 \sim U[0, 1]$ можно получить две н.о.р.с.в. $X, Y \sim \mathcal{N}(0, 1)$:

$$\begin{aligned} X &= -\sqrt{-2 \ln \eta_1} \cos(2\pi\eta_2), \\ Y &= -\sqrt{-2 \ln \eta_1} \sin(2\pi\eta_2). \end{aligned}$$

Результат смоделированной таким способом стандартной нормальной случайной величины см. 8.

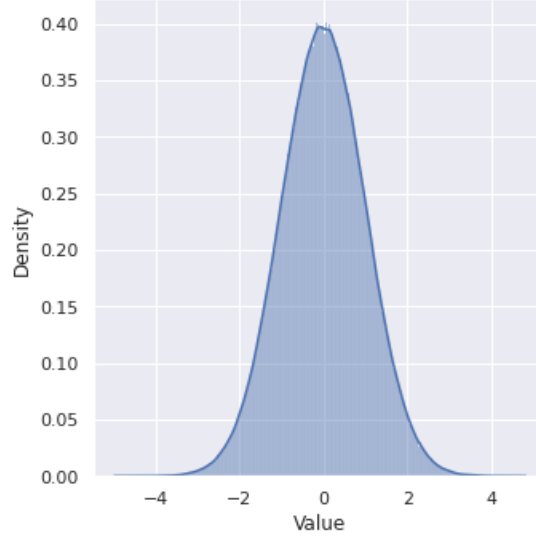


Рис. 8: Э.ф.п. нормального распределения

Для проверки однородности двух независимых нормальных выборок (в нашем случае это X и Y) используют критерии Фишера и Стьюдента. Оба критерия имеют двустороннюю критическую области, т.о. при заданном уровне значимости α нулевая гипотеза принимается в случае если статистика приняла значение между $(\alpha/2)$ и $(1 - \alpha/2)$ квантилями соответствующего распределения.

Критерий Фишера служит для проверки гипотезы о соответствии дисперсий двух распределений. Его статистика имеет вид

$$\frac{S_1^2}{S_2^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{m-1} \sum_{j=1}^m (Y_j - \bar{Y})^2}, \quad (5)$$

и распределена по закону $F_{n-1, m-1}$, т.е так же как и случайная величина $\zeta = (\frac{1}{n-1}\xi)/(\frac{1}{m-1}\eta)$, где $\xi \sim \chi_{n-1}^2, \eta \sim \chi_{m-1}^2$, ξ и η независимы.

В результате вычисления статистики (5) гипотеза о соответствии дисперсий принимается с уровнем значимости $\alpha = 5\%$.

Критерий Стьюдента позволяет проверить гипотезу о соответствии математических ожиданий. Статистика в данном случае имеет вид

$$\frac{\sqrt{\frac{nm}{n+m}}(\bar{X} - \bar{Y})(n+m-2)}{[(n-1)S_1^2 + (m-1)S_2^2]} \sim t_{n+m-2}, \quad (6)$$

где t_{n+m-2} — распределение Стьюдента с $(n+m-2)$ степенями свободы.

Значение статистики Стьюдента позволяет нам принять гипотезу о равенстве мат. ожиданий с уровнем значимости $\alpha = 5\%$, что, в совокупности с уже установленным равенством дисперсий дает нам сделать вывод об однородности нормальных выборок X и Y с тем же уровнем значимости.

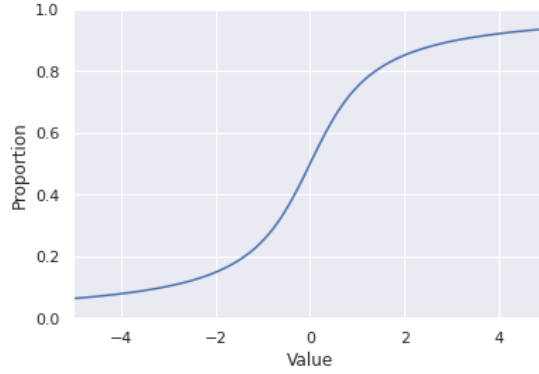


Рис. 9: Э.ф.р. датчика Коши

4 Задание 4

1. Построить датчик распределения Коши.
2. На основе датчика распределения Коши с помощью метода фон Неймана построить датчик стандартного нормального распределения. При помощи функции `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 Датчик распределения Коши

Распределение Коши — абсолютно непрерывное распределение с функцией распределения $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctg(\frac{x-x_0}{\gamma})$.

Как и в пункте 3.1 воспользуемся методом обратной функции:

$$\xi \sim F^{-1}(\eta),$$

где $\eta \sim U(0, 1)$, $F^{-1}(y) = x_0 + \gamma \operatorname{tg}[\pi(x - \frac{1}{2})]$.

Пример результата работы полученного датчика представлен на рис. 9.

4.2 Датчик стандартного нормального распределения 2

Будем генерировать стандартную нормальную случайную величину методом Фон-Неймана используя стандартное распределение Коши ($\eta \sim \text{Cauchy}(0, 1)$) и распределение Бернулли.

Требуемая выборка $\{X_i\}_{i=1}^n$ получается следующим образом. Для очередного i генерируется некоторое значение x из закона $\eta \sim \text{Cauchy}(0, 1)$ до тех пор, пока результат проведенного затем испытания Бернулли $\nu(x)$ с вероятностью успеха

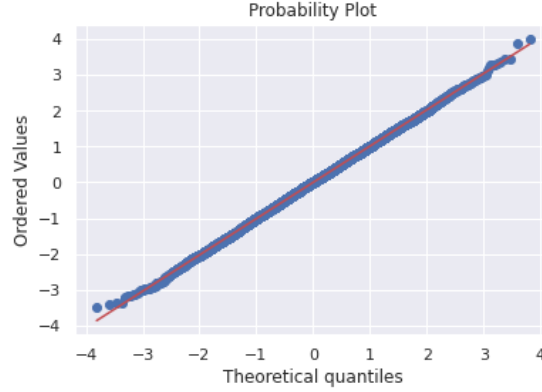


Рис. 10

$\frac{\sqrt{e}}{2}e^{-\frac{x^2}{2}}(x^2 + 1)$ не будет положительным. Тогда значение элемента выборки X_i принимается равным x .

Формула вероятности успеха в испытании Бернулли возникает как отношение "вероятностей попадания в точку x " моделируемого распределения (нормального) и распределения с помощью которого моделируют (в нашем случае Коши), т.е. отношение плотностей упомянутых распределений $\frac{p_1(x)}{p_2(x)}$. Здесь $p_1(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $p_2(x) = \frac{1}{\pi} \frac{1}{x^2+1}$. Для ускорения работы моделирующего алгоритма это отношение стараются приблизить к единице. В данном случае это достигается домножением на коэффициент $\frac{1}{k}$, где $k = \sqrt{\frac{2\pi}{e}}$. В итоге получаем $p = \frac{p_1(x)}{kp_2(x)} = \frac{\sqrt{e}}{2}e^{-\frac{x^2}{2}}(x^2 + 1)$.

Для проверки корректности построенного датчика воспользуемся функцией `scipy.stats.probplot`, которая строит теоретическую (для стандартного нормального распределения) и эмпирическую (для проверяемого распределения) функцию квантилей распределения. Поскольку нормальное распределение для разных параметров линейно выражается через стандартное, неудивительно что для разных μ и σ график будет меняться линейно. График для датчика стандартного нормального распределения см. рис. 10.

4.3 Сравнение реализаций датчика стандартного нормального распределения

По результатам 1000 запусков функций генерации выборки размера 1000 среднее время работы составляет

- $555\mu s \pm 28.6\mu s$ для датчика из 4.2
- $83.6\mu s \pm 5.61\mu s$ для датчика из 3.4

5 Задание 5

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

$$\frac{S_n - \mu n}{\sigma \sqrt{n}}.$$

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

5.1 Проверка ЗБЧ и ЦПТ

Действительно, для выборки $\{X_i\} \sim \mathcal{N}(\mu, \sigma^2)$ значение $\frac{S_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$ ведет себя в соответствии с законом больших чисел (см. рис. 11a), т.е

$$\frac{S_n}{n} \rightarrow \mu, \text{ при } n \rightarrow \infty.$$

Результаты опыта соответствуют и центральной предельной теореме, т.е

$$\frac{S_n - \mu n}{\sigma \sqrt{n}} \xrightarrow{d} \eta \sim \mathcal{N}(0, 1) \text{ при } n \rightarrow \infty. \text{ (см. рис. 11b)}$$

5.2 Доверительные интервалы

Пусть имеется выборка $X_i \sim \mathcal{N}(\mu, \sigma^2)$, где μ и σ^2 — неизвестны. Исходя из

Теорема. Для нормальной выборки $X_i \sim \mathcal{N}(\mu, \sigma^2)$ выборочное среднее $\bar{X} = \frac{1}{n} \sum X_i$ и выборочная дисперсия $S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$ независимы, причем $nS^2/\sigma^2 \sim \chi_{n-1}^2$, а $\sqrt{n-1}(\bar{X} - \mu)/S \sim t_{n-1}$.

можно построить доверительные интервалы

- для параметра сдвига μ :

$$\mathbb{P} \left(\bar{X} - \frac{y_{1-\alpha/2} S}{\sqrt{n-1}} < \mu < \bar{X} - \frac{y_{\alpha/2} S}{\sqrt{n-1}} \right) = \mathbb{P} \left(y_{\alpha/2} < \frac{\sqrt{n-1}(\bar{X} - \mu)}{S} < y_{1-\alpha/2} \right) = 1 - \alpha,$$

где y_p — p -квантиль распределения Стьюдента t_{n-1} (в силу симметрии закона $y_{\alpha/2} = -y_{1-\alpha/2}$);

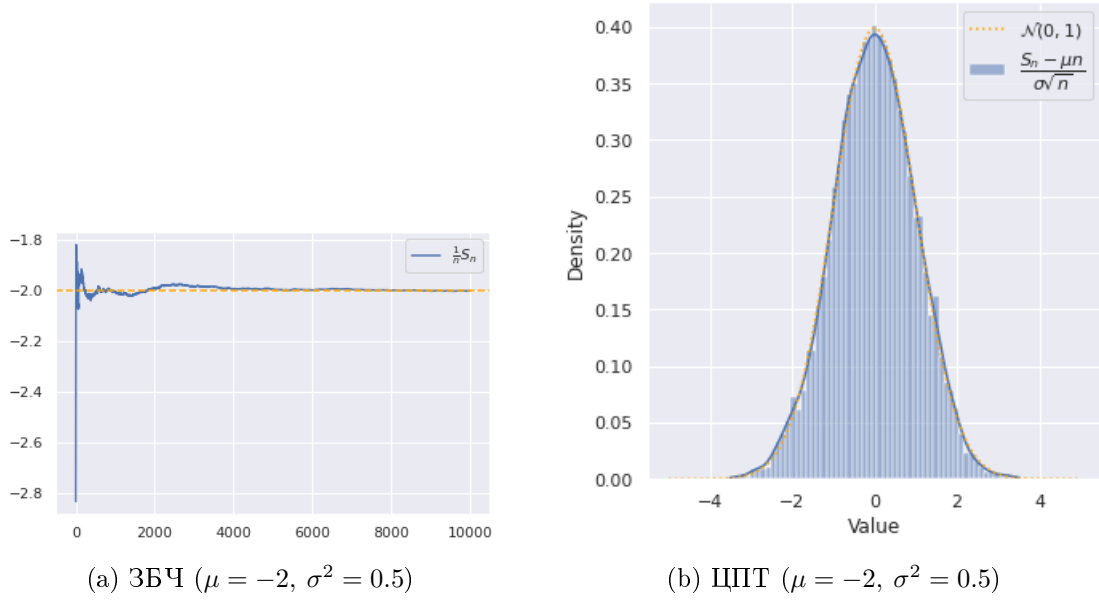


Рис. 11

- и для параметра масштаба σ :

$$\mathbb{P}\left(\frac{\sqrt{n}S}{\sqrt{z_{1-\alpha/2}}} < \sigma < \frac{\sqrt{n}S}{\sqrt{z_{\alpha/2}}}\right) = \mathbb{P}\left(z_{\alpha/2} < \frac{nS^2}{\sigma^2} < z_{1-\alpha/2}\right) = 1 - \alpha,$$

где z_p — p -квантиль закона χ_{n-1}^2 .

Доверительные интервалы для некоторой выборки из $\mathcal{N}(-2, 0.5)$ для разных n приведены на рис. 12

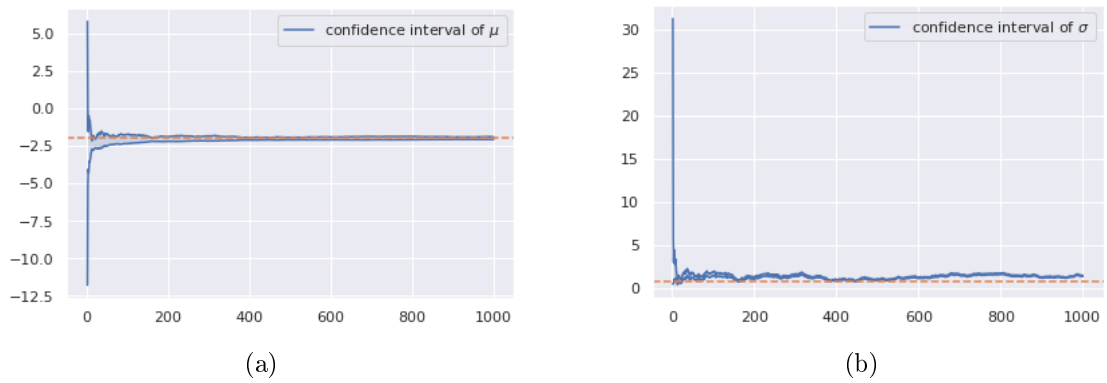


Рис. 12

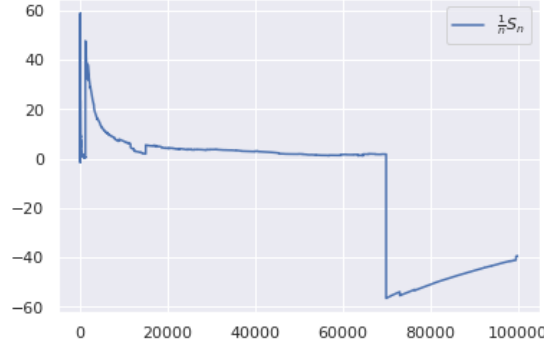


Рис. 13: $a = 0, b = 3$

5.3 Поведение сумм S_n/n для распределения Коши

Как известно распределение Коши не имеет математического ожидания (ни конечного, ни бесконечного), а потому для него не выполняется ЗБЧ, и сходимость выборочного среднего S_n/n не имеет места. Тем не менее распределение $\text{Cauchy}(a, b)$ симметрично относительно a и интеграл $\int_{-\infty}^{\infty} x p_c(x) dx$ сходится в смысле главного значения к a . Этими наблюдениями объясняется полученное на опыте поведение сумм S_n/n : их значение в основном держится около некоторой средней величины определяемой значением a (сходимость *v.p.*) и "периодическими" сильными выбросами (расходимость в целом) от a ее отклонившими. График сумм S_n/n для некоторой выборки X_i представлен на рис. 13.

Ввиду отсутствия математического ожидания не выполнена и ЦПТ, и сходимость S_n/n по распределению с ее помощью доказать не получится. Вместо этого воспользуемся аппаратом характеристических функций. Характеристическая функция распределения Коши с параметрами a, b имеет вид

$$\psi(t) = \int_{-\infty}^{\infty} \frac{e^{itx}}{\pi(b^2 + (x-a)^2)} dx = e^{ait - b|t|}.$$

Характеристическая функция суммы независимых случайных величин есть произведение их характеристических функций, а потому

$$\psi_{\frac{S_n}{n}}(t) = \psi_{\sum \frac{1}{n} X_i}(t) = \prod \psi_{\frac{1}{n} X_i}(t) = \prod \psi_{X_i}\left(\frac{t}{n}\right) = e^{n(ai\frac{t}{n} - b|\frac{t}{n}|)} = e^{ait - b|t|}. \quad (7)$$

Поскольку хар. функция однозначно задает распределение, равенство (7) доказывает что величина $\frac{S_n}{n}$ (так же как и X_i) распределена по Коши с параметрами a и b .

6 Задание 6

1. Посчитать интеграл

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{e^{-\left(x_1^2 + \dots + x_{10}^2 + \frac{1}{2^7 \cdot x_1^2 \cdots x_{10}^2}\right)}}{x_1^2 \cdots x_{10}^2} dx_1 dx_2 \dots dx_{10} \quad (8)$$

— методом Монте-Карло

— методом квадратур, сводя задачу к вычислению собственного интеграла Римана

2. Для каждого случая оценить точность вычислений.

6.1 Численное интегрирование

Перепишем (8) следующим образом

$$\begin{aligned} & \int \cdots \int_{\mathbb{R}^{10}} \frac{e^{-\left(x_1^2 + \dots + x_{10}^2 + \frac{1}{2^7 \cdot x_1^2 \cdots x_{10}^2}\right)}}{x_1^2 \cdots x_{10}^2} dx_1 dx_2 \dots dx_{10} = \\ & \int \cdots \int_{\mathbb{R}^{10}} \frac{\pi^5 e^{-\frac{1}{2^7 \cdot x_1^2 \cdots x_{10}^2}}}{x_1^2 \cdots x_{10}^2} \cdot \frac{e^{-(x_1^2 + \dots + x_{10}^2)}}{\pi^5} dx_1 dx_2 \dots dx_{10} = \int \cdots \int_{\mathbb{R}^{10}} f(x) \cdot p(x) dx, \end{aligned}$$

где $p(x)$ — плотность многомерного нормального распределения $\mathcal{N}(0, \frac{1}{2}E)$, $E \in \mathbb{R}^{10 \times 10}$. Таким образом

$$\int \cdots \int_{\mathbb{R}^{10}} f(x) \cdot p(x) dx = \mathbb{E} f(\eta), \quad \eta \sim \mathcal{N}(0, \frac{1}{2}E).$$

По усиленному закону больших чисел имеем

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n f(\eta_i) \xrightarrow{\text{п.н.}} \mathbb{E} f(\eta) = \int \cdots \int_{\mathbb{R}^{10}} f(x) \cdot p(x) dx = I,$$

что дает нам основания использовать метод Монте-Карло для вычисления интеграла (8).

Другим способом вычисления может служить метод квадратур, для реализации которого проведем замену

$$x_i = \operatorname{tg}\left(\frac{\pi}{2}t_i\right), \quad t_i \in [0, 1], \quad i = \overline{1, 10}.$$

Тогда (8) примет вид

$$I = \pi^{10} \int \cdots \int_{[0,1]^{10}} \frac{e^{-\left(\sum_{i=1}^{10} \operatorname{tg}\left(\frac{\pi}{2}t_i\right)^2 + \frac{1}{2^7 \prod_{i=1}^{10} \operatorname{tg}\left(\frac{\pi}{2}t_i\right)^2}\right)}}{\prod_{i=1}^{10} \operatorname{tg}\left(\frac{\pi}{2}t_i\right)^2 \cdot \prod_{i=1}^{10} \cos\left(\frac{\pi}{2}t_i\right)^2} dt. \quad (9)$$

Интеграл (9) уже можно вычислить, например, стандартным методом прямоугольников на равномерной сетке.

6.2 Точность вычислений

В соответствии с ЦПТ и правилом трех сигм, при достаточно больших n погрешность метода Монте-Карло с вероятностью около 0.997 составляет

$$\psi_n = 3 \frac{\sqrt{\mathbb{D} f(\eta)}}{\sqrt{n}} \geq |\hat{I}_n - I|.$$

Для вычисления погрешности будем приближать значение $\mathbb{D} f(\eta)$ выборочной дисперсией

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n f^2(x_i) - \left(\frac{1}{n} \sum_{i=1}^n f(x_i) \right)^2.$$

Результаты работы программы представлены на таблице 1.

Объем выборки	Результат	Величина ошибки	Время работы (сек.)
$n = 10^3$	120.0206	13.7449	0.021
$n = 10^4$	122.5785	3.3040	0.416
$n = 10^5$	124.7199	1.0937	3.67

Таблица 1

Как известно погрешность метода прямоугольников составляет

$$\psi_n = \frac{h^2}{24} (b - a) \sum_{i,j=1}^{10} \max |f''_{x_i x_j}| = \frac{1}{6n^2} \sum_{i,j=1}^{10} \max |f''_{x_i x_j}|.$$

Результаты работы программы см. на таблице 2.

Мощность сетки	Результат	Время работы (сек.)
$n = 10^3$	1572.5830	0.03
$n = 10^4$	1247.9712	0.474
$n = 10^5$	794.7241	5.19

Таблица 2

7 Задание 7

1. Методом случайного поиска найти минимальное значение функции f на множестве $A = \{x_1, x_2 : x_1^2 + x_2^2 \leq 1\}$, т.е. $y = \min f(x)$, где

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

при $x_1 \neq 0$ и $x_2 \neq 0$, функция доопределяется по непрерывности при $x_1 = 0$ или $x_2 = 0$.

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.

7.1 Случайный поиск

Случайный поиск реализуем следующим образом. Равномерно разбросаем на единичном круге A точки $(X_i, Y_i) \sim \text{Uniform}(A)$, т.е.

$$\begin{aligned} X_i &= \sqrt{r} \cos(\varphi), \\ Y_i &= \sqrt{r} \sin(\varphi), \end{aligned}$$

где $r \sim \text{Uniform}[0, 1]$, $\varphi \sim \text{Uniform}[0, 2\pi]$. Из полученной выборки возьмем точку (X_k, Y_k) доставляющую минимум функции f . Она и будет результатом работы алгоритма. Испытания см. таб. 3.

Объем выборки	argmin	\min
$n = 10^3$	(0.4135, -0.9091)	-1.2351
$n = 10^4$	(-0.3691, -0.9256)	-1.2557
$n = 10^5$	(-0.3472, -0.9372)	-1.2832
$n = 10^6$	(-0.3576, -0.9338)	-1.2883

Таблица 3

7.2 Отжиг

Алгоритм метода принимает некоторую точку x_0 как исходные данные. Затем строится минимизирующая последовательность $\{x_i\}$. Точка x_{i+1} получается на основе текущей точки x_i , а именно: случайно генерируется точка x^* , например из распределения $\mathcal{N}(x_i, q_i \sigma^2 E)$, после чего точка x^* с некоторой вероятностью становится

точкой x_{i+1}

$$\mathbb{P}(x^* \rightarrow x_{i+1} | x_i) = \begin{cases} 1, & F(x^*) \leq F(x_i), \\ \exp\left(-\frac{F(x^*) - F(x_i)}{q_i}\right), & F(x^*) \geq F(x_i). \end{cases}$$

В качестве $\{q_i\}$ берется обычно некоторая убывающая последовательность, например геометрическая прогрессия со знаменателем меньшим единицы.

За 100 шагов алгоритм с параметрами: $x_0 = (0, 0)$, $\sigma = 2$, $q_0 = 1000$, $k = 0.93$ нашел $\min = 0.0031$ в точке $x_{100} = (1.0382, 1.0820)$.

7.3 Оценка точности

Оценим точность случайного поиска. Пусть $x = (x_1, x_2)$ — точка минимума, $\hat{x} = (\hat{x}_1, \hat{x}_2)$ — результат работы алгоритма.

Вероятность того что хотя бы одна точка попадет в ε -окрестность x составляет $1 - (1 - 2\varepsilon^2)^n$. Таким образом

$$\mathbb{P}(|x - \hat{x}| < \varepsilon) = 1 - (1 - 2\varepsilon^2)^n.$$

Ясно, что $|f(x) - f(\hat{x})| < \|f'\| |x - \hat{x}| < \max_A \|f'\| |x - \hat{x}|$. Можно оценить $\max_A \|f'\|$ как

$$\|f'\| = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2} \leq 34.26.$$

Итого мы можем оценить ошибку $\psi_n = |f(x) - f(\hat{x})|$ сверху величиной ε , с вероятностью $1 - (1 - 2(\frac{\varepsilon}{34.26})^2)^n$.

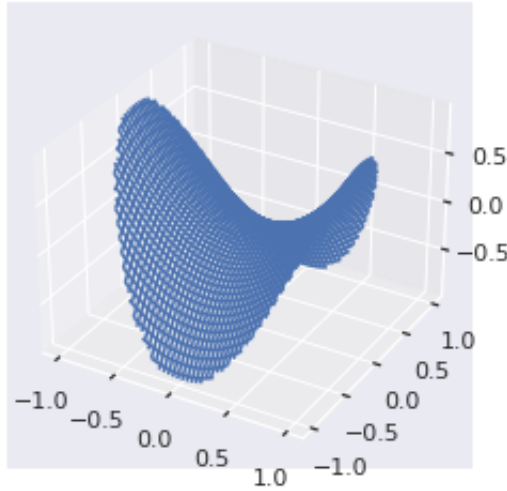
8 Задание 8

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

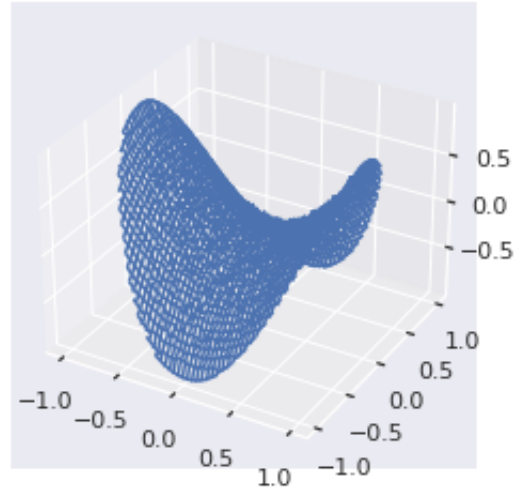
$$\begin{cases} \Delta u = 0, (x, y) \in D, \\ u|_{\delta D} = f(x, y), \\ u \in C^2(D), f \in C(\delta D), \\ D = \{x, y : x^2 + y^2 \leq 1\} \end{cases} \quad (10)$$

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

Численное решение уравнения Дирихле (10) в круге D можно получить таким способом. Для каждой точки M из заданной заранее сетки N раз проводится следующая процедура. Сначала строится окружность с центром в точке M и максимальным радиусом, таким что эта окружность еще принадлежит замыканию D . На



(a) Аналитическое решение



(b) Численное решение ($N = 1000$)

Рис. 14

окружности разыгрывается случайная равномерно распределенная точка M_1 . Если $\rho(M_1, \partial D) < \varepsilon$, то процесс обрывается и выбирается граничное значение $f(Q^1)$, где Q^1 — ближайшая граничная точка, иначе M_1 принимается за следующую точку и процедура продолжается.

После того как такой процесс блуждания производится N раз считается среднее арифметическое

$$\frac{f(Q^1) + f(Q^2) + \dots + f(Q^N)}{N},$$

являющееся приближенным значением искомого решения $u(M)$.

Аналитическим решением (10) для $f(x, y) = x^2 - y^2$ является, очевидно, $u = x^2 - y^2$. Сравните аналитическое решение с результатом работы описанного выше алгоритма на рис. 14

9 Задание 9

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. Начальные значения X_0 генерируются случайным образом так, чтобы полученный процесс был стационарным.

Для данных гауссовских процессов

1. Найти ковариационную функцию и переходные вероятности.

2. Моделировать 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 Винеровский процесс

По определению ([1]) ковариационная функция Винеровского процесса W_t имеет вид $\text{cov}(t, s) = \min(t, s)$. Плотность многомерного нормального распределения $\mathcal{N}(m, R)$

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|R|}} e^{-\frac{1}{2}(x-m)^T R^{-1}(x-m)},$$

где R — ковариационная матрица. Для моделирования Винеровского процесса достаточно знать, что $W_0 = 0, W_1 \sim \mathcal{N}(0, 1)$ и для $t_0, t_1, \alpha \in (0, 1)$ условное распределение в момент $t = (1 - \alpha)t_0 + \alpha t_1$ имеет плотность

$$p_{W_t|W_{t_0}, W_{t_1}}(x | x_0, x_1) = \frac{p_{W_{t_0}, W_t, W_{t_1}}(x_0, x, x_1)}{p_{W_{t_0}, W_{t_1}}(x_0, x_1)},$$

где

$$\begin{aligned} p_{W_{t_0}, W_t, W_{t_1}} &= \frac{1}{(2\pi)^{\frac{3}{2}} \sqrt{|R_3|}} e^{-\frac{1}{2}x^T R_3^{-1}x}, \\ p_{W_{t_0}, W_{t_1}} &= \frac{1}{(2\pi) \sqrt{|R_2|}} e^{-\frac{1}{2}x^T R_2^{-1}x}, \\ R_3 &= \begin{pmatrix} t_0 & t_0 & t_0 \\ t_0 & t & t \\ t_0 & t & t_1 \end{pmatrix}, R_2 = \begin{pmatrix} t_0 & t_0 \\ t_0 & t_1 \end{pmatrix}. \end{aligned}$$

Отсюда

$$p_{W_t|W_{t_0}, W_{t_1}}(x | x_0, x_1) = \frac{1}{\sqrt{2\pi\alpha(1-\alpha)(t_1 - t_0)}} e^{-\frac{(x - ((1-\alpha)x_0 + \alpha x_1))^2}{2\alpha(1-\alpha)(t_1 - t_0)}},$$

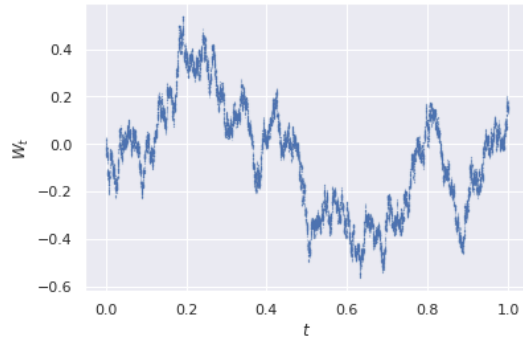
т.е. $W_t \sim \mathcal{N}(x, \alpha(1-\alpha)(t_1 - t_0))$. В случае $\alpha = \frac{1}{2}$ $W_t \sim \mathcal{N}(x, \frac{t_1 - t_0}{4})$

Таким образом, для моделирования Винеровского процесса достаточно получить некоторые значения $W_0 = 0, W_1 \sim \mathcal{N}(0, 1)$, и затем последовательно методом бисекции разыгрывать величины $W_t \sim \mathcal{N}(x, \frac{t_1 - t_0}{4})$.

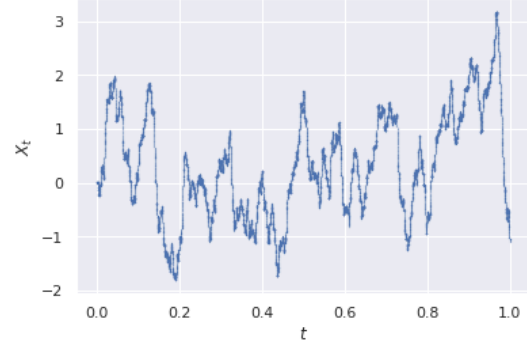
9.2 Процесс Орнштейна-Уленбека

Из стационарности процесса Орнштейна-Уленбека X_t следует, что $\mathbb{E} X_t = a$, $\mathbb{D} X_t = \sigma^2$ и

$$\text{cov}(X_t, X_s) = \sigma^2 \rho(x, y) = \sigma^2 \rho(x) \rho(y) = \sigma^2 e^{-\theta(x+y)}$$



(a) Винеровский процесс



(b) Процесс Орнштейна-Уленбека

Рис. 15

Условное распределение имеет вид $(X_t | X_s = x) \sim \mathcal{N}(xe^{\theta|t-s|}, \sigma^2(1 - e^{-2\theta|t-s|}))$.

Процесс Орнштейна-Уленбека моделируется так же как и Винеровский, лишь с учетом того, что $X_0 \sim \mathcal{N}(a, \sigma^2)$, $X_1 \sim \mathcal{N}(x_0 e^{-\theta T}, \sigma^2(1 - e^{-2\theta T}))$, и

$$X_{\frac{t_0+t_1}{2}} \sim \mathcal{N}\left((x_0 + x_1) \frac{e^{-\frac{\theta(t_1-t_0)}{2}}}{1 + e^{-\theta(t_1-t_0)}}, \sigma^2 \frac{1 - e^{-\theta(t_1-t_0)}}{1 + e^{-\theta(t_1-t_0)}}\right)$$

Результат моделирования процессов см. рис. 15

10 Задание 10

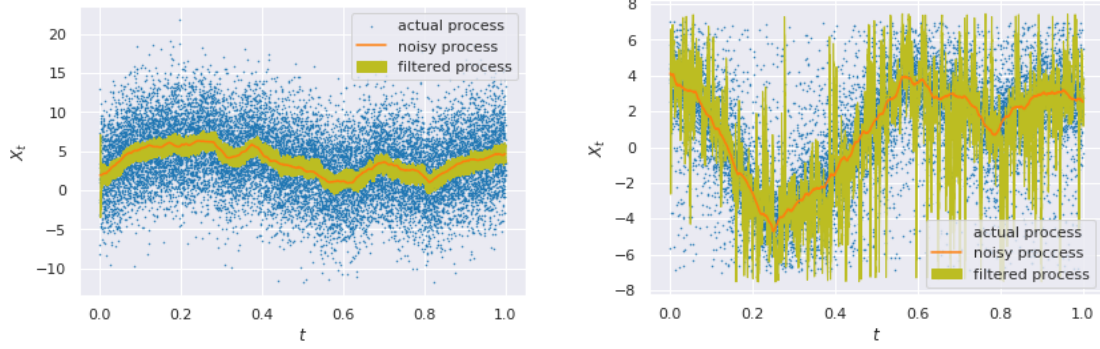
Произвести фильтрацию одномерного процесса Орнштейна-Уленбека:

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.

Белый шум для случайного процесса $x_{k+1} = A_k x_k$ будем моделировать линейным стохастическим уравнением

$$x_{k+1} = A_k x_k + \omega_k, \quad (11)$$

где ω_k — случайная помеха, распределенная нормально или по Коши.



(a) Гауссовский шум

(b) Шум распределенный по Коши

Рис. 16: Фильтрация Калмана

Фильтр Калмана для задачи (11) имеет вид

$$\begin{cases} \hat{x}_{k|k} = \hat{x}_{k|k-1} + R_{k|k-1} C_k^T (C_k R_{k|k-1} C_k^T + N_k)^{-1} (y_k - C_k \hat{x}_{k|k-1} - \mathbb{E} v_k), \\ \hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + \mathbb{E} \omega_k, \\ R_{k|k} = R_{k|k-1} - R_{k|k-1} C_k^T (C_k R_{k|k-1} C_k^T + N_k)^{-1} C_k R_{k|k-1}, \\ R_{k+1|k} = A_k R_{k|k} A_k^T + M_k, \\ \hat{x}_{0|-1} = \bar{x}_0, R_{0|-1} = s, \end{cases} \quad (12)$$

где $y_k = C_k x_k + v_k$ — измерения, $x_{k+1} = A_k x_k + \omega_k$ — шаги. В случае процесса Орнштейна-Уленбека с параметрами λ, σ система (12) принимает вид

$$\begin{cases} \hat{x}_{k|k} = \hat{x}_{k|k-1} + R_{k|k-1} (R_{k|k-1} \sigma_v^2)^{-1} (y_k - \hat{x}_{k|k-1}), \\ \hat{x}_{k+1|k} = e^{-\lambda \Delta t} \hat{x}_{k|k}, \\ R_{k|k} = R_{k|k-1} - R_{k|k-1} (R_{k|k-1} + \sigma_v^2)^{-1} R_{k|k-1}, \\ R_{k+1|k} = e^{-\lambda \Delta t} R_{k|k} A_k^T + \sigma^2 (1 - e^{-2\lambda \Delta t}), \\ \hat{x}_{0|-1} = 0, R_{0|-1} = \sigma^2, \end{cases} \quad (13)$$

Доверительный интервал фильтрации составляет

$$-k_{1-\frac{\alpha}{2}} \sqrt{R_{k|k}} < \hat{x}_{k|k} < k_{1-\frac{\alpha}{2}} \sqrt{R_{k|k}}, \quad (14)$$

где $k_{1-\frac{\alpha}{2}}$ — $(1 - \frac{\alpha}{2})$ -квантиль нормального распределения.

Пример работы программы см рис. 16.

11 Задание 11

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 её обслуживания (распределение χ^2 с 10-ю степенями свободы).
2. Вторая интерпретация: система массового обслуживания с циклической интенсивностью $\lambda(t) = \lambda_0(1 + \cos(t))$ и единичными скачками. Свести данную задачу моделирования неоднородного пуассоновского процесса при помощи метода Льюиса и Шедлера 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 всевозможные случаи поведения системы в зависимости от значения параметров.

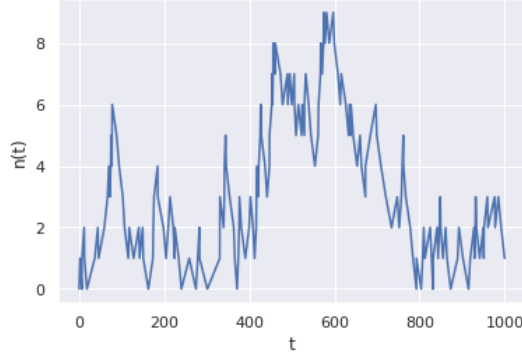
11.1 СМО

Систему массового обслуживания будем моделировать временами поступления заявок t_i : $t_i - t_{i-1} \sim \text{Exp}(\lambda)$, где λ — интенсивность потока заявок, и временами обработки заявок s_i распределенных по закону χ^2 с 10 степенями свободы. Для каждой заявки будет считаться время ее исполнения, т.е. если следующая заявка пришла раньше чем прошло время обработки заявки то будет накапливаться очередь. Поскольку $\mathbb{E} s_i = 10$, в зависимости того, больше λ чем $\frac{1}{10}$ или меньше, очередь будет в среднем накапливаться или сокращаться соответственно.

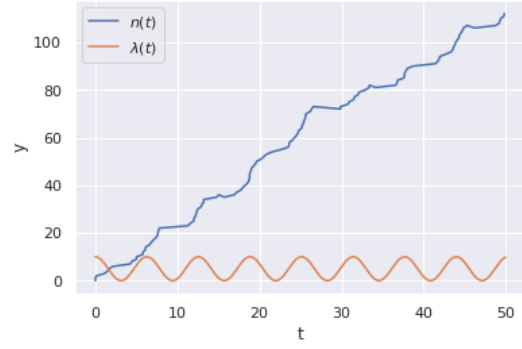
11.2 СМО с циклической интенсивностью

Имеющийся неоднородный процесс с помощью вышеупомянутого алгоритма сводится к двумерному пуассоновскому полю, а именно, генерируется стационарный пуассоновский процесс X_t , из множества точек роста которого случайно (с бернуллиевским распределением) выбирается подмножество T . Совокупность полученных величин образует искомое пуассоновское поле.

Иллюстрации работы моделей СМО представлены на рис. 17



(a) с постоянной интенсивностью



(b) с циклической интенсивностью

Рис. 17: СМО

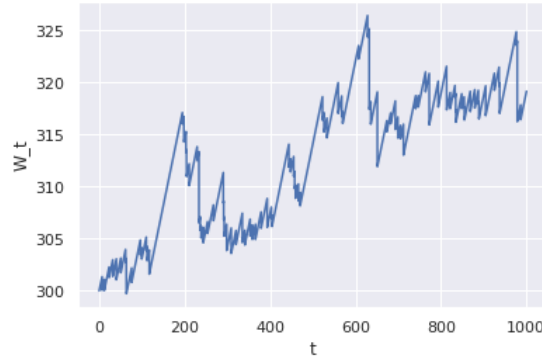


Рис. 18: работа страховой компании

11.3 Работа страховой компании

Для построения модели работы страховой компании генерируются времена поступления страховых случаев t_i : $t_i - t_{i-1} \sim \text{Exp}(\lambda)$ и ущерб страхового случая s_i распределенный по Парето с параметрами x_m и k . По ним можно вычислить величину капитала компании в момент t : $W(t) = W_0 + cy - \sum_{i:t_i < t} s_i$ и среднюю скорость прироста капитала $(\mathbb{E} W(t))' = c - \frac{\lambda k x_m}{k-1}$. Из чего можно заключить, что страховая компания будет расти при $c(k-1) - \lambda k x_m > 0$, и обанкротится при $c(k-1) - \lambda k x_m < 0$.

Список литературы

- [1] Смирнов С. Н. кафедральный курс «*Стохастический анализ и моделирование*», 2021.
- [2] Ширяев А. Н. «*Вероятность*». Изд-во Наука, Москва, 1979.
- [3] Лагутин М. Б. «*Наглядная математическая статистика*». Изд-во Бином, Москва, 2009.
- [4] Бусленко Н. П., Шрейдер Ю. А. «*Метод статистических испытаний*». Изд-во физ-мат литературы, Москва, 1961.