Теория вероятностей

Курс

ML-РАЗРАБОТЧИК

01 ЦИФРОВАЯ КАФЕДРА / МГТУ ИМ. Н.Э. БАУМАНА

Введение

Случайность окружает нас повсюду. **Теория вероятностей** — это математический аппарат, который позволяет анализировать случайные события логически обоснованным образом.

Вероятность события — это число, показывающее, насколько вероятно наступление этого события. Это число всегда находится в диапазоне от 0 до 1, где 0 означает невозможность, а 1 — полную уверенность.

Теория вероятностей изучает закономерности, возникающие в **случайных экспериментах** (точнее, в их математических моделях). Случайные эксперименты в теории вероятностей моделируются с помощью *случайных величин*.

Случайные величины

Случайная величина — это *функция*, которая каждому возможному результату эксперимента присваивает определённое действительное число.

Её можно рассматривать как способ «кодирования» исходов эксперимента.

Нас будет интересовать набор значений случайной величины и вероятности того, что она принимает те или иные значения.

Вероятности событий для случайных величин

Мы рассматриваем вероятность событий, связанных со случайной величиной X. Возможные события, которые нас могут интересовать:

- Событие, при котором случайная величина принимает конкретное значение: $\{X = a\},\ a \in \mathbb{R}$;
- Событие, при котором случайная величина меньше заданного значения: $\{X < a\},\ a \in \mathbb{R}$;
- Событие, при котором случайная величина лежит в определённом интервале: $\{a \leq X < b\},\ a, b \in \mathbb{R},\ a < b$.

В общем случае событие может быть представлено как:

$\{X \in A\}$, где $A \subset \mathbb{R}$.

Классификация случайных величин

Случайные величины можно классифицировать в зависимости от мощности их множества значений:

- **Дискретные случайные величины** имеют конечное или счётное множество значений.

  Примеры: $\{1, \ldots, n\}$, $\mathbb{N}$, $\mathbb{Z}$.

- **Непрерывные случайные величины** имеют несчётное множество значений.

  Примеры: $[0, 1]$, $\mathbb{R}$.

Классификация случайных величин

Дискретные случайные величины

Для задания распределения дискретной случайной величины необходимо:

- перечислить возможные значения $a_1, a_2, a_3, \ldots;$
- задать вероятности $p_1, p_2, p_3, \ldots,$ такие что:

$$p_i \geq 0, \quad \sum_{i=1}^{\infty} p_i = 1;$$

- определить $\mathbb{P}(X = a_i) = p_i.$

Удобно представить это в таблице:

X	a_1	a_2	a_3	$\ldots$
$\mathbb{P}$	p_1	p_2	p_3	$\ldots$

Вероятность события A рассчитывается по формуле: $\mathbb{P}(X \in A) = \sum_{i: a_i \in A} p_i.$

Пример дискретной случайной величины

1. Константа.
2. Распределение Бернулли.
3. Равномерное распределение на конечном множестве.
4. Биномиальное распределение.
5. Распределение Пуассона.

Распределение Бернулли

Представьте, что у нас есть сайт, на котором размещен баннер, и мы хотим извлечь из данных что-то полезное для себя.

Например, в отношении баннера мы можем построить модель, что каждый посетитель может совершить только два типа действий: «успешное» (клик) и «неуспешное» (проигнорировать баннер).

Такое событие можно описать распределением Бернулли, так как существует только два возможных исхода.

Распределение Бернулли $\mathbb{B}_p$

Случайная величина X подчиняется распределению Бернулли с параметром $p \in [0, 1]$ если:

$X = 1$ с вероятностью p;

$X = 0$ с вероятностью 1 - p.

Таблица распределения:

X	0	1
$\mathbb{P}$	$1 - p$	p

Параметр p называют вероятностью «успеха». Для обозначения будем писать:

$$X \sim \mathbb{B}_{1/2}, \quad X \sim \mathbb{B}_{1/4}, \quad X \sim \mathbb{B}_1, \ldots$$

Примеры дискретных случайных величин: Распределение Бернулли $\mathbb{B}_p$

Примеры распределения Бернулли:

- Орёл/Решка при подбрасывании монеты.
- Выиграл/Проиграл в игре.
- Купил/Не купил товар.
- Попал в цель/Не попал в цель.
- Любое другое бинарное событие.

Биномиальное распределение

Вернемся к нашему сайту. Мы знаем, что какой-то пользователь заходит на сайт каждый день, и видит там баннер. Мы хотим узнать, сколько кликов он совершит за 100 дней.

Мы знаем вероятность клика посетителя и знаем, что каждый заход на сайт не зависит от предыдущих заходов. В этом случае общее количество кликов за 100 дней можно описать биномиальным распределением.

Биномиальное распределение моделирует ситуацию, где проводится n независимых экспериментов (в данном случае 100), каждый из которых имеет два исхода («клик» или «не клик»), с фиксированной вероятностью успеха p.

Примеры дискретных случайных величин: Биномиальное распределение $\mathbb{B}_{n,p}$

Случайная величина Y имеет биномиальное распределение $\mathbb{B}_{n,p}$

с параметрами $n \in \mathbb{N}$ и $p \in [0, 1]$,

если: $Y = X_1 + X_2 + \cdots + X_n$,

где $X_1, X_2, \ldots, X_n \sim \mathbb{B}_p$ — независимые случайные величины.

Интерпретация: Y — это количество «успехов» в n независимых испытаниях Бернулли.

Вероятность распределения: Вероятность того, что будет k «успехов» в n испытаниях, равна:

$$\mathbb{P}(Y = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

где C_n^k — биномиальный коэффициент: $$C_n^k = \frac{n!}{k!(n-k)!}.$$

Примеры дискретных случайных величин: Биномиальное распределение $\mathbb{B}_{n,p}$

Примеры применения биномиального распределения:

- Количество орлов в n бросках монеты.
- Число покупок, совершённых n посетителями магазина.
- Любое количество «успехов» в n независимых испытаниях Бернулли.

Равномерное распределение на конечном множестве

Случайная величина X имеет равномерное распределение на множестве $\{a_1, a_2, ..., a_n\}$, если она принимает каждое из этих значений с одинаковой вероятностью 1/n.

Таблица распределения вероятностей в этом случае имеет вид:

X	a_1	a_2	...	a_n
P	1/n	1/n	...	1/n

Примеры равномерного распределения:

- Бросок игральной кости (n=6)
- Случайный выбор карты из колоды (n=52)
- Случайный выбор элемента из массива фиксированной длины
- Генерация случайного числа в заданном диапазоне целых чисел

Важные свойства:

- Все значения равновероятны
- Сумма вероятностей равна 1
- Каждое значение имеет вероятность 1/n

Равномерное распределение на конечном множестве

Случайная величина X имеет равномерное распределение на множестве $\{a_1, a_2, ..., a_n\}$, если она принимает каждое из этих значений с одинаковой вероятностью 1/n.

Таблица распределения вероятностей в этом случае имеет вид:

X	a_1	a_2	...	a_n
P	1/n	1/n	...	1/n

Примеры равномерного распределения:

- Бросок игральной кости (n=6)
- Случайный выбор карты из колоды (n=52)
- Случайный выбор элемента из массива фиксированной длины
- Генерация случайного числа в заданном диапазоне целых чисел

Важные свойства:

- Все значения равновероятны
- Сумма вероятностей равна 1
- Каждое значение имеет вероятность 1/n

Распределение Пуассона $\Pi_\lambda, \lambda > 0$

Случайная величина X имеет распределение Пуассона с параметром λ>0, если X принимает значения k=0,1,2,… (и только их) с вероятностями:

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Свойства и применение:

- Параметр λ представляет собой интенсивность или среднее число событий
- Описывает количество событий в фиксированном временном интервале или пространственной области
- Является предельным распределением для биномиального распределения $\mathbb{B}_{n,p}$ при $p \to 0$, $np \to \lambda$

Распределение Пуассона $\Pi_\lambda, \lambda > 0$

Примеры использования:

- Количество опечаток на странице текста
- Число звонков в колл-центр за час
- Количество дефектов на участке ткани
- Число частиц, испускаемых радиоактивным веществом за единицу времени
- Количество снарядов, попадающих в заданную область (при условии случайного распределения попаданий)

Непрерывные случайные величины. Определение плотности

Функция $f : \mathbb{R} \to \mathbb{R}$ называется плотностью на $\mathbb{R}$, если выполняются следующие условия:

1. $f(u) \geq 0$ для любого $u \in \mathbb{R}$ *(неотрицательность)*;
2. $\displaystyle\int_{-\infty}^{+\infty} f(u)\, du = 1$ *(нормировка)*.

Непрерывные случайные величины. Общие правила перехода

При переходе от дискретного к непрерывному случаю используются следующие правила замены в формулах:

$$\sum_i \longrightarrow \int_{\mathbb{R}},$$
$$a_i \longrightarrow u,$$
$$p_i \longrightarrow f(u)du$$

Пример: Условие нормировки в дискретном и непрерывном случаях:

> $\sum_i p_i = 1 \longrightarrow \int_{\mathbb{R}} f(u)du = 1$

Эти правила позволяют легко переносить формулы из дискретного случая в непрерывный, заменяя суммы на интегралы и соответствующие величины на их непрерывные аналоги.

Как определить непрерывную случайную величину?

Для задания распределения непрерывной случайной величины X необходимо определить её функцию плотности.

Если X имеет плотность распределения $f(u)$, то вероятность того, что X принадлежит подмножеству $A \subset \mathbb{R}$, вычисляется по формуле:

$$\mathbb{P}(X \in A) = \int_A f(u)\, du.$$

Функция $f(u)$ позволяет вычислять вероятности для любого подмножества A множества действительных чисел.

Частные случаи вероятностей для непрерывных случайных величин

Для любых $a, b \in \mathbb{R}$, где $a < b$, вероятность событий вычисляется следующим образом:

$A = (-\infty, a]$, тогда $\mathbb{P}(X \in A) = \mathbb{P}(X \le a) = \int_{-\infty}^{a} f(u)\,du$;

$A = [a, +\infty)$, тогда $\mathbb{P}(X \in A) = \mathbb{P}(X \ge a) = \int_{a}^{+\infty} f(u)\,du$;

$A = [a, b]$, тогда $\mathbb{P}(X \in A) = \mathbb{P}(a \le X \le b) = \int_{a}^{b} f(u)\,du$;

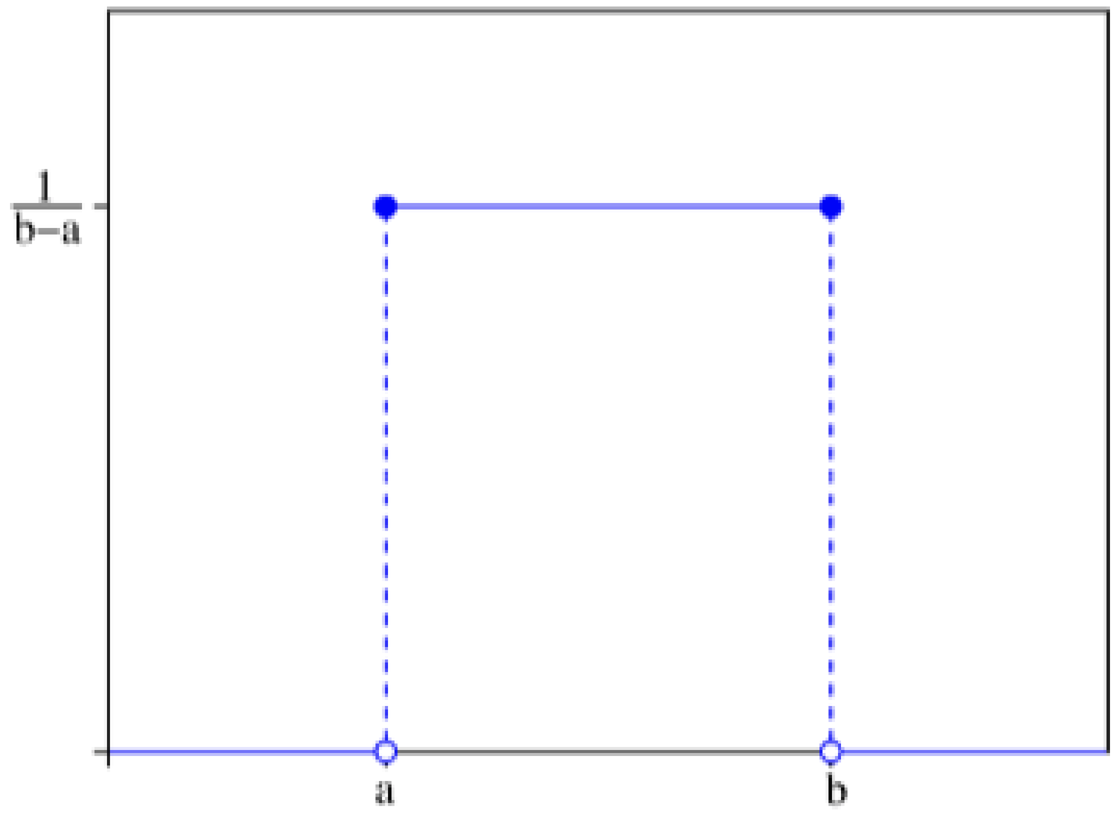
$A = \{a\}$, тогда $\mathbb{P}(X \in A) = \mathbb{P}(X = a) = \int_{a}^{a} f(u)\,du = 0$.

Замечание: По свойствам определённого интеграла, вероятность того, что непрерывная случайная величина *X* с плотностью *f* принимает точное значение $a \in \mathbb{R}$,равна нулю:

$$\mathbb{P}(X = a) = 0.$$

Равномерное распределение на отрезке $[a,b]$

Случайная величина X имеет равномерное распределение на отрезке $[a,b]$, если её плотность $f(u)$ имеет вид:



$$f(u) = \begin{cases} \frac{1}{b-a}, & u \in [a,b], \\ 0, & u \notin [a,b]. \end{cases}$$

Равномерное распределение на отрезке $[a, b]$

Примеры применения:

- Погрешность округления числа
- Угол поворота стрелки в ЧГК
- Равномерное распределение на [0, 1] является базовым для генерации псевдослучайных чисел в компьютере
- Априорное распределение на параметры модели в условиях отсутствия априорной информации.

Экспоненциальное распределение

Представьте, что вы стоите на автобусной остановке и ждете прибытия следующего автобуса. Если автобусы ходят нерегулярно, время между их прибытием может быть описано **экспоненциальным распределением**.

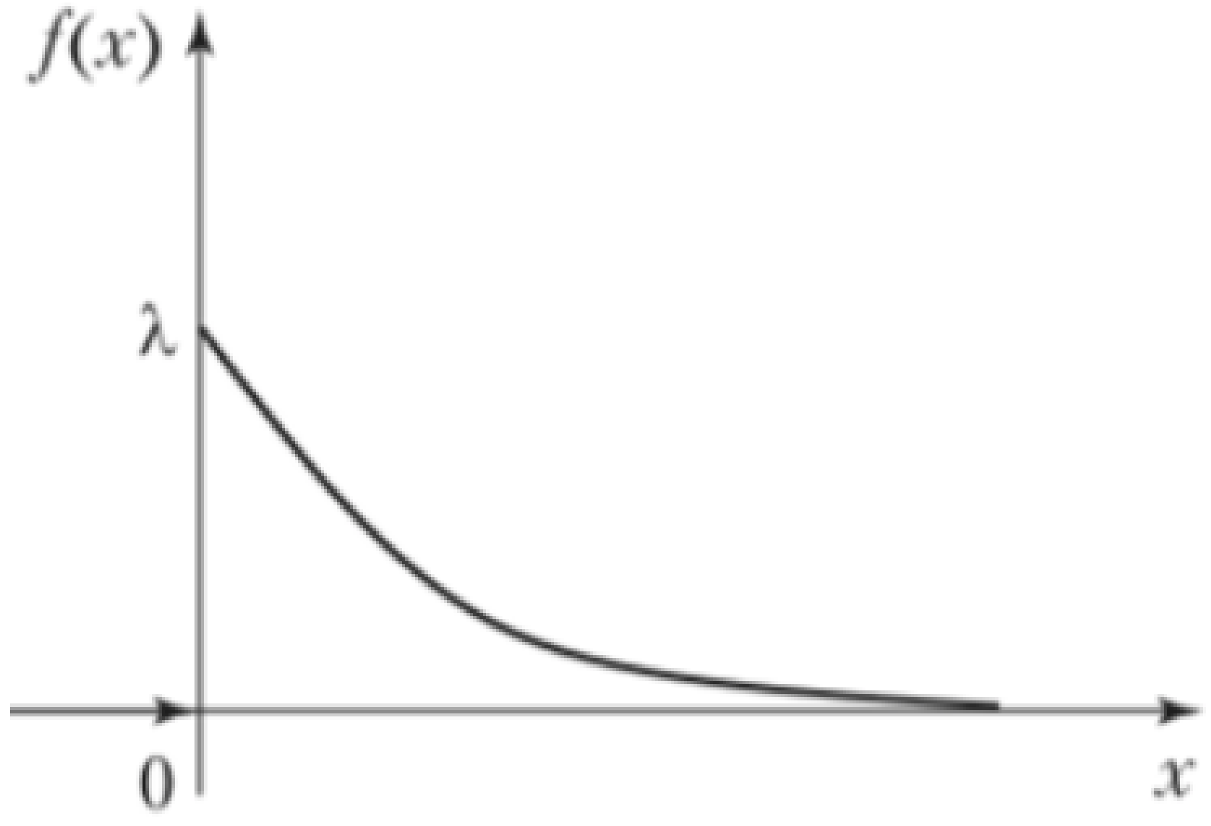
- **Свойство отсутствия памяти:** Если вы уже ждали определенное время, это не влияет на вероятность того, что автобус прибудет в ближайшие минуты.
- **Интенсивность:** При этом мы знаем, что на одном маршруте есть 20 автобусов, а на другом 10. Это значит, что в среднем мы будем ждать автобус на одном маршруте дольше, чем на другом, но при этом временные промежутки между прибытиями автобусов будут случайными.

Экспоненциальное распределение описывает случайные процессы, в которых события происходят с постоянной средней интенсивностью.

Экспоненциальное распределение $\text{Exp}(\lambda),\ \lambda > 0$

Случайная величина X подчиняется экспоненциальному (показательному) распределению с параметром $\lambda > 0$, если её функция плотности задана следующим образом:

$$f(u) = \begin{cases} \lambda e^{-\lambda u}, & u \geq 0, \\ 0, & u < 0. \end{cases}$$



На графике показана функция плотности $f(u)$ экспоненциального распределения, где $\lambda > 0$.

Экспоненциальное распределение $\text{Exp}(\lambda), \lambda > 0$

Распределение Пуассона описывает количество событий, происходящих за фиксированный интервал времени (или пространства), при условии, что события происходят независимо с постоянной средней интенсивностью λ.

Экспоненциальное распределение описывает время между двумя последовательными событиями, которые происходят по распределению Пуассона. Это распределение характеризуется параметром λ, который также является средней интенсивностью событий.

Примеры непрерывных случайных величин: Экспоненциальное распределение $\mathrm{Exp}(\lambda),\ \lambda > 0$

Примеры использования экспоненциального распределения:

- Время ожидания:
  - время до прибытия нужного автобуса;
  - время между покупками в магазине;
  - время между звонками в колл-центр.
- Срок эксплуатации:
  - время работы лампочки до перегорания;
  - время работы других приборов до выхода из строя.

Нормальное распределение

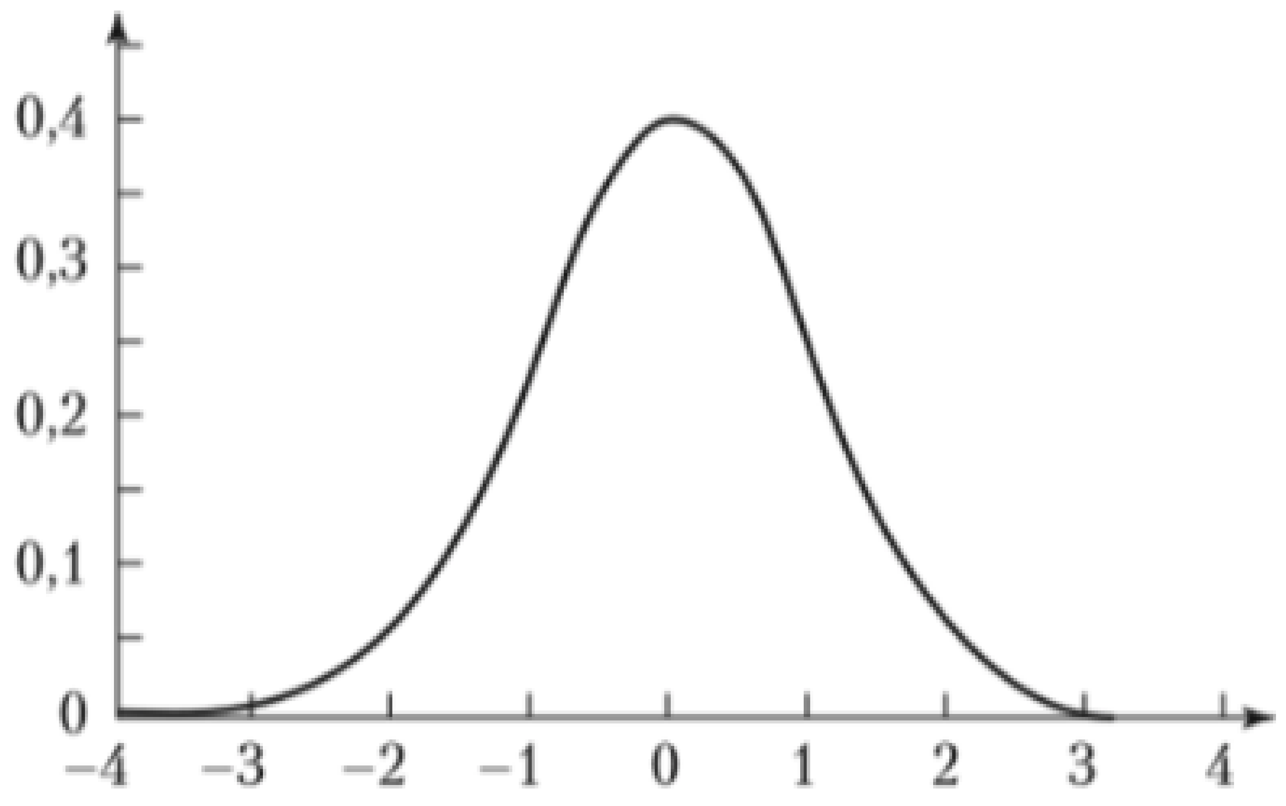
Представим ситуацию с ростом взрослых мужчин в определённой стране. Обычно, если взять достаточную по размеру выборку мужского населения, их рост будет распределён примерно по нормальному закону:

- Большинство мужчин будет иметь рост, близкий к некоторому среднему значению, например, около 175 см.
- По мере удаления от среднего значения количество людей, соответствующих такому росту, будет уменьшаться. Очень высоких или очень низких людей намного меньше, чем тех, чей рост близок к среднему.
- Таким образом, если мы построим гистограмму распределения роста по выборке, мы увидим «колоколобразную» кривую, характерную для нормального распределения: пик в районе среднего роста и плавное убывание числа людей по мере перехода к крайним значениям.

Нормальное распределение $\mathcal{N}(0,1)$

Случайная величина X имеет стандартное нормальное распределение (или стандартное гауссовское), если её плотность задаётся следующим образом:

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$



Обозначение: $X \sim \mathcal{N}(0,1)$.

Нормальное распределение $\mathcal{N}(a, \sigma^2)$

В общем случае случайная величина Y имеет нормальное распределение (или гауссовское) с параметрами $a \in \mathbb{R}$ и $\sigma > 0$,

если: $Y = a + \sigma X$,

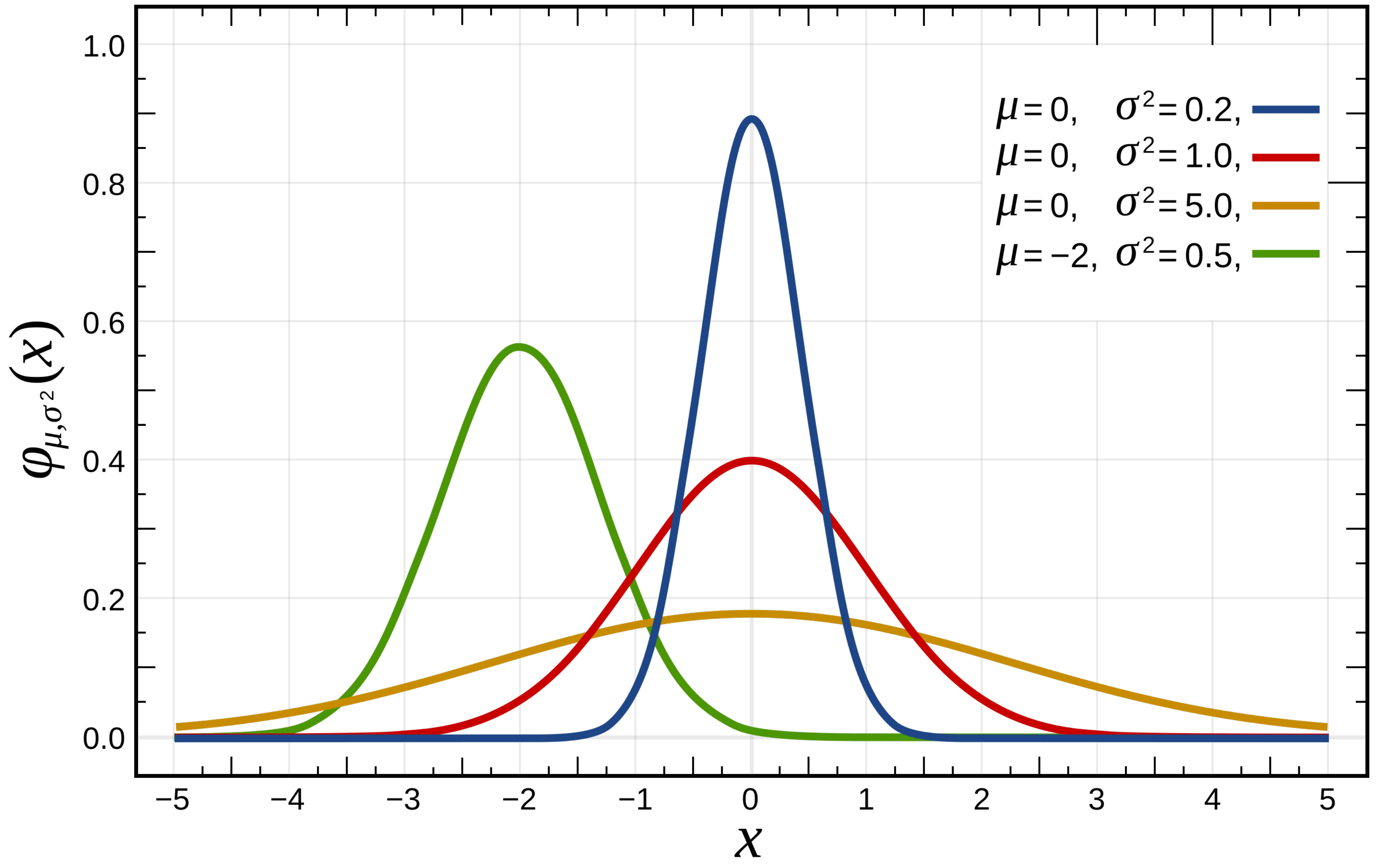
где $X \sim \mathcal{N}(0, 1)$.

Нормальное распределение $\mathcal{N}(a, \sigma^2)$

Плотность Y задаётся формулой:

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-a)^2}{2\sigma^2}}.$$

Обозначение: $Y \sim \mathcal{N}(a, \sigma^2)$.



Примеры непрерывных случайных величин: Нормальное распределение $\mathcal{N}(a, \sigma^2)$

Примеры применения нормального распределения:

- Отклонение снаряда от цели.
- Ошибки измерений.
- Центральная предельная теорема: суммы одинаковых независимых случайных величин часто распределены нормально.

Условная вероятность

Игральная кость подбрасывается один раз. Известно, что выпало более трёх очков. Какова вероятность того, что выпало нечётное число очков?

Распределение исходов:

X	1	2	3	4	5	6
$\mathbb{P}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Условная вероятность

Игральная кость подбрасывается один раз. Известно, что выпало более трёх очков. Какова вероятность того, что выпало нечётное число очков?

Распределение исходов:

X	1	2	3	4	5	6
$\mathbb{P}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Решение:

- Возможные исходы при условии $X > 3$: X = 4, 5, 6;
- Нечётные исходы среди них: $X = 5$;
- Условная вероятность:

$$\mathbb{P}(\text{нечётное} \mid X > 3) = \frac{\mathbb{P}(X = 5)}{\mathbb{P}(X > 3)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Ответ: $\frac{1}{3}$.

Условная вероятность

Определение: Условной вероятностью события S при условии, что произошло событие H, называется величина:

$$\mathbb{P}(S \mid H) = \frac{\mathbb{P}(S \cap H)}{\mathbb{P}(H)},$$

где $\mathbb{P}(H) > 0.$

Примечание: Если $\mathbb{P}(H) = 0$, то условная вероятность $\mathbb{P}(S \mid H)$ считается неопределённой.

Условная вероятность: Формула полной вероятности

Разбиением называется конечный или счётный набор попарно непересекающихся событий $H_1, H_2, \ldots,$ объединение которых составляет тождественное событие (множество всех возможных результатов эксперимента).

События $H_1, H_2, \ldots,$ образующие разбиение, называют **гипотезами.**

Обычно для некоторого события S известны вероятности $\mathbb{P}(S \mid H_i)$ и $\mathbb{P}(H_i)$. Возникает вопрос: как, используя эти данные, вычислить вероятность события S?

Формула полной вероятности

Разложим событие S в сумму попарно непересекающихся событий:

$$S = S \cap (H_1 \cup H_2 \cup \ldots) = SH_1 \cup SH_2 \cup \ldots$$

Далее:

$$\mathbb{P}(S) \overset{\text{св-ва вер.}}{=} \sum_i \mathbb{P}(SH_i) \overset{\text{опр. усл. вер.}}{=} \sum_i \mathbb{P}(H_i)\mathbb{P}(S|H_i).$$

Вывод: Это и есть **формула полной вероятности!**

Формула полной вероятности

Пример: Два стрелка

Два стрелка подбрасывают (симметричную) монетку и выбирают, кто из них будет стрелять по мишени. Первый стрелок попадает по мишени с вероятностью 1, второй стрелок — с вероятностью 10^{-5}.

Вопрос: С какой вероятностью в результате эксперимента пуля попадет в мишень?

Формула полной вероятности

Пример: Два стрелка

Решение:

1. Пусть событие S = "пуля попала в мишень"
2. Возможные исходы подбрасывания монеты:
   - H_1: "выбран первый стрелок" с $P(H_1) = 1/2$
   - H_2: "выбран второй стрелок" с $P(H_2) = 1/2$
3. Условные вероятности:
   - $P(S|H_1) = 1$
   - $P(S|H_2) = 10^{-5}$
4. По формуле полной вероятности:

$$\begin{aligned} P(S) &= P(H_1)P(S|H_1) + P(H_2)P(S|H_2) \\ &= \left(\frac{1}{2} \times 1\right) + \left(\frac{1}{2} \times 10^{-5}\right) \\ &= 0.5 + 0.5 \times 10^{-5} \\ &= 0.500005 \end{aligned}$$

Ответ: Вероятность попадания в мишень равна 0.500005 или примерно 50.0005%

Формула Байеса

Формула Байеса позволяет пересчитать вероятности гипотез после получения данных о результате эксперимента.

Формула:

$$\mathbb{P}(H_k \mid S) = \frac{\mathbb{P}(H_k \cap S)}{\mathbb{P}(S)} \quad \text{(определение условной вероятности).}$$

Используя формулу полной вероятности, получаем:

$$\mathbb{P}(H_k \mid S) = \frac{\mathbb{P}(H_k)\mathbb{P}(S \mid H_k)}{\sum_i \mathbb{P}(H_i)\mathbb{P}(S \mid H_i)}.$$

Интерпретация: Формула Байеса используется для уточнения вероятностей гипотез H_k с учётом новой информации S.

Формула Байеса

Пример: Два стрелка (продолжение)

Два стрелка подбрасывают (симметричную) монетку и выбирают, кто из них будет стрелять по мишени. Первый стрелок попадает по мишени с вероятностью 1, второй стрелок — с вероятностью 10^{-5}.

Вопрос: По мишени попали. С какой вероятностью это сделал первый стрелок? А второй?

Формула Байеса

Пример: Два стрелка (продолжение)

Решение:

1. Пусть событие S = "пуля попала в мишень"
2. Нужно найти $P(H_1|S)$ и $P(H_2|S)$
3. По формуле Байеса:

$$P(H_1|S) = \frac{P(H_1)P(S|H_1)}{P(S)} = \frac{0.5 \times 1}{0.500005} \approx 0.999990$$

$$P(H_2|S) = \frac{P(H_2)P(S|H_2)}{P(S)} = \frac{0.5 \times 10^{-5}}{0.500005} \approx 0.000010$$

Ответ:

- Вероятность того, что попал первый стрелок ≈ 0.999990 (≈ 99.999%)
- Вероятность того, что попал второй стрелок ≈ 0.000010 (≈ 0.001%)

Замечание: Несмотря на то, что стрелков выбирали с равной вероятностью, тот факт, что мишень поражена, делает гораздо более вероятным, что стрелял первый стрелок.

Независимость случайных величин

Перед тем как перейти к численным характеристикам случайных величин, определим понятие **независимости**.

Пусть даны две независимые случайные величины X и Y, их распределения заданы следующим образом:

$$
\begin{array}{c|ccc}
X & a_1 & a_2 & a_3 \\
\ldots & & & \\
\hline
\mathbb{P} & p_1 & p_2 & p_3 \\
\ldots & & &
\end{array}
\qquad
\begin{array}{c|ccc}
Y & b_1 & b_2 & b_3 \\
\ldots & & & \\
\hline
\mathbb{P} & q_1 & q_2 & q_3 \\
\ldots & & &
\end{array}
$$

Для произвольных i и j, чему равна вероятность совместного события:

$$\mathbb{P}(X = a_i \mid Y = b_j) = ?$$

Независимость случайных величин

Логично определить независимость случайных величин X и Y

так, чтобы для произвольных i и j :

$$\mathbb{P}(X = a_i \mid Y = b_j) = \mathbb{P}(X = a_i).$$

Общее определение для дискретных и непрерывных случаев

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B).$$

Числовые характеристики случайных величин

1. Математическое ожидание.

Рассмотрим сначала дискретный случай.

Для дискретной случайной величины X с таблицей распределения

X	a_1	a_2	a_3	$\dots$
$\mathbb{P}$	p_1	p_2	p_3	$\dots$

Математическое ожидание (или среднее значение)

определяется как число $\mathbb{E}[X]$, вычисляемое по формуле:

$$\mathbb{E}[X] = \sum_i a_i \cdot \mathbb{P}(X = a_i) = \sum_i a_i p_i.$$

Если ряд $\sum_i |a_i| p_i$ не сходится абсолютно, то есть если $\sum_i |a_i| p_i = \infty$, то говорят, что математическое ожидание не существует.

Числовые характеристики случайных величин

1. Математическое ожидание.

Для непрерывной случайной величины X с плотностью распределения $f(u)$ **математическое ожидание** $\mathbb{E}[X]$ вычисляется по формуле:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} u \cdot f(u)\, du.$$

Числовые характеристики случайных величин

1. Математическое ожидание.

Определения для дискретного и непрерывного случаев согласуются по следующим правилам перехода:

$$\sum_i \leftrightarrow \int_{-\infty}^{+\infty}, \quad a_i \leftrightarrow u, \quad p_i \leftrightarrow f(u)\,du.$$

Соответственно:

$$\sum_i a_i \cdot p_i \leftrightarrow \int_{-\infty}^{+\infty} u \cdot f(u)\,du.$$

1. Математическое ожидание: Смысл

1) Физический смысл: Если распределить единичную массу вдоль невесомого стержня:

- В дискретном случае: в точке a_i масса p_i,
- В непрерывном случае: по плотности $f(u)$,

то точка $\mathbb{E}[X]$ совпадёт с координатой центра тяжести.

2) Вероятностный смысл: При многократном повторении эксперимента:

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mathbb{E}[X], \quad \text{при } n \to \infty.$$

Этот факт называется *Законом больших чисел.*

1. Свойства математического ожидания

Основные свойства математического ожидания:

1. $\mathbb{E}[c] = c$, где $c \in \mathbb{R}$;
2. $\mathbb{E}[cX] = c\mathbb{E}[X]$, где $c \in \mathbb{R}$;
3. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$;
4. $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, если X и Y независимы;
5. Если $X \geq 0$, то $\mathbb{E}[X] \geq 0$.

1. Математическое ожидание: Основные распределения

1. Распределение Бернулли $\mathbb{B}_p$:

- $\mathbb{E}[X] = p$
- Пример: среднее число успехов в одном испытании

2. Биномиальное распределение $\mathbb{B}_{n,p}$:

- $\mathbb{E}[X] = np$
- Пример: среднее число успехов в n испытаниях

3. Распределение Пуассона Π_λ:

- $\mathbb{E}[X] = \lambda$
- Пример: среднее число событий за фиксированный интервал

1.Математическое ожидание: Основные распределения

4. Равномерное распределение на конечном множестве $a_1, \ldots, a_n$:

- $\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} a_i$
- Пример: среднее значение на игральной кости = 3.5

5. Равномерное распределение на отрезке $[a, b]$:

- $\mathbb{E}[X] = \frac{a+b}{2}$
- Пример: среднее значение при равномерном распределении на $[0, 1]$ равно 0.5

6. Экспоненциальное распределение $\mathrm{Exp}(\lambda)$:

- $\mathbb{E}[X] = \frac{1}{\lambda}$
- Пример: среднее время ожидания между событиями

7. Нормальное распределение $\mathcal{N}(a, \sigma^2)$:

- $\mathbb{E}[X] = a$
- Пример: среднее значение роста взрослых мужчин в популяции
- Замечание: параметр a непосредственно задаёт математическое ожидание

2. Дисперсия

Дисперсия случайной величины X определяется как

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Также можно записать:

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

Данное определение применимо как для дискретных, так и для непрерывных распределений. Обратите внимание, что математическое ожидание $\mathbb{E}[X]$ вычисляется по-разному в этих случаях.

2. Дисперсия: Смысл

Смысл дисперсии:

Дисперсия $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ — это среднее значение квадрата отклонений случайной величины *X* от её математического ожидания.

Она характеризует степень разброса значений случайной величины вокруг её среднего значения.

2. Дисперсия: Свойства

Свойства дисперсии:

- (V1) $\mathrm{Var}(c) = 0$ для любого $c \in \mathbb{R}$;
- (V2) $\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$ для любого $c \in \mathbb{R}$;
- (V3) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$, если X и Y независимы;
- (V4) $\mathrm{Var}(X + c) = \mathrm{Var}(X)$ для любого $c \in \mathbb{R}$;
- (V5) $\mathrm{Var}(X) \geq 0$.

2. Дисперсия: Стандартное отклонение

Стандартное отклонение (σ) определяется как квадратный корень из дисперсии:

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Разделив случайную величину *X* на σ, получим нормированную случайную величину с $\mathrm{Var}(X/\sigma) = 1$. Этот процесс называется *нормировкой*.

2. Дисперсия: Основные распределения

1. Распределение Бернулли $\mathbb{B}_p$:

- $\mathrm{Var}(X) = p(1-p)$
- Максимальное значение $\frac{1}{4}$ достигается при $p = \frac{1}{2}$

2. Биномиальное распределение $\mathbb{B}_{n,p}$:

- $\mathrm{Var}(X) = np(1-p)$
- Следует из независимости испытаний Бернулли

3. Распределение Пуассона Π_λ:

- $\mathrm{Var}(X) = \lambda$
- Совпадает с математическим ожиданием

2. Дисперсия: Основные распределения

4. Равномерное распределение на конечном множестве $a_1, \dots, a_n$:

- $\mathrm{Var}(X) = \frac{1}{n} \sum_{i=1}^{n} (a_i - \mathbb{E}[X])^2$
- Для множества $\{1, 2, \dots, n\}$: $\mathrm{Var}(X) = \frac{n^2-1}{12}$

5. Равномерное распределение на отрезке $[a, b]$:

- $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
- Для $[0, 1]$: $\mathrm{Var}(X) = \frac{1}{12}$

6. Экспоненциальное распределение $\mathrm{Exp}(\lambda)$:

- $\mathrm{Var}(X) = \frac{1}{\lambda^2}$
- Квадрат математического ожидания

7. Нормальное распределение $\mathcal{N}(a, \sigma^2)$:

- $\mathrm{Var}(X) = \sigma^2$
- Параметр σ^2 непосредственно задаёт дисперсию

3. Моменты старших порядков

- $\mathbb{E}[X^k]$ — k-й момент X;
- $\mathbb{E}[|X|^k]$ — абсолютный k-й момент X;
- $\mathbb{E}[(X - \mathbb{E}[X])^k]$ — центральный k-й момент X;
- $\mathbb{E}[|X - \mathbb{E}[X]|^k]$ — абсолютный центральный k-й момент X.

Эти моменты характеризуют распределение случайной величины и часто используются в задачах концентрации.

4. Коэффициент асимметрии (Skewness)

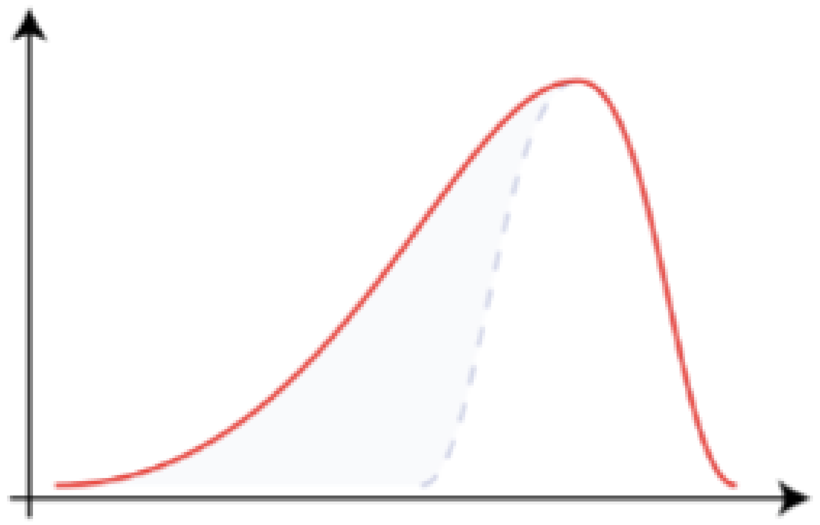
Коэффициент асимметрии γ_1 описывает степень асимметрии распределения:

$$\gamma_1 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^3]}{(\mathrm{Var}(X))^{3/2}}.$$

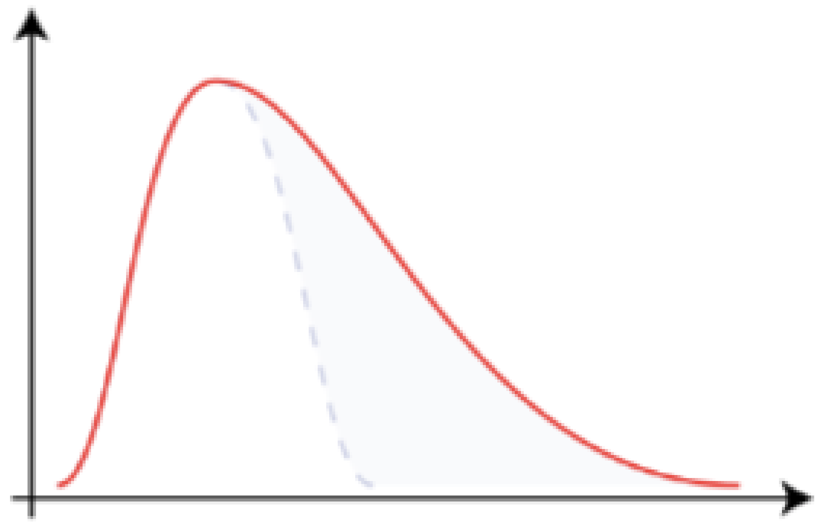
γ_1 — нормированный центральный момент третьего порядка.

Примеры:

- *Negative Skew:* длинный левый хвост.
- *Positive Skew:* длинный правый хвост.



Negative Skew



Positive Skew

4. Коэффициент эксцесса (Kurtosis)

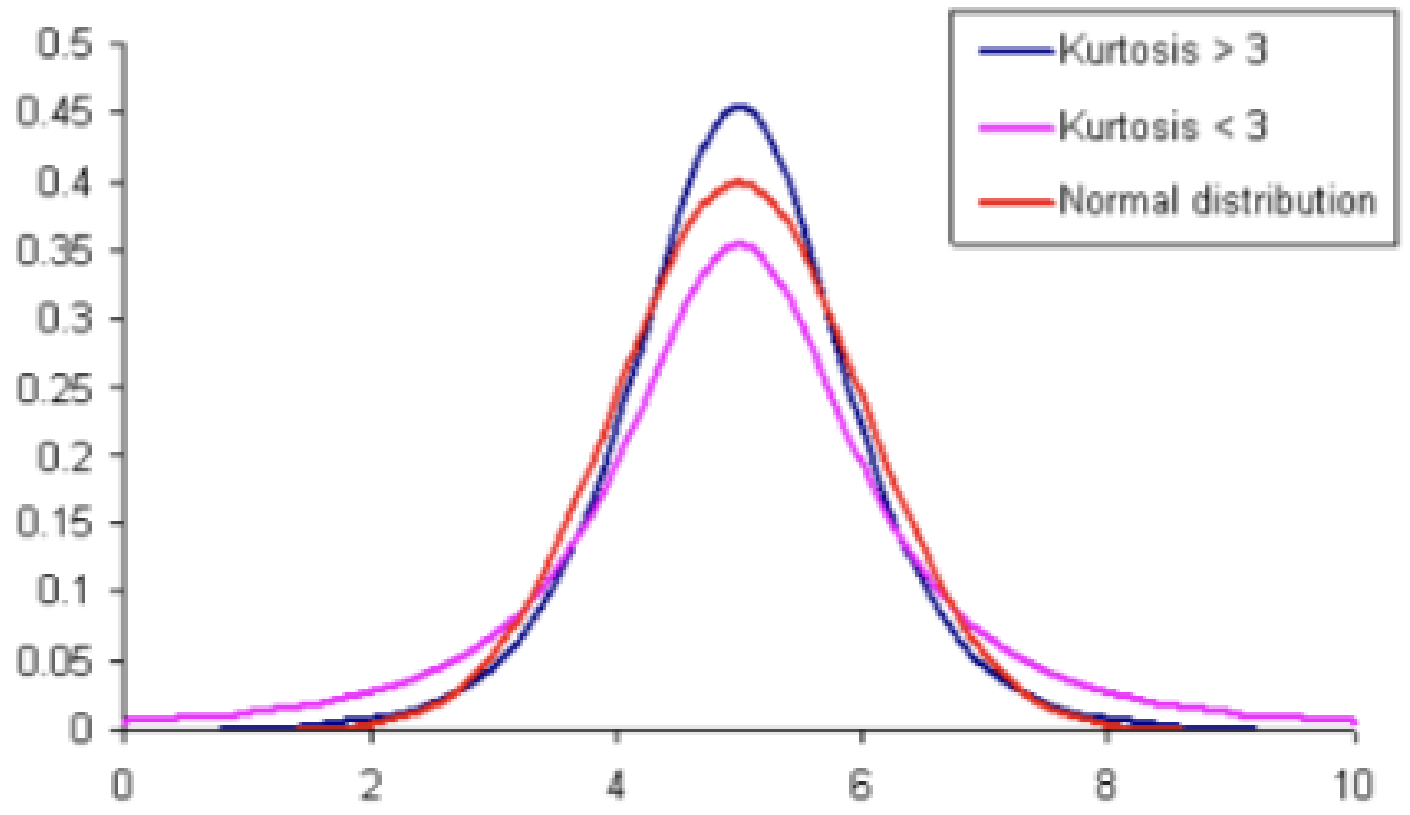
Коэффициент эксцесса γ_2 измеряет остроту пика распределения:

$$\gamma_2 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{(\mathrm{Var}(X))^2}.$$

γ_2 – нормированный центральный момент четвёртого порядка.

Интерпретация:

- $\gamma_2 > 3$: острое распределение (*leptokurtic*).
- $\gamma_2 = 3$: нормальное распределение (*mesokurtic*).
- $\gamma_2 < 3$: плоское распределение (*platykurtic*).



Метод максимального правдоподобия

Постановка задачи:

Дана выборка из распределения с известной формой, но неизвестными параметрами.

Вопрос:

Как оценить параметры распределения на основе имеющихся данных?

Существует множество методов оценки параметров: метод моментов, метод максимального правдоподобия, метод спейсингов и др.

Мы рассмотрим наиболее популярный и интуитивно понятный метод — **метод максимального правдоподобия**.

Метод максимального правдоподобия

В теории оценивания неизвестные параметры обозначаются как

$$\theta_1, \theta_2, \ldots, \theta_d.$$

Для упрощения будем считать параметр многомерным:

$$\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \Theta \subseteq \mathbb{R}^d.$$

Основная идея метода построения оценок

Идея: Для оценки *d* неизвестных параметров модели составляют *d* уравнений.

Метод максимального правдоподобия: Найти максимум функции правдоподобия, приравняв частные производные по *d* параметрам к нулю.

Метод максимального правдоподобия

Дана выборка $x_1, \ldots, x_n$ из распределения с неизвестным параметром θ.

Обозначения:

- $\mathbb{P}_\theta(X = u)$ — вероятность для дискретного случая;
- $f_\theta(u)$ — плотность распределения для непрерывного случая.

Метод максимального правдоподобия

Введем величину:

- Вероятность или плотность:

$$p(u, \theta) = \begin{cases} \mathbb{P}_\theta(X = u), & \text{дискретный случай;} \\ f_\theta(u), & \text{непрерывный случай.} \end{cases}$$

- **Функция правдоподобия:**

$$L(\theta) = p(x_1, \theta) \cdot \ldots \cdot p(x_n, \theta).$$

- В дискретном случае $L(\theta)$ — вероятность наблюдения выборки $x_1, \ldots, x_n$.
- В общем случае $L(\theta)$ описывает вероятность или плотность выборки при параметре θ.

Метод максимального правдоподобия: Основная идея

Представляется разумным в качестве оценки параметра θ взять наиболее правдоподобное значение, которое получается при максимизации функции $L(\theta)$.

Замечание 1: Часто проще искать точку максимума функции $\ln L(\theta)$, которая совпадает с максимумом $L(\theta)$ в силу монотонности логарифма.

Замечание 2: В случае, если функция $L(\theta)$ не является непрерывно дифференцируемой, необходимо дополнительно анализировать окрестности точек разрыва.

Метод максимального правдоподобия: Основная идея

Алгоритм:

1. Составить функцию правдоподобия $L(\theta)$
2. Взять логарифм: $\ln L(\theta)$
3. Найти производную и приравнять к нулю
4. Проверить, что найденная точка действительно является максимумом
5. Проверить точки разрыва, если они есть

Преимущества логарифмирования:

- Произведение переходит в сумму
- Числа становятся более удобными для вычислений
- Экспоненты переходят в линейные выражения

Пример: Оценка параметра распределения Бернулли методом максимального правдоподобия

Рассмотрим практический пример применения метода максимального правдоподобия для оценки параметра p в распределении Бернулли.

Задача: Имеется выборка из n независимых испытаний Бернулли $x_1, x_2, \ldots, x_n$, где $x_i \in \{0, 1\}$. Требуется оценить неизвестный параметр p (вероятность успеха).

Решение:

1. **Составим функцию правдоподобия:**

   Для распределения Бернулли вероятность получить значение x равна: $\mathbb{P}_p(X = x) = p^x (1-p)^{1-x}$, где $x \in \{0, 1\}$

   Функция правдоподобия для выборки: $L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$

2. **Логарифмируем функцию правдоподобия:**

   $\ln L(p) = \sum_{i=1}^{n} x_i \ln p + (n - \sum_{i=1}^{n} x_i) \ln(1-p)$

3. **Найдем производную и приравняем к нулю:**

   $\frac{d}{dp} \ln L(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0$

Пример: Оценка параметра распределения Бернулли методом максимального правдоподобия

4. **Решим уравнение:**

   $\frac{\sum_{i=1}^{n} x_i}{p} = \frac{n-\sum_{i=1}^{n} x_i}{1-p}$

   $\sum_{i=1}^{n} x_i(1-p) = (n - \sum_{i=1}^{n} x_i)p$

   $\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i = np - p\sum_{i=1}^{n} x_i$

   $\sum_{i=1}^{n} x_i = np$

   $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$

5. **Проверим, что это максимум:**

   Вторая производная: $\frac{d^2}{dp^2} \ln L(p) = -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{n-\sum_{i=1}^{n} x_i}{(1-p)^2} < 0$

   Так как вторая производная отрицательна, точка $\hat{p}$ является точкой максимума.

Пример: Оценка параметра распределения Бернулли методом максимального правдоподобия

Вывод: Оценка максимального правдоподобия для параметра p распределения Бернулли равна $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$, что соответствует доле успехов в выборке.

Интерпретация: Если мы провели n испытаний Бернулли и получили k успехов, то наилучшая оценка вероятности успеха равна $\hat{p} = \frac{k}{n}$.

Пример: Монета подбрасывается 100 раз, и орел выпадает 45 раз. Оценка максимального правдоподобия для вероятности выпадения орла равна $\hat{p} = \frac{45}{100} = 0.45$.