



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

1 Задание 1	3
1.1 Реализация схемы Бернулли и биномиального распределения	3
1.2 Геометрическое распределение	3
1.3 Игра в орлянку	4
2 Задание 2	5
2.1 Датчик для канторова распределения	5
2.2 Свойства симметричности и самоподобия	7
2.3 Математическое ожидание и дисперсия	8
3 Задание 3	9
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	12
4 Задание 4	12
4.1 Датчик распределения Коши	13
4.2 Датчик стандартного нормального распределения 2	13
4.3 Сравнение реализаций датчика стандартного нормального распределения	13
5 Задание 5	13
5.1 Проверка ЗБЧ и ЦПТ	14
5.2 Доверительные интервалы	14
5.3 Поведение сумм S_n/n для распределения Коши	14
6 Задание 6	14
6.1 Численное интегрирование	14
6.2 Точность вычислений	14

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 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 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 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

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. γ означает, что

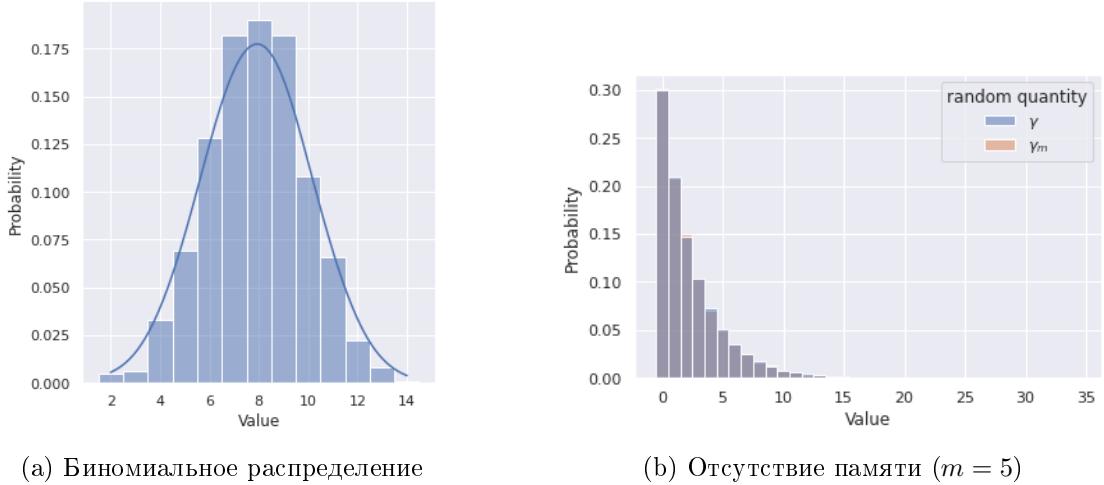


Рис. 1

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

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 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.

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

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

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

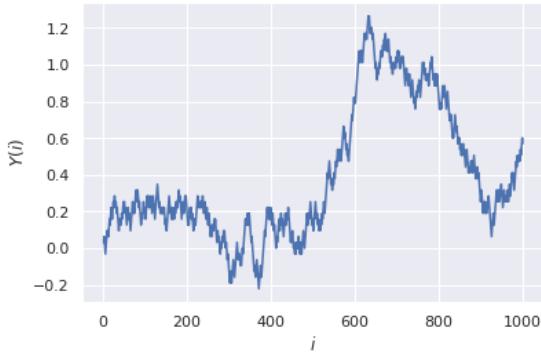


Рис. 2: Игра в орлянку ($n = 1000$)

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

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

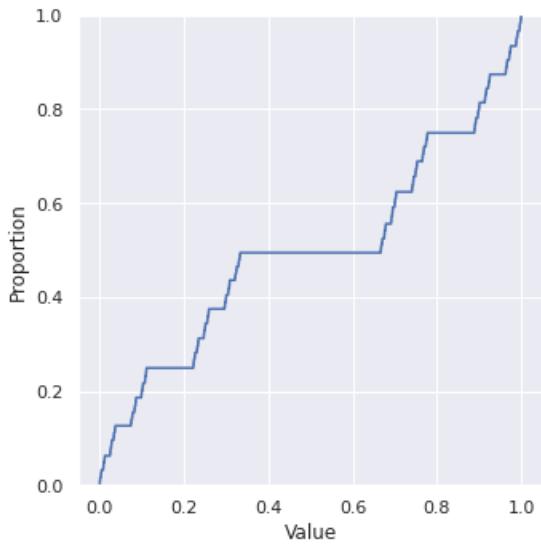


Рис. 3: Э.д.ф. канторовой сл.в.

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

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

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 $x > 0$*

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

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

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

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 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 $1 - K(\sqrt{n}D_n) = \alpha(X)$ равно 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*

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

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 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 одинаковой 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

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

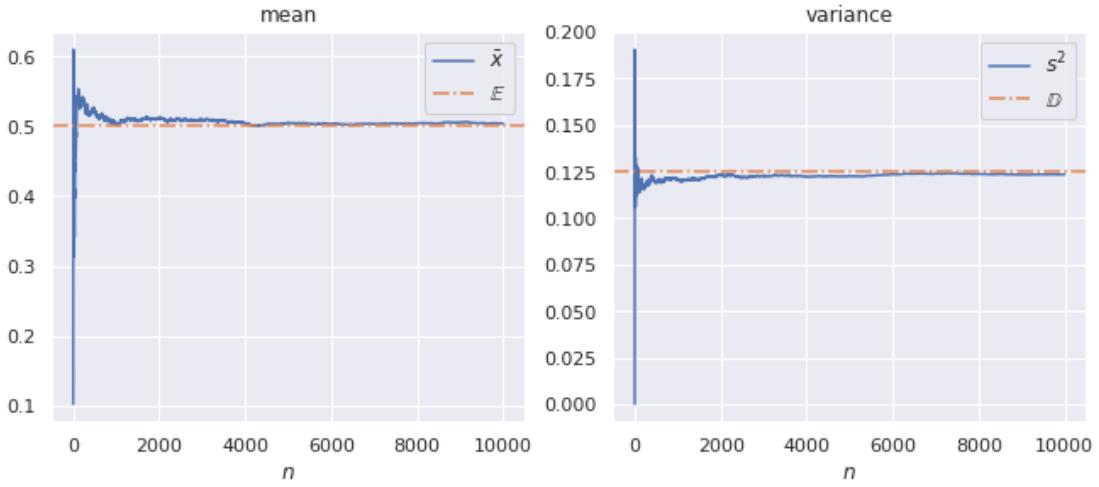


Рис. 4

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

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

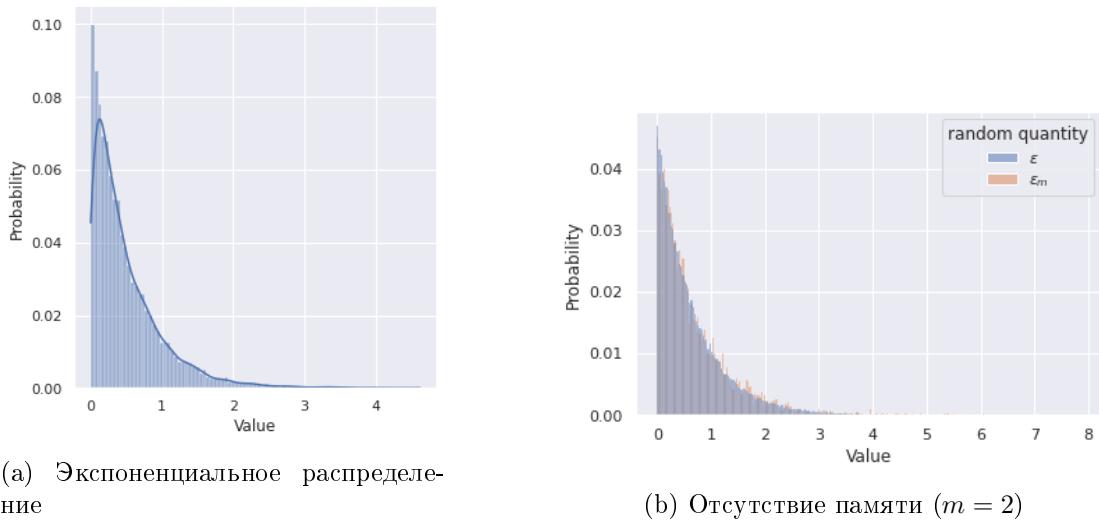


Рис. 5

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 ¹

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

Результат моделирования см. рис. 5а

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

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

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

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

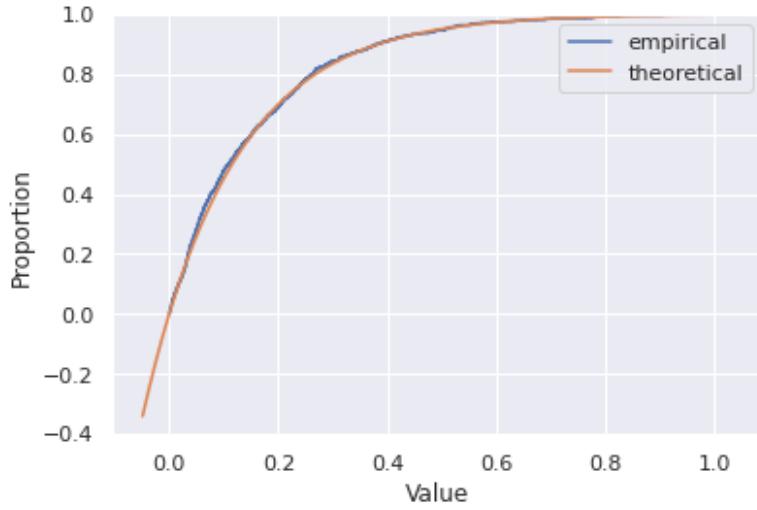


Рис. 6

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

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

Таким образом

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

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

3.2 Датчик пуссоновского распределения 1

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. 7

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

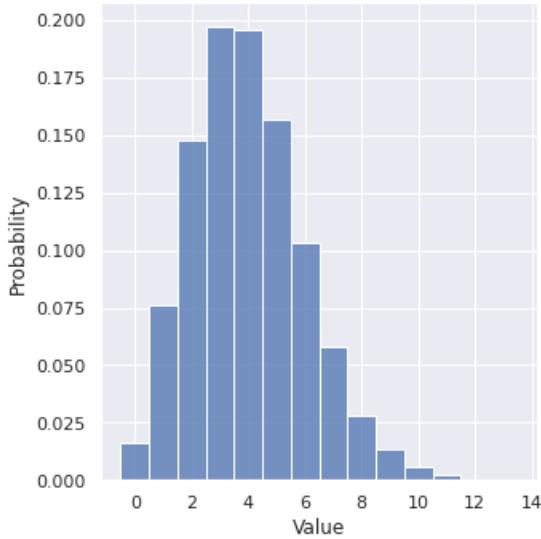


Рис. 7

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 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 .

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 χ_{N-1}^2 .

Промежутки Δ_j зададим как набор равновероятных интервалов на прямой, N положим равным приблизительно $\log_2 n$.

3.4 Датчик стандартного нормального распределения 1

4 Задание 4

- Построить датчик распределения Коши.

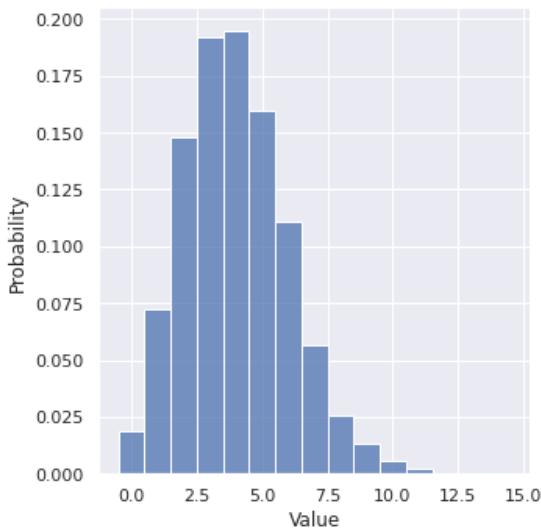


Рис. 8

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

4.2 Датчик стандартного нормального распределения 2

4.3 Сравнение реализаций датчика стандартного нормального распределения

5 Задание 5

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

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

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

5.2 Доверительные интервалы

5.3 Поведение сумм S_n/n для распределения Коши

6 Задание 6

1. Посчитать интеграл

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

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

2. Для каждого случая оценить точность вычислений.

6.1 Численное интегрирование

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.