

# ПРИКЛАДНАЯ СТАТИСТИКА



## К чёрту условности!\*

22 февраля 2021 г.

Ты здоров на 91%, когда все вокруг говорят, что ты болен на 99%

### 1. Условные вероятности

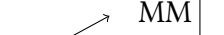
Поговорим о гороскопах. У тёти Маши — двое детей разных возрастов. Вероятности рождения мальчика и девочки равны.

- а. Какова вероятность того, что у тёти Маши оба ребёнка — мальчики, если известно, что у неё хотя бы один ребёнок — мальчик?
- б. Какова вероятность того, что у тёти Маши оба ребёнка — мальчики, если известно, что у неё хотя бы один ребёнок — мальчик да ещё и стрелец по гороскопу?

Давайте вспомним семинары по терверу. Скорее всего, для таких задач, читатель обычно рисовал дерево. Не будем далеко ходить и сделаем то же самое.

---

\*Эта pdf-ка, по факту, представляет из себя кусочек недописанной виньетки по Байесовским методам:  
[https://github.com/FUlyankin/book\\_about\\_bayes](https://github.com/FUlyankin/book_about_bayes)



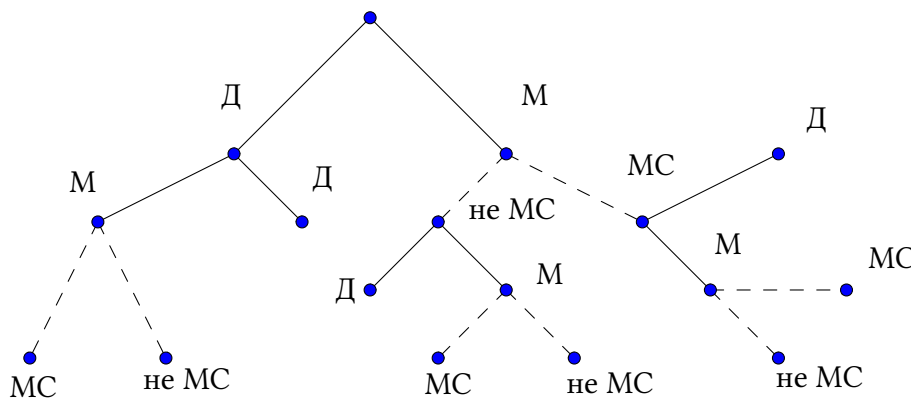
ММ	0.25	$1/3$
МД	0.25	$1/3$
ДМ	0.25	$1/3$
ДД	0.25	0

До поступления информации о том, что в семье имеется один мальчик, вероятности получить каждую пару детей были одинаковыми. После того, как стало известно, что хотя бы один ребёнок является мальчиком, вероятность исхода ДД занулилась. Остальные вероятности, в связи с поступлением новой информации, должны как-то поменяться, чтобы в сумме давать единицу. Поскольку, как мальчик, так и девочка в качестве второго ребёнка равновероятны, то они возрастают пропорционально до  $\frac{1}{3}$ . Условие, словно садовник, срезает одну из ветвей дерева.

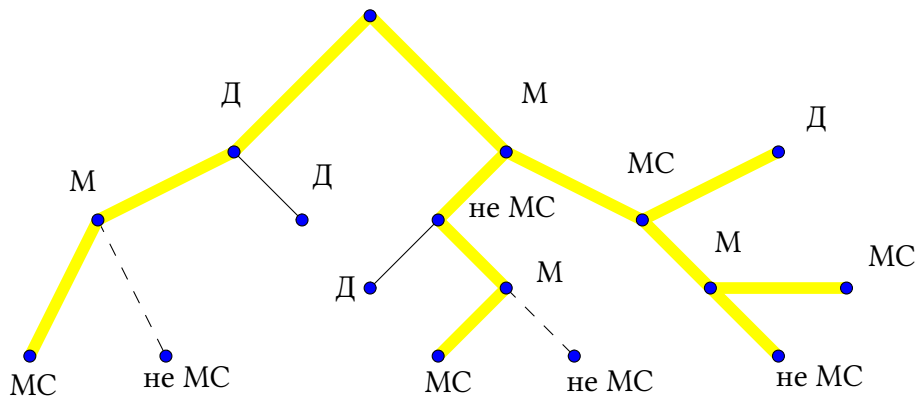
Можно посчитать всё это по-честному. Вспомним формулу условной вероятности

$$\mathbb{P}(\text{ММ} \mid \text{Хотя бы один М}) = \frac{\mathbb{P}(\text{ММ}, \text{Хотя бы один М})}{\mathbb{P}(\text{Хотя бы один М})} = \frac{0.25}{1 - 0.25} = \frac{1}{3}.$$

Хорошо. Теперь найдём вероятность того, что у тёти Маши два мальчика при условии, что один из них стрелец. Для этого снова нарисуем дерево.



Новое дерево получилось развесистым. Развесистости добавляют пунктирные ветки, отвечающие за гороскоп. По ветке МС мы проходим с вероятностью  $\frac{1}{12}$ . По обычным, не пунктирным, веткам мы проходим с вероятностью  $\frac{1}{2}$ . У тёти Маши есть один мальчик, и он стрелец. Это означает, что мы хотя бы разочек должны пройти по пунктирной ветке с надписью МС. Это ограничение состригает с дерева некоторые ветки и даёт нам пять траекторий.



Просуммируем вероятности пройти по каждой из траекторий и получим вероятность того, что у нас родится хотя бы один мальчик стрелец

$$\mathbb{P}(\text{хотя бы один МС}) = 0.5 \cdot 0.5 \cdot \frac{1}{12} + 0.5 \cdot \left( \frac{11}{12} \cdot 0.5 \cdot \frac{1}{12} + \frac{1}{12} \cdot \left( 0.5 + 0.5 \cdot \left( \frac{1}{12} + \frac{11}{12} \right) \right) \right) \approx 0.0815.$$

Фактически, мы сейчас использовали формулу полной вероятности, чтобы найти вероятность события «хотя бы один МС», просуммировав его появление при всех возможных гипотезах: МС и Д, МС и не МС, МС и Д, МС и МС, МС и не МС.

Из рассмотренных пяти траекторий только три приводят нас в лист, где пересекаются события хотя бы один МС и ММ. Сумма вероятностей прохода по этим траекториям составит 0.0399. Остаётся только применить формулу условной вероятности:

$$\mathbb{P}(\text{ММ} \mid \text{Хотя бы один МС}) = \frac{\mathbb{P}(\text{ММ}, \text{Хотя бы один МС})}{\mathbb{P}(\text{Хотя бы один МС})} = \frac{0.0399}{0.0815} \approx 0.489.$$

В конечном итоге получаем что-то близкое к  $\frac{1}{2}$ . Удивительно, но информация о том, что мальчик, родившийся у тёти Маши — стрелец, существенно изменяет вероятность того, что родятся два мальчика. Появление на свет стрельца — это настолько редкое событие, что для его наступления надо нарожать много мальчиков. Событие ММ происходит гораздо чаще, чем ММС.

Интуитивно объяснить полученный результат можно, выбросив из рассмотрения событие, при котором рождается два мальчика стрельца. Оно редкое, зачем обращать на него внимание?

$$\begin{array}{ccccccc} \text{МД} & \text{ДМ} & \underline{\text{ММ}} & \Rightarrow & \frac{1}{3} \\ \text{МСД} & \text{ДМС} & \cancel{\text{МСМС}} & \underline{\text{МС НМС}} & \underline{\text{МНС МС}} & \Rightarrow & \frac{1}{2} \end{array}$$

После удаления мелкого варианта как раз и получится наша  $\frac{1}{2}$ . Мораль первого чуда состоит в том, что дополнительная информация стрижет дерево и перераспределяет вероятностную массу между вариантами. Иногда перераспределение выходит очень неожиданным.

На самом деле, мы с вами только что использовали формулу Байеса. Давайте чуть более формально вспомним откуда она берётся. **Совместная вероятность** — это вероятность одновременного наступления двух событий  $P(A, B)$ . События называются **независимыми**, если вероятность совместного наступления событий равна произведению их вероятностей:

$$P(A, B) = P(A) \cdot P(B).$$

**Условная вероятность** — это вероятность наступления одного события, если известно, что произошло другое,  $P(A | B)$ . Она определяется как

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

Внимательно приглядевшись к формуле, можно заметить, что для независимых событий  $P(A | B) = P(A)$ . Если немного переписать условную вероятность, можно получить, что:

$$P(A, B) = P(A | B) \cdot P(B) = P(B | A) \cdot P(A),$$

из чего и вытекает формула Байеса:

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

Эта формула позволяет переоценивать наши текущие представления о мире (в формуле это  $P(B)$ ), используя новую, поступившую к нам информацию (в формуле это  $P(A | B)$ ), в качестве переоценки получая новое состояние наших представлений,  $P(B | A)$ .

Другой задачкой, которая иллюстрирует идею переоценки наших знаний, является медицинская диагностика.

## Упражнение 1 (Доктор, что со мной?)

Предположим, что по планете путешествует чрезвычайно смертоносное заболевание. И имя ему ковионавивус. Им заражается один человек из 10000. Медицинская диагностика очень хорошо

развита и люди умеют находить это заболевание с высокой точностью, они обнаруживают его в 99% случаев. Когда заболевания у человека нет, вероятность того, что его найдут составляет 1%. Иначе говоря, 1% это вероятность как ошибки первого рода (анализы нашли болезнь, когда её нет), так и ошибки второго рода (анализы не нашли болезнь, когда она есть).

Случилось страшное! Обследование показало, что человек болен. Какова вероятность того, что болезни у него на самом деле нет?

## Решение:

Можно попробовать нарисовать дерево и посмотреть как информация о диагнозе состригает лишние ветки, а мы, на этот раз, посчитаем всё по формулам.

Найдём вероятность того, что у человека найдут болезнь. Это может произойти, если он реально болен и если он не болен, но произошла ошибка. Пусть событие  $A$  означает, что болезнь обнаружили, а событие  $B$  означает, что человек реально болен. Тогда событие  $B^c$  будет означать, что у человека нет болезни<sup>1</sup>.

$$\mathbb{P}(A) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B) + \mathbb{P}(A \mid B^c) \cdot \mathbb{P}(B^c) = 0.99 \cdot \frac{1}{10000} + 0.01 \cdot \frac{9999}{10000} \approx 0.01.$$

Обратите ещё раз внимание на то, что мы делаем. Мы ищем вероятность того, что человек болен при всех возможных гипотезах. Так как гипотез дискретное число, мы суммируем вероятности, соответствующие им. В непрерывном случае мы будем поступать похожим образом. Интегрируя плотность  $f(y \mid x)$  по всем возможным значениям, которые принимает случайная величина  $X$ , мы будем получать плотность  $f(y)$ . Интеграл это то же самое, что обычная сумма, но для непрерывного случая.

В конечном итоге, используя формулу байеса, мы можем пересчитать вероятность того, что человек реально болен, при условии, что у него нашли болезнь:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)} = \frac{0.99 \cdot \frac{1}{10000}}{0.01} \approx 0.01.$$

Читатель может попробовать нарисовать дерево и посмотреть как информация о диагнозе состригает лишние ветки, а мы, на этот раз, посчитаем всё по формулам.

Найдём вероятность того, что у человека найдут болезнь. Это может произойти, если он реально болен и если он не болен, но произошла ошибка. Пусть событие  $A$  означает, что болезнь обнаружили, а событие  $B$  означает, что человек реально болен. Тогда событие  $B^c$  будет означать, что у человека нет болезни<sup>2</sup>.

<sup>1</sup>Обычно символом  $B^c$  обозначают событие, которое является дополнением к событию  $B$ . Для таких событий  $\mathbb{P}(B) + \mathbb{P}(B^c) = 1$ .

<sup>2</sup>Обычно символом  $B^c$  обозначают событие, которое является дополнением к событию  $B$ . Для таких событий  $\mathbb{P}(B) + \mathbb{P}(B^c) = 1$ .

$$\mathbb{P}(A) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B) + \mathbb{P}(A \mid B^c) \cdot \mathbb{P}(B^c) = 0.99 \cdot \frac{1}{10000} + 0.01 \cdot \frac{9999}{10000} \approx 0.01.$$

Обратите ещё раз внимание на то, что мы делаем. Мы ищем вероятность того, что человек болен при всех возможных гипотезах. Так как гипотез дискретное число, мы суммируем вероятности, соответствующие им. В непрерывном случае мы будем поступать похожим образом. Интегрируя плотность  $f(y \mid x)$  по всем возможным значениям, которые принимает случайная величина  $X$ , мы будем получать плотность  $f(y)$ . Интеграл это то же самое, что обычная сумма, но для непрерывного случая.

В конечном итоге, используя формулу байеса, мы можем пересчитать вероятность того, что человек реально болен, при условии, что у него нашли болезнь:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)} = \frac{0.99 \cdot \frac{1}{10000}}{0.01} \approx 0.01.$$

Это что же — воскликнет возмущённый читатель, — теперь не верить врачам и лечиться средствами традиционной медицины? Нет. Врачам нужно продолжать доверять, но всегда надо иметь в виду, что в случае, если болезнь довольно слабо распространена, точность диагноза должна быть не 99%, а гораздо выше.

Задачку про медицину и болезнь можно встретить почти во всех пособиях и лекциях по байсовским методам. Она является неплохим примером обратной задачи теории вероятностей.

**Прямые задачи** возникают, когда дано описание некоторого вероятностного процесса, и мы должны найти вероятность того или иного события, то есть по модели предсказать поведение. Например, перед нами две урны, в каждой по пять шаров, но известно, что в одной три красных, а в другой два. Остальные — чёрные. Мы не видели содержимое урн и вытаскиваем из одной из них шар. Какова вероятность выбрать красный?

**Обратные задачи**, наоборот, по известному результату, просят построить вероятностную модель. Например, известно, что мы вытащили красный шар. Насколько вероятно, что мы вытащили его из первой урны? До эксперимента мы не различали урны между собой и могли сказать лишь то, что в каждой из них может оказаться по три красных шара с вероятностью  $\frac{1}{2}$ . После того, как мы извлекли шар, наши изначальные представления должны пересчитаться в соответствии с новой информацией.

Неудивительно, что формула Байеса пронизывает насквозь всю современную статистику. Слишком часто в ней нам приходится по уже имеющимся данным восстанавливать неизвестную вероятностную модель. **Условная вероятность — садовник, подстригающий деревья, а болеть — плохо.**

## 2. Условные плотности

Сформулируем те дискретные формулы, которые мы использовали выше, для непрерывного случая. Для этого возьмём и заменим все вероятности на плотности.

Случайные величины  $X$  и  $Y$  независимы, если их совместная плотность распределения  $f(x, y)$  равна произведению частных плотностей  $f(x) \cdot f(y)$ . По-другому частные плотности иногда называют **маргинальными**.

Если случайные величины  $X$  и  $Y$  зависимы, их совместная плотность может быть представлена через произведение условного распределения одной случайной величины и безусловного другой. При этом сделать это можно целыми двумя способами

$$f(x, y) = f(x | y) \cdot f(y) = f(y | x) \cdot f(x). \quad (1)$$

Если у нас есть целых три случайных величины,  $X$ ,  $Y$  и  $Z$ , то начинает работать цепное правило

$$f(x, y, z) = f(x | y, z) \cdot f(y | z) \cdot f(z).$$

Коль скоро случайных величин у нас три, совместную плотность можно представить через произведение шестью способами: первую случайную величину выбираем тремя способами, вторую двумя, третью одним.

Найти условную плотность можно, посмотрев на всё это дело с обратной стороны

$$f(x | y) = \frac{f(x, y)}{f(y)}. \quad (2)$$

Значения, которые принимает случайная величина  $Y$  влияют на то, как именно распределена случайная величина  $X$ . Чуть ниже мы посмотрим на это на конкретных примерах.

Важно не забывать, что при этом  $\int f(x | y) dx = 1$ , так как это плотность распределения случайной величины  $X$ , но при этом  $\int f(x | y) dy$  единице равняться не обязан.

Случайную величину  $X$  можно очистить от зависимости и получить из условной функции распределения,  $f(x | y)$ , безусловную,  $f(x)$ . Такая процедура называется умным словом **маргинализация**. В дискретном случае мы это делали с помощью формулы полной вероятности. Пусть, например, у нас есть совместное распределение для двух дискретных случайных величин

	$Y = -1$	$Y = 1$
$X = -1$	0.4	0.3
$X = 1$	0.2	0.1

Случайная величина  $Y$  может принять значение  $-1$  в двух ситуациях:  $X = -1$  и  $X = 1$ . Чтобы найти вероятность  $\mathbb{P}(Y = 1)$ , нам нужно сложить вероятности в первых двух строках. Такая операция будет эквивалентна использованию формулы полной вероятности. Мы найдём вероятность того, что  $Y = 1$ , рассмотрев все наши гипотезы относительно величины  $X$

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = 1 \mid X = 1) \cdot \mathbb{P}(X = 1) + \mathbb{P}(Y = 1 \mid X = -1) \cdot \mathbb{P}(X = -1);$$

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = 1, X = 1) + \mathbb{P}(Y = 1, X = -1) = 0.4 + 0.2 = 0.6.$$

В итоге получится маргинальное распределение для  $Y$ :

$Y$	$-1$	$1$
$\mathbb{P}$	$0.6$	$0.4$

Если бы случайная величина  $X$  принимала бы  $n$  значений, то сумма была бы покрупнее:

$$\mathbb{P}(Y = 1) = \sum_{i=1}^n \mathbb{P}(Y = 1 \mid X = i) \cdot \mathbb{P}(X = i).$$

В случае непрерывной случайной величины надо поступить точно также. Правда сумму заменит её непрерывный аналог, интеграл. Возьмём совместную плотность распределения и выинтегрируем из неё все значения  $X$ :

$$f(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \int_{-\infty}^{+\infty} f(y \mid x) \cdot f(x) dx = \mathbb{E}_X f(y \mid x) \quad (3)$$

Дальше мы будем довольно часто использовать такой приём для очистки случайной величины от зависимости. Обратите внимание на то, что фактически мы ищем математическое ожидание от  $f(y \mid x)$ , то есть как-бы усредняем условную плотность по всем значениям случайной величины  $X$ .

Если разложить в нашей новой формуле (2)  $f(x, y)$  по правилу (1), то легко можно перейти от одной условной плотности к другой

$$f(x \mid y) = \frac{f(y \mid x) \cdot f(x)}{f(y)}.$$

Если в добавок ещё и вспомнить о формуле полной вероятности (3), то можно получить формулу Байеса:

$$f(x \mid y) = \frac{f(y \mid x) \cdot f(x)}{\int_{-\infty}^{+\infty} f(y \mid x) \cdot f(x) dx}.$$



Выше мы уже использовали эту формулу для переоценки вероятности гипотез после наступления какого-то события. Однако внизу вместо интеграла стояла сумма, полученная по формуле полной вероятности. **Главная проблема непрерывного аналога — интеграл в знаменателе. Очень часто он не берётся.** Более того, если мы ищем совместное распределение на вектор параметров, этот интеграл оказывается многомерным. С этим явлением приходится бороться с помощью компьютеров.

Давайте посмотрим как работают все эти замечательный формулы на примере нормально-го распределения. Оно довольно часто будет встречаться нам по ходу книги. Уже сейчас нужно начать привыкать к его громоздкости. В общем виде для двумерного нормального распределения

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left[ \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X & \rho_{XY} \\ \rho_{XY} & \sigma_Y \end{pmatrix} \right]$$

плотность будет иметь вид

$$f(x, y) = \frac{1}{2\pi\sigma_X^2\sigma_Y^2\sqrt{1-\rho_{XY}^2}} \cdot e^{-\frac{1}{2 \cdot (1-\rho_{XY}^2)} \cdot \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{XY} \cdot \frac{x-\mu_X}{\sigma_X} \cdot \frac{y-\mu_Y}{\sigma_Y} + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]}.$$

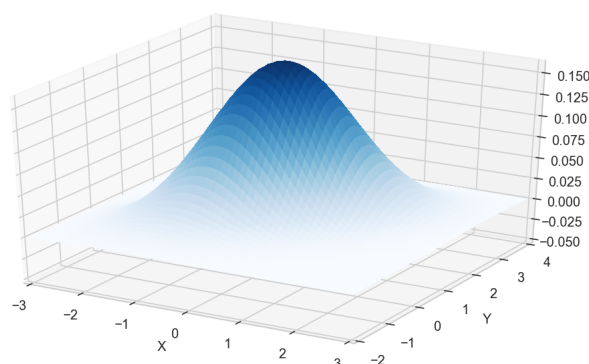
Давайте для простоты рассмотрим какой-нибудь конкретный случай. Например,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix} \right]$$

с плотностью

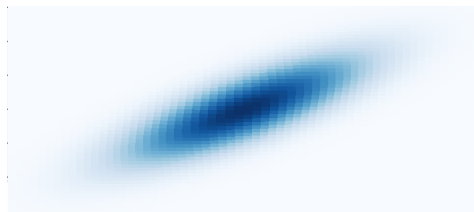
$$f(x, y) = \frac{1}{4\pi\sqrt{0.75}} \cdot e^{-\frac{1}{2 \cdot 0.75} \cdot \left[ x^2 - x \cdot \frac{y-1}{2} + \left( \frac{y-1}{2} \right)^2 \right]}$$

Из-за того, что случайная величина у нас двумерная, её плотность распределения рисуется в трёхмерном пространстве.



Плотность двумерного нормального распределения — это одеяло, зависшее в воздухе. При этом, под этим одеялом, по свойствам плотности распределения, содержится объём равный единице. Если ваш объём меньше единицы, то вы можете залезть под одеяло.

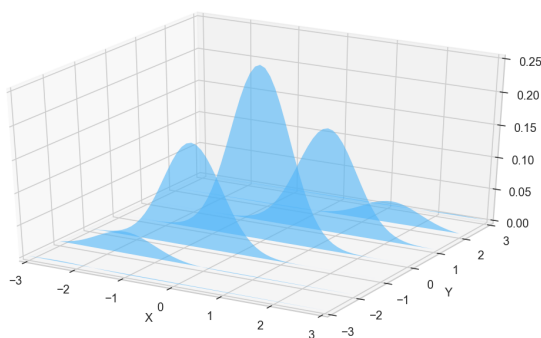
Случайные величины  $X$  и  $Y$  зависимы,  $\text{Corr}(X, Y) = 0.5$ . Чем большие значения принимает  $X$ , тем, в среднем, большие значения принимает  $Y$ . Из-за этого их совместная плотность будет вытянута вдоль одной из осей. Если взглянуть на одеяло сверху, можно заметить это.



От этой зависимости можно избавиться, и найти маргинальную плотность для  $X$ . Воспользуемся для этого формулой полной вероятности:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \\ &= \int_{-\infty}^{+\infty} \frac{1}{4\pi\sqrt{0.75}} \cdot e^{-\frac{1}{2 \cdot 0.75} \cdot \left[ x^2 - x \cdot \frac{y-1}{2} + \left( \frac{y-1}{2} \right)^2 \right]} dy = \\ &= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}. \end{aligned}$$

Когда мы выинтегрировали из совместной плотности  $y$ , мы получили  $N(0, 2^2)$ . Вполне ожидаемый результат. Если зафиксировать случайную величину  $Y$  и сделать по зафиксированному значению срез, можно получить условную плотность для случайной величины  $X$ . Внешний вид, а также характеристики новой плотности будут зависеть от того, где именно мы произвели срез. Если мы это сделали в районе пика,  $X | Y = 0$ , новоиспечённая случайная величина будет обладать нулевым математическим ожиданием. По мере отдаления среза от пика математическое ожидание будет сдвигаться.



В формуле условной плотности можно увидеть, как именно характеристики условного распределения зависят от точки среза

$$f_X(x | Y = y) = \frac{f(x, y)}{f_Y(y)} = \frac{1}{\sqrt{2 \cdot 0.75 \cdot \pi}} e^{-\frac{1}{2 \cdot 0.75} \cdot (x - 0.25 \cdot (y-1))^2}$$

Если посчитать условную плотность для общего случая, можно получить, что

$$X | Y = y \sim N \left( \mu_X + \rho \cdot \frac{\sigma_X}{\sigma_Y} \cdot (y - \mu_Y), \quad \sigma_X \cdot \sqrt{1 - \rho_{XY}^2} \right).$$

В этой формуле видно, как математическое ожидание зависит от точки среза. Дисперсия зависит только от корреляции. То есть при всех срезах она одинаковая. На картинке выше это не наблюдается, так как мы произвели срез, но не сделали нормировку площади под ними к единице. После неё плотности, в плане разброса, будут выглядеть абсолютно одинаково.

## Упражнение 1 (Азартный Жора)

Жорик взял случайную величину  $X$  с плотностью распределения

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{иначе} \end{cases}$$

и как следует подкинул её. Выпавшее число  $x$  Виталик использовал для генерации равномерной случайной величины  $Y \sim U[-x, x]$ . Найдите:  $f_{Y|X}(y)$ ,  $f_{XY}(x, y)$ ,  $f_Y(y)$ ,  $f_{X|Y}(x)$ .

**Решение:**

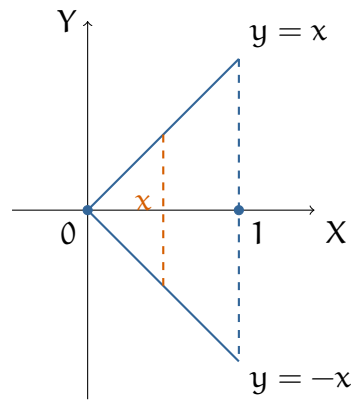
**Делай раз!** Распределение случайной величины  $Y$ , при условии, что  $X$  приняла значение  $x$  будет равномерным. Плотность будет зависеть от того какое именно значение приняла случайная величина  $X$ :

$$f_{Y|X}(y) = \frac{1}{2x}.$$

**Делай два!** Воспользуемся формулой условной вероятности и найдём совместное распределение случайных величин

$$f_{XY}(x, y) = f_X(x) \cdot f_{Y|X}(y) = \begin{cases} 2x \cdot \frac{1}{2x} = 1, & \text{если } 0 \leq x \leq 1 \text{ и } -x \leq y \leq x \\ 0, & \text{иначе.} \end{cases}$$

**Делай три!** Из совместной плотности распределения можно найти  $f_Y(y)$ , выинтегрировав из неё  $X$ . Чтобы грамотно сделать это, нужно разобраться в структуре двумерной случайной величины.



Все свои значения случайная величина принимает внутри треугольника. За его пределами она зануляется. Обратите внимание на красную пунктирную линию. Когда виталик подкинул  $X$  и она зафиксировалась,  $Y$  будет варьироваться только в пределах этой красной линии. Чтобы выинтыгрировать  $X$  из совместной плотности, мы должны пройти по всем подобным красным линиям и очистить от них случайную величину  $Y$ . Интеграл будет разбиваться на два кусочка. В первом случае  $y > 0$  и  $x$  бежит от  $x = y$  до 1. Во втором случае  $y < 0$  и  $x$  бежит от  $x = -y$  до 1.

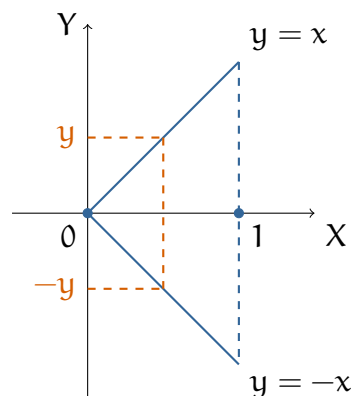
$$f_Y(y) = \begin{cases} \int_{-y}^1 1 \, dx = 1 + y, & -1 \leq y \leq 0 \\ \int_y^1 1 \, dx = 1 - y, & 0 \leq y \leq 1 \\ 0, & \text{иначе} \end{cases}$$

Любители модуля могут переписать первые две строчки как  $1 - |y|$ . При этом  $x$  будет изменяться в диапазоне  $|y| \leq x \leq 1$ .

**Делай четыре!** По формуле Байеса найдём оставшуюся условную плотность.

$$f_{X|Y}(x) = \frac{f_{Y|X}(y) \cdot f_X(x)}{f_Y(y)} = \frac{\frac{1}{2x} \cdot 2x}{1 - |y|} = \frac{1}{1 - |y|}$$

Величину  $y$  мы при этом можем выбирать на отрезке от  $-1$  до  $1$ , а  $x$  будет изменяться от  $|y|$  до  $1$ . Попробуем на картинке понять почему так будет происходить.



Когда мы фиксируем  $y$ , у нас появляется красная пунктирная линия, которая срезает для случайной величины  $X$  все значения, находящиеся левее неё. При этом всё, что справа, случайная величина по-прежнему может принимать, и значение  $y$  в знаменателе плотности распределения регулирует то, как именно при фиксации  $y$  будет происходить перераспределение вероятностной массы в условной плотности.

### 3. Снова о формуле Байеса

Наше погружение в чертоги разума почти окончено. Осталось вытащить наружу последние два волшебных словечка и ещё пару раз переписать формулу Байеса. Первое слово — **априорный**. Априорными называются знания, которыми мы располагаем до эксперимента. Второе слово — **апостериорный**. Апостериорными называются знания, которыми мы обладаем после эксперимента. Новая информация уточняет наши априорные представления и трансформирует их в апостериорные с помощью формулы Байеса.

Теорема Байеса — это основной, центральный инструмент современной статистики, на котором держится огромное число рассуждений. Чтобы увидеть это, давайте напишем формулу Байеса в том виде, в котором будем использовать по ходу книги:

$$f(\beta \mid y_1, \dots, y_n) = \frac{f(\beta) \cdot f(y_1, \dots, y_n \mid \beta)}{f(y_1, \dots, y_n)}.$$

Здесь  $y_1, \dots, y_n$  — это данные, которые нам удалось собрать,  $\beta$  — это параметры модели, которые мы хотим оценить.

Плотность  $f(\beta)$  отражает наши априорные знания о параметре  $\beta$  (prior). Она является математической формализацией нашей интуиции и всех тех вещей, которые мы знали о  $\beta$  до проведения эксперимента.

Плотность  $f(\beta \mid y_1, \dots, y_n)$  — это то, что мы в итоге хотим найти, распределение вероятностей параметров модели после того, как мы приняли во внимание данные, апостериорная плотность распределения (posterior).

Функция  $f(y_1, \dots, y_n \mid \beta)$  называется правдоподобием (likelihood). Из курса матстата вы должны помнить, что это вероятность появления наших данных, при условии, что все параметры модели нам известны. Мы не будем здесь вдаваться в подробности метода максимального правдоподобия. В конце главы есть несколько задачек, которые помогут читателю освежить знания.

Плотность  $f(y_1, \dots, y_n)$  — это вероятность данных, усреднённая по всем гипотезам (evidence). Для того, чтобы найти её, часто приходится брать интегралы. Это самая страшная часть формулы Байеса.

В концептуальном виде теорему Байеса можно переписать следующим образом

$$\text{posterior} = \frac{\text{prior} \cdot \text{likelihood}}{\text{evidence}}.$$

Давайте ещё раз вспомним пресловутую задачу про шары и коробки. Перед нами стоит две коробки. В каждой по пять шаров. В одной из них три красных, в другой два. Мы не знаем в какой из них сколько. Наше незнание можно описать с помощью априорных вероятностей, сказав, что с вероятностью 0.5 в правой коробке два красных шара, и с вероятностью 0.5 три. Как только мы засунули руку в правую коробку и вытянули из неё красный шар, мы получили наблюдение  $y_1$ . Для него мы можем выписать правдоподобие.

После поступления дополнительной информации, мы пересчитываем вероятности начальных гипотез. В самом первом чуде, пересчитывая их, мы стригли деревья. Из-за этой стрижки, для того, чтобы в сумме вероятности по-прежнему давали единицу, нам нужно было делить их на полную вероятность события, то есть как раз на  $\text{evidence}$ . К счастью, самая сложная часть формулы представляет из себя нормировочную константу, и в ходе расчётов её очень часто можно игнорировать. Ничего не мешает нам вспомнить про нормировку только в самый последний момент.

**Формула Байеса конвертирует информацию, которую мы пронаблюдали и наше априорное представление о природном явлении в апостериорные представления.**