

Главная » Рациональность: от ИИ до зомби » Машина в призраке

Просмотр » Редактировать » Обсуждение » Редакции

## Наглядное объяснение теоремы Байеса

*Элиезер Юдковский*

*Примечание редактора книги:* Это сокращённая версия [исходного эссе](#)<sup>#</sup> со множеством интерактивных элементов.

*Примечание редактора сайта:* Упомянутые выше интерактивные элементы, увы, уже давно не поддерживаются ни в каких современных браузерах. Кроме того, некоторое время назад Юдковский добавил в исходное эссе примечание, что он считает его устаревшим и рекомендует вместо него читать [руководство по теореме Байеса на Арбитале](#)<sup>#</sup> (на русский язык не переводилось).

Исходное эссе в своё время [переводилось на русский язык](#) Сергеем Щегловым.

\*\*\*

Ваши друзья и сослуживцы обсуждают какую-то «теорему Байеса» или даже говорят о каком-то «байесианском мышлении». Судя по всему, эта штука им всерьёз нравится, а потому вы лезете в Гугл, находите страницу про «теорему Байеса» и…

Оказывается, это уравнение. И всё. Просто уравнение. Страница, которую вы нашли, даёт математические определения, но не объясняет, что это такое по существу, как это использовать и почему это так нравится вашим друзьям. Просто формула из теории вероятности.

Как получилось, что математическая идея вызывает столь удивительный энтузиазм у людей, которые её изучают? Почему так называемая Байесианская Революция проходит катком по наукам и объявляет своим частным случаем даже экспериментальный метод? Какой секрет известен приверженцам Байеса? Какой свет узрели они?

Скоро вы узнаете. Скоро вы станете одним из нас.

В Сети есть парочка объяснений теоремы Байеса. Однако я пробовал знакомить других с байесианским мышлением и пришёл к выводу, что эти объяснения слишком абстрактны. Байесианское мышление очень контринтуитивно. Люди не применяют байесианское мышление автоматически, им сложно его изучить и легко забыть после завершения обучения. Причём это касается как новичков-студентов, так и опытных профессионалов с многолетним опытом. Судя по всему, байесианское мышление — одна из тех вещей, которые, подобно квантовой механике или селективному тесту Вейсона, по самой своей природе трудно ухватываются нашими встроенными мыслительными способностями.

Во всяком случае так считается. В этом эссе я попытаюсь разъяснить байесианское мышление наглядно. Устроить беспощадно щадящее знакомство, задействующее все возможные пути для понимания чисел, от естественных частот до пространственных визуализаций. Я намерен передать вам не абстрактные правила манипуляции числами, но смысл этих чисел, и понимание: почему правила именно таковы (и не могут быть иными). Когда вы закончите читать эту страницу, байесианские задачи будут вам сниться.

Давайте же начнём.

\*\*\*

Вот задача о ситуации, с которой часто сталкиваются врачи:

У 1% женщин в возрасте 40 лет, участвующих в регулярных обследованиях, есть рак груди. 80% женщин с раком груди получают положительный результат маммографии. 9.6% здоровых женщин также получают положительный результат (маммография, как любые измерения, не даёт стопроцентных результатов). Женщина-пациент из этой возрастной группы получила положительный результат на регулярном обследовании. Какова вероятность того, что она действительно больна раком груди?

Каков ответ с вашей точки зрения? Если вы раньше не сталкивались с такого рода задачами, пожалуйста, подумайте над собственным ответом, прежде чем читать дальше.

\*\*\*

Теперь предположим, что я скажу вам: большинство докторов дают неверный ответ. Обычно лишь около 15% врачей решают эту задачу правильно. («В самом деле? 15%? Это реальные данные, или городская легенда, основанная на опросах в Интернете?» Да, это реальные данные. См. Casscells, Schoenberger, and Grayboys 1978<sup>1</sup>; Eddy 1982<sup>2</sup>; Gigerenzer and Hoffrage 1995<sup>3</sup>; и многие другие статьи. Это удивительный результат, который однако легко воспроизводится, и потому воспроизводится в широких масштабах.)

В случаях с приведённой выше задачей большинство врачей оценивают вероятность между 70% и 80%, что совершенно неправильно.

Вот другая версия той же задачи, с которой врачи справляются несколько лучше:

У 10 из 1000 женщин в возрасте 40 лет, участвующих в регулярных обследованиях, есть рак груди. 800 из 1000 женщин с раком груди получают положительный результат маммографии. 96 из 1000 здоровых женщин также получают положительный результат маммографии. Если 1000 женщин данного возраста пройдут регулярное обследование, какая часть из получивших положительный результат маммографии будет реально больна раком груди?

И наконец, вот задача, с которой врачи справляются лучше всего. В 46% — почти половине! — случаев они дают правильный ответ.

У 100 из 10 000 женщин в возрасте 40 лет, участвующих в регулярных обследованиях, есть рак груди. 80 из каждых 100 женщин с раком груди получают положительный результат маммографии. 950 из оставшихся 9 900 здоровых женщин также получают положительный результат маммографии. Если 10 000 женщин этого возраста пройдут обследование, какая часть из получивших положительный результат будет реально больна раком груди?

Правильный ответ: 7,8%. Получается он так. Из 10 000 женщин 100 больны раком груди, 80 из этих 100 имеет положительные маммограммы. Из тех же 10 000 женщин, 9 900 не имеют рака груди, и из этих 9 900 женщин 950 тоже получат положительные маммограммы. Таким образом, общее число женщин с положительными маммограммами 950 + 80 то есть 1030. Из этих 1 030 женщин с положительными маммограммами, 80 реально больны раком. Таким образом мы получаем отношение 80/1030, или 0,07767, то есть 7,8%.

Иначе говоря, перед маммографическим обследованием 10 000 женщин можно разделить на две группы:

- Группа 1: 100 женщин, больных раком груди.
- Группа 2: 9 900 женщин, не больных раком груди.

В сумме эти группы составляют 10 000 пациенток. Мы никого не потеряли при подсчётах. После маммографии женщин можно разделить уже на четыре группы:

- Группа A: 80 женщин, больных раком груди, и с положительной маммограммой.
- Группа B: 20 женщин, больных раком груди, и с отрицательной маммограммой.
- Группа C: 950 женщин, не больных раком груди, и с положительной маммограммой.
- Группа D: 8 950 женщин, не больных раком груди, и с отрицательной маммограммой.

Сумма групп A и B, групп больных раком груди, соответствует группе 1. Сумма групп C и D, групп здоровых женщин, соответствует группе 2. Если вы примените маммографию к 10 000 пациенток, вы получите 1030 положительных результатов, лишь 80 из которых будут принадлежать пациенткам, реально больным раком. Это правильный ответ, ответ, который врачи должны давать пациенткам с положительными маммограммами, когда они спрашивают о своих шансах на наличие рака груди. Если тринадцать пациенток задают этот вопрос, приблизительно 1 из этих 13 будет больна раком.

\*\*\*

Люди очень часто совершают следующую ошибку: они не учитывают какая доля женщин больна раком и какая доля женщин без рака получает ложноположительные результаты, и фокусируются лишь на группе больных, получивших положительные результаты. Например, абсолютное большинство врачей в уже упоминавшихся исследованиях полагают, что если около 80% женщин с раком груди имеют положительные маммограммы, то и вероятность для женщины с положительной маммограммой быть больной раком тоже около 80%.

Чтобы получить правильный ответ нужно учесть все три условия: процент женщин с раком груди, процент здоровых женщин с ложноположительными результатами исследования и процент женщин с раком груди, получивших истинно положительные результаты.

Исходная доля пациенток с раком груди называется *априорной вероятностью*. Шанс, что пациентка с раком груди получит положительную маммограмму, и шанс, что пациентка без рака получит положительную маммограмму, называются *условными вероятностями*. В совокупности, вся исходная информация называется *априорной*. Результат — ожидаемая вероятность, что пациентка больна раком груди, если её маммограмма положительна, — называется уточнённой вероятностью или *апостериорной вероятностью*. Только что мы видели, что апостериорная вероятность зависит отчасти от априорной вероятности.

Понять, что окончательный ответ всегда зависит от исходной доли женщин с раком груди, можно на примере альтернативной вселенной, где только одна женщина на миллион больна этим раком. Если маммография и в этом мире выявляет рак груди в 8 случаях из 10, давая одновременно ложноположительный результат только в 1 случае из 10, это будет означать сотни тысяч ложноположительных результатов на каждый реально диагностированный рак. Первоначальная вероятный, что женщина больна раком груди, настолько мала, что хотя положительный результат маммографии и увеличивает ожидаемую вероятность, эта вероятность не увеличивается до уверенности или хотя бы до «заметного шанса». Вероятность растёт лишь с 1:1 000 000 до 1:100 000.

Таким образом мы видим, что результаты маммографии не заменяют



предыдущую информацию о шансах пациента оказаться больной раком. Маммография лишь сдвигает ожидаемую вероятность в направлении своего результата. Положительный результат сдвигает первоначальную вероятность вверх, отрицательный - вниз. Например, в нашей первой задаче, где 1% женщин болеют раком, 80% больных раком получают положительные результаты маммографии, и 9.6% здоровых женщин получают ложноположительные результаты, положительный результат маммографии сдвигает 1% шанс до 7.8% шанса.

Большинство же людей, впервые столкнувшись с подобными задачами, просто в уме заменяют исходную вероятность в 1% на вероятность в 80% для больных раком получить положительную маммограмму. Интуитивно это кажется хорошей идеей, но на самом деле это не так. «Вероятность того, что у женщины с положительной маммограммой есть рак груди» не то же самое, что «вероятность для женщины с раком груди получить положительный результат маммографии». Эти величины разные, как яблоки и сыр.

**\*\*\***

#### Почему байесианец перешёл дорогу?

Нужно больше информации, чтобы ответить на этот вопрос.

**\*\*\***

Предположим, что в бочке находится множество маленьких пластиковых капсул. Некоторые капсулы окрашены в красный цвет, некоторые - в синий. У 40% от всех капсул внутри жемчужина, 60% пусты. В синий цвет окрашены 30% капсул, содержащих жемчужины, и 10% пустых капсул. Какова вероятность, что синяя капсула содержит жемчужину? В этом примере числа достаточно просты, чтобы найти решение в уме, и я предлагаю вам попробовать это сделать.

Эту же задачу можно записать более компактно:

*P*(жемчуг) = 40%
*P* (синий|жемчуг) = 30%
*P* (синий|¬жемчуг) = 10%
*P* (жемчуг|синий) = ?

«¬» здесь сокращение для «нет», так что ¬жемчуг читается как «нет жемчуга».

*P*(синий|жемчуг) — это сокращённая запись выражения «вероятность синего при условии жемчужина» или «вероятность того, что капсула синяя, если внутри находится жемчужина». Справа от вертикальной черты записывается то, что вы уже знаете — условие, а слева — следствие или вывод. Если у нас *P*(синий|жемчуг) = 30%, и мы уже знаем, что внутри некоторой капсулы есть жемчужина, то мы делаем вывод, что эта капсула будет синей с вероятностью 30%. Таким образом, мы ищем ответ на вопрос: «какова вероятность, что внутри синей капсулы находится жемчужина». Записывается как *P*(жемчужина|синий).

Теперь вернёмся к задаче. Мы знаем, что 40% капсул содержат жемчужины, а 60% капсул совершенно пусты. 30% капсул, содержащих жемчужины, синего цвета, то есть 12% капсул от общего числа содержат жемчужины и синие. 10% от пустых капсул синего цвета, то есть от общего количества 6% капсул пустые и синие. Всего у нас получается 18% синих капсул, и 12% синих и содержащих жемчужины капсул, так что шанс для синей капсулы содержать жемчужину равен 12/18 = 2/3 = примерно 67%.

Как и в предыдущем случае мы можем рассмотреть предельные случаи и убедиться в необходимости всех трёх условий. В гигантской бочке, где лишь одна капсула из тысячи содержит жемчужину, знание о том, что капсула синяя, увеличивает наши шансы с 0,1% до 0,3% (а не с 40% до 67%). Аналогично, если 999 из 1000 капсул содержат жемчужины, знание, что капсула синяя, увеличивает шансы с 99,9% до 99,966%. Вероятность, что капсула не содержит жемчужины, меняется с 1/1000 до примерно 1/3000.

В задаче с капсулами и жемчужинами большинство опрошиваемых, не знакомых с байесианским мышлением, ответили бы, что вероятность для синей капсулы содержать жемчужину будет 30%, или возможно 20% (30% шансов на истинно-положительный результат минус 10% шансов на ложно-положительный). Даже если эта мысленная операция представляется хорошей идеей, она не имеет смысла в контексте предложенной задачи. Это все равно как если бы вы спросили у первоклассника: «Если восемнадцать человек сели в автобус, а потом ещё семь человек сели в автобус, сколько лет водителю?» Многие дети ответят: «Двадцать пять». Они понимают, что получили возможность использовать специальную процедуру (сложение), но не вполне осознают связь этой процедуры с реальностью. Точно так же для нахождения вероятности, что пациентка с положительной маммограммой больна раком груди, абсолютно бессмысленно заменять исходную вероятность заболевания на вероятность, что женщина с раком груди получит положительную маммограмму. Столь же бессмысленно вычитать вероятность ложноположительного результата из вероятности истинно-положительного. Эти операции здесь настолько же неуместны, как подсчёт пассажиров в автобусе для определения возраста водителя.

**\*\*\***

Исследование Гигеренцера и Хоффраге 1995 года показывает, что некоторые формулировки задач лучше помогают «включить» байесианское мышление<sup>[4]</sup>. Хуже всего работают формулировки с вероятностями. Чуть лучше работает использование частот вместо вероятности: задача не меняется, но в условии говорится не «1% женщин болен раком груди», а «1 из 100 женщин больна раком груди», «80 из 100 женщин с раком груди получают положительные маммограммы» и так далее. Почему при таких формулировках больше испытуемых начинают думать по байесиански? Вероятно, слова «одна из ста женщин» побуждают представить X женщин с раком и подготавливают к тому, чтобы потом представить X женщин с раком и положительной маммограммой, и так далее.

Эффективнее же всего (из найденного на данный момент) работают «естественные частоты»: формулировки, что 40 из 100 капсул содержат жемчужины, что 12 из 40 капсул с жемчужинами — синие, и что 6 из оставшихся 60 пустых капсул тоже синие. При описании через естественные частоты информация об априорных вероятностях включается в описание условных вероятностей. Если вы захотите узнать условные вероятности в результате естественного эксперимента — вскрывая все капсулы — вы обнаружите, 40 капсул с жемчужинами, из которых 12 — синие, и 60 пустых, из которых 6 — синие. Таким образом, вы увидите, что синие капсулы с жемчужинами встречаются в два раза чаще, чем синие капсулы без жемчужин.

К сожалению, хотя естественные частоты — это шаг в правильном направлении, их всё же недостаточно. Когда задачи формулируются через естественные частоты, доля людей, использующие байесианское мышление, возрастает примерно до половины. Это значительное улучшение, но недостаточное, когда речь идёт о реальных врачах и реальных пациентах.

**\*\*\***

#### Где мне брать априорные вероятности для моей задачи?

Многие априорные вероятности можно найти в справочниках по химии и физике.

#### Откуда вообще изначально берутся априорные вероятности?

Никогда не задавайте этот вопрос.

#### Ага. Тогда откуда учёные берут априорные вероятности?

Априорные вероятности для научных задач устанавливаются ежегодным голосованием [Американской Ассоциации Содействия Развитию Науки](#)<sup>[ⓘ]</sup>. В последние годы голосования становятся всё более нервными. Царит всеобщая язвительность, ассоциация раскалывается на фракции, произошло даже несколько политических убийств. Возможно, это следствия конфликтов в Байесианском совете, а может быть у дискутирующих слишком много свободного времени. Никто не знает точно.

#### Понятно. А где все остальные берут априорные вероятности?

Загружают свои априорные вероятности из [Kazaa](#)<sup>[ⓘ]</sup>.

#### Но что если нужны мне априорные вероятности недоступны на Kazaa?

На задворках китайского квартала Сан-Франциско есть маленький тесный антикварный магазинчик. Ни в коем случае не спрашивайте про бронзовую крысу.

Вообще-то, априорные вероятности могут быть истинными и ложными — как и конечное решение. Они отражают реальность и о них можно судить, сравнивая их с реальностью. Например, если вы думаете, что 920 из 10000 женщин в выборке больны раком груди, а настоящее количество 100 из 10000, то ваша априорная вероятность неверна. Например, для нашей задачи априорные вероятности могли появиться из результатов трёх исследований: исследование историй болезни женщин с раком груди, чтобы узнать, сколько из них получили положительный результат маммографии, исследование женщин без рака груди, чтобы узнать, сколько из них получили положительный результат маммографии, и эпидемиологическое исследование распространённости рака груди в некотором демографическом срезе.

**\*\*\***

Вероятность *P*(*A*,*B*) равна *P*(*B*,*A*), но вероятность *P*(*A*|*B*) не равна *P*(*B*|*A*) и уж точно *P*(*A*,*B*) не равна *P*(*A*|*B*). Довольно часто эти вероятности путают.

Чтобы познакомиться поближе с этими величинами и отношениями между ними, сыграем в игру «посчитаем количество степеней свободы». Например, у величины *P*(рак) и *P*(¬рак) совместно одна степень свободы, поскольку *P*(*A*) + *P*(¬*A*) = 1. Если вы знаете, что *P*(¬рак) = 0,99, вы можете получить *P*(рак) = 1 - *P*(¬рак) = 0,01.

*P*(положительный|рак) и *P*(¬положительный|рак) тоже имеют между собой только одну степень свободы; каждая женщина с раком груди может либо получить положительную маммограмму, либо не получить. С другой стороны, *P*(положительный|рак) и *P*(положительный|¬рак) связаны двумя степенями свободы. Вы можете можете иметь тест, который даёт положительный результат для 80% больных раком и для 9.6% здоровых, или тест, дающий 70% для больных и 2% для здоровых, или даже тест, возвращающий «положительный» результат для 30% больных раком и для 92% здоровых. Две эти величины — результаты теста для больных и для здоровых — математически независимы; ни одна из них не может быть вычислена на основании другой каким-либо способом, и это означает, что они имеют две степени свободы между собой.

Что насчёт *P*(положительный, рак), *P*(положительный|рак) и *P*(рак)? У нас здесь три величины. Сколько же степеней свободы? В этом случае должно соблюдаться равенство:

*P*(положительный, рак) = *P*(положительный|рак) × *P*(рак).

Это равенство уменьшает количество степеней свободы на единицу. Если мы знаем долю пациенток с раком, а также вероятность, что у пациентки с раком будет положительная маммограмма, мы можем перемножить эти числа и получить долю пациенток с раком груди и положительной маммограммой.



Аналогично, если мы знаем количество пациенток с раком груди и положительными маммограммами, а также общее число пациенток с раком груди, мы можем оценить шансы, что женщина с раком груди получит положительную маммограмму, простым делением:  $P(\text{положительный}|\text{рак}) = P(\text{положительный, рак}) / P(\text{рак})$ . Фактически, именно так и калибруют медицинские тесты; вы проводите исследование на 8520 женщинах с раком груди, получаете в результате 6816 (или около) женщин с раком груди и положительными маммограммами, после чего делите 6816 на 8250, чтобы найти 80% вероятность для женщины, больной раком, получить положительную маммограмму. (Между прочим, если вы случайно поделите 8250 на 6816, вместо того чтобы наоборот, ваши вычисления начнут давать странные результаты, такие как утверждение, что 125% женщин с раком груди и положительной маммограммой больны раком. По моему опыту, это довольно распространённая ошибка в расчётах по байесовской арифметике.) И в заключении, если вы знаете  $P(\text{положительный,рак})$  и  $P(\text{положительный}|\text{рак})$ , вы можете вывести исходную долю больных раком пациенток. Итого, у этих трёх величин две степени свободы: если вы знаете две из них, вы можете рассчитать третью.

А как связаны  $P(\text{положительный})$ ,  $P(\text{положительный,рак})$  и  $P(\text{положительный,¬рак})$ ? Снова у нас только две степени свободы для трёх переменных. Уравнение, устраняющее ещё одну степень свободы:

$$P(\text{положительный}) = P(\text{положительный,рак}) + P(\text{положительный,¬рак})$$

Для начала, вот как рассчитывается  $P(\text{положительный})$ : мы берём количество женщин с раком груди и положительными маммограммами, прибавляем к нему количество женщин без рака груди и положительными маммограммами, и получаем вместе искомое количество женщин с положительными маммограммами. Конечно, весьма странно проводить исследование для определения числа женщин с положительными маммограммами - только одного числа и ничего больше - но в теории вы можете это сделать. И если потом вы проведёте другое исследование, и найдёте количество с положительными маммограммами и раком груди, вы также узнаете и количество женщин с положительными маммограммами и без рака груди — поскольку женщина с положительной маммограммой или больна раком груди, или нет. В общем,  $P(A,B) + P(A,¬B) = P(A)$ . Аналогично,  $P(A,B) + P(¬A,B) = P(B)$ .

А что у нас с  $P(\text{положительный,рак})$ ,  $P(\text{положительный,¬рак})$ ,  $P(\text{¬положительный,рак})$  и  $P(\text{¬положительный,¬рак})$ ? Поначалу соблазнительно решить, что здесь только две степени свободы для четырёх переменных — что возможно, например, получить  $P(\text{положительный,¬рак})$  через умножение  $P(\text{положительный}) \times P(\text{¬рак})$ , и тогда все четыре значения могут быть найдены через две независимые переменные —  $P(\text{положительный})$  и  $P(\text{рак})$ . Но это не тот случай!  $P(\text{положительный,¬рак}) = P(\text{положительный}) \times P(\text{¬рак})$  верно только для двух вероятностей, которые статистически независимы — если бы шансы, что женщина больна раком груди, никак не были бы связаны с тем, что у неё положительная маммограмма. Но как вы уже знаете, такое возможно лишь в случае, если обе условные вероятности одинаковы - требование, которое может устранить одну степень свободы. Если вы вспомните, что эти четыре величины являются группами A, B, C и D, вы сможете посмотреть на эти четыре группы и осознать, что теоретически вы можете поместить любое число женщин в каждую из этих групп. Если вы начнете с группы 80 женщин с раком груди и положительными маммограммами, нет никаких причин, почему вы не могли бы добавить следующую группу из 500 женщин с раком груди и отрицательными маммограммами, затем группу из 3 женщин без рака груди и отрицательными маммограммами, и так далее. Теперь может показаться, что эти четыре величины имеют четыре степени свободы. Так оно и есть, за исключением того момента, что когда они записываются как вероятности, мы должны нормализовать их до долей единицы от общей группы, что добавит следующее ограничение:  $P(\text{положительный,рак}) + P(\text{положительный,¬рак}) + P(\text{¬положительный,рак}) + P(\text{¬положительный,¬рак}) = 1$ . Это уравнение забирает одну из степеней свободы, оставляя в итоге три степени для четырёх величин. Если вы зададите доли единицы для женщин в группах A, B, и D, вы сможете вывести долю единицы для женщин в группе C.

Располагая четырьмя группами A, B, C, и D, очень просто вычислить все прочее:

$$P(\text{рак}) = \frac{A + B}{A + B + C + D}$$

$$P(\text{¬положительный} | \text{рак}) = \frac{B}{A + B}$$

И так далее. Поскольку набор {A, B, C, D} содержит три степени свободы, отсюда следует, что полный набор из 16 вероятностей также содержит только три степени свободы. Вспоминаем, что в наших задачах мы всегда нуждаемся в трёх частях информации — одной априорной и двух условных вероятностях, — которые и в самом деле имеют три степени свободы (будучи независимыми друг от друга). Собственно, в байесовских задачах любые три величины с тремя степенями свободы будут исчерпывающе описывать ситуацию.

Вероятность, что тест даёт истинно положительный результат, делённая на вероятность, что тест даёт ложноположительный результат называется степенью правдоподобия этого теста. Степень правдоподобия положительного результата показывает в итоге, насколько положительный результат теста сместит априорную вероятность. Достаточно ли одной степени правдоподобия, чтобы сказать, что мы знаем всё о полезности данного теста?

Нет, недостаточно! Степень правдоподобия говорит всё, что требуется о значении положительного результата. Но она ничего не говорит о значении отрицательного результата, и не может сказать, как часто тест оказывается полезным. Например, маммография 80% истинно положительных и 9,6% ложноположительных результатов имеет ту же степень правдоподобия, что и тест с 8% истинно положительных и 0,96% ложноположительных результатов. Но несмотря на одинаковость степеней правдоподобия, первый тест намного более полезен — он чаще определяет болезнь, а его отрицательный результат служит более сильным признаком здоровья.

\*\*\*

Предположим, что вы используете два теста для рака один за другим — скажем, стандартную маммографию и ещё какой-нибудь тест, совершенно независимый по отношению к маммографии. Поскольку я не знаю ни одного такого теста (реально не зависящего от маммографии), я просто выдумаю такой тест для нашей задачи и назову его «Разделительный тест Темза-Брейлора», предположив, что он позволяет обнаружить, что некоторые клетки делятся быстрее чем другие. Теперь предположим, что тест Темза-Брейлора даёт истинно положительные результаты для 90% пациенток с раком груди, и ложноположительные результаты для 5% здоровых пациенток. Априорная вероятность рака груди пусть будет все тот же 1%. Если пациентка получает положительные результаты маммографии и теста Темза-Брейлора, какова вероятность, что она больна раком груди?

Один из способов решить эту задачу: взять полученную вероятность для положительной маммограммы, которую мы уже посчитали и получили 7,8%, и подставить в тест Темза-Брейлора как его априорную вероятность. В этом случае мы получим 60%.

Предположим, что априорная распространённость рака груди в некоторой группе населения — 1%. Предположим, что у нас, как у врачей, есть три независимых теста для рака груди. Первый тест (A) - это маммография со степенью правдоподобия в 80%/9,6% = 8,33. У второго теста (B) степень правдоподобия — 18,0 (например, 90% против 5%), у третьего (C) — 3,5 (например, 70% против 20% или 35% против 10%, что совершенно одинаково). Предположим, пациентка получила положительные результаты всех трёх тестов. Какова вероятность, что у неё рак груди?

Привожу интересный трюк для упрощения расчётов. Если априорная распространённость рака груди в группе населения 1%, то 1 из 100 женщин больны раком груди, а 99 женщин здоровы. Тогда, если мы перепишем вероятность в 1% как отношение шансов, отношение получится 1:99.

Теперь запишем степени правдоподобия для тестов A, B и C:

$$\begin{aligned} 8,33 : 1 &= 25 : 3 \\ 18,0 : 1 &= 18 : 1 \\ 3,5 : 1 &= 7 : 2 \end{aligned}$$

Шансы для женщины с раком груди, получившей положительные результаты всех трёх тестов, против женщины без рака груди, также получившей положительные результаты, составят:

$$(1 \times 25 \times 18 \times 7) : (99 \times 3 \times 1 \times 2) = 3150 : 594 .$$

Чтобы вернуться от шансов к вероятностям, просто напишем:

$$3150 / (3150 + 594) = 84\% .$$

Этот приём работает независимо от способа записи отношения шансов, то есть 8,33:1 то же самое, что 25:3 или 75:9. Также нет разницы, в каком порядке выполняются тесты, или в каком порядке подсчитываются результаты. Читатель может доказать это в качестве упражнения.

\*\*\*

Э. Т. Джейнс в своей книге «Теория вероятностей в науке и инженерном деле» предлагает измерять правдоподобие и силу свидетельства в децибелах<sup>5</sup>.

Почему в децибелах?

Децибелы используются для измерения экспоненциальных различий интенсивности. Например, если звук автомобильной сирены несёт в 10000 раз больше энергии (на квадратный метр в секунду), чем звонок будильника, то сирена на 40 децибел громче. Звук чириканья птицы может нести в 1000 раз меньше энергии, чем звук будильника, следовательно, он будет на 30 децибел тише. Для получения количества децибел нужно взять десятичный логарифм интенсивности и умножить на 10.

$$\text{д е ц и б е л ы} = 10 \cdot \log_{10} \text{и н т е н с и в н о с т ь}$$

или

$$\text{и н т е н с и в н о с т ь} = 10^{\text{д е ц и б е л ы} / 10}$$

Когда вы решаете задачу с априорной вероятностью в 1%, что даёт отношение шансов 1:99, и тремя тестами со степенями правдоподобия в 25:3, 18:1 и 7:2, вы можете перемножать эти числа... или же просто складывать их логарифмы:

$$10 \cdot \log_{10} \frac{1}{99} \approx -20$$

$$10 \cdot \log_{10} \frac{25}{3} \approx 9$$

$$10 \cdot \log_{10} \frac{18}{1} \approx 13$$

$$10 \cdot \log_{10} \frac{7}{2} \approx 5$$

Изначально достаточно маловероятно, что пациентка больна раком груди — и наш уровень правдоподобия составляет -20 децибел. Далее три теста дают результаты, соответствующие 9, 13 и 5 децибелам доказательности. Это поднимает уровень правдоподобия на 27 децибел, и в результате априорное правдоподобие в -20 превращается в апостериорные 7 децибел. Соответственно шансы изменяются с 1:99 до 5:1, а вероятность - с 1% до 83%.

\*\*\*

Вы — механик по игрушкам. Когда игрушка ломается, это в 30% случаев происходит из-за засорения трубки. Если трубка игрушки засорена, существует 45% вероятность, что игрушка будет искрить при попытке её завести. Если трубка не засорена, есть только 5% шансов, что появятся искры. Покупатель принёс вам неисправную игрушку. вы попробовали её завести и увидели искры. Какова вероятность, что у этой искрящей игрушки засорена трубка?

Какую последовательность арифметических операций вы проделаете для решения этой задачи?

$$(45\% \times 30\%) / (45\% \times 30\% + 5\% \times 70\%)$$

Аналогично, чтобы получить вероятность того, что у женщины с положительной маммограммой есть рак груди, мы вычисляли:

$$\frac{P(\text{положительный | рак}) \times P(\text{рак})}{P(\text{положительный | рак}) \times P(\text{рак}) + P(\text{положительный | ¬рак}) \times P(\text{¬рак})}$$

что равно

$$\frac{P(\text{положительный, рак})}{P(\text{положительный, рак}) + P(\text{положительный, ¬рак})}$$

что в свою очередь равно

$$\frac{P(\text{положительный, рак})}{P(\text{положительный})}$$

или

$$P(\text{рак} | \text{положительный}).$$

В общем случае эти вычисления называются теоремой Байеса или правилом Байеса.

**Теорема Байеса:**

$$P(A | X) = \frac{P(X | A) \times P(A)}{P(X | A) \times P(A) + P(X | \neg A) \times P(\neg A)}$$

Когда мы хотим исследовать некоторое явление А и у нас есть наблюдение Х, которое что-то говорит про А (например, как в предыдущем примере, А — это рак груди, а Х — положительный результат маммографии), теорема Байеса говорит нам, как именно мы должны изменить нашу оценку вероятности А после получения нового свидетельства Х.

Возможно, сейчас теорема уже кажется вопиюще очевидной или даже тавтологичной, а вовсе не новой и удивительной. В этом случае это введение выполнило своё предназначение.

\*\*\*

Теорема Байеса описывает, что такое свидетельство и насколько сильное это свидетельство. Чтобы оценить статистические модели, их сравнивают с методом Байеса, ведь в статистике метод Байеса показывает, какую максимальную пользу можно получить из свидетельства — точно так же, как термодинамика показывает, максимальное количество работы, которое можно извлечь из разницы температур. Именно поэтому учёные-когнитивисты говорят об агентах, рассуждающих по Байесу. В когнитивной науке термином «байесовский агент» фактически обозначается рациональный разум.

Теорема Байеса также помогает лучше понять некоторые эвристики человеческого мышления.

Например, когда когнитивные психологи обсуждают теорему Байеса, можно услышать, что люди не учитывают априорные вероятности. В смысле, когда люди решают задачу, где некоторое свидетельство Х говорит что-то об истинности условия А, они судят о правдоподобности А исключительно по силе свидетельства Х в пользу А без учёта априорной вероятности А. Если, например, в задаче с маммограммой вы посчитали, что у женщины есть рак груди с вероятностью 70-80%, то вы не учли априорную вероятность. Для такого стиля мышления нет разницы, у сколько вообще женщин в популяции есть рак груди — у 1% или у 10%. Чтобы частично компенсировать встроенные в нас искажения, нам нужно чаще обращать внимание на априорные вероятности!

Также люди, рассуждая о том, насколько сильно Х свидетельствует в пользу А, часто сбрасывают повышенное внимание на  $P(X|A)$  и недостаточное на  $P(X|\neg A)$ . Сила свидетельства Х для события А зависит не только о того, насколько мы ожидаем обнаружить Х, если А — истинно, но также от того, насколько мы ожидаем не обнаружить Х, если А — ложно. Например, если идёт дождь, то можно ожидать, что трава мокрая —  $P(\text{мокрая трава} | \text{дождь}) \approx 1$ . Однако мокрая трава не обязательно означает, что шёл дождь. Возможно, был включён автополив, возможно вы видите утреннюю росу. Поскольку  $P(\text{мокрая трава} | \neg \text{дождь})$  значительно больше нуля, то  $P(\text{дождь} | \text{мокрая трава})$  существенно меньше единицы. С другой стороны, если трава мокнет только под дождём и никак иначе, тогда знание, что трава мокрая всегда даёт понять, что прошёл дождь.  $P(\text{дождь} | \text{мокрая трава}) \approx 1$ , даже если  $P(\text{мокрая трава} | \text{дождь}) = 50\%$ , то есть даже если трава становится мокрой лишь в половине случаев, когда идёт дождь. Свидетельство всегда зависит от того, насколько отличаются эти две условные вероятности. Сильное свидетельство получается не в результате очень высокой вероятности того, что А влечёт за собой Х, а в результате того, что очень мала вероятность, что не-А влечёт за собой Х.

Байесовская революция в науке набирает силу не только из-за того, что всё больше учёных-когнитивистов замечают, что у наших мыслительных процессов байесовская структура. И не только из-за того, что во всех областях учёные начинают оценивать статистические методы, сравнивая их с байесовским подходом. В первую очередь байесовскую революцию движет идея, что наука сама по себе есть частный случай теоремы Байеса: экспериментальные факты — всего лишь байесовские свидетельства. Революционеры-байесовцы настаивают, что когда вы производите эксперимент и получаете свидетельства, которые «подтверждают» или «опровергают» Вашу теорию, эти подтверждения и опровержения подчиняются правилам Байеса. Например, вы должны принимать во внимание не только то, предсказывает ли ваша теория некоторое явление, но и предсказывают ли это явление другие теории.

Самой популярной философией науки в прошлом, видимо, следует считать фальсификационизм Карла Поппера. Сегодня байесовская революция сбрасывает эту философию с трона. Идея Карла Поппера, что теории могут быть достоверно фальсифицированы, но никогда не могут быть достоверно подтверждены, - всего лишь частный случай теоремы Байеса. Если  $P(X|A) \approx 1$  — теория А даёт однозначное предсказание Х, — то наблюдение ¬Х очень сильно фальсифицирует А. С другой стороны, если  $P(X|A) \approx 1$ , и мы в очередной раз наблюдаем Х, это не является каким-то подтверждением теории А, поскольку может существовать другое условие В, для которого  $P(X|B) \approx 1$ , и в этом случае наблюдение Х не позволяет выбрать между А и В. Чтобы наблюдение Х достоверно подтвердило А, мы должны знать не что  $P(X|A) \approx 1$ , а что  $P(X|\neg A) \approx 0$ , чего мы знать не можем, поскольку не в состоянии перебрать все возможные альтернативные объяснения. К примеру, когда эйнштейновская общая теория относительности сменила невероятно хорошо подтверждённую теорию гравитации Ньютона, выяснилось, что все ньютоновские предсказания являются частным случаем эйнштейновских.

Попперовскую философию можно даже записать математически. Соотношение правдоподобия для Х —  $P(X|A) / P(X|\neg A)$  — определяет, насколько наблюдение Х сдвигает вероятность истинности А, соотношение правдоподобия говорит нам, насколько сильно Х как свидетельство. Так вот, в своей теории А вы можете предсказать Х с вероятностью в 1. Но вы не можете контролировать знаменатель степени правдоподобия,  $P(X|\neg A)$ , поскольку всегда могут найтись другие теории, которые тоже предсказывают Х. И хотя мы соглашаемся с простейшей теорией, которая согласуется с имеющимися свидетельствами, однажды мы можем обнаружить свидетельство, которое другая теория предсказывает, а ваша — нет. Так незаметная до поры погрешность опрокинула ньютоновскую теорию гравитации. Поэтому есть предел достоверности, дальше которого с помощью успешных предсказаний пройти нельзя. Существует предел для соотношения правдоподобия, которое можно получить на основании подтверждающих фактов.

С другой стороны, если вы обнаружите некоторое свидетельство Y, которое ваша теория однозначно не допускает, это чрезвычайно сильно свидетельство против вашей теории. Если  $P(Y|A)$  стремится к нулю, то и степень правдоподобия стремится к нулю. Например, если  $P(Y|A) = 0.0001\%$ , и  $P(Y|\neg A)$  составляет 1%, то степень правдоподобия  $P(Y|A) / P(Y|\neg A)$  будет 1:10000. -40 децибел достоверности! Или в обратную сторону, если  $P(Y|A)$  является очень малой, то  $P(Y|\neg A) / P(Y|A)$  будет очень большой, что означает намного большую вероятность обнаружить ¬А, нежели А. Фальсификация намного сильнее подтверждения. Это прямо следует из предшествующих соображений, что очень сильное свидетельство возникает не от высокой вероятности, с которой из А следует Х, а из очень низкой вероятности, что не-А может привести к Х. В основе попперовской эвристики о фальсифицируемости лежит точное правило Байеса.

Схожим образом попперовский тезис, что любящая идея должна быть фальсифицируемой, может быть интерпретирован как воплощение байесовского Закона Сохранения Вероятностей: если результат Х является положительным свидетельством для теории, то результат ¬Х будет опровергать теорию до некоторой степени. Если вы попытаетесь интерпретировать оба результата — Х и ¬Х — как «подтверждения» теории, то правило Байеса скажет, что это невозможно! Для увеличения вероятности теории вы должны простить её события, которые могут уменьшить её вероятность. Это не просто способ выявлять возможных мошенников в науке, но ещё и прямое следствие из байесовской теории вероятности. В то же время, попперовская идея, что есть только фальсификация и нет такой штуки как подтверждение оказывается неверной. Теорема Байеса показывает, что фальсификация действительно очень сильное свидетельство в сравнении с подтверждением, но фальсификация все равно имеет вероятностную природу; она не подчиняется каким-то отличным от подтверждения правилам, как утверждал Поппер.



Итак, мы обнаружили, что многие явления в когнитивных науках, а также используемые учёными статистические методы, а также научный метод сам по себе — это частные случаи теоремы Байеса. Вот она, байесовская революция.

\* \* \*

Записав теорему Байеса в виде формулы, мы можем подробно обсудить её компоненты.

$$P(A|X) = \frac{P(X|A) \times P(A)}{(P(X|A) \times P(A) + P(X|\neg A) \times P(\neg A))}$$

Начнём с  $P(A|X)$ . Если вы сомневаетесь, что такое А и что такое X в Теореме Байеса, начинайте с  $P(A|X)$  в левой части уравнения; это самая простая часть для понимания. А это штука, насчёт которой мы хотим что-то узнать. X — это то, как мы её видим. X — это факт, который мы используем, чтобы вынести суждение насчёт А. Запомните, что в любом выражении вида  $P(Q|P)$  мы хотим узнать вероятность Q, которую даёт ему P, степень, в которой P предполагает Q - в более вразумительной записи, которую уже поздно предлагать статистикам, это выглядело бы как  $P(Q \leftarrow P)$ .

$P(Q|P)$  тесно связано с  $P(Q,P)$ , но это не одно и то же. Выраженное как вероятность или доля,  $P(Q,P)$  представляет собой отношение вещей, обладающих свойством Q и свойством P среди всех вещей; например, отношение «женщин с раком груди и положительной маммограммой» к численности всех женщин. Если общее число женщин 10000, и 80 женщин имеют рак груди и положительную маммограмму, то  $P(Q,P)$  будет  $80/10,000 = 0,8\%$ . Вы можете видеть, что абсолютное количество, 80, преобразуется в вероятность через отношение к группе всех женщин. Чтобы сделать это еще понятнее, предположим что имеется группа из 641 женщин с раком груди и положительными маммограммами внутри общей выборки из 89031 женщин. 641 - абсолютное количество. Если вы возьмёте случайную женщину из всей выборки, то вероятность что это будет женщина с раком груди и положительной маммограммой, равна  $P(Q,P)$ , или 0,72% в этом примере.

С другой стороны,  $P(Q|P)$  — это отношение количества объектов, обладающих свойствами Q и P, к количеству объектов со свойством P. Например, доля женщин с раком груди и положительной маммограммой в группе всех женщин с положительной маммограммой. Если у нас есть 641 женщина с раком груди и положительной маммограммой, 7915 женщин с положительными маммограммами, и 89031 женщин во всей выборке, то  $P(Q,P)$  это вероятность получить одну из 641 этих женщин при случайном выборе из всех 89031, в то время как  $P(Q|P)$  это вероятность выбрать одну из 641 женщин из куда меньшей группы в 7915 человек.

На самом деле,  $P(Q|P)$  означает ровно то же самое, что и  $P(Q,P|P)$ , но писать постоянно ещё одно P — излишняя роскошь. Вы и так знаете, что ваши объекты обладают свойством P, а теперь исследуете свойство Q, хотя при этом вы исследуете размер группы (Q, P) в составе группы P, а не на размер группы Q в составе группы P (что было бы абсурдом). P в скобках в записи  $P(Q|P)$  означает, что свойство P — задано, вы работаете только с объектом, обладающих этим свойством. Когда вы фокусируете своё внимание лишь на этой меньшей, чем целое, группе, многие другие вероятности изменяются. Если вы берете свойство P как заданное, то  $P(Q\&P)$  становится равным просто  $P(Q)$  — по крайней мере, по отношению к группе P. При этом старое  $P(Q)$ , частота «вещей, которые имеют свойство Q во всей выборке», пересматривается к новой частоте «вещей, которые имеют свойство Q в части выборки, обладающей свойством P». Когда P задано, то P становится всем нашим миром, и в нем искать (Q,P) - то же самое, что искать просто Q.

Если вы сосредоточите ваше внимание только на множестве капсул синего цвета, то «вероятность, что капсула содержит жемчужину» сразу же изменится: количество жемчужин во множестве синих капсул отличается от количества жемчужин во всех капсулах. Условие задачи, свойство, на котором фокусируется наше внимание, всегда стоит в правой части выражения  $P(Q|P)$ ; это P становится нашим миром, всем, что мы видим, и это означает, что «заданное» P всегда имеет вероятность 1 — именно потому, что оно задано. Таким образом,  $P(Q|P)$  означает «Если вероятность P равна 1, какова вероятность Q?», или «Если мы примем во внимание только вещи или события, для которых P истинно, какой будет вероятность Q?». Q, находящаяся с другой стороны выражения, не является чем-то известным — его вероятность может быть 10%, или 90%, или любой другой. Так что когда вы применяете теорему Байеса, и пишете в левой части  $P(A|X)$ , вы делаете это с целью уточнить вероятность А после обнаружения X, найти новую вероятность А, при условии что вы знаете X, степень, в которой X влечёт за собой А. Можно сказать, что X всегда наблюдение или факт, а А — это предмет исследования, то, о чём мы хотим что-то узнать.

\* \* \*

Правая часть формулы Байеса получается из левой путём следующих преобразований:

$$P(A|X) = P(A|X)$$

$$P(A|X) = \frac{P(X, A)}{P(X)}$$

$$P(A|X) = \frac{P(X, A)}{P(X, A) + P(X, \neg A)}$$

$$P(A|X) = \frac{P(X|A) \times P(A)}{P(X|A) \times P(A) + P(X|\neg A) \times P(\neg A)}$$

В итоге в правой части уравнения условные вероятности имеют вид  $P(X|A)$  или  $P(X|\neg A)$ . В левой части находится условная вероятность  $P(A|X)$ . Симметрия возникает потому, что исходные причинно-следственные связи обычно направлены от фактов к наблюдениям, например, от рака груди к положительным результатам маммографии. Рассуждения же наоборот обычно идут от наблюдений к фактам, например, от результатов маммографии к наличию рака. Левая часть формулы Байеса — это элементарный логический шаг от наблюдения положительных результатов маммографии к заключению об увеличении вероятности рака груди. Следование записывается справа налево, так что мы пишем  $P(\text{рак}| \text{положительный})$  в левой части уравнения. Правая часть формулы Байеса описывает причинно-следственную связь — например, переход от рака груди к положительной маммограмме, — поэтому условные вероятности в правой части выглядят как  $P(\text{положительный}|\text{рак})$  и  $P(\text{положительный}|\neg\text{рак})$ .

Вот что такое теорема Байеса. Рациональный вывод в левой части, физическая причинность в правой. Разум с одной стороны, реальность — с другой. Помните, как научный метод оказался частным случаем теоремы Байеса? Поэтически можно выразиться, что теорема Байеса привязывает мышление к реальной вселенной.

Отлично, мы закончили.

\* \* \*

Преподобный Байес говорит:



Теперь вы — посвящённый Байесовского заговора.

1. Ward Casscells, Arno Schoenberger, and Thomas Graboys, «Interpretation by Physicians of Clinical Laboratory Results», New England Journal of Medicine 299 (1978): 999-1001.
2. David M. Eddy, «Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities», in Judgement Under Uncertainty: Heuristics and Biases, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (Cambridge University Press, 1982).
3. Gerd Gigerenzer and Ulrich Hoffrage, «How to Improve Bayesian Reasoning without Instruction: Frequency Formats», Psychological Review 102 (1995): 684-704.
4. Там же.
5. Edwin T. Jaynes, «Probability Theory, with Applications in Science and Engineering», Unpublished manuscript (1974).

Ссылка на оригинал: [An Intuitive Explanation of Bayes's Theorem](#)<sup>Ⓔ</sup>

Перевод: sepremento, Alaric

Ссылка на оригинал на [readthesequences.com](#):

[An Intuitive Explanation of Bayes's Theorem](#)<sup>Ⓔ</sup>

Оцените качество перевода:

★★★★★  
Средняя оценка: 4.4 (7 votes)



[← Когда слова ошибочны](#)

[Вверх](#)

[Просто реальность >](#)

[Версия для печати](#) [Добавить комментарий](#) [Обсуждение](#)

- [Короткая ссылка сюда: lesswrong.ru/642](#)