ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

Факультет Технологий Искусственного Интеллекта

Теория вероятностей и продвинутая математическая статистика

Расчетно-графическая работа №1

Выполнили студенты группы J3211:
Нимеева Ангелина Борисовна
ИСУ: 466898
Эль-Сибаи Амир Назихович
ИСУ: 468154
Преподаватель практики:
Кононов Иван Александрович

Санкт-Петербург
2025

Содержание

1 Задача №1

1.1 Постановка задачи

Требуется аналитически решить задачу. Помимо этого нужно написать программу, которая будет считать данные вероятности приблизительно с помощью метода Монте-Карло. Особо любознательные могут задаться поиском (аналитическим) оптимального количества итераций, которых будет достаточно для получения ответа с достаточно «разумной» точностью. При решении задач могут быть полезны принципы произведения вероятностей и включений-исключений.

1.2 Условие варианта 5

Уходя из детского сада, каждый из n случайно берет один левый и один правый ботинок. Найти вероятность того, что каждый их них возьмет не свой левый и не свой правый ботинок.

1.3 Аналитическое решение

1. Общее число исходов (N)
Общее число способов распределить n левых ботинок среди n детей равно числу перестановок n элементов, то есть $n!$. Общее число способов распределить n правых ботинок среди n детей также равно $n!$.
Поскольку выбор левых и правых ботинок независим, общее число исходов равно:

$$\mathbf{N = n! \cdot n! = (n!)^2}$$

2. Число перестановок, при которых никто не берет свой ботинок (D_n)
Мы используем принцип включения-исключения (ПВИ) для нахождения числа перестановок, в которых ни один элемент не остается на своем месте.
Пусть A_i — событие, что ребенок i взял свой левый ботинок. Мы ищем $D_n = |\overline{A_1} \cap \cdots \cap \overline{A_n}| = n! - |A_1 \cup \cdots \cup A_n|$. Сумма k пересечений по ПВИ:

$$\sum_{1 \le i_1 < \cdots < i_k \le n} |A_{i_1} \cap \cdots \cap A_{i_k}| = \binom{n}{k} \cdot (n-k)!$$

Раскладывая биномиальный коэффициент:

$$\binom{n}{k} \cdot (n-k)! = \frac{n!}{k!(n-k)!} \cdot (n-k)! = \frac{n!}{k!}$$

Применяем ПВИ:

$$|A_1 \cup \cdots \cup A_n| = \sum_{k=1}^{n} (-1)^{k-1} \frac{n!}{k!}$$

Число искомых перестановок D_n:

$$D_n = n! - \sum_{k=1}^{n} (-1)^{k-1} \frac{n!}{k!}$$

$$D_n = n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} \right)$$

Объединяя в одну сумму:

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

3. Искомая вероятность (P)
Число благоприятных исходов M равно произведению числа способов для левых и правых ботинок: $M = D_n \cdot D_n = (D_n)^2$.
Вероятность $P = \frac{M}{N}$:

$$P = \frac{(D_n)^2}{(n!)^2} = \frac{\left(n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}\right)^2}{(n!)^2}$$

Сокращая $(n!)^2$, получаем окончательную формулу:

$$\mathbf{P} = \left(\sum_{\mathbf{k=0}}^{\mathbf{n}} \frac{(-\mathbf{1})^{\mathbf{k}}}{\mathbf{k}!}\right)^{\mathbf{2}}$$

1.4 Приближенное решение Монте-Карло

Решение реализовано в блокноте 1.ipynb.

2 Задача №2

2.1 Постановка задачи

Тоже как и в первом упражнении нужно решить аналитически и написать скрипт, вычисляющий вероятность приближенно по методу Монте-Карло. Дополнительный вопрос о поиске числа итерация тоже имеет место быть.

2.2 Условие варианта 2

На окружности случайно ставятся три точки A, B, C. С какой вероятностью треугольник ABC будет остроугольным? Тупоугольным? А прямоугольным?

2.3 Аналитическое решение

Пусть длина окружности равна 1: $L = 1$. Фиксируем точку A. Положения B и C задаются их угловыми координатами относительно A. Пусть x — дуга AB, а y — дуга AC. Наше пространство элементарных исходов — это квадрат $\Omega : 0 \le x \le 1, 0 \le y \le 1. S_\Omega = 1 \cdot 1 = 1$.
Треугольник ABC не зависит от того, как мы назвали точки B и C. То есть, точка (x, y) и (y, x) дают один и тот же треугольник. Чтобы избежать двойного счета, смотрим только на ту часть пространства, где $x \le y$. Получаем треугольную область $\Delta : 0 \le x \le y \le 1$. Площадь нашего рабочего пространства:

$$S_\Delta = \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{2}$$

Вероятность P — это отношение площади благоприятной области S_G к S_Δ:

$$P(G) = \frac{S_G}{1/2} = 2S_G$$

1. Условие остроугольности
$\triangle ABC$ остроугольный, если и только если он содержит центр окружности (описанной). Для этого нужно, чтобы все три дуги между точками были меньше половины окружности. Пусть l_1, l_2, l_3 — длины дуг:

$$\begin{cases} l_1 = x \\ l_2 = y - x \\ l_3 = 1 - y \end{cases}$$

Условие остроугольности (событие G):

$$\begin{cases} l_1 < \frac{1}{2} \\ l_2 < \frac{1}{2} \\ l_3 < \frac{1}{2} \end{cases} \Leftrightarrow \begin{cases} x < \frac{1}{2} \quad \text{(Дуга AB)} \\ y - x < \frac{1}{2} \Rightarrow y < x + \frac{1}{2} \quad \text{(Дуга BC)} \\ 1 - y < \frac{1}{2} \Rightarrow y > \frac{1}{2} \quad \text{(Дуга CA)} \end{cases}$$

С учетом нашего рабочего треугольника Δ $(0 \le x \le y \le 1)$, область G выглядит так:

$$G : \begin{cases} 0 \le x < \frac{1}{2} \\ \frac{1}{2} < y \le 1 \\ y < x + \frac{1}{2} \end{cases}$$

Визуально это трапеция в области Δ.
3. Нахождение площади S_G
Считаем интеграл, чтобы найти площадь G:

$$S_G = \int_{x=0}^{x=1/2} (y_{\text{верх}} - y_{\text{ниж}})\, dx = \int_0^{1/2} \left((x + \frac{1}{2}) - \frac{1}{2} \right) dx$$

Получаем:

$$S_G = \int_0^{1/2} x\, dx = \frac{x^2}{2}\Big|_0^{1/2} = \frac{(1/2)^2}{2} - 0 = \frac{1/4}{2} = \frac{1}{8}$$

4. Считаем вероятности
Остроугольный треугольник

$$P(\text{остроугольный}) = \frac{S_G}{S_\Delta} = \frac{1/8}{1/2} = \frac{1}{4}$$

Прямоугольный треугольник
$\triangle ABC$ прямоугольный, если одна из его сторон является диаметром окружности. Это значит, что одна из дуг должна быть ровно 1/2.

$$l_1 = \frac{1}{2} \quad \text{или} \quad l_2 = \frac{1}{2} \quad \text{или} \quad l_3 = \frac{1}{2}$$

Это линии $x = 1/2$, $y = x + 1/2$, $y = 1/2$ в двумерном пространстве. Площадь (мера) линии в двумерном пространстве равна нулю.

$$P(\text{прямоугольный}) = 0$$

Тупоугольный треугольник
Треугольник тупоугольный (событие T), если он не остроугольный и не прямоугольный. В нашем случае, если $\triangle ABC$ не содержит центр окружности, т.е. хотя бы одна дуга больше 1/2.Событие T — это дополнение к $G \cup R$, где R — прямоугольный.

$$P(\text{тупоугольный}) = 1 - P(\text{остроугольный}) - P(\text{прямоугольный})$$

$$P(\text{тупоугольный}) = 1 - \frac{1}{4} - 0 = \frac{3}{4}$$

2.4 Приближенное решение Монте-Карло

Решение реализовано в блокноте 2.ipynb.

3 Задача №3

3.1 Постановка задачи

Аналитически решить задачу

3.2 Условие варианта 3

Введем события $A_i = \{X = i\}$, $B_i = \{Y = i\}$, $i \geq 0$. Известно, что для любых $i \geq 0$ и $j \geq 0$ события A_i и B_j — независимы, при этом

$$\mathrm{P}(X = i) = \mathrm{P}(Y = i) = (1-p)^i p \quad i \geq 0,$$

где $p \in (0, 1)$. Найти $\mathrm{P}(X = i | X + Y = j)$.

3.3 Аналитическое решение

1. Вычисление вероятности $P(X = i)$ и $P(Y = i)$
По условию:

$$P(X = i) = (1-p)^i p, \quad i \geq 0$$

$$P(Y = i) = (1-p)^i p, \quad i \geq 0$$

2. Вычисление вероятности $P(X = i, X + Y = j)$
Мы ищем $P(X = i | X + Y = j)$. По определению условной вероятности:

$$P(X = i | X + Y = j) = \frac{P(X = i, X + Y = j)}{P(X + Y = j)}$$

Событие $\{X = i, X + Y = j\}$ эквивалентно событию $\{X = i, Y = j - i\}$. Для этого требуется, чтобы $i \geq 0$ и $j - i \geq 0$, то есть $0 \leq i \leq j$. Если $i < 0$ или $i > j$, то $P(X = i, X + Y = j) = 0$.
Так как X и Y независимы (события $A_i = \{X = i\}$ и $B_k = \{Y = k\}$ независимы для любых $i, k \geq 0$):

$$P(X = i, Y = j - i) = P(X = i)P(Y = j - i)$$

Подставляем формулы вероятностей:

$$P(X = i, Y = j - i) = \left[(1-p)^i p\right] \cdot \left[(1-p)^{j-i} p\right]$$

$$P(X = i, Y = j - i) = (1-p)^{i+(j-i)} p^2$$

$$P(X = i, Y = j - i) = (1-p)^j p^2 \quad \text{для } 0 \leq i \leq j$$

3. Вычисление вероятности $P(X + Y = j)$
Случайная величина $Z = X + Y$ принимает значение j тогда и только тогда, когда $X = i$ и $Y = j - i$ для некоторого i такого, что $0 \leq i \leq j$. События $\{X = i, Y = j - i\}$ для разных i несовместны.

$$P(X + Y = j) = \sum_{i=0}^{j} P(X = i, Y = j - i)$$

Используя результат из пункта 2:

$$P(X + Y = j) = \sum_{i=0}^{j} (1-p)^j p^2$$

Так как $(1-p)^j p^2$ не зависит от i:

$$P(X+Y=j)=(j+1)(1-p)^j p^2$$

4. Вычисление условной вероятности $P(X=i|X+Y=j)$
Подставляем результаты из пунктов 2 и 3 в формулу условной вероятности. Опять же, мы предполагаем $0 \le i \le j$, иначе вероятность равна 0.

$$P(X=i|X+Y=j)=\frac{P(X=i,X+Y=j)}{P(X+Y=j)}=\frac{(1-p)^j p^2}{(j+1)(1-p)^j p^2}$$

После сокращения:

$$P(X=i|X+Y=j)=\frac{1}{j+1}$$

Таким образом, условная вероятность $P(X=i|X+Y=j)$ равна:

$$P(X=i|X+Y=j)=\begin{cases}\frac{1}{j+1}, & \text{если } 0 \le i \le j\\ 0, & \text{если } i<0 \text{ или } i>j\end{cases}$$

4 Задача №4

4.1 Постановка задачи

Представьте, что вам нужно быстро создать некий программный продукт, в котором будут использоваться вероятностные модели. Возможно, вы будете использовать Python и модуль SciPy, в котором реализовано достаточно много математических алгоритмов и моделей, в том числе и вероятностно-статистических (подмодуль stats). Допустим, в вашей модели есть "нестандартное" распределение (например, не реализованное в SciPy), для которого известна плотность, и вам нужно быстро реализовать распределение, то есть определить функцию распределения, квантили и прочие методы. Тогда идея унаследоваться от класса rv_continious кажется разумной: ведь достаточно переопределить метод, соответствующий плотности. Казалось бы отличная идея, плюс применили объектно-ориентированный подход. Но допустим, что вам понадобилось сгенерировать случайные числа (на самом деле псевдослучайные) из данного распределения, но вдруг что-то пойдёт не так, например, будет слишком долго работать... И вдруг нужно будет сообразить другие методы.
Само задание.

- Для заданной плотности p наследуйтесь от класса rv_continuous и сгенерируйте с помощью реализованного подкласса для вашего распределения n случайных чисел из данного распределения. Посмотрите, как будет работать алгоритм при росте n (сразу большим n не делайте). Убедитесь эмпирически, что распределение сгенерированных чисел согласуется с данным вероятностным законом.

- Реализуйте генерацию для вашего распределения с помощью обратной к функции распределения (при написании функции для F^{-1} вызов специальных функций, например, квантили стандартного нормального закона разрешается; не забыть аналитический вывод F^{-1}). Проведите тот же эксперимент.

- Выберите еще один метод для генерации случайных чисел (можно, например, rejecting sampling, ratio of uniforms или другой метод). Опишите его математическое обоснование. Воспользуйтесь реализацией или сами реализуйте. Проведите тот же эксперимент.

4.2 Условие варианта 2

$$p(x) = \frac{3x^2}{\sqrt{2\pi}} e^{-(5-x^3)^2/2}$$

4.3 Метод rejecting sampling

Метод выборки с отклонением состоит в том, чтобы генерировать случайные выборки из сложного целевого распределения $f(x)$, из которого трудно семплировать напрямую, используя при этом более простое вспомогательное распределение $g(x)$, из которого семплировать легко, и которое удовлетворяет следующему условию:

$$\forall x \; f(x) \leq cg(x), \quad c > 1$$

Алгоритм

1. Взять семпл x по распределению $g(x)$;
2. Выбрать случайное число u равномерно из отрезка $[0, cg(x)]$;
3. Вычислить $f(x)$;
4. Если $u \leq f(x)$, то x добавляется к семплам;
5. Если $u > f(x)$, то x отклоняется (отсюда и название метода).

Обоснование
Пусть Y — случайная величина с плотностью распределения $g(y)$, а U — случайная величина, равномерно распределённая на отрезке $[0, cg(x)]$. Мы хотим показать, что условная плотность Y при условии события $U \leq f(Y)$ равна искомой плотности $f(x)$.

$$P(Y = y \mid U \leq f(Y)) = \frac{P(Y = y \cap U \leq f(Y))}{P(U \leq f(Y))}$$

Рассмотрим знаменатель. Заметим, что U и Y независимы, так как U выбирается независимо от Y.

$$P(U \leq f(Y)) = \int_{-\infty}^{\infty} P(U \leq f(y))g(y)dy =$$

$$= \int_{-\infty}^{\infty} \frac{f(y)}{cg(y)} g(y)dy = \frac{1}{c} \int_{-\infty}^{\infty} f(y)dy = \frac{1}{c}$$

Рассмотрим числитель.

$$P(Y = y \cap U \leq f(Y)) = P(U \leq f(Y) \mid Y = y) \cdot P(Y = y) = P(U \leq f(y) \cdot g(y)$$

$$= \frac{f(y)}{cg(y)} \cdot g(y) = \frac{f(y)}{c}$$

Теперь объединим и получим искомый результат:

$$P(Y = y \mid U \leq f(Y)) = \frac{f(y)}{c} \cdot c = f(y)$$

4.4 Эмпирическое решение

Решение реализовано в блокноте 4.ipynb.

5 Задача №5

5.1 Постановка задачи

Для заданного распределения требуется найти:

- «нормирующие» константы, если требуется
- функцию распределения/плотность (зависит от исходных данных)
- математическое ожидание, дисперсию, среднее квадратическое отклонение, моду, медиану, асимметрию, эксцесс (если они определен)
- функцию распределения и плотность функции от исходной случайной величины

5.2 Условие варианта 6

$$F(x) = \frac{c}{x^8}\mathbf{1}(|x| \geq 2), \quad Y = X^4$$

5.3 Решение

1. Нормирующая константа Нормировка требует

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = 2c\int_2^{\infty} x^{-8}\,dx.$$

Так как

$$\int_2^{\infty} x^{-8}\,dx = \left[\frac{x^{-7}}{-7}\right]_2^{\infty} = \frac{1}{7\cdot 2^7} = \frac{1}{896},$$

получаем

$$2c\cdot\frac{1}{896} = 1 \quad \Rightarrow \quad c = 448.$$

2. Функция распределения $F_X(x)$ Для разных областей x:

$$\begin{aligned} x < -2: F_X(x) &= \int_{-\infty}^{x} \frac{448}{t^8}\,\mathbf{1}(|t| \geq 2)\,dt = \frac{448}{7|x|^7} = \frac{64}{|x|^7}, \\ -2 \leq x < 2: F_X(x) &= P(X \leq -2) = \tfrac{1}{2}, \\ x \geq 2: F_X(x) &= \tfrac{1}{2} + 448\int_2^{x} t^{-8}\,dt = 1 - \frac{448}{7x^7} = 1 - \frac{64}{x^7}. \end{aligned}$$

(Проверка в точках: $F_X(-2) = F_X(2^-) = 1/2$, $F_X(2) = 1 - 64/2^7 = 1 - 64/128 = 1/2$; при $x \to \infty$ — 1.)

3. Математическое ожидание и моменты Из симметрии плотности относительно нуля:

$$E[X] = 0.$$

Обобщённая формула для конечных моментов при целых $m < 7$:

$$E[X^m] = 2\cdot 448\int_2^{\infty} x^{m-8}\,dx = \frac{448\cdot 2^{\,m-6}}{7-m} \quad (m < 7).$$

(Здесь знаменатель положителен, потому что $7 - m > 0$.)

В частности:

$$E[X^2] = \frac{28}{5} = 5.6, \qquad E[X^4] = \frac{112}{3} \approx 37.333 \qquad E[X^6] = 448.$$

Моменты порядка $m \geq 7$ (включая $m = 8$) расходятся (бесконечны).

4. Дисперсия и стандартное отклонение

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = E[X^2] = \frac{28}{5}, \qquad \sigma_X = \sqrt{\frac{28}{5}}.$$

5. Мода и медиана Плотность $f_X(x) = 448/x^8$ убывает по $|x|$ на области $|x| \geq 2$, следовательно максимумы достигаются при минимальном $|x|$, т.е.

$$\text{моды: } x = \pm 2.$$

Медиана — любой m с $F_X(m) \geq 1/2$ и $F_X(m^-) \leq 1/2$. Так как $F_X(x) = 1/2$ для всех $x \in [-2, 2)$, то технически любая точка из отрезка $[-2, 2]$ является медианой. По симметрии обычно берут $m = 0$.

6. Асимметрия (скью) и эксцесс (кортозис) Из симметрии все нечётные центральные моменты нулевые, значит skewness $= 0$. Классический (избыточный) эксцесс:

$$\text{кортозис} = \frac{E[X^4]}{\sigma_X^4}, \qquad \text{excess} = \text{кортозис} - 3.$$

Подставляя найденные значения:

$$\sigma_X^4 = \left(\tfrac{28}{5}\right)^2 = \frac{784}{25}, \qquad E[X^4] = \frac{112}{3}.$$

Поэтому

$$\text{кортозис} = \frac{112/3}{784/25} = \frac{25}{21} \approx 1.19048, \qquad \text{excess} = \frac{25}{21} - 3 = -\frac{38}{21} \approx -1.8095.$$

(Эксцесс определён, т.к. $E[X^4] < \infty$.)

7. Распределение $Y = X^4$ Область значений: поскольку $|X| \geq 2$, имеем $Y \geq 2^4 = 16$. Отображение $x \mapsto y = x^4$ даёт два прообраза $x = \pm y^{1/4}$ для $y \geq 16$. По формуле для преобразования плотности:

$$f_Y(y) = \sum_{x:\, x^4 = y} \frac{f_X(x)}{|g'(x)|} = f_X(y^{1/4})\frac{1}{4}y^{-3/4} + f_X(-y^{1/4})\frac{1}{4}y^{-3/4}.$$

Так как $f_X(\pm y^{1/4}) = 448/(y^{1/4})^8 = 448\, y^{-2}$, получаем

$$f_Y(y) = 2 \cdot 448\, y^{-2} \cdot \frac{1}{4} y^{-3/4} = \frac{448}{2}\, y^{-11/4} = 224\, y^{-11/4}, \qquad y \geq 16.$$

Проверка нормировки:

$$\int_{16}^{\infty} 224\, y^{-11/4} dy = 224 \cdot \frac{4}{7}\, 16^{-7/4} = 128 \cdot 16^{-7/4} = 128 \cdot (2^{-7}) = 1.$$

Функция распределения $F_Y(y)$:

$$F_Y(y) = \begin{cases} 0, & y < 16, \\ \displaystyle\int_{16}^{y} 224\, t^{-11/4} dt = 1 - 128\, y^{-7/4}, & y \geq 16. \end{cases}$$

8. Моменты и характеристики Y

- $E[Y] = E[X^4] = \dfrac{112}{3}$ (конечно).
- $E[Y^2] = E[X^8]$ — расходится (бесконечно), поэтому $\mathrm{Var}(Y) = \infty$ и все центральные моменты порядка ≥ 2 не определены (бесконечны). Отсюда для Y дисперсия, стандартное отклонение, коэффициенты асимметрии и эксцесса не определены.

5.4 Краткие итоговые формулы (для записи)

$$\boxed{c = 448}, \qquad \boxed{f_X(x) = \frac{448}{x^8}\mathbf{1}(|x| \geq 2)}$$

$$\boxed{F_X(x) = \begin{cases} \frac{64}{|x|^7}, & x < -2, \\ \frac{1}{2}, & -2 \leq x < 2, \\ 1 - \dfrac{64}{x^7}, & x \geq 2, \end{cases}}$$

$$E[X] = 0, \quad \mathrm{Var}(X) = \frac{28}{5}, \quad \sigma_X = \sqrt{\frac{28}{5}}, \quad \text{modes } x = \pm 2, \quad \text{median any } m \in [-2, 2].$$

$$f_Y(y) = 224\, y^{-11/4}\mathbf{1}(y \geq 16), \qquad F_Y(y) = \begin{cases} 0, & y < 16, \\ 1 - 128\, y^{-7/4}, & y \geq 16. \end{cases}$$

5.5 Эмпирическое решение

Решение реализовано в блокноте 5.ipynb.

6 Задача №6

6.1 Постановка задачи

Для решения этой задачи будет нелишним следующее утверждение.
Пусть $g : U \to V$ – диффеоморфизм (гладкая биекция), $U, V \subset R^n$, $X = (X_1, \dots, X_n)^T$ – случайный вектор с плотностью $p_X(x_1, \dots, x_n) = p_X(\mathbf{x})$. Рассмотрим случайный вектор $Y = (Y_1, \dots, Y_n)^T = g(X) = g(X_1, \dots, X_n)$. Тогда для вычисления плотности случайного вектора Y можно воспользоваться формулой (сразу записывается в векторном виде):

$$p_Y(\mathbf{y}) = p_X(g^{-1}(\mathbf{y})) \left|\det \frac{Dg^{-1}(\mathbf{y})}{D\mathbf{y}}\right| = \frac{p_X(g^{-1}(\mathbf{y}))}{\left|\det \frac{Dg(\mathbf{x})}{D\mathbf{x}}\right|_{\mathbf{x}=g^{-1}(\mathbf{y})}},$$

где $D(\cdot)$ – матрица Якоби и в последнем равенстве мы воспользовались теоремой об обратном отображении.
Кроме того проведите эксперимент и эмпирически убедитесь, что полученные величины действительно являются гауссовскими.

6.2 Условие варианта 2

Пусть случайная величина R имеет **распределение Рэлея**, то есть распределение с плотностью

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathbf{1}(x \geq 0),$$

где $\sigma > 0$ – масштабирующий параметр, и пусть $\Theta \sim U[0, 2\pi]$.
Показать, что величины $X = R\cos\Theta$ и $Y = R\sin\Theta$ независимы и распределены нормально с нулевым математическим ожиданием и дисперсией σ^2, если R и Θ независимы.

6.3 Аналитическое решение

Так как R и Θ независимы, имеем:

$$p_{R,\Theta}(r, \theta) = f_R(r) \cdot \frac{1}{2\pi} = \frac{1}{2\pi} \cdot \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \geq 0,\ \theta \in [0, 2\pi).$$

Переход к (X, Y)
Пусть $x = r\cos\theta$, $\quad y = r\sin\theta$. Тогда обратное преобразование задается формулами:

$$r = \sqrt{x^2 + y^2}, \quad \theta = \begin{cases} \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right), & y \geq 0, \\ 2\pi - \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right), & y < 0. \end{cases}$$

Якобиан перехода $(r, \theta) \mapsto (x, y)$ равен $|\det D| = r$, поэтому для обратного преобразования $(x, y) \mapsto (r, \theta)$ имеем

$$|\det D_{g^{-1}}(x, y)| = \frac{1}{r}.$$

По формуле:

$$p_{X,Y}(x, y) = p_{R,\Theta}(r(x, y), \theta(x, y)) \cdot |\det D_{g^{-1}}(x, y)|.$$

Подставим $r = \sqrt{x^2 + y^2}$:

$$p_{X,Y}(x, y) = \frac{1}{2\pi} \cdot \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \cdot \frac{1}{r}$$

$$= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),$$

для всех $(x, y) \in R^2$. Маргинальная плотность $p_X(x)$:

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x, y)\, dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) dy$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/(2\sigma^2)}\, dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} \cdot \sqrt{2\pi\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/(2\sigma^2)}\, dy$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)} \cdot \sqrt{2\pi\sigma^2} \cdot 1 = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}.$$

Аналогично,

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/(2\sigma^2)}.$$

Тогда

$$p_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/(2\sigma^2)} = p_X(x) p_Y(y).$$

Следовательно, X и Y независимы. Из вида маргинальных плотностей следует:

$$X \text{ имеет плотность } p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}.$$

$$Y \text{ имеет плотность } p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/(2\sigma^2)}.$$

Эти плотности соответствуют нормальному распределению $N(0, \sigma^2)$. Следовательно,

$$E[X] = E[Y] = 0, \quad \mathrm{Var}(X) = \mathrm{Var}(Y) = \sigma^2.$$

6.4 Эмпирическое решение

Решение реализовано в блокноте 6.ipynb.

7 Задача №7

7.1 Постановка задачи

Дано совместное распределение дискретных случайных величин (X, Y). Требуется найти:

- Пропущенное значение вероятности
- Маргинальные распределения для X и Y
- Распределение случайной величины $U = g(X, Y)$ для заданной функции g
- Математическое ожидание, дисперсию, среднее квадратичное отклонение, моду, медиану, коэффициенты асимметрии и эксцесса случайных величин X и Y
- Ковариацию, коэффициент корреляции между X, Y
- Исследовать на независимость

В решении задания следует указать общие формулы/соображения, из которых считаются искомые показатели, однако для их вычисления можно написать скрипт (**функции соответствующие реализовать самостоятельно**).

7.2 Условие варианта 2

Таблица совместного распределения вероятностей $p_{ij} = P(Y = y_j, X = x_i)$

Y \X	**1**	**2**	**9**	**15**
-8	0.04	0.10	0.01	0.11
-5	NULL	0.01	0.02	0.07
0	0.06	0.08	0.11	0.03
2	0.04	0.05	0.08	0.09
6	0.01	0.02	0.04	0.01

7.3 Решение

1. Пропущенное значение вероятности

$$\sum p_{ij}^{\text{известные}} = 0.26 + 0.10 + 0.28 + 0.26 + 0.08 = 0.98$$

Следовательно, пропущенное значение равно:

$$p(Y = -5, X = 1) = 1 - \sum p_{ij}^{\text{известные}} = 1 - 0.98 = 0.02$$

2. Маргинальные распределения для X и Y

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$$

$$P(X = 1) = 0.04 + 0.02 + 0.06 + 0.04 + 0.01 = 0.17$$

$$P(X = 2) = 0.10 + 0.01 + 0.08 + 0.05 + 0.02 = 0.26$$

$$P(X = 9) = 0.01 + 0.02 + 0.11 + 0.08 + 0.04 = 0.26$$

$$P(X = 15) = 0.11 + 0.07 + 0.03 + 0.09 + 0.01 = 0.31$$

Проверка: $0.17 + 0.26 + 0.26 + 0.31 = 1.00$.
Маргинальное распределение Y

$$P(Y = y_j) = \sum_i P(X = x_i, Y = y_j)$$

$$P(Y = -8) = 0.04 + 0.10 + 0.01 + 0.11 = 0.26$$

$$P(Y = -5) = 0.02 + 0.01 + 0.02 + 0.07 = 0.12$$

$$P(Y = 0) = 0.06 + 0.08 + 0.11 + 0.03 = 0.28$$

$$P(Y = 2) = 0.04 + 0.05 + 0.08 + 0.09 = 0.26$$

$$P(Y = 6) = 0.01 + 0.02 + 0.04 + 0.01 = 0.08$$

Проверка: $0.26 + 0.12 + 0.28 + 0.26 + 0.08 = 1.00$.

3. Распределение $\mathbf{U = (XY)\%5}$

Остаток должен быть целым числом в диапазоне $\{0, 1, 2, 3, 4\}$.
Строка $Y = -8$:

$$X = 1 : XY = 1 \times (-8) = -8$$

$$X = 2 : XY = 2 \times (-8) = -16$$

$$X = 9 : XY = 9 \times (-8) = -72$$

$$X = 15 : XY = 15 \times (-8) = -120$$

и так далее...

Y \X	**1**	**2**	**9**	**15**
-8	2	4	3	0
-5	0	0	0	0
0	0	0	0	0
2	2	4	3	0
6	1	2	4	0

Суммируем вероятности p_{ij} для каждого значения U:

$$P(U = 0) = 0.11 + 0.12 + 0.28 + 0.09 + 0.01 = 0.61$$

$$P(U = 1) = 0.01 = 0.01$$

$$P(U = 2) = 0.04 + 0.04 + 0.02 = 0.10$$

$$P(U = 3) = 0.01 + 0.08 = 0.09$$

$$P(U = 4) = 0.10 + 0.05 + 0.04 = 0.19$$

U (u_k)	**0**	**1**	**2**	**3**	**4**
P(U $= u_k$	0.61	0.01	0.10	0.09	0.19

4. Числовые характеристики X и Y

Для X:

$$E(X) = 1(0.17) + 2(0.26) + 9(0.26) + 15(0.31) = 7.68$$

$$E(X^2) = 1^2(0.17) + 2^2(0.26) + 9^2(0.26) + 15^2(0.31) = 92.02$$

$$D(X) = E(X^2) - (E(X))^2 = 92.02 - (7.68)^2 = 33.04$$

$$\sigma(X) = \sqrt{33.04} = 5.75$$

$$Mo(X) = 15(\text{макс. } P(X) = 0.31)$$

$$Me(X) = 9(P(X \leq 9) = 0.69)$$

Коэффициент ассиметрии

$$As_X = \frac{E[(X - E(X))^3]}{\sigma_X^3}$$

Вычислим $E[(X - E(X))^3]$:

$$E[(X - E(X))^3] = \sum_i (x_i - E(X)^3)P(X = x_i) =$$

$$E[(X-E(X))^3] = (1-7.68)^3(0.17)+(2-7.68)^3(0.26)+(9-7.68)^3(0.26)+(15-7.68)^3(0.31)$$

$$= -50.673 - 47.645 + 0.597 + 121.589 = 23.868$$

$$As_X = \frac{E[(X - E(X))^3]}{\sigma_X^3} = \frac{23.868}{(5.748)^3} = \frac{23.868}{189.911} = 0.126$$

Коэффициент эксцессии

$$Ex_X = \frac{E[(X - E(X))^4]}{\sigma_X^4} - 3$$

Вычислим $E[(X - E(X))^4]$:

$$E[(X - E(X))^4] = (-6.68)^4(0.17) + (-5.68)^4(0.26) + (1.32)^4(0.26) + (7.32)^4(0.31)$$

$$= 338.497 + 270.624 + 0.789 + 890.033 = 1499.913$$

$$Ex_X = \frac{E[(X - E(X))^4]}{\sigma_X^4} - 3 = \frac{1499.913}{1091.48} - 3 = -1.628$$

Для Y:

$$E(Y) = (-8)(0.26) + (-5)(0.12) + 0(0.28) + 2(0.26) + 6(0.08) = -1.68$$

$$E(Y^2) = (-8)^2(0.26) + (-5)^2(0.12) + 0^2(0.28) + 2^2(0.26) + 6^2(0.08) = 23.56$$

$$D(Y) = E(Y^2) - (E(Y))^2 = 23.56 - (-1.68)^2 = 20.73$$

$$\sigma(Y) = \sqrt{20.73} = 4.55$$

$$Mo(Y) = 0(\text{макс. } P(Y) = 0.28)$$

$$Me(Y) = 0(P(Y \leq 0) = 0.66)$$

Коэффициент ассиметрии

$$As_X = \frac{E[(Y - E(Y))^3]}{\sigma_Y^3}$$

Вычислим $E[(Y-E(Y))^3]$:

$$E[(Y-E(Y))^3] = (-8+1.68)^3(0.26) + (-5+1.68)^3(0.12) + (0+1.68)^3(0.28)$$

$$+(2+1.68)^3(0.26)+(6+1.68)^3(0.08) = -65.551-4.396+1.328+13.170+36.142 = -19.307$$

$$As_Y = \frac{E[(Y-E(Y))^3]}{\sigma_Y^3} \approx \frac{-19.307}{(4.554)^3} \approx \frac{-19.307}{94.48} \approx -0.204$$

Коэффициент эксцессии

$$Ex_Y = \frac{E[(Y-E(Y))^4]}{\sigma_Y^4} - 3 = \frac{756.0}{(20.7376)^2} - 3 = -1.242$$

5. Ковариация и корреляция

$$E(XY) = \sum_i \sum_j x_i y_j p_{ij} = -14.41$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = -14.41 - (-12.80) = -1.61$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} = \frac{-1.61}{5.74 \times 4.55} = -0.06$$

6. Исследование на независимость

Случайные величины X и Y независимы тогда и только тогда, когда для всех значений x_i и y_j выполняется:

$$p_{ij} = P(Y = y_j, X = x_i) = P(Y = y_j)P(X = x_i).$$

1) Ячейка $(Y = -8, X = 1)$:

$$P(Y = -8, X = 1) = 0.04$$

Произведение маргинальных вероятностей:

$$P(Y = -8)P(X = 1) = 0.26 \times 0.17 = 0.0442$$

Так как $0.04 \neq 0.0442$, равенство $p_{ij} = P(Y = y_j)P(X = x_i)$ нарушено уже для этой ячейки.

2) Ячейка $(Y = -5, X = 15)$:

$$P(Y = -5, X = 15) = p_{24} = 0.07$$

Произведение маргинальных вероятностей:

$$P(Y = -5)P(X = 15) = 0.12 \times 0.31 = 0.0372$$

Снова $0.07 \neq 0.0372$. Поскольку существует по крайней мере одна пара (y_j, x_i), для которой

$$P(Y = y_j, X = x_i) \neq P(Y = y_j)P(X = x_i),$$

случайные величины X и Y не являются независимыми.

7.4 Эмпирическое решение

Решение реализовано в блокноте 7.ipynb.