

Глава 2

Случайные величины, распределения, моменты

Содержание

2.1.	Случайность и случайные величины	20
2.2.	Средние, дисперсии и моменты	21
2.3.	Базовые распределения	22
2.4.	Счёт успехов: биномиальная модель	22
2.5.	Счёт редких событий: пуассоновская модель	25
2.6.	Непрерывные ошибки: нормальная модель	27
2.7.	Сравнение трёх распределений	29
2.8.	Задачи	30
	Литература	31

Аннотация

В этой главе мы подробнее разберём случайные величины и их распределения. Введём функцию вероятностей для дискретной величины, плотность вероятности и кумулятивную функцию распределения для непрерывной. Затем определим моменты и на примерах выведем среднее и дисперсию для биномиального, пуассоновского и нормального распределений.

2.1. Случайность и случайные величины

Откуда в физических данных возникает случайность? В классической физике её обычно связывают с неполным знанием начальных условий и большим числом степеней свободы. Брошенная монета движется по законам механики, но для предсказания результата нужно было бы знать её начальное положение, скорость, вращение, сопротивление воздуха и ещё множество деталей.

В квантовой физике вероятностный характер заложен в саму теорию. Даже если начальное состояние приготовлено полностью, теория предсказывает вероятности возможных результатов измерения.

Для статистического анализа в обоих случаях эксперимент даёт одну конкретную реализацию случайного процесса. При повторении опыта результат меняется, и это изменение должно быть описано вероятностной моделью.

В физике **случайной величиной** X будем называть экспериментально измеряемую величину, значение которой нельзя знать до проведения эксперимента. В математической формулировке X каждому возможному исходу эксперимента сопоставляет число [1].

Например, N может обозначать число нейтрино, зарегистрированных детектором за сутки. До начала измерения N — случайная величина. После его окончания мы получим определённое число, скажем

$$n = 17.$$

Чтобы описать N , нужно указать все возможные значения и вероятности их получения. Вместе они образуют **распределение случайной величины**.

Если X принимает отдельные значения x_k , распределение задаётся **функцией вероятностей** (*probability mass function*, PMF):

$$p_k = P(X = x_k), \quad \sum_k p_k = 1. \quad (2.1)$$

Для непрерывной случайной величины вводится **плотность вероятности** (*probability density function*, PDF):

$$P(a \leq X < b) = \int_a^b f(x) dx, \quad \int_{-\infty}^{+\infty} f(x) dx = 1. \quad (2.2)$$

Сумма в формуле 2.1 и второй интеграл в формуле 2.2 выражают условие нормировки: один из возможных результатов обязательно произойдёт.

Вероятность для непрерывной величины определяется интегралом плотности. Поэтому $P(X = x) = 0$, а вероятность попадания в конечный интервал может быть ненулевой. Для короткого интервала ширины Δx эту вероятность можно записать как $P(x \leq X < x + \Delta x) \approx f(x) \Delta x$. Если x измеряется в МэВ, плотность $f(x)$ имеет размерность МэВ⁻¹, и произведение $f(x) \Delta x$ остаётся безразмерным.

2.1.1. Кумулятивная функция распределения (CDF)

Нас обычно интересует вероятность получить результат меньше некоторого значения a . Для этого вводится **кумулятивная функция распределения**, или CDF (*cumulative distribution function*):

$$F(a) = P(X < a). \quad (2.3)$$

Для непрерывной случайной величины она равна интегралу плотности:

$$F(a) = \int_{-\infty}^a f(x) dx. \quad (2.4)$$

При движении a слева направо функция $F(a)$ монотонно растёт от нуля до единицы. Если плотность непрерывна, то

$$\frac{dF(a)}{da} = f(a). \quad (2.5)$$

Вероятность попадания в интервал $[a, b)$ теперь вычисляется одной строкой:

$$P(a \leq X < b) = F(b) - F(a). \quad (2.6)$$

Для непрерывного распределения со строго возрастающей F **квантиль** x_q определяется условием

$$F(x_q) = q. \quad (2.7)$$

Например, медиана $x_{0.5}$ делит такое распределение пополам: вероятность получить значение слева или справа от неё равна $1/2$. Квантили понадобятся нам при построении доверительных интервалов.

2.2. Средние, дисперсии и моменты

Распределение содержит всю вероятностную информацию о случайной величине. На практике часто хочется начать с нескольких чисел: где расположен центр распределения и насколько широко разбросаны его значения.

2.2.1. Математическое ожидание

Для произвольной функции $g(X)$ математическое ожидание определяется формулой

$$E[g(X)] = \begin{cases} \sum g(x_k)p_k, & X - \text{дискретная величина,} \\ \int_{-\infty}^{+\infty} g(x)f(x) dx, & X - \text{непрерывная величина.} \end{cases} \quad (2.8)$$

Положив $g(X) = X^n$, получим n -й **начальный момент** распределения:

$$\langle x^n \rangle = E[X^n]. \quad (2.9)$$

Первый момент — это среднее значение

$$\mu = \langle x \rangle = E[X]. \quad (2.10)$$

Формула напоминает определение центра масс. Если представить $f(x)$ как плотность единичной массы на числовой оси, то точка μ действительно будет её центром масс.

2.2.2. Дисперсия

Среднее указывает центр распределения, но ничего не говорит о его ширине. Для описания разброса используется **дисперсия**:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X - \mu)^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \langle x^2 \rangle - \langle x \rangle^2.\end{aligned}\quad (2.11)$$

Квадрат не позволяет положительным и отрицательным отклонениям взаимно уничтожиться. Корень из дисперсии называется **стандартным отклонением**:

$$\sigma_X = \sqrt{\text{Var}(X)}.\quad (2.12)$$

Стандартное отклонение имеет ту же размерность, что и сама величина X .

Для функции $g(X)$ определение остаётся тем же:

$$\text{Var}(g(X)) = \mathbb{E}[(g(X) - \mathbb{E}[g(X)])^2].\quad (2.13)$$

Все эти формулы предполагают, что соответствующие сумма или интеграл сходятся. У распределения Коши, которое мы встретим в следующей главе, математического ожидания и дисперсии нет. Более того, даже полный набор моментов не всегда определяет распределение однозначно. Поэтому моменты полезны, но не заменяют саму функцию распределения.

2.3. Базовые распределения

В этой главе нам понадобятся три распределения.

- Биномиальное распределение описывает число успехов в фиксированном числе испытаний.
- Распределение Пуассона описывает число событий за фиксированное время или в фиксированной области пространства.
- Нормальное распределение описывает непрерывную величину и часто возникает при сложении большого числа малых независимых вкладов.

Разберём их по порядку.

2.4. Счёт успехов: биномиальная модель

Предположим, что в эксперименте проводится n независимых испытаний. В каждом из них возможны два исхода, которые условно назовём успехом и неудачей. Вероятность успеха равна p и остаётся одной и той же во всех испытаниях.

Обозначим через K полное число успехов. Вероятность любой определённой последовательности из k успехов и $n - k$ неудач равна

$$p^k(1 - p)^{n-k}.$$

Успехи можно расположить среди n испытаний

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

различными способами. Поэтому

Биномиальное распределение задаётся функцией вероятностей

$$P(K = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (2.14)$$

Короткая запись имеет вид

$$K \sim \text{Bin}(n, p).$$

Условие нормировки следует непосредственно из бинома Ньютона:

$$\sum_{k=0}^n P(K = k) = [p + (1-p)]^n = 1.$$

2.4.1. Пример биномиального распределения

Пусть зарегистрировано n распадов W -бозона. Каждый распад либо происходит по каналу $W \rightarrow \mu\nu$, либо нет. Если вероятность этого канала равна p , то число распадов по данному каналу имеет биномиальное распределение.

На рисунке 2.1 показаны три распределения при $p = 0.3$ и $n = 5, 20$ и 100 .

```
from math import comb
import numpy as np

p = 0.3
for n in [5, 20, 100]:
    k = np.arange(n + 1)
    pmf = np.array([comb(n, i) * p**i * (1-p)**(n-i) for i in k])
```

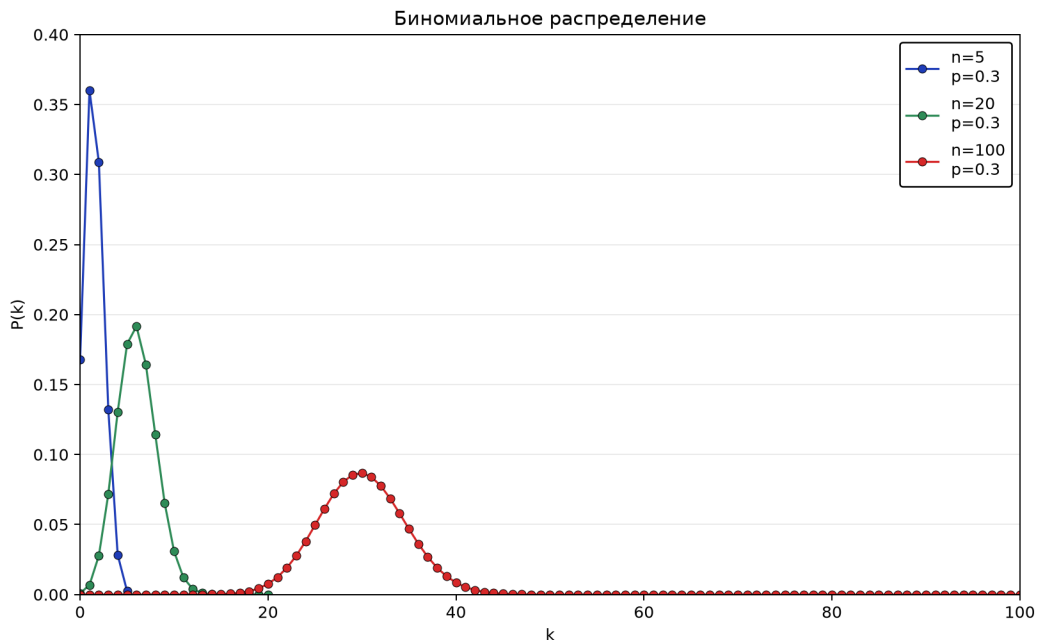


Рис. 2.1. Биномиальные распределения при $p = 0.3$ и трёх значениях числа испытаний: $n = 5, 20$ и 100

С ростом n максимум смещается вслед за средним значением np . Абсолютная ширина распределения растёт как $\sqrt{np(1-p)}$, а относительная ширина уменьшается

как $1/\sqrt{n}$. При больших np и $n(1-p)$ форма становится близкой к гауссовой, хотя величина K по-прежнему принимает только целые значения.

2.4.2. Моменты биномиального распределения

Для биномиального распределения

$$\mathbb{E}[K] = np, \quad \text{Var}(K) = np(1-p). \quad (2.15)$$

Выведем обе формулы. Этот вывод заодно показывает полезный приём с факториальными моментами.

Вывод

2.4.2.1. Среднее

Начнём с определения математического ожидания:

$$\mathbb{E}[K] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}.$$

Используем тождество

$$k \binom{n}{k} = n \binom{n-1}{k-1}.$$

Тогда

$$\begin{aligned} \mathbb{E}[K] &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\ &= np. \end{aligned} \quad (2.16)$$

Последняя сумма равна единице по биному Ньютона.

Вывод

2.4.2.2. Дисперсия

Для вычисления дисперсии удобно начать с факториального момента $K(K-1)$:

$$\mathbb{E}[K(K-1)] = \sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k}.$$

Поскольку

$$k(k-1) \binom{n}{k} = n(n-1) \binom{n-2}{k-2},$$

тот же приём даёт

$$\mathbb{E}[K(K-1)] = n(n-1)p^2. \quad (2.17)$$

Обычный второй момент связан с факториальным тождеством

$$K^2 = K(K-1) + K.$$

Поэтому

$$\begin{aligned}
 \text{Var}(K) &= E[K^2] - E[K]^2 \\
 &= n(n-1)p^2 + np - n^2p^2 \\
 &= np(1-p).
 \end{aligned}
 \tag{2.18}$$

2.5. Счёт редких событий: пуассоновская модель

Теперь рассмотрим другой эксперимент. Радиоактивный источник работает в течение заданного времени Δt , а детектор считает распады. Число испытаний заранее не фиксировано. Задана средняя частота регистрации r , и среднее число событий за время наблюдения равно

$$\lambda = r \Delta t. \tag{2.19}$$

Если интенсивность не меняется со временем, а события происходят независимо, число зарегистрированных событий N подчиняется распределению Пуассона.

Распределение Пуассона задаётся функцией вероятностей

$$P(N = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots, \tag{2.20}$$

где $\lambda > 0$ — среднее число событий. Короткая запись имеет вид

$$N \sim \text{Pois}(\lambda).$$



БИОГРАФИЯ

Симеон Дени Пуассон

1781–1840

Симеон Дени Пуассон — французский математик и физик. Он работал над теорией вероятностей, механикой, теорией упругости, электричеством и магнетизмом. Распределение, носящее его имя, появилось в исследовании вероятностей судебных решений и сегодня постоянно используется при подсчёте событий в физическом эксперименте.

Источник: MacTutor; фото MacTutor

2.5.1. Пример распределения Пуассона

На рисунке 2.2 показаны распределения Пуассона при $\lambda = 1, 4$ и 10 .

```

from math import factorial
import numpy as np

for lam in [1, 4, 10]:
    k = np.arange(26)
    pmf = np.exp(-lam) * lam**k / np.array([factorial(i) for i in k])

```

При $\lambda = 1$ основная вероятность сосредоточена около нуля и распределение заметно асимметрично. С ростом λ максимум смещается вправо, ширина увеличивается, а форма становится всё более симметричной. При больших λ распределение Пуассона можно приближать нормальным распределением. Условия

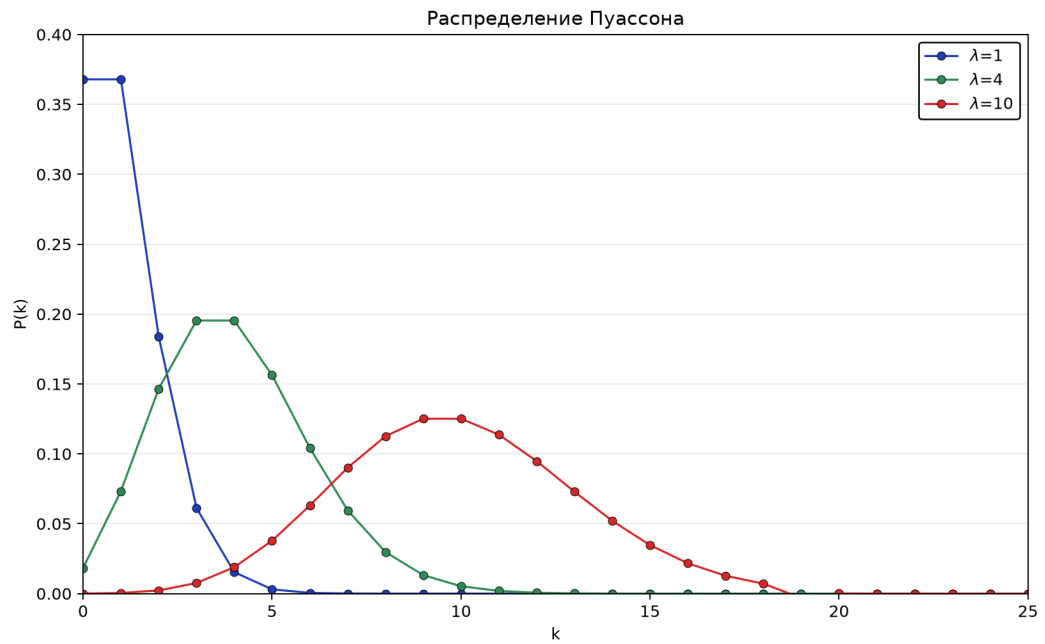


Рис. 2.2. Распределения Пуассона при среднем числе событий $\lambda = 1, 4$ и 10

и точность такого приближения мы обсудим в главе о центральной предельной теореме.

2.5.2. Моменты распределения Пуассона

У распределения Пуассона среднее и дисперсия равны одному и тому же параметру:

$$E[N] = \lambda, \quad \text{Var}(N) = \lambda. \quad (2.21)$$

Вывод

2.5.2.1. Среднее

По определению математического ожидания

$$E[N] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!}.$$

Слагаемое при $k = 0$ равно нулю. В остальных слагаемых вынесем один множитель λ и сдвинем индекс $j = k - 1$:

$$\begin{aligned} E[N] &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned} \quad (2.22)$$

Вывод

2.5.2.2. Дисперсия

Снова начнём со второго факториального момента:

$$E[N(N-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!}.$$

Первые два слагаемых равны нулю, а при $k \geq 2$

$$\frac{k(k-1)}{k!} = \frac{1}{(k-2)!}.$$

Отсюда

$$\begin{aligned} E[N(N-1)] &= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda^2. \end{aligned} \quad (2.23)$$

Используя $N^2 = N(N-1) + N$, получаем

$$\begin{aligned} \text{Var}(N) &= E[N^2] - E[N]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned} \quad (2.24)$$

2.6. Непрерывные ошибки: нормальная модель

Биномиальное и пуассоновское распределения описывают целое число событий. Результат измерения энергии, времени или координаты обычно рассматривается как непрерывная величина. Самая известная модель для такой величины — нормальное, или гауссово, распределение.

Нормальное распределение с параметрами μ и $\sigma > 0$ имеет плотность

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \quad (2.25)$$

Здесь μ — среднее значение, а σ — стандартное отклонение. Короткая запись имеет вид

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Почему это распределение называется нормальным? Карл Пирсон, который ввёл это название в широкое употребление, позднее заметил его неудачную сторону:

Много лет назад я назвал кривую Лапласа-Гаусса нормальной кривой. Это название, избегая международного спора о приоритете, имеет недостаток: оно заставляет думать, что все прочие распределения частот в том или ином смысле «ненормальны».

К. Пирсон, 1920 [2]

Причина частого появления гауссовой формы связана с центральной предельной теоремой. Сумма большого числа малых независимых вкладов при достаточно общих условиях приближается нормальным распределением. Этому будет посвящена отдельная глава.

2.6.1. Пример нормального распределения

На рисунке 2.3 показаны три плотности с параметрами $(\mu, \sigma) = (0, 1)$, $(0, 2)$ и $(2, 1)$.

```
import numpy as np

x = np.linspace(-6, 6, 400)
for mu, sigma in [(0, 1), (0, 2), (2, 1)]:
    pdf = np.exp(-0.5*((x-mu)/sigma)**2)/(sigma*np.sqrt(2*np.pi))
```



Рис. 2.3. Плотности нормального распределения при $(\mu, \sigma) = (0, 1)$, $(0, 2)$ и $(2, 1)$

Синие и красные кривые имеют одинаковую ширину $\sigma = 1$ и разные средние: изменение μ сдвигает распределение вдоль оси x . У зелёной кривой $\sigma = 2$, поэтому она шире и ниже. Площадь под каждой кривой по-прежнему равна единице.

2.6.2. Моменты распределения Гаусса

Параметры нормального распределения совпадают с его средним и стандартным отклонением:

$$E[X] = \mu, \quad \text{Var}(X) = \sigma^2. \quad (2.26)$$

Вывод

2.6.2.1. Среднее

Перейдём к стандартной нормальной величине

$$z = \frac{x - \mu}{\sigma}, \quad x = \mu + \sigma z, \quad dx = \sigma dz,$$

плотность которой обозначим

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \quad (2.27)$$

Тогда

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{+\infty} (\mu + \sigma z) \varphi(z) dz \\
&= \mu \int_{-\infty}^{+\infty} \varphi(z) dz + \sigma \int_{-\infty}^{+\infty} z \varphi(z) dz \\
&= \mu.
\end{aligned} \tag{2.28}$$

Первый интеграл равен единице по условию нормировки. Второй равен нулю, потому что $z\varphi(z)$ — нечётная функция.

Вывод

2.6.2.2. Дисперсия

Та же замена переменной даёт

$$\text{Var}(X) = \sigma^2 \int_{-\infty}^{+\infty} z^2 \varphi(z) dz.$$

Оставшийся интеграл вычислим по частям. Поскольку

$$\varphi'(z) = -z\varphi(z),$$

получаем

$$\begin{aligned}
\int_{-\infty}^{+\infty} z^2 \varphi(z) dz &= [-z\varphi(z)]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \varphi(z) dz \\
&= 1.
\end{aligned}$$

Следовательно,

$$\text{Var}(X) = \sigma^2. \tag{2.29}$$

2.7. Сравнение трёх распределений

В таблице 2.1 собраны три рассмотренные модели. В последнем столбце перечислены условия, которые относятся к устройству повторяемого эксперимента.

Таблица 2.1. Основные свойства биномиального, пуассоновского и нормального распределений

Распределение	Среднее	Дисперсия	Условия применения
$K \sim \text{Bin}(n, p)$	np	$np(1 - p)$	Фиксированное n ; независимые испытания; постоянное p
$N \sim \text{Pois}(\lambda)$	λ	λ	Счёт событий в фиксированном интервале; постоянная интенсивность; независимость

Распределение	Среднее	Дисперсия	Условия применения
$X \sim \mathcal{N}(\mu, \sigma^2)$	μ	σ^2	Непрерывная величина; приближение для суммы многих малых независимых вкладов

Итоги главы

- Случайная величина описывается распределением. Для дискретной величины оно задаёт вероятности отдельных значений, для непрерывной — плотность вероятности.
- Математическое ожидание указывает центр распределения, а дисперсия измеряет средний квадрат отклонения от этого центра. Эти величины существуют только при сходимости соответствующих суммы или интеграла.
- Биномиальное распределение относится к фиксированному числу испытаний, распределение Пуассона — к числу событий в фиксированном интервале, а нормальное распределение — к непрерывной величине.
- Выбор распределения включает условия, при которых повторяется эксперимент: независимость испытаний, постоянство вероятности или интенсивности и способ набора данных.

2.8. Задачи

Задача 1. Предел биномиального распределения

Покажите, что при

$$n \rightarrow \infty, \quad p \rightarrow 0, \quad np = \lambda = \text{const}$$

биномиальное распределение переходит в распределение Пуассона:

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{\lambda^k e^{-\lambda}}{k!}.$$

Подсказка. Положите $p = \lambda/n$ и рассмотрите предел при фиксированном k .

Задача 2. Предел Пуассона к гауссову распределению

Используя формулу Стирлинга, покажите, что при больших λ

$$P(N = n) \approx \frac{1}{\sqrt{2\pi\lambda}} \exp\left[-\frac{(n - \lambda)^2}{2\lambda}\right].$$

Задача 3. Моменты равномерного распределения

Пусть $X \sim U(-1, 1)$. Найдите:

1. $E[X]$;
2. $E[X^2]$;
3. $\text{Var}(X)$.

Задача 4. Когда моменты однозначно задают распределение

Пусть

$$E[X^{2n}] = 1, \quad E[X^{2n+1}] = 0.$$

Покажите, что $X = \pm 1$ с равными вероятностями.

Подсказка. Используйте $E[(X^2 - 1)^2] = 0$.

Задача 5. Восстановление распределения на конечном носителе

Пусть X принимает значения $-1, 0, 1$, причём

$$E[X] = 0, \quad E[X^2] = \frac{1}{2}.$$

Найдите вероятности трёх значений.

Задача 6. Факториальные моменты распределения Пуассона

Для $N \sim \text{Pois}(\lambda)$ покажите, что

$$E[N(N-1) \cdots (N-r+1)] = \lambda^r.$$

Литература

- [1] Feller, W. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons. 1971.
[2] Pearson, K. Notes on the History of Correlation. *Biometrika*, **13**, 25–45. 1920. doi:10.1093/biomet/13.1.25.