

Глава 1

Чудеса условной вероятности

Тут должна быть абсурдная цитата
про чудеса

автор цитаты

Во время изучения любой науки нам периодически приходится обращаться к своим чертогам разума и искать в них уже давно усвоенную, а после успешно забытую, информацию. Байесовский подход в этом плане не будет исключением. По мере его изучения нам придётся иногда залезать в самые глубины чертог и выковыривать из них кучу разных вероятностей и распределений. С каких-то понятий придётся сдуть довольно большой слой пыли. От моды и медианы даже придётся прогнать пауков.

В этой главе мы сделаем первый шаг по направлению к байесовскому подходу и откопаем несколько чудес, связанных с условной вероятностью. Изучая их, мы вспомним кучу разных формул. Глава ни в коей мере не претендует на краткое изложение курса по теории вероятностей. Она всего лишь поднимает некоторые аспекты, которые понадобятся нам для байесовских методов и служит небольшой напоминкой.

Мы искренне надеемся, что наш усердный читатель сначала сам постарается разобраться в чудесах главы и решить их, и только по-

сле будет читать готовое решение. В конце главы нужно попробовать решить ещё несколько упражнений. Если возникают какие-то затруднения, можно посмотреть готовые решения.

1.1 О чудесах, русалках и леших

Чудо номер 1. У тёти Маши — двое детей разных возрастов. Вероятности рождения мальчика и девочки равны.

- 1. Какова вероятность того, что у тёти Маши оба ребёнка — мальчики, если известно, что у неё хотя бы один ребёнок — мальчик?
- 2. 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 задачек, читатель обычно рисовал дерево. Не будем далеко ходить и сделаем то же самое.

		Хотя бы один М	
	ММ	0.25	$\frac{1}{3}$
	МД	0.25	$\frac{1}{3}$
	ДМ	0.25	$\frac{1}{3}$
	ДД	0.25	0

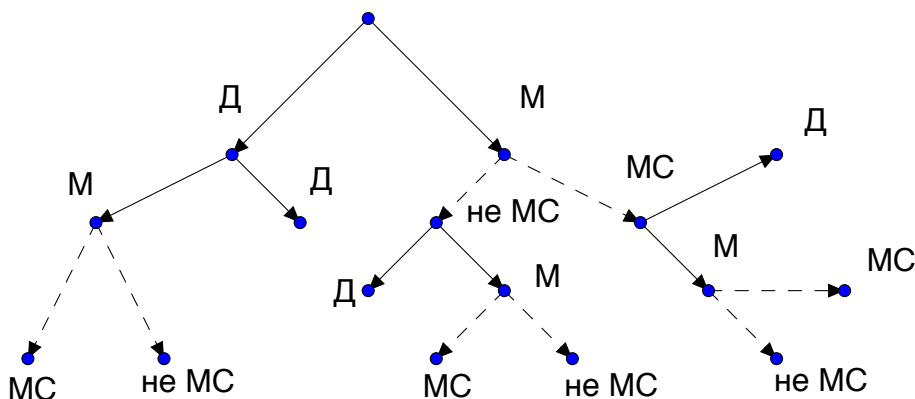
До поступления информации о том, что в семье имеется один мальчик, вероятности получить каждую пару детей были одинаковыми. После того, как стало известно, что хотя бы один ребёнок является мальчиком, вероятность исхода ДД занулилась. Остальные вероятности,

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.

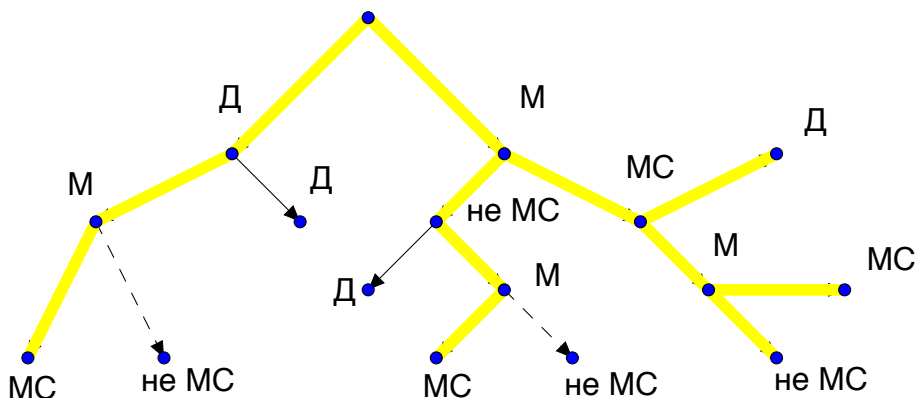
Можно посчитать всё это по-честному. Вспомним формулу условной вероятности

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

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

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

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. Остается только применить формулу условной вероятности:

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

В конечном итоге получаем что-то близкое к $\frac{1}{2}$. Удивительно, но информация о том, что мальчик, родившийся у тётки 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?

$$\text{МД} \quad \text{ДМ} \quad \underline{\text{ММ}} \Rightarrow \frac{1}{3}$$

$$\text{МСД} \quad \text{ДМС} \quad \cancel{\text{МСМС}} \quad \underline{\text{МС НМС}} \quad \underline{\text{МНС МС}} \Rightarrow \frac{1}{2}$$

После удаления мелкого варианта как раз и получится наша $\frac{1}{2}$. Мораль первого чуда состоит в том, что дополнительная информация стрищет 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, $P(B | A)$. Именно об этом мы и будем говорить в течение всей книги.

Другим чудом, иллюстрирующим идею переоценки наших знаний является чудо медицинской диагностики. Предположим, что по планете путешествует чрезвычайно смертоносное заболевание. 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? Напоминаем, что сначала надо попробовать решить задачу самому. 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 В означает, что человек реально болен. Тогда событие B^c будет означать, что у человека нет болезни¹.

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

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 $f(y | x)$ 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:

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

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.²

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

¹Обычно символом B^c обозначают событие, которое является дополнением к событию В. Для таких событий $P(B) + P(B^c) = 1$.

²Люди довольно плохо понимаю условные вероятности. Например,

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.

Другим классическим примером применения формулы Байеса, про который довольно часто спрашивают на собеседованиях, является «парадокс» Монти Холла (в кавычках, потому что тут никакого парадокса нет).

Чудо номер 2. Есть три закрытых двери. За двумя из них — по козе, за третьей автомобиль. Вы выбираете одну из дверей. Допустим, вы выбрали дверь номер один. Ведущий шоу, чтобы поддержать интригу, не открывает сразу выбранную вами дверь. Сначала он открывает одну из дверей, не выбранных вами, причем ради интриги ведущий не открывает сразу дверь с автомобилем. Допустим, ведущий открыл дверь номер два. И в этот момент он предлагает вам изменить ваш выбор двери. Имеет ли смысл изменить свой выбор?

Пусть $X_i = 1$, если за дверью i оказалась машина. За Y обозначим дверь, которую выбрал ведущий. Изначально

$$P(X_1 = 1) = P(X_1 = 2) = P(X_1 = 3) = \frac{1}{3}.$$

Если бы ведущий открыл одну из дверей наугад, то после этого события местонахождение машины было бы равновероятно распределено между оставшимися дверьми.

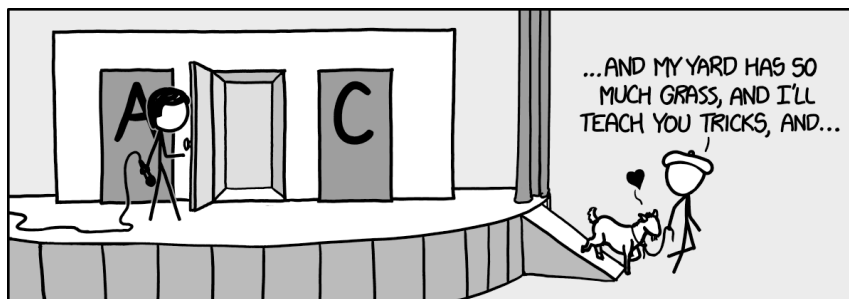


Рис. 1.1. A few minutes later, the goat from behind door C drives away in the car.
<http://xkcd.com/1282/>

Однако структура эксперимента устроена сложнее. Ведущий открывает дверь неслучайно. Он открывает одну из дверей, не выбрав игрок. Из-за этого возникает зависимость между событиями. Предположим, мы выбрали первую дверь. Так как мы можем нумеровать двери как захотим, эта предпосылка не повлияет на решение задачи и мы не потеряем общность. Также будем предполагать, что если машина находится за выбранной нами дверью, то одну из оставшихся дверей ведущий открывает равновероятно. Тогда

$$P(X_1 = 1, Y = 2) = P(Y = 2 | X_1 = 1) \cdot P(X_1 = 1) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6},$$

потому что машина за нашей дверью и ведущий выбирает одну из оставшихся дверей равновероятно. В свою очередь

$$P(X_2 = 1, Y = 3) = P(Y = 3 | X_2 = 1) \cdot P(X_2 = 1) = 1 \cdot \frac{1}{3} = \frac{1}{3},$$

потому что мы выбрали первую дверь, за ней коза. Машина лежит за второй дверью, ведущий не может её открыть, значит с вероятностью 1 он откроет третью дверь. Рассуждая по аналогии, можно найти все остальные вероятности:

$$P(X_1 = 1, Y = 2) = \frac{1}{6} \quad P(X_2 = 1, Y = 2) = 0 \quad P(X_3 = 1, Y = 2) = \frac{1}{3}$$

$$P(X_1 = 1, Y = 2) = \frac{1}{6} \quad P(X_2 = 1, Y = 2) = \frac{1}{3} \quad P(X_3 = 1, Y = 2) = 0.$$

Дело осталось за малым, вычислить по формуле Байеса интересующие нас вероятности. Вероятность того, что машина находится за нашей дверью составит

$$P(X_1 = 1 | Y = 2) = \frac{P(X_1 = 1, Y = 2)}{P(X_1 = 1, Y = 2) + P(X_2 = 1, Y = 2) + P(X_3 = 1, Y = 2)}.$$

Подставим все вероятности, найденные нами выше и получим ответ

$$P(X_1 = 1 | Y = 2) = \frac{1/6}{1/6 + 0 + 1/3} = \frac{1}{3}.$$

Вероятность того, что машина за другой дверью составит

$$P(X_1 = 0 | Y = 2) = \frac{P(X_1 = 0, Y = 2)}{P(X_1 = 1, Y = 2) + P(X_2 = 1, Y = 2) + P(X_3 = 1, Y = 2)}.$$

Получается, что

$$P(X_1 = 0 | Y = 2) = \frac{1/3}{1/6 + 0 + 1/3} = \frac{2}{3}.$$

Ведущий, открыв пустую дверь, посылает дополнительную информацию. И нам выгодно поменять выбранную дверь.

Чудо номер 3. Предположим, что вам предложили два конверта. Суммы в них относятся друг к другу как 1 : 2. Вам предлагают выбрать один из конвертов и открыть его. Вы открыли конверт и увидели в нём 100 рублей. У вас есть две возможности: либо забрать деньги с собой, либо поменять конверт на другой и забрать деньги из него. Как поступить в такой ситуации?

Так это же легко! Давайте просто найдём математическое ожидание, да и дело с концом. Пусть X_1 — сумма в первом конверте, а X_2 — сумма во втором конверте, тогда

$$E(X_2) = 0.5 \cdot 50 + 0.5 \cdot 200 = 125 > 100.$$

Надо менять. В среднем мы от этого выиграем больше. Секундочку. Если бы всё было бы так просто, здесь бы не оказалось этого чуда. Загвоздка состоит в том, что мы не можем так сделать. Мы не знаем настоящего закона распределения денег в конверте. Например, если бы в конверте могли бы изначально, до вытягивания лежать пары $(25, 50)$, $(50, 100)$, $(12.5, 25)$ с вероятностями $\frac{1}{3}$, то после вытягивания конверта с сотней рублей мы бы получили, что конверт менять не нужно

$$E(X_2 | X_1 = 100) = 50.$$

При вытягивании других конвертов мы бы получили бы ещё два математических ожидания:

$$E(X_2 | X_1 = 12.5) = 25,$$

$$E(X_2 | X_1 = 50) = 0.5 \cdot 25 + 0.5 \cdot 100 = 62.5.$$

В первой ситуации нужно поменять конверт, во второй нет. Получается, что при таком распределении иногда замена имеет смысл, а иногда нет. Давайте посмотрим на ещё одно распределение, например, вот такое:

$(3, 9)$	$(9, 27)$	$(27, 81)$...
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$...

Пусть в открытом конверте лежит 3 рубля. Получается, что

$$E(X_2 | X_1 = 3) = 9,$$

и нам нужно поменять конверт. Если в открытом конверте лежит 27 рублей, то $P(X_2 = 9 | X_1 = 27) = \frac{0.25}{0.25+0.125} = \frac{2}{3}$, то есть

$$E(X_2 | X_1 = 27) = \frac{2}{3} \cdot 9 + \frac{1}{3} \cdot 81 = 33 > 27.$$

И снова нам нужно поменять конверт. Для любого другого веса конверт также нужно будет поменять. Получается забавная ситуация. При таком распределении, какой бы конверт мы не открыли, мы должны его поменять. Попробуйте придумать распределение, при котором конверт невыгодно менять никогда.

Чудо номер 4. Спящая Красавица согласилась принять участие в научном эксперименте. В воскресенье её специально уколют веретеном. Как только она заснёт, будет подброшена правильная монетка.

Если монетка выпадет орлом, то Спящую Красавицу разбудят в понедельник и спросят о том, как выпала монетка и эксперимент завершится.

Если монетка выпадет решкой, то спящую Красавицу разбудят в понедельник, спросят о монетке, снова уколют веретеном, разбудят во вторник и снова спросят о монетке. Укол вызывает лёгкую амнезию, и Красавица не сможет определить просыпается ли она в первый раз или во второй. Внимание: монетка подкидывается только один раз!

Красавица только-только проснулась. Вспомнила правила эксперимента, и услышала вопрос исследователя: «Ваше высочество, так как же выпала монетка?»

1. Как следует ответить Красавице, если за каждый верный ответ ей дарят молодильное яблоко?
2. Как следует отвечать Красавице, если за неверный ответ её тут же превращают в тыкву навсегда?
3. Отвечая на вопрос исследователя, спящая Красавица задумалась, а какова вероятность того, что сегодня понедельник. А действительно, какова?

Помните о чём мы договаривались в самом начале? Сначала попробуйте разобраться с задачей сами, а уже потом изучайте решение.

Чтобы Красавице ответить на первые два вопроса, надо понять какова вероятность того, что монетка выпала орлом. Как не странно, эта вероятность равна $\frac{1}{2}$. То, что Красавица не помнит сколько раз она пробуждалась, никак не влияет на то какой стороной упадёт монетка. Эта вероятность всегда будет оставаться равной $\frac{1}{2}$.

Если за каждый правильный ответ Красавица получает молодильное яблоко, то Красавица может получить от нуля до двух яблок. Если выпал орёл, Красавице вопрос зададут только один раз. Если выпала решка, Красавице вопрос зададут дважды.

Ежели Красавица говорит, что выпал орёл, тогда она с вероятностью 0.5 получает одно яблоко. Ежели Красавица говорит, что выпала решка, тогда она с вероятностью 0.5 получает два яблока, так как вопрос ей задают целых два раза. Получается, что выгодно говорить, что выпала решка. Так мы заработаем большее ожидаемое число яблок.

Если за неправильный ответ Красавицу сразу же превращают в тыкву навсегда, то если мы говорим орёл, мы превращаемся в тыкву с вероятностью 0.5. Если решка, то также. Тыквой можно стать только один раз. Выходит, что нам всё-равно как отвечать на этот вопрос.

При разных функциях потерь Красавице выгодно действовать по-разному. Можно сказать, что Красавицы реагируют на стимулы. В будущем перед нами встанет вопрос: «А как прогнозировать?». Ответ на него зависит от введённой системы наказаний, то есть от функции потерь. В этом состоит главная мораль текущего чуда.

Вопрос «Какова вероятность того, что сегодня понедельник?» некорректен. На него нельзя дать ответ. Мы никак не можем выразить понедельник через события, происходящие в эксперименте. Монетка подбрасывается единожды. «Сегодня понедельник» — это не событие, а функция от времени. Тем не менее, мы можем решить как правильно отвечать на вопрос «Какой сегодня день недели?».

Пусть за угаданный день недели Красавица получает яблоки, тогда выгодно говорить, что сегодня понедельник. Если выпал орёл, Красавица получит одно яблоко, если выпала решка, она также получит одно яблоко. Одно яблоко ей гарантируется. Если Красавица будет утверждать, что сегодня вторник, она получит одно яблоко с вероятностью 0.5.

Пусть за неверно названный день недели Красавицу превращают в тыкву. Так как понедельник наступает раньше вторника, выгодно отвечать, что сегодня понедельник. Так Красавица с вероятностью 0.5 останется сама собой. Если она будет говорить, что сегодня вторник, то она превратится в тыкву почти наверное, то есть с вероятностью единица. При такой постановке вопроса Красавице выгодно давать один и тот же ответ.

Чудо номер 5. Поговорим о событиях и независимости: условной и безусловной. Пусть у нас есть две монетки: правильная и неправильная. Неправильная выпадает орлом с вероятностью $\frac{2}{3}$. Мы равновероятно выбираем одну из этих монеток и подбрасываем её дважды. Пусть событие A — орёл выпал в первый раз, событие B — орёл выпал во второй раз. Правда ли, что события A и B зависимые? Если да, то придумайте событие D такое, что события $A \mid D$ и $B \mid D$ независимы.

Если первое подбрасывание закончилось выпадением орла, у нас в голове рождается мысль: « A не обманывают ли нас?». Появление орла говорит нам, что, скорее всего, была взята неправильная монетка, а интуиция подсказывает, что события зависимы. Попробуем почестному посчитать условную вероятность и посмотреть совпадает ли она с безусловной.

Безусловную вероятность того, что в первый раз выпал орёл, можно найти по формуле полной вероятности:

$$P(A) = 1/2 \cdot 1/2 + 1/2 \cdot 2/3 = 7/12.$$

Условную вероятность можно найти по формуле условной вероятности

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/2 \cdot 1/4 + 1/2 \cdot 4/9}{1/2 \cdot 1/2 + 1/2 \cdot 2/3} \neq 7/12.$$

Видим, что эти вероятности не совпадают, значит события зависимые. Какое событие D сделает их независимыми?

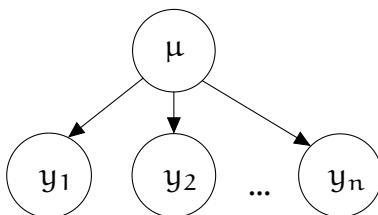
Если мы получим информацию о том, какая монетка была выбрана, события $A \mid D$ и $B \mid D$ станут независимыми. Чтобы убедиться в этом,

можно пересчитать вероятности и увидеть, что $P(A \mid B \cap D) = P(A \mid D) = \frac{2}{3}$. При получении дополнительной информации события стали независимы.

Для того, чтобы ещё больше прочувствовать всё это дело, представим себе, что Некто пришёл в ресторан, чтобы попробовать рыбу фугу. Если такая рыба была неправильно приготовлена, то от неё можно умереть. Пусть событие A — Некто ел рыбу, а событие B — Некто умер. Логично, что события A и B , если мы не располагаем никакой другой информацией, зависимы. Некто вполне мог скончаться из-за рыбы. При этом, если в наше распоряжение поступит информация о событии D — рыба была приготовлена правильно, то события $A \mid D$ и $B \mid D$ станут условно независимыми. В то же самое время, события $A \mid D^c$ и $B \mid D^c$, где D^c — отрицание события D , окажутся условно зависимыми. Если рыба была плохо приготовлена, то тогда смерть может быть связана с её поеданием. Новая информация творит чудеса!

В дальнейшем в наших руках будет оказываться куча разных случайных величин. Очень часто нам будет хотеться сделать их условно независимыми. Это будет довольно сильно упрощать расчёты.

Например, пусть в нашем распоряжении оказалось нормальное распределение с неизвестным математическим ожиданием μ и дисперсией 4. Также у нас в руках оказалась выборка из этого распределения y_1, \dots, y_n . Раньше мы смотрели на параметр μ как на константу и оценивали её методом максимального правдоподобия. В байесовских методах всё изменится. Параметр μ мы будем трактовать как неизвестную случайную величину. Зависимости между величинами будут устроены как-то вот так:



Наше знание о том, какое значение приняла случайная величина y_1 приносит нам кусочек информации о том, какое именно значение принимает скрытая от нас случайная величина μ . Эта информация, в свою очередь, подсказывает нам какое значение примет случайная величина y_2 .

То, что от значения величины μ зависят значения случайных величин y_1, \dots, y_n вполне логично. Это же выборка из распределения с математическим ожиданием μ . Получается, что случайные величины y_1, \dots, y_n зависимы, и по определению независимости

$$f(y_1, \dots, y_n) \neq f(y_1) \cdot \dots \cdot f(y_n).$$

В то же самое время, если бы мы узнали какое значение принимает случайная величина μ , все стрелочки в нашем графе взаимосвязей исчезли бы. Выпадение величины y_1 ничего бы не подсказывало нам о значении y_2 через значение μ , потому что оно и так известно, и величины $y_1 \mid \mu, \dots, y_n \mid \mu$ были бы условно независимыми, и мы могли бы написать, что

$$f(y_1, \dots, y_n \mid \mu) = f(y_1 \mid \mu) \cdot \dots \cdot f(y_n \mid \mu).$$

В будущем мы будем пользоваться такими финтами. Иногда мы будем придумывать скрытую от наших глаз, латентную, переменную, знание которой будет давать нам условную независимость, и упрощать вычисления.

В этом разделе мы рассмотрели несколько чудес. Иногда, в будущем, мы будем вспоминать про них, восхищаться ими и покорять новые вершины. Мы вспомнили как работать с событиями и дискретными случайными величинами. Наверное, мы никого не удивим, если скажем, что в случае непрерывных случайных величин работают те же самые правила, но формулируются они уже в терминах плотностей и интегралов. Об этом подробнее мы поговорим в следующем разделе. Сейчас же давайте закрепим мораль чудес:

- Условная вероятность — садовник, подстригающий деревья, а болеть — плохо.

- Дверь нужно поменять. Особенно, если в квартире лежат деньги, а ключи потеряны. Иначе деньги окажутся у кого-нибудь в конверте.
- Красавицы реагируют на стимулы.
- Новая информация делает многие вещи независимыми. Шерлок Холмс называл это дедукцией и использовал байесовские методы, чтобы раскрывать преступления, когда они ещё не были мейнстримом.

1.2 О том, что творится в чертогах разума

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**.

Ежели 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

$$f(x, y) = f(x | y) \cdot f(y) = f(y | x) \cdot f(x). \quad (1.1)$$

Ежели 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, третьей одним.

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

$$f(x | y) = \frac{f(x, y)}{f(y)}. \quad (1.2)$$

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 $P(Y = 1)$, 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

$$P(Y = 1) = P(Y = 1 | X = 1) \cdot P(X = 1) + P(Y = 1 | X = -1) \cdot P(X = -1);$$

$$P(Y = 1) = P(Y = 1, X = 1) + P(Y = 1, X = -1) = 0.4 + 0.2 = 0.6.$$

В итоге получится маргинальное распределение для Y :

Y	-1	1
P	0.6	0.4

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

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

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 :

$$f(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \int_{-\infty}^{+\infty} f(y | x) \cdot f(x) dx = E_X f(y | x) \quad (1.3)$$

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 $f(y | x)$, то есть как-бы усредняем условную плотность по всем значениям случайной величины X .

Если разложить в нашей новой формуле (1.2) $f(x, y)$ по правилу (1.1), то легко можно перейти от одной условной плотности к другой

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

Если в добавок ещё и вспомнить о формуле полной вероятности (1.3), то можно получить главную формулу этой книги, формулу Байеса:

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

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 нормального 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

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

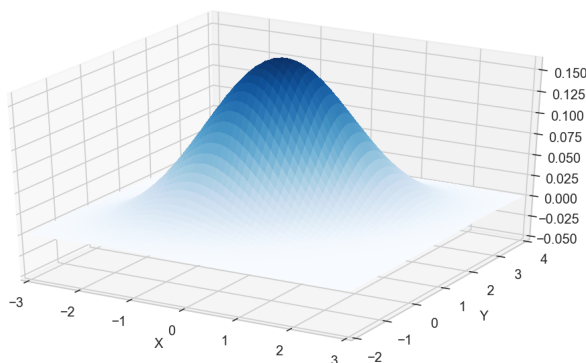
Давайте для простоты рассмотрим какой-нибудь конкретный случай. Например,

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

с плотностью

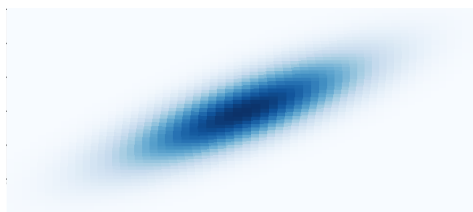
$$f(x, y) = \frac{1}{4\pi\sqrt{0.75}} \cdot e^{-\frac{1}{2 \cdot 0.75} \cdot \left[x^2 - x \cdot \frac{y-1}{2} + \left(\frac{y-1}{2} \right)^2 \right]}$$

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.

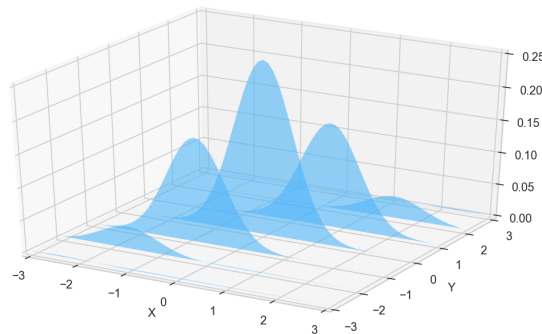
Случайные величины X и Y зависимы, $\text{Corr}(X, Y) = 0.5$. Чем большие значения принимает X , тем, в среднем, большие значения принимает Y . Из-за этого их совместная плотность будет вытянута вдоль одной из осей. Если взглянуть на одеяло сверху, можно заметить это.



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

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

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.

1.3 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 напомним формулу Байеса в том виде, в котором будем использовать по ходу книги:

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

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

Плотность $f(\beta)$ отражает наши априорные знания о параметре β (prior). Она является математической формализацией нашей интуи-

ции и всех тех вещей, которые мы знали о β 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 задачеч, 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 evidence. К счастью, самая сложная часть формулы представляет из себя нормировочную константу, и в ходе расчётов её очень часто можно игно-

рировать. 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. Эта формула наше последнее чудо.

Последнее? Скорее всего, когда вы изучали теорию вероятностей, вы решили много разных задач на пересчёт вероятностей с теми же самыми шарами. Ещё вам, наверное, говорили о том, что эту формулу можно использовать для пристрелки, если конечно же у вас есть пушка, которую можно калибровать. Потом вы решали ещё с десятков разнообразных задач, сдали контрольную работу и мир байеса для вас закончился. И теперь, когда ваши друзья разговаривают о чём-то, под названием «теорема байеса» или «байесовское мышление», у вас в голове всплывает одна единственная формула. И всё. Может быть, вы прекрасно понимаете это уравнение и можете его использовать для решения задач, но при этом вы не можете понять почему все вокруг интересуются какой-то одной случайной штукой из статистики и верят, что в ней заложен секрет вселенной. Какая, к чёрту, философия может стоять за одним маленьким уравнением?!

Садитесь в кресло. Закройте глаза. Расслабьтесь. Откройте глаза. Не удивляйтесь, когда увидите перед собой Морфеуса. Он протягивает вам руку и раскрывает кулак. На ладони лежит две таблетки — голубая и красная.

Выберете голубую, и история закончится прямо здесь, на чудесах условной вероятности, на паре концепций, которые вы и так знали и паре задачек, которые вы итак могли решить. Сразу после этого выбора имеет смысл захлопнуть книгу и даже не дочитывать этот абзац.

Выберете красную, и погрузитесь в новый для себя мир, почувствуете себя Алисой, падающей в кроличью нору. Теорема байеса повсюду. Она окружает нас. Даже сейчас она с нами рядом. Вы видите её, когда смотрите на восход солнца или подбрасываете монету. Вы ощущаете её, когда играете в покер, выбираете покупки в магазине и спорите о том, кто же победит на следующих выборах. Вы должны увидеть это сами. И раз уж вы дочитали этот абзац, у вас нет пути назад. Добро пожаловать в страну чудес. В страну, где, по определению, нет

последнего чуда. **И для начала, никаких формул. Только философия и идеи.**

1.4 Ещё задачи

В этом разделе приведены 10 задач для того, чтобы вспомнить базовые понятия из теории вероятностей, которыми мы будем оперировать в дальнейшем, а также для того, чтобы чуть лучше освоиться в работе с условными распределениями.

Упражнение 1. Представим себе такую ситуацию. На улице уже темно, вас всё ещё нет дома, мама пытается дозвониться до вас, но вы не берёте трубку. Мама начинает волноваться и представлять себе как на вас, на улице, нападают джентельмены с угрюмыми лицами и револьверами и требуют отдать все свои сбережения. Несколько часов переживаний, и читатель наконец приходите домой целый и невредимый. Переживания были пустыми, вы просто засиделись с друзьями в баре. Обоснованны ли переживания мамы с точки зрения формулы байеса?

Упражнение 2. Чёрная кошка одержима местью. Она разыскивает Джерома Валеску, чтобы убить его. Будучи близко к своей цели, она случайно нарвалась на Харли Квин, которая предложила ей следующую игру. Есть 6-патронный револьвер. В него вставлено два патрона подряд. Харли прокручивает барабан и стреляет себе в голову. Она остаётся в живых и отдаёт револьвер Кошке.

Что выгоднее сделать Кошке: сразу выстрелить или прокрутить заново барабан? Если Кошка после выбора оптимальной стратегии выживет, что выгодно сделать Харли Квин: прокрутить барабан или стрелять сразу?

Изначально возможно 6 различных исходов. Когда Харли пытается выпустить пулю себе в голову, она ничего не знает про текущее состояние ячеек в барабане. Патрон мог оказаться в дуле с вероятностью $\frac{2}{6}$. После выстрела, Кошка понимает, что одна пустая ячейка израсходована. Также она знает, что две пули в пушке идут друг за другом

подряд. Её интересует в какой именно комбинации в барабане расположена последовательность из двух пуль. Первая пуля может оказаться в одной из четырёх позиций. Вторая будет следовать за ней. Все возможные исходы приведены на картинке **??**. Красным крестом обозначено положение ячейки, которую уже использовала Харли. Сразу же за ней идёт ячейка, которая будет использована Кошкой. Получается, что Кошка умрёт с вероятностью $\frac{1}{4}$.

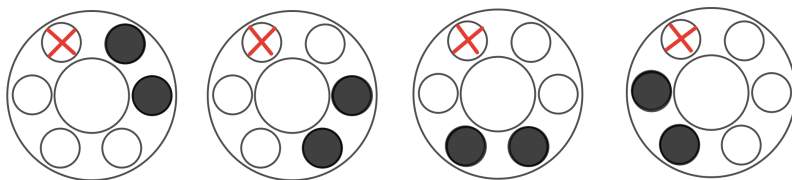


Рис. 1.2. Варианты барабана для кошки

Если Кошка прокрутит барабан, она анулирует всю информацию, накопленную в результате выстрела Харли. Будет непонятно находится ли в дуле патрон и Кошка умрёт с вероятностью $\frac{2}{6}$. Кошке выгодно сразу выстрелить себе в голову без прокрутки барабана. Так вероятность того, что она выживет выше.

Следующий ход делает Харли. Она располагает информацией о том, что израсходованы две пустые ячейки. Последовательность из двух пуль может оказаться в одной из трёх позиций. Харли умрёт с вероятностью $\frac{1}{3}$. При прокрутке барабана она умрёт ровно с такой же вероятностью. Для неё подойдёт любая стратегия. Если она выберет вариант не сбрасывать барабан, то к Кошке поступит информация о том, что было израсходовано слишком много пустых ячеек и барабан пора бы скинуть.

Упражнение 3. При полевых испытаниях имперские маркетологи выяснили, что вероятность бесперебойной работы Звезды Смерти при отсутствии негативных факторов составляет 0.99. При атаке повстанцев вероятность работы Звезды Смерти равна 0.95. При появлении

внеочередного последнего (крайнего) джедая вероятность работы Звезды Смерти равна 0.9. Вероятность работы Звезды Смерти при одновременном появлении крайнего джедая и атаке повстанцев равна 0.8. В далёкой-далёкой галактике повстанцы нападают с вероятностью 0.2, а крайний джедай появляется с вероятностью 0.1.

1. Какова вероятность аварийного отключения Звезды Смерти, если считать, что атаки повстанцев и возникновение крайнего джедая не зависят друг от друга?
2. Докажите, что вероятность аварийного отключения Звезды Смерти без предположения о том, что атаки повстанцев и возникновение джедая независимы, заключена в пределах от 0,027 до 0,033.

Пусть событие B_1 состоит в том, что возник крайний джедай. Событие B_2 состоит в том, что повстанцы напали. По условию первого пункта, эти события независимы. Событие A состоит в том, что Звезда отключилась. Её отключение может произойти при четырёх разных гипотезах:

$$\begin{array}{ll} H_1 = B_1^C \cap B_2^C & P(H_1) = 0.9 \cdot 0.8 \\ H_2 = B_1^C \cap B_2 & P(H_2) = 0.9 \cdot 0.2 \\ H_3 = B_1 \cap B_2^C & P(H_3) = 0.1 \cdot 0.8 \\ H_4 = B_1 \cap B_2 & P(H_4) = 0.1 \cdot 0.2 \end{array}$$

Полная вероятность отключения Звезды можно найти по формуле полной вероятности:

$$P(A) = 0.9 \cdot 0.8 \cdot 0.01 + 0.9 \cdot 0.2 \cdot 0.05 + 0.1 \cdot 0.8 \cdot 0.1 + 0.1 \cdot 0.2 \cdot 0.2 = 0.028$$

Теперь будем работать в парадигме, что между событиями может быть зависимость. Попробуем выразить через вероятность $x = P(B_2 | B_1)$ все оставшиеся. С первыми двумя вероятностями всё оказывается довольно просто.

$$P(B_1 \cap B_2) = P(B_1) \cdot P(B_2 | B_1) = 0.1 \cdot x$$

$$P(B_1 \cap B_2^C) = P(B_1) \cdot P(B_2^C | B_1) = P(B_1) \cdot (1 - P(B_2 | B_1)) = 0.1(1 - x)$$

Дальше нужно обратить внимание на то, что

$$P(B_2) = P(B_2 | B_1) \cdot P(B_1) + P(B_2 | B_1^C) \cdot P(B_1^C).$$

Из этого равенства можно выразить $P(B_2 | B_1^C)$ через всё остальное. Получится, что

$$P(B_1^C \cap B_2) = (1 - P(B_1)) \left(\frac{P(B_2) - P(B_2 | B_1^C) \cdot P(B_1^C)}{P(B_1)} \right) = 0.9 \cdot \frac{0.2 - 0.1x}{0.9}$$

По аналогии получаем, что

$$\begin{aligned} P(B_1^C \cap B_2^C) &= (1 - P(B_1)) \left(1 - \frac{P(B_2) - P(B_2 | B_1^C) \cdot P(B_1^C)}{P(B_1)} \right) = \\ &= 0.9 \cdot \left(1 - \frac{0.2 - 0.1x}{0.9} \right) \end{aligned}$$

Выпишем вероятность отключения Звезды Смерти:

$$\begin{aligned} P(A) &= 0.1 \cdot x \cdot 0.2 + 0.1 \cdot (1 - x) \cdot 0.1 + \\ &\quad + 0.9 \cdot \frac{0.2 - 0.1x}{0.9} \cdot 0.05 + 0.9 \cdot \left(1 - \frac{0.2 - 0.1x}{0.9} \right) \cdot 0.01 \end{aligned}$$

Мы договорились, что x это вероятность $P(B_2 | B_1)$. Она может принимать любые значения на отрезке $[0; 1]$. Подставив концы этого отрезка в формулу полной вероятности, мы можем получить её границы. Подставив ноль, получим 0.027. Подставив единицу получим 0.33. Выходит, что если события B_1 и B_2 зависимы, вероятность отключения Звезды Смерти лежит между 0.027 и 0.33. Если мы подставим в формулу $x = P(B_2 | B_1) = P(B_2)$, получим решение первого пункта.

Упражнение 4. Дан совместный закон распределения случайных величин X и Y :

	$Y = 0$	$Y = 1$
$X = 0$	$\frac{1}{5}$	$\frac{2}{5}$
$X = 1$	$\frac{2}{5}$	0

1. Найдите маргинальные распределения случайных величин X и Y .
2. Являются ли случайные величины X и Y независимыми?
3. Найдите условные распределения $X | Y = 0$ и $X | Y = 1$.
4. Пусть $Z = E(X | Y)$, найдите распределение этой случайной величины, найдите $E(Z)$.
1. Чтобы найти маргинальное распределение X , высуммируем Y и наоборот. Например,

$$\begin{aligned}
 P(X = 0) &= P(X = 0 | Y = 0) \cdot P(Y = 0) + P(X = 0 | Y = 1) \cdot P(Y = 1) = \\
 &= P(X = 0, Y = 0) + P(X = 0, Y = 1) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}.
 \end{aligned}$$

По факту все манипуляции сводятся к суммированию по столбцам для Y и суммированию по строкам для X .

Y	$Y = 0$	$Y = 1$
$P(Y = k)$	$\frac{3}{5}$	$\frac{2}{5}$

X	$X = 0$	$X = 1$
$P(X = k)$	$\frac{3}{5}$	$\frac{2}{5}$

2. Если для любых двух клеток $P(X = i) \cdot P(Y = j) = P(X = i, Y = j)$, случайные величины независимы. Если это хотя бы для одной клетки не так, они зависимы. Сразу же получаем, что они зависимы: $\frac{3}{5} \cdot \frac{3}{5} \neq \frac{1}{5}$.

3. По-честному найдём какую-нибудь одну вероятность для условного распределения

$$P(X = 0 \mid Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = \frac{1/5}{3/5} = \frac{1}{3}.$$

Остальные клетки можно найти по аналогии. На самом деле дополнительные условия обрезают у нашей изначальной таблицы определённые куски. Если мы работаем в рамках условия $Y = 0$, значит мы оказываемся в первом столбце. Вероятности, которые находятся в нём нужно отнормировать к единице. На выходе после такого среза получится условное распределение.

$X \mid Y = 0$	0	1
$P(\dots)$	$\frac{1}{3}$	$\frac{2}{3}$

$X \mid Y = 1$	0	1
$P(\dots)$	1	0

4. Последний шаг — условное математическое ожидание. Это тоже случайная величина. В зависимости от того, какое значение приняла величина Y , оно будет разным

$$E(X \mid Y = 0) = \frac{2}{3} \quad E(X \mid Y = 1) = 0$$

Конкретное значение Z зависит от того, как именно выпала Y , то есть

Z	$\frac{2}{3}$	0
$P(Z = k)$	$\frac{3}{5}$	$\frac{2}{5}$

Найти $E(Z)$ можно либо по-честному, либо вспомнив, что $E(E(X \mid Y)) = E(X) = \frac{2}{5}$.

Упражнение 5. Пусть число посетителей ресторана $N \sim \text{Poiss}(\lambda)$. Предположим, что каждый посетитель покупает напиток с вероятностью p независимо от остальных покупателей и их числа. Пусть X — число посетителей, которые купили напиток, а Y — число посетителей, которые не стали покупать напиток.

1. Найдите распределение $X | N$.
2. Найдите $E(X | N)$.
3. Найдите $E(X)$.
4. Найдите маргинальные распределения для X и Y .
5. Найдите совместное распределение величин X и Y . Являются ли случайные величины X и Y зависимыми?
6. Найдите $E(X^2 Y^2)$
1. Число людей в очереди фиксировано и равно N . Каждый может либо купить либо не купить водичку с вероятностью p , значит условное распределение будет биномиальным, то есть $X | N \sim \text{Bin}(p, N)$. Вероятность, того, что $X | N$ примет конкретное значение, можно найти по формуле Бернулли:

$$P(X | N = k) = C_N^k \cdot p^k \cdot (1 - p)^{N-k}$$

2. Для биномиального распределения $E(X | N) = N \cdot p$
3. По закону повторного ожидания

$$E(X) = E(E(X | N)) = E(N \cdot p) = \lambda \cdot p.$$

К этому же факту можно прийти немного иначе, просто «свернув» все условные математические ожидания по всем значениям случайной величины N

$$\begin{aligned} E(X) &= \sum_{n=0}^{\infty} E(X | N = n) \cdot P(N = n) = \\ &= \sum_{n=0}^{\infty} n \cdot p \cdot \frac{e^{-\lambda} \cdot \lambda^n}{n!} = p \cdot E(N) = p \cdot \lambda. \end{aligned}$$

4. Чтобы найти маргинальное распределение для X , нужно воспользоваться формулой полной вероятности и свернуть распределение $X | N$ по распределению N

$$\begin{aligned}
 P(X = k) &= \sum_{n=0}^{\infty} P(X = k | N = n) \cdot P(N = n) = \\
 &= \sum_{n=k}^{\infty} C_n^k p^k (1-p)^{n-k} \cdot \frac{e^{-\lambda} \cdot \lambda^n}{n!} = \sum_{n=k}^{\infty} \frac{p^k (1-p)^{n-k} e^{-\lambda} \lambda^n}{k! (n-k)!} = \\
 &= \frac{e^{-\lambda} \cdot (\lambda p)^k}{k!} \cdot \sum_{n=k}^{\infty} \frac{(\lambda \cdot (1-p))^{n-k}}{(n-k)!} = \frac{e^{-\lambda} \cdot (\lambda p)^k}{k!} \cdot e^{-\lambda(1-p)} = \\
 &= \frac{e^{-\lambda p} \cdot (\lambda p)^k}{k!}.
 \end{aligned}$$

Выходит, что $X \sim \text{Poiss}(\lambda p)$. По аналогии можно показать, что $Y \sim \text{Poiss}(\lambda(1-p))$.

5. Найдём совместное распределение случайных величин X и Y :

$$P(X = i, Y = j) = \sum_{n=0}^{\infty} P(X = i, Y = j | N = n)$$

Посмотрим чуть внимательнее на совместную вероятность. Мы знаем, что $X + Y = N$, так как человек в очереди либо купил водичку, либо не купил. Зная это немного упростим формулу

$$\begin{aligned}
 P(X = i, Y = j | N = n) &= P(X = i | N = i + j) \cdot P(N = i + j) = \\
 &= C_{i+j}^i \cdot p^i \cdot (1-p)^j \cdot e^{-\lambda} \cdot \frac{\lambda^{i+j}}{(i+j)!} = \\
 &= \frac{e^{-\lambda p} \cdot (\lambda p)^i}{i!} \cdot \frac{e^{-\lambda(1-p)} \cdot (\lambda(1-p))^j}{j!} = P(X = i) \cdot P(Y = j).
 \end{aligned}$$

При вычислении мы получили, что совместная вероятность для X и Y всегда равна произведению их частных вероятностей. Значит случайные величины независимы.

6. Зная, что случайные величины независимы, найти математическое ожидание довольно легко

$$E(X^2 Y^2) = E(X^2) \cdot E(Y^2) = (\lambda p + \lambda^2 p^2) \cdot (\lambda(1-p) + \lambda^2(1-p)^2).$$

При расчёте мы воспользовались тем, что $E(X^2) = \text{Var}(X) + (E(X))^2$.

Упражнение 6. Пусть $X \sim \text{Exp}(1)$.

1. Найдите условные плотность и функцию распределения случайной величины X при ограничении $X > 1$.
2. Найдите $E(X | X > 1)$ и $\text{Var}(X | X > 1)$.

Вспомним формулу условной вероятности

$$F_{X|X>1}(x) = P(X \leq x | X > 1) = \frac{P(X \leq x, X > 1)}{P(X > 1)}$$

Найдём все её частички

$$P(X > 1) = \int_1^{\infty} e^{-x} dx = \frac{1}{e}$$

$$P(X \leq x, X > 1) = \int_1^x e^{-x} dx = -(e^{-x} - e^{-1})$$

$$P(X \leq x | X > 1) = 1 - e^{-(x-1)}$$

Последний штрих!

$$f_{X|X>1}(x) = F'_{X|X>1}(x) = \begin{cases} e^{-x+1}, & x > 1 \\ 0, & \text{иначе.} \end{cases}$$

Зная плотность, легко найти, что $E(X | X > 1) = 2$, а $\text{Var}(X | X > 1) = 1$. На самом деле, такое распределение называется экспоненциальным распределением со сдвигом. Его дисперсия остаётся такой же, как и для обычного экспоненциального распределения, а математическое ожидание возрастает на величину сдвига.

Упражнение 7. Виталик 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: $f_{Y|X}(y)$, $f_{XY}(x, y)$, $f_Y(y)$, $f_{X|Y}(x)$.

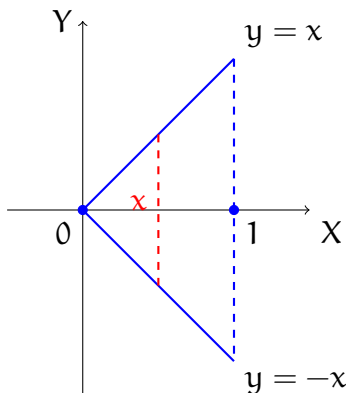
Делай раз! Распределение случайной величины Y , при условии, что X приняла значение x будет равномерным. Плотность будет зависеть от того какое именно значение приняла случайная величина X :

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

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.

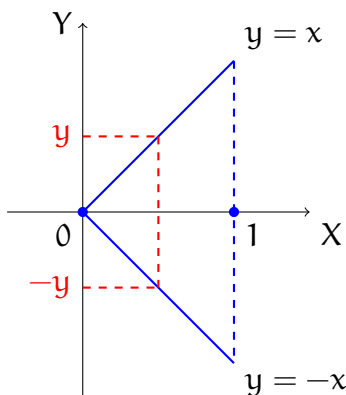
$$f_Y(y) = \begin{cases} \int_{-y}^1 1 \, dx = 1 + y, & -1 \leq y \leq 0 \\ \int_y^1 1 \, dx = 1 - y, & 0 \leq y \leq 1 \\ 0, & \text{иначе} \end{cases}$$

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 будет происходить перераспределение вероятностной массы в условной плотности.

Упражнение 8. Пусть $X \sim U[1, 2]$ и $Y | X = x \sim \text{Exp}(x)$. Найдите $E(Y)$ и $\text{Var}(Y)$.

Первый способ: вспомнить про полное математическое ожидание!

$$E(Y) = E(E(Y | X)) = E\left(\frac{1}{X}\right) = \int_1^2 \frac{1}{x} dx = \ln(2).$$

Второй путь по-честному свернуть условное математическое ожидание по всем значениям случайной величины X . По факту это ровно то же самое, что и применить формулу полного ожидания

$$E(Y) = \int_{-\infty}^{+\infty} E(Y | X = x) f_X(x) dx = \int_1^2 \frac{1}{x} dx = \ln(2).$$

Фактически, равенство, которое мы выписали, доказывает формулу $E(X) = E(E(X | Y))$, так как мы усредняем значения $E(X | Y)$ по плотности случайной величины X . По аналогии можно поступить

с $E(Y^2)$, а после по формуле $\text{Var}(Y) = E(Y^2) - (E(Y))^2$ выяснить, что $\text{Var}(Y) = 1 - (\ln(2))^2$.

Упражнение 9. Пусть мы знаем совместную плотность распределения трёх случайных величин $f(x, y, z)$. Пусть мы наблюдаем переменную X и хотим по ней спрогнозировать переменную Y . Переменная Z ненаблюдаема для нас. Как можно оценить $f(y | x)$, зная $f(x, y, z)$?

По формуле условной вероятности

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

Мы не знаем $f(x, y)$, но можем получить её из $f(x, y, z)$, выинтегрировав всё лишнее:

$$f(x, y) = \int f(x, y, z) dz.$$

По аналогии можно получить из $f(x, y, z)$ маргинальную плотность $f(x)$:

$$f(x) = \iint f(x, y, z) dz dy.$$

В итоговой формуле получается дробь из двух интегралов, один из которых двойной.

Упражнение 10. В течение книги мы часто будем вспоминать про метод максимального правдоподобия. Давайте вспомним как он работает!

1. Представим себе, что мы приехали на выходные в какой-нибудь жаркий южный город и пошли гулять по главной площади. На ней мы обнаружили фонтан. Он работает. Оцените с помощью метода максимального правдоподобия, как часто работает этот фонтан.
2. Немного чисто технических упражнений. Пусть $x_1, \dots, x_n \sim \text{iidPois}(\lambda)$. Найдите оценку максимального правдоподобия для λ . Найдите $\hat{\text{Var}}\hat{\lambda}$. Постройте 95% доверительный интервал.

3. Что-нибудь непрерывное (экспоненциальное или нормальное - ?)

1. Формализуем задачу. Параметр β , которые мы должны оценить — «работоспособность фонтана». Он может принимать различные значения. Например:

- фонтан работает каждый день;
- фонтан работает только по выходным;
- фонтан работает только сегодня.

В наших руках есть только одно наблюдение: мы приехали в конкретный день и увидели, что фонтан работает. Правдоподобие — это вероятность пронаблюдать данные, при условии, что выбрано конкретное значение параметра. То значение, при котором вероятность получить данные максимальна и будет нашей оценкой.

Какова вероятность того, что мы увидим работающий фонтан, если он работает ежедневно? Она равна 1. По аналогии, вероятность увидеть работающий фонтан, если он работает один раз в году составит $\frac{1}{365}$. Вероятность того, что он работает каждые выходные окажется около $\frac{1}{3}$. Выходит, что гипотеза, состоящая в том, что фонтан работает каждый день, самая правдоподобная. Именно это значение параметра «работоспособность фонтана» является оценкой максимального правдоподобия.

2.

Упражнение 11. Каждый из группы второкурсников, пришедших на пару, подкинул монетку два раза и никому не рассказывал, что у него выпало. В зависимости от того, что выпало на монетках, а также от того пробовал он наркотики или нет, он сказал одну из двух фраз.

	выпали два орла	другая комбинация
пробовал наркотики	ололо	юппи
не пробовал наркотики	юппи	ололо

Таким образом, если человек ололо, либо он курил травку и выпали два орла, либо он не курил травку и выпала другая комбинация. Всегда можно отмазаться. Получается, что при такой форме опроса нет смысла врать, и мы получим из него намного больше информации, чем если бы спрашивали у человека напрямик. «Ололо» сказали 10 человек, «юпи» сказали 4 человека.

1. Оцените методом максимального правдоподобия долю второкурсников, пробовавших наркотики;
2. Найдите $\hat{Var}(\hat{p})$.
3. Постройте 80 – % доверительный интервал для доли пробовавших наркотики. Можно ли доверять полученным оценкам? Что означает такой доверительный интервал?

Наш эксперимент устроен таким образом, что мы наблюдаем не совсем то событие, которое нам нужно. В данные вносится элемент случайности с целью получить более правдивый результат. Мы видим, что человек говорит либо «ололо» либо «юппи», хотя нам хотелось бы узнать кто употребляет наркотики.³ Вероятность того, что человек скажет «ололо» можно найти по формуле полной вероятности:

$$P(\text{ололо}) = P(\text{ололо} \mid \text{пробовал}) \cdot P(\text{пробовал}) + \\ + P(\text{ололо} \mid \text{не пробовал}) \cdot P(\text{не пробовал}) = \frac{1}{4} \cdot p + \frac{3}{4} \cdot (1 - p).$$

Вероятность того, что человек скажет «ололо», при условии что он пробовал наркотики совпадает с тем, что на монетке выпало два орла. Вероятность того, что человек пробовал наркотики мы хотим оценить. Обозначим её буквой p . Точно также мы можем выписать полную вероятность того, что человек скажет «юппи»:

³Нет, тезис «авторы употребляют наркотики» не является решением задачи.

$$\begin{aligned} P(\text{юппи}) &= P(\text{юппи} \mid \text{пробовал}) \cdot P(\text{пробовал}) + \\ &+ P(\text{юппи} \mid \text{не пробовал}) \cdot P(\text{не пробовал}) = \frac{3}{4} \cdot p + \frac{1}{4} \cdot (1 - p). \end{aligned}$$

Теперь мы можем выписывать функцию правдоподобия. Давайте выпишем её в общем виде. Пусть у нас есть n_1 «ололо» и n_2 «юппи», тогда:

$$L = P(\text{ололо})^{n_1} \cdot P(\text{юппи})^{n_2} = \left(\frac{3}{4} - \frac{1}{2} \cdot p \right)^{n_1} \cdot \left(\frac{1}{4} + \frac{1}{2} \cdot p \right)^{n_2}.$$

Логарифмируем правдоподобие

$$\ln L = n_1 \cdot \left(\frac{3}{4} - \frac{1}{2} \cdot p \right) + n_2 \cdot \left(\frac{1}{4} + \frac{1}{2} \cdot p \right).$$

Упражнение 12. Про Чебурашку и шляпы