



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, 2021

## Содержание

<b>1</b>	<b>Задание 1</b>	<b>3</b>
1.1	Реализация схемы Бернулли и биномиального распределения . . . . .	3
1.2	Геометрическое распределение . . . . .	3
1.3	Игра в орлянку . . . . .	4
<b>2</b>	<b>Задание 2</b>	<b>5</b>
2.1	Датчик для канторова распределения . . . . .	5
2.2	Свойства симметричности и самоподобия . . . . .	7
2.3	Математическое ожидание и дисперсия . . . . .	8
<b>3</b>	<b>Задание 3</b>	<b>9</b>
3.1	Экспоненциальное распределение . . . . .	10
3.2	Датчик пуассоновского распределения 1 . . . . .	11
3.3	Датчик пуассоновского распределения 2 . . . . .	12
3.4	Датчик стандартного нормального распределения 1 . . . . .	13
<b>4</b>	<b>Задание 4</b>	<b>14</b>
4.1	Датчик распределения Коши . . . . .	15
4.2	Датчик стандартного нормального распределения 2 . . . . .	15
4.3	Сравнение реализаций датчика стандартного нормального распределения . . . . .	16
<b>5</b>	<b>Задание 5</b>	<b>16</b>
5.1	Проверка ЗБЧ и ЦПТ . . . . .	17
5.2	Доверительные интервалы . . . . .	17
5.3	Поведение сумм $S_n/n$ для распределения Коши . . . . .	18
<b>6</b>	<b>Задание 6</b>	<b>19</b>
6.1	Численное интегрирование . . . . .	19
6.2	Точность вычислений . . . . .	20

## 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

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  $\xi_i$  через  $\eta$  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. 1a).

### 1.2 Геометрическое распределение

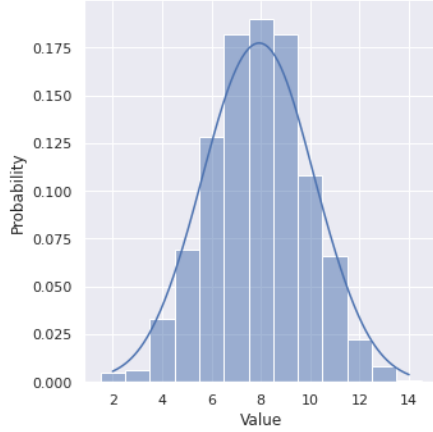
Случайная величина  $\gamma \sim \text{Geom}(p)$  представима как

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

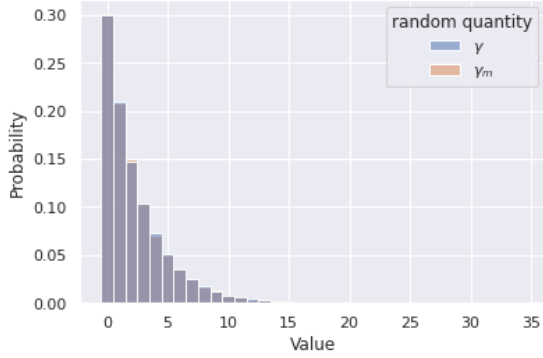
Геометрическое распределение обладает свойством отсутствия памяти, т.е.

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

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.  $\gamma$  означает, что



(a) Биномиальное распределение



(b) Отсутствие памяти ( $m = 5$ )

Рис. 1

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

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  $\gamma$  отличается ровно на константу  $m$  от распределения сл.в.  $\gamma$ , индуцированной на вероятностное подпространство  $\Omega_m$ .

Проверка свойства отсутствия памяти проведена численным моделированием распределений сл.в.  $\gamma$  и  $\gamma_m$  (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.

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

**Теорема** (Центральная предельная теорема). Пусть  $\xi_1, \xi_2, \dots$  — последовательность независимых одинаково распределенных (невыврожденных) случайных величин с  $\mathbb{E}\xi_1^2 \leq \infty$  и  $S_n = \xi_1 + \dots + \xi_n$ . Тогда

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

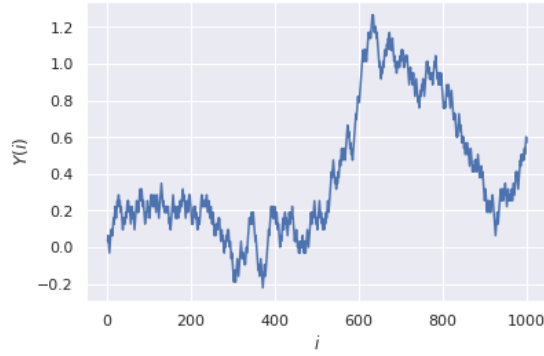


Рис. 2: Игра в орлянку ( $n = 1000$ )

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

Носителем канторова распределения является счетное пересечение множеств

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

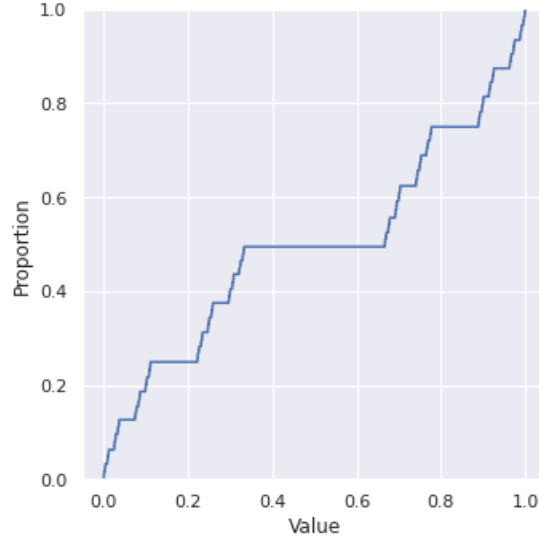


Рис. 3: Э.д.ф. канторовой сл.в.

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  $\xi \sim \text{Bern}(0.5)$

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

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*  $\alpha$  отвергается, если на полученной выборке  $(X_1, \dots, X_n)$  значение статистики неправдоподобно велико, т.е.

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

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 (1) 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  $\alpha$  больший чем величина  $\alpha(X)$ . И обратно, гипотеза *может быть принята* для любого уровня значимости  $\alpha$  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  $\delta$  статистика  $D_n$  составляет приблизительно 0.0087. Величина  $1 - K(\sqrt{n}D_n) = \alpha(X)$  равно приблизительно 0.9999, что даёт нам основания принять гипотезу о корректности построенного датчика  $\delta$ , 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

$$\begin{aligned} D_{n,m} &= \sup_x \left| \hat{F}_n(x) - \hat{G}_n(x) \right|, \\ \text{где } \hat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x), \quad \hat{G}_n(x) = \frac{1}{m} \sum_{j=1}^m \mathbb{I}(Y_j \leq x). \end{aligned}$$

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

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

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

В результате компьютерных вычислений для случайных величин  $\delta$  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  $\delta$  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

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



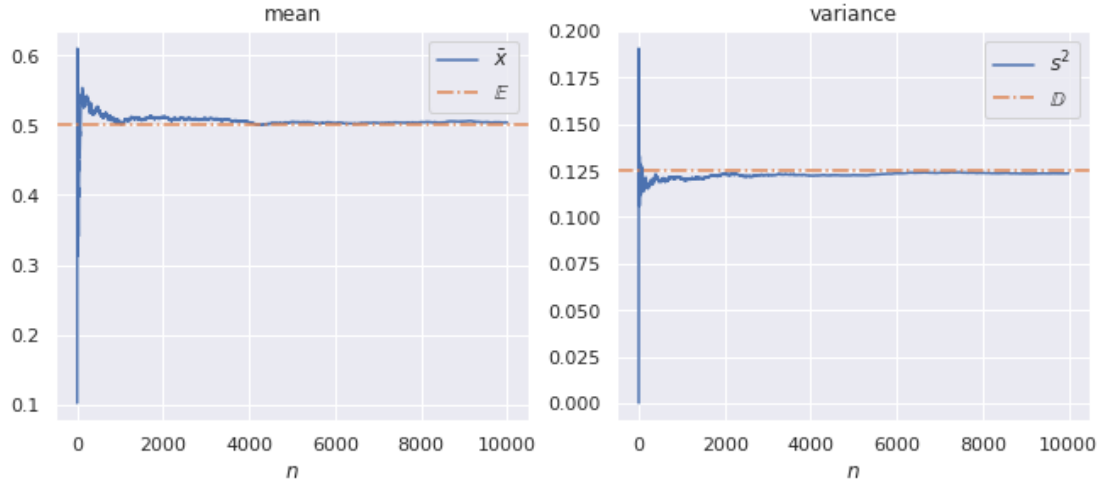


Рис. 4

Следовательно

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

Аналогично дисперсия:

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

Т.е.

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

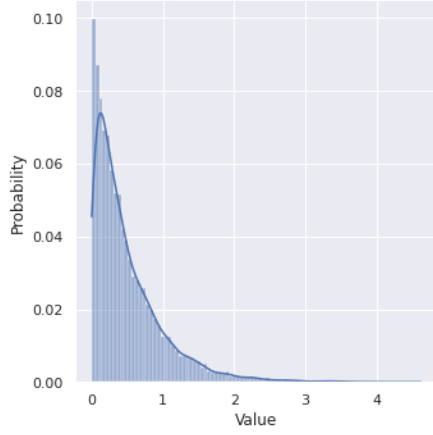
и

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

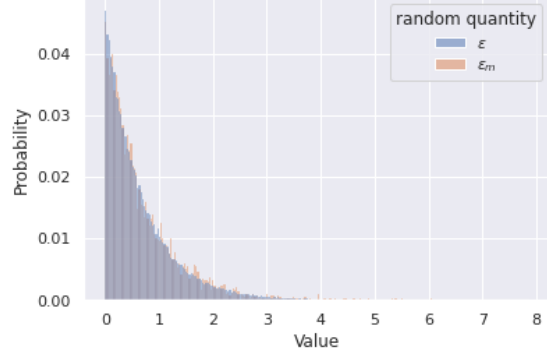
Сходимость эмпирических показателей к теоретическим показана на рис. 4.

### 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  $Y = \min(X_1, X_2, \dots, X_n)$ .



(a) Экспоненциальное распределение



(b) Отсутствие памяти ( $m = 2$ )

Рис. 5

2. На основе датчика экспоненциального распределения построить датчик пуассоновского распределения.
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

Напомним, что  $\eta \sim \text{Uni}(0, 1)$ , тогда для  $\varepsilon \sim \text{Exp}(\lambda)$  справедливо <sup>1</sup>

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

Результат моделирования см. рис. 5а

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

<sup>1</sup>см. [3] гл. 4 §1 “Метод обратной функции”.

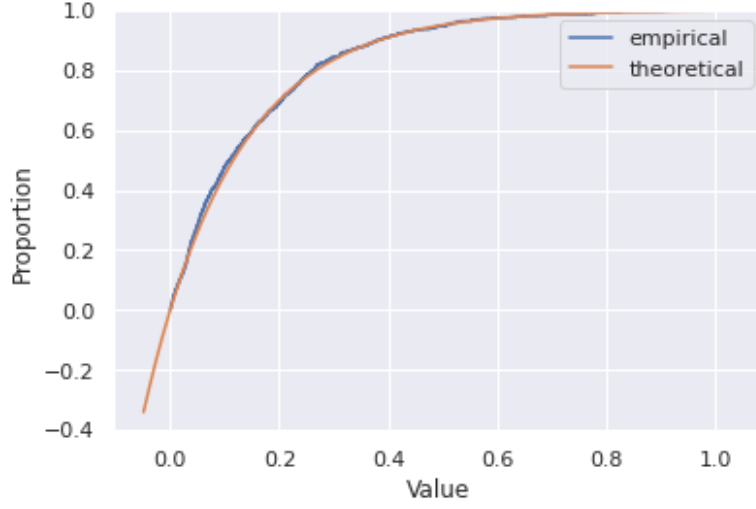


Рис. 6

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

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

Таким образом

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

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

### 3.2 Датчик пуассоновского распределения 1

Для  $\pi \sim \text{Pois}(\lambda)$  верно следующее представление <sup>2</sup>

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

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

Результат моделирования см. 7

---

<sup>2</sup>см. [3] гл. 5 §1 “Моделирование дискретных величин”.

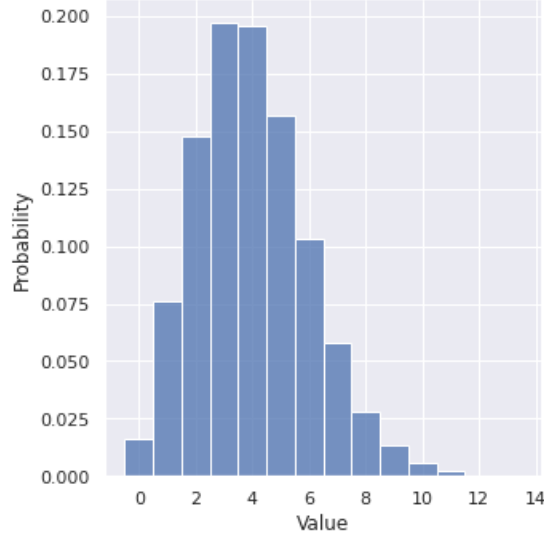


Рис. 7

### 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. 8.

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

Пусть  $X_1, \dots, X_n$  — выборка. Разобьем множество значений  $\xi_1$  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  $\nu_j$  — равными количеству элементов выборки в  $\Delta_j$ .

Так, статистикой критерия Пирсона служит величина

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

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

Поскольку распределение Пуассона дискретно, промежутки  $\Delta_j$  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%.

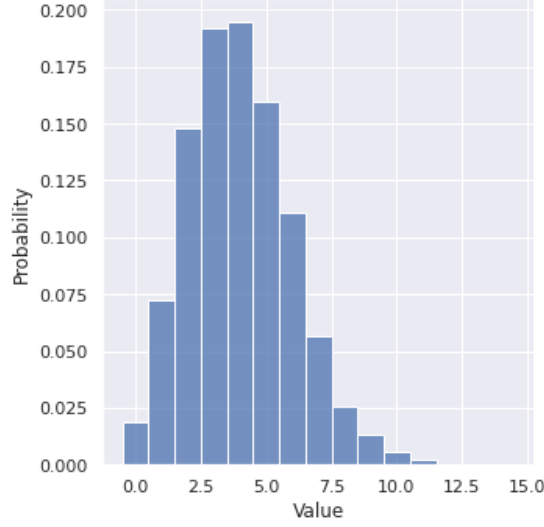


Рис. 8

### 3.4 Датчик стандартного нормального распределения 1

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. 9.

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  $\alpha$  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

$$\frac{S_1^2}{S_2^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{m-1} \sum_{j=1}^m (Y_j - \bar{Y})^2}, \quad (3)$$

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  $\xi \sim \chi_{n-1}^2, \eta \sim \chi_{m-1}^2$ ,  $\xi$  и  $\eta$  независимы.

В результате вычисления статистики (3) гипотеза о соответствии дисперсий принимается с уровнем значимости  $\alpha = 5\%$ .

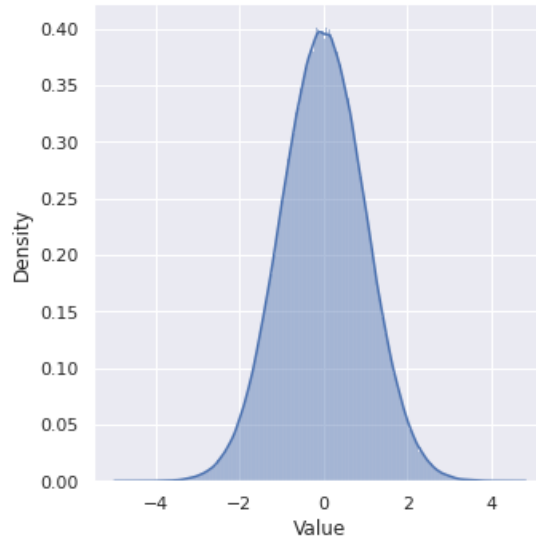


Рис. 9

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

$$\frac{\sqrt{\frac{nm}{n+m}}(\bar{X} - \bar{Y})(n + m - 2)}{[(n - 1)S_1^2 + (m - 1)S_2^2]} \sim t_{n+m-2}, \quad (4)$$

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.

## 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 probabity plot` убедиться в корректности построенного датчика и обосновать наблюдаемую линейную зависимость.
3. Сравнить скорость моделирования стандартного нормального распределения в заданиях 3 и 4.

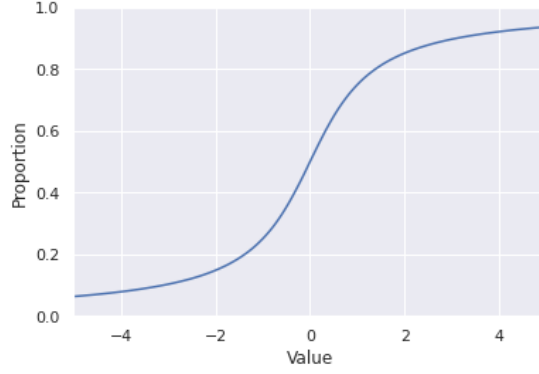


Рис. 10

#### 4.1 Датчик распределения Коши

Для моделирования абсолютно непрерывных распределений используется метод обратной функции, основанный на том факте, что случайная величина

$$\xi = F^{-1}(\eta), \text{ где } \eta \sim U(0, 1)$$

имеет функцию распределения  $F(x)$ . ( $F(x)$  — непрерывная, строго возрастающая функция распределения.)

Функция распределения Коши  $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctg(\frac{x-x_0}{\gamma})$ . Обратная к ней  $F^{-1}(y) = x_0 + \gamma \operatorname{tg}[\pi(x - \frac{1}{2})]$ . Таким образом искомая формула для моделирования  $\xi \sim \text{Cauchy}(x_0, \gamma)$  имеет вид

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

Пример результата работы полученного датчика представлен на рис.10.

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

Результат моделирования см. на рис. 11.

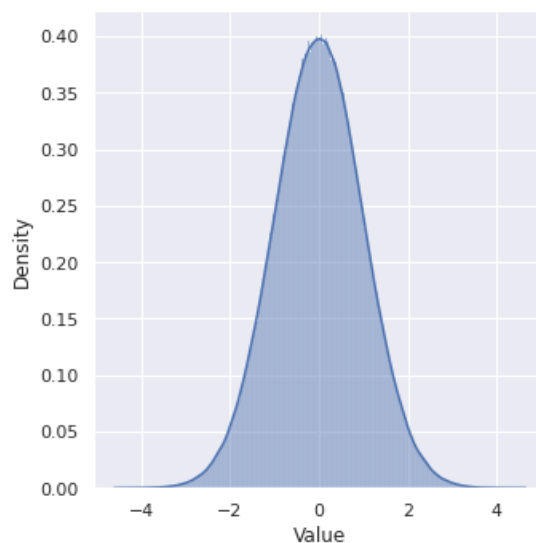


Рис. 11

### 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  $\mu$  и  $\sigma^2$  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.



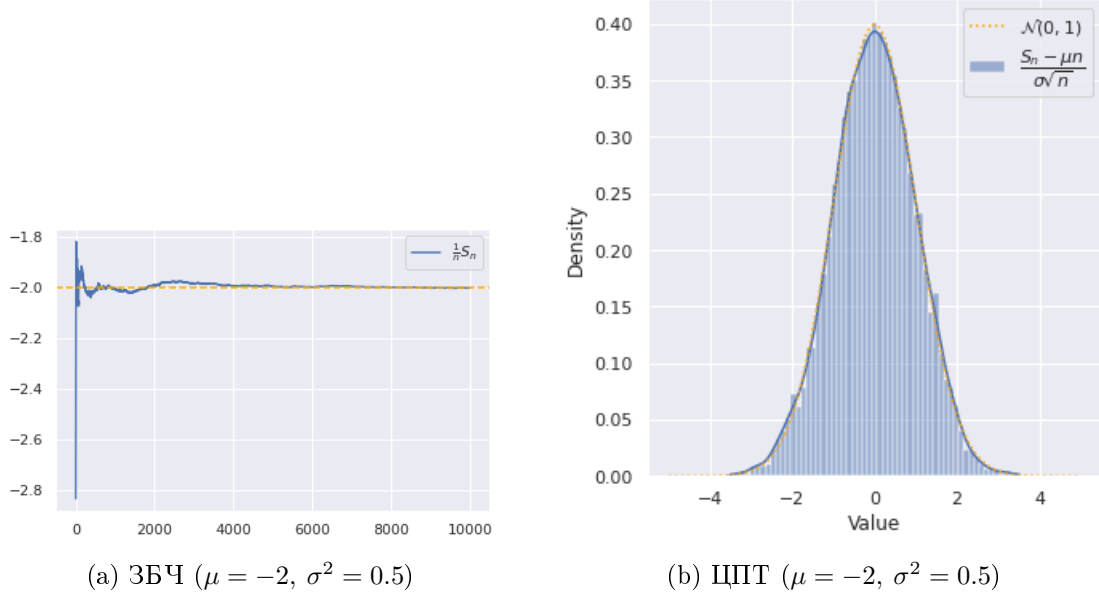


Рис. 12

## 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. 12a), т.е

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

Результаты опыта соответствуют и центральной предельной теореме, т.е

$$\frac{S_n - \mu n}{\sigma \sqrt{n}} \xrightarrow{d} \eta \sim \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \text{ (см. рис. 12b)}$$

## 5.2 Доверительные интервалы

Пусть имеется выборка  $X_i \sim \mathcal{N}(\mu, \sigma^2)$ , где  $\mu$  и  $\sigma^2$  — 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  $\mu$ :

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

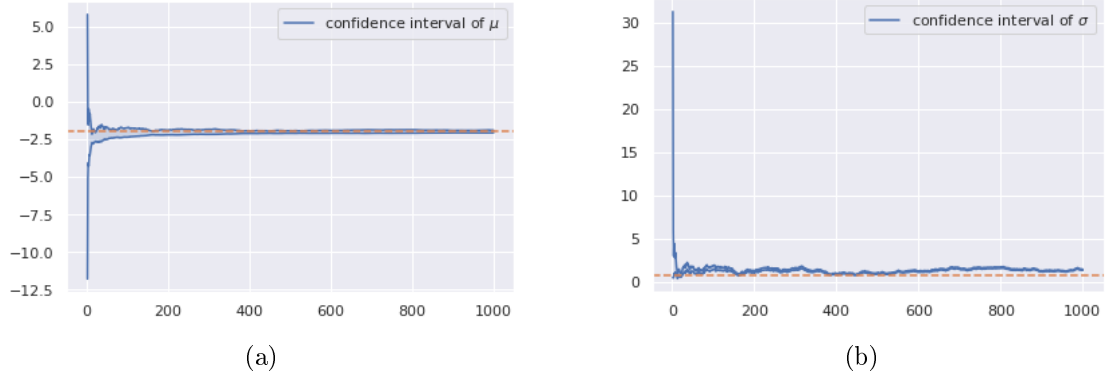


Рис. 13

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

- и для параметра масштаба  $\sigma$ :

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

где  $z_p$  —  $p$ -квантиль закона  $\chi^2_{n-1}$ .

Доверительные интервалы для некоторой выборки из  $\mathcal{N}(-2, 0.5)$  для разных  $n$  приведены на рис. 13

### 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 выбороками (расходимость в целом) от  $a$  ее отклонившими. График сумм  $S_n/n$  для некоторой выборки  $X_i$  представлен на рис. 14.

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

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

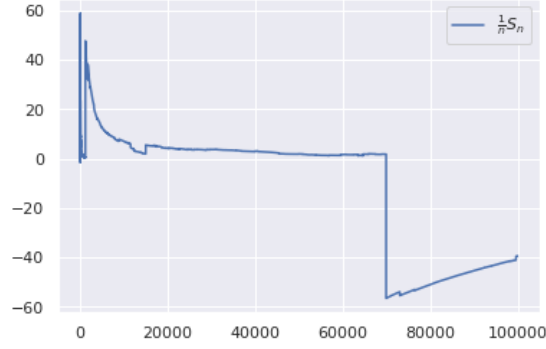


Рис. 14:  $a = 0, b = 3$

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

$$\psi_{\frac{S_n}{n}}(t) = \psi_{\sum \frac{1}{n} X_i}(t) = \prod \psi_{\frac{1}{n} X_i}(t) = \prod \psi_{X_i}\left(\frac{t}{n}\right) = e^{n(ai\frac{t}{n} - b|\frac{t}{n}|)} = e^{ait - b|t|}. \quad (5)$$

Поскольку хар. функция однозначно задает распределение, равенство (5) 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

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{e^{-\left(x_1^2 + \dots + x_{10}^2 + \frac{1}{2^7 \cdot x_1^2 \cdot \dots \cdot x_{10}^2}\right)}}{x_1^2 \cdot \dots \cdot x_{10}^2} dx_1 dx_2 \dots dx_{10} \quad (6)$$

— методом Монте-Карло

— методом квадратур, сводя задачу к вычислению собственного интеграла Римана

2. Для каждого случая оценить точность вычислений.

### 6.1 Численное интегрирование

Перепишем (6) следующим образом

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

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 (6).

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

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

Тогда (6) примет вид

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

Интеграл (7) уже можно вычислить, например, стандартным методом прямоугольников на равномерной сетке.

## 6.2 Точность вычислений

## 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.