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

Считается доступным лишь генератор равномерно распределенной на отрезке $[0,1]$ случайной величины $\eta \sim \mathrm{Uni}(0,1)$. 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 испытаниями 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 $i = 1, \ldots, n$ для одной отдельно взятой траектории. Дать теоритическую оценку для $Y(n)$ при $n \to \infty$.

1.1 Реализация схемы Бернулли и биномиального распределения

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

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

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

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

1.2 Геометрическое распределение

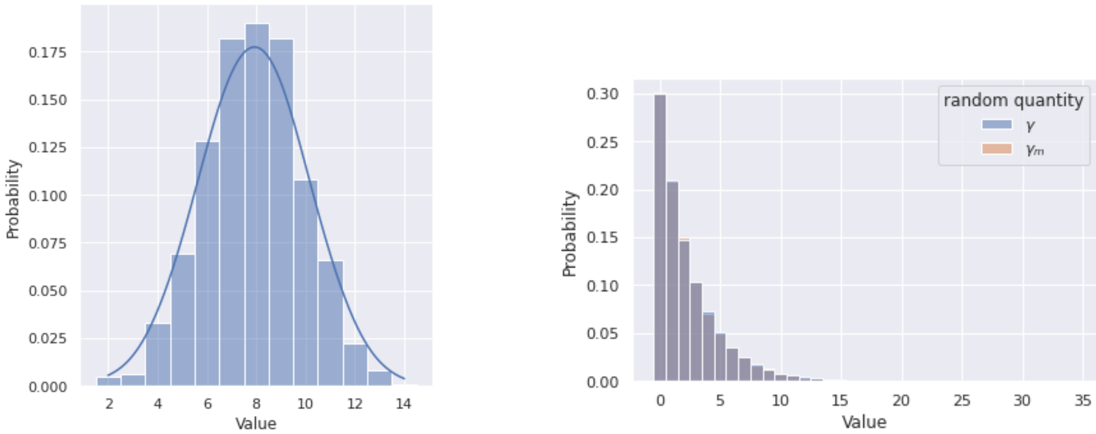
Случайная величина $\gamma \sim \mathrm{Geom}(p)$ представима как

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

Геометрическое распределение обладает свойством отсутствия памяти, т.е.

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

Это свойство можно переформулировать. Пусть $\gamma \sim \mathrm{Geom}(p)$ — случайная величина, определенная на вероятностном пространстве $(\Omega, \mathcal{A}, \mathbb{P})$. Свойство отсутствия памяти сл. в. γ 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 $\gamma_m := (\gamma\big|_{\Omega_m} - m), \quad \Omega_m = \gamma^{-1}(\gamma \geqslant m) \in \mathcal{A}$.

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.

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

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 $n \to \infty$, нам потребуется

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

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

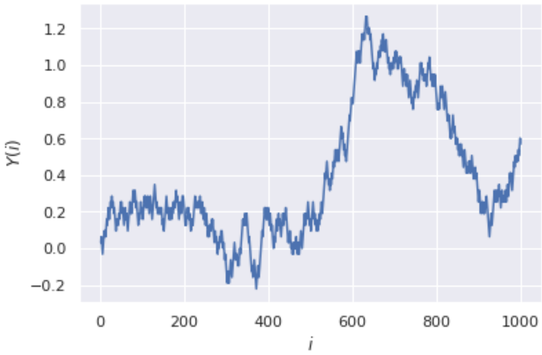


Рис. 2: Игра в орлянку ($n = 1000$)

Действительно

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

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 $Y \in \left[0, \frac{1}{3}\right]$ 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

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

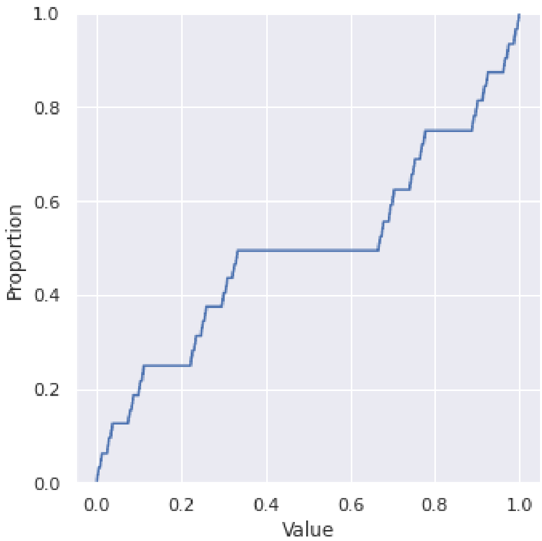


Рис. 3: Э.д.ф. канторовой сл.в.

Причем $\mathbb{P}([a_i, b_i]) = \frac{1}{2^n}$, $\forall i$, что дает естественный способ разложения сл.в. $\delta \sim$ Cant в ряд по $\xi \sim$ 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 $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leqslant x)$ — эмпирическая функция распределения.

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

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

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

$$\sqrt{n} D_n(X_1, \ldots, X_n) \geqslant x_{1-\alpha}, \tag{1}$$

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

$$\mathbb{P}\left(\sqrt{n}D_n \geqslant x_{1-\alpha}\right) \leqslant \alpha.$$

В силу теоремы Колмогорова "в пределе"

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

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

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

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

$$\begin{gathered} x_{1-\alpha(X)} \geqslant x_{1-\alpha} \Leftrightarrow \\ 1-\alpha(X) \geqslant 1-\alpha \Leftrightarrow \\ \alpha \geqslant \alpha(x) \end{gathered}$$

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

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

Посчитанная таким образом на некотрой выборке $(X_1,\ldots,X_n)$ 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

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.