

БАЙЕСОВСКИЙ ВЫВОД ДЛЯ МОНЕТКИ

Сергей Николенко

СПбГУ — Санкт-Петербург

11 сентября 2026 г.

Random facts:



- 11 сентября 1943 г. на стекольном заводе в Гусь-Хрустальном был выпущен первый советский гранёный стакан; форму, по легенде, придумала Вера Мухина
- 11 сентября 1962 г. The Beatles закончили запись первого сингла «P.S. I Love You»; на барабанах был сессионный музыкант Энди Уайт, а Ринго Старру дали марacas
- 11 сентября 1973 г. в результате переворота в Чили погиб президент Альенде; вместе с ним закончился и «Киберсин» — проект централизованного управления экономикой, который строил там британский кибернетик Стаффорд Бир
- 11 сентября — дата невесёлая: ураган на Ньюфаундленде в 1847 г., землетрясение в Крыму в 1927 г., пожары в Пакистане в 2012 г. и, конечно, 2001 год
- 11 сентября — Национальный день Каталонии, учреждённый в память о взятии Барселоны во время войны за испанское наследство в 1713 г.
- 11 сентября — важная дата для флота Петра I: в 1714 г. в этот день произошло Гангутское сражение, а в 1720 г. — сражение при Гренгаме

БАЙЕСОВСКИЙ ПОДХОД

- Нам не понадобятся математические определения сигма-алгебры, вероятностной меры, борелевских множеств и т.п.
- Достаточно понимать, что бывают дискретные случайные величины (неотрицательные вероятности исходов в сумме дают единицу) и непрерывные случайные величины (интеграл неотрицательной функции плотности равен единице).

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

- Совместная вероятность – вероятность одновременного наступления двух событий, $p(x, y)$; маргинализация:

$$p(x) = \sum_y p(x, y).$$

- Условная вероятность – вероятность наступления одного события, если известно, что произошло другое, $p(x | y)$:

$$p(x, y) = p(x | y)p(y) = p(y | x)p(x).$$

- Теорема Байеса – из предыдущей формулы:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_{y'} p(x|y')p(y')}.$$

- Независимость: x и y независимы, если

$$p(x, y) = p(x)p(y).$$

ПРЯМЫЕ И ОБРАТНЫЕ ЗАДАЧИ

- Прямая задача: в урне лежат 10 шаров, из них 3 чёрных. Какова вероятность выбрать чёрный шар?
- Или: в урне лежат 10 шаров с номерами от 1 до 10. Какова вероятность того, что номера трёх последовательно выбранных шаров дадут в сумме 12?
- Обратная задача: перед нами две урны, в каждой по 10 шаров, но в одной 3 чёрных, а в другой — 6. Кто-то взял из какой-то урны шар, и он оказался чёрным. Насколько вероятно, что он брал шар из первой урны?
- Заметьте, что в обратной задаче вероятности сразу стали байесовскими (хоть здесь и можно переформулировать через частоты).

- Иначе говоря, прямые задачи теории вероятностей описывают некий вероятностный процесс или модель и просят подсчитать ту или иную вероятность (т.е. фактически по модели предсказать поведение).
- Обратные задачи содержат *скрытые переменные* (в примере — номер урны, из которой брали шар). Они часто просят по известному поведению построить вероятностную модель.
- Задачи машинного обучения обычно являются задачами второй категории.

- Обычно в классической теории вероятностей, происходящей из физики, вероятность понимается как предел отношения количества определённого результата эксперимента к общему количеству экспериментов.
- Стандартный пример: бросание монетки.

- Мы можем рассуждать о том, «насколько вероятно» то, что
 - сборная России победит на чемпионате мира по футболу в 2026 году;
 - «Одиссею» написала женщина;
 - Александр Фёдорович Керенский бежал за границу в женском платье;
 - ...
- Но о «стремящемся к бесконечности количестве экспериментов» говорить бессмысленно — эксперимент здесь ровно один.

- Здесь вероятности уже выступают как *степени доверия* (degrees of belief). Это байесовский подход к вероятностям (Томас Байес так понимал).
- К счастью, и те, и другие вероятности подчиняются одним и тем же законам; есть результаты о том, что вполне естественные аксиомы вероятностной логики тут же приводят к весьма узкому классу функций (Cox, 19).

- Запишем теорему Байеса:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}.$$

- Здесь $p(\theta)$ — априорная вероятность (prior probability), $p(D|\theta)$ — правдоподобие (likelihood), $p(\theta|D)$ — апостериорная вероятность (posterior probability), $p(D) = \int p(D | \theta)p(\theta)d\theta$ — вероятность данных (evidence).



- В статистике обычно ищут *гипотезу максимального правдоподобия* (maximum likelihood):

$$\theta_{ML} = \arg \max_{\theta} p(D | \theta).$$

- В байесовском подходе ищут *апостериорное распределение* (posterior)

$$p(\theta|D) \propto p(D|\theta)p(\theta)$$

и, возможно, *максимальную апостериорную гипотезу* (maximum a posteriori):

$$\theta_{MAP} = \arg \max_{\theta} p(\theta | D) = \arg \max_{\theta} p(D | \theta)p(\theta).$$

ФОРМУЛА БАЙЕСА НА ПРИМЕРАХ

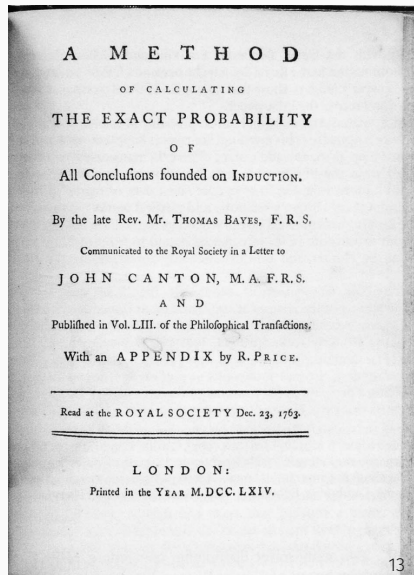
- Приведём классический пример из классической области применения статистики — медицины.
- Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95% (т.е. 5% — вероятность как позитивной, так и негативной ошибки).
- Всего болезнь имеется у 1% респондентов (отложим на время то, что они разного возраста и профессий).
- Пусть некий человек получил позитивный результат теста (тест говорит, что он болен). С какой вероятностью он действительно болен?

- Приведём классический пример из классической области применения статистики — медицины.
- Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95% (т.е. 5% — вероятность как позитивной, так и негативной ошибки).
- Всего болезнь имеется у 1% респондентов (отложим на время то, что они разного возраста и профессий).
- Пусть некий человек получил позитивный результат теста (тест говорит, что он болен). С какой вероятностью он действительно болен?
- Ответ: 16%.

- Обозначим через t результат теста, через d — наличие болезни.
- $p(t = 1) = p(t = 1|d = 1)p(d = 1) + p(t = 1|d = 0)p(d = 0)$.
- Используем теорему Байеса:

$$\begin{aligned} p(d = 1|t = 1) &= \\ &= \frac{p(t = 1|d = 1)p(d = 1)}{p(t = 1|d = 1)p(d = 1) + p(t = 1|d = 0)p(d = 0)} = \\ &= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16. \end{aligned}$$

- Вот такие задачи составляют суть вероятностного вывода (probabilistic inference).
- Поскольку они обычно основаны на теореме Байеса, вывод часто называют байесовским (Bayesian inference).
- Но не только поэтому.

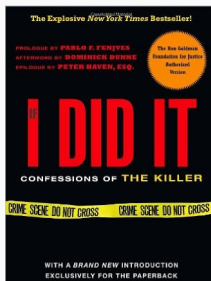


1. У моего знакомого два ребёнка. Будем предполагать, что пол ребёнка выбирается независимо и равновероятно, с вероятностью $\frac{1}{2}$. Две постановки вопроса:
 - (1) я спросил, есть ли у него мальчики, и он ответил «да»; какова вероятность того, что один из детей – девочка?
 - (2) я встретил одного из его детей, и это мальчик; какова вероятность того, что второй ребёнок – девочка?

2. Произошло убийство. На месте убийства найдена кровь, которая явно принадлежит убийце. Кровь принадлежит редкой группе, которая присутствует у 1% населения, в том числе у подсудимого.
- (1) Прокурор говорит: «Шанс, что у подсудимого была бы именно такая группа крови, если бы он был невиновен – всего 1%; значит, с вероятностью 99% он виновен». В чём не прав прокурор?
- (2) Адвокат говорит: «В городе живёт миллион человек, то есть у 10000 из них такая группа крови. Значит, всё, что говорит нам эта кровь – это что подсудимый совершил убийство с вероятностью 0.01%; никакое это не доказательство». В чём не прав адвокат?

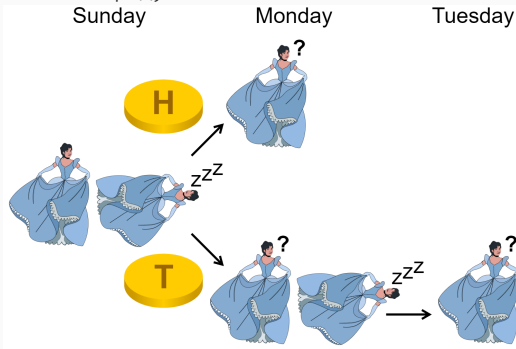
3. Реальные случаи.

- (1) Прокурор указал, что O.J. Simpson уже бил жену в прошлом. Адвокат ответил: «Убивают только одну из 2500 женщин, подвергавшихся семейному насилию, так что это вообще нерелевантно». Суд согласился с адвокатом; верно ли это рассуждение?
- (2) У Sally Clark погибли два младенца; прокурор указал, что вероятность двух случаев SIDS в одной семье, которую он получил из статистики одиночных случаев, — около 1 из 73 миллионов; в чём он не прав?



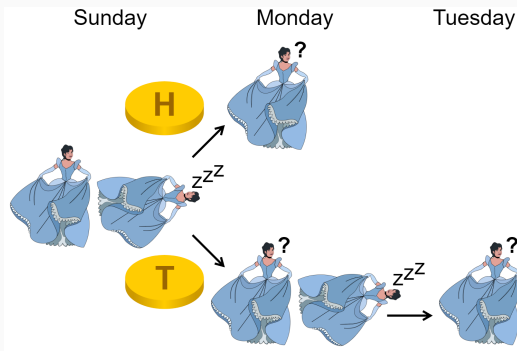
ЗАДАЧА О СПЯЩЕЙ КРАСАВИЦЕ

- А вот парадокс, который до сих пор вызывает споры (Zuboff, 1990; Elga, 2000). В воскресенье Спящую красавицу усыпляют и бросают монетку:
 - если выпадает орёл, её будят в понедельник и опять усыпляют до среды
 - если решка, то будят в понедельник, делают укол, от которого она забывает о пробуждении, и будят ещё раз во вторник, а потом так же в среду



ЗАДАЧА О СПЯЩЕЙ КРАСАВИЦЕ

- Спящая красавица знает всю процедуру эксперимента, но не может различить разные пробуждения и отличить понедельник от вторника, когда просыпается
- Представьте себя на месте Спящей красавицы; вас разбудили; какова вероятность того, что монета упала орлом?



БАЙЕСОВСКИЙ ВЫВОД ДЛЯ МОНЕТКИ

ПОСТАНОВКА ЗАДАЧИ

- Простая задача вывода: дана нечестная монетка, она подброшена N раз, имеется последовательность результатов падения монетки. Надо определить её «нечестность» и предсказать, чем она выпадет в следующий раз.

- Если у нас есть вероятность $p_h = \theta$ того, что монетка выпадет орлом (вероятность решки $p_t = 1 - \theta$), то вероятность того, что выпадет последовательность s , которая содержит n орлов и m решек, равна

$$p(s|p_h = \theta) = \theta^n (1 - \theta)^m.$$

- Сделаем предположение: будем считать, что вероятность монетки до экспериментов для нас распределена равномерно, т.е. у нас нет априорных предпочтений насчёт θ .
- Теперь нужно использовать теорему Байеса и вычислить скрытые параметры.

- Правдоподобие: $p(\theta|s) = \frac{p(s|\theta)p(\theta)}{p(s)}$.
- Здесь $p(\theta)$ следует понимать как непрерывную случайную величину, сосредоточенную на интервале $[0, 1]$, коей она и является. Наше предположение о равномерном распределении в данном случае значит, что априорная вероятность $p(\theta) = 1, \theta \in [0, 1]$ (т.е. априори мы не знаем, насколько нечестна монетка, и предполагаем это равновероятным). А $p(s|\theta)$ мы уже знаем.
- Итого получается:

$$p(\theta|s) = \frac{\theta^n(1-\theta)^m}{p(s)}.$$

- Итого получается:

$$p(\theta|s) = \frac{\theta^n(1-\theta)^m}{p(s)}.$$

- $p(s)$ можно подсчитать как

$$\begin{aligned} p(s) &= \int_0^1 \theta^n(1-\theta)^m d\theta = \\ &= \frac{\Gamma(n+1)\Gamma(m+1)}{\Gamma(n+m+2)} = \frac{n!m!}{(n+m+1)!}, \end{aligned}$$

но найти $\arg \max_{\theta} p(\theta | s) = \frac{n}{n+m}$ можно и без этого.

- Итого получается:

$$p(\theta|s) = \frac{\theta^n(1-\theta)^m}{p(s)}.$$

- Но это ещё не всё. Чтобы предсказать следующий исход, надо найти $p(\text{heads}|s)$:

$$\begin{aligned} p(\text{heads}|s) &= \int_0^1 p(\text{heads}|\theta)p(\theta|s)d\theta = \\ &= \int_0^1 \frac{\theta^{n+1}(1-\theta)^m}{p(s)} dp_h = \\ &= \frac{(n+1)!m!}{(n+m+2)!} \cdot \frac{(n+m+1)!}{n!m!} = \frac{n+1}{n+m+2}. \end{aligned}$$

- Получили правило Лапласа.

- Итого получается:

$$p(\theta|s) = \frac{\theta^n(1-\theta)^m}{p(s)}.$$

- Это была иллюстрация двух основных задач байесовского вывода:
 - найти апостериорное распределение (posterior distribution) на гипотезах/параметрах:

$$p(\theta | D) \propto p(D|\theta)p(\theta)$$

и/или найти максимальную апостериорную гипотезу (maximum a posteriori) $\theta_{\text{MAP}} = \arg \max_{\theta} p(\theta | D)$;

- найти предсказательное распределение (predictive distribution) исходов дальнейших экспериментов:

$$p(x | D) \propto \int_{\theta \in \Theta} p(x | \theta)p(D|\theta)p(\theta)d\theta.$$

СПАСИБО!

Спасибо за внимание!

