

## Введение. Основные понятия теории вероятностей

В теории вероятностей всё начинается со случайного эксперимента.

Под **случайным экспериментом** понимается такой эксперимент, результат которого не детерминирован изначально.

**Элементарный исход** — это любой возможный исход случайного эксперимента.

Всё множество элементарных исходов носит название **пространством элементарных исходов** и обычно обозначается буквой  $\Omega$  (омега).

## Классическое определение вероятности. Правила сложения и умножения

**Вероятностью случайного события A** называется отношение числа  $n$  равновероятных элементарных исходов, составляющих событие A, к числу всех возможных элементарных исходов  $N$ :

$$P(A) = \frac{n}{N}$$

Элементарные исходы, составляющие событие A, также очень часто называют исходами, **благоприятными** или **благоприятствующими** для события A.

Можно также сказать, что вероятность — это число, которое оценивает степень возможности наступления того или иного случайного события. Вероятности всегда находятся в диапазоне от 0 до 1 включительно:

$$0 \leq P(A) \leq 1$$

Чем больше полученное значение, тем более вероятно, что событие произойдёт. Вероятность 0 означает, что событие никогда не случится. Такое событие называют **невозможным**:

$$P(A) = P(\emptyset) = \frac{n}{N} = \frac{0}{N} = 0$$

Вероятность 1 означает, что событие произойдёт в любом случае — такое событие называют **достоверным**:

$$P(A) = P(\Omega) = \frac{n}{N} = \frac{N}{N} = 1$$

Все остальные значения от 0 до 1 представляют различные уровни вероятности.

**Дополнение события** — это подмножество таких исходов во всём пространстве исходов, что они не благоприятствуют событию  $A$ . Дополнение события  $A$  само по себе тоже является событием и обозначается как  $\bar{A}$ .

Важно понимать, что **у события и дополнения к нему нет общих исходов**, то есть они взаимоисключающие или, как это обычно называют в теории вероятностей, **несовместные**. Также событие и дополнение к нему содержат в сумме абсолютно все исходы из пространства исходов. Из этого следует, что сумма их вероятностей равняется одному:

$$P(A) + P(\bar{A}) = 1$$

Если у нас есть любое количество взаимоисключающих событий, которые описывают абсолютно все возможные элементарные исходы, то сумма их вероятностей равна 1:

$$P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$$

### Правило суммы

Если события  $A$  и  $B$  являются несовместными, то вероятность для объединения этих событий вычисляется по следующей формуле:

$$P(A \cup B) = P(A) + P(B)$$

### Обобщённое правило суммы

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

### Правило умножения

Правило произведения используется для нахождения вероятности пересечения событий:

$$P(A \cap B) = P(A) \times P(B)$$

## Условная вероятность

**Условная вероятность** события В при условии А определяется как вероятность того, что событие В произойдёт после того, как событие А уже произошло, и обозначается следующим образом:

$$P(B | A)$$

## Независимость событий

События А и В называются **независимыми**, если вероятность их пересечения равна произведению вероятностей

$$P(A \cap B) = P(A) \cdot P(B)$$

## Полная вероятность

**Разбиение вероятностного пространства** — это взаимоисключающие и совместно исчерпывающие события. Проще говоря, это события, которые не пересекаются (т. е. не имеют общих исходов), но в объединении дают все возможные исходы.

**Формула полной вероятности:**

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i)$$

В данной формуле:

- $P(B)$  — вероятность наступления события В;
- $P(A_i)$  — вероятность наступления события  $A_i$ , которое является условием для события В;
- $P(B | A_i)$  — условная вероятность наступления события В, если известно, что произошло событие  $A_i$ .



Предположим, что два производителя, **A** и **B**, поставляют двигатели для гоночных автомобилей *Formula 1* со следующими характеристиками:

- 99 % двигателей завода **A** служат более 5 000 км;
- завод **B** производит двигатели, которые прослужат более 5 000 км с вероятностью 95 %;
- 70 % двигателей произведены заводом **A**, а остальные — заводом **B**.

Найдите вероятность того, что двигатель прослужит более 5 000 км.

- Вероятность, что двигатель с завода **A** и прослужит более 5000 км, равна  $0.7 \cdot 0.99 = 0.693$
- Вероятность, что двигатель с завода **B** и прослужит более 5000 км, равна  $0.3 \cdot 0.95 = 0.285$

Итоговая вероятность:  $0.693 + 0.285 = 0.978$ .

Получается, что случайно выбранный двигатель прослужит более 5000 км с вероятностью 0.978.

## Теорема Байеса

Теорема/формула Байеса:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$



1 % клиентов постоянно жалуется на сервис — из них 60 % уйдут. Ещё 10 % клиентов периодически жалуются на сервис — из них 10 % уйдут. Оставшиеся 89 % клиентов никогда не жалуются на сервис — из них 3 % уйдут.

Найдите долю довольных клиентов среди ушедших.

Мы, зная информацию об ушедших клиентах среди довольных, можем выразить долю довольных клиентов среди ушедших, то есть получить одну условную вероятность через другую.

Для начала найдём общую долю оттока клиентов. В целом, мы уже делали это в предыдущем юните, так что просто повторим уже известные вам вычисления:

$$\begin{aligned} P(C) &= P(C | CUC) \cdot P(CUC) + P(C | DC) \cdot P(DC) + P(C | CSC) \cdot P(CSC) = \\ &= 0.01 \cdot 0.6 + 0.1 \cdot 0.1 + 0.89 \cdot 0.03 = 0.0427 \end{aligned}$$

Если мы говорим про отток довольных клиентов, то в общей формуле можно найти слагаемое, в котором вычисляется именно доля оттока довольных клиентов:

$$P(C) = P(C | CUC) \cdot P(CUC) + P(C | DC) \cdot P(DC) + P(C | CSC) \cdot P(CSC)$$

Тогда получаем, что доля довольных клиентов среди всех ушедших равна отношению доли ушедших и довольных клиентов к доле всех ушедших:

$$P(CSC | C) = \frac{P(C | CSC)P(CSC)}{P(C)} = \frac{0.03 \cdot 0.89}{0.0427} \approx 0.63$$

Под **априорными вероятностями** понимаются безусловные вероятности, то есть они фиксированы и не зависят от вероятностей наступления других событий.

**Апостериорные вероятности**, напротив, обусловлены вероятностями наступления каких-то ещё случайных событий.

## Наивный байесовский классификатор

**Наивный байесовский классификатор (НБК, англ. Naive Bayes Classifier, NBC)** решает задачу классификации объектов по типам. Большим преимуществом этого алгоритма является его простота, как идейная, так и алгоритмическая.

Наивная байесовская классификация — это достаточно простой вероятностный алгоритм, основанный на том, что все признаки модели независимы.

Если, например, мы говорим про задачу классификации спам-писем и обычных писем, то в этом контексте мы считаем, что каждое слово в

сообщении не зависит от всех других слов, то есть каждое слово мы учитываем, не обращая внимания на контекст.

Возьмём формулу Байеса и применим её к нашей задаче:

$$P(\text{Спам} \mid w_1, w_2, \dots, w_n) \propto P(\text{Спам}) \cdot \prod_{i=1}^n P(w_i \mid \text{Спам})$$

Вероятность того, что письмо является спамом при условии, что в нём есть определённые слова (которые мы обозначили), пропорциональна произведению двух значений:

- вероятности получения спама в целом (по сути, это доля спама в выборке);
- произведения вероятностей, что в письме есть некоторое слово  $w_i$ , если письмо является спамом, для всех слов выборки.

Давайте разберёмся с этим подробнее. Для каждого слова в сообщении мы рассчитываем вероятность того, что это слово окажется в спаме. В рамках нашей задачи рассматриваем следующие значения:

- $P(spam)$  — вероятность, что случайно взятое письмо будет спамом (также это доля спам-сообщений в нашем наборе данных);
- $P(w_i \mid spam)$  — вероятность того, что в сообщении будет определённое слово, если это письмо является спамом.

По той же логике можем определить:

- $P(not\ spam)$  — доля сообщений, которые не являются спамом;
- $P(w_i \mid not\ spam)$  — вероятность того, что в сообщении будет определённое слово, если это письмо не является спамом.

Теперь необходимо понять, как рассчитать вероятности каждого слова. Для этого в алгоритме используется следующая формула:

$$P(w_i \mid Spam) = \frac{N_{w_i|Spam} + \alpha}{N_{Spam} + \alpha \cdot N_{Vocabulary}}$$

В этой формуле:

- $N_{Vocabulary}$  — количество уникальных слов во всём наборе данных;
- $N_{Spam}$  — общее количество слов в спам-сообщениях;
- $N_{w_i | Spam}$  — количество повторов слова во всех спам-сообщениях;
- $\alpha$  — коэффициент для случаев, когда слово в сообщении отсутствует в нашем наборе данных.

Кратко это можно объяснить так: вероятность того, что это слово встретится в спам сообщении, — это частота этого слова в «спамовой части» нашего набора данных (но с добавлением «сглаживания»), чтобы учитывать ситуации, когда попадают слова, которых не было в обучающей выборке).

## Случайная величина и её характеристики

Случайные величины удобно вводить для численного описания результатов случайного эксперимента и случайных событий, например для описания количества продаж в магазине на этой неделе или количества новых клиентов в компании за предыдущий год.

Случайные величины бывают двух видов — **дискретные** и **непрерывные**.

Величина является **дискретной**, если может принимать значения только из конечного или счётного (бесконечного) множества.

Например, если мы описываем количество клиентов, то это, разумеется, дискретная величина, так как оно может быть выражено только целым неотрицательным числом (не может быть 1.5 клиента или 0.2 клиента).

Если величина может принимать любое значение, то она является **непрерывной**.

К примеру, человеческий рост — это непрерывная величина, так как может принимать бесконечно много значений: 170 см, 170 см и 1 см, 170 см 1 см и 0001 мм и так далее.

Для того чтобы задать дискретную случайную величину, необходимо задать список её значений и список вероятностей каждого из этих значений.

Предположим, что нам необходимо задать дискретную случайную величину, равную количеству выпавших орлов при подбрасывании одной или двух идеальных монеток.

X	0	1
Исходы	P	O
P	0.5	0.5

**Математическое ожидание случайной величины** — это взвешенное среднее значений этой величины с учётом вероятностей возможных значений. Чаще всего оно обозначается буквой E (от англ. Expectation), ещё одно общепринятое обозначение —  $\mu$  (греческая буква «мю»). В русскоязычной литературе принято обозначать математическое ожидание буквой M.

Оно вычисляется по следующей формуле:

$$E(X) = \mu_X = \sum x_i \cdot P(X = x_i)$$

Здесь:

- $i$  — индекс переменной;
- $x_i$  — значение случайной величины с индексом  $i$ ;
- $P(X = x_i)$  — вероятность того, что случайная величина принимает значение  $x_i$ .

Для того чтобы численно оценить разброс, используют **дисперсию**. Она вычисляется по следующей формуле:

$$V(X) = D(X) = \sigma_X^2 = E(X^2) - (E(X))^2$$

Дисперсию принято обозначать буквой V (от англ. Variance) или буквой D. Иногда встречается обозначение  $\sigma_X^2$ .



## Дискретные распределения

**Равномерное распределение** — это распределение вероятностей, в качестве значений которого могут выступать любые целые числа от 1 до N, а вероятности их появления одинаковы.

Равномерное распределение применяется там, где нужны равновероятные модели.

Например, равномерное распределение пригодится, если для розыгрыша призов необходимо реализовать генератор случайных чисел, который будет с одинаковыми вероятностями выбирать натуральные числа в заданном интервале.

Для равномерного распределения уже известны математическое ожидание

$$EX = \frac{N + 1}{2}$$

и стандартное отклонение

$$\sigma_X = \sqrt{\frac{N^2 - 1}{12}}$$

### Распределение Бернулли

**Распределение Бернулли**, по сути, моделирует однократное подбрасывание «фальшивой» монеты. Это распределение вероятностей случайной величины, принимающей только два значения: 1 («успех») и 0 («неудача») с вероятностями  $p$  и  $1 - p$  соответственно.

Основной параметр, определяющий это распределение, —  $p$  (вероятность «успеха»).

Распределение Бернулли формально описывается следующим образом:

$$P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

Для этого распределения также известны математическое ожидание

$$EX = 1 \cdot p + 0 \cdot (1 - p) = p$$

и стандартное отклонение

$$\sigma_X = \sqrt{p(1 - p)}$$

## Биномиальное распределение

**Биномиальным** называют распределение, при котором возможны только два исхода (успех или неудача, выигрыш или проигрыш) и вероятность успеха и неудачи одинакова для всех испытаний. Однако исходы не обязательно должны быть равновероятными, и каждое испытание не зависит от других.

Параметры биномиального распределения —  $n$  и  $p$ , где  $n$  — общее количество испытаний, а  $p$  — вероятность успеха в каждом испытании.

Для того чтобы оценить вероятность, что среди  $n$  успехов будет  $k$  испытаний, используют следующую формулу:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\binom{n}{k}$$

называют **биномиальным коэффициентом**, и он вычисляется следующим образом:

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

Математическое ожидание биномиального распределения равно:

$$EX = np$$

Стандартное отклонение:

$$\sigma_X = \sqrt{np(1 - p)}$$

## Распределение Пуассона

**Распределение Пуассона** — это дискретное распределение вероятностей числа событий, происходящих в данный период времени, с учётом среднего количества раз, когда событие происходит за этот период времени.

В распределении Пуассона значение случайной величины может быть любым неотрицательным числом. Случайная величина будет обладать следующими характеристиками:

$$EX = \lambda$$

$$\sigma_X = \sqrt{\lambda},$$

где  $\lambda$  — ожидаемое число событий за период времени.

Чтобы рассчитать вероятность того, что за период времени произойдёт  $k$  событий, можно пользоваться следующей формулой:

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

## Непрерывные распределения

В непрерывной модели распределение вероятностей задаётся не таблицей, а с помощью функции плотности вероятности.

Плотностью может быть любая неотрицательная функция, площадь которой под графиком равна 1 (так как сумма вероятностей всех исходов равна 1) и которая удовлетворяет некоторым условиям.

### Равномерное распределение

Непрерывная случайная величина  $X$  распределяется равномерно, если вероятность попадания  $X$  в интервал пропорциональна длине этого интервала.

Случайная величина, которая распределена по непрерывному равномерному закону, может принимать в качестве значения любое вещественное число в

заданном отрезке от  $a$  до  $b$  и имеет следующее математическое ожидание и дисперсию:

$$EX = \frac{a + b}{2}$$

$$\sigma_X = \frac{b - a}{\sqrt{12}}$$

Плотность распределения постоянна и имеет следующий вид:

$$f(x) = \begin{cases} 0, & \text{при } x < a, x > b \\ \frac{1}{b - a}, & \text{при } a \leq x \leq b \end{cases}$$

## Нормальное распределение

**Нормальное распределение** (его ещё называют **гауссовским** или **распределением Гаусса**) представляет собой распределение вероятностей, обычно используемое для моделирования таких явлений, как физические характеристики (например, рост, вес и т. д.) или результаты тестирований.

Обычно гауссовское распределение определяется через математическое ожидание и стандартное отклонение — это записывается следующим образом:

$$X \sim \text{Normal}(\mu, \sigma^2)$$

Плотность нормального распределения задаётся формулой:

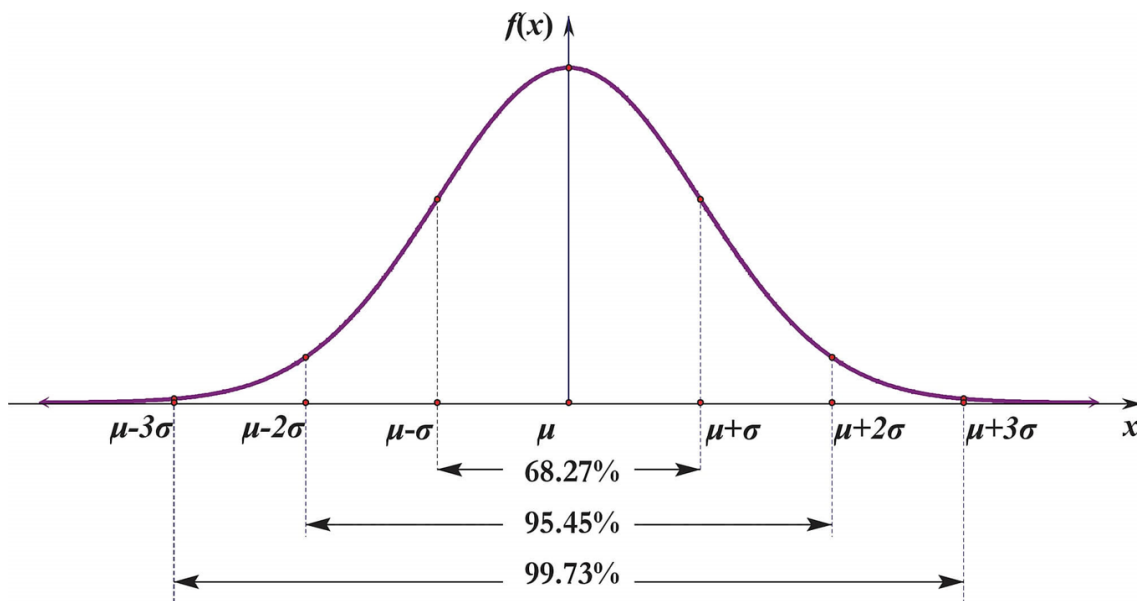
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Существует **эмпирическое правило трёх сигм**, которое гласит:

- Если отступить от среднего на одно стандартное отклонение в меньшую и большую сторону, то в этих пределах будет 68 % данных.

- Если отступить от среднего на два стандартных отклонения в меньшую и большую сторону, то в этих пределах будет 95 % данных.
- Если отступить от среднего на три стандартных отклонения в меньшую и большую сторону, то в этих пределах будет 99.7% данных.

Эта идея также проиллюстрирована ниже:



## Экспоненциальное распределение

**Экспоненциальное распределение** — это распределение вероятностей времени между событиями в процессе Пуассона.

Плотность экспоненциального распределения задаётся следующей формулой:

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

В качестве параметра  $\lambda$  здесь берётся ожидаемое число событий за единицу времени.

Математическое ожидание для этого распределения вычисляется по следующей формуле:

$$EX = 1/\lambda$$

Стандартное отклонение можно найти ровно так же:

$$\sigma_X = 1/\lambda$$

Вероятность того, что значение случайной величины будет не больше X, равна:

$$1 - e^{-\lambda x}$$