



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

Практикум

# «Стохастический анализ и моделирование»

*Студент 415 группы*  
К. Ю. Егоров

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

Москва, 2019

## Содержание

<b>1</b>	<b>Задание №1</b>	<b>3</b>
1.1	Задача №1 . . . . .	3
1.2	Задача №2 . . . . .	4
1.3	Задача №3 . . . . .	6
<b>2</b>	<b>Задание №2</b>	<b>7</b>
2.1	Задача №1 . . . . .	7
2.2	Задача №2 . . . . .	9
2.3	Задача №3 . . . . .	9

## 1 Задание №1

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

### 1.1 Задача №1

**Определение 1.1.** *Схемой Бернулли* называется последовательность испытаний, в каждом из которых возможны два исхода — «успех» и «неудача», при этом «успех» в каждом испытании происходит с одной и той же вероятностью  $p \in (0, 1)$ , а «неудача» — с вероятностью  $q \equiv 1 - p$ . На испытания в схеме Бернулли налагаются следующие требования: отсутствие взаимного влияния, воспроизводимость, а также сходные — но не идентичные — условия проведения.

**Определение 1.2.** Будем говорить, что случайная величина  $X$  имеет *распределение Бернулли*, если она принимает всего два значения: 1 и 0 с вероятностями  $p$  и  $q \equiv 1 - p$  соответственно. Таким образом,

$$\mathbb{P}(X = 1) = p \quad \text{и} \quad \mathbb{P}(X = 0) = q,$$

то есть событие  $\{X = 1\}$  соответствует «успеху», а  $\{X = 0\}$  — «неудаче». Будем обозначать такую случайную величину

$$X \sim \text{Bern}(p).$$

Реализуем генератор схемы Бернулли с заданной вероятностью успеха  $p$  следующим образом: пусть нам дана случайная величина  $\xi$ , равномерно распределённая на отрезке  $[0, 1]$ . Тогда случайная величина  $X \sim \text{Bern}(p)$  задаётся следующим образом:

$$X = \mathbb{I}(\xi < p) = \begin{cases} 1, & \text{при } 0 \leq \xi < p, \\ 0, & \text{при } p \leq \xi \leq 1. \end{cases}$$

**Определение 1.3.** Будем говорить, что случайная величина  $X$  имеет *биномиальное распределение* с параметрами  $n$  и  $p$ , если

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad \text{где } k \in \mathbb{N}_0.$$

В таком случае  $X$  интерпретируют как число «успехов» в серии из  $n$  испытаний схемы Бернулли с вероятностью успеха  $p$ . Будем обозначать такую случайную величину

$$X \sim \text{Bi}(n, p).$$

Пусть теперь  $X \sim \text{Bi}(n, p)$ , а  $Y_i \sim \text{Bern}(p), i = \overline{1, n}$ . Тогда, как видно из интерпретации биномиального распределения, датчик биномиальной случайной величины будет иметь вид:

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

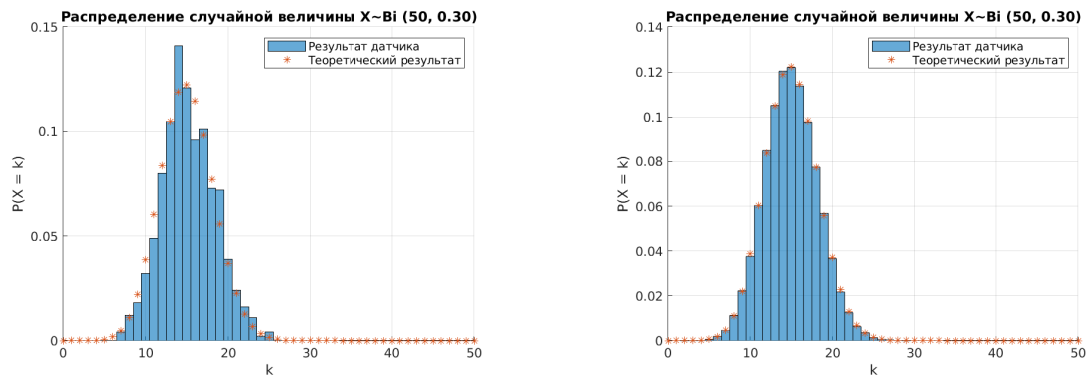


Рис. 1.1: Гистограмма биномиального распределения при  $10^3$  (слева) и  $10^5$  (справа) испытаний.

## 1.2 Задача №2



Рис. 1.2: Гистограмма геометрического распределения при  $10^5$  испытаний.

**Определение 1.4.** Будем говорить, что случайная величина  $X$  имеет *геометрическое распределение*, если

$$\mathbb{P}(X = k) = (1 - p)^k p = q^k p, \quad \text{где } k \in \mathbb{N}_0.$$

Геометрически распределенная случайная величина интерпретируется как количество «неудач» до первого «успеха» в схеме испытаний Бернулли с вероятностью  $p$ . Будем обозначать такие случайные величины

$$X \sim \text{Geom}(p).$$

Зная интерпретацию, мы легко строим датчик и для геометрического распределения.

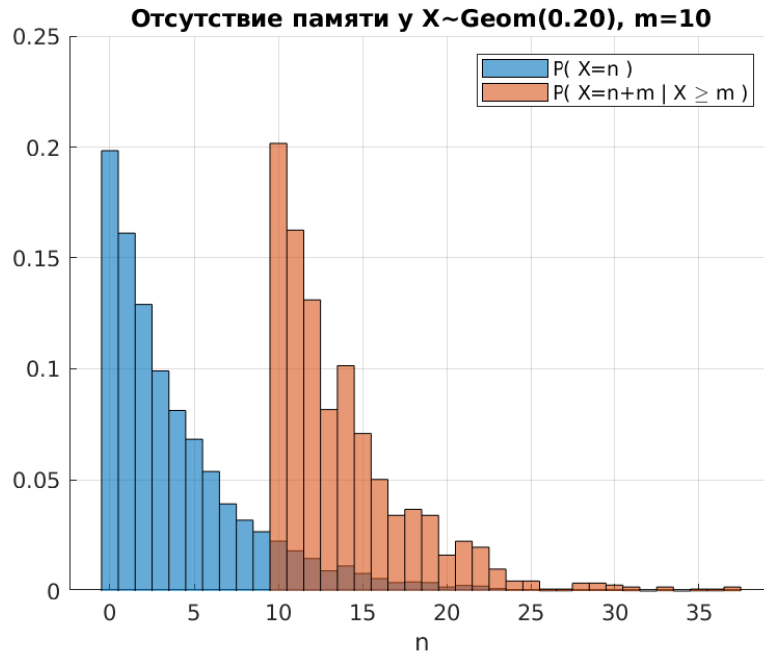


Рис. 1.3: Гистограмма геометрического распределения, демонстрирующая его свойство отсутствия памяти.

**Утверждение 1.1** (Свойство отсутствия памяти). Пусть  $X \sim \text{Geom}(p)$ , тогда для любых  $n, m \in \mathbb{N}_0$  справедливо

$$\mathbb{P}(X \geq m + n \mid X \geq m) = \mathbb{P}(X \geq n),$$

то есть количество прошлых «неудач» не влияет на количество будущих «неудач».

**Доказательство.** Рассмотрим левую часть равенства из условия утверждения:

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

Теперь рассмотрим правую часть равенства:

$$\mathbb{P}(X \geq n) = \sum_{i=n}^{\infty} q^i p = p \frac{q^n}{1-q} = q^n.$$

Таким образом, утверждение доказано. ■

### 1.3 Задача №3

Рассмотрим игру Орлянка, правила которой описаны в формулировке задания и построим траекторию заданного процесса  $Y(i)$ .



Рис. 1.4: Иллюстрация варианта поведения нормированной суммы  $Y(i)$ .

Ну, вот и финал. Как построить оценку мне не ясно.

$$X_i \sim \text{Bern}\left(\frac{1}{2}\right)$$

$$Y(n) = \frac{2 \sum_{i=1}^n (X_i - \frac{1}{2})}{\sqrt{n}} = \frac{2}{\sqrt{n}} \left( Z - \frac{n}{2} \right) = 2 \frac{Z}{\sqrt{n}} - \sqrt{n} =$$

$$Z \sim \text{Bi}\left(n, \frac{1}{2}\right)$$

## 2 Задание №2

1. Построить датчик сингулярного распределения, имеющий в качестве функции распределения канторову лестницу. С помощью критерия Колмогорова убедиться в корректности работы датчика.
2. Для канторовых случайных величин 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  $\frac{Y}{3}$ ) с помощью критерия Смирнова.
3. Вычислить значение математического ожидания и дисперсии с эмпирическими для разного объема выборок. Проиллюстрировать сходимость.

### 2.1 Задача №1

**Определение 2.1.** Пусть дано вероятностное пространство  $(\mathbb{R}, \mathcal{F}, \mathbb{P})$ , и на нем определена случайная величина  $\xi$  с распределением  $\mathbb{P}_\xi$ . Тогда *функцией распределения* случайной величины  $X$  называется функция  $F_\xi : \mathbb{R} \rightarrow [0, 1]$ , задаваемая формулой:

$$F_\xi(x) = \mathbb{P}(\xi \leq x) \equiv \mathbb{P}_\xi((-\infty, x]).$$

**Определение 2.2.** Функция распределения некоторой случайной величины называется *сингулярной*, если она непрерывна и ее множество точек роста имеет нулевую меру Лебега.

**Определение 2.3.** Из единичного отрезка  $C_0 = [0, 1]$  удалим интервал  $(\frac{1}{3}, \frac{2}{3})$ . Оставшееся множество обозначим за  $C_1$ . Множество  $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$  состоит из двух отрезков; удалим теперь из каждого отрезка его среднюю часть, и оставшееся множество обозначим за  $C_2$ . Повторив данную процедуру, то есть удаляя средние трети у всех четырех отрезков, получим  $C_3$ . Дальше таким же образом получаем последовательность замкнутых множеств  $C_0 \supset C_1 \supset C_2 \supset \dots \supset C_i \supset \dots$ . Пересечение

$$C = \bigcap_{i=0}^{\infty} C_i$$

называется *канторовым множеством*.

*Замечание 2.1.* Канторово множество так же можно определить как множество всех чисел от нуля до единицы, которые можно представить в троичной записи при помощи только нулей и двоек. То есть

$$C = \{0, \alpha_1 \alpha_2 \dots \alpha_i \dots_3 \mid \alpha_i = 0, 2\}.$$

**Утверждение 2.1.** Канторово множество имеет нулевую меру Лебега. [1]

**Определение 2.4.** Рассмотрим функцию  $K(x)$  такую, что в точках 0 и 1 значение функции принимается равным соответственно 0 и 1. Далее интервал  $(0, 1)$  разбивается на три равные части  $(0, \frac{1}{3})$ ,  $(\frac{1}{3}, \frac{2}{3})$  и  $(\frac{2}{3}, 1)$ . На среднем сегменте полагаем  $K(x) = \frac{1}{2}$ . Оставшиеся два сегмента снова разбиваются на три равные части каждый, и на средних сегментах  $K(x)$  полагается равной  $\frac{1}{4}$  и  $\frac{3}{4}$ . Каждый из оставшихся сегментов снова делится

на три части, и на внутренних сегментах  $K(x)$  определяется как постоянная, равная среднему арифметическому между соседними, уже определенными значениями  $K(x)$ . На остальных точках единичного отрезка определяется по непрерывности. Полученная функция называется *канторовой лестницей*.

*Замечание 2.2.* Из определения канторовой лестницы  $K(x)$  следует, что она действует на точки из канторова множества  $C$  по следующему правилу:

$$K(0, \alpha_1 \alpha_2 \dots \alpha_i \dots)_3 = 0, \frac{\alpha_1}{2} \frac{\alpha_2}{2} \dots \frac{\alpha_i}{2} \dots_2.$$

Теперь рассмотрим случайную величину

$$Y = 0, \xi_1 \xi_2 \dots \xi_k \dots_2 = \sum_{k=1}^{\infty} \frac{\xi_k}{2^k}, \quad \text{где } \xi_k \sim \text{Bern}\left(\frac{1}{2}\right).$$

Такая случайная величина имеет равномерное распределение на отрезке  $[0, 1]$ , так как мы равновероятным образом выбираем знаки разложения числа  $\xi_k$  в двоичном представлении. Теперь рассмотрим искомую случайную величину  $X$ , имеющую в качестве функции распределения  $F_X(x)$  канторову лестницу  $K(x)$ . Образ каждой случайной величины  $Y$  для такой функции будет равен

$$K^{-1}(Y) = \sum_{k=1}^{\infty} \frac{2\xi_k}{3^k}.$$

Эта точка лежит в канторовом множестве.

**Теорема 2.1.** Пусть некоторая функция распределения  $F$  имеет обратную  $F^{-1}$ . Тогда функцией распределения случайной величины

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

является  $F$ .

**Д о к а з а т е л ь с т в о.** Найдем функцию распределения случайной величины  $\eta$ :

$$F_{\eta}(x) = \mathbb{P}(\eta \leq x) = \mathbb{P}(F^{-1}(\xi) \leq x) = \mathbb{P}(\xi \leq F(x)) = F(x).$$

Таким образом, теорема доказана. ■

Из теоремы вытекает, что при помощи построенного ранее (см. раздел 1) генератора схемы Бернулли мы можем смоделировать случайную величину  $X$ , принимающую с вероятностью 1 значения из канторова множества  $C$  и имеющую канторову лестницу  $K(x)$  в качестве функции распределения  $F_X(x)$ , следующим образом:

$$X = \sum_{k=1}^{\infty} \frac{2\xi_k}{3^k}, \quad \text{где } \xi_k \sim \text{Bern}\left(\frac{1}{2}\right).$$

В программной реализации будем рассматривать частичные суммы. Для этого этого введем погрешность  $\varepsilon$  и найдем такое число  $n$ , при котором частичная сумма будет отличаться от бесконечной не более, чем на заданную погрешность.

$$\sum_{k=n}^{\infty} \frac{2\xi_k}{3^k} \leq 2 \sum_{k=n}^{\infty} \frac{1}{3^k} = \frac{1}{3^{n-1}} \leq \varepsilon,$$



↓

$$n \geq 1 - \lceil \log_3 \varepsilon \rceil \quad \forall \varepsilon < 1.$$

*Замечание 2.3.* Из выведенной формулы также видно, что для столь малой погрешности как  $\varepsilon = 10^{-9}$  достаточно использовать всего  $n = 20$  первых членов ряда.

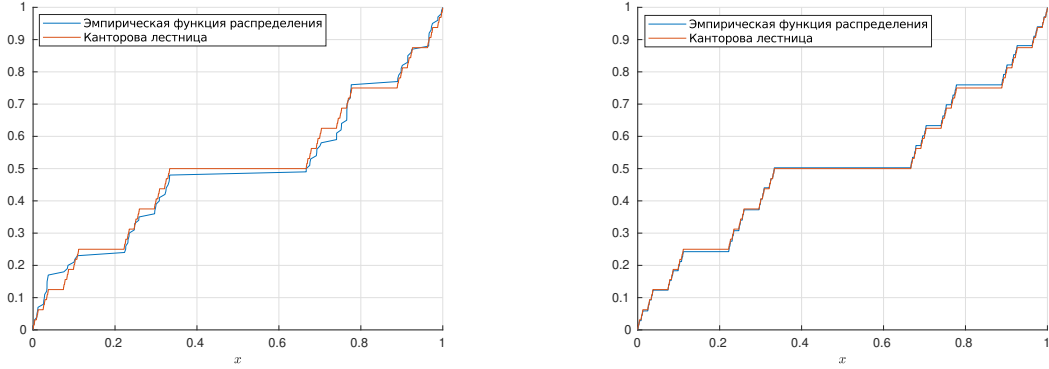


Рис. 2.1: Эмпирическая и теоретическая функции распределения «канторовой» случайной величины  $X$  при выборке из 100 испытаний (слева) и  $10^4$  испытаний (справа).

Дописать проверку того, является ли функция распределения  $X$  канторовой лестницей при помощи критерия Колмогорова.

## 2.2 Задача №2

## 2.3 Задача №3

Вычислим значение математического ожидания для построенной случайной величины  $X$ :

$$\mathbb{E} X = \mathbb{E} \sum_{k=1}^{\infty} \frac{2\xi_k}{3^k} = \sum_{k=1}^{\infty} \frac{2}{3^k} \mathbb{E} \xi_k = \sum_{k=1}^{\infty} \frac{2}{3^k} \cdot \frac{1}{2} = \frac{\frac{1}{3}}{1 - \frac{1}{3}} = \frac{1}{2}.$$

Теперь, помня о независимости случайных величин  $\xi_k$   $k \in \mathbb{N}$ , вычислим значение дисперсии

$$\mathbb{V}ar X = \mathbb{V}ar \sum_{k=1}^{\infty} \frac{2\xi_k}{3^k} = \sum_{k=1}^{\infty} \left( \frac{2}{3^k} \right)^2 \mathbb{V}ar \xi_k = \sum_{k=1}^{\infty} \frac{4}{9^k} \cdot \frac{1}{4} = \frac{\frac{1}{9}}{1 - \frac{1}{9}} = \frac{1}{8}.$$

*Замечание 2.4.* При подсчете мы использовали известные значения для математического ожидания и дисперсии бернуллиевой случайной величины  $\xi \sim \text{Bern}(p)$ :

$$\mathbb{E} \xi = p \quad \text{и} \quad \mathbb{V}ar \xi = p(1 - p).$$

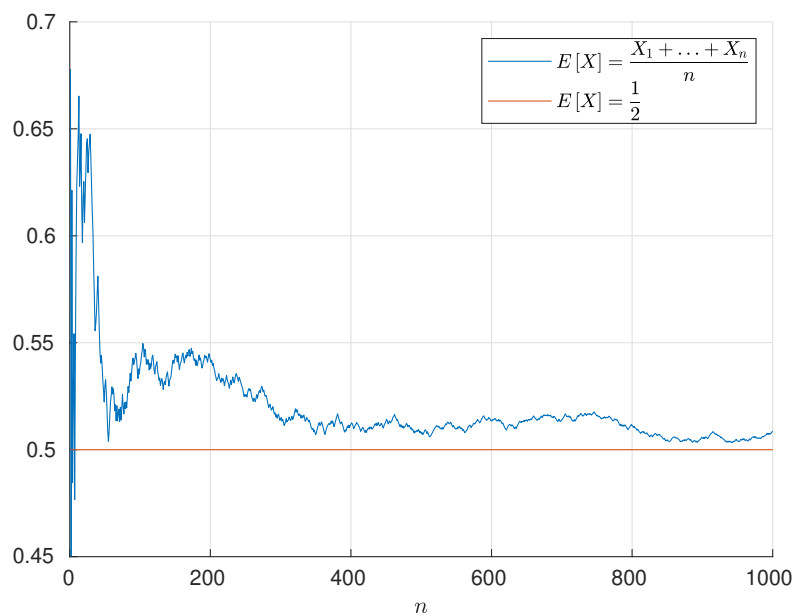


Рис. 2.2: Эмпирическое значение математического ожидания «канторовой» случайной величины  $X$ .

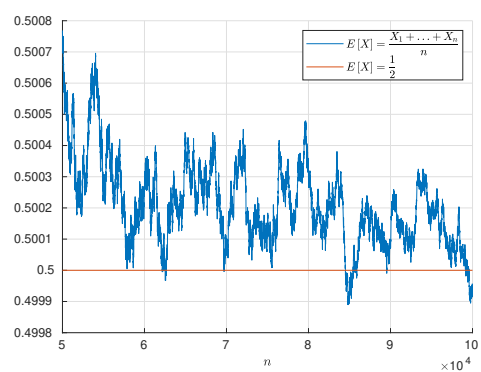
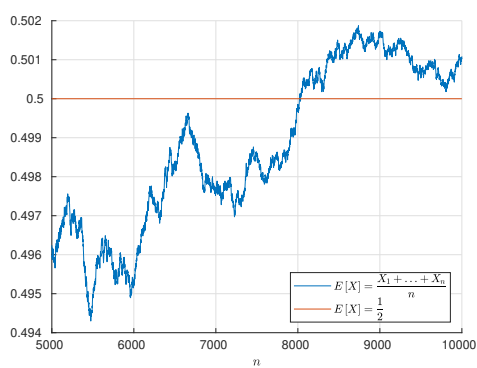


Рис. 2.3: Изменение эмпирическое значения математического ожидания «канторовой» случайной величины  $X$  в зависимости от количества экспериментов:  $5 \cdot 10^3$ – $10^4$  — слева,  $5 \cdot 10^4$ – $10^5$  — справа.

## Список литературы

- [1] Ширяев А. Н. Вероятность, в 2-х кн. — 4-е изд., переработанное и дополненное — М.: МЦНМО, 2007.