**Лекция 11**

**Случайные переменные и их распределения**

*(Как математика описывает хаос: от кубиков до Big Data)*

**1. Что такое случайная переменная?**

**Случайная переменная (величина)** — это числовая функция, которая ставит в соответствие каждому исходу эксперимента определённое значение.

**Примеры:**

* Число орлов при 10 подбрасываниях монеты.
* Время ожидания автобуса на остановке.
* Рост случайно выбранного человека.

**Типы:**

1. **Дискретные** — принимают отдельные значения (например, целые числа).
2. **Непрерывные** — принимают любые значения в интервале (например, вещественные числа).

**2. Дискретные распределения**

**2.1. Биномиальное распределение**

**Когда использовать?**  
Моделирует число успехов в серии из **n** независимых испытаний с вероятностью успеха **p** в каждом.

**Формула:**

P(k)=C\_n^k⋅p^k⋅(1−p)^(n−k)

**Пример:**  
Вероятность выпадения 3 орлов при 5 подбрасываниях монеты (p=0.5):

P(3)=C\_5^3⋅0.53⋅0.52=10⋅0.125⋅0.25=0.3125.

**Код на Python:**

from scipy.stats import binom

n = 5

p = 0.5

k = 3

prob = binom.pmf(k, n, p)

print(f"Вероятность: {prob:.4f}") # 0.3125

**2.2. Распределение Пуассона**

**Когда использовать?**  
Моделирует число редких событий за фиксированный интервал времени (например, звонки в кол-центр за час).

**Формула:**

P(k)=(λ^k⋅e^−λ) / k!

где λ — среднее число событий.

**Пример:**  
Если в магазин за час приходит в среднем 4 клиента (λ=4), какова вероятность прихода 2 клиентов?

P(2) = 4^2⋅e^−4 / 2!=16⋅0.01832≈0.1465

**Код на Python:**

from scipy.stats import poisson

lambda\_ = 4

k = 2

prob = poisson.pmf(k, lambda\_)

print(f"Вероятность: {prob:.4f}") # 0.1465

**2.3. Геометрическое распределение**

**Когда использовать?**  
Моделирует число испытаний до первого успеха.

**Формула:**

P(k)=(1−p)k−1⋅p

**Пример:**  
Вероятность, что первый орёл выпадет на 3-м подбрасывании (p=0.5):

P(3)=0.52⋅0.5=0.125.

**Код на Python:**

from scipy.stats import geom

p = 0.5

k = 3

prob = geom.pmf(k, p)

print(f"Вероятность: {prob:.4f}") # 0.125

**3. Непрерывные распределения**

**3.1. Нормальное распределение (Гаусса)**

**Когда использовать?**  
Моделирует естественные процессы (рост, ошибки измерений, оценки IQ).

**Пример:**  
Рост взрослых мужчин (μ = 175 см, σ = 10 см). Вероятность, что рост человека между 165 и 185 см:

from scipy.stats import norm

mu = 175

sigma = 10

prob = norm.cdf(185, mu, sigma) - norm.cdf(165, mu, sigma)

print(f"Вероятность: {prob:.4f}") # ~0.6827 (правило 68-95-99.7)

**3.2. Экспоненциальное распределение**

**Когда использовать?**  
Моделирует время между событиями (например, время до поломки прибора).

**Формула плотности:**

f(x)=λ⋅e^(−λx)

**Пример:**  
Если среднее время между заказами в интернет-магазине 10 минут (λ=0.1*λ*=0.1), какова вероятность, что следующий заказ придёт в течение 5 минут?

from scipy.stats import expon

lambda\_ = 0.1

prob = expon.cdf(5, scale=1/lambda\_)

print(f"Вероятность: {prob:.4f}") # ~0.3935

**3.3. Равномерное распределение**

**Когда использовать?**  
Все исходы в интервале [a,b][*a*,*b*] равновероятны.

**Формула плотности:**

f(x)=1b−a для a≤x≤b

**Пример:**  
Генерация случайного числа от 2 до 5:

import numpy as np

a, b = 2, 5

sample = np.random.uniform(a, b)

print(f"Случайное число: {sample:.2f}") # Например, 3.78

**4. Как выбрать распределение?**

1. **Анализируйте природу данных:**
   * Дискретные события (биномиальное, Пуассона).
   * Время/расстояния (экспоненциальное, нормальное).
2. **Стройте гистограммы** для визуальной проверки.
3. **Используйте статистические тесты** (например, критерий Колмогорова-Смирнова).

**5. Визуализация распределений**

**Пример с нормальным распределением:**

import matplotlib.pyplot as plt

import numpy as np

mu, sigma = 0, 1

data = np.random.normal(mu, sigma, 1000)

plt.hist(data, bins=30, density=True, alpha=0.6)

x = np.linspace(-4, 4, 100)

plt.plot(x, norm.pdf(x, mu, sigma), 'r-', lw=2)

plt.title("Нормальное распределение")

plt.show()

**6. Практические задачи**

**6.1. Задача о качестве продукции**

**Условие:** Завод производит микросхемы с вероятностью брака 5%. Какова вероятность, что в партии из 100 микросхем будет не более 3 бракованных?

**Решение (биномиальное распределение):**

prob = binom.cdf(3, 100, 0.05)

print(f"Вероятность: {prob:.4f}") # ~0.2578

**6.2. Задача о времени отклика сервера**

**Условие:** Среднее время отклика сервера — 2 секунды. Какова вероятность, что время отклика превысит 5 секунд?

**Решение (экспоненциальное распределение):**

lambda\_ = 1/2

prob = 1 - expon.cdf(5, scale=1/lambda\_)

print(f"Вероятность: {prob:.4f}") # ~0.0821