# Урок 7. Условная вероятность и формула Байеса

**Вероятность и парадоксы · ~35 минут**

Тебе сделали тест на редкую болезнь. Тест точный — ошибается всего в 1% случаев. Результат положительный. Насколько ты болен(а)? Интуиция говорит: «на 99%!». А правильный ответ — меньше 10%. Это, пожалуй, самый важный парадокс курса: его непонимание приводит к неверным медицинским решениям в реальной жизни. Разберёмся.

## 🎯 Что ты узнаешь

- Что такое условная вероятность `P(A|B)`.
- Как работает формула Байеса «по-простому».
- Почему точный тест на редкую болезнь даёт много ложных тревог.

## 📖 Разбираемся в теме

### Условная вероятность

`P(A|B)` — вероятность события `A` **при условии, что** `B` уже произошло. Читается «вероятность A при условии B».

Пример: бросили кубик. `P(выпало 2) = 1/6`. Но если известно, что **выпало чётное** (событие B = {2,4,6}), то

$$P(2 \mid \text{чёт}) = \frac{1}{3}.$$

Информация «выпало чётное» сузила мир до трёх исходов, и вероятность двойки выросла. Формула:

$$P(A \mid B) = \frac{P(A \text{ и } B)}{P(B)}.$$

### Идея формулы Байеса

Часто мы знаем `P(B|A)`, а хотим найти `P(A|B)` — «перевернуть» условие. Например, врач знает `P(тест «+» | болен)` (точность теста), а пациенту важно `P(болен | тест «+»)`. Это разные вещи!

**Метод естественных частот** — самый понятный способ. Вместо дробей возьмём конкретную большую группу людей и просто посчитаем их «по головам». Это и есть Байес, только наглядно.

> 📌 **Запомни:** `P(тест «+» | болен)` и `P(болен | тест «+»)` — это НЕ одно и то же. Путать их — классическая и опасная ошибка.

### Разбор парадокса: тест на редкую болезнь

Пусть:

- болезнью страдает **1 человек из 1000** (0,1%);
- тест **всегда** выявляет больного (нет ложноотрицательных);
- тест ошибается на здоровых в **1%** случаев (ложноположительные).

Тест точный на 99% — казалось бы, положительный результат почти наверняка означает болезнь. Проверим методом естественных частот. Возьмём **100 000** человек:

- Больных: `100000 · 0,001 = 100` человек. Все получат «+»: **100 истинно положительных**.
- Здоровых: `100000 − 100 = 99900`. Из них 1% ошибочно получат «+»: `99900 · 0,01 = 999` — **999 ложноположительных**.

Всего положительных результатов: `100 + 999 = 1099`. Из них реально больны только 100. Значит

$$P(\text{болен} \mid \text{тест «+»}) = \frac{100}{1099} \approx 0{,}091 \approx 9{,}1\%.$$

Меньше 10%! Хотя тест «точный». В чём фокус? Болезнь **очень редкая**, поэтому здоровых людей огромное большинство, и даже маленький процент их ошибок

(1%) даёт больше ложных «+», чем настоящих больных.

> 💡 Именно поэтому при положительном тесте на редкое заболевание назначают **повторный, другой** тест. Второй положительный результат резко повышает вероятность — база «подозреваемых» теперь уже не всё население, а те самые 1099 человек.

> ⚠️ Ключевой урок: результат теста надо всегда рассматривать вместе с тем, **насколько болезнь редка** (её «базовой частотой»). Игнорировать базовую частоту — это ошибка, которую совершают даже врачи.

### Формула Байеса (для любознательных)

Тот же ответ можно получить формулой:

$$P(A \mid B) = \frac{P(B \mid A),P(A)}{P(B)}.$$

Здесь `A` = «болен», `B` = «тест положителен». Подставим: `P(A) = 0,001`, `P(B|A) = 1`, а `P(B) = 1·0,001 + 0,01·0,999 = 0,01099`. Тогда

$$P(A|B) = \frac{1 \cdot 0{,}001}{0{,}01099} \approx 0{,}091.$$

Тот же самый результат — 9,1%. Метод естественных частот и формула Байеса — это одно и то же, просто записанное по-разному.

## ✍️ Разбор примера

**Задача.** В городе 90% такси жёлтые и 10% синие. Свидетель ДТП говорит, что такси было синим. Известно, что свидетель различает цвета правильно в 80% случаев. Какова вероятность, что такси действительно было синим?

Возьмём 1000 такси:

- Синих: `1000 · 0,1 = 100`. Свидетель правильно назовёт синими 80%: `100 · 0,8 = 80` скажет «синее».

- Жёлтых: `900`. Ошибётся в 20% и назовёт синими: `900 · 0,2 = 180` скажет «синее».

Всего «сказал синее»: `80 + 180 = 260`. Из них реально синих — 80.

$$P(\text{синее} \mid \text{«синее»}) = \frac{80}{260} \approx 0{,}308 \approx 30{,}8\%.$$

Хотя свидетель надёжен на 80%, вероятность, что такси было синим, — всего около 31%! Причина та же: синих такси изначально мало.

**Ответ:** ≈ 30,8%.

## 📝 Задачи

1. Бросили кубик. Известно, что выпало число больше 3. Какова вероятность, что выпало 6?
2. В колоде 36 карт вытянули карту, и известно, что она красной масти (18 красных). Какова вероятность, что это дама червей?
3. Болезнь у 1 человека из 200. Тест выявляет всех больных, но ошибается на 2% здоровых (ложные «+»). При положительном тесте — какова вероятность болезни? (Возьми группу 20000 человек.)
4. Почему `P(тест «+» | болен)` и `P(болен | тест «+»)` — разные величины? Приведи объяснение своими словами.
5. В задаче про такси пусть свидетель надёжен на 90% (вместо 80%), синих такси по-прежнему 10%. Какова вероятность, что такси было синим? (Группа 1000 такси.)
6. Почему при положительном результате теста на редкую болезнь врачи назначают повторный тест? Объясни через изменение «базовой частоты».
7. На фабрике два станка: первый делает 60% деталей с браком 2%, второй — 40% деталей с браком 5%. Наугад взяли деталь, она бракованная. Какова вероятность, что её сделал **второй** станок? (Возьми 1000 деталей.)