

Основы теории вероятностей

Недавно прочитал книгу [Вэй Цзи Ма. Байесовские модели восприятия и действия](#), и мне понравилось введение в теорию вероятностей, вынесенное в приложение. Кратко и оригинально.

Алгебра помогает понять математику, а теория вероятностей помогает понять жизнь. К сожалению, теория вероятностей не является стандартным компонентом большинства учебных программ бакалавриата. Мы надеемся, что это изменится, а пока предлагаем ознакомиться с некоторыми основами теории вероятностей. Это ни в коем случае не исчерпывающее введение, которое можно найти в учебнике по теории вероятностей. Это краткий справочник, который сосредоточен только на концепциях и расчетах, используемых в книге.

Объективная и субъективная вероятности

Вероятность – это степень возможности. В самом узком смысле вероятность можно определить как ожидаемую частоту исхода повторяющегося события, например вероятность того, что монета выпадет орлом, или вероятность того, что кто-то выбросит 5 на игральном кубике. Эти события могут повторяться произвольно большое количество раз, и можно подсчитать частоты долгосрочных результатов. Если доля подбрасываний, при которых монета выпадает орлом, сходится к 0,5, по мере того как число подбрасываний приближается к бесконечности, мы можем утверждать, что монета имеет вероятность выпадения орла, равную 0,5. Этот тип вероятности иногда называют объективной вероятностью, и это единственный допустимый тип вероятности в соответствии со строгим частотным взглядом на вероятность.

Намного более широкое и, как мы полагаем, более полезное толкование вероятности – это степень веры в возможность исхода. Иногда ее называют субъективной вероятностью. Повседневные термины «уверенность» и «неопределенность» относятся к субъективным вероятностям. Если я знаю, что у кубика шанс выпадения 5 равен $1/6$, то моя уверенность в том, что выпадет 5, равна $1/6$. Этот конкретный пример тривиален, потому что он основан на простом преобразовании объективной вероятности (частоты долгосрочных результатов) в утверждение убеждения. Однако гораздо более широкая применимость субъективного толкования становится очевидной, когда мы рассматриваем степень уверенности в результатах, которые невозможно повторить, например вероятность того, что кандидат А победит кандидата В на предстоящих выборах.

Это не та вероятность, которую можно получить, повторяя одно и то же событие много раз, но тем не менее мы можем иметь надежное предсказание относительно результата. В самом деле, повседневная жизнь изобилует примерами субъективных вероятностей, которые нельзя сформулировать как частоты долгосрочных результатов: какова вероятность того, что сегодня пойдет дождь? Какова вероятность того, что мне понравится книга? Многие научные вопросы также можно сформулировать только в терминах субъективных, а не объективных вероятностей: какова вероятность того, что масса Сатурна находится между 10^{25} и 10^{26} кг? Какова вероятность того, что болезнь X вызвана вирусом?

Различие между объективной и субъективной вероятностями имеет долгую историю и не всегда очевидно. Например, чтобы определить вероятность того, что сегодня пойдет дождь, синоптик может запустить большое количество симуляций, начиная с текущего состояния атмосферы, каждая из которых использует разные экземпляры стохастических факторов в модели, и записать частоту дождливых дней в этой последовательности результатов запуска модели. Хотя результирующая вероятность является субъективной, она была получена «объективным» способом, а именно путем подсчета. Точно так же, если я наблюдаю на небе темные тучи и высказываю свое мнение о высокой вероятности дождя, я выражаю субъективное вероятностное суждение, но основываю это суждение на большом количестве ранее наблюдавшихся, подобных (хотя и неидентичных) состояний неба.

Байесовский вывод рассматривает субъективные и объективные вероятности одинаково: одни и те же математические отношения (правило Байеса, маргинализация и т. д.) одинаково применимы к обоим типам вероятностей. Поэтому байесовский вывод находит широкое применение. Однако байесовские модели восприятия основаны в основном на субъективной вероятности: существует только одно истинное состояние мира, но с точки зрения организма, пытающегося сделать вывод, существует множество возможностей, и этим возможностям можно присвоить степени уверенности.

Смысловое толкование вероятности

Мы называем множество всех рассматриваемых возможностей пространством событий. Событие или гипотеза – это подмножество пространства событий. Термин «событие» обычно используется при обсуждении объективной вероятности, а термин «гипотеза» – при обсуждении субъективной вероятности. Пространством событий могут быть «все возможные числа, которые я могу выбросить на кости» или «все возможные погодные условия, которые могут наблюдаться сегодня».

Относительно первого пространства событием может быть выпадение четного числа. Относительно второго пространства можно рассматривать, например, гипотезу «сегодня будет дождь».

Вероятность события или гипотезы – это вещественное число от 0 до 1, указывающее степень вероятности события или гипотезы. Событие, которое гарантированно случается, имеет вероятность 1, а невозможное событие имеет вероятность 0. Для событий вероятность можно представить как частоту, с которой событие происходит среди очень большого числа случайных выборок из пространства. Например, вероятность выпадения четного числа на шестигранной кости равна 0,5. Эту вероятность можно также рассматривать как степень уверенности. К гипотезам понятие частоты обычно не применяется, но вероятность по-прежнему представляет собой степень уверенности; например, степень уверенности в том, что сегодня будет дождь, может быть равна 0,35. Мы обозначаем вероятность события или гипотезы X через $p(X)$. Например, $p(\text{монета выпадет орлом}) = 0,50$.

Дополняющее событие

Для рассматриваемого события или гипотезы дополняющим событием или гипотезой является сценарий, при котором первое событие или гипотеза не имеет места или является ложным. Например, дополняющим событием к выпадению 1 на кубике является событие «выпадение 2, 3, 4, 5 или 6 на кубике». Если событие или гипотеза обозначены X , то их дополняющее событие или гипотеза обозначаются $\neg X$ (читается «не X »). Здесь \neg – символ логического отрицания. Вероятность дополняющего события или гипотезы равна 1 минус вероятность основного события или гипотезы: $p(\neg X) = 1 - p(X)$.

В некоторых ситуациях легче рассчитать вероятность дополняющего события или гипотезы, чем самого события или гипотезы. Например, если вас просят вычислить вероятность того, что сумма очков на двух игральных костях не меньше 3, проще сначала вычислить вероятность того, что сумма очков меньше 3, и вычесть ее из 1 (ответ 35).

Представление в виде диаграммы Венна

События и гипотезы могут быть представлены графически с помощью диаграмм Венна. Строго говоря, площадные диаграммы, которые мы показываем в этой книге, являются диаграммами Эйлера, поскольку они представляют только отношения, которые действительно имеют мест. По определению диаграммы Венна представляют все возможные логические отношения между множествами. Однако мы следуем общепринятому соглашению называть любые площадно-вероятностные диаграммы диаграммами Венна.

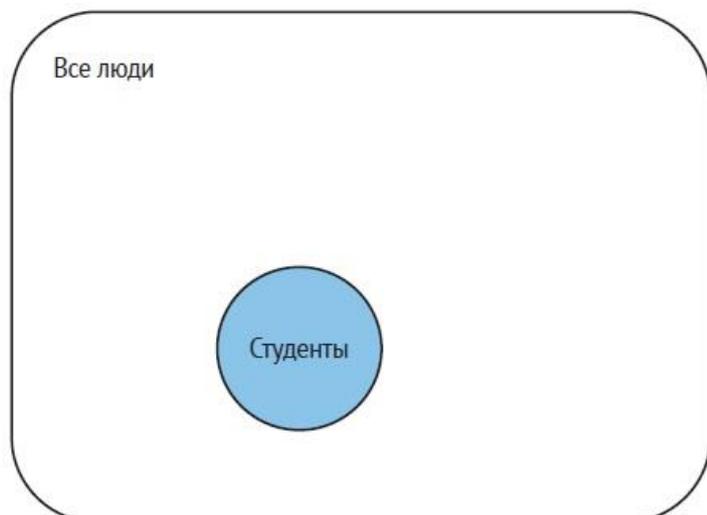


Рис. 1. Пример диаграммы Венна, используемой для представления вероятностей

Сначала нарисуем большой прямоугольник, внутренняя часть которого представляет все возможные результаты, то есть пространство выборки. Присвоим площади этого прямоугольника значение 1, представляющее суммарную вероятность 1. Затем нарисуем внутри этого прямоугольника круг, представляющий все результаты, согласующиеся с конкретным событием или гипотезой.

Например, прямоугольник может представлять всех людей в определенной группе, а круг – всех студентов среди них. Площадь, заключенная в круг, является частью площади, заключенной в прямоугольник; эта доля представляет собой вероятность события или гипотезы – в нашем примере вероятность того, что случайно выбранный человек является студентом. Дополняющее событие или гипотеза представлено точками, которые находятся внутри прямоугольника, но снаружи круга. Его площадь, деленная на площадь прямоугольника, представляет собой вероятность того, что случайно выбранный человек в группе не является студентом.

Случайные величины и их распределения

Случайная величина – это переменная, значения которой не могут быть известны с уверенностью. Примеры случайной величины: число, выпавшее на кубике, дата рождения человека, размер обуви случайного человека на вашей улице, время, которое требуется, чтобы добраться от дома до работы, количество избирателей, которые будут участвовать в предстоящих выборах, или цена акции завтра. Противоположностью случайной величине является переменная, значение которой достоверно известно. Примерами неслучайных величин являются количество планет между нами и Солнцем (две), количество дней в неделе (семь), отношение длины окружности к диаметру круга (π) и расстояние между двумя соседними сантиметровыми отметками на линейке (1 см).

Это не абсолютное различие. Переменные, которые кажутся неслучайными, могут быть подвержены шуму производства или измерения, что делает их случайными. Например, расстояние между двумя соседними отметками на линейке может слегка различаться, поскольку машина, производящая линейки, вероятно, была запрограммирована компьютером, который устанавливал отметки в сантиметрах. Однако числа, сгенерированные компьютером, имеют только конечное число знаков после запятой, возможно десять. Как следствие сантиметровые отметки никогда не достигнут фемтометровой точности.¹ Кроме того, краска, используемая для маркировки, не будет прилипать к поверхности одинаковым образом при изготовлении каждой линейки. Следовательно, расстояние между двумя соседними метками можно рассматривать как случайную величину. По причинам, подобным этим, может быть полезно рассматривать все переменные как случайные, только имеющие очень низкую неопределенность.

Случайность, также называемая изменчивостью, шумом или стохастичностью, часто является следствием недостатка знаний. Если бы перед броском кубика вам удалось каким-то образом записать точное положение, направление и скорость, с которой кость покидает руку, то вы смогли бы точно смоделировать взаимодействие кости с воздухом и столом и предсказать исход броска. Поскольку никому неизвестны значения всех этих переменных, результат броска кубика считается случайным. Существует ли истинная случайность – это философский вопрос, который выходит за рамки книги.

Сравнение дискретных и непрерывных случайных величин

Случайные переменные можно классифицировать по значениям, которые они могут принимать. Наиболее существенное различие наблюдается между дискретными и непрерывными случайными величинами. *Дискретная случайная величина* принимает набор значений, которые можно посчитать, даже если их может быть бесконечно много. Примерами могут служить количество детей в семье, количество потенциалов действия, запускаемых нейроном, количество ходов в шахматной партии, возраст человека, если его считать целыми годами, и идентичность произнесенного слова. Дискретная случайная величина, принимающая только два возможных значения, называется бинарной.

¹ Вот что пояснил ChatGPT по этому поводу. Фемтометровая точность означает измерение с точностью до фемтометра, что равно 10^{-15} метра. Фемтометр (фм) – это единица измерения длины в системе Международных единиц (СИ), используемая для измерения очень маленьких объектов, таких как размеры атомных ядер или длины волн в высокоэнергетической физике. Точность в фемтометрах требует очень высокоточных измерительных приборов и методов, и она является одним из высших уровней точности в научных и инженерных измерениях.

Непрерывная случайная величина принимает значения в континууме. Примерами являются длина отрезка, направление, в котором можно идти в открытом поле, время ожидания перед красным светом, скорость автомобиля и частота звука. Непрерывную переменную можно рассматривать как дискретную, но со значениями, поступающими с бесконечно малыми приращениями. Например, расстояние является непрерывной переменной, но когда его измеряют в миллиметрах, оно является дискретной переменной. В компьютерной программе истинно непрерывных переменных не существует; они всегда дискретизированы с конечной точностью.

Суммарная вероятность всегда равна 1

Суммарная вероятность всех возможных значений случайной величины равна 1. Эта сумма является отражением того факта, что возможности взаимоисключающие. Если увеличить вероятность одного значения, то вероятность хотя бы одного другого значения должна уменьшиться.

Дискретные распределения вероятности

Дискретные распределения вероятности – это функции, которые присваивают вероятность каждому возможному значению дискретной случайной величины. Распределение вероятности по дискретной случайной величине также называется *функцией распределения массы вероятности*. Дискретная случайная величина X , принимающая определенное значение x , является событием, обозначаемым $p(X = x)$. Поскольку теперь мы варьируем x по всем его возможным значениям, мы получаем функцию от x . Это функция распределения массы вероятности, которую мы будем обозначать через $p_x(x)$ или через $p(x)$, если нет двусмысленности в отношении идентичности случайной величины: $p_x(x) = p(X = x)$.

Это равенство означает, что функция распределения массы вероятности, вычисленная в точке x , равна вероятности того, что случайная величина примет это значение. Мы используем индекс X (верхний регистр) для обозначения случайной величины, а аргумент x (нижний регистр) – для конкретного значения этой случайной величины. Термин «масса» заимствован из физики. Грубо говоря, он основан на использовании материи как метафоры возможности (точки в пространстве событий). Чем больше вероятность события или гипотезы, тем больше масса «материи».

Например, пусть случайная величина X – это число, выпавшее на игральном кубике. Его возможные значения x равны от 1 до 6. Если кубик честный (шансы выпадения всех граней равны), вероятность каждого из этих значений равна $1/6$, то есть $p(X = x) = 1/6$ для каждого x . Это пример дискретного равномерного распределения. Если кубик «нечестный», вероятность хотя бы двух значений будет отличаться от $1/6$, и распределение больше не будет однородным.

Для дискретных случайных величин общая вероятность вычисляется путем суммирования всех возможных значений и в итоге должна быть равна 1. Это свойство записывается следующим образом:

$$(1) \sum_x p(x) = 1$$

В конкретном случае, когда даны возможные значения X , мы можем поставить их выше и ниже знака суммы. Например, общая вероятность броска кости будет равна

$$(2) \sum_{x=1}^6 p(x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

Бинарные случайные величины являются частным случаем дискретных случайных величин.

Предположим, что бинарная случайная величина X может принимать значения x_1 и x_2 . Мы знаем, что общая вероятность равна 1. Следовательно, вероятность x_2 равна 1 минус вероятность x_1 , то есть $p(X = x_2) = 1 - p(X = x_1)$.

Непрерывные распределения вероятности

Какова вероятность того, что рост человека точно равен 160 см? Она равна нулю, поскольку «точно» означает, что рост указан с точностью до бесконечного числа знаков после запятой. Эта проблема характерна для непрерывных случайных величин и говорит о том, что функции распределения массы вероятности, которые хорошо работали для дискретных распределений, должны быть заменены другим понятием, чтобы учесть свойства непрерывных переменных.

Предположим, нас интересует распределение вероятности роста взрослого человека (рис. 2). В первом приближении можно рассмотреть возможные значения в интервалах с шагом 10 см: от 120 до 130 см, от 130 до 140 см и т. д. С каждым интервалом связана вероятность, и таким образом мы можем построить функцию распределения масс вероятностей. Однако может возникнуть необходимость описать рост более точно, скажем в интервалах по 5 см: между 122,5 и 127,5 см и т. д. Таким образом, каждый исходный интервал заменяется двумя новыми, каждый из которых имеет в среднем половину вероятностной массы исходного интервала. Таким образом, новая функция распределения массы вероятности масштабируется примерно до половины высоты исходной.

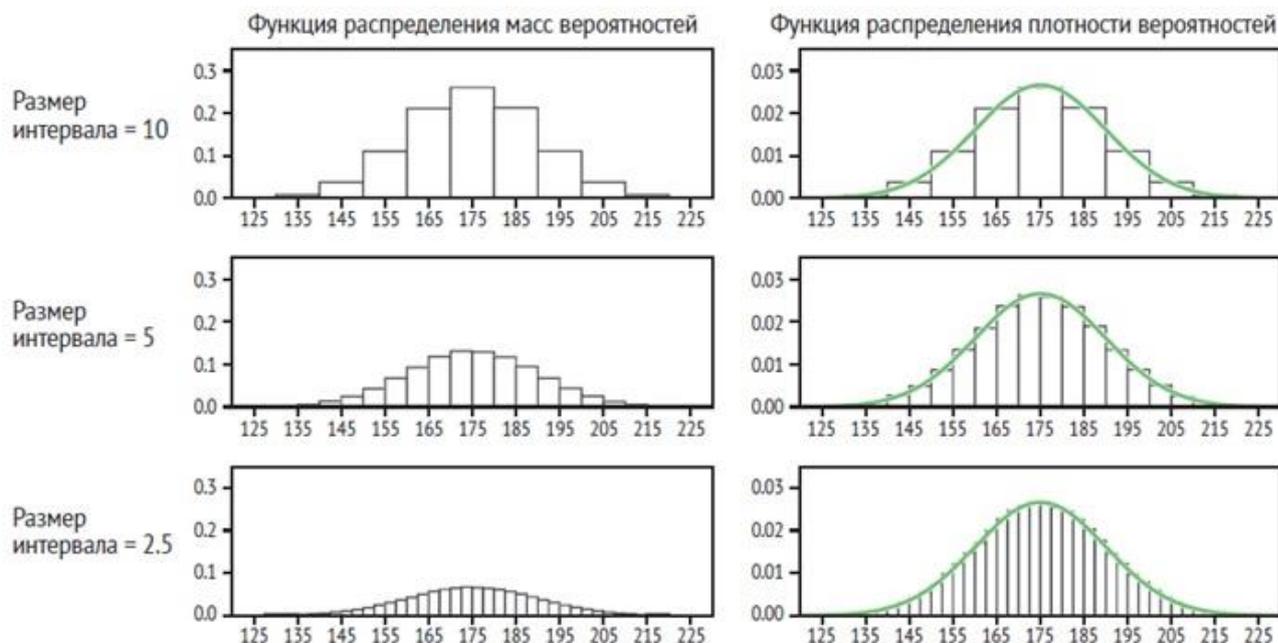


Рис. 2. Если случайная величина может принимать континуум значений (ось x), функция распределения массы вероятности определена только тогда, когда результаты разбиты по интервалам. Значения функции будут уменьшаться с уменьшением размера интервала (левый столбец). Функции распределения плотности вероятности получаются путем деления значений массы вероятности на размер интервала. Это дает значения, которые не зависят от размера интервала. Процесс уменьшения размера интервалов можно продолжать до тех пор, пока они не станут бесконечно малы. Получится непрерывная функция распределения плотности вероятности, показанная зеленым цветом

Поскольку мы продолжаем уменьшать ширину интервала, чтобы повысить точность, вероятностная масса на интервал также продолжает уменьшаться – она может стать сколь угодно малой. Это не очень хорошо. Есть ли способ предотвратить «исчезновение» массы вероятности? Да, это можно сделать, разделив массу вероятности в интервале на ширину интервала. При этом функция несильно меняется по мере уменьшения ширины интервала – она становится только более точной. Результатом этого процесса, когда ширина ячейки приближается к нулю (зеленая кривая), является *функция плотности вероятности*. Опять же, здесь просматривается аналогия с физикой: если вероятность в интервале рассматривается как масса, то деление этой вероятности на ширину интервала аналогично вычислению линейной плотности: масса на единицу длины оси x .

Сходство в обозначениях между функцией распределения массы вероятности для дискретных переменных и функцией распределения плотности вероятности для непрерывных переменных, которые записываются как $p_x(x)$ или $p(x)$, вводит в заблуждение, поскольку между ними есть некоторые важные концептуальные различия. Для дискретного распределения масса вероятности отдельной точки никогда не превышает 1, поскольку значения массы вероятности должны в сумме равняться 1. Для непрерывного распределения плотность вероятности в одной точке не имеет смысла и может принимать сколь угодно большие значения.

Рассмотрим, например, равномерное распределение на интервале $[0, 0,01]$. Оно будет иметь плотность вероятности 100 в каждой точке. Только интеграл по интервалу всегда будет меньше или равен 1. Иными словами, для дискретного распределения вероятность $p(X = x)$ – это значимое число,

которое может принимать любое значение от 0 до 1, и на самом деле тождественно $p(x)$. Для непрерывного распределения $p(X = x)$ всегда равна 0, и имеют смысл только вероятности вида $p(a \leq X \leq b)$ с произвольными числами a и b . Например, вероятность того, что вы попадете абсолютно точно в центр мишени бесконечно тонким дротиком, равна 0.

Мы будем использовать такие термины, как *функция распределения вероятности* (probability distribution function, PDF), распределение вероятности или просто распределение для обозначения функции распределения массы вероятности дискретной случайной величины или функции распределения плотности вероятности непрерывной случайной величины.

Так же, как и для дискретных переменных, полная вероятность всех значений непрерывной переменной равна 1. Полная вероятность для непрерывной переменной вычисляется не как сумма, а как интеграл. Интеграл непрерывной функции плотности вероятности, как определено выше, представляет собой ширину интервала, умноженную на значение функции в этом интервале и суммированную по всем интервалам в пределе, когда ширина интервала приближается к нулю. В алгебре существуют методы вычисления интегралов определенных функций. Правило полной вероятности для непрерывной переменной X записывается как

$$(3) \int p(x)dx = 1$$

По сути, dx – это ширина очень маленького интервала (приращения), а знак интеграла – растянутая буква S от слова «сумма».

Наиболее важным непрерывным распределением является нормальное распределение, которое мы подробно обсудим ниже. Другим важным распределением является равномерное распределение. Равномерное распределение имеет постоянную плотность распределения на интервале $[a, b]$:

$$(4) p(x) = \frac{1}{b - a}$$

Существуют и другие распределения, широко распространенные в приложениях теории вероятностей. Экспоненциальное распределение задается выражением $p(x) = \lambda e^{-\lambda x}$, где λ – константа, x – положительное вещественное число. Распределение по степенному закону задается как $p(x) \propto x^{-a}$, где a – константа.

Формальное определение функции плотности вероятности

Рассмотрим непрерывную случайную величину X , такую как, например, время ожидания в очереди. Вероятность того, что значение X меньше или равно x , обозначается как $p(X \leq x)$. Это *интегральная функция распределения* (cumulative distribution function, CDF) X в точке x , обозначаемая $P_X(x)$: $P_X(x) = p(X \leq x)$.

По определению это монотонно возрастающая функция, принимающая значения между 0 (при приближении x к 0) и 1 (при приближении x к ∞). PDF X теперь является производной этой функции:

$$(5) p_X(x) = \frac{dP_X}{dx}$$

Пример CDF и PDF:

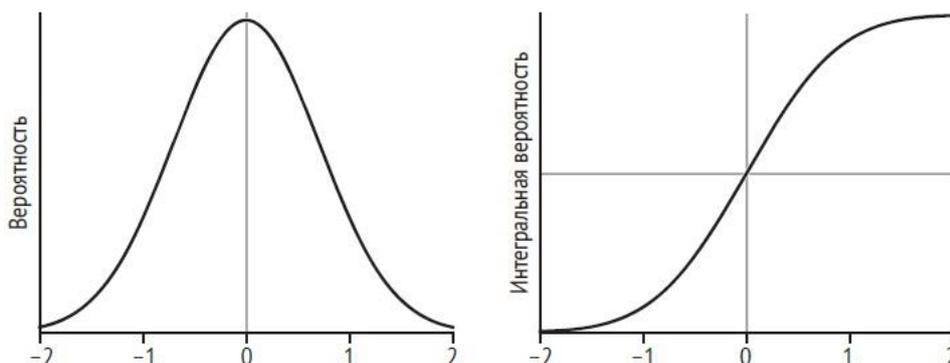


Рис. 3. Функция плотности вероятности и интегральная функция распределения

Чтобы вернуться из PDF в CDF, мы интегрируем:

$$(6) P_X(x) = p(X \leq x) = \int_{-\infty}^x p_X(y) dy$$

$p_X(y)$ означает плотность вероятности случайной величины X , оцененную при значении y фиктивной переменной. Физический смысл этого выражения состоит в том, что интеграл по плотности представляет собой массу. Отсюда сразу следует, что вероятность того, что X принимает значения в интервале (x_1, x_2) , может быть получена путем интегрирования $p_X(x)$ между x_1 и x_2 :

$$(7) p(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p_X(x) dx$$

Нормирование

Функцию можно превратить в распределение вероятностей, разделив каждое значение на общее значение во всей области, при условии что это общее значение конечно. В результате распределение вероятностей будет интегрироваться (или суммироваться) до 1. Этот процесс называется нормированием. Если общее значение во всей области бесконечно, нормирование невозможно.

Упражнение В.1. Докажите, что экспоненциальное распределение нормировано.

Функция плотности вероятности экспоненциального распределения, определенного на области $[0, \infty)$

$$(8) p_X(x) = \lambda e^{-\lambda x}$$

Интегральная функция экспоненциального распределения

$$(9) P_X(x) = \int_0^{\infty} \lambda e^{-\lambda x} dx$$

Интегрируем

$$(10) \int_0^{\infty} \lambda e^{-\lambda x} dx = \lim_{a \rightarrow \infty} \int_0^a \lambda e^{-\lambda x} dx = \lim_{a \rightarrow \infty} \left[-e^{-\lambda x} \Big|_0^a \right] = \lim_{a \rightarrow \infty} (-e^{-\lambda a} + 1)$$

При a стремящемся к бесконечности $-e^{-\lambda a}$ стремится к нулю. Таким образом интеграл (9) равен 1, что и требовалось доказать.

Упражнение В.2. Нормируйте степенное распределение $p(x) = x^{-a}$ и найдите условие для a , при соблюдении которого возможно нормирование.

Итак, мы хотим проинтегрировать степенной закон:

$$(11) \int_0^{\infty} x^{-a} dx$$

Исключим $x = 0$, так как в этой точке интеграл не сходится.

Произведем интегрирование:

$$(12) \int_0^{\infty} x^{-a} dx = \lim_{b \rightarrow \infty} \left[\frac{x^{-a+1}}{-a+1} \Big|_0^b \right] = \lim_{b \rightarrow \infty} \left[-\frac{1}{(a-1)b^{a-1}} \right]$$

Если $a > 1$, то знаменатель уходит в бесконечность при $b \rightarrow \infty$, и весь предел равен нулю. Таким образом, интеграл сходится и равен некоторому конечному значению, что соответствует нормировке функции распределения.

Среднее значение, дисперсия и математическое ожидание

Для дискретной случайной величины X среднее или ожидание X равно

$$(13) E[X] = \sum_x xp(x)$$

Дисперсия, которая является мерой разброса вокруг среднего, определяется как

$$(14) Var[X] = \sum_x (x - E[X])^2 p(x)$$

Стандартное отклонение – это квадратный корень из дисперсии. Среднее и дисперсия являются частными случаями математического ожидания любой функции случайной величины. Если мы обозначим функцию через f , то ожидание f равно

$$(15) \mathbb{E}[f(X)] = \sum_x f(x)p(x)$$

Таким образом, среднее значение является ожиданием X , а дисперсия равна

$$(16) \text{Var}[X] = \mathbb{E}[(x - \mathbb{E}[X])^2]$$

Для непрерывной случайной величины X с плотностью вероятности $p(x)$ аналогичные выражения получают заменой сумм интегралами:

$$(17) \mathbb{E}[X] = \int_{-\infty}^{\infty} xp(x)dx$$

$$(18) \text{Var}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 p(x)dx$$

$$(19) \mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)dx$$

Как для дискретных, так и для непрерывных переменных существует общее эквивалентное выражение для дисперсии, а именно разность между средним значением квадратов и квадратом среднего:

$$(20) \text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Нормальное распределение

Наиболее важным непрерывным распределением в приложениях теории вероятностей является нормальное, или гауссово, распределение. Его функция распределения вероятностей имеет вид:

$$(21) p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Иногда мы используем сокращенную запись $p(x) = N(x; \mu, \sigma^2)$. Параметры μ и σ^2 не имеют априорного значения (они просто обозначены как предположение), но оказываются равными среднему и дисперсии распределения соответственно. Множитель перед экспонентой нужен для нормирования.

Стандартное нормальное распределение – это нормальное распределение со средним значением 0 и стандартным отклонением 1.

Центральная предельная теорема

Важность нормального распределения вытекает главным образом из центральной предельной теоремы. Говоря упрощенно, центральная предельная теорема утверждает, что среднее значение большого числа независимых случайных величин с идентичными распределениями вероятностей будет следовать приблизительно нормальному распределению, независимо от распределения исходных переменных. Вся мощь теоремы заключается в ее последней части: распределение исходных переменных не имеет значения. Ограничение теоремы можно смягчить, чтобы учесть независимые, но не одинаково распределенные переменные.

В математических моделях восприятия центральная предельная теорема всегда присутствует на заднем плане: всякий раз, когда мы предполагаем, что шум, искажающий стимул, распределен нормально, мы, по сути, обосновываем это с помощью центральной предельной теоремы. Случайная величина, которая описывает шум, искажающий стимул, может быть суммой большого числа независимых шумовых процессов.

Произведение двух нормальных распределений

Рассмотрим произведение двух гауссовых распределений вероятностей по одной и той же случайной величине X . У одного распределения среднее μ_1 и дисперсия σ_1^2 , а у другого среднее μ_2 и дисперсия σ_2^2 . Перемножим эти два распределения так же, как мы умножаем обычные функции, а затем

нормируем результат (поскольку произведение не нормализуется автоматически). Каким будет итоговое распределение вероятности? Сначала запишем выражения для двух функций плотности вероятности:

$$(22) p(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$$(23) p(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

Умножение этих двух функций сводится к суммированию показателей степени. Сумма показателей степени

$$(24) -\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2} = -\frac{1}{2} \left(\frac{x^2 - 2\mu_1x + \mu_1^2}{\sigma_1^2} + \frac{x^2 - 2\mu_2x + \mu_2^2}{\sigma_2^2} \right)$$

Выполним перестановку, сгруппировав все члены, содержащие x^2 , и все содержащие x . Сумма показателей степени

$$(25) -\frac{1}{2} \left[\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) x^2 - 2x \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right) + \dots \right] = -\frac{1}{2} [(J_1 + J_2)x^2 - 2x(J_1\mu_1 + J_2\mu_2) + \dots]$$

где мы используем следующие обозначения точности:

$$(26) J_1 \equiv \frac{1}{\sigma_1^2}, J_2 \equiv \frac{1}{\sigma_2^2}$$

Кроме того, здесь и далее точками обозначены все слагаемые, не зависящие от x . При возведении в степень эти члены становятся постоянными множителями, не зависящими от x . Поскольку результат перемножения распределений должен быть нормирован в конце вычислений, любой постоянный множитель, который мы добавляем или оставляем до этого момента, не имеет значения.

Таким образом, произведение распределений (23 и 24) равно

$$(27) p_1(x)p_2(x) \propto e^{-\frac{1}{2(J_1+J_2)^{-1}} \left(x^2 - 2x \frac{J_1\mu_1 + J_2\mu_2}{J_1+J_2} + \dots \right)} = e^{-\frac{1}{2(J_1+J_2)^{-1}} \left(x - \frac{J_1\mu_1 + J_2\mu_2}{J_1+J_2} \right)^2}$$

где знак пропорциональности поглощает все множители, не зависящие от x . Мы рассматриваем результат как другое нормальное распределение, теперь со средним значением

$$(28) \frac{J_1\mu_1 + J_2\mu_2}{J_1 + J_2}$$

и дисперсией

$$(29) \frac{1}{J_1 + J_2}$$

Интегральное нормальное распределение

Интегральная функция гауссова распределения не является элементарной функцией (то есть функцией, построенной из композиции таких элементарных функций, как степенные с любым действительным показателем, показательные и логарифмические, тригонометрические и обратные тригонометрические с использованием операций сложения, вычитания, умножения и деления). Однако у нее есть специальное название просто потому, что она часто встречается. Мы определяем интегральное распределение стандартной нормальной плотности вероятности как

$$(30) \Phi_{standard}(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{x^2}{2}} dx$$

Эта функция принимает значения от 0 до 1.

Упражнение В.5. Покажите, что $\Phi_{standard}(0) = 0,5$.

Распределение фон Мизеса

Некоторые переменные, такие как ориентация и направление движения, естественным образом имеют круговую (периодическую) область. Для таких переменных гауссово распределение не имеет смысла. Одним из решений является выбор распределения фон Мизеса. Его можно рассматривать как круговой аналог гауссова распределения. Оно имеет два параметра: круговое среднее и параметр концентрации, аналогичный обратной величине дисперсии гауссова распределения. Распределение фон Мизеса по круговой переменной s с областью определения $[0, 2\pi)$, круговым средним μ и параметром концентрации k имеет вид:

$$(31) p(s) = \frac{1}{2\pi I_0(k)} e^{k \cos(s-\mu)}$$

Здесь I_0 – модифицированная функция Бесселя первого рода порядка 0. Это так называемая специальная функция – определяемая через интеграл или бесконечный ряд. Для нас важно лишь то, что $2\pi I_0(k)$ нормирует распределение фон Мизеса. Когда $k = 0$, $I_0(k) = 1$, и распределение Мизеса становится равномерным на окружности. Чем выше k , тем больше распределение фон Мизеса похоже на нормальное распределение.

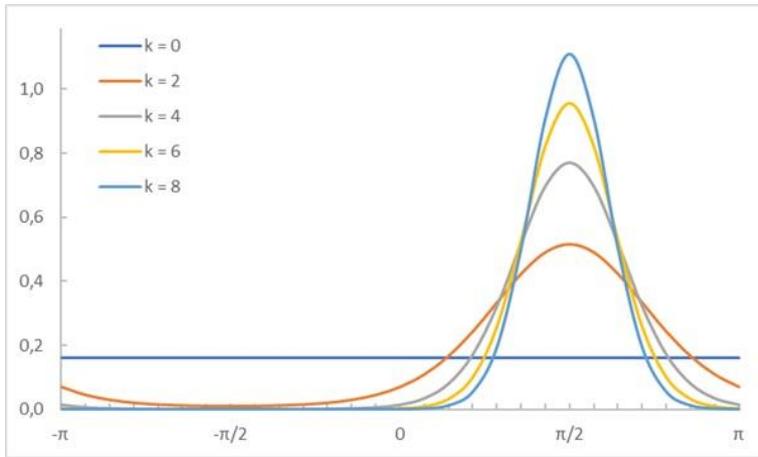


Рис. 4. Примеры распределений фон Мизеса. Область определения случайной величины является периодической. Для построения распределения в Excel использовалась формула (31), $I_0(k) = \text{БЕССЕЛЬ.}I(k;0)$, $\mu = \pi/2$

Упражнение В.7. Мы сказали, что μ – это круговое среднее, но как бы вы определили среднее значение для круговой переменной?

Вот что об этом думает ChatGPT. Круговое среднее представляет собой угол, который характеризует центр масс (центр распределения) круговых данных. Пусть у нас есть набор угловых данных $\theta_1, \theta_2, \dots, \theta_m$, измеряемых в радианах. Круговое среднее обозначается как μ и вычисляется по формуле:

$$(32) \mu = \text{atan2} \left(\frac{1}{n} \sum_{i=1}^n \sin(\theta_i), \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \right)$$

где atan2 – это функция арктангенса с двумя аргументами, что предотвращает проблемы с выбором квадранта.

Это выражение представляет угол, соответствующий центру массы всех угловых точек на круге. По сути, это среднее направление углов.

Упражнение В.8. Покажите аналитически, что в пределе больших k распределение фон Мизеса становится нормальным распределением. (Подсказка: используйте разложение косинуса в ряд Тейлора.) Кроме того, покажите, что точность этого нормального распределения равна k .



ChatGPT

Давайте рассмотрим функцию плотности вероятности распределения фон Мизеса:

$$f(\theta|\mu, \kappa) = \frac{e^{\kappa \cos(\theta-\mu)}}{2\pi I_0(\kappa)}$$

где μ - центр распределения, κ - параметр концентрации, а $I_0(\kappa)$ - модифицированная функция Бесселя первого рода и нулевого порядка.

Разложим косинус в ряд Тейлора до второго порядка:

$$\cos(\theta - \mu) \approx 1 - \frac{(\theta-\mu)^2}{2}$$

Теперь подставим это приближение в функцию плотности вероятности:

$$f(\theta|\mu, \kappa) \approx \frac{e^{\kappa(1-(\theta-\mu)^2/2)}}{2\pi I_0(\kappa)}$$

Упростим выражение:

$$f(\theta|\mu, \kappa) \approx \frac{e^{\kappa-\kappa(\theta-\mu)^2/2}}{2\pi I_0(\kappa)}$$

$$f(\theta|\mu, \kappa) \approx \frac{e^{\kappa}}{2\pi I_0(\kappa)} e^{-\kappa(\theta-\mu)^2/2}$$

Теперь функция плотности вероятности стала похожей на функцию плотности вероятности нормального распределения с параметрами μ и $\frac{1}{\kappa}$. В пределе, когда κ становится большим, распределение фон Мизеса приближается к нормальному распределению с параметрами μ и $\frac{1}{\kappa}$.

Чтобы показать, что точность этого нормального распределения равна κ , мы можем рассмотреть асимптотическое разложение модифицированной функции Бесселя $I_0(\kappa)$ при больших значениях κ :

$$I_0(\kappa) \approx \frac{e^{\kappa}}{\sqrt{2\pi\kappa}} \left(1 - \frac{1}{8\kappa} + \frac{1}{128\kappa^2} - \dots\right)$$

Таким образом, точность аппроксимации фон Мизеса нормальным распределением пропорциональна κ , как видно из дополнительного члена $\frac{1}{\kappa}$ в разложении $I_0(\kappa)$.

Рис. 5. Ответ ChatGPT на упражнение В.8

Дельта-функция

Случайная величина особого типа – это случайная величина, принимающая только одно возможное значение. Существует специальное обозначение распределения вероятностей такой случайной величины. Если X – непрерывная случайная величина, которая всегда принимает значение $X = a$, то ее распределение можно записать так: $p(x) = \delta(x - a)$.

Здесь δ – дельта-функция Дирака. Она возвращает 0, если аргумент не равен 0, а в ином случае возвращает бесконечность. Конечно, бесконечность не является числом, и поэтому дельта-функция Дирака, строго говоря, не является обычной функцией. Для нас это не имеет значения, поскольку единственное место, где мы будем использовать дельта-функцию, – внутри интеграла. Там любая функция $f(x)$ обладает следующим свойством:

$$(33) \int \delta(x - a)f(x)dx = f(a)$$

где мы предполагаем, что a входит в область интегрирования. Фактически уравнение (33) является определяющим свойством дельты Дирака. Дельта-функция равносильна вычислению функции f внутри интеграла в одной точке a .

Нам представляется удобным использовать для дискретных переменных те же обозначения, что и для непрерывных, т. е. писать $\delta(x - a)$ вместо δ_{xa} . Тогда дискретный аналог уравнения (33) имеет вид

$$(34) \sum_x \delta(x - a)f(x) = f(a)$$

Распределение Пуассона

Дискретное распределение вероятностей, которое используем для описания нейронной активности, – это распределение Пуассона. Возможные значения случайной величины Пуассона: 0, 1, 2, 3, ... (верхний предел отсутствует). Распределение Пуассона имеет свободный параметр λ . Распределение вероятностей X определяется выражением

$$(35) p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Распределение Пуассона является дискретным и определяется только целыми неотрицательными числами. Множитель $e^{-\lambda}$ действует как нормировочный. В отличие от x , λ не обязательно должно быть целым числом. Среднее значение и дисперсия переменной с распределением Пуассона равны λ .

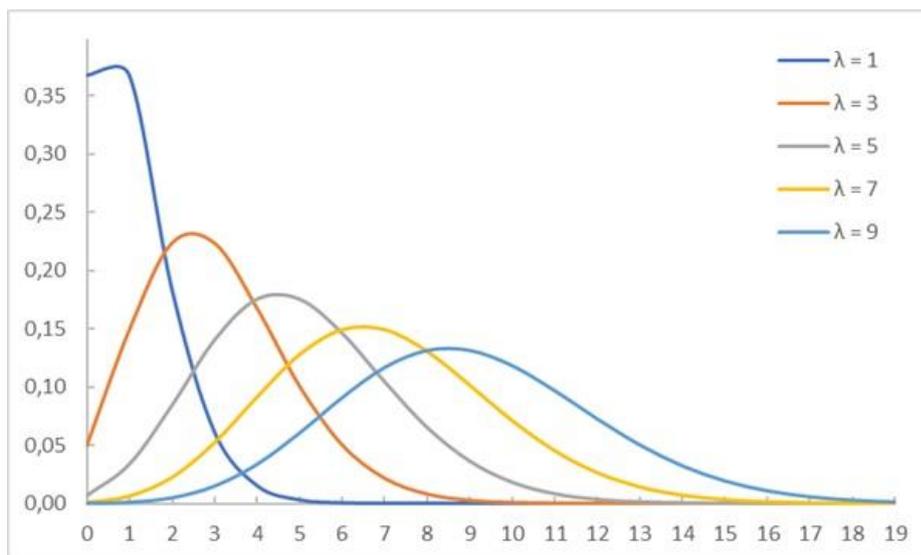


Рис. 6. Распределение Пуассона

Выборка из распределения вероятностей

В вероятностном моделировании нам часто приходится получать случайные числа в соответствии с заданным распределением вероятностей. Эти числа также называются выборками. На самом деле это отнюдь не тривиальная задача, но, к счастью, большинство программных пакетов имеют встроенные генераторы случайных чисел для наиболее распространенных распределений. Мы можем использовать эти функции для написания пользовательского кода, извлекающего из вероятностных распределений случайные значения, которые не были запрограммированы заранее.

Распределения, включающие несколько переменных

Для нас может представлять интерес зависимость случайных переменных друг от друга. Эта зависимость отражена в совместных и условных распределениях вероятностей, а также в правиле Байеса. Концепции, обсуждаемые в этом разделе, применимы как к непрерывным, так и к дискретным переменным. Таким образом, p может обозначать распределение либо массы вероятности, либо плотность вероятности. Поскольку мы рассматриваем несколько переменных одновременно, мы обычно будем использовать нижний индекс у p , чтобы обозначить, к какой случайной переменной(ым) принадлежит распределение вероятностей.

Совместная вероятность

Совместное распределение вероятностей случайных величин X и Y обозначается $p_{X,Y}(x, y)$, или, сокращенно, $p(x, y)$. Это вероятность значений x и y как пары. Суммирование по x и y дает 1:

$$(36) \sum_x \sum_y p(x, y) = 1$$

Для непрерывных переменных интеграл по обоим переменным равен 1:

$$(37) \iint p(x, y) dx dy = 1$$

Совместная вероятность симметрична: $p(x, y) = p(y, x)$

Если X и Y представляют события, совместная вероятность X и Y – это вероятность того, что произойдут оба события, обозначаемая как $p(X, Y)$ или $p(X \cap Y)$. В форме диаграммы Венна мы представляем Y в виде второй окружности, пересекающей первую. Совместная вероятность X и Y равна площади пересечения. Она всегда меньше или равна площади каждого отдельного круга. Это свойство выражают отношения $p(X, Y) \leq p(X)$ и $p(X, Y) \leq p(Y)$. Например, вероятность того, что в данный день идет дождь и вы будете на работе вовремя, меньше, чем вероятность того, что идет дождь. Эти отношения справедливы только для дискретных переменных.

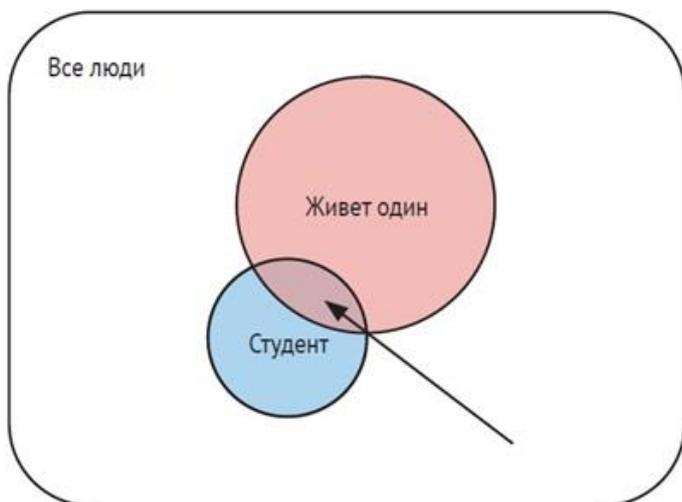


Рис. 7. Совместная вероятность событий «быть студентом» и «жить отдельно» представлена площадью пересечения, указанной стрелкой

Маргинализация

Представьте, что у вас есть кошка и собака. Вы тщательно отслеживаете, какова вероятность того, что в течение дня в гостиной находится только ваша кошка, только ваша собака, ни та, ни другая или они обе. Эти вероятности показаны в таблице, которая называется таблицей сопряженности (contingency table) и представляет вероятности совместных исходов. Предельные вероятности – это вероятности того, что кошка присутствует или отсутствует независимо от собаки, и вероятности того, что собака присутствует или отсутствует независимо от кошки.

	Кошки нет	Кошка есть	Итого
Собаки нет	0.40	0.05	0.45
Собака есть	0.30	0.25	0.55
Итого	0.70	0.30	1

Рис. 8. Частоты сочетаний двух случайных величин

Маргинализация – это операция получения из совместного распределения по нескольким переменным распределения по подмножеству этих переменных. Например, если $p(x, y)$ является совместным распределением X и Y , то суммирование по Y дает распределение только X :

$$(38) \sum_y p(x, y) = p(x)$$

Аналог для непрерывных переменных получается заменой суммы интегралом:

$$(39) \int p(x, y) dy = p(x)$$

Это суммирование, или интегрирование, называется маргинализацией, потому что $p(x)$ и $p(y)$ называются маргиналами $p(x, y)$. Если вы представляете (x, y) как точку в двумерном пространстве и совместное распределение обеспечивает z значений в этом пространстве, то маргиналы – это распределения, полученные путем суммирования по одному из измерений. Это приводит к двум одномерным распределениям, которые находятся на «краях» (margin) исходного двумерного распределения.

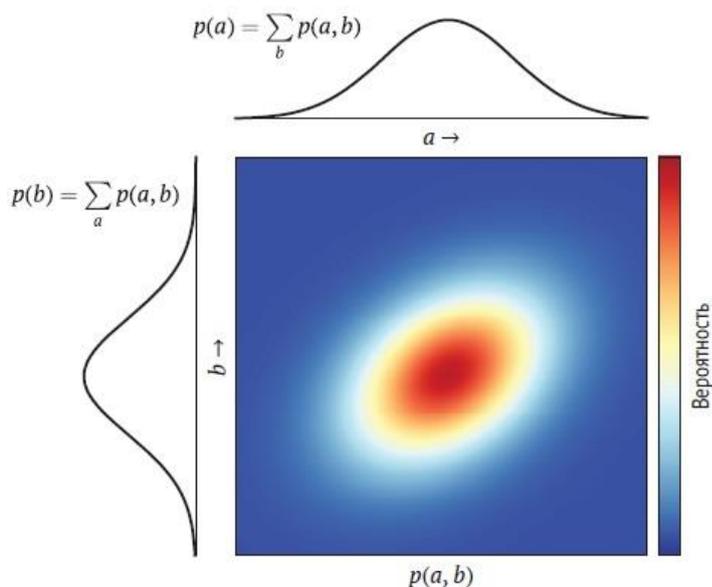


Рис. 9. Совместное распределение вероятностей двух случайных величин А и В. Черные кривые представляют маргиналы, полученные путем суммирования по одной из двух переменных

Условная вероятность

Для событий X и Y условная вероятность $p(X|Y)$ является ответом на вопрос: «Из всех исходов, согласующихся с событием Y , какая часть также согласуется с событием X ?» Условная вероятность события всегда находится между 0 и 1. В представлении диаграммы Венна вероятность $p(X|Y)$ равна площади пересечения, деленной на площадь второго круга. Точно так же вероятность того, что произойдет Y , при условии что произойдет X , равна площади пересечения, деленной на площадь первого круга.

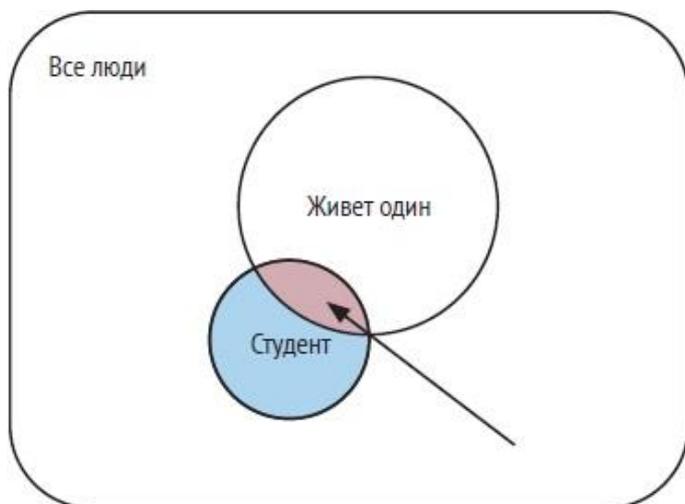


Рис. 10. Условная вероятность «живет один» при условии «является студентом» представлена площадью пересечения, деленной на площадь синего диска

Три примера условной вероятности:

- вероятность того, что на кубике выпало 6, если выпало четное число, равна $\frac{1}{6} / \frac{1}{2} = \frac{1}{3}$

- если вероятность того, что сегодня идет дождь и вы успеете на работу вовремя, равна 0,4, а вероятность того, что сегодня пойдет дождь, равна 0,5, то вероятность того, что вы придете на работу вовремя, при условии что идет дождь, равна $0,4/0,5 = 0,8$;
- в определенной стране в каждом штате разная доля водителей такси среди населения. Вероятность $p(x|y)$ того, что человек, случайно выбранный из определенного штата, является водителем такси, равна доле граждан страны, которые живут в этом штате и являются водителями такси $p(x, y)$, деленной на долю людей, живущих в этом штате $p(y)$;

Из таблицы сопряженности (рис. 8) можно вычислить условные вероятности. Например, вероятность присутствия кошки при наличии собаки равна 0,25 (присутствуют и кошка, и собака), деленным на $0,30 + 0,25 = 0,55$ (присутствие собаки).

Помните, что $p(X|Y)$ не равно $p(Y|X)$.

Если X и Y являются случайными величинами, распределение вероятностей X при заданном Y обозначается $p_{X|Y}(x|y)$ или $p(x|y)$, если нет неоднозначности в отношении переменных. Знак $|$ читается как «обусловленный» или «при условии что». Он определяется как вероятность пары x и y , деленная на вероятность y :

$$(40) p(x|y) = \frac{p(x, y)}{p(y)}$$

Упражнение В.9. Покажите, что $p(x|y)$ нормирована как функция от x . Приведите пример, показывающий, что $p(x|y)$ не нормируется как функция y .

Теперь объединим понятие маргинализации с определением условной вероятности.

Упражнение В.10. Покажите, что

$$(41) p(x) = \sum_y p(x|y)p(y)$$

Интегральная форма этого уравнения:

$$(42) p(x) = \int p(x|y)p(y)dy$$

Продолжая пример с таксистом: предположим, вас интересует вероятность того, что случайно выбранный гражданин страны является таксистом. Вы знаете долю таксистов в населении каждого штата. Вы также знаете долю всех граждан, проживающих в каждом штате. Чтобы получить ответ, вы перемножаете эти две доли для каждого штата, а затем суммируете по всем штатам.

Мы можем обусловить каждую вероятность в уравнениях (38) и (41) третьей случайной величиной z . Тогда мы получаем:

$$(43) p(x|z) = \sum_y p(x, y|z) = \sum_y p(x|y, z)p(y|z)$$

или в интегральной форме:

$$(44) p(x|z) = \int p(x, y|z)p(y|z)dy$$

Условные распределения не ограничиваются одной случайной величиной до и после знака вертикальной черты. Например, можно рассмотреть распределение X и Y при заданных Z и W , обозначаемое как $p(x, y|z, w)$.

Независимость

Две случайные величины X и Y называются независимыми, если их совместное распределение разлагается на маргиналы, т. е. если $p(x, y) = p(x)p(y)$ для всех x и y . Например, вероятность того, что на кубике выпадет 6 и на монете выпадет орел, является произведением вероятностей обоих событий, взятых по отдельности. Независимость можно изобразить графически (рис. 11). Можно реконструировать совместное распределение путем умножения маргиналов. Понятие независимости тесно связано с понятием корреляции: две независимые случайные величины также некоррелированы. Обратное неверно.



Рис. 11. Диаграмма Венна, изображающая независимость двух случайных величин

Упражнение В.11. Почему обратное неверно?

Упражнение В.12. Если X и Y независимы, что можно сказать об условных распределениях $p(x | y)$ и $p(y | x)$?

Если X , Y и Z обозначают три случайные величины, то X и Y условно независимы при заданном Z , если $p(x,y/z) = p(x/z)p(y/z)$ для любых значений x , y и z .

Правило Байеса

Правило Байеса связывает условные вероятности $p(x/y)$ и $p(y/x)$ друг с другом:

$$(45) p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Здесь $p(y/x)$ как функция x – функция правдоподобия по x , $p(x)$ – априорное распределение по x , а $p(x/y)$ – апостериорное распределение по x .

Доказательство. Из (40) мы знаем, что $p(x,y) = p(x/y)p(y)$. Меняя местами x и y , мы также получаем $p(y,x) = p(y/x)p(x)$. Совместная вероятность симметрична, т. е. $p(x,y) = p(y,x)$. Из этих трех уравнений следует, что $p(x/y)p(y) = p(y/x)p(x)$. Разделив обе части на $p(y)$, мы получаем правило Байеса.

Упражнение В.14. Докажите, что правая часть уравнения (В.54) нормирована по x .

 ChatGPT

Давайте докажем, что правая часть уравнения Байеса нормирована по событию A , то есть сумма вероятностей по всем возможным гипотезам A равна 1.

Имеем уравнение Байеса:

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

Теперь суммируем по всем возможным значениям гипотезы A :

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

Поскольку $P(B)$ не зависит от гипотезы A , мы можем вынести его за знак суммирования:

$$\sum_A P(A|B) = \frac{1}{P(B)} \sum_A P(B|A) \cdot P(A)$$

Теперь заметим, что сумма произведений $P(B|A) \cdot P(A)$ по всем гипотезам A равна вероятности $P(B)$:

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

Таким образом, у нас остается:

$$\sum_A P(A|B) = \frac{1}{P(B)} \cdot P(B)$$

Получаем:

$$\sum_A P(A|B) = 1$$

Таким образом, правая часть уравнения Байеса нормирована по событию A .

Рис. 12. Доказательство нормирования правой части уравнения Байеса от ChatGPT

Интерпретация диаграммой Венна правила Байеса для событий X и Y заключается в том, что площадь перекрытия может быть рассчитана двумя способами: как доля площади круга X , умноженная на площадь круга Y , или как доля площади круга Y , умноженная на площадь круга X . Поскольку результаты должны быть идентичными, это означает, что две дроби могут быть выражены друг через друга, если известно отношение площадей кругов X и Y .

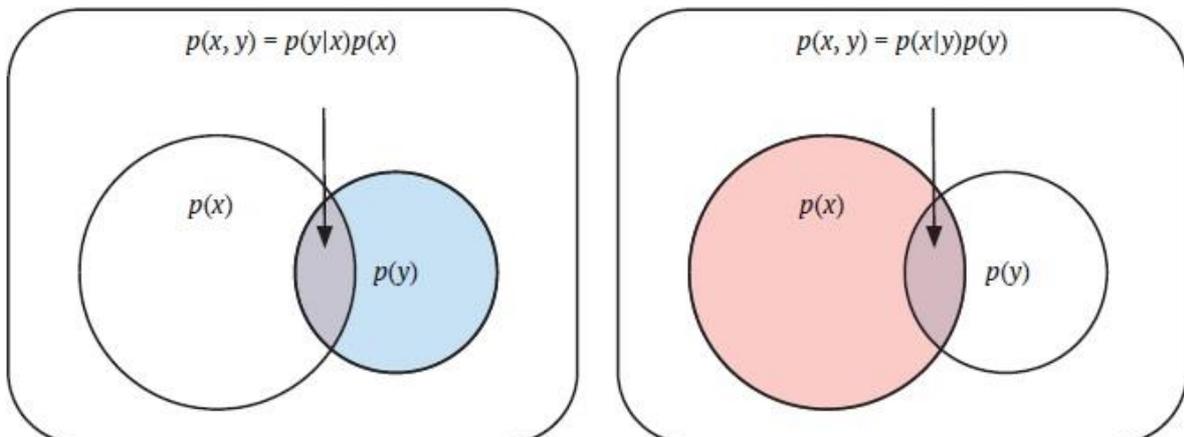


Рис. 13. Правило Байеса получается путем записи площади пересечения двумя разными способами и приравнивания двух выражений

Предположим, что 1 из 100 000 человек – профессиональный баскетболист, 1 из 100 человек – высокий и 95% профессиональных баскетболистов – высокие. Какова вероятность того, что высокий человек является профессиональным баскетболистом? Мы решаем эту задачу, применяя правило Байеса напрямую: если X – это «профессиональный баскетболист», а Y – «высокий рост», то $p(X) = 0,00001$, $p(Y) = 0,01$ и $p(Y|X) = 0,95$. Отсюда следует, что $p(X|Y) = 0,95 * 0,00001/0,01 = 0,0095$, или примерно 1 из 1000. Еще раз: среди всех людей профессиональный баскетболист встречается 1 из 100 000 человек, а среди высоких – 1 из 1000!

Упражнение В.15. Докажите другую форму правила Байеса:

$$(46) p(x|y) = \frac{p(y|x)p(x)}{\sum_x p(y|x)p(x)}$$

Последний способ записать правило Байеса - использовать знак пропорциональности: $p(x|y) \propto p(y|x)p(x)$.

Функции случайных величин

Функции одной переменной: изменение переменных

Если X – случайная величина с распределением вероятностей $p_X(x)$, а $Y = f(X)$ – новая случайная величина, полученная путем применения функции или преобразования f к X , каково распределение Y ? Например, X – равномерна распределена на $[0, 1]$. $Y = X^2$ – новая случайная величина, полученная путем возведения в квадрат результатов X . Каково распределение Y ?

Простой, но неправильный ответ заключается в том, что поскольку X следует равномерному распределению, то и Y тоже. Если число x лежит между 0 и 1, возведение его в квадрат всегда будет давать меньший результат. Таким образом, хотя значения Y также будут лежать между 0 и 1, более низкие значения в этом диапазоне будут иметь большую плотность вероятности, чем высокие значения. Рассмотрим интегральные функции распределения X и Y , которые обозначим $P_X(x)$ и $P_Y(y)$:

$$(47) P_Y(y) = p(Y \leq y) = p(X^2 \leq y) = p(X \leq \sqrt{y}) = P_X(\sqrt{y})$$

Воспользуемся тем фактом, что плотность распределения вероятности PDF, обозначенная здесь как $p_Y(y)$, является производной интегральной функции распределения CDF:

$$(48) p_Y(y) = \frac{dP_Y}{dy} = \frac{d}{dy} P_X(\sqrt{y}) \xrightarrow{\text{цепное правило}} \frac{dP_X}{dx} \Big|_{x=\sqrt{y}} \cdot \frac{d}{dy} \sqrt{y} = P_X(\sqrt{y}) \frac{1}{2\sqrt{y}} = 1 \cdot \frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}$$

Результирующее распределение $p_Y(y)$ нормировано (проверьте это) и соответствует нашему первоначальному предположению: плотность вероятности выше для более низких значений y . Проверим результат с помощью симуляции: возьмем множество случайных чисел из равномерного распределения между 0 и 1, возведем их в квадрат, и построим нормированную гистограмму квадратов с мелкими интервалами, а поверх нее проведем функцию (48):

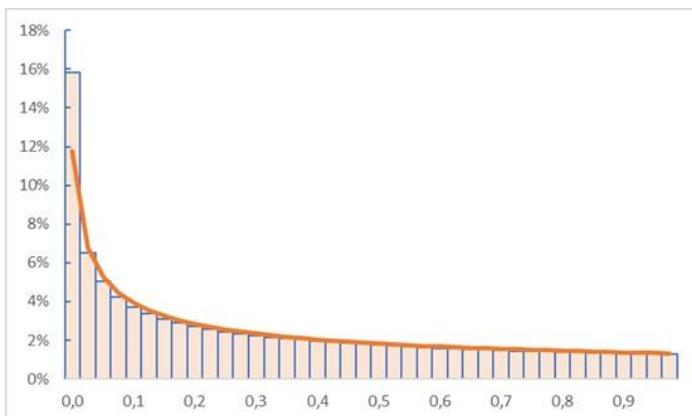


Рис. 14. Симуляция (гистограмма) и функция (48)

Мы могли бы сформулировать ту же задачу с произвольным распределением $p_X(x)$ вместо равномерного. Вычисления идентичны, за исключением последнего шага. Далее мы находим, что

$$(49) p_Y(y) = p_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}$$

Таким образом, распределение квадрата переменной является произведением исходного распределения, оцениваемого по значению x , которое отображается в y , $p_X(\sqrt{y})$, и дополнительного множителя. Дополнительный множитель, называемый якобианом, равен производной отображения от u к x . Было бы неправильно не учитывать якобиан и утверждать, что $p_Y(y) = p_X(\sqrt{y})$

Якобиан фигурирует не только в этом примере возведения в квадрат случайной величины, но и в нашей первоначальной общей задаче. Предположим, что X – случайная величина с распределением вероятностей $p_X(x)$ и $Y = f(X)$, где f – монотонно возрастающая функция. Каково распределение Y ? Сначала мы определим обратную функцию f^{-1} как функцию от y , которая «отменяет» действие f , другими словами, $f^{-1}(f(x)) = x$. Эта обратная функция определена вполне обоснованно, поскольку мы предполагали, что f монотонно возрастает. Заманчивым, но неправильным способом получения распределения Y было бы подставить обратную функцию в $p(X)$: $p_Y(y) = p(X)(f^{-1}(y))$ или предположить, что операция, применяемая к распределению, такая же, как и распределение, примененное к операции, $p_Y(y) = f(p_X(y))$. Правильный подход снова состоит в том, чтобы вычислить интегральное распределение y :

$$(50) P_Y(y) = p(Y \leq y) = p(f(X) \leq y) = p(X \leq f^{-1}(y)) = P_X(f^{-1}(y))$$

и, следовательно, функция плотности вероятности y :

$$(51) p_Y(y) = \frac{dP_Y}{dy} = \frac{d}{dy} P_X(f^{-1}(y)) \xrightarrow{\text{цепное правило}} \frac{dP_X}{dx} \Big|_{x=f^{-1}(y)} \cdot \frac{df^{-1}}{dy} = P_X(f^{-1}(y)) \frac{df^{-1}}{dy}$$

До сих пор мы рассматривали монотонно возрастающую функцию f . Когда f вместо этого монотонно убывает, окончательное выражение для $p_Y(y)$ приобретает дополнительный знак минус.

Упражнение В.16. Покажите это.

Пример с яблоками

В качестве иллюстрации процедуры замены переменных предположим, что вы собираетесь посетить яблоневый сад. Вы очень мало знаете о том, как быстро растут яблоки, или о продолжительности вегетационного периода в этом регионе, и вы не знаете, какие яблоки растут в саду. Если друг спросит вас, какого, по вашему мнению, размера будут яблоки в саду, вы можете сначала ответить, что понятия не имеете. После более обстоятельных размышлений и на основе ограниченных знаний о яблоках в целом у вас возникает убеждение в однородном априорном распределении плотности вероятности по диаметру яблок в саду от 3 до 13 см. Какова же тогда ваша первоначальная плотность распределения относительно объема яблока?

Прежде чем мы получим ответ, давайте обсудим задачу. Равномерное априорное распределение относительно диаметра яблока означает, что, например, вы считаете равновероятным, что диаметр яблока будет лежать как между 5 и 6 см, так и между 10 и 11 см. Если мы аппроксимируем яблоки сферами, то объем яблока $v = \pi w^3/6$, где w – диаметр яблока. Таким образом, объемы, соответствующие диаметрам 5, 6, 10 и 11 см, равны (с точностью до целого числа) 65, 113, 524 и 697 см³. Это означает, что вы считаете равновероятным (точнее, с 10%-ной вероятностью), что объем яблока находится между 65 и 113 см³ (диапазон 48 см³), как и между 524 и 697 см³ (диапазон 173 см³). Таким образом, априорное распределение плотность вероятности относительно объема яблок отнюдь не плоское. Наоборот, плотность вероятности будет выше при меньших объемах.

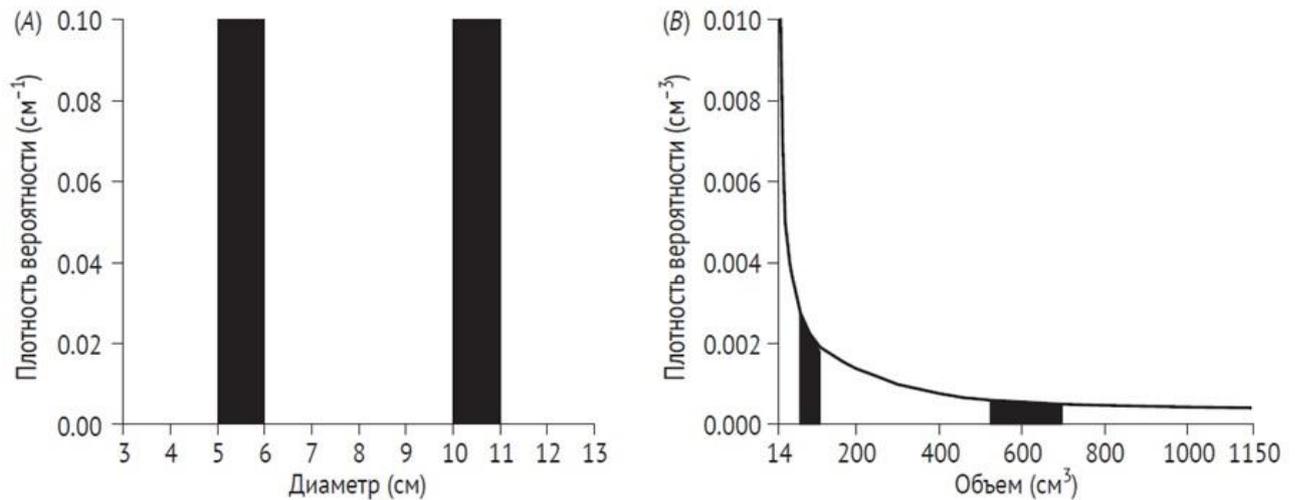


Рис. 15. Замена переменных: (А) равномерное априорное распределение вероятности для диаметра яблока от 3 до 13 см (диапазон 10 см). Априорное распределение плотности вероятности представляет собой линию высотой $0,1 \text{ см}^{-1}$, потому что общая площадь под плотностью должна быть равна 1. Вероятность того, что диаметр находится между 5 и 6 см, равна 0,1, как и вероятность того, что диаметр находится между 10 и 11 см (заполненные прямоугольные области); (В) априорное распределение плотности вероятности для объема яблока. Каждая закрашенная прямоугольная область – это вероятность того, что объем яблока находится в диапазоне, соответствующем диаметрам яблок, покрытым закрашенными прямоугольниками в А. Опять же, общая площадь под плотностью равна 1, а площадь каждого закрашенного прямоугольника равна 0,1 (обратите внимание на различия в масштабах по оси y)

Чтобы получить плотность вероятности по объему, мы отмечаем, что

$$(52) \quad w = \sqrt[3]{\frac{6v}{\pi}}$$

Для производной находим, что

$$(53) \quad \frac{dw}{dv} = \sqrt[3]{\frac{2}{9\pi v^2}}$$

Следовательно:

$$(54) \quad p_V(v) = p_W(w) \left| \frac{dw}{dv} \right| = p_W(w) \sqrt[3]{\frac{2}{9\pi v^2}} = \frac{1}{10 \text{ см}} \sqrt[3]{\frac{2}{9\pi v^2}}$$

Это кривая, изображенная на рис. 15В.

Обязательство неведения

Интересное последствие процедур замены переменных, проиллюстрированное примером с яблоневым садом, состоит в том, что невозможно быть в неведении относительно всех нюансов задачи. Например, невозможно абсолютно ничего не знать о размере яблока в его общепринятом определении. Как вы только что видели, если мы заявляем, что ничего не знаем о диаметре яблока в том смысле, что принимаем равновероятным широкий диапазон диаметров, этим допущением мы демонстрируем определенное априорное знание объема яблока!

При выполнении байесовского статистического анализа исследователь может постараться включить как можно меньше априорных убеждений в анализ, о котором, по его мнению, у него почти нет соответствующих фоновых знаний. Как лучше всего это сделать, если, указывая на свое незнание параметра, он тем самым демонстрирует знание о связанных параметрах? Например, если исследователь «не знает» о стандартном отклонении σ случайной величины, он может выбрать использование плоского априорного значения в очень широком диапазоне σ , но тогда он неявно задает неравномерное априорное распределение для дисперсии σ^2 . Поиск подходящих значений по

умолчанию или эталонных априорных распределений для таких ситуаций является интересной темой в области байесовского статистического анализа. В байесовских моделях восприятия и действия это обычно не проблема, поскольку предполагается, что априорное значение получено либо из экспериментальной, либо из естественной статистики.

Функции нескольких переменных

Предположим, вы бросаете две игральные кости и складываете результаты. Каково распределение вероятностей суммы? Простой подсчет дает ответ: исход 2 может быть достигнут только одним способом (1 + 1) и, следовательно, имеет вероятность $1/36$. Исход 3 может быть достигнут двумя способами (1 + 2 и 2 + 1) и, следовательно, имеет вероятность $2/36$ и т. д. Это приводит к следующему распределению вероятностей:

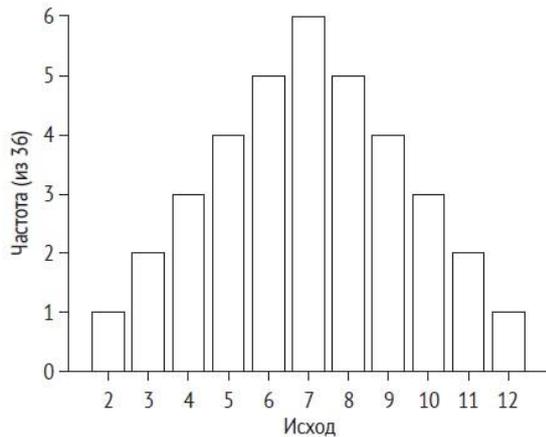


Рис. 16. Распределение вероятностей суммы двух бросков кости

Упражнение В.19. Если X и Y независимы и имеют равномерное распределение на $[0, 1]$, вычислите распределение $Z = X + Y$. Ответ – частный случай распределения [Ирвина-Холла](#).

Упражнение В.20. Если X и Y независимы и имеют нормальное распределение, докажите, что $Z = X + Y$ также имеет нормальное распределение.

Задачи

Задача В.1. В цветнике 20% цветов составляют тюльпаны. Из них четверть красные. Какова вероятность того, что случайно выбранным цветком в этом саду окажется тюльпан, но не красный? Хотя эта задача легко решается с помощью логических рассуждений, примените формулы правил вероятности.

Задача В.2. Четыре игрока, сидящих вокруг стола, собираются сыграть в игру. Чтобы определить, кто начинает, один игрок бросает две кости. Сумма двух чисел определяет, кто начинает, при этом отсчет идет по часовой стрелке, начиная с бросающего кубик, которому присвоен номер 1 (таким образом, при сумме 5 или 9 начинает бросающий). Какова вероятность начать игру для каждого игрока?

Задача В.3. Вы и я бросаем кубик по одному разу.

- Какова вероятность того, что у одного из нас выпадет 6, а у другого – нечетное число?
- Какова вероятность того, что хотя бы у одного из нас выпадет 6?
- Какова вероятность того, что у вас выпадет больше, чем у меня?
- Какова вероятность того, что наша сумма больше 8?

Задача В.4. Вы учитесь в группе из тридцати человек.

- Какова вероятность того, что день рождения конкретного одноклассника совпадает с вашим днем рождения?
- Какова вероятность того, что день рождения хоть кого-то из одноклассников совпадает с вашим днем рождения?
- Какова вероятность того, что у любых двух студентов совпадают дни рождения?

Задача В.6. Мы с вами поочередно подбрасываем монету. Вы начинаете. Если у вас выпадает решка, вы выигрываете. Если у вас выпадает орел, бросок переходит ко мне. Если у меня выпадает решка, я выигрываю. Если у меня выпадает орел, бросок снова переходит к вам. Игра продолжается до тех пор, пока один из нас не выиграет. Каков ваш шанс выиграть в этой игре?

Задача В.8. Эта задача связана с известным парадоксом Монти Холла. Вы участвуете в игровом шоу. Ведущий показывает вам три двери. За одной из них спрятан приз. Вы выбираете одну дверь (она остается закрытой). Ведущий, который знает, за какой дверью спрятан приз, открывает из оставшихся дверей ту, за которой нет приза. Затем ведущий дает вам возможность переключить свой выбор на оставшуюся неоткрытую дверь или остаться с вашим первоначальным выбором. Выбранная вами дверь открывается, и вы получаете приз, если он там есть.

- (a) Что нужно сделать, чтобы максимизировать вероятность получения приза?
- (b) Если есть N дверей и ведущий открывает m из них (где $m < n - 1$, какова вероятность получить приз при наилучшей стратегии?
- (c) Изменился бы ответ на вопрос (a), если бы ведущий не знал, в какой из двух оставшихся дверей находится приз, а в той, которую он открывает, случайно не оказалось приза? Обоснуйте свой ответ.
- (d) Подумайте, почему большинство людей считают, что не имеет значения, сохраните ли вы первоначальный выбор или выберете другую дверь.

Задача В.9. Функция плотности вероятности гауссовой случайной величины X задается уравнением 21. Покажите, что среднее значение и дисперсия этой случайной величины равны μ и σ^2 соответственно.

Задача В.10. Если X – экспоненциально распределенная случайная величина, где областью определения X является положительная вещественная числовая прямая, каковы область определения и распределение $Y = e^X$?

Задача В.11

- (a) Кто-то трижды подбрасывает честную монету. Вы наблюдаете за результатом одного броска, в котором выпадает орел. Какова вероятность того, что во всех трех подбрасываниях орел выпадет чаще, чем решка?
- (b) Кто-то подбрасывает честную монету N раз. Вы наблюдаете за результатом одного броска, в котором выпадает орел. Какова вероятность того, что во всех N подбрасываниях орел выпадет чаще, чем решка? (Для четного N интерпретируйте «больше» как «строго больше».)

Задача В.13. Если X и Y являются независимыми переменными со стандартным нормальным распределением, покажите, что случайная величина $Z = Y/X$ имеет распределение Коши, т. е.

$$(55) p_Z(z) = \frac{1}{\pi(1 + z^2)}$$

Задача В.14. Если X и Y независимы и имеют равномерное распределение вероятностей на $[0, 1]$ ($0 \leq x \leq 1$), покажите, что произведение случайных величин $Z = XY$ имеет распределение $p_Z(z) = -\log z$. Убедитесь, что это распределение нормировано, хотя плотность в точке 0 равна бесконечности. Этот пример показывает, как распределение произведения может сильно отличаться от распределения каждого из множителей.

Задача В.15. Возьмем два распределения фон Мизеса для одной и той же случайной величины, одно со средним значением μ_1 и параметром концентрации k_1 , другое со средним значением μ_2 и параметром концентрации k_2 . Покажите, что нормированное произведение этих распределений снова является распределением фон Мизеса, и вычислите его среднее значение и параметр концентрации.